A GENERAL COVARIANCE-BASED OPTIMIZATION FRAMEWORK USING ORTHOGONAL PROJECTIONS

Raphael Hunger, David A. Schmidt, Michael Joham, and Wolfgang Utschick

Associate Institute for Signal Processing
Technische Universität München, 80290 Munich, Germany
email: `hunger@tum.de`

ABSTRACT

We present a general framework for the minimization of a function which is parametrized by a set of covariance matrices over a constraint set. Since all covariance matrices have to obey the property of being positive semidefinite, this characteristic has to be reflected in the constraint set. In addition, the sum of all traces of the covariance matrices shall be upper bounded. Using a preconditioned gradient descent algorithm, we derive an orthogonal projection onto this constraint set in an easy to follow monolithic way such that it directly results from the definition of the projection. Interestingly, this projection allows for a descriptive water-spilling interpretation in the style of the well-known water-filling algorithm. Two possible applications are investigated: the sum mean-square-error minimization and the weighted sum-rate maximization for the MIMO broadcast channel. Simulations finally reveal the excellent performance of the proposed framework.

1. INTRODUCTION

Many optimizations arising in the context of information theory and signal processing feature the nice property that the precoding matrices always appear as outer products in the involved expressions. Congruously, all terms can be represented by functions of the *covariance* matrices. Prominent examples for such expressions are *mean-square-error* (MSE) terms (e.g., [1, 2, 3]), rate terms (e.g., [4, 5, 6, 7]), and signal-to-noise ratio terms (e.g., [8, 9]). For the optimization of any of such covariance-based objectives either in the MIMO *multiple access channel* (MAC) or in the MIMO *broadcast channel* (BC), we present a general framework consisting of three parts: an unconstrained gradient descent in combination with a proper step-size rule, a preconditioning alleviating the influence of different transmit powers, and an orthogonal projection which optimally maps the unconstrained gradient update back onto the set of feasible covariance matrices. Indeed, the final result of this mapping was also independently applied in [10]. However, our derivation of the optimum projection is based on an optimization following from the definition of the orthogonal projection. Moreover, it offers an easy to follow interpretation in the style of the well known water-filling algorithm.

Considering an arbitrary objective $\Psi(\cdot)$ depending on several covariance matrices, we first derive the three aforementioned parts our general framework consists of. In the sequel, we present a clearly laid out pseudo-code algorithm which can easily be implemented in Matlab for example. This algorithm acts as the composition of those three elements termed as the preconditioned projected gradient algorithm. In addition, if the objective is convex, convergence to the global optimum is ensured. Having derived our concept for general cost functions, we investigate its performance when applied to the sum-MSE minimization and the weighted sum-rate maximization. Both optimizations clearly fall in the category of convex covariance-based optimizations. Besides the observation of an extremely quick convergence, the simulations also confirm the conjecture that covariance-based optimizations outperform the respective precoder-based ones in terms of iterations until convergence.

2. COVARIANCE-BASED GRADIENT DESCENT

We focus on the minimization of a function $\Psi(\cdot)$ which is jointly convex in every of its arguments. In particular, we turn our attention to the case where the arguments resemble a set $\{\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K\}$ of K covariance matrices each of which has to fulfill the positive semidefiniteness constraint $\boldsymbol{Q}_k \succcurlyeq \mathbf{0}$, $\forall k\in\{1,\ldots,K\}$. Upper bounding the sum of the traces by the constant P_{tx}, the resulting optimization reads as

$$\underset{\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K}{\operatorname{minimize}}\ \Psi(\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K)\quad \text{s.t.: } \boldsymbol{Q}_k\succcurlyeq\mathbf{0}\ \forall k\in\{1,\ldots,K\},\quad \sum_{k=1}^{K}\operatorname{tr}(\boldsymbol{Q}_k)\le P_{\mathrm{tx}}. \tag{1}$$

In general, a closed form solution to above minimization (1) is not feasible thus necessitating the use of an iterative algorithm. In this paper we apply the *scaled projected gradient* algorithm [11] which performs a preconditioned gradient descent step followed by an *orthogonal* projection onto the constraint set. The gradient descent step for the covariance matrix $\boldsymbol{Q}_k\in\mathbb{C}^{r_k\times r_k}$ ignoring the constraints can be expressed as

$$\boldsymbol{Q}_k' = \boldsymbol{Q}_k - p\cdot s\cdot\frac{\partial\Psi(\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K)}{\partial\boldsymbol{Q}_k^{\mathrm{T}}}, \tag{2}$$

where $s\in\mathbb{R}_+$ denotes the (iteration-dependent) step-size. The iteration-dependent preconditioning scalar $p>0$ increases the speed of convergence by normalizing the sum of the gradient traces to the same order of magnitude as the transmit power: Thus, the gradient is almost independent of the current transmit signal-to-noise ratio.

$$p=\frac{P_{\mathrm{tx}}}{\left|\sum_{k=1}^{K}\operatorname{tr}\left[\frac{\partial\Psi(\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K)}{\partial\boldsymbol{Q}_k^{\mathrm{T}}}\right]\right|}.$$

For the computation of the Wirtinger derivatives, see [12]. Any objective $\Psi(\cdot)$ which we will investigate in the sequel has the property that the gradients with respect to any k are negative semidefinite:

$$\frac{\partial\Psi(\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K)}{\partial\boldsymbol{Q}_k^{\mathrm{T}}}\preccurlyeq\mathbf{0}\ \forall k. \tag{3}$$

As a consequence, the unconstrained gradient descent update in (2) yields positive semidefinite matrices $\boldsymbol{Q}_k' \succcurlyeq \mathbf{0}\ \forall k$ and the traces satisfy the inequality

$$\operatorname{tr}(\boldsymbol{Q}_k') \geq \operatorname{tr}(\boldsymbol{Q}_k)\ \forall k. \tag{4}$$

3. PROJECTION ONTO THE CONSTRAINT SET

Obviously, the updated temporary covariance matrices $\boldsymbol{Q}_1',\ldots,\boldsymbol{Q}_K'$ do not comply with the constraint set in (1). As a consequence, they have to be mapped to the constraint set $\mathcal{C}$ which is defined by

$$\mathcal{C} = \Big\{\boldsymbol{C}_1,\ldots,\boldsymbol{C}_K\ \Big|\ \boldsymbol{C}_k \succcurlyeq \mathbf{0}\ \forall k,\ \sum_{k=1}^{K}\operatorname{tr}(\boldsymbol{C}_k) \leq P_{\mathrm{tx}}\Big\}. \tag{5}$$

Due to the sum-trace constraint $\sum_{k=1}^K \operatorname{tr}(\boldsymbol{C}_k) \leq P_{\mathrm{tx}}$, all covariances are coupled. Hence, the nonlinear projection has to map all matrices $\boldsymbol{Q}_1',\ldots,\boldsymbol{Q}_K'$ simultaneously to the set $\mathcal{C}$ in (5). Composing the blockdiagonal matrix

$$\boldsymbol{Q}' = \operatorname{blockdiag}\left\{\boldsymbol{Q}_k'\right\}_{k=1}^K \in \mathbb{C}^{R\times R},$$

where $R=\sum_{k=1}^K r_k$, the orthogonal projection of $\boldsymbol{Q}'$ onto $\mathcal{C}$ yielding the blockdiagonal covariance matrix $\boldsymbol{C}$ is achieved by the operation $\boldsymbol{C} = (\boldsymbol{Q}')_\perp$.

3.1. The Naive Projection Approach

In our precoder-based projection algorithm in [13], the objective is a function depending on the *precoders* $\boldsymbol{T}_1,\ldots\boldsymbol{T}_K$ instead of the covariance matrices. There, the constraint set $\sum_{k=1}^K \|\boldsymbol{T}_k\|_{\mathrm{F}}^2 \leq P_{\mathrm{tx}}$ is simply a ball, as the covariances $\boldsymbol{Q}_k = \boldsymbol{T}_k\boldsymbol{T}_k^{\mathrm{H}}$ are positive semidefinite by construction. The *orthogonal* projection onto the sphere is obtained by a common scaling of all precoders such that the sum-power constraint is met with equality. Since we are working with the covariances, the constraint set (5) does not simply represent a ball any longer. But due to the fact that the unprojected gradient update $\boldsymbol{Q}_k'$ in (2) already fulfills the positive semidefiniteness constraint $\boldsymbol{Q}_k' \succcurlyeq \mathbf{0}\ \forall k$ (cf. Eq. 3), one might think of scaling all covariance matrices according to

$$\boldsymbol{C} = (\boldsymbol{Q}')_\perp = \frac{P_{\mathrm{tx}}}{\operatorname{tr}(\boldsymbol{Q}')}\boldsymbol{Q}' = \frac{P_{\mathrm{tx}}}{\sum_{k=1}^K \operatorname{tr}(\boldsymbol{Q}_k')}\boldsymbol{Q}'$$

in order to let $\boldsymbol{C} = (\boldsymbol{Q}')_\perp \in \mathcal{C}$ hold, similar to our precoder-based design in [13]. However, this kind of projection is *not* orthogonal, and the scaled projected gradient algorithm employing this kind of naive projection inherited from the precoder-based approach fails to converge to a point that fulfills the KKT optimality conditions of (1).

3.2. The Orthogonal Projection Approach

In case of an *orthogonal* projection and by means of a step-size adaptation convergence of the iterative algorithm can be ensured, cf. [11]. The orthogonal projection of $\boldsymbol{Q}'$ onto the constraint set $\mathcal{C}$ in (5) minimizes the distance between $\boldsymbol{Q}'$ and $\boldsymbol{C} \in \mathcal{C}$ [11, 14]. Here, we use the Frobenius norm as a distance measure. What follows is a solid and easy-to-follow derivation of the orthogonal projection [14]

$$\boldsymbol{C} = \left(\boldsymbol{Q}'\right)_\perp = \boldsymbol{Q}' + \boldsymbol{E}_\perp, \tag{6}$$

where $\boldsymbol{E}_\perp$ is found via

$$\boldsymbol{E}_\perp = \operatorname*{argmin}_{\boldsymbol{E}} \|\boldsymbol{E}\|_{\mathrm{F}}^2 \ \text{ s.t.: }\ \operatorname{tr}(\boldsymbol{Q}'+\boldsymbol{E}) \leq P_{\mathrm{tx}},\ \ \boldsymbol{Q}'+\boldsymbol{E} \succcurlyeq \mathbf{0}. \tag{7}$$

Note that the blockdiagonal matrix $\boldsymbol{Q}'$ is positive semidefinite and satisfies $\operatorname{tr}(\boldsymbol{Q}') > \operatorname{tr}(\boldsymbol{Q}) = P_{\mathrm{tx}}$ according to (4). Additionally, the gradient $\partial\Psi(\cdot)/\partial\boldsymbol{Q}_k^{\mathrm{T}}$ does not vanish in the optimum, i.e., not all inequalities in (4) can hold with equality. This implies a blockdiagonal structure of $\boldsymbol{E}_\perp$ with $\boldsymbol{E}_\perp$ being Hermitian and negative semidefinite. Therefore, we can ignore the blockdiagonal structure for the derivation in the following without loss of generality.

Theorem 1. *The orthogonal projection of the matrix $\boldsymbol{Q}'$ with eigenvalue decomposition $\boldsymbol{Q}' = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ onto the constraint set $\mathcal{C}$ reads as $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{U}$ is the unitary eigenbasis of $\boldsymbol{Q}'$ and $\boldsymbol{\Lambda}$ contains the eigenvalues $\lambda_1,\ldots,\lambda_R$. The diagonal entries of the diagonal matrix $\boldsymbol{D}$ are $d_i = [\lambda_i - \mu]_+$, and the spilling level μ follows from $\sum_{i=1}^R [\lambda_i - \mu]_+ = P_{\mathrm{tx}}$, where the operator $[\cdot]_+$ is defined by $[\cdot]_+ = \max(0,\cdot)$.*

Proof. Assigning the Lagrangian function

$$\mathcal{L} = \operatorname{tr}(\boldsymbol{E}\boldsymbol{E}^{\mathrm{H}}) + 2\mu[\operatorname{tr}(\boldsymbol{Q}'+\boldsymbol{E}) - P_{\mathrm{tx}}] - 2\operatorname{tr}[\boldsymbol{S}(\boldsymbol{Q}'+\boldsymbol{E})] \tag{8}$$

to the minimization in (7), the Lagrangian multipliers feature the properties $\mu \geq 0$ and $\boldsymbol{S} \succcurlyeq \mathbf{0}$. From the derivative of (8) with respect to $\boldsymbol{E}^{\mathrm{T}}$, we find

$$\boldsymbol{E}_\perp = \boldsymbol{S} - \mu\mathbf{I}_R. \tag{9}$$

Furthermore, the second KKT condition $\boldsymbol{S}(\boldsymbol{Q}'+\boldsymbol{E}_\perp) = \mathbf{0}$ first implies that

$$\boldsymbol{S}\boldsymbol{Q}' + \boldsymbol{S}^2 - \mu\boldsymbol{S} = \mathbf{0},$$

and second, that $\boldsymbol{S}$ and $\boldsymbol{Q}'$ must have the same eigenbasis. Assuming the eigenvalue decompositions $\boldsymbol{Q}' = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ and $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}}$, where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ are diagonal with non-negative diagonal entries, this leads to

$$\boldsymbol{\Sigma}\boldsymbol{\Lambda} + \boldsymbol{\Sigma}^2 - \mu\boldsymbol{\Sigma} = \mathbf{0}.$$

In scalar form, these R equations read as

$$\sigma_i^2 + \sigma_i(\lambda_i - \mu) = 0,\ i \in \{1,\ldots,R\},$$

having the two possible solutions $\sigma_i = 0$ or $\sigma_i = \mu - \lambda_i$ with $\mu_i > \lambda_i$ to ensure $\boldsymbol{\Sigma} \succcurlyeq \mathbf{0}$. A compact notation is therefore

$$\sigma_i = [\mu - \lambda_i]_+.$$

Hence, the summand $\boldsymbol{E}_\perp$ generating the matrix $\boldsymbol{C}$ according to (6) reads as (cf. Eq. 9)

$$\boldsymbol{E}_\perp = \boldsymbol{U}\left[\mu\mathbf{I}_R - \boldsymbol{\Lambda}\right]_+ \boldsymbol{U}^{\mathrm{H}} - \mu\mathbf{I}_R,$$

and the projected matrix $\boldsymbol{C} = \boldsymbol{Q}' + \boldsymbol{E}_\perp = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$ features the eigenvalues

$$d_i = \lambda_i + \sigma_i - \mu = \lambda_i + [\mu - \lambda_i]_+ - \mu = [\lambda_i - \mu]_+. \tag{10}$$

Summing up, the orthogonally projected matrix $\boldsymbol{C}$ reads as

$$\boldsymbol{C} = \boldsymbol{U}[\boldsymbol{\Lambda} - \mu\mathbf{I}_R]_+\boldsymbol{U}^{\mathrm{H}}. \tag{11}$$

□

It remains to determine the *spilling level* μ.

Corollary 1. *The spilling level is $\mu = \frac{1}{L}(\sum_{i=1}^L \lambda_i - P_{\mathrm{tx}})$ if all λ_i are sorted non-increasingly. The number of active streams L is found by initializing L with $R = \sum_{k=1}^K r_k$ and checking afterwards, if the termination criterion $L\lambda_L - \sum_{i=1}^L \lambda_i + P_{\mathrm{tx}} > 0$ holds. If so, the optimum L has been found, otherwise, L is repeatedly reduced by one until the termination criterion is met.*

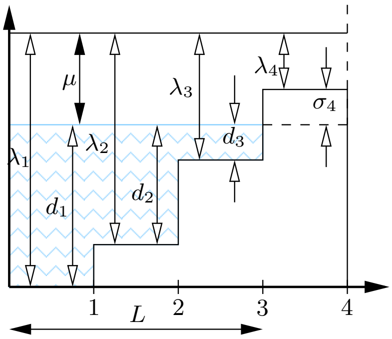

Fig. 1. Water-spilling interpretation of the projection.

Proof. The number L of active streams after the projection is defined by $\lambda_L - \mu(L) > 0$ and $\lambda_{L+1} - \mu(L) \leq 0$. Note that $\mu(L) = 0$ would imply $\boldsymbol{\Sigma} = \mathbf{0}$, $\boldsymbol{E} = \mathbf{0}$, and therefore $\boldsymbol{Q} = \boldsymbol{Q}'$. But since $\mathrm{tr}(\boldsymbol{Q}') > P_{\mathrm{tx}}$, $\mu(L) = 0$ is not possible and the complementary slackness condition says that the full power P_{tx} has to be consumed by $\boldsymbol{Q}$. Then, we have $\mu(L) = (\sum_{i=1}^{L} \lambda_i - P_{\mathrm{tx}})/L$. In order to ensure that $\lambda_L - \mu(L) > 0$, we start with $L = R$ and check if this inequality holds. If not, L is decreased until $\lambda_L - \mu(L) > 0$ holds. Otherwise, the optimum L has been found because $\lambda_{L+1} - \mu(L) \leq 0$ is automatically fulfilled since the inequality $\lambda_{L+1} - \mu(L+1) > 0$ was not fulfilled in the previous iteration. □

Note that the solution in (11) represents the unique global optimum of the projection. This follows from the closest point theorem due to the convexity of the set $\mathcal{C}$. The same result was independently found in [10] almost at the same time. However, we derive the orthogonal projection *monolithically* as a requirement for the projected gradient algorithm. In contrast to [10], we come up with a compact and easy to follow derivation directly following from the KKT conditions of the orthogonal projection.

3.3. Water-Spilling Interpretation of the Projection

Due to the similarity of Eq. (10) and the standard water-filling solution, we present a graphical interpretation for the orthogonal projection. Now, the eigenvalues λ_i of $\boldsymbol{Q}'$ represent the powers of the individual modes of the *unprojected* gradient update (2). In sum, they correspond to the amount of water before the projection. Having applied the projection, the energies d_i of the modes follow from spilling water with the level μ, which is illustrated in Figure 1, such that the total water mass $\sum_{i=1}^{L} d_i$ equals P_{tx}.

4. THE COMPLETE COVARIANCE-BASED PROJECTED GRADIENT ALGORITHM

Combining the unconstrained preconditioned gradient descent in (2) and the orthogonal projection from Section 3.2, we end up at the covariance-based preconditioned projected gradient algorithm, which is depicted in Alg. 1 in detail. Given an accuracy ε and a transmit power P_{tx}, the algorithm starts with the initialization of all covariance matrices $\boldsymbol{Q}_k$ $\forall k$ with scaled identities in Line 1. Having set up the inverse step-size d, the current objective under white signaling is evaluated in Line 3. Lines 5 to 18 correspond to one outer iteration. After the computation of the gradient in Line 5, the preconditioning scalar p which reduces the sensitivity to the transmit power P_{tx} is determined in Line 6. Lines 8 to 15 are executed

Algorithm 1 Covariance-based preconditioned projected gradient algorithm for a general cost function $\Psi(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K)$.

Require: Accuracy ε, transmit power P_{tx}
1: $\boldsymbol{Q}_k \leftarrow \frac{P_{\mathrm{tx}}}{\sum_{k=1}^{K} r_k} \mathbf{I}_{r_k}$ $\forall k$ *initialize all covariance matrices*
2: $d \leftarrow 1$ *initialize inverse step-size*
3: old_cost $\leftarrow \Psi(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K)$ *evaluate objective*
4: **repeat**
5: $\quad \boldsymbol{G}_k \leftarrow \frac{\partial \Psi(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K)}{\partial \boldsymbol{Q}_k^{\mathrm{T}}}$ $\forall k$ *gradient computation for all users*
6: $\quad p \leftarrow \frac{P_{\mathrm{tx}}}{\sum_{k=1}^{K} \mathrm{tr}(\boldsymbol{G}_k)}$ *preconditioning scalar computation*
7: $\quad$ **repeat**
8: $\quad\quad s \leftarrow \frac{1}{d}$ *set step-size*
9: $\quad\quad \boldsymbol{Q}_k' \leftarrow \boldsymbol{Q}_k - p \cdot s \cdot \boldsymbol{G}_k$ *unconstr. gradient update (2)*
10: $\quad\quad \boldsymbol{C} \leftarrow (\boldsymbol{Q}')_{\perp}$ *simult. projection of blockdiagonal $\boldsymbol{Q}'$ (11)*
11: $\quad\quad$ new_cost $\leftarrow \Psi(\boldsymbol{C}_1, \ldots, \boldsymbol{C}_K)$ *evaluate objective*
12: $\quad\quad$ cost_reduction $\leftarrow$ old_cost $-$ new_cost
13: $\quad\quad$ **if** cost_reduction ≤ 0 **then**
14: $\quad\quad\quad d \leftarrow d + 1$ *decrease step-size*
15: $\quad\quad$ **end if**
16: $\quad$ **until** cost_reduction > 0
17: $\quad \boldsymbol{Q}_k \leftarrow \boldsymbol{C}_k$ $\forall k$ *save new covariances*
18: $\quad$ old_cost $\leftarrow$ new_cost *save new objective*
19: **until** cost_reduction $\leq \varepsilon$

until the projected gradient descent leads to a reduction of the objective. First, the unconstrained gradient update according to (2) is performed in Line 9. Then, the result is simultaneously projected onto the convex set $\mathcal{C}$ from (5) via (11). If the resulting covariance matrices $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_K$ bring a reduction of the cost function, these covariances are saved in Line 17. Otherwise, the step-size is reduced in Line 14 by increasing d by one.

5. APPLICATIONS SUITED FOR THE COVARIANCE-BASED PROJECTED GRADIENT ALGORITHM

Having derived the covariance-based gradient descent step in combination with the orthogonal projection onto the convex constraint set, we point out two well known optimizations in the MIMO broadcast channel which can be solved using above general framework: sum-MSE minimization with linear pre- and decoding and weighted sum-rate maximization utilizing nonlinear dirty paper coding [15].

5.1. Sum-MSE Minimization

Based on the MSE duality results in [16, 17, 18] stating that the MIMO BC and the MIMO MAC feature the same MSE region under a sum-power constraint with linear filtering, any MSE-based optimization in the BC can conveniently be solved in the dual MAC and afterwards be transformed back to the BC. Thus, we can exploit the hidden convexity [16] of the BC sum-MSE minimization and solve the equivalent optimization in the uplink, where it turns out to be convex. Converting the solution of the dual MAC problem to the downlink BC turns out to be very easy entailing a very low complexity since only a single scalar has to be computed, cf. [16]. Assuming MMSE receivers in the dual MAC, the sum-MSE cost function $\Psi(\cdot)$ reads as

$$\Psi(\boldsymbol{Q}_1, \ldots, \boldsymbol{Q}_K) = \mathrm{tr}\left[\left(\mathbf{I}_N + \sigma_\eta^{-2} \sum_{\ell=1}^{K} \boldsymbol{H}_\ell \boldsymbol{Q}_\ell \boldsymbol{H}_\ell^{\mathrm{H}}\right)^{-1}\right] + R - N,$$

with $\boldsymbol{H}_k \in \mathbb{C}^{N\times r_k}$ denoting the channel matrix describing the transmission from user k to the base station, N refers to the number of antennas at the base station, and σ_η^2 is the noise variance at each receive antenna. Above objective involves the Wirtinger derivatives

$$\frac{\partial \Psi(\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K)}{\partial \boldsymbol{Q}_k^{\mathrm{T}}} = -\sigma_\eta^{-2}\boldsymbol{H}_k^{\mathrm{H}}\Big(\mathbf{I}_N+\sigma_\eta^{-2}\sum_{\ell=1}^{K}\boldsymbol{H}_\ell\boldsymbol{Q}_\ell\boldsymbol{H}_\ell^{\mathrm{H}}\Big)^{-2}\boldsymbol{H}_k$$

which obviously satisfy (3) such that the orthogonal projection described in Section 3.2 is optimum. If the gradient update (2) yielded an indefinite matrix $\boldsymbol{Q}_k'$, the projection would look different. Because of the joint convexity of the cost function with respect to the covariance matrices $\boldsymbol{Q}_1,\ldots,\boldsymbol{Q}_K$ and due to the convexity of the constraint set, the algorithm described in Section 4 converges to the global optimum.

5.2. Weighted Sum-Rate Maximization

Similar to the MSE regions for linear filtering in the previous section, the capacity regions of the BC and the MAC obtained by nonlinear processing exactly coincide. The linking element between uplink and downlink is again resembled by duality transformations [5]. Every rate tuple feasible in the BC can also be achieved in the MAC and vice versa, and the conversion from one domain to the other is obtained by means of singular-value-decomposition-based transformations resulting from the effective/flipped channel framework. In turn, this duality converts the weighted sum-rate to a concave function in the dual MAC. There, the rate of user $\pi[k]$ reads as (cf. [4])

$$R_{\pi[k]} = \log_2\frac{\left|\mathbf{I}_N+\sigma_\eta^{-2}\sum_{\ell=1}^{k}\boldsymbol{H}_{\pi[\ell]}\boldsymbol{Q}_{\pi[\ell]}\boldsymbol{H}_{\pi[\ell]}^{\mathrm{H}}\right|}{\left|\mathbf{I}_N+\sigma_\eta^{-2}\sum_{\ell=1}^{k-1}\boldsymbol{H}_{\pi[\ell]}\boldsymbol{Q}_{\pi[\ell]}\boldsymbol{H}_{\pi[\ell]}^{\mathrm{H}}\right|}, \qquad (12)$$

where $\pi[1]$ is the index of the user who is decoded last. From the polymatroidal structure of the capacity region it follows that the decoding order $\pi[\cdot]$ is optimal, if the rate weights $w_1,\ldots,w_K$ are sorted in a nonincreasing order [19], i.e., $w_{\pi[1]} \geq w_{\pi[2]} \geq \ldots \geq w_{\pi[K]}$. W.l.o.g. we assume in the following a renumbering of the users for a clearer notation such that $w_1 \geq w_2 \geq \ldots \geq w_K$ and $\pi[\cdot]$ becomes the identity mapping. Let $\alpha_k = w_k - w_{k+1}$ for $k \in \{1,\ldots,K-1\}$ and $\alpha_K = w_K$. The cost function $\Psi(\cdot)$ corresponding to the negative weighted sum-rate $-\sum_{k=1}^{K} w_k R_k$ then reads as (cf. Eq. 12)

$$\Psi(\boldsymbol{Q}_1,\ldots\boldsymbol{Q}_K) = -\sum_{k=1}^{K}\alpha_k\log_2\left|\mathbf{I}_N+\sigma_\eta^{-2}\sum_{\ell=1}^{k}\boldsymbol{H}_\ell\boldsymbol{Q}_\ell\boldsymbol{H}_\ell^{\mathrm{H}}\right|. \quad (13)$$

The Wirtinger derivatives can be computed via

$$\frac{\partial \Psi(\boldsymbol{Q})}{\partial \boldsymbol{Q}_k^{\mathrm{T}}} = -\frac{1}{\ln 2}\sum_{i=k}^{K}\alpha_i\boldsymbol{H}_k^{\mathrm{H}}\Big(\mathbf{I}_N+\sigma_\eta^{-2}\sum_{\ell=1}^{i}\boldsymbol{H}_\ell\boldsymbol{Q}_\ell\boldsymbol{H}_\ell^{\mathrm{H}}\Big)^{-1}\boldsymbol{H}_k.$$

Again, (3) holds and due to the convexity of (13), the iterative algorithm in Section 4 reaches the global optimum.

6. SIMULATION RESULTS

6.1. Sum-MSE Minimization

We investigate the number of iterations which are required to let different iterative algorithms targeted at minimizing the sum-MSE achieve an MSE which is smaller than $(1+\varepsilon)$ times the total MMSE with an accuracy of $\varepsilon = 10^{-4}$. The total MMSE is assumed to be reached after 100 iterations of our proposed algorithm since our algorithm has converged then long before. Fig. 2 shows the number of iterations to reach the MMSE up to a fraction of ε versus the relative frequency for three different algorithms. For this setup, $K = 4$ two-antenna users ($r_k = 2\ \forall k$) are served by a base station with $N = 4$ antennas, and the transmit power is set to $P_{\mathrm{tx}} = 10$, whereas the noise variance is $\sigma_\eta^2 = 1$. Moreover, we averaged over 10000 i.i.d. channel realizations where each entry of $\boldsymbol{H}_k$ has a zero-mean complex Gaussian distribution with variance one.

The alternating optimization approach in [20] optimizes the transmit and receive filters in an alternating fashion directly in the downlink and is resembled by the red bars. Unfortunately, this algorithm is very sensitive to the ratio $P_{\mathrm{tx}}/\sigma_\eta^2$ and a higher ratio results in a slower speed of convergence. As a consequence, the relative frequencies have a nonzero support between 20 and 200 approximately, and the average number of iterations to achieve the target MSE is around 75. Increasing the number of antennas N at the base station leads to a reduced number of iterations. In [17] the authors present a projected gradient descent algorithm working on the *precoders*. Thus, the orthogonal projection onto the sum-power constraint is simply a rescaling of the matrices obtained by the unconstrained gradient descent. The corresponding green bars show a much denser histogram at much fewer iterations yielding an average number of iterations around 16. Finally, our proposed approach performing the covariance-based projection is depicted by the blue curve. The average number of iterations necessary to converge within a tolerance of ε reduces to about 6, i.e., less than half of the number of iterations are required compared to the precoder-based approach in [17], and less than ten percent compared to the alternating optimization in [20]. Note that the last two algorithms have been initialized with scaled identity covariance matrices in the uplink and solve the optimization in this dual MAC whereas the alternating optimization approach directly operates in the downlink. There, the precoders have been initialized such that they correspond to a white power allocation in the uplink, i.e., according to [16], the precoders are scaled Hermitian MMSE receivers of the downlink. This initialization yields much better results than a white power allocation in the downlink.

6.2. Weighted Sum-Rate Maximization

We measure the number of iterations that are needed in order to let the weighted sum-rate grow above $(1-\varepsilon)$ times the maximum weighted sum-rate. This time, the relative accuracy $\varepsilon = 10^{-3}$ is slightly larger than before due to the fact that the algorithm in [21] takes an enormous amount of time to reach this target rate. All other parameters are left unchanged, and the weight vector $\boldsymbol{w}$ is arbitrarily set to $\boldsymbol{w} = [1,2,3,4]^{\mathrm{T}}$. We compare four different algorithms: The covariance-based rank-one covariance matrix update in [21], which can be regarded as the first one tackling the problem of maximizing the weighted sum-rate. Second, the precoder-based conjugate gradient approach in [22], third, the covariance-based conjugate gradient approach from Liu et al. in [10], and finally, our proposed method which is also mentioned in [22] to some extent.

The right plot of Fig. 3 shows that the algorithm in [21] converges extremely slowly. Since the covariance matrix updates have only rank one, an average number of about 450 iterations is necessary to reach the desired target rate. Moreover, every single iteration has a critically high complexity since many function evaluations have to be executed for the one-dimensional bisection which determines the optimum step-size. The most recent work from Liu et al. in [10] combines a conjugate gradient ascent with an Armijo rule for the *outer* step-size. However, convergence cannot be guaranteed this

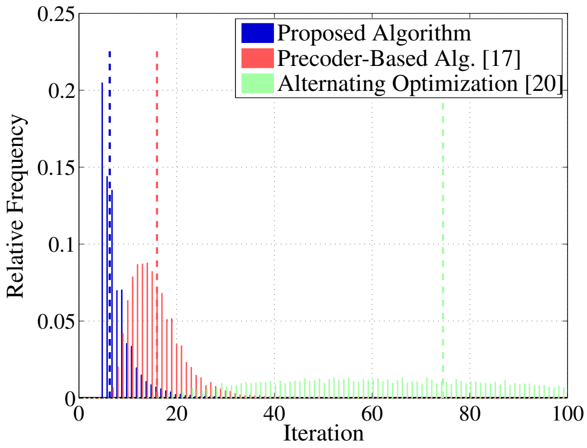

Fig. 2. Relative frequency of the number of iterations required to achieve an MSE smaller than $(1+\varepsilon)$ times the MMSE, where $\varepsilon = 10^{-4}$. The mean values correspond to the dashed vertical lines.

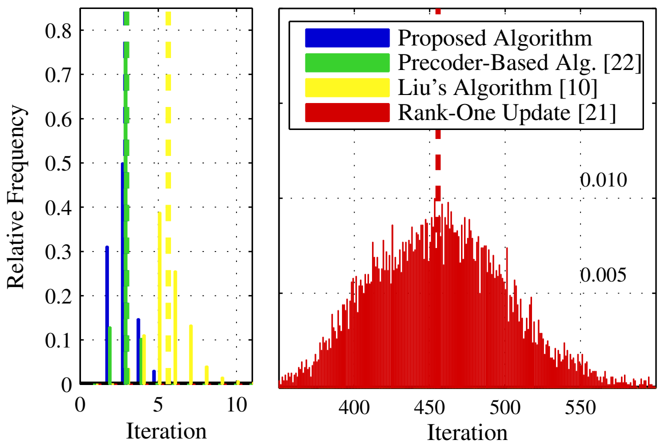

Fig. 3. Relative frequency of the number of iterations required to achieve a weighted sum-rate larger than $(1-\varepsilon)$ times the maximum weighted sum-rate, where $\varepsilon = 10^{-3}$.

way since this kind of step-size rule should be applied to the inner step-size which is held constant in [10]. Its performance is depicted in the left plot of Fig. 3, yielding an average number of around 6 iterations to reach the target rate. Both the conjugate gradient precoder based approach from Böhnke et al. in [22] and our proposed covariance-based algorithm obtain an even smaller average number of iterations, which is about 3. The latter one needs only two iterations in 30 percent of the cases whereas the first one features a smaller standard deviation and requires only 3 iterations in about 75 percent of the cases.

7. REFERENCES

[1] S. Serbetli and A. Yener, "Transceiver Optimization for Multiuser MIMO Systems," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 214–226, January 2004.

[2] A.J. Tenenbaum and R.S. Adve, "Joint Multiuser Transmit-Receive Optimization Using Linear Processing," in *IEEE International Conference on Communications*, Paris, France, June 2004, vol. 1, pp. 588–592.

[3] M. Codreanu, A. Tolli, M. Juntti, and M. Latva-aho, "Weighted Sum MSE Minimization for MIMO Broadcast Channel," in *17th International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, September 2006.

[4] W. Yu, W. Rhee, S. Boyd, and J.M. Cioffi, "Iterative Water-Filling for Gaussian Vector Multiple-Access Channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, January 2004.

[5] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of MIMO Broadcast Channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2658–2668, October 2003.

[6] J. Lee and N. Jindal, "Symmetric Capacity of MIMO Downlink Channels," in *IEEE International Symposium on Information Theory (ISIT)*, July 2006, pp. 1031–1035.

[7] W. Yu and J. M. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels.," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, September 2004.

[8] R. Doostnejad, T. J. Lim, and E. Sousa, "Precoding for the MIMO Broadcast Channels with Multiple Antennas at Each Receiver," in *in Proc. of The Johns Hopkins University Conf. on Information Sciences and Systems*, Baltimore, MD, Mar. 2005.

[9] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear recoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, January 2006.

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

[11] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computations*, Prentice-Hall, 1989.

[12] R. Hunger, "An Introduction to Complex Differentials and Complex Differentiability," Tech. Rep. TUM-LNS-TR-07-06, Technische Universität München, October 2007.

[13] R. Hunger, D. A. Schmidt, and W. Utschick, "Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, June 2007.

[14] N. J. Higham, "Matrix Nearness Problems and Applications," in *Applications of Matrix Theory*, M. J. C. Gover and S. Barnett, Eds. 1989, pp. 1–27, Oxford University Press.

[15] M. Costa, "Writing on Dirty Paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[16] R. Hunger, M. Joham, and W. Utschick, "On the MSE-Duality of the Broadcast Channel and the Multiple Access Channel," *Submitted to IEEE Transactions on Signal Processing*.

[17] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver Design for Multi-User MIMO Systems," in *Proc. ITG/IEEE WSA 2006*, March 2006.

[18] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE Transceiver Optimization for Multiuser MIMO Systems: Duality and Sum-MSE Minimization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5436–5446, November 2007.

[19] D. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov 1998.

[20] R. Hunger, W. Utschick, D. A. Schmidt, and M. Joham, "Alternating Optimization for MMSE Broadcasting," in *Proc. ICASSP 2006*, May 2006, vol. IV, pp. IV–757–IV–760.

[21] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink Capacity Evaluation of Cellular Networks With Known-Interference Cancellation," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 802–811, June 2003.

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