Title	Efficient Zero-Forcing Precoder Design for Weighted Sum-Rate Maximization With Per-Antenna Power Constraint
Authors(s)	Pham, Thuy M.; Farrell, Ronan; Dooley, John; Tran, Le-Nam; et al.
Publication date	2017-11-22
Publication information	IEEE Transactions on Vehicular Technology, 67 (4): 3640-3645
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/10388
Publisher's statement	© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/TVT.2017.2776356

Efficient Zero-forcing Precoder Design for Weighted Sum-rate Maximization with Per-antenna Power Constraint

Thuy M. Pham, Ronan Farrell, John Dooley, Eryk Dutkiewicz, Diep N. Nguyen, and Le-Nam Tran

***Abstract*—This paper proposes an efficient (semi-closed-form) zero-forcing (ZF) precoder design for the weighted sum-rate maximization problem under per-antenna power constraint (PAPC). Existing approaches for this problem are based on either interior-point methods that do not favorably scale with the problem size or subgradient methods that are widely known to converge slowly. To address these shortcomings, our proposed method is derived from three elements: minimax duality, alternating optimization (AO), and successive convex approximation (SCA). Specifically, the minimax duality is invoked to transform the considered problem into an equivalent minimax problem, for which we then recruit AO and SCA to find a saddle-point. That enables us to take advantages of closed-form expressions and hence achieve fast convergence rate. Moreover, the complexity of the proposed method scales linearly with the number of users, compared to cubically for the standard interior-point methods. We provide an analytical proof for the convergence of the proposed method and numerical results to demonstrate its superior performance over existing approaches. Our proposed method offers a powerful tool to characterize the achievable rate region of ZF schemes under PAPC for massive MIMO.**

***Index Terms*—MIMO, zero-forcing, alternating optimization, minimax duality, closed-form, successive convex approximation.**

I. Introduction

For a Gaussian multiple-input multiple-output (MIMO) broadcast channel (BC), dirty paper coding (DPC) was proved to achieve the full capacity region [1]. However, it is very challenging to apply DPC in reality due to its nonlinear encoding and decoding nature. As a result, linear precoding strategies such as block diagonalization (BD) or zero-forcing (ZF) [2], [3] which can strike a good balance between the achievable rate region and complexity has drawn much attention in the design of multi-user MIMO systems, especially from a viewpoint of massive MIMO.

One of the most fundamental problems in wireless communications design is the weighted sum-rate maximization (WSRMax). In this regard, the vast majority of previous studies consider a sum power constraint (SPC) across all the transmit antennas. Despite its simplicity and mathematical tractability, SPC is too ideal to be applied in practical MIMO transceivers. This is because each transmit antenna has its own power amplifier, hence per-antenna power constraint (PAPC) is more realistic. The PAPC consideration in MIMO systems has recently been investigated extensively in the literature, e.g., [3]–[10]. In relation to ZF methods, an interesting result in [3] showed that the pseudo-inverse-based precoder is no longer optimal for PAPC. Furthermore, the authors in [3] transformed the sum-rate maximization (SRMax) problem under PAPC into a determinant maximization (MAXDET) program, which was then solved by off-the-shelf convex solvers. To the best of our knowledge, the first attempt to solve the WSRMax problem by closed-form expressions was made in [11]. However, this is achieved by a dual subgradient method which generally suffers from slow convergence.

This paper proposes a semi-closed-form solution to the WSRMax problem subject to PAPC for ZF method. Unlike the dual subgradient approach in [11] which was derived directly in the BC, we first convert the BC's WSRMax problem into a *minimax problem* in its dual multiple access channel (MAC) by leveraging the duality of BC and MAC channels. Our main contributions include the following:

- We recruit the alternating optimization (AO) and successive convex approximation (SCA) methods to derive an iterative algorithm to find a saddle-point for the minimax problem. These two optimization techniques are combined in such a way that monotonic convergence is achieved. To the best of our knowledge, such novel combination of AO and SCA has not been reported elsewhere.
- By exploiting the specific structure of the considered problem, we can solve the subproblem at each iteration of the proposed method by water-filling-like algorithms. Thus, the proposed method can deal with the ZF precoder design in large-scale MIMO systems that are beyond the capability of state-of-the-art convex solvers.
- We numerically demonstrate that the proposed algorithm can converge much faster compared to existing approaches.
- We show that the proposed algorithm can be easily modified to treat other types of power constraint and can also be generalized to cope with other precoding methods.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

T. M. Pham, R. Farrell and J. Dooley are with Department of Electronic Engineering, Maynooth University, Ireland. Email: {minhthuy.pham, ronan.farrell, john.dooley}@mu.ie.

E. Dutkiewicz and D. N. Nguyen are with Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. Email: {eryk.dutkiewicz, diep.nguyen}@uts.edu.au.

L.-N. Tran is with School of Electrical and Electronic Engineering, University College Dublin, Ireland. Email: nam.tran@ucd.ie.

Notation: Standard notations are used in this paper. Bold lower and upper case letters represent vectors and matrices, respectively. $\mathbf{I}$ defines an identity matrix, of which the size can be easily inferred from the context; $\mathbb{C}^{M\times N}$ denotes the space of $M\times N$ complex matrices; $\mathbf{H}^H$ and $\mathbf{H}^T$ are Hermitian and normal transpose of $\mathbf{H}$, respectively; $\mathbf{H}_{i,j}$ is the (i,j)th entry of $\mathbf{H}$; $|\mathbf{H}|$ is the determinant of $\mathbf{H}$; $\mathrm{rank}(\mathbf{H})$ stands for the rank of $\mathbf{H}$; $\mathrm{diag}(\mathbf{x})$ denotes the diagonal matrix with diagonal elements being $\mathbf{x}$ and $\mathbb{E}[.]$ is the expectation value of a random variable; $\mathrm{diag}(\mathbf{H})$, where $\mathbf{H}$ is a square matrix, returns the vector of diagonal elements of $\mathbf{H}$. Furthermore, we denote $[x]_+ = \max(x,0)$, $||\cdot||$ to be the Euclidean norm.

II. System Model

Consider a K-user MIMO BC where the base station and each user have N and M_k antennas, respectively. Let $\mathbf{H}_k$ be the channel matrix for user k. Then, the received signal at user k is given as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{s}_k + \sum\nolimits_{j\neq k}\mathbf{H}_k\mathbf{s}_j + \mathbf{z}_k \tag{1}$$

where $\mathbf{s}_k$ is the downlink signal and $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_M)$ refers to the noise for the kth user. For linear ZF precoding, we can express $\mathbf{s}_k$ as $\mathbf{s}_k = \mathbf{R}_k\mathbf{x}_k$, where $\mathbf{R}_k$ and $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_M)$ denote the precoding matrix and information-bearing signal, respectively. For user k, the interference from other users in the system is suppressed by designing $\mathbf{R}_k$ such that $\mathbf{H}_j\mathbf{R}_k = 0$ for all $j\neq k$. The WSRMax problem for ZF precoding with PAPC is formulated as [11]

$$\begin{array}{ll}\underset{\{\mathbf{X}_k\succeq\mathbf{0}\}}{\text{maximize}} & \sum_{k=1}^K w_k \log|\mathbf{I}+\mathbf{H}_k\mathbf{X}_k\mathbf{H}_k^H| \\ \text{subject to} & \mathbf{H}_j\mathbf{X}_k\mathbf{H}_j^H = 0,\ \forall j\neq k \\ & \sum_{k=1}^K[\mathbf{X}_k]_{i,i}\le P_i,\ i=1,2,\dots,N\end{array} \tag{2}$$

where $\mathbf{X}_k = \mathbb{E}[\mathbf{s}_k\mathbf{s}_k^H] = \mathbf{R}_k\mathbf{R}_k^H$ is the input covariance matrix for user k, P_i is the power constraint on antenna i, and w_k is the positive weighting factor assigned to the kth user. In the above formulation we have omitted the rank constraint $\mathrm{rank}(\mathbf{X}_k)\le M_k$ but this step does not affect the optimality as proved in [11]. We also remark that this rank constraint will be automatically satisfied the proposed solution presented next.

III. Proposed Algorithm

A. Algorithm Description

In this section, we derive an efficient algorithm to solve (2) using minimax duality, AO, and SCA. Assuming that $N > \sum M_k - \min\{M_k\}$, let $\check{\mathbf{H}}_k$ be the channel matrix of all users, except for user k, i.e. $\check{\mathbf{H}}_k = [\mathbf{H}_1^H,\dots\mathbf{H}_{k-1}^H,\mathbf{H}_{k+1}^H,\dots\mathbf{H}_K^H]^H$, and $\mathbf{B}_k$ be a basis of the null space of $\check{\mathbf{H}}_k$. Then (2) reduces to the following problem

$$\begin{array}{ll}\underset{\{\tilde{\mathbf{X}}_k\succeq\mathbf{0}\}}{\text{maximize}} & \sum_{k=1}^K w_k\log|\mathbf{I}+\mathbf{H}_k\mathbf{B}_k\tilde{\mathbf{X}}_k\mathbf{B}_k^H\mathbf{H}_k^H| \\ \text{subject to} & \sum_{k=1}^K[\mathbf{B}_k\tilde{\mathbf{X}}_k\mathbf{B}_k^H]_{i,i}\le P_i, i=1,\dots,N.\end{array} \tag{3}$$

For the special case of SRMax problem (i.e., $w_1 = w_2 = \cdots w_K$), (3) becomes a MAXDET program as mentioned in [3]. We further note that for this special case, (3) can be recast as a semidefinite program. For the general case of WSRMax problem, the optimization package SDPT3 is a dedicated solver. However, solving (3) by generic convex solvers is not practically appealing for a large number of antennas N and/or a large number of users K. A closed-form solution for (3) was proposed in [11], but it was found by leveraging the subgradient method whose convergence rate is typically slow.

To overcome the aforementioned drawbacks, by extending Theorem 2 of [12], we first transform (3) into a minimax problem in the dual MAC as

$$\begin{array}{ll}\underset{\mathbf{\Lambda}\succeq\mathbf{0}}{\min}\ \underset{\{\bar{\mathbf{X}}_k\succeq\mathbf{0}\}}{\max} & \sum_{k=1}^K w_k\log\frac{|\mathbf{B}_k^H\mathbf{\Lambda}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k\tilde{\mathbf{H}}_k|}{|\mathbf{B}_k^H\mathbf{\Lambda}\mathbf{B}_k|} \\ \text{subject to} & \sum_{k=1}^K\mathrm{tr}(\bar{\mathbf{X}}_k)=P;\mathrm{tr}(\mathbf{\Lambda}\mathbf{P})=P,\mathbf{\Lambda}:\mathsf{diagonal}\end{array} \tag{4}$$

where $\tilde{\mathbf{H}}_k = \mathbf{H}_k\mathbf{B}_k$. Then the optimal solution $\tilde{\mathbf{X}}_k^*$ of (3) is given by

$$\tilde{\mathbf{X}}_k^* = (\mathbf{B}_k^H\mathbf{\Lambda}^*\mathbf{B}_k)^{-\frac{1}{2}}\mathbf{U}_k\mathbf{V}_k^H\bar{\mathbf{X}}_k^*\mathbf{V}_k\mathbf{U}_k^H(\mathbf{B}_k^H\mathbf{\Lambda}^*\mathbf{B}_k)^{-\frac{1}{2}} \tag{5}$$

where $\{\bar{\mathbf{X}}_k^*\}$ and $\mathbf{\Lambda}^*$ are a saddle-point of (4), and $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H$ is the economy-size singular value decomposition of $(\mathbf{B}_k^H\mathbf{\Lambda}^*\mathbf{B}_k)^{-1/2}\tilde{\mathbf{H}}_k^H$. A proof of this transformation is given in Appendix A.

The problem now is to find a saddle-point of (4). For a general minimax optimization, one may alternate between minimization and maximization but the convergence of such a method is not guaranteed. A more common approach to tackle (4) is based on Newton's method, e.g., [8]. However, the complexity of this method increases rapidly with the problem size. In the sequel, we show that (4) can be solved efficiently by combining AO and SCA to derive closed-form expressions.

Let $\{\bar{\mathbf{X}}_k^n\}$ be the optimal value of the following maximization in the nth iteration

$$\begin{array}{ll}\max & \sum_{k=1}^K w_k\log|\mathbf{B}_k^H\mathbf{\Lambda}^n\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k\tilde{\mathbf{H}}_k| \\ \text{s.t.} & \sum_{k=1}^K\mathrm{tr}(\bar{\mathbf{X}}_k)=P;\bar{\mathbf{X}}_k\succeq\mathbf{0},k=1,\dots,K\end{array} \tag{6}$$

Note that the above problem admits the water-filling solution which is skipped here for the sake of brevity.

Now, we turn our attention to the minimization of $\mathbf{\Lambda}$ for given $\{\bar{\mathbf{X}}_k^n\}$. To achieve monotonic convergence, instead of minimizing the objective of (4), we construct and then minimize an upper bound of it. This step is inspired by the concept of SCA which has received growing attention recently. To this end, we recall the following inequality which results from the concavity of the log-determinant function [13, p. 73]

$$\log|\mathbf{B}_k^H\mathbf{\Lambda}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k^n\tilde{\mathbf{H}}_k|\le\log|\mathbf{\Phi}_k^n|\,\mathrm{tr}\big(\mathbf{B}_k\mathbf{\Phi}_k^{-n}\mathbf{B}_k^H\big(\mathbf{\Lambda}-\mathbf{\Lambda}^n\big)\big) \tag{7}$$

where $\mathbf{\Phi}_k^n \triangleq \mathbf{B}_k^H\mathbf{\Lambda}^n\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k^n\tilde{\mathbf{H}}_k$, and $\mathbf{\Phi}_k^{-n}\triangleq\big(\mathbf{\Phi}_k^n\big)^{-1}$. Thus, in the $(n+1)$th iteration of the proposed algorithm, $\mathbf{\Lambda}^{n+1}$ is the solution to the following problem

$$\begin{array}{ll}\min & \sum_{k=1}^K w_k\big(\mathrm{tr}\big(\mathbf{B}_k\mathbf{\Phi}_k^{-n}\mathbf{B}_k^H\mathbf{\Lambda}\big)-\log|\mathbf{B}_k^H\mathbf{\Lambda}\mathbf{B}_k|\big) \\ \text{s.t.} & \mathrm{tr}(\mathbf{\Lambda}\mathbf{P})=P,\mathbf{\Lambda}:\mathsf{diagonal};\mathbf{\Lambda}\succeq\mathbf{0}\end{array} \tag{8}$$

We remark that the inequality in (7) is not entirely new. In fact it has been appeared in previous studies such as [14]–[16]. Our contributions in this regard are twofold. Firstly, the use of (7) allows us to analytically prove that the proposed algorithm converges monotonically to a saddle-point of (4). Secondly, we show that (8) can be solved by closed-form expressions as follows.

Since $\mathbf{\Lambda}$ is diagonal, (8) reduces to the following problem

$$\underset{\boldsymbol{\lambda}\in\Theta}{\text{minimize}}\ \boldsymbol{\alpha}^T\boldsymbol{\lambda} - \textstyle\sum_{k=1}^{K} w_k \log|\mathbf{B}_k^H \operatorname{diag}(\boldsymbol{\lambda})\mathbf{B}_k| \qquad (9)$$

where $\boldsymbol{\alpha} = \sum_{k=1}^{K} w_k \big(\operatorname{diag}(\mathbf{B}_k \mathbf{\Phi}_k^{-n} \mathbf{B}_k^H)\big)$ and $\Theta \triangleq \{\mathbf{p}^T\boldsymbol{\lambda} = P; \boldsymbol{\lambda} \geq 0\}$. From the above, we observe that (i) Θ is a simplex, and (ii) projection onto a simplex can be computed by a water-filling-like algorithm as shown shortly. These observations lead to the proposed gradient projection method to solve (9), which is outlined in Algorithm 1.

Algorithm 1: The proposed gradient projection algorithm for solving (9).

Input: $\boldsymbol{\lambda}_0$, $m = 0, \epsilon_1 > 0, \tau_1 = 1 + \epsilon_1$.

1 **repeat**

2 Calculate the gradient $\tilde{\mathbf{g}}_m = -\nabla f(\boldsymbol{\lambda}_m) = \sum_{k=1}^{K} w_k \operatorname{diag}(\mathbf{B}_k(\mathbf{B}_k^H \operatorname{diag}(\boldsymbol{\lambda}_m)\mathbf{B}_k)^{-1}\mathbf{B}_k^H) - \boldsymbol{\alpha}$.

3 Choose an appropriate positive scalar ρ_m and create $\tilde{\boldsymbol{\lambda}}_m = \boldsymbol{\lambda}_m + \rho_m \tilde{\mathbf{g}}_m$.

4 Project $\tilde{\boldsymbol{\lambda}}_m$ onto Θ to obtain $\bar{\boldsymbol{\lambda}}_m$.

5 Choose appropriate step size ν_m using the Armijo rule [17] and set $\boldsymbol{\lambda}_{m+1} = \boldsymbol{\lambda}_m + \nu_m(\bar{\boldsymbol{\lambda}}_m - \boldsymbol{\lambda}_m)$.

6 $m := m + 1$.

7 **until** $\tau_1 = |\nabla f(\boldsymbol{\lambda}_m)^T(\boldsymbol{\lambda}_{m+1} - \boldsymbol{\lambda}_m)| < \epsilon_1$;

Output: $\boldsymbol{\lambda}_m$ as the optimal solution to (9).

In Algorithm 1, the subscript m denotes the iteration index. The main operation of Algorithm 1 is the projection of $\tilde{\boldsymbol{\lambda}}_m$ onto Θ which can be formulated as

$$\begin{array}{ll} \text{minimize} & ||\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_m||^2 \\ \text{subject to} & \mathbf{p}^T\boldsymbol{\lambda} = P; \boldsymbol{\lambda} \geq \mathbf{0}. \end{array} \qquad (10)$$

It is easy to see that (10) can be solved efficiently by water-filling-like algorithm. Note that when an equal power constraint is considered, Θ becomes a canonical simplex for which more efficient algorithms for projection are available [18]. Moreover, Algorithm 1 can be easily modified into a conjugate gradient projection method. The overall algorithm to solve the WSRMax problem for ZF precoding with PAPC is summarized in Algorithm 2, and its convergence proof is provided in the Appendix. Note that the residual error τ_2 is only computed for $n \geq 1$

Algorithm 2: Proposed algorithm for solving (3).

Input: $\mathbf{\Lambda} := \mathbf{\Lambda}^0$, $n := 0$, $\epsilon_2 > 0, \tau_2 = 1 + \epsilon_2$

1 **repeat**

2 Apply the water-filling algorithm to solve (6). Denote the optimal solution by $\{\bar{\mathbf{X}}_k^n\}$.

3 For each k, set $\mathbf{\Phi}_k^n = (\mathbf{B}_k^H \mathbf{\Lambda}^n \mathbf{B}_k + \tilde{\mathbf{H}}_k^H \bar{\mathbf{X}}_k^n \tilde{\mathbf{H}}_k)$.

4 Find $\mathbf{\Lambda}^{n+1}$ using Algorithm 1.

5 $n := n + 1$.

6 **until** $\tau_2 = |f(\mathbf{\Lambda}^n, \bar{\mathbf{X}}^n) - f(\mathbf{\Lambda}^{n-1}, \bar{\mathbf{X}}^{n-1})| < \epsilon_2$;

Output: $\{\bar{\mathbf{X}}_k^n\}_{k=1}^K$ and apply the BC-MAC transformation to compute optimal $\{\tilde{\mathbf{X}}_k^n\}_{k=1}^K$.

B. Complexity Analysis

In this section, we provide the complexity analysis of the proposed algorithm in terms of the number of flops. The flop count for related operations is taken from [19] and [20]. For convenience, we assume that all the receivers have the same number of antennas i.e. $M_k = M$. To solve the SDP problem for K covariance matrices of $N \times N$ by the interior-point-based approach (e.g., [3]), the complexity is $\mathcal{O}(K^3N^6)$ [13], [21]. As explained earlier, Algorithm 2 performs the water-filling algorithm and eigenvalue decomposition to solve (6) , which needs $K(N-(K-1)M)^3 + K(4(N-(K-1)M)^2M - 8(N-(K-1)M)M^2)$ flops [20]. To find $\mathbf{\Lambda}$, Algorithm 1 requires $K(N-(K-1)M)^3$ flops for gradient computation (cf. line 2), while the complexity of the projection on a simplex (cf. line 4) and of other steps is negligible, and therefore is ignored. Thus, the total per-iteration complexity of Algorithm 2 is $\mathcal{O}(KN^3)$ flops. For the same problem, the subgradient-based method in [11] has a similar per-iteration complexity. However, the subgradient method is generally known for slow convergence, and thus potentially results in higher overall computation time that is investigated in the next section.

IV. Numerical Results

This section numerically evaluates the performance of the proposed algorithms. The step size ρ_m in Algorithm 1 is fixed at $\rho_m = 1$, which is empirically found to achieve fastest convergence rate for all the considered simulation scenarios. The initial value $\mathbf{\Lambda}^0$ in Algorithm 2 is set to the identity matrix for all simulations unless otherwise stated. The power constraint is set equally for all the transmit antennas, i.e., $P_i = P/N$, for $i = 1, \ldots, N$. Other relevant simulation parameters are specified for each setup. The codes are built in MATLAB and executed on a 64-bit desktop that supports 8 Gbyte RAM and Intel CORE i7.

In the first simulation, we investigate the convergence behavior of Algorithm 1. To find $\mathbf{\Lambda}^n$, we use $\mathbf{\Lambda}^{n-1}$ as the input of Algorithm 1. As can be seen in Fig. 1, Algorithm 1 achieves monotonic convergence as a result of employing the gradient projection method. Also, Algorithm 1 converges faster and obtains superlinear convergence rate after some first iterations. The reason is that when $\mathbf{\Lambda}^n$ is close to optimal, the gradient projection method tends to converge superlinearly [22].

Next, we compare the convergence rate of Algorithm 2 and the method in [11] for two different initial points: one taken as all-one vector and the other generated randomly. For fair comparison, the residual error for both methods is defined as the difference between the value of the current objective and the globally optimal one obtained by YALMIP [25] with MOSEK [23] as internal solver. We can clearly see in Fig. 2 that Algorithm 2 always achieves monotonic convergence which is in complete agreement with our proof in the Appendix. Moreover, Algorithm 2 converges much faster than the subgradient method in [11] for both cases.

To obtain a more comprehensive comparison, we report the average run time for solving (3) by several approaches over 1000 channel realizations. As mentioned earlier, we

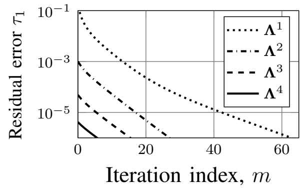

Fig. 1. Convergence result of Algorithm 1 for finding $\mathbf{\Lambda}^1$, $\mathbf{\Lambda}^2$, $\mathbf{\Lambda}^3$, and $\mathbf{\Lambda}^4$. The total power is $P = 10$ dBW. We simply set $\rho_m = 1$ during the whole iterative process.

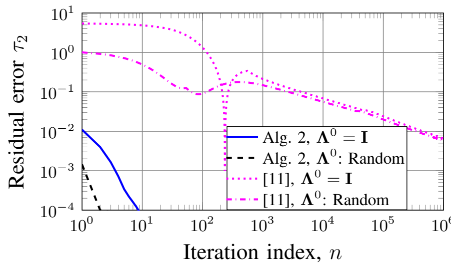

Fig. 2. Convergence comparison of different iterative methods with $N = 8$, $M = 2$ and $K = 4$.

can use generic convex solvers to solve (3) optimally. Here we compare Algorithm 2 to MOSEK [23] and SDPT3 [24] through the parser YALMIP [25]. In Table I, '$\times$' stands for either a computer crash or extremely large computation time. Table I clearly shows that Algorithm 2 requires the lowest computation time. Recall that in the aforementioned complexity analysis Algorithm 2 and [11] have similar per-iteration complexity order. However, the subgradient-based algorithm needs much more time to solve (3), due to slow convergence rate as illustrated in Fig. 2. Off-the-shelf solvers i.e. MOSEK and SDPT3 work relatively effectively for small N, but fail for large N. This result can be explained by the fact that interior-point-based solvers do not scale well with the problem size. Note that our simulation codes are built on MATLAB environment which is by no means real-time implementation. Thus the run time reported in Table I is mainly for relative benchmarking purpose. Real-time implementation of the proposed algorithm is beyond the scope of the paper and is left as future work. However, it is normally expected that embedded implementation can speed up the efficiency of a MATLAB code by several orders of magnitude. We further remark that the coherence time in massive MIMO systems is much larger than that in conventional ones due to the channel hardening effect [27], [28]. Thus these two facts may indicate good embedded implementation of the proposed algorithm is likely suitable for real-time massive MIMO applications.

TABLE I

AVERAGE RUN TIME (SECONDS) COMPARISON FOR $P = 10$ dBW, $M = 2$, $K = 8$. THE RUN TIME IS AVERAGED OVER 1000 CHANNEL REALIZATIONS.

	No. of Tx. antennas N	16	32	64	128
SRMax	Algorithm 2	**0.097**	**3.84**	**115.79**	**175.96**
	[11]	0.85	610.92	> 1 hr	$\times$
	SDPT3	0.32	11.36	$\times$	$\times$
	MOSEK	0.23	11.01	> 1 hr	$\times$
WSRMax	Algorithm 2	**0.10**	**2.24**	**92.47**	**147.18**
	[11]	1.02	89.27	> 1 hr	$\times$
	SDPT3	0.32	7.99	$\times$	$\times$

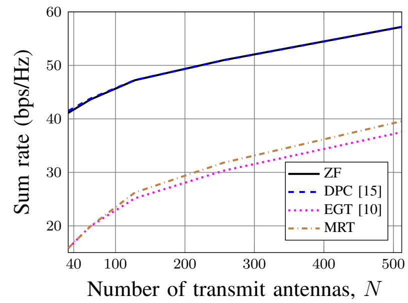

Fig. 3. Sum rate of linear precoding (ZF, EGT, MRT) versus nonlinear precoding (DPC) in massive MIMO system with PAPC, $K = 4$ users, $M = 1$ antenna.

Finally we utilize Algorithm 2 to evaluate the sum-rate performance of ZF methods in a massive MIMO setting [26]. Small-scale fading in massive MIMO systems tends to be deterministic due to the so-called channel hardening effect [27], [28], and thus a popular strategy to optimize performance is to perform power control merely accounting for large-scale fading. An example using this design philosophy was presented in [10], where a precoding scheme based on the principle of equal gain transmission (EGT) was proposed under PAPC. We remark that Algorithm 2 takes into account the both small- and large-scale fading when computing optimal ZF precoders, and thus offers the best performance. Our purpose in the last numerical experiment is to quantify how channel hardening affects the performance of large-scale MIMO in the particular case of ZF precoding. Towards this end we consider a single-cell under the default typical urban micro-cell B1 WINNER II channel model [29]. The base station is located at the center of the cell and 4 single-antenna users are distributed with a distance ranging from 70 meters to 212 meters around the base station. Moreover, we only take path loss into account and ignore the shadowing. The transmit power and the noise power are set to 30 dBm and -94 dBm over a bandwidth of 100 MHz, respectively. The results were averaged over 1000 channel realizations.

Fig. 3 plots the sum rate as a function of the number of transmit antennas. As can be seen clearly, when both large- and small-scale fadings are considered, ZF achieves a sum-rate performance very close to that of DPC, especially when N is very large. In fact, this observation again confirms the widely known fact that ZF is a much-lower complexity alternative to DPC. In Fig. 3 we also report the sum rate of the well-known maximal ratio transmission (MRT) scheme subject to PAPC. It is clear from Fig. 3 that if simple linear methods such as EGT and MRT are employed, there is still a big gap between their performance and that of optimal ZF precoding even for a relatively large number of antennas. This observation basically implies that small-scale fading is still beneficial and may be taken into account in future practical massive MIMO designs.

V. CONCLUSIONS

We have solved the WSRMax problem for ZF methods subject to PAPC. The proposed algorithm is based on the BC-

MAC duality together with AO and SCA. We have carried out numerical experiments to demonstrate the superior performance of the proposed algorithm over existing approaches. For considered large-scale scenarios, the proposed method is able to compute the optimal solution after relatively short time, while other methods of comparison fail or take much longer time. Consequently, our proposed method provides a powerful tool to characterize the achievable rate region of ZF schemes under PAPC for large-scale MIMO systems. In particular, we have utilized the proposed method to evaluate the performance of known simpler linear algorithms such as EGT and MRT that mainly focus on large-scale fading. It has been shown that the performance of these simpler methods is quite far from optimal. Our conclusion is that small-scale fading should be accounted for to take full advantage of massive MIMO. We remark that the proposed algorithm can be easily modified to deal with the general case of MIMO optimization with multiple transmit covariance constraints [30]. In this case, the resulting minimax problem in the dual MAC has the same structure as (4). A thorough study on this regards is left as future work.

APPENDIX A
DUALITY TRANSFORMATION PROOF

The duality transformation in (4) can be proved using the same arguments as those in [12]. First, we write the partial Lagrangian function of (3) as

$$\mathcal{L}(\{\tilde{\mathbf{X}}_k\},\mathbf{A}) = \sum\nolimits_{k=1}^{K}(w_k \log|\mathbf{I}+\tilde{\mathbf{H}}_k\tilde{\mathbf{X}}_k\tilde{\mathbf{H}}_k^H| - \mathrm{tr}(\mathbf{C}_k\tilde{\mathbf{X}})) + \mathrm{tr}(\mathbf{AP}) \tag{11}$$

where $\mathbf{C}_k = \mathbf{B}_k^H\mathbf{A}\mathbf{B}_k$, $\mathbf{A} = \mathrm{diag}(a_1, a_2, \dots, a_i, \dots, a_N)$. Let $\hat{\mathbf{X}}_k = \mathbf{C}_k^{1/2}\tilde{\mathbf{X}}_k\mathbf{C}_k^{1/2}$. Then $\mathcal{L}(\{\tilde{\mathbf{X}}_k\},\mathbf{A})$ is equal to

$$\mathcal{L}(\{\tilde{\mathbf{X}}_k\},\mathbf{A}) = \sum\nolimits_{k=1}^{K}(w_k \log|\mathbf{I}+\tilde{\mathbf{H}}_k\mathbf{C}_k^{-1/2}\hat{\mathbf{X}}_k\mathbf{C}_k^{-1/2}\tilde{\mathbf{H}}_k^H| - \mathrm{tr}(\hat{\mathbf{X}}_k)) + \mathrm{tr}(\mathbf{AP}). \tag{12}$$

Denote $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H$ to be the singular value decomposition of $\tilde{\mathbf{H}}_k\mathbf{C}_k^{-1/2}$, i.e., $\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^H = \tilde{\mathbf{H}}_k\mathbf{C}_k^{-1/2}$. By the so-called channel flipping effect, we can express the dual objective as

$$\mathcal{D}(\mathbf{A}) = \max_{\dot{\mathbf{X}}_k\succeq 0}\sum\nolimits_{k=1}^{K}(w_k\log\frac{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\dot{\mathbf{X}}_k\tilde{\mathbf{H}}_k|}{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k|} - \mathrm{tr}(\dot{\mathbf{X}}_k)) + \mathrm{tr}(\mathbf{AP}) \tag{13}$$

where $\hat{\mathbf{X}}_k = \mathbf{V}_k\mathbf{U}_k^H\dot{\mathbf{X}}_k\mathbf{U}_k\mathbf{V}_k^H$. Now the dual problem of (3) is

$$\min_{\mathbf{A}\succeq\mathbf{0}}\max_{\{\dot{\mathbf{X}}_k\}\succeq 0}\sum\nolimits_{k=1}^{K}(w_k\log\frac{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\dot{\mathbf{X}}_k\tilde{\mathbf{H}}_k|}{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k|} - \mathrm{tr}(\dot{\mathbf{X}}_k)) + \mathrm{tr}(\mathbf{AP}). \tag{14}$$

By introducing new optimization variable $\delta > 0$, we can rewrite the above problem as

$$\begin{aligned}\min_{\mathbf{A}\succeq\mathbf{0},\delta>0}\max_{\{\dot{\mathbf{X}}_k\}\succeq\mathbf{0}} & \sum\nolimits_{k=1}^{K} w_k\log\frac{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\dot{\mathbf{X}}_k\tilde{\mathbf{H}}_k|}{|\mathbf{B}_k^H\mathbf{A}\mathbf{B}_k|} - \delta P + \mathrm{tr}(\mathbf{AP})\\ \text{subject to} & \sum\nolimits_{k=1}^{K}\mathrm{tr}(\dot{\mathbf{X}}) \le \delta P.\end{aligned} \tag{15}$$

Note that we can again change the optimization variables as

$$\bar{\mathbf{X}}_k = \frac{\dot{\mathbf{X}}_k}{\delta};\, \bar{\mathbf{A}} = \frac{\mathbf{A}}{\delta}. \tag{16}$$

Thus, (15) is equivalent to

$$\begin{aligned}\min_{\bar{\mathbf{A}}\succeq\mathbf{0}}\max_{\{\bar{\mathbf{X}}_k\}\succeq\mathbf{0}} & \sum\nolimits_{k=1}^{K} w_k\log\frac{|\mathbf{B}_k^H\bar{\mathbf{A}}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k\tilde{\mathbf{H}}_k|}{|\mathbf{B}_k^H\bar{\mathbf{A}}\mathbf{B}_k|}\\ \text{subject to} & \sum\nolimits_{k=1}^{K}\mathrm{tr}(\bar{\mathbf{X}}_k)\le P; \mathrm{tr}(\bar{\mathbf{A}}\mathbf{P})\le P.\end{aligned} \tag{17}$$

which is the form given in (4) and thus completes the proof.

APPENDIX B
CONVERGENCE PROOF OF ALGORITHM 2

Let us define $\Omega \triangleq \{\mathbf{\Lambda}|\mathbf{\Lambda}: \mathsf{diagonal}, \mathbf{\Lambda}\succeq\mathbf{0}, \mathrm{tr}(\mathbf{\Lambda P}) = P\}$ and $\mathcal{X} = \{\bar{\mathbf{X}}_k|\bar{\mathbf{X}}_k\succeq\mathbf{0}, \sum_{k=1}^{K}\mathrm{tr}(\bar{\mathbf{X}}_k) = P, k = 1,\dots,K.\}$. We note that the sets Ω and $\mathcal{X}$ are compact convex. We first show that Algorithm 2 yields a *decreasing* objective $f(\mathbf{\Lambda}^n,\{\bar{\mathbf{X}}_k^n\})$ following similar arguments in [15], [16]. Since $\mathbf{\Lambda}^{n+1}$ is the optimal solution to the minimization problem (8), the inequality below holds

$$f(\mathbf{\Lambda}^n,\{\bar{\mathbf{X}}_k^n\}) = \sum_{k=1}^{K} w_k\Big(\log|\mathbf{\Phi}_k^n| - \log|\mathbf{B}_k^H\mathbf{\Lambda}^n\mathbf{B}_k|\Big) \ge \sum_{k=1}^{K} w_k \Big(\log|\mathbf{\Phi}_k^n| + \mathrm{tr}\big(\mathbf{B}_k\mathbf{\Phi}_k^{-n}\mathbf{B}_k^H\left(\mathbf{\Lambda}^{n+1}-\mathbf{\Lambda}^n\right)\big) - \log|\mathbf{B}_k^H\mathbf{\Lambda}^{n+1}\mathbf{B}_k|\Big). \tag{18}$$

In addition, $\log|\mathbf{B}_k^H\mathbf{\Lambda}\mathbf{B}_k+\tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k\tilde{\mathbf{H}}_k|$ is jointly concave with $\mathbf{\Lambda}$ and $\bar{\mathbf{X}}_k$ and note that $\bar{\mathbf{X}}_k$ is the optimal solution to (6), we can easily prove that

$$\sum_{k=1}^{K} w_k\Big(\log|\mathbf{\Phi}_k^n| + \mathrm{tr}\big(\mathbf{B}_k\mathbf{\Phi}_k^{-n}\mathbf{B}_k^H\left(\mathbf{\Lambda}^{n+1}-\mathbf{\Lambda}^n\right)\big) - \log|\mathbf{B}_k^H\mathbf{\Lambda}^{n+1}\mathbf{B}_k|\Big) \ge \underbrace{\sum_{k=1}^{K} w_k\Big(\log|\mathbf{B}_k^H\mathbf{\Lambda}^{n+1}\mathbf{B}_k + \tilde{\mathbf{H}}_k^H\bar{\mathbf{X}}_k^{n+1}\tilde{\mathbf{H}}_k| - \log|\mathbf{B}_k^H\mathbf{\Lambda}^{n+1}\mathbf{B}_k|\Big)}_{f(\mathbf{\Lambda}^{n+1},\{\bar{\mathbf{X}}_k^{n+1}\})}. \tag{19}$$

Combining (18) and (19) results in

$$f(\mathbf{\Lambda}^n,\{\bar{\mathbf{X}}_k^n\}) \ge \sum_{k=1}^{K} w_k\Big(\log|\mathbf{\Phi}_k^n| + \mathrm{tr}\big(\mathbf{B}_k\mathbf{\Phi}_k^{-n}\mathbf{B}_k^H\left(\mathbf{\Lambda}^{n+1}-\mathbf{\Lambda}^n\right)\big) - \log|\mathbf{B}_k^H\mathbf{\Lambda}^{n+1}\mathbf{B}_k|\Big) \overset{(a)}{\ge} f(\mathbf{\Lambda}^{n+1},\{\bar{\mathbf{X}}_k^{n+1}\}). \tag{20}$$

We remark that the inequality (a) is strict if $\mathbf{\Lambda}^n \ne \mathbf{\Lambda}^{n+1}$. Thus, the sequence $\{f(\mathbf{\Lambda}^n,\{\bar{\mathbf{X}}_k^n\})\}$ is *strictly decreasing* unless it is convergent. Moreover, the continuity of $f(\cdot)$ and the compactness of $\mathcal{X}$ and Ω imply $\lim_{n\to\infty} f(\mathbf{\Lambda}^n,\{\bar{\mathbf{X}}_k^n\}) = f(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})$.

Now let $\{(\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\})\}$ be the subsequence converging to the limit point. Next we shall show that $\{(\mathbf{\Lambda}^{n_i+1},\{\bar{\mathbf{X}}_k^{n_i+1}\})\}\to(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})$. In fact, it is sufficient to prove that $\mathbf{\Lambda}^{n_i+1}\to\mathbf{\Lambda}^*$ which can be done by contradiction. Assume the contrary that $\mathbf{\Lambda}^{n_i+1}$ does not converge to $\mathbf{\Lambda}^*$. Consequently, there exists a $d > 0$ such that

$$d \le d^{n_i} = ||\mathbf{\Lambda}^{n_i+1} - \mathbf{\Lambda}^{n_i}||, \forall n_i \tag{21}$$

where $||\cdot||$ stands for arbitrary norm. We have

$$\begin{aligned} f(\mathbf{\Lambda}^{n_i+1},\{\bar{\mathbf{X}}_k^{n_i+1}\}) &\leq F(\mathbf{\Lambda}^{n_i}+d^{n_i}\mathbf{\Gamma}^{n_i};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) \\ &\leq F(\mathbf{\Lambda}^{n_i}+\delta d\mathbf{\Gamma}^{n_i};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}),\forall\delta\in[0,1] \\ &\leq F(\mathbf{\Lambda}^{n_i};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) = f(\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) \end{aligned} \quad (22)$$

where $F(\mathbf{\Lambda}^{n_i+1};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) = \sum_{k=1}^{K} w_k\Big(\log|\mathbf{\Phi}_k^{n_i}| + \operatorname{tr}\Big(\mathbf{B}_k\mathbf{\Phi}_k^{-n_i}\mathbf{B}_k^H\big(\mathbf{\Lambda}^{n_i+1}-\mathbf{\Lambda}^{n_i}\big)\Big) - \log|\mathbf{B}_k^H\mathbf{\Lambda}^{n_i+1}\mathbf{B}_k|\Big)$. Letting $k\to\infty$ leads to

$$f(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}) = F(\mathbf{\Lambda}^*+\delta d\mathbf{\Gamma}^*;\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}),\forall\delta\in[0,1]. \quad (23)$$

Furthermore

$$\begin{aligned} F(\mathbf{\Lambda}^{n_{i+1}};\mathbf{\Lambda}^{n_{i+1}},\{\bar{\mathbf{X}}_k^{n_{i+1}}\}) &= f(\mathbf{\Lambda}^{n_{i+1}},\{\bar{\mathbf{X}}_k^{n_{i+1}}\}) \\ \leq f(\mathbf{\Lambda}^{n_i+1},\{\bar{\mathbf{X}}_k^{n_i+1}\}) &\leq F(\mathbf{\Lambda}^{n_i+1},\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) \\ &\leq F(\mathbf{\Lambda};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}),\forall\mathbf{\Lambda}\in\Omega. \end{aligned} \quad (24)$$

Letting $k\to\infty$ we obtain

$$F(\mathbf{\Lambda}^*;\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}) \leq F(\mathbf{\Lambda};\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}),\forall\mathbf{\Lambda}\in\Omega \quad (25)$$

which further implies that $\mathbf{\Lambda}^*$ is the minimizer of $F(\cdot;\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})$. Since $\mathbf{\Lambda}^{n_i+1} = \underset{\mathbf{\Lambda}\in\Omega}{\arg\min} F(\mathbf{\Lambda};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\})$ it follows that

$$F(\mathbf{\Lambda}^{n_i+1};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}) \leq F(\mathbf{\Lambda};\mathbf{\Lambda}^{n_i},\{\bar{\mathbf{X}}_k^{n_i}\}),\forall\mathbf{\Lambda}\in\Omega. \quad (26)$$

Letting $k\to\infty$ implies

$$F(\mathbf{\Lambda}^*;\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}) \leq F(\mathbf{\Lambda};\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}),\forall\mathbf{\Lambda}\in\Omega. \quad (27)$$

That is

$$\langle\nabla_{\mathbf{\Lambda}}F(\mathbf{\Lambda};\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})|_{\mathbf{\Lambda}=\mathbf{\Lambda}^*},\mathbf{W}-\mathbf{\Lambda}^*\rangle \geq 0,\forall\mathbf{W}\in\Omega \quad (28)$$

where $\langle . \rangle$ denotes the inner product. Recall that $F(\cdot;\mathbf{\Lambda},\{\bar{\mathbf{X}}_k\})$ is the first order of $f(\mathbf{\Lambda},\{\bar{\mathbf{X}}_k\})$. Thus it is easy to see that

$$\nabla_{\mathbf{\Lambda}}F(\mathbf{\Lambda};\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})|_{\mathbf{\Lambda}=\mathbf{\Lambda}^*} = \nabla f(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}) \quad (29)$$

and thus (28) is equivalent to

$$\langle\nabla_{\mathbf{\Lambda}}f(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}),\mathbf{W}-\mathbf{\Lambda}^*\rangle \geq 0,\forall\mathbf{W}\in\Omega. \quad (30)$$

In the same way we can show that

$$\langle\nabla_{\bar{\mathbf{X}}_k}f(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\}),\mathbf{A}_k-\bar{\mathbf{X}}_k^*\rangle \leq 0,\forall\mathbf{A}_k\in\mathcal{X}. \quad (31)$$

Two above inequalities imply that $(\mathbf{\Lambda}^*,\{\bar{\mathbf{X}}_k^*\})$ is a saddle point of (4), which completes the proof.

References

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

[2] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[3] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409 – 4418, Sep. 2008.

[4] M. Vu, "MIMO capacity with per-antenna power constraint," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1 – 5.

[5] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Successive zero-forcing DPC with sum power constraint: Low-complexity optimal precoders," in *Proc. IEEE ICC*, Jun. 2012, pp. 4857–4861.

[6] S. Shi, M. Schubert, and H. Boche, "Per-antenna power constrained rate optimization for multiuser MIMO systems," in *Proc. WSA*, Feb. 2008, pp. 270 – 277.

[7] A. Tolli, M. Codreanu, and M. Juntti, "Linear multiuser MIMO transceiver design with quality of service and per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3049–3055, Jul. 2008.

[8] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[9] D. Tuninetti, "On the capacity of the AWGN MIMO channel under per-antenna power constraints," in *Proc. IEEE ICC*, Jun. 2014, pp. 2153 – 2157.

[10] S. Zhang, R. Zhang, and T. J. Lim, "Massive MIMO with per-antenna power constraint," in *Proc. IEEE GlobalSIP*, Dec. 2014, pp. 642 – 646.

[11] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[12] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Beamformer designs for MISO broadcast channels with zero-forcing dirty paper coding," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1173–1185, Mar. 2013.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[14] Chris T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, Dec. 2010.

[15] T. M. Pham, R. Farrell, and L.-N. Tran, "Alternating optimization for capacity region of Gaussian MIMO broadcast channels with per-antenna power constraint," in *Proc. IEEE VTC-Spring*, 2017. [Online]. Available: http://arxiv.org/abs/1704.01473

[16] ——, "Low-complexity approaches for MIMO capacity with per-antenna power constraint," in *Proc. IEEE VTC-Spring*, 2017. [Online]. Available: https://arxiv.org/abs/1704.01470

[17] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives." *Pacific J. Math.*, vol. 16, no. 1, pp. 1 – 3, 1966.

[18] L. Condat, "Fast projection onto the simplex and the ℓ_1 ball," *Mathematical Programming, Series A*, vol. 158, no. 1, pp. 575 – 585, Jul. 2016.

[19] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. The John Hopkins Univ. Press, 1996.

[20] Z. Shen, R. Chen, J. Andrews, J. Heath, R.W., and B. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.

[21] B. Li, C. Z. Wu, H. H. Dam, A. Cantoni, and K. L. Teo, "A parallel low complexity zero-forcing beamformer design for multiuser MIMO systems via a regularized dual decomposition method," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4179 – 4190, Aug. 2015.

[22] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

[23] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015.

[24] K. C. Toh, M. J. Todd, and R. Tutuncu, "SDPT3— a Matlab software package for semidefinite programming," *Optimization Methods and Software*, Nov. 1999.

[25] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. the CACSD Conference*, 2004.

[26] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[27] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893 – 1909, Sep. 2004.

[28] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921 – 2935, May, 2017.

[29] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzić, M. Milojević, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. de Jong, and T. Rautiainen, "WINNER II channel models," *tech. rep. D1.1.2 V1.2, IST-4-027756 WINNER II*, 2007.

[30] L. Zhang and R. Zhang and Y. C. Liang and Y. Xin and H. V. Poor, "On Gaussian MIMO BC-MAC duality With multiple transmit covariance constraints," *IEEE Trans. Inf. Theory*,vol. 58, no. 4, pp. 2064–2078, Mar. 2012.