Successive Zero-forcing DPC with Per-Antenna Power Constraint: Optimal and Suboptimal Designs

Le-Nam Tran[*], Markku Juntti[*], Mats Bengtsson[†], and Björn Ottersten[†‡]
[*]Centre for Wireless Communications and Dept. Commun. Eng., University of Oulu, Oulu, Finland
Email: {le.nam.tran;markku.juntti}@ee.oulu.fi
[†]ACCESS Linnaeus Center, Royal Institute of Technology (KTH), Stockholm, Sweden
Email: {mats.bengtsson,bjorn.ottersten}@ee.kth.se
[‡]Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg

Abstract—This paper considers the precoder designs for successive zero-forcing dirty paper coding (SZF-DPC), a suboptimal transmission technique for MIMO broadcast channels (MIMO BCs). Existing precoder designs for SZF-DPC often consider a sum power constraint. In this paper, we address the precoder design for SZF-DPC with per-antenna power constraints (PAPCs), which has not been well studied. First, we formulate the precoder design as a rank-constrained optimization problem, which is generally difficult to handle. To solve this problem, we follow a relaxation approach, and prove that the optimal solution of the relaxed problem is also optimal for the original problem. Considering the relaxed problem, we propose a numerically efficient algorithm to find the optimal solution, which exhibits a fast convergence rate. Suboptimal precoder designs, with lower computational complexity, are also presented, and compared with the optimal ones in terms of achievable sum rate and computational complexity.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) transmission techniques are capable of increasing the channel capacity without the need of additional bandwidth or power [1]. In this work, we consider the downlink of a single cell, where a multi-antenna base station (BS) wants to send data to multiple multi-antenna receivers simultaneously, i.e., a MIMO broadcast channel (MIMO BC). It is well known that that the capacity region of a MIMO BC is achieved by dirty paper coding (DPC) [2]. However, finding the optimal transmit covariances for DPC requires iterative algorithms of high computational complexity [3], [4]. Thus, it is of particular interest to develop suboptimal solutions to DPC.

SZF-DPC, introduced in [5], is a suboptimal alternative to DPC for MIMO BCs. In fact, SZF-DPC is a generalization of the zero-forcing dirty paper coding (ZF-DPC) in [6], which was devised for MISO BCs, i.e. for single-antenna receivers. Related works for SZF-DPC often assume a sum power constraint when designing the precoders, where optimal precoders can be found analytically, i.e., by closed-form expressions [5]. In a companion paper, we propose several low-complexity

optimal precoder designs for SZF-DPC under the sum power constraint [7]. Herein, we consider the precoder design for SZF-DPC with per-antenna power constraints, which has not been extensively investigated.

In practice, individual per-antenna power constraints are more realistic than the sum power constraint, since each antenna is equipped with its own power amplifier. In fact, the precoder designs for channel inversion or block diagonalization with PAPCs in [8], [9] are applicable to SZF-DPC since SZF-DPC can be viewed as a relaxation of these schemes. Particularly, a numerical algorithm based on dual decomposition method was proposed in [9], using a subgradient method to find the optimal precoders for block diagonalized systems. Generally, subgradient methods are slow to converge to high accuracy. Herein, we propose a specialized numerical algorithm to compute the optimal precoders for SZF-DPC with PAPCs.

The precoder design with PAPCs is difficult to solve because closed-form solutions may not exist. For these cases, numerical algorithms should be developed to find the optimal solutions. First, we formulate the precoder design as a rank-constrained optimization problem. We further show that the relaxed problem (by dropping the rank constraints) is a convex optimization problem, whose all optimal solutions *always satisfy the rank constraints*. That is, both the original and the relaxed problems are equivalent. Then, we propose a numerical algorithm to solve the relaxed problem, based on a barrier method. As a part of the proposed algorithm, we observe that all the power constraints must be active at the optimum. This facilitates the numerical method to find the optimal solutions since equality constraints are easier to be handled. The proposed algorithm is numerically shown to have a better convergence behavior, compared to the two-stage iterative algorithm in [9]. Finally, we present suboptimal precoder designs which have lower computational complexity, and perform very close to the optimal solutions.

The remainder of the paper is organized as follow. In Section II, we briefly review the system model and the precoder design for SZF-DPC. The precoder design with per-antenna power constraints is addressed in Section III, where we present a specialized numerical algorithm to the find

the optimal precoders. We also introduce suboptimal designs which have lower computational complexity, but perform close to the optimal one. Numerical results are given in Section IV, followed by conclusions in Section V.

Notation: Standard notations are used in this paper. Bold lower and upper case letters represent vectors and matrices, respectively; $\boldsymbol{H}^H$ and $\boldsymbol{H}^T$ are Hermitian and normal transpose of $\boldsymbol{H}$, respectively; $||\boldsymbol{H}||_F$ and $|\boldsymbol{H}|$ are the Frobenius norm and determinant of $\boldsymbol{H}$, respectively. $\boldsymbol{I}_M$ represents an $M \times M$ identity matrix. $\mathcal{N}(\boldsymbol{H})$ denotes the null space of $\boldsymbol{H}$. $\mathrm{diag}(\boldsymbol{x})$, where $\boldsymbol{x}$ is a vector, denotes a diagonal matrix with elements $\boldsymbol{x}$; $\mathrm{diag}(\boldsymbol{H})$, where $\boldsymbol{H}$ is a square matrix, denotes a vector of its diagonal elements.

II. MIMO BCs AND SZF-DPC

Consider a single-cell MIMO BC with a base station (BS) and K multiple-antenna users. The channel between the BS and the kth user is modeled by a matrix $\boldsymbol{H}_k \in \mathbb{C}^{n_k \times N}$, where N and $n_k \geqslant 1$ are the number of antennas at the BS and the kth user, respectively. The received signal at the kth user is given by

$$\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{W}_k \boldsymbol{s}_k + \sum_{j<k} \boldsymbol{H}_k \boldsymbol{W}_j \boldsymbol{s}_j + \sum_{j>k} \boldsymbol{H}_k \boldsymbol{W}_j \boldsymbol{s}_j + \boldsymbol{z}_k. \quad (1)$$

where $\boldsymbol{y}_k$ is the received sample, $\boldsymbol{W}_k \in \mathbb{C}^{N \times L_k}$ is the precoder, and $\boldsymbol{s}_k \in \mathbb{C}^{L_k \times 1}$, $L_k \leq \min(N, n_k)$ is the vector of transmitted symbols of the kth user, respectively. The background noise $\boldsymbol{z}_k \in \mathbb{C}^{n_k \times 1}$ is assumed to be white complex-Gaussian with zero mean and covariance matrix $\boldsymbol{I}_{n_k}$. It is well-known that DPC is a capacity achieving transmission strategy for MIMO BCs. For the kth user, the BS views the interference term $\sum_{j<k} \boldsymbol{H}_k \boldsymbol{W}_j \boldsymbol{s}_j$ as known non-causally, and can be perfectly eliminated by DPC. As a result, the resulting data rate of the kth user is given by

$$R_k^{\mathrm{dpc}} = \log \frac{|\boldsymbol{I} + \sum_{j \leq k} \boldsymbol{H}_k \boldsymbol{S}_j \boldsymbol{H}_k^H|}{|\boldsymbol{I} + \sum_{j < k} \boldsymbol{H}_k \boldsymbol{S}_j \boldsymbol{H}_k^H|}, \quad (2)$$

where $\boldsymbol{S}_j = \boldsymbol{W}_j \boldsymbol{W}_j^H$ is the transmit covariance matrix for the jth user. Since (2) is non-convex with respect to $\boldsymbol{S}_j$, it is generally difficult to be handled. In [5], SZF-DPC was proposed as a low-complexity alternative to DPC, which admits a closed-form solution for optimal precoders. For SZF-DPC, the interference term $\sum_{j<k} \boldsymbol{H}_k \boldsymbol{W}_j \boldsymbol{s}_j$ in (1) is canceled by DPC, and the interference term $\sum_{j>k} \boldsymbol{H}_k \boldsymbol{W}_j \boldsymbol{s}_j$ is eliminated by designing $\boldsymbol{W}_j$ such that

$$\boldsymbol{H}_k \boldsymbol{W}_j = \boldsymbol{0} \text{ for all } j > k. \quad (3)$$

Accordingly, the resulting data rate of the kth user for SZF-DPC is written by

$$R_k^{\mathrm{szf\text{-}dpc}} = \log |\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{S}_k \boldsymbol{H}_k^H|. \quad (4)$$

In this paper, we consider the sum rate maximization of SZF-DPC under PAPCs, which is presented next.

III. SZF-DPC WITH PER-ANTENNA POWER CONSTRAINTS

While the precoder designs for SZF-DPC with a sum power constraint can be found easily using the water-filling algorithm [5], that with PAPCs has not been extensively studied. In practice, PAPCs are more relevant since each antenna has its own power amplifier [8], [10], [11]. The sum rate maximization for SZF-DPC under PAPCs is formulated as

$$\underset{\boldsymbol{W}_k}{\text{maximize}} \sum_{k=1}^{K} \log |\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{W}_k \boldsymbol{W}_k^H \boldsymbol{H}_k^H| \quad (5a)$$

$$\text{subject to } \boldsymbol{H}_j \boldsymbol{W}_k = \boldsymbol{0}, \ \forall j < k \quad (5b)$$

$$\sum_{k=1}^{K} [\boldsymbol{W}_k \boldsymbol{W}_k^H]_{n,n} \leq P_n, \ \forall n = 1, \ldots, N, \quad (5c)$$

where P_n is the power constraint for the nth transmit antenna. Note that (5a) is neither a convex nor a concave function with respect to $\boldsymbol{W}_k$. A convenient way to solve (5) is to define $\boldsymbol{S}_k = \boldsymbol{W}_k \boldsymbol{W}_k^H \succeq \boldsymbol{0}$, and rewrite (5) as

$$\begin{aligned} \underset{\boldsymbol{S}_k}{\text{maximize}} \quad & \sum_{k=1}^{K} \log |\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{S}_k \boldsymbol{H}_k^H| \\ \text{subject to} \quad & \boldsymbol{H}_j \boldsymbol{S}_k \boldsymbol{H}_j^H = \boldsymbol{0}, \ \forall j < k \\ & \sum_{k=1}^{K} [\boldsymbol{S}_k]_{n,n} \leq P_n, \ \forall n \\ & \mathrm{rank}(\boldsymbol{S}_k) \leq L_k, \ \forall k, \end{aligned} \quad (6)$$

where $L_k \leq n_k$. The rank constraints, which are non-convex, make (6) difficult to solve. In the following, we propose a method to find the optimal solutions for (6) based on a relaxation technique.

A. Optimal precoder design

We begin by omitting the rank constraints in (6), and consider the following relaxed problem

$$\begin{aligned} \underset{\boldsymbol{S}_k}{\text{maximize}} \quad & \sum_{k=1}^{K} \log |\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{S}_k \boldsymbol{H}_k^H| \\ \text{subject to} \quad & \boldsymbol{H}_j \boldsymbol{S}_k \boldsymbol{H}_j^H = \boldsymbol{0}, \ \forall j < k \\ & \sum_{k=1}^{K} [\boldsymbol{S}_k]_{n,n} \leq P_n, \ \forall n. \end{aligned} \quad (7)$$

We can transform (7) into an equivalent problem by removing the ZF constraints. First, for each k, define a matrix $\bar{\boldsymbol{H}}_k$ as

$$\bar{\boldsymbol{H}}_k = [\boldsymbol{H}_1^H \ \boldsymbol{H}_2^H \ \cdots \ \boldsymbol{H}_{k-1}^H]^H \in \mathbb{C}^{\sum_{i=1}^{k-1} n_i \times N}. \quad (8)$$

Let $\bar{\boldsymbol{V}}_k^{(0)} \in \mathbb{C}^{N \times \bar{n}_k}$, $\bar{n}_k = N - \sum_{i=1}^{k-1} n_i$, be a basis of $\mathcal{N}(\bar{\boldsymbol{H}}_k)$, which can be found, e.g., using the singular value decomposition or QR decomposition. Then, solving (7) amounts to solving the following optimization

$$\begin{aligned} \underset{\boldsymbol{\Omega}_k \succeq \boldsymbol{0}}{\text{maximize}} \quad & \sum_{k=1}^{K} \log |\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H| \\ \text{subject to} \quad & \sum_{k=1}^{K} [\bar{\boldsymbol{V}}_k^{(0)} \boldsymbol{\Omega}_k (\bar{\boldsymbol{V}}_k^{(0)})^H]_{n,n} \leq P_n, \ \forall n \end{aligned} \quad (9)$$

where $\boldsymbol{\Omega}_k \in \mathbb{C}^{\bar{n}_k \times \bar{n}_k}$, and $\tilde{\boldsymbol{H}}_k = \boldsymbol{H}_k \bar{\boldsymbol{V}}_k^{(0)} \in \mathbb{C}^{n_k \times \bar{n}_k}$ is called the effective channel matrix of the kth user. Problem (9) is a determinant maximization (MAXDET) program subject to linear matrix inequalities [12], which can be solved efficiently by standard numerical optimization tools, for example CVX [13] or YALMIP [14]. Regarding the relaxation technique, an important question to ask is whether the optimal solution to the relaxed problem is also optimal to the original problem. Interestingly, we now show that (6) and (9) are equivalent,

and, thus, the optimal solution to (9) is also optimal to (6). The proof is an intermediate consequence of the following lemma.

Lemma 1. *The optimal solutions* $\{\mathbf{\Omega}_k\}$ *to* (9) *satisfy* $\text{rank}(\mathbf{\Omega}_k) \le \min(n_k, \bar{n}_k)$.

Proof: Please refer to the Appendix. ∎

As stated earlier, problem (9) can be solved by a general purpose optimization package, but it does not offer useful insights into the optimal solutions. Herein, we develop a specialized algorithm to solve (9), using the barrier method. It is also worth pointing out that using the Lagrangian duality, [9] proposed a two-stage iterative algorithm for the precoder design of block diagonalization scheme based on a subgradient method, which can be applied to solve (9). Specifically, consider the Lagrangian function of (9), which is given by

$$\mathcal{L}(\mathbf{\Omega}_k, \boldsymbol{\lambda}) = \sum_{k=1}^{K} \log |\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H| \\ - \sum_{n=1}^{N} \lambda_n (\sum_{k=1}^{K} [\bar{\mathbf{V}}_k^{(0)} \mathbf{\Omega}_k (\bar{\mathbf{V}}_k^{(0)})^H]_{n,n} - P_n), \quad (10)$$

where λ_n is the dual variable associated with the nth per-antenna power constraint. Since strong duality holds for (9), its optimal solution can be found via the following convex-concave optimization problem

$$\underset{\boldsymbol{\lambda} \ge 0}{\text{minimize}} \; \underset{\mathbf{\Omega}_k \succeq \mathbf{0}}{\text{maximize}} \; \mathcal{L}(\mathbf{\Omega}_k, \boldsymbol{\lambda}). \quad (11)$$

The two-stage iterative algorithm in [9] works as follows. For a set of fixed dual variables $\{\lambda_n\}$, the set of covariance matrices $\{\mathbf{\Omega}_k\}$ that maximizes $\mathcal{L}(\mathbf{\Omega}_k, \boldsymbol{\lambda})$ can be obtained by the water-filling algorithm. Next, for a set of given $\{\mathbf{\Omega}_k\}$, the iterative algorithm updates $\{\lambda_n\}$ to minimize $\mathcal{L}(\mathbf{\Omega}_k, \boldsymbol{\lambda})$ based on the subgradient method. However, the convergence rate of the subgradient method for the minimization problem over $\boldsymbol{\lambda}$ is generally slow. It is well known that for a minimax optimization problem, the infeasible-start Newton's method [15] that solves the maximization and the minimization at the same time has a faster convergence rate [11], [15]. In the following, we propose a numerical algorithm to solve (9), which exhibits a better convergence behavior.

First, we observe that the constraints in (9) are active at the optimum. As proof, suppose the ith constraint is inactive, i.e. $\sum_{k=1}^{K} [\bar{\mathbf{V}}_k^{(0)} \mathbf{\Omega}_k (\bar{\mathbf{V}}_k^{(0)})^H]_{i,i} = [\mathbf{\Omega}_1]_{i,i} + \sum_{k=2}^{K} [\bar{\mathbf{V}}_k^{(0)} \mathbf{\Omega}_k (\bar{\mathbf{V}}_k^{(0)})^H]_{i,i} < P_i$. There exists $\epsilon > 0$ to be small enough such that $[\mathbf{\Omega}_1]_{i,i} + \sum_{k=2}^{K} [\bar{\mathbf{V}}_k^{(0)} \mathbf{\Omega}_k (\bar{\mathbf{V}}_k^{(0)})^H]_{i,i} + \epsilon < P_i$. Replacing $[\mathbf{\Omega}_1]_{i,i}$ by $([\mathbf{\Omega}_1]_{i,i} + \epsilon)$ yields a larger objective value in (9), which contradicts the assumption that $\{\mathbf{\Omega}_k\}$ are optimal. This observation is computationally useful since the inequality constraints in (9) can be replaced by equality ones, which are easier to be handled. In other words,

(9) is equivalent to

$$\begin{aligned} \text{maximize} \quad & \sum_{k=1}^{K} \log |\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H| \\ \text{subject to} \quad & \sum_{k=1}^{K} [\bar{\mathbf{V}}_k^{(0)} \mathbf{\Omega}_k (\bar{\mathbf{V}}_k^{(0)})^H]_{n,n} = P_n, \; \forall n \quad (12) \\ & \mathbf{\Omega}_k \succeq \mathbf{0}, \; \forall k. \end{aligned}$$

The proposed algorithm is based on the barrier method to solve (12). As a standard step, we define a modified objective function

$$f(t, \{\mathbf{\Omega}_k\}) = -(t \sum_{k=1}^{K} \log |\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H| + \sum_{k=1}^{K} \log |\mathbf{\Omega}_k|), \quad (13)$$

where $\log |\mathbf{\Omega}_k|$ is the logarithmic barrier function to account for the positive semidefinite constraint $\mathbf{\Omega}_k \succeq \mathbf{0}$, and $t > 0$ is a parameter that controls the logarithmic barrier terms. For mathematical convenience, let

$$\mathbf{A}_k^{(n)} = (\bar{\mathbf{V}}_k^{(0)})^H \text{diag}(\underbrace{0, \dots, 0}_{n-1}, 1, \underbrace{0, \dots, 0}_{N-n}) \bar{\mathbf{V}}_k^{(0)}, \quad (14)$$

and consider a standard equality constrained minimization problem

$$\begin{aligned} \text{minimize} \quad & f(\{\mathbf{\Omega}_k\}, t) \\ \text{subject to} \quad & \sum_{k=1}^{K} \text{tr}[\mathbf{A}_k^{(n)} \mathbf{\Omega}_k] = P_n, \; \forall n. \end{aligned} \quad (15)$$

The general idea of a barrier method is that, for a fixed t, we find the optimal solutions $\{\mathbf{\Omega}_k(t)\}$ to (15) (which is known as the centering step), and increase t until the algorithm converges. In our work, we employ the infeasible start Newton method to find the optimal solutions to (15). The use of the infeasible start Newton method also simplifies the initialization of $\{\mathbf{\Omega}_k\}$. We start with the optimal conditions (i.e. the KKT conditions) for (15), which are given by

$$-t\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k (\mathbf{I} + \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k)^{-1} - \mathbf{\Omega}_k^{-1} + \sum_{n=1}^{N} \mu_n \mathbf{A}_k^{(n)} = \mathbf{0}, \forall k \; (16)$$

$$\sum_{k=1}^{K} \text{tr}[\mathbf{A}_k^{(n)} \mathbf{\Omega}_k] = P_n, \forall n \quad (17)$$

where μ_n are the dual variables. In (17) we have used the fact that the gradient of $\log |\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H|$ with respect to $\mathbf{\Omega}_k$ is given by $\nabla_{\mathbf{\Omega}_k} \log |\mathbf{I} + \tilde{\mathbf{H}}_k \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H| = \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k (\mathbf{I} + \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k)^{-1}$. The crucial part in a Newton method is to calculate the Newton step. To do this, we replace $\mathbf{\Omega}_k$ by $\mathbf{\Omega}_k + \triangle \mathbf{\Omega}_k$ and μ_n by $\mu_n + \triangle \mu_n$ in the KKT conditions, which yields the KKT system for the Newton step

$$t\tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k (\mathbf{I} + \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k + \triangle \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k)^{-1} \\ + (\mathbf{\Omega}_k + \triangle \mathbf{\Omega}_k)^{-1} - \sum_{n=1}^{N} (\mu_n + \triangle \mu_n) \mathbf{A}_k^{(n)} = \mathbf{0}, \; \forall k \; (18)$$

$$\sum_{k=1}^{K} \text{tr}[\mathbf{A}_k^{(n)} \triangle \mathbf{\Omega}_k] = P_n - \sum_{k=1}^{K} \text{tr}[\mathbf{A}_k^{(n)} \mathbf{\Omega}_k], \; \forall n. \quad (19)$$

Denote $\tilde{\mathbf{\Omega}}_k = \mathbf{I} + \mathbf{\Omega}_k \tilde{\mathbf{H}}_k^H \tilde{\mathbf{H}}_k$. Since $\tilde{\mathbf{\Omega}}_k$ is invertible, and $(\mathbf{A} + \mathbf{B})^{-1} \approx \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$ for small $\mathbf{B}$, we can write

(18) as

$$t\tilde{\boldsymbol{H}}_k^H \tilde{\boldsymbol{H}}_k(\tilde{\boldsymbol{\Omega}}_k^{-1} - \tilde{\boldsymbol{\Omega}}_k^{-1}\triangle\boldsymbol{\Omega}_k\tilde{\boldsymbol{H}}_k^H \tilde{\boldsymbol{H}}_k\tilde{\boldsymbol{\Omega}}_k^{-1})$$

$$+(\boldsymbol{\Omega}_k^{-1} - \boldsymbol{\Omega}_k^{-1}\triangle\boldsymbol{\Omega}_k\boldsymbol{\Omega}_k^{-1}) - \sum_{n=1}^{N}(\mu_n + \triangle\mu_n)\boldsymbol{A}_k^{(n)} = \boldsymbol{0} \quad (20)$$

or equivalently as

$$t\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}\triangle\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}_k\boldsymbol{\Omega}_k + \triangle\boldsymbol{\Omega}_k$$

$$= t\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}_k\boldsymbol{\Omega}_k + \boldsymbol{\Omega}_k - \sum_{n=1}^{N}(\mu_n + \triangle\mu_n)\boldsymbol{\Omega}_k\boldsymbol{A}_k^{(n)}\boldsymbol{\Omega}_k, \ \forall k \quad (21)$$

where $\dot{\boldsymbol{H}}_k = \tilde{\boldsymbol{H}}_k^H \tilde{\boldsymbol{H}}_k\tilde{\boldsymbol{\Omega}}_k^{-1}$. The traditional way to find $\{\triangle\boldsymbol{\Omega}_k\}$ and $\{\triangle\mu_n\}$ is to vectorize $\triangle\boldsymbol{\Omega}_k$ as a vector $\boldsymbol{\omega}_k$ of length $\bar{n}_k(\bar{n}_k+1)/2$ for each k, transform (19) and (21) into a form of a linear system of $(N+\sum_{k=1}^{K} \bar{n}_k(\bar{n}_k+1)/2)$ variables, and use a generic method to solve the resulting linear system. However, the complexity of such a method is $\mathcal{O}(N^6)$. In this paper, we present a low-complexity method to find $\{\triangle\boldsymbol{\Omega}_k\}$ and $\{\triangle\mu_n\}$, using block elimination method [15]. Specifically, we express $\triangle\boldsymbol{\Omega}_k$ as

$$\triangle\boldsymbol{\Omega}_k = \boldsymbol{\Xi}_k^{(0)} + \sum_{i=1}^{N}\triangle\mu_i\boldsymbol{\Xi}_k^{(i)}. \quad (22)$$

Substituting (22) into (21) yields a system of $N + 1$ discrete-time Sylvester equations

$$t\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}\boldsymbol{\Xi}_k^{(0)}\dot{\boldsymbol{H}}_k\boldsymbol{\Omega}_k + \boldsymbol{\Xi}_k^{(0)}$$

$$= t\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}_k\boldsymbol{\Omega}_k + \boldsymbol{\Omega}_k - \sum_{n=1}^{N}\mu_n\boldsymbol{\Omega}_k\boldsymbol{A}_k^{(n)}\boldsymbol{\Omega}_k, \quad (23)$$

$$t\boldsymbol{\Omega}_k\dot{\boldsymbol{H}}\boldsymbol{\Xi}_k^{(i)}\dot{\boldsymbol{H}}_k\boldsymbol{\Omega}_k + \boldsymbol{\Xi}_k^{(i)} = -\boldsymbol{\Omega}_k\boldsymbol{A}_k^{(i)}\boldsymbol{\Omega}_k,$$

$$\text{for } 1 \leq i \leq N. \quad (24)$$

Numerical methods to solve the discrete-time Sylvester equations in (23) and (24) with complexity $\mathcal{O}(\bar{n}_k^3)$ can be found e.g. in [16]. To compute the Newton step for the dual variable $\boldsymbol{\mu}$, we plug $\triangle\boldsymbol{\Omega}_k$ from (22) into (19), which results in a linear system

$$\boldsymbol{\Psi}\triangle\boldsymbol{\mu} = \boldsymbol{\psi}, \quad (25)$$

where $[\boldsymbol{\Psi}]_{i,j} = \sum_{k=1}^{K}\mathrm{tr}(\boldsymbol{A}_k^{(i)}\boldsymbol{\Xi}_k^{(j)})$, and $[\boldsymbol{\psi}]_i = P_i - \sum_{k=1}^{K}\mathrm{tr}(\boldsymbol{A}_k^{(i)}\boldsymbol{\Xi}_k^{(0)} + \boldsymbol{A}_k^{(i)}\boldsymbol{\Omega}_k)$. We define the residual norm at $\{\boldsymbol{\Omega}_k\}$ and $\boldsymbol{\mu}$, which is used in the backtracking line search, as

$$r(\{\boldsymbol{\Omega}_k\}, \boldsymbol{\mu}) = \sum_{k=1}^{K} || - t\tilde{\boldsymbol{H}}_k^H \tilde{\boldsymbol{H}}_k(\boldsymbol{I} + \boldsymbol{\Omega}_k\tilde{\boldsymbol{H}}_k^H \tilde{\boldsymbol{H}}_k)^{-1}$$

$$- \boldsymbol{\Omega}_k^{-1} + \sum_{n=1}^{N}\mu_n\boldsymbol{A}_k^{(n)}||_F + ||\boldsymbol{v}||_2, \quad (26)$$

where $v_i = P_i - \sum_{k=1}^{K}\mathrm{tr}[\boldsymbol{A}_k^{(i)}\boldsymbol{\Omega}_k]$ for $i = 1, 2, \ldots, N$. The proposed algorithm to solve (9) is outlined in Algorithm 1. The proposed algorithm can be easily modified to be a primal-dual interior point method which often shows faster convergence rate. However, we skip the details for the sake of simplicity.

Algorithm 1 The proposed numerical algorithm to solve (9)

Initialization: $\boldsymbol{\Omega}_k = I_{\bar{n}_k}$, $\boldsymbol{\mu} = \boldsymbol{0}$, $t = t_0$, and γ and tolerance $\epsilon > 0$

1: **repeat** {Outer iteration}
2:   **repeat** {Inner iteration (centering step)}
3:     Compute the Newton step $\triangle\boldsymbol{\Omega}_k$ and dual step $\triangle\boldsymbol{\mu}$ from (22) and (25), respectively
4:     Backtracking line search
6:       $s = 1$
7:       **while** $(r(\{\boldsymbol{\Omega}_k\} + s\{\triangle\boldsymbol{\Omega}_k\}, \boldsymbol{\mu} + s\triangle\boldsymbol{\mu}) > (1 - \alpha s)r(\{\boldsymbol{\Omega}_k\}, \boldsymbol{\mu}))$ **and** $(\{\boldsymbol{\Omega}_k\} + s\{\triangle\boldsymbol{\Omega}_k\} \succeq 0)$ **do**
8:         $s = \beta s$
9:       **end while**
10:      Update primal and dual variables: $\boldsymbol{\Omega}_k = \boldsymbol{\Omega}_k + s\triangle\boldsymbol{\Omega}_k$; $\boldsymbol{\mu} = \boldsymbol{\mu} + s\triangle\boldsymbol{\mu}$
11: **until** $r(\{\boldsymbol{\Omega}_k\}, \boldsymbol{\mu}) < \epsilon$
12: Increase t. $t = \gamma t$
13: **until** t is sufficiently large to tolerate the duality gap.

B. Suboptimal designs

In general, it is of practial interest to find suboptimal designs that can achieve a significant fraction of the optimal performance, but require lower complexity. In this subsection we propose two suboptimal precoder designs, and briefly discuss their computational complexity, compared to (9).

The first suboptimal precoder design is derived from the QR decomposition (QRD). First, stack the channel matrix of all users in a matrix $\boldsymbol{H}$ defined as

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1^H & \boldsymbol{H}_2^H & \cdots & \boldsymbol{H}_K^H \end{bmatrix}^H \in \mathbb{C}^{n_R \times N}, \quad (27)$$

where $n_R = \sum_{k=1}^{K} n_k$ is the total number of receive antennas, and consider a QRD of $\boldsymbol{H}$ given as

$$\boldsymbol{H} = \boldsymbol{L}\boldsymbol{Q} \quad (28)$$

where $\boldsymbol{L}$ is a lower triangular matrix. We can further partition $\boldsymbol{Q}$ into

$$\boldsymbol{Q} = [\boldsymbol{Q}_1 \, \boldsymbol{Q}_2 \cdots \boldsymbol{Q}_K]^H \quad (29)$$

where $\boldsymbol{Q}_k \in \mathbb{C}^{N \times n_k}$ satisfies $\boldsymbol{H}_j\boldsymbol{Q}_k = \boldsymbol{0}$, $\forall j < k$, i.e. $\bar{\boldsymbol{H}}_k\boldsymbol{Q}_k = \boldsymbol{0}$, and $\boldsymbol{Q}_k^H\boldsymbol{Q}_k = \boldsymbol{I}_{n_k}$. To satisfy ZF constraints, the covariance matrices of the first suboptimal design is given by

$$\boldsymbol{S}_k = \boldsymbol{Q}_k\boldsymbol{\Omega}_k\boldsymbol{Q}_k^H, \quad (30)$$

where $\boldsymbol{\Omega}_k \in \mathbb{C}^{n_k \times n_k}$ is the solution of the following problem

$$\begin{aligned} \underset{\boldsymbol{\Omega}_k}{\text{maximize}} \quad & \sum_{k=1}^{K}\log|\boldsymbol{I} + \boldsymbol{H}_k\boldsymbol{Q}_k\boldsymbol{\Omega}_k\boldsymbol{Q}_k^H\boldsymbol{H}_k^H| \\ \text{subject to} \quad & \sum_{k=1}^{K}[\boldsymbol{Q}_k\boldsymbol{\Omega}_k\boldsymbol{Q}_k^H]_{n,n} \leq P_n, \ \forall n. \end{aligned} \quad (31)$$

Note that columns of $\boldsymbol{Q}_k$ does not span $\mathcal{N}(\bar{\boldsymbol{H}}_k)$. Thus, (31) is a suboptimal solution of (9). However, the dimension of each $\boldsymbol{\Omega}_k$ in (31) is just $n_k(n_k + 1)/2$, which is smaller than $\bar{n}_k(\bar{n}_k + 1)/2$ for each optimal $\boldsymbol{\Omega}_k$ in (9). Thus, solving (31) requires much lower complexity, especially when N is large, and n_k is small. The simulation results show that the sum-rate
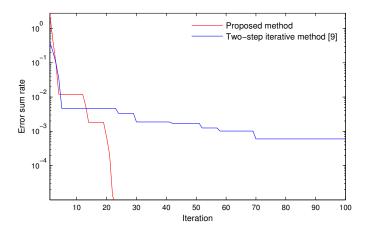
Fig. 1. Convergence behavior, $N = 6$, $n_k = 2$ for $k = 1, 2, 3$.

Fig. 2. Sum rate comparison of optimal and suboptimal designs for SZF-DPC schemes, $N = 8$, $K = 4$, $n_k = 2$, $\forall k = 1, 2, \ldots, 4$.

gap between the optimal design in (9) and suboptimal design in (31) is negligible.

The second suboptimal design follows a heuristic manner, which is obtained by scaling each column of the optimal precoders for the sum power constraint to meet the PAPCs. To be specific, let $\boldsymbol{W}_k^\star$ be the optimal precoder for the kth user with a sum power constraint, which can be analytically found using the water-filling algorithm [5]. Then a suboptimal precoder for the kth user can be given by $\boldsymbol{W}_k = \boldsymbol{W}_k^\star \operatorname{diag}(\sqrt{\boldsymbol{\alpha}_k})$, where $\boldsymbol{\alpha}_k = [\alpha_{k,1}\, \alpha_{k,2} \cdots \alpha_{k,n_k}]$, and $\alpha_{k,i}$ is a factor used to the scale the ith column of $\boldsymbol{W}_k^\star$ to satisfy the PAPCs. Mathematically, $\boldsymbol{\alpha}_k$ is found through solving the following problem

$$
\begin{aligned}
\underset{\boldsymbol{\alpha}_k \geq 0}{\operatorname{maximize}} \quad & \sum_{k=1}^{K} \log |\boldsymbol{I} + \boldsymbol{H}_k \boldsymbol{W}_k^\star \operatorname{diag}(\boldsymbol{\alpha}_k) \boldsymbol{W}_k^{\star H} \boldsymbol{H}_k^H| \\
\text{subject to} \quad & \sum_{k=1}^{K} [\boldsymbol{W}_k^\star \operatorname{diag}(\boldsymbol{\alpha}_k) \boldsymbol{W}_k^{\star H}]_{n,n} \leq P_n, \ \forall n.
\end{aligned}
\tag{32}
$$

This suboptimal design requires lower complexity, compared with the first one, since the dimension of each $\boldsymbol{\alpha}_k$ is just n_k.

IV. NUMERICAL RESULTS

In this section, we provide numerical examples to demonstrate the results in this paper. In the first numerical experiment, we compare the convergence rate of Algorithm 1 and the two-stage iterative method in [9]. A MIMO BC with $N = 6$ transmit antennas, $K = 3$ users, each with 2 receive antennas is simulated. The tolerance (for each centering step) is set to be $\epsilon = 10^{-5}$. The barrier method parameters t_0 and γ are set to 50 and 1, respectively. The effects of the initial value of γ and t_0 are discussed in [15]. The backtracking line search parameters in Algorithm 1 are $\alpha = 0.01$, and $\beta = 0.5$. Fig. 1 plots the error in the sum rate versus the number of iterations for a random realization of channel matrices. Note that, to update $\boldsymbol{\Omega}_k$ in each iteration of the two-stage iterative method, we need to compute the singular value decomposition (SVD) of an $n_k \times \bar{n}_k$ matrix, and the inverse of an $\bar{n}_k \times \bar{n}_k$ matrix, of which the complexity is $\mathcal{O}(\bar{n}_k^2 n_k)$ and $\mathcal{O}(\bar{n}_k^3)$, respectively. That is, the complexity of each iteration of the two-stage iterative method is of the same order as that of the proposed
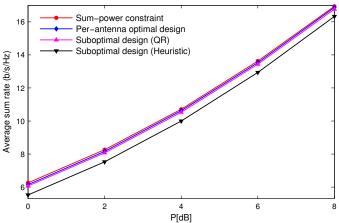
method. For the first iterations, the two-stage iterative method converges faster than the proposed method to the optimal solution. However, the proposed method performs better when precoders approach the optimum. Simulation results obtained with other randomly generated channel matrices illustrate the same convergence behavior of the two methods. Recall that a subgradient need not be a descent direction. Thus, an iteration can even increase the objective function. Moreover, the convergence rate of subgradient methods relies strongly on the problem size. For example, we observe that the two-stage iterative method fails to converge within three thousand iterations when $N = 16$, and $K = 8$, while Algorithm 1 still converges to the optimal solution within tens of iterations. As a conclusion, the proposed numerical algorithm demonstrates better convergence rate than the two-stage iterative algorithm in [9].

In Fig. 2, we plot the average sum rate of optimal and suboptimal precoder design methods for SZF-DPC schemes as a function of P, the total transmit power. The resulting power constraint for each antenna (when considering the PAPCs) is $P_n = P/N$ for all n. A quasi-static fading model is used in our simulation, where independent realizations of $\boldsymbol{H}_k$ are generated as zero mean and unit variance complex Gaussian random variables for each snapshot. Fig. 2 considers a scenario with $N = 8$, $K = 4$, and $n_i = 2$ for $i = 1, 2, \ldots, 4$. The optimal precoder designs with per-antenna power constraints yield a slightly lower sum rate than those with the corresponding sum power constraint. The sum-rate gap between optimal and suboptimal designs are small, and decreases as P increases. The suboptimal design based on QRD performs slightly better than the suboptimal design based on a heuristic manner. We can expect that these precoder designs give the same performance as P approaches to infinity, since the equal power allocation is proved to be optimal in the high SNR regime.

V. CONCLUSIONS

This paper addresses the precoder design of SZF-DPC for MIMO BCs with PAPCs. The precoder design is first

formulated as a rank-constrained optimization problem. Then we consider a relaxed problem, which is obtained by dropping the rank constraints. For the SZF-DPC scheme, we prove that the relaxed and original problems are equivalent. More explicitly, we show that optimal solutions of the relaxed problem always satisfy the rank constraints. Next, we propose an efficient numerical method based on a barrier method to solve the relaxed problem. The proposed numerical method is shown to have a superior convergence behavior, compared with the two-stage iterative method based on the dual subgradient method.

APPENDIX
PROOF OF LEMMA 1

In this appendix, we prove that the rank of optimal solutions to (9) is less than or equal to L_k. We begin by reformulating (9) as

$$
\begin{aligned}
\underset{\{\boldsymbol{\Omega}_k\}}{\text{maximize}} \quad & \sum_{k=1}^{K} \log |\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H| \\
\text{subject to} \quad & \sum_{k=1}^{K} \text{tr}(\boldsymbol{\Omega}_k \boldsymbol{A}_k^{(n)}) \leq \bar{P}, \ \forall n \\
& \boldsymbol{\Omega}_k \succeq \boldsymbol{0}, \ \forall k
\end{aligned} \tag{33}
$$

where $\tilde{\boldsymbol{H}}_k = \boldsymbol{H}_k \bar{\boldsymbol{V}}_k^{(0)} \in \mathbb{C}^{n_k \times \bar{n}_k}$, and $\boldsymbol{A}_k^{(n)}$ is defined in (14). The Lagrangian function of (33) is given by

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Omega}_k, \boldsymbol{\lambda}, \boldsymbol{\Phi}_k) = & \sum_{k=1}^{K} \log |\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H| \\
& - \sum_{n=1}^{N} \lambda_n \left(\sum_{k=1}^{K} \text{tr}(\boldsymbol{\Omega}_k \boldsymbol{A}_k^{(n)}) - \bar{P} \right) + \text{tr}(\boldsymbol{\Phi}_k \boldsymbol{\Omega}_k) \quad (34)
\end{aligned}
$$

where $\{\lambda_n\}$ are dual variables associated with the per-antenna power constraints, and $\boldsymbol{\Phi}_k \succeq \boldsymbol{0}$ is the dual variable for the positive semidefinite constraint. Denote $\boldsymbol{P} = \text{diag}(\bar{P})$, $\boldsymbol{\Lambda} = (\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$, and $\boldsymbol{\Lambda}_k = (\bar{\boldsymbol{V}}_k^{(0)})^H \boldsymbol{\Lambda} \bar{\boldsymbol{V}}_k^{(0)}$. We can then rewrite (34) as

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Omega}_k, \boldsymbol{\lambda}, \boldsymbol{\Phi}_k) = & \sum_{k=1}^{K} \log |\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H| \\
& - \text{tr}[\boldsymbol{\Lambda}_k \boldsymbol{\Omega}_k - \boldsymbol{\Lambda} \boldsymbol{P}] + \text{tr}(\boldsymbol{\Phi}_k \boldsymbol{\Omega}_k) \quad (35)
\end{aligned}
$$

At the optimum, we have

$$
\tilde{\boldsymbol{H}}_k^H (\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H)^{-1} \tilde{\boldsymbol{H}}_k - \boldsymbol{\Lambda}_k + \boldsymbol{\Phi}_k = \boldsymbol{0} \tag{36}
$$

Using the complementary slackness property $\boldsymbol{\Phi}_k \boldsymbol{\Omega}_k = \boldsymbol{0}$, we obtain

$$
\tilde{\boldsymbol{H}}_k^H (\boldsymbol{I} + \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k \tilde{\boldsymbol{H}}_k^H)^{-1} \tilde{\boldsymbol{H}}_k \boldsymbol{\Omega}_k = \boldsymbol{\Lambda}_k \boldsymbol{\Omega}_k \tag{37}
$$

We now show that the dual optimal variables of (9) are strictly positive, $\lambda_n > 0$ for all $1 \leq n \leq N$. As proof, consider the dual objective of (9), which can be expressed as,

$$
g(\boldsymbol{\lambda}, \boldsymbol{\Phi}_k) = \max_{\boldsymbol{\Omega}_k \succeq \boldsymbol{0}} \mathcal{L}(\boldsymbol{\Omega}_k, \boldsymbol{\lambda}, \boldsymbol{\Phi}_k) \tag{38}
$$

By contradiction, suppose $\lambda_i = 0$ for some $1 \leq i \leq N$. We construct a set of $\boldsymbol{\Omega}_k$ such that $\boldsymbol{\Omega}_k = \text{diag}(\underbrace{0, \ldots, 0}_{i-1}, \alpha, \underbrace{0, \ldots, 0}_{N-i})$, and $\boldsymbol{\Omega}_k = \boldsymbol{0}$ for $2 \leq k \leq K$. Then, the objective function in (38) becomes

$$
\mathcal{L}(\boldsymbol{\Omega}_k, \boldsymbol{\lambda}, \boldsymbol{\Phi}_k) = \log(1 + \alpha \|[\tilde{\boldsymbol{H}}_1]_i\|_2^2) + \text{tr}(\boldsymbol{\Phi}_1 \boldsymbol{\Omega}_1), \tag{39}
$$

where $[\tilde{\boldsymbol{H}}_1^H]_i$ is the ith column of $\tilde{\boldsymbol{H}}_1^H$. We can see that the objective function in (39) is unbounded above if $\alpha \to \infty$. Since we are only interested in the case where $g(\boldsymbol{\lambda}, \boldsymbol{\Phi}_k)$ is finite, we conclude that $\lambda_i > 0$ for all $1 \leq i \leq N$. Thus, $\boldsymbol{\Lambda}$ must be positive definite, and $\boldsymbol{\Lambda}_k$ is invertible. It follows from (37) that $\text{rank}(\boldsymbol{\Omega}_k^\star) \leq \text{rank}(\tilde{\boldsymbol{H}}_k) = L_k$, which completes the proof.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun*, vol. 10, pp. 585–598, Nov. 1999.

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

[3] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[4] W. Yu, "Sum-capacity computation for the Gaussian vector broadcast channel via dual decomposition," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 754–759, Feb. 2006.

[5] A. Dabbagh and D. Love, "Precoding for multiple antenna Gaussian broadcast channels with successive zero-forcing," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3837–3850, Jul. 2007.

[6] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[7] Le-Nam Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Successive zero-forcing DPC with sum power constraint: Low-complexity optimal precoders," *IEEE ICC 2012*, to appear.

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

[9] 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.

[10] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," in *IEEE SPAWC 2006*, 2006, pp. 1–5.

[11] 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.

[12] L. Vandenberghe, S. Boyd, and S.-P. Wu, "Determinant maximization with linear matrix inequality constraints," *SIAM J. on Matrix Anal. and Appl.*, vol. 19, no. 2, pp. 499–533, 1998.

[13] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," http://cvxr.com/cvx, Apr. 2011.

[14] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan., 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[15] S. Bold and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed. Society for Industrial and Applied Mathematics, 2002.