Decomposition by Successive Convex Approximation: A Unifying Approach for Linear Transceiver Design in Heterogeneous Networks

Mingyi Hong, Qiang Li, Ya-Feng Liu

Abstract

In this work, we study the downlink linear precoder design problem in a multi-cell heterogeneous network (HetNet), in which macro/pico base stations (BSs) are densely deployed within each cell. The problem is formulated in a very general setting as the users’ sum-utility maximization problem. Our formulation includes many practical precoder design problems such as multi-cell coordinated linear precoding, full and partial per-cell coordinated multi-point (ComP) transmission, zero-forcing precoding and joint BS clustering and beamforming/precoding as special cases.

The sum-utility maximization problem is difficult due to its non-convexity and the tight coupling of the users’ precoders. We propose a novel convex approximation technique to approximate the original problem by a series of convex subproblems, each of which decomposes across all the cells. The convexity of the subproblems allows for efficient computation, while their decomposability leads to distributed implementation. Our approach is made possible by the identification of certain key convexity properties of the sum-utility objective. Simulation experiments show that the proposed framework is quite effective for solving interference management problems in many practical settings.

I. INTRODUCTION

Heterogeneous network (HetNet) has recently emerged as a promising wireless network architecture capable of accommodating the explosive demand for wireless data [1]. In HetNet, the upper tier high-
power BSs such as Macro BSs provide per-cell interference management and blanket coverage, while the lower tier low-power access points such as micro/pico/femto BSs are densely deployed to provide capacity extension. This new paradigm of network design brings the transmitters and receivers closer to each other, thus is able to provide high link quality with low transmission power [2].

Due to the large number of potential interfering nodes in the network, one of the key challenges in the design of the HetNet is to properly mitigate both the inter-cell and intra-cell multiuser interference. Interference management for the HetNet, or for general interference networks, has been under extensive research recently [3], [4]. Introducing appropriate coordination among the nodes in the network, either in the physical layer or in higher layers, is effective for such purpose [5]. For the network with the nodes equipped with multiple antennas, major approaches for physical layer coordinated interference management include coordinated beamforming (CB) and joint processing (JP) [5]. The CB approach allows the nodes to coordinate in the beamformer/precoder level [6]–[8]. The JP approach, also known as the coordinated multi-point (ComP) transmission, optimizes the transceiver structures assuming that the users’ data are available at all the BSs. For example, in a downlink network, a single virtual BS can be formed if the users’ data is shared among all transmitting BSs. In this way, transmission schemes designed for single cell MIMO broadcasting channels can be used for coordinated transmission. However, due to the high signaling overhead associated with implementing the full JP, dynamic combination of JP and CB is usually adopted. For example, JP can be performed in a per-cell basis to cancel intra-cell interference, while the CB is used to mitigate inter-cell interference [5], [9].

Interference management is usually formulated as problems that optimize certain system utility functions, which are directly related to the users’ individual rates [4]. These utility functions, when chosen properly, can well balance the network spectrum efficiency and the fairness level among the users. Unfortunately, for a large class of utility functions, the associated optimization problems are difficult to solve (except for a few special cases, see [8], [10]–[13]). Therefore, many low-complexity algorithms that compute high-quality solutions to the interference management problems have been recently developed. The key to designing practical algorithms for interference management is to recognize certain convexity and/or decomposability of the underlying utility maximization problems. Convexity leads to efficient computation, while decomposability is crucial for distributed implementation [14].

Despite the nonconvexity of the overall utility maximization problems, researchers have long been exploring hidden convexity structures in order to tackle the problem. For example, reference [15] is the first to recognize that in a MIMO interfering channel (IC), individual users’ achievable rate is concave in its own transmit covariance while at the same time convex in all other interfering users’ transmit...
covariances. Similar concave-convex properties have since been leveraged heavily to design resource allocation algorithms in various network settings [6], [16]–[19]. The convex-concave property allows one to obtain, for each individual user, an approximated version of the objective by linearizing its convex part while keeping its concave part unchanged. In this way, the users can successively optimize their transmit strategies by solving a series of convex subproblems. Different convex approximation approaches that are not based on the convex-concave properties are also possible, see e.g., [8], [20]–[23]. For example reference [23] proposes to first approximate the sum rate problem using certain Geometric Program (GP), and then use subgradient method to solve the resulting problem. In a series of recent works, the authors of [21], [22], [24] propose a very general parallel decomposition scheme for distributed optimization in multi-agent systems, including interfering wireless networks as a special case. However, most the schemes cited above (except for the parallel scheme [21], [22], [24]) do not decompose well across the nodes: for the schemes based on the convex-concave structure, the convexification procedures can only be done one node at each time, thus only a single node can update its transmit strategy in each iteration; for the algorithm proposed in [20], the convex subproblem is still coupled among all the users. Moreover, most of these algorithms are designed to handle peer-to-peer networks, with each transmitter dedicated to transmit to a single receiver, and/or each receiver receives from a single transmitter. Hence they may not be suitable for the HetNet setting where each BS can transmit to multiple users while at the same time each user can receive from multiple BSs as well. There are a few recent works that have attempted to address these drawbacks, e.g., [25]. However, there is no theoretical convergence analysis for these algorithms.

Nevertheless, decomposability structure of interference management problems is highly desirable. When judiciously exploited, it can lead to efficient distributed implementation. This fact has long been recognized for other important large-scale network optimization problems such as the network utility maximization (NUM) problems; see [14]. However, unlike the NUM problems, in the interfering networks the nodes are tightly coupled in a nonlinear manner through multi-user interference. As a result, even in simple networks with single antenna transceiver pairs communicating simultaneously, decomposability structure is difficult to come by. Recently, a weighted minimum mean square error (WMMSE) algorithm is proposed in [26] for general utility optimization problems in interfering broadcast channels (IBC). In each of its steps, the computation is completely decoupled among the interfering BSs in network. Stationary solutions of the latter problem can then be obtained via solving three subproblems alternately. In [21], [24] a different decomposition technique based on partial linearization is proposed. The scheme can also decompose a sum rate maximization problem in an IBC channel to the BSs in the network. However, it is not clear how
these algorithms can apply to general HetNet setting with other cooperation strategies and/or physical layer beamforming schemes. In this work, we propose to achieve decomposability by means of *successive convex approximation*. Central to our approach is a key observation that reveals certain hidden convexity for a wide range of sum-utility maximization problems. The identified property allows us to approximate the original non-convex problem by a series of convex subproblems, each of which is completely decoupled among the nodes in the network. Based on different ways of solving the convex subproblems, two low-complexity algorithms are proposed, each having wide applicability in interference management problems.

The rest of the paper is organized as follows. In Section II, we outline a general system model for interference management in HetNet, and describe its wide applicability. In Section III, we present a key convexity structure of the considered utility maximization problem, which leads to two general successive convex approximation algorithms. In Section IV, the proposed algorithms are specialized to various interference management scenarios. Numerical results are given in Section V and concluding remarks are provided in Section VI.

*Notations:* For a symmetric matrix $X$, $X \succeq 0$ signifies that $X$ is Hermitian positive semi-definite. We use $\text{Tr}[X]$, $|X|$, $X^H$, $\rho(X)$, $\|X\|_F$ and $\text{Rank}(X)$ to denote the trace, determinant, Hermitian, spectral radius, Frobenius norm, and the rank of a matrix, respectively. We use $\langle \cdot, \cdot \rangle$ to denote the inner product operation. The $(m, n)$-th element of a matrix $X$ is denoted by $X[m, n]$. We use $I_n$ to denote an $n \times n$ identity matrix. Moreover, we let $\mathbb{R}^{N \times M}$ and $\mathbb{C}^{N \times M}$ denote the set of real and complex $N \times M$ matrices, and use $\mathbb{S}^N$, $\mathbb{S}_+^N$, $\mathbb{S}_{++}^N$ to denote the set of $N \times N$ Hermitian, Hermitian positive semi-definite and Hermitian positive definite matrices, respectively.

**II. System Model and Problem Formulation**

We consider a downlink multi-cell HetNet which consists of a set $\mathcal{K} := \{1, \cdots, K\}$ of cells. Within each cell $k$ there is a set of $\mathcal{Q}_k := \{1, \cdots, Q_k\}$ distributed base stations (BS) such as macro/micro/pico BSs which provide service to users located in different areas of the cell. Assume that in each cell $k$, there is a low-latency backhaul network connecting the set of BSs $\mathcal{Q}_k$ to a central controller (usually the macro BS), who makes the resource allocation decisions for all BSs within the cell. Furthermore, this central entity has access to the data signals of all the users in its cell. Let $\mathcal{I}_k := \{1, \cdots, I_k\}$ denote the users associated with cell $k$. Each of the users $i_k \in \mathcal{I}_k$ is served jointly by a subset of BSs in $\mathcal{Q}_k$. Let $\mathcal{I}$ and $\mathcal{Q}$ denote the set of all the users and all the BSs, respectively. Assume that each BS has $M$ transmit antennas, and each user has $N$ receive antennas. Let $H_{i_k}^{\mu_l} \in \mathbb{C}^{N \times M}$ denote the channel matrix between

August 19, 2014 DRAFT
the \( q \)-th BS in the \( \ell \)-th cell and the \( i \)-th user in the \( k \)-th cell. Similarly, we use \( \mathbf{H}_{ik}^\ell \) to denote the channel matrix between all the BSs in the \( \ell \)-th cell to the user \( i_k \), i.e., \( \mathbf{H}_{ik}^\ell := \{ \mathbf{h}_{ik}^q \}_{q \in \mathcal{Q}_k} \in \mathbb{C}^{N \times MQ^k} \).

Suppose that it is possible to transmit \( d_{ik} \leq \min\{M, N\} \) parallel data streams to user \( i_k \). Let \( \mathbf{V}_{ik}^q \in \mathbb{C}^{M \times d_{ik}} \) denote the transmit precoder that BS \( q_k \) uses to transmit data \( \mathbf{s}_{ik} \in \mathbb{C}^{d_{ik}} \) to user \( i_k \). Define \( \mathbf{V}_{ik} := \{ \mathbf{V}_{ik}^q \}_{q \in \mathcal{Q}_k} \), \( \mathbf{V}^k := \{ \mathbf{V}_{ik}^q \}_{i \in \mathcal{I}_k} \) and \( \mathbf{V}^k := \{ \mathbf{V}_{ik} \}_{i \in \mathcal{I}_k} \) as the collection of all precoders intended for user \( i_k \), the collection of all the precoder belong to BS \( q_k \), and the collection of all precoders in cell \( k \), respectively. Let \( \mathbf{V} := \{ \mathbf{V}_{ik} \}_{i \in \mathcal{I}} \). Further assume that all the BSs in cell \( k \) form a single virtual BS that jointly transmit to user \( i_k \in \mathcal{I}_k \), then \( \mathbf{V}_{ik} \) can be viewed as the \textit{virtual precoder} for user \( i_k \).

Using the above definitions, we can express the transmitted signal of BS \( q_k \) as well as the combined transmitted signal for all the BSs in cell \( k \) as:

\[
\mathbf{x}^q = \sum_{i_k \in \mathcal{I}_k} \mathbf{V}_{ik}^q \mathbf{s}_{ik} \in \mathbb{C}^{M \times 1}, \quad \mathbf{x}^k = \sum_{i_k \in \mathcal{I}_k} \mathbf{V}_{ik} \mathbf{s}_{ik} \in \mathbb{C}^{MQ_k \times 1}.
\]

The received signal \( \mathbf{y}_{ik} \in \mathbb{C}^{N \times 1} \) of user \( i_k \) is

\[
\mathbf{y}_{ik} = \mathbf{H}_{ik}^\ell \mathbf{V}_{ik} \mathbf{s}_{ik} + \sum_{j_k \neq i_k} \mathbf{H}_{ik}^j \mathbf{V}_{jk} \mathbf{s}_{jk} + \sum_{\ell \neq k} \sum_{j \in \mathcal{L}_\ell} \mathbf{H}_{ik}^\ell \mathbf{V}_{j\ell} \mathbf{s}_{j\ell} + \mathbf{z}_{ik}
\]

where \( \mathbf{z}_{ik} \in \mathbb{C}^{N \times 1} \) is the additive white complex Gaussian noise with distribution \( \mathcal{CN}(0, \sigma^2_{ik} \mathbf{I}_N) \).

Let \( \mathbf{U}_{ik} \in \mathbb{C}^{N \times d_{ik}} \) denote the linear white complex Gaussian noise with distribution \( \mathcal{CN}(0, \sigma^2_{ik} \mathbf{I}_{N}) \).

The Minimum MSE (MMSE) receiver minimizes user \( i_k \)'s MSE, and can be expressed as \[27\]

\[
\mathbf{U}_{ik}^{\text{mmse}} = \left( \sum_{(\ell, j) \neq (k, i)} \mathbf{H}_{ik}^\ell \mathbf{V}_{j\ell} \mathbf{V}_{j\ell}^H (\mathbf{H}_{ik}^\ell)^H + \sigma^2_{ik} \mathbf{I}_N \right)^{-1} \mathbf{H}_{ik}^\ell \mathbf{V}_{ik}
\]

where \( \mathbf{C}_{ik} \in \mathbb{S}_+^N \) denotes user \( i_k \)'s received signal covariance matrix. When the MMSE receiver is used, the MMSE matrix \( \mathbf{U}_{ik}^{\text{mmse}} \) is reduced to

\[
\mathbf{E}_{ik}^{\text{mmse}} = \mathbf{I}_{d_{ik}} - \mathbf{V}_{ik}^H (\mathbf{H}_{ik}^\ell)^H \mathbf{C}_{ik}^{-1} \mathbf{H}_{ik}^\ell \mathbf{V}_{ik} \succeq 0.
\]

Clearly we also have \( \mathbf{I}_{d_{ik}} - \mathbf{E}_{ik}^{\text{mmse}} \succeq 0 \).
Let us assume that Gaussian signaling is used and the interference is treated as noise. The achievable rate for user $i_k$ is given by [28]

$$R_{i_k} = \log \left| I_N + H_{i_k}^{H} V_{i_k} H_{i_k}^{H} (H_{i_k}^{H})^{-1} Y_k \right| = -\log |E_{i_k}^{\text{mmse}}|$$  \hspace{1cm} (4)

where we have defined the matrix $Y_{i_k} := \sum_{(\ell,j) \neq (k,i)} H_{i_k}^{\ell} V_{j_k}^{\ell} H_{j_k}^{H} (H_{i_k}^{H})^{-1} + \sigma_{i_k}^2 I_N$ as the interference matrix for user $i_k$; the last equality is the well-known relationship between the transmission rate and the MMSE matrix (see e.g., [29]). We will occasionally use the notations $R_{i_k}(V)$, $C_{i_k}(V)$ and $E_{i_k}^{\text{mmse}}(V)$ to make their dependencies on $V$ explicit.

Let $f_{i_k} : \mathbb{R}_+ \rightarrow \mathbb{R}$ denote the utility function of user $i_k$’s data rate, and assume that it satisfies the following assumptions:

A-1) $f_{i_k}(x)$ is a concave non-decreasing function in $x$ for all $x \geq 0$;

A-2) $f_{i_k}(-\log(|X|))$ is convex in $X$, for all $I \succeq X \succeq 0$;

A-3) $f_{i_k}(x)$ is continuously differentiable (i.e., a smooth function).

Note that this family of utility functions includes well-known utilities such as the weighted sum rate, the geometric mean of one plus rate and the harmonic mean rate utility functions (see [26]). They differ considerably with those studied in references [30], [31] which, although admit concave representations, are not directly related to individual users’ rates.

Let $s_{i_k}^q(\cdot)$ denote a penalty term for the precoder $V_{i_k}^{q_k}$. As will be shown shortly, such penalization is useful in inducing certain structure on the precoders. We make the following assumption on $s_{i_k}^q(\cdot)$:

B-1) $s_{i_k}^q(V_{i_k}^{q_k})$ is a convex, continuous, but possibly nonsmooth function.

In this paper, we consider the general system-level sum utilities maximization problem in the following form

$$\max \quad u(V) := f(V) - s(V)$$ \hspace{1cm} (P_{\text{SYSTEM}})

s.t. \quad f(V) := \sum_{k \in K} \sum_{i_k \in I_k} f_{i_k}(R_{i_k}(V))$$

$$s(V) := \sum_{k \in K} \sum_{i_k \in I_k} \sum_{q_k \in Q_k} s_{i_k}^q(V_{i_k}^{q_k})$$

$$V_{i_k}^{q_k} \in \mathcal{Y}_{i_k}^{q_k}, \quad \forall \ q_k \in Q, \quad V_k \in \mathcal{Y}_k, \quad \forall \ k \in K$$

where $\mathcal{Y}_{i_k}^{q_k}$ and $\mathcal{Y}_k$ are the feasible sets for $V_{i_k}^{q_k}$ and $V_k$, respectively. Let $\mathcal{V}$ denote the feasible set for $V$. When the functions $f(V)$ and $s(V)$ as well as the feasible sets $\{\mathcal{Y}_{i_k}^{q_k}\}$ and $\{\mathcal{Y}_k\}$ are properly specified, the general problem (P_{\text{SYSTEM}}) can cover a wide range of transceiver design problems in...
multicell networks; see a few examples given below.

1) MIMO IBC/IMAC/IC channels with inter-BS CB [8], [19], [25], [29], [32]: Each cell $k$ has a single BS serving all the users $I_k$. In this case no penalty term $s(V)$ is needed in the objective, and the constraint set $\mathcal{V}^k$ becomes the same as the constraint set $\mathcal{V}^k$, which is given by the following sum-power constrained set ($\bar{P}_k$ denotes the power budget for cell $k$)

$$\mathcal{V}^k = \left\{ V^k : \sum_{i_k \in I_k} \text{Tr}[V_{i_k} V_{i_k}^H] \leq \bar{P}_k \right\}. \quad (5)$$

2) Multicell MIMO network with intra-cell ComP and inter-cell CB [33], [34]: The BSs in different cells cooperate in the precoder level, while the BSs in the same cell share the users’ data and perform joint transmission. Each BS $q_k$ has a separate power constraint ($\bar{P}_{q_k}$ denotes the power budget for BS $q_k$):

$$\mathcal{V}^{q_k} = \left\{ V^{q_k} : \sum_{i_k \in I_k} \text{Tr}[V_{i_k} V_{i_k}^{q_k}] \leq \bar{P}_{q_k} \right\}. \quad (6)$$

This model generalizes those for the IBC model (e.g., [25], [26]), in the sense that the sum-power constraint (5) is replaced by the set of per group of antennas power constraints.

3) Multicell MIMO network with intra-cell partial ComP and inter-cell CB [34], [35]: Performing full ComP in each cell can achieve huge improvement of the overall spectrum efficiency, while suffering from excessive overhead in the backhaul network [5]. A practical alternative is to implement a partial ComP strategy, in which each user is served by not all, but only a few BSs in each cell. In this case, the BSs in the same cell are grouped into different (possibly overlapping) clusters with small sizes, within which they fully cooperate for joint transmission. Clearly, besides precoder design, the cluster membership of the BSs needs to be decided. This task can be done jointly with precoder design by properly specifying the penalty term $s(V)$. The requirement that each user is served by a few BSs translates to the restriction that for each use $i_k \in I_k$, its precoder $V_{i_k} = \{V_{i_k}^{q_k}\}_{q_k \in Q_k}$ should contain only a few nonzero block components [34], [36]. To induce such block sparsity, the penalty term can take the form of

$$s_{i_k}^{q_k}(V_{i_k}^{q_k}) = \gamma_{i_k}^{q_k} \|V_{i_k}^{q_k}\|_F, \quad (7)$$

with $\gamma_{i_k}^{q_k} \geq 0$ being some constant. See [34], [36] and the references therein for motivation of using (7). The constraint set for this problem is the same as that in (6). Note that most of the existing decomposition based algorithms [8], [20]–[22] cannot handle such nonsmooth problem.

4) Multicell MIMO network with intra-cell ZF and inter-cell CB [9], [37]: In this setting, all the BSs in the same cell jointly perform ZF precoder, or equivalently the block-diagonalization (BD)
precoding \cite{38}, while the BSs in different cells perform CB. The feasible sets are given as

\[ V^k = \{ V^k : H^H_{j_k} V_{i_k} (V_{i_k})^H H_{j_k} = 0, \forall j_k \neq i_k, j_k, i_k \in I_k \} \]

\[ V^{q_k} = \{ V^{q_k} : \sum_{i_k \in I_k} \text{Tr}[V^{q_k}_{i_k} (V^{q_k}_{i_k})^H] \leq \bar{P}_{q_k}, \forall q_k \in Q_k \}. \]

Despite the wide applicability of (\text{SYSTEM}), solving it to its global optimality is often difficult, even without the nonsmooth penalty term \( s(V) \). In fact, a set of recent works \cite{8}, \cite{10}, \cite{12}, \cite{13} have rigorously established the level of difficulties of various forms of problem (\text{SYSTEM}). The main message is that except for a very few cases such as MISO/SIMO min utility maximization problem, solving the problem (\text{SYSTEM}) is generally NP-hard. We refer the readers to \cite{3} for a summary of these complexity results. The NP-hardness of the problem indicates that it is unlikely even to find an equivalent convex reformulation for it (unless P=NP). Therefore the best that one can do is to seek efficient algorithms that provide approximately optimal solutions. Unfortunately, due to the fact that the variables \( \{ V^k \}_{k=1}^K \) that belong to different cells are tightly coupled in the objective via multi-user interference, even finding efficient approximate algorithm for this family of problems can be challenging.

### III. A Successive Convex Approximation Approach

In this section, we will present our main approach for computing a high quality solution for the general problem (\text{SYSTEM}). Our approach is to solve, possibly in an inexact manner, a series of convex subproblems each of which is a local approximation of (\text{SYSTEM}). A desirable feature of our approach is that each of the obtained subproblems is not only convex with respect to the transmit precoder \( V \), but is completely separable among the precoders of different cells. Consequently, the computation of the resource allocation in different stages of the algorithm can be carried out in a parallel fashion by individual cells.

#### A. A Local Approximation for (\text{SYSTEM})

To approximate the objective function of (\text{SYSTEM}), we begin by deriving a simple local approximation for the individual users’ utility function \( f_{i_k}(\cdot) \) using the convexity assumed in assumption A-2). More specifically, let \( \hat{V} \in V \) be a feasible solution to problem (\text{SYSTEM}). Let \( \hat{E}_{i_k} := E^\text{mmse}_{i_k}(\hat{V}) \) denote the MMSE evaluated at \( \hat{V} \). From the relationship of the individual users’ rates and their MMSE matrices
we can express \( f_{ik}(\cdot) \) as a function of \( \mathbf{E}_{ik}^{\text{mmse}} \) only. In particular, we have

\[
\begin{align*}
  f_{ik}(R_{ik}(\mathbf{V})) &= f_{ik}(\log|\mathbf{E}_{ik}^{\text{mmse}}|) \\
  &\geq f_{ik}(\log|\mathbf{E}_{ik}|) - \frac{\partial f_{ik}(x)}{\partial x} \bigg|_{x=\log|\mathbf{E}_{ik}|} \text{Tr} \left[ \mathbf{E}_{ik}^{-1} \mathbf{E}_{ik}^{\text{mmse}} - \mathbf{E}_{ik} \right] \\
  &\quad - \beta^k \text{Tr} \left[ (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik}) (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik})^H \right] \\
  &=: \tilde{a}_{ik} - \tilde{c}_{ik} \text{Tr} \left[ \mathbf{E}_{ik}^{-1} \left( \mathbf{E}_{ik}^{\text{mmse}} - \mathbf{E}_{ik} \right) \right] \\
  &\quad - \beta^k \text{Tr} \left[ (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik}) (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik})^H \right] \\
  &=: \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \tag{8}
\end{align*}
\]

where \( \beta^k \geq 0 \) is some constant that can be tuned by the algorithm; \( \tilde{c}_{ik} \) and \( \tilde{a}_{ik} \) are two constants that are not related to either \( \mathbf{E}_{ik}^{\text{mmse}} \) or \( \mathbf{V} \); we have \( \tilde{c}_{ik} \geq 0 \) due to the non-decreasing property of \( f_{ik}(\cdot) \)’s assumed in A-1); the inequality in (i) is due to the fact that the convex function is always lower bounded by its local linear approximation, and the nonnegativity of the following term

\[
\text{Tr} \left[ (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik}) (\mathbf{V}_{ik} - \mathbf{\hat{V}}_{ik})^H \right] \geq 0. \tag{9}
\]

Note that the above is a proximal quadratic term, adding which to the lower bound provides more flexibility in the algorithm design stage, as will be seen shortly. Note that the function \( \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \) is a locally approximated version of \( f_{ik}(\cdot) \) at the point \( (\mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \). In fact, the approximation is a global underestimator of \( f_{ik}(\cdot) \), and it is tight at the point \( (\mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \). That is, for all feasible \( (\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik}) \) and feasible \( (\mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \), and all \( \beta^k \geq 0 \), we have

\[
\begin{align*}
  \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) &\leq f_{ik}(\log|\mathbf{E}_{ik}^{\text{mmse}}|), \\
  \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) &= f_{ik}(\log|\mathbf{E}_{ik}|).
\end{align*}
\]

Unfortunately, the above approximation does not simplify the problem. Although the sum of the derived approximated functions, \( \sum_{k \in \mathcal{K}} \sum_{i_k \in I_k} \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \), can be viewed as a locally approximated version of \( f(\mathbf{V}) \), it is still non-convex with respect to the system precoder \( \mathbf{V} \) (cf. (3)).

Our next step is therefore to perform a second-stage approximation, in which we further approximate \( \tilde{h}_i^{\beta_k}(\mathbf{E}_{ik}^{\text{mmse}}, \mathbf{V}_{ik} : \mathbf{E}_{ik}, \mathbf{\hat{V}}_{ik}) \) by a concave function of \( \mathbf{V} \). To this end, we need a key lemma that explores some hidden convexity property of the function \( \tilde{h}_i^{\beta_k}(\cdot) \).
Lemma 1: The following function

$$\text{Tr} \left[ \hat{E}^{-1}_{ik} V^H_i (H^k_{ik})^H C^{-1}_{ik} H^k_{ik} V_{ik} \right]$$

is jointly convex with respect to the pair of variables \((V_{ik}, C_{ik})\) in the feasible region \((\mathbb{C}^{MQ_k \times d_{ik}}, S^N_{++})\).

Proof: Let us define

$$l_{ik}(V_{ik}, C_{ik}) := \text{Tr} \left[ \hat{E}^{-1}_{ik} V^H_i (H^k_{ik})^H C^{-1}_{ik} H^k_{ik} V_{ik} \right].$$

Consider the epigraph of \(l_{ik}(V_{ik}, C_{ik})\), i.e.,

$$\{(V_{ik}, C_{ik}, t) \mid l_{ik}(V_{ik}, C_{ik}) \leq t\}.$$ 

It suffices to show that the epigraph \((V_{ik}, C_{ik}, t)\) is a convex set [39, Chapter 3]. To this end, let us consider the following extended set (with \(Z_{ik} \succeq 0\) being a slack variable):

$$\{(V_{ik}, C_{ik}, Z_{ik}, t) \mid \text{Tr}[Z_{ik}] \leq t, Z_{ik} \succeq 0, Z_{ik} \succeq \hat{E}^{-1/2}_{ik} V^H_i (H^k_{ik})^H C^{-1}_{ik} H^k_{ik} V_{ik} \hat{E}^{-1/2}_{ik}\}.$$ (10)

It is not hard to show that \((V_{ik}, C_{ik}, t)\) is just a projection of the set defined by (10). Therefore, if the extended set \((V_{ik}, C_{ik}, Z_{ik}, t)\) is convex, then \((V_{ik}, C_{ik}, t)\) is also convex. By applying Schur’s complement to \(Z_{ik} \succeq \hat{E}^{-1/2}_{ik} V^H_i (H^k_{ik})^H C^{-1}_{ik} H^k_{ik} V_{ik} \hat{E}^{-1/2}_{ik}\), we have

$$Z_{ik} \succeq \hat{E}^{-1/2}_{ik} V^H_i (H^k_{ik})^H C^{-1}_{ik} H^k_{ik} V_{ik} \hat{E}^{-1/2}_{ik} \iff \begin{bmatrix} Z_{ik} & \hat{E}^{-1/2}_{ik} V^H_i (H^k_{ik})^H C_{ik} \\ H^k_{ik} V_{ik} \hat{E}^{-1/2}_{ik} & C_{ik} \end{bmatrix} \succeq 0.$$ (11)

This shows that (10) is a convex set whenever \(C_{ik} \succeq 0\).

The result provided by Lemma 1 allows us to further approximate the function \(\hat{h}_k^{\beta k}(E_{ik}^{\text{mmse}}, V_{ik}, \hat{E}_{ik}, \hat{V}_{ik})\).
Specifically, define $\hat{C}_{ik} = C_{ik}(\hat{V})$ as user $i_k$’s received covariance evaluated at point $\hat{V}$, then we have

\[
\begin{align*}
l_{ik}(V_{ik}, C_{ik}) = & \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k V_{ik}\right] \\
\geq & \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}\right] \\
& + \frac{dl_{ik}(\hat{V}_{ik} + t(V_{ik} - \hat{V}_{ik}), \hat{C}_{ik})}{dt} \bigg|_{t=0} \\
& + \frac{dl_{ik}(\hat{V}_{ik}, \hat{C}_{ik} + t(C_{ik} - \hat{C}_{ik}))}{dt} \bigg|_{t=0} \\
\equiv & \operatorname{Tr}[\bar{E}_{ik}^{-1}] - d_{ik} + \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k (V_{ik} - \hat{V}_{ik})\right] \\
& + \operatorname{Tr}\left[\bar{E}_{ik}^{-1}(V_{ik} - \hat{V}_{ik})^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}\right] \\
& - \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}(C_{ik} - \bar{C}_{ik})\bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}\right]
\end{align*}
\]

where (i) is due to the fact that convex functions are lower bounded by their linear approximations; (ii) simply uses the definition of directional derivative; see Appendix A for detailed derivation. The above inequality combined with the definition in (8) yields:

\[
\begin{align*}
\bar{h}_{ik}^{\beta_k}(E_{ik}^{\text{mmse}}, V_{ik}, \bar{E}_{ik}, \hat{V}_{ik}) = & \bar{a}_{ik} + \bar{c}_{ik} \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k V_{ik}\right] + \bar{c}_{ik} d_{ik} - \bar{c}_{ik} \operatorname{Tr}[\bar{E}_{ik}^{-1}] \\
\geq & \bar{a}_{ik} + \bar{c}_{ik} \operatorname{Tr}[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k (V_{ik} - \hat{V}_{ik})] + \\
& \bar{c}_{ik} \operatorname{Tr}[\bar{E}_{ik}^{-1}(V_{ik} - \hat{V}_{ik})^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}] \\
& - \bar{c}_{ik} \operatorname{Tr}\left[\bar{E}_{ik}^{-1}V_{ik}^H(H_{ik}^k)^H\bar{C}_{ik}^{-1}(C_{ik} - \bar{C}_{ik})\bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}\right] \\
& - \beta_k \operatorname{Tr}\left[\left(V_{ik} - \hat{V}_{ik}\right)\left(V_{ik} - \hat{V}_{ik}\right)^H\right] \\
\equiv & \bar{a}_{ik} + \bar{c}_{ik} \operatorname{Tr}\left[\bar{E}_{ik}^{-1}\left(\bar{U}_{ik}^H H_{ik}^k V_{ik} + V_{ik}^H(H_{ik}^k)^H \bar{U}_{ik}\right)\right] \\
& - \bar{U}_{ik}^H \left(\sum_{(i,j) \neq (i,k)} H_{ij}^k V_{jk}(V_{jk})^H(H_{ik}^k)^H \bar{U}_{ik}\right) \\
& - \beta_k \operatorname{Tr}\left[V_{ik} V_{ik}^H - V_{ik} \hat{V}_{ik}^H - \hat{V}_{ik} V_{ik}^H\right] \\
:= & h_{ik}^{\beta_k}(V; \hat{V})
\end{align*}
\]

where (i) simply use the definition of the MMSE matrix (5); (ii) uses the inequality (12); in (iii), we have defined $\bar{U}_{ik} := \bar{C}_{ik}^{-1}H_{ik}^k \hat{V}_{ik}$ and used the definition of $C_{ik}$ in (2); $\bar{a}_{ik}$ collects all the terms that
\[ h^\beta(\mathbf{V}; \hat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}} \left( \hat{a}_{i_k} + \text{Tr} \left( 2 \left( \hat{c}_{i_k} \hat{E}^{-1}_{i_k} \hat{U}^H_{i_k} \mathbf{H}_{i_k}^k + \beta^k \hat{V}^H_{i_k} \right) \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^H \mathbf{J}^k_{i_k} \mathbf{V}_{i_k}^* \right) \right) = \sum_{i_k \in \mathcal{I}} g^\beta_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}}). \] (16)

are not dependent on \( \mathbf{V} \). It is interesting to observe that the newly defined quantity \( \hat{U}_{i_k} \) is in fact the MMSE receiver for user \( i_k \) when the system precoder is given by \( \hat{\mathbf{V}} \) (cf. (2)).

Combining (8) and (13), we see that \( h^\beta_{i_k}(\mathbf{V}; \hat{\mathbf{V}}) \) becomes a locally approximated version of \( f_{i_k}(R_{i_k}(\mathbf{V})) \) that satisfies the following two properties for all feasible \( \mathbf{V} \) and \( \hat{\mathbf{V}} \):

\[ h^\beta_{i_k}(\hat{\mathbf{V}}; \hat{\mathbf{V}}) = f_{i_k}(R_{i_k}(\hat{\mathbf{V}})), \quad h^\beta_{i_k}(\mathbf{V}; \hat{\mathbf{V}}) \leq f_{i_k}(R_{i_k}(\mathbf{V})). \] (14)

Let us define \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) := \sum_{i_k \in \mathcal{I}} h^\beta_{i_k}(\mathbf{V}; \hat{\mathbf{V}}). \) The above results further imply that for all feasible \( \mathbf{V} \) and \( \hat{\mathbf{V}} \), and all \( \beta^k \geq 0 \)

\[ h^\beta_v(\hat{\mathbf{V}}; \hat{\mathbf{V}}) = f(\hat{\mathbf{V}}), \quad h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) \leq f(\mathbf{V}). \] (15)

That is, the function \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) \) is a locally tight, universal lower bound for the sum-utility function \( f(\mathbf{V}) \). A direct consequence of this observation is that the function \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) - s(\mathbf{V}) \) is a universal lower bound for \( u(\mathbf{V}) \).

Besides the universal lower bound property, \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) \) is in fact a quadratic concave function of \( \mathbf{V} \), and it is separable over all design variables \( \{\mathbf{V}_{i_k}\} \). To see this, we expand \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) \), and rearrange terms to obtain (16), where we have defined

\[ \mathbf{J}^k := \sum_{j_i \in \mathcal{I}} \hat{c}_{j_i}(\mathbf{H}_{j_i}^k)^H \hat{U}_{j_i} \hat{E}^{-1}_{j_i} \hat{U}^H_{j_i} \mathbf{H}_{j_i}^k + \beta^k \mathbf{I}_{MQ_k} \in \mathbb{S}^{MQ_k}_{++}. \] (17)

Clearly the approximation function \( h^\beta_v(\mathbf{V}; \hat{\mathbf{V}}) \) consists of \( |\mathcal{I}| \) component functions, each of which only has to do with a single variable \( \mathbf{V}_{i_k} \) (user \( i_k \)'s precoder). Put it differently, our two-stage approximation not only convexifies the objective function \( u(\mathbf{V}) \), but more importantly decomposes the nonlinearly coupled objective \( u(\mathbf{V}) \). This property of the sum-utility function \( u(\mathbf{V}) \) will be leveraged heavily for designing efficient distributed algorithms.

**Remark 1:** When \( \beta^k = 0 \) for all \( k \), the first stage approximation expressed in (8) is well-known and has been used in existing works such as [26] [29]; also see [40] Section VIII-A] for discussion. However, none of those works articulated the fact that there exists a separable quadratic function of \( \mathbf{V} \) in the form of (16) which serves as a global lower bound for the original system utility; cf. (15). This property has been made clear in our derivation precisely due to the use of the second stage approximation. Moreover, the introduction of the coefficients \( \beta^k \)'s further adds flexibility to the algorithm. The latter point will be
made clear shortly.

\[ \text{B. The Successive Convex Approximation Algorithm} \]

The lower bounds developed in the previous subsection are crucial for our design of efficient algorithms for problem (\text{P}_{\text{SYSTEM}}). In this subsection, we will develop two general algorithms and analyze their properties. In the sections that follow, we will see how these general algorithms can be effectively implemented and tailored for special cases of \text{P}_{\text{SYSTEM}}.

Our approach is to successively approximate \( f(V) \) using \( h^\beta(V; \hat{V}) \) to obtain progressively improved solutions. Let us use \( (t) \) to denote the iteration index. The first algorithm, referred to as the successive convex approximation (SCA) algorithm, consists of three main steps; also see Fig. 1 for a graphical illustration.

**Step 1:** Suppose that \( V(t-1) \) is a feasible solution to \( \text{P}_{\text{SYSTEM}} \). At iteration \( t \), the following convex optimization problem is solved to obtain \( V(t) \)

\[
\begin{align*}
\max_V & \quad h^\beta(V; V(t-1)) - s(V) \\
\text{s.t.} & \quad V^{q_k} \in \mathcal{V}^{q_k}, \forall q_k \in Q, \quad V^k \in \mathcal{V}^k, \forall k \in \mathcal{K}.
\end{align*}
\]  

**Step 2:** For each user \( i_k \in \mathcal{I} \), compute

\[
\begin{align*}
C_{ik}(t) &= \sum_{(i,j)} H_{ik}^f V_{ji}(t) V_{ji}^H(t) (H_{ik}^f)^H + \sigma^2_i I_N \\
U_{ik}(t) &= C_{ik}^{-1}(t) H_{ik}^k V_{ik}(t) \\
E_{ik}(t) &= \mathbf{I}_{d_{ik}} - V_{ik}^H(t) (H_{ik}^k)^H C_{ik}^{-1}(t) H_{ik}^k V_{ik}(t) \\
c_{ik}(t) &= \frac{\partial f_{ik}(x)}{\partial x} \bigg|_{x = -\log |E_{ik}(t)|} \\
J^k(t) &= \sum_{j_i \in \mathcal{I}} c_{ji}(t) (H_{ji}^k)^H U_{ji}(t) E_{ji}^{-1}(t) U_{ji}^H(t) H_{ji}^k + \beta^k \mathbf{I}_{M Q_k}.
\end{align*}
\]

**Step 3:** Form the updated function \( h^\beta(V; V(t)) \) according to (16). Let \( t = t + 1 \), and go to Step 1.

The key step in the SCA algorithm is to solve the lower-bound maximization problem \( \text{P}_{\text{Lower-Bound}} \) exactly to global optimality. In practice, especially for very large-scale networks, exactly solving \( \text{P}_{\text{Lower-Bound}} \) may be computationally very demanding. Therefore it is desirable to perform an inexact optimization for \( \text{P}_{\text{Lower-Bound}} \). The following alternative, termed inexact successive convex approximation (In-SCA) algorithm, replaces Step 1 of SCA by the following step

**Step 1 of In-SCA:** Suppose that \( V(t-1) \) is a feasible solution to \( \text{P}_{\text{SYSTEM}} \). At iteration \( t \), obtain \( V(t) \) that satisfies conditions (20).
(h_β^V(t); V(t-1)) - s(V(t)) - (h_β^V(t-1); V(t-1)) - s(V(t-1)) ≥ η(t)\|V(t-1) - V(t)\|_F^2
\]

\[ V(t) = V(t-1) \implies V(t-1) = \arg\max_{V} h_β^V(V; (V(t-1)) - s(V). \]

(20)

Here η(t) > η > 0 is some iteration dependent constant that is bounded away from zero. The first condition requires that the new precoder should sufficiently improve the lower bound, and such improvement is measured by the size of the successive difference of the precoders. The second condition restrict the way that the iterates should be generated by the algorithm. It says that if the iteration stops, then a fixed point of the SCA algorithm must have been reached. Note that the In-SCA algorithm is rather a family of algorithms that generates the precoders that satisfy the conditions specified in (20).

We will soon see some examples of the In-SCA algorithm in the subsequent section.

Next we analyze the convergence properties of the two proposed algorithms. To proceed, the following definitions are needed:

- **Directional derivative**: Let \( f: \mathcal{V} \to \mathbb{R} \) be a function where \( \mathcal{V} \) is a convex set. The directional derivative of \( f \) at point \( x \in \mathcal{V} \) in direction \( d \) is defined by

\[
f'_d(x) := \liminf_{r \downarrow 0} \frac{f(x + rd) - f(x)}{r}.
\]

- **Stationary points of a function**: Let \( f: \mathcal{V} \to \mathbb{R} \) where \( \mathcal{V} \) is a convex set. The point \( x \) is a stationary point of \( f(\cdot) \) if \( f'_d(x) \leq 0 \) for all \( d \) such that \( x + d \in \mathcal{V} \).

![Fig. 1. A graphical illustration of the proposed algorithms. Left: The SCA Algorithm. Right: The In-SCA Algorithm.](image)

We have the following convergence theorem. The proof is given in Appendix B.

**Theorem 1**: Suppose the following conditions hold:

1) Assumptions A-1)–A-3) and B-1) are all satisfied;
2) The sets \( \{V^k\}_{k \in K} \) and \( \{V^q_k\}_{q_k \in Q} \) are all convex, closed and compact.

Then the precoders \( \{V(t)\} \) generated by either the SCA or the In-SCA algorithm globally converge to a stationary solution of problem \( (P_{\text{SYSTEM}}) \).

It is important to note that for both algorithms, efficiently performing Step 1 is the key for the low-complexity implementation. It is always desirable to develop customized algorithms tailored for specific problems, especially for those that have nonsmooth penalization terms. This topic will be treated in the next section.

IV. CUSTOMIZED ALGORITHMS FOR PRECODER DESIGN

In this section, we customize the SCA and the In-SCA algorithms described in the previous section to different network settings.

A. Linear Precoder Design for IBC Model

First let us consider the IBC model in which there is a single BS in each cell. The objective is to design the optimal linear precoders subject to the sum power constraint. As each cell consists of a single BS, \( V_i^k \) is used to denote the transmit precoder used by the BS in cell \( k \) to transmit to user \( i_k \). In this case, there is no penalty term \( s(V) \), and the sets \( V^q_k \) and \( V^k \) collapse to a single feasible set given by \( V^k = \sum_{i_k \in I_k} \text{Tr}(V_{i_k}^k V_{i_k}^H) \leq \bar{P}_k \). Hence the subproblem \( (P_{\text{Lower-Bound}}) \) is given by

\[
\max_{\{V^k\}_{k \in K}} \sum_{k \in K} \sum_{i_k \in I_k} g_{i_k}^k (V_{i_k}, \hat{V}) \quad \text{(P_{IBC})}
\]

subject to

\[
\sum_{i_k \in I_k} \text{Tr}(V_{i_k}^k V_{i_k}^H) \leq \bar{P}_k, \quad k \in K.
\]

Clearly both the constraints and the objective of the above problem are separable among the BSs. Therefore we can decompose this problem into \( K \) independent subproblems, with the \( k \)-th subproblem taking the following form

\[
\max_{V^k_{i_k} \in I_k} \sum_{i_k \in I_k} g_{i_k}^k (V_{i_k}, \hat{V}) \quad \text{(P_{IBC-SUB})}
\]

subject to

\[
\sum_{i_k \in I_k} \text{Tr}(V_{i_k}^k V_{i_k}^H) \leq \bar{P}_k.
\]

Let \( \lambda^k \geq 0 \) denote the Lagrangian multiplier associated with the power constraint. Then the Lagrangian function for problem \( (P_{\text{IBC-SUB}}) \) is given by (21) (excluding constants that are not related to \( (\lambda^k, V^k) \)).

Clearly the Slater Condition holds, so the optimal primal-dual pair \( ((V^k)^*, (\lambda^k)^*) \) should satisfy the
$$L_k(\mathbf{V}^k, \lambda^k; \hat{\mathbf{V}}) = \sum_{i_k \in I_k} \left( \text{Tr} \left[ 2(\hat{c}_{i_k} \hat{E}_{i_k}^{-1} \hat{U}_{i_k}^H \hat{H}_{i_k}) \mathbf{V}_{i_k} - \mathbf{V}_{i_k}^H \hat{J}_k \mathbf{V}_{i_k} \right] - \lambda^k \left( \sum_{i_k \in I_k} \text{Tr} [\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H] - \bar{P}_k \right) \right).$$

(21)

KKT optimality conditions

$$\begin{align*}
(V^k)^* &= \arg \max_{V^k} L_k(V^k; (\lambda^k)^*; \hat{V}), \quad (22a) \\
(\lambda^k)^* \left( \bar{P}_k - \sum_{i_k \in I_k} \text{Tr}[\mathbf{V}_{i_k}^* \mathbf{V}_{i_k}^{*H}] \right) &= 0, \quad (22b) \\
(\lambda^k)^* &\geq 0, \quad \bar{P}_k - \sum_{i_k \in I_k} \text{Tr}[\mathbf{V}_{i_k}^* \mathbf{V}_{i_k}^{*H}] \geq 0. \quad (22c)
\end{align*}$$

For fixed $\lambda^k \geq 0$, the solution for the unconstrained problem $\max_{\mathbf{V}^k} L(\mathbf{V}^k; \lambda^k; \hat{\mathbf{V}})$, denoted as $\{\mathbf{V}_{i_k}^* (\lambda^k)\}_{i_k \in I_k}$, can be expressed as

$$\mathbf{V}_{i_k}^* (\lambda^k) = \left( \hat{J}_k + \lambda^k \mathbf{I}_M \right)^{-1} \left( \hat{c}_{i_k} (\mathbf{H}_{i_k}^H)^H \hat{U}_{i_k} \hat{E}_{i_k}^{-1} + \beta^k \hat{V}_{i_k} \right), \quad \forall i_k. \quad (23)$$

The above solution can be written down in closed form because $L(\mathbf{V}^k, \lambda^k; \hat{\mathbf{V}})$ is simply a quadratic function with respect to $\mathbf{V}^k$.

To find the optimal multiplier $(\lambda^k)^*$ that satisfies the complementarity condition $(22b)$, we utilize a general result on penalty method for optimization, e.g., [41, Section 12.1, Lemma 1]. This result asserts that for the solution $\{\mathbf{V}_{i_k}^* (\lambda^k)\}_{i_k \in I_k}$, the penalized term $\sum_{i_k \in I_k} \text{Tr}[\mathbf{V}_{i_k}^* (\lambda^k)(\mathbf{V}_{i_k}^* (\lambda^k))^H]$ must be monotonically decreasing with respect to $\lambda^k$. Such monotone property suggests that we can find the optimal multiplier that satisfies $(22b)$ by a simple bisection search procedure; see e.g., [3, Section 3.3.1] for detailed description of the bisection procedure.

The algorithm discussed in this subsection is summarized in Table I.

Remark 2: When we set $\beta^k = 0$ for all $k$, the algorithm listed in Table II recovers the weighted-MMSE (WMMSE) algorithm proposed in [26] for utility maximization problem in IBC networks. It is interesting to understand the difference between the WMMSE approach and the SCA approach studied here. Reference [26] starts with a utility optimization problem with the objective $u(\mathbf{V})$ (\mathbf{V} is the users’ precoder). Then it uses a “dimension lifting” approach that adds two extra variables $\mathbf{W}$ and $\mathbf{U}$, and consider a different weighted MSE minimization problem $\min \bar{u}(\mathbf{U}, \mathbf{W}, \mathbf{V})$. The authors of [26] shows that these two problems are equivalent, so they apply a BCD algorithm to solve the reformulated three-
TABLE I
SCA FOR (P SYSTEM) IN IBC SETTING

S1): Initialization Obtain a feasible solution $V_{ik}(0)$ for all $i_k$
S2): For each $k$, compute $C_{ik}(t)$, $E_{ik}(t)$, $U_{ik}(t)$, $c_{ik}(t)$, according to (19a)–(19d), for all $i_k \in I_k$
S3): For each $k$, compute $J_k(t)$ by (19e);
For each $k$, compute $V_k(t)$ by
$$V_{ik}(t) = (J_k(t) + (\lambda^k)^{-1} I_M) V_{ik}(t-1) + \beta^k V_{ik}(t-1)$$
where $(\lambda^k)^{-1}$ is computed by a bisection procedure
S4) Until some stopping criterion is met

block problem. However, it is by no means clear what is the intuition behind adding these two new variables. Our SCA framework well explains this question: the new variable $W$ is in fact the coefficients derived from the first stage linear approximation, while the new variable $U$ is the coefficients derived from the second stage convex approximation. Note that in this paper we arrive at the WMMSE algorithm by specializing the SCA algorithm to the IBC setting. Thus the SCA algorithm is more general and includes the WMMSE as a special case.

□

B. Intra-cell ZF and Inter-cell CB for IBC Model

We then consider a similar IBC model as in the previous subsection, but with each BS employing a ZF precoder to cancel the intra-cell interference among the users. We assume that certain user-selection within each cell has already been performed, which ensures that the per-cell zero forcing constraint
$$H_{jk} V_{ik} V_{ik}^H (H_{jk})^H = 0, \forall j_k \neq i_k, j_k \in I_k$$
is always feasible (one easily checkable condition that guarantees feasibility is $|I_k| \times N \leq M$, see [9], [38]). Note that in this network setting, the inter-cell interference is still present, despite the fact that the intra-cell interference is canceled by the use of ZF precoder. Therefore the original problem (P SYSTEM) is still difficult to solve. To apply the SCA algorithm, specialize the subproblem (P Lower–Bound) to the following

$$\max_{\mathbf{V}} \sum_{k \in K} \sum_{i_k \in I_k} g_{ik}^k (V_{ik}; \hat{V})$$
subject to
$$\sum_{i_k \in I_k} \text{Tr}[V_{ik} V_{ik}^H] \leq \bar{P}_k, k \in K$$
$$H_{jk} V_{ik} V_{ik}^H (H_{jk})^H = 0, \forall j_k \neq i_k, j_k \in I_k.$$
To remove the ZF constraints (24b), let $L := N(|I_k| - 1)$, and define a set of concatenated channel matrices

$$Q_{k} := \{H_{ik}^k\}_{j_k \in I_k \{i_k\}} \in \mathbb{C}^{L \times M}, \forall i_k \in I_k. \tag{25}$$

Let $Q_{ik} = L_{ik} \Sigma_{ik} R_{ik}$ denote the singular value decomposition of $Q_{ik}$, where $L_{ik}$ and $R_{ik}$ are two unitary matrices, with $L_{ik} \in \mathbb{C}^{L \times L}$ and $R_{ik} \in \mathbb{C}^{L \times M}$, and $\Sigma_{ik}$ being an $L \times L$ diagonal matrix. Define $P_{ik} := (I - R_{ik} R_{ik}^H)$ as a projection matrix to the space orthogonal to the one spanned by $R_{ik}$. Let $P_{ik} = \tilde{R}_{ik} \tilde{R}_{ik}^H$, where $\tilde{R}_{ik} \in \mathbb{C}^{M \times (M - L)}$ is composed of the orthogonal basis that satisfies $R_{ik} \tilde{R}_{ik} = 0$ and $\tilde{R}_{ik}^H \tilde{R}_{ik} = I$. Then [37] Lemma 3.1 asserts that the optimal solution of problem $P_{IBC-ZF}$ must be of the form: $V_{ik} = \tilde{R}_{ik} W_{ik}$, where $W_{ik} \in \mathbb{C}^{(M - L) \times d_{ik}}$.

Utilizing this structure of the optimal solution, the problem $P_{IBC-ZF}$ can be equivalently written as

$$\begin{align*}
\max_{W_k} & \quad \sum_{i_k \in I_k} \sum_{k} \tilde{g}^{\beta_k}_{i_k}(W_{ik}; \hat{V}) \\
\text{s.t.} & \quad \sum_{i_k \in I_k} \text{Tr}(\tilde{R}_{ik} W_{ik} (\tilde{R}_{ik} W_{ik})^H) \leq \tilde{P}_k, \quad k \in K,
\end{align*} \tag{P_{IBC-ZF2}}$$

where the function $\tilde{g}^{\beta_k}_{i_k} (\cdot)$ is the same as the original objective $g^{\beta_k}_{i_k} (V_{ik}; \hat{V})$, except that $V_{ik}$ is replaced by $\tilde{R}_{ik} W_{ik}$. Again both the constraints and the objective of this problem are separable among the BSs, and we can further decompose this problem into $K$ independent subproblems of the form

$$\begin{align*}
\max_{W_k} & \quad \sum_{i_k \in I_k} \tilde{g}^{\beta_k}_{i_k}(W_{ik}; \hat{V}) \\
\text{s.t.} & \quad \sum_{i_k \in I_k} \text{Tr} \left[\tilde{R}_{ik} W_{ik} (\tilde{R}_{ik} W_{ik})^H\right] \leq \tilde{P}_k.
\end{align*} \tag{P_{IBC-ZF-SUB}}$$

Let us use $\lambda_k$ to denote the Lagrangian multiplier associated with the power constraint. Then by using similar steps leading to (23), we can show that the optimal solution for problem $P_{IBC-ZF-SUB}$ is of the form

$$W_{ik}^* = \left( \sum_{(i,j)} \bar{c}_{ij} (H_{jk}^k) \tilde{U}_{ij} \tilde{E}_{ij}^{-1} \tilde{U}_{ij}^H H_{jk}^k \tilde{R}_{ik} + \lambda_k^* I_{M-L} \right)^{-1}$$

$$\times \left( \tilde{R}_{ik}^H \left[ \tilde{c}_{ik} (H_{jk}^k) \tilde{U}_{ik} \tilde{E}_{ik}^{-1} + \beta_k^* \hat{V}_{ik} \right] \right), \forall i_k \in I_k. \tag{28}$$

where $\lambda_k^*$ can be again computed by the bisection method. The detailed steps are summarized in Table II.
**TABLE II**

SCA FOR $\text{P}_{\text{SYSTEM}}$ IN IBC SETTING WITH INTRA-CELL ZF AND INTER-CELL CB

| S1) Initialization                                                                 |
|---------------------------------|----------------------------------|
| For each $k$ and $i_k$, compute: |
| $Q_{i_k} = L_{i_k} \Sigma_{i_k} R_{i_k}$, $P_{i_k} = (I - R_{i_k} R_{i_k}^H)$, $P_{i_k} = \tilde{R}_{i_k} \tilde{R}_{i_k}^H$, |
| Obtain a feasible solution $V_{i_k}(0)$ for all $i_k$; |
| S2): For each BS $k$, compute $C_{i_k}(t)$, $E_{i_k}(t)$, $U_{i_k}(t)$ and $c_{i_k}(t)$, according to (19a)–(19d), for all $i_k$; |
| S3): Compute $W_{i_k}(t)$ according to (28). |
| S4) Until some stopping criterion is met. |

**C. Linear Precoder Design for HetNet with Intra-Cell Full ComP and Inter-Cell CB**

Consider a HetNet setting, in which there are a set of $Q_k$ BSs in each cell, and they form a single virtual BS to transmit to the users. In this case, $Y_{qk}$ becomes $Y_{qk} = \{V_{qk} : \sum_{i_k \in I_k} \text{Tr}[V_{i_k} Q_{i_k}] \leq \tilde{P}_{qk}\}$, representing the power constraint for BS $q_k$. This scenario also covers the multi-cell IBC scenario with per group of antenna power constraints, see e.g., [42].

Assume for now that there is no penalty term $s(V)$. Then the subproblem $P_{\text{Lower-Bound}}$ is again decomposed into $K$ independent subproblems:

$$\max_{V_k} \sum_{i_k \in I_k} g^{i_k}(V_{i_k}, \hat{V})$$

subject to

$$\sum_{i_k \in I_k} \text{Tr}[V_{i_k} Q_{i_k}] \leq \tilde{P}_{q_k}, \forall q_k \in Q_k.$$  

Differently from problem $P_{\text{IBC-SUB}}$ discussed in Section IV-A, the above problem has $Q_k = |Q_k|$ separable constraints (each constraining a subset of variables), hence $Q_k$ Lagrangian multipliers $\{\lambda_{q_k}\}_{q_k \in Q_k}$. Therefore the bisection algorithm on a single multiplier does not work anymore.

Fortunately, the constraints for this problem are separable among different block variables $\{V_{qk}\}_{q_k \in Q_k}$. The leads to a natural block coordinate descent (BCD) algorithm (see [43], [44]) to compute the optimal solution for problem $P_{\text{VIBC-SUB}}$. The BCD algorithm updates one block variable $V_{qk}$ at a time while holding the remaining block variables fixed. To capitalize the block structure of the problem, the following definitions are needed. Let

$$\tilde{S}_{i_k} := \tilde{c}_{i_k}(H_{i_k}^H \tilde{U}_{i_k} \tilde{E}_{i_k}^{-1} + \beta k \tilde{V}_{i_k}) \in \mathbb{C}^{M_k Q_k \times d_k}, \forall i_k \in I_k.$$  

(29)
Partition \( \hat{J}^k \) (cf. (17)) and \( \hat{S}_{ik} \) into the following form

\[
\hat{j}^k = \begin{bmatrix}
\hat{j}^k[1, 1], & \cdots, & \hat{j}^k[1, Q_k] \\
\vdots & \ddots & \vdots \\
\hat{j}^k[Q_k, 1] & \cdots & \hat{j}^k[Q_k, Q_k]
\end{bmatrix}
\]

\[
\hat{s}_{ik} = \left[ \hat{s}_{ik}^H[1], \cdots, \hat{s}_{ik}^H(Q_k) \right]^H
\]

where \( \hat{j}^k[q, p] \in \mathbb{C}^{M \times M}, \forall (q, p) \in Q_k \times Q_k \), and \( \hat{s}_{ik}[q] \in \mathbb{C}^{M \times d_k}, \forall q \in Q_k \). Then the function \( g_{ik}^{\delta^k}(V_{ik}; \hat{V}) \) defined in (16) can be alternatively expressed as

\[
g_{ik}^{\delta^k}(V_{ik}; \hat{V}) = \alpha_{ik} + \sum_{p_k \in Q_k} 2 \text{Tr} \left[ \hat{s}_{ik}^H[p_k] V_{ik}^p \right] - \sum_{p_k, q_k \in Q_k} \text{Tr} \left[ (V_{ik}^{q_k})^H \hat{j}^k[q_k, p_k] V_{ik}^p \right]
\]

Now it becomes clear that \( \sum_{i_k \in I_k} g_{ik}^{\delta^k}(V_{ik}; \hat{V}) \) is again a quadratic function with respect to one particular block variable, say \( V_{m_k}^{m_k} \). It follows that the per-block problem, written in the following form, can be efficiently solved in closed form for each block \( m_k \in Q_k \)

\[
\begin{align*}
\max_{V_{m_k}} & \sum_{i_k \in I_k} g_{ik}^{\delta^k}(V_{ik}; \hat{V}) \\
\text{s.t.} & \sum_{i_k \in I_k} \text{Tr} [V_{m_k}^{m_k} (V_{m_k}^{m_k})^H] \leq M_{m_k}.
\end{align*}
\]

(P_VIBC-BLK)

Let \( \lambda_{m_k}^m \geq 0 \) denote the Lagrangian multiplier associated with the power constraint of the \( m_k \)-th subproblem. Following the same derivation in Section IV-A the optimal solution \( (V_{m_k}^{m_k})^* \) for problem (P_VIBC-BLK) can be expressed as

\[
V_{m_k}^{m_k*} = \left( \hat{j}^k[|m_k, m_k] + \lambda_{m_k}^m \lambda_M \right)^{-1}
\]

\[
\times \left( \hat{s}_{ik}^H[|m_k] - \sum_{p_k \neq m_k} \hat{j}^k[|m_k, p_k] V_{ik}^p \right), \quad \forall i_k \in I_k
\]

where the optimal multiplier can be computed again using a bisection search.

The above observations leads to a natural two-layer algorithm: \( i \) the outer layer updates \( E_{i_k}(t), U_{i_k}(t), c_{i_k}(t), J^{k}(t) \) and \( S^{k}(t) \); \( ii \) the inner layer updates each \( V^{k} \) by a BCD algorithm with blocks given by \( \{V^{q_k}\}_{q_k \in Q_k} \). See Table III for the detailed description. Once again the convergence of the algorithm is a direct consequence of Theorem I.

The key question here is whether it is necessary to solve the inner problem (P_VIBC-SUB) exactly before we can update the outer layer. There are several drawbacks with this approach, which we list below:
S1): Initialization Obtain a feasible solution $V_{i_k}(0)$ for all $i_k$
S2): For each BS $k$, compute $C_{i_k}(t)$, $E_{i_k}(t)$, $U_{i_k}(t)$ and $c_{i_k}(t)$ according to (19a)–(19d), for all $i_k \in \mathcal{I}_k$
S3): For each BS $k$, compute $J_k(t)$ by (19e);
For each BS $k$, compute $S_{i_k}(t) = c_{i_k}(t)(H_{i_k}^H U_{i_k}(t)E_{i_k}^{-1}(t)\nonumber$
$+ \beta^k V_{i_k}(t-1), \forall i_k \in \mathcal{I}_k.$
S4): For each BS $k$, compute the precoders $V^k(t)$ by
Repeat Cyclically pick $m_k \in Q_k$
Compute $V_{m_k}^{i_k}$ using (32), $\forall i_k \in \mathcal{I}_k$.
where $\lambda_{m_k}^*$ is computed by a bisection procedure
Until An optimal solution of $(P_{VIBC-SUB})$ is obtained.
Let $V_{m_k}^{i_k}(t) = V_{m_k}^{i_k}, \forall i_k, m_k$
S5) Until some stopping criteria is met.

1) It is usually difficult to check whether the inner iteration has indeed reached the optimality;
2) Before reaching the optimality for the subproblem $(P_{VIBC-SUB})$, the marginal benefit of the precoder updates in the inner iteration decreases as the iteration progresses. This effect is manifested in particular in the first few outer iterations, in which even the inner problem is solved exactly, the precoders obtained are still far away from the optimal ones.

Somewhat surprisingly, by utilizing the In-SCA algorithm, a single inner BCD iteration is sufficient to guarantee the convergence of the overall algorithm. The benefit of such inexact algorithm is quite obvious from our preceding discussion. It allows one to solve the subproblems approximately at the beginning, and more accurately later as the iteration progresses. To be more specific, all the steps of the resulting algorithm will be the same as before, except for Step S4), where the inexact algorithm only requires that each block is picked at least once; see Table IV for detailed description.

Next we show the convergence of the above inexact algorithm, by appealing to Theorem I.

**Corollary 1:** Suppose $\beta^k > 0$ for all $k$. Then the inexact algorithm given in Table IV is a special case of the In-SCA algorithm, hence it converges to a stationary solution of problem $(P_{SYSTEM})$ with intra-cell ComP and inter-cell CB.

**Proof:** We will focus on the slightly special case that each coordinate $m_k \in Q_k$ is updated precisely once in Step 3. The general case in which each coordinate $m_k \in Q_k$ is updated at least once is a straightforward extension.

We only need to verify that the two conditions specified in the In-SCA algorithm are satisfied. To this
end, define the following short-handed notation

\[
\mathbf{V}^{-m_k}_{ik} := \{ \mathbf{V}_{ik}^{p_k} \}_{p_k \neq m_k}, \quad \mathbf{V}^{-m_k} := \{ \mathbf{V}^{p_k} \}_{p_k \neq m_k}.
\]

Also define \( \nabla_m g^\beta_{ik} (\mathbf{V}^{m_k}_{ik}, \mathbf{V}^{-m_k}_{ik}; \hat{\mathbf{V}}) \) as the gradient of \( g^\beta_{ik} \) with respect to the block \( \mathbf{V}^{m_k}_{ik} \).

At any given \( \hat{\mathbf{V}} \), we fix \( \mathbf{V}^{-m_k} \) and update \( \mathbf{V}^{m_k} \) by using (32). First observe that the optimality condition for the per-block problem (\( P^{\text{VIBC–BLK}} \)) is given by

\[
\sum_{i_k \in I_k} \left\langle \nabla_m g_{ik}^\beta \left( \mathbf{V}^{m_k}_{ik}, \mathbf{V}^{-m_k}_{ik}; \hat{\mathbf{V}} \right), \mathbf{V}^{m_k} - \mathbf{V}^{-m_k}_{ik} \right\rangle \geq 0, \forall \hat{\mathbf{V}} \in \mathbf{V}^{m_k}.
\]

Utilizing this property, we have

\[
\sum_{i_k \in I_k} g_{ik}^\beta \left( \mathbf{V}^{m_k}_{ik}, \mathbf{V}^{-m_k}_{ik}; \hat{\mathbf{V}} \right) - g_{ik}^\beta \left( \mathbf{V}^{m_k}_{ik}, \mathbf{V}^{-m_k}_{ik}; \hat{\mathbf{V}} \right) + \sum_{i_k \in I_k} \frac{\beta^k}{2} \left\| \mathbf{V}^{m_k}_{ik} - \mathbf{V}^{-m_k}_{ik} \right\|_F^2
\]

\[
\geq \sum_{i_k \in I_k} \frac{\beta^k}{2} \left\| \mathbf{V}^{m_k}_{ik} - \hat{\mathbf{V}}_{ik} \right\|_F^2,
\]

where (i) is from the strong concavity of \( g_{ik}^\beta \left( \mathbf{V}^{m_k}_{ik}, \mathbf{V}^{-m_k}_{ik}; \hat{\mathbf{V}} \right) \) with respect to each \( \mathbf{V}^{m_k}_{ik} \), and the fact

\[
\mathbf{S}_{ik}(t) = c_{ik}(t) (\mathbf{H}_{ik})^H \mathbf{U}_{ik}(t) \mathbf{E}^{-1}(t)
\]

\[
+ \beta^k \mathbf{V}_{ik}(t-1), \forall i_k \in I_k.
\]
that the modulus of the strong concavity is at least $\beta_k$. (ii) comes from the optimality condition (33).

Then after one round of the update, we must have
\[
\begin{align*}
&\quad h^{\beta}(V^*; \hat{V}) - h^{\beta}(\hat{V}; \hat{V}) \\
&\geq \sum_{k \in K} \sum_{i_k \in I_k} \sum_{m_k \in Q_k} \frac{\beta_k^2}{2} \| V_{i_k}^{m_k*} - \hat{V}_{i_k}^{m_k} \|^2_F \\
&\geq \frac{\min_k \beta_k^2}{2} \| V^* - \hat{V} \|^2_F
\end{align*}
\]
which proves the first condition in (20).

To show the second condition in (20), we fix $V^{m_k} = \hat{V}^{m_k}$ and consider $V^{m_k*}$ obtained by using (32). Clearly if $\hat{V}^{m_k} = V^{m_k*}$ is true, then we have
\[
\hat{V}^{m_k} = \arg \max_{V^{m_k}} \sum_{i_k \in I_k} g^{\beta_k}_{i_k}(V_{i_k}^{m_k}, \hat{V}_{i_k}^{m_k}; \hat{V}). \tag{34}
\]
Clearly the above relation is true for each block $m_k \in Q_k$. Writing down the first order optimality condition for these $|Q_k|$ relations (cf. (33)), we can see that collectively they imply the first order optimality condition of the following convex problem
\[
\hat{V}^k = \arg \max_{V^k} \sum_{i_k \in I_k} g^{\beta_k}_{i_k}(V_{i_k}; \hat{V}). \tag{35}
\]
Enumerating over all cells $k \in K$, we obtain
\[
\hat{V} = \arg \max_{V} \sum_{k \in K} \sum_{i_k \in I_k} g^{\beta_k}_{i_k}(V_{i_k}; \hat{V}) \\
= \arg \max_{V} h^{\beta}(V; \hat{V}). \tag{36}
\]
We have proved that the inexact algorithm in Table IV is a special case of In-SCA algorithm. The convergence then follows as a direct consequence of Theorem 1.

Remark 3: From the previous proof it is clear why we have introduced the quadratic term
\[
-\beta_k^2 \text{Tr} \left[ (V_{i_k} - \hat{V}_{i_k}) (V_{i_k} - \hat{V}_{i_k})^H \right]
\]
in the lower bound expression $g^{\beta_k}_{i_k}$. As long as $\beta_k$ is bounded away from zero, the lower bound function is strongly convex with respect to $V_{i_k}$, which in turn guarantees the sufficient descent property (20).

\footnote{Note that the matrix $\hat{J}^k$ is positive definite with the smallest eigenvalue being at least $\beta^k$.}
Intuitively, a larger $\beta_k$ means that a new iterate will be closer to the old iterate, which may result in slower convergence. Therefore in practice, $\beta_k$ is often chosen to be a very small number.

**Remark 4:** The algorithms proposed in this section can be easily extended to the case of per-cell partial ComP.

Assume that the BS clustering structure is known, and we let $S^q_k \subseteq \mathcal{I}_k$ denote the set of users served by BS $q_k$, then we only need to slightly modify the proposed algorithm in Table III and IV by the following:

- In S1), for each BS $m_k \in Q_k$, set $V_{ik}^{m_k}(0) = 0$ for all $i_k \notin S^m_k$;
- In S4), let each BS $m_k \in Q_k$ compute $V_{ik}^{m_k}(t)$ using (32), $\forall i_k \in S^m_k$ (rather than using all $i_k \in \mathcal{I}_k$).

In this way, only the precoders of the subset of users served by each BS will be updated in each iteration. Once again, the computation in each iteration admits a closed-form solution, while in related works such as [33], general purpose convex solvers need to be used for solving the subproblems.

Moreover, when the BSs’ clustering structure needs to be designed jointly with the precoders, we can include the penalty term $s(V)$ into the objective to induce certain block-sparsity in the precoder $V^{q_k}$. Except for this additional term in the objective, which leads to different solution to the associated subproblem, the algorithm for the joint BS clustering and precoder design problem has the same structure as the one presented in Table III. Specifically, the per-block subproblem $\text{P}_{\text{VIBC-BLK}}$ takes the following form

$$
\max_{V^{m_k}} \sum_{i_k \in \mathcal{I}_k} g_{ik}(V_{ik}, \hat{V}) - \sum_{i_k \in \mathcal{I}_k} s_{ik}^{m_k}(V_{ik}^{m_k}) \quad (37)
$$

$$\text{s.t.} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}(V_{ik}^{m_k}(V_{ik}^{m_k})^H) \leq \tilde{P}^{m_k}.
$$

When we let $s_{ik}^{m_k}(V_{ik}^{m_k}) = \gamma_{ik}^{m_k} \|V_{ik}^{m_k}\|_F$, this subproblem becomes a well known quadratic group-LASSO problem [45] (with an additional quadratic constraint), which can be solved using an iterative procedure; see [34], [36] for details.

**Remark 5:** Certainly one can reformulate each subproblem $\text{P}_{\text{VIBC-BLK}}$ into a second order cone program (SOCP, see, e.g., [46]), and solve it using general purpose solvers. Unfortunately by simply applying the analysis of the SCA algorithm, it is again not clear when the inner optimization should stop, as theoretically we need the inner problem to be solved exactly. More importantly this approach does not sufficiently utilize problem structure, therefore can be quite inefficient when the size of problem is large, and it may not be able to produce sparse solutions even in the presence of nonsmooth penalization.
Remark 6: For the problem discussed in this section, the In-SCA algorithm can take other forms as well. For example one can utilize the BSCA algorithm proposed in [40], or the CGD method proposed in [43], [47]. All one needs to do is to verify that the conditions set out in [20] are satisfied. However, it appears that the BCD-based In-SCA proposed in Table IV takes a much simpler form, and is much easier to implement and analyze.

V. Numerical Results

In this section we conduct experiments to validate the effectiveness of the proposed algorithms. Both the exact and inexact SCA algorithms are tested for three main settings: 1) Multicell downlink linear precoder design (i.e., the IBC model); 2) HetNet downlink linear precoder design with inter-cell CB and intra-cell JP; 3) HetNet downlink joint clustering and linear precoder design with intra-cell partial ComP.

The general setup for the experiments are given as follows. We consider a multicell network of up to 10 cells. The distance of the centers of two adjacent cells is set to be 500 meters (representing a HetNet with densely deployed cells). Both the BSs and the users are randomly placed in each cell. Let \( y_{q\ell i_k} \) denote the distance between BS \( q\ell \) and user \( i_k \). The channel coefficients between user \( i_k \) and BS \( q\ell \) are modeled as zero mean circularly symmetric complex Gaussian vector with \((200/y_{q\ell i_k}^3 L_{q\ell i_k}^q)\) as variance for both real and imaginary dimensions, where \( 10\log_{10}(L_{q\ell i_k}^q) \sim \mathcal{N}(0,64) \) is a real Gaussian random variable modeling the shadowing effect. We set the noise power \( \sigma_{i_k}^2 = 1 \) for all \( i_k \), and uniformly randomly generate the power budget \( \bar{P}_{q\ell k} \in (0, C_k] \) for all \( q\ell k \in Q_k \), where the constant \( C_k \) is decided such that the sum of power budget of all \( q\ell k \in Q_k \) is given by \( P_{tot}^k \).

The stopping criteria are chosen as follows. The single time-scale algorithm (i.e. the In-SCA in Table IV or the SCA described in Table I) as well as the outer loop of the double time-scale algorithm (i.e., the SCA described in Table III) stop when \( |u(t+1) - u(t)|/|u(t)| \leq 10^{-3} \). The inner loop of the two-time scale algorithm stops while the relative increase of the objective value for the related subproblem (i.e., problem \( P_{\text{VIBC-BLK}} \)) is less than \( 10^{-3} \) after performing one round of update by all the BSs in the cell.

A. HetNet and Multicell Downlink Setting

In this section, the performance of the following algorithms will be demonstrated and compared:

1) WMMSE [26]: This algorithm is the one described in Table I with \( \beta_k = 0 \) for all \( k \).
Fig. 2. Comparison of the system throughput of different algorithms in HetNet setting. $K = [2, 4, 6, 8, 10]$, $P_{k_{\text{tot}}} = 20$ dB, $|Q_k| = 6$, $|I_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 1$ or $d_{i_k} = 3$ for all $i_k \in I$.

2) **SCA for HetNet**: This algorithm is the two time-scale algorithm described in Table III with the difference that the inner problem Step S4) is solved using general purpose solvers.

3) **In-SCA for HetNet**: This algorithm is the inexact algorithm described in Table IV where we choose $\beta_k = \beta$ for all $k$, and each block is updated only once in Step S4).

4) **ZF-SCA for IBC**: This algorithm is the intra-cell ZF plus inter-cell CB algorithm described in Table II with $\beta_k = \beta = 0$ for all $k$;

5) **Per-Cell ZF for HetNet** [37]: This algorithm is Algorithm 2 proposed in [37]. It performs the intra-cell ZF for the HetNet setting with BS power constraint, while completely ignoring the inter-cell

---

2) In particular, we first formulate the inner problem into a SOCP, and then solved using MOSEK; see Remark 5.
interference.

All the algorithms considered in this subsection use the system sum rate as the users’ utility function. The plots to be shown represent the averaged performance of the algorithms running over 100 independent network generations.

![Fig. 3. Comparison of the CPU Time and the total iteration number required for computation in HetNet setting with different sizes of the network. K = [2, 4, 6, 8, 10], P_{tot}^k = 20dB, |Q_k| = 6, |I_k| = 10, M = 5, N = 3, d_{i_k} = 3 for all i_k ∈ I.](image)

Our first set of experiments compare the performance of the first three algorithms listed above. In Fig. 2–Fig. 3 the averaged system sum rate achieved by different algorithms as well as the averaged CPU time and total number of iterations used are compared for a network with |Q_k| = 6, M = 5, N = 3, |I_k| = 10 and P_{tot}^k = 20dB. For the WMMSE, the per-BS power constraint is completely ignored. Instead, a single per-cell power budget is assumed. Several interesting observations can be made. First, in the HetNet setting the In-SCA is much more efficient than the SCA. Second, the In-SCA uses almost the same number of iterations as the SCA algorithm, which implies that they also require similar amount of message exchanges among the cells. Note that to ensure convergence of the In-SCA algorithm, we have chosen some small but positive β’s. Our experiments suggest that In-SCA is not quite sensitive to the choice of β. It even works well when β = 0, although we have omitted the latter result due to its similarity to the cases when β = {0.01, 0.05}.

The second set of experiments compare the general linear precoding and the ZF precoding. In Fig. 4 we show the performance of the algorithms with K = 6 cells, |Q_k| = 6, |I_k| = \{2,4,\cdots,12\}, M = \{4,6\}, N = 2 and d_{i_k} = 2, P_{tot}^k = 20dB. Note that when increasing the number of users |I_k|, more resources are dedicated to eliminating the intra-cell interference. However, as suggested in Fig. 4...
this is not an ideal strategy to deal with interference in densely deployed HetNet. The performance of the ZF-based strategies degrades as $|I_k|$ increases. Further, when $M = 4$ and when $|I_k|$ approaches the maximum number of allowable users for which the ZF strategy is still feasible (12 in this case), the performance degradation is even more severe. The reason is that when the cells and the BSs are densely deployed, inter-cell interference is equally detrimental as the intra-cell interference on the users’ rates. The general linear precoding approach, without pre-specifying the source of interference to be mitigated, appears to be a more balanced way of dealing with the interference.

![Comparison of system sum rate achieved for different algorithms with different number of users per cell.](image)

**Fig. 4.** Comparison of system sum rate achieved for different algorithms with different number of users per cell. $K = 6$, $P_k^\text{tot} = 20\text{dB}$, $|Q_k| = 6$, $|I_k| = [2, 4, 6, \cdots, 12]$, $N = 2$, $d_{ik} = 2$ for all $i_k \in I$. Left: performance with $M = 6$. Right: performance with $M = 4$.

### B. Partial ComP in HetNet

In this section, we jointly design the clustering and linear precoding schemes in a partial ComP setting. To induce the desired clustering structure, we specialize the penalty terms in the objective to take the following form \cite{34, 36}:

$$s_{ik}(V) = \lambda \|V_{ik}^q\|_F, \ orall i_k, q_k$$

where $\lambda > 0$ is chosen appropriately to balance the resulting group size and the throughput performance. To provide certain level of fairness among the users, we use the geometric mean utility of one plus the user’s rate, i.e.,

$$u_{ik}(R_{ik}) = \log(1 + R_{ik}).$$

We compare the performance of the following algorithms: (1) WMMSE, the baseline algorithm that optimizes the precoders by treating all the BSs in each cell as a single virtual BS; (2) The algorithm proposed in \cite{34}, which can be viewed as a special case of the SCA algorithm for solving the penalized utility maximization problem, see Remark \cite{4}; (3) In-SCA algorithm in Table IV with $\beta = 0.001$. 

August 19, 2014

DRAFT
The results are summarized in Fig. 5 - 6 as well as in Table V. We see that both the SCA and the In-SCA based approaches are able to keep a large portion of the system sum rate achieved by the full per-cell cooperation, while using only small cluster sizes. In contrast, the precoders generated by the WMMSE are void of the desired structure: they always mandate all the BSs to transmit to each user. SCA and In-SCA use similar number of iterations and achieve similar performance. The advantage of using the inexact version is mainly in the efficiency of computation.

![Comparison of system performance](image)

**Fig. 5.** Comparison of the system performance of different algorithms in HetNet. $K = [1, 2, \cdots, 10]$, $P_{k}^{\text{tot}} = 20\text{dB}$, $|Q_k| = 10$, $|I_k| = [10, 20]$, $M = 4$, $N = 2$, $d_{ik} = 1$ for all $i_k \in I$. $\lambda = 0.1$ when $|I_k| = 10$, and $\lambda = 0.05$ when $|I_k| = 20$. Left: System throughput. Right: Averaged cluster size.

**TABLE V**

| CPU TIME NEEDED FOR DIFFERENT ALGORITHMS (UNIT: SECOND) |
|----------------------------------------------------------|
| K=2 | K=4 | K=6 | K=8 | K=10 |
| WMMSE ($I_k = 20$) | 20.1 | 61.3 | 73.4 | 89.9 | 120.3 |
| SCA ($I_k = 20$) | 103.1 | 200.1 | 263.3 | 368.4 | 450.5 |
| IN-SCA ($I_k = 20$) | 57.9 | 88.2 | 109.9 | 153.1 | 189.1 |
| WMMSE ($I_k = 10$) | 5.2 | 15.2 | 31.2 | 39.3 | 58.1 |
| SCA ($I_k = 10$) | 15.8 | 40.1 | 67.3 | 136.1 | 170.3 |
| IN-SCA ($I_k = 10$) | 7.4 | 19.6 | 40.1 | 45.3 | 72.3 |

**VI. CONCLUSION**

In this paper we have addressed an important family of interference management problems arising in the heterogeneous networks. The main novelty of this work lies in the proposal to achieve decomposition.
across the interference-coupled networks by using the technique of successive convex approximation. Our proposed approach is of low computational complexity, as each of the subproblems to be solved is convex and completely decomposes across the cells. Depending on whether the subproblems are solved exactly or inexactly, two general algorithms have been proposed, both of which find applications in practical interference management problems. We believe that the framework studied in this paper is extendable to many other problems for interference management beyond those mentioned in this work.

VII. ACKNOWLEDGEMENT

The authors wish thank Prof. Zhi-Quan Luo from University of Minnesota and Dr. Zi Xu from Shanghai University for helpful discussions in many technical issues in this paper.
\[
\frac{dl_{ik}(V_{ik} + tD_{ik}, C_{ik} + tM_{ik})}{dt} = d\text{Tr}\left[\hat{E}_{ik}^{-1}(V_{ik} + tD_{ik})^{H}(H_{ik}^{k})^{H}(C_{ik} + tM_{ik})^{-1}H_{ik}^{k}(V_{ik} + tD_{ik})\right] \\
= \text{Tr}\left[\hat{E}_{ik}^{-1}(D_{ik})^{H}(H_{ik}^{k})^{H}(C_{ik} + tM_{ik})^{-1}H_{ik}^{k}(V_{ik} + tD_{ik})\right] \\
+ \text{Tr}\left[\hat{E}_{ik}^{-1}(V_{ik} + tD_{ik})^{H}(H_{ik}^{k})^{H}(C_{ik} + tM_{ik})^{-1}M_{ik}(C_{ik} + tM_{ik})^{-1}H_{ik}^{k}(V_{ik} + tD_{ik})\right] \\
- \text{Tr}\left[\hat{E}_{ik}^{-1}(V_{ik} + tD_{ik})^{H}(H_{ik}^{k})^{H}(C_{ik} + tM_{ik})^{-1}M_{ik}(C_{ik} + tM_{ik})^{-1}H_{ik}^{k}(V_{ik} + tD_{ik})\right].
\]

\[\text{(38)}\]

**APPENDIX**

**A. Derivation of the equality in (12)**

Let us investigate the first order directional derivative of the function \(l_{ik}\). We can first obtain (38).

Let \(V_{ik} = \tilde{V}_{ik}, D_{ik} = (V_{ik} - \tilde{V}_{ik}), C_{ik} = \tilde{C}_{ik}\) and \(M_{ik} = 0\), we obtain the expression

\[
\frac{dl_{ik}(\tilde{V}_{ik} + t(V_{ik} - \tilde{V}_{ik}), \tilde{C}_{ik})}{dt}\bigg|_{t=0} = \text{Tr}\left[\hat{E}_{ik}^{-1}(V_{ik} - \tilde{V}_{ik})^{H}(H_{ik}^{k})^{H}\tilde{C}_{ik}^{-1}H_{ik}^{k}\tilde{V}_{ik}\right] \\
+ \text{Tr}\left[\hat{E}_{ik}^{-1}\tilde{V}_{ik}^{H}(H_{ik}^{k})^{H}\tilde{C}_{ik}^{-1}H_{ik}^{k}(V_{ik} - \tilde{V}_{ik})\right].
\]

Similarly, let \(V_{ik} = \tilde{V}_{ik}, D_{ik} = 0, C_{ik} = \tilde{C}_{ik}\) and \(M_{ik} = (C_{ik} - \tilde{C}_{ik})\), we obtain the expression

\[
\frac{dl_{ik}(\tilde{V}_{ik}, \tilde{C}_{ik} + t(C_{ik} - \tilde{C}_{ik}))}{dt}\bigg|_{t=0} = -\text{Tr}\left[\hat{E}_{ik}^{-1}\tilde{V}_{ik}^{H}(H_{ik}^{k})^{H}\tilde{C}_{ik}^{-1}(C_{ik} - \tilde{C}_{ik})\tilde{C}_{ik}^{-1}H_{ik}^{k}\tilde{V}_{ik}\right].
\]

This completes the proof.

**B. Proof of Theorem 7**

We first prove the convergence for the SCA algorithm. It is easy to observe that the objective value of the problem \(P_{\text{SYSTEM}}\) is monotonically nondecreasing, i.e., we have

\[
u(V(t + 1)) \overset{(i)}{\geq} h^3(V(t + 1); V(t)) - s(V(t + 1)) \\
\overset{(ii)}{\geq} h^3(V(t); V(t)) - s(V(t)) \overset{(iii)}{=} u(V(t)).
\]

\[\text{(39)}\]
where step (i) is from [15]; step (ii) is from the fact that $V(t + 1)$ is the solution to the problem (PLower-Bound); step (iii) is because of [15]. As both $f(V)$ and $s(V)$ are upper bounded for all $V$ in the feasible set, it follows that the sequence $\{u(V(t))\}$ converges. Let $\bar{u}$ denote its limit. This result combined with (39) implies that

$$\lim_{t \to \infty} h^\beta(V(t + 1); V(t) - s(V(t + 1)) = \bar{u}. \quad (40)$$

Using a similar argument as in [40, Lemma 1], and use the fact that $f(\cdot)$ and $h(\cdot; \tilde{V})$ satisfy [15], we can show that for any feasible $\tilde{V}$ the directional derivative of $f(\cdot)$ at the point $\tilde{V}$ equals that of $h^\beta(\cdot; \tilde{V})$ at the point $\tilde{V}$, i.e.,

$$\lim_{r \to 0} \left( \frac{f(\tilde{V} + rD) - f(\tilde{V})}{r} \right) = \lim_{r \to 0} \left( \frac{h^\beta(\tilde{V} + rD; \tilde{V}) - h^\beta(\tilde{V})}{r} \right) \quad (41)$$

where $D$ is any feasible direction that satisfies $\tilde{V} + D \in V$.

Let $\{V(t_m)\}_{m=1}^\infty$ be a converging subsequence of $V(t)$, and denote its limit as $V^*$. Clearly we have

$$h^\beta(V^*; V^*) - s(V^*) = u(V^*) = \bar{u}.$$  

Then by Step 1 of the algorithm, the optimality of $V(t)$ to the problem (PLower-Bound) implies that

$$h^\beta(V(t_m + 1); V(t_m)) - s(V(t_m + 1)) \geq h^\beta(V; V(t_m)) - s(V), \quad \forall V \in V. \quad (42)$$

Taking limit on both sides and use (40), we obtain $\bar{u} \geq h^\beta(V; V^*) - s(V), \forall V$. It follows that

$$h^\beta(V^*; V^*) - s(V^*) \geq h^\beta(V; V^*) - s(V)$$

for all feasible $V$, which further implies that

$$h^\beta_D(V; V^*) \bigg|_{V=V^*} - s^*_D(V) \bigg|_{V=V^*} \leq 0, \quad \forall D + V^* \in V. \quad (43)$$

Utilizing (41), we obtain

$$u^*_D(V) \bigg|_{V=V^*} = f^*_D(V) \bigg|_{V=V^*} - s^*_D(V) \bigg|_{V=V^*} \leq 0,$$

$$\forall D + V^* \in V,$$

(44)

which says that $V^*$ is a stationary solution to the problem (Psystem).
We then prove the result for the IN-SCA algorithm. Similarly as the previous case, we have

\[
u(V(t+1)) = \begin{cases} h^\beta(V(t); V(t)) - s(V(t+1)) & \text{if } (i) \\ \geq h^\beta(V(t); V(t)) - s(V(t)) + \eta \|V(t+1) - V(t)\|^2_F & \text{else} \end{cases} \]

\[
= u(V(t)) + \eta \|V(t+1) - V(t)\|^2_F.
\]

(45)

where (i) is due to condition (20). The above inequality implies that

\[
u(V(t+1)) - \nu(V(0)) \geq \sum_{t=1}^t \eta \|V(t+1) - V(t)\|^2_F.
\]

(46)

Taking limit on both sides, and using the fact that \(\nu(V)\) is upper-bounded for all \(V \in \mathcal{V}\), we obtain

\[
\lim_{t \to \infty} \|V(t+1) - V(t)\|^2_F = 0,
\]

\[
\lim_{t \to \infty} \nu(V(t+1)) = \bar{\nu}.
\]

(47)

Let \(\{V(t_m)\}_{m=1}^\infty\) be a converging subsequence of \(V(t)\), and denote its limit as \(V^*\). Combining the first equation in (47) and the second condition in (20), we have

\[
V^* = \arg \min_V h^\beta(V; V^*) - s(V).
\]

(48)

or equivalently

\[
h^\beta(V^*; V^*) - s(V^*) \geq h^\beta(V; V^*) - s(V), \quad \forall \ V \in \mathcal{V}.
\]

(49)

Then the conclusion follows from a similar argument as in the first part of the proof.

REFERENCES

[1] 3GPP, “Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN): overall description,” 2011, 3GPP TS 36.300, V8.9.0.

[2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, “A survey on 3GPP heterogeneous networks,” IEEE Wireless Communications, vol. 18, pp. 10 –21, 2011.

[3] M. Hong and Z.-Q. Luo, “Signal processing and optimal resource allocation for the interference channel,” in Academic Press Library in Signal Processing, Academic Press, 2013.

[4] E. Bjornson and E. Jorswieck, “Optimal resource allocation in coordinated multi-cell systems,” Foundations and Trends in Communications and Information Theory, vol. 9, 2013.

[5] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” IEEE Journal on Selected Areas in Communications, vol. 28, no. 9, pp. 1380 –1408, 2010.
[6] Z. K. M. Ho and D. Gesbert, “Balancing egoism and altruism on interference channel: The MIMO case,” in *2010 IEEE International Conference on Communications (ICC)*, 2010, pp. 1 –5.

[7] E. Jorswieck and E. Larsson, “The MISO interference channel from a game-theoretic perspective: A combination of selfishness and altruism achieves pareto optimality,” in *IEEE ICASSP*, 2008, pp. 5364 –5367.

[8] Y.-F Liu, Y.-H. Dai, and Z.-Q. Luo, “Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 1142 –1157, 2011.

[9] J. Zhang, R. Chen, J.G. Andrews, A. Ghosh, and R.W. Heath, “Networked MIMO with clustered linear precoding,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 8, pp. 1910–1921, 2009.

[10] Z.-Q. Luo and S. Zhang, “Dynamic spectrum management: Complexity and duality,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, 2008.

[11] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness,” *Signal Processing*, vol. 93, no. 12, pp. 3327–3340, 2013.

[12] Y.-F Liu, Y.-H. Dai, and Z.-Q. Luo, “Max-min fairness linear transceiver design for a multi-user MIMO interference channel,” *IEEE Transactions on Signal Processing*, vol. 61, no. 9, pp. 2413–2423, 2013.

[13] Y.-F Liu, M. Hong, and Y.-H. Dai, “Max-min fairness linear transceiver design problem for a multi-user SIMO interference channel is polynomial time solvable,” *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 27 –30, 2013.

[14] D. P. Palomar and M. Chiang, “A tutorial on decomposition methods for network utility maximization,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439 –1451, 2006.

[15] S. Ye and R.S. Blum, “Optimized signaling for MIMO interference systems with feedback,” *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2839 –2848, 2003.

[16] C. Shi, R. A. Berry, and M. L. Honig, “Monotonic convergence of distributed interference pricing in wireless networks,” in *Proc. of IEEE international conference on Symposium on Information Theory*, 2009.

[17] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen, “Improved dual decomposition based optimization for DSL dynamic spectrum management,” *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2230 –2245, 2010.

[18] M. Hong and Z.-Q. Luo, “Joint linear precoder optimization and base station selection for an uplink MIMO network: A game theoretic approach,” in *the Proceedings of the IEEE ICASSP*, 2012.

[19] S.-J. Kim and G. B. Giannakis, “Optimal resource allocation for MIMO Ad Hoc Cognitive Radio Networks,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3117 –3131, 2011.

[20] J. Papandriopoulos and J. S. Evans, “SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks,” *IEEE Transactions on Information Theory*, vol. 55, no. 8, 2009.

[21] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 641–656, 2014.

[22] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, “Distributed dynamic pricing for MIMO interfering multiuser systems: A unified approach,” in *Int. Conf. on NETwork Games, COntrol and OPtimization (NetGCooP 2011)*, Oct. 2011, pp. 12 –14.

[23] P. C. Weeraddana, M. Codreanu, M. Latva-aho, A. Ephremides, and C. Fischione, “Weighted sum-rate maximization in wireless networks: A review,” *Foundations and Trends in Networking*, vol. 6, 2012.

[24] G. Scutari, F. Facchinei, P. Song, D. Palomar, and Jong-Shi Pang, “Decomposition by partial linearization in multiuser systems,” in *Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 13)*, 2013.
[25] L. Venturino, N. Prasad, and X. Wang, “Coordinated linear beamforming in downlink multicell wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1451–1461, 2010.

[26] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[27] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory, second edition*, Wiley, 2005.

[29] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.

[30] H. Boche, S. Naik, and T. Alpcan, “Characterization of convex and concave resource allocation problems in interference coupled wireless systems,” *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2382–2394, 2011.

[31] H. Boche and M. Schubert, “A unifying approach to interference modeling for wireless networks,” *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3282–3297, 2010.

[32] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, “Distributed interference pricing for the MIMO interference channel,” in *IEEE International Conference on Communications*, 2009, pp. 1–5.

[33] C. T. K. Ng and H. Huang, “Linear precoding in cooperative MIMO cellular networks with limited coordination clusters,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1446–1454, 2010.

[34] M. Hong, R. Sun, H. Baligh, and Z.-Q. Luo, “Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks,” *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 226–240, 2013.

[35] S.-J. Kim, S. Jainand, and G.B. Giannakis, “Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering,” in *2012 IEEE SPWAC*, 2012.

[36] M. Hong, M. Razaviyayn, R. Sun, and Z.-Q. Luo, “Joint transceiver design and base station clustering for heterogeneous networks,” in *Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 574–578.

[37] R. Zhang, “Cooperative multi-cell block diagonalization with per-base-station power constraints,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1435–1445, 2010.

[38] Q. Spencer, A. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.

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

[40] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[41] David G. Luenberger, *Linear and Nonlinear Programming, Second Edition*, Springer, 1984.

[42] W. Yu and T. Lan, “Transmitter optimization for the multi-antenna downlink with per-antenna power constraints,” *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646–2660, 2007.

[43] P. Tseng and S. Yun, “Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization,” *Journal of Optimization Theory and Applications*, vol. 140, pp. 513–535, 2009.

[44] P. Tseng, “Convergence of a block coordinate descent method for nondifferentiable minimization,” *Journal of Optimization Theory and Applications*, vol. 103, no. 9, pp. 475–494, 2001.
[45] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, no. 1, pp. 49–67, 2006.

[46] H. Zhu, N. Prasad, and S. Rangarajan, “Precoder design for physical layer multicasting,” IEEE Transactions on Signal Processing, vol. 60, no. 11, pp. 5932–5947, Nov 2012.

[47] P. Tseng and S. Yun, “A coordinate gradient descent method for nonsmooth separable minimization,” Mathematical Programming, vol. 117, pp. 387–423, 2009.
This figure "figBSClustering.png" is available in "png" format from:

http://arxiv.org/ps/1210.1507v3