IOWA STATE UNIVERSITY
Digital Repository

Industrial and Manufacturing Systems Engineering Publications | Industrial and Manufacturing Systems Engineering

2-2016

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

Mingyi Hong
Iowa State University, mingyi@iastate.edu

Qiang Li
University of Electronic Science and Technology of China

Ya-Feng Liu
Chinese Academy of Sciences

Follow this and additional works at: http://lib.dr.iastate.edu/imse_pubs

Part of the Industrial Engineering Commons, Systems Architecture Commons, and the Systems Engineering Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/imse_pubs/86. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Industrial and Manufacturing Systems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Industrial and Manufacturing Systems Engineering Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

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

Mingyi Hong, Qiang Li, Ya-Feng Liu

arXiv:1210.1507v4 [cs.IT] 20 Oct 2015

***Abstract*—We study the downlink linear precoder design problem in a multi-cell dense heterogeneous network (HetNet). The problem is formulated as a general sum-utility maximization (SUM) problem, which includes as special cases many practical precoder design problems such as multi-cell coordinated linear precoding, full and partial per-cell coordinated multi-point transmission, zero-forcing precoding and joint BS clustering and beamforming/precoding. The SUM problem is difficult due to its non-convexity and the tight coupling of the users' precoders. In this paper 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 hinges upon the identification of certain key convexity properties of the sum-utility objective, which allows us to transform the problem into a form that can be solved using a popular algorithmic framework called BSUM (Block Successive Upper-Bound Minimization). Simulation experiments show that the proposed framework is effective for solving interference management problems in large HetNet.**

I. Introduction

Heterogeneous network (HetNet) has recently emerged as a promising wireless network architecture [1]. In HetNet, the upper tier high-power Base Stations (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, enabling high link quality with low transmission power [2].

Due to the presence of a large number of potential interfering nodes, 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 interfering networks, has been under extensive research recently [3], [4]. An effective approach is to introduce appropriate coordination among the network nodes, either in the physical layer or in higher layers [5]. For example, physical-layer coordination can take the form of coordinated beamforming (CB) or joint processing (JP) [5]. The CB coordinates the nodes in the beamformer/precoder level [6], [7], while the JP (a.k.a. the coordinated multi-point transmission, CoMP) optimizes the transceivers assuming that the users' data are available at all the BSs. In practice, dynamic combination of the two techniques is usually adopted for signaling overhead reduction [5], [8].

Mathematically, interference management is commonly formulated as some form of sum-utility maximization (SUM) problem [3]. The 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 popular utility functions, the associated optimization problems are difficult to solve (except for a few special cases, see [7], [9]–[11]). Therefore, many low-complexity algorithms have been recently developed. Often, the key in finding a good algorithm is to recognize certain convexity and decomposability of the SUM problem, as the former leads to efficient computation, while the latter is critical for distributed implementation [12].

The hidden convexity structures of various SUM problems have been extensively investigated in the past. For example, reference [13] is the first to recognize the now well-known concave-convex property for a MIMO interfering channel (IC), i.e., each users' achievable rate is concave in its own transmit covariance while being convex in all other interfering users' transmit covariances. Similar concave-convex properties have since been leveraged heavily to design resource allocation algorithms in different network settings [6], [7], [14]–[21]. However, most of the schemes cited above 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 [18], 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 transmiting to a single receiver, and/or with each receiver receiving from a single transmitter. Hence they are not directly applicable to the HetNet setting where each BS or user can communicate with multiple nodes. Further, all the schemes mentioned above deal with smooth SUM problems, therefore they are not applicable to formulations that require nonsmooth penalizations [22]. Classical subgradient (SG) method [23] can potentially be used to solve such nonsmooth optimization, but in practice the SG can be very slow and its performance

M. Hong is with the Department of Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, USA (email: mingyi@iastate.edu). Q. Li is with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, China (email: lq@uestc.edu.cn). Y.-F. Liu is with the State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China (email: yafliu@lsec.cc.ac.cn). M. Hong is supported in part by NSF under Grant CCF-1526078. Q. Li is supported in part by the National Natural Science Foundation of China under Grant 61401073 and 61531009. Y.-F. Liu is partially supported by the National Natural Science Foundation of China, Grants 11301516 and 11331012.

depends critically on the choice of the stepsizes.

Nevertheless, decomposability structure of interference management problems is highly desirable. When judiciously exploited, it leads to efficient distributed implementation. This fact has long been recognized and leveraged in other important large-scale network optimization problems such as the network utility maximization (NUM) problems; see [12]. However, unlike the NUM, the nodes in the interfering networks are tightly coupled in a nonlinear manner through multi-user interference. Therefore even in simple peer-to-peer interfering networks, decomposability structure is difficult to come by. To address this issue, reference [24] proposes a weighted minimum mean square error (WMMSE) algorithm to solve the SUM problem in interfering broadcast channels (IBC). The most desirable feature of WMMSE is that each of its steps completely decouples among the interfering BSs. Stationary solutions of the SUM problem can be obtained via solving three subproblems alternatingly. Recently, there are a few works [19], [20], [25], [26] developing general parallel schemes for distributed optimization in multi-agent systems, which cover interfering wireless networks as a special case. For example, in [19] an interesting parallelization technique based on partial linearization is proposed. Generally speaking, there are subtle differences between parallelization and decomposition. The parallelization techniques such as those developed in [19], [20], [25], [26] gear toward solving a general multi-block problem a few blocks a time, but all requiring careful stepsize tuning. The decomposition techniques such as those used in [12] and the ones to be discussed in this paper are more problem specific: they often find an alternative representation of the problem in an iterative manner, so that the variables automatically become independent (hence the name decomposable). We will elaborate more on the distinctions between these two types of approaches in the following presentation.

In this work, we propose to achieve decomposability by means of *successive convex approximation*. The key novelty of our work is the identification of an interesting hidden convexity for a wide range of sum-utility maximization problems, which allows us to interpret and generalize many existing algorithms under the recently developed BSUM (Block Successive Upper-Bound Minimization) framework [27], [28]. By using BSUM, the original non-convex problem is approximated by a series of convex subproblems, each of which is completely decoupled among the network nodes. 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 describe the system model and problem formulation. In Section III, we present a key convexity structure which leads to two general SCA algorithms. In Section IV, the proposed algorithms are specialized to various interference management scenarios. Numerical results are given in Section V, followed by concluding remarks.

Notations: For a symmetric matrix $\mathbf{X}$, $\mathbf{X} \succeq 0$ signifies that $\mathbf{X}$ is Hermitian positive semi-definite. We use $\mathrm{Tr}[\mathbf{X}]$, $|\mathbf{X}|$, $\mathbf{X}^H$, $\rho(\mathbf{X})$, $\|\mathbf{X}\|_F$ and $\mathrm{Rank}(\mathbf{X})$ 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 $\mathbf{X}$ is denoted by $\mathbf{X}[m,n]$. We use $\mathbf{I}_n$ 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. A list of other main notations are given in the following table.

II. System Model and Problem Formulation

A. Problem Description

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 BSs such as macro/micro/pico BSs which provide service to the users. 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. The central controller 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, and let $I = |\mathcal{I}|$ and $Q = |\mathcal{Q}|$. Assume that each BS has M transmit antennas, and each user has N receive antennas. Let $\mathbf{H}^{q_\ell}_{i_k} \in \mathbb{C}^{N\times M}$ denote the channel matrix between the q-th BS in the ℓ-th cell and the i-th user in the k-th cell. Similarly, we use $\mathbf{H}^{\ell}_{i_k}$ to denote the channel matrix between all the BSs in the ℓ-th cell to the user i_k, i.e., $\mathbf{H}^{\ell}_{i_k} := \left\{\mathbf{H}^{q_\ell}_{i_k}\right\}_{q_\ell \in \mathcal{Q}_\ell} \in \mathbb{C}^{N\times MQ_\ell}$.

Suppose that $d \le \min\{M, N\}$ data streams are transmitted to user i_k (for notational simplicity we assume that all users have the same number of streams). Let $\mathbf{V}^{q_k}_{i_k} \in \mathbb{C}^{M\times d}$ denote the transmit precoder that BS q_k uses to transmit data $\mathbf{s}_{i_k} \in \mathbb{C}^d$ to user i_k. Define

$$\mathbf{V}_{i_k} := \left\{\mathbf{V}^{q_k}_{i_k}\right\}_{q_k\in\mathcal{Q}_k},\ \mathbf{V}^{q_k} := \left\{\mathbf{V}^{q_k}_{i_k}\right\}_{i_k\in\mathcal{I}_k},\ \mathbf{V}^{k} := \left\{\mathbf{V}_{i_k}\right\}_{i_k\in\mathcal{I}_k}$$

as the collection of all precoders intended for user i_k, all precoders belong to BS q_k, and all precoders in cell k, respectively. Let $\mathbf{V} := \{\mathbf{V}_{i_k}\}_{i_k\in\mathcal{I}}$. Further assume that all the BSs in cell k form a single virtual BS, and they jointly transmit to user $i_k \in \mathcal{I}_k$. Then $\mathbf{V}_{i_k}$ can be viewed as the *virtual precoder* for user i_k. Using these definitions, we can express the transmitted signal of BS q_k and the combined transmitted signal for all the BSs in cell k as:

$$\mathbf{x}^{q_k} = \sum_{i_k\in\mathcal{I}_k} \mathbf{V}^{q_k}_{i_k}\mathbf{s}_{i_k} \in \mathbb{C}^{M\times 1}, \qquad \mathbf{x}^{k} = \sum_{i_k\in\mathcal{I}_k} \mathbf{V}_{i_k}\mathbf{s}_{i_k} \in \mathbb{C}^{MQ_k\times 1}.$$

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

$$\mathbf{y}_{i_k} = \mathbf{H}^{k}_{i_k}\mathbf{V}_{i_k}\mathbf{s}_{i_k} + \sum_{j_k\in\mathcal{I}_k\setminus i_k} \mathbf{H}^{k}_{i_k}\mathbf{V}_{j_k}\mathbf{s}_{j_k} + \sum_{\ell\neq k}\sum_{j_\ell\in\mathcal{I}_\ell} \mathbf{H}^{\ell}_{i_k}\mathbf{V}_{j_\ell}\mathbf{s}_{j_\ell} + \mathbf{z}_{i_k}$$

where $\mathbf{z}_{i_k} \in \mathbb{C}^{N\times 1}$ is the complex Gaussian noise with distribution $\mathcal{CN}(0, \sigma^2\mathbf{I}_N)$. Let $\mathbf{U}_{i_k} \in \mathbb{C}^{N\times d}$ denote the receiver used by user i_k to decode the intended signal. Then the estimated signal for user i_k is: $\widehat{\mathbf{s}}_{i_k} = \mathbf{U}^H_{i_k}\mathbf{y}_{i_k}$. The mean

K	# of cells	Q_k	# of BSs in cell k
i_k	i-th user in k-th cell	q_k	q-th BS in k-th cell
$\mathbf{H}_{i_k}^{q_\ell}$	channel matrix between the q_ℓ-th BS and i_k-th user	I_k	# of users in cell k
$\mathbf{H}_{i_k}^{\ell}$	channel matrix between the all BSs in cell ℓ and i_k-th user	d	# of data streams for each user
$\mathbf{V}_{i_k}^{q_k}$	precoder used by q_k-th BS for i_k-th user	$\mathbf{U}_{i_k}$	The receiver used by i_k-th user
$\mathbf{V}_{i_k}$	precoder used by all BSs in cell k for i_k-th user	$\mathbf{E}_{i_k}$	The MSE matrix for i_k-th user
$\mathbf{C}_{i_k}$ ($\mathbf{Y}_{i_k}$)	total received signal (interference) covariance matrix of user i_k	R_{i_k}	transmit rate for user i_k

TABLE I
A LIST OF NOTATIONS

square error (MSE) for user i_k can be written as

$$\begin{aligned}\mathbf{E}_{i_k} &:= \mathbb{E}[(\mathbf{s}_{i_k} - \widehat{\mathbf{s}}_{i_k})(\mathbf{s}_{i_k} - \widehat{\mathbf{s}}_{i_k})^H] \\ &= (\mathbf{I}_d - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k})(\mathbf{I}_d - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k})^H \\ &\quad + \sum_{(\ell,j)\neq(k,i)} \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} \mathbf{V}_{j_\ell}^H (\mathbf{H}_{i_k}^\ell)^H \mathbf{U}_{i_k} + \sigma_{i_k}^2 \mathbf{U}_{i_k}^H \mathbf{U}_{i_k}.\end{aligned} \tag{1}$$

The Minimum MSE (MMSE) receiver is given by [29]

$$\begin{aligned}\mathbf{U}_{i_k}^{\mathrm{mmse}} &= \left(\sum_{(\ell,j)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} \mathbf{V}_{j_\ell}^H (\mathbf{H}_{i_k}^\ell)^H + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \\ &:= \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k}\end{aligned} \tag{2}$$

where $\mathbf{C}_{i_k} \in \mathbb{S}_{++}^N$ is user i_k's received signal covariance matrix. Plugging (2) into (1) we obtain

$$\mathbf{E}_{i_k}^{\mathrm{mmse}} = \mathbf{I}_d - \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \succeq 0. \tag{3}$$

Clearly we also have $\mathbf{I}_d - \mathbf{E}_{i_k}^{\mathrm{mmse}} \succeq 0$.

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 [30]

$$R_{i_k} = \log\left|\mathbf{I}_N + \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{Y}_{i_k}^{-1}\right| = -\log\left|\mathbf{E}_{i_k}^{\mathrm{mmse}}\right| \tag{4}$$

where we have defined the matrix $\mathbf{Y}_{i_k} := \sum_{(\ell,j)\neq(k,i)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} \mathbf{V}_{j_\ell}^H (\mathbf{H}_{i_k}^\ell)^H + \sigma^2 \mathbf{I}_N$ as the received interference covariance matrix for user i_k; the last equality is a well-known relationship between the rate and the MMSE matrix (see e.g., [31]). We will occasionally use the notations $R_{i_k}(\mathbf{V})$, $\mathbf{C}_{i_k}(\mathbf{V})$ and $\mathbf{E}_{i_k}^{\mathrm{mmse}}(\mathbf{V})$ to make their dependencies on $\mathbf{V}$ explicit.

Let $f_{i_k}(\bullet) : \mathbb{R}_+ \to \mathbb{R}$ denote the utility function of user i_k's data rate. Let $s_{i_k}^{q_k}(\bullet)$ denote a penalty term for $\mathbf{V}_{i_k}^{q_k}$, which is useful in inducing certain structures on the precoders such as sparsity or group sparsity [22]. We assume that these functions satisfy the following assumptions:

Assumption A.

- **A-1)** $f_{i_k}(x)$ is a concave non-decreasing function in x for all $x \geq 0$;
- **A-2)** $f_{i_k}(-\log(|\mathbf{X}|))$ is convex in $\mathbf{X}$, for all $\mathbf{I} \succeq \mathbf{X} \succeq 0$;
- **A-3)** $f_{i_k}(x)$ is continuously differentiable;
- **A-4)** $s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k})$ is a convex, continuous, but possibly non-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 [24]).

We consider the general system-level sum utilities maximization problem given below

$$\begin{aligned}\max_{\mathbf{V}} \quad & u(\mathbf{V}) := f(\mathbf{V}) - s(\mathbf{V}) \qquad (\mathrm{P}_{\mathrm{SYSTEM}}) \\ \text{s.t.} \quad & f(\mathbf{V}) := \sum_{k\in\mathcal{K}} \sum_{i_k\in\mathcal{I}_k} f_{i_k}(R_{i_k}(\mathbf{V})) \\ & s(\mathbf{V}) := \sum_{k\in\mathcal{K}} \sum_{i_k\in\mathcal{I}_k} \sum_{q_k\in\mathcal{Q}_k} s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k}) \\ & \mathbf{V}^{q_k} \in \mathcal{V}^{q_k},\ \forall\ q_k \in \mathcal{Q}, \\ & \mathbf{V}^k \in \mathcal{V}^k,\ \forall\ k \in \mathcal{K}\end{aligned}$$

where $\mathcal{V}^{q_k}$ and $\mathcal{V}^k$ are some feasible sets for $\mathbf{V}^{q_k}$ and $\mathbf{V}^k$, respectively. Let $\mathcal{V}$ denote the feasible set for $\mathbf{V}$. In the next subsection we provide a few transceiver design problems arising in HetNet that are covered by the model (P_{SYSTEM}).

B. Applications

1) **MIMO IBC/IMAC/IC channels with inter-BS CB [7], [17], [31]–[33]:** Consider a network with $|\mathcal{Q}_k| = 1\ \forall\ k$, i.e., each cell has a single BS. In this case $s(\mathbf{V}) \equiv 0$, and the constraint set $\mathcal{V}^{q_k}$ is the same as $\mathcal{V}^k$, which is given by the following sum-power constrained set ($\bar{P}_k$ denotes the power budget for cell k)

$$\mathcal{V}^k = \left\{\mathbf{V}^k : \sum_{i_k\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H] \leq \bar{P}_k\right\}. \tag{5}$$

2) **Multicell MIMO with intra-cell CoMP and inter-cell CB [22], [34]:** The BSs in different cells cooperate in the precoder level, while those in the same cell perform JP. Each BS q_k has a separate power budget, denoted as $\bar{P}^{q_k}$:

$$\mathcal{V}^{q_k} = \left\{\mathbf{V}^{q_k} : \sum_{i_k\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k}\right\}. \tag{6}$$

This model generalizes those for the IBC model (e.g., [24], [33]), in that the sum-power constraint (5) is replaced by the set of *per-group of antennas* power constraints.

3) **Multicell MIMO with intra-cell partial CoMP and inter-cell CB [22], [35]:** A practical alternative to full CoMP is to implement a *partial CoMP* strategy, where each user is served by 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. To jointly design the BS clusters and the precoders, one needs to properly utilize the penalty term $s(\mathbf{V})$. The requirement that each user is served only by *a small number* of BSs translates to certain *block sparse* structure of $\mathbf{V}_{i_k} = \{\mathbf{V}_{i_k}^{q_k}\}_{q_k \in \mathcal{Q}_k}$, see [22], [36]–[38]. To induce such block sparsity, the penalty term can take the following form

$$s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k}) = \gamma_{i_k}^{q_k} \|\mathbf{V}_{i_k}^{q_k}\|_F, \tag{7}$$

with $\gamma_{i_k}^{q_k} \geq 0$ being some constant. See [22], [36]–[38] and the references therein for motivation of using (7). It is worth noting that most of the existing decomposition based algorithms [7], [18]–[20] cannot handle such nonsmooth problem.

4) **Multicell MIMO network with intra-cell ZF and inter-cell CB [8], [39]:** In this case all the BSs in the same cell jointly perform the ZF precoding (a.k.a. the block-diagonalization precoding) [40], while the BSs in different cells perform CB. The feasible sets are given as

$$\mathcal{V}^k = \left\{\mathbf{V}^k : \mathbf{H}_{j_k}^k \mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H (\mathbf{H}_{j_k}^k)^H = \mathbf{0}, \forall\, j_k \neq i_k,\ j_k, i_k \in \mathcal{I}_k\right\}$$

$$\mathcal{V}^{q_k} = \left\{\mathbf{V}^{q_k} : \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}^{q_k} (\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k}, \forall\, q_k \in \mathcal{Q}_k\right\}.$$

Despite the wide applicability of ($\mathrm{P_{SYSTEM}}$), solving it to its global optimality is often difficult (i.e., NP-hard) [7], [9]–[11]. The NP-hardness of the problem indicates that it is even unlikely 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, the tight coupling of the variables $\{\mathbf{V}^k\}_{k=1}^K$ in the objective often makes such task challenging. We mention that existing general optimization frameworks are not directly applicable to solve ($\mathrm{P_{SYSTEM}}$). For example the partial linearization approach proposed in [19], [20] cannot handle nonsmooth terms in the objective. The algorithms developed in [25], [26] are parallelization schemes that can potentially handle non-convex and nonsmooth objective. However when applied to ($\mathrm{P_{SYSTEM}}$), it requires that each subproblem (say for optimizing BS q_k's precoder $\mathbf{V}^{q_k}$) or its approximation is easy to solve. It is not clear how to construct an effective approximation for ($\mathrm{P_{SYSTEM}}$) here. Further a careful stepsize tuning is needed to enable parallel update.

III. A Successive Convex Approximation Approach

In this section, we present our main approach for computing a high quality solution for the general problem ($\mathrm{P_{SYSTEM}}$). Our approach is to solve, possibly in an inexact manner, a series of convex subproblems each of which is a local approximation of ($\mathrm{P_{SYSTEM}}$).

We first highlight the main steps of our subsequent development. Our first step is to derive an approximation of the objective function $u(\mathbf{V}) = f(\mathbf{V}) - s(\mathbf{V})$ at a given feasible point $\widehat{\mathbf{V}}$, denoted as $h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) - s(\mathbf{V})$, for any feasible $\mathbf{V}$ ($\beta > 0$ is some constant parameter). The construction of the function $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ requires a novel two-stage convex approximation. It turns out that for every fixed input argument $\widehat{\mathbf{V}}$, our approximation is a global lower bound of $u(\mathbf{V})$, and it is easy to optimize with respect to $\mathbf{V}$. More importantly, this lower bound decouples completely among the variables, that is, we have the following decomposition structure

$$h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}} g_{i_k}^\beta(\mathbf{V}_{i_k}; \widehat{\mathbf{V}}), \tag{8}$$

for some $g_{i_k}^\beta(\bullet; \widehat{\mathbf{V}})$ which is only a function of $\mathbf{V}_{i_k}$. Our second step constructs algorithms that iteratively optimize $h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) - s(\mathbf{V})$ with respect to $\mathbf{V}$, while taking the previous iterate $\widehat{\mathbf{V}}$ as the input argument. Thanks to the decomposability structure (8), the resulting computation can be completely distributed into each cell.

A. A Local Approximation for ($\mathrm{P_{SYSTEM}}$)

We begin by deriving a simple local approximation for the individual users' utility function $f_{i_k}(\bullet)$ using the convexity assumption A-2). Let $\widehat{\mathbf{V}} \in \mathcal{V}$ denote a feasible solution to problem ($\mathrm{P_{SYSTEM}}$). Let $\widehat{\mathbf{E}}_{i_k} := \mathbf{E}_{i_k}^{\mathrm{mmse}}(\widehat{\mathbf{V}})$ denote the MMSE evaluated at $\widehat{\mathbf{V}}$. Leveraging (4), we can express $f_{i_k}(\bullet)$ as a function of $\mathbf{E}_{i_k}^{\mathrm{mmse}}$ only:

$$\begin{aligned}
&f_{i_k}(R_{i_k}(\mathbf{V})) = f_{i_k}(-\log|\mathbf{E}_{i_k}^{\mathrm{mmse}}|)\\
&\overset{\mathrm{(i)}}{\geq} f_{i_k}(-\log|\widehat{\mathbf{E}}_{i_k}|) - \left.\frac{\partial f_{i_k}(x)}{\partial x}\right|_{x=-\log|\widehat{\mathbf{E}}_{i_k}|} \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{E}_{i_k}^{\mathrm{mmse}} - \widehat{\mathbf{E}}_{i_k})\right]\\
&:= \widehat{a}_{i_k} - \widehat{c}_{i_k} \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{E}_{i_k}^{\mathrm{mmse}} - \widehat{\mathbf{E}}_{i_k})\right]\\
&:= \bar{h}_{i_k}\left(\mathbf{E}_{i_k}^{\mathrm{mmse}}, \mathbf{V}_{i_k}; \widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k}\right)
\end{aligned} \tag{9}$$

where $\widehat{c}_{i_k}$ and $\widehat{a}_{i_k}$ are two constants not related to either $\mathbf{E}_{i_k}^{\mathrm{mmse}}$ or $\mathbf{V}$; the inequality in (i) is due to the fact that the convex function is always lower bounded by its local linear approximation. Note that $\widehat{c}_{i_k} \geq 0$ due to the non-decreasing property of $f_{i_k}(\bullet)$'s assumed in A-1).

In fact, from the above derivation it is readily seen that $\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\mathrm{mmse}}, \mathbf{V}_{i_k}; \widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k})$ is a global lower bound of $f_{i_k}(\bullet)$ at the point $(\widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k})$, i.e., for all feasible $(\mathbf{E}_{i_k}^{\mathrm{mmse}}, \mathbf{V}_{i_k})$ and feasible $(\widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k})$ we have

$$\begin{aligned}
\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\mathrm{mmse}}, \mathbf{V}_{i_k}; \widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k}) &\leq f_{i_k}(-\log|\mathbf{E}_{i_k}^{\mathrm{mmse}}|)\\
\bar{h}_{i_k}(\widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k}; \widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k}) &= f_{i_k}(-\log|\widehat{\mathbf{E}}_{i_k}|).
\end{aligned}$$

Unfortunately, such approximation does not simplify the problem, as summing $\bar{h}_{i_k}(\bullet)$ over the users still results in a non-convex non-separable function with respect to $\mathbf{V}$ (cf. (3)).

Our next step performs a *second-stage approximation*, in which we further approximate $\bar{h}_{i_k}(\bullet)$ by a concave function of $\mathbf{V}$. To this end, we need a key lemma that explores some hidden convexity property of the function $\bar{h}_{i_k}(\bullet)$.

Lemma 1: The following function

$$l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) := \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k}\right]$$

is jointly convex with respect to the pair of variables $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$ *in the feasible region* $(\mathbb{C}^{MQ_k \times d}, \mathbb{S}_{++}^N)$.

The proof is given in Appendix A. The result provided by Lemma 1 allows us to further approximate the following function $\bar{h}_{i_k}(\bullet)$. To describe such second-stage approximation,

we define $\widehat{\mathbf{C}}_{i_k} := \mathbf{C}_{i_k}(\widehat{\mathbf{V}})$, and note

$$\begin{aligned}
&l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) = \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\mathbf{V}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\mathbf{C}_{i_k}^{-1}\mathbf{H}_{i_k}^k\mathbf{V}_{i_k}\right] \\
&\overset{\text{(i)}}{\geq} \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right] \\
&\quad + \left.\frac{dl_{i_k}(\widehat{\mathbf{V}}_{i_k} + t(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k}), \widehat{\mathbf{C}}_{i_k})}{dt}\right|_{t=0} \\
&\quad + \left.\frac{dl_{i_k}(\widehat{\mathbf{V}}_{i_k}, \widehat{\mathbf{C}}_{i_k} + t(\mathbf{C}_{i_k} - \widehat{\mathbf{C}}_{i_k}))}{dt}\right|_{t=0} \\
&\overset{\text{(ii)}}{=} \mathrm{Tr}[\widehat{\mathbf{E}}_{i_k}^{-1}] - d + \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k})\right] \\
&\quad + \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k})^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right] \\
&\quad - \mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}(\mathbf{C}_{i_k} - \widehat{\mathbf{C}}_{i_k})\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right] \qquad (10)
\end{aligned}$$

where (i) is due to the convexity of $l_{i_k}(\bullet)$ derived in Lemma 1; (ii) simply uses the definition of directional derivatives; see Appendix B for detailed derivation. The above inequality combined with the definition in (9) yields:

$$\begin{aligned}
&\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\mathrm{mmse}}, \mathbf{V}_{i_k}; \widehat{\mathbf{E}}_{i_k}, \widehat{\mathbf{V}}_{i_k}) \\
&\overset{\text{(i)}}{=} \widehat{a}_{i_k} + \widehat{c}_{i_k}\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\mathbf{V}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\mathbf{C}_{i_k}^{-1}\mathbf{H}_{i_k}^k\mathbf{V}_{i_k}\right] + \widehat{c}_{i_k}d - \widehat{c}_{i_k}\mathrm{Tr}[\widehat{\mathbf{E}}_{i_k}^{-1}] \\
&\overset{\text{(ii)}}{\geq} \widehat{a}_{i_k} + \widehat{c}_{i_k}\mathrm{Tr}[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k})] \\
&\quad + \widehat{c}_{i_k}\mathrm{Tr}[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k})^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}] \\
&\quad - \widehat{c}_{i_k}\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}(\mathbf{C}_{i_k} - \widehat{\mathbf{C}}_{i_k})\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right] \\
&\quad - \beta\left\|\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k}\right\|_F^2 \\
&\overset{\text{(iii)}}{=} \tilde{a}_{i_k} + \widehat{c}_{i_k}\mathrm{Tr}\Bigg[\widehat{\mathbf{E}}_{i_k}^{-1}\Bigg(\widehat{\mathbf{U}}_{i_k}^H\mathbf{H}_{i_k}^k\mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{U}}_{i_k} \\
&\quad - \widehat{\mathbf{U}}_{i_k}^H\Big(\sum_{(\ell,j)}\mathbf{H}_{i_k}^\ell\mathbf{V}_{j_\ell}(\mathbf{V}_{j_\ell})^H(\mathbf{H}_{i_k}^\ell)^H\Big)\widehat{\mathbf{U}}_{i_k}\Bigg)\Bigg] \\
&\quad - \beta\mathrm{Tr}\left[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H - \mathbf{V}_{i_k}\widehat{\mathbf{V}}_{i_k}^H - \widehat{\mathbf{V}}_{i_k}\mathbf{V}_{i_k}^H\right] := h_{i_k}^\beta(\mathbf{V}; \widehat{\mathbf{V}}) \qquad (11)
\end{aligned}$$

where (i) is due to the definition of the MMSE matrix (3); (ii) uses the inequality (10) and the fact that for any $\beta \geq 0$, the proximal term $-\beta\left\|\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k}\right\|_F^2$ is always nonnegative; in (iii), we have defined $\widehat{\mathbf{U}}_{i_k} := \widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}$ and used the definition of $\mathbf{C}_{i_k}$ in (2); $\tilde{a}_{i_k}$ collects all the terms that are not dependent on $\mathbf{V}$. It is interesting to observe that the newly defined quantity $\widehat{\mathbf{U}}_{i_k}$ is in fact the MMSE receiver for user i_k when the system precoder is given by $\widehat{\mathbf{V}}$ (cf. (2)).

We note that in the approximation (11) a proximal term $-\beta\left\|\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k}\right\|_F^2$ has been introduced, where $\beta \geq 0$ is some nonnegative constant. The reasons for introducing such term are twofold. First, it makes each approximation function $h_{i_k}^\beta$ strongly convex over $\mathbf{V}_{i_k}$, therefore easier to optimize (with a unique solution). Second, it helps to construct an algorithm with provable convergence guarantee. The latter point will be clarified shortly in Remark 4.

Combining (9) and (11), we see that $h_{i_k}^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ is a local approximation of $f_{i_k}(R_{i_k}(\mathbf{V}))$, which satisfies the following two properties:

$$\begin{aligned}
&h_{i_k}^\beta(\widehat{\mathbf{V}}; \widehat{\mathbf{V}}) = f_{i_k}(R_{i_k}(\widehat{\mathbf{V}})) \\
&h_{i_k}^\beta(\mathbf{V}; \widehat{\mathbf{V}}) \leq f_{i_k}(R_{i_k}(\mathbf{V})),\ \forall\ \widehat{\mathbf{V}}, \mathbf{V} \in \mathcal{V}. \qquad (12)
\end{aligned}$$

Let us define $h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) := \sum_{i_k \in \mathcal{I}} h_{i_k}^\beta(\mathbf{V}; \widehat{\mathbf{V}})$. Then we readily have

$$\begin{aligned}
&h^\beta(\widehat{\mathbf{V}}; \widehat{\mathbf{V}}) = f(\widehat{\mathbf{V}}), \qquad h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) \leq f(\mathbf{V}),\ \forall\ \widehat{\mathbf{V}}, \mathbf{V} \in \mathcal{V}, \\
&h^\beta(\widehat{\mathbf{V}}; \widehat{\mathbf{V}}) - s(\mathbf{V}) = u(\widehat{\mathbf{V}}), h^\beta(\mathbf{V}; \widehat{\mathbf{V}}) - s(\widehat{\mathbf{V}}) \leq u(\mathbf{V}),\ \forall\ \widehat{\mathbf{V}}, \mathbf{V} \in \mathcal{V}. \qquad (13)
\end{aligned}$$

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

Another key property of $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ is that it is a *quadratic* function *separable* over all variables $\{\mathbf{V}_{i_k}\}$. To see this, we can expand $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ to obtain (14), where we have defined

$$\widehat{\mathbf{J}}^k := \sum_{j_l \in \mathcal{I}} \widehat{c}_{j_l}(\mathbf{H}_{j_l}^k)^H\widehat{\mathbf{U}}_{j_l}\widehat{\mathbf{E}}_{j_l}^{-1}\widehat{\mathbf{U}}_{j_l}^H\mathbf{H}_{j_l}^k + \beta\mathbf{I}_{MQ_k} \in \mathbb{S}_{++}^{MQ_k}. \qquad (15)$$

Clearly the approximation function $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ consists of $|\mathcal{I}|$ component functions, each of which is a function of a *single* variable $\mathbf{V}_{i_k}$ (user i_k's precoder).

Let us briefly summarize what we have so far. 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 $\beta = 0$, the first-stage approximation expressed in (9) *is well-known and has been used in existing works such as [24] [31]; also see [27, Section VIII-A] for discussion. However, none of those works has articulated the fact that there exists a separable quadratic function of* $\mathbf{V}$ *in the form of* (14) *which serves as a global lower bound for the original system utility; cf.* (13)*. The latter property has been made clear in our derivation precisely due to the use of the two-stage approximation.* □

Remark 2: The main message here is that by performing the two-stage approximation, the resulting $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$, which can be viewed as an alternative representation of the original objective $u(\mathbf{V})$, completely decomposes across the optimization variable $\mathbf{V}$. Such decomposability is intrinsic to the problem and independent of any algorithm to be used. Our method is therefore philosophically different from the general schemes discussed in [19], [20], [25], [26], where parallelization is made possible by appropriate choices of algorithms and their parameters. □

B. The Successive Convex Approximation Algorithm

In this subsection, we develop two algorithms and analyze their properties.

Our approach is to successively approximate $f(\mathbf{V})$ using $h^\beta(\mathbf{V}; \widehat{\mathbf{V}})$ 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 $\mathbf{V}(t-1)$ is a feasible solution to $(\mathrm{P}_{\mathrm{SYSTEM}})$. At iteration t, the following convex optimization

$$h^{\beta}(\mathbf{V};\widehat{\mathbf{V}}) = \sum_{i_k\in\mathcal{I}}\left(\tilde{a}_{i_k} + \mathrm{Tr}\left[2\left(\widehat{c}_{i_k}\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{U}}_{i_k}^H\mathbf{H}_{i_k}^k + \beta\widehat{\mathbf{V}}_{i_k}^H\right)\mathbf{V}_{i_k} - \mathbf{V}_{i_k}^H\widehat{\mathbf{J}}^k\mathbf{V}_{i_k}\right]\right) := \sum_{i_k\in\mathcal{I}} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}). \tag{14}$$

problem is solved to obtain $\mathbf{V}(t)$

$$\begin{aligned} \max_{\mathbf{V}} \quad & h^{\beta}(\mathbf{V};\mathbf{V}(t-1)) - s(\mathbf{V}) \qquad (\mathrm{P}_{\mathrm{Lower-Bound}}) \\ \text{s.t.} \quad & \mathbf{V}^{q_k}\in\mathcal{V}^{q_k}, \forall\, q_k\in\mathcal{Q}, \quad \mathbf{V}^k\in\mathcal{V}^k, \forall\, k\in\mathcal{K}. \end{aligned}$$

Step 2: For each user $i_k\in\mathcal{I}$, compute

$$\mathbf{C}_{i_k}(t) = \sum_{(\ell,j)}\mathbf{H}_{i_k}^{\ell}\mathbf{V}_{j_\ell}(t)\mathbf{V}_{j_\ell}^H(t)(\mathbf{H}_{i_k}^{\ell})^H + \sigma^2\mathbf{I}_N \tag{17a}$$

$$\mathbf{U}_{i_k}(t) = \mathbf{C}_{i_k}^{-1}(t)\mathbf{H}_{i_k}^k\mathbf{V}_{i_k}(t) \tag{17b}$$

$$\mathbf{E}_{i_k}(t) = \mathbf{I}_d - \mathbf{V}_{i_k}^H(t)(\mathbf{H}_{i_k}^k)^H\mathbf{C}_{i_k}^{-1}(t)\mathbf{H}_{i_k}^k\mathbf{V}_{i_k}(t) \tag{17c}$$

$$c_{i_k}(t) = \frac{\partial f_{i_k}(x)}{\partial x}\Big|_{x=-\log|\mathbf{E}_{i_k}(t)|} \tag{17d}$$

$$\mathbf{J}^k(t) = \sum_{j_l\in\mathcal{I}} c_{j_l}(t)(\mathbf{H}_{j_l}^k)^H\mathbf{U}_{j_l}(t)\mathbf{E}_{j_l}^{-1}(t)\mathbf{U}_{j_l}^H(t)\mathbf{H}_{j_l}^k + \beta\mathbf{I}_{MQ_k}. \tag{17e}$$

Step 3: Form the updated function $h^{\beta}(\mathbf{V};\mathbf{V}(t))$ according to (14). Let $t=t+1$, go to **Step 1**.

The key step in the SCA is to solve the lower-bound maximization subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$) to global optimality. In practice, solving ($\mathrm{P}_{\mathrm{Lower-Bound}}$) can be computationally expensive. Therefore it is desirable to inexactly optimize the subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$). The following alternative, termed as *inexact successive convex approximation* (In-SCA) algorithm, replaces Step 1 of the SCA by the following step:

Step 1 of In-SCA: Suppose $\mathbf{V}(t-1)$ is feasible to ($\mathrm{P}_{\mathrm{SYSTEM}}$). At iteration t, approximately solve the subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$) to obtain $\mathbf{V}(t)$, which satisfies conditions (18).

Here $\eta(t) > \eta > 0$ is some iteration dependent constant bounded away from zero. The first condition (18a) requires that the new precoder should *sufficiently* improve the lower bound, while the second condition (18b) restricts the way in which the iterates should be generated by the algorithm. More specifically (18b) 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 satisfying the conditions (18).

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 v is defined by
$$f_v'(x) := \liminf_{r\downarrow 0}\frac{f(x+rv)-f(x)}{r}.$$
- **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 $f(\bullet)$ if $f_v'(x)\le 0$ for all v such that $x+v\in\mathcal{V}$.

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

Theorem 1: Suppose the following conditions hold:

1) *Assumptions A-1)– A-4) are all satisfied;*
2) *The sets $\{\mathcal{V}^k\}_{k\in\mathcal{K}}$ and $\{\mathcal{V}^{q_k}\}_{q_k\in\mathcal{Q}}$ are all convex, closed and compact.*

Then the precoders $\{\mathbf{V}(t)\}$ generated by either the SCA or the In-SCA globally converge to the set of stationary solutions of problem ($\mathrm{P}_{\mathrm{SYSTEM}}$).

It is useful to compare the SCA and In-SCA with various existing approaches for parallel optimization. Essentially, SCA falls in the BSUM framework [27], [28], but with a *single* block variable $\mathbf{V}$. To see this, recall that $h^{\beta}(\mathbf{V};\widehat{\mathbf{V}}) - s(\mathbf{V})$ is a globally tight lower bound of the utility function $u(\mathbf{V})$ (cf. (13)), therefore it satisfies the condition A1-A4 in [27]. In this sense, the SCA can also be viewed as closely related to the FLEXA (Flexible Parallel Algorithm) and the PSCA (Parallel Successive Convex Approximation) algorithms in [25], [26], again with a *single* block variable. The difference with PSCA and FLEXA is that the decomposability of our algorithms is an intrinsic feature of the approximation function $h^{\beta}(\mathbf{V};\widehat{\mathbf{V}}) - s(\mathbf{V})$, not a result of algorithm design. Therefore our algorithms are stepsize-free, while both FLEXA and PSCA requires careful stepsize tuning.

Here we note that the In-SCA algorithm possesses the same convergence guarantee (to the set of stationary solutions) as the SCA algorithm, despite the fact that its subproblems are solved inexactly. Also note that in Step 1 of both algorithms there are K independent subproblems, one for each cell k. This is a result of the decomposability of the approximation function $h^{\beta}(\bullet)$. Clearly, efficiently solving these per-cell problems is the key to the low-complexity implementation of our algorithm. In the next section we shall develop customized algorithms tailored for the per-cell problems of different applications.

IV. Customized Algorithms for Precoder Design

In this section, we customize the SCA and the In-SCA algorithms 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. As each cell consists of a single BS, $\mathbf{V}_{i_k}$ is used to denote the transmit precoder intended for user i_k. In this case, $s(\mathbf{V})\equiv 0$, and the sets $\mathcal{V}^{q_k}$ and $\mathcal{V}^k$ collapse to a single set, cf. (5). The subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$) is given by

$$\begin{aligned} \max_{\{\mathbf{V}^k\}_{k\in\mathcal{K}}} \quad & \sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}), \qquad (\mathrm{P}_{\mathrm{IBC}}) \\ \text{s.t.} \quad & \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}(\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H) \le \bar{P}_k,\ k\in\mathcal{K}. \end{aligned}$$

Clearly both the constraints and the objective of the above problem are separable among the BSs. Therefore we can

$$\big(h^{\beta}(\mathbf{V}(t);\mathbf{V}(t-1))-s(\mathbf{V}(t))\big)-\big(h^{\beta}(\mathbf{V}(t-1);\mathbf{V}(t-1))-s(\mathbf{V}(t-1))\big)\geq \eta(t)\|\mathbf{V}(t-1)-\mathbf{V}(t)\|_F^2 \tag{18a}$$

$$\text{if}\quad \mathbf{V}(t)=\mathbf{V}(t-1),\quad \text{then}\quad \mathbf{V}(t-1)=\arg\max_{\mathbf{V}}\quad h^{\beta}(\mathbf{V};\mathbf{V}(t-1))-s(\mathbf{V}). \tag{18b}$$

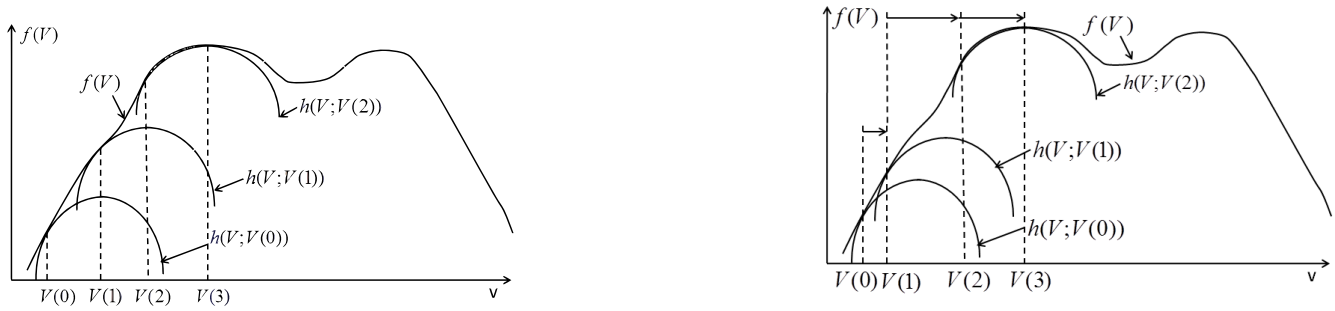


Fig. 1. A graphical illustration of the proposed algorithms. Left: The SCA Algorithm. Right: The In-SCA Algorithm.

decompose this problem into K independent subproblems of the form

$$\begin{aligned}\max_{\mathbf{V}^k}\quad & \sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k},\widehat{\mathbf{V}}), \qquad (\mathrm{P}_{\mathrm{IBC-SUB}})\\ \text{s.t.}\quad & \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right]\leq \bar{P}_k.\end{aligned}$$

Let $\lambda^k \geq 0$ denote the Lagrangian multiplier associated with the power constraint. Then the Lagrangian function for problem ($\mathrm{P}_{\mathrm{IBC-SUB}}$) is given by (19).

Clearly the Slater Condition holds, so the optimal primal-dual pair $((\mathbf{V}^k)^*,(\lambda^k)^*)$ satisfy the KKT optimality conditions

$$\mathbf{V}^{k^*}=\arg\max_{\mathbf{V}^k} L_k(\mathbf{V}^k,\lambda^{k*};\widehat{\mathbf{V}}),\quad \lambda^{k*}\geq 0$$

$$\bar{P}_k-\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}[\mathbf{V}_{i_k}^*\mathbf{V}_{i_k}^{*H}]\geq 0, \tag{20a}$$

$$\lambda^{k*}\left(\bar{P}_k-\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}[\mathbf{V}_{i_k}^*\mathbf{V}_{i_k}^{*H})]\right)=0. \tag{20b}$$

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

$$\mathbf{V}_{i_k}^*(\lambda^k)=\left(\widehat{\mathbf{J}}^k+\lambda^k\mathbf{I}_M\right)^{-1}\left(\widehat{c}_{i_k}(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{U}}_{i_k}\widehat{\mathbf{E}}_{i_k}^{-1}+\beta\widehat{\mathbf{V}}_{i_k}\right),\ \forall\ i_k. \tag{21}$$

To find the optimal multiplier $(\lambda^k)^*$ that satisfies the complementarity condition (20b), we utilize a result on penalty method for optimization, e.g., [41, Section 12.1, Lemma 1]. This result asserts that for the solution $\{\mathbf{V}_{i_k}^*(\lambda^k)\}_{i_k\in\mathcal{I}_k}$, the penalized term $\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}[\mathbf{V}_{i_k}^*(\lambda^k)(\mathbf{V}_{i_k}^*(\lambda^k))^H]$ must be monotonically decreasing with respect to λ^k. It follows that we can find the optimal multiplier by a simple bisection search procedure; see e.g., [3, Section 3.3.1] for details on such procedure.

The algorithm discussed in this subsection is summarized in Table II. We can verify that the overall per-iteration computational complexity is given by (assuming $d=N$)

$$\mathcal{O}(I^2NM^2+IM^3+IM^2N+IN^2M+IN^3).$$

S1): **Initialization** Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all i_k
S2): For each k, compute $\mathbf{C}_{i_k}(t),\mathbf{E}_{i_k}(t),\mathbf{U}_{i_k}(t),c_{i_k}(t)$, according to (17a)–(17d), for all $i_k\in\mathcal{I}_k$
S3): For each k, compute $\mathbf{J}^k(t)$ by (17e); For each k, and $i_k\in\mathcal{I}_k$, compute $\mathbf{V}^k(t)$ by $\mathbf{V}_{i_k}(t)=\big(\mathbf{J}^k(t)+(\lambda^k)^*\mathbf{I}_M\big)^{-1}\times\big(c_{i_k}(t)(\mathbf{H}_{i_k}^k)^H\mathbf{U}_{i_k}(t)\mathbf{E}_{i_k}^{-1}(t)+\beta\mathbf{V}_{i_k}(t-1)\big)$. where $(\lambda^k)^*$ is computed by a bisection procedure
S4) Until some stopping criterion is met

TABLE II
SCA FOR ($\mathrm{P}_{\mathrm{SYSTEM}}$) IN IBC SETTING

Remark 3: When we set $\beta=0$, the algorithm listed in Table II recovers the WMMSE algorithm proposed in [24]. It is interesting to understand the difference between the WMMSE approach and the SCA approach studied here. Reference [24] starts with a SUM problem with the objective $u(\mathbf{V})$. Then it uses a "dimension lifting" approach that adds two extra variables $\mathbf{W}$ and $\mathbf{U}$, and considers a different problem $\min\ \bar{u}(\mathbf{U},\mathbf{W},\mathbf{V})$. The authors show that these two problems are equivalent, so they apply a block coordinate descent (BCD) algorithm to solve the reformulated three-block problem. However, it is by no means clear what is the intuition behind adding these two variables. Our SCA framework well explains this question: $\mathbf{W}$ is in fact the coefficients derived from the first-stage linear approximation, while $\mathbf{U}$ is the coefficients derived from the second-stage convex approximation. Note that in this paper we arrive at the WMMSE by specializing the SCA algorithm to the IBC setting. Thus the SCA algorithm is more general and covers the WMMSE as a special case. □

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

Consider an IBC model in which each BS employs a ZF precoder to cancel the intra-cell interference. We assume that certain user selection within each cell has already taken place, so that the following zero-forcing constraint is always feasible

$$\mathbf{H}_{j_k}^k\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H(\mathbf{H}_{j_k}^k)^H=\mathbf{0},\forall\ j_k\neq i_k,\ j_k\in\mathcal{I}_k. \tag{22}$$

Note that in this network setting, the inter-cell interference is still present, despite the fact that the intra-cell interference

$$L_k(\mathbf{V}^k,\lambda^k;\widehat{\mathbf{V}}) = \sum_{i_k\in\mathcal{I}_k}\left(\mathrm{Tr}\left[2(\widehat{c}_{i_k}\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{U}}_{i_k}^H\mathbf{H}_{i_k}^k+\beta\widehat{\mathbf{V}}_{i_k}^H)\mathbf{V}_{i_k}-\mathbf{V}_{i_k}^H\widehat{\mathbf{J}}^k\mathbf{V}_{i_k}\right]\right)-\lambda^k\left(\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right]-\bar{P}_k\right). \quad (19)$$

is canceled by the use of ZF precoder. Therefore the original problem ($\mathrm{P}_{\mathrm{SYSTEM}}$) is still difficult to solve. To apply the SCA algorithm, we first specialize the subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$) to the following

$$\max_{\mathbf{V}} \quad \sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}) \quad (\mathrm{P}_{\mathrm{IBC-ZF1}})$$

$$\text{s.t.} \quad \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left[\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right]\le\bar{P}_k,\ k\in\mathcal{K} \quad (23a)$$

$$\mathbf{H}_{j_k}^k\mathbf{V}_{i_k}\mathbf{V}_{i_k}^H(\mathbf{H}_{j_k}^k)^H=\mathbf{0},\forall\ j_k\neq i_k,\ j_k\in\mathcal{I}_k. \quad (23b)$$

To remove the ZF constraints (23b), let $L:=N(|\mathcal{I}_k|-1)$, and define new channel matrices as:

$$\mathbf{Q}_{i_k}:=\{\mathbf{H}_{j_k}^k\}_{j_k\in\mathcal{I}_k\setminus\{i_k\}}\in\mathbb{C}^{L\times M},\ \forall\ i_k\in\mathcal{I}_k. \quad (24)$$

Let $\mathbf{Q}_{i_k}=\mathbf{L}_{i_k}\mathbf{\Sigma}_{i_k}\mathbf{R}_{i_k}^H$ denote the singular value decomposition of $\mathbf{Q}_{i_k}$, where $\mathbf{L}_{i_k}\in\mathbb{C}^{L\times L}$ and $\mathbf{R}_{i_k}\in\mathbb{C}^{M\times L}$ are two unitary matrices, and $\mathbf{\Sigma}_{i_k}\in\mathbb{R}^{L\times L}$ is a nonnegative diagonal matrix. Define $\mathbf{P}_{i_k}:=(\mathbf{I}-\mathbf{R}_{i_k}\mathbf{R}_{i_k}^H)$ as a projection matrix to the space orthogonal to the one spanned by $\mathbf{R}_{i_k}$. Let $\mathbf{P}_{i_k}=\widetilde{\mathbf{R}}_{i_k}\widetilde{\mathbf{R}}_{i_k}^H$, where $\widetilde{\mathbf{R}}_{i_k}\in\mathbb{C}^{M\times(M-L)}$ is composed of the orthogonal basis that satisfies $\mathbf{R}_{i_k}^H\widetilde{\mathbf{R}}_{i_k}=\mathbf{0}$ and $\widetilde{\mathbf{R}}_{i_k}^H\widetilde{\mathbf{R}}_{i_k}=\mathbf{I}$. Then [39, Lemma 3.1] asserts that the optimal solution of problem ($\mathrm{P}_{\mathrm{IBC-ZF1}}$) must be of the form: $\mathbf{V}_{i_k}=\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k}$, with $\mathbf{W}_{i_k}\in\mathbb{C}^{(M-L)\times d}$. The structure of the optimal solution implies problem ($\mathrm{P}_{\mathrm{IBC-ZF1}}$) can be equivalently written as

$$\max_{\mathbf{W}} \quad \sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k}\widetilde{g}_{i_k}^{\beta}(\mathbf{W}_{i_k};\widehat{\mathbf{V}}) \quad (\mathrm{P}_{\mathrm{IBC-ZF2}})$$

$$\text{s.t.} \quad \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left[\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k}(\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k})^H\right]\le\bar{P}_k,\ k\in\mathcal{K},$$

where the function $\widetilde{g}_{i_k}^{\beta}(\bullet)$ is the same as the original objective $g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}})$, except that $\mathbf{V}_{i_k}$ is replaced by $\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k}$. Once again both the constraints and the objective are separable among the BSs, so the problem further decomposes into K independent subproblems:

$$\max_{\mathbf{W}^k} \quad \sum_{i_k\in\mathcal{I}_k}\widetilde{g}_{i_k}^{\beta}(\mathbf{W}_{i_k};\widehat{\mathbf{V}}) \quad (\mathrm{P}_{\mathrm{IBC-ZF-SUB}})$$

$$\text{s.t.} \quad \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left[\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k}(\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k})^H\right]\le\bar{P}_k.$$

Let us use λ_k to denote the Lagrangian multiplier associated with the power constraint. Then by using similar steps leading to (21), we can show that the optimal solution for problem ($\mathrm{P}_{\mathrm{IBC-ZF-SUB}}$) is of the form

$$\mathbf{W}_{i_k}^*=\left(\sum_{(\ell,j)}\widehat{c}_{j_\ell}(\mathbf{H}_{j_\ell}^k\widetilde{\mathbf{R}}_{i_k})^H\widehat{\mathbf{U}}_{\ell_j}\widehat{\mathbf{E}}_{j_\ell}^{-1}\widehat{\mathbf{U}}_{\ell_j}^H\mathbf{H}_{j_\ell}^k\widetilde{\mathbf{R}}_{i_k}+\lambda_k^*\mathbf{I}_{M-L}\right)^{-1}\times\widetilde{\mathbf{R}}_{i_k}^H\left(\widehat{c}_{i_k}(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{U}}_{i_k}\widehat{\mathbf{E}}_{i_k}^{-1}+\beta\widehat{\mathbf{V}}_{i_k}\right), \quad (27)$$

where λ_k^* can be computed by the bisection method. The algorithm is summarized in Table III.

S1): **Initialization** For each k and i_k, compute: $\mathbf{Q}_{i_k}=\mathbf{L}_{i_k}\mathbf{\Sigma}_{i_k}\mathbf{R}_{i_k}^H$ $\mathbf{P}_{i_k}=(\mathbf{I}-\mathbf{R}_{i_k}\mathbf{R}_{i_k}^H)$, $\mathbf{P}_{i_k}=\widetilde{\mathbf{R}}_{i_k}\widetilde{\mathbf{R}}_{i_k}^H$; Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all i_k;
S2): For each BS k, compute $\mathbf{C}_{i_k}(t),\mathbf{E}_{i_k}(t),\mathbf{U}_{i_k}(t),c_{i_k}(t)$, according to (17a)–(17d), for all i_k;
S3): Compute $\mathbf{W}_{i_k}(t)$ according to (27). Let $\mathbf{V}_{i_k}(t)=\widetilde{\mathbf{R}}_{i_k}\mathbf{W}_{i_k}(t)$;
S4) Until some stopping criterion is met.

TABLE III
SCA FOR ($\mathrm{P}_{\mathrm{SYSTEM}}$) IN IBC SETTING WITH INTRA-CELL ZF AND INTER-CELL CB

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 $\mathcal{Q}_k$ BSs in each cell, and they form a single *virtual* BS to transmit to the users. In this case, $\mathcal{V}^{q_k}$ is given by (6), i.e., each BS q_k has its own power constraint. 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 $s(\mathbf{V})$. Then the subproblem ($\mathrm{P}_{\mathrm{Lower-Bound}}$) again decomposes into K independent subproblems:

$$\max_{\mathbf{V}^k} \quad \sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k},\widehat{\mathbf{V}}) \quad (\mathrm{P}_{\mathrm{VIBC-SUB}})$$

$$\text{s.t.} \quad \sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}[\mathbf{V}_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k})^H]\le\bar{P}^{q_k},\ \forall\ q_k\in\mathcal{Q}_k.$$

Differently from problem ($\mathrm{P}_{\mathrm{IBC-SUB}}$), the above problem has $Q_k=|\mathcal{Q}_k|$ separable constraints (each constraining a subset of variables), hence Q_k Lagrangian multipliers $\{\lambda_k^{q_k}\}_{q_k\in\mathcal{Q}_k}$. Therefore the bisection algorithm on a single multiplier no longer works.

Fortunately, the constraints for this problem are separable among different block variables $\{\mathbf{V}^{q_k}\}_{q_k\in\mathcal{Q}_k}$. This leads to a block coordinate descent (BCD) algorithm (see [43], [44]), in which one block variable $\mathbf{V}^{q_k}$ is updated at a time while holding the remaining block variables fixed. To capitalize the block structure of the problem, the following definitions are helpful. Let

$$\widehat{\mathbf{S}}_{i_k}:=\widehat{c}_{i_k}(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{U}}_{i_k}\widehat{\mathbf{E}}_{i_k}^{-1}+\beta\widehat{\mathbf{V}}_{i_k}\in\mathbb{C}^{MQ_k\times d_k},\ \forall\ i_k\in\mathcal{I}_k. \quad (28)$$

Partition $\widehat{\mathbf{J}}^k$ (cf. (15)) and $\widehat{\mathbf{S}}_{i_k}$ into the following form

$$\widehat{\mathbf{J}}^k=\begin{bmatrix}\widehat{\mathbf{J}}^k[1,1], & \cdots, & \widehat{\mathbf{J}}^k[1,Q_k]\\ \vdots & \ddots & \vdots\\ \widehat{\mathbf{J}}^k[Q_k,1] & \cdots & \widehat{\mathbf{J}}^k[Q_k,Q_k]\end{bmatrix},$$

$$\widehat{\mathbf{S}}_{i_k}=\left[\widehat{\mathbf{S}}_{i_k}^H[1],\cdots,\widehat{\mathbf{S}}_{i_k}^H[Q_k]\right]^H \quad (29)$$

where $\widehat{\mathbf{J}}^k[q,p]\in\mathbb{C}^{M\times M},\ \forall\ (q,p)\in\mathcal{Q}_k\times\mathcal{Q}_k$, and $\widehat{\mathbf{S}}_{i_k}[q]\in\mathbb{C}^{M\times d},\forall\ q\in\mathcal{Q}_k$. Then the function $g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}})$ defined in

(14) can be alternatively expressed as

$$g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}) = \widetilde{a}_{i_k} + \sum_{p_k\in\mathcal{Q}_k} 2\mathrm{Tr}\left[\widehat{\mathbf{S}}_{i_k}^H[p_k]\mathbf{V}_{i_k}^{p_k}\right] - \sum_{p_k,q_k\in\mathcal{Q}_k} \mathrm{Tr}\left[(\mathbf{V}_{i_k}^{q_k})^H\widehat{\mathbf{J}}^k[q_k,p_k]\mathbf{V}_{i_k}^{p_k}\right]. \quad (30)$$

Now it becomes clear that the objective function of problem ($\mathrm{P}_{\mathrm{VIBC-SUB}}$), $\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}})$, is a quadratic function with respect to $\mathbf{V}^{m_k}$, the precoder used by BS $m_k\in\mathcal{Q}_k$. It follows that the per-block problem, written in the following form, can be efficiently solved in closed form for each block $m_k\in\mathcal{Q}_k$

$$\max_{\mathbf{V}^{m_k}} \sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}) \qquad (\mathrm{P}_{\mathrm{VIBC-BLK}})$$
$$\text{s.t.} \quad \sum_{i_k\in\mathcal{I}_k} \mathrm{Tr}[\mathbf{V}_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k})^H] \le \bar{P}^{m_k}.$$

Let $\lambda^{m_k}\ge 0$ 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 $\mathbf{V}^{m_k*}$ for problem ($\mathrm{P}_{\mathrm{VIBC-BLK}}$) can be expressed as

$$\mathbf{V}_{i_k}^{m_k*} = \left(\widehat{\mathbf{J}}^k[m_k,m_k] + \lambda^{m_k*}\mathbf{I}_M\right)^{-1} \times \left(\widehat{\mathbf{S}}_{i_k}[m_k] - \sum_{p_k\neq m_k}\widehat{\mathbf{J}}^k[m_k,p_k]\mathbf{V}_{i_k}^{p_k}\right),\ \forall\ i_k\in\mathcal{I}_k \quad (31)$$

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

The above observation leads to a natural two-layer algorithm: *i)* the outer layer updates $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$, $c_{i_k}(t)$, $\mathbf{J}^k(t)$ and $\mathbf{S}^k(t)$; *ii)* the inner layer updates each $\mathbf{V}^k$ by a BCD algorithm with blocks given by $\{\mathbf{V}^{q_k}\}_{q_k\in\mathcal{Q}_k}$. See Table IV for the detailed description. Note that in the table $\beta>0$ is some proximal parameter introduced to regularize the iterates. Once again the convergence of the algorithm is a direct consequence of Theorem 1.

S1): **Initialization** Obtain a feasible solution $\mathbf{V}_{i_k}(0)$, $\forall\ i_k$
S2): For each BS k, compute $\mathbf{C}_{i_k}(t),\mathbf{E}_{i_k}(t),\mathbf{U}_{i_k}(t),c_{i_k}(t)$, according to (17a)–(17d), for all $i_k\in\mathcal{I}_k$
S3): For each BS k and $i_k\in\mathcal{I}_k$, compute $\mathbf{J}^k(t)$ by (17e);
$\mathbf{S}_{i_k}(t) = c_{i_k}(t)(\mathbf{H}_{i_k}^k)^H\mathbf{U}_{i_k}(t)\mathbf{E}_{i_k}^{-1}(t) + \beta\mathbf{V}_{i_k}(t-1)$.
S4): For each BS k, compute the precoders $\mathbf{V}^k(t)$ by
Repeat Cyclically pick $m_k\in\mathcal{Q}_k$
Compute $\mathbf{V}_{i_k}^{m_k*}$ using (31), $\forall\ i_k\in\mathcal{I}_k$, where λ^{m_k*} is computed by a bisection procedure
Until An optimal solution of ($\mathrm{P}_{\mathrm{VIBC-SUB}}$) is obtained.
Let $\mathbf{V}_{i_k}^{m_k}(t) = \mathbf{V}_{i_k}^{m_k*}$, $\forall\ i_k,m_k$
S5) Until some stopping criteria is met.

TABLE IV
SCA FOR SOLVING ($\mathrm{P}_{\mathrm{SYSTEM}}$) WITH INTRA-CELL COMP AND INTER-CELL CB

The key question here is whether in Step S4) one needs to solve the inner problem ($\mathrm{P}_{\mathrm{VIBC-SUB}}$) *exactly* before we can update the outer layer. There are several drawbacks with this approach:

1) It is usually difficult to check whether the inner iteration has indeed reached the optimality;
2) Before reaching the optimality for the subproblem ($\mathrm{P}_{\mathrm{VIBC-SUB}}$), the marginal benefit of the precoder updates in the inner iteration decreases as the iteration progresses. This effect is manifested at 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.

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 each block be picked at least once; see Table V. It can be verified that the overall per-iteration complexity is given by (assuming $d=N$)

$$\mathcal{O}\left(QI(M^3+M^2N+QM^2N)+I^2NM^2+IN^3+IMN^2\right).$$

S1): **Initialization** Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all i_k
S2): For each BS k, compute $\mathbf{C}_{i_k}(t),\mathbf{E}_{i_k}(t),\mathbf{U}_{i_k}(t),c_{i_k}(t)$, according to (17a)–(17d), for all $i_k\in\mathcal{I}_k$
S3): For each BS k, compute $\mathbf{J}^k(t)$ by (17e);
For each BS k, compute
$\mathbf{S}_{i_k}(t) = c_{i_k}(t)(\mathbf{H}_{i_k}^k)^H\mathbf{U}_{i_k}(t)\mathbf{E}_{i_k}^{-1}(t) + \beta\mathbf{V}_{i_k}(t-1),\ \forall\ i_k\in\mathcal{I}_k$.
S4): For each BS k, compute the precoders $\mathbf{V}^k(t)$ by
Repeat Cyclically pick $m_k\in\mathcal{Q}_k$
Compute $\mathbf{V}_{i_k}^{m_k*}$ using (31), $\forall\ i_k\in\mathcal{I}_k$, where λ^{m_k*} is computed by a bisection procedure
Until Each m_k is picked at least once
Let $\mathbf{V}_{i_k}^{m_k}(t) = \mathbf{V}_{i_k}^{m_k*}$, $\forall\ i_k,m_k$
S5) Until some stopping criteria is met.

TABLE V
IN-SCA FOR SOLVING ($\mathrm{P}_{\mathrm{SYSTEM}}$) WITH INTRA-CELL COMP AND INTER-CELL CB

Next we show the convergence of the above inexact algorithm, by appealing to Theorem 1. The proof is given in Appendix D.

Corollary 1: Consider problem ($\mathrm{P}_{\mathrm{SYSTEM}}$) *specialized to the setting given in this section. Suppose* $\beta>0$, *then the inexact algorithm in Table V converges to the set of stationary solutions.*

Remark 4: The proof of Corollary 1 hinges on the fact that each $\mathbf{V}^{m_k}$ problem is strongly convex, cf. (41). This is precisely the reason that we have introduced the proximal term $-\beta\left\|\mathbf{V}_{i_k}-\widehat{\mathbf{V}}_{i_k}\right\|_F^2$ in the first place. As long as β is bounded away from zero, the lower bound function $\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}})$ is strongly convex with respect to each $\mathbf{V}^{m_k}$, providing the desired sufficient descent property (18a). □

Remark 5: 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 $\mathcal{S}^{q_k}\subseteq\mathcal{I}_k$

denote the set of users served by BS q_k. Then we only need to slightly modify the algorithm in Table IV and V by the following:

- In S1), for each BS $m_k \in \mathcal{Q}_k$, set $\mathbf{V}_{i_k}^{m_k}(0) = \mathbf{0}$ for all $i_k \notin \mathcal{S}^{m_k}$;
- In S4), let each BS $m_k \in \mathcal{Q}_k$ compute $\mathbf{V}_{i_k}^{m_k}(t)$ using (31), $\forall\ i_k \in \mathcal{S}^{m_k}$ (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 at each iteration. Once again, the update at each iteration is closed-form, while in related works such as [34], general purpose convex solvers are required.

Moreover, when the BSs' clustering structure needs to be designed jointly with the precoders, we can include appropriate penalty term $s(\mathbf{V})$ into the objective. Specifically, the per-block subproblem ($\mathrm{P}_{\mathrm{VIBC-BLK}}$) takes the following form

$$\begin{aligned} \max_{\mathbf{V}^{m_k}} \quad & \sum_{i_k \in \mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k}, \widehat{\mathbf{V}}) - \sum_{i_k \in \mathcal{I}_k} s_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k}) \\ \text{s.t.} \quad & \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}\left[\mathbf{V}_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k})^H\right] \leq \bar{P}^{m_k}. \end{aligned}$$

When we let $s_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k}) = \gamma_{i_k}^{m_k}\|\mathbf{V}_{i_k}^{m_k}\|_F$, this subproblem becomes a well known quadratic group-LASSO problem [45] (with an additional quadratic constraint), which can be solved using a bisection method; see [22], [36] for details. □

Remark 6: For the problem discussed in this section, the In-SCA can take other forms as well. For example one can utilize the BSCA proposed in [27], or the CGD method proposed in [43], [46]. All that is needed is to verify that the conditions (18) are satisfied. However, it appears that the BCD-based In-SCA proposed in Table V takes a much simpler form, and it is much easier to implement and analyze. □

V. Numerical Results

In this section we conduct experiments to validate the effectiveness of the proposed algorithms. We consider 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 multi-cell network of up to 10 hexagonal cells in a square grid. 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 $y_{i_k}^{q_\ell}$ denote the distance between BS q_ℓ and user i_k. The channel coefficients between user i_k and BS q_ℓ are modeled as zero mean circularly symmetric complex Gaussian vector with $\left(200/y_{i_k}^{q_\ell}\right)^3 L_{i_k}^{q_\ell}$ as variance for both real and imaginary dimensions, where $10\log_{10}(L_{i_k}^{q_\ell}) \sim \mathcal{N}(0, 64)$ is a real Gaussian random variable modeling the shadowing effect [1]. We set the noise power $\sigma^2 = 1$, and uniformly randomly generate the power budget $\bar{P}^{q_k} \in (0,\ \mathrm{P}_k^{\mathrm{tot}}/|\mathcal{Q}_k|]$ for all $q_k \in \mathcal{Q}_k$, where the $\mathrm{P}_k^{\mathrm{tot}} > 0$ represents an upper bound of the power budget for cell k.

[1]The choice of the channel parameters follow those provided in [22], [33], [38]. Note that commonly accepted standard deviation of shadowing is between 6dB and 12dB, and our choice is 8dB.

The stopping criteria are chosen as follows. The single time-scale algorithm (i.e. the In-SCA in Table V, or the SCA described in Table II, III) as well as the outer loop of the double time-scale algorithm (i.e., the SCA described in Table IV) stop when $\frac{|u(t+1)-u(t)|}{|u(t)|} \leq 10^{-3}$. 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 ($\mathrm{P}_{\mathrm{VIBC-BLK}}$)) is less than 10^{-3} after performing one round of update by all the BSs in the cell.

A. HetNet and Multicell Downlink Setting

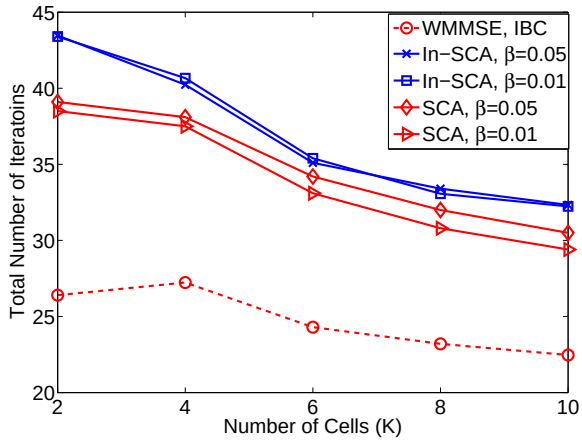
In this section, the performance of the following algorithms is compared:

1) **WMMSE** [24]: The algorithm described in Table II with $\beta = 0$.
2) **SCA for HetNet**: The two time-scale algorithm given in Table IV, with the difference that the inner problem Step S4) is solved using general purpose solvers.
3) **In-SCA for HetNet**: The inexact algorithm described in Table V, where each block is updated *only once* in Step S4).
4) **ZF-SCA for IBC**: The intra-cell ZF plus inter-cell CB algorithm described in Table III;
5) **Per-Cell ZF for HetNet** [39]: Algorithm 2 proposed in [39], which performs intra-cell ZF for the HetNet setting with BS power constraint, while ignoring the inter-cell interference.

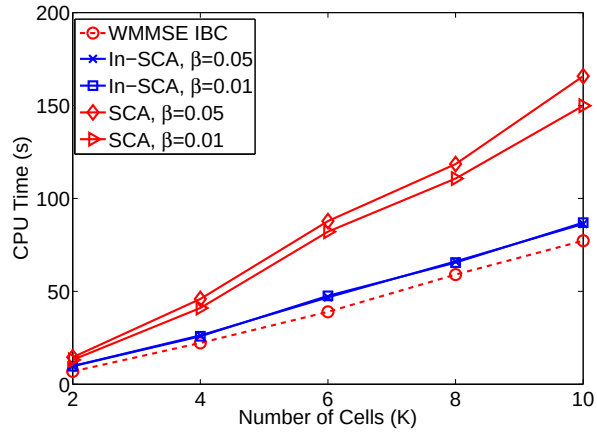
All algorithms considered in this subsection use the sum rate utility. The plots to be shown represent the averaged performance over 100 independent network generations.

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, the averaged CPU time and the total number of iterations used by different algorithms are compared for a network with $|\mathcal{Q}_k| = 6$, $M = 5$, $N = 3$, $|\mathcal{I}_k| = 10$ and $P_k^{\mathrm{tot}} = 20$dB. 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. Third, our experiments suggest that In-SCA is not quite sensitive to the choice of β. It even works well when $\beta = 0$, although this result is not displayed due to its similarity to the cases when $\beta = \{0.01, 0.05\}$.

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 $K = 6$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = \{2, 4, \cdots, 12\}$, $M = \{4, 6\}$, $N = 2$ and $d = 2$, $P_k^{\mathrm{tot}} = 20$dB. Note that when increasing the number of users $|\mathcal{I}_k|$, 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 $|\mathcal{I}_k|$ increases. Further, when $M = 4$ and when



(a) Comparing number of iterations



(b) Comparing CPU time

Fig. 3. Comparison of the CPU Time and the total iteration number required for computation in HetNet setting with different sizes of the network. $K = [2, 4, 6, 8, 10]$, $P_k^{\text{tot}} = 20$dB, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d = 3$ for all $i_k \in \mathcal{I}$.

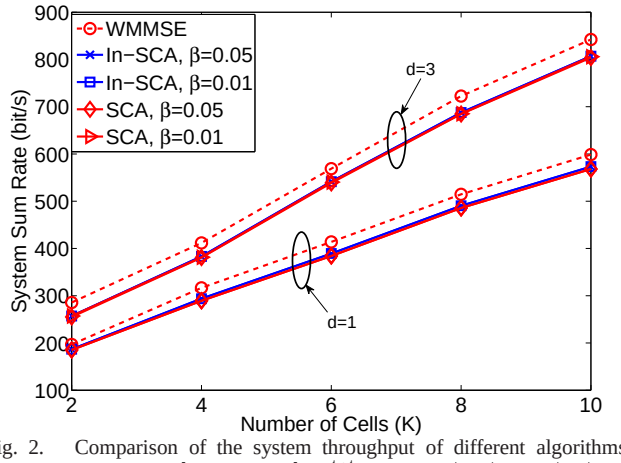


Fig. 2. Comparison of the system throughput of different algorithms in HetNet setting. $K = [2, 4, 6, 8, 10]$, $P_k^{\text{tot}} = 20$dB, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d = 1$ or $d = 3$ for all $i_k \in \mathcal{I}$.

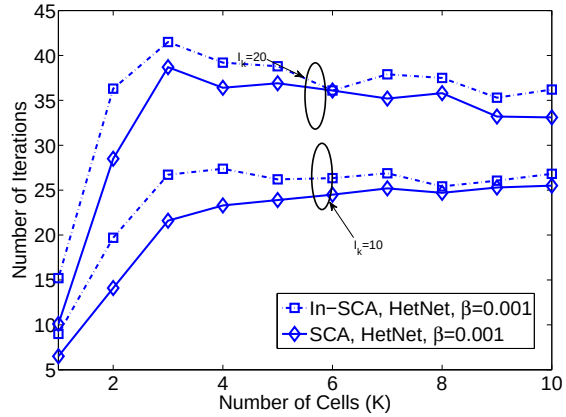


Fig. 6. Comparison of the system performance of different algorithms in HetNet. $K = [1, 2, \cdots, 10]$, $P_k^{\text{tot}} = 20$dB, $|\mathcal{Q}_k| = 10$, $|\mathcal{I}_k| = [10, 20]$, $M = 4$, $N = 2$, $d = 1$ for all $i_k \in \mathcal{I}$. $\lambda = 0.1$ when $|\mathcal{I}_k| = 10$, and $\lambda = 0.05$ when $|\mathcal{I}_k| = 20$.

$|\mathcal{I}_k|$ approaches the maximum number of allowable users for which the ZF strategy is still feasible (12 in this case), the performance degradation is more severe. The reason is that when the nodes densely deployed, inter-cell interference becomes equally detrimental as the intra-cell interference. The general linear precoding does not pre-specify the source of interference to be mitigated, therefore appears to be a better candidate for dealing with interference.

B. Partial CoMP in HetNet

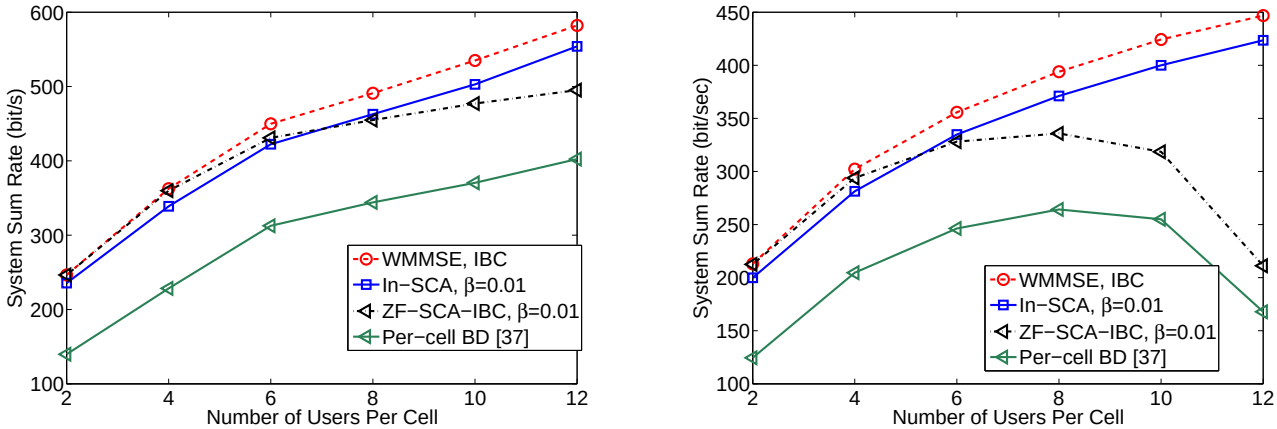
In this section, we jointly design the clustering and linear precoding schemes in a partial CoMP setting. To induce the desired clustering structure, we specialize the penalty terms in the objective to take the following form [22], [36]: $s_{i_k}^{q_k}(\mathbf{V}) = \lambda \|\mathbf{V}_{i_k}^{q_k}\|_F$, $\forall\ i_k, q_k$ 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., $u_{i_k}(R_{i_k}) = \log(1 + R_{i_k})$.

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 [22] (which can be viewed as a special case of the SCA algorithm for solving the penalized utility maximization problem, see Remark 5), modified by adding the proximal term with $\beta = 0.001$; **(3)**: In-SCA algorithm in Table V, with $\beta = 0.001$.

The results are summarized in Fig. 5–6 and Table V-B. 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 WMMSE always mandates all the BSs to transmit to each user. The SCA and In-SCA use similar number of iterations and achieve similar performance. The advantage of using the inexact version is in the efficiency of computation.

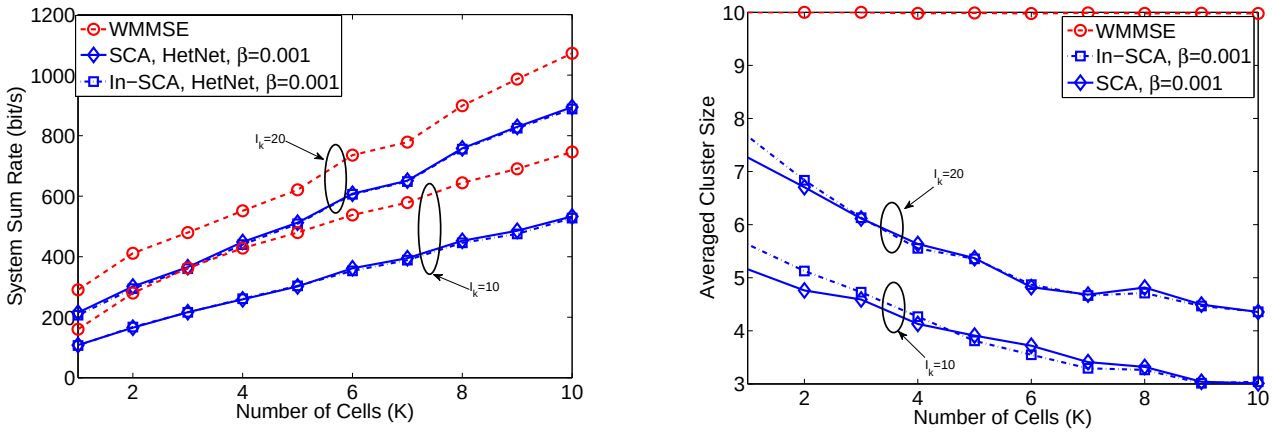
VI. Conclusion

In this work we study an important family of interference management problems arising in the HetNet. The main novelty of this work lies in the proposal to achieve decomposition across the interference-coupled networks by using successive convex approximation. Our proposed approach has low computational complexity, as each of the subproblems to be



(a) $M = 4$ (b) $M = 6$

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



(a) System throughput (b) Averaged cluster size

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

	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 ($I_k = 20$)	105.2	204.4	258.1	359.4	443.7
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 ($I_k = 10$)	13.9	38.2	69.3	140.5	164.5
IN-SCA ($I_k = 10$)	7.4	19.6	40.1	45.3	72.3

TABLE VI
CPU TIME NEEDED FOR DIFFERENT ALGORITHMS (UNIT: SECOND)

solved is convex and decomposes across the cells. Depending on whether the subproblems are solved exactly, two general algorithms have been proposed, each having applications in a few interference management problems. In the future, we plan to extend our framework to other problems for interference management beyond those mentioned in this work.

APPENDIX

A. Proof of Lemma 1

Consider the epigraph of $l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$, i.e., $\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t) \mid l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) \le t\}$. It suffices to show that the epigraph $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is a convex set [47, Chapter 3]. To this end, let us consider the following extended set (with $\mathbf{Z}_{i_k} \succeq \mathbf{0}$ being a slack variable):

$$\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, \mathbf{Z}_{i_k}, t) \mid \mathrm{Tr}[\mathbf{Z}_{i_k}] \le t,\ \mathbf{Z}_{i_k} \succeq \mathbf{0},\ \mathbf{Z}_{i_k} \succeq \widehat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \widehat{\mathbf{E}}_{i_k}^{-1/2}\}. \quad (32)$$

It is not hard to show that $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is just a projection of the set defined by (32). Therefore, if the extended set $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, \mathbf{Z}_{i_k}, t)$ is convex, then $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is also convex. By applying Schur's complement, we have

$$\mathbf{Z}_{i_k} \succeq \widehat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \widehat{\mathbf{E}}_{i_k}^{-1/2} \iff \begin{bmatrix} \mathbf{Z}_{i_k} & \widehat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \\ \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \widehat{\mathbf{E}}_{i_k}^{-1/2} & \mathbf{C}_{i_k} \end{bmatrix} \succeq \mathbf{0}. \quad (33)$$

This shows that (32) is a convex set whenever $\mathbf{C}_{i_k} \succeq 0$.

B. Derivation of the equality in (10)

Let us investigate the directional derivative of the function l_{i_k}. We can first obtain (34).

Let $\mathbf{V}_{i_k} = \widehat{\mathbf{V}}_{i_k}$, $\mathbf{d} = \left(\mathbf{V}_{i_k} - \widehat{\mathbf{V}}_{i_k}\right)$, $\mathbf{C}_{i_k} = \widehat{\mathbf{C}}_{i_k}$ and $\mathbf{M}_{i_k} =$

$$\frac{dl_{i_k}(\mathbf{V}_{i_k}+t\mathbf{d},\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})}{dt}=\frac{d\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k}+t\mathbf{d})^H(\mathbf{H}_{i_k}^k)^H(\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k}+t\mathbf{d})\right]}{dt}$$
$$=\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{d})^H(\mathbf{H}_{i_k}^k)^H(\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k}+t\mathbf{d})\right]+\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k}+t\mathbf{d})^H(\mathbf{H}_{i_k}^k)^H(\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})^{-1}\mathbf{H}_{i_k}^k\mathbf{d}\right]$$
$$-\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k}+t\mathbf{d})^H(\mathbf{H}_{i_k}^k)^H(\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})^{-1}\mathbf{M}_{i_k}(\mathbf{C}_{i_k}+t\mathbf{M}_{i_k})^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k}+t\mathbf{d})\right]. \tag{34}$$

$\mathbf{0}$, we obtain the expression

$$\left.\frac{dl_{i_k}(\widehat{\mathbf{V}}_{i_k}+t(\mathbf{V}_{i_k}-\widehat{\mathbf{V}}_{i_k}),\widehat{\mathbf{C}}_{i_k})}{dt}\right|_{t=0}$$
$$=\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}(\mathbf{V}_{i_k}-\widehat{\mathbf{V}}_{i_k})^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right]$$
$$+\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k(\mathbf{V}_{i_k}-\widehat{\mathbf{V}}_{i_k})\right].$$

Similarly, let $\mathbf{V}_{i_k}=\widehat{\mathbf{V}}_{i_k}$, $\mathbf{d}=\mathbf{0}$, $\mathbf{C}_{i_k}=\widehat{\mathbf{C}}_{i_k}$ and $\mathbf{M}_{i_k}=\left(\mathbf{C}_{i_k}-\widehat{\mathbf{C}}_{i_k}\right)$, we obtain the expression

$$\left.\frac{dl_{i_k}(\widehat{\mathbf{V}}_{i_k},\widehat{\mathbf{C}}_{i_k}+t(\mathbf{C}_{i_k}-\widehat{\mathbf{C}}_{i_k}))}{dt}\right|_{t=0}$$
$$=-\mathrm{Tr}\left[\widehat{\mathbf{E}}_{i_k}^{-1}\widehat{\mathbf{V}}_{i_k}^H(\mathbf{H}_{i_k}^k)^H\widehat{\mathbf{C}}_{i_k}^{-1}(\mathbf{C}_{i_k}-\widehat{\mathbf{C}}_{i_k})\widehat{\mathbf{C}}_{i_k}^{-1}\mathbf{H}_{i_k}^k\widehat{\mathbf{V}}_{i_k}\right].$$

This completes the proof.

C. Proof of Theorem 1

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

$$u(\mathbf{V}(t+1))\overset{\text{(i)}}{\geq}h^\beta(\mathbf{V}(t+1);\mathbf{V}(t))-s(\mathbf{V}(t+1))$$
$$\overset{\text{(ii)}}{\geq}h^\beta(\mathbf{V}(t);\mathbf{V}(t))-s(\mathbf{V}(t))\overset{\text{(iii)}}{=}u(\mathbf{V}(t)) \tag{35}$$

where step (i) is from (13); step (ii) is from the fact that $\mathbf{V}(t+1)$ is the solution to the problem ($\mathrm{P}_{\mathrm{Lower-Bound}}$); step (iii) is because of (13). As both $f(\mathbf{V})$ and $s(\mathbf{V})$ are upper bounded for all $\mathbf{V}$ in the feasible set, it follows that the sequence $\{u(\mathbf{V}(t))\}$ converges. Let $\bar{u}$ denote its limit. This result combined with (35) implies that

$$\lim_{t\to\infty}h^\beta(\mathbf{V}(t+1);\mathbf{V}(t))-s(\mathbf{V}(t+1))=\bar{u}. \tag{36}$$

Using a similar argument as in [27, Lemma 1], and use the fact that $f(\bullet)$ and $h(\bullet;\widehat{\mathbf{V}})$ satisfy (13), we can show that for any feasible $\widehat{\mathbf{V}}$ the directional derivative of $f(\bullet)$ at the point $\widehat{\mathbf{V}}$ equals that of $h^\beta(\bullet;\widehat{\mathbf{V}})$ at the point $\widehat{\mathbf{V}}$, i.e.,

$$\lim_{r\to0}\frac{f(\widehat{\mathbf{V}}+r\mathbf{D})-f(\widehat{\mathbf{V}})}{r}=\lim_{r\to0}\frac{h^\beta(\widehat{\mathbf{V}}+r\mathbf{D};\widehat{\mathbf{V}})-h^\beta(\widehat{\mathbf{V}})}{r} \tag{37}$$

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

Let $\{\mathbf{V}(t_m)\}_{m=1}^\infty$ be a converging subsequence of $\mathbf{V}(t)$ with limit $\mathbf{V}^*$. Clearly we have

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

Then by Step 1 of the algorithm, the optimality of $\mathbf{V}(t)$ implies

$$h^\beta(\mathbf{V}(t_m+1);\mathbf{V}(t_m))-s(\mathbf{V}(t_m+1))$$
$$\geq h^\beta(\mathbf{V};\mathbf{V}(t_m))-s(\mathbf{V}),\ \forall\ \mathbf{V}\in\mathcal{V}.$$

Taking limit on both sides and use (36), we obtain

$$\bar{u}\geq h^\beta(\mathbf{V};\mathbf{V}^*)-s(\mathbf{V}),\quad\forall\ \mathbf{V}.$$

It follows that $h^\beta(\mathbf{V}^*;\mathbf{V}^*)-s(\mathbf{V}^*)\geq h^\beta(\mathbf{V};\mathbf{V}^*)-s(\mathbf{V})$ for all feasible $\mathbf{V}$, which further implies that

$$\left(h_{\mathbf{D}}^{'}(\mathbf{V};\mathbf{V}^*)-s_{\mathbf{D}}^{'}(\mathbf{V})\right)\big|_{\mathbf{V}=\mathbf{V}^*}\leq0,\ \forall\ \mathbf{D}+\mathbf{V}^*\in\mathcal{V}.$$

Utilizing (37), we obtain

$$u_{\mathbf{D}}^{'}(\mathbf{V})\big|_{\mathbf{V}=\mathbf{V}^*}=\left(f_{\mathbf{D}}^{'}(\mathbf{V})-s_{\mathbf{D}}^{'}(\mathbf{V})\right)\big|_{\mathbf{V}=\mathbf{V}^*}\leq0,\forall\ \mathbf{D}+\mathbf{V}^*\in\mathcal{V},$$

which says that $\mathbf{V}^*$ is a stationary solution to the problem ($\mathrm{P}_{\mathrm{SYSTEM}}$).

We then prove the result for the IN-SCA algorithm. Similarly as the previous case, we have

$$u(\mathbf{V}(t+1))\geq h^\beta(\mathbf{V}(t+1);\mathbf{V}(t))-s(\mathbf{V}(t+1))$$
$$\overset{\text{(i)}}{\geq}h^\beta(\mathbf{V}(t);\mathbf{V}(t))-s(\mathbf{V}(t))+\eta\|\mathbf{V}(t+1)-\mathbf{V}(t)\|_F^2$$
$$=u(\mathbf{V}(t))+\eta\|\mathbf{V}(t+1)-\mathbf{V}(t)\|_F^2 \tag{38}$$

where (i) is due to condition (18). The above inequality implies that

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

Taking limit on both sides, and by the upper-boundedness of $u(\mathbf{V})$ for all $\mathbf{V}\in\mathcal{V}$, we obtain

$$\lim_{t\to\infty}\|\mathbf{V}(t+1)-\mathbf{V}(t)\|_F^2=0,\quad\lim_{t\to\infty}u(\mathbf{V}(t+1))=\bar{u}. \tag{39}$$

Let $\{\mathbf{V}(t_m)\}_{m=1}^\infty$ be a converging subsequence of $\mathbf{V}(t)$, and denote its limit as $\mathbf{V}^*$. Combining the first equation in (39) and the second condition in (18), we have

$$\mathbf{V}^*=\arg\min_{\mathbf{V}}h^\beta(\mathbf{V};\mathbf{V}^*)-s(\mathbf{V})$$
$$\Longleftrightarrow h^\beta(\mathbf{V}^*;\mathbf{V}^*)-s(\mathbf{V}^*)\geq h^\beta(\mathbf{V};\mathbf{V}^*)-s(\mathbf{V}),\ \forall\ \mathbf{V}\in\mathcal{V}.$$

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

D. Proof of Corollary 1

We focus on the case that each coordinate $m_k\in\mathcal{Q}_k$ is updated precisely once in Step 3. The case in which each coordinate is updated *at least* once is a straightforward extension.

We only need to verify the two conditions (18). Define the following short-handed notation

$$\mathbf{V}_{i_k}^{-m_k}:=\{\mathbf{V}_{i_k}^{p_k}\}_{p_k\neq m_k},\ \mathbf{V}^{-m_k}:=\{\mathbf{V}^{p_k}\}_{p_k\neq m_k}.$$

Also define $\nabla_{m_k}g_{i_k}^\beta(\mathbf{V}_{i_k}^{m_k},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}})$ as the gradient of $g_{i_k}^\beta(\bullet)$ with respect to the block $\mathbf{V}_{i_k}^{m_k}$.

At any given $\widehat{\mathbf{V}}$, we fix $\mathbf{V}^{-m_k}$ and update $\mathbf{V}^{m_k}$ by using (31). First observe that the optimality condition for the per-

block problem ($\mathrm{P}_{\mathrm{VIBC-BLK}}$) is given by

$$\sum_{i_k\in\mathcal{I}_k}\left\langle \nabla_{m_k} g_{i_k}^{\beta}\left(\mathbf{V}_{i_k}^{m_k*},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right),\mathbf{V}_{i_k}^{m*}-\widetilde{\mathbf{V}}_{i_k}^{m_k}\right\rangle\geq 0,$$
$$\forall\ \widetilde{\mathbf{V}}^{m_k}\in\mathcal{V}^{m_k}. \tag{40}$$

Utilizing this property, we have

$$\begin{aligned}&\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}\left(\mathbf{V}_{i_k}^{m_k*},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right)-g_{i_k}^{\beta}\left(\widehat{\mathbf{V}}_{i_k}^{m_k},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right) \qquad (41)\\ &\overset{\text{(i)}}{\geq}\sum_{i_k\in\mathcal{I}_k}\left\langle \nabla_{m_k} g_{i_k}^{\beta}\left(\mathbf{V}_{i_k}^{m*},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right),\mathbf{V}_{i_k}^{m*}-\widehat{\mathbf{V}}_{i_k}^{m_k}\right\rangle\\ &\quad+\sum_{i_k\in\mathcal{I}_k}\frac{\beta}{2}\|\mathbf{V}_{i_k}^{m_k*}-\widehat{\mathbf{V}}_{i_k}^{m_k}\|_F^2\\ &\overset{\text{(ii)}}{\geq}\sum_{i_k\in\mathcal{I}_k}\frac{\beta}{2}\|\mathbf{V}_{i_k}^{m_k*}-\widehat{\mathbf{V}}_{i_k}^{m_k}\|_F^2\end{aligned}$$

where (i) is from the strong concavity of $g_{i_k}^{\beta}\left(\mathbf{V}_{i_k}^{m_k},\mathbf{V}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right)$ with respect to $\mathbf{V}_{i_k}^{m_k}$ (with modulus β); (ii) comes from the optimality condition (40). After one round of the update where each m_k is updated once, we have

$$\begin{aligned}&h^{\beta}(\mathbf{V}^*;\widehat{\mathbf{V}})-h^{\beta}(\widehat{\mathbf{V}};\widehat{\mathbf{V}})\\ &\geq\sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k}\sum_{m_k\in\mathcal{Q}_k}\frac{\beta}{2}\|\mathbf{V}_{i_k}^{m_k*}-\widehat{\mathbf{V}}_{i_k}^{m_k}\|_F^2\geq\frac{\beta}{2}\|\mathbf{V}^*-\widehat{\mathbf{V}}\|_F^2\end{aligned}$$

which proves the first condition in (18).

To show the second condition in (18), we fix $\mathbf{V}^{-m_k}=\widehat{\mathbf{V}}^{-m_k}$ and consider $\mathbf{V}^{m_k*}$ obtained by using (31). Clearly if for each $m_k\in\mathcal{Q}_k$, $\widehat{\mathbf{V}}^{m_k}=\mathbf{V}^{m_k*}$ is true, then we have

$$\widehat{\mathbf{V}}^{m_k}=\arg\max_{\mathbf{V}^{m_k}}\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k}^{m_k},\widehat{\mathbf{V}}_{i_k}^{-m_k};\widehat{\mathbf{V}}),\ \forall\ m_k\in\mathcal{Q}_k.$$

The optimality conditions of these $|\mathcal{Q}_k|$ problems are given below

$$\sum_{i_k\in\mathcal{I}_k}\left\langle \nabla_{m_k} g_{i_k}^{\beta}\left(\widehat{\mathbf{V}}_{i_k}^{m_k},\widehat{\mathbf{V}}_{i_k}^{-m_k};\widehat{\mathbf{V}}\right),\widehat{\mathbf{V}}_{i_k}^{m_k}-\widetilde{\mathbf{V}}_{i_k}^{m_k}\right\rangle\geq 0,$$
$$\forall\ \widetilde{\mathbf{V}}^{m_k}\in\mathcal{V}^{m_k},\ \forall\ m_k\in\mathcal{Q}_k. \tag{42}$$

which collectively imply the optimality condition of the following problem

$$\widehat{\mathbf{V}}^{k}=\arg\max_{\mathbf{V}^{k}}\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}}).$$

Enumerating over all cells $k\in\mathcal{K}$, we obtain

$$\widehat{\mathbf{V}}=\arg\max_{\mathbf{V}}\sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k} g_{i_k}^{\beta}(\mathbf{V}_{i_k};\widehat{\mathbf{V}})=\arg\max_{\mathbf{V}}\ h^{\beta}(\mathbf{V};\widehat{\mathbf{V}}).$$

which satisfies (18b). The proof is completed.

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] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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, pp. 3711–3724, 2009.

[19] 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.

[20] 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.

[21] 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.

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

[23] D. P. Bertsekas, *Nonlinear Programming, 2nd ed*, Athena Scientific, Belmont, MA, 1999.

[24] 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.

[25] M. Razaviyayn, M. Hong, Z.-Q. Luo, and J. S. Pang, "Parallel successive convex approximation for nonsmooth nonconvex optimization," in *the Proceedings of the Neural Information Processing (NIPS)*, 2014.

[26] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, 2015.

[27] 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.
[28] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data," 2015, accepted by IEEE Signal Processing Magazine.
[29] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.
[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory, second edition*, Wiley, 2005.
[31] 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.
[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] 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.
[34] 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.
[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] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access, Special issue on Cloud Radio-Access Networks*, 2014, to appear.
[38] W.-C. Liao, M. Hong, Y.-F. Liu, and Z.-Q. Luo, "Base station activation and linear transceiver design for optimal resource management in heterogeneous networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3939–3952, 2014.
[39] 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.
[40] 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.
[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] P. Tseng and S. Yun, "A coordinate gradient descent method for nonsmooth separable minimization," *Mathematical Programming*, vol. 117, pp. 387–423, 2009.
[47] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.