Self-Organizing Algorithms for Interference Coordination in Small Cell Networks

Furqan Ahmed, *Member, IEEE,* Alexis A. Dowhuszko, *Member, IEEE,* and Olav Tirkkonen, *Member, IEEE*

***Abstract*—This paper discusses novel joint (intra-cell and inter-cell) resource allocation algorithms for self-organized interference coordination in multi-carrier multiple-input multiple-output (MIMO) small cell networks (SCNs). The proposed algorithms enable interference coordination autonomously, over multiple degrees of freedom, such as base station transmit powers, transmit precoders, and user scheduling weights. A generic α-fair utility maximization framework is considered to analyze performance-fairness trade-off, and to quantify the gains achievable in interference-limited networks. The proposed scheme involves limited inter-base station signaling in the form of two step (power and precoder) pricing. Based on this decentralized coordination, autonomous power and precoder update decision rules are considered, leading to algorithms with different characteristics in terms of user data rates, signaling load, and convergence speed. Simulation results in a practical setting show that the proposed pricing-based self-organization can achieve up to 100% improvement in cell-edge data rates, when compared to baseline optimization strategies. Furthermore, the convergence of the proposed algorithms is also proved theoretically.**

***Index Terms*—Self-organizing networks, autonomous algorithms, interference coordination, resource allocation, multiple-input multiple-output, co-channel interference, small cell networks, network utility maximization.**

I. Introduction

EFFECTIVE interference mitigation in spectrum sharing wireless networks entails efficient allocation of resources across the available degrees of freedom (time, frequency, and space) in a non-interfering way, thereby reducing the co-channel interference and leading to higher spectral efficiency and better network level performance [1]. For cellular networks, this translates to improved quality of service (QoS) for all served users, provided that the fairness issues are considered at network level. The fairness among users is influenced by a multitude of factors, such as asymmetric channel conditions, disparity in transmitter powers, and spectrum allocation. These issues are especially relevant to the currently emerging small cell paradigm, where a large number of low power base stations (BSs) are deployed in conjunction with macro BSs to increase the capacity and improve the coverage of the network. The small cells are inherently different from typical macro cells, and lead to new design challenges related to self-organized interference management [2].

This work was funded in part by the Academy of Finland (grant 284725).

F. Ahmed is with KTH Royal Institute of Technology, 164 40 Kista, Sweden (e-mail: furqanah@kth.se). A. A. Dowhuszko is with the Centre Tecnolgic de les Telecomunicacions de Catalunya, Barcelona 08860, Spain (e-mail: alexis.dowhuszko@cttc.es). O. Tirkkonen is with Aalto University, Espoo 02150, Finland (e-mail: olav.tirkkonen@aalto.fi).

In particular, the interference management for small cell networks (SCNs) and heterogeneous networks is considered as an important aspect of self-organized networking (SON) concept, and therefore requires scalable and distributed approaches [3]–[5]. For a detailed overview on SON in contemporary cellular networks, see e.g. [6] and the references therein. In particular, distributed approaches for generic self-organized resource allocation problems in SCNs have been proposed in [7], [8]. Recent approaches for interference coordination in SCNs mostly focus on time domain [9], [10], frequency domain [11]–[13], and power control techniques [14]–[17]. In addition, a number of papers focus on joint allocation of resources across multiple dimensions [18]–[22].

In this paper, we propose joint optimization of frequency, power, and spatial resources, for interference coordination in orthogonal frequency division multiple access (OFDMA)-based SCNs. This entails both intra-cell and inter-cell resource allocation, in a self-organized way. In an OFDMA wireless network, the subcarriers can be partitioned into disjoint carriers or subchannels, thereby reducing the feedback overhead related to the resource allocation. In this paper, *inter-cell resource allocation* refers to the problem of allocating both BS transmit powers over carriers and BS transmit precoders to mitigate the inter-cell interference. Our approach is based on a pricing exchange mechanism, where BSs cooperate to maximize the performance of the whole network. On the other hand, *intra-cell resource allocation* refers to the scheduling of the served users on the active carriers of a given cell, where multiple users share the resources of a carrier in an orthogonal way. Moreover, it is assumed that each carrier is infinitely divisible and, therefore, can be shared among served users with very fine granularity. The main objective is to optimize the network level utility over BS transmit powers, transmit precoders, and scheduling weights. To this end, a pricing exchange mechanism is designed, where the prices reflect the impact that co-channel interference has on the achieved satisfaction level of users in a given cell. Therefore, the possibility of cooperation among BSs over the backhaul links is exploited to optimize both inter-cell and the intra-cell resource allocation, in a distributed way. The inter-cell resource allocation mitigates the mutual interference among cells by adapting transmit powers and precoders over carriers, enabling both frequency and spatial diversity gains. On the other hand, intra-cell resource allocation yields a multi-user diversity gain through channel-aware scheduling. We formulate the optimization problem using a network utility maximization (NUM) framework, where the level of satisfaction of

each user in the network is captured through a utility function, which is a function of its achievable data rate.

A. Background and Related work

Self-organized resource allocation constitutes an important approach for enabling interference coordination in SCNs. Most of the existing work in this direction focuses on channel selection [11]–[13] and power control schemes [14], [15]. In [11], a switched-based scheme for orthogonal channel selection is proposed for single-antenna overloaded SCNs. Autonomous mechanisms for component carrier selection problem based on background interference matrices is considered in [12]. Game-theoretic approach for fully distributed channel selection are presented in [13], [23]. Joint allocation of channel and transmit powers is discussed in [14]. Joint allocation of multiple resources is considered in [18]–[22], [24], [25]. A decentralized algorithm for user scheduling and power allocation is proposed in [24]. Joint carrier and transmit power allocation is discussed in [25]. In [19], joint resource allocation and user association is addressed, whereas [20] discusses joint resource management and interference coordination. User scheduling and almost blank resource block allocation is proposed in [18]. The authors of [21] focus on a game theoretical approach for joint spectral and energy efficiency, and in [22] transmit power allocation and user scheduling is considered.

In this paper, the concept of joint resource optimization in a multi-carrier multiple-input multiple-output (MIMO) SCN is extended towards multiple dimensions, which include frequency carriers, user scheduling, and spatial beams. Joint optimization of these resources is a challenging problem and, to the best of our knowledge, has not been considered before.

B. Contributions

The current paper generalizes the analysis presented in [26], where a distributed algorithm is discussed for transmit power and scheduling weight allocation in a multi-carrier SCN. The generalization is in two directions. First, we generalize NUM algorithms so that convergence can be guaranteed for utility functions that are not Lipschitz continuous. Commonly used utility functions with quantified fairness, such as the proportionally fair, harmonic mean, and max-min-rate utility functions fall into this class. We consider NUM with α-fair utility, which subsumes all these non-Lipschitz continuous utility functions. For these, NUM with quantified fairness cannot be reliably treated by using known iterative algorithms, from e.g. [27], in a conventional manner. Second, in addition to power and scheduling weight allocation, we extend the analysis to the multi-antenna domain, where the multiple antennas are considered not only to provide multi-user diversity gain, but also to provide spatial multiplexing using the same precoders to serve multiple users per carrier. Thus, a generic NUM framework with α-fair utility is considered for joint power, scheduling weight, and precoder allocation when serving multiple users per carrier. In order to analyze the performance of proposed algorithms, a realistic Urban Microcell (UMi) scenario [28] is used for simulation purposes. Furthermore, convergence analysis of the algorithms based on both theoretical proofs as well as simulation results is also presented. The main aim is to design distributed and self-organizing algorithms to find a joint solution of the following three related subproblems: 1) power allocation, 2) precoder allocation, and 3) multi-user scheduling. To the best of our knowledge, the joint optimization over these parameters has not been addressed in the literature before. In this paper, we study the allocation of these resources to maximize the network utility. The proposed algorithms are based on finding the solution of Karush-Kuhn-Tucker (KKT) optimality conditions of the NUM problem. Two special cases of the α-fair utility function considered in the simulations are the Maximum-Rate (Max-Rate) utility function ($\alpha = 0$) [29], and the Proportionally Fair-Rate (PF-Rate) utility function ($\alpha = 1$) [30] [31]. The inter-cell resource allocation requires only the exchange of the power and precoder prices. A schematic illustration of the proposed pricing concept for MIMO SCNs is given in Fig. 1, where two BSs are communicating to their respective mobile stations (MSs), while exchanging interference prices to mitigate the mutual interference. Furthermore, it is assumed that channel gains are estimated periodically through common downlink pilot signals.

C. Notation

The vectors and matrices are represented using boldface lowercase and uppercase letters, respectively. The trace operation is represented by $\mathrm{Tr}(\cdot)$, the conjugate transpose by $(\cdot)^{\mathrm{H}}$, and the matrix inverse by $(\cdot)^{-1}$. The identity matrix is denoted by $\mathbf{I}$ (with dimensions clear from the context), and the determinant of the matrix by $\det(\cdot)$. For any matrices $\mathbf{X}, \mathbf{Y}$, $\mathbf{X} \succeq \mathbf{Y}$ is an order relation, indicating that the matrix $\mathbf{X} - \mathbf{Y}$ is a positive semidefinite matrix. The cardinality of set $\mathcal{X}$ is denoted by $|\mathcal{X}|$, whereas $[\cdot]_{\mathcal{Y}}$ represents the projection on constraints set $\mathcal{Y}$. Moreover, $\mathbb{C}$ and $\mathbb{R}$ symbolize the sets of complex and real numbers, respectively.

The rest of the paper is organized as follows: Section II discusses the system model and formulates the NUM based resource optimization problem. Section III introduces the decomposition using distributed pricing concept, and discusses different algorithmic approaches for resource allocation in a multi-carrier SCN. Section IV discusses the method based on Gauss-Seidel gradient projection, whereas Section V focuses on an approach based on nonlinear Gauss-Seidel method. In Section VI, comparison of the proposed algorithms is presented by an analysis of the simulations carried out in a standard SCN scenario. Finally, conclusions are given in Section VII.

II. System Model and Problem Formulation

A. System Model

Consider a downlink multi-cell scenario where a number of low-power MIMO BSs with indices $i \in \mathcal{I}$ are deployed within a certain area to serve a group of MSs (users) denoted by $k \in \mathcal{K}$. A MS is served by the BS with minimum path loss. The association of MSs to BSs is fixed and known *a priori*,

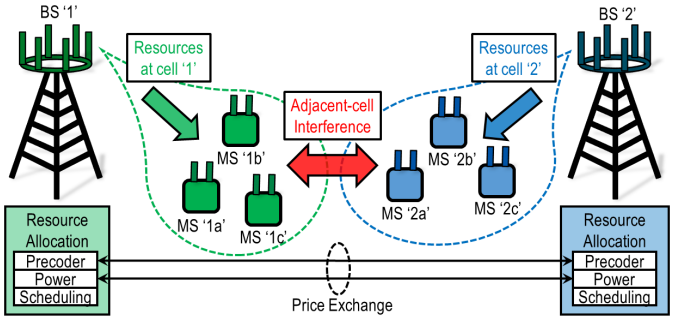

Fig. 1. Pricing based distributed resource allocation can enable interference mitigation in multi-cell MIMO networks.

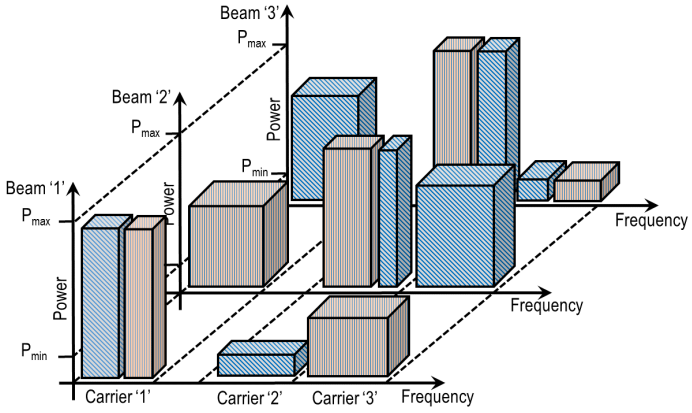

Fig. 2. Multi-cell MIMO networks involve multiple degrees of freedom that can be exploited via efficient resource allocation for interference mitigation leading to network utility maximization.

where $\{\mathcal{K}_i : i = 1, \ldots, |\mathcal{I}|\}$ represents a partition of set $\mathcal{K}$ so that each MS is served by a unique BS, i.e.,

$$\mathcal{K} = \bigcup_{i \in \mathcal{I}} \mathcal{K}_i, \qquad \mathcal{K}_i \cap \mathcal{K}_j = \emptyset \quad i \neq j. \tag{1}$$

The total bandwidth B is re-used by all cells, and is divided into M equal bandwidth (non-)contiguous carriers. Each BS has N_{T} transmit antennas and each MS has N_{R} receive antennas. Let $\mathbf{x}_i^{(m)}$ be the transmit signal vector of BS i on carrier m, and let $\mathbf{n}_k^{(m)}$ be complex Gaussian noise at MS k with covariance matrix $\mathbf{R}_{n_{i,k}}^{(m)}$. Then, the received signal vector at carrier m of MS k is

$$\mathbf{y}_k^{(m)} = \mathbf{H}_{i,k}^{(m)} \mathbf{x}_i^{(m)} + \sum_{j \neq i} \mathbf{H}_{j,k}^{(m)} \mathbf{x}_j^{(m)} + \mathbf{n}_k^{(m)} \qquad k \in \mathcal{K}_i, \tag{2}$$

where $\mathbf{H}_{i,k}^{(m)} \in \mathbb{C}^{N_{\mathrm{T}} \times N_{\mathrm{R}}}$ is the direct channel gain matrix between BS i and MS k on carrier m, which takes into account distance dependent path loss and shadow fading components. We divide the transmit signal covariance matrix into two parts – a normalized covariance matrix and a scalar average power. Thus, for BS i and carrier m, we have $\mathrm{E}\{\mathbf{x}_i^{(m)} \mathbf{x}_i^{(m)\mathrm{H}}\} = p_i^{(m)} \mathbf{Q}_i^{(m)}$, where $\mathrm{Tr}\left(\mathbf{Q}_i^{(m)}\right) = 1$ and $p_i^{(m)}$ is the transmit power of BS i on carrier m. The normalized covariance matrix $\mathbf{Q}_i^{(m)}$ is positive semidefinite. The carrier-specific covariance matrices $\mathbf{Q}_i^{(m)}$ of BS i are stacked to a matrix $\mathbf{Q}_i = \left[\mathbf{Q}_i^{(1)} \ldots \mathbf{Q}_i^{(m)} \ldots \mathbf{Q}_i^{(|\mathcal{M}|)}\right]$, and all covariance matrices in the system are stacked to a large matrix $\mathbf{Q}$. We choose to separate these variables so that the multi-channel power allocation problem is governed by the variables p_i, and the per-carrier MIMO precoding and power allocation is governed by $\mathbf{Q}$. The total power that BS i can allocate on its carriers is bounded between $P_{\min}$ and $P_{\max}$. The powers of all BSs $\mathcal{I}$ on all carriers $\mathcal{M}$ is denoted by $\mathbf{P} = \left[\mathbf{p}_1 \ldots \mathbf{p}_i \ldots \mathbf{p}_{|\mathcal{I}|}\right]$, where $\mathbf{p}_i$ is vector comprising of transmit powers of BS i on carriers $m \in \mathcal{M}$.

When optimizing the allocation of resources, there are three problems to address. The design of elements of $\mathbf{P}$ constitutes the *power allocation* problem. The design of $\mathbf{Q}$ is the network level *precoder allocation* problem. It is assumed that BSs schedule their resources for orthogonal transmission to the users served by the BS. It should be noted that all the interference is caused by the neighboring cells, as MSs within the serving cell are sharing resources orthogonally. The inter-cell scheduling resources are assumed infinitely divisible, so that $a_{i,k}^{(m)}$ is the scheduling weight that BS i allocates to MS k on carrier m, and $\mathbf{a}_i^{(m)}$ is a vector with scheduling weights of all MSs on carrier m. The scheduling weights of BS i over all carriers $\mathcal{M}$ are stacked to the matrix $\mathbf{A}_i = \{\mathbf{a}_i^{(1)} \ldots \mathbf{a}_i^{(m)} \ldots \mathbf{a}_i^{(|\mathcal{M}|)}\}$, and scheduling weights of all BSs in the network are stacked to matrix $\mathbf{A}$. Thus, design of $\mathbf{A}$ is considered as the *multi-user scheduling* problem in individual cells. There is only one precoding matrix per carrier, which is used irrespectively of the scheduled user. This enables multi-cell optimization of the resource allocation, and decouples the power and precoder optimization problems from scheduling decisions of other cells. With this formulation, it is possible to allocate resources across multiple degrees of freedom in the considered multi-carrier multi-cell MIMO SCN. Figure 2 delineates this for a given MIMO BS, which has the flexibility to allocate its transmit power over carriers, and can also adjust the direction of its beams to direct power in the spatial domain.

We model the achievable data rate of a user with the well-known MIMO mutual information formula, assuming that the interference can be treated as noise. Therefore, the achievable rate of user k in cell i (in nats/s) is

$$r_k = \sum_{m \in \mathcal{M}} \frac{B}{|\mathcal{M}|} a_{i,k}^{(m)} \log \left\{ \det \left(\mathbf{I} + \mathbf{Z}_{i,k}^{(m)} \mathbf{X}_{i,k}^{-1(m)} \right) \right\}, \tag{3}$$

where

$$\mathbf{Z}_{i,k}^{(m)} = p_i^{(m)} \mathbf{H}_{i,k}^{(m)} \mathbf{Q}_i^{(m)} \mathbf{H}_{i,k}^{(m)\mathrm{H}}, \tag{4}$$

and the noise-plus-interference covariance matrix is

$$\mathbf{X}_{i,k}^{(m)} = \left(\mathbf{R}_{n_{i,k}} + \textstyle\sum_{j \neq i} p_j^{(m)} \mathbf{H}_{j,k}^{(m)} \mathbf{Q}_j^{(m)} \mathbf{H}_{j,k}^{(m)\mathrm{H}} \right). \tag{5}$$

B. Problem Formulation

The aim is to maximize the network utility in the SCN downlink over BS transmit powers $\mathbf{P}$, transmit precoders

covariances $\mathbf{Q}$, and scheduling weights $\mathbf{A}$. The network utility is defined as

$$U_{\text{sum}}\left(\mathbf{P},\mathbf{Q},\mathbf{A}\right)=\sum_{i\,\in\,\mathcal{I}}U_i\left(\mathbf{P},\mathbf{Q},\mathbf{A}_i\right), \tag{6}$$

where $U_i\left(\mathbf{P},\mathbf{Q},\mathbf{A}_i\right)$ is the utility of BS i. Note that the utility functions of the BSs are coupled through both $\mathbf{P}$ and $\mathbf{Q}$. This means that the utility of BS i will be affected by any change either in the power or the precoder allocation of BS j, for all $j\neq i$. The sum-utility of cell i, which is the sum of utilities of the users served by BS i, can be expressed as

$$U_i\left(\mathbf{P},\mathbf{Q},\mathbf{A}_i\right)=\sum_{k\,\in\,\mathcal{K}_i}u_k, \tag{7}$$

where u_k is the utility function of user k, which is a function of the aggregate user rate r_k. The utility function is chosen according to the performance metric to be optimized. A trade-off exists between network level performance and fairness among users. An α-fair formulation to explore this trade-off is given by [32], i.e.,

$$u\left(r\right)=\begin{cases}\frac{1}{1-\alpha}\left(r\right)^{1-\alpha} & \alpha\neq 1,\\ \log\left(r\right) & \alpha=1,\end{cases} \tag{8}$$

where r is the rate of the user. Cases of special interest are: (1) $\alpha=0$ or the Max-Rate utility function, (2) $\alpha=1$, or the PF-Rate utility function, which leads to a more balanced approach in terms of individual user rates, (3) α = 2, or the harmonic mean utility function, and (4) $\alpha\rightarrow\infty$, which yields the Max-Min Rate utility function. Taking $\alpha>0$ allows to improve the QoS of the users that have low signal to interference-plus-noise ratios (SINRs), due to high co-channel interference emanating from nearby BSs, at the expense of reducing the data rate of users with high SINR. This makes the utility function *non-separable* [33], which increases the complexity of the decomposition. Moreover, for a generic α-fair utility function, the derivative is $u'(r)=r^{\alpha}$. The rates r are non-negative real numbers. For $\alpha>0$, the derivative is unbounded when $r\rightarrow 0$. This causes problems in convergence of conventional algorithms, as the gradients of the network utility functions are not Lipschitz continuous in the optimization variables, and thus differing from the functions considered in [27], [31]. Note that for $0<\alpha<1$, the utility function itself is not singular when $r\rightarrow 0$, whereas the derivative is. Irrespective of this, we call points in configuration space where one or more users have $r=0$ utility singularities. For $\alpha>0$, it is essential for converging algorithms to avoid neighborhoods of utility singularities.

The network level optimization is constrained by cell-specific constraints in the set

$$\begin{aligned}\mathcal{C}_i \;=\; \Big\{ &\sum_{m\in\mathcal{M}} p_i^{(m)}\leq P_{\max}\;,\quad \sum_{m\in\mathcal{M}} p_i^{(m)}\geq P_{\min}\;,\\ &p_i^{(m)}\geq 0\;,\,\mathrm{Tr}\Big(\mathbf{Q}_i^{(m)}\Big)=1\;,\quad \mathbf{Q}_i^{(m)}\succeq 0\;,\\ &a_{i,k}^{(m)}\geq 0\;,\quad \sum_{k\in\mathcal{K}_i} a_{i,k}^{(m)}=1\;\Big\},\end{aligned} \tag{9}$$

where all free indexes $m\in\mathcal{M}$ and $k\in\mathcal{K}_i$. The Power-Precoder-Scheduling ($\mathbf{P}-\mathbf{Q}-\mathbf{A}$) NUM problem that we aim to solve is

$$\begin{aligned}\underset{\mathbf{P},\,\mathbf{Q},\,\mathbf{A}}{\text{maximize}}\quad & \textstyle\sum_{i\in\mathcal{I}} U_i\left(\mathbf{P},\mathbf{Q},\mathbf{A}_i\right)\\ \text{subject to}\quad & \text{constraints } \mathcal{C}_i\;,\quad i\in\mathcal{I}.\end{aligned} \tag{P1}$$

The objective function is the network-wide utility, which is the sum of utility of all users served by the BSs, and the optimization is carried out over powers, precoders and scheduling weights. The sum of scheduling weights on each carrier equals 1, and the BS power over carriers is constrained by a maximum and minimum transmit power limit, given by $P_{\max}$ and $P_{\min}$, respectively. Also, there are non-negativity constraints on powers and scheduling weights. This is generically a non-convex problem in both transmit powers and transmit precoder covariances, and therefore, it is difficult to solve, even in a centralized setting [34] [35]. The non-convexity arises because of the mutual interference that couples SINRs of MSs served by different BSs. We aim to find a (local) optimum of (P1) in a distributed way by decomposing (P1) into $|\mathcal{I}|$ subproblems, one per BS. This set of $|\mathcal{I}|$ subproblems can be considered as a distributed version of the network level optimization problem, which can then be solved with the help of pricing information that is exchanged between BSs.

C. Computational Complexity

For some optimization problems, changes of variables exist, which reveal a convex formulation of a seemingly non-convex problem. To understand the nature of a problem, it is important to understand its underlying computational complexity. When $|\mathcal{M}|>2$, it is possible to show that (P1) is NP-hard, by using reduction. In order to give a sketch of the proof, let us simplify the problem by considering a SCN comprising of multiple cells with single antenna BSs. Moreover, each BS has only one user in its cell. The user is scheduled on full resources of the cell, i.e. on all carriers. This reduces the problem to transmit power allocation problem over multiple channels for multiple interfering links, where a link represents a BS-MS pair in a given cell. This problem can be shown to be NP-hard [36], by reducing it to a maximum independent set problem for $\alpha=0$, and to a graph coloring problem for other values of α. Thus, the NP-hardness of (P1) with generic parameters follows for the generic α-fair utility.

III. Power-Precoder-Scheduling NUM

We first derive the KKT optimality conditions of problem (P1), followed by a pricing-based algorithm to compute the solution in a distributed way.

A. KKT Optimality Conditions

Let $\Psi_i^{(m)}\in\mathbb{R}$ and $\mathbf{\Theta}_i^{(m)}\in\mathbb{R}^{N_{\mathrm{T}}\times N_{\mathrm{R}}}$ be the Lagrange multipliers associated with the constraints on precoder covariance matrices, and $\Phi_i^{(m)}\in\mathbb{R}$ the Lagrange multipliers defined for the constraint on sum of BS scheduling weights on carrier m of BS i. For notational ease, we define stacked matrices $\mathbf{\Psi}\triangleq\{\{\Psi_i^{(m)}\}_{m\in\mathcal{M}}\}_{i\in\mathcal{I}}$, $\mathbf{\Theta}\triangleq\{\{\mathbf{\Theta}_i^{(m)}\}_{m\in\mathcal{M}}\}_{i\in\mathcal{I}}$, and $\mathbf{\Phi}\triangleq\{\{\Phi_i^{(m)}\}_{m\in\mathcal{M}}\}_{i\in\mathcal{I}}$, respectively. Likewise, we introduce Lagrange multipliers $\Lambda_i\in\mathbb{R}$ and $\Upsilon_i\in\mathbb{R}$ for the

$$\mathcal{L}=\sum_{i\in\mathcal{I}}\Bigg(U_i-\sum_{m\in\mathcal{M}}\Bigg[\Psi_i^{(m)}\left(\mathrm{Tr}\mathbf{Q}_i^{(m)}-1\right)-\mathrm{Tr}\Big(\mathbf{\Theta}_i^{(m)}\mathbf{Q}_i^{(m)}\Big)+\Phi_i^{(m)}\left(\sum_{k\in\mathcal{K}_i}a_{i,k}^{(m)}-1\right)$$
$$-\lambda_i^{(m)}p_i^{(m)}-\sum_{k\in\mathcal{K}_i}\phi_{i,k}^{(m)}a_{i,k}^{(m)}\Bigg]-\Lambda_i\Big(\sum_{m\in\mathcal{M}}p_i^{(m)}-P_{\max}\Big)-\Upsilon_i\Big(P_{\min}-\sum_{m\in\mathcal{M}}p_i^{(m)}\Big)\Bigg) \tag{10}$$

minimum and maximum constraints on sum power, whereas we define the Lagrange multipliers $\lambda_i^{(m)}\in\mathbb{R}$ and $\phi_i^{(m)}\in\mathbb{R}$ to handle the non-negativity constraints related to BS power and scheduling weights, respectively. Again, these multipliers can be stacked as matrices $\mathbf{\Lambda}\triangleq\{\Lambda_i\}_{i\in\mathcal{I}}$,$\mathbf{\Upsilon}\triangleq\{\Upsilon_i\}_{i\in\mathcal{I}}$, $\boldsymbol{\lambda}\triangleq\{\{\lambda_i^{(m)}\}_{m\in\mathcal{M}}\}_{i\in\mathcal{I}}$, and $\boldsymbol{\phi}\triangleq\{\{\phi_i^{(m)}\}_{m\in\mathcal{M}}\}_{i\in\mathcal{I}}$, respectively. Then, the Lagrangian of problem (P1) is given by $\mathcal{L}=\mathcal{L}(\mathbf{P},\mathbf{Q},\mathbf{A},\mathbf{\Psi},\mathbf{\Theta},\mathbf{\Phi},\mathbf{\Lambda},\mathbf{\Upsilon},\boldsymbol{\lambda},\boldsymbol{\phi})$, as expanded in (10). For any local maximum of problem (P1), there exists a set of unique Lagrange multipliers such that the following KKT necessary conditions hold for all $i\in\mathcal{I}$ and $m\in\mathcal{M}$:

Stationarity:

$$\begin{aligned}&\tfrac{\partial U_i(\mathbf{P},\mathbf{Q},\mathbf{A}_i)}{\partial p_i^{(m)}}+\textstyle\sum_{j\neq i}\tfrac{\partial U_j(\mathbf{P},\mathbf{Q},\mathbf{A}_j)}{\partial p_i^{(m)}}-\Lambda_i+\Upsilon_i+\lambda_i^{(m)}=0,\\&\tfrac{\partial U_i(\mathbf{P},\mathbf{Q},\mathbf{A}_i)}{\partial \mathbf{Q}_i^{(m)}}+\textstyle\sum_{j\neq i}\tfrac{\partial U_j(\mathbf{P},\mathbf{Q},\mathbf{A}_j)}{\partial \mathbf{Q}_i^{(m)}}-\Psi_i^{(m)}\mathbf{I}+\mathbf{\Theta}_i^{(m)}=0,\\&\tfrac{\partial U_i(\mathbf{P},\mathbf{Q},\mathbf{A}_i)}{\partial a_{i,k}^{(m)}}-\Phi_{i,k}^{(m)}+\phi_{i,k}^{(m)}=0,\end{aligned} \tag{11}$$

Primal Feasibility:

$$\begin{aligned}&\textstyle\sum_{m\in\mathcal{M}}p_i^{(m)}\leq P_{\max},\quad \sum_{m\in\mathcal{M}}p_i^{(m)}\geq P_{\min},\\&p_i^{(m)}\geq 0,\qquad \mathrm{Tr}\Big(\mathbf{Q}_i^{(m)}\Big)=1,\qquad \mathbf{Q}_i^{(m)}\succeq 0,\\&\textstyle\sum_{k\in\mathcal{K}_i}a_{i,k}^{(m)}=1,\qquad\qquad\qquad a_{i,k}^{(m)}\geq 0.\end{aligned} \tag{12}$$

For the sake of brevity, we omit the conditions pertaining to dual feasibility and complementary slackness. Let us define the following term given in KKT conditions as *power price*:

$$\begin{aligned}\pi_{j,i}^{(m)}&=-\tfrac{\partial U_j(\mathbf{P},\mathbf{Q},\mathbf{A}_j)}{\partial p_i^{(m)}}\\&=\textstyle\sum_{k\in\mathcal{K}_j}r_k{}^{-\alpha}a_{j,k}^{(m)}\mathrm{Tr}\Bigg(\Big(\mathbf{M}_{j,k}^{(m)}\Big)^{-1}\mathbf{Z}_{j,k}^{(m)}\mathbf{V}_{j,k}^{(m)}\Bigg),\quad j\neq i,\end{aligned} \tag{13}$$

where $\mathbf{V}_{j,k}^{(m)}=\left(\mathbf{X}_{j,k}^{(m)}\right)^{-1}\mathbf{H}_{i,k}^{(m)}\mathbf{Q}_i^{(m)}\mathbf{H}_{i,k}^{(m)\mathrm{H}}\left(\mathbf{X}_{j,k}^{(m)}\right)^{-1}$ and $\mathbf{M}_{i,k}^{(m)}=\mathbf{I}+\mathbf{Z}_{i,k}^{(m)}\mathbf{X}_{i,k}^{-1(m)}$. Similarly, one can define a *precoder price* as follows:

$$\begin{aligned}\mathbf{\Pi}_{j,i}^{(m)}&=-\tfrac{\partial U_j(\mathbf{P},\mathbf{Q},\mathbf{A}_j)}{\partial \mathbf{Q}_i^{(m)}}\\&=\textstyle\sum_{k\in\mathcal{K}_j}(r_k)^{-\alpha}a_{j,k}^{(m)}\tilde{\mathbf{V}}_{j,k}^{(m)},\quad j\neq i,\end{aligned} \tag{14}$$

where$\tilde{\mathbf{V}}_{j,k}^{(m)}=\mathbf{H}_{i,k}^{(m)\mathrm{H}}\Big(\mathbf{X}_{j,k}^{(m)}+\mathbf{Z}_{j,k}^{(m)}\Big)^{-1}\mathbf{Z}_{j,k}^{(m)}\left(\mathbf{X}_{j,k}^{(m)}\right)^{-1}\mathbf{H}_{i,k}^{(m)}$. The prices are calculated in a way that all MSs in a given cell calculate their prices towards interfering BSs, and then they report this pricing to the serving BS. The serving BS then computes the aggregate price for all interfering BSs and communicates it to them. However, this requires the knowledge of cross channel gains from interfering BSs. These channel gains are estimated through periodic transmission of pilot signals, where all BSs transmit a pilot signal so that the MSs can measure the channel gains. However, the exchange of prices can take place over low-rate links that connect the BSs. With the channel gains and prices available, each BS can allocate its resources to maximize the network utility in a distributed way. This distributed problem is formalized by decomposition of the NUM problem, as discussed in the following section.

B. Distributed Formulation

Following the decomposition procedure [26], [34], we formalize a distributed problem that has same KKT conditions as the network level optimization problem (P1), for all $i\in\mathcal{I}$, provided that optimization variables at cell $j\neq i$ remain fixed, during the resource allocation update at cell i. Let us define the surplus function as

$$s_i=U_i-\sum_{m\in\mathcal{M}}\sum_{j\neq i}\Big(p_i^{(m)}\pi_{j,i}^{(m)}+\mathrm{Tr}\Big(\mathbf{Q}_i^{(m)}\mathbf{\Pi}_{j,i}^{(m)}\Big)\Big)\ . \tag{15}$$

Then, the distributed problem becomes

$$\begin{aligned}&\text{maximize}\quad s_i\\&\mathbf{p}_i,\mathbf{Q}_i,\mathbf{A}_i\\&\text{subject to}\quad \text{constraints }\mathcal{C}_i\ .\end{aligned} \tag{16}$$

Accordingly, each BS $i\in\mathcal{I}$ solves its individual subproblem, which involves joint optimization over $\mathbf{p}_i$, $\mathbf{Q}_i$, and $\mathbf{A}_i$. Following the approach presented in [37], it is possible to separate the joint optimization for each BS $i\in\mathcal{I}$, so that the optimizations over $\mathbf{p}_i$, $\mathbf{Q}_i$, and $\mathbf{A}_i$ are carried out separately. The power allocation and precoder allocation problems couple the decisions between cells i, whereas the multi-user scheduling problem does not directly affect the interference experienced in neighboring cells, but it does affect the interference prices reported to the neighbors.

C. Algorithms: Background and New Perspectives

In the following sections, we design distributed network optimization algorithms based on Gauss-Seidel gradient projection (GSGP) and non-linear Gauss-Seidel (NLGS) approaches [27]. However, our approach differs in a number of important ways. In particular, the treatment in [27] is focused on a simplistic scenario with Lipschitz continuous utility functions, whereas we consider generic non-Lipschitz continuous utility functions in a distributed NUM setting, which poses significant challenges for distributed optimization. To this end, we devise a method that is based on the direct solution of KKT equations, with guaranteed convergence even in the case where the utility functions may have infinite gradients.

IV. GSGP FOR $\mathbf{P}-\mathbf{Q}-\mathbf{A}$ NUM

The gradient of network utility can be calculated using the power and precoder prices that each BS receives from the neighboring BSs. Therefore, it is possible to formulate a gradient projection algorithm based on Gauss-Seidel updates for both power and precoder covariance matrices. In a given iteration round, each BS updates its powers, precoders, and scheduling weights, while taking into account the prices from the other BSs. The updated values are calculated by taking a fixed step towards the gradient direction.

A. Gradients

To derive the update rule, we first calculate the gradient of the network utility function with respect to optimization variables (i.e. transmit powers, covariance matrices of transmit precoders, and scheduling weights). To this end, the gradients of the network utility with respect to power $p_i^{(m)}$ and covariance matrix $\mathbf{Q}_i^{(m)}$ can be calculated in terms of pricing expressions $\pi_{j,i}^{(m)}$ and $\mathbf{\Pi}_{j,i}^{(m)}$, respectively. For the transmit power case, gradient results

$$\nabla_{p_i^{(m)}} U_{\text{sum}} \quad = \pi_{i,i}^{(m)} - \textstyle\sum_{j\neq i} \pi_{j,i}^{(m)}, \tag{17}$$

where $\pi_{i,i}^{(m)} = \frac{\partial U_i}{\partial p_i^{(m)}} = \sum_{k\in\mathcal{K}_i} a_{i,k}^{(m)} {r_k}^{-\alpha} \text{Tr}\left(\mathbf{Z}_{i,k}^{(m)}\right)$ and $\mathbf{Z}_{i,k}^{(m)} = \left(\mathbf{M}_{i,k}^{(m)}\right)^{-1} \mathbf{H}_{i,k}^{(m)} \mathbf{Q}_i^{(m)} \mathbf{H}_{i,k}^{(m)\text{H}} \left(\mathbf{X}_{i,k}^{(m)}\right)^{-1}$. Similarly, for the precoding covariance matrix case, the gradient attains the form

$$\nabla_{\mathbf{Q}_i^{(m)}} U_{\text{sum}} \quad = \mathbf{\Pi}_{i,i}^{(m)} - \textstyle\sum_{j\neq i} \mathbf{\Pi}_{j,i}^{(m)}, \tag{18}$$

where $\mathbf{\Pi}_{i,i}^{(m)} = \frac{\partial U_i}{\partial \mathbf{Q}_i^{(m)}} = \sum_{k\in\mathcal{K}_i} a_{i,k}^{(m)} {r_k}^{-\alpha} \mathbf{Y}_{i,k}^{(m)} p_i^{(m)} \mathbf{H}_{i,k}^{(m)}$ and $\mathbf{Y}_{i,k}^{(m)} = \mathbf{H}_{i,k}^{(m)\text{H}} \left(\mathbf{M}_{i,k}^{(m)}\right)^{-1} \left(\mathbf{X}_{i,k}^{(m)}\right)^{-1}$. Moreover, for scheduling weights, the gradient is given by

$$\nabla_{\mathbf{a}_i^{(m)}} U_{\text{sum}} = \frac{\partial U_i}{\partial \mathbf{a}_i^{(m)}} = \left\{ {r_k}^{-\alpha} \left[\log_e \det \left(\mathbf{M}_{i,k}^{(m)} \right) \right] \right\}_{k\in\mathcal{K}_i}. \tag{19}$$

With the gradients available as pricing terms, it is straightforward to devise a distributed GSGP iterative algorithm, for the maximization of the network utility. The precoder optimization part in GSGP is discussed in [35], for a simplistic single user case.

B. GSGP Update Rule

When the pricing information is available at a given BS, the updates can be carried out in a simple way using the GSGP principle. The gradient-based update involves a small step in the direction of gradient, followed by the projection on the given constraint set. To formulate the GSGP rule, we consider the constraint sets for BS i for powers, covariance matrices of precoders, and scheduling weights allocation subproblems. The feasible set of power vectors for a BS is

$$\mathcal{P} \triangleq \left\{ \mathbf{p} \in \mathbb{R}^{|\mathcal{M}|} \; : \; \begin{array}{l} P_{\min} \leq \sum_{m\in\mathcal{M}} p^{(m)} \leq P_{\max} \\ p^{(m)} \geq 0, \quad m \in \mathcal{M} \end{array} \right\}, \tag{20}$$

where $\mathbf{p}$ is the power vector, and a feasible set of precoder covariance matrices on a carrier is

$$\mathcal{Q} \triangleq \left\{ \mathbf{Q} \in \mathbb{C}^{N_{\text{T}}\times N_{\text{R}}} \; : \text{Tr}(\mathbf{Q}) = 1, \; \mathbf{Q} \succeq 0 \right\}. \tag{21}$$

Similarly, for BS i, the set of feasible scheduling weights on a carrier is

$$\mathcal{A}_i \triangleq \left\{ \mathbf{a} \in \mathbb{R}^{|\mathcal{K}_i|} \; : \sum_{k\in\mathcal{K}_i} a_k = 1, \; a_k \geq 0 \right\}. \tag{22}$$

The set of multi-carrier scheduling weights in cell i is the M-fold direct product $\mathcal{A}_i^{\text{tot}} = \mathcal{A}_i \times \ldots \times \mathcal{A}_i$, and the set of multi-carrier precoders the direct product $\mathcal{Q}^{\text{tot}} = \mathcal{Q} \times \ldots \times \mathcal{Q}$. For $\alpha > 0$, we restrict the feasible sets, by carving out open neighborhoods of the utility singularities, leading to the sets $\widetilde{\mathcal{P}}$, $\widetilde{\mathcal{A}}_i$, $\widetilde{\mathcal{Q}}$, $\widetilde{\mathcal{A}}_i^{\text{tot}}$ and $\widetilde{\mathcal{Q}}_i^{\text{tot}}$, see the appendix for details. The stepsizes for GSGP updates of transmit power, precoder covariance matrices and scheduling weights are denoted by δ^P, δ^Q and δ^A, respectively.

1) Power Allocation: The GSGP based power update for all carriers $m \in \mathcal{M}$ is given by:

$$p_i^{*(m)} \quad = \left[p_i^{(m)} + \delta^P \nabla_{p_i^{(m)}} U_{\text{sum}} \right]_{\widetilde{\mathcal{P}}}. \tag{23}$$

2) Precoder Allocation: The next step is the precoder update for all carriers $m \in \mathcal{M}$:

$$\mathbf{Q}_i^{*(m)} \quad = \left[\mathbf{Q}_i^{(m)} + \delta^Q \nabla_{\mathbf{Q}_i^{(m)}} U_{\text{sum}} \right]_{\widetilde{Q}}. \tag{24}$$

3) Scheduling Allocation: Similarly, for the scheduling weight update on all carriers $m \in \mathcal{M}$:

$$\mathbf{a}_i^{*(m)} \quad = \left[\mathbf{a}_i^{(m)} + \delta^A \nabla_{\mathbf{a}_i^{(m)}} U_{\text{sum}} \right]_{\widetilde{\mathcal{A}}_i}. \tag{25}$$

To solve the distributed subproblems at each BS, an asynchronous and periodic update procedure is considered, where in a given iteration n, each BS i performs its updates only once at a unique time instant $t_i[n]$, and at this instant no other BS changes its resource allocation. The GSGP algorithm is summarized in the flowchart given in Fig. 3, along with the NLGS algorithm which is discussed next.

V. NLGS FOR $\mathbf{P}-\mathbf{Q}-\mathbf{A}$ NUM

Like GSGP, NLGS is a generic approach for solving the non-linear optimization problems in an iterative manner based on Gauss-Seidel updates. In [27], non-linear algorithms were considered, where the full non-linear dependence of the objective function is known to the agent controlling a variable. When using NLGS methodology for a distributed solution of P − Q − A NUM, the objective function is partially linearized though exchange of prices to enable distributed operation. In [33], distributed multi-carrier power control was addressed, where the effect that increasing power in one cell has on the utility of a neighboring cell was linearized through power prices. Bounded Lipschitz-continuous non-separable utility functions were considered. Here, we have generic utility functions with unbounded gradients, and in addition to multi-carrier power allocation, we consider both

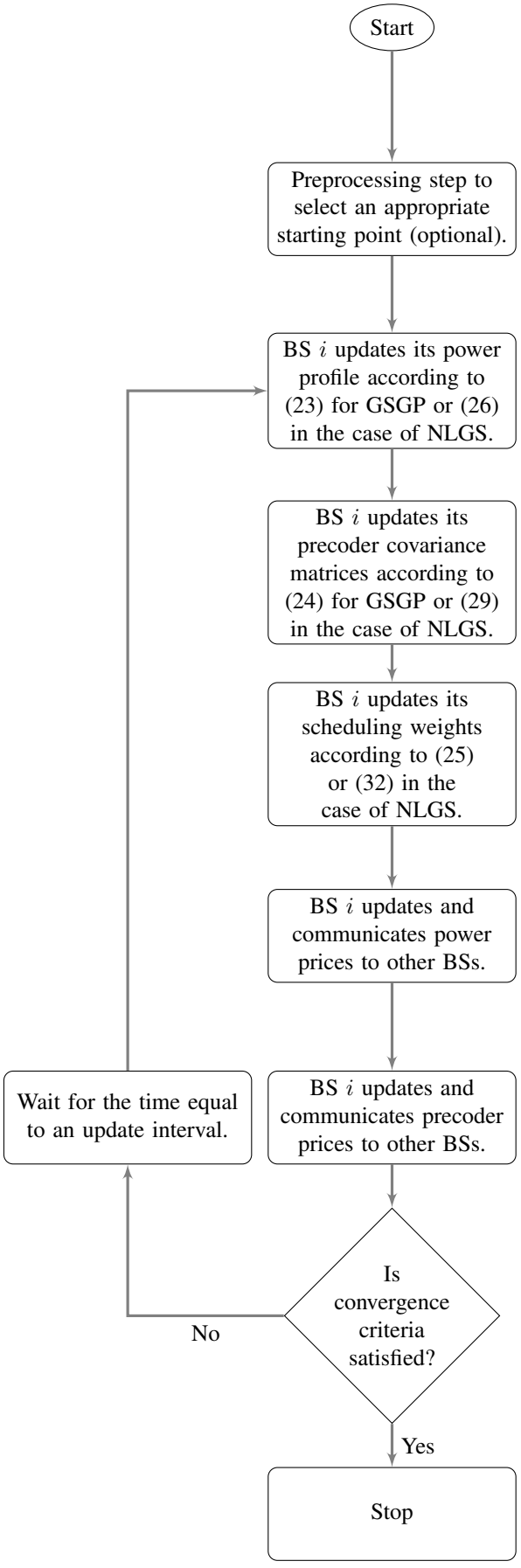

Fig. 3. Flow chart summarizing the GSGP and NLGS algorithms that run in a distributed manner over each BS i of the SCN.

multi-user scheduling and MIMO precoder allocation. In a given iteration, each BS solves a set of optimization problems while taking into account the power and precoder prices. The order in which these problems are solved is the same as GSGP. However, the update rules for subproblems are different, as discussed next.

A. Power Allocation

To optimize the sum data rate over power allocation on carriers in a distributed way, we consider the following update rule for BS $i \in \mathcal{I}$, at time instant n:

$$\mathbf{p}_i\left(n+1\right) = \left(1-\beta^P\right)\mathbf{p}_i\left(n\right) + \beta^P \mathbf{p}_i^*, \tag{26}$$

where

$$\begin{aligned} \mathbf{p}_i^* = \arg\max_{\mathbf{p}_i}\ & s_i^P(\mathbf{p}_i;\mathbf{P}_{-i},\mathbf{Q},\mathbf{A}_i) \\ \text{subject to } & \mathbf{p}_i \in \widetilde{\mathcal{P}}, \end{aligned} \tag{27}$$

while the *power surplus* function is given by

$$\begin{aligned} s_i^P(\mathbf{p}_i;\mathbf{P}_{-i},\mathbf{Q},\mathbf{A}) = U_i\left(\mathbf{p}_i;\mathbf{P}_{-i},\mathbf{Q},\mathbf{A}\right) \\ - \textstyle\sum_{m\in\mathcal{M}} p_i^{(m)}\left(\sum_{j\neq i}\pi_{j,i}^{(m)}\right). \end{aligned} \tag{28}$$

The power prices $\pi_{j,i}^{(m)}$ are given in (13). Note that s_i^P is a concave function in power $\mathbf{p}_i$, provided that the rest of the parameters remain constant. This optimization problem can be solved using gradient projection method, by using the gradients calculated in the previous section. Thus, the BS power is updated using the fixed stepsize β^P, with a convex combination of the current and the new updated power vectors. The stepsize can be chosen in different ways, like e.g. constant stepsize considered here, or Armijo rule [35]. Performance and convergence rate may differ from case to case. Next, we discuss the precoder allocation followed by multi-user scheduling.

B. Precoder Allocation

For optimization of precoder allocation, we have the following update for all BSs $i \in \mathcal{I}$:

$$\mathbf{Q}_i\left(n+1\right) = \left(1-\beta^Q\right)\mathbf{Q}_i\left(n\right) + \beta^Q \mathbf{Q}_i^*. \tag{29}$$

Here, the update is done with stepsize β^Q over all the carriers simultaneously, where

$$\begin{aligned} \mathbf{Q}_i^* = \arg\max_{\mathbf{Q}_i}\ & s_i^Q(\mathbf{Q}_i;\mathbf{P}_{-i},\mathbf{Q}_{-i},\mathbf{A}_{-i}) \\ \text{subject to } & \mathbf{Q}_i \in \widetilde{\mathcal{Q}}, \end{aligned} \tag{30}$$

and the surplus function is given by

$$\begin{aligned} s_i^Q(\mathbf{Q}_i;\mathbf{P}_{-i},\mathbf{Q}_{-i},\mathbf{A}_{-i}) = U_i\left(\mathbf{Q}_i;\mathbf{P},\mathbf{Q}_{-i},\mathbf{A}_i\right) \\ - \textstyle\sum_{m\in\mathcal{M}} \mathrm{Tr}\left(\mathbf{Q}_i^{(m)}\sum_{j\neq i}\mathbf{\Pi}_{j,i}^{(m)}\right). \end{aligned} \tag{31}$$

The precoder prices $\mathbf{\Pi}_{j,i}^{(m)}$ are given in (14). With prices and precoders of all BSs $j \neq i$ fixed, the surplus maximization problem in (30) is convex and can be solved by the gradient projection method with fixed stepsize. This requires gradient of function s_i^Q with respect to $\mathbf{Q}_i^{(m)}$ for all $m \in \mathcal{M}$, and the

projection on the set of positive semidefinite matrices. The projection is carried out following the approach given in [35], which involves adjusting the eigenvalues of the matrix, while keeping the same eigenvectors.

C. Multi-user Scheduling

Following the same procedure for optimizing the surplus function over scheduling weight allocation, we have the following scheduling weights update with stepsize β^A for all BSs $i \in \mathcal{I}$:

$$\mathbf{A}_i\left(n+1\right) = \left(1-\beta^A\right)\mathbf{A}_i\left(n\right) + \beta^A \mathbf{A}_i^*. \tag{32}$$

The surplus maximization problem for calculating $\mathbf{A}_i^*$ is

$$\begin{aligned} \mathbf{A}_i{}^* = \arg\max_{\mathbf{A}_i} \; & U_i\left(\mathbf{A}_i;\mathbf{P},\mathbf{Q}\right) \\ \text{subject to } & \mathbf{A}_i \in \widetilde{\mathcal{A}}_i. \end{aligned} \tag{33}$$

The scheduling decisions do not affect the interference seen by neighbors, and thus prices have no role in (33). Scheduling weight optimization is a cell-internal operation, and the surplus function is just the own-cell utility. The algorithm is summarized by the flowchart in Fig. 3. Note that both GSGP and NLGS approaches are based on the Gauss-Seidel update procedure. Alternatively, one may also consider synchronous or Jacobi update procedure, which would be a special case where all updates are carried out simultaneously. It is also pertinent to mention that both algorithms are distributed and scalable, and therefore well-suited for dense SCNs scenarios. The only information exchange required for their implementation is the pricing terms. This is substantially much less feedback than the one required for power (and/or precoder) allocation algorithms that rely on complete information regarding the whole network; see e.g. [38]. The information exchange required here is clearly dependent upon the iterations that it takes to converge. The iterations required are significantly lesser in NLGS when compared to GSGP, especially with an appropriately chosen stepsize. The stepsize required for convergence in GSGP are usually very small compared to the NLGS stepsizes. Apart from that, the direction taken for the update is quite different in both algorithms, leading to different (local) optimum solutions.

D. Convergence Analysis

The convergence of GSGP and NLGS discussed below applies for any utility function that is concave in rate, and which has utility singularities, i.e. unbounded utility gradients, at most when $r \to 0$. When $\alpha > 0$, convergence can be proven if the feasible sets of optimization are restricted so that open neighborhoods of utility singularities are removed. Utility singularities are at the boundary of the original feasible sets, which makes it possible to prove convergence in the redefined feasible sets.

Proposition 1: There exists feasible sets $\tilde{\mathcal{A}}_i^{\text{tot}}, \tilde{\mathcal{P}}$, and $\tilde{\mathcal{Q}}^{\text{tot}}$, and stepsizes $\delta^A > 0$, $\delta^P > 0$ and $\delta^Q > 0$, so that GSGP applied in the feasible sets always converges to a fixed point.

Proposition 2: There exists feasible sets $\tilde{\mathcal{A}}_i^{\text{tot}}, \tilde{\mathcal{P}}$, and $\tilde{\mathcal{Q}}^{\text{tot}}$, and stepsizes $\beta^P > 0$, $\beta^Q > 0$ so that NLGS with $\beta^A = 1$, applied in the feasible sets always converges to a fixed point. The proofs can be found in the Appendix.

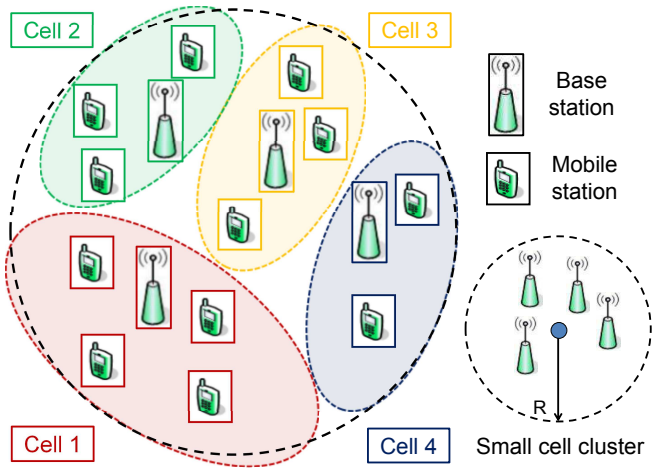

Fig. 4. SCN scenario used for simulations, where $|\mathcal{I}| = 4$ small cells, $|\mathcal{K}| = 12$ MSs, $|\mathcal{M}| = 2$ carriers, $N_{\text{T}} = N_{\text{R}} = 2$ antennas at both BSs and MSs.

VI. Performance Analysis

We study the performance of discussed resource allocation schemes by providing a comparison of different algorithms across various parameters in a practical SCN. The simulation scenario is explained first, followed by the discussion on numerical results obtained through simulations. We concentrate on Max-Rate ($\alpha = 0$) and PF-Rate ($\alpha = 1$) utilities.

A. Simulation Scenario

The considered high interference scenario comprises of a cluster of $|\mathcal{I}| = 4$ small cell BSs and $|\mathcal{K}| = 12$ MSs uniformly distributed on the surface of circular region with radius $R = 15$ m, as shown in Fig. 4. There is a minimum inter-BS separation of $d_{\min\,(\text{bs}-\text{bs})} = 5$ m. Similarly, the MSs are also deployed randomly with a minimum distance to the closest BS $d_{\min\,(\text{ms}-\text{bs})} = 5$ m. The cell association of each MS is based on the mean received signal power from BS in the downlink. Under the assumption that all BSs have the same transmit power, each MS is served by the BS from which it sees the minimum path loss (PL) attenuation. The total bandwidth of the system is $B = 10$ MHz, and it is divided equally into $|\mathcal{M}| = 2$ contiguous carriers with center frequency $f_c = 3.5$ GHz. The antenna heights at BSs and MSs are $h_{\text{bs}} = 10$ m and $h_{\text{ms}} = 1.5$ m, respectively. The distance dependent path loss and shadow fading parameters are detailed in Table I. These are modeled according to the UMi scenario specified in [28]. The channel gains between antennas of each BS-MS pair are modeled as independent and identically distributed (i.i.d) zero-mean circularly symmetric complex Gaussian variables with standard deviation dependent on path loss. The frequency response of each carrier is considered to be flat. The total transmit power of a BS is $P_{\max} = \sum_{m\in\mathcal{M}} p_i^{(m)} = 26$ dBm, and is fixed for all BSs. The equal distribution of powers over the carriers is the starting point for all the algorithms in the simulations, where pricing is exchanged on all carriers to maximize the total network utility.

TABLE I
URBAN MICRO (UMI) PATH LOSS MODEL

Path loss models [dB] (f_c given in GHz, d in meters)	**Shadow fading**	**Applicability range values**
LOS		
$\mathrm{PL} = 22\log_{10}(d) + 28.0 + 20\log_{10}(f_c)$	$\sigma = 3$ dB	$10\,\mathrm{m} \le d \le d_{\mathrm{bp}}$
$\mathrm{PL} = 40\log_{10}(d) + 7.8 - 18\log_{10}(h_{\mathrm{bs}} - 1) - 18\log_{10}(h_{\mathrm{ms}} - 1) + 2\log_{10}(f_c)$	$\sigma = 3$ dB	$d_{\mathrm{bp}} \le d \le 5000\,\mathrm{m}$
NLOS		
$\mathrm{PL} = 36.7\log_{10}(d) + 22.7 + 26\log_{10}(f_c)$	$\sigma = 4$ dB	$10\,\mathrm{m} \le d \le 2000\,\mathrm{m}$
Break point distance: $d_{\mathrm{bp}} = 4\,(h_{\mathrm{bs}} - 1)\,(h_{\mathrm{ms}} - 1)/\lambda_c$ Carrier wavelength (λ_c) and antenna heights (h_{bs}, h_{ms}) are given in meters		
LOS probability: $\mathrm{P_{LOS}} = \min(18/d, 1)\,\big(1 - \exp(-d/36)\big) + \exp(-d/36)$		

As such, to quantify the gain by pricing exchange, we compare the following four *inter-cell resource allocation strategies* for the proposed algorithms:

- Non-cooperative power allocation and Non-cooperative precoder covariance matrix allocation (GSGP/NLGS with power prices set to zero and precoder prices set to the null matrix).
- Cooperative power allocation and Non-cooperative precoder covariance matrix allocation (GSGP/NLGS with precoder prices set to the null matrix).
- Non-cooperative power allocation and Cooperative precoder allocation (GSGP/NLGS with power prices set to zero).
- Cooperative power allocation and Cooperative precoder allocation (GSGP/NLGS).

Similarly, to evaluate the performance enhancement by intra-cell resource allocation, different strategies for scheduling weight optimization are considered. Recall that intra-cell resource allocation is orthogonal, and hence, there is no intra-cell interference between MSs. To this end, we consider Max-Rate and PF-Rate scheduling as *intra-cell resource allocation strategies*. To analyze the performance, we compare different combinations of inter-cell and intra-cell resource allocation strategies. The statistics are gathered over 100 network instances, generated according to the parameters specified in Table I. For a given network instance, each algorithm is run until convergence is achieved or the maximum number of iterations is reached. The condition of convergence is that the change in total network utility is less than 0.1% in successive iterations. The stepsizes guaranteeing convergence for NLGS according to Proposition 2 would be rather small, and for GSGP according to Proposition 1 very small. In practice, larger stepsizes can be used. Here, stepsizes $\delta^P = \delta^Q = \delta^A = 0.005$ are used for all simulations of GSGP, and $\beta^P = \beta^Q = 0.7$ and $\beta^A = 1$ for NLGS. These stepsizes are chosen to obtain a balance between the convergence probability and the required iterations for convergence. The maximum number of iterations for GSGP is set to 2500, whereas for NLGS it is 250.

Furthermore, we are interested in comparing the quality of solutions found by these distributed algorithms against a global optimum. This is accomplished by employing a multi-start framework for global optimization [39]. It consists of an algorithm that attempts to find a global optimum by starting a local solver from many different initial points in the feasible region, while keeping track of the best solution found. Theoretically, this method converges to a global optimum with probability one, when the number of initial points approaches infinity. In addition to the high computational effort, this method is difficult to implement in a distributed manner. The generation of new initial points and the tracking of the best solution would require a centralized controller and a high signaling load. Due to these drawbacks, it is not suitable for resource allocation problems in self-organizing SCNs. In our simulations for finding a global optimum, we use NLGS as a local solver and set the number of random initial points to ten, for each network instance.

B. Performance of GSGP and NLGS approaches

To compare the performance of different algorithms in terms of total network data rate, the cumulative distribution function (CDF) of normalized data rate of the network summed over all users is considered. A comparison of GSGP with NLGS reveals that NLGS performs better and achieves a higher network utility than GSGP. This can be seen in Fig. 5, in which network data rate CDFs of NLGS and GSGP are illustrated for both utility functions. Moreover, global optimization performs the best, as expected.

Significantly less iterations are required for convergence in NLGS, with higher convergence probability. The achieved network utility, as expected, does not vary significantly with the change in stepsize. The difference in performance is due to the very small stepsize of GSGP, which leads to slow growth in utility towards convergence. NLGS uses full knowledge of the consequences of decisions on own-cell utility, and thus performs much better. The number of iterations required for convergence is given in Table II for all variants of GSGP and NLGS for both utility functions. These are averaged over instances that converged within the maximum number of iterations. The GSGP algorithm requires a much higher number of iterations to converge due to the gradient projection update rule that consists of a very small stepsize. For the MR utility, the number of iterations required for convergence for GSGP is approximately 10 times larger than that of NLGS. The number of iterations for NLGS may be reduced further by increasing the stepsize. However, it could jeopardize the convergence, as big changes in the values that optimization variables take may result in the inaccuracy of their respective interference prices.

The convergence probabilities are reported in Table III. The non-convergence of GSGP in some instances is the result of an oscillation around the solution, caused by too long stepsizes. These may be controlled by reducing the stepsize even further, which leads to an increase in required iterations for convergence. Note that the number of iterations directly translates to the amount of signaling between BSs. NLGS performs significantly better than GSGP both in terms of the required information exchange, and convergence probability.

TABLE II
AVERAGE NUMBER OF ITERATIONS FOR CONVERGENCE

Algorithms	Self	P	Q	Joint
GSGP (MR)	244.14	364.22	516.60	529.53
GSGP (PF)	201.30	463.97	406.04	633.24
NLGS (MR)	59.07	36.82	47.23	44.54
NLGS (PF)	81.32	81.56	91.74	114.38

TABLE III
CONVERGENCE PROBABILITIES

Algorithms	Self	P	Q	Joint
GSGP (MR)	1	1	1	0.94
GSGP (PF)	1	0.92	1	0.74
NLGS (MR)	1	1	1	1
NLGS (PF)	1	1	1	1

C. Comparison of Max Rate and PF Rate Utilities

A comparison of network data rate for Max-Rate and PF-Rate utilities is given in Fig. 6. The baseline case (Self) is the fully non-cooperative or selfish scheme in which no prices are exchanged, where Self-MR and Self-PF are the variants for Max-Rate (MR) and PF-Rate (PF) utilities respectively. It can be seen that Self-MR significantly outperforms Self-PF in terms of network data rate. This is due to the fact that the MR utility function aims at maximizing the network data rate directly over all resources. On the other hand, PF-Rate maximizes the logarithmic function of rate to ensure fairness among the users at a network level, taking into account all allocations across all resources. Note that in these two cases, Self-MR and Self-PF no pricing is being exchanged; therefore, surplus maximization steps in NLGS algorithms are done with power prices set to zero and precoder prices set to null matrix. This enables us to quantify the gain achievable by the use of pricing for powers ($\mathbf{P}$) and precoders ($\mathbf{Q}$). First partially

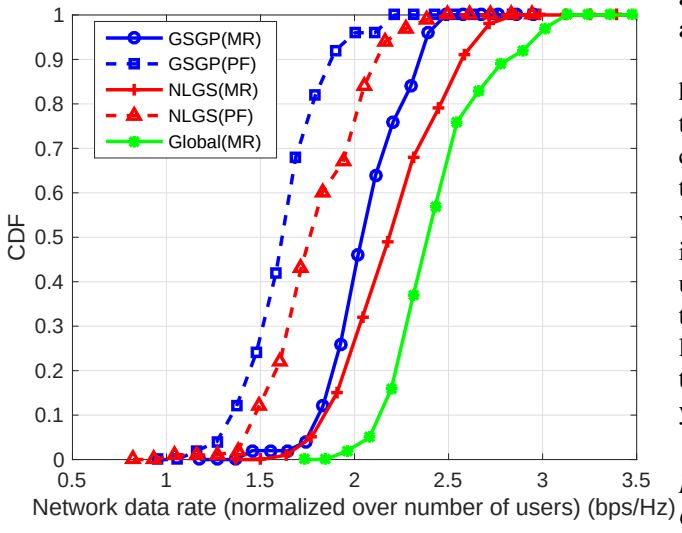

Fig. 5. Comparison of GSGP with NLGS for Max-Rate (solid lines) and PF-Rate (dashed lines) utility functions. Result for global optimum based on multi-start method is also shown. The stepsizes are $\delta^P = \delta^Q = \delta^A = 0.005$ for GSGP algorithm, and $\beta^P = \beta^Q = 0.7$ and $\beta^A = 1$ for NLGS algorithm.

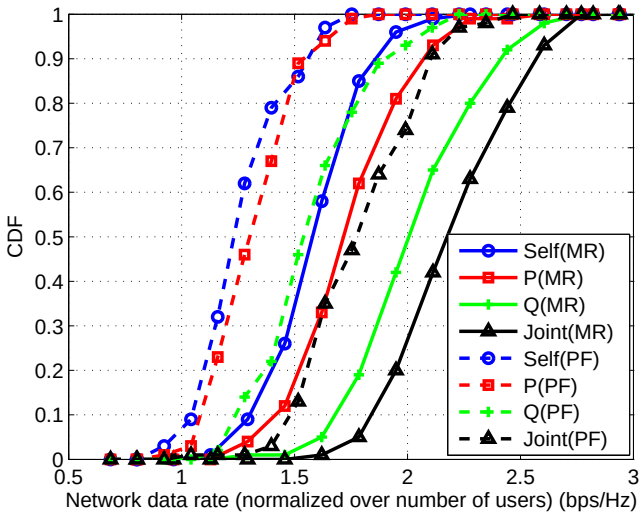

Fig. 6. Comparison of NLGS variants for Max-Rate (solid lines) and PF-Rate (dashed lines) utility functions. CDFs of network data rates normalized over number of users are shown for different cooperative/non-cooperative strategies.

cooperative pricing alternatives such as P(MR) and Q(MR) are considered, in which the pricing is exchanged over one degree of freedom only. Thus, P(MR) means pricing is used for surplus maximization in the power allocation step of NLGS, while the surplus maximization for precoder allocation step is done without prices. Similarly, the converse case is Q(MR), in which cooperation or pricing exchange only over the precoder, where the power allocation step is carried out without pricing. The fully cooperative case is when the resources are optimized jointly (Joint), with prices exchanged over both power and precoders. This variant is the exact NLGS algorithm. All four variants are considered for both utility functions. Note that, in all above cases the scheduling allocation is always carried out, as it is an intra-cell resource allocation and does not requires any pricing exchange.

As shown in the Fig. 6, MR variants achieve (as expected) higher data rates than their PF counterparts. The net gain in terms of network data rate of the pricing (Joint) over the selfish case (Self) is around 40% for both utility functions. In Fig. 7, the complementary CDFs of the user data rates for all the variants are illustrated. The expected advantage of PF-Rate is the increased fairness in terms of data rates of individual users. The difference is more profound in the lower end of the CDF (shown in the inset), where it can be seen that MR leads to a large outage of comprising of more than 50% of the users. For the PF case, the gain that fully cooperative case yields is approximately 100% for the 10-th percentile users.

D. Statistical Analysis of the Gap Between Local and Global Optimum

Due to the non-convex nature of the problem, the initial point of the algorithm has an impact on the solution obtained by the algorithms. A compelling way to improve the performance in such cases consists in adding a preprocessing step

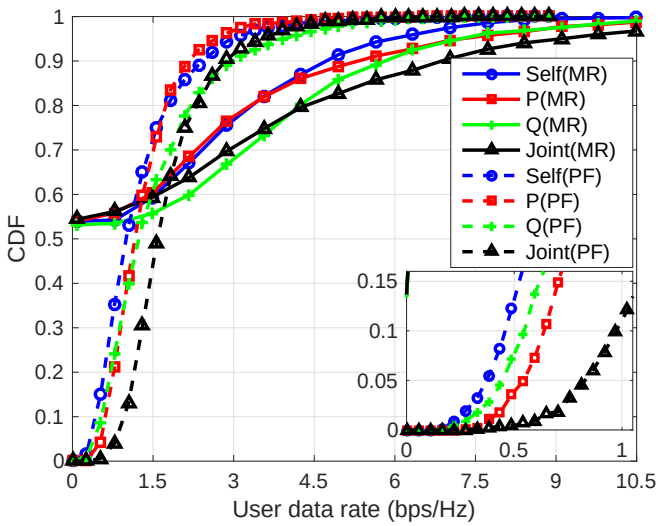

Fig. 7. Comparison of NLGS variants for Max-Rate (solid lines) and PF-Rate (dashed lines) utility functions. CDFs of user data rates are shown for different cooperative/non-cooperative strategies.

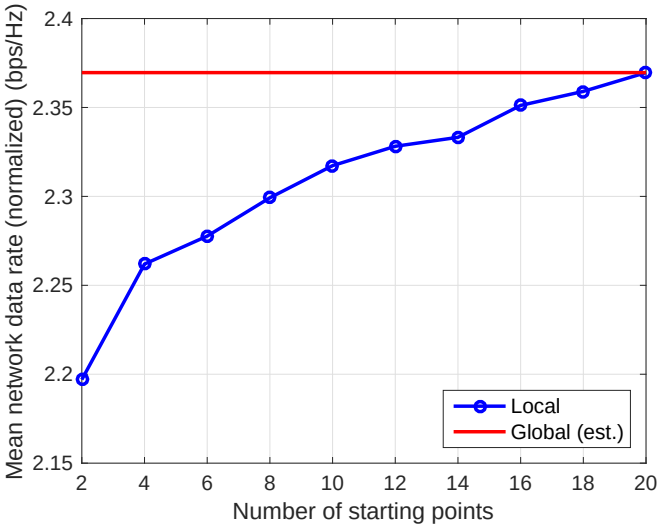

Fig. 8. Comparison of local (multi-start) optima to an estimated global optimum for different number of starting points with NLGS (Max-Rate) algorithm.

that finds a better starting point for the algorithm. In our approach, this is enabled by applying as initialization selfish variants of the GSGP and NLGS algorithms, as illustrated in Fig. 3. These variants do not involve an exchange of pricing and are able to provide an improved starting point without incurring any signaling overhead. In principle, the initialization with a suitable starting point helps to reduce the gap that is observed between obtained solution and the global optimum, as illustrated in Fig. 8. The estimation of the global optimum presented here is obtained using the scatter-search method discussed in Section VI A with all starting points, whereas the local optima are obtained taking into account a subset of these starting points. The results shown in Fig. 8 are obtained after averaging out 50 randomly generated network instances when using the NLGS algorithm with Max Rate. It can be seen that the obtained local optima are very close to the estimated global optimum, and that the difference vanishes rapidly as the number of starting points increase. This increase leads to a higher probability of having a good initial point, thereby improving the odds of getting closer to the global optimum.

VII. Conclusion

This paper has discussed autonomous algorithms for joint resource allocation enabled by an exchange of pricing signals for inter-cell interference mitigation in MIMO SCN scenarios. Optimization parameters include transmit power on carriers, MIMO precoders, and scheduling weights. Distributed pricing algorithms based on solving the multi-dimensional network level optimization problem for MIMO SCN scenarios were derived, and their performance has been analyzed. The algorithms work by iteratively increasing the total network utility by optimizing the resource allocation, thereby improving the data rates of users served by the SCN, subject to an underlying α-fairness criterion. A nonlinear algorithm (based on NLGS), where the effects of resource reallocation in the own cell is fully taken into account, is observed to significantly outperform its gradient projection counterpart (based on GSGP) in terms of both total utility and convergence probability. Both algorithms are distributed, and involve simple updates and pricing exchange. In simulations it is seen that the overall gain from adding message passing to a fully self-organized method, where all cells act selfishly, is dramatic. Furthermore, the combined use of power and precoder prices results in significant gains compared to the case where only power or only precoder prices are exchanged, the joint gain being larger than the sum of the two separate gains. Fairness among users can be ensured by selecting an appropriate utility function that enables network level fairness, in the overall allocation of both inter- and intra-cell resources. Rapid convergence to a (local) optimum along with the scalability features underscore the potential of NLGS for practical systems such as dense self-organizing SCNs.

Appendix: Convergence Proofs

Preliminaries: First, we observe that step 4 in both resource allocation algorithms (i.e. GSGP and NLGS), where BS i updates its internal scheduling weights, is a convex optimization step. The prices of other cells do not affect the dependency of the sum utility or the surplus function on $\mathbf{A}_i$. Utility singularities are at the boundaries of the scheduling weight configuration space, in configurations where at least one of the users is not scheduled at all. To guarantee convergence, we redefine the set of scheduling weights $\mathcal{A}_i$ of (22) so that open neighborhoods of the singular points are removed.

On the other hand, the power and precoder variables of BS i affect the rates of users in i, and the rates experienced at its neighbors $j \neq i$. When users in neighboring cells j have a finite received own-cell power, the rates at the neighbors are finite for all power and precoder allocations in cell i, as a

consequence of the power and the precoder trace constraints. Thus changing power and/or precoder variables of cell i does not cause singularities in other cells. Singularities in own cell utilities may occur when some of the power variables $\mathbf{p}_i$ vanish, or when precoding happens towards a null-space of a user, which is possible if $\mathrm{rank}\left(\mathbf{H}_{i,k}^{(m)}\right) < N_{\mathrm{T}}$. These singularities can be avoided by redefining $\mathcal{P}$ of (20) and $\mathcal{Q}$ of (21). When proving the convergence of NLGS, we shall need convexity of the feasible sets. To remove utility singularities due to precoding, keeping convexity, we restrict certain precoders to have full rank. Feasible domains of the variables, where utility singularities have been carved out, are defined as follows.

Lemma 1: In each cell, on at least one carrier, take scheduling weights from $\tilde{\mathcal{A}}_i^{\mathrm{tot}} = \tilde{\mathcal{A}}_i^{(1)} \times \ldots \times \tilde{\mathcal{A}}_i^{(M)}$, where each user $k \in \mathcal{K}_i$ is guaranteed a non-zero scheduling weight $a_{i,k}^{(m)} \geq \epsilon_a > 0$ on at least one carrier. The set of carriers in cell i, where at least one user has a guaranteed non-zero scheduling weight is $\mathcal{M}_i^{\mathrm{gnz}}$. Take powers from $\tilde{\mathcal{P}}_i$ where a minimum power $p_i^{(m)} \geq \epsilon_p > 0$ is guaranteed on each carrier in $\mathcal{M}_i^{\mathrm{gnz}}$. Further, on each carrier in $\mathcal{M}_i^{\mathrm{gnz}}$ take covariance matrices from $\tilde{\mathcal{Q}} = \left\{\epsilon_q \mathbf{I} + \tilde{\mathbf{Q}} : \mathrm{Tr}\tilde{\mathbf{Q}} = 1 - N_{\mathrm{T}}\epsilon_q, \tilde{\mathbf{Q}} \succeq 0\right\}$ with $0 < \epsilon_q < 1/N_{\mathrm{T}}$. The defined set of feasible values in all cells is compact and convex. Then, all users have a non-vanishing rate in this domain, U_{sum} is twice continuously differentiable with respect to all variables, and its gradient has finite norm for all α.

Proof: All users are scheduled on a carrier with a full-rank precoder, and accordingly have a strictly positive rate. Accordingly, U_{sum} is twice continuously differentiable in the feasible set, and gradients are finite. The set $\tilde{\mathcal{Q}}$ is a translation and dilatation of the compact convex set $\mathcal{Q}$ and thus compact and convex. $\tilde{\mathcal{P}}_i$ and $\tilde{\mathcal{A}}_i$ are compact by definition. They are hypercubes sliced by hyperplanes, just as $\mathcal{P}_i$ and $\mathcal{A}_i$, and thus convex. ∎

Note that more than simple ϵ_a-neighborhoods of utility singularities have been removed from $\tilde{\mathcal{A}}_i$. Removing only ϵ_a-neighborhoods would render $\tilde{\mathcal{A}}_i$ non-convex. The multi-carrier precoding space in cell i is the direct product of $\mathcal{Q}$ for each carrier not in $\mathcal{M}_i^{\mathrm{gnz}}$ and of $\tilde{\mathcal{Q}}$ for each carrier in $\mathcal{M}_i^{\mathrm{gnz}}$. We call this space $\tilde{\mathcal{Q}}_i^{\mathrm{tot}}$. We call the points in $\tilde{\mathcal{A}}_i^{\mathrm{tot}}, \tilde{\mathcal{P}}_i$ and $\tilde{\mathcal{Q}}_i^{\mathrm{tot}}$, where one of the coordinates reach the value $\epsilon_a, \epsilon_p, \epsilon_q$, respectively, *interior boundaries*. The boundaries shared with $\mathcal{A}_i^{\mathrm{tot}}, \mathcal{P}_i$ and $\mathcal{Q}_i^{\mathrm{tot}}$, induced by the power, precoding and scheduling weight constraints, are *exterior boundaries*. The next step is to formulate stepsize control mechanisms that make GSGP and NLGS ascent algorithms. We concentrate on the update of cell i. The variables controlled by a cell are the scheduling weights $\mathbf{A}_i \in \tilde{\mathcal{A}}_i^{\mathrm{tot}}$, affecting only own cell utility, and the variables $\mathbf{p}_i \in \tilde{\mathcal{P}}_i$, and $\mathbf{Q}_i \in \tilde{\mathcal{Q}}_i^{\mathrm{tot}}$, affecting both own-cell utility and inter-cell interference. The domains are as defined in Lemma 1. An arbitrary one of the three sets of variables controlled by i is denoted by $\mathbf{x}_i$, taking values in $\mathcal{X}_i$. All other variables controlled by cell i and cells $j \neq i$ are grouped to $\mathbf{y}_i \in \mathcal{Y}_i$. We start by considering utility sums $U_{\mathcal{S}} = \sum_{j\in\mathcal{S}} U_j$ of arbitrary subsets $\mathcal{S}$ of cells. All U_j are functions of $\mathbf{x}_i$ and $\mathbf{y}_i$, but we leave the latter dependence implicit in notations. The operator norm induced by the l_2 vector norm of a matrix $\mathbf{M}$, i.e. the largest singular value of $\mathbf{M}$ (see e.g. [40, Theorem 5.6.2]), is denoted by $\|\mathbf{M}\|_{\mathrm{op}}$.

Lemma 2: The operator norm $\left\|\nabla^2_{\mathbf{x}_i} U_{\mathcal{S}}\right\|_{\mathrm{op}}$ of the Hessian of $U_{\mathcal{S}}$ is bounded from above by a finite L_i, when $\mathbf{x}_i \in \mathcal{X}_i$ and $\mathbf{y}_i \in \mathcal{Y}_i$. The gradient is locally Lipschitz continuous with local Lipschitz constant $L_{\mathcal{S},\,i}$; then we have

$$\left|U_{\mathcal{S}}(\mathbf{x}_2) - U_{\mathcal{S}}(\mathbf{x}_1) - (\mathbf{x}_2 - \mathbf{x}_1)^{\mathrm{H}} \nabla_{\mathbf{x}_i} U_{\mathcal{S}}(\mathbf{x}_1)\right| \leq \frac{L_{\mathcal{S},\,i}}{2} \|\mathbf{x}_2 - \mathbf{x}_1\|^2 \tag{34}$$

and

$$\|\nabla_{\mathbf{x}_i} U_{\mathcal{S}}(\mathbf{x}_2) - \nabla_{\mathbf{x}_i} U_{\mathcal{S}}(\mathbf{x}_1)\| \leq L_{\mathcal{S},\,i} \|\mathbf{x}_2 - \mathbf{x}_1\| \tag{35}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_i$ and $\mathbf{y}_i \in \mathcal{Y}_i$.

Proof: First consider the utility $U_j(\mathbf{x}_i)$ in cell $j \in \mathcal{S}$ as a function of $\mathbf{x}_i$. According to Lemma 1, the utility $U_j(\mathbf{x}_i)$ is twice continuously differentiable, and elements in the Hessian $\nabla^2_{\mathbf{x}_i} U_j$ are finite for all $\mathbf{x}_i \in \mathcal{X}_i$ and $\mathbf{y} \in \mathcal{Y}$. The singular values of a matrix with finite entries is finite. Both $\mathcal{X}_i$ and $\mathcal{Y}$ are compact. As a function of $\mathbf{x}_i$ and $\mathbf{y}_i$, the largest singular value of $\nabla^2_{\mathbf{x}_i} U_j$ is a scalar function on a compact domain. According to the extreme value theorem, it has a finite maximum value L_{ji} on this domain. Then, a bound on the operator norm of the sum Hessian $\nabla^2_{\mathbf{x}_i} U_{\mathcal{S}} = \sum_{j\in\mathcal{S}} \nabla^2_{\mathbf{x}_i} U_j(\mathbf{x}_i)$, directly follows from the triangle inequality: $\left\|\nabla^2_{\mathbf{x}_i} U_{\mathcal{S}}\right\|_{\mathrm{op}} \leq \sum_{j\in\mathcal{S}} L_{ji} \equiv L_{\mathcal{S},\,i}$. The local Lipschitz continuity of the gradient conventionally follows from Taylor's theorem. ∎

Proof of Proposition 1: We use Lemma 2 for $U_{\mathcal{S}} = U_{\mathrm{sum}}$, and for $\mathbf{x}_i$ being $\mathbf{A}_i, \mathbf{p}_i, \mathbf{Q}_i$, respectively, to get local Lipschitz constants L_i^A, L_i^p, L_i^Q. Taking $L^x = \max_i L_i^x$, we get bounds on stepsizes $0 < \delta^x < \frac{2}{L^x}$ for $x = A, p, Q$. Finite stepsizes exist as a consequence of finiteness of the L_i. Convergence then follows from a conventional ascent and Cauchy sequence argument [27, Section 3.3.2]. Compactness guarantees that limit points are within the feasible domain. ∎

Note that Proposition 1 does not guarantee that upon convergence, the gradient projected to the original feasible sets $\mathcal{A}_i^{\mathrm{tot}}, \mathcal{P}_i, \mathcal{Q}_i^{\mathrm{tot}}$ would vanish. If convergence is to a point on an interior boundary, some users may have unnecessarily non-zero weights on some carriers, some carriers may have unnecessarily non-zero powers, and/or some precoders may have unnecessarily high rank. Improved algorithms that increase the feasible sets can be devised, if convergence is to an interior boundary created according to Lemma 1. The regularization constants ϵ and, accordingly, the stepsizes would be changed online. Such stepsize control would guarantee convergence in the original feasible sets.

Proof of Proposition 2: We consider a NLGS update, where a variable $\mathbf{x}_i$ controlled by cell i is updated. For simplicity, we suppress the index i from $\mathbf{x}_i$. The update is

$$\mathbf{x}_{\mathrm{u}}(\beta) = \beta \mathbf{x}^* + (1 - \beta)\mathbf{x}_0 \ , \tag{36}$$

where $\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathcal{X}_i} s_i(\mathbf{x})$, and $\mathbf{x}_0 \in \mathcal{X}_i$ is the value of the variable before the update. As a consequence of convexity of the feasible set, proven in Lemma 1, $\mathbf{x}_{\mathrm{u}} \in \mathcal{X}_i$ for all $0 \leq \beta \leq 1$. The network utility (6) is divided into two parts,

$U_{\text{sum}} = U_i(\mathbf{x}) + U_{\neg i}(\mathbf{x})$, where $U_{\neg i}(\mathbf{x}) = \sum_{j \neq i} U_j(\mathbf{x})$. Cell i has complete information about $U_i(\mathbf{x})$, but only linearized information of $U_{\neg i}(\mathbf{x})$, except when $\mathbf{x}$ is $\mathbf{A}_i$ and $U_{\neg i}$ does not depend on $\mathbf{x}$ at all. As a sum of concave functions, the own-cell utility $U_i(\mathbf{x})$ is a concave function in $\mathbf{x}$. We use Lemma 2 to bound the effect the update in $\mathbf{x}$ has on $U_{\neg i}(\mathbf{x})$. Applying Taylor's theorem to $U_{\neg i}(\mathbf{x})$, we have

$$U_{\text{sum}}\left(\mathbf{x}_{\text{u}}\right) - U_{\text{sum}}\left(\mathbf{x}_0\right) = s_i\left(\mathbf{x}_{\text{u}}\right) - s_i\left(\mathbf{x}_0\right) + R_{\neg i}, \quad (37)$$

where s_i is the surplus function of cell i corresponding to $\mathbf{x}$, and the remainder is bounded according to (34) as $|R_{\neg i}| \leq \frac{L_{\neg i}}{2} \left\|\mathbf{x}_{\text{u}} - \mathbf{x}_0\right\|^2 = \frac{\beta^2 L_{\neg i}}{2} \left\|\mathbf{x}^* - \mathbf{x}_0\right\|^2$, where $L_{\neg i}$ the local Lipschitz constant of $\nabla_{\mathbf{x}} U_{\neg i}(\mathbf{x})$. To guarantee that the update (37) is an ascent, we thus require

$$s_i\left(\mathbf{x}_{\text{u}}(\beta)\right) - s_i\left(\mathbf{x}_0\right) \geq \frac{\beta^2 L_{\neg i}}{2} \left\|\mathbf{x}^* - \mathbf{x}_0\right\|^2 \ . \quad (38)$$

Fastest convergence is achieved by selecting the largest β fulfilling this condition. If $\mathbf{x}_0 \neq \mathbf{x}^*$, a strictly positive β exists. This can be seen by applying Taylor's theorem to $U_i(\mathbf{x})$ in s_i:

$$s_i(\mathbf{x}_{\text{u}}) - s_i(\mathbf{x}_0) = \left(\mathbf{x}_{\text{u}} - \mathbf{x}_0\right)^{\text{H}} \nabla_{\mathbf{x}} s_i(\mathbf{x}_0) + R_i \geq 0. \quad (39)$$

The inequality follows from the definition of $\mathbf{x}^*$, and concavity of $U_i(\mathbf{x})$. As $U_i(\mathbf{x})$ is concave, we have $R_i \leq 0$, and $\left(\mathbf{x}_{\text{u}} - \mathbf{x}_0\right)^{\text{H}} \nabla_{\mathbf{x}} s_i(\mathbf{x}_0) \geq |R_i| \geq 0$. The remainder is bounded according to (34) as $|R_i| \leq \frac{L_i}{2} \left\|\mathbf{x}_{\text{u}} - \mathbf{x}_0\right\|^2$, where L_i is the local Lipschitz constant of $\nabla_{\mathbf{x}} U_i(\mathbf{x})$. Applying the linear bound (39) to (38) and recalling that $\mathbf{x}_{\text{u}} - \mathbf{x}_0 = \beta(\mathbf{x}^* - \mathbf{x}_0)$, we find that ascent is guaranteed at least for

$$\beta \leq \frac{2}{L_i + L_{\neg i}} \cdot \frac{\left(\mathbf{x}^* - \mathbf{x}_0\right)^{\text{H}} \nabla_{\mathbf{x}} s_i(\mathbf{x}_0)}{\left\|\mathbf{x}^* - \mathbf{x}_0\right\|^2} \quad (40)$$

which allows for strictly positive β. Convergence then follows from conventional Cauchy sequence argument. Finding the largest β that fulfills the inequality (38) or (40) for all initial points $\mathbf{x}_0$ and all feasible interference prices in s_i, and for all cells i, yields a global β. For scheduling weight optimization, $\beta^A = 1$ can be chosen, as scheduling does not affect other cells. ■

Note that as in the above proof of Proposition 1, convergence may be to an interior boundary. In such a case, the feasible sets $\tilde{\mathcal{A}}_i^{\text{tot}}, \tilde{\mathcal{P}}_i, \tilde{\mathcal{Q}}_i^{\text{tot}}$ can be replaced by some other neighborhoods of the converged point. Such changes of feasible sets, similarly to searches for an update-specific β, are purely cell-internal things, which do not require communication.

REFERENCES

[1] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Distributed resource allocation schemes," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 53–63, Sept. 2009.

[2] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas in Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.

[3] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 41–48, Sept. 2009.

[4] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, June 2011.

[5] C. Kosta, B. Hunt, A. Quddus, and R. Tafazolli, "On interference avoidance through inter-cell interference coordination (ICIC) based on OFDMA mobile systems," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 3, pp. 973–995, 3Q 2013.

[6] O. Aliu, A. Imran, M. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Commun. Surveys & Tutorials*, vol. 15, no. 1, pp. 336–361, Jan. 2013.

[7] F. Ahmed, O. Tirkkonen, M. Peltomäki, J.-M. Koljonen, C.-H. Yu, and M. Alava, "Distributed graph coloring for self-organization in LTE networks," *J. Elec. Computer Engineering*, pp. 1–10, 2010.

[8] F. Ahmed and O. Tirkkonen, "Simulated annealing variants for self-organized resource allocation in small cell networks," *J. Applied Soft Computing*, vol. 38, pp. 762–770, 2016.

[9] M. Khawer, J. Tang, and F. Han, "usICIC–a proactive small cell interference mitigation strategy for improving spectral efficiency of LTE networks in the unlicensed spectrum," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2303–2311, Mar. 2016.

[10] M. Simsek, M. Bennis, and I. Güvenc, "Learning based frequency- and time-domain inter-cell interference coordination in HetNets," *IEEE Trans. Veh. Tech.*, vol. 64, no. 10, pp. 4589–4602, Oct. 2015.

[11] F. Gaaloul, R. Radaydeh, and M. Alouini, "Performance improvement of switched-based interference mitigation for channel assignment in overloaded small-cell networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2091–2103, May 2013.

[12] L. Garcia, I. Kovacs, K. Pedersen, G. Costa, and P. Mogensen, "Autonomous component carrier selection for 4G femtocells: A fresh look at an old problem," *IEEE J. Sel. Areas in Commun.*, vol. 30, no. 3, pp. 525–537, Apr. 2012.

[13] J. Zheng, Y. Cai, and A. Anpalagan, "A stochastic game-theoretic approach for interference mitigation in small cell networks," *IEEE Commun. Letters*, vol. 19, no. 2, pp. 251–254, Feb. 2015.

[14] M. Bennis, S. Perlaza, P. Blasco, Z. Han, and H. Poor, "Self-organization in small cell networks: A reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3202–3212, July 2013.

[15] M. Kasparick and G. Wunder, "Autonomous algorithms for centralized and distributed interference coordination: a virtual layer-based approach," *EURASIP J. Wireless Commun. and Netw.*, vol. 2014, no. 1, pp. 1–20, July 2014.

[16] F. Ahmed, A. A. Dowhuszko, O. Tirkkonen, and R. Berry, "A distributed algorithm for network power minimization in multicarrier systems," in *Proc. IEEE Int'l. Symp. Personal Indoor and Mobile Radio Commun.*, Sep. 2013, pp. 1914–1918.

[17] F. Ahmed and O. Tirkkonen, "Local optimum based power allocation approach for spectrum sharing in unlicensed bands," in *Self-Organizing Systems*, ser. Lecture Notes in Computer Science, T. Spyropoulos and K. Hummel, Eds. Springer Berlin / Heidelberg, 2009, pp. 238–243.

[18] A. Liu, V. Lau, L. Ruan, J. Chen, and D. Xiao, "Hierarchical radio resource optimization for heterogeneous networks with enhanced inter-cell interference coordination (eICIC)," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1684–1693, Apr. 2014.

[19] S. Deb, P. Monogioudis, J. Miernik, and J. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 137–150, Feb. 2014.

[20] M. Feng, T. Jiang, D. Chen, and S. Mao, "Cooperative small cell networks: high capacity for hotspots with interference mitigation," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 108–116, Dec. 2014.

[21] C. Yang, J. Li, A. Anpalagan, and M. Guizani, "Joint power coordination for spectral-and-energy efficiency in heterogeneous small cell networks: A bargaining game-theoretic perspective," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1364–1376, Feb. 2016.

[22] A. Sediq, R. Schoenen, H. Yanikomeroglu, and G. Senarath, "Optimized distributed inter-cell interference coordination (ICIC) scheme using projected subgradient and network flow optimization," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 107–124, Jan. 2015.

[23] J. Zheng, Y. Cai, Y. Xu, and A. Anpalagan, "Distributed channel selection for interference mitigation in dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Veh. Tech.*, vol. 63, no. 9, pp. 4757–4762, Nov. 2014.

[24] J. Zheng, Y. Cai, Y. Liu, Y. Xu, B. Duan, and X. Shen, "Optimal power allocation and user scheduling in multicell networks: Base station cooperation using a game-theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6928–6942, Dec. 2014.

[25] J. Xu, J. Wang, Y. Zhu, Y. Yang, X. Zheng, S. Wang, L. Liu, K. Horneman, and Y. Teng, "Cooperative distributed optimization for the hyper-dense small cell deployment," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 61–67, May 2014.

[26] F. Ahmed, A. A. Dowhuszko, and O. Tirkkonen, "Distributed algorithm for downlink resource allocation in multicarrier small cell networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2012, pp. 6802–6808.

[27] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA: Athena Scientific, 1997.

[28] 3GPP, "Evaluation assumptions for small cell enhancements - physical layer," Tech. Rep. R1-130856, 2013.

[29] R. Kwan, C. Leung, and J. Zhang, "Multiuser scheduling on the downlink of an LTE cellular system," *Res. Letters Commun.*, vol. 2008, pp. 31–34, Jan. 2008.

[30] L. Li, M. Pal, and Y. Yang, "Proportional fairness in multi-rate wireless LANs," in *Proc. IEEE Conf. Computer Commun.*, Apr. 2008.

[31] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," in *Proc. IEEE Conf. Computer Commun.*, Apr. 2009, pp. 2611–2615.

[32] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.

[33] C. Shi, R. Berry, and M. Honig, "Distributed interference pricing for OFDM wireless networks with non-separable utilities," in *Proc. Conf. Inform. Sciences and Systems*, Mar. 2008, pp. 755–760.

[34] J. Huang, R. Berry, and M. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas in Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[35] S. Ye and R. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.

[36] Z. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[37] J.-W. Cho, J. Mo, and S. Chong, "Joint network-wide opportunistic scheduling and power control in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1520–1531, Mar. 2009.

[38] Y. Zhao, S. Diggavi, A. Goldsmith, and H. Poor, "Convex optimization for precoder design in MIMO interference networks," in *Proceedings IEEE Annual Allerton Conf. Commun., Control, and Computing*, Oct. 2012, pp. 1213–1219.

[39] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, "Scatter search and local NLP solvers: A multistart framework for global optimization," *INFORMS J. Computing*, vol. 19, no. 3, pp. 328–340, Jul. 2007.

[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 2012.