# PARALLEL AND DISTRIBUTED METHODS FOR NONCONVEX OPTIMIZATION

*G. Scutari,*[1] *F. Facchinei,*[2] *L. Lampariello,*[2] *and P. Song*[1]

[1]Dpt. of Electrical Eng., State University of New York (SUNY) at Buffalo, Buffalo, NY 14260, USA.

[2]Dpt. of Computer, Control, and Management Eng., University of Rome La Sapienza, Roma, Italy.

Emails: <gesualdo,peiranso>@buffalo.edu, <facchinei,lampariello>@dis.uniroma1.it.

## ABSTRACT

We propose a general algorithmic framework for the minimization of a nonconvex smooth function subject to nonconvex smooth constraints. The algorithm solves a sequence of *(separable) strongly convex* problems. Convergence to a stationary solution of the original nonconvex optimization is established. Our framework is very general and flexible; it unifies several existing Successive Convex Approximation (SCA)-based algorithms such as (proximal) gradient or Newton type methods, block coordinate (parallel) descent schemes, difference of convex functions methods, and improves on their convergence properties. More importantly, and differently from current SCA schemes, it naturally leads to *distributed and parallelizable* schemes for a large class of nonconvex problems. The new method is applied to the solution of a new rate profile optimization problem over Interference Broadcast Channels (IBCs); numerical results show that it outperforms existing ad-hoc algorithms.

***Index Terms—*** Nonconvex problems, Parallel & distributed optimization, Successive convex approximation.

## 1. INTRODUCTION

The minimization of a nonconvex function $U$ subject to some convex constraints $\mathcal{K}$ and nonconvex ones $g_j(\mathbf{x}) \leq 0$

$$\left.\begin{array}{ll}\min\limits_{\mathbf{x}} & U(\mathbf{x}) \\ \text{s. t.} & g_j(\mathbf{x}) \leq 0,\ j=1,\ldots,m \\ & \mathbf{x} \in \mathcal{K},\end{array}\right\} \triangleq \mathcal{X}, \tag{P}$$

with $U : \mathcal{K} \to \mathbb{R}$ and $g_j : \mathcal{K} \to \mathbb{R}$ smooth, is an ubiquitous problem that arises in many fields, ranging from signal processing to communication, networking, machine learning, etc.

It is hardly possible here to even summarize the huge amount of solution methods that have been proposed for Problem (P). Our focus in this paper is on *distributed* algorithms converging to stationary solutions of (P) while *preserving the feasibility of the iterates*. While the former feature needs no further comment, the latter is motivated by several reasons. First, in many cases the objective function $U$ is not even defined outside the feasible set; second, in some applications one may have to interrupt calculations before a solution has been reached and it is then important that the current iterate is feasible; and third, in on-line implementations it is mandatory that some constraints are satisfied by every iterate (e.g., think of power budget or interference constraints). As far as we are aware of, there exists no method for the solution of (1) in its full generality that is both feasible *and* distributed.

Existing efforts pursuing the above design criteria can be roughly divided into three main categories: 1) Feasible Sequential Quadratic Programming (FSQP) methods (e.g., [1]); 2) Parallel Variable Distribution (PVD) schemes (e.g., [2, 3, 4]); and 3) SCA algorithms (in the spirit of [5, 6, 7, 8, 9, 10]). FSQP methods [1] mantain feasibility throughout the iterations, but are centralized and computationally expensive. PVD schemes are suitable for implementation over parallel architectures but they require an amount of information exchange/knowledge that is not compatible with a distributed architecture (for example they cannot be applied to the case study discussed in Section 3). Furthermore, when applied to problem (P), they call for the solution of possibly difficult nonconvex (smaller) subproblems; and convergence has been established only for convex [2, 4] or nonconvex but block separable $g_j$'s [3]. Finally, standard SCA methods are centralized [5, 6, 10], with the exception of [8, 9] and some instances of [7] that lead instead to distributed schemes. However convergence conditions have been established only in the case of strongly convex $U$ [6] or convex and separable $g_j$'s [7, 8, 9].

In this paper we propose a new convergent algorithmic framework for the general formulation (P) which on the one hand maintains feasibility and, on the other hand, leads, under very mild additional assumptions, to parallel and distributed solution methods. More specifically, the method solves a sequence of *strongly convex* inner approximations of (P) that, under some mild assumptions, can be solved in a distributed way using standard primal/dual decomposition techniques (e.g., [11, 12]). Additional key features of the proposed method are: i) it includes as special cases several classical SCA-based algorithms, such as (proximal) gradient or Newton type methods, block coordinate (parallel) descent schemes, difference of convex functions methods; ii) our convergence conditions unify and extend to the general class (P) those of current (centralized) SCA methods; and iii) it provides new efficient algorithms for old problems, e.g. power control problems in cellular systems [13, 14, 15, 16]; MIMO relay optimization [17], dynamic spectrum management in DSL systems [18, 19], sum-rate maximization, proportional-fairness and max-min optimization of SISO/MISO/MIMO ad-hoc networks [8, 20, 21, 22, 23, 24, 25], robust optimization of CR networks [26, 27, 28]. As a case study, our method is applied to a novel rate profile maximization problem over IBCs. Numerical results show that the new method compares favorably to existing ad-hoc algorithms.

## 2. MAIN RESULTS

### 2.1. Technical preliminaries and main idea

Consider Problem (P), whose feasible set is denoted by $\mathcal{X}$.

**Assumption 1.** We make the blanket assumptions:

A1) $\mathcal{K} \subseteq \mathbb{R}^n$ is closed and convex;

A2) $U$ and each $g_j$ are continuously differentiable on $\mathcal{K}$;

A3) $\nabla_{\mathbf{x}} U$ is Lipschitz continuous on $\mathcal{K}$ with constant $L_{\nabla U}$.

A4) For some $\mathbf{x}^0 \in \mathcal{X}$, $\{\mathbf{x} \in \mathcal{X} : U(\mathbf{x}) \leq U(\mathbf{x}^0)\}$ is compact.

The assumptions above are quite standard and are satisfied by a large class of problems of practical interest. In particular, condition A4 guarantees that the social problem has a solution, even when the feasible set $\mathcal{X}$ is not bounded. Note that we do not assume convexity of the functions $U$ and $g_1, \ldots, g_m$; w.l.o.g., convex constraints, if present, are accommodated in the set $\mathcal{K}$.

Our goal is to efficiently compute locally optimal solutions of (P), possibly in a distributed way while preserving the feasibility of the iterates. Building on the idea of SCA methods, the proposed approach consists in solving a sequence of *strongly convex inner* approximations of (P) in the form: given $\mathbf{x}^\nu \in \mathcal{X}$,

The work of Scutari and Song was supported by the USA NSF under Grants CMS 1218717 and CAREER Award No. 1254739.

$$\left.\begin{array}{ll}\min\limits_{\mathbf{x}} & \tilde{U}(\mathbf{x};\mathbf{x}^\nu)\\ \text{s.t.} & \tilde{g}_j(\mathbf{x};\mathbf{x}^\nu)\le 0,\ j=1,\dots,m\\ & \mathbf{x}\in\mathcal{K}.\end{array}\right\}\triangleq \mathcal{X}(\mathbf{x}^\nu), \qquad (\mathrm{P}^\nu)$$

where $\tilde{U}(\mathbf{x};\mathbf{x}^\nu)$ and $\tilde{g}_j(\mathbf{x};\mathbf{x}^\nu)$ represent approximations of $U(\mathbf{x})$ and $g_j(\mathbf{x})$ at the current iterate $\mathbf{x}^\nu$, respectively; and $\mathcal{X}(\mathbf{x}^\nu)$ denotes the feasible set. We make the following assumptions; weaker requirements can be found in [29].

**Assumption 2 (On $\tilde{U}$).** Let $\tilde{U}:\mathcal{K}\times\mathcal{X}\to\mathbb{R}$ satisfy the following:
B1) $\tilde{U}(\bullet,\mathbf{y})$ is uniformly strongly convex on $\mathcal{K}$ with constant $c_{\tilde{U}}>0$;
B2) $\nabla_{\mathbf{x}}\tilde{U}(\mathbf{y};\mathbf{y})=\nabla_{\mathbf{x}}U(\mathbf{y})$, for all $\mathbf{y}\in\mathcal{X}$;
B3) $\nabla_{\mathbf{x}}\tilde{U}(\bullet;\bullet)$ is Lipschitz continuous on $\mathcal{K}\times\mathcal{X}$;
where $\nabla_{\mathbf{x}}\tilde{U}(\mathbf{y};\mathbf{y})$ denotes the partial gradient of $\tilde{U}(\mathbf{x};\mathbf{y})$ with respect to $\mathbf{x}$ evaluated at $(\mathbf{y};\mathbf{y})$.

**Assumption 3 (On $\tilde{g}_j$'s).** Let each $\tilde{g}_j:\mathcal{K}\times\mathcal{X}\to\mathbb{R}$ satisfy
C1) $\tilde{g}_j(\bullet,\mathbf{y})$ is convex on $\mathcal{K}$ for all $\mathbf{y}\in\mathcal{X}$;
C2) $\tilde{g}_j(\mathbf{x};\mathbf{x})=g_j(\mathbf{x})$, for all $\mathbf{x}\in\mathcal{X}$;
C3) $g_j(\mathbf{x})\le\tilde{g}_j(\mathbf{x};\mathbf{y})$ for all $\mathbf{x}\in\mathcal{K}$ and $\mathbf{y}\in\mathcal{X}$;
C4) $\tilde{g}_j(\bullet;\bullet)$ is Lipschitz continuous on $\mathcal{K}\times\mathcal{X}$;
C5) $\nabla_{\mathbf{x}}g_j(\mathbf{y})=\nabla_{\mathbf{x}}\tilde{g}_j(\mathbf{y};\mathbf{y})$, for all $\mathbf{y}\in\mathcal{X}$;
C6) $\nabla_{\mathbf{x}}\tilde{g}_j(\bullet;\bullet)$ is continuous on $\mathcal{K}\times\mathcal{X}$;
where $\nabla_{\mathbf{x}}\tilde{g}_j(\mathbf{y},\mathbf{y})$ denotes the partial gradient of $\tilde{g}_j(\mathbf{x},\mathbf{y})$ with respect to $\mathbf{x}$ evaluated at $(\mathbf{y},\mathbf{y})$.

The key assumptions are B1, C1, and C3: B1 and C1 make ($\mathrm{P}^\nu$) strongly convex, whereas C3 guarantees $\mathcal{X}(\mathbf{x}^\nu)\subseteq\mathcal{X}$ (iterate feasibility). The others are technical conditions (easy to be satisfied in practice) ensuring that the approximations have locally the same first order behavior of the original functions. In the next section we provide some examples of approximate functions that automatically satisfy Assumptions 2 and 3. As a final remark, we point out that Assumptions 1 and 2 are in many ways similar *but generally weaker* than those proposed in the literature to solve special cases of problem (P) [5, 6, 7, 8, 9]. For instance, [7, 8, 9] studied the simpler case of convex constraints, and [7] requires the convex approximation $\tilde{U}(\bullet;\mathbf{x}^\nu)$ to be a global upper bound of $\tilde{U}(\bullet)$, whereas we do not. The upper bound condition C3 is assumed also in [5, 6] but, differently from those works, we are able to handle also nonconvex objective functions (rather than only convex ones).

Under the above assumptions, associated with each strongly convex subproblem ($\mathrm{P}^\nu$), we define its unique solution $\hat{\mathbf{x}}(\mathbf{x}^\nu)$ (a function of $\mathbf{x}^\nu$)

$$\hat{\mathbf{x}}(\mathbf{x}^\nu)\triangleq\operatorname*{argmin}_{\mathbf{x}\in\mathcal{X}(\mathbf{x}^\nu)}\tilde{U}(\mathbf{x};\mathbf{x}^\nu). \qquad (1)$$

that will be used to formally introduce our algorithm.

*Regularity conditions*. We conclude this section mentioning certain standard regularity conditions on the stationary points of constrained optimization problems. These conditions are needed in the study of the convergence properties of our method. A stationary point $\bar{\mathbf{x}}$ of (P) (or ($\mathrm{P}^\nu$)) is called *regular* if the Mangasarian-Fromovitz Constraint Qualification (MFCQ) holds at $\bar{\mathbf{x}}$; see [30, Sec. 3.2]. If the feasible set is convex, as it is in ($\mathrm{P}^\nu$), the MFCQ holds if and only if the Slater's Constraints Qualification (CQ) holds [30, Sec. 3.2]. Stronger, but easier to be checked CQs can also be used, see [29] for more details.

### 2.2. Algorithmic framework

We are now ready to introduce a formal description of the proposed solution method for (P), which is given in Algorithm 1. Its convergence properties are stated in Theorem 1; see [29] for the proof and weaker convergence conditions.

**Algorithm 1: Inner SCA Algorithm for (P)**

**Data**: $\gamma^\nu\in(0,1]$, $\mathbf{x}^0\in\mathcal{X}$. Set $\nu=0$.
`(S.1)` : If $\mathbf{x}^\nu$ is a stationary solution of (P): `STOP`.
`(S.2)` : Compute the solution $\hat{\mathbf{x}}(\mathbf{x}^\nu)$ of problem ($\mathrm{P}^\nu$) [cf. (1)].
`(S.3)` : Set $\mathbf{x}^{\nu+1}=\mathbf{x}^\nu+\gamma^\nu(\hat{\mathbf{x}}(\mathbf{x}^\nu)-\mathbf{x}^\nu)$.
`(S.4)` : $\nu\leftarrow\nu+1$ and go to step `(S.1)`.

**Theorem 1** *Given the nonconvex problem (P) under Assumptions 1-3, suppose that one of the two following conditions holds.*
*(a): The step-size $\gamma^\nu$ is such that $0<\inf_\nu\gamma^\nu\le\sup_\nu\gamma^\nu\le\gamma^{\max}\le 1$ and $2c_{\tilde{U}}\ge\gamma^{\max}L_{\nabla U}$.*
*(b): i) $\mathcal{X}$ is compact, ii) $\hat{\mathbf{x}}(\mathbf{y})\in\mathcal{X}(\mathbf{y})$ is regular for every $\mathbf{y}\in\mathcal{X}$; and iii) the step-size $\gamma^\nu$ is such that $\gamma^\nu\in(0,1]$, $\gamma^\nu\to 0$, and $\sum_\nu\gamma^\nu=+\infty$.*
*Then every regular limit point of $\{\mathbf{x}^\nu\}$ is a stationary solution of* (P)*. Furthermore, none of such points is a local maximum of $U$.*

**On the generality of the algorithm**. Algorithm 1 implements a novel family of inner SCA methods for problem (P). Roughly speaking, it consists in solving the sequence of strongly convex problems ($\mathrm{P}^\nu$) wherein the original objective function $U$ is replaced by the strongly convex approximation $\tilde{U}$ and the nonconvex constraints $g_j$'s with the convex upper estimates $\tilde{g}_j$; convex constraints, if any, are kept unaltered. Note that in Step 3 we allow a memory in the update of the iterates $\mathbf{x}^\nu$ (Step 3), in the form of a convex combination via $\gamma^\nu\in(0,1]$. Convergence is guaranteed under mild assumptions that offer a lot of flexibility in the choice of the approximation functions and free parameters [cf. Theorem 1a) and b)], making the proposed scheme appealing for many applications. Note that $\hat{\mathbf{x}}(\mathbf{y})\in\mathcal{X}(\mathbf{y})$ in Theorem 1b) is regular if Slater's CQ holds on $\mathcal{X}(\mathbf{y})$.

**On the approximations $\tilde{g}_j$'s**. As already mentioned, while assumption C3 might look rather elusive, in many practical cases, an upper approximate function for the nonconvex constraints $g_j$'s is close at hand. Some examples of $\tilde{g}_j$ satisfying Assumption 3 (and in particular C3) are given next; many others are discussed in [29].

*Example #1– Nonconvex constraints with DC structure*. Suppose that $g_j$ has a DC structure, that is, $g_j(\mathbf{x})=g_j^+(\mathbf{x})-g_j^-(\mathbf{x})$ is the difference of two convex and continuously differentiable functions $g_j^+$ and $g_j^-$. By linearizing the concave part $-g_j^-$ and keeping the convex part $g_j^+$ unchanged, we obtain the following convex upper approximation of $g_j$: for all $\mathbf{x}\in\mathcal{K}$ and $\mathbf{y}\in\mathcal{X}$,

$$\tilde{g}_j(\mathbf{x},\mathbf{y})\triangleq g_j^+(\mathbf{x})-g_j^-(\mathbf{y})-\nabla_{\mathbf{x}}g_j^-(\mathbf{y})^T(\mathbf{x}-\mathbf{y})\ge g_j(\mathbf{x}). \qquad (2)$$

*Example #2– Nonconvex constraints with Lipschitz gradients*. If the gradient of the nonconvex function $g_j$ is Lipschitz continuous on $\mathcal{K}$ with constant $L_{\nabla g_j}$, the following convex approximation function is a global upper bound of $g_j$: for all $\mathbf{x}\in\mathcal{K}$ and $\mathbf{y}\in\mathcal{X}$,

$$\tilde{g}_j(\mathbf{x},\mathbf{y})\triangleq g_j(\mathbf{y})+\nabla_{\mathbf{x}}g_j(\mathbf{y})^T(\mathbf{x}-\mathbf{y})+\frac{L_{\nabla g_j}}{2}\|\mathbf{x}-\mathbf{y}\|^2\ge g_j(\mathbf{x}). \qquad (3)$$

**On the approximation $\tilde{U}$**. The approximation $\tilde{U}(\mathbf{x};\mathbf{y})$ of $U$ need not be a tight global upper bound of $U(\mathbf{x})$ for every $\mathbf{y}\in\mathcal{X}$ [cf. Assumption 2]. This represents a turning point in the literature of SCA methods [proposed for special cases of (P)] [5, 7]. In fact, conditions on $\tilde{U}$ as in Assumption 2 are relatively weak, enlarging significantly the class of utility functions $U$ which the proposed solution method is applicable to. For instance, all the approximations proposed in [8] can be used also here. Some instances of valid $\tilde{U}$ for specific $U$ are given next; see [29] for more examples.

*Example #3– (Not jointly) convex $U(\mathbf{x}_1,\dots,\mathbf{x}_n)$*. In many applications, the vector of variables $\mathbf{x}$ is partitioned in blocks $\mathbf{x}=(\mathbf{x}_i)_{i=1}^I$

and the function $U$ is convex in each block $\mathbf{x}_i$ separately, but not jointly. A natural approximation for such a $U$ exploring its "partia" convexity is $\tilde{U}(\mathbf{x};\mathbf{y}) = \sum_{i=1}^I \tilde{U}_i(\mathbf{x}_i;\mathbf{y})$, with

$$\tilde{U}_i(\mathbf{x}_i;\mathbf{y}) \triangleq U(\mathbf{x}_i,\mathbf{y}_{-i}) + \frac{\tau_i}{2}(\mathbf{x}_i-\mathbf{y}_i)^T\mathbf{H}_i(\mathbf{y})(\mathbf{x}_i-\mathbf{y}_i), \quad (4)$$

where $\mathbf{y} \triangleq (\mathbf{y}_i)_{i=1}^I$, $\mathbf{y}_{-i} \triangleq (\mathbf{y}_j)_{j\neq i}$, and $\mathbf{H}_i(\mathbf{y})$ is any uniformly positive definite matrix (possibly depended on $\mathbf{y}$). Note that the quadratic term in (4) can be set to zero if $U(\mathbf{x}_i,\mathbf{y}_{-i})$ is strongly convex in $\mathbf{x}_i$, uniformly for all feasible $\mathbf{y}_{-i}$. Mimicking (quasi-) Newton-like schemes, an alternative choice for $\tilde{U}_i(\mathbf{x}_i;\mathbf{y})$ is

$$\begin{aligned}\tilde{U}_i(\mathbf{x}_i;\mathbf{y}) \triangleq{}& \nabla_{\mathbf{x}_i}U(\mathbf{y})^T(\mathbf{x}_i-\mathbf{y}_i)\\ &+\frac{1}{2}(\mathbf{x}_i-\mathbf{y}_i)^T\nabla^2_{\mathbf{x}_i}U(\mathbf{y})(\mathbf{x}_i-\mathbf{y}_i)+\frac{\tau_i}{2}\|\mathbf{x}_i-\mathbf{y}_i\|^2,\end{aligned}$$

where $\nabla^2_{\mathbf{x}_i}U(\mathbf{y})$ is the Hessian of $U$ w.r.t. $\mathbf{x}_i$ evaluated in $\mathbf{y}$. One can also use any positive definite "approximation" of $\nabla^2_{\mathbf{x}_i}U(\mathbf{y})$.

Needless to say, if $U(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ is *jointly* convex in all the block variables, then $\tilde{U}(\mathbf{x};\mathbf{y})$ can be chosen $\tilde{U}(\mathbf{x};\mathbf{y}) \triangleq U(\mathbf{x}) + \sum_i \frac{\tau_i}{2}\|\mathbf{x}_i-\mathbf{y}_i\|^2$, where $\frac{\tau_i}{2}\|\mathbf{x}_i-\mathbf{y}_i\|^2$ is not needed if $U(\mathbf{x}_i,\mathbf{x}_{-i})$ is strongly convex in $\mathbf{x}_i$, uniformly for all feasible $\mathbf{x}_{-i}$.

*Example #4−(Proximal) gradient-like approximations*. If no convexity whatsoever is present in $U$, mimicking proximal-gradient methods, a valid choice of $\tilde{U}$ is the first order approximation of $U$, that is, $\tilde{U}(\mathbf{x};\mathbf{y}) = \sum_{i=1}^I \tilde{U}_i(\mathbf{x}_i;\mathbf{y})$, with each $\tilde{U}_i(\mathbf{x}_i;\mathbf{y}) \triangleq \nabla_{\mathbf{x}_i}U(\mathbf{y})^T(\mathbf{x}_i-\mathbf{y}_i)+\frac{\tau_i}{2}\|\mathbf{x}_i-\mathbf{y}_i\|^2$. Note that even though classical (proximal) gradient descent methods (see, e.g., [31]) share the same approximation function, they are not applicable to problem (P), due to the noncovexity of the feasible set.

*Example #5− Sum-utility function*. In multi-agent scenarios, the objective function $U$ is generally written as $U(\mathbf{x}) \triangleq \sum_{i=1}^I f_i(\mathbf{x}_1,\ldots,\mathbf{x}_I)$, that is, the sum of the utilities $f_i(\mathbf{x}_1,\ldots,\mathbf{x}_I)$ of $I$ agents, each of them controlling the variables $\mathbf{x}_i$. A typical situation is when $f_i$ are convex in some agents' variables. To capture this property, let us define by $\mathcal{S}_i \triangleq \{j=1,\ldots,I : f_j(\bullet,\mathbf{x}_{-i})$ is convex in $\mathbf{x}_i$ $\forall(\mathbf{x}_i,\mathbf{x}_{-i})\in\mathcal{K}\}$ the set of indices of all the functions $f_j(\mathbf{x}_i,\mathbf{x}_{-i})$ that are convex in $\mathbf{x}_i$, for any feasible $\mathbf{x}_{-i}$, and let $\mathcal{C}_i \subseteq \mathcal{S}_i$ be any subset of $\mathcal{S}_i$. Then, the following approximation function $\tilde{U}(\mathbf{x};\mathbf{y})$ satisfies Assumption 2 while exploiting the partial convexity of $U$ (if any): $\tilde{U}(\mathbf{x};\mathbf{y}) = \sum_{i=1}^n \tilde{U}_{\mathcal{C}_i}(\mathbf{x}_i;\mathbf{y})$, with each $\tilde{U}_{\mathcal{C}_i}$ defined as

$$\begin{aligned}\tilde{U}_{\mathcal{C}_i}(\mathbf{x}_i;\mathbf{y}) \triangleq{}& \textstyle\sum_{j\in\mathcal{C}_i} f_j(\mathbf{x}_i,\mathbf{y}_{-i}) + \sum_{k\notin\mathcal{C}_i}\nabla_{\mathbf{x}_i}f_k(\mathbf{y})^T(\mathbf{x}_i-\mathbf{y}_i)\\ &+\tfrac{\tau_i}{2}(\mathbf{x}_i-\mathbf{y}_i)^T\mathbf{H}_i(\mathbf{y})(\mathbf{x}_i-\mathbf{y}_i),\end{aligned}$$

where $\mathbf{H}_i(\mathbf{y})$ is any uniformly positive definite matrix.

**On the choice of the step-size rule**. Theorem 1 states that Algorithm 1 converges either employing a constant step-size rule [case a)] or a diminishing step-size rule [case b)]. Some effective guidelines in the choice of the free parameters can be found in [8]; we omit more details because of space limitation, see also [29].

### 2.3. Distributed implementation

In many applications, e.g., multiuser systems or distributed networks, it is desirable to keep users coordination and communication overhead as low as possible. Here, we briefly discuss how to decompose Algorithm 1; we focus on the following (still very large) subclass of problems (P) allowing for distributed computation.

**Assumption 4 (Decomposabilty)**. Given (P), the following holds:
D1) The set $\mathcal{K}$ has a Cartesian structure, i.e., $\mathcal{K} = \mathcal{K}_1\times\cdots\times\mathcal{K}_I$, with each $\mathcal{K}_i \subset \mathbb{R}^{n_i}$, and $\sum_i n_i = n$; and $\mathbf{x} \triangleq (\mathbf{x}_i)_{i=1}^I$ is partitioned accordingly, with each $\mathbf{x}_i \in \mathcal{K}_i$.
D2) The approximate function $\tilde{U}(\mathbf{x};\mathbf{y})$ satisfying Assumption 2 is chosen so that $\tilde{U}(\mathbf{x};\mathbf{y}) = \sum_i \tilde{U}_i(\mathbf{x}_i;\mathbf{y})$.
D3) The approximate functions $\tilde{g}_j(\mathbf{x};\mathbf{y})$ satisfying Assumption 3 are (block) separable in the $\mathbf{x}$-variables, for any given $\mathbf{y}$, that is, each $\tilde{g}_j(\mathbf{x};\mathbf{y}) = \sum_i \tilde{g}_j^i(\mathbf{x}_i;\mathbf{y})$, for some $\tilde{g}_j^i : \mathcal{K}_i\times\mathcal{X} \to \mathbb{R}$.

Condition D2 still offers many choices for $\tilde{U}$. For instance, any of the $\tilde{U}$ presented in Sec. 2.2 are usable. Some examples where condition D3 can be readily satisfied are:
*−Individual nonconvex constraints*: Each $g_j$ (still nonconvex) depends only on one of the block variables $\mathbf{x}_1,\ldots,\mathbf{x}_I$, i.e, $g_j(\mathbf{x}) = g_j^i(\mathbf{x}_i)$, for some $g_j^i : \mathcal{K}_i \to \mathbb{R}$ and $i$;
*−Separable nonconvex constraints*: Each $g_j$ has the form $g_j(\mathbf{x}) = \sum_i g_j^i(\mathbf{x}_i)$, with $g_j^i : \mathcal{K}_i \to \mathbb{R}$;
*−Nonconvex constraints with Lipschitz gradients*: Each $g_j$ is not necessarily separable but has Lipshitz gradient on $\mathcal{K}$. In this case one can choose, e.g., the approximation $\tilde{g}_j$ as in (3).

Under Assumptions 1−4, each subproblem (P$^\nu$) becomes

$$\begin{aligned}\min_{\mathbf{x}}\quad & \textstyle\sum_{i=1}^I \tilde{U}_i(\mathbf{x}_i;\mathbf{x}^\nu)\\ \text{s.t.}\quad & \textstyle\sum_i \tilde{g}_j^i(\mathbf{x}_i;\mathbf{x}^\nu)\le 0,\ j=1,\ldots,m\\ & \mathbf{x}_i\in\mathcal{K}_i,\quad i=1,\ldots,I.\end{aligned} \qquad (\tilde{P}^\nu)$$

The block separable structure of the objective functions as well as that of the constraints lends itself to a parallel decomposition of the subproblems $(\tilde{P}^\nu)$ in the primal or dual domain, resulting in a distributed implementation of Step 2 of Algorithm 1. For instance, one can use standard primal/dual decomposition techniques [11, 12]. To the best of our knowledge, this is the first attempt to obtain distributed algorithms for (P) in its generality. Because of space limitation, we omit further details and refer to [29]. Here we only observe that, if there are only individual constraints in (P), given $\mathbf{x}^\nu$, each $(\tilde{P}^\nu)$ can be split in $I$ independent subproblems in the variables $\mathbf{x}_i$, even if the original nonconvex $U$ is *not separable*.

## 3. CASE STUDY: MAX-MIN FAIRNESS IN IB NETWORKS

In this section we show how to customize the proposed framework to a novel rate profile maximization problem over IBCs; see [29] for many other applications.

**System model**. Consider a broadcast cellular system composed of $K$ cells; each cell $k\in\mathcal{K}_{\text{BS}} \triangleq \{1,\ldots,K\}$, contains one Base Station (BS) equipped with $T_k$ transmit antennas and serving $I_k$ Mobile Terminals (MTs). We denote by $i_k$ the $i$-th user in cell $k$, equipped with $M_{i_k}$ antennas; the set of users in cell $k$ and the set of all the users are $\mathcal{I}_k \triangleq \{i_k : 0\le i\le I_k\}$ and $\mathcal{I} \triangleq \{i_k : k\in\mathcal{K}_{\text{BS}},\ i_k\in\mathcal{I}_k\}$, respectively. The optimization variables of each BS $k$ are the covariance matrices $\mathbf{Q}_k \triangleq \{\mathbf{Q}_{i_k}\}_{i_k\in\mathcal{I}_k}$ of the signals transmitted to the $I_k$ users in the cell, with each $\mathbf{Q}_{i_k}\in\mathbb{C}^{T_k\times T_k}$ being the covariance matrix of the information symbols of user $i_k$.

The power budget constraint of each BS is $\sum_{i=1}^{I_k}\text{tr}(\mathbf{Q}_{i_k}) \le P_k$. Additional constraints on $\mathbf{Q}_k$, such as interference, per-antenna peak power, null constraints, etc., may also be considered, and will be written in the general form $\mathbf{Q}_k\in\mathcal{Q}_k$, where $\mathcal{Q}_k$ is a given convex and closed set (with nonempty relative interior [32]). The set of all the constraints of each BS $k\in\mathcal{K}_{\text{BS}}$ will be denoted by $\mathcal{Z}_k \triangleq \{\mathbf{Q}_k \triangleq \{\mathbf{Q}_{i_k}\}_{i_k\in\mathcal{I}_k} : \mathbf{Q}_{i_k}\succeq\mathbf{0},\ \mathbf{Q}_k\in\mathcal{Q}_k,\ \sum_{i=1}^{I_k}\text{tr}(\mathbf{Q}_{i_k})\le P_k\}$.

Treating the intra-cell and inter-cell interference at each MT as noise, the maximum achievable rate of each user $i$ in the cell $k$ is

$$R_{i_k}(\mathbf{Q}) \triangleq \log\det\left(\mathbf{I}+\mathbf{H}_{i_kk}\mathbf{Q}_{i_k}\mathbf{H}_{i_kk}^H\bar{\mathbf{R}}_{i_k}(\mathbf{Q}_{-i_k})^{-1}\right) \qquad (5)$$

where $\mathbf{H}_{i_kl}\in\mathbb{C}^{M_{i_k}\times T_l}$ represents the channel matrix between BS $l$ and MT $i_k$; $\bar{\mathbf{R}}_{i_k}(\mathbf{Q}_{-i_k}) \triangleq \mathbf{R}_{w_{i_k}} + \sum_{j\neq i}\mathbf{H}_{i_kk}\mathbf{Q}_{j_k}\mathbf{H}_{i_kk}^H + \sum_{l\neq k}\sum_{j\in\mathcal{I}_l}\mathbf{H}_{i_kl}\mathbf{Q}_{jl}\mathbf{H}_{i_kl}^H$ is the covariance matrix of the Gaus-

sian thermal noise (assumed to be full-rank) plus the intra-cell (second term) and inter-cell (last term) interference; and we used the notation $\mathbf{Q}_{-i_k} \triangleq (\mathbf{Q}_{jl})_{(j,l)\neq(i,k)}$ and $\mathbf{Q} \triangleq \{\mathbf{Q}_{i_k}\}_{i_k\in\mathcal{I}}$.

**Problem formulation**. Providing max-min fairness has long been considered an important design criterion for wireless networks. Here we propose the following rate profile optimization problem: given the profile $(\alpha_{i_k})_{i_k\in\mathcal{I}}$, with each $\alpha_{i_k} > 0$ and $\sum_{i,k}\alpha_{i_k} = 1$, let

$$\begin{array}{ll}\max\limits_{\mathbf{Q}=(\mathbf{Q}_k)_{k\in\mathcal{K}_{\rm BS}},\bar{R}} & \bar{R}\\ \text{s. t.} & \mathbf{Q}_k = (\mathbf{Q}_{i_k})_{i_k\in\mathcal{I}_k} \in \mathcal{Z}_k,\ \forall k\in\mathcal{K}_{\rm BS}\\ & R_{i_k}(\mathbf{Q}) \geq \alpha_{i_k}\bar{R},\quad \forall i_k\in\mathcal{I}.\end{array} \tag{6}$$

Special cases of (6) have already been studied in the literature; [23, 24] studied the rate profile optimization over MISO or *single stream* MIMO ICs (which are simpler models than the IBC); the max-min fairness problem (corresponding to equal $\alpha_{i_k}$) has been recently considered in [25] under standard power constrains (i.e., without the additional constraints $\mathcal{Q}_{i_k}$). The algorithms proposed in the aforementioned works are thus not applicable to solve the more general formulation (6); moreover they are *all centralized*.

**Algorithmic design**. Problem (6) is an instance of (P); we can then readily use Algorithm 1. We briefly discuss next two approximations for the nonconvex constraints $R_{i_k}(\mathbf{Q}) \geq \alpha_{i_k}\bar{R}$, both satisfying Assumption 3 but leading to two alternative SCA algorithms.

*Approximation # 1*. Since each rate function has the DC structure:

$$R_{i_k}(\mathbf{Q}) = f^+_{i_k}(\mathbf{Q}) - f^-_{i_k}(\mathbf{Q}_{-i_k}), \tag{7}$$

with $f^+_{i_k}(\mathbf{Q}) = \log\det\left(\bar{\mathbf{R}}_{i_k}(\mathbf{Q}_{-i_k}) + \mathbf{H}_{i_k k}\mathbf{Q}_{i_k}\mathbf{H}^H_{i_k k}\right)$ and $f^-_{i_k}(\mathbf{Q}_{-i_k}) = \log\det(\bar{\mathbf{R}}_{i_k}(\mathbf{Q}_{-i_k}))$, a natural tight lower bound of $R_{i_k}(\mathbf{Q})$ can be obtained by linearizing $f^-_{i_k}$, which leads to the following rate approximation functions [cf. (2)]: given $\mathbf{Q}^\nu \succeq \mathbf{0}$,

$$R_{i_k}(\mathbf{Q}) \geq \widetilde{R}_{i_k}(\mathbf{Q};\mathbf{Q}^\nu) \triangleq f^+_{i_k}(\mathbf{Q}) - \tilde{f}^-_{i_k}(\mathbf{Q}_{-i_k};\mathbf{Q}^\nu) \tag{8}$$

where, introducing $\langle\mathbf{A},\mathbf{B}\rangle \triangleq \mathrm{Re}\{\mathrm{tr}(\mathbf{A}^H\mathbf{B})\}$, $\tilde{f}^-_{i_k}(\mathbf{Q}_{-i_k})$ is given by

$$\tilde{f}^-_{i_k}(\mathbf{Q}_{-i_k};\mathbf{Q}^\nu) \triangleq f^-_{i_k}(\mathbf{Q}^\nu_{-i_k}) + \sum_{(j,l)\neq(i,k)}\left\langle\mathbf{\Pi}^-_{i_k\,jl}(\mathbf{Q}^\nu_{-i_k}),\mathbf{Q}_{jl}-\mathbf{Q}^\nu_{jl}\right\rangle, \tag{9}$$

with $\mathbf{\Pi}^-_{i_k\,jl}(\mathbf{Q}^\nu_{-i_k}) \triangleq \nabla_{\mathbf{Q}^*_{jl}} f^-_{i_k}(\mathbf{Q}^\nu_{-i_k}) = \mathbf{H}^H_{ikl}\bar{\mathbf{R}}^{-1}_{i_k}(\mathbf{Q}^\nu_{-i_k})\mathbf{H}_{ikl}$ [33]. Using $\widetilde{R}_{i_k}(\mathbf{Q};\mathbf{Q}^\nu)$, the $\nu$-th approximate strongly convex problem (P$^\nu$) becomes: given $\mathbf{X}^\nu \triangleq (\mathbf{Q}^\nu,\bar{R}^\nu)$ and $\tau_{\bar{R}} > 0$,

$$\begin{array}{ll}\hat{\mathbf{X}}(\mathbf{X}^\nu) \triangleq \underset{\mathbf{Q},\bar{R}}{\mathrm{argmax}} & \left\{\bar{R} - \dfrac{\tau_{\bar{R}}}{2}(\bar{R}-\bar{R}^\nu)^2 - \tau_{\mathbf{Q}}\left\|\mathbf{Q}-\mathbf{Q}^\nu\right\|^2\right\}\\ \text{s. t.} & \mathbf{Q}_k = (\mathbf{Q}_{i_k})_{i_k\in\mathcal{I}_k} \in \mathcal{Z}_k,\ \forall k\in\mathcal{K}_{\rm BS}\\ & \widetilde{R}_{i_k}(\mathbf{Q};\mathbf{Q}^\nu) \geq \alpha_{i_k}\bar{R},\quad \forall i_k\in\mathcal{I},\end{array} \tag{$\tilde{P}^\nu$}$$

whose solution $\hat{\mathbf{X}}(\mathbf{X}^\nu)$ can be efficiently computed using conventional optimization packages. Given $\hat{\mathbf{X}}(\mathbf{X}^\nu)$, one can now use Algorithm 1, whose convergence is guaranteed by Theorem 1 (one can check that every limit point of $\{\mathbf{X}^\nu\}$ whose $Q$-part is not all zero is regular). It is important to remark that $\hat{\mathbf{X}}(\mathbf{X}^\nu)$ can also be computed in a distributed way by introducing proper slack (duplicating) variables and hinging on the Alternating Direction Method of Multipliers (ADMM); see [29] for details.

*Approximation # 2*. An alternative *distributed* solution method to the ADMM for (6) can be obtained exploring a different approximation of the rates (7). Invoking the Lipschitz continuity of each $\nabla_{\mathbf{Q}^*_{jl}} f^+_{i_k}(\mathbf{Q})$ with constant $L_{\nabla_{i_k\,jl}}$ (whose explicit expression is omitted, see [29]), the following lower bound holds for each rate [cf. (3)] and given $\mathbf{Q}^\nu \succeq \mathbf{0}$: $R_{i_k}(\mathbf{Q}) \geq \widetilde{g}_{i_k}(\mathbf{Q};\mathbf{Q}^\nu) \triangleq \tilde{f}^+_{i_k}(\mathbf{Q};\mathbf{Q}^\nu) - \tilde{f}^-_{i_k}(\mathbf{Q}_{-i_k};\mathbf{Q}^\nu) - \sum_{j,l} L_{\nabla_{i_k\,jl}}\|\mathbf{Q}_{jl}-\mathbf{Q}^\nu_{jl}\|^2$, where $\tilde{f}^-_{i_k}(\mathbf{Q}_{-i_k};\mathbf{Q}^\nu)$ is defined in (9), and

$$\tilde{f}^+_{i_k}(\mathbf{Q};\mathbf{Q}^\nu) \triangleq f^+_{i_k}(\mathbf{Q}^\nu) + \sum_{j,l}\left\langle\mathbf{\Pi}^+_{i_k\,jl}(\mathbf{Q}^\nu),\mathbf{Q}_{jl}-\mathbf{Q}^\nu_{jl}\right\rangle, \tag{10}$$

with $\mathbf{\Pi}^+_{i_k\,jl}(\mathbf{Q}^\nu) = \mathbf{H}^H_{ikl}\left(\bar{\mathbf{R}}_{i_k}(\mathbf{Q}^\nu_{-i_k}) + \mathbf{H}_{i_k k}\mathbf{Q}^\nu_{i_k}\mathbf{H}^H_{i_k k}\right)^{-1}\mathbf{H}_{i_k l}$.

Using $\widetilde{g}_{i_k}$ we can now define a different sequence of subproblems (P$^\nu$) having the form of ($\tilde{P}^\nu$) but where the approximate functions $\widetilde{R}_{i_k}$ in the constraints are replaced by $\widetilde{g}_{i_k}$. Since $\widetilde{g}_{i_k}$ is separable in the users covariance matrices $(\mathbf{Q}_{i_k})_{i,k}$, this problem can be solved in a distributed way using standard decomposition techniques; we omit further details because of space limitation, see [29].

**Numerical results**. We present now some numerical experiments comparing five different approaches for IBCs, namely: 1) the Max-Min WMMSE [25] for maximizing the minimum rate of the system [a special case of (6)]; 2) our Algorithm 1 based on the best-response $\hat{\mathbf{X}}(\mathbf{X}^\nu)$ in ($\tilde{P}^\nu$); 3) the WMMSE algorithm [34] and the partial linearization-based algorithm (termed SJBR) [8] proposed for the maximization of the system sum-rate; 4) the partial linearization-based algorithm [8] maximizing the geometric mean of the rates (the proportional fairness utility function), termed GSJBR. To allow the comparison, we consider a special case of (6) as in [25]. We simulated a 4 cell IBC with 3 randomly placed active MTs per cell; the BSs and MTs are equipped with 4 antennas. Channels are Rayleigh fading, whose path-loss are generated using the 3GPP(TR 36.814) methodology [35]. We assume white zero-mean Gaussian noise at each receiver, with variance $\sigma^2$, and same power budget $P$ for all the BSs; the SNR is then snr $= P/\sigma^2$. Algorithm 1 is simulated using $\tau_{\bar{R}}\tau_{\mathbf{Q}} = 1\mathrm{e}{-7}$ and the step-size rule $\gamma^\nu = \gamma^{\nu-1}(1-10^{-3}\gamma^{\nu-1})$, with $\gamma^0 = 1$. The same step-size rule is used for SJBR and GSJBR (the former with no proximal regularization). In Fig. 1 we plot the minimum rate versus snr achieved by the aforementioned algorithms. All results are averaged over 300 independent channel/topology realizations. The figures show that our algorithm yields substantially more fair rate allocation in the system than all the other algorithms (i.e., larger minimum rates). As expected, we observed that SJBR and WMMSE achieve higher sum-rates (not reported in the figure) while sacrificing the fairness; indeed Fig. 1 shows that SJBR and WMMSE can shut off some users (the associated minimum rate is zero).

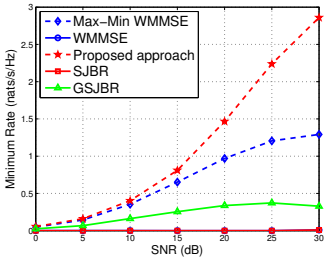


**Fig. 1**. Min rate vs. snr.

## 4. CONCLUSIONS

In this paper, we proposed a general algorithmic framework based on SCA for the solution of nonconvex smooth optimization problems. Some key new features of our scheme are: i) it maintains feasibility *and* leads to *parallel and distributed* solution methods for a very general class of nonconvex problems; ii) it includes as special cases several classical SCA-based algorithms and improves on their convergence properties; and iii) it provides new efficient algorithms also for old problems. Finally, when customized to a novel rate profile maximization problem over IBCs, the algorithm was shown to outperform *ad-hoc* existing schemes.

## 5. REFERENCES


[1] C. Lawrence and A. Tits, "A computationally efficient feasible sequential quadratic programming algorithm," *SIAM Journal on Optimization*, vol. 11, no. 4, 2001.

[2] M. Ferris and O. Mangasarian, "Parallel variable distribution," *SIAM Journal on Optimization*, vol. 4, no. 4, pp. 815–832, 1994.

[3] C. Sagastizábal and M. Solodov, "Parallel variable distribution for constrained optimization," *Computational Optimization and Applications*, vol. 22, no. 1, pp. 111–131, 2002.

[4] M. Solodov, "On the convergence of constrained parallel variable distribution algorithms," *SIAM Journal on Optimization*, vol. 8, no. 1, pp. 187–196, 1998.

[5] B. Marks and G. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[6] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. of Global Optimization*, vol. 47, no. 1, pp. 29–51, 2010. [Online]. Available: http://dx.doi.org/10.1007/s10898-009-9456-5

[7] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[8] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multiuser systems," *IEEE Trans. on Signal Processing*, vol. 63, no. 3, pp. 641–656, Feb. 2014.

[9] A. Alvarado, G. Scutari, and J.-S. Pang, "A new decomposition method for multiuser dc-programming and its applications," *IEEE Trans. on Signal Processing*, (under review), 2013. [Online]. Available: http://arxiv.org/abs/1308.3521.

[10] T. Quoc and M. Diehl, "Sequential convex programming methods for solving nonlinear optimization problems with dc constraints," *ArXiv.org http://adsabs.harvard.edu/abs/2011arXiv1107.5841D*, 2011.

[11] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. on Automatic Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[12] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, 2nd ed. Athena Scientific Press, 1989.

[13] K. Phan, S. Vorobyov, C. Telambura, and T. Le-Ngoc, "Power control for wireless cellular systems via D.C. programming," in *IEEE/SP 14th Workshop on Statistical Signal Processing*, 2007, pp. 507–511.

[14] H. Al-Shatri and T. Weber, "Achieving the maximum sum rate using D.C. programming in cellular networks," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1331–1341, March 2012.

[15] N. Vucic, S. Shi, and M. Schubert, "DC programming approach for resource allocation in wireless networks," in *8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 2010, pp. 380–386.

[16] M. Chiang, C. Tan, D. Palomar, D. O. Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, July 2007.

[17] A. Khabbazibasmenj, F. Roemer, S. Vorobyov, and M. Haardt, "Sum-rate maximization in two-way AF MIMO relaying: Polynomial time solutions to a class of DC programming problems," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5478–5493, 2012.

[18] Y. Xu, T. Le-Ngoc, and S. Panigrahi, "Global concave minimization for optimal spectrum balancing in multi-user DSL networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2875–2885, July 2008.

[19] P. Tsiaflakis, M. Diehl, and M. Moonen, "Distributed spectrum management algorithms for multiuser dsl networks," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4825–4843, Oct. 2008.

[20] D. Schmidt, C. Shi, R. Berry, M. Honig, and W. Utschick, "Distributed resource allocation schemes: Pricing algorithms for power control and beamformer design in interference networks," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 53–63, Sept. 2009.

[21] M. Hong and Z.-Q. Luo, "Signal processing and optimal resource allocation for the interference channel," *Elsevier e-Reference-Signal Processing*, 2013. Available at http://arxiv.org/pdf/1206.5144v1.pdf.

[22] M. Hong, Q. Li, Y.-F. Liu, and Z.-Q. Luo, "Decomposition by successive convex approximation: A unifying approach for linear transceiver design in interfering heterogeneous networks," *arXiv preprint arXiv:1210.1507*, 2012.

[23] R. Mochaourab, P. Cao, and E. Jorswieck, "Alternating rate profile optimization in single stream mimo interference channels," *arXiv preprint*, 2013. [Online]. Available: http://arxiv.org/abs/1303.4683

[24] J. Qiu, R. Zhang, Z.-Q. Luo, and S. Cui, "Optimal distributed beamforming for miso interference channels," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5638–5643, Nov. 2011.

[25] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "Linear transceiver design for a mimo interfering broadcast channel achieving max-min fairness," *Signal Processing*, vol. 93, pp. 3327–3340, Dec. 2013.

[26] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. on Information Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[27] Y. Zhang, E. Dall'Anese, and G. B. Giannakis, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6495–6508, Dec. 2012.

[28] Y. Yang, G. Scutari, and D. Palomar, "Robust mimo cognitive radio under interference temperature constraints," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2465–2482, Nov. 2013.

[29] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and distrubuted methods for nonconvex optimization−Part I&II: Theory & Applications," *IEEE Trans. Signal Process.*, (in preparation). [Online]. Available: Preliminary report at http://www.eng.buffalo.edu/∼gesualdo/FLSTechRep13.pdf.

[30] F. Facchinei and J.-S. Pang, "Exact penalty functions for generalized Nash problems," *G. Di Pillo and M. Roma (eds): Large Scale Nonlinear Optimization*, pp. 115–126, 2006.

[31] D. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 2th Ed., 1999.

[32] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific, 2003.

[33] G. Scutari, F. Facchieni, J.-S. Pang, and D. P. Palomar, "Real and complex monotone communication games," *IEEE Trans. on Information Theory*, (submitted, Nov. 2012). [Online]. Available: http://arxiv.org/abs/1212.6235.

[34] 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 Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sept. 2011.

[35] "3GPP TR 36.814." [Online]. Available: http://www.3gpp.org/ftp/Specs/archive/36_serier/36.814