DECOMPOSITION BY PARTIAL LINEARIZATION IN MULTIUSER SYSTEMS

G. Scutari[1], *F. Facchinei*[2], *P. Song*[1], *D. P. Palomar*[3], *and J.-S. Pang*[4]

[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.

[3]Dpt. of Electronic and Computer Eng., Hong Kong Univ. of Science and Technology, Hong Kong.

[4]Dpt. of Industrial and Enterprise Systems Eng., University of Illinois, Urbana, Illinois 61801, U.S.A.

ABSTRACT

We propose a decomposition framework for the distributed optimization of *general* nonconvex sum-utility functions arising in the design of wireless multi-user interfering systems. Our main contributions are: i) the development of the first provably convergent Jacobi best-response algorithm, where all users simultaneously solve a suitably convexified version of the original sum-utility optimization problem; ii) the derivation of a general dynamic pricing mechanism that provides a unified view of existing pricing schemes that are based, instead, on heuristics; and iii) a framework that can be easily particularized to well-known applications, giving rise to practical algorithms that outperform all existing ad-hoc methods proposed for very specific problems. Our framework contains as special cases well-known gradient algorithms for nonconvex sum-utility problems, and many block-coordinate descents schemes for convex functions.

***Index Terms*—** Nonconvex social problems, parallel & distributed optimization, successive convex approximation.

1. INTRODUCTION

Wireless networks are composed of users that may have different objectives and generate interference, when no multiplexing scheme is imposed to regulate the transmissions; examples are peer-to-peer networks, cognitive radio systems, and ad-hoc networks. A usual and convenient way of designing such multiuser systems is by optimizing the "social function", i.e. the (weighted) sum of the users' objective functions. When the social problem is a sum-separable *convex* programming, many distributed algorithms have been proposed; see, e.g., [1, 2, 3] and references therein. In this paper we address the more frequent and difficult case in which the social function is nonconvex. It is well known that the problem of finding a global minimum of the social function is, in general, an NP hard problem [4], and even centralized solution methods are in jeopardy. As a consequence, most current research efforts have been focused on finding efficiently high quality suboptimal solutions via low complexity (possibly) distributed algorithms. Several *sequential* decomposition algorithms have been proposed in the literature, e.g., [5, 6, 7, 8, 9, 10] wherein only one user at a time is allowed to update his optimization variables; a fact that in large scale networks may lead to excessive communication overhead and slow convergence.

Our aim in this paper is instead to devise more appealing *simultaneous distributed* best-response-like algorithms for *general* nonconvex sum-utility problems, where *all* users can update their variables at the same time. The design of such algorithms is difficult, as witnessed by the scarcity of results in the literature. Besides the application of the classical gradient projection algorithm to the sum-rate maximization problem over MIMO ICs [11], parallel iterative algorithms (with message passing) were proposed in [12, 13] and [14] for DSL/ad-hoc SISO networks and MIMO broadcast interfering channels, respectively. Unfortunately, the gradient schemes [11] suffer from slow convergence; [12, 13] hinge crucially on the special log-structure of the users' rate functions; and [14] is based on the connection with a weighted MMSE problem. This makes [12, 13, 14] not applicable to different sum-utility problems.

Building on the idea first introduced in [15], in this paper, we propose a new decomposition method that: *i)* converges to stationary points of a large class of social problems, encompassing most sum-utility functions of practical interest (including functions of complex variables); *ii)* decomposes well across the users, resulting in the *parallel* solution of *convex* subproblems, one for each user; *iii)* contains as special case the gradient algorithms for nonconvex sum-utility problems, and many block-coordinate descent schemes for convex functions. Moreover, it can be easily particularized to well-known applications, such as [5, 6, 8, 9, 13], giving rise in a unified way to distributed simultaneous algorithms that outperform existing, specialized, methods both theoretically and numerically. We remark that while we follow the seminal ideas put forward in [15], in this paper, we i) consider a much wider class of social-problems and algorithms, including [15] as special cases, ii) discuss in detail the case of functions of complex variables, and iii) compare numerically to state-of-the-art alternative methods.

On the one hand, our approach draws on the Successive Convex Approximation (SCA) paradigm, but relaxes the key requirement that the convex approximation must be a tight global upper bound of the social function, as instead in [10, 12, 16]. On the other hand, the new method also sheds new light on pricing mechanism widely used in the literature: indeed, our method can be viewed as a dynamic pricing algorithm where the pricing rule derives from a deep understanding of the problem characteristics and is not obtained on an ad-hoc basis, as in [5, 6, 17]. We conclude this review by mentioning the recent work [18], where the authors, developing ideas contained in [10, 14], proposed parallel SCA-based schemes that are applicable (only) to sum-utility problems for which a connection with a MMSE formulation can be established. Note that [10, 18], which share some ideas with our approach, appeared after [15].

2. MAIN RESULTS

We consider the design of a multiuser system composed by I coupled users $\mathcal{I} \triangleq \{1, \ldots, I\}$. Each user i makes decisions on his own n_i-dimensional real strategy vector $\mathbf{x}_i$, which belongs to the feasible set $\mathcal{K}_i$; the vector variables of the other users is denoted by $\mathbf{x}_{-i} \triangleq (\mathbf{x}_j)_{j \neq i} \in \mathcal{K}_{-i} \triangleq \prod_{j \neq i} \mathcal{K}_j$; the users' strategy profile is $\mathbf{x} = (\mathbf{x}_i)_{i=1}^{I}$, and the joint strategy set of the users is $\mathcal{K} \triangleq \prod_{j \in \mathcal{I}} \mathcal{K}_j$.

The work of Scutari and Song, Palomar, and Pang is supported by the NSF grant No. CNS-1218717, the HK grant No. RGC 617810, and the U.S. NSF grant No. CMMI 0969600, respectively.

The system design is formulated as:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & U(\mathbf{x}) \triangleq \sum_{i \in \mathcal{I}_f} f_i(\mathbf{x}) \\ \text{subject to} & \mathbf{x}_i \in \mathcal{K}_i, \quad \forall i \in \mathcal{I}, \end{array} \tag{1}$$

with $\mathcal{I}_f \triangleq \{1, \ldots, I_f\}$. Observe that, in principle, the set $\mathcal{I}_f$ of objective functions is different from the set $\mathcal{I}$ of users; we show shortly how to explore this extra degree of freedom to good effect. Of course, (1) contains the most common case where there is exactly one function for each user, i.e. $I = I_f$.

Assumptions. We make the following blanket assumptions:

A1) Each $\mathcal{K}_i$ is closed and convex;

A2) Each f_i is continuously differentiable on $\mathcal{K}$;

A3) $U(\mathbf{x})$ has Lipschitz gradient on $\mathcal{K}$, with constant L_∇;

A4) The lower level set $\mathcal{L}(\mathbf{x}^0) \triangleq \{\mathbf{x} \in \mathcal{K} : U(\mathbf{x}) \leq U(\mathbf{x}^0)\}$ of the social function $U(\mathbf{x})$ is compact for some $\mathbf{x}^0 \in \mathcal{K}$.

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 $\mathcal{K}$ is not bounded; if $\mathcal{K}$ is bounded A4) is trivially satisfied.

At the basis of our decomposition technique there is the attempt to properly exploit any degree of convexity that might be present in the problem (recall that we did not make any convexity assumption so far). To capture this idea, for each user $i \in \mathcal{I}$, let $\mathcal{S}_i \subseteq \mathcal{I}_f$ be the set of indices of all the functions $f_j(\mathbf{x}_i, \mathbf{x}_{-i})$ that are convex in $\mathbf{x}_i \in \mathcal{K}_i$, for any $\mathbf{x}_{-i} \in \mathcal{K}_{-i}$:

$$\mathcal{S}_i \triangleq \{j \in \mathcal{I}_f : f_j(\bullet, \mathbf{x}_{-i}) \text{ is convex on } \mathcal{K}_i, \forall \mathbf{x}_{-i} \in \mathcal{K}_{-i}\}, \tag{2}$$

and let $\mathcal{C}_i \subseteq \mathcal{S}_i$ be a given subset of $\mathcal{S}_i$. Note that we allow the possibility that $\mathcal{S}_i = \emptyset$, even if we "hope" that $\mathcal{S}_i \neq \emptyset$, and actually this latter case occurs in most of the applications of interest, see Sec. 4. For each user $i \in \mathcal{I}$, we can introduce the following convex approximation of $U(\mathbf{x})$ at $\mathbf{x}^n \in \mathcal{K}$:

$$\begin{aligned} \widetilde{f}_{\mathcal{C}_i}(\mathbf{x}_i; \mathbf{x}^n) \quad \triangleq \quad & \sum_{j \in \mathcal{C}_i} f_j(\mathbf{x}_i, \mathbf{x}_{-i}^n) + \boldsymbol{\pi}_{\mathcal{C}_i}(\mathbf{x}^n)^T (\mathbf{x}_i - \mathbf{x}_i^n) \\ & + \frac{\tau_i}{2} \, (\mathbf{x}_i - \mathbf{x}_i^n)^T \, \mathbf{H}_i(\mathbf{x}^n) \, (\mathbf{x}_i - \mathbf{x}_i^n) \end{aligned} \tag{3}$$

where $\boldsymbol{\pi}_{\mathcal{C}_i}(\mathbf{x}^n) \triangleq \sum_{j \in \mathcal{C}_{-i}} \nabla_{\mathbf{x}_i} f_j(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^n}$, with $\mathcal{C}_{-i} \triangleq \mathcal{I}_f \backslash \mathcal{C}_i$ being the complement of $\mathcal{C}_i$; $\tau_i > 0$ is a given positive constant; and $\mathbf{H}_i(\mathbf{x}^n)$ are $n_i \times n_i$ uniformly positive definite matrices (possibly dependent on $\mathbf{x}^n$), i.e. $\mathbf{H}_i(\mathbf{x}^n) - c_{H_i}\mathbf{I} \succeq \mathbf{0}$, for some $c_{H_i} > 0$. Associated with each $\widetilde{f}_{\mathcal{C}_i}(\mathbf{x}_i; \mathbf{x}^n)$, we can introduce the following "best response" map that, under A1)-A4), is well-defined:

$$\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \tau_i) \triangleq \underset{\mathbf{x}_i \in \mathcal{K}_i}{\operatorname{argmin}} \, \widetilde{f}_{\mathcal{C}_i}(\mathbf{x}_i; \mathbf{x}^n). \tag{4}$$

Let $\widehat{\mathbf{x}}_{\mathcal{C}}(\mathbf{x}^n, \boldsymbol{\tau}) \triangleq (\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \tau_i))_{i=1}^I$ be the overall best-response map, with $\boldsymbol{\tau} \triangleq (\tau_i)_{i=1}^I$. We can now introduce our decomposition scheme, Algorithm 1, whose convergence is stated in Theorem 1; we omit the proof because of the space limitation; see [19].

Algorithm 1: Exact Jacobi SCA Algorithm

Data : $\boldsymbol{\tau} > \mathbf{0}$, $\{\gamma^n\} > 0$, $\mathbf{x}^0 \in \mathcal{K}$.

`(S.0)` : Set $n = 0$;

`(S.1)` : If $\mathbf{x}^n$ satisfies a termination criterion: STOP;

`(S.2)` : For all $i \in \mathcal{I}$, compute $\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \boldsymbol{\tau})$ according to (4);

`(S.3)` : Set $\mathbf{x}^{n+1} \triangleq \mathbf{x}^n + \gamma^n (\widehat{\mathbf{x}}(\mathbf{x}^n, \boldsymbol{\tau}) - \mathbf{x}^n)$;

`(S.4)` : $n \leftarrow n + 1$, and go to `(S.1)`.

Theorem 1 *Given the social problem (1) under A1)-A4), suppose that one of the two following conditions is satisfied*

(a) For each i, $\mathbf{H}_i(\mathbf{x}) - c_{H_i}\mathbf{I} \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{K}$ and some $c_{H_i} > 0$, and the sequence $\{\gamma^n\}$ and $\boldsymbol{\tau} > \mathbf{0}$ are chosen so that

$$0 < \inf_n \gamma^n \leq \sup_n \gamma^n \leq \gamma^{\max} \leq 1 \text{ and } 2\, c_{\boldsymbol{\tau}} \geq \gamma^{\max} L_{\nabla U}, \tag{5}$$

where $c_{\boldsymbol{\tau}} \triangleq \min_{i \in \mathcal{I}}\{\tau_i \cdot c_{H_i} + \min_{\mathbf{x}_{-i} \in \mathcal{K}_{-i}} c_{f_{\mathcal{C}_i}}(\mathbf{x}_{-i})\}$ and $c_{f_{\mathcal{C}_i}}(\mathbf{x}_{-i})$ is the constant of strong convexity of $\sum_{j \in \mathcal{C}_i} f_j(\bullet, \mathbf{x}_{-i})$ on $\mathcal{K}_i$.[1]

(b) For each i, $\mathbf{H}_i(\mathbf{x}) - c_{H_i}\mathbf{I} \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{K}$ and some $c_{H_i} > 0$, $\boldsymbol{\tau} > \mathbf{0}$, and the sequence $\{\gamma^n\}$ is chosen so that

$$\gamma^n \in (0, 1], \quad \gamma^n \to 0, \quad \text{and} \quad \sum_n \gamma^n = +\infty. \tag{6}$$

Then, either Algorithm 1 converges in a finite number of iterations to a stationary solution of (1) or every limit point of the sequence $\{\mathbf{x}^n\}_{n=1}^\infty$ (at least one such points exists) is a stationary solution of (1). Moreover, none of such points is a local maximum of U.

Main features of Algorithm 1. The algorithm implements a novel *distributed* SCA decomposition: all the users solve *in parallel* the sequence of *decoupled* strongly convex optimization problems (4). It is expected to perform better than classical gradient-based schemes at the cost of no extra signaling, because the convexity of the objective functions, if any, is better exploited. It is guaranteed to converge under the weakest assumptions available in the literature while offering some flexibility in the choice of the free parameters [see conditions a) or b) of Theorem 1]. This degree of freedom can be exploited, e.g., to achieve the desired tradeoff between signaling and convergence speed, as discussed next. Note that *inexact* versions of Algorithm 1 are possible [where each user need not solve (4) exactly], but we do not discuss them here for lack of space; see [19].

On the choice of the free parameters. Convergence of Algorithm 1 is guaranteed either using a constant step-size rule [cf. (5)] or a diminishing step-size rule [cf. (6)]. Moreover, different feasible choices of $\{\mathcal{C}_i\}$ are possible for a given social function, resulting in different best-response functions and signaling among the users.

1) Constant step-size. In this case, $\gamma^n = \gamma \leq \gamma^{\max}$ for all n, where $\gamma^{\max} \in (0, 1]$ is chosen together with $\boldsymbol{\tau} > \mathbf{0}$ and $(\mathbf{H}_i(\mathbf{y}))_{i=1}^I$ so that the condition $2\, c_{\boldsymbol{\tau}} \geq \gamma^{\max} L_{\nabla U}$ is satisfied. This can be done in several ways. A case worth mentioning is: $\gamma = \gamma^{\max} = 1$ for all n, $\mathbf{H}_i(\mathbf{y}) = \mathbf{I}$ for all $i \in \mathcal{I}$, and $\boldsymbol{\tau} > \mathbf{0}$ large enough so that $2\, c_{\boldsymbol{\tau}} \geq L_{\nabla U}$. This choice leads to the more classical proximal-Jacobi best-response scheme: at each iteration n, $\mathbf{x}_i^{n+1} = \widehat{\mathbf{x}}_i(\mathbf{x}^n, \boldsymbol{\tau})$, $\forall i \in \mathcal{I}$. To the best of our knowledge, this algorithm along with its convergence conditions [Theorem 1a)] is new in the optimization literature; indeed classic best-response nonlinear Jacobi schemes require much stronger (sufficient) conditions to converge (implying contraction) [3, Ch. 3.3.5]. Note that the choice of the τ_i's to guarantee convergence [i.e., $2\, c_{\boldsymbol{\tau}} \geq L_{\nabla U}$] can be done locally by each user with no signaling exchange, if $L_{\nabla U}$ is known.

2) Variable step-size. In scenarios where $L_{\nabla U}$ is not available, one can use the diminishing step-size rule (6), whose implementation does not require any information on the system parameters. Moreover, under a diminishing step-size rule, convergence is guaranteed for *any* choice of $\boldsymbol{\tau} > \mathbf{0}$ [if the function $\sum_{j \in \mathcal{C}_i} f_j(\bullet, \mathbf{x}_{-i})$ are strongly convex for any $\mathbf{x}_{-i} \in \mathcal{K}_{-i}$, we can also set $\tau_i = 0$] and $\mathbf{H}_i(\mathbf{x}) - c_{H_i}\mathbf{I} \succeq \mathbf{0}$, which provides a further degree of flexibility. An example of step-size rule satisfying (6) is: given $\gamma^0 = 1$, let

$$\gamma^n = \gamma^{n-1}\left(1 - \epsilon\, \gamma^{n-1}\right), \quad n = 1, \ldots, \tag{7}$$

[1] If $\sum_{j \in \mathcal{C}_i} f_j(\bullet, \mathbf{x}_{-i})$ is convex but not strongly convex, $c_{f_{\mathcal{C}_i}} = 0$.

where $\epsilon \in (0,1)$ is a given constant; see [19] for others rules.

Another issue to discuss is the choice of the free positive definite matrices $\mathbf{H}_i(\mathbf{x}^n)$. Mimicking (quasi-)Newton-like schemes [20], a possible choice is to consider a proper (diagonal) uniformly positive definite "approximation" of the Hessian matrix $\nabla^2_{\mathbf{x}_i} U(\mathbf{x}^n)$. The exact expression to consider depends on the amount of signaling and computational power available to calculate such a $\mathbf{H}_i(\mathbf{x}^n)$, and thus varies with the specific problem under considerations [19].

3) On the choice of $\mathcal{C}_i$'s. Note that more than one feasible choice of $\{\mathcal{C}_i\}$ is in general possible for a given social function. Some illustrative examples are discussed next.

Example #1−(Proximal) gradient algorithms: If each $\mathcal{C}_i = \emptyset$ and $I = I_f$, $\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \tau_i)$ reduces to the gradient response (possibly with a proximal regularization). Therefore (exact and inexact) gradient algorithms along with their convergence conditions are special cases of our framework. Note that if $\mathcal{S}_i = \emptyset$ for every i this is the only possible choice. And indeed, if no convexity whatsoever is present in U our approach reduces to a form of gradient method. On the other hand, as soon as at least one $\mathcal{S}_i \neq \emptyset$ we depart from the gradient method and exploit in the best possible way the available convexity.

Example #2−Algorithms in [15]: Suppose that $I = I_f$, and each $\mathcal{S}_i = \{i\}$ (implying that $f_i(\bullet, \mathbf{x}_{-i})$ is convex on $\mathcal{K}_i$ for any $\mathbf{x}_{-i} \in \mathcal{K}_{-i}$). By taking each $\mathcal{C}_i = \{i\}$, we obtain the algorithms in [15].

Example #3−(Proximal) Jacobi algorithms for a single jointly convex function: Suppose that the social function is a single (jointly) convex function $f(\mathbf{x}_1, \ldots, \mathbf{x}_I)$ on $\mathcal{K} = \prod_i \mathcal{K}_i$. Of course, this optimization problem is a special case of our setting, with $\mathcal{C}_i \equiv \mathcal{I}_f = \{1\}$ for all $i \in \mathcal{I}$ and $f_1(\mathbf{x}) = f(\mathbf{x})$. Then, (4) reduces to

$$\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \tau_i) \triangleq \underset{\mathbf{x}_i \in \mathcal{K}_i}{\operatorname{argmin}}\, f(\mathbf{x}_i, \mathbf{x}^n_{-i}) + \frac{\tau_i}{2}\left\|\mathbf{x}_i - \mathbf{x}^n_i\right\|^2, \qquad (8)$$

where we set $\mathbf{H}_i(\mathbf{x}^n) = \mathbf{I}$. Algorithm 1 based on such a $\widehat{\mathbf{x}}_{\mathcal{C}_i}(\mathbf{x}^n, \tau_i)$ reads as an exact (proximal) block-Jacobi scheme converging to the global minima of $f(\mathbf{x}_1, \ldots, \mathbf{x}_I)$ over $\mathcal{K}$. To the best of our knowledge, this is a new algorithm in the literature; moreover its convergence conditions enlarge current ones; see, e.g., [3, Sec. 3.2.4].

Other special cases of our framework (such as the application to D.C. programming) are discussed in [19].

3. THE COMPLEX CASE

In this section we show how to extend our framework to sum-utility problems where the users' optimization variables are complex matrices. This will allow us to deal with the design of MIMO multiuser systems. Let us consider the following sum-utility optimization:

$$\begin{array}{ll} \underset{\mathbf{X}_1, \ldots, \mathbf{X}_I}{\operatorname{minimize}} & U(\mathbf{X}) \triangleq \displaystyle\sum_{i \in \mathcal{I}_f} f_i(\mathbf{X}_i,\, \mathbf{X}_{-i}) \\ \text{subject to} & \mathbf{X}_i \in \mathcal{K}_i, \quad \forall i \in \mathcal{I}, \end{array} \qquad (9)$$

where now $\mathbf{X}_i \in \mathbb{C}^{n_i \times m_i}$, $\mathcal{K}_i \subseteq \mathbb{C}^{n_i \times m_i}$, and $f_i : \mathcal{K} \to \mathbb{R}$. We study (9) under the same assumptions A1)-A4) stated for the real case, where in A2) the differentiability condition is now replaced by the $\mathbb{R}$-differentiability (see, e.g., [21, 22]), and in A3) $U(\mathbf{X})$ is required to have Lipschitz *conjugate-gradient* $\nabla_{\mathbf{X}^*} U(\mathbf{X})$ on $\mathcal{K}$, with constant $L^{\mathbb{C}}_{\nabla U}$, where $\mathbf{X}^*$ is the conjugate of $\mathbf{X}$.

At the basis of the proposed decomposition techniques for (9) there is the (second order) Taylor expansion of a continuously $\mathbb{R}$-differentiable function $f : \mathbb{C}^{n \times m} \to \mathbb{R}$ [23]:

$$\begin{aligned} f(\mathbf{X} + \Delta\mathbf{X}) - f(\mathbf{X}) &\approx 2\left\langle \Delta\mathbf{X},\, \nabla_{\mathbf{X}^*} f(\mathbf{X})\right\rangle \\ &+ \frac{1}{2}\operatorname{vec}([\boldsymbol{\Delta}\mathbf{X}, \boldsymbol{\Delta}\mathbf{X}^*])^H \mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X}) \operatorname{vec}([\boldsymbol{\Delta}\mathbf{X}, \boldsymbol{\Delta}\mathbf{X}^*]), \end{aligned}$$

where $\langle \mathbf{A}, \mathbf{B}\rangle \triangleq \operatorname{tr}(\mathbf{A}^H\mathbf{B})$, $\operatorname{vec}(\bullet)$ denotes the "vec" operator, and $\mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X})$ is the so-called *augmented Hessian* of f,

$$\mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X}) \triangleq \frac{\partial}{\partial \operatorname{vec}([\mathbf{X}, \mathbf{X}^*])^T}\left(\frac{\partial f(\mathbf{X})}{\partial \operatorname{vec}([\mathbf{X}^*, \mathbf{X}])^T}\right)^T. \qquad (10)$$

In [23], we proved that $\mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X})$ plays the role of the Hessian matrix for functions of real variables. In particular, f is strongly convex on $\mathbb{C}^{n \times m}$ if and only if there exists a $c_{f^{\mathbb{C}}} > 0$, the constant of strong convexity, such that

$$\operatorname{vec}([\mathbf{Y}, \mathbf{Y}^*])^H \mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X}) \operatorname{vec}([\mathbf{Y}, \mathbf{Y}^*]) \geq c_{f^{\mathbb{C}}} \left\|\mathbf{Y}\right\|_F^2, \qquad (11)$$

for all $\mathbf{X} \in \mathbb{C}^{n \times m}$ and $\mathbf{Y} \in \mathbb{C}^{n \times m}$, where $\|\bullet\|_F$ denotes the Frobenius norm. When (11) holds, we say that $\mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X})$ is *augmented* uniformly positive definite, and write $\mathcal{H}_{\mathbf{X}\mathbf{X}^*} f(\mathbf{X}) - c_{f^{\mathbb{C}}}\mathbf{I} \overset{\mathcal{A}}{\succeq} \mathbf{0}$ [23]. When f is only convex but not strongly convex, then $c_{f^{\mathbb{C}}} = 0$.

Motivated by the Taylor expansion above, and using the same symbols $\mathcal{S}_i$ and $\mathcal{C}_i$ to denote the complex counterparts of $\mathcal{S}_i$ and $\mathcal{C}_i$ introduced for the real case [cf. (2)], let us consider for each user i the following convex approximation of $U(\mathbf{X})$ at $\mathbf{X}^n$: denoting by $\Delta\mathbf{X}_i \triangleq \mathbf{X}_i - \mathbf{X}^n_i$,

$$\begin{aligned} \widetilde{f}_{\mathcal{C}_i}(\mathbf{X}_i; \mathbf{X}^n) &\triangleq \sum_{j \in \mathcal{C}_i} f_j(\mathbf{X}_i,\, \mathbf{X}^n_{-i}) + \left\langle \boldsymbol{\Pi}_{\mathcal{C}_i}(\mathbf{X}^n),\, \Delta\mathbf{X}_i\right\rangle \\ &+ \frac{\tau_i}{2}\operatorname{vec}([\Delta\mathbf{X}_i, \Delta\mathbf{X}^*_i])^H \mathcal{H}_i(\mathbf{X}^n) \operatorname{vec}([\Delta\mathbf{X}_i, \Delta\mathbf{X}^*_i]) \end{aligned} \qquad (12)$$

with $\boldsymbol{\Pi}_{\mathcal{C}_i}(\mathbf{X}^n) \triangleq \sum_{j \in \mathcal{C}_{-i}} \nabla_{\mathbf{X}^*_i} f_j(\mathbf{X})\big|_{\mathbf{X} = \mathbf{X}^n}$, where $\mathcal{H}_i(\mathbf{X}^n)$ is any given $2nm \times 2nm$ matrix such that $\mathcal{H}_i(\mathbf{X}) - c_{\mathcal{H}_i}\mathbf{I} \overset{\mathcal{A}}{\succeq} \mathbf{0}$, for all $\mathbf{X} \in \mathcal{X}$ and some $c_{\mathcal{H}_i} > 0$. Note that if $\mathcal{H}_i(\mathbf{X}) = \mathbf{I}$, the quadratic term in (12) reduces to the standard proximal regularization $\tau_i \left\|\mathbf{X}_i - \mathbf{X}^n_i\right\|_F^2$. Then, the best-response matrix function of each user $i \in \mathcal{I}$ is

$$\widehat{\mathbf{X}}_{\mathcal{C}_i}(\mathbf{X}^n, \tau_i) \triangleq \underset{\mathbf{X}_i \in \mathcal{K}_i}{\operatorname{argmin}}\, \widetilde{f}_{\mathcal{C}_i}(\mathbf{X}_i; \mathbf{X}^n). \qquad (13)$$

Decomposition algorithms for (9) are formally the same as those proposed in Sec. 2 for (1), where the real-valued best-response map $\widehat{\mathbf{x}}(\mathbf{x}^n, \boldsymbol{\tau})$ is replaced by the complex-valued counterpart $\widehat{\mathbf{X}}_{\mathcal{C}}(\mathbf{X}^n, \boldsymbol{\tau}) \triangleq (\widehat{\mathbf{X}}_{\mathcal{C}_i}(\mathbf{X}^n, \tau_i))_{i=1}^I$. Convergence conditions read as those in Theorem 1, under the following changes [19]: i) $L_{\nabla U}$ becomes $L^{\mathbb{C}}_{\nabla U}$; ii) $\mathbf{H}_i(\mathbf{x}) - c_{H_i}\mathbf{I} \succeq \mathbf{0}$ reads $\mathcal{H}_i(\mathbf{X}) - c_{\mathcal{H}_i}\mathbf{I} \overset{\mathcal{A}}{\succeq} \mathbf{0}$; and iii) in the definition of $c_{\boldsymbol{\tau}}$ in Theorem 1a), c_{H_i} and $c_{f_i}(\mathbf{x}_{-i})$ are replaced by $c_{\mathcal{H}_i}$ and $c_{f_{\mathcal{C}_i}}(\mathbf{X}_{-i})$, respectively, where $c_{f_{\mathcal{C}_i}}(\mathbf{X}_{-i})$ is the constant of strong convexity of $\sum_{j \in \mathcal{C}_i} f_j(\bullet,\, \mathbf{X}^n_{-i})$ on $\mathcal{K}_i$.

4. THE SUM-RATE MAXIMIZATION PROBLEM

In this section, we customize the proposed decomposition framework to a sum-rate maximization problem over MIMO Gaussian ICs, and compare the resulting new algorithm with state-of-the-art schemes [8, 9, 14]. Consider the sum-rate maximization problem

$$\begin{array}{ll} \underset{\mathbf{Q}_1, \ldots, \mathbf{Q}_I}{\operatorname{maximize}} & \displaystyle\sum_{i \in \mathcal{I}} w_i\, r_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \\ \text{subject to} & \mathbf{Q}_i \in \mathcal{Q}_i, \quad \forall i \in \mathcal{I}. \end{array} \qquad (14)$$

where $r(\mathbf{Q}_i, \mathbf{Q}_{-i})$ is the rate over the MIMO link i,

$$r_i(\mathbf{Q}_i, \mathbf{Q}_{-i}) \triangleq \log\det\left(\mathbf{I} + \mathbf{H}^H_{ii}\mathbf{R}_i(\mathbf{Q}_{-i})^{-1}\mathbf{H}_{ii}\mathbf{Q}_i\right), \qquad (15)$$

$\mathbf{Q}_i$ is the covariance matrix of transmitter i, $\mathbf{R}_i(\mathbf{Q}_{-i}) \triangleq \mathbf{R}_{n_i} + \sum_{j \neq i} \mathbf{H}_{ij}\mathbf{Q}_j\mathbf{H}^H_{ij}$ is the covariance matrix of the multiuser interference plus the thermal noise $\mathbf{R}_{n_i}$ (assumed to be full-rank), with $\mathbf{Q}_{-i} \triangleq (\mathbf{Q}_j)_{j\neq i}$, $\mathbf{H}_{ij}$ is the channel matrix between the j-th transmitter and the i-th receiver, and $\mathcal{Q}_i$ is the feasible set of user i,

$$\mathcal{Q}_i \triangleq \left\{ \mathbf{Q}_i \in \mathbb{C}^{n_i \times n_i} \,:\, \mathbf{Q}_i \succeq \mathbf{0},\, \text{tr}(\mathbf{Q}_i) \leq P_i,\, \mathbf{Q}_i \in \mathcal{Z}_i \right\}, \quad (16)$$

where P_i is the transmit power budget. In $\mathcal{Q}_i$ we also included an arbitrary convex and closed set $\mathcal{Z}_i$, which allows us to add additional constraints, such as: i) null constraints $\mathbf{U}_i^H \mathbf{Q}_i = \mathbf{0}$, where $\mathbf{U}$ is a tall full-column rank matrix; ii) soft-shaping constraints $\mathtt{tr}\left(\mathbf{G}_i^H \mathbf{Q}_i \mathbf{G}_i\right) \leq I_i^{\text{ave}}$, where $\mathbf{G}_i$ is an arbitrary given matrix; iii) peak-power constraints $\lambda_{\max}\left(\mathbf{F}_i^H \mathbf{Q}_i \mathbf{F}_i\right) \leq I_i^{\text{peak}}$; and iv) per-antenna constraints $[\mathbf{Q}_i]_{kk} \leq \alpha_{ik}$. Note that the sum-rate maximization problems studied, e.g., in [8, 9, 11, 14] are special cases of (14).

This problem is of course an instance of (9), and alternative decompositions are possible [19]. Since each $r_i(\mathbf{Q}_i, \mathbf{Q}_{-i})$ is concave in $\mathbf{Q}_i \in \mathcal{Q}_i$, a natural choice is $I_f = I$ and $\mathcal{C}_i = \{i\}$, which leads to the following class of strongly concave subproblems [cf. (4)]: given $\mathbf{Q}^n \in \mathcal{Q}$ and choosing $\mathcal{H}_i(\mathbf{Q}^n) = \mathbf{I}$, the best-response of user i is

$$\hat{\mathbf{Q}}_i(\mathbf{Q}^n, \tau_i) \triangleq \underset{\mathbf{Q}_i \in \mathcal{Q}_i}{\text{argmax}} \left\{ w_i\, r_i(\mathbf{Q}_i, \mathbf{Q}_{-i}^n) - \langle \mathbf{\Pi}_i(\mathbf{Q}^n), \mathbf{Q}_i \rangle - \tau_i \left\| \mathbf{Q}_i - \mathbf{Q}_i^n \right\|_F^2 \right\}$$

where $\mathbf{\Pi}_i(\mathbf{Q}^n) \triangleq \sum_{j \in \mathcal{N}_i} w_j\, \mathbf{H}_{ji}^H \widetilde{\mathbf{R}}_j(\mathbf{Q}^n) \mathbf{H}_{ij}$, $\mathcal{N}_i$ is the set of neighbors of user i, i.e., the set of users j's which user i interferers with, and $\widetilde{\mathbf{R}}_j(\mathbf{Q}^n) \triangleq \mathbf{R}_j(\mathbf{Q}_{-j}^n)^{-1} - \left(\mathbf{R}_j(\mathbf{Q}_{-j}^n) + \mathbf{H}_{jj}\mathbf{Q}_j^n \mathbf{H}_{jj}^H\right)^{-1}$.

Given $\hat{\mathbf{Q}}_i(\mathbf{Q}^n, \tau_i)$, one can now use any instance of Algorithm 1. For example, a good candidate is the scheme with diminishing step-size, whose convergence is guaranteed if, e.g., the rule in (7) is used for the sequence $\{\gamma^n\}$ [Theorem 1b)]. Note that the proposed algorithm is fairly distributed. Indeed, given the covariance matrix of the interference generated by the other users and the current interference price matrix $\mathbf{\Pi}_i(\mathbf{Q}^n)$, each user can efficiently and locally compute his best-response $\hat{\mathbf{Q}}_i(\mathbf{Q}^n, \tau_i)$ solving a strongly concave optimization problem. Moreover, for some specific structures of the feasible sets $\mathcal{Q}_i$, full-column rank channel matrices $\mathbf{H}_i$, and $\tau_i = 0$, a solution in closed form (up to the multipliers associated with the power budget constraints) is also available [9]. The estimation of the prices $\mathbf{\Pi}_i(\mathbf{Q}^n)$ requires however some signaling among nearby receivers. Quite interestingly, the pricing expression and thus the resulting signaling overhead necessary to compute it coincides with that in [8, 9, 14]. But, because of their sequential nature, algorithms in [8, 9] require more CSI exchange in the network than the proposed *simultaneous* schemes.

Numerical Example. We compare now our Algorithm 1 based on $\hat{\mathbf{Q}}_i(\mathbf{Q}^n, \tau_i)$ and the rule in (7) (termed SJBR) with those proposed in [8, 9] (termed MDP), and [14] (termed WMMSE). Since the methods in [8, 9, 14] deal only with power budget constraints, we simplified the sum-rate maximization problem (14) accordingly; we set $w_i = 1$, $P_i = P$ and $\mathbf{R}_{n_i} = \sigma^2 \mathbf{I}$ for all i, and $\mathtt{snr} \triangleq P/\sigma^2 = 3$dB. All the transmitters/receivers are equipped with 4 antenna. We simulated uncorrelated fading channels, whose coefficients are generated as i.i.d. Gaussian random variables with zero mean and variance $1/d_{ij}^3$, where d_{ij} is the distance between the transmitter j and the receiver i; for the sake of simplicity, we set $d \triangleq d_{ij}/d_{ii}$, with $d_{ij} = d_{ji}$ and $d_{ii} = d_{jj}$, for all i and $j \neq i$. In (7) we chose $\epsilon = 1e-5$. In Fig. 1, we plot the average number of iterations required by the aforementioned algorithms to converge (under the same termination criterion set to $1e-6$) versus the number of users. The average is taken over 100 independent channel realizations and $d = 3$. In Table 1, we report the average number of iterations required to converge for different (normalized) distances d, number of users, and termination accuracy equal to $1e$-3. In our experiments, all the algorithms reach the same average sum-rate.

The analysis of the numerical results shows that the proposed SJBR outperforms all the others schemes in terms of iterations, while having similar (or even better) computational complexity (see [19] for details on the complexity analysis). Interestingly, the iteration gap with the other schemes reduces with the distance and the termination accuracy. More specifically: i) SJBR seems to be much faster than all the other schemes (about one order of magnitude) when $d = 3$ [say low interference scenarios], and just a bit faster (or comparable to WMMSE) when $d = 1$ [say high interference scenarios]; and ii) SJBR is much faster than all the others, if an high termination accuracy is set (compare Fig. 1 with Table I). Also, the convergence speed of SJBR is not affected too much by the number of users. Finally, in our experiments, we also observed that the performance of SJBR are not affected too much by the choice of the parameter ϵ in the (7): a change of ϵ of many orders of magnitude leads to a difference in the average number of iterations which is within 5%; we refer the reader to [24] for details, where one can also find a comparison of several other step-size rules. We must stress however that MDP and WMMSE do not need any tuning, which is an advantage with respect to our method.

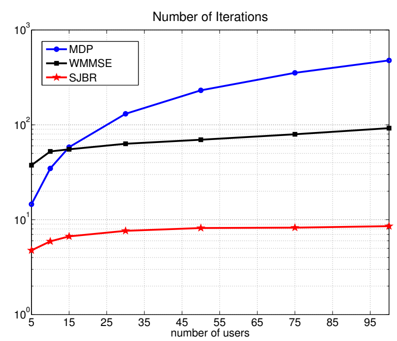

Fig. 1. Average number of iterations versus number of users (termination accuracy= $1e-6$).

	# of users = 10			# of users = 50			# of users = 100		
	d=1	d=2	d=3	d=1	d=2	d=3	d=1	d=2	d=3
MDP	429.4	74.3	32.8	1739.5	465.5	202	3733	882	442.6
WMMSE	51.6	19.2	14.7	59.6	24.9	16.3	69.8	26.0	19.2
SJBR	48.6	9.4	4.0	46.9	12.6	5.1	49.7	12	5.5

Table 1. Average number of iterations (termination accuracy=$1e$-3)

5. CONCLUSIONS

In this paper, we proposed a novel decomposition framework, based on SCA, to compute stationary solutions of general nonconvex sum-utility problems (including social functions of complex variables). The main result is a new class of convergent *distributed Jacobi* best-response algorithms, where all users *simultaneously* solve a suitably convexified version of the original social problem. Quite interestingly, our framework contains as special cases many decomposition methods already proposed in the literature, such as gradient algorithms, and many block-coordinate descents schemes for convex functions. Finally, we tested our methodology on the sum-rate maximization problem over MIMO ICs; our experiments show that our algorithms are (much) faster than ad-hoc state-of-the-art methods [8, 9, 14]. We are currently investigating how to adaptively choose the step-size rule (so that no a-priori tuning is needed), and how to generalize our framework to scenarios where only long-term channel statistics are available.

6. REFERENCES

[1] 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.

[2] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, “Layering as optimization decomposition: A mathematical theory of network architectures,” *Proc. of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

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

[4] Z.-Q. Luo and S. Zhang, “Spectrum management: Complexity and duality,” *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–72, Feb. 2008.

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

[6] F. Wang, M. Krunz, and S. Cui, “Price-based spectrum management in cognitive radio networks,” *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 74–87, Feb. 2008.

[7] C. Shi, R. Berry, and M. Honig, “Distributed interference pricing for ofdm wireless networks with non-separable utilities,” in *42nd Annu. Conf. Information Sciences and Systems*, Princeton, NJ, USA, Mar. 2008.

[8] 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.

[9] 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.

[10] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “A unified convergence analysis of block successive minimization methods for nonsmooth optimization,” *Arxiv.org*, Oct. 2012. http://arxiv.org/abs/1209.2385v1.

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

[12] M. Chiang, C. W. Tan, D. P. 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.

[13] J. Papandriopoulos and J. S. Evans, “Scale: a low-complexity distributed protocol for spectrum balancing in multiuser dsl networks,” *IEEE Trans. on Information Theory*, vol. 55, no. 8, pp. 3711–3724, Aug. 2009.

[14] 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.

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

[16] B. R. Marks and G. P. Wright, “A general inner approximation algorithm for nonconvex mathematical programs,” *Operations Research*, vol. 26, no. 2, pp. 681–683, 1978.

[17] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, “Distributed interference pricing for the MIMO interference channel,” in *Int. Conf. on Comm. (ICC)*, Princeton, NJ, USA, June 14-18 2009.

[18] 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 heterogenous networks,” *Arxiv.org*, Oct. 2012. http://arxiv.org/abs/1210.1507v1.

[19] 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*, (submitted, Jan. 2013). [Online]. Available: http://arxiv.org/abs/1302.0756.

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

[21] K. Kreutz-Delgado, “The complex gradient operator and the CR-calculus,” Electrical and Computer Engineering, Jacobs School of Engineering, University of California, San Diego, Tech. Rep., July 2009.

[22] A. Hjorungnes, *Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. London: Cambridge University Press, May 2011.

[23] 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

[24] P. Song, “Numerical comparison of decomposition algorithms for nonconvex sum-utility problems,” Dept. of Elect. Eng., Univ. at Buffalo, the State Univ. of New York, Tech. Rep., Jan. 2013. [Online]. Available: http://www.eng.buffalo.edu/∼peiranso/SongTechRepNumCompJan13.pdf.