

Distributed constrained convex optimization and consensus via dual decomposition and proximal minimization

Alessandro Falsone, Kostas Margellos, Simone Garatti, Maria Prandini

Abstract— We consider a general class of convex optimization problems over time-varying, multi-agent networks, that naturally arise in many application domains like energy systems and wireless networks. In particular, we focus on programs with separable objective functions, local (possibly different) constraint sets and a coupling inequality constraint expressed as the non-negativity of the sum of convex functions, each corresponding to one agent. We propose a novel distributed algorithm to deal with such problems based on a combination of dual decomposition and proximal minimization. Our approach is based on an iterative scheme that enables agents to reach consensus with respect to the dual variables, while preserving information privacy. Specifically, agents are not required to disclose information about their local objective and constraint functions, nor to assume knowledge of the coupling constraint. Our analysis can be thought of as a generalization of dual gradient/subgradient algorithms to a distributed set-up. We show convergence of the proposed algorithm to some optimal dual solution of the centralized problem counterpart, while the primal iterates generated by the algorithm converge to the set of optimal primal solutions. A numerical example demonstrating the efficacy of the proposed algorithm is also provided.

I. INTRODUCTION

Optimization in multi-agent networks has attracted significant interest from both the control and the operations research community, and has already found numerous applications in different domains, like power systems [1], [2], wireless networks [3], [4], robotics [5], etc. Typically, agents cooperate to reach agreement/consensus on a common decision, while optimizing a given performance criterion. From a centralized perspective, this task can be represented as an optimization problem defined over the entire network, but the resulting mathematical program is often of large size, making numerical computations prohibitive for large scale systems, and/or requires the presence of a central entity to have access to agent specific information, e.g., agents' utility/objective and constraint functions.

Distributed optimization offers the means to bypass these limitations, that are inherent in centralized approaches, allowing agents to keep information about their objective and constraint functions private, while distributing computation,

thus leading to computational savings compared to centralized paradigms. Typical implementations involve applying an iterative procedure, where at each iteration agents perform some local computation. They then exchange the outcome of this computation (but not their private information) with neighboring agents, and the process is then repeated on the basis of the information received.

A notable research activity oriented to the development of distributed optimization algorithms over time-varying multi-agent networks for general classes of convex problems has flourished in recent years. In particular, in [6], [7], [8], [9] a gradient/subgradient based consensus approach is followed to address problems where agents with their own objective functions and constraints are coupled via a common decision vector. In [10] the authors revisit this problem from a proximal minimization perspective, addressing also the case where the agents' constraints may be affected by uncertainty. Another class of problems, which is the one considered in this paper and which has attracted considerable interest, involves programs with separable objective functions each agent having its own decision vector, local (possibly different) constraint sets, and a coupling inequality constraint expressed as the non-negativity of the sum of convex functions, each corresponding to one agent. Applying the methodologies of the aforementioned references to this problem, though possible, would unnecessarily increase the computational and communication effort, since it would require each agent to maintain an estimate of the decision vectors of all other agents when solving its local optimization program, and to communicate it to its neighboring agents.

To exploit the particular problem structure and alleviate these difficulties, dual decomposition techniques (see [11], and references therein), or approaches based on the alternating direction method of multipliers [12], are often employed, relying on the separable structure of the problem after dualizing the coupling constraint. These methods are based on time-invariant, connected networks, and require a central update step for the dual variables, that should be then communicated to all agents that are coupled via the constraints. The latter, however, may not be possible in time-varying connectivity set-ups. Standard incremental gradient/subgradient algorithms [13], [14], [15] constitute an alternative to dual decomposition, however, they require agents to perform updates sequentially, in a cyclic or randomized order, and hence do not allow for parallelizable computations. Recently these techniques have been extended to allow for distributed computation under the assumption that the underlying network is time-invariant and the agents

Research was supported by the European Commission under the project UnCoVerCPS, grant number 643921.

Alessandro Falsone, Simone Garatti and Maria Prandini are with the Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, e-mail: {alessandro.falsone, simone.garatti, maria.prandini}@polimi.it

Kostas Margellos is with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom, e-mail: kostas.margellos@eng.ox.ac.uk

have memory capabilities [16]. Other extensions of such incremental algorithms are provided in [17], [18], [19], [20], though addressing the problem under study using the approaches proposed in the aforementioned references would require all agents to store and exchange copies of their local decision variables with their neighbors. This would result in an unnecessary increase of the amount of communication and, moreover, it requires an exchange of private information. Another research direction involves primal-dual subgradient based consensus algorithms [21], whereas in [22] a perturbation variant with superior performance is adopted. However, in the former the coupling constraint is assumed to be known to all agents, whereas in the latter each agents' objective function is required to be differentiable.

In this paper we propose a novel distributed algorithm to deal with optimization problems that exhibit the aforementioned structure, based on a combination of dual decomposition and proximal minimization. In particular, the contributions of our paper can be summarized as follows: 1) We extend dual decomposition based algorithms to a distributed setting, accounting for possibly time-varying network connectivity. 2) We respect agents' information privacy, with agents not being required to share information about their local objective function and constraint set, nor about the constraint function that encodes their contribution to the coupling constraint. In particular, agents are not required to share their tentative estimates for the primal decision variables, but only for the dual ones. 3) We provide a proximal minimization perspective to gradient/subgradient algorithms, that allows us to bypass the differentiability assumptions on the primal objective functions, which is at the basis of such algorithms, and/or the requirement for gradient/subgradient computation.

The remainder of the paper unfolds as follows: Section II provides a statement of the problem under study, introduces the proposed algorithm, states the main results of the paper, and provides a sketch of their proofs. In Section III we demonstrate the efficacy of the proposed algorithm on a numerical example. Finally, Section IV concludes the paper and provides some directions for future work. Complete proofs of the main statements, as well as some intermediate results, are omitted in the interest of space; they are, however, available in [23].

II. DISTRIBUTED CONSTRAINED OPTIMIZATION

A. Problem statement and proposed solution

Consider a time-varying network of m agents that communicate to solve the following optimization program

$$\begin{aligned} \mathcal{P} : \quad & \min_{\{x_i \in X_i\}_{i=1}^m} \sum_{i=1}^m f_i(x_i) \\ & \text{subject to: } \sum_{i=1}^m g_i(x_i) \leq 0, \end{aligned} \quad (1)$$

where for each $i = 1, \dots, m$, $x_i \in \mathbb{R}^{n_i}$ is the vector of n_i decision variables of agent i , $f_i(\cdot) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ is its objective function, $X_i \subseteq \mathbb{R}^{n_i}$ its local constraint set, and $g_i(\cdot) : \mathbb{R}^{n_i} \rightarrow$

\mathbb{R}^p is a function which represents the contribution of agent i to the coupling constraint¹ $\sum_{i=1}^m g_i(x_i) \leq 0$.

Solving \mathcal{P} in a centralized fashion, would likely result in a computationally intensive program, especially in the case where the number of interacting agents is high. Moreover, agent i may not be willing to share information about $f_i(\cdot)$, X_i , and $g_i(\cdot)$, with other agents, due to privacy issues. To account for information privacy and facilitate the development of a computationally tractable solution, we seek for a distributed strategy. Let $x = [x_1^\top \cdots x_m^\top]^\top \in \mathbb{R}^n$, with $n = \sum_{i=1}^m n_i$, and $X = X_1 \times \cdots \times X_m$. Motivated by the separable structure of \mathcal{P} , consider its dual problem, which is given by

$$\mathcal{D} : \max_{\lambda \geq 0} \min_{x \in X} L(x, \lambda), \quad (2)$$

where λ is the vector of Lagrange multipliers, and $\lambda \geq 0$ stands for $\lambda \in \mathbb{R}_+^p$, where \mathbb{R}_+^p denotes the p -th dimensional non-negative orthant. The Lagrangian function $L(x, \lambda) : \mathbb{R}^n \times \mathbb{R}_+^p \rightarrow \mathbb{R}$ is given by

$$L(x, \lambda) = \sum_{i=1}^m L_i(x_i, \lambda) = \sum_{i=1}^m \{f_i(x_i) + \lambda^\top g_i(x_i)\}. \quad (3)$$

The dual function can be then defined as

$$\varphi(\lambda) = \min_{x \in X} L(x, \lambda). \quad (4)$$

Since it is the point-wise minimum of affine functions, $\varphi(\cdot)$ is a concave function. Notice that, due to the separable structure of the objective and the constraint functions in \mathcal{P} ,

$$\varphi(\lambda) = \sum_{i=1}^m \varphi_i(\lambda) = \sum_{i=1}^m \min_{x_i \in X_i} L_i(x_i, \lambda). \quad (5)$$

Therefore, we can equivalently write \mathcal{D} as

$$\mathcal{D} : \max_{\lambda \geq 0} \sum_{i=1}^m \varphi_i(\lambda), \quad (6)$$

in which each agent i has its own dual function $\varphi_i(\lambda)$, and the coupling between agents arises due to the fact that they should all agree on the same vector λ .

Although, in principle, (6) fits the framework of algorithms like [7], [10], the concave function $\varphi_i(\cdot)$ is implicitly defined through an optimization problem parametric in λ . This would require each agent to handle a max-min optimization program which, apart from specific cases, is in general difficult to solve.

We develop a distributed algorithm that is specifically tailored to the resolution of \mathcal{D} . Its basic steps are summarized in Algorithm 1. At the initialization step, each agent i , $i = 1, \dots, m$, considers an estimate of its local decision vector such that $\hat{x}_i(0) \in X_i$, and an estimate of what the solution of \mathcal{D} is believed to be, i.e., $\lambda_i(0) \in \mathbb{R}_+^p$ (step 3 and step 4 of Algorithm 1, respectively). A sensible choice is to set $\hat{x}_i(0)$ as $\hat{x}_i(0) \in \arg \min_{x_i \in X_i} f_i(x_i)$, and to take

¹Inequality here is meant component-wise. It is perhaps worth noticing that this setup comprises also equality coupling constraints like $\sum_{i=1}^m \hat{g}_i(x_i) = 0$. To this purpose it is enough to define $g_i = [\hat{g}_i^\top - \bar{g}_i^\top]^\top$.

Algorithm 1 Distributed algorithm

- 1: **Initialization**
 - 2: $k = 0$.
 - 3: Consider $\hat{x}_i(0) \in X_i$, for all $i = 1, \dots, m$.
 - 4: Consider $\lambda_i(0) \in \mathbb{R}_+^p$, for all $i = 1, \dots, m$.
 - 5: **For** $i = 1, \dots, m$ **repeat until convergence**
 - 6: $\ell_i(k) = \sum_{j=1}^m a_j^i(k) \lambda_j(k)$.
 - 7: $x_i(k+1) \in \arg \min_{x_i \in X_i} f_i(x_i) + \ell_i(k)^\top g_i(x_i)$.
 - 8: $\lambda_i(k+1) = \arg \max_{\lambda_i \geq 0} \left\{ g_i(x_i(k+1))^\top \lambda_i - \frac{1}{2c(k)} \|\lambda_i - \ell_i(k)\|^2 \right\}$
 - 9: $\hat{x}_i(k+1) = \hat{x}_i(k) + \frac{c(k)}{\sum_{r=0}^k c(r)} (x_i(k+1) - \hat{x}_i(k))$.
 - 10: $k \leftarrow k + 1$.
-

$\lambda_i(0) = 0$, $i = 1, \dots, m$. At every iteration k , each agent constructs a weighted average $\ell_i(k)$ of the solutions $\lambda_j(k)$, $j = 1, \dots, m$, of the other agents and its local one (step 6). Coefficient $a_j^i(k)$ is the weight that agent i attributes to the solution of agent j at iteration k ; $a_j^i(k) = 0$ means that agent j does not send any information to agent i at iteration k .

The optimization program in (6) exhibits the same structure of the problem addressed in [10], but involves dual, as opposed to primal, variables. This would motivate the use of a proximal maximization, as opposed to minimization, step, for agent i to update its local estimate $\lambda_i(k+1)$, i.e., $\lambda_i(k+1) = \arg \max_{\lambda_i \geq 0} \min_{x_i \in X_i} \left\{ L_i(x_i, \lambda_i) - \frac{1}{2c(k)} \|\lambda_i - \ell_i(k)\|^2 \right\}$, $\|\cdot\|$ being the standard Euclidean norm. However, this problem is in general not easy to solve because it requires the solution of a max-min program. Therefore, we alternate between a primal and a dual update step. In particular, the update of the local primal vector $x_i(k+1)$ (step 7) is the same as in dual decomposition, whereas, in contrast to dual decomposition, the update of the dual vector (step 8) involves also a proximal term, which facilitates consensus among the agents. Step 9 of Algorithm 1 returns an update $\hat{x}_i(k+1)$ for the auxiliary primal iterates. It can be easily shown that $\hat{x}_i(k+1)$ can be equivalently written as

$$\hat{x}_i(k+1) = \frac{\sum_{r=0}^k c(r) x_i(r+1)}{\sum_{r=0}^k c(r)}, \quad (7)$$

i.e., it is a weighted running average of $\{x_i(r+1)\}_{r=0}^k$. Such an auxiliary sequence is referred to as primal recovery procedure and it is often used in dual decomposition methods, since it has better convergence properties compared to $\{x_i(k)\}_{k \geq 0}$ [24], [22], [21].

Note that, since the maximization program in step 8 of Algorithm 1 is quadratic with respect to λ_i , an explicit resolution is possible, and step 8 can be equivalently written as $\lambda_i(k+1) = [\ell_i(k) + c(k)g_i(x_i(k+1))]_+$, where $[\cdot]_+$ denotes the projection of its argument on \mathbb{R}_+^p . The aforementioned representation resembles the structure of a projected subgradient step, where $g_i(x_i(k+1))$ constitutes a subgradient of $\varphi_i(\cdot)$ evaluated at $\ell_i(k)$, and $c(k)$ plays the role of the gradient step. Throughout the manuscript

we will use both representations according to convenience; however, the proximal perspective in step 8 of Algorithm 1, and some of its related properties that will be shown in the sequel, are crucial in the convergence analysis of the proposed algorithm, enabling us to extend the approach of [24] to the distributed case, and overcome the requirement imposed in [21] for the coupling constraint to be known to all agents (notice that in our set-up agent i needs to know only its contribution $g_i(\cdot)$ to the coupling constraint).

B. Structural and communication assumptions

We impose the following assumptions.

Assumption 1. [Convexity] For each $i = 1, \dots, m$, function $f_i(\cdot) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ and the components of $g_i(\cdot) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^p$ are convex; moreover, set $X_i \subseteq \mathbb{R}^{n_i}$ is convex too.

Assumption 2. [Compactness] For each $i = 1, \dots, m$, the set $X_i \subseteq \mathbb{R}^{n_i}$ is compact.

Note that, under Assumptions 1 and 2, $\|g_i(x_i)\|$ is finite for any $x_i \in X_i$. Therefore we have that $\|g_i(x_i)\| \leq G$, where $G = \max_{i=1, \dots, m} \max_{x_i \in X_i} \|g_i(x_i)\|$.

Assumption 3. [Constraint qualification] Problem \mathcal{P} satisfies the Slater's condition, i.e., there exists a vector $\tilde{x} = [\tilde{x}_1^\top \dots \tilde{x}_m^\top]^\top \in X$ and $\rho \in \mathbb{R}_+$ with $\rho \neq 0$, such that $\{x \in \mathbb{R}^n : \|x - \tilde{x}\| \leq \rho\} \subset X$ and $\sum_{i=1}^m g_i(\tilde{x}_i) < 0$. An equality in the latter condition is admitted only for those components that are linear.

Assumptions 1-3 are sufficient conditions for strong duality to hold, and for an optimal primal-dual pair (x^*, λ^*) to exist, where $x^* = [x_1^{*\top} \dots x_m^{*\top}]^\top$. Moreover, for any optimal pair (x^*, λ^*) the Saddle-Point Theorem [25] holds, i.e.,

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \quad (8)$$

for any $\lambda \in \mathbb{R}_+^p$ and any $x \in X$.

Denote by $X^* \times \Lambda^* = X_1^* \times \dots \times X_m^* \times \Lambda^*$ the set of all optimal primal-dual pairs. Under Assumptions 1-3, it was shown in [24] that the set of optimal dual solutions Λ^* is bounded, and hence $\sup_{\lambda \in \Lambda^*} \|\lambda\|$ is finite.

We impose the following assumption on the time-varying coefficient $c(k)$.

Assumption 4. [Coefficient $c(k)$] Assume that for all $k \geq 0$, $c(k) \in \mathbb{R}_+ \setminus \{0\}$ and $\{c(k)\}_{k \geq 0}$ is a non-increasing sequence, i.e., $c(k) \leq c(r)$ for all $k \geq r$, with $r \geq 0$. Moreover,

- 1) $\sum_{k=0}^{\infty} c(k) = \infty$,
- 2) $\sum_{k=0}^{\infty} c(k)^2 < \infty$.

One possible choice for $\{c(k)\}_{k \geq 0}$ that satisfies Assumption 4 is to take $c(k) = \beta/(k+1)$ for some $\beta \in \mathbb{R}_+ \setminus \{0\}$. This assumption is analogous to the one imposed by the authors of [10], [7], [21].

In line with [26], [27], [28] we impose the following assumptions on the information exchange between agents and on the connectivity of the network.

Assumption 5. [Weight coefficients] There exists $\eta \in (0, 1)$ such that for all $i, j \in \{1, \dots, m\}$ and all $k \geq 0$, $a_j^i(k) \in [0, 1)$, $a_i^i(k) \geq \eta$, and $a_j^i(k) > 0$ implies that $a_j^i(k) \geq \eta$. Moreover, for all $k \geq 0$,

- 1) $\sum_{j=1}^m a_j^i(k) = 1$ for all $i = 1, \dots, m$,
- 2) $\sum_{i=1}^m a_j^i(k) = 1$ for all $j = 1, \dots, m$.

For each $k \geq 0$ the information exchange between the m agents can be represented by a directed graph (V, E_k) , where the agents are the nodes $V = \{1, \dots, m\}$, and the set E_k of directed edges is defined as

$$E_k = \{(j, i) : a_j^i(k) > 0\}, \quad (9)$$

i.e., at time k the link (j, i) is present if agent j sends information to agent i and agent i weight this information with $a_j^i(k)$. If the communication link is not present we set $a_j^i(k) = 0$, otherwise if $a_j^i(k) > 0$ we say that j is a neighbor of agent of i at time k . In Algorithm 1 at each iteration each agent exchanges information with its neighbors only, thus accounting for a fully distributed setup.

Let $E_\infty = \{(j, i) : (j, i) \in E_k \text{ for infinitely many } k\}$ denote the set of edges (j, i) that represent agent pairs that communicate directly infinitely often. We then impose the following connectivity and communication assumption.

Assumption 6. [Connectivity and communication] The graph (V, E_∞) is strongly connected, i.e., for any two nodes there exists a path of directed edges that connects them. Moreover, there exists $T \geq 1$ such that for every $(j, i) \in E_\infty$, agent i receives information from a neighboring agent j at least once every consecutive T iterations.

For details about the interpretation of Assumptions 5 and 6, the reader is referred to [6], [10], [7].

C. Statement of the main results

Under Assumptions 1-6, Algorithm 1 converges and agents reach consensus to a common vector of Lagrange multipliers. In particular, their local estimates $\lambda_i(k)$ converge to some optimal dual solution, while the vector $\hat{x}(k) = [\hat{x}_1(k)^\top \dots \hat{x}_m(k)^\top]^\top$ converges to the set of optimal primal solutions X^* .

This is formally stated in the following theorems.

Theorem 1. [Dual Optimality] Consider Assumptions 1-6. We have that, for some $\lambda^* \in \Lambda^*$,

$$\lim_{k \rightarrow \infty} \|\lambda_i(k) - \lambda^*\| = 0, \text{ for all } i = 1, \dots, m. \quad (10)$$

Theorem 2. [Primal Optimality] Consider Assumptions 1-6. We have that

$$\lim_{k \rightarrow \infty} \text{dist}(\hat{x}(k), X^*) = 0, \quad (11)$$

where $\text{dist}(y, Z) = \min_{z \in Z} \|y - z\|$ denotes the distance between y and the set Z .

D. Sketch of the proof

The proofs of Theorems 1 and 2 are quite technical and require the derivation of several intermediate results,

therefore they are omitted in the interest of space. In the following we provide a sketch of the main idea behind their proofs, while for more details the interested reader is referred to [23].

Let $v(k) = \frac{1}{m} \sum_i^m \lambda_i(k)$ be the average of all tentative Lagrange multipliers at iteration k , and $e_i(k+1) = \lambda_i(k+1) - \ell_i(k+1)$ the consensus error. Assumptions 4-6 can be exploited to prove certain relations among the quantities $\|\lambda_i(k+1) - v(k+1)\|$, $\|e_i(k+1)\|$, and $c(k)$. Assumptions 1-3 allow us to embed the obtained results in an inequality that relates consecutive terms of the sequence $\{\sum_i^m \|\lambda_i(k) - \lambda^*\|^2\}_{k \geq 0}$. Specifically, it can be shown that

$$\begin{aligned} \sum_{i=1}^m \|\lambda_i(k+1) - \lambda^*\|^2 &\leq \sum_{i=1}^m \|\lambda_i(k) - \lambda^*\|^2 \\ &- \gamma_1 \sum_{i=1}^m \|e_i(k+1)\|^2 + \gamma_2 c(k)^2 \\ &+ \gamma_3 c(k) \sum_{i=1}^m \|\lambda_i(k+1) - v(k+1)\|, \end{aligned} \quad (12)$$

where γ_1 , γ_2 , and γ_3 are appropriate positive constants, together with the fact that $\sum_{k=0}^{\infty} c(k) \sum_{i=1}^m \|\lambda_i(k+1) - v(k+1)\| < \infty$. Such a relationship can be exploited to show that the consensus error vanishes, i.e. $\lim_{k \rightarrow \infty} \|e_i(k)\| = 0$, and that consensus is achieved, i.e. $\lim_{k \rightarrow \infty} \|\lambda_i(k) - v(k)\| = 0$, for all $i = 1, \dots, m$. Furthermore, (12) can also be exploited to show convergence of the $\{\sum_i^m \|\lambda_i(k) - \lambda^*\|^2\}_{k \geq 0}$ sequence due to the deterministic version of the supermartingale convergence theorem [14] (Proposition 8.2.10, p. 489) and the fact that the last two terms in the right hand side of (12) are summable.

Once convergence has been proven it is sufficient to show that there exist a subsequence for which the quantity $\sum_i^m \|\lambda_i(k) - \lambda^*\|^2$ goes to 0 to prove Theorem 1. This is established by showing that the sequence $\{\lambda_i(k)\}_{k \geq 0}$ achieves the optimal value of the dual function across a subsequence, and relies on Proposition 4 of [7].

Finally, the proof of Theorem 2 follows from [24], extending its derivations to deal with the considered distributed context.

III. NUMERICAL EXAMPLE

In this section we present a numerical example to show the validity of the proposed approach. Consider a network of $m = 8$ agents connected as in Figure 1. For the sake of simplicity we assumed that the network does not change across iterations. Each agent i , $i = 1, \dots, m$, has $n_i = 2$ optimization variables, a local objective function defined as $f_i(x_i) = \xi_i^\top x_i$, and its own constraint set $X_i = [-5, 5] \times [-5, 5]$. The coefficients ξ_1, \dots, ξ_m are independently extracted at random from a Gaussian probability distribution with zero mean and covariance matrix equal to $25I_2$, I_2 being the identity matrix of order 2. Consider now a coupling constraint given by $\sum_{i=1}^m \|x_i\|^2 \leq b$, where $b = 25m$. The

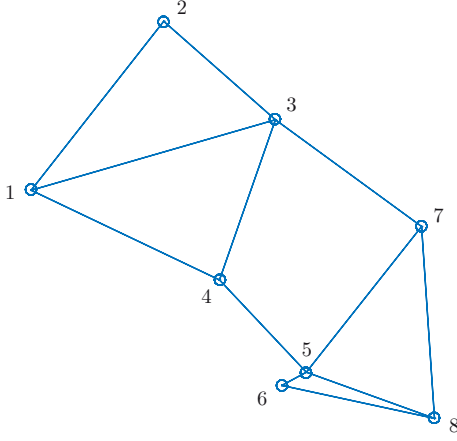


Fig. 1. Network of $m = 8$ agents.

resulting optimization program is given by

$$\begin{aligned} \min_{\{x_i \in X_i\}_{i=1}^m} & \sum_{i=1}^m f_i(x_i) \\ \text{subject to:} & \sum_{i=1}^m \|x_i\|^2 \leq b. \end{aligned} \quad (13)$$

To put (13) in the form of \mathcal{P} it suffices to set

$$g_i(x_i) = \|x_i\|^2 - \frac{b}{m}, \quad (14)$$

for all $i = 1, \dots, m$, which is quadratic and convex. Since problem (13) has a unique coupling constraint, there is just one Lagrange multiplier $\lambda \in \mathbb{R}_+$.

We ran the Algorithm 1 for 1000 iterations. Figure 2 shows the evolution of the agents' estimates $\lambda_i(k)$, $i = 1, \dots, m$, of the optimal value of λ (blue dotted lines) and their arithmetic average $v(k)$ (orange solid line), as the algorithm progresses. By

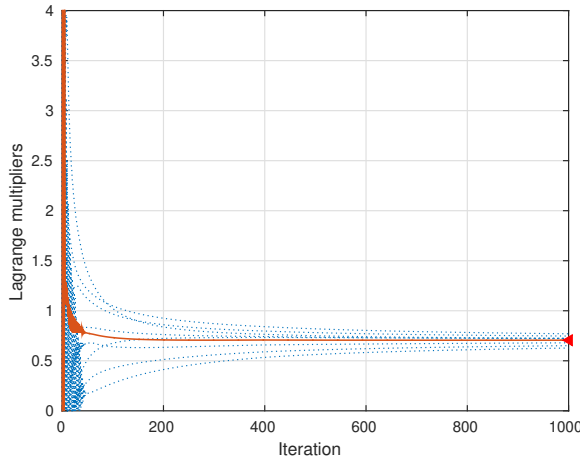


Fig. 2. Evolution of the agents' estimates $\lambda_i(k)$, $i = 1, \dots, m$, of the vector λ (blue dotted lines), and their arithmetic average $v(k)$ (orange solid line). Red triangles represent the optimal dual solution.

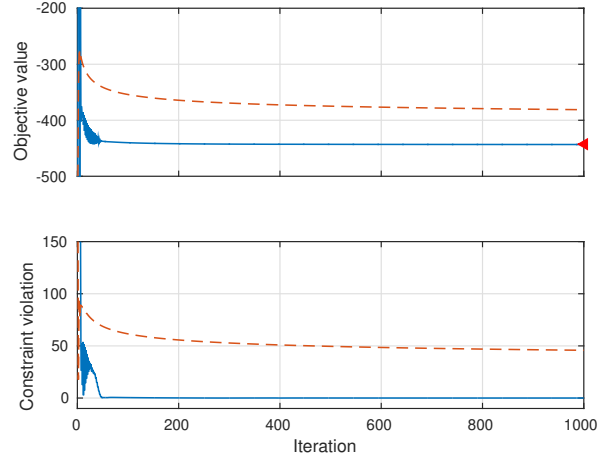


Fig. 3. Evolution of primal objective $\sum_{i=1}^m f_i(x_i)$ (upper plot) and constraint violation $\|\sum_{i=1}^m g_i(x_i)\|_\infty$ (lower plot) as a function of $x_i(k)$ (blue solid lines) and $\hat{x}_i(k)$ (orange dashed lines).

inspection of Figure 2, the average converge quite fast to the optimal Lagrange multipliers of (13) (red triangles), whereas all agents gradually reach consensus on those values.

Figure 3 shows the evolution of the primal objective value $\sum_{i=1}^m f_i(x_i)$ (upper plot), and constraint violation in terms of $\|\sum_{i=1}^m g_i(x_i)\|_\infty$ (lower plot), where x_i is replaced by two different sequences: $x_i(k)$ (blue solid lines), and $\hat{x}_i(k)$ (orange dashed lines), where the latter is given by (7).

For the sake of completeness we also show in Figure 4 the evolution of the sequences $\{x_i(k)\}_{k \geq 0}$ (solid lines), and $\{\hat{x}_i(k)\}_{k \geq 0}$ (dashed lines) for the first agent ($i = 1$); the evolution for other agents is similar.

IV. CONCLUDING REMARKS

In this paper, a novel distributed algorithm to deal with a class of convex optimization programs that exhibit a separable structure was developed. We considered an iterative

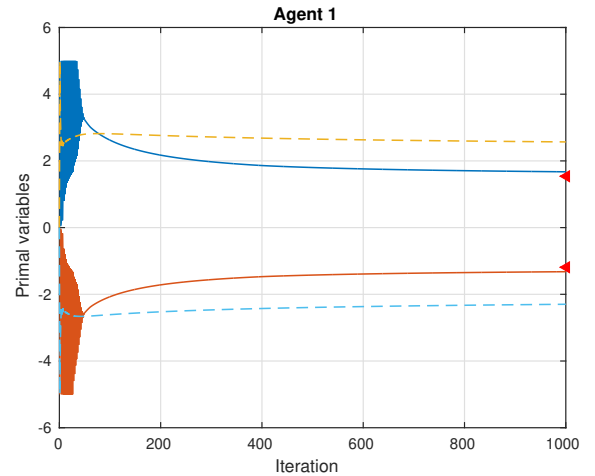


Fig. 4. Sequences $\{\hat{x}_1(k)\}_{k \geq 0}$ (dashed lines) and $\{\tilde{x}_1(k)\}_{k \geq 0}$ (solid lines). Red triangles represent the optimal primal solution for agent 1.

scheme based on a combination of dual decomposition and proximal minimization, and we showed that this scheme converges to some optimal dual solution of the centralized problem counterpart, while the primal iterates generated by the algorithm converge to the set of optimal primal solutions.

Current work concentrates on a convergence rate analysis and further comparison with gradient/subgradient methods. Moreover, we aim at relaxing the convexity assumption by extending the results of [29] to a distributed set-up, quantifying the duality gap incurred in case of mixed-integer programs. From an application point of view, the main focus is on applying the proposed algorithm to the problem of energy efficient control of a building network [30].

REFERENCES

- [1] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Transactions on Automatic Control*, vol. 60, no. 4, pp. 966–981, 2015.
- [2] Y. Zhang and G. Giannakis, "Distributed stochastic market clearing with high-penetration wind power and large-scale demand response," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 895–906, 2016.
- [3] G. Mateos and G. Giannakis, "Distributed recursive least-squares: Stability and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3740–3754, 2012.
- [4] B. Baingana, G. Mateos, and G. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 563–575, 2014.
- [5] S. Martinez, F. Bullo, J. Cortez, and E. Frazzoli, "On synchronous robotic networks - Part I: Models, tasks, and complexity," *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2199–2213, 2007.
- [6] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [7] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
- [8] S. Ram, A. Nedic, and V. Veeravalli, "A new class of distributed optimization algorithm: Application of regression of distributed data," *Optimization Methods & Software*, vol. 27, no. 1, pp. 71–88, 2012.
- [9] S. Lee and A. Nedic, "Distributed random projection algorithm for convex optimization," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 221–229, 2013.
- [10] K. Margellos, A. Falsone, S. Garatti, and M. Prandini, "Distributed constrained optimization and consensus in uncertain networks via proximal minimization," *IEEE Transactions on Automatic Control*, under review, pp. 1–15, 2016. [Online]. Available: <http://arxiv.org/abs/1603.02239>
- [11] B. Yang and M. Johansson, "Distributed Optimization and Games: a Tutorial Overview," *Chapter I in Networked Control Systems*, Springer London, pp. 109–148, 2010.
- [12] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [13] A. Nedic and D. Bertsekas, "Incremental subgradient methods for non-differentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.
- [14] D. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex analysis and optimization*. Athena Scientific, 2003.
- [15] D. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Mathematical Programming*, vol. 129, no. 6, pp. 163–195, 2011.
- [16] P. Bianchi, W. Hachem, and F. Iutzeler, "Coordinate descent primal-dual algorithm and application to distributed asynchronous optimization," *IEEE Transactions on Automatic Control (to appear)*, pp. 1–11, 2016.
- [17] I. Notarnicola and G. Notarstefano, "Asynchronous distributed optimization via randomized dual proximal gradient," *arXiv preprint arXiv:1509.08373*, 2015.
- [18] P. Bianchi, W. Hachem, and F. Iutzeler, "A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization," *arXiv preprint arXiv:1407.0898*, 2014.
- [19] P. Richtárik and M. Takáč, "Parallel coordinate descent methods for big data optimization," *Mathematical Programming*, vol. 156, no. 1-2, pp. 433–484, 2016.
- [20] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, 2015.
- [21] M. Zhu and S. Martinez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.
- [22] T.-H. Chang, A. Nedic, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *Automatic Control, IEEE Transactions on*, vol. 59, no. 6, pp. 1524–1538, 2014.
- [23] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition and proximal minimization for multi-agent distributed optimization with coupling constraints," *arXiv preprint arXiv:1607.00600*, 2016.
- [24] A. Nedic and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.
- [25] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [26] J. Tsitsiklis, *Problems in decentralized decision making and computation*. Ph.D. Dissertation, MIT, Cambridge, MA, 1984.
- [27] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.
- [28] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed convergence and averaging," *SIAM Review*, vol. 53, no. 4, pp. 747–772, 2011.
- [29] R. Vujanic, P. Mohajerin, P. Goulart, S. Mariethoz, and M. Morari, "A decomposition method for large scale MILPs, with performance guarantees and a power system application," *Automatica*, vol. 67, no. 5, pp. 144–156, 2016.
- [30] D. Ioli, A. Falsone, and M. Prandini, "Optimal energy management of a building cooling system with thermal storage: A convex formulation," *IFAC International Symposium on Advanced Control of Chemical Processes*, pp. 1151 – 1156, 2015.