A Distributed Primal Decomposition Scheme for Nonconvex Optimization

Published Version:
Camisa, A., Notarstefano, G. (2019). A Distributed Primal Decomposition Scheme for Nonconvex Optimization. RADARWEG 29, 1043 NX AMSTERDAM : Elsevier [10.1016/j.ifacol.2019.12.174].

Availability:
This version is available at: https://hdl.handle.net/11585/729714 since: 2020-02-20

Published:
DOI: http://doi.org/10.1016/j.ifacol.2019.12.174

A Distributed Primal Decomposition Scheme for Nonconvex Optimization

Andrea Camisa and Giuseppe Notarstefano

Department of Electrical, Electronic and Information Engineering
University of Bologna, Bologna, Italy
`a.camisa@unibo.it, giuseppe.notarstefano@unibo.it`

Abstract: In this paper, we deal with large-scale nonconvex optimization problems, typically arising in distributed nonlinear optimal control, that must be solved by agents in a network. Each agent is equipped with a local cost function, depending only on a local variable. The variables must satisfy private *nonconvex* constraints and global coupling constraints. We propose a distributed algorithm for the fast computation of a feasible solution of the nonconvex problem in finite time, through a distributed primal decomposition framework. The method exploits the solution of a convexified version of the problem, with restricted coupling constraints, to compute a feasible solution of the original problem. Numerical computations corroborate the results.

Keywords: Distributed optimization and MPC; multi-agent systems

1. INTRODUCTION

In distributed systems, several estimation, learning and control tasks can be formulated as optimization problems that must be solved by a network of processors without any central coordinator. A relevant scenario arises in distributed optimal control, where independent systems must be controlled, while satisfying coupling constraints. If the systems are nonlinear, the resulting optimal control problem is a nonconvex optimization problem, which is mainly an open research area.

We split the relevant references in two groups, namely distributed nonlinear optimal control and general distributed nonconvex optimization. As for the first group, in Necoara et al. (2009) a decentralized optimal control algorithm, based on linearized dynamics and dual decomposition, is proposed, whereas in Spedicato and Notarstefano (2018) a cloud-assisted distributed optimal control algorithm based on the projection operator for systems with coupled dynamics is considered. In Raimondo et al. (2009), an iterative decentralized Model Predictive Control (MPC) algorithm with local tube-based MPC controllers is considered. A suboptimal approach for distributed MPC of coupled systems, with linearization of the dynamics, is investigated in Grancharova and Johansen (2011), while Lucia et al. (2015) proposes a so-called contract-based distributed MPC for coupled systems. In Müller et al. (2012), a sequential distributed MPC algorithm for decoupled systems with coupling constraints is discussed.

Regarding general distributed nonconvex optimization, we first consider the case in which nonconvexity is in the cost function. In Bianchi and Jakubowicz (2013) a distributed stochastic gradient method with gossip communication is analyzed. A distributed Frank-Wolfe algorithm with convergence rate analysis is provided in Wai et al. (2017), while Di Lorenzo and Scutari (2016) and Scutari and Sun (2019) propose distributed gradient tracking algorithms, based on successive convex approximations, for undirected and directed networks (respectively). A more general set-up occurs when nonconvexity is also in the feasible set. In Zhu and Martínez (2013), a distributed dual subgradient method to obtain a an approximate solution is considered, whereas Dinh et al. (2013) proposes a parallel sequential quadratic programming algorithm when nonconvexity is in the feasible set only. Most works on distributed nonconvex optimization deal with problems that are coupled in the cost function, while only few references (e.g., Dinh et al. (2013)) consider the set-up of distributed control, in which the coupling among the systems is in the constraints.

The contributions of this paper are as follows. Motivated by distributed nonlinear optimal control, in this paper we focus on large-scale nonconvex optimization problems, where each variable has individual nonconvex constraints, and all the variables must satisfy coupling constraints. This set-up, with many optimization variables, subject to local and coupling constraints, is very challenging in a distributed framework, and nonconvexity makes it even more difficult to solve. Because of the problem complexity, the exact computation of an optimal solution is typically not affordable, and it is not always necessary when implemented in MPC schemes. We propose a distributed method that allows for the fast computation of a feasible (suboptimal) solution in finite time. To this end, we extend a primal decomposition framework, introduced in Camisa et al. (2018) for mixed-integer linear programs, to the *general nonconvex* set-up. The framework exploits a convexified version of the original problem, with tightened coupling constraints, to compute a solution satisfying both the nonconvex constraints and the coupling constraints of the original problem. Our distributed algorithm enjoys

⋆ This result is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 638992 - OPT4SMART).

finite-time feasibility of the solution, and opens the way to new distributed nonlinear optimal control frameworks. A major feature of our algorithm is that a globally feasible solution is obtained from small local nonconvex problems, that however do *not* need to be solved exactly: any feasible solution is sufficient. The theoretical discussion is followed by a numerical example to show its validity.

The paper is organized as follows. In Section 2, we formalize the optimization set-up and recall some preliminaries. In Section 3, we formalize our method to compute a feasible solution of the nonconvex optimization problem. The distributed algorithm is then formalized in Section 4, while Section 5 provides numerical computations to highlight the main features of the algorithm. For the sake of space, all the proofs are omitted and will be provided in a forthcoming document.

2. OPTIMIZATION SET-UP AND PRELIMINARIES

In this section we introduce the set-up considered in the paper. First, we formalize a motivating distributed control scenario. The general optimization set-up is then formalized and a relaxation of the optimization problem is considered. Finally, a decomposition method for the relaxed problem is reviewed.

2.1 Motivation: Distributed Nonlinear Optimal Control

Let us consider a network of N dynamical systems. Each system i is assumed to have its own *nonlinear* dynamics $z_i(k+1) = h_i(z_i(k), u_i(k))$, where $z_i(k) \in \mathbb{R}^{z_i}$ denotes the system's state at time $k \in \mathbb{N}$, $u_i(k) \in \mathbb{R}^{u_i}$ denotes the input fed to the system at time k and $h_i : \mathbb{R}^{z_i+u_i} \to \mathbb{R}^{u_i}$ is a continuous function. We assume that each system i initially has state equal to $z_i(0)$ and we assume the state and the inputs have to satisfy constraints $z_i \in \mathcal{Z}_i, u_i \in \mathcal{U}_i$ for some compact sets $\mathcal{Z}_i, \mathcal{U}_i$. We further assume that the system states and/or must satisfy coupling constraints that are, e.g., linear, $\sum_{i=1}^{N} \big(Z_i z_i(k) + U_i u_i(k)\big) \le b$, where $b \in \mathbb{R}^S$ and the matrices Z_i, U_i have appropriate dimensions. The goal for the agents is to compute a control input by solving an optimal control problem, involving the whole group of dynamical systems, of the form

$$
\begin{aligned}
\min_{\{z_i(k+1),u_i(k)\}_{k,i}} \quad & \sum_{i=1}^{N} \sum_{k=0}^{K-1} \ell_i\big(z_i(k), u_i(k)\big) + V_i\big(z_i(K)\big) \\
\text{subj. to} \quad & z_i(k+1) = h_i(z_i(k), u_i(k)), \qquad \forall\, k, i \\
& z_i(k+1) \in \mathcal{Z}_i, \;\; u_i(k) \in \mathcal{U}_i, \qquad \forall\, k, i \\
& \sum_{i=1}^{N} \big(Z_i z_i(k) + U_i u_i(k)\big) \le b, \qquad \forall\, k,
\end{aligned} \tag{1}
$$

where $K \in \mathbb{N}$ is the *prediction horizon*, $\ell_i : \mathbb{R}^{z_i+u_i} \to \mathbb{R}$ is the i-th system stage cost and $V_i : \mathbb{R}^{z_i} \to \mathbb{R}$ is the i-th terminal cost. In distributed MPC, problem (1) is repeatedly solved in a receding horizon fashion.

It is important to note that problem (1), in general, is nonconvex. Indeed, even when the stage cost ℓ_i and the functions h_i are convex, the feasible set is nonconvex due to the constraints $z_i(k+1) = h_i(z_i(k), u_i(k))$. In order to lighten the notation and give a general discussion, in the next subsection we formalize a distributed optimization set-up that encloses (1) as a special case.

2.2 Distributed Nonconvex Optimization Set-up

Let us consider a network of N agents that aim to cooperatively solve the optimization problem

$$
\begin{aligned}
\min_{\mathbf{x}_1,\dots,\mathbf{x}_N} \quad & \sum_{i=1}^{N} f_i(\mathbf{x}_i) \\
\text{subj. to} \quad & \sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0} \\
& \mathbf{x}_i \in X_i, \qquad i \in \{1, \dots, N\},
\end{aligned} \tag{2}
$$

where $\mathbf{x}_1 \in \mathbb{R}^{n_1}, \dots, \mathbf{x}_N \in \mathbb{R}^{n_N}$ are the decision variables (one for each agent), and each $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is the cost function associated to $\mathbf{x}_i$. Each variable $\mathbf{x}_i$ must satisfy individual constraints $\mathbf{x}_i \in X_i$, where $X_i \subset \mathbb{R}^{n_i}$ can be nonconvex. Moreover, the variables are intertwined by means of $S \in \mathbb{N}$ coupling constraints $\sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0}$, where each $\mathbf{g}_i : \mathbb{R}^{n_i} \to \mathbb{R}^S$ is used to model the contribution of $\mathbf{x}_i$ to the coupling constraints, and $\mathbf{0}$ denotes the vector with all zero entries. We use the symbols $\le$ and $\ge$ to denote component-wise inequalities. We focus on large-scale instances of problem (2) where the number of agents is considerably larger than the number of coupling constraints, i.e., $N \gg S$, which is a challenging scenario in distributed control applications.

The optimal control problem (1) is a special case of problem (2): indeed, each variable $\mathbf{x}_i$ consists of the trajectory $\{z_i(k+1), u_i(k)\}_k$ associated to the i-th dynamical system, the sets X_i take into account the i-th system dynamics (i.e., X_i contains the constraints $z_i(k+1) = h_i(z_i(k), u_i(k))$, $z_i(k) \in \mathcal{Z}_i$, $u_i(k) \in \mathcal{U}_i$), and the functions $\mathbf{g}_i$ encode the coupling among the systems.

Throughout the paper, we assume that each agent i knows only its local constraint X_i, its local cost function f_i and its own contribution $\mathbf{g}_i$ to the coupling constraints. We make the following standing assumption.

Assumption 1. Problem (2) is feasible. Moreover, for all $i \in \{1, \dots, N\}$, *(i)* the set X_i is compact, *(ii)* the function f_i and each component of $\mathbf{g}_i$ are convex. □

The analysis carried out in this paper can be extended to the case in which f_i and $\mathbf{g}_i$ are not convex, by using their convex closure (this is subject of current investigation). In the remainder of this section, we consider a convex relaxation of problem (2) and we recall a decomposition strategy that allows for the design of distributed algorithms to solve the convexified problem.

2.3 Problem Relaxation

Let us now consider the following relaxed version of problem (2),

$$
\begin{aligned}
\min_{\mathbf{x}_1,\dots,\mathbf{x}_N} \quad & \sum_{i=1}^{N} f_i(\mathbf{x}_i) \\
\text{subj. to} \quad & \sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0} \\
& \mathbf{x}_i \in \operatorname{conv}(X_i), \qquad i \in \{1, \dots, N\},
\end{aligned} \tag{3}
$$

where $\operatorname{conv}(X_i)$ denotes the convex hull of X_i. Thus, under Assumption 1, problem (3) is convex.

The significance of problem (3) lies in two facts. On the one hand, there exist distributed methods for the solution of convex problems in the form (3). On the other hand, it can be shown that its optimal solutions have a very special structure that pave the way for a distributed suboptimal approach for problem (2). Indeed, under the assumption of uniqueness, the optimal solution of the convex problem (3) largely satisfies the nonconvex constraints $\mathbf{x}_i \in X_i$, except for a small number of solution portions. The following theorem formalizes this fact.

Proposition 2. Let Assumption 1 hold and let $(\mathbf{x}_1^\star, \ldots, \mathbf{x}_N^\star)$ be the unique solution of problem (3). There exists a set $I \subset \{1, \ldots, N\}$, with cardinality at most $S+1$, such that

$$\begin{aligned} \mathbf{x}_i^\star &\notin X_i, \qquad \text{for all } i \in I, \text{ and} \\ \mathbf{x}_i^\star &\in X_i, \qquad \text{for all } i \notin I. \end{aligned} \qquad \square$$

Notice that if the optimal control problem (1) has quadratic cost with positive definite matrices, the assumption of unique optimal solution of (3) is readily satisfied.

2.4 Primal Decomposition Review

In this subsection, we recall a strategy that allows for the decomposition of problem (3). Such method allows us to utilize the distributed algorithm in Notarnicola and Notarstefano (2019) to solve the relaxed problem (3).

Specifically, we consider a *primal* decomposition scheme, also called right-hand side allocation (Silverman (1972); Bertsekas (1999)). With this approach, the coupling constraints $\sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0}$ are treated as a (given) limited resource. Then, a two-level hierarchical structure is formulated, with N independent *subproblems* having a fixed, local allocation of resource, and a *master problem*, coordinating the overall resource allocation process.

Formally, for all $i \in \{1, \ldots, N\}$ and $\mathbf{y}_i \in \mathbb{R}^S$, the i-th *subproblem* is

$$\begin{aligned} p_i(\mathbf{y}_i) = \min_{\mathbf{x}_i} \;& f_i(\mathbf{x}_i) \\ \text{subj. to } & \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i \\ & \mathbf{x}_i \in \text{conv}(X_i), \end{aligned} \tag{4}$$

where $\mathbf{y}_i$ is a local *allocation* for subproblem i and $p_i : \mathbb{R}^S \to \mathbb{R}$ is the function associating each $\mathbf{y}_i$ to the optimal cost of the corresponding subproblem. The local allocations are coordinated by a *master* problem, i.e.,

$$\begin{aligned} \min_{\mathbf{y}_1, \ldots, \mathbf{y}_N} \;& \sum_{i=1}^N p_i(\mathbf{y}_i) \\ \text{subj. to } & \sum_{i=1}^N \mathbf{y}_i = \mathbf{0} \\ & \mathbf{y}_i \in Y_i, \qquad i \in \{1, \ldots, N\}, \end{aligned} \tag{5}$$

where, for all $i \in \{1, \ldots, N\}$, the set $Y_i \subseteq \mathbb{R}^S$ is the domain of the local allocations for subproblem i (i.e., the set of $\mathbf{y}_i$ that make problem (4) feasible). The following lemma establishes the equivalence between problems (5) and (3).

Lemma 2.1. (Silverman (1972)). Let Assumption 1 hold. Then, problems (3) and (5) are equivalent, in the sense that *(i)* the optimal costs are equal, *(ii)* if $(\mathbf{y}_1^\star, \ldots, \mathbf{y}_N^\star)$ is an optimal solution of (5) and $\mathbf{x}_i^\star$ is an optimal solution of (4) with $\mathbf{y}_i = \mathbf{y}_i^\star$ for all i, then $(\mathbf{x}_1^\star, \ldots, \mathbf{x}_N^\star)$ is an optimal solution of (3). $\square$

There is a nice correspondence between the primal decomposition scheme and the distributed information structure associated to problem (3). Indeed, each subproblem i involves only the information that is known to agent i, and if the agents are provided with their own optimal local allocation, by Lemma 2.1 they are able to compute their own portion of optimal solution of problem (3).

3. APPROACH FOR NONCONVEX SOLUTION

In this section, we extend a framework for feasible solution, formerly introduced in Camisa et al. (2018) for mixed-integer linear programs, to the nonconvex set-up (2). The method is a building block for the distributed algorithm in Section 4. First, we introduce a modified version of problem (3), then we formalize and analyze our method.

Approach idea: adapt the solution of (3), which satisfies most of the local nonconvex constraints (cf. Proposition 2), in order to obtain a feasible solution of (2).

3.1 Restricted Problem

By exploiting the result in Proposition 2, the main idea is to change only those portions of optimal solution of (3) that do not *already* satisfy the local nonconvex constraints. This is obtained via a local correction procedure executed at each node (see Section 3.2). Since, as seen in Section 3.2, the corrected solution may result into a violation of the coupling constraints $\sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0}$, we replace problem (3) with a *restricted* version,

$$\begin{aligned} \min_{\mathbf{x}_1, \ldots, \mathbf{x}_N} \;& \sum_{i=1}^N f_i(\mathbf{x}_i) \\ \text{subj. to } & \sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \le -\boldsymbol{\sigma} \\ & \mathbf{x}_i \in \text{conv}(X_i), \qquad i \in \{1, \ldots, N\}, \end{aligned} \tag{6}$$

where the value of the restriction vector $\boldsymbol{\sigma} \in \mathbb{R}^S$, with non-negative components, will be specified in Section 3.3. In order to apply Proposition 2 to problem (6), we make the following assumption.

Assumption 3. Problem (6) is feasible and its optimal solution is unique. $\square$

To formulate the primal decomposition scheme for problem (6), it is sufficient to replace the master problem (5) with the following restricted version, i.e.,

$$\begin{aligned} \min_{\mathbf{y}_1, \ldots, \mathbf{y}_N} \;& \sum_{i=1}^N p_i(\mathbf{y}_i) \\ \text{subj. to } & \sum_{i=1}^N \mathbf{y}_i = -\boldsymbol{\sigma} \\ & \mathbf{y}_i \in Y_i, \qquad i \in \{1, \ldots, N\}, \end{aligned} \tag{7}$$

where the subproblems have the form (4). Clearly, Lemma 2.1 holds true also for problems (7) and (6).

3.2 Local Procedure for Solution Correction

In this subsection, we formalize the local procedure to correct the local solutions at each node. This procedure is a building block for the distributed algorithm in Section 4. We assume that each agent i is provided with a local allocation, which we denote by $\mathbf{y}_i^{\text{END}} \in \mathbb{R}^S$. We will show in the next that, even though the agents may violate the restricted coupling constraints of (6), an appropriate choice of $\boldsymbol{\sigma}$ will result in a feasible solution of the original problem (2). The local procedure is formalized in the following table from the perspective of agent i.

Local Procedure GET-NONCONVEX-SOL

Input: $\mathbf{y}_i^{\text{END}}$, i-th allocation

Compute $\mathbf{x}_i^{\text{CONV}}$ as optimal solution of (4) with $\mathbf{y}_i^{\text{END}}$

If $\mathbf{x}_i^{\text{CONV}} \in X_i$ **then output** $\mathbf{x}_i^{\text{OUT}} = \mathbf{x}_i^{\text{CONV}}$

Else

Compute $\mathbf{x}_i^{\text{NC}}$ as a feasible solution of

$$\begin{aligned} \min_{\mathbf{x}_i} \;\; & f_i(\mathbf{x}_i) \\ \text{subj. to } \; & \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i^{\text{END}} \\ & \mathbf{x}_i \in X_i \end{aligned} \tag{8}$$

If (8) is feasible **then output** $\mathbf{x}_i^{\text{OUT}} = \mathbf{x}_i^{\text{NC}}$

Else output $\mathbf{x}_i^{\text{OUT}} = \mathbf{x}_i^{\text{NC\_VIOL}}$ as a feasible solution of

$$\begin{aligned} \min_{\mathbf{x}_i} \;\; & f_i(\mathbf{x}_i) \\ \text{subj. to } \; & \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i^{\text{END}} + \rho_i \mathbf{1} \\ & \mathbf{x}_i \in X_i \end{aligned} \tag{9}$$

with minimal violation $\rho_i > 0$

End If

End If

The procedure always yields a vector satisfying (by construction) the local nonconvex constraints $\mathbf{x}_i \in X_i$. It is completely local: no communication is needed to perform the computation, and the connection with the rest of the problem depends on the input $\mathbf{y}_i^{\text{END}}$, which is provided by the distributed algorithm.

Let us comment on the structure of GET-NONCONVEX-SOL. The computation of $\mathbf{x}_i^{\text{CONV}}$ requires the solution of a convex problem, and for most of the agents (cf. Proposition 2) the procedure will end with $\mathbf{x}_i^{\text{CONV}}$ without solving any nonconvex problem. If $\mathbf{x}_i^{\text{CONV}}$ does not satisfy the constraints, the agents first try to solve problem (8), a nonconvex version of the local problem (4). However, since $X_i \subset \text{conv}(X_i)$, problem (8) might be infeasible. In this case, problem (9) is solved instead, where a violation $\rho\mathbf{1}$ of the local coupling constraints $\mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i^{\text{END}}$ is allowed. The solution of problem (9), to be carried out after a minimal violation ρ_i is determined, will result into a (controlled) violation of the local allocation and will produce a global violation of the coupling constraints. In order to compute ρ_i, agents can solve the problem

$$\begin{aligned} \min_{\mathbf{x}_i, \rho_i} \;\; & \rho_i \\ \text{subj. to } \; & \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i^{\text{END}} + \rho_i \mathbf{1} \\ & \mathbf{x}_i \in X_i. \end{aligned}$$

An optimal solution of the nonconvex problems (8) and (9) is desirable to improve the overall cost, however note that a feasible solution is sufficient for the distributed algorithm to produce a feasible solution to the original problem.

3.3 Restriction Vector and Preliminary Analysis

In order to make sure the overall solution $(\mathbf{x}_1^{\text{OUT}}, \ldots, \mathbf{x}_N^{\text{OUT}})$ is feasible for the original problem (2), it is necessary to guarantee that the coupling constraints are satisfied. The restriction vector $\boldsymbol{\sigma}$ is meant to compensate for possible violations of the local coupling constraints, so that, overall, it holds $\sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i^{\text{OUT}}) \le \mathbf{0}$.

For all $i \in \{1, \ldots, N\}$, let us define a vector $\mathbf{L}_i$, representing a resource *lower bound*, with components

$$\big[\mathbf{L}_i\big]_s \triangleq \min_{\mathbf{x}_i \in \text{conv}(X_i)} \big[\mathbf{g}_i(\mathbf{x}_i)\big]_s \qquad s \in \{1, \ldots, S\},$$

where the notation $[\cdot]_s$ indicates the s-th component of the vector in brackets. Then, let us define $\mathbf{x}_i^{\text{L}} \in X_i$ as the vector with *minimal resource usage*, i.e.,

$$\mathbf{x}_i^{\text{L}} \in \operatorname*{argmin}_{\mathbf{x}_i \in X_i} \max_{s \in \{1,\ldots,S\}} \big[\mathbf{g}_i(\mathbf{x}_i) - \mathbf{L}_i\big]_s.$$

Note that, by construction, the optimal cost of the preceding problem represents a minimal allocation required by agent i to compute a feasible vector in X_i. A restriction vector $\boldsymbol{\sigma}$, representing the worst-case overall violation, is

$$\boldsymbol{\sigma} = (S+1) \cdot \max_{i \in \{1,\ldots,N\}} \max_{s \in \{1,\ldots,S\}} \big[\mathbf{g}_i(\mathbf{x}_i^{\text{L}}) - \mathbf{L}_i\big]_s \mathbf{1}, \tag{10}$$

where the inner maximization is the worst-case violation for each agent i and the coefficient $S+1$ is due to the maximum number of agents that can simultaneously violate (by Proposition 2). We point out that the restriction (10) can be computed in a distributed way by using a max-consensus algorithm.

To conclude this section, we give a preliminary analysis of the framework introduced so far. To this end, let us consider the simplified case in which the local procedure is initialized with the optimal solution of problem (7) (the general case is discussed in Section 4.2). The first result of the paper is summarized in the next theorem, by which we assess that the solution computed by GET-NONCONVEX-SOL under the preceding assumption is feasible for the original problem (2).

Theorem 4. Let Assumptions 1 and 3 hold and let $(\mathbf{y}_1^\star, \ldots, \mathbf{y}_N^\star)$ be an optimal solution of problem (7), with $\boldsymbol{\sigma}$ equal to (10). Let $\mathbf{x}^{\text{OUT}} = (\mathbf{x}_1^{\text{OUT}}, \ldots, \mathbf{x}_N^{\text{OUT}})$, where each $\mathbf{x}_i^{\text{OUT}}$ is the output of GET-NONCONVEX-SOL with input $\mathbf{y}_i^\star$. Then, $\mathbf{x}^{\text{OUT}}$ is feasible for the original problem (2). $\square$

4. DISTRIBUTED PRIMAL DECOMPOSITION FOR NONCONVEX PROBLEMS

In the remainder of the paper, we formalize our distributed algorithm to compute a feasible solution to the nonconvex problem (2). The algorithm is obtained by integrating the framework of Section 3 with the distributed algorithm in Notarnicola and Notarstefano (2019). First, we formally describe the algorithm and implementation features. Then, we discuss its theoretical properties.

4.1 Algorithm Description and Implementation

Let us consider a network of N agents communicating according to a *connected* and *undirected* graph $\mathcal{G} = (\{1, \ldots, N\}, \mathcal{E})$, where $\mathcal{E} \subseteq \{1, \ldots, N\} \times \{1, \ldots, N\}$ is the set of edges. If $(i, j) \in \mathcal{E}$, then agents i and j can exchange information (and in fact also $(j,i) \in \mathcal{E}$). We denote by $\mathcal{N}_i$ the set of *neighbors* of agent i in $\mathcal{G}$, i.e., $\mathcal{N}_i = \{j \in \{1,\dots,N\} \mid (i,j) \in \mathcal{E}\}$.

To achieve finite-time feasibility of the sequences computed by the distributed algorithm, we consider the following problem,

$$\begin{aligned} \min_{\mathbf{x}_1,\dots,\mathbf{x}_N} \quad & \sum_{i=1}^N f_i(\mathbf{x}_i) \\ \text{subj. to} \quad & \sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \le -(\boldsymbol{\sigma} + \delta \mathbf{1}) \\ & \mathbf{x}_i \in \text{conv}(X_i), \qquad i \in \{1,\dots,N\}, \end{aligned} \tag{11}$$

which is obtained from problem (6) by further restricting the coupling constraint components by an arbitrary $\delta > 0$. We will assume that Assumption 3 holds true for problem (11), whose primal decomposition can be formulated by properly adapting problem (7).

As for the notation, $t \in \mathbb{N}$ denotes a universal time index, while GET-NONCONVEX-SOL$(\mathbf{y}_i^t)$ is the output of the local procedure with input equal to $\mathbf{y}_i^t$, the i-th allocation at time t. In the following table we summarize our Distributed Primal Decomposition for Nonconvex problems (DiP-Nonconvex) from the perspective of agent i, where $M > 0$ is a scalar and $\{\alpha^t\}_{t\ge0}$ is the step-size sequence.

Distributed Algorithm DiP-Nonconvex

Initialization: $\mathbf{y}_i^0$ such that $\sum_{i=1}^N \mathbf{y}_i^0 = -(\boldsymbol{\sigma} + \delta \mathbf{1})$

Evolution: FOR $t = 0, 1, \dots$

Compute $\boldsymbol{\mu}_i^t$ as a Lagrange multiplier of

$$\begin{aligned} \min_{\mathbf{x}_i,\rho_i} \quad & f_i(\mathbf{x}_i) + M\rho_i \\ \text{subj. to} \quad & \boldsymbol{\mu}_i : \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{y}_i^t + \rho_i \mathbf{1} \\ & \mathbf{x}_i \in \text{conv}(X_i), \;\; \rho_i \ge 0 \end{aligned} \tag{12}$$

Gather $\boldsymbol{\mu}_j^t$ from $j \in \mathcal{N}_i^t$ and update

$$\mathbf{y}_i^{t+1} = \mathbf{y}_i^t + \alpha^t \sum_{j \in \mathcal{N}_i^t} \left(\boldsymbol{\mu}_i^t - \boldsymbol{\mu}_j^t\right) \tag{13}$$

Compute $\mathbf{x}_i^t$ as

$$\mathbf{x}_i^t = \text{GET-NONCONVEX-SOL}(\mathbf{y}_i^t) \tag{14}$$

The algorithm is fully distributed, in the sense that at every iteration t the computation performed by each agent i involves only local information and the information gathered from its neighbors to perform (13). The algorithm is initialized such that $\sum_{i=1}^N \mathbf{y}_i^0 = -(\boldsymbol{\sigma} + \delta\mathbf{1})$, in order to take into account both restrictions $\boldsymbol{\sigma}$ and $\delta\mathbf{1}$. The first two steps (12)-(13) implement the distributed algorithm in Notarnicola and Notarstefano (2019), while the last step (14) implements the framework of Section 3 to compute a solution satisfying the nonconvex constraints. The satisfaction of the coupling constraints is taken into account by the restriction in the initialization of the algorithm. A remarkable property of our algorithm is that it only requires the solution of convex problems in order to evolve (indeed problem (12) is convex), while, as already noted in Section 3.2, nonconvex problems in GET-NONCONVEX-SOL can also be solved suboptimally.

In order to carry out the step (12) numerically, agents may need an explicit description of $\text{conv}(X_i)$ in terms of inequalities. This issue is problem dependent, and a couple of special cases are discussed here. If the cost f_i and the coupling constraint function $\mathbf{g}_i$ are both linear, a Lagrange multiplier of problem (12) can be computed by locally running a dual subgradient algorithm involving the solution of small nonconvex problems. In Section 5, we instead consider a numerical example in which the description of $\text{conv}(X_i)$ can be obtained by simply replacing equalities with inequalities.

4.2 Theoretical Properties of the Algorithm

In order to formulate the theoretical results, we make the following assumption on problem (11).

Assumption 5. There exist vectors $\bar{\mathbf{x}}_1 \in \text{conv}(X_1)$, ..., $\bar{\mathbf{x}}_N \in \text{conv}(X_N)$ such that $\sum_{i=1}^N \mathbf{g}_i(\bar{\mathbf{x}}_i) < -(\boldsymbol{\sigma} + \delta\mathbf{1})$. □

Assumption 5 is Slater's constraint qualification and is a standard condition for the application of duality. As for the step-size, we make the following assumption, which is common for duality-based algorithms.

Assumption 6. The step-size sequence $\{\alpha^t\}_{t\ge0}$, with each $\alpha^t \ge 0$, satisfies $\sum_{t=0}^\infty \alpha^t = \infty$, $\sum_{t=0}^\infty \left(\alpha^t\right)^2 < \infty$. □

Under the preceding assumptions, we are able to show the following theorem, in which we assess that in finite time the solution sequence computed by DiP-Nonconvex is feasible for the original problem (2).

Theorem 7. Let Assumptions 1, 5 and 6 hold and let Assumption 3 hold for problem (11), where $\boldsymbol{\sigma}$ is equal to (10) and $\delta > 0$ is arbitrary. Moreover, let the local allocation vectors $\mathbf{y}_i^0$ be initialized such that $\sum_{i=1}^N \mathbf{y}_i^0 = -(\boldsymbol{\sigma} + \delta\mathbf{1})$. Then, there exists a sufficiently large $M > 0$ and $T_\delta > 0$ for which DiP-Nonconvex generates a sequence $\{\mathbf{x}_1^t, \dots, \mathbf{x}_N^t\}_{t\ge0}$ such that the vector $(\mathbf{x}_1^t, \dots, \mathbf{x}_N^t)$ is a feasible solution for problem (2) for all $t \ge T_\delta$. □

Finite-time feasibility of DiP-Nonconvex is an appealing feature for model predictive control applications, since it can ensure recursive feasibility of the control algorithm. As for the parameter M, it must be greater than $\|\boldsymbol{\mu}^\star\|_1$, where $\boldsymbol{\mu}^\star$ denotes a dual optimal solution of problem (11) (see Notarnicola and Notarstefano (2019)). In practice, it suffices to choose M sufficiently large.

Remark 8. We point out that it is possible to compute guaranteed suboptimality bounds of the solution computed by DiP-Nonconvex. For the sake of space, we omit the discussion, but an explicit formula can be obtained by extending the results in Camisa et al. (2018) to the general nonconvex case. □

5. NUMERICAL EXAMPLE

In this section, we provide numerical computations performed with the MATLAB software to corroborate the theoretical results and to highlight the main features of our algorithm. We consider a simplified scenario in which we are able to express $\text{conv}(X_i)$ explicitly.

Formally, consider a network of $N = 50$ agents, whose aim is to cooperatively find a feasible solution to an optimal control of the type (1). For simplicity, we consider 1-step predictions of the dynamics. Each dynamical system i has 1-dimensional state and input, with dynamics $z_i(k+1) = z_i(k)^2 + q_i u_i(k)^2 + r_i$, where the parameters q_i and r_i are randomly drawn from $[1,5]$ and $[-4,0]$ respectively. The state and input constraints $\mathcal{Z}_i$ and $\mathcal{U}_i$ are box constraints (i.e., $z_i^{\text{LB}} \leq z_i(k) \leq z_i^{\text{UB}}$ and similarly for the set $\mathcal{U}_i$) where the lower and upper bounds have entries in $[-10,-5]$ and $[5,10]$ respectively. We assume that the systems are initialized in the origin, i.e., $z_i(0) = 0$ for all i. Therefore, the local nonconvex feasible set X_i is a clipped parabola in $\mathbb{R}^2$, and $\text{conv}(X_i)$ can be obtained by replacing the dynamics constraints with the inequality version $z_i(k+1) \geq q_i u_i(k)^2 + r_i$. The agents must further satisfy $S = 3$ coupling constraints, where the matrices Z_i and U_i have entries in $[0,1]$ and the vector b has entries in $[-3,7]$. As for the cost functions, we assume that $\ell_i(z_i, u_i)$ and $V_i(z_i)$ are linear with random entries in $[-5,5]$.

The communication graph is a random Erdős-Rényi graph with edge probability 0.2. A random problem has been generated, and a local minimum has been found using a centralized solver (FMINCON). In order to check whether the instance is meaningful, we make sure it has a duality gap by solving the dual problem with a dual subgradient algorithm. We perform a simulation of the distributed algorithm with $\delta = 1$. The restriction $\boldsymbol{\sigma}$ has ∞-norm equal to 3.5, and agents computed in finite time a feasible solution to the nonconvex problem (1) (as expected from Theorem 7). In Figure 1 the distributed utilization of the coupling constraints is shown, where $\mathbf{x}_i^t$ denotes the stack of the local optimization variables, obtained as the output of GET-NONCONVEX-SOL with allocation equal to $\mathbf{y}_i^t$, and $\mathbf{g}_i(\mathbf{x}_i) = U_i u_i(0) + Z_i z_i(1) - b/N$.

Notably, the solution is feasible since the first iteration of the distributed algorithm and has 19% suboptimality with respect to the solution computed by FMINCON.

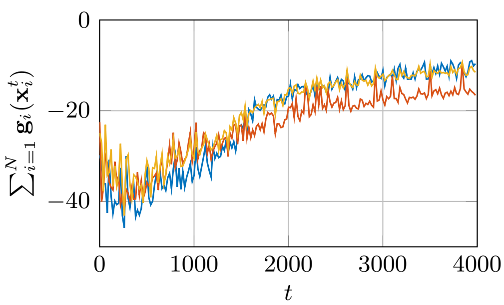

Fig. 1. Evolution of the coupling constraint utilization ($S = 3$ components). The solution computed by the algorithm is feasible for the coupling constraints since the first iteration, indeed the maximum value is below 0 in the whole graph.

6. CONCLUSIONS

In this paper, we considered a large-scale optimization set-up, arising in distributed nonlinear optimal control, in which the optimization variables must satisfy individual nonconvex constraints and coupling constraints. We proposed a distributed algorithm, based on a primal decomposition framework and on a convexified version of the problem, that computes a feasible solution of the original problem in a finite number of iterations. A numerical example shows the main features of the algorithm.

REFERENCES

Bertsekas, D.P. (1999). *Nonlinear programming*. Athena Scientific.

Bianchi, P. and Jakubowicz, J. (2013). Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization. *IEEE Transactions on Automatic Control*, 58(2), 391–405.

Camisa, A., Notarnicola, I., and Notarstefano, G. (2018). A primal decomposition method with suboptimality bounds for distributed mixed-integer linear programming. In *IEEE Conference on Decision and Control*, 3391–3396.

Di Lorenzo, P. and Scutari, G. (2016). Next: In-network nonconvex optimization. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2), 120–136.

Dinh, Q.T., Necoara, I., and Diehl, M. (2013). A dual decomposition algorithm for separable nonconvex optimization using the penalty function framework. In *IEEE Conference on Decision and Control*, 2372–2377.

Grancharova, A. and Johansen, T.A. (2011). Distributed quasi-nonlinear model predictive control by dual decomposition. *IFAC Proceedings Volumes*, 44(1), 1429–1434.

Lucia, S., Kögel, M., and Findeisen, R. (2015). Contract-based predictive control of distributed systems with plug and play capabilities. *IFAC-PapersOnLine*, 48(23), 205–211.

Müller, M.A., Reble, M., and Allgöwer, F. (2012). Cooperative control of dynamically decoupled systems via distributed model predictive control. *International Journal of Robust and Nonlinear Control*, 22(12), 1376–1397.

Necoara, I., Savorgnan, C., Tran, D.Q., Suykens, J., and Diehl, M. (2009). Distributed nonlinear optimal control using sequential convex programming and smoothing techniques. In *IEEE Conference on Decision and Control held jointly with Chinese Control Conference*, 543–548.

Notarnicola, I. and Notarstefano, G. (2019). Constraint-coupled distributed optimization: a relaxation and duality approach. *IEEE Transactions on Control of Network Systems*, PP(99), 1–10.

Raimondo, D.M., Hokayem, P., Lygeros, J., and Morari, M. (2009). An iterative decentralized MPC algorithm for large-scale nonlinear systems. *IFAC Proceedings Volumes*, 42(20), 162–167.

Scutari, G. and Sun, Y. (2019). Distributed nonconvex constrained optimization over time-varying digraphs. *Mathematical Programming*, 176(1-2), 497–544.

Silverman, G.J. (1972). Primal decomposition of mathematical programs by resource allocation: I – basic theory and a direction-finding procedure. *Operations Research*, 20(1), 58–74.

Spedicato, S. and Notarstefano, G. (2018). Cloud-assisted distributed nonlinear optimal control for dynamics over graph. *IFAC-PapersOnLine*, 51(23), 361–366.

Wai, H.T., Lafond, J., Scaglione, A., and Moulines, E. (2017). Decentralized frank–wolfe algorithm for convex and nonconvex problems. *IEEE Transactions on Automatic Control*, 62(11), 5522–5537.

Zhu, M. and Martínez, S. (2013). An approximate dual subgradient algorithm for multi-agent non-convex optimization. *IEEE Transactions on Automatic Control*, 58(6), 1534–1539.