# A distributed iterative algorithm for multi-agent MILPs: finite-time feasibility and performance characterization

Alessandro Falsone, Kostas Margellos, Maria Prandini

***Abstract*—We deal with decision making in a large-scale multi-agent system, where each agent aims at minimizing a local cost function subject to local constraints, and the local decision variables of all agents are coupled through a global constraint. We consider a cooperative framework where the multi-agent decision problem is formulated as a constrained optimization program with the sum of the local costs as global cost to be minimized with respect to the local decision variables of all agents, subject to both local and global constraints. We focus on a non-convex linear set-up where all costs and constraints are linear but local decision variables are discrete or include a discrete component, and propose a distributed iterative scheme based on dual decomposition and consensus to solve the resulting Mixed Integer Linear Program (MILP). Our approach extends recent results in the literature to a distributed set-up with a time-varying communication network and allows to: reduce the computational and communication effort, achieve resilience to communication failures, and also preserve privacy of local information. The approach is demonstrated on a numerical example of optimal charging of plug-in electric vehicles.**

***Index Terms*—Optimization algorithms, distributed control, agents-based systems**

## I. Introduction

We consider a system composed of $m$ agents, where each agent $i$, $i = 1, \dots, m$, is optimizing its local (scalar) cost function

$$J_i(x_i) = c_i^\top x_i$$

with respect to its decision vector $x_i$ of dimension $n_i$, which has its first $n_{c,i}$ components continuous and the remaining $n_{d,i} = n_i - n_{c,i}$ ones discrete. The decision vector $x_i$ is subject to the following constraints

$$D_i x_i \leq d_i,$$
$$\sum_{i=1}^m A_i x_i \leq b,$$

where $d_i$ and $b$ are vectors of dimension $k_i$ and $p$, respectively, matrices $D_i$ and $A_i$ have appropriate dimensions, and inequalities have to be interpreted component-wise. The first constraint is local, while the second one is global since it involves the decision variables of all the agents, which makes the decision making problem coupled.

We assume that agents act cooperatively and they hence jointly aim at solving the resulting Mixed Integer Linear Program (MILP):

$$\begin{aligned} \min_{x_1,\dots,x_m} \quad & \sum_{i=1}^m J_i(x_i) = \sum_{i=1}^m c_i^\top x_i \\ \text{subject to:} \quad & \sum_{i=1}^m A_i x_i \leq b, \\ & x_i \in X_i, \ i = 1, \dots, m, \end{aligned} \tag{$\mathcal{P}$}$$

where for each agent $i$, $i = 1, \dots, m$, we set

$$X_i = \{x_i \in \mathbb{R}^{n_{c,i}} \times \mathbb{Z}^{n_{d,i}} : D_i x_i \leq d_i\}, \tag{1}$$

and by $\mathbb{R}^{n_{c,i}} \times \mathbb{Z}^{n_{d,i}}$ we imply a vector space involving both continuous and discrete components.

When the number of agents is large, solving the constrained optimization problem $\mathcal{P}$ centrally is hard, mainly due to the presence of discrete decision variables. Also, agents may not be willing to disclose their private information coded in their local cost and constraint.

Methods for efficiently solving problem $\mathcal{P}$ by exploiting its almost-separable structure and decomposing it into $m$ (smaller) problems solved in parallel by the agents under the coordination of a central unit are proposed in [1], [2]. Both [1], [2] adopt a dualization approach with tightening of the coupling constraint to determine a feasible solution and provide a bound on the duality gap. The introduction of an adaptive tightening in [2] allows for the formulation of an iterative decentralized algorithm with finite-time feasibility properties and a performance that is not worse than the one of the asymptotically feasible solution in [1].

Optimal charging of a fleet of electric vehicles is considered as numerical case study in both papers. Optimization problems that fit the structure of $\mathcal{P}$ can be found in further application domains, like, e.g., supply chain management [3], portfolio optimization [4], and energy systems, see e.g. [5], [6]. In all applications, the agents decisions are implementable in practice only if they are feasible for the associated optimization problem $\mathcal{P}$.

In this paper we extend the decentralized algorithm in [2] to a distributed framework where agents exchange information

This work is partially supported by the European Commission under the project UnCoVerCPS with grant number 643921, and by EPSRC, U.K., under Grant EP/P03277X/1.

Alessandro Falsone and Maria Prandini are with Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, e-mail:{alessandro.falsone, 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

only with their neighbors and no central unit is present, while preserving its finite-time feasibility property and performance guarantees.

In [2], a central unit is in charge of enforcing the coupling constraint by iteratively solving the dual problem

$$\max_{\lambda \geq 0} \; -\lambda^\top (b-\rho) + \sum_{i=1}^{m} \min_{x_i \in X_i} (c_i^\top + \lambda^\top A_i) x_i, \qquad (\mathcal{D}_\rho)$$

of the tightened LP problem:

$$\begin{aligned} \min_{x_1,\dots,x_m} \quad & \sum_{i=1}^{m} c_i^\top x_i & (\mathcal{P}_{\mathrm{LP},\rho}) \\ \text{subject to:} \quad & \sum_{i=1}^{m} A_i x_i \leq b - \rho & \\ & x_i \in \operatorname{conv}(X_i), \; i = 1,\dots,m, & \end{aligned}$$

where $\operatorname{conv}(X_i)$ is the convex hull of all points of $X_i$ and the tightening coefficient $\rho \in \mathbb{R}^p$ satisfies $\rho \geq 0$ and is updated at every step based on the tentative primal solutions computed by the agents. More specifically, at each iteration, the central unit broadcasts a tentative solution for the dual problem, the agents solve in parallel the (smaller) MILP

$$x_i(\lambda) \in \arg \min_{x_i \in \operatorname{vert}(X_i)} (c_i^\top + \lambda^\top A_i) x_i, \qquad (2)$$

based on the broadcasted information, where $\operatorname{vert}(X_i)$ denotes the set of vertices of $\operatorname{conv}(X_i)$, and, then, the central unit collects the contribution of each agent to the joint constraint so as to update both the tightening coefficient $\rho$ and the dual variables $\lambda$.

Here, we eliminate the need of a central unit by integrating within the iterative decentralized scheme in [2] a max-consensus algorithm on the tightening coefficient and the distributed approach for updating the dual variables proposed in [7]. In [7] problems of a more general form than $\mathcal{P}$ but without discrete decision variables are addressed. No tightening is introduced. However, differently from the present paper where finite time feasibility is proven, in [7] feasibility is guaranteed only asymptotically.

The rest of the paper is organized as follows. In Section II the proposed distributed algorithm is described and the main results are stated together with the relevant assumptions. Proofs are confined to Section III. An application to electric vehicles optimal charging is presented in Section IV. Some concluding remarks are given in Section V.

## II. Proposed approach

In this section, we introduce the iterative Algorithm 1 for the distributed computation of a solution to $\mathcal{P}$, and analyze its properties.

Note that the operators $\max$ and $\min$ appearing in steps 10, 11, 12, and 13 of Algorithm 1 with arguments in $\mathbb{R}^p$ are meant to be applied component-wise. Moreover, if $\arg\min_{x_i \in \operatorname{vert}(X_i)} (c_i^\top + \ell_i(k)^\top A_i) x_i$ in step 9 is a set of cardinality larger than 1, then, a deterministic tie-break rule is applied to choose a value for $x_i(k+1)$. Finally, $[\,\cdot\,]_+$ in step 14 denotes the projection operator onto the $p$-dimensional non-negative orthant $\mathbb{R}^p_+$.

**Algorithm 1** Distributed algorithm for solving $\mathcal{P}$

1: **Initialization**
2: $\quad k = 0$
3: $\quad$ Set $\lambda_i(0) = 0$, for all $i = 1,\dots,m$
4: $\quad$ Set $\bar{s}_i(0) = -\infty$, for all $i = 1,\dots,m$
5: $\quad$ Set $\underline{s}_i(0) = +\infty$, for all $i = 1,\dots,m$
6: $\quad$ Set $\rho_i(0) = 0$, for all $i = 1,\dots,m$
7: **For** $i = 1,\dots,m$ **repeat**
8: $\quad \ell_i(k) = \sum_{j \in \mathcal{N}_i(k)} a^i_j(k) \lambda_j(k)$
9: $\quad x_i(k+1) \leftarrow \arg\min_{x_i \in \operatorname{vert}(X_i)} (c_i^\top + \ell_i(k)^\top A_i) x_i$
10: $\quad \varrho_i(k) = \max_{j \in \mathcal{N}_i(k)} \{\rho_j(k)\}$
11: $\quad \bar{s}_i(k+1) = \max\{\bar{s}_i(k), A_i x_i(k+1)\}$
12: $\quad \underline{s}_i(k+1) = \min\{\underline{s}_i(k), A_i x_i(k+1)\}$
13: $\quad \rho_i(k+1) = \max\{\varrho_i(k), p\,(\bar{s}_i(k+1) - \underline{s}_i(k+1))\}$
14: $\quad \lambda_i(k+1)$
$$= \Big[\ell_i(k) + \alpha(k)\Big(A_i x_i(k+1) - \tfrac{b - \rho_i(k+1)}{m}\Big)\Big]_+$$
15: $\quad k \leftarrow k+1$

Algorithm 1 integrates the distributed approach in [7] for problems of the form $\mathcal{P}$ but without discrete decision variables within the iterative decentralized scheme in [2], where adaptive tightening of the joint constraint is used to account for discrete decision variables and recover feasibility.

Three steps in Algorithm 1 are based on [7]: step 8, in which agent $i$ constructs an average $\ell_i(k)$ of the dual iterates $\lambda_j(k)$ received from his neighbors $j \in \mathcal{N}_i(k)$, weighted through the coefficients $a^i_j(k)$ that satisfy $a^i_j(k) > 0$ if and only if agent $j$ sends information to agent $i$ at time $k$ and are used to define the neighboring set as $\mathcal{N}_i(k) = \{j : a^i_j(k) > 0\}$; step 9, in which a tentative solution for the primal problem is computed by fixing the dual variables $\lambda$ in (2) to $\ell_i(k)$ and then performing the minimization; and step 14, which involves a dual subgradient step, based on the tentative primal solution, with step size equal to $\alpha(k)$, and a projection onto the non-negative orthant.

Four additional steps are based on [2]: steps 11 and 12, which iteratively construct worst-case contributions $A_i x_i(k+1)$ of agent $i$ to the joint constraint, based on its tentative primal solution $x_i(k+1)$; step 13, which updates the current value of the tightening coefficient $\rho_i(k+1)$ based on the new information of agent $i$; and step 14, in which the resource vector $b$ is reduced by the current tightening $\rho_i(k+1)$.

Step 10 in Algorithm 1 is the max-consensus update of the tightening coefficient, in which agent $i$ takes the maximum among its current tightening coefficient value $\rho_i(k)$ and those of its neighbors, i.e., $\rho_j(k)$, $j \in \mathcal{N}_i(k)$, $j \neq i$.

Thanks to steps 8 and 10, no central unit is needed to coordinate the agents, thus making the decentralized approach in [2] completely distributed. Note that parameter $b$ appearing in Algorithm 1 needs to be known to every agent in order to run the algorithm. However, this is not a restrictive condition, since $b$ appears in the coupling constraint of problem $\mathcal{P}$, which are due to shared resources with limited capacity, and quantifies the corresponding capacity limits.

As for the initialization of Algorithm 1, $\lambda_i(0)$ is set equal to 0 for all $i = 1,\dots,m$, so that $\ell_i(0) = 0$, $i = 1,\dots,m$, and each agent computes its locally optimal solution at step 9. As for $\rho_i(0)$, a sensible choice is to set it to 0 so as not to impose any a-priori defined tightening of the coupling constraint.

Finally, it is remarkable that the agents do not need to disclose any private information regarding the primal problem, but only their tentative Lagrange multipliers $\lambda_i(k)$ and tightening coefficient $\rho_i(k)$.

### *A. Statement of the main results*

Before stating Theorems 1 and 2 on the feasibility and performance guarantees of the solution to $\mathcal{P}$ computed by Algorithm 1, we need to introduce some quantities and assumptions, and a preliminary result that is instrumental to the proof of Theorems 1 and 2.

In order to guarantee that the solution to step 9 in Algorithm 1 is well-defined, we impose the following assumption on the agents local constraint sets.

**Assumption 1** (Boundedness)**.** *The sets $X_i$, $i = 1, \dots, m$, defined in* (1) *are bounded and non-empty.* □

For any agent $i$, $i = 1, \dots, m$, and $k \geq 0$, we introduce the following quantities:

$$y_i(k) = \max_{j \in \mathcal{N}_i(k)} \{\gamma_j(k)\} \tag{3}$$

$$\begin{aligned}\gamma_i(k+1)& \\ = \max &\Big\{ y_i(k), p\Big( \max_{r \leq k+1} c_i^\top x_i(r) - \min_{r \leq k+1} c_i^\top x_i(r) \Big) \Big\},\end{aligned} \tag{4}$$

where $\{x_i(r)\}_{r \geq 1}$ are the tentative primal solutions generated at step 9 and $\gamma_i(0) = 0$.

The following proposition states that after a finite number of iterations the agents reach consensus on the value of the tightening coefficient and of $\gamma_i(k)$ defined in (3)-(4).

**Proposition 1** (Max-consensus)**.** *Under Assumption 1, there exists a finite iteration index $K$ such that, for all $k \geq K$, $\rho_i(k) = \bar{\rho}$ and $\gamma_i(k) = \bar{\gamma}$ for all $i = 1, \dots, m$.* □

In order to prove that the integration of the distributed approach in [7] within Algorithm 1 is effective, we need to introduce some assumptions on the step size coefficient $\alpha(k)$ and on the communication network properties as defined through the $a^i_j(k)$ coefficients.

**Assumption 2** (Step size)**.** *The sequence $\{\alpha(k)\}_{k \geq 0}$ is positive and non-increasing. Furthermore, it satisfies $\sum_{k=0}^{\infty} \alpha(k) = +\infty$ and $\sum_{k=0}^{\infty} \alpha(k)^2 < +\infty$.* □

A typical choice for $\alpha(k)$ satisfying Assumption 2 is $\alpha(k) = \alpha_1/(k+1)^{\alpha_2}$, with $\alpha_1 > 0$ and $\alpha_2 \in (0.5, 1]$. Note that, differently from [2], we additionally require the $\{\alpha(k)\}_{k \geq 0}$ sequence to be square-summable.

**Assumption 3** (Weight coefficients)**.** *For all $k \geq 0$ and all $i, j = 1, \dots, m$, $a^i_j(k) \in \mathbb{R}_+$, and there exists $l \in (0, 1)$ such that $a^i_i(k) \geq l$ and $a^i_j(k) > 0$ implies $a^i_j(k) > l$. Moreover, for all $k \geq 0$*

- $\sum_{j=1}^m a^i_j(k) = 1$, $i = 1, \dots, m$,
- $\sum_{i=1}^m a^i_j(k) = 1$, $j = 1, \dots, m$. □

Assumption 3 requires the agents to agree on an infinite sequence of doubly stochastic matrices. In case the communication graph is undirected, then a simple distributed procedure can be employed to ensure doubly stochasticity at each iteration, see [8, Assumption 6]. Note that Assumption 3 can be relaxed to requiring row stochasticity only by substituting the consensus update in step 8 with the so-called "push-sum" mechanism, see e.g. [9]. The push-sum mechanism has then to be accounted for explicitly in the proof of the properties of the distributed scheme. This is not addressed here and is left for future work.

**Assumption 4** (Connectivity)**.** *The graph $(V, E_\infty)$ with nodes $V = \{1, \dots, m\}$ and edges $E_\infty = \{(j, i) : a^i_j(k) > 0$ for infinitely many $k\}$ 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.* □

Assumptions 2-4 are quite standard in the distributed optimization literature and some details on their interpretation can be found, e.g., in [7], [10]–[12].

Let $\overline{\mathcal{P}}_{\mathrm{LP}}$ and $\overline{\mathcal{D}}$ denote the primal-dual pair of optimization problems obtained by setting $\rho$ appearing in $\mathcal{P}_{\mathrm{LP},\rho}$ and $\mathcal{D}_\rho$ equal to the consensus value $\bar{\rho}$ for the tightening coefficient defined in Proposition 1.

**Assumption 5** (Uniqueness)**.** *Problems $\overline{\mathcal{P}}_{\mathrm{LP}}$ and $\overline{\mathcal{D}}$ have unique solutions $\bar{x}^\star_{\mathrm{LP}}$ and $\bar{\lambda}^\star$.* □

Note that in case $\overline{\mathcal{P}}_{\mathrm{LP}}$ has multiple solutions (e.g., when it exhibits a high degree of symmetry), then a small perturbation in its cost coefficients will render its solution unique, thus making Assumption 5 fulfilled again. We refer the reader to [1] for a more in-depth discussion about Assumption 5.

**Assumption 6** (Slater)**.** *There exists a scalar $\zeta > 0$ and $\hat{x}_i \in \mathrm{conv}(X_i)$, $i = 1, \dots, m$, such that $\sum_{i=1}^m A_i \hat{x}_i \leq b - \bar{\rho} - m\zeta\mathbf{1}$.* □

We can now state the two main results of this paper.

**Theorem 1** (Finite-time feasibility)**.** *Under Assumptions 1–5, there exists a finite iteration index $K_f$ such that, for all $k \geq K_f$, $x(k) = [x_1(k)^\top \cdots x_m(k)^\top]^\top$, where $x_i(k)$, $i = 1, \dots, m$, are computed by Algorithm 1, is a feasible solution for problem $\mathcal{P}$, i.e., $\sum_{i=1}^m A_i x_i(k) \leq b$ and $x_i(k) \in X_i$, $i = 1, \dots, m$, $k \geq K_f$.* □

**Theorem 2** (Performance guarantees)**.** *Under Assumptions 1–6, there exists a finite iteration index $K_p$ such that, for all $k \geq K_p$, $x(k) = [x_1(k)^\top \cdots x_m(k)^\top]^\top$, where $x_i(k)$, $i = 1, \dots, m$, are computed by Algorithm 1, is a feasible solution for problem $\mathcal{P}$ that satisfies the following performance bound:*

$$\sum_{i=1}^m c_i^\top x_i(k) - J^\star_{\mathcal{P}} \leq \bar{\gamma} + \frac{\|\bar{\rho}\|_\infty}{p\zeta}\tilde{\gamma}, \tag{5}$$

*where $J^\star_{\mathcal{P}}$ is the optimal value of $\mathcal{P}$ and*

$$\tilde{\gamma} = p \max_{i \in \{1, \dots, m\}} \left\{ \max_{x_i \in X_i} c_i^\top x_i - \min_{x_i \in X_i} c_i^\top x_i \right\}.$$

□

As previously mentioned, we retain the finite-time feasibility feature and the performance bound of the work in [2], but without the need of a coordinating unit.

## III. Proof of the main results

*Proof of Proposition 1.* Consider agent $i$. Due to Assumption 1, $\mathrm{conv}(X_i)$ is a bounded non-empty polyhedron. Then, due to Corollaries 2.1 and 2.2 and Theorem 2.3 in [13, Chapter 2], $\mathrm{vert}(X_i)$ is a non-empty finite set. As a consequence, the sequences $\{\bar{s}_i(k)\}_{k\geq 0}$ and $\{\underline{s}_i(k)\}_{k\geq 0}$ defined in steps 11 and 12 take values in a finite set. Since they are component-wise monotonically non-decreasing and non-increasing sequences, respectively, they converge after a finite number of iterations, say $K_i$, to some vectors $\bar{s}_i$ and $\underline{s}_i$, for all $i = 1, \ldots, m$. Let $\bar{\rho}_i = p(\bar{s}_i - \underline{s}_i)$.

For all $k \geq \bar{K} = \max\{K_1, \ldots, K_m\}$ steps 10 and 13 reduce to

$$\rho_i(k+1) = \max_{j\in\mathcal{N}_i(k)}\{\rho_j(k)\}, \tag{6}$$

which is a max-consensus algorithm.

Let $E_k = \{(j,i) : \; a^i_j(k) > 0\}$ be the set of directed edges of the communication graph which are active at iteration $k$. Due to Assumption 4, for every $k \geq 0$ we have that

$$E_\infty \subseteq E^T_k = \bigcup_{r=k}^{k+T-1} E_r. \tag{7}$$

Suppose that agent $i$ is such that $\rho_i(\bar{K}) = \bar{\rho} \geq \bar{\rho}_j$ for all $j = 1, \ldots, m$. Then, running (6) for $T$ consecutive iterations will result in $\rho_{i'}(\bar{K}+T) = \bar{\rho}$ for all $i'$ such that $(i, i') \in E_\infty \subseteq E^T_{\bar{K}}$. After other $T$ consecutive iterations $\rho_{i''}(\bar{K}+2T) = \bar{\rho}$ for all $i''$ such that $(i', i'') \in E_\infty \subseteq E^T_{\bar{K}+T}$. Through a recursive argument one can show that, after at most $dT \leq (m-1)T$ iterations ($d$ being the diameter of the graph $(V, E_\infty)$), every agent $j$ that can receive information from agent $i$ through a path of directed edges in $E_\infty$, will have $\rho_j(\bar{K}+(m-1)T) = \bar{\rho}$.

Since, by Assumption 4, $(V, E_\infty)$ is strongly connected, there exists a path of directed edges connecting agent $i$ to every agent $j$, $j = 1, \ldots, m$. Therefore, setting $K_\rho = \bar{K}+(m-1)T$ we have that $\rho_j(K_\rho) = \bar{\rho}$ for every $j = 1, \ldots, m$.

Using a similar reasoning, we have that the $\{\gamma_i(k)\}_{k\geq 0}$ sequences generated by (3)-(4) converge after a finite number of iterations, say $K_\gamma$, to some values $\bar{\gamma}$, for every $i = 1, \ldots, m$.

Taking $K = \max\{K_\rho, K_\gamma\}$ concludes the proof. □

**Proposition 2** (Dual optimality)**.** *Under Assumptions 1-5 we have that*

$$\lim_{k\to\infty} \|\lambda_i(k) - \bar{\lambda}^\star\| = 0, \textit{ for all } i = 1, \ldots, m. \tag{8}$$

*Proof.* By Proposition 1, there exists a $K \in \mathbb{N}$ such that, for all $k \geq K$, $\rho_i(k) = \bar{\rho}$ for all $i = 1, \ldots, m$. Therefore, for any $k \geq K$ Algorithm 1 reduces to the algorithm in [7] with $f_i(x_i) = c_i^\top x_i$, $g_i(x_i) = A_i x_i - \frac{b-\bar{\rho}}{m}$, and the local constraint sets $X_i$ replaced by $\mathrm{conv}(X_i)$.

Clearly, $f_i(\cdot)$ and $g_i(\cdot)$ are linear and the sets $\mathrm{conv}(X_i)$ are convex, therefore, Assumption 1 in [7] is verified. Furthermore, due to the linearity of $g_i(\cdot)$, Assumption 3 in [7] reduces to a feasibility of $\overline{\mathcal{P}}_{\mathrm{LP}}$, which is implied by Assumption 5. Finally, Assumptions 1, 2, 3, and 4 are the same of Assumption 2, 4, 5, and 6 in [7]. Therefore, recalling Theorem 1 in [7], we immediately get (8), thus concluding the proof. □

**Proposition 3** (Primal finite-time set convergence)**.** *Under Assumptions 1-5, there exists a finite $K$ such that agent $i$ tentative primal solution $x_i(k)$, $i = 1, \ldots, m$, generated by Algorithm 1 satisfies*

$$x_i(k) \in \arg\min_{x_i\in\mathrm{vert}(X_i)} (c_i^\top + \bar{\lambda}^{\star\top} A_i)x_i, \quad k \geq K. \tag{9}$$

*Proof.* From step 8, together with the fact that $a^i_j(k) = 0$ for all $j \notin \mathcal{N}_i(k)$, we have that

$$\begin{aligned}\ell_i(k) - \bar{\lambda}^\star &= \sum_{j\in\mathcal{N}_i(k)} a^i_j(k)\lambda_j(k) - \bar{\lambda}^\star \\ &= \sum_{j=1}^m a^i_j(k)\lambda_j(k) - \bar{\lambda}^\star \\ &= \sum_{j=1}^m a^i_j(k)\left(\lambda_j(k) - \bar{\lambda}^\star\right),\end{aligned} \tag{10}$$

where the last equality is given by the fact that $\sum_{j=1}^m a^i_j(k) = 1$ under Assumption 3. Taking the norm on both sides of (10) we get

$$\begin{aligned}\|\ell_i(k) - \bar{\lambda}^\star\| &= \|\sum_{j=1}^m a^i_j(k)\left(\lambda_j(k) - \bar{\lambda}^\star\right)\| \\ &\leq \sum_{j=1}^m a^i_j(k)\|\lambda_j(k) - \bar{\lambda}^\star\|,\end{aligned} \tag{11}$$

where the inequality is given by the convexity of $\|\cdot\|$ and Assumption 3. Taking the limit for $k \to \infty$ on both sides of (11) together with Proposition 2 yields

$$\lim_{k\to\infty} \|\ell_i(k) - \bar{\lambda}^\star\| = 0, \text{ for all } i = 1, \ldots, m. \tag{12}$$

We can therefore follow the same steps in the proof of Proposition 6 in [2] with $\ell_i(k)$ in place of $\lambda(k)$ and the fact that the sequence $\{\ell_i(k)\}_{k\geq 0}$ generated by Algorithm 1 converges to $\bar{\lambda}^\star$, to get the desired result, thus concluding the proof. □

The proof of the main results of this paper are then obtained by following the same steps as of the proof of Theorem 3 and 4 in [2], respectively. Consistent notations were given to this purpose. Propositions 5 and 6, and Assumptions 3, 4, and 5 used in the proof of Theorem 3 and 4 in [2] are substituted by Propositions 2 and 3, and Assumptions 1, 5, and 6 in the present paper, respectively. References to [14, pag. 117] in [2] should be replaced by [7, Theorem 2].

**Remark 1.** *An upper bound for $K_p$ in Theorem 2 can be obtained as a sum of two contributions: i) the worst-case number of iteration required by the network to explore all the different combinations of $x_i \in X_i$ and agree upon a common $\bar{\rho}$, and ii) the number of iterations required by the algorithm in [7] to enter a ball of a certain radius centered in $\bar{\lambda}^\star$ (see the proof of Proposition 6 in [7]). The latter contribution is not straightforward to derive, and providing an expression of the bound on $K_p$ needs some further work.* □

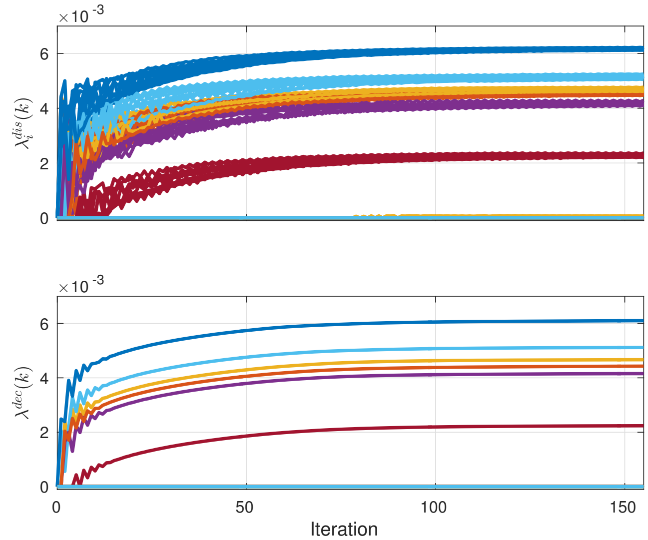


Fig. 1. Lagrange multipliers sequences for Algorithm 1 (upper plot) and [2, Algorithm 1] (lower plot). Different colours represent different components of the Lagrange multipliers vector, and different curves with the same colour represent estimates of different agents for the same Lagrange multiplier.

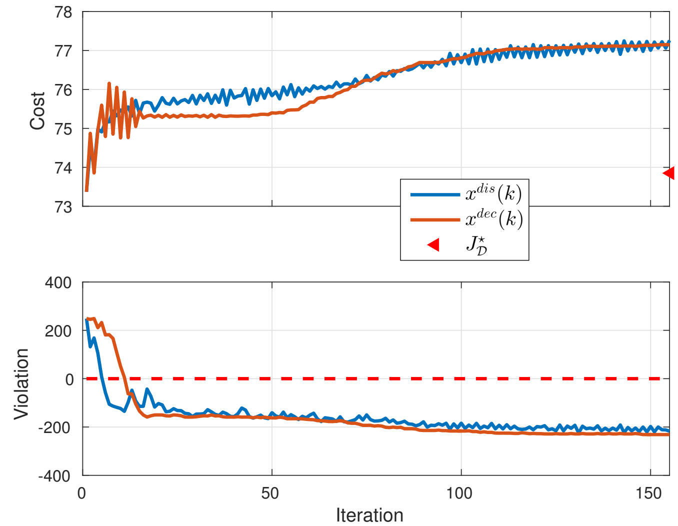


Fig. 2. Cost (upper plot) and violation (lower plot) of the solution $x^{dis}(k)$ generated by Algorithm 1 (blue) and of $x^{dec}(k)$ generated by [2, Algorithm 1] (orange) across iterations. The red triangle denote the optimal value $J^\star_{\mathcal{D}}$ of $\mathcal{D}_\rho$ with $\rho = 0$.

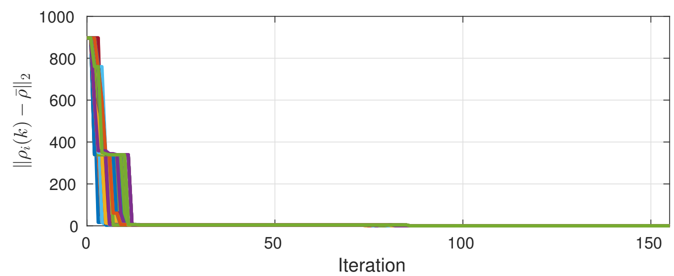


Fig. 3. Distance between the $\rho_i(k)$ generated by Algorithm 1 and their limiting value $\bar{\rho}$. Different colours are associated to different agents.

## IV. Application: Plug-in Electric Vehicles Optimal Charging

Algorithm 1 has been tested on the Plug-in Electric Vehicles (PEVs) optimal charging problem introduced in [1] and further analyzed in [2]. The problem consists in finding an overnight charging schedule for a fleet of $m$ vehicles so as to minimize the charging costs while, at the same time, guaranteeing a desired state of charge for the morning after and complying with local constraints and limitations of the electric network.

Focusing on what is referred to as the "vehicle to grid" set-up in [1], the 8 hour long time horizon is discretized into 24 time slots of 20 minutes each. For every time slot each vehicle has to decide whether to charge/discharge its internal battery at a (fixed) given rate or not to do anything. Such a charging schedule constitutes the optimization vector $x_i$ of vehicle $i$. The battery dynamics is modeled as a discrete-time integrator with charging/discharging losses and is subject to state and input constraints. All these constraints are coded into vehicle's $i$ local constraint set $X_i$. The cost function $c_i^\top x_i$ of each vehicle is the cost of charging the internal battery minus the revenues for selling part of their energy to the network. Finally, the network-wide constraints $\sum_{i=1}^m A_i x_i \leq b$ to be satisfied represent the maximum amount of energy that can be exchanged with the electric network at any given time slot. The reader is referred to [1] for a complete description of the problem and all quantities involved.

We considered an instance of the problem with $m = 250$ vehicles and we applied both Algorithm 1 and the decentralized method in [2, Algorithm 1]. For Algorithm 1 we considered a time-varying communication network constructed as follows: first, the edges of a randomly generated undirected and connected graph were divided into two groups and activated alternately, thus satisfying Assumption 4 with $T = 2$; then, for each of the two subnetworks a set of coefficients $a_j^i$ satisfying Assumption 3 was generated. For [2, Algorithm 1] we set $\alpha(k) = \frac{10^{-5}}{k+1}$, while for Algorithm 1 we set $\alpha(k) = m\frac{10^{-5}}{k+1}$. This is because step 14 in Algorithm 1 contains only the contribution of agent $i$ to the joint constraint, whereas step 13 in [2, Algorithm 1] contains the contribution of all the $m$ agents.

In Figure 1 we report a comparison between Algorithm 1 and [2, Algorithm 1] in terms of evolution of the Lagrange multipliers. Note that, despite the fact that Algorithm 1 is completely distributed, as opposed to [2, Algorithm 1] which is only decentralized, the two methods exhibit a similar behavior towards convergence.

In Figure 2, we also report the behavior of the tentative solutions obtained by the two algorithms in terms of primal cost $J(x(k)) = \sum_{i=1}^m c_i^\top x_i(k)$ and joint constraint violation across iterations. Violation is measured in terms of $\max_{j=1,\dots,p}\{\sum_{i=1}^m [A_i]_j x_i - [b]_j\}$, which is negative if the solution is feasible. From the figure we can observe that both algorithms exhibit a similar behavior and that feasibility is reached within the first 12 iterations.

To quantify the performance achieved by the two solutions, we also report in the upper plot of Figure 2 the optimal value

$J^\star_{\mathcal{D}}$ of $\mathcal{D}_\rho$ with $\rho = 0$, which constitutes a lower bound for the optimal value $J^\star_{\mathcal{P}}$ of $\mathcal{P}$. Since, for $k > 12$, $x(k)$ is a feasible solution (see the lower plot of Figure 2) we have that $J^\star_{\mathcal{P}} \leq J(x(k))$. Therefore $J^\star_{\mathcal{D}} \leq J^\star_{\mathcal{P}} \leq J(x(k))$, and we can conclude that the performance achieved by both algorithms is within a $4\%$ distance from the optimum.

Furthermore, the two approaches converge to a very similar value of $\bar{\rho}$, thus exhibiting about the same level of conservativeness. Finally, in Figure 3 we report the distance between the values of the $\{\rho_i(k)\}_{k\geq 0}$ sequences and their limiting value $\bar{\rho}$, showing that convergence to $\bar{\rho}$ is achieved after a finite number of iterations (around $80$).

## V. Concluding remarks

In this paper we proposed a new algorithm which is able to compute, after a finite number of iterations and in a distributed way, a feasible solution to a large-scale mixed integer linear program. The method exploits the separable structure of the original problem to decompose it into smaller programs, with the additional side-effect of preserving privacy of the agents local information in terms of local cost function and constraint set.

The work retains the state-of-the-art features of finite-time feasibility and sub-optimality bounds of the algorithm in [2], which is extended to a completely distributed set-up accounting also for time-varying communication among agents, thus removing the need for a central unit and broadening the applicability of the approach. The proposed method has been tested on a plug-in electric vehicles optimal charging problem.

Finite-horizon optimal control problems for discrete time Mixed Logical Dynamical (MLD) systems introduced in [14] with linear objective functions can be formulated as an MILP fitting $\mathcal{P}$, since an MLD system is a hybrid system that is modeled via a set of linear inequalities involving both discrete and continuous inputs and state variables. This allows for further applications of the proposed distributed MILP scheme.

Future research directions include the generalization of our results to the case of non-convex cost functions, based on the recent work in [15], and the computation of an upper bound for the number of iterations required to obtain a feasible solution.

## References


[1] R. Vujanic, P. M. Esfahani, P. J. Goulart, S. Mariéthoz, and M. Morari, "A decomposition method for large scale MILPs, with performance guarantees and a power system application," *Automatica*, vol. 67, pp. 144–156, 2016.

[2] A. Falsone, K. Margellos, and M. Prandini, "A decentralized approach to multi-agent MILPs: finite-time feasibility and performance guarantees," *Automatica*, 2018, provisionally accepted as regular paper on January 6, 2018. Available online arXiv:1706.08788.

[3] M. Dawande, S. Gavirneni, and S. Tayur, "Effective heuristics for multiproduct partial shipment models," *Operations research*, vol. 54, no. 2, pp. 337–352, 2006.

[4] P. Baumann and N. Trautmann, "Portfolio-optimization models for small investors," *Mathematical Methods of Operations Research*, vol. 77, no. 3, pp. 345–356, 2013.

[5] H. Y. Yamin, "Review on methods of generation scheduling in electric power systems," *Electric Power Systems Research*, vol. 69, no. 2, pp. 227–248, 2004.

[6] D. Ioli, A. Falsone, and M. Prandini, "An iterative scheme to hierarchically structured optimal energy management of a microgrid," in *54th Conference on Decision and Control (CDC2015)*, 2015, pp. 5227–5232.

[7] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, 2017.

[8] 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.

[9] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proceedings of the 51st IEEE Conference on Decision and Control*. IEEE, 2012, pp. 5453–5458.

[10] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[11] 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.

[12] 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*, vol. 63, no. 5, pp. 1372–1387, 2018.

[13] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*. Athena Scientific Belmont, MA, 1997, vol. 6.

[14] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, no. 3, pp. 407–427, 1999.

[15] M. Udell and S. Boyd, "Bounding duality gap for separable problems with linear constraints," *Computational Optimization and Applications*, vol. 64, no. 2, pp. 355–378, 2016.