Constraint-Coupled Distributed Optimization: a Relaxation and Duality Approach

Ivano Notarnicola, *Member, IEEE* and Giuseppe Notarstefano, *Member, IEEE*

***Abstract*—In this paper we consider a general, challenging distributed optimization set-up arising in several important network control applications. Agents of a network want to minimize the sum of local cost functions, each one depending on a local variable, subject to local and coupling constraints, with the latter involving all the decision variables. We propose a novel fully distributed algorithm based on a relaxation of the primal problem and an elegant exploration of duality theory. Despite its complex derivation, based on several duality steps, the distributed algorithm has a very simple and intuitive structure. That is, each node finds a primal-dual optimal solution pair of a local, relaxed version of the original problem, and then updates suitable auxiliary local variables. We prove that agents asymptotically compute their portion of an optimal (feasible) solution of the original problem. This primal recovery property is obtained without any averaging mechanism typically used in dual decomposition methods. To corroborate the theoretical results, we show how the methodology applies to an instance of a Distributed Model Predictive Control scheme in a microgrid control scenario.**

I. Introduction

In the last decade distributed optimization has received significant attention. Literature has mainly focused on *cost-coupled* optimization problems in which the cost to be minimized is the sum of local functions depending on a common decision variable, see [2]–[10] and references therein for an overview. A different, more general optimization set-up amenable to distributed computation is the minimization of the sum of local cost functions, each one depending on a *local* variable, subject to a local constraint for each variable and a coupling constraint involving all the decision variables. In this problem, the global optimal solution is obtained by stacking all the local variables. This feature leads easily to so-called big-data problems having a very highly dimensional decision variable that grows with the network size. However, since agents are typically interested in computing only their (small) portion of the optimal solution, novel tailored methods need to be developed to address these challenges. We call this framework *constraint-coupled* optimization set-up.

Several scenarios of interest in Controls and Robotics as well as Communication and Signal Processing strongly motivate the investigation of such a problem. Example set-ups include resource allocation (e.g., in Communication or Cooperative Robotics) or network flow optimization (e.g., in smart grid energy management). A set-up which is particularly relevant in our community is distributed Model Predictive Control (MPC) in which the goal is to design a feedback law for a (spatially distributed) network of dynamical systems based on the MPC concept. In such a scheme several optimization problems need to be iteratively solved. The local decision variable of each agent corresponds to its state-input trajectory, while the local constraints encode its dynamics, which is usually independent of other agents. A constraint that couples agents' states, inputs or outputs needs to be taken into account in order to enforce cooperative tasks as, e.g., formation control, or to take into account common bounds, e.g., due to shared resources. Distributed MPC approaches are mainly classified into non-cooperative and cooperative schemes [11]. While in non-cooperative schemes the main focus is to guarantee recursive feasibility and stability, in cooperative approaches agents care also about optimality when solving the global, constraint-coupled problems arising in each time window and, thus, call for tailored distributed optimization algorithms.

Parallel methods for constraint-coupled problems have been developed mainly in the context of cooperative MPC. They are based on a master-subproblem architecture that traces back to late 90s [12]. Duality is a widely used tool to decompose the problem and design optimization algorithms as shown, e.g., in the tutorial papers [13], [14]. In [15] an accelerated dual decomposition is proposed to solve a MPC problem. In [16] dual decomposition is combined with a penalty approach to solve separable nonconvex problems. A linear convergence rate for a dual gradient algorithm for linearly constrained separable convex problems is proven in [17]. In [18] an inexact dual decomposition scheme combined with smoothing techniques is proposed. In [19] a primal-dual, real-time strategy is proposed to solve parametric nonconvex programs usually arising in nonlinear MPC. Recently, parallel augmented Lagrangian methods have been proposed in [20], [21] to solve nonconvex problem instances with linear coupling constraints. Although sometimes termed as "distributed" algorithms, they require a centralized unit performing some steps in the proposed strategies. When further sparsity is assumed in the problem, e.g., the sub-systems have coupled dynamics with their neighbors only [14]–[17], then the parallel scheme can be implemented over a network without a central authority. In this paper we propose a purely distributed algorithm also for general coupling constraints that involve the entire set of agents.

Distributed optimization algorithms for special versions of the constraint-coupled set-up, arising in the context of resource allocation problems, have been proposed in [22]–[25].

The general constraint-coupled set-up we consider in this paper has not received extensive investigation in a purely distributed framework and only few works are available,

A preliminary short version of this paper has appeared as [1]: the current article includes a much improved comprehensive treatment, all the theoretical proofs and a numerical example on a microgrid control problem.

Ivano Notarnicola and Giuseppe Notarstefano are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy, name.lastname@unibo.it. 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).

i.e., [26]–[30]. In [26] a consensus-based primal-dual perturbation algorithm is proposed to solve smooth constraint-coupled optimization problems. Very recently, in [27] and [28] distributed algorithms are proposed based on a consensus-based dual decomposition and a dual proximal optimization approach, respectively. A class of min-max optimization problems, strictly related to the constraint-coupled set-up, is addressed in [29] using a Laplacian-based saddle-point subgradient scheme. A well-known drawback in methods based on dual decomposition, is that primal feasibility is not easily retrieved from dual solutions. Thus, primal recovery mechanisms are devised in the papers above in order to recover a primal solution by suitably applying running average schemes borrowed from the centralized literature, see, e.g., [31]. A special coupling associated to peak minimization problems arising in demand-side management is considered in [30], where a (simplified) algorithmic approach, similar to the one proposed in this paper, is proposed for that set-up. An alternative approach to constraint-coupled problems has been proposed in [32], and customized to MPC in [33], [34], where agents employ a cutting-plane consensus scheme to iteratively approximate their local problems. This approach enforces agents to eventually agree on the complete solution vector and this may be an undesirable feature in some applications.

The main paper contributions are as follows. We propose a novel, distributed optimization algorithm to solve constraint-coupled optimization problems over networks. Overall, our distributed algorithm enjoys three appealing features: (i) local computations at each node involve only the local decision variable and, thus, scale nicely with respect to the dimension of the decision vector, (ii) privacy is preserved since agents do not communicate, and thus disclose, their estimates of local decision variable, cost or constraints, and (iii) an estimate of a primal optimal solution component is directly computed by each agent without any averaging mechanism, which results in a faster algorithm.

The proposed approach combines a proper relaxation of the original problem with an elegant exploration of duality theory. The resulting distributed algorithm is a two-step procedure in which each agent iteratively performs a (primal) constrained and small-sized optimization, followed by a dual update. The local problems involve the local cost function and the local constraint of the agent. Also, a local inequality constraint, which is adjusted at each step, accounts for the coupling constraint involving all the agents. Although this constraint dynamically changes over the iterations, we do not need to assume a priori feasibility of local problems, but rather, thanks to the proposed relaxation approach, local violations are allowed and proven to be asymptotically vanishing. Each local solution estimate is guaranteed to asymptotically converge to the component of an optimal (and, thus, feasible) solution of the original problem. Such primal convergence of local estimates, known in the literature as primal recovery, is non-obvious in duality-based methods applied to (merely) convex programs. In our distributed algorithm, this property results from the methodology we employed, without resorting to any (commonly used) running averaging mechanism. Moreover, this key feature has an even stronger impact on our scheme since it allows primal quantities to directly inherit the convergence rate of a "centralized" subgradient iteration, while, in general, running averages further degrade such a rate. Finally, since no particular initialization is required, our distributed algorithm can be implemented in a dynamic context in which the problem may change or nodes can appear or disappear.

The paper unfolds as follows. In Section II we formalize the set-up and introduce our distributed optimization algorithm. In Section III we give a constructive derivation of of the algorithm and in Section IV we conclude the analysis by proving the its convergence properties. In Section V we corroborate the theoretical results by showing how the methodology applies to an instance of a distributed MPC controller for a microgrid.

II. Distributed Set-up and Optimization Algorithm

In this section we formally state the general constraint-coupled problem we aim at investigating in this paper as strongly motivated by control applications discussed in the introduction. Moreover, we introduce the proposed distributed algorithm along with its convergence theorem.

A. Constraint-Coupled Set-up

Consider the following optimization problem

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

where for all $i \in \{1,\dots,N\}$, the set $X_i \subseteq \mathbb{R}^{n_i}$ with $n_i \in \mathbb{N}$, the functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ and $\mathbf{g}_i : \mathbb{R}^{n_i} \to \mathbb{R}^S$ with $S \in \mathbb{N}$. The symbol $\leq$ (and, consistently, for other sides) means that the inequality holds component-wise and $\mathbf{0} \triangleq [0,\dots,0]^\top \in \mathbb{R}^S$.

Assumption II.1. *For all $i \in \{1,\dots,N\}$, each function f_i is convex, and each X_i is a non-empty, compact, convex set. Moreover, each $\mathbf{g}_i$ is a component-wise convex function, i.e., for all $s \in \{1,\dots,S\}$ each component $\mathbf{g}_{is} : \mathbb{R}^{n_i} \to \mathbb{R}$ of $\mathbf{g}_i$ is a convex function.* □

The following assumption is the well-known Slater's constraint qualification.

Assumption II.2. *There exist $\bar{\mathbf{x}}_1 \in X_1, \dots, \bar{\mathbf{x}}_N \in X_N$ such that $\sum_{i=1}^{N} \mathbf{g}_i(\bar{\mathbf{x}}_i) < \mathbf{0}$.* □

These assumptions are quite standard and guarantee that problem (1) admits (at least) an optimal solution $(\mathbf{x}_1^\star, \dots, \mathbf{x}_N^\star)$ such that its optimal cost is $\sum_{i=1}^{N} f_i(\mathbf{x}_i^\star) = f^\star$. Moreover, a dual approach can be applied since strong duality holds. Notice that we assumed that $\sum_{i=1}^{N} \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}$ admits a strictly feasible point, while each $\mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}$ may not.

We consider a network of N processors communicating according to a *connected* and *undirected* graph $\mathcal{G} = (\{1,\dots,N\}, \mathcal{E})$, where $\mathcal{E} \subseteq \{1,\dots,N\} \times \{1,\dots,N\}$ is the set of edges. Edge (i,j) models the fact that node i sends information to j. Note that, being the graph undirected, for each $(i,j) \in \mathcal{E}$, then also $(j,i) \in \mathcal{E}$. We denote by $|\mathcal{E}|$ the

cardinality of $\mathcal{E}$ and by $\mathcal{N}_i$ the set of *neighbors* of node i in $\mathcal{G}$, i.e., $\mathcal{N}_i = \{j \in \{1,\dots,N\} \mid (i,j) \in \mathcal{E}\}$. Each node i knows only f_i, $\mathbf{g}_i$ and X_i, and aims at estimating its portion $\mathbf{x}_i^\star$ of an optimal solution $(\mathbf{x}_1^\star,\dots,\mathbf{x}_N^\star)$ of (1) by means of local communication only.

Remark II.3. *Problem* (1) *is a special case of the general set-up* $\min_{\mathbf{x}\in X}\sum_i f_i(\mathbf{x})$. *However, addressing the solution of* (1) *by means of existing distributed optimization algorithms tailored for that general set-up incurs in two main issues: first, each agent* i *should know the entire coupling constraint* $\sum_i g_i$ *(while in some applications* i *only knows* g_i*); second, all agents eventually converge to the stack of optimal solutions while in general they can be interested (e.g., for privacy reasons) in computing their portion* $\mathbf{x}_i^\star$ *only.* □

B. Relaxation and Successive Distributed Decomposition

In this subsection we present our Relaxation and Successive Distributed Decomposition method (RSDD) which is a novel strategy to solve problem (1) over networks.

Informally, the algorithm consists of an iterative two-step procedure. First, each node $i \in \{1,\dots,N\}$ stores a set of variables $((\mathbf{x}_i,\rho_i),\boldsymbol{\mu}_i) \in \mathbb{R}^{n_i}\times\mathbb{R}\times\mathbb{R}^S$ obtained as a primal-dual optimal solution pair of problem (2). The vector $\boldsymbol{\mu}_i$ is the multiplier associated to the inequality constraint in (2). We notice that (2) mimics a local version of the centralized problem (1), where the coupling with the other nodes in the original formulation is replaced by a local term depending only on neighboring variables $\boldsymbol{\lambda}_{ij} \in \mathbb{R}^S$ and $\boldsymbol{\lambda}_{ji} \in \mathbb{R}^S$, $j \in \mathcal{N}_i$. Moreover, this local version of the coupling constraint is also relaxed, i.e., a positive violation $\rho_i\mathbf{1}$ is allowed, where $\mathbf{1} \triangleq [1,\dots,1]^\top \in \mathbb{R}^S$. Finally, instead of minimizing only the local f_i, the (scaled) violation $M\rho_i$, $M>0$, enters the cost function as well. The auxiliary variables $\boldsymbol{\lambda}_{ij}$, $j \in \mathcal{N}_i$, are updated in the second step according to a suitable linear law that combines neighboring $\boldsymbol{\mu}_i$ as in (3). Nodes use a step-size denoted by γ^t and can initialize the variables $\boldsymbol{\lambda}_{ij}$, $j \in \mathcal{N}_i$ to *arbitrary* values. In the following table we formally state our distributed algorithm from the perspective of node i.

Distributed Algorithm RSDD

Processor states: $\mathbf{x}_i$, ρ_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{\lambda}_{ij}$ for $j \in \mathcal{N}_i$

Initialization: $\boldsymbol{\lambda}_{ij}^0$ arbitrary for $j \in \mathcal{N}_i$, M sufficiently large

Evolution:

Gather $\boldsymbol{\lambda}_{ji}^t$ from $j \in \mathcal{N}_i$

Compute $\big((\mathbf{x}_i^{t+1},\rho_i^{t+1}),\boldsymbol{\mu}_i^{t+1}\big)$ as a primal-dual optimal solution pair of

$$\begin{aligned} \min_{\mathbf{x}_i,\rho_i}\ & f_i(\mathbf{x}_i) + M\rho_i \\ \text{subj. to}\ & \rho_i \geq 0,\ \ \mathbf{x}_i \in X_i \\ & \mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i}\big(\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t\big) \leq \rho_i\mathbf{1} \end{aligned} \tag{2}$$

Gather $\boldsymbol{\mu}_j^{t+1}$ from $j \in \mathcal{N}_i$

Update for all $j \in \mathcal{N}_i$

$$\boldsymbol{\lambda}_{ij}^{t+1} = \boldsymbol{\lambda}_{ij}^t - \gamma^t\big(\boldsymbol{\mu}_i^{t+1} - \boldsymbol{\mu}_j^{t+1}\big) \tag{3}$$

As already mentioned, each agent i aims at computing an optimal strategy by means of local interaction only. The proposed distributed algorithm exploits the nice structure of the constraints in the problem formulation to derive a protocol in which agents exchange only the vectors $\boldsymbol{\mu}_i^t$ and $\boldsymbol{\lambda}_{ij}^t$ without explicitly communicating the current estimates of their local decision variables $\mathbf{x}_i^t$, costs f_i or constraints $\mathbf{g}_i$.

This is an important, appealing feature of the RSDD distributed algorithm since it guarantees privacy in the network.

Remark II.4. *Since the initialization is arbitrary, then the algorithm can run in a flexible scenario in which agents can join or leave the network and problem data can change. These events induce a new optimization problem and trigger a transient for the distributed algorithm. Provided that the algorithm runs for sufficiently long time before the triggering event, then practical convergence is preserved also for the new problem instance.* □

In order to gain more intuition about the algorithmic evolution, at this point we provide an informal interpretation, supported by Figure 1, of the local optimization step in (2).

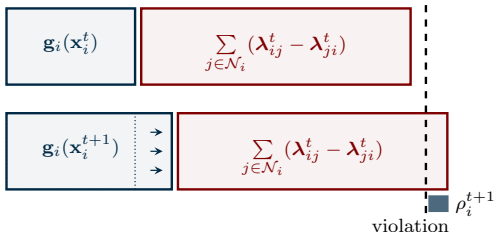

Figure 1. Graphical representation of the local constraint relaxation for a scalar coupling constraint.

Agent i, due to its partial knowledge, can only optimize with respect to its own decision variable $\mathbf{x}_i$. Thus, it can solve an instance of problem (1) in which all the other variables in the network have a given value $\mathbf{x}_j^t$ for $j \in \{1,\dots,N\}\setminus\{i\}$. Thus, the cost function reduces to f_i only. As for the coupling constraint, it describes how the resources are allocated to all the agents at each iteration t. In the figure, we show a possible instance of a feasible allocation: in blue we depicted the resource assigned to agent i while in red an estimate of the resources currently allocated to all the other agents. When agent i optimizes its local variable $\mathbf{x}_i$ only, it can "play" with the "available resource slot" given by $-\sum_{j\neq i}\mathbf{g}_j(\mathbf{x}_j^t)$. Since the current allocation is in general not optimal, this constraint might be too restrictive. In fact, it can slow down (and even stall) the optimization process by easily leading to infeasibility of the local problem (2) when ρ_i is set to 0. On this regard, recall that we do not assume feasibility of every $\mathbf{g}_i$ independently. Also, it is worth noting that such "available resource slot" depends on the entire network's variables, and, thus, it is not an easily available information in a distributed scenario. Thus, we propose a strategy in which at each iteration t, each agent i replaces the term $\sum_{j\neq i}\mathbf{g}_j(\mathbf{x}_j^t)$ in the coupling with $\sum_{j\in\mathcal{N}_i}\big(\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t\big)$. Notice that this term can be computed *locally* at each node by communicating with neighboring agents only. This term is then iteratively adjusted along the

algorithmic evolution in order to eventually obtain an optimal solution. Finally, each agent i is allowed for a violation $\rho_i \mathbf{1}$ of the local version of the coupling constraint. At the same time, this violation is penalized in order to encourage it to eventually converge to zero. This intuitive description will be rigorously derived and proven in the following sections.

We are now ready to state the main result of the paper, namely the convergence of the RSDD distributed algorithm. We start by formalizing the assumption that the step-size should satisfy.

Assumption II.5. *The sequence $\{\gamma^t\}_{t\geq 0}$, with each $\gamma^t \geq 0$, satisfies the conditions $\sum_{t=0}^{\infty} \gamma^t = \infty$, $\sum_{t=0}^{\infty} \left(\gamma^t\right)^2 < \infty$.* □

The convergence theorem is stated below.

Theorem II.6. *Let Assumption II.1 and II.2 hold, and let the step-size γ^t satisfy Assumption II.5. Moreover, letting $\boldsymbol{\mu}^\star$ be an optimal solution of the dual of problem* (1)*, assume $M > \|\boldsymbol{\mu}^\star\|_1$. Consider a sequence $\left\{\mathbf{x}_i^t, \rho_i^t\right\}_{t\geq 0}$, $i \in \{1,\dots,N\}$, generated by the RSDD distributed algorithm with an arbitrary initialization. Then, the following holds:*

(i) $\left\{\sum_{i=1}^N \left(f_i(\mathbf{x}_i^t) + M\rho_i^t\right)\right\}_{t\geq 0}$ converges to the optimal cost $f^\star$ of (1)*;*

(ii) every limit point of $\left\{\mathbf{x}_i^t\right\}_{t\geq 0}$, $i \in \{1,\dots,N\}$, is a primal optimal (feasible) solution of (1)*.*

Proof. The proof is given in Section IV-B. □

We point out that Theorem II.6 does not require/claim any asymptotic consensus among the $\boldsymbol{\mu}_i^t$, $i \in \{1,\dots,N\}$.

Remark II.7. *When $f_i = 0$ for all $i \in \{1,\dots,N\}$, then our RSDD becomes a distributed algorithm for solving a feasibility problem, i.e., find $(\mathbf{x}_1,\dots,\mathbf{x}_N)$ such that $\mathbf{x}_i \in X_i$, for all $i \in \{1,\dots,N\}$ and $\sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}$.* □

III. Algorithm Analysis: Relaxation and Duality Tour

In this section we present a constructive derivation of our distributed algorithm. The methodology relies on a proper relaxation of the original problem and on the derivation of a sequence of equivalent problems. We point out that, although based on a relaxation, the proposed algorithm exploits such a relaxation to solve *exactly* the original problem formulation in a distributed way.

A. First Duality Step and Relaxation Approach

We start our duality tour by deriving the dual problem of (1) and a restricted version necessary for the algorithm derivation.

Let $\boldsymbol{\mu} \geq \mathbf{0} \in \mathbb{R}^S$, be a multiplier associated to the inequality constraint $\sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \leq \mathbf{0}$ in (1). Then, the dual of problem (1) is given by

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^S} \ & \sum_{i=1}^N q_i(\boldsymbol{\mu}) \\ \text{subj. to } & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{4}$$

where each term q_i of the dual function $q(\boldsymbol{\mu}) = \sum_{i=1}^N q_i(\boldsymbol{\mu})$ is defined as

$$q_i(\boldsymbol{\mu}) = \min_{\mathbf{x}_i \in X_i} \Big(f_i(\mathbf{x}_i) + \boldsymbol{\mu}^\top \mathbf{g}_i(\mathbf{x}_i)\Big), \tag{5}$$

for all $i \in \{1,\dots,N\}$. Let $q^\star$ be the optimal cost of (4). Notice that we do not need to assume uniqueness of the solution of problem (4). As already mentioned, in light of Assumptions II.1 and II.2, problem (1) is feasible and has finite optimal cost $f^\star$. Moreover, the Slater's condition holds and, thus, the strong duality theorem for convex inequality constraints, [35, Proposition 5.3.1], applies, ensuring that strong duality holds, i.e., problems (1) and (4) have the same optimal cost $f^\star = q^\star$. Moreover, $q^\star$ is attained at some $\boldsymbol{\mu}^\star \geq \mathbf{0}$, i.e., $q(\boldsymbol{\mu}^\star) = q^\star$, cf. [35, Proposition 5.1.4]. Finally, we recall that since $\sum_{i=1}^N q_i(\boldsymbol{\mu})$ is the dual function of (1), then it is concave on its convex domain $\boldsymbol{\mu} \geq \mathbf{0}$. With the dual problem at hand, several existing algorithms can be applied to directly solve (4) in a distributed way, see e.g., the distributed projected subgradient [2]. However, as pointed out in the introduction such dual approaches do not guarantee primal recovery and additional schemes must be employed to regain it, e.g., averaging mechanisms.

In this paper we propose an alternative approach that relies on a further duality step that gives rise to the RSDD distributed algorithm, which overtakes these issues. Let us introduce an optimization problem similar to (4), given by

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^S} \ & \sum_{i=1}^N q_i(\boldsymbol{\mu}) \\ \text{subj. to } & \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\mu}^\top \mathbf{1} \leq M, \end{aligned} \tag{6}$$

where M is a positive scalar and $\mathbf{1} = [1,\dots,1]^\top$. This problem is a *restricted* version of problem (4). Here, in fact an additional constraint, namely $\boldsymbol{\mu}^\top \mathbf{1} \leq M$, has been added to (4). It is worth mentioning that this restriction makes the constraint set of (6) a compact set. Although this step may seem not motivated at this point of the paper, its necessity will be clear from the following steps of the analysis, see also Section IV-C for a dedicated discussion.

Notice that, if M is sufficiently large, the presence of the constraint $\boldsymbol{\mu}^\top \mathbf{1} \leq M$ in (6) will not alter the optimal solutions of the unrestricted problem (4). The next result formally establishes the relationship between problems (6) and (4).

Lemma III.1. *Let $\boldsymbol{\mu}^\star$ be an optimal solution of problem* (4) *and M be a positive scalar satisfying $M > \|\boldsymbol{\mu}^\star\|_1$. Then, problems* (6) *and* (4) *have the same optimal cost, namely $q^\star = f^\star$. Moreover, $\boldsymbol{\mu}^\star$ is an optimal solution also for problem* (6)*.*

Proof. The constraint set $\{\boldsymbol{\mu} \geq \mathbf{0} \mid \boldsymbol{\mu}^\top \mathbf{1} \leq M\}$ is a restriction of the constraint set $\boldsymbol{\mu} \geq \mathbf{0}$ of problem (4) containing $\boldsymbol{\mu}^\star$. Thus the optimal cost of (6) is, in general, greater than or equal to the optimal cost of (4). Since the domain of (6) contains at least one optimal solution of problem (4), namely $\boldsymbol{\mu}^\star$, then the optimal cost of problem (6) is $q^\star$ and is (at least) attained at $\boldsymbol{\mu}^\star$, so that the proof follows. □

With the dual problem (4) and its restricted version (6) at hand, one may wonder about the connection between

their primal counterparts. Next, we show that the restricted problem (6) is the dual of a *relaxed* version of the original primal problem (1).

Lemma III.2. *Problem* (6) *is the dual of the following optimization problem*

$$\begin{aligned} \min_{\mathbf{x}_1,\dots,\mathbf{x}_N,\rho} \ & \sum_{i=1}^N f_i(\mathbf{x}_i) + M\rho \\ \text{subj. to } & \rho \geq 0, \ \ \mathbf{x}_i \in X_i, \quad i \in \{1,\dots,N\} \\ & \sum_{i=1}^N \mathbf{g}_i(\mathbf{x}_i) \leq \rho \mathbf{1}, \end{aligned} \tag{7}$$

and strong duality holds.

Proof. The dual function of (7) is given by

$$\begin{aligned} q_R(\boldsymbol{\mu}) &= \inf_{\substack{\mathbf{x}_1\in X_1,\dots,\mathbf{x}_N\in X_N \\ \rho\geq 0}} \sum_{i=1}^N \Big(f_i(\mathbf{x}_i)+\boldsymbol{\mu}^\top \mathbf{g}_i(\mathbf{x}_i)\Big)+\rho\big(M-\boldsymbol{\mu}^\top \mathbf{1}\big) \\ &= \begin{cases} \sum_{i=1}^N \underbrace{\min_{\mathbf{x}_i\in X_i}\Big(f_i(\mathbf{x}_i)+\boldsymbol{\mu}^\top \mathbf{g}_i(\mathbf{x}_i)\Big)}_{q_i(\boldsymbol{\mu})}, & \text{if } M-\boldsymbol{\mu}^\top\mathbf{1}\geq 0 \\ -\infty, & \text{otherwise} \end{cases} \end{aligned}$$

where each $q_i(\boldsymbol{\mu})$ is the same defined in (5). The maximization of the dual function $q_R(\boldsymbol{\mu})$ on its domain turns out to be the maximization of $\sum_{i=1}^N q_i(\boldsymbol{\mu})$ over $\{\boldsymbol{\mu} \geq \mathbf{0} \mid \boldsymbol{\mu}^\top \mathbf{1} \leq M\}$, which is problem (6), and the proof follows. □

Notice that problem (7) is a relaxation of problem (1) since we allow for a positive violation of the coupling constraint. At the same time, the violation ρ is penalized with a scaling factor M in order to discourage it. The variable ρ resembles the ρ_i introduced in the distributed RSDD algorithm. However, as it will be clear from the forthcoming analysis, ρ_i is *not* a local estimate (or copy) of ρ, but it rather represents the local contribution of agent i to the common violation ρ.

The following result characterizes how the original primal problem (1) and its relaxed version (7) are related.

Proposition III.3. *Let M be such that $M > \|\boldsymbol{\mu}^\star\|_1$ with $\boldsymbol{\mu}^\star$ an optimal solution of the dual of problem* (1)*. The optimal solutions of the relaxed problem* (7) *are in the form $(\mathbf{x}_1^\star,\dots,\mathbf{x}_N^\star,0)$, where $(\mathbf{x}_1^\star,\dots,\mathbf{x}_N^\star)$ is an optimal solution of* (1)*, i.e., solutions of* (7) *must have $\rho^\star = 0$.*

Proof. First, we notice that problem (7) is the epigraph formulation of

$$\begin{aligned} \min_{\mathbf{x}_1,\dots,\mathbf{x}_N} & \sum_{i=1}^N f_i(\mathbf{x}_i) + M \max\Big\{0, \sum_{i=1}^N \mathbf{g}_{i1}(\mathbf{x}_i), \dots, \sum_{i=1}^N \mathbf{g}_{iS}(\mathbf{x}_i)\Big\} \\ \text{subj. to } & \mathbf{x}_i \in X_i, \ i \in \{1,\dots,N\}, \end{aligned} \tag{8}$$

where $\mathbf{g}_{is}$ denotes the s-th component of $\mathbf{g}_i$. Problems (1) and (8) enjoy the same structure as the ones considered in [36, Proposition 5.25]. By Assumption II.2, problem (7) (and thus (8)) satisfies the assumptions for strong duality. Thus, following [36, p. 364], we consider the penalty function $M\max\{0, \sum_{i=1}^N \mathbf{g}_{i1}(\mathbf{x}_i), \dots, \sum_{i=1}^N \mathbf{g}_{iS}(\mathbf{x}_i)\}$, with $M > \|\boldsymbol{\mu}^\star\|_1$, so that we can apply [36, Proposition 5.25] to conclude that problems (1) and (8) have the same optimal solutions, thus completing the proof. □

Remark III.4 (Alternative restrictions)**.** *Other choices for the restriction of the domain $\boldsymbol{\mu} \geq \mathbf{0}$ of* (4) *can be considered. For instance, one can consider upper bounds in the form $\boldsymbol{\mu} \leq M\mathbf{1}$ or $\boldsymbol{\mu} \leq [M_1,\dots,M_S]^\top$. As one might expect, the specific constraint restriction gives rise to different forms of the relaxed primal problem* (7)*.* □

B. Second Dual Problem Derivation

At this point, we continue our duality tour in order to design an algorithm that solves problem (6) instead of the unrestricted dual problem (4).

In order to make problem (6) amenable for a distributed solution, we enforce a sparsity structure that matches the network. To this end, we introduce copies of the common optimization variable $\boldsymbol{\mu}$ and we copy also its domain. Moreover, we enforce coherence constraints among the copies $\boldsymbol{\mu}_i$ having the sparsity of the connected graph $\mathcal{G}$, thus obtaining

$$\begin{aligned} \max_{\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_N} & \sum_{i=1}^N q_i(\boldsymbol{\mu}_i) \\ \text{subj. to } & \boldsymbol{\mu}_i \geq \mathbf{0}, \boldsymbol{\mu}_i^\top \mathbf{1} \leq M, \quad i \in \{1,\dots,N\} \\ & \boldsymbol{\mu}_i = \boldsymbol{\mu}_j, \qquad\qquad (i,j) \in \mathcal{E}. \end{aligned} \tag{9}$$

Being problem (9) an equivalent version of problem (6), it has the same optimal cost $q^\star = f^\star$.

On problem (9) we would like to use a dual decomposition approach with the aim of obtaining a distributed algorithm. That is, the leading idea is to derive the dual of problem (9) and apply a subgradient method to solve it.

We start deriving the dual problem of (9) by dualizing only the coherence constraints. Consider the partial Lagrangian

$$\mathcal{L}(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_N,\boldsymbol{\Lambda}) = \sum_{i=1}^N \Big(q_i(\boldsymbol{\mu}_i) + \sum_{j\in\mathcal{N}_i} \boldsymbol{\lambda}_{ij}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)\Big), \tag{10}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{S\cdot|\mathcal{E}|}$ is the vector stacking each Lagrange multiplier $\boldsymbol{\lambda}_{ij} \in \mathbb{R}^S$, with $(i,j) \in \mathcal{E}$, associated to the constraint $\boldsymbol{\mu}_i - \boldsymbol{\mu}_j = \mathbf{0}$. Notice that we have not dualized the local constraints $\{\boldsymbol{\mu}_i \geq \mathbf{0} \mid \boldsymbol{\mu}_i^\top \mathbf{1} \leq M\}$.

Since the communication graph $\mathcal{G}$ is undirected, we can exploit the symmetry of the constraints. Indeed, for each $(i,j) \in \mathcal{E}$ we also have $(j,i) \in \mathcal{E}$, and, expanding all the terms in (10), for given i and j, we always have both the terms $\boldsymbol{\lambda}_{ij}^\top(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$ and $\boldsymbol{\lambda}_{ji}^\top(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)$. Thus, after some simple algebraic manipulations, we can rephrase (10) as $\mathcal{L}(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_N,\boldsymbol{\Lambda}) = \sum_{i=1}^N \Big(q_i(\boldsymbol{\mu}_i) + \boldsymbol{\mu}_i^\top \sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}_{ij} - \boldsymbol{\lambda}_{ji})\Big)$, which is separable with respect to $\boldsymbol{\mu}_i$, $i \in \{1,\dots,N\}$. Thus, the dual function of (9) is

$$\eta(\boldsymbol{\Lambda}) = \sum_{i=1}^N \eta_i\big(\{\boldsymbol{\lambda}_{ij},\boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big), \tag{11}$$

where, for all $i \in \{1,\dots,N\}$,

$$\eta_i\big(\{\boldsymbol{\lambda}_{ij},\boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big) = \sup_{\substack{\boldsymbol{\mu}_i\geq\mathbf{0}, \\ \boldsymbol{\mu}_i^\top\mathbf{1}\leq M}} \Big(q_i(\boldsymbol{\mu}_i) + \boldsymbol{\mu}_i^\top \sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}_{ij} - \boldsymbol{\lambda}_{ji})\Big). \tag{12}$$

Finally, by denoting the domain of η as $D_\eta = \{\boldsymbol{\Lambda} \in \mathbb{R}^{S\cdot|\mathcal{E}|} \mid \eta(\boldsymbol{\Lambda}) < +\infty\}$, the dual of problem (9) reads

$$\min_{\boldsymbol{\Lambda}\in D_\eta} \eta(\boldsymbol{\Lambda}) = \min_{\boldsymbol{\Lambda}\in D_\eta} \sum_{i=1}^{N} \eta_i\big(\{\boldsymbol{\lambda}_{ij}, \boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big). \tag{13}$$

Since problem (13) is a dual program, then it is a convex (constrained) problem. Moreover, its cost function $\eta(\boldsymbol{\Lambda})$ is very structured since it is a sum of contributions η_i and each of them depends only on neighboring variables. In the next lemma we characterize the domain of problem (13).

Lemma III.5. *The domain D_η of η in* (11) *is $\mathbb{R}^{S\cdot|\mathcal{E}|}$, thus optimization problem* (13) *is unconstrained.*

Proof. We show that each $\eta_i\big(\{\boldsymbol{\lambda}_{ij}, \boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big)$ is finite for all $\{\boldsymbol{\lambda}_{ij}, \boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}$. Each function $q_i(\boldsymbol{\mu}_i)$ is concave on its domain $\boldsymbol{\mu}_i \geq \mathbf{0}$ for all $i \in \{1,\dots,N\}$. In fact, from the definition of q_i in (5), we notice that it is obtained as minimization over a nonempty compact set X_i of the function $f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \mathbf{g}_i(\mathbf{x}_i)$. Such a function is concave (in fact linear) in $\boldsymbol{\mu}_i$, thus, following the proof of [35, Proposition 5.1.2], we can conclude that every q_i is concave over its convex domain $\boldsymbol{\mu}_i \geq \mathbf{0}$. For each $i \in \{1,\dots,N\}$, the function η_i as defined in (12) is obtained by maximizing a (concave) continuous function (q_i plus a linear term) over a compact set and, thus, has always a finite value, so that the proof follows. □

It is worth noting that Lemma III.5 strongly relies on the compactness of $\{\boldsymbol{\mu}_i \geq \mathbf{0} \mid \boldsymbol{\mu}_i^\top \mathbf{1} \leq M\}$. This means that without the primal relaxation, D_η is not guaranteed to be $\mathbb{R}^{S\cdot|\mathcal{E}|}$. In Section IV-C, we better clarify this aspect.

Next we characterize the optimization problem (13).

Lemma III.6. *Let M be such that $M > \|\boldsymbol{\mu}^\star\|_1$ with $\boldsymbol{\mu}^\star$ an optimal solution of the dual of problem* (1)*. Problem* (13) *has a bounded optimal cost, call it $\eta^\star$, and a nonempty optimal solution set. Moreover, it enjoys strong duality with* (9)*. Also, it holds $\eta^\star = f^\star$, where $f^\star$ is the optimal solution of the original primal problem* (1)*.*

Proof. Since (9) is equivalent to (4), then by Lemma III.1 its optimal cost is finite and equal to $q^\star$. Since each q_i is concave as shown in the proof of Lemma III.5, then also $\sum_{i=1}^N q_i(\boldsymbol{\mu}_i)$ is a concave function of $(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_N)$. Thus, since the domain $\{\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_N \mid \boldsymbol{\mu}_i \geq \mathbf{0} \mid \boldsymbol{\mu}_i^\top\mathbf{1} \leq M, i \in \{1,\dots,N\}\}$, is polyhedral, by [35, Proposition 5.2.1] strong duality between problem (9) and its dual (13) holds, i.e., $\eta^\star$ is finite since it holds $\eta^\star = q^\star$. From the same proposition, we have that the optimal solution set of (13) is nonempty. The equality $\eta^\star = f^\star$ follows readily by strong duality between problems (1) and (4), which concludes the proof. □

C. Distributed Subgradient Method

We detail in this subsection how to explicitly design a distributed dual decomposition algorithm to solve problem (6) based on a subgradient iteration applied to problem (13).

Exploiting the separability of η in (11), we recall how to compute each component of a subgradient of η at a given $\boldsymbol{\Lambda} \in \mathbb{R}^{S\cdot|\mathcal{E}|}$, see e.g., [35, Section 6.1] That is, it holds

$$\frac{\tilde{\partial}\eta(\boldsymbol{\Lambda})}{\partial \boldsymbol{\lambda}_{ij}} = \boldsymbol{\mu}_i^\star - \boldsymbol{\mu}_j^\star, \tag{14}$$

where $\frac{\tilde{\partial}\eta(\cdot)}{\partial\boldsymbol{\lambda}_{ij}}$ denotes the component associated to the variable $\boldsymbol{\lambda}_{ij}$ of a subgradient of η, and

$$\boldsymbol{\mu}_k^\star \in \operatorname*{argmax}_{\boldsymbol{\mu}_k\geq\mathbf{0}, \boldsymbol{\mu}_k^\top\mathbf{1}\leq M} \Big(q_k(\boldsymbol{\mu}_k) + \boldsymbol{\mu}_k^\top \sum_{h\in\mathcal{N}_k} (\boldsymbol{\lambda}_{kh} - \boldsymbol{\lambda}_{hk})\Big), \tag{15}$$

for $k = i, j$.

Having recalled how to compute subgradients of η, we are ready to summarize how the subgradient method reads when applied to problem (13). At each iteration t, each node i:

(S1) receives $\boldsymbol{\lambda}_{ji}^t$, $j \in \mathcal{N}_i$, and computes $\boldsymbol{\mu}_i^{t+1}$ as an optimal solution of

$$\max_{\boldsymbol{\mu}_i\geq\mathbf{0}, \boldsymbol{\mu}_i^\top\mathbf{1}\leq M} \Big(q_i(\boldsymbol{\mu}_i) + \boldsymbol{\mu}_i^\top \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t)\Big); \tag{16}$$

(S2) receives the updated $\boldsymbol{\mu}_j^{t+1}$, $j \in \mathcal{N}_i$ and updates $\boldsymbol{\lambda}_{ij}$, $j \in \mathcal{N}_i$, via

$$\boldsymbol{\lambda}_{ij}^{t+1} = \boldsymbol{\lambda}_{ij}^t - \gamma^t(\boldsymbol{\mu}_i^{t+1} - \boldsymbol{\mu}_j^{t+1}),$$

where γ^t is the step-size.

Notice that (S1)–(S2) is a distributed algorithm, i.e., it can be implemented by means of local computations and communications without any centralized step. However, we want to stress that the algorithm is *not* implementable as it is written, since the functions q_i are still not explicitly available.

The next lemma states a property on the subdifferential of the convex function η.

Lemma III.7. *The subgradients of η are uniformly bounded, i.e., there exists a positive constant C such that for every $\boldsymbol{\Lambda} \in \mathbb{R}^{S\cdot|\mathcal{E}|}$ with components $\boldsymbol{\lambda}_{ij} \in \mathbb{R}^S$, $\forall\,(i,j) \in \mathcal{E}$, any subgradient $\tilde{\partial}\eta(\boldsymbol{\Lambda})/\partial\boldsymbol{\Lambda}$ satisfies $\|\frac{\tilde{\partial}\eta(\boldsymbol{\Lambda})}{\partial\boldsymbol{\Lambda}}\| \leq C$.*

Proof. To prove the lemma, we show that each component $\frac{\tilde{\partial}\eta(\boldsymbol{\Lambda})}{\partial\boldsymbol{\lambda}_{ij}}$ of $\frac{\tilde{\partial}\eta(\boldsymbol{\Lambda})}{\partial\boldsymbol{\Lambda}}$ is bounded. Using (14), it is sufficient to show that $\boldsymbol{\mu}_i^\star$ and $\boldsymbol{\mu}_j^\star$, associated to the given $\boldsymbol{\Lambda}$, are uniformly bounded and, hence, their difference. Since, from equation (15) the two are obtained as maxima of a concave function over a compact domain, the proof follows. □

IV. Algorithm Analysis: Convergence Proof

This section is devoted to prove the convergence of the RSDD distributed algorithm formally stated in Theorem II.6.

A. Preparatory Results

We give two intermediate results that represent building blocks for the convergence proof given in Section IV-B. The next lemma is instrumental to the second one.

Lemma IV.1. *Consider the optimization problem*

$$\begin{aligned} \max_{\boldsymbol{\mu}_i}\ & f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top\Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t)\Big) \\ \text{subj. to }\ & \boldsymbol{\mu}_i \geq \mathbf{0}, \boldsymbol{\mu}_i^\top\mathbf{1} \leq M, \end{aligned} \tag{17}$$

with given $\mathbf{x}_i$, $\boldsymbol{\lambda}_{ij}^t$ *and* $\boldsymbol{\lambda}_{ji}^t$, $j \in \mathcal{N}_i$, *and* $M > 0$. *Then, its dual problem is*

$$\begin{aligned} \min_{\rho_i} \ & f_i(\mathbf{x}_i) + M\rho_i \\ \text{subj. to } & \rho_i \geq 0 \\ & \mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \leq \rho_i \mathbf{1}, \end{aligned} \tag{18}$$

and strong duality holds.

Proof. First, since $\mathbf{x}_i$, $\boldsymbol{\lambda}_{ij}^t$ and $\boldsymbol{\lambda}_{ji}^t$ are given, problem (17) is a feasible *linear* program (the box constraint is nonempty) with compact domain. Thus, both problem (17) and its dual have finite optimal cost and strong duality holds.

In order to show that (18) is the dual of (17), we introduce a multiplier $\rho_i \geq 0$ associated to the constraint $M - \boldsymbol{\mu}_i^\top \mathbf{1} \geq 0$. Then the dual function of (17) is defined as $\max_{\boldsymbol{\mu}_i \geq 0} f_i(\mathbf{x}_i) + M\rho_i + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) - \rho_i \mathbf{1} \Big)$. It is equal to $f_i(\mathbf{x}_i) + M\rho_i$ if $\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) - \rho_i \mathbf{1} \leq \mathbf{0}$ and $+\infty$ otherwise. Finally, the minimization of the dual function on its domain with respect to $\rho_i \geq 0$ gives problem (18) and concludes the proof. □

In the following, we propose a technique to make step (16) explicit. By plugging in (16) the definition of q_i, given in (5), the following max-min optimization problem is obtained:

$$\max_{\substack{\boldsymbol{\mu}_i \geq \mathbf{0}, \\ \boldsymbol{\mu}_i^\top \mathbf{1} \leq M}} \ \min_{\mathbf{x}_i \in X_i} \Bigg(f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \Big) \Bigg). \tag{19}$$

The next lemma allows us to recast problem (19) in a more convenient formulation from a computational point of view.

Lemma IV.2. *Consider the optimization problem*

$$\begin{aligned} \min_{\mathbf{x}_i, \rho_i} \ & f_i(\mathbf{x}_i) + M\rho_i \\ \text{subj. to } & \rho_i \geq 0, \ \ \mathbf{x}_i \in X_i \\ & \mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} \big(\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t\big) \leq \rho_i \mathbf{1}. \end{aligned} \tag{20}$$

A finite primal-dual optimal solution pair of (20)*, call it* $\big((\mathbf{x}_i^{t+1}, \rho_i^{t+1}), \boldsymbol{\mu}_i^{t+1}\big)$*, does exist and* $(\mathbf{x}_i^{t+1}, \boldsymbol{\mu}_i^{t+1})$ *is a solution of* (19)*.*

Proof. Problem (20) is a feasible convex program, in fact $f_i(\mathbf{x}_i) + M\rho_i$ is convex, the set X_i is nonempty, convex and compact, the constraint $\rho_i \geq 0$ is convex as well as the inequality constraint $\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} \big(\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t\big) - \rho_i \mathbf{1} \leq \mathbf{0}$. Then, by choosing a sufficiently large ρ_i, we can show that the Slater's constraint qualification is satisfied and, thus, strong duality holds. Therefore, a primal-dual optimal solution pair $(\mathbf{x}_i^{t+1}, \rho_i^{t+1}, \boldsymbol{\mu}_i^{t+1})$ of (20) exists. Moreover, problem (20) can be recast as

$$\min_{\mathbf{x}_i \in X_i} \Bigg(\min_{\rho_i \geq 0,\, \mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \leq \rho_i \mathbf{1}} f_i(\mathbf{x}_i) + M\rho_i \Bigg).$$

By Lemma IV.1, we can substitute the inner minimization with its equivalent dual maximization obtaining

$$\min_{\mathbf{x}_i \in X_i} \ \max_{\substack{\boldsymbol{\mu}_i \geq \mathbf{0}, \\ \boldsymbol{\mu}_i^\top \mathbf{1} \leq M}} \Bigg(f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \Big) \Bigg). \tag{21}$$

Let $\phi(\mathbf{x}_i, \boldsymbol{\mu}_i) = f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \Big)$ and observe that (i) $\phi(\cdot, \boldsymbol{\mu}_i)$ is closed and convex for all $\boldsymbol{\mu}_i \geq \mathbf{0}$ (affine transformation of a convex function with compact domain X_i) and (ii) $\phi(\mathbf{x}_i, \cdot)$ is closed and concave since it is a linear function with compact domain $(\{\boldsymbol{\mu}_i \geq \mathbf{0} \mid \boldsymbol{\mu}_i^\top \mathbf{1} \leq M\})$, for all $\mathbf{x}_i \in \mathbb{R}^S$. Thus, we can invoke [37, Propositions 4.3] to switch min and max operators in (21), and write

$$\begin{aligned} & \min_{\mathbf{x}_i \in X_i} \ \max_{\substack{\boldsymbol{\mu}_i \geq \mathbf{0}, \\ \boldsymbol{\mu}_i^\top \mathbf{1} \leq M}} \Bigg(f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \Big) \Bigg) \\ & = \max_{\substack{\boldsymbol{\mu}_i \geq \mathbf{0}, \\ \boldsymbol{\mu}_i^\top \mathbf{1} \leq M}} \ \min_{\mathbf{x}_i \in X_i} \Bigg(f_i(\mathbf{x}_i) + \boldsymbol{\mu}_i^\top \Big(\mathbf{g}_i(\mathbf{x}_i) + \sum_{j\in\mathcal{N}_i} (\boldsymbol{\lambda}_{ij}^t - \boldsymbol{\lambda}_{ji}^t) \Big) \Bigg). \end{aligned} \tag{22}$$

which is (19), thus concluding the proof. □

We highlight that problem (20) is the local optimization step (2) in the RSDD distributed algorithm.

Finally, the next corollary makes a connection between the optimal cost of i-th problem (20) and the value of the i-th local term η_i (defined in (12)) of the second dual function η (defined in (11)).

Corollary IV.3. *Let* $(\mathbf{x}_i^{t+1}, \rho_i^{t+1})$ *be a solution of* (20) *with given* $\boldsymbol{\lambda}_{ij}^t$ *and* $\boldsymbol{\lambda}_{ji}^t$ *for* $j \in \mathcal{N}_i$*. Then*

$$\eta_i\big(\{\boldsymbol{\lambda}_{ij}^t, \boldsymbol{\lambda}_{ji}^t\}_{j\in\mathcal{N}_i}\big) = f_i(\mathbf{x}_i^{t+1}) + M\rho_i^{t+1}, \tag{23}$$

with η_i *defined in* (12)*.*

Proof. In the proof of Lemma IV.2, we have shown that condition (22) holds for all $t \geq 0$. Its left hand side has optimal cost $f_i(\mathbf{x}_i{}^{t+1}) + M\rho_i^{t+1}$, while the one of the right hand side is exactly the definition of $\eta_i\big(\{\boldsymbol{\lambda}_{ij}^t, \boldsymbol{\lambda}_{ji}^t\}_{j\in\mathcal{N}_i}\big)$ in (12). Thus, equation (22) can be rewritten as

$$f_i(\mathbf{x}_i^{t+1}) + M\rho_i^{t+1} = \eta_i\big(\{\boldsymbol{\lambda}_{ij}^t, \boldsymbol{\lambda}_{ji}^t\}_{j\in\mathcal{N}_i}\big),$$

for all $i \in \{1, \ldots, N\}$, concluding the proof. □

B. Proof of Theorem II.6

To prove statement *(i)*, we show that the RSDD distributed algorithm is an operative way to implement the subgradient method (S1)–(S2) and, that (S1)–(S2) solves problem (6).

First, let $\{\boldsymbol{\mu}_i^t\}_{t\geq 0}$ and $\{\boldsymbol{\lambda}_{ij}^t\}_{t\geq 0}$, $j \in \mathcal{N}_i$, be the auxiliary sequences generated by the RSDD distributed algorithm associated to $\{(\mathbf{x}_i^t, \rho_i^t)\}_{t\geq 0}$, for each $i \in \{1, \ldots, N\}$. From Lemma IV.2, a primal-dual optimal solution pair $\big((\mathbf{x}_i^t, \rho_i^t), \boldsymbol{\mu}_i^t\big)$ of (2) in fact exists at each iteration t, so that the algorithm is well-posed. Second, to show that RSDD implements (S1)–(S2) we notice that update (3) and (S2) are trivially identical. As for (S1), we have shown in the discussion after Lemma IV.1, that equation (19) is an explicit expression for (16) in (S1). Thus, by invoking Lemma IV.2, we can conclude that finding the dual part of a primal-dual optimal solution pair of (2) corresponds to performing (S1). Therefore, the sequences $\{\boldsymbol{\lambda}_{ij}^t\}_{t\geq 0}$, $(i,j) \in \mathcal{E}$ generated by RSDD and by (S1)–(S2) coincide. Third, we show that RSDD solves problem (13). By Lemma III.6 the optimal solution set of (13) is nonempty and by Lemma III.7 the subgradients of η are uniformly bounded. Since the step-size γ^t satisfies Assumption II.5, we can

invoke [38, Proposition 3.2.6] to conclude that the sequence $\{\boldsymbol{\lambda}_{ij}^t\}_{t\geq 0}$, $(i,j)\in\mathcal{E}$ generated by RSDD (or equivalently by (S1)–(S2)) converges to an optimal solution of (13). Then, we use (23) in Corollary IV.3 and take the limit as $t\to\infty$, thus obtaining

$$\begin{aligned}\lim_{t\to\infty}\sum_{i=1}^N\Big(f_i(\mathbf{x}_i^{t+1})+M\rho_i^{t+1}\Big)&=\lim_{t\to\infty}\sum_{i=1}^N\eta_i\big(\{\boldsymbol{\lambda}_{ij}^t,\boldsymbol{\lambda}_{ji}^t\}_{j\in\mathcal{N}_i}\big)\\&=\eta^\star=f^\star,\end{aligned}$$

where the last equality follows by Lemma III.6, so that the proof of the first statement is complete.

To prove statement *(ii)*, i.e., the primal recovery property, we start by studying the properties of the aggregated vector $(\mathbf{x}_1^t,\dots,\mathbf{x}_N^t,\rho_1^t,\dots,\rho_N^t)$. By construction, for all $t\geq 0$ each pair $(\mathbf{x}_i^t,\rho_i^t)$ satisfies $\mathbf{x}_i^t\in X_i$, $\rho_i^t\geq 0$ and $\mathbf{g}_i(\mathbf{x}_i^t)+\sum_{j\in\mathcal{N}_i}\big(\boldsymbol{\lambda}_{ij}^{t-1}-\boldsymbol{\lambda}_{ji}^{t-1}\big)\leq\rho_i^t\mathbf{1}$. Summing over $i\in\{1,\dots,N\}$ the previous condition, it follows that

$$\sum_{i=1}^N\mathbf{g}_i(\mathbf{x}_i^t)+\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}\big(\boldsymbol{\lambda}_{ij}^{t-1}-\boldsymbol{\lambda}_{ji}^{t-1}\big)\leq\sum_{i=1}^N\rho_i^t\mathbf{1},\tag{24}$$

for all $t\geq 0$. Let us denote by a_{ij} the (i,j)-th entry of the adjacency matrix associated to the undirected graph $\mathcal{G}$. Then, we can write

$$\begin{aligned}\sum_{i=1}^N\sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}_{ij}^t-\boldsymbol{\lambda}_{ji}^t)&\overset{(a)}{=}\sum_{i=1}^N\sum_{j=1}^N a_{ij}(\boldsymbol{\lambda}_{ij}^t-\boldsymbol{\lambda}_{ji}^t)\\&=\sum_{i=1}^N\sum_{j=1}^N a_{ij}\boldsymbol{\lambda}_{ij}^t-\sum_{i=1}^N\sum_{j=1}^N a_{ij}\boldsymbol{\lambda}_{ji}^t\overset{(b)}{=}0,\end{aligned}$$

where (a) follows by writing the sum over neighboring agents in terms of the adjacency matrix, while (b) holds since $\mathcal{G}$ is undirected (so that $a_{ij}=a_{ji}$ for all $(i,j)\in\mathcal{E}$), which implies that the two summations in the second line are identical for all $t\geq 0$. Hence, equation (24) reduces to

$$\sum_{i=1}^N\mathbf{g}_i(\mathbf{x}_i^t)\leq\sum_{i=1}^N\rho_i^t\mathbf{1},\tag{25}$$

for all $t\geq 0$. Equation (25) shows that for all $t\geq 0$ the aggregate vector $(\mathbf{x}_1^t,\dots,\mathbf{x}_N^t,\rho_1^t,\dots,\rho_N^t)$ is feasible for the following optimization problem

$$\begin{aligned}\min_{\substack{\mathbf{x}_1,\dots,\mathbf{x}_N\\ \rho_1,\dots,\rho_N}}\ &\sum_{i=1}^N f_i(\mathbf{x}_i)+M\sum_{i=1}^N\rho_i\\ \text{subj. to }\ &\rho_i\geq 0,\ \ \mathbf{x}_i\in X_i,\ \ i\in\{1,\dots,N\}\\ &\sum_{i=1}^N\mathbf{g}_i(\mathbf{x}_i)\leq\sum_{i=1}^N\rho_i\mathbf{1}.\end{aligned}\tag{26}$$

Notice that, by defining $\rho=\sum_{i=1}^N\rho_i$, problem (26) is equivalent to problem (7). Thus, at each iteration t the point $(\mathbf{x}_1^t,\dots,\mathbf{x}_N^t,\sum_{i=1}^N\rho_i^t)$ is feasible for problem (7). The equivalence also shows that ρ_i is not a copy of ρ, but it is the i-th contribution to ρ.

We now show that every limit point of the sequence $\{\mathbf{x}_1^t,\dots,\mathbf{x}_N^t,\sum_{i=1}^N\rho_i^t\}_{t\geq 0}$ is feasible for problem (7). By construction, each $\mathbf{x}_i^t\in X_i$ for all $i\in\{1,\dots,N\}$, so that $\{\mathbf{x}_i^t\}_{t\geq 0}$ is bounded. Moreover, from the statement *(i)* of the theorem, also the sequence $\{\sum_{i=1}^N\rho_i^t\}_{t\geq 0}$ is bounded since $\{\sum_{i=1}^N f_i(\mathbf{x}_i^t)+M\sum_{i=1}^N\rho_i^t\}_{t\geq 0}$ converges to a finite value $f^\star$. Since the sequence of vectors $\{(\mathbf{x}_1^t,\dots,\mathbf{x}_N^t,\sum_{i=1}^N\rho_i^t)\}_{t\geq 0}$ is bounded, then there exists a sub-sequence of indices $\{t_n\}_{n\geq 0}\subseteq\{t\}_{t\geq 0}$ such that the sub-sequence $\{(\mathbf{x}_1^{t_n},\dots,\mathbf{x}_N^{t_n},\sum_{i=1}^N\rho_i^{t_n})\}_{n\geq 0}$ converges to a limit point $(\bar{\mathbf{x}}_1,\dots,\bar{\mathbf{x}}_N,\bar\rho)$. From the first statement of the theorem we have that $(\bar{\mathbf{x}}_1,\dots,\bar{\mathbf{x}}_N,\bar\rho)$ satisfies

$$\sum_{i=1}^N f_i(\bar{\mathbf{x}}_i)+M\bar\rho=f^\star.$$

Moreover, since each component of $\mathbf{g}_i$ is a (finite) convex function over $\mathbb{R}^{n_i}$, it is also continuous over any compact subset of $\mathbb{R}^{n_i}$. Thus, by taking the limit as $n\to\infty$ in (25) with $t=t_n$, it also holds

$$\sum_{i=1}^N\mathbf{g}_i(\bar{\mathbf{x}}_i)\leq\bar\rho\,\mathbf{1}.\tag{27}$$

By Proposition III.3 it must hold that $(\bar{\mathbf{x}}_1,\dots,\bar{\mathbf{x}}_N,\bar\rho)=(\bar{\mathbf{x}}_1,\dots,\bar{\mathbf{x}}_N,0)$, i.e., $\bar\rho=0$. Thus, (27) holds with $\bar\rho=0$ and, thus, guarantees that every limit point of $(\mathbf{x}_1^t,\dots,\mathbf{x}_N^t)$ is feasible for the (not relaxed) coupling constraint in the original problem (1) and thus optimal for that problem. So that the proof follows.

C. Discussion on the Necessity of the Relaxation

In this subsection we show how our approach reads when no dual restriction is applied. This will further highlight the strength of the proposed strategy.

Suppose we do not restrict the original dual problem (4), but we still apply the same formal derivation given in the previous sections. Then, the counterpart of (11) is $\eta^{\text{NR}}(\boldsymbol{\Lambda})=\sum_{i=1}^N\eta_i^{\text{NR}}\big(\{\boldsymbol{\lambda}_{ij},\boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big)$ with, for all $i\in\{1,\dots,N\}$,

$$\eta_i^{\text{NR}}\big(\{\boldsymbol{\lambda}_{ij},\boldsymbol{\lambda}_{ji}\}_{j\in\mathcal{N}_i}\big)=\sup_{\boldsymbol{\mu}_i\geq\mathbf{0}}\Big(q_i(\boldsymbol{\mu}_i)+\boldsymbol{\mu}_i^\top\sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}_{ij}-\boldsymbol{\lambda}_{ji})\Big).$$

Finally, by denoting the domain of η^{NR} as D_η^{NR}, we have that the counterpart of problem (13) is $\min_{\boldsymbol{\Lambda}\in D_\eta^{\text{NR}}}\eta^{\text{NR}}(\boldsymbol{\Lambda})$. Notice that this problem is a constrained minimization since, differently from the relaxed case, the domain D_η^{NR} does not always coincide with the entire space $\mathbb{R}^{S\cdot|\mathcal{E}|}$ (Cf. Lemma III.5). Thus, to apply the subgradient method we need to adapt (S1)–(S2) by appending an additional projection step, i.e., $\boldsymbol{\Lambda}^{t+1}=\big[\widetilde{\boldsymbol{\Lambda}}^{t+1}\big]_{D_\eta^{\text{NR}}}$, where each component $\widetilde{\boldsymbol{\lambda}}_{ij}^{t+1}$ of $\widetilde{\boldsymbol{\Lambda}}^{t+1}$ is the result of (S2) and $[\,\cdot\,]_{D_\eta^{\text{NR}}}$ denotes the Euclidean projection onto D_η^{NR}. Notice that the projection onto D_η^{NR} of the entire $\widetilde{\boldsymbol{\Lambda}}^{t+1}$ prevents the distributed implementation of the algorithm.

It is worth noting that, being the set $\{\boldsymbol{\mu}_i\in\mathbb{R}^S\mid\boldsymbol{\mu}_i\geq\mathbf{0}\}$ not compact, then Lemma IV.2 would not hold. The theoretical issue is that switching $\min$ and $\max$ operators in (22) may not be possible due to the non-compact domains (Cf. [37, Propositions 4.3]. Moreover, differently from the relaxed case, Lemma III.7 does not hold anymore so that no guarantees about the boundedness of subgradients of η^{NR} can be established. Thus, the convergence result about the centralized subgradient method cannot be invoked.

V. Application to Distributed Microgrid Control

In this section we present a computational study of our RSDD distributed algorithm tailored for an optimization problem to be solved within a distributed MPC scheme for microgrid control. We consider a simplified microgrid model (with both static and dynamical units) to compute an optimal power profile in a given horizon $[0, T]$.

A. Microgrid Model

A microgrid consists of generators, controllable loads, storage devices and a connection to the main grid [39]. Generators are collected in the set $\mathtt{GEN}$, storage devices are collected in $\mathtt{STOR}$, controllable loads are collected in $\mathtt{CONL}$ while the device $\mathtt{tr}$ is the connection node with the main grid. Formally, the constraint-coupled optimization problem to be solved is

$$\begin{aligned}\min \sum_{\tau=0}^{T} \Bigg(& \sum_{i\in\mathtt{GEN}} f^{\tau}_{\mathtt{gen},i}(p^{\tau}_{\mathtt{gen},i}) \\ & + \sum_{i\in\mathtt{CONL}} f^{\tau}_{\mathtt{conl},i}(p^{\tau}_{\mathtt{conl},i}) + f^{\tau}_{\mathtt{tr}}(p^{\tau}_{\mathtt{tr}}) \Bigg) \\ \text{subj. to } & p_{\mathtt{gen},i} \in X_{\mathtt{gen}},\ i \in \mathtt{GEN},\ p_{\mathtt{stor},i} \in X_{\mathtt{stor}},\ i \in \mathtt{STOR} \\ & p_{\mathtt{conl},i} \in X_{\mathtt{conl}},\ i \in \mathtt{CONL},\ |p^{\tau}_{\mathtt{tr}}| \le E,\ \tau \in [0,T] \\ & \sum_{i\in\mathtt{GEN}} p^{\tau}_{\mathtt{gen},i} + \sum_{i\in\mathtt{STOR}} p^{\tau}_{\mathtt{stor},i} \\ & \qquad + \sum_{i\in\mathtt{CONL}} p^{\tau}_{\mathtt{conl},i} + p^{\tau}_{\mathtt{tr}} - D^{\tau} \le 0,\ \tau \in [0,T],\end{aligned}$$

where the optimization variables are $p_{\mathtt{gen},i}$, $p_{\mathtt{stor},i}$, $p_{\mathtt{conl},i}$, $p_{\mathtt{tr}} \in \mathbb{R}^{T+1}$. For the generators, at each time instant τ, the power $p^{\tau}_{\mathtt{gen},i}$ must satisfy magnitude and rate bounds, $\underline{p} \le p^{\tau}_{\mathtt{gen},i} \le \bar{p}$ and $\underline{r} \le p^{\tau+1}_{\mathtt{gen},i} - p^{\tau}_{\mathtt{gen},i} \le \bar{r}$, with $\tau \in [0, T-1]$, for given positive scalars $\underline{p}$, $\bar{p}$, $\underline{r}$ and $\bar{r}$; the cost to produce power by a generator is modeled as a quadratic function $f^{\tau}_{\mathtt{gen},i}(p^{\tau}_{\mathtt{gen},i}) = \alpha_1 p^{\tau}_{\mathtt{gen},i} + \alpha_2 (p^{\tau}_{\mathtt{gen},i})^2$ for some $\alpha_1 > 0$ and $\alpha_2 > 0$. The power of each storage device $p^{\tau}_{\mathtt{stor},i}$ satisfies bounds and a dynamical constraint given by $-d_{\mathtt{stor}} \le p^{\tau}_{\mathtt{stor},i} \le c_{\mathtt{stor}}$, $\tau \in [0,T]$, $q^{\tau+1}_{\mathtt{stor},i} = q^{\tau}_{\mathtt{stor},i} + p^{\tau}_{\mathtt{stor},i}$, $\tau \in [0, T-1]$, and $0 \le q^{\tau}_{\mathtt{stor},i} \le q_{\max}$, $\tau \in [0,T]$, where $q^{\tau}_{\mathtt{stor},i}, \tau \in [0,T]$, describes the storage capacity and the initial value $q^{0}_{\mathtt{stor},i}$ is given and $d_{\mathtt{stor}}$, $c_{\mathtt{stor}}$ and $q_{\max}$ are positive scalars. There are no costs associated with the stored power. The power of controllable loads $p^{\tau}_{\mathtt{conl},i}$ must follow a desired load profile $p^{\tau}_{\mathtt{des},i}$ for $p^{\tau}_{\mathtt{conl},i}$. The controllable load incurs in a cost $f^{\tau}_{\mathtt{conl},i}(p^{\tau}_{\mathtt{conl},i}) = \beta \max\{0, p^{\tau}_{\mathtt{des},i} - p^{\tau}_{\mathtt{conl},i}\}$, $\beta \ge 0$, if the desired load is not matched. Upper and lower bounds on the $p^{\tau}_{\mathtt{conl},i}$ are imposed to avoid meaningless power values. The power of the connection node $p^{\tau}_{\mathtt{tr}}$ incurs in a trading cost modeled as $f^{\tau}_{\mathtt{tr}}(p^{\tau}_{\mathtt{tr}}) = -c_1 p^{\tau}_{\mathtt{tr}} + c_2 |p^{\tau}_{\mathtt{tr}}|$, with $c_1 > 0$ and $c_2 > 0$ being respectively the price and the transaction cost. Finally, the power network must satisfy a given power demand D^{τ} modeled by the coupling constraint. Reasonably, we assume D^{τ} to be known only by the connection node $\mathtt{tr}$.

B. Numerical Results

We consider a heterogeneous network of $N = 10$ units with 4 generators, 3 storage devices, 2 controllable loads and 1 connection to the main grid. We assume that in the distributed MPC scheme each unit predicts its power generation strategy over a horizon of $T = 12$ slots. In order to fit the microgrid control problem in our set-up, we let each $\mathbf{x}_i$ be the whole trajectory over the prediction horizon $[0, T]$, e.g., $\mathbf{x}_i = [p^{0}_{\mathtt{gen},i}, \dots, p^{T}_{\mathtt{gen},i}]^{\top}$, for all the generators $i \in \mathtt{GEN}$ and, consistently, for the other device types. As for the cost functions we define $f_i(\mathbf{x}_i) = \sum_{\tau=0}^{T} f^{\tau}_{\mathtt{gen},i}(p^{\tau}_{\mathtt{gen},i})$ for $i \in \mathtt{GEN}$ and, similarly, for the other device types. Each local constraint X_i encodes the heterogeneous dynamics and bounds on the state of the units. We set $M = 10$, where $\boldsymbol{\mu}^{\star}$ is a dual optimal solution of the problem computed in a centralized way and $\gamma^t = 0.1 \cdot (1/t)^{0.7}$.

In Figure 2 we show the algorithmic evolution of the sum of the penalty parameters ρ^t_i and the maximum violation of the coupling constraint at each iteration t. As claimed in Theorem II.6, $\sum_{i=1}^{N} \rho^t_i$ asymptotically goes to zero. In this particular instance we also notice that, after the very first iterations, the generated points are strictly feasible for the coupling constraints and hit the boundary in the limit from the interior. We point out that feasibility of the coupling constraint is obtained during the transient, even if some ρ^t_i are positive, so that the algorithm would not work without the relaxation strategy. In Figure 3 we show how $\sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}^t_{ij} - \boldsymbol{\lambda}^t_{ji})$ compares with the unknown part of the coupling constraint of each agent i, namely $\sum_{j\neq i} \mathbf{g}_j(\mathbf{x}^t_j)$. The picture highlights that $\sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}^t_{ij} - \boldsymbol{\lambda}^t_{ji})$ actually distributedly "tracks" the maximum of the contribution in the coupling constraint due to all the other agents $j \neq i$ in the network.

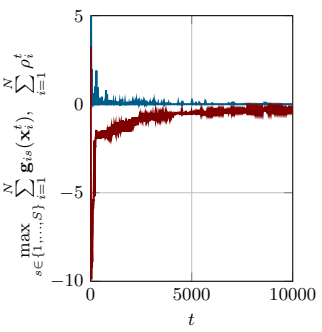

Figure 2. Asymptotically vanishing behavior of the sum of local violations (blue). Evolution of the maximum violation of coupling constraints showing feasibility of primal sequences (red).

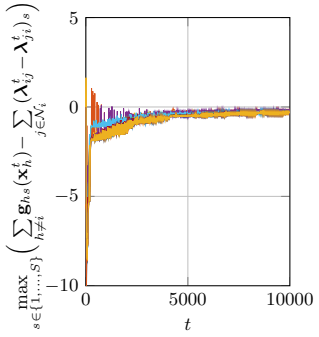

Figure 3. Evolution of the maximum difference (over the components $s \in \{1, \dots, S\}$) between $\sum_{h\neq i} \mathbf{g}_{hs}(\mathbf{x}^t_h)$ and $\sum_{j\in\mathcal{N}_i}(\boldsymbol{\lambda}^t_{ij} - \boldsymbol{\lambda}^t_{ji})$, for all i.

Finally, in Figure 4 we compare the convergence rate of the RSDD and the distributed subgradient [2] applied to (13) equipped with a running average to recover a primal feasible solution, [28]. That is, we plot the difference between the optimal cost $\eta^{\star} = f^{\star}$ and the sum of local costs $\sum_{i=1}^{N} f_i(\mathbf{x}^t_i$ normalized by $|f^{\star}|$. It can be seen that both algorithms converge to the optimal cost $f^{\star}$ in a non-monotone fashion with a sublinear rate. However our algorithm is faster since no averaging mechanisms are used.

VI. Conclusions

In this paper we have proposed a novel distributed method to solve constraint-coupled convex optimization problems in which a separable cost function is minimized subject to both local constraints involving one component of the decision vector and coupling constraints involving all the components.

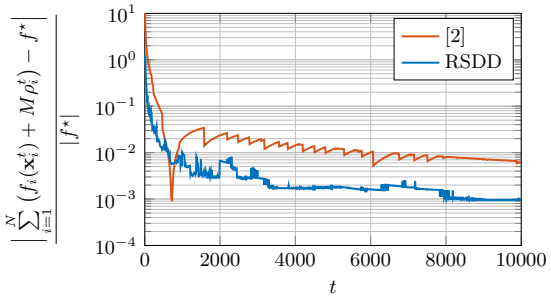

Figure 4. Evolution of the cost error $|\sum_{i=1}^{N}\left(f_i(\mathbf{x}_i^t)+M\rho_i^t\right)-f^\star|/|f^\star|$ that shows convergence to the optimal cost.

While the algorithm has a very simple structure (a local optimization and a linear update), its analysis involves a relaxation approach and a deep tour into duality theory showing both the convergence to the optimal cost and the primal recovery. In particular, this last property allows each node to compute its portion of the optimal solution without resorting to any averaging mechanism, which is instead commonly required in methods based on dual decomposition. Numerical computations on an instance of a cooperative Distributed Model Predictive Control scheme in smart microgrids have corroborated the theoretical results.

Acknowledgment

The authors would like to thank Andrea Camisa for the deep discussions and useful suggestions.

References

[1] I. Notarnicola and G. Notarstefano, "A duality-based approach for distributed optimization with coupling constraints," in *IFAC World Congress*, 2017, pp. 14 891–14 896.

[2] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. on Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.

[3] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. on Autom. Control*, vol. 55, no. 4, pp. 922–938, 2010.

[4] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. on Autom. Control*, vol. 57, no. 3, pp. 592–606, 2012.

[5] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Trans. on Autom. Control*, vol. 57, no. 1, pp. 151–164, 2012.

[6] D. Jakovetic, J. Xavier, and J. M. Moura, "Fast distributed gradient methods," *IEEE Trans. on Autom. Control*, vol. 59, no. 5, pp. 1131–1146, 2014.

[7] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. on Autom. Control*, vol. 60, no. 3, pp. 601–615, 2015.

[8] P. Di Lorenzo and G. Scutari, "Next: In-network nonconvex optimization," *IEEE Trans. on Signal and Inform. Proc. over Networks*, vol. 2, no. 2, pp. 120–136, 2016.

[9] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," *IEEE Trans. on Autom. Control*, vol. 61, no. 4, pp. 994–1009, 2016.

[10] I. Notarnicola and G. Notarstefano, "Asynchronous distributed optimization via randomized dual proximal gradient," *IEEE Trans. on Autom. Control*, vol. 62, no. 5, pp. 2095–2106, 2017.

[11] P. D. Christofides, R. Scattolini, D. M. de la Peña, and J. Liu, "Distributed model predictive control: A tutorial review and future research directions," *Computers & Chemical Engin.*, vol. 51, pp. 21–41, 2013.

[12] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice Hall Englewood Cliffs, NJ, 1989, vol. 23.

[13] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.

[14] I. Necoara, V. Nedelcu, and I. Dumitrache, "Parallel and distributed optimization methods for estimation and control in networks," *Journal of Process Control*, vol. 21, no. 5, pp. 756–766, 2011.

[15] P. Giselsson, M. D. Doan, T. Keviczky, B. De Schutter, and A. Rantzer, "Accelerated gradient methods and dual decomposition in distributed model predictive control," *Automatica*, vol. 49, no. 3, pp. 829–833, 2013.

[16] Q. Tran-Dinh, I. Necoara, and M. Diehl, "A dual decomposition algorithm for separable nonconvex optimization using the penalty function framework," in *IEEE Conf. on Dec. and Control (CDC)*, 2013, pp. 2372–2377.

[17] I. Necoara and V. Nedelcu, "On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems," *Automatica*, vol. 55, pp. 209–216, 2015.

[18] Q. Tran-Dinh, I. Necoara, and M. Diehl, "Fast inexact decomposition algorithms for large-scale separable convex optimization," *Optimization*, vol. 65, no. 2, pp. 325–356, 2016.

[19] J.-H. Hours and C. N. Jones, "A parametric nonconvex decomposition algorithm for real-time and distributed NMPC," *IEEE Trans. on Autom. Control*, vol. 61, no. 2, pp. 287–302, 2016.

[20] B. Houska, J. Frasch, and M. Diehl, "An augmented Lagrangian based algorithm for distributed nonconvex optimization," *SIAM J. on Optimization*, vol. 26, no. 2, pp. 1101–1127, 2016.

[21] N. Chatzipanagiotis and M. M. Zavlanos, "On the convergence of a distributed augmented Lagrangian method for non-convex optimization," *IEEE Trans. on Autom. Control*, vol. 62, no. 9, pp. 4405–4420, 2017.

[22] H. Lakshmanan and D. P. De Farias, "Decentralized resource allocation in dynamic networks of agents," *SIAM J. on Optimization*, vol. 19, no. 2, pp. 911–940, 2008.

[23] A. Simonetto, T. Keviczky, and M. Johansson, "A regularized saddle-point algorithm for networked optimization with resource allocation constraints," in *IEEE Conf. on Dec. and Control (CDC)*, 2012, pp. 7476–7481.

[24] I. Necoara, "Random coordinate descent algorithms for multi-agent convex optimization over networks," *IEEE Trans. on Autom. Control*, vol. 58, no. 8, pp. 2001–2012, 2013.

[25] A. Cherukuri and J. Cortés, "Distributed generator coordination for initialization and anytime optimization in economic dispatch," *IEEE Trans. on Control of Network Systems*, vol. 2, no. 3, pp. 226–237, 2015.

[26] T.-H. Chang, A. Nedić, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Trans. on Autom. Control*, vol. 59, no. 6, pp. 1524–1538, 2014.

[27] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," *J. of Optim. Theory and Applications*, vol. 168, no. 1, pp. 172–197, 2016.

[28] 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.

[29] D. Mateos-Núñez and J. Cortés, "Distributed saddle-point subgradient algorithms with laplacian averaging," *IEEE Trans. on Autom. Control*, vol. 62, no. 6, pp. 2720–2735, 2017.

[30] I. Notarnicola, M. Franceschelli, and G. Notarstefano, "A duality-based approach for distributed min-max optimization with application to demand side management," in *IEEE Conf. on Dec. and Control (CDC)*, 2016, pp. 1877–1882.

[31] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.

[32] M. Bürger, G. Notarstefano, and F. Allgöwer, "A polyhedral approximation framework for convex and robust distributed optimization," *IEEE Trans. on Autom. Control*, vol. 59, no. 2, pp. 384–395, 2014.

[33] ——, "From non-cooperative to cooperative distributed MPC: A simplicial approximation perspective," in *European Control Conf. (ECC)*, 2013, pp. 2795–2800.

[34] M. Lorenzen, M. Bürger, G. Notarstefano, and F. Allgöwer, "A distributed solution to the adjustable robust economic dispatch problem," *IFAC Proceedings Volumes*, vol. 46, no. 27, pp. 75–80, 2013.

[35] D. P. Bertsekas, *Nonlinear programming*. Athena scientific, 1999.

[36] ——, *Constrained optimization and Lagrange multiplier methods*. Academic Press, 1982.

[37] ——, "Min common/max crossing duality: a geometric view of conjugacy in convex optimization," *Tech. Rep. LIDS-P-2796*, 2009.

[38] ——, *Convex Optimization Algorithms*. Athena Scientific, 2015.

[39] R. Zamora and A. K. Srivastava, "Controls for microgrids with storage: Review, challenges, and research needs," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 7, pp. 2009–2018, 2010.