

This is the post peer-review accepted manuscript of:

I. Notarnicola, M. Franceschelli and G. Notarstefano, "A duality-based approach for distributed min-max optimization with application to demand side management," 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, 2016, pp. 1877-1882.

The published version is available online at:

<https://doi.org/10.1109/CDC.2016.7798538>

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A duality-based approach for distributed min-max optimization with application to demand side management

Ivano Notarnicola¹, Mauro Franceschelli², Giuseppe Notarstefano¹

Abstract—In this paper we consider a distributed optimization scenario in which a set of processors aims at minimizing the maximum of a collection of “separable convex functions” subject to local constraints. This set-up is motivated by peak-demand minimization problems in smart grids. Here, the goal is to minimize the peak value over a finite horizon with: (i) the demand at each time instant being the sum of contributions from different devices, and (ii) the local states at different time instants being coupled through local dynamics. The min-max structure and the double coupling (through the devices and over the time horizon) makes this problem challenging in a distributed set-up (e.g., well-known distributed dual decomposition approaches cannot be applied). We propose a distributed algorithm based on the combination of duality methods and properties from min-max optimization. Specifically, we derive a series of equivalent problems by introducing ad-hoc slack variables and by going back and forth from primal and dual formulations. On the resulting problem we apply a dual subgradient method, which turns out to be a distributed algorithm. We prove the correctness of the proposed algorithm and show its effectiveness via numerical computations.

I. INTRODUCTION

The addition of processing, measurement, communication and control capability to the electric power grid is leading to smart grids, in which smart generators, accumulators and loads can cooperate to execute Demand Side Management (DSM) programs [1]. The goal is to reduce the hourly and daily variations and peaks of electric demand by optimizing generation, storage and consumption. A widely adopted objective in DSM programs is Peak-to-Average Ratio (PAR), defined as the ratio between peak-daily and average-daily power demands. PAR minimization gives raise to a min-max optimization problem if the average daily electric load is assumed not to be affected by the demand response strategy.

In [2] the authors propose a game-theoretic model for PAR minimization and provide a distributed energy-cost-based strategy for the users. A noncooperative-game approach is also proposed in [3], where optimal strategies are characterized and a distributed scheme is designed based on a proximal decomposition algorithm. A key difference of the

The research leading to these results 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) and from the Italian grant SIR “Scientific Independence of young Researchers”, project CoNetDomeSys, code RBSI14OF6H, funded by the Italian Ministry of Research and Education (MIUR).

¹Ivano Notarnicola and Giuseppe Notarstefano are with the Department of Engineering, Università del Salento, Via Monteroni, 73100 Lecce, Italy, ivan.notarnicola@unisalento.it.

²Mauro Franceschelli (corresponding author) is with the Department of Electrical and Electronic Engineering, University of Cagliari, Piazza D’Armi, 09123 Cagliari, Italy, mauro.franceschelli@diee.unica.it.

set-up in [2], [3], compared to the one proposed in our paper, is that in those works each agent needs to know the total load and tariffs in the power distribution system. Moreover, the agents do not cooperate to compute the strategy. In [4] a Model Predictive Control scheme is proposed to optimize micro-grid operations while satisfying a time-varying request and operation constraints using a mixed-integer linear model.

In this paper we propose a novel distributed optimization framework for min-max optimization problems commonly found in DSM problems. Differently from the references above, we consider a cooperative, distributed computation model in which the agents in the network do not have knowledge of aggregate quantities, communicate only with neighboring agents and perform local computations (with no central coordinator) to solve the optimization problem.

Duality is a widely used tool for distributed optimization algorithms as shown, e.g., in the tutorials [5], [6]. These standard approaches do not apply to the framework considered in this paper. In [7] a distributed consensus-based primal-dual algorithm is proposed to solve optimization problems with coupled global cost function and inequality constraints.

Min-max optimization is strictly related to saddle-point problems. In [8] the authors propose a subgradient method to generate approximate saddle-points. A min-max problem is also considered in [9] and a distributed algorithm based on a suitable penalty approach has been proposed. Another class of algorithms exploits the exchange of active constraints among the network nodes to solve constrained optimization problems which include min-max problems, [10], [11]. Although they work under asynchronous, directed communication they do not scale in set-ups as the one in this paper in which the terms of the max function are coupled. Very recently, in [12] the authors proposed a distributed projected subgradient method to solve constrained saddle-point problems with agreement constraints. Although our problem set-up fits in those considered in [12], our algorithmic approach and the analysis are different. In [13], [14] saddle point dynamics are used to design distributed algorithms for standard separable optimization problems.

The contribution of this paper is twofold. First, we propose a novel distributed optimization framework which is strongly motivated by peak power-demand minimization in DSM. The optimization problem has a min-max structure with local constraints at each node. Each term in the max function represents a daily cost (so that the maximum over a given horizon needs to be minimized), while the local constraints are due to the local dynamics and input bounds of the subsystems in the smart grid. The problem is challenging

when approached in a distributed way since it is *doubly coupled* (each term of the max function is coupled among the agents, while the local constraints impose a coupling between different “days” in the time-horizon).

Second, as main paper contribution, we propose a distributed algorithm to solve this class of min-max optimization problems. The algorithm has a very simple and clean structure in which a primal minimization and a dual ascent step are performed. The primal problem has a similar structure to the centralized one. Despite this simple structure, which resembles standard distributed dual methods, the algorithm is not a standard decomposition scheme and the derivation of the algorithm is non-obvious. Specifically, the algorithm is derived by heavily resorting to duality theory and properties of min-max optimization (or saddle-point) problems. In particular, a sequence of equivalent problems is derived in order to decompose the originally coupled problem into locally-coupled subproblems, and thus being able to design a distributed algorithm. An interesting feature of the algorithm is its expression in terms of dual variables of two different problems and of the original primal variables. Since we apply duality more than once and on different problems, this property, although apparently intuitive, was not obvious a priori. Another appealing feature of the algorithm is that every limit point of the primal sequence at each node is a (feasible) optimal solution of the original optimization problem (although this is only convex and not strictly convex). This property is obtained by the minimizing sequence of the local primal subproblems without resorting to averaging schemes, [15]. Finally, since each node only computes the decision variable of interest, our algorithm can solve both large-scale (many agents are present) and big-data (a large horizon is considered) problems.

The paper is structured as follows. In Section II we provide some useful preliminaries on optimization, duality theory and subgradient methods. In Section III we formalize our distributed min-max optimization set-up and present the main contribution of the paper, a novel, duality based distributed optimization method. In Section IV we characterize its convergence properties. Finally, in Section V we corroborate the theoretical results with a numerical example involving peak power minimization in a smart-grid scenario.

Due to space constraints all proofs are omitted in this paper and will be provided in a forthcoming document.

II. PRELIMINARIES

A. Optimization and Duality

Consider a constrained optimization problem, addressed as primal problem, having the form

$$\begin{aligned} \min_{z \in Z} f(z) \\ \text{subj. to } g(z) \preceq 0 \end{aligned} \quad (1)$$

where $Z \subseteq \mathbb{R}^N$ is a convex and compact set, $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is a convex function and each component $g_s : \mathbb{R}^N \rightarrow \mathbb{R}$, $s \in \{1, \dots, S\}$, of g is a convex function.

The following optimization problem

$$\begin{aligned} \max_{\mu} q(\mu) \\ \text{subj. to } \mu \succeq 0 \end{aligned} \quad (2)$$

is called the dual of problem (1), where $q : \mathbb{R}^S \rightarrow \mathbb{R}$ is obtained by minimizing with respect to $z \in Z$ the Lagrangian function $\mathcal{L}(z, \mu) := f(z) + \mu^\top g(z)$, i.e., $q(\mu) = \min_{z \in Z} \mathcal{L}(z, \mu)$. Problem (2) is well posed since the domain of q is convex and q is concave on its domain.

It can be shown that the following inequality holds

$$\inf_{z \in Z} \sup_{\mu \succeq 0} \mathcal{L}(z, \mu) \geq \sup_{\mu \succeq 0} \inf_{z \in Z} \mathcal{L}(z, \mu), \quad (3)$$

which is called weak duality. When in (3) the equality holds, then we say that strong duality holds and, thus, solving the primal problem (1) is equivalent to solving its dual formulation (2). In this case the right-hand-side problem in (3) is referred to as *saddle-point problem* of (1).

Definition 2.1: A pair (z^*, μ^*) is called an primal-dual optimal solution of problem (1) if $z^* \in Z$ and $\mu^* \succeq 0$, and (z^*, μ^*) is a saddle point of the Lagrangian, i.e.,

$$\mathcal{L}(z^*, \mu) \leq \mathcal{L}(z^*, \mu^*) \leq \mathcal{L}(z, \mu^*)$$

for all $z \in Z$ and $\mu \succeq 0$. \square

A more general min-max property can be stated. Let $Z \subseteq \mathbb{R}^N$ and $W \subseteq \mathbb{R}^S$ be nonempty convex sets. Let $\phi : Z \times W \rightarrow \mathbb{R}$, then the following inequality

$$\inf_{z \in Z} \sup_{w \in W} \phi(z, w) \geq \sup_{w \in W} \inf_{z \in Z} \phi(z, w)$$

holds true and is called the *max-min* inequality. When the equality holds, then we say that ϕ , Z and W satisfy the *strong max-min* property or the *saddle-point* property.

The following theorem gives a sufficient condition for the strong max-min property to hold.

Proposition 2.2 ([16, Propositions 4.3]): Let ϕ be such that (i) $\phi(\cdot, w) : Z \rightarrow \mathbb{R}$ is convex and closed for each $w \in W$, and (ii) $-\phi(z, \cdot) : W \rightarrow \mathbb{R}$ is convex and closed for each $z \in Z$. Assume further that W and Z are convex and compact sets. Then

$$\sup_{w \in W} \inf_{z \in Z} \phi(z, w) = \inf_{z \in Z} \sup_{w \in W} \phi(z, w)$$

and the set of saddle points is nonempty and compact. \square

B. Subgradient Method

Consider the following (constrained) optimization problem

$$\min_{z \in Z} f(z) \quad (4)$$

with $Z \subseteq \mathbb{R}^N$ a closed convex set and $f : \mathbb{R}^N \rightarrow \mathbb{R}$ convex. The (projected) subgradient method is the iterative algorithm

$$z(t+1) = P_Z \left(z(t) - \gamma(t) \tilde{\nabla} f(z(t)) \right) \quad (5)$$

where $t \in \mathbb{N}$ denotes the iteration index, $\gamma(t)$ is the step-size, $\tilde{\nabla} f(z(t))$ denotes a subgradient of f at $z(t)$, and $P_Z(\cdot)$ is the Euclidean projection onto Z .

Assumption 1: The step-size $\gamma(t) \geq 0$ satisfies the following diminishing condition

$$\lim_{t \rightarrow \infty} \gamma(t) = 0, \quad \sum_{t=1}^{\infty} \gamma(t) = \infty, \quad \sum_{t=1}^{\infty} \gamma(t)^2 < \infty. \quad \square$$

Proposition 2.3 ([17, Proposition 3.2.6]): Assume that the subgradients $\nabla f(z)$ are bounded for all $z \in Z$ and the set of optimal solutions is nonempty. Let the step-size $\gamma(t) \geq 0$ satisfy the diminishing condition in Assumption 1. Then the subgradient method in (5) applied to problem (4) converges in objective value and sequence $z(t)$ converges to an optimal solution. \square

III. PROBLEM SET-UP AND DISTRIBUTED OPTIMIZATION ALGORITHM

In this section we set-up the distributed min-max optimization problem and propose a distributed algorithm to solve it.

A. Distributed min-max optimization set-up

We consider a network of N processors which communicate according to a *connected, 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. That is, the edge (i, j) models the fact that node i and j exchange information. We denote by \mathcal{N}_i the set of *neighbors* of node i in the fixed graph \mathcal{G} , i.e., $\mathcal{N}_i := \{j \in \{1, \dots, N\} \mid (i, j) \in \mathcal{E}\}$.

Motivated by applications in Demand Side Management of Smart Grids, we introduce a min-max optimization problem to be solved by the network processors in a distributed way. Specifically, we associate to each processor i a decision vector $x^i = [x_1^i, \dots, x_S^i]^\top \in \mathbb{R}^S$, a constraint set $X_i \subseteq \mathbb{R}^S$ and local cost functions g_{is} , $s \in \{1, \dots, S\}$, and set-up the following optimization problem

$$\begin{aligned} \min_{x^1, \dots, x^N} \quad & \max_{s \in \{1, \dots, S\}} \sum_{i=1}^N g_{is}(x_s^i) \\ \text{subj. to} \quad & x^i \in X_i, \quad i \in \{1, \dots, N\} \end{aligned} \quad (6)$$

where for each $i \in \{1, \dots, N\}$ the set $X_i \subseteq \mathbb{R}^S$ is nonempty, convex and compact, and the functions $g_{is} : \mathbb{R} \rightarrow \mathbb{R}$, $s \in \{1, \dots, S\}$, are convex.

Note that we use the superscript $i \in \{1, \dots, N\}$ to indicate that a vector $x^i \in \mathbb{R}^S$ belongs to node i , while we use the subscript to identify a vector component, i.e., x_s^i , $s \in \{1, \dots, S\}$, is the s -th component of x^i .

Using a standard approach for min-max problems, we introduce an auxiliary variable P to write the so called epigraph representation of problem (6), given by

$$\begin{aligned} \min_{x^1, \dots, x^N, P} \quad & P \\ \text{subj. to} \quad & x^i \in X_i, \quad i \in \{1, \dots, N\} \\ & \sum_{i=1}^N g_{is}(x_s^i) \leq P, \quad s \in \{1, \dots, S\}. \end{aligned} \quad (7)$$

Notice that, this problem is convex, but not strictly convex. This means that it is not guaranteed to have a unique solution. This impacts on dual approaches when trying to recover a primal optimal solution, see e.g., [15] and references therein.

B. Algorithm description

Next, we introduce our distributed optimization algorithm. Informally, the algorithm consists of a two-step procedure. First, each node $i \in \{1, \dots, N\}$ stores a set of variables $((x^i, \rho^i), \mu^i)$ obtained as the primal-dual optimal solution pair of a local min-max optimization problem with a structure similar to the centralized problem. The coupling with the other nodes in the original formulation is replaced by a term depending on neighboring variables λ^{ij} , $j \in \mathcal{N}_i$. These variables are updated in the second step according to a suitable linear law weighting the difference of neighboring μ^i . Nodes use a diminishing step-size denoted by $\gamma(t)$ and can initialize the variables λ^{ij} , $j \in \mathcal{N}_i$ to zero. In the next table we formally state our Primal Min-Max Dual Subgradient distributed algorithm from the perspective of node i .

Distributed Algorithm Primal Min-Max Dual Subgradient

Processor states: (x^i, ρ^i) , μ^i and λ^{ij} for $j \in \mathcal{N}_i$

Evolution:

Gather $\lambda^{ji}(t)$ from $j \in \mathcal{N}_i$

Compute $((x^i(t+1), \rho^i(t+1)), \mu^i(t+1))$ as a primal-dual optimal solution pair of

$$\begin{aligned} \min_{x^i, \rho^i} \quad & \rho^i \\ \text{subj. to} \quad & x^i \in X_i \\ & g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \leq \rho^i, \quad (8) \\ & s \in \{1, \dots, S\} \end{aligned}$$

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

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

$$\lambda^{ij}(t+1) = \lambda^{ij}(t) - \gamma(t)(\mu^i(t+1) - \mu^j(t+1)) \quad (9)$$

The structure of the algorithm and the meaning of the updates will be clear in the constructive analysis carried out in the next section. At this point we want to point out that although problem (8) has the same min-max structure of problem (7), ρ^i is not a copy of the centralized cost P , but rather a local contribution to that cost. That is, as we will see, the total cost P will be the sum of the ρ^i s.

IV. ALGORITHM ANALYSIS

The analysis of the proposed Primal Min-Max Dual Subgradient distributed algorithm is constructive and heavily relies on duality theory tools.

We start by deriving the equivalent dual problem of (7) which is formally stated in the next lemma.

Lemma 4.1: The optimization problem

$$\begin{aligned} \max_{\mu \in \mathbb{R}^S} \quad & \sum_{i=1}^N q_i(\mu) \\ \text{subj. to} \quad & \mathbf{1}^\top \mu = 1, \quad \mu \succeq 0 \end{aligned} \quad (10)$$

where $\mathbf{1} := [1, \dots, 1]^\top \in \mathbb{R}^S$ and

$$q_i(\mu) := \min_{x^i \in X_i} \sum_{s=1}^S \mu_s g_{is}(x_s^i), \quad i \in \{1, \dots, N\}, \quad (11)$$

is the dual of problem (7) and strong duality holds. \square

In order to make problem (10) amenable for a distributed solution, we can rewrite it in an equivalent form. To this end, we introduce copies of the common optimization variable μ and coherence constraints having the sparsity of the connected graph \mathcal{G} , obtaining

$$\begin{aligned} & \max_{\mu^1, \dots, \mu^N} \sum_{i=1}^N q_i(\mu^i) \\ & \text{subj. to } \mathbf{1}^\top \mu^i = 1, \mu^i \geq 0, i \in \{1, \dots, N\} \\ & \quad \mu^i = \mu^j, \text{ for all } (i, j) \in \mathcal{E}. \end{aligned} \quad (12)$$

Notice that we have also duplicated the simplex constraint so that it becomes a local constraint for each node.

To solve this problem we can use a dual decomposition approach by designing a dual subgradient algorithm. This can be done since the constraints are convex and the cost function concave. A dual subgradient algorithm applied to problem (12) would immediately result into a distributed algorithm if functions q_i were available in a closed form.

Intuition suggests that deriving the dual of a dual problem would somehow bring back to a primal formulation. However, we want to stress that:

- (i) problem (12) is dualized rather than problem (10),
- (ii) different constraints are dualized, namely the coherence constraints rather than the simplex ones.

We start deriving the dual subgradient algorithm by dualizing only the coherence constraints. Thus, we write the partial Lagrangian

$$\begin{aligned} \mathcal{L}_2(\mu^1, \dots, \mu^N, \{\lambda^{ij}\}_{(i,j) \in \mathcal{E}}) \\ = \sum_{i=1}^N \left(q_i(\mu^i) + \sum_{j \in \mathcal{N}_i} \lambda^{ij \top} (\mu^i - \mu^j) \right) \end{aligned} \quad (13)$$

where $\lambda^{ij} \in \mathbb{R}^S$ for all $(i, j) \in \mathcal{E}$ are Lagrange multipliers associated to the constraints $\mu^i - \mu^j = 0$. By exploiting the undirected nature and the connectivity of communication graph \mathcal{G} , after some algebraic manipulations, we get

$$\begin{aligned} \mathcal{L}_2(\mu^1, \dots, \mu^N, \{\lambda^{ij}\}_{(i,j) \in \mathcal{E}}) \\ = \sum_{i=1}^N \left(q_i(\mu^i) + \mu^{i \top} \sum_{j \in \mathcal{N}_i} (\lambda^{ij} - \lambda^{ji}) \right), \end{aligned} \quad (14)$$

which is separable with respect to $\mu^i, i \in \{1, \dots, N\}$.

The dual of problem (12) is thus

$$\min_{\{\lambda^{ij}\}_{(i,j) \in \mathcal{E}}} \eta(\{\lambda^{ij}\}_{(i,j) \in \mathcal{E}}) = \sum_{i=1}^N \eta_i(\{\lambda^{ij}, \lambda^{ji}\}_{j \in \mathcal{N}_i}), \quad (15)$$

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

$$\eta_i(\{\lambda^{ij}, \lambda^{ji}\}_{j \in \mathcal{N}_i}) = \max_{\mathbf{1}^\top \mu^i = 1, \mu^i \geq 0} q_i(\mu^i) + \mu^{i \top} \sum_{j \in \mathcal{N}_i} (\lambda^{ij} - \lambda^{ji}).$$

In order to apply a subgradient method to problem (15), we recall, [18, Section 6.1], that

$$\frac{\tilde{\partial} \eta(\{\lambda^{ij}\}_{(i,j) \in \mathcal{E}})}{\partial \lambda^{ij}} = \mu^{i*} - \mu^{j*}, \quad (16)$$

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

$$\mu^{k*} \in \operatorname{argmax}_{\mathbf{1}^\top \mu^k = 1, \mu^k \geq 0} \left(q_k(\mu^k) + \mu^{k \top} \sum_{h \in \mathcal{N}_k} (\lambda^{kh} - \lambda^{hk}) \right),$$

for $k = i, j$. The dual subgradient algorithm for problem (12) can be summarized as follows, for each node $i \in \{1, \dots, N\}$:

- (S1) receive $\lambda^{ji}(t), j \in \mathcal{N}_i$ and compute a subgradient $\mu^i(t+1)$ by solving

$$\max_{\mu^i} q_i(\mu^i) + \mu^{i \top} \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t)) \quad (17)$$

- subj. to $\mathbf{1}^\top \mu^i = 1, \mu^i \geq 0$.

- (S2) exchange with neighbors the updated $\mu^j(t+1), j \in \mathcal{N}_i$, and update $\lambda^{ij}, j \in \mathcal{N}_i$, via

$$\lambda^{ij}(t+1) = \lambda^{ij}(t) - \gamma(t)(\mu^i(t+1) - \mu^j(t+1)).$$

where $\gamma(t)$ denotes the step-size.

It is worth noting that in (17) the value of $\lambda^{ij}(t)$ and $\lambda^{ji}(t)$, for $j \in \mathcal{N}_i$, is fixed as highlighted by the index t . Moreover, we want to stress, once again, that the algorithm is *not* implementable as it is written, since functions q_i are not available in closed form. On this regard, here we slightly abuse notation since in (S1)-(S2) we use $\mu^i(t)$ as in the Primal Min-Max Dual Subgradient algorithm, but we have not proven the equivalence yet. Since we will prove it in the next lemmas we preferred not to overweight the notation.

Lemma 4.2: The dual subgradient updates (S1)-(S2), with step-size $\gamma(t)$ satisfying Assumption 1, generate sequences $\{\lambda^{ij}(t)\}, (i, j) \in \mathcal{E}$ that converge in objective value to $\eta^* = q^* = P^*$, optimal costs of (15), (10) and (7), respectively. \square

We can explicitly rephrase update (17) by plugging in the definition of q_i , given in (11), thus obtaining the following max-min optimization problem

$$\max_{\mathbf{1}^\top \mu^i = 1, \mu^i \geq 0} \left(\min_{x^i \in X_i} \sum_{s=1}^S \mu_s^i \left(g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \right) \right). \quad (18)$$

Notice that this is a local problem at each node i once the value for $\lambda^{ij}(t)$ and $\lambda^{ji}(t)$ for all $j \in \mathcal{N}_i$ are given.

Lemma 4.3: Max-min optimization problem (18) is the saddle point problem associated to problem (8). Moreover, a primal-dual optimal solution pair of (8), call it $\{(x^i(t+1), \rho^i(t+1)), \mu^i(t+1)\}$, exists and $(x^i(t+1), \mu^i(t+1))$ is a solution of (18).

Proof: We give a constructive proof which clarifies how problem (8) is derived from (18). Define

$$\phi(x^i, \mu^i) := \sum_{s=1}^S \mu_s^i \left(g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \right) \quad (19)$$

and note that (i) $\phi(\cdot, \mu^i)$ is closed and convex for all $\mu^i \succeq 0$ and (ii) $\phi(x^i, \cdot)$ is closed and concave (linear over the compact $\mathbf{1}^\top \mu^i = 1, \mu^i \succeq 0$), for all $x^i \in \mathbb{R}^S$. Thus we can invoke the saddle point Proposition 2.2 which allows us to switch the max and min operators, and write

$$\begin{aligned} & \mathbf{1}^\top \max_{\mu^i=1, \mu^i \succeq 0} \left(\min_{x^i \in X_i} \sum_{s=1}^S \mu_s^i \left(g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \right) \right) \\ &= \min_{x^i \in X_i} \left(\mathbf{1}^\top \max_{\mu^i=1, \mu^i \succeq 0} \sum_{s=1}^S \mu_s^i \left(g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \right) \right). \end{aligned} \quad (20)$$

Since the inner maximization problem depends nonlinearly on x^i (which is itself an optimization variable), the solution cannot be obtained without considering the optimization also on x^i . We overcome this issue by substituting the inner maximization problem with its equivalent dual. Notice that the inner problem is a linear program when x^i are kept fixed, and thus strong duality can be exploited. Introducing a scalar multiplier ρ^i associated to the simplex constraint, we have

$$\max_{\mu^i} \sum_{s=1}^S \mu_s^i \left(g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \right) \quad (21)$$

$$\text{subj. to } \mathbf{1}^\top \mu^i = 1, \mu^i \succeq 0$$

is equivalent to its dual

$$\min_{\rho^i} \rho^i \quad (22)$$

$$\text{subj. to } g_{is}(x_s^i) + \sum_{j \in \mathcal{N}_i} (\lambda^{ij}(t) - \lambda^{ji}(t))_s \leq \rho^i, s \in \{1, \dots, S\}$$

where the S inequality constraints follow from the minimization of the partial Lagrangian of (21) with respect to $\mu^i \succeq 0$. Plugging formulation (22) in place of the inner maximization in (20), we can write a *joint* minimization, i.e., minimize simultaneously with respect to x^i and ρ^i , which leads to (8).

To prove the second part, notice that problem (8) is convex. Then, the problem satisfies the Slater's constraint qualification and, thus, strong duality holds. Therefore, a primal-dual optimal solution pair $(x^i(t+1), \rho^i(t+1), \mu^i(t+1))$ exists and from the previous arguments the proof follows. \square

We point out that the previous lemma shows that performing minimization in (8) turns out to be equivalent to performing step (S1).

We are now ready to state the main result of the paper, namely the convergence of the Primal Min-Max Dual Subgradient distributed algorithm.

Theorem 4.4: Let $\{(x^i(t), \rho^i(t))\}, i \in \{1, \dots, N\}$, be the sequence generated by the Primal Min-Max Dual Subgradient distributed algorithm, with $\gamma(t)$ satisfying Assumption 1. Then, the sequence $\{\sum_{i=1}^N \rho^i(t)\}$ converges to the optimal cost P^* of (6) and every limit point of the sequence $\{x^i(t)\}, i \in \{1, \dots, N\}$, is an optimal (feasible) solution of (6). \square

Remark 4.5: From condition (20) it can be shown that each $\rho^i(t)$ is equal to $\eta_i(\{\lambda^{ij}, \lambda^{ji}\}_{j \in \mathcal{N}_i})$ for all $t \geq 0$. Since the optimal cost of (15) is equal to the optimal

primal cost P^* , then we have that $\lim_{t \rightarrow \infty} \sum_i \rho^i(t) = \lim_{t \rightarrow \infty} \sum_i \eta_i(t) = P^*$. \square

V. NUMERICAL SIMULATIONS

In this section we propose a numerical example in which we apply the proposed method to a network of Thermostatically Controlled Loads (TCLs) (such as air conditioners, heat pumps, electric water heaters), [19].

The dynamical model of the i -th device is given by

$$\dot{T}^i(\tau) = -\alpha (T^i(\tau) - T_{out}^i(\tau)) + Qx^i(\tau), \quad (23)$$

where $T_i(\tau) \geq 0$ is the temperature, $\alpha > 0$ is a parameter depending on geometric and thermal characteristics, $T_{out}^i(\tau)$ is the air temperature outside the device, $x^i(\tau) \in [0, 1]$ is the control input, and $Q > 0$ is a scaling factor.

We consider a discretized version of the system with constant input over the sampling interval $\Delta\tau$, i.e., $x^i(\tau) = x_s^i$ for $\tau \in [s\Delta\tau, (s+1)\Delta\tau)$, and sampled state T_s^i ,

$$T_{s+1}^i = T_s^i e^{-\alpha\Delta\tau} + (1 - e^{-\alpha\Delta\tau}) \left(\frac{Q}{\alpha} x_s^i - T_{out,s}^i \right). \quad (24)$$

We assume that the power consumption $g_{is}(x_s^i)$ of the i -th device in the s -th slot $[s\Delta\tau, (s+1)\Delta\tau]$ is directly proportional to x_s^i . For the sake of simplicity we consider $g_{is}(x_s^i) = x_s^i$ in the numerical example proposed in this section. Thus, optimization problem (6) for this scenario is

$$\begin{aligned} & \min_{x^1, \dots, x^N} \max_{s \in \{1, \dots, S\}} \sum_{i=1}^N x_s^i \\ & \text{subj. to } x^i \in X_i, \quad i \in \{1, \dots, N\} \end{aligned} \quad (25)$$

where $X_i := \{x^i \in \mathbb{R}^S \mid A_i x^i \preceq b_i \text{ and } x^i \in [0, 1]^S\}$, with A_i and b_i obtained by enforcing the dynamics constraints (24) and temperature constraints $T_s^i \in [T_{min}, T_{max}]$.

In the proposed numerical example we consider $N = 15$ agents communicating according to an undirected connected Erdős-Rényi random graph \mathcal{G} with parameter 0.2. We consider a horizon of $S = 50$. Finally, a diminishing step-size sequence $\gamma(t) = (\frac{1}{t})^{0.8}$ at iteration t , which satisfies Assumption 1, is used.

In Figure 1 we show the evolution at each algorithm iteration t of the local objective functions $\rho^i(t), i \in \{1, \dots, N\}$, (solid lines) which converge to stationary values. We also plot their sum $\sum_{i=1}^N \rho^i(t)$ (dotted line) that asymptotically converges to the centralized optimal cost P^* of problem (25) (see Remark 4.5).

In Figure 2 it is shown the profile of an optimal consumption of the devices, i.e., $\sum_{i=1}^N x_s^{i*}$, over the horizon $s = 1, \dots, S$. It can be seen that the proposed method effectively shaves off the peak power demand. In the same figure it also shown an optimal consumption strategy x^{i*} that each single device locally computes.

Finally, in Figure 3 it is shown the convergence rate of the distributed algorithm, i.e., the difference between the centralized optimal cost P^* and the sum of the local costs $\sum_{i=1}^N \rho^i(t)$, in logarithmic scale. It can be seen that the proposed algorithm converges to the optimal cost with sublinear rate $O(1/\sqrt{t})$ as expected for a subgradient method.

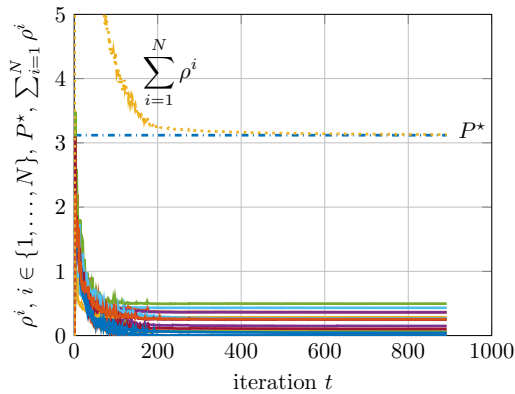


Fig. 1. Evolution of ρ^i , $i \in \{1, \dots, N\}$, (solid lines), their sum $\sum_{i=1}^N \rho^i$ (dotted line), and (centralized) optimal cost P^* (dash-dotted line).

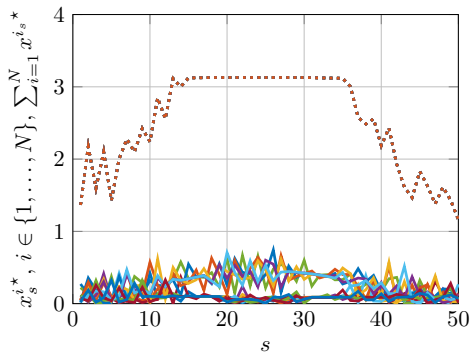


Fig. 2. Profile of optimal solutions x_s^{i*} , $i \in \{1, \dots, N\}$, (solid lines), and $\sum_{i=1}^N x_s^{i*}$ (dotted line) on the optimization horizon $\{1, \dots, S\}$.

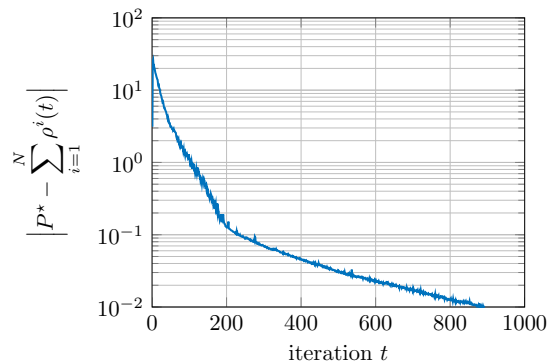


Fig. 3. Evolution of the cost error, in logarithmic scale.

VI. CONCLUSIONS

In this paper we have introduced a novel distributed min-max optimization framework motivated by peak minimization problems in Demand Side Management. Standard distributed optimization algorithms cannot be applied to this problem set-up due to a highly nontrivial coupling in the objective function and in the constraints. We proposed a distributed algorithm based on the combination of duality methods and properties from min-max optimization. We proved the correctness of the proposed algorithm and corroborated the theoretical results with a numerical example.

REFERENCES

- [1] M. Alizadeh, X. Li, Z. Wang, A. Scaglione, and R. Melton, "Demand-side management in the smart grid: Information processing for the power switch," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 55–67, 2012.
- [2] A.-H. Mohsenian-Rad, V. W. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, 2010.
- [3] I. Atzeni, L. G. Ordóñez, G. Scutari, D. P. Palomar, and J. R. Fonollosa, "Demand-side management via distributed energy generation and storage optimization," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 866–876, 2013.
- [4] A. Parisio, E. Rikos, and L. Glielmo, "A model predictive control approach to microgrid operation optimization," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 5, pp. 1813–1827, 2014.
- [5] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.
- [6] B. Yang and M. Johansson, "Distributed optimization and games: A tutorial overview," in *Networked Control Systems*. Springer, 2010, pp. 109–148.
- [7] T.-H. Chang, A. Nedić, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1524–1538, 2014.
- [8] A. Nedić and A. Ozdaglar, "Subgradient methods for saddle-point problems," *Journal of optimization theory and applications*, vol. 142, no. 1, pp. 205–228, 2009.
- [9] K. Srivastava, A. Nedić, and D. Stipanović, "Distributed min-max optimization in networks," in *IEEE 17th International Conference on Digital Signal Processing (DSP)*, 2011, pp. 1–8.
- [10] G. Notarstefano and F. Bullo, "Distributed abstract optimization via constraints consensus: Theory and applications," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2247–2261, October 2011.
- [11] M. Bürger, G. Notarstefano, and F. Allgöwer, "A polyhedral approximation framework for convex and robust distributed optimization," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 384–395, 2014.
- [12] D. Mateos-Núñez and J. Cortés, "Distributed subgradient methods for saddle-point problems," in *IEEE 54th Conference on Decision and Control (CDC)*, 2015.
- [13] A. Simonetto, T. Keviczky, and M. Johansson, "A regularized saddle-point algorithm for networked optimization with resource allocation constraints," in *IEEE 51st Conference on Decision and Control (CDC)*, 2012, pp. 7476–7481.
- [14] A. Koppel, F. Y. Jakubiec, and A. Ribeiro, "Regret bounds of a distributed saddle point algorithm," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 2969–2973.
- [15] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.
- [16] D. P. Bertsekas, "Min common/max crossing duality: A geometric view of conjugacy in convex optimization," *Lab. for Information and Decision Systems, MIT, Tech. Rep. Report LIDS-P-2796*, 2009.
- [17] —, *Convex Optimization Algorithms*. Athena Scientific, 2015.
- [18] —, *Nonlinear programming*. Athena scientific, 1999.
- [19] M. Alizadeh, A. Scaglione, A. Applebaum, G. Kesidis, and K. Levitt, "Reduced-order load models for large populations of flexible appliances," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1758–1774, 2015.