Distributed optimization for structured programs and its application to energy management in a building district[☆]

Fabio Belluschi[a], Alessandro Falsone[a,*], Daniele Ioli[a], Kostas Margellos[b], Simone Garatti[a], Maria Prandini[a]

[a] *Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy*
[b] *Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom*

Abstract

This paper deals with structured multi-agent optimization problems that involve coupled local and global decision variables. We propose an iterative distributed algorithm that explicitly accounts for this structure, and requires the agents to communicate only their tentative solutions for the global variables throughout iterations. Our approach extends to structured multi-agent optimization a proximal-based distributed methodology that has recently appeared in the literature. Privacy of local information is preserved and communication effort is reduced with respect to alternative distributed solutions where local and global optimization variables are grouped together and treated as a single decision vector. Multi-agent optimization problems with the considered structural properties appear in various contexts. In this paper, we apply our approach to energy management in a district where multiple buildings can communicate over a possibly time-varying network and aim at optimizing the use of shared and local resources. We illustrate the efficacy of the resulting distributed energy management algorithm by means of a detailed simulation study on a cooling

[☆]This work was supported by the European Commission, H2020, under the project UnCoVerCPS, grant number 643921, and by EPSRC UK under the grant EP/P03277X/1.

[*]Corresponding author

Email addresses: fabio.belluschi@mail.polimi.it (Fabio Belluschi), alessandro.falsone@polimi.it (Alessandro Falsone), daniele.ioli@polimi.it (Daniele Ioli), kostas.margellos@eng.ox.ac.uk (Kostas Margellos), simone.garatti@polimi.it (Simone Garatti), maria.prandini@polimi.it (Maria Prandini)

problem.

Keywords: Distributed optimization, proximal minimization, energy management, building control.

1. Introduction

This paper addresses optimization problems where multiple agents are connected through a possibly time-varying network and aim at optimizing their local performance indices subject to heterogeneous constraints. A key feature of the considered set-up is that it involves both local and global optimization variables, and the latter variables generate a coupling in the decisions of the agents. More specifically, we consider m agents that need to agree on a global decision vector $x \in \mathbb{R}^n$, while also deciding their own individual decision vector $u_i \in \mathbb{R}^{n_i}$, $i = 1, \dots, m$, so as to minimize the sum of their local cost functions $f_i(x, u_i) : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}$, $i = 1, \dots, m$, while satisfying the local constraints $(x, u_i) \in V_i$, $i = 1, \dots, m$, where $V_i \subseteq \mathbb{R}^{n+n_i}$ is the constraint set of agent i. The tuple (u_i, f_i, V_i) constitutes private information that agent i is not willing to share with the other agents. As for the communication structure, the network connecting the agents is modeled as a directed graph (N, E_k), where the node set $N = \{1, \dots, m\}$ represents the agents and the edge set $E_k \subseteq N \times N$ represents the communication links that are active at time step k. More specifically, $(j, i) \in E_k$ if agent j can communicate with agent i at k.

Our goal is devising an iterative algorithm over such a time-varying communication network so that the agents jointly solve the following constrained optimization problem

$$\mathcal{P} : \min_{x \in \mathbb{R}^n, \{u_i \in \mathbb{R}^{n_i}\}_{i=1}^m} \sum_{i=1}^{m} f_i(x, u_i) \tag{1a}$$

$$\text{subject to: } \quad (x, u_i) \in V_i, \text{ for all } i = 1, \dots, m, \tag{1b}$$

while not sharing with each other their private information represented by the tuple (u_i, f_i, V_i) for each agent i, $i = 1, \dots, m$.

Note that the presence of a structured time-varying network and of information privacy constraints prevents us from solving $\mathcal{P}$ in a centralized fashion. Moreover, even in the case when all this information were made available, a centralized solution to $\mathcal{P}$ would be computationally intense and not scalable for problems with a high number of agents.

In case when a star communication graph is present, with one of the agent acting as a central authority/aggregator and collecting information from all agents, a possibility to preserve privacy and obtain a scalable solution would be to adopt a decentralized[1] paradigm, where the central authority/aggregator collects from all agents their estimates on the global decision vector x, performs some aggregation/computation and broadcasts some update to all agents. Under this regime, each agent has then to solve a problem with fewer decision variables and constraints compared to the original one, [1].

Several decentralized algorithms have been proposed in the literature; among those the alternating direction method of multipliers (ADMM) has attracted particular attention, [1, 2]. To render $\mathcal{P}$ amenable to ADMM each agent should create a copy of the global decision vector, thus giving rise to a separable objective function and constraint sets in (1a) and (1b), respectively. However, we would need to introduce the so called consistency constraints to guarantee that all these copies should be the same. Applying then ADMM, which involves running a primal-dual scheme (as opposed to the primal algorithm presented in the sequel), the dual variables associated with the consistency constraints would need to be exchanged with all agents via the central authority. This would entail exchanging nm variables, where n is the number of variables in each copy and m is the number of agents (one consistency constraint per agent).

[1] In certain research domains the term distributed is used instead. Motivated by the convention adopted by the majority of the control systems community we use the term *decentralized* when a central authority is present, while we use the term *distributed* to indicate that a central authority is absent and communication is restricted only among agents considered as neighbors according to an underlying communication protocol.

Here, in view of reducing the amount of information exchange and respecting the privacy constraints, we aim at a protocol that does not involve a central authority, restricting communication only among neighboring agents. Given the time-varying nature of the communication graph (N, E_k) and, hence, of the notion of neighbors, the proposed communication protocol should be time-varying. Therefore, we resort to distributed optimization algorithms that allow for time-varying communication in contrast to distributed implementations of ADMM (see [2] for consensus developments). In this realm, the gradient/subgradient algorithms of [3, 4, 5, 6], or the proximal minimization based algorithm of [7, 8], could be adopted. The drawback, however, of all these distributed methods is that agents need to share their local decisions u_i which constitute private information. Alternatively, the primal-dual scheme proposed in [9, 10] based on dual decomposition could be adopted. However, such a choice would entail introducing consistency constraints as in ADMM and as a result an excess of communication involving exchanging nm variables.

In this paper, we propose a proximal based distributed algorithm that builds on [8], as the latter imposes fewer assumptions and does not require differentiability of the objective function or computation of subgradients as in other approaches in the literature. At the same time, it is a primal based scheme that overcomes the need for a communication exchange that increases with the number of agents as in [9, 10], and exploits the structure of the problem requiring to exchange information related only to the global decision vector. The value taken by the global optimization vector will in turn affect the optimization of the local ones, which, however, will be performed locally to each agent, without sharing with any other its individual objective function and constraint set (privacy preserving algorithm), and without the need of enlarging its optimization vector so as to include the local decision vectors of the other agents. The proposed solution is then scalable in the number of agents since the size of the global optimization vector over which the agents need to reach consensus is fixed, and the computation of the local ones is made individually by each agent, without the need of providing any related information to the others. Note that

Information exchange \ Algorithm	[7, 8, 3, 4, 5, 6]	[9, 10]	Algorithm 1
Dual variables	–	nm	–
Local variables u_i per agent	n_i	–	–
Global variables x	n	–	n

Table 1: Information exchange in the proposed algorithm and the most relevant approaches in the literature.

this nested optimization scheme was suggested in [8, Remark 1] but with reference to the special (significantly easier) case where local and global decisions are coupled only through the performance index and the feasibility region for the local decision variables is independent of the value taken by the global ones. It is actually the coupling via the constraint that makes the problem difficult to solve.

Our algorithm extends the one in [7, 8] to the considered structured framework, and has the advantage with respect to [9, 10] of reducing the exchange of information, and with respect to [3, 4, 5, 6, 7, 8] of preserving privacy of local decisions, besides achieving significant communication savings. The theoretical guarantees provided in [7, 8] on the convergence to an optimizer of the centralized counterpart of the problem are shown to still hold in our structured setting. The proposed approach is particularly convenient when there is a high number of local optimization variables and only a few global ones. Table 1 classifies the main features of the proposed algorithm (Algorithm 1) with respect to the most relevant distributed optimization approaches over time-varying networks, [3, 4, 5, 6, 7, 8, 10, 9]. All these algorithms perform local computations of the same complexity; we thus report their difference in terms of the amount and nature of information that needs to be exchanged.

Optimization problems for multi-agent systems exhibiting the considered structural properties can be found in various application domains. Here, we focus on energy management in buildings connected over a network sharing common resources, as it naturally fits the class of structured programs with each building/agent having several decision variables related to temperature set-points that are local and should not be shared with other agents, while having some global variables related to the usage of shared resources. Optimal energy management in buildings has attracted significant attention worldwide, since recent studies [11] have shown that more than 30% of the total electricity consumption in Europe and in the United States is related to buildings and half of that to climate control. Constructing algorithms for optimal energy management in building networks will allow demand modulation through intelligent control and coordination of certain appliances, or demand deferrability by appropriate use of the storage devices. To achieve this, not only conventional energy management methods need to be revisited, but also conceptually different control and coordination schemes have to be designed.

Towards this direction, optimization based algorithms have been already successfully applied to the problem of energy management in buildings, due to their ability to handle the multi-objective nature of the problem (e.g., minimize energy costs, maximize building utility), while taking physical and/or technological constraints (e.g., storage limits, comfort constraints) into account. Studies in this direction include, but are not limited to, [12, 13, 14, 15, 16, 17]. Recently, in [18, 19], a compositional perspective is adopted, allowing for smart-grid control that involves multiple buildings, chiller plants, storage devices, co-generation plants, etc., interacting with each other, whereas in [20] an energy-hub perspective is adopted, investigating the problem of managing a collection of buildings in a cooperative manner. However, the network encoding the interaction among the different modules is considered to be time-invariant, and the problem is solved in a centralized fashion. In [21], a decentralized scheme for scheduling smart appliances in a residential district with a shared energy storage system is described, with an aggregator playing the role of the central entity coordinating the buildings demand and managing the exchanges with the grid and the shared energy storage system. In [22], a hierarchical scheme implementing a decentralized heuristic solution is proposed, which accounts also for the on-off switching of devices. In [23] a decentralized control methodology is applied to a home energy management problem. In all cases the underlying network topology is assumed to be time-invariant.

In this paper, we deal in particular with the problem of cooling of a building district, where buildings connected over a (possibly) time-varying network, are equipped with individual chiller plants and are connected to a cooling network through which they can exchange cooling energy. The aim is minimizing the district electrical energy costs over a given time horizon, while guaranteeing comfort conditions for the building occupants.

Note that building energy management applications have been studied recently in [24] and [25] according to a multi-agent perspective. In contrast with these references, we allow for a more general formulation where the cost function is only required to be convex, and propose an algorithm that is completely distributed since it does not require any central authority. As for the application, we provide a more accurate modeling for the building, since we explicitly account for its thermal inertia, and for the chiller unit, since we consider its efficiency as a function of the cooling energy request.

Our contributions can then be summarized as follows:

- we propose a scalable distributed algorithm which extends [7, 8] to structured multi-agent optimization problems, preserving its optimality guarantees, without requiring agents to disclose their local decision variables and limiting the amount of exchanged information to that related to the global decision variables, which have a fixed size, independent on the number of agents;
- we show how the proposed algorithm can be applied to resource sharing in energy management of a building district, and perform a detailed simulation-based study.

The rest of the paper unfolds as follows: In Section 2 we formulate the structured multi-agent optimization problem and introduce our distributed solution, including a detailed convergence analysis. In Section 3 we describe the energy management problem over a building district. Section 4 provides a simulation based study, whereas Section 5 concludes the paper.

2. Multi-agent optimization with local decisions

In Section 2.1, we provide a distributed iterative procedure to solve the optimization problem $\mathcal{P}$ in (1) and then analyze its convergence properties in Section 2.2.

At every iteration of the proposed distributed algorithm, each agent i solves an appropriate local optimization problem and then exchanges information with other agents only regarding the tentatively obtained value for the common decision vector x. In this way, one can account for information privacy, because agents are not required to share their own cost function f_i, constraint set V_i, and decision vector u_i, $i = 1, \ldots, m$. Moreover, even though all the necessary information could be exchanged, solving $\mathcal{P}$ in a centralized fashion may be computationally intensive and our distributed algorithm is also a means to alleviate this issue.

Under certain structural and communication assumptions, the proposed algorithm converges, and agents reach consensus to a common value for the global decision vector x that, together with the converged values for the local decision vectors u_i, $i = 1, \ldots, m$, forms an optimal solution of $\mathcal{P}$ (note that $\mathcal{P}$ does not necessarily admit a unique solution).

2.1. Distributed algorithm

The pseudo-code of the proposed distributed procedure is given in Algorithm 1. In the remainder of this subsection we provide some explanations of the algorithm steps.

Initially, each agent i, $i = 1, \ldots, m$, starts with some tentative values $u_i(0)$ and $x_i(0)$ for its local decision vector and the global decision vector, respectively. The latter constitutes an estimate of agent i (this justifies the subscript i in x_i) of what the value of the global decision vector might be. Those

Algorithm 1 Distributed algorithm for structured optimization

1: **Initialization**

2: $k = 0$

3: $(x_i(0), u_i(0)) \in V_i$, for all $i = 1, \ldots, m$

4: **For** $i = 1, \ldots, m$ **repeat until convergence**

5: $\bar{x}_i(k) = \sum_{j=1}^{m} a_j^i(k) x_j(k)$

6: $(x_i(k+1), u_i(k+1)) \in \arg\min_{(x_i, u_i) \in V_i} f_i(x_i, u_i) + \frac{1}{2c(k)} \|\bar{x}_i(k) - x_i\|^2$

7: $k \leftarrow k + 1$

tentative values are chosen arbitrarily from the set of feasible solutions, i.e., $(x_i(0), u_i(0)) \in V_i$ (step 3). One sensible choice for $(x_i(0), u_i(0))$ is to set it such that $(x_i(0), u_i(0)) \in \arg\min_{(x_i, u_i) \in V_i,} f_i(x_i, u_i)$, as it guarantees local constraint satisfaction for each agent. However, the convergence analysis presented in the sequel does not depend on the initialization of the algorithm. At iteration k, each agent i constructs a weighted average $\bar{x}_i(k)$ of the solutions $x_j(k)$, $j = 1, \ldots, m$ communicated by its neighboring agents and its own one (step 5). Coefficient $a_j^i(k) \geq 0$, indicates how agent i weights the solution received by agent j at iteration k. If $a_j^i(k) = 0$, agent j does not use information related to agent i at iteration k (this is necessarily the case if $(j, i) \notin E_k$). The coefficients $a_j^i(k)$ are chosen by the user but they are required to satisfy some assumptions specified in Section 2.2. Agent i solves then a local minimization problem, seeking the optimal solution pair (x_i, u_i) within V_i that minimizes a performance criterion, which is defined as a linear combination of the local objective function $f_i(x_i, u_i)$ and a quadratic term[2], penalizing the difference from $\bar{x}_i(k)$ (step 6). The relative importance of these two terms is dictated by $c(k) > 0$, which act as the step-size parameter of a gradient-like method. Similarly to the $a_j^i(k)$ coefficients, also the sequence $\{c(k)\}_{k \geq 0}$ is a design parameter but it is subject to restrictions described in Section 2.2. Since multiple minimizers may exist, we assume that at every iteration the same deterministic tie-break rule (as e.g.

[2]Throughout the paper, $\| \cdot \|$ denotes Euclidean norm.

that implemented by a deterministic numerical solver) is used.

Algorithm 1 is closely related to the distributed methodology that has been recently proposed in [7, 8]. However, in Algorithm 1, neighboring agents need to exchange at every iteration their tentative estimates for the value of the global decision vector only, while, as discussed in the introduction, the distributed algorithm in [7, 8] requires to exchange both the global and the local decision vectors. When the dimension of the local decision vector is high compared to the global one, this would unnecessarily increase the amount of information that needs to be exchanged. Algorithm 1 alleviates this issue by exploiting the particular structure of $\mathcal{P}$, where the objective functions and the constraint sets are coupled only by means of x.

2.2. Algorithm analysis

In this section we study the convergence properties of Algorithm 1. To this end, we shall first introduce some assumptions on optimization problem $\mathcal{P}$.

Assumption 1. *For all $i = 1, \dots, m$, the function $f_i(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}$ is jointly convex with respect to its arguments. Moreover, for all $i = 1, \dots, m$, $f_i(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}$ is jointly Lipschitz continuous with respect to its arguments.*

Note that under Assumption 1, and due to the presence of the quadratic penalty term, the objective function in the optimization problem at step 6 of Algorithm 1 is strictly convex with respect to x_i. Therefore, a unique solution for x_i is admitted; this is not the case for u_i.

Assumption 2. *For all $i = 1, \dots, m$, the set $V_i \subseteq \mathbb{R}^{n+n_i}$ is compact and convex. Moreover, $\bigcap_{i=1}^m V_i$ has non-empty interior.*

For all $i = 1, \dots, m$, for any $x \in \mathbb{R}^n$, consider the set

$$U_i(x) = \big\{u_i \in \mathbb{R}^{n_i} : \ (x, u_i) \in V_i\big\}. \tag{2}$$

Moreover, for all $i = 1, \ldots, m$, consider the projection of V_i on the x domain, i.e.,

$$X_i = \big\{x \in \mathbb{R}^n : \ \exists u_i \in \mathbb{R}^{n_i} \text{ such that } (x, u_i) \in V_i\big\}. \tag{3}$$

A direct consequence of the first part of Assumption 2 is that, for all $i = 1, \ldots, m$, X_i and $U_i(x)$ for any $x \in X_i$ are all compact and convex. By the second part of Assumption 2, we also have that $\bigcap_{i=1}^m X_i$, and hence also X_i, $i = 1, \ldots, m$, has a non-empty interior. Moreover, $U_i(x)$ is non-empty for any $x \in X_i$, $i = 1, \ldots, m$. The fact that V_i is both convex and compact implies that the set-valued mapping $U_i(\cdot)$ is continuous on X_i, see [26]. In the following assumption we further require that $U_i(\cdot)$ is Lipschitz continuous.

Assumption 3. *For all $i = 1, \ldots, m$, the set-valued mapping $U_i(\cdot) : X_i \rightrightarrows \mathbb{R}^{n_i}$ is Lipschitz continuous, i.e., there exists $L_i \in \mathbb{R}$, $L_i > 0$, such that*

$$d_H(U_i(x), U_i(x')) \leq L_i \|x - x'\|, \ \textit{for all } x, x' \in X_i, \tag{4}$$

where

$$d_H(U_i(x), U_i(x')) = \sup_{u_i \in \mathbb{R}^{n_i}} \left| \min_{v_i \in U_i(x)} \|u_i - v_i\| - \min_{v_i' \in U_i(x')} \|u_i - v_i'\| \right|, \tag{5}$$

denotes the Pompeiu-Hausdorff distance (see p. 272 in [27]) between the sets $U_i(x)$ and $U_i(x')$.

Besides Assumptions 1–3, which need to be verified from problem to problem, we also impose the following restrictions on the choices of the penalty parameter sequence $\{c(k)\}_{k \geq 0}$ and on the communication weights $a_j^i(k)$, $i, j = 1, \ldots, m$ and $k \geq 0$.

Assumption 4. *$\{c(k)\}_{k \geq 0}$ is a non-increasing sequence with $c(k) > 0$ for all k. Moreover, $\sum_{k=0}^{\infty} c(k) = \infty$ and $\sum_{k=0}^{\infty} c(k)^2 < \infty$.*

A direct consequence of the last part of Assumption 4 is that $\lim_{k \to \infty} c(k) = 0$, meaning that the relative importance of the quadratic penalty term over the local cost function $f_i(\cdot, \cdot)$ is progressively increased to force consensus of the different $x_i(k)$'s. A possible choice for $\{c(k)\}_{k\geq 0}$ that satisfies Assumption 4 is to select it from the class of generalized harmonic series, e.g., $c(k) = \alpha/(k+1)$ for some $\alpha > 0$. Given this choice, a small value of α will drive the agents to quickly reach consensus on x and then seek optimality, whereas a high value of α will let the agents minimize their local cost function first and then adjust their decisions to agree on a common x. To get an insight on the role of $c(k)$ we refer the reader to an example in Appendix B.

Assumption 5. *There exists $\eta \in (0,1)$ such that for all $i, j \in \{1, \ldots, m\}$ and all $k \geq 0$, $a^i_j(k) \geq 0$ and $a^i_j(k) > 0$ implies that $a^i_j(k) \geq \eta$. Moreover, for all $k \geq 0$,*

1. $\sum_{j=1}^m a^i_j(k) = 1$ *for all* $i = 1, \ldots, m$,
2. $\sum_{i=1}^m a^i_j(k) = 1$ *for all* $j = 1, \ldots, m$.

The interpretation of having a uniform lower bound η, independent of k, for the (non-zero) coefficients $a^i_j(k)$ in Assumption 5 is that it ensures that each agent is weighting information received by other agents at a non-diminishing rate (as η is strictly greater than zero) as iterations progress, [4]. Moreover, points 1 and 2 ensure that this weighting is a convex combination of the other agent estimates and the local estimate, where a non-zero weight is assigned to this latter since $a^i_i(k) \geq \eta$.

Let $E_\infty = \{(j,i) : \ a^i_j(k) > 0 \text{ for infinitely many } k\}$ denote the set of edges (j,i) such that agent j uses information provided by agent i infinitely often. The following connectivity and communication assumption is eventually enforced.

Assumption 6. *The graph (N, E_∞) is strongly connected, i.e., for any two nodes there exists a path of directed edges that connects them. Moreover, there exists $\bar{k} \geq 1$ such that for every $(j,i) \in E_\infty$, agent i uses information from a neighboring agent j at least once every consecutive $\bar{k}$ iterations.*

Assumption 6 guarantees that any pair of agents communicates at least indirectly infinitely often, and the intercommunication interval is bounded. For further details the reader is referred to [8, 3]. It should be emphasized that allowing for iteration-varying topology is typically referred to as time-varying communication in the distributed optimization literature [4]. However, at each iteration of the algorithm a finite horizon optimization program has often to be solved, as in the energy management application presented in the sequel. Therefore, the absence of communication at any given iteration of the algorithm does not necessarily imply absence of communication at a given step of the finite horizon problem, but rather at the time of local computation.

Problem $\mathcal{P}$ can be equivalently written as

$$\min_{x \in \bigcap_{i=1}^m X_i} \sum_{i=1}^m g_i(x), \tag{6}$$

where, for all $i = 1, \ldots, m$, and for any $x \in \mathbb{R}^n$,

$$g_i(x) = \min_{u_i \in U_i(x)} f_i(x, u_i). \tag{7}$$

Note that for all $x \in X_i$ the minimum in (7) exists due to the Weierstrass' theorem (Proposition A.8, p. 625 in [1]), since $U_i(x)$ is compact by Assumption 2 and $f_i(\cdot, \cdot)$ is continuous due to Assumption 1. We then have the following auxiliary lemmas, which are crucial for the proof of Theorem 2. Their proofs are provided in Appendix A.

Lemma 1. *Under Assumptions 1 and 2, it holds that $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is convex on X_i, for all $i = 1, \ldots, m$.*

Lemma 2. *Under Assumptions 1, 2, and 3, it holds that $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous on X_i, for all $i = 1, \ldots, m$.*

Consider now Algorithm 1, and, according to (7), re-write step 6 as

$$x_i(k+1) = \arg\min_{x_i \in X_i} g_i(x_i) + \frac{1}{2c(k)} \|\bar{x}_i(k) - x_i\|^2. \tag{8}$$

Note that, since $g_i(\cdot)$ is convex on X_i (Lemma 1) and the quadratic penalty term in (8) is strictly convex, $x_i(k+1)$ is univocally defined.

Le us now recall Theorem 1 in [8], which is crucial for proving convergence of Algorithm 1.

Theorem 1 ([8]). *Consider problem* (6)*. If X_i are convex and compact, $g_i(\cdot)$ are convex on $\mathbb{R}^n$, and $\bigcap_{i=1}^m X_i$ has non-empty interior, then, under Assumptions 4–6, there exists a minimizer $x^\star$ of* (6) *such that the sequence $\{x_i(k)\}_{k\geq 0}$ satisfies $\lim_{k\to\infty} \|x_i(k) - x^\star\| = 0$, for all $i = 1, \ldots, m$.*

What is missing for a direct application of Theorem 1 in our context is the convexity of functions $g_i(\cdot)$ over the whole $\mathbb{R}^n$. Yet, building on Theorem 1, we prove in Theorem 2 below that Algorithm 1 converges to a minimizer of $\mathcal{P}$. More precisely, we are able to show that there exists a minimizing global decision vector $x^\star$ of $\mathcal{P}$ such that the values $\{x_i(k)\}_{k\geq 0}$ generated by Algorithm 1 converge to $x^\star$, for all $i = 1, \ldots, m$ (i.e. agents reach consensus on the value of the global decision vector). Moreover, though the local decision vector $\{u_i(k)\}_{k\geq 0}$, $i = 1, \ldots, m$, generated by Algorithm 1 may exhibit an oscillatory behavior, all their limit points will form together with $x^\star$ a minimizer of $\mathcal{P}$.

Theorem 2. *Let $\{x_i(k)\}_{k\geq 0}$, $\{u_i(k)\}_{k\geq 0}$, $i = 1, \ldots, m$, be the sequences of estimates generated by Algorithm 1. Under Assumptions 1-6:*

1. *there exists a minimizing vector $x^\star$ of $\mathcal{P}$, such that $\lim_{k\to\infty} \|x_i(k) - x^\star\| = 0$, for all $i = 1, \ldots, m$;*
2. *any limit point $(u_1^\star, \ldots, u_m^\star)$ of the sequence $\{(u_1(k), \ldots, u_m(k))\}_{k\geq 0}$, is such that $(x^\star, u_1^\star, \ldots, u_m^\star)$ is a minimizer of $\mathcal{P}$.*

Proof. Consider Algorithm 1 with step 6 rewritten as in (8). Thanks to Assumptions 1-3 and thanks to Lemmas 1 and 2, it holds that

p.1 X_i is convex and compact, for all $i = 1, \ldots, m$;

p.2 $\bigcap_{i=1}^m X_i$ has non-empty interior;

p.3 $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is convex on X_i, for all $i = 1, \ldots, m$;

p.4 $g_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous on X_i, for all $i = 1, \ldots, m$.

Under Assumptions 4-6 and given properties p.1 and p.2, if $g_i(\cdot)$ in (7) were convex on $\mathbb{R}^n$, one could invoke Theorem 1 to directly conclude that there exists a minimizer $x^\star$ of (6), such that $\lim_{k\to\infty} \|x_i(k) - x^\star\| = 0$, for all $i = 1, \dots, m$. Here, we do not have convexity of $g_i(\cdot)$ over the whole $\mathbb{R}^n$, but only the characterization of $g_i(\cdot)$ through conditions p.3 and p.4. On the other hand, p.3 and p.4 constitute a weaker set of conditions on $g_i(\cdot)$ (as a matter of fact, convexity over the whole $\mathbb{R}^n$ together with the compactness condition in p.1 implies p.3 and p.4), and, as it follows from the discussion below Assumption 3 in [8], it is easy to see that Assumptions 4-6 and properties p.1-p.4 suffice to draw the same conclusion that $\lim_{k\to\infty} \|x_i(k) - x^\star\| = 0$, for all $i = 1, \dots, m$, where $x^\star$ is a minimizer of (6). The fact that a minimizer $x^\star$ of (6) is also the x-component of the minimizer of $\mathcal{P}$ (by the equivalence between problem $\mathcal{P}$ and (6)), concludes then the proof of the first part of the theorem.

The second part follows along lines akin to the proof of point (b) of Theorem 1.17 in [28]. Specifically, let $(u_1^\star, \dots, u_m^\star)$ be any limit point of the sequence $\{(u_1(k), \dots, u_m(k))\}_{k\geq 0}$, which exists thanks to the compactness Assumption 2. Thanks to Assumption 2 it also holds that $(x^\star, u_1^\star, \dots, u_m^\star)$ is feasible for $\mathcal{P}$. Given the definition of g_i and that of $u_i(k)$, recalling that $\lim_{k\to\infty} \|x_i(k) - x^\star\| = 0$ for all $i = 1, \dots, m$, and thanks to the continuity of $g_i(\cdot)$ as assured by Lemma 2, for any given $\epsilon > 0$ it holds that

$$\sum_{i=1}^{m} f_i(x_i(k), u_i(k)) = \sum_{i=1}^{m} g_i(x_i(k)) \leq \sum_{i=1}^{m} g_i(x^\star) + \epsilon$$

for k large enough. This in turn implies that $\sum_{i=1}^{m} f_i(x^\star, u_i^\star) \leq \sum_{i=1}^{m} g_i(x^\star) + \epsilon$. Being ϵ arbitrary, it follows that $\sum_{i=1}^{m} f_i(x^\star, u_i^\star) \leq \sum_{i=1}^{m} g_i(x^\star)$, which, given the equivalence between problem $\mathcal{P}$ and (6), shows that $(x^\star, u_1^\star, \dots, u_m^\star)$ is a minimizer of $\mathcal{P}$. □

Remark 1 (satisfaction of the algorithm assumptions)**.** *Assumptions 4 and 5 imply that Algorithm 1 is synchronous, and agents need to agree prior to the execution of the algorithm on $\{c(k)\}_{k\geq 0}$ and the weight coefficients $\{a_j^i(k)\}_{k\geq 0}$, $i, j = 1, \dots, m$. For every iteration k, these weights should form a doubly stochastic matrix. A distributed methodology to construct doubly stochastic matrices can be found in [29, 3]. Assumption 6 is standard in distributed optimiza-*

tion algorithms over networks, and is satisfied for a wide class of time-varying network structures. In particular, periodic absence of communication links, as in the case study of Section 4, falls in the proposed framework.

As for the other assumptions, even though it is relatively straightforward to verify Assumptions 1 and 2, note that the compactness requirement of Assumption 2 is not restrictive from a practical point of view, as numerical computation is typically performed over compact domains (enclosing the region where decision variables take values from). Moreover, most practical problems involve decisions/actuation that is subject to limitations, thus ensuring compactness. However, it is in general difficult to verify Assumption 3. This is due to the fact that existence of a uniform Lipschitz constant, such that the set-valued continuity condition (4) *is satisfied, is hard to verify even numerically. In [30], the authors determine a Lipschitz constant for Assumption 3 to hold, for the case where the set-valued function* $U_i(x)$ *in* (2) *admits a representation as a product of a Lipschitz continuous single-valued function and a convex, compact and non-empty set (see Lemma 3.5 and Remark 2.7 therein). This is the case if the constraint sets* V_i*,* $i = 1, \dots, m$*, are polytopic. This opens the road for approximation procedures for problems where Assumption 3 is hard to verify. To this end, a piece-wise affine approximation of general convex constraint sets could be constructed, thus replacing the original problem with one that has polytopic constraints sets, for which Assumption 3 is satisfied [30].*

3. Energy management of a building district

In this section, we describe the cooling problem of a building district. Each building can set the temperatures of its thermally controlled zones within some appropriate range, can operate on its own chiller unit, and can exchange energy with a cooling network shared among the other buildings in the district, so as to satisfy its cooling load and minimize electric energy costs. Given that the shared resource has a limited capacity, some coordination is needed among buildings and this is realized via a (possibly) time-varying communication network so as to model temporary failures.

In this set-up, global decision variables are the energy exchanges of all buildings with the cooling network, and local decision variables are the zone temperature set-points of each building, which affect its cooling load and have to be chosen compatibly with the actuation capabilities of the chiller unit of the building. The resulting constraint on the local decision variables finally depends on the global decision variables since the cooling load has to match the sum of the cooling energy produced by the chiller and that drawn from the cooling network. Clearly, the electrical energy cost depends on both the energy exchanges with the cooling network and the zone temperature set-points.

3.1. Modeling of the components

We next present the models of a building, a chiller plant, and a cooling network, which constitute the basic components of the considered district network.

We focus on energy management over a finite time horizon, divided into n_t time slots, each of them having duration $\Delta \in \mathbb{R}$. Models describe the energy contribution of each component per time slot t, $t = 1, \dots, n_t$, and are taken from [19], where a compositional modeling framework for energy management of a district network is presented.

Building

We adopt the convex formulation proposed in [31] to model each individual building in the network. The adopted building model was validated in [19] according to the ANSI-ASHRAE (American Society for Heating Refrigerating and Air-conditioning Engineers) 140 standard.

Temperature set-points are control input variables and the actual building temperatures are assumed to track the imposed profiles. This entails the presence of a lower level control system able to effectively track the set-points. Constraints are enforced on the maximum cooling energy request and thus indirectly on the admissible temperatures and temperatures variation rate, so as to make it a reasonable assumption.

We consider a building composed of n_z zones and denote by $\widetilde{T}_z(t) \in \mathbb{R}$ the temperature of zone z, $z = 1, \ldots, n_z$, at the end of time slot t, $t = 1, \ldots, n_t$. Then, we can collect all control inputs in vector $\widetilde{T} = [\widetilde{T}(1) \cdots \widetilde{T}(n_t)]^\top \in \mathbb{R}^{n_z n_t}$, where we set $\widetilde{T}(t) = [\widetilde{T}_1(t) \cdots \widetilde{T}_{n_z}(t)]^\top \in \mathbb{R}^{n_z}$.

Let $E_{B,z}(t) \in \mathbb{R}$ denote the cooling energy request of building zone z during time slot t, in order to track a given zone temperature profile, with $t = 1, \ldots, n_t$ and $z = 1, \ldots, n_z$. $E_{B,z}(t)$ constitutes of four energy contributions, namely

$$E_{B,z}(t) = E_{\text{walls},z}(t) + E_{\text{people},z}(t) + E_{\text{internal},z}(t) + E_{\text{inertia},z}(t), \tag{9}$$

where $E_{\text{walls},z}(t) \in \mathbb{R}$ is the amount of thermal energy exchanged between zone z and its adjacent walls over time slot t, $E_{\text{people},z}(t) \in \mathbb{R}$ and $E_{\text{internal},z}(t) \in \mathbb{R}$ are the thermal energy produced by people and by other internal heat sources in zone z, respectively, and $E_{\text{inertia},z}(t) \in \mathbb{R}$ is the energy contribution due to the thermal inertia of zone z, over time slot t. The energy request of the building over the time slot t is given by

$$E_B(t) = \sum_{z=1}^{n_z} E_{B,z}(t),$$

and $\mathbf{E_B} = [E_B(1) \cdots E_B(n_t)]^\top$ describes the cooling energy requested by the building to track the temperature set-points of every zone over the time horizon $[1, n_t]$. In [19, Section 2.1] it is shown that the following expression holds for $\mathbf{E_B}$:

$$\mathbf{E_B} = AT(0) + B(d)\widetilde{T} + C(d) + Dd, \tag{10}$$

where $T(0)$ represents the building thermal state at time 0 and vector d collects all the disturbances affecting the system, i.e., the outside ambient temperature, the incoming shortwave and longwave solar radiation and people occupancy. Matrices A, D, $B(d)$, and $C(d)$ have appropriate dimensions and the last two depend on the disturbance vector d. Also, the building thermal state $T(n_t)$ at the end of the time horizon is given by a similar expression but with different matrices, i.e.,

$$T(n_t) = \tilde{A}T(0) + \tilde{B}(d)\widetilde{T} + \tilde{C}(d) + \tilde{D}d.$$

Chiller plant

A chiller plant converts electric energy into cooling energy. The cooling energy is then transferred to the building via, e.g., the chilled water circuit. The electrical energy $E_{\text{chiller},e}(t)$ needed to produce a certain amount $E_{\text{chiller},c}(t)$ of cooling energy during time slot t can be obtained as a bi-quadratic convex approximation of the Ng-Gordon model, [32]:

$$E_{\text{chiller},e}(t) = c_2 E^4_{\text{chiller},c}(t) + c_1 E^2_{\text{chiller},c}(t) + c_0, \tag{11}$$

where the parameters c_0, c_1, c_2 are determined using weighted least squares to best fit the most relevant points, i.e, those that correspond to zero energy request and to the maximum value of the Coefficient Of Performance (COP), which is the ratio between $E_{\text{chiller},c}(t)$ and $E_{\text{chiller},e}(t)$. Derivations are reported in [19, Section 2.2].

Cooling network

Since the cooling network has a high thermal inertia, it acts as a thermal storage, whose energy content can be described as a first-order dynamical system, with the energy exchange (drawn or inserted) as input and the thermal energy stored as state:

$$E_{\text{stored}}(t+1) = aE_{\text{stored}}(t) - \sum_{i=1}^{m} e^i_s(t), \tag{12}$$

where $E_{\text{stored}}(t) \in \mathbb{R}$ is the amount of cooling energy stored, see [19, Section 2.3]. In view of the multi-building problem considered in the next section we assume that the cooling network is shared among m buildings, and denote by $e^i_s(t) \in \mathbb{R}$ the cooling energy exchanged ($e^i_s(t) > 0$ if the cooling network is discharged, and $e^i_s(t) < 0$ if it is charged), with building i in time slot t. The coefficient $a \in (0, 1)$ is introduced to model energy losses.

3.2. Building district problem formulation

Consider a district of m buildings, each of them equipped with a different chiller plant, that share a common cooling network. To this end, append to all quantities introduced in the previous section the superscript i, to denote that they correspond to building i, $i = 1, \ldots, m$, e.g., $E^i_{\text{chiller},e}(t)$ denotes the cooling energy of the chiller at building i at time slot t, $\widetilde{T}^i$ denotes the vector of zone temperatures at building i, etc. For each i and t, the electric energy request of building i over the time slot t is given by the chiller electric energy request $E^i_{\text{chiller},e}(t)$. Our objective is to minimize the total electric energy cost for the m building network, across a horizon of n_t steps. To achieve this, for each building i we will schedule the zone temperature set-points $\widetilde{T}^i(t)$ and the energy exchange $e^i_s(t)$ with the cooling network. Therefore, we seek to solve the following minimization problem:

$$\min_{\substack{\left\{\left\{\widetilde{T}^i(t)\in\mathbb{R}^{n_z}, e^i_s(t)\in\mathbb{R}\right\}_{t=1}^{n_t}\right\}_{i=1}^{m}, \\ \left\{T^i(0)\in\mathbb{R}^{n_w}\right\}_{i=1}^{m}, E_{\text{stored}}(1)\in\mathbb{R}}} \sum_{i=1}^{m}\sum_{t=1}^{n_t} \psi^i(t) E^i_{\text{chiller},e}(t), \tag{13}$$

where $\psi^i(t) \in \mathbb{R}$ is the electric energy price for building i over the time slot t and $E^i_{\text{chiller},e}(t)$ is its electric energy request (computed according to (11)) within the same time slot.

This minimization is subject to the following constraints, that must hold for each time slot t and every building i:

- Energy balance equation: the chiller cooling energy request $E^i_{\text{chiller},c}(t)$ is given by

  $$E^i_{\text{chiller},c}(t) = E^i_B(t) - e^i_s(t), \tag{14}$$

  where $E^i_B(t)$ is the cooling energy requested by the building in the time slot t and is one of the component of vector $\mathbf{E^i_B}$ shown in (10), whereas $e^i_s(t)$ is the energy exchange between building i and the cooling network in the same time slot.

- Electric energy limits: the electric energy drawn from the grid is limited to $E^i_{\max} \in \mathbb{R}$, as an effect of the chiller unit size and maximum capability, thus giving rise to

  $$0 \leq E_{\text{chiller},e}(t) \leq E^i_{\max}. \tag{15}$$

- Cooling energy limits: zone z cooling energy request $E^i_{B,z}(t)$, is non-negative, i.e.,

$$E^i_{B,z}(t) \geq 0 \qquad \forall z = 1, \ldots, n_z. \tag{16}$$

- Comfort constraints: the zone temperature set-point is kept within certain comfort limits, i.e.,

$$\widetilde{T}^i(t) \in [\widetilde{T}^i_{\min}(t),\ \widetilde{T}^i_{\max}(t)], \tag{17}$$

  where $\widetilde{T}^i_{\min}(t) \in \mathbb{R}^{n_z}$, $\widetilde{T}^i_{\max}(t) \in \mathbb{R}^{n_z}$ denote the minimum and maximum comfort temperatures.

- Cooling network energy limits: the amount of cooling energy stored should be non-negative and within the energy cooling network capacity limit $E_{s,\max} \in \mathbb{R}$, i.e.,

$$E_{\text{stored}}(t) \in [0,\ E_{s,\max}]. \tag{18}$$

- Cooling network energy exchange limits: the energy exchanged with the cooling network is subject to

$$e^i_s(t) \in [-e^i_{s,\max},\ e^i_{s,\max}], \tag{19}$$

  where $e^i_{s,\max} \in \mathbb{R}$ denotes the maximum value of energy that the building can exchange with the cooling network per time slot.

- Final value constraints: the zone temperature and the building thermal state at the beginning and at the end of the planning horizon should be equal, i.e.,

$$\widetilde{T}^i(n_t) = \widetilde{T}^i(0) \ \wedge \ T^i(n_t) = T^i(0). \tag{20}$$

  To ensure that the cooling network does not get empty at the end of the horizon, we impose the constraint

$$E_{\text{stored}}(n_t) \geq E_{\text{stored}}(1). \tag{21}$$

Constraint (21) is of particular importance in case of a receding horizon implementation of the proposed scheme.

Note that the quantity $E^i_{\text{chiller},e}(t)$ in (13) is a function of the decision variables $\{\{\widetilde{T}^i(t), e^i_s(t)\}_{t=1}^{n_t}\}_{i=1}^m$, $\{T^i(0)\}_{i=1}^m$ (see (14) and (10)) and $E_{\text{stored}}(1)$ affects the admissible range for $e^i_s(t)$, $t = 1, \dots n_t$, $i = 1, \dots, m$, through (18) and (12), so that here we optimize it as well.

If we define now vectors u_i, $i = 1, \dots, m$, and x as follows:

$$u_i = \left[\widetilde{T}^i(1), \dots, \widetilde{T}^i(n_t), T^i(0)\right]^\top \in \mathbb{R}^{n_t n_z + n_w}, \tag{22a}$$

$$x = \left[\bar{e}^1_s, \dots, \bar{e}^m_s, E_{\text{stored}}(1)\right]^\top \in \mathbb{R}^{m n_t + 1}, \tag{22b}$$

where $\bar{e}^i_s = \left[e^i_s(1), \dots, e^i_s(n_t)\right]^\top \in \mathbb{R}^{n_t}$, then, u_i, $i = 1, \dots, m$, can be thus thought of as a local decision vector related to the comfort and actuation constraints of each chiller plant, that can be enforced locally, whereas x can be treated as a global decision vector which is related to the energy exchange of the building district with the common cooling network. Given (22a) and (22b), the energy management in (13)-(21) is an instance of problem (1).

Remark 2 (alternative closed-loop strategy)**.** *Note that the proposed energy management solution consists of an open-loop strategy that is pre-computed offline, where the zone temperature set-points and the energy exchange with the cooling network for the whole one-day reference time horizon are set based on some nominal profile for the disturbances. Alternatively, one could adopt a model predictive control approach where the optimal value for the zone temperature set-points and the energy exchange with the cooling network is determined online, by solving problem* (13) *on some finite-length time window, applying only the values corresponding to the current time instant, shifting the time window one step ahead in time, recomputing the optimal value for the decision variables on the shifted time window, and so on (receding horizon strategy). The advantage is that one can exploit state measurements and the possibly updated profile of the disturbances, thus getting a closed-loop strategy that is better tailored to the actual disturbance realizations. Computations can still be run in a distributed* *way through Algorithm 1. Constraints are however imposed on the time required for computing the optimal solution, which calls for further investigations on its convergence rate.*

3.3. Satisfaction of algorithm assumptions

The energy management problem of Section 3.2 is a convex minimization program. Assumption 2 is satisfied, as an effect of the physical and technological constraints imposed in (14)-(21). Even if this were not the case, all numerical calculations are performed on compact domains, hence satisfaction of Assumption 2 is not an issue. Due to convexity and compactness, it can be also easily verified that the objective function is Lipschitz continuous with respect to all decision variables, thus satisfying Assumption 1.

Assumptions 4 and 5 can be imposed by defining appropriately $\{c(\cdot)\}_{k\geq 0}$ and the weight coefficients $\{a_j^i(k)\}_{k\geq 0}$, $i, j = 1, \dots, m$ in Algorithm 1. Assumption 6 is satisfied in the case of periodic absence of communication links, as in the case study of Section 4, which falls in the proposed framework. Verifying Assumption 3 is instead generally difficult. However, applying Algorithm 1 to the case study of Section 4 we verified numerically that the assertions of Theorem 2 are valid (by comparing the achieved results with the optimal solution of the centralized problem), even though we were not able to formally verify satisfaction of Assumption 3.

4. Simulation results

Consider a network of $m = 3$, identical, three-storey buildings, as schematically illustrated in Figure 1. Each building is divided into $n_z = 3$ thermal zones (one per floor) and is equipped with its own chiller, namely, building 1 has a medium-size chiller, building 2 a small one, and building 3 a large one for which the COP curves as a function of the cooling energy request $E_{\text{chiller},c}$ are shown in Figure 2. Parameters values of the bi-quadratic approximations (11) can be found in [19, Section 2.2].

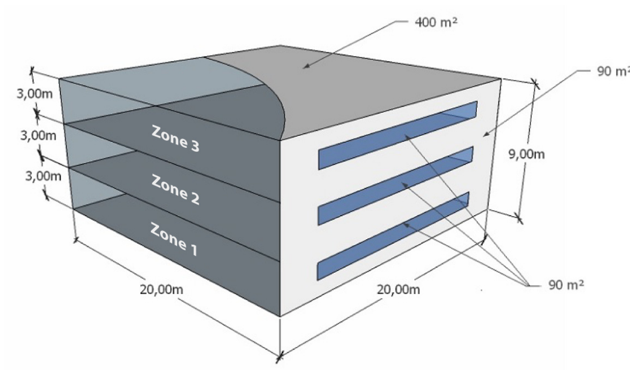

Figure 1: Structure of each building.

We considered a time horizon of 24 hours discretized in $n_t = 144$ time slots of $\Delta = 10$min each. The external disturbances affecting the buildings are reported in Figure 3. The three buildings are supposed to be subject to the same disturbance profiles, and the occupancy shall be intended per building and equally partitioned among the zones. The period in which the occupancy is greater than zero is referred to as "occupancy period" and it is within the "working hours" range 7AM to 6PM. In all buildings, temperature constraints are set to $\widetilde{T}^i_{\min} = 20°$C and $\widetilde{T}^i_{\max} = 24°$C during working hours and to $\widetilde{T}^i_{\min} = 16°$C and $\widetilde{T}^i_{\max} = 30°$C otherwise. Figure 4 represents the profile of the energy price, which is assumed to be identical for all buildings, during the 24 hours time horizon.

We assessed the performance of the algorithm for two different choices of the weights $\{a^i_j(k)\}_{k\geq 0}$, $i, j = 1, \ldots, m$, defining the communication protocol over the same bi-directional communication graph where all buildings are connected together. In the first communication protocol, buildings 1 and 3 exchange information only with building 2 but not with each other and the communication scheme is kept fixed across iterations with link weights equal to 1/3. In the second one, at each iteration k, only two buildings communicate and weights of active links are set equal to 1/2. The order in which the links are activated within the period is: $(1, 2)$, $(2, 3)$, and $(1, 3)$. We applied Algorithm 1 with

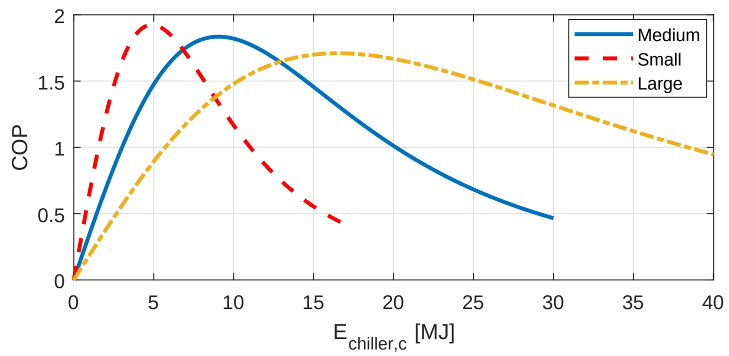

Figure 2: Chiller COP: maximum cooling energies are 18, 30 and 40 MJ.

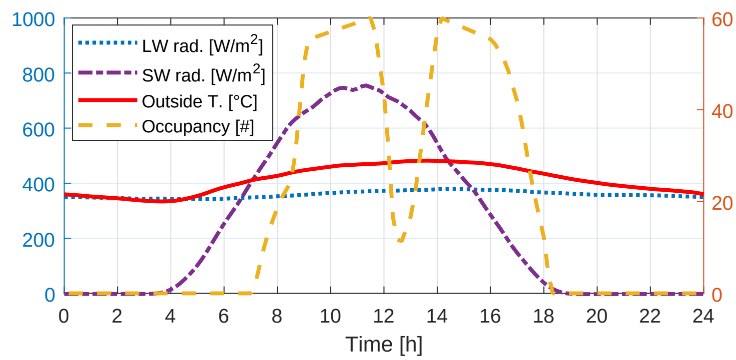

Figure 3: Disturbance profiles.

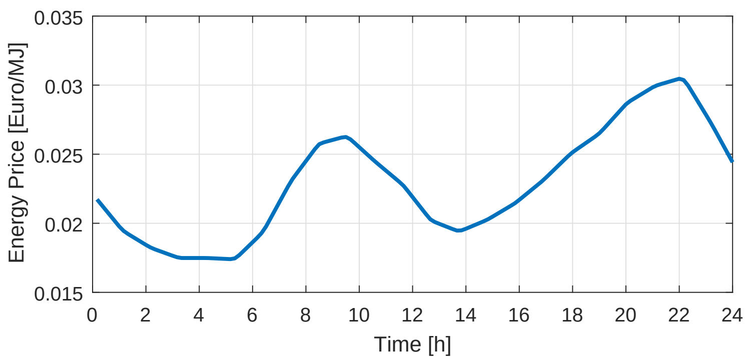

Figure 4: Energy price profile along the one-day time horizon.

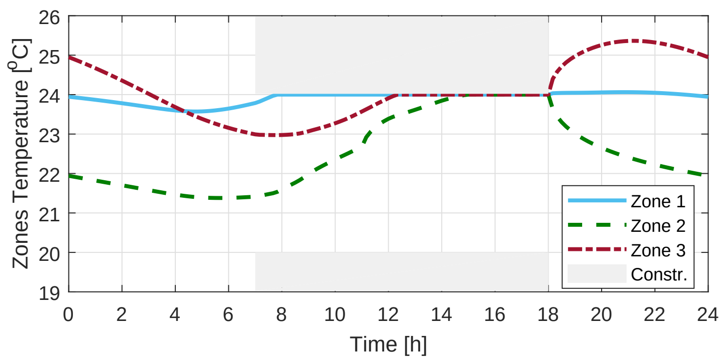

Figure 5: Optimal zone temperature profiles of building 1. The temperature of zone 2 (at the middle) is always the lowest, it acts as a passive thermal storage draining heat of the other zones through floor and ceiling.

the two communication protocols and in both cases the proposed distributed approach was able to retrieve the optimal solution.

Figure 5 shows the optimal temperature profiles for the three zones of building 1. It can be observed that, while the profiles of zones 1 and 2 are kept close to the maximum temperature bound of the working hours comfort range (outside the grey area), the temperature of zone 2 is always lower than the other two. Zone 2 is indeed subject to a pre-cooling phase before the occupancy period so as to cool down the building, acting as a passive thermal storage to drain the heat of the other zones through floor and ceiling. The temperature profiles of the other two buildings are very similar to that of building 1, and hence are not reported here.

In Figures 6 and 7 we report the cooling network exchange profiles computed by building 1 at iteration $k = 1$ and at consensus (when Algorithm 1 converges), respectively. From Figure 6 it is clear that, at the beginning, building 1 acts in a "selfish" manner and its optimal strategy is to constantly withdraw cooling energy from the cooling network ($e_s^1 > 0$, solid line), thus forcing buildings 2 and 3 to charge the cooling network ($e_s^2 < 0$ and $e_s^3 < 0$, dashed and dot-dashed lines, respectively). The stored energy is shown with the black dotted line. The consensus solution depicted in Figure 7 is instead cooperative. Building 3, which

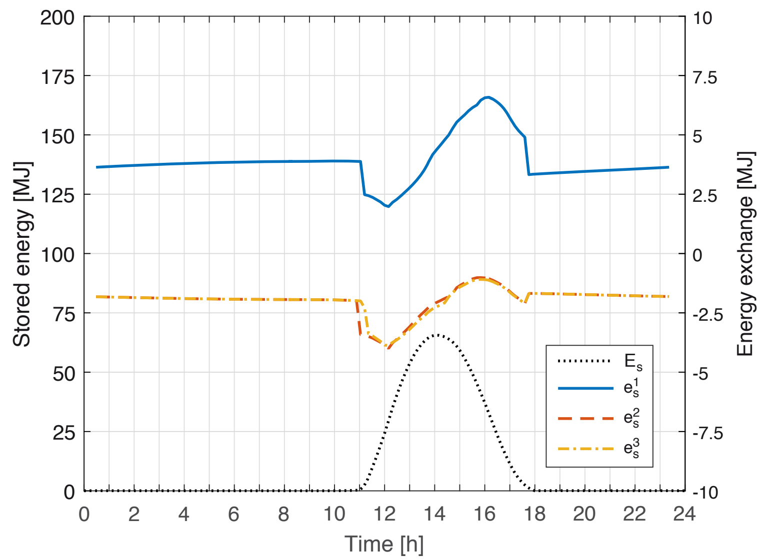

Figure 6: Cooling network exchange profiles at iteration $k = 1$. Building 1 is constantly withdrawing cooling energy ($e_s^1 > 0$, solid line), thus forcing buildings 2 and 3 to charge the cooling network ($e_s^2 < 0$ and $e_s^3 < 0$, dashed and dot-dashed lines, respectively). The stored energy is shown with the black dotted line.

has the biggest chiller, is constantly providing cooling energy ($e_s^3 < 0$) to the shared cooling network; building 2, which has the smallest chiller, is constantly withdrawing energy ($e_s^2 > 0$) from it; and building 1 provides/retrieves energy to/from the cooling network depending on the time slot. In this way, differences in the chiller sizes are compensated through the cooling network.

The number of iterations needed to achieve consensus are 278 for the fixed topology and 1032 for the time-varying topology, where we considered the solution to be at consensus if either the absolute or the relative difference between the solutions of the agents across two consecutive iterations was less than a given threshold, which was taken to be 10^{-3}.

We compared the performance achieved in the considered set-up of three buildings sharing a cooling network with that of a baseline setting where each building uses one third of the cooling network capacity, without exchanging

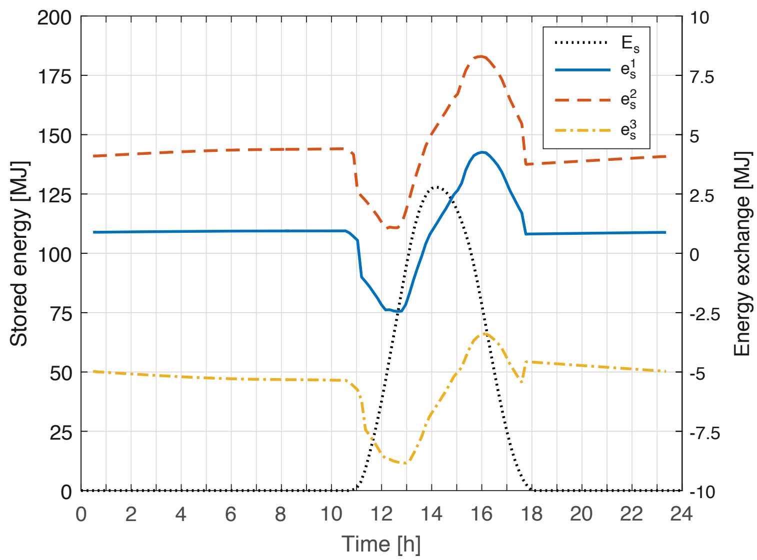

Figure 7: Cooling network exchange profiles at consensus. Cooperative solution, with building 3 constantly providing cooling energy ($e_s^3 < 0$) to the shared cooling network; building 2 constantly withdrawing energy ($e_s^2 > 0$) from it; and building 1 providing/withdrawing energy depending on the time slot. The stored energy is shown with the black dotted line.

cooling energy with the other buildings. This is the case when buildings do not communicate, and each one optimizes the use of its own share of energy in the cooling network. In the baseline setting, the electrical energy cost for the district is 56.83 euros whereas it reduces to 48.28 euros in the case when buildings are sharing the common cooling network, with a saving of about 15%.

Figure 8 shows the COP coefficient of the chillers of the three buildings for the baseline setting. Figure 9 shows the same quantities but for the case when the cooling energy in the cooling network is shared. In Figure 8 the chiller of building 1 is clearly better performing with respect to the other two, whereas the consensus solution reported in Figure 9 shows that the efficiency of the two other chillers is increased significantly. As for the individual costs, in the baseline setting, buildings 1, 2, and 3 spend 15.43, 20.72, and 20.68, respectively, whereas if they share the cooling network their costs become 13.63, 7.17, and 27.48, with an increase of the amount spent by building 3 (the one that owns the large chiller) that is largely compensated by the decrease of those of the other two buildings.

5. Concluding remarks

In this paper we proposed a distributed scheme for structured multi-agent decision making problems over a time-varying communication network, involving both local and global decision variables, with the feasibility domain of the local decision variables depending on the global ones, and the individual objective functions depending on the global decision variables. In particular, a proximal minimization based approach was adopted, and a theoretical extension to an algorithm that recently appeared in [8] was provided. The proposed scheme does not require for agents to reveal information that is considered as private, and overcomes the communication and computational challenges imposed by centralized or decentralized optimization paradigms.

The efficacy of the proposed distributed algorithm was illustrated by means of a detailed simulation study on the cooling of multiple buildings in a district

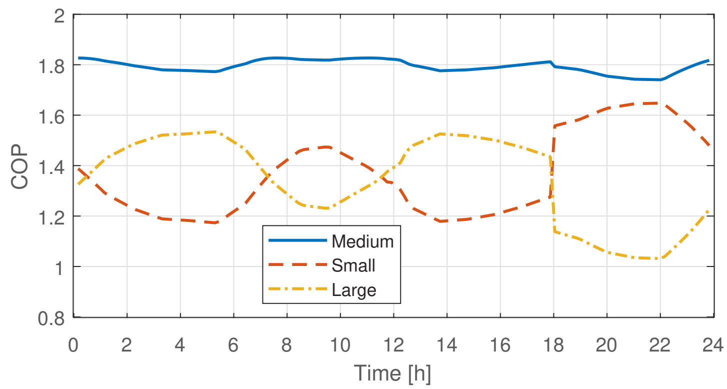

Figure 8: COP profiles for the baseline setting.

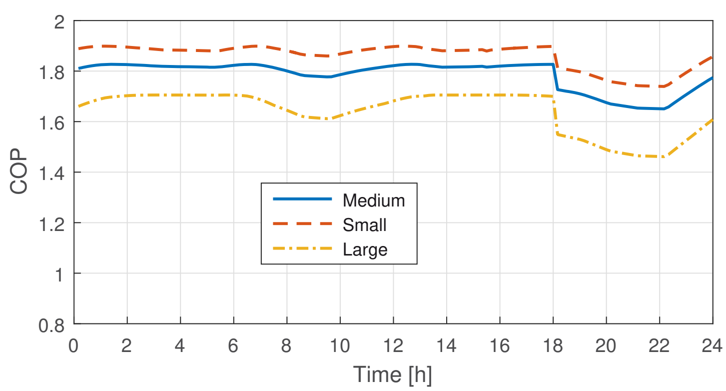

Figure 9: COP profiles when the energy in the cooling network is shared.

sharing a cooling network. The considered case study refers to a simple set-up, which, however, could be extended to a more realistic one by including local generation capabilities, intermittent sources, electrical storage systems. As long as convexity is preserved, nonlinearity will not be an issue. To relax the convexity requirement of Assumption 1, convex relaxations techniques with guaranteed relaxation gap may be employed. To this end, we aim at investigating the case where integer variables related, e.g., to on/off device switching, are introduced. The combinatorial nature of the resulting mixed integer optimization problem makes distributed schemes involving the solution of lower dimensional problems very appealing. However, devising effective distributed solutions is far more challenging and, indeed, only a limited number of results are available in the literature on decentralized, [33, 34], and distributed, [35], mixed integer optimization.

Current work concentrates on extending the proposed distributed optimization scheme for structured decision making problems to the stochastic case, based on the scenario-based solution to stochastic distributed optimization proposed in [8]. This extension will enable the design of a distributed energy management algorithm that is robust against the uncertainty on the disturbances affecting the thermal dynamics of a building.

Appendix A. Proofs

Proof of Lemma 1. For each $i = 1, \dots, m$, fix any $x, x' \in X_i$ and $\lambda \in [0, 1]$. By (7), let

$$u_i^\star(x) \in \arg \min_{u_i \in U_i(x)} f_i(x, u_i), \tag{A.1}$$

$$u_i^\star(x') \in \arg \min_{u_i \in U_i(x')} f_i(x', u_i). \tag{A.2}$$

Note that the existence of such minimizers is guaranteed by Weierstrass' theorem (Proposition A.8, p. 625 in [1]), since $U_i(x), U_i(x')$ are compact and non-empty (Assumption 2), and $f_i(\cdot, \cdot)$ is continuous (Assumption 1).

Since $u_i^\star(x) \in U_i(x)$ and $u_i^\star(x') \in U_i(x')$, we have that $(x, u_i^\star(x)) \in V_i$ and $(x', u_i^\star(x')) \in V_i$, which, given the convexity of V_i (Assumption 2), implies that

$$\big(\lambda x + (1-\lambda)x', \lambda u_i^\star(x) + (1-\lambda)u_i^\star(x')\big) \in V_i. \tag{A.3}$$

This also implies that $\lambda u_i^\star(x) + (1-\lambda)u_i^\star(x') \in U_i(\lambda x + (1-\lambda)x')$ (see (2)). We then have

$$\begin{aligned} g_i(\lambda x + (1-\lambda)x') &= \min_{u_i \in U_i(\lambda x + (1-\lambda)x')} f_i(\lambda x + (1-\lambda)x', u_i) \\ &\leq f_i(\lambda x + (1-\lambda)x', \lambda u_i^\star(x) + (1-\lambda)u_i^\star(x')) \\ &\leq \lambda f_i(x, u_i^\star(x)) + (1-\lambda) f_i(x', u_i^\star(x')) \\ &= \lambda g_i(x) + (1-\lambda) g_i(x'), \end{aligned} \tag{A.4}$$

where the first inequality follows because $\lambda u_i^\star(x) + (1-\lambda)u_i^\star(x') \in U_i(\lambda x + (1-\lambda)x')$ and the definition of min, the second inequality because $f_i(\cdot,\cdot)$ is jointly convex with respect to its arguments (Assumption 1), whereas the last equality because (A.1), (A.2) and the definition of $g_i(\cdot)$ in (7). Since (A.4) holds for any $x, x' \in X_i$, and for any $\lambda \in [0,1]$, the convexity of $g_i(\cdot)$ on X_i remains proven. □

Proof of Lemma 2. The proof is inspired by the proof of Corollary 3.5 of [27]. For each $i = 1, \ldots, m$, fix any $x, x' \in X_i$. Let also $u_i^\star(x) \in U_i(x)$, $u_i^\star(x') \in U_i(x')$, be as in (A.1) and (A.2), respectively. By Assumption 3, we have for all $u_i \in \mathbb{R}^{n_i}$ that

$$\left| \min_{v_i \in U_i(x)} \|u_i - v_i\| - \min_{v_i' \in U_i(x')} \|u_i - v_i'\| \right| \leq L_i \|x - x'\|. \tag{A.5}$$

Take $u_i = u_i^\star(x)$. We then have that

$$\begin{aligned} \min_{v_i' \in U_i(x')} \|u_i^\star(x) - v_i'\| &\leq \min_{v_i \in U_i(x)} \|u_i^\star(x) - v_i\| + L_i\|x - x'\| \\ &\leq L_i \|x - x'\|, \end{aligned} \tag{A.6}$$

where the last inequality holds because $u_i^\star(x) \in U_i(x)$.

Letting $\bar{v}_i' \in \arg\min_{v_i' \in U_i(x')} \|u_i^\star(x) - v_i'\|$, (A.6) is equivalent to

$$\|u_i^\star(x) - \bar{v}_i'\| \leq L_i \|x - x'\|. \tag{A.7}$$

Similarly, taking $u_i = u_i^\star(x')$ in (A.5) gives that $\min_{v_i \in U_i(x)} \|u_i^\star(x') - v_i\| \leq L_i\|x - x'\|$, which, letting $\bar{v}_i \in \arg\min_{v_i \in U_i(x)} \|u_i^\star(x') - v_i\|$ is equivalent to

$$\|u_i^\star(x') - \bar{v}_i\| \leq L_i\|x - x'\|. \tag{A.8}$$

Note that $\bar{v}_i, \bar{v}_i'$, exist due to the Weierstrass' theorem (Proposition A.8, p. 625 in [1]), since $U_i(x)$, $U_i(x')$ are compact and non-empty due to Assumption 2, and since $\|u_i^\star(x') - v_i\|$ and $\|u_i^\star(x) - v_i'\|$ are continuous with respect to v_i and v_i', respectively.

By Assumption 1, $f_i(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^{n_i} \to \mathbb{R}$ is Lipschitz continuous. Denoting its Lipschitz constant by $C_i \in \mathbb{R}$, $C_i > 0$, we have that

$$\begin{aligned} f_i(x', \bar{v}_i') &\leq f_i(x, u_i^\star(x)) + C_i\|x - x'\| + C_i\|u_i^\star(x) - \bar{v}_i'\| \\ &\leq f_i(x, u_i^\star(x)) + C_i(1 + L_i)\|x - x'\|, \end{aligned} \tag{A.9}$$

where the last inequality follows in view of (A.7). Since $\bar{v}_i' \in U_i(x')$ and since $u_i^\star(x')$ minimizes $f_i(x', \cdot)$ over $U_i(x')$, (A.9) yields

$$f_i(x', u_i^\star(x')) \leq f_i(x, u_i^\star(x)) + C_i(L_i + 1)\|x - x'\|. \tag{A.10}$$

Similarly, by the Lipschitz continuity of $f_i(\cdot,\cdot)$ and by using (A.8), we have that

$$\begin{aligned} f_i(x, \bar{v}_i) &\leq f_i(x', u_i^\star(x')) + C_i\|x - x'\| + C_i\|u_i^\star(x') - \bar{v}_i\| \\ &\leq f_i(x', u_i^\star(x')) + C_i(1 + L_i)\|x - x'\|. \end{aligned} \tag{A.11}$$

Since $\bar{v}_i \in U_i(x)$ and since $u_i^\star(x)$ minimizes $f_i(x, \cdot)$ over $U_i(x)$, (A.11) in turn gives that

$$f_i(x, u_i^\star(x)) \leq f_i(x', u_i^\star(x')) + C_i(L_i + 1)\|x - x'\|. \tag{A.12}$$

Combining (A.10) and (A.12) we have that $|f_i(x, u_i^\star(x)) - f_i(x', u_i^\star(x'))| \leq C_i(L_i + 1)\|x - x'\|$, which is equivalent to $|g_i(x) - g_i(x')| \leq C_i(L_i + 1)\|x - x'\|$, being $g_i(x) = f_i(x, u_i^\star(x))$ and $g_i(x') = f_i(x', u_i^\star(x'))$.

Hence, $g_i(\cdot)$ is Lipschitz continuous on X_i with Lipschitz constant $C_i(L_i+1)$. This concludes the proof. □

Appendix B. Step-size sequence

In this section we provide a numerical example that admits an analytic solution, and offers insight on how the $\{c(k)\}_{k\geq 0}$ sequence affects the algorithm convergence. To this end consider problem $\mathcal{P}$ with $m = 2$, $n = 1$, $n_i = 0$, $f_i(x, u_i) = f_i(x) = \gamma(x + s_i)$, with $\gamma > 0$ and $s_1 = 1$, $s_2 = -1$, and $V_i = [-M, M]$ with $1 < M < \infty$, for $i = 1, 2$. This amounts to solving the following optimization problem

$$\begin{aligned} \min_{x\in\mathbb{R}} \quad & \gamma(x+1)^2 + \gamma(x-1)^2 \\ \text{subject to} \quad & x \in [-M, M], \end{aligned} \tag{B.1}$$

where the cost function is split among two agents. Clearly, the optimal solution of (B.1) is achieved when $x = x^\star = 0$.

Let us apply Algorithm 1 to solve (B.1) in a distributed manner, where as far as the communication structure is concerned we assume that it is fully connected and choose $a^i_j(k) = 1/2$ for all $k \geq 0$, for $i, j = 1, 2$. Step 6 of Algorithm 1 simplifies to the following iteration

$$x_i(k+1) = \begin{cases} \min(\hat{x}_i(k+1), M), & \text{if } \hat{x}_i(k+1) \geq 0 \\ \max(\hat{x}_i(k+1), -M), & \text{otherwise,} \end{cases} \tag{B.2}$$

where

$$\hat{x}_i(k+1) = \frac{\bar{x}_i(k) - s_i 2\gamma c(k)}{2\gamma c(k) + 1}, \tag{B.3}$$

is the unconstrained minimizer of the optimization program of agent i. Therefore, (B.2) encodes the projection of $\hat{x}_i(k)$ on $[-M, M]$.

If we initialize the algorithm with $x_i(0) = \arg\min_{x_i \in V_i} = -s_i$, $i = 1, 2$, as suggested in Section 2.1, we have that $\bar{x}_i(0) = 0$, for $i = 1, 2$. It can then be easily shown using induction that $\bar{x}_i(k) = 0$ for all $k \geq 0$, for $i = 1, 2$, thus (B.2) reduces to

$$x_i(k+1) = -s_i \frac{2\gamma c(k)}{2\gamma c(k) + 1}. \tag{B.4}$$

By inspection of (B.4), we can note that the constraints are satisfied but are never active since $M > 1$. Moreover, we have that

- $c(k)$ has to converge to zero (as required by Assumption 4) for $x_i(k)$ to converge to $x^\star = 0$,
- the convergence rate of $x_i(k)$ is dictated by that of $c(k)$.

References

[1] D. Bertsekas, J. Tsitsiklis, Parallel and distributed computation: Numerical methods, Athena Scientific (republished in 1997), 1989.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning 3 (1) (2010) 1–122.

[3] A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization, IEEE Transactions on Automatic Control 54 (1) (2009) 48–61.

[4] A. Nedic, A. Ozdaglar, P. Parrilo, Constrained consensus and optimization in multi-agent networks, IEEE Transactions on Automatic Control 55 (4) (2010) 922–938.

[5] M. Zhu, S. Martinez, On distributed convex optimization under inequality and equality constraints, IEEE Transactions on Automatic Control 57 (1) (2012) 151–164.

[6] S. Lee, A. Nedic, Distributed random projection algorithm for convex optimization, IEEE Journal on Selected Topics in Signal Processing 7 (2) (2013) 221–229.

[7] K. Margellos, A. Falsone, S. Garatti, M. Prandini, Proximal minimization based distributed convex optimization, in: 2016 American Control Conference (ACC), 2016, pp. 2466–2471.

[8] K. Margellos, A. Falsone, S. Garatti, M. Prandini, Distributed constrained optimization and consensus in uncertain networks via proximal minimization, IEEE Transactions on Automatic Control 63 (5) (2018) 1372–1387.

[9] T.-H. Chang, A. Nedic, A. Scaglione, Distributed Constrained Optimization by Consensus-Based Primal-Dual Perturbation Method, IEEE Transactions on Automatic Control 59 (6) (2014) 1524–1538.

[10] A. Falsone, K. Margellos, S. Garatti, M. Prandini, Dual decomposition for multi-agent distributed optimization with coupling constraints, Automatica 84 (2017) 149 – 158.

[11] J. Lausten, Energy efficiency requirements in building codes, energy efficiency policies for new buildings, International Energy Agency (IEA) (2008) 477–488.

[12] G. Henze, D. Kalz, S. Liu, C. Felsmann, Experimental analysis of model-based predictive optimal control for active and passive building thermal storage inventory, International Journal of HVAC & Research 11 (2) (2005) 189–214.

[13] Y. Ma, F. Borrelli, B. Hencey, A. Packard, S. Bortoff, Model predictive control of thermal energy storage in building cooling systems, IEEE Conference on Decision and Control (2009) 392–397.

[14] J. Siroky, F. Oldewurtel, J. Cigler, S. Privara, Experimental analysis of model predictive control for an energy efficient building heating system, Applied Energy 88 (9) (2011) 3079–3087.

[15] Y. Ma, A. Kelman, A. Daly, F. Borrelli, Predictive control for energy efficient buildings with thermal storage: Modeling, simulation and experiments, IEEE Control Systems Magazine 1 (1) (2012) 45–66.

[16] F. Oldewurtel, A. Parisio, C. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, M. Morari, Experimental analysis of model predictive control

for an energy efficient building heating system, Energy and Buildings 45 (2) (2012) 15–27.

[17] K. Deng, Y. Sun, A. Chakraborty, Y. Lu, J. Brouwer, P. Mehta, Optimal scheduling of chiller plant with thermal energy storage using mixed integer linear programming, American Control Conference (2013) 2958 – 2963.

[18] D. Ioli, A. Falsone, S. Schuler, M. Prandini, A compositional framework for energy management of a smart grid: A scalable stochastic hybrid model for cooling of a district network, in: 2016 12th IEEE International Conference on Control and Automation (ICCA), 2016, pp. 389–394.

[19] D. Ioli, A. Falsone, A. V. Papadopoulos, M. Prandini, A compositional modeling framework for the optimal energy management of a district network, Journal of Process ControlAvailable online 6 November 2017. doi:https://doi.org/10.1016/j.jprocont.2017.10.005.

[20] G. Darivianakis, A. Georghiou, R. Smith, J. Lygeros, A Stochastic Optimization Approach to Cooperative Building Energy Management via an Energy Hub, IEEE Conference on Decision and Control (2015) 1 – 6.

[21] K. Paridari, A. Parisio, H. Sandberg, K. H. Johansson, Demand response for aggregated residential consumers with energy storage sharing, in: 2015 54th IEEE Conference on Decision and Control (CDC), 2015, pp. 2024–2030.

[22] D. Ioli, A. Falsone, M. Prandini, An iterative scheme to hierarchically structured optimal energy management of a microgrid, in: IEEE Conference on Decision and Control, Osaka, Japan, 2015, pp. 5227–5232.

[23] T.-H. Chang, M. Alizadeh, A. Scaglione, Real-time power balancing via decentralized coordinated home energy scheduling, IEEE Transactions on Smart Grid 4 (3) (2013) 1490–1504.

[24] S. Koehler, C. Danielson, F. Borrelli, A primal-dual active-set method for distributed model predictive control, Optimal Control Applications and Methods 38 (3) (2017) 399–419.

[25] A. Parisio, S. P. Gutierrez, Distributed model predictive control for building demand-side management, in: European Control Conference (ECC), IEEE, 2018, pp. 2549–2554.

[26] J. P. Aubin, Viability theory, Birkhäuser Boston, 1991.

[27] R. Wets, Lipschitz Continuity of inf-Projections, Computational Optimization and Applications 25 (1) (2003) 269–282.

[28] R. Rockafellar, R.-B. Wets, Variational analysis, Springer, 2009.

[29] R. Sinkhorn, P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, Pacific Journal of Mathematics 21 (2) (1967) 343–348.

[30] R. Baier, E. Farkhi, Regularity of set-valued maps and their selections through set differences. Part 1: Lipschitz continuity, Serdica Mathematical Journal 39 (3-4) (2013) 365–390.
URL `https://hal.inria.fr/hal-01067206`

[31] D. Ioli, A. Falsone, M. Prandini, Optimal energy management of a building cooling system with thermal storage: A convex formulation, IFAC International Symposium on Advanced Control of Chemical Processes (2015) 1151 – 1156.

[32] M. Gordon, K. C. Ng, Cool thermodynamics: The Engineering and Physics of Predictive, Diagnostic and Optimization Methods for Cooling Systems, Cambridge International Science Publishing, 2000.

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

[34] A. Falsone, K. Margellos, M. Prandini, A decentralized approach to multi-agent milps: finite-time feasibility and performance guarantees, Automatica 103 (2019) 141–150.

[35] A. Falsone, K. Margellos, M. Prandini, A distributed iterative algorithm for multi-agent MILPs: Finite-time feasibility and performance characterization, IEEE Control Systems Letters 2 (4) (2018) 563–568.