Distributed Constrained Optimization and Consensus in Uncertain Networks via Proximal Minimization

Kostas Margellos [ID] , Alessandro Falsone [ID] , Simone Garatti [ID] , and Maria Prandini [ID]

Abstract—We provide a unifying framework for distributed convex optimization over time-varying networks, in the presence of constraints and uncertainty, features that are typically treated separately in the literature. We adopt a proximal minimization perspective and show that this set-up allows us to bypass the difficulties of existing algorithms while simplifying the underlying mathematical analysis. We develop an iterative algorithm and show the convergence of the resulting scheme to some optimizer of the centralized problem. To deal with the case where the agents' constraint sets are affected by a possibly common uncertainty vector, we follow a scenario-based methodology and offer probabilistic guarantees regarding the feasibility properties of the resulting solution. To this end, we provide a distributed implementation of the scenario approach, allowing agents to use a different set of uncertainty scenarios in their local optimization programs. The efficacy of our algorithm is demonstrated by means of a numerical example related to a regression problem subject to regularization.

Index Terms—Consensus, distributed optimization, proximal minimization, scenario approach, uncertain systems.

I. INTRODUCTION

OPTIMIZATION in multiagent networks has attracted significant attention in the control and signal processing literature, due to its applicability in different domains like power systems [2], [3], wireless networks [4], [5], robotics [6], etc. Typically, agents solve a local decision-making problem, communicate their decisions with other agents, and repeat the process on the basis of the new information received. The main objective of this cooperative set-up is for agents to agree on a common decision that optimizes a certain performance criterion for the overall multiagent system while satisfying local constraints. This distributed optimization scheme leads to computational and communication savings compared to centralized paradigms, while allowing agents to keep privacy by exchanging partial information only.

A. Contributions of this Work

In this paper, we deal with distributed convex optimization problems over time-varying networks, under a possibly different constraint set per agent, and in the presence of uncertainty. Focusing first on the deterministic case, we construct an iterative, proximal minimization-based algorithm. Proximal minimization, where a penalty term (proxy) is introduced in the objective function of each agent's local decision problem, serves as an alternative to (sub)gradient methods. This is interesting per se, since it constitutes the multiagent counterpart of connections between proximal algorithms and gradient methods that have been established in the literature for single-agent problems (see [7]). Moreover, as observed in [8] with reference to incremental algorithms, the proximal minimization approach leads to numerically more stable algorithms compared to their gradient-based counterparts. A rigorous and detailed analysis is provided, showing that the proposed iterative scheme converges to an optimizer of the centralized problem counterpart. This is achieved without imposing differentiability assumptions or requiring excessive memory capabilities as other methods in the literature.

We move then to the case where constraints depend on an uncertain parameter and should be robustly satisfied for all values that this parameter may take. This poses additional challenges when devising a distributed solution methodology. Here, we exploit results on scenario-based optimization [9]–[14]. In particular, we assume that each agent is provided with its own set of uncertainty realizations (scenarios) and enforces the constraints corresponding to these scenarios only. We then show that our distributed algorithm is applicable and that the converged solution is feasible in a probabilistic sense for the constraints of the centralized problem, i.e., it satisfies with high probability all agents' constraints when an unseen uncertainty instance is realized. To achieve this, we rely on the novel contribution of [15], which leads to a sharper result compared to the one that would be obtained by a direct application of the basic scenario theory [10]. Our approach can be thought of as the data-driven counterpart of robust or worst-case optimization paradigms, enabling us to provide *a priori* guarantees on the probability of constraint satisfaction without imposing any assumptions on the underlying distribution of the uncertainty and its moments, and/or the geometry of the uncertainty sets (e.g., [16, Chapters

6, 7]); however, providing the overall feasibility statement with a certain confidence. The proposed distributed implementation of the scenario approach, which is instead typically performed in a centralized fashion, allows for a reduction of the communication burden and the satisfaction of privacy requirements regarding the available knowledge on the uncertain parameter.

B. Related Work

Most literature builds on the seminal work of [7], [17], and [18] (see also [19], [20], and references therein for a more recent problem exposition), where a wide range of decentralized optimization problems is considered, using techniques based on gradient descent, dual decomposition, and the method of multipliers. The recent work of [21] deals with similar problems but from a game theoretic perspective.

Distributed optimization problems, in the absence of constraints though, have been considered in [22]–[31]. In most of these references the underlying network is allowed to be time varying. In the presence of constraints, Newton-based or gradient/subgradient-based approaches are adopted in [20], [32]–[34], which show asymptotic agreement of the agents' solutions to an optimizer of the centralized problem; in [35] and [36], a distributed alternating direction method of multipliers approach is adopted and its convergence properties are analyzed, whereas in [37] and [38] a constraints consensus approach is adopted. In these contributions, however, the underlying network is time-invariant, while agents are required to have certain memory capabilities. In a time-varying environment, as that considered in this paper, Nedíc *et al.* [39] and Zhu and Martínez [40] propose a projected subgradient methodology to solve distributed convex optimization problems in the presence of constraints. In [39], however, the particular case where the agents' constraint sets are all identical is considered. As a result, the computational complexity of each agents' local optimization program is the same as that of the centralized algorithm. Our approach, which allows for different constraint sets per agent, is most closely related to the work of [40], but we adopt a proximal minimization instead of a subgradient-based perspective, thus avoiding the requirement for gradient/subgradient computation.

In most of the aforementioned references, a deterministic set-up is considered. Results taking into account both constraints and uncertainty have recently appeared in [41]–[43]. In [42], a penalty-based approach is adopted and the convergence of the proposed scheme is shown under the assumption that the algorithm is initialized with some feasible solution, which, however, can be difficult to compute. This is not required in the approach proposed in this paper. In [43], an asynchronous algorithm is developed for a quite particular communication protocol that involves gossiping, i.e., pairwise communication, under stronger regularity conditions (strong convexity of the agents' objective function).

Our set-up is closely related, albeit different from the approach of [41], which proposes a projected gradient descent approach, where at every iteration a random extraction of each agents' constraints is performed. In [41] almost sure convergence is proved, but this requires that different scenarios are extracted at every iteration, and these scenarios must be independent from each other, and independent across iterations. This creates difficulties in accounting for temporal correlation of the uncertain parameter, and poses challenges if sampling from the underlying distribution is computationally expensive. On the contrary, in our algorithm each agent is provided with a

TABLE I
CLASSIFICATION OF RELATED WORK

Agents' decision vectors	Network	time-invariant	time-varying
unconstrained		-	[17], [18], [23]–[31]
same constraints		-	[39]
different constraints	deterministic	[20], [32]–[39]	[40]
	uncertain	-	[41], our work

given number of scenarios and the same uncertainty scenarios are used at every iteration. In this case, convergence in [41] is not guaranteed, whilst our scenario-based approach provides probabilistic feasibility, as opposed to almost sure feasibility, guarantees. This probabilistic treatment of uncertainty, which is particularly suited to data-based optimization, does not appear, to the best of our knowledge, in any of the aforementioned references. Moreover, differently from [41], our proximal minimization perspective allows us to bypass the requirement for gradient computations, rendering the developed programs amenable to existing numerical solvers, and does not impose differentiability assumptions on the agents' objective functions and Lipschitz continuity of the objective gradients.

Finally, it is perhaps worth mentioning that our approach is fundamentally different from the randomized algorithm of [37], which is based on iteratively exchanging active constraints over a time-invariant network; in our case, the network is time varying and we do not require for constraint exchange, thus reducing the communication requirements.

For a quick overview, Table I provides a classification of the literature most closely related to our work in terms of communication requirements (which is related to whether the underlying network is time varying or not) and their ability to deal with different types of constraints (which is also related to the overall computational effort as explained before).

All the aforementioned references, and our work as well, are concerned with static optimization problems, or problems with discrete-time dynamics. As for distributed optimization for continuous-time systems, the interested reader is referred to [44]–[49], and references therein.

C. Structure of the Paper

In Section II, we provide a formal statement of the problem under study, and, focusing on the deterministic case, formulate the proposed distributed algorithm based on proximal minimization; convergence and optimality are also discussed, but to streamline the presentation all proofs, along with some preparatory results and useful relations regarding the agents' local solutions, are deferred to Section V. Section III deals with the stochastic case, where constraints are affected by uncertainty, following a scenario-based methodology. To illustrate the efficacy of our algorithm, Section IV provides a distributed implementation of a regression problem subject to $L1$-regularization. Finally, Section VI concludes the paper and provides some directions for future work.

C. Notation

$\mathbb{R}$, $\mathbb{R}_+$ denote the real and positive real numbers, and $\mathbb{N}$, $\mathbb{N}_+$ the natural and positive natural numbers, respectively. For any $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm of x, whereas

for a scalar $a \in \mathbb{R}$, $|a|$ denotes its absolute value. Moreover, $x^{\top}$ denotes the transpose vector of x. For a continuously differentiable function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ is the gradient of $f(x)$. Given a set X, we denote by $\mathrm{co}(X)$ its convex hull. We write $\mathrm{dist}(y, X)$ to denote the Euclidean distance of a vector y from a set X, i.e., $\mathrm{dist}(y, X) = \inf_{x \in X} \|y - x\|$.

A vector $a \in \mathbb{R}^m$ is said to be a stochastic vector if all its components a_j are nonnegative and sum up to one, i.e., $\sum_{j=1}^{m} a_j = 1$. Consider a square matrix $A \in \mathbb{R}^{m \times m}$ and denote its ith column by $a^i \in \mathbb{R}^m$, $i = 1, \ldots, m$. A is said to be doubly stochastic if both its rows and columns are stochastic vectors, i.e., $\sum_{i=1}^{m} a_j^i = 1$ for all $j = 1, \ldots, m$, and $\sum_{j=1}^{m} a_j^i = 1$ for all $i = 1, \ldots, m$.

II. DISTRIBUTED CONSTRAINED CONVEX OPTIMIZATION

A. Problem Set-Up

We consider a time-varying network of m agents that communicate to cooperatively solve

$$\mathcal{P}_\delta : \quad \min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} f_i(x)$$

$$\text{subject to } x \in \bigcap_{\delta \in \Delta} \bigcap_{i=1}^{m} X_i(\delta) \quad (1)$$

where $x \in \mathbb{R}^n$ represents a vector of n decision variables, and $\delta \in \Delta$. We assume that Δ is endowed with a σ-algebra $\mathcal{D}$ and that $\mathbb{P}$ is a fixed, but possibly unknown, probability measure defined over $\mathcal{D}$. For each $i = 1, \ldots, m$, $f_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is the objective function of agent i, whereas, for any $\delta \in \Delta$, $X_i(\delta) \subseteq \mathbb{R}^n$ is its constraint set.[1]

Problem $\mathcal{P}_\delta$ is a robust program, where any feasible solution x should belong to $\bigcap_{i=1}^{m} X_i(\delta)$ for all realizations $\delta \in \Delta$ of the uncertainty. Note that the fact that uncertainty appears only in the constraints and not in the objective functions is without loss of generality; in the opposite case, an epigraphic reformulation would recast the problem in the form of $\mathcal{P}_\delta$.

Due to the presence of uncertainty, problem $\mathcal{P}_\delta$ may be very difficult to solve, especially when Δ is a continuous set. Hence, a proper way to deal with uncertainty must be introduced. Moreover, our perspective is that $f_i(\cdot)$ and X_i represent private information, available only to agent i and/or even though the whole information were available to all agents, imposing all the constraints in one shot, would result in a computationally intensive program. This motivates the use of a distributed algorithm.

To ease the exposition of our distributed algorithm, we focus first on the following deterministic variant of $\mathcal{P}_\delta$ with constraint sets being independent of δ:

$$\mathcal{P} : \quad \min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} f_i(x)$$

$$\text{subject to } x \in \bigcap_{i=1}^{m} X_i. \quad (2)$$

$\mathcal{P}_\delta$ will be revisited in Section III, where we will specify how to deal with the presence of uncertainty. We impose the following standing assumption.

Algorithm 1: Distributed Algorithm.

1: **Initialization**
2:     Set $\{a_j^i(k)\}_{k \geq 0}$, for all $i, j = 1, \ldots, m$.
3:     Set $\{c(k)\}_{k \geq 0}$.
4:     $k = 0$.
5:     Consider $x_i(0) \in X_i$, for all $i = 1, \ldots, m$.
6: **For** $i = 1, \ldots, m$ **repeat until convergence**
7:     $z_i(k) = \sum_{j=1}^{m} a_j^i(k) x_j(k)$.
8:     $x_i(k+1) = \arg \min_{x_i \in X_i} f_i(x_i) + \frac{1}{2c(k)} \|z_i(k) - x_i\|^2$.
9:     $k \leftarrow k + 1$.

Assumption 1 (Convexity and Compactness): For each $i = 1, \ldots, m$, the function $f_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is convex, and the set $X_i \subseteq \mathbb{R}^n$ is convex and compact.

B. New Proximal Minimization-Based Algorithm

The pseudocode of the proposed proximal minimization-based iterative approach is given in Algorithm 1. Initially, each agent i, $i = 1, \ldots, m$, starts with some tentative value $x_i(0)$ which belongs to the local constraint set X_i of agent i, but not necessarily to $\bigcap_{i=1}^{m} X_i$. One sensible choice for $x_i(0)$ is to set it such that $x_i(0) \in \arg \min_{x_i \in X_i} f_i(x_i)$. At iteration k, each agent i constructs a weighted average $z_i(k)$ of the solutions communicated by the other agents and its local one (step 7, Algorithm 1, where $a_j^i(k)$ are the weights). Then, each agent solves a local minimization problem, involving its local objective function $f_i(x_i)$ and a quadratic term, penalizing the difference from $z_i(k)$ (step 8, Algorithm 1, where the coefficient $c(k)$, which is assumed to be nonincreasing with k, regulates the relative importance of the two terms). Note that, unlike $\mathcal{P}$, under Assumption 1 and due to the presence of the quadratic penalty term, the resulting problem is strictly convex with respect to x_i, and hence admits a unique solution.

For each $k \geq 0$, the information exchange between the m agents can be represented by a directed graph (V, E_k), where the nodes $V = \{1, \ldots, m\}$ are the agents and the set E_k of directed edges (j, i) indicating that at time k agent i receives information from agent j is given by

$$E_k = \{(j, i) : a_j^i(k) > 0\}. \quad (3)$$

From (3), we set $a_j^i(k) = 0$ in the absence of communication. If $(j, i) \in E_k$, we say that j is a neighboring agent of i at time k. Under this set-up, Algorithm 1 provides a fully distributed implementation, where at iteration k each agent $i = 1, \ldots, m$ receives information only from out-neighbors. Moreover, this information exchange is time varying and may be occasionally absent. However, the following connectivity and communication assumption is made, where $E_\infty = \{(j, i) : (j, i) \in E_k$ for infinitely many $k\}$ denotes the set of edges (j, i) representing agent pairs that communicate directly infinitely often.

Assumption 2 (Connectivity and Communication): The graph (V, E_∞) is strongly connected, i.e., for any two nodes there exists a path of directed edges that connects them. Moreover, there exists $T \geq 1$ such that for every $(j, i) \in E_\infty$, agent i receives information from a neighboring agent j at least once every consecutive T iterations.

Assumption 2 guarantees that any pair of agents communicates directly infinitely often, and the intercommunication interval is bounded. For further interpretation of the imposed network structure, the reader is referred to [28], [39].

Algorithm 1 terminates if the iterates maintained by all agents converge. From an implementation point of view, agent i, $i = 1, \ldots, m$, will terminate its update process if the absolute difference (relative difference can also be used) between two consecutive iterates $\|x_i(k+1) - x_i(k)\|$ keeps below some user-defined tolerance for a number of iterations equal to T (see Assumption 2) times the diameter of the graph (i.e., the greatest distance between any pair of nodes connected via an edge in E_∞). This is the worst-case number of iterations required for an agent to communicate with all others in the network; note that if an agent terminated the process at the first iteration where the desired tolerance is met, then convergence would not be guaranteed since its solution may still change as an effect of other agents updating their solutions.

The proposed iterative methodology resembles the structure of proximal minimization for constrained convex optimization [7, Chapter 3.4.3]. The difference, however, is that our set-up is distributed and the quadratic term in step 8 does not penalize the deviation of x_i from the previous iterate $x_i(k)$, but from an appropriately weighted average $z_i(k)$. Note that, in contrast with the inspiring work in [39]–[41] addressing $\mathcal{P}$ under a similar set-up but following a projected subgradient approach, our proximal minimization-based approach allows for an intuitive economic interpretation: at every iteration k, we penalize a consensus residual proxy by the time-varying coefficient $1/(2c(k))$, which progressively increases. This can be thought of as a pricing settling mechanism, where the more we delay to achieve consensus the higher the price is.

In the case where $a_j^i(k) = 1/m$ for all $i, j = 1, \ldots, m$, for all $k \geq 0$, that corresponds to a decentralized control paradigm, the solution of our proximal minimization approach coincides with the one obtained when the alternating direction of multipliers [7], [19], is applied to this problem (see [7, eq. (4.72)–(4.74), p. 254]). In the latter, the quadratic penalty term is not added to the local objective function as in step 8 of Algorithm 1, but to the Lagrangian function of an equivalent problem, and the coefficient $c(k)$ is an arbitrary constant independent of k; however, a dual-update step is required. Formal connections between penalty methods and the method of multipliers have been established in [50].

Remark 1 (Application to a Specific Problem Structure): Algorithm 1 can be simplified when the underlying optimization problem exhibits a specific structure, namely agents need to agree on a common vector $y \in \mathbb{R}^{\bar{n}}$, but each of them decides upon a local vector $u_i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, m$ as well:

$$\min_{y \in \mathbb{R}^{\bar{n}}, \{u_i \in \mathbb{R}^{n_i}\}_{i=1}^m} \sum_{i=1}^m f_i(y, u_i)$$

$$\text{subject to } y \in \bigcap_{i=1}^m Y_i, \ u_i \in U_i, \ i = 1, \ldots, m \quad (4)$$

where $Y_i \in \mathbb{R}^{\bar{n}}$ and $U_i \subseteq \mathbb{R}^{n_i}$, for all $i = 1, \ldots, m$. Provided that Assumptions 1-5 hold for problem (4) with $x = (y, u_1, \ldots, u_m)$ and $X_i = Y_i \times \mathbb{R}^{n_1} \times \cdots \times U_i \times \cdots \times \mathbb{R}^{n_m}$, we can rewrite it as $\min_{y \in \mathbb{R}^{\bar{n}}} \sum_{i=1}^m g_i(y)$ subject to $y \in \bigcap_{i=1}^m Y_i$, where $g_i(y) = \min_{u_i \in U_i} f_i(y, u_i)$ and simplify

Algorithm 1 by replacing steps 7 and 8 with

$$z_i(k) = \sum_{j=1}^m a_j^i(k) y_j(k),$$

$$\big(y_i(k+1), u_i(k+1)\big)$$

$$= \arg \min_{y_i \in Y_i, u_i \in U_i} f_i(y_i, u_i) + \frac{1}{2c(k)} \|z_i(k) - y_i\|^2.$$

This entails that agents only need to communicate their local estimates $y_i(k)$, $i = 1, \ldots, m$, of y while the local solutions related to u_i, $i = 1, \ldots, m$, need not be exchanged.

C. Further Structural Assumptions and Communication Requirements

We impose some additional assumptions on the structure of problem $\mathcal{P}$ in (2) and the communication set-up that is considered in this paper. These assumptions will play a crucial role in the proof of convergence of Section V.

Note that due to Assumption 1, co $\left(\bigcup_{i=1}^m X_i\right)$ is also compact. Let then $D \in \mathbb{R}_+$ be such that $\|x\| \leq D$ for all $x \in$ co $\left(\bigcup_{i=1}^m X_i\right)$. Moreover, due to Assumption 1, $f_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous on X_i with Lipschitz constant $L_i \in \mathbb{R}_+$, i.e., for all $i = 1, \ldots, m$

$$|f_i(x) - f_i(y)| \leq L_i \|x - y\|, \text{ for all } x, y \in X_i. \quad (5)$$

Assumption 3 (Interior Point): The feasibility region $\bigcap_{i=1}^m X_i$ of $\mathcal{P}$ has a nonempty interior, i.e., there exists $\bar{x} \in \bigcap_{i=1}^m X_i$ and $\rho \in \mathbb{R}_+$, such that $\{x \in \mathbb{R}^n : \|x - \bar{x}\| < \rho\} \subset \bigcap_{i=1}^m X_i$.

Due to Assumption 3, by Weierstrass' theorem ([7, Proposition A.8, p. 625]), $\mathcal{P}$ admits at least one optimal solution. Therefore, if we denote by $X^* \subseteq \bigcap_{i=1}^m X_i$ the set of optimizers of $\mathcal{P}$, then X^* is nonempty. Notice also that $f_i(\cdot)$, $i = 1, \ldots, m$, is continuous due to the convexity condition of Assumption 1; the addition of the compactness part of the assumption is to imply Lipschitz continuity. However, $f_i(\cdot)$, $i = 1, \ldots, m$, is not required to be differentiable.

We impose the following assumption on the coefficients $\{c(k)\}_{k \geq 0}$, that appear in step 8 of Algorithm 1.

Assumption 4 (Coefficient $\{c(k)\}_{k \geq 0}$): Assume that for all $k \geq 0$, $c(k) \in \mathbb{R}_+$ and $\{c(k)\}_{k \geq 0}$ is a nonincreasing sequence, i.e., $c(k) \leq c(r)$ for all $k \geq r$, with $r \geq 0$. Moreover,
 1) $\sum_{k=0}^\infty c(k) = \infty$;
 2) $\sum_{k=0}^\infty c(k)^2 < \infty$.

In standard proximal minimization [7] convergence is highly dependent on the appropriate choice of $c(k)$. Assumption 4 is in fact needed to guarantee the convergence of Algorithm 1. A direct consequence of the last part of Assumption 4 is that $\lim_{k \to \infty} c(k) = 0$. One choice for $\{c(k)\}_{k \geq 0}$ that satisfies the conditions of Assumption 4 is to select it from the class of generalized harmonic series, e.g., $c(k) = \alpha/(k+1)$ for some $\alpha \in \mathbb{R}_+$. Note that Assumption 4 is in a sense analogous to the conditions that Nedic *et al.* [39] and Zhu and Martínez [40] impose on the step size of their subgradient algorithms. It should be also noted that our set-up is synchronous, using the same $c(k)$ for all agents, at every iteration k. Extension to an asynchronous implementation is a topic for future work.

In line with [17], [18], and [29], we impose the following assumptions on the information exchange between the agents.

Assumption 5 (Weight Coefficients): There exists $\eta \in (0,1)$ such that for all $i,j \in \{1,\ldots,m\}$ and all $k \geq 0$, $a_j^i(k) \in \mathbb{R}_+ \cup \{0\}$, $a_j^i(k) \geq \eta$, and $a_j^i(k) > 0$ implies that $a_j^i(k) \geq \eta$. Moreover, for all $k \geq 0$

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

Assumptions 2 and 5 are identical to [39, Assumptions 2 and 5] (the same assumptions are also imposed in [40]), but were reported also here to ease the reader and facilitate the exposition of our results. Note that these are rather standard for distributed optimization and consensus problems; for possible relaxations, the reader is referred to [29] and [51]. The interpretation of having a uniform lower bound η, independent of k, for the coefficients $a_j^i(k)$ in Assumption 5 is that it ensures that each agent is mixing information received by other agents at a nondiminishing rate in time [39]. Moreover, points 1) and 2) in Assumption 5 ensure that this mixing is a convex combination of the other agent estimates, assigning a nonzero weight to its local one since $a_i^i(k) \geq \eta$. Note that satisfying Assumption 5 requires agents to agree on an infinite sequence of doubly stochastic matrices (double stochasticity arises due to conditions 1 and 2 in Assumption 5), where $a_j^i(k)$ would be element (i,j) of the matrix at iteration k. This agreement should be performed prior to the execution of the algorithm in a centralized manner, and the resulting matrices have to be communicated to all agents via some consensus scheme; this is standard in distributed optimization algorithms of this type (see also [29], [39], and [40]). It would be of interest to construct doubly stochastic matrices in a distributed manner using the machinery of [52]; however, exploiting these results requires further investigation and is outside the scope of the paper.

D. Statement of the Main Convergence Result

Under the structural assumptions and the communication set-up imposed in the previous section, Algorithm 1 converges and agents reach consensus, in the sense that their local estimates $x_i(k)$, $i = 1,\ldots,m$, converge to some minimizer of problem $\mathcal{P}$. This is formally stated in the following theorem, which constitutes one main contribution of our paper.

Theorem 1: Consider Assumptions 1–5 and Algorithm 1. We have that, for some minimizer $x^* \in X^*$ of $\mathcal{P}$

$$\lim_{k \to \infty} \|x_i(k) - x^*\| = 0, \text{ for all } i = 1,\ldots,m. \qquad (6)$$

To streamline the contribution of the paper, the rather technical proof of this statement is deferred to Section V-B2.

III. DEALING WITH UNCERTAINTY

In this section, we revisit problem $\mathcal{P}_\delta$ in (1), and give a methodology to deal with the presence of uncertainty. Motivated by data driven considerations, we assume that each agent i, $i = 1,\ldots,m$, is provided with a fixed number of realizations of δ, referred to as scenarios, extracted according to the underlying probability measure $\mathbb{P}$ with which δ takes values in Δ. According to the information about the scenarios that agents possess, two cases are distinguished in the sequel, scenarios as a common resource versus scenarios as a private resource.

Throughout, the following modifications to Assumptions 1 and 3 are imposed.

1) For each $i = 1,\ldots,m$, $X_i(\delta)$ is a convex set for any $\delta \in \Delta$.

2) For each $i = 1,\ldots,m$, and for any finite set S of values for δ, $\bigcap_{\delta \in S} X_i(\delta)$ is compact.
3) For any finite set S of values for δ, $\bigcap_{i=1}^m \bigcap_{\delta \in S} X_i(\delta)$ has a nonempty interior.

For the subsequent analysis, note that for any $N \in \mathbb{N}_+$, $\mathbb{P}^N$ denotes the corresponding product measure. We assume measurability of all the involved functions and sets.

A. Probabilistic Feasibility—Scenarios as a Common Resource

We first consider the case where all agents are provided with the same scenarios of δ, i.e., scenarios can be thought of as a common resource for the agents. This is the case if all the agents have access to the same set of historical data for δ, or if agents communicate the scenarios with each other. The latter case, however, increases the communication requirements.

Let $\bar{N} \in \mathbb{N}_+$ denote the number of scenarios, and $\bar{S} = \{\delta^{(1)},\ldots,\delta^{(\bar{N})}\} \subset \Delta$ be the set of scenarios available to all agents. The scenarios are independently and identically distributed (i.i.d.) according to $\mathbb{P}$. Consider then the following optimization program $\mathcal{P}_{\bar{N}}$, where the subscript $\bar{N}$ is introduced to emphasize the dependency with respect to the uncertainty scenarios

$$\mathcal{P}_{\bar{N}} : \quad \min_{x \in \mathbb{R}^n} \sum_{i=1}^m f_i(x)$$

$$\text{subject to } x \in \bigcap_{\delta \in \bar{S}} \bigcap_{i=1}^m X_i(\delta). \qquad (7)$$

Clearly, $x \in \bigcap_{\delta \in \bar{S}} \bigcap_{i=1}^m X_i(\delta)$ is equivalent to $x \in \bigcap_{i=1}^m \bigcap_{\delta \in \bar{S}} X_i(\delta)$, and $\mathcal{P}_{\bar{N}}$ is amenable to be solved via the distributed algorithm of Section II-A. In fact, one can apply Algorithm 1 with $\bigcap_{\delta \in \bar{S}} X_i(\delta)$ in place of X_i, for all $i = 1,\ldots,m$. Let $X_{\bar{N}}^* \subseteq \bigcap_{i=1}^m \bigcap_{\delta \in \bar{S}} X_i(\delta)$ be the set of minimizers of $\mathcal{P}_{\bar{N}}$. We then have the following corollary of Theorem 1.

Corollary 1: Consider Assumptions 1–5 with the modifications stated in Section III and Algorithm 1. We have that, for some $x_{\bar{N}}^* \in X_{\bar{N}}^*$

$$\lim_{k \to \infty} \|x_{i,\bar{N}}(k) - x_{\bar{N}}^*\| = 0, \text{ for all } i = 1,\ldots,m \qquad (8)$$

where $x_{i,\bar{N}}(k)$ denotes the solution generated at iteration k, step 8 of Algorithm 1, when X_i is replaced by $\bigcap_{\delta \in \bar{S}} X_i(\delta)$.

We address the problem of quantifying the robustness of the minimizer $x_{\bar{N}}^*$ of $\mathcal{P}_{\bar{N}}$ to which our iterative scheme converges according to Corollary 1. In the current set-up, a complete answer is given by the scenario approach theory [9], [10], which shows that $x_{\bar{N}}^*$ is feasible for $\mathcal{P}_\delta$ up to a quantifiable level $\bar{\varepsilon}$. This result is based on the notion of support constraints (see also [9, Definition 4]), and in particular on the notion of support set [15] (also referred to as compression scheme in [14]). Given an optimization program, we say that a subset of the constraints constitutes a support set, if it is the minimal cardinality subset of the constraints such that by solving the optimization problem considering only this subset of constraints, we obtain the same solution to the original problem, where all the constraints are enforced. As a consequence, all constraints that do not belong to the support set are in a sense redundant since their removal leaves the optimal solution unaffected.

By [9, Theorem 3], for any convex optimization program, the cardinality of the support set is at most equal to the number

of decision variables n, whereas in [53], a refined bound is provided. The subsequent result is valid for any given bound on the cardinality of the support set. Therefore, and since $\mathcal{P}_{\bar{N}}$ is convex, let $d \in \mathbb{N}_+$ be a known upper bound for the cardinality of its support set. A direct application of the scenario approach theory in [9] leads to the following result.

Theorem 2: Fix $\beta \in (0, 1)$ and let

$$\bar{\varepsilon} = 1 - \sqrt[\bar{N}-d]{\frac{\beta}{\binom{\bar{N}}{d}}}. \tag{9}$$

We then have that

$$\mathbb{P}^{\bar{N}} \left\{ \bar{S} \in \Delta^{\bar{N}} : \right.$$
$$\left. \mathbb{P} \left\{ \delta \in \Delta : x_{\bar{N}}^* \notin \bigcap_{i=1}^m X_i(\delta) \right\} \leq \bar{\varepsilon} \right\} \geq 1 - \beta. \tag{10}$$

In words, Theorem 2 implies that with confidence at least $1 - \beta$, $x_{\bar{N}}^*$ is feasible for $\mathcal{P}_\delta$ apart from a set of uncertainty instances with measure at most $\bar{\varepsilon}$. Notice that $\bar{\varepsilon}$ is in fact a function of $\bar{N}, \beta$, and d. We suppress this dependency though to simplify notation. Note that even though $\mathcal{P}_{\bar{N}}$ does not necessarily have a unique solution, Theorem 2 still holds for the solution returned by Algorithm 1 (assuming convergence), since it is a deterministic algorithm and hence serves as a tie-break rule to select among the possibly multiple minimizers.

Following [10], (9) could be replaced with an improved $\bar{\varepsilon}$, obtained as the solution of $\sum_{k=0}^{d-1} \binom{\bar{N}}{k} \bar{\varepsilon}^k (1 - \bar{\varepsilon})^{\bar{N}-k} = \beta$. However, we use (9) since it gives an explicit relation expression for $\bar{\varepsilon}$, and also renders (10) directly comparable with the results provided in the next section.

In case $\bar{\varepsilon}$ exceeds one, the result becomes trivial. However, note that Theorem 2 can be also reversed (as in experiment design) to compute the number $\bar{N}$ of scenarios that is required for (10) to hold for given $\bar{\varepsilon}, \beta \in (0, 1)$. This can be determined by solving (9) with respect to $\bar{N}$ with the chosen $\bar{\varepsilon}$ fixed (e.g., using numerical inversion). The reader is referred to Theorem 1 of [9] for an explicit expression of $\bar{N}$.

B. Probabilistic Feasibility—Scenarios as a Private Resource

We now consider the case where the information carried by the scenarios is distributed, that is, each agent has its own set of scenarios, which constitute agents' private information. Specifically, assume that each agent i, $i = 1, \ldots, m$, is provided with a set $S_i = \{\delta_i^{(1)}, \ldots, \delta_i^{(N_i)}\} \subset \Delta$ of $N_i \in \mathbb{N}_+$ i.i.d. scenarios of δ, extracted according to the underlying probability measure $\mathbb{P}$. Here, $\delta_i^{(j)}$ denotes scenario j of agent i, $j = 1, \ldots, N_i$, $i = 1, \ldots, m$. The scenarios across the different sets S_i, $i = 1, \ldots, m$, are independent from each other. The total number of scenarios is $N = \sum_{i=1}^m N_i$. Consider then the following optimization program $\mathcal{P}_N$, where each agent has its own scenario set

$$\mathcal{P}_N : \min_{x \in \mathbb{R}^n} \sum_{i=1}^m f_i(x)$$

$$\text{subject to } x \in \bigcap_{i=1}^m \bigcap_{\delta \in S_i} X_i(\delta). \tag{11}$$

Program $\mathcal{P}_N$ can be solved via the distributed algorithm of Section II-A, so that a solution is obtained without exchanging any private information regarding the scenarios. In fact, one can apply Algorithm 1 with $\bigcap_{\delta \in S_i} X_i(\delta)$ in place of X_i, for all $i = 1, \ldots, m$.

Similarly to Corollary 1, letting $X_N^* \subseteq \bigcap_{i=1}^m \bigcap_{\delta \in S_i} X_i(\delta)$ be the set of minimizers of $\mathcal{P}_N$, we have the following corollary of Theorem 1.

Corollary 2: Consider Assumptions 1–5 with the modifications stated in Section III, and Algorithm 1. We have that, for some $x_N^* \in X_N^*$

$$\lim_{k \to \infty} \|x_{i,N}(k) - x_N^*\| = 0, \text{ for all } i = 1, \ldots, m \tag{12}$$

where $x_{i,N}(k)$ denotes the solution generated at iteration k, step 8 of Algorithm 1, when X_i is replaced by $\bigcap_{\delta \in S_i} X_i(\delta)$.

As in Section III-A, we show that the minimizer x_N^* of $\mathcal{P}_N$ to which our iterative scheme converges according to Corollary 2 is feasible in a probabilistic sense for $\mathcal{P}_\delta$. Here, a difficulty arises, since we seek to quantify the probability that x_N^* satisfies the global constraint $\bigcap_{i=1}^m X_i(\delta)$, where δ is a common parameter to all $X_i(\delta)$, $i = 1, \ldots, m$, while x_N^* has been computed considering $X_i(\delta)$ for uncertainty scenarios that are independent from those of $X_j(\delta)$, $j \neq i$, $i = 1, \ldots, m$.

Let $S = \{S_i\}_{i=1}^m$ be a collection of the scenarios of all agents. Similarly to the previous case, we denote by $d \in \mathbb{N}_+$ a known upper bound for the cardinality of the support set of $\mathcal{P}_N$. However, the way the constraints of the support set are split among the agents depends on the specific S employed. Therefore, for each set of scenarios S and for $i = 1, \ldots, m$, denote by $d_{i,N}(S) \in \mathbb{N}$ (possibly equal to zero) the number of constraints that belong to both the support set of $\mathcal{P}_N$ and S_i, i.e., the constraints of agent i. We then have that $\sum_{i=1}^m d_{i,N}(S) \leq d$, for any $S \in \Delta^N$. We will write $d_{i,N}$ in short instead of $d_{i,N}(S)$ and make the dependency on S explicit only when necessary.

1) Naive Result: For any collection of agents' scenarios, it clearly holds that $d_{i,N} \leq d$ for all $i = 1, \ldots, m$, for any scenario set. Thus, for each $i = 1, \ldots, m$, Theorem 2 can be applied conditionally to the scenarios of all other agents to obtain a local, in the sense that it holds only for the constraints of agent i, feasibility characterization. Fix $\beta_i \in (0, 1)$ and let

$$\widetilde{\varepsilon}_i = 1 - \sqrt[N_i-d]{\frac{\beta_i}{\binom{N_i}{d}}}. \tag{13}$$

We then have that

$$\mathbb{P}^N \left\{ S \in \Delta^N : \mathbb{P} \left\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \right\} \leq \widetilde{\varepsilon}_i \right\} \geq 1 - \beta_i. \tag{14}$$

By the subadditivity of $\mathbb{P}^N$ and $\mathbb{P}$, (14) can be used to quantify the probabilistic feasibility of x_N^* with respect to the global constraint $\bigcap_{i=1}^m X_i(\delta)$. Following the proof of [54, Corollary 1], where a similar argument is provided, we have that

$$\mathbb{P}^N \left\{ S \in \Delta^N : \mathbb{P} \left\{ \delta \in \Delta : x_N^* \notin \bigcap_{i=1}^m X_i(\delta) \right\} \leq \sum_{i=1}^m \widetilde{\varepsilon}_i \right\}$$

$$= \mathbb{P}^N \left\{ S \in \Delta^N : \mathbb{P} \left\{ \delta \in \Delta : \exists i \in \{1, \ldots, m\}, x_N^* \notin X_i(\delta) \right\} \right.$$

$$\leq \sum_{i=1}^m \widetilde{\varepsilon}_i \right\}$$

$$= \mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\Big\{ \bigcup_{i=1}^m \big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} \Big\} \leq \sum_{i=1}^m \widetilde{\varepsilon}_i \Big\}$$

$$\geq \mathbb{P}^N \Big\{ S \in \Delta^N : \sum_{i=1}^m \mathbb{P}\big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} \leq \sum_{i=1}^m \widetilde{\varepsilon}_i \Big\}$$

$$\geq \mathbb{P}^N \Big\{ \bigcap_{i=1}^m \big\{ S \in \Delta^N : \mathbb{P}\big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} \leq \widetilde{\varepsilon}_i \big\} \Big\}$$

$$\geq 1 - \sum_{i=1}^m \mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} > \widetilde{\varepsilon}_i \Big\}$$

$$\geq 1 - \sum_{i=1}^m \beta_i \tag{15}$$

which leads to the following proposition.

Proposition 1: Fix $\beta \in (0,1)$ and choose β_i, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \beta_i = \beta$. For each $i = 1, \ldots, m$, let $\widetilde{\varepsilon}_i$ be as in (13) and set $\widetilde{\varepsilon} = \sum_{i=1}^m \widetilde{\varepsilon}_i$. We then have that

$$\mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\Big\{ \delta \in \Delta : x_N^* \notin \bigcap_{i=1}^m X_i(\delta) \Big\} \leq \widetilde{\varepsilon} \Big\} \geq 1 - \beta. \tag{16}$$

Proposition 1 implies that with confidence at least $1 - \beta$, x_N^* is feasible for $\mathcal{P}_\delta$ apart from a set with measure at most $\widetilde{\varepsilon}$. This result, however, tends to be very conservative thus prohibiting its applicability to problems with a high number of agents. This can be seen by comparing $\widetilde{\varepsilon}$ with $\bar{\varepsilon}$, where the latter corresponds to the case where scenarios are treated as a common resource. To this end, consider the particular set-up where $N_i = \bar{N}$ and $\beta_i = \beta/m$, for all $i = 1, \ldots, m$. By (9) and (13), it follows that $\widetilde{\varepsilon} = m\widetilde{\varepsilon}_i \approx m\bar{\varepsilon}$, thus growing approximately (we do not have exact equality since $\beta_i = \beta/m$) linearly with the number of agents. This can be also observed in the numerical comparison of Section III-B2 (see Fig. 1). The issue with Proposition 1 is that it accounts for a worst-case setting, where $d_{i,N} = d$ for all $i = 1, \ldots, m$; however, this cannot occur, since $\sum_{i=1}^m d_{i,N} \leq d$ implies that if $d_{i,N} = d$ for some i, then $d_{j,N} = 0$, for all $j \neq i$, $i = 1, \ldots, m$.

2) Tighter Result: To alleviate the conservatism of Proposition 1, and exploit the fact that $\sum_{i=1}^m d_{i,N} \leq d$, we use the recent results of [15].

For each $i = 1, \ldots, m$, fix $\beta_i \in (0,1)$ and consider a function $\varepsilon_i(\cdot)$ defined as follows:

$$\varepsilon_i(k) = 1 - \sqrt[N_i - k]{\frac{\beta_i}{(d+1)\binom{N_i}{k}}}, \quad \text{for all } k = 0, \ldots, d. \tag{17}$$

Notice that $\varepsilon_i(\cdot)$ is also a function of N_i, β_i, and d, but this dependency is suppressed to simplify notation. For each $i = 1, \ldots, m$, working conditionally with respect to the scenarios $S \setminus S_i$ of all other agents, [15, Theorem 1] entails that

$$\mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} \leq \varepsilon_i(d_{i,N})$$
$$\Big| \big\{ S \setminus S_i \in \Delta^{N-N_i} \big\} \Big\} \geq 1 - \beta_i. \tag{18}$$
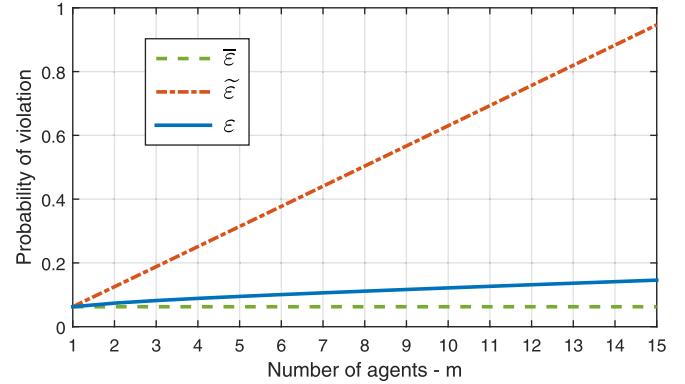

Fig. 1. Probability of constraint violation as a function of the number of agents, for the case where $d = 50$, $\beta = 10^{-5}$, $N_i = \bar{N} = 4500$, and $\beta_i = \beta/m$, for all $i = 1, \ldots, m$. The probability of violation $\bar{\varepsilon}$ (green dashed line) for the case of Section III-A is independent of m, so it remains constant as the number of agents m increases. For the case of Section III-B1, $\widetilde{\varepsilon} \approx m\bar{\varepsilon}$ (red dotted–dashed line) for the considered set-up, so it grows approximately linearly with m. For the case of Section III-B2, ε (blue solid line) is moderately increasing with m, thus offering a less conservative result compared to the approach of Section III-B1, while, in contrast to the approach of Section III-A, it allows for distributed information about the scenarios.

Integrating (18) with respect to the probability of realizing the scenarios $S \setminus S_i$, we have that

$$\mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\big\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \big\} \leq \varepsilon_i(d_{i,N}) \Big\}$$
$$\geq 1 - \beta_i. \tag{19}$$

The statement in (19) implies that for each agent $i = 1, \ldots, m$, with confidence at least $1 - \beta_i$, the probability that x_N^* does not belong to the constraint set $X_i(\delta)$ of agent i is at most equal to $\varepsilon_i(d_{i,N})$.

Note, however, that (19) is very different from (14), which is obtained by means of the basic scenario approach theory, since $d_{i,N}$ is not known *a priori* but depends on the extracted scenarios. Using (19) in place of (14) in the derivations of (15), by the subadditivity of $\mathbb{P}^N$ and $\mathbb{P}$, we have that

$$\mathbb{P}^N \Big\{ S \in \Delta^N : \mathbb{P}\Big\{ \delta \in \Delta : x_N^* \notin \bigcap_{i=1}^m X_i(\delta) \Big\}$$
$$\leq \sum_{i=1}^m \varepsilon_i(d_{i,N}) \Big\} \geq 1 - \sum_{i=1}^m \beta_i. \tag{20}$$

Unlike (10) and (16), (20) is an *a posteriori* statement due to the dependency of $\varepsilon_i(d_{i,N})$ on the extracted scenarios. However, the sought *a priori* result can be obtained by considering the worst-case value for $\sum_{i=1}^m \varepsilon_i(d_{i,N})$, with respect to the different combinations of $d_{i,N}$, $i = 1, \ldots, m$, satisfying $\sum_{i=1}^m d_{i,N} \leq d$. This can be achieved by means of the following maximization problem:

$$\varepsilon = \max_{\{d_i \in \mathbb{N}_+\}_{i=1}^m} \sum_{i=1}^m \varepsilon_i(d_i)$$
$$\text{subject to } \sum_{i=1}^m d_i \leq d. \tag{21}$$

Problem (21) is an integer optimization program. It can be solved numerically to obtain ε. The optimal value ε of the problem above depends on $\{N_i, \beta_i\}_{i=1}^m$ and d, but this dependency is suppressed to simplify notation. Notice the slight abuse of notation, since $\{d_i\}_{i=1}^m$ in (21) are integer decision variables and should not be related to $\{d_{i,N}\}_{i=1}^m$.

Theorem 3: Fix $\beta \in (0, 1)$ and choose β_i, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \beta_i = \beta$. Set ε according to (21). We then have that

$$\mathbb{P}^N \left\{ S \in \Delta^N : \mathbb{P} \left\{ \delta \in \Delta : x_N^* \notin \bigcap_{i=1}^m X_i(\delta) \right\} \leq \varepsilon \right\} \geq 1 - \beta. \tag{22}$$

Proof: Fix $\beta \in (0, 1)$ and choose β_i, $i = 1, \ldots, m$, such that $\sum_{i=1}^m \beta_i = \beta$. Consider any set S of scenarios and notice that $\sum_{i=1}^m d_{i,N}(S) \leq d$. This implies that $\{d_{i,N}(S)\}_{i=1}^m$ constitute a feasible solution of (21). Due to the fact that ε is the optimal value of (21), $\sum_{i=1}^m \varepsilon_i(d_{i,N}(S)) \leq \varepsilon$ for any S, which together with (20), leads to (22) and hence concludes the proof. ∎

The result of Theorem 3 can be significantly less conservative compared to that of Proposition 1, since we explicitly account for the fact that $\sum_{i=1}^m d_{i,N} \leq d$ in the maximization problem in (21). This can be also observed by means of the numerical example of Fig. 1, where we investigate how $\bar{\varepsilon}$, $\widetilde{\varepsilon}$, and ε change as a function of the number of agents m. We consider a particular case, where $d = 50$, $\beta = 10^{-5}$, $N_i = \bar{N} = 4500$, and $\beta_i = \beta/m$, for all $i = 1, \ldots, m$. For this set-up, where β is split evenly among agents and all agents have the same number of scenarios, it turned out that the maximum value ε in (21) is achieved for $d_i = d/m$, $i = 1, \ldots, m$. As it can be seen from Fig. 1, $\bar{\varepsilon}$ (green dashed line) for the case of Section III-A is independent of m, so it remains constant as the number of agents m increases. For the case of Section III-B1, $\widetilde{\varepsilon}$ (red dotted–dashed line) rows approximately linearly with m (see also discussion at the end of Section III-B1). For the case of Section III-B2, ε (blue solid line) is moderately increasing with m, thus offering a less conservative result compared to the approach of Section III-B1, while, in contrast to the approach of Section III-A, it allows for distributed information about the uncertainty scenarios.

In certain cases (e.g., when the number of agents is high), ε may still exceed one and hence the result of Theorem 3 becomes trivial (the same for Proposition 1 in such cases). Similarly to the discussion at the end of Section III-A, Theorem 3 can be reversed to compute the number of scenarios N_i that need to be extracted by agent i, $i = 1, \ldots, m$, for a given value of $\varepsilon, \beta \in (0, 1)$. This can be achieved by numerically seeking for values of N_i, $i = 1, \ldots, m$, that lead to a solution of (21) that attains the desired ε.

IV. NUMERICAL EXAMPLE

We address a multiagent regression problem subject to L_1-regularization, which is inspired by [55, Example 1]. Specifically, we consider m functions $s_i(\delta)$, $i = 1, \ldots, m$, which can, for instance, represent the effect of the same phenomenon at different locations of m agents. The functions are unknown, and each agent i has access to a (private) dataset $\{(\delta_i^{(j)}, s_i(\delta_i^{(j)})), j = 1, \ldots, N_i\}$ of measurements of function $s_i(\delta)$ only.

The agents seek to determine the magnitude of d cosinusoids at given frequencies, so that their superposition provides a central approximation of all the $s_i(\cdot)$, $i = 1, \ldots, m$. To this end, letting $x = [x^{[1]}, \ldots, x^{[d]}, x^{[d+1]}] \in \mathbb{R}^{d+1}$, the following
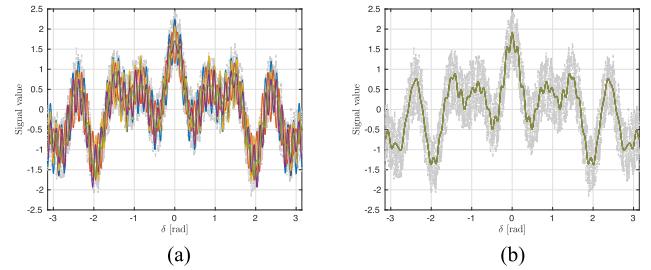

Fig. 2. Data points (grey crosses), and the functions (solid lines) corresponding to the local solutions returned by Algorithm 1 (a) at the initialization and (b) after 150 iterations.

program is considered:

$$\min_{x \in X \subset \mathbb{R}^{d+1}} x^{[d+1]} + \lambda \|x\|_1$$

$$\text{subject to } \left| \sum_{\ell=1}^d x^{[\ell]} \cos(\ell \delta_i^{(j)}) - s_i(\delta_i^{(j)}) \right| \leq x^{[d+1]},$$

$$\text{for all } j = 1, \ldots, N_i, \text{ for all } i = 1, \ldots, m. \tag{23}$$

In (23), one minimizes $x^{[d+1]}$, which is the worst-case approximation error over the data-points of all agents, plus a regularization term $\lambda \|x\|_1$, which induces sparsity in the solution. The set X is a hyperrectangular with high enough edge length so that the solution remains the same as in the unconstrained case, and it is introduced to ensure compactness so that Algorithm 1 can be applied (in fact this set could be different per agent, and does not need to be agreed upfront). By setting $f_i(x) = (1/m)(x^{[d+1]} + \lambda \|x\|_1)$, $X_i(\delta) = \{x \in X : |\sum_{\ell=1}^d x^{[\ell]} \cos(\ell \delta) - s_i(\delta)| \leq x^{[d+1]}\}$, and $S_i = \{\delta_i^{(1)}, \ldots, \delta_i^{(N_i)}\}$, $i = 1, \ldots, m$, it is seen that problem (23) is in the form of $\mathcal{P}_N$, and, moreover, it satisfies the assumptions of Corollary 2. Hence, the distributed Algorithm 1 can be employed to compute the optimal solution of (23). Notice that x_i in Algorithm 1 corresponds to a copy of x maintained by agent i and should not be confused with $x^{[\ell]}$, which is the ℓth component of x. Each objective function $f_i(x)$ is nondifferentiable. In our simulation, we considered $m = 6$ agents on a ring of alternating communicating pairs (time-varying communication graph), and assigned at each step the same weight to both the local solution and that transmitted by the active neighbor. Moreover, we set $n = d + 1 = 51$, $\lambda = 0.001$, and $N_i = N = 4500$ for all $i = 1, \ldots, m$. All samples $\delta_i^{(j)}$, $i = 1, \ldots, m$, $j = 1, \ldots, N$, were independently drawn from a uniform distribution with support $[-\pi, \pi]$, while, mimicking [55], each $s_i(\delta_i^{(j)})$ was obtained by evaluating the sum of a certain number of randomly shifted cosinusoids. Finally, Algorithm 1 was initialized with the solutions satisfying the local constraints only and $c(k) = 0.05/(k + 1)$. Fig. 2 shows the data points for each agent (grey dots) and the functions $\sum_{\ell=1}^d x_i^{[\ell]} \cos(\ell \delta)$ corresponding to the agents' solutions returned by Algorithm 1 (a) at the initialization and (b) after 150 iterations. As it appears, in conformity to Corollary 2, all local solutions converge to a unique solution. The fact that this solution is also optimal can be experimentally inspected from Fig. 3, where the objective values corresponding to the agent local solutions as iterations progress are displayed against the optimal objective value of problem (23)
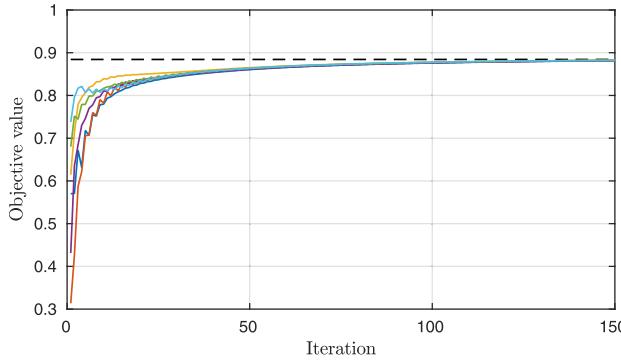
Fig. 3. Objective values corresponding to the agent local solutions as iterations progress (solid lines) versus optimal value of problem (23) computed via a centralized algorithm (dashed line).

computed via a centralized algorithm for comparison purposes. The value to which $x_i^{[d+1]}$ converged was 0.88. In our simulations, scenarios were treated as private resources as each agent's scenarios are independent of the scenarios of other agents. Nonetheless, for a newly seen observation δ, one may be interested in assessing the joint-constraint violation probability $\mathbb{P}\left\{ \delta \in \Delta : x_N^* \notin X_i(\delta) \right\}$, which in the present example corresponds to the probability of being apart from the obtained central function $\sum_{\ell=1}^d x_i^{[\ell]} \cos(\ell\delta)$ more than 0.88 for at least one of the function $s_i(\delta)$, $i = 1, \ldots, m$. Using 80 000 new scenarios (different from those used in the optimization process), this probability was empirically estimated as 0.01. Using $\beta = 10^{-5}$ and $d = 50$ (the bound on the dimension of the support set is $d = 50$ and not $d + 1$, since we do not need to account for the epigraphic variable $x^{[d+1]}$, see [53]), Proposition 1 and Theorem 3 give $\widetilde{\varepsilon} = 0.37$ and $\varepsilon = 0.097$, respectively. As it can be seen, the novel bound of Theorem 3 provides a much tighter guaranteed upper bound for the probability of joint-constraint violation compared to $\widetilde{\varepsilon}$, while not requiring agents to have access to the same set of scenarios. Other runs of the example, with new observations extracted, always gave an estimate of the joint-constraint violation probability smaller than 0.09, as it was expected given the high-confidence $1 - 10^{-5}$ with which the bound is guaranteed.

V. CONVERGENCE ANALYSIS AND PROOF OF THEOREM 1

A. Preparatory Results

We establish several relations between the difference of the agent estimates from certain average quantities. At the end of this section, we provide a summability result that is fundamental for the proof of Theorem 1 in Section V-B. Proofs for the results of this subsection V-A are provided in the appendix.

Let

$$v(k) = \frac{1}{m} \sum_{i=1}^m x_i(k), \text{ for all } k \geq 0. \tag{24}$$

Under Assumptions 1 and 3, (compactness of X_i can be relaxed to closedness), it is shown in [39, Lemma 2] that

$$\bar{v}(k) = \frac{\epsilon(k)}{\epsilon(k) + \rho} \bar{x} + \frac{\rho}{\epsilon(k) + \rho} v(k) \in \bigcap_{i=1}^m X_i,$$

$$\text{for all } k \geq 0 \tag{25}$$

where $\epsilon(k) = \sum_{i=1}^m \text{dist}(v(k), X_i)$, and $\bar{x} \in \mathbb{R}^n$, $\rho \in \mathbb{R}_+$ are as in Assumption 3. Note that unlike $x_i(k)$ and $v(k)$, which do not necessarily belong to $\bigcap_{i=1}^m X_i$, for $\bar{v}(k)$ this is always the case, thus providing a feasible solution of $\mathcal{P}$.

For each $i = 1, \ldots, m$, denote by

$$e_i(k+1) = x_i(k+1) - z_i(k), \text{ for all } k \geq 0 \tag{26}$$

the error between the values computed at steps 7 and 8 of Algorithm 1, i.e., the difference of the weighted average $z_i(k)$ computed by agent i at time k from its local update $x_i(k+1)$.

1) Error Relations: We provide some intermediate results that form the basis of the subsequent summability result.

Lemma 1: Consider Assumptions 1 and 3. For all $k \geq 0$

$$\sum_{i=1}^m \|x_i(k) - \bar{v}(k)\| \leq \mu \sum_{i=1}^m \|x_i(k) - v(k)\| \tag{27}$$

where $\mu = (2/\rho)mD + 1$, with D as given in Section II-C.

From step 7 of Algorithm 1, we have that for all $k \geq 0$, for all $i = 1, \ldots, m$

$$x_i(k+1) = \sum_{j=1}^m a_j^i(k)x_j(k) + x_i(k+1) - z_i(k)$$

$$= \sum_{j=1}^m a_j^i(k)x_j(k) + e_i(k+1) \tag{28}$$

where the last equality follows from (26).

Following [28], for each $k \geq 0$ consider a matrix $A(k) \in \mathbb{R}_+^{m \times m}$, such that $a_j^i(k)$ is the jth element of its ith column. For all k, s with $k \geq s$, let $\Phi(k, s) = A(s)A(s+1)\ldots A(k-1)A(k)$, with $\Phi(k, k) = A(k)$ for all $k \geq 0$. Denote by $\left[\Phi(k, s)\right]_j^i$ element j of column i of $\Phi(k, s)$. It is then shown in [28] that, under Assumption 5, $\Phi(k, s)$ is doubly stochastic. Similarly to [28], by propagating (28) in time, it can be shown that for all $k > s$ (the inequality is strict for convenience of the subsequent derivations), for all $i = 1, \ldots, m$

$$x_i(k+1) = \sum_{j=1}^m \left[\Phi(k, s)\right]_j^i x_j(s)$$

$$+ \sum_{r=s}^{k-1} \sum_{j=1}^m \left[\Phi(k, r+1)\right]_j^i e_j(r+1) + e_i(k+1). \tag{29}$$

For all $k > s$, the last statement, together with (24) and the fact that $\Phi(k, s)$ is a doubly stochastic matrix, leads to

$$v(k+1) = \frac{1}{m} \sum_{j=1}^m x_j(s)$$

$$+ \frac{1}{m} \sum_{r=s}^{k-1} \sum_{j=1}^m e_j(r+1) + \frac{1}{m} \sum_{j=1}^m e_i(k+1). \tag{30}$$

We then have the following lemma, which relates $\|x_i(k+1) - v(k+1)\|$ to $\|e_i(k+1)\|$, $i = 1, \ldots, m$. Its proof follows from [39, Lemma 8].

Lemma 2: Consider Assumptions 2 and 5. For all k, s with $s \geq 0$, $k > s$, and for all $i = 1, \ldots, m$

$$\|x_i(k+1) - v(k+1)\| \leq \lambda q^{k-s} \sum_{j=1}^{m} \|x_j(s)\|$$

$$+ \sum_{r=s}^{k-1} \lambda q^{k-r-1} \sum_{j=1}^{m} \|e_j(r+1)\|$$

$$+ \|e_i(k+1)\| + \frac{1}{m} \sum_{j=1}^{m} \|e_j(k+1)\| \qquad (31)$$

where $\lambda = 2(1 + \eta^{-(m-1)T})/(1 - \eta^{(m-1)T}) \in \mathbb{R}_+$ and $q = (1 - \eta^{(m-1)T})^{\frac{1}{(m-1)T}} \in (0, 1)$.

2) Summability Relation: Let $N \in \mathbb{N}_+$ and consider

$$2\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\| \qquad (32)$$

where $\bar{L} = \max_{i=1,\ldots,m} L_i$ with L_i defined according to (5). We will show that (32) has an interesting relation with $\sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2$ and will come back to it often in the next section to establish certain summability results.

Consider Lemma 1 with $k + 1$ in place of k and Lemma 2, summing both sides of (31) with respect to $i = 1, \ldots, m$ and setting $s = 0$. After some algebraic manipulations and index changes, we have that

$$2\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\|$$

$$\leq 2m\mu\lambda\bar{L} \sum_{k=1}^{N} c(k) q^k \sum_{i=1}^{m} \|x_i(0)\|$$

$$+ 2m\mu\lambda\bar{L} \sum_{k=1}^{N} \sum_{r=0}^{k-1} c(k) q^{k-r-1} \sum_{i=1}^{m} \|e_i(r+1)\|$$

$$+ 4\mu\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|e_i(k+1)\|. \qquad (33)$$

Lemma 3: Consider Assumptions 1–5. Fix any $\alpha_1 \in (0, 1)$, and consider (24)–(26). We then have that for any $N \in \mathbb{N}_+$

$$2\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\|$$

$$< \alpha_1 \sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2 + \alpha_2 \sum_{k=1}^{N} c(k)^2 + \alpha_3 \qquad (34)$$

where

$$\alpha_2 = \frac{2}{\alpha_1} m\mu^2\bar{L}^2 \left(m^2\lambda^2 \frac{1}{(1-q)^2} + 4\right),$$

$$\alpha_3 = 2m^3\mu^2\lambda^2\bar{L}^2 c(0)^2 \frac{1}{\alpha_1(1-q)^2}$$

$$+ 2m^2\mu\lambda\bar{L}Dc(1)\frac{q}{1-q} + 2\alpha_1 mD^2. \qquad (35)$$

B. Algorithm Analysis

1) Error Convergence: We prove convergence properties for the error in (26), which are instrumental to the proof of Theorem 1. We use the following result, which is proven in [7, Lemma 4.1, p. 257] for the case where the constraint sets are polyhedral. As mentioned in [7, p. 662], the assertion of the lemma remains valid also in the case of general convex constraint sets. For the latter, we refer the reader to [56], and to [57, Lemma 9] for a recent use of the lemma in case of convex constraint sets.

Lemma 4 (Lemma 4.1 in [7], p. 257): If $y^* = \arg\min_{y \in Y} J_1(y) + J_2(y)$ (assuming uniqueness of the minimizer), where $Y \subseteq \mathbb{R}^n$ is a closed, convex set, $J_1(\cdot), J_2(\cdot) : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $J_2(\cdot)$ is continuously differentiable, then $y^* = \arg\min_{y \in Y} J_1(y) + \nabla J_2(y^*)^\top y$, where $\nabla J_2(y^*)$ is the gradient of $J_2(y)$ with respect to y, evaluated at y^*.

Consider step 8 of Algorithm 1. By Assumptions 1 and 3, and the fact that $(1/(2c(k)))\|z_i(k) - x_i\|^2$ is continuously differentiable with respect to x_i, Lemma 4 can be applied to the problem with x_i, X_i in place of y, Y, respectively, $f_i(x_i)$ in place of $J_1(y)$ and $(1/(2c(k)))\|z_i(k) - x_i\|^2$ in place of $J_2(y)$. We have that

$$x_i(k+1) = \arg\min_{x_i \in X_i} f_i(x_i)$$

$$- \frac{1}{c(k)}(z_i(k) - x_i(k+1))^\top x_i \qquad (36)$$

where in the second term of (36), $-(1/c(k))(z_i(k) - x_i(k+1))$, is the gradient of $(1/(2c(k)))\|z_i(k) - x_i\|^2$ with respect to x_i, evaluated at $x_i(k+1)$. We then have the following lemma, which provides a useful relation between the consecutive algorithm iterates $x_i(k+1)$ and $x_i(k)$, and we will be using it extensively in the subsequent results. The subsequent proof strongly depends on the use of Lemma 4, and deviates from the proofs of the basic iterate relations in [39, Lemma 6] and [41, Lemma 5]; it is motivated by the proof of the alternating direction method of multipliers ([7, Proposition 4.2], [19, Appendix A]), and relies on our proximal minimization perspective.

Lemma 5: Consider Assumptions 1, 3, and 5. We then have that for any $k \in \mathbb{N}_+$, for any $x^* \in X^*$

$$2c(k) \sum_{i=1}^{m} f_i(\bar{v}(k+1)) + \sum_{i=1}^{m} \|e_i(k+1)\|^2$$

$$+ \sum_{i=1}^{m} \|x_i(k+1) - x^*\|^2$$

$$\leq 2c(k) \sum_{i=1}^{m} f_i(x^*) + \sum_{i=1}^{m} \|x_i(k) - x^*\|^2$$

$$+ 2\bar{L}c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\| \qquad (37)$$

where $e_i(k+1)$ is given as in (26).

Proof: By Lemma 4, (36) holds true. Since $x_i(k+1) \in X_i$ is the minimizer of the optimization problem that appears in the

right-hand side of (36), we have that

$$f_i(x_i(k+1)) - \frac{1}{c(k)}(z_i(k) - x_i(k+1))^\top x_i(k+1)$$

$$\leq f_i(x) - \frac{1}{c(k)}(z_i(k) - x_i(k+1))^\top x,$$

$$\text{for all } x \in X_i. \tag{38}$$

Since the last statement holds for any $x \in X_i$, it will also hold for any minimizer $x^* \in X^* \subseteq \bigcap_{i=1}^m X_i$ of problem $\mathcal{P}$ in (2).

We have that for any $x^* \in X^*$

$$-(z_i(k) - x_i(k+1))^\top (x_i(k+1) - x^*)$$

$$= \frac{1}{2}\|x_i(k+1) - z_i(k)\|^2 + \frac{1}{2}\|x_i(k+1) - x^*\|^2$$

$$- \frac{1}{2}\|z_i(k) - x^*\|^2. \tag{39}$$

By (38) and (39), we have that for any $x^* \in X^*$

$$f_i(x_i(k+1)) + \frac{1}{2c(k)}\|x_i(k+1) - z_i(k)\|^2$$

$$+ \frac{1}{2c(k)}\|x_i(k+1) - x^*\|^2$$

$$\leq f_i(x^*) + \frac{1}{2c(k)}\|z_i(k) - x^*\|^2$$

$$\leq f_i(x^*) + \frac{1}{2c(k)}\sum_{j=1}^m a_j^i(k)\|x_j(k) - x^*\|^2 \tag{40}$$

where the last inequality follows by the definition of $z_i(k)$ (see step 7 of Algorithm 1), the fact that, under Assumption 5, $\|\sum_{j=1}^m a_j^i(k)x_j(k) - x^*\|^2 = \|\sum_{j=1}^m a_j^i(k)(x_j(k) - x^*)\|^2$ and the convexity of $\|\cdot\|^2$.

Multiply both sides of (40) by $2c(k)$, sum with respect to $i = 1, \ldots, m$, and notice that for any $k \geq 0$, under the double stochasticity condition of Assumption 5, $\sum_{i=1}^m a_j^i(k) = 1$. We then have that

$$2c(k)\sum_{i=1}^m f_i(x_i(k+1)) + \sum_{i=1}^m \|x_i(k+1) - z_i(k)\|^2$$

$$+ \sum_{i=1}^m \|x_i(k+1) - x^*\|^2$$

$$\leq 2c(k)\sum_{i=1}^m f_i(x^*) + \sum_{i=1}^m \|x_i(k) - x^*\|^2. \tag{41}$$

Consider Assumption 3, and let $\bar{v}(k)$ be as in (25). Under Assumption 1, by (5), we have that $f_i(x_i(k+1)) \geq f_i(\bar{v}(k+1)) - L\|x_i(k+1) - \bar{v}(k+1)\|$, where $L = \max_{i=1,\ldots,m} L_i$. Recall also that $\|x_i(k+1) - z_i(k)\| = \|e_i(k+1)\|$ by (26). Therefore, for any $x^* \in X^*$, the last statements together with (41), lead to (37) and hence concludes the proof. ∎

Proposition 2: Consider Assumptions 1–5 and Algorithm 1. We have that
1) $\sum_{k=1}^\infty \sum_{i=1}^m \|e_i(k)\|^2 < \infty$;
2) $\lim_{k\to\infty} \|e_i(k)\| = 0$, for all $i = 1, \ldots, m$
where $e_i(k)$ is given as in (26).

Proof: By Lemma 5, (37) holds. Fix any $\alpha_1 \in (0, 1)$. Under the compactness part of Assumption 1 and Assumption 2, let α_2, α_3 as in (35), and consider (34). Sum then (37) with respect to $k = 1, \ldots, N$ for an arbitrary $N \in \mathbb{N}_+$, and upper bound the resulting last term in the right-hand side of (37) using (34). We then have that, for all $x^* \in X^*$

$$2\sum_{k=1}^N c(k) \sum_{i=1}^m \left(f_i(\bar{v}(k+1)) - f_i(x^*) \right)$$

$$+ (1 - \alpha_1)\sum_{k=1}^N \sum_{i=1}^m \|e_i(k+1)\|^2$$

$$\leq \sum_{i=1}^m \|x_i(1) - x^*\|^2 - \sum_{i=1}^m \|x_i(N+1) - x^*\|^2$$

$$+ \alpha_2 \sum_{k=1}^N c(k)^2 + \alpha_3. \tag{42}$$

Since $\bar{v}(k+1) \in \bigcap_{i=1}^m X_i$ for all $k \geq 0$, and x^* is a minimizer of $\mathcal{P}$, $2\sum_{k=1}^N c(k) \sum_{i=1}^m (f_i(\bar{v}(k+1)) - f_i(x^*)) \geq 0$. Moreover, $\sum_{i=1}^m \|x_i(N+1) - x^*\|^2 \geq 0$, hence these two terms can be dropped from (42). Therefore,

$$(1 - \alpha_1)\sum_{k=1}^N \sum_{i=1}^m \|e_i(k+1)\|^2$$

$$\leq \sum_{i=1}^m \|x_i(1) - x^*\|^2 + \alpha_2 \sum_{k=1}^N c(k)^2 + \alpha_3. \tag{43}$$

Let now $N \to \infty$. Since $\alpha_1 \in (0, 1)$ and $\sum_{i=1}^m \|x_i(1) - x^*\|^2 + \alpha_2 \sum_{k=1}^\infty c(k)^2 + \alpha_3 < \infty$, by the compactness part of Assumption 1 and Assumption 4, (43) implies that $\sum_{k=1}^\infty \sum_{i=1}^m \|e_i(k+1)\|^2 < \infty$, and hence also $\sum_{k=1}^\infty \sum_{i=1}^m \|e_i(k)\|^2 < \infty$, thus establishing the first part of the proposition. The second part directly follows from the fact that $\sum_{k=1}^\infty \sum_{i=1}^m \|e_i(k)\|^2 < \infty$ and $\|e_i(k)\|$ is a nonnegative quantity, thus concluding the proof. ∎

2) Average Tracking: We show that the agents' estimates $x_i(k), i = 1, \ldots, m$, track their arithmetic average $v(k)$, in the sense that $\lim_{k\to\infty} \|x_i(k) - v(k)\| = 0$ for all $i = 1, \ldots, m$. This is summarized in the following proposition. The proof follows from [39, Lemma 8], however, we include it also here for completeness.

Proposition 3: Consider Assumptions 1–5 and Algorithm 1. We have that

$$\lim_{k\to\infty} \|x_i(k) - v(k)\| = 0, \text{ for all } i = 1, \ldots, m \tag{44}$$

where $v(k)$ is given by (24).

Proof: Under Assumptions 1–5, by the second part of Proposition 2, we have that $\lim_{k\to\infty} \|e_i(k)\| = 0$, for all $i = 1, \ldots, m$. Then, for any $\epsilon > 0$, we can choose $s > 0$, such that $\|e_i(k)\| \leq \epsilon$ for all $k > s$, for all $i = 1, \ldots, m$.

By (31) of Lemma 2, we then have that for all $i = 1, \ldots, m$,

$$\|x_i(k+1) - v(k+1)\| \leq \lambda q^{k-s} \sum_{j=1}^{m} \|x_j(s)\|$$

$$+ m\lambda\epsilon \sum_{r=s}^{k-1} q^{k-r-1} + 2\epsilon$$

$$= \lambda q^{k-s} \sum_{j=1}^{m} \|x_j(s)\| + m\lambda\epsilon \sum_{t=0}^{k-s-1} q^t + 2\epsilon$$

$$< \lambda q^{k-s} \sum_{j=1}^{m} \|x_j(s)\| + m\lambda\epsilon \sum_{t=0}^{\infty} q^t + 2\epsilon$$

$$\leq m\lambda D q^{k-s} + m\lambda \frac{1}{1-q}\epsilon + 2\epsilon \qquad (45)$$

where the equality is due to a change of the summation limits, and the last inequality is due to the compactness part of Assumption 1 and the fact that $q \in (0, 1)$.

Taking limit superior in both sides of (45) as $k \to \infty$

$$\limsup_{k\to\infty} \|x_i(k+1) - v(k+1)\| \leq m\lambda \frac{1}{1-q}\epsilon + 2\epsilon. \qquad (46)$$

Note that taking the limit superior as $k \to \infty$ is well defined, since ϵ is assumed to be fixed, and hence also s. Notice also that the resulting quantity in the right-hand side of (46) no longer depends on s. Since $\epsilon > 0$ is arbitrary, relation (46) implies that $\lim_{k\to\infty} \|x_i(k+1) - v(k+1)\| = 0$, and hence $\lim_{k\to\infty} \|x_i(k) - v(k)\| = 0$, for all $i = 1, \ldots, m$, thus concluding the proof. ∎

3) Convergence and Optimality: In this last part of the section, we will provide a proof of Theorem 1. To achieve this, we will first show an intermediate convergence result. Notice that by the first part of Proposition 2 (under Assumptions 1–5), $\sum_{k=1}^{\infty} \sum_{i=1}^{m} \|e_i(k)\|^2 < \infty$. Letting then $N \to \infty$ in (34) leads to the following summability result, which states that

$$2\bar{L} \sum_{k=1}^{\infty} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\| < \infty. \qquad (47)$$

We can then show the following convergence result.

Theorem 4: Consider Assumptions 1–5 and Algorithm 1. We have that, for any minimizer $x^* \in X^*$, the sequence $\{\|x_i(k) - x^*\|\}_{k\geq 0}$ is convergent for all $i = 1, \ldots, m$.

Proof: By Lemma 5, (37) holds. Summing then (37) with respect to $k = M, \ldots, N$ for arbitrary $M, N \in \mathbb{N}_+$, we have that, for all $x^* \in X^*$

$$2 \sum_{k=M}^{N} c(k) \sum_{i=1}^{m} \left(f_i(\bar{v}(k+1)) - f_i(x^*) \right)$$

$$+ \sum_{k=M}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2 + \sum_{i=1}^{m} \|x_i(N+1) - x^*\|^2$$

$$\leq \sum_{i=1}^{m} \|x_i(M) - x^*\|^2$$

$$+ 2\bar{L} \sum_{k=M}^{N} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\|. \qquad (48)$$

As in the proof of Proposition 2, notice that since $\bar{v}(k+1) \in \bigcap_{i=1}^{m} X_i$ for all $k \geq 0$, and x^* is a minimizer of $\mathcal{P}$, $2\sum_{k=M}^{N} c(k) \sum_{i=1}^{m}(f_i(\bar{v}(k+1)) - f_i(x^*)) \geq 0$. Moreover, $\sum_{k=M}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2 \geq 0$, hence these two terms can be dropped from the left-hand side of (48). Therefore, by (48) we have that

$$\sum_{i=1}^{m} \|x_i(N+1) - x^*\|^2 \leq \sum_{i=1}^{m} \|x_i(M) - x^*\|^2$$

$$+ 2\bar{L} \sum_{k=M}^{N} c(k) \sum_{i=1}^{m} \|x_i(k+1) - \bar{v}(k+1)\|. \qquad (49)$$

Notice that, under Assumptions 1–5, the summability statement of (47) holds. Taking then in (49) the limit superior as $N \to \infty$ and the limit inferior as $M \to \infty$, we have that

$$\limsup_{N\to\infty} \sum_{i=1}^{m} \|x_i(N+1) - x^*\|^2$$

$$\leq \liminf_{M\to\infty} \sum_{i=1}^{m} \|x_i(M) - x^*\|^2. \qquad (50)$$

The last statement, together with the fact that the sequence $\{\sum_{i=1}^{m} \|x_i(k) - x^*\|\}_{k\geq 0}$ is bounded due to the compactness part of Assumption 1, implies that $\{\sum_{i=1}^{m} \|x_i(k) - x^*\|\}_{k\geq 0}$ converges for all $x^* \in X^*$.

Consider now $v(k) = \frac{1}{m} \sum_{i=1}^{m} x_i(k)$. We have that

$$\|v(k) - x^*\| = \|\frac{1}{m} \sum_{i=1}^{m} x_i(k) - x^*\| \leq \frac{1}{m} \sum_{i=1}^{m} \|x_i(k) - x^*\|. \qquad (51)$$

Moreover, from

$$\|x_i(k) - x^*\| \leq \|v(k) - x^*\| + \|x_i(k) - v(k)\| \qquad (52)$$

it also holds that $\|v(k) - x^*\| \geq \frac{1}{m} \sum_{i=1}^{m} \|x_i(k) - x^*\| - \frac{1}{m} \sum_{i=1}^{m} \|x_i(k) - v(k)\|$, which, together with (51), and since, for all $i = 1, \ldots, m$, $\lim_{k\to\infty} \|x_i(k) - v(k)\| = 0$ by Proposition 3, and $\{\sum_{i=1}^{m} \|x_i(k) - x^*\|\}_{k\geq 0}$ is convergent for all $x^* \in X^*$, gives $\|v(k) - x^*\|_{k\geq 0}$ is also convergent for any $x^* \in X^*$. From (52) and

$$\|v(k) - x^*\| - \|x_i(k) - v(k)\| \leq \|x_i(k) - x^*\|$$

the convergence of $\|v(k) - x^*\|$, along with $\lim_{k\to\infty} \|x_i(k) - v(k)\| = 0$, for all $i = 1, \ldots, m$, gives the statement of the theorem. ∎

We are now in a position to prove Theorem 1 of Section II-D, showing that there exists some minimizer $x^* \in X^*$ of $\mathcal{P}$, such that $\lim_{k\to\infty} \|x_i(k) - x^*\| = 0$, for all $i = 1, \ldots, m$, i.e., all agents reach consensus to a common minimizer of $\mathcal{P}$.

Proof of Theorem 1: By Lemma 5, (37) holds. Fix any $\alpha_1 \in (0, 1)$. Under Assumptions 1–5, let α_2, α_3 as in (35), and consider (34). As in the proof of Proposition 2, sum (37) with respect to $k = 1, \ldots, N$ for an arbitrary $N \in \mathbb{N}_+$, and upper bound the resulting last term in the right-hand side of (37) using (34). We then have that, for all $x^* \in X^*$, (42) holds.

Since $\sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2 \geq 0$ and $\sum_{i=1}^{m} \|x_i(N+1) - x^*\|^2 \geq 0$, we can drop the two terms in the left-hand side

of (42). Therefore, by (42), we have that

$$2\sum_{k=1}^{N} c(k) \sum_{i=1}^{m} (f_i(\bar{v}(k+1)) - f_i(x^*))$$

$$\leq \sum_{i=1}^{m} \|x_i(1) - x^*\|^2 + \alpha_2 \sum_{k=1}^{N} c(k)^2 + \alpha_3. \quad (53)$$

Let now $N \to \infty$. Notice that, by the compactness part of Assumption 1 and Assumption 4, $\sum_{i=1}^{m} \|x_i(1) - x^*\|^2 + \alpha_2 \sum_{k=1}^{\infty} c(k)^2 + \alpha_3 < \infty$. Therefore, $2\sum_{k=1}^{\infty} c(k) \sum_{i=1}^{m} (f_i(\bar{v}(k+1)) - f_i(x^*)) < \infty$; however, $\sum_{k=0}^{\infty} c(k) = \infty$, by Assumption 4. Hence

$$\liminf_{k \to \infty} \sum_{i=1}^{m} (f_i(\bar{v}(k+1)) - f_i(x^*)) = 0. \quad (54)$$

Due to the continuity of $f_i(\cdot), i = 1, \dots, m$, under the convexity requirement of Assumption 1, (54) implies that there exists some $\bar{x}^* \in X^*$, such that

$$\liminf_{k \to \infty} \|\bar{v}(k) - \bar{x}^*\| = 0. \quad (55)$$

In other words, $\{\|\bar{v}(k) - \bar{x}^*\|\}_{k \geq 0}$ converges to 0 across a subsequence.

By Proposition 3 and Lemma 1, we have that $\lim_{k \to \infty} \|x_i(k) - \bar{v}(k)\| = 0$, for all $i = 1, \dots, m$. Therefore, and since $\|x_i(k) - \bar{x}^*\| \leq \|\bar{v}(k) - \bar{x}^*\| + \|x_i(k) - \bar{v}(k)\|$, by (55) we have that, for all $i = 1, \dots, m$

$$\liminf_{k \to \infty} \|x_i(k) - \bar{x}^*\| = 0. \quad (56)$$

On the other hand, it was shown in Theorem 4 that, for all $i = 1, \dots, m$, $\{\|x_i(k) - x^*\|\}_{k \geq 0}$ converges for all $x^* \in X^*$, and hence also for $\bar{x}^*$. Hence, it must be $\lim_{k \to \infty} \|x_i(k) - \bar{x}^*\| = 0$, for all $i = 1, \dots, m$, which concludes the proof. ∎

Note that a direct byproduct of Proposition 3, Theorem 1, and Lemma 1, is that there exists $x^* \in X^*$, such that $\lim_{k \to \infty} \|x_i(k) - x^*\| = \lim_{k \to \infty} \|v(k) - x^*\| = \lim_{k \to \infty} \|\bar{v}(k) - x^*\| = 0$, for all $i = 1, \dots, m$.

VI. Conclusion

In this paper, a unifying framework for distributed convex optimization over time-varying networks, in the presence of constraints and uncertainty is provided. We constructed an iterative, proximal minimization-based algorithm, and analyzed its convergence and optimality properties. To deal with the case where the agents' constraint sets are affected by a possibly common uncertainty vector, a scenario-based methodology was adopted, allowing agents to use a different set of uncertainty scenarios.

Current work concentrates on three main directions:

1) Investigating the convergence rate properties of the developed algorithm, and the potential of an asynchronous implementation.
2) Developing rolling horizon implementations, extending the work of [58] to the case where constraints are also present.
3) Analyzing the quality of the scenario-based solutions, providing confidence intervals connecting the optimal values of $\mathcal{P}_{\bar{N}}$, $\mathcal{P}_N$ with the one of $\mathcal{P}_\delta$ by exploiting the results of [59] and [60].

From an application point of view, the main focus is on applying the proposed algorithm to the problem of energy efficient control of a building network [61].

Appendix

Proof of Lemma 1: We have that for all $k \geq 0$

$$\|x_i(k) - \bar{v}(k)\|$$

$$= \left\| \frac{\epsilon(k) + \rho}{\epsilon(k) + \rho} x_i(k) - \frac{\epsilon(k)}{\epsilon(k) + \rho} \bar{x} - \frac{\rho}{\epsilon(k) + \rho} v(k) \right\|$$

$$\leq \frac{1}{\epsilon(k) + \rho} \big(\epsilon(k)\|x_i(k) - \bar{x}\| + \rho\|x_i(k) - v(k)\| \big)$$

$$\leq \frac{1}{\rho} \big(\epsilon(k)\|x_i(k) - \bar{x}\| + \rho\|x_i(k) - v(k)\| \big) \quad (57)$$

where the last inequality is due to the fact that $\epsilon(k) \geq 0$.

By the definition of $\operatorname{dist}(\cdot, \cdot)$, and since $x_i(k) \in X_i$ for all $i = 1, \dots, m$, we have that for all $k \geq 0$

$$\epsilon(k) = \sum_{i=1}^{m} \operatorname{dist}(v(k), X_i) \leq \sum_{i=1}^{m} \|x_i(k) - v(k)\|. \quad (58)$$

By (57) and (58), we have that

$$\|x_i(k) - \bar{v}(k)\| \leq \frac{1}{\rho} \left(\sum_{i=1}^{m} \|x_i(k) - v(k)\| \right) \|x_i(k) - \bar{x}\|$$

$$+ \|x_i(k) - v(k)\|. \quad (59)$$

Summing both sides of (59) with respect to $i = 1, \dots, m$

$$\sum_{i=1}^{m} \|x_i(k) - \bar{v}(k)\|$$

$$\leq \frac{1}{\rho} \left(\sum_{i=1}^{m} \|x_i(k) - v(k)\| \right) \left(\sum_{i=1}^{m} \|x_i(k) - \bar{x}\| \right)$$

$$+ \sum_{i=1}^{m} \|x_i(k) - v(k)\|$$

$$\leq \left(\frac{2}{\rho} m D + 1 \right) \sum_{i=1}^{m} \|x_i(k) - v(k)\| \quad (60)$$

where the last inequality is since $\|x_i(k) - \bar{x}\| \leq \|x_i(k)\| + \|\bar{x}\| \leq 2D$ for all $i = 1, \dots, m$, [D as defined above (5)], by the compactness part of Assumption 1. ∎

Proof of Lemma 2: By (29) and (30), for all k, s with $s \geq 0$, $k > s$, and for all $i = 1, \dots, m$ we have that

$$\|x_i(k+1) - v(k+1)\| = \left\| \sum_{j=1}^{m} \left([\Phi(k,s)]_j^i - \frac{1}{m} \right) x_j(s) \right.$$

$$+ \sum_{r=s}^{k-1} \sum_{j=1}^{m} \left([\Phi(k, r+1)]_j^i - \frac{1}{m} \right) e_j(r+1)$$

$$\left. + e_i(k+1) - \frac{1}{m} \sum_{j=1}^{m} e_j(k+1) \right\|$$

$$\leq \sum_{j=1}^{m} \left| [\Phi(k,s)]_j^i - \frac{1}{m} \right| \|x_j(s)\|$$

$$+ \sum_{r=s}^{k-1} \sum_{j=1}^{m} \left| [\Phi(k,r+1)]_j^i - \frac{1}{m} \right| \|e_j(r+1)\|$$

$$+ \|e_i(k+1)\| + \frac{1}{m} \sum_{j=1}^{m} \|e_j(k+1)\|. \tag{61}$$

Under Assumptions 2 and 5, by [28, Lemma 4], for all k, s with $s \geq 0$, $k \geq s$, we have that

$$\left| [\Phi(k,s)]_j^i - \frac{1}{m} \right| \leq 2 \frac{1 + \eta^{-(m-1)T}}{1 - \eta^{(m-1)T}} \left(1 - \eta^{(m-1)T} \right)^{\frac{k-s}{(m-1)T}}. \tag{62}$$

Setting $\lambda = 2(1 + \eta^{-(m-1)T})/(1 - \eta^{(m-1)T})$ and $q = (1 - \eta^{(m-1)T})^{\frac{1}{(m-1)T}}$, (62) implies that $|[\Phi(k,s)]_j^i - \frac{1}{m}| \leq \lambda q^{k-s}$, for all $k \geq s$. Noticing that $q \in (0,1)$, since $\eta \in (0,1)$, (61) and (62) lead to (31), thus concluding the proof. ∎

Proof of Lemma 3: Fix any $N \in \mathbb{N}_+$ and, under Assumptions 1–5, consider (33). To show (34), we treat each of the three terms in the right-hand side of (33) separately.

Term 1: $2m\mu\lambda\bar{L} \sum_{k=1}^{N} c(k)q^k \sum_{i=1}^{m} \|x_i(0)\|$.

Due to the compactness part of Assumption 1, $\|x_i(0)\| \leq D$, for all $i = 1, \ldots, m$. Therefore, $\sum_{i=1}^{m} \|x_i(0)\| \leq mD$. The last statement together with the fact that, under Assumption 4, $c(k) \leq c(1)$, leads to

$$2m\mu\lambda\bar{L} \sum_{k=1}^{N} c(k)q^k \sum_{i=1}^{m} \|x_i(0)\| \leq 2m^2\mu\lambda\bar{L}Dc(1) \sum_{k=1}^{N} q^k$$

$$< \frac{2m^2\mu\lambda\bar{L}Dc(1)q}{1-q} \tag{63}$$

where the last step is due to the fact that, by Lemma 2, $q \in (0,1)$ and hence $\sum_{k=1}^{\infty} q^k = q \sum_{k=0}^{\infty} q^k = q/(1-q)$.

Term 2: $2m\mu\lambda\bar{L} \sum_{k=1}^{N} \sum_{r=0}^{k-1} c(k)q^{k-r-1} \sum_{i=1}^{m} \|e_i(r+1)\|$.

Fix any $\alpha_1 \in (0,1)$. We then have that

$$2m\mu\lambda\bar{L} \sum_{k=1}^{N} \sum_{r=0}^{k-1} c(k)q^{k-r-1} \sum_{i=1}^{m} \|e_i(r+1)\|$$

$$= \sum_{i=1}^{m} \sum_{k=1}^{N} \sum_{r=0}^{k-1} 2 \left(m\mu\lambda\bar{L} \sqrt{\frac{2}{\alpha_1(1-q)}} c(k) \right)$$

$$\left(\sqrt{\frac{\alpha_1(1-q)}{2}} \|e_i(r+1)\| \right) q^{k-r-1}$$

$$\leq \sum_{i=1}^{m} \sum_{k=1}^{N} \sum_{r=0}^{k-1} m^2\mu^2\lambda^2\bar{L}^2 \frac{2}{\alpha_1(1-q)} q^{k-r-1}c(k)^2$$

$$+ \sum_{i=1}^{m} \sum_{k=1}^{N} \sum_{r=0}^{k-1} \frac{\alpha_1(1-q)}{2} q^{k-r-1} \|e_i(r+1)\|^2 \tag{64}$$

where in the last step we used the fact that $2xy \leq x^2 + y^2$ for all $x, y \in \mathbb{R}$. We have that

$$\sum_{k=1}^{N} \sum_{r=0}^{k-1} q^{k-r-1}c(k)^2$$

$$\leq \sum_{k=1}^{N} \sum_{r=0}^{k-1} q^{k-r-1}c(r)^2 = \sum_{r=0}^{N-1} c(r)^2 \sum_{t=0}^{N-r-1} q^t$$

$$< \sum_{r=0}^{N-1} c(r)^2 \sum_{t=0}^{\infty} q^t = \sum_{k=0}^{N-1} \frac{1}{1-q} c(k)^2$$

$$< \frac{1}{1-q} c(0)^2 + \sum_{k=1}^{N} \frac{1}{1-q} c(k)^2 \tag{65}$$

where the first inequality is due to the fact that, under Assumption 4, $c(k) \leq c(r)$ since $k > r$. The first equality is due to series convolution, in the last equality, we performed an index change from r to k and the last inequality is included to introduce the desired summation limits.

Repeating the same derivation as in (65) with $\|e_i(r+1)\|^2$ in place of $c(r)^2$ leads to

$$\sum_{k=1}^{N} \sum_{r=0}^{k-1} q^{k-r-1} \|e_i(r+1)\|^2$$

$$< \frac{1}{1-q} \|e_i(1)\|^2 + \sum_{k=1}^{N} \frac{1}{1-q} \|e_i(k+1)\|^2$$

$$\leq \frac{4}{1-q} D^2 + \sum_{k=1}^{N} \frac{1}{1-q} \|e_i(k+1)\|^2 \tag{66}$$

where the last inequality is due to the fact that $\|e_i(1)\| \leq 2D$ under the compactness part of Assumption 1.

By (64), (65), and (66), and noticing that some terms are independent of i, we have that

$$2m\mu\lambda\bar{L} \sum_{k=1}^{N} \sum_{r=0}^{k-1} c(k)q^{k-r-1} \sum_{i=1}^{m} \|e_i(r+1)\|$$

$$< 2m^3\mu^2\lambda^2\bar{L}^2 \frac{1}{\alpha_1(1-q)^2} c(0)^2 + 2\alpha_1 mD^2$$

$$+ 2m^3\mu^2\lambda^2\bar{L}^2 \frac{1}{\alpha_1(1-q)^2} \sum_{k=1}^{N} c(k)^2$$

$$+ \frac{\alpha_1}{2} \sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2. \tag{67}$$

Term 3: $4\mu\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|e_i(k+1)\|$.

We have that

$$4\mu\bar{L} \sum_{k=1}^{N} c(k) \sum_{i=1}^{m} \|e_i(k+1)\|$$

$$= \sum_{k=1}^{N} \sum_{i=1}^{m} 2 \left(2\sqrt{\frac{2}{\alpha_1}} \mu\bar{L}c(k) \right) \left(\sqrt{\frac{\alpha_1}{2}} \|e_i(k+1)\| \right)$$

$$\leq \sum_{k=1}^{N} \sum_{i=1}^{m} \frac{8}{\alpha_1} \mu^2 \bar{L}^2 c(k)^2 + \frac{\alpha_1}{2} \sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2$$

$$= \frac{8}{\alpha_1} m \mu^2 \bar{L}^2 \sum_{k=1}^{N} c(k)^2 + \frac{\alpha_1}{2} \sum_{k=1}^{N} \sum_{i=1}^{m} \|e_i(k+1)\|^2. \quad (68)$$

where for the first inequality we follow the same reasoning with the last step of (64), and the second equality is since the first term of the first inequality is independent of i.

We are now in a position to show (34). Substituting (63), (67), and (68) in (33), and setting α_2, α_3 according to (35), leads to (34) (the inequality is strict since the inequalities in (63), (67) are also strict), thus concluding the proof. ∎

REFERENCES

[1] K. Margellos, A. Falsone, S. Garatti, and M. Prandini, "Proximal minimization based distributed convex optimization," in *Proc. Amer. Control Conf.*, 2016, pp. 2466–2471.

[2] S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Trans. Automat. Control*, vol. 60, no. 4, pp. 966–981, Apr. 2015.

[3] Y. Zhang and G. Giannakis, "Distributed stochastic market clearing with high-penetration wind power," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 895–906, Mar. 2016.

[4] G. Mateos and G. Giannakis, "Distributed recursive least-squares: Stability and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3740–3754, Jul. 2012.

[5] B. Baingana, G. Mateos, and G. Giannakis, "Proximal-gradient algorithms for tracking cascades over social networks," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 4, pp. 563–575, Aug. 2014.

[6] S. Martínez, F. Bullo, J. Cortés, and E. Frazzoli, "On synchronous robotic networks - Part I: Models, tasks, and complexity," *IEEE Trans. Automat. Control*, vol. 52, no. 12, pp. 2199–2213, Dec. 2007.

[7] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA, USA: Athena Scientific 1989 (republished in 1997).

[8] D. Bertsekas, "Incremental proximal methods for large scale convex optimization," *Math. Programm.*, vol. 129, no. 2, pp. 163–195, 2011.

[9] G. Calafiore and M. Campi, "The scenario approach to robust control design," *IEEE Trans. Automat. Control*, vol. 51, no. 5, pp. 742–753, May 2006.

[10] M. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.

[11] M. Campi, S. Garatti, and M. Prandini, "The scenario approach for systems and control design," *Annu. Rev. Control*, vol. 33, no. 2, pp. 149–157, 2009.

[12] M. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality," *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.

[13] S. Garatti and M. Campi, "Modulating robustness in control design: Principles and algorithms," *IEEE Control Syst.*, vol. 33, no. 2, pp. 36–51, Apr. 2013.

[14] K. Margellos, M. Prandini, and J. Lygeros, "On the connection between compression learning and scenario based single-stage and cascading optimization problems," *IEEE Trans. Automat. Control*, vol. 60, no. 10, pp. 2716–2721, Oct. 2015.

[15] M. Campi, S. Garatti, and F. Ramponi, "Non-convex scenario optimization with application to system identification," in *Proc. 54th IEEE Conf. Decision Control*, Osaka, Japan, 2015, pp. 4023–4028.

[16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[17] J. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, MIT, Cambridge, MA, USA, 1984.

[18] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automat. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

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

[21] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to Nash equilibria in constrained deterministic mean field control," *IEEE Trans. Automat. Control*, vol. 61, no. 11, pp. 3315–3329, Nov. 2016.

[22] T. Vicsek, A. Czirok, E. Ben-Jacob, I. Cohen, and O. Shochet, "Novel type of phase transitions in a system of self-driven particles," *Phys. Rev. Lett.*, vol. 75, no. 6, pp. 1226–1229, 1995.

[23] A. Jadbabaie, J. Lin, and S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automat. Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.

[24] R. Olfati-Saber and R. Murray, "Consensus problems in networks of agents with switching topology and time delays," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

[25] M. Cao, D. Spielman, and S. Morse, "A lower bound on convergence of a distributed network convergence algorithm," in *Proc. 44th IEEE Conf. Decision Control Eur. Control Conf.*, Seville, Spain, 2005, pp. 2356–2361.

[26] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[27] B. Johansson, M. Rabi, and M. Johansson, "A simple peer-to-peer algorithm for distributed optimization in sensor networks," in *Proc. 46th IEEE Conf Decision Control*, New Orleans, LA, USA, 2007, pp. 4705–4710.

[28] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Automat. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[29] A. Olshevsky and J. Tsitsiklis, "Convergence speed in distributed convergence and averaging," *SIAM Rev.*, vol. 53, no. 4, pp. 747–772, 2011.

[30] A. Nedić and A. Olshevsky, "Distributed optimization over timevarying directed graphs," *IEEE Trans. Automat. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

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

[32] B. Johansson, T. Keviczky, M. Johansson, and K. Johansson, "An interiorpoint Lagrangian decomposition method for separable convex optimization," in *Proc. 47th IEEE Conf. Decision Control*, Cancun, Mexico, 2008, pp. 4185–4190.

[33] I. Necoara and J. Suykens, "An interior-point Lagrangian decomposition method for separable convex optimization," *J. Optim. Theory Appl.*, vol. 143, no. 3, pp. 567–588, 2009.

[34] Y. Pu, M. Zeilinger, and C. Jones, "Quantization design for distributed optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 5, pp. 2107–2120, 2017.

[35] E. Wei and A. Ozdaglar, "On the $1/k$ convergence if asynchronous alternating direction method of multipliers over networks," Tech. Rep. MIT, shorter version at *Proc. IEEE Global Conf. Signal Inf. Process.*, pp. 1–4, 2013.

[36] P. Bianchi, W. Hachem, and F. Iutzeler, "A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization," *IEEE Trans. Automat. Control*, vol. 61, no. 10, pp. 2947–2957, Oct. 2016.

[37] L. Carlone, V. Srivastava, F. Bullo, and G. Calafiore, "Distributed random convex programming via constraints consensus," *SIAM J. Control Optim.*, vol. 52, no. 1, pp. 629–662, 2014.

[38] M. Bürger, G. Notarstefano, and F. Allgöwer, "A polyhedral approximation framework for convex and robust distributed optimization," *IEEE Trans. Automat. Control*, vol. 59, no. 2, pp. 3740–3754, Feb. 2014.

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

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

[41] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE J. Sel. Top. Signal Process.*, vol. 7, no. 2, pp. 221–229, Apr. 2013.

[42] Z. Towfic and A. Sayed, "Adaptive penalty-based distributed stochastic convex optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3924–3938, Aug. 2014.

[43] S. Lee and A. Nedić, "Asynchronous gossip-based random projection algorithms over networks," *IEEE Trans. Automat. Control*, vol. 61, no. 4, pp. 953–968, Apr. 2016.

[44] J. Cortés and F. Bullo, "Coordination and geometric optimization via distributed dynamical systems," *SIAM J. Control Optim.*, vol. 44, no. 5, pp. 1543–1574, 2005.

[45] J. Cortés and F. Bullo, "Nonsmooth coordination and geometric optimization via distributed dynamical systems," *SIAM Rev.*, vol. 51, no. 1, pp. 163–189, 2009.

[46] J. Wang and N. Elia, "A control perspective for centralized and distributed convex optimization," in *Proc. 50th IEEE Conf. Decision Control Eur. Control Conf.*, Orlando, FL, USA, 2011, pp. 3800–3805.

[47] S. Kia, "A distributed dynamical solver for an optimal resource allocation problem over networked systems," in *Proc. 54th IEEE Conf. Decision Control*, Osaka, Japan, 2015, pp. 7482–7487.

[48] S. Kia, J. Cortés, and S. Martínez, "Dynamic average consensus under limited control authority and privacy requirements," *Int. J. Robust Nonlinear Control*, vol. 25, no. 13, pp. 1946–1966, 2015.

[49] A. Cherukuri and J. Cortés, "Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment," *Automatica*, vol. 74, pp. 183–193, 2016.

[50] D. Bertsekas, "Multiplier methods: A survey," *Automatica*, vol. 12, no. 2, pp. 133–145, 1976.

[51] A. Nedíc, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Trans. Automat. Control*, vol. 54, no. 11, pp. 2506–2517, Nov. 2009.

[52] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pac. J. Math.*, vol. 21, no. 2, pp. 343–348, 1967.

[53] G. Schildbach, L. Fagiano, and M. Morari, "Randomized solutions to convex programs with multiple chance constraints," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2479–2501, 2013.

[54] N. Kariotoglou, K. Margellos, and J. Lygeros, "On the computational complexity and generalization properties of multi-stage and stage-wise coupled scenario programs," *Syst. Control Lett.*, vol. 94, pp. 63–69, 2016.

[55] M. Campi and A. Caré, "Random convex programs with $L1$-regularization: sparsity and generalization," *SIAM J. Control Optim.*, vol. 51, no. 5, pp. 3532–3557, 2013.

[56] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[57] T. Chang, A. Nedíc, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Trans. Automat. Control*, vol. 59, no. 6, pp. 1524–1538, Jun. 2014.

[58] J. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Trans. Automat. Control*, vol. 29, no. 1, pp. 42–50, Jan. 1984.

[59] T. Kanamori and A. Takeda, "Worst-case violation of sampled convex programs for optimization under uncertainty," *J. Optim. Theory Appl.*, vol. 152, no. 1, pp. 171–197, 2012.

[60] P. Mohajerin, T. Sutter, and J. Lygeros, "Performance bounds for the scenario approach and an extension to a class of non-convex programs," *IEEE Trans. Automat. Control*, vol. 60, no. 1, pp. 46–58, Jan. 2015.

[61] D. Ioli, A. Falsone, and M. Prandini, "Optimal energy management of a building cooling system with thermal storage: A convex formulation," in *Proc. 9th IFAC Symp. Adv. Control Chem. Processes*, Whistler, BC, Canada, 2015, pp. 1151–1156.

Alessandro Falsone was born in Milan, Italy, in 1989. He received the Bachelor of Science degree and the Master of Science (*cum laude*) degree both in automation and control engineering from Politecnico di Milano, Milano, Italy, in 2011 and 2013, respectively. Since November 2014, he has been working toward the Ph.D. degree in the System and Control Division, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano.

From November 2013 to October 2014, he worked as a Research Assistant in the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His current research interests include nonlinear model identification, distributed optimization and control, randomized algorithms, and stochastic hybrid systems.

Simone Garatti received the Laurea degree in computer science engineering and the Ph.D. degree in information technology engineering both from the Politecnico di Milano, Milano, Italy, in 2000 and 2004, respectively.

He is an Associate Professor in the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy. He has been a Visiting Scholar at the Lund University of Technology, Lund, Sweden, at the University of California San Diego, San Diego, CA, USA, and at the Massachusetts Institute of Technology and the Northeastern University, Boston, MA, USA. His research interests include system identification and model quality assessment, identification of interval predictor models, and stochastic and data-based optimization for problems in systems and control.

Dr. Garatti is a member of the IEEE Technical Committee on Computational Aspects of Control System Design and of the IFAC Technical Committee on Modeling, Identification and Signal Processing.

Maria Prandini received the Laurea degree in electrical engineering (summa cum laude) from Politecnico di Milano, Milano, Italy, in 1994, and the Ph.D. degree in information technology from Università degli Studi di Brescia, Brescia, Italy, in 1998.

From 1998 to 2000, she was a Postdoctoral Researcher in the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley. She also held visiting positions at Delft University of Technology (1998), Cambridge University (2000), University of California at Berkeley (2005), and Swiss Federal Institute of Technology Zurich (2006). In 2002, she started as an Assistant Professor in systems and control at Politecnico di Milano, where she is currently an Associate Professor. Her research interests include stochastic hybrid systems, randomized algorithms, constrained control, system abstraction and verification, nonlinear identification, distributed optimization, and the application of control theory to air traffic management and energy systems.

Dr. Prandini serves on the editorial board of Cyber-Physical Systems, and previously of *European Journal of Control*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY and *Nonlinear Analysis: Hybrid Systems*. From 2013 to 2105, she has been editor for Electronic Publications of the IEEE Control Systems Society. She has been member of the IEEE CSS Board of Governors for a triennium ending 2017, and she has been the CSS Vice-President for Conference Activities in 2016 and 2017.

Kostas Margellos received the Diploma in electrical engineering from the University of Patras, Patras, Greece, in 2008, and the Ph.D. degree in control engineering from ETH Zurich, Zurich, Switzerland, in 2012.

He spent 2013, 2014 and 2015 as a Postdoctoral Researcher at ETH Zurich, UC Berkeley, and Politecnico di Milano, respectively. In 2016, he joined the Control Group, Department of Engineering Science, University of Oxford, Oxford, U.K., where he is currently an Associate Professor. He is also a Lecturer at Worcester College, Oxford. His research interests include optimization and control of complex uncertain systems, with applications to generation and load side control for power networks.