Annual Reviews in Control 56 (2023) 100925

Contents lists available at ScienceDirect

Annual Reviews in Control

journal homepage: www.elsevier.com/locate/arcontrol

Full length article

Probabilistic feasibility in data-driven multi-agent non-convex optimization ☆

Lucrezia Manieri, Alessandro Falsone, Maria Prandini *

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Giuseppe Ponzio, 34/35, 20133, Milano, Italy

ARTICLE INFO

Keywords:
Data-driven optimization
Uncertain systems
Multi-agent non-convex optimization
Statistical learning theory

ABSTRACT

In this paper, we focus on the optimal operation of a multi-agent system affected by uncertainty. In particular, we consider a cooperative setting where agents jointly optimize a performance index compatibly with individual constraints on their discrete and continuous decision variables and with coupling global constraints. We assume that individual constraints are affected by uncertainty, which is known to each agent via a private set of data that cannot be shared with others. Exploiting tools from statistical learning theory, we provide data-based probabilistic feasibility guarantees for a (possibly sub-optimal) solution of the multi-agent problem that is obtained via a decentralized/distributed scheme that preserves the privacy of the local information. The generalization properties of the data-based solution are shown to depend on the size of each local dataset and on the complexity of the uncertain individual constraint sets. Explicit bounds are derived in the case of linear individual constraints. A comparative analysis with the cases of a common dataset and of local uncertainties that are independent is performed.

1. Introduction

This paper addresses the optimal operation of a system composed of multiple interacting agents that are affected by uncertainty. We consider a cooperative framework where agents aim at optimizing some performance index (either a loss to be minimized or a reward to be maximized) compatibly with their local operational limitations and some global constraints, typically related to shared resources, that are coupling their decisions. Problems in this form arise naturally in many engineering applications, from industrial multi-type energy generation plants (Shen, Zhao, Du, Zhong, & Qian, 2019) to optimal collaborative operation of microgrids (Zhang, Li, Wang, & Feng, 2018), and cooperative adaptive cruise control of modern vehicle automation (Bevly et al., 2016). In many problems, the agents are characterized by both continuous and discrete components, which causes the resulting optimization problem to be a *Mixed-Integer Program* (MIP). If the cost function and the constraints are all linear in the decision variables, then the MIP problem becomes a *Mixed-Integer Linear Program* (MILP). This is the case for the class of Mixed Logical Dynamical systems, (Bemporad & Morari, 1999) with a linear cost function.

The computational effort required to solve a MIP to optimality grows exponentially in the number of discrete variables and becomes prohibitive for large-scale systems. In many cases, even determining a feasible solution may be challenging due to the inherent combinatorial complexity of the problem. Multi-agent MILPs, however, exhibit a partially separable structure that can be leveraged to decompose the problem into lower-dimensional MILPs and distribute the computational load among the agents according to a distributed or decentralized paradigm. Recent works (Falsone, Margellos, & Prandini, 2018, 2019; Manieri, Falsone, & Prandini, 2023a, 2023b) propose solution-seeking algorithms for constraint-coupled multi-agent MILPs that combine dual decomposition (to recover separability) and constraint tightening (to guarantee the feasibility of the computed solution). In particular, they lift the coupling constraints (that prevent separability) to the cost function by means of a set of non-negative weights (the Lagrange multipliers), and then compute a solution of this *relaxed* problem by means of a decentralized or distributed iterative scheme where each agent solves a lower-dimensional MILP at each iteration. Satisfaction of the coupling constraint is enforced by adding a fictitious tightening, adjusted adaptively throughout the iterations. The application to prosumers aggregation for providing balancing services to the grid in La Bella, Falsone, Ioli, Prandini, and Scattolini (2021) showcases the scalability of the decomposition approach whose theoretical guarantees are proven in Manieri et al. (2023a).

In Camisa, Notarnicola, and Notarstefano (2022), primal decomposition is used to derive a master-sub-problem architecture where the master problem handles the coupling by assigning a portion of the shared resources to each agent, whilst each agent solves a sub-problem to retrieve the best solution compatible with such resource allocation.

☆ This paper is supported by PNRR-PE-AI FAIR project funded by the NextGeneration EU program and the PRIN project CUP:D53D23001450006.
* Corresponding author.
E-mail addresses: lucrezia.manieri@polimi.it (L. Manieri), alessandro.falsone@polimi.it (A. Falsone), maria.prandini@polimi.it (M. Prandini).

https://doi.org/10.1016/j.arcontrol.2023.100925
Received 17 July 2023; Accepted 5 November 2023
Available online 18 November 2023

Even in this case, a tightening of the shared resource is needed to enforce the feasibility of the coupling constraints.

All these approaches focus on deterministic MILPs and are not applicable in contexts where the agents are operating in uncertain environments affecting their operational constraints.

The presence of uncertainty introduces an additional challenge and calls for a suitable notion of solution. According to the *robust* paradigm (Ben-Tal & Nemirovski, 1998; Kwakernaak, 1993; Li et al., 2020), feasibility is required to hold for every and each uncertainty realization (even those that are unlikely to occur), thus typically resulting in a conservative solution. Also, knowledge of the set of values that the uncertainty can assume is required, and the resulting semi-infinite optimization problem is often hard to solve.

On the other hand, the *probabilistic* paradigm calls for a characterization of the uncertainty through a probability distribution and requires feasibility to hold for most of the uncertainty realizations except for a set of predefined (small) probability. The resulting chance-constrained optimization problem is often computationally challenging even when the uncertainty is fully characterized (Ben-Tal & Nemirovski, 1997), which is rarely the case in practice.

The lack of a characterization of the uncertainty affecting the problem can be addressed by collecting data and formulating a data-based optimization problem, where constraints are imposed only on the seen uncertainty instances. This, however, adds another layer of uncertainty since data-based solutions inevitably depend on the available samples and the generalization properties to unseen uncertainty instances can then hold only with a certain probability.

For uncertain convex optimization problems, the *scenario approach* (Campi, Garatti, & Prandini, 2009) provides a-priori distribution-free probabilistic guarantees in terms of a lower bound on the number of samples (scenarios) needed to ensure – with tunable *high* confidence – that a given data-based solution violates the uncertain constraints with small probability. Data-based multi-agent convex problems with a partially decoupled structure (either constraint-coupled or decision-coupled) were considered in Falsone, Margellos, Prandini, and Garatti (2020), by extending the scenario theory to decentralized and distributed schemes preserving privacy of the local information.

As for the non-convex case, a-priori feasibility guarantees are obtained based on the scenario approach in Esfahani, Sutter, and Lygeros (2014) for MILPs and in Calafiore, Lyons, and Fagiano (2012) for MIPs with constraints with a convex but not necessarily linear continuous counterpart. More recently, a-posteriori (distribution-free) probabilistic guarantees based on the available data and the resulting support constraints have been derived for the solution to general non-convex problems (Campi, Garatti, & Ramponi, 2018). The above results are, however, not directly applicable to a multi-agent privacy-preserving non-convex context for two main reasons: (i) agents are not willing to share with the others their data, so that in the scenario problem formulation local constraints cannot be imposed on the same scenarios of the uncertainty affecting the multi-agent system, and (ii) solutions obtained via state-of-the-art decentralized/distributed algorithms for multi-agent non-convex problems like those in Camisa et al. (2022), Falsone et al. (2018, 2019), Manieri et al. (2023a) are only guaranteed to be feasible but not necessarily optimal.

In this paper, we exploit tools from statistical learning theory (Alamo, Tempo, & Camacho, 2009) and extend the results in Falsone, Margellos, Prandini, and Garatti (2020) to non-convex multi-agent optimization problems with uncertain linear local constraints. We derive probabilistic feasibility guarantees that preserve the privacy of the local information of the agents, including that on the uncertainty, while allowing for scalability in the number of agents. Differently from the work in Falsone, Molinari, and Prandini (2020) on uncertain multi-agent MILPs, we do not require the agents to use the same uncertainty realizations and provide improved bounds when the agents are subject to independent uncertainty sources.

Note that in a parallel stream of work, Pantazis, Fele, and Margellos (2022) provides probabilistic feasibility guarantees to data-based multi-agent problems with convex – not necessarily linear – local constraints. However, guarantees are obtained a-posteriori and require a procedure to determine the minimal support samples for the feasibility region, which is computationally challenging in general. In the same vein, a-posteriori feasibility guarantees are provided in Pantazis, Fele, and Margellos (2021) for uncertain polytopic constraints in a non-cooperative multi-agent setting. Though limited to linear local constraints, our guarantees are instead a-priori and do not require any assessment on the support samples of the feasibility region.

The remainder of the paper is structured as follows. The addressed problem is formally stated in Section 2, where we highlight its main challenges. We derive the main result of the paper in Section 3, where we provide probabilistic feasibility guarantees for the case where agents have private datasets that cannot be shared with the others and re-derive the results in Falsone, Molinari, and Prandini (2020) for the common dataset case. The two results are then compared in Section 4. Section 5 concludes the paper.

Notation. We denote with $\mathbb{Z}$ the set of integer numbers and with $\mathbb{R}$ the set of real numbers. For a vector v, $v^\top$ denotes its transpose and $[v]_r$ its rth component. Symbols $\wedge$ and $\vee$ denote the AND and OR logical operators, respectively. The symbol $\emptyset$ denotes the empty set, while $\cap$, $\cup$, and $\setminus$ denote set intersection, union, and difference, respectively. Given a probability measure $\mathbb{P}$ over a set Δ, we denote with $\mathbb{P}^N$ the product probability measure to describe the joint distribution of N independent variables, each one taking value on Δ according to $\mathbb{P}$.

2. Problem setting and data-driven formulation

Let us consider a multi-agent system composed of m cooperating agents, where each agent i has $n_{c,i}$ continuous optimization variables and $n_{d,i}$ integer decision variables, collected in a decision vector $x_i \in \mathbb{X}_i = \mathbb{R}^{n_{c,i}} \times \mathbb{Z}^{n_{d,i}}$ with $n_i = n_{c,i} + n_{d,i}$ elements.

The scalar-valued cost function of the multi-agent system is denoted as $J(x_1, \dots, x_m)$, while the local constraint set is a mixed-integer polyhedral set

$$X_i(\delta) = \left\{ x_i \in \mathbb{X}_i : \; D_i(\delta) x_i \le d_i(\delta) \right\},$$

defined by means of a set of linear inequalities where matrix $D_i(\delta)$ and vector $d_i(\delta)$ are affected by some uncertain parameter vector $\delta \in \mathbb{R}^p$ taking values in a set $\Delta \subseteq \mathbb{R}^p$ according to some probability distribution $\mathbb{P}$. The agents' decisions are coupled by k_v global constraints possibly modeling the presence of some shared resources or the need to reach a consensus, and expressed as

$$v(x_1, \dots, x_m) \le 0,$$

where $v : \mathbb{X}_1 \times \cdots \times \mathbb{X}_m \to \mathbb{R}^{k_v}$, with the understanding that the inequality has to be intended component-wise. Global equality constraints can also be accounted for without modifications.

The decisions that minimize the objective function subject to local and global constraints can be obtained by solving the following multi-agent MIP

$$\min_{x_1,\dots,x_m} \quad J(x_1, \dots, x_m) \tag{1a}$$

$$\text{subject to:} \quad v(x_1, \dots, x_m) \le 0 \tag{1b}$$

$$x_i \in X_i(\delta), \qquad i = 1, \dots, m. \tag{1c}$$

Since the value of the uncertain parameter vector δ in (1c) is not known at decision time, then some alternative problem formulation must be adopted. A possibility would be to replace δ in (1c) with some nominal value $\bar{\delta}$, which would, however, lead to a solution that is guaranteed to be feasible only for $\delta = \bar{\delta}$. Another possibility is to enforce the local constraints for all possible values that the uncertain parameters δ may take, thus opting for a robust paradigm and replacing constraint (1c) with

$$x_i \in X_i(\delta), \quad i = 1, \dots, m, \quad \delta \in \Delta.$$

However, this approach requires knowledge of Δ and typically attains poor performance, as it also considers realizations of δ that are very unlikely to occur. In between these two extremes, the chance-constraint formulation leverages that δ is a random vector from Δ distributed according to $\mathbb{P}$ to enforce that the uncertain constraints are satisfied with a certain probability. Formally, this is achieved substituting (1c) with the following chance-constraint

$$\mathbb{P}\left\{\delta \in \Delta : \ x_i \in X_i(\delta), \ i = 1, \dots, m\right\} \geq 1 - \varepsilon, \tag{2}$$

where ε (also called *violation parameter*) tunes the risk of violating constraint $X_i(\delta)$ for at least one i. This is certainly the most flexible approach, but it relies on the knowledge of $\mathbb{P}$ (and implicitly of Δ), which may not be available. In addition, even when $\mathbb{P}$ and Δ are known, solving the resulting chance-constrained problem can be hard even when all decision variables are continuous.

If realizations of the uncertainty are available, they can be directly embedded in the optimization problem as a proxy for Δ and $\mathbb{P}$ to approximate (2), thus obtaining a standard MIP. Specifically, let us assume that each agent i has access to a collection

$$\mathcal{D}_i = \left\{\delta_i^{(1)}, \delta_i^{(2)}, \dots, \delta_i^{(N_i)}\right\} \in \Delta^{N_i}$$

of N_i samples of the uncertain vector δ, each extracted from Δ according to $\mathbb{P}$. Then, by enforcing the local constraints $X_i(\delta)$ in (1c) for all values of δ in $\mathcal{D}_i$, we obtain the following data-based optimization program

$$\min_{x_1,\dots,x_m} \quad J(x_1, \dots, x_m) \tag{3a}$$

$$\text{subject to:} \quad v(x_1, \dots, x_m) \leq 0 \tag{3b}$$

$$x_i \in X_i(\delta_i), \qquad \delta_i \in \mathcal{D}_i, \ i = 1 \dots, m, \tag{3c}$$

which is still a multi-agent MIP, but it is now deterministic. Its resolution does not require explicit knowledge of the domain Δ where uncertainty takes values nor its probability distribution $\mathbb{P}$. Any solution to (3) depends on the specific instances of the datasets $\mathcal{D}_1, \dots, \mathcal{D}_m$. Intuition suggests that as the number of samples N_i increases, a solution to (3) becomes more and more "robust" and more likely to belong to the local constraints $X_i(\delta)$ for an unseen value of $\delta \in \Delta$. This intuition rests on solid ground, as the generalization properties of data-based solutions have been extensively analyzed both in the convex and in the non-convex cases (Alamo et al., 2009; Campi et al., 2009, 2018), albeit the multi-agent framework has been considered only recently (see Falsone, Margellos, Prandini, and Garatti (2020), Falsone, Molinari, and Prandini (2020), Pantazis et al. (2022)).

Note that even if (3) is a deterministic problem, it can be difficult to solve in a centralized fashion, especially when the number of agents m is large. Additionally, a centralized resolution scheme requires agents to share their dataset, which may overload the communication network and also create privacy concerns. Computational tractability can be recovered when (3) exhibits a partially decomposable structure like in multi-agent MILPs where decentralized algorithms like Falsone et al. (2019), Manieri et al. (2023a) can be used to find a feasible (possibly suboptimal) solution, while preserving privacy.

In the next section, we show that by suitably selecting N_i, $i = 1, \dots, m$, we can equip any feasible solution of (3) with probabilistic feasibility guarantees for the chance-constraint in (2). Notably, such guarantees are valid even if agents collect data independently and keep them private, and also irrespective of the algorithm used to compute the solution, which can then be freely chosen by the user.

3. Probabilistic feasibility guarantees

In this section, we present our main results. We start by characterizing local feasibility guarantees at the level of the single agent in Section 3.1, which are then used to provide feasibility guarantees that hold for all agents jointly (and hence for problem (3)) in Section 3.2.

3.1. Local feasibility guarantees

Consider a generic agent $i \in \{1, \dots, m\}$ along with its local uncertain constraint set $X_i(\delta)$ and impose the following assumption.

Assumption 1. Agent i has access to a *private* dataset $\mathcal{D}_i$ of N_i realizations $\delta_i^{(1)}, \dots, \delta_i^{(N_i)}$ of the uncertain parameter δ, where each $\delta_i^{(j)}$ is extracted independently at random from Δ according to $\mathbb{P}$.

To ease the notation, let us denote the data-based local constraint set of agent i as

$$\mathcal{X}_i = \bigcap_{\delta_i \in \mathcal{D}_i} X_i(\delta_i).$$

Note that, since $\mathcal{X}_i$ depends on the random dataset $\mathcal{D}_i$, any decision $x_i \in \mathcal{X}_i$ selected based on the knowledge of $\mathcal{X}_i$ will be a random quantity itself and its properties will then hold with a certain confidence.

For a generic agent i, the following result relates the violation level of a feasible decision $x_i \in \mathcal{X}_i$ with the number of samples N_i in the dataset $\mathcal{D}_i$ and the confidence with which a local chance-constraint is satisfied by x_i.

Theorem 1. *Under Assumption 1, with confidence no smaller than $1 - \beta_i$, either $\mathcal{X}_i$ is empty, or any decision $x_i \in \mathcal{X}_i$ satisfies*

$$\mathbb{P}\left\{\delta \in \Delta : \ x_i \in X_i(\delta)\right\} \geq 1 - \varepsilon_i, \tag{4}$$

if N_i is such that

$$N_i \geq \frac{5}{\varepsilon_i}\left[2n_{c,i}\log_2(4ek_{c,i})\ln\left(\frac{40}{\varepsilon_i}\right) + \ln\left(\frac{4}{\beta_i}\right) + \ln(k_{d,i})\right], \tag{5}$$

where $n_{c,i}$ is the number of continuous optimization variables of agent i, $k_{c,i}$ is the number of linear inequality constraints of agent i involving continuous variables and affected by the uncertain parameter δ, and $k_{d,i}$ is the number of combinations for the discrete variables of agent i.

Proof. Let $g_i : \mathbb{R}^{n_{c,i}} \times \mathbb{Z}^{n_{d,i}} \times \Delta \to \{0,1\}$ be the binary measurable function

$$g_i(x_{c,i}, x_{d,i}, \delta) = \begin{cases} 0 & \text{if } x_i \in X_i(\delta) \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

describing the violation of local constraint of agent i evaluated at decision $x_i = [x_{c,i}^\top \ \ x_{d,i}^\top]^\top$ and uncertainty instance $\delta \in \Delta$, where subvector $x_{c,i} \in \mathbb{R}^{n_{c,i}}$ collects all the continuous decision variables, and vector $x_{d,i} \in \mathbb{Z}^{n_{d,i}}$ all the discrete variables of agent i. Then, from Alamo et al. (2009, Definition 1), the probability of violation of x_i for the function $g_i : \mathbb{R}^{n_{c,i}} \times \mathbb{Z}^{n_{d,i}} \times \Delta \to \{0,1\}$ is defined as

$$\begin{aligned} V_{g_i}(x_i) &= V_{g_i}(x_{c,i}, x_{d,i}) \\ &= \mathbb{P}\left\{\delta \in \Delta : \ g_i(x_{c,i}, x_{d,i}, \delta) = 1\right\}. \end{aligned} \tag{7}$$

We are interested in computing the probability of extracting a dataset $\mathcal{D}_i$ and finding a solution $x_i \in \mathcal{X}_i$ which has violation bigger than ε_i. Formally, we want to estimate the probability of one-sided constraint failure (cf. Alamo et al., 2009, Definition 3)

$$\begin{aligned} p_{g_i}(N_i, \varepsilon_i) = \mathbb{P}^{N_i}\big\{&\mathcal{D}_i \in \Delta^{N_i} : \ \exists x_{c,i}, x_{d,i} : \\ &(g_i(x_{c,i}, x_{d,i}, \delta) = 0, \ \delta \in \mathcal{D}_i) \wedge (V_{g_i}(x_{c,i}, x_{d,i}) > \varepsilon_i)\big\}, \end{aligned} \tag{8}$$

which can be upper bounded as follows

$$p_{g_i}(N_i, \varepsilon_i) \leq \sum_{x_{d,i} \in \Xi_i} p_{g_i, x_{d,i}}(N_i, \varepsilon_i), \tag{9}$$

where Ξ_i represents the set of all the possible combinations that discrete variables $x_{d,i}$ can take, and $p_{g_{i,x_{d,i}}}$ is similar to (8) but for a fixed value of $x_{d,i}$, i.e.,

$$p_{g_{i,x_{d,i}}}(N_i,\varepsilon_i)=\mathbb{P}^{N_i}\{\mathcal{D}_i\in\Delta^{N_i}\ :\ \exists x_{c,i}\ :\ (\underbrace{g_i(x_{c,i},x_{d,i},\delta)}_{g_{i,x_{d,i}}(x_{c,i},\delta)}=0,\ \delta\in\mathcal{D}_i)\wedge(\underbrace{V_{g_i}(x_{c,i},x_{d,i})}_{V_{g_{i,x_{d,i}}}(x_{c,i})}>\varepsilon_i)\}. \tag{10}$$

Based on Alamo et al. (2009, Definition 4) and following Alamo et al. (2009, Theorem 1 with $\rho=0$), the probability of one-sided constraint failure $p_{g_{i,x_{d,i}}}(N_i,\varepsilon_i)$ can be upper bounded by the probability of relative difference failure (cf. Alamo et al., 2009, Definition 4)

$$r_{g_{i,x_{d,i}}}\left(N_i,\sqrt{\varepsilon_i}\right)=\mathbb{P}^{N_i}\left\{\mathcal{D}_i\in\Delta^{N_i}\ :\ \sup_{x_{c,i}\in\mathbb{R}^{n_{c,i}}}\sqrt{V_{g_{i,x_{d,i}}}(x_{c,i})}>\sqrt{\varepsilon_i}\right\} \tag{11}$$

as

$$p_{g_{i,x_{d,i}}}(N_i,\varepsilon_i)\le r_{g_{i,x_{d,i}}}\left(N_i,\sqrt{\varepsilon_i}\right). \tag{12}$$

Then, according to Alamo et al. (2009, Theorem 5) we have

$$r_{g_{i,x_{d,i}}}\left(N_i,\sqrt{\varepsilon_i}\right)<4\pi_{g_{i,x_{d,i}}}(2N_i)e^{-N_i\varepsilon_i/4} \tag{13}$$

where $\pi_{g_{i,x_{d,i}}}(\cdot)$ is the growth function (cf. Alamo et al., 2009, Definition 5) associated with $g_{i,x_{d,i}}$ and is defined as follows. Given the function $g_{i,x_{d,i}}\ :\ \mathbb{R}^{n_{c,i}}\times\Delta\to\{0,1\}$ and the dataset $\mathcal{D}_i=\left\{\delta_i^{(1)},\dots,\delta_i^{(N_i)}\right\}\in\Delta^{N_i}$, if $\phi_{g_{i,x_{d,i}}}(\mathcal{D}_i)$ denotes the number of distinct binary vectors

$$\left\{g_{i,x_{d,i}}\left(x_{c,i},\delta_i^{(1)}\right),\dots,g_{i,x_{d,i}}\left(x_{c,i},\delta_i^{(N_i)}\right)\right\}\in\{0,1\}^{N_i} \tag{14}$$

that can be obtained letting $x_{c,i}$ vary in $\mathbb{R}^{n_{c,i}}$, then the growth function $\pi_{g_{i,x_{d,i}}}(N_i)$ is defined as

$$\pi_{g_{i,x_{d,i}}}(N_i)=\sup_{\mathcal{D}_i\in\Delta^{N_i}}\phi_{g_{i,x_{d,i}}}(\mathcal{D}_i). \tag{15}$$

By Alamo et al. (2009, Lemma 1), if the family of functions $\{g_{i,x_{d,i}}(x_{c,i},\delta),\ x_{c,i}\in\mathbb{R}^{n_{c,i}}\}$ has a finite Vapnik–Chervonenkis dimension (VC-dimension, cf. Alamo et al., 2009, Definition 6) $\mathcal{VC}_{g_{i,x_{d,i}}}<\infty$, then, for any $N_i>\mathcal{VC}_{g_{i,x_{d,i}}}$, it holds that

$$\pi_{g_{i,x_{d,i}}}(2N_i)\le\left(\frac{2eN_i}{\mathcal{VC}_{g_{i,x_{d,i}}}}\right)^{\mathcal{VC}_{g_{i,x_{d,i}}}}. \tag{16}$$

Using (12), (13) and (16) in (9), we can finally estimate $p_{g_i}(N_i,\varepsilon_i)$ as

$$p_{g_i}(N_i,\varepsilon_i)<\sum_{x_{d,i}\in\Xi_i}4\left(\frac{2eN_i}{\mathcal{VC}_{g_{i,x_{d,i}}}}\right)^{\mathcal{VC}_{g_{i,x_{d,i}}}}e^{-N_i\varepsilon_i/4}. \tag{17}$$

If we upper bound $\mathcal{VC}_{g_{i,x_{d,i}}}$ with some value $\overline{\mathcal{VC}}_i$ independent of $x_{d,i}$, the right hand side is just a summation of the same quantity over all possible combinations Ξ_i that $x_{d,i}$ may take. Thus we can enforce $p_{g_i}(N_i,\varepsilon_i)\le\beta_i$ by requiring

$$4\left(\frac{2eN_i}{\overline{\mathcal{VC}}_i}\right)^{\overline{\mathcal{VC}}_i}e^{-N_i\varepsilon_i/4}\le\frac{\beta_i}{k_{d,i}}. \tag{18}$$

which, by Alamo et al. (2009, Theorem 6 with $a=4$, $b=\varepsilon_i/4$, $c=2$, $\delta=\beta_i/k_{d,i}$, and $\mu=5$), can be made explicit in N_i as

$$N_i\ge\frac{5}{\varepsilon_i}\left(\overline{\mathcal{VC}}_i\ln\left(\frac{40}{\varepsilon_i}\right)+\ln\left(\frac{4k_{d,i}}{\beta_i}\right)\right). \tag{19}$$

Since $X_i(\delta)$ is a mixed-integer polyhedral set, if we fix a combination for the discrete variables $x_{d,i}$, then the remaining continuous variables $x_{c,i}$ are constrained to belong to the polyhedral set

$$X_i^c(x_{d,i},\delta)=\left\{x_{c,i}\in\mathbb{R}^{n_{c,i}}\ :\ D_i^c(\delta)x_{c,i}\le d_i(\delta)-D_i^d(\delta)x_{d,i}\right\}$$

and the function $g_{i,x_{d,i}}(x_{c,i},\delta)$ can be expressed as

$$g_{i,x_{d,i}}(x_{c,i},\delta)=\bigvee_{r=1}^{k_{c,i}}\begin{cases}0 & \text{if } \left[D_i^c(\delta)\,x_{c,i}\right]_r\le\left[d_i(\delta)-D_i^d(\delta)\,x_{d,i}\right]_r\\ 1 & \text{otherwise}\end{cases} \tag{20}$$

where $[\,\cdot\,]_r$ denotes the rth component of its vector argument. Relation (20) shows that $g_{i,x_{d,i}}(x_{c,i},\delta)$ can be expressed as a $(1,k_{c,i})$-Boolean function (Alamo et al., 2009, Definition 7), as it is the result of a Boolean operation among $k_{c,i}$ polynomials in $x_{c,i}$ with maximum degree equal to 1. Thus, we can leverage Alamo et al. (2009, Lemma 2) to get

$$\mathcal{VC}_{g_{i,x_{d,i}}}\le 2n_{c,i}\log_2(4ek_{c,i}). \tag{21}$$

Setting $\overline{\mathcal{VC}}_i=2n_{c,i}\log_2(4ek_{c,i})$ in (19) yields (5), thus concluding the proof. □

Note that the confidence parameter β_i in (5) appears inside a logarithm, meaning that we can push β_i very close to zero (and have that (4) holds with almost certainty) without increasing N_i much.

3.2. Global feasibility guarantees

Building upon the above result for a single agent, we can now derive feasibility guarantees for the joint decision of the agents. To this end, we impose the following assumption.

Assumption 2. Let Assumption 1 hold for all $i=1,\dots,m$ and assume that samples collected by different agents are independent.

The following result (inspired by Falsone, Margellos, Prandini, and Garatti (2020, Proposition 1 and Theorem 3) for the case of convex problems) confirms the intuition mentioned in Section 2 and relates the cardinality of each agent dataset with the violation level of a (feasible for (3)) joint decision $x=[x_1^\top\ \cdots\ x_m^\top]^\top$ and the confidence with which the chance-constraint in (2) is satisfied by x.

Theorem 2. *Let $\varepsilon=\sum_{i=1}^m\varepsilon_i$ and $\beta=\sum_{i=1}^m\beta_i$. Under Assumption 2, if, for each agent $i=1,\dots,m$, N_i satisfies (5) in Theorem 1, then, with confidence at least $1-\beta$, either problem (3) is infeasible, or any feasible solution $x=[x_1^\top\ \cdots\ x_m^\top]^\top$ satisfies*

$$\mathbb{P}\left\{\delta\in\Delta\ :\ \ x_i\in X_i(\delta),\ i=1,\dots,m\right\}\ge 1-\varepsilon. \tag{22}$$

Proof. Recall that $\mathcal{X}_i=\bigcap_{\delta\in\mathcal{D}_i}X_i(\delta)$, let $\mathcal{D}=\bigcup_{i=1}^m\mathcal{D}_i$ and define $N=\sum_{i=1}^m N_i$. The statement of the theorem can be equivalently written as

$$\mathbb{P}^N\left\{\mathcal{D}\in\Delta^N\ :\ (3)\text{ is feasible }\wedge\ \mathbb{P}\{\delta\in\Delta\ :\ \exists i\in\{1,\dots,m\}\ :\ x_i\notin X_i(\delta)\}>\varepsilon\right\}\le\beta. \tag{23}$$

To show that (23) holds, let us focus on the left-hand side. Since requiring that (3) is feasible is stricter than requiring $\mathcal{X}_i$ to be nonempty for all $i=1,\dots,m$, the left-hand side of (23) is upper bounded by

$$\mathbb{P}^N\left\{\mathcal{D}\in\Delta^N\ :\ \mathcal{X}_i\neq\emptyset\text{ for all }i=1,\dots,m\ \wedge\ \mathbb{P}\{\delta\in\Delta\ :\ \exists i\in\{1,\dots,m\}\ :\ x_i\notin X_i(\delta)\}>\varepsilon\right\}. \tag{24}$$

Let us denote with $E_i=\left\{\delta\in\Delta\ :\ x_i\notin X_i(\delta)\right\}$ the event that we extract a $\delta\in\Delta$ and x_i is not feasible for $X_i(\delta)$. Then (24) can be compactly written as

$$\mathbb{P}^N\left\{\mathcal{D}\in\Delta^N\ :\ \{\mathcal{X}_i\neq\emptyset,\ \forall i\}\ \wedge\ \mathbb{P}\{\textstyle\bigcup_{i=1}^m E_i\}>\varepsilon\right\}. \tag{25}$$

From the sub-additivity property of probability measures (cf. Papoulis & Pillai, 2002, Theorem 2.3), we have

$$\mathbb{P}\{\textstyle\bigcup_{i=1}^m E_i\}\le\sum_{i=1}^m\mathbb{P}\{E_i\},$$

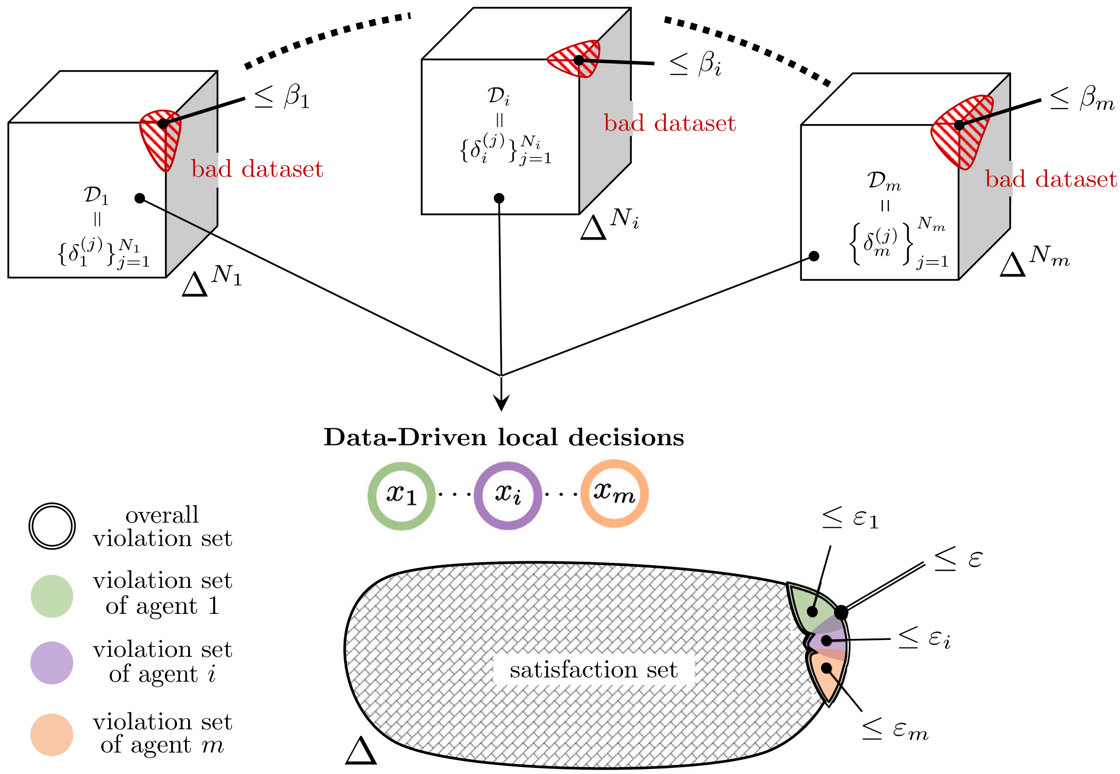

Fig. 1. Pictorial representation of Theorems 1 and 2.

meaning that $\mathbb{P}\{\bigcup_{i=1}^m E_i\} > \varepsilon$ is less likely to occur than $\sum_{i=1}^m \mathbb{P}\{E_i\} > \varepsilon$ and, hence,

$$(24) = (25)$$

$$\leq \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \{\mathcal{X}_i \neq \emptyset,\ \forall i\} \wedge \sum_{i=1}^m \mathbb{P}\{E_i\} > \sum_{i=1}^m \varepsilon_i\Big\}, \tag{26}$$

where in the last expression we have used the definition $\varepsilon = \sum_{i=1}^m \varepsilon_i$. Moreover, since $\sum_{i=1}^m \mathbb{P}\{E_i\} > \sum_{i=1}^m \varepsilon_i$ implies that there exist at least one $i \in \{1,\ldots,m\}$ for which $\mathbb{P}\{E_i\} > \varepsilon_i$, then

$$\begin{aligned}(24) \leq (26) &\leq \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \{\mathcal{X}_i \neq \emptyset,\ \forall i\} \wedge \{\exists i :\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\}\\ &\leq \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \exists i :\ \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\}\\ &= \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \bigcup_{i=1}^m \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\},\end{aligned} \tag{27}$$

where the last inequality is due to the fact that the event is not requiring $\mathcal{X}_j \neq \emptyset$ for $j \neq i$. Using again sub-additivity (cf. Papoulis & Pillai, 2002, Theorem 2.3), we have

$$(24) \leq (27) \leq \sum_{i=1}^m \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\}. \tag{28}$$

Under Assumption 2, for each agent $i = 1,\ldots,m$, Theorem 1 holds irrespective of the samples extracted by the other agents, so that we can rewrite the statement of Theorem 1 as

$$\begin{aligned}\beta_i &\geq \mathbb{P}^{N_i}\Big\{\mathcal{D}_i\in\Delta^{N_i} : \ \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\}\\ &= \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\ \Big|\ \mathcal{D}\setminus\mathcal{D}_i \in \Delta^{N-N_i}\Big\}\\ &= \mathbb{P}^N\Big\{\mathcal{D}\in\Delta^N : \ \{\mathcal{X}_i \neq \emptyset\ \wedge\ \mathbb{P}\{E_i\} > \varepsilon_i\}\Big\},\end{aligned} \tag{29}$$

where the last equality is obtained integrating over $\mathcal{D}\setminus\mathcal{D}_i$. Using (29) in (28) and recalling that $\beta = \sum_{i=1}^m \beta_i$ we obtain (24) $\leq \beta$. Recalling that (24) is an upper bound of the left-hand side of (23) finally yields (23), thus concluding the proof. $\square$

Note that Theorem 2 holds irrespective of how ε and β are split among $\varepsilon_1,\ldots,\varepsilon_m$ and $\beta_1,\ldots,\beta_m$. It is, therefore, intuitive to choose ε_i and β_i in a way that minimizes the overall number of samples $\sum_{i=1}^m N_i$ required. This can be achieved by solving the following optimization problem

$$\begin{aligned}\min_{\{(\varepsilon_i,\beta_i,N_i)\}_{i=1}^m} \quad & \sum_{i=1}^m N_i\\ \text{subject to:} \quad & \sum_{i=1}^m \varepsilon_i = \varepsilon, \quad \sum_{i=1}^m \beta_i = \beta\\ & (5), \quad i = 1,\ldots,m\\ & (\varepsilon_i,\beta_i) \in [0,1]^2, \quad i = 1,\ldots,m,\end{aligned} \tag{30}$$

which is a convex problem since the right-hand side of (5) is convex in ε_i and β_i as an effect of being a positive sum of three terms: $\frac{1}{\varepsilon_i}\ln\frac{1}{\varepsilon_i}$, $\frac{1}{\varepsilon_i}\ln\frac{1}{\beta_i}$, and $\frac{1}{\varepsilon_i}$, which are all convex when $(\varepsilon_i,\beta_i) \in [0,1]^2$.

Inspired by Campi et al. (2009), the results in Theorems 1 and 2 can be depicted as in Fig. 1. Each cube at the top represents the (private) N_i-dimensional space containing all possible datasets each agent may have, and the cubes are separated because these datasets are assumed to be independent. The extracted datasets are then used to formulate problem (3) and obtain a feasible solution $x = [x_1^\top \cdots x_m^\top]^\top$. Each agent component x_i partitions the uncertainty space Δ into two regions: a satisfaction set (those δ's for which $x_i \in X_i(\delta)$) and a violation set (those δ's for which $x_i \notin X_i(\delta)$). Theorem 1 ensures that the violation set of each agent i has a $\mathbb{P}$-probability measure of at most ε_i, provided that N_i is sufficiently high. Finally, Theorem 2 leverages the previous result to ensure that the union of the violation sets of all agents has a $\mathbb{P}$-probability measure of at most ε, provided that $N_1,\ldots,N_m$ are all sufficiently high. These local and global guarantees hold with a certain confidence since the extracted dataset might not be informative enough to generalize to unseen uncertainty instances. These "bad datasets" are confined in the red region of each cube whose $\mathbb{P}^{N_i}$-probability measure can be made arbitrarily small (by suitably increasing N_i) to make it almost impossible to extract datasets for which (3) is feasible but the union of the violation sets of all agents exceeds ε.

Since Theorem 2 relies on Theorem 1, also in this case, we can push β to have a very small value so that chance-constraint (2) is satisfied with almost certainty. However, as the number of agents m increases, in order to keep $\varepsilon = \sum_{i=1}^{m} \varepsilon_i$ small, we need to reduce every ε_i, which in turn, according to Theorem 1, increases the number of samples N_i each agent must have. In case N_i's are fixed (because we have access to only a finite number of data) ε_i are fixed as well and, therefore, ε increases with m, eventually leading to an $\varepsilon > 1$ and no generalization.

Finally, notice that, since the results in Theorem 2 hold for any feasible solution of (3), they can be used to certify the (feasible) solution returned by any distributed or decentralized solution algorithm.

Next, we show how the results in this section apply to the case where agents are still affected by the same uncertainty but have access to a common dataset and to the case where each agent is affected by independent uncertainties.

3.3. Common dataset case

Let us now consider the framework of Falsone, Molinari, and Prandini (2020), where each agent has access to the same dataset $\bar{\mathcal{D}}$ of $\bar{N}$ uncertainty instances. According to the notation introduced in the previous section, this corresponds to setting $N_i = \bar{N}$ and $\mathcal{D}_i = \bar{\mathcal{D}}$, for all $i = 1, \dots, m$, and the resulting data-based MIP reads as

$$\min_{x_1,\dots,x_m} \quad J(x_1,\dots,x_m) \tag{31a}$$

$$\text{subject to:} \quad v(x_1,\dots,x_m) \le 0 \tag{31b}$$

$$x_i \in X_i(\delta), \qquad \delta \in \bar{\mathcal{D}},\ i = 1,\dots,m. \tag{31c}$$

In Falsone, Molinari, and Prandini (2020), Theorem 1 ensures that any feasible solution of (31) satisfies chance-constraint (2) with confidence at least $1-\beta$, if the dataset size $\bar{N}$ satisfies

$$\bar{N} \ge \frac{5}{\varepsilon}\left[2n_c \log_2(4ek_c)\ln\left(\frac{40}{\varepsilon}\right) + \ln\left(\frac{4}{\beta}\right) + \ln(k_d)\right], \tag{32}$$

where n_c is the total number of continuous decision variables, k_c is the total number of linear inequality constraints that involve continuous variables and are affected by the uncertainty, and k_d is the total number of admissible combinations of the discrete variables, where total means regarding (31) as a whole.

The fact that (32) is similar to (5) should not be surprising. Indeed, Falsone, Molinari, and Prandini (2020, Theorem 1) can be interpreted as a straightforward application of Theorem 1, where all decision variables and local constraints of problem (31) are fictitiously assigned to a single agent. In fact, if we define $\bar{x} = [x_1^\top \cdots x_m^\top]^\top$, and $\bar{X}(\delta) = X_1(\delta) \times \cdots \times X_m(\delta)$, then problem (31) becomes

$$\min_{\bar{x}} \quad J(\bar{x}) \tag{33a}$$

$$\text{subject to:} \quad v(\bar{x}) \le 0 \tag{33b}$$

$$\bar{x} \in \bar{X}(\delta), \qquad \delta \in \bar{\mathcal{D}}, \tag{33c}$$

which fits (3) with $m = 1$. We can thus readily apply Theorem 1 and substitute $n_{c,i} = n_c$, $k_{c,i} = k_c$, $k_{d,i} = k_d$, $\varepsilon_i = \varepsilon$, $\beta_i = \beta$, and $N_i = \bar{N}$ in (5) (which is equal to (32)) to get that, with confidence no smaller than $1-\beta$, either $\bar{\mathcal{X}} = \bigcap_{\delta\in\bar{\mathcal{D}}} \bar{X}(\delta)$ is empty, or any decision $\bar{x} \in \bar{\mathcal{X}}$ satisfies

$$\mathbb{P}\left\{\delta \in \Delta :\ \bar{x} \in \bar{X}(\delta)\right\} \ge 1-\varepsilon.$$

The statement in Falsone, Molinari, and Prandini (2020, Theorem 1) is then recovered noticing (similarly to the beginning of the proof of Theorem 2) that requiring (33) to be feasible is stricter than requiring $\bar{\mathcal{X}} \neq \emptyset$. Nonetheless, this result holds only when agents have access to a common dataset.

3.4. Independent local uncertainty case

Another interesting case is the one in which each agent is affected by independent uncertainty parameters. In this case, the parameter vector δ can be decomposed as $\delta = [\delta_1^\top \cdots \delta_m^\top]^\top$, δ_i and δ_j being independent random vectors for all $i \neq j$, and agent i local set $X_i(\delta) = X_i(\delta_i)$ depends on its local uncertainty parameters δ_i only, for all $i = 1,\dots,m$. If we assume that each agent has access to a collection of N_i samples $\delta_{i,i}^{(1)},\dots,\delta_{i,i}^{(N_i)}$ of its uncertain parameters δ_i, then the data-based problem to be solved is given by

$$\min_{x_1,\dots,x_m} \quad J(x_1,\dots,x_m) \tag{34a}$$

$$\text{subject to:} \quad v(x_1,\dots,x_m) \le 0 \tag{34b}$$

$$x_i \in X(\delta_{i,i}^{(\ell)}), \quad \ell = 1,\dots,N_i,\ i = 1,\dots,m. \tag{34c}$$

It is easy to show that (34) fits the framework presented in Section 2. To see this, let us construct for each agent i a fictitious dataset

$$\mathcal{D}_i = \left\{\begin{bmatrix}\delta_{1,i}^{(1)}\\ \vdots \\ \delta_{i,i}^{(1)} \\ \vdots \\ \delta_{m,i}^{(1)}\end{bmatrix},\dots,\begin{bmatrix}\delta_{1,i}^{(N_i)}\\ \vdots \\ \delta_{i,i}^{(N_i)} \\ \vdots \\ \delta_{m,i}^{(N_i)}\end{bmatrix}\right\},$$

complementing the N_i realizations of its own uncertainty parameters δ_i with N_i realizations $\delta_{j,i}^{(1)},\dots,\delta_{j,i}^{(N_i)}$ of the uncertainty parameters δ_j, $j = 1,\dots,m$ and $j \neq i$, of the other agents, extracted independently. In this way, each agent has an independent dataset of independent samples of the overall uncertainty vector δ, thus satisfying Assumptions 1 and 2. Since we assumed that each agent is affected by its local uncertainty parameters δ_i only, then $X_i(\delta) = X_i(\delta_i)$ and

$$\bigcap_{\ell=1}^{N_i} X_i(\delta_{i,i}^{(\ell)}) = \bigcap_{\delta_i\in\mathcal{D}_i} X_i(\delta_i)$$

for all $i = 1,\dots,m$, which shows that (34) is actually an instance of (3) and, as such, inherits the guarantees granted by Theorems 1 and 2.

We would like to stress that, in practice, agent i does *not* actually need to extract samples of the uncertain parameters of the other agents because its local constraints $X_i(\delta_i)$ are not affected by δ_j for $j \neq i$.

This case where the agents are affected by independent uncertain parameters can also be seen as a special case of the common dataset case of Section 3.3. To this end, let us further assume that $N_i = \bar{N}$ for all $i = 1,\dots,m$. To see that, in this case, (34) fits (31) it is sufficient to define

$$\bar{\mathcal{D}} = \left\{\begin{bmatrix}\delta_{1,1}^{(1)}\\ \vdots \\ \delta_{i,i}^{(1)} \\ \vdots \\ \delta_{m,m}^{(1)}\end{bmatrix},\dots,\begin{bmatrix}\delta_{1,1}^{(\bar{N})}\\ \vdots \\ \delta_{i,i}^{(\bar{N})} \\ \vdots \\ \delta_{m,m}^{(\bar{N})}\end{bmatrix}\right\},$$

which is equivalent to stacking the $\delta_{i,i}^{(1)},\dots,\delta_{i,i}^{(\bar{N})}$ realizations of all agents to build a common dataset of $\bar{N}$ realizations of the overall uncertainty vector $\delta = [\delta_1^\top \cdots \delta_m^\top]^\top$, and, then, note that

$$\bigcap_{\ell=1}^{N_i} X_i(\delta_{i,i}^{(\ell)}) = \bigcap_{\delta\in\bar{\mathcal{D}}} X_i(\delta).$$

This shows that also the results of Falsone, Molinari, and Prandini (2020) can be applied to this case, but with the additional restriction that all agents must have the same number $\bar{N}$ of realizations of their local uncertainty parameters. This is, instead, not required for the results presented in this paper.

4. Comparison with the common dataset case

Since we have two special cases (cf. Sections 3.3 and 3.4) in which both the results of this paper and those of Falsone, Molinari, and Prandini (2020) can be applied, it is interesting to compare the sample size of the twos with ε and β fixed and a varying number of agents m.

At first glance, the fact that Theorem 2 requires $\sum_{i=1}^m \varepsilon_i = \varepsilon$ suggests that there is a significant price to pay to allow the agents to have private datasets since we need to decrease ε_i if m increases, but a closer look at (5) and (32) reveals that the right-hand-side of (5) is typically much lower than that of (32) since the complexity (in terms of variables and constraints) associated with a single agent is just a fraction of the complexity of the overall problem, balancing the decrease in ε_i.

This observation can be made more precise if we consider the case of homogeneous agents (simpler in notation). To this end, assume that the agents have the same number of continuous and discrete decision variables and the same number and type of constraints so that

$$n_{c,i} = \bar{n}_c, \quad k_{d,i} = \bar{k}_d, \quad k_{c,i} = \bar{k}_c, \quad n_c = m\,\bar{n}_c, \quad k_d = \bar{k}_d^m, \quad k_c = m\,\bar{k}_c.$$

Under these conditions, by strict convexity of the right-hand side of (5), it is intuitive to uniformly split ε and β across the agents by setting $\varepsilon_i = \varepsilon/m$ and $\beta_i = \beta/m$ to minimize $\sum_{i=1}^m N_i$ as prescribed by (30).

Evaluating (5) and (32) in this homogeneous agents case, we obtain

$$N_i \geq \frac{5m}{\varepsilon}\left[2\bar{n}_c \log_2(4e\bar{k}_c)\ln\left(\frac{40m}{\varepsilon}\right) + \ln\left(\frac{4m}{\beta}\right) + \ln(\bar{k}_d)\right], \tag{35}$$

and

$$\bar{N} \geq \frac{5}{\varepsilon}\left[2m\bar{n}_c \log_2(4em\bar{k}_c)\ln\left(\frac{40}{\varepsilon}\right) + \ln\left(\frac{4}{\beta}\right) + m\ln(\bar{k}_d)\right], \tag{36}$$

whose right-hand sides can be manipulated using $\log_c(ab) = \log_c(a) + \log_c(b)$ and $\log_a(c)\log_b(d) = \log_b(c)\log_a(d)$ to get

$$N_i \geq \frac{5}{\varepsilon}\Bigg[2m\bar{n}_c \log_2(4e\bar{k}_c)\ln\left(\frac{40}{\varepsilon}\right) + m\ln(\bar{k}_d) + 2m\bar{n}_c \log_2(4e\bar{k}_c)\ln(m) + m\ln\left(\frac{4m}{\beta}\right)\Bigg], \tag{37}$$

and

$$\bar{N} \geq \frac{5}{\varepsilon}\Bigg[2m\bar{n}_c \log_2(4e\bar{k}_c)\ln\left(\frac{40}{\varepsilon}\right) + m\ln(\bar{k}_d) + 2m\bar{n}_c \log_2\left(\frac{40}{\varepsilon}\right)\ln(m) + \ln\left(\frac{4}{\beta}\right)\Bigg]. \tag{38}$$

Inspecting (37) and (38) it is immediately clear that the first two terms inside the square brackets are the same. It is also clear that the term containing the confidence parameter is different, and the one in (37) scales as $m\ln(m)$ while the one in (38) does not scale with m. This is intuitive because whilst in the common dataset case there is only one dataset accessible to all agents, in the case of private datasets there is an increased chance of extracting a bad dataset that must be taken into account. The analysis of the remaining term is a bit more subtle since, as a function of m, it scales in the same way in both expressions but with a different coefficient depending on $\bar{k}_c$ and ε. Specifically, the term in (37) is lower than the corresponding one in (38) if and only if

$$\log_2(4e\bar{k}_c) < \log_2\left(\frac{40}{\varepsilon}\right) \iff \bar{k}_c\,\varepsilon < \frac{10}{e}, \tag{39}$$

which can be interpreted as follows: asking for a low violation parameter ε is less demanding for less complicated sets (low $\bar{k}_c$), hence splitting the complexity of the overall problem into m sub-problems requires fewer samples for the local constraints. As a matter of fact, this term could even compensate for the difference in the terms depending on the confidence parameters, effectively yielding $N_i < \bar{N}$.

Indeed, by further expressing $m\ln(4m/\beta)$ as $m\ln(4/\beta) + m\ln(m)$ and using the property $\log_c(a) - \log_c(b) = \log_c(a/b)$, the difference between the right-hand-side of (37) and (38) is given by

$$\begin{aligned}
&\frac{5}{\varepsilon}\left[\left(2\bar{n}_c \log_2\left(\frac{4e\bar{k}_c\varepsilon}{40}\right) + 1\right) m\ln(m) + (m-1)\ln\left(\frac{4}{\beta}\right)\right] \\
&= \frac{5m}{\varepsilon}\left[\left(2\bar{n}_c \log_2\left(\frac{4e\bar{k}_c\varepsilon}{40}\right) + 1\right)\ln(m) + \frac{m-1}{m}\ln\left(\frac{4}{\beta}\right)\right] \\
&\leq \frac{5m}{\varepsilon}\left[\left(2\bar{n}_c \log_2\left(\frac{4e\bar{k}_c\varepsilon}{40}\right) + 1\right)\ln(m) + \ln\left(\frac{4}{\beta}\right)\right],
\end{aligned}$$

which will be negative (hence implying $N_i < \bar{N}$) for sufficiently high m whenever

$$2\bar{n}_c \log_2\left(\frac{4e\bar{k}_c\varepsilon}{40}\right) + 1 < 0 \iff \bar{k}_c\,\varepsilon < \frac{1}{\sqrt[2\bar{n}_c]{2}}\frac{10}{e}, \tag{40}$$

which is slightly stricter than (39), but approaches it as $\bar{n}_c$ increases. The interpretation of condition (40) is very similar to that of condition (39).

Up to now, we compared the sample complexity for each agent versus the sample complexity of a common dataset. However, the overall number of samples for the case of the private datasets is given by $\sum_{i=1}^m N_i = m\,N_i$ for the homogeneous case, which introduces an additional m factor. Yet, such a comparison may seem somewhat unfair to further consideration since (regardless of the privacy requirements) each agent must have access to a dataset, and in the common dataset case every agent uses the same, while in the case of private datasets each agent has to (unavoidably) have its own.

To support the previous discussion, we report in the sequel a numerical comparison of (35) and (36) where we set the problem parameters as

$$\bar{n}_c = 10, \quad \bar{k}_d = 2^5, \quad \bar{k}_c = 15, \quad \varepsilon = 0.01, \quad \beta = 10^{-6},$$

and we let m vary between 1 and 100. We checked numerically that $\varepsilon_i = \varepsilon/m$ and $\beta_i = \beta/m$ was indeed the optimal solution to (30) for the considered case.

In Fig. 2, we compare the minimum number of samples for the common dataset case $\bar{N}$ and the minimum number of samples N_i for each agent for the private dataset case. As expected, since $\bar{k}_c\,\varepsilon = 0.5 \ll \frac{10}{e}$, N_i is smaller than $\bar{N}$ irrespective of the number of agents considered, meaning that the sample complexity of each local problem is slightly lower for the private dataset case than for the common dataset case.

Clearly, the common dataset case outperforms the private dataset case if we compare, instead, the overall sample complexity $\bar{N}$ vs. $N = \sum_{i=1}^m N_i$, as reported in the upper plot of Fig. 3, which shows that N is between 1 and 2 orders of magnitude bigger than $\bar{N}$ (as $1 \leq m \leq 100$), thus requiring far more data to get the same feasibility guarantees. The ratio $N/\bar{N}$ is reported in the lower plot of Fig. 3 (blue dots) for different values of m. It grows linearly with the number of agents but is generally such that $N < m\,\bar{N}$.

Conversely, Fig. 4 shows how different values of ε affect the sample complexity for a fixed number of agents $m = 100$. For values of the violation parameter ε smaller than $\frac{1}{\bar{k}_c}\frac{1}{\sqrt[2\bar{n}_c]{2}}\frac{10}{e} = 0.2539$ (denoted in the plot with a vertical dash-dotted black line), the term $\log_2(4e\bar{k}_c)$ in (37) is lower enough than the corresponding term $\log_2(\frac{40}{\varepsilon})$ in (38), so as to compensate also for the confidence term in (37), yielding a local sample size N_i smaller than $\bar{N}$ and thus fewer local constraints in (3). The behavior of N and $\bar{N}$ as a function of ε is reported for the sake of completeness in Fig. 5. Not surprisingly, the overall number of samples for the private dataset is still two orders of magnitude (roughly m times) bigger than the number of samples required by the common dataset case.

To summarize, enforcing the privacy of the agents' local datasets takes a toll in terms of the total number of samples required, which

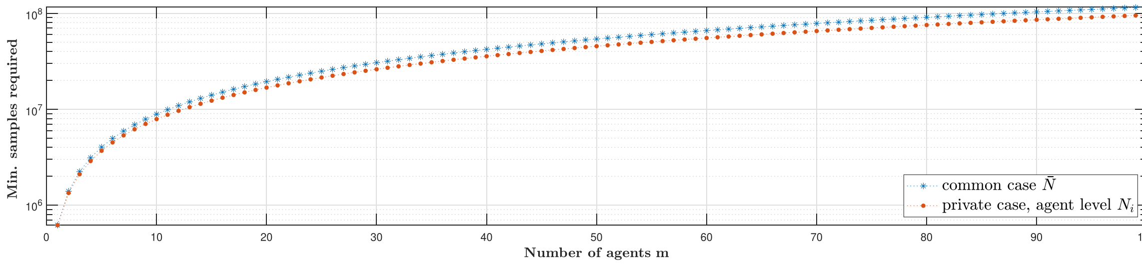

Fig. 2. $\bar{N}$ (blue asterisks) vs. $N_i = \tilde{N}$ (red circles) for different values of the number of agents m.

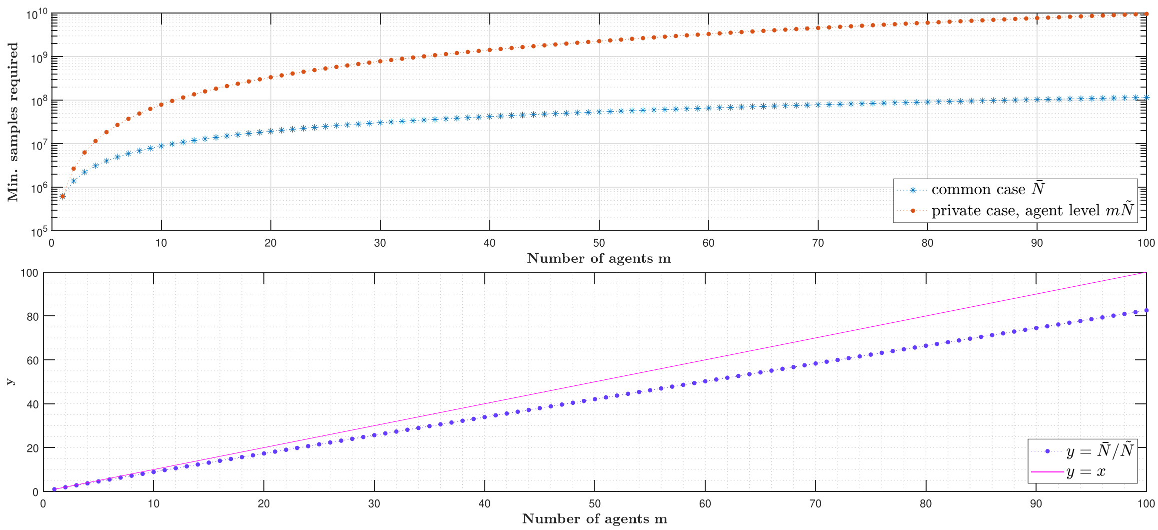

Fig. 3. Upper plot: $\bar{N}$ (blue asterisks) vs. $N = m\tilde{N}$ (red circles) for different values of the number of agents m. Lower plot: ratio $N/\bar{N}$ (blue asterisks) vs. m (magenta line) for different values of the number of agents m.

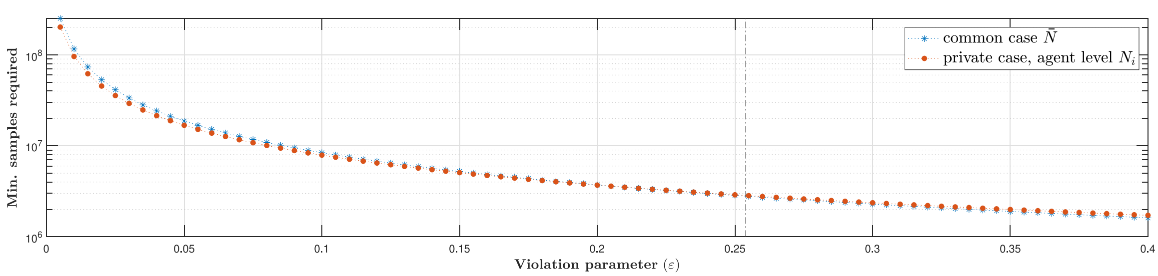

Fig. 4. $\bar{N}$ (blue asterisks) vs. $N_i = \tilde{N}$ (red circles) for different values of ε with $m = 100$.

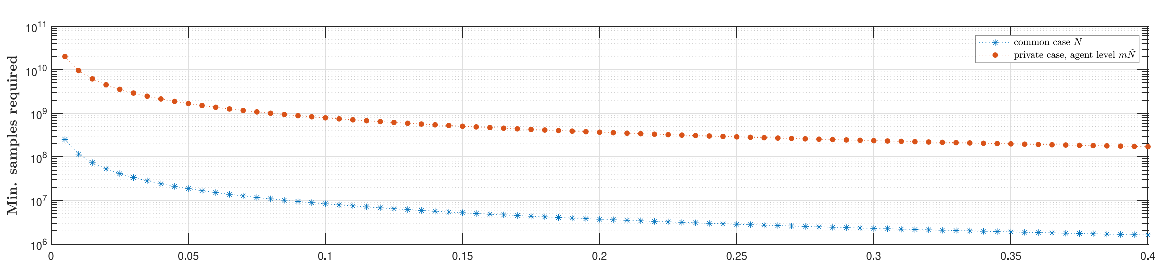

Fig. 5. $\bar{N}$ (blue asterisks) vs. $N = m\tilde{N}$ (red circles) for different values of ε with $m = 100$.

increases as the system grows in scale. However, the complexity of the local constraint set that each agent has to enforce when the dataset is private is comparable and, in some cases, smaller than the case of a common dataset, thus yielding a computational advantage.

5. Conclusions

This paper extends results of statistical learning theory to constraint-coupled multi-agent MIPs affected by uncertainty. Differently from the existing approaches in the literature, we derived a-priori probabilistic feasibility guarantees for any feasible (as opposed to optimal) data-based solution. The proposed guarantees can thus be combined with any decentralized or distributed solution-seeking algorithm originally devised for deterministic MIPs and applied to the data-based formulation of an optimization problem affected by uncertainty. Since the data-based formulation preserves the multi-agent structure, in the case when the problem presents a partially separable structure the computational effort can be eventually reduced by distributing the computational load while preserving the privacy of the local information and uncertainty data.

A comparison with existing guarantees for set-ups with a common dataset suggests that enforcing the privacy of the agents' information generally results in a higher number of total samples to be collected, yet it reduces the number of constraints of the data-driven local problems, and thus their complexity.

Ongoing research focuses on deriving similar results for uncertain decision-coupled multi-agent MIPs where the decision vectors of all agents are identical, without exploiting their reformulation as a constraint-coupled multi-agent MIP by creating m copies of the common decision vector and imposing that they are identical via global coupling constraints.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Alamo, T., Tempo, R., & Camacho, E. (2009). Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, *54*(11), 2545–2559.

Bemporad, A., & Morari, M. (1999). Control of systems integrating logic, dynamics, and constraints. *Automatica*, *35*(3), 407–427.

Ben-Tal, A., & Nemirovski, A. (1997). Robust truss topology design via semidefinite programming. *SIAM Journal on Optimization*, *7*(4), 991–1016.

Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of Operations Research*, *23*(4), 769–805.

Bevly, D., Cao, X., Gordon, M., Ozbilgin, G., Kari, D., Nelson, B., et al. (2016). Lane change and merge maneuvers for connected and automated vehicles: A survey. *IEEE Transactions on Intelligent Vehicles*, *1*(1), 105–120.

Calafiore, G. C., Lyons, D., & Fagiano, L. (2012). On mixed-integer random convex programs. In *2012 IEEE 51st IEEE conference on decision and control* (pp. 3508–3513).

Camisa, A., Notarnicola, I., & Notarstefano, G. (2022). Distributed primal decomposition for large-scale MILPs. *IEEE Transactions on Automatic Control*, *67*(1), 413–420.

Campi, M., Garatti, S., & Prandini, M. (2009). The scenario approach for system and control design. *Elsevier, Annual Reviews in Control*, *33*, 149–150.

Campi, M. C., Garatti, S., & Ramponi, F. A. (2018). A general scenario theory for nonconvex optimization and decision making. *IEEE Transactions on Automatic Control*, *63*(12), 4067–4078.

Esfahani, P. M., Sutter, T., & Lygeros, J. (2014). Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Transactions on Automatic Control*, *60*(1), 46–58.

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

Falsone, A., Margellos, K., & Prandini, M. (2019). A decentralized approach to multi-agent MILPs: Finite-time feasibility and performance guarantees. *Automatica*, *103*, 141–150.

Falsone, A., Margellos, K., Prandini, M., & Garatti, S. (2020). A scenario-based approach to multi-agent optimization with distributed information. *IFAC-PapersOnLine*, *53*(2), 20–25, 21st IFAC World Congress.

Falsone, A., Molinari, F., & Prandini, M. (2020). Uncertain multi-agent MILPs: A data-driven decentralized solution with probabilistic feasibility guarantees. In *Proceedings of machine learning research*: *vol. 120*, *Proceedings of the 2nd conference on learning for dynamics and control* (pp. 1000–1009).

Kwakernaak, H. (1993). Robust control and H-infinity optimization tutorial paper. *Automatica*, *29*(2), 255–273.

La Bella, A., Falsone, A., Ioli, D., Prandini, M., & Scattolini, R. (2021). A mixed-integer distributed approach to prosumers aggregation for providing balancing services. *International Journal of Electrical Power & Energy Systems*, *133*, Article 107228.

Li, K., Li, S. E., Gao, F., Lin, Z., Li, J., & Sun, Q. (2020). Robust distributed consensus control of uncertain multiagents interacted by eigenvalue-bounded topologies. *IEEE Internet of Things Journal*, *7*(5), 3790–3798.

Manieri, L., Falsone, A., & Prandini, M. (2023a). Handling complexity in large scale cyber-physical systems through distributed computation. In *Computation-aware algorithmic design for cyber-physical systems* (pp. 55–78). Springer Nature Switzerland (in press).

Manieri, L., Falsone, A., & Prandini, M. (2023b). A novel decentralized approach to large-scale multi-agent MILPs. In *Proceedings of 22nd World congress of the international federation of automatic control* (pp. 6452–6457) (in press) IFAC PoL.

Pantazis, G., Fele, F., & Margellos, K. (2021). A posteriori probabilistic feasibility guarantees for Nash equilibria in uncertain multi-agent games. In *IFAC world congress* (pp. 3403–3408).

Pantazis, G., Fele, F., & Margellos, K. (2022). On the probabilistic feasibility of solutions in multi-agent optimization problems under uncertainty. *European Journal of Control*, *63*, 186–195.

Papoulis, A., & Pillai, S. (2002). *Probability, random variables and stochastic processes* (4th ed.). McGraw-Hill.

Shen, F., Zhao, L., Du, W., Zhong, W., & Qian, F. (2019). Large-scale industrial energy systems optimization under uncertainty: A data-driven robust optimization approach. *Elsevier, Applied Energy*.

Zhang, B., Li, Q., Wang, L., & Feng, W. (2018). Robust optimization for energy transactions in multi-microgrids under uncertainty. *Elsevier, Applied Energy*.