



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE  
DELLA RICERCA

## Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Distributed Online Aggregative Optimization for Dynamic Multirobot Coordination

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Carnevale, G., Camisa, A., Notarstefano, G. (2023). Distributed Online Aggregative Optimization for Dynamic Multirobot Coordination. IEEE TRANSACTIONS ON AUTOMATIC CONTROL, 68(6), 3736-3743 [10.1109/TAC.2022.3196627].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/907052> since: 2026-05-21

*Published:*

DOI: <http://doi.org/10.1109/TAC.2022.3196627>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

# Distributed Online Aggregative Optimization for Dynamic Multi-robot Coordination

Guido Carnevale, *Graduate Student Member, IEEE*, Andrea Camisa, *Member, IEEE*,  
Giuseppe Notarstefano, *Member, IEEE*

**Abstract**— This paper focuses on an online version of the emerging distributed constrained aggregative optimization framework, which is particularly suited for applications arising in cooperative robotics. Agents in a network want to minimize the sum of local cost functions, each one depending both on a local optimization variable, subject to a local constraint, and on an aggregated version of all the variables (e.g., the mean). We focus on a challenging online scenario in which the cost, the aggregation functions and the constraints can all change over time, thus enlarging the class of captured applications. Inspired by an existing scheme, we propose a distributed algorithm with constant step size, named Projected Aggregative Tracking, to solve the online optimization problem. We prove that the dynamic regret is bounded by a constant term and a term related to time variations. Moreover, in the static case (i.e., with constant cost and constraints), the solution estimates are proved to converge with a linear rate to the optimal solution. Finally, numerical examples show the efficacy of the proposed approach on a robotic surveillance scenario.

**Index Terms**— Cooperative Control, Distributed Optimization, Optimization algorithms

## I. INTRODUCTION

Distributed optimization captures a variety of estimation and learning problems over networks, including distributed data classification and localization in smart sensor networks, to name a few. The term “online” refers to scenarios in which the problem data is not available a-priori, but rather it arrives dynamically while the optimization process is executed. In this paper, we consider an online distributed optimization set-up in which agents in a network must cooperatively minimize the sum of local cost functions that depend both on a local optimization variable and on a global variable obtained by performing some kind of aggregation of all the local variables (as, e.g., the mean). This *aggregative optimization* set-up was introduced in the pioneering work [1]. The framework is fairly general and embraces several control applications of interest such as cooperative robotics and multi-vehicle surveillance. Originally, it stems from distributed aggregative games [2]–[6], where however the objective is to compute a (generalized) Nash equilibrium rather than an optimal solution cooperatively.

G. Carnevale, A. Camisa and G. Notarstefano are with the Department of Electrical, Electronic and Information Engineering, University of Bologna, Bologna, Italy, {guido.carnevale, a.camisa, giuseppe.notarstefano}@unibo.it. This result is part of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 638992 - OPT4SMART).

There exists a vast literature on distributed online optimization, which addresses two main optimization set-ups known as cost-coupled (or consensus optimization) and constraint-coupled, see, e.g., [7]. In the cost-coupled framework the goal is to optimize a cost function given by the sum of several local functions with a common decision variable. Distributed online algorithms based on subgradient schemes are proposed in [8]–[10], while primal-dual or dual approaches are proposed in [11]–[13]. The use of a model for the minimum variation is considered in [14], where a mirror descent algorithm is proposed. Distributed online optimization is used in [15] to handle the distribution grids problem. In [16] an online algorithm based on the alternating direction method of multipliers is proposed. In the more recent constraint-coupled optimization framework, each local objective function depends on a local decision variable, but all the variables are coupled through separable coupling constraints. In [17] this set-up is addressed in an online setting and a sublinear regret bound is ensured by using a distributed primal dual algorithm. The same result is achieved in [18] by introducing a push-sum mechanism to allow for directed graph topologies. Time-varying inequality constraints have been taken into account in [19], in which a distributed primal dual mirror descent algorithm is proposed.

We recall that the above works are suited for cost-coupled and constraint-coupled set-ups. The aggregative optimization framework addressed in this paper has been recently introduced in works [1], [20] which consider respectively a static unconstrained framework and an online constrained one. The distributed algorithms proposed in these two papers leverage a tracking action to reconstruct both the aggregative variable and the gradient of the whole cost function. This “tracking action” is based on dynamic average consensus (see [21], [22]) and has been introduced in the gradient tracking scheme for cost-coupled optimization [23]–[31]. The gradient tracking has been applied to online optimization in [32], in [33] where partially unknown cost functions are considered, and in [34] where adaptive momenta are used.

The contributions of this paper are as follows. We focus on an online constrained aggregative optimization set-up over peer-to-peer networks of agents inspired to the set-up studied in the seminal work [20]. Despite the fact that we do not assume boundedness of the gradients and the feasible sets, we demonstrate stronger theoretical results than the state of art, i.e., tighter regret bounds and, by using our method in a time-invariant problem, linear convergence rate (instead of a sublinear one) to the optimal solution. Additionally, the

optimization set-up we consider enlarges the scope of previous works. Indeed, we consider a wider time-varying framework in which also the feasible sets and the aggregation rules vary over time. These generalizations introduce additional terms in the regret analysis and thus pose new challenges that must be appropriately handled. In Projected Aggregative Tracking, each agent projects the updated local solution estimate on its time-varying constraint set and then performs a convex combination with the current estimate. Moreover, the trackers of the aggregative variable are generalized to handle time-variation of the aggregation rules. Under mild assumptions on the cost and constraint variations, we provide a bound about the dynamic regret for the proposed scheme. We additionally provide a regret result on the violation of the time-varying constraints. In order to obtain this result we study a dynamical system describing the algorithmic evolution of: (i) the error of the solution estimate with respect to the minimum, (ii) the consensus error of the aggregative variable trackers, and (iii) the consensus error of the global gradient trackers. Such dynamics is characterized by a Schur system matrix, by which we are then able to draw conclusions on the dynamic regret. For the static case (with constant cost and constraints), we show that the algorithm iterates converge to the optimal solution with linear rate. Notably, our algorithm allows for a constant step-size. To corroborate the theoretical analysis, we show numerical simulations from a cooperative robotics scenario in which robots have to accomplish a surveillance task. The proposed scenario is dynamic and intrinsically characterized by time-varying cost functions, aggregation functions and constraints, thus it cannot be addressed by using state-of-art techniques.

The rest of the paper is organized as follows. Section II describes the distributed online aggregative optimization framework and presents the distributed algorithm with its convergence properties. The algorithm analysis is performed in Section III. Finally, Section IV shows the effectiveness of our method.

*Notation:* We use  $\text{COL}(v_1, \dots, v_n)$  to denote the vertical concatenation of the column vectors  $v_1, \dots, v_n$ . We use  $\text{blkdiag}(M_1, \dots, M_N)$  to denote the block diagonal matrix where the  $i$ -th diagonal block is given by the matrix  $M_i \in \mathbb{R}^{n_i \times m_i}$  for all  $i \in \{1, \dots, N\}$ . The Kronecker product is denoted by  $\otimes$ . The identity matrix in  $\mathbb{R}^{m \times m}$  is  $I_m$ , while  $0_m$  is the zero matrix in  $\mathbb{R}^{m \times m}$ . The column vector of  $N$  ones is denoted by  $\mathbf{1}_N$  and we define  $\mathbf{1} \triangleq \mathbf{1}_N \otimes I_d$ . Dimensions are omitted whenever they are clear from the context. Given a closed and convex set  $X$ , we use  $P_X[y]$  to denote the projection of a vector  $y$  on a  $X$ , namely  $P_X[y] = \arg \min_{x \in X} \|x - y\|$ , while we use  $\text{dist}(y, X)$  to denote its distance from the set, namely  $\text{dist}(y, X) = \min_{x \in X} \|x - y\|$ . Given  $x \in \mathbb{R}^n$ , we use  $[x]^+$  to denote  $\max\{0, x\}$  in a component-wise sense. Let  $M \in \mathbb{R}^{n \times n}$ , then we denote as  $\rho_{\max}(M)$  its spectral radius.

## II. PROBLEM FORMULATION AND ALGORITHM DESCRIPTION

In this paper, we consider distributed online aggregative optimization problems that can be written as

$$\min_{(x_1, \dots, x_N) \in X_t} \sum_{i=1}^N f_{i,t}(x_i, \sigma_t(x)) \quad (1)$$

in which  $x := \text{COL}(x_1, \dots, x_N) \in \mathbb{R}^n$  is the global decision vector, with each  $x_i \in \mathbb{R}^{n_i}$  and  $n = \sum_{i=1}^N n_i$ . The global decision vector at time  $t$  is constrained to belong to a set  $X_t \subseteq \mathbb{R}^n$  that can be written as  $X_t = (X_{1,t} \times \dots \times X_{N,t})$ , where each  $X_{i,t} \subseteq \mathbb{R}^{n_i}$ . The functions  $f_{i,t} : \mathbb{R}^{n_i} \times \mathbb{R}^d \rightarrow \mathbb{R}$  represent the local objective functions at time  $t$ , while the *aggregation function*  $\sigma_t(x)$  has the form

$$\sigma_t(x) := \frac{\sum_{i=1}^N \phi_{i,t}(x_i)}{N}, \quad (2)$$

where each  $\phi_{i,t} : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^d$  is the  $i$ -th contribution to the aggregative variable at time  $t$ . We compactly denote the cost function of problem (1) as  $f_t(x, \sigma_t(x)) := \sum_{i=1}^N f_{i,t}(x_i, \sigma_t(x))$ . In problem (1),  $f_t(\cdot, \sigma_t(\cdot))$  is not known to any agent: each of them can only privately access  $f_{i,t}$ ,  $X_{i,t}$ , and  $\phi_{i,t}$ . We remark that each agent  $i$  accesses its private information  $f_{i,t}$ , and  $\phi_{i,t}$  only once its estimate  $x_{i,t}$  has been computed. The idea is to solve problem (1) in a distributed way over a network of  $N$  agents communicating according to a graph  $\mathcal{G} := (\{1, \dots, N\}, \mathcal{E}, \mathcal{A})$ , where  $\{1, \dots, N\}$  is the set of agents,  $\mathcal{E} \in \{1, \dots, N\} \times \{1, \dots, N\}$  is the edges set, and  $\mathcal{A} \in \mathbb{R}^{N \times N}$  is the weighted adjacency matrix. Each agent  $i$  can exchange data only with its neighbors defined by  $\mathcal{E}$ .

The goal is to design distributed algorithms to seek a minimum for problem (1). Next, we will denote as  $\nabla_1 f_{i,t}(\cdot, \cdot)$  and as  $\nabla_2 f_{i,t}(\cdot, \cdot)$  the gradient of  $f_{i,t}$  with respect to respectively the first argument and the second argument. Moreover, we also introduce  $G_t : \mathbb{R}^n \times \mathbb{R}^{Nd} \rightarrow \mathbb{R}^n$  defined as  $G_t(x, s) := \nabla_1 f_t(x, s) + \nabla \phi(x) \frac{1}{N} \sum_{i=1}^N f_{i,t}(x_i, s_i)$ , where  $x := \text{COL}(x_1, \dots, x_N) \in \mathbb{R}^n$ ,  $s := \text{COL}(s_1, \dots, s_N) \in \mathbb{R}^{Nd}$  with each  $x_i \in \mathbb{R}^{n_i}$ ,  $s_i \in \mathbb{R}^d$  for all  $i \in \{1, \dots, N\}$ ,  $\nabla_1 f_t(x, s) := \text{COL}(\nabla_1 f_{1,t}(x_1, s_1), \dots, \nabla_1 f_{N,t}(x_N, s_N))$ , and  $\nabla \phi(x) := \text{blkdiag}(\nabla \phi_1(x_1), \dots, \nabla \phi_N(x_N)) \in \mathbb{R}^{n \times Nd}$ .

Let  $x_{i,t}$  be the solution estimate of the problem at time  $t$  maintained by agent  $i$ , and let  $x_t^*$  be the (unique) minimizer of  $f_t(x, \sigma_t(x))$  over the set  $X_t$ . Indeed, as we will formalize within Assumption 2, strong convexity of  $f_t(x, \sigma_t(x))$  guarantees existence (and uniqueness) of  $x_t^*$ . Then, given a finite value  $T > 1$ , the agents want to minimize the dynamic regret:

$$R_T := \sum_{t=1}^T f_t(x_t, \sigma_t(x_t)) - \sum_{t=1}^T f_t(x_t^*, \sigma_t(x_t^*)). \quad (3)$$

Another popular metric is the so-called static regret [12]. However, as done in most of the literature, we focus on (3), which is more challenging to handle. To this end, we propose our Projected Aggregative Tracking algorithm. Each agent  $i$  maintains for each time instant  $t$  an estimate  $x_{i,t}$  of the component  $i$  of a minimum  $x_t^*$  of problem (1). In order to reconstruct the descent direction and use it to update the estimate  $x_{i,t}$ , agent  $i$  needs to reconstruct the global information  $\sum_{i=1}^N \frac{\phi_{i,t}(x_{i,t})}{N}$  and  $\sum_{i=1}^N \nabla_2 f_{i,t} \left( x_{i,t}, \sum_{j=1}^N \frac{\phi_{j,t}(x_{j,t})}{N} \right)$ , which are not locally available. To overcome this lack of information, agent  $i$  maintains auxiliary variables  $s_{i,t}$  and  $y_{i,t}$  and iteratively updates them according to a perturbed consensus mechanism. A pseudo-code of the Projected Aggregative Tracking algorithm is reported in Algorithm 1 from the perspective of agent  $i$ , in which  $\alpha$  is a positive constant step-size,  $\delta \in (0, 1)$  is a constant

algorithm parameter, and each element  $a_{ij}$  represents the  $(i, j)$  entry of the weighted adjacency matrix  $\mathcal{A}$  of the network.

---

**Algorithm 1** Projected Aggregative Tracking (Agent  $i$ )

---

initialization:

$$x_{i,0} \in X_{i,0}, \quad s_{i,0} = \phi_{i,0}(x_{i,0}), \quad y_{i,0} = \nabla_2 f_{i,0}(x_{i,0}, s_{i,0})$$

**for**  $t = 0, 1, \dots$  **do**

$$\tilde{x}_{i,t} = P_{X_{i,t}}[x_{i,t} - \alpha(\nabla_1 f_{i,t}(x_{i,t}, s_{i,t}) + \nabla \phi_{i,t}(x_{i,t})y_{i,t})]$$

$$x_{i,t+1} = x_{i,t} + \delta(\tilde{x}_{i,t} - x_{i,t})$$

$$s_{i,t+1} = \sum_{j=1}^N a_{ij} s_{j,t} + \phi_{i,t+1}(x_{i,t+1}) - \phi_{i,t}(x_{i,t})$$

$$y_{i,t+1} = \sum_{j=1}^N a_{ij} y_{j,t} + \nabla_2 f_{i,t+1}(x_{i,t+1}, s_{i,t+1}) - \nabla_2 f_{i,t}(x_{i,t}, s_{i,t})$$

**end for**

---

### III. CONVERGENCE ANALYSIS

This section gives the convergence properties of the proposed distributed algorithm.

#### A. Projected Aggregative Tracking Reformulation and Assumptions

First of all let us rewrite all the agents' updates, according to Algorithm 1, in a stacked vector form as

$$\tilde{x}_t = P_{X_t}[x_t - \alpha(\nabla_1 f_t(x_t, s_t) + \nabla \phi_t(x_t)y_t)] \quad (4a)$$

$$x_{t+1} = x_t + \delta(\tilde{x}_t - x_t) \quad (4b)$$

$$s_{t+1} = A s_t + \phi_{t+1}(x_{t+1}) - \phi_t(x_t) \quad (4c)$$

$$y_{t+1} = A y_t + \nabla_2 f_{t+1}(x_{t+1}, s_{t+1}) - \nabla_2 f_t(x_t, s_t), \quad (4d)$$

where we used the notation  $x_t := \text{COL}(x_{1,t}, \dots, x_{N,t})$ ,  $s_t := \text{COL}(s_{1,t}, \dots, s_{N,t})$ , and  $y_t := \text{COL}(y_{1,t}, \dots, y_{N,t})$ . Moreover, we also introduced the symbols  $\nabla_2 f_t(x_t, s_t) := \text{COL}(\nabla_2 f_{1,t}(x_{1,t}, s_{1,t}), \dots, \nabla_2 f_{N,t}(x_{N,t}, s_{N,t}))$ ,  $\nabla \phi_t(x_t) := \text{blkdiag}(\nabla \phi_{1,t}(x_{1,t}), \dots, \nabla \phi_{N,t}(x_{N,t}))$  and  $A := \mathcal{A} \otimes I$ . In order to perform the convergence analysis, we derive bounds for the quantities  $\|x_{t+1} - x_t\|$ ,  $\|x_{t+1} - x_{t+1}^*\|$ ,  $\|y_{t+1} - \mathbf{1}\bar{y}_{t+1}\|$ , and  $\|s_{t+1} - \mathbf{1}\bar{s}_{t+1}\|$ , in which  $\bar{y}_t := \frac{1}{N} \sum_{i=1}^N y_{i,t}$  and  $\bar{s}_t := \frac{1}{N} \sum_{i=1}^N s_{i,t}$  denote the mean vectors of  $y_t$  and  $s_t$ , respectively. Let  $z_t$  be the vector stacking the above quantities

$$z_t := \begin{bmatrix} \|x_t - x_t^*\| \\ \|s_t - \mathbf{1}\bar{s}_t\| \\ \|y_t - \mathbf{1}\bar{y}_t\| \end{bmatrix}. \quad (5)$$

Moreover, also the following variables will be useful to provide the main result of the paper, namely

$$\eta_t := \sup_{x \in \mathbb{R}^n, z \in \mathbb{R}^{Nd}} \|\nabla_2 f_{t+1}(x, z) - \nabla_2 f_t(x, z)\| \quad (6a)$$

$$\omega_t := \sup_{x \in \mathbb{R}^n} \|\phi_{t+1}(x) - \phi_t(x)\| \quad (6b)$$

$$\gamma_t := \sup_{x \in \mathbb{R}^n} |\text{dist}(x, X_{t+1}) - \text{dist}(x, X_t)| \quad (6c)$$

$$\zeta_t := \|x_{t+1}^* - x_t^*\|, \quad (6d)$$

where we recall that  $x_t^*$  is the optimal solution of  $f_t(x_t, \sigma_t(x_t))$ . Next, we state the assumptions of our framework.

**Assumption 1** (Communication graph). *The graph  $\mathcal{G}$  is undirected and connected and  $\mathcal{A}$  is doubly stochastic.*  $\square$

**Assumption 2** (Convexity). *For all  $i \in \{1, \dots, N\}$  and all  $t \geq 0$ ,  $X_{i,t} \subseteq \mathbb{R}^{n_i}$  is nonempty, closed and convex, while the global objective function  $f_t(x, \sigma_t(x))$  is  $\mu$ -strongly convex.*  $\square$

**Assumption 3** (Function Regularity). *For all  $t \geq 0$ , the function  $f_t(x, \sigma_t(x))$  is differentiable with  $L_1$ -Lipschitz continuous gradients, and  $G_t(x, s)$ ,  $\nabla_2 f_t(x, s)$  are Lipschitz continuous with constants  $L_1, L_2 > 0$ , respectively. For all  $i \in \{1, \dots, N\}$  and  $t \geq 0$ , the aggregation function  $\phi_{i,t}(x_i)$  is differentiable and  $L_3$ -Lipschitz continuous, and  $\eta_t$  and  $\omega_t$  are finite.*  $\square$

We start by noting that

$$\bar{s}_{t+1} = \bar{s}_t + \frac{\mathbf{1}^\top}{N} (\phi_{t+1}(x_{t+1}) - \phi_t(x_t)) \quad (7a)$$

$$\bar{y}_{t+1} = \bar{y}_t + \frac{\mathbf{1}^\top}{N} (\nabla_2 f_{t+1}(x_{t+1}, s_{t+1}) - \nabla_2 f_t(x_t, s_t)). \quad (7b)$$

Then, if we initialize  $\sigma$  and  $y$  as  $\sigma_0 := \phi_0(x_0)$  and  $y_0 := \nabla_2 f_0(x_0, s_0)$ , from (7a) and (7b), it holds for all  $t \geq 0$

$$\bar{s}_t = \frac{1}{N} \sum_{i=1}^N \phi_t(x_{i,t}) := \sigma_t(x_t) \quad (8a)$$

$$\bar{y}_t = \frac{1}{N} \sum_{i=1}^N \nabla_2 f_{i,t}(x_{i,t}, s_{i,t}). \quad (8b)$$

#### B. Preparatory Lemmas

Here we present four preparatory Lemmas that we need to prove Theorem 1. For brevity, we will use  $d_t$  to denote the descent direction used within the update (4a), i.e.,

$$d_t := \nabla_1 f_t(x_t, s_t) + \nabla \phi_t(x_t)y_t. \quad (9)$$

**Lemma 1.** *Let Assumptions 1, 2, and 3 hold. If  $\alpha \leq \frac{1}{L_1}$ , then*

$$\|x_{t+1} - x_{t+1}^*\| \leq (1 - \delta\mu\alpha) \|x_t - x_t^*\| + \delta\alpha L_1 \|s_t - \mathbf{1}\bar{s}_t\| + \delta\alpha L_3 \|y_t - \mathbf{1}\bar{y}_t\| + \zeta_t.$$

*Proof.* The proof is provided in Appendix A.  $\square$

**Lemma 2.** *Let Assumptions 1, 2, and 3 hold. Then*

$$\|x_{t+1} - x_t\| \leq \delta(2 + \alpha L_1 + \alpha L_1 L_3) \|x_t - x_t^*\| + \delta\alpha L_1 \|s_t - \mathbf{1}\bar{s}_t\| + \delta\alpha L_3 \|y_t - \mathbf{1}\bar{y}_t\|.$$

*Proof.* The proof is provided in Appendix B.  $\square$

**Lemma 3.** *Let Assumptions 1, 2, and 3 hold. Then*

$$\|s_{t+1} - \mathbf{1}\bar{s}_{t+1}\| \leq \Lambda \|s_t - \mathbf{1}\bar{s}_t\| + \delta\alpha L_1 L_3 \|s_t - \mathbf{1}\bar{s}_t\| + \delta(2L_3 + \alpha L_1 L_3 + \alpha L_1 L_3^2) \|x_t - x_t^*\| + \delta\alpha L_3^2 \|y_t - \mathbf{1}\bar{y}_t\| + \omega_t,$$

where  $\Lambda$  is the maximum eigenvalue of the matrix  $A - \frac{\mathbf{1}\mathbf{1}^\top}{N}$ .

*Proof.* The proof is provided in Appendix C.  $\square$

**Lemma 4.** *Let Assumptions 1, 2, and 3 hold. Then*

$$\begin{aligned} \|y_{t+1} - \mathbf{1}\bar{y}_{t+1}\| &\leq \Lambda \|y_t - \mathbf{1}\bar{y}_t\| \\ &+ \delta\alpha L_3(L_2 + L_2 L_3) \|y_t - \mathbf{1}\bar{y}_t\| \\ &+ \delta(2 + \alpha L_1 + \alpha L_1)(L_2 + L_2 L_3) \|x_t - x_t^*\| \\ &+ \delta\alpha L_1(L_2 + L_2 L_3) \|s_t - \mathbf{1}\bar{s}_t\| \\ &+ 2L_2 \|s_t - \mathbf{1}\bar{s}_t\| + L_2\omega_t + \eta_t, \end{aligned}$$

where  $\Lambda$  is the maximum eigenvalue of the matrix  $A - \frac{\mathbf{1}\mathbf{1}^\top}{N}$ .

*Proof.* The proof is provided in Appendix D.  $\square$

### C. Regret analysis and linear rate in static set-up

Now, we state the main theoretical results of the paper. Next theorem provides a bound on the dynamic regret of the iterates generated by the Projected Aggregative Tracking distributed algorithm in the general, online set-up (1).

**Theorem 1.** *Consider Projected Aggregative Tracking as given in Algorithm 1. Let Assumptions 1, 2, and 3 hold. Then, there exists  $\lambda, \bar{\delta} > 0$  and  $\bar{\rho} \in (0, 1)$  so that, if  $\alpha \leq \frac{1}{L_1}$  and  $\delta \in (0, \bar{\delta})$ , it holds*

$$R_T \leq \frac{L_1\lambda^2}{2} \left( \frac{\|z_0\|^2}{1 - \bar{\rho}^2} + 2\|z_0\| U_T + Q_T \right), \quad (10)$$

where  $R_T$  is defined as in (3) and

$$U_T := \sum_{t=1}^T \sum_{k=0}^{t-1} \bar{\rho}^{t+k} \left( \|\zeta_{t-k-1}\| + 2\|\eta_{t-k-1}\| + (1 + L_2) \|\omega_{t-k-1}\| \right), \quad (11a)$$

$$Q_T := \sum_{t=1}^T \sum_{k=0}^{t-1} \bar{\rho}^{2k} (\zeta_{t-k-1}^2 + 2\eta_{t-k-1}^2 + (1 + L_2)\omega_{t-k-1}^2). \quad (11b)$$

Moreover, if  $\gamma_t$  (cf. (6c)) is finite for all  $t \geq 0$ , then the constraint violation is bounded by

$$\begin{aligned} \sum_{t=1}^T \text{dist}(x_t, X_t) &\leq \frac{1}{1 - (1 - \delta)^T} \text{dist}(x_0, X_0) \\ &+ \sum_{t=1}^T \sum_{k=0}^{t-1} (1 - \delta)^k \gamma_{t-k-1}. \end{aligned} \quad (12)$$

*Proof.* The proof is provided in Appendix E.  $\square$

Operatively, in order to choose an appropriate value of the parameter  $\delta$ , it is necessary to first estimate the upper bound  $\bar{\delta}$ . As it emerges from the proof of Theorem 1, this can be done as follows: (i) compute a matrix  $M(\delta)$  (cf. (21)), which depends on the various problem constants and on  $\delta$ , (ii) compute  $\bar{\delta}$  as the maximum value of  $\delta$  such that all the eigenvalues of  $M(\delta)$  are strictly in the unit circle. We observe that Theorem 1 improves the dynamic regret bound provided in [20], which demonstrates a bound of the type  $O(T) + O(\sqrt{T}V_T)$  (where  $V_T$  is a term capturing variations of the problem). The authors also show that there exists a particular, constant step-size that allows to tighten the first term to  $O(\sqrt{T})$ . However, the choice

of the the step-size requires a prior knowledge of  $T$  and  $V_T$ . In both cases, we improve the first term, which is replaced by the constant  $\frac{L_1\lambda^2\|z_0\|^2}{2(1-\bar{\rho}^2)}$ , while in our terms  $U_T$  and  $Q_T$  (cf. (11)) the variations of the problem are scaled by  $\bar{\rho}^t$ , i.e., an exponentially decaying quantity since  $\bar{\rho} \in (0, 1)$ .

**Remark 1** (Average Regret). *Let us consider the case in which the problem variations are bounded by a constant, i.e., suppose there exists  $C > 0$  so that  $\zeta_t, \eta_t, \omega_t \leq C$  for all  $t \geq 0$ . In this case, by using the definitions of  $U_T$  and  $Q_T$  (cf. (11)) and recalling that  $\bar{\rho} \in (0, 1)$ , we can use the geometric series property to get*

$$U_T \leq \frac{(4 + L_2)C}{1 - \bar{\rho}^T}, \quad \text{and} \quad Q_T \leq \frac{(3 + (1 + L_2))C^2 T}{1 - \bar{\rho}^2}.$$

In this case, the average regret approaches a constant value,

$$\lim_{T \rightarrow \infty} R_T/T = \frac{L_1\lambda^2(4 + L_2)C^2}{2(1 - \bar{\rho}^2)}. \quad \square$$

**Remark 2** (Inequality constraints). *Consider the case in which  $X_{i,t}$  can be expressed in terms of inequality constraints, namely*

$$X_{i,t} := \{x_i \in \mathbb{R}^{n_i} \mid h_{i,t}(x_i) \leq 0_{m_i}\},$$

with  $h_{i,t} : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{m_i}$  for all  $i \in \{1, \dots, N\}$  and  $t \geq 0$ . In this case, in place of the distance function  $\text{dist}(x, X_t)$ , one can use  $\| [h_t(x)]^+ \|$  as a metric to characterize the constraint violation, where  $h_t(x) := \text{COL}(h_{1,t}(x_1), \dots, h_{N,t}(x_N))$ . By repeating similar arguments as in the proof of Theorem 1, one obtains similarly that  $\sum_{t=1}^T \| [h_t(x_t)]^+ \| \leq \frac{1}{1 - (1 - \delta)^T} \| [h_t(x_0)]^+ \| + \sum_{t=1}^T \sum_{k=0}^{t-1} (1 - \delta)^k \gamma_{t-k-1}$ .  $\square$

In the following corollary, we assess that in the static case the Projected Aggregative Tracking distributed algorithm converges to the (fixed) optimal solution  $x^*$  with a linear rate.<sup>1</sup>

**Corollary 1** (Static set-up). *Under the same assumptions of Theorem 1, if it holds  $f_t = f$ ,  $\phi_t = \phi$ , and  $X_{i,t} = X_i$  for all  $i \in \{1, \dots, N\}$  and all  $t \geq 0$ , then there exists  $\lambda, \bar{\delta} > 0$  and  $\bar{\rho} \in (0, 1)$  so that, if  $\alpha \leq \frac{1}{L_1}$  and  $\delta \in (0, \bar{\delta})$ , it holds*

$$f(x_t, \sigma(x_t)) - f(x^*, \sigma(x^*)) \leq \bar{\rho}^{2t} \frac{L_1\lambda^2}{2} \|z_0\|^2.$$

*Proof.* The proof is provided in Appendix F.  $\square$

## IV. NUMERICAL EXPERIMENTS

In this section we show the effectiveness of Projected Aggregative Tracking on a multi-robot surveillance scenario.

**1) Online set-up:** Let us consider a network of cooperating robots that aim to protect a target with location  $b_t \in \mathbb{R}^2$  at time  $t$  from some intruders. The optimization variables  $x_{i,t} \in \mathbb{R}^2$  represent the position of robots at each time  $t$  and each robot  $i$  is able to move from  $x_{i,t}$  to  $x_{i,t+1}$  using a local controller. We associate to each robot  $i$  an intruder located at  $p_{i,t} \in \mathbb{R}^2$  at time  $t$ . The dynamic protection strategy applied by each robot consists of staying simultaneously close to the protected target and to the associated intruder. Meanwhile, the whole team

<sup>1</sup>A sequence  $\{x_t\}$  converges linearly to  $\bar{x}$  if there exists a number  $\eta \in (0, 1)$  such that  $\frac{\|x_t - \bar{x}\|}{\|x_{t-1} - \bar{x}\|} \rightarrow \eta$  as  $t \rightarrow \infty$ .

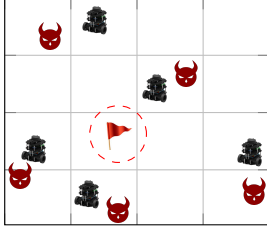


Fig. 1: Multi-robot surveillance scenario - Robot icons denote agents, devil icons denote intruders, while the flag is the target to be protected.

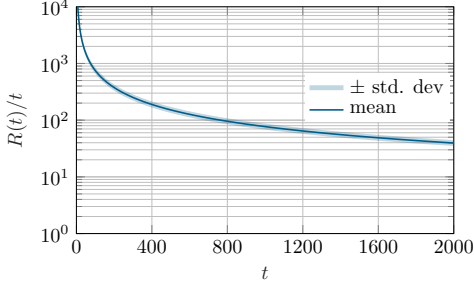


Fig. 2: Online case – Mean of the average Dynamic regret and 1-standard deviation band over 100 Monte Carlo trials.

of robots tries to keep its weighted center of mass rotating close to the target. A concept of this scenario is given in Fig. (1). This strategy is obtained by solving problem (1) with the cost functions  $f_{i,t}(x_i, \sigma_t(x)) = \frac{1}{2} \|x_i - p_{i,t}\|^2 + \frac{\gamma_1}{2} \|x_i - b_t\|^2 + \frac{\gamma_2}{2N} \|\sigma_t(x) - b_t\|^2$ , with  $\gamma_1 = 1$ ,  $\gamma_2 = 10$  and the aggregation rules  $\phi_{i,t}(x_i) = \beta_i x_i + a_t$ , where  $\beta_i > 0$  and  $a_t \in \mathbb{R}^2$  represents a time-varying offset which follows the law  $a_t = r \text{COL}(\cos(t/(2\pi\tau)), \sin(t/(2\pi\tau)))$  for some  $r, \tau > 0$ . In this way, the center of mass  $\frac{1}{N} \sum_{i=1}^N x_i^t$  is forced to rotate around the target position  $b_t$ .

We address a scenario with  $N = 50$  agents and intruders. As regards the constraints, we consider a common time-varying box  $X_{i,t} = \{x \in \mathbb{R}^2 \mid 0 \leq x \leq u_t\}$  for all  $i$ , where  $u_t \in \mathbb{R}^2$  starts from  $[20, 20]$  and linearly increases at each iteration. In this way, the agents initially stay closer to the target and then they move toward the associated intruders. Each intruder  $i$  moves along a circle of radius  $r = 1$  according to the law  $p_{i,t} = p_{i,c} + r \text{COL}(\cos(t/100), \sin(t/100))$ , where  $p_{i,c} \in \mathbb{R}^2$  is randomly generated. The target  $b_t$  and the offset  $a_t$  follow similar laws. In this setup, being the sinusoidal functions bounded, the constants  $\eta_t$  and  $\omega_t$  introduced in (6) can be uniformly bounded as  $\eta_t \leq \gamma_2 \sqrt{N} r$  and  $\omega_t \leq \sqrt{N} r$  for all  $t \geq 0$ . Moreover, the vector  $u_t$  defining the box  $X_{i,t}$  changes linearly with respect time and, thus, also the constant  $\gamma_t$  (cf. (6c)) can be uniformly bounded. As regards the algorithm parameters, we set  $\alpha = 1$  and  $\delta = 0.5$ . We performed 100 Monte Carlo trials that differ in the problem parameters and agents' initial conditions. Fig. 2 shows that the behavior of the algorithm does not depend on the generated instances. Indeed, the achieved average dynamic regret, as predicted in Remark 1, converges asymptotically to a constant.

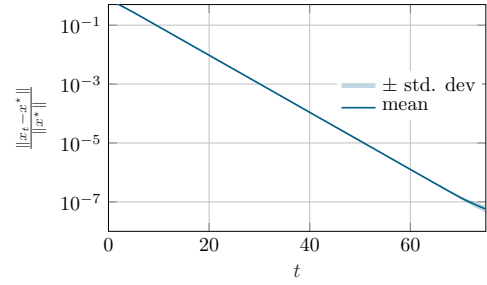


Fig. 3: Static case – Mean of the relative error and 1-standard deviation band obtained with 100 Monte Carlo trials.

2) *Static set-up*: Now we address a static instance of the problem. Namely, we fix  $X_t$  and the positions of the intruders and of the target. We perform a Monte Carlo simulation consisting of 100 trials on the same network of  $N = 50$  agents with the same algorithm parameters. As predicted by Corollary 1, Fig. 3 shows an exponential decay of  $\frac{\|x_t - x^*\|}{\|x^*\|}$ .

## V. CONCLUSIONS

In this paper, we focused on online instances of the distributed constrained aggregative optimization framework. We proposed Projected Aggregative Tracking, a distributed algorithm allows for time-varying feasible sets and aggregation rules. We perform a regret analysis of the scheme by which we conclude that the dynamic regret is bounded by a constant term and a term related to time variations, while in the static case, the solution estimates linearly converge to the optimal solution. Numerical computations confirmed our findings.

## APPENDIX

### A. Proof of Lemma 1

We begin by using (4b), which leads to

$$\begin{aligned} \|x_{t+1} - x_{t+1}^*\| &= \|x_t + \delta(\tilde{x}_t - x_t) - x_{t+1}^*\| \\ &\stackrel{(a)}{\leq} \|x_t + \delta(\tilde{x}_t - x_t) - x_t^*\| + \|x_{t+1}^* - x_t^*\| \\ &\stackrel{(b)}{\leq} \|x_t + \delta(\tilde{x}_t - x_t) - x_t^*\| + \zeta_t, \end{aligned} \quad (13)$$

where in (a) we add and subtract the term  $x_t^*$  and use the triangle inequality, and in (b) we use  $\zeta_t$  (cf (6d)). Being  $x_t^*$  the minimizer of  $f_t$  over  $X_t$ , then it holds  $P_{X_t}[x_t^* - \alpha f_t(x_t^*, \sigma_t(x_t^*))] = x_t^*$ . Then, we add the null term  $\delta(P_{X_t}[x_t^* - \alpha \nabla f_t(x_t^*, \sigma_t(x_t^*))] - x_t^*)$  in the first norm of (13) and we apply the triangle inequality and (4a) to write

$$\begin{aligned} \|x_{t+1} - x_{t+1}^*\| &\leq (1 - \delta) \|x_t - x_t^*\| \\ &\quad + \delta \|P_{X_t}[x_t - \alpha d_t] - P_{X_t}[x_t^* - \alpha \nabla f_t(x_t^*, \sigma_t(x_t^*))]\| + \zeta_t \\ &\stackrel{(a)}{\leq} (1 - \delta) \|x_t - x_t^*\| \\ &\quad + \delta \|x_t - \alpha d_t - (x_t^* - \alpha \nabla f_t(x_t^*, \sigma_t(x_t^*)))\| + \zeta_t, \end{aligned} \quad (14)$$

where (a) uses the non-expansiveness of the projection, see [35]. Add and subtract within the second norm  $\alpha \nabla f_t(x_t, \sigma_t(x_t))$

and apply the triangle inequality to rewrite (14) as

$$\begin{aligned} & \|x_{t+1} - x_{t+1}^*\| \leq (1 - \delta) \|x_t - x_t^*\| \\ & + \delta \|x_t - \alpha \nabla f_t(x_t, \sigma_t(x_t)) - (x_t^* - \alpha f_t(x_t^*, \sigma_t(x_t^*)))\| \\ & + \delta \alpha \|d_t - \nabla f_t(x_t, \sigma_t(x_t))\| + \zeta_t \\ & \stackrel{(a)}{\leq} (1 - \delta \mu \alpha) \|x_t - x_t^*\| + \delta \alpha \|d_t - \nabla f_t(x_t, \sigma_t(x_t))\| + \zeta_t, \end{aligned}$$

where (a) uses [1, Lemma 3]. Add and subtract into the second norm  $\nabla \phi_t(x_t) \mathbf{1} \frac{1}{N} \sum_{i=1}^N \nabla_2 f_{i,t}(x_{i,t}, s_{i,t})$ , and rearrange as

$$\begin{aligned} & \|x_{t+1} - x_{t+1}^*\| \leq (1 - \delta \mu \alpha) \|x_t - x_t^*\| \\ & + \delta \alpha \|G_t(x_t, s_t) - \nabla f(x_t)\| \\ & + \delta \alpha \left\| \nabla \phi_t(x_t) \left( y_t - \mathbf{1} \frac{1}{N} \sum_{i=1}^N \nabla_2 f_{i,t}(x_{i,t}, s_{i,t}) \right) \right\| + \zeta_t \\ & \stackrel{(a)}{=} (1 - \delta \mu \alpha) \|x_t - x_t^*\| + \delta \alpha \|G_t(x_t, s_t) - \nabla f_t(x_t, \sigma_t(x_t))\| \\ & + \delta \alpha \|\nabla \phi_t(x_t) (y_t - \mathbf{1} \bar{y}_t)\| + \zeta_t, \end{aligned} \quad (15)$$

where in (a) we use (8b). Consider the term  $\|G_t(x_t, s_t) - \nabla f_t(x_t, \sigma_t(x_t))\|$ . The definition of  $G_t$  and (8a) give

$$\begin{aligned} & \|G_t(x_t, s_t) - \nabla f_t(x_t, \sigma_t(x_t))\| \\ & = \|G_t(x_t, s_t) - \nabla f_t(x_t, \bar{s}_t)\| \stackrel{(a)}{\leq} L_1 \|s_t - \mathbf{1} \bar{s}_t\|, \end{aligned} \quad (16)$$

where (a) uses the Lipschitz continuity of  $G_t$  (cf. Assumption 3). The proof follows by (15), (16), and  $\|\nabla \phi_t(x)\| \leq L_3$  for all  $x \in \mathbb{R}^n$  (which is derived from Assumption 3).  $\square$

## B. Proof of Lemma 2

We can use (4b) to write

$$\begin{aligned} \|x_{t+1} - x_t\| & = \|x_t + \delta(\tilde{x}_t - x_t) - x_t\| \\ & = \delta \|\tilde{x}_t - x_t\| \stackrel{(a)}{=} \delta \|P_{X_t}[x_t - \alpha d_t] - x_t\|, \end{aligned}$$

where in (a) we have used the update (4a). By adding the null quantity  $(P_{X_t}[x_t^* - \alpha \nabla f_t(x_t^*, \sigma_t(x_t^*))] - x_t^*)$  within the norm and applying the triangle inequality, we get

$$\begin{aligned} & \|x_{t+1} - x_t\| \\ & \leq \delta \|P_{X_t}[x_t - \alpha d_t] - P_{X_t}[x_t^* - \alpha \nabla f_t(x_t^*, \sigma_t(x_t^*))]\| \\ & + \delta \|x_t - x_t^*\| \\ & \stackrel{(a)}{\leq} 2\delta \|x_t - x_t^*\| + \delta \alpha \|d_t - \nabla f_t(x_t^*, \sigma_t(x_t^*))\|, \end{aligned}$$

where in (a) we use a projection property and the triangle inequality. We add and subtract within the norm the term  $\nabla \phi_t(x_t) \mathbf{1} \sum_{i=1}^N \nabla_2 f_{i,t}(x_{i,t}, s_{i,t})$  and use the expression of  $d_t$  and  $G_t$  and the triangle inequality to write

$$\begin{aligned} & \|x_{t+1} - x_t\| \leq 2\delta \|x_t - x_t^*\| \\ & + \delta \alpha \|G_t(x_t, s_t) - \nabla f_t(x_t^*, \sigma_t(x_t^*))\| \\ & + \delta \alpha \left\| \nabla \phi_t(x_t) \left( y_t - \mathbf{1} \sum_{i=1}^N \nabla_2 f_{i,t}(x_{i,t}, s_{i,t}) \right) \right\| \\ & \stackrel{(a)}{=} 2\delta \|x_t - x_t^*\| + \delta \alpha \|G_t(x_t, s_t) - \nabla f_t(x_t^*, \sigma_t(x_t^*))\| \\ & + \delta \alpha L_3 \|y_t - \mathbf{1} \bar{y}_t\|, \end{aligned} \quad (17)$$

where in (a) we use (8b) and  $\|\nabla \phi_t(x)\| \leq L_3$ . The definition of  $G_t$  and its Lipschitz continuity (cf. Assumption 3) imply

$$\begin{aligned} & \|G_t(x_t, s_t) - \nabla f_t(x_t^*, \sigma_t(x_t^*))\| \\ & \leq L_1 \|x_t - x_t^*\| + L_1 \|s_t - \mathbf{1} \sigma_t(x_t^*)\|. \end{aligned} \quad (18)$$

By combining (17) with (18), we get

$$\begin{aligned} & \|x_{t+1} - x_t\| \leq \delta(2 + \alpha L_1) \|x_t - x_t^*\| + \delta \alpha L_1 \|s_t - \mathbf{1} \sigma_t(x_t^*)\| \\ & + \delta \alpha L_3 \|y_t - \mathbf{1} \bar{y}_t\| \\ & \stackrel{(a)}{\leq} \delta(2 + \alpha L_1) \|x_t - x_t^*\| + \delta \alpha L_1 \|s_t - \mathbf{1} \bar{s}_t\| \\ & + \delta \alpha L_1 \|\mathbf{1} \bar{s}_t - \mathbf{1} \sigma_t(x_t^*)\| + \delta \alpha L_3 \|y_t - \mathbf{1} \bar{y}_t\|, \end{aligned} \quad (19)$$

where in (a) we add and subtract  $\mathbf{1} \bar{s}_t$  and we apply the triangle inequality. Now, consider the term  $\|\mathbf{1} \bar{s}_t - \mathbf{1} \sigma_t(x_t^*)\|$ . By using the definition of  $\sigma_t$  and the Lipschitz continuity of  $\phi_{i,t}$  (cf. Assumption 3), it can be seen that (see also [1]),

$$\|\mathbf{1} \bar{s}_t - \mathbf{1} \sigma_t(x_t^*)\| \leq L_3 \|x_t - x_t^*\|. \quad (20)$$

By combining the results (19) and (20), the proof is given.  $\square$

## C. Proof of Lemma 3

By applying (4c) and (7a), we can write

$$\begin{aligned} & \|s_{t+1} - \mathbf{1} \bar{s}_{t+1}\| = \left\| A s_t - \mathbf{1} \bar{s}_t + \tilde{I}(\phi_{t+1}(x_{t+1}) - \phi_t(x_t)) \right\| \\ & \stackrel{(a)}{\leq} \left\| \left( A - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (s_t - \mathbf{1} \bar{s}_t) \right\| + \left\| \tilde{I}(\phi_{t+1}(x_{t+1}) - \phi_t(x_t)) \right\|, \end{aligned}$$

where (a) applies the triangle inequality, introduces  $\tilde{I} := I - \frac{\mathbf{1} \mathbf{1}^\top}{N}$ , and uses the fact that  $\mathbf{1} \in \ker \left( A - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right)$ . Now, we add and subtract within the second norm the term  $\phi_{t+1}(x_t)$  and we apply the triangle inequality obtaining

$$\begin{aligned} & \|s_{t+1} - \mathbf{1} \bar{s}_{t+1}\| \leq \left\| \left( A - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (s_t - \mathbf{1} \bar{s}_t) \right\| \\ & + \left\| \tilde{I}(\phi_{t+1}(x_{t+1}) - \phi_{t+1}(x_t)) \right\| + \left\| \tilde{I}(\phi_{t+1}(x_t) - \phi_t(x_t)) \right\| \\ & \stackrel{(a)}{\leq} \Lambda \|s_t - \mathbf{1} \bar{s}_t\| + L_3 \|x_{t+1} - x_t\| + \omega_t, \end{aligned}$$

where in (a) we use the maximum eigenvalue  $\Lambda$  of the matrix  $A - \frac{\mathbf{1} \mathbf{1}^\top}{N}$ , Assumption 3,  $\omega_t$  (cf (6b)), and  $\|\tilde{I}\| = 1$ . By using Lemma 2 to bound  $\|x_{t+1} - x_t\|$ , the proof follows.  $\square$

## D. Proof of Lemma 4

We use (4d) and (7b) to write

$$\begin{aligned} & \|y_{t+1} - \mathbf{1} \bar{y}_{t+1}\| \leq \|A y_t - \bar{y}_t\| \\ & + \left\| \left( I - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (\nabla_2 f_{t+1}(x_{t+1}, s_{t+1}) - \nabla_2 f_{t+1}(x_t, s_t)) \right\| \\ & \stackrel{(a)}{\leq} \left\| \left( A - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (y_t - \mathbf{1} \bar{y}_t) \right\| \\ & + \left\| \left( I - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (\nabla_2 f_{t+1}(x_{t+1}, s_{t+1}) - \nabla_2 f_{t+1}(x_t, s_t)) \right\| \\ & + \left\| \left( I - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right) (\nabla_2 f_{t+1}(x_t, s_t) - \nabla_2 f_t(x_t, s_t)) \right\|, \end{aligned}$$

where (a) uses  $\mathbf{1} \in \ker \left( A - \frac{\mathbf{1} \mathbf{1}^\top}{N} \right)$  and applies the triangle inequality after adding and subtracting  $(I - \frac{\mathbf{1} \mathbf{1}^\top}{N}) \nabla_2 f_{t+1}(x_t, s_t)$

within the norm. By using the maximum eigenvalue  $\Lambda$  of  $A - \frac{\mathbf{1}\mathbf{1}^\top}{N}$ , Assumption 3, and  $\eta_t$  (cf. (6a)), we get

$$\begin{aligned} & \|y_{t+1} - \mathbf{1}\bar{y}_{t+1}\| \\ & \leq \Lambda \|y_t - \mathbf{1}\bar{y}_t\| + L_2 \|x_{t+1} - x_t\| + L_2 \|s_{t+1} - s_t\| + \eta_t, \end{aligned}$$

Now, we can use (4d) to get

$$\begin{aligned} & \|y_{t+1} - \mathbf{1}\bar{y}_{t+1}\| \leq \Lambda \|y_t - \mathbf{1}\bar{y}_t\| + L_2 \|x_{t+1} - x_t\| \\ & \quad + L_2 \|(A - I)s_t + \phi_{t+1}(x_{t+1}) - \phi_t(x_t)\| + \eta_t \\ & \stackrel{(a)}{\leq} \Lambda \|y_t - \mathbf{1}\bar{y}_t\| + L_2 \|x_{t+1} - x_t\| + L_2 \|(A - I)(s_t - \mathbf{1}\bar{s}_t)\| \\ & \quad + L_2 \|\phi_{t+1}(x_{t+1}) - \phi_t(x_t)\| + \eta_t, \end{aligned}$$

where in (a) we apply the triangle inequality and the fact that  $\mathbf{1} \in \ker\left(A - \frac{\mathbf{1}\mathbf{1}^\top}{N}\right)$ . We add and subtract within the norm the term  $\phi_{t+1}(x_t)$  and apply the triangle inequality, obtaining

$$\begin{aligned} & \|y_{t+1} - \mathbf{1}\bar{y}_{t+1}\| \leq \Lambda \|y_t - \mathbf{1}\bar{y}_t\| + L_2 \|x_{t+1} - x_t\| \\ & \quad + L_2 \|(A - I)(s_t - \mathbf{1}\bar{s}_t)\| + L_2 \|\phi_{t+1}(x_{t+1}) - \phi_{t+1}(x_t)\| \\ & \quad + L_2 \|\phi_{t+1}(x_t) - \phi_t(x_t)\| + \eta_t \\ & \stackrel{(a)}{\leq} \rho \|y_t - \mathbf{1}\bar{y}_t\| + L_2 \|x_{t+1} - x_t\| + L_2 \|(A - I)(s_t - \mathbf{1}\bar{s}_t)\| \\ & \quad + L_2 L_3 \|x_{t+1} - x_t\| + L_2 \omega_t + \eta_t, \end{aligned}$$

where (a) uses Assumption 3 and  $\omega_t$  (cf. (6b)). The proof follows by  $\|A - I\| \leq 2$ , and by applying Lemma 2.  $\square$

### E. Proof of Theorem 1

Let us introduce  $u_t$  to denote  $u_t := \text{COL}(\zeta_t, \eta_t, \omega_t)$ . Then, by combining Lemma 1, 3, and 4, we bound the evolution of  $z_t$  (defined in (5)) through the following dynamical system

$$z_{t+1} \leq M(\delta)z_t + Bu_t, \quad (21)$$

in which

$$M(\delta) := M_0 + \delta E, \quad B := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & L_2 \end{bmatrix},$$

where

$$M_0 := \begin{bmatrix} 1 & 0 & 0 \\ 0 & \Lambda & 0 \\ 0 & 2L_2 & \Lambda \end{bmatrix}, \quad E := \begin{bmatrix} -\mu\alpha & \alpha L_1 & \alpha L_3 \\ E_{21} & \alpha L_1 L_3 & \alpha L_3^2 \\ E_{31} & E_{32} & E_{33} \end{bmatrix},$$

with  $E_{21} := 2L_3 + \alpha L_1 L_3 + \alpha L_1 L_3^2$ ,  $E_{31} := (2 + \alpha L_1 + \alpha L_1)(L_2 + L_2 L_3)$ ,  $E_{32} := \alpha L_1(L_2 + L_2 L_3)$ , and  $E_{33} := \alpha L_3(L_2 + L_2 L_3)$ . Being  $M_0$  triangular, its spectral radius is 1 since  $\Lambda \in (0, 1)$  as implied by Assumption 1. Denote by  $\chi(\delta)$  the eigenvalues of  $M(\delta)$  as a function of  $\delta$ . Call  $v$  and  $w$  respectively the right and left eigenvectors of  $A_0$  associated to 1. Then,  $v = [1 \ 0 \ 0]^\top$ ,  $w = [1 \ 0 \ 0]^\top$ . Being 1 a simple eigenvalue of  $M(0)$ , from [36, Theorem 6.3.12] it holds

$$\left. \frac{d\chi(\delta)}{d\delta} \right|_{\chi=1, \delta=0} = \frac{w^\top E v}{w^\top v} = -\mu\alpha < 0.$$

Then, by continuity of eigenvalues with respect to the matrix entries, there exists  $\bar{\delta} > 0$  so that  $\rho_{\max}(M(\delta)) < 1$  for any  $\delta \in (0, \bar{\delta})$ . From now on we will omit the dependency of  $M$

and its eigenvalues from  $\delta$ . Since  $z_t \geq 0$  for all  $t$ , and  $M$  and  $Bu_t$  have only non-negative entries, one can use (21) to write

$$z_t \leq M^t z_0 + \sum_{k=0}^{t-1} M^k Bu_k. \quad (22)$$

Pick  $\theta \in (0, 1 - \rho_{\max}(M))$  and define  $\tilde{\rho} \triangleq \rho_{\max}(M) + \theta$ . Then, by [36, Lemma 5.6.10], there exists a matrix norm<sup>2</sup>, which we denote as  $\|\cdot\|_\theta$ , such that  $\|M\|_\theta \leq \rho_{\max}(M) + \theta < 1$ . Moreover, by applying [36, Theorem 5.7.13], there exists a vector norm, which we denote by  $\|\cdot\|_\theta$ , which is compatible with the corresponding matrix norm, i.e., such that  $\|Mv\|_\theta \leq \|M\|_\theta \|v\|_\theta$  for any matrix  $M \in \mathbb{R}^{3 \times 3}$  and  $v \in \mathbb{R}^3$ . Using this fact, we use the norm  $\|\cdot\|_\theta$  on both sides of (22) and we apply the triangle inequality to get

$$\begin{aligned} \|z_t\|_\theta & \leq \|M^t z_0\|_\theta + \left\| \sum_{k=0}^{t-1} M^k Bu_{t-k-1} \right\|_\theta \\ & \leq \tilde{\rho}^t \|z_0\|_\theta + \sum_{k=0}^{t-1} \tilde{\rho}^k \|Bu_{t-k-1}\|_\theta. \end{aligned} \quad (23)$$

Being  $\nabla f_t$  Lipschitz continuous (cf. Assumption 3), it holds

$$\begin{aligned} f_t(x_t, \sigma_t(x_t)) - f_t(x_t^*, \sigma_t(x_t^*)) & \leq \frac{L_1}{2} \|x_t - x_t^*\|^2 \\ & \stackrel{(a)}{\leq} \frac{L_1}{2} \|z_t\|^2, \end{aligned} \quad (24)$$

where in (a) uses the fact that  $\|x_t - x_t^*\|$  is a component of  $z_t$ . Recalling that all norms are equivalent on finite-dimensional vector spaces, there always exist  $\lambda_1 > 0$  and  $\lambda_2 > 0$  such that  $\|\cdot\| \leq \lambda_1 \|\cdot\|_\theta$  and  $\|\cdot\|_\theta \leq \lambda_2 \|\cdot\|$ . Thus, by exploiting the square norm and combining the results (23) with the equivalence of the norms, we can bound (24) as

$$\begin{aligned} f_t(x_t, \sigma_t(x_t)) - f_t(x_t^*, \sigma_t(x_t^*)) & \leq \frac{L_1 \lambda_1^2}{2} \left( \tilde{\rho}^{2t} \|z_0\|_\theta^2 \right. \\ & \quad \left. + 2\tilde{\rho}^t \|z_0\|_\theta \sum_{k=0}^{t-1} \tilde{\rho}^k \|Bu_{t-k-1}\|_\theta + \sum_{k=0}^{t-1} \tilde{\rho}^{2k} \|Bu_{t-k-1}\|_\theta^2 \right), \end{aligned}$$

which, combined with the definitions of  $R_T$  (cf. (3)),  $U_T$ , and  $Q_T$  (cf. (11)), and the equivalence of the norms, leads to

$$\begin{aligned} R_T & \leq \frac{L_1 \lambda^2}{2} \left( \sum_{t=1}^T \tilde{\rho}^{2t} \|z_0\| + 2\|z_0\| U_T + Q_T \right) \\ & \stackrel{(a)}{\leq} \frac{L_1 \lambda^2}{2} \left( \frac{\|z_0\|^2}{1 - \tilde{\rho}^2} + 2\|z_0\| U_T + Q_T \right), \end{aligned} \quad (25)$$

where  $\lambda = \lambda_1 \lambda_2$  and (a) uses the geometric series property.

As regards the result (12), we use (14) to write

$$\begin{aligned} \text{dist}(x_{t+1}, X_{t+1}) & = \text{dist}(x_t + \delta(\tilde{x}_t - x_t), X_{t+1}) \\ & \stackrel{(a)}{\leq} \text{dist}(x_t + \delta(\tilde{x}_t - x_t), X_t) + \gamma_t, \end{aligned} \quad (26)$$

where in (a) we add and subtract the term  $\text{dist}(x_t + \delta(\tilde{x}_t - x_t), X_t)$  and we introduce  $\gamma_t$  (cf. (6c)). Now, we recall that

$$\text{dist}(x_t + \delta(\tilde{x}_t - x_t), X_t) = \min_{y \in X_t} \|x_t + \delta(\tilde{x}_t - x_t) - y\|.$$

<sup>2</sup>An expression of  $\|\cdot\|_\theta$  can be found in the proof of [36, Lemma 5.6.10].

Thus, by adding and subtracting within the norm the term  $(1 - \delta)v_t$  with  $v_t \in X_t$  so that  $\|x_t - v_t\| = \text{dist}(x_t, X_t)$ , we can use the triangle inequality and the definition of min to get

$$\begin{aligned} & \text{dist}(x_t + \delta(\tilde{x}_t - x_t), X_t) \\ & \leq (1 - \delta) \|x_t - v_t\| + \min_{y \in X_t} \|v_t + \delta(\tilde{x}_t - v_t) - y\| \\ & = (1 - \delta) \text{dist}(x_t, X_t) + \text{dist}(v_t + \delta(\tilde{x}_t - v_t), X_t), \end{aligned}$$

which allows us to rewrite (26) as

$$\begin{aligned} \text{dist}(x_{t+1}, X_{t+1}) & \leq (1 - \delta) \text{dist}(x_t, X_t) \\ & \quad + \text{dist}(v_t + \delta(\tilde{x}_t - v_t), X_t) + \gamma_t. \end{aligned} \quad (27)$$

Notice that  $v_t, \tilde{x}_t \in X_t$  and  $0 < \delta < 1$ , then  $v_t + \delta(\tilde{x}_t - v_t) \in X_t$  and the second term of (27) is null and (27) becomes

$$\text{dist}(x_{t+1}, X_{t+1}) \leq (1 - \delta) \text{dist}(x_t, X_t) + \gamma_t. \quad (28)$$

Both members of (28) are always positive, then (28) leads to

$$\text{dist}(x_t, X_t) \leq (1 - \delta)^t \text{dist}(x_0, X_0) + \sum_{k=0}^{t-1} (1 - \delta)^k \gamma_{t-k-1}.$$

By summing the latter for  $t = 1$  up to  $t = T$  and using the geometric series property the proof follows.  $\square$

## F. Proof of Corollary 1

Here, for all  $t \geq 0$ , it holds  $u_t \equiv 0$  and  $x_t^* = x^*$ . By the same arguments of Theorem 1, we use the Lipschitz continuity of  $\nabla f_t$  (cf. Assumption 3), (23) with  $u_t \equiv 0$ , and the equivalence of the norms to get  $f(x_t, \sigma(x_t)) - f(x^*, \sigma(x^*)) \leq \tilde{\rho}^{2t} \frac{L_1 \lambda_1^2}{2} \|z^0\|_{\theta}^2$ . The proof follows by using again the equivalence of the norms and setting  $\lambda := \lambda_1 \lambda_2$ .  $\square$

## REFERENCES

- [1] X. Li, L. Xie, and Y. Hong, "Distributed aggregative optimization over multi-agent networks," *IEEE Transactions on Automatic Control*, 2021.
- [2] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, no. 3, pp. 680–704, 2016.
- [3] S. Liang, P. Yi, and Y. Hong, "Distributed nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.
- [4] D. Gadjov and L. Pavel, "A passivity-based approach to nash equilibrium seeking over networks," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1077–1092, 2018.
- [5] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.
- [6] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2061–2075, 2020.
- [7] G. Notarstefano, I. Notarnicola, and A. Camisa, "Distributed optimization for smart cyber-physical networks," *Foundations and Trends® in Systems and Control*, vol. 7, no. 3, pp. 253–383, 2019.
- [8] R. L. Cavalcante and S. Stanczak, "A distributed subgradient method for dynamic convex optimization problems under noisy information exchange," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 243–256, 2013.
- [9] Z. J. Towfic and A. H. Sayed, "Adaptive penalty-based distributed stochastic convex optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 15, pp. 3924–3938, 2014.
- [10] M. Akbari, B. Ghahesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 3, pp. 417–428, 2015.
- [11] D. Mateos-Núñez and J. Cortés, "Distributed online convex optimization over jointly connected digraphs," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 23–37, 2014.
- [12] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed convex optimization on dynamic networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3545–3550, 2016.
- [13] D. Yuan, D. W. Ho, and G.-P. Jiang, "An adaptive primal-dual subgradient algorithm for online distributed constrained optimization," *IEEE transactions on cybernetics*, vol. 48, no. 11, pp. 3045–3055, 2017.
- [14] S. Shahrampour and A. Jadbabaie, "Distributed online optimization in dynamic environments using mirror descent," *IEEE Transactions on Automatic Control*, vol. 63, no. 3, pp. 714–725, 2017.
- [15] X. Zhou, E. Dall'Anese, L. Chen, and A. Simonetto, "An incentive-based online optimization framework for distribution grids," *IEEE transactions on Automatic Control*, vol. 63, no. 7, pp. 2019–2031, 2017.
- [16] M. Akbari, B. Ghahesifard, and T. Linder, "Individual regret bounds for the distributed online alternating direction method of multipliers," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1746–1752, 2019.
- [17] S. Lee and M. M. Zavlanos, "On the sublinear regret of distributed primal-dual algorithms for online constrained optimization," *arXiv preprint arXiv:1705.11128*, 2017.
- [18] X. Li, X. Yi, and L. Xie, "Distributed online optimization for multi-agent networks with coupled inequality constraints," *IEEE Transactions on Automatic Control*, 2020.
- [19] X. Yi, X. Li, L. Xie, and K. H. Johansson, "Distributed online convex optimization with time-varying coupled inequality constraints," *IEEE Transactions on Signal Processing*, vol. 68, pp. 731–746, 2020.
- [20] X. Li, X. Yi, and L. Xie, "Distributed online convex optimization with an aggregative variable," *IEEE Transactions on Control of Network Systems*, 2021.
- [21] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.
- [22] S. S. Kia, B. Van Scoy, J. Cortes, R. A. Freeman, K. M. Lynch, and S. Martínez, "Tutorial on dynamic average consensus: The problem, its applications, and the algorithms," *IEEE Control Systems Magazine*, vol. 39, no. 3, pp. 40–72, 2019.
- [23] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.
- [24] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, "Newton-Raphson consensus for distributed convex optimization," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 994–1009, 2015.
- [25] P. Di Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 120–136, 2016.
- [26] A. Nedić, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM Journal on Optimization*, vol. 27, no. 4, pp. 2597–2633, 2017.
- [27] G. Qu and N. Li, "Harnessing Smoothness to Accelerate Distributed Optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1245–1260, 2018.
- [28] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Convergence of asynchronous distributed gradient methods over stochastic networks," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 434–448, 2017.
- [29] C. Xi, R. Xin, and U. A. Khan, "ADD-OPT: Accelerated distributed directed optimization," *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1329–1339, 2017.
- [30] R. Xin and U. A. Khan, "A linear algorithm for optimization over directed graphs with geometric convergence," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 315–320, 2018.
- [31] G. Scutari and Y. Sun, "Distributed nonconvex constrained optimization over time-varying digraphs," *Mathematical Programming*, vol. 176, no. 1-2, pp. 497–544, 2019.
- [32] Y. Zhang, R. J. Ravier, M. M. Zavlanos, and V. Tarokh, "A distributed online convex optimization algorithm with improved dynamic regret," in *IEEE Conference on Decision and Control (CDC)*, 2019, pp. 2449–2454.
- [33] I. Notarnicola, A. Simonetto, F. Farina, and G. Notarstefano, "Distributed personalized gradient tracking with convex parametric models," *IEEE Transactions on Automatic Control*, 2022.
- [34] G. Carnevale, F. Farina, I. Notarnicola, and G. Notarstefano, "Distributed online optimization via gradient tracking with adaptive momentum," *arXiv preprint arXiv:2009.01745*, 2020.
- [35] D. P. Bertsekas and A. Scientific, *Convex optimization algorithms*. Athena Scientific Belmont, 2015.
- [36] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.