

Random Coordinate Descent for Resource Allocation in Open Multi-Agent Systems

Charles Monnoyer de Galland, Renato Vizuete, Julien M. Hendrickx, Elena Panteley, and Paolo Frasca

Abstract—We propose a method for analyzing the distributed random coordinate descent algorithm for solving separable resource allocation problems in the context of an open multi-agent system, where agents can be replaced during the process. In particular, we characterize the evolution of the distance to the minimizer in expectation by following a time-varying optimization approach which builds on two components. First, we establish the linear convergence of the algorithm in closed systems, in terms of the estimate towards the minimizer, for general graphs and appropriate step-size. Second, we estimate the change of the optimal solution after a replacement, in order to evaluate its effect on the distance between the current estimate and the minimizer. From these two elements, we derive stability conditions in open systems and establish the linear convergence of the algorithm towards a steady-state expected error. Our results enable to characterize the trade-off between speed of convergence and robustness to agent replacements, under the assumptions that local functions are smooth, strongly convex, and have their minimizers located in a given ball. The approach proposed in this paper can moreover be extended to other algorithms guaranteeing linear convergence in closed system.

Index Terms—Open multi-agent systems, distributed optimization, gradient methods, agents and autonomous systems.

I. INTRODUCTION

Resource allocation is an important optimization problem, where a fixed amount of resources must be distributed among a specific number of activities or agents in an optimal way [1, 2]. In multi-agent systems, this problem is formulated as the minimization of an objective function f that is separable in local costs $f_i: \mathbb{R}^d \rightarrow \mathbb{R}$ held by the agents, subject to an equality constraint on the weighted sum of the states $x_i \in \mathbb{R}^d$ with respect to the budget $b \in \mathbb{R}^d$. The problem can then be written as

$$\min_{x \in \mathbb{R}^{nd}} f(x) = \sum_{i=1}^n f_i(x_i) \quad \text{subject to} \quad \sum_{i=1}^n a_i x_i = b, \quad (1)$$

Research supported by the “RevealFlight” ARC at UCLouvain, by the *Incentive Grant for Scientific Research (MIS)* “Learning from Pairwise Data” of the F.R.S.-FNRS and in part by the Agence Nationale de la Recherche (ANR) via grant “Hybrid And Networked Dynamical sYstems” (HANDY), number ANR-18-CE40-0010.

C. Monnoyer de Galland and R. Vizuete equally contributed to this work. C. Monnoyer de Galland, R. Vizuete and J. M. Hendrickx are with the ICTEAM institute, UCLouvain, Louvain-la-Neuve, Belgium. R. Vizuete was with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, France, and C. Monnoyer de Galland was a FRIA fellow (F.R.S.-FNRS) during the elaboration of this work. E. Panteley is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France. P. Frasca is with Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France. (E-mail addresses: charles.monnoyer@uclouvain.be; renato.vizueteharo@uclouvain.be; julien.hendrickx@uclouvain.be; elena.panteley@l2s.centralesupelec.fr; paolo.frasca@gipsa-lab.fr).

where $a_i > 0$ is the weight of agent i to satisfying the constraint. Each agent thus aims at minimizing its local cost while guaranteeing the budget, which requires a certain level of collaboration between them.

Problem (1) appears in different applications, including smart grids [3], power systems [4], actuator networks [5], and games [6]. Most of these existing studies assume that the composition of the multi-agent system remains fixed during the entire process. Yet, with the growing size of systems nowadays, arrivals and departures of agents are expected to happen more frequently, giving rise to *open multi-agent systems*, where agents join and leave the system at a time-scale similar to that of the studied process. Consider, for instance, the case of distributed energy resources where a fixed amount of energy must be supplied by a network of devices [7], and where some of the devices might experience failures with higher probability as the system size increases, or change their operating point due to environmental conditions.

In the framework of open systems, a fixed solution for (1) cannot be obtained as in general, the size is not fixed and the cost functions keep changing, such that the goal of the agents is to track the time-varying solution of (1) as well as possible at all times. Moreover, as the size of such systems reaches large values, global optimization methods relying, *e.g.*, on the computation of the whole gradient of f are not suited since the computational complexity would be high and in some cases, it would not be practical to gather the whole gradient as agents may have entered/left in the meantime. In fact, most of the algorithms used to solve (1) in a decentralized way are gradient-based, such as in [8], where the authors use a weighted version of the well-known *Gradient Descent* algorithm with an appropriate choice of weights to preserve the constraint. Yet, this type of methods requires significant computational resources and, therefore, is not suitable for handling open systems. For this reason, it is important to consider optimization algorithms based on local interactions, since they are more flexible. An alternative type of algorithms that allow to considerably reduce the computational complexity is the so-called *Coordinate Descent* algorithm introduced by Nesterov, where the optimization is performed only along one direction at each iteration [9]. For multi-agent systems, the selection of one coordinate is equivalent to the choice of a particular edge of the network to perform the optimization. In such algorithms, the sequence of edges is crucial, and hence a randomized choice denoted as *Random Coordinate Descent* algorithm (RCD) was studied in [10], where convergence of the cost functions is proved under standard assumptions when only pairwise interactions are considered, so that the algorithm

requires only the computation of a pair of local gradients per iteration.

In general, interactions in a multi-agent system are characterized by an underlying network, so that agents can only communicate with a limited number of other neighbors determined by the communication network topology. This feature of the network plays an important role in the analysis and design of algorithms, since the performance can be different in sparse and dense graphs [11].

Furthermore, problems of the type of (1) often assume that the contributions of the agent to the constraint is homogeneous (*i.e.*, that $a_i = 1$ for each agent i). However, this is not always the case, as for example in energy supply, where priority may be given to renewable energy sources while fossil-fuel plants ought to contribute less to the amount of energy required by each region [12].

Our goal is to analyze the RCD algorithm applied to the most general possible version (1) (*i.e.*, with arbitrary graph topologies and non-homogeneous contributions of the agents) in open systems. In particular, we focus on systems subject to only replacements, and hence of fixed size, so that the main challenge to handle is the variations of local cost functions, such as *e.g.*, in the context of energy distribution where such changes can be triggered by time-varying environmental conditions.

A. Optimization in open multi-agent systems

Algorithms for open multi-agent systems have recently been studied in several contexts. In the case of consensus, [13]–[15] analyzed the behavior and performance of gossip interactions, [16, 17] studied dynamic consensus in terms of stability, and [18, 19] focused on consensus with stochastic interactions.

Optimization problems in open multi-agent systems scenarios, characterized, among others, by time-varying objective functions, start getting attention as well. In [20], the behavior and the stability of gradient descent was studied in a setting where agents can be replaced. In [21], an algorithm based on dual averaging was proposed to minimize a global cost function that depends on a time-varying set of active agents in a fixed size network.

Time-varying objective functions are also considered in an alternative field of work called *online optimization* [22, 23], where a common approach is to minimize, over a finite period of time T , the *dynamic regret* defined as

$$\text{Reg}_T^d := \sum_{t=1}^T f^t(x^t) - \sum_{t=1}^T f^t(x^{*,t}), \quad (2)$$

where f^t is a sequence of cost functions, x^t are the estimates and $x^{*,t} := \arg \min_x f^t(x)$ is the minimizer of the global function f^t at time t . The objective of online optimization is thus to determine the sequence of estimates x^t that keeps Reg_T^d as small as possible over the time period, under some assumptions about the possible sequences of time-varying cost functions.

Instead, in our problem, replacements of agents occur without any regularity and in this case, it is not possible to obtain a sublinear regret which is the usual objective of

online optimization [24]. Furthermore, we can observe from (2) that the computation of the regret implies an accumulation of errors from the time instant $t = 1$, which does not seem appropriate for the case of open multi-agent systems, where the replacement of an agent implies that all the past information of the replaced agent is not longer available, since this agent left the system. For this reason, we perform the analysis of the resource allocation problem in open multi-agent systems considering a *time-varying optimization* approach [25], where the objective of the algorithms is to be at all times as close as possible to the instantaneous minimizer. This objective is more suitable for open multi-agent systems where replacements may be infrequent and agents try to reach the best performance at all time instants without regarding a performance in a potential future horizon since even if the cost functions belong to the same class, there is no regularity in the way they change [26]–[28]. Nevertheless, even if replacements are not frequent, they can impact the performance of the algorithms since they modify the location of the global minimizer.

The framework of time-varying optimization has been used in many works, including the resource allocation problem. Exact convergence can be obtained under restrictive assumptions like identical Hessian matrices [29] or local cost functions with a time-independent part [30], that generally guarantee the continuity of the time-varying minimizer. When exact convergence is not possible, the main challenge is the derivation of an upper bound for the error [25]. This approach has been used, for instance, in the case of quadratic cost functions [31]. However, changes of the cost functions due to replacements of agents without establishing further restrictions on the structure of the cost functions have never been explored, and this particular characteristic is the motivation of this work and plays an important role in the formulation of the problem and the derivation of an explicit bound for the error.

B. Preliminary version and contribution

We study the Random Coordinate Descent algorithm (RCD) to solve the resource allocation problem in an open system where agents get replaced during the process, relying on a decoupled analysis of the RCD algorithm and of replacements of single agents. A preliminary version of this work was presented in [32], where the problem was analyzed for homogeneous agents holding one-dimensional local cost functions, interacting in a fully connected network, and with uniform probabilities in the selection of edges for the updates. Moreover, replacements of single agents were studied by analyzing the case where possibly all agents can get replaced at once.

By contrast, in this work, we focus on heterogeneous agents holding d -dimensional local cost functions, interacting in a general graph topology. We consider an arbitrary distribution for the probabilities associated to the choice of edges and we derive an upper bound for the convergence of the algorithm following a similar approach in terms of a norm induced by a matrix associated to the network. Moreover, we now directly study the replacement of a single agent instead of considering the possibility for all agents to be replaced at once, yielding tighter bounds for that case, independent of the system size.

The remainder of this article is organized as follows. In Section II we introduce the problem statement and the necessary preliminaries. In Section III we study the impact of the replacements on the location of the minimizers of the system and their distance with respect to the estimate. Section IV focuses on the linear convergence of the RCD in a closed system considering an appropriate norm. Section V presents the analysis of the RCD in an open multi-agent scenario. Finally, conclusions and future work are exposed in Section VI.

II. PROBLEM STATEMENT

In this section, we formulate the constrained resource allocation problem in open systems. First, we introduce the notation used along this work. Then, we present the problem in closed systems and we describe how the problem changes because of replacements, next we detail how the time is sampled. Finally, we present the Random Coordinate Descent algorithm which is considered in this work.

A. Notation

For two vectors $x, y \in \mathbb{R}^n$, $\langle x, y \rangle = x^\top y = \sum_{i=1}^n x_i y_i$ denotes the usual Euclidean inner product and $\|x\| = \sqrt{x^\top x}$ the Euclidean norm. The 1-norm of a vector x is denoted as $\|x\|_1$. We denote the vector of size n constituted of only ones by $\mathbf{1}_n$ and the identity matrix of dimension n by I_n . The vector of size n constituted of only zeros is denoted by $\mathbf{0}_n$. We use $B(x, r) = \{y : \|x - y\| \leq r\}$ to denote the closed ball of radius $r \geq 0$ centered at x . For a positive (semipositive) definite matrix A , we denote by $\|x\|_A = (x^\top A x)^{1/2}$ the vector norm (seminorm) induced by A . The Kronecker product is denoted by $\otimes$.

B. Resource allocation problem

We consider the resource allocation problem defined in (1), where a budget b must be distributed among n agents according to some positive weight distribution $a \in \mathbb{R}^n$ (i.e., with $a_i > 0$ for $i = 1, \dots, n$). For the sake of simplicity, we first describe it in closed system (i.e., where the set of agents remains the same); we will see in the next subsection that this formulation directly extends to open systems.

The constraint in (1) can be equivalently expressed as $(a^\top \otimes I_d)x = b$, where $\otimes$ denotes the Kronecker product. The feasible set of (1) is thus given by

$$\mathcal{S}_{a,b} := \left\{ x \in \mathbb{R}^{nd} \mid (a^\top \otimes I_d)x = b \right\}. \quad (3)$$

For the particular case $d = 1$, the resource allocation constraint can be expressed as $\langle a, x \rangle = b$. We make the following classical assumption on the local cost functions.

Assumption 1. Each function f_i is continuously differentiable, α -strongly convex (i.e., $f_i(x) - \frac{\alpha}{2}\|x\|^2$ is convex) and β -smooth (i.e., $\|\nabla f_i(x) - \nabla f_i(y)\| \leq \beta\|x - y\|$, $\forall x, y \in \mathbb{R}^d$).

Assumption 1 provides an upper and a lower bound to the curvature of the functions. The value $\kappa := \frac{\beta}{\alpha} \geq 1$ is called the *condition number* of the functions. The set of the functions satisfying Assumption 1 is denoted by $\mathcal{F}_{\alpha,\beta}$.

Proposition 1. If $f_1, \dots, f_n \in \mathcal{F}_{\alpha,\beta}$, then the global cost function f from (1) satisfies $f \in \mathcal{F}_{\alpha,\beta}$.

Since Assumption 1 guarantees that f is α -strongly convex, the solution of the problem (1) denoted by x^* is unique [33]. By using KKT conditions we obtain that there exists a $\lambda^* \in \mathbb{R}^d$, such that a necessary and sufficient condition for the optimality of x^* is

$$\nabla f(x^*) = (a^\top \otimes I_d)^\top \lambda^* \quad (4)$$

which equivalently reads $\nabla f(x^*) = a \otimes \lambda^*$, where λ^* is a Lagrange multiplier vector [33].

C. Network description and open system

In addition to problem (1), we assume to have an undirected and connected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the set of nodes is given by $\mathcal{V} = \{1, \dots, n\}$ and the set of edges by $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Each agent $i \in \mathcal{V}$ has access to a local cost function $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$ and to a local variable $x_i \in \mathbb{R}^d$. Agents can exchange information at random times through pairwise interactions according to the network $\mathcal{G}$. Whenever an interaction happens in the system, an edge $(i, j) \in \mathcal{E}$ is selected with some fixed probability $p_{ij} > 0$ and agents i and j can then exchange information in a bidirectional manner to update their respective estimates.

Moreover, we consider that *replacements of agents* happen in the system, making it *open*. Each agent $i \in \mathcal{V}$ gets replaced at random time instants, resulting in the change of its local cost function, and hence of the global minimizer x^* . Following the approach in [20], we restrict the location of the minimizers of the local functions:

Assumption 2. There exists $c > 0$, such that for all $i \in \mathcal{V}$, the minimizer of f_i denoted as $\bar{x}_i^* := \arg \min_x f_i(x)$ satisfies $\bar{x}_i^* \in B(\mathbf{0}_d, c)$. Moreover, without loss of generality $f_i(\bar{x}_i^*) = 0$ for all $i \in \mathcal{V}$.

Assumption 2 guarantees a certain level of uniformity among the local cost functions. In particular, it prevents arbitrary changes of functions, and thus of x^* , during replacements. Also, since our objective is to study the convergence of the minimizer and how it is affected by the replacements, we do not use the actual values of f_i .

For the sake of simplicity, we assume that when agent i is replaced, the joining agent that takes its place retrieves its label i and its estimate x_i , so that the constraint $(a^\top \otimes I_d)x = b$ is preserved, but receives a new local cost function satisfying Assumptions 1 and 2. Denoting f_i^k the local cost function held by agent i at the time instant k , we can then reformulate (1) as the time-varying resource allocation problem

$$\min_{x \in \mathcal{S}_{a,b}} f^k(x) := \sum_{i=1}^n f_i^k(x_i), \quad (5)$$

where the changes of the cost functions are due to replacements. The solution of (5) can thus differ from a time instant t_k to another, and we denote it by $x^{*,k} := \arg \min_{x \in \mathcal{S}_{a,b}} f^k(x)$. The objective of the agents is to track $x^{*,k}$ as well as possible even though replacements happen in the system.

D. Discrete-event modelling

The evolution of the open network presented in the previous section is characterized by the instantaneous occurrence at random time instants of either pairwise interactions or replacements. With a slight abuse of language, we call “event” such occurrence which results in a modification of the system depending on its nature. We moreover define the “event set” of the system from which that nature is drawn as

$$\Xi = U \cup R, \quad (6)$$

where $U := \bigcup_{(i,j) \in \mathcal{E}} U_{ij}$ is the set of all possible events U_{ij} , i.e., the pairwise interaction between two connected agents i and j , and $R := \bigcup_{i \in \mathcal{V}} R_i$ is the set of all events R_i , i.e., the replacement of a single agent i in the system. This formulation actually relates with *discrete event systems* (see e.g., [34]).

We assume that two events never occur simultaneously, so that we can consider a discrete evolution of the time where each time-step $k \in \mathbb{N}$ corresponds to the time instant at which the k -th event takes place. For all $k \in \mathbb{N}$, we then define the random variable $\xi_k \in \Xi$ which characterizes the nature of the event taking place at the time-step k . We moreover consider the following assumption that guarantees that replacements and interactions are independent processes, so that the event happening at time-step k is a pairwise interaction (i.e., $\xi_k \in U$) with fixed probability p_U , and a replacement (i.e., $\xi_k \in R$) with fixed probability $p_R = 1 - p_U$.

Assumption 3. For all $k \in \mathbb{N}$, ξ_k is independent of ξ_j for all $j \neq k$, and of any variable in the system prior to time-step k , such as the estimates or local cost functions.

Our goal is to characterize the evolution of the distance between the estimates held by the agents x^k and the instantaneous minimizer $x^{*,k}$. A choice of measure for this distance is, e.g., the squared Euclidean norm $\|x^k - x^{*,k}\|^2$, although we will see later that this metric might not be appropriate for general graph topologies. Assumption 3 allows for performing this analysis by studying independently the effect of each type of event on our metric in order to characterize its evolution in expectation over a single time step.

E. Random Coordinate Descent (RCD) algorithm

To compute the solution of (5) we consider the Random Coordinate Descent (RCD) algorithm introduced in [10]. This algorithm involves the update of the states of only a pair of neighbouring agents at each iteration, so that it is distributed and its computational complexity is cheap. Hence, in the event U_{ij} , i.e., whenever the pair of agents $(i, j) \in \mathcal{E}$ is selected with probability p_{ij} during a pairwise update event, they perform an *RCD update*, which is defined as follows for some nonnegative step-size $h \geq 0$:

$$x^+ = x - h \mathbf{Q}^{ij} \nabla f(x), \quad (7)$$

where $\mathbf{Q}^{ij}$ is the $nd \times nd$ matrix defined as $\mathbf{Q}^{ij} = Q^{ij} \otimes I_d$, with Q^{ij} the $n \times n$ matrix filled with zeroes except for the four

following entries:

$$\begin{aligned} [Q^{ij}]_{i,i} &= \frac{a_j^2}{a_i^2 + a_j^2}; & [Q^{ij}]_{i,j} &= -\frac{a_i a_j}{a_i^2 + a_j^2}; \\ [Q^{ij}]_{j,i} &= -\frac{a_i a_j}{a_i^2 + a_j^2}; & [Q^{ij}]_{j,j} &= \frac{a_i^2}{a_i^2 + a_j^2}. \end{aligned}$$

With the update rule (7), only agents i and j update their estimates while all the other agents keep it the same. For agents i and j , (7) essentially amounts to perform a gradient step on the function $f_i(x_i) + f_j(x_j)$ under the constraint that $a_i x_i + a_j x_j$ remains constant. This ensures that the resource allocation constraint is preserved as long as the starting point satisfies it. In particular, in the case of homogeneous agents (i.e., where $a = \mathbb{1}_n$), then one shows that $x_i^+ = x_i - \frac{h}{2}(\nabla f_i(x_i) - \nabla f_j(x_j))$, so that the update follows both gradients with equal weight while preserving the constraint.

Observe that the method presented here requires (i) the initial point to be feasible (which is rather standard for such methods), and (ii) that the estimates are maintained during replacements (which is assumed in Section II-C). Otherwise one would need to design a process to run in parallel of the optimization process to meet the constraint. This is, however, out of the scope of this paper.

Remark 1. The update rule (7) can be formally obtained by solving the following optimization problem, which corresponds to the interpretation given above (we refer to [10] for details):

$$\arg \min_{s_i, s_j \in \mathbb{R}^d: a_i s_i + a_j s_j = 0} \left\langle \begin{bmatrix} \nabla f_i(x_i) \\ \nabla f_j(x_j) \end{bmatrix}, \begin{bmatrix} s_i \\ s_j \end{bmatrix} \right\rangle + \frac{\beta}{2} \left\| \begin{bmatrix} s_i \\ s_j \end{bmatrix} \right\|^2. \quad (8)$$

Based on the approach of [10], one can then show that the optimal step-size that solves (8) is given by $h = 1/\beta$.

We also introduce the following matrix that builds on the definition of the update rule (7) and that will be used later:

$$\mathbf{L}_p = \sum_{(i,j) \in \mathcal{E}} p_{ij} \mathbf{Q}^{ij} = \left(\sum_{(i,j) \in \mathcal{E}} p_{ij} Q^{ij} \right) \otimes I_d = L_p \otimes I_d. \quad (9)$$

This matrix appears in the dynamics corresponding to the conditional expectation:

$$\mathbb{E}[x(k+1)|x(k)] = x(k) - h \mathbf{L}_p \nabla f(x(k)), \quad (10)$$

and will be used for the definition of an appropriate norm for the analysis of the RCD algorithm. Observe that by definition of Q^{ij} and L_p , we have

$$L_p a = Q^{ij} a = \mathbf{0}_n, \quad (11)$$

which means that zero is an eigenvalue of both Q^{ij} and L_p with corresponding eigenvector a . We denote by λ_2 and λ_n respectively the second smallest and the largest eigenvalues of L_p . Since L_p is symmetric, all the eigenvalues are real and satisfy $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$ when the graph $\mathcal{G}$ is connected (we refer to Lemma 3.3 of [10] for a detailed proof).

Remark 2. For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, when $a = \mathbb{1}_n$ (homogeneous agents) and the probabilities p_{ij} are uniformly distributed, we have $L_p = \frac{1}{2|\mathcal{E}|} L$, where L is the usual Laplacian of the graph. Hence, we refer to L_p as a “scaled Laplacian”, as it enjoys similar properties, especially in terms of eigenvalues.

III. EFFECT OF REPLACEMENTS

In this section, we bound the distance by which the minimizer of f can change after the replacement of a single agent, *i.e.*, the modification of a single function. Our first two results concern the location of the minimizer: Lemma 1 is a generalisation of the analysis performed in [32], and Lemma 2 studies the location of the minimizer held by each individual agent.

Lemma 1. *Let $x^* := \arg \min_{x \in \mathcal{S}_{a,b}} \sum_{i=1}^n f_i(x_i)$. If all f_i satisfy Assumptions 1 and 2, then $x^* \in B(\mathbf{0}_{nd}, R_{b,\kappa})$ with*

$$R_{b,\kappa} = \sqrt{n\kappa} \left(c + \frac{c}{\sqrt{\kappa}} + \frac{\|b\|}{\sqrt{n}\|a\|} \right), \quad (12)$$

where c was defined in Assumption 2.

Proof. The proof is left to Appendix A. $\square$

Lemma 2. *Let $x^* := \arg \min_{x \in \mathcal{S}_{a,b}} \sum_{i=1}^n f_i(x_i)$. If f_i satisfies Assumptions 1 and 2 for all $i = 1, \dots, n$, then for λ^* defined in (4)*

$$\|\lambda^*\| \leq \frac{\beta}{\|a\|^2} (\|b\| + c\|a\|_1); \quad (13)$$

and

$$\|x_i^*\| \leq \frac{a_i}{\|a\|^2} \kappa (\|b\| + c\|a\|_1) + c. \quad (14)$$

Proof. The proof is left to Appendix B. $\square$

We can now use these two lemmas to characterize the evolution of the distance between the estimate x^k and the minimizer $x^{*,k}$ after a replacement event. Without loss of generality, we assume that agent n , and hence f_n , is replaced, and for the $n+1$ functions $f_1, f_2, \dots, f_{n-1}, f_n^{(1)}, f_n^{(2)}$ satisfying Assumptions 1 and 2 we define the minimizer before a replacement $x^{(1)}$, and after a replacement $x^{(2)}$ as

$$\begin{aligned} x^{(1)} &:= \arg \min_{x \in \mathcal{S}_{a,b}} \left(\sum_{i=1}^{n-1} f_i(x_i) + f_n^{(1)}(x_n) \right); \\ x^{(2)} &:= \arg \min_{x \in \mathcal{S}_{a,b}} \left(\sum_{i=1}^{n-1} f_i(x_i) + f_n^{(2)}(x_n) \right). \end{aligned} \quad (15)$$

Proposition 2. *Consider $x^{(1)}$ and $x^{(2)}$ as defined in (15), let a_+ and a_- respectively denote the largest and smallest values in a , and let $\rho_a := \frac{a_+^2}{\|a\|^2 - a_-^2}$, then*

$$\|x^{(1)} - x^{(2)}\|^2 \leq \min\{\psi_{n,\kappa}, \chi_{n,\kappa}, \theta_{n,\kappa}\} =: \bar{M}_{n,\kappa}^2, \quad (16)$$

with

$$\psi_{n,\kappa} = 4n\kappa \left(c + \frac{c}{\sqrt{\kappa}} + \frac{\|b\|}{\sqrt{n}\|a\|} \right)^2; \quad (17)$$

$$\chi_{n,\kappa} = 8 \left(\frac{a_+}{\|a\|^2} \kappa (\|b\| + c\|a\|_1) + c \right)^2; \quad (18)$$

$$\theta_{n,\kappa} = 4 \left(1 + \frac{(\kappa+1)^2}{4\kappa} \rho_a \right) \left(\frac{a_+}{\|a\|^2} \kappa (\|b\| + c\|a\|_1) + c \right)^2. \quad (19)$$

Proof. The proof is left to Appendix C. $\square$

The bound $\bar{M}_{n,\kappa}^2$ from Proposition 2 is obtained by taking the minimum between three quantities: $\psi_{n,\kappa}$, $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$. The first one follows from the largest possible distance existing between two minimizers, defined by the region in which they

can be located. The second and third ones rely on the largest possible distance between the local minimizers corresponding to the replaced agents. While $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ are derived using inequalities associated with α -strongly convex functions, the proof of $\theta_{n,\kappa}$ also involves the use of additional properties corresponding to β -smooth functions and the determination of the maximum value of a concave function. The bound $\theta_{n,\kappa}$ shows a strong dependence on the weights of the agents through the coefficient ρ_a , which is not present in the other two bounds. Notice that the bounds $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ coincide when

$$\frac{(\kappa+1)^2}{4\kappa} \rho_a = 1.$$

Let $\bar{a}$ and $\overline{a^2}$ respectively stand for the average value and average of the squared values of a . One can more generally highlight the dependencies of the three quantities with the parameters using standard algebraic manipulations, yielding

$$\psi_{n,\kappa} \leq 4n\kappa \left(2c + \frac{\|b\|}{n\bar{a}} \right)^2 = O(n\kappa);$$

$$\chi_{n,\kappa} \leq 8 \left(\frac{a_+^2}{a^2} \left(\frac{\|b\|}{n\bar{a}} + 2c \right) \kappa \right)^2 = O(\kappa^2);$$

$$\theta_{n,\kappa} \leq 4 \left(\frac{a_+^2}{a^2} \left(\frac{\kappa}{2(n-1)} + 2 \right) \right) \left(\frac{a_+^2}{a^2} \left(\frac{\|b\|}{n\bar{a}} + 2c \right) \kappa \right)^2 = O \left(\kappa^2 + \frac{\kappa^3}{n} \right).$$

The linear scaling of $\psi_{n,\kappa}$ in both n and κ and the higher order scaling of both $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ in only κ suggest that $\psi_{n,\kappa}$ is tighter for small values of n and large values of κ , whereas $\theta_{n,\kappa}$ and $\chi_{n,\kappa}$ are tighter otherwise. The main difference between $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ lies in a multiplicative factor, constant for the former, and depending of the parameters and the values in a for the latter. In general, $\chi_{n,\kappa}$ tends to be tighter than $\theta_{n,\kappa}$ as κ gets large and n small. This difference becomes significant in heterogeneous settings, where it can get tighter than $\psi_{n,\kappa}$ as well. These behaviors are illustrated in Fig. 1.

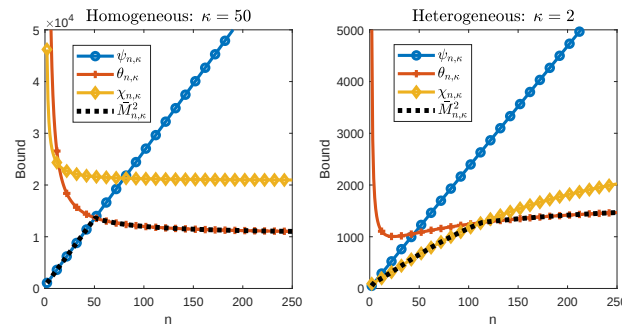


Fig. 1. Bounds of Proposition 2 with respect to the system size n for $b=1$, $c=1$, respectively for $\kappa=50$ with homogeneous agents ($a_i=1$ for all i) on the left, and $\kappa=2$ with heterogeneous agents ($a_1=10$, $a_i=1$ for $i>1$) on the right. The plots show all three quantities $\psi_{n,\kappa}$, $\theta_{n,\kappa}$ and $\chi_{n,\kappa}$ as well as the final bound $\bar{M}_{n,\kappa}^2$ for both cases.

Remark 3. *The interpretation of the quantities $\psi_{n,\kappa}$, $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ actually depends on the implicit assumption that $\|b\|$ is fixed and $\|a\|_1$ scales with n (*i.e.*, $\bar{a}$ is fixed). This particular modelling choice is arbitrary, and implies that the solution held by an agent $x_i^{*,k}$ becomes smaller for large values of n . Other choices might have different implications*

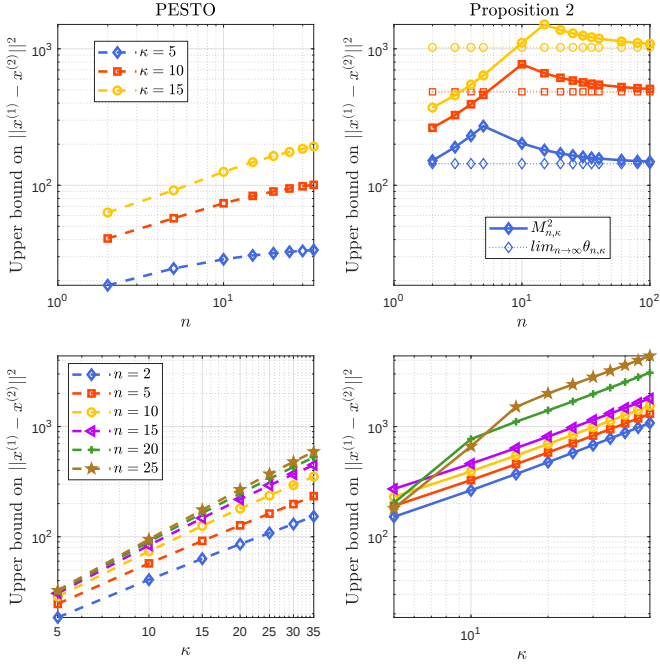


Fig. 2. The upper bound (16) on $\|x^{(1)} - x^{(2)}\|^2$ respectively with respect to n with homogeneous agents for several values of κ (top) and with respect to κ for several values of n (bottom). For each plot the bound obtained in Proposition 2 (right) is compared with the empirical upper bound derived using PESTO in the same settings (left). The top-right plot also shows the asymptotic value expected to be reached by $\theta_{n,\kappa}$ as $n \rightarrow \infty$ based on (19).

on the interpretation, and in particular on the scaling of these quantities. For instance one could choose to either fix $\|b\|$ and $\|a\|_1$, or that both $\|b\|$ and $\|a\|_1$ scale with n , so that the $x_i^{*,k}$ remain mostly the same no matter n (observe that the latter choice yields the same scalings than those presented above).

The result of Proposition 2 can be analyzed with respect to empirical results derived with the PESTO toolbox [35], which allows computing exact empirical bounds for quantities related to convex functions. A similar analysis was performed in [32], and we thus refer to Appendix B of [32] for details about the PESTO simulation. For the sake of simplicity, the analysis here is only done for the homogeneous case, and consequently does not involve $\chi_{n,\kappa}$; similar conclusions could however be drawn the same way using heterogeneous agents.

We can observe in Fig. 2 that even though there is some gap between the theoretical result and that obtained using PESTO, the scaling of the bounds with respect to n and κ is well captured. In particular, the top plot shows that $\bar{M}_{n,\kappa}^2 = \theta_{n,\kappa}$ when n becomes large, resulting in the convergence of $\bar{M}_{n,\kappa}^2$ towards a constant, consistently with the result obtained with PESTO. In parallel, the bottom plot suggests that the bounds from PESTO asymptotically grows linearly with κ , consistently with the evolution of $\psi_{n,\kappa}$, which is the value taken by $\bar{M}_{n,\kappa}^2$ for large values of κ .

IV. LINEAR CONVERGENCE OF RCD IN CLOSED SYSTEM

We now analyze the effect of the second type of events happening in the system, *i.e.*, pairwise interactions resulting

in RCD updates. This corresponds to studying the linear convergence of the RCD Algorithm in closed system.

A. Linear convergence and $\mathbf{L}_p^+$ -seminorm

In this section, we derive the constant of convergence of the RCD algorithm in terms of the distance to the minimizer with the objective of characterizing the effect of a single RCD step on that expected distance at interaction events. We introduce the following standard definitions [36].

Definition 1 (Q-Linear Convergence). Let $\{x^k\}$ be the sequence of points converging to some point $x^* \in \mathbb{R}^d$ generated by some algorithm. For any norm $\|\cdot\|$, we say the convergence is Q-linear if there exists $r \in (0, 1)$ such that for all k

$$\|x^{k+1} - x^*\| \leq r \|x^k - x^*\|.$$

The number r is called the constant of convergence.

Definition 2 (R-Linear Convergence). Let $\{x^k\}$ be the sequence of points converging to some point $x^* \in \mathbb{R}^d$ generated by some algorithm. For any norm $\|\cdot\|$, we say the convergence is R-linear if there exists $r \in (0, 1)$ and some positive constant C such that for all k

$$\|x^k - x^*\| \leq Cr^k.$$

In the rest of the work, we will refer to Q-linear convergence just as linear convergence. R-linear convergence is typically referred as exponential convergence in control systems theory. Clearly, R-linear convergence is weaker than linear convergence since it is concerned with the overall rate of decrease in the error, rather than the decrease over each individual iteration of the algorithm [37].

In [10], the author proves linear convergence of the RCD algorithm in expectation in terms of the function value, *i.e.*, $f(x) - f(x^*)$. Hence, from the inequalities corresponding to smooth functions and strong convexity [38, 39], it is straightforward to prove R-linear convergence of the algorithm from [10, Eq. (26)]:

$$\mathbb{E} [\|x^k - x^*\|] \leq \kappa(1 - \alpha\lambda_2)^k \|x_0 - x^*\|. \quad (20)$$

However, due to the alternation of updates and replacements, our analysis in open systems requires the strict contraction of some metric after each iteration. The linear convergence of the RCD algorithm was established in the preliminary work [32] for the Euclidean norm under the assumption of a complete communication graph with homogeneous agents and uniform probabilities p_{ij} . Nevertheless, the following example shows that such contraction no longer holds for the Euclidean norm for general graphs.

Example. Consider a line graph with 3 agents satisfying the constraint $\langle \mathbf{1}, x \rangle = -3$ with probabilities $p_{12} = 0.9$, $p_{23} = 0.1$ (and hence $p_{13} = 0$), and whose local cost functions and estimates at iteration k are:

i	$f_i(x_i)$	x_i^*	x_i^k
1	$50(x_1 - 2)^2$	2	10
2	$20(x_2 + 2)^2$	-2	7
3	$(x_3 + 3)^2$	-3	-20

Starting from x^k the expected result of the RCD operation with step-size $h = 1/\beta = 0.01$ is

$$\mathbb{E}[\|x^{k+1} - x^*\|^2] = 437.204 > 434 = \|x^k - x^*\|^2, \quad (21)$$

and hence linear convergence cannot be achieved.

Remark 4 (Weighted gradient descent). Notice that the expected behavior of the RCD algorithm (10) is linked with the weighted gradient descent, whose convergence has been studied e.g., in [8, 40, 41]. Those works, however, do not prove linear convergence, which is required for the approach we follow in this paper.

For this reason, we propose to study the problem in a different norm associated with the algorithm. Since the RCD is performed along a network of agents, a natural choice is to consider norms induced by associated matrices as in [42]. In this case, we focus on the seminorm induced by the Moore-Penrose inverse of the matrix $\mathbf{L}_p$ introduced in (9), denoted by $\mathbf{L}_p^\dagger$, and defined as follows for some $x \in \mathbb{R}^{nd}$:

$$\|x\|_{\mathbf{L}_p^\dagger} := \sqrt{x^\top \mathbf{L}_p^\dagger x}. \quad (22)$$

We show with the next proposition that this seminorm is a norm on $\mathcal{S}_{a,0}$, where we recall that $\mathcal{S}_{a,0}$ is the feasible set defined in (3) when $b = \mathbf{0}_d$ and corresponds to the kernel of $a^\top \otimes I_d$. For the particular case $d = 1$, $\mathcal{S}_{a,0}$ is the orthogonal complement of a .

Proposition 3. The seminorm $\|\cdot\|_{\mathbf{L}_p^\dagger}$ is a norm on $\mathcal{S}_{a,0}$.

Proof. By (22), $\|x\|_{\mathbf{L}_p^\dagger} = 0$ implies that x must be in the kernel of the matrix $\mathbf{L}_p^\dagger$. From (9) we have $\mathbf{L}_p^\dagger = L_p^\dagger \otimes I_d$, and since the kernel of $\mathbf{L}_p^\dagger$ is spanned by all the eigenvectors of $\mathbf{L}_p^\dagger$ corresponding to a zero eigenvalue, we get

$$\ker(\mathbf{L}_p^\dagger) = \{x \in \mathbb{R}^{nd} \mid x = a \otimes w, w \in \mathbb{R}^d\}.$$

Since $x \in \mathcal{S}_{a,0}$, it must satisfy $(a^\top \otimes I_d)x = \mathbf{0}_{nd}$ and we have:

$$(a^\top \otimes I_d)x = (a^\top \otimes I_d)(a \otimes w) = \|a\|^2 w,$$

which is equal to $\mathbf{0}_d$ only for $w = \mathbf{0}_d$. $\square$

If $x, y \in \mathcal{S}_{a,b}$, then $z = x - y$ belongs to $\mathcal{S}_{a,0}$, so that the norm $\|\cdot\|_{\mathbf{L}_p^\dagger}$ can be used to measure the distance between two vectors in the context of this work.

B. Contraction of an iteration in closed system

Let us remind the update rule of the RCD algorithm defined in (7) for some positive step-size h as

$$x^+ = x - h\mathbf{Q}^{ij}\nabla f(x). \quad (23)$$

In the following proposition, we analyze the convergence of (23) with respect to the norm induced by $\mathbf{L}_p^\dagger$ defined in the previous section.

Proposition 4. Let a function $f(x) := \sum_{i=1}^n f_i(x_i)$ and $x^* := \arg \min_{x \in \mathcal{S}_{a,b}} f(x)$. Under Assumption 1, for any positive scalar

$$h \leq \frac{\lambda_2}{\lambda_n} \frac{2}{\alpha + \beta}, \quad (24)$$

and for any initial point $x \in \mathcal{S}_{a,b}$, then the update rule (23) applied on the randomly selected pair of agents $(i, j) \in \mathcal{E}$ satisfies

$$\mathbb{E}[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2] \leq (1 - 2h\alpha\lambda_2 + h^2\alpha^2\lambda_n) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (25)$$

Proof. By definition:

$$\begin{aligned} \mathbb{E}[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2] &= \sum_{(i,j) \in \mathcal{E}} p_{ij} \|x - h\mathbf{Q}^{ij}\nabla f(x) - x^*\|_{\mathbf{L}_p^\dagger}^2 \\ &= \|x - x^*\|_{\mathbf{L}_p^\dagger}^2 + h^2 \sum_{(i,j) \in \mathcal{E}} p_{ij} \|\mathbf{Q}^{ij}\nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 \\ &\quad - 2h \sum_{(i,j) \in \mathcal{E}} p_{ij} \langle \mathbf{Q}^{ij}\nabla f(x), \mathbf{L}_p^\dagger(x - x^*) \rangle. \end{aligned} \quad (26)$$

Since $\mathbf{L}_p = \sum_{(i,j) \in \mathcal{E}} p_{ij} \mathbf{Q}^{ij}$, it follows that

$$\begin{aligned} \mathbb{E}[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2] &= \|x - x^*\|_{\mathbf{L}_p^\dagger}^2 + h^2 \sum_{(i,j) \in \mathcal{E}} p_{ij} \|\mathbf{Q}^{ij}\nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 \\ &\quad - 2h \langle \mathbf{L}_p \nabla f(x), \mathbf{L}_p^\dagger(x - x^*) \rangle. \end{aligned} \quad (27)$$

We first treat the second term of the right-hand side of (27). Remember from (11) that $\mathbf{Q}^{ij}a = \mathbf{0}_n$, and from (4) that $\nabla f(x^*) = a \otimes \lambda^*$ for some $\lambda^* \in \mathbb{R}^d$. Hence, since $\mathbf{Q}^{ij} = \mathbf{Q}^{ij} \otimes I_d$ by definition:

$$\mathbf{Q}^{ij}\nabla f(x^*) = (\mathbf{Q}^{ij} \otimes I_d)(a \otimes \lambda^*) = (\mathbf{Q}^{ij}a) \otimes \lambda^* = \mathbf{0}_{nd}. \quad (28)$$

It thus follows that

$$\begin{aligned} \|\mathbf{Q}^{ij}\nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 &= \|\mathbf{Q}^{ij}(\nabla f(x) - \nabla f(x^*))\|_{\mathbf{L}_p^\dagger}^2 \\ &\leq \frac{1}{\lambda_2} \|\mathbf{Q}^{ij}(\nabla f(x) - \nabla f(x^*))\|^2, \end{aligned} \quad (29)$$

where the inequality follows from the fact that the eigenvalues of $\mathbf{L}_p$ are exactly those of L_p repeated d times (by Theorem 13.12 of [43]), so that the smallest and largest nonzero eigenvalues of $\mathbf{L}_p^\dagger$ are respectively $1/\lambda_n$ and $1/\lambda_2$, yielding for all $z \in \mathbb{R}^{nd}$:

$$\|z\|_{\mathbf{L}_p^\dagger}^2 \leq \frac{1}{\lambda_2} \|z\|^2. \quad (30)$$

Therefore, since $\mathbf{Q}^{ij} = (\mathbf{Q}^{ij})^\top = (\mathbf{Q}^{ij})^2$, and using the fact that $\|z\|_{\mathbf{L}_p}^2 \leq \lambda_n \|z\|^2$ for all $z \in \mathbb{R}^{nd}$, it follows from (29):

$$\begin{aligned} \sum_{(i,j) \in \mathcal{E}} p_{ij} \|\mathbf{Q}^{ij}\nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 &\leq \frac{1}{\lambda_2} \|\nabla f(x) - \nabla f(x^*)\|_{\mathbf{L}_p}^2 \\ &\leq \frac{\lambda_n}{\lambda_2} \|\nabla f(x) - \nabla f(x^*)\|^2. \end{aligned} \quad (31)$$

We now analyze the third term of the right-hand side of (27). From (28) we get

$$\mathbf{L}_p \nabla f(x^*) = \sum_{(i,j) \in \mathcal{E}} p_{ij} \mathbf{Q}^{ij} \nabla f(x^*) = \mathbf{0}_{nd}, \quad (32)$$

yielding [38, Thm. 2.1.12]:

$$\langle \nabla f(x) - \nabla f(x^*), x - x^* \rangle \geq \frac{\beta^{-1} \|\nabla f(x) - \nabla f(x^*)\|^2}{1 + \kappa^{-1}} + \frac{\alpha \|x - x^*\|^2}{1 + \kappa^{-1}}.$$

Hence, using the result above and (30), it follows that

$$\begin{aligned} &-2h \langle \mathbf{L}_p \nabla f(x), \mathbf{L}_p^\dagger(x - x^*) \rangle \\ &\leq -2h \frac{\beta^{-1} \|\nabla f(x) - \nabla f(x^*)\|^2}{1 + \kappa^{-1}} - 2h \frac{\alpha \lambda_2 \|x - x^*\|_{\mathbf{L}_p^\dagger}^2}{1 + \kappa^{-1}}. \end{aligned} \quad (33)$$

Injecting (31) and (33) into (26) yields

$$\mathbb{E} \left[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \right] \leq \left(1 - 2h \frac{\alpha \lambda_2}{1 + \kappa^{-1}} \right) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2 + \left(h^2 \frac{\lambda_n}{\lambda_2} - 2h \frac{\beta^{-1}}{1 + \kappa^{-1}} \right) \|\nabla f(x) - \nabla f(x^*)\|^2. \quad (34)$$

Observe that if $h \leq \frac{\lambda_2}{\lambda_n} \frac{2}{\alpha + \beta}$ then $h^2 \frac{\lambda_n}{\lambda_2} - 2h \frac{\beta^{-1}}{1 + \kappa^{-1}} \leq 0$. As a consequence, (34) can be upper bounded using the definition of α -strongly convex functions, and more specifically

$$\|\nabla f(x) - \nabla f(x^*)\|^2 \geq \alpha^2 \lambda_2 \|x - x^*\|_{\mathbf{L}_p^\dagger}^2,$$

where we used the fact that $\|x - x^*\|^2 \geq \lambda_2 \|x - x^*\|_{\mathbf{L}_p^\dagger}^2$. It follows that

$$\mathbb{E} \left[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \right] \leq \|x - x^*\|_{\mathbf{L}_p^\dagger}^2 + (h^2 \alpha^2 \lambda_n - 2h \alpha \lambda_2) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2,$$

which concludes the proof. $\square$

It is clear that the constant of convergence is less than one if $h \leq \frac{2\lambda_2}{\alpha \lambda_n}$, which is thus satisfied on all its range of validity since $h \leq \frac{2\lambda_2}{(\alpha + \beta)\lambda_n} \leq \frac{2\lambda_2}{\alpha \lambda_n}$. We can then find the step-size which minimizes (25) and the corresponding constant of convergence.

Corollary 1. *The optimal constant of convergence in (25) under (24) is achieved for $h^* = \frac{2\lambda_2}{(\alpha + \beta)\lambda_n}$ which yields*

$$\mathbb{E} \left[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \right] \leq \left(1 - \frac{\lambda_2^2}{\lambda_n \kappa} \right) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (35)$$

Interestingly, Proposition 4 shows that linear convergence can be achieved by the RCD algorithm with respect to the norm induced by $\mathbf{L}_p^\dagger$ with a constant of convergence similar to that of classical algorithms based on gradient descent [38, 39].

Remark 5 (Complete graph). *For the particular case of a complete graph with 1-dimensional homogeneous agents and uniform probabilities, the eigenvalues of L_p are $\lambda_2 = \lambda_n = \frac{1}{n-1}$ and the $\mathbf{L}_p^\dagger$ -norm coincides with the Euclidean norm for all $z = x - y$, where $x, y \in \mathcal{S}_{a,b}$. Then, the result of Proposition 4 becomes*

$$\mathbb{E} [\|x^+ - x^*\|^2] \leq \left(1 - \frac{\alpha h}{n-1} (2 - \alpha h) \right) \|x - x^*\|^2.$$

Since in that case by definition $h \leq \frac{2}{\alpha + \beta} \leq \frac{1}{\alpha}$, it follows that

$$\mathbb{E} [\|x^+ - x^*\|^2] \leq \left(1 - \frac{\alpha h}{n-1} \right) \|x - x^*\|^2,$$

which coincides with [32, Eq. (13)].

Remark 6 (Alternative rate). *Starting from (34) in the proof of Proposition 4, one can use a similar argument to derive the following alternative constant of convergence, valid for $\frac{\lambda_2}{\lambda_n} \frac{2}{\alpha + \beta} \leq h \leq \frac{\kappa^{-1} + \kappa_L}{\kappa^{-1} + 1} \frac{\lambda_2}{\lambda_n^2 \beta}$, with $\kappa_L = \frac{\lambda_n}{\lambda_2}$:*

$$\mathbb{E} \left[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \right] \leq \left(1 - 2\beta \lambda_2 h + h^2 \beta^2 \frac{\lambda_n^2}{\lambda_2} \right) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (36)$$

This result could be used in the rest of this work the same way as that of Proposition 4 for the corresponding step-size. This development is however omitted in this work.

C. Homogeneous agents and uniform probabilities

Assumption 4. *The agents are homogeneous (i.e., $a = \mathbb{1}_n$) and the probabilities of selecting the edges during the implementation of the RCD algorithm are uniform (i.e., $p_{ij} = p$).*

For the particular case of homogeneous agents and uniform probabilities, the matrix L_p can be expressed as $L_p = \frac{p}{2} L$ where L is the usual Laplacian matrix. In this case, the matrix L_p can be associated to an electrical circuit [44], and we can use the concept of effective resistance to find an upper bound for the step size of the algorithm independently of λ_2 .

Hence, the following proposition provides an alternative bound for the convergence of the RCD algorithm in the specific case described above, and can be used the same way as that of Proposition 4 in the remainder of this work for that case. However, for the sake of generality, we express the main result in the next section only in terms of Proposition 4.

Proposition 5. *Let a function $f(x) := \sum_{i=1}^n f_i(x_i)$ and $x^* := \arg \min_{x \in \mathcal{S}_b} f(x)$. Under Assumptions 1 and 4, for any positive scalar*

$$h \leq \frac{2p}{\lambda_n} \frac{2}{\alpha + \beta}, \quad (37)$$

and for any initial point $x \in \mathcal{S}_{a,b}$, then the update rule (23) applied on the randomly selected pair of agents $(i, j) \in \mathcal{E}$ satisfies

$$\mathbb{E} \left[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \right] \leq \left(1 - 2h\alpha\lambda_2 + \frac{h^2\alpha^2\lambda_2\lambda_n}{2p} \right) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (38)$$

Proof. Since the matrices $\mathbf{Q}^{ij}$ are idempotent, the summation term of the second element in (27) can be expressed as:

$$\|\mathbf{Q}^{ij} \nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 = \langle \mathbf{Q}^{ij} \nabla f(x), \mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij} \mathbf{Q}^{ij} \nabla f(x) \rangle$$

Then we can use an upper bound for the quadratic form and we obtain for each term:

$$\|\mathbf{Q}^{ij} \nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 \leq \|\mathbf{Q}^{ij} \nabla f(x)\|^2 \lambda_{\max}(\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij}).$$

Now, the matrix $\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij}$ is given by:

$$\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij} = (\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij}) \otimes I_d,$$

which implies that $\lambda_{\max}(\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij}) = \lambda_{\max}(\mathbf{Q}^{ij} \mathbf{L}_p^\dagger \mathbf{Q}^{ij})$. Then, we have:

$$\mathbf{Q}^{ij} \mathbf{L}^\dagger \mathbf{Q}^{ij} = \frac{1}{2} \left([\mathbf{L}^\dagger]_{ii} + [\mathbf{L}^\dagger]_{jj} - 2[\mathbf{L}^\dagger]_{ij} \right) \mathbf{Q}^{ij} = \frac{1}{2} r_{ij} \mathbf{Q}^{ij},$$

where r_{ij} is the effective resistance between the agents i and j . Since there is an edge between i and j , we have $r_{ij} \leq 1/p$. Then we have the following upper bound for the largest eigenvalue:

$$\lambda_{\max}(\mathbf{Q}^{ij} \mathbf{L}^\dagger \mathbf{Q}^{ij}) \leq \frac{1}{2p} \text{ for all } (i, j) \in E, \quad (39)$$

and we get:

$$\begin{aligned} \sum_{(i,j) \in \mathcal{E}} p_{ij} \|\mathbf{Q}^{ij} \nabla f(x)\|_{\mathbf{L}_p^\dagger}^2 &\leq \frac{1}{2p} \|\nabla f(x) - \nabla f(x^*)\|_{\mathbf{L}_p}^2 \\ &\leq \frac{\lambda_n}{2p} \|\nabla f(x) - \nabla f(x^*)\|^2, \end{aligned} \quad (40)$$

which replaces (31). The rest of the proof follows the same steps as in the proof of Proposition 4. $\square$

Similarly to Proposition 4, the constant of convergence is strictly decreasing if $h \leq \frac{4p}{\alpha\lambda_n}$, which is always satisfied since $h \leq \frac{4p}{(\alpha+\beta)\lambda_n} \leq \frac{4p}{\alpha\lambda_n}$. Hence, we can also find the optimal step-size for the algorithm, and the corresponding constant of convergence.

Corollary 2. *The optimal constant of convergence in (38) under (37) is achieved for $h^* = \frac{4p}{(\alpha+\beta)\lambda_n}$ which yields*

$$\mathbb{E}[\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2] \leq \left(1 - \frac{2p\lambda_2}{\lambda_n} \frac{1}{\kappa}\right) \|x - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (41)$$

Remark 7. *The upper bound for the step size derived in Proposition 5 is better suited for graphs with a small λ_2 (also known as algebraic connectivity), that is, non-robust networks that can be easily disconnected [45]. If we denote by μ the eigenvalues of L , which satisfy $\lambda = \frac{p}{2}\mu$, then we have that for Proposition 4 the step size must satisfy $h \leq \frac{\mu_2}{\mu_n} \frac{2}{\alpha+\beta}$ while for Proposition 5 the step size is upper bounded by $h \leq \frac{4}{\mu_n} \frac{2}{\alpha+\beta}$.*

V. CONVERGENCE OF RCD IN OPEN SYSTEM

Let us define the ratio

$$\rho_R := \frac{1 - p_U}{p_U}, \quad (42)$$

which characterizes the expected number of replacements happening in the system between two consecutive RCD updates. In particular, when $\rho_R \rightarrow 0$, then the system converges to a closed system, and when $\rho_R \rightarrow \infty$, then replacements become so frequent that RCD updates are negligible.

In the following theorem we present the main result of this work, in which we derive the constant of convergence for the error achieved by the RCD algorithm in expectation in open system, under the form of an upper bound on that error. The derivation of this result relies on the separate analysis of the effects of replacements and of RCD updates, which is enabled by Assumption 3.

Theorem 1. *Let $M_{n,\kappa} = \frac{1}{\lambda_2} \bar{M}_{n,\kappa}$ with $\bar{M}_{n,\kappa}$ defined in (16), and let $\bar{\eta} := \rho_R \frac{M_{n,\kappa}}{\alpha h(2\lambda_2 - \lambda_n \alpha h)}$. In the setting described in Section II, the sequence of estimates x^k generated by the iteration rule (23) with $h \leq \frac{\lambda_2}{\lambda_n} \frac{2}{\alpha+\beta}$ satisfies for all k and for any $\eta > \bar{\eta}$:*

$$\mathbb{E}[\|x^{k+1} - x^{*,k+1}\|_{\mathbf{L}_p^\dagger}^2] - \Gamma_\eta \leq A_\eta \left(\mathbb{E}[\|x^k - x^{*,k}\|_{\mathbf{L}_p^\dagger}^2] - \Gamma_\eta \right), \quad (43)$$

with

$$A_\eta := 1 - p_U \alpha h(2\lambda_2 - \alpha \lambda_n h) + (1 - p_U) \frac{M_{n,\kappa}}{\eta}; \quad (44)$$

$$\Gamma_\eta := \frac{(1 - p_U) M_{n,\kappa} (\eta + M_{n,\kappa}) \eta}{p_U \eta \alpha h(2\lambda_2 - \alpha \lambda_n h) - (1 - p_U) M_{n,\kappa}}. \quad (45)$$

Proof. Let us denote $C^k := \|x^k - x^{*,k}\|_{\mathbf{L}_p^\dagger}^2$. From Assumption 3, we have

$$\mathbb{E}[C^{k+1}] = p_U \mathbb{E}[C^{k+1}|U] + (1 - p_U) \mathbb{E}[C^{k+1}|R], \quad (46)$$

where U and R respectively stand for the occurrence of an RCD update and a replacement event. Proposition 4 then yields for $h \leq \frac{\lambda_2}{\lambda_n} \frac{2}{\alpha+\beta}$

$$\mathbb{E}[C^{k+1}|U] \leq (1 - 2\lambda_2 \alpha h + \lambda_n \alpha^2 h^2) \mathbb{E}[C^k]. \quad (47)$$

Under a replacement event, we have $x^{k+1} = x^k$, and hence Proposition 2 yields

$$\begin{aligned} \mathbb{E}[C^{k+1}|R] &= \mathbb{E}[\|x^{k+1} - x^{k,*} + x^{k,*} - x^{k+1,*}\|_{\mathbf{L}_p^\dagger}^2] \\ &\leq \mathbb{E}\left[\left(\|x^k - x^{k,*}\|_{\mathbf{L}_p^\dagger} + \|x^{k,*} - x^{k+1,*}\|_{\mathbf{L}_p^\dagger}\right)^2\right] \\ &= \mathbb{E}\left[\left(\sqrt{C^k} + M_{n,\kappa}\right)^2\right]. \end{aligned} \quad (48)$$

Injecting (47) and (48) into (46) then yields the following nonlinear recurrence:

$$\begin{aligned} \mathbb{E}[C^{k+1}] &\leq (1 - p_U \alpha h(2\lambda_2 - \lambda_n \alpha h)) \mathbb{E}[C^k] \\ &\quad + (1 - p_U) \left(2M_{n,\kappa} \mathbb{E}[\sqrt{C^k}] + M_{n,\kappa}^2\right). \end{aligned} \quad (49)$$

Since $2x \leq \eta + \frac{x^2}{\eta}$ holds for all $x \geq 0$ and $\eta > 0$, we obtain $\sqrt{C^k} \leq \frac{\eta}{2} + \frac{C^k}{2\eta}$ for any $\eta > 0$. Hence, it follows that

$$\begin{aligned} \mathbb{E}[C^{k+1}] &\leq \left(1 - p_U \alpha h(2\lambda_2 - \lambda_n \alpha h) + (1 - p_U) \frac{M_{n,\kappa}}{\eta}\right) \mathbb{E}[C^k] \\ &\quad + (1 - p_U) M_{n,\kappa} (\eta + M_{n,\kappa}). \end{aligned}$$

Observe that $1 - p_U \alpha h(2\lambda_2 - \lambda_n \alpha h) + (1 - p_U) \frac{M_{n,\kappa}}{\eta} < 1$ for $\eta > \bar{\eta}$, so that solving the linear recurrence yields the conclusion. $\square$

Notice that the setting of Theorem 1 guarantees that the contraction rate satisfies $A_\eta < 1$ (which is ensured for any $\eta > \bar{\eta}$) for any parametrization of the system (as long as $\rho_R < \infty$, i.e., if updates happen). Hence, Γ_η provides an upper bound on the asymptotic expected error, and a few algebraic manipulations yield

$$\limsup_{k \rightarrow \infty} \mathbb{E}[\|x^k - x^{*,k}\|_{\mathbf{L}_p^\dagger}^2] \leq \Gamma_\eta = \frac{\bar{\eta} (M_{n,\kappa} + \eta)}{1 - \bar{\eta}/\eta}. \quad (50)$$

When $\eta \rightarrow \infty$, then the contraction rate becomes minimal, i.e., $A_\eta \rightarrow 1 - p_U \alpha h(2\lambda_2 - \lambda_n \alpha h)$, and the asymptotic error diverges. Observe that for $\eta > \bar{\eta}$, then Γ_η is convex, and one can determine the value of η that minimizes the upper bound on the asymptotic expected error Γ_η , as presented in the following corollary.

Corollary 3. *When $\eta = \eta^* = \bar{\eta} \left(1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}\right)$, the convergence of the RCD algorithm in open system is guaranteed with minimal upper bound on the asymptotic error Γ_{η^*} , and:*

$$A_{\eta^*} = 1 - p_U \alpha h(2\lambda_2 - \lambda_n \alpha h) \frac{\sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}}{1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}}; \quad (51)$$

$$\Gamma_{\eta^*} = (\eta^*)^2 = \bar{\eta}^2 \left(1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}\right)^2, \quad (52)$$

where we recall that $\bar{\eta} := \rho_R \frac{M_{n,\kappa}}{\alpha h(2\lambda_2 - \lambda_n \alpha h)}$.

Proof. Observe that Γ_η is convex for $\eta > \bar{\eta}$, and we have $\frac{d}{d\eta}\Gamma|_{\eta=\eta^*} = 0$ with $\eta^* = \arg \min_{\eta > \bar{\eta}} \Gamma_\eta$. Hence we compute

$$\frac{d}{d\eta}\Gamma_\eta = \frac{\bar{\eta}}{(\eta - \bar{\eta})^2}(\eta^2 - 2\bar{\eta}\eta - M_{n,\kappa}\bar{\eta}) = 0, \quad (53)$$

which is satisfied for

$$\eta_1^* = \bar{\eta} + \sqrt{\bar{\eta}^2 + M_{n,\kappa}\bar{\eta}} \quad ; \quad \eta_2^* = \bar{\eta} - \sqrt{\bar{\eta}^2 + M_{n,\kappa}\bar{\eta}}.$$

Since $\eta_2^* < \bar{\eta}$, it must be rejected, and it follows that

$$\eta^* = \bar{\eta} + \sqrt{\bar{\eta}^2 + M_{n,\kappa}\bar{\eta}} = \bar{\eta} \left(1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}} \right).$$

We can then compute

$$A_{\eta^*} = 1 - p_U \alpha h (2\lambda_2 - \lambda_n \alpha h) + \frac{(1 - p_U)M_{n,\kappa}}{\bar{\eta} (1 + \sqrt{1 + M_{n,\kappa}/\bar{\eta}})},$$

and a few algebraic manipulations yield (51). Now observe that $\Gamma_\eta = \eta^2$ if and only if

$$\eta^2 - 2\bar{\eta}\eta - M_{n,\kappa}\bar{\eta} = 0,$$

which is equivalent to (53) for $\eta > \bar{\eta}$, so that the solution is η^* . Hence, $\Gamma_{\eta^*} = (\eta^*)^2$, which yields (52). $\square$

Remark 8. The bound Γ_η in (50) depends on $M_{n,\kappa}$ and $\bar{\eta}$, which itself depends on $M_{n,\kappa}$ and other parameters as well. In fact, $\bar{\eta}$ can be interpreted as the ratio between the spurious effect of replacements and the advantageous effect of RCD steps on the error. Briefly, this means that the greater the effect of replacements and the smaller that of updates (as discussed in the previous sections), then the larger the asymptotic error.

Observe that Theorem 1 and Corollary 3 provide *upper bounds* on the expected error of the algorithm, and therefore induce a certain conservatism with respect to the actual error. This essentially follows from Proposition 2 whose aim is to bound the additive error injected at one single replacement, whereas tighter bounds might be derived on the sum of those additive errors. This is especially true as ρ_R grows.

When no replacements happen, *i.e.*, $\rho_R \rightarrow 0$, then the system behaves as a closed system, and we retrieve the corresponding convergence behavior: the expected asymptotic error $\Gamma_\eta \rightarrow 0$ and the contraction rate $A_\eta \rightarrow 1 - \alpha h (2\lambda_2 - \lambda_n \alpha h)$ for all $\eta > 0$, consistently with the constant of convergence of the RCD in closed system derived in (25). By contrast, as ρ_R gets larger, *i.e.*, as replacements become more frequent, then the expected asymptotic error increases, and the contraction rate A_{η^*} gets closer to 1 (observe that $A_{\eta^*} < 1$ remains true as long as $\rho_R < \infty$). In the particular limit case where $\rho_R \rightarrow \infty$, then the minimal upper bound on the expected asymptotic error becomes $\Gamma_{\eta^*} \rightarrow 4\bar{\eta} \rightarrow \infty$ and $A_{\eta^*} \rightarrow 1$.

Interestingly, within the allowed range of h , Γ_{η^*} decays as the step-size h increases, suggesting that choosing h as large as possible leads to the smallest value for the upper bound on the expected asymptotic error Γ_{η^*} . This means that the only limitation on the choice of the step-size comes from the analysis of the algorithm in closed system ($h \leq \frac{\lambda_2}{\lambda_n} \frac{2}{\alpha + \beta}$ from Proposition 4 in our case), and that no particular precaution should be taken regarding the open character of the system.

Remark 9. The methodology we used in this section can easily be extended to other algorithms than the RCD algorithm. In particular the results of Theorem 1 and Corollary 3 can be adapted to any algorithm with linear convergence in closed systems, that is, such that

$$\|x^+ - x^*\|_{\mathbf{L}_p^\dagger}^2 \leq K \|x - x^*\|_{\mathbf{L}_p^\dagger}^2, \quad (54)$$

with some positive $K < 1$. In that case, the same constant of convergence as that presented in Theorem 1 is obtained with

$$A_\eta := 1 - p_U(1 - K) + (1 - p_U) \frac{M_{n,\kappa}}{\eta}; \quad (55)$$

$$\Gamma_\eta := \frac{(1 - p_U)M_{n,\kappa}(\eta + M_{n,\kappa})\eta}{p_U\eta(1 - K) - (1 - p_U)M_{n,\kappa}}. \quad (56)$$

We can show that convergence can be guaranteed in open system following a similar argument as that used to prove Theorem 1 if $K < 1$. Hence, this analysis can be applied *e.g.*, to the results presented in Proposition 5 or in Remark 6.

To illustrate the results of Theorem 1, we consider systems with piecewise quadratic local cost functions f_i : for $\phi_{i1}, \phi_{i2} \in [\frac{\alpha}{2}, \frac{\beta}{2}]$, the cost function f_i is given by

$$f_i(x_i) = \begin{cases} \phi_{i1}(x_i - v_i)^2, & \text{if } x_i < v_i \\ \phi_{i2}(x_i - v_i)^2, & \text{if } x_i \geq v_i \end{cases}, \quad (57)$$

where v_i is the minimizer of f_i satisfying Assumption 2. Such function therefore satisfies Assumption 1 as well. Observe that no assumption on the way we choose the local cost function f_i of a joining agent at replacements is required in the derivation of our results. Hence, we consider two possible cases for that choice: *random*, where the parameters ϕ_{i1} and ϕ_{i2} are uniformly randomly chosen in $[\frac{\alpha}{2}, \frac{\beta}{2}]$, and *adversarial*, where these parameters are arbitrarily chosen to maximize the error $\|x^k - x^{*,k}\|_{\mathbf{L}_p^\dagger}^2$ after the replacement among 100 realizations of such uniform random choice.

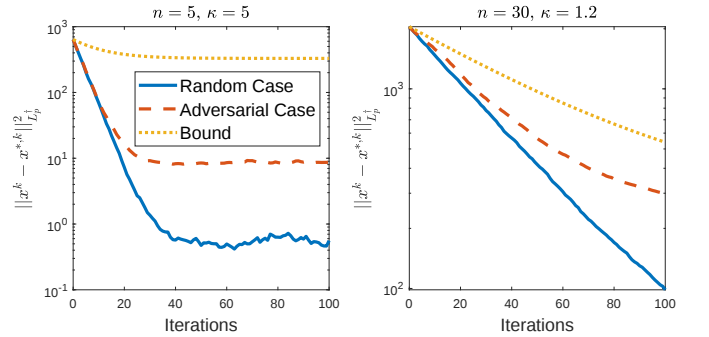


Fig. 3. Performance of the RCD algorithm in a complete graph constituted of respectively $n = 5$ agents with $\kappa = 5$ (left) and $n = 30$ agents with $\kappa = 1.2$ (right), with $p_U = 0.95$ and $b = 1$, and where each local objective function is defined by (57). The plain blue and red dashed lines represent the actual performance of the algorithm averaged over 500 realizations of the process, respectively for the random and adversarial replacements cases. The yellow dotted line is the upper bound (43) obtained from Corollary 3.

In Fig. 3, we show the evolution of the expected error $\mathbb{E}[\|x^k - x^{*,k}\|_{\mathbf{L}_p^\dagger}^2]$ simulated for a network with interconnections

defined by a complete graph, homogeneous agents and uniform probabilities. We consider two parametrizations of κ and n in both random and adversarial replacement cases, and we compare the simulations with (43) using the values given by Corollary 3. The figure shows that convergence is indeed guaranteed for the RCD in the presented settings and that the result of Corollary 3 shows some conservatism, which is inherited from Proposition 2. It is interesting to point out that these settings respectively make use of $\bar{M}_{n,\kappa}^2 = \theta_{n,\kappa}$ for $n = 30$, $\kappa = 1.2$, and of $\bar{M}_{n,\kappa}^2 = \psi_{n,\kappa}$ for $n = 5$, $\kappa = 5$, consistently with the description of $\bar{M}_{n,\kappa}^2$ in the homogeneous case of Section III. This highlights the impact of those parameters in the tightness of the bound $\bar{M}_{n,\kappa}^2$ used to derive our main results.

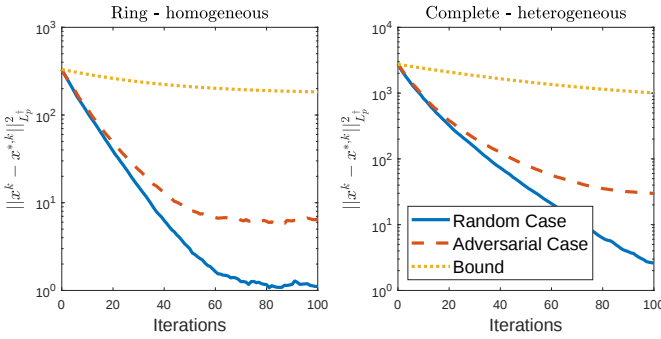


Fig. 4. Performance of the RCD algorithm with $n = 5$ agents, $\kappa = 1.2$, $p_U = 0.95$ and $b = 1$, respectively in (left) a ring graph with homogeneous agents (i.e., $a_i = 1$ for all i) and (right) a complete graph with heterogeneous agents (i.e., $a_1 = 10$, $a_i = 1$ for $i > 1$), and where each local objective function is defined by (57). The plain blue and red dashed lines represent the actual performance of the algorithm averaged over 500 realizations of the process, respectively for the random and adversarial replacements cases. The yellow dotted line is the upper bound (43) obtained from Corollary 3.

In Fig. 4, we compare the simulated performance in both replacement cases with the upper bound from Corollary 3 for a ring graph with homogeneous agents and a complete graph with heterogeneous agents. By contrast with the previous illustrations, the ring graph setting implies a different, sparse, topology which thus reduces the range of validity for the step-size h due to the small value of λ_2 that does not scale with n . Similarly, the heterogeneous setting impacts λ_2 and consequently reduces the range of h , due to the imbalance in L_p . Those moreover affect the behavior of the norm $\|\cdot\|_{L_p}^2$. Furthermore, the heterogeneous setting influences $\bar{M}_{n,\kappa}$ as well, and hence increases the effect of replacements on the bounds. Nevertheless, even though they differ quantitatively, the results of Fig. 4 are qualitatively similar to the case of the complete graph with homogeneous agents presented in Fig. 3.

VI. CONCLUSION

We have studied the behavior of the distance to the minimizer for the resource allocation problem and proved linear convergence of the Random Coordinate Descent algorithm in an appropriate norm for the closed system. We analyzed the algorithm for a general graph topology and possible heterogeneous agents in an open multi-agent systems scenario when agents can be replaced during the iterations. Under

replacement events, we proved that for an appropriate step-size, the algorithm cannot converge to the instantaneous minimizer due to the perturbations generated by the replacements but is stable. We derived an upper bound for the error in expectation which depends on the variation of the minimizer due to replacements and the frequency of these events.

A natural continuation of this work would be to handle the case where the budget and weights in the constraint can vary in time. Also, it would be interesting to consider more general equality constraints between the states of the agents and block updates at each iteration [46], such that the optimization is performed along more than one edge. Finally, a possible varying size of the system is an interesting direction for future work, where agents could join and leave the network independently of the current state of the system. This extension would however introduce a significant amount of new challenges, especially regarding the impact of (dis)connections of agents in terms of both graph properties and vector dimensions in our analysis.

REFERENCES

- [1] T. Ibaraki and N. Katoh, *Resource Allocation Problems: Algorithmic Approaches*. MIT press, 1988.
- [2] M. Patriksson, "A survey on the continuous nonlinear resource allocation problem," *European Journal of Operational Research*, vol. 185, 2008.
- [3] P. Dai, W. Yu, and D. Chen, "Distributed Q-learning algorithm for dynamic resource allocation with unknown objective functions and application to microgrid," *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 12 340–12 350, 2022.
- [4] P. Yi, Y. Hong, and F. Liu, "Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems," *Automatica*, vol. 74, pp. 259–269, 2016.
- [5] A. Teixeira, J. Araújo, H. Sandberg, and K. H. Johansson, "Distributed actuator reconfiguration in networked control systems," *IFAC Proceedings Volumes*, vol. 46, no. 27, pp. 61–68, 2013.
- [6] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.
- [7] A. D. Dominguez-Garcia, S. T. Cady, and C. N. Hadjicostis, "Decentralized optimal dispatch of distributed energy resources," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3688–3693.
- [8] L. Xiao and S. Boyd, "Optimal scaling of a gradient method for distributed resource allocation," *Journal of Optimization Theory and Applications*, vol. 129, no. 3, pp. 469–488, 2006.
- [9] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM Journal on Optimization*, vol. 22, no. 2, p. 341–362, 2012.
- [10] I. Necoara, "Random coordinate descent algorithms for multi-agent convex optimization over networks," *IEEE Transactions on Automatic Control*, vol. 58, no. 8, pp. 2001–2012, 2013.
- [11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT press, 2009.
- [12] L. Bird, M. Milligan, and D. Lew, "Integrating variable renewable energy: Challenges and solutions," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2013.
- [13] J. M. Hendrickx and S. Martin, "Open multi-agent systems: Gossiping with random arrivals and departures," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 763–768.
- [14] C. Monnuyer de Galland and J. M. Hendrickx, "Fundamental performance limitations for average consensus in open multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 68, no. 2, pp. 646–659, 2023.
- [15] C. Monnuyer de Galland, S. Martin, and J. M. Hendrickx, "Modelling gossip interactions in open multi-agent systems," *arXiv preprint arXiv:2009.02970*, 2020.
- [16] M. Franceschelli and P. Frasca, "Stability of open multi-agent systems and applications to dynamic consensus," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2326–2331, 2021.

- [17] Z. A. Z. S. Dashti, C. Seatzu, and M. Franceschelli, "Dynamic consensus on the median value in open multi-agent systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019.
- [18] R. Vizuete, P. Frasca, and E. Panteley, "On the influence of noise in randomized consensus algorithms," *IEEE Control Systems Letters*, vol. 5, no. 3, pp. 1025–1030, 2021.
- [19] V. S. Varma, I.-C. Morărescu, and D. Nešić, "Open multi-agent systems with discrete states and stochastic interactions," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 375–380, 2018.
- [20] J. M. Hendrickx and M. G. Rabbat, "Stability of decentralized gradient descent in open multi-agent systems," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 4885–4890.
- [21] Y.-G. Hsieh, F. Iutzeler, J. Malick, and P. Mertikopoulos, "Optimization in open networks via dual averaging," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 514–520.
- [22] E. Hazan, "Introduction to online convex optimization," *Foundations and Trends in Optimization*, vol. 2, pp. 157–325, 2016.
- [23] 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, 2018.
- [24] X. Li, L. Xie, and N. Li, "A survey of decentralized online learning," *arXiv preprint arXiv:2205.00473*, 2022.
- [25] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 2032–2048, 2020.
- [26] T. Yang, L. Zhang, R. Jin, and J. Yi, "Tracking slowly moving clairvoyant: Optimal dynamic regret of online learning with true and noisy gradient," in *Proceedings of The 33rd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. F. Balcan and K. Q. Weinberger, Eds., vol. 48. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 449–457. [Online]. Available: <https://proceedings.mlr.press/v48/yangb16.html>
- [27] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online Optimization : Competing with Dynamic Comparators," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, G. Lebanon and S. V. N. Vishwanathan, Eds., vol. 38. San Diego, California, USA: PMLR, 09–12 May 2015, pp. 398–406. [Online]. Available: <https://proceedings.mlr.press/v38/jadbabaie15.html>
- [28] X. Li, L. Xie, and N. Li, "A survey on distributed online optimization and game," *arXiv preprint arXiv:2205.00473*, 2023.
- [29] B. Wang, Q. Fei, and Q. Wu, "Distributed time-varying resource allocation optimization based on finite-time consensus approach," *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 599–604, 2021.
- [30] M. Doostmohammadian, A. Aghasi, M. Vrakopoulou, and T. Charalambous, "1st-order dynamics on nonlinear agents for resource allocation over uniformly-connected networks," in *2022 IEEE Conference on Control Technology and Applications (CCTA)*, 2022, pp. 1184–1189.
- [31] A.-S. Esteki and S. S. Kia, "Distributed optimal resource allocation with time-varying quadratic cost functions and resources over switching agents," in *2022 European Control Conference (ECC)*, 2022, pp. 441–446.
- [32] C. Monnayer de Galland, R. Vizuete, J. M. Hendrickx, P. Frasca, and E. Panteley, "Random coordinate descent algorithm for open multi-agent systems with complete topology and homogeneous agents," in *2021 IEEE 60th Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 1701–1708.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [34] A. Kurve, K. Kotobi, and G. Kesidis, "An agent-based framework for performance modeling of an optimistic parallel discrete event simulator," *Complex Adapt Syst Model*, vol. 1, no. 12, pp. 3206–1–12, 2013.
- [35] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Performance estimation toolbox (PESTO): Automated worst-case analysis of first-order optimization methods," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 1278–1283.
- [36] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, 2000.
- [37] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Science & Business Media, 2006.
- [38] Y. Nesterov, *Lectures on Convex Optimization*. Springer, 2018.
- [39] S. Bubeck, "Convex optimization: Algorithms and complexity," *Foundations and Trends in Machine Learning*, vol. 8, no. 3-4, 2015.
- [40] H. Lakshmanan and D. P. De Farias, "Decentralized resource allocation in dynamic networks of agents," *SIAM Journal on Optimization*, vol. 19, no. 2, pp. 911–940, 2008.
- [41] A. Cherukuri and J. Cortés, "Distributed generator coordination for initialization and anytime optimization in economic dispatch," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 226–237, 2015.
- [42] B. Wang, D. Zou, Q. Gu, and S. J. Osher, "Laplacian smoothing stochastic gradient Markov Chain Monte Carlo," *SIAM Journal on Scientific Computing*, vol. 43, no. 1, pp. A26–A53, 2021.
- [43] A. J. Laub, *Matrix Analysis For Scientists And Engineers*. USA: Society for Industrial and Applied Mathematics, 2004.
- [44] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Electrical networks and algebraic graph theory: Models, properties, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 977–1005, 2018.
- [45] A. Jamakovic and P. Van Mieghem, "On the robustness of complex networks by using the algebraic connectivity," in *International conference on research in networking*. Springer, 2008, pp. 183–194.
- [46] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Mathematical Programming*, vol. 144, no. 1, pp. 1–38, 2014.

APPENDIX

A. Proof of Lemma 1

Let us consider some $x \in \mathcal{S}_{a,b}$ such that $x \notin B(0, R_{b,\kappa})$, and let $\bar{x}^* = \arg \min_{x \in \mathbb{R}^{nd}} f(x)$ denote the global minimizer without constraint. We have $\|x\| > R_{b,\kappa}$ by definition and $\|\bar{x}^*\| \leq \sqrt{nc}$ since $\bar{x}^* \in B(0, c)^n$, and it thus follows that $\|x - \bar{x}^*\| > R_{b,\kappa} - \sqrt{nc}$. Hence, since f is α -strongly convex, and since $f(\bar{x}^*) = 0$ from Assumption 2, we have

$$f(x) \geq \frac{\alpha}{2} \|x - \bar{x}^*\|^2 > \frac{\alpha}{2} \kappa \left(\sqrt{nc} + \frac{\|b\|}{\|a\|} \right)^2 = \frac{\beta}{2} \left(\sqrt{nc} + \frac{\|b\|}{\|a\|} \right)^2.$$

Now let us define $x_b := \frac{1}{\|a\|^2} (a \otimes I_d) b$, which is a feasible point with norm $\|x_b\| = \frac{\|b\|}{\|a\|}$. Since f is β -smooth, and since $f(\bar{x}^*) = 0$ from Assumption 2, we get

$$f(x_b) \leq \frac{\beta}{2} \|x_b - \bar{x}^*\|^2 \leq \frac{\beta}{2} \left(\sqrt{nc} + \frac{\|b\|}{\|a\|} \right)^2.$$

Finally, since $x_b \in \mathcal{S}_{a,b}$, we have $f(x^*) \leq f(x_b)$ by definition. Combining all the inequalities above then yields

$$f(x) > \frac{\beta}{2} \left(\sqrt{nc} + \frac{\|b\|}{\|a\|} \right)^2 \geq f(x_b) \geq f(x^*),$$

which implies that x cannot be the minimizer of the problem and concludes the proof.

B. Proof of Lemma 2

Let us denote $\bar{x}_i^*$ the minimizer of f_i without constraint which satisfies $f_i(\bar{x}_i^*) = 0$ and $\nabla f_i(\bar{x}_i^*) = 0_d$. From β -smoothness of the local functions we have [38, Thm. 2.1.5]:

$$\|\nabla f_i(x_i^*)\|^2 \leq \beta \langle \nabla f_i(x_i^*), x_i^* - \bar{x}_i^* \rangle.$$

Then, from the optimality condition (4):

$$a_i^2 \|\lambda^*\|^2 \leq \beta \langle \lambda^*, a_i(x_i^* - \bar{x}_i^*) \rangle.$$

By summing over all the i , we obtain:

$$\|a\|^2 \|\lambda^*\|^2 \leq \beta \langle \lambda^*, \sum_{i=1}^n a_i(x_i^* - \bar{x}_i^*) \rangle.$$

We use the Cauchy-Schwarz inequality to get:

$$\|a\|^2 \|\lambda^*\|^2 \leq \beta \|\lambda^*\| \sum_{i=1}^n a_i \|x_i^* - \bar{x}_i^*\|,$$

and by using the triangle inequality and the fact that $\sum_{i=1}^n a_i x_i^* = b$ we obtain:

$$\begin{aligned}\|\lambda^*\| &\leq \frac{\beta}{\|a\|^2} \left(\|b\| + \left\| \sum_{i=1}^n a_i \bar{x}_i^* \right\| \right) \\ &\leq \frac{\beta}{\|a\|^2} (\|b\| + c\|a\|_1),\end{aligned}$$

which corresponds to (13). From α -strongly convexity of the local functions we have:

$$\alpha\|x_i - \bar{x}_i^*\| \leq \|\nabla f(x_i)\|.$$

By using the reverse triangle inequality and the optimality condition we get

$$\|x_i^*\| \leq \frac{1}{\alpha} \|a_i \lambda^*\| + \|\bar{x}_i^*\| \leq \frac{a_i}{\alpha} \|\lambda^*\| + c. \quad (58)$$

Equation (14) then follows from combining (13) and (58).

C. Proof of Proposition 2

a) Proof that $\psi_{n,\kappa}$ is an upper bound: Observe that $x^{(1)}, x^{(2)} \in B(\mathbf{0}_{nd}, R_{b,\kappa})$ from Lemma 1, so that:

$$\|x^{(1)} - x^{(2)}\|^2 \leq 2 \left(\|x^{(1)}\|^2 + \|x^{(2)}\|^2 \right) \leq 4R_{b,\kappa}^2,$$

which yields that $\psi_{n,\kappa}$ is an upper bound.

b) Proof that $\chi_{n,\kappa}$ is an upper bound: We remind the reader that for $i = 1, \dots, n-1$, we have $\nabla f_i(x_i^{(q)}) = a_i \lambda^{(q)}$, with $q = 1, 2$. From α -strongly convexity of the local functions, it follows that for all $i = 1, \dots, n-1$:

$$a_i \langle \lambda^{(1)} - \lambda^{(2)}, x_i^{(1)} - x_i^{(2)} \rangle \geq \alpha \|x_i^{(1)} - x_i^{(2)}\|^2.$$

Let us define $y^{(q)} \in \mathbb{R}^{d(n-1)}$ the vector such that $y_i^{(q)} = x_i^{(q)}$ for $q = 1, 2$ and for $i = 1, \dots, n-1$. Using the fact that $\sum_{i=1}^n a_i x_i^{(q)} = b$ for $q = 1, 2$ and summing up the above inequalities over all $i = 1, \dots, n-1$ yields

$$a_n \langle \lambda^{(1)} - \lambda^{(2)}, x_n^{(2)} - x_n^{(1)} \rangle \geq \alpha \|y^{(1)} - y^{(2)}\|^2,$$

where we used the fact that $\sum_{i=1}^{n-1} a_i x_i^{(q)} + a_n x_n^{(q)} = b$. By using Cauchy-Schwarz inequality and triangle inequality we obtain

$$\|y^{(1)} - y^{(2)}\|^2 \leq \frac{a_n}{\alpha} \left(\|\lambda^{(1)}\| + \|\lambda^{(2)}\| \right) \left(\|x_n^{(1)}\| + \|x_n^{(2)}\| \right)$$

Then, we use (13) and (14) to get

$$\|y^{(1)} - y^{(2)}\|^2 \leq 4 \left(\frac{a_n}{\|a\|^2} \kappa (\|b\| + c\|a\|_1) + c \right). \quad (59)$$

By definition we have

$$\|x^{(1)} - x^{(2)}\|^2 = \|y^{(1)} - y^{(2)}\|^2 + \|x_n^{(1)} - x_n^{(2)}\|^2. \quad (60)$$

We apply triangle inequality and (14) to obtain

$$\begin{aligned}\|x^{(1)} - x^{(2)}\|^2 &\leq \|y^{(1)} - y^{(2)}\|^2 + \left(\|x_n^{(1)}\| + \|x_n^{(2)}\| \right)^2 \\ &\leq \|y^{(1)} - y^{(2)}\|^2 + 2 \left(\|x_n^{(1)}\|^2 + \|x_n^{(2)}\|^2 \right) \\ &\leq \|y^{(1)} - y^{(2)}\|^2 + 4 \left(\frac{a_n}{\|a\|^2} \kappa (\|b\| + c\|a\|_1) + c \right)\end{aligned} \quad (61)$$

Finally, the result (18) yields by combining (59) and (61) and using the fact that $a_n \leq a_+$.

c) Proof that $\theta_{n,\kappa}$ is an upper bound: Since the functions are α -strongly convex and β -smooth, we have that for all $i = 1, \dots, n-1$:

$$\begin{aligned}a_i(1 + \kappa^{-1}) \langle \lambda^{(1)} - \lambda^{(2)}, x_i^{(1)} - x_i^{(2)} \rangle &\geq \\ \beta^{-1} a_i^2 \|\lambda^{(1)} - \lambda^{(2)}\|^2 + \alpha \|x_i^{(1)} - x_i^{(2)}\|^2.\end{aligned}$$

By summing up the above inequalities over all $i = 1, \dots, n-1$ yields

$$\begin{aligned}a_n(1 + \kappa^{-1}) \langle \lambda^{(1)} - \lambda^{(2)}, x_n^{(2)} - x_n^{(1)} \rangle &\geq \\ m\beta^{-1} \|\lambda^{(1)} - \lambda^{(2)}\|^2 + \alpha \|y^{(1)} - y^{(2)}\|^2,\end{aligned}$$

where $m = \sum_{i=1}^{n-1} a_i^2$. By using Cauchy-Schwarz inequality we obtain:

$$\begin{aligned}a_n(1 + \kappa^{-1}) \|\lambda^{(1)} - \lambda^{(2)}\| \|x_n^{(2)} - x_n^{(1)}\| &\geq \\ m\beta^{-1} \|\lambda^{(1)} - \lambda^{(2)}\|^2 + \alpha \|y^{(1)} - y^{(2)}\|^2.\end{aligned}$$

This can be written as follows

$$\alpha \|y^{(1)} - y^{(2)}\|^2 \leq \phi(\|\lambda^{(1)} - \lambda^{(2)}\|), \quad (62)$$

where

$$\phi(z) = -m\beta^{-1}z^2 + a_n(1 + \kappa^{-1})\|x_n^{(2)} - x_n^{(1)}\|z.$$

Since ϕ is a concave parabola, there exists $\phi^* = \max_z \phi(z) < \infty$ such that $\phi(z) \leq \phi^*$ for all z given by

$$\phi^* = \frac{a_n^2(1 + \kappa^{-1})^2 \|x_n^{(2)} - x_n^{(1)}\|^2}{4m\beta^{-1}},$$

and it follows by using (62) that

$$\begin{aligned}\|y^{(1)} - y^{(2)}\|^2 &\leq \frac{a_n^2(1 + \kappa^{-1})^2 \|x_n^{(1)} - x_n^{(2)}\|^2}{\alpha\beta^{-1} 4m} \\ &= \left(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} \right)^2 \frac{a_+^2 \|x_n^{(1)} - x_n^{(2)}\|^2}{4(\|a\|^2 - a_+^2)}.\end{aligned}$$

Equation (19) then follows from (60) using the fact that $x_n^{(1)}$ and $x_n^{(2)}$ are upper bounded by (14), which thus yields $\theta_{n,\kappa}$, and the conclusion follows.



Charles Monnoyer de Galland is a postdoctoral researcher at UCLouvain in the ICTeAM Institute.

He obtained an engineering degree in applied mathematics (2018), and the PhD degree in mathematical engineering (2022) as a FRIA fellow (F.R.S.-FNRS) from the same university. His research interests are centered around the analysis of open multi-agent systems and decentralized optimization. He was the recipient of the Networks and Communication Systems TC Outstanding Student Paper Prize of the IEEE Control Systems Society in 2022.



Renato Vizuite received the M.S. degree (très bien) in Systems, Control, and Information Technologies from Université Grenoble Alpes, France (2019), and the PhD degree in Automatic Control from Université Paris-Saclay, France (2022). He is currently a postdoctoral researcher at UCLouvain, Belgium, in the ICTEAM Institute, holding a postdoctoral scholarship from the same university. His research interests include multi-agent systems, distributed optimization, control theory and networked control systems. He was the recipient of the Networks and

Communication Systems TC Outstanding Student Paper Prize of the IEEE Control Systems Society in 2022, and the Second Thesis Prize in the category Impact Science of the Fondation CentraleSupélec in 2023.



Julien M. Hendrickx is professor of mathematical engineering at UCLouvain, in the Ecole Polytechnique de Louvain since 2010.

He obtained an engineering degree in applied mathematics (2004) and a PhD in mathematical engineering (2008) from the same university. He has been a visiting researcher at the University of Illinois at Urbana Champaign in 2003-2004, at the National ICT Australia in 2005 and 2006, and at the Massachusetts Institute of Technology in 2006 and 2008. He was a postdoctoral fellow at the Laboratory

for Information and Decision Systems of the Massachusetts Institute of Technology 2009 and 2010, holding postdoctoral fellowships of the F.R.S.-FNRS (Fund for Scientific Research) and of Belgian American Education Foundation. He was also resident scholar at the Center for Information and Systems Engineering (Boston University) in 2018-2019, holding a WBI World excellence fellowship.

Doctor Hendrickx is the recipient of the 2008 EECI award for the best PhD thesis in Europe in the field of Embedded and Networked Control, and of the Alcatel-Lucent-Bell 2009 award for a PhD thesis on original new concepts or application in the domain of information or communication technologies.



Elena Panteley received the M.Sc. and Ph.D. degrees in applied mathematics from the State University of St. Petersburg, St. Petersburg, Russia, in 1986 and 1997, respectively. From 1986 to 1998, she held a research position with the Institute for Problem of Mechanical Engineering, Russian Academy of Science, St. Petersburg. Since 2004 she holds a tenure position as Senior Researcher of the French National Centre of Scientific Research (CNRS), at the Laboratoire de signaux et systèmes, France. She is also associate researcher of ITMO University, St

Petersbourg Russia, since 2014. Her research interests include stability and control of nonlinear dynamical systems, network systems with applications to electromechanical and neuronal systems.



Paolo Frasca (M'13, SM'18) received the Ph.D. degree in Mathematics for Engineering Sciences from Politecnico di Torino, Torino, Italy, in 2009. From 2013 to 2016, he was an Assistant Professor at the University of Twente in Enschede, the Netherlands. Since October 2016 he is a CNRS Researcher affiliated with GIPSA-lab, Grenoble, France. His research interests are in the theory of networks and control systems, with main applications to transportation and social networks.