Contributions to open multi-agent systems : consensus, optimization and epidemics

Renato Sebastian Vizuete Haro

université PARIS-SACLAY

Contributions to open multi-agent systems: consensus, optimization and epidemics

Contributions aux systèmes multi-agents ouverts : consensus, optimisation et épidémies

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°580, Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Automatique
Graduate School : Sciences de l'ingénierie et des systèmes
Référent : CentraleSupélec

Thèse préparée dans le **Laboratoire des signaux et systèmes** (Université Paris-Saclay, CNRS, CentraleSupélec), sous la direction de **Elena PANTELEY**, Directrice de recherche, et le co-encadrement de **Paolo FRASCA**, Chargé de recherche.

Thèse soutenue à Paris-Saclay, le 23 septembre 2022, par

Renato Sebastian VIZUETE HARO

Composition du jury

Antoine GIRARD Directeur de Recherche, CNRS, L2S	Président & Examinateur
Giacomo COMO Professeur, Politecnico di Torino	Rapporteur & Examinateur
Giuseppe NOTARSTEFANO Professeur, Università di Bologna	Rapporteur & Examinateur
Franck IUTZELER Maître de conférences, Université Grenoble-Alpes	Rapporteur & Examinateur
Federica GARIN Chargée de recherche, INRIA, GIPSA-lab	Examinatrice
Elena PANTELEY Directrice de Recherche, CNRS, L2S	Directrice de thèse

THESE DE DOCTORAT

NNT : 2022UPAST114

Abstract

In this thesis we address several problems formulated in an open multi-agent system (OMAS) scenario where the set of agents can change in time, independently of the evolution of the dynamics associated with the system. We analyze the case of OMAS formulated in a fixed finite size network and we use two approaches for the analysis of the systems. In the first approach we consider scenarios characterized by activation/deactivation of agents such that at each time instant a different subset of active agents can interact in the system. In the second approach we study scenarios characterized by replacements of agents where at a specific time instant, an agent can be replaced while the rest of the agents remain the same. In this case, all the agents are able to interact at all time.

Three different problems are considered in this thesis: randomized consensus, resource allocation problem and epidemics. First, we analyze the problem of randomized consensus subject to additive noise where different subset of agents exchange information at each iteration. We define a noise index based on the expected mean squared error and we derive upper bounds. Then, we consider the resource allocation problem where agents can be replaced during the implementation of an optimization algorithm. For this optimization problem, we analyze two different algorithms: weighted gradient descent and random coordinate descent. For the weighted gradient descent, we evaluate the performance of the algorithm in an OMAS subject to packet losses by defining appropriate performance metrics. For the random coordinate descent algorithm, we study the convergence to the minimizer in an OMAS and we provide an alternative analysis using tools inspired from online optimization. Finally, we study a SIS epidemic in continuous time subject to replacements of agents during the evolution of the disease. We perform the analysis using an aggregate function to condensate the information of the system and we derive upper bounds for its asymptotic behavior.

Aperçu de la thèse

Systèmes multi-agents ouverts (OMAS)

Les systèmes multi-agents sont composés d'entités individuelles interagissant entre elles dans un environnement commun pour accomplir une tâche collective ou atteindre un objectif commun. Les interactions sont généralement définies à l'aide d'un graphe composé d'un ensemble de nœuds représentant les agents et d'un ensemble d'arêtes modélisant les échanges d'information entre les agents. En général, les chercheurs considèrent un ensemble fixe d'agents pour la dérivation des résultats, tandis que l'ensemble d'arêtes peut varier dans le temps. Cependant, dans de nombreux scénarios, l'ensemble d'agents peut également changer puisque des agents peuvent rejoindre ou quitter le système ou simplement être remplacés. Un système multi-agents soumis à des ajouts, des suppressions ou des remplacements d'agents est appelé un système multi-agents ouvert (OMAS). Si ces changements potentiels dans l'ensemble des agents ne sont pas fréquents, l'étude des systèmes multi-agents peut être effectuée sous l'hypothèse que les agents restent les mêmes dans le temps, même si les connexions peuvent varier. Cependant, si des ajouts, des suppressions ou des remplacements se produisent à une échelle de temps similaire à celle du système dynamique sur le réseau, il est nécessaire de considérer des OMAS pour l'analyser. Parmi les applications où l'usage des OMAS peut être envisagée, on trouve par exemple les réseaux sociaux, les épidémies, les convois de véhicules, les *smart grids*, la robotique en essaim, *etc.*

Les systèmes multi-agents soumis à des remplacements sont un cas particulier d'OMAS, où le départ d'un agent est immédiatement suivi de l'arrivée d'un nouvel agent de sorte que la taille du système ne change pas. Lorsque l'intervalle de temps entre un départ et une arrivée est négligeable, l'occurrence de ces deux événements peut être considérée comme simultanée, et l'analyse peut être effectuée en ne considérant que les remplacements. Ce comportement particulier peut également être vu comme une approximation de l'OMAS où les taux d'arrivées et de départs sont similaires.

Principaux défis dans les OMAS

Les changements possibles dans la cardinalité de l'ensemble des agents augmentent considérablement la complexité d'OMAS et posent de nombreux défis pour l'analyse de ces systèmes.

Espaces d'états et trajectoires : lorsque l'ensemble des agents change de cardinalité, les agents peuvent appartenir à différents espaces d'états et la définition des trajectoires ne peut pas être facilement formulée.

Modèles d'événements : au cours de l'évolution du système, les OMAS se caractérisent par la survenance de trois types d'événements : les arrivées, les départs et les remplacements. Une identification parfaite des instants temporels auxquels ces événements se produisent ne semble pas plausible et des hypothèses appropriées doivent être considérées pour modéliser des OMAS et garder l'analyse traitable.

Topologies de graphes : il existe quelques cas où les topologies de graphes suivent un modèle spécifique qui peut être complètement étendu pour n'importe quel nombre d'agents. Cependant, il est peu probable que les réseaux réels suivent des topologies de graphes restrictives et conservent également les mêmes connexions à tout moment.

Formulation en temps : la plupart des travaux antérieurs sur les OMAS considèrent une formulation en temps discret, où les agents n'interagissent qu'à des instants de temps précis. Cependant, de nombreux phénomènes complexes, notamment associés aux systèmes chaotiques, ne peuvent pas être reproduits par les systèmes en temps discret et des formulations en temps continu sont nécessaires.

Approches potentielles de l'OMAS

Après avoir identifié les enjeux des OMAS, nous passons maintenant en revue les principales approches qui pourraient être utilisées dans leur étude.

Sur-ensemble fini invariant dans le temps : des OMAS peuvent être sous-approximés en considérant un sur-ensemble fini d'agents où l'ouverture du système est caractérisée par la activation et la désactivation des agents, de sorte qu'à chaque instant un nombre différent d'agents peut interagir dans le système. Cette approche sera considerée dans le Chapitre 2.

Espace d'état fixe : lorsqu'un OMAS est soumis uniquement à des remplacements, la dimension du système ne change pas avec le temps, et lorsqu'un agent quitte le réseau, un nouvel agent rejoint immédiatement le système en gardant le même label. Cette approche sera poursuivie dans les Chapitres 3, 4 et 5.

Systèmes multidimensionnels multimodes : étant donné que la taille du système peut changer en raison de l'occurrence d'événements liés aux OMAS, des systèmes de commutation entre des espaces de différentes dimensions peuvent être utilisés pour l'analyse.

Systèmes dynamiques de dimension infinie : le système dynamique de dimension infinie peut être utilisé comme sur-approximation de OMAS. L'objectif est de résoudre le problème en dimension infinie puis d'utiliser des techniques d'approximation pour obtenir la solution au cas d'OMAS avec une certaine marge d'approximation.

Principales contributions de cette thèse

L'intérêt principal de cette thèse est l'analyse d'OMAS en utilisant deux approches différentes : le sur-ensemble fini invariant dans le temps et l'espace d'état fixe. Nous considérons trois problèmes différents : l'influence du bruit dans le consensus randomisé, le problème d'allocation des ressources et la propagation d'une maladie. Nous formulons ces problèmes en utilisant une approche stochastique et nous donnons des bornes pour des quantités scalaires qui sont définies dans le but d'évaluer les impacts des activations/désactivations ou des remplacements d'agents dans le système.

Dans le Chapitre 2 nous analysons le problème du consensus randomisé perturbé par un bruit additif. Nous considérons la formulation en temps discret d'un algorithme de consensus randomisé linéaire où les matrices d'interaction sont symétriques et tirées d'une distribution commune. Ensuite, nous analysons le cas où le bruit additif modifie les états des agents et nous dérivons une expression sous forme fermée de l'erreur quadratique moyenne induite par le bruit, ainsi que des bornes supérieure et inférieure plus simples à évaluer. Ensuite, nous considérons le cas d'OMAS où les matrices à chaque instant correspondent au sous-ensemble d'agents actifs dans le système. L'activation des agents est déterminée par des variables aléatoires de Bernoulli indépendantes associées à chaque agent potentiel du système. Pour les OMAS, nous exprimons les bornes en utilisant les valeurs propres de la matrice laplacienne du graphe sous-jacent ou la résistance effective moyenne du graphe, prouvant ainsi leur qualité. Enfin, nous dérivons des expressions pour les bornes sur quelques exemples de graphes et les évaluons numériquement.

Le Chapitre 3 se concentre sur l'analyse de l'algorithme *Weighted Gradient Descent* pour résoudre le problème d'allocation de ressources (RA) sous pertes de paquets. Nous considérons le cas particulier des fonctions de coût quadratiques par morceaux et définissons deux métriques de performance qui mesurent, respectivement, l'écart à la contrainte et l'erreur sur la fonction de coût attendue. Nous dérivons des bornes supérieures sur les deux métriques qui sont proportionnelles à la différence entre la fonction de coût initiale et la fonction de coût évaluée au minimiseur. Ensuite, nous étendons l'analyse de la violation de contrainte aux OMAS où les agents peuvent être remplacés, ce qui implique un changement sur la fonction de coût associée. Nous montrons que la combinaison des remplacements et des pertes fait diverger l'erreur de violation de contrainte avec le temps.

Le Chapitre 4 est consacré à l'analyse de l'algorithme *Random Coordinate Descent* (RCD) pour résoudre le problème RA dans des OMAS sous l'hypothèse que les fonctions locales sont lisses fortement convexes et ont leurs minimiseurs situés dans une boule donnée. Nous établissons la convergence linéaire de l'algorithme dans les systèmes fermés, en termes d'estimation vers le minimiseur, pour des graphes généraux et des pas appropriés et nous estimons le changement de la solution optimale après un remplacement pour évaluer son effet sur la distance entre les estimation et le minimiseur. A partir de ces deux éléments, nous dérivons des conditions de stabilité dans des OMAS où les agents peuvent être remplacés et établissons la convergence linéaire de l'algorithme vers une erreur en espérance en régime permanent. De plus, en considérant des paramètres plus simples, nous analysons l'algorithme RCD dans des OMAS à l'aide d'outils similaires à ceux couramment utilisés dans l'optimisation en ligne. En particulier, nous comparons la solution issue de l'algorithme RCD avec la solution optimale et la stratégie égoïste. Ce chapitre a été développé en collaboration avec Charles Monnoyer de Galland et Julien Hendrickx de l'Université Catholique de Louvain (UCLouvain).

Dans le Chapitre 5 nous fournissons une analyse préliminaire de la propagation d'une maladie en temps continu dans des OMAS soumis à des remplacements. Nous considérons le modèle épidémique SIS sur un graphe et définissons une fonction agrégée pour l'analyse du système. Premièrement, nous considérons le cas des remplacements sur un graphe donné fixe où l'occurrence des remplacements est déterminée par un processus de Poisson. Ensuite, nous étudions le cas de graphes échantillonnés à partir de graphons où nous dérivons des bornes supérieures qui dépendent des propriétés du graphon. Enfin, nous fournissons une analyse préliminaire de scénarios où les connexions changent également lors des remplacements selon un graphon, et nous formulons une conjecture pour le comportement asymptotique de la fonction d'agrégation.

Acknowledgments

First, I would like to thank God for giving me the strength to keep going.

I would like to express my sincere gratitude to my supervisors Elena Panteley and Paolo Frasca for guiding me along these three years of research and for their genuine dedication at all levels. I learned a lot from our interactions and thanks to you I always found the motivation to continue with the research despite the difficulties. Thank you for your advice, help and support.

I would like to express my appreciation to Profs. Giacomo Como, Giuseppe Notarstefano and Franck Iutzeler for your reviews that helped me improving this manuscript. Thanks also to Antoine Girard and Federica Garin for accepting being members of the jury for my defense.

My thanks also go to Prof. Julien Hendrickx for giving me the opportunity of visiting his research lab in the Université Catholique de Louvain. It was a wonderful experience to spend those weeks with him and his fantastic research team.

I would like to thank to my colleague and friend Charles Monnoyer de Galland, with whom I had the opportunity of working together for a long period of time. I really enjoyed our exchanges on math, research and life, together with Iona. Thank you for your guidance during my stay in Louvain-la-Neuve.

I thank all my fellows labmates and friends in GIPSA-lab, whom I had the chance to meet during the years I stayed in the lab. I really enjoyed our discussions and jokes during the lunch and the different activities and games that we had together. Thank you for making the lab a pleasant environment to work in, especially during the epidemic. I wish I had been with you until the end of my doctorate.

Let me also express my thanks to my colleagues and friends in L2S, with whom I spent the last months of my doctorate having pleasant conversations during the lunch.

Last but not least, I would like to thank to my family. My parents, Roberto and Dolores, who always gave me their love and support to overcome all the difficulties in my life. I also would like to thank my cousin Pablo and his wife Margarita, my family in Europe, who always welcomed me warmly during my visits to Madrid.

Contents

Appendix

List of Figures

List of symbols and acronyms

$\mathbb{R}$	Set of real numbers
$\mathbb{R}^n$	Set of n-dimensional real vectors
$\mathbb{R}^n_{\geq 0}$	Set of n-dimensional real vectors with nonnegative entries
$\mathbb{R}^{m \times n}$	Set of $m \times n$-dimensional real matrices
$\mathbb{Z}_{\geq 0}$	Set of nonnegative integers
$\mathbb{Z}_{>0}$	Set of positive integers
I_n	$n \times n$ identity matrix
$\mathbf{0}_n$	n-dimensional vector of all zeros
$\mathbb{1}_n$	n-dimensional vector of all ones
$\otimes$	Kronecker product
$\odot$	Hadamard product
$(\cdot)^T$	Transpose of a matrix or a vector
A^{-1}	Inverse matrix of a matrix A
$A^\dagger$	Moore-Penrose inverse matrix of a matrix A
$\langle \cdot, \cdot \rangle$	Inner product of two vectors
$O(\cdot)$	Order of magnitude
$\lambda_i(L)$	i-th eigenvalue of a symmetric Laplacian matrix L given in non-decreasing order of magnitude $0 = \lambda_1(L) \leq \ldots \leq \lambda_n(L)$
$\lambda_i(A)$	i-th eigenvalue of a symmetric matrix A given in non-increasing order of magnitude $\lambda_1(A) \geq \ldots \geq \lambda_n(A)$
$\|\|\cdot\|\|$	2-norm of a vector or matrix in Euclidean space
$\|\|\cdot\|\|_1$	1-norm of a vector or matrix in Euclidean space
$\|\|\cdot\|\|_A$	Norm induced by a positive definite matrix A
$\|\|\cdot\|\|_F$	Frobenius norm of a matrix
$\|\|\cdot\|\|_{S_p}$	Schatten p-norm of a matrix or operator (p.19)
$\|\|\cdot\|\|_\square$	Cut norm of a matrix or measurable function $f : R^2 \to R$ (p.127)
$\|\|\cdot\|\|_{L^p}$	L^p norm of a measurable function (p.127)
$\|\|\|\cdot\|\|\|$	Operator norm
$B(x, r)$	Ball of radius $r \geq 0$ centered at x
$\mathbb{E}[X]$	Expectation of a random variable X
$\mathrm{Var}(X)$	Variance of a random variable X
∇f	Gradient of a function f

Acronyms

i.i.d.	independent and identically distributed
ODE	ordinary differential equation
OMAS	open multi-agent system
PDMP	piecewise deterministic Markov process (p.104)
RA	resource allocation (p.33)
RCD	random coordinate descent (p.54)
RIDL	randomly induced discretized Laplacian (p.21)
SDE	stochastic differential equation (p.102)

Chapter 1

Introduction

1.1 Classical multi-agent systems

In the last years, an increasing number of real-world problems are being understood as constantly evolving towards a networked environment where single entities interact with each other to accomplish a collective task or achieve a common objective. Systems of this type are called *multi-agent systems* and have been extensively studied in the scientific community considering different types of dynamics, coupling, goals and time formulation [1].

Regarding the dynamics, each agent is in itself a dynamical system that can be modeled by ordinary differential equations (ODE). At the most basic level, each agent i is commonly considered as a single integrator of the form:

$$\dot{x}_i(t) = u_i(t),$$

where $u_i(t)$ is the input of the agent. However, this type of models does not reflect the behavior of more complex systems. For this reason, researchers have considered the use of more complex dynamics to model each single agent. In this line, models such as double integrators, Lagrangian systems, etc. [2, 3], have been studied with the objective of extending well-known results to more complex dynamics.

With respect to the goals, researchers have considered several problems associated with multi-agent systems including: stability, consensus, synchronization and formation control. Stability corresponds to the analysis of the equilibrium points of the system. Consensus is a special case of coordination of multi-agent systems where the objective is to reach a common value for all the agents which must be static [4]. As a generalization of consensus, many researchers also consider the problem of synchronization where agents must reach a common value, but this value is time-varying. The synchronization problem is often considered in networks of oscillators, where the trajectories of agents are more chaotic [5]. In

the case of formation control, agents must reach specific locations determined by desired geometric patterns [6].

Furthermore, in the last years, the community has paid attention to the extension of optimization problems to the case of multi-agent system. In addition to its states, an agent has a local cost function depending only on the state of the agent, and the objective is the minimization of the sum of the cost functions of all the agents while agents exchange information only with the neighbors [1]. Additional constraints can be added depending on the nature of the problem. Generally, most of the algorithms make use of the well known *subgradients* of the cost fuctions to solve the optimization problem [7]. However, other methods like primal duals have been adapted to multi-agent systems [8].

While the links among the agents can be modified by nonlinear functions as in the case of oscillators [5], the connections in most of the multi-agent systems are characterized by diffusive couplings. For this particular case, the network can be represented by a matrix and tools of Linear Algebra can be used for the analysis of multi-agent systems. Furthermore, due to the link between specific type of matrices and graphs, *Spectral Graph Theory* [9] plays an important role in the study of interactions of agents over a network .

In many scenarios, the links in multi-agent systems do not remain static and change with time. For this reason, researchers focused on the analysis of classical problems applied to time-varying graphs, where under several assumptions on the connectivity of the possible graph topologies in a time interval, many results were easily extended [10–12]. Furthermore, in many cases, a complete knowledge of a sequence of graph topologies is not available and the attention was focused on randomized interactions, where the graph topologies are sampled from a common distribution and tools from Probability Theory can be used [13–16]. In this way, the complexity of multi-agent systems regarding the network topology were explored considering only possible changes in the current set of edges while the set of agents was assumed to be always static.

1.2 Open multi-agent systems (OMAS)

Even if models based on time-varying switching topologies provide a good approach for studying the behavior of systems where communications among the agents do not remain static, they do not encompass possible changes in the set of agents that could significantly affect the evolution of several systems. In the real world, agents can join and leave the network, and therefore the size of many multi-agent systems is changing with time. A multi-agent system subject to additions, removals

[1] A neighbor of an agent i in a multi-agent system is an agent that is directly connected to i.

or replacements of agents is called an *open multi-agent system* (OMAS) [17, 18]. If these potential changes in the set of agents are not frequent, the study of the multi-agent system can be performed under the assumption that agents remain the same in time, even if the connections may vary. However, if additions, removals or replacements occur in a time-scale similar to that of the dynamical system on the network, it is necessary to consider OMAS for the analysis of the dynamics. Among the potential applications where OMAS can be used we have:

Social networks: users enter and leave communities, discussions, forums, etc., modifying the group behavior.

Epidemics: infected individuals can enter or leave towns, villages, cities, etc., affecting the spread of diseases.

Vehicle platoons: during transportation, vehicles may join or leave the platoon according to their particular destinations.

Smart-grids: smart appliances are constantly switched on and off in a network of electrical devices, modifying the number of active devices.

Swarm robotics: the size of a group of robots can change depending on the environmental conditions and goals of the team.

Historically, computer science is one of the first research fields where the concept of OMAS has been considered in the analysis of trust and security, however, no dynamics are associated with the individual agents [19,20]. Also, in the field of game theory, researchers have analyzed the problem of price of anarchy bounds in dynamic populations where players can exit and be replaced by new participants [21]. Regarding dynamical systems, several works have indirectly considered potential cases of addition/removal of agents as an adverse scenario to analyze the robustness of algorithms [22, 23].

Preliminary works focused on consensus in OMAS consider the analysis of the pairwise gossiping algorithm in a complete graph [17, 24], where a discrete time formulation was used to analyze replacements and arrivals of agents determined by deterministic and random time instants. Furthermore, in [25], authors extended the results to the case of arrivals/departures when the time instants are determined by Poisson processes. In these works, the variance of the system was the object of study, which condensates the information of the system in a single value like an *aggregate function*. Following this line of research, in [26], the authors provided lower bounds on the mean squared error for any averaging algorithm applied to

an OMAS. Another work based on gossip interactions is [27], where the MAX-consensus in an OMAS over a complete graph is analyzed.

Regarding deterministic formulations of OMAS, one of the preliminary works is [28], where authors study the dynamic consensus in a contractive OMAS under the assumption that the arrival process is bounded. Other works related to the analysis of the dynamic consensus in OMAS are [29] and [30] where the median value and the max value are the object of the study respectively and the assumption of a dwell time is used for the derivation of the results. In [31], the authors analyze the stability of multi-dimensional switched systems, where restrictions are imposed on the switching signal to guarantee practical consensus.

Other works include the analysis of OMAS under stochastic interactions in a finite superset of agents where the activation of agents are determined by Bernoulli random variables [32]. Using a similar approach, 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 [33]. The study of the Friedkin-Johnsen models where agents exchange roles at different time instant may also be encompassed in OMAS since the set of agents can be seen as time-varying due to the changes in roles [34].

1.2.1 The case of replacements

Multi-agent systems subject to replacements is a special case of OMAS, where the departure of an agent is immediately followed by the arrival of a new agent such that the size of the system does not change. When the time interval between a departure and an arrival is negligible, the occurrence of these two events can be considered as simultaneous, and the analysis can be performed considering only replacements. This particular behavior can also be seen as an approximation of OMAS where the rate of arrivals and departures are similar.

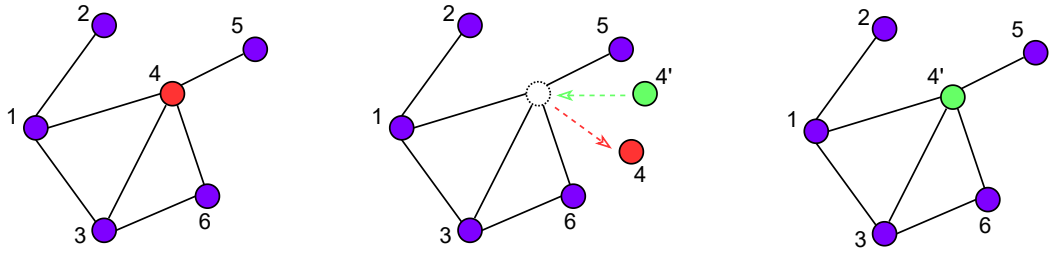


Figure 1.1: Replacement of an agent in OMAS, while preserving the topology.

The particular case of replacements has been studied in opinion dynamics, where researchers analyzed scenarios when the states of agents are reset [35]. In

the analysis of consensus in OMAS, replacements have been studied in [17, 24], where the authors considered the variance for the analysis of agents interacting in a complete graph. Regarding replacements in optimization, in [36], the authors analyzed the gradient descent algorithm in a network subject to replacements of the cost functions of the agents.

1.2.2 Main challenges in OMAS

Possible changes in the cardinality of the set of agents increase considerably the complexity of OMAS and pose many challenges for the analysis of these systems.

State spaces and trajectories: the concept of state spaces and trajectories are fundamental in the study of classical multi-agent systems where it is assumed that the agents always remain in the same state-space and they have an associated trajectory for all time [1]. However, when the set of agents is changing in cardinality, agents can belong to different state-spaces and the definition of trajectories cannot be easily formulated since the states of the agents may not be well-defined for some time intervals. For instance, during the removal of an agent, its associated trajectory is completely removed while when an agent joins the system, a new trajectory is created. Even if only replacements occur in the system, trajectories experiment jumps, which can drastically change the behavior of the current state.

Models of events: during the evolution of the system, OMAS are characterized by the occurrence of three types of events: arrivals, departures and replacements [17, 28]. A perfect identification of the time instants at which these events take place do not seem plausible in a real application, and appropriate assumptions must be considered to model OMAS and keep the analysis tractable. In this line, while the use of dwell times can guarantee some lower bounds on the time interval between the occurrence of two distinct events, the use of stochastic processes could model richer behaviors of OMAS that may be closer to real applications.

Network topologies: there are few specific cases where the graph topologies follow a specific pattern that can be completely extended for any number of agents such as: line, cycle, grids, complete or bipartite graphs [37]. However, it is unlikely that real networks will follow graph topologies that are restrictive, and also keep the same connections at all time. For this reason, new graph topologies must be explored, which could include random graphs [38] and graphons [39].

Time formulation: unlike few works in continuous time such as [29, 31], most of the previous works in OMAS consider a discrete time formulation, where agents

only interact at discrete time instants. Even if a discrete time formulation can in many cases reproduce the behaviors of classical dynamical systems formulated as ordinary differential equations (ODE), many complex phenomena, specially associated with chaotic systems, cannot be modeled with discrete time systems [40]. For this reason, a continuous time formulation of OMAS is necessary to study other types of dynamics, which implies the study of more complex processes, where the state of the system also changes between arrivals, departures and replacements.

1.2.3 Potential approaches to OMAS: pros and cons

After having identified the challenges in OMAS, we now proceed to review the main approaches that could be used in their study.

Time-invariant finite superset: OMAS can be under-approximated by considering a finite superset where it is assumed that we have the information about all the potential agents [32, 33]. The openness of the system is characterized by the activation and deactivation of agents, such that at each time instant different number of agents may interact in the system. Since the dimension of the global system is constant, standard tools from dynamical systems can be used for the analysis. However, this approach is restricted to systems where a full knowledge of all the possible agents is available such that unexpected agents cannot join the network. This approach will be pursued in Chapter 2.

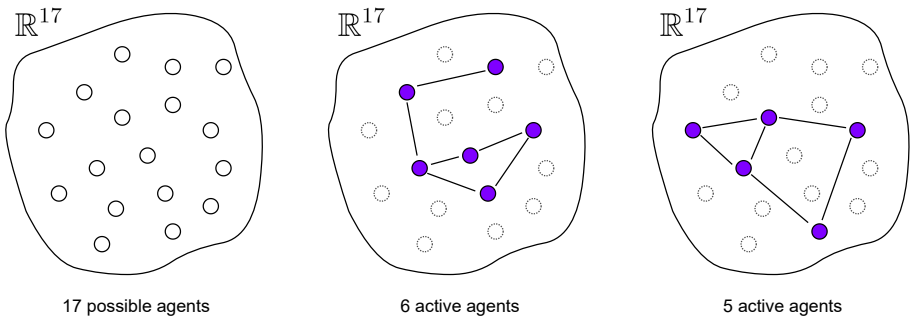


Figure 1.2: Time-invariant finite superset of agents characterized by different active agents.

Fixed state-space: when an OMAS is subject only to replacements, the dimension of the system does not change with time, and the state of the system always remains in the same state-space. Unlike the time-invariant finite superset, all the agents interact at the same time instants, but when an agent leaves the network,

a new agent will join the system immediately keeping the same label. Trajectories are well-defined but they may experience jumps during the replacements. Since the state-space does not change, standard concepts of dynamical systems can be applied. However, this approach is restricted to replacements and cannot be used in the case of decoupled arrivals and departures. This approach will be pursued in Chapters 3, 4 and 5.

Multi-mode multi-dimensional systems: since the size of the system may change due to the occurrence of arrivals or departures, switching systems between spaces of different dimensions may be used for the analysis. Preliminary works in this field correspond to [41, 42], where the author studied the switching between a fixed number of spaces of different dimensions. The family of possible vector spaces to which the state belongs is defined as a discrete vector bundle and concepts of Topology are required for the analysis. For this framework, the concept of trajectories is meaningful and a global equilibrium for the system can be defined, allowing the use of some tools of hybrid systems [43]. However, the classical concept of stability cannot be applied and the approach can be rigid for many applications since it is required to know all the possible vector spaces to define the discrete vector bundle.

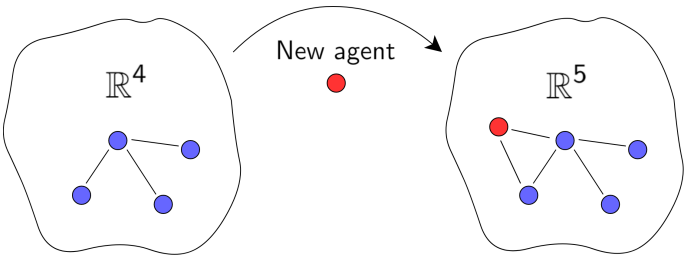


Figure 1.3: Multi-mode multi-dimensional systems as a tool to study OMAS.

Infinite dimensional dynamical systems: infinite dimensional dynamical system may be used as an over-approximation of OMAS. In this case, the problem can be formulated as a dynamical system in a space of sequences when we consider a infinite network [44] or in a space of mesurable functions by using a graphon [45]. The objective is to solve the problem in infinite dimensions and then use approximation techniques to obtain the solution to the case of OMAS with a certain margin of approximation. The main advantages of this approach is that all the possible agents can be considered in the analysis of the system and advanced tools from infinite dimensional systems can be used. However, the definition of appropriate mappings from the finite network to the infinite system and vice versa could make it difficult to use this approach.

1.3 Main contributions of this thesis

The main interest of this work is the analysis of OMAS using two different approaches: time-invariant finite superset and fixed state-space. We consider three different problems: influence of noise in randomized consensus, resource allocation problem and the spread of a disease. We formulate the problems using a stochastic approach and we provide bounds for scalar quantities that are defined with the objective of evaluating the impacts of activations/deactivations or replacements of agents in the system.

In Chapter 2 we analyze the problem of randomized consensus perturbed by additive noise. We consider the formulation in discrete time of a linear randomized consensus algorithm where the interaction matrices are symmetric and drawn from a common distribution. Then, we analyze the case where additive noise modifies the states of the agents and we derive a closed form expression for the mean squared error induced by the noise, together with upper and lower bounds that are simpler to evaluate. Next, we consider the case of OMAS where the interaction matrices at each time instant are generated by the subset of active agents in the system. The activation of agents are determined by independent Bernoulli random variables associated to each potential agent in the system. For OMAS, we express the bounds by using the eigenvalues of the Laplacian matrix of the underlying graph and the graph's average effective resistance, thereby proving their tightness. Finally, we derive expressions for the bounds for particular graph topologies and numerically evaluate them.

Chapter 3 focuses on the analysis of the weighted gradient descent algorithm to solve the resource allocation (RA) problem under packet losses. We consider the particular case of piecewise quadratic cost functions and define two performance metrics that measure, respectively, the deviation from the constraint and the error on the expected cost function. We derive upper bounds on both metrics which are proportional to the difference between the initial cost function and the cost function evaluated at the minimizer. Then, we extend the analysis of the constraint violation to OMAS where agents can be replaced, which implies a change on the associated cost function. We show that the combination of replacements and losses makes the constraint violation error diverge with time.

Chapter 4 is devoted to the analysis of the Random Coordinate Descent (RCD) algorithm to solve the RA problem in OMAS under the assumptions that local functions are smooth strongly convex and have their minimizers located in a given ball. 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 and we estimate the change of the optimal solution after a replacement to evaluate its effect on the distance between the estimate and the minimizer. From these two elements, we derive stability conditions in OMAS where agents can be

replaced and establish the linear convergence of the algorithm towards a steady state expected error. Additionally, by considering a simple setting, we analyze the RCD algorithm in OMAS using tools similar to those commonly used in online optimization. In particular, we study the accumulated errors that compare solutions issued from the RCD algorithm and the optimal solution or the non-collaborating selfish strategy and we derive some bounds in expectation for these accumulated errors. This chapter was developed in collaboration with Charles Monnoyer de Galland and Julien Hendrickx of the Université Catholique de Louvain (UCLouvain).

In Chapter 5 we provide a preliminary analysis of the spread of a disease in continuous time in OMAS subject to replacements. We study the SIS epidemic model over a network and inspired by the seminal work in [17], we define an aggregate function for the analysis of the system based on a macroscopic description. First, we consider the case of replacements over a fixed given graph where the occurrence of replacements follows a homogeneous Poisson process and the new values of the replaced agents are determined by a random variable. We analyze the expectation and variance of the aggregate function and we provide some upper bounds. Then, we study the case of graphs sampled from graphons where we derive upper bounds which depend on properties of the graphon. Finally, we provide a preliminary analysis of scenarios where connections also change during the replacements according to a graphon, and we formulate a conjecture for the asymptotic behavior of the aggregate function.

Thesis outline. Chapter 2 analyzes the influence of noise in randomized consensus with an application to OMAS considering a finite superset of agents. Chapters 3 and 4 focus on the resource allocation problem in OMAS where agents can be replaced at each time instant. While Chapter 3 analyzes the weighted gradient descent algorithm under packet losses, Chapter 4 studies the stability of the RCD algorithm and considers the analysis in a particular setting using tools inspired by online optimization. Chapter 5 provides a preliminary analysis of a SIS epidemic in continuous time in an OMAS subject to replacements determined by a Poisson process. In Chapter 6 we summarize the results of the thesis and propose directions for future research. Appendix A introduces some mathematical concepts and results related to graphs, graphons and Poisson processes.

1.4 List of publications

The following is an exhaustive list of publications written during these three years of thesis, that are either published, under review or still in preparation. It also contains the publications related to graphons, whose objective is to develop tools

for the analysis of OMAS using an infinite dimensional approach.

Journal papers

J1. R. Vizuete, P. Frasca and E. Panteley, "On the Influence of Noise in Randomized Consensus Algorithms," in IEEE Control Systems Letters, vol. 5, no. 3, pp. 1025-1030, 2021. DOI: http://doi.org/10.1109/LCSYS.2020.3009035. Corresponds to Chapter 2 of this thesis.

J2. R. Vizuete, C. M. de Galland, J. M. Hendrickx, E. Panteley and P. Frasca, "Random coordinate descent for resource allocation in open multi-agent systems," submitted to IEEE Transactions on Automatic Control. Available at: https://arxiv.org/abs/2205.10259 [2]. Corresponds to Chapter 4 of this thesis.

Conference papers

C1. C. M. 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," 2021 60th IEEE Conference on Decision and Control (CDC), 2021, pp. 1701-1708. DOI: http://doi.org/10.1109/CDC45484.2021.9683049. The contents of this conference paper correspond to a special case covered by the journal paper J2 and contribute to Chapter 4 of this thesis.

C2. R. Vizuete, P. Frasca and E. Panteley, "Gradient descent for resource allocation with packet loss," NecSys 2022 - 9th IFAC Conference on Networked Systems, 2022. Available at: http://doi.org/10.1016/j.ifacol.2022.07.244. Corresponds to Chapter 3 of this thesis.

C3. R. Vizuete, C. M. de Galland, J. M. Hendrickx, P. Frasca and E. Panteley, "Resource allocation in open multi-agent systems: an online optimization approach," accepted for the 2022 61th IEEE Conference on Decision and Control (CDC). Available at: https://arxiv.org/abs/2207.09316. Corresponds to the Section 4.5 of this thesis.

In preparation

P1. R. Vizuete, P. Frasca and E. Panteley, "Analysis of a SIS epidemic in open multi-agent systems". Corresponds to Chapter 5 of this thesis.

[2] R. Vizuete and C. Monnoyer de Galland equally contributed to this paper and can be considered as the first author.

Journal papers related to graphons

During these three years of thesis, two papers related to graphons were published, which partially originated in the internship corresponding to the Master Degree in Systems, Control and Information Technologies at the Université Grenoble Alpes. Their objective is to study the properties of graphons that can be used for the analysis of multi-agent systems where the network topology is sampled from a graphon. Although some results are used in this thesis, specially in Section 2.3, Chapter 5 and Appendix A, their contents are not specifically focused on the analysis of OMAS and therefore are not included into this thesis.

J3. R. Vizuete, P. Frasca and F. Garin, "Graphon-Based Sensitivity Analysis of SIS Epidemics," in IEEE Control Systems Letters, vol. 4, no. 3, pp. 542-547, 2020. DOI: `http://doi.org/10.1109/LCSYS.2020.2971021`

J4. R. Vizuete, F. Garin and P. Frasca, "The Laplacian Spectrum of Large Graphs Sampled From Graphons," in IEEE Transactions on Network Science and Engineering, vol. 8, no. 2, pp. 1711-1721, 2021. DOI: `http://doi.org/10.1109/TNSE.2021.3069675`

Chapter 2

Randomized consensus with activation/deactivation of agents subject to noise disturbance

When the set of all the possible agents in the systems is available, OMAS can be analyzed using a superset. The openness of the system is then characterized by a continuous activation and deactivation of agents [32], with a network topology determined by the induced subgraph. Moreover, if the time-scale is correctly chosen such that the dependence of the activation/deactivation of agents between iterations is negligible, the interaction matrices at different time instants can be considered as independent events and the behavior of the system can be analyzed using tools associated with randomized algorithms. This corresponds for example to the case of smart devices connected to a large electrical network in a town or city, where plugging/unplugging a device can be considered as an action independent of the state of other devices on the network.

Randomized interactions have been extensively studied in consensus systems [46, 47], with a wide range of applications including social networks [48], sensor networks [49], and clock synchronization [50]. Even if consensus in random networks has been analyzed for a long time, researchers have mostly considered ideal, noiseless interactions [14]. Instead, more realistic dynamical models should at least include noise. The influence of noise has indeed been investigated in several works [51–53] regarding deterministic consensus, both in discrete-time [54] and in continuous-time [55, 56] but, relatively little is known about the effects of noise on randomized consensus. For instance, the authors in [57] studied additive noise in random consensus over directed graphs, showing that, due to asymmetric loss of links, uncommon phenomena such as Lévy flights can arise in the evolution of the system. However, the emergence of this phenomenon is not possible for undirected graphs, provided the links are deactivated in a symmetric way.

In the case of OMAS, in addition to the variation of the number of active agents interacting at each time instant, the effect of noise adds an additional perturbation to the behavior of the system. It is clear that noise prevents the states of the nodes from reaching consensus. For this reason, it is important to quantify the effects of noise by a noise index, which is simply the steady-state normalized mean squared error between the states of the nodes and their average. It is well known that for symmetric matrices, this type of indexes are closely related to the spectrum of the interaction matrices, such that closed forms and bounds are generally expressed as a function of the eigenvalues [54].

In this chapter we study the steady-state normalized mean squared error in a randomized consensus perturbed by additive noise considering agents with scalar states. However, the majority of the results can be extended to agents with higher dimensional states. In Section 2.1 we define a noise index and obtain an explicit expression for it. Since this closed form involves the calculation of an n^2-dimensional matrix, we derive an upper bound and a lower bound that depend on the eigenvalues of relevant n-dimensional matrices. These results generalize well-known results about deterministic consensus with noise [54, 56]. Section 2.2 presents the analysis of randomized consensus with noise in OMAS characterized by activations and deactivations of agents. We refine our analysis for a specific class of random update matrices related to OMAS, which we refer to as Randomly Induced Discretized Laplacians (RIDL). Given an underlying large graph, these matrices are generated by sampling a subset of active nodes and considering the subgraph induced by the active nodes. When the update matrices are RIDLs, our bounds can be expressed as functions of the Laplacian eigenvalues of the underlying graph. Further rewriting of the bounds as functions of the graph's average effective resistance reveals that they are asymptotically tight. In Section 2.3 we illustrate the behavior of the upper and lower bounds for several sparse and dense graphs. Finally, Section 2.4 presents the conclusions and future work of this chapter.

2.1 Randomized consensus with additive noise

We consider a system composed by n agents holding scalar values $x_i \in \mathbb{R}$, $i = 1, \dots, n$, that interact through a randomized averaging algorithm in discrete time given by:

$$x(k+1) = P(k)x(k), \quad \text{for all } k \in \mathbb{Z}_{\geq 0} \tag{2.1}$$

where $x(k) = [x_1(k), \dots, x_n(k)]^T$ is a vector with the states of the nodes of the network and $P(k)$ is a stochastic matrix for all k. One of the most important results on dynamics (2.1) is the determination of sufficient conditions to achieve probabilistic consensus with independent and identically distributed (i.i.d.) matrices $P(k)$.

Lemma 2.1 (Corollary 3.2 [14]) *The dynamics* (2.1) *with i.i.d. matrices* $P(k)$ *reaches consensus with probability 1 if:*

$$P_{ii}(k) > 0 \text{ with probability 1 for all } i \in V, \tag{2.2a}$$

$$G_{\bar{P}} \text{ has a globally reachable node}, \tag{2.2b}$$

where $V = \{1, \dots, n\}$ *and* $G_{\bar{P}}$ *is the graph associated*[1] *with the expected matrix* $\bar{P} = \mathbb{E}[P(k)]$.

Condition (2.2a) is easy to satisfy in most cases since each agent usually has access to its own state. For undirected graphs, condition (2.2b) is equivalent to having a connected network. Moreover, for undirected graphs, the matrices $P(k)$ of dynamics (2.1) are usually symmetric and, therefore, doubly stochastic: if matrices $P(k)$ are doubly stochastic, then dynamics (2.1) reaches average consensus with probability 1.

Even if under ideal conditions it is possible to reach average consensus for undirected graphs, noise can affect the performance of the system. In order to study its effects, additive noise can be included in the dynamics [37,54], by defining:

$$x(k+1) = P(k)x(k) + w(k), \tag{2.3}$$

where $w(k)$ is a vector of noise. It is natural to define a performance index to measure the deviation from the consensus point due to the perturbations as:

$$J_{\text{noise}} := \frac{1}{n} \lim_{k \to +\infty} \mathbb{E}\left[\left\| x(k) - \frac{1}{n} \mathbb{1}\mathbb{1}^T x(k) \right\|^2 \right], \tag{2.4}$$

where $\mathbb{1}$ is a vector of ones. The purpose is studying this index, under the following standing assumption.

Assumption 2.1 *The stochastic matrices* $P(k)$ *are i.i.d., symmetric, and satisfy* (2.2a)-(2.2b) *for all* k.

Our first contribution in this direction is showing that, under suitable conditions on the noise vector $w(t)$, index J_{noise} can be expressed in closed form.

Theorem 2.1 (Mean squared error index) *Consider system* (2.3) *under Assumption 2.1 with a noise vector* $w(k)$ *uncorrelated w.r.t. both* i *and* k, *with zero mean and variance* σ^2. *Then, the noise index* (2.4) *can be expressed as:*

$$J_{\text{noise}} = \frac{\sigma^2}{n} \left(\text{vec}^T(I_n)(I_{n^2} - K)^{-1} \text{vec}(I_n) - 1 \right), \tag{2.5}$$

where $K = \mathbb{E}[P(k)\Omega \otimes P(k)]$, $\Omega = I_n - \frac{1}{n}\mathbb{1}\mathbb{1}^T$ *and* I_n *is the identity matrix of size* n.

[1] Graph $G_{\bar{P}}$ is formed by adding edges between two nodes i, j in V when $\bar{P}_{ij} > 0$, for all i, j, including possible self-loops.

Proof: For every k we have:

$$x(k) = Q(0,k-1)x(0) + \sum_{s=0}^{k-1} Q(s+1,k-1)w(s),$$

where $Q(\ell,m) = P(m)P(m-1)\cdots P(\ell+1)P(\ell)$. Let us denote $H = \frac{1}{n}\mathbb{1}\mathbb{1}^T$ and $\delta(k) = x(k) - \frac{1}{n}\mathbb{1}\mathbb{1}^T x(k) = x(k) - Hx(k)$. Then,

$$\delta(k) = (Q(0,k-1) - H)x(0) + \sum_{s=0}^{k-1}(Q(s+1,k-1) - H)w(s).$$

The expectation of its squared norm can be expressed as:

$$\begin{aligned}\mathbb{E}[\|\delta(k)\|^2] &= \mathbb{E}\left[\|(Q(0,k-1)-H)x(0)\|^2\right] \\ &+ 2\sum_{s=0}^{k-1} x^T(0)(Q(0,k-1)-H)^T(Q(s+1,k-1)-H)\mathbb{E}[w(s)] \\ &+ \sum_{s,r=0}^{k-1}\mathbb{E}\left[[(Q(s+1,k-1)-H)w(s)]^T[(Q(r+1,k-1)-H)w(r)]\right].\end{aligned} \tag{2.6}$$

First, we analyze the third term of (2.6) which we denote by U. By applying the trace on the scalar U, we obtain

$$U = \sum_{s,r=0}^{k-1} \mathrm{tr}\left(\mathbb{E}\left[(Q^T(s+1,k-1)Q(r+1,k-1) - H)\right]\mathbb{E}[w(r)w^T(s)]\right).$$

Due to the characteristics of the noise vector, we have:

$$\mathbb{E}[w(r)w^T(s)] = \begin{cases}\sigma^2 I_n, & r = s \\ 0, & r \neq s\end{cases},$$

which yields:

$$\begin{aligned}U &= \sigma^2 \sum_{s=0}^{k-1} \mathrm{tr}\left(\mathbb{E}\left[Q^T(s+1,k-1)Q(s+1,k-1) - H\right]\right) \\ &= \sigma^2 \sum_{s=0}^{k-1} \mathrm{tr}\left(\mathbb{E}\left[Q^T(s+1,k-1)\Omega Q(s+1,k-1)\right]\right).\end{aligned} \tag{2.7}$$

Then, we express the trace using the vectorization operator ($\mathrm{tr}(A^TB) = \mathrm{vec}^T(A)\mathrm{vec}(B)$) as:

$$U = \sigma^2 \sum_{s=0}^{k-1} \mathrm{vec}^T(I_n)\mathrm{vec}\left(\mathbb{E}\left[Q^T(s+1,k-1)\Omega Q(s+1,k-1)\right]\right).$$

Due to the idempotence of Ω and commutativity of the product $P(s)\Omega$, we can write the matrix $Q^T(s+1,k-1)\Omega Q(s+1,k-1)$ as:

$$\begin{aligned} Q^T(s+1,k-1)\Omega Q(s+1,k-1) &= P(k-1)\cdots P(s+1)\Omega P(s+1)\cdots P(k-1) \\ &= P(k-1)\cdots P(s+1)\Omega^2 P(s+1)\cdots P(k-1) \\ &= P(k-1)\cdots P(s+1)\Omega P(s+1)\Omega\cdots P(k-1) \\ &\vdots \\ &= P(k-1)\cdots P(s+1)P(s+1)\Omega\cdots P(k-1)\Omega. \end{aligned} \tag{2.8}$$

Then, we can apply the property $\mathrm{vec}(ABC) = (C^T \otimes A)\mathrm{vec}(B)$ and we get:

$$U = \sigma^2 \mathrm{vec}^T(I_n)\Bigg((I_n \otimes \Omega)\mathrm{vec}(I_n) + \sum_{s=0}^{k-2} \mathbb{E}\left[(P(k-1)\Omega \otimes P(k-1))\mathrm{vec}(P(k-2)\dots P^2(s+1)\Omega \dots P(k-2)\Omega) \right] \Bigg).$$

Following an iterative procedure we obtain:

$$U = \sigma^2 \mathrm{vec}^T(I_n)\Bigg((I_n \otimes \Omega) + \sum_{s=0}^{k-2} \mathbb{E}\left[(P(k-1)\Omega \otimes P(k-1)) \dots (P(s+1)\Omega \otimes P(s+1)) \right] \Bigg) \mathrm{vec}(I_n).$$

Since the matrices $P(s)$ are i.i.d., we get:

$$\begin{aligned} U &= \sigma^2 \mathrm{vec}^T(I_n) \left(\sum_{s=1}^{k-1} \mathbb{E}[P\Omega \otimes P]^s + (I_n \otimes \Omega) \right) \mathrm{vec}(I_n) \\ &= \sigma^2 \mathrm{vec}^T(I_n) \left(\sum_{s=0}^{k-1} \mathbb{E}[P\Omega \otimes P]^s + (I_n \otimes \Omega) - I_{n^2} \right) \mathrm{vec}(I_n), \end{aligned}$$

where we have avoided including the time dependence of matrices $P(k)$. This abuse of notation will be made multiple times in the rest of this chapter. Notice that $(I_n \otimes \Omega) - I_{n^2} = -(I_n \otimes \frac{1}{n}\mathbb{1}\mathbb{1}^T)$ and $\mathrm{vec}^T(I_n)(I_n \otimes \frac{1}{n}\mathbb{1}\mathbb{1}^T)\mathrm{vec}(I_n) = 1$. Let us recall $K = \mathbb{E}[P\Omega \otimes P]$ and consider the limit $k \to +\infty$. By applying Jensen inequality to the spectral norm $\|\cdot\|_2$, we can see that

$$\|\mathbb{E}[P\Omega \otimes P]\|_2^2 \leq \mathbb{E}[\|P\Omega \otimes P\|_2^2] = \mathbb{E}[\|P\Omega\|_2^2 \|P\|_2^2] < 1.$$

Therefore, $\sum_{s=0}^{\infty} K^s = (I_{n^2} - K)^{-1}$, and we have:

$$\lim_{k\to+\infty} U = \sigma^2 \left(\text{vec}^T(I_n)(I_{n^2} - K)^{-1}\text{vec}(I_n) - 1\right).$$

The second term of (2.6) is zero since the noise process has zero mean. The first term of (2.6), denoted by $F = \mathbb{E}\left[\|(Q(0,k-1) - H)x(0)\|^2\right]$, can be expressed as:

$$\begin{aligned} F &= \mathbb{E}\left[\left[(Q(0,k-1) - H)x(0)\right]^T \left[(Q(0,k-1) - H)x(0)\right]\right] \\ &= \mathbb{E}\left[x^T(0)(Q^T(0,k-1)Q(0,k-1) - H)x(0)\right]. \end{aligned}$$

Being $Q^T(0,k-1)Q(0,k-1)$ the product of doubly stochastic matrices, per Theorem 4 in [46] it converges to H with probability 1, so that $\lim_{n\to\infty} F = 0$ by applying the Lebesgue Dominated Convergence Theorem. Finally, J_{noise} is given by (2.5). ■

Remark 2.1 (Alternate expressions for J_{noise}) *Since Ω is idempotent and commutes with any stochastic matrix, expression* (2.8) *could be arranged in two other ways, so that instead of* (2.5) *one would obtain the analogous expression where the symmetric matrix $K = \mathbb{E}[P(k)\Omega \otimes P(k)]$ is replaced by $\mathbb{E}[P(k)\Omega \otimes P(k)\Omega]$ or by $\mathbb{E}[P(k) \otimes P(k)\Omega]$.*

2.1.1 Upper and lower bounds on the noise index

The calculation of (2.5) requires the determination of the $n^2 \times n^2$ matrix K which is impractical either for theoretical or numerical analysis. For this reason, we derive a lower bound J_{LB} and an upper bound J_{UB} for the noise index, which depend on the spectrum of symmetric n-dimensional matrices.

Theorem 2.2 (Upper and lower bounds) *For the noise index* (2.4) *we have:*

$$\frac{\sigma^2}{n}\sum_{i=2}^{n}\frac{1}{1-\lambda_i^2(\bar{P})} \le J_{\text{noise}} \le \frac{\sigma^2}{n}\sum_{i=2}^{n}\frac{1}{1-\lambda_i(\bar{\bar{P}})}, \tag{2.9}$$

where $\bar{P} = \mathbb{E}[P(k)]$ and $\bar{\bar{P}} = \mathbb{E}[P^2(k)]$.

Proof: In order to obtain the lower bound for the noise index, we can express the term (2.7) as:

$$\begin{aligned} U &= \sigma^2 \sum_{s=0}^{k-1} \text{tr}\left(\mathbb{E}\left[Q^T(s+1,k-1)\Omega Q(s+1,k-1)\right]\right) \\ &= \sigma^2 \sum_{s=0}^{k-1} \mathbb{E}\left[\|Q(s+1,k-1)\Omega\|_F^2\right], \end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius norm. Then we can apply Jensen inequality obtaining:

$$U \geq \sigma^2 \sum_{s=0}^{k-1} \|\mathbb{E}\left[Q(s+1,k-1)\Omega\right]\|_F^2.$$

Since the matrices $P(s)$ are i.i.d., we can see that $\mathbb{E}[Q(s+1,k-1)\Omega] = \bar{P}^s\Omega$ such that:

$$U \geq \sigma^2 \sum_{s=1}^{k-1} \|\bar{P}^s\Omega\|_F^2 + \|\Omega\|_F^2.$$

For a symmetric matrix P we have $\|P\|_F^2 = \|P^2\|_{S_1}$, where $\|A\|_{S_1} = \sum_{i=1}^n |\lambda_i(A)|$ is the Schatten 1-norm. Thus:

$$U \geq \sigma^2 \sum_{s=1}^{k-1} \|(\bar{P}^2\Omega)^s\|_{S_1} + \|\Omega\|_{S_1} \geq \sigma^2 \left\| \sum_{s=1}^{k-1} (\bar{P}^2\Omega)^s + \Omega \right\|_{S_1}.$$

When $k \to +\infty$, we have:

$$\begin{aligned}\lim_{k\to+\infty} U &\geq \sigma^2 \left\| \sum_{s=0}^{\infty} (\bar{P}^2\Omega)^s + \Omega - I_n \right\|_{S_1} \\ &= \sigma^2 \left\| (I_n - \bar{P}^2\Omega)^{-1} - \frac{1}{n}\mathbb{1}\mathbb{1}^T \right\|_{S_1}.\end{aligned}$$

By using the definition of the Schatten 1-norm we get:

$$\lim_{k\to+\infty} U \geq \sigma^2 \sum_{i=2}^{n} \frac{1}{1-\lambda_i^2(\bar{P})},$$

which is the desired lower bound.
For the upper bound, we consider the summation element in (2.7), denoted by S:

$$\begin{aligned}S &= \operatorname{tr}\left(\mathbb{E}\left[Q^T(s+1,k-1)(I_n - H)Q(s+1,k-1)\right]\right) \\ &= \operatorname{tr}\left(\mathbb{E}\left[Q^T(s+1,k-1)Q(s+1,k-1)\right]\right) - \operatorname{tr}(H).\end{aligned}$$

At this point, we can use the inequality

$$\operatorname{tr}(\mathbb{E}[Q^T(s+1,k-1)Q(s+1,k-1)]) \leq \operatorname{tr}(\mathbb{E}[P^2]^{k-s-1}),$$

proved in [58, Cor. 5], which yields

$$U \leq \sum_{s=0}^{k-1} \left(\operatorname{tr}(\bar{\bar{P}}^{k-s-1}) - \operatorname{tr}(H)\right) = \sum_{s=0}^{k-1} \left(\operatorname{tr}(\bar{\bar{P}}^s) - \operatorname{tr}(H)\right)$$

Then, we obtain:

$$\begin{aligned} S &\leq \mathrm{tr}(\bar{\bar{P}}^s) - \mathrm{tr}\,(H) \\ &= \mathrm{tr}\left(\bar{\bar{P}}^s - H\right) \\ &= \mathrm{tr}\left(\mathbb{E}\left[P^2(0)P^2(1)\cdots P^2(s-1) - H\right]\right) \\ &= \mathrm{tr}\left(\mathbb{E}\left[P^2(0)P^2(1)\cdots P^2(s-1)\Omega\right]\right). \end{aligned}$$

Since Ω is idempotent and $P(s)\Omega = \Omega P(s)$, we can express the latter as:

$$\begin{aligned} S &\leq \mathrm{tr}\left(\mathbb{E}\left[P(0)\Omega P(0)\cdots P(s-1)\Omega P(s-1)\right]\right) \\ &= \mathrm{tr}\left(\mathbb{E}\left[P(0)\Omega P(0)\right]\cdots\mathbb{E}\left[P(s-1)\Omega P(s-1)\right]\right) \\ &= \mathrm{tr}\left(\mathbb{E}\left[P(k)\Omega P(k)\right]^s\right). \end{aligned}$$

Then, the sum in (2.7) is upper bounded by:

$$U \leq \sigma^2 \sum_{s=1}^{k-1} \mathrm{tr}\left(\mathbb{E}\left[P(0)\Omega P(0)\right]^s\right).$$

When $k \to +\infty$, we get:

$$\begin{aligned} \lim_{k\to+\infty} U &\leq \sigma^2 \mathrm{tr}\left(\sum_{s=0}^{\infty} \mathbb{E}\left[P\Omega P\right]^s - H\right) \\ &= \sigma^2 \mathrm{tr}\left(\left(I_n - \mathbb{E}\left[P\Omega P\right]\right)^{-1} - H\right) \\ &= \sigma^2 \sum_{i=2}^{n} \frac{1}{1-\lambda_i(\bar{\bar{P}})}, \end{aligned}$$

which gives the desired upper bound. ■

Remark 2.2 (Deterministic case) *For the deterministic dynamics $x(k+1) = \bar{P}x(k)+w(t)$, that is, $P(k) = \bar{P}$, the noise index is given by $J_{\mathrm{noise}} = \frac{\sigma^2}{n}\sum_{i=2}^{n} \frac{1}{1-\lambda_i^2(\bar{P})}$. In this case, the upper and lower bounds of Theorem 2.2 coincide and recover the well-known results about deterministic consensus [37, 54]. These questions have also been extensively studied in continuous-time by the closely-related notion of network coherence [55, 56].*

2.2 Noise index in OMAS

When studying OMAS, one is confronted with the issue of agents continuously joining and leaving the system. When the pool of possible participating agents

is known, arrivals and departures can be equivalently seen as activations and deactivations [32]. In such a case, one can further assume that the interaction network is a subgraph, induced by the active nodes, of a larger network of potential interactions and specific sampling procedures can be used to generate the stochastic matrices.

Let a set V of n nodes be given, together with an undirected graph $\bar{G}$ that has V as node set and whose adjacency matrix is denoted by $\bar{A}$. At each time k, every node has an independent probability p of being *active*; at each time k, the current graph $G(k)$ is the subgraph of $\bar{G}$ that is induced by the set of active nodes. Equivalently, the current adjacency matrix $A(k)$ is generated as

$$A(k) = \Gamma(k)\,\bar{A}\,\Gamma(k), \tag{2.10}$$

where $\Gamma(k) = \mathrm{diag}[\gamma_1(k), \ldots, \gamma_n(k)]$ is a diagonal matrix with n independent identically distributed Bernoulli random variables corresponding to each node. The value of the random variable $\gamma_i(k)$ indicates the state of agent i at time k such that when $\gamma_i(k) = 1$, the node is active.

Given the adjacency matrix $A(k)$, we then define the time-varying doubly stochastic matrix given by:

$$P(k) = I_n - \epsilon(D(k) - A(k)), \tag{2.11}$$

where $D(k)$ is the diagonal matrix of the degrees of the induced graph $G(k)$, $0 < \epsilon < \frac{1}{d_{\max}(\bar{G})}$, and $d_{\max}(\bar{G})$ is the maximum degree of the graph $\bar{G}$. After defining the time-varying Laplacian $L(k) = D(k) - A(k)$, it becomes natural to refer to stochastic matrices sampled according to (2.11)-(2.10) as *Randomly Induced Discretized Laplacians (RIDL)*. It is clear that RIDL can be used to study OMAS under the assumption that the activation of each agent is an independent event, as illustrated in Fig. 2.1, where at each iteration we have a different number of active agents n_a.

We now study J_{noise} for RIDL matrices. For an exact calculation as per (2.5), it would be necessary to compute the matrix $K = \mathbb{E}[P(k)\Omega \otimes P(k)]$. To this end, we can observe that

$$\begin{aligned}
K &= \mathbb{E}[P(k)\Omega \otimes P(k)] \\
&= \mathbb{E}\left[P(k)\left(I_n - \frac{1}{n}\mathbb{1}\mathbb{1}^T\right) \otimes P(k)\right] \\
&= \mathbb{E}[P(k) \otimes P(k)] - \mathbb{E}\left[\frac{1}{n}\mathbb{1}\mathbb{1}^T \otimes P(k)\right] \\
&= \mathbb{E}[P(k) \otimes P(k)] - \frac{1}{n}\mathbb{1}\mathbb{1}^T \otimes \bar{P}.
\end{aligned}$$

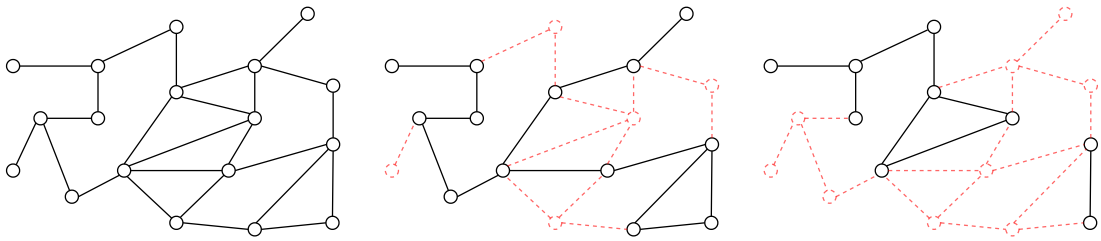


Figure 2.1: Random interactions with different numbers of active agents n_a in a network with $n = 18$ possible agents.

The second term only depends on $\bar{P}$, which is given by

$$\bar{P} = I_n - \epsilon p^2(\bar{D} - \bar{A}) = I_n - \epsilon p^2 \bar{L}, \tag{2.12}$$

where $\bar{D}$ and $\bar{L}$ are the degree and Laplacian matrices of graph $\bar{G}$, but the first term

$$\mathbb{E}[P(k) \otimes P(k)] = I_{n^2} - \epsilon p^2(I_n \otimes \bar{L} + \bar{L} \otimes I_n) + \epsilon^2 \mathbb{E}[L(k) \otimes L(k)]$$

cannot be written as a function of $\bar{P}$ or $\bar{L}$. Hence, J_{noise} cannot be written as a function of $\bar{P}$. Nevertheless, the upper and lower bounds of Theorem 2 can be expressed as functions of the eigenvalues of $\bar{P}$.

Proposition 2.1 (Bounds for Discretized Laplacian) *For the time-varying stochastic matrix* (2.11) *generated with* (2.10)*, the lower bound* J_{LB} *and upper bound* J_{UB} *in* (2.9) *become:*

$$J_{\text{LB}} = \frac{\sigma^2}{\epsilon p^2 n} \sum_{i=2}^{n} \frac{1}{2\lambda_i(\bar{L}) - \epsilon p^2 \lambda_i^2(\bar{L})}, \tag{2.13}$$

$$J_{\text{UB}} = \frac{\sigma^2}{\epsilon p^2 n} \sum_{i=2}^{n} \frac{1}{2(1 + \epsilon p - \epsilon)\lambda_i(\bar{L}) - \epsilon p \lambda_i^2(\bar{L})}, \tag{2.14}$$

where $0 = \lambda_1(\bar{L}) < \lambda_2(\bar{L}) \leq \ldots \leq \lambda_n(\bar{L})$ *are the eigenvalues of* $\bar{L}$*.*

Proof: For the lower bound, we consider the expression (2.12). Since $\bar{L}$ is symmetric, the eigenvalues of $\bar{P}$ are given by $\lambda_i(\bar{P}) = 1 - \epsilon p^2 \lambda_i(\bar{L})$ and we obtain (2.13).

For the upper bound, $\bar{\bar{P}}$ can be written as:

$$\begin{aligned}\bar{\bar{P}} &= \mathbb{E}[(I_n - \epsilon L)^2] = I_n - 2\epsilon p^2 \bar{L} + \epsilon^2 \mathbb{E}[L^2] \\ &= I_n - 2\epsilon p^2 \bar{L} + \epsilon^2(\mathbb{E}[D^2] - \mathbb{E}[DA] - \mathbb{E}[AD] + \mathbb{E}[A^2])\end{aligned}$$

Since $\mathbb{E}[\gamma_i(k)] = \mathbb{E}[\gamma_i^2(k)] = p$, we have:

$$\mathbb{E}[D^2] = (p^2 - p^3)\bar{D} + p^3\bar{D}^2, \quad \mathbb{E}[DA] = p^3\bar{D}\bar{A} + (p^2 - p^3)\bar{A},$$
$$\mathbb{E}[AD] = p^3\bar{A}\bar{D} + (p^2 - p^3)\bar{A}, \quad \mathbb{E}[A^2] = (p^2 - p^3)\bar{D} + p^3\bar{A}^2,$$

which yields:

$$\bar{\bar{P}} = I_n + 2\epsilon p^2(\epsilon - \epsilon p - 1)\bar{L} + \epsilon^2 p^3 \bar{L}^2.$$

The eigenvalues of $\bar{\bar{P}}$ are $\lambda_i(\bar{\bar{P}}) = 1 + 2\epsilon p^2(\epsilon - \epsilon p - 1)\lambda_i(\bar{L}) + \epsilon^2 p^3 \lambda_i^2(\bar{L})$ and we obtain (2.14). ■

From expressions (2.13) and (2.14) we can observe that for the computation of the bounds for J_{noise}, we only need to know the topology of $G_{\bar{P}}$ through $\bar{L}$.

Furthermore, inspired by the work in [59], we relate J_{noise} with the well-studied *Average Effective Resistance* (or *Kirchhoff Index*) of the graph. The latter can indeed be expressed [37, Chapter 5] as a function of Laplacian eigenvalues by

$$R_{\text{ave}} := \frac{1}{n}\sum_{i=2}^{n} \frac{1}{\lambda_i(\bar{L})}.$$

Proposition 2.2 (Bounds and average effective resistance) *For the noise index* (2.4) *with the time-varying stochastic matrix* (2.11) *generated by* (2.10) *we have:*

$$\frac{\sigma^2}{2p^2}\frac{R_{\text{ave}}}{\epsilon} \leq J_{\text{noise}} \leq \frac{\sigma^2}{2p^3(1 - r_\epsilon)}\frac{R_{\text{ave}}}{\epsilon},$$

where $r_\epsilon = \epsilon d_{\max}(\bar{G}) < 1$.

Proof: For the lower bound, we immediately observe from (2.13) that

$$J_{\text{LB}} \geq \frac{\sigma^2}{\epsilon p^2 n}\sum_{i=2}^{n}\frac{1}{2\lambda_i(\bar{L})} = \frac{\sigma^2}{2\epsilon p^2}R_{\text{ave}}.$$

For the upper bound, we remind $\lambda_i(\bar{L}) \leq 2d_{\max}(\bar{G})$ and obtain:

$$\begin{aligned} 2(1 + \epsilon p - \epsilon)\lambda_i(\bar{L}) - \epsilon p\lambda_i^2(\bar{L}) &= 2\epsilon(p-1)\lambda_i(\bar{L}) + (2 - \epsilon p\lambda_i(\bar{L}))\lambda_i(\bar{L}) \\ &\geq 2\epsilon(p-1)\lambda_i(\bar{L}) + \left(2 - \epsilon p\left(2d_{\max}(\bar{G})\right)\right)\lambda_i(\bar{L}) \\ &\geq 2p(1 - r_\epsilon)\lambda_i(\bar{L}), \end{aligned}$$

which yields:

$$J_{\text{UB}} \leq \frac{\sigma^2}{\epsilon p^2 n}\sum_{i=2}^{n}\frac{1}{2p(1 - r_\epsilon)\lambda_i(\bar{L})} = \frac{\sigma^2}{2\epsilon p^3(1 - r_\epsilon)}R_{\text{ave}},$$

thereby completing the proof. ■

The bounds of Proposition 2.2 can also be equivalently expressed as

$$\frac{\sigma^2}{2p^2 r_\epsilon} d_{\max}(\bar{G}) R_{\text{ave}} \leq J_{\text{noise}} \leq \frac{\sigma^2}{2p^3 r_\epsilon (1-r_\epsilon)} d_{\max}(\bar{G}) R_{\text{ave}}.$$

Since only $d_{\max}(\bar{G})$ and R_{ave} depend on the graph (and therefore on n), this form of the bounds implies that

$$J_{\text{noise}} = \Theta(d_{\max}(\bar{G}) R_{\text{ave}}) \quad \text{as} \quad n \to \infty,$$

where $\Theta(\cdot)$ corresponds to the Θ-notation [2], and proves that the bounds are asymptotically tight for RIDLs. For instance, for sparse graphs with bounded degree, the rate of growth of J_{noise} is equal to the rate of R_{ave}. Instead, for dense graphs where $d_{\max}(\bar{G})$ is linear in n, the rate of growth of J_{noise} is the rate of nR_{ave}.

2.3 Particular graph topologies

In this section, we illustrate the behavior of the noise index for RIDL matrices associated to graphs with particular structures, which allow for the explicit computation of the bounds in Theorem 2.2 and Proposition 2.1. We consider $\epsilon = r_\epsilon / d_{\max}(\bar{G})$ with $r_\epsilon \in (0,1)$ and we perform computations for networks with $3 \leq n \leq 100$, $p = 0.9$, and $\sigma^2 = 1$.

2.3.1 Sparse graphs

Star graphs: for the star graph S_n we have $d_{\max}(S_n) = n-1$ and we choose $\epsilon = r_\epsilon/(n-1)$. The eigenvalues of the Laplacian matrix are given by $\lambda_i(\bar{L}) = 1$ for $i = 2, \ldots, n-1$ and $\lambda_n(\bar{L}) = n$, so that the bounds are:

$$J_{\text{LB}} = \frac{\sigma^2}{\epsilon p^2 n} \left(\frac{n-2}{2-\epsilon p^2} + \frac{1}{n(2-\epsilon p^2 n)} \right); \tag{2.15}$$

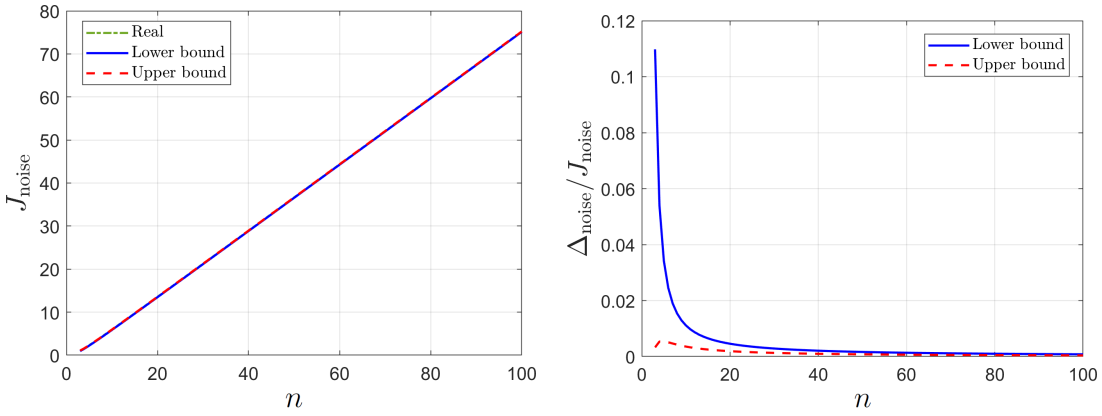
$$J_{\text{UB}} = \frac{\sigma^2}{\epsilon p^2 n} \left(\frac{n-2}{\epsilon p + 2 - 2\epsilon} + \frac{1}{n(2\epsilon p + 2 - 2\epsilon - \epsilon p n)} \right), \tag{2.16}$$

implying that

$$J_{\text{noise}} = \frac{\sigma^2}{2 r_\epsilon p^2} n + o(n) \quad \text{as} \quad n \to \infty,$$

[2] For a given function $g(n)$, we denote by $\Theta(g(n))$ the set of functions [60]

$$\Theta(g(n)) = \{f(n) : \text{there exist positive constants } c_1,\ c_2 \text{ and } n_0 \text{ such that } 0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n) \text{ for all } n \geq n_0\}.$$



(a) J_{noise} (b) Relative error

Figure 2.2: Computation of J_{noise} and the relative error for a star graph S_n with growing n. In the left plot, the solid blue line corresponds to the lower bound (2.15) and the dashed red line is the upper bound (2.16).

where $o(\cdot)$ corresponds to the o-notation [3]. Figures 2.2(a) and 2.2(b) present the results of the computations for the star graph with $r_\epsilon = 0.8$.

Line and grids: for the line graph P_n, $d_{\max}(P_n) = 2$ and we select $\epsilon = 0.8/2$ for all n. The eigenvalues of the Laplacian matrix are given by $\lambda_i(\bar{L}) = 2 - 2\cos\left(\frac{\pi(i-1)}{n}\right)$ for $i = 2, \ldots, n$ such that the corresponding bounds are:

$$J_{\text{LB}} = \frac{\sigma^2}{4\epsilon p^2 n} \sum_{i=1}^{n-1} \frac{1}{1 - \epsilon p^2 - \epsilon p^2 \cos^2\left(\frac{\pi i}{n}\right) + (2\epsilon p^2 - 1)\cos\left(\frac{\pi i}{n}\right)}; \tag{2.17}$$

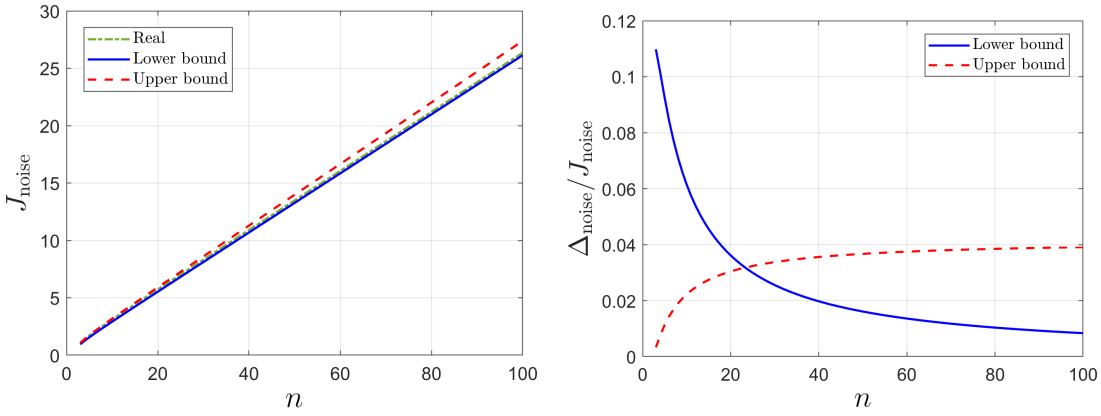
$$J_{\text{UB}} = \frac{\sigma^2}{4\epsilon p^2 n} \sum_{i=1}^{n-1} \frac{1}{1 - \epsilon - \epsilon p \cos^2\left(\frac{\pi i}{n}\right) + (\epsilon p + \epsilon - 1)\cos\left(\frac{\pi i}{n}\right)}, \tag{2.18}$$

and the rate of growth of the noise index is $J_{\text{noise}} = \Theta(n)$. Fig. 2.3(a) and 2.3(b) present the results of the computations for the line graph.

The maximum degree for a 2-D grid is $d_{\max}(\text{2-D}) = 4$ and for a 3-D grid is $d_{\max}(\text{3-D}) = 6$. The spectrum of the Laplacian matrix for a m-D grid are given by:

[3] We define $o(g(n))$ as the set [60]

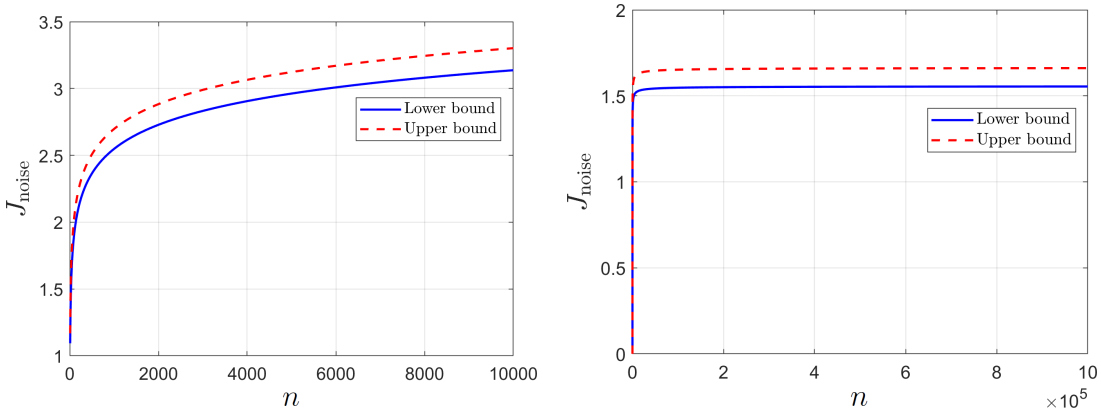
$$o(g(n)) = \{f(n) : \text{for any positive constant } c > 0, \text{ there exists a constant } n_0 > 0 \text{ such that } 0 \leq f(n) < cg(n) \text{ for all } n \geq n_0\}.$$



(a) J_{noise} (b) Relative error

Figure 2.3: Computation of J_{noise} and the relative error for a line graph P_n with growing n. In the left plot, the solid blue line corresponds to the lower bound (2.17) and the dashed red line is the upper bound (2.18).

$2m - 2\sum_{i=1}^{m} \cos(\frac{\pi}{n} h_i)$ with $h_i \in \{0, \ldots, n-1\}$. Due to the complexity of dealing with the closed form from Theorem 1, we present only the behavior of the bounds for the 2-D grid in Fig 2.4(a) and for the 3-D grid in Fig 2.4(b). These examples also confirm the theoretical analysis.



(a) 2-D grid (b) 3-D grid

Figure 2.4: Computation of J_{noise} for 2−D and 3−D grids with growing n.

2.3.2 Dense graphs

Complete graph: since for a complete graph K_n, $d_{\max}(K_n) = n-1$, we select $\epsilon = r_\epsilon/(n-1)$. The eigenvalues of the Laplacian matrix of the complete graph K_n are $\lambda_i(\bar{L}) = n$ for $i = 2, \ldots, n$ such that the bounds for J_{noise} are:

$$J_{\text{LB}} = \frac{\sigma^2(n-1)}{\epsilon p^2 n^2(2-\epsilon p^2 n)}; \tag{2.19}$$

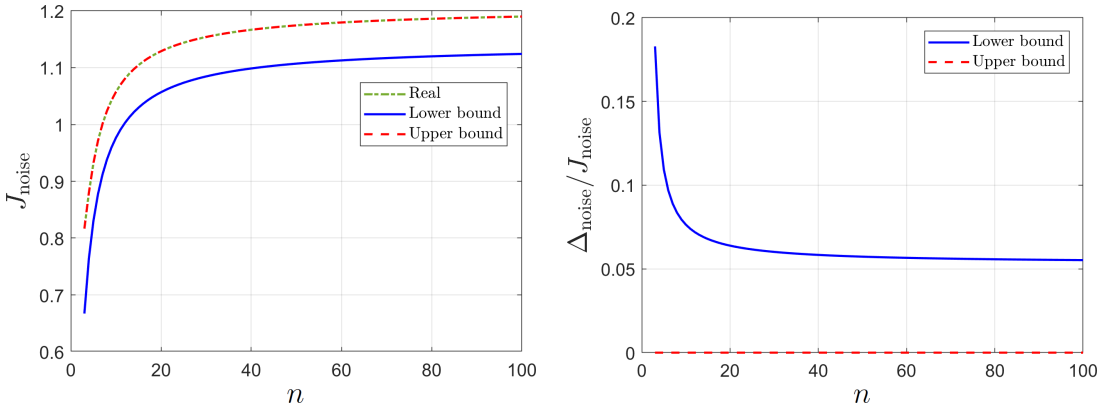
$$J_{\text{UB}} = \frac{\sigma^2(n-1)}{\epsilon p^2 n^2(2+2\epsilon p-2\epsilon-\epsilon p n)}, \tag{2.20}$$

implying a rate of growth $J_{\text{noise}} = \Theta(1)$. The limit when $n \to \infty$ is given by:

$$\lim_{n\to\infty} J_{\text{LB}} = \frac{\sigma^2}{p^2 r_\epsilon(2-p^2 r_\epsilon)};$$

$$\lim_{n\to\infty} J_{\text{UB}} = \frac{\sigma^2}{p^2 r_\epsilon(2-p r_\epsilon)}.$$

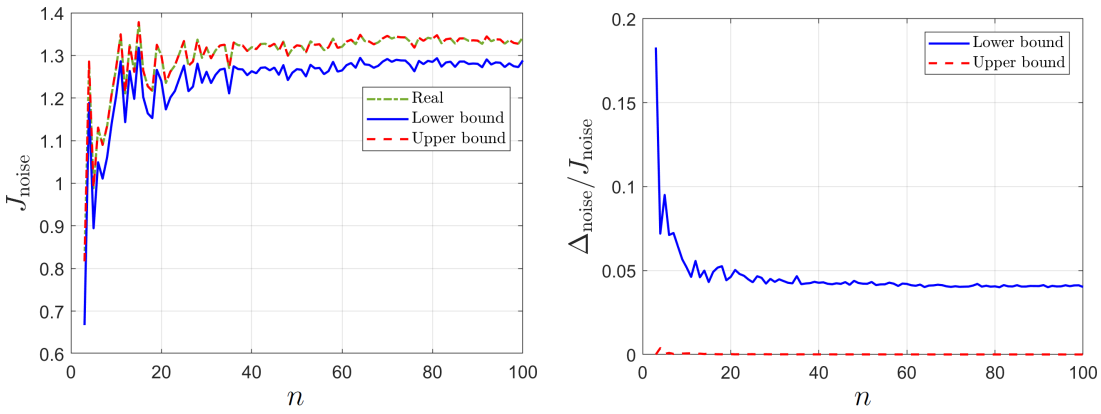
For the computations we choose $r_\epsilon = 0.8$ and we obtain the results in Figures 2.5(a) and 2.5(b): the latter implies that the upper bound is equal to the noise index.



(a) J_{noise} (b) Relative error

Figure 2.5: Computation of J_{noise} and the relative error for a complete graph K_n with growing n. In the left plot, the solid blue line corresponds to the lower bound (2.19) and the dashed red line is the upper bound (2.20).

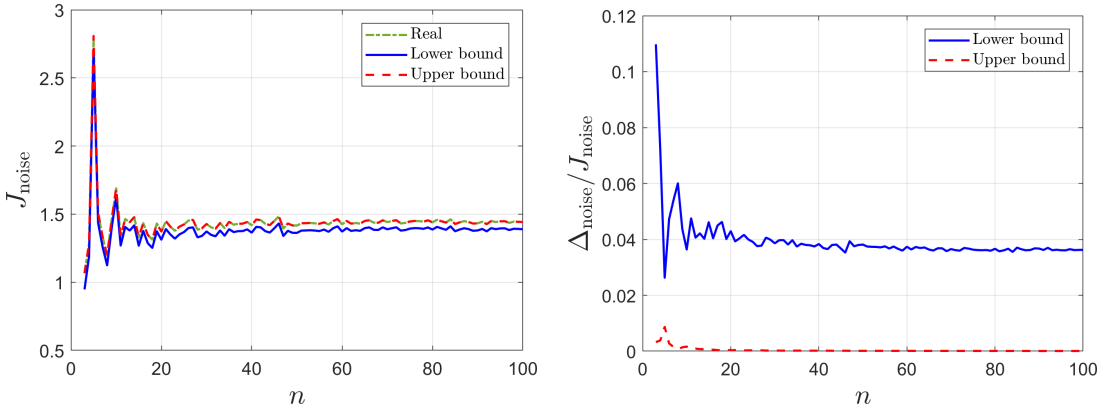
Erdős-Rényi graph: each edge between two nodes is generated with a probability p_{er}, independently of any other edge. As an example, figures 2.6(a) and 2.6(b) present the results of the computations for the Erdős-Rényi graph $p_{\text{er}} = 0.8$ and $\epsilon = 0.8/(n-1)$.



(a) J_{noise} (b) Relative error

Figure 2.6: Computation of J_{noise} and the relative error for a sequence of Erdős-Rényi graphs with growing n.

Graphon: we consider the graphon $W(x, y) = 1 - xy$ corresponding to the limit of a *randomly grown ranked attachment graph sequence* [4] [61]. We generate graphs from the graphon according to the Definition A.1 and we select $\epsilon = 0.8/(n-1)$. The results of the computations for the graphon $W(x, y) = 1 - xy$ are presented in Fig. 2.7(a) and 2.7(b).

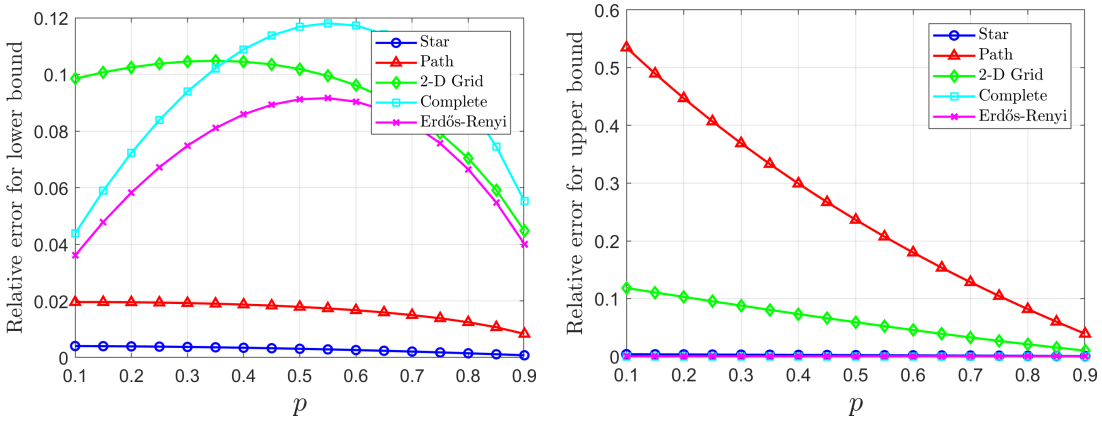


(a) J_{noise} (b) Relative error

Figure 2.7: Computation of J_{noise} and the relative error for a sequence of graphs sampled from the graphon $W(x, y) = 1 - xy$ with growing n.

[4] The sequence starts with a single node. At the n-th iteration, a new node is added and it is connected to node i with probability $1 - i/n$. Then, every pair of nonadjacent nodes is connected with probability $2/n$.

Variation of p: figures 2.8(a) and 2.8(b) present the behavior of the relative error for the bounds of Theorem 2.2 for graphs with $n = 100$, $\epsilon = 0.8/d_{\max}(\bar{G})$ and $0.1 \leq p \leq 0.9$.



(a) Lower bound (b) Upper bound

Figure 2.8: Computations of the relative error for different values of p for graphs with $n = 100$.

2.4 Conclusion

In this chapter, we analyzed the influence of noise in random consensus with an application to OMAS by considering the random sampling of active nodes. We derived an expression for a noise index that measures the deviation from the consensus point and an upper and lower bound that depend on the spectrum of expected matrices related to the network. For particular cases of OMAS characterized by activations and deactivations of agents, we defined Randomly Induced Discretized Laplacians and expressed the bounds as functions of the eigenvalues of the expected Laplacian matrix and, more specifically, of its average effective resistance. For sparse graphs, such as stars and paths, the noise index grows linearly in n and the lower bound presents a better approximation of the real value. For dense graphs, such as complete, Erdős-Rényi and graphs sampled from graphons, the noise index is bounded in n and the upper bound is a better approximation.

For future work, we identify three main possible directions of the research. Regarding the rich family of graphs sampled from graphons, the eigenvalues of the Laplacian matrix are related to the degree function of the graphon [62]. It would be interesting to consider approximations of the bounds by using appropriate definitions that include the degree function of the graphon as in the case of the average effective resistance in [62].

A key assumption for the derivation of the results in this chapter is the independence of active sets across time. However, it would be important to consider the scenario where only few agents change their states from active to inactive or vice versa while the activation state of other agents remain invariant. In this case, independence in the stochastic matrices will not hold and a certain degree of statistical dependence in the generation of the stochastic matrices should be considered.

The metric (2.4) used for the analysis of OMAS was normalized considering the total number of potential agents in the system, i.e., n. We would like to analyze an alternative definition of this index function that could be more informative for OMAS, by considering only the current number of active agents at each iteration, i.e., n_a, as the normalization factor.

Chapter 3

Weighted gradient descent with packet losses and agent replacements

The *resource allocation* (RA) is an important problem in optimization where a budget must be optimally distributed among multiple entities or activities [63, 64]. In the case of multi-agent systems, the objective is to find an optimal distribution of the budget among the agents while each agent aims to minimize the whole cost function but it can only access its own local cost and state and those of its neighbors [65, 66]. Applications include energy resources [67], smart grids [68], actuator networks [69] games [70], and distributed computer systems [71]. The majority of the algorithms are designed for ideal scenarios with perfect communication and a symmetric exchange of information through the network that maintain the budget fixed (keeping the budget fixed is indeed the constraint of the optimization problem). However, the RA problem can also be solved in digraphs by introducing additional variables to preserve the constraint, which increases the complexity of the algorithms [72, 73]. One of the most important algorithms used to solve the RA problem in a multi-agent system is the weighted gradient descent proposed in [74], where the update rule makes use of a matrix associated to the network that preserves the constraint.

The classical formulation of the RA problem assumes that the set of agents remains invariant during the optimization process; however, this assumption could not be valid for some scenarios. Consider, for instance, the case of distributed energy resources where a fixed amount of energy must be supplied by a network of devices in an optimal way [67]. Nevertheless, in real life, some of the devices might experience failures with higher probability as the system size increases, or change their operating point due to environmental conditions, generating an open system where the cost functions of agents can change. Optimization in OMAS is a recent topic research, where two different types of formulations have been considered. In [33, 75], the authors study the case of different active agents interacting in the

network for the implementation of an optimization algorithm, while the size of the system is fixed and all the possible cost functions are known. In [36], the openness of the system is characterized by replacements, where new cost functions can be assigned to the agents, implying a possible change of the global minimizer.

In addition to potential changes in the set of agents, communication networks can suffer from errors in the communications that cause the loss of information. In this case, even if the graph associated with the communication network was originally undirected, interactions between the agents can effectively be asymmetric when the packet loss occurs only in one direction. The symmetry of the system being broken, relevant quantities may fail to be preserved in many scenarios [16, 76, 77]. In the case of the average estimation under packet losses, [78] derived upper bounds for the mean squared error to evaluate the deviation from the initial average for general graph topologies. While a possible deviation of the initial average value is not necessarily problematic in the average estimation where the main objective is to reach consensus in most cases, for optimization problems, similar errors can be more consequential, since they may imply the violation of the constraint. For example in the case of the RA problem, agents could perform unnecessary tasks or the sum of the individual tasks may not satisfy the global demand. Indeed, the analysis of the constraint violation in complex scenarios has been the object of many works, specially in the framework of online optimization, where it is assumed that cost functions change in time [79].

In this chapter, we analyze the effects of packet losses in the performance of the weighted gradient descent algorithm to solve the RA problem. In Section 3.1 we introduce the formal formulation of the RA problem, including the assumptions used for the derivation of the results of this chapter. Section 3.2 presents the problem of packet losses in the weighted gradient descent algorithm and the conditions to guarantee convergence of the algorithm to the minimizer at least in expectation. We define appropriate metrics to measure the deviation from the constraint and the error of the expected value of the cost function and we derive upper bounds. In Section 3.3 we analyze the weighted gradient descent in OMAS where replacements of the agents can take place and we show that the constraint error can diverge for this particular scenario. Finally, conclusions and future work are exposed in Section 3.4.

3.1 Resource Allocation Problem

In multi-agent systems, the RA problem is formulated as the minimization of an objective function f that is separable in local costs $f_i : \mathbb{R}^d \to \mathbb{R}$ 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 a budget composed by the sum of the demands of the agents $d_i \in \mathbb{R}^d$.

The problem can then be written as

$$\min_{x=[x_1^T,\dots,x_n^T]^T\in\mathbb{R}^{nd}} f(x)=\sum_{i=1}^{n} f_i(x_i) \qquad \text{subject to} \qquad \sum_{i=1}^{n} a_i x_i=\sum_{i=1}^{n} d_i, \tag{3.1}$$

where $a_i > 0$ is the weight of agent i to satisfying the constraint. When the agents are isolated, their objective is to minimize the corresponding local cost functions f_i. In the RA problem (3.1), each agent aims at minimizing the global cost function f while guaranteeing the budget. However, since each agent only has access to local information, they need to exchange the current states of the gradients to achieve the goal.

If the budget is fixed in time $\sum_{i=1}^{n} d_i = b$, then the constraint in (3.1) can be equivalently expressed as $\left(a^\top \otimes I_d\right) x = b$, where $a = [a_1, \dots, a_n]^T$ and we remind that $\otimes$ denotes the Kronecker product. The feasible set of (3.1) is thus given by

$$\mathcal{S}_{a,b} := \left\{ x \in \mathbb{R}^{nd} \middle| \left(a^\top \otimes I_d\right) x = b \right\}. \tag{3.2}$$

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 3.1 *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 3.1 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 3.1 is denoted by $\mathcal{F}_{\alpha,\beta}$.

Since all the local cost functions f_i are α-strongly convex and β-smooth, the global cost function f from (3.1) also satisfies $f \in \mathcal{F}_{\alpha,\beta}$. As a consequence, Assumption 3.1 guarantees that the solution of the problem (3.1) denoted by x^* is unique. Therefore, for some $\zeta^* \in \mathbb{R}^d$, a necessary and sufficient condition for the optimality of x^* is

$$\nabla f(x^*) = (a^\top \otimes I_d)^\top \zeta^* = a \otimes \zeta^*. \tag{3.3}$$

In this chapter, we restrict to the specific case defined by the following assumptions, while in Chapter 4 we consider the more general case of the RA problem.

Assumption 3.2 (1-D functions) *The local cost function of any agent at any time is one-dimensional: $f_i^t : \mathbb{R} \to \mathbb{R}$.*

Assumption 3.3 (Fixed demand) *The total demand associated with the system at any time is $\sum_{i=1}^{n} d_i = b$.*

Assumption 3.4 (Homogeneous agents) *The weight associated with any agent i at any time is $a_i = 1$.*

For this particular setting, the RA problem (3.1) can be formulated as

$$\min_{x=[x_1,\ldots,x_n]\in\mathbb{R}^n} f(x) = \sum_{i=1}^{n} f_i(x_i) \qquad \text{subject to} \qquad \sum_{i=1}^{n} x_i = b, \tag{3.4}$$

and the optimality condition implies that $\mathbb{1}^T x^* = b$ and $\nabla f^* = \zeta^* \mathbb{1}$ for some $\zeta^* \in \mathbb{R}$ where $\nabla f^* = \nabla f(x^*)$. The feasible set of (3.4) is given by

$$\mathcal{S}_{\mathbb{1},b} := \left\{ x \in \mathbb{R}^n | \mathbb{1}^T x = b \right\}. \tag{3.5}$$

3.2 Packet losses

We consider that agents interact through a connected undirected graph $G = (V, E)$, where $V = \{1, \ldots, n\}$ is the set of agents and $E \subseteq V \times V$ is the set of edges. To the graph G we associate an adjacency matrix $A = [a_{ij}]$ where $a_{ij} = 1$ if $(i, j) \in \mathbb{E}$ and $a_{ij} = 0$ otherwise. We assume that the graph does not have self-loops (i.e., $a_{ii} = 0$). The Laplacian matrix of the graph G is defined as $L = D - A$ where $D = \text{diag}[d_1, \ldots, d_n]$ and d_i is the degree of agent i (i.e., d_i is the ith row-sum of A). Notice that L is a symmetric matrix that satisfies $L = L^T$.

To solve problem (3.4) at each iteration we apply the weighted gradient descent algorithm with positive step-size h [74]:

$$x(k+1) = x(k) - hL\nabla f(x(k)), \tag{3.6}$$

where L determines the exchange of information between the agents according to the network topology.

For the RA problem formulated in 3.4 with local cost functions satisfying Assumption 3.1 we provide an upper bound for the step-size h to guarantee linear convergence of the cost function.

Proposition 3.1 *Under Assumptions 3.1 to 3.4, for any initial point $x_0 \in \mathcal{S}_{\mathbb{1},b}$, the sequence $\{x(k), k > 0\}$, produced by (3.6) with $h \leq \frac{2}{\beta\lambda_n(L)}$, satisfies:*

$$f(x(k+1)) - f(x^*) \leq (1 - \alpha h \lambda_2(L)(2 - h\beta\lambda_n(L))) \left(f(x(k)) - f(x^*)\right), \tag{3.7}$$

where $\lambda_n(L)$ and $\lambda_2(L)$ are the largest and the second smallest eigenvalues of L respectively.

Proof: Since the cost function is β-smooth we have:

$$\begin{aligned} f(x(k+1)) &\leq f(x(k)) + \langle \nabla f(x(k)), x(k+1) - x(k) \rangle + \frac{\beta}{2} \left|\left| x(k+1) - x(k) \right|\right|^2 \\ &= f(x(k)) + \langle \nabla f(x(k)), -hL\nabla f(x(k)) \rangle + \frac{\beta}{2} \left|\left| hL\nabla f(x(k)) \right|\right|^2 \\ &\leq f(x(k)) - h \left|\left| L^{1/2}\nabla f(x(k)) \right|\right|^2 + \frac{\beta h^2 \lambda_n(L)}{2} \left|\left| L^{1/2}\nabla f(x(k)) \right|\right|^2 \\ &= f(x(k)) - \frac{h}{2}(2 - \beta h \lambda_n(L)) \left|\left| L^{1/2}\nabla f(x(k)) \right|\right|^2 . \end{aligned}$$

Using a vector projection on $\mathbb{1}$, the gradient can be expressed as: $\nabla f(x(k)) = \xi_\perp \mathbb{1} + \nabla f(x(k))_\perp$, where $\xi_\perp = \frac{1}{n}\langle \nabla f(x(k)), \mathbb{1} \rangle$. Since $h \leq \frac{2}{\beta \lambda_n(L)}$, we can upper bound the last expression by:

$$\begin{aligned} f(x(k+1)) &\leq f(x(k)) - \frac{h}{2}(2 - \beta h \lambda_n(L)) \left|\left| \xi_\perp L^{1/2}\mathbb{1} + L^{1/2}\nabla f(x(k))_\perp \right|\right|^2 \\ &\leq f(x(k)) - \frac{h}{2}(2 - \beta h \lambda_n(L)) \left|\left| L^{1/2}\nabla f(x(k))_\perp \right|\right|^2 \\ &\leq f(x(k)) - \frac{h\lambda_2(L)}{2}(2 - \beta h \lambda_n(L)) \left|\left| \nabla f(x(k))_\perp \right|\right|^2 , \end{aligned} \tag{3.8}$$

where we used the fact that $L^{1/2}\mathbb{1} = \mathbf{0}$ since $\mathbb{1}$ belongs to the kernel of $L^{1/2}$. The expression (3.8) shows the convergence of the algorithm when the step-size satisfies $h \leq \frac{2}{\beta \lambda_n(L)}$. Furthermore, by using the properties of α-strongly convex functions, we have the following inequality [80, Proof of Theorem 4.1]:

$$f(x(k)) - f(x^*) \leq \frac{1}{2\alpha} \left|\left| \nabla f(x(k))_\perp \right|\right|^2 . \tag{3.9}$$

Finally, by combining (3.8) and (3.9) we obtain:

$$\begin{aligned} f(x(k+1)) - f(x^*) &\leq f(x(k)) - f(x^*) - \frac{h\lambda_2(L)}{2}(2 - \beta h \lambda_n(L)) \left|\left| \nabla f(x(k))_\perp \right|\right|^2 \\ &\leq f(x(k)) - f(x^*) - \alpha h \lambda_2(L)(2 - \beta h \lambda_n(L)) \left(f(x(k)) - f(x^*) \right) . \end{aligned}$$

■

In this analysis, we consider that the communication network is unreliable and some messages can be lost. Following the approach in [78] we assume that at each time instant $k \in \mathbb{Z}_{\geq 0}$, we have a matrix $\Gamma(k) = [\Gamma(k)^{ij}] \in \{0,1\}^{n \times n}$ composed of Bernoulli random variables γ_k^{ij} independent across k, i, j with $\mathbb{P}(\gamma_k^{ij} = 1) = p$ and $p \in (0,1)$. Then, for an edge $a_{ij} = 1$, if $\gamma_k^{ij} = 1$ then the link a_{ij} is active, otherwise there was a failure of the link a_{ij} and the data transmitted was lost. If $a_{ij} = 0$, the value of γ_k^{ij} does not have any effect.

The topology of the network with possible packet losses at each time instant is thus defined by:

$$A(k) = A \odot \Gamma(k); \tag{3.10}$$

$$L(k) = D(k) - A(k), \tag{3.11}$$

and the weighted gradient descent algorithm becomes

$$x(k+1) = x(k) - hL(k)\nabla f(x(k)). \tag{3.12}$$

Due to the asymmetric loss of information, the graph topology at each time instant is effectively directed with a non-symmetric Laplacian matrix $L(k)$. Because of this lack of symmetry, the constraint can be violated during an update of the states of the network using (3.12) and hence, the minimizer of the problem (3.1) will not be reached.

Nevertheless, it is desirable that algorithm (3.12) solves (3.1) at least in expectation. Since $\Gamma(k)$ is independent in time and $\mathbb{E}[L(k)] = pL$, the expected dynamics is given by:

$$\mathbb{E}[x(k+1)] = \mathbb{E}[x(k)] - hpL\mathbb{E}[\nabla f(x(k))]. \tag{3.13}$$

Dynamics (3.13) preserves the constraint in expectation since:

$$\mathbb{1}^T\mathbb{E}[x(k+1)] = \mathbb{1}^T(\mathbb{E}[x(k)] - hpL\mathbb{E}[\nabla f(x(k))]) = \mathbb{1}^T\mathbb{E}[x(k)]. \tag{3.14}$$

However, the minimizer x^* is not necessarily a stationary point of (3.13), as shown by the following discussion. Let us assume that $\mathbb{E}[x(k)] = x^*$, so that we have

$$\mathbb{E}[x(k+1)] = x^* - hpL\mathbb{E}[\nabla f(x(k))].$$

Under the assumption that the local functions are β-smooth and α-strongly convex we cannot guarantee that $\mathbb{E}[\nabla f(x(k))] = \nabla(f(\mathbb{E}[x(k)])) = \zeta\mathbb{1}$. For this reason, we restrict the set of possible cost functions to continuous differentiable piecewise quadratic functions, which are β-smooth and α-strongly convex.

Definition 3.1 (Piecewise quadratic function) *A continuous differentiable function $f : \mathbb{R} \to \mathbb{R}$ is piecewise quadratic if there exists finitely many quadratic functions $\{q_i\}_{i=1}^I$ such that $f(x) \in \{q_i(x)\}_{i=1}^I$ for all $x \in \mathbb{R}$.*

Assumption 3.5 (Piecewise quadratic functions) *The local cost function of any agent at any time is piecewise quadratic.*

In this case, the gradient is a vector of piecewise linear functions $[\nabla f(x)]_i = m_i x_i + c_i$ and we have

$$\mathbb{E}\left[\nabla f(x(k))\right] = \nabla f\left(\mathbb{E}\left[x(k)\right]\right). \tag{3.15}$$

Then, the expected dynamics (3.13) becomes:

$$\mathbb{E}\left[x(k+1)\right] = \mathbb{E}\left[x(k)\right] - hpL\nabla f\left(\mathbb{E}\left[x(k)\right]\right), \tag{3.16}$$

which corresponds to the original weighted gradient descent, thereby guaranteeing that the constraint is preserved at each time instant and that the minimizer is reached such that

$$\lim_{k\to\infty} \mathbb{E}\left[x(k)\right] = x^* \quad \text{and} \quad \lim_{k\to\infty} f\left(\mathbb{E}\left[x(k)\right]\right) = f^*,$$

where $f^* = f(x^*)$. However, a similar equality is not true for the expected value of the cost function. Since the functions are convex, by using Jensen's inequality we obtain:

$$\mathbb{E}\left[f(x(k))\right] \geq f\left(\mathbb{E}\left[x(k)\right]\right),$$

and by taking the limit (whose existence will be discussed in the next section) we get:

$$\lim_{k\to\infty} \mathbb{E}\left[f(x(k))\right] \geq f^*, \tag{3.17}$$

where the inequality is in general strict.

3.2.1 Constraint violation

Although the constraint is preserved in expectation according to (3.14), this fact does not imply that the constraint is also preserved for any particular realization of the process. We thus aim to measure the deviation from the constraint by deriving an upper bound for the constraint violation metric:

$$J_{\text{constr}} := \lim_{k\to\infty} \mathbb{E}\left[\left(\mathbb{1}^T x(k) - b\right)^2\right], \tag{3.18}$$

which corresponds to the asymptotic mean squared error.

Before presenting the main result of this subsection, we recall a lemma derived in [78].

Lemma 3.1 *For $L(k)$ given by* (3.10) *and* (3.11)*:*

$$\mathbb{E}\left[L(k)^T L(k)\right] = p^2 L^2 + 2p(1-p)L \tag{3.19}$$

$$\mathbb{E}\left[L(k)^T \mathbb{1}\mathbb{1}^T L(k)\right] = 2p(1-p)L. \tag{3.20}$$

Theorem 3.1 *Consider the weighted gradient descent given by* (3.10), (3.11) *and* (3.12). *Under Assumption 3.5, for any positive scalar* $h \leq \frac{2}{\beta(p\lambda_n(L)+2-2p)}$, *the constraint violation metric satisfies:*

$$J_{\text{constr}} \leq \frac{4h(1-p)}{2-h\beta(p\lambda_n(L)-2p+2)}\left(f(x_0)-f^*\right). \tag{3.21}$$

Proof: Let us denote by $H(k) = \left(\mathbb{1}^T x(k) - b\right)$. For (3.12) we denote by $\{\mathcal{F}_k\}_{k\in\mathbb{Z}_{\geq 0}}$ the filtration of σ-algebras generated by the process $x(k)$. We compute the expectation conditioned upon the filtration generated by $x(k)$:

$$\begin{aligned}
\mathbb{E}\left[H^2(k+1)|\mathcal{F}_k\right] &= \mathbb{E}\left[\left(\mathbb{1}^T\left(x(k) - hL(k)\nabla f(x(k))\right) - b\right)^2\right] \\
&= H^2(k) + h^2\nabla f(x(k))^T\mathbb{E}\left[L(k)^T\mathbb{1}\mathbb{1}^T L(k)\right]\nabla f(x(k)) \\
&\quad - hH(k)\mathbb{1}^T\mathbb{E}\left[L(k)\right]\nabla f(x(k)) \\
&= H^2(k) + h^2\nabla f(x(k))^T\mathbb{E}\left[L(k)^T\mathbb{1}\mathbb{1}^T L(k)\right]\nabla f(x(k)),
\end{aligned}$$

where we have used the fact that $\mathbb{1}^T L = 0$. Then, we use (3.20) to obtain:

$$\begin{aligned}
\mathbb{E}\left[H^2(k+1)|\mathcal{F}_k\right] &= H^2(k) + 2h^2p(1-p)\nabla f(x(k))^T L\nabla f(x(k)) \\
&= H^2(k) + 2h^2p(1-p)\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2.
\end{aligned}$$

If we take the total expectation we get:

$$\mathbb{E}\left[\mathbb{E}\left[H^2(k+1)|\mathcal{F}_k\right]\right] = \mathbb{E}\left[H^2(k)\right] + 2h^2p(1-p)\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2\right].$$

By using a recursive argument we have that at the time instant k:

$$\begin{aligned}
\mathbb{E}\left[H^2(k)\right] &= H_0^2 + 2h^2p(1-p)\sum_{i=0}^{k-1}\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x_i)\right|\right|^2\right] \\
&= 2h^2p(1-p)\sum_{i=0}^{k-1}\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x_i)\right|\right|^2\right],
\end{aligned} \tag{3.22}$$

where we used the fact that at the initial time $k = 0$, the constraint is not violated $H_0 = 0$. Now, let us consider the inequality corresponding to β-smooth functions:

$$\begin{aligned}
f(x(k+1)) &\leq f(x(k)) + \frac{\beta}{2}\left|\left|x(k) - x(k+1)\right|\right|^2 + \langle\nabla f(x(k)), x(k+1) - x(k)\rangle \\
&= f(x(k)) + \frac{h^2\beta}{2}\left|\left|L(k)\nabla f(x(k))\right|\right|^2 - h\langle\nabla f(x(k)), L(k)\nabla f(x(k))\rangle \\
&= f(x(k)) + \frac{h^2\beta}{2}\langle\nabla f(x(k)), L(k)^T L(k)\nabla f(x(k))\rangle \\
&\quad - h\langle\nabla f(x(k)), L(k)\nabla f(x(k))\rangle.
\end{aligned}$$

We compute the expectation given the filtration generated by $x(k)$ and we use (3.19) to get:

$$\begin{aligned}\mathbb{E}\left[f(x(k+1))|\mathcal{F}_k\right] &\leq f(x(k)) + \frac{h^2\beta}{2}\langle\nabla f(x(k)), \mathbb{E}\left[L(k)^T L(k)\right]\nabla f(x(k))\rangle \\ &\quad - h\langle\nabla f(x(k)), \mathbb{E}\left[L(k)\right]\nabla f(x(k))\rangle \\ &= f(x(k)) + \frac{h^2\beta}{2}\langle\nabla f(x(k)), (p^2L^2 + 2p(1-p)L)\nabla f(x(k))\rangle \\ &\quad - h\langle\nabla f(x(k)), pL\nabla f(x(k))\rangle\end{aligned}$$

We use the upper bound $||L\nabla f(x(k))||^2 \leq \lambda_n(L)\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2$ and we obtain:

$$\mathbb{E}\left[f(x(k+1))|\mathcal{F}_k\right] \leq f(x(k)) - hp\left(1 - h\beta(1-p) - \frac{h\beta p\lambda_n(L)}{2}\right)\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2.$$

We denote $\xi = hp\left(1 - h\beta(1-p) - \frac{h\beta p\lambda_n(L)}{2}\right)$ and we compute the total expectation to get:

$$\mathbb{E}\left[\mathbb{E}\left[f(x(k+1))|\mathcal{F}_k\right]\right] \leq \mathbb{E}\left[f(x(k))\right] - \xi\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2\right]. \tag{3.23}$$

Since $h \leq \frac{2}{\beta(p\lambda_n(L)+2-2p)}$ we guarantee that $\mathbb{E}\left[f(x(k+1))\right] \leq \mathbb{E}\left[f(x(k))\right]$ such that:

$$\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2\right] \leq \frac{1}{\xi}\left(\mathbb{E}\left[f(x(k))\right] - \mathbb{E}\left[f(x(k+1))\right]\right).$$

If we take the sum over the time instants s we have:

$$\sum_{s=0}^{k-1}\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(s))\right|\right|^2\right] \leq \frac{1}{\xi}\left(f(x_0) - \mathbb{E}\left[f(x(k))\right]\right).$$

Notice that since $\mathbb{E}\left[f(x(k))\right]$ is monotonically decreasing and $\mathbb{E}\left[f(x(k))\right] \geq f(\bar{x}^*)$, where $\bar{x}^*$ is the minimizer of $f(x)$ without constraints, the limit $\lim_{k\to\infty}\mathbb{E}\left[f(x(k))\right]$ exists and we get:

$$\begin{aligned}\lim_{k\to\infty}\sum_{s=0}^{k-1}\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(s))\right|\right|^2\right] &\leq \frac{1}{\xi}\left(f(x_0) - \lim_{k\to\infty}\mathbb{E}\left[f(x(k))\right]\right) \\ &\leq \frac{1}{\xi}\left(f(x_0) - f^*\right),\end{aligned} \tag{3.24}$$

where we used the fact that $\mathbb{E}\left[f(x(k))\right] \geq f\left(\mathbb{E}\left[x(k)\right]\right)$. Finally, by using (3.24) in (3.22), we complete the proof. ■

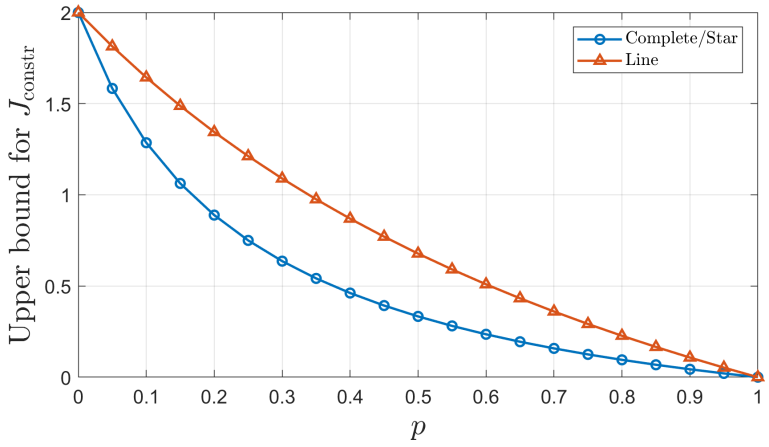


Figure 3.1: Computation of the upper bound for J_{constr} for a complete, star and line graph with $n = 10$, $\beta = 1$, $\epsilon = 0.5$ and $f(x_0) - f^* = 1$.

By inspecting the right-hand side of (3.21), we can make several observations. The upper bound is decreasing in p and

$$\lim_{p \to 1} J_{\text{constr}} = 0, \tag{3.25}$$

which is consistent with the fact that the constraint is always preserved with no communication losses. In the case of small values of p, which corresponds to a network exposed to many failures, if we consider $h = \frac{2\epsilon}{\beta(p\lambda_n(L)+2-2p)}$ with $\epsilon \in (0, 1)$ we obtain:

$$\lim_{p \to 0} J_{\text{constr}} \leq \frac{2\epsilon}{\beta(1-\epsilon)} \left(f(x_0) - f^*\right). \tag{3.26}$$

From (3.26), we can observe that large values of ϵ, corresponding to large values of the step-size, increase the potential violation of the constraint. Figure 3.1 presents the computation of the upper bound for the constraint violation metric in a complete, star and line graph with $n = 10$ agents, $\beta = 1$ and $\epsilon = 0.5$, where we can observe the behavior of the upper bound and the limit values determined by (3.25) and (3.26).

3.2.2 Error on the expected cost function

Due to the packet losses, $f(x(k))$ is a random process whose evolution depends on the different realizations of $L(k)$. When the algorithm (3.12) is executed, it is important to know the distribution of $f(x(k))$, to evaluate the impact of the perturbations on the performance of (3.12) and identify the additional cost corresponding to the deviation from the ideal cost f^*. From (3.17) we can observe

that this quantity is greater than f^*. Since the objective is to remain as close as possible to f^*, we define the cost function metric:

$$J_{\text{funct}} := \lim_{k\to\infty} \mathbb{E}\left[f(x(k))\right] - f(x^*), \tag{3.27}$$

which aims to measure this gap due to the packet losses during the application of the weighted gradient descent algorithm. It is important to notice that (3.27) corresponds to the limit of the so called *Jensen gap* [81].

Theorem 3.2 *Consider the weighted gradient descent given by* (3.10), (3.11) *and* (3.12). *Under Assumption 3.5, for any positive scalar* $h \leq \frac{2}{\beta(p\lambda_n(L)+2-2p)}$, *the cost function metric satisfies:*

$$J_{\text{funct}} \leq \left(1 - \frac{2 - h\beta\left(p\lambda_n(L) - 2p + 2\right)}{2 - h\alpha p\lambda_2(L)}\right)\left(f(x_0) - f^*\right).$$

Proof: From (3.23), we have that the total expectation of $f(x(k+1))$ satisfies:

$$\mathbb{E}\left[f(x(k+1))\right] \leq \mathbb{E}\left[f(x(k))\right] - \xi\mathbb{E}\left[\left|\left|L^{1/2}\nabla f(x(k))\right|\right|^2\right],$$

and by using Jensen's inequality and (3.15) we get:

$$\begin{aligned}\mathbb{E}\left[f(x(k+1))\right] &\leq \mathbb{E}\left[f(x(k))\right] - \xi\left|\left|\mathbb{E}\left[L^{1/2}\nabla f(x(k))\right]\right|\right|^2 \\ &= \mathbb{E}\left[f(x(k))\right] - \xi\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2.\end{aligned}$$

If we take the sum over the time instants s we obtain:

$$\mathbb{E}\left[f(x(k))\right] \leq f(x_0) - \xi\sum_{s=0}^{k-1}\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(s)\right])\right|\right|^2,$$

and the limit satisfies:

$$\lim_{k\to\infty}\mathbb{E}\left[f(x(k))\right] \leq f(x_0) - \xi\lim_{k\to\infty}\sum_{s=0}^{k-1}\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(s)\right])\right|\right|^2. \tag{3.28}$$

Now, we consider the dynamics in expectation given by (3.13), which satisfies the inequality of α-strongly convex functions:

$$\begin{aligned}f(\mathbb{E}\left[x(k+1)\right]) &\geq f(\mathbb{E}\left[x(k)\right]) + \langle\nabla f(\mathbb{E}\left[x(k)\right]), \mathbb{E}\left[x(k+1)\right] - \mathbb{E}\left[x(k)\right]\rangle \\ &\quad + \frac{\alpha}{2}\left|\left|\mathbb{E}\left[x(k+1)\right] - \mathbb{E}\left[x(k)\right]\right|\right|^2 \\ &= f(\mathbb{E}\left[x(k)\right]) + \langle\nabla f(\mathbb{E}\left[x(k)\right]), -hpL\nabla f(\mathbb{E}\left[x(k)\right])\rangle \\ &\quad + \frac{\alpha}{2}\left|\left|hpL\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 \\ &= f(\mathbb{E}\left[x(k)\right]) - hp\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 + \frac{h^2\alpha p^2}{2}\left|\left|L\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2\end{aligned}$$

Notice that due to the Min-Max Theorem, for any vector $z \in \mathbb{R}^n$, it holds

$$||Lz||^2 = \left|\left|L^{1/2}L^{1/2}z\right|\right|^2 \geq \lambda_2(L)\left|\left|L^{1/2}z\right|\right|^2, \tag{3.29}$$

since $L^{1/2}z$ is orthogonal to $\mathbb{1}$. Hence we obtain:

$$\begin{aligned} f(\mathbb{E}\left[x(k+1)\right]) &\geq f(\mathbb{E}\left[x(k)\right]) - hp\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 \\ &\quad + \frac{h^2\alpha p^2\lambda_2(L)}{2}\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 \\ &= f(\mathbb{E}\left[x(k)\right]) - hp\left(1 - \frac{h\alpha p\lambda_2(L)}{2}\right)\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 \end{aligned}$$

We denote $\delta = hp\left(1 - \frac{h\alpha p\lambda_2(L)}{2}\right)$, which is non-negative if $h \leq \frac{2}{\alpha\lambda_2(L)}$. Since $\frac{2}{\beta(p\lambda_n(L)+2-2p)} \leq \frac{2}{\alpha\lambda_2(L)}$ we get:

$$\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(k)\right])\right|\right|^2 \geq \frac{1}{\delta}\left(f(\mathbb{E}\left[x(k)\right]) - f(\mathbb{E}\left[x(k+1)\right])\right).$$

If we take the sum over the time instants s we obtain:

$$\sum_{s=0}^{k-1}\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(s)\right])\right|\right|^2 \geq \frac{1}{\delta}\left(f(x_0) - f(\mathbb{E}\left[x(k)\right])\right),$$

and for $p > 0$ the limit satisfies:

$$\lim_{k\to\infty}\sum_{s=0}^{k-1}\left|\left|L^{1/2}\nabla f(\mathbb{E}\left[x(s)\right])\right|\right|^2 \geq \frac{1}{\delta}\left(f(x_0) - f^*\right). \tag{3.30}$$

We use (3.30) in (3.28) to get:

$$\lim_{k\to\infty}\mathbb{E}\left[f(x(k))\right] \leq f(x_0) - \frac{\xi}{\delta}\left(f(x_0) - f^*\right),$$

which yields

$$\lim_{k\to\infty}\mathbb{E}\left[f(x(k))\right] - f(x^*) \leq \left(1 - \frac{\xi}{\delta}\right)\left(f(x_0) - f^*\right).$$

■

To analyze the asymptotic behavior of the upper bound for J_{funct}, let us consider a step-size given by $h = \frac{2\epsilon}{\beta(p\lambda_n(L)+2-2p)}$ with $\epsilon \in (0,1)$. For low values of probabilities corresponding to non-robust networks we have:

$$\lim_{p\to 0} J_{\text{funct}} \leq \epsilon\left(f(x_0) - f^*\right), \tag{3.31}$$

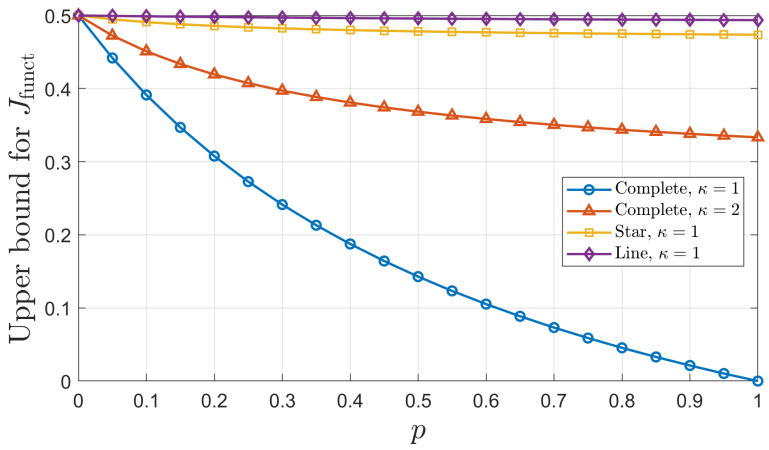


Figure 3.2: Computations of the upper bound for J_{funct} for a complete, star and line graph with $n = 10$, $\epsilon = 0.5$, $f(x_0) - f^* = 1$ and two values of κ.

which implies that small step-sizes, characterized by small values of ϵ, reduce more the gap between the expected value of the cost function and f^*. In the case of large values of probabilities, we have the following asymptotic behavior:

$$\lim_{p \to 1} J_{\text{funct}} \leq \frac{\epsilon(1 - \kappa^{-1}\kappa_\lambda^{-1})}{1 - \epsilon\kappa^{-1}\kappa_\lambda^{-1}} \left(f(x_0) - f^*\right), \tag{3.32}$$

where $\kappa = \frac{\beta}{\alpha}$ is the condition number of the cost functions f_i and $\kappa_\lambda = \frac{\lambda_n(L)}{\lambda_2(L)}$. We can observe that only for the particular case of a complete graph (i.e., $\lambda_2(L) = \lambda_n(L)$) and the particular value $\kappa = 1$ (i.e., $\alpha = \beta$) we obtain

$$\lim_{p \to 1} J_{\text{funct}} = 0, \tag{3.33}$$

which implies a certain level of conservatism in the upper bound since (3.33) should hold for any choice of κ and κ_λ in order to recover the behavior of a network without packet losses. Figure 3.2 shows the computation of the upper bound for J_{funct} for a complete, star and line graph with $n = 10$ agents, $\epsilon = 0.5$, $f(x_0) - f^* = 1$ and two values of κ where we can appreciate the behavior of the upper bound and the limit values defined by (3.31) and (3.32).

3.3 Packet losses in OMAS

In this subsection, we consider the case when *replacements of agents* happen in the system, making it *open*. Each agent $i \in V$ gets replaced at random time instants,

resulting in the change of its local cost function, and hence of the global minimizer of the system x^*.

For the sake of simplicity, we assume that when an agent i is replaced, the joining agent that takes its place retrieves its label i and its estimate x_i, so that the constraint $\mathbb{1}^T x = b$ is preserved, but receives a new local cost function satisfying Assumptions 3.5. Denoting f_i^k the local cost function held by agent i at the time instant k, we can then reformulate (3.1) as

$$\min_{x \in \mathcal{S}_{\mathbb{1},b}} f^k(x) := \sum_{i=1}^{n} f_i^k(x_i). \tag{3.34}$$

The solution of (3.34) can thus differ from a time instant t_k to another, and we denote it by $x^*(k) := \arg\min_{x \in \mathcal{S}_{\mathbb{1},b}} f^k(x)$.

3.3.1 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 interactions of agents according to the network topology or replacements. We call such occurrences *events*, and we define the *event set* of the system, which contains all events that can happen in the system, as

$$\Xi = U \cup R, \tag{3.35}$$

where U is the update event corresponding to the application of the weighted gradient descent, and $R := \bigcup_{i \in V} R_i$ is the set of all events R_i, i.e., the replacement of a single agent i in the system.

We assume that two distinct events never happen simultaneously, and following the above characterization, we consider a discrete evolution of the time where each time-step k corresponds to the random time instant at which the k-th event of the system $\xi_k \in \Xi$ takes place. Moreover, we consider the following assumption that guarantees that replacements and interactions are independent processes, so that at any iteration k, the event ξ_k is an 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 3.6 *An event ξ_k is independent of any other event ξ_j, with $j \neq k$, and of any information in the system prior to time-step k, such as the estimates or local cost functions.*

More precisely, we assume that at each time instant k we can have an update with probability p_U or a replacement with probability $p_R = 1 - p_U$. If an update occurs, the weighted gradient descent is applied and $f^{k+1} = f^k$. If a replacement occurs, a single agent i is uniformly randomly selected and receives a new (piecewise quadratic) cost function.

3.3.2 Constraint violation

When the open system does not suffer from packet losses, the constraint is always preserved but the global minimizer $x^*(k)$ is time-varying since the cost functions change in time. In this analysis, we assume that the open system also suffers of packet losses such that the weighted gradient descent algorithm becomes:

$$x(k+1) = x(k) - hL(k)\nabla f^k(x(k)). \tag{3.36}$$

Similarly to the closed system we analyze the constraint violation metric J_{constr} defined in (3.18).

Proposition 3.2 *Consider the weighted gradient descent given by* (3.10), (3.11) *and* (3.36). *Under Assumption 3.5, for any positive scalar* $h \leq \frac{2}{\beta(p\lambda_n(L)+2-2p)}$, *the constraint violation metric satisfies:*

$$J_{\text{constr}} \geq \frac{4p_U h(1-p)}{2-\alpha h(p\lambda_2(L)+2-2p)} \lim_{k\to\infty} \sum_{s=0}^{k-1} \mathbb{E}\left[f^s(x(s)) - f^s(x(s+1))\right].$$

Proof: Let us recall the notation in the proof of Theorem 3.1 for the error $H(k) = \left(\mathbb{1}^\top x(k) - b\right)$ and the natural filtration $\{\mathcal{F}_k\}_{k\in\mathbb{Z}_{\geq 0}}$. From (3.22) we have that when an update U happens

$$\mathbb{E}\left[H^2(k+1)|U\right] = \mathbb{E}\left[H^2(k)\right] + 2h^2p(1-p)\mathbb{E}\left[\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2\right]. \tag{3.37}$$

When a replacement R occurs, the constraint is not violated at that event

$$\mathbb{E}\left[H^2(k+1)|R\right] = \mathbb{E}\left[H^2(k)\right].$$

Then, we compute the total expectation and we obtain:

$$\begin{aligned}\mathbb{E}\left[H^2(k+1)\right] &= p_U\mathbb{E}\left[H^2(k+1)|U\right] + p_R\mathbb{E}\left[H^2(k+1)|R\right] \\ &= \mathbb{E}\left[H^2(k)\right] + 2p_U h^2 p(1-p)\mathbb{E}\left[\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2\right].\end{aligned}$$

We use a recursive argument to get:

$$\mathbb{E}\left[H^2(k)\right] = 2p_U h^2 p(1-p)\sum_{s=0}^{k-1}\mathbb{E}\left[\left|\left|L^{1/2}\nabla f^s(x(s))\right|\right|^2\right]. \tag{3.38}$$

Since the cost functions f^k are α-strongly convex we have:

$$\begin{aligned}f^k(x(k+1)) &\geq f^k(x(k)) + \frac{\alpha}{2}\left|\left|x(k)-x(k+1)\right|\right|^2 + \langle\nabla f^k(x(k)), x(k+1)-x(k)\rangle \\ &= f^k(x(k)) + \frac{h^2\alpha}{2}\langle\nabla f^k(x(k)), L(k)^T L(k)\nabla f^k(x(k))\rangle \\ &\quad - h\langle\nabla f^k(x(k)), L(k)\nabla f^k(x(k))\rangle.\end{aligned}$$

We compute the expectation given the filtration generated by $x(k)$ and we use (3.19) to get:

$$\mathbb{E}\left[f^k(x(k+1))|\mathcal{F}_k\right] \geq f^k(x(k)) + \frac{h^2\alpha}{2}\langle \nabla f^k(x(k)), (p^2L^2 + 2p(1-p)L)\nabla f^k(x(k))\rangle \\ - h\langle \nabla f^k(x(k)), pL\nabla f^k(x(k))\rangle.$$

By using (3.29), we have the lower bound $\left|\left|L\nabla f^k(x(k))\right|\right|^2 \geq \lambda_2(L)\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2$ and we obtain:

$$\mathbb{E}\left[f^k(x(k+1))|\mathcal{F}_k\right] \geq f^k(x(k)) - hp\left(1 - h\alpha(1-p) - \frac{h\alpha p\lambda_2(L)}{2}\right)\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2.$$

We denote $\theta = hp\left(1 - h\alpha(1-p) - \frac{h\alpha p\lambda_2(L)}{2}\right)$ and we compute the total expectation to get:

$$\mathbb{E}\left[\mathbb{E}\left[f^k(x(k+1))|\mathcal{F}_k\right]\right] \geq \mathbb{E}\left[f^k(x(k))\right] - \theta\mathbb{E}\left[\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2\right].$$

Since $h \leq \frac{2}{\beta(p\lambda_n(L)+2-2p)} \leq \frac{2}{\alpha(p\lambda_2(L)+2-2p)}$ we guarantee that $\mathbb{E}\left[f^k(x(k+1))\right] \leq \mathbb{E}\left[f^k(x(k))\right]$ such that:

$$\mathbb{E}\left[\left|\left|L^{1/2}\nabla f^k(x(k))\right|\right|^2\right] \geq \frac{1}{\theta}\left(\mathbb{E}\left[f^k(x(k))\right] - \mathbb{E}\left[f^k(x(k+1))\right]\right). \tag{3.39}$$

By using (3.39) in (3.38) we complete the proof. ■

Unlike a closed system, we observe that in general $\mathbb{E}\left[f^s(x(s+1))\right] \neq \mathbb{E}\left[f^{s+1}(x(s+1))\right]$ due to the possible replacements that take place in the system. This mismatch implies that terms are not canceled in the series in the lower bound and the error accumulates in time, so that J_{constr} grows to infinity. This intuition is corroborated by the simulations of $\mathbb{E}\left[(\mathbb{1}^T x(k) - b)^2\right]$ for a complete graph with $n = 7$ agents that are shown in Fig. 3.3. We can observe in the plot that both the constraint violation metric and the lower bound diverge. During the transient state, the accumulation of errors is considerable and the increase of the constraint violation metric is significant, while in the steady state the accumulation of errors is small and the metric increases with a lower rate. This difference in the growth rates can be explained by the distance between the initial state $x(0)$ and the minimizer x^*. During the transient state, agents exchange large values to reach the minimizer, such that a loss of information has a high impact in the system. Nevertheless, in the steady state, the states of the agents are close to the minimizer and the values exchanged are small, which implies that the potential loss of information is not significant.

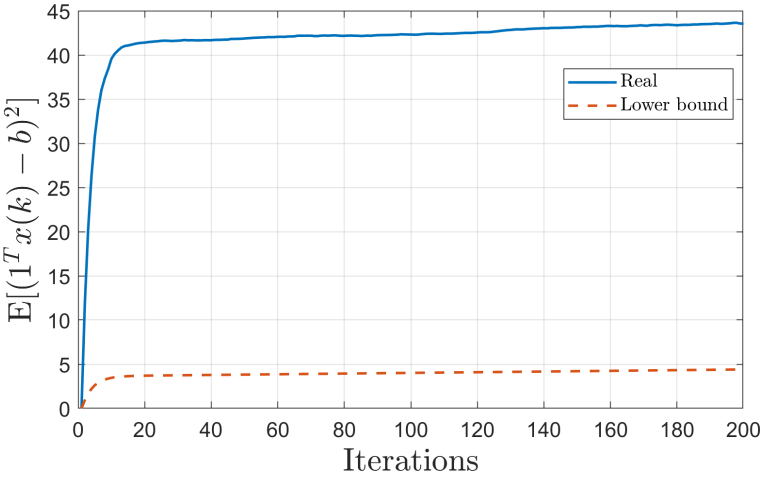


Figure 3.3: Simulation of $\mathbb{E}\left[(\mathbb{1}^T x(k) - b)^2\right]$ and of the lower bound in Proposition 3.2 for a complete graph under replacements with $n = 7$, $\alpha = 1$, $\beta = 10$, $p = 0.8$, $p_U = 0.5$ by considering 10000 realizations of the process.

When replacements do not take place, the lower bound of Proposition 3.2 is given by:

$$J_{\text{constr}} \geq \frac{4h(1-p)}{2 - \alpha h(p\lambda_2(L) + 2 - 2p)} \left(f(x(0)) - \lim_{k\to\infty} \mathbb{E}\left[f(x(k))\right] \right).$$

In Fig. 3.4 we present the behavior of J_{constr} and the lower bound of Proposition 3.36 for a system without replacements considering the same parameters of Fig. 3.3 where we observe a similar behavior for the transient state characterized by a fast increase. However, unlike the case of replacements, during the steady state, J_{constr} converges to a value and remains bounded, which is also guaranteed by Theorem 3.1.

When $p = 1$, corresponding to an ideal system without packet losses, even if replacements take place in the system (i.e., $p_U > 0$), J_{constr} is zero since $\mathbb{E}\left[H^2(k+1)|U\right] = \mathbb{E}\left[H^2(k)\right]$ as we can see in (3.37). This can also be derived from (3.14), since the algorithm preserves the constraint independently of the cost functions considered for the computation of $x(k+1)$.

3.4 Conclusion

In this chapter, we analyzed the effect of packet losses in the performance of the weighted gradient descent algorithm. We defined two performance metrics to measure the deviation from the constraint and the error of the expected cost

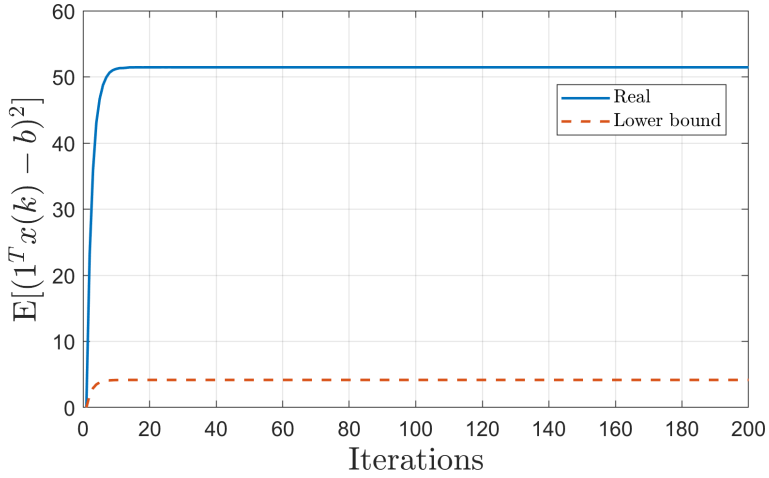


Figure 3.4: Simulation of $\mathbb{E}\left[(\mathbb{1}^T x(k) - b)^2\right]$ and of the lower bound in Proposition 3.2 for a complete graph with no replacements with $n = 7$, $\alpha = 1$, $\beta = 10$, $p = 0.8$, $p_U = 1$ by considering 10000 realizations of the process.

function and we derived an upper bound. We extended the analysis of packet losses to OMAS and we showed that the constraint violation metric may diverge in such a case due to the accumulation of errors during the replacements.

We showed that when the weighted gradient descent algorithm is applied in a system subject only to packet losses, the constraint violation metric remains bounded while in a system subject only to replacements, the constraint violation metric is zero since the constraint is preserved during replacements. However, when updates using the weighted gradient descent algorithm with packet losses and replacements occur in the same system, the constraint violation metric diverges, showing the fragility of the algorithm in OMAS. This result presents a similarity with the case of Lévy flights studied in [57], where an asymmetric loss of information together with delays and noise can generate complex behaviors.

As for future work, natural continuations of the paper in closed and open systems can be explored. In the closed case, although an upper bound was derived for the cost function metric, the expression presents a certain level of conservatism. It would be important to improve the bound to obtain a tighter result that guarantees a zero value of the metric when $p = 1$ (no replacements). Regarding the open case, it is necessary to improve the lower bound for the constraint violation metric so that it does not depend on the trajectories but on the properties of the system. Also, the error of the expected cost function should be studied in OMAS to determine if it is bounded or if it could diverge.

Finally, regarding the objective of achieving an optimal resource allocation even

under the perturbations generated by packet losses and replacements of agents, the design of more complex algorithms to obtain better performances remains open. One of the most interesting approaches is online optimization, where the problem of time-varying functions is also analyzed from a different perspective and algorithms based on several stages are proposed with the aim of achieving a good performance in average [82, 83].

Chapter 4

Random coordinate descent with replacements

In OMAS, a fixed solution for the resource allocation problem (3.1) cannot be obtained as in general, the size of the system is not fixed and the cost functions keep changing, such that the goal of the agents is to track the time-varying solution of (3.1) as well as possible at all times using an optimization algorithm. However, 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. 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 [84]. 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 [80], 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 the context of optimization, problems such as (3.1) typically imply a minimization process on a long period, and thus in many situations, the cost is expected to be paid on a regular basis. In such a setting, a natural way of measuring the performance of an algorithm is to compute its accumulated error with respect to a given strategy over a finite number of iterations. Similar metrics occur in the context of online optimization [85], where the objective is to minimize the so-called

regret, commonly defined as the accumulated error of the estimate $x(k)$ with respect to $x^* := \arg\min_x \sum_{k=1}^{T} f^k(x)$, or sometimes with respect to the time-varying solution of (3.1) $x^*(k) := \arg\min_x f^k(x)$ such as e.g., in [86].

In this chapter, we study the Random Coordinate Descent algorithm (RCD) to solve the resource allocation (RA) problem in OMAS where agents get replaced during the process, relying on a decoupled analysis of the RCD algorithm and of replacements of single agents. In the first part, we analyze the RCD algorithm using a time-varying optimization framework [87], where we focus in the convergence of the algorithm towards the minimizer and its stability. In Section 4.1 we introduce a more general setting for the RA problem where agents can be heterogeneous (i.e., that a_i may be different) and hold d-dimensional local cost functions. In addition, we consider an arbitrary distribution for the probabilities associated with the choice of edges during the execution of the RCD algorithm. Section 4.2 presents an analysis of the behavior of the minimizer during a replacement. We derive three different upper bounds for the difference between two possible minimizers using different techniques. In Section 4.3 we prove the linear convergence of the RCD algorithm in a closed system in a norm induced by a matrix associated with the network. In Section 4.4 we analyze the RCD algorithm in an open system subject to replacements and we provide conditions for the convergence of the algorithm inside a ball, whose size depends on parameters of the system. In the second part, we analyze the RCD algorithm using tools similar to those used in Online Optimization. Section 4.5 presents a simplified setting, where homogeneous agents, holding 1-dimensional cost functions, interact in a complete graph. We introduce the performance metrics used in this work, which are inspired by the loss accumulated by the RCD algorithm with respect to the time-varying optimal solution $x^*(k)$ over a finite number of iterations, and its gain with respect to the *selfish strategy* $x^s(k)$, which consists in the absence of collaboration between the agents (i.e., $x_i^s(k) = d_i^k$). We derive upper bounds for the performance metrics and show that the metrics scale linearly with time. In addition, the case of quadratic cost functions is analyzed, for which tighter results are obtained. Finally, the conclusions and future work are exposed in Section 4.6.

4.1 Problem statement

4.1.1 Network description and open system

We consider the resource allocation problem (3.1) with a fixed budget b:

$$\min_{x=[x_1^T,\dots,x_n^T]^T \in \mathbb{R}^{nd}} f(x) = \sum_{i=1}^{n} f_i(x_i) \qquad \text{subject to} \qquad \sum_{i=1}^{n} a_i x_i = b, \tag{4.1}$$

where the local cost functions $f_i : \mathbb{R}^d \to \mathbb{R}$ satisfy Assumption 3.1 (i.e., f_i is continuously differentiable, α-strongly convex and β-smooth).

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

Moreover, we consider that *replacements of agents* happen in the system, making it *open*. Each agent $i \in V$ gets replaced at random time instants, resulting in the change of its local cost function, and hence of the global minimizer x^*.

The evolution of the open network presented in this section is characterized by the instantaneous occurrence at random time instants of either pairwise interactions or replacements. We call such occurrences *events*, and we define the *event set* of the system, which contains all events that can happen in the system, as

$$\Xi = U \cup R, \tag{4.2}$$

where $U = \bigcup_{(i,j)\in E} U_{ij}$ is the set of all possible events U_{ij}, i.e., the pairwise interaction between two connected agents i and j, and $R := \bigcup_{i\in V} R_i$ is the set of all events R_i, i.e., the replacement of a single agent i in the system.

Similarly to Section 3.3.1, we consider a discrete evolution of the time where each time-step k corresponds to the random time instant at which the k-th event of the system $\xi_k \in \Xi$ takes place. Moreover, we consider that the events ξ_k satisfy the Assumption 3.6 that guarantees that replacements and interactions are independent processes, so that at any iteration k, the event ξ_k is an interaction, i.e., $\xi_k \in U$, with fixed probability p_U, and a replacement, i.e., $\xi_k \in R$, with fixed probability $p_R = 1 - p_U$.

Following the approach in [88], we restrict the location of the minimizers of the local functions:

Assumption 4.1 *For all $i \in V$, the minimizer of f_i denoted as $\bar{x}_i^* := \arg\min_x f_i(x)$ satisfies $\bar{x}_i^* \in B(\mathbf{0}_d, c)$ for some $c > 0$. Moreover, without loss of generality $f_i(\bar{x}_i^*) = 0$ for all $i \in V$.*

Assumption 4.1 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.

In the first part of this chapter, 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 $\left(a^\top \otimes I_d\right) x = b$ is preserved, but receives a new local cost function satisfying Assumptions 3.1 and 4.1 [1]. This is consistent with e.g., the case of energy distribution where the sources maintain the same value of generated energy but with a different cost function. Denoting f_i^k the local cost function held by agent i at the time instant k, we can then reformulate the RA problem (4.1) as

$$\min_{x \in \mathcal{S}_{a,b}} f^k(x) := \sum_{i=1}^{n} f_i^k(x_i), \tag{4.3}$$

where $\mathcal{S}_{a,b} = \left\{x \in \mathbb{R}^{nd} | \left(a^\top \otimes I_d\right) x = b\right\}$ is the feasible set. The solution of (4.3) 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.

4.1.2 Random Coordinate Descent (RCD) algorithm

To compute the solution of (4.3) we consider the Random Coordinate Descent (RCD) algorithm introduced in [80]. This algorithm involves the update of the states of only a pair of neighbouring agents at each iteration, so that the computational complexity is cheap. Hence, whenever two agents i and j interact (i.e., in the event U_{ij}), they perform an *RCD update*, which is defined as follows for some nonnegative step-size $h \geq 0$:

$$x(k+1) = x(k) - h\mathbf{Q}^{ij} \nabla f(x(k)), \tag{4.4}$$

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:

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

With the update rule (4.4) only agents i and j update their estimates while all the other agents keep it the same. For agents i and j, (4.4) 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 $x_i^+ = x_i -$

[1] This assumption about the preservation of the value of x_i by the new agent will be relaxed in Section 4.5 corresponding to the second part of this chapter where tools inspired by online optimization are used for the analysis.

$\frac{h}{2}(\nabla f_i(x_i) - \nabla f_j(x_j))$, so that the update follows both gradients with equal weight while preserving the constraint.

Remark 4.1 *The update rule* (4.4) *can be formally obtained by solving the following optimization problem, which corresponds to the interpretation given above (we refer to [80] for details):*

$$\arg \min_{s_i, s_j \in \mathbb{R}^d : a_i s_i + a_j s_j = 0} \left\langle \begin{bmatrix} \nabla f_i(x_i) \\ \nabla f_j(x_j) \end{bmatrix}, \begin{bmatrix} s_i \\ s_j \end{bmatrix} \right\rangle + \frac{\beta}{2} \left\| \begin{bmatrix} s_i \\ s_j \end{bmatrix} \right\|^2. \tag{4.5}$$

Based on the approach of [80], one can then show that the optimal step-size that solves (4.5) *is given by* $h = 1/\beta$.

We also introduce the following matrix that builds on the definition of the update rule (4.4):

$$\mathbf{L}_p = \sum_{(i,j)\in E} p_{ij} \mathbf{Q}^{ij} = \left(\sum_{(i,j)\in E} p_{ij} Q^{ij} \right) \otimes I_d = L_p \otimes I_d. \tag{4.6}$$

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

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

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, one has

$$L_p a = Q^{ij} a = \mathbf{0}_n, \tag{4.7}$$

which means that zero is an eigenvalue of both Q^{ij} and L_p with corresponding eigenvector a. We denote by $\lambda_2(L_p)$ and $\lambda_n(L_p)$ respectively the second smallest and the largest eigenvalues of L_p. One can show that $\lambda_2(L_p) > 0$ since the graph G is connected (we refer to Lemma 3.3 of [80] for a detailed proof).

Remark 4.2 *For a graph* $G = (V, E)$*, when* $a = \mathbb{1}_n$ *(homogeneous agents) and the probabilities* p_{ij} *are uniformly distributed, one has* $L_p = \frac{1}{2|E|} L$*, 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.*

4.2 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 4.1 corresponds to an analysis of the location of the global minimizer of the system, and Lemma 4.2 is an analysis of the location of the minimizer held by each individual agent.

Lemma 4.1 *Let $x^* := \arg\min_{x\in\mathcal{S}_{a,b}} \sum_{i=1}^n f_i(x_i)$. If all f_i satisfy Assumptions 3.1 and 4.1, then $x^* \in B(\mathbf{0}_{nd}, R_{b,\kappa})$ with*

$$R_{b,\kappa} = \sqrt{n\kappa}\left(c + \frac{c}{\sqrt{\kappa}} + \frac{||b||}{\sqrt{n}\,||a||}\right), \tag{4.8}$$

where c was defined in Assumption 4.1.

Proof: We prove by contradiction. 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 $||\bar{x}^*|| \leq \sqrt{n}c$ since $\bar{x}^* \in B(\mathbf{0}_d, c)^n$, and it thus follows that $||x - \bar{x}^*|| > R_{b,\kappa} - \sqrt{n}c$. Hence, since f is α-strongly convex, and reminding that $f(\bar{x}^*) = 0$ from Assumption 4.1, there holds

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

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 4.1, there holds

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

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

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

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

Lemma 4.2 *Let $x^* := \arg\min_{x\in\mathcal{S}_{a,b}} f(x) = \sum_{i=1}^n f_i(x_i)$. If f_i satisfies Assumptions 3.1 and 4.1 for all $i = 1, \ldots, n$, then for ζ^* defined in* (3.3) *there holds*

$$||\zeta^*|| \leq \frac{\beta}{||a||^2}\left(||b|| + c\,||a||_1\right); \tag{4.9}$$

and

$$||x_i^*|| \leq \frac{a_i}{||a||^2}\kappa\left(||b|| + c\,||a||_1\right) + c. \tag{4.10}$$

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

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

Then, from the optimality condition (3.3), there holds

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

By summing over all the i, we obtain:

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

We use the Cauchy-Schwarz inequality to get:

$$||a||^2 \, ||\zeta^*||^2 \leq \beta \, ||\zeta^*|| \left|\left| \sum_{i=1}^{n} a_i \, (x_i^* - \bar{x}_i^*) \right|\right|,$$

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

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

which corresponds to (4.9). 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

$$||x_i^*|| \leq \frac{1}{\alpha} \, ||a_i \zeta^*|| + ||\bar{x}_i^*|| \leq \frac{a_i}{\alpha} \, ||\zeta^*|| + c. \tag{4.11}$$

Equation (4.10) then follows from combining (4.9) and (4.11). ■

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 $f_1, f_2, \ldots, f_{n-1}, f_n^{(1)}, f_n^{(2)}$ satisfying Assumptions 3.1 and 4.1 we define the minimizer before a replacement $x^{(1)}$, and after a replacement $x^{(2)}$ as

$$\begin{aligned} x^{(1)} &:= \underset{x \in \mathcal{S}_{a,b}}{\arg\min} \left(\sum\nolimits_{i=1}^{n-1} f_i(x_i) + f_n^{(1)}(x_n) \right); \\ x^{(2)} &:= \underset{x \in \mathcal{S}_{a,b}}{\arg\min} \left(\sum\nolimits_{i=1}^{n-1} f_i(x_i) + f_n^{(2)}(x_n) \right). \end{aligned} \tag{4.12}$$

Proposition 4.1 *Consider $x^{(1)}$ and $x^{(2)}$ as defined in (4.12), 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 there holds*

$$\left|\left|x^{(1)} - x^{(2)}\right|\right|^2 \leq \min\{\psi_{n,\kappa}, \chi_{n,\kappa}, \theta_{n,\kappa}\} =: \bar{M}_{n,\kappa}^2, \tag{4.13}$$

with

$$\psi_{n,\kappa} = 4n\kappa\left(c + \frac{c}{\sqrt{\kappa}} + \frac{||b||}{\sqrt{n}\,||a||}\right)^2; \tag{4.14}$$

$$\chi_{n,\kappa} = 8\left(\frac{a_+}{||a||^2}\kappa(||b|| + c\,||a||_1) + c\right)^2; \tag{4.15}$$

$$\theta_{n,\kappa} = 4\left(1 + \frac{(\kappa+1)^2}{4\kappa}\rho_a\right)\left(\frac{a_+}{||a||^2}\kappa(||b|| + c\,||a||_1) + c\right)^2, \tag{4.16}$$

Proof:

a) Proof of $\psi_{n,\kappa}$: Observe that $x^{(1)}, x^{(2)} \in B(\mathbf{0}_{nd}, R_{b,\kappa})$ from Lemma 4.1, so that there holds

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

which yields (4.14).

b) Proof of $\chi_{n,\kappa}$: We remind that for $i = 1, \ldots, n-1$, there holds $\nabla f_i(x_i^{(q)}) = a_i\zeta^{(q)}$, with $q = 1, 2$. From α-strongly convexity of the local cost functions, there holds that for all $i = 1, \ldots, n-1$:

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

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 $i = 1, \ldots, n-1$. 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 $i = 1, \ldots, n-1$ yields

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

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

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

Then, we use (4.9) and (4.10) to get

$$\left|\left|y^{(1)} - y^{(2)}\right|\right|^2 \leq 4\left(\frac{a_n}{||a||^2}\kappa\left(||b|| + c\,||a||_1\right) + c\right). \tag{4.17}$$

By definition we have

$$\left|\left|x^{(1)} - x^{(2)}\right|\right|^2 = \left|\left|y^{(1)} - y^{(2)}\right|\right|^2 + \left|\left|x_n^{(1)} - x_n^{(2)}\right|\right|^2. \tag{4.18}$$

We apply triangle inequality and (4.10) to obtain

$$\begin{aligned}\left|\left|x^{(1)} - x^{(2)}\right|\right|^2 &\leq \left|\left|y^{(1)} - y^{(2)}\right|\right|^2 + \left(\left|\left|x_n^{(1)}\right|\right| + \left|\left|x_n^{(2)}\right|\right|\right)^2 \\ &\leq \left|\left|y^{(1)} - y^{(2)}\right|\right|^2 + 2\left(\left|\left|x_n^{(1)}\right|\right|^2 + \left|\left|x_n^{(2)}\right|\right|^2\right) \\ &\leq \left|\left|y^{(1)} - y^{(2)}\right|\right|^2 + 4\left(\frac{a_n}{||a||^2}\kappa\left(||b|| + c\,||a||_1\right) + c\right) \end{aligned} \tag{4.19}$$

Finally, the result (4.15) yields by combining (4.17) and (4.19) and using the fact that $a_n \leq a_+$.

c) Proof of $\theta_{n,\kappa}$: Since the functions are α-strongly convex and β-smooth, there holds that for all $i = 1, \ldots, n-1$:

$$a_i(1+\kappa^{-1})\langle \zeta^{(1)} - \zeta^{(2)}, x_i^{(1)} - x_i^{(2)}\rangle \geq \beta^{-1}a_i^2\left|\left|\zeta^{(1)} - \zeta^{(2)}\right|\right|^2 + \alpha\left|\left|x_i^{(1)} - x_i^{(2)}\right|\right|^2.$$

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

$$a_n(1+\kappa^{-1})\langle \zeta^{(1)} - \zeta^{(2)}, x_n^{(2)} - x_n^{(1)}\rangle \geq m\beta^{-1}\left|\left|\zeta^{(1)} - \zeta^{(2)}\right|\right|^2 + \alpha\left|\left|y^{(1)} - y^{(2)}\right|\right|^2,$$

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

$$a_n(1+\kappa^{-1})\left|\left|\zeta^{(1)} - \zeta^{(2)}\right|\right|\left|\left|x_n^{(2)} - x_n^{(1)}\right|\right| \geq m\beta^{-1}\left|\left|\zeta^{(1)} - \zeta^{(2)}\right|\right|^2 + \alpha\left|\left|y^{(1)} - y^{(2)}\right|\right|^2.$$

This can be written as follows

$$\alpha\left|\left|y^{(1)} - y^{(2)}\right|\right|^2 \leq \phi(\left|\left|\zeta^{(1)} - \zeta^{(2)}\right|\right|), \tag{4.20}$$

where

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

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

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

and it follows by using (4.20) that

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

Equation (4.16) then follows from (4.18) using the fact that $x_n^{(1)}$ and $x_n^{(2)}$ are upper bounded by (4.10), which thus yields $\theta_{n,\kappa}$, and the conclusion follows. ■

The bound $\bar{M}_{n,\kappa}^2$ from Proposition 4.1 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

$$\psi_{n,\kappa} \leq 4n\kappa\left(2c+\frac{||b||}{n\bar{a}}\right)^2 = O(n\kappa);$$
$$\chi_{n,\kappa} \leq 8\left(\frac{a_+^2}{\overline{a^2}}\left(\frac{||b||}{n\bar{a}}+2c\right)\kappa\right)^2 = O(\kappa^2);$$
$$\theta_{n,\kappa} \leq 4\left(\frac{a_+^2}{a_-^2}\left(\frac{\kappa}{2(n-1)}+2\right)\right)\left(\frac{a_+^2}{\overline{a^2}}\left(\frac{||b||}{n\bar{a}}+2c\right)\kappa\right)^2 = O\left(\kappa^2+\frac{\kappa^3}{n}\right),$$

where $O(\cdot)$ corresponds to the O-notation [2]. The linear scaling of $\psi_{n,\kappa}$ in both n

[2]For a given function $g(n)$, we denote by $O(g(n))$ the set of functions [60]

$$O(g(n)) = \{f(n) : \text{there exist positive constants } c \text{ and } n_0 \text{ such that } 0 \leq f(n) \leq cg(n) \text{ for all } n \geq n_0\}.$$

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. 4.1.

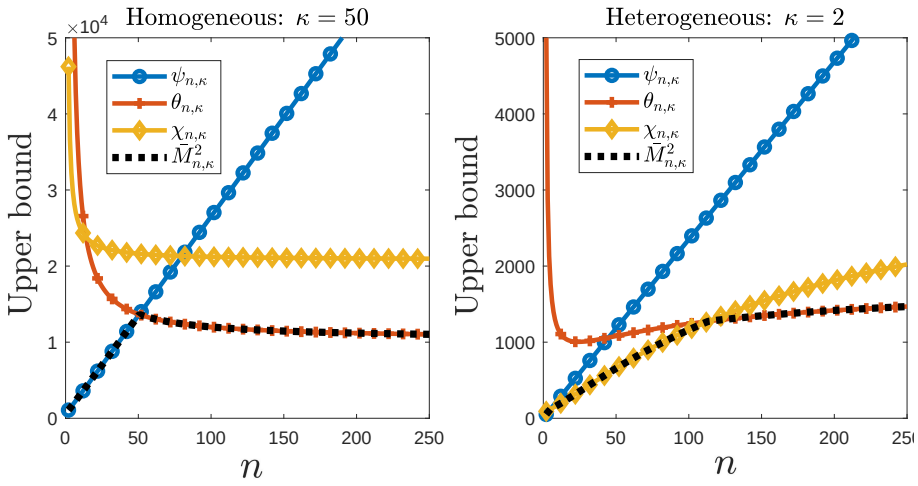


Figure 4.1: Bound of Proposition 4.1 with respect to the system size n for $b = 1$, $c = 1$, respectively for $\kappa = 50$ with homogeneous agents ($a_i = 1$ for all i) on the left, and $\kappa = 2$ with heterogeneous agents ($a_1 = 10$, $a_i = 1$ for $i > 1$) on the right. The figures show all three quantities $\psi_{n,\kappa}$, $\theta_{n,\kappa}$ and $\chi_{n,\kappa}$ as well as the final bound $\bar{M}^2_{n,\kappa}$ for both cases.

Remark 4.3 *The interpretation of the quantities $\psi_{n,\kappa}$, $\chi_{n,\kappa}$ and $\theta_{n,\kappa}$ actually strongly 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 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 yields the same scalings than those presented above).*

The result of Proposition 4.1 can be analyzed with respect to empirical results derived with the PESTO toolbox [89], which allows computing exact empirical

bounds for questions related to convex functions [3]. 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 e.g., using heterogeneous agents.

We can observe in Fig. 4.2, 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 $\bar{M}^2_{n,\kappa} = \theta_{n,\kappa}$ when n becomes large, resulting in the convergence of $\bar{M}^2_{n,\kappa}$ towards a constant, consistently with the result obtained with PESTO. In parallel, the bottom plot shows that the bounds from PESTO scale linearly with κ, consistently with the evolution of $\psi_{n,\kappa}$, which is the value taken by $\bar{M}^2_{n,\kappa}$ for large values of κ.

4.3 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 convergence of the RCD Algorithm in closed system.

4.3.1 Linear convergence and $\mathbf{L}^{\dagger}_{p}$-seminorm

In this section, we derive the convergence rate 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 [90].

Definition 4.1 (Linear Convergence) *Let $\{x(k)\}$ be the sequence of points converging to some point $x^* \in \mathbb{R}^d$ generated by some algorithm. We say the convergence is* 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 rate of convergence.

Definition 4.2 (Exponential Convergence) *Let $\{x(k)\}$ be the sequence of points converging to some point $x^* \in \mathbb{R}^d$ generated by some algorithm. We say*

[3] We consider the setting of Proposition 4.1 with $\arg\min_x f_i(x) \in B(0,1)$. We use PESTO to evaluate $\max \left|\left|x^{(2)} - x^{(1)}\right|\right|^2$, where $x^{(1)}$ and $x^{(2)}$ are defined as in (4.12) and where we impose $\sum_i x_i^{(1)} = \sum_i x_i^{(2)} = v_b$, for some vector v_b satisfying $||v_b|| = b$. The variables of the problem are the functions f_i, the vector v_b and the decision variables x_i, and PESTO derives the performance achieved by the empirical worst-case instance of the problem.

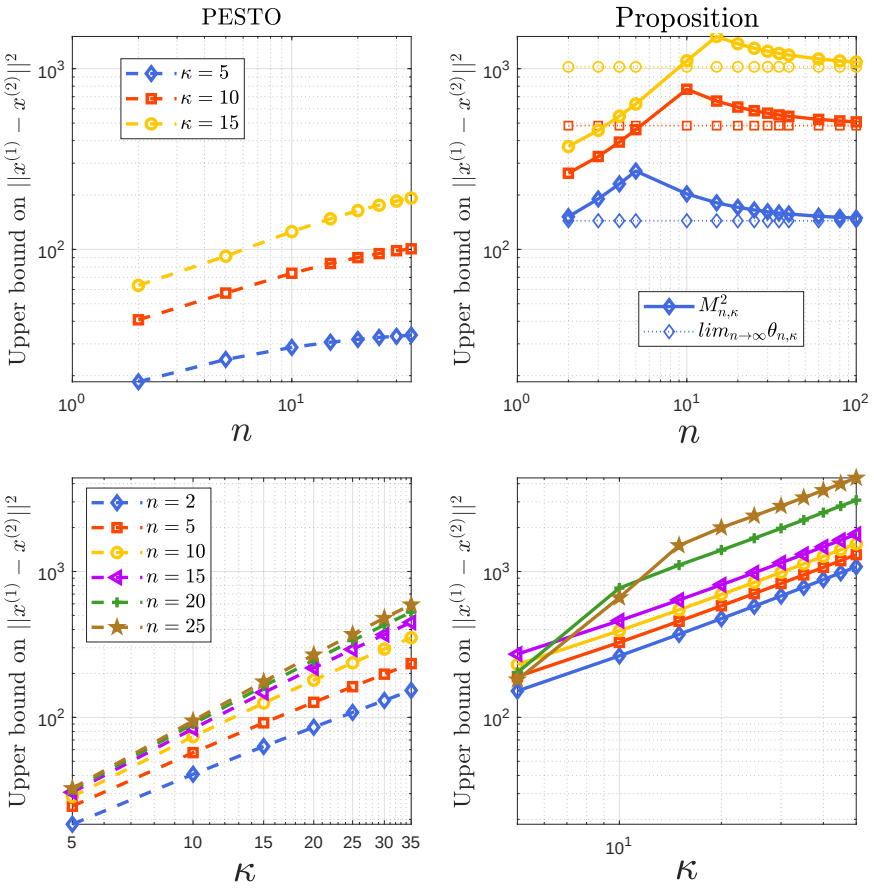


Figure 4.2: Evolution of the upper bound on $\left|\left|x^{(1)} - x^{(2)}\right|\right|^2$ respectively with respect to n for several values of κ (top) and with respect to κ for several values of n (bottom). For each plot the bound obtained in Proposition 4.1 (right) is compared with the empirical upper 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 $n \to \infty$ based on (4.16).

the convergence is exponential if there exists $r \in (0, 1)$ *and some positive constant* C *such that for all* k

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

In optimization, exponential convergence is also known as *R-linear convergence*. Clearly, exponential convergence is weaker than linear convergence since it is concerned with the overall rate of decrease in the error, rather that the decrease over each individual iteration of the algorithm [91].

In [80], the author proves linear convergence of the RCD algorithm in expec-

tation in terms of the function value, i.e., $f(x) - f^*$. Hence, from the inequalities corresponding to smooth functions and strong convexity [92, 93], it is straightforward to prove exponential convergence of the algorithm from [80, Eq. (26)]:

$$\mathbb{E}\left[||x(k) - x^*||\right] \leq \kappa(1 - \alpha\lambda_2(L_p))^k \, ||x_0 - x^*|| \,. \tag{4.21}$$

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 was established in the preliminary work [94] 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 4.1 *Consider a line graph with 3 agents satisfying the constraint* $\langle \mathbb{1}, x \rangle = -3$ *with probabilities* $p_{12} = 0.9$, $p_{23} = 0.1$ *(and hence* $p_{13} = 0$*), and whose local cost functions and initial conditions are:*

i	$f_i(x_i)$	x_i^*	$x_i(0)$
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 $x(0)$ *the expected result of the RCD operation with step-size* $h = 1/\beta = 0.01$ *is*

$$\mathbb{E}\left[||x(1) - x^*||^2\right] = 437.204 > 434 = ||x(0) - x^*||^2 \,, \tag{4.22}$$

and hence linear convergence cannot be achieved at least in this case.

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 [95]. In this case, we focus on the seminorm induced by the Moore-Penrose inverse of the matrix $\mathbf{L}_p$ defined in (4.6), denoted by $\mathbf{L}_p^\dagger$, which we remind is defined as follows for some $x \in \mathbb{R}^{nd}$:

$$\|x\|_{\mathbf{L}_p^\dagger} := \sqrt{x^\top \mathbf{L}_p^\dagger x}. \tag{4.23}$$

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 (3.2) when $b = \mathbf{0}_d$ and corresponds to the kernel of $a^T \otimes I_d$. For the particular case $d = 1$, $\mathcal{S}_{a,0}$ is the orthogonal complement of a.

Proposition 4.2 *The seminorm* $\|\cdot\|_{\mathbf{L}_p^\dagger}$ *is a norm on* $\mathcal{S}_{a,0}$.

Proof: By (4.23), $\|x\|_{\mathbf{L}_p^\dagger} = 0$ implies that x must be in the kernel of $\mathbf{L}_p^\dagger$, defined as $\ker(\mathbf{L}_p^\dagger) = \{x \in \mathbb{R}^{nd} \,|\, x = a \otimes w, w \in \mathbb{R}^d\}$. Since $x \in \mathcal{S}_{a,0}$, it must satisfy $(a^\top \otimes I_d)x = \mathbf{0}_{nd}$ and we have:

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

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

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.

4.3.2 Contraction of an iteration in closed system

Let us remind the update rule of the RCD algorithm defined in (4.4) for some positive step-size h as

$$x(k+1) = x(k) - h\mathbf{Q}^{ij}\nabla f(x(k)). \tag{4.24}$$

In the following proposition, we analyze the convergence of (4.24) with respect to the norm induced by $\mathbf{L}_p^\dagger$ defined tn the previous section.

Proposition 4.3 *Let $f(x) := \sum_{i=1}^n f_i(x_i)$ and $x^* := \arg\min_{x \in \mathcal{S}_{a,b}} f(x)$. Under Assumption 3.1, for any positive scalar*

$$h \leq \frac{\lambda_2(L_p)}{\lambda_n(L_p)} \frac{2}{\alpha + \beta}, \tag{4.25}$$

and for any initial point $x \in \mathcal{S}_{a,b}$, then the update rule (4.24) applied on the randomly selected pair of agents $(i,j) \in E$ satisfies

$$\mathbb{E}\left[\|x(k+1) - x^*\|_{\mathbf{L}_p^\dagger}^2\right] \leq \left(1 - 2h\alpha\lambda_2(L_p) + h^2\alpha^2\lambda_n(L_p)\right)\|x(k) - x^*\|_{\mathbf{L}_p^\dagger}^2. \tag{4.26}$$

Proof: There holds

$$\begin{aligned}
\mathbb{E}\left[\|x(k+1) - x^*\|_{\mathbf{L}_p^\dagger}^2\right] &= \sum_{(i,j)\in E} p_{ij}\|x(k) - h\mathbf{Q}^{ij}\nabla f(x(k)) - x^*\|_{\mathbf{L}_p^\dagger}^2 \\
&= \|x(k) - x^*\|_{\mathbf{L}_p^\dagger}^2 + h^2 \sum_{(i,j)\in E} p_{ij}\|\mathbf{Q}^{ij}\nabla f(x(k))\|_{\mathbf{L}_p^\dagger}^2 \\
&\quad - 2h \sum_{(i,j)\in E} p_{ij}\langle \mathbf{Q}^{ij}\nabla f(x(k)), L_p^\dagger(x(k) - x^*)\rangle.
\end{aligned} \tag{4.27}$$

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

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] = \|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger} + h^2 \sum_{(i,j)\in E} p_{ij}\|\mathbf{Q}^{ij}\nabla f(x(k))\|^2_{\mathbf{L}_p^\dagger} - 2h\langle \mathbf{L}_p\nabla f(x(k)), \mathbf{L}_p^\dagger(x(k)-x^*)\rangle. \tag{4.28}$$

We first treat the second term of the right-hand side of (4.28). Remember from (4.7) that $Q^{ij}a = \mathbf{0}_n$, and from (3.3) that $\nabla f^* = a\otimes\zeta^*$ for some $\zeta^*\in\mathbb{R}^d$. Hence, since $\mathbf{Q}^{ij} = Q^{ij}\otimes I_d$ by definition, there holds

$$\mathbf{Q}^{ij}\nabla f^* = (Q^{ij}\otimes I_d)(a\otimes\zeta^*) = (Q^{ij}a)\otimes\zeta^* = \mathbf{0}_{nd}. \tag{4.29}$$

It thus follows that

$$\begin{aligned}\|\mathbf{Q}^{ij}\nabla f(x(k))\|^2_{\mathbf{L}_p^\dagger} &= \|\mathbf{Q}^{ij}(\nabla f(x(k))-\nabla f^*)\|^2_{\mathbf{L}_p^\dagger}\\ &\le \frac{1}{\lambda_2(L_p)}\left|\left|\mathbf{Q}^{ij}(\nabla f(x(k))-\nabla f^*)\right|\right|^2,\end{aligned} \tag{4.30}$$

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 [96]), so that the smallest and largest nonzero eigenvalues of $\mathbf{L}_p^\dagger$ are respectively $1/\lambda_n(L_p)$ and $1/\lambda_2(L_p)$, yielding for all $z\in\mathbb{R}^{nd}$:

$$\|z\|^2_{\mathbf{L}_p^\dagger} \le \frac{1}{\lambda_2(L_p)}\left|\left|z\right|\right|^2. \tag{4.31}$$

Therefore, since $\mathbf{Q}^{ij} = (\mathbf{Q}^{ij})^\top = (\mathbf{Q}^{ij})^2$, and using the fact that $||z||^2_{\mathbf{L}_p} \le \lambda_n(L_p)\left|\left|z\right|\right|^2$ for all $z\in\mathbb{R}^{nd}$, there holds from (4.30):

$$\begin{aligned}\sum_{(i,j)\in E} p_{ij}\|\mathbf{Q}^{ij}\nabla f(x(k))\|^2_{\mathbf{L}_p^\dagger} &\le \frac{1}{\lambda_2(L_p)}\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2_{\mathbf{L}_p}\\ &\le \frac{\lambda_n(L_p)}{\lambda_2(L_p)}\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2.\end{aligned} \tag{4.32}$$

We now analyze the third term of the right-hand side of (4.28). From (4.29) there holds

$$\mathbf{L}_p\nabla f^* = \sum_{(i,j)\in E} p_{ij}\mathbf{Q}^{ij}\nabla f^* = \mathbf{0}_{nd}, \tag{4.33}$$

yielding

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

Hence, using the result above and (4.31), it follows that

$$-2h\langle \mathbf{L}_p \nabla f(x(k)), \mathbf{L}_p^\dagger (x(k)-x^*)\rangle \leq -2h\frac{\beta^{-1}\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2}{1+\kappa^{-1}} - 2h\frac{\alpha\lambda_2(L_p)\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger}}{1+\kappa^{-1}}. \tag{4.34}$$

Injecting (4.32) and (4.34) into (4.27) yields

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] \leq \left(1-2h\frac{\alpha\lambda_2(L_p)}{1+\kappa^{-1}}\right)\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger} + \left(h^2\frac{\lambda_n(L_p)}{\lambda_2(L_p)}-2h\frac{\beta^{-1}}{1+\kappa^{-1}}\right)\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2. \tag{4.35}$$

Observe that if $h \leq \frac{\lambda_2(L_p)}{\lambda_n(L_p)}\frac{2}{\alpha+\beta}$ then $h^2\frac{\lambda_n(L_p)}{\lambda_2(L_p)} - 2h\frac{\beta^{-1}}{1+\kappa^{-1}} \leq 0$, and hence we can use the definition of α-strongly convex functions to find that

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

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

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] \leq \|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger} + (h^2\alpha^2\lambda_n(L_p)-2h\alpha\lambda_2(L_p))\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger},$$

which concludes the proof. ■

It is clear that the rate of convergence is less than one if $h \leq \frac{2\lambda_2(L_p)}{\alpha\lambda_n(L_p)}$, which is thus satisfied on all its range of validity since $h \leq \frac{2\lambda_2(L_p)}{(\alpha+\beta)\lambda_n(L_p)} \leq \frac{2\lambda_2(L_p)}{\alpha\lambda_n(L_p)}$. We can then find the step-size which minimizes (4.26) and the corresponding convergence rate.

Corollary 4.1 *The optimal rate of convergence in* (4.26) *under* (4.25) *is achieved for* $h^* = \frac{2\lambda_2(L_p)}{(\alpha+\beta)\lambda_n(L_p)}$ *which yields*

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] \leq \left(1-\frac{\lambda_2(L_p)^2}{\lambda_n(L_p)}\frac{1}{\kappa}\right)\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger}. \tag{4.36}$$

Interestingly, Proposition 4.3 shows that linear convergence can be achieved by the RCD algorithm with respect to the norm induced by $\mathbf{L}_p^\dagger$ with a convergence rate similar to that of classical algorithms based on the gradient descent [92, 93].

Remark 4.4 (Complete graph) *For the particular case of a complete graph with 1-dimensional homogeneous agents and uniform probabilities, the eigenvalues of L_p are $\lambda_2(L_p) = \lambda_n(L_p) = \frac{1}{n-1}$ 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 4.3 becomes*

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

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

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

which coincides with [94, Eq. (13)].

Remark 4.5 (Alternative rate) *Starting from* (4.35) *in the proof of Proposition 4.3, one can use a similar argument to derive the following alternative convergence rate, valid for $\frac{\lambda_2(L_p)}{\lambda_n(L_p)}\frac{2}{\alpha+\beta} \leq h \leq \frac{\kappa^{-1}+\kappa_L}{\kappa^{-1}+1}\frac{\lambda_2(L_p)}{\beta\lambda_n^2(L_p)}$, with $\kappa_L = \frac{\lambda_n(L_p)}{\lambda_2(L_p)}$:*

$$\mathbb{E}\left[\|x(k+1) - x^*\|_{\mathbf{L}_p^\dagger}^2\right] \leq \left(1 - 2\beta\lambda_2(L_p)h + h^2\beta^2\frac{\lambda_n^2(L_p)}{\lambda_2(L_p)}\right) \|x(k) - x^*\|_{\mathbf{L}_p^\dagger}^2. \quad (4.37)$$

This result could be used in the rest of this chapter the same way as that of Proposition 4.3 for the corresponding step-size. This development is however omitted as it is only marginal progress.

4.3.3 Homogeneous agents and uniform probabilities

For the particular case of homogeneous agents (i.e., $a = \mathbb{1}_n$) and uniform probabilities $p_{ij} = p$, 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 [97], and we can use the concept of effective resistance to find an upper bound for the step-size of the algorithm independently of $\lambda_2(L_p)$.

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 4.3 in the remainder of this chapter for that case. However, for the sake of generality, we express the main result in the next section only in terms of Proposition 4.3.

Proposition 4.4 *Let a function $f(x) := \sum_{i=1}^n f_i(x_i)$ and $x^* := \arg\min_{x \in S_b} f(x)$. Under Assumption 3.1, for any positive scalar*

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

and for any initial point $x \in \mathcal{S}_{a,b}$, then the update rule (4.24) *applied on the randomly selected pair of agents $(i,j) \in E$ satisfies*

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] \leq \left(1-2h\alpha\lambda_2(L_p)+\frac{h^2\alpha^2\lambda_2(L_p)\lambda_n(L_p)}{2p}\right)\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger}. \tag{4.38}$$

Proof: Since the matrices $\mathbf{Q}^{ij}$ are idempotent, the summation term of the second element in (4.28) can be expressed as:

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

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

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

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

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

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

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

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:

$$\lambda_{\max}(Q^{ij}L^\dagger Q^{ij}) \leq \frac{1}{2p} \text{ for all } (i,j) \in E, \tag{4.39}$$

and we get:

$$\begin{aligned}\sum_{(i,j)\in E} p_{ij}\|\mathbf{Q}^{ij}\nabla f(x(k))\|^2_{\mathbf{L}_p^\dagger} &\leq \frac{1}{2p}\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2_{\mathbf{L}_p} \\ &\leq \frac{\lambda_n(L_p)}{2p}\left|\left|\nabla f(x(k))-\nabla f^*\right|\right|^2, \end{aligned}\tag{4.40}$$

which replaces (4.32). The rest of the proof follows the same steps as in the proof of Proposition 4.3. ■

Similarly to Proposition 4.3, the rate of convergence is strictly decreasing if $h \leq \frac{4p}{\alpha\lambda_n(L_p)}$, which is always satisfied since $h \leq \frac{4p}{(\alpha+\beta)\lambda_n(L_p)} \leq \frac{4p}{\alpha\lambda_n(L_p)}$. Hence, we can also find the optimal step-size for the algorithm, and the corresponding convergence rate.

Corollary 4.2 *The optimal rate of convergence for the RCD algorithm is achieved for* $h^* = \frac{4p}{(\alpha+\beta)\lambda_n(L_p)}$ *which yields*

$$\mathbb{E}\left[\|x(k+1)-x^*\|^2_{\mathbf{L}_p^\dagger}\right] \leq \left(1-\frac{2p\lambda_2(L_p)}{\lambda_n(L_p)}\frac{1}{\kappa}\right)\|x(k)-x^*\|^2_{\mathbf{L}_p^\dagger}. \tag{4.41}$$

The upper bound for the step-size derived in Proposition 4.4 is better suited for graphs with a small $\lambda_2(L_p)$ (also known as *algebraic connectivity*), that is non-robust networks that can be easily disconnected [98]. If we denote by μ the eigenvalues of L, which satisfy $\lambda = \frac{p}{2}\mu$, we have that for Proposition 4.3 the step-size must satisfy $h \leq \frac{\mu_2}{\mu_n}\frac{2}{\alpha+\beta}$ while for Proposition 4.4 the step-size is upper bounded by $h \leq \frac{4}{\mu_n}\frac{2}{\alpha+\beta}$.

4.4 Convergence of RCD in open systems

In the following theorem we present the convergence rate of the RCD algorithm in expectation in an open system. The derivation of this result relies on the separate analysis of the effects of replacements and of RCD updates, which we remind is allowed by Assumption 3.6.

Theorem 4.1 *In the setting described in Section 4.1, the iteration rule* (4.24) *with* $h \leq \frac{\lambda_2(L_p)}{\lambda_n(L_p)}\frac{2}{\alpha+\beta}$ *generates a sequence of estimates* $x(k)$ *satisfying the following for all* k *and for any* $\eta > 0$*:*

$$\mathbb{E}\left[\|x(k+1)-x^*(k+1)\|^2_{\mathbf{L}_p^\dagger}\right] - \Gamma_\eta \leq A_\eta\left(\mathbb{E}\left[\|x(k)-x^*(k)\|^2_{\mathbf{L}_p^\dagger}\right] - \Gamma_\eta\right), \tag{4.42}$$

with

$$A_\eta := 1 - p_U\alpha h(2\lambda_2(L_p) - \alpha h\lambda_n(L_p)) + (1-p_U)\frac{M_{n,\kappa}}{\eta}; \tag{4.43}$$

$$\Gamma_\eta := \frac{(1-p_U)M_{n,\kappa}(\eta+M_{n,\kappa})\eta}{p_U\eta\alpha h(2\lambda_2(L_p)-\alpha h\lambda_n(L_p)) - (1-p_U)M_{n,\kappa}}, \tag{4.44}$$

where $M_{n,\kappa} = \frac{1}{\lambda_2(L_p)}\bar{M}_{n,\kappa}$*, with* $\bar{M}_{n,\kappa}$ *defined in* (4.13)*.*

Proof: Let us denote $C(k) := \|x(k)-x^*(k)\|^2_{\mathbf{L}_p^\dagger}$. From Assumption 3.6, there holds

$$\mathbb{E}\left[C(k+1)\right] = p_U\mathbb{E}\left[C(k+1)|U\right] + (1-p_U)\mathbb{E}\left[C(k+1)|R\right], \tag{4.45}$$

where U and R respectively stand for the occurrence of an RCD update and a replacement event. Proposition 4.3 then yields for $h \leq \frac{\lambda_2(L_p)}{\lambda_n(L_p)} \frac{2}{\alpha+\beta}$

$$\mathbb{E}\left[C(k+1)|U\right] \leq \left(1 - 2h\alpha\lambda_2(L_p) + \lambda_n(L_p)\alpha^2h^2\right)\mathbb{E}\left[C(k)\right]. \tag{4.46}$$

Under a replacement event, there holds $x(k+1) = x(k)$, and hence Proposition 4.1 yields

$$\begin{aligned}\mathbb{E}\left[C(k+1)|R\right] &= \mathbb{E}\left[\|x(k+1) - x^*(k) + x^*(k) - x^*(k+1)\|^2_{\mathbf{L}_p^\dagger}\right] \\ &\leq \mathbb{E}\left[\left(\|x(k) - x^*(k)\|_{\mathbf{L}_p^\dagger} + \|x^*(k) - x^*(k+1)\|_{\mathbf{L}_p^\dagger}\right)^2\right] \\ &= \mathbb{E}\left[\left(\sqrt{C(k)} + M_{n,\kappa}\right)^2\right]. \end{aligned} \tag{4.47}$$

Injecting (4.46) and (4.47) into (4.45) then yields the following nonlinear recurrence:

$$\begin{aligned}\mathbb{E}\left[C(k+1)\right] \leq &\left(1 - p_U\alpha h(2\lambda_2(L_p) - \alpha h\lambda_n(L_p))\right)\mathbb{E}\left[C(k)\right] \\ &+ (1-p_U)\left(2M_{n,\kappa}\mathbb{E}\left[\sqrt{C(k)}\right] + M^2_{n,\kappa}\right). \end{aligned} \tag{4.48}$$

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

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

Solving the linear recurrence then yields the conclusion. ■

Let us now define the ratio

$$\rho_R := \frac{1-p_U}{p_U}, \tag{4.49}$$

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

Corollary 4.3 *Let $\bar{\eta} := \rho_R \frac{M_{n,\kappa}}{\alpha h(2\lambda_2(L_p) - \alpha h\lambda_n(L_p))}$, then for any $\eta > \bar{\eta}$ one has $A_\eta < 1$. Moreover,*

$$\limsup_{k\to\infty} \mathbb{E}\left[\|x(k) - x^*(k)\|^2_{\mathbf{L}_p^\dagger}\right] \leq \Gamma_\eta = \frac{\bar{\eta}(M_{n,\kappa} + \eta)}{1 - \bar{\eta}/\eta}. \tag{4.50}$$

Proof: The proof follows from studying

$$A_\eta < 1 \Longleftrightarrow p_U \alpha h(2\lambda_2(L_p) - \alpha h \lambda_n(L_p)) + (1 - p_U) M_{n,\kappa} < 0,$$

and from rewriting Γ_η as

$$\Gamma_\eta = \frac{\rho_R M_{n,\kappa}(M_{n,\kappa} + \eta)\eta}{\alpha h(2\lambda_2(L_p) - \lambda_n(L_p)\alpha h)\eta - \rho_R M_{n,\kappa}} = \frac{\frac{\rho_R M_{n,\kappa}}{\alpha h(2\lambda_2(L_p) - \alpha h \lambda_n(L_p))}(M_{n,\kappa} + \eta)}{1 - \frac{\rho_R M_{n,\kappa}}{\alpha h(2\lambda_2(L_p) - \alpha h \lambda_n(L_p))\eta}},$$

which directly yields the conclusion. ■

Corollary 4.3 guarantees that one can always choose η such that the contraction rate satisfies $A_\eta < 1$ for any parametrization of the system (as long as $\rho_R < \infty$, i.e., if updates happen). This guarantees convergence to the corresponding Γ_η, defined in (4.50). When $\eta \to \infty$, then the contraction rate becomes minimal, i.e., $A_\eta \to 1 - p_U \alpha h(2\lambda_2(L_p) - \lambda_n(L_p)\alpha h)$, and the asymptotic error becomes extremely large, i.e., $\Gamma_\eta \to \infty$. Observe that for $\eta > \bar{\eta}$, then Γ_η is convex, and one can determine the value of η that minimizes the asymptotic expected error Γ_η, as presented in the following corollary.

Corollary 4.4 *When $\eta = \eta^* = \bar{\eta}\left(1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}\right)$, then the convergence of the RCD algorithm in open system is guaranteed with minimal asymptotic error Γ_{η^*}, and there holds:*

$$A_{\eta^*} = 1 - p_U \alpha h(2\lambda_2(L_p) - \alpha h \lambda_n(L_p)) \frac{\sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}}{1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}}; \tag{4.51}$$

$$\Gamma_{\eta^*} = (\eta^*)^2 = \bar{\eta}^2 \left(1 + \sqrt{1 + \frac{M_{n,\kappa}}{\bar{\eta}}}\right)^2, \tag{4.52}$$

where we remind $\bar{\eta} := \frac{\rho_R M_{n,\kappa}}{\alpha h(2\lambda_2(L_p) - \alpha h \lambda_n(L_p))}$ from Corollary 4.3.

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

$$\frac{d}{d\eta}\Gamma_\eta = \frac{\bar{\eta}}{(\eta - \bar{\eta})^2}(\eta^2 - 2\bar{\eta}\eta - M_{n,\kappa}\bar{\eta}) = 0, \tag{4.53}$$

which is satisfied for

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

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

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

and a few algebraic manipulations yield (4.51). 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 (4.53) for $\eta > \bar{\eta}$, so that the solution is η^*. Hence, $\Gamma_{\eta^*} = (\eta^*)^2$, which yields (4.52). ■

The results of Theorem 4.1 and of Corollary 4.4, and in particular the minimal expected asymptotic error and corresponding convergence rate, allow for some interpretation.

When no replacements happen, i.e., $\rho_R \to 0$, then the system behaves as a closed system, and we retrieve the corresponding convergence behavior: the expected asymptotic error $\Gamma_\eta \to 0$ and the contraction rate $A_\eta \to 1 - \alpha h(2\lambda_2(L_p) - \alpha h\lambda_n(L_p))$ for all $\eta > 0$, consistently with the convergence rate of the RCD in closed system derived in (4.26). 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 $\rho_R \to \infty$, then the minimal expected asymptotic error becomes $\Gamma_{\eta^*} \to 4\bar{\eta} \to \infty$ and $A_{\eta^*} \to 1$. The conservatism of the results essentially follow from Proposition 4.1, 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 highlights one limitation of our model.

Interestingly, within the allowed range of h, the minimal expected asymptotic error Γ_{η^*} decays as the step-size h increases, suggesting that choosing h as large as possible leads to the smallest value of 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 (i.e., in our case $h \leq \frac{\lambda_2(L_p)}{\lambda_n(L_p)}\frac{2}{\alpha+\beta}$ from Proposition 4.3), and that no particular precaution should be taken regarding the open character of the system.

Remark 4.6 *The methodology we used in this section can easily be extended to other algorithms than the RCD algorithm. In particular the results of Theorem 4.1,*

and hence of Corollaries 4.3 and 4.4, can be adapted to any algorithm with linear convergence in closed system, that is, such that

$$\|x(k+1) - x^*\|^2_{\mathbf{L}_p^\dagger} \leq K\|x(k) - x^*\|^2_{\mathbf{L}_p^\dagger}, \tag{4.54}$$

with some positive $K < 1$. In that case, the same convergence rate as that presented in Theorem 4.1 is obtained with

$$A_\eta := 1 - p_U(1-K) + (1-p_U)\frac{M_{n,\kappa}}{\eta}; \tag{4.55}$$

$$\Gamma_\eta := \frac{(1-p_U)M_{n,\kappa}(\eta + M_{n,\kappa})\eta}{p_U\eta(1-K) - (1-p_U)M_{n,\kappa}}. \tag{4.56}$$

We can show that convergence can be guaranteed in open system following a similar argument as that used to prove Corollary 4.3 if $K < 1$. Hence, this analysis can be applied e.g., to the results presented in Proposition 4.4 or in Remark 4.5.

To illustrate the results of Theorem 4.1, we consider systems with piecewise quadratic local cost functions f_i: for $\varphi_{i1}, \varphi_{i2} \in \left[\frac{\alpha}{2}, \frac{\beta}{2}\right]$, the cost function f_i is given by

$$f_i(x_i) = \begin{cases} \varphi_{i1}(x_i - \nu_i)^2, & \text{if } x_i < \nu_i \\ \varphi_{i2}(x_i - \nu_i)^2, & \text{if } x_i \geq \nu_i \end{cases}, \tag{4.57}$$

where ν_i is the minimizer of f_i satisfying Assumption 4.1. Such function therefore satisfies Assumption 3.1 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 $\left[\frac{\alpha}{2}, \frac{\beta}{2}\right]$, and *adversarial*, where these parameters are arbitrarily chosen to maximize the error $\|x(k) - x^*(k)\|^2_{\mathbf{L}_p^\dagger}$ after the replacement among 100 realizations of such uniform random choice.

In Fig. 4.3, we show the evolution of the expected error $\mathbb{E}\left[\|x(k) - x^*(k)\|^2_{\mathbf{L}_p^\dagger}\right]$ 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 (4.42) using the values given by Corollary 4.4. The figure shows that convergence is indeed guaranteed for the RCD in the presented settings and that the result of Corollary 4.4 shows some conservatism, which is inherited from Proposition 4.1. It is interesting to point out that these settings respectively make use of $\bar{M}^2_{n,\kappa} = \theta_{n,\kappa}$ for $n = 30$, $\kappa = 1.2$, and of $\bar{M}^2_{n,\kappa} = \psi_{n,\kappa}$ for $n = 5$, $\kappa = 5$, consistently with the description of $\bar{M}^2_{n,\kappa}$ in the homogeneous case of Section 4.2.

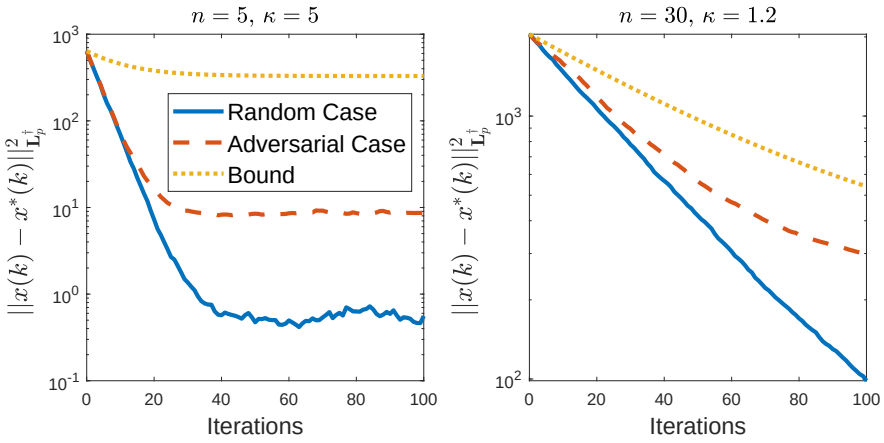


Figure 4.3: Performance of the RCD algorithm in a complete graph constituted of respectively $n = 5$ agents with $\kappa = 5$ (left) and $n = 30$ agents with $\kappa = 1.2$ (right), with $p_U = 0.95$ and $b = 1$, and where each local objective function is defined by (4.57). The solid blue and red dashed 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 (4.42) obtained from Corollary 4.4.

This highlights the impact of those parameters in the tightness of the bound $\bar{M}^2_{n,\kappa}$ used to derive our main results.

In Fig. 4.4, we compare the simulated performance in both replacement cases with the upper bound from Corollary 4.4 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 $\lambda_2(L)$ that does not scale with n. Similarly, the heterogeneous setting impacts $\lambda_2(L)$ and consequently reduces the range of h, due to the imbalance in $\mathbf{L}_p$. Those moreover affect the behavior of the norm $\|\cdot\|^2_{\mathbf{L}_p^\dagger}$. 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. 4.4 are qualitatively similar to the case of the complete graph with homogeneous agents presented in Fig. 4.3.

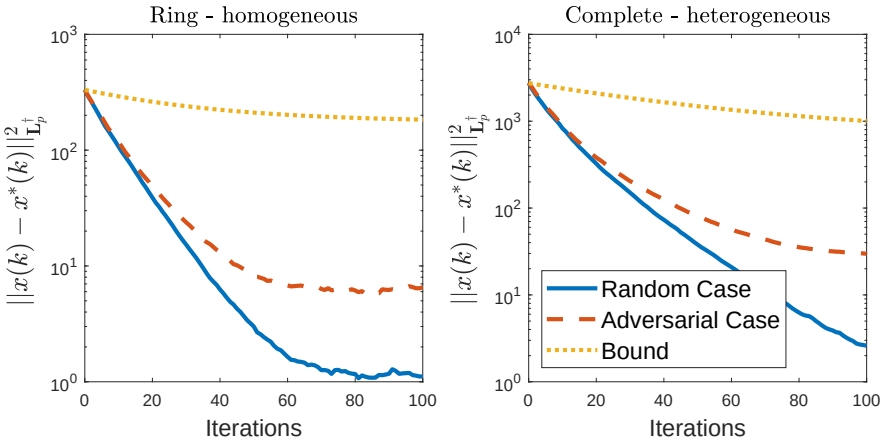


Figure 4.4: Performance of the RCD algorithm with $n = 5$ agents, $\kappa = 1.2$, $p_U = 0.95$ and $b = 1$, respectively in (left) a ring graph with homogeneous agents (i.e., $a_i = 1$ for all i) and (right) a complete graph with heterogeneous agents (i.e., $a_1 = 10$, $a_i = 1$ for $i > 1$), and where each local objective function is defined by (4.57). The solid 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 (4.42) obtained from Corollary 4.4.

4.5 Online optimization analysis

In this section we perform an analysis of the RA problem in OMAS by using performance metrics inspired by online optimization. Due to the potential replacements that occur in the system, the local cost functions of the agents can change in time. Time-varying objective functions can remind an alternative line of work called *online optimization* [85,86], where the approach is to minimize a *regret function* over a finite period of time T, commonly defined as

$$\text{Reg}_T^s := \sum_{k=1}^{T} f^k(x(k)) - \min_x \sum_{k=1}^{T} f^k(x), \tag{4.58}$$

where the superscript s indicates that the minimum of the sum is computed over a *static* variable x [86]. The objective in online optimization is thus to determine the sequence of estimates x^t that keeps Reg_T^s as small as possible over the time period under some assumptions about the possible sequences of time-varying cost functions.

In contrast, the objective of the algorithms in our work is to be at all times as close as possible to the instantaneous minimizer while the possible changes

of the cost functions do not follow a specific rule [87]. Nevertheless, even if the minimization of the regret is not the objective of optimization in OMAS, a related metric called *dynamical regret*, which is defined in (4.61), can be used along with two other similar metrics to evaluate the performance of the algorithms. These quantities are defined based on the assumption that the cost of performing an operation by an agent (i.e., the cost function) must be paid at each iteration. In this case, it is natural to consider the accumulated cost in time as a measure of performance of the system.

We present a preliminary analysis by considering a special case of the RA problem in a simplified setting.

4.5.1 Simplified setting

Regarding the RA problem defined in (3.1), where the local cost functions satisfy the Assumption 3.1, we consider the following assumptions:

Assumption 4.2 (Nonnegative 1-D functions) *The local cost function of any agent at any time is nonnegative one-dimensional:* $f_i^k : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$.

Assumption 4.3 (Minimizer) *The local cost function of any agent satisfies* $\arg\min_{x \in \mathbb{R}^p_{\geq 0}} f_i(x) = \mathbf{0}_p$ *and* $f_i(\mathbf{0}_p) = 0$.

Assumption 4.4 (Homogeneous demand) *The demand associated with any agent* i *at any time* k *is* $d_i^k = 1$ *such that* $b = n$.

Assumption 4.5 *The graph* $G = (V, E)$ *is complete.*

Let $\mathcal{S}_{\mathbb{1},n} := \left\{ x \in \mathbb{R}^n_{\geq 0} : \mathbb{1}^\top x = n \right\}$ be the feasible set, we can now express (3.1) in our setting under the assumptions of this section:

$$\min_{x \in \mathcal{S}_{\mathbb{1},n}} f^k(x) = \sum_{i=1}^{n} f_i^k(x_i). \tag{4.59}$$

4.5.2 Performance metrics

Natural indexes for measuring the performance of an algorithm in our setting consist in evaluating its accumulated error over a finite number of iterations with respect to a given strategy. We define the two following strategies of interest in the context of the RA problem:

- **Perfect collaboration**: at each time instant k the agents know the optimal solution of (4.59) denoted $x^*(k)$;

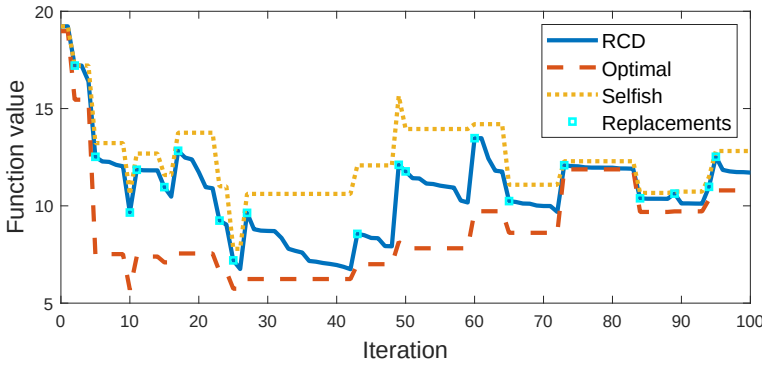


Figure 4.5: Evolution of the function value f^k evaluated with the RCD algorithm $x(k)$ defined in (4.64), the optimal solution $x^*(k)$, and the selfish strategy $x^s(k)$, in a system subject to replacements of agents (i.e., simultaneous departures and arrivals) on average once every 4 RCD steps.

- **Selfish players**: the agents do not collaborate to minimize f^k, and they operate at their individual desired point so that $x^s(k) = \mathbb{1}_n$ at all k.

Hence, for any T, the estimate $x(k)$ obtained with a well-designed algorithm is expected to satisfy

$$\sum_{k=1}^{T} f^k(x^*(k)) \leq \sum_{k=1}^{T} f^k(x(k)) \leq \sum_{k=1}^{T} f^k(x^s(k)). \tag{4.60}$$

The evolution of these strategies, compared with that of a given algorithm, is illustrated in Fig. 4.5.

We define the following performance metrics to analyze the value provided by the RCD algorithm $x(k)$ with respect to the strategies above:

$$\textbf{\textit{Dynamical Regret:}} \qquad \mathrm{Reg}_T := \sum_{k=1}^{T} \left(f^k(x(k)) - f^k(x^*(k)) \right); \tag{4.61}$$

$$\textbf{\textit{Benefit:}} \qquad \mathrm{Ben}_T := \sum_{k=1}^{T} \left(f^k(x^s(k)) - f^k(x(k)) \right); \tag{4.62}$$

$$\textbf{\textit{Potential Benefit:}} \qquad \mathrm{Pot}_T := \sum_{k=1}^{T} \left(f^k(x^s(k)) - f^k(x^*(k)) \right). \tag{4.63}$$

The *dynamical regret* and *benefit* respectively measure the accumulated error from using a given algorithm with respect to the optimal solution $x^*(k)$ and the accumulated gain from using it instead of the selfish strategy $x^s(k)$. The *potential benefit* is independent of the algorithm; it represents the accumulated advantage

of the optimal strategy with respect to the selfish one, and satisfies $\text{Pot}_T = \text{Ben}_T + \text{Reg}_T$.

Remark 4.7 *The regret* (4.58)*, also known as static regret, which is commonly used in online optimization, typically compares $x(k)$ with the overall optimal solution taken over all the iterations, i.e., $x^* = \arg\min_{x\in\mathcal{S}_{\mathbb{1},n}} \sum_{t=1}^{T} f^k(x)$. In that sense, it differs from the dynamical regret in* (4.61)*, which compares $x(k)$ with the time-varying instantaneous optimal solution $x^*(k) = \arg\min_{x\in\mathcal{S}_{\mathbb{1},n}} f^k(x)$ at each iteration, such as e.g., in [86].*

4.5.3 RCD algorithm and replacements

For the particular case $d = 1$, the RCD algorithm (4.4) can be expressed as:

$$\begin{aligned} x_i(k+1) &= x_i(k) - \frac{1}{\beta}(f_i'(x_i(k)) - f_j'(x_j(k))) \\ x_j(k+1) &= x_j(k) - \frac{1}{\beta}(f_j'(x_j(k)) - f_i'(x_i(k))). \end{aligned} \tag{4.64}$$

Unlike the assumption used in Section 4.1.1 where the new agent retrieves the value of the replaced agent, in this second part of the chapter we assume that whenever an agent *in* joins the system, it initializes its estimate as

$$x_{in} = d_{in} = 1, \tag{4.65}$$

that corresponds to the selfish strategy of the agent, and whenever an agent *out* leaves the system, it sends a last message to all its neighbours (i.e., all the other agents in our setting) with its current estimate x_{out} and its demand d_{out} so the agents $i \neq out$ update their estimates as

$$x_i(k+1) = x_i(k) + \frac{x_{out}(k) - x_i(k)}{n} = \left(1 - \frac{1}{n}\right) x_i(k) + \frac{1}{n} x_{out}(k). \tag{4.66}$$

We show in the following proposition that RCD iterations, arrivals and departures as they are defined in (4.64) to (4.66) guarantee that as long as the initial estimate x_0 is feasible, then all the estimates remain feasible.

Proposition 4.5 (Well-posedness) *The event set* (4.2) *guarantees that if $x_0 \in \mathcal{S}_{\mathbb{1},n}$, then $x(k) \in \mathcal{S}_{\mathbb{1},n}$ for all k.*

Proof: We first consider arrivals: the nonnegativity of x_i and preservation of the constraint is a direct consequence of (4.65). In the case of departures, the nonnegativity of x_i is a direct consequence of (4.66). Moreover, if the constraint is

satisfied at iteration k with $n_k = n$, then under the departure of the agent labelled *out* we have

$$\sum_{i \neq out} x_i(k+1) = n - x_{out}(k) + \frac{x_{out}(k)}{n}(n-1) - \frac{n - x_{out}(k)}{n} = n - 1.$$

We finally consider iterations of the RCD algorithm. From Assumptions 3.1, 4.2 and 4.3, it follows that for any $x_i \geq 0$

$$x_i(k) f_i'(x_i(k)) \geq \alpha \left| x_i(k) \right|^2 \geq 0,$$

so that $f_i'(x_i(k)) \geq 0$. Moreover, since f_i is β-smooth, one has $f_i'(x_i(k)) \leq \beta x_i(k)$, and therefore at each update of the RCD algorithm between agents i and j there holds

$$x_i(k+1) = x_i(k) - \frac{1}{2\beta}\left(f_i'(x_i(k)) - f_j'(x_j(k))\right) \geq x_i(k) - \frac{1}{2\beta}(\beta x_i(k)) = \frac{x_i(k)}{2},$$

establishing the nonnegativity of $x_i(k)$. A similar analysis can be used for $x_j(k)$. Due to the symmetry of the update rule, the constraint is always preserved and we conclude the proof. ■

Our goal is to analyze the performance of the RCD algorithm (4.64) with the arrival and departure rules (4.65) and (4.66) in the setting described in Section 4.5.1 using the metrics defined in Section 4.5.2 in expectation.

4.5.4 Upper bounds on the performance metrics

We now derive upper bounds on the evolution of the Potential Benefit and the Dynamical Regret respectively defined in (4.63) and (4.61) in expectation. Whereas the former is only related to the problem itself, the latter actually depends on the algorithm we consider.

We first provide the following lemmas, where Lemma 4.3 directly follows from the equivalence of the norms.

Lemma 4.3 *Let $x \in \mathcal{S}_{\mathbb{1},n}$, then $n \leq ||x||^2 \leq n^2$.*

Lemma 4.4 *Let $f(x) = \sum_{i=1}^n f_i(x_i)$, where all f_i satisfy Assumptions 3.1, 4.2 and 4.3, then for any $x \in \mathcal{S}_{\mathbb{1},n}$ there holds*

$$\frac{\alpha}{2} n \leq f(x) \leq \frac{\beta}{2} n^2. \tag{4.67}$$

Proof: From Assumptions 3.1, 4.2 and 4.3, $f(0) = f'(0) = 0$. Hence, since f is β-smooth and using Lemma 4.3, there holds $f(x) \leq \frac{\beta}{2} ||x||^2 \leq \frac{\beta}{2} n^2$ which establishes the upper bound. Similarly, since f is α-strongly convex and by using Lemma 4.3, it follows that $f(x) \geq \frac{\alpha}{2} ||x||^2 \geq \frac{\alpha}{2} n$, which establishes the lower bound, and concludes the proof. ■

Lemma 4.4 provides a global upper bound on the difference between any two solutions $x, y \in \mathcal{S}_{\mathbb{1},n}$:

$$\left| f^k(x(k)) - f^k(y(k)) \right| \leq \frac{n}{2}(n\beta - \alpha). \tag{4.68}$$

This can be used to derive upper bounds on any of the metrics defined in Section 4.5.2, e.g., $\text{Ben}_T \leq \frac{n}{2}(n\beta - \alpha)T$.

Potential Benefit

We first obtain in the following theorem an upper bound on the expected value of the potential benefit, which we remind quantifies the accumulated advantage of using the optimal strategy rather than not collaborating at all.

Theorem 4.2 *In the setting of Section 4.5.1, there holds*

$$Pot_T \leq \frac{n}{2} \alpha \left(\kappa - 1\right) T, \tag{4.69}$$

and in particular

$$\lim_{T \to \infty} \frac{Pot_T}{T} \leq \frac{n}{2} \alpha (\kappa - 1). \tag{4.70}$$

Proof: Remember that $x^s(k) = \mathbb{1}_n$ by definition, and that $f^k(\mathbf{0}_n) = 0$ and $\nabla f^k(\mathbf{0}_n) = \mathbf{0}_n$ from Assumption 4.3. Hence, there holds from the β-smoothness of f^k

$$f^k(x^s(k)) \leq \nabla f^k(\mathbf{0}_n)^\top (x^s(k)) + \frac{\beta}{2} ||x^s(k)||^2 = \frac{\beta}{2} ||\mathbb{1}_n||^2 = \frac{\beta}{2} n.$$

Similarly, since f^k is α-strongly convex, we get

$$f^k(x^*(k)) \geq \nabla f^k(\mathbf{0}_n)^\top (x^*(k)) + \frac{\alpha}{2} ||x^*(k)||^2 = \frac{\alpha}{2} ||x^*(k)||^2 \geq \frac{\alpha}{2} n,$$

where the last inequality follows from Lemma 4.3. Hence

$$f^k(x^s(k)) - f^k(x^*(k)) \leq \frac{\beta}{2} n - \frac{\alpha}{2} n = \frac{n}{2}(\beta - \alpha)$$

holds, and injecting it into (4.63) yields (4.69). The last result then follows from dividing (4.69) by T. ■

Notice that when the constraint is always preserved, the dynamical regret is nonnegative and the bounds (4.69) and (4.70) also hold for the benefit since $\text{Ben}_T = \text{Pot}_T - \text{Reg}_T \leq \text{Pot}_T$.

Dynamical Regret

We now obtain an upper bound on the expected dynamical regret defined in (4.61), where we remind $x(k)$ is obtained with the RCD algorithm defined in Section 4.5.3. For that purpose, we first introduce the following intermediate quantities:

$$C(k) := f^k(x(k)) - f^k(x^*(k)); \tag{4.71}$$

$$\Delta f(k) := f^{k+1}(x(k+1)) - f^k(x(k)); \tag{4.72}$$

$$\Delta f^*(k) := f^{k+1}(x^*(k+1)) - f^k(x^*(k)). \tag{4.73}$$

Thus, $C(k)$ corresponds to the instantaneous loss of the RCD algorithm with respect to the optimal solution at iteration k, and $\Delta f(k)$ and $\Delta f^*(k)$ respectively stand for the instantaneous variation at one iteration of the total estimated cost and optimal cost, such that $C(k+1) = C(k) + \Delta f(k) - \Delta f^*(k)$.

In the following proposition, we study the effect or replacements on $\Delta f(k)$ in order to later characterize $C(k)$ in expectation, and consequently the expected dynamical regret.

Proposition 4.6 *In the setting of Section 4.5.1 the replacement of an agent, denoted R, results in*

$$\mathbb{E}\left[\Delta f(k) \mid R\right] \leq \frac{5}{2}\beta - \frac{3}{2}\alpha. \tag{4.74}$$

Proof: We analyze the effects of arrivals and departures separately. Let g denote the local cost function of the joining agent at an arrival, then $f^{k+1}(x(k+1)) = f^k(x(k)) + g(1)$ and

$$\Delta f(k) = f^k(x(k)) + g(1) - f^k(x(k)) = g(1) \leq \frac{\beta}{2}, \tag{4.75}$$

where the last inequality follows from Assumptions 3.1 and 4.3, and in particular the β-smoothness of g.

Consider now a departure, and let ℓ denote the label of the leaving agent, such that $f^{k+1}(x(k+1)) = \sum_{i \neq \ell} f_i^k(x_i(k+1))$, with $x_i(k+1) = x_i(k) + \frac{x_i(k)+x_\ell(k)}{n}$ from (4.66). From the definition of departures, ℓ is uniformly selected among the n agents in the system and by taking the expected value $\Delta f(k)$ over the leaving agent, one gets the following, where we omit the reference to time to lighten the

notation:

$$\begin{aligned}\mathbb{E}\left[\Delta f(k)\right] &= \sum_{\ell=1}^{n} \frac{1}{n}\left(\sum_{i\neq\ell} f_i\left(x_i + \frac{x_\ell - x_i}{n}\right) - f(x)\right) \\ &= \frac{1}{n}\sum_{\ell=1}^{n}\left(f\left(\frac{n-1}{n}x + \frac{x_\ell}{n}\mathbb{1}_n\right) - f_\ell(x_\ell)\right) - f(x) \\ &= \frac{1}{n}\sum_{\ell=1}^{n} f\left(\frac{n-1}{n}x + \frac{x_\ell}{n}\mathbb{1}_n\right) - \frac{n+1}{n}f(x).\end{aligned}$$

Since f is β-smooth from Assumption 3.1, one has

$$f\left(\frac{n-1}{n}x + \frac{x_\ell}{n}\mathbb{1}_n\right) \leq f(x) + \frac{1}{n}\langle\nabla f(x), x_\ell\mathbb{1}_n - x\rangle + \frac{\beta}{2n^2}\left|\left|x_\ell\mathbb{1}_n - x\right|\right|^2,$$

and it follows that

$$\mathbb{E}\left[\Delta f(k)\right] \leq \frac{1}{n^2}\sum_{\ell=1}^{n}\langle\nabla f(x), x_\ell\mathbb{1}_n - x\rangle + \frac{\beta}{2n^3}\sum_{\ell=1}^{n}\left|\left|x_\ell\mathbb{1}_n - x\right|\right|^2 - \frac{1}{n}f(x). \quad (4.76)$$

From Assumption 3.1, in particular since f is β-smooth and α-strongly convex, it satisfies $\alpha\left|\left|x\right|\right|^2 \leq \langle\nabla f(x), x\rangle \leq \beta\left|\left|x\right|\right|^2$ for any x. Hence, reminding that $\sum_{\ell=1}^{n} x_\ell = n$, the first sum of (4.76) can be upper bounded by

$$\begin{aligned}\sum_{\ell=1}^{n}\langle\nabla f(x), x_\ell\mathbb{1}_n - x\rangle &= n\langle\nabla f(x), \mathbb{1}_n - x\rangle \\ &\leq n(\beta n - \alpha\left|\left|x\right|\right|^2).\end{aligned}$$

The second sum of (4.76) can be expressed as:

$$\begin{aligned}\sum_{\ell=1}^{n}\left|\left|x_\ell\mathbb{1}_n - x\right|\right|^2 &= \sum_{\ell=1}^{n}\sum_{i=1}^{n}\left(x_\ell^2 - 2x_\ell x_i + x_i^2\right) \\ &= \sum_{\ell=1}^{n}\left(nx_\ell^2 - 2nx_\ell + \left|\left|x\right|\right|^2\right) \\ &= 2n\left(\left|\left|x\right|\right|^2 - n\right).\end{aligned}$$

Then (4.76) is upper bounded by

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

Lemmas 4.3 and 4.4 yield $||x||^2 \leq n^2$ and $f(x) \geq \frac{\alpha}{2}n$, so that

$$\mathbb{E}\left[\Delta f(k)\right] \leq 2\beta - \frac{3}{2}\alpha. \tag{4.77}$$

The conclusion follows from adding (4.75) and (4.77). ■

We can now use Proposition 4.6 to study the evolution of the expected dynamical regret in the following theorem.

Theorem 4.3 *In the setting of Section 4.5.1, there holds*

$$\mathbb{E}\left[Reg_T\right] \leq C_0 \sum_{k=1}^{T} \eta^k + \sum_{k=0}^{T-1} \eta^k \left(M_f + (1-p_U)(T-k)\theta\right), \tag{4.78}$$

where $\eta = 1 - \frac{p_U}{\kappa(n-1)}$ *(with* p_U *the probability that a given event is an update from Assumption 3.6),* $M_f = \frac{n}{2}(\beta n - \alpha)$ *and* $\theta = \frac{5}{2}\beta - \frac{3}{2}\alpha$.

Proof: Let $\gamma = 1 - \frac{1}{\kappa(n-1)}$ denote the contraction rate of the RCD algorithm as defined in (4.64) [80]. Remember that there holds $C(k+1) = C(k) + \Delta f(k) + \Delta f^*(k)$ for all times k. Hence, at any time-step k one has

$$\mathbb{E}\left[C(k+1)\right] = \mathbb{E}\left[C(k) + \Delta f(k) - \Delta f^*(k)\right]. \tag{4.79}$$

From Assumption 3.6 the event at iteration k is an update, denoted U_k, with probability p_U, or a replacement, denoted R_k, with probability $1 - p_U$.
In the case of an update, we have $x^*(k+1) = x^*(k)$, so that $\Delta f^*(k) = 0$. Hence, we have $\Delta f(k) = C(k+1) - C(k)$, and since $\mathbb{E}\left[C(k+1)|C(k), U_k\right] \leq \gamma C(k)$ with the RCD algorithm from [80]:

$$\mathbb{E}\left[\Delta f(k)|, U_k\right] = \mathbb{E}\left[C(k+1) - C(k)|U_k\right] \leq (\gamma - 1)\mathbb{E}\left[C(k)\right]. \tag{4.80}$$

In the replacement case, there holds

$$\mathbb{E}\left[\Delta f(k)|R_k\right] \leq \theta = \frac{5}{2}\beta - \frac{3}{2}\alpha, \tag{4.81}$$

where θ comes from Proposition 4.6. Injecting (4.80) and (4.81) into (4.79) then yields

$$\begin{aligned}\mathbb{E}\left[C(k+1)\right] &\leq \mathbb{E}\left[C(k)\right] + p_U(\gamma - 1)\mathbb{E}\left[C(k)\right] + (1-p_U)\theta - \mathbb{E}\left[\Delta f^*(k)\right] \\ &= \eta\mathbb{E}\left[C(k)\right] + (1-p_U)\theta - \mathbb{E}\left[\Delta f^*(k)\right],\end{aligned} \tag{4.82}$$

where $\eta = 1 + p_U(\gamma - 1)$. Expression (4.82) actually describes the evolution of a discrete-time dynamical system of the type $v(k+1) \leq Av(k) + Bu(k)$. Standard results on that framework yield $v(k) \leq A^k v(0) + \sum_{j=0}^{k-1} A^{k-j-1} Bu(j)$, and we obtain

$$\mathbb{E}\left[C(k)\right] \leq \eta^k C_0 + \sum_{j=0}^{k-1} \eta^{k-j-1} \left((1-p_U)\theta - \mathbb{E}\left[\Delta f^*(k)\right]\right). \tag{4.83}$$

Injecting this last result into (4.61) then yields

$$\begin{aligned}\mathbb{E}\left[\mathrm{Reg}_T\right] &= \sum_{k=1}^{T} \mathbb{E}\left[C(k)\right] \\ &\leq C_0 \sum_{k=1}^{T} \eta^k + \sum_{k=1}^{T} \left(\sum_{j=0}^{k-1} \eta^{k-j-1} \left((1-p_U)\theta - \mathbb{E}\left[\Delta f^*(k)\right]\right) \right).\end{aligned}$$

After some term re-organization, it becomes

$$\mathbb{E}\left[\mathrm{Reg}_T\right] \leq C_0 \sum_{k=1}^{T} \eta^k + (1-p_U) \sum_{k=0}^{T-1} (T-k)\eta^k \theta - \sum_{k=0}^{T-1} \eta^k \left(\sum_{j=0}^{T-k} \mathbb{E}\left[\Delta f_j^*\right] \right).$$

Finally, using Lemma 4.4, one concludes that

$$\begin{aligned}-\sum_{j=0}^{T-k} \mathbb{E}\left[\Delta f_j^*\right] &= -\mathbb{E}\left[\sum_{j=0}^{T-k} \Delta f_j^*\right] \\ &= -\mathbb{E}\left[f^{T-k+1}(x^*(T-k+1)) - f^1(x^*(1))\right] \\ &\leq \frac{n}{2}(\beta n - \alpha),\end{aligned}$$

and $M_f = \frac{n}{2}(\beta n - \alpha)$ yields the conclusion. ■

We now analyze the asymptotic behavior of the averaged regret in the following corollary.

Corollary 4.5 *Let $\rho_R := \frac{1-p_U}{p_U}$. In the same setting as that of Theorem 4.3, there holds*

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[Reg_T\right]}{T} \leq \rho_R (n-1)\beta \left(\frac{5\kappa - 3}{2}\right). \tag{4.84}$$

Proof: Starting from (4.78), we have

$$\frac{\mathbb{E}\left[\mathrm{Reg}_T\right]}{T} \leq C_0 \sum_{k=1}^{T} \frac{\eta^k}{T} + \sum_{k=0}^{T-1} \left(M_f \frac{\eta^k}{T} + (1-p_U)\left(1 - \frac{k}{T}\right)\eta^k \theta \right).$$

Remember that $\eta = 1 - \frac{p_U}{\kappa(n-1)} \leq 1$, so that $\sum_{k=1}^{T} \eta^k < T$, and $\lim_{T\to\infty} \sum_{k=1}^{T} \frac{\eta^k}{T} = 0$. Hence

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\mathrm{Reg}_T\right]}{T} \leq \lim_{T\to\infty} (1-p_U) \sum_{k=0}^{T-1} \left(1 - \frac{k}{T}\right) \eta^k \theta.$$

Moreover, for $\eta < 1$, one shows $\sum_{j=0}^{\infty} \eta^j = \frac{1}{1-\eta}$ and $\sum_{j=0}^{\infty} j\eta^j = \frac{\eta}{(1-\eta)^2}$. The latter implies that $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} k\eta^k = 0$, and therefore there holds

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\mathrm{Reg}_T\right]}{T} \leq (1-p_U) \frac{1}{1-\eta} \theta = \frac{1-p_U}{p_U} \kappa(n-1)\theta,$$

and the conclusion follows from the definitions of θ in Theorem 4.3, and from $\rho_R := \frac{1-p_U}{p_U}$. ■

The upper bounds for the potential benefit (4.69) and the dynamical regret (4.78) linearly scale with T. This behavior is rather natural for the former, which does not depend on the algorithm. For the latter, it is most likely unavoidable due to the introduction at each replacement of perturbations of non-decaying magnitude and without any regularity [99]. Interestingly, this behavior contrasts with standard results in online optimization, where a sublinear growth in T is desired to cancel the asymptotic averaged regret [85, Ch. 1.1]. However, these results usually apply on the *static regret* (4.58), where $x(k)$ is compared with an overall time-independent strategy x^* computed over all T iterations, in opposition with $x^*(k)$ which is optimal for each iteration. In the case of the dynamical regret, it has been showed that without any regularity in the variations of the cost functions, a sublinear behavior cannot be achieved [99].

Moreover, the corresponding asymptotic upper bounds linearly grow with n and α for (4.70), and with $n-1$ and β for (4.84), consistently with their expected behavior. In particular, the scaling of (4.84) with $n-1$ follows from the convergence rate of the RCD algorithm $\gamma = 1 - \frac{1}{\kappa(n-1)}$. Interestingly, (4.84) is proportional to $\rho_R(n-1) = (1-p_U)\frac{n-1}{p_U}$, and the bound can thus be seen as the ratio between the probability for a given agent to be involved in a RCD update $\frac{p_U}{n-1}$ (involved in γ), and the impact of replacements at the system level $1-p_U$, independently of n. This is consistent with the bound on the impact of replacements in (4.74) which is independent of n (by contrast, alternative situations such as e.g., if all agents were to be reset at each replacement are expected to generate an impact growing with n). Hence, for small values of ρ_R (i.e., rare replacements), the bound guarantees that the asymptotic dynamical regret remains reasonably bounded, and decays to zero when $\rho_R \to 0$, i.e., for closed systems.

Finally, observe that (4.84) is proportional to $\frac{5\kappa-3}{2}$, consistently with the fact that a larger interval for the possible curvature of the cost functions should generate a larger potential error at replacements. This factor is a potential source of

conservatism, e.g., with respect to (4.70) where the scaling is in $\frac{1}{2}(\kappa - 1)$. More generally, it is not clear yet whether other algorithms than the RCD might provide tighter bounds.

Remark 4.8 *The proofs of Theorem 4.3 and Corollary 4.5 can directly be adapted to any contraction rate $\gamma < 1$, and are thus easily generalized to any other algorithm that guarantees linear convergence; in particular* $\lim_{T\to\infty} \frac{\mathbb{E}[Reg_T]}{T} \leq \rho_R \frac{\theta}{1-\gamma}$.

4.5.5 The case of quadratic functions

The bound on the expected dynamical regret can be refined for the particular case where all local functions are quadratic, i.e., satisfy the following additional assumption.

Assumption 4.6 (Quadratic functions) *The local cost function of any agent i at time t is of the form*

$$f_i(x_i) = \phi_i x_i^2, \qquad \phi_i \in \left[\tfrac{\alpha}{2}, \tfrac{\beta}{2}\right]. \tag{4.85}$$

The parameter ϕ_i is randomly chosen according to a distribution with a finite support determined by the interval $\left[\frac{\alpha}{2}, \frac{\beta}{2}\right]$. Observe that functions satisfying Assumption 4.6 necessarily satisfy Assumption 3.1 as well.

Under Assumption 4.6, we can obtain a tighter bound than that of Proposition 4.6, presented in the following proposition.

Proposition 4.7 *In the setting of Section 4.5.1, and under Assumption 4.6, the replacement of an agent R results in*

$$\mathbb{E}\left[\Delta f(k) \mid R\right] \leq \frac{3n^2 - 3n + 1}{2n^2} (\beta - \alpha). \tag{4.86}$$

Proof: The arrival case is treated the same was as in the proof of Proposition 4.6, resulting in $\Delta f(k) \leq \frac{\beta}{2}$. The departure case follows the same first steps with $f_i(x) = \phi_i x^2$, and

$$\begin{aligned}
\mathbb{E}\left[\Delta f(k)\right] &= \sum_{\ell=1}^{n} \frac{1}{n} \left(\sum_{i\neq\ell} \phi_i \left(x_i + \frac{x_\ell - x_i}{n} \right)^2 - f(x) \right) \\
&= \frac{1}{n} \sum_{\ell=1}^{n} \left(\sum_{i\neq\ell} \phi_i \left(\left(x_i + \frac{x_\ell - x_i}{n} \right)^2 - x_i^2 \right) - \phi_\ell x_\ell^2 \right) \\
&= \frac{1}{n^2} \sum_{\ell=1}^{n} \sum_{i\neq\ell} \phi_i \left(\frac{1-2n}{n} x_i^2 + 2 x_i x_\ell \frac{n-1}{n} + \frac{x_\ell^2}{n} \right) - \frac{f(x)}{n}.
\end{aligned}$$

Using the fact that $\sum_{\ell=1}^{n}\sum_{i\neq\ell}\phi_i x_i^2 = (n-1)f(x)$ and that $\phi_i \leq \frac{\beta}{2}$ for all i, we obtain

$$\mathbb{E}\left[\Delta f(k)\right] \leq \left(\frac{1}{n^2}\frac{1-2n}{n}(n-1) - \frac{1}{n}\right) f(x) + \frac{\beta}{2n^2}\sum_{\ell=1}^{n}\sum_{i\neq\ell}\left(2x_i x_\ell \frac{n-1}{n} + \frac{x_\ell^2}{n}\right).$$

Observe that $\sum_{\ell=1}^{n}\sum_{i\neq l} x_i x_\ell = n^2 - ||x||^2$ and $\sum_{\ell=1}^{n}\sum_{i\neq\ell} x_\ell^2 = (n-1)\,||x||^2$, so that a few algebraic manipulations yield

$$\mathbb{E}\left[\Delta f(k)\right] \leq -\frac{3n^2-3n+1}{n^3} f(x) + \beta\frac{n-1}{2n^3}\left(2n^2 - ||x||^2\right).$$

Using $||x||^2 \geq n$ (from Lemma 4.3) and $f(x) \geq \frac{\alpha}{2}n$ (from Lemma 4.4) then yields

$$\mathbb{E}\left[\Delta f(k)\right] \leq -\frac{3n^2-3n+1}{2n^3}\alpha + \beta\frac{2n^2-3n+1}{2n^3},$$

and combining with the arrival case concludes the proof. ■

The result above allows us stating the following theorem, which improves Theorem 4.3 and Corollary 4.5 respectively for the case of quadratic functions.

Theorem 4.4 *In the setting of Section 4.5.1, and under Assumption 4.6, there holds*

$$\mathbb{E}\left[Reg_T\right] \leq C_0\sum_{k=1}^{T}\eta^k + \sum_{k=0}^{T-1}\eta^k\left(M_f + (1-p_U)(T-k)\theta\right), \tag{4.87}$$

where $\eta = 1 - \frac{p_U}{\kappa(n-1)}$ *(with* p_U *the probability that a given event is an update from Assumption 3.6),* $M_f = \frac{n}{2}(\beta n - \alpha)$ *and* $\theta = (\beta-\alpha)\frac{3n^2-3n+1}{2n^2}$*. In particular,*

$$\lim_{T\to\infty}\frac{\mathbb{E}\left[Reg_T\right]}{T} \leq \rho_R(n-1)\frac{3n^2-3n+1}{2n^2}\beta\left(\kappa-1\right). \tag{4.88}$$

Proof: The proof follows the exact same steps as those of Theorem 4.3 and Corollary 4.5 where Proposition 4.7 is used instead of Proposition 4.6. ■

The upper bound for the quadratic case is qualitatively better than that of the general case; it was derived based on an additional information of the cost function, thus resulting in tighter bounds. In this case, the dependence of (4.88) is in $\frac{3}{2}(\kappa-1)$, which is consistent with the result derived for the potential benefit (4.70). In particular, for n becoming large, (4.88) and (4.84) become equivalent up to a constant β. Moreover, (4.88) becomes 0 when $\kappa = 1$, consistently with the expected behavior of the RCD algorithm for quadratic functions since all the cost functions would then be the same.

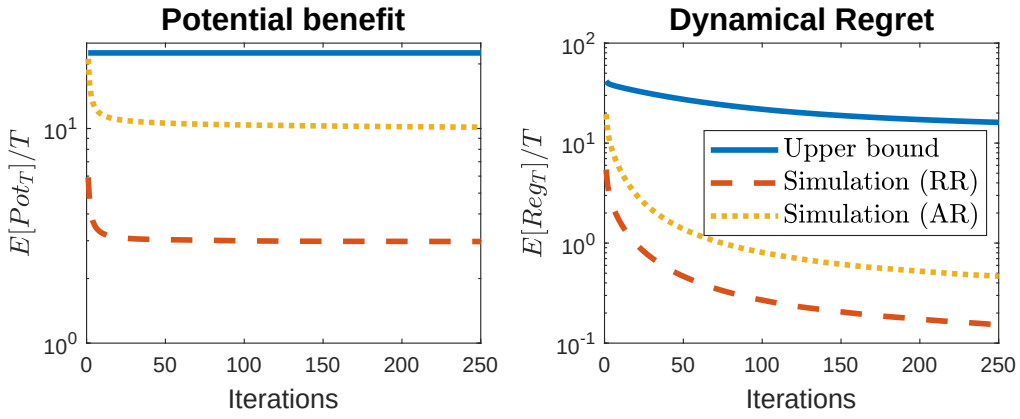


Figure 4.6: Evolution of the averaged asymptotic expected Potential Benefit (on the left) and dynamical regret (on the right) in a system of 5 agents with $\rho_R = 0.0125$ and $\kappa = 10$. Each plot compares the upper bounds, respectively from (4.70) and (4.84), with simulated results, either with random replacements (RR) or adversarial replacements (AR).

4.5.6 Numerical results

To illustrate the results of Theorems 4.2 to 4.3, we consider a system of 5 agents with $\kappa = 10$ and $\rho_R = 0.0125$, the latter implies that on average there is one replacement every 80 events. We consider two possibilities: *random replacements* (RR) where the local function is randomly uniformly chosen among the set of piecewise quadratic functions satisfying Assumptions 3.1, 4.2 and 4.3, and *adversarial replacements* (AR), where these functions are quadratic functions $\phi_i x^2$, with $\phi_i \in \left\{\frac{\alpha}{2}, \frac{\beta}{2}\right\}$. The AR setting is expected to be less favorable than the RR setting, since replacements might result in the largest change of local functions. Notice that the bounds (4.69) and (4.78) are independent of the distribution from which the local cost functions are assigned to the agents when they join the system, so that they hold for any such assignment rule.

Fig. 4.6 compares the results of Theorem 4.2 and Corollary 4.5 with simulations for both random and adversarial replacements in the setting described above. Even though the theoretical bounds are conservative, they capture well the qualitative behavior of these metrics. In particular, consistently with Pot_T and $\mathbb{E}\left[\mathrm{Reg}_T\right]$ that grow linearly with T, the bounds in the figure do not converge to zero, and a remaining asymptotic error is observed. Our bounds are tighter for the adversarial replacement case than the random replacement case, this suggests that our bounds might be tight for some particular choice of the joining functions at replacements, especially that on the potential benefit.

Fig. 4.7 compares the results of Theorems 4.3 and 4.4 with simulations in

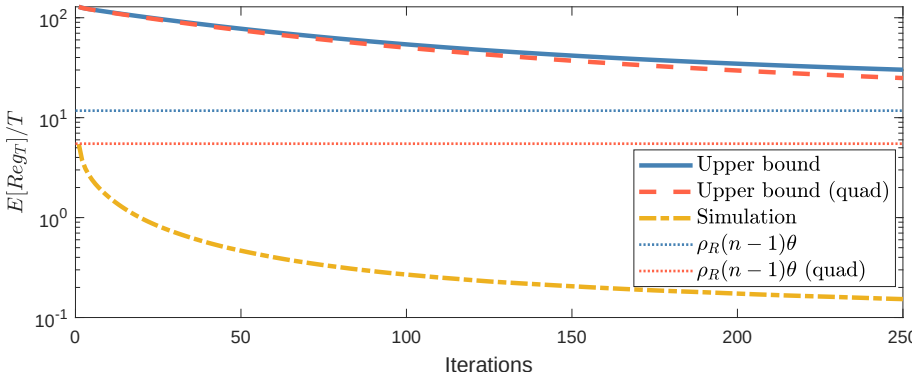


Figure 4.7: Evolution of the expected averaged regret for a system of 5 agents holding quadratic functions with $\rho_R = 0.0125$ and $\kappa = 10$. The solid blue line and the dash-dotted red line respectively correspond to the upper bounds of Theorems 4.3 and 4.4 respectively. The dotted yellow line corresponds to simulation where we consider random replacements (RR).

the same setting as described previously, with random replacements (RR). The figure shows that bound (4.88) is tighter for quadratic functions, following the fact that we have access to more information regarding the local cost functions, thus improving the estimation of the effect of replacements on the expected regret.

4.6 Conclusion

In this chapter, we studied the Random Coordinate Descent algorithm in a resource allocation problem by analyzing the distance to the minimizer and by using metrics inspired from online optimization. Regarding the behavior of the distance to the minimizer, we proved linear convergence of the RCD algorithm in an appropriate norm for the closed system. We analyzed the algorithm for a general graph topology and possible heterogeneous agents in OMAS 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.

Inspired by the similarities with the online optimization, we considered a simple preliminary setting where the budget is homogeneous and the graph is complete, and show that it is not possible to achieve convergence to the optimal solution with the RCD algorithm in expectation in OMAS, but that the error is expected to remain reasonable. We have derived upper bounds on the evolution of the

dynamical regret and the potential benefit in expectation and showed that due to the random choice of the new local cost function during replacements, an error is expected to be accumulated with time and cannot be compensated. A natural continuation of this work would be to extend the results to more general graph topologies, improve the current bounds for the three metrics and derive lower bounds that allow us to support the analysis regarding the linear behavior of the metrics.

This chapter opens several questions for future research. Scenarios where the budget in the constraint is time-varying looks more appropriate for OMAS where the occurrence of a replacement may imply a modification of the constraint in the optimization problem. In this case, it would be interesting to analyze possible violations of the constraint as in Chapter 3. In addition, the case of agents replaced with new values of the corresponding weights a_i is also linked to this future line of research, where the contributions of new agents to the budget can be different.

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. In this scenario, it would be important to specify the potential changes in the constraint due to arrivals and departures since the budget of the system would be defined by the sum of the states of a different number of agents.

In this chapter, the constraint in the RA problem was simply defined as the weighted sum of the states of the agents. However, it would be interesting to analyze scenarios where the constraint has the form $Ax = b$, where the matrix A is used to model more complex relationships between the state variables of the agents. Finally, block updates at each iteration [100], looks also as an interesting direction for future work, where at each iteration the optimization is performed along more than one edge, such that the convergence of the algorithm is faster.

Chapter 5

Contagion in OMAS

In this chapter, we consider a continuous time formulation of a SIS epidemic over a network of individuals and perform an analysis of the system when replacements take place. Section 5.1 introduces the general characteristics of epidemic models used to model the spread of diseases. In Section 5.2 we present the nonlinear model of a SIS epidemic over a network and the aggregate function used for the analysis. We also present the linearized version of the model, which will be used for the derivation of an upper bound for the variance of the aggregate function. In Section 5.3 we analyze the SIS epidemic when replacements of agents take place during the evolution of the disease. We consider that the time instants at which replacements occur are determined by a homogeneous Poisson process. The expectation and variance of the aggregate function for the system are studied by deriving an upper bound for their asymptotic values when the network of connections is given. Additionally, we provide an analysis of the expectation of the aggregate function when the graph is sampled from a piecewise Lipschitz graphon. We give some insights into the analysis of more complex scenarios where the connections of the network can change during the replacements according to a piecewise Lipschitz graphon, and we formulate a conjecture on a potential upper bound for the asymptotic value of the aggregate function in expectation. Finally, in Section 5.4 we present the conclusions and future work of this chapter.

5.1 Epidemic models

Epidemic models are dynamical systems used to model the spread of a disease in a community of individuals. The importance of epidemic models has grown in the last years, specially due to the COVID-19 outbreak that impacted the entire world. Different types of mathematical models were used to describe the spread of epidemics. For instance, mean field approximations, non-markovian epidemics, per-

colation process, partial differential equations, among others, [101], which model specific characteristics and properties of the spread of the disease. Depending on the nature of the disease, the models can include different compartments (division of the population); for instance, M for immune class, S for susceptible class, E for exposed class, I for infective class and R for recovered class [102].

One of the most important models is the SIS (Susceptible-Infected-Susceptible) epidemic model, which describes diseases where the agents, after being recovered, can acquire the disease again. The dynamics of this epidemic is determined by the following equation [101]:

$$\dot{x}(t) = \beta x(t)(1 - x(t)) - \delta x(t),$$

where $x(t)$ is the fraction of a fixed population infected at time t.

In real life, the evolution of the epidemic is not the same in all the infected agents since each individual may be exposed to different external conditions that can have a significant impact in the infection and recovery processes. One of the most important factors that contribute to the propagation of a disease is the network of contacts of an individual, which allow to model the influence of social interactions in the evolution of the disease. SIS models have been extended to the spread of diseases over networks where nodes can have two different interpretations. If the node i is considered as a small population, $x_i(t)$ represents the fraction of infected people at time t. If the node i is considered as an individual, $x_i(t)$ represents its probability of infection at time t [103].

Most of the stability results derived using network models depend on the spectral radius of the adjacency matrix associated with the network, assuming that the composition of the system is fixed [104]. Nevertheless, in more realistic environments, individuals do not stay in a fixed place for a long time and are constantly moving between different locations. Different works begin to consider the mobility as an important aspect that must be taken into account to obtain more accurate results in the study of the spread of diseases [105,106]. In this case, open multi-agent systems appear as an important tool for the analysis of more realistic epidemics where the set of agents is time-varying. Even, if the total number of individuals in a specific place could remain almost invariant, it is unrealistic to assume that exactly the same agents stay in a fixed location during all the evolution of the epidemic. If we consider a similar rate of departures and arrivals, which preserve the number of individuals, epidemics under these conditions can be modeled as a dynamical system subject to replacements of individuals.

5.2 SIS networked model

We consider a system composed by n agents interacting through a connected undirected graph $G = (V, E)$. The SIS epidemic model over a network is given by [104]:

$$\dot{x}_i(t) = -\delta x_i(t) + \sum_{j=1}^{n} a_{ij}\beta x_j(t)\left(1 - x_i(t)\right),$$

where $x_i(t) \in [0, 1]$ is the probability of infection of an individual at time t, δ is the recovery rate, β is the infection rate, $a_{ij} = 1$ if there is an interaction between agents i and j, and $a_{ij} = 0$ otherwise.

The model of all the network can be expressed as:

$$\dot{x}(t) = \left(\beta A - \delta I\right) x(t) - \beta X(t) A x(t), \tag{5.1}$$

where $A = [a_{ij}]$ is the adjacency matrix and $X(t) = \text{diag}[x_1(t), \ldots, x_n(t)]$.

Proposition 5.1 (Stability [104]) *For the SIS epidemic* (5.1)*, the equilibrium* $x = 0$ *is globally asymptotically stable if and only if*

$$\lambda_1(A)\frac{\beta}{\delta} < 1, \tag{5.2}$$

where $\lambda_1(A)$ *is the largest eigenvalue of* A*.*

Since in OMAS the size of the system may change in time due to decoupled arrivals and departures, it is desirable to perform the analysis with the use of an aggregate function $V(x(t))$ that condensates the associated process in a scalar value, independently of the size of the system. In the case of SIS epidemics, a natural choice is the Lyapunov function $V : [0, 1]^n \to \mathbb{R}_{\geq 0}$:

$$V(x(t)) = \frac{1}{n}\left|\left|x(t)\right|\right|^2, \tag{5.3}$$

which satisfies

$$\lim_{t\to\infty} V(x(t)) = 0, \tag{5.4}$$

when the epidemic is stable.

This particular definition of $V(x(t))$ provides uniform bounds for the aggregate function, independently of the size of the system:

$$0 \leq V(x(t)) \leq 1 \quad \text{for all } t. \tag{5.5}$$

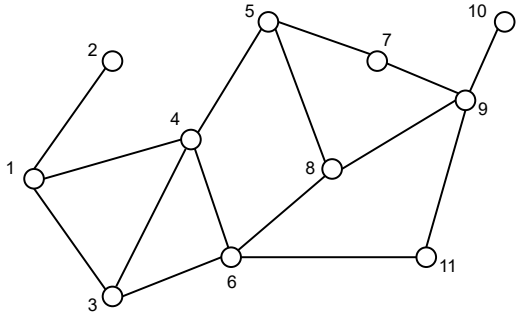


Figure 5.1: Graph G.

In a closed system, where the evolution of the states of the agents is fully determined by (5.1), the dynamics of the aggregate function $V(x(t))$ can be upper bounded by [107]:

$$\begin{aligned}\dot{V}(x(t)) &= \frac{2}{n}x^T(t)\dot{x}(t) \\ &= \frac{2}{n}x^T(t)\left(\beta A - \delta I\right)x(t) - \beta x^T(t)X(t)Ax(t) \qquad (5.6)\\ &\leq \frac{2}{n}x^T(t)\left(\beta A - \delta I\right)x(t) \\ &\leq 2(\beta\lambda_1(A) - \delta)V(x(t)), \qquad (5.7)\end{aligned}$$

where we used the fact that $x^T(t)X(t)Ax(t) \geq 0$ since $x_i \in [0,1]$. Clearly, the right-hand size of (5.7) is negative if $\beta\lambda_1(A) - \delta < 0$, which corresponds to the stability condition (5.2). Finally, by applying the Comparison Lemma, we have that $V(x(t))$ satisfies:

$$V(x(t)) \leq V_0 + \int_0^t 2(\beta\lambda_1(A) - \delta)V(x(\tau))d\tau, \qquad (5.8)$$

where V_0 denotes the value of $V(x(t))$ at the time $t = 0$.

To illustrate the behavior of the upper bound (5.8), we consider a graph with $n = 11$ agents represented in Fig. 5.1, whose adjacency matrix is given by

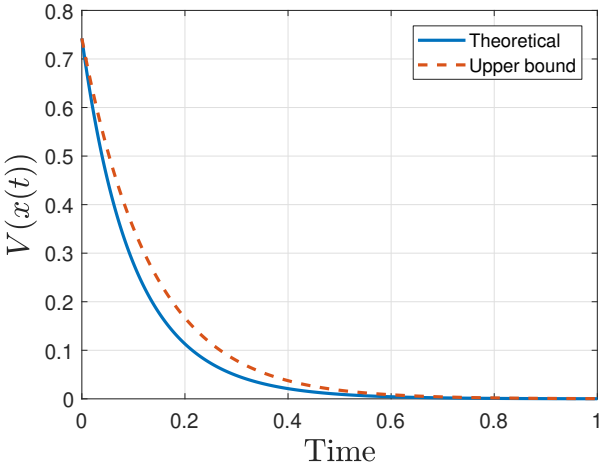


Figure 5.2: Evolution of the aggregate function $V(x(t))$ defined in (5.3) in continuous time for the graph determined by (5.9) with $n = 11$ agents, $\beta = 0.5$ and $\delta = 5.3$. The solid blue line corresponds to the simulation of $V(x(t))$ while the dashed red line is the upper bound (5.8).

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.9}$$

Figure 5.2 presents the evolution of $V(x(t))$ for the graph determined by (5.9) with $\beta = 0.5$ and $\delta = 5.3$, where the dashed red line corresponds to the right-hand side of (5.8).

5.2.1 Linearized model

If the probabilities of infection of the agents $x_i(t)$ remain close to the equilibrium point $x = 0$, it is possible to analyze the behavior of the epidemic by using the linearization of (5.1) near the origin [108]:

$$\dot{x}(t) = (\beta A - \delta I)\, x(t). \tag{5.10}$$

This linearization is exponentially stable [109] under condition (5.2). Another motivation to consider the linear model in the analysis is that the frequency of occurrence of arrivals, departures and replacements is low compared to the natural evolution of the epidemic. For this reason, the system will remain close to the equilibrium point most of the time and it will only be perturbed at specific time instants. However, the jumps must be small since the analysis is local.

Similar to the nonlinear model, the aggregate function $V(x(t))$ for (5.10) also satisfies (5.7) and (5.8). However, unlike the nonlinear case, the dynamics of $V(x(t))$ can also be lower bounded by:

$$\begin{aligned}\dot{V}(x(t)) &= \frac{2}{n}x^T(t)\dot{x}(t) \\ &= \frac{2}{n}x^T(t)\left(\beta A - \delta I\right)x(t) \\ &\geq 2(\beta\lambda_n(A) - \delta)V(x(t)),\end{aligned} \tag{5.11}$$

where $\lambda_n(A)$ is the smallest eigenvalue of A. By applying the Comparison Lemma, the aggregate function $V(x(t))$ also satisfies:

$$V(x(t)) \geq V_0 + \int_0^t 2(\beta\lambda_n(A) - \delta)V(x(\tau))d\tau.$$

The lower bound (5.11) will be used for the derivation of a tighter upper bound for the variance of $V(x(t))$ when the linearized model (5.10) is considered for the analysis.

5.3 Replacements

5.3.1 What is a replacement?

In this subsection, we analyze the case when in addition to the continuous evolution of the states of the individuals according to (5.1), the system also experiments replacements of individuals consisting of a departure of an agent, which stops spreading the infection in the network, followed immediately by the addition of a new infected agent.

In Chapters 3 and 4 we define the replacements of agents as the change of the cost function associated with the agents. However, in the case of a SIS epidemic model, the replacement of an agent is related to the change of the probability of infection associated with the agent [1].

[1]The state $x_i(t)$ in a SIS epidemic can also be interpreted as the fraction of a population infected at time t. However, under this interpretation, the replacement of x_i may not represent the real behavior of a epidemic since a change of the state could imply that a considerable fraction of the population became healthy abruptly.

Definition 5.1 *We say that an agent $j \in \{1, \dots, n\}$ is replaced at time t if:*

$$x^+ = [x_1^-, \cdots, x_{j-1}^-, x_r, x_{j+1}^-, \cdots, x_n^-]^T, \qquad x_r \neq x_j^-,$$

where $x^- = x(t^-)$ is the state of the system before the replacement, $x^+ = x(t^+)$ is the state of the system after the replacement and x_r is the value of the new agent.

Assumption 5.1 *The set of edges E of the network remains invariant for all time t.*

In the following proposition we analyze the variation of the aggregate function $V(x(t))$ under a replacement event where the replaced agent is chosen uniformly and the value of the new agent is determined by a continuous distribution with support in the interval $[0, 1]$, with mean m and variance σ^2.

Proposition 5.2 *During the replacement of an agent, the aggregate function $V(x(t))$ defined in* (5.3) *satisfies:*

$$\mathbb{E}\left[V(x(t^+))\right] - \mathbb{E}\left[V(x(t^-))\right] = \frac{1}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t^-))\right]\right). \tag{5.12}$$

Proof: Let us assume that an agent j is replaced. We begin by computing the conditional expectation of $V(x(t^+))$ given $x(t^-)$ and the value of the replaced agent Θ. Then, it holds

$$\begin{aligned}
\mathbb{E}\left[V(x(t^+))|x(t^-), \Theta\right] &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i} x_j^2(t^-) + \frac{1}{n}\Theta^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left(\left|\left|x(t^-)\right|\right|^2 - x_i^2 + \Theta^2\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\left|\left|x(t^-)\right|\right|^2 - \frac{1}{n^2}\sum_{i=1}^{n} x_i^2(t^-) + \frac{1}{n^2}\sum_{i=1}^{n}\Theta^2 \\
&= \frac{1}{n}\left|\left|x(t^-)\right|\right|^2 - \frac{1}{n^2}\left|\left|x(t^-)\right|\right|^2 + \frac{\Theta^2}{n} \\
&= V(x(t^-)) - \frac{1}{n}V(x(t^-)) + \frac{\Theta^2}{n}.
\end{aligned}$$

Then, we compute the total expectation and we obtain

$$\mathbb{E}\left[\mathbb{E}\left[V(x(t^+))|x(t^-), \Theta\right]\right] = \mathbb{E}\left[V(x(t^-))\right] - \frac{1}{n}\mathbb{E}\left[V(x(t^-))\right] + \frac{\sigma^2 + m^2}{n},$$

which completes the proof. ■

Remark 5.1 (Other systems) *The proof of Proposition 5.2 did not use the equation* (5.1)*, only the definition of* $V(x(t))$*. Therefore, this proof remains valid for other systems.*

Along this work, we make the following assumption about the occurrence of replacements in the evolution of the epidemic.

Assumption 5.2 (Replacement process) *The replacements instants are determined by a homogeneous Poisson process with rate* $\lambda_r > 0$*. During a replacement, an agent* $j \in \{1, \dots, n\}$ *is chosen uniformly and it is assigned a new state* x_r *determined by a random variable* Θ*, which takes values according to a continuous distribution with support in the interval* $[0, 1]$*, with mean* m *and variance* σ^2*.*

5.3.2 Pure replacements $\dot{x}(t) = 0$

To analyze the behavior of the aggregate function during the jumps, first, we consider the setting where $V(x(t))$ does not change between the jumps. Notice that the jumps of the state $x(t)$ corresponds to the jump of the aggregate function $V(x(t))$. In this case, the value of $V(x(t))$ is given by:

$$\begin{aligned} V(x(t)) &= V_0 + \Delta V_1 + \Delta V_2 + \cdots + \Delta V_{N_t} \\ &= V_0 + \sum_{k=1}^{N_t} \Delta V_k \\ &= V_0 + \sum_{k=1}^{N_t} \left(V(x(T_k^+)) - V(x(T_k^-)) \right) \end{aligned}$$

where ΔV_k is the size of the jump of the function $V(x(t))$ at the T_k jump time of the Poisson process N_t and

$$V(x(T_k^-)) = \lim_{t \uparrow T_k} V(x(t)) \qquad \text{and} \qquad V(x(T_k^+)) = \lim_{t \downarrow T_k} V(x(t)).$$

The aggregate function satisfies $V(x(t^-)) = V(x(t^+))$ for almost all t, except in a countable number of time jumps. Figure 5.3 presents a sample path of $V(x(t))$ where the jumps are determined by the Poisson process N_t. Then, we have that the aggregate function can be expressed as [110]:

$$V(x(t)) = V_0 + \int_0^t (V(x(\tau^+)) - V(x(\tau^-)))dN_\tau := V_0 + \sum_{k=1}^{N_t} \left(V(x(T_k^+)) - V(x(T_k^-)) \right).$$

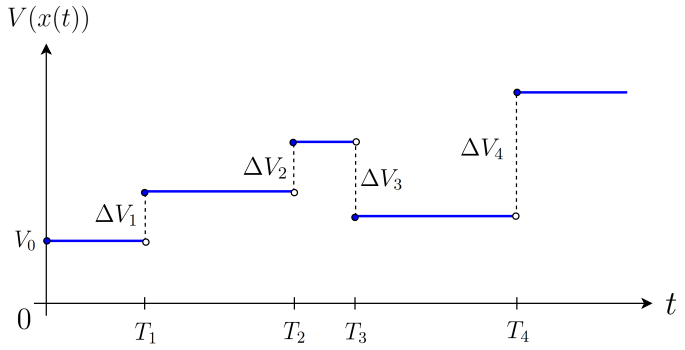


Figure 5.3: Sample path of $V(x(t))$.

An homogeneous Poisson process satisfies $\mathbb{E}[N_t] = \lambda_r t$, which gives us $\mathbb{E}[dN_t] = d\mathbb{E}[N_t] = \lambda_r dt$ [110]. Then, it holds

$$\mathbb{E}[V(x(t))] = V_0 + \int_0^t \mathbb{E}\left[V(x(\tau^+)) - V(x(\tau^-))\right] \lambda_r d\tau,$$

which can be expressed as the ODE:

$$d\mathbb{E}[V(x(t))] = \lambda_r \mathbb{E}\left[V(x(t^+)) - V(x(t^-))\right] dt. \tag{5.13}$$

Proposition 5.3 *Under Assumption 5.2 and inter-event dynamics $\dot{x}(t) = 0$, the aggregate function $V(x(t))$ defined in* (5.3) *satisfies:*

$$\mathbb{E}[V(x(t))] = \left(V_0 - \sigma^2 - m^2\right) e^{-\frac{\lambda_r}{n}t} + (\sigma^2 + m^2). \tag{5.14}$$

Proof: The result is obtained by applying (5.12) in (5.13) and solving the corresponding ODE:

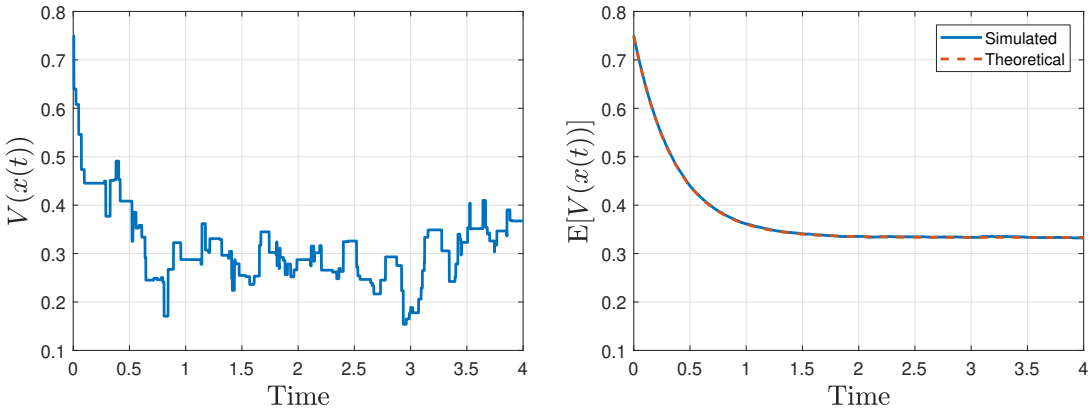
$$\frac{d}{dt}\mathbb{E}[V(x(t))] = -\frac{\lambda_r}{n}\mathbb{E}[V(x(t))] + \frac{\lambda_r}{n}(\sigma^2 + m^2).$$

■

Proposition 5.3 shows that in a system subject only to replacements, the expected value of the aggregate function is bounded and its asymptotic value is:

$$\lim_{t\to\infty} \mathbb{E}[V(x(t))] = \sigma^2 + m^2, \tag{5.15}$$

which corresponds to the second moment of the distribution used to generate the new values of the replaced agents. The rate of the Poisson process λ_r only has an influence in the rate of convergence of $V(x(t))$ such that a large value of λ_r will



(a) One realization (b) Expectation

Figure 5.4: Evolution of the aggregate function $V(x(t))$ defined in (5.3) in a system subject only to replacements with $\lambda_r = 30$. In the right plot, the solid blue line corresponds to the computation of $V(x(t))$ considering 10000 realizations of the process while the dashed red line is (5.14).

guarantee a fast convergence. The asymptotic value of $V(x(t))$ is independent of the rate λ_r.

Figure 5.4(a) presents the evolution of $V(x(t))$ for one realization of a replacement process following a Poisson process with intensity $\lambda_r = 30$ and where the new value of the agent is taken from a uniform distribution $\mathcal{U}[0, 1]$. Figure 5.4(b) presents the evolution of $\mathbb{E}[V(x(t))]$ computed considering 10000 realizations of the process.

5.3.3 SIS epidemic with replacements on a given graph

Now, we consider the behavior of the aggregate function in continuous time subject to replacements at time instants determined by a Poisson process. Since the solution to (5.6) is unique and the jumps are determined by a homogeneous Poisson process, the aggregate function $V(x(t))$ is given by [111]:

$$V(x(t)) = V_0 + \int_0^t \frac{2}{n} \left(x^T(\tau) \left(\beta A - \delta I \right) x(\tau) - \beta x^T(\tau) X(\tau) A x(\tau) \right) d\tau + \int_0^t \Delta V_k dN_\tau, \tag{5.16}$$

which can be expressed as the SDE

$$dV(x(t)) = \frac{2}{n} \left(x^T(t) \left(\beta A - \delta I \right) x(t) - \beta x^T(t) X(t) A x(t) \right) dt + \Delta V_k dN_t. \tag{5.17}$$

Notice that by denoting $\omega(x(t)) = \frac{2}{n}\left(x^T(t)\left(\beta A - \delta I\right)x(t) - \beta x^T(t)X(t)Ax(t)\right)$, (5.16) can be expressed as

$$V(x(t)) = V_0 + \int_0^{T_1} \omega(x(\tau))d\tau + \int_{T_1}^{T_2} \omega(x(\tau))d\tau + \cdots + \sum_{k=1}^{N_t} \Delta V_k,$$

which shows that $V(x(t))$ jumps at the time instants T_k determined by the Poisson process, corresponding to the sum, and between these time intervals, $V(x(t))$ evolves according to (5.6), corresponding to the integrals.

Even if the solution of (5.17) is unique, every realization of the stochastic process can be completely different and the analysis of the asymptotic behavior is not appropriate because the limit $\lim_{t\to\infty} V(x(t))$ does not exist due to the continuous replacements. For this reason, we are interested in the statistical properties of (5.17), corresponding to $\mathbb{E}\left[V(x(t))\right]$.

Theorem 5.1 (Expectation) *Consider a SIS epidemic satisfying the stability condition* (5.2)*. Under Assumptions 5.1 and 5.2, the aggregate function* $V(x(t))$ *defined in* (5.3) *satisfies:*

$$\limsup_{t\to\infty} \mathbb{E}\left[V(x(t))\right] \leq \frac{\lambda_r(\sigma^2 + m^2)}{\lambda_r + 2\delta n - 2\beta n\lambda_1(A)}. \tag{5.18}$$

Proof: We take the expectation of both sides of (5.17) and we obtain:

$$\begin{aligned} d\mathbb{E}\left[V(x(t))\right] =& \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta A - \delta I\right)x(t) - \beta x^T(t)X(t)Ax(t)\right)\right]dt \\ &+ \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right)dt, \end{aligned}$$

where we applied the Dominated Convergence Theorem to interchange the expectation and the derivative $\mathbb{E}\left[dV\right] = d\mathbb{E}\left[V\right]$ and we used the fact that $\mathbb{E}\left[\Delta V\right] = \frac{1}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V\right]\right)$ according to Proposition 5.2. Then we obtain the following ODE:

$$\begin{aligned} \frac{d}{dt}\mathbb{E}\left[V(x(t))\right] &= \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta A - \delta I\right)x(t) - \beta x^T(t)X(t)Ax(t)\right)\right] \\ &\quad + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\ &\leq 2(\beta\lambda_1(A) - \delta)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \qquad (5.19) \\ &= \left(\frac{2\beta n\lambda_1(A) - 2\delta n - \lambda_r}{n}\right)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r(\sigma^2 + m^2)}{n}. \qquad (5.20) \end{aligned}$$

By the Comparison Lemma, we can guarantee that $\mathbb{E}[V(x(t))]$ is upper bounded by the solution of the right-hand side of (5.20). Then, it holds

$$\begin{aligned}
\mathbb{E}[V(x(t))] &\leq \mathbb{E}[V_0]\, e^{\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)t} \\
&\quad + e^{\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)t}\int_0^t \frac{\lambda_r(\sigma^2+m^2)}{n} e^{-\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)\tau} d\tau \\
&= \mathbb{E}[V_0]\, e^{\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)t} \\
&\quad + \frac{\lambda_r(\sigma^2+m^2)}{\lambda_r+2\delta n-2\beta n\lambda_1(A)} e^{\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)t}\left[e^{-\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)\tau}\right]\Big|_0^t \\
&= \left(\mathbb{E}[V_0] - \frac{\lambda_r(\sigma^2+m^2)}{\lambda_r+2\delta n-2\beta n\lambda_1(A)}\right) e^{\left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right)t} \\
&\quad + \frac{\lambda_r(\sigma^2+m^2)}{\lambda_r+2\delta n-2\beta n\lambda_1(A)}. \qquad (5.21)
\end{aligned}$$

Finally, the result follows by taking the limit $t \to \infty$ in (5.21). ■

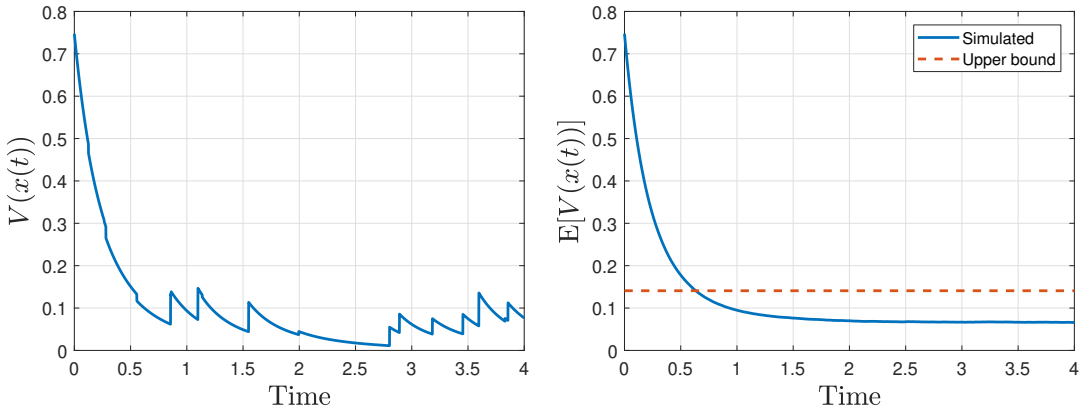
Clearly, the bound (5.18) is a function of the parameters of the SIS epidemic (i.e., β, δ and $\lambda_1(A)$) and the parameters of the replacement process (i.e., λ_r, σ^2 and m^2). Similarly to (5.15), the bound is proportional to σ^2+m^2, but in this case the rate of the Poisson process λ_r also appears as a multiplication factor. The upper bound for the rate of convergence of the SIS epidemic in a closed system determined by (5.7) appears in the denominator of (5.18), but this value is increased by the rate of the Poisson process λ_r. From (5.21), it can be seen that a large value of λ_r will increase the rate of convergence of $V(x(t))$, such that the convergence will be fast, but the asymptotic value will also increase. When $\lambda_r \to \infty$, the bound (5.18) is given by σ^2+m^2, which coincides with the result of Proposition 5.3 since the process will be characterized only by replacements.

Notice that the right-hand side in (5.19) corrresponds to the expectation of a SDE of the form:

$$dV = f(V)dt + g(V)dN_t, \qquad (5.22)$$

which is also known as a *Poisson Counter driven Stochastic Differential Equation* [111,112]. The solution of (5.19) is a cadlag function corresponding to a Piecewise-deterministic Markov Process (PDMP) and can also be studied using the theory of PDMP [113].

Figure 5.5 presents the computation of $V(x(t))$ for the graph determined by (5.9) with $n = 11$ agents, $\beta = 0.5$, $\delta = 2$ and $\lambda_r = 7$. The states of the replaced agents are generated with a uniform distribution with $m = 1/2$ and $\sigma^2 = 1/12$. The expectation of $V(x(t))$ has been computed by considering 10000 realizations of the process.



(a) One realization (b) Expectation

Figure 5.5: Evolution of the aggregate function $V(x(t))$ defined in (5.3) for the graph determined by (5.9) with $n = 11$ agents, $\beta = 0.5$, $\delta = 2$ and $\lambda_r = 7$. In the right plot, the solid blue line corresponds to the computation of $\mathbb{E}[V(x(t))]$ considering 10000 realizations of the process while the dashed red line is the upper bound (5.18).

Second order analysis

Although the asymptotic behavior of the expected aggregate function $\mathbb{E}[V(x(t))]$ provides useful information of a system subject to the replacements of agents, it does not give any information about the deviation of single realizations from the expected value. For this reason, it is necessary to analyze the variance of the aggregate function, denoted by $\mathrm{Var}(V(x(t)))$, and derive an upper bound. In this case, we use the Itô Lemma formulated for SDEs of the form (5.22) [111, Section 2.3].

Lemma 5.1 (Itô Lemma) *For a continuous differentiable function $\phi : \mathbb{R} \to \mathbb{R}$:*

$$d\phi(V) = \langle \frac{d\phi}{dV}, f(V) \rangle dt + [\phi(V + g(V)) - \phi(V)]\, dN_t,$$

where $f(V)$ and $g(V)$ are the mappings in the SDE (5.22) and N_t is a Poisson process.

We first analyze the change of $\mathbb{E}[V^2(x(t))]$ during a replacement.

Lemma 5.2 *During the replacement of an agent, the aggregate function $V(x(t))$ defined in* (5.3) *satisfies:*

$$\mathbb{E}\left[V^2(x(t^+))\right] - \mathbb{E}\left[V^2(x(t^-))\right] \leq -\frac{2}{n}\mathbb{E}\left[V^2(x(t^-))\right] + \frac{\mathbb{E}\left[\Theta^4\right]}{n^2} + \left(\frac{2\mathbb{E}\left[\Theta^2\right](n-1)+1}{n^2}\right)\mathbb{E}\left[V(x(t^-))\right],$$

where Θ is the new value of a replaced agent according to Assumption 5.2.

Proof: During the replacement of an agent x_j, the function V^2 satisfies:

$$\begin{aligned} V^2(x(t^+)) &= \left(V(x(t^-)) + \frac{\Theta^2}{n} - \frac{x_j^2(t^-)}{n}\right)^2 \\ &= V^2(x(t^-)) + \frac{\Theta^4}{n^2} + \frac{x_j^4(t^-)}{n^2} + \frac{2\Theta^2}{n}V(x(t^-)) \\ &\quad - \frac{2x_j^2(t^-)}{n}V(x(t^-)) - \frac{2\Theta^2 x_j^2(t^-)}{n^2}. \end{aligned}$$

Then, we compute the conditional expectation:

$$\begin{aligned} \mathbb{E}\left[V^2(x(t^+))|x(t^-),\Theta\right] &= \frac{1}{n}\sum_{j=1}^{n}\left(V^2(x(t^-)) + \frac{\Theta^4}{n^2} + \frac{x_j^4(t^-)}{n^2} + \frac{2\Theta^2}{n}V(x(t^-))\right. \\ &\quad \left. - \frac{2x_j^2(t^-)}{n}V(x(t^-)) - \frac{2\Theta^2 x_j^2(t^-)}{n^2}\right) \\ &= V^2(x(t^-)) + \frac{\Theta^4}{n^2} + \frac{1}{n^3}\sum_{j=1}^{n}x_j^4(t^-) + \frac{2\Theta^2}{n}V(x(t^-)) \\ &\quad - \frac{2V(x(t^-))}{n^2}\sum_{j=1}^{n}x_j^2(t^-) - \frac{2\Theta^2}{n^3}\sum_{j=1}^{n}x_j^2(t^-) \\ &\leq V^2(x(t^-)) + \frac{\Theta^4}{n^2} + \frac{1}{n^3}\sum_{j=1}^{n}x_j^2(t^-) + \frac{2\Theta^2}{n}V(x(t^-)) \\ &\quad - \frac{2V^2(x(t^-))}{n} - \frac{2\Theta^2}{n^2}V(x(t^-)) \\ &= V^2(x(t^-)) - \frac{2V^2(x(t^-))}{n} + \frac{\Theta^4}{n^2} \\ &\quad + \left(\frac{1}{n^2} + \frac{2\Theta^2}{n} - \frac{2\Theta^2}{n^2}\right)V(x(t^-)) \end{aligned}$$

Finally, we compute the total expectation to get the desired result. ■

Proposition 5.4 (Second moment) *Consider a SIS epidemic satisfying the stability condition* (5.2)*. Under Assumptions 5.1 and 5.2, the aggregate function* $V(x(t))$ *defined in* (5.3) *satisfies:*

$$\limsup_{t\to\infty} \mathbb{E}\left[V^2(x(t))\right] \leq \frac{\lambda_r \mathbb{E}\left[\Theta^4\right](\lambda_r + 2\delta n - 2\beta n \lambda_1(A)) + \lambda_r^2 \mathbb{E}\left[\Theta^2\right](2\mathbb{E}\left[\Theta^2\right](n-1)+1)}{2n(\lambda_r + 2\delta n - 2\beta n\lambda_1(A))^2}. \tag{5.23}$$

Proof: By applying Itô Lemma with $\phi(V) = V^2$, we obtain:

$$\begin{aligned} dV^2(x(t)) &\leq \langle 2V(x(t)), 2(\beta\lambda_1(A) - \delta)V(x(t))\rangle dt + \left(V^2(x(t_k^+)) - V^2(x(t_k^-))\right) dN_t \\ &= 4(\beta\lambda_1(A) - \delta)V^2(x(t))dt + \left(V^2(x(t_k^+)) - V^2(x(t_k^-))\right) dN_t. \end{aligned} \tag{5.24}$$

We take the expectation in (5.24) and we get:

$$\begin{aligned} d\mathbb{E}\left[V^2(x(t))\right] &\leq 4(\beta\lambda_1(A) - \delta)\mathbb{E}\left[V^2(x(t))\right] dt \\ &\quad + \left(\mathbb{E}\left[V^2(x(t^+))\right] - \mathbb{E}\left[V^2(x(t^-))\right]\right) \lambda_r dt \\ &= 4(\beta\lambda_1(A) - \delta)\mathbb{E}\left[V^2(x(t))\right] dt \\ &\quad + \left(\left(\frac{2\mathbb{E}\left[\Theta^2\right](n-1)+1}{n^2}\right)\mathbb{E}\left[V(x(t))\right] + \frac{\mathbb{E}\left[\Theta^4\right]}{n^2} - \frac{2}{n}\mathbb{E}\left[V^2(x(t))\right]\right)\lambda_r dt \\ &= 2\left(\frac{2\beta n\lambda_1(A) - 2\delta n - \lambda_r}{n}\right)\mathbb{E}\left[V^2(x(t))\right] dt \\ &\quad + \lambda_r\left(\frac{2\mathbb{E}\left[\Theta^2\right](n-1)+1}{n^2}\right)\mathbb{E}\left[V(x(t))\right] dt + \frac{\lambda_r \mathbb{E}\left[\Theta^4\right]}{n^2} dt, \end{aligned}$$

where we use the Dominated Convergence Theorem to interchange the expectation and the derivative of $V^2(x(t))$ and we apply Lemma 5.2. Let us denote $a_V = 2\left(\frac{2\beta n\lambda_1(A) - 2\delta n - \lambda_r}{n}\right)$, $b_V = \lambda_r\left(\frac{2\mathbb{E}\left[\Theta^2\right](n-1)+1}{n^2}\right)$ and $c_V = \frac{\lambda_r \mathbb{E}\left[\Theta^4\right]}{n^2}$, so that we have

$$\frac{d}{dt}\mathbb{E}\left[V^2(x(t))\right] \leq a_V \mathbb{E}\left[V^2(x(t))\right] + b_V \mathbb{E}\left[V(x(t))\right] + c_V \tag{5.25}$$

The solution of (5.25) is given by:

$$\begin{aligned} \mathbb{E}\left[V^2(x(t))\right] &\leq e^{a_V t}\mathbb{E}\left[V_0^2\right] + e^{a_V t}\int_0^t e^{-a_V \tau}\left(b_V \mathbb{E}\left[V(x(\tau))\right] + c_V\right) d\tau \\ &= e^{a_V t}\mathbb{E}\left[V_0^2\right] + b_V e^{a_V t}\int_0^t e^{-a_V\tau}\mathbb{E}\left[V(x(\tau))\right] d\tau + c_V e^{a_V t}\int_0^t e^{-a_V\tau} d\tau \\ &= e^{a_V t}\mathbb{E}\left[V_0^2\right] + b_V e^{a_V t}\int_0^t e^{-a_V\tau}\mathbb{E}\left[V(x(\tau))\right] d\tau - \frac{c_V}{a_V}\left(1 - e^{a_V t}\right). \end{aligned} \tag{5.26}$$

We recall that by (5.21), the function $\mathbb{E}[V]$ satisfies:

$$\mathbb{E}[V(x(t))] \leq d_V e^{h_V t} + m_V,$$

where $d_V = \mathbb{E}[V_0] - \frac{\lambda_r(\sigma^2+m^2)}{\lambda_r+2\delta n-2\beta n\lambda_1(A)}$, $h_V = \left(\frac{2\beta n\lambda_1(A)-2\delta n-\lambda_r}{n}\right) t$, $m_V = \frac{\lambda_r(\sigma^2+m^2)}{\lambda_r+2\delta n-2\beta n\lambda_1(A)}$. Then $\mathbb{E}[V^2(x(t))]$ satisfies:

$$\begin{aligned}
\mathbb{E}\left[V^2(x(t))\right] &\leq e^{a_V t}\mathbb{E}\left[V_0^2\right] - \frac{c_V}{a_V}\left(1-e^{a_V t}\right) + b_V e^{a_V t}\int_0^t e^{-a_V\tau}\left(d_V e^{h_V\tau}+m_V\right)d\tau \\
&= e^{a_V t}\mathbb{E}\left[V_0^2\right] - \frac{c_V}{a_V}\left(1-e^{a_V t}\right) + b_V d_V e^{a_V t}\int_0^t e^{(h_V-a_V)\tau}d\tau \\
&\quad + b_V m_V e^{a_V t}\int_0^t e^{-a_V\tau}d\tau \\
&= e^{a_V t}\mathbb{E}\left[V_0^2\right] - \frac{c_V}{a_V}\left(1-e^{a_V t}\right) + \frac{b_V d_V}{h_V-a_V}\left(e^{h_V t}-e^{a_V t}\right) \\
&\quad - \frac{b_V m_V}{a_V}\left(1-e^{a_V t}\right) \\
&= \left(\mathbb{E}\left[V_0^2\right] + \frac{c_V}{a_V} - \frac{b_V d_V}{h_V-a_V} + \frac{b_V m_V}{a_V}\right)e^{a_V t} + \frac{b_V d_V}{h_V-a_V}e^{h_V t} \\
&\quad - \left(\frac{c_V}{a_V}+\frac{b_V m_V}{a_V}\right). \qquad (5.27)
\end{aligned}$$

Finally, the result follows by taking the limit $t \to \infty$ in (5.27), where we recall that a_V and h_V are negative. ■

The bound (5.23) is derived using (5.18), which corresponds to the asymptotic behavior of $\mathbb{E}[V(x(t))]$. Unlike the previous result (5.18), the value of (5.23) is also affected by the fourth moment of Θ, while (5.18) only depends on the second moment ($\mathbb{E}[\Theta^2] = \sigma^2 + m^2$).

Proposition 5.4 can be used to derive an upper bound for the variance:

$$\begin{aligned}
\limsup_{t\to\infty} \operatorname{Var}(V(x(t))) &= \limsup_{t\to\infty}\left(\mathbb{E}\left[V^2(x(t))\right] - \left(\mathbb{E}[V(x(t))]\right)^2\right) \\
&\leq \limsup_{t\to\infty}\mathbb{E}\left[V^2(x(t))\right].
\end{aligned}$$

Although the result is a bit conservative since we are neglecting the effect of $-\left(\mathbb{E}[V(x(t))]\right)^2$, we have an estimate of the deviation of the realizations from the expected value. However, a stronger bound can be obtained for the linearized model (5.10) by using the lower bound (5.11).

Corollary 5.1 (Variance) *Consider the linearized model of a SIS epidemic satisfying the stability condition* (5.2)*. Under Assumptions 5.1 and 5.2, the aggregate*

function $V(x(t))$ *defined in* (5.3) *satisfies:*

$$\limsup_{t\to\infty} \mathrm{Var}(V(x(t))) \leq \frac{\lambda_r \mathbb{E}\left[\Theta^4\right](\lambda_r + 2\delta n - 2\beta n \lambda_1(A)) + \lambda_r^2 \mathbb{E}\left[\Theta^2\right](2\mathbb{E}\left[\Theta^2\right](n-1)+1)}{2n(\lambda_r + 2\delta n - 2\beta n \lambda_1(A))^2} - \frac{\lambda_r^2 \left(\mathbb{E}\left[\Theta^2\right]\right)^2}{(\lambda_r + 2\delta n - 2\beta n \lambda_n(A))^2}. \tag{5.28}$$

Proof: We compute the expectation of the aggregate funciton and we obtain

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[V(x(t))\right] &= \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta A - \delta I\right)x(t) - \beta x^T(t) X(t) A x(t)\right)\right] \\
&\quad + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\
&\geq 2(\beta\lambda_n(A) - \delta)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\
&= \left(\frac{2\beta n \lambda_n(A) - 2\delta n - \lambda_r}{n}\right)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r(\sigma^2 + m^2)}{n},
\end{aligned} \tag{5.29}$$

where we used (5.11) for the lower bound. By the Comparison Lemma, we can guarantee that $\mathbb{E}\left[V\right]$ is lower bounded by the solution of the right-hand side of (5.29). We follow similar steps as in the proof of Theorem 5.1 and we obtain:

$$\liminf_{t\to\infty} \mathbb{E}\left[V(x(t))\right] \geq \frac{\lambda_r(\sigma^2 + m^2)}{\lambda_r + 2\delta n - 2\beta n \lambda_n(A)}. \tag{5.30}$$

The result follows by applying (5.23) and (5.30) in the definition of the variance:

$$\mathrm{Var}(V(x(t))) = \mathbb{E}\left[V^2(x(t))\right] - \left(\mathbb{E}\left[V(x(t))\right]\right)^2,$$

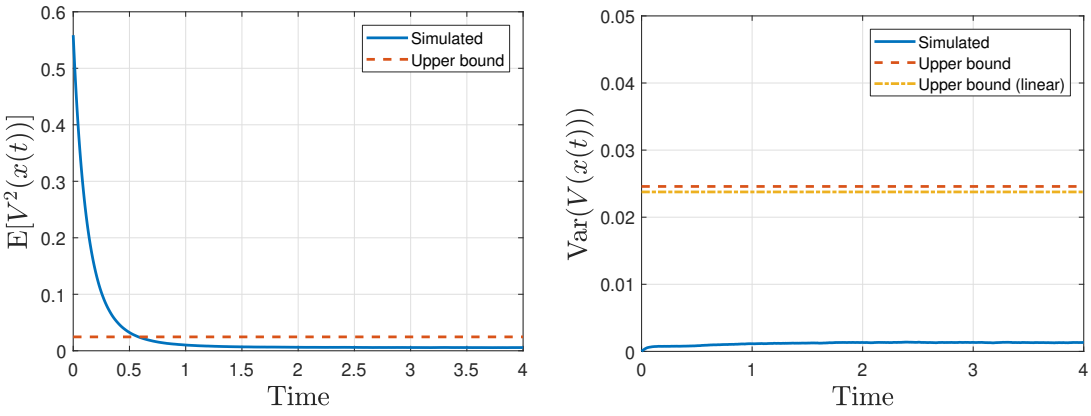
and remembering that $\mathbb{E}\left[V^2(x(t))\right] \geq \left(\mathbb{E}\left[V(x(t))\right]\right)^2$. ■

Notice that the right-hand side of (5.30) is 0 only when $\lambda_r = 0$, so that this lower bound can be used to prove that the linearized model of a SIS epidemic under replacements is not stochastically stable in the sense that it does not satisfy [114]:

$$\int_0^\infty \mathbb{E}\left[V(x(t))\right] dt < \infty.$$

In Figure 5.6, we present the computations of the second moments of $V(x(t))$ for the graph determined by (5.9) with $n = 11$ agents, $\beta = 0.5$, $\delta = 2$ and $\lambda_r = 7$ considering 10000 realizations of the process and a uniform distribution with $m = 1/2$ and $\sigma^2 = 1/12$ for the states of the replaced agents. Figure 5.6(a) shows the evolution of $\mathbb{E}\left[V^2(x(t))\right]$ where the dashed red line corresponds to the upper bound (5.23). Finally, Fig. 5.6(b) presents the evolution of $\mathrm{Var}(V(x(t)))$

where the dashed red line is the upper bound (5.23) and the dash-dotted yellow line corresponds to the upper bound (5.28) considering the linearized model of the SIS epidemic. Notice that $\mathrm{Var}(V(x(t)))$ does not go to zero with t and that the upper bound obtained using the linearized model is also valid for the nonlinear system for this particular network topology and parameters of the infection.



(a) Second moment: $\mathbb{E}\left[V^2(x(t))\right]$ (b) Second central moment: $\mathrm{Var}(V(x(t)))$

Figure 5.6: Evolution of the second moments of the aggregate function $V(x(t))$ defined in (5.3) for the graph determined by (5.9) with $n = 11$ agents, $\beta = 0.5$, $\delta = 2$ and $\lambda_r = 7$. In the left plot, the solid blue line corresponds to the estimation of $\mathbb{E}\left[V^2(x(t))\right]$ while the dashed red line is the upper bound (5.23). In the right plot, the solid blue line corresponds to the estimation of $\mathrm{Var}(V(x(t)))$ while the dashed red line is the upper bound (5.23) and the dash-dotted yellow line is the upper bound (5.28) considering the linearized model. The simulated values were obtained considering 10000 realizations of the process.

5.3.4 SIS epidemic with replacements on a graph sampled from graphon

From Proposition 5.1 and Theorem 5.1, it is clear that the analysis of the stability and behavior of a SIS epidemic in OMAS is linked to the largest eigenvalue $\lambda_1(A)$ of the adjacency matrix of the graph. However, in the case of large graphs, it is difficult to obtain a complete representation of the network because of the presence of noise and errors in data and the constant evolution of links and nodes. Also, even if it is possible to obtain a good knowledge of network topology, their sheer size prevents the full simulation or analysis of the dynamics, or the computation of relevant network properties, because of limitations in computational resources.

In this subsection, we consider that the graphs over which the epidemic takes place are sampled from graphons using the Definition A.1 and we derive results equivalent to Proposition 5.1 and Theorem 5.1 based on the spectrum of the graphon operator. (For a description of the most important properties of graphons see Section A.2). In the perspective approximating graphs with graphons, it is convenient to consider sequences of graphs parametrized by their size n. Therefore, it is reasonable to assume that parameters δ or β be also dependent on n: upon need, we shall emphasize this dependence by writing δ_n and β_n.

Proposition 5.5 ([115]) *Consider a SIS epidemic over a graph G with n nodes sampled from a piecewise Lipschitz graphon W. Then for n large enough, the SIS epidemic will reach the disease-free state with probability at least $1-\nu$ if:*

$$\delta_n > n\beta_n(\lambda_1(T_W) + \phi_n), \tag{5.31}$$

where $\lambda_1(T_W)$ is the largest eigenvalue of the graphon operator T_W.

Proof: Condition (5.2) is equivalent to $\delta > \beta\lambda_1(A)$. By Lemma A.1, with n large enough and probability at least $1-\nu$:

$$|n\lambda_1(A) - \lambda_1(T_W)| \leq \phi_n,$$

$$\lambda_1(A) \leq n(\lambda_1(T_W) + \phi_n),$$

which yields the desired result. ■

Since almost surely $\lambda_1(A)$ grows linearly with n (except for the trivial graphon $W = 0$), in order to ensure a uniform bound on $\lambda_1(A)\beta_n/\delta_n$, we will assume that the infection strength β_n/δ_n satisfies

$$\frac{\beta_n}{\delta_n} = \frac{1}{n}\frac{\bar{\beta}}{\bar{\delta}} \tag{5.32}$$

for some positive constants $\bar{\beta}$ and $\bar{\delta}$.

Assuming that $\delta_n = \bar{\delta}$ is constant and $\beta_n = n^{-1}\bar{\beta}$ means that, as the graph grows in size, the healing rate (which depends on each individual) remains constant, whereas the infection rate decreases. This natural scaling law is also chosen in [116]. Indeed, on dense graphs this assumption means that in larger graphs, even though there are more potential interactions, the average strength of the connections is suitably adjusted: this fact accounts for natural limitations in the rates of contact between individuals.

In the case of replacements, it is natural to assume that the frequency of these events increases with the number of agents in the system. For this reason, we assume:

$$\lambda_r = n\bar{\lambda}_r, \tag{5.33}$$

for some positive constant $\bar{\lambda}_r$.

Theorem 5.2 *Consider a SIS epidemic over a graph G with n nodes sampled from a piecewise Lipschitz graphon W with $\delta_n = \bar{\delta}$ and $\beta_n = n^{-1}\bar{\beta}$ satisfying condition* (5.31)*. Under Assumptions 5.1 and 5.2, for n large enough, with probability at least $1-\nu$, the aggregate function $V(x(t))$ defined in* (5.3) *satisfies:*

$$\limsup_{t\to\infty} \mathbb{E}\left[V(x(t))\right] \leq \frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r + 2\bar{\delta} - 2\bar{\beta}(\lambda_1(T_W)+\phi_n)}. \tag{5.34}$$

Proof: The aggregate function $V(x(t))$ satisfies:

$$dV(x(t)) = \frac{2}{n}\left(x^T(t)\left(\beta_n A - \delta_n I\right)x(t) - \beta_n x^T(t)X(t)Ax(t)\right)dt + \Delta V_k dN_t \tag{5.35}$$

We compute the expectation and we obtain

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[V(x(t))\right] &= \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta_n A - \delta_n I\right)x(t) - \beta_n x^T(t)X(t)Ax(t)\right)\right] \\
&\quad + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\
&\leq \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta_n A - \delta_n I\right)x(t)\right)\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\
&\leq 2(\beta_n\lambda_1(A) - \delta_n)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right)
\end{aligned} \tag{5.36}$$

Then, with probability at least $1-\nu$:

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[V(x(t))\right] &\leq 2\left(\beta_n n(\lambda_1(T_W)+\phi_n) - \delta_n\right)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r}{n}\left(\sigma^2+m^2-\mathbb{E}\left[V(x(t))\right]\right) \\
&= \left(\frac{2\beta_n n^2(\lambda_1(T_W)+\phi_n) - 2\delta_n n - \lambda_r}{n}\right)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r(\sigma^2+m^2)}{n}, \\
&= \left(2\bar{\beta}(\lambda_1(T_W)+\phi_n) - 2\bar{\delta} - \bar{\lambda}_r\right)\mathbb{E}\left[V(x(t))\right] + \bar{\lambda}_r(\sigma^2+m^2),
\end{aligned} \tag{5.37}$$

which replaces 5.20. The rest of the proof follows the same step as in the proof of Theorem 5.1 ■

Remark 5.2 (Given graph) *In the case of a given graph with a Poisson rate λ_r given by* (5.33)*, the upper bound* (5.18) *is equivalent to:*

$$\limsup_{t\to\infty} \mathbb{E}\left[V(x(t))\right] \leq \frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r + 2\delta - 2\beta\lambda_1(A)},$$

which is independent of the number of agents n.

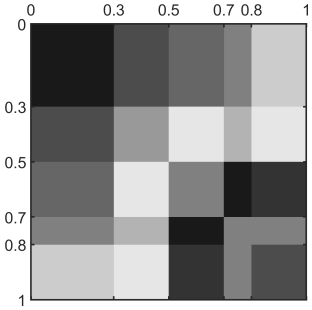


Figure 5.7: Pixel diagram of the stochastic block model graphon.

Remark 5.3 (Asymptotics for $n \to \infty$) *When we let n go to infinity, if ν is constant or if $\nu = 1/n^{\alpha}$ for some constant α, we can see that the upper bound given in Theorem 5.2 goes to $\frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r+2\bar{\delta}-2\bar{\beta}\lambda_1(T_W)}$. Hence, by choosing $\alpha > 1$ and applying Borel-Cantelli Lemma, we obtain that for n large enough, with probability 1:*

$$\mathbb{E}\left[V(x(\infty))\right]_n \leq \frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r+2\bar{\delta}-2\bar{\beta}(\lambda_1(T_W)+\phi_n)} \to \frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r+2\bar{\delta}-2\bar{\beta}\lambda_1(T_W)} \text{ as } n \to \infty,$$

where $\mathbb{E}\left[V(x(\infty))\right]_n := \limsup_{t\to\infty} \mathbb{E}\left[V(x(t))\right]$.

Remark 5.4 (Stochastic latent variables) *An alternative procedure for the generation of the complete weighted graph $\bar{G}$ from graphons is the use of stochastic latent variables proposed in [117]. All the results of this section easily extend to this stochastic sampling by considering the appropriate values of the probability ν and the bound ϕ_n, which are established in [117].*

To illustrate the results on graphons, we consider the stochastic block model graphon W_{SB} with pixel diagram in Fig. 5.7, where the values of the blocks are:

$$W_{\text{SB}} = \begin{bmatrix} 0.9 & 0.7 & 0.6 & 0.5 & 0.2 \\ 0.7 & 0.4 & 0.1 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.5 & 0.9 & 0.8 \\ 0.5 & 0.3 & 0.9 & 0.5 & 0.5 \\ 0.2 & 0.1 & 0.8 & 0.5 & 0.7 \end{bmatrix}.$$

The largest eigenvalue of the graphon operator $T_{W_{\text{SB}}}$ is $\rho_{W_{\text{SB}}} = 0.5275$. We assume $\beta_n = \bar{\beta}/n$, with $\bar{\beta} = 1.5$, and $\nu = 0.02$, which satisfies condition (A.2c) for $n \geq 40$. The value of the rate of replacements is $\bar{\lambda}_r = 2$ while $\delta_n = \bar{\delta} = \bar{\beta}(\lambda_1(T_W) + \phi_{40})$ is selected, satisfying condition (5.31). We compute the value of the upper bound $\mathbb{E}\left[V(x(\infty))\right]_n$ for $40 \leq n \leq 1000$ and we present the results in Figure 5.8 where the dashed red line corresponds to the limit when $n \to \infty$.

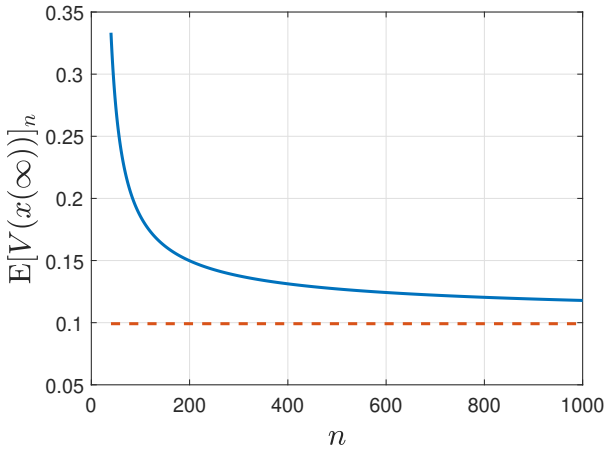


Figure 5.8: Upper bound $\mathbb{E}\left[V(x(\infty))\right]_n$ as a function of n for the stochastic block model graphon W_{SB} with $\bar{\beta} = 1.5$, $\bar{\delta} = \bar{\beta}(\lambda_1(T_W) + \phi_{40})$ and $\bar{\lambda}_r = 2$. The solid blue line corresponds to $\mathbb{E}\left[V(x(\infty))\right]_n$ while the dashed red line corresponds to the limit when $n \to \infty$.

5.3.5 SIS epidemic with replacements with new neighborhoods determined by graphons

Now, we consider the case of a SIS epidemic where during a replacement, the topology of the network may also change. The dynamics of the system is given by

$$\dot{x}(t) = \left(\beta_n A(t) - \delta_n I\right) x(t) - \beta_n X(t) A(t) x(t), \tag{5.38}$$

where the adjacency matrix $A(t)$ is time-varying.

We assume that at time $t = 0$, the graph of the network is sampled from a graphon W according to Definition A.1 and the network interconnections are preserved between the switching times determined by a Poisson process with intensity λ_r, which implies that the dynamics (5.38) also remains invariant. However, we consider that during the replacement of an agent the topology of the network also changes.

Assumption 5.3 *During the replacement of an agent, the connections of all the agents are sampled again according to the piecewise Lipschitz graphon W.*

The motivation for this particular choice of the change of the network topology is that stochastic block model graphons are used to model community structures in social networks. In this case, the replacement of an agent may impact all the network such that connections between individuals may also change but preserve the same pattern determined by the graphon.

By definition, the aggregate function $V(x(t))$ satisfies:

$$dV(x(t)) = \frac{2}{n}\left(x^T(t)\left(\beta_n A(t) - \delta_n I\right)x(t) - \beta_n x^T(t)X(t)A(t)x(t)\right)dt + \Delta V_k dN_t,$$

where $A(t)$ is sampled from a graphon. Unlike Theorem 5.2 which is based on the computation of the expectation given a simple graph G sampled from a graphon, in this section, we compute the expectation of all the graphs that can form the sequence G_t on which the process takes place. Then, the expectation is given by

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[V(x(t))\right] = \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta_n A(t) - \delta_n I\right)x(t) - \beta_n x^T(t)X(t)A(t)x(t)\right)\right] \\ + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right). \end{aligned} \tag{5.39}$$

We consider an approximation of (5.39) by assuming independence between the evolution of $x(t)$ and the adjacency matrix $A(t)$ such that we have:

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\left[V(x(t))\right] &= \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta_n \bar{A}_r - \delta_n I\right)x(t) - \beta_n x^T(t)X(t)\bar{A}_r x(t)\right)\right] \\ &\quad + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \qquad (5.40) \\ &\leq \mathbb{E}\left[\frac{2}{n}\left(x^T(t)\left(\beta_n \bar{A}_r - \delta_n I\right)x(t)\right)\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right) \\ &\leq 2(\beta_n \lambda_1(\bar{A}_r) - \delta_n)\mathbb{E}\left[V(x(t))\right] + \frac{\lambda_r}{n}\left(\sigma^2 + m^2 - \mathbb{E}\left[V(x(t))\right]\right), \qquad (5.41)\end{aligned}$$

where $\bar{A}_r = \mathbb{E}\left[A(t)\right]$. Notice that $\bar{A}_r = \bar{A} - \bar{D}$, where $\bar{A}$ is the adjacency matrix of the complete weighted graph $\bar{G}$ according to Definition A.1 and $\bar{D} = \text{diag}[W(u_i, u_i)]$. Then by applying Lemma A.1, we have that for n large enough:

$$\lambda_1(\bar{A}_r) = \left|\left|\bar{A}_r\right|\right| \leq \left|\left|\bar{A}\right|\right| + \left|\left|\bar{D}\right|\right| \leq n\lambda_1(T_W) + n\vartheta_n + 1,$$

where $\vartheta_n = \frac{2\sqrt{L^2 - K^2 + Kn}}{n}$. Under this approximation, the stability condition for the SIS epidemic in expectation is given by:

$$\delta_n > n\beta_n(\lambda_1(T_W) + \vartheta_n + 1/n). \tag{5.42}$$

By following similar steps in the proof of Theorem 5.2 with (5.41) instead of (5.36), we formulate the following conjecture.

Conjecture 5.1 *Consider a SIS epidemic over a sequence of graphs G_t with n nodes sampled from a piecewise Lipschitz graphon W with $\delta_n = \bar{\delta}$ and $\beta_n = n^{-1}\bar{\beta}$*

satisfying condition (5.42). *Under Assumptions 5.2 and 5.3, for n large enough, the aggregate function* $V(x(t))$ *defined in* (5.3) *satisfies:*

$$\limsup_{t\to\infty} \mathbb{E}\left[V(x(t))\right] \leq \frac{\bar{\lambda}_r(\sigma^2+m^2)}{\bar{\lambda}_r + 2\bar{\delta} - 2\bar{\beta}(\lambda_1(T_W)+\vartheta_n+1/n)}. \tag{5.43}$$

Theorem 5.2 holds with probability $1-\nu$ because of the upper bound for $\lambda_1(A)$ in terms of the spectral radius of the graphon operator T_W, as per Lemma A.1. However, in the case of Conjecture 5.1, it can be seen from (5.40), that the potential result is based on an epidemic process with replacements over a time-invariant graph with adjacency matrix $\bar{A}_r$, which is sampled from the graphon using deterministic latent variables and does not involve the use of random variables.

Figure 5.9 presents the computation of $V(x(t))$ for graphs sampled from the stochastic block model W_{SB} with $n = 50$ agents, $\bar{\beta} = 1.5$, $\bar{\delta} = \bar{\beta}(\lambda_1(T_W) + \vartheta_{50} + 1/50)$ and $\bar{\lambda}_r = 2$ where the states of new agents are drawn from a uniform distribution with $m = 1/2$ and $\sigma^2 = 1/12$, and the expected values were calculated considering 1000 realizations of the process. We can observe that the behavior of the approximated model (5.40) in dashed red line is similar to the behavior of the real model (5.39) in solid blue line, and both seem to converge to the same limit when $t \to \infty$. Furthermore, the upper bound (5.43) in dash-dotted yellow line is valid for both processes, which supports the conjecture.

5.4 Conclusion

In this chapter, we analyzed a SIS epidemic in OMAS when replacements occur at time instants determined by a Poisson process. We defined an aggregate function to analyze the epidemic and derived upper bounds for the expectation and variance when $t \to \infty$. We derived a similar result in expectation when graphs are sampled from graphons. When the connections of the network can change according to a graphon during the occurrence of replacements, we performed a preliminary analysis based on the assumption of independence of some random variables and formulated a conjecture for the asymptotic value in expectation of the aggregate function, which is corroborated by simulations.

Due to the versatility of graphons to generate different graph topologies with a given number of nodes, they are an important tool to be used in the case of OMAS over dynamic networks, which are more realistic, since it is unlikely that in a large network, during the replacement of an agent, the connections remain invariant. Graphons also appear to be a promising tool for analyzing OMAS under arrivals of agents, where the connections of the new agents can be well defined by generating edges based on the graphon and the values of the latent variables associated with the agents of the network.

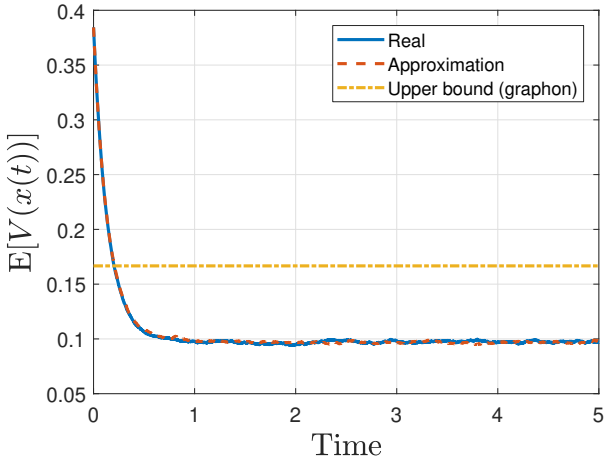


Figure 5.9: Evolution of the aggregate function $V(x(t))$ defined in (5.3) for graphs sampled from the stochastic block model graphon W_{SB} with $n = 50$ agents, $\bar{\beta} = 1.5$, $\bar{\delta} = \bar{\beta}(\lambda_1(T_W) + \vartheta_{50} + 1/50)$ and $\bar{\lambda}_r = 2$. The solid blue line corresponds to the estimation of $\mathbb{E}\left[V(x(t))\right]$ while the dashed red line corresponds to the approximation (5.40) of the real process and the dash-dotted yellow line is the conjectured upper bound (5.43). The expected values were computed by considering 1000 realizations of the process.

For future work, it would be important to extend the results by considering potential departures and arrivals of agents, which would imply a time-varying size of the network. In this case, the arrivals and departures of agents can be studied as *birth* and *death* events respectively and the theory of birth-death process [118] can provide useful tools for the analysis of more general situations. This approach has been used in [25], where a linear death process has been considered for the departure of agents. In this line of research, it is necessary to determine if the aggregate function used for replacement remains valid for possible changes on the number of agents of the system or if a new definition is needed to condensate the behavior of the system in a more appropriate way.

In the case of graphs sampled from graphons, we only considered the case of deterministic latent variables. However, the use of stochastic latent variables could provide a more appropriate sampling procedure during arrivals where the connections of the current agents in the network are preserved and new connections are established only with new agents.

Chapter 6

Conclusion and Future Work

Open multi-agent systems (OMAS) allow to model more realistic dynamic networks where agents can enter, leave or being replaced in the system. Unlike single changes in the network topology, these variations in the set of agents make the analysis more complex since the dimension of the system may change and traditional concepts associated to dynamical systems might not be applied in many cases. In this thesis, we analyzed OMAS considering different problems associated with classical multi-agent systems. In all of these problems, we selected a set of assumptions to make the problem realistic and tractable at the same time, and performed an analysis based on the study of an appropriate scalar quantity independent of the size of the system.

The problem of consensus was analyzed in Chapter 2. We considered a randomized algorithm where the interaction matrix is drawn from a common distribution and we studied the effect of additive noise in the performance of the algorithm. To perform the analysis, we considered the mean squared error and derived a closed form and bounds based on the eigenvalues of the expected matrices. This approach allowed us to study the problem in an OMAS when we consider a finite superset of agents, where at each time instant, only a subset of agents is active, determined by independent Bernoulli random variables associated with each agent in the system. We expressed the bounds derived for the general case as a function of the parameters associated with the OMAS and provided some expressions for particular graph topologies.

In Chapter 3, we focused on the weighted gradient descent algorithm to solve the well-known problem of resource allocation (RA) in multi-agent systems where the transmission of information is subject to packet losses. First, we proposed a set of assumptions to guarantee that the problem can be solved at least in expectation, which justifies the application of the algorithm. Then, we proposed two metrics to measure the deviation from the constraint and the ideal expected cost, and we derived bounds for the metrics which are proportional to the difference between

the initial cost function and the ideal cost function. Next, we extended the results to the case of OMAS, where the cost functions of the agents can be replaced with a certain probability. Unlike the closed system, we showed that the constraint violation error diverges and does not remain bounded due to the accumulation of errors. This result showed that a well-defined metric, which is bounded in a closed system, could diverge in an OMAS.

Chapter 4 continued the analysis of the RA problem in OMAS but considering the application of the RCD algorithm to perform the optimization. In this chapter we analyzed the problem from a perspective of dynamical systems, focusing on the stability of the time-varying minimizer. We proved that for a general graph topology, it is not possible to guarantee linear convergence in a closed system, and we performed an analysis based on a norm induced by a matrix associated with the network. Some bounds for the difference between the minimizers where proved using different approaches related to the properties of β-smooth and α-strongly convex functions. We proved that the RCD algorithm applied to an OMAS subject to replacements will converge inside a region determined by the parameters of the system. Additionally, by considering a simple setting, we proposed an alternative approach based on online optimization, where the states of the agents may also change during the replacements. We defined three metrics based on a cooperative and selfish strategy and derived some bounds for them which grow linearly with time. We provided a detailed analysis to justify this linear behavior, which is mainly due to the accumulation of errors during the replacements, similar to the constraint violation metric studied in Chapter 3.

In Chapter 5, we analyzed a SIS epidemic in OMAS where agents can be replaced during the evolution of the dynamics. Unlike the previous chapters, we used a continuous time formulation of the system, which made the analysis more complex since the states of the agents can change between replacements. Inspired by Lyapunov functions, we defined an appropriate aggregate function for the analysis of the system and used a stochastic differential equation (SDE) to model its evolution. We performed an analysis in expectation of the aggregate function and derived an upper bound for its variance. An extension of the results was provided when the graphs are sampled from graphons. In the case of possible changes of the connections during replacements following a sampling procedure from graphons, we presented a preliminary analysis of this scenario and formulated a conjecture about the asymptotic behavior of the aggregate function that was corroborated by simulations.

6.1 Future research

Although different types of problems and applications of OMAS were analyzed in this thesis, there are still many problems, questions, extensions and improvements that can be considered to develop a further understanding of the main characteristics of OMAS and the best tools to address more complex scenarios.

Regarding Chapter 2, the results were derived by considering independence in the stochastic matrices. However, scenarios where only few agents change their states from active to inactive or vice versa still need to be explored. In addition, it would be important to analyze an alternative performance metric where the normalization term includes only the number of active agents, which could provide a more informative result for OMAS.

In Chapter 3, considering that the upper bound derived for the cost function metric presents a certain level of conservatism, it would be important to improve this bound to obtain a tighter result. Regarding open systems, it would be important to improve the lower bound for the constraint violation metric. Finally, an analysis of the cost function metric for OMAS needs to be developed.

Chapter 4 analyzed the RCD algorithm in OMAS only in the case of possible replacements of agents while the budget remains the same. The case of time-varying budget and weights is a natural continuation of this work. A possible time-varying size of the system is also an interesting direction for future work, where the size of the system can change in time due to arrivals and departures of agents. In the case of the metrics inspired from online optimization, it would be important to extend the results for more general graph topologies and develop the corresponding lower bounds.

In Chapter 5, we analyzed a SIS epidemic in OMAS subject to replacements. It would be natural to extend the work to the case of scenarios where arrivals and departures can take place such that the size of the system can change in time [25]. In the case of graphons, there is a conjecture that still needs to be verified to validate the results obtained in simulations. Also, for graphs sampled from graphons, it would be important to consider more complex sampling procedures, to study scenarios where the connections of the new agent do not follow a specific pattern and the connections of the other agents are preserved.

Finally, regarding OMAS in general, many questions and problems remain open. For instance, in this thesis we have considered only two approaches for the analysis of OMAS: time-invariant finite superset and fixed state-space. In the case of multi-mode multi-dimensional systems, preliminary works only considered the case of a fixed number of spaces with different dimensions [41, 42], which does not seem suitable for a general OMAS where the dimension of the system could take an arbitrary value. For this reason, further work is needed in this field to extend this approach for systems with arbitrary dimensions. Regarding infinite

dimensional dynamical systems, during this thesis two works focused on the use of graphons for the study of epidemics [115] and the analysis of the spectrum of the Laplacian matrix corresponding to sampled graphs [62] were developed. Since most of the dynamics in multi-agent systems are formulated in terms of the Laplacian matrix, it is important to extend the theory of graphons for this matrix through the definition of an appropriate operator. In this line of research, the authors in [119] studied the case of consensus in first-order graphon models, while in [120], the authors analyzed the spectrum of the Laplacian matrix for particular types of graphons. However, the relation between the spectrum of the Laplacian operator associated with an arbitrary graphon and the eigenvalues of the Laplacian matrix of a sampled graph is still under analysis.

In most of the works, OMAS have been analyzed only in the case of undirected graphs. In [31], the authors analyzed the case of switching between multi-agent systems of different dimensions where agents interact through digraphs. However, the approach used in this work disregards the impacts of the processes associated with arrivals, departures and replacements, limiting the potential applications of the work. Since most of the results derived form OMAS rely on the symmetry of the Laplacian matrix, it seems possible that new techniques of analysis are required for the extension of the results to the case of digraphs. In addition, the definition of appropriate graph topologies is critical in this case, specially in the case of arrivals where the connections of the new agents could play an important role in the preservation of the connectivity of the network.

Agents with more general dynamics like double integrators or nonlinear dynamics (Kuramoto oscillators, Lotka-Volterra, etc.) should be analyzed in the context of arrivals/departures of agents, which would extend the analysis of OMAS to more complex phenomena. In this line of research, the approach used in Chapter 5 seems promising for these cases, where Lyapunov functions could be used to study the systems in terms of a SDE. Alternatively, the theory of stochastic hybrid systems appear as an important framework that can be used for the analysis of OMAS when the objective is the determination of conditions to guarantee the stability of OMAS [121, 122].

Heterogeneity is an important characteristic that should be explored in future works. In simple settings, it is assumed that all the agents that interact in the system have the same dynamics. However, in many scenarios, this assumption could not be appropriate as in the case of robotics, where different types of robots need to interact to achieve a desired goal [123]. Several approaches could be used for this analysis, including the study of emergent dynamics [124, 125], which has provided promising results in the case of switching systems with heterogeneous agents [126].

Finally, as it was remarked in [127, 128], OMAS look as a promising tool to handle opinion dynamics where agents are constantly entering and leaving the system as in the case of social networks, forums, discussions, etc. Models based on opinion-dependent connectivity like the Hegselmann-Krause model [129–131] and k-nearest neighbors [132] seem to be more suitable for OMAS where the interactions are determined only by the values of the opinions. This type of interconnections simplifies the definition of the network topology in OMAS in the case of arrivals and departures, since it is not necessary to define the connections of new agents or reconfigure the connections of the network when agents leave the system.

Appendix A

Mathematical background

A.1 Graphs

A graph can be defined as a pair $G = (V, E)$ where $V = \{1, \ldots, n\}$ is a set of vertices or nodes and $E \subseteq \{(i, j) \in V \times V : i \neq j\}$ is the set of edges. An undirected graph has edges with no direction and a simple graph does not contain self-loops or multiedges. A graph is connected if each node is reachable from any other node via a path (a sequence of adjacent edges).

The adjacency matrix of a graph $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ is defined by $a_{ij} > 0$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. This matrix is real symmetric non-negative with real eigenvalues [133]. If the graph has weights w_{ij} in the edges, we design the entries of the adjacency matrix as $a_{ij} = w_{ij}$. If we do not consider weights in the edges, we define the entries of the adjacency matrix as $a_{ij} = 1$ for all the edges. We consider the order of the eigenvalues of the adjacency matrix as: $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$.

Another matrix associated with a graph is the Laplacian matrix defined as $L = [\ell_{ij}]$, where $\ell_{ij} = \sum_{j=1}^{n} a_{ij}$ if $i = j$, and $\ell_{ij} = -a_{ij}$ if $i \neq j$. According to its definition, L is diagonally dominant ($\ell_{ii} \geq \sum_{j=1}^{n} |\ell_{ij}|$), and all the eigenvalues $\lambda_i(L)$ have nonnegative real part. Furthermore, 0 is always an eigenvalue of the Laplacian matrix. We consider the order of the eigenvalues of the Laplacian matrix as: $\lambda_1(L) = 0 \leq \lambda_2(L) \leq \ldots \leq \lambda_n(L)$. Specifically, this matrix is symmetric and positive semidefinite if the graph is undirected. If the graph is connected, we have $\lambda_2(L) > 0$. The Laplacian matrix can be expressed as:

$$L = D - A, \tag{A.1}$$

where $D = \text{diag}[d_1, \ldots, d_n]$ and d_i is the degree of the node i (i.e., d_i is the ith row-sum of A).

A.2 Graphons

We denote by $\mathcal{W}$ the space of all bounded symmetric measurable functions $W : [0,1]^2 \to \mathbb{R}$. The elements of this space are called *kernels* given their connection to integral operators. The set of all kernels $W \in \mathcal{W}$ such that $0 \leq W \leq 1$ is denoted by $\mathcal{W}_0$ and its elements are called *graphons*, whose name is a contraction of graph-function. By analogy with degrees in finite graphs, the *degree* function of a graphon is defined as $d_W(x) := \int_0^1 W(x,y)\,dy$. In order to consider differences between graphons, we shall sometimes work in the set $\mathcal{W}_1$ of kernels W such that $-1 \leq W \leq 1$.

Every function $W \in \mathcal{W}$ defines an integral operator $T_W : L^2[0,1] \to L^2[0,1]$ by:

$$(T_W f)(x) := \int_0^1 W(x,y) f(y)\,dy.$$

This operator is compact and has a discrete spectrum with 0 as the only accumulation point. Every nonzero eigenvalue has finite multiplicity [39]. A graphon W is said to have finite rank if the spectrum of the associated operator contains a finite number of nonzero eigenvalues [39].

A *step graphon* is a graphon defined as a step function. A function is called a *step function* if there is a partition $S_1 \cup \cdots \cup S_k$ of $[0,1]$ into measurable sets such that W is constant on every product set $S_i \times S_j$ where the sets S_i are the steps of W. This type of graphon is also called *stochastic block model graphon* because of its relation to stochastic block models [134]. Step graphons are finite rank graphons with a rank at most equal to the number of steps. Also, graphons expressed as a finite sum of products of integrable functions have finite rank [117].

Each graph G has an associated step graphon W_G obtained by considering a uniform partition of $[0,1]$ into the intervals B_i^n, where $B_i^n = [(i-1)/n, i/n)$ for $i = 1, \ldots, n-1$ and $B_n^n = [(n-1)/n, 1]$ such that:

$$W_G(x,y) := \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \mathbb{1}_{B_i^n}(x) \mathbb{1}_{B_j^n}(y),$$

where $\mathbb{1}_A(x)$ is the indicator function. The operator associated to the step graphon is

$$(T_{W_G} f)(x) := \sum_{j=1}^{n} a_{ij} \int_{B_j^n} f(y)\,dy \;\text{ for any } x \in B_i^n$$

and the spectrum of T_{W_G} consists of the normalized spectrum of the graph (i.e., $\lambda_i(T_{W_G}) = \lambda_i(A)/n$), together with infinitely many zeros.

A graphon is usually visualized with a pixel picture, where each point $(x,y) \in [0,1]^2$ is colored with a grey level representing $W(x,y)$. For a step graphon asso-

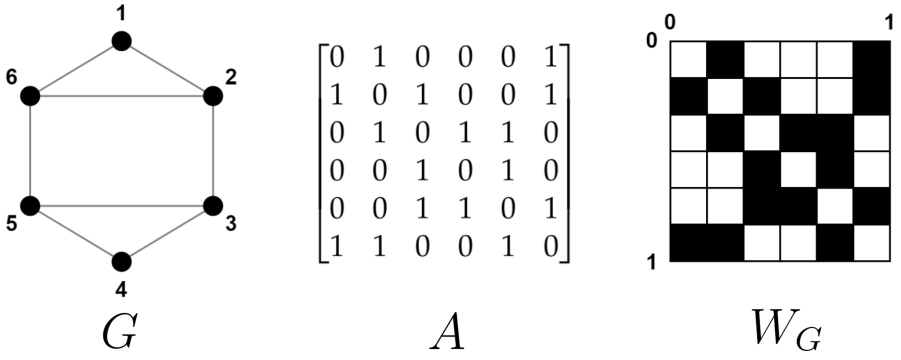


Figure A.1: Graph G, adjacency matrix A and step graphon W_G.

ciated to a graph G, we visualize a 0 as a small white square and a 1 as a small black square as we can appreciate in Fig. A.1.

A.2.1 Norms

In the study of kernels, various norms are relevant to consider [39, 117, 135]. For $1 \leq p < \infty$, we define the L^p norm of a kernel as

$$\|W\|_{L^p} := \left(\int_{[0,1]^2} |W(x,y)|^p dx\, dy \right)^{1/p}$$

and its *cut norm* by

$$\|W\|_{\square} := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y)\, dx\, dy \right|.$$

For $W \in \mathcal{W}_1$, we have the following inequalities between L^p norms and the cut norm:

$$\|W\|_{\square} \leq \|W\|_{L^1} \leq \|W\|_{L^2} \leq \|W\|_{L^1}^{1/2} \leq 1.$$

By considering the operator T_W associated to a kernel $W \in \mathcal{W}$, we can define the operator norm:

$$|||T_W||| := \sup_{\substack{f \in L^2[0,1] \\ \|f\|_{L^2}=1}} \|T_W f\|_{L^2}.$$

For graphons, the operator norm is equal to the largest eigenvalue of the operator: $|||T_W||| = \lambda_1(T_W)$. For the elements of $\mathcal{W}_1$, the cut and operator norms are related by:

$$\|W\|_{\square} \leq |||T_W||| \leq \sqrt{8} \|W\|_{\square}^{1/2}.$$

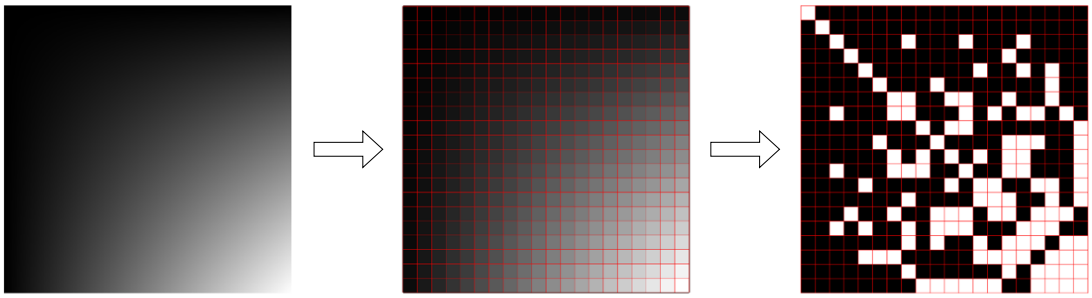


Figure A.2: Sampling graphon procedure.

Finally, we can define the Hilbert-Schmidt norm of the operator as:

$$\|T_W\|_{\text{HS}}^2 := \sum_i |\lambda_i(T_W)|^2,$$

which corresponds to the Schatten 2-norm of T_W. For all $W \in \mathcal{W}$, $\|T_W\|_{\text{HS}}$ is finite (i.e., kernel operators are Hilbert-Schmidt operators), and moreover $\|T_W\|_{\text{HS}} = \|W\|_{L^2}$.

A.2.2 Sampling and Approximation

A graphon W can be used to generate random graphs using a sampling method [39].

Definition A.1 (Sampled Graph [117]) *Given a graphon W and a size $n \in \mathbb{Z}_{>0}$, we say that the graph G is sampled from W if it is obtained through:*

1. ***Complete weighted graph** $\bar{G}$: fixing deterministic latent variables $\{u_i = \frac{i}{n}\}_{i=1}^n$, we generate the complete weighted graph $\bar{G}$ with n vertices, whose adjacency matrix is defined as: $\bar{A}(i,j) = W(u_i, u_j)$ for all i, j in $\{1, \dots, n\}$.*

2. ***Simple graph** G: taking n vertices $\{1, \dots, n\}$ and randomly adding undirected edges between vertices i and j independently with probability $W(u_i, u_j)$ for all $i > j$.*

Definition A.2 (Piecewise Lipschitz graphon [117]) *Graphon W is said to be piecewise Lipschitz if there exists a constant L and a sequence of non-overlapping intervals $I_k = [\alpha_{k-1}, \alpha_k)$ defined by $0 = \alpha_0 < \cdots < \alpha_{K+1} = 1$, for a finite*

non-negative integer K, such that for any k, l, any set $I_{kl} = I_k \times I_l$ and pairs (x_1, y_1) and $(x_2, y_2) \in I_{kl}$ we have that:

$$|W(x_1, y_1) - W(x_2, y_2)| \leq L(|x_1 - x_2| + |y_1 - y_2|).$$

Definition A.3 (Large enough n [117]) *Given a piecewise Lipschitz graphon W (as per Definition A.2) and $\nu < e^{-1}$, n is large enough if n satisfies the following conditions:*

$$\frac{2}{n} < \min_{k \in \{1,\dots,K+1\}} (\alpha_k - \alpha_{k-1}), \tag{A.2a}$$

$$\frac{1}{n} \log\left(\frac{2n}{\nu}\right) + \frac{1}{n}(2K + 3L) < \max_x d_W(x), \tag{A.2b}$$

$$ne^{-n/5} < \nu. \tag{A.2c}$$

Lemma A.1 (Theorem 1 [117]) *Let W be a piecewise Lipschitz graphon (as per Definition A.2) and G a graph with n nodes sampled from W. Then for n large enough with probability 1:*

$$\left|\left|\left|T_{W_{\bar{G}}} - T_W\right|\right|\right| \leq \frac{2\sqrt{L^2 - K^2 + Kn}}{n} =: \vartheta_n,$$

and with probability at least $1 - \nu$:

$$|||T_{W_G} - T_W||| \leq \sqrt{\frac{4\log(2n/\nu)}{n}} + \frac{2\sqrt{L^2 - K^2 + Kn}}{n} =: \phi_n.$$

By considering a constant value of ν, the difference between the graphons in the operator norm is bounded by: $|||T_{W_G} - T_W||| = O\left((\log n/n)^{1/2}\right)$.

Lemma A.2 *Let W be a piecewise Lipschitz graphon and G a graph with n nodes sampled from W. Then for n large enough, with probability at least $1 - \nu$:*

$$\|W - W_G\|_{L^2} \leq \sqrt[4]{2n}\sqrt{\phi_n}.$$

Proof: First, we have

$$\|W - W_G\|_\square \leq |||T_W - T_{W_G}||| \leq \phi_n.$$

It is easy to see that W_G is a step graphon that takes values in $\{0, 1\}$ and we can apply the inequality derived in [135, Remark 10.8], so that:

$$\|W - W_G\|_{L^1} \leq \sqrt{2n}\|W - W_G\|_\square \leq \sqrt{2n}\phi_n.$$

Applying the relation $\|W - W_G\|_{L^2} \leq \|W - W_G\|_{L^1}^{1/2}$ yields the desired result. ■

A.3 Poisson processes

A stochastic process $(N_t)_{t\geq 0}$ is a *counting process* if N_t represents the total number of *events* that have occurred up to time t. More specifically, a stochastic process is said to be a counting process if [136]:

1. $N_t \geq 0$
2. $N_t \in \mathbb{Z}_{\geq 0}$
3. If $s < t$, then $N_s \leq N_t$
4. For $s < t$, $N_t - N_s$ equals the number of events that have occurred in the interval $(s, t]$

A counting process has *independent increments* if the numbers of events that occur in disjoint time intervals are independent. A counting process has *stationary increments* if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval [136].

Definition A.4 (Poisson process [136]) *The counting process $(N_t)_{t\geq 0}$ is said to be a homogeneous Poisson process with rate $\lambda > 0$ if*

1. *$N_0 = 0$*
2. *The process has independent increments*
3. *The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$,*

$$P\left\{N_{t+s} - N_s = n\right\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1,$$

The last property implies that a Poisson process has stationary increments and that

$$\mathbb{E}\left[N_t\right] = \lambda t.$$

By construction, a homogeneous Poisson process only has a finite number of jumps on any interval and is a Markov counting process (i.e., has the Markov property), as the exponential distribution has the memoryless property.

Bibliography

[1] F. Bullo, *Lectures on Network Systems*, 1.6 ed. Kindle Direct Publishing, 2022.

[2] W. Ren and R. W. Beard, *Distributed Consensus in Multi-vehicle Cooperative Control*. London, UK: Springer, 2008, vol. 27, no. 2.

[3] W. Ren and Y. Cao, *Distributed Coordination of Multi-agent Networks: Emergent Problems, Models, and Issues*. London, UK: Springer, 2011, vol. 1.

[4] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[5] F. Dörfler and F. Bullo, "Synchronization in complex networks of phase oscillators: A survey," *Automatica*, vol. 50, no. 6, pp. 1539–1564, 2014.

[6] K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, 2015.

[7] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[8] 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.

[9] D. M. Cvetković, P. Rowlinson, and S. Simić, *An Introduction to the Theory of Graph Spectra*. Cambridge University Press, 2009.

[10] L. Moreau, "Stability of continuous-time distributed consensus algorithms," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 4, 2004, pp. 3998–4003.

[11] ——, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, 2005.

[12] Z. Lin, B. Francis, and M. Maggiore, "State agreement for continuous-time coupled nonlinear systems," *SIAM Journal on Control and Optimization*, vol. 46, no. 1, pp. 288–307, 2007.

[13] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[14] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, 2008.

[15] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 791–795, 2008.

[16] P. Frasca and J. M. Hendrickx, "On the mean square error of randomized averaging algorithms," *Automatica*, vol. 49, no. 8, pp. 2496–2501, 2013.

[17] J. M. Hendrickx and S. Martin, "Open multi-agent systems: Gossiping with deterministic arrivals and departures," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2016, pp. 1094–1101.

[18] M. Franceschelli and P. Frasca, "Proportional dynamic consensus in open multi-agent systems," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 900–905.

[19] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, "An integrated trust and reputation model for open multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 13, no. 2, pp. 119–154, 2006.

[20] I. Pinyol and J. Sabater-Mir, "Computational trust and reputation models for open multi-agent systems: a review," *Artificial Intelligence Review*, vol. 40, no. 1, pp. 1–25, 2013.

[21] T. Lykouris, V. Syrgkanis, and É. Tardos, "Learning and efficiency in games with dynamic population," in *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 2016, pp. 120–129.

[22] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

[23] R. Patel, P. Frasca, J. W. Durham, R. Carli, and F. Bullo, "Dynamic partitioning and coverage control with asynchronous one-to-base-station communication," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 1, pp. 24–33, 2016.

[24] 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)*, 2017, pp. 763–768.

[25] C. Monnoyer de Galland, S. Martin, and J. M. Hendrickx, "Modelling gossip interactions in open multi-agent systems," *arXiv preprint arXiv:2009.02970*, 2020.

[26] C. Monnoyer de Galland and J. M. Hendrickx, "Fundamental performance limitations for average consensus in open multi-agent systems," *IEEE Transactions on Automatic Control*, 2022.

[27] M. Abdelrahim, J. M. Hendrickx, and W. Heemels, "Max-consensus in open multi-agent systems with gossip interactions," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 4753–4758.

[28] M. Franceschelli and P. Frasca, "Stability of open multiagent systems and applications to dynamic consensus," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2326–2331, 2021.

[29] Z. A. Z. Sanai 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)*, 2019, pp. 3691–3697.

[30] D. Deplano, M. Franceschelli, and A. Giua, "Dynamic min and max consensus and size estimation of anonymous multi-agent networks," *IEEE Transactions on Automatic Control*, 2021.

[31] M. Xue, Y. Tang, W. Ren, and F. Qian, "Stability of multi-dimensional switched systems with an application to open multi-agent systems," *arXiv preprint arXiv:2001.00435*, 2020.

[32] 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.

[33] 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)*, 2021, pp. 514–520.

[34] L. Wang, C. Bernardo, Y. Hong, F. Vasca, G. Shi, and C. Altafini, "Achieving consensus in spite of stubbornness: time-varying concatenated Friedkin-Johnsen models," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 4964–4969.

[35] S. Grauwin and P. Jensen, "Opinion group formation and dynamics: Structures that last from nonlasting entities," *Phys. Rev. E*, vol. 85, p. 066113, Jun 2012.

[36] 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)*, 2020, pp. 4885–4890.

[37] F. Fagnani and P. Frasca, *Introduction to Averaging Dynamics over Networks*. Cham, Switzerland: Springer, 2018.

[38] B. Bollobás, *Random Graphs*, 2nd ed. Cambridge University Press, 2001.

[39] L. Lovász, *Large Networks and Graph Limits*. American Mathematical Society, 2012, vol. 60.

[40] H. Krivine, A. Lesné, and J. Treiner, "Discrete-time and continuous-time modelling: some bridges and gaps," *Mathematical Structures in Computer Science*, vol. 17, no. 2, pp. 261–276, 2007.

[41] E. I. Verriest, "Multi-mode multi-dimensional systems," in *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006, pp. 1268–1274.

[42] ——, "Pseudo-continuous multi-dimensional multi-mode systems," *Discrete Event Dynamic Systems*, vol. 22, no. 1, pp. 27–59, 2012.

[43] R. Goebel, R. G. Sanfelice, and A. R. Teel, *Hybrid dynamical systems*. Princeton University Press, 2012.

[44] N. Noroozi, A. Mironchenko, C. Kawan, and M. Zamani, "Small-gain theorem for stability, cooperative control and distributed observation of infinite networks," *arXiv preprint arXiv:2002.07085*, 2020.

[45] S. Gao and P. E. Caines, "Graphon control of large-scale networks of linear systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 10, pp. 4090–4105, 2020.

[46] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 225–230, 2009.

[47] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii, "Ergodic randomized algorithms and dynamics over networks," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 1, pp. 78–87, 2015.

[48] D. Acemoğlu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Mathematics of Operations Research*, vol. 38, no. 1, pp. 1–27, 2013.

[49] S. Kar and J. M. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, 2008.

[50] S. Bolognani, R. Carli, E. Lovisari, and S. Zampieri, "A randomized linear algorithm for clock synchronization in multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 7, pp. 1711–1726, 2015.

[51] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of parallel and distributed computing*, vol. 67, no. 1, pp. 33–46, 2007.

[52] A. Jadbabaie and A. Olshevsky, "Scaling laws for consensus protocols subject to noise," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1389–1402, 2018.

[53] Y. Yazıcıoğlu, W. Abbas, and M. Shabbir, "Structural robustness to noise in consensus networks: Impact of average degrees and average distances," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5444–5449.

[54] F. Garin and L. Schenato, "A survey on distributed estimation and control applications using linear consensus algorithms," in *Networked Control Systems*. London, U.K.: Springer, 2010, pp. 75–107.

[55] S. Patterson and B. Bamieh, "Leader selection for optimal network coherence," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 2692–2697.

[56] B. Bamieh, M. R. Jovanovic, P. Mitra, and S. Patterson, "Coherence in large-scale networks: Dimension-dependent limitations of local feedback," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2235–2249, 2012.

[57] J. Wang and N. Elia, "Distributed averaging under constraints on information exchange: emergence of Lévy flights," *IEEE Transactions on Automatic Control*, vol. 57, no. 10, pp. 2435–2449, 2012.

[58] S. Del Favero and S. Zampieri, "A majorization inequality and its application to distributed Kalman filtering," *Automatica*, vol. 47, no. 11, pp. 2438–2443, 2011.

[59] E. Lovisari, F. Garin, and S. Zampieri, "Resistance-based performance analysis of the consensus algorithm over geometric graphs," *SIAM Journal on Control and Optimization*, vol. 51, no. 5, pp. 3918–3945, 2013.

[60] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT press, 2022.

[61] C. Borgs, J. Chayes, L. Lovász, V. Sós, and K. Vesztergombi, "Limits of randomly grown graph sequences," *European Journal of Combinatorics*, vol. 32, no. 7, pp. 985–999, 2011.

[62] R. Vizuete, F. Garin, and P. Frasca, "The Laplacian spectrum of large graphs sampled from graphons," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1711–1721, 2021.

[63] T. Ibaraki and N. Katoh, *Resource Allocation Problems: Algorithmic Approaches*. MIT press, 1988.

[64] M. Patriksson, "A survey on the continuous nonlinear resource allocation problem," *European Journal of Operational Research*, vol. 185, 2008.

[65] A. Nedić, A. Olshevsky, and W. Shi, "Improved convergence rates for distributed resource allocation," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 172–177.

[66] T. T. Doan and C. L. Beck, "Distributed resource allocation over dynamic networks with uncertainty," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4378–4384, 2021.

[67] 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.

[68] 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*, 2021.

[69] 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.

[70] S. Liang, P. Yi, and Y. Hong, "Distributed Nash equilibrium seeking for aggregative games with coupled constraints," *Automatica*, vol. 85, pp. 179–185, 2017.

[71] J. F. Kurose and R. Simha, "A microeconomic approach to optimal resource allocation in distributed computer systems," *IEEE Transactions on Computers*, vol. 38, no. 5, pp. 705–717, 1989.

[72] Y. Xu, T. Han, K. Cai, Z. Lin, G. Yan, and M. Fu, "A distributed algorithm for resource allocation over dynamic digraphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2600–2612, 2017.

[73] Z. Deng, S. Liang, and Y. Hong, "Distributed continuous-time algorithms for resource allocation problems over weight-balanced digraphs," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3116–3125, 2018.

[74] 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.

[75] Y.-G. Hsieh, F. Iutzeler, J. Malick, and P. Mertikopoulos, "Multi-agent online optimization with delays: Asynchronicity, adaptivity, and optimism," *Journal of Machine Learning Research*, vol. 23, no. 78, pp. 1–49, 2022.

[76] S. Patterson, B. Bamieh, and A. El Abbadi, "Distributed average consensus with stochastic communication failures," in *2007 46th IEEE Conference on Decision and Control*. IEEE, 2007, pp. 4215–4220.

[77] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 102–133, 2009.

[78] P. Frasca and J. M. Hendrickx, "Large network consensus is robust to packet losses and interferences," in *2013 European Control Conference (ECC)*. IEEE, 2013, pp. 1782–1787.

[79] X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. Johansson, "Regret and cumulative constraint violation analysis for online convex optimization with long term constraints," in *International Conference on Machine Learning*. PMLR, 2021, pp. 11 998–12 008.

[80] 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.

[81] S. Abramovich and L.-E. Persson, "Some new estimates of the 'Jensen gap'," *Journal of Inequalities and Applications*, vol. 2016, no. 1, pp. 1–9, 2016.

[82] G. Carnevale, F. Farina, I. Notarnicola, and G. Notarstefano, "GTAdam: Gradient tracking with adaptive momentum for distributed online optimization," *arXiv preprint arXiv:2009.01745*, 2020.

[83] G. Carnevale, A. Camisa, and G. Notarstefano, "Distributed online aggregative optimization for dynamic multi-robot coordination," *IEEE Transactions on Automatic Control*, 2022.

[84] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM Journal on Optimization*, vol. 22, no. 2, p. 341–362, 2012.

[85] E. Hazan, "Introduction to Online Convex Optimization," *Foundations and Trends® in Optimization*, vol. 2, pp. 157–325, 01 2016.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, 2000.

[91] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Science & Business Media, 2006.

[92] Y. Nesterov, *Lectures on Convex Optimization.* Cham, Switzerland: Springer, 2018.

[93] S. Bubeck, "Convex Optimization: Algorithms and Complexity," *Foundations and Trends® in Machine Learning*, vol. 8, no. 3-4, 2015.

[94] C. Monnoyer 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 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 1701–1708.

[95] 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.

[96] A. J. Laub, *Matrix Analysis For Scientists And Engineers.* USA: Society for Industrial and Applied Mathematics, 2004.

[97] 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.

[98] 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.

[99] X. Li, L. Xie, and N. Li, "A survey of decentralized online learning," *arXiv preprint arXiv:2205.00473*, 2022.

[100] 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.

[101] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks.* Cham, Switzerland: Springer, 2017.

[102] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[103] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.

[104] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, 2016.

[105] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, "On the use of human mobility proxies for modeling epidemics," *PLoS computational biology*, vol. 10, no. 7, p. e1003716, 2014.

[106] S. Hazarie, D. Soriano-Paños, A. Arenas, J. Gómez-Gardeñes, and G. Ghoshal, "Interplay between population density and mobility in determining the spread of epidemics in cities," *Communications Physics*, vol. 4, no. 1, pp. 1–10, 2021.

[107] A. Khanafer, T. Başar, and B. Gharesifard, "Stability properties of infection diffusion dynamics over directed networks," in *53rd IEEE Conference on Decision and Control*, 2014, pp. 6215–6220.

[108] A. Khanafer and T. Basar, "An optimal control problem over infected networks," in *Proceedings of the International Conference of Control, Dynamic Systems, and Robotics, Ottawa, Ontario, Canada*, 2014.

[109] W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo, "On the dynamics of deterministic epidemic propagation over networks," *Annual Reviews in Control*, vol. 44, pp. 116–128, 2017.

[110] N. Privault, *Stochastic Finance: An Introduction with Market Examples*. CRC Press, 2013.

[111] R. Brockett, "Stochastic control," *Lecture Notes, Harvard University*, 2009.

[112] R. Brockett, W. Gong, and Y. Guo, "Stochastic analysis for fluid queueing systems," in *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No.99CH36304)*, vol. 3, 1999, pp. 3077–3082 vol.3.

[113] M. Jacobsen, *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser Boston, 2006.

[114] O. L. V. Costa, M. D. Fragoso, and M. G. Todorov, *Continuous-Time Markov Jump Linear Systems*. Springer Science & Business Media, 2012.

[115] R. Vizuete, P. Frasca, and F. Garin, "Graphon-based sensitivity analysis of SIS epidemics," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 542–547, 2020.

[116] S. Gao and P. E. Caines, "Spectral representations of graphons in very large network systems control," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 5068–5075.

[117] M. Avella-Medina, F. Parise, M. T. Schaub, and S. Segarra, "Centrality measures for graphons: Accounting for uncertainty in networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 520–537, 2020.

[118] E. Renshaw, *Modelling Biological Populations in Space and Time*. Cambridge University Press, 1991.

[119] B. Bonnet, N. P. Duteil, and M. Sigalotti, "Consensus formation in first-order graphon models with time-varying topologies," *Mathematical Models and Methods in Applied Sciences*, 2022.

[120] J. Bramburger and M. Holzer, "Pattern formation in random networks using graphons," *arXiv preprint arXiv:2110.14018*, 2021.

[121] A. R. Teel, A. Subbaraman, and A. Sferlazza, "Stability analysis for stochastic hybrid systems: A survey," *Automatica*, vol. 50, no. 10, pp. 2435–2456, 2014.

[122] A. R. Teel and J. P. Hespanha, "Stochastic hybrid systems: A modeling and stability theory tutorial," in *2015 54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 3116–3136.

[123] J. Abad Torres, P. J. Cruz, R. Vizuete, and R. Fierro, "On resilience and heterogeneity in robotic networks," in *Cooperative Localization and Navigation*. CRC Press, 2019, pp. 141–158.

[124] E. Panteley and A. Loría, "Synchronization and dynamic consensus of heterogeneous networked systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3758–3773, 2017.

[125] M. Maghenem, E. Panteley, and A. Loría, "Singular-perturbations-based analysis of dynamic consensus in directed networks of heterogeneous nonlinear systems," *arXiv preprint arXiv:2205.15646*, 2022.

[126] B. Adhikari, I.-C. Morărescu, and E. Panteley, "An emerging dynamics approach for synchronization of linear heterogeneous agents interconnected over switching topologies," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 43–48, 2021.

[127] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annual Reviews in Control*, vol. 45, pp. 166–190, 2018.

[128] C. Ravazzi, F. Dabbene, C. Lagoa, and A. V. Proskurnikov, "Learning Hidden Influences in Large-Scale Dynamical Social Networks: A Data-Driven Sparsity-Based Approach, in Memory of Roberto Tempo," *IEEE Control Systems Magazine*, vol. 41, no. 5, pp. 61–103, 2021.

[129] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis, and simulation," *Journal of artificial societies and social simulation*, vol. 5, no. 3, 2002.

[130] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "Continuous-time average-preserving opinion dynamics with opinion-dependent communications," *SIAM Journal on Control and Optimization*, vol. 48, no. 8, pp. 5214–5240, 2010.

[131] F. Ceragioli and P. Frasca, "Continuous and discontinuous opinion dynamics with bounded confidence," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 3, pp. 1239–1251, 2012.

[132] W. S. Rossi and P. Frasca, "Asynchronous opinion dynamics on the k-nearest-neighbors graph," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 3648–3653.

[133] P. Van Mieghem, *Graph Spectra for Complex Networks*. Cambridge University Press, 2010.

[134] E. M. Airoldi, T. B. Costa, and S. H. Chan, "Stochastic blockmodel approximation of a graphon: Theory and consistent estimation," in *Advances in Neural Information Processing Systems*, 2013, pp. 692–700.

[135] S. Janson, "Graphons, cut norm and distance, couplings and rearrangements," *New York J. Math. Monographs*, vol. 4, pp. 1–76, 2013.

[136] S. M. Ross, *Stochastic Processes*. Wiley New York, 1996.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Contributions aux systèmes multi-agents ouverts : consensus, optimisation et épidémies
Mots clés: Systèmes multi-agents ouverts, optimisation distribuée, consensus, agents et systèmes autonomes

Résumé: Dans cette thèse, nous abordons différents problèmes formulés dans un scénario de système multi-agents ouvert (OMAS) où l'ensemble des agents peut changer dans le temps, indépendamment de l'évolution de la dynamique associée au système. Nous considérons des OMAS formulés à l'aide d'un réseau fixe de taille finie, et utilisons deux approches distinctes pour les analyser. Dans la première approche, nous considérons des scénarios caractérisés par l'activation et la désactivation d'agents, de sorte qu'à chaque instant un sous-ensemble différent d'agents actifs peut interagir dans le système. Dans la deuxième approche, nous étudions des scénarios caractérisés par des remplacements d'agents où, à un instant donné, un agent peut être remplacé alors que les autres ne changent pas. Dans ce cas, tous les agents peuvent interagir à tout moment.

Trois problèmes différents sont considérés dans cette thèse : le consensus randomisé, le problème d'allocation des ressources et les épidémies. Premièrement, nous analysons le problème du consensus randomisé soumis à un bruit additif où différents sous-ensembles d'agents échangent des informations à chaque itération. Nous définissons un indice de bruit basé sur l'erreur quadratique moyenne attendue et nous dérivons des bornes supérieures. Ensuite, nous considérons le problème d'allocation de ressources où les agents peuvent être remplacés lors de l'implémentation d'un algorithme d'optimisation. Pour ce problème d'optimisation, nous analysons deux algorithmes différents : la descente de gradient pondérée (*weighted gradient descent*) et la descente de coordonnées aléatoires (*random coordinate descent*). Pour la descente de gradient pondérée, nous évaluons les performances de l'algorithme dans un OMAS soumis à des pertes de paquets en définissant des métriques de performances appropriées. Pour l'algorithme de descente de coordonnées aléatoires, nous étudions la convergence vers le minimiseur dans un OMAS et nous proposons une analyse alternative à l'aide d'outils inspirés de l'optimisation en ligne. Enfin, nous étudions une épidémie SIS en temps continu sujette à des remplacements d'agents lors de sa propagation. Nous effectuons l'analyse en utilisant une fonction d'agrégation et en dérivant des bornes supérieures pour son comportement asymptotique.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title: Contributions to open multi-agent systems: consensus, optimization and epidemics
Keywords: Open multi-agent systems, distributed optimization, consensus, agents and autonomous systems

Abstract: In this thesis we address several problems formulated in an open multi-agent system (OMAS) scenario where the set of agents can change in time, independently of the evolution of the dynamics associated with the system. We analyze the case of OMAS formulated in a fixed finite size network and we use two approaches for the analysis of the systems. In the first approach we consider scenarios characterized by activation/deactivation of agents such that at each time instant a different subset of active agents can interact in the system. In the second approach we study scenarios characterized by replacements of agents where at a specific time instant, an agent can be replaced while the rest of the agents remain the same. In this case, all the agents are able to interact at all time.

Three different problems are considered in this thesis: randomized consensus, resource allocation problem and epidemics. First, we analyze the problem of randomized consensus subject to additive noise where different subset of agents exchange information at each iteration. We define a noise index based on the expected mean squared error and we derive upper bounds. Then, we consider the resource allocation problem where agents can be replaced during the implementation of an optimization algorithm. For this problem, we analyze two different algorithms: weighted gradient descent and random coordinate descent. For the weighted gradient descent, we evaluate the performance of the algorithm in an OMAS subject to packet losses by defining appropriate performance metrics. For the random coordinate descent algorithm, we study the convergence to the minimizer in an OMAS and we provide an alternative analysis using tools inspired from online optimization. Finally, we study a SIS epidemic in continuous time subject to replacements of agents during the evolution of the disease. We perform the analysis using an aggregate function and deriving upper bounds for its asymptotic behavior.