



HAL
open science

Gradient descent for resource allocation with packet loss

Renato Vizquete, Paolo Frasca, Elena Panteley

► **To cite this version:**

Renato Vizquete, Paolo Frasca, Elena Panteley. Gradient descent for resource allocation with packet loss. NecSys 2022 - 9th IFAC Conference on Networked Systems, Jul 2022, Zurich, Switzerland. pp.109-114, 10.1016/j.ifacol.2022.07.244 . hal-03705202

HAL Id: hal-03705202

<https://hal.science/hal-03705202>

Submitted on 27 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gradient Descent for Resource Allocation with Packet Loss [★]

Renato Vizuete ^{***} Paolo Frasca ^{**} Elena Panteley ^{*}

^{*} *Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France (e-mail: {renato.vizuete,elena.panteley}@l2s.centralesupelec.fr)*

^{**} *Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, GIPSA-lab, F-38000 Grenoble, France (e-mail: paolo.frasca@gipsa-lab.fr)*

Abstract: This paper studies the effect of packet loss during the application of the weighted gradient descent to solve a resource allocation problem with piecewise quadratic cost functions in a multi-agent system. We 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: both bounds 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 open multi-agent systems where agents are replaced: based on a preliminary result and simulations we show that the combination of replacements and losses makes the constraint violation error diverge with time.

Keywords: Distributed optimization, multi-agent systems, packet loss, gradient descent, resource allocation, open multi-agent systems.

1. INTRODUCTION

The optimal distribution of a budget among multiple entities or activities is one of the most important problems in optimization (Ibaraki and Katoh, 1988). 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 a local cost function and communications are restricted by the network topology (Nedić et al., 2018; Doan and Beck, 2021). Applications include energy resources (Dominguez-Garcia et al., 2012), games (Liang et al., 2017), and distributed computer systems (Kurose and Simha, 1989). In general, algorithms assume perfect communications and rely on a symmetric exchange of information through the network to maintain the budget fixed (keeping the budget fixed is indeed the constraint of the optimization problem). One of the most important algorithms used to solve the resource allocation problem in a multi-agent system is the weighted gradient descent proposed in (Xiao and Boyd, 2006), where the update rule makes use of a matrix associated to the network that preserves the constraint.

However, 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 (Patterson et al., 2007; Fag-

nani and Zampieri, 2009; Frasca and Hendrickx, 2013b). In the case of the average estimation under packet losses, (Frasca and Hendrickx, 2013a) derived upper bounds for the mean square 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 resource allocation 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 (Yi et al., 2021).

Moreover, in large systems, the set of agents can change due to arrivals, departures or replacements, giving rise to *open multi-agent systems*. Since the size of the system may change with time, it is important to define performance metrics independent of the dimension like the variance (Monnoyer de Galland et al., 2020) or the normalized mean square error (Vizuete et al., 2021). In the case of optimization in open systems, it is assumed that agents can be replaced during the process, implying a change of the local cost functions (Hendrickx and Rabbat, 2020; Monnoyer de Galland et al., 2021). Due to the possible replacements, the minimizer is time-varying and in many scenarios the original constraint may not be preserved even if communications do not suffer of potential losses.

In this paper, we analyze the effects of packet losses in the performance of the weighted gradient descent algorithm

[★] Research supported in part by the Agence Nationale de la Recherche (ANR) via grant “Hybrid And Networked Dynamical sYstems” (HANDY), number ANR-18-CE40-0010.

to solve the resource allocation problem. First, assuming possible packet losses, we find conditions to guarantee convergence of the algorithm to the minimizer at least in expectation. Then, we derive upper bounds on the deviation from the constraint and on the error of the expected value of the cost function. Finally, we present an analysis in the case of open multi-agent systems where replacements of the agents can take place and we show, by a preliminary result and simulations, that the constraint error can diverge for this particular scenario. This fact highlights the difference between closed and open multi-agent systems with packet loss.

2. PROBLEM FORMULATION

The set of real numbers is denoted by \mathbb{R} and the set of non negative integers by $\mathbb{Z}_{\geq 0}$. For two vectors $x, y \in \mathbb{R}^n$, we denote the Euclidean inner product by $\langle x, y \rangle = x^T y$ and the Euclidean norm by $\|x\| = (x^T x)^{1/2}$. The vector constituted of only ones is denoted by $\mathbf{1}$. The Hadamard or element-wise product between two matrices $A, B \in \mathbb{R}^{n \times n}$ is denoted by $A \odot B$.

2.1 Resource Allocation Problem

The resource allocation problem in a system of n agents is formulated as:

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

where $f_i : \mathbb{R} \rightarrow \mathbb{R}$ is the local cost function associated to agent i and b is the budget. Following standard assumptions, we consider that each local cost function is continuously differentiable, β -smooth (i.e., $|f'_i(x) - f'_i(y)| \leq \beta |x - y|, \forall x, y$) and α -strongly convex (i.e., $f_i(x) - \frac{\alpha}{2} |x|^2$ is convex) (see (Nesterov, 2018)). Hence, the global cost function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is also β -smooth and α -strongly convex and the optimal solution denoted by x^* is unique. Furthermore, the optimality condition implies that $\mathbf{1}^T x^* = b$ and $\nabla f^* = \zeta \mathbf{1}$ for some $\zeta \in \mathbb{R}$ where $f^* = f(x^*)$.

We consider that agents interact through a connected undirected graph $G = (V, E)$, where $V = \{1, \dots, n\}$ 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 $(i, j) \in E$ 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 $D = \text{diag}[d_1, \dots, d_n]$ and d_i is the degree of agent i (i.e., d_i is the i th row-sum of A). Notice that L is a symmetric matrix that satisfies $L = L^T$.

To solve problem (1) at each iteration we apply the weighted gradient descent algorithm with positive step-size h :

$$x_{t+1} = x_t - hL\nabla f(x_t), \quad (2)$$

where L determines the exchange of information between the agents according to the network topology. It is well-known that for any initial condition x_0 satisfying the constraint $\mathbf{1}^T x_0 = b$, algorithm (2) solves problem (1) with an appropriate choice of the step-size h and guarantees linear convergence of the cost function.

2.2 Packet Losses

In this paper, we consider that the communication network is unreliable and some messages can be lost. Following the approach in (Frasca and Hendrickx, 2013a) we assume that at each time instant $t \in \mathbb{Z}_{\geq 0}$, we have a matrix $\Gamma_t = [\gamma_t^{ij}] \in \{0, 1\}^{n \times n}$ composed of Bernoulli random variables γ_t^{ij} independent across t, i, j with $\mathbb{P}(\gamma_t^{ij} = 1) = p$ for some $p \in (0, 1)$. Under this assumption, a link $a_{ij} = 1$ is active if $\gamma_t^{ij} = 1$ and inactive if $\gamma_t^{ij} = 0$. If $a_{ij} = 0$, the value of γ_t^{ij} does not have any effect.

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

$$A_t = A \odot \Gamma_t, \quad (3)$$

$$L_t = D_t - A_t, \quad (4)$$

and the weighted gradient descent algorithm becomes

$$x_{t+1} = x_t - hL_t \nabla f(x_t). \quad (5)$$

Due to the asymmetric loss of information, the graph topology at each time instant is directed and hence the Laplacian matrix L_t is non-symmetric. The constraint can be violated during an update of the states of the network using (5) and hence, the minimizer of the problem (1) will not be reached.

At least, the algorithm (5) must solve (1) in expectation. Since Γ_t is independent in time and $\mathbb{E}[L_t] = pL$, the expected dynamics is given by:

$$\mathbb{E}[x_{t+1}] = \mathbb{E}[x_t] - hpL\mathbb{E}[\nabla f(x_t)]. \quad (6)$$

Notice that similar to (2), the dynamics (6) preserves the constraint in expectation since:

$$\mathbf{1}^T \mathbb{E}[x_{t+1}] = \mathbf{1}^T (\mathbb{E}[x_t] - hpL\mathbb{E}[\nabla f(x_t)]) = \mathbf{1}^T \mathbb{E}[x_t]. \quad (7)$$

However, the minimizer x^* is not necessarily a stationary point of (6). Indeed, assume that $\mathbb{E}[x_t] = x^*$, then we have

$$\mathbb{E}[x_{t+1}] = x^* - hL\mathbb{E}[\nabla f(x_t)].$$

Under the assumption that the local functions are β -smooth and α -strongly convex we cannot guarantee that $\mathbb{E}[\nabla f(x_t)] = \nabla(f(\mathbb{E}[x_t])) = \zeta \mathbf{1}$.

For this reason, we restrict the set of possible cost functions to piecewise quadratic functions, which are β -smooth and α -strongly convex.

Assumption 1. Each local cost function $f_i(x_i)$ is continuous differentiable piecewise quadratic (i.e., there exists finitely many quadratic functions $\{q_j\}_{j=1}^J$ such that $f_i(x_i) \in \{q_j(x_i)\}_{j=1}^J$ for all $x_i \in \mathbb{R}$).

Remark that under Assumption 1, the gradient is a vector of piecewise linear functions $[\nabla f(x)]_i = \frac{\partial f(x)}{\partial x_i} = m_{ij}x_i + c_{ij}$ and we have

$$\mathbb{E}[\nabla f(x_t)] = \nabla f(\mathbb{E}[x_t]). \quad (8)$$

Then, the expected dynamics (6) becomes:

$$\mathbb{E}[x_{t+1}] = \mathbb{E}[x_t] - hpL\nabla f(\mathbb{E}[x_t]), \quad (9)$$

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

$$\lim_{t \rightarrow \infty} \mathbb{E}[x_t] = x^* \quad \text{and} \quad \lim_{t \rightarrow \infty} f(\mathbb{E}[x_t]) = f^*.$$

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:

$$\mathbb{E}[f(x_t)] \geq f(\mathbb{E}[x_t]),$$

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

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(x_t)] \geq f^*, \quad (10)$$

where the inequality is in general strict.

Performance of the gradient descent algorithm (5) in the presence of packet losses can be characterized by the following two quantities: errors due to the violation of the constraint and errors due to the deviation of the expected cost from the ideal cost f^* . In the following sections, we analyze these two errors by defining appropriate performance metrics.

3. CONSTRAINT VIOLATION

Although the constraint is preserved in expectation according to (7), this fact does not imply that the constraint is also preserved for any particular realization of the process. In this section we estimate the deviation from the constraint by deriving an upper bound for the constraint violation metric:

$$J_{\text{constr}} := \limsup_{t \rightarrow \infty} \mathbb{E} \left[(\mathbf{1}^T x_t - b)^2 \right], \quad (11)$$

where the initial condition x_0 satisfies $\mathbf{1}^T x_0 = b$. Notice that (11) corresponds to the asymptotic mean square error.

Theorem 1. Consider the weighted gradient descent given by (3), (4) and (5). Under Assumption 1, for any positive scalar $h \leq \frac{2}{\beta(p\lambda_n + 2 - 2p)}$, the constraint violation metric satisfies:

$$J_{\text{constr}} \leq \frac{4h(1-p)}{2 - h\beta(p\lambda_n - 2p + 2)} (f(x_0) - f^*), \quad (12)$$

where λ_n is the largest eigenvalue of L .

Before presenting the proof of Theorem 1, we recall a lemma derived in (Frasca and Hendrickx, 2013a).

Lemma 1. For L_t given by (3) and (4):

$$\mathbb{E}[L_t^T L_t] = p^2 L^2 + 2p(1-p)L \quad (13)$$

$$\mathbb{E}[L_t^T \mathbf{1} \mathbf{1}^T L_t] = 2p(1-p)L. \quad (14)$$

Proof of Theorem 1. Let us denote $H_t = (\mathbf{1}^T x_t - b)$. For (5) we denote by $\{\mathcal{F}_t\}_{t \in \mathbb{Z}_{>0}}$ the filtration of σ -algebras generated by the process x_t . We compute the expectation conditioned upon the filtration generated by x_t :

$$\begin{aligned} \mathbb{E}[H_{t+1}^2 | \mathcal{F}_t] &= \mathbb{E} \left[(\mathbf{1}^T (x_t - hL_t \nabla f(x_t)) - b)^2 \right] \\ &= H_t^2 + h^2 \nabla f(x_t)^T \mathbb{E}[L_t^T \mathbf{1} \mathbf{1}^T L_t] \nabla f(x_t) \\ &\quad - hH_t \mathbf{1}^T \mathbb{E}[L_t] \nabla f(x_t) \\ &= H_t^2 + h^2 \nabla f(x_t)^T \mathbb{E}[L_t^T \mathbf{1} \mathbf{1}^T L_t] \nabla f(x_t), \end{aligned}$$

where we have used the fact that $\mathbf{1}^T L = 0$. Then, we use (14) to obtain:

$$\mathbb{E}[H_{t+1}^2 | \mathcal{F}_t] = H_t^2 + 2h^2 p(1-p) \left\| L^{1/2} \nabla f(x_t) \right\|^2.$$

If we take the total expectation we get:

$$\mathbb{E}[\mathbb{E}[H_{t+1}^2 | \mathcal{F}_t]] = \mathbb{E}[H_t^2] + 2h^2 p(1-p) \mathbb{E} \left[\left\| L^{1/2} \nabla f(x_t) \right\|^2 \right].$$

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

$$\begin{aligned} \mathbb{E}[H_t^2] &= H_0^2 + 2h^2 p(1-p) \sum_{s=0}^{t-1} \mathbb{E} \left[\left\| L^{1/2} \nabla f(x_s) \right\|^2 \right] \\ &= 2h^2 p(1-p) \sum_{s=0}^{t-1} \mathbb{E} \left[\left\| L^{1/2} \nabla f(x_s) \right\|^2 \right], \quad (15) \end{aligned}$$

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

$$\begin{aligned} f(x_{t+1}) &\leq f(x_t) + \frac{\beta}{2} \|x_t - x_{t+1}\|^2 + \langle \nabla f(x_t), x_{t+1} - x_t \rangle \\ &= f(x_t) + \frac{h^2 \beta}{2} \langle \nabla f(x_t), L_t^T L_t \nabla f(x_t) \rangle \\ &\quad - h \langle \nabla f(x_t), L_t \nabla f(x_t) \rangle. \end{aligned}$$

We compute the expectation given the filtration generated by x_t and we use (13) to get:

$$\begin{aligned} \mathbb{E}[f(x_{t+1}) | \mathcal{F}_t] &\leq f(x_t) - h \langle \nabla f(x_t), pL \nabla f(x_t) \rangle \\ &\quad + \frac{h^2 \beta}{2} \langle \nabla f(x_t), (p^2 L^2 + 2p(1-p)L) \nabla f(x_t) \rangle. \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[f(x_{t+1}) | \mathcal{F}_t] &\leq f(x_t) \\ &\quad - hp \left(1 - h\beta(1-p) - \frac{h\beta p \lambda_n}{2} \right) \left\| L^{1/2} \nabla f(x_t) \right\|^2. \end{aligned}$$

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

$$\mathbb{E}[\mathbb{E}[f(x_{t+1}) | \mathcal{F}_t]] \leq \mathbb{E}[f(x_t)] - \xi \mathbb{E} \left[\left\| L^{1/2} \nabla f(x_t) \right\|^2 \right]. \quad (16)$$

Since $h \leq \frac{2}{\beta(p\lambda_n + 2 - 2p)}$ we guarantee that $\mathbb{E}[f(x_{t+1})] \leq \mathbb{E}[f(x_t)]$ such that:

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

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

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

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

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_{s=0}^{t-1} \mathbb{E} \left[\left\| L^{1/2} \nabla f(x_s) \right\|^2 \right] &\leq \frac{1}{\xi} \left(f(x_0) - \lim_{t \rightarrow \infty} \mathbb{E}[f(x_t)] \right) \\ &\leq \frac{1}{\xi} (f(x_0) - f^*), \quad (17) \end{aligned}$$

where we used the fact that $\mathbb{E}[f(x_t)] \geq f(\mathbb{E}[x_t])$. Finally, by using (17) in (15), we complete the proof. \square

By inspecting the right-hand side of (12), we can make several observations. The upper bound decreases when p increases and

$$\lim_{p \rightarrow 1} J_{\text{constr}} = 0, \quad (18)$$

which is consistent with the fact that the constraint is always preserved with no communication losses. In the

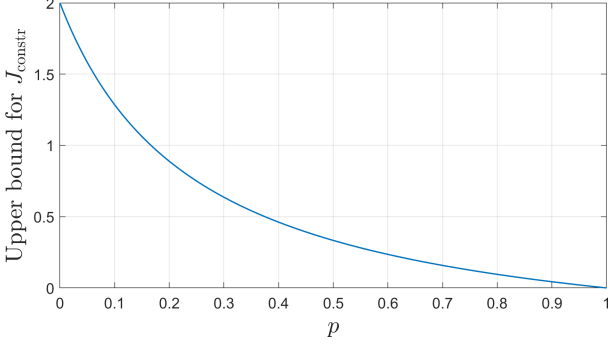


Fig. 1. Computation of the upper bound (12) for J_{constr} for a complete graph with $n = 10$, $\beta = 1$, $\epsilon = 0.5$ and $f(x_0) - f^* = 1$.

case of multiple failures, which correspond to p small, we parametrize the step size as $h = \frac{2\epsilon}{\beta(p\lambda_n + 2 - 2p)}$ with $\epsilon \in (0, 1)$ and using (12) we obtain:

$$\lim_{p \rightarrow 0} J_{\text{constr}} \leq \frac{2\epsilon}{\beta(1 - \epsilon)} (f(x_0) - f^*). \quad (19)$$

From (19), we see that for small values of p , large values of ϵ , corresponding to large values of the step-size, increase the potential violation of the constraint. Figure 1 presents the computation of the upper bound for the constraint violation metric in a complete 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 (18) and (19).

4. ERROR ON THE EXPECTED COST FUNCTION

Due to the packet losses, $f(x_t)$ is a random process whose evolution depends on the different realizations of L_t . When the *degraded* version of the weighted gradient descent algorithm (5) is executed, it is important to know the distribution of $f(x_t)$, to evaluate the impact of the perturbations on the performance of (5) and identify the additional cost corresponding to the deviation from the ideal cost f^* . From (10) we can observe that $\mathbb{E}[f(x_t)]$ is greater than f^* . Since the objective is to remain as close as possible to f^* , we define the cost function metric:

$$J_{\text{funct}} := \limsup_{t \rightarrow \infty} \mathbb{E}[f(x_t)] - f(x^*), \quad (20)$$

where the initial condition x_0 satisfies $\mathbf{1}^T x_0 = b$. The metric (20) estimates the gap in the cost function due to the packet losses during the application of the weighted gradient descent algorithm. It is important to notice that (20) corresponds to the limit of the so called *Jensen gap* (Abramovich and Persson, 2016).

Theorem 2. Consider the weighted gradient descent given by (3), (4) and (5). Under Assumption 1, for any positive scalar $h \leq \frac{2}{\beta(p\lambda_n + 2 - 2p)}$, the cost function metric satisfies:

$$J_{\text{funct}} \leq \left(1 - \frac{2 - h\beta(p\lambda_n - 2p + 2)}{2 - h\alpha p \lambda_2}\right) (f(x_0) - f^*), \quad (21)$$

where λ_2 is the second smallest eigenvalue of L .

Proof. From (16), we have that the total expectation of $f(x_{t+1})$ satisfies:

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

and by using Jensen's inequality and (8) we get:

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

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

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

and the limit satisfies:

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(x_t)] \leq f(x_0) - \xi \lim_{t \rightarrow \infty} \sum_{s=0}^{t-1} \left\| L^{1/2} \nabla f(\mathbb{E}[x_s]) \right\|^2. \quad (22)$$

Now, we consider the dynamics in expectation given by (6), which satisfies the inequality of α -strongly convex functions:

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

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

$$\|Lz\|^2 = \left\| L^{1/2} L^{1/2} z \right\|^2 \geq \lambda_2 \left\| L^{1/2} z \right\|^2, \quad (23)$$

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

$$f(\mathbb{E}[x_{t+1}]) \geq f(\mathbb{E}[x_t]) - hp \left(1 - \frac{h\alpha p \lambda_2}{2}\right) \left\| L^{1/2} \nabla f(\mathbb{E}[x_t]) \right\|^2.$$

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

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

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

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

and for $p > 0$ the limit satisfies:

$$\lim_{t \rightarrow \infty} \sum_{s=0}^{t-1} \left\| L^{1/2} \nabla f(\mathbb{E}[x_s]) \right\|^2 \geq \frac{1}{\delta} (f(x_0) - f^*). \quad (24)$$

We use (24) in (22) to get:

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(x_t)] \leq f(x_0) - \frac{\xi}{\delta} (f(x_0) - f^*),$$

which yields

$$\lim_{t \rightarrow \infty} \mathbb{E}[f(x_t)] - f(x^*) \leq \left(1 - \frac{\xi}{\delta}\right) (f(x_0) - f^*). \quad \square$$

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

$$\lim_{p \rightarrow 0} J_{\text{funct}} \leq \epsilon (f(x_0) - f^*), \quad (25)$$

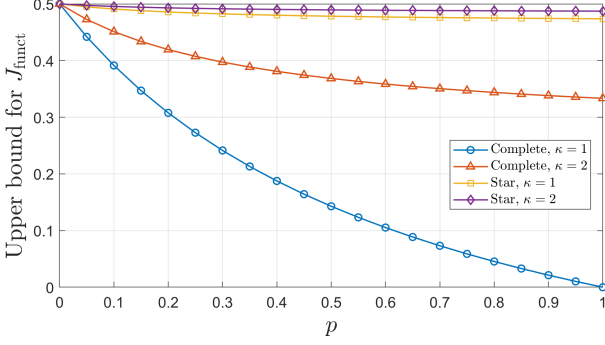


Fig. 2. Computations of the upper bound (21) for J_{funct} for a complete and star graph with $n = 10$, $\epsilon = 0.5$, $f(x_0) - f^* = 1$ and two values of κ .

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:

$$\lim_{p \rightarrow 1} J_{\text{funct}} \leq \frac{\epsilon(1 - \kappa^{-1}\kappa_\lambda^{-1})}{1 - \epsilon\kappa^{-1}\kappa_\lambda^{-1}} (f(x_0) - f^*), \quad (26)$$

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

$$\lim_{p \rightarrow 1} J_{\text{funct}} = 0, \quad (27)$$

which implies a certain level of conservatism in the upper bound since (27) should hold for any choice of κ and κ_λ in order to recover the behavior of a network without packet losses. Figure 2 shows the computation of the upper bound for J_{funct} for a complete graph and a star 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 (25) and (26).

5. OPEN MULTI-AGENT SYSTEMS

Let us consider the open scenario when agents can be replaced at any time step. In order to preserve the constraint of problem (1) at least in expectation, we shall consider only replacements such that the local cost function of an agent f_i can change but the state x_i is kept. More precisely, we assume that at each time instant t we can have an update with probability p_U or a replacement with probability $p_R = 1 - p_U$. To simplify notations, we will denote f_t , the global cost function at the time instant t . If an update occurs, the weighted gradient descent is applied and $f_{t+1} = f_t$. If a replacement occurs, a single agent i is uniformly randomly selected and receives a new (piecewise quadratic) cost function.

When the open system does not suffer from packet losses, the constraint is always preserved but the global minimizer x_t^* 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:

$$x_{t+1} = x_t - hL_t \nabla f_t(x_t). \quad (28)$$

Similarly to the closed system we analyze the constraint violation metric J_{constr} defined in (11).

Proposition 1. Consider the weighted gradient descent given by (3), (4) and (28). Under Assumption 1, for any positive scalar $h \leq \frac{2}{\beta(p\lambda_n + 2 - 2p)}$, the constraint violation metric satisfies:

$$J_{\text{constr}} \geq \frac{4p_U h(1-p)}{2 - \alpha h(p\lambda_2 + 2 - 2p)} \lim_{t \rightarrow \infty} \sum_{s=0}^{t-1} \mathbb{E} [f_s(x_s) - f_s(x_{s+1})]. \quad (29)$$

Proof. As in the proof of Theorem 1, we use the notation for the error $H_t = (\mathbf{1}^\top x_t - b)$ and the natural filtration $\{\mathcal{F}_t\}_{t \in \mathbb{Z}_{\geq 0}}$. If at the moment t , an update U happens, then from (15) we have at $t + 1$:

$$\mathbb{E} [H_{t+1}^2 | U] = \mathbb{E} [H_t^2] + 2h^2 p(1-p) \mathbb{E} \left[\left\| L^{1/2} \nabla f_t(x_t) \right\|^2 \right]. \quad (30)$$

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

$$\mathbb{E} [H_{t+1}^2 | R] = \mathbb{E} [H_t^2].$$

Then, the total expectation at the time instant $t + 1$ is defined as:

$$\begin{aligned} \mathbb{E} [H_{t+1}^2] &= p_U \mathbb{E} [H_{t+1}^2 | U] + p_R \mathbb{E} [H_{t+1}^2 | R] \\ &= \mathbb{E} [H_t^2] + 2p_U h^2 p(1-p) \mathbb{E} \left[\left\| L^{1/2} \nabla f_t(x_t) \right\|^2 \right]. \end{aligned}$$

Using a recursive argument we obtain:

$$\mathbb{E} [H_t^2] = 2p_U h^2 p(1-p) \sum_{s=0}^{t-1} \mathbb{E} \left[\left\| L^{1/2} \nabla f_s(x_s) \right\|^2 \right]. \quad (31)$$

Since the cost functions f_t are α -strongly convex we have:

$$\begin{aligned} f_t(x_{t+1}) &\geq f_t(x_t) + \frac{\alpha}{2} \|x_t - x_{t+1}\|^2 + \langle \nabla f_t(x_t), x_{t+1} - x_t \rangle \\ &= f_t(x_t) + \frac{h^2 \alpha}{2} \langle \nabla f_t(x_t), L_t^T L_t \nabla f_t(x_t) \rangle \\ &\quad - h \langle \nabla f_t(x_t), L_t \nabla f_t(x_t) \rangle. \end{aligned}$$

We compute the expectation given the filtration generated by x_t and we use (13) to get:

$$\begin{aligned} \mathbb{E} [f_t(x_{t+1}) | \mathcal{F}_t] &\geq f_t(x_t) - h \langle \nabla f_t(x_t), pL \nabla f_t(x_t) \rangle \\ &\quad + \frac{h^2 \alpha}{2} \langle \nabla f_t(x_t), (p^2 L^2 + 2p(1-p)L) \nabla f_t(x_t) \rangle. \end{aligned}$$

By using (23), we have the lower bound $\|L \nabla f_t(x_t)\|^2 \geq \lambda_2 \left\| L^{1/2} \nabla f_t(x_t) \right\|^2$ and we obtain:

$$\begin{aligned} \mathbb{E} [f_t(x_{t+1}) | \mathcal{F}_t] &\geq f_t(x_t) \\ &\quad - hp \left(1 - h\alpha(1-p) - \frac{h\alpha p \lambda_2}{2} \right) \left\| L^{1/2} \nabla f_t(x_t) \right\|^2. \end{aligned}$$

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

$$\mathbb{E} [\mathbb{E} [f_t(x_{t+1}) | \mathcal{F}_t]] \geq \mathbb{E} [f_t(x_t)] - \theta \mathbb{E} \left[\left\| L^{1/2} \nabla f_t(x_t) \right\|^2 \right].$$

Since $h \leq \frac{2}{\beta(p\lambda_n + 2 - 2p)} \leq \frac{2}{\alpha(p\lambda_2 + 2 - 2p)}$ we guarantee that $\mathbb{E} [f_t(x_{t+1})] \leq \mathbb{E} [f_t(x_t)]$ and therefore

$$\mathbb{E} \left[\left\| L^{1/2} \nabla f_t(x_t) \right\|^2 \right] \geq \frac{1}{\theta} (\mathbb{E} [f_t(x_t)] - \mathbb{E} [f_t(x_{t+1})]). \quad (32)$$

By using (32) in (31) we complete the proof. \square

Unlike a closed system, we observe that due to the possible replacements that take place in the system, in general

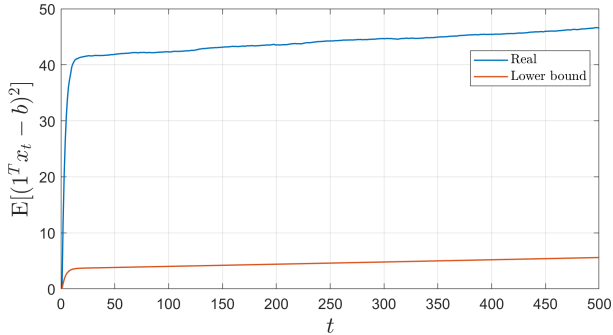


Fig. 3. Simulation of $\mathbb{E}[(\mathbf{1}^T x_t - b)^2]$ and of the lower bound (29) for a complete graph under replacements with $n = 7$, $\alpha = 1$, $\beta = 10$, $p = 0.8$, $p_U = 0.5$ by considering 10000 realizations of the process.

$\mathbb{E}[f_s(x_{s+1})] \neq \mathbb{E}[f_{s+1}(x_{s+1})]$. 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 $\mathbb{E}[(\mathbf{1}^T x_t - b)^2]$ for a complete graph with $n = 7$ agents that are shown in Figure 3. We can observe in the plot that both the constraint violation metric and the lower bound diverge. When $p = 1$, corresponding to an ideal system without packet losses, even if replacements happen in the system (i.e., $p_U > 0$), J_{constr} is zero since $\mathbb{E}[H_{t+1}^2|U] = \mathbb{E}[H_t^2]$ as we can see in (30). This can also be derived from (7), since the algorithm preserves the constraint independently of the cost functions.

6. CONCLUSION

In this paper, 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 function and we derived an upper bound. We extended the analysis of packet losses to open multi-agent systems and we showed that the constraint violation metric may diverge in such a case.

Several questions remain open as future work. First, it would be useful to improve the upper bound for the cost function metric since the current result suffers from a certain level of conservatism. Second, the analysis of open case is very preliminary, even though it has already shown that its properties are very different from those of the closed case. Relevant questions in this direction include the improvement of the lower bound for the constraint violation metric and the analysis of the cost function metric. Finally, it would be interesting to extend the analysis of this work to algorithms based on the computation of the gradient of a subset of agents like the random coordinate descent algorithm (Necoara, 2013; Monnoyer de Galland et al., 2021).

REFERENCES

Abramovich, S. and Persson, L.E. (2016). Some new estimates of the ‘Jensen gap’. *Journal of Inequalities and Applications*, 2016(1), 1–9.

Doan, T.T. and Beck, C.L. (2021). Distributed resource allocation over dynamic networks with uncertainty. *IEEE Transactions on Automatic Control*, 66(9), 4378–4384.

Dominguez-Garcia, A.D., Cady, S.T., and Hadjicostis, C.N. (2012). Decentralized optimal dispatch of distributed energy resources. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 3688–3693. IEEE.

Fagnani, F. and Zampieri, S. (2009). Average consensus with packet drop communication. *SIAM Journal on Control and Optimization*, 48(1), 102–133.

Frasca, P. and Hendrickx, J.M. (2013a). Large network consensus is robust to packet losses and interferences. In *2013 European Control Conference (ECC)*, 1782–1787. IEEE.

Frasca, P. and Hendrickx, J.M. (2013b). On the mean square error of randomized averaging algorithms. *Automatica*, 49(8), 2496–2501.

Hendrickx, J.M. and Rabbat, M.G. (2020). Stability of decentralized gradient descent in open multi-agent systems. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 4885–4890. IEEE.

Ibaraki, T. and Katoh, N. (1988). *Resource Allocation Problems: Algorithmic Approaches*. MIT press.

Kurose, J.F. and Simha, R. (1989). A microeconomic approach to optimal resource allocation in distributed computer systems. *IEEE Transactions on Computers*, 38(5), 705–717.

Liang, S., Yi, P., and Hong, Y. (2017). Distributed Nash equilibrium seeking for aggregative games with coupled constraints. *Automatica*, 85, 179–185.

Monnoyer de Galland, C., Martin, S., and Hendrickx, J.M. (2020). Modelling gossip interactions in open multi-agent systems. *arXiv preprint arXiv:2009.02970*.

Monnoyer de Galland, C., Vizuete, R., Hendrickx, J.M., Frasca, P., and Panteley, E. (2021). Random coordinate descent algorithm for open multi-agent systems with complete topology and homogeneous agents. In *2021 IEEE 60th Conference on Decision and Control (CDC)*, 1701–1708. IEEE.

Necoara, I. (2013). Random coordinate descent algorithms for multi-agent convex optimization over networks. *IEEE Transactions on Automatic Control*, 58(8), 2001–2012.

Nedić, A., Olshevsky, A., and Shi, W. (2018). Improved convergence rates for distributed resource allocation. In *2018 IEEE Conference on Decision and Control (CDC)*, 172–177.

Nesterov, Y. (2018). *Lectures on Convex Optimization*. Springer.

Patterson, S., Bamieh, B., and El Abbadi, A. (2007). Distributed average consensus with stochastic communication failures. In *2007 46th IEEE Conference on Decision and Control*, 4215–4220. IEEE.

Vizuete, R., Frasca, P., and Panteley, E. (2021). On the influence of noise in randomized consensus algorithms. *IEEE Control Systems Letters*, 5(3), 1025–1030.

Xiao, L. and Boyd, S. (2006). Optimal scaling of a gradient method for distributed resource allocation. *Journal of Optimization Theory and Applications*, 129(3), 469–488.

Yi, X., Li, X., Yang, T., Xie, L., Chai, T., and Johansson, K. (2021). Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. In *International Conference on Machine Learning*, 11998–12008. PMLR.