DISTRIBUTED NETWORK PROCESSING AND OPTIMIZATION UNDER COMMUNICATION CONSTRAINT

by

Chang-Shen Lee

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Electrical and Computer Engineering

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL STATEMENT OF COMMITTEE APPROVAL

Dr. Nicolò Michelusi, Co-chair

School of Electrical, Computer and Energy Engineering, Arizona State University

Dr. Gesualdo Scutari, Co-chair

School of Industrial Engineering, Purdue University

Dr. Shreyas Sundaram

School of Electrical and Computer Engineering, Purdue University

Dr. Xiaojun Lin

School of Electrical and Computer Engineering, Purdue University

Approved by:

Dr. Dimitrios Peroulis

To my parents, Chia-Hua & Chi-Wei,
and my better half, Wei-Yun.

ACKNOWLEDGMENTS

I am very grateful to my advisors, Professor Nicolò Michelusi and Professor Gesualdo Scutari, for their guidance and comments on my research work. In particular, they have provided me invaluable comments in the following aspects: 1) writing reader-friendly while rigorous mathematical proofs; 2) generalizing mathematical results; and 3) writing concise research articles.

I am also very grateful to the other two professors in my academic and doctoral advisory committee, Professor Shreyas Sundaram and Professor Xiaojun Lin, for their inspiring comments on my research and attending my preliminary and final examinations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In recent years, the amount of data in the information processing systems has significantly increased, which is also referred to as *big-data*. The design of systems handling big-data calls for a scalable approach, which brings distributed systems into the picture. In contrast to centralized systems, data are spread across the network of agents in the distributed system, and agents cooperatively complete tasks through local communications and local computations. However, the design and analysis of distributed systems, in which no central coordinators with complete information are present, are challenging tasks. In order to support communication among agents to enable multi-agent coordination among others, practical communication constraints should be taken into consideration in the design and analysis of such systems. The focus of this dissertation is to provide design and analysis of distributed network processing using *finite-rate communications* among agents. In particular, we address the following open questions: 1) can one design algorithms balancing a graph weight matrix using finite-rate and simplex communications among agents? 2) can one design algorithms computing the average of agents' states using finite-rate and simplex communications? and 3) going beyond of ad-hoc algorithmic designs, can one design a black-box mechanism transforming a general class of algorithms with unquantized communication to their finite-bit quantized counterparts?

This dissertation addresses the above questions. First, we propose novel distributed algorithms solving the weight-balancing and average consensus problems using only finite-rate simplex communications among agents, compliant to the directed nature of the network topology. A novel convergence analysis is put forth, based on a new metric inspired by the positional system representations. In the second half of this dissertation, distributed optimization subject to quantized communications is studied. Specifically, we consider a general class of linearly convergent distributed algorithms cast as fixed-point iterate, and propose a novel black-box quantization mechanism. In the proposed mechanism, a novel quantizer preserving linear convergence is proposed, which is proved to be more communication efficient than state-of-the-art quantization mechanisms. Extensive numerical results validate our theoretical findings.

1. INTRODUCTION

1.1 Motivation

In recent years, distributed processing for large-scale networks have gained increasing attentions. In particular, consensus-based distributed optimization problems, where agents in a network seek to cooperatively solve a partially shared problem, have found applications in a variety of areas including network information processing and decision making [1], resource allocation in communication networks [2], distributed spectrum sensing [3], distributed machine learning [4], and so on. For the above and other applications, centralized approaches are less preferable or even infeasible due to the following reasons.

First, centralized approaches are inefficient or unfeasible when it comes to deal with *big-data* applications: 1) the central coordinator may not have the capability to process the overwhelming data, or it may take too much time to perform such tasks; 2) in many applications, data are sensitive and should not be shared across agents, which makes collecting data at the central node less preferable and even prohibited; 3) collecting huge amount of data from the sources to the central coordinator may be time and energy consuming and impose huge burden on the communication networks. This problem is particularly important when some agents have limited energy (e.g., battery powered sensors in sensor networks) and/or wireless communication is adopted among agents (e.g., wireless sensor networks); 4) in some applications, where agents are homogeneous e.g., formation control for unmanned aerial vehicles, the selection of central coordinator will result in the imbalance workload among agents; and 5) centralized approaches are less robust to failure of central coordinator and external attacks. This problem is particularly important in some cyber-physical systems. Therefore, a promising approach is to distribute the data and workload across the network of agents, by cooperatively solving the optimization problem through local processing and local communications, which results in distributed in-network processing.

However, the design of distributed solution should take the following challenges into account:

1. **Information incompleteness:** each agent only has partial information on the problem for making decisions. To tackle this issue, agents need to exchange information

through local communications to get the information of the global problem. Hence, local communications among neighboring agents is an important part in the design of distributed algorithms.

2. **Communications constraints**: in the design of local communication protocols, communication constraints, such as the data rate constraint, need to be accounted in the design and analysis, as elaborated next.

Finite Rate Communications

From information theory, the maximum achievable communication rate of a link is the Shannon capacity of the channel, which depends on the signal/noise power and the channel bandwidth, and more importantly, is finite [5]. In other words, agents can only exchange encoded information using finite-bit representation instead of the exact information. To this end, *quantization*, which maps the input from an uncountable (continuous) set to the output in a countable (discrete) set, is a common approach for encoding the information. For instance, deterministic quantization [6] and probabilistic quantization [7] both map information from real numbers to countable sets. To represent the quantized information using finite bits, the cardinality (i.e., size) of its image set (i.e., the set of all possible outputs) must be finite. As a concrete example, the output of uniform quantizer [6], which maps the input to its nearest integers, cannot be represented by finite bits, since the set of all integers is of infinite size. Therefore, among all quantizations, *finite-bit quantizations*, where the output can be represented by finite bits, are required in the design of local communication in distributed algorithms.

To date, the majority of distributed algorithms do not address the finite rate constraint, i.e, they require exchange of *exact* (i.e., real-valued) information. Among those adopt some form of quantization in the design of local communication, some of them adopt aforementioned quantizations requiring infinite bits to implement. Hence, the implementation and analysis of the above algorithms becomes questionable when taking the finite-rate constraint into account. To the best of our knowledge, the design and analysis for distributed approaches addressing the finite-rate communication constraint has not been explored for problems in many areas, which motivates this work.

1.2 Research contributions

The contribution of this dissertation is to provide *the first* design/analysis of distributed algorithms using finite-rate communication for the following problems:

1) Weight balancing problem for directed graphs (digraphs);

2) Average consensus problem over digraphs;

3) The strongly convex optimization problems.

In the weight balancing problem, the goal is to balance the graph weight matrix, in the sense that the sum of incoming weights and that of outgoing weights are equal, for every agent in the network modelled as a digraph. This problem finds applications in many consensus-based algorithms over digraphs. In the average consensus problem, the goal is to compute the average of agents' initial state, which is important per se and is an essential building block for many consensus-based algorithms. In the strongly convex optimization problem, the goal is to find the solution of a general class of (possibly nonsmooth and/or constrained) strongly convex problems using first-order information, and such an algorithm is expect to converge linearly.

In the analyses of these proposed algorithms, this dissertation provides convergence guarantee as well as convergence rate for each algorithm. In addition, the communication costs are also analyzed. In a nutshell, this dissertation opens up the research of distributed approaches for some important problems while addressing the finite-rate communication constraint.

1.3 Outline of the dissertation

In Chapter 2, we propose the first distributed graph weight balancing algorithm using only finite rate and simplex communications among agents, and provide a novel convergence analysis. Building on this result, we propose the first distributed average consensus algorithm over digraphs, using finite rate communications. In Chapter 3, we study a general class of linearly convergent algorithms which can be cast as fixed point iterates, and propose a black-box finite-bit quantization scheme, which preserves linear convergence of the unquantized modelled algorithm. Our analysis shows that the proposed scheme is more communication efficient than the state-of-the-art quantization schemes, and achieves the lower complexity bounds for the convex quadratic optimization problem over a two-agent network. To the best of our knowledge, this is the first time that such limit is achieved. Finally, Chapter 4 summarize this dissertation.

Notation: Throughout the dissertation, we will use the following notation and conventions. The sets of real, integer, nonnegative integer, and positive integer numbers are denoted by $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{Z}_+$, and $\mathbb{Z}_{++}$, respectively. For any positive integer a, we define $[a] \triangleq \{1, \cdots, a\}$. Vectors are denoted as $\mathbf{x}$ (lowercase, boldface), matrices as $\mathbf{X}$ (uppercase, boldface). We denote by $\mathbf{0}, \mathbf{1}, \mathbf{O}$, and $\mathbf{I}$ denote the vector of all zeros, the vector of all ones, the matrix of all zeros, and the identity matrix, respectively. For vectors $\mathbf{c}_1, \cdots, \mathbf{c}_m$ and a set $\mathcal{S} \subseteq [m]$, define $\mathbf{c}_{\mathcal{S}} \triangleq \{\mathbf{c}_i : i \in \mathcal{S}\}$. We define the floor and ceiling functions $\lfloor x \rfloor \triangleq \max\{y \in \mathbb{Z} : y \leq x\}$, $\lceil x \rceil \triangleq \min\{y \in \mathbb{Z} : y \geq x\}$. We denote by $\mathbb{1}\{\bullet\}$ the indicator function, returning 1 if the input argument is true and 0 otherwise; and define the sign of x by $\operatorname{sgn}(x) = x/|x|, \forall x \neq 0$, $\operatorname{sgn}(0) = 0$. We use $\|\cdot\|$ to denote a norm in the Euclidean space (whose dimension will be clear from the context); when a specific norm is used, such as ℓ_2-norm or ℓ_∞, we will append the associate subscript to $\|\cdot\|$. The ith eigenvalue of matrix $\mathbf{G}$ is denoted by $\rho_i(\mathbf{G})$; and we order the eigenvalues of any real, symmetric matrix in nonincreasing order such that $\rho_1(\mathbf{G}) \geq \ldots \rho_i(\mathbf{G}) \geq \rho_{i+1}(\mathbf{G})$. Finally, asymptotic behaviors of functions is captured by the standard big-$\mathcal{O}, \Theta$, and Ω notations, namely: 1) $g(x) = \mathcal{O}(h(x))$ as $x \to x_0$ iff. $\limsup_{x \to x_0} |g(x)/h(x)| \in [0, \infty)$; 2) $g(x) = \Omega(h(x))$ iff. $h(x) = \mathcal{O}(g(x))$; and 3) $g(x) = \Theta(h(x))$ iff. $g(x) = \mathcal{O}(h(x))$ and $g(x) = \Omega(h(x))$. We will use superscript to denote iteration counters of sequences generated algorithms, for instance, x^k will denote the value of the x-sequence at iteration k. We will instead use $(x)^k$ for the k-power. We use L.x, C.x, D.x, T.x, P.x, A.x and App.x for Lemma x, Corollary x, Definition x, Theorem x, Proposition x, Assumption x and Appendix x, respectively.

2. FINITE RATE DISTRIBUTED WEIGHT-BALANCING AND AVERAGE CONSENSUS OVER DIGRAPHS

In this chapter, we first study the distributed weight balancing problem, and proposes the first distributed algorithm using only finite rate and simplex communications among agents, compliant with the directed nature of the graph edges. It is proved that the algorithm converges to a weight-balanced solution at sublinear rate. The analysis builds upon a new metric inspired by positional system representations, which characterizes the dynamics of information exchange over the network, and on a novel step-size rule. Building on this result, a novel distributed algorithm is proposed that solves the average consensus problem over digraphs, using, at each timeslot, finite rate simplex communications between adjacent agents – some bits for the weight-balancing problem and others for the average consensus. Convergence of the proposed quantized consensus algorithm to the average of the agent's unquantized initial values is established, both almost surely and in the moment generating function of the error; and a sublinear convergence rate is proved for sufficiently large step-sizes. Numerical results validate our theoretical findings.

The novel results of this chapter have been published in

- C.-S. Lee, N. Michelusi, and G. Scutari, "Topology-agnostic average consensus in sensor networks with limited data rate," in *Proc. 51st ACSSC*, Oct. 2017.
- C.-S. Lee, N. Michelusi, and G. Scutari, "Distributed quantized weight-balancing and average consensus over digraphs," in *Proc. 57th IEEE CDC*, Dec. 2018, pp. 5857–5862.
- C.-S. Lee, N. Michelusi and G. Scutari, "Finite Rate Distributed Weight-Balancing and Average Consensus Over Digraphs," *IEEE Trans. Autom. Control (Early Access)*, pp. 1-1, 2020.

2.1 Introduction

Digraphs play a key role in a number of network applications, such as distributed optimization [8], distributed flow-balancing [9], distributed averaging and cooperative control

[10], to name a few. In particular, distributed average consensus, whereby agents aim at agreeing on the sample average of their local values, has received considerable attention over the years; some applications include load-balancing [11], vehicle formation [12], and sensor networks [13]. Several of the these distributed algorithms, when run on digraphs, require some form of graph regularity, such as the *weight-balanced* property [14]: at each agent, the sum of the outgoing edge weights equals that of the incoming edge weights.

Several centralized algorithms have been proposed to balance a digraph; see, e.g., [15] and references therein. In this chapter, we are interested in the design of *distributed* algorithms that solve the weight-balancing and average consensus problems over digraphs, using only *quantized* information, *simplex communications*,[1] and without knowledge of the graph topology other than the direct neighbor. This problem is motivated by realistic scenarios, such as wireless sensor networks, where channels may be asymmetric due to different transmit powers of agents and interference, and where communications are subject to finite rate constraints. To date, the design of such algorithms in distributed settings remains a challenging and open problem, as documented next.

2.1.1 Related works

Distributed weight-balancing algorithms were proposed in [9], [14], [16]–[18] (see Table 2.1). With the exception of [14, Sec. IV], [17], [18], all these algorithms require *infinite bits* in each communication round, since agents need to exchange either real valued, or integer but unbounded quantities. Although [14, Sec. IV] and [18] use a finite number of bits at each iteration, this number cannot be arbitrarily chosen (e.g. to satisfy some transmission constraints), it is instead the result of the algorithmic trajectory and thus it is not known a-priori. In addition, these works adopt *unicast* communications, whereby agents transmit different signals to different out-neighbors. To reduce signaling overhead, *broadcast* communications are preferable in dense networks. Finally, while compliant with prescribed finite rate constraints, the distributed integer weight-balancing algorithm [17] requires *full-duplex* edge communications–each agent must exchange information with *both*

[1] ↑One way, as opposed to duplex, two ways communications.

Table 2.1. Related works on graph weight-balancing.

Reference	Broadcast	Digraph	# Bits/Timeslot
[9]	✓		Infinite
[16]		✓	Infinite
[14, Sec. III]	✓	✓	Infinite
[17]	✓		Any
[14, Sec. IV], [18]		✓	Problem-Dependent
Proposed	✓	✓	Any

Table 2.2. Related works on quantized consensus. problem-dependent: depends on the problem setting, e.g., initial values, weight matrix; trajectory-dependent: depends on trajectory of the algorithm during execution; informative: cf. A.4.

Ref.	Quantization	Digraph	Convergence	Limit Point	# Bits/Timeslot	Initial Value
[6]	Deterministic			Neighborhood	Problem-Dependent	Integer
[19]	Deterministic		✓	Neighborhood	Problem-Dependent	Integer
[20]	Deterministic			Neighborhood	Problem-Dependent	Box
[21]	Deterministic			Neighborhood	Infinite	Any
[22]	Deterministic			Neighborhood	Any	Any
[7]	Probabilistic		✓	Neighborhood	Any	Any
[23]	Probabilistic		✓	Neighborhood	Any	Box
[24], [25]	Deterministic		✓	Average	Any	Box
[26]	Probabilistic		✓	Average	Any	Box
[27]	Deterministic	✓(Balanced)	✓	Average	Any	Box
[28]	Probabilistic	✓(Balanced)	✓	Average	Any	Informative
[29], [30]	Deterministic	✓	✓	Weighted Average	Any	Box
[31], [32]	Deterministic	✓	✓	Average	Trajectory-Dependent	Integer
[33]	Deterministic	✓		Neighborhood	Infinite	Box
[34]	Probabilistic	✓	✓	Average	Infinite	Any
Proposed	Probabilistic	✓	✓	Average	Any	Informative

its out- and in-neighbors–which does not comply with simplex constraints. To the best of our knowledge, the design of distributed algorithms that solve the weight-balancing problem using a prescribed *finite rate and simplex* communications is an open problem.

Distributed average consensus algorithms have a long history, tracing back to the seminal works [1], [10], [35]. These early works assumed that agents can reliably exchange unquantized information over undirected networks. To cope with limited data rates, quantization was later introduced, and its effect analyzed for both undirected [6], [7], [19]–[26] or directed graphs [27]–[34], as documented in Table 2.2. The quantized average consensus problem based on deterministic uniform quantization and dithered (probabilistic) quantization has been considered in [6], [19]–[22] and [7], [23], respectively. However, these schemes do not achieve exact consensus but converge to a neighborhood of the average. *Exact* average consensus is proved in [24], [25] for deterministic quantization and in [26] for probabilistic quantization. However, all these algorithms consider *undirected* graphs, which can be easily weight-balanced (e.g. using the Metropolis weights [36]). While the extensions of the above deterministic and probabilistic schemes to digraphs were studied in [27], [29], [30] and [28], respectively, all these works only achieve exact average convergence over *balanced* digraph. However, the weight matrices of digraphs are inherently unbalanced, thus requiring specific weight-balancing algorithms, as documented earlier; they thus suffer from the same limitations of distributed weight balancing schemes.

To address unbalanced digraphs, the idea adopted in the seminal work [37] is to estimate and compensate the bias caused by the unbalanced weights, via the so-called *push-sum algorithm* [37]. This algorithm requires unquantized communication. Unfortunately, applying naively a finite-bit quantization to the push-sum scheme does not lead to convergence, as we will demonstrate numerically in Sec. 2.6 (cf. Q-Push-Sum). Extensions of push-sum employing quantization have been developed in [31]–[34]. However, [33], [34] consider unbounded quantization intervals, which necessitates *infinite* bits to encode the signal whereas [31], [32] impose integer constraints on the initial values of the consensus signals and necessitate a trajectory-dependent number of bits. Besides quantization, other instances of imperfect communications in average consensus problems over digraphs were investigated in [38], [39] (asynchrony) and [40], [41] (link failure).

2.1.2 Summary of the main contributions

The above literature review shows that there are no distributed algorithms solving the weight-balancing and the *exact* average consensus problems for real initial values over digraphs, using *finite-bit quantized information with a prescribed number of bits* and simplex communications. This chapter provides an answer to these open questions.

1) Distributed quantized weight-balancing: The first contribution is a novel distributed quantized weight-balancing algorithm whereby agents transfer part of their balance–the difference between the out-going and the incoming sum-weights, which should be zero for a weight-balanced graph–to their out-neighbors via quantized simplex communications; by doing so, the balance is transferred from high imbalance to low imbalance agents, provably converging to a weight-balanced solution at sublinear rate. Differently from existing quantized weight-balancing schemes [14], [16], [18], the proposed algorithm can use at each iteration a prescribed number of bits (possibly, time-varying). The convergence analysis is also a novel technical contribution of this chapter:

i) First, we identify necessary and sufficient conditions under which the total imbalance decreases, denoted by the *decreasing event* (see D.2.4.1). Roughly speaking, this event occurs when an agent transfers its balance to a neighbor with balance of opposite sign. Hence, agents closer to agents with balance of opposite sign more directly contribute to trigger the *decreasing event* and thus reduce the total imbalance, and are therefore more *important* than those farther away.

ii) The next step is to prove that the decreasing events occur often enough that the total imbalance asymptotically vanishes at sublinear rate. To this end, we show that the time interval between two consecutive occurrences of a decreasing event is *uniformly bounded*. This is proved by introducing a sophisticated metric, a non-negative integer-valued function of the imbalances of agents and of their importance, which *strictly* increases every time there is a transfer of balance from less important agents to more important ones, *up until* the next decreasing event occurs. By proving that this function is uniformly bounded, we conclude that the decreasing events occur infinitely often.

   To build such a function, we use the idea of positional system representation: the value of the function at each timeslot is expressed by a number whose hth digit represents the sum-imbalance of the hth most important *agents*. By doing so, every transfer of balance from agents of lower importance towards those of higher importance causes this function to increase, as it induces a “carry” operation from a digit to the next more significant one in its positional representation.

iii) We introduce a novel diminishing step-size rule, which guarantees that the balance at each agent is expressed as an integer multiple of the current step-size. This choice greatly facilitates the convergence analysis, since it allows one to tightly control the amount of decrement of the total imbalance at each timeslot.

2) Distributed average quantized consensus: Building on the proposed weight-balancing scheme, we introduce a novel distributed algorithm that performs average consensus and weight-balancing *on the same time scale* with finite-bit simplex communications–some bits for consensus and some to balance the digraph. For instance, one may perform one-bit (simplex) communication per channel use, by exchanging weight-balancing and consensus information alternately. The key idea behind the algorithm is to preserve the average of the variables over time, while gradually weight-balancing the graph. We prove convergence of the agents' local variables to the *exact* average of the initial values, both almost surely and in the moment generating function of the error. A sublinear convergence rate is proved for sufficiently large step-sizes.

The rest of this chapter is organized as follows. In Sec. 2.2, we introduce some preliminary definitions. Sec. 2.3 introduces the ideas of the proposed distributed quantized weight-balancing and average consensus algorithms, whose details are discussed in Sec. 2.4 and Sec. 2.5, respectively. Some numerical results are discussed in Sec. 2.6, while Sec. 2.7 draws some conclusions. The proof of auxiliary lemmas is provided in the Sec. 2.8.

Notation: In addition to the notation defined in Chapter 1. In this chapter, we will use the following notations. We define the clip function by $\text{clip}_{[l,u]}(x) = \min\{\max\{x, l\}, u\}$. All equalities and inequalities involving random variables are tacitly assumed to hold almost surely (i.e., with probability 1), unless otherwise stated. The rest of the symbols used in this chapter are summarized in Table 2.3.

2.2 Background

2.2.1 Basic graph-related definitions

Consider a network with m agents, modeled as a static, directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = [m]$ is the set of vertices (the agents), and $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$ is the set of edges (the communication links). A directed edge from i to j is denoted by $(i, j) \in \mathcal{E}$, so that information flows from i to j. We assume $(i, i) \notin \mathcal{E}, \forall i \in \mathcal{V}$, and denote the set of *in-* and *out-neighbors* of

Table 2.3. Notation used in the chapter

Symbol	**Description**
$\mathcal{N}_{+,i}, \mathcal{N}_{-,i}$	Out- (+) & in-neighbors (−) of agent i
$D_{+,i}, D_{-,i}$	Out- (+) & in-degrees (−) of agent i
k	Timeslot Index
γ^k	Step-size (for weight-balancing)
α^k	Step-size (for consensus)
$\mathbf{W}^k = (w_{ij}^k)_{i,j=1}^m$	Weight matrix
$S_{+,i}^k, S_{-,i}^k$	Sum of outgoing (+) & incoming (−) weights at agent i
$\mathbf{b}^k = (b_i^k)_{i=1}^m$	(Weight) balance
$\mathbf{L}_+^k, \mathbf{L}_-^k$	Graph Laplacian matrices
$[q_{\min}, q_{\max}]$	Quantization range for consensus
$\mathbf{y}^k = (y_i^k)_{i=1}^m$	Local estimate (cf. (2.9))
$\tilde{\mathbf{y}}^k = (\tilde{y}_i^k)_{i=1}^m$	Clipped local estimate (cf. (2.8))
$\mathbf{y}^0 = (y_i^0)_{i=1}^m$	Initial measurements
$B_{\mathrm{w},i}^k$	Number of bits to quantize b_i^k
$B_{\mathrm{c},i}^k$	Number of bits to quantize y_i^k
$\Delta(B)$	Distance between consecutive quantization points
n_i^k	Weight-balancing signal sent by agent i
$\mathcal{D}^k$,$\mathcal{U}^k$	Decreasing & Update events

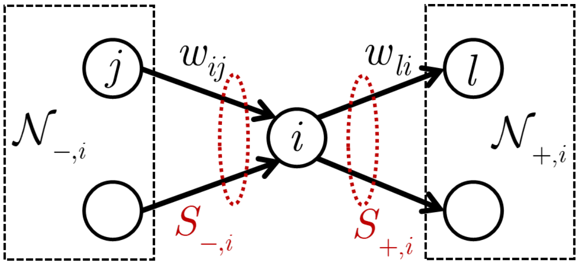


Figure 2.1. Some basic graph definitions.

agent i as $\mathcal{N}_{-,i} = \{j : (j,i) \in \mathcal{E}\}$ and $\mathcal{N}_{+,i} = \{j : (i,j) \in \mathcal{E}\}$, with cardinality $D_{-,i}$ (*in-degree*) and $D_{+,i}$ (*out-degree*), respectively. We will consider strongly connected digraphs.

Definition 2.2.1. *A digraph $\mathcal{G}$ is strongly connected if, $\forall i, j \in \mathcal{V}$ with $i \neq j$, there exists a directed path from i to j.*

Associated with the digraph $\mathcal{G}$, we define a weight matrix $\mathbf{W} \triangleq (w_{ij})_{i,j=1}^{m} \in \mathbb{R}^{m\times m}$ such that

$$\begin{cases} w_{ij} > 0, & \text{if } (j,i) \in \mathcal{E}; \\ w_{ij} = 0, & \text{otherwise}; \end{cases} \quad \forall i, j \in \mathcal{V}, \tag{2.1}$$

along with the following quantities (cf. Fig. 2.1) instrumental to formulate the weight-balancing problem.

Definition 2.2.2 (In-flow, out-flow and weight-balance)**.** *Given a digraph $\mathcal{G}$ with weight matrix $\mathbf{W}$, the in-flow of agent i is defined as $S_{-,i} \triangleq \sum_{j\in\mathcal{N}_{-,i}} w_{ij}$ while the out-flow is $S_{+,i} \triangleq \sum_{j\in\mathcal{N}_{+,i}} w_{ji}$. The weight-balance of agent i is defined as $b_i \triangleq S_{-,i} - S_{+,i}$; and the overall weight-balance vector is $\mathbf{b} \triangleq (b_i)_{i=1}^{m} = (\mathbf{W} - \mathbf{W}^\top)\mathbf{1}$.*

Definition 2.2.3 (Weight-balanced digraph)**.** *A weight matrix $\mathbf{W} \geq \mathbf{0}, \mathbf{W} \neq \mathbf{O}$, associated to the digraph $\mathcal{G}$, is said to be weight-balanced if it induces zero balance, i.e., $\mathbf{b} = (\mathbf{W} - \mathbf{W}^\top)\mathbf{1} = \mathbf{0}$.*

2.3 Summary of The Proposed Algorithms

In this section, we introduce the proposed distributed quantized weight-balancing and consensus algorithms; their detailed analysis will be carried out in Sec. 2.4 and Sec. 2.5.

2.3.1 System model and problem formulation

Average consensus problem

Each agent i controls and iteratively updates a local variable y_i, whose initial value is set to y_i^0. The *average consensus problem* consists in the following iterative algorithm (or variations of it): given $\mathbf{y}^k = (y_i^k)_{i\in\mathcal{V}}$ at time k, let

$$\mathbf{y}^{k+1} = \mathbf{W}\mathbf{y}^k, \tag{2.2}$$

where $\mathbf{W}$ is a suitably chosen weight matrix compliant with the graph [cf.(2.1)]. The goal is to locally estimate the average of the initial values

$$\bar{y}^0 \triangleq \frac{1}{m}\sum_{i=1}^{m} y_i^0, \tag{2.3}$$

i.e., $\|\mathbf{y}^k - \bar{y}^0\mathbf{1}\| \to 0$ as $k \to \infty$.

We consider a setting where: i) communications among the agents are quantized using a *finite* number of bits; and ii) information exchanges flows according to the edge directions of the graph $\mathcal{G}$ (simplex communications). This puts in jeopardy the convergence of the vanilla consensus algorithm (2.2), as communications therein subsume an infinite number of bits and $\mathbf{W}$ needs to be *balanced* [28], a condition that cannot be enforced a priori without using a centralized controller with knowledge of $\mathcal{G}$. To cope with these two issues, we first introduce a distributed quantized weight-balancing algorithm solving the weight-balancing problem (cf. Sec, 2.3.2); and then we integrate this algorithm with a distributed consensus algorithm using quantized simplex communications solving the average consensus problem (cf. Sec. 2.3.3).

2.3.2 Distributed quantized weight-balancing

We propose a distributed, iterative algorithm to solve the weight-balancing problem over a strongly connected digraph $\mathcal{G}$ using only quantized information and simplex communications. Note that strong connectivity guarantees the existence of a matrix, compliant to the digraph

$\mathcal{G}$ (cf. D.2.2.1) that is weight-balanced (cf. D.2.2.3) [15]. The proposed algorithm is formally stated in Algorithm 1 and discussed next.

Each agent i controls the in-neighbors weights $(w_{ij}^k)_{j \in \mathcal{N}_{-,i}}$. In `WB.1`, each agent i quantizes the local balance b_i^k via (2.4), using $B_{\mathrm{w},i}^k$ bits (a B-bit quantizer has $(2)^B + 1$ quantization levels), and broadcasts the quantized signal n_i^k to its out-neighbors. In `WB.2`, each agent i collects the signals from its in-neighbors, and updates the corresponding weights according to (2.5). The balance of each agent is then updated according to (2.6). Roughly speaking, by (2.5)-(2.6) there is a *transfer* of the balance among agents in the network: the quantity $\gamma^k \, D_{+,i} \, n_i^k$ (with γ^k denoting the step-size) is subtracted from the balance b_i^k of agent i [cf. (2.6)], and equally divided among its out-neighbors $j \in \mathcal{N}_{+,i}$, which will increase their incoming weight w_{ji}^k by $\gamma^k \, n_i^k$ [cf. (2.5)]. Note that 1) although n_i^k may be negative, $\mathbf{W}^k$ remains compliant to $\mathcal{G}$, which will be shown in T.2.4.1; 2) Algorithm 1 is fully distributed: each agent i only needs to know its in- and out-degrees $D_{+,i}$ and $D_{-,i}$, and to agree on a common step-size rule $\{\gamma^k\}_{k \in \mathbb{Z}_+}$. This assumption, along with knowledge of $D_{+,i}$ or its equivalent information, is commonly used in distributed algorithms over directed graphs; see, e.g., [14], [18], [33], [37], [42]. Convergence of Algorithm 1 is studied in Sec. 2.4.

2.3.3 Distributed quantized average consensus algorithm

We now introduce the proposed distributed quantized average consensus algorithm over non-balanced digraphs, as described in Algorithm 2. The algorithm combines Algorithm 1 with a variation of the quantized average consensus protocol based on probabilistic quantization, which we recently proposed in [28]. The algorithm is designed so that these two building blocks run *on the same time-scale.*

More specifically, each agent i controls two set of variables, namely: i) the in-neighbors weights $(w_{ij}^k)_{j \in \mathcal{N}_{-,i}}$; and ii) the local estimate y_i^k. The goal is to update these variables so that asymptotically the average consensus problem is solved while the weights converge to a balanced matrix. At each iteration k, agent i quantizes its local estimate y_i^k, by first clipping it within the quantization range $[q_{\min}, q_{\max}]$ [cf. (2.8)], followed by the probabilistic quantization (2.7) with $B_{\mathrm{c},i}^k$ bits; it then transmits the resulting quantized signal x_i^k (along

Algorithm 1 Distributed Quantized Weight-Balancing

```
Require: (WB.0) $\mathbf{W}^0$; $\{\gamma^k, (B^k_{\mathrm{w},i})_{i\in\mathcal{V}}\}_{k\in\mathbb{Z}_+}$.
  Set $k = 0$. Repeat (WB.1)-(WB.2) for $k = 1, 2, \dots$ until a termination criterion is satisfied;
  (WB.1) Each agent $i$ broadcasts the $n^k_i$ to $\mathcal{N}_{+,i}$, where
          $$n^k_i = \mathrm{sgn}(b^k_i)\min\left\{\frac{2}{(2)^{B^k_{\mathrm{w},i}}}\left[\left\lceil\frac{(2)^{B^k_{\mathrm{w},i}}|b^k_i|}{2D_{+,i}\gamma^k}\right\rceil - 1\right], 1\right\}. \quad (2.4)$$
  (WB.2) Each agent $i$ collects $n^k_j$ from its in-neighbors $j \in \mathcal{N}_{-,i}$, and updates
          $$w^{k+1}_{ij} = w^k_{ij} + \gamma^k\, n^k_j, \qquad \forall j \in \mathcal{N}_{-,i}, \quad (2.5)$$
          $$b^{k+1}_i = b^k_i - \gamma^k\, D_{+,i}\, n^k_i + \gamma^k \sum_{j\in\mathcal{N}_{-,i}} n^k_j. \quad (2.6)$$
```

with n^k_i for the weight-balancing) to its out-neighbors (AC.1). Upon receiving the signals $(n^k_j, x^k_j)_{j\in\mathcal{N}_{-,i}}$ from its in-neighbors, agent i updates its weights $(w^k_{ij})_{j\in\mathcal{N}_{-,i}}$ using (2.5), and the local variable y^k_i according to (2.9). The update in (2.9) aims at forcing a consensus on the average $\bar{y}^0$ among the local variables y^k_i. In fact, the third term in (2.9) is instrumental to align the local copies y^k_i, while the second term $+\alpha^k b^k_i x^k_i$ is a correction needed to preserve the average of the iterates, i.e., $(1/m)\sum_i y^{k+1}_i = (1/m)\sum_i y^k_i$, for all $k \in \mathbb{Z}_+$ [cf. (2.53)]. Hence, if all y^k_i are asymptotically consensual, it must be $\left|y^k_i - (1/m)\sum_i y^k_i\right| = \left|y^k_i - (1/m)\sum_i y^0_i\right| \xrightarrow[k\to\infty]{} 0$.

Convergence of Algorithm 2 is studied in Sec. 2.5.

2.4 Distributed Quantized Weight-Balancing

We study convergence of Algorithm 1 under the following mild assumptions.[2]

[2] ↑The analysis can be extended to the case in which each agent uses its own step-size $\{\gamma^k_i\}$, provided that: 1) every agent knows the step-size of its in-neighbors, and 2) every $\{\gamma^k_i\}$ satisfies A.2.

Algorithm 2 Distributed Quantized Average Consensus

Require: `(AC.0)` Init. Algorithm 1 as in `(WB.0)`; $q_{\min}, q_{\max}$; $\{\alpha^k, (B_{\mathrm{c},i}^k)_{i\in\mathcal{V}}\}_{k\in\mathbb{Z}_+}$; and $\mathbf{y}^0$.

Set $k = 0$. Repeat `(AC.1)-(AC.3)` for $k = 1, 2, \ldots$ until a termination criterion is satisfied;

`(AC.1)` Each agent i broadcasts n_i^k (cf. (2.4)) and x_i^k to $\mathcal{N}_{+,i}$, where $x_i^k = 0$ if $B_{\mathrm{c},i}^k = 0$ and, if $B_{\mathrm{c},i}^k > 0$,

$$x_i^k = \begin{cases} q_{\min} + \left\lceil \dfrac{\tilde{y}_i^k - q_{\min}}{\Delta(B_{\mathrm{c},i}^k)} \right\rceil \Delta(B_{\mathrm{c},i}^k), & \text{w.p. } p_i^k; \\[2ex] q_{\min} + \left\lfloor \dfrac{\tilde{y}_i^k - q_{\min}}{\Delta(B_{\mathrm{c},i}^k)} \right\rfloor \Delta(B_{\mathrm{c},i}^k), & \text{w.p. } 1 - p_i^k, \end{cases} \tag{2.7}$$

where $\tilde{y}_i^k = \mathrm{clip}(y_i^k; q_{\min}, q_{\max})$,

$$\begin{aligned} \Delta(B) &= \frac{q_{\max} - q_{\min}}{(2)^B - 1}, \forall B \in \mathbb{Z}_{++}, \\ p_i^k &= \frac{\tilde{y}_i^k - q_{\min}}{\Delta(B_{\mathrm{c},i}^k)} - \left\lfloor \frac{\tilde{y}_i^k - q_{\min}}{\Delta(B_{\mathrm{c},i}^k)} \right\rfloor. \end{aligned} \tag{2.8}$$

`(AC.2)` Each agent i collects (n_j^k, x_j^k) from its in-neighbors $j \in \mathcal{N}_{-,i}$, updates $(w_{ij}^{k+1})_{j\in\mathcal{N}_{-,i}}$ (cf. (2.5)) and

$$y_i^{k+1} = y_i^k + \alpha^k b_i^k x_i^k + \alpha^k \sum_{j\in\mathcal{N}_{-,i}} w_{ij}^k \, (x_j^k - x_i^k). \tag{2.9}$$

Assumption 1. *Let $\{B_{\mathrm{w}}^k\}_{k\in\mathbb{Z}_+}$ be a sequence satisfying $B_{\mathrm{w}}^k \in \{0, 1\}$ and $\sum_{t=nW}^{(n+1)W-1} B_{\mathrm{w}}^t \geq 1$, for all $k, n \in \mathbb{Z}_+$ and some $W \in \mathbb{Z}_{++}$. Then, there exists $B_{\max} \in \mathbb{Z}_+$ such that the number of bits $\{B_{\mathrm{w},i}^k\}_{k\in\mathbb{Z}_+}$ satisfies: for all $i \in \mathcal{V}$,*

$$\begin{cases} B_{\max} \geq B_{\mathrm{w},i}^k \geq B_{\mathrm{w}}^k, & \textit{if } B_{\mathrm{w}}^k = 1; \\ B_{\mathrm{w},i}^k = 0, & \textit{else.} \end{cases}$$

Assumption 2. *The step-size $\{\gamma^k\}_{k\in\mathbb{Z}_+}$ and initial weight matrix $\mathbf{W}^0 \triangleq (w_{ij}^0)_{i,j=1}^m$ satisfy:*

$$\begin{aligned} &\gamma^k = (c_1)^{-n}, \ \textit{with } n \in \mathbb{Z}_+ : ((c_1)^n - 1)c_2 \leq k \leq ((c_1)^{n+1} - 1)c_2 - 1; \\ &\textit{and} \quad w_{ij}^0 = \mathbb{1}\{(j, i) \in \mathcal{E}\}, \end{aligned} \tag{2.10}$$

respectively, where $c_1 \in \mathbb{Z}, c_1 \geq 2$, *and* $c_2 \in \mathbb{R}_{++}$.

Define $\bar{\gamma}^k \triangleq \gamma^k (2)^{1-B_{\max}}$, $k \in \mathbb{Z}_+$.

Note that the step-size satisfying A.2 is vanishing and non-summable, as shown below.

Lemma 1. *If* $\{\gamma^k\}_{k\in\mathbb{Z}_+}$ *satisfies A.2, then*

$$\frac{c_2}{k+c_2} \leq \gamma^k \leq \frac{c_1\, c_2}{k+c_2}, \quad \forall k \in \mathbb{Z}_+. \tag{2.11}$$

Proof. Let $n \in \mathbb{Z}_+$. Note that $\gamma^k = (c_1)^{-n}, \forall k : [(c_1)^n - 1]c_2 \leq k \leq [(c_1)^{n+1} - 1]c_2 - 1$. Then, the upper and lower bounds on γ^k are obtained by bounding $(c_1)^n$ with respect to this interval. □

A.2 is consistent with similar choices adopted in stochastic optimization [43], such as $\gamma^k = 1/(k+1)$. However the diminishing and non-summability properties alone are not sufficient to prove convergence of Algorithm 1; A.2 further guarantees that the balance at each agent is always an integer multiple of the current step-size (L.7, cf. App.2.8.1), which will be shown to be a key property to prove that $\|\mathbf{b}^k\|_1$ is asymptotically vanishing. An instance of $\{\gamma^k\}_{k\in\mathbb{Z}_+}$ satisfying A.2 is [44]

$$\{\gamma^k\}_{k\in\mathbb{Z}_+} = \left\{1, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \cdots\right\}. \tag{2.12}$$

Note that a fixed step-size $\gamma^k = \gamma$, for all k, may fail to achieve convergence. In fact, it is possible that each $0 < |b_i^k| \leq D_{+,i}\gamma$, so that $n_i^k = 0$ $\forall i$ and there is no further transfer of balance, resulting still in an unbalanced digraph.

A.1 states that at least once over a time window of duration W, *all* agents are simultaneously communicating at least one bit to their out-neighbors. This offers some flexibility in the design of the communication protocol. For instance, agents can transmit one bit at each time slot [44], yielding one bit per channel use, or transmit one bit every $W > 1$ time slots, resulting in a lower effective rate of $1/W$ bits per channel use.

We are now ready to state our main convergence result.

Theorem 2.4.1. *Let $\{\mathbf{W}^k\}_{k\in\mathbb{Z}_+}$ be the sequence generated by Algorithm 1 under A.1 and A.2. Then, there hold:*

$$\text{(a)}\|\mathbf{b}^k\|_1 = \mathcal{O}\Big(\frac{1}{k}\Big), \quad \text{(b)} \lim_{k\to\infty} \mathbf{W}^k = \mathbf{W}^*, \quad \text{(c)} 0 < w_{\min} \leq w_{ij}^k \leq w_{\max}, \forall (j,i) \in \mathcal{E}, \forall k \in \mathbb{Z}_+,$$

where $\mathbf{W}^*$ is weight-balanced and $w_{\min}$, $w_{\max}$ are defined in (2.28).

2.4.1 Proof of Theorem 2.4.1

Proof of statement (a)

We begin by highlighting the main steps of the proof, with the help of Fig. 2.2. **Step 1:** We show that $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$ is non-increasing. Furthermore, we identify two key events affecting the dynamics of $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$, namely: the so-called "decreasing event" $\mathcal{D}^k$ and "update event" $\mathcal{U}^k$. $\mathcal{D}^k$, formally defined in (2.13), occurs if at timeslot k either one of the following two facts happen: 1) an agent transfers its nonzero balance to an out-neighbor with balance of opposite sign; 2) two agents, with balance of opposite sign, transfer their balances to a common out-neighbor. On the other hand, $\mathcal{U}^k$, formally defined in D.2.4.2, occurs if at timeslot k an agent transfers its balance to its out-neighbors (i.e., $n_i^k \neq 0$, for some $i \in \mathcal{V}$) but $\mathcal{D}^k$ does *not* occur. Note that it can happen that neither $\mathcal{D}^k$ nor $\mathcal{U}^k$ occur at some k; this is the case when $|b_i^k|$ is "too small" or all agents are inactive ($B_{\mathrm{w}}^k = 0$). We show that $\|\mathbf{b}^k\|_1$ decreases by at least $2\bar{\gamma}^k$, with $\bar{\gamma}^k \triangleq \gamma^k(2)^{1-B_{\max}}$, iff. $\mathcal{D}^k$ occurs, and remains unchanged otherwise (cf. L.2). **Step 2:** To guarantee that $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$ vanishes, the decreasing event must occur *sufficiently often.* Towards this end, we prove two key properties of the decreasing and update events, namely:

P1) there are at most $\bar{U} \triangleq (m)^{2m-2}$ update events between two consecutive decreasing events;

P2) if $\|\mathbf{b}^k\|_1 \geq (m)^2\gamma^k$ (roughly speaking, if $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$ does not decrease sufficiently fast), there are at most $2W - 2$ timeslots between two consecutive update events.

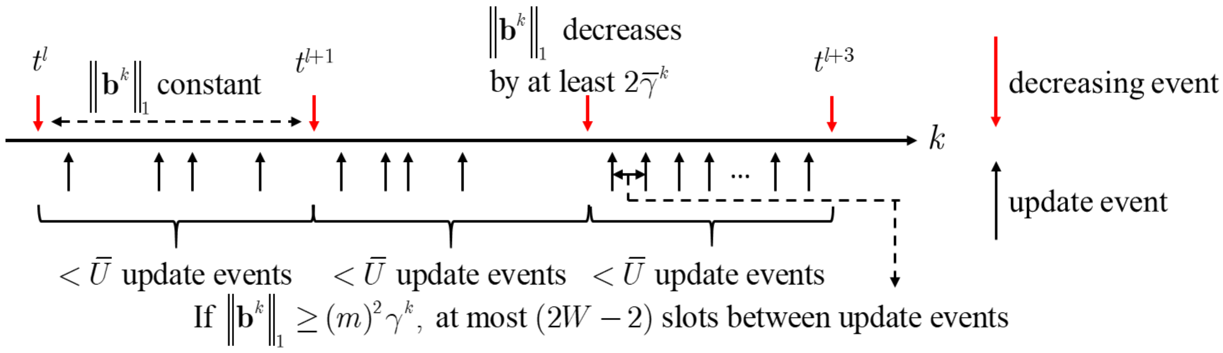

t^l
$\|\mathbf{b}^k\|_1$ constant
t^{l+1}
$\|\mathbf{b}^k\|_1$ decreases
by at least $2\bar{\gamma}^k$
t^{l+3}
k
decreasing event
update event
$< \bar{U}$ update events
$< \bar{U}$ update events
$< \bar{U}$ update events
If $\|\mathbf{b}^k\|_1 \geq (m)^2\gamma^k$, at most $(2W-2)$ slots between update events

Figure 2.2. Key properties of $\|\mathbf{b}^k\|_1$ in Algorithm 1.

The decreasing step-size together with P2 guarantee that update events occur within bounded time; this property combined with P1 guarantees that decreasing events occur at uniformly bounded time intervals. Finally, **Step 3** builds on the above results to prove statement (a) of the theorem. Roughly speaking, one can infer that: either 1) $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$ is below the diminishing threshold (Step 1), causing it to vanish (since the step-size vanishes); or 2) it exceeds the threshold at some timeslots, causing it to be suppressed by the decreasing event (Step 2) until it falls again below the vanishing threshold.

We proceed next with the formal proof.

Step 1: We begin by introducing the definition of $\mathcal{D}^k$ and $\mathcal{U}^k$.

Definition 2.4.1 (Decreasing event $\mathcal{D}^k$ and its occurrence time t^l)**.** *Let $\mathcal{D}^k$, $k \in \mathbb{Z}_+$, be defined as*

$$\exists i, (j, j') \in [\mathcal{N}_{-,i}]^2 : b_i^k n_j^k < 0 \vee n_{j'}^k n_j^k < 0. \tag{2.13}$$

Furthermore, let t^l, $l \in \mathbb{Z}_+$, be the timeslot of occurrence of the lth decreasing event; recursively, $t^0 \triangleq -1$ and $l \in \mathbb{Z}_{++}$,

$$t^l \triangleq \min\{k > t^{l-1} : \mathbb{1}\{\mathcal{D}^k\} = 1\}, \ (possibly,\ t^l = \infty). \tag{2.14}$$

Definition 2.4.2 (Update event $\mathcal{U}^k$)**.** *Let $\mathcal{U}^k$, $k \in \mathbb{Z}_+$, be defined as $\exists i \in \mathcal{V} : n_i^k \neq 0 \wedge \mathbb{1}(\mathcal{D}^k) = 0$.*

We show next that $\{\|\mathbf{b}^k\|_1\}_{k\in\mathbb{Z}_+}$ decreases by at least $2\bar{\gamma}^k$ iff. $\mathcal{D}^k$ occurs, and remains unchanged otherwise.

Lemma 2. *There holds*

$$\begin{cases} \|\mathbf{b}^{k+1}\|_1 \leq \|\mathbf{b}^k\|_1 - 2\,\bar{\gamma}^k, & \textit{if } \mathbb{1}(\mathcal{D}^k) = 1; \\ \|\mathbf{b}^{k+1}\|_1 = \|\mathbf{b}^k\|_1, & \textit{otherwise.} \end{cases} \tag{2.15}$$

Proof. See App.A-II. □

Step 2: This step characterizes "how often" $\mathcal{D}^k$ and $\mathcal{U}^k$ occur (properties P1 and P2), and the implication on $\|\mathbf{b}^k\|_1$. We first provide some intuition motivating our approach.

• **Intuition:** Let us look at the balance transfer within two consecutive decreasing events at (finite) times t^l and t^{l+1}. Let

$$\mathcal{V}_+^k \triangleq \{i : b_i^k > 0\},\ \mathcal{V}_-^k \triangleq \{i : b_i^k < 0\} \tag{2.16}$$

be the set of agents with positive and negative balance. Note that $\mathcal{V}_+^k, \mathcal{V}_-^k \neq \emptyset$ iff. $\|\mathbf{b}^k\|_1 > 0$, since $\sum_{i=1}^m b_i^k = 0$. For the next decreasing event $\mathcal{D}_{t^{l+1}}$ to occur at time t^{l+1}: either 1) an agent $i \in \mathcal{V}_+^{t^{l+1}}$ (resp. $i \in \mathcal{V}_-^{t^{l+1}}$) has enough balance (i.e., $n_i(t^{l+1}) \neq 0$) to trigger the update to an out-neighbor in $\mathcal{V}_-^{t^{l+1}}$ (resp. $\mathcal{V}_+^{t^{l+1}}$); or 2) two agents $j \in \mathcal{V}_+^{t^{l+1}}$ and $j' \in \mathcal{V}_-^{t^{l+1}}$ trigger an update to an out-neighbor in $\mathcal{N}_{+,j} \bigcap \mathcal{N}_{+,j'}$ ($\neq \emptyset$). This balance is built-up throughout the update events within (t^l, t^{l+1}), during which agents in $\mathcal{V}_+^k$ (resp. $\mathcal{V}_-^k$), $k \in (t^l, t^{l+1})$, keep transferring part of their balance towards the out-neighbors in $\mathcal{V}_+^k$ (resp. $\mathcal{V}_-^k$) that are *closer* to agents outside $\mathcal{V}_+^k$ (resp. $\mathcal{V}_-^k$). Hence, one can expect that the decreasing event $\mathcal{D}_{t^{l+1}}$ will occur after a certain number of update events, specifically when a sufficient amount of balance has been transferred to some agents having out-neighbors in $\mathcal{V} \setminus \mathcal{V}_+^k$ (resp. $\mathcal{V} \setminus \mathcal{V}_-^k$). To characterize this number and show that it is bounded over each interval (t^l, t^{l+1}), the proposed idea is to construct a *nonnegative, integer-valued* function of $\mathbf{b}^k, k \in (t^l, t^{l+1})$, denoted by U^k, which (a) *strictly increases* whenever $\mathcal{U}^k$ occurs; and (b) is uniformly upper *bounded* on (t^l, t^{l+1}). These properties guarantee that the number of update events within (t^l, t^{l+1}) is bounded, which proves P1. The same function U will be also used to prove P2 (cf. P.2.4.1 & C.1). Next, we build U^k and prove P1 and P2.

- **Building the function U^k:** Let $l \in \mathbb{Z}_+$, $t^l < \infty$,

$$U_h \triangleq (m)^{2(m-1-h)}, \quad h \in \{1, 2, \ldots, m-1\}. \tag{2.17}$$

We define the set (possibly empty) of agents that are h directed hops away from an agent with opposite sign of balance as

$$\mathcal{V}_h^k \triangleq \left\{ i \in \mathcal{V}_+^k \ : \ \min_{j \in \mathcal{V}_-^k} d(i,j) = h \right\} \cup \left\{ i \in \mathcal{V}_-^k \ : \ \min_{j \in \mathcal{V}_+^k} d(i,j) = h \right\}, \tag{2.18}$$

where $d(i,j)$ is the directed distance between i and $j \in \mathcal{V}$. Then, we define the function

$$U^k \triangleq \sum_{h=1}^{m-1} U_h \sum_{i \in \mathcal{V}_h^k} \min\left\{ \frac{|b_i^k|}{\bar{\gamma}^k}, m \right\}, \ \ k \in (t^l, t^{l+1}]. \tag{2.19}$$

Based on the above discussion, agents in $\mathcal{V}_1^k$ are "more important" than agents in $\mathcal{V}_2^k$, in the sense that they will more immediately trigger the next decreasing event; agents in $\mathcal{V}_2^k$ are more important than those in $\mathcal{V}_3^k$, and so on. The function U^k aims at capturing this hierarchical transfer of balance along the chain $\mathcal{V}_{m-1}^k \to \cdots \mathcal{V}_h^k \cdots \to \mathcal{V}_1^k$ during each update event, up until $\mathcal{D}_{t^{l+1}}$ occurs. In particular, we want U^k to increase its value by (at least) one (integer) unit every time one of such transfers happens (i.e., $\mathcal{U}^k$ occurs).

To motivate the choice of (2.19), let us look at the balance transfer during an update event $\mathcal{U}^k$ at time $k \in (t^l, t^{l+1})$; for the sake of simplicity, say $\mathcal{U}^k$ is triggered by agent $i \in \mathcal{V}_{h+1}^k$. As a result, agent i transfers part of its balance to its out-neighbors $\mathcal{N}_{+,i} \cap \mathcal{V}_h^k$,[3] according to (2.6). In (2.6), $\bar{\gamma}^k$ can be regarded as the unit of balance and $|b_i^k|/\bar{\gamma}^k$ is the normalized imbalance, an integer number (cf. L.7, App.2.8.1). Such an agent $i \in \mathcal{V}_{h+1}^k$ experiences a decrease of its normalized imbalance by $D_{+,i}$ units while the normalized imbalance of $j \in \mathcal{V}_h^k$ increases by at least one unit. To encode this balance transfer as an increase of U^k by at least one integer, we can associate it with the "carry on" operation from a digit to the next

[3] ↑Note that $\mathcal{N}_{+,i} \cap \mathcal{V}_h^k \neq \emptyset$ as, by definition of $\mathcal{V}_h^k$, at least one agent in $\mathcal{N}_{+,i}$ is one step closer to agents with balance of opposite sign.

more significant one in a positional notational representation of the U^k value. Specifically, U^k is expressed in radix-$(m)^2$ notation wherein the sum normalized imbalance of agents in $\mathcal{V}_1$ contributes to the most significant digit, the one of agents in $\mathcal{V}_2$ contributes to the second most significant digit, and so on. By doing so, when $\mathcal{U}^k$ occurs as above, the aforementioned exchange of balance $\mathcal{V}_{h+1}^k \to \mathcal{V}_h^k$ triggers the transfer of one unit from the $(h+1)th$ most representative digit to the hth one, so that U^k increases by at least one unit.

- **Proof of P1 and P2:** P.2.4.1 below states the desired properties of U^k and proves P2 as a by-product; P1 follows from C.1.

Proposition 2.4.1 (Properties of U^k). *Let $k \in (t^l, t^{l+1})$. Then:*

(i) $0 \leq U^k \leq (m)^{2m-1}$ *is nondecreasing;*

(ii) $U^{k+1} \geq U^k + m\mathbb{1}(\mathcal{U}^k)$;

(iii) If $\|\mathbf{b}^k\|_1 \geq (m)^2\gamma^k$, then an update or decreasing event occurs within the next $2W-1$ slots; hence,
$U^{k+2W-1} \geq U^k + m, \forall k \leq t^{l+1} - 2W + 1.$[4]

Proof. See Appendix A-III. □

Corollary 1. *The following hold.*

(i) There are at most $\bar{U} = (m)^{2m-2}$ update events between two consecutive decreasing events.

(ii) If $\|\mathbf{b}^k\|_1 \geq (m)^2\gamma^k, \forall k \in \mathbb{Z}_+$, then $\|\mathbf{b}^{k+(2W-1)\bar{U}}\|_1 \leq \|\mathbf{b}^k\|_1 - 2\bar{\gamma}^{k+(2W-1)\bar{U}}$.

Proof. (i) is a direct result of P.2.4.1(i)-(ii) and the fact that $U^k \geq 2$ iff. $\mathbf{b}^k \neq \mathbf{0}$. *(ii)*: let k such that $\|\mathbf{b}^k\|_1 \geq (m)^2\gamma^k$, and let $t^{l+1} = \{\tau \geq k : \mathcal{I}\{\mathcal{D}^\tau\} = 1\}$ be the next decreasing event at or after k (possibly, $t^l = -1$ and/or $t^{l+1} = \infty$). Invoking P.2.4.1, we infer that i) there are at most $\bar{U}$ update events in (t^l, t^{l+1}); ii) $\|\mathbf{b}^\tau\|_1 \geq (m)^2\gamma^\tau, \forall \tau \in (t^l, t^{l+1}]$ (γ^τ is non-increasing and $\|\mathbf{b}^\tau\|_1$ only decreases after decreasing events), so that the first update or decreasing event after t^l occurs within $2W-1$ timeslots and subsequent ones are separated

[4] ↑Since we are interested in the variations of U^k within $(t^l, t^{l+1}]$, the case $k > t^{l+1} - 2W + 1$ is irrelevant.

by at most $2W-2$ timeslots until the next decreasing event at t^{l+1}. These two facts together imply $t^{l+1} \leq k + (2W-1)\bar{U} - 1$; therefore,

$$\left\|\mathbf{b}^{k+(2W-1)\bar{U}}\right\|_1 \overset{(2.15)}{\leq} \left\|\mathbf{b}^{t^{l+1}+1}\right\|_1 \overset{(2.15)}{\leq} \left\|\mathbf{b}^{t^{l+1}}\right\|_1 - 2\bar{\gamma}^{t^{l+1}} \overset{(A.2)}{\leq} \left\|\mathbf{b}^{k}\right\|_1 - 2\bar{\gamma}^{k+(2W-1)\bar{U}}.$$

□

Step 3: We now prove that $\|\mathbf{b}^k\|_1 = \mathcal{O}(1/k)$. Equivalently,

$$\exists\ 0 < M < \infty \ \text{ and } \bar{k} \in \mathbb{Z}_{++} : k \cdot \|\mathbf{b}^k\|_1 \leq M, \forall k \geq \bar{k}. \tag{2.20}$$

To this end, note that it suffices to show that

$$\|\mathbf{b}^k\|_1 \leq (m)^2 c_1 \gamma^k, \quad \forall k \geq \bar{k}. \tag{2.21}$$

In fact, using L.1, (2.21) implies

$$k\|\mathbf{b}^k\|_1 \leq \frac{(m)^2 c_1^2 c_2 k}{k + c_2} = \mathcal{O}((m)^2 c_1^2 c_2),$$

so that (2.20) readily follows with $M \triangleq (m)^2 c_1^2 c_2$. To prove (2.21), let $\tilde{k}_n \triangleq ((c_1)^n - 1)c_2, n \in \mathbb{Z}_+$; we define $\bar{k}$ as

$$\bar{k} = \min\left\{\tau \geq \tilde{k}_n : \|\mathbf{b}^\tau\|_1 < (m)^2 \gamma^\tau\right\}, \tag{2.22}$$

for some sufficiently large $n \in \mathbb{Z}_+$ to be determined. The existence of such $\bar{k}$ is guaranteed by L.6 in App.2.8.1. Let

$$p = \min\{n' > n : \tilde{k}_{n'} > \bar{k}\}.$$

Then, it readily follows that, for all $k \in [\bar{k}, \tilde{k}_p - 1]$,

$$\|\mathbf{b}^k\|_1 \overset{(2.15)}{\leq} \|\mathbf{b}^{\bar{k}}\|_1 \overset{(2.22)}{<} (m)^2 \gamma^{\bar{k}} \overset{(A.2)}{=} (m)^2 \gamma^k,$$

so that (2.21) holds for $k \in [\bar{k}, \tilde{k}_p - 1]$. It remains to prove that it holds for $k \geq \tilde{k}_p$. We do so by induction. Assume that it holds at $k \in [\bar{k}, \tilde{k}_{n'} - 1]$, for some $n' \geq p$, and that

$$\|\mathbf{b}^{\tilde{k}_{n'}-1}\|_1 \leq (m)^2 \gamma^{\tilde{k}_{n'}-1}. \tag{2.23}$$

Clearly, this is true for $n' = p$. We show next that this condition implies that (2.21) holds for $k \in [\bar{k}, \tilde{k}_{n'+1} - 1]$, and

$$\|\mathbf{b}^{\tilde{k}_{n'+1}-1}\|_1 \leq (m)^2 \gamma(\tilde{k}_{n'+1} - 1). \tag{2.24}$$

Therefore, (2.21) holds $\forall k \geq \bar{k}$. To show the induction step, note that (2.23) implies (2.21), $\forall k \in [\tilde{k}_{n'}, \tilde{k}_{n'+1} - 1]$, since

$$\|\mathbf{b}^k\|_1 \overset{(L.2)}{\leq} \|\mathbf{b}^{\tilde{k}_{n'}-1}\|_1 \overset{(2.23)}{\leq} (m)^2 \gamma^{\tilde{k}_{n'}-1} \overset{(A.1)}{=} (m)^2 c_1 \gamma^k, \; \forall k \in [\tilde{k}_{n'}, \tilde{k}_{n'+1} - 1]. \tag{2.25}$$

It remains to prove (2.24); we do it by contradiction. Assume (2.23) and (2.25) hold but (2.24) does not. Then,

$$(m)^2 c_1 \gamma^k \overset{(2.25)}{\geq} \|\mathbf{b}^k\|_1 \overset{(L.2)}{\geq} \|\mathbf{b}^{\tilde{k}_{n'+1}-1}\|_1 > (m)^2 \gamma^{\tilde{k}_{n'+1}-1} \overset{(A.1)}{=} (m)^2 \gamma^k, \; \forall k \in [\tilde{k}_{n'}, \tilde{k}_{n'+1} - 1].$$

Choosing n large enough so that, for $T \triangleq (2)^{B_{\max}-2}(m)^2(c_1 - 1)$,

$$\tilde{k}_{n'} + T(2W-1)\bar{U} \leq \tilde{k}_{n'+1} - 1, \; \forall n' \geq p > n, \tag{2.26}$$

(this is possible since $\tilde{k}_{n'+1} - \tilde{k}_{n'} \geq \tilde{k}_{n+1} - \tilde{k}_n = c_2(c_1 - 1)(c_1)^n$) we can then apply C.1*(ii)* recursively T times, yielding

$$\|\mathbf{b}^{\tilde{k}_{n'+1}-1}\|_1 \overset{(2.26),(L.2)}{\leq} \|\mathbf{b}^{\tilde{k}_{n'}+T(2W-1)\bar{U}}\|_1 \leq \|\mathbf{b}^{\tilde{k}_{n'}}\|_1 - \sum_{j=0}^{T-1} 2\bar{\gamma}^{\tilde{k}_{n'}+j(2W-1)\bar{U}} \overset{(a)}{\leq} (m)^2 \gamma^{\tilde{k}_{n'+1}-1},$$

where (a) follows from A.2 and (2.25). This proves the contradiction, hence $\|\mathbf{b}^k\|_1 = \mathcal{O}(1/k)$.

Proof of statement (b)

Convergence of $\{\mathbf{W}^k\}_{k\in\mathbb{Z}_+}$ to $\mathbf{W}^*$ is a consequence of the following Lemma.

Lemma 3. *The sequence $\{\mathbf{W}^k\}_{k\in\mathbb{Z}_+}$ is a Cauchy sequence.*

Proof. See Appendix 2.8.1. □

Proof of statement (c)

First, using the fact that for any $x \in \mathbb{R}_+$, there exists $\beta \in [0, 1)$ such that $\lceil x \rceil - 1 = \beta x$, it follows that

$$n_i^k = \beta_i^k \frac{b_i^k}{D_{+,i}\gamma^k}, \text{ for some } \beta_i^k \in [0, 1). \tag{2.27}$$

Note that (2.10) and (2.5) imply $w_{ji}^k, \forall j \in \mathcal{N}_i^+$ have the same value. Let $\mathbf{S}_+^0 = \text{diag}\{S_+^k, i \in \mathcal{V}\}, \boldsymbol{\beta}^k = \text{diag}\{\beta_i^k, i \in \mathcal{V}\}, \mathbf{w}^k = (w_i^k)_{i=1}^m$, where $w_i^k = w_{ji}^k, j \in \mathcal{N}_i^+$. Applying (2.27) to the update of w_{ij}^k, it follows that

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \boldsymbol{\beta}^k(\mathbf{S}_+^0)^{-1}(\mathbf{W}^0 - \mathbf{S}_+^0)\mathbf{w}^k,$$

where we use the fact $\mathbf{b}^k = \mathbf{S}_-^k - \mathbf{S}_+^k = \mathbf{W}^0\mathbf{w}^k - \mathbf{S}_+^0\mathbf{w}^k$, which implies that

$$\mathbf{w}^{k+1} = \left[\mathbf{I} - \boldsymbol{\beta}^k + \boldsymbol{\beta}^k(\mathbf{S}_+^0)^{-1}\mathbf{W}^0\right]\mathbf{w}^k.$$

Let $\mathbf{P}^k \triangleq \mathbf{I} - \boldsymbol{\beta}^k + \boldsymbol{\beta}^k(\mathbf{S}_+^0)^{-1}\mathbf{W}^0, \mathbf{L}_+^0 \triangleq \mathbf{S}_+^0 - \mathbf{W}^0$, and $\mathbf{w} \triangleq (w_i)_{i=1}^m \neq \mathbf{0}$ be a non-trivial weight-balancing solution, i.e., $\mathbf{w} \in \{\boldsymbol{\omega} : \mathbf{L}_+^0\boldsymbol{\omega} = \mathbf{0}\}$. It follows that $\mathbf{P}^k$ and $\mathbf{a}$ satisfy the following properties

1. $\mathbf{P}^k \geq \mathbf{0}$ since $\beta_i^k \in [0, 1), \forall i \in \mathcal{V}, \forall k \in \mathbb{Z}_+$;

2. $\mathbf{w} > \mathbf{0}$. Since $\mathbf{L}_+^0$ is an irreducible singular M-matrix [45], it follows from [45, T.4.31] that 1) $\mathbf{L}_+^0$ has rank $m - 1$, and 2) $\exists \tilde{\boldsymbol{\omega}} > \mathbf{0} : \mathbf{L}_+^0\tilde{\boldsymbol{\omega}} = \mathbf{0}$, and thus $\mathbf{w} > \mathbf{0}$;

3. $\mathbf{P}^k\mathbf{w} = \mathbf{w}, \forall k \in \mathbb{Z}_+$ since $\mathbf{S}_+^0\mathbf{w} = \mathbf{W}^0\mathbf{w}$.

Let $\mathbf{w}_{\min} \triangleq (w_{\min,i})_{i=1}^m, \mathbf{w}_{\max} \triangleq (w_{\max,i})_{i=1}^m \in \{\boldsymbol{\omega} : \mathbf{L}_+^0 \boldsymbol{\omega} = \mathbf{0}\}$ with $\max_{i\in\mathcal{V}} w_{\min,i} = \min_{i\in\mathcal{V}} w_{\max,i} = 1$, $\mathbf{e}_{\min} \triangleq \mathbf{1} - \mathbf{w}_{\min} = \mathbf{w}^0 - \mathbf{w}_{\min} \geq \mathbf{0}$ and $\mathbf{e}_{\max} \triangleq \mathbf{1} - \mathbf{w}_{\max} \leq \mathbf{0}$. It follows that

$$\mathbf{w}^{k+1} = \mathbf{P}^k \mathbf{w}^k = \left(\prod_{t=0}^{k} \mathbf{P}^t\right)\mathbf{w}^0 = \left(\prod_{t=0}^{k} \mathbf{P}^t\right)(\mathbf{w}_{\min} + \mathbf{e}_{\min}) \overset{(a)}{=} \mathbf{w}_{\min} + \left(\prod_{t=0}^{k} \mathbf{P}^t\right)\mathbf{e}_{\min} \overset{(b)}{\geq} \mathbf{w}_{\min},$$

where (a) follows from $\mathbf{P}^k \mathbf{w}_{\min} = \mathbf{w}_{\min}, \forall k \in \mathbb{Z}_+$ and (b) follows from $\mathbf{P}^k \geq \mathbf{0}, \forall k \in \mathbb{Z}_+, \mathbf{e}_{\min} \geq \mathbf{0}$. Similarly,

$$\mathbf{w}^{k+1} = \mathbf{w}_{\max} + \left(\prod_{t=0}^{k} \mathbf{P}^t\right)\mathbf{e}_{\max} \leq \mathbf{w}_{\max},$$

which proves the desired results with

$$w_{\min} = \min_{i\in\mathcal{V}} w_{\min,i}, \quad \text{and} \quad w_{\max} = \max_{i\in\mathcal{V}} w_{\max,i}. \tag{2.28}$$

2.5 Distributed Quantized Average Consensus

In this section, we study convergence of Algorithm 2. We introduce the following mild assumptions.

The first condition is on the number of bits used to quantize the consensus variables at each iteration.

Assumption 3. *Let $\{B_{\mathrm{c}}^k\}_{k\in\mathbb{Z}_+}$ be an activation sequence satisfying $B_{\mathrm{c}}^k \in \{0,1\}$ and $\sum_{t=nW}^{(n+1)W-1} B_{\mathrm{c}}^t \geq 1$, for all $k, n \in \mathbb{Z}_+$ and some given $W \in \mathbb{Z}_{++}$. The number of bits $\{B_{\mathrm{c},i}^k\}_{k\in\mathbb{Z}_+}$ used by each agent i satisfies*

$$\begin{cases} B_{\mathrm{c},i}^k \geq B_{\mathrm{c}}^k, & \textit{if } B_{\mathrm{c}}^k = 1; \\ B_{\mathrm{c},i}^k = 0, & \textit{else.} \end{cases}$$

The above condition is almost the same as the one used in the weight-balancing algorithm (cf. A.1) except for the global upper bound, and can be coupled with it. For example, agents can communicate for weight-balancing using one bit at odd time slots, and for average consensus using one bit at even time slots, yielding one bit per channel use. Lower effective data rates can be achieved using intermittent communications.

We next introduce the assumption on $\bar{y}^0$ and the step-size used in the consensus updates.

Assumption 4 (Informative $\bar{y}^0$)**.** *The average $\bar{y}^0$ [cf. (2.3)] satisfies $\bar{y}^0 \in [q_{\min}, q_{\max}]$.*

Assumption 5. *The step-sizes $\{\alpha^k\}_{k\in\mathbb{Z}_+}$ satisfy $0 < \alpha^{k+1} \leq \alpha^k, \forall k \in \mathbb{Z}_+, \sum_{k=1}^{\infty} \alpha^k = \infty, \sum_{k=1}^{\infty} (\alpha^k)^2 < \infty$.*

It is important to remark that A.4 neither requires y_i^0 to be confined within the quantization range nor its to be known. This is a major departure from the literature, which calls for y_i^0 to be within the quantization range – see, e.g. [24]–[26], [29]. We require instead the *average* $\bar{y}^0$ to fall within the quantization interval $[q_{\min}, q_{\max}]$, which is a less restrictive condition. For example, if agents are estimating a common unknown parameter θ via noisy measurements $y_i^0 = \theta + \omega_i$ corrupted by zero mean Gaussian noise ω_i, i.i.d. across agents, then $\bar{y}^0$ is the sample mean estimate across the agents. In this case, a bound on y_i^0 is hard to obtain (theoretically it is unbounded), but the bound of the parameter, $\theta \in [\theta_{\mathrm{m}}, \theta_{\max}]$, is known in many cases. Even worse, $\max_{i\in\mathcal{V}} |y_i^0| \to \infty$ for $m \to \infty$, whereas the sample average $\bar{y}^0 \to \theta$, so that it becomes more and more informative for large m, whereas the initial local measurements become larger and larger. In this example, agents can simply set $(q_{\min}, q_{\max}) = (\theta_{\mathrm{m}}, \theta_{\max})$, so that $\bar{y}^0$ is informative with high probability. Herein, we are not interested in non-informative $\bar{y}^0$, which, as the name suggests, does not provide any information to estimate θ.

We are now ready to state the convergence of Algorithm 2.

Theorem 2.5.1. *Let $\left\{\mathbf{y}^k = (y_i^k)_{i=1}^m\right\}_{k\in\mathbb{Z}_+}$ be the sequence generated by Algorithm 2 under A.2-5. Then:*

(a) Almost sure convergence:

$$\mathbb{P}\left(\lim_{k\to\infty} \mathbf{y}^k = \bar{y}^0 \cdot \mathbf{1}\right) = 1. \tag{2.29}$$

(b) Convergence in the moment generating function:

$$\lim_{k\to\infty} \mathbb{E}\left[e^{r\|\mathbf{y}^k-\bar{y}^0\cdot\mathbf{1}\|}\right] = 1, \quad \forall r \in \mathbb{R}. \tag{2.30}$$

Furthermore, if $\alpha^k = \mathcal{O}(1/k)$ and $\exists m > 0 : \alpha^k \geq m/(k+1), \forall k$, then:

(c) Convergence rate:

$$V^k \triangleq \mathbb{E}[V(\mathbf{y}^k)] \leq \begin{cases} \mathcal{O}(1/k) & \rho > 1, \\ \mathcal{O}(\ln(k)/k) & \rho = 1, \\ \mathcal{O}(1/k^\rho) & \rho < 1, \end{cases} \tag{2.31}$$

where $\rho \triangleq 2\xi_1 m > 0$ with $\xi_1 > 0$ defined in L.10.

Proof. Let $(\mathbf{y}^k, \mathbf{x}^k, \tilde{\mathbf{y}}^k) \triangleq (y_i^k, x_i^k, \tilde{y}_i^k)_{i=1}^m$ with $\tilde{y}_i = \text{clip}(y_i; q_{\min}, q_{\max})$. Using (2.9), the y-updates become

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \alpha^k \mathbf{L}_+^k \mathbf{x}^k. \tag{2.32}$$

To study the dynamics of the consensus error, we define

$$V(\mathbf{y}) \triangleq \|\mathbf{y} - \bar{y}^0\mathbf{1}\|^2, \tag{2.33}$$

and prove that the sequence $\{V(\mathbf{y}^k)\}_{k\in\mathbb{Z}_+}$ satisfies the conditions of [46, T.1], sufficient to prove our theorem.

Intermediate results: We begin by introducing some properties of $V(\mathbf{y})$, instrumental for the sequel of the proof.

Lemma 4. *In the setting of T.2.5.1, $\forall \mathbf{y} \in \mathbb{R}^m$ there holds*

$$\mathbb{E}\left[V(\mathbf{y}^{k+1})|\mathbf{y}^k = \mathbf{y}\right] = \begin{cases} V(\mathbf{y}) - 2\alpha^k \mathbf{y}^\top \mathbf{L}_+^k \tilde{\mathbf{y}} + (\alpha^k)^2 \mathbb{E}\left[\|\mathbf{L}_+^k \mathbf{x}^k\|^2 | \mathbf{y}^k = \mathbf{y}\right], & \textit{if } B_\text{c}^k = 1, \\ V(\mathbf{y}), & \textit{otherwise.} \end{cases}$$

The case $B_{\rm c}^k = 0$ holds trivially since $\mathbf{y}^{k+1} = \mathbf{y}$. Otherwise ($B_{\rm c}^k = 1$) L.4 follows from the dynamics (2.32) and the fact that $\mathbb{E}[\mathbf{x}^k|\mathbf{y}^k] = \tilde{\mathbf{y}}^k$ with probabilistic quantization. To bound these dynamics when $B_{\rm c}^k > 0$, we use the fact that $\mathbf{y}^k$ is uniformly bounded within a bounded set $\mathcal{S}$ with probability 1 (cf. L.9) and L.10 to obtain

$$\begin{aligned}\mathbb{E}\left[V(\mathbf{y}^{k+1})|\mathbf{y}^k = \mathbf{y}\right] &\leq V(\mathbf{y}) - 2\alpha^k\xi_1 V(\mathbf{y}) + 2\alpha^k\xi_2\|\mathbf{b}^k\|_1 + (\alpha^k)^2\|\mathbf{L}_+^k\|_2^2\mathbb{E}\left[\|\mathbf{x}^k\|^2|\mathbf{y}^k = \mathbf{y}\right] \\ &\overset{(a)}{\leq} V(\mathbf{y}) - 2\xi_1\alpha^k\left[V(\mathbf{y}) - c^k\right], \forall \mathbf{y} \in \mathcal{S}, \end{aligned} \tag{2.34}$$

where ξ_1, ξ_2 are constants defined in L.10 and in (a) we defined

$$c^k \triangleq \frac{\xi_2}{\xi_1}\|\mathbf{b}^k\|_1 + \alpha^k\frac{\xi_3}{2\xi_1}, \tag{2.35}$$

for some constant $\xi_3 \geq \|\mathbf{L}_+^k\|_2^2\mathbb{E}\left[\|\mathbf{x}^k\|^2|\mathbf{y}^k = \mathbf{y}\right] > 0$. Note that the boundedness of $\mathbf{W}^k$ (cf. L.3), and thus of $\mathbf{L}_+^k$, and that of $\mathbf{x}^k$ (being the output of a finite rate quantizer), guarantee that $\xi_3 < \infty$.

We are now ready to prove T.2.5.1.

Proof of statement (a): Define $\tilde{\mathcal{K}} \triangleq \{k : B_{\rm c}^k = 1\}$, $\{\mathbf{y}_{\tilde{\mathcal{K}}}^k\} \triangleq \{\mathbf{y}^{\tilde{k}}\}_{\tilde{k}\in\tilde{\mathcal{K}}}$, $\{\mathbf{c}_{\tilde{\mathcal{K}}}^k\} \triangleq \{\mathbf{c}^{\tilde{k}}\}_{\tilde{k}\in\tilde{\mathcal{K}}}$ and $\{\alpha_{\tilde{\mathcal{K}}}^k\} \triangleq \{\alpha(\tilde{k})\}_{\tilde{k}\in\tilde{\mathcal{K}}}$. It is sufficient to show that V in (2.33) satisfies the conditions of [46, T.1], namely:

$$\begin{aligned}&1) \inf_{\|\mathbf{y}-\bar{y}^0\mathbf{1}\|\geq\epsilon} V(\mathbf{y}) > 0, \forall\epsilon > 0, V(\bar{y}^0\cdot\mathbf{1}) = 0, \text{ and } \lim_{\mathbf{y}\to\bar{y}^0\cdot\mathbf{1}}\sup V(\mathbf{y}) = 0; \\ &2) \mathbb{E}\left[V(\mathbf{y}_{\tilde{K}}^{k+1})|\mathbf{y}_{\tilde{K}}^k = \mathbf{y}\right] - V(\mathbf{y}) \leq g^k\left[1 + V(\mathbf{y})\right] - \alpha_{\tilde{\mathcal{K}}}^k\phi(\mathbf{y}),\end{aligned}$$

where $\phi(\mathbf{y}) \geq 0$ such that $\inf_{\|\mathbf{y}-\bar{y}^0\mathbf{1}\|\geq\epsilon}\phi(\mathbf{y}) > 0, \forall\epsilon > 0$; and $\alpha_{\tilde{\mathcal{K}}}^k$ and g^k satisfy

$$\alpha_{\tilde{\mathcal{K}}}^k > 0, \quad \sum_{k=1}^{\infty}\alpha_{\tilde{\mathcal{K}}}^k = \infty, \quad g^k > 0, \quad \sum_{k=1}^{\infty}g^k < \infty.$$

Conditions in 1) are trivially satisfied by definition [cf. (2.33)]. To prove the condition in 2), we use $\mathbf{1}^\top \mathbf{y}_{\tilde{K}}^{k+1} = \mathbf{1}^\top \mathbf{y}_{\tilde{K}}^{k}$, L.9 in App.2.8.2, and (2.34), yielding,

$$\mathbb{E}\left[V(\mathbf{y}_{\tilde{K}}^{k+1})|\mathbf{y}_{\tilde{K}}^{k} = \mathbf{y}\right] - V(\mathbf{y}) \leq g^k - 2\xi_1 \alpha_{\tilde{K}}^k V(\mathbf{y}),$$

with $g^k = 2\xi_1 \alpha_{\tilde{K}}^k c_{\tilde{K}}^k$ and $\phi(\mathbf{y}) = 2\xi_1 V(\mathbf{y})$. Moreover,

$$\sum_{k\geq 0} g^k \leq 2\xi_1 \sqrt{\left[\sum_{k\in\tilde{\mathcal{K}}} (\alpha^k)^2\right]\left[\sum_{k\in\tilde{\mathcal{K}}} (c^k)^2\right]} \overset{(a)}{<} \infty,$$

where (a) we used $\sum_{k\in\tilde{\mathcal{K}}} (\alpha^k)^2 < \infty$ (cf. A.5); and $\sum_{k\in\tilde{\mathcal{K}}} (c^k)^2 < \infty$, due to (2.35), A.5, and T.2.4.1. Therefore, the condition in 2) holds.

Overall, we have shown that all the conditions of [46, T.1] are satisfied, implying that $\mathbb{P}(\lim_{k\to\infty, k\in\tilde{\mathcal{K}}} \mathbf{y}^k = \bar{y}^0 \cdot \mathbf{1}) = 1$. Since $|\tilde{\mathcal{K}}| = \infty$ and $\mathbf{y}^{k+1} = \mathbf{y}^k$, for all $k \notin \tilde{\mathcal{K}}$, statement (a) of the theorem follows.

Proof of statement (b): Since $\|\mathbf{y}^k - \bar{y}^0\mathbf{1}\|^r < \infty$ and $\mathbb{P}(\lim_{k\to\infty} \|\mathbf{y}^k - \bar{y}^0\mathbf{1}\|^r = 0) = 1$, for all $r \in \mathbb{Z}_{++}$ (recall that $|y_i^k - \bar{y}^0|$ is bounded for all $i \in \mathcal{V}$, cf. L.9 in App.2.8.2), it follows from the dominated convergence theorem (cf. [47, T.1.6.7]) that $\lim_{k\to\infty} \mathbb{E}[\|\mathbf{y}^k - \bar{y}^0\mathbf{1}\|^r] = 0, \forall r > 0$, which implies statement (b).

Proof of statement (c): For simplicity, we assume that $B_{\mathrm{c}}^k = 1, \forall k \in \mathbb{Z}_+$, and the proof can be easily generalized to the case that B_{c}^k satisfying A.3. Since $\alpha^k = \mathcal{O}(1/k)$ and $\|\mathbf{b}^k\|_1 = \mathcal{O}(1/k)$, it follows that $c^k = \mathcal{O}(1/k)$, and there exist $M > 0, C > 0$ such that

$$\alpha^k \leq M/(k+1),\ c^k \leq C/(k+1), \forall k.$$

Under the conditions of the theorem, (2.34) holds, which implies

$$V^{k+1} \leq \left(1 - \frac{\rho}{k+1}\right) V^k + \gamma \frac{1}{(k+1)^2},\ \forall k,$$

where $\gamma \triangleq 2\xi_1 MC$. Let $\bar{k} \triangleq \lceil \rho \rceil - 1$. By induction, we can show $V^k \leq [(\bar{k}+1)V^{\bar{k}} + \gamma]\beta(\bar{k}, k) + \gamma \sum_{t=\bar{k}+1}^{k-1} \beta(t,k), \forall k > \bar{k}$, where $\beta(t,k) \triangleq \frac{1}{(t+1)^2} \prod_{i=t+1}^{k-1}(1 - \frac{\rho}{i+1})$. Note that

$$\begin{aligned}
\ln \beta(t,k) &\leq \int_{t+2}^{k+1} \ln\left(1 - \frac{\rho}{x}\right) \mathrm{d}x - 2\ln(t+1) \\
&= (k+1-\rho)\ln(1-\rho/(k+1)) - (t+2-\rho)\ln(1-\rho/(t+2)) \\
&+ 2\ln((t+2)/(t+1)) + (\rho-2)\ln(t+2) - \rho\ln(k+1).
\end{aligned}$$

The first three terms are bounded, since $\ln((n+2)/(n+1)) \to 0$ and $(n-\rho)\ln(1-\rho/n) \to -\rho$ for $n \to \infty$. It follows that $\beta(t,k) \leq (e)^Q (t+2)^{\rho-2}(k+1)^{-\rho}$, for some $Q < \infty$. Letting $S^k \triangleq (k+1)^{-\rho} \sum_{t=\bar{k}+3}^{k+1} (t)^{\rho-2}$, it follows that

$$V^k \leq A(k+1)^{-\rho} + \gamma(e)^Q S^k, \ \forall k > \bar{k}, \tag{2.36}$$

for some $A < \infty$. To conclude, note that $A(k+1)^{-\rho} = \mathcal{O}((k)^{-\rho})$,

$$S^k \leq (k+1)^{-\rho} \int_1^{k+2} (x)^{\rho-2} \mathrm{d}x = \begin{cases} \mathcal{O}(1/k), & \rho > 1, \\ \mathcal{O}(\ln(k)/k), & \rho = 1, \\ \mathcal{O}(1/(k)^\rho), & 0 < \rho < 1, \end{cases}$$

which proves the desired result.

□

2.6 Numerical Results

In this section, we present some numerical results to validate our theoretical findings on strongly connected digraphs with $m = 50$ agents constructed by the following procedure: a directed ring links all the agents, to ensure strong connectivity (cf. Fig. 2.3). Then directed edges are randomly added, with probability 0.2 on each pair of agents.

2.6.1 Quantized weight-balancing

We adopt (2.12) for $\{\gamma^k\}_{k\in\mathbb{Z}_+}$. We compare the total imbalance $\|\mathbf{b}^k\|_1$ of our proposed scheme with the integer weight-balancing and real weight-balancing schemes in [14]. The real weight-balancing scheme uses real valued communications; the integer weight-balancing scheme uses *unicast* transmissions to each of its out-neighbors to communicate the associated edge weight, and cannot use a prescribed number of bits. As we will see numerically, these features allow the scheme to converge to a weight-balanced solution within finite time. In contrast, our scheme uses *broadcast* communications with a prescribed number of bits per channel use, which in general does not guarantee convergence within finite time. The simulation results are averaged over 100 graph realizations.

Fig. 2.4 shows the total imbalance of Algorithm 1 with 1-bit and 5-bit of information exchange, as well as of the other two benchmark schemes. Note that in the integer weight-balancing scheme, the maximum weights in the 100 realizations are between 88 to 250, implying that 7 to 8 bits are required *per edge* per timeslot, which implies 7 or $8 \times (1 + 0.2 \times 49) = 75.6$ or 86.4 bits per agent per timeslot. It is shown that the metric $\|\mathbf{b}^k\|_1$ is non-increasing for the proposed schemes, which is consistent with our analytical results (cf. L.2). In addition, one can see that the curve of $\|\mathbf{b}^k\|_1$ can be partitioned into nearly flat and steep line segments, for both schemes. The rationale behind this behavior is that the total imbalance decreases only when decreasing events occur (steep line segments); in between, the imbalance may be transferred within the network, but without causing the total imbalance to decrease. Compared with the two benchmark schemes, it shows that the proposed scheme with 50 bits outperforms the real weight-balancing scheme [14], which requires infinite rate communications. On the other hand, The comparison between the proposed 7-bit scheme and the integer weight-balancing scheme shows that, initially, the proposed scheme has better performance. However, as noted earlier, the integer weight-balancing scheme later outperforms the proposed 7-bit scheme since it is guaranteed to converge to a weight-balanced solution in finite timeslots.

2.6.2 Quantized average consensus

We compare our proposed algorithm with the following state-of-the-art schemes: 1) Q-Push-Sum, where we straightforwardly apply the finite-bit probabilistic quantization to the original push-sum algorithm in [37], i.e., $z_i^0 = s_i, \psi_i^0 = 1, \forall i \in \mathcal{V}$,

$$\psi_i^{k+1} = \psi_i^k + \alpha^k \sum_{j \in \mathcal{N}_{-,i}} w_{ij}^k \left[\mathcal{Q}(\psi_j^k) - \mathcal{Q}(\psi_i^k)\right];$$

$$z_i^{k+1} = z_i^k + \alpha^k \sum_{j \in \mathcal{N}_{-,i}} w_{ij}^k \left[\mathcal{Q}(z_j^k) - \mathcal{Q}(z_i^k)\right];$$

and $y_i^k = z_i^k / \psi_i^k$ is the estimate of the initial average, where $\mathcal{Q}(\bullet)$ is the quantization defined in (2.7); note that Q-Push-Sum can be regarded as the generalization of [31] to real valued initialization and finite rate communications; 2) Q-Run-Avg, where we apply the finite-bit probabilistic quantization to the algorithm in [34];[5] 3) Q-Monte-Carlo, where we apply the B-bit quantization $r_\beta(x; B) = (1+\beta)^{\max\left\{\lfloor \log_{1+\beta} x \rfloor, (2)^B\right\}}$ to the Monte-Carlo based algorithm in [33, Sec. 4]. In this algorithm agents exchange quantized random values sampled from the exponential distribution with parameter related to their current states. Note that exact convergence can be achieved by [33], [34] using infinite-bit quantized communications, [37] using real value communications, and [31] using integer value communications. However, there is no theoretical guarantee for all these benchmark schemes with finite bit quantization: our proposed scheme is the first algorithm solving the distributed average consensus over unbalanced digraphs with a prescribed finite rate communications.

We adopt the mean square error (MSE) $\text{MSE}^k = V(\mathbf{y}^k)/m$ as defined in (2.33) as performance metric. The simulation results are averaged over 100 graph realizations and 100 initial value realizations, i.e., totally 10000 realizations.

For the proposed algorithm, we adopt: $q_{\min} = 0, q_{\max} = 1$, $B_{\mathrm{w},i}^k = B_{\mathrm{c},i}^k = 50, \forall i, \forall k$; (2.12) is adopted for $\{\gamma^k\}_{k \in \mathbb{Z}_+}$, and $\alpha^k = 1/(k+1), \forall k \in \mathbb{Z}_+$, which satisfies A.5; for Q-Push-Sum we use 50 bits and $q_{\min} = 0, q_{\max} = m$ to quantize ψ, and 50 bits and $q_{\min} = 0, q_{\max} = 1$ to quantize z; for Q-Run-Avg we quantize each element of $\mathbf{z} \in \mathbb{R}^m$ (the estimate of the left

[5] ↑Note: this algorithm requires $\mathcal{O}(m)$ memory space to store the estimate of the eigenvector of graph Laplacian at each agent.

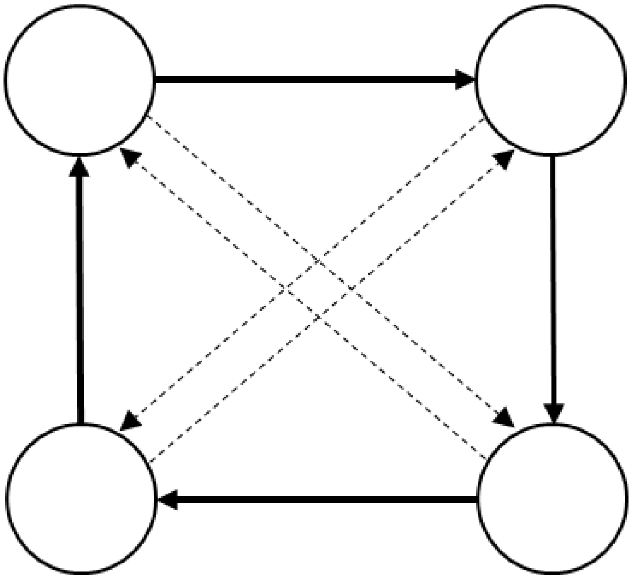


Figure 2.3. Illustration of the random graph model for $m = 4$, where dashed arrows represent potential directed links depending on the realizations.

eigenvector at 0 of graph Laplacian constructed by a row stochastic weight matrix) using 4 bits (i.e., totally $4 \times m = 80$ bits required for quantizing $\mathbf{z}$) and $q_{\min} = 0, q_{\max} = (m)^{\kappa}$ with $\kappa = 1.15$, and y is quantized using 20 bits and $q_{\min} = 0, q_{\max} = 1$; for Q-Monte-Carlo, we use 50 bits to quantize both X and Y, and other parameters are: $a = 0, b = 1, \varepsilon = 10^{-3}$. Note that the communication resource budget per agent per timeslot is 100 bits in all schemes.

Fig. 2.5 shows the MSE performance of Algorithm 2 as well as other benchmark schemes. It is shown that only the proposed scheme and the Q-Monte-Carlo are reaching the average consensus, among all finite rate schemes. Note that Q-Run-Avg and Q-Push-Sum seem also converge for some realizations, cf. Fig. 2.6. However, only the proposed scheme has theoretical convergence guarantees.

Fig. 2.7 shows the communication cost (left y-axis) and delay (right y-axis) needed by Algorithm 2 to reach a target MSE of 1×10^{-3} and 5×10^{-3}, versus the total number of bits per channel use. The communication cost is defined as the product of the total number of bits per agent per timeslot and the number of timeslots. For each parameter setting, we run 50 graph realizations and 10 initial value realizations. To avoid the average results affected by the outliers, we select the best 95% of results to perform averaging. We observe that increasing the total number of bits reduces the number of timeslots required. On the other hand, there exists an optimal number of bits that minimizes the communication cost. Using more bits does not appear to be beneficial, since the communication cost becomes larger, and it is only marginally compensated by the reduction of the number of timeslots required.

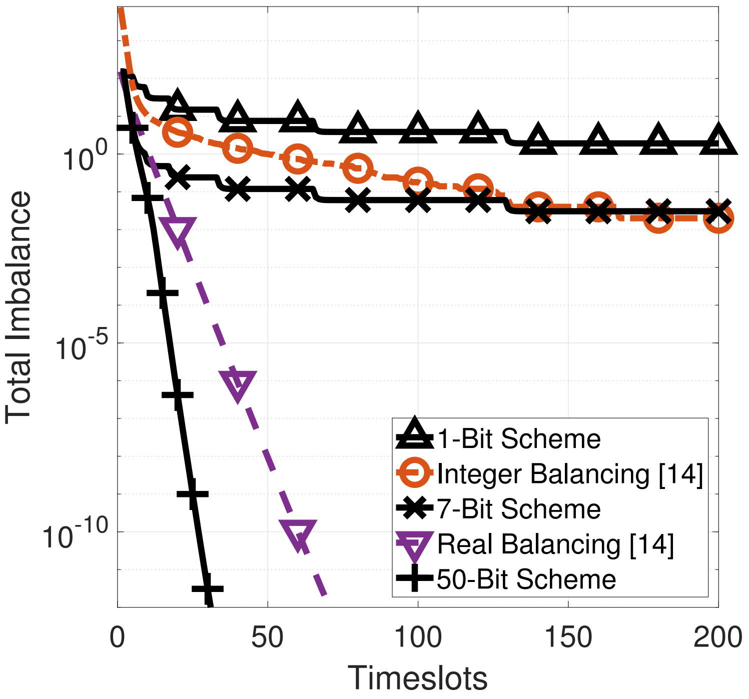


Figure 2.4. Quantized weight-balancing problem: Total imbalance $\|\mathbf{b}^k\|_1$ the propsoed algorithm with 1-bit , 7-bit, and 50-bit, as well as the integer and real weight-balancing schemes [14].

2.7 Conclusions

In this chapter, we introduced a novel distributed algorithm that solves the weight-balancing problem using only quantized information and simplex communications. Building on this scheme, a second contribution of this chapter was a novel distributed average con-

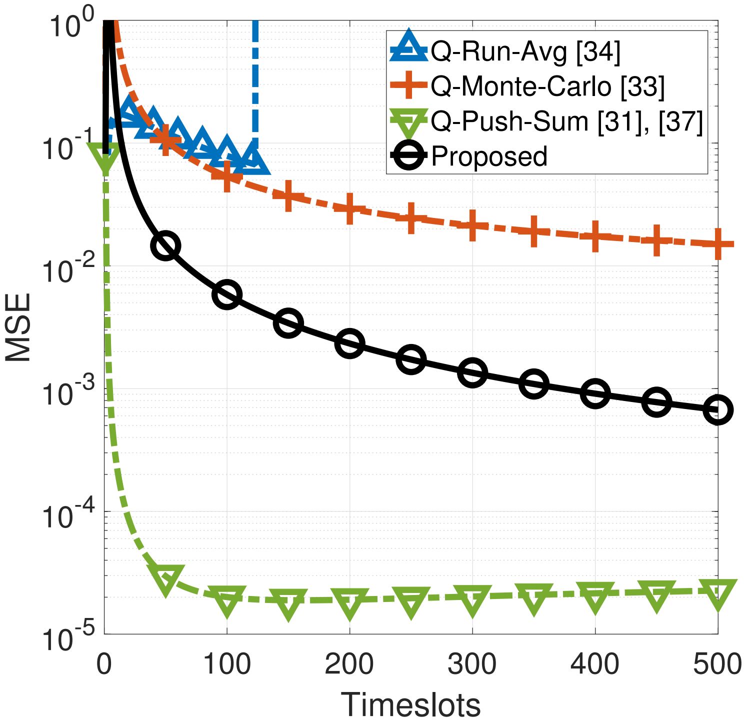


Figure 2.5. Quantized average consensus problem: MSE of average consensus algorithms average over 10000 realizations.

sensus algorithm over non-balanced digraphs that uses quantized simplex communications. Convergence of the algorithm was proved using a novel line of analysis, based on a novel metric inspired by the positional system representation and a new step-size rule. Finally, numerical results validated our theoretical findings.

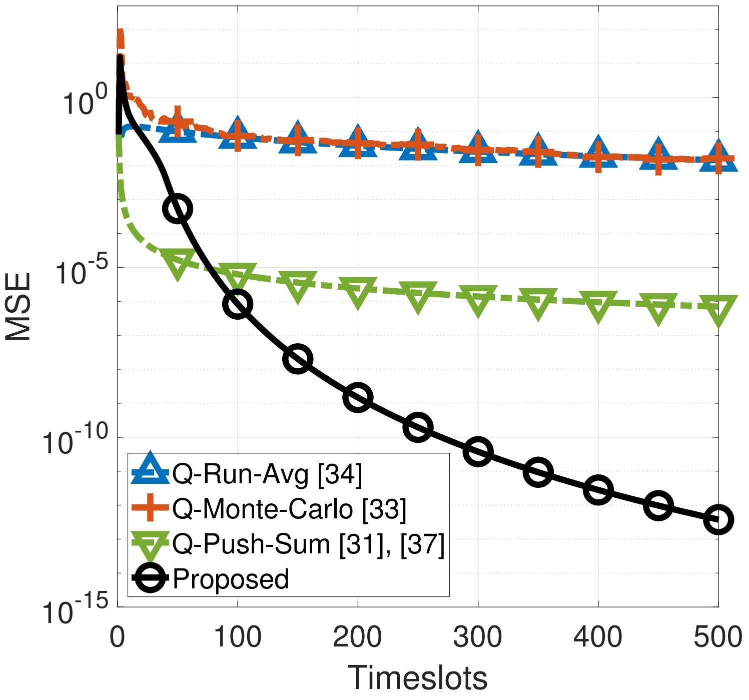


Figure 2.6. Quantized average consensus problem: MSE of average consensus algorithms for a particular realization.

2.8 Appendix: Proofs of Theorems

2.8.1 Intermediate Results in the Proof of Theorem 2.4.1

Preliminary definitions and results

Throughout the proof, we write the updates of $S_{+,i}^k$, $S_{-,i}^k$, b_i^k of Algorithm 1 as

$$S_{+,i}^{k+1} \triangleq \sum_{j=1}^{m} w_{ji}^{k+1} \overset{(2.5)}{=} \sum_{j\in\mathcal{N}_{+,i}} (w_{ji}^k + \gamma^k n_i^k) = S_{+,i}^k + D_{+,i}\gamma^k n_i^k, \tag{2.37}$$

$$S_{-,i}^{k+1} \triangleq \sum_{j=1}^{m} w_{ij}^{k+1} \overset{(2.5)}{=} \sum_{j\in\mathcal{N}_{-,i}} (w_{ij}^k + \gamma^k n_j^k) = S_{-,i}^k + \sum_{j\in\mathcal{N}_{-,i}} \gamma^k n_j^k, \tag{2.38}$$

$$b_i^{k+1} \triangleq S_{-,i}^{k+1} - S_{+,i}^{k+1} = b_i^k - D_{+,i}\gamma^k n_i^k + \sum_{j\in\mathcal{N}_{-,i}} \gamma^k n_j^k. \tag{2.39}$$

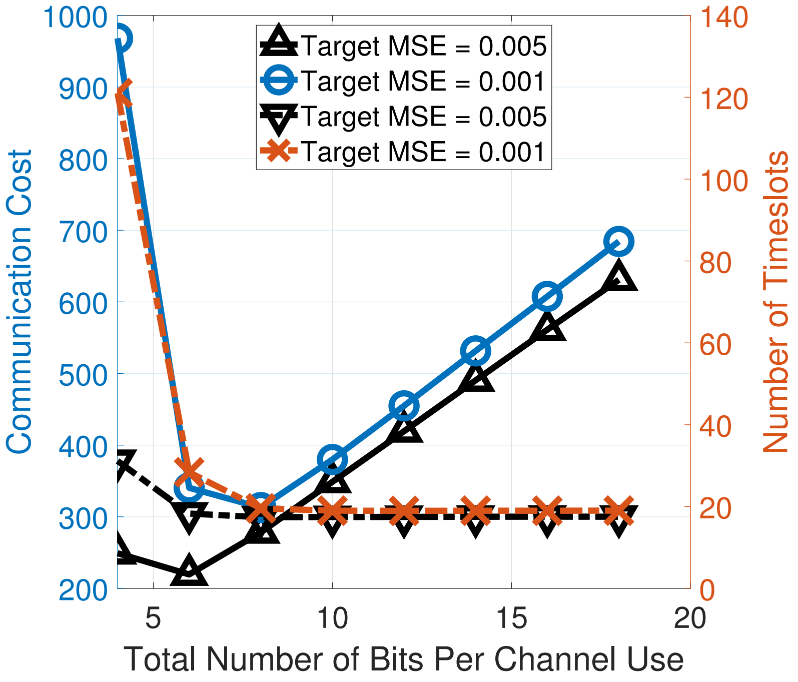


Figure 2.7. Quantized consensus problem: Communication cost (left y-axis, solid lines) and number of timeslots (right y-axis, dashed lines) needed to reach the target MSE, versus the total number of bits per channel use.

Lemma 5. *Given $\mathcal{V}_+^k$ and $\mathcal{V}_-^k$, defined in (2.16), it holds:*

$$\mathbb{1}(\mathcal{D}^k) = 0 \;\Rightarrow\; \mathcal{V}_+^{k+1} \supseteq \mathcal{V}_+^k \textit{ and } \mathcal{V}_-^{k+1} \supseteq \mathcal{V}_-^k.$$

Proof. Let $\mathbb{1}\{\mathcal{D}^k\} = 0$ and consider $i \in \mathcal{V}_-^k$. Then, (2.4) and $b_i^k < 0$ imply $\gamma^k n_i^k > b_i^k / D_{+,i}$; $\mathbb{1}\{\mathcal{D}^k\} = 0$ (see (2.13)) implies $n_j^k \leq 0, \forall j \in \mathcal{N}_{-,i}$. $b_i^{k+1} < 0$ then follows from (2.39), so that $i \in \mathcal{V}_-^{k+1}$; hence $\mathcal{V}_-^{k+1} \supseteq \mathcal{V}_-^k$. $\mathcal{V}_+^{k+1} \supseteq \mathcal{V}_+^k$ follows from a similar argument on $i \in \mathcal{V}_+^k$. □

Lemma 6. $\forall k \in \mathbb{Z}_+, \exists \tau \geq k : \|\mathbf{b}^\tau\|_1 < (m)^2 \gamma^\tau$.

Proof. We prove it by contradiction. Let $\bar{T}_0 \triangleq (2W-1)\bar{U}$. Suppose $\exists k \in \mathbb{Z}_+ : \|\mathbf{b}^\tau\|_1 \geq (m)^2\gamma^\tau$, for all $\tau \geq k$. Invoking C.1.*(ii)* recursively m times and taking $m \to \infty$ yields

$0 \leq \|\mathbf{b}^{k+m\bar{T}_0}\|_1 \leq \|\mathbf{b}^k\|_1 - 2\sum_{n=1}^m \gamma^{k+n\bar{T}_0}$, a contradiction since $\sum_{n=1}^m \gamma^{k+n\bar{T}_0} \to \infty$ due to (2.11). □

Proof of Lemma 3

By the definition of Cauchy sequence applied to each entry of $\mathbf{W}^k$, we need to prove that, $\forall \epsilon > 0, \exists k_\epsilon \in \mathbb{Z}_+$ such that $\max_{i,j} |w_{ij}^{k'} - w_{ij}^k| < \epsilon, \forall k', k \geq k_\epsilon$. To this end, let $\epsilon > 0$ and define k_ϵ as[6]

$$k_\epsilon = \min\{k : \|\mathbf{b}^k\|_1 < \epsilon/(2\bar{U}), \gamma^k < \epsilon/(4\bar{U})\}. \tag{2.40}$$

Since w_{ji}^k is updated only at update or decreasing events, using (2.5) recursively, we infer $|w_{ji}^k - w_{ji}^{k_\epsilon}| \leq \sum_{\ell=k_\epsilon}^\infty \mathbb{1}\{\mathcal{U}^\ell \vee \mathcal{D}^\ell\}\gamma^\ell$, $\forall k \geq k_\epsilon$. With t^l defined as in (2.14) and letting $L_\epsilon \triangleq \min\{l \in \mathbb{Z}_+ : t^l \geq k_\epsilon\}$, we can further upper bound $|w_{ji}^k - w_{ji}^{k_\epsilon}| \leq \gamma^{k_\epsilon} \sum_{\ell=k_\epsilon}^{t_{L_\epsilon}-1} \mathbb{1}\{\mathcal{U}^\ell \vee \mathcal{D}^\ell\} + \sum_{l=L_\epsilon}^\infty \gamma^{t^l} \sum_{\ell=t^l}^{t^{l+1}-1} \mathbb{1}\{\mathcal{D}^\ell \vee \mathcal{U}^\ell\}$. Since there are at most $\bar{U}$ update events between the two consecutive decreasing events at times t^l and t^{l+1} (cf. C.1), it follows that $\sum_{\ell=t^l}^{t^{l+1}-1} \mathbb{1}\{\mathcal{D}^\ell \vee \mathcal{U}^\ell\} \leq \bar{U}$, hence

$$|w_{ji}^k - w_{ji}^{k_\epsilon}| \leq \Big(\gamma^{k_\epsilon} + \sum_{l=L_\epsilon}^\infty \gamma^{t^l}\Big)\bar{U}. \tag{2.41}$$

To bound $\sum_{l=L_\epsilon}^\infty \gamma^{t^l}$, we apply recursively L.2,

$$-\|\mathbf{b}^{k_\epsilon}\|_1 = \sum_{\ell=k_\epsilon}^\infty (\|\mathbf{b}^{\ell+1} - \|\mathbf{b}^\ell\|_1) \leq -2\sum_{\ell=k_\epsilon}^\infty \gamma^\ell \mathbb{1}\{\mathcal{D}^\ell\} = -2\sum_{l=L_\epsilon}^\infty \gamma^{t^l}, \tag{2.42}$$

hence $\sum_{l=L_\epsilon}^\infty \gamma^{t^l} \leq \|\mathbf{b}^{k_\epsilon}\|_1/2 \leq \epsilon/(2\bar{U})$. By combining (2.41) with (2.42) and (2.40), we finally obtain, $\forall k \geq k_\epsilon$,

$$|w_{ji}^k - w_{ji}^{k_\epsilon}| \leq \Big(\gamma^{k_\epsilon} + \frac{\|\mathbf{b}^{k_\epsilon}\|_1}{2}\Big)\bar{U} < \epsilon/2,$$

[6] ↑Note that $k_\epsilon < \infty$ since $\|\mathbf{b}^k\|_1 \to 0$ and $\gamma^k \to 0$, see (2.11).

and, $\forall k, k' \geq k_\epsilon$,

$$|w_{ji}^{k} - w_{ji}^{k'}| \leq |w_{ji}^{n} - w_{ji}^{k_\epsilon}| + |w_{ji}^{n'} - w_{ji}^{k_\epsilon}| < \epsilon,$$

which proves that $\{\mathbf{W}^k\}_{k\in\mathbb{Z}_+}$ is a Cauchy sequence.

Proof of Lemma 2

We first introduce the following intermediate result.

Lemma 7. *Let $\{\mathbf{b}^k\}_{k\in\mathbb{Z}_+}$ be the sequence generated by Algorithm 2. Then, $b_i^k/\bar{\gamma}^k \in \mathbb{Z}, \forall i \in \mathcal{V}$ and $k \in \mathbb{Z}_+$.*

Proof. We prove this lemma by induction using (2.39). The induction hypothesis holds at $k = 0$, since $w_{ij}^0 = 1, \forall (j,i) \in \mathcal{E}$ and $\bar{\gamma}^0 = (2)^{1-B_{\max}}$ (cf. A.2). Suppose that it holds at $k \geq 0$, i.e., $b_i^k/\bar{\gamma}^k \in \mathbb{Z}, \forall i$. Then, since $\bar{\gamma}^{k+1} = \bar{\gamma}^k/m^k$, with $m^k \in \{1, c_1\} \subset \mathbb{Z}_{++}$ (A.2), it follows that $b_i^k/\bar{\gamma}^{k+1} = m^k b_i^k/\bar{\gamma}^k \in \mathbb{Z}$ and $\gamma^k n_i^k/\bar{\gamma}^{k+1} = m^k n_i^k (2)^{B_{\max}-1} \in \mathbb{Z}$. Therefore, by (2.39), one can infer that $b_i^{k+1}/\bar{\gamma}^{k+1} \in \mathbb{Z}$, proving the induction step and completing the proof. □

Proof of Lemma 2: Let $i \in \mathcal{V}$ and $k \in \mathbb{Z}_+$. Using (2.39), we find

$$\frac{|b_i^{k+1}|}{\gamma^k} = \left| \frac{b_i^k}{\gamma^k} - D_{+,i} n_i^k + \sum_{j\in\mathcal{N}_{-,i}} n_j^k \right|. \tag{2.43}$$

It will be useful to note that, as can be seen from (2.4), b_i^k, n_i^k (possibly, $n_i^k = 0$) and $b_i^k - D_{+,i} n_i^k$ have the same signs, yielding the following inequality for all i,

$$\frac{|b_i^{k+1}|}{\gamma^k} \overset{\triangle}{\leq} \frac{|b_i^k|}{\gamma^k} - D_{+,i}|n_i^k| + \sum_{j\in\mathcal{N}_{-,i}} |n_j^k|, \tag{2.44}$$

where $\triangle$ stands for the triangle inequality. We now distinguish the two cases $\mathbb{1}\{\mathcal{D}^k\} = 0$ and $\mathbb{1}\{\mathcal{D}^k\} = 1$. If $\mathbb{1}\{\mathcal{D}^k\} = 0$, from the negation of $\mathcal{D}^k$ in (2.13) and from (2.4), it follows that $b_i^k, n_i^k, b_i^k - D_{+,i} n_i^k$, and $n_j^k, \forall j \in \mathcal{N}_{-,i}$ have the same signs, so that (2.44) holds with equality, $\forall i$.

Conversely, if $\mathbb{1}\{\mathcal{D}^k\} = 1$, there exists ℓ and $(j, j') \in [\mathcal{N}_{-,\ell}]^2$ such that either 1) $b_\ell^k \neq 0$, $n_\ell^k \neq 0$ and $\operatorname{sgn}(n_j^k) = -\operatorname{sgn}(b_\ell^k)$; or 2) $b_\ell^k = 0$, $n_\ell^k = 0$ and $\operatorname{sgn}(n_j^k) = -\operatorname{sgn}(n_{j'}^k)$. In the first case ($b_\ell^k \neq 0$, $n_\ell^k \neq 0$ and $\operatorname{sgn}(n_j^k) = -\operatorname{sgn}(b_\ell^k)$), we bound (2.43) as

$$\begin{aligned}
\frac{|b_\ell^{k+1}|}{\gamma^k} &\overset{\triangle}{\leq} \left| \frac{b_\ell^k}{\gamma^k} - d_\ell^+ n_\ell^k + n_j^k \right| + \sum_{i \in \mathcal{N}_{-,\ell}, i \neq j} |n_i^k| \\
&\leq \left| \frac{|b_\ell^k|}{\gamma^k} - d_\ell^+ |n_\ell^k| - |n_j^k| \right| + \sum_{i \in \mathcal{N}_{-,\ell}, i \neq j} |n_i^k| \\
&= \frac{|b_\ell^k|}{\gamma^k} - d_\ell^+ |n_\ell^k| - 2 \min \left\{ |n_j^k|, \frac{|b_\ell^k|}{\gamma^k} - d_\ell^+ |n_\ell^k| \right\} + \sum_{i \in \mathcal{N}_{-,\ell}} |n_i^k|.
\end{aligned}$$

In the second case ($b_\ell^k = 0$, $n_\ell^k = 0$, $\operatorname{sgn}(n_j^k) = -\operatorname{sgn}(n_{j'}^k)$ and, without loss of generality, $|n_j^k| \geq |n_{j'}^k| \geq (2)^{1-B_{\max}}$), we bound instead

$$\frac{|b_\ell^{k+1}|}{\gamma^k} \overset{\triangle}{\leq} \left| \frac{b_\ell^k}{\gamma^k} - d_\ell^+ n_\ell^k \right| + \left| n_j^k + n_{j'}^k \right| + \sum_{i \in \mathcal{N}_{-,\ell}, i \neq j, j'} |n_i^k| \leq \frac{|b_\ell^k|}{\gamma^k} - d_\ell^+ |n_\ell^k| + \sum_{i \in \mathcal{N}_{-,\ell}} |n_i^k| - 2|n_{j'}^k|.$$

In both cases, since $|n_{j'}^k| \geq (2)^{1-B_{\max}}$, $|n_j^k| \geq (2)^{1-B_{\max}}$ and $\gamma^k |n_\ell^k| \leq |b_\ell^k| / d_\ell^+$ (see (2.4)), we further bound

$$\frac{|b_\ell^{k+1}|}{\gamma^k} \leq \frac{|b_\ell^k|}{\gamma^k} - d_\ell^+ |n_\ell^k| + \sum_{i \in \mathcal{N}_{-,\ell}} |n_i^k| - (2)^{2-B_{\max}}. \tag{2.45}$$

By summing (2.44) (with strict equality if $\mathbb{1}\{\mathcal{D}^k\} = 0$) and (2.45) over $i \in \mathcal{V}$, it holds

$$\|\mathbf{b}^{k+1}\|_1 \begin{cases} = \|\mathbf{b}^k\|_1, & \text{if } \mathbb{1}\{\mathcal{D}^k\} = 0 \\ \leq \|\mathbf{b}^k\|_1 - 2\bar{\gamma}^k, & \text{otherwise,} \end{cases}$$

after noticing that $j \in \mathcal{N}_{-,i} \Leftrightarrow i \in \mathcal{N}_{+,j}$, hence $\sum_{i, j \in \mathcal{N}_{-,i}} |n_j^k| = \sum_{j, i \in \mathcal{N}_{+,j}} |n_j^k| = \sum_{j=1}^m |n_j^k| d_j^+$. This completes the proof.

Proof of Proposition 2.4.1

Property (i): Note that $U^k \geq 0$ since $|b_i^k| \geq 0$, for all $i \in \mathcal{V}$. To show that it is upper bounded, we use $U_h \leq U_1, \forall h \geq 1$ and $\cup_h \mathcal{V}_h^k \subseteq \mathcal{V}$, and write

$$U^k \leq \sum_{h=1}^{m-1} U_h \sum_{i \in \mathcal{V}_h^k} m \leq U_1 m \sum_{h=1}^{m-1} |\mathcal{V}_h^k| \leq (m)^{2(m-1)}.$$

We prove that U^k is nondecreasing as by product of the proof of Property (ii), as given below.

Property (ii): Since $\bar{\gamma}^{k+1} \leq \bar{\gamma}^k$, it follows that $U^{k+1} \geq \sum_{h=1}^{m-1} U_h \sum_{j \in \mathcal{V}_h^{k+1}} \min\{|b_j^{k+1}|/\bar{\gamma}^k, m\}$. **Case 1:** $\mathbb{1}(\mathcal{D}^k) = \mathbb{1}(\mathcal{U}^k) = 0$. We have $b_j^{k+1} = b_j^k, \forall j$ and $\mathcal{V}_h^{k+1} \triangleq \mathcal{V}_h^k, \forall n$, which implies $U^{k+1} \geq U^k$. **Case 2:** $\mathbb{1}(\mathcal{U}^k) = 1$. From the discussion following (2.44), (2.44) holds with equality:

$$\frac{|b_i^{k+1}|}{\gamma^k} \triangleq \frac{|b_i^k|}{\gamma^k} - D_{+,i} n_i^k + \sum_{j \in \mathcal{N}_{-,i}} |n_j^k|, \ \forall i. \tag{2.46}$$

Moreover, $\exists i \in \mathcal{V} : n_i^k \neq 0$; this implies that there exists a non-empty set of agents that receive at least one update from their in-neighbors, defined as

$$\mathcal{R}^k = \{i \in \mathcal{V} : n_j^k \neq 0, \exists j \in \mathcal{N}_{-,i}\}.$$

It is straightforward to show that

$$\mathcal{R}^k \subseteq \mathcal{V}_+^{k+1} \cup \mathcal{V}_-^{k+1}. \tag{2.47}$$

In fact, if $i \notin \mathcal{V}_+^{k+1} \cup \mathcal{V}_-^{k+1}$ (i.e., $b_i^{k+1} = 0$), it follows that $i \notin \mathcal{V}_+^k \cup \mathcal{V}_-^k$ (i.e., $b_i^k = 0$ and $n_i^k = 0$, cf. L.5); therefore, setting $b_i^{k+1} = b_i^k = 0$ and $n_i^k = 0$ in (2.46), we find that $n_j^k = 0, \ \forall j \in \mathcal{N}_{-,i}$, so that $i \notin \mathcal{R}^k$ and (2.47) follows. With this definition, let

$$h^* = \min\{h \in \{1, 2, \ldots, m-1\} : |\mathcal{V}_h^{k+1} \cap \mathcal{R}^k| > 1\}$$

be the distance of the agent closest to those of opposite sign of balance at $k+1$ to receive the update, and let $\ell \in \mathcal{V}_{h^*}^{k+1} \cap \mathcal{R}^k$ be one of such agents. Then, we have

$$n_i^k = 0,\ \forall i \in \mathcal{V}_h^{k+1},\ \forall h \leq h^*. \tag{2.48}$$

In fact, if $n_i^k \neq 0$ for some of such i, then $\exists j \in \mathcal{N}_i^+ \cap \mathcal{V}_{h-1}^{k+1} \cap \mathcal{R}^k$ receiving the update, which contradicts the definition of h^*. Reading (2.46) at $i \in \mathcal{V}_h^{k+1}, h \leq h^*$, yields

$$\frac{|b_i^{k+1}|}{\bar{\gamma}^k} = \frac{|b_i^k|}{\bar{\gamma}^k} + \sum_{j \in \mathcal{N}_{-,i}} \frac{|n_j^k|}{(2)^{1-B_{\max}}} \geq \frac{|b_i^k|}{\bar{\gamma}^k} + \mathbb{1}\{i \in \mathcal{R}^k\}.$$

We can then further lower bound U^{k+1} as

$$U^{k+1} \geq \sum_{h=1}^{h^*} U_h \sum_{j \in \mathcal{V}_h^{k+1}} \min\left\{\frac{|b_j^k|}{\bar{\gamma}^k}, m\right\} \mathbb{1}\{(h,j) \neq (h^*, \ell)\} + U_{h^*} \min\left\{\frac{|b_i^k|}{\bar{\gamma}^k} + 1, m\right\}, \tag{2.49}$$

where we neglected the non-negative terms associated to $\mathcal{V}_h^{k+1}, h > h^*$. To further bound this quantity, note that $n_\ell^k = 0$ (cf. (2.48)), which, together with $B_{\mathrm{w},i}^k > 0, \forall i$ for an update event to occur, implies $b_\ell^k/\bar{\gamma}^k \leq D_{+,i} \leq m-1$ (cf. (2.4)). Therefore $\min\{|b_\ell^k|/\bar{\gamma}^k + 1, m\} = \min\{|b_\ell^k|/\bar{\gamma}^k, m\} + 1$. Finally, we use the fact that

$$0 \geq \sum_{h=h^*+1}^{m-1} U_h \sum_{j \in \mathcal{V}_h^{k+1}} \left(\min\left\{\frac{|b_j^k|}{\bar{\gamma}^k}, m\right\} - m\right),$$

yielding

$$U^{k+1} \geq \sum_{j \in \mathcal{V}} \min\left\{\frac{|b_j^k|}{\bar{\gamma}^k}, m\right\} \sum_{h=1}^{m-1} U_h \mathbb{1}\{j \in \mathcal{V}_h^{k+1}\} + U_{h^*} - m \sum_{h=h^*+1}^{m-1} U_h |\mathcal{V}_h^{k+1}|. \tag{2.50}$$

Now, using $U_h \leq U_{h^*+1}, \forall h > h^*$ and $\cup_{h=h^*+1}^{m-1} \mathcal{V}_h^{k+1} \subseteq \mathcal{V} \setminus \{\ell\}$, we obtain

$$U_{h^*} - m \sum_{h=h^*+1}^{m-1} U_h |\mathcal{V}_h^{k+1}| \geq U_{h^*} - m U_{h^*+1} \sum_{h=h^*+1}^{m-1} |\mathcal{V}_h^{k+1}| \geq (m)^{2(m-h^*)-3} \geq m. \tag{2.51}$$

In the last inequality, we used the fact that $h^* \leq m-2$. In fact, if $h^* = m-1$, then (2.48) implies that $n_i^k = 0, \forall i$, which contradicts the occurrence of the update event. Finally, note that $\mathcal{V}_+^k \subseteq \mathcal{V}_+^{k+1}$ and $\mathcal{V}_-^k \subseteq \mathcal{V}_-^{k+1}$ (cf. L.5), hence $j \in \mathcal{V}_h^k \Rightarrow j \in \cup_{n=1}^h \mathcal{V}_n^{k+1}$, i.e., j gets closer to agents of opposite sign. Together with $U_h > U_{h+1}$, it implies

$$\sum_{h=1}^{m-1} U_h \mathbb{1}\{j \in \mathcal{V}_h^{k+1}\} \geq \sum_{h=1}^{m-1} U_h \mathbb{1}\{j \in \mathcal{V}_h^k\}. \tag{2.52}$$

The desired result follows by using (2.51)-(2.52) in (2.50).

Property (iii): We prove it by contradiction. Assume that $t_1, t_2 \in [k, t^{l+1})$ such that $\mathcal{U}^{t_1}$ and $\mathcal{U}^{t_2}$ are two consecutive update events with $t_2 - t_1 > 2W-1$, and $\|\mathbf{b}^t\|_1 \geq (m)^2\gamma^t, \forall t \in [k, t^{l+1})$. It follows that $\exists i \in \mathcal{V}$ such that $|b_i^{t_1}| \geq m\gamma^{t_1} \overset{(A.2)}{\geq} m\gamma^t > D_{+,i}\gamma^t, \forall t \in [t_1, t_1 + 2W - 1]$, which implies that $\exists t \in [t_1, t_1 + 2W - 1] : \mathbb{1}\{\mathcal{U}^t\} = 1$ due to A.1, which contradicts the assumption that $\mathcal{U}^{t_1}$ and $\mathcal{U}^{t_2}$ are two consecutive update events since $t_1 + 2W - 1 < t_2$. Hence, it follows from property *(ii)* that $U^{k+2W-1} \geq U^k + m$, which proves property *(iii)*. ■

2.8.2 Auxiliary Results for Theorem 2.5.1

Lemma 8. *Let $\{\mathbf{y}^k\}_{k \in \mathbb{Z}_+}$ be the sequence generated by Algorithm 2, in the setting of T.2.5.1. Then, $\mathbf{1}^\top \mathbf{y}^k = \mathbf{1}^\top \mathbf{y}^0$.*

Proof. From (2.32) and $\mathbf{L}_+^k = \mathbf{S}_+^k - \mathbf{W}^k$, it follows

$$\mathbf{1}^\top \mathbf{y}^{k+1} = \mathbf{1}^\top \mathbf{y}^k - \alpha^k \mathbf{1}^\top (\mathbf{S}_+^k - \mathbf{W}^k)\mathbf{x}^k, \tag{2.53}$$

so that the statement of the lemma readily follows after noticing that $\mathbf{1}^\top \mathbf{W}^k = \mathbf{S}_+^k$. □

Lemma 9. *Let $\{\mathbf{y}^k\}_{k \in \mathbb{Z}_+}$ be the sequence generated by Algorithm 2, in the setting of T.2.5.1. Then,*

$$y_{\min,i} \leq y_{\min,i}^k \leq y_i^k \leq y_{\max,i}^k \leq y_{\max,i}, \;\; \forall k \in \mathbb{Z}_+,$$

where $y_{\max,i} \triangleq \lim_{t\to\infty} y^t_{\max,i} < \infty, y_{\min,i} \triangleq \lim_{t\to\infty} y^t_{\min,i} > -\infty,\ q^* \triangleq \max\{|q_{\min}|, |q_{\max}|\}$, $S_{\max} \triangleq \sup_{k\in\mathbb{Z}_+} S^k_{-,i} < \infty$,

$$y^k_{\max,i} \triangleq \max\{q_{\max}, y^0_i\} + \alpha^0\|\mathbf{b}^0\|_1 q^* + \alpha^0 S_{\max}\left(q_{\max} - q_{\min}\right) + q^* \sum_{t=0}^{k-1} \alpha^t |b^t_i|,$$
$$y^k_{\min,i} \triangleq \min\{q_{\min}, y^0_i\} - \alpha^0\|\mathbf{b}^0\|_1 q^* - \alpha^0 S_{\max}\left(q_{\max} - q_{\min}\right) - q^* \sum_{t=0}^{k-1} \alpha^t |b^t_i|.$$

Proof. We first show that $|y_{\max,i}|, |y_{\min,i}| < \infty$. From T.2.4.1, we know that $\mathbf{W}^k$ is bounded for all $k \in \mathbb{Z}_+$, which implies $S^k_{-,i} \le S_{\max} < \infty$. On the other hand, since $|b^t_i| \le \|\mathbf{b}^t\|_1 = \mathcal{O}(1/t)$ and $\sum_{t\in\mathbb{Z}_+} (\alpha^t)^2 < \infty$, one can verify using Cauchy-Schwarz inequality that $\sum_{t=0}^{\infty} \alpha^t |b^t_i| < \infty$ and thus $|y_{\max,i}|, |y_{\min,i}| < \infty$. By inspection, it is also clear that $y^k_{\max,i} \le y_{\max,i}$ and $y^k_{\min,i} \ge y_{\min,i}, \forall k$. We now prove $y^k_i \in [y^k_{\min,i}, y^k_{\max,i}]$ by induction. Clearly, it for $k = 0$. Now, assume it holds for some $k \ge 0$, we prove that this implies $y^{k+1}_i \in [y^{k+1}_{\min,i}, y^{k+1}_{\max,i}]$ (induction step). We have:

1) If $y^k_i \le q_{\max}$, then

$$\begin{aligned} y^{k+1}_i &= y^k_i + \alpha^k b^k_i x^k_i + \alpha^k \sum_{j\in\mathcal{N}_{-,i}} w^k_{ij}(x^k_j - x^k_i) \\ &\le q_{\max} + \alpha^k |b^k_i| q^* + \alpha^k S^k_{-,i}\left(q_{\max} - q_{\min}\right) \\ &\le \max\{q_{\max}, y^0_i\} + \alpha^0\|\mathbf{b}^0\|_1 q^* + \alpha^0 S_{\max}\left(q_{\max} - q_{\min}\right) = y^0_{\max,i} \le y^{k+1}_{\max,i}. \end{aligned}$$

2) If $y^k_i > q_{\max}$, then $x^k_i = q_{\max}$, so (2.9) yields $y^{k+1}_i \le y^k_i + \alpha^k b^k_i q_{\max} \le y^k_{\max,i} + \alpha(k)|b_i(k)|q^* = y^{k+1}_{\max,i}$.

3) If $y^k_i \ge q_{\min}$ then $y^{k+1}_i \ge \min\{q_{\min}, y^0_i\} - \alpha^0\|\mathbf{b}^0\|_1 q^* - \alpha^0 S_{\max}\left(q_{\max} - q_{\min}\right) = y^0_{\min,i} \ge y^{k+1}_{\min,i}$.

4) If $y^k_i < q_{\min}$, then $x^k_i = q_{\min}$ so (2.9) yields $y^{k+1}_i \ge y^k_i - \alpha^k|b^k_i|q^* \ge y^k_{\min,i} - \alpha^k|b^k_i|q^* = y^{k+1}_{\min,i}$. □

Lemma 10. *Let* $\{\mathbf{y}^k\}_{k\in\mathbb{Z}_+}$ *be the sequence generated by Algorithm 2, in the setting of T.2.5.1. Then,* $\exists \xi_1, \xi_2 > 0$ *such that*

$$\mathbf{y}^{k\top} \mathbf{L}^k_+ \tilde{\mathbf{y}}^k \ge \xi_1 V(\mathbf{y}^k) - \xi_2 \|\mathbf{b}^k\|_1. \tag{2.54}$$

Proof. Let $\hat{\mathbf{e}}^k = \mathbf{y}^k - \tilde{\mathbf{y}}^k$ be the saturation error, $\mathbf{S}^k_\pm = \text{diag}\{S^k_{\pm,i}, \forall i\}$, $\mathbf{B}^k = \text{diag}\{\mathbf{b}^k\} = \mathbf{S}^k_- - \mathbf{S}^k_+$, $\mathbf{L}^k_\pm = \mathbf{S}^k_\pm - \mathbf{W}^k$, $\mathbf{L}^k = (\mathbf{S}^k_+ + \mathbf{S}^k_-) - (\mathbf{W}^k + \mathbf{W}^{k\top})$. The proof contains three steps:
Step 1: We will lower bound $\mathbf{y}^{k\top}\mathbf{L}^k_+\tilde{\mathbf{y}}^k$ as

$$\mathbf{y}^{k\top}\mathbf{L}^k_+\tilde{\mathbf{y}}^k \geq -\mathbf{y}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \frac{1}{2}\tilde{\mathbf{y}}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \frac{1}{2}\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k. \tag{2.55}$$

Step 2: we will show that the last term of the RHS in Step 1 satisfies, for some $\xi_4 > 0$,

$$\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k \geq \xi_4 V(\mathbf{y}^k). \tag{2.56}$$

Step 3: by combining the above results, we will show that, for some constants $\xi_1, \xi_2 > 0$,

$$\mathbf{y}^{k\top}\mathbf{L}^k_+\tilde{\mathbf{y}}^k \geq \xi_1 V(\mathbf{y}^k) - \xi_2\|\mathbf{b}^k\|_1.$$

In the following, we provide detailed derivations of each step. **Step 1:** It is easy to show that

$$\mathbf{y}(k)^{k\top}\mathbf{L}^k_+\tilde{\mathbf{y}}^k = -\mathbf{y}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \mathbf{y}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k.$$

The term $\mathbf{y}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k$ can be lower bounded as

$$\begin{aligned}\mathbf{y}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k = \hat{\mathbf{e}}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k + \tilde{\mathbf{y}}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k \overset{(a)}{\geq} \tilde{\mathbf{y}}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k &= \frac{1}{2}\tilde{\mathbf{y}}^{k\top}(\mathbf{L}^k_- + \mathbf{L}^{k\top}_-)\tilde{\mathbf{y}}^k \\ &= \frac{1}{2}\tilde{\mathbf{y}}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \frac{1}{2}\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k,\end{aligned}$$

where (a) comes from the fact that

$$\hat{\mathbf{e}}^{k\top}\mathbf{L}^k_-\tilde{\mathbf{y}}^k = \hat{\mathbf{e}}^{k\top}[\mathbf{S}^k_- - \mathbf{W}^k]\tilde{\mathbf{y}}^k = \sum_{i=1}^m \left[\hat{e}^k_i \sum_{j=1}^m w^k_{ij}(\tilde{y}^k_i - \tilde{y}^k_j)\right] \geq 0,$$

where the last inequality comes from the fact that (i) if $y^k_i \in [q_{\min}, q_{\max}]$, then $\hat{e}^k_i = 0$; (ii) if $y^k_i > q_{\max}$, then $\hat{e}^k_i > 0$ and $\tilde{y}^k_i - \tilde{y}^k_j = q_{\max} - \tilde{y}^k_j \geq 0, \forall j \in \mathcal{V}$; and (iii) if $y^k_i < q_{\min}$, then

$\hat{e}_i^k < 0$ and $\tilde{y}_i^k - \tilde{y}_j^k = q_{\min} - \tilde{y}_j^k \leq 0, \forall j \in \mathcal{V}$.

Step 2: First, one can verify that

$$\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k = \frac{1}{2}\sum_{i,j=1}^{m}(w_{ij}^k + w_{ji}^k)(\tilde{y}_i^k - \tilde{y}_j^k)^2.$$

Let $i^* \in \arg\max_i\{y_i^k\}$, $j^* \in \arg\min_i\{y_i^k\}$, $j^* \neq i^*$. Note that $y_{i^*}^k \geq \bar{y}^0 \geq y_{j^*}^k$ to preserve the average (L.8). Since $\mathcal{G}$ is strongly connected, there exists a path from i^* to j^*. Let $\{i_1, \cdots, i_p\}$ be the set of agents in the shortest path from i^* to j^*, with $i_1 = i^*, i_p = j^*$ and $i_{n+1} \in \mathcal{N}_{+,i_n}, \forall n \in [p-1]$. We have

$$\begin{aligned}
\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k &= \frac{1}{2}\sum_{i,j=1}^{m}(w_{ij}^k + w_{ji}^k)(\tilde{y}_i^k - \tilde{y}_j^k)^2 \\
&\geq \frac{1}{2}\sum_{l=1}^{p-1}(w_{i_l i_{l+1}}^k + w_{i_{l+1} i_l}^k)(\tilde{y}_{i_l}^k - \tilde{y}_{i_{l+1}^k})^2 \\
&\overset{(a)}{\geq} \frac{w_{\min}}{2}\sum_{l=1}^{p-1}(\tilde{y}_{i_l}^k - \tilde{y}_{i_{l+1}^k})^2 \\
&\overset{(b)}{\geq} \frac{w_{\min}}{2(p-1)}\Big[\sum_{l=1}^{p-1}(\tilde{y}_{i_l}^k - \tilde{y}_{i_{l+1}}^k)\Big]^2 \geq \frac{w_{\min}}{2(m-1)}(\tilde{y}_{i^*}^k - \tilde{y}_{j^*}^k)^2,
\end{aligned} \tag{2.57}$$

where (a) follows from $w_{i_{l+1}i_l}^k \geq w_{\min}, \forall l \in [1, p), \forall k \in \mathbb{Z}_+$ (T.2.4.1(iii)); (b) comes from Cauchy-Schwarz inequality. To further bound this quantity, note that

$$\frac{1}{m}V(\mathbf{y}) = \frac{1}{m}\sum_i(y_i^k - \bar{y}^0)^2 \leq \max_{i \in \{i^*, j^*\}}(y_i^k - \bar{y}^0)^2 \leq (y_{i^*}^k - y_{j^*}^k)^2. \tag{2.58}$$

On the other hand, since the consensus algorithm preserves the average, it follows

$$y_{i^*}^k - \bar{y}^0 \leq (m-1)(\bar{y}^0 - y_{j^*}^k), \quad \bar{y}^0 - y_{j^*}^k \leq (m-1)(y_{i^*}^k - \bar{y}^0), \tag{2.59}$$

so that the first inequality in (2.58) is upper bounded as $\frac{1}{m}V(\mathbf{y}) \leq (m-1)^2 \min_{i \in \{i^*, j^*\}}[y_i^k - \bar{y}^0]^2$. Consider the following two cases:

(i) $y_i^k \in [q_{\min}, q_{\max}], \forall i$, so that $\tilde{y}_i^k = y_i^k, \forall i$ and $\tilde{y}_{i^*}^k - \tilde{y}_{j^*}^k = y_{i^*}^k - y_{j^*}^k \geq \frac{1}{\sqrt{m}}\sqrt{V(\mathbf{y})}$.

(ii) $y_{i^*}^k > q_{\max}$ ($y_{j^*}^k < q_{\min}$ can be solved similarly) so that $\tilde{y}_{i^*}^k = q_{\max}$: since $\tilde{y}_{j^*}^k \leq \max\{y_{j^*}^k, q_{\min}\}$, using (2.59) and $\bar{y}^0 \leq q_{\max}$ it follows

$$\tilde{y}_{i^*}^k - \tilde{y}_{j^*}^k \geq \min\left\{\frac{y_{i^*}^k - \bar{y}^0}{m-1}, q_{\max} - q_{\min}\right\} \geq \min\left\{\frac{\sqrt{V(\mathbf{y})}}{\sqrt{m}(m-1)^2}, q_{\max} - q_{\min}\right\}.$$

From (i), (ii) and (2.57), there exists some $\xi_4 > 0$ such that

$$\tilde{\mathbf{y}}^{k\top}\mathbf{L}^k\tilde{\mathbf{y}}^k \geq \frac{w_{\min}}{2(m-1)}(\tilde{y}_{i^*}^k - \tilde{y}_{j^*}^k)^2 \geq \frac{w_{\min}/2}{(m-1)} \min\left\{\frac{V(\mathbf{y}^k)}{m(m-1)^4}, (q_{\max} - q_{\min})^2\right\} \geq \xi_4 V(\mathbf{y}^k),$$

since y_i^k and thus $V(\mathbf{y}^k)$ is bounded (L.9).

Step 3: Let $y^* = \max_i\{\max\{|y_{\max,i}|, |y_{\min,i}|\}\}$. By combining (2.55) and (2.56), we get

$$\begin{aligned}\mathbf{y}^{k\top}\mathbf{L}_+^k\tilde{\mathbf{y}}^k &\geq -\mathbf{y}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \frac{1}{2}\tilde{\mathbf{y}}^{k\top}\mathbf{B}^k\tilde{\mathbf{y}}^k + \xi_1 V(\mathbf{y}^k)\\ &\geq -\sum_{i=1}^m |b_i^k y_i^k \tilde{y}_i^k| - \frac{1}{2}\sum_{i=1}^m |b_i^k(\tilde{y}_i^k)^2| + \xi_1 V(\mathbf{y}^k) \overset{(a)}{\geq} -q^*\Big(y^* + \frac{1}{2}q^*\Big)\|\mathbf{b}^k\|_1 + \xi_1 V(\mathbf{y}^k),\end{aligned}$$

with $\xi_1 = \xi_4/2 > 0$, where (a) comes from the facts $|y_i^k| \leq y^*$, $|\tilde{y}_i^k| \leq q^*$, and $\|\mathbf{b}^k\|_1 = \sum_{i=1}^m |b_i^k|$. □

3. FINITE-BIT QUANTIZATION FOR DISTRIBUTED ALGORITHM WITH LINEAR CONVERGENCE

This chapter studies distributed algorithms for (strongly convex) composite optimization problems over mesh networks, subject to quantized communications. Instead of focusing on a specific algorithmic design, we propose a black-box model casting distributed algorithms in the form of fixed-point iterates, converging at linear rate. The algorithmic model is coupled with a novel (random) Biased Compression (BC-)rule on the quantizer design, which preserves *linear* convergence. A new quantizer coupled with a communication-efficient encoding scheme is also proposed, which efficiently implements the BC-rule using a *finite* number of bits. This contrasts with most of existing quantization rules, whose implementation calls for an *infinite* number of bits. A unified communication complexity analysis is developed for the black-box model, determining the average number of bit required to reach a solution of the optimization problem within the required accuracy. Numerical results validate our theoretical findings and show that distributed algorithms equipped with the proposed quantizer have more favorable communication complexity than algorithms using existing quantization rules.

The novel results of this chapter have been published in

- C.-S. Lee, N. Michelusi and G. Scutari, ”Finite rate quantized distributed optimization with geometric convergence”, in *Proc. 52nd ACSSC*, pp. 1876-1880, Oct. 2018.
- C.-S. Lee, N. Michelusi and G. Scutari, ”Finite-bit quantization for distributed algorithms with linear convergence” *submitted to IEEE Trans. Inf. Theory*, Jul. 2021, Available [online]: https://arxiv.org/abs/2107.11304.

3.1 Introduction

We study distributed optimization over a network of m agents modeled as an undirected (connected) graph. We consider mesh networks, that is, arbitrary topologies with no central hub connected to all the other agents, where each agent can communicate with its immediate neighbors (master/worker architectures will be treated as a special case). The m agents aim at solving cooperatively the optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} \quad \underbrace{\frac{1}{m}\sum_{i=1}^{m} f_i(\mathbf{x})}_{=F(\mathbf{x})} + r(\mathbf{x}), \tag{P}$$

where each f_i is the local cost function of agent i, assumed to be smooth, convex, and known only to the agent; $r : \mathbb{R}^d \to [-\infty, \infty]$ is a nonsmooth, convex (extended-value) function known to all agents, which can be used to force shared constraints or some structure on the solution (e.g., sparsity); and the global loss $F : \mathbb{R}^d \to \mathbb{R}$ is assumed to be strongly convex on the domain of r. This setting is fairly general and finds applications in several areas, including network information processing, telecommunications, multi-agent control, and machine learning (e.g., [48]–[50]).

Since the functions f_i can be accessed only locally and routing local data to other agents is infeasible or highly inefficient, solving (P) calls for the design of distributed algorithms that alternate between a local computation procedure at each agent's side and some rounds of communication among neighboring nodes. While most existing works focus on *ad-hoc* solution methods, here we consider a *general* distributed algorithmic framework, encompassing algorithms whose dynamics are modeled by the fixed-point iteration

$$\mathbf{z}^{k+1} = \tilde{\mathcal{A}}\big(\mathbf{z}^k\big), \tag{3.1}$$

where $\mathbf{z}^k$ is the updating variable at iteration k and $\tilde{\mathcal{A}}$ is a mapping that embeds the local computation and communication steps, whose fixed point typically coincides with the solutions of (P). This model encompasses several distributed algorithms over different network architectures, each one corresponding to a specific expression of $\mathbf{z}$ and $\tilde{\mathcal{A}}$–see Sec. 3.2 for some examples.

By assuming that F is strongly convex, with (3.1) we explicitly target distributed schemes converging to solutions of (P) at *linear rate*. Furthermore, since the cost of communications is often the bottleneck for distributed computing when compared with local (possibly parallel) computations (e.g., [51], [52]), we achieve communication efficiency by embedding the iterates (3.1) with quantized communication protocols. Our goal is to design a black-box quantization mechanism for the class of distributed algorithms (3.1) that preserves their linear convergence while employing *finite-bit* quantized communications.

To our knowledge, this is an open problem, since there exists no *linearly* convergent distributed algorithmic *framework* for the general class of the composite (constrained) optimization problems (P) employing *finite-bit* communications. While we defer to Sec. 3.1.2 for a detailed literature review, here we only point out that existing distributed schemes employing some form of quantization of the communications are applicable only to smooth, unconstrained instances of (P) (i.e., $r = 0$) [53]–[61]. Furthermore, the majority of such algorithms require *infinite*-bit communications [53]–[58]. The exceptions are [59], [60] and our work [61]; yet, the quantization schemes developed in these papers are tailored to a specific distributed algorithm, namely, a primal-dual scheme in [59] and NEXT [8], [62], [63] in [60], [61].

3.1.1 Summary of main contributions

Our major contributions are summarized next–see also Table 3.1.

- **A black-box quantization model for (3.1):** We propose a novel black-box model that introduces quantization in the communication steps of all distributed algorithms cast in the form (3.1). Our approach paves the way to a *unified* design of quantization rules and analysis of their impact on the convergence rate of a gamut of distributed algorithms. This constitutes a major departure from the majority of existing studies focusing on ad-hoc algorithms and quantization rules, which in fact are special instances of our framework. Furthermore, our model brings for the first time quantization to distributed algorithms applicable to composite optimization (P) (i.e., with $r \neq 0$).
- **Preserving linear convergence of (3.1) under quantization:** We provide a novel *biased compression* rule (the BC-rule) on the quantizer design equipping the proposed black-box model, which preserves *linear* convergence of the distributed algorithms while using a *finite* number of bits and *without* altering their original tuning. Our condition encompasses

several deterministic and random quantization rules, new and old [6], [7], [19]–[21], [23]–[28], [34], [44], [59]–[61], [64]. Furthermore, our analysis reveals that, despite common wisdom, several rules proposed in the literature for signal compression and used in particular for quantization [53]–[58], [65]–[74] cannot be implemented using a finite number of bits.

• **A novel finite-bit quantizer:** To make the BC-rule practical, we also propose a novel *finite-bit* quantizer fulfilling the BC-rule along with a communication-efficient bit-encoding/decoding rule which enables transmissions on digital channels; it is termed *Adaptive encoding Nonuniform Quantization* (ANQ). ANQ is a deterministic quantizer that adapts the number of bits of the output (discrete representation) based upon the input signal. By doing so, it achieves a more communication-efficient design than existing quantizers that encode the signal by using a fixed number of bits based on the worst-case range of the input signal (a predetermined fixed range in [6], [7], [19]–[21], [23], [26], [28], [34], [44], [64], or a shrinking one in [24], [25], [27], [59]–[61]).

• **Communication complexity:** We derive the first communication complexity for quantized distributed algorithms over *mesh networks* (see. Table 3.1), in terms of average number of bits required to reach an ε-solution of (P) (using a proper optimality measure–see T.3.5.2) by *any* distributed algorithm belonging to our black-box model. This also sheds light on the dependence of the convergence rate and communication cost on the quantization design parameters.

Finally, we validate numerically our theoretical findings on regularized least square and logistic regression problems. Among others, our evaluations show that 1) linear convergence of all distributed algorithms is preserved under finite-bit quantization based upon the proposed BC-rule; as predicted by our analysis, the rate approaches the one of their unquantized counterpart scheme when a sufficient number of bits is used; 2) the proposed ANQ rule outperforms existing finite-bit quantization rules; and 3) a benchmark of several distributed schemes under quantization is provided, for which convergence guarantees are established for the first time in this work.

3.1.2 Related works

The literature on distributed algorithms is vast; here, we review relevant works employing some form of quantization with linear convergence guarantees [53]–[57], [59]–[61], categorized into those requiring infinite or finite number of bits.

1) Infinite-bit quantization schemes [53]–[57]: Distributed algorithms employing quantization in the agents' communications are proposed in [53]–[57] for special instances of (P) with $r = 0$ (i.e., smooth and unconstrained optimization). In these schemes, quantization is implemented by compressing the signal $\mathbf{x} \in \mathbb{R}^d$ through a (random or deterministic[1]) *compression operator* $\mathbf{x} \mapsto \mathcal{Q}(\mathbf{x})$, that satisfies the *compression rule*

$$\sqrt{\mathbb{E}[\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\|_2^2]} \leq \omega \|\mathbf{x}\|_2, \quad \text{for some} \quad \omega \in (0,1). \tag{3.2}$$

Despite common wisdom, we prove that all the quantization rules derived from (3.2)–hence those in [53]–[57]–can only be implemented using an *infinite* number of bits (see C.2). This calls for the development of new compression rules using a finite number of bits. The proposed BC-rule provides a positive answer to this question.

2) Finite-bit quantization schemes [59]–[61]: While finite-rate quantization has been extensively studied for average consensus schemes (e.g., [7], [19], [24]–[28], [44], [64]), their extension to optimization algorithms over mesh networks is less explored [59]–[61]. Specifically, in our conference work [61], we equip the NEXT algorithm [8], [62] with a finite-bit deterministic quantization to solve (P) with $r = 0$; to preserve linear convergence, the quantizer shrinks its input range linearly. An expression of the convergence rate of the scheme in [61] has been later determined in [60] along with its scaling properties with respect to problem, network, and quantization parameters.

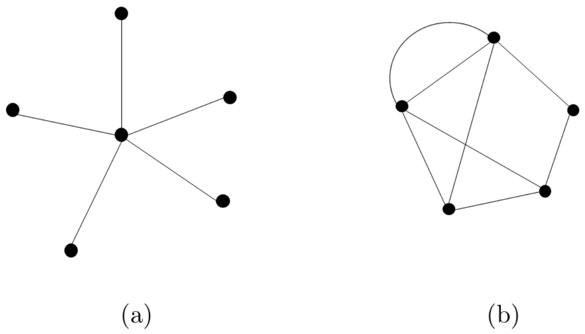
The closest paper to our work is [59], where the authors proposed a finite-bit quantization mechanism preserving linear convergence of a sub-class of algorithms cast as (3.1). Yet, there are several key differences between [59] and our work. First, the convergence analysis in [59] is applicable only to algorithms solving smooth, unconstrained optimization problems, and

[1] ↑We treat compression rules using deterministic mappings $\mathcal{Q}^k$ as special cases of the random ones; in this case, the expected value operator will just return the deterministic value argument.

Table 3.1. Comparison with the state-of-the-art distributed algorithms using some form of quantization; λ is the convergence rate of the input (unquantized) algorithm, and d is the dimension of $\mathbf{x}$. The scheme proposed in this chapter is applicable to the distributed algorithms listed in the table and, in addition, to the following: general primal-dual-based methods [80], EXTRA [76], NEXT [8], [62], AugDGM [81], DIGing [78], the scheme in [79], NIDS [82], Exact Diffusion [83], and some of their proximal counterpart as those in [84] and [80].

Ref.	**Problem**	**# of bits/agent to ε accuracy**	**Algorithms**
[61]	(P) with $r = 0$	N/A	ad-hoc (NEXT [8], [62])
[60]	(P) with $r = 0$	N/A	ad-hoc (NEXT [8], [62])
[59]	(P) with $r = 0$	$\mathcal{O}\Big(\log_2\left(1+\frac{d}{1-\lambda}\right)\frac{d}{1-\lambda}\log_2(d/\varepsilon)\Big)$	GD over star networks
		N/A	ad-hoc (primal-dual [75])
This chapter	(P)	$\mathcal{O}\Big(\log_2\left(1+\frac{1}{1-\lambda}\right)\frac{d}{1-\lambda}\log_2(d/\varepsilon)\Big)$	All the schemes listed in the caption of the table

thus not to Problem (P) with $r \neq 0$. Second, linear convergence under finite-bit quantization is explicitly proved in [59] only for schemes whose updates utilize current iterate information, namely: gradient descent (GD) over star networks and the primal-dual algorithm in [75] over mesh networks. This leaves open the question whether distributed algorithms using historical information–e.g., in the form of gradient tracking or dual variables–are linearly convergent under finite-bit quantization, and under which conditions; renowned examples include EXTRA [76], AugDGM [77], DIGing [78], Harnessing [79], and NEXT [8], [62]. Our work provides a positive answer to these open questions. Third, communication complexity of the scheme in [59] is not provided over mesh networks, which instead is a novel contribution of this work for a wide class of distributed algorithms–see Table 3.1 and Sec. 3.5. Fourth, [59] proposed an ad-hoc deterministic quantization rule while the proposed BC-rule encompasses several deterministic and random quantizations (including that in [59] as a special case), possibly using a variable number of bits (adapted to the input signal). As a result, even when customized to the setting/algorithms in [59], the BC-rule leads to more communication-efficient schemes, both analytically (see Sec. 3.5) and numerically (see Sec. 3.6)–see also Table 3.1.



(a) (b)

Figure 3.1. Examples of star network (a) versus mesh topology (b).

3.1.3 Organization and notation

The remainder of this chapter is organized as follows. Sec. 3.2 introduces the proposed black-box model for casting distributed algorithms in the form (3.1). Sec. 3.3 embeds quantized communications, introduces the proposed BC-rule, and analyzes the convergence properties. Sec. 3.4 describes the proposed quantizer, the ANQ, and studies communication complexity. Sec. 3.5 customizes the proposed framework and convergence guarantees to several existing distributed algorithms, equipping them with the ANQ rule. Sec. 3.6 provides some numerical results, while Sec. 3.7 draws some conclusions. All the proofs of our results are presented in the appendix.

Notation: Throughout the chapter, we model a network of m agents as a fixed, undirected, connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = [m]$ is the set of vertices (agents) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (communication links); $(i, j) \in \mathcal{E}$ if there is a link between agents i and j, so that the two can send information to each other. We let $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}\}$ be the set of neighbors of agent i, and assume that $(i, i) \in \mathcal{E}$, i.e., $i \in \mathcal{N}_i$. Master/workers architectures will be considered as special cases–see Fig. 3.1.

3.2 A General Distributed Algorithmic Framework: Exact Communications

In this section, we show how to cast distributed algorithms for (P) in the form (3.1). As a warm-up, we begin with schemes using only current information to produce the next update (cf. Sec. 3.2.1). We then generalize the model to capture distributed algorithms using historical information via multiple rounds of communications between computation steps (cf. Sec. 3.2.2).

3.2.1 Warm-up: A class of distributed algorithms

We cast distributed algorithms in the form (3.1) by incorporating computations and communications as two separate steps. We use state variable $\mathbf{z}_i$ to capture local information owned by agent i (including optimization variables) and $\hat{\mathbf{c}}_i$ to denote the signal transmitted by agent i to its neighbors.[2] Similarly to [59], the updates of the $z, \hat{c}$-variables read: for agent $i \in [m]$,

$$\begin{aligned} \hat{\mathbf{c}}_i^k &= \mathcal{C}_i\Big(\mathbf{z}_i^k\Big), && \text{(communication step)} \\ \mathbf{z}_i^{k+1} &= \mathcal{A}_i\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^k\Big), && \text{(computation step)} \end{aligned} \tag{M0}$$

where the function $\mathbf{z}_i \mapsto \mathcal{C}_i(\mathbf{z}_i)$ models the processing on the local information $\mathbf{z}_i^k$ at the current iterate, generating the signal $\hat{\mathbf{c}}_i^k$ to be transmitted to agent i's neighbors; and the function $(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i}) \mapsto \mathcal{A}_i(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i})$ is the function producing the update of the agent i's state variable $\mathbf{z}_i$, based upon the local information at iteration k (including the signals received by its neighbors).

Some examples: The algorithmic model (M0) captures a variety of distributed algorithms that build updates using single rounds of communications; examples include the renewed DGD [85], NIDS [82], and the primal-dual scheme [75]. To show a concrete example, consider DGD, which aims at solving a special instance of (P) with $r = 0$; agents' updates read

$$\mathbf{x}_i^{k+1} = \Bigg(\sum_{j=1}^{m} w_{ij}\mathbf{x}_j^k\Bigg) - \gamma \nabla f_i\Big(\mathbf{x}_i^k\Big), \quad i \in [m],$$

where $\mathbf{x}_i^k$ is the local copy owned by agent i at iteration k of the optimization variables $\mathbf{x}$, γ is a step-size, and w_{ij}'s are nonnegative weights properly chosen and compliant with the

[2] ↑Dimensions of these vectors are algorithm-dependent and omitted for simplicity, and will be clear from the context.

graph $\mathcal{G}$ (i.e., $w_{ij} > 0$ if $(i,j) \in \mathcal{E}$; and $w_{ij} = 0$ otherwise). It is not difficult to check that DGD can be rewritten in the form (M0) by letting

$$\mathbf{z}_i^k = \hat{\mathbf{c}}_i^k = \mathbf{x}_i^k \quad \text{and} \quad \mathcal{A}_i(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^k) = \left(\sum_{j=1}^m w_{ij} \hat{\mathbf{c}}_j^k\right) - \gamma \nabla f_i(\mathbf{z}_i^k).$$

Despite its generality, model (M0) leaves out several important distributed algorithms, specifically, the majority of schemes employing correction of the gradient direction based on past state information–these are the best performing algorithms to date. Examples include EXTRA [76], DIGing [78] and their proximal version, NEXT/SONATA [8], [62], [63], and the ABC framework [80], just to name a few. Consider for instance NEXT/SONATA:

$$\mathbf{x}_i^{k+1} = \sum_{j \in \mathcal{N}_i} w_{ij}\Big(\mathbf{x}_j^k - \gamma \mathbf{y}_j^k\Big) \quad \text{and} \quad \mathbf{y}_i^{k+1} = \sum_{j \in \mathcal{N}_i} w_{ij}\Big(\mathbf{y}_j^k + \nabla f_j(\mathbf{x}_j^{k+1}) - \nabla f_j(\mathbf{x}_j^k)\Big). \tag{3.3}$$

Clearly, this does not fit model (M0): the update of the y-variable uses information from two iteration ages (k and $k+1$). This calls for a more general model, introduced next.

3.2.2 Proposed general model (using historical information)

We generalize the algorithmic model (M0) as follows: for all $i \in [m]$,

$$\left.\begin{aligned} \hat{\mathbf{c}}_i^{k,1} &= \mathcal{C}_i^1\Big(\mathbf{z}_i^k, \mathbf{0}_{\mathcal{N}_i}\Big), \\ &\;\vdots \\ \hat{\mathbf{c}}_i^{k,R} &= \mathcal{C}_i^R\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,R-1}\Big), \end{aligned}\right\} \text{(multiple communication rounds)}$$
$$\mathbf{z}_i^{k+1} = \mathcal{A}_i\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \cdots, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,R}\Big), \quad \text{(computation step)} \tag{M}$$

which embeds $R \geq 1$ rounds of local communications, via the functions $(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i}^{s-1}) \mapsto \mathcal{C}_i^s(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i}^{s-1})$; and the function $(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i}^1, \cdots, \hat{\mathbf{c}}_{\mathcal{N}_i}^R) \mapsto \mathcal{A}_i(\mathbf{z}_i, \hat{\mathbf{c}}_{\mathcal{N}_i}^1, \cdots, \hat{\mathbf{c}}_{\mathcal{N}_i}^R)$ updates the local state by possibly using the signals $\hat{\mathbf{c}}_{\mathcal{N}_i}$'s received from all neighbors during all R rounds of communications, along with $\mathbf{z}_i$. Stacking agents' state-variables $\mathbf{z}_i$, communication signals $\hat{\mathbf{c}}_i$, and mappings $\mathcal{C}_i^s$ and $\mathcal{A}_i$ into the respective vectors $\mathbf{z}$, $\hat{\mathbf{c}}$, $\mathcal{C}^s$ and $\mathcal{A}$, we can rewrite (M) in compact form as

$$\left.\begin{aligned} \hat{\mathbf{c}}^{k,0} &= \mathbf{0}, \\ \hat{\mathbf{c}}^{k,s} &= \mathcal{C}^s\Big(\mathbf{z}^k, \hat{\mathbf{c}}^{k,s-1}\Big), \quad s \in [R], \end{aligned}\right\} \text{(multiple communication rounds)}$$
$$\mathbf{z}^{k+1} = \mathcal{A}\Big(\mathbf{z}^k, \hat{\mathbf{c}}^{k,1}, \cdots, \hat{\mathbf{c}}^{k,R}\Big). \qquad \text{(computation step)}$$

Absorbing the communication signals $\hat{\mathbf{c}}^{k,s}$ in the mapping $\mathcal{A}$, we can finally write the above system as a fixed-point iterate on the z-variables only:

$$\mathbf{z}^{k+1} = \tilde{\mathcal{A}}(\mathbf{z}^k) \triangleq \mathcal{A}\Big(\mathbf{z}^k, \mathcal{C}^1\Big(\mathbf{z}^k, \mathbf{0}\Big), \cdots, \mathcal{C}^R\Big(\mathbf{z}^k, \mathcal{C}^{R-1}\Big(\mathbf{z}^k, \cdots \mathcal{C}^1\Big(\mathbf{z}^k, \mathbf{0}\Big)\cdots\Big)\Big)\Big). \tag{M'}$$

Under suitable conditions, the iterates (M') convergence to fixed-points $\mathbf{z}^\infty = \tilde{\mathcal{A}}(\mathbf{z}^\infty)$ of the mapping $\tilde{\mathcal{A}}$, possibly constrained to a set $\mathcal{Z} \ni \mathbf{z}^\infty$. The convergence rate depends on the properties of $\tilde{\mathcal{A}}$; here we focus on *linear* convergence, which can be established under the following standard condition.

Assumption 6. *Let $\tilde{\mathcal{A}} : \mathcal{Z} \to \mathcal{Z}$; the following hold: (i) $\tilde{\mathcal{A}}$ admits a fixed-point $\mathbf{z}^\infty$; and (ii) $\tilde{\mathcal{A}}$ is λ-pseudo-contractive on $\mathcal{Z}$ w.r.t. some norm $\| \bullet \|$, that is, there exists $\lambda \in (0,1)$ such that*

$$\|\tilde{\mathcal{A}}(\mathbf{z}) - \mathbf{z}^\infty\| \leq \lambda \cdot \|\mathbf{z} - \mathbf{z}^\infty\|, \quad \forall \mathbf{z} \in \mathcal{Z}.$$

Without loss of generality, the norm $\| \bullet \|$ is scaled such that $\| \bullet \|_2 \leq \| \bullet \|$.[3]

The following convergence result follows readily from 6 and [43, Ch. 3, P.1.2].

Theorem 3.2.1. *Let $\tilde{\mathcal{A}} : \mathcal{Z} \to \mathcal{Z}$ satisfy A.6. Then: i) the fixed point $\mathbf{z}^\infty$ is unique; and ii) the sequence $\{\mathbf{z}^k\}$ generated by the update (M') converges Q-linearly to $\mathbf{z}^\infty$ w.r.t. the norm $\| \bullet \|$ at rate λ, i.e., $\left\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\right\| \leq \lambda \cdot \left\|\mathbf{z}^k - \mathbf{z}^\infty\right\|$.*

Discussion: The algorithmic framework (M) encompasses a variety of distributed algorithms, while T.3.2.1 captures their convergence properties; in addition to the schemes

[3] ↑ This is always possible since $\| \bullet \|$ is a norm defined on a finite-dimensional field.

covered by (M0), (M) can also represent EXTRA [76] and its proximal version [80], NEXT [8], [62], [63], DIGing [78], NIDS [82], and primal-dual schemes such as [75]. App.3.8.4 provides specific expressions for the mappings $\mathcal{A}$ and $\mathcal{C}^s$ for each of the above algorithms, along with their convergence properties under T.3.2.1; here, we elaborate on the NEXT algorithm (3.3) as an example. It can be rewritten in the form (M) by using $R = 2$ rounds of communications and letting

$$\mathbf{z}_i^k = \begin{bmatrix} \mathbf{x}_i^k \\ \mathbf{y}_i^k \end{bmatrix}, \quad \hat{\mathbf{c}}_i^{k,1} = \mathbf{x}_i^k - \gamma \mathbf{y}_i^k, \quad \hat{\mathbf{c}}_i^{k,2} = \mathbf{y}_i^k + \nabla f_i\Bigg(\sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1}\Bigg) - \nabla f_i(\mathbf{x}_i^k), \text{ and}$$

$$\mathcal{A}_i(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}) = \begin{bmatrix} \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1} \\ \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,2} \end{bmatrix}.$$

3.3 A General Distributed Algorithmic Framework: Quantized Communications

In this section, we equip the distributed algorithmic framework (M) with quantized communications. The communication channel between any two agents is modeled as a noiseless digital channel; only quantized signals are received with no errors. This means that, in each of the communication rounds, the signals $\hat{\mathbf{c}}_j^{k,1}, \ldots, \hat{\mathbf{c}}_j^{k,R}$, $j \in \mathcal{N}_i$, received by agent i no longer coincide with the intended, unquantized ones $\mathcal{C}_j^1(\mathbf{z}_j^k, \mathbf{0}_{\mathcal{N}_j}), \ldots, \mathcal{C}_j^R(\mathbf{z}_j^k, \hat{\mathbf{c}}_{\mathcal{N}_j}^{k,R-1})$, generated at the transmitter side of agents $j \in \mathcal{N}_i$. This calls for a proper encoding/decoding mechanism that transfers, via quantized communications, the aforementioned unquantized signals at the receiver sides with limited distortion. Here, we leverage differential encoding/decoding techniques [24] coupled with a novel finite-level quantization mechanism.

We begin recalling the idea of quantized differential encoding/decoding in the context of a point-to-point communication–the same mechanism will be then embedded in the communication of the distributed multi-agent framework (M). Consider a transmitter-receiver pair; let $\mathbf{c}^k$ be the unquantized information generated at iteration k, intended to be transferred to the receiver over the digital channel, and let $\hat{\mathbf{c}}^k$ be the estimate of $\mathbf{c}^k$, built using quantized information. The differential encoding/decoding rule reads: $\hat{\mathbf{c}}^0 = \mathbf{0}$, and for $k = 1, \ldots,$

$$\begin{cases} \mathbf{q}^k & = \mathcal{Q}^k(\mathbf{c}^k - \hat{\mathbf{c}}^{k-1}), \\ \hat{\mathbf{c}}^k & = \hat{\mathbf{c}}^{k-1} + \mathbf{q}^k, \end{cases} \tag{3.4}$$

where $\mathcal{Q}^k$ is the quantization operator (a map from real numbers to the set of quantized points), possibly dependent on iteration k. In words, the encoder quantizes at each iteration the "prediction" error $\mathbf{c}^k - \hat{\mathbf{c}}^{k-1}$ rather than the current estimate $\mathbf{c}^k$, generating the quantized signal $\mathbf{q}^k$, which is then transmitted over the digital channel. The estimate $\hat{\mathbf{c}}^k$ of $\mathbf{c}^k$ is built from $\mathbf{q}^k$ using a one-step prediction rule. The rationale of this decoding rule is that, for negligible quantization errors $\mathbf{q}^k = \mathcal{Q}^k(\mathbf{c}^k - \hat{\mathbf{c}}^{k-1}) \approx \mathbf{c}^k - \hat{\mathbf{c}}^{k-1}$, the estimate reads $\hat{\mathbf{c}}^k = \hat{\mathbf{c}}^{k-1} + \mathbf{q}^k \approx \hat{\mathbf{c}}^{k-1} + \mathbf{c}^k - \hat{\mathbf{c}}^{k-1} = \mathbf{c}^k$. Note that, since $\mathbf{q}^k$ is received unaltered, $\hat{\mathbf{c}}^k$ is identical at the transmitter's and receiver's sides.

We can now introduce our distributed algorithmic framework using quantized communications, as described in Algorithm 3; it embeds the differential enconding/decoding rule (3.4) in each communication round of model (M). The fixed-point based formulation of Algorithm 3 then reads: for $i \in [m]$,

$$\begin{aligned} &\left.\begin{aligned} \mathbf{c}_i^{k,1} &= \mathcal{C}_i^1\Big(\mathbf{z}_i^k, \mathbf{0}_{\mathcal{N}_i}\Big), \\ \hat{\mathbf{c}}_i^{k,1} &= \hat{\mathbf{c}}_i^{k-1,1} + \mathcal{Q}_i^k\Big(\mathbf{c}_i^{k,1} - \hat{\mathbf{c}}_i^{k-1,1}\Big), \\ &\;\vdots \\ \mathbf{c}_i^{k,R} &= \mathcal{C}_i^R\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,R-1}\Big), \\ \hat{\mathbf{c}}_i^{k,R} &= \hat{\mathbf{c}}_i^{k-1,R} + \mathcal{Q}_i^k\Big(\mathbf{c}_i^{k,R} - \hat{\mathbf{c}}_i^{k-1,R}\Big), \end{aligned}\right\} \quad \text{(multiple communication rounds)} \\ &\mathbf{z}_i^{k+1} = \mathcal{A}_i\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \cdots, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,R}\Big), \qquad\qquad \text{(computation step).} \end{aligned} \tag{3.5}$$

Algorithm 3 Distributed Algorithmic Framework with Quantized Communications

```
Require: $\hat{\mathbf{c}}^{-1,s} \triangleq \mathbf{0}$, for all $s \in [R]$; and $\mathbf{z}^0 \in \mathcal{Z}$. Set $k = 0$;
  Iteration $k \to k+1$
  (S.1): Multiple communication rounds
    for $s = 1, \ldots, R$, each agent $i$:
      • Computes $\mathbf{c}_i^{k,s} = \mathcal{C}_i^s(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,s-1})$ [with $\hat{\mathbf{c}}_i^{k,0} \triangleq \mathbf{0}$];
      • Generates $\mathbf{q}_i^{k,s} = \mathcal{Q}_i^k(\mathbf{c}_i^{k,s} - \hat{\mathbf{c}}_i^{k-1,s})$ and broadcasts it to its neighbors $j \in \mathcal{N}_i$;
      • Upon receiving the signals $\mathbf{q}_j^{k,s}$ from its neighbors $j \in \mathcal{N}_i$, it reconstructs $\hat{\mathbf{c}}_j^{k,s}$ as
            $$\hat{\mathbf{c}}_j^{k,s} = \hat{\mathbf{c}}_j^{k-1,s} + \mathbf{q}_j^{k,s}, \quad j \in \mathcal{N}_i;$$
    end
  (S.2): Computation Step
    Each agent $i$ updates its own $\mathbf{z}_i^{k+1}$ according to
            $$\mathbf{z}_i^{k+1} = \mathcal{A}_i(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \cdots, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,R}).$$
```

Stacking agents' state-variables $\mathbf{z}_i$, signals $\mathbf{c}_i$ and $\hat{\mathbf{c}}_i$, and mappings $\mathcal{C}_i^s$, $\mathcal{A}_i$, and $\mathcal{Q}_i^k$ into the respective vectors $\mathbf{z}$, $\mathbf{c}$, $\hat{\mathbf{c}}$, $\mathcal{C}^s$, $\mathcal{A}$, and $\mathcal{Q}^k$, we can rewrite (3.5) in compact form as

$$\left.\begin{aligned} \hat{\mathbf{c}}^{k,0} &= \mathbf{0}, \\ \mathbf{c}^{k,s} &= \mathcal{C}^s\Big(\mathbf{z}^k, \hat{\mathbf{c}}^{k,s-1}\Big), \\ \hat{\mathbf{c}}^{k,s} &= \hat{\mathbf{c}}^{k-1,s} + \mathcal{Q}^k\Big(\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\Big), \quad s \in [R], \end{aligned}\right\} \quad \text{(multiple communication rounds)}$$
$$\mathbf{z}^{k+1} = \mathcal{A}\Big(\mathbf{z}^k, \hat{\mathbf{c}}^{k,1}, \cdots, \hat{\mathbf{c}}^{k,R}\Big). \qquad \text{(computation step)}. \tag{Q-M}$$

Model (Q-M) paves the way to a unified design and convergence analysis of several distributed algorithms–all the schemes cast in the form (M)–employing quantization in the communications, as elaborated next.

3.3.1 Convergence Analysis

We begin by establishing sufficient conditions on the quantized mapping $\mathcal{Q}^k$ and algorithmic functions $\mathcal{A}$ and $\mathcal{C}$ in (Q-M) to preserve linear convergence

• **On the quantization mapping $\mathcal{Q}^k$.** A first critical choice is the quantizer $\mathcal{Q}$ (we omit the dependence on k for notation simplicity), including both random and deterministic quantization rules (the latter as special cases of the former). For random quantization, the function $\mathcal{Q}_i(\mathbf{x})$, $i \in [m]$, is a random variable for any given $\mathbf{x} \in \mathbb{R}^d$, defined on a suitable probability space (generally dependent on $\mathbf{x}$). We propose the following novel *biased compression rule* (BC-rule), for each agent i.

Definition 3.3.1 (Biased compression rule). *Given $\mathbf{x} \in \mathbb{R}^d$, $\mathcal{Q}(\mathbf{x})$ (possibly, a random variable defined on a suitable probability space) satisfies the BC-rule with* bias $\eta \geq 0$ *and* compression rate $\omega \in [0, 1)$ *if*

$$\sqrt{\mathbb{E}\left[\left\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\right\|_2^2\right]} \leq \sqrt{d}\,\eta + \omega\|\mathbf{x}\|_2, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{3.6}$$

When $\mathcal{Q}(\mathbf{x})$ is a deterministic map, (3.6) reduces to

$$\left\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\right\|_2 \leq \sqrt{d}\,\eta + \omega\|\mathbf{x}\|_2, \quad \forall \mathbf{x} \in \mathbb{R}^d. \tag{3.7}$$

Roughly speaking, the bias η determines the basic spacing between quantization points, uniform across the entire domain. On the other hand, the compression term ω adds a nonuniform spacing between quantization points: quantization points farther away from $\mathbf{0}$ will have more separation.

The BC-rule encompasses and generalizes several existing compression and quantization rules proposed in the literature for specific algorithms, deterministic [6], [19]–[21], [24], [25], [27], [57], [59]–[61], [66]–[69], [71], [73] and random [7], [23], [26], [28], [34], [44], [53]–[57], [64], [65], [68]–[70], [72]–[74] ones. Specifically, **(i)** the compression rules proposed in [53]–[57], [65], [68]–[70], [72]–[74] (resp. [57], [66]–[69], [71], [73]) can be interpreted as *unbiased* instances of (3.6) [resp. (3.7)], i.e., corresponding to $\eta = 0$. The proof of L.11 in Sec. 3.4 will show that such special instances can only be implemented using *infinite* quantization points (hence number of bits). **(ii)** On the other hand, the quantization rules in [7], [23], [26], [28], [34], [44], [64] (resp. [6], [19]–[21], [24], [25], [27], [59]–[61]) are special cases of the BC-rule (3.6) [resp. (3.7)], with $\omega = 0$. While they can be implemented using a finite number of bits, they do not take advantage of the degree of freedom offered by the compression rate ω, a fact that will be numerically shown to lead to more communication-efficient schemes.

• **On the algorithmic mappings $\mathcal{A}$ and $\mathcal{C}^s$.** Our analysis will require some standard conditions on the mappings $\mathcal{A}$ and $\mathcal{C}^s$ (in addition to A.6) to preserve linear convergence under quantization. Roughly speaking, the functions $\mathcal{A}$ and $\mathcal{C}^s$ should vary smoothly with respect to perturbations in their arguments, so that small quantization errors result in small deviations from the trajectory of the unquantized algorithm. Specifically, we postulate the following.[4]

Assumption 7. *There exists a constant $L_A \geq 0$ such that, for every $s \in [R]$, it holds*

$$\left\|\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \cdots, \mathbf{c}^{s-1}, \mathbf{c}^s, \mathbf{c}^{s+1}, \cdots, \mathbf{c}^R) - \mathcal{A}(\mathbf{z}, \mathbf{c}^1, \cdots, \mathbf{c}^{s-1}, \mathbf{c}^{s\prime}, \mathbf{c}^{s+1}, \cdots, \mathbf{c}^R)\right\| \leq L_A \|\mathbf{c}^s - \mathbf{c}^{s\prime}\|_2, \tag{3.8}$$

for all $\mathbf{c}^s, \mathbf{c}^{s\prime} \in \mathbb{R}^{md}$, uniformly with respect to $\mathbf{z} \in \mathcal{Z}$, and $\mathbf{c}^1, \ldots, \mathbf{c}^{s-1}, \mathbf{c}^{s+1}, \ldots \mathbf{c}^R \in \mathbb{R}^{md}$.

Assumption 8. *There exist constants $L_C, L_Z \geq 0$ such that*

$$\|\mathcal{C}^s(\mathbf{z}, \mathbf{c}) - \mathcal{C}^s(\mathbf{z}, \mathbf{c}')\|_2 \leq L_C \|\mathbf{c} - \mathbf{c}'\|_2, \quad \forall \mathbf{c}, \mathbf{c}' \in \mathbb{R}^{md}, \tag{3.9}$$

$$\|\mathcal{C}^s(\mathbf{z}, \mathbf{c}) - \mathcal{C}^s(\mathbf{z}', \mathbf{c})\|_2 \leq L_Z \|\mathbf{z} - \mathbf{z}'\|_2, \quad \forall \mathbf{z}, \mathbf{z}' \in \mathcal{Z}, \tag{3.10}$$

uniformly with respect to $\mathbf{z} \in \mathcal{Z}$ and $\mathbf{c} \in \mathbb{R}^{md}$, respectively.

These assumptions are quite mild, and satisfied by a variety of existing distributed algorithms, as we will show in App.3.8.4. We are now ready to introduce our main convergence result.

[4] ↑For the sake of notation, the constants L_A, L_C and L_Z defined in A.7 and A.8 are assumed to be independent of the index s (communication round). Our convergence results can be readily extended to constants dependent on s.

Theorem 3.3.1. *Let $\{\mathbf{z}^k\}_{k\in\mathbb{Z}_+}$ be the sequence generated by Algorithm 3 under A.6-8, with $\mathcal{Q}^k$ satisfying the BC-rule (3.6) with bias $\eta = \eta^0 \cdot (\sigma)^k$ and compression rate $\omega \in [0, \bar{\omega}(\sigma))$, for some $\sigma \in (\lambda, 1)$ and $\eta^0 > 0$, where $\bar{\omega}(\sigma)$ is defined as*

$$\bar{\omega}(\sigma) \triangleq \frac{\sigma}{R} \cdot \frac{\sigma - \lambda}{\sigma - \lambda + 2L_A L_Z [R \max\{1, (2L_C)^{R-1}\}]^2}. \tag{3.11}$$

Then,

$$\sqrt{\mathbb{E}[\|\mathbf{z}^k - \mathbf{z}^\infty\|_2^2]} \leq V_0 \cdot (\sigma)^k, \quad k \in \mathbb{Z}_+,$$

where V_0 is a positive constant, whose expression is given in (3.35), App.3.8.1.

Proof. See App.3.8.1. □

Note that, when the deterministic instance of the BC-rule is used [see (3.7)], the convergence rate reads $\|\mathbf{z}^k - \mathbf{z}^\infty\|_2 \leq V_0 \cdot (\sigma)^k$, for all $k \in \mathbb{Z}_+$.

T.3.3.1 shows that linear convergence is achievable when quantized communications are performed in distributed optimization, provided that the bias η and compression rate ω of the BC-rule are chosen suitably. Specifically, the bias η should shrink linearly (at rate σ) and the compression rate ω should be sufficiently small, so that the quantization errors along the iterates will not accumulate disruptively. The final linear convergence rate is determined by σ, which is larger than rate λ achievable by the same scheme that does not use any quantization. As expected, there is a tension between the amount of quantization/compression of the transmitted signals (measured by η and ω) and the resulting linear convergence rate σ: the slower the decay of η (resulting in less stringent quantization requirements), the slower the convergence rate of the algorithm (the larger σ). However, as we will see in T.3.5.1, the price to pay to achieve faster convergence is communication cost.

T.3.3.1 certifies linear convergence in terms of number of iterations; building on this result, in the forthcoming sections we study the communication complexity of the schemes (Q-M)–the total number of bits needed to reach an ε-solution of problem (P). This depends on the specific quantizer used in the algorithms. The next section introduces a novel quantizer that satisfies the BC-rule whi using the minimum number of quantization points, and a communication-efficient bit-encoding/decoding scheme. When embedded in (Q-M), the proposed quantization leads to linearly convergent distributed algorithms whose communication complexity compares favorably with that of existing ad-hoc schemes (Sec. 3.5).

3.4 Non-Uniform Quantizer with Adaptive Encoding/Decoding

As discussed in Sec. 3.3, the BC-rule encompasses a variety of quantizer designs. In this section, we propose a scalar quantizer that fulfills the BC-rule with minimum number of quantization points (Sec. 3.4.1). The quantizer is then coupled with a communication-efficient bit-encoding/decoding rule which enables transmission on the digital channel (Sec. 3.4.2). We refer to the proposed quantizer coupled with the encoding/decoding scheme as *Adaptive encoding Non-uniform* Quantization (ANQ).

3.4.1 Quantizer design

Since no information is assumed on the distribution of the input signal, a natural approach is to quantize each vector signal component-wise. We design such a scalar quantizer $\mathcal{Q} : [-\delta, \delta] \to \mathbb{Q}$ under the BC-rule by minimizing the number of quantization points $|\mathbb{Q}|$ for a fixed input dynamic δ. Equivalently, we seek $\mathcal{Q}$ that maximizes δ, for a given number $N = |\mathbb{Q}|$ of quantization points while satisfying the BC-rule. The optimal scalar deterministic and probabilistic quantizer designs satisfying the BC-rule are provided in L.11 and L.12, respectively. For convenience, we focus on the case of N odd; the case of N even is provided in App.3.8.2.

Lemma 11 (Deterministic Quantizer)**.** *Let $\mathcal{Q} : [-\delta, \delta] \to \mathbb{Q}$. The maximum range δ that can be quantized using $|\mathbb{Q}| = N$ points while fulfilling the BC-rule (3.7) with bias $\eta \geq 0$ and compression rate $\omega \in [0, 1)$ is*

$$\delta(\eta, \omega, N) = \frac{q_{(N-1)/2} + q_{(N+1)/2}}{2}, \tag{3.12}$$

with quantization points

$$q_\ell = -q_{-\ell} = \frac{\eta}{\omega}\left[\left(\frac{1+\omega}{1-\omega}\right)^{\ell} - 1\right], \quad \ell \geq 0. \tag{3.13}$$

The resulting optimal quantization rule reads: $x \mapsto \mathcal{Q}(x) = q_{\ell(x)}$, *with*

$$\ell(x) = \mathrm{sgn}(x) \cdot \left\lceil \frac{\ln(1-\omega) + \ln(1+\frac{\omega}{\eta}|x|)}{\ln(1+\omega) - \ln(1-\omega)} \right\rfloor. \tag{3.14}$$

Proof. See App.3.8.2. □

From L.11, one infers that the optimal quantizer has quantization points non-uniformly spaced–hence the name ANQ–and maps inputs x to the nearest q_ℓ.

We next study the optimal probabilistic quantizer design under the BC-rule (3.6).

Lemma 12 (Probabilistic Quantizer)**.** *For any given* $x \in [-\delta, \delta]$, *let* $\mathcal{Q}(x) \in \mathbb{Q}$ *be a random variable defined on a suitable probability space. The maximum range* δ *that can be quantized using* $|\mathbb{Q}| = N$ *points while fulfilling the BC-rule* (3.6) *with* $\mathbb{E}[\mathcal{Q}(x)] = x$ *with bias* $\eta \geq 0$ *and compression rate* $\omega \in [0, 1)$ *is*

$$\delta(\eta, \omega, N) = q_{(N-1)/2}, \tag{3.15}$$

with quantization points

$$q_\ell = -q_{-\ell} = \frac{\eta}{\omega}\left[\left(\sqrt{1+(\omega)^2} + \omega\right)^{2\ell} - 1\right], \quad \ell \geq 0. \tag{3.16}$$

The resulting optimal quantization rule reads: $x \mapsto \mathcal{Q}(x) = q_{\ell(x)}$, *with*

$$\ell(x) = \begin{cases} \ell - 1, & w.p.\ \frac{q_\ell - x}{q_\ell - q_{\ell-1}}; \\ \ell, & w.p.\ \frac{x - q_{\ell-1}}{q_\ell - q_{\ell-1}}, \end{cases} \quad \text{and} \quad \ell = \mathrm{sgn}(x)\left\lceil \frac{\ln(1+\frac{\omega}{\eta}|x|)}{2\ln\left(\sqrt{1+(\omega)^2} + \omega\right)} \right\rceil. \tag{3.17}$$

Proof. See App.3.8.2. □

The probabilistic quantizer above has quantization points non-uniformly spaced, and maps x to one of the two nearest quantization points, selected randomly such that $\mathbb{E}[\mathcal{Q}(x)] = x$.

Note that for the proposed deterministic and probabilistic quantizers, the index $\ell(x)$ is sufficient information to infer the quantization point $q_{\ell(x)}$. In Sec. 3.4.2 we present a communication-efficient finite bit-encoding/decoding scheme to transmit $\ell(x)$ over the digital channel.

From L.11 and L.12, it is clear that the optimal quantizer uses a finite number of quantization points (and thus of bits) when $\eta > 0$. This contrasts with the compression rule (3.2), which cannot be implemented using a *finite* number of quantization points (in fact, in this case $\delta(0, \omega, N) = 0$ for any finite N). The next corollary formalizes this negative result.

Corollary 2 (Converse)**.** *No quantizer using a finite number of quantization points can satisfy the compression rule* (3.2)*. Therefore, the compression rules in [53]–[57], [65]–[71], [74] cannot be implemented using a finite number of bits.*

Proof. See App.3.8.2. □

3.4.2 Adaptive encoding scheme

It remains to design an encoding/decoding scheme mapping the index $\ell(x)$ into a finite-bit representation, to be transmitted over the digital channel. To do so, we adopt an adaptive number of bits, based upon the value of $\ell(x)$, as detailed next. We assume that a constellation $\mathbb{S} = [S] \cup \{0\}$ of $S+1$ symbols is used, with $S \geq 2$ (this might be obtained as $\mathbb{S} \equiv [\tilde{\mathbb{S}}]^w$, by concatenating sequences of w symbols from a smaller constellation $\tilde{\mathbb{S}}$). We use the symbol 0 to indicate the end of an information sequence, and the remaining S symbols $[S]$ to encode the value of $\ell(x)$. Defining $\tilde{\mathcal{L}}_{-1} \equiv \emptyset$, let

$$\tilde{\mathcal{L}}_b \equiv \left\{ -\left\lceil \frac{(S)^{b+1}-1}{2(S-1)} \right\rceil + 1, \dots, \left\lfloor \frac{(S)^{b+1}-1}{2(S-1)} \right\rfloor \right\}, \quad b \in \mathbb{Z}_+,$$

and

$$\mathcal{L}_b = \tilde{\mathcal{L}}_b \setminus \tilde{\mathcal{L}}_{b-1}, \quad b = 0, 1, \dots. \tag{3.18}$$

It is not difficult to check that $\{\mathcal{L}_b : b = 0, 1, \dots\}$ creates a partition of $\mathbb{Z}$ and $|\mathcal{L}_b| = (S)^b$. Therefore, a natural way to encode $\ell(x)$ is to use a unique sequence of b symbols from $[S]$, i.e., $[s_1, \dots, s_b] \in [S]^b$, where b is the unique integer such that $\ell(x) \in \mathcal{L}_b$. The transmitted sequence coding $\ell(x)$ reads then $[s_1, \ldots, s_b, 0]$, where 0 marks the end of the information sequence.

Upon receiving this sequence, the receiver can detect the start and end of the information symbols, and decode the associated $\ell(x)$ by inverting the symbol-mapping. The communication cost to transmit the index $\ell(x) \in \mathcal{L}_b$ is thus $b+1$ (symbols), which leads to the following upper bound on the overall communication cost incurred by each agent i to quantize and encode a d-dimensional vector $\mathbf{x}$. Again, we focus on the case when N is odd; the other case is provided in the proof in App.3.8.3.

Lemma 13. *The number of bits $C(\mathbf{x})$ required by the ANQ with bias $\eta \geq 0$ and compression rate $\omega \geq 0$ and constellation of $S+1$ symbols to quantize and encode an input signal $\mathbf{x} \in \mathbb{R}^d$ is upper bounded by*

(i) Deterministic quantizer:

$$C(\mathbf{x}) \leq \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{\ln(1-\omega) + \ln\left(1 + \frac{\omega\|\mathbf{x}\|_2}{\sqrt{d}\eta}\right)}{\ln(1+\omega) - \ln(1-\omega)}\right)\right] \quad \text{bits;} \tag{3.19}$$

(ii) Probabilistic quantizer with $\mathbb{E}[\mathcal{Q}(\mathbf{x})] = \mathbf{x}$:

$$C(\mathbf{x}) \leq \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{\ln\left(1 + \frac{\omega\|\mathbf{x}\|_2}{\sqrt{d}\eta}\right)}{2\ln\left(\sqrt{1+(\omega)^2} + \omega\right)}\right)\right] \quad \text{bits,} \quad a.s.. \tag{3.20}$$

Proof. See App.3.8.3. □

Compared with existing deterministic quantizers [59], [60] that are special cases of the BC-rule (with $\omega = 0$), the proposed ANQ adapts the number of bits to the input signal–less bits for smaller input signals (mapped to smaller ℓ) and more bits for larger ones (mapped to larger ℓ)–rather than using a fixed number of bits determined by the worst-case input signal [59], [60]. This leads to more communication-efficient schemes, as certified by T.3.5.2.

Comparing the communication cost for the probabilistic and deterministic quantizers, it can be shown that the former incurs in a larger cost than the latter. This is due to the fact that in the probabilistic ANQ, the need to enforce the constraint $\mathbb{E}[\mathcal{Q}(x)] = x$ makes the solution less optimal than the one for deterministic ANQ.

3.5 Communication Complexity of (Q-M) using the ANQ Rule

We now study communication complexity of the distributed schemes falling within the framework (Q-M) and using the ANQ to quantize communications. Our results complement T.3.3.1 and are of two types: (i) first, we determine the number of bits/agent to be used by the ANQ at each iteration to guarantee linear convergence rate (in terms of iterations) of any algorithm within (Q-M) (T.3.5.1); (ii) then, we provide the communication complexity in the setting of (i), that is, the total number of bits/agent transmitted to achieve an ε-solution of (P) (T.3.5.2). Finally, we customize (ii) to some specific distributed algorithms within (Q-M) (Sec. 3.5.1).

Throughout this section, all the results stated in terms of $\mathcal{O}$-notation are meant asymptotically when $m, d \to \infty$. Also, the following additional mild assumption is postulated, which is satisfied by a variety of existing algorithms, see App.3.8.4.

Assumption 9. *The constants L_A, L_C, L_Z and the initial conditions $\|\mathcal{C}^s(\mathbf{z}^0, \mathbf{0})\|_2$ and $\|\mathbf{z}^0 - \mathbf{z}^\infty\|$ satisfy*

$$\begin{aligned}&L_A L_Z = \mathcal{O}(1), \quad L_C = \mathcal{O}(1), \quad \|\mathcal{C}^s(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md}),\ \forall s \in [R],\\ &\textit{and} \quad \|\mathbf{z}^0 - \mathbf{z}^\infty\| = \mathcal{O}(\sqrt{md}).\end{aligned}$$

Our first result on the number of bits transmitted at each iteration to sustain linear convergence is summarized next.

Theorem 3.5.1. *Instate the setting of T.3.3.1, under the additional A.9. Furthermore, suppose that the deterministic ANQ (or probabilistic ANQ with $\mathbb{E}[\mathcal{Q}(\mathbf{x})] = \mathbf{x}$) is used to quantize all the communications in (Q-M), with $\eta^0 = \Theta(L_Z(\sigma - \lambda))$ and ω such that $1 - \omega/\bar{\omega}(\sigma) = \Omega(1)$. Then, linear convergence $\sqrt{\mathbb{E}[\|\mathbf{z}^k - \mathbf{z}^\infty\|_2^2]} = \mathcal{O}(\sqrt{md} \cdot (\sigma)^k)$, $k \in \mathbb{Z}_+$, is achieved with an average number of bits/agent at every iteration k given by*

$$\mathcal{O}\left(d \log_2\left(1 + \frac{1}{\sigma(\sigma - \lambda)}\right)\right). \tag{3.21}$$

Proof. See App.3.8.3. □

The following comments are in order.

(i) As expected, the faster the quantized algorithm (smaller σ), the larger the communication cost incurred per iteration; in particular, when $\sigma \to \lambda$, the number of bits required to sustain linear convergence at rate σ grows indefinitely. In other words, an infinite number of bits is required if a quantized distributed scheme (Q-M) wants to match the convergence rate of its unquantized counterpart.

(ii) It is interesting to contrast the communication efficiency (bits transmitted per iteration) of the proposed model (Q-M) equipped with the ANQ with that of existing schemes. Specifically, the schemes in [59], [60] use $\mathcal{O}(d \log_2(1+\frac{\sqrt{md}}{\sigma(\sigma-\lambda)}))$ bits/agent/iteration over mesh networks while $\mathcal{O}(d \log_2(1+\frac{\sqrt{d}}{\sigma(\sigma-\lambda)}))$ bits/agent/iteration are required in the analysis of [59] over star networks. Both are less favorable than (3.21). This can be attributed to the fact that the ANQ adapts the number of bits to the input signal rather than adopting a constant number of bits for any input signal as [59], [60].

(iii) T.3.5.1 reveals a tension between convergence rate (the closer σ to λ, the faster the algorithm) and number of transmitted bits per iteration (the larger σ, the smaller the cost). We provide a favorable choice of σ to exploit this trade-off and determine consequently the total number of transmitted bits/agent to achieve a target ε-accuracy.

Theorem 3.5.2. *Instate the setting of T.3.5.1, with $R = \mathcal{O}(1)$ and σ chosen so that $\frac{(1-\lambda)^2}{(1-\sigma)(\sigma-\lambda)} = \mathcal{O}(1)$. Then, the following average number of (transmitted) bits/agent is sufficient for $\{\mathbf{z}^k\}_{k\in\mathbb{Z}_+}$ to achieve $(1/m)\mathbb{E}[\|\mathbf{z}^k - \mathbf{z}^\infty\|_2^2] \leq \varepsilon$:*

$$\mathcal{O}\biggl(d \log_2\left(1+\frac{1}{1-\lambda}\right)\frac{1}{1-\lambda}\log_2(d/\varepsilon)\biggr) \quad \text{bits/agent.} \tag{3.22}$$

Proof. See App.3.8.3. □

The following comments are in order.

(i) Intuitively, T.3.5.2 provides a range of values of σ to balance the conflicting effect on the convergence rate and communication cost per iteration resulting from too small or too large values of σ. The condition of the theorem can be satisfied, e.g., by choosing $\sigma = (1+\lambda)/2$.

(ii) The term $d \cdot \log_2(1 + 1/(1-\lambda))$ in (3.22) represents the number of bits/iteration, as postulated by T.3.5.2, under the additional restriction on σ: the faster the unquantized algorithm (i.e., the smaller λ), the fewer bits are required. The second term $(1-\lambda)^{-1}\log_2(d/\varepsilon)$ represents the total number of iterations required to achieve ε accuracy for the unquantized algorithm, with $\log_2(d)$ capturing the gap of the initial point from the fixed point; as expected, the number of iterations increases as the unquantized algorithm slows down (λ increases), the dimension d increases, and/or the target error ε decreases.

Nest, we customize T.3.5.2 to some distributed algorithms within (Q-M).

3.5.1 Special cases of (Q-M) using the ANQ rule

1) GD over star-networks: Our first case study is the GD algorithm solving (P) (with $r = 0$) over star networks. The unquantized scheme reads

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{\gamma}{m}\sum_{i=1}^{m} \nabla f_i(\mathbf{x}^k), \tag{3.23}$$

with $\gamma \in (0, 2/L)$. In [59], it is shown that, when employed with the quantization scheme proposed therein, the resulting quantized GD achieves an ε-solution by transmitting on average

$$\mathcal{O}\bigg(\log_2\Big(1 + \frac{d}{1-\lambda}\Big)\frac{d}{1-\lambda}\log_2\big(d/\varepsilon\big)\bigg) \quad \text{bits/agent.} \tag{3.24}$$

The GD (3.23) is an instance of (M); therefore, we can employ quantization in the master-workers' communications and cast the resulting quantized algorithm in (Q-M). When the ANQ is used, a direct application of T.3.5.2 leads to the following communication complexity.

Corollary 3 (GD over star networks)**.** *The GD algorithm* (3.23) *employing quantization in the master-workers' communications using the ANQ with tuning as in T.3.5.2 requires on average of*

$$\mathcal{O}\bigg(\log_2\Big(1 + \frac{1}{1-\lambda}\Big)\frac{d}{1-\lambda}\log_2\big(d/\varepsilon\big)\bigg) \quad \text{bits/agent,} \tag{3.25}$$

to reach an ε-solution of (P) *(with $r = 0$).*

A direct comparison of (3.24) and (3.25) shows that the proposed ANQ improves on the state-of-the-art deterministic quantizer with shrinking range developed in [59].

2) Distributed algorithms employing gradient correction: Our second example deals with distributed algorithms solving (P) (now possibly with $r \neq 0$) over mesh networks. We consider among the most popular ones, employing gradient correction in the optimization direction. Their computational complexity when using the ANQ is summarized next–see App.3.8.4 for a description of such algorithms.

Corollary 4 (mesh networks)**.** *Consider any of the following algorithms using the ANQ in agents' communications, with parameters as specified in T.3.5.2: the primal-dual algorithm (solving* (P) *with* $r = 0$*) [59]; the (Prox-)EXTRA, (Prox-)NEXT, (Prox-)DIGing, and (Prox-)NIDS [80] (for general* (P) *with* $r \neq 0$*). An* ε*-solution is achieved by transmitting on average*

$$\mathcal{O}\left(\log_2\left(1+\frac{1}{1-\lambda}\right)\frac{d}{1-\lambda}\log_2\left(d/\varepsilon\right)\right) \quad \text{bits/agent.}$$

To our knowledge, C.4 provides the first analytical result on the communication cost to achieve an ε-solution of Problem (P) by distributed algorithms over mesh networks. In particular, the schemes employing the proposed ANQ are the first algorithms using finite rate communications when applied to such a general class of optimization problems (with $r \neq 0$) and networks (mesh topology).

3.6 Numerical Results

In this section, we validate numerically our theoretical findings and compare different distributed algorithms using quantization. We simulate two instances of (P): a least square and a logistic regression problem. The communication network is modelled by an undirected graph of $m = 20$ agents, generated by the Erdos-Renyi model with edge activating probability of 0.6. We measure performance of the algorithms using the following two metrics:

$$\text{MSE}^k \triangleq \frac{\sum_{i=1}^m \|\mathbf{x}_i^k - \mathbf{x}^*\|_2^2}{m\|\mathbf{x}^*\|_2^2} \quad \text{and} \quad \text{C}_{\text{cm}}(\varepsilon) \triangleq \sum_{k=0}^{k_\varepsilon}\sum_{s=1}^{R}\sum_{i=1}^{m} b_i^{k,s}, \tag{3.26}$$

where $k_\varepsilon \triangleq \min_{k\in\mathbb{Z}_+} \mathrm{MSE}^k \leq \varepsilon$ and $b_i^{k,s}$ is the number of bits used by the quantizer for encoding the sth transmitted signal by agent i at iteration k.

3.6.1 Least square problem

Problem setting: Consider the following least square problem [instance of (P)] over networks:

$$f_i(\mathbf{x}) = \frac{1}{2}\|\mathbf{U}_i\mathbf{x} - \mathbf{v}_i\|_2^2 + \frac{0.01}{2}\|\mathbf{x}\|_2^2 \quad \text{and} \quad r(\mathbf{x}) = \alpha\|\mathbf{x}\|_1, \tag{3.27}$$

where $\mathbf{U}_i \in \mathbb{R}^{20\times 40}$ and $\mathbf{v}_i \in \mathbb{R}^{1\times 40}$ are the feature vector and observation measurements, respectively, accessible only by agent i. These are generated as follows [86]: $\mathbf{U}_1 \sim \mathcal{N}(\mathbf{0}, \frac{1}{\sqrt{1-\beta^2}}\mathbf{I})$ and, for $i > 1$, $\mathbf{U}_i|\mathbf{U}_{i-1} \sim \mathcal{N}(\beta\mathbf{U}_{i-1}, \mathbf{I})$, where $\beta = 0.3$. In this way, each row of $\mathbf{U} = [\mathbf{U}_1, \cdots \mathbf{U}_{40}]$ is a Gaussian random vector with zero mean and covariance depending on β: larger β generates more ill-conditioned covariance matrices. Then, letting $\mathbf{x}_0 \in \mathbb{R}^{40}$ be the ground truth vector, generated as a sparse vector with 70% zero entries, and i.i.d. nonzero entries drawn from $\mathcal{N}(0,1)$, we generate $\mathbf{v} = [\mathbf{v}_1, \cdots \mathbf{v}_{40}]$ as $\mathbf{v}|(\mathbf{U}, \mathbf{x}_0) \sim \mathcal{N}(\mathbf{U}\mathbf{x}_0, 0.04\mathbf{I})$. We let μ, L be the strong convexity and smoothness parameters of f_i, respectively.[5]

We test several distributed algorithms considering either smooth ($\alpha = 0$) or nonsmooooth instances of the least square problem (3.27). In fact, most of the existing schemes are applicable only to smooth optimization problems. The free parameters of these algorithms are tuned as recommended in the original papers, unless otherwise stated; the weight matrix $\tilde{\mathbf{W}}$ used to mix the received signals is constructed according to the Metropolis-Hastings rule [35]; the number of bits transmitted by each scheme as reported below is per agent, per dimension, per iteration. For each quantized algorithm, we choose $\sigma = 0.99 \cdot \lambda + 0.01$, where $\lambda = (\mathrm{MSE}^{100}/\mathrm{MSE}^{50})^{0.01}$ is the numerical estimate of the convergence rate of its unquantized counterpart.

Smooth least square (Fig. 3.2a): We begin by considering the smooth least square problem and the following quantized schemes:

[5] ↑When f_i is μ_i-strongly convex and L_i-smooth, we let $\mu = \min_i \mu_i$ and $L = \max_i L_i$.

1) `Q-Dual` [59]: parameters are chosen as in [59, T.1]–the averaged number of transmitted bits is 13.

2) `ANQ-Dual`: this is the primal-dual algorithm [75] equipped with the proposed deterministic ANQ (see App.12), with $\eta^0 = 0.01$ and $\omega = \bar{\omega}/2$ [recall that $\bar{\omega}$ is defined in (3.11)]–the average transmitted number of bits is 6.63.

3) `Q-NEXT` [60], with quantization as in [60, T.4]–the average transmitted number of bits is 78.

4) `ANQ-NEXT`: this is the NEXT algorithm [8], [62], [79] quantized using the deterministic ANQ with $\eta^0 = 0.029$ and $\omega = \bar{\omega}/2$. The average transmitted number of bits is 15.69.

5) `ANQ-NIDS`: this is an instance of the NIDS algorithm [80], [82] equipped with the deterministic ANQ (see App.12) with parameters $\eta^0 = 0.001$ and $\omega = \bar{\omega}/2$. The transmitted average number of bits is 8.69.

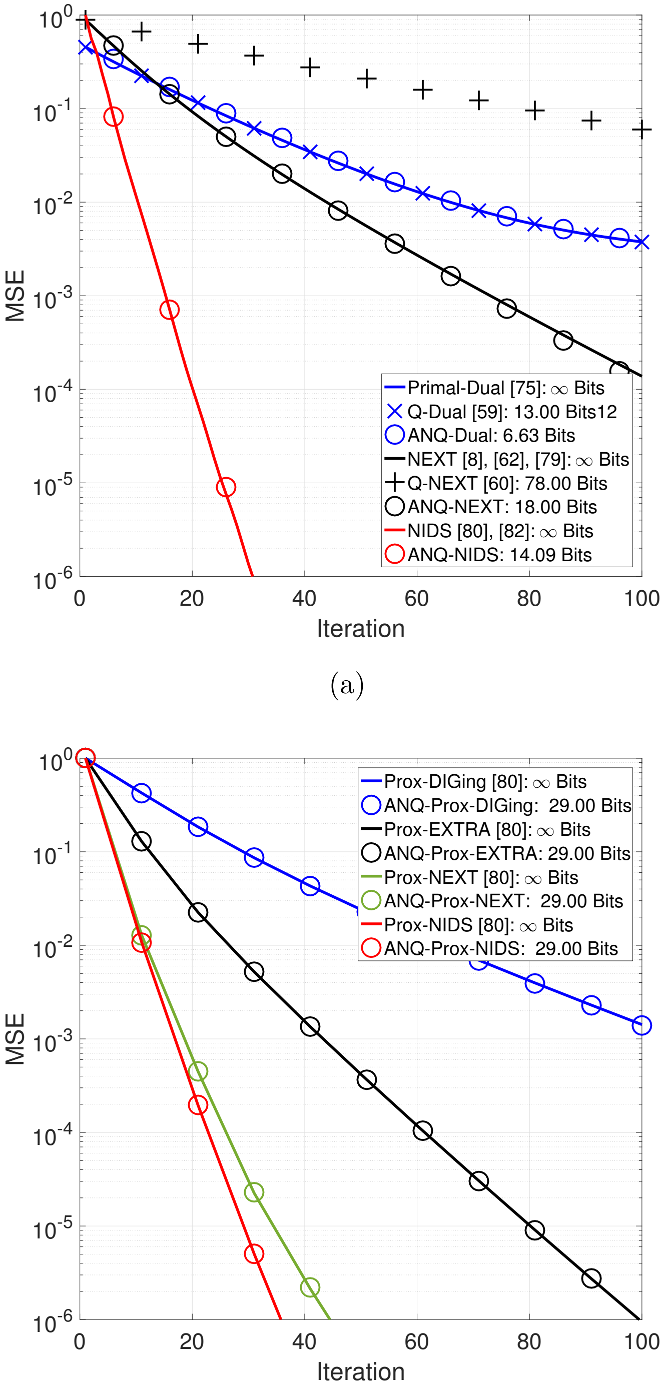
As benchmark, we also simulated some of the unquantized instances of the algorithms listed above, namely:

6) `Primal-Dual` [75] with step-size $\gamma = 2L\mu/(\mu\rho_{m-1}(\mathbf{L}) + L\rho_1(\mathbf{L}))$ ([59, P.2]), where $\mathbf{L}$ is the graph Laplacian matrix associated with the graph.

7) `NEXT` [8], [62], [79] with step-size $\gamma = 0.0029$, manually tuned for fastest practical convergence.

8) `NIDS` [80], [82] with step-size $\gamma = \frac{2}{L+\mu}$ and mixing matrix $\mathbf{W} = [(1+\nu)\mathbf{I}+(1-\nu)\tilde{\mathbf{W}}]/2$, with $\nu = 0.001$.

In Fig. 3.2a, we plot the MSE versus iteration index k. Remarkably, all algorithms, when equipped with ANQ, incur in a negligible loss of convergence speed with respect to their unquantized counterpart. Comparing ANQ with the state-of-the-art quantized algorithms, we notice that ANQ is more communication-efficient than Q-NEXT and Q-Dual that instead



(a)

(b)

Figure 3.2. Least square problem (3.27): MSE versus iterations for the smooth (a) and non-smooth (b) cases. Solid curves and markers refer to algorithms with exact and quantized communications, respectively.

use deterministic uniform quantizers with shrinking range: ANQ-NEXT (18 bits) and ANQ-Dual (6.63 bits) use fewer bits per iteration than Q-NEXT (13 bits) and Q-Dual (78 bits), respectively, despite converging faster.

Nonsmooth least square (Fig. 3.2b): We now move to the nonsmooth instance of (3.27), with $\alpha = 10^{-4}$. We tested the following quantized algorithms:

1) `ANQ-Prox-EXTRA`: this is an instance of the Prox-EXTRA algorithm [80] equipped with the deterministic ANQ (see App.12) with parameters $\eta^0 = 2.68 \times 10^{-5}$ and $\omega = \bar{\omega}/2$.

2) `ANQ-Prox-NEXT`: this is the Prox-NEXT algorithm [80] equipped with the deterministic ANQ (see App.12) with parameters $\eta^0 = 2.85 \times 10^{-3}$ and $\omega = \bar{\omega}/2$.

3) `ANQ-Prox-DIGing`: this is the Prox-DIGing algorithm [80] equipped with the deterministic ANQ (see App.12) with parameters $\eta^0 = 3.65 \times 10^{-3}$ and $\omega = \bar{\omega}/2$.

4) `ANQ-Prox-NIDS`: this is the Prox-NIDS algorithm in [80] equipped with the deterministic ANQ (see App.12) with parameters $\eta^0 = 3.08 \times 10^{-5}$ and $\omega = \bar{\omega}/2$.

As benchmark, we also included the unquantized counterparts of the above algorithms; in all these schemes we used the weight matrix $\mathbf{W} = [(1+\nu)\mathbf{I}+(1-\nu)\tilde{\mathbf{W}}]/2$ with $\nu = 0.001$; the step-size is chosen according to [80], namely: $\gamma = \frac{2\rho_m(\mathbf{W})}{L+\mu\rho_m(\mathbf{W})}$ for Prox-EXTRA, $\gamma = \frac{2\rho_m(\mathbf{W}^2)}{L+\mu\rho_m(\mathbf{W}^2)}$ for Prox-DIGing, and $\gamma = \frac{2}{L+\mu}$ for Prox-NEXT and Prox-NIDS.

Fig. 3.2b plots the MSE achieved by all the algorithms versus the iteration index. As predicted, all four quantized schemes converge linearly. Remarkably, all of the ANQ-equipped algorithms incur in a negligible loss of convergence speed with respect to their unquantized counterparts, while transmitting only 29 bits per agent/dimension/iteration.

3.6.2 Logistic regression

We now consider the distributed logistic regression problem using the MNIST dataset [87]. This is an instance of (P) with

$$f_i(\mathbf{x}) = \frac{0.01}{2}\|\mathbf{x}\|_2^2 + \frac{1}{3000}\sum_{p=1}^{3000} \ln\Big(1+\exp\big(-v_{i,p}\mathbf{u}_{i,p}^\top\mathbf{x}\big)\Big), \quad \text{and} \quad r(\mathbf{x}) = \alpha\|\mathbf{x}\|_1, \tag{3.28}$$

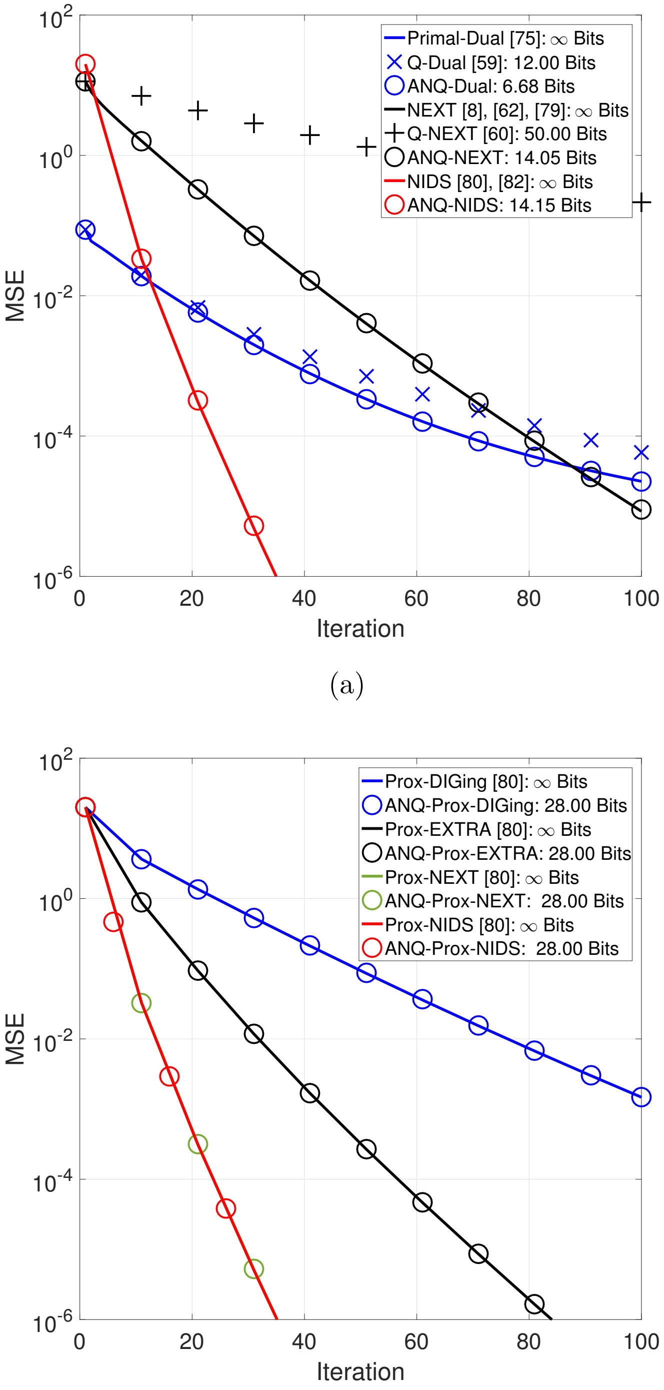
where $\mathbf{u}_{i,p} \in \mathbb{R}^{784\times 1}$ and $v_{i,p} \in \{-1, 1\}$ are the feature vector and labels, respectively, only accessible by agent i. Here we implement the one-vs.-all scheme, i.e., the goal is to distinguish the data of label '0' from others. To generate $\mathbf{u}_{i,p}$, we first flatten each picture of size 28×28 in MNIST into a real feature vector of length $28 \times 28 = 784$, and then normalize it to unit l_2 norm. We then allocate equal number of feature vectors and labels to each agent.

Smooth logistic regression (Fig. 3.3a): We begin by considering the smooth logistic regression problem (3.28), with $\alpha = 0$. We tested the same algorithms (with the same tuning) as described in Sec. 3.6.1 for the smooth least square problem. In Fig. 3.3a, we plot the MSE versus iteration index k. Consistently with the results in Fig. 3.2a, we notice the following facts. ANQ-NIDS achieves the fastest convergence, followed by ANQ-NEXT, ANQ-Dual, Q-Dual and Q-NEXT. Comparing our quantization method with existing ones on the same algorithm, we notice that the proposed ANQ is more communication-efficient than Q-NEXT and Q-Dual: ANQ-NEXT (14 bits) and ANQ-Dual (6.7 bits) use fewer bits per iteration than Q-NEXT (50 bits) and Q-Dual (12 bits), despite converging faster.

Nonsmooth logistic regression (Fig. 3.3b): We now consider the nonsmooth instance of the logistic regression problem (3.28), with $\alpha = 10^{-4}$. We tested the same algorithms (with the same tuning) as described in Sec. 3.6.1 for the nonsmooth least square problem. Fig. 3.3b plots the MSE achieved by all the algorithms versus iterations k. The results confirm the trends already commented in Fig. 3.2b.

3.6.3 Communication cost

We now study the effect of dimension d on the communication cost for different algorithms solving the nonsmooth least square problem (3.27), with $\alpha = 10^{-4}$. Note that the rate of an unquantized algorithm λ depends on both the weight matrix and the condition number $\kappa = L/\mu$, which depends itself on d. We chose the coefficient of the l_2 regularizer so as to make κ remain fixed across different d. The rest of the settings are the same as in Fig. 3.2b. Fig. 3.4 plots the communication cost versus d, for a target MSE-accuracy $\varepsilon = 10^{-8}$. One can observe that the communication cost for all algorithms scales roughly linearly with respect to the dimension, which is consistent with T.3.5.2.



(a)

(b)

Figure 3.3. Logistic regression (3.28): MSE versus iterations for the smooth (a) and non-smooth (b) cases. Solid curves and markers refer to algorithms with exact and quantized communications, respectively.

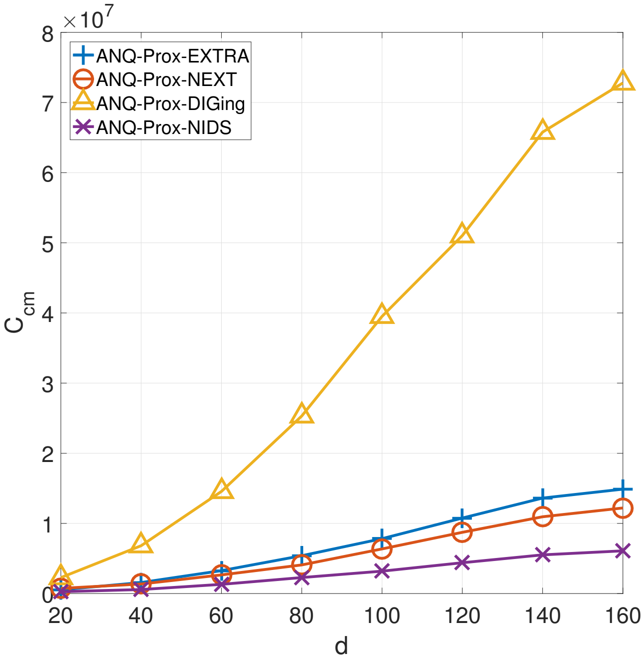
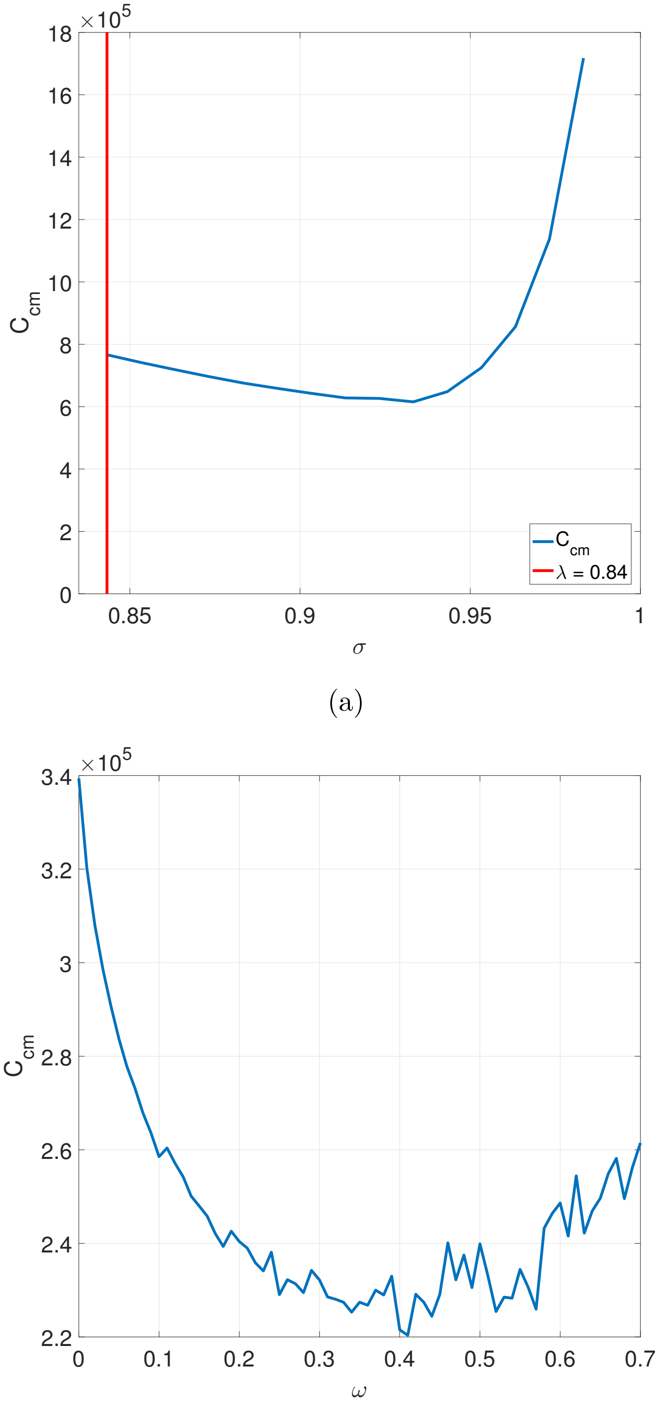


Figure 3.4. Communication cost evaluation on nonsmooth least square problem with $\alpha = 10^{-4}$ versus d.

Finally, we investigate numerically the effect of σ and ω on the communication cost as defined in (3.26), for a target MSE-accuracy $\varepsilon = 10^{-8}$. We consider the ANQ-NIDS algorithm solving the least square problem (3.27), with $\alpha = 0$. Fig. 3.5a plots the communication cost versus σ with $\omega = \bar{\omega}/2$. Note that there is a sweet spot for σ, resulting in a saving of 64% with respect to the largest communication cost, which justifies the discussion in Sec. 3.5 that σ should be chosen away from λ and 1 in order to save on communication cost. Fig. 3.5a plots the communication cost (3.26) versus ω with $\sigma = 0.99 \times \lambda + 0.01$. It can be seen that, by optimizing the compression rate ω, a saving of 30% in communication cost can be obtained over a quantization scheme that employs no compression ($\omega = 0$). This observation numerically supports our BC-rule, which is more general than the deterministic/probabilistic quantizers that have no compression term.



(a)

(b)

Figure 3.5. Communication cost evaluation of ANQ-NIDS on smooth least square problem: (a) effect of σ with $\omega = \bar{\omega}/2$; and (b) ω with $\sigma = 0.99 \times \lambda + 0.01$.

3.7 Conclusions

This chapter proposes a black-box model and unified convergence analysis for a general class of linearly convergent algorithms subject to quantized communications. This generalizes existing algorithmic frameworks, which cannot deal with composite optimization problems and model distributed algorithms using historical information (e.g., EXTRA [76] and NEXT [8]). Quantizaton is addressed by proposing a novel *biased compression (BC-)rule* that preserves linear convergence of distributed algorithms while using a *finite* number of bits in each communication. In fact, we proved that most of existing quantization rules can be implemented only using an infinite number of bits. As special instance of the BC-rule, we also proposed a new (random) quantizer, the ANQ, coupled with a communication-efficient encoding scheme. The communication cost of a variety of distributed algorithms equipped with the ANQ was analyzed (in a unified fashion), showing favorable performance analytically and numerically with respect to existing quantization rules and ad-hoc distributed algorithms.

3.8 Appendix

3.8.1 Proof of Theorem 3.3.1

In this appendix we prove T.3.3.1. We begin by introducing some preliminary results, whose proofs are deferred to the end of this section. Throughout this section, we make the blanket assumption that the conditions in T.3.3.1 are satisfied. In particular, $\sigma \in (\lambda, 1)$ and $\omega \in [0, \bar{\omega}(\sigma))$, with $\bar{\omega}(\sigma)$ defined in (3.11). Due to the possibly random nature of the quantizer, $\{\mathbf{z}^k, \mathbf{c}^{k,s}, \hat{\mathbf{c}}^{k,s}\}_{k \in \mathbb{Z}_+, s \in [R]}$ is a stochastic process defined on a proper probability space; we denote by $\mathcal{F}^{k,s}$ the σ-algebra generated by $\{\mathbf{z}^k, \mathbf{c}^{k,s}, \hat{\mathbf{c}}^{k,s}\}_{k<k, s \in [R]} \cup \{\mathbf{z}^k, \mathbf{c}^{k,s}, \hat{\mathbf{c}}^{k,s-1}\}_{s \leq s}$ ($\hat{\mathbf{c}}^{k,s}$ excluded).

Preliminaries

The idea of the proof is to show by induction that both the optimization error $\|\mathbf{z}^k - \mathbf{z}^\infty\|$ and the input to the quantizer, $\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\|_2$, are linearly convergent (in expectation) at rate σ, i.e.,

$$\sqrt{\mathbb{E}[\left\|\mathbf{z}^k - \mathbf{z}^\infty\right\|^2]} \leq V_0 \cdot (\sigma)^k, \tag{3.29}$$

$$\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\right\|_2^2]} \leq F^s \cdot (\sigma)^k, \quad \forall s \in [R] \cup \{0\}, \tag{3.30}$$

where $F^0 = 0$, and V_0, F^s, $s \in [R]$ satisfy

$$V_0 \geq \max\left\{\|\mathbf{z}^0 - \mathbf{z}^\infty\|, \frac{\sqrt{md}R\eta^0 + \omega\mathbf{F}^\top\mathbf{1}}{\sigma - \lambda}\tilde{L}_A\right\}, \tag{3.31}$$

$$\begin{aligned} F^s \geq \max\Big\{ & L_Z c^* + L_C\sqrt{md}\eta^0 + L_C(1+\omega)F^{s-1}, \\ & \frac{\sqrt{md}\eta^0(1 + L_C\sigma) + L_C\sigma(1+\omega)F^{s-1} + L_Z(1+\sigma)V_0}{\sigma - \omega}\Big\}, \ \forall s \in [R], \end{aligned} \tag{3.32}$$

and we have defined $\mathbf{F} \triangleq (F^s)_{s\in[R]}$,

$$c^* \triangleq \frac{1}{L_Z}\max_{s\in[R]}\|\mathcal{C}^s(\mathbf{z}^0, \mathbf{0})\|_2, \tag{3.33}$$

$$\tilde{L}_A \triangleq L_A\sum_{s=0}^{R-1}(L_C)^s. \tag{3.34}$$

The existence of such V_0 and F^s, $s \in [R]$, is proved in the following lemma.

Lemma 14. *Let $\omega \in [0, \bar{\omega}(\sigma))$. Then, (3.31) and (3.32) are satisfied by*

$$V_0 = \max\left\{c^*, \|\mathbf{z}^0 - \mathbf{z}^\infty\|\right\} + \frac{L_A\psi\sqrt{md}R^2\eta^0}{\sigma - \lambda}\frac{1 + \frac{R\omega}{\sigma}[(1 + L_C\sigma)\psi - 1]}{1 - \omega/\bar{\omega}}, \tag{3.35}$$

$$F^s = \frac{\sqrt{md}\eta^0(1 + L_C\sigma) + 2L_Z V_0(\sigma, \omega, \eta^0)}{\sigma - \omega}\sum_{s=0}^{s-1}\left(\frac{2L_C}{1 - \omega/\sigma}\right)^s, \quad s \in [R], \tag{3.36}$$

where $\psi \triangleq \max\{1, (2L_C)^{R-1}\}$.

Since the effect of $\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\|_2$ on $\|\mathbf{z}^k - \mathbf{z}^\infty\|$ is through quantization, we need the following bound on the quantization error (its proof follows readily by the Cauchy–Schwarz inequality).

Lemma 15. *Let $\mathcal{Q}_i$, $i \in [m]$, satisfy the BC-rule (3.6) with bias $\eta \geq 0$ and compression rate $\omega \in [0, \bar{\omega}(\sigma))$. Then the following holds for the stack $\mathcal{Q} \triangleq [\mathcal{Q}_1, \ldots, Q_m]^\top$:*

$$\sqrt{\mathbb{E}[\left\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\right\|_2^2]} \leq \sqrt{md}\,\eta + \omega\|\mathbf{x}\|_2, \quad \forall \mathbf{x} = [\mathbf{x}_1^\top, \ldots, \mathbf{x}_m^\top]^\top \in \mathbb{R}^{md}. \tag{3.37}$$

A direct application of L.15 leads to the following bound on the quantizer's input, which we use recurrently in the proofs:

$$\begin{aligned}\sqrt{\mathbb{E}\Big[\left\|\hat{\mathbf{c}}^{k,s} - \mathbf{c}^{k,s}\right\|_2^2 |\mathcal{F}^{k,s}\Big]} &\overset{\text{(Q-M)}}{=} \sqrt{\mathbb{E}\Big[\left\|\mathcal{Q}^k(\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}) - (\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s})\right\|_2^2 |\mathcal{F}^{k,s}\Big]} \\ &\overset{(3.37)}{\leq} \sqrt{md}\,\eta^0 \cdot (\sigma)^k + \omega\left\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\right\|_2, \quad a.s.,\end{aligned} \tag{3.38}$$

for all $s \in [R]$ and $k \in \mathbb{Z}_+$, where we used $\eta^k = \eta^0 \cdot (\sigma)^k$.

The following lemma bounds the distortion introduced by quantization in one iteration of (Q-M).

Lemma 16. *There holds: for all $k \in \mathbb{Z}_+$,*

$$\sqrt{\mathbb{E}[\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\|^2]} \leq \lambda\sqrt{\mathbb{E}[\left\|\mathbf{z}^k - \mathbf{z}^\infty\right\|^2]} + \tilde{L}_A\sqrt{md}R\eta^0 \cdot (\sigma)^k + \tilde{L}_A\omega\sum_{s=1}^R \sqrt{\mathbb{E}[\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\|_2^2]},$$

a.s., where $\tilde{L}_A$ is defined in (3.34).

We conclude this section of preliminaries with the following useful result.

Lemma 17. *Let $\{X_t : t \in [T]\} \subset \mathbb{R}$ be a collection of random variables. Then,*

$$\sqrt{\mathbb{E}\Big[\Big(\sum_{t=1}^T X_t\Big)^2\Big]} \leq \sum_{t=1}^T \sqrt{\mathbb{E}[X_t^2]}.$$

Proof. It can be proved by developing the square within the expectation on the left hand side expression, and by using $\mathbb{E}[X_t X_u] \leq \sqrt{\mathbb{E}[X_t^2]}\sqrt{\mathbb{E}[X_u^2]}$. □

Proof of Theorem 3.3.1

We prove (3.29) and (3.30) by induction. Let V_0 and F^s, $s \in [R]$, satisfy (3.31) and (3.32). Since $\|\mathbf{z}^0 - \mathbf{z}^\infty\| \le V_0$ (see (3.31)) and $\mathbf{c}^{0,0} = \hat{\mathbf{c}}^{-1,0} = \mathbf{0}$, (3.29) holds for $k = 0$ and (3.30) holds trivially for $k = 0$ and $s = 0$. We now use induction to prove that (3.30) holds for $k = 0$ and $s \in [R]$. Assume that (3.30) holds for $k = 0$ and $s < R$. Then, it follows that

$$\begin{aligned}
&\left\|\mathbf{c}^{0,s+1} - \hat{\mathbf{c}}^{-1,s+1}\right\|_2 = \left\|\mathbf{c}^{0,s+1}\right\|_2 \\
&\overset{(a)}{\le} \left\|\mathcal{C}^{s+1}\Big(\mathbf{z}^0, \mathbf{0}\Big)\right\|_2 + \left\|\mathcal{C}^{s+1}\Big(\mathbf{z}^0, \hat{\mathbf{c}}^{0,s}\Big) - \mathcal{C}^{s+1}\Big(\mathbf{z}^0, \mathbf{c}^{0,s}\Big)\right\|_2 + \left\|\mathcal{C}^{s+1}\Big(\mathbf{z}^0, \mathbf{c}^{0,s}\Big) - \mathcal{C}^{s+1}\Big(\mathbf{z}^0, \mathbf{0}\Big)\right\|_2 \\
&\overset{(3.9),(3.33)}{\le} L_Z c^* + L_C \left\|\hat{\mathbf{c}}^{0,s} - \mathbf{c}^{0,s}\right\|_2 + L_C \left\|\mathbf{c}^{0,s} - \hat{\mathbf{c}}^{-1,s}\right\|_2, \quad a.s.,
\end{aligned}$$

where in (a) we used the triangle inequality and $\mathbf{c}^{0,s+1} = \mathcal{C}^{s+1}\Big(\mathbf{z}^0, \hat{\mathbf{c}}^{0,s}\Big)$. Taking the conditional expectation on both sides and using L.17 yield

$$\begin{aligned}
\sqrt{\mathbb{E}\Big[\left\|\mathbf{c}^{0,s+1} - \hat{\mathbf{c}}^{-1,s+1}\right\|_2^2 | \mathcal{F}^{0,s}\Big]} &\le L_Z c^* + L_C \sqrt{\mathbb{E}\Big[\left\|\hat{\mathbf{c}}^{0,s} - \mathbf{c}^{0,s}\right\|_2^2 | \mathcal{F}^{0,s}\Big]} + L_C \left\|\mathbf{c}^{0,s} - \hat{\mathbf{c}}^{-1,s}\right\|_2 \\
&\overset{(3.38)}{\le} L_Z c^* + L_C \sqrt{md}\eta^0 + L_C(1+\omega)\left\|\mathbf{c}^{0,s} - \hat{\mathbf{c}}^{-1,s}\right\|_2, \quad a.s..
\end{aligned}$$

Taking the unconditional expectation on both sides and using again L.17 yield

$$\begin{aligned}
\sqrt{\mathbb{E}\Big[\left\|\mathbf{c}^{0,s+1} - \hat{\mathbf{c}}^{-1,s+1}\right\|_2^2\Big]} &\le L_Z c^* + L_C \sqrt{md}\eta^0 + L_C(1+\omega)\sqrt{\mathbb{E}[\left\|\mathbf{c}^{0,s} - \hat{\mathbf{c}}^{-1,s}\right\|_2^2]} \\
&\overset{(3.30)}{\le} L_Z c^* + L_C \sqrt{md}\eta^0 + L_C(1+\omega)F^s \overset{(3.32)}{\le} F^{s+1},
\end{aligned}$$

which completes the induction proof of (3.30) for $k = 0$ and $s \in [R]$. Now, let us assume that (3.29) and (3.30) hold for a generic $k \in \mathbb{Z}_+$; we prove that they hold at $k+1$. We begin with (3.29). Invoking L.16, L.17, and using the induction hypotheses (3.29) and (3.30) at k, yield

$$\sqrt{\mathbb{E}[\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\|^2]} \le \lambda V_0 \cdot (\sigma)^k + \tilde{L}_A\Big(\sqrt{md}R\eta^0 + \omega \mathbf{F}^\top \mathbf{1}\Big) \cdot (\sigma)^k \le V_0 \cdot (\sigma)^{k+1},$$

where the last inequality follows from the definition of V_0 in (3.31), which concludes the induction argument for (3.29).

We now prove that (3.30) holds for $k+1$, by induction over $s \in [R]$. First, note that (3.30) holds trivially for $k+1$ and $s=0$, since $\mathbf{c}^{k+1,0} = \hat{\mathbf{c}}^{k,0} = \mathbf{0}$. Now, assume that (3.30) holds at iteration $k+1$ for $s < R$. Then,

$$\begin{aligned}
&\left\|\mathbf{c}^{k+1,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2 \overset{\text{(Q-M)}}{=} \left\|\mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \hat{\mathbf{c}}^{k+1,s}\right) - \mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \mathbf{c}^{k+1,s}\right)\right. \\
&\left. + \mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \mathbf{c}^{k+1,s}\right) - \mathcal{C}^{s+1}\left(\mathbf{z}^{k}, \hat{\mathbf{c}}^{k,s}\right) + \mathbf{c}^{k,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2 \\
&\overset{(a)}{\leq} \left\|\mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \hat{\mathbf{c}}^{k+1,s}\right) - \mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \mathbf{c}^{k+1,s}\right)\right\|_2 + \left\|\mathcal{C}^{s+1}\left(\mathbf{z}^{k+1}, \mathbf{c}^{k+1,s}\right) - \mathcal{C}^{s+1}\left(\mathbf{z}^{k}, \hat{\mathbf{c}}^{k,s}\right)\right\|_2 \\
&+ \left\|\mathbf{c}^{k,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2 \\
&\overset{(3.9),(3.10)}{\leq} L_C\left\|\hat{\mathbf{c}}^{k+1,s} - \mathbf{c}^{k+1,s}\right\|_2 + L_C\left\|\mathbf{c}^{k+1,s} - \hat{\mathbf{c}}^{k,s}\right\|_2 + L_Z\left\|\mathbf{z}^{k+1} - \mathbf{z}^{\infty}\right\|_2 \\
&+ L_Z\left\|\mathbf{z}^{k} - \mathbf{z}^{\infty}\right\|_2 + \left\|\hat{\mathbf{c}}^{k,s+1} - \mathbf{c}^{k,s+1}\right\|_2, \quad a.s..
\end{aligned}$$

Then, taking the conditional expectation on $\mathcal{F}^{k+1,s}$ and invoking L.17 and (3.38) to bound $\sqrt{\mathbb{E}\left[\left\|\hat{\mathbf{c}}^{k+1,s} - \mathbf{c}^{k+1,s}\right\|_2^2 | \mathcal{F}^{k+1,s}\right]}$, yield

$$\begin{aligned}
&\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k+1,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2^2 | \mathcal{F}^{k+1,s}]} \leq L_C\sqrt{md}\eta^0 \cdot (\sigma)^{k+1} + L_C(1+\omega)\left\|\mathbf{c}^{k+1,s} - \hat{\mathbf{c}}^{k,s}\right\|_2 \\
&+ L_Z\left\|\mathbf{z}^{k+1} - \mathbf{z}^{\infty}\right\|_2 + L_Z\left\|\mathbf{z}^{k} - \mathbf{z}^{\infty}\right\|_2 + \left\|\hat{\mathbf{c}}^{k,s+1} - \mathbf{c}^{k,s+1}\right\|_2, \quad a.s..
\end{aligned}$$

Now, taking the conditional expectation on $\mathcal{F}^{k,s+1} \subseteq \mathcal{F}^{k+1,s}$, invoking L.17, and (3.38) to bound $\sqrt{\mathbb{E}\left[\left\|\hat{\mathbf{c}}^{k,s+1} - \mathbf{c}^{k,s+1}\right\|_2^2 | \mathcal{F}^{k,s+1}\right]}$, yield

$$\begin{aligned}
&\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k+1,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2^2 | \mathcal{F}^{k,s+1}]} \\
&\leq \sqrt{md}\eta^0(1 + L_C\sigma) \cdot (\sigma)^k + L_C(1+\omega)\sqrt{\mathbb{E}[\|\mathbf{c}^{k+1,s} - \hat{\mathbf{c}}^{k,s}\|_2^2 | \mathcal{F}^{k,s+1}]} \\
&+ L_Z\sqrt{\mathbb{E}[\left\|\mathbf{z}^{k+1} - \mathbf{z}^{\infty}\right\|_2^2 | \mathcal{F}^{k,s+1}]} + L_Z\sqrt{\mathbb{E}[\left\|\mathbf{z}^{k} - \mathbf{z}^{\infty}\right\|_2^2 | \mathcal{F}^{k,s+1}]} + \omega\left\|\mathbf{c}^{k,s+1} - \hat{\mathbf{c}}^{k-1,s+1}\right\|_2, \quad a.s..
\end{aligned}$$

Taking the unconditional expectation and invoking L.17 again yield

$$\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k+1,s+1} - \hat{\mathbf{c}}^{k,s+1}\right\|_2^2]}$$

$$
\begin{aligned}
&\leq \sqrt{md}\eta^0(1+L_C\sigma)\cdot(\sigma)^k + L_C(1+\omega)\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k+1,s}-\hat{\mathbf{c}}^{k,s}\right\|_2^2]} \\
&+ L_Z\sqrt{\mathbb{E}[\left\|\mathbf{z}^{k+1}-\mathbf{z}^{\infty}\right\|_2^2]} + L_Z\sqrt{\mathbb{E}[\left\|\mathbf{z}^{k}-\mathbf{z}^{\infty}\right\|_2^2]} + \omega\sqrt{\mathbb{E}[\left\|\mathbf{c}^{k,s+1}-\hat{\mathbf{c}}^{k-1,s+1}\right\|_2^2]} \\
&\overset{(a)}{\leq} \sqrt{md}\eta^0(1+L_C\sigma)\cdot(\sigma)^k + L_C(1+\omega)F^s\cdot(\sigma)^{k+1} + \omega F^{s+1}\cdot(\sigma)^k + L_Z(1+\sigma)V_0\cdot(\sigma)^k \\
&\overset{(3.32)}{\leq} F^{s+1}\cdot(\sigma)^{k+1},
\end{aligned}
$$

where in (a) we used the induction hypotheses (3.30) (applied to the second and last terms) and (3.29) (applied to the third and fourth terms). This proves the induction for (3.30), and the theorem.

Proof of Lemma 14

It is not difficult to check that conditions (3.31) and (3.32) can be satisfied by choosing

$$
\begin{aligned}
&V_0 \geq \max\left\{c^*, \|\mathbf{z}^0-\mathbf{z}^{\infty}\|, \frac{\sqrt{md}R\eta^0+\omega\mathbf{F}^{\top}\mathbf{1}}{\sigma-\lambda}\tilde{L}_A\right\}, \\
&F^s \geq \frac{\sqrt{md}\eta^0(1+L_C\sigma)+L_C\sigma(1+\omega)F^{s-1}+L_Z(1+\sigma)V_0}{\sigma-\omega}, \ \forall s\in[R],
\end{aligned}
$$

where $c^*, \tilde{L}_A$ are defined in (3.33) and (3.34), respectively. Moreover, since $\omega<\sigma<1$, it is sufficient to choose

$$
F^s = \frac{1}{\sigma}\frac{\sqrt{md}\eta^0(1+L_C\sigma)+2L_C\sigma F^{s-1}+2L_ZV_0}{1-\omega/\sigma}, \quad \forall s\in[R].
$$

Solving this expression recursively yields (3.36).

We now prove (3.35). We begin noting that F^s is a non-decreasing function of s, hence $F^s \leq F^R$. Moreover, F^R is an affine function of V_0. Using the facts that $(\frac{2L_C}{1-\omega/\sigma})^s \leq \psi(1-\omega/\sigma)^{-s}$, where $\psi \triangleq \max\{1,(2L_C)^{R-1}\}$, and $(1-\omega/\sigma)^{-s} \leq (1-\omega/\sigma)(1-R\omega/\sigma)^{-1}$, for all $s\in[R-1]\cup\{0\}$ and $\omega<\bar{\omega}(\sigma)<\sigma/R$, we can upper bound F^s as

$$
F^s \leq F^R \leq a_1 + V_0a_2 \triangleq \bar{F}^R, \tag{3.39}
$$

where

$$a_1 \triangleq \psi\sqrt{md}\eta^0(1+L_C\sigma)\cdot\frac{R/\sigma}{1-R\omega/\sigma}, \tag{3.40}$$

$$a_2 \triangleq 2L_Z\psi\cdot\frac{R/\sigma}{1-R\omega/\sigma}. \tag{3.41}$$

Furthermore, since $\mathbf{F}^\top\mathbf{1} \leq RF^R \leq R(a_1+V_0a_2)$ and $\tilde{L}_A = L_A\sum_{s=0}^{R-1}(L_C)^s \leq L_A\psi R$, to satisfy (3.31), it is sufficient to choose V_0 as

$$V_0 \geq \max\left\{c^*, \|\mathbf{z}^0-\mathbf{z}^\infty\|, L_A\psi R^2\frac{\sqrt{md}\eta^0+\omega(a_1+V_0a_2)}{\sigma-\lambda}\right\}.$$

Using $x \geq \max\{c, a+bx\} \Leftrightarrow x \geq \max\{c, a/(1-b)\}$, under $b<1$, the above condition is equivalent to

$$V_0 \geq \max\left\{c^*, \|\mathbf{z}^0-\mathbf{z}^\infty\|, \frac{L_A\psi R^2(\sqrt{md}\eta^0+\omega a_1)}{\sigma-\lambda-L_A\psi R^2\omega a_2}\right\}, \tag{3.42}$$

as long as $L_A\psi R^2\omega a_2/(\sigma-\lambda)<1$. Solving with respect to ω (note that a_2 is a function of ω), this condition is equivalent to $\omega\in[0,\bar{\omega}(\sigma))$ with $\bar{\omega}(\sigma)$ given by (3.11), hence it holds by assumption. Substituting the values of a_1, a_2 in (3.42) and using $\max\{a,b\}\leq a+b$ (for $a,b\geq 0$), yields (3.35). □.

Proof of Lemma 16

At iteration k, let ζ_s be defined as

$$\zeta^s = \mathcal{A}\left(\mathbf{z}^k, \hat{\mathbf{c}}^{k,1}, \ldots, \hat{\mathbf{c}}^{k,s}, \tilde{\mathbf{c}}_s^{s+1}, \ldots, \tilde{\mathbf{c}}_s^R\right),$$

where

$$\tilde{\mathbf{c}}_s^s = \hat{\mathbf{c}}^{k,s} \quad \text{and} \quad \tilde{\mathbf{c}}_s^{\ell+1} \triangleq \mathcal{C}^{\ell+1}\left(\mathbf{z}^k, \tilde{\mathbf{c}}_s^\ell\right), \forall \ell \geq s.$$

In other words, $\tilde{\mathbf{c}}_s^\ell, \zeta^s$ are the communication signals at round ℓ and the updated computation state, respectively, obtained by applying the unquantized communication mapping after round s and the quantized one before round s. Clearly, $\mathbf{z}^{k+1} = \mathcal{A}(\mathbf{z}^k, \hat{\mathbf{c}}^{k,1}, \cdots, \hat{\mathbf{c}}^{k,R}) = \zeta^R$ and $\tilde{\mathcal{A}}(\mathbf{z}^k) = \zeta^0$ (unquantized update of the computation state). It then follows that

$$\mathbf{z}^{k+1} = \zeta^R = \tilde{\mathcal{A}}(\mathbf{z}^k) + \sum_{s=1}^{R} \left(\zeta^s - \zeta^{s-1}\right), \quad a.s..$$

Invoking the triangle inequality yields

$$\left\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\right\| \leq \left\|\tilde{\mathcal{A}}(\mathbf{z}^k) - \mathbf{z}^\infty\right\| + \sum_{s=1}^{R} \|\zeta^s - \zeta^{s-1}\|, \quad a.s.. \tag{3.43}$$

We now study the second term. From the Lipschitz continuity of $\mathcal{A}$ (A.7) and the definition of ζ^s, it holds that

$$\|\zeta^s - \zeta^{s-1}\| \leq L_A \sum_{\ell=s}^{R} \|\tilde{\mathbf{c}}_s^\ell - \tilde{\mathbf{c}}_{s-1}^\ell\|_2, \quad a.s..$$

Furthermore, $\|\tilde{\mathbf{c}}_s^s - \tilde{\mathbf{c}}_{s-1}^s\|_2 = \|\hat{\mathbf{c}}^{k,s} - \mathcal{C}^s(\mathbf{z}^k, \hat{\mathbf{c}}^{k,s-1})\|_2 = \|\hat{\mathbf{c}}^{k,s} - \mathbf{c}^{k,s}\|_2$ a.s., and, for $\ell > s$,

$$\|\tilde{\mathbf{c}}_s^\ell - \tilde{\mathbf{c}}_{s-1}^\ell\|_2 = \|\mathcal{C}^\ell(\mathbf{z}^k, \tilde{\mathbf{c}}_s^{\ell-1}) - \mathcal{C}^\ell(\mathbf{z}^k, \tilde{\mathbf{c}}_{s-1}^{\ell-1})\|_2 \leq L_C \|\tilde{\mathbf{c}}_s^{\ell-1} - \tilde{\mathbf{c}}_{s-1}^{\ell-1}\|_2 \leq \cdots \leq (L_C)^{\ell-s} \|\hat{\mathbf{c}}^{k,s} - \mathbf{c}^{k,s}\|_2,$$

a.s., where the last step follows from induction over ℓ. Replacing these bounds in (3.43), we finally obtain

$$\begin{aligned}\left\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\right\| &\leq \left\|\tilde{\mathcal{A}}(\mathbf{z}^k) - \mathbf{z}^\infty\right\| + L_A \sum_{s=1}^{R} \sum_{\ell=0}^{R-s} (L_C)^\ell \|\hat{\mathbf{c}}^{k,s} - \mathbf{c}^{k,s}\|_2 \\ &\leq \lambda \left\|\mathbf{z}^k - \mathbf{z}^\infty\right\| + \tilde{L}_A \sum_{s=1}^{R} \|\hat{\mathbf{c}}^{k,s} - \mathbf{c}^{k,s}\|_2, \quad a.s.,\end{aligned}$$

where $\tilde{L}_A$ is defined in (3.34) and we used A.6. Taking the conditional expectation on the filtration $\mathcal{F}^{k,s}$ while applying L.17 and (3.38), starting from $s = R, R-1, \ldots, 1$, it follows that

$$\sqrt{\mathbb{E}\left[\left\|\mathbf{z}^{k+1} - \mathbf{z}^\infty\right\|^2 |\mathcal{F}^{k,1}\right]}$$

$$\leq \lambda \left\| \mathbf{z}^k - \mathbf{z}^\infty \right\| + \tilde{L}_A \sqrt{md} R \eta^0 \cdot (\sigma)^k + \tilde{L}_A \omega \sum_{s=1}^{R} \sqrt{\mathbb{E}\left[\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\|_2^2 | \mathcal{F}^{k,1} \right]}, \quad a.s..$$

Finally, taking unconditional expectation and using L.17 concludes the proof.

3.8.2 Proof of Lemmata 11, 12, and Corollary 2

Proof of Lemma 11

Let $\mathcal{Q}(\bullet) : [-\delta, \delta]^d \to \mathbb{Q}^d$ be a component-wise quantizer, with the nth quantizer $\mathcal{Q}_n(\bullet)$ mapping points in the interval $[-\delta, \delta]$ to quantization points in the set $\mathbb{Q}$ (we assume that the same quantizer is applied across all n, since each component is optimized with the same range and number of quantization points). The goal is to define a quantizer $\mathcal{Q}$ which satisfies the BC-rule within $\mathbf{x} \in [-\delta, \delta]^d$ with maximal range δ. To this end, note that a necessary and sufficient condition is

$$|\mathcal{Q}_n(x) - x| \leq \eta + \omega |x|, \ \forall x \in [-\delta, \delta], \ \forall n \in [d]. \tag{3.44}$$

The sufficiency can be proved using Cauchy–Schwarz inequality. To prove the necessity, assume that (3.44) is violated for some $x \in [-\delta, \delta]$, i.e., $|\mathcal{Q}_n(x) - x| > \eta + \omega |x|$, and let $\mathbf{x} = x\mathbf{1}$. It follows that

$$\|\mathcal{Q}(\mathbf{x}) - \mathbf{x}\|_2 = \sqrt{d} |\mathcal{Q}_n(x) - x| > \sqrt{d}\eta + \omega \sqrt{d} |x| = \sqrt{d}\eta + \omega \|\mathbf{x}\|_2,$$

implying the BC-rule is not satisfied at $\mathbf{x}$.

Hence, we now focus on the design of a component-wise quantizer $\mathcal{Q}_n$ satisfying (3.44) with maximal range δ. In the following, we omit the dependence on n for convenience.

Assume that $N = |\mathbb{Q}|$ is odd (the case N even can be studied in a similar fashion, and is provided at the end of this proof for completeness), and let $\mathbb{Q} \triangleq \cup_{\ell=0}^{(N-1)/2} \{\tilde{q}_\ell, -\tilde{q}_\ell\}$ be the set of quantization points, with $0 = \tilde{q}_0 < \tilde{q}_1 < \ldots, \tilde{q}_\ell < \tilde{q}_{\ell+1} < \ldots$. Note that we restrict to a symmetric quantizer since the error metric is symmetric around 0 (the detailed proof on the optimality of symmetric quantizers is omitted due to space constraints). We then aim to solve

$$\begin{aligned} \max_{\delta \geq 0, \mathcal{Q}} \quad & \delta \\ \text{s.t.} \quad & |\mathcal{Q}(x) - x| \leq \eta + \omega x, \ \forall x \in [0, \delta], \end{aligned} \tag{3.45}$$

where the constraint (3.44) is imposed only to $x \in [0, \delta]$ since the quantizer is symmetric around 0. Since the quantization error in (3.45) is measured in Euclidean distance, it is optimal to restrict the quantization points to $\mathbb{Q} \subset [-\delta, \delta]$ and to map the input to the nearest quantization point (ties may be resolved arbitrarily). Then, letting $\mathcal{X}_\ell = ((\tilde{q}_{\ell-1} + \tilde{q}_\ell)/2, (\tilde{q}_\ell + \tilde{q}_{\ell+1})/2]$, with $\tilde{q}_{-1} = 0$ and $\tilde{q}_{(N+1)/2} = 2\delta - \tilde{q}_{(N-1)/2}$, it follows that $[0, \delta] \equiv \cup_{\ell=0}^{(N-1)/2} \mathcal{X}_\ell$ and $\mathcal{Q}(x) = \tilde{q}_\ell, \forall x \in \mathcal{X}_\ell$. Therefore, the optimization problem (3.45) can be expressed equivalently as

$$\begin{aligned} \max_{\delta \geq 0, \tilde{\mathbf{q}}} \quad & \delta \\ \text{s.t.} \quad & (\tilde{q}_\ell - x)^2 \leq (\eta + \omega x)^2, \ \forall x \in \mathcal{X}_\ell, \quad \forall \ell = 0, 1, \ldots, (N-1)/2, \\ & 0 = \tilde{q}_0 \leq \cdots \leq \tilde{q}_{(N+1)/2} = 2\delta - \tilde{q}_{(N-1)/2}. \end{aligned}$$

Equivalently,

$$\begin{aligned} \max_{\delta \geq 0, \tilde{\mathbf{q}}} \quad & \delta \\ \text{s.t.} \quad & \max\nolimits_{x \in \mathcal{X}_\ell} (\tilde{q}_\ell - x)^2 - (\eta + \omega x)^2 \leq 0, \quad \forall \ell = 0, 1, \ldots, (N-1)/2, \\ & 0 = \tilde{q}_0 \leq \cdots \leq \tilde{q}_{(N+1)/2} = 2\delta - \tilde{q}_{(N-1)/2}, \end{aligned}$$

and solving the maximization with respect to $x \in \mathcal{X}_\ell$ (note that the quadratic function is convex in x, hence it is maximized at the margins of $\mathcal{X}_\ell$), we obtain

$$\begin{aligned} \max_{\delta \geq 0, \tilde{\mathbf{q}}} \quad & \tfrac{\tilde{q}_{(N-1)/2} + \tilde{q}_{(N+1)/2}}{2} \\ \text{s.t.} \quad & \tilde{q}_\ell \leq \tilde{q}_{\ell-1} \left(\tfrac{1+\omega}{1-\omega} \right) + \tfrac{2\eta}{1-\omega}, \ \forall \ell \in [(N+1)/2], \\ & 0 = \tilde{q}_0 < \tilde{q}_1 < \cdots < \tilde{q}_{(N+1)/2}. \end{aligned}$$

Solving this problem with respect to $\tilde{\mathbf{q}}$ yields $q_0 = 0$ and

$$q_\ell = q_{\ell-1}\Big(\frac{1+\omega}{1-\omega}\Big) + \frac{2\eta}{1-\omega}, \quad \forall \ell \geq 1.$$

Solving by induction, we obtain q_ℓ as in (3.13), $\delta(\eta, \omega, N)$ as in (3.12), and

$$\ell(x) = \operatorname{sgn}(x) \cdot \min\Big\{\ell \in \mathbb{Z}_+ : \frac{q_\ell + q_{\ell+1}}{2} \geq |x|\Big\},$$

yielding (3.14) after solving with the expression of q_ℓ. A similar technique can be proved for the case when N is even. In this case, the quantization points are given by

$$q_\ell = -q_{-\ell} = \frac{\eta}{\omega}\Big[\frac{(1+\omega)^\ell}{(1-\omega)^{\ell-1}} - 1\Big], \quad \forall \ell \geq 1,$$

and $\delta(\eta, \omega, N) = (q_{N/2} + q_{N/2+1})/2$, which concludes the proof.

Proof of Lemma 12

Using a similar technique as in App.3.8.2 when N is odd, using the fact that $[0, \delta] = \cup_{\ell \in [(N-1)/2]} [\tilde{q}_{\ell-1}, \tilde{q}_\ell]$ and $\delta = \tilde{q}_{N/2}$ it suffices to solve

$$\begin{aligned} \max_{\delta \geq 0, \tilde{\mathbf{q}}} \quad & q_{(N-1)/2} \\ \text{s.t.} \quad & \mathbb{E}[|\mathcal{Q}(x) - x|^2] \leq (\eta + \omega x)^2, \ \forall x \in [\tilde{q}_{\ell-1}, \tilde{q}_\ell], \quad \forall \ell = [(N-1)/2], \\ & \mathbb{E}[\mathcal{Q}(x)] = x, \quad 0 = \tilde{q}_0 \leq \cdots \leq \tilde{q}_{(N-1)/2} = \delta. \end{aligned}$$

Furthermore, since $x \in [\tilde{q}_{\ell-1}, \tilde{q}_\ell]$ is mapped to $\tilde{q}_{\ell-1}$ w.p. $(\tilde{q}_\ell - x)/(\tilde{q}_\ell - \tilde{q}_{\ell-1})$ and to $\tilde{q}_\ell$ w.p. $(x - \tilde{q}_{\ell-1})/(\tilde{q}_\ell - \tilde{q}_{\ell-1})$ to satisfy $\mathbb{E}[\mathcal{Q}(x)] = x$, the problem can be expressed equivalently as

$$\begin{aligned} \max_{\delta \geq 0, \tilde{\mathbf{q}}} \quad & q_{(N-1)/2} \\ \text{s.t.} \quad & (x - \tilde{q}_\ell)(x - \tilde{q}_{\ell-1}) + (\eta + \omega x)^2 \geq 0, \quad \forall x \in [\tilde{q}_{\ell-1}, \tilde{q}_\ell], \quad \forall \ell = [(N-1)/2], \\ & 0 = \tilde{q}_0 \leq \cdots \leq \tilde{q}_{(N-1)/2} = \delta, \end{aligned}$$

or equivalently

$$\begin{aligned} \max_{\delta\geq 0,\tilde{\mathbf{q}}} \quad & q_{(N-1)/2} \\ \text{s.t.} \quad & \min_{x\in[\tilde{q}_{\ell-1},\tilde{q}_{\ell}]}(x-\tilde{q}_{\ell})(x-\tilde{q}_{\ell-1})+(\eta+\omega x)^2 \geq 0, \quad \forall \ell = [(N-1)/2], \\ & 0=\tilde{q}_0\leq\cdots\leq\tilde{q}_{(N-1)/2}=\delta. \end{aligned}$$

Solving the minimization over $x\in[\tilde{q}_{\ell-1},\tilde{q}_{\ell}]$ and solving with respect to $\tilde{\mathbf{q}}$ yields the following optimal quantization points: $q_0=0$ and

$$q_\ell = q_{\ell-1}(\sqrt{1+(\omega)^2}+\omega)^2+2\eta(\sqrt{1+(\omega)^2}+\omega), \quad \forall \ell\geq 1.$$

Solving by induction, we obtain q_ℓ as in (3.16), $\delta(\eta,\omega,N)$ as in (3.15), and the probabilistic qantization rule as in (3.17), with ℓ given by

$$\ell = \operatorname{sgn}(x)\cdot\min\Big\{\ell\in\mathbb{Z}_+ : q_\ell\geq|x|\Big\},$$

yielding (3.17) after solving with the expression of q_ℓ.

A similar technique can be proved for the case when N is even. In this case, the quantization points are given by

$$q_\ell = \frac{\eta}{\omega}\left[\frac{(\sqrt{1+(\omega)^2}+\omega)^{2\ell-1}}{\sqrt{1+(\omega)^2}}-1\right], \quad \forall\ell\geq 1.$$

and $\delta(\eta,\omega,N)=q_{N/2}$, which concludes the proof.

Proof of Corollary 2

Let $\mathcal{Q}(\mathbf{x})$ be a generic deterministic or probabilistic quantizer with domain $[-\delta,\delta]^d$ and codomain $\mathbb{Q}\in\mathbb{R}^d$ with $|\mathbb{Q}|<\infty$, that satisfies the BC-rule with $\eta=0$. It follows that

$$\omega\|\mathbf{x}\|_2 \geq \sqrt{\mathbb{E}[\|\mathcal{Q}(\mathbf{x})-\mathbf{x}\|_2^2]} \geq \min_{\mathbf{q}\in\mathbb{Q}}\|\mathbf{q}-\mathbf{x}\|_2 = \|\mathcal{Q}_{\text{det}}(\mathbf{x})-\mathbf{x}\|_2, \forall\mathbf{x}\in[-\delta,\delta]^d, \tag{3.46}$$

where the lower bound is achievable by a deterministic quantizer that maps $\mathbf{x}$ to the nearest quantization point, denoted as $\mathcal{Q}_{\text{det}}(\mathbf{x})$. Let $\mathcal{Q}_n(x) = \mathbf{e}_n^\top \mathcal{Q}_{\text{det}}(x\mathbf{e}_n)$ be the projection of $\mathcal{Q}_{\text{det}}$ on its nth element, where $\mathbf{e}_n$ is the nth canonical vector. Since $\|\mathcal{Q}_{\text{det}}(x\mathbf{e}_n) - x\mathbf{e}_n\|_2 \geq |\mathcal{Q}_n(x) - x|$, from (3.46) it follows that

$$\omega\|x\mathbf{e}_n\|_2 = \omega|x| \geq \|\mathcal{Q}_{\text{det}}(x\mathbf{e}_n) - x\mathbf{e}_n\|_2 \geq |\mathcal{Q}_n(x) - x|, \quad \forall x \in [-\delta, \delta],$$

hence $\mathcal{Q}_n$ satisfies the BC-rule with $\eta = 0$ as well. Note that $\mathcal{Q}_n$ is a scalar quantizer with $N_n \leq |\mathbb{Q}|$ quantization points. However, L.11 dictates that $\delta = 0$ for this quantizer, hence the contradiction, and we have proved the statement for both deterministic and probabilistic compression rules.

3.8.3 Proof of Theorems 3.5.1 and 3.5.2

Proof of Theorem 3.5.1

We first present some preliminary results instrumental to prove T.3.5.1, whose proofs are deferred to the end of this section.

The idea of the proof is to study the asymptotic behavior of an upper bound on the number of bits required per iteration, provided in the following lemma.

Lemma 18. *Under the same setting as T.3.3.1, and the proposed ANQ satisfying the BC rule, the average number of bits required per agent at the kth iteration, B^k, is upper bounded as*

$$\mathbb{E}[B^k] \leq \log_2(S+1)\left[3dR + dR\log_S\left(3 + \frac{\bar{F}^R(\sigma,\omega,\eta^0)}{\sqrt{md}\eta^0}\right)\right] \quad \text{bits}, \ \forall k \in \mathbb{Z}_+, \tag{3.47}$$

where $\bar{F}^R(\sigma,\omega,\eta^0)$ is defined in (3.39).

In addition, we need the following lemma to connect the asymptotic results of the logarithmic function and its argument.

Lemma 19. *For positive functions f, g, it holds: $\ln f(x) = \mathcal{O}(\ln g(x))$ as $x \to x_0$ if $\liminf_{x\to x_0} g(x) > 1$, and $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$.*

We are now ready to prove the main theorem. From L.18, the average number of bits per agent per iteration is upper bounded by

$$\mathbb{E}[B_k] \leq \log_2(S+1)\left[3dR + dR\log_S\left(3 + \frac{\bar{F}^R(\sigma,\omega,\eta^0)}{\sqrt{md}\eta^0}\right)\right], \qquad \forall k \in \mathbb{Z}_+,$$

where

$$\bar{F}^R(\sigma,\omega,\eta^0) = a_1 + a_2 V_0,$$

a_1 and a_2 are defined in (3.40) and (3.41), respectively; and V_0 is given in (3.35). We want to prove that this is $\mathbb{E}[B_k] = \mathcal{O}(d\ln(1+\frac{1}{\sigma(\sigma-\lambda)}))$ under A.9 and conditions $\eta^0 = \Theta(L_Z(\sigma-\lambda)), 1-\omega/\bar{\omega}(\sigma) = \Omega(1)$. Using the fact that $\bar{F}^R(\sigma,\omega,\eta^0)/\eta^0 = a_1/\eta^0 + a_2V_0/\eta^0$, it is sufficient to show that $a_1/(\sqrt{md}\eta^0) = \mathcal{O}(1+\frac{1}{\sigma})$ and $a_2V_0/(\sqrt{md}\eta^0) = \mathcal{O}(1+\frac{1}{\sigma(\sigma-\lambda)})$. In fact, using $R/\sigma \leq 1/\bar{\omega}(\sigma)$, we can bound a_1 and a_2 as

$$\begin{aligned} a_1/(\sqrt{md}\eta^0) &\leq \frac{1}{\sigma}(1+L_C\sigma)\max\{1,(2L_C)^{R-1}\}\frac{R}{1-\omega/\bar{\omega}(\sigma)}, \\ a_2 &\leq \frac{2L_Z}{\sigma}\max\{1,(2L_C)^{R-1}\}\frac{R}{1-\omega/\bar{\omega}(\sigma)}. \end{aligned}$$

Clearly, $a_1/(\sqrt{md}\eta^0) = \mathcal{O}(1+\frac{1}{\sigma})$ and $a_2 = \mathcal{O}(\frac{L_Z}{\sigma})$ since $L_C, R = \mathcal{O}(1)$ and $1-\omega/\bar{\omega}(\sigma) = \Omega(1)$. We next study V_0. From its expression in (3.35), we notice that $V_0 = \mathcal{O}(\sqrt{md})$ since $\omega/\sigma \leq 1$, $\max\{c^*, \|\mathbf{z}^0 - \mathbf{z}^\infty\|_2\} = \mathcal{O}(\sqrt{md})$, $\eta^0 = \Theta(L_Z(\sigma-\lambda))$, $L_AL_Z, L_C, R = \mathcal{O}(1)$, and $1-\omega/\bar{\omega} = \Omega(1)$. Therefore, it follows that $a_2V_0/(\sqrt{md}\eta^0) = \mathcal{O}(\frac{1}{\sigma(\sigma-\lambda)}) = \mathcal{O}(1+\frac{1}{\sigma(\sigma-\lambda)})$, and the proof is completed by invoking L.19.

Proof of Lemma 18

Let $\Delta\mathbf{c}_i^{k,s} \triangleq \mathbf{c}_i^{k,s} - \hat{\mathbf{c}}_i^{k-1,s}$ be the input to the quantizer for agent i, at iteration k and communication round s. We now study the average number of bits required for 1) the deterministic quantizer, and 2) the probabilistic quantizer with $\mathbb{E}[\mathcal{Q}(x)] = x$.

i) Deterministic quantizer: The average number of bits required is bounded as (see (3.19), one can also verify that the following also holds for even N)

$$b_i^{k,s} \le \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{\ln\left(1+\frac{\omega\|\Delta\mathbf{c}_i^{k,s}\|_2}{\sqrt{d}\eta^0\cdot(\sigma)^k}\right)}{\ln(1+\omega)-\ln(1-\omega)}\right)\right] \quad \text{bits.}$$

We now upper bound the argument inside the second logarithm. Since it is a decreasing function of ω, it is maximized in the limit $\omega \to 0$, yielding

$$\frac{\ln\left(1+\frac{\omega\|\Delta\mathbf{c}_i^{k,s}\|_2}{\sqrt{d}\eta^0\cdot(\sigma)^k}\right)}{\ln(1+\omega)-\ln(1-\omega)} \le \frac{\|\Delta\mathbf{c}_i^{k,s}\|_2}{2\sqrt{d}\eta^0\cdot(\sigma)^k}.$$

With this upper bound, we can then upper bound the average number of bits per agent at communication round s, iteration k, as

$$\begin{aligned}\mathbb{E}[b^{k,s}] \triangleq \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[b_i^{k,s}] &\overset{(a)}{\le} \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{\sqrt{\mathbb{E}[\|\mathbf{c}^{k,s} - \hat{\mathbf{c}}^{k-1,s}\|_2^2]}}{2\sqrt{md}\eta^0\cdot(\sigma)^k}\right)\right]\\ &\overset{(b)}{\le} \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{F^s(\sigma,\omega,\eta^0)}{2\sqrt{md}\eta^0}\right)\right]\\ &\overset{(c)}{\le} \log_2(S+1)\left[3d + d\log_S\left(2 + \frac{\bar{F}^R(\sigma,\omega,\eta^0)}{2\sqrt{md}\eta^0}\right)\right],\end{aligned}$$

where (a) follows from Cauchy–Schwarz inequality, Jensen's inequality, and the definition of $\Delta\mathbf{c}^{k,s}$; (b) follows from (3.30); and (c) follows from $F^s \le \bar{F}^R$ (see (3.39)).

ii) Probabilistic quantizer with $\mathbb{E}[\mathcal{Q}(x)] = x$: Using the same technique as in i), along with the inequality $1 - 1/x \le \ln(x) \le x - 1$ for $x > 1$ to bound the argument inside the second logarithm of (3.20),

$$\mathbb{E}[b^{k,s}] \le \log_2(S+1)\left[3d + d\log_S\left(3 + \frac{\bar{F}^R(\sigma,\omega,\eta^0)}{\sqrt{md}\eta^0}\right)\right].$$

One can also verify that it also holds for even N.

Finally, for both the deterministic and probabilistic cases, the proof is completed by summing over $s \in [R]$ to get the average communication cost per agent at iteration k.

Proof of Lemma 19

If $f(x) = \mathcal{O}(g(x))$ as $x \to x_0$ and $\liminf_{x\to x_0} g(x) > 1$, then

$$\limsup_{x\to x_0}\left|\frac{\ln f(x)}{\ln g(x)}\right| \le 1 + \limsup_{x\to x_0}\left|\frac{\ln(f(x)/g(x))}{\ln g(x)}\right| \overset{(a)}{\le} 1 + \frac{\limsup_{x\to x_0}|\ln(f(x)/g(x))|}{|\liminf_{x\to x_0}\ln g(x)|} < \infty,$$

where (a) follows from $\liminf_{x\to x_0} g(x) > 1$. This completes the proof.

Proof of Theorem 3.5.2

Since $\sqrt{\mathbb{E}[\|\mathbf{z}^k - \mathbf{z}^\infty\|^2]} \le V_0 \cdot (\sigma)^k$ (cf. T.3.3.1), the ε-accuracy is achieved if $k[-\ln(\sigma)] \ge \ln(V_0/\sqrt{m\varepsilon})$, which yields $k(1-\sigma) \ge \ln(V_0/\sqrt{m\varepsilon})$ since $-\ln(\sigma) \ge 1-\sigma$. Hence, ε-accuracy is achieved if all conditions in T.3.3.1 hold and $k \ge k_\varepsilon \triangleq \left\lceil \frac{1}{1-\sigma}\ln\frac{V_0}{\sqrt{m\varepsilon}}\right\rceil$. Hence, to compute the upper bound of the communication cost $\sum_{k=0}^{k_\varepsilon-1}\mathbb{E}[B^k]$, we need the upper bounds for $\mathbb{E}[B^k]$, $\frac{1}{1-\sigma}$ and V_0. In the proof of T.3.5.1, we found that, under A.9 and the conditions $1-\omega/\bar{\omega}(\sigma) = \Omega(1), \eta^0 = \Theta(L_Z(\sigma-\lambda))$,

$$\mathbb{E}[B^k] = \mathcal{O}\Big(d\log_2\Big(1+\frac{1}{\sigma(\sigma-\lambda)}\Big)\Big),\ V_0 = \mathcal{O}(\sqrt{md}).$$

Moreover, it can be shown that $1/\sigma \le \frac{(1-\lambda)^2}{(1-\sigma)(\sigma-\lambda)}$ for $\sigma \in (\lambda, 1)$, and therefore

$$\frac{1}{\sigma(\sigma-\lambda)} \le \frac{1}{(1-\lambda)}\left[\frac{(1-\lambda)^2}{(1-\sigma)(\sigma-\lambda)}\right]^2\frac{1-\sigma}{1-\lambda} = \mathcal{O}\Big(\frac{1}{1-\lambda}\Big),$$

where we used $\frac{(1-\lambda)^2}{(1-\sigma)(\sigma-\lambda)} = \mathcal{O}(1)$. It then follows from L.19 that $\mathbb{E}[B^k] = \mathcal{O}(d\log_2(1+\frac{1}{1-\lambda}))$. On the other hand, we can bound k_ε as

$$k_\varepsilon \le \frac{1}{1-\lambda}\Big(1-\lambda+\frac{1-\lambda}{1-\sigma}\log_2\frac{V_0}{\sqrt{m\varepsilon}}\Big) = \mathcal{O}\Big(\frac{1}{1-\lambda}\log_2\Big(\frac{d}{\varepsilon}\Big)\Big),$$

since $\frac{1-\lambda}{1-\sigma} = \mathcal{O}(1)$ and $V_0 = \mathcal{O}(\sqrt{md})$, Therefore, the communication cost satisfies

$$\sum_{k=0}^{k_\varepsilon-1}\mathbb{E}[B^k] = \mathcal{O}\Big(dk_\varepsilon\log_2\Big(1+\frac{1}{1-\lambda}\Big)\Big) = \mathcal{O}\Big(\frac{d}{1-\lambda}\log_2\Big(\frac{d}{\varepsilon}\Big)\log_2\Big(1+\frac{1}{1-\lambda}\Big)\Big),$$

which completes the proof.

Proof of Lemma 13

Consider $\ell \geq 0$. Using (3.18), the number of information symbols required to encode ℓ is

$$b_\ell^* = \min\left\{b \in \mathbb{Z}_+ : \ell \leq \left\lfloor \frac{(S)^{b+1} - 1}{2(S-1)} \right\rfloor\right\}.$$

Similarly, for $\ell < 0$,

$$b_\ell^* = \min\left\{b \in \mathbb{Z}_+ : -\ell \leq \left\lceil \frac{(S)^{b+1} - 1}{2(S-1)} \right\rceil - 1\right\}.$$

Since $\lfloor \frac{(S)^{b+1}-1}{2(S-1)} \rfloor \geq \lceil \frac{(S)^{b+1}-1}{2(S-1)} \rceil - 1 \geq \frac{(S)^{b+1}-1}{2(S-1)} - 1$, we can then upper bound $b_\ell, \ell \in \mathbb{Z}$, as

$$\begin{aligned} b_\ell^* &\leq \min\{b \geq 1 : 1 + 2(S-1)(1 + |\ell|) \leq (S)^{b+1}\} \\ &= \min\{b \geq 1 : b \geq \log_S(2 - 1/S + 2(1 - 1/S)|\ell|)\} = \lceil \log_S(2 - 1/S + 2(1 - 1/S)|\ell|)\rceil. \end{aligned}$$

Using $\lceil x \rceil \leq x + 1$ We can then further upper bound

$$b_\ell^* \leq 1 + \log_S(2 - 1/S + 2(1 - 1/S)|\ell|) \leq \log_S(2S + 2S|\ell|) \leq 2 + \log_S(1 + |\ell|),$$

resulting the upper bound to the communication cost (including the termination symbol)

$$\bar{C}_{\text{comm}}(\ell) \leq 3 + \log_S(1 + |\ell|) \quad \text{symbols.} \tag{3.48}$$

Let $\mathbf{x} = (x_n)_{n=1}^d$. We now study the result with 1) the deterministic quantizer, and 2) the probabilistic quantizer with $\mathbb{E}[\mathcal{Q}(x)] = x$.

Note that for the deterministic and probabilistic quantizers, we can express $\ell(x)$ as

$$|\ell(x)| \leq \lceil c_1 + c_2 \ln(1 + \omega/\eta|x|)\rceil,$$

for some $c_1 \leq 0$, $c_2 > 0$ (see (3.14) and (3.17) for a closed-form expression of c_1 and c_2). Invoking (3.48) and $\bar{C}_{\text{comm}}(\mathbf{x}) = \sum_{n=1}^{d} \bar{C}_{\text{comm}}(\ell_n)$ yields

$$\begin{aligned} C(\mathbf{x}) &\leq 3d + \sum_{n=1}^{d} \log_S \left(1 + \left\lceil c_1 + c_2 \ln\left(1 + \frac{\omega |x_n|}{\eta}\right)\right\rceil\right) \\ &\overset{(a)}{\leq} 3d + d \log_S \left(2 + c_1 + c_2 \ln\left(1 + \frac{\omega \|\mathbf{x}\|_2}{\sqrt{d}\eta}\right)\right) \quad \text{symbols} \\ &= \log_2(S+1)\left[3d + d \log_S \left(2 + c_1 + c_2 \ln\left(1 + \frac{\omega \|\mathbf{x}\|_2}{\sqrt{d}\eta}\right)\right)\right] \quad \text{bits,} \end{aligned} \tag{3.49}$$

where (a) follows from $\lceil x \rceil \leq x + 1$, Jensen's inequality and Cauchy–Schwarz inequality, in order. Invoking the expressions of c_1 and c_2 from (3.14) and (3.17), respectively, yield the result for the deterministic and probabilistic quantizers with odd N. Similar techniques can be used to find $\ell(x)$ and thus the result for quantizers with even N.

3.8.4 Examples of (M)

In this section, we will show that (M) contains a gamut of distributed algorithms, corresponding to different choices of $R, \mathcal{C}_i^s$, and $\mathcal{A}_i$. Given (P), we will assume that each f_i is L-smooth and μ-strongly convex.

Every distributed algorithm on mesh networks we will describe below alternates one step of optimization with possibly multiple rounds of communications. In each communication round, every agent i combines linearly the signal received by its neighbors using weights $(w_{ij})_{j \in \mathcal{N}_i}$; let $\mathbf{W} = (w_{ij})_{i,j=1}^{m}$. Consistently with the undirected graph $\mathcal{G}$, we will tacitly assume that $\mathbf{W}$ is symmetric and doubly stochastic, i.e., $\mathbf{W} = \mathbf{W}^\top$ and $\mathbf{W}\mathbf{1} = \mathbf{1}$, with $w_{ij} > 0$ if $(j,i) \in \mathcal{E}$, and $w_{ij} = 0$ otherwise. We assume that the eigenvalues of $\mathbf{W}$ are in $[\nu, 1]$, with $\nu > 0$.[6] Note that this condition can be achieved by design: in fact, given a doubly stochastic weight matrix $\tilde{\mathbf{W}} = (\tilde{w}_{ij})_{i,j \in [m]}$, each agent i can set $w_{ii} = [(1+\nu) + (1-\nu)\tilde{w}_{ii}]/2$ and $w_{ij} = (1-\nu)\tilde{w}_{ij}/2, \forall j \neq i$ for a design parameter $\nu \in (0,1]$. Note that, for any given $\mathbf{z}^0$ and $\mathbf{z}^\infty$ with bounded entries, it holds $\|\mathbf{z}^0 - \mathbf{z}^\infty\|_2 = \mathcal{O}(\sqrt{md})$.

[6] ↑This assumption is also required in [80] for prox-EXTRA, prox-NEXT, prox-DIGing, and prox-NIDS for achieving $\|\mathbf{z}^k - \mathbf{z}^\infty\| = \mathcal{O}(\sqrt{md}(\lambda)^k)$.

Finally, in the rest of this section, we will adopt the following notations: $\mathbf{x}^k = (\mathbf{x}_i^k)_{i=1}^m, \mathbf{y}^k = (\mathbf{y}_i^k)_{i=1}^m, \mathbf{w}^k = (\mathbf{w}_i^k)_{i=1}^m, \mathbf{x} = (\mathbf{x}_i)_{i=1}^m, \mathbf{y} = (\mathbf{y}_i)_{i=1}^m$ and $\mathbf{w} = (\mathbf{w}_i)_{i=1}^m$, $\hat{\mathbf{W}} = \mathbf{W} \otimes \mathbf{I}_d$, and $\mathbf{G}^\dagger$ is the pseudo-inverse of matrix $\mathbf{G}$. Given $\mathbf{x}^k = (\mathbf{x}_i^k)_{i=1}^m$, we also define $\nabla f(\mathbf{x}^k) \triangleq (\nabla f_i(\mathbf{x}_i^k))_{i=1}^m$. For any function $g : \mathbb{R}^d \to \mathbb{R}$ and positive semi-definite matrix $\mathbf{G}$, define $\|\mathbf{x}\|_{\mathbf{G}} \triangleq \sqrt{\mathbf{x}^\top \mathbf{G}\mathbf{x}}$ and

$$\operatorname{prox}_{\mathbf{G},g}(\mathbf{x}) \triangleq \underset{\mathbf{z}\in\mathbb{R}^d}{\arg\min} \quad g(\mathbf{z}) + \frac{1}{2}\|\mathbf{z}-\mathbf{x}\|^2_{\mathbf{G}^{-1}}.$$

(Prox-)EXTRA [80]

The update of prox-EXTRA solving (P) reads

$$\begin{aligned}
\mathbf{x}^k &= \operatorname{prox}_{\gamma\mathbf{I},r}(\mathbf{w}^k),\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \left(\mathbf{I} - \hat{\mathbf{W}}\right)\mathbf{w}^{k+1},\\
\mathbf{w}^{k+1} &= \hat{\mathbf{W}}\mathbf{x}^k - \gamma\nabla f(\mathbf{x}^k) - \mathbf{y}^k,
\end{aligned}$$

with $\mathbf{y}^0 = \mathbf{0}$ and $\mathbf{w}^0 \in \mathbb{R}^{md}$.

Prox-EXTRA can be cast as (M) with $R = 2$ rounds of communications with

$$\begin{aligned}
\mathbf{z}^\top &= [\mathbf{y}^\top, \mathbf{w}^\top],\\
\hat{\mathbf{c}}_i^{k,1} &= \mathcal{C}_i^1\left(\mathbf{z}_i^k, \mathbf{0}\right) = \operatorname{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k), \quad (3.50)\\
\hat{\mathbf{c}}_i^{k,2} &= \mathcal{C}_i^2\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}\right) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1} - \gamma\nabla f_i(\hat{\mathbf{c}}_i^{k,1}) - \mathbf{y}_i^k, \quad (3.51)\\
\mathbf{z}_i^{k+1} &= \mathcal{A}_i\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}\right) = \begin{bmatrix} \mathbf{y}_i^k + \hat{\mathbf{c}}_i^{k,2} - \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,2} \\ \hat{\mathbf{c}}_i^{k,2} \end{bmatrix}. \quad (3.52)
\end{aligned}$$

We show next that the above instance of (M) satisfies A.6-9.

• **On A.6:** Using [80] it is not difficult to check that, if $\gamma = \frac{2\rho_m(\mathbf{W})}{L+\mu\rho_m(\mathbf{W})}$, then prox-EXTRA satisfies A.6, with some $\lambda < 1$ and the norm $\|\bullet\|$ defined as

$$\|\mathbf{z}\|^2 = \mathbf{w}^\top\hat{\mathbf{W}}^{-1}\mathbf{w} + \mathbf{y}^\top(\mathbf{I}-\hat{\mathbf{W}})^\dagger\mathbf{y}.$$

Note that $\|\mathbf{z}\|^2 \geq \|\mathbf{z}\|_2^2$, due to $\rho_i(\mathbf{W}) \in [\nu, 1]$, $i \in [m]$.

- **On A.7-9:** Based on (3.50), (3.51) and (3.52), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2) = \begin{bmatrix} \mathbf{y} + (\mathbf{I} - \hat{\mathbf{W}})\mathbf{c}^2 \\ \mathbf{c}^2 \end{bmatrix},$$

$$\mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}), \quad \text{and} \quad \mathcal{C}^2(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c} - \gamma \nabla f(\mathbf{c}) - \mathbf{y},$$

respectively; and $\mathcal{Z} = \operatorname{span}(\mathbf{I} - \hat{\mathbf{W}}) \times \mathbb{R}^{md}$, where we defined (with a slight abuse of notation) $\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}) = (\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}_i))_{i=1}^m$.

We show that prox-EXTRA satisfies A.9. Note that

$$\begin{aligned} \left\| \mathcal{A}\left(\mathbf{z}, \mathbf{c}^1, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}^1, \mathbf{c}\right) \right\|^2 &= \|\sqrt{\hat{\mathbf{W}}^{-1}}(\mathbf{c} - \mathbf{c})\|_2^2 + \left\| \sqrt{\mathbf{I} - \hat{\mathbf{W}}}(\mathbf{c} - \mathbf{c}) \right\|_2^2 \\ &\leq (1 + \nu^{-1}) \|\mathbf{c} - \mathbf{c}\|_2^2, \end{aligned}$$

and $\mathcal{A}\left(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2\right)$ is constant with respect to $\mathbf{c}^1$, hence A.7 holds with $L_A = \sqrt{1 + \nu^{-1}}$.

We next derive L_C and L_Z. Since the proximal mapping is non-expansive [88], it follows that

$$\begin{aligned} \left\| \mathcal{C}^1(\mathbf{z}, \mathbf{0}) - \mathcal{C}^1(\mathbf{z}, \mathbf{0}) \right\|_2 &= \|\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}) - \operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})\|_2 \leq \|\mathbf{w} - \mathbf{w}\|_2 \leq \|\mathbf{z} - \mathbf{z}\|_2, \\ \left\| \mathcal{C}^2(\mathbf{z}, \mathbf{c}) - \mathcal{C}^2(\mathbf{z}, \mathbf{c}) \right\|_2 &= \left\| \hat{\mathbf{W}}(\mathbf{c} - \mathbf{c}) - \gamma(\nabla f(\mathbf{c}) - \nabla f(\mathbf{c})) - (\mathbf{y} - \mathbf{y}) \right\|_2 \\ &\leq \|\mathbf{c} - \mathbf{c}\|_2 + \gamma L \|\mathbf{c} - \mathbf{c}\|_2 + \|\mathbf{z} - \mathbf{z}\|_2, \end{aligned}$$

A.8 holds with $L_C = 1 + \gamma L$ and $L_Z = 1$. Since $\gamma = \mathcal{O}(1/L)$, it follows that $L_C = \mathcal{O}(1)$. For the initial conditions, we have

$$\begin{aligned} \|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 = \|\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}^0)\|_2 &\overset{(a)}{\leq} \|\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}^0) - \operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}^\infty)\|_2 + \|\operatorname{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}^\infty)\|_2 \\ &\overset{(b)}{\leq} \|\mathbf{w}^0 - \mathbf{w}^\infty\|_2 + \|\mathbf{w}^\infty\|_2 = \mathcal{O}(\sqrt{md}), \\ \|\mathcal{C}^2(\mathbf{z}^0, \mathbf{0})\|_2 = \|\mathbf{y}^0\|_2 \leq \|\mathbf{y}^0 - \mathbf{y}^\infty\|_2 + \|\mathbf{y}^\infty\|_2 = \mathcal{O}(\sqrt{md}), \end{aligned}$$

where (a) follows from the triangle inequality; and (b) follows from the non-expansive property of the proximal mapping and $\mathbf{w}^\infty$ is a fixed point of the proximal mapping [80].

Therefore, $L_A L_Z = \mathcal{O}(1), L_C = \mathcal{O}(1)$, $\|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$ and $\|\mathcal{C}^2(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$; hence A.9 holds.

(Prox-)NEXT [80]

The update of prox-NEXT solving (P) reads

$$\begin{aligned}
\mathbf{x}^k &= \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}^k),\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + (\mathbf{I} - \hat{\mathbf{W}})^2\mathbf{w}^{k+1},\\
\mathbf{w}^{k+1} &= \hat{\mathbf{W}}^2\Big(\mathbf{x}^k - \gamma\nabla f(\mathbf{x}^k)\Big) - \mathbf{y}^k,
\end{aligned}$$

with $\mathbf{y}^0 = \mathbf{0}$ and $\mathbf{w}^0 \in \mathbb{R}^{md}$.

Prox-NEXT can be cast as (M) with $R = 4$ rounds of communications, using the following definitions:

$$\begin{aligned}
\mathbf{z}^\top &= [\mathbf{y}^\top, \mathbf{w}^\top],\\
\hat{\mathbf{c}}_i^{k,1} &= \mathcal{C}_i^1\Big(\mathbf{z}_i^k, \mathbf{0}\Big) = \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k) - \gamma\nabla f_i(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k)), && (3.53)\\
\hat{\mathbf{c}}_i^{k,2} &= \mathcal{C}_i^2\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}\Big) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1}, && (3.54)\\
\hat{\mathbf{c}}_i^{k,3} &= \mathcal{C}_i^3\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}\Big) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,2} - \mathbf{y}_i^k, && (3.55)\\
\hat{\mathbf{c}}_i^{k,4} &= \mathcal{C}_i^4\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,3}\Big) = \sum_{j\in\mathcal{N}_i} w_{ij}\Big(\hat{\mathbf{c}}_i^{k,3} - \hat{\mathbf{c}}_j^{k,3}\Big), && (3.56)\\
\mathbf{z}_i^{k+1} &= \mathcal{A}_i\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,3}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,4}\Big) = \begin{bmatrix} \mathbf{y}_i^k + \sum_{j\in\mathcal{N}_i} w_{ij}\Big(\hat{\mathbf{c}}_i^{k,4} - \hat{\mathbf{c}}_j^{k,4}\Big) \\ \hat{\mathbf{c}}_i^{k,3} \end{bmatrix}. && (3.57)
\end{aligned}$$

We show next that the above instance of (M) satisfies A.6-9.

• **On A.6:** Using [80] it is not difficult to check that, if $\gamma = 2/(\mu + L)$, then prox-NEXT satisfies A.6, with $\lambda < 1$ and the norm $\| \bullet \|$ defined as

$$\|\mathbf{z}\|^2 = \left\|\hat{\mathbf{W}}^{-2}\mathbf{w}\right\|^2_{\mathbf{I}-(\mathbf{I}-\hat{\mathbf{W}})^2} + \mathbf{y}\Big(\hat{\mathbf{W}}^{-2}\Big((\mathbf{I}-\hat{\mathbf{W}})^2\Big)^{\dagger}\hat{\mathbf{W}}^{-2}\Big)\mathbf{y}.$$

Again, note that $\|\mathbf{z}\|^2 \geq \|\mathbf{z}\|_2^2$.

• **On A.7-9:** Based on (3.53)-(3.57), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\begin{aligned}
&\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4) = \begin{bmatrix} \mathbf{y} + (\mathbf{I} - \hat{\mathbf{W}})\mathbf{c}^4 \\ \mathbf{c}^3 \end{bmatrix}, \\
&\mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}) - \gamma\nabla f(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})), \quad \mathcal{C}^2(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c}, \\
&\mathcal{C}^3(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c} - \mathbf{y}, \quad \text{and} \quad \mathcal{C}^4(\mathbf{z}, \mathbf{c}) = (\mathbf{I} - \hat{\mathbf{W}})\mathbf{c},
\end{aligned}$$

respectively; and $\mathcal{Z} = \mathrm{span}(\mathbf{I} - \hat{\mathbf{W}}) \times \mathbb{R}^{md}$.

We now show that prox-NEXT satisfies A.9. Note that

$$\begin{aligned}
&\|\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}, \mathbf{c}^4) - \mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}, \mathbf{c}^4)\|^2 \\
&= (\mathbf{c} - \mathbf{c})^{\top}\hat{\mathbf{W}}^{-4}[\mathbf{I} - (\mathbf{I} - \hat{\mathbf{W}})^2](\mathbf{c} - \mathbf{c}) \leq \nu^{-2}(2\nu^{-1} - 1)\|\mathbf{c} - \mathbf{c}\|_2^2
\end{aligned}$$

and

$$\left\|\mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}\right)\right\|^2 = \|\hat{\mathbf{W}}^{-2}(\mathbf{c} - \mathbf{c})\|_2^2 \leq \nu^{-4}\|\mathbf{c} - \mathbf{c}\|_2^2.$$

Moreover, $\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4)$ is constant with respect to $\mathbf{c}^1, \mathbf{c}^2$. Therefore, A.7 holds with $L_A = \sqrt{2}\nu^{-2}$.

We next derive L_C and L_Z. Using the non-expansive property, it follows that

$$\begin{aligned}
&\|\mathcal{C}^1(\mathbf{z}, \mathbf{0}) - \mathcal{C}^1(\mathbf{z}, \mathbf{0})\|_2 \\
&= \|(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}) - \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})) - \gamma(\nabla f(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})) - \nabla f(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})))\|_2 \\
&\leq (1 + \gamma L)\|\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}) - \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})\|_2 \leq (1 + \gamma L)\|\mathbf{w} - \mathbf{w}\|_2 \leq (1 + \gamma L)\|\mathbf{z} - \mathbf{z}\|_2,
\end{aligned}$$

$$
\begin{aligned}
&\|\mathcal{C}^2(\mathbf{c},\mathbf{z}) - \mathcal{C}^2(\mathbf{c},\mathbf{z})\|_2 = \|\hat{\mathbf{W}}(\mathbf{c}-\mathbf{c})\|_2 \leq \|\mathbf{c}-\mathbf{c}\|_2,\\
&\|\mathcal{C}^3(\mathbf{c},\mathbf{z}) - \mathcal{C}^3(\mathbf{c},\mathbf{z})\|_2 \leq \|\hat{\mathbf{W}}(\mathbf{c}-\mathbf{c})\|_2 + \|\mathbf{y}-\mathbf{y}\|_2 \leq \|\mathbf{c}-\mathbf{c}\|_2 + \|\mathbf{z}-\mathbf{z}\|_2,\\
&\|\mathcal{C}^4(\mathbf{c},\mathbf{z}) - \mathcal{C}^4(\mathbf{c},\mathbf{z})\|_2 = \|(\mathbf{I}-\hat{\mathbf{W}})(\mathbf{c}-\mathbf{c})\|_2 \leq \|\mathbf{c}-\mathbf{c}\|_2,
\end{aligned}
$$

which implies that A.8 holds with $L_C = 1$ and $L_Z = 1 + \gamma L$. Since $\gamma = \mathcal{O}(1/L)$, it follows that $L_Z = \mathcal{O}(1)$. For the initial conditions, we have $\|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 \leq (1+\gamma L)(\|\mathbf{w}^0 - \mathbf{w}^\infty\|_2 + \|\mathbf{w}^\infty\|_2) = \mathcal{O}(\sqrt{md})$ and $\|\mathcal{C}^3(\mathbf{z}^0, \mathbf{0})\|_2 \leq \|\mathbf{y}^0 - \mathbf{y}^\infty\|_2 + \|\mathbf{y}^\infty\|_2 = \mathcal{O}(\sqrt{md})$.

Therefore, $L_A L_Z = \mathcal{O}(1)$, $L_C = \mathcal{O}(1)$, $\|\mathcal{C}^1(\mathbf{z}^0,\mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$, $\|\mathcal{C}^2(\mathbf{z}^0,\mathbf{0})\|_2 = 0$, $\|\mathcal{C}^3(\mathbf{z}^0,\mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$, and $\|\mathcal{C}^4(\mathbf{z}^0,\mathbf{0})\|_2 = 0$; hence A.9 holds.

(Prox-)DIGing [80]

The update of prox-DIGing solving (P), reads

$$
\begin{aligned}
\mathbf{x}^k &= \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}^k),\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + (\mathbf{I}-\hat{\mathbf{W}})^2\mathbf{w}^{k+1},\\
\mathbf{w}^{k+1} &= \hat{\mathbf{W}}^2\mathbf{x}^k - \gamma\nabla f(\mathbf{x}^k) - \mathbf{y}^k,
\end{aligned}
$$

with $\mathbf{y}^0 = \mathbf{0}$ and $\mathbf{w}^0 \in \mathbb{R}^{md}$.

Prox-DIGing can be cast as (M) with $R = 4$ rounds of communications, using the following definitions:

$$
\begin{aligned}
\mathbf{z}^\top &= [\mathbf{y}^\top, \mathbf{w}^\top],\\
\hat{\mathbf{c}}_i^{k,1} &= \mathcal{C}_i^1\left(\mathbf{z}_i^k, \mathbf{0}\right) = \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k), &(3.58)\\
\hat{\mathbf{c}}_i^{k,2} &= \mathcal{C}_i^2\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}\right) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1}, &(3.59)\\
\hat{\mathbf{c}}_i^{k,3} &= \mathcal{C}_i^3\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}\right) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,2} - \gamma\nabla f_i(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k)) - \mathbf{y}_i^k, &(3.60)\\
\hat{\mathbf{c}}_i^{k,4} &= \mathcal{C}_i^4\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,3}\right) = \sum_{j\in\mathcal{N}_i} w_{ij}\left(\hat{\mathbf{c}}_j^{k,3} - \hat{\mathbf{c}}_i^{k,3}\right), &(3.61)
\end{aligned}
$$

$$\mathbf{z}_i^{k+1} = \mathcal{A}_i\Big(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,3}, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,4}\Big) = \begin{bmatrix} \mathbf{y}_i^k + \sum_{j\in\mathcal{N}_i} w_{ij}\Big(\hat{\mathbf{c}}_i^{k,4} - \hat{\mathbf{c}}_j^{k,4}\Big) \\ \hat{\mathbf{c}}_i^{k,3} \end{bmatrix}. \tag{3.62}$$

We show next that the above instance of (M) satisfies A.6-9.

- **On A.6:** Using [80] it is not difficult to check that, if $\gamma = \frac{2\rho_1(\mathbf{W})}{L+\mu\rho_1(\mathbf{W})}$, then prox-DIGing satisfies A.6, with $\lambda < 1$ and the norm $\|\bullet\|$ defined as

$$\|\mathbf{z}\|^2 = \frac{1}{2\nu - \nu^2}\Big(\|\mathbf{w}\|^2_{\mathbf{I}-(\mathbf{I}-\hat{\mathbf{W}})^2} + \mathbf{y}^\top \Big((\mathbf{I}-\hat{\mathbf{W}})^2\Big)^\dagger \mathbf{y}\Big).$$

Note that $\|\mathbf{z}\|^2 \geq \|\mathbf{z}\|_2^2$.

- **On A.7-9:** Based on (3.58)-(3.62), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4) = \begin{bmatrix} \mathbf{y} + (\mathbf{I}-\hat{\mathbf{W}})\mathbf{c}^4 \\ \mathbf{c}^3 \end{bmatrix},$$
$$\mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}), \quad \mathcal{C}^2(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c},$$
$$\mathcal{C}^3(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c} - \gamma\nabla f(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w})) - \mathbf{y}, \quad \text{and} \quad \mathcal{C}^4(\mathbf{z}, \mathbf{c}) = (\mathbf{I}-\hat{\mathbf{W}})\mathbf{c},$$

respectively; and $\mathcal{Z} = \mathrm{span}(\mathbf{I}-\hat{\mathbf{W}}) \times \mathbb{R}^{md}$. We now show that prox-DIGing satisfies A.9. Note that

$$\begin{aligned}\Big\|\mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}, \mathbf{c}^4\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}, \mathbf{c}^4\right)\Big\|^2 &= \frac{1}{2\nu-\nu^2}(\mathbf{c}-\mathbf{c})^\top[\mathbf{I}-(\mathbf{I}-\hat{\mathbf{W}})^2](\mathbf{c}-\mathbf{c}) \\ &\leq \frac{1}{2\nu-\nu^2}\|\mathbf{c}-\mathbf{c}\|_2^2,\end{aligned}$$

and

$$\Big\|\mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}\right)\Big\|^2 = \frac{1}{2\nu-\nu^2}\|\mathbf{c}-\mathbf{c}\|_2^2.$$

Moreover, $\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4)$ is constant with respect to $\mathbf{c}^1, \mathbf{c}^2$. Therefore, A.7 holds with $L_A = \sqrt{1/(2\nu-\nu^2)}$.

We next derive L_C and L_Z. We have

$$
\begin{aligned}
&\|\mathcal{C}^1(\mathbf{z},\mathbf{0})-\mathcal{C}^1(\mathbf{z},\mathbf{0})\|_2 \le (1+\gamma L)\|\mathbf{z}-\mathbf{z}\|_2,\\
&\|\mathcal{C}^2(\mathbf{z},\mathbf{c})-\mathcal{C}^2(\mathbf{z},\mathbf{c})\|_2 = \|\hat{\mathbf{W}}(\mathbf{c}-\mathbf{c})\|_2 \le \|\mathbf{c}-\mathbf{c}\|_2,\\
&\|\mathcal{C}^3(\mathbf{z},\mathbf{c})-\mathcal{C}^3(\mathbf{z},\mathbf{c})\|_2 \le \|\mathbf{c}-\mathbf{c}\|_2 + (1+\gamma L)\|\mathbf{z}-\mathbf{z}\|_2,\\
&\|\mathcal{C}^4(\mathbf{z},\mathbf{c})-\mathcal{C}^4(\mathbf{z},\mathbf{c})\|_2 = \|(\mathbf{I}-\hat{\mathbf{W}})\mathbf{c}-\mathbf{c}\|_2 \le \|\mathbf{c}-\mathbf{c}\|_2,
\end{aligned}
$$

which implies that A.8 holds with $L_C = 1$ and $L_Z = 1+\gamma L$. Since $\gamma = \mathcal{O}(1/L)$, it follows that $L_Z = \mathcal{O}(1)$. For the initial conditions, we have $\|\mathcal{C}^1(\mathbf{z}^0,\mathbf{0})\|_2 \le \|\mathbf{w}^0-\mathbf{w}^\infty\|_2 + \|\mathbf{w}^\infty\|_2 = \mathcal{O}(\sqrt{md})$ and $\|\mathcal{C}^3(\mathbf{z}^0,\mathbf{0})\|_2 \le \gamma L(\|\mathbf{w}^0-\mathbf{w}^\infty\|_2 + \|\mathbf{w}^\infty\|_2) + \|\mathbf{y}^0-\mathbf{y}^\infty\|_2 + \|\mathbf{y}^\infty\|_2 = \mathcal{O}(\sqrt{md})$.

Therefore, $L_A L_Z = \mathcal{O}(1), L_C = \mathcal{O}(1), \|\mathcal{C}^1(\mathbf{z}^0,\mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md}), \|\mathcal{C}^2(\mathbf{z}^0,\mathbf{0})\|_2 = 0, \|\mathcal{C}^3(\mathbf{z}^0,\mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$, and $\|\mathcal{C}^4(\mathbf{z}^0,\mathbf{0})\|_2 = 0$; hence A.9 holds.

(Prox-)NIDS [80]

The update of prox-NIDS solving (P), reads

$$
\begin{aligned}
\mathbf{x}^k &= \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}^k),\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + (\mathbf{I}-\hat{\mathbf{W}})\mathbf{w}^{k+1},\\
\mathbf{w}^{k+1} &= \hat{\mathbf{W}}\Big(\mathbf{x}^k - \gamma\nabla f(\mathbf{x}^k)\Big) - \mathbf{y}^k,
\end{aligned}
$$

with $\mathbf{y}^0 = \mathbf{0}$ and $\mathbf{w}^0 \in \mathbb{R}^{md}$.

Prox-NIDS can be cast as (M) with $R = 2$ rounds of communications, using the following:

$$
\begin{aligned}
&\mathbf{z}^\top = [\mathbf{y}^\top, \mathbf{w}^\top]\\
&\mathcal{C}_i^1\Big(\mathbf{z}_i^k,\mathbf{0}\Big) = \mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k) - \gamma\nabla f_i(\mathrm{prox}_{\gamma\mathbf{I},r}(\mathbf{w}_i^k)) &\quad (3.63)\\
&\mathcal{C}_i^2\Big(\mathbf{z}_i^k,\hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}\Big) = \sum_{j\in\mathcal{N}_i} w_{ij}\hat{\mathbf{c}}_j^{k,1} - \mathbf{y}_i^k, &\quad (3.64)\\
&\mathbf{z}_i^{k+1} = \mathcal{A}_i\Big(\mathbf{z}_i^k,\hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1},\hat{\mathbf{c}}_{\mathcal{N}_i}^{k,2}\Big) = \begin{bmatrix} \mathbf{y}_i^k - \sum_{j\in\mathcal{N}_i} w_{ij}\left(\hat{\mathbf{c}}_j^{k,2} - \hat{\mathbf{c}}_i^{k,2}\right) \\ \hat{\mathbf{c}}_i^{k,2} \end{bmatrix}. &\quad (3.65)
\end{aligned}
$$

We show next that the above instance of (M) satisfies A.6-9.

- **On A.6:** Using [80] it is not difficult to check that, if $\gamma = 2/(\mu + L)$, then prox-NIDS satisfies A.6, with $\lambda < 1$ and the norm $\| \bullet \|$ defined as

$$\|\mathbf{z}\|^2 = \mathbf{w}^\top \hat{\mathbf{W}}^{-1}\mathbf{w} + \mathbf{y}^\top \hat{\mathbf{W}}^{-1}(\mathbf{I} - \hat{\mathbf{W}})^\dagger \hat{\mathbf{W}}^{-1}\mathbf{y}.$$

Note that $\|\mathbf{z}\|^2 \geq \|\mathbf{z}\|_2^2$.

- **On A.7-9:** Based on (3.63), (3.64), and (3.65), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2) = \begin{bmatrix} \mathbf{y} + (\mathbf{I} - \hat{\mathbf{W}})\mathbf{c}^2 \\ \mathbf{c}^2 \end{bmatrix},$$
$$\mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}) - \gamma \nabla f(\mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})), \quad \text{and} \quad \mathcal{C}^2(\mathbf{z}, \mathbf{c}) = \hat{\mathbf{W}}\mathbf{c} - \mathbf{y},$$

respectively; and $\mathcal{Z} = \mathrm{span}(\mathbf{I} - \hat{\mathbf{W}}) \times \mathbb{R}^{md}$. We now show that prox-NIDS satisfies A.9. Note that

$$\left\| \mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}_1, \mathbf{c}\right) \right\|^2 = \|\sqrt{\mathbf{I} - \hat{\mathbf{W}}}\hat{\mathbf{W}}^{-1}(\mathbf{c} - \mathbf{c})\|_2^2 \leq (\nu^{-2} - \nu^{-1})\|\mathbf{c} - \mathbf{c}\|_2^2,$$

and $\mathcal{A}(\mathbf{z}, \mathbf{c}^1, \mathbf{c}^2)$ is constant with respect to $\mathbf{c}^1$. It follows that A.7 holds with $L_A = \sqrt{\nu^{-2} - \nu^{-1}}$.

We next derive L_C and L_Z. Using the non-expansive property of the proximal mapping, it holds

$$\begin{aligned} &\|\mathcal{C}^1(\mathbf{z}, \mathbf{0}) - \mathcal{C}^1(\mathbf{z}, \mathbf{0})\|_2 \\ &= \|(\mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}) - \mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})) - \gamma(\nabla f(\mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})) - \nabla f(\mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})))\|_2 \\ &\leq (1 + \gamma L)\|\mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w}) - \mathrm{prox}_{\gamma \mathbf{I}, r}(\mathbf{w})\|_2 \leq (1 + \gamma L)\|\mathbf{w} - \mathbf{w}\|_2 \leq (1 + \gamma L)\|\mathbf{z} - \mathbf{z}\|_2, \\ &\|\mathcal{C}^2(\mathbf{z}, \mathbf{c}) - \mathcal{C}^2(\mathbf{z}, \mathbf{c})\|_2 = \|\hat{\mathbf{W}}(\mathbf{c} - \mathbf{c})\|_2 + \|\mathbf{z} - \mathbf{z}\|_2 \leq \|\mathbf{c} - \mathbf{c}\|_2 + \|\mathbf{z} - \mathbf{z}\|_2, \end{aligned}$$

which implies that A.8 holds with $L_C = 1$ and $L_Z = 1 + \gamma L$. Since $\gamma = \mathcal{O}(1/L)$, it follows that $L_Z = \mathcal{O}(1)$. For the initial conditions, we have $\|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 \leq (1 + \gamma L)(\|\mathbf{w}^0 - \mathbf{w}^\infty\|_2 + \|\mathbf{w}^\infty\|_2) = \mathcal{O}(\sqrt{md})$ and $\|\mathcal{C}^2(\mathbf{z}^0, \mathbf{0})\|_2 \leq \|\mathbf{y}^0 - \mathbf{y}^\infty\|_2 + \|\mathbf{y}^\infty\|_2 = \mathcal{O}(\sqrt{md})$.

Therefore, $L_A L_Z = \mathcal{O}(1), L_C = \mathcal{O}(1), \|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$, and $\|\mathcal{C}^2(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$; hence A.9 holds.

GD over star networks [89]

Consider Problem (P) with $r = 0$ over a master/workers system. The GD update

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{\gamma}{m}\sum_{i=1}^m \nabla f_i\left(\mathbf{x}^k\right), \tag{3.66}$$

with $\mathbf{x}^0 \in \mathbb{R}^d$, is implemented at the master node as follows: at iteration k, the server broadcasts $\mathbf{x}^k$ to the m agents; each agent i then computes its own gradient $\nabla f_i(\mathbf{x}^k)$ and sends it back to the master; upon collecting all local gradients, the server updates the variable $\mathbf{x}^{k+1}$ according to (3.66).

The GD (3.66) can be cast as (M) with $R = 1$ round of communications, using the following:

$$\begin{aligned}
&\mathbf{z} = \mathbf{1}_m \otimes \mathbf{x}, \\
&\hat{\mathbf{c}}_i^{k,1} = \mathcal{C}_i^1\left(\mathbf{z}_i^k, \mathbf{0}\right) = \nabla f_i\left(\mathbf{x}_i^k\right), && (3.67)\\
&\mathbf{z}^{k+1} = \mathcal{A}\left(\mathbf{z}^k, \hat{\mathbf{c}}_{[m]}^{k,1}\right) = \mathbf{x}^k - \frac{\gamma}{m}\sum_{i=1}^m \mathbf{1}_m \otimes \hat{\mathbf{c}}_i^{k,1}. && (3.68)
\end{aligned}$$

We show next that the above instance of (M) satisfies A.6-9.

- **On A.6:** Using [89] it is not difficult to check that, if $\gamma = 2/(\mu + L)$, then GD over star networks satisfies A.6, with $\lambda < 1$ and the norm $\| \bullet \|$ defined as

$$\|\mathbf{z}\| = \|\mathbf{x}\|_2.$$

Note that $\|\mathbf{z}\| \geq \|\mathbf{z}\|_2$.

- **On A.7-9:** Based on (3.67) and (3.68), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\mathcal{A}(\mathbf{z}, \mathbf{c}^1) = \mathbf{z} - \frac{\gamma}{m} \sum_{i=1}^{m} \mathbf{1}_m \otimes \mathbf{c}_i, \quad \mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \gamma \nabla f(\mathbf{z}),$$

respectively; and $\mathcal{Z} = \{\mathbf{1}_m \otimes \mathbf{x} : \mathbf{x} \in \mathbb{R}^d\}$.

We now show that GD over star networks satisfies A.9. Since

$$\mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) = -\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_m \otimes \left(\mathbf{c}_i - \mathbf{c}_i\right),$$

we have

$$\left\| \mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) \right\|^2 = \frac{1}{m} \left\| \sum_{i=1}^{m} \mathbf{c}_i - \mathbf{c}_i \right\|_2^2 \leq \left\| \mathbf{c} - \mathbf{c} \right\|_2^2.$$

Hence, A.7 holds with $L_A = 1$.

We now derive L_C and L_Z. Note that

$$\|\mathcal{C}^1(\mathbf{z}, \mathbf{0}) - \mathcal{C}^1(\mathbf{z}, \mathbf{0})\|_2^2 \leq \gamma^2 L^2 \left\| \mathbf{z} - \mathbf{z} \right\|_2^2,$$

which implies that A.8 holds with $L_C = 0$ and $L_Z = \gamma L$. Since $\gamma = \mathcal{O}(1/L)$, it follows that $L_A L_Z = \mathcal{O}(1)$ and $\|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 \leq \gamma L \|\mathbf{x} - \tilde{\mathbf{x}}^*\|_2 = \mathcal{O}(\sqrt{md})$, where $\tilde{\mathbf{x}}^* = (\tilde{\mathbf{x}}_i^*)_{i=1}^m$ with $\tilde{\mathbf{x}}_i^* = \arg\min_{\mathbf{x}_i} f_i(\mathbf{x}_i)$.

Therefore, $L_A L_Z = \mathcal{O}(1), L_C = \mathcal{O}(1), \|\mathcal{C}^1(\mathbf{z}^0, \mathbf{0})\|_2 = \mathcal{O}(L_Z \sqrt{md})$; hence A.9 holds.

Primal-dual algorithm [59], [75]

Let $\mathbf{L} = (l_{ij})_{i,j=1}^m$ be the Laplacian matrix associated with the 0-1 adjacency matrix of $\mathcal{G}$, i.e., $l_{ii} = |\mathcal{N}_i \setminus \{i\}|, i \in [m]$; and $l_{ij} = -\mathbb{1}\{(i,j) \in \mathcal{E}\}, i \neq j \in [m]$; and $\hat{\mathbf{L}} = \mathbf{L} \otimes \mathbf{I}_d$.

The primal-dual algorithm solving (P), with $r = 0$, reads [59], [75]

$$\begin{aligned}
\mathbf{y}_i^{k+1} &= \mathbf{y}_i^k + \gamma \sum_{j \in \mathcal{N}_i} l_{ij} \mathbf{x}_j^k, \\
\mathbf{x}_i^{k+1} &= \arg\min_{\mathbf{x}_i} \; f_i(\mathbf{x}_i) + \mathbf{x}_i^\top \mathbf{y}_i^{k+1},
\end{aligned}$$

with $\mathbf{y}_i^0 = \mathbf{0}$ and $\mathbf{x}_i^0 = \arg\min_{\mathbf{x}_i} f_i(\mathbf{x}_i)$.

The primal-dual algorithm can be cast in the form (M), with $R = 1$ round of communications, using the following:

$$\begin{aligned}
&\mathbf{z} = \mathbf{y} \\
&\mathcal{C}_i^1\left(\mathbf{z}_i^k, \mathbf{0}\right) = \underset{\mathbf{c}_i}{\arg\min}\; f_i(\mathbf{c}_i) + \mathbf{c}_i^\top \mathbf{y}_i^k, && (3.69) \\
&\mathbf{z}_i^{k+1} = \mathcal{A}_i\left(\mathbf{z}_i^k, \hat{\mathbf{c}}_{\mathcal{N}_i}^{k,1}\right) = \mathbf{y}_i^k + \gamma \sum_{j \in \mathcal{N}_i} l_{ij} \hat{\mathbf{c}}_j^{k,1}. && (3.70)
\end{aligned}$$

We show next that the above instance of (M) satisfies A.6-9.

• **On A.6:** Define $\mathbf{M} = \sqrt{\mathbf{\Sigma}}\mathbf{Q}$, where $\hat{\mathbf{L}} = \mathbf{Q}^\top \mathbf{\Sigma} \mathbf{Q}$ is its eigenvalue decomposition, with $\mathbf{\Sigma}$ being diagonal with elements sorted in descending order; and let $\bar{\mathbf{M}}$ be the matrix containing the non-zero rows of $\mathbf{M}$. Using [59] it is not difficult to check that, if $\gamma = \frac{2L\mu}{\mu\rho_2(\mathbf{L}) + L\rho_m(\mathbf{L})}$, the primal-dual algorithm satisfies A.6, with $\lambda < 1$ and the norm $\| \bullet \|$ defined as

$$\|\mathbf{z}\| = \sqrt{\rho_2(\mathbf{L})} \left\| \left(\bar{\mathbf{M}}\bar{\mathbf{M}}^\top\right)^{-1} \bar{\mathbf{M}}\mathbf{z} \right\|_2 .$$

Note that $\|\mathbf{z}\| \geq \|\mathbf{z}\|_2$.

• **On A.7-9:** Based on (3.69) and (3.70), the mappings $\mathcal{A}$ and $\mathcal{C}$ read

$$\mathcal{A}(\mathbf{z}, \mathbf{c}^1) = \mathbf{z} + \gamma \hat{\mathbf{L}} \mathbf{c}^1 \quad \text{and} \quad \mathcal{C}^1(\mathbf{z}, \mathbf{0}) = \begin{bmatrix} \underset{\mathbf{c}}{\arg\min}\; f_1(\mathbf{c}) + \mathbf{c}^\top \mathbf{z}_1 \\ \vdots \\ \underset{\mathbf{c}}{\arg\min}\; f_m(\mathbf{c}) + \mathbf{c}^\top \mathbf{z}_m \end{bmatrix},$$

respectively; and $\mathcal{Z} = \operatorname{span}(\mathbf{L})$. We now show that the primal-dual algorithm satisfies A.9. Note that

$$\mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) - \mathcal{A}\left(\mathbf{z}, \mathbf{c}\right) = \gamma \hat{\mathbf{L}} \left(\mathbf{c} - \mathbf{c}\right).$$

It follows that

$$\begin{aligned}\left\|\mathcal{A}\left(\mathbf{z},\mathbf{c}\right)-\mathcal{A}\left(\mathbf{z},\mathbf{c}\right)\right\|^2 &= \gamma^2\rho_2(\mathbf{L})\left\|\left(\bar{\mathbf{M}}\bar{\mathbf{M}}^\top\right)^{-1}\bar{\mathbf{M}}\hat{\mathbf{L}}\left(\mathbf{c}-\mathbf{c}\right)\right\|_2^2\\ &\leq \gamma^2\rho_2(\mathbf{L})\rho_m\left(\mathbf{L}\right)\left\|\mathbf{c}-\mathbf{c}\right\|_2^2,\end{aligned}$$

which implies that A.7 holds with $L_A = \gamma\sqrt{\rho_2(\mathbf{L})\rho_m(\mathbf{L})}$. Using the expression of γ, it follows that $L_A = \mathcal{O}(\mu)$. Hence, for all the objective functions of (P) such that $\mu = \mathcal{O}(1)$, we also have $L_A = \mathcal{O}(1)$. For instance, this is the typical case of several machine learning problems where a regularization $\mu/2\|\mathbf{x}\|^2$ is enforced on the objective function to make it strongly convex, with $\mu = \mathcal{O}(1)$.

We now derive L_C and L_Z. Since

$$\|\mathcal{C}^1(\mathbf{z},\mathbf{0})-\mathcal{C}^1(\mathbf{z},\mathbf{0})\|_2 = \|\mathbf{z}-\mathbf{z}\|_2,$$

A.8 holds with $L_C = 0$ and $L_Z = 1$. For the initial conditions, since $\mathbf{z}^0 = \mathbf{0}$ we have $\|\mathcal{C}^1(\mathbf{z}^0,\mathbf{0})\|_2 = \|\tilde{\mathbf{x}}^*\|_2 = \mathcal{O}(\sqrt{md})$, where $\tilde{\mathbf{x}}_i^* = (\tilde{\mathbf{x}}_i^*)_{i=1}^m$ with $\tilde{\mathbf{x}}_i^* \triangleq \arg\min_{\mathbf{x}_i} f_i(\mathbf{x}_i)$.

Therefore, $L_A L_Z = \mathcal{O}(1), L_C = \mathcal{O}(1)$, and $\|\mathcal{C}^1(\mathbf{z}^0,\mathbf{0})\|_2 = \mathcal{O}(L_Z\sqrt{md})$, when $\mu = \mathcal{O}(1)$; hence A.9 holds.

4. SUMMARY

In this dissertation, we have studied the design and analysis of distributed algorithms employing finite-bit quantized communication among agents for several important problems. First, we have proposed the first distributed algorithms solving the weight-balancing and average consensus problems addressing the simplex and finite-rate communication among agents. The proposed algorithms are compatible with *any rate constraint*, i.e., they have convergence guarantee when using 1-bit communication. In the analysis, we have proposed a novel metric, which together with the proposed step-size rule, greatly facilitates the convergence analysis. We have also characterized the convergence rate of the proposed algorithms.

In the second half of the dissertation, we have proposed a black-box quantization mechanism which can be employed to a general class of linearly convergent distributed algorithms which can be cast as fixed-point iterates. In fact, the proposed mechanism generates the first distributed algorithms employing finite-bit quantization solving the composite optimization problem. In the analysis, 1) we have shown that linear rate can be preserved and characterized the effect of quantization on the convergence rate; and 2) we have shown that the compression rule, a special instance of the proposed condition, requires infinite bits to implement; 3) we have characterized the communication cost for the proposed scheme, which is the first results for distributed algorithms over mesh networks.

An interesting extension of this dissertation is the answer to the open question: can one design a *universal* quantization mechanism which can transform *any* unquantized algorithm into its finite-bit quantized counterpart, while maintaining the same order of convergence rate?

REFERENCES

[1] J. N. Tsitsiklis, “Problems in decentralized decision making and computation,” Ph.D. dissertation, Mass. Inst. Technol., Cambridge, 1984.

[2] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J. S. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.

[3] J. A. Bazerque and G. B. Giannakis, “Distributed spectrum sensing for cognitive radio networks by exploiting sparsity,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.

[4] P. Forero, A. Cano, and G. B. Giannakis, “Consensus-based distributed support vector machines,” *Journal of Machine Learning Research*, vol. 59, pp. 1663–1707, May 2010.

[5] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[6] A. Kashyap, T. Basar, and R. Srikant, “Quantized consensus,” *Automatica*, vol. 43, no. 7, pp. 1192–1203, May 2007.

[7] T. C. Aysal, M. J. Coates, and M. G. Rabbat, “Distributed average consensus with dithered quantization,” *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 4905–4918, Oct. 2008.

[8] P. D. Lorenzo and G. Scutari, “Next: In-network nonconvex optimization,” *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 2, pp. 120–136, Jun. 2016.

[9] C. N. Hadjicostis and A. D. Domínguez-García, “Distributed balancing of commodity networks under flow interval constraints,” *IEEE Trans. Autom. Control*, vol. 64, no. 1, pp. 51–65, Jan. 2019.

[10] R. Olfati-Saber and R. M. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

[11] G. Cybenko, “Dynamic load balancing for distributed memory multiprocessors,” *J. Parallel and Distrib. Comput.*, vol. 7, no. 2, pp. 279–301, Oct. 1989.

[12] J. A. Fax and R. M. Murray, “Information flow and cooperative control of vehicle formations,” *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1465–1476, Sep. 2004.

[13] G. Scutari, S. Barbarossa, and L. Pescosolido, “Distributed decision through self-synchronizing sensor networks in the presence of propagation delays and asymmetric channels,” *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1667–1684, Apr. 2008.

[14] A. I. Rikos, T. Charalambous, and C. N. Hadjicostis, “Distributed weight balancing over digraphs,” *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 2, pp. 190–201, Jun. 2014.

[15] L. Hooi-Tong, “On a class of directed graphs-with an application to traffic-flow problems,” *Operations Res.*, vol. 18, no. 1, pp. 87–94, 1970.

[16] B. Gharesifard and J. Cortés, “Distributed strategies for generating weight-balanced and doubly stochastic digraphs,” *Eur. J. Contr.*, vol. 18, no. 6, pp. 539–557, 2012.

[17] A. I. Rikos and C. N. Hadjicostis, “Distributed balancing with constrained integer weights,” *IEEE Trans. Autom. Control*, vol. 64, no. 6, pp. 2553–2558, Jun. 2019.

[18] A. I. Rikos and C. N. Hadjicostis, “Distributed integer weight balancing in the presence of time delays in directed graphs,” *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1300–1309, Sep. 2018.

[19] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, “On distributed averaging algorithms and quantization effects,” *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2506–2517, Oct. 2009.

[20] J. Lavaei and R. M. Murray, “Quantized consensus by means of gossip algorithm,” *IEEE Trans. Autom. Control*, vol. 57, no. 1, pp. 19–32, Jan. 2012.

[21] M. El Chamie, J. Liu, and T. Basar, “Design and analysis of distributed averaging with quantized communication,” *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3870–3884, Dec. 2016.

[22] S. Zhu and B. Chen, “Distributed average consensus with bounded quantization,” in *Proc. IEEE 17th SPAWC*, Jul. 2016, pp. 1–6.

[23] S. Kar and J. M. F. Moura, “Distributed consensus algorithms in sensor networks: Quantized data and random link failures,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.

[24] T. Li, M. Fu, L. Xie, and J.-F. Zhang, “Distributed consensus with limited communication data rate,” *IEEE Trans. Autom. Control*, vol. 56, no. 2, pp. 279–292, Feb. 2011.

[25] D. Thanou, E. Kokiopoulou, Y. Pu, and P. Frossard, “Distributed average consensus with quantization refinement,” *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 194–205, Jan. 2013.

[26] R. Rajagopal and M. J. Wainwright, “Network-based consensus averaging with general noisy channels,” *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 373–385, Jan. 2011.

[27] H. Li, G. Chen, T. Huang, and Z. Dong, “High-performance consensus control in networked systems with limited bandwidth communication and time-varying directed topologies,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1043–1054, May 2017.

[28] C.-S. Lee, N. Michelusi, and G. Scutari, “Topology-agnostic average consensus in sensor networks with limited data rate,” in *Proc. 51st ACSSC*, Oct. 2017.

[29] Y. Wang, Q. Wu, and Y. Wang, “Quantized consensus with finite data rate under directed topologies,” in *Proc. 50th IEEE CDC and ECC*, Dec. 2011, pp. 6427–6432.

[30] Z. Chen, J. Ma, and X. Yu, “Consensus of general linear multi-agent systems under directed communication graph with limited data rate,” in *Proc. 3rd ISAS*, May 2019, pp. 394–399.

[31] A. I. Rikos and C. N. Hadjicostis, “Distributed average consensus under quantized communication via event- triggered mass summation,” in *Proc. 57th IEEE CDC*, Dec. 2018, pp. 894–899.

[32] A. I. Rikos and C. N. Hadjicostis, “Distributed average consensus under quantized communication via event-triggered mass splitting,” *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 2957–2962, 2020, 21th IFAC World Congress.

[33] B. Charron-Bost and P. Lambein-Monette, “Randomization and quantization for average consensus,” in *Proc. 57th IEEE CDC*, Dec. 2018, pp. 3716–3721.

[34] S. Zhu, Y. C. Soh, and L. Xie, “Distributed parameter estimation with quantized communication via running average,” *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4634–4646, Sep. 2015.

[35] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[36] L. Xiao, S. Boyd, and S. Lall, “A scheme for robust distributed sensor fusion based on average consensus,” in *Proc. 4th Int. Symp. on Information Processing in Sensor Networks*, Apr. 2005, pp. 63–70.

[37] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th IEEE FOCS*, Oct. 2003, pp. 482–491.

[38] F. Blasa, S. Cafiero, G. Fortino, and G. Fatta, "Symmetric push-sum protocol for decentralised aggregation," in *Proc. 3rd AP2PS*, Jan. 2011, pp. 27–32.

[39] A. Olshevsky, I. C. Paschalidis, and A. Spiridonoff, "Fully asynchronous push-sum with growing intercommunication intervals," in *Proc. ACC*, Jun. 2018, pp. 591–596.

[40] F. Fagnani and S. Zampieri, "Average consensus with packet drop communication," *SIAM J. on Control and Optim.*, vol. 48, no. 1, pp. 102–133, 2009.

[41] B. Gerencsér and J. M. Hendrickx, "Push-sum with transmission failures," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1019–1033, Mar. 2019.

[42] J. M. Hendrickx and J. N. Tsitsiklis, "Fundamental limitations for anonymous distributed systems with broadcast communications," in *Proc. 53rd ALLERTON*, Sep. 2015, pp. 9–16.

[43] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation:Numerical Methods*. Belmont, MA, USA: Athena Scientific, 1989.

[44] C.-S. Lee, N. Michelusi, and G. Scutari, "Distributed quantized weight-balancing and average consensus over digraphs," in *Proc. 57th IEEE CDC*, Dec. 2018.

[45] Z. Qu, *Cooperative Control of Dynamical Systems: Applications to Autonomous Vehicles*, 1st. Springer Publishing Company, Incorporated, 2009.

[46] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[47] R. Durrett, *Probability: Theory and Examples*, 4.1th. New York, NY, USA: Cambridge University Press (4th ed.), 2013.

[48] D. K. Molzahn, F. Dörfler, H. Sandberg, *et al.*, "A survey of distributed optimization and control algorithms for electric power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2941–2962, Nov. 2017.

[49] A. Nedić, J.-S. Pang, G. Scutari, and Y. Sun, *Multi-agent Optimization*, 1st ed. Springer, Cham, 2018.

[50] T. Yang, X. Yi, J. Wu, *et al.*, "A survey of distributed optimization," *Annu Rev Control*, vol. 47, pp. 278–305, 2019.

[51] R. Bekkerman, M. Bilenko, and J. Langford, *Scaling up Machine Learning: Parallel and Distributed Approaches*. Cambridge University Press, 2011.

[52] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu., "Can decentralized algorithms outperform centralized algorithms? A case study for decentralized parallel stochastic gradient descent," in *Proc. 31st NeurIPS*, Dec. 2017.

[53] H. Tang, S. Gan, C. Zhang, T. Zhang, and J. Liu, "Communication compression for decentralized training," in *Proc. 32nd NeurIPS*, Dec. 2018, pp. 7663–7673.

[54] H. Taheri, A. Mokhtari, H. Hassani, and R. Pedarsani, "Quantized push-sum for gossip and decentralized optimization over directed graphs," *arXiv:2002.09964v5*, Dec. 2020.

[55] A. Beznosikov, S. Horváth, P. Richtárik, and M. Safaryan, "On biased compression for distributed learning," *arXiv:2002.12410v1*, Feb. 2020.

[56] D. Kovalev, A. Koloskova, M. Jaggi, P. Richtarik, and S. U. Stich, "A linearly convergent algorithm for decentralized optimization: Sending less bits for free!" *arXiv:2011.01697v1*, Nov. 2020.

[57] Y. Liao, Z. Li, K. Huang, and S. Pu, "Compressed gradient tracking methods for decentralized optimization with linear convergence," *arXiv:2103.13748v3*, Jun. 2021.

[58] X. Liu, Y. Li, R. Wang, J. Tang, and M. Yan, "Linear convergent decentralized optimization with compression," in *Proc. 9th ICLR*, May 2021.

[59] S. Magnússon, H. Shokri-Ghadikolaei, and N. Li, "On maintaining linear convergence of distributed learning and optimization under limited communication," *IEEE Trans. Signal Process.*, vol. 68, pp. 6101–6116, Oct. 2020.

[60] Y. Kajiyama, N. Hayashi, and S. Takai, "Linear convergence of consensus-based quantized optimization for smooth and strongly convex cost functions," *IEEE Trans. Autom. Control (Early Access)*, pp. 1–1, 2020.

[61] C.-S. Lee, N. Michelusi, and G. Scutari, "Finite rate quantized distributed optimization with geometric convergence," in *Proc. 52nd ACSSC*, Oct. 2018, pp. 1876–1880.

[62] Y. Sun, A. Daneshmand, and G. Scutari, "Convergence rate of distributed optimization algorithms based on gradient tracking," *arXiv:1905.02637v1*, May 2019.

[63] G. Scutari and Y. Sun, "Distributed nonconvex constrained optimization over time-varying digraphs," *Math. Program.*, vol. 176, pp. 497–544, Feb. 2019.

[64] C.-S. Lee, N. Michelusi, and G. Scutari, “Finite rate distributed weight-balancing and average consensus over digraphs,” *IEEE Trans. Autom. Control (Early Access)*, pp. 1–1, 2020.

[65] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, “Qsgd: Communication-efficient sgd via gradient quantization and encoding,” in *Proc. 31st NeurIPS*, Dec. 2017.

[66] D. Alistarh, T. Hoefler, M. Johansson, N. Konstantinov, S. Khirirat, and C. Renggli, “The convergence of sparsified gradient methods,” in *Proc. 32nd NeurIPS*, Dec. 2018.

[67] P. Karimireddy, Q. Rebjock, S. Stich, and M. Jaggi, “Error feedback fixes signsgd and other gradient compression schemes,” in *Proc. 36th ICML*, Jul. 2018.

[68] S. U. Stich, J.-B. Cordonnier, and M. Jaggi, “Sparsified sgd with memory,” in *Proc. 32nd NeurIPS*, Dec. 2018.

[69] M. J. Anastasia Koloskova Sebastian U. Stich, “Decentralized stochastic optimization and gossip algorithms with compressed communication,” *arXiv:1902.00340v1*, Feb. 2019.

[70] X. Zhang, J. Liu, Z. Zhu, and E. S. Bentley, “Compressed distributed gradient descent: Communication-efficient consensus over networks,” in *Proc. IEEE INFOCOM*, Apr. 2019, pp. 2431–2439.

[71] S. Zheng, Z. Huang, and J. Kwok, “Communication-efficient distributed blockwise momentum sgd with error-feedback,” in *Proc. 33rd NeurIPS*, Dec. 2019.

[72] E. Gorbunov, F. Hanzely, and P. Richtárik, “A unified theory of sgd: Variance reduction, sampling, quantization and coordinate descent,” in *Proc. 23rd AISTATS*, Apr. 2020.

[73] S. U. Stich, “On communication compression for distributed optimization on heterogeneous data,” *arXiv:2009.02388v2*, Dec. 2020.

[74] F. Haddadpour, M. M. Kamani, A. Mokhtari, and M. Mahdavi, “Federated learning with compression: Unified analysis and sharp guarantees,” in *Proc. 24th AISTATS*, Apr. 2021.

[75] C. A. Uribe, S. Lee, A. Gasnikov, and A. Nedić, “A dual approach for optimal algorithms in distributed optimization over networks,” *Optim. Methods Softw.*, vol. 36, no. 1, pp. 171–210, 2021.

[76] W. Shi, Q. Ling, G. Wu, and W. Yin, "Extra: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, pp. 944–966, May 2015.

[77] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Convergence of asynchronous distributed gradient methods over stochastic networks," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 434–448, Feb. 2018.

[78] A. Nedic, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, Dec. 2017.

[79] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Trans. Control. Netw. Syst.*, vol. 5, no. 3, pp. 1245–1260, Sep. 2018.

[80] J. Xu, Y. Tian, Y. Sun, and G. Scutari, "Distributed algorithms for composite optimization: Unified and tight convergence analysis," *arXiv:2002.11534v2*, Mar. 2020.

[81] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *Proc. 54th IEEE CDC*, 2015, pp. 2055–2060.

[82] Z. Li, W. Shi, and M. Yan, "A decentralized proximal-gradient method with network independent step-sizes and separated convergence rates," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4494–4506, Sep. 2019.

[83] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning—part i: Algorithm development," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 708–723, Feb. 2019.

[84] S. A. Alghunaim, E. Ryu, K. Yuan, and A. H. Sayed, "Decentralized proximal gradient algorithms with linear convergence rates," *IEEE Trans. Autom. Control (Early Access)*, pp. 1–1, 2020.

[85] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Automat. Contr.*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[86] A. Agarwal, S. Negahban, and M. J. Wainwright, "Fast global convergence rates of gradient methods for high-dimensional statistical recovery," in *Proc. 24th NeurIPS*, Dec. 2010, pp. 37–45.

[87] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[88] J. J. Moreau, “Proximité et dualité dans un espace hilbertien,” *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[89] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, 1st ed. Springer Publishing Company, Incorporated, 2014, ISBN: 1461346916.

[90] C.-S. Lee, N. Michelusi, and G. Scutari, “Finite-bit quantization for distributed algorithms with linear convergence,” *submitted to IEEE Trans. Inf. Theory*, Jul. 2021, Available [online]: https://arxiv.org/abs/2107.11304.

VITA

Chang-Shen Lee received the B.Sc. degree in the electrical engineering and computer science honor program and the M.Sc. in communications engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2012 and 2014, respectively. He joined the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 2016, as a Ph.D. student. From 2015 to 2016, he was a Research Assistant with the Research Center for Information Technology Innovation, Academia Sinica, Taiwan. His research interests include machine learning, distributed computing, and optimization.