Convergence of Limited Communication Gradient Methods

Sindri Magnússon, Chinwendu Enyioha, *Member, IEEE*, Na Li, *Member, IEEE*, Carlo Fischione, *Member, IEEE*, and Vahid Tarokh, *Fellow, IEEE*

***Abstract*—Distributed optimization increasingly plays a central role in economical and sustainable operation of cyber-physical systems. Nevertheless, the complete potential of the technology has not yet been fully exploited in practice due to communication limitations posed by the real-world infrastructures. This work investigates fundamental properties of distributed optimization based on gradient methods, where gradient information is communicated using a limited number of bits. In particular, a general class of quantized gradient methods are studied, where the gradient direction is approximated by a finite quantization set. Sufficient and necessary conditions are provided on such a quantization set to guarantee that the methods minimize any convex objective function with Lipschitz continuous gradient and a nonempty and bounded set of optimizers. A lower bound on the cardinality of the quantization set is provided, along with specific examples of minimal quantizations. Convergence rate results are established that connect the fineness of the quantization and the number of iterations needed to reach a predefined solution accuracy. Generalizations of the results to a relevant class of constrained problems using projections are considered. Finally, the results are illustrated by simulations of practical systems.**

***Index Terms*—Cyberphysical systems, distributed optimization, limited communication.**

I. Introduction

RECENT advances in distributed optimization have enabled more economical and sustainable control and operation of cyber-physical systems. However, these systems usually assume the availability of high-performing communication infrastructures, which is often not practically possible. For example, although large-scale cyber-physical systems such as power networks are equipped with a natural communication infrastructure given by the power lines [1], such a communication network has a limited bandwidth. Instead, research efforts in distributed operation of power networks usually assume high data rates and low-latency communication technologies that are, unfortunately, not affordable or available today. Another example is given by wireless sensor networks [2], where efficient usage of communication plays a central role. In fact, these networks are powered by battery sources for communication over wireless links; hence, they are constrained in how much transmission they engage. These communication limitations are especially harsh in underwater networks, where acoustic channels are generally used, which have strong bandwidth limits [3]. Light communications are also essential in coordinating data networks [4], where the control channels that support the data channels are obviously allocated limited bandwidth. Another relevant example is within the emerging technology of extremely low latency networking or tactile internet [5], where information, especially for real-time control applications, will be transmitted with very low latencies over wireless and wired networks. However, this comes at a cost of using short packets containing limited information.

In all the cyber-physical systems mentioned above, distributed optimization plays a central role. These systems are networks of nodes whose operations have to be optimized by local decisions, yet the coordination information can only go through constrained communication channels. In this paper, we restrict ourselves to one of the most prominent distributed optimization methods, decomposition based on the gradient method, and we investigate the fundamental properties of such a method in terms of coordination limitations and optimality.

A. Related Literature

Decomposition methods in optimization have been widely investigated in wired/wireless communication [6]–[9], power networks [10], [11], and wireless sensor networks [12], among others. These methods are typically based on communicating gradient information from a set of source nodes to users, which then solve a simple local subproblem. The procedure can be performed using 1) one-way communication where the source nodes estimate the gradient using available information [7], [13], [14] or 2) two-way communication where users and sources coordinate to evaluate the gradient. We investigate the performance of such methods using one-way communication, where the number of bits per coordination step is limited.

Manuscript received March 9, 2017; revised June 13, 2017; accepted August 7, 2017. Date of publication August 23, 2017; date of current version April 24, 2018. This work was supported by the VR Chromos Project, NSF 1548204, 1608509, and Career 1553407, and ARPA-E under the NODES program. Recommended by Associate Editor Z. Gao. *(Corresponding author: Sindri Magnússon.)*

S. Magnússon and C. Fischione are with the School of Electrical Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden (e-mail: sindrim@kth.se; carlofi@kth.se).

C. Enyioha, N. Li, and V. Tarokh are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: cenyioha@seas.harvard.edu; nali@seas.harvard.edu; vahid@seas.harvard.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2743678

Bandwidth constrained optimization has already received attention in the literature [15]–[20]. Initial studies are found in [15], where Tsitsiklis and Luo provide lower bounds on the number of bits that two processors need to communicate to (approximately) minimize the sum of two convex functions, each of which is only accessed by one processor. More recently, the authors of [16] consider a variant of incremental gradient methods [21] over networks where each node projects its iterate to a grid before sending the iterate to the next node. Similar quantization ideas are considered [17]–[19] in the context of consensus-type subgradient methods [22]. The work in [20] studies the convergence of standard interference function methods for power control in cellular wireless systems, where base stations send binary signals to the users optimizing the transmit radio power. Those papers consider only the original primal optimization problem, without introducing its dual problem, where quantized primal variables are communicated. However, many network and resource sharing/allocation optimization problems are naturally decomposed using duality theory, where it is the dual *gradients* that are communicated. This motivates our studies of limited communication gradient methods.

B. Statement of Contributions

The main contribution of this paper is to investigate the convergence of gradient methods, where gradients are communicated using limited bandwidth. We first consider gradient methods where each coordinate of the gradient is communicated using only 1 bit per iteration. This setup is motivated by dual decomposition applications, where a single entity maintains each dual variable, e.g., in the transmission control protocol (TCP) control in [7], each dual variable is maintained by one flow line. Since dual problems are either unconstrained or constrained to the positive orthant $\mathbb{R}^N_+$,[1] we consider both unconstrained problems and problems constrained to $\mathbb{R}^N_+$. We prove that when the step size $\gamma > 0$ of the gradient method is fixed, then the iterates converge approximately to the set of optimal solutions within some $\epsilon > 0$ accuracy in a finite number of iterations, where ϵ tends to 0 as γ converges to 0. Moreover, we provide an upper complexity bound on the number of iterations needed to reach any $\epsilon > 0$ accuracy. This upper bound grows proportionally to $1/\epsilon^2$ as ϵ goes to zero for unconstrained problems, and proportionally to $1/\epsilon^4$ for problems constrained by $\mathbb{R}^N_+$. We also prove that if the step sizes $\gamma(t)$ are nonsummable and converge to 0, then the iterates converge to the set of optimal solutions.

The second contribution of this paper is to investigate the convergence of more general class of quantized gradient methods (QGMs), where the gradient direction is quantized at every iteration. We start by identifying necessary and sufficient conditions on the quantization so that the QGMs can minimize any convex objective function with Lipschitz continuous gradients and a nonempty and bounded set of optimizers. We show that the minimal quantizations that satisfy these conditions have the cardinality $N+1$, where N is the problem dimension. We prove that for fixed step sizes $\gamma > 0$, the iterates converge approximately to the set of optimal solutions within some ϵ-accuracy, where ϵ converges to 0 as γ converges to 0. We provide an upper complexity bound on the number of iterations T needed to reach any solution accuracy $\epsilon > 0$. This upper bound depends on the fineness of the quantization. Moreover, we show that the solution accuracy $\epsilon > 0$ converges to zero at a rate proportional to $1/\sqrt{T}$ or $1/\sqrt{b}$ where T and b are the numbers of iterations and communicated bits, respectively. We show that when the step sizes $\gamma(t)$ are nonsummable and converge to zero, then the iterates asymptotically converge to the set of optimal solutions.

A conference version of this work including parts of Sections III and V appeared in [23], but without most of the proofs. The rest of the work is presented here for the first time. Our previous papers [24], [25] consider similar resource allocation problems as in this paper without communication constraints.

C. Notation

Vectors and matrices are represented by boldface lower and upper case letters, respectively. The set of real, positive real, and natural numbers, are denoted by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$, respectively. The set of real and positive n vectors and $n \times m$ matrices are denoted by $\mathbb{R}^n$, $\mathbb{R}^n_+$, and $\mathbb{R}^{n\times m}$, respectively. Other sets are represented by calligraphy letters. We denote by $\mathcal{S}^{N-1}$, $\mathcal{S}^{N-1}(\mathbf{x}, R)$, and $\mathcal{B}^N(\mathbf{x}, R)$, respectively, the unit sphere in $\mathbb{R}^N$ and the sphere and open ball centred at $\mathbf{x}$ with radius R in $\mathbb{R}^N$. The superscript $(\cdot)^{\mathbf{T}}$ stands for transpose. We let $[\mathbf{x}]_{\mathcal{X}}$ and $\lceil \mathbf{x} \rceil_+$ denote the projection of $\mathbf{x}$ to the sets $\mathcal{X}$ and $\mathbb{R}_+$. $||\cdot||$ denotes the 2-norm. ∇f is the gradient of f. The distance between a vector $\mathbf{x} \in \mathbb{R}^N$ and a set $\mathcal{X} \subseteq \mathbb{R}^N$ is denoted by $\texttt{dist}\,(\mathbf{x}, \mathcal{X}) = \inf_{\mathbf{v}\in\mathcal{X}} ||\mathbf{v} - \mathbf{x}||$.

II. Preliminaries and Motivating Examples

In this paper, we consider optimization problems of the form

$$\begin{aligned} \underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{1}$$

where $f{:}\mathbb{R}^N \to \mathbb{R}$. We denote by $f^\star$ and $\mathcal{X}^\star \subseteq \mathcal{X}$ the optimal value and the set of optimizers to Problem (1), respectively. We consider the following class of optimization problems.

Definition 1: Let $\mathcal{F}_L(\mathcal{X})$ denote the set of optimization problems of the form of (1), where the function f is convex and differentiable with L-Lipschitz continuous gradient, $\mathcal{X}$ is closed and convex set, and the optimal solution set $\mathcal{X}^\star$ is nonempty and bounded. We write $f \in \mathcal{F}_L(\mathcal{X})$ to indicate that the optimization Problem (1) with the objective function f in the class $\mathcal{F}_L(\mathcal{X})$.

For $f \in \mathcal{F}_L(\mathcal{X})$, it is well known that the gradient method

$$\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma(t)\nabla f(\mathbf{x}(t))]_{\mathcal{X}} \tag{2}$$

converges to $\mathcal{X}^\star$ under appropriate step-size rules [26]. When only the gradient *direction* is known, the above iterates become

$$\mathbf{x}(t+1) = \left[\mathbf{x}(t) - \gamma(t)\frac{\nabla f(\mathbf{x}(t))}{||\nabla f(\mathbf{x}(t))||}\right]_{\mathcal{X}} \tag{3}$$

where we set $\mathbf{x}(t+1) = \mathbf{x}(t)$ if $||\nabla f(\mathbf{x}(t))|| = 0$. For appropriate diminishing step-size rules, the iterates converge to $\mathcal{X}^\star$, and for fixed step size, the stopping condition $f(\mathbf{x}(t)) - f^\star < \epsilon$ can be achieved for any $\epsilon > 0$ [27].

[1]Depending on whether the primal problem has inequality or equality constraints.

Problems on the form of (1) often appear as dual problems used to decompose optimization problems with coupling constraints [8], [9]. There, a distributed solution approach is achieved by solving the dual problem using gradient methods as given in (2) and (3). The dual gradient $\nabla(f(\mathbf{x}(t))$ can often be measured over the course of the algorithm, as it is the constraint violation in the primal problem, given dual variable $\mathbf{x}(t)$ (see, e.g., [7], [13], and [14]). To perform the gradient update in (2) and (3), the gradient or gradient direction must be communicated, as illustrated in the following examples. However, since communication resources are scarce in many networks, we consider another variant of the gradient method in (2). That is, $\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma(t)\mathbf{d}(t)]_{\mathcal{X}}$, where $\mathbf{d}(t)$ is a quantized gradient direction coded using limited number of bits. Before introducing the details of our quantized methods, we provide some application examples.

A. TCP Flow Control

Consider a communication network with N undirected links and S data sources. Let $\mathcal{L}$ and $\mathcal{S}$ denote the ordered sets $\{1,\dots,N\}$ and $\{1,\dots,S\}$. Denote the capacity of link $l \in \mathcal{L}$ by $c_l > 0$ and the transmission rate of source $s \in \mathcal{S}$ by $q_s \in [m_s, M_s]$, where m_s and M_s are upper and lower bound on the source. Source $s \in \mathcal{S}$ has utility function $U_s : [m_s, M_s] \longrightarrow \mathbb{R}$. The data from source $s \in \mathcal{S}$ flow through a path consisting of links $\mathcal{L}_s \subseteq \mathcal{L}$ to its destination. We denote by $\mathcal{S}_l \subseteq \mathcal{S}$ the sources that use link $l \in \mathcal{L}$, i.e., $\mathcal{S}_l := \{s \in \mathcal{S} | l \in \mathcal{L}_s\}$. Then, the TCP flow control is to find data rates q_s, $s \in \mathcal{S}$, that solve the following optimization problem [7], [28]–[30]:

$$\begin{aligned} \underset{q_1,\dots,q_S}{\text{maximize}} \quad & \sum_{s=1}^{S} U_s(q_s) \\ \text{subject to} \quad & \sum_{s\in\mathcal{S}_l} q_s \le c_l, \qquad \text{for } l = 1,\dots,N \\ & q_s \in [m_s, M_s], \qquad \text{for } s = 1,\dots,S. \end{aligned} \tag{4}$$

For notational ease, we write $\mathbf{q} = (q_s)_{s\in\mathcal{S}}$, $\mathcal{Q} = \prod_{s\in\mathcal{S}}[m_s, M_s]$, $\mathbf{c} = (c_l)_{l\in\mathcal{L}}$, and $\mathbf{A} \in \mathbb{R}^{N\times S}$, where

$$\mathbf{A}_{ls} = \begin{cases} 1, & \text{if } l \in \mathcal{S}_s \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The dual problem of (4) is of the form (1), where $\mathcal{X} = \mathbb{R}^N_+$ and the dual function $f : \mathbb{R}^N \to \mathbb{R}$ is given by

$$f(\mathbf{x}) = \underset{\mathbf{q}\in\mathcal{Q}}{\text{maximize}}\, L(\mathbf{q}, \mathbf{x}) = L(\mathbf{q}(\mathbf{x}), \mathbf{x})$$

where

$$L(\mathbf{q}, \mathbf{x}) = \sum_{i=1}^{M} U_i(q_i) - \mathbf{x}^{\mathbf{T}}(\mathbf{A}\mathbf{q} - \mathbf{c})$$

$$q_i(\mathbf{x}) = \underset{q_i\in[m_i,M_i]}{\text{argmax}}\; U_i(q_i) - q_i \sum_{l\in\mathcal{L}_s} x_l. \tag{6}$$

The dual gradient is given by $\nabla f(\mathbf{x}) = \mathbf{c} - \mathbf{A}\mathbf{q}(\mathbf{x})$. The dual gradient is bounded since the set $\mathcal{Q}$ is compact. Moreover, the set of optimal dual variables is bounded from [31, Lemma 1], and the dual gradient $\nabla f(\cdot)$ is L-Lipschitz continuous from [7, Lemma 3], provided that $U_i(\cdot)$ are strongly concave. Therefore, the dual iterates $\mathbf{x}(t)$ in (2) or (3), and the associated primal iterates $\mathbf{q}(t) = \mathbf{q}(\mathbf{x}(t))$ converge to the optimal primal/dual solution of the optimization Problem (4), provided that $\gamma(t)$ are chosen properly.

Dual gradient methods are desirable because they entail distributed solution to Problem (4), since Subproblems (6) can be solved without any coordination between the sources $\mathcal{S}$. Moreover, the gradient component $\nabla_l f(\mathbf{x}) = c_l - \sum_{s\in\mathcal{L}(l)} q_s(x_l)$ can often be measured at the data link l, since it is simply the difference between the link capacity, c_l, and the data transferred through the link [7]. Therefore, the algorithm can be accomplished using one-way communication, where each iteration t consists of the following steps.

1) The links broadcast $\mathbf{x}(t)$ to the sources.
2) The sources solve the local Subproblem (6) and then transfer the source at the data rate $q_i(\mathbf{x}(t))$.
3) The links measure the dual gradient $\nabla_l f(\mathbf{x}(t))$, the data flow through the line, to make the update (2) or (3).

The control channels used to coordinate communication networks are often bandwidth limited. Hence, it is not practically feasible to broadcast the real-valued vector $\nabla f(\mathbf{x}(t))$ to the users. The questions we address in this paper are these: Can we still solve the optimization problem by communicating quantized versions of the gradient? And what are the tradeoffs between optimality and quantization? This motivates our investigation of limited communication gradient methods.

B. Optimal Network Flow

Consider a directed network $(\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1,\dots,N\}$ and $\mathcal{E} = \{1,\dots,E\}$ denote the sets of nodes and edges, respectively. Let v_e denote the flow through the edge $e \in \mathcal{E}$. The flow through the network can then be expressed by the matrix $\mathbf{A} \in \mathbb{R}^{N\times E}$ defined as

$$\mathbf{A}_{ne} := \begin{cases} 1, & \text{if edge } e \text{ leaves node } n \\ -1, & \text{if edge } e \text{ enters node } n \\ 0, & \text{otherwise.} \end{cases}$$

Component $n \in \mathcal{N}$ of $\mathbf{A}\mathbf{v}$ indicates the flow injection/consumption at node n, where $\mathbf{v} = (v_1,\dots,v_E)$. We assume that the flow injection ($c_n > 0$) or consumption ($c_n \le 0$) of node $n \in \mathcal{N}$ is $c_n \in \mathbb{R}$ and set $\mathbf{c} = (c_1,\dots,c_N) \in \mathbb{R}^N$. Then, the optimal network flow problem is [32]–[34]

$$\begin{aligned} \underset{v_1,\dots,v_E}{\text{maximize}} \quad & \sum_{e\in\mathcal{E}} -C_e(v_e) \\ \text{subject to} \quad & \mathbf{A}\mathbf{v} = \mathbf{c} \end{aligned}$$

where $C_e : \mathbb{R} \to \mathbb{R}$ are cost functions of the flow through edge $e \in \mathcal{E}$. If C_e are μ-strongly convex, then the dual gradient is L-Lipschitz continuous with $L = \lambda_{\max}(\mathbf{A}\mathbf{A}^{\mathbf{T}})/\mu$; see [33, Lemma 1]. Then, a similar one-way communication dual decomposition algorithm can be performed as in Section II-A. In contrast to Section II-A, the dual problem is unconstrained, i.e., $\mathcal{X} = \mathbb{R}^N$. In addition, the dual variables are maintained by

the nodes, so it is the nodes that broadcast the dual gradients, while the edges solve the local subproblems. Nevertheless, the dual gradient $\nabla f(\mathbf{x})$ can be measured at the nodes, as the component $\nabla_n f(\mathbf{x})$ is simply the flow injection/consumption of node n for a given $\mathbf{x}$.

C. Task Allocation

Consider the problem of continuous allocation of N tasks between K machines. The sets of tasks and machines are denoted by $\mathcal{N} = \{1, \dots, N\}$ and $\mathcal{K} = \{1 \dots, K\}$, respectively. Let $\mathbf{c} \in \mathbb{R}^N$ denote the total amount of each task that needs to be completed. The amount of each task done by machine $k \in \mathcal{K}$ is represented by the vector $\mathbf{w}_k \in \mathcal{W}_k \subseteq \mathbb{R}^N$, where $\mathcal{W}_k$ is a local constraint of machine k. Then, the goal is to find the task allocation that minimizes the cost:

$$\begin{aligned} \underset{\mathbf{w}_1,\dots,\mathbf{w}_K}{\text{maximize}} \quad & \sum_{k\in\mathcal{K}} -C_k(\mathbf{w}_k) \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} \mathbf{w}_k = \mathbf{c} \\ & \mathbf{w}_k \in \mathcal{W}_k, \ \text{for } k \in \mathcal{K} \end{aligned} \tag{7}$$

where $C_k : \mathbb{R}^N \longrightarrow \mathbb{R}$ is the cost of performing the different tasks on machine $k \in \mathcal{K}$. If C_k are μ-strongly convex, then the dual gradient is L-Lipschitz continuous with $L = (K+1)/\mu$; see [30, Lemma 1]. Therefore, dual gradient methods (2) and (3) can solve the problem. If the task manager can measure the total amount done of each task, i.e., the dual gradient, then a similar one-way communication coordination scheme as in Sections II-A and II-B can solve Problem (7).

As shown later, the Lipschitz constant L will be used to characterize several complexity bounds. Since L on the dual gradient of the examples above is a function of primal problem parameters such as the number of users and the concavity parameter μ, those parameters affect the complexity bounds as well.

III. Quantized Gradient Descent Methods

We consider general QGMs of the form

$$\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma(t)\mathbf{d}(t)]_{\mathcal{X}} \tag{8}$$

where $\mathbf{d}(t) \in \mathcal{D} \subseteq \mathcal{S}^{N-1}$ is a finite set of quantized gradient directions. The following relation is between the cardinality of $\mathcal{D}$ and communicated bits of each Iteration (8).

Remark 1: The set $\mathcal{D}$ can be coded using $\log_2(|\mathcal{D}|)$ bits.

We now introduce the quantization sets $\mathcal{D}$ used in this paper.

A. Binary Quantization

In the application examples in Sections II-A–II-C, each dual variable is associated with a single problem component, i.e., a link, user, or task, respectively. For example, in the TCP control example in Section II-A, the dual variable $\mathbf{x}_n$ is associated with link n. Therefore, to achieve the dual gradient Algorithm (2), each link $l \in \mathcal{L}$ can measure its flow, i.e., the dual gradient component $\nabla_l f(\mathbf{x}(t))$, and then broadcast $\nabla_l f(\mathbf{x}(t))$ to the sources that use link l. However, it might be infeasible to broadcast the full dual gradient when the bandwidth is limited. An alternative approach is to have the links broadcast a binary signal indicating whether the associated dual variable is to be increased or decreased, i.e., link l broadcasts $\texttt{sign}\,(\nabla_l f(\mathbf{x}(t)))$. Similarly, in the network flow problem in Section II-B, each node can measure the flow through the node and then broadcast a binary signal indicating the direction of the associated dual gradient component. This quantization can be formally expressed as follows.

Example 1 (Signs of the gradients): Consider the QGM in (8). Set $\mathcal{D} = \{(1/\sqrt{N})(e_1, e_2, \dots, e_N) \big| e_i \in \{-1, 1\}\}$ and take $\mathbf{d}(t) = \texttt{sign}\,(\nabla f(\mathbf{x}(t)))$.

By using this binary quantization, we prove the convergence of the Iterates (8) when $\mathcal{X} = \mathbb{R}^N$ and $\mathcal{X} = \mathbb{R}^N_+$. Therefore, our results cover both the case when the optimization Problem (1) is a dual problem associated with equality and inequality constrained primal problems. Our results show that the Iterates (8) using the quantization in Example 1 converge 1) approximately to the set of optimal values when the step sizes are fixed and 2) asymptotically when the step sizes are diminishing and nonsummable. In Section IV, we prove the convergence in the constrained case when $\mathcal{X} = \mathbb{R}^N_+$. The convergence in the unconstrained case $\mathcal{X} = \mathbb{R}^N$ is a special case of the more general convergence results in Section V.

B. Fundamental Limit: Proper Quantization

When the quantization in Example 1 is used in the TCP problem, then there is no collaboration between the network links (or the nodes in the Network flow problem). As a result, $|\mathcal{D}| = 2^N$ and $\log_2(2^N) = N$ bits are used to broadcast the quantized gradient direction per iteration. However, in many applications [8], [9], the dual problem is maintained by a single coordinator. Therefore, an interesting question is: Whether even fewer than N bits can be used per iteration when a single coordinator maintains the dual problem? In that case, what is the minimal quantization $|\mathcal{D}|$ so the Iterates (8) can solve the optimization Problem (1)? More generally, for what quantization sets $\mathcal{D}$ do the Iterates (8) converge to optimal solution to the Problem (1)? To answer such questions, we now formalize how a quantization set $\mathcal{D}$ enables the Iterates (8) to solve the optimization Problem (1).

Definition 2: Consider Iterations (8). A finite set $\mathcal{D}$ is a *proper quantization* for the problem class $\mathcal{F}_L(\mathcal{X})$ if for every $f \in \mathcal{F}_L(\mathcal{X})$ and every initialization $\mathbf{x}(0) \in \mathcal{X}$, we can choose $\mathbf{d}(t) \in \mathcal{D}$ and $\gamma(t) \in \mathbb{R}_+$, for all $t \in \mathbb{N}$, such that $\lim_{t\to\infty} \texttt{dist}\,(\mathbf{x}(t), \mathcal{X}^\star) = 0$.

Using Definition 2, we investigate the following questions.

A) Are there equivalent constructive conditions that can be used to determine whether $\mathcal{D}$ is a *proper quantization* or to construct such quantization sets?

B) What is the minimal quantization, i.e., size $|\mathcal{D}|$, for which $\mathcal{D}$ is a *proper quantization*?

C) What are the connections between the fineness of the quantization, i.e., the size of $|\mathcal{D}|$, and the possible convergence rate of the algorithm?

For the class $\mathcal{F}_L(\mathbb{R}^N)$ of unconstrained problems, the next few paragraphs answer all these questions. However, as shown in Section III-C, a proper quantization set $\mathcal{D}$ for the class $\mathcal{F}_L(\mathbb{R}^N)$ might not be a proper quantization for $\mathcal{F}_L(\mathcal{X})$ when $\mathcal{X}$ is a proper subset of $\mathbb{R}^N$.

Question A): Consider the following definition.

Definition 3: The finite set $\mathcal{D}$ is a θ-cover if $\theta \in [0, \pi/2)$ and for all $\mathbf{g} \in \mathcal{S}^{N-1}$, there is $\mathbf{d} \in \mathcal{D}$ such that $\texttt{ang}(\mathbf{g}, \mathbf{d}) = \cos^{-1}(\langle \mathbf{g}, \mathbf{d} \rangle) \leq \theta$. Equivalently, for all $\mathbf{g} \in \mathcal{S}^{N-1}$, there is $\mathbf{d} \in \mathcal{D}$ such that $\cos(\texttt{ang}(\mathbf{g}, \mathbf{d})) \geq \cos(\theta) > 0$.

Informally, $\mathcal{D}$ is a θ-cover if for any nonzero vector in $\mathbb{R}^N$, there exists some element in $\mathcal{D}$ such that the angle between them is smaller than or equal to θ. The following theorem shows that Definition 3 of θ-cover is actually equivalent to Definition 2 of proper quantization for the problem class $\mathcal{F}_L(\mathbb{R}^N)$.

Theorem 1: Consider a finite set $\mathcal{D} \subseteq \mathcal{S}^{N-1}$. $\mathcal{D}$ is a proper quantization for the class $\mathcal{F}_L(\mathbb{R}^N)$ (see Definition 2) if and only if there exists $\theta \in [0, 2\pi)$ such that $\mathcal{D}$ is a θ-cover (see Definition 3).

Proof: The proof is found in Appendix A. ■

Unlike Definition 2 of proper quantization, Definition 3 is *constructive* in the sense that it can be used to determine if a set $\mathcal{D}$ is a proper quantization and to construct such sets. For example, we can use Definition 3 and Theorem 1 to show that the quantization scheme in Example 1 is a proper quantization for the problem class $\mathcal{F}_L(\mathbb{R}^N)$.

Lemma 1: The quantization in Example 1 is a θ-cover with $\cos(\theta) = 1/\sqrt{N}$.

The lemma follows from the fact that for any $\mathbf{x} \in \mathcal{S}^{N-1}$, if we choose $\mathbf{d} = (1/\sqrt{N})\,\texttt{sign}(\mathbf{x})$, then

$$\cos(\texttt{ang}(\mathbf{x}, \mathbf{d})) = \langle \mathbf{x}, \mathbf{d} \rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{x}_i \cdot \texttt{sign}(\mathbf{x}_i)$$

$$\geq \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{x}_i^2 = \frac{1}{\sqrt{N}} ||\mathbf{x}||^2 = \frac{1}{\sqrt{N}}.$$

Lemma 1 proves that the quantization in Example 1 is a proper quantization. We give some other interesting θ-covers now.

Example 2 (Minimal proper quantization: $|\mathcal{D}| = N + 1$): Set $\mathcal{D} = \{\mathbf{e}_1, \ldots, \mathbf{e}_N, -\mathbf{1}/\sqrt{N}\}$, where $\mathbf{e}_i$ is the ith element of the normal basis and $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^N$. Clearly, $|\mathcal{D}| = N + 1$, so $\mathcal{D}$ can be coded using only $\log_2(N + 1)$ bits. We show below in Theorem 2 that this is a minimal proper quantization, as there does not exist proper quantization $\mathcal{D}$ with $|\mathcal{D}| \leq N$. We show in Lemma 7 in Appendix J that $\mathcal{D}$ is a θ-cover with

$$\cos(\theta) = \frac{1}{\sqrt{N^2 + 2\sqrt{N}(N-1)}}. \tag{9}$$

Example 3 (Example in $\mathbb{R}^2$: $|\mathcal{D}| = n$): For every $n \in \mathbb{N}$ set

$$\mathcal{D}_n = \left\{ \begin{bmatrix} \cos(2\pi k/n) \\ \sin(2\pi k/n) \end{bmatrix} \in \mathbb{R}^2 \middle| k = 0, 1, \ldots, n-1 \right\}.$$

Clearly, if $n \geq 3$, $\mathcal{D}_n$ is a θ-cover with $\theta = \pi/n$.

Example 4 (Normal basis: $|\mathcal{D}| = 2N$): Let $\mathcal{D} = \{\mathbf{e}_1, -\mathbf{e}_1, \mathbf{e}_2, -\mathbf{e}_2, \ldots, \mathbf{e}_N, -\mathbf{e}_N\}$. $|\mathcal{D}| = 2N$ and $\mathcal{D}$ is a θ-cover with $\cos(\theta) = 1/\sqrt{N}$ since for all $\mathbf{x} \in \mathcal{S}^{N-1}$, if we choose $\mathbf{d} = \texttt{sign}(\mathbf{x}_i)\mathbf{e}_i$, where $i = \text{argmax}_{j=1,\ldots,N} |\mathbf{x}_j|$, then $\cos(\texttt{ang}(\mathbf{x}, \mathbf{d})) = \langle \mathbf{x}, \mathbf{d} \rangle = \mathbf{x}_i \cdot \texttt{sign}(\mathbf{x}_i) = |\mathbf{x}_i| \geq 1/\sqrt{N}$.

For constant $\theta \in (0, \pi/2)$, it can be of interest to find the θ-cover $\mathcal{D}$, which has minimal cardinality $|\mathcal{D}|$. This problem has been investigated in the coding theory literature [35], [36].

Question B): The minimal proper quantization for the problem class $\mathcal{F}_L(\mathbb{R}^N)$ is $|\mathcal{D}| = N + 1$. We already have a proper quantization with $|\mathcal{D}| = N + 1$; see Example 2. The following result shows that there does not exist a quantization set $\mathcal{D}$ with fewer elements than $N + 1$.

Theorem 2: Suppose that $\mathcal{D} \subseteq \mathcal{S}^{N-1}$ and $|\mathcal{D}| \leq N$. Then, $\mathcal{D}$ is not a proper quantization.

Proof: The proof is in Appendix B. ■

This result shows that the minimum data rate needed for the algorithm to converge is $\log_2(N + 1)$ bits/iteration. To the best of our knowledge, there are no similar results on minimal quantizations for distributed optimization methods in the existing literature.

Question C): In Section V, we study the convergence of Iterates (8) for the problem class $\mathcal{F}_L(\mathbb{R}^N)$ when $\mathcal{D}$ is a θ-cover. When the step size is constant, i.e., $\gamma(t) = \gamma$, then we show that any solution accuracy $||\nabla f(\mathbf{x})|| \leq \epsilon$ and $f(\mathbf{x}) - f^\star \leq \epsilon$ can be achieved, for $\epsilon > 0$. We also give an upper bound on the number of iterations/bits needed to achieve that accuracy. An implication of the results is that after $T \in \mathbb{N}$ iterations, the accuracy $||\nabla f(\mathbf{x})|| \leq \frac{M}{\cos(\theta)\sqrt{T}}$ can be reached using appropriate constant step-size choice, where $M > 0$ is some constant. Finally, we show how to choose the step sizes so that every limit point of the algorithm is an optimizer of Problem (1).

C. Fundamental Limit: What If There Are Constraints?

We now show by simple examples why a θ-cover $\mathcal{D}$ might not be a proper quantization for the problem class $\mathcal{F}_L(\mathcal{X})$ when the feasible set $\mathcal{X}$ is a proper subset of $\mathbb{R}^N$. These examples are illustrated in Fig. 1(a) and (b). Both figures demonstrate scenarios, where a single step of Iteration (8) is taken from $\mathbf{x}$. In both figures, $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ is a θ-cover with $\theta = \pi/3$. The feasible region is depicted by gray color. The curves depict the contours of the objective function f. The dotted lines depict the angle θ.

Fig. 1(a) depicts a scenario, where Iteration (8) may have a nonoptimal stationary point, even though $\mathcal{D}$ is a θ-cover. The point $\mathbf{x}$ is a stationary point, since $-\mathbf{d}_1$ is orthogonal to the constraint. However, $\mathbf{x}$ is not an optimal solution of Problem (1), since the gradient $\nabla f(\mathbf{x})$ is not orthogonal to the constraint. This example shows that the equivalence established in Theorem 1 does not generalize to the constrained case. Fig. 1(b) shows that the Iterates (8) can go in the opposite direction of the optimal solution. The Iterate (8) is a not a descent direction; hence, the objective function value is increasing. The optimal solution of Problem (1) and the Iterate (8) are denoted by $\mathbf{x}^\star$ and $[\mathbf{x} - \gamma \mathbf{d}_1]_{\mathcal{X}}$, respectively.

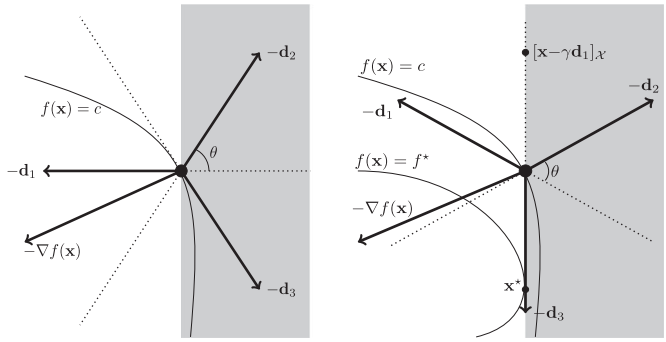

Fig. 1. Iteration of (8) starting at $\mathbf{x}$. $\mathbf{x}$ is in the middle of the figure marked by a big filled circle. The feasible set $\mathcal{X}$ is marked by the shaded region. $\mathbf{x}^\star$ and $f^\star$ are the optimizers and optimal value. $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ is a θ-cover with $\theta = \pi/3$. In (a), $\mathbf{x}$ is a fixed point of the algorithm. In (b), the algorithm takes the step $[\mathbf{x} - \gamma \mathbf{d}_1]_{\mathcal{X}}$.

IV. Convergence—Binary Quantization

In this section, we investigate the convergence of the QGM

$$\mathbf{x}(t+1) = \left[\mathbf{x}(t) - \frac{\gamma(t)}{\sqrt{N}}\, \mathtt{sign}\left(\nabla f(\mathbf{x}(t))\right)\right]^+ \tag{10}$$

for solving the optimization Problem (1) when $\mathcal{X} = \mathbb{R}_+^N$. We make the additional assumption that the gradients ∇f are B-bounded, i.e., $||\nabla f(\mathbf{x})|| \leq B$ for all $\mathbf{x} \in \mathbb{R}_+^N$.[2] In the next section, which considers unconstrained problems, we allow the gradients to be unbounded.

In the analysis, we take advantage of the following property of optimal solution $\mathbf{x}^\star$.

Lemma 2: Consider optimization Problem (1) with $\mathcal{X} = \mathbb{R}_+^N$. For $\alpha > 0$, define the function

$$L_\alpha(\mathbf{x}) = ||\mathbf{x} - \lceil \mathbf{x} - \alpha \nabla f(\mathbf{x}) \rceil^+||. \tag{11}$$

Then, $\mathbf{x} \in \mathcal{X}^\star$ if and only if $L_\alpha(\mathbf{x}) = 0$.

Proof: The proof is found in Appendix C. ■

We investigate the convergence of the iterates in (10) when the step sizes are fixed in Section IV-A and when the step sizes are diminishing in Section IV-B.

A. Constant Step Size

In this section, we study the convergence of Iterates (10) when the step size $\gamma(t)$ is constant, i.e., when $\gamma(t) = \gamma$ for all t. We show that the Iterates (10) can approximately solve optimization Problem (1) up to any ϵ-accuracy, provided that the step size $\gamma > 0$ is small enough. By approximately solving (1), we mean that we can find $\mathbf{x} \in \mathbb{R}_+^N$ that approximately satisfies certain optimality conditions. In particular, we consider the following two types of optimality conditions for Problem (1):

$$\text{Type-1:} \qquad L_\alpha(\mathbf{x}) \leq \epsilon \tag{12}$$

$$\text{Type-2:} \qquad f(\mathbf{x}) - f^\star \leq \epsilon. \tag{13}$$

[2]This assumption is only needed in this section. The dual gradient is generally bounded if the primal problem is strongly convex and has bounded feasible set; see [26, Proposition 6.1.1]. For example, the dual gradient is bounded in the TCP problem in Section II-A.

A point $\mathbf{x}$ is an optimal solution to optimization Problem (1) if and only if (12) [or (13)] holds with $\epsilon = 0$. The Type-1 optimality condition is a generalization of the optimality condition that $||\nabla f(\mathbf{x})|| = 0$ for unconstrained problems. The Type-2 optimality condition simply states that the distance from the obtained objective value to the optimal value is less than ϵ. We now show that both Type-1 and Type-2 approximate optimality conditions can be reached in finite number of iterations.

1) Stopping Condition of Type-1: We start by showing that the Type-1 optimality condition can be reached for all $\epsilon > 0$ in a finite number of iterations. The following lemma is essential in proving the result.

Lemma 3: Suppose $f \in \mathcal{F}_L(\mathbb{R}_+^N)$ and ∇f is B-bounded. Suppose $\epsilon > 0$ and $\alpha \geq 0$ are such that

$$L_\alpha(\mathbf{x}) = ||\mathbf{x} - \lceil \mathbf{x} - \alpha \nabla f(\mathbf{x}) \rceil^+|| \geq \epsilon. \tag{14}$$

Then, the following holds:

$$f\left(\left[\mathbf{x} - \frac{\gamma}{\sqrt{N}}\, \mathtt{sign}\left(\nabla f(\mathbf{x})\right)\right]^+\right) \leq f(\mathbf{x}) - \bar{\delta}(\epsilon, \alpha, \gamma)$$

where $\bar{\delta}(\epsilon, \alpha, \gamma) = (\frac{2\epsilon^2}{L\alpha^2 BN^{3/2}} - \gamma)\frac{L}{2}\gamma$.

Proof: The proof is provided in Appendix D. ■

The lemma shows that if $L_\alpha(\mathbf{x}) > 0$ for some $\mathbf{x} \in \mathbb{R}_+^N$, then the objective function value $f(\mathbf{x})$ will decrease with Iterates (10), provided that the step size $\gamma(t) > 0$ is small enough. We use this intuition to provide an upper bound on the number of iterations needed to achieve the Type-1 approximate optimality.

Theorem 3: Suppose $f \in \mathcal{F}_L(\mathbb{R}_+^N)$, ∇f is B-bounded, $\mathbf{x}(t)$ are generated by (10), and define the set

$$\bar{\mathcal{X}}_\alpha(\epsilon) := \{\mathbf{x} \in \mathbb{R}_+^N \,|\, L_\alpha(\mathbf{x}) \leq \epsilon\}. \tag{15}$$

Then, for any $\epsilon > 0$ and $\gamma \in (0, 2\epsilon^2/(\alpha^2 BN^{3/2}))$, there exists $T \in \mathbb{N} \cup \{0\}$ such that $\mathbf{x}(T) \in \bar{\mathcal{X}}_\alpha(\epsilon)$, with T bounded by

$$T \leq \left\lceil \frac{2(f(\mathbf{x}(0)) - f^\star)\alpha^2 BN^{3/2}}{\gamma(2\epsilon^2 - L\gamma\alpha^2 BN^{3/2})} \right\rceil. \tag{16}$$

The upper bound in (16) is minimized when the step size $\gamma^\star = \epsilon^2/(L\alpha^2 BN^{3/2})$ is used.

Proof: Let $\epsilon > 0$ be given and choose any $\gamma \in (0, 2\epsilon^2/(L\alpha^2 BN^{3/2}))$. Then, from Lemma 3, we have for all $\mathbf{x}(t) \notin \bar{\mathcal{X}}_\alpha(\epsilon)$ that

$$f(\mathbf{x}(t+1)) \leq f(\mathbf{x}(t)) - \bar{\delta}(\epsilon, \alpha, \gamma) \tag{17}$$

where $\bar{\delta}(\epsilon, \alpha, \gamma) > 0$. By recursively applying (17), it follows that if $\mathbf{x}(t) \notin \bar{\mathcal{X}}_\alpha(\epsilon)$ for all $t < s$, then

$$0 \leq f(\mathbf{x}(s)) - f^\star \leq f(\mathbf{x}(0)) - f^\star - s\,\bar{\delta}(\epsilon, \gamma, \theta). \tag{18}$$

Therefore, there must exist $T \leq \lceil (f(\mathbf{x}(0)) - f^\star)/\bar{\delta}(\epsilon, \alpha, \gamma) \rceil$ such that $\mathbf{x}(T) \in \bar{\mathcal{X}}_\alpha(\epsilon)$; otherwise, we can use (18) with $s = \lceil (f(\mathbf{x}(0)) - f^\star)/\bar{\delta}(\epsilon, \alpha, \gamma) \rceil + 1$ to get the contradiction that $f(\mathbf{x}(s)) < f^\star$. By rearranging $T \leq \lceil (f(\mathbf{x}(0)) - f^\star)/\bar{\delta}(\epsilon, \alpha, \gamma) \rceil$, we obtain the bound in (16). The optimal step size $\gamma^\star = \epsilon^2/(L\alpha^2 BN^{3/2})$ comes by maximizing the denominator in (16). ■

The theorem shows that the Type-1 approximate optimality condition can be reached in a finite number of iterations. Moreover, (16) provides an upper complexity bound on the algorithm,

showing the number of iterations needed to reach any $\epsilon > 0$ accuracy. When the step size is $\gamma^\star = \epsilon^2/(L\alpha^2 BN^{3/2})$, then the upper bound in (16) increases proportionally to $1/\epsilon^4$ as ϵ goes to zero. In Section V-A, we show that this bound can be improved for unconstrained problems (where it increases proportionally to $1/\epsilon^2$ as ϵ goes to zero).

2) Stopping Condition of Type-2: We now show that the Type-2 approximate optimality can be reached for any ϵ accuracy. The following lemma is used to obtain the result.

Lemma 4: Suppose that $f \in \mathcal{F}_L(\mathbb{R}^N_+)$, ∇f is B-bounded, and the iterates $\mathbf{x}(t)$ are generated by (10). Then, for all $\epsilon > 0$, $\alpha > 0$, $\gamma(t) \in (0, \bar{\gamma})$, with $\bar{\gamma} = 2\epsilon^2/(\alpha^2 BN^{3/2})$, and $T \in \mathbb{N}$ such that $\mathbf{x}(T) \in \bar{\mathcal{X}}_\alpha(\epsilon)$

$$f(\mathbf{x}(t)) \leq \bar{F}_\alpha(\epsilon) + \frac{L}{2}\bar{\gamma}^2, \text{ for all } t \geq T \tag{19}$$

where $\bar{F}_\alpha : \mathbb{R}_+ \to \mathbb{R} \cup \{\infty\}$ is given by

$$\bar{F}_\alpha(\kappa) = \sup\{f(\mathbf{x}) | \mathbf{x} \in \bar{\mathcal{X}}_\alpha(\kappa)\}. \tag{20}$$

Moreover, there exists $\kappa > 0$ such that for all $\epsilon \in [0, \kappa]$, the following holds: 1) $\bar{\mathcal{X}}_\alpha(\epsilon)$ is bounded; and 2) $\bar{F}_\alpha(\epsilon) < \infty$. It also holds that $\lim_{\epsilon \to 0^+} \bar{F}_\alpha(\epsilon) = f^\star$.

Proof: The proof is provided in Appendix E. ■

The lemma is useful in deriving Type-2 optimality conditions, since it connects the results from Theorem 3 to the quantity $f(\mathbf{x}(t)) - f^\star$ via the function $\bar{F}_\alpha(\kappa)$ defined in (20). In particular, the lemma provides a bound on $f(\mathbf{x}(t))$ that depends on $\bar{F}_\alpha(\epsilon)$ and the step size. Therefore, since $\lim_{\epsilon \to 0^+} \bar{F}_\alpha(\epsilon) = f^\star$, we can enforce $f(\mathbf{x}(t)) - f^\star$ to be arbitrarily small after some time T, i.e., for all $t \geq T$, by choosing a small enough step size. The idea is now formalized.

Theorem 4: Suppose that $f \in \mathcal{F}_L(\mathbb{R}^N_+)$, ∇f is B-bounded, and the iterates $\mathbf{x}(t)$ are generated by (10). Then, for any $\epsilon > 0$, there exists step size $\gamma > 0$ and $T \in \mathbb{N}$ such that $f(\mathbf{x}(t)) - f^\star \leq \epsilon$ for all $t \geq T$.

Proof: The result follows directly from Lemma 4 and Theorem 3. ■

Theorem 4 proves that the Type-2 optimality condition [see (13)] can be achieved for any $\epsilon > 0$ in a finite number of iterations. However, unlike for the Type-1 optimality condition, we could not provide an explicit step-size choice or a bound on the number of iterations needed to achieve the $\epsilon > 0$ accuracy.

B. Diminishing Step Size

We now consider the diminishing step-size case. The following result shows that the step sizes can be chosen so that Iterates (10) converge asymptotically to the optimal solutions of Problem (1).

Theorem 5: Suppose that $f \in \mathcal{F}_L(\mathbb{R}^N_+)$, ∇f is B-bounded, and the iterates $\mathbf{x}(t)$ are generated by (10), where $\lim_{t\to\infty} \gamma(t) = 0$ and $\sum_{t=0}^{\infty} \gamma(t) = \infty$. Then, $\mathbf{x}(t)$ converges to the set of optimal solutions of the optimization Problem (1), i.e., $\lim_{t\to\infty} \texttt{dist}\,(\mathbf{x}(t), \mathcal{X}^\star) = 0$.

Proof: The proof is provided in Appendix F. ■

The step-size choice in the theorem is also necessary to ensure that $\lim_{t\to\infty} \texttt{dist}\,(\mathbf{x}(t), \mathcal{X}^\star) = 0$ holds for all $f \in \mathcal{F}_+(\mathbb{R}^N_+)$, with ∇f being B-bounded. To see why, consider the scalar function

$$f(x) = \begin{cases} 0.5(x-1)^2, & \text{if } |x-1| \leq 1 \\ |x-1| - 0.5, & \text{otherwise.} \end{cases} \tag{21}$$

Then, f has the unique optimizer $x^\star = 1$ and satisfies the assumptions of Theorem 5.

Let us first show that if $\lim_{t\to\infty} \gamma(t) \neq 0$, then $\lim_{t\to\infty} \texttt{dist}\,(x(t), \mathcal{X}^\star) \neq 0$. If $\gamma(t) > 0$ does not converge to zero, then there exists $I \in (0,1)$ and a subsequence t_k such that $\gamma(t_k) \geq I$ for infinitely many $t \in \mathbb{N}$. Then, either $|x(t_k+1) - x^\star|$ or $|x(t_k) - x^\star|$ must be larger than or equal to $I/2$ for all $k \in \mathbb{N}$ because if $|x(t_k) - x^\star| \leq I/2$, then $|x(t_k+1) - x^\star| \geq I/2$. Therefore, $|x(t) - x^\star| \geq I/2$ for infinitely many $t \in \mathbb{N}$, so $\limsup_{t\to\infty} \texttt{dist}\,(x(t), \mathcal{X}^\star) \geq I/2$. Let us next show that if $\sum \gamma(t) < \infty$, then $\lim_{t\to\infty} \texttt{dist}\,(x(t), \mathcal{X}^\star) \neq 0$. Take $x(0) = 2 + \sum_{t=0}^{\infty} \gamma(t)$. Then, $x(t) \geq \mathbf{x}(0) - \sum_{\tau=0}^{t} \gamma(t) \geq 2$ for all $t \in \mathbb{N}$, so $\liminf_{t\to\infty} \texttt{dist}\,(x(t), \mathcal{X}^\star) \geq 1$.

The advantage of using diminishing step sizes, as in Theorem 5, is that the algorithm can asymptotically converge to the set of optimal values. Moreover, diminishing step-size rules can be implemented even if problem parameters such as the Lipschitz constant L are unknown, unlike when the step size is constant. On the other hand, it is more complicated to characterize the convergence rate, similar to (16), when diminishing step sizes are used.

V. Convergence—General Quantization

In the previous section, we studied QGMs, where a particular quantization based on the sign of the gradient was used for constrained optimization problems. As we discussed in Section III-B, for unconstrained problems, a more general class of quantizations called θ-covers (see Definition 3) ensures that the QGMs can minimize any $f \in \mathcal{F}_L(\mathbb{R}^N)$. In this section, we formally prove this, i.e., if the quantization is a θ-cover, then the QGMs converge 1) approximately to an optimal solution when the step sizes are constant and 2) asymptotically to an optimal solution when the step sizes are nonsummable and converge to zero. Moreover, we study how the quantization fineness, i.e., θ, affects the algorithm convergence. We first consider the case when the step sizes are fixed, i.e., $\gamma(t) = \gamma$, in Section V-A. Then, in Section V-B, we consider diminishing step sizes.

A. Constant Step Size

Similar to Section IV-A, we consider the following two types of approximate optimality conditions:

$$\text{Type-1:} \qquad ||\nabla f(\mathbf{x})|| \leq \epsilon \tag{22}$$

$$\text{Type-2:} \qquad f(\mathbf{x}(t)) - f^\star \leq \epsilon. \tag{23}$$

1) Stopping Condition of Type-1: We start by showing that Type-1 approximate optimality can be achieved for any $\epsilon > 0$ in a finite number of iterations. Furthermore, we provide a lower and an upper bound on the number of iterations needed to achieve the ϵ-accuracy (that depends on θ). A key result used to obtain the result is the following lemma.

Lemma 5: Suppose $f \in \mathcal{F}_L(\mathbb{R}^N)$, $\epsilon > 0$, $\theta \in [0, \pi/2)$, $\mathbf{x} \in \mathbb{R}^N$, $||\nabla f(\mathbf{x})|| > \epsilon$, and $\mathbf{d} \in \mathcal{S}^{N-1}$, where $\texttt{ang}(\nabla f(\mathbf{x}), \mathbf{d}) \leq \theta$. Then

$$f(\mathbf{x} - \gamma \mathbf{d}) \leq f(\mathbf{x}) - \delta(\epsilon, \gamma, \theta) \tag{24}$$

where $\delta(\epsilon, \gamma, \theta) = (\frac{2\cos(\theta)\epsilon}{L} - \gamma)\frac{L}{2}\gamma$. Clearly, $\delta(\epsilon, \gamma, \theta) > 0$ when $\gamma \in (0, 2\cos(\theta)\epsilon/L)$.

Proof: The proof is provided in Appendix G. ■

The lemma shows that if $||\nabla f(\mathbf{x})|| > 0$ and $\texttt{ang}(\nabla f(\mathbf{x}), \mathbf{d}) \leq \theta < \pi/2$ for some $\mathbf{x} \in \mathbb{R}^N$, then the objective function value can be decreased by taking a step in the direction $-\mathbf{d}$, i.e., $f(\mathbf{x} - \gamma \mathbf{d}) \leq f(\mathbf{x})$ for small enough $\gamma > 0$. Therefore, if $\mathcal{D}$ is a θ-cover, then we can always find $\mathbf{d} \in \mathcal{D}$ and a step size $\gamma > 0$ such that $f(\mathbf{x} - \gamma \mathbf{d}) \leq f(\mathbf{x})$. We now use this intuition from Lemma 5 to provide the upper and lower bounds on the number of iterations that are needed to reach the Type-1 optimality condition.

Theorem 6: Suppose that $f \in \mathcal{F}_L(\mathbb{R}^N)$, $\mathcal{D}$ is a θ-cover (see Definition 3), the iterates $\mathbf{x}(t)$ are generated by (8), and define the set

$$\mathcal{X}(\epsilon) = \{\mathbf{x} \in \mathbb{R}^N \,|\, ||\nabla f(\mathbf{x})|| \leq \epsilon\}. \tag{25}$$

Then, the following holds:

a) For any $\epsilon > 0$, if $\gamma \in (0, 2\cos(\theta)\epsilon/L)$, then there exists $T \in \mathbb{N} \cup \{0\}$ such that $\mathbf{x}(T) \in \mathcal{X}(\epsilon)$, with T bounded by

$$T \leq \left\lceil \frac{2(f(\mathbf{x}(0)) - f^\star)}{\gamma(2\cos(\theta)\epsilon - L\gamma)} \right\rceil. \tag{26}$$

The upper bound in (27) is minimized with the optimal step size $\gamma^\star = \cos(\theta)\epsilon/L$.

b) Given a fixed step size $\gamma > 0$ and scalar $\kappa > 0$, if we choose $\epsilon(\kappa, \gamma) = \kappa + \gamma L/(2\cos\theta)$ then there exists $T \in \mathbb{N} \cup \{0\}$ such that $\mathbf{x}(T) \in \mathcal{X}(\epsilon(\kappa, \gamma))$, with T bounded by

$$T \leq \left\lceil \frac{f(\mathbf{x}(0)) - f^\star}{\cos(\theta)\gamma\kappa} \right\rceil. \tag{27}$$

c) (Lower Bound on T) For any step size $\gamma > 0$ and $\epsilon > 0$ if $\mathbf{x}(T) \in \mathcal{X}(\epsilon)$, then $\frac{||\nabla f(\mathbf{x}(0))|| - \epsilon}{\gamma L} \leq T$.

Proof: a) The proof follows the same arguments as the proof of Theorem 3 using Lemma 5 in place of Lemma 3.

b) The result can be obtained by substituting $\epsilon(\kappa, \gamma)$ in (27).

c) Using the fact that the gradient ∇f is L-Lipschitz continuous, we have

$$||\nabla f(\mathbf{x}(t)) - \nabla f(\mathbf{x}(t+1))|| \leq L||\mathbf{x}(t) - \mathbf{x}(t+1)|| \leq L\gamma.$$

Therefore, using the triangle inequality, we have

$$||\nabla f(\mathbf{x}(t))|| - L\gamma \leq ||\nabla f(\mathbf{x}(t+1))|| \quad \text{for all } t \in \mathbb{N}.$$

Recursively applying the inequality gives

$$||\nabla f(\mathbf{x}(0))|| - L\gamma t \leq ||\nabla f(\mathbf{x}(t))||.$$

Hence, $||\nabla f(\mathbf{x}(t))|| \leq \epsilon$ can only hold when $t \geq (||\nabla f(\mathbf{x}(0))|| - \epsilon)/(L\gamma)$. ■

Theorem 6(a) proves that if $\mathcal{D}$ is a θ-cover, then the Type-1 optimality condition [see (22)] can be achieved with ϵ-accuracy in a finite number of iterations, for all $\epsilon > 0$. Moreover, the theorem gives an upper bound on the number of iterations needed to achieve such ϵ-accuracy. This bound decreases as θ decreases, i.e., as the quantization becomes finer. Even though the bound in (27) is on the number of iteration, since $\log_2(|\mathcal{D}|)$ bits are communicated per iteration, the results shows that in the worst-case scenario

$$\left\lceil \frac{2(f(\mathbf{x}(0)) - f^\star)}{\gamma(2\cos(\theta)\epsilon - L\gamma)} \right\rceil \log_2(|\mathcal{D}|) \text{ bits}$$

are needed to find $\mathbf{x} \in \mathbb{R}^N$ such that $||\nabla f(\mathbf{x})|| \leq \epsilon$.

Theorem 6(b) demonstrates what ϵ-accuracy can be achieved for a given step size. The parameter κ captures a tradeoff between the ϵ-accuracy and the number of iterations needed to achieve that ϵ-accuracy. By optimizing over both γ and κ in Theorem 6(b), we can find an optimal bound on the accuracy ϵ that can be guaranteed in T iterations. That is to find γ and κ that solve the following optimization problem:

$$\begin{aligned} \underset{\kappa,\gamma}{\text{minimize}} \quad & \epsilon(\kappa, \gamma) = \kappa + \frac{L}{2\cos(\theta)}\gamma \\ \text{subject to} \quad & \frac{f(\mathbf{x}(0)) - f^\star}{\cos(\theta)\gamma\kappa} \leq T, \\ & \gamma, \kappa > 0. \end{aligned} \tag{28}$$

Formally, this bound is given as follows.

Corollary 1: For any $T \in \mathbb{N}$, we have the following.

i) The minimal bound $\epsilon(\kappa, \gamma)$ achieved in T iterations, i.e., the solution to Problem (29), is

$$\epsilon^\star = \frac{\sqrt{2L(f(\mathbf{x}(0)) - f^\star)}}{\cos(\theta)\sqrt{T}} \tag{29}$$

where the corresponding optimal γ and κ are

$$\gamma^\star = \sqrt{\frac{2(f(\mathbf{x}(0)) - f^\star)}{LT}} \text{ and } \kappa^\star = \frac{\sqrt{L(f(\mathbf{x}(0)) - f^\star)}}{\cos(\theta)\sqrt{2T}}. \tag{30}$$

ii) If the step size γ is chosen as $\gamma^\star$ in (31), then

$$||\nabla f(\mathbf{x}^\star(T))|| \leq \epsilon^\star \tag{31}$$

where $\epsilon^\star$ is as in (30) and

$$\mathbf{x}^\star(T) = \underset{\mathbf{x} \in \{\mathbf{x}(t) | t = 0, \ldots, T\}}{\text{argmin}} ||\nabla f(\mathbf{x})||. \tag{32}$$

Proof: i) First note that Problem (29) is convex, which can be seen by equivalently writing it as

$$\begin{aligned} \underset{\kappa,\gamma}{\text{minimize}} \quad & \kappa + \frac{L}{2\cos(\theta)}\gamma \\ \text{subject to} \quad & \left(\frac{f(\mathbf{x}(0)) - f^\star}{\cos(\theta)T}\right)\frac{1}{\gamma} - \kappa \leq 0, \\ & \gamma, \kappa > 0 \end{aligned} \tag{33}$$

and recalling that the reciprocal $1/\gamma$ is a convex function for $\gamma > 0$. It can be checked that $\kappa^\star$ and $\gamma^\star$ satisfy the Karush–Kuhn–Tucker (KKT) condition with the Lagrangian multiplier $\lambda^\star = 1$.

ii) Follows directly from part i). ■

In addition to minimizing the bound $\epsilon(\kappa,\gamma)$, Corollary 1 gives insights into the convergence of Iterations (8). For example, when T is fixed, then the upper bound in (30) gets larger as θ decreases. As a result, when the quantization set $\mathcal{D}$ becomes coarser, then less accuracy can be ensured. Moreover, the results show that $||\nabla f(\mathbf{x}^\star(T))||$ converges, in the worst case, at the rate $1/\sqrt{T}$ to 0.

In Corollary 1, we used the step size $\gamma^\star$ given in (31). To compute $\gamma^\star$, the optimal objective function value $f^\star$ is needed, which is usually not available prior to solving Problem (1). However, some bounds on the quantity $f(\mathbf{x}(0))-f^\star$ are often available. Any such upper bound $K\in\mathbb{R}$, with $K\geq f(\mathbf{x}(0))-f^\star$, can be used to obtain similar results as to those in Corollary 1 by replacing $f(\mathbf{x}(0))-f^\star$ by K.

Corollary 2: Take $T\in\mathbb{N}$ and $K\in\mathbb{R}$ such that $K\geq f(\mathbf{x}(0))-f^\star$. If we choose the step size as $\gamma=2K/(LT)$, then

$$||\nabla f(\mathbf{x}^\star(T))||\leq\frac{\sqrt{2LK}}{\cos(\theta)\sqrt{T}} \tag{34}$$

where $\mathbf{x}^\star(t)$ is chosen as in (33).

We next demonstrate how the convergence results translate to Type-2 stopping conditions [see (23)].

2) Stopping Condition of Type-2: We now show that the Type-2 approximate optimality [see (23)] can be achieved for any accuracy $\epsilon>0$. The result is based on the following lemma.

Lemma 6: Suppose $f\in\mathcal{F}_L(\mathbb{R}^N)$, $\mathcal{D}$ is a θ-cover, and the iterates $\mathbf{x}(t)$ are generated by (8), then:

a) for any $\epsilon>0$, $\gamma(t)\in(0,\bar{\gamma})$, where $\bar{\gamma}=2\cos(\theta)\epsilon/L$, and $T\in\mathbb{N}$ such that $\mathbf{x}(T)\in\mathcal{X}(\epsilon)$, the following holds:

$$f(\mathbf{x}(t))\leq F(\epsilon)+\frac{L}{2}\gamma^2,\quad\text{for all } t\geq T \tag{35}$$

where $F:\mathbb{R}_+\to\mathbb{R}\cup\{\infty\}$ is given by

$$F(\kappa)=\sup\{f(\mathbf{x})|\mathbf{x}\in\mathcal{X}(\kappa)\}. \tag{36}$$

There exists $\kappa>0$ such that for all $\epsilon\in[0,\kappa]$, $\mathcal{X}(\epsilon)$ is bounded and $F(\epsilon)<\infty$. Moreover, $\lim_{\epsilon\to0^+}F(\epsilon)=f^\star$.

b) if f is μ-strongly convex, then we have

$$F(\epsilon)\leq f^\star+\epsilon^2/(2\mu). \tag{37}$$

Proof: The proof is provided in Appendix H. ■

Lemma 6 is useful in obtaining Type-2 approximate optimality, as it connects the quantity $f(\mathbf{x}(t))-f^\star$ to Theorem 6 via the function $F(\epsilon)$ in (37). In particular, part a) of Lemma 6 bounds $f(\mathbf{x}(t))$ by a constant that depends on $F(\cdot)$, where $F(\epsilon)$ converges to $f^\star$ as ϵ converges to 0. Therefore, by using the dependence of $F(\epsilon)$ on $\mathcal{X}(\epsilon)$, defined in (26), we can connect $f(\mathbf{x}(t))-f^\star$ to the convergence result in Theorem 6 to ensure that the Type-2 stopping condition can be achieved for any $\epsilon>0$. Part b) of Lemma 6 then illustrates how the upper bound on $f(\mathbf{x}(t))$ depending on $F(\epsilon)$ can be further improved when f is μ-strongly convex. These ideas are formally illustrated in the following theorem.

Theorem 7: Suppose $f\in\mathcal{F}_L(\mathbb{R}^N)$, $\mathcal{D}$ is a θ-cover, and $\mathbf{x}(t)$ are generated by (8). Then, for any $\epsilon>0$:

a) there exists a step size $\gamma>0$ and $T\in\mathbb{N}$ such that $f(\mathbf{x}(t))-f^\star<\epsilon$ for all $t\geq T$,

b) moreover, if f is μ-strongly convex and $\gamma\in(0,\bar{\gamma})$, where

$$\bar{\gamma}=\min\left\{\frac{2\cos(\theta)\sqrt{\mu\epsilon}}{L},\sqrt{\frac{\epsilon}{L}}\right\}$$

then $f(\mathbf{x}(t))-f^\star\leq\epsilon$ for all $t\geq T$, where

$$T\leq\left\lceil\frac{2(f(\mathbf{x}(0))-f^\star)}{\gamma(2\cos(\theta)\sqrt{\mu\epsilon}-L\gamma)}\right\rceil. \tag{38}$$

Proof: a) The result follows directly from Lemma 6(a) and Theorem 6(a).

b) Since $\gamma\in(0,2\cos(\theta)\sqrt{\mu\epsilon/L})$, it follows from Theorem 6(a) that there exists T bounded as in (27) such that $\mathbf{x}(T)\in\mathcal{X}(\sqrt{\mu\epsilon})$. From Lemma 6(a), for all $t\geq T$ that

$$f(\mathbf{x}(t))-f^\star\leq F(\sqrt{\mu\epsilon})-f^\star+\frac{L}{2}\gamma^2\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon,$$

where the second inequality follows from Lemma 6(b) and that $\gamma\leq\sqrt{\epsilon/L}$. ■

Theorem 7 shows that when $\mathcal{D}$ is a θ-cover, the Type-2 optimality condition [see (23)] can be achieved in a finite number of iterations. For general functions, the theorem does not provide a step size γ_ϵ that can achieve any particular ϵ-accuracy, even though such γ_ϵ always exist. This is challenging in general, as it can be difficult to bound the function $F(\cdot)$ for general convex functions $f(\cdot)$. Nevertheless, part b) of the proof shows that when f is μ-strongly convex, then a range of step sizes that ensure a given $\epsilon>0$ accuracy is provided. Moreover, when f is μ-strongly convex, then we can obtain similar bound on number of iterations needed to achieve that ϵ-accuracy as in (27) in Theorem 6(a).

B. Diminishing Step Size

We now consider the diminishing step-size case. The following result shows that the step sizes can actually be chosen so Iterates (8) converges to the optimal solution to Problem (1).

Theorem 8: Suppose that $f\in\mathcal{F}_L(\mathbb{R}^N)$, $\mathcal{D}$ is a θ-cover, and that the iterates $\mathbf{x}(t)$ are generated by (8). If the step sizes $\gamma(t)\geq0$ are chosen so that $\lim_{t\to0}\gamma(t)=0$ and $\sum_{t=0}^N\gamma(t)=\infty$, then $\lim_{t\to\infty}\texttt{dist}\,(\mathbf{x}(t),\mathcal{X}^\star)=0$.

Proof: The proof is provided in Appendix I. ■

The step-size choice in the theorem is necessary to ensure that $\lim_{t\to\infty}\texttt{dist}\,(\mathbf{x}(t),\mathcal{X}^\star)=0$ for all $f\in\mathcal{F}_L(\mathbb{R}^N)$, consider the scalar function f defined in (21).

Theorem 8 shows that when $\mathcal{D}$ is a θ-cover, then there exists a step-size rule such that every limit point of the QGMs is an optimal solution to Problem (1). A particular implication of this result is that every θ-cover is a proper quantization; see Definition 2. Therefore, Theorem 8 actually proves one direction of the equivalence established in Theorem 1.

VI. Numerical Illustrations

We now illustrate how the studied algorithms perform on two of the application examples discussed in Section II. We compare the numerical performance with some of the theoretical results

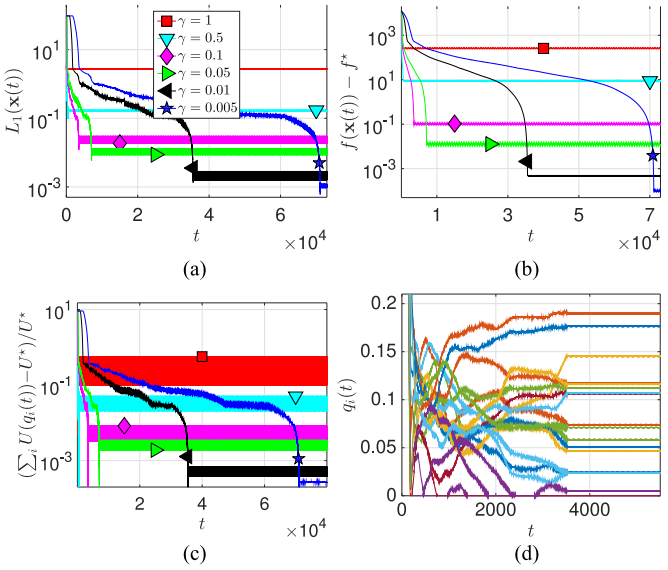

Fig. 2. Optimality measure $L_1(\cdot)$ from (11) (a), dual (b), and primal (c) objective function value at every iteration for different step sizes. The resource allocations $q_i(t)$ (d) at every iteration when $\gamma = 0.1$. (a) Optimality condition [see (11)]. (b) Dual objective function. (c) Relative primal objective value. (d) Allocation $\mathbf{q}(t)$ when $\gamma = 0.1$.

in this paper and with algorithms that use perfect communication with no quantization.

A. TCP Flow Control With Binary Feedback

We illustrate the convergence of Iterates (10) on the TCP flow control in Section II-A. We consider a network with $S = 20$ sources and $N = 100$ links. We use the same utility functions as in Experiment 1 of [7, Sec. VI-B], i.e., $U_s(q_s) = 1000 \log(1 + q_s)$, for $s \in \mathcal{S}$. Similarly, as in [28], we generate the network matrix $\mathbf{A}$ [see (5)] randomly so each entry of $\mathbf{A}$ is 1 with the probability $1/2$ and 0 otherwise. We set $c_l = 1$ for all $l \in \mathcal{L}$. The local constraint of each source is $[0, 1]$. We consider the step sizes $\gamma = 0.005$, 0.01, 0.05, 0.1, 0.5, and 1. Note that in the figures described below, some of lines that appear to be thick lines actually show some fluctuations.

Fig. 2(a) depicts the optimality measure $L_1(\mathbf{x}(t))$; see (11) in Lemma 2. From Lemma 2, $\mathbf{x} \in \mathbb{R}^N_+$ is an optimal solution to Problem (1) if and only if $L_1(\mathbf{x}) = 0$. For all step sizes γ, the measure $L_1(\mathbf{x}(t))$ converges to some small error floor and then fluctuates slightly there. For smaller step sizes γ, the optimality measure $L_1(\mathbf{x}(t))$ converges to smaller values, roughly to 2.6, 0.16, 0.019, 0.009, 0.002, 0.001 for $\gamma = 1$, 0.5, 0.01, 0.005, 0.001, and 0.0005. These results show that the step-size choices γ in Theorem 3 are conservative. For example, to ensure $\epsilon = 0.1$ accuracy in Theorem 3, the step sizes should be $\gamma \in (0, 0.00002)$, but in this example, the step size chooses $\gamma \leq 0.1$ and achieves the $\epsilon = 0.1$ accuracy.

Fig. 2(b) and (c) depicts the dual and primal objective function values at every iteration. The figures demonstrate a similar convergence behavior of the primal/dual objective function values as in the optimality measure $L_1(\mathbf{x}(t))$ in Fig. 2(a). For the dual objective value, these results agree with the results in Theorem 4. Finally, Fig. 2(d) illustrates the convergence of the data rate allocation to each source when γ. The results show that the all the data rate allocations converge after roughly 3500 iterations and then fluctuate slightly there.

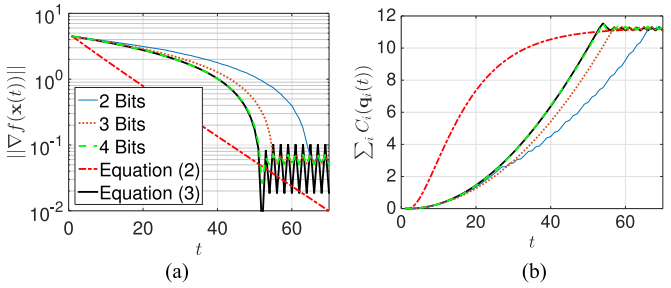

Fig. 3. (a) Dual gradient and (b) primal objective function value at every iteration when 2, 3, and 4 bits are communicated per iteration. The results are compared with the algorithms in (2) and (3). (a) Dual gradient. (b) Objective function.

B. Task Allocation

We now illustrate the performance of the QGMs on the Task Allocation Problem (4) from Section II-A with $K = 4$ machines and $N = 2$ tasks. For each machine $i = 1, 2, 3, 4$, we have the cost function $C_i(\mathbf{q}_i) = a_i \mathbf{q}_{i,1}^2 + b_i \mathbf{q}_{i,2}^2$, where a_i and b_i are uniform random variables on the interval $[1, 5]$. The private constraint of machine $i = 1, 2, 3, 4$ is $\mathcal{Q}_i = \{(x, y) \in \mathbb{R}^2 | x, y \geq 0, x + y \leq 3\}$. Clearly, $-C_i$ are strongly concave with concavity parameters $\mu_i = \min\{a_i, b_i\} \geq 1$. It can be verified that the dual gradient is L-Lipschitz continuous with $L = 4/\mu$, where $\mu = \min\{\mu_1, \ldots, \mu_4\}$. The step size is $\gamma = 0.1$ and the initialization is $\mathbf{x}(0) = (0, 0)$ (recall that $\mathbf{x}$ is the dual variable). We use the quantization set $\mathcal{D}$ from Example 3 when 2, 3, and 4 bits are communicated per iteration, i.e., when $|\mathcal{D}| = 4, 8, 16$, see Remark 1.

Fig. 3 depicts the norm of the gradient and the primal objective function at every iteration of the algorithm. The norm of the gradient $||\nabla f||$ reaches the accuracy $\epsilon = 0.1$ in roughly 51, 56, and 65 iterations using 204, 168, and 130 bits when 2, 3, and 4 bits are communicated per iteration, respectively. We compare the results to Iterations (2) and (3), where no quantization is done, i.e., infinite bandwidth is used. Fig. 3(a) shows that by using 4 bits per iteration, the results achieved by the QGM are almost as good as when the full gradient direction is communicated using Iterations (3). However, the QGMs do not perform as well as Iterations (2); this is to be expected, since in Iterations (2) the full direction and magnitude of the gradient is known. These results illustrate that we can dramatically reduce the number of bits communicated without sacrificing much in performance.

VII. Conclusion and Future Work

This paper studied gradient methods where the gradient direction is quantized at every iteration of the algorithm. Such methods are of interest, for example, in distributed optimization, where the gradient can often be measured but has to be com-

municated to accomplish the algorithm. An instance of such a procedure is dual decomposition, where primal problems that are scattered between different entities are solved in a distributed fashion by performing a gradient descent on the dual problem; see Section II for examples. Our results show that a variant of the projected dual descent taking the sign of the dual gradients, i.e.,

$$\mathbf{x}(t+1) = [\mathbf{x}(t) - \gamma(t)\ \mathtt{sign}\left(\nabla f(\mathbf{x}(t))\right]_{\mathcal{X}} \tag{39}$$

converges to the optimal solution under mild conditions on f when $\mathcal{X} = \mathbb{R}^N$ and $\mathcal{X} = \mathbb{R}^N_+$, i.e., for dual problems associated with primal problems with equality and inequality constraints, respectively. Therefore, when different entities maintain the components of the variable $\mathbf{x}$, then each entity only needs to broadcast one bit per iteration to ensure convergence to the optimal solution. Our results also show that when a single entity maintains $\mathbf{x}$, then the minimal quantization has cardinality $N+1$; for smaller quantizations, there exists an optimization problem that the QGMs cannot solve. Therefore, only $\log_2(N+1)$ bits/iteration are communicated instead of N bits/iteration as when the components of $\mathbf{x}$ are maintained by different entities in (40). We also connect fineness of the quantization to the convergence rate of the algorithm to the available bandwidth (bits/iteration). The convergence rate improves as the bandwidth is increased.

Future work will consider how to additionally quantize the magnitude of the gradient to get a better tradeoff between convergence rate and available bandwidth. Moreover, it is interesting to see if the results can be generalized to nonsmooth optimization problems.

Appendix A
Proof of Theorem 1

Proof: Let us start by showing by contradiction that $\mathcal{D}$ being a proper quantization for the problem class $\mathcal{F}_L(\mathbb{R}^N)$ implies that there exists $\theta \in [0, \pi/2)$ such that $\mathcal{D}$ is a θ-cover. Suppose there does not exists such θ. Then, $\min_{\mathbf{a}\in\mathcal{S}^{N-1}} \max_{\mathbf{d}\in\mathcal{D}} \cos(\ \mathtt{ang}\left(\mathbf{a},\mathbf{d}\right))\leq 0$, since the function $g(\mathbf{a}) = \max_{\mathbf{d}\in\mathcal{D}} \cos(\ \mathtt{ang}\left(\mathbf{a},\mathbf{d}\right)) = \max_{\mathbf{d}\in\mathcal{D}}\langle \mathbf{a},\mathbf{d}\rangle$ is continuous and $\mathcal{S}^{N-1}$ is compact. Therefore, there exists $\mathbf{a}\in\mathcal{S}^{N-1}$ such that $\cos(\ \mathtt{ang}\left(\mathbf{a},\mathbf{d}\right))\leq 0$ for all $\mathbf{d}\in\mathcal{D}$. In particular, we have for all $\mathbf{d}\in\mathcal{D}$ that $\langle \mathbf{a},\mathbf{d}\rangle = \cos(\ \mathtt{ang}\left(\mathbf{a},\mathbf{d}\right))\leq 0$. By choosing $\mathbf{x}(0)=\mathbf{a}$, using Iterations (8) and Cauchy–Schwarz inequality, we conclude that for all $t\in\mathbb{N}$

$$||\mathbf{x}(t)|| \geq \langle \mathbf{a},\mathbf{x}(t)\rangle = \langle \mathbf{a},\mathbf{a}\rangle - \sum_{i=0}^{t-1}\gamma(t)\langle \mathbf{a},\mathbf{d}(t)\rangle \geq 1$$

where the inequality follows from the fact that $||\mathbf{a}||=1$ and that for all $\mathbf{d}\in\mathcal{D}$, we have $\langle \mathbf{a},\mathbf{d}\rangle\leq 0$. If we choose $f(\mathbf{x}) = (L/2)||\mathbf{x}||^2$, then $f\in\mathcal{F}_L(\mathbb{R}^N)$ and f has the unique optimizer $\mathbf{x}^\star=\mathbf{0}$, but $\mathtt{dist}\left(\mathbf{x}(t),\mathcal{X}^\star\right) = ||\mathbf{x}(t)||\geq 1$, for all $t\in\mathbb{N}$. Since $\mathtt{dist}\left(\mathbf{x}(t),\mathcal{X}^\star\right)\geq 1$ for all $\mathbf{d}(t)\in\mathcal{D}$ and $\gamma(t)\in\mathbb{R}_+$, we can conclude that $\mathcal{D}$ is not a proper quantization.

The fact that $\mathcal{D}$ being a θ-cover implies that $\mathcal{D}$ is a proper quantization follows from Theorem 8 in Section V-B, where we showed that for all $f\in\mathcal{F}_L(\mathbb{R}^N)$, we can choose $\mathbf{d}(t)\in\mathcal{D}$ and $\gamma(t)\in\mathbb{R}_+$ such that $\lim_{t\to\infty}\ \mathtt{dist}\left(\mathbf{x}(t),\mathcal{X}^\star\right)=0$. ■

Appendix B
Proof of Theorem 2

Proof: First, consider the case where either $|\mathcal{D}|<N$ or $|\mathcal{D}|=N$ and the elements of $\mathcal{D}$ are linearly dependent. Then, $\mathtt{Span}(\mathcal{D})$ is a proper subspace of $\mathbb{R}^N$, so there exists a normal $\mathbf{a}\in\mathcal{S}^{N-1}$ such that $\cos(\ \mathtt{ang}\left(\mathbf{a},\mathbf{d}\right)) = \langle \mathbf{a},\mathbf{d}\rangle\leq 0$ for all $\mathbf{d}\in\mathtt{Span}(\mathcal{D})$. Since $\mathcal{D}\subseteq\mathtt{Span}(\mathcal{D})$, $\mathcal{D}$ is not a θ-cover for any $\theta\in[0,\pi/2)$, and the result follows from Theorem 1.

Let us next consider the other case, where $|\mathcal{D}|=N$ and the vectors of $\mathcal{D}$ are linearly independent, i.e., $\mathtt{Span}(\mathcal{D})=\mathbb{R}^N$. Define $\mathbf{D}\in\mathbb{R}^{N\times N}$ such that for $i=1,\dots,N$, row i in $\mathbf{D}$ is the ith element of $\mathcal{D}$, where the elements have some arbitrary order. Then, $\mathbf{D}$ is invertible, and we can choose $\mathbf{a}=-\mathbf{D}^{-1}\mathbf{1}$, where $\mathbf{1}\in\mathbb{R}^N$ is a vector of all ones. Then, we have for $i=1,\dots,N$ that $\langle \mathbf{d}_i,\mathbf{a}\rangle = -\mathbf{d}_i\mathbf{D}^{-1}\mathbf{1} = -1$. Hence, as in the previous case, we get that $\langle \mathbf{a},\mathbf{d}\rangle\leq 0$ for all $\mathbf{d}\in\mathcal{D}$ implying that $\mathcal{D}$ cannot be a θ-cover for any $\theta\in[0,\pi/2)$, and the result follows from Theorem 1. ■

Appendix C
Proof of Lemma 2

Proof: Since f and $\mathcal{X}$ are convex, $\mathbf{x}\in\mathcal{X}^\star$ if and only if the KKT optimality conditions hold for $\mathbf{x}$ [32, Sec. 5.5]. It can be checked that since $\mathcal{X}=\mathbb{R}^N_+$, the optimal dual variable associated with $\mathbf{x}\in\mathcal{X}^\star$ is $\boldsymbol{\lambda}=\nabla f(\mathbf{x})$. Therefore, the KKT conditions reduce to the following three conditions holding for all $i=1,\dots,N$: (i) $\nabla_i f(\mathbf{x})\mathbf{x}_i=0$; (ii) $\nabla_i f(\mathbf{x})\geq 0$; and (iii) $\mathbf{x}_i\geq 0$. We now show both directions of the proof.

First, assume that $L(\mathbf{x})=0$. We show that (i)–(iii) hold, so $\mathbf{x}\in\mathcal{X}^\star$. We have $\mathbf{x}=\lceil \mathbf{x}-\alpha\nabla f(\mathbf{x})\rceil^+$ or $\mathbf{x}_i=\lceil \mathbf{x}_i-\alpha\nabla_i f(\mathbf{x})\rceil^+$, for $i=1,\dots,N$. So (i) holds because if $\mathbf{x}_i\neq 0$, then $\nabla_i f(\mathbf{x})=0$, and if $\nabla_i f(\mathbf{x})\neq 0$, then $\mathbf{x}_i=0$. Similarly, (ii) holds because if $\mathbf{x}_i\neq 0$, then $\nabla_i f(\mathbf{x})=0$, and if $\mathbf{x}_i=0$, then $\nabla_i f(\mathbf{x})\geq 0$. Finally, (iii) holds because $\lceil\cdot\rceil^+$ is the projection to $\mathbb{R}_+$.

Now, assume $\mathbf{x}\in\mathcal{X}^\star$, so (i)–(iii) above hold. If $\mathbf{x}_i=0$ for some $i=1,\dots,N$, then $\mathbf{x}_i=\lceil \mathbf{x}_i-\alpha\nabla_i f(\mathbf{x})\rceil^+$, since by (ii), $\nabla_i f(\mathbf{x})\geq 0$. Otherwise, if $\mathbf{x}_i>0$, then $\nabla_i f(\mathbf{x})=0$ by (i), so $\mathbf{x}_i=\lceil \mathbf{x}_i-\alpha\nabla_i f(\mathbf{x})\rceil^+$. ■

Appendix D
Proof of Lemma 3

Proof: For all $\beta\in[0,1]$, we have

$$\frac{\beta\epsilon}{\sqrt{N}} \leq \frac{\beta}{\sqrt{N}}||\mathbf{x}-\lceil \mathbf{x}-\alpha\nabla f(\mathbf{x})\rceil^+|| \tag{40}$$

$$\leq \beta||\mathbf{x}-\lceil \mathbf{x}-\alpha\nabla f(\mathbf{x})\rceil^+||_\infty \tag{41}$$

$$= \beta|\mathbf{x}_i-\lceil \mathbf{x}_i-\alpha\nabla_i f(\mathbf{x})\rceil^+| \tag{42}$$

$$\leq |\mathbf{x}_i-\lceil \mathbf{x}_i-\alpha\beta B\ \mathtt{sign}\left(\nabla_i f(\mathbf{x})\right)\rceil^+| \tag{43}$$

where, in (42), $i=\mathrm{argmax}_j|x_j(t)-\lceil x_j(t)-\alpha\nabla_j f(x(t))\rceil^+|$, (40) comes from the bound in (14), (41) comes from the equiva-

lence of the 2- and ∞-norms, (42) comes by using the definition of i, (43) comes from (63) and (64) in Lemma 8 in Appendix J, and that $||\nabla f|| \leq B$. By taking $\beta = \gamma/(\alpha B\sqrt{N})$ in (44), we get

$$\frac{\epsilon\gamma}{\alpha BN} \leq |\mathbf{x}_i - \lceil \bar{\mathbf{x}}_i \rceil^+| \tag{44}$$

where $\bar{\mathbf{x}} = \mathbf{x} - (\gamma/\sqrt{N})\,\mathtt{sign}\,(\nabla f(\mathbf{x}))$. Moreover, from (42) and the nonexpansiveness of projections (see [26, Proposition B.11-c)]), we have

$$\frac{\epsilon}{\alpha\sqrt{N}} \leq |\nabla_i f(\mathbf{x})|. \tag{45}$$

Therefore, we have

$$\begin{aligned}\langle \nabla f(\mathbf{x}), \mathbf{x} - \lceil \bar{\mathbf{x}} \rceil^+ \rangle &= \sum_{j=1}^{N} \nabla_j f(\mathbf{x})(\mathbf{x}_j - \lceil \bar{\mathbf{x}}_j \rceil^+) && (46)\\ &\geq \nabla_i f(\mathbf{x})(\mathbf{x}_i - \lceil \bar{\mathbf{x}}_i \rceil^+) && (47)\\ &\geq \frac{\epsilon^2\gamma}{\alpha^2 BN^{3/2}} && (48)\end{aligned}$$

where (47) comes from the fact that every component of the sum is nonnegative (see (65) in Lemma 8 in Appendix J), and (48) comes from using the bound in (44) and (45). Inequality (48) and the descent lemma [37, eq. (2.1.6)] yield

$$\begin{aligned} f(\lceil \bar{\mathbf{x}} \rceil^+) &\leq f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{x} - \lceil \bar{\mathbf{x}} \rceil^+ \rangle + \frac{L}{2}||\mathbf{x} - \lceil \bar{\mathbf{x}} \rceil^+||^2 \\ &\leq f(\mathbf{x}) - \frac{\epsilon^2\gamma}{\alpha^2 BN^{3/2}} + \frac{L}{2}\gamma^2 \end{aligned}$$

where the last term comes from the fact that $||\mathbf{x} - \lceil \bar{\mathbf{x}} \rceil^+|| \leq \gamma$ by the nonexpansiveness of the projection. ■

Appendix E
Proof of Lemma 4

Proof: Step 1: We prove by induction that (19) holds for all $t \geq T$. When $t = T$, then (19) holds because $f(\mathbf{x}(T)) \leq \bar{F}_\alpha(\epsilon)$ by definition of $\bar{F}_\alpha$ [see (20)]. Now, suppose (19) holds for $t \geq T$. We will show that (19) also holds for $t+1$. Consider first the case when $\mathbf{x}(t) \in \bar{\mathcal{X}}_\alpha(\epsilon)$. Then, from [37, eq. (2.1.6)]

$$\begin{aligned} f(\mathbf{x}(t+1)) &\leq f(\mathbf{x}(t)) - \langle \nabla f(\mathbf{x}(t)), \mathbf{x}(t) - \lceil \bar{\mathbf{x}}(t) \rceil^+ \rangle \\ &\quad + \frac{L}{2}||\mathbf{x}(t) - \lceil \bar{\mathbf{x}}(t) \rceil^+||^2 \\ &\leq \bar{F}_\alpha(\epsilon) + \frac{L}{2}\bar{\gamma}^2 \end{aligned}$$

where $\bar{\mathbf{x}}(t) = \mathbf{x}(t) - (\gamma(t)/\sqrt{N})\,\mathtt{sign}\,(\nabla f(\mathbf{x}(t)))$, the second inequality comes from the fact that (i) that $f(\mathbf{x}(t)) \leq \bar{F}_\alpha(\epsilon)$ since $\mathbf{x}(t) \in \bar{\mathcal{X}}_\alpha(\epsilon)$, (ii) that the inner product term is nonnegative because every term of the sum [see (46)] is nonnegative following (65) in Lemma 8, and (iii) that $||\mathbf{x}(t) - \lceil \bar{\mathbf{x}}(t) \rceil^+|| \leq \bar{\gamma}$ because of the nonexpansiveness of the projection $\lceil \cdot \rceil^+$ (see [26, Proposition B.11-c)]). Otherwise, if $\mathbf{x}(t) \notin \bar{\mathcal{X}}_\alpha(\epsilon)$, then $f(\mathbf{x}(t+1)) \leq f(\mathbf{x}(t))$ by Lemma 3, yielding the result.

Step 2: We will prove that there exists $\kappa > 0$ such that (i) $\bar{\mathcal{X}}_\alpha(\epsilon)$ is bounded set and (ii) $\bar{F}_\alpha(\epsilon) < \infty$ for all $\epsilon \in [0, \kappa]$. Part (i) follows directly from Lemma 9 in Appendix J. To prove part (ii), note that $\bar{\mathcal{X}}_\alpha(\epsilon)$ is closed set and also bounded for all $\epsilon \in [0, \kappa]$ for some $\kappa > 0$ from part (i). In particular, $\bar{\mathcal{X}}_\alpha(\epsilon)$ is compact set for all $\epsilon \in [0, \kappa]$ so the supremum in (20) is attained, and hence, $\bar{F}_\alpha(\epsilon) < \infty$.

Step 3: We will prove that $\lim_{\epsilon \to 0^+} \bar{F}_\alpha(\epsilon) = f^\star$. In particular, we show that $\bar{F}_\alpha$ is continuous at 0, which implies the result, since $\bar{F}_\alpha(0) = f^\star$. Take any sequence $(\epsilon_k)_{k\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\lim_{k\to\infty} \epsilon_k = 0$. Then, there exists $K \in \mathbb{N}$ and a sequence $(\mathbf{x}(k))_{k\in\mathbb{N}}$ such that $f(\mathbf{x}(k)) = \bar{F}_\alpha(\epsilon)$ holds for all $k \geq K$, since $\bar{\mathcal{X}}_\alpha(\epsilon)$ is compact for all $\epsilon \in [0, \kappa]$, where κ is chosen as in *Step 2*. Moreover, by the definition of $\bar{\mathcal{X}}_\alpha(\epsilon)$, we have that $\lim_{k\to\infty} L_\alpha(\mathbf{x}(k)) = 0$. Now, since L_α is a continuous function, we can conclude that for every limit point $\bar{\mathbf{x}}$ of $(\mathbf{x}(k))_{k\in\mathbb{N}}$, it holds that $L_\alpha(\bar{\mathbf{x}}) = 0$, i.e., $\bar{\mathbf{x}} \in \mathcal{X}^\star$ or $f(\bar{\mathbf{x}}) = f^\star$. Since $f(\bar{\mathbf{x}}) = f^\star$ holds for every limit point of $(\mathbf{x}(k))_{k\in\mathbb{N}}$ and f is continuous, we can conclude that $\lim_{k\to\infty} f(\mathbf{x}(k)) = \lim_{k\to\infty} \bar{F}_\alpha(\epsilon_k) = f^\star$. ■

Appendix F
Proof of Theorem 5

Proof: Step 1: We will prove by contradiction that for any $\alpha > 0$

$$I := \liminf_{t\to\infty} L_\alpha(\mathbf{x}(t)) = 0. \tag{49}$$

Suppose, to the contrary, that $I > 0$. Choose T such that $L_\alpha(\mathbf{x}(t)) \geq I/2$ and $\gamma(t) < \min\{1, I^2/(4L\alpha^2 BN^{3/2})\}$ for all $t \geq T$. Then, by Lemma 3, we get for all $t \geq T$

$$\begin{aligned} f(\mathbf{x}(t+1)) &\leq f(\mathbf{x}(t)) - \frac{I^2\gamma(t)}{4\alpha^2 BN^{3/2}} + \frac{L}{2}\gamma(t)^2 && (50)\\ &= f(\mathbf{x}(t)) - \frac{I^2\gamma(t)}{8\alpha^2 BN^{3/2}} \\ &\quad + \left(\gamma(t) - \frac{I^2}{4L\alpha^2 BN^{3/2}}\right)\frac{L}{2}\gamma(t) \\ &\leq f(\mathbf{x}(t)) - \frac{I^2\gamma(t)}{8\alpha^2 BN^{3/2}}. && (51)\end{aligned}$$

Since (52) holds for all $t \geq T$, we obtain

$$f(\mathbf{x}(t)) \leq f(\mathbf{x}(T)) - \frac{I^2}{8\alpha^2 BN^{3/2}} \sum_{\tau=T}^{t-1} \gamma(\tau), \text{ for } t \geq T. \tag{52}$$

Since $\gamma(t)$ is nonsummable (52) implies that $\lim_{t\to\infty} f(\mathbf{x}(t)) = -\infty$, which contradicts the fact that $\mathcal{X}^\star$ is nonempty. Therefore, we can conclude that $I = 0$ [see (50)].

Step 2: We will prove that $\lim_{t\to\infty} f(\mathbf{x}(t)) = f^\star$. Let $\epsilon > 0$ be given. Choose $\kappa > 0$ such that $\bar{F}_\alpha(\kappa) < f^\star + \epsilon/2$, where $\bar{F}_\alpha(\kappa)$ is defined in (20) of Lemma 4, such κ exists since $\lim_{\epsilon\to 0^+} F(\epsilon) = f^\star$. Now, choose T such that $\mathbf{x}(T) \in \bar{\mathcal{X}}_\alpha(\kappa)$ [see (15)] and for all $t \geq T$, it holds that $\gamma(t) \leq \bar{\gamma} := \sqrt{\epsilon/L}$, such T exists because of (49) and that $\lim_{t\to\infty} \gamma(t) = 0$. Then, from (19) in Lemma 4, we have for all $t \geq T$ that

$$f(\mathbf{x}(t)) - f^\star \leq \bar{F}_\alpha(\kappa) - f^\star + \frac{L}{2}\bar{\gamma}^2 \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Step 3: We will prove that the sequence $\mathbf{x}(t)$ is bounded. Take $\kappa > 0$ such that $\bar{F}_\alpha(\kappa) < \infty$, where $\bar{F}_\alpha$ is defined in (20), such κ exists by Lemma 4. From (49), $\mathbf{x}(t) \in \bar{\mathcal{X}}_\alpha(\kappa)$ holds for infinitely many $t \in \mathbb{N}$. Let $\mathbf{x}(t_k)$, with $k \in \mathbb{N}$, be the subsequence of all $\mathbf{x}(t) \in \bar{\mathcal{X}}_\alpha(\kappa)$. Choose $T \in \mathbb{N}$ such that $\gamma(t) \leq \epsilon^2/(L\alpha^2 BN^{3/2})$ for all $t \geq T$. Then, by following the same steps as used to obtain (51) and (52) and using the fact that $f(\mathbf{x}(t_k)) \leq \bar{F}_\alpha(\kappa)$, we have for every $k \in \mathbb{N}$ such that $t_k \geq T$ and all $t \in \mathbb{N}$ such that $t_k < t < t_{k+1}$ that

$$f(\mathbf{x}(t)) \leq \bar{F}_\alpha(\kappa) - \frac{\epsilon^2}{2\alpha^2 BN^{3/2}} \sum_{\tau=t_k}^{t} \gamma(\tau).$$

Therefore, since $f^\star \leq f(\mathbf{x}(t))$, we have that

$$\sum_{\tau=t_k}^{t_{k+1}} \gamma(t) \leq \frac{2\alpha BN^{3/2}}{\epsilon^2}(\bar{F}_\alpha(\kappa) - f^\star). \tag{53}$$

We also have from Lemma 4 that $\bar{\mathcal{X}}_\alpha(\kappa)$ is bounded, so there exists $A \in \mathbb{R}_+$ such that $||\mathbf{x}|| \leq A$ for all $\mathbf{x} \in \bar{\mathcal{X}}_\alpha(\kappa)$. As a result, for all $t \geq T$ and $k(t) = \max\{k \in \mathbb{N} | t_k \leq t\}$, we get

$$\begin{aligned} ||\mathbf{x}(t)|| &\leq \sum_{\tau=t_{k(t)}}^{t-1} ||\mathbf{x}(\tau+1) - \mathbf{x}(\tau)|| + ||\mathbf{x}(t_k)|| \\ &\leq \sum_{\tau=t_{k(t)}}^{t} \gamma(\tau) + A \end{aligned}$$

where the first inequality comes by writing $\mathbf{x}(t)$ as a telescoping series starting at $\mathbf{x}(t_k)$ together with the triangle inequality and the second inequality comes from the relation

$$||\mathbf{x}(t+1) - \mathbf{x}(t)|| = \left|\left|\lceil \bar{\mathbf{x}}(t) \rceil^+ - \mathbf{x}(t)\right|\right| \leq \gamma(t)$$

for all $t \in \mathbb{N}$, where $\bar{\mathbf{x}}(t) = \mathbf{x}(t) - (\gamma(t)/\sqrt{N})\,\mathtt{sign}\,(\nabla f(\mathbf{x}(t)))$. Thus, from (53), we can conclude that the sequence $\mathbf{x}(t)$ is bounded.

Step 4: We will prove $\lim_{t\to\infty} \mathtt{dist}\,(\mathbf{x}(t), \mathcal{X}^\star) = 0$ by contradiction. Suppose that there exists $\epsilon > 0$ and a subsequence $\mathbf{x}(t_k)$ such that $\mathtt{dist}\,(\mathbf{x}(t_k), \mathcal{X}^\star) \geq \epsilon$ for all $k \in \mathbb{N}$. Then, since $\mathbf{x}(t)$ is bounded, so we can without loss of generality restrict $\mathbf{x}(t_k)$ to a convergent subsequence to some point $\bar{\mathbf{x}}$, so $\lim_{k\to\infty} \mathbf{x}(t_k) = \bar{\mathbf{x}}$. Now, since f is continuous and $\lim_{t\to\infty} f(\mathbf{x}(t)) = f^\star$, we can conclude that $f(\bar{\mathbf{x}}) = f^\star$ and $\bar{\mathbf{x}} \in \mathcal{X}^\star$. Then, $\lim_{t\to\infty} \bar{\mathbf{x}}(t_k) = \mathbf{x} \in \mathcal{X}^\star$ contradicts that $\mathtt{dist}\,(\mathbf{x}(t_k), \mathcal{X}^\star) \geq \epsilon$ for all $k \in \mathbb{N}$. ■

Appendix G
Proof of Lemma 5

Proof: By using that the gradients of f are L-Lipschitz continuous, we can apply the descent lemma (see, for example, [37, eq. (2.1.6)] or [26, Proposition A.24]). The descent lemma states that for all γ, we have

$$\begin{aligned} f(\mathbf{x} - \gamma\mathbf{d}) &\leq f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{d}\rangle\gamma + \frac{L}{2}||\mathbf{d}||^2\gamma^2 && (54)\\ &= f(\mathbf{x}) + \left(\frac{L}{2}\gamma - \langle \nabla f(\mathbf{x}), \mathbf{d}(t)\rangle\right)\gamma && (55)\\ &\leq f(\mathbf{x}) + \left(\frac{L}{2}\gamma - \cos(\theta)\epsilon\right)\gamma && (56)\\ &= f(\mathbf{x}) - \delta(\epsilon, \gamma, \theta) && (57) \end{aligned}$$

where (55) comes from that $||\mathbf{d}|| = 1$, (56) comes from that $\mathtt{ang}\,(\nabla f(\mathbf{x}), \mathbf{d}) \leq \theta$, $||\nabla f(\mathbf{p})|| \geq \epsilon$, since $\mathbf{x} \notin \mathcal{X}(\epsilon)$, and $\langle \nabla f(\mathbf{x}), \mathbf{d}\rangle = ||\nabla f(\mathbf{x})|| \cos(\,\mathtt{ang}\,(\mathbf{d}, \nabla f(\mathbf{x})))$. ■

Appendix H
Proof of Lemma 6

Proof: a) The result can be proved using *Steps 1–3* used to prove Lemma 4, using $||\nabla f(\cdot)||$, $\mathcal{X}(\epsilon)$, $F(\epsilon)$, and Lemma 5 in place of $L_\alpha(\cdot)$, $\bar{\mathcal{X}}_\alpha(\alpha)$, $\bar{F}_\alpha(\epsilon)$ and Lemma 3, respectively.

b) For any $\mathbf{x} \in \mathcal{X}(\epsilon)$ [37, eq. (2.1.19) in Th. 2.1.10]

$$f(\mathbf{x}) \leq f^\star + \frac{1}{2\mu}||\nabla f(\mathbf{x})||^2 \leq f^\star + \frac{\epsilon^2}{2\mu}$$

where we have used that $\nabla f(\mathbf{x}^\star) = \mathbf{0}$ for all $\mathbf{x}^\star \in \mathcal{X}^\star$. ■

Appendix I
Proof of Theorem 8

Proof: The results can be proved using *Steps 1–4* used to prove Theorem 5. The main difference is that here *Step 1* is to prove that $\liminf_{t\to\infty} ||\nabla f(\mathbf{x}(t))|| = 0$, instead of $\liminf_{t\to\infty} L_\alpha(\mathbf{x}(t)) = 0$ as in the proof of Theorem 5. Moreover, here, we use $||\nabla f(\cdot)||$, $\mathcal{X}(\epsilon)$, $F(\epsilon)$, Lemma 5, and Lemma 6(a) in place of $L_\alpha(\cdot)$, $\bar{\mathcal{X}}_\alpha(\alpha)$, $\bar{F}_\alpha(\epsilon)$, Lemma 3, and Lemma 4, respectively. ■

Appendix J
Additional Lemmas

Lemma 7: Consider $\mathcal{D}$ defined in Example 2 of Section III-B. $\mathcal{D}$ is a θ-cover with the θ in (9).

Proof: We show that for θ defined in (9), it holds for any $\mathbf{x} \in \mathcal{S}^{N-1}$ that there exists $\mathbf{d} \in \mathcal{D}_1$ such that (3) holds.

First, consider the case where $\mathbf{x}_j \geq \cos(\theta)$ for some component j. Then, for $\mathbf{e}_j \in \mathcal{D}_1$, we get $\cos(\,\mathtt{ang}\,(\mathbf{x}, \mathbf{e}_j)) = \langle \mathbf{x}, \mathbf{e}_j\rangle = \mathbf{x}_j \geq \cos(\theta)$. Therefore, we finalize the proof by showing that if $\mathbf{x} \in \mathcal{S}^{N-1}$ and $\mathbf{x}_i \leq \cos(\theta)$ for $i = 1, \ldots, N$, then

$$\cos\left(\mathtt{ang}\left(\mathbf{x}, -\frac{1}{\sqrt{N}}\mathbf{1}\right)\right) = \frac{-1}{\sqrt{N}}\sum_{i=1}^{N} \mathbf{x}_i \geq \cos(\theta)$$

Without loss of generality, let the components of $\mathbf{x}$ be ordered so that $\mathbf{x}_i \geq 0$ if $i = 1, \ldots, K$ and $\mathbf{x}_i < 0$ if $i = K+1, \ldots, N$,

where K is the number of positive components of $\mathbf{x}$. Then

$$\frac{-1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{x}_i \geq -\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{K}\mathbf{x}_i - \sqrt{1-\sum_{i=1}^{K}\mathbf{x}_i^2}\right) \quad (58)$$

$$\geq -\frac{1}{\sqrt{N}}\left(K\cos(\theta) - \sqrt{1-K\cos(\theta)^2}\right) \quad (59)$$

where (58) comes by using that $\sum_{i=1}^{N}\mathbf{x}_i^2 = 1$ and the inequality between the 1 and 2 norm, i.e.,

$$\sum_{i=K+1}^{N}|\mathbf{x}_i| \geq \sqrt{\sum_{i=K+1}^{N}\mathbf{x}_i^2} = \sqrt{1-\sum_{i=1}^{K}\mathbf{x}_i^2}$$

and (59) comes by noting that (58) is decreasing and that $\mathbf{x}_i \leq \cos(\theta)$ for all i. Now, by inserting our choice of $\cos(\theta)$ from (9) in (59), we get

$$\frac{-1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{x}_i \geq -\frac{K-\sqrt{N^2+2\sqrt{N}(N-1)-K}}{\sqrt{N}\sqrt{N^2+2\sqrt{N}(N-1)}} \quad (60)$$

$$\geq -\frac{N-1-\sqrt{N^2+2\sqrt{N}(N-1)-(N-1)}}{\sqrt{N}\sqrt{N^2+2\sqrt{N}(N-1)}} \quad (61)$$

$$= \frac{\sqrt{N}}{\sqrt{N}\sqrt{N^2+2\sqrt{N}(N-1)}} = \cos(\theta) \quad (62)$$

where (61) comes from the fact that (60) is decreasing in K and $K \leq N-1$, and (62) comes by using that $N^2+2\sqrt{N}(N-1)-(N-1) = ((N-1)+\sqrt{N})^2$. ■

Lemma 8: For all $\beta \in [0,1]$, $z \in \mathbb{R}$ and $x, \alpha_1, \alpha_2 \in \mathbb{R}_+$ with $\alpha_1 \leq \alpha_2$, the following holds:

$$\beta|x - \lceil x - z\rceil^+| \leq |x - \lceil x - \beta z\rceil^+| \quad (63)$$

$$|x - \lceil x - \alpha_1 z\rceil^+| \leq |x - \lceil x - \alpha_2 z\rceil^+| \quad (64)$$

$$0 \leq z(x - \lceil x - \alpha_1 z\rceil^+). \quad (65)$$

Proof: We first prove (63). Direct inspection shows that

$$\phi_1(x,z,\beta) := \beta|x - \lceil x - z\rceil^+| = \begin{cases}\beta|z|, & \text{if } x \geq z\\ \beta x, & \text{if } x \leq z\end{cases}$$

$$\phi_2(x,z,\beta) := |x - \lceil x - \beta z\rceil^+| = \begin{cases}\beta|z|, & \text{if } x \geq \beta z\\ x, & \text{if } x \leq \beta z.\end{cases}$$

Therefore, for $z \in \mathbb{R}_+$, we have $\phi_1(x,z,\beta) = \beta|z| = \phi_2(x,z,\beta)$ if $x \in [z,\infty)$, $\phi_1(x,z,\beta) = \beta x \leq \beta|z| = \phi_2(x,z,\beta)$ if $x \in [\beta z, z]$, and $\phi_1(x,z,\beta) = \beta x \leq x = \phi_2(x,z,\beta)$ if $x \in [0,\beta z]$. So $\phi_1(x,z,\beta) \leq \phi_2(x,z,\beta)$ for all $x,z \in \mathbb{R}_+$ and $\beta \in [0,1]$, which yields (63).

Equation (64) follows directly from using the definition of $\phi_2(\cdot)$. To prove (65), we use the fact that $\texttt{sign}(z)\lceil x - \alpha_1 z\rceil^+ \leq \texttt{sign}(z)x$ or by rearranging $0 \leq \texttt{sign}(z)(x - \lceil x - \alpha_1 z\rceil^+)$. By multiplying $|z|$ on both sides, we obtain (65). ■

Lemma 9: Suppose $\mathcal{X}^\star$ is bounded. Then: (i) there exists $\kappa>0$ such that $\mathcal{X}(\epsilon)$ defined in (26) is bounded for all $\epsilon<\kappa$. (ii) If $||\nabla f(\mathbf{x})|| \leq B$ for all $\mathbf{x}\in\mathbb{R}_+^N$, then there exists $\kappa>0$ such that $\bar{\mathcal{X}}_\alpha(\epsilon)$ in (15) is bounded for all $\epsilon<\kappa$.

Proof: (i) Take any $\mathbf{x}^\star \in \mathcal{X}^\star$ and choose $R>0$ so that $\mathcal{X}^\star \subseteq \mathcal{B}^N(\mathbf{x}^\star, R)$. Take $\kappa_1 > 0$ given by

$$\kappa_1 = \frac{1}{L}\underset{\mathbf{x}\in\mathcal{S}^{N-1}(\mathbf{x}^\star,R)}{\text{minimize}} ||\nabla f(\mathbf{x})||^2. \quad (66)$$

Note that such a κ_1 exists since $\mathcal{S}^{N-1}(\mathbf{x}^\star, R)$ is compact and $\kappa_1 > 0$ since $\mathcal{S}^{N-1}(\mathbf{x}^\star, R) \cap \mathcal{X}^\star$ is empty. Moreover, using [37, (2.1.8) in Th. 2.1.5], that $\nabla f(\mathbf{x}^\star) = \mathbf{0}$, and (66), we have for all $\mathbf{x} \in \mathcal{S}^{N-1}(\mathbf{x}^\star, R)$ that

$$\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^\star\rangle \geq (1/L)||\nabla f(\mathbf{x})||^2 \geq \kappa_1. \quad (67)$$

We now show that for all $\mathbf{x} \in \mathbb{R}^N \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$, we have $||\nabla f(\mathbf{x})|| \geq \kappa$, where $\kappa = \kappa_1/R$. Take some $\mathbf{x} \in \mathbb{R}^N \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$, and let $\bar{\mathbf{x}}$ denote the unique point in the intersection of the line segment $[\mathbf{x}^\star, \mathbf{x}]$ and $\mathcal{S}^{N-1}(\mathbf{x}^\star, R)$, such an $\bar{\mathbf{x}}$ exists because $\mathbf{x} \in \mathbb{R}^N \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$ and $\mathbf{x}^\star \in \mathcal{B}^N(\mathbf{x}^\star, R)$. Consider now the function $G : [0,\infty) \to \mathbb{R}^N$ with

$$G(\tau) = \nabla f(\mathbf{x}^\star + \tau(\bar{\mathbf{x}} - \mathbf{x}^\star)). \quad (68)$$

Clearly, $G(0) = \nabla f(\mathbf{x}^\star) = \mathbf{0}$, $G(1) = \nabla f(\bar{\mathbf{x}})$, and there exists $\hat{\tau} \geq 1$ such that $G(\hat{\tau}) = \nabla f(\mathbf{x})$. By using that gradients of convex functions are monotone, i.e., for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$, it holds that $\langle \nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_2), \mathbf{x}_1 - \mathbf{x}_2\rangle \geq 0$, we conclude that for $\tau_1, \tau_2 \in \mathbb{R}_+$ with $\tau_1 \geq \tau_2$, it holds that $\langle G(\tau_1) - G(\tau_2), (\tau_1 - \tau_2)(\bar{\mathbf{x}} - \mathbf{x}^\star)\rangle \geq 0$. Rearranging this,

$$\langle G(\tau_1), (\bar{\mathbf{x}} - \mathbf{x}^\star)\rangle \geq \langle G(\tau_2), (\bar{\mathbf{x}} - \mathbf{x}^\star)\rangle, \quad \text{for } \tau_1 \geq \tau_2. \quad (69)$$

By combining (67) and (69), we get that

$$\langle G(\hat{\tau}), (\bar{\mathbf{x}} - \mathbf{x}^\star)\rangle \geq \langle G(1), (\bar{\mathbf{x}} - \mathbf{x}^\star)\rangle \geq \kappa_1. \quad (70)$$

Hence, by the Cauchy–Schwarz inequality, we have $||\nabla f(\mathbf{x})||R = ||G(\hat{\tau})||R \geq \kappa_1$ and by rearranging we get $||\nabla f(\mathbf{x})|| \geq \kappa_1/R = \kappa$. Since $||\nabla f(\mathbf{x})|| \geq \kappa$ holds for all $\mathbf{x} \in \mathbb{R}^N \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$, we can conclude that $\mathcal{X}(\epsilon)$ is bounded for $\epsilon < \kappa$.

(ii) We prove the result by contradiction. Suppose $\bar{\mathcal{X}}_\alpha(\kappa)$ is unbounded for all $\kappa > 0$. Then, there exists a sequence $\mathbf{x}^k \in \mathbb{R}_+^N$ such that $\lim_{k\to\infty}||\mathbf{x}^k|| = \infty$ and $\lim_{k\to\infty} L_\alpha(\mathbf{x}^k) = 0$. We prove the contraction in the following steps.

Step 1: We will show that there exists $\bar{\kappa} > 0$ and R such that $||\nabla f(\mathbf{x})|| \geq \bar{\kappa}$ holds for all $\mathbf{x} \in \mathbb{R}_+^N$ and $||\mathbf{x} - \mathbf{x}^\star|| \geq R$. If there exists $\mathbf{x}^\star \in \mathcal{X}^\star$ such that $||\nabla f(\mathbf{x}^\star)|| = 0$, then the result follows from part (i). Therefore, without loss of generality, suppose we can take $\mathbf{x}^\star \in \mathcal{X}^\star$ with $||\nabla f(\mathbf{x}^\star)|| > 0$. Then, the set $\mathcal{J} := \{j = 1,\ldots,N, |\nabla_j f(\mathbf{x}^\star) \neq 0\}$ is nonempty. We also have, using the KKT conditions [32, Sec. 5.9.2], that $\mathbf{x} \in \mathcal{X}^\star$ if and only if the following three conditions hold: (A) $\mathbf{x} \in \mathbb{R}_+^N$; (B) $\nabla_i f(\mathbf{x}) \geq 0$ for $i = 1,\ldots,N$; and (C) $\nabla_i f(\mathbf{x})\mathbf{x}_i = 0$ for $i = 1,\ldots,N$.

We first show that $\langle \nabla f(\mathbf{x}), \mathbf{x} - \mathbf{x}^\star\rangle > 0$ for all $\mathbf{x} \in \mathbb{R}_+^N \setminus \mathcal{X}^\star$. Consider first the case when $\mathbf{x} \in \mathbb{R}_+^N \setminus \mathcal{X}^\star$ and $\mathbf{x}_j > 0$ for

some $j \in \mathcal{J}$. Then, we have

$$\langle \nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^\star \rangle \geq \langle \nabla f(\mathbf{x}^\star), \mathbf{x}-\mathbf{x}^\star \rangle \geq \sum_{i=1}^{N} \nabla_i f(\mathbf{x}^\star)\mathbf{x}_i > 0$$

where the first inequality comes by the monotonicity of ∇f, the second inequality comes by the optimality condition (C), and the final inequality comes by the optimality condition (B), the fact that $\nabla_i f(\mathbf{x}^\star) > 0$ for all $j \in \mathcal{J}$, and that $\mathbf{x}_j > 0$ for some $j \in \mathcal{J}$. Consider next the case when $\mathbf{x} \in \mathbb{R}^N_+ \setminus \mathcal{X}^\star$ and $\mathbf{x}_j = 0$ for all $j \in \mathcal{J}$. Then, $\nabla_i f(\mathbf{x}) \neq \nabla_i f(\mathbf{x}^\star)$ for some i, because otherwise the optimality conditions (A), (B), and (C) hold for $\mathbf{x}$ so $\mathbf{x} \in \mathcal{X}^\star$. In particular, $||\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^\star)|| > 0$. Therefore, we have [37, eq. (2.1.8)]

$$\langle \nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^\star \rangle \geq \langle \nabla f(\mathbf{x}^\star), \mathbf{x}-\mathbf{x}^\star \rangle + \frac{1}{L}||\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^\star)||^2$$

which is strictly larger than 0 since $\langle \nabla f(\mathbf{x}^\star), \mathbf{x}-\mathbf{x}^\star \rangle \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N_+$ and that $||\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^\star)|| > 0$.

Now, take $R > 0$ such that $\mathcal{X}^\star \subseteq \mathcal{B}^N(\mathbf{x}^\star, R)$. Then, since $\mathcal{S}^{N-1}(\mathbf{x}^\star, R) \cap \mathbb{R}^N_+$ is compact, there exists $\kappa_1 = \min_{\mathbf{x} \in \mathcal{S}^{N-1}(\mathbf{x}^\star, R) \cap \mathbb{R}^N_+} \langle \nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^\star \rangle > 0$. We can now follow same arguments as in the proof of part (i) to show that $||\nabla f(\mathbf{x})|| \geq \bar{\kappa}$, where $\bar{\kappa} = \kappa_1 / R$.

Step 2: We will show that the following inequality holds for all $\mathbf{x} \in \mathcal{R}^N_+ \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$:

$$\cos(\,\mathtt{ang}\,(\nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^\star)) \geq \frac{\bar{\kappa}}{B} \tag{71}$$

where R and $\bar{\kappa}$ are defined as in *Step 1*. Take some $\mathbf{x} \in \mathbb{R}^N_+ \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$. Similarly as in part (i), let $\bar{\mathbf{x}}$ denote the unique point in the intersection of the line segment $[\mathbf{x}^\star, \mathbf{x}]$ and $\mathcal{S}^{N-1}(\mathbf{x}^\star, R)$. Moreover, take $\hat{\tau} > 1$ such that $\mathbf{x} = \mathbf{x}^\star + \hat{\tau}(\bar{\mathbf{x}} - \mathbf{x}^\star)$ and define $G : [0, \infty) \to \mathbb{R}^N$ as in (68). Then, by rearranging (70) and multiplying both sides with $1/R$

$$\cos(\,\mathtt{ang}\,(\nabla f(\mathbf{x}), \mathbf{x}-\mathbf{x}^\star)) \geq \frac{\kappa_1}{R}\frac{1}{||G(\hat{\tau})||} \geq \frac{\bar{\kappa}}{B}$$

where κ_1 and $\bar{\kappa}$ are defined as in part (i) and *Step 1*.

Step 3: We will show that the subsequence $\mathbf{x}^k$ can be restricted so that (a) $\lim_{k\to\infty} \nabla f(\mathbf{x}^k) = \mathbf{f}$ for some $\mathbf{f} \in \mathbb{R}^N$ and (b) for each component $i = 1, \ldots, N$ either $\lim_{k\to\infty} \mathbf{x}^k_i = 0$ or $\mathbf{x}_i \geq W$, for some $W > 0$. We first show (a). Since ∇f is bounded by B, the sequence $\nabla f(\mathbf{x}^k)$ is bounded. Therefore, we can restrict the sequence $\mathbf{x}^k$ so that $\nabla f(\mathbf{x}^k)$ is a convergent subsequence with $\lim_{k\to\infty} \nabla f(\mathbf{x}^k) = \mathbf{f}$. To show (b), for each component $i = 1, \ldots, N$, we restrict the sequence $\mathbf{x}^k$ so that $\mathbf{x}^k_i \geq W_i$ if $\mathbf{x}^k_i$ does not converge to 0 and taking $W = \max W_i$.

Step 4: We will prove that $\mathbf{f}_i = 0$ for $i \notin \mathcal{I} := \{i = 1, \ldots, N \,|\, \lim_{k\to\infty} \mathbf{x}^k_i = 0\}$. We prove the result by contradiction. Without loss of generality, suppose $\mathbf{f}_i > 0$ for some $i \notin \mathcal{I}$, the case when $\mathbf{f}_i < 0$ follows same arguments. Then, there exists $K \in \mathbb{N}$ such that $\nabla_i f(\mathbf{x}^k) \geq \eta_0 := \mathbf{f}_i/2 > 0$ for all $k \geq K$. This together with that $\mathbf{x}^k_i \geq W$ implies that $|\mathbf{x}^k_i - \lceil \mathbf{x}^k_i - \nabla_i f(\mathbf{x}^k) \rceil_+| \geq \min\{\eta_0, W\} =: \eta > 0$. Therefore, $L_\alpha(\mathbf{x}^k) \geq \eta$ for all $k \geq K$, contradicting that $\lim_{t\to\infty} L_\alpha(\mathbf{x}^k) = 0$.

Step 5: We will prove contradiction when $\mathcal{I}$ is empty. From *Step 4*, we have $\mathbf{f} = \mathbf{0}$. However, we also have from *Step 1* that $||\nabla f(\mathbf{x})|| \geq \bar{\kappa}$ for all $\mathbf{x} \in \mathbb{R}^N_+ \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$. Since $\mathcal{B}^N(\mathbf{x}^\star, R)$ is bounded and $\lim_{t\to\infty} ||\mathbf{x}(t)|| = \infty$, we have that $||\mathbf{f}|| \geq \bar{\kappa} > 0$, which contradicts that $\mathbf{f} = \mathbf{0}$.

Step 6: We will prove contradiction when $\mathcal{I}$ is nonempty. Consider the sequence $\mathbf{z}^k = (\mathbf{x}^k - \mathbf{x}^\star)/(||\mathbf{x}^k - \mathbf{x}^\star||)$. Since $\mathbf{z}^k$ is bounded, we can restrict the subsequence $\mathbf{x}^k$ so that $\mathbf{z}^k$ has a convergent subsequence, with the limit $\mathbf{z}$. We have $\mathbf{z}_i = 0$ for $i \in \mathcal{I}$ and $||\mathbf{z}|| = 1$, since $||\cdot||$ is continuous function and $\mathbf{z}^k$ convergent sequence. Therefore, as both $\nabla f(\mathbf{x}^k)$ and $\mathbf{z}^k$ are convergent sequences and the inner product $\langle \cdot, \cdot \rangle$ is continuous function, the sequence $\langle \nabla f(\mathbf{x}^k), \mathbf{z}^k \rangle$ is convergent and has the limit $\langle \mathbf{f}, \mathbf{z} \rangle = 0$. However, for all $\mathbf{x}^k \in \mathcal{R}^N_+ \setminus \mathcal{B}^N(\mathbf{x}^\star, R)$

$$\langle \nabla f(\mathbf{x}^k), \mathbf{z}^k \rangle = ||\nabla f(\mathbf{x}^k)|| \cos(\,\mathtt{ang}\,(\nabla f(\mathbf{x}^k), \mathbf{z}^k)) \geq \frac{\bar{\kappa}^2}{B}$$

where the inequality comes by (71) in *Step 2*. ■

REFERENCES

[1] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proc. IEEE*, vol. 99, no. 6, pp. 998–1027, Jun. 2011.

[2] C. Fischione, "Fast-Lipschitz optimization with wireless sensor networks applications," *IEEE Trans. Automat. Control*, vol. 56, no. 10, pp. 2319–2331, Oct. 2011.

[3] J. Partan, J. Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," *ACM SIGMOBILE Mobile Comput Commun. Rev.*, vol. 11, no. 4, pp. 23–33, 2007.

[4] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data Networks*, vol. 2. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.

[5] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[6] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. pp. 237–252, 1998.

[7] S. Low and D. Lapsley, "Optimization flow control. I. Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

[8] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[9] D. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Automat. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[10] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2011, pp. 1–8.

[11] L. Chen, N. Li, S. H. Low, and J. C. Doyle, "Two market models for demand response in power networks," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, 2010, pp. 397–402.

[12] R. Madan and S. Lall, "Distributed algorithms for maximum lifetime routing in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2185–2193, Aug. 2006.

[13] C. Zhao, U. Topcu, N. Li, and S. Low, "Design and stability of load-side primary frequency control in power systems," *IEEE Trans. Automat. Control*, vol. 59, no. 5, pp. 1177–1189, May 2014.

[14] E. Dall'Anese, S. V. Dhople, and G. B. Giannakis, "Photovoltaic inverter controllers seeking ac optimal power flow solutions," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2809–2823, Jul. 2016.

[15] J. N. Tsitsiklis and Z. Q. Luo, "Communication complexity of convex optimization," *J. Complexity*, vol. 3, pp. 231–243, 1987.

[16] M. Rabbat and R. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 798–808, Apr. 2005.

[17] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Proc. IEEE Conf. Decision Control*, Dec. 2008, pp. 4177–4184.

[18] Y. Pu, M. N. Zeilinger, and C. N. Jones, "Quantization design for unconstrained distributed optimization," in *Proc. IEEE Amer. Control Conf.*, 2015, pp. 1229–1234.
[19] P. Yi and Y. Hong, "Quantized subgradient algorithm and data-rate analysis for distributed optimization," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 4, pp. 380–392, Dec. 2014.
[20] J. Herdtner and E. Chong, "Analysis of a class of distributed asynchronous power control algorithms for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 436–446, Mar. 2000.
[21] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. Optim.*, vol. 12, no. 1, pp. 109–138, 2001.
[22] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Automat. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
[23] S. Magnússon, K. Heal, C. Enyioha, N. Li, C. Fischione, and V. Tarokh, "Convergence of limited communications gradient methods," in *Proc. IEEE Amer. Control Conf.*, 2016, pp. 1421–1426.
[24] C. Enyioha, S. Magnússon, K. Heal, N. Li, C. Fischione, and V. Tarokh, "Robustness analysis for an online decentralized descent power allocation algorithm," in *Proc. IEEE Inf. Theory Appl. Workshop*, Feb. 2016, pp. 1–8.
[25] S. Magnússon, C. Enyioha, K. Heal, N. Li, C. Fischione, and V. Tarokh, "Distributed resource allocation using one-way communication with applications to power networks," in *Proc. IEEE Annu. Conf. Inf. Sci. Syst.*, Mar. 2015, pp. 1–6.
[26] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.
[27] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. New York, NY, USA: Springer, 1985.
[28] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A distributed Newton method for network utility maximization I: Algorithm," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2162–2175, Sep. 2013.
[29] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A distributed Newton method for network utility maximization II: Convergence," *IEEE Trans. Automat. Control*, vol. 58, no. 9, pp. 2176–2188, Sep. 2013.
[30] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle, "An O(1/k) gradient method for network resource allocation problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 64–73, Mar. 2014.
[31] A. Nedic and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1757–1780, 2009.
[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
[33] E. Ghadimi, I. Shames, and M. Johansson, "Multi-step gradient methods for networked optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5417–5429, Nov. 2013.
[34] M. Zargham, A. Ribeiro, A. Ozdaglar, and A. Jadbabaie, "Accelerated dual descent for network flow optimization," *IEEE Trans. Automat. Control*, vol. 59, no. 4, pp. 905–920, Apr. 2014.
[35] A. D. Wyner, "Random packings and coverings of the unit n-sphere," *Bell Syst. Tech. J.*, vol. 46, no. 9, pp. 2111–2118, Nov. 1967.
[36] P. Solé, "The covering radius of spherical designs," *Eur. J. Combinatorics*, vol. 12, no. 5, pp. 423–431, 1991.
[37] Y. Nesterov, *Introductory Lectures on Convex Optimization*. New York, NY, USA: Springer, 2004.

Sindri Magnússon received the B.Sc. degree in mathematics from the University of Iceland, Reykjavík, Iceland, in 2011, and the master's degree in mathematics and the Ph.D. degree in electrical engineering from KTH Royal Institute of Technology, Stockholm, Sweden, in 2013 and 2017, respectively.

He is currently a Postdoctoral Researcher with the Department of Electrical Engineering, KTH Royal Institute of Technology. He was a Visiting Researcher with Harvard University, Cambridge, MA, USA, for nine months in 2015 and 2016. His research interests include distributed optimization, both theory and applications.

Chinwendu Enyioha (M'14) received the B.Sc. degree in mathematics from Gardner–Webb University, Boiling Springs, NC, USA, in 2008, and the Ph.D. degree in electrical and systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 2014.

He is currently a Postdoctoral Research Fellow with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. Prior to arriving Harvard, he was a Postdoctoral Researcher with the GRASP Lab, University of Pennsylvania. His research interests include the area of design, optimization, and control of distributed networked systems, with applications to power systems and cyber-physical networks.

Dr. Enyioha is a Fellow of the Ford Foundation, was named a William Fontaine Scholar at the University of Pennsylvania, and has received the Mathematical Association of America Patterson award.

Na Li (M'13) received the B.S. degree in mathematics and applied mathematics from Zhejiang University, Hangzhou, China, in 2007, and the Ph.D. degree in control and dynamical systems from California Institute of Technology, Pasadena, CA, USA, in 2013.

She is an Assistant Professor with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. She was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology. Her research interests include the design, analysis, optimization, and control of distributed network systems, with particular applications to power networks and systems biology/physiology.

Dr. Li was a Best Student Paper Award finalist at the 2011 IEEE Conference on Decision and Control. She received the NSF CAREER Award in 2016 and the Air Force Young Investigator Award in 2017.

Carlo Fischione (M'05) received the Laurea degree (*summa cum laude*) in electronic engineering and the Ph.D. degree in electrical and information engineering from the University of L'Aquila, L'Aquila, Italy, in April 2001 and May 2005, respectively.

He is currently a tenured Associate Professor with the School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden. He was a Visiting Professor at Massachusetts Institute of Technology, Cambridge, MA, USA, in 2015; an Associate at Harvard University, Cambridge, in 2015; and a Visiting Scholar in 2004–2005 and a Research Associate in 2007–2008 at the University of California at Berkeley, CA, USA. His research interests include optimization with applications to wireless sensor networks, networked control systems, wireless networks, security, and privacy.

Dr. Fischione received or co-received a number of awards, including the Best Paper Award from the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS in 2007. He is an Ordinary Member of the academy of history Deputazione Abruzzese di Storia Patria.

Vahid Tarokh (F'09) received the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada.

He then worked with AT&T Labs-Research and AT&T wireless services until August 2000, where he was the Head of the Department of Wireless Communications and Signal Processing. In September 2000, he joined the Department of Electrical Engineering and Computer Sciences, at Massachusetts Institute of Technology, as an Associate Professor. In June 2002, he joined Harvard University, Cambridge, MA, USA, where he is a Professor of applied mathematics.

Dr. Tarokh has received a Guggenheim Fellowship in Applied Mathematics and holds four honorary degrees.