Communication Complexity of Distributed Resource Allocation Optimization

Sindri Magnússon, Chinwendu Enyioha, Na Li, Carlo Fischione, and Vahid Tarokh

***Abstract*—This paper investigates the communication complexity of distributed resource allocation problems. Communication complexity, which has emerged from computer science, measures the minimal number of communicated bits needed to solve some classes of problems regardless of the used algorithm, when the communication is lossless. We introduce two complexity measures for finding primal/dual saddle point solutions of resource allocation problems. The first measure is ϵ-complexity, which quantifies the minimal number of bits needed to find an ϵ-accurate solution. The second measure is b-complexity, which quantifies the best possible solution accuracy that can be achieved from communicating b bits. We find the exact ϵ- and b-complexity of a class of resource allocation problems where a single supplier allocates resources to multiple users. For both the primal and dual problems, the ϵ-complexity grows proportionally to $\log_2(1/\epsilon)$ and the b-complexity proportionally to $1/2^b$. We also introduce a variant of the ϵ- and b-complexity measures where only algorithms that ensure primal feasibility of the iterates are allowed. Such algorithms are often desirable because overuse of the resources can overload the respective systems, e.g., by causing blackouts in power systems. We provide upper and lower bounds on the convergence rate of these primal feasible complexity measures. In particular, we show that the b-complexity cannot converge at a faster rate than $\mathcal{O}(1/b)$. Therefore, the results demonstrate a trade-off between fast convergence and primal feasibility. We illustrate the result by numerical studies.**

I. INTRODUCTION

The allocation of shared resources is a fundamental task in most networks. For example, power networks are responsible for allocating electric power, and communication networks for allocating data. While the algorithms used to allocate resources in traditional networks are well established [1]–[4], the emergence of the Internet of Things and Cyber-physical systems means that our networks are becoming more diverse, with varying levels of communication capabilities. In particular, many of these networks will have limited data rates. For example, future communication networks will offer extremely low latencies that can only be achieved at the cost of low data-rates [5]. Likewise, data-rate is limited in networks with high degree of collision and interference, such as in dense wireless networks and in networks relying on communication through power lines [6], [7]. Furthermore, underwater networks have data-rate limits as they communicate via acoustic signals [8]. Motivated by these challenges, we study how resources can be allocated in networks with limited data rates.

This work was supported by the VR Chromos Project and NSF grant No. 1548204 and NSF CAREER 1553407.

S. Magnússon and C. Fischione are with Electrical Engineering School, Access Linnaeus Center, KTH Royal Institute of Technology, 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 USA. (email: cenyioha@seas.harvard.edu; nali@seas.harvard.edu)

In coordinating the solutions to resource allocation problems in networks, dual decomposition methods have been among the most prominent approaches [1], [2]. They play an important role in, e.g., communication networks [3], [4] and power networks [9]–[13]. These methods are based on coordinating dual variables, often interpreted as prices, to synchronize the supply and demand in networks, with the purpose of maximizing some global objective function. However, bandwidth limited dual decomposition methods have not received much attention in the literature, despite the emerging need to efficiently use communication bandwidth in many networks.

The theory of communication complexity has been conceived to answer questions about how many bits must be communicated to solve problems that require coordination between entities [14], [15]. Informally, communication complexity measures the minimal number of bits needed to solve the most "communication intense" problem in some class of problems, regardless of the used algorithm, when the communication is lossless. For example, in [14] Yao studies the minimal number of bits that two processors must communicate to compute the output of a class of functions that take inputs from both processors. More recent work has considered the communication complexity of finding Nash equilibrium in games [16], [17], of statistical inference in networks with sparsely located data [18], [19], and of minimizing the sum of two convex functions [20].

Besides the work in [20], communication complexity of distributed optimization problems has not received much attention in the literature. Nevertheless, some interesting papers have studied various types of distributed optimization algorithms in bandwidth limited networks [21]–[24]. However, these papers do not study algorithm invariant quantities such as communication complexity. Moreover, the work in [20]–[24] considers optimization problems where there are no shared resources or coupling constraints, as opposed to what we do in this paper. Nevertheless, many important optimization problems revolve around allocating shared resources and have coupling constraints [1]–[4], [9]–[13]. Those problem are naturally decomposed using duality theory. This gives new analytical challenges compared to the earlier work [20]–[24].

Bandwidth limited dual decomposition algorithms are considered in [25] and [26], two papers most closely related to our work. The work in [25] studies the convergence of general gradient methods, including many dual decomposition methods, when the gradients are quantized to limited number of bits. However, [25] provides no lower bound complexity results for these algorithms. The work in [26] studies quantized dual decomposition algorithms for a class of resource allocation problems that have strongly convex dual problems.

The authors of [26] also provide a lower bound on the convergence of the algorithm based on differential entropy. However, this lower bound is only valid when the initialization of the algorithm is randomly generated. In contrast to [26], we consider resource allocation problems whose dual problems are not necessarily strongly convex. We also provide different type of lower bounds than [26] based on communication complexity, which are valid even if the initialization is not randomly generated. Moreover, unlike [25] and [26], we study how the primal feasibility of bandwidth limited dual decomposition methods can be ensured, which is essential in on-line implementations of these algorithms.

A. Contributions of This Work

The main contribution of this paper is to introduce and explore the communication complexity of distributed resource allocation problems. We consider a class $\mathcal{R}$ of optimzation problems where a single supplier allocates some resource, such as power or data, to multiple users. To solve the problems in $\mathcal{R}$ the users and the supplier can coordinate by using any iterative algorithm that is implementable using the following one-way communication model: at each iteration i) the supplier broadcasts a coordination message containing finite number of bits over a lossless communication channel, ii) the users consume some amount of the resources that depends on the coordination message and the supplier measures the total consumption. This communication model naturally appears in many networks, e.g., in power networks [11] and in communication networks [2].

Using the communication model above, we introduce two measures of communication complexity of the problem class $\mathcal{R}$, ϵ-complexity and b-complexity. Informally, ϵ-complexity is the minimal number of bits needed to achieve an ϵ-accurate solution to the most "difficult" problem in $\mathcal{R}$. Similarly, the b-complexity is the best solution accuracy that can be achieved to the most "difficult" problem in $\mathcal{R}$ when b bits are communicated. We provide an exact ϵ-complexity and b-complexity as a function of ϵ and b, respectively. The ϵ-complexity grows proportionally to $\log_2(1/\epsilon)$ as ϵ decreases and the b-complexity decreases proportionally to $1/2^b$ as b grows. Since the users do not communicate to the supplier, but instead take the measurable action of consuming resources, it is desirable to ensure primal feasibility at every algorithm iteration. Otherwise, the users might consume more resources than are available and overload the system, which can cause blackouts in power networks, or outages in wireless networks. This has motivated us to consider also the class of primal feasible algorithms that are feasible at every iteration for all problems in $\mathcal{R}$. We characterize that class of primal feasible algorithms and explore the ϵ-complexity and b-complexity measures when only such algorithms are allowed. In this case, we show that the b-complexity cannot converge at a faster rate than $\mathcal{O}(1/b)$. Therefore, the results demonstrate a trade-off between fast convergence and primal feasibility.

Some preliminary studies of this work appeared in [13], [27]. However, all the theoretical and numerical results in this paper are appearing here for the first time.

B. Notation and Definitions

The set of real, positive real, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$. We set $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$. The projections of $z \in \mathbb{R}$ to $\mathbb{R}_+$ and $[m, M]$ are denoted by $\lceil z \rceil^+$ and $[z]_m^M$. $||\cdot||$ is the 2-norm and $\operatorname{dist}(\mathbf{x}, \mathcal{X}) = \inf_{\mathbf{z} \in \mathcal{Z}} ||\mathbf{x} - \mathbf{z}||$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is μ-strongly-concave, or just μ-concave, on $\mathcal{X} \subseteq \mathbb{R}^n$ if

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle - \frac{\mu}{2} ||\mathbf{y} - \mathbf{x}||^2,$$

holds for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous on $\mathcal{X} \subseteq \mathbb{R}^n$ if

$$||f(\mathbf{y}) - f(\mathbf{x})|| \leq L||\mathbf{y} - \mathbf{x}||, \text{ for all } \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

II. Problem Formulation and Related Background

We now introduce some important background information needed to obtain the main results of this paper. First in Section II-A, we introduce the class of resource allocation problems that we consider. Then in Section II-B, we introduce the considered communications schemes. Finally in Sections II-C, we introduce the concept of communication complexity for resource allocation problems.

A. Resource Allocation and Dual Decomposition

Consider a network with N users $\mathcal{N} = \{1, \dots, N\}$ and a single supplier of some resource, e.g., electricity. The resource allocated to user $i \in \mathcal{N}$ is denoted by $x_i \in \mathbb{R}_+$ and the total supply capacity is $C \in \mathbb{R}_+$. The *Resource Allocation* problem that models the resource distribution is given by [1], [2]

$$\begin{aligned} \underset{x_1,\dots,x_N}{\text{maximize}} \quad & U(\mathbf{x}) := \sum_{i=1}^{N} U_i(x_i), \\ \text{subject to} \quad & \sum_{i=1}^{N} x_i \leq C, \\ & x_i \in [m_i, M_i], \end{aligned} \tag{1}$$

where $U_i : \mathbb{R} \to \mathbb{R}$ is a utility function, $\mathbf{x} = (x_1, \dots, x_N)$, and m_i and M_i are lower and upper bounds on the demand of user $i \in \mathcal{N}$. We consider the following class of problems.

Definition 1. *Let $\mathcal{R}_{\mu,N,P}$ denote the class of all Problems* (1) *where the following assumptions hold: $U_i(x_i)$ is μ-concave for all $i \in \mathcal{N}$,*

$$\sum_{i=1}^{N} m_i \leq C \leq \sum_{i=1}^{N} M_i, \tag{2}$$

and $U'(m_i) \leq P$ for all $i \in \mathcal{N}$. We write $R = (\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ to indicate that Problem (1) *with the parameters $\mathbf{U} = (U_1, \dots, U_N)$, C, $\mathbf{m} = (m_1, \dots, m_N)$, and $\mathbf{M} = (M_1, \dots, M_N)$, is in $\mathcal{R}_{\mu,N,P}$. We denote by $\mathbf{x}_R^\star$ the optimal solution to the problem $R \in R_{\mu,N,P}$.*

Equation (2) simply states that Problem (1) is feasible and that the constraint $\sum_{i=1}^{N} x_i \leq C$ is not redundant.

Resource allocation problems are usually solved by distributed methods based on duality theory [1], [2]. The dual problems of primal Problem (1) are on the form

$$\begin{aligned} \underset{p}{\text{minimize}} \quad & D(p), \\ \text{subject to} \quad & p \geq 0, \end{aligned} \tag{3}$$

where D and p are the dual function and dual variables, see [28, Chapter 5]. Formally, D is given by

$$D(p) = \underset{\mathbf{x} \in \prod_{i=1}^{N}[m_i, M_i]}{\text{maximize}} \sum_{i=1}^{N} U_i(x_i) - p\left(\sum_{i=1}^{N} x_i - C\right). \tag{4}$$

Lemma 5 in Appendix A shows that all dual problems associated to primal problems in the class $\mathcal{R}_{\mu,N,P}$ are in the following problem class.

Definition 2. *Let $\mathcal{D}_{L,P}$ denote the set of all (dual) optimization problems of the form*

$$\begin{aligned} \underset{p}{\textit{minimize}} \quad & D(p) \\ \textit{subject to} \quad & p \in [0, P], \end{aligned} \tag{5}$$

where i) $D : \mathbb{R}_+ \to \mathbb{R}$ is convex and has L-Lipschitz continuous gradient $D'(\cdot)$, and ii) at least one optimal solution $p^\star$ to Problem (5) *is also an optimal solution to Problem* (3). *We write $D \in \mathcal{D}_{L,P}$ to indicate that Problem* (5) *with the objective function D is in $\mathcal{D}_{L,P}$.*

Due to Lemma 5 in Appendix A, the dual problem of every primal problem $R \in \mathcal{R}_{\mu,N,P}$ is in the class $\mathcal{D}_{L,P}$ where $L = N/\mu$.[1] Then the solution to the primal and dual problems can be coordinated by using dual decomposition:

$$\begin{aligned} x_i(t) =& x_i(p(t)) := \underset{x_i \in [m_i, M_i]}{\text{argmax}} \ U_i(x_i) - p(t) \ x_i \\ &= \left[(U_i')^{-1}(p(t))\right]_{m_i}^{M_i}, \end{aligned} \tag{6a}$$

$$D'(p(t)) = C - \sum_{i=1}^{N} x_i(t), \tag{6b}$$

$$p(t+1) = \left[p(t) - \gamma(t) D'(p(t))\right]_0^P, \tag{6c}$$

where $\gamma(t) \in \mathbb{R}_+$ is a step-size. This algorithm converges to optimal primal and dual solutions under mild conditions [2, Theorem 1].

B. Communication Model and Algorithms

In many real-world networks, the supplier of a resource can measure the dual gradient $D'(p)$ since it is the difference between supply and demand. However, the supplier has to communicate some coordination signals to the users, often over bandwidth limited channels. Formally, this exchange of information is captured by coordination algorithms that, at each iteration t, follow these two operations:

- **Feedback Information**: After the users $i \in \mathcal{N}$ take their actions $x_i(t)$, then the supplier can measure deviation between total consumption and supply, i.e., the dual gradient $D'(p(t)) = C - \sum_{i=1}^{N} x_i(t)$.
- **One-Way Communication**: The suppliers transmit $r(t) > 0$ bits of information at iteration t to the users over a lossless communication channel.

[1] Since at least on optimal solution to the dual Problem (3) is in $[0, P]$ we can restrict it to the interval $[0, P]$ by considering Problem (5). Problem (5) has bounded feasibility set, a property used to derive some of the results.

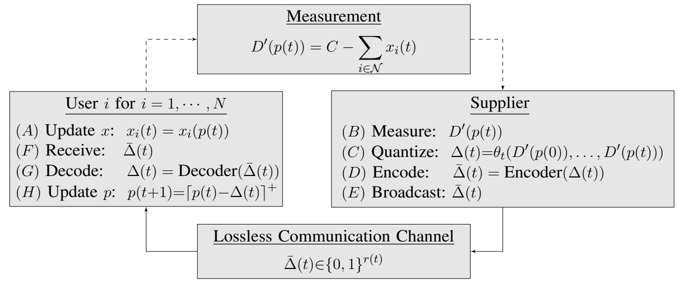


Fig. 1: The OneWay-DD algorithm.

Using this exchange of information, we want to find the algorithm that can solve the resource allocation problems by using the fewest number of bits in the one-way communication. In particular, since it is the dual gradients $D'(p(t))$ that are available at the supplier and are needed so that the users can do their local action [see Eq. (6a)], we need to find the best way to quantize the information in $D'(p)$. In other words, at every iteration t, we need to find the best way to quantize the information in $D'(p(0)), \ldots, D'(p(t))$. For that purpose, we now define the class of all such quantization schemes.

Definition 3. *We call a pair $(p(0), \theta_t(\cdot))$ a quantizion scheme if $p(0) \in [0, P]$ and $\theta_t : \mathbb{R}^{t+1} \to \mathcal{V}_t$, where $\mathcal{V}_t \subseteq \mathbb{R}$. For all t, we have $|\mathcal{V}_t| = 2^{r(t)} \in \mathbb{N}$ where $r(t) > 0$ is the communication rate at iteration t. We denote the set of all quantization schemes by $\mathcal{Q}_P$.*

Remark 1. *We assume that $r(t) > 0$ and $|\mathcal{V}_t| = 2^{r(t)} \in \mathbb{N}$, i.e., so the users receive at least one bit of information from the supplier at each iteration. This is a natural assumption, since if the users receive no new information at some iteration then they have no reason to change their update from the previous iteration.*

The value $p(0)$ is an initialization of the dual variable and should be from the interval $[0, P]$, see Definition 2 and the discussion that follows it. The function $\theta_t : \mathbb{R}^{t+1} \to \mathcal{V}_t$ is a $r(t)$-bit quantization of the information that is available at the supplier until iteration t, i.e., the values of the dual gradients $D'(p(0)), \ldots, D'(p(t))$. We wish to find the quantization scheme $(p(0), \theta_t(\cdot)) \in \mathcal{Q}_P$ that produces the "best" algorithm:

$$x_i(t) = \left[(U_i')^{-1}(p(t))\right]_{m_i}^{M_i}, \tag{7a}$$

$$\Delta(t) = \theta_t(D'(p(0)), \ldots, D'(p(t))), \tag{7b}$$

$$p(t+1) = \left[p(t) - \Delta(t)\right]_0^P, \tag{7c}$$

for $t \in \mathbb{N}_0$, to solve the resource allocation problems from the class $\mathcal{R}_{\mu,N,P}$. In Equation (7c), we can project the dual iterates to the set $[0, P]$ because the set $[0, P]$ contains an optimal dual variable, see Lemma 5 in Appendix A. We will sometimes use the notation $\mathbf{x}_R^q(t)$ and $p_D^q(t)$ to denote the iterates in Equation (7) when the quantization scheme $q \in \mathcal{Q}_P$ is used to solve the problems $R \in \mathcal{R}_{\mu,N,P}$ and $D \in \mathcal{D}_{\mu,L}$, respectively. For each quantization scheme $(p(0), \theta_t(\cdot))$, the algorithm can be implemented using the steps in following table and in Figure 1.

ONEWAY-DD: One-way communication dual decomposition
Step 1 Each user $i \in \mathcal{N}$: (A) updates its consumption $x_i(t)$ [see Eq. (6a)]. **Step** 2 The Supplier: (B) measures $D'(p(t))$ [see Eq. (6b)], (C) quantizes the information in $[D'(p(i))]_{i=0,\dots,t}$ into $\Delta(t)$ [see Eq. (7b)], (D) encodes $\Delta(t)$: $\bar{\Delta}(t)$=Encoder$(\Delta(t))$, (E) broadcasts $\bar{\Delta}(t)$ using $r(t)$ bit message through the communication channel. **Step** 3 Each user $i \in \mathcal{N}$: (F) receives the encoded message $\bar{\Delta}(t)$, (G) decodes $\bar{\Delta}(t)$: $\Delta(t) =$ Decoder$(\bar{\Delta}(t))$, (H) updates the dual variable $p(t+1)$ [see Eq. (7)].

To find the "best" quantization scheme from $\mathcal{Q}_P$ we now introduce the concept of communication complexity in our resource allocation setting.

C. Communication Complexity

In this paper, we study the communication complexity of distributed resource allocation optimization problems by seeking answers to the following questions.

i) What are the fewest number of communicated bits needed to obtain an optimal allocation with a given accuracy?
ii) What is the best solution accuracy that can be achieved given a fixed number of communicated bits?

To answer Question i) we introduce the notation of ϵ-communication-complexity, or in short ϵ-complexity. Before giving a formal definition, we introduce some notation. We say that the primal variable $\mathbf{x}$ is an ϵ-optimizer of the primal problem $R = (\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ if $||\mathbf{x} - \mathbf{x}_R^\star|| \leq \epsilon$ where $\mathbf{x}_R^\star$ is the optimal solution of the primal problem (1). Similarly, we say that the dual variable p is an ϵ-optimizer to the dual problem $D \in \mathcal{D}_{L,P}$ if $\text{dist}(p, \mathcal{P}_D^\star) \leq \epsilon$ where $\mathcal{P}_D^\star$ is the set of optimal dual solutions.[2] We need to quantify how many iterations it takes to obtain an ϵ-optimizer for each $q = (p(0), \theta_t(\cdot)) \in \mathcal{Q}_P$. For that purpose we consider the following quantities

$$T^{\text{Prim}}(q, R, \epsilon) = \min\left\{t \in \mathbb{N}_0 \middle| ||\mathbf{x}_R^q(t) - \mathbf{x}_P^\star|| \leq \epsilon\right\},$$
$$T^{\text{Dual}}(q, D, \epsilon) = \min\left\{t \in \mathbb{N}_0 \middle| \text{dist}(p_D^q(t), \mathcal{P}_D^\star) \leq \epsilon\right\},$$

for $R \in \mathcal{R}_{\mu,N,P}$, $D \in \mathcal{D}_{L,P}$, $q = (p(0), \theta_t(\cdot)) \in \mathcal{Q}_P$, and $\epsilon \in \mathbb{R}_+$. We then define the ϵ-complexity as follows.

[2]Since the utility function in the primal problem is strongly concave, the optimal solution is unique. However, there can be more than one optimal dual solution.

Definition 4 (ϵ-Complexity)**.** *For $\epsilon \in \mathbb{R}_+$, the ϵ-complexity of the resource allocation problems $\mathcal{R}_{\mu,N,P}$ and the dual problems $\mathcal{D}_{L,P}$ using the quantization schemes $\bar{\mathcal{Q}} \subseteq \mathcal{Q}_P$ is*

$$\mathcal{E}_{\mu,N,P}^{\text{Prim}}(\bar{\mathcal{Q}}, \epsilon) := \min_{q \in \bar{\mathcal{Q}}} \max_{R \in \mathcal{R}_{\mu,N,P}} \sum_{t=0}^{T^{\text{Prim}}(q,R,\epsilon)} r_q(t),$$
$$\mathcal{E}_{L,P}^{\text{Dual}}(\bar{\mathcal{Q}}, \epsilon) := \min_{q \in \bar{\mathcal{Q}}} \max_{D \in \mathcal{D}_{L,P}} \sum_{t=0}^{T^{\text{Dual}}(q,D,\epsilon)} r_q(t),$$

where $r_q(t)$ is the communication rate at iteration t of the quantization scheme q. We call the quantization schemes that solve the above problems optimal quantization schemes.

The ϵ-complexities $\mathcal{E}_{\mu,N,P}^{\text{Prim}}(\bar{\mathcal{Q}}, \epsilon)$ and $\mathcal{E}_{L,P}^{\text{Dual}}(\bar{\mathcal{Q}}, \epsilon)$ measure the fewest number of communicated bits needed to find an ϵ-optimizer of the most "difficult" problems from the classes $\mathcal{R}_{\mu,N,P}$ and $\mathcal{D}_{L,P}$, respectively, using the quantization scheme $\bar{\mathcal{Q}} \subseteq \mathcal{Q}_P$.

To answer Question-ii), we introduce the notation of b-communication-complexity, or just b-complexity. We need the following quantity

$$T(q, b) := \min\left\{T \in \mathbb{N}_0 \middle| \sum_{t=0}^{T-1} r_q(t) \leq b < \sum_{t=0}^{T} r_q(t)\right\},$$

where we define $\sum_{t=0}^{T-1} r_q(0) = 0$. The quantity $T(q, b)$ maps the quantization scheme $q \in \mathcal{Q}_P$ and the number bits $b \in \mathbb{N}_0$ to number of iterations that are attained when b bits are transfered using the quantization scheme q.

Definition 5 (b-Complexity)**.** *For $b \in \mathbb{N}$, the b-complexity of the resource allocation problems $\mathcal{R}_{\mu,N,P}$ and the dual problems $\mathcal{D}_{L,P}$ using the quantization schemes $\bar{\mathcal{Q}} \subseteq \mathcal{Q}_P$ is*

$$\mathcal{B}_{\mu,N,P}^{\text{Prim}}(\bar{\mathcal{Q}}, b) := \min_{q \in \bar{\mathcal{Q}}} \max_{R \in \mathcal{R}_{\mu,N,P}} ||\mathbf{x}_R^q(T(q,b)) - \mathbf{x}_R^\star||,$$
$$\mathcal{B}_{L,P}^{\text{Dual}}(\bar{\mathcal{Q}}, b) := \min_{q \in \bar{\mathcal{Q}}} \max_{D \in \mathcal{D}_{L,P}} \text{dist}(p_D^q(T(q,b)), \mathcal{P}_D^\star),$$

We call the quantization schemes that minimize the above upper bounds optimal quantizers.

The b-complexity measures the best possible solution accuracy that can be achieved from communicating b bits in total.

III. Communication Complexity of Resource Allocation Optimization Problems

We now find the general communication complexity of the resource allocation problems in Section II. That is the communication complexity when there are no restrictions on the quantization scheme, i.e., all quantization schemes in $\mathcal{Q}_P$ are considered. We start by providing the exact complexity in in Section III-A. Then in Sections III-B and III-C, we illustrate how the communication complexity is obtained by identifying tight lower and upper bounds, respectively.

A. General Communication Complexity

The general communication complexity is given in the following theorem.

Theorem 1. *Set $L = N/\mu$, then*

$$\mathcal{E}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}_P,\epsilon) = B_1(\epsilon) := \log_2\left\lceil \frac{P\sqrt{N}}{2\mu\epsilon}\right\rceil \text{ bits}, \quad (8a)$$

$$\mathcal{E}^{\text{Dual}}_{L,P}(\mathcal{Q}_P,\epsilon) = B_1(\epsilon) := \log_2\left\lceil \frac{P}{2\epsilon}\right\rceil \text{ bits}, \quad (8b)$$

$$\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}_P,b) = \frac{P\sqrt{N}}{\mu 2^{b+1}}, \quad (8c)$$

$$\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}_P,b) = \frac{P}{2^{b+1}}. \quad (8d)$$

Proof: The proof follows from Lemmas 1 and 2. ■

The results of the theorem are obtained by providing lower bounds and equal upper bounds on the communication complexities. The lower bound is obtained by finding, for each quantization scheme $q \in \mathcal{Q}_P$, the most difficult primal and dual problems from $\mathcal{R}_{\mu,N,P}$ and $\mathcal{D}_{L,P}$ for the quantization q. On the other hand, the upper bound is obtained by finding a particular good quantization that achieves the lower bound. We demonstrate this process in the following two sections.

B. General Lower Bounds

We now provide lower bounds on the general communication complexity. The lower bound is obtained by constructing particular primal and dual problems that are difficult for each quantization. To construct these difficult problems, we build on the following example.

Example 1. *Consider the following primal Problem* (1) *from the class $\mathcal{R}_{\mu,N,P}$ whose associated dual problem is in the class $\mathcal{D}_{L,P}$. Set $m_i = 0$ and $M_i = M := PL/N$ for all $i \in \mathcal{N}$. Set*

$$U_i(x_i) = -\frac{N}{2L}(x_i - M)^2, \text{ for all } i \in \mathcal{N}.$$

Then in view of Equation (6a) *we have*

$$x_i(p) = [U_i']^{-1}(p) = M - (L/N)p, \quad \text{for all } p \in [0,P],$$

and the dual function and its gradient are

$$D(p) = \frac{L}{2}p^2 + (C - MN)p, \quad \text{for all } p \in [0,P] \quad (9)$$

$$D'(p) = Lp + (C - MN), \quad \text{for all } p \in [0,P]. \quad (10)$$

The optimal solution to the dual problem is $p^\star = (MN - C)/L$. Therefore, given $C \in [0, NM]$ the optimal primal/dual solutions are $x_i^\star(C) = C/N$ and $p^\star(C) = (MN - C)/L$.

By using Example 1 we can prove the following lemma.

Lemma 1 (Lower Bound)**.** *Suppose $L = N/\mu$. Then for every quantization scheme $q \in \mathcal{Q}_P$ following holds:*

*a) ϵ-**compelxity:** For every $\epsilon > 0$, there exists a primal problem $R \in \mathcal{R}_{\mu,N,P}$ and an associated dual problem $D \in \mathcal{D}_{L,P}$ such that*

$$||\mathbf{x}^q_R(T(q,b)) - \mathbf{x}^\star_R|| > \epsilon \quad \text{for all } b < B_1(\epsilon) \quad (11a)$$

$$\text{dist}(p^q_D(T(q,b)), \mathcal{P}^\star_D) > \epsilon \quad \text{for all } b < B_2(\epsilon), \quad (11b)$$

where $B_1(\epsilon)$ and $B_2(\epsilon)$ are defined in Equations (8a) *and* (8b)*, respectively.*

*b) b-**compelxity:** For every $b \in \mathbb{N}_0$ there exists a primal problem $R \in \mathcal{R}_{\mu,N,P}$ and an associated dual problem $D \in \mathcal{D}_{L,P}$ such that*

$$\frac{P\sqrt{N}}{\mu 2^{b+1}} < ||\mathbf{x}^q_R(T(q,\beta)) - \mathbf{x}^\star_R|| \quad \text{for all } \beta < b, \quad (12a)$$

$$\frac{P}{2^{b+1}} < \text{dist}(p^q_D(T(q,\beta)), \mathcal{P}^\star_D) \quad \text{for all } \beta < b. \quad (12b)$$

To proof the lemma, we find for each quantization scheme $q \in \mathcal{Q}_P$ an instance of Example 1 so that the above lower bounds hold for q. This process is now formalized.

Proof: a) Given a quantization scheme $q \in \mathcal{Q}_P$. Take some $T \in \mathbb{N}_0$ and denote by $\mathcal{A}_T \subseteq [0,P]$ the set of all possible outcomes of $p(T) \in [0,P]$, that is

$$\mathcal{A}_T = \left\{ p^q_D(t) \in [0,P] \middle| D \in \mathcal{D}_{L,P} \text{ and } t = T \right\}.$$

We have the following bound on the cardinality of $\mathcal{A}_T$

$$|\mathcal{A}_T| \leq 2^{\sum_{t=0}^{T-1} r_q(t)},$$

since $\sum_{t=0}^{T-1} r_q(t)$ binary signals have been transfered at iteration T.

We first prove Equation (11a), by showing that if

$$|\mathcal{A}_T| < 2^{B_1(\epsilon)} = \left\lceil \frac{P\sqrt{N}}{2\mu\epsilon} \right\rceil, \quad (13)$$

then there exists $R \in \mathcal{R}_{\mu,N,P}$ such that $||\mathbf{x}_R(p) - \mathbf{x}^\star_R|| > \epsilon$ for all $p \in \mathcal{A}_T$. From Lemma 7 in Appendix A and Equation (13), there exists $p^\star \in [0,P]$ such that $|p - p^\star| > \mu\epsilon/\sqrt{N}$ for all $p \in \mathcal{A}_T$. Let $R \in \mathcal{R}_{\mu,N,P}$ be the primal problem given in Example 1 with $C = MN - p^\star L$. We have $x_i(p) = M - (L/N)p$ for all $p \in [0,P]$. Therefore, for all $p \in \mathcal{A}_T$ we have

$$||\mathbf{x}_R(p) - \mathbf{x}^\star_R|| = \sqrt{\sum_{i=1}^{N}(x_i(p) - x_i^\star)^2} = \sqrt{\frac{L^2}{N}(p - p^\star)^2}$$
$$= \frac{L}{\sqrt{N}}|p - p^\star| > \frac{L}{\sqrt{N}}\frac{\mu\epsilon}{\sqrt{N}} = \epsilon,$$

where the inequality comes from Lemma 7 and the fact that $L = N/\mu$.

We can now prove Equation (11b) by showing that if $|\mathcal{A}_T| < 2^{B_2(\epsilon)} = \lceil P/(2\epsilon)\rceil$, then there is $D \in \mathcal{D}_{P,L}$ such that $\text{dist}(p, \mathcal{P}^\star_D) > \epsilon$ for all $p \in \mathcal{A}_T$. Since $|\mathcal{A}_T| < \lceil P/(2\epsilon)\rceil$, it follows from Lemma 7 that there exists $p^\star \in [0,P]$ such that $|p - p^\star| > \epsilon$ for all $p \in \mathcal{A}_T$. Let $D \in \mathcal{D}_{L,P}$ be the dual problem given in Example 1 with $C = MN - p^\star L$. Then $p^\star$ is the unique optimal solution of D. Therefore, $\text{dist}(p, \mathcal{P}^\star) = |p - p^\star| > \epsilon$ for all $p \in \mathcal{A}_T$.

b) The proof follows similar arguments as the proof of Lemma 1-a). Let the quantization scheme $q \in \mathcal{Q}_P$ and $\beta < b$ be given. Similarly as before, consider a subset $\mathcal{A}_\beta \subseteq [0,P]$ of all possible outcomes of $p(T(q,\beta)) \in [0,P]$. Since β bits have been communicated we have that $|\mathcal{A}_\beta| = 2^\beta < 2^b = \lceil P/2\delta\rceil$, where $\delta = P/2^{b+1}$. Therefore, by Lemma 7 in Appendix A there exists $p^\star \in [0,P]$ such that $|p - p^\star| > \delta$ for all $p \in \mathcal{A}_\beta$. The proof now follows the same steps as used to prove parts a), where we use the primal problem given in Example 1 with

$C = MN - p^\star L$ and the associated dual problem to obtain Equations (12a) and (12b). ■

As shown in the proof, the lower bound is obtained by constructing specific primal and dual problems that are difficult for each quantization scheme. We now show that these lower bounds are tight by providing matching upper bounds.

C. General Upper Bounds

We now provide upper bounds on the general communication complexity. The upper bounds are obtained by constructing a particular quantization scheme that performs well for all primal and dual problem. In particular, we consider the following quantization scheme.

Lemma 2 (Upper Bound)**.** *Consider the quantization scheme* $q = (p(0), \theta_t(\cdot)) \in \mathcal{Q}_P$ *given by*

$$p(0) = P/2, \tag{14a}$$

$$\theta_t(d_0, \dots, d_t) = \begin{cases} P/2^{t+2} & \text{if } d_t \geq 0 \\ -P/2^{t+2} & \text{if } d_t < 0 \end{cases} \quad \text{for } t \in \mathbb{N}. \tag{14b}$$

Then for all $R \in \mathcal{R}_{\mu,N,P}$ *and* $D \in \mathcal{D}_{L,P}$*, where* $L = N/\mu$*, following holds:*

$$||\mathbf{x}_R^q(T(q,b)) - \mathbf{x}_R^\star|| \leq \epsilon \text{ if } b \geq B_1(\epsilon) \text{ for all } \epsilon > 0, \tag{15a}$$

$$\text{dist}(p_D^q(T(q,b)), \mathcal{P}_D^\star) \leq \epsilon \text{ if } b \geq B_2(\epsilon) \text{ for all } \epsilon > 0, \tag{15b}$$

and

$$||\mathbf{x}_R^q(T(q,b)) - \mathbf{x}_R^\star|| \leq \frac{P\sqrt{N}}{\mu 2^{b+1}} \quad \text{for all } b \in \mathbb{N}_0, \tag{16a}$$

$$\text{dist}(p_D^q(T(q,b)), \mathcal{P}_D^\star) \leq \frac{P}{2^{b+1}} \quad \text{for all } b \in \mathbb{N}_0. \tag{16b}$$

Proof: We start by showing that for the quantization scheme q in Equation (14) and any $D \in \mathcal{D}_{L,P}$ we have

$$\text{dist}(p_D^q(t), \mathcal{P}_D^\star) \leq \frac{P}{2^{t+1}}, \quad \text{for all } t \in \mathbb{N}_0. \tag{17}$$

We need to consider two cases, i) there exists $p^\star \in \mathcal{P}_D^\star$ such that $D'(p^\star) = 0$ and ii) no such $p^\star \in \mathcal{P}_D^\star$ exists. In case i) the quantization scheme reduces to the bisection method for finding a root of the function $D'(\cdot)$ on the interval $[0, P]$, yielding Equation (17). In case ii) $\mathcal{P}_D^\star$ only contains the element $p^\star = 0$; this is because $p^\star$ must be on the boundary of $[0, P]$ and $p^\star = P$ is not possible by Definition 2 of the class $\mathcal{D}_{L,P}$. Since $p^\star = 0$ and $D'(p^\star) \neq 0$, we have from Lemma 6 in Appendix A that $D'(p) > 0$ for all $p \in [0, P]$. Therefore, the algorithm reduces to $p(t) = P/2^{t+1}$ which together with the fact that $p^\star = 0$ yields Equation (17).

We can now prove Equation (15a). Given a $R = (\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$, set $\mathbf{x}_R(p) = (x_1(p), \dots, x_N(p))$ where $x_i(p) = [(U_i')^{-1}(p)]_{m_i}^{M_i}$ [cf. Equation (6a)]. Then the function $x_i(p)$ is $1/\mu$ Lipschitz continuous. This follows from the fact that $(U_i')^{-1}(p)$ is $1/\mu$ Lipschitz continuous [29, Equation (2.1.20)] and from the non-expansive property of the projection $[\cdot]_{m_i}^{M_i}$ [30, Proposition B.11 (c)]. Therefore, for all $p^\star \in \mathcal{P}^\star$ we have

$$||\mathbf{x}_R(p) - \mathbf{x}_R^*|| = \sqrt{\sum_{i=1}^N (x_i(p) - x_i^\star)^2} \leq \frac{\sqrt{N}}{\mu} |p - p^\star|,$$

and

$$||\mathbf{x}_R^q(t) - \mathbf{x}_R^*|| \leq \frac{\sqrt{N}}{\mu} \min_{p^\star \in \mathcal{P}^\star} |p(t) - p^\star| \leq \frac{\sqrt{N} \cdot P}{\mu \cdot 2^{t+1}}. \tag{18}$$

As a result, $||\mathbf{x}_R^q(t) - \mathbf{x}_R^*|| \leq \epsilon$ for all $t \in \mathbb{N}_0$ such that

$$t \geq \log_2 \left(\frac{\sqrt{N} P}{2\mu\epsilon} \right).$$

Equations (15b), (16a), and (16b) follow by rearranging Equations (17) and (18). ■

In the optimal quantization scheme in Lemma 2 only one bit of information is communicated per iteration and hence $T(q, b) = t$. Moreover, the optimal quantizers $\theta_t(d_0, \dots, d_t)$ do not rely on the parameters μ, N, L or on any history information, since $\theta_t(d_0, \dots, d_t)$ only depends on d_t. However, we will need to use the history information to generate optimal primal feasible quantization schemes, which are studied in Section IV.

This fast convergence rate of the upper bound of the dual iterates in Equation (16b) of Lemma 2 might look too good. In fact, it is better than the best possible convergence rate for general gradient methods when minimizing convex functions with Lipschitz continuous gradients, see Theorem 2.1.7 in [29]. However, in the proof we have exploited the fact that the dual problem is one-dimensional to achieve the convergence rate; whereas the lower bounds in [29] hold for multi-dimensional problems.

IV. Primal Feasible Quantization Schemes

When implementing the resource allocation algorithms from Section II-B in physical systems, it is essential that primal feasibility is ensured at each iteration. Otherwise, if primal feasibility is not ensured, then users may consume more resource than is available and overload the system. This is unacceptable because it causes blackouts in power networks, or outages in wireless networks. This motivates the following study of Primal-Feasible (PF) quantization schemes.

A. Primal Feasible (PF) Quantization Schemes

We are interested in the subset of all quantization schemes $q \in \mathcal{Q}_P$ that ensure the feasibility of the primal iterates at every iteration and for all primal problems in $\mathcal{R}_{\mu,N,P}$. Such quantization schemes are formally defined as follows.

Definition 6 (Primal Feasible (PF) Quantization Schemes)**.** *We say that the quantization scheme* $q \in \mathcal{Q}_P$ *is* Primal Feasible (PF)*, with respect to the problem class* $\mathcal{R}_{\mu,N,P}$*, if for every* $R = (\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ *the iterates* $\mathbf{x}_R^q(t)$ *are feasible to the problem* R *for all* $t \in \mathbb{N}_0$*. That is*

$$\sum_{i=1}^N x_{i,R}^q(t) \leq C \quad \text{for all } t \in \mathbb{N}_0,$$

where $\mathbf{x}_R^q(t) = (x_{1,R}^q(t), \dots, x_{N,R}^q(t))$*. We denote by* $\mathcal{Q}_{\mu,N,P}^{PF}$ *the set of all PF-quantization schemes with respect to* $\mathcal{R}_{\mu,N,P}$*.*

PF-quantization schemes are practically desirable since they ensure that the users do not overuse the resources as the algorithm runs. However, it is still unclear how to how to design such quantization schemes. The following result demonstrates a key sufficient condition used later to design PF-quantization schemes.

Theorem 2. *Set $L = N/\mu$. If $p(0) = P$ and the following inequality holds for all $t \in \mathbb{N}_0$*

$$\theta_t(d_0, \dots, d_t) \le \frac{1}{L} d_t, \quad \textit{for all } d_t \ge 0, \tag{19}$$

then $(p(0), \theta_t(\cdot)) \in \mathcal{Q}^{PF}_{\mu,N,P}$.

Proof: See Appendix B-A. ∎

The theorem provides us with an easy-to-use design mechanism to produce PF-quantization schemes. Moreover, the first condition of the theorem, that $p(0) = P$, necessarily holds for all PF-quantization schemes.

Proposition 1. *A quantization scheme $(p(0), \theta_t(\cdot)) \in \mathcal{Q}_P$ is not a PF-quantization if $p(0) \in [0, P)$.*

Proof: See Appendix B-B. ∎

We now study the communication complexity of PF-quantization schemes.

B. Communication Complexity of PF-Quantization Schemes

We now study the communication complexity of PF-quantization schemes. We present the main results in this section, but leave many details to Sections IV-C and IV-D and to Appendix B.

When considering the communication complexity of PF-quantization schemes, the following complexity measures will be useful in addition to those defined in Section II-C.

Definition 7. *For each $\epsilon > 0$, $b \in \mathbb{N}$, and a quantization class $\bar{\mathcal{Q}} \subseteq \mathcal{Q}_P$, consider the following quantities*

$$\begin{aligned}\mathcal{E}^{\text{DualObj}}_{L,P}(\bar{\mathcal{Q}}, \epsilon) &= \min_{q \in \bar{\mathcal{Q}}} \max_{D \in \mathcal{D}_{L,P}} \sum_{t=0}^{T^{\text{DualObj}}(q,D,\epsilon)} r_q(t) \\ \mathcal{B}^{\text{DualObj}}_{L,P}(\bar{\mathcal{Q}}, b) &= \min_{q \in \bar{\mathcal{Q}}} \max_{D \in \mathcal{D}_{L,P}} D(p^q_D(T(q,b))) - D^\star.\end{aligned}$$

where $T^{\text{DualObj}}(q, D, \epsilon) = \min\left\{t \in \mathbb{N}_0 \middle| D(p^q_D(t)) - D^\star \le \epsilon\right\}$.

The measures $\mathcal{E}^{\text{DualObj}}_{L,P}(\bar{\mathcal{Q}}, \epsilon)$ and $\mathcal{B}^{\text{DualObj}}_{L,P}(\bar{\mathcal{Q}}, b)$ are the same as the ϵ- and b-complexities defined in Section II-C, except they use the optimality criterion $D(p) - D^\star$.

Because of the primal feasibility requirement, it is difficult to reduce the communication complexity of PF-quantization schemes to a single number, as we did for general quantization schemes in Section III. Nevertheless, we can give lower and upper bounds on the asymptotic behaviour of the complexity measures for PF-quantization schemes. We start by giving lower bounds. For that purpose we use the following big-Ω notation:

$$\begin{aligned}\Omega(g(b)) = \{h : \mathbb{N} \to \mathbb{R} \;|Bg(b) \le h(b) \text{ for all } b \ge b_0 \\ \text{and some } b_0 \in \mathbb{N} \text{ and } B \in \mathbb{R}_+\}.\end{aligned}$$

Informally, $h(b) \in \Omega(g(b))$ means that the function $h(b)$ dominates $g(b)$ up to some constant factor, for large enough b. We now provide lower bounds on the b-complexity for PF-quantization schemes.

Theorem 3 (PF-Lower Bound Complexity)**.** *For $k \in \mathbb{N}$ with $k \ge 4$ we have the following lower bounds on the b-complexity:*

$$\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{PF}_{\mu,N,P}, b) \in \Omega\left(1/b^{\frac{k}{k-1}}\right) \tag{20a}$$

$$\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P}, b) \in \Omega\left(1\right) \tag{20b}$$

$$\mathcal{B}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P}, b) \in \Omega\left(1/b^{\frac{k+1}{k-1}}\right). \tag{20c}$$

Proof: See Appendix B-C. ∎

Equations (20a)-(20c) give a lower bound on the convergence rate of the three b-complexities. For example, Equation (20b) shows that there exists a constant $K > 0$ such that for any PF-quantization q and any $b \in \mathbb{N}$, we can find a dual problem $D \in \mathcal{D}_{L,P}$ such that $\text{dist}(p^q_D(T(q,b)), \mathcal{P}^\star_D) \ge K$. Similarly, Equation (20a) shows that for any PF-quantization q there exists a primal problem $R \in \mathcal{R}_{\mu,N,P}$ such that the optimality criterion $||\mathbf{x}^q_R(T(q,b)) - \mathbf{x}^\star_R||$ converges at a rate that is at best proportional to $1/b^{\frac{k}{k-1}}$. Equation (20c) shows similar behaviour for the optimality criterion $D(p^q_D(T(q,b))) - D^\star$.

Theorem 3 also shows that the complexity measures $\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b)$ and $\mathcal{B}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b)$ cannot be upper bounded by a function on the form $g(b) = C/b^a$ where $a > 1$ and $C \in \mathbb{R}_+$. To formally assert this result we consider the following big-$\mathcal{O}$ notation.

$$\begin{aligned}\mathcal{O}(g(b)) = \{h : \mathbb{N} \to \mathbb{R} \;|h(b) \le Bg(b) \text{ for all } b \ge b_0 \\ \text{and some } b_0 \in \mathbb{N} \text{ and } B \in \mathbb{R}_+\}.\end{aligned}$$

Informally, $h(b) \in \mathcal{O}(g(b))$ means that the function $h(b)$ is dominated by $g(b)$ up to some constant factor, for large enough b. We have the following consequence of Theorem 3.

Corollary 1. *For all $a > 1$ we have $\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{PF}_{\mu,N,P}, b) \notin \mathcal{O}\left(1/b^a\right)$ and $\mathcal{B}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P}, b) \notin \mathcal{O}\left(1/b^a\right)$.*

Proof: See Appendix B-D. ∎

The result in the corollary clearly also holds for the complexity measure $\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b)$, because of Equation (20b) in Theorem 3. Corollary 1 shows that for any PF-quantization scheme the convergence rate of the quantities $||\mathbf{x}(T(q,b)) - \mathbf{x}^\star||$ and $D(p(T(q,b))) - D^\star$ as they approach zero, can be no better than $\mathcal{O}(1/b)$. We now provide big-$\mathcal{O}$ upper bounds on both the b- and ϵ-complexities. For the ϵ-complexity, we need the following big-$\mathcal{O}$ notation that describes the convergence behaviour when ϵ is close to 0.

$$\begin{aligned}\mathcal{O}_0(g(\epsilon)) = \{h : \mathbb{N} \to \mathbb{R} \;|h(\epsilon) \le Bg(\epsilon) \text{ for all } \epsilon \in [0, \epsilon_0] \\ \text{and some } \epsilon_0 > 0 \text{ and } B \in \mathbb{R}_+\}.\end{aligned}$$

Informally, $h(\epsilon) \in \mathcal{O}(g(\epsilon))$ means that the function $h(\epsilon)$ is dominated by $g(\epsilon)$ up to some constant factor for small enough ϵ. We have the following upper bounds on the communication complexity.

Theorem 4 (PF-Upper Bound Complexity)**.** *We have the following upper bounds on the ϵ-complexity:*

$$\mathcal{E}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{PF}_{\mu,N,P},\epsilon) \in \mathcal{O}_0\left(1/\epsilon^2\right) \tag{21a}$$
$$\mathcal{E}^{\text{Dual}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P},\epsilon) = \infty \qquad \textit{for all} \quad \epsilon < P \tag{21b}$$
$$\mathcal{E}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P},\epsilon) \in \mathcal{O}_0\left(1/\epsilon\right). \tag{21c}$$

We have the following bounds on the b-complexity:

$$\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{PF}_{\mu,N,P},b) \in \mathcal{O}\left(1/\sqrt{b}\right) \tag{22a}$$
$$\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P},b) \in \mathcal{O}\left(1\right) \tag{22b}$$
$$\mathcal{B}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{PF}_{\mu,N,P},b) \in \mathcal{O}\left(1/b\right). \tag{22c}$$

Proof: See Appendix B-E. ∎

The upper bound in Equations (21a) and (21c) show that there exists a PF-quantization scheme q that can find primal and dual variables $\mathbf{x}$ and p such that $||\mathbf{x}-\mathbf{x}^\star|| \leq \epsilon$ and $D(p)-D^\star \leq \epsilon$ by using number of bits that is proportional to $1/\epsilon^2$ and $1/\epsilon$, respectively, for small enough ϵ. Similarly, the upper bound in Equations (22a) and (22c) show that there exists a PF-quantization scheme q such that the accuracy of the primal iterates $||\mathbf{x}(T(q,b))-\mathbf{x}^\star||$ and the dual objective function $D(p(T(q,b)))-D^\star$ converge to zero at a rate proportional to $1/\sqrt{b}$ and $1/b$, respectively, as the number of bits b diverges to infinity. On the other hand, the upper bounds in Equations (21b) and (22b) on the accuracy of the dual iterates $\text{dist}(p(t),\mathcal{P}^\star)$ are less promising. They show that the dual iterates $p(t)$ can converge at an arbitrarily slow rate to $p^\star$. This is consistent with the fact that iterates of gradient methods can converge at an arbitrarily slow rate to an optimal solution when minimizing convex functions with Lipschitz continuous gradients, while the objective function can converge at the rate $\mathcal{O}(1/t^2)$ [29, Theorem 2.1.7].

These results illustrate a trade-off between ensuring primal feasibility and the convergence rate. To ensure primal feasibility, the convergence rate of the quantity $||\mathbf{x}(T(q,b))-\mathbf{x}^\star||$ can be no better than $\mathcal{O}(1/b)$, as showed in Corollary 1. On the other hand, when there are no primal feasibility requirements then the convergence rate of $||\mathbf{x}(T(q,b))-\mathbf{x}^\star||$ is $\mathcal{O}(1/2^b)$, as proved in Section III.

C. Lower Bounds for PF-Quantization Schemes

To prove the lower bounds in Theorem 3 we use the following result.

Lemma 3. *Take $P \in \mathbb{R}_+$, $\mu > 0$, $N \in \mathbb{N}$, and $L = N/\mu$. Then for every $k \in \mathbb{N}$ with $k \geq 4$ there exists a primal problem $R_k \in \mathcal{R}_{\mu,N,P}$ and an associated dual problem $D_k \in \mathcal{D}_{L,P}$ such that the following holds for all $q \in \mathcal{Q}^{PF}_{\mu,N,P}$:*

$$||\mathbf{x}^q_{R_k}(t)-\mathbf{x}^\star_{R_k}|| \geq \frac{P\sqrt{N}}{k\mu}\left(\frac{1}{1+2(t{+}1)}\right)^{\frac{k}{k-1}}, \tag{23a}$$
$$\text{dist}(p^q_{D_k}(t),\mathcal{P}^\star_{D_k}) \geq P\left(\frac{1}{1+2(t{+}1)}\right)^{\frac{1}{k-1}}, \tag{23b}$$
$$D(p^q_{D_k}(t)) - D^\star_k \geq (D_k(p(0)) - D^\star_k)\left(\frac{1}{1{+}2(t{+}1)}\right)^{\frac{k+1}{k-1}}. \tag{23c}$$

Proof: See Appendix B-F. ∎

Note that the lower bounds in Equations (23a)-(23c) are given in number of iterations t, not in number of bits b. Therefore, it does not matter how many bits are communicated per iteration, these lower bound always hold. In particular, since at least one bit is communicated at each iteration (since $r(t) > 0$ from Definition 3), t can be replaced by the number of bits b.

Lemma 3 holds even when no quantization is done, i.e., for standard dual gradient methods. Therefore, these results also show a trade-off between ensuring primal feasibility and convergence rate of standard dual gradient methods when dual gradient information is used. When ensuring primal feasibility, the best convergence rate of $D(p(t))-D^\star$ that can be achieved is $\mathcal{O}(1/t)$.[3] However, if primal feasibility is not required, the literature shows that gradient methods that minimize general convex dual functions D with Lipschitz continuous gradient can have the convergence rate $\mathcal{O}(1/t^2)$ [29, Chapter 2]. Moreover, for the dual problem studied in this paper a linear convergence can be ensured, as proved in Section III.

D. Upper Bounds for PF-Quantization Schemes

The proof of Theorem 4 is obtained by using the following PF-quantization scheme.

Lemma 4. *Consider the quantization scheme $q = (p(0),\theta(\cdot)) \in \mathcal{Q}_P$ with $p(0) = P$, and*

$$Q_0 = P, \tag{24a}$$
$$\theta_t(d_0,\ldots,d_t) = \begin{cases} 0 & \textit{if } d_t < Q_t L \\ Q_t & \textit{if } d_t \geq Q_t L \end{cases} \tag{24b}$$
$$Q_{t+1} = \begin{cases} \frac{Q_t}{2} & \textit{if } d_t < Q_t L \\ Q_t & \textit{if } d_t \geq Q_t L. \end{cases} \tag{24c}$$

Then $q \in \mathcal{Q}^{PF}_{\mu,N,P}$. Moreover, for all $R \in \mathcal{R}_{\mu,N,P}$ and $D \in \mathcal{D}_{L,P}$, where $L = N/\mu$, the following holds:

$$||\mathbf{x}^q_R(t)-\mathbf{x}^\star_R|| \leq \frac{4\sqrt{N}P}{\mu\sqrt{t}} \quad \textit{for all} \quad t \geq 1, \tag{25a}$$
$$D(p^q_D(t)) - D^\star \leq \frac{16LP^2}{t} \quad \textit{for all } t \geq 1. \tag{25b}$$

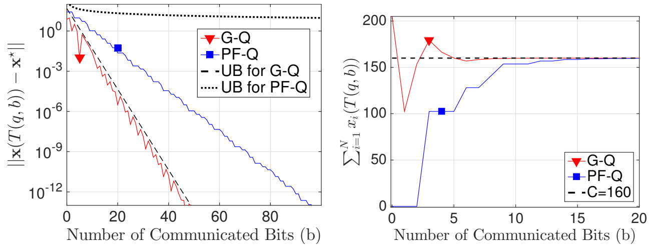
Proof: See Appendix B-G ∎

The number of iterations t can be replaced by the number of bits b, since the quantization scheme uses 1 bit communication at each iteration. With the quantization scheme in Lemma 4, the algorithm in Equation (7) takes the gradient step $\Delta(t) = Q_t$ provided that $Q_t \leq D'(p^q_D(t))/L$, which ensures primal feasibility because of Lemma 2. Otherwise, if primal feasibility cannot be assured with the step Q_t then $\Delta(t) = 0$ and Q_t is halved, i.e., $Q_{t+1} = Q_t/2$. A key property of this quantizatin that is used to prove the lemma is that every time a step is taken, i.e., when $\Delta(t) = Q_t$, then the step length is proportional to $D'(p^q_D(t))/L$ (See Equation (39) in the proof of Lemma 4).

V. Numerical Studies

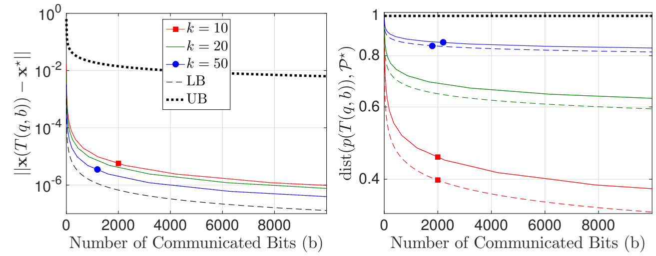
We now illustrate the results with two numerical studies.

[3] This can be proved by following the steps in the proof of Corollary 1.



(a) Primal iterates (b) Feasibility of the iterates

Fig. 2: Convergence behaviour of the quantization schemes in Lemma 2 (G-Q) and Lemma 4 (PF-Q) for solving Problem (27): (a) shows the primal iterates and (b) the feasibility of the primal iterates as a function of communicated bits.



(a) Primal iterates (b) Dual Iterates

Fig. 3: Convergence behaviour of the quantization schemes in Lemma 4 (PF-Q) for solving the problem in Example 2: (a) shows the primal iterates and (b) the dual iterates as function of communicated bits.

A. Proportionally Fair Power Allocation in Micro Grids

Consider a power supplier in a micro grid with the task of supplying N users (or devices) with C units of power. Denote by $d_i > 0$ and $x_i \in [0, d_i]$, respectively, the power demand and the actual power allocation of user $i = 1, \dots, N$. The supplier's task is to allocate the power in an efficient and a proportionally fair manner [1], [31]–[34]:

Definition 8. *A power allocation* $\mathbf{x}^\star \in \mathbb{R}^N$ *is* proportionally fair *if there exists* $\gamma \in (0, 1]$ *such that*

$$\frac{d_i - x_i^\star}{d_i} = \gamma, \quad \textit{for all} \quad i = 1, \dots, N. \tag{26}$$

A power allocation $\mathbf{x}^\star \in \mathbb{R}^N$ *is* efficient *if either* $x_i^\star = d_i$ *for all* $i = 1, \dots, N$ *or* $\sum_{i=1}^N x_i^\star = C$.

In other words, a power allocation is proportionally fair if the deviation of the power allocation of each user from its demand is proportional the same. Similarly, a power allocation is efficient if either the demand of every user is satisfied or the power is used up. An efficient and a proportionally fair power allocation can be found by solving the following optimization problem (see Lemma 8 in Appendix A)

$$\begin{aligned} \underset{x_1, \cdots, x_N}{\text{maximize}} \quad & U(\mathbf{x}) := \sum_{i=1}^N -\frac{1}{2d_i}(d_i - x_i)^2 \\ \text{subject to} \quad & \sum_{i=1}^N x_i \le C, \\ & x_i \in [0, d_i], \end{aligned} \tag{27}$$

Figure 2 plots the results of solving Problem (27) when $N = 40$, $C = 160$, and d_i is taken uniformly at random from the interval $[5, 15]$, for each $i = 1, \dots, N$. This problem is in the class $\mathcal{R}_{\mu,N,P}$ where $\mu = 1/15$ and $P = 1$. Therefore, the dual problem is in the class $\mathcal{D}_{L,P}$ where $L = 15N$. To solve this problem, we use the general quantization scheme in Lemma 2 indicated by G-Q (red curves marked with a filled circle) and the primal feasible quantization scheme in Lemma 4 indicated by PF-Q (blue curves marked with a filled square).

Figure 2a shows the convergence of the primal iterates as a function of the number of communicated bits. The upper bounds on the quantization schemes G-Q and PF-Q in Lemmas 2 and 4, respectively, are demonstrated by the black dashed and dotted lines. The figure shows that the convergence is faster for G-Q than for PF-Q. However, as shown in Figure 2b, for G-Q, the primal variables are infeasible at some iterations of the algorithm, which in practice could cause a blackout in the micro grid. On the other hand, for PF-Q, the primal iterates are feasible at every iteration of the algorithm, which is not surprising since PF-Q is a primal feasible quantization scheme (see Definition 6 and Lemma 4 in Section IV). Therefore, G-F archives fast convergence at the cost of the security of the micro grid.

B. Extreme Example

Figure 3 illustrates the results of using the PF-quantization scheme in Lemma 4 when solving the problem given in Example 2 in Appendix B-F when $P = L = 1$, $N = 40$, $\alpha = 0$, and $k = 10, 20, 50$. The convergence of the primal and dual iterates are, respectively, illustrated in Figure 3a and Figure 3b. The dotted lines illustrate upper bounds on the convergence, the upper bound in Figure 3a comes from Lemma 4 and the upper bound in Figure 3b is obtained from that $\text{dist}(p, \mathcal{P}^\star) \le P$ for all $p \in [0, P]$. The dashed lines illustrate lower bounds on the convergence obtained from Lemma 3. In Figure 3a, we have only plotted the smallest lower bound achieved when $k = 50$, since the lower bounds for $k = 10, 20$ are all in similar range.

Figure 3a shows that the primal iterates are upper and lower bounded by sequences $A_1/\sqrt{b}$ and $A/b^{k/(k-1)}$, respectively, for some constants a $A_1, A_2 \in \mathbb{R}_+$ and $k = 10, 20, 50$. Since the lower bound can be achieved for any $k \ge 4$, these results show that the convergence rate of the primal iterates can be no better than $\mathcal{O}(1/b)$, as we proved in Section III-B. Figure 3b shows that the convergence of the dual iterates gets slower as k increases. This shows why the communication complexity of finding approximately optimal dual variables p with respect to the optimality measure $\text{dist}(p, \mathcal{P}^\star)$ is so bad, e.g., $\mathcal{E}_{L,P}^{\text{Dual}}(\mathcal{Q}_{\mu,N,P}^{\text{PF}}, \epsilon) = \infty$ for $\epsilon < P$ in Theorem 4.

VI. Conclusions

This paper studies the communication complexity of resource allocation problems. The results show that for the considered problem class the number of bits needed to find an ϵ solution accuracy grows proportionally to $\log_2(1/\epsilon)$.

Similarly, the smallest solution accuracy that can be found when b bits are used decreases proportionally to $1/2^b$. We also consider the communication complexity for a stricter communication model that allows only algorithms that ensure the problem's feasibility at every iteration. This restriction is motivated by that resource allocation algorithms are often executed online, where the users consume the resources while the algorithm is running. Therefore, if the users consume more resources than available, then they can overload the system, which could for example cause blackouts in power networks. With this feasibility requirement, we showed that the best possible convergence rate of the solution accuracy is proportional to $1/b$, where b is the number of communicated bits. Therefore, these results illustrate a trade-off between feasibility and fast convergence. Future work will consider the communication complexity of more general classes of resource allocation problems.

Appendix A
Important Lemmas

Lemma 5. *For all* $(\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ *and the associated dual function* D*, the following holds:*

a) *D has L-Lipschitz continuous gradient, where $L = N/\mu$.*
b) *The dual Problem* (3) *has an optimal solution in the interval* $[0, P]$.

Proof: a) See Lemma II.2 in [4].

b) Strong duality holds between the primal Problem (1) and the dual Problem (3), see [13, Lemma 1]. Therefore, D an optimal solution $p^\star \in \mathbb{R}_+$. We now prove that if $p^\star \geq P$ then P is also an optimal solution to (3). Since $P \geq \max_{i\in\mathcal{N}} U_i'(m_i)$, we have $P \geq U_i'(m_i)$ for all $i \in \mathcal{N}$. Then $m_i \geq (U_i')^{-1}(P)$ for all $i \in \mathcal{N}$ since $(U_i')^{-1}(\cdot)$ is a decreasing function, due to the concavity of U_i and the fact that the inverse of decreasing function is decreasing. Then $x_i(P) = [(U_i')^{-1}(P)]_{m_i}^{M_i} = m_i$ and hence $x_i(p) = m_i$ for $p \geq P$, since $(U_i')^{-1}(\cdot)$ is decreasing. Therefore, if $p^\star \geq P$ then P is an optimal solution to (3). So D must have an optimal solution in $[0, P]$. ■

Lemma 6. *Consider the primal and dual problems* (1) *and* (3)*. Let $\mathcal{P}^\star = [\underline{p}^\star, \bar{p}^\star]$ be the set of dual optimizers. Then for $p \in \mathbb{R}_+$, the following three conditions are equivalent:*

a) *$(x_i(p))_{i\in\mathcal{N}}$ is primal feasible,*
b) *$D'(p) \geq 0$, and*
c) *$p \geq \underline{p}^\star$.*

Proof: a) $\iff$ b): $D'(p) = C - \sum_{i=1}^N x_i(p)$ from Equation (6b). Therefore, if $D'(p) < 0$ then $C < \sum_{i=1}^N x_i(p)$ and if $0 \leq D'(p)$ then $\sum_{i=1}^N x_i(p) \leq C$.

b) $\iff$ c): For $p^\star \in \mathcal{P}^\star$ then $D'(p^\star) = 0$ if $p^\star > 0$ and $D'(p^\star) \geq 0$ if $p^\star = 0$. Therefore, if $p \geq \underline{p}^\star$ then $D'(p) \geq D'(\underline{p}^\star) \geq 0$, because $D'(\cdot)$ is increasing since D is convex. Similarly, if $p \in \mathbb{R}_+$ and $p < \underline{p}^\star$ then $D'(p) < D'(\underline{p}^\star) = 0$, since $0 < \underline{p}^\star$. ■

Lemma 7. *Let $\mathcal{A}$ be a subset of the interval $[0, P]$ with the cardinality $|\mathcal{A}| < \lceil P/(2\delta) \rceil$ for some $\delta > 0$. Then there exists $\bar{p} \in [0, P]$ such that $|p - \bar{p}| > \delta$ for all $p \in \mathcal{A}$.*

Proof: Suppose that $\mathcal{A} \subseteq [0, P]$ with $|\mathcal{A}| < \lceil P/(2\delta) \rceil$. Then $2\delta|\mathcal{A}| < P$ so the intervals $[p-\delta, p+\delta]$ for all $p \in \mathcal{A}$ do not cover the interval $[0, P]$. Therefore, there exist $\bar{p} \in [0, P]$ such that $p \notin \bigcup_{p\in\mathcal{P}} [p - \delta, p + \delta]$. ■

Lemma 8. *The unique optimal solution* $\mathbf{x}^\star$ *of Problem* (27) *is efficient and proportionally fair power allocation.*

Proof: Consider first the case when $\sum_{i=1}^N d_i \leq C$. Then $x_i^\star = d_i$ is the optimal solution to Problem (27), hence $\mathbf{x}^\star$ is efficient power allocation. Moreover, since $x_i^\star = d_i$, Equation (26) holds with $\gamma = 0$. Therefore, $\mathbf{x}^\star$ is a proportionally fair allocation.

Consider next the case when $\sum_{i=1}^N d_i > C$. Then the KKT optimality conditions of Problem (27) show that the optimal solution is $x_i^\star = d_i(1 - p^\star)$, where $p^\star = (\sum_{i=1}^N d_i - C)/(\sum_{i=1}^N d_i)$, is the optimal dual variable associated to the constraint $\sum_{i=1}^N x_i \leq C$. Then $\mathbf{x}^\star$ is a proportionally fair allocation with $\gamma = p^\star$. Moreover, we have $\sum_{i=1}^N x_i^\star = \sum_{i=1}^N d_i(1 - p^\star) = \sum_{i=1}^N d_i - \sum_{i=1}^N d_i + C = C$, which proves that $\mathbf{x}^\star$ is an efficient power allocation. ■

Appendix B
Proofs for Section IV

A. Proof of Theorem 2

Proof: We prove this result by induction. Let some primal problem $R \in \mathcal{R}_{\mu,N,P}$ and the associated dual problem $D \in \mathcal{D}_{L,P}$ be given. Then the initial value $\mathbf{x}_R^q(0) = [x_i(P)]_{i\in\mathcal{N}}$ is feasible by Lemma 5 and Lemma 6-a) in Appendix A. We next show that if $\mathbf{x}_R^q(t)$ is feasible then $\mathbf{x}_R^q(t+1)$ is also feasible. We consider separately the two cases when $\Delta(t) \leq 0$ and $\Delta(t) > 0$, where $\Delta(t) = \theta_t(D'(p_D^q(0)), \dots, D'(p_D^q(t)))$ is defined in Equation (7b).

Suppose that $\Delta(t) \leq 0$. Then we have from Equation (7c) that $p_D^q(t+1) \geq p_D^q(t)$. Moreover, we have from Lemma 6-c) in Appendix A that $p_D^q(t) \geq p^\star := \min \mathcal{P}_R^\star$. Therefore, $p_D^q(t+1) \geq p_D^q(t) \geq p^\star$ so $\mathbf{x}_R^q(t+1)$ is feasible by Lemma 6-c) in Appendix A.

Suppose that $\Delta(t) > 0$. Then $p_D^q(t) > p_D^q(t+1)$ so

$$\begin{aligned} D'(p_D^q(t)) - D'(p_D^q(t+1)) &\leq L\ (p_D^q(t) - p_D^q(t+1)) \\ &\leq L\ \theta_t(D'(p_D^q(0)), \dots, D'(p_D^q(t))) \\ &\leq D'(p_D^q(t)), \end{aligned} \tag{28}$$

where the first inequality is obtained by using that $D'(\cdot)$ is L-Lipschitz continuous and that $p_D^q(t) - p_D^q(t+1) > 0$, the second inequality is obtained by using Equation (7c) and the fact that $\Delta(t) > 0$, and the final inequality is obtained by using Equation (19) and that $D'(p_D^q(t)) \geq 0$ by Lemma 6-b) in Appendix A. By rearranging (28) we get that $D'(p_D^q(t+1)) \geq 0$. Therefore, $\mathbf{x}_R^q(t)$ is feasible by Lemma 6-b). ■

B. Proof of Proposition 1

Proof: Take some $p(0) \in [0, P)$ and consider the primal problem in Example 1 with $C = MN - L(p(0)+P)/2$. Then the dual problem has the unique solution $p^\star = (p(0) + P)/2$. Since $p(0) < p^\star$ it follows from Lemma 6 in Appendix A that $[x_i(p(0))]_{i\in\mathcal{N}}$ is infeasible. ■

C. Proof of Theorem 3

Proof: The proof follows directly from Lemma 3 in Section IV-C. For example, to prove Equation (20a) for any $k \geq 4$, we consider the primal problem $R_k \in \mathcal{R}_{\mu,N,P}$ from Lemma 3. Then from Equation (23a) in Lemma 3 we have

$$\begin{aligned}\mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b) =& \min_{q\in\mathcal{Q}^{\text{PF}}_{\mu,N,P}} \max_{R\in\mathcal{R}_{\mu,N,P}} ||\mathbf{x}^q_R(T(q,b)) - \mathbf{x}^\star_R||,\\ \geq& \min_{q\in\mathcal{Q}^{\text{PF}}_{\mu,N,P}} ||\mathbf{x}^q_{R_k}(T(q,b)) - \mathbf{x}^\star_{R_k}||\\ \geq& \frac{P\sqrt{N}}{k\mu}\left(\frac{1}{1+2(b+1)}\right)^{\frac{k}{k-1}},\end{aligned}$$

where we have used that $T(q,b) \leq b$ since at least one bit is used per iteration. The lower bound in Equation (20c) can be proved in the same way, except by using Equation (23c) from Lemma 3 instead of Equation (23a).

To prove Equation (20b), we can use Lemma 3 to prove that for all $\epsilon \in [0, P)$ following holds

$$\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b) \geq \epsilon, \quad \text{for all } b \in \mathbb{N}_0.$$

Let $b \in \mathbb{N}_0$ be given. Consider $D_k \in \mathcal{D}_{L,P}$ from Lemma 3, where $k \in \mathbb{N}$ is such that $k > \max\{4, \log_{P/\epsilon}(3+2b)+1\}$. Then from Equation (23b) in Lemma 3 we have

$$\begin{aligned}\mathcal{B}^{\text{Dual}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b) =& \min_{q\in\mathcal{Q}^{\text{PF}}_{\mu,N,P}} \max_{D\in\mathcal{D}_{L,P}} \text{dist}(p^q_D(T(q,b)), \mathcal{P}^\star_D),\\ \geq& \min_{q\in\mathcal{Q}^{\text{PF}}_{\mu,N,P}} \text{dist}(p^q_{D_k}(T(q,b)), \mathcal{P}^\star_{D_k}),\\ \geq& P\left(\frac{1}{3+2b}\right)^{\frac{1}{k-1}} \geq P\left(\frac{1}{3+2b}\right)^{\frac{1}{\log_{P/\epsilon}(3+2b)}}\\ =& P\left(\frac{1}{(3+2b)^{\log_{3+2b}(P/\epsilon)}}\right) = P\frac{1}{P/\epsilon} = \epsilon,\end{aligned}$$

where we have used that $1/\log_a(b) = \log_b(a)$ for $a, b > 0$. ■

D. Proof of Corollary 1.

Proof: We prove the result by contradiction. Set $h(b) := \mathcal{B}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b)$ and suppose that $h(b) \in \mathcal{O}(1/b^a)$, for some $a > 1$. Then there exists $A_1 \in \mathbb{R}_+$ and $B_1 \in \mathbb{N}$ such that $h(b) \leq A_1/b^a$ for all $b \geq B_1$. Now choose $k \geq 4$ so that $a_2 := k/(k-1) < a$, such k exists since $\lim_{k\to\infty} k/(k-1) = 1$ and $a > 1$. From Theorem 3, $h(b) \in \Omega\left(1/b^{\frac{k}{k-1}}\right)$. So there exists $A_2 \in \mathbb{R}_+$ and $B_2 \in \mathbb{N}$ such that $h(b) \geq A_2/b^{a_2}$ for all $b \geq B_2$. Now choose $B = \max\{B_1, B_2, (A_1/A_2)^{1/(a-a_2)}\}$. Then by rearranging the inequality $b \geq B \geq (A_1/A_2)^{1/(a-a_2)}$ we have $h(b) \geq A_2/b^{a_2} \geq A_1/b^a$ for all $b \geq B$. This contradicts the fact that $h(b) \leq A_1/b^a$ for all $b \geq B_1$. Therefore, no such A_1 and B_1 can exists so $h \notin \mathcal{O}(1/b^a)$. The proof for $\mathcal{B}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, b)$ follows the same steps. ■

E. Proof of Theorem 4

Proof: The proof follows directly from Lemma 4 in Section IV-D. For example, to prove Equations (21a) and (21c), consider the quantization scheme $q \in \mathcal{Q}^{\text{PF}}_{\mu,N,P}$ given in Lemma 4. For any $\epsilon > 0$ define

$$b_1(\epsilon) = \left\lceil \frac{16LP^2}{\mu\epsilon^2} \right\rceil \quad \text{and} \quad b_2(\epsilon) = \left\lceil \frac{16LP^2}{\epsilon} \right\rceil,$$

Then from Lemma 4, we have

$$\begin{aligned}||\mathbf{x}^q_R(T(q,b_1(\epsilon)) - \mathbf{x}^\star_R|| \leq& \epsilon \quad \text{for all } R \in \mathcal{R}_{\mu,N,P}\\ D(p^q_R(T(q,b_2(\epsilon))) - D^\star \leq& \epsilon \quad \text{for all } D \in \mathcal{D}_{L,P}.\end{aligned}$$

Therefore we have

$$\begin{aligned}\mathcal{E}^{\text{Prim}}_{\mu,N,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, \epsilon) \leq& \left\lceil \frac{16LP^2}{\mu\epsilon^2} \right\rceil \leq \frac{32LP^2}{\mu\epsilon^2} \quad \text{for all } \epsilon \leq \frac{4\sqrt{L}P}{\sqrt{\mu}}\\ \mathcal{E}^{\text{DualObj}}_{L,P}(\mathcal{Q}^{\text{PF}}_{\mu,N,P}, \epsilon) \leq& \left\lceil \frac{16LP^2}{\epsilon} \right\rceil \leq \frac{32LP^2}{\mu\epsilon} \quad \text{for all } \epsilon \leq 16LP^2,\end{aligned}$$

where we have used that $\lceil z \rceil \leq 2z$ for $z \geq 1$. Equation (21a) can be proved by following similar arguments as used to prove Equation (20b) in Theorem 3.

To prove Equation (22a) we also consider the quantization scheme $q \in \mathcal{Q}^{\text{PF}}_{\mu,N,P}$. Then the upper bounds in Equation (22a) come directly by Equation (25a) in Lemma 4 and the fact that only one bit is used per iteration in the quantization scheme q. The upper bound in Equation (22b) can be proved by using that $\text{dist}(p^q_D(t), \mathcal{P}^\star_D) \leq P$ for all $D \in \mathcal{D}_{L,P}$, since $p^q_D(t) \in [0, P]$ and every dual problem in $\mathcal{D}_{L,P}$ has a optimal solution in $[0, P]$ from Lemma 5 in Appendix A. The upper bound in Equation (22c) can be proved similarly as the upper bound in Equation (22a). ■

F. Proof of Lemma 3

The prove is obtained by considering the primal and dual problems in the following example.

Example 2. *Take some $N \in \mathbb{N}_0$, $L, P \in \mathbb{R}_+$, $\alpha \in [0, P]$, and $k \in \mathbb{N}_0$, where $L > 0$ and $k \geq 4$. Set $C = PL/k$,*

$$\beta = \frac{C}{N}\left(1 - \frac{\alpha^k}{P^k}\right), \qquad M = \frac{C}{N}\left(1 - \frac{\alpha}{P}\right),$$

$$A_1 = \frac{Ck\alpha^{k+1}}{N(k+1)P^k} - \frac{2L\alpha^2}{N}, \qquad A_2 = \frac{L\alpha^2}{2} - \frac{L\alpha^{k+1}}{(k+1)P^k},$$

$m_i = 0$ and $M_i = M$, for $i = 1, \dots, N$. Let the utility function of each user $i = 1, \dots, N$ be given by

$$U_i(x_i) = \begin{cases} -\frac{PCk}{N(k+1)}\left(1 - \frac{N}{C}x_i\right)^{\frac{k+1}{k}} + A_1 & \text{if } x \in [0, \beta]\\ -\frac{2}{NL}\left(C + L\alpha - \frac{C}{P^k}\alpha^k - Nx_i\right)^2 & \text{if } x \in [\beta, M]\end{cases}$$

$$U'_i(x_i) = \begin{cases} P\left(1 - \frac{N}{C}x_i\right)^{\frac{1}{k}} & \text{if } x \in [0, \beta]\\ \frac{1}{L}\left(C + L\alpha - \frac{C}{P^k}\alpha^k - Nx_i\right) & \text{if } x \in [\beta, M].\end{cases}$$

It can be verified that $U_i(\cdot)$ is μ-strongly concave with $\mu = N/L$. Therefore, $(\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ and the associated dual problem is in the class $\mathcal{D}_{L,P}$. We also have that

$$x_i(p) = \begin{cases} \frac{C}{N}\left(1 - \frac{\alpha^k}{P^k}\right) + \frac{L}{N}(1\alpha - p) & \text{if } p \in [0, \alpha]\\ \frac{C}{N}\left(1 - \frac{p^k}{P^k}\right) & \text{if } p \in [\alpha, P],\end{cases} \tag{29}$$

and the dual function and its derivative on the range $[0, P]$ are given by

$$D(p) = \begin{cases} \frac{L}{2}p^2 - \left(L\alpha - \frac{L}{kP^{k-1}}\alpha^k\right)p + A_2 & \text{if } p \in [0, \alpha]\\ \frac{L}{(k+1)kP^{k-1}}p^{k+1} & \text{if } p \in [\alpha, P].\end{cases}$$

$$D'(p) = \begin{cases} Lp - L\alpha + \frac{L}{kP^{k-1}}\alpha^k & \text{if } p \in [0, \alpha]\\ \frac{L}{kP^{k-1}}p^k & \text{if } p \in [\alpha, P].\end{cases}$$

To prove the lower bound in the lemma, we consider the dual function D in the example with $\alpha = 0$. The optimal dual solution is $p^\star = 0$ and the first k derivatives of $D(\cdot)$ are zero. Therefore, the convergence rate of minimizing D using dual gradient steps becomes increasingly bad as k increases. Moreover, by tuning $\alpha \in [0, P]$, we can ensure that for the associated dual function $D_\alpha(\cdot)$ following holds i) $D_\alpha(p) = D(p)$ for all $p \in [\alpha, P]$ and ii) the primal iterates associated to the dual variables $p < \alpha - (1/L)D'_\alpha(\alpha)$ are infeasible. We now use this intuition to prove the lemma.

Proof: Consider the primal/dual problems $R = (\mathbf{U}, C, \mathbf{m}, \mathbf{M}) \in \mathcal{R}_{\mu,N,P}$ and $D \in \mathcal{D}_{L,P}$ given in Example 2 for some $k \geq 4$ with $\alpha = 0$. The optimal solutions are $\mathbf{x}_R^\star = (C/N, \ldots, C/N)$ and $p_D^\star = 0$. Take any $q = (P, \theta_t(\cdot)) \in \mathcal{Q}^{\text{PF}}_{\mu,N,P}$. The proof now follows from the following 3 steps:

Step 1 Prove the lower bound

$$z(t) \leq \min_{\tau=0,\ldots,t} p_D^q(\tau) \quad \text{for all } t \in N, \tag{30}$$

where $z(t)$ is defined by the recursive relation

$$z(t+1) = z(t) - \frac{1}{kP^{k-1}} z(t)^k, \quad z(0) = P. \tag{31}$$

Step 2 Prove the following lower bound on $z(t)$

$$P \left(\frac{1}{1+2\left(\frac{k-1}{k}\right)(t+1)} \right)^{\frac{1}{k-1}} \leq z(t) \ \text{ for } \ t \in \mathbb{N}_0. \tag{32}$$

Step 3 Use **Steps** 1 and 2 to to prove Equations (23a)-(23c).

Proof of **Step** 1: We prove Equation (30) by induction. For $t = 0$ the the result holds from Proposition 1. Now suppose that the result holds for some $t \in \mathbb{N}_0$. Set $r = \min_{\tau=0,\ldots,t} p_D^q(\tau)$. Consider the dual problem $\bar{D} \in \mathcal{D}_{L,P}$ given in Example 2 with $\alpha = r$. Then $\bar{D}(p) = D(p)$ for all $p \geq r$, so we have $D'(p_D^q(\tau)) = \bar{D}'(p_D^q(\tau))$ for $\tau = 0, \ldots, t$, and hence

$$\theta_\tau(D'(p(0)), \ldots, D'(p(\tau))) = \theta_\tau(\bar{D}'(p(0)), \ldots, \bar{D}'(p(\tau)))$$

and $p_D^q(\tau) = p_{\bar{D}}^q(\tau)$, for all $\tau = 0, \ldots, t$. Since q is primal feasible we have

$$p_D^q(t+1) = p_{\bar{D}}^q(t+1) \geq r - \frac{1}{kP^{k-1}} r^k,$$

because $\bar{D}$ is infeasible for all $p < r - r^k/(kP^{k-1})$. Now by using that the function $g(r) = r - r^k/(kP^{k-1})$ is increasing for $r \leq P$ and by the fact that Equation (30) holds for t, we get

$$p_D^q(t+1) \geq r - \frac{1}{kP^{k-1}} r^k \geq z(t) - \frac{1}{kP^{k-1}} z(t)^k = z(t+1).$$

Proof of **Step** 2: Consider the scaled sequence $\bar{z}(t) = z(t)/P$, which can be expressed by the following recursive relation

$$\bar{z}(t+1) = \bar{z}(t) - \frac{1}{k} \bar{z}(t)^k, \qquad \bar{z}(0) = 1.$$

Then the inequality in Equation (32) is equivalent to

$$v(t+1) \leq \bar{z}(t), \qquad \text{for all } t \in \mathbb{N}_0, \tag{33}$$

where

$$v(t) = \left(\frac{1}{1 + 2\left(\frac{k-1}{k}\right) t} \right)^{\frac{1}{k-1}}. \tag{34}$$

In the sequel, we prove the lemma by proving the inequality in Equation (33) by induction for all $t \in N$. Direct inspection shows that Equation (33) holds true when $t = 0$. In the rest of the proof we suppose that $v(t+1) \leq \bar{z}(t)$ holds for some $t \in \mathbb{N}_0$ and prove that then $v(t+2) \leq \bar{z}(t+1)$ also holds.

To proof the result we need the following two equations

$$\dot{v}(t) = -\frac{2}{k} v(t)^k, \quad v(0) = 1, \tag{35}$$

$$2v(t+1)^k \geq \ v(t)^k, \quad \text{for all } t \geq 1 \text{ and } k \geq 4. \tag{36}$$

Equation (35) can by obtain by differentiating $v(t)$ in Equation (34). Equation (36) is proved in the next paragraph. Using Equation (35) and Equation (36) we get

$$\begin{aligned} v(t+2) &= v(t+1) + \int_{t+1}^{t+2} \dot{v}(\tau) d\tau = v(t+1) - \frac{2}{k} \int_{t+1}^{t+2} v(\tau)^k d\tau \\ &\leq v(t+1) - \frac{2}{k} v(t+2)^k \leq v(t+1) - \frac{1}{k} v(t+1)^k \\ &\leq \bar{z}(t) - \frac{1}{k} \bar{z}(t)^k = \bar{z}(t+1), \end{aligned}$$

where the equality comes from Equation (35), the first inequality comes from the fact that the function $v(\cdot)$ is monotone decreasing, the second inequality comes from Equation (36), and the third inequality comes from the fact that $v(t+1) \leq \bar{z}(t)$ and that the function $g(a) = a - (1/k)a^k$ is monotone increasing on the interval $[0, 1]$. We next finish the proof by proving Equation (36).

Proof of Equation (36): For $k \geq 4$, following holds

$$\frac{1}{2^{\frac{k-1}{k}} - 1} - \frac{1}{2} \frac{k}{k-1} \leq \frac{1}{2^{\frac{3}{4}} - 1} - \frac{1}{2} \leq 1, \tag{37}$$

where the first inequality is obtained by noting that $1/(2^{\frac{k-1}{k}} - 1)$ is decreasing in k and that $k/(k-1) \geq 1$. By using the inequality in Equation (37) together with the fact that $t \geq 1$ we obtain

$$\frac{1}{2^{\frac{k-1}{k}} - 1} \leq \frac{1}{2} \frac{k}{k-1} + t,$$

or by rearranging

$$2^{\frac{k-1}{k}} \geq 1 + \frac{1}{\frac{k}{2(k-1)} + t} = \frac{1 + 2\left(\frac{k-1}{k}\right)(t+1)}{1 + 2\left(\frac{k-1}{k}\right) t},$$

which yields

$$2 \geq \left(\frac{1 + 2\left(\frac{k-1}{k}\right)(t+1)}{1 + 2\left(\frac{k-1}{k}\right) t} \right)^{\frac{k}{k-1}} = \frac{v(t)^k}{v(t+1)^k}.$$

Rearranging the above inequality yields Equation (36).

Proof of **Step** 3: From Equations (30) and (32) we have

$$\text{dist}(p_D^q(t), \mathcal{P}^\star) \geq z(t) \geq P \left(\frac{1}{1+2\left(\frac{k-1}{k}\right)(t+1)} \right)^{\frac{1}{k-1}},$$

which proves Equation (23b). Then we have from Equation (29) and the equality $C = PN/(k\mu)$ that

$$||\mathbf{x}_R^q(t) - \mathbf{x}_R^\star|| = \sqrt{\sum_{i=1}^{N}(x_i(t) - x_i^\star)^2} = \sqrt{N\left(\frac{C}{NP^k}p_D^q(t)^k\right)^2}$$
$$\geq \frac{P\sqrt{N}}{\mu k}\left(\frac{1}{1+2\left(\frac{k-1}{k}\right)(t+1)}\right)^{\frac{k}{k-1}},$$

which proves Equation (23a). Finally, by using that $D(p(0)) - D^\star = P^2L/((k+1)k)$ we get that

$$D(p_D^q(t)) - D^\star \geq \frac{L}{(k+1)kP^{k-1}}p_D^q(t)^{k+1}$$
$$= \frac{P^2L}{(k+1)k}\left(\frac{1}{1+2\left(\frac{k-1}{k}\right)(t+1)}\right)^{\frac{k+1}{k-1}}$$
$$\geq (D(p(0)) - D^\star)\left(\frac{1}{1+2\left(\frac{k-1}{k}\right)(t+1)}\right)^{\frac{k+1}{k-1}},$$

which proves Equation (23c). ∎

G. Proof of Lemma 4

Proof: The fact that q is a primal feasible quantization scheme follows from Theorem 2 in Section IV-A and Equation (39) proved below. We now prove the upper bounds in Equations (25a) and (25b).

a) Proof of Equation (25a): Consider any primal problem $R \in \mathcal{R}_{\mu,N,P}$ and the associated dual problem $D \in \mathcal{D}_{L,P}$. We start by showing that

$$||\mathbf{x}_R^q(t) - \mathbf{x}_R^\star|| \leq \sqrt{\frac{2}{\mu}(D(p_D^q(t)) - D^\star)},$$

by following similar arguments as used in the proof of Theorem 1 in [4]. Define the Lagrangian function as

$$\mathcal{L}(\mathbf{x}, p) = \sum_{i=1}^{N} U_i(x_i) - p\left(\sum_{i=1}^{N} x_i - C\right).$$

The Lagrangian function $\mathcal{L}(\mathbf{x}, p)$ is μ-strongly concave in $\mathbf{x}$, since $U(\cdot)$ is μ-strongly concave and the sum of a μ-strongly concave and a concave function is μ-strongly, see Lemma 2.1.4 in [29]. Therefore, we have

$$D(p_D^q(t)) - D^\star = \mathcal{L}(\mathbf{x}_R^q(t), p_D^q(t)) - U(\mathbf{x}_R^\star)$$
$$\geq \mathcal{L}(\mathbf{x}_R^q(t), p_D^q(t)) - \mathcal{L}(\mathbf{x}_R^\star, p_D^q(t))$$
$$\geq \frac{\mu}{2}||\mathbf{x}_R^q(t) - \mathbf{x}_R^\star||^2,$$

where the equality comes by the strong duality, the first inequality comes by using that $p_D^q(t) \geq 0$ and that the primal optimal solution $\mathbf{x}_R^\star$ is feasible, and the second inequality comes by using that $\mathcal{L}(\mathbf{x}, p)$ is μ-strongly concave in $\mathbf{x}$ and the fact that $\mathbf{x}_R^q(t)$ minimizes the function $\mathcal{L}(\cdot, p_D^q(t))$ on the set $\prod_{i=1}^{N}[m_i, M_i]$ (see Equation (6a)). Now the result follows directly from part b) of the theorem, which we prove now.

b) Proof of Equation (25b): Let $t \in \mathbb{N}$ and the dual problem $D \in \mathcal{D}_{L,P}$ be given. Without loss of generality, we assume that $p_D^q(t) \notin \mathcal{P}_D^\star$, since otherwise $\mathrm{dist}(p_D^q(t), \mathcal{P}_D^\star) = 0$ so then the result trivially holds. The proof follows the following 6 steps:

Step 1 Use the fact that $p_D^q(t) \notin \mathcal{P}_D^\star$ to prove that

$$p_D^q(i+1) = p_D^q(i) - \Delta(i), \quad \text{for } i = 0, \dots, t-1, \tag{38}$$

where $\Delta(i) = \theta_i(D'(p(0)), \dots, D'(p(i))$.

Step 2 Use the fact that $p_D^q(t) \notin \mathcal{P}_D^\star$ to prove that $D'(p_D^q(0)) \leq LP$.

Step 3 Use **Step** 2 to prove that for each $i = 0, \dots, t$, either $\Delta(i) = 0$ or the following inequality holds

$$\frac{1}{2L}D'(p(i)) \leq \Delta(i) \leq \frac{1}{L}D'(p(i)). \tag{39}$$

Step 4 Use **Step** 3 to prove to the inequality

$$D(p_D^q(t)) - D^\star \leq \frac{2LP^2}{2^\tau} \tag{40}$$

where $\tau := |\{i = 0, \dots, t-1 \;|\Delta(i) = 0\}|$.

Step 5 Use **Steps** 1-3 to prove that if $\tau < t$ then the inequality

$$D(p_D^q(t)) - D^\star \leq \frac{8LP^2}{t - \tau} \tag{41}$$

where $\tau := |\{i = 0, \dots, t-1 \;|\Delta(i) = 0\}|$.

Step 6 Use **Step** 4 and **Step** 5 to prove the inequality in Equation (25b), that is

$$D(p_D^q(t)) - D^\star \leq \frac{32LP^2}{t} \quad \text{for all } t \geq 1.$$

Proof of **Step** 1: From Equation (7) we have that

$$p_D^q(i+1) = \lceil p_D^q(i) - \Delta(i) \rceil_+, \quad \text{for } i = 0, \dots, t-1.$$

If $p_D^q(i) - \Delta(i) \geq 0$ for all $i = 0, \dots, t-1$ then Equation (38) trivially holds. Therefore, we suppose that $p_D^q(i) - \Delta(i) < 0$ for some $i \in \mathbb{N}_0$. Then $p_D^q(i+1) = 0$. It follows that $p_D^q(i+1) \in \mathcal{P}^\star$ since $D'(p_D^q(i+1)) \geq 0$ from Lemma 6-b) in Appendix A (see the optimality conditions for constrained convex optimization problems in [30, Proposition 2.1.2.]). Moreover, then $p_D^q(j) \in \mathcal{P}_D^\star$ for all $j \geq i+1$, because $\Delta(j) \geq 0$ for all $j \in \mathbb{N}_0$. In particular, $p_D^q(t) \in \mathcal{P}_D^\star$ which contradicts that $p_D^q(t) \notin \mathcal{P}^\star$.

Proof of **Step** 2: We prove this by contradiction. Suppose that $D'(p_D^q(0)) > LP$, then from Equations (24a) and (24b) we have $p_D^q(1) = \lceil p_D^q(0) - \Delta(0) \rceil_+ = \lceil P - P \rceil_+ = 0 \in \mathcal{P}_D^\star$. Therefore, $p_D^q(1) \in \mathcal{P}_D^\star$ implying that $p_D^q(t) \in \mathcal{P}_D^\star$.

Proof of **Step** 3: We first prove the upper bound in Equation (39). If $\Delta(i) \neq 0$, then from Equations (24b) and (24c) we have that $\Delta(i) = Q_t$ and $(1/L)D'(p_D^q(t)) \geq Q_t = \Delta(t)$.

We now prove by induction over i that either $\Delta(i) = 0$ or the lower bound in Equation (39) holds for i, for all $i = 0, \dots, t$. Consider the case when $i = 0$. If $\Delta(0) \neq 0$, then form Equation (24) we have that $\Delta(0) = P$ and from **Step** 2 that $D'(p_D^q(0)) \leq LP$, so the lower bound in Equation (39) holds. Now suppose that either $\Delta(i) = 0$ or the lower bound in Equation (39) holds for $i+1$. Suppose also that $\Delta(i+1) \neq 0$, since otherwise the result already holds. Consider first the case when $\Delta(i) = 0$. Then from Equations (24b) and (24c), we have that $D'(p_D^q(i)) < Q_iL$, $p_D^q(i+1) = p_D^q(i)$, and $\Delta(i+1) = Q_{i+1} = Q_i/2$. Therefore, we have

$$D'(p_D^q(i+1)) = D'(p_D^q(i)) < Q_iL = 2Q_{i+1}L = 2L\Delta(i+1),$$

or by rearranging $(1/2L)D'(p_D^q(i+1)) < \Delta(i+1)$. Consider next the case when $\Delta(i) \neq 0$. Then from Equations (24b) and (24c), we have that $\Delta(i) = \Delta(i+1)$ and $p_D^q(i+1) = p_D^q(i) - \Delta(i) \leq p_D^q(i)$. Therefore, because of the monotonicity of D' we have

$$\frac{1}{2L}D'(p_D^q(i+1)) \leq \frac{1}{2L}D'(p_D^q(i)) \leq \Delta(i) = \Delta(i+1).$$

Proof of **Step** 4: We first show that

$$D'(p_D^q(t)) \leq \frac{2LP}{2^\tau}. \tag{42}$$

From (24) we have that $Q_t = P/2^\tau$. Therefore, if $\Delta(t) = 0$ then the inequality follows from Equation (24b). Otherwise, if $\Delta(t) \neq 0$ then the inequality follows from Equation (39) in **Step** 3. From Equation (42) and from the convexity of $D(\cdot)$, we have for any $p^\star \in \mathcal{P}_D^\star$ that

$$D(p_D^q(t)) - D^\star \leq D'(p_D^q(t))(p_D^q(t) - p^\star) \leq \frac{2LP^2}{2^\tau}, \tag{43}$$

where we have used that $P \geq p_D^q(t) - p^\star \geq 0$ from Lemma 6-b) in Appendix A.

Proof of **Step** 5: The proof of this step is similar to the proof of Theorem 2.1.14 in [29] that shows the $\mathcal{O}(1/t)$ convergence rate of gradient methods with constant step-size when minimizing convex functions with Lipschitz continuous gradients. However, there are some key differences since the gradient information is quantized here.

Consider now the sequence $(c_i)_{i=1,\dots,t-\tau}$ of the elements in the set $\{j = 0, \dots, t-1 \mid \Delta(j) \neq 0\}$, an increasing order. From **Step** 1 and the fact that $D(\cdot)$ is convex and L-Lipschitz continuous we have [29, Theorem 2.1.5]

$$\begin{aligned} D(p_D^q(c_{i+1})) &\leq D(p_D^q(c_i)) + D'(p_D^q(c_i))(-\Delta(c_i)) + \frac{L}{2}\Delta(c_i)^2 \\ &\leq D(p_D^q(c_i)) - \frac{L}{2}\Delta(c_i)^2, \end{aligned} \tag{44}$$

where the second inequality comes by using Equation (39). Now consider the sequence $\omega_i = D(p_D^q(c_i)) - D^\star$. From Equation (39) and the convexity of $D(\cdot)$ we have

$$\omega_i \leq D'(p_D^q(c_i))(p_D^q(c_i) - p^\star) \leq 2LP\Delta(c_i),$$

where $p^\star \in \mathcal{P}_D^\star$. This together with Equation (44) yields

$$\omega_{i+1} \leq \omega_i - \frac{1}{8P^2L}\omega_i^2, \quad \text{for } i = 1, \dots, t-\tau-1,$$

and by multiplying $1/(\omega_i\omega_{i+1})$ on both sides and rearranging we have

$$\frac{1}{\omega_{i+1}} \geq \frac{1}{\omega_i} + \frac{1}{8P^2L}\frac{\omega_i}{\omega_{i+1}} \geq \frac{1}{\omega_i} + \frac{1}{8P^2L},$$

where we have used that $\omega_{i+1} \geq \omega_i$. Then by summing over i we get

$$\frac{1}{\omega_{i+1}} \geq \frac{1}{\omega_0} + \frac{1}{8P^2L}(i+1),$$

or by rearranging

$$\omega_{i+1} \leq \frac{8P^2L(D(p_D^q(0)) - D^\star)}{8P^2L + (D(p_D^q(0)) - D^\star)(i+1)} \leq \frac{8P^2L}{i+1}.$$

Now the inequality in Equation (41) follows by setting $i = t - \tau - 1$ in the inequality above and by the inequality in Equation (43).

Proof of **Step** 6: If $\tau > t/2$ then we have from **Step** 4 that

$$D(p_D^q(t)) - D^\star \leq \frac{2P^2L}{2^\tau} \leq \frac{P^2L}{2^{t/2}} \leq \frac{32P^2L}{t},$$

where the last inequality comes from that $16 \times 2^{t/2} \geq t$ for $t \in \mathbb{N}_0$, which follows from that $h(t) := 16 \times 2^{t/2} - t \geq 0$ for $t \in \mathbb{R}_+$ since $h(0) = 16$ and $h'(t) > 0$ on $\mathbb{R}_+$. If $\tau \leq t/2$ then $t - \tau \geq t/2$ and the result holds from **Step** 5. ■

References

[1] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research society*, vol. 49, no. 3, pp. 237–252, 1998.

[2] S. Low and D. Lapsley, "Optimization flow control. i. basic algorithm and convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, pp. 861–874, Dec 1999.

[3] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, Aug 2006.

[4] A. Beck, A. Nedic, A. Ozdaglar, and M. Teboulle, "An O(1/k) gradient method for network resource allocation problems," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 64–73, March 2014.

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

[6] E. Biglieri, "Coding and modulation for a horrible channel," *IEEE Communications Magazine*, vol. 41, no. 5, pp. 92–98, 2003.

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

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

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

[10] L. Chen, N. Li, S. H. Low, and J. C. Doyle, "Two market models for demand response in power networks," *IEEE SmartGridComm*, vol. 10, pp. 397–402, 2010.

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

[12] C. Enyioha, S. Magnússon, K. Heal, N. Li, C. Fischione, and V. Tarokh, "Robustness analysis for an online decentralized descent power allocation algorithm," in *IEEE Information Theory and Applications Workshop (ITA)*, February 2016.

[13] S. Magnusson, C. Enyioha, K. Heal, N. Li, C. Fischione, and V. Tarokh, "Distributed resource allocation using one-way communication with applications to power networks," in *IEEE Annual Conference on Information Science and Systems (CISS)*, March 2016, pp. 631–636.

[14] A. C.-C. Yao, "Some complexity questions related to distributive computing (preliminary report)," in *Proceedings of the eleventh annual ACM symposium on Theory of computing*. ACM, 1979, pp. 209–213.

[15] E. Kushilevitz and N. Nisan, *Communication complexity*. Cambridge University Press, 1997.

[16] S. Hart and Y. Mansour, "How long to equilibrium? the communication complexity of uncoupled equilibrium procedures," *Games and Economic Behavior*, vol. 69, no. 1, pp. 107–126, 2010.

[17] V. Conitzer and T. Sandholm, "Communication complexity as a lower bound for learning in games," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 24.

[18] M.-F. Balcan, A. Blum, S. Fine, and Y. Mansour, "Distributed learning, communication complexity and privacy." in *COLT*, vol. 23, 2012, pp. 26–1.

[19] Y. Zhang, J. Duchi, M. I. Jordan, and M. J. Wainwright, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints," in *Advances in Neural Information Processing Systems*, 2013, pp. 2328–2336.

[20] J. N. Tsitsiklis and Z.-Q. Luo, "Communication complexity of convex optimization," *Journal of Complexity*, vol. 3, no. 3, pp. 231–243, 1987.

[21] M. Rabbat and R. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798–808, April 2005.

[22] A. Nedic, A. Olshevsky, A. Ozdaglar, and J. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, Dec. 2008, pp. 4177–4184.

[23] P. Yi and Y. Hong, "Quantized subgradient algorithm and data-rate analysis for distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 380–392, Dec. 2014.

[24] Y. Pu, M. N. Zeilinger, and C. N. Jones, "Quantization design for unconstrained distributed optimization," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 1229–1234.

[25] S. Magnússon, K. Heal, C. Enyioha, N. Li, C. Fischione, and V. Tarokh, "Convergence of limited communications gradient methods," in *IEEE American Control Conference (ACC)*, June 2016, pp. 1421–1426.

[26] E. Nekouei, T. Alpcan, G. Nair, and R. Evans, "Convergence analysis of quantized primal-dual algorithms in network utility maximization problems," *IEEE Transactions on Control of Network Systems*.

[27] S. Magnusson, C. Enyioha, N. Li, and C. Fischione, "Practical coding schemes for bandwidth limited one-way communication resource allocation," in *IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 221–226.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[29] Y. Nesterov, *Introductory Lectures on Convex Optimization*. Springer, 2004.

[30] D. P. Bertsekas, *Nonlinear Programming: 2nd Edition*. Athena Scientific, 1999.

[31] Y. W. Chen, X. Chen, and N. Maxemchuk, "The fair allocation of power to air conditioners on a smart grid," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 2188–2195, Dec 2012.

[32] Z. Fan, "A distributed demand response algorithm and its application to phev charging in smart grids," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1280–1290, Sept 2012.

[33] J. J. Escudero-Garzas, A. Garcia-Armada, and G. Seco-Granados, "Fair design of plug-in electric vehicles aggregator for v2g regulation," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3406–3419, Oct 2012.

[34] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Basar, "Dependable demand response management in the smart grid: A stackelberg game approach," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 120–132, March 2013.