Practical Coding Schemes For Bandwidth Limited One-Way Communication Resource Allocation

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

Abstract— This paper investigates resource allocation algorithms that use limited communication – where the supplier of a resource broadcasts a coordinating signal using *one bit* of information to users per iteration. Rather than relay anticipated consumption to the supplier, the users locally compute their allocation, while the supplier measures the total resource consumption. Since the users do not compare their local consumption against the supplier's capacity at each iteration, they can easily overload the system and cause an outage (for example blackout in power networks). To address this challenge, this paper investigates pragmatic coding schemes, called PF-codes (Primal-Feasible codes), that not only allow the restriction of communication to a single bit of information, but also avoid system overload due to users' heavy consumption. We derive a worst case lower bound on the number of bits needed to achieve any desired accuracy using PF-codes. In addition, we demonstrate how to construct time-invariant and time-varying PF-codes. We provide an upper bound on the number of bits needed to achieve any desired solution accuracy using time-invariant PF-codes. Remarkably, the difference between the upper and lower bound is only 2 bits. It is proved that the time-varying PF-codes asymptotically converge to the true primal/dual optimal solution. Simulations demonstrating accuracy of our theoretical analyses are presented.

I. INTRODUCTION

A fundamental task in most networked systems is to allocate shared resources between the network entities, e.g., allocate power in electrical distribution and data rates in communication networks. Most real-world networks are now growing at unprecedented rates, in part, due to advances in smart sensor/actuator technologies. It is, thus, becoming more important that the protocols used to allocate the resources appropriately scale with the growing network size. In particular, the communication architecture used to coordinate these algorithms must be managed efficiently to reduce excessive bandwidth consumption [1]. Motivated by these challenges, we investigate algorithms for resource allocation amongst spatially distributed network entities that operate using limited communication overhead.

Resource allocation optimization problems and related decomposition optimization algorithms have had much success in efficient operation of communication networks [2]–[5]. More recent work has also demonstrated the potential of using similar ideas to coordinate and economically operate smart power grids [6]–[9]. However, distributed optimization algorithms that use limited communication have not received much attention, despite their significant benefits, especially given bandwidth constraints on most communication networks. Some interesting work [10]–[13] on distributed optimization using limited bandwidth can be found in the literature. The problems addressed in [10]–[13] consider bandwidth-limited coordination of primal problem iterates, in the setting of consensus/incremental subgradient methods and cellular power control, where no dual variables are introduced. However, many interesting problems have coupled constraints and are therefore more naturally decomposed using duality theory [2]–[9]. This motivates us to consider dual decomposition where the dual gradients are communicated using limited bandwidth. This gives new analytical challenges compared to the earlier work.

A. Contributions of This Work

This paper investigates resource allocation algorithms based on dual decomposition using one-way communication from supplier to users. By one-way communication, we mean that the supplier iteratively broadcasts a coordination signal (dual variable) to the users, which the users use to locally compute their optimal allocations, without having relayed their anticipated demands to the suppliers. Instead, the users consume an amount of the resources that depends on the coordination signal. Then the supplier measures the total consumption before updating the coordination signal. Such physical measurements are easily achievable in many real-world networks, e.g., in power networks [14].

Unlike our previous work [9], where a real valued coordination signal (thus requiring infinite bits) is iteratively broadcast, in this paper only *one bit* is broadcast per iteration. The generalization to more than one bit per iteration is given in [9]. The main contribution of this paper is to investigate and design practical codes, or quantizers that allow only one bit of information per iteration to be broadcast, while guaranteeing feasibility of the primal iterates at every iteration. If primal feasibility is not ensured, users may consume more resources than are available and overload the system. This is clearly unacceptable because it causes blackouts in power networks, or outages in wireless networks. We investigate the main properties of such primal feasible (PF) codes. We provide a lower bound on the performance of PF-codes, i.e., a worst case lower bound on the bits needed to achieve any ϵ-solution accuracy. We demonstrate how to construct a) time-invariant and b) time-varying PF-codes. For time invariant PF-codes, we provide an upper bound on the number bits

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 and N. Li are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA USA. (email: cenyioha@seas.harvard.edu; nali@seas.harvard.edu)

needed to achieve any ϵ-solution accuracy. The difference between the upper and lower bound is 2 bits. We show that the time varying PF-codes do asymptotically converge to the true primal/dual solution. Unlike our earlier work [8], [9] the current paper studies bandwidth limited communication. Compared to [15], which considered how to quantize *only* the direction of a high dimensional gradient, this paper investigates how to quantize the magnitude of a scalar dual derivative while simultaneously ensuring primal feasibility of the iterates.

B. Notation and Definitions

The set of real, positive real, and natural numbers are denoted by $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{N}$. $|\mathcal{A}|$ 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. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is, respectively, L-smooth and μ-concave on $\mathcal{X} \subseteq \mathbb{R}^n$ if ∇f is L-Lipschitz continuous on $\mathcal{X}$ and if $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle - \frac{\mu}{2} ||\mathbf{y} - \mathbf{x}||^2$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{X}$.

II. PROBLEM FORMULATION AND RELATED BACKGROUND

In this section we introduce the problem background. Subsection II-A presents the resource allocation problem studied in the paper and how they can be solved using dual decomposition. Subsection II-B demonstrates how the bandwidth limited one-way communication is used, via dual decomposition, to solve the resource allocation problem.

A. Resource Allocation and Dual Decomposition

Consider a network with N users $\mathcal{N} = \{1, \cdots, N\}$ 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 [2]

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

where $U_i : \mathbb{R} \to \mathbb{R}$ is a utility function and m_i and M_i are the lower and upper bounds on the demand of user $i \in \mathcal{N}$. We make the following standard assumptions on (1):

Assumption 1. *(Convex)* $U_i(x_i)$ *is* μ-*concave on* $[m_i, M_i]$.

Assumption 2. *(Well Posed) We have that* $\sum_{i=1}^N m_i \leq C \leq \sum_{i=1}^N M_i$ *In other words, Problem* (1) *is feasible and the constraint* $\sum_{i=1}^N x_i \leq C$ *is not redundant.*

This paper considers distributed algorithms for solving (1) based on duality theory [2], [3]. In particular, the solution of (1) is obtained by solving the dual problem:

$$
\begin{aligned}
\underset{p}{\text{minimize}} \quad & D(p), \\
\text{subject to} \quad & p \geq 0,
\end{aligned}
\tag{2}
$$

where D and p are the dual function and dual variables [16, Chapter 5], respectively. D is given by

$$
\begin{aligned}
D(p) &= \underset{\mathbf{x} \in \prod_{i=1}^N [m_i, M_i]}{\text{maximize}} \sum_{i=1}^N U_i(x_i) - p\left(\sum_{i=1}^N x_i - C\right) \\
&= \sum_{i=1}^N U_i(x_i(p)) - p\left(\sum_{i=1}^N x_i(p) - C\right)
\end{aligned}
\tag{3}
$$

where $\mathbf{x} = (x_1, \ldots, x_N)^T$ and

$$
x_i(p) = \underset{x_i \in [m_i, M_i]}{\text{argmax}} \, U_i(x_i) - p \, x_i = \left[(U_i')^{-1}(p)\right]_{m_i}^{M_i}.
\tag{4}
$$

The dual derivative is

$$
D'(p) = C - \sum_{i=1}^N x_i(p).
\tag{5}
$$

The following result establishes a relationship between (1) and (2) [9, Lemma 1]:

Lemma 1. *(Strong Duality) Suppose Assumptions 1 and 2 hold. If* $p^\star$ *is an optimal solution of* (2)*, then* $\mathbf{x}(p^\star) = [x_i(p^\star)]_{i \in \mathcal{N}}$ *[cf. (4)] is the optimal solution to* (1)*.*

The following result is proved in the Appendix:

Proposition 1. *Suppose Assumption 1 and 2 hold, then i)* $D(\cdot)$ *is differentiable and* N/μ-*smooth and ii)* (2) *has an optimal solution in the interval* $[0, P]$ *where* $P = \max_{i \in \mathcal{N}} U_i'(m_i)$.

Motivated by Proposition 1, the theory developed in this paper considers dual problems that come from the following class of convex optimization problems:

Definition 1. *Let* $\mathcal{D}_{P,L}$ *denote the set of all (dual) optimization problems of the form*

$$
\begin{aligned}
\underset{p}{\text{minimize}} \quad & D(p), \\
\text{subject to} \quad & p \in [0, P],
\end{aligned}
\tag{6}
$$

where $D : \mathbb{R}_+ \to \mathbb{R}$ *is convex and* L-*smooth and at least one optimal solution* $p^\star$ *to* (6) *is also an optimal solution to the relaxed problem, i.e., if* $p^\star = P$ *then* $D'(p^\star) = 0$.

Due to Proposition 1, all dual problems of Problem (1), where Assumptions 1 and 2 hold are in the class $\mathcal{D}_{P,L}$ with $L = N/\mu$ and $P = \max_{i \in \mathcal{N}} U_i'(m_i)$. To ensure that the suppliers know to which class $\mathcal{D}_{P,L}$ their dual problems belongs, we make the following mild assumption.

Assumption 3. *There exists a* P *with* $P \geq U_i'(m_i) \, \forall \, i \in \mathcal{N}$.

The primal problem (1) can be solved in a distributed way via dual decomposition. Then, instead of directly solving the coupled primal problem, the dual problem is solved using the gradient descent method. Specifically, the supplier updates the dual variable $p(t)$ using the gradient descent iterate

$$
p(t+1) = \lceil p(t) - \gamma D'(p(t)) \rceil^+,
\tag{7}
$$

where $\gamma > 0$ is the step-size and the dual derivative $D'(p(t))$ [cf. (5)] can be computed in a distributed fashion from (4).
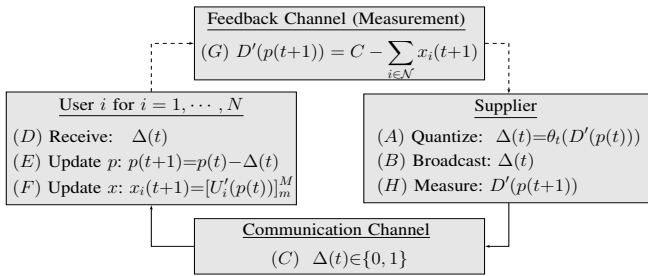
Fig. 1: The OneWay-DD Algorithm.

The following result on the convergence of the recursion (7) is standard in the literature [2, Theorem 1].

Proposition 2. *If $\gamma \in (0, 2\mu/N)$ in the recursions given in Equation* (7)*, then every limit point of $p(t)$ is a solution to* (2) *and $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^\star$ where $\mathbf{x}^\star$ is the solution to* (1)*.*

The dual gradient (5) is simply the difference between supply and demand, which can be measured by the supplier, e.g., in power networks [14]. Based on this observation, we suggest a one-way communication model for solving (1) that is well suited for real-world power networks.

B. Communication Model: One-Way Communication

We consider bandwidth limited one-way communication protocols for solving (1) that use:

- **One-Way Communication**: The suppliers communicate with users via a noiseless, binary channel with bandwidth of 1 bit per iteration.
- **Feedback Information**: After broadcasting the dual variable $p(t)$ to the users, the supplier can measure deviation between total power consumption and supply, i.e., the dual gradient $D'(p(t)) = C - \sum_{i=1}^{N} x_i(t)$.

For the one-way communication, a channel is needed between the supplier and users. We consider minimal bandwidth where 1 bit is communicated per iteration. Therefore, the results are useful for "horrible" communication channels [17] with very limited bandwidth, e.g., power line communication. The feedback information obtained from measuring $D'(p(t))$, that is, the difference between supply and demand can be obtained in many realistic engineering scenarios, e.g., in power networks [14].

Figure 1 depicts the one-way communication model and steps of the dual decomposition algorithm studied in the rest of the paper. At each iteration t of the algorithm, the supplier broadcasts one binary signal indicating the change of the dual variable, i.e., quantized dual gradient information $\theta_t(D'(p(t)))$, where $\theta_t : \mathbb{R} \to \mathcal{R}_t$, and $\mathcal{R}_t \subseteq \mathbb{R}$ with cardinality $|\mathcal{R}_t| = 2$. Then each user $i \in \mathcal{N}$ can update its local copy of the dual variable $p_i(t+1)$ as in Equation (7).[1] In particular, the dual variable is updated as

$$p(t+1) = \lceil p(t) - \theta_t(D'(p(t))) \rceil^+. \tag{8}$$

---

[1]In the rest of the paper we use the notation $p(t)$ instead of $p_i(t)$ since $p_i(t) = p_j(t)$ for all $i \in \mathcal{N}$ and $t \in \mathbb{N}$.

The users update their usage based on the dual variable $p(t+1)$ following Equation (4). Then the supplier can measure the total consumption, i.e., the dual gradient (5). The steps of the algorithm (A)-(G) can be described as follows:

ONEWAY-DD: One way communication dual decomposition

(A) The supplier quantizes $D'(p(t))$: $\Delta(t) = \theta_t(D'(p(t)))$.
(B) The supplier broadcasts $\Delta(t)$ to all users.
(C) $\Delta(t)$ is carried over the binary channel.
(D) Each user receives the message $\Delta(t)$.
(E) The users locally update the dual variable $p(t)$ using (8).
(F) The users update their power demands using (4).
(G) The dual gradient $D'(p(t))$ [cf. (5)] is available.
(H) The supplier measures $D'(p(t))$ [cf. (5)].

We consider different codes for the ONEWAY-DD Algorithm, where codes are formally defined as follows:

Definition 2. *We call a pair $(p(0), \theta_t(\cdot))$ a code for the* ONEWAY-DD *Algorithm if $p(0) \in [0, P]$ and $\theta_t : \mathbb{R} \to \mathcal{R}_t$, with $|\mathcal{R}_t| = 2$.*

Due to the one-way communication, the primal iterates $x_i(t)$ are not communicated to the supplier. Instead the users locally compute their allocations $x_i(t)$ and their aggregate allocation is measured by the supplier. Therefore, it is essential that the primal problem (1) is feasible during every iteration of the algorithm, i.e. $\sum_{i=1}^{N} x_i(t) \leq C$ for all $t \in \mathbb{N}$. Otherwise, a heightened demand of power in the network can result in a system overload, e.g., blackouts in power systems.

In what follows, we study the convergence of the ONEWAY-DD Algorithm and how to ensure that the primal problem is feasible during every iteration.

III. HOW TO AVOID BLACKOUTS? PRIMAL FEASIBLE CODES

In this section, we investigate how to ensure transient primal feasibility of the ONEWAY-DD Algorithm. In Subsection III-A, we introduce codes for the ONEWAY-DD Algorithm that ensure primal feasibility at every iteration and for all considered resource allocation problems – so called *Primal Feasible* (PF) codes. In Subsection III-B, we provide worst case lower bounds on the performance of the PF-codes in terms of number of bits needed to achieve given accuracy.

The following result, proved in the Appendix, is used throughout the rest of the paper.

Lemma 2. *Consider the primal and dual problems* (1) *and* (2)*. Let $\mathcal{P}^\star = [\underline{p}^\star, \bar{p}^\star]$ be the set of dual optimizers. Then the following three condition are equivalent: (i) $(x_i(p))_{i\in\mathcal{N}}$ is primal feasible, (ii) $D'(p) \geq 0$, and (iii) $p \geq \underline{p}^\star$.*

A. Primal Feasible (PF) Codes

We are interested in codes that ensure primal feasibility of Problem (1), where Assumptions 1, 2, and 3 hold. We formally define such codes as follows.

Definition 3 (Primal Feasible (PF) Codes)**.** *We say that a code $(p(0), \theta_t(\cdot))$ is* Primal Feasible (PF) *if for every resource allocation problem* (1) *where Assumptions 1, 2, and 3*

hold, the iterates $[x_i(p(t))]_{i\in\mathcal{N}}$ are feasible for Problem (1), for $t = 1, 2\cdots$, and $p(t+1) = p(t) - \theta_t(p(t))$. We denote the set of all PF-codes by $\mathcal{C}_{PF}$.

PF-codes are practically desirable since they ensure that the users $i \in \mathcal{N}$ do not overuse the resources as the algorithm runs. However, we wish to arrive at a design method for such codes. The following result demonstrates a key property used later to design PF-codes for the ONEWAY-DD Algorithm.

Theorem 1. $(p(0), \theta_t(\cdot))$ *is a PF-code if* $p(0) = P$ *and* $\theta_t(z) \leq (1/L)z$ *for all* $z \geq 0$.

Proof. We first show that $[x_i(P)]_{i\in\mathcal{N}}$ is feasible to all resource allocation problems (1) where Assumptions 1 and 2 hold. By Assumptions 2 and 3, for any dual function $D(\cdot) \in \mathcal{D}_{P,L}$ there is $p^\star \in [0, P]$ with $D'(p^\star) = 0$. In particular, $P \geq p^\star$ and hence by Lemma 2-iii) $[x_i(P)]_{i\in\mathcal{N}}$ is feasible.

We next show that if $[x_i(p(t))]_{i\in\mathcal{N}}$ is feasible for Problem (1) and $\theta_t(z) \leq (1/L)z$, $z \geq 0$, then $[x_i(p(t+1))]_{i\in\mathcal{N}}$ is also feasible for Problem (1), where $p(t+1) = p(t) - \theta_t(D'(p(t)))$. From Lemma 2-ii) we have for all $D\in\mathcal{D}_{P,L}$ that $[x_i(p)]_{i\in\mathcal{N}}$ is feasible for $p\in[0, P]$ if and only if $D'(p)\geq0$. Hence, $D'(p(t))\geq0$ and by the L-Lipschitz continuity of D' we have that $D'(p(t)) - D'(p(t+1)) \leq L\theta_t(D'(p(t)))\leq D'(p(t))$, or by rearranging $0\leq D'(p(t+1))$. Therefore, $[x_i(p(t+1))]_{i\in\mathcal{N}}$ is feasible for Problem (1) by Lemma 2-ii). $\square$

The following example demonstrates that the conditions that $p(0) = P$ and $\theta_t(z) \leq (1/L)z$, for $z \geq 0$, are generally also necessary for $(p(0), \theta_t)$ to be a PF-code.

Example 1. *Consider Problem* (1)*, where the utility function of user* $i\in\mathcal{N}$ *is* $U_i(x_i)=-\frac{N}{2L}(x_i-M)^2$*, where* $m_i=m$*,* $M_i = M$*, and* $M-m=PL/N$*. Direct inspection shows that* $U'(m) \leq P$*, Assumption 3 holds, and* $x_i(p) = [U_i']^{-1}(p) = M - (L/N)p$ *[cf. (4)] for* $p \in [0, P]$ *and* $x_i(p) = m_i$ *for* $p > P$*. The dual function is* $D(p) = \frac{L}{2}p^2 + (C - MN)p$*, for* $p \in [0, P]$*.* $D'(p) = Lp + (C - MN)$ *is L-Lipschitz continuous and* $D'(p) = 0$ *when* $p^\star = (MN - C)/L$*. Therefore, given* $C \in [Nm, NM]$ *the optimal primal/dual solutions are* $x_i^\star(C) = C/N$ *and* $p^\star(C) = (MN - C)/L$*.*

From Example 1, we see that $p(0) = P$ is necessary for $(p(0), \theta_t)$ to be a PF-code. Since for any $p \in [0, P)$, $[x_i(p)]_{i\in\mathcal{N}}$ is infeasible to the the primal problem given in Example 1 with $C = MN-L(p+P)/2$. This fact is established by Lemma 2 iii) and $p^\star = (p+P)/2$ is then the unique optimal solution to the dual problem. Similarly, if $p(t)-p(t+1) = \theta_t(D'(p(t)) > (1/L)D'(p(t))$, then it possible to obtain iterates such that $[x_i(p(t))]_{i\in\mathcal{N}}$ is feasible for Problem (1) but $[x_i(p(t+1))]_{i\in\mathcal{N}}$ is not.

We next provide a lower bound on the number of bits needed to achieve ϵ-accuracy using the primal feasible ONEWAY-DD Algorithm.

B. Lower Bounds on Primal Feasible-Codes

The following result provides a lower bound on the number of bits needed to ensure that the dual problem is solved up to any given ϵ-accuracy.

Theorem 2 (Lower Bound)**.** *Consider the* ONEWAY-DD *Algorithm and let* $\epsilon > 0$ *be given. Then for every PF-code* $(p(0), \theta_t)$ *there exists* $D \in \mathcal{D}_{P,L}$ *such that at least*

$$B_{PF}(\epsilon) := \left\lceil \frac{PL}{\epsilon} \right\rceil - 2 \text{ bits} \qquad (9)$$

are needed to find $p \in [0, P]$ *for which* $|D'(p(T))| \leq \epsilon$*. In other words, for any PF-code* $(p(0), \theta_t)$*,* $B_{PF}(\epsilon)$ *is a lower bound on the following minimax problem:*

$$\begin{aligned} \min_{T\in\mathbb{N}} \max_{D\in\mathcal{D}_{P,L}} \quad & T \\ \text{subject to} \quad & |D'(p(T))| \leq \epsilon, \\ & (p(0), \theta_t) \in \mathcal{C}_{PF}. \end{aligned} \qquad (10)$$

Proof. We can assume without loss of generality that $p(t+1) \leq p(t)$ for all $t \in \mathbb{N}$, i.e., we only consider PF-codes with $\theta_t(z) \geq 0$ for $z \geq 0$. Otherwise, if $p(t+1) > p(t)$ for the PF-code $(p(0), \theta_t(\cdot))$ and some t, then we can choose another PF-code $(p(0), \hat{\theta}_t(\cdot))$ that reaches $p(t+1)$ in fewer iterations. Moreover, $p(0) = P$ is a necessary condition for $(p(0), \theta_t(\cdot))$ to be a PF-code, as outlined in the discussion following Example 1; PF-codes have the form $(P, \theta_t(\cdot))$.

To prove the result, we consider a primal problem with a dual problem $D(\cdot) \in \mathcal{D}_{P,L}$ that is difficult for all PF-codes. For any $\delta > 0$, consider Example 2 (defined below) with $\alpha = (\epsilon+\delta)/L$ and $C = MN$. Then the dual function and dual derivative can be obtained by inserting $\alpha = (\epsilon+\delta)/L$ and $C = MN$ into Equations (11) and (12).

We start by showing that for all PF-codes $(P, \theta_t(\cdot))$ and $D(\cdot)$ it holds that $\theta_t(D'(p(t))) \leq (\epsilon + \delta)/L$ for all $t \in \mathbb{N}$. In particular, we show that if $\theta_{t_0}(\epsilon + \delta) > (\epsilon + \delta)/L$ for some $t_0 \in \mathbb{N}$ then $(P, \theta_t(\cdot))$ is not a PF-code by constructing a primal problem with dual problem $\hat{D}(\cdot) \in \mathcal{D}_{P,L}$ such that $[x_i(p(t_0))]_{i\in\mathcal{N}}$ is feasible and $[x_i(p(t_0+1))]_{i\in\mathcal{N}}$ is infeasible, where $p(t_0+1) = p(t_0) - \theta_t(\hat{D}(p(t_0)))$. Consider Example 2 (defined below) with $\alpha = p(t_0)$ and $C = MN+\epsilon+\delta-Lp(t_0)$. Then $\hat{D}'(p) = \epsilon+\delta$ for all $p \geq p(t_0)$, $p(t) \in [p(t_0), P]$ for $t \leq t_0$ (since $p(t+1) \leq p(t)$), and $\hat{D}'(p^\star) = 0$ if and only if $p^\star = p(t) - (\epsilon+\delta)/L$. As a result, $p(t_0+1) = p(t_0) - \theta_t(\hat{D}(p(t_0))) < p(t) - (\epsilon + \delta)/L = p^\star$. Hence, $[x_i(p(t_0+1))]_{i\in\mathcal{N}}$ is infeasible by Lemma 2-iii).

From above, the iterations $p(t)=p(t-1)-\theta_t(D'(p(t-1)))$ can be lower bounded by $p(t) \geq P - t(\epsilon + \delta)/L$ for $t \leq \lfloor LP/(\epsilon + \delta)\rfloor - 1$. In particular, $D'(p(t)) = \epsilon+\delta > \epsilon$ for all $t \leq \lfloor LP/(\epsilon+\delta)\rfloor - 1$. Then the result follows by noting that for $\delta > 0$ sufficiently small, $\lfloor LP/(\epsilon+\delta)\rfloor \leq \lceil LP/\epsilon \rceil -1$. $\square$

The following example is used in the proof of Theorem 2 to provide a problem that is difficult for all PF-codes.

Example 2. *Let* $\alpha > 0$ *and* C *be given and consider Problem* (1)*, where the utility of user* $i \in \mathcal{N}$ *is* $U_i(x_i) = -\frac{N}{2L}(x_i-M)^2$*,* $m_i = m$*,* $M_i = M$*, and* $m = M-(L/N)\alpha$*. It can be checked that* $U'(m) \leq P$*, Assumption 3 holds, and* $x_i(p)=M-(L/N)p$ *if* $p\in[0,\alpha]$ *and* $x_i(p)=m$ *if* $p\in[\alpha, P]$*. The dual function and dual derivative are*

$$D(p) = \begin{cases} Lp^2/2 + (C - NM)p & \text{if } p\in[0,\alpha] \\ L\alpha(p - \alpha/2)+(C-NM)p & \text{if } p\in[\alpha, P], \end{cases} \qquad (11)$$

$$D'(p) = \begin{cases} L\ p + (C - NM) & \text{if } p \in [0, \alpha] \\ L\alpha + (C - NM) & \text{if } p \in [\alpha, P]. \end{cases} \quad (12)$$

Theorem 2 shows that for any $\epsilon > 0$ and PF-code $(P, \theta_t(\cdot))$ there exists a primal problem and an associated dual problem $D(\cdot) \in \mathcal{D}_{P,L}$ such that at least $B_{PF}(\epsilon)$ bits are needed to satisfy $D'(p(t)) \leq \epsilon$. In the next section, we demonstrate how to construct practical PF-codes, including codes that almost achieve that lower bound $B_{PF}(\epsilon)$ in Equation (9).

IV. PRACTICAL PRIMAL FEASIBLE (PF) CODES

This section provides two practical PF-codes. In Subsections IV-A and IV-B we respectively introduce time-invariant and time-varying PF-codes, and study their properties.

A. Coding scheme 1: Time-Invariant PF-Codes

We now introduce a class of PF-codes that almost reach the lower bound in Theorem 2. The difference between the upper and lower bound is two bits.

Definition 4 (Binary Time-Invariant Codes). (P, θ_t) *is a binary time-invariant code if* $\theta_t = \theta : \mathbb{R} \to \mathcal{R}$, $t \in \mathbb{N}$ *and* $|\mathcal{R}| = 2$.

We consider binary time-invariant codes $(P, \theta_\gamma^\kappa)$ generated by the following class of quantizers

$$\theta_\gamma^\kappa(z) = \begin{cases} \gamma & \text{if } z \geq \kappa \\ 0 & \text{otherwise,} \end{cases} \quad (13)$$

$\gamma, \kappa > 0$. The quantizer (13) ensures that at every iteration of the ONEWAY-DD Algorithm the dual variable sequence $p(t)$ is non-increasing. The following result demonstrates how to choose γ and κ such that i) $(P, \theta_\gamma^\kappa(\cdot))$ is a PF-code and ii) $D'(\cdot)$ can be made arbitrarily small.

Theorem 3. *Consider the quantizer* $\theta_\gamma^\kappa(\cdot)$ *[cf. (13)], then:*
i) $(P, \theta_\gamma^\kappa(\cdot))$ *is a PF-code if* $\gamma > 0$ *and* $\kappa \geq \gamma L$.
ii) For any $\epsilon > 0$, *if* $\gamma = \epsilon/L$ *and* $\kappa = \epsilon$ *then* $(P, \theta_\gamma^\kappa(\cdot))$ *is a PF-code and* $|D'(p(t))| \leq \epsilon$ *for all* $t \geq T$ *where* $T \leq \lceil PL/\epsilon \rceil$.

Proof. i) If $\gamma > 0$ and $\kappa \geq \gamma L$ then $\theta_\gamma^\kappa(z) \leq (1/L)z$ for all $z \geq 0$. Therefore, the result follows directly from Theorem 1.

ii) The fact that $(P, \theta_\gamma^\kappa(\cdot))$ is a PF-code follows from i); hence, $D'(p(t)) \geq 0$, for $t \in \mathbb{N}$ [see Lemma 2-ii)]. Moreover, we have $D'(p(t+1)) \leq D'(p(t))$, for all $t \in \mathbb{N}$, since $p(t)$ is decreasing, due to $\gamma > 0$, and $D'(\cdot)$ is monotone, due to the convexity of $D(\cdot)$. Therefore, $p(t) = \lceil P - t\gamma \rceil^+$ for $t = 1, \cdots, T$ and $p(t) = p(T)$ for $t > T$, where $T = \min\{t \in \mathbb{N} | \epsilon \leq |D'(P - t\gamma)|\}$. It holds that $0 \leq D'(p(T)) \leq \kappa = \epsilon$ since $D'(p^\star) = 0$ for some $p^\star \in [0, P]$, Assumption 2, and $D'(p(t)) \geq 0$ for all $t \in \mathbb{N}$, $(P, \theta_\gamma^\kappa(\cdot))$ is a PF-code. Hence, $T \leq \lceil P/\gamma \rceil = \lceil PL/\epsilon \rceil$ as $\lceil P - t\gamma \rceil^+ = 0$ for $t > P/\gamma$. $\square$

The PF-code $(P, \theta_\gamma^\kappa(\cdot))$ possesses the practical advantage that for all considered primal/dual problems, at most $\lceil PL/\epsilon \rceil$ bits/iterations are needed to reach any ϵ-accuracy, even on the most extreme problems; that is, only 2 bits more than the lower bound provided in Theorem 2. However, the bound obtained in Theorem 2 is a lower bound on the number of bits needed to solve the most difficult problem in the class $\mathcal{D}_{P,L}$. In particular, it was obtained by considering the

extreme Example 2, where for given $\epsilon > 0$, the dual gradient was $D'(p) = \epsilon + \delta$ for $p \in [\alpha, P]$, with $\delta, \alpha > 0$ small. Practical problems are likely to be less stringent and the dual derivative $D'(\cdot)$ is generally not a constant $\epsilon + \delta$ on $[\alpha, P]$, for the chosen $\epsilon > 0$. Most problems are likely to have a dual gradient that changes more over $[0, P]$. Therefore, it is desirable to consider time-varying PF-codes.

B. Coding scheme 2: Time-Varying PF-Codes

We now consider time varying PF-codes, which can have faster convergence for many moderate resource allocation problems, though they may not perform as well on extreme worst-case problems. For example, in the initial steps when the dual derivate $D'(\cdot)$ is likely to be large, then large steps can be taken in the dual descent step to accelerate the convergence. Then, *en route* to convergence the dual gradient decreases and it is natural to take more conservative steps. Another advantage of time varying codes is that they can asymptotically converge to the true primal/dual optimal solution. We consider time varying quantizers of the form

$$\theta_t(z) = \begin{cases} \gamma(t) & \text{if } z \geq \kappa(t) \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

We now show that $(P, \theta_t(\cdot))$ with $\theta_t(\cdot)$ in (14) is a PF-code and prove the asymptotic convergence of such codes.

Theorem 4. *Consider the quantizer* θ_t *in* (14) *with* $\gamma(t), \kappa(t) > 0$ *for* $t \in \mathbb{N}$. *Then*
i) $(P, \theta_t(\cdot))$ *is a PF-code if* $\kappa(t) \leq \gamma(t)L$ *for* $t \in \mathbb{N}$.
ii) If $\kappa(t) = \gamma(t)L$, $\lim_{t \to \infty} \gamma(t) = 0$, *and* $\sum_{t=1}^\infty \gamma(t) = \infty$ *then* $\lim_{t \to \infty} p(t) = \bar{p}^\star$ *and* $\lim_{t \to \infty} x_i(t) = x_i^\star$, *where* $[\underline{p}^\star, \bar{p}^\star]$ *is the set of optimal solutions to the dual problem and* $[x_i^\star]_{i=1,\cdots,N}$ *is the optimal solution to the primal problem.*

Proof. i) As in the proof of Theorem 3-i), $\theta_t(z) \leq (1/L)z$ for all $z \geq 0$, so the result follows from Theorem 1.

ii) Since $\gamma(t) > 0$, the sequence $p(t)$ is decreasing, i.e., $p(t+1) \leq p(t)$ for all $t \in \mathbb{N}$. Since $\theta_t(z) \leq (1/L)z$ holds for all $z \geq 0$, $p(t)$ is bounded below by $\bar{p}^\star$, i.e., $\bar{p}^\star \leq p(t)$ for $t \in \mathbb{N}$. Therefore, the limit $\hat{p} := \lim_{t \to \infty} p(t)$ exists and $\bar{p}^\star \leq \hat{p}$. If $\bar{p}^\star = \hat{p}$ then $\lim_{t \to \infty} x_i(p(t)) = x_i^\star$, yielding the result. To conclude the proof, we now show that $\bar{p}^\star = \hat{p}$.

We prove $\bar{p}^\star = \hat{p}$ by contradiction: Suppose $\bar{p}^\star < \hat{p}$, then $D'(\hat{p}) > 0$ and $\lim_{t \to \infty} \kappa(t) = 0$. Hence, $T \in \mathbb{N}$ can be chosen so that $D'(\hat{p}) \geq \kappa(t)$ for $t \geq T$. Therefore, $D'(p(t)) \geq \kappa(t)$ for $t \geq T$, $D'(\cdot)$ is monotone since $D(\cdot)$ is convex. As a result for $t \geq T$ we have that $\theta_t(D'(p(t))) = \gamma(t)$ and $p(t+1) = \left[p(T) - \sum_{k=T}^t \gamma(t) \right]^+$. Since $\sum_{k=T}^t \gamma(t) = \infty$, we have $\hat{p} = \lim_{t \to \infty} p(t) = 0$, which contradicts $\hat{p} > \bar{p}^\star \geq 0$. $\square$

We next illustrate the results in simulations.

V. NUMERICAL SIMULATIONS

Consider a power supplier in a micro grid with the task of supplying $N = 40$ users (or devices) with $C = 200$ units of power. Each user $i \in \mathcal{N}$ has the power demand d_i and the supplier's task is to fairly allocate the limited power among the users by solving Problem (1) with the

(a) Dual derivative.

(b) Primal/Dual objective values

(c) Primal variables.
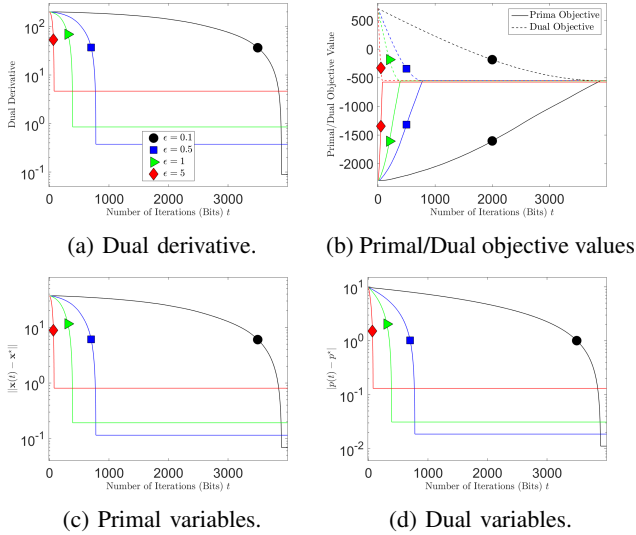

(d) Dual variables.

Fig. 2: Convergence of the ONEWAY-DD Algorithm for the time-invariant PF-code (13) with $\gamma = \epsilon/L$ and $\kappa = \epsilon$ [as in Theorem 3-ii)] where $\epsilon = 0.1, 0.5, 1, 5$.

utility functions given by $U_i(x_i) = -0.5(x_i - d_i)^2$. The power demand d_i of each user $i \in \mathcal{N}$ is taken uniformly at random from the interval $[5, 15]$. The local constraint of each user $i \in \mathcal{N}$ is $[m_i, M_i] = [0, 15]$. Direct inspection shows that Assumptions 2 holds and that U_i is 1-concave and $U_i'(0) \leq P := 15$, i.e., Assumptions 1 and 3 hold. Therefore, the dual function of the problem is L-smooth with $L = N$, by Proposition 1. We apply the ONEWAY-DD Algorithm on the problem and use the time-invariant PF-code (13) with $\gamma = \epsilon/L$ and $\kappa = \epsilon$ [as in Theorem 3-ii)] to achieve $\epsilon > 0$ solution accuracy with $\epsilon = 0.1, 0.5, 1, 5$.

Figure 2a depicts the dual derivative $D'(p(t))$ at every iteration. The results show that the $\epsilon = 0.1, 0.5, 1, 5$ accuracy is reached using 3899, 780, 390, and 78 bits compared with the theoretical upper bound 6000, 1200, 600, and 120 from Theorem 3-ii). Figure 2b depicts the primal and dual objective function values at every iteration. The duality gap after the ϵ-accuracy is reached, i.e., $D(p(t)) - \sum_{i \in \mathcal{N}} U_i(x_i)$ with $t \geq T$ [Theorem 3-ii)], is 0.47, 1.97, 4.48, and 24.99, for $\epsilon = 0.1, 0.5, 1$, and 5, respectively. Since a PF-code is used, $[x_i(p(t))]_{i \in \mathcal{N}}$ are always primal feasible; hence the duality gap shows how far the feasible (suboptimal) points $[x_i(p(t))]_{i \in \mathcal{N}}$ are from true optimal value. Figures 2c and 2d show how the primal/dual iterates converge to the optimal primal/dual point, with good accuracy in all cases.

APPENDIX

Proof of Proposition 1. i) See Lemma II.2 in [5].

ii) Clearly, (2) has an optimal solution $p^\star \in \mathbb{R}_+$ due to Lemma 1. We prove the result by showing that if there is an optimal solution $p^\star$ such that $p^\star \geq P$, then P is also an optimal solution to (2). Since $P = \max_{i \in \mathcal{N}} U_i'(m_i)$, we have $U_i'(m_i) \leq P$ for all $i \in \mathcal{N}$. Moreover, $m_i \geq (U_i')^{-1}(P)$ 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. In particular, $x_i(P) = [(U_i')^{-1}(P)]_{m_i}^{M_i} = m_i$ and hence $x_i(p) = m_i$ for $p \geq P$. As a result, if (2) has an optimal solution $p^\star$ with $p^\star \geq P$ then P is also an optimal solution to (2). ∎

Proof of Lemma 2. (i) $\Longleftrightarrow$ (ii): $D'(p) = C - \sum_{i=1}^{N} x_i(p)$ from Equation (5). Therefore, if $D'(p) < 0$ then $C < \sum_{i=1}^{N} x_i(p)$ and if $0 \leq D'(p)$ then $\sum_{i=1}^{N} x_i(p) \leq C$.

(ii) $\Longleftrightarrow$ (iii): For $p^\star \in \mathcal{P}^\star$, we have $D'(p^\star) = 0$. Since D is a convex function, $D'(p) = C - \sum_{i=1}^{N} x_i(p)$ is increasing. Therefore, if $p^\star \in \mathcal{P}^\star$ then $0 = D'(p^\star) \leq D'(p) = C - \sum_{i=1}^{N} x_i(p)$ for $p \geq p^\star$. Similarly, if $p < p^\star$ then $D'(p) < 0$. ∎

REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2619–2692, 1998.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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 *Information Sciences and Systems (CISS), 50th IEEE Annual Conference on*, March 2016.

[10] J. Herdtner and E. Chong, "Analysis of a class of distributed asynchronous power control algorithms for cellular wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 436–446, March 2000.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

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

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