

Primal Recovery from Consensus-Based Dual Decomposition for Distributed Convex Optimization

Andrea Simonetto¹ · Hadi Jamali-Rad¹

Received: 28 August 2014 / Accepted: 13 May 2015 / Published online: 29 May 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Dual decomposition has been successfully employed in a variety of distributed convex optimization problems solved by a network of computing and communicating nodes. Often, when the cost function is separable but the constraints are coupled, the dual decomposition scheme involves local parallel subgradient calculations and a global subgradient update performed by a master node. In this paper, we propose a consensus-based dual decomposition to remove the need for such a master node and still enable the computing nodes to generate an approximate dual solution for the underlying convex optimization problem. In addition, we provide a primal recovery mechanism to allow the nodes to have access to approximate near-optimal primal solutions. Our scheme is based on a constant stepsize choice, and the dual and primal objective convergence are achieved up to a bounded error floor dependent on the stepsize and on the number of consensus steps among the nodes.

Keywords Distributed convex optimization · Dual decomposition · Primal recovery · Consensus algorithm · Subgradient optimization · Epsilon-subgradient · Ergodic convergence

Mathematics Subject Classification 90C25 · 90C30 · 90C46 · 90C59

✉ Andrea Simonetto
a.simonetto@tudelft.nl
Hadi Jamali-Rad
h.jamalirad@tudelft.nl

¹ Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, 2826 CD Delft, The Netherlands

1 Introduction

Lagrangian relaxation and dual decomposition are extremely effective in solving large-scale convex optimization problems [1–6]. Dual decomposition has also been employed successfully in the field of distributed convex optimization, where the optimization problem requires to be decomposed among cooperative computing entities (called in the following simply by nodes). In this case, the optimization problem is generally divided into two steps, a first step pertaining the calculation of the local subgradients of the Lagrangian dual function, and a second step consisting of the global update of the dual variables by projected subgradient ascent. The first step can typically be performed in parallel on the nodes, whereas the second step has often to be performed centrally, by a so-called *master* node (or data-gathering node, or fusion center), which combines the local subgradient information.

Even though by solving the dual problem, one obtains a lower bound on the optimal value of the original convex problem, in practical situations, one would also like to have access to an approximate primal solution. However, even with the availability of an approximate dual optimal solution, a primal one cannot be easily obtained. The reason is that the Lagrangian dual function is generally nonsmooth at an optimal point, thus an optimal primal solution is not a trivial combination of the extreme subproblem solutions. Methods to recover approximate (near-optimal) primal solutions from the information coming from dual decomposition have been proposed in the past [4, 7–13] (and references therein). In one way or another, all these methods use a combination of all the approximate primal solutions that are generated, while the dual decomposition scheme converges to a near-optimal dual solution. A possible choice for the combination is the ergodic mean [4, 11, 14].

Among the dual decomposition schemes with primal recovery mechanism available in the literature, we are interested here in the ones that employ a constant stepsize in the projected dual subgradient update. The reasons are twofold. First of all, a constant stepsize yields faster convergence to a bounded error floor, which is fundamental in real-time applications (e.g., control of networked systems). In addition, the error floor can be tuned by trading-off the number of iterations required and the value of the stepsize. The second reason is that in many situations, the underlying convex optimization problem is not stationary, but changes over time. Having in mind the development of methods to update the dual variables while the optimization problem varies [15–17], it is of key importance to employ a constant stepsize. In this way, the capability of the subgradient scheme to track the dual optimal solutions does not change over time due to a vanishing stepsize approach.

In this paper, we propose a way to remove the need for a *master* node to collect the local subgradient information coming from the different nodes and generate a global subgradient. The reason is that in distributed systems, the nodes are connected via an ad-hoc network and the communication is often limited to geographically nearby nodes. It is therefore impractical to collect the local subgradient information in one physical location, whereas it is advisable to enable the nodes themselves to have access to a suitable approximation of the global subgradient. We use consensus-based mechanisms to construct such an approximation. Consensus-based mechanisms have been used in the primal domain both with constant stepsizes [18, 19] and with vanishing

ones [19–21], however, to the best of the authors' knowledge, they have not been used in the dual domain, and not together with primal recovery. An interesting, but different, approach applying consensus on the cutting-plane algorithm to solve the master problem has been very recently proposed in [22]. Our main contributions can be described as follows.

First, we develop a constant stepsize consensus-based dual decomposition. Our method enables the different nodes to generate a sequence of approximate dual optimal solutions whose dual cost eventually converges to the optimal dual cost within a bounded error floor. Under the assumptions of convexity, compactness of the feasible set, and Slater's condition, the convergence goes as $O(1/k)$, where k is the number of iterations. The error depends on the stepsize and on the number of consensus steps between subsequent iterations k . Furthermore, the nodes are exchanging subgradient information only with their nearby neighboring nodes.

Then, since in our method, each node maintains its own approximate dual sequence, we provide an upper bound on the disagreement among the nodes and we prove its convergences to a bounded value.

Finally, we propose a primal recovery scheme to generate approximate primal solutions from consensus-based dual decomposition. This scheme is proven to converge to the optimal primal cost up to a bounded error floor. Once again, under the same assumptions, the convergence goes as $O(1/k)$ and the error depends on the stepsize and on the number of consensus steps.

Organization Section 2 describes the problem setting, our main research question, and some sample applications. In Sect. 3, we cover the basics of dual decomposition to pinpoint the main limitation, i.e., the need for a *master* node. We propose, develop, and investigate the convergence results of our algorithm in Sects. 4 and 5. All the proofs are contained in Sects. 6 and 7. In Sect. 8, we collect numerical simulation results. Future research questions and conclusions are discussed in Sects. 9 and 10, respectively.

2 Problem Formulation

Notation For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the standard inner product is indicated as $\langle \mathbf{x}, \mathbf{y} \rangle$, while its induced (Euclidean) norm is represented as $\|\mathbf{x}\|_2$. A vector \mathbf{x} belongs to \mathbb{R}_+^n iff it is of size n , and all its components are nonnegative (i.e., \mathbb{R}_+^n is the nonnegative orthant). For any vectors $\mathbf{x} \in \mathbb{R}^n$, its components are indicated by x_i , $i \in \{1, \dots, n\}$. The vector $\mathbf{1}_n$ is the column vector of length n containing only ones. We indicate by \mathbf{I}_n the identity matrix of size n . For any real-valued squared matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, we say $\mathbf{X} \succeq 0$ or $\mathbf{X} \preceq 0$ iff the matrix is positive semi-definite or negative semi-definite, respectively. We also write $\mathbf{X} \in \mathbb{S}_+^n$, iff $\mathbf{X} \succeq 0$. For any real-valued squared matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$, the norm $\|\mathbf{X}\|_F$ represents the Frobenius norm, while the trace is indicated by $\text{tr}[\mathbf{X}]$. The symbol $(\cdot)^T$ is the transpose operator, \otimes represents the Kronecker product, \circ stands for map composition, $\text{conv}[\cdot]$ is the convex hull, $\text{vec}(\cdot)$ is the vectorization operator, while $\mathbf{P}_X[\cdot]$ is the projection operator onto the set X . The ϵ -subgradient of a *concave* function $q(\mathbf{x}) : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$, for the nonnegative scalar $\epsilon \geq 0$, at $\mathbf{x}' \in X$ is a vector $\tilde{\mathbf{g}} \in \mathbb{R}^n$ such that

$$\langle \tilde{g}, y - x' \rangle \geq q(y) - q(x') - \epsilon, \quad \forall y \in X. \tag{1}$$

Furthermore, the collection of ϵ -subgradients of $q(x)$ at x' is called the ϵ -subdifferential set, denoted by $\partial_\epsilon q_x(x')$. If $\epsilon = 0$, the ϵ -subgradient is the regular subgradient and we drop the ϵ in the notation of the subdifferential.

Formulation We consider a convex optimization problem defined on a network of computing and communicating nodes. Let the nodes be labeled with $i \in V = \{1, \dots, n\}$, and we equip each of them with the local (private) convex function $f_i(x_i) : \mathbb{R} \rightarrow \mathbb{R}$. Let x be the stacked vector of all the local decision variables, i.e., $x = (x_1, \dots, x_n)^T$. Let the functions $g_i(x_i) : \mathbb{R} \rightarrow \mathbb{R}, i \in V$ be convex. Let $A_0, A_i, i \in V$ be $d \times d$ real-valued square and symmetric matrices. Let $X_i \subset \mathbb{R}, i \in V$ be convex and compact sets, and let $X := \prod_{i \in V} X_i$. We are interested in solving decomposable convex optimization problems of the form,

$$\underset{x_i \in X_i, i \in V}{\text{minimize}} \quad f(x) := \sum_{i \in V} f_i(x_i) \tag{2a}$$

$$\text{subject to} \quad \sum_{i \in V} g_i(x_i) \leq 0, \tag{2b}$$

$$A_0 + \sum_{i \in V} A_i x_i \geq 0. \tag{2c}$$

In order to simplify our notation (and without loss of generality), we have chosen to work with scalar decision variables x_i , with one scalar inequality and with one linear matrix inequality. The following assumptions are in place.

Assumption 2.1 (*Convexity and compactness*) The cost functions $f_i(x_i)$ and the constraint functions $g_i(x_i)$ are convex in x_i for each i . The sets X_i are convex and compact (thus, bounded). The matrices $A_0, A_i, i \in V$ are real-valued square and symmetric.

Assumption 2.2 (*Existence of solution*) The feasible set $F := \{x \in X \mid (2b) \text{ and } (2c)\}$ is nonempty; for all $x \in F$, the cost function $f(x) > -\infty$, and there exists a vector $x \in F$ such that $f(x) < \infty$.

Assumption 2.3 (*Slater condition*) There exists a vector $\bar{x} \in \mathbb{R}^n$ that is strictly feasible for problem (2), i.e.,

$$\sum_{i \in V} g_i(\bar{x}_i) < 0, \text{ and } A_0 + \sum_{i \in V} A_i \bar{x}_i > 0.$$

Assumption 2.4 (*Communication network*) The computing nodes communicate synchronously via undirected time-invariant communication links.

Assumption 2.1 is required to ensure a convex program with compact feasible set. Assumption 2.2 ensures the existence of a solution for the optimization problem (2). Let x^* be such a (possibly not unique) solution (i.e., a minimizer) and let f^* be the unique minimum. Assumption 2.3 is often required in dual decomposition approaches

in order to guarantee *zero duality gap* and to be able to derive the optimal value of the optimization problem (2) by solving its dual. In addition, Slater condition helps in bounding the dual variables, which is crucial in our convergence analysis. Assumption 2.4 is required to simplify the convergence analysis. One might be able to loosen it and require only asynchronous communications, but this is left for future research since it is not the core idea of this paper. By Assumption 2.4, we can define an undirected communication graph \mathcal{G} consisting of a vertex set V as well as an edge set E . For each node i , we call *neighborhood*, or N_i , the set of the nodes it can communicate with.

The main research problem we tackle in this paper can be stated as follows.

Research Problem: We would like to devise an algorithm that enables each node, by communicating with their neighbors only, to construct a sequence of approximate local optimizers $\{x_i^k\}$, for which their primal objective sequence $\{f(x^k)\}$ eventually converges to f^* (possibly) up to a bounded error floor.

Our approach toward this problem is to devise a consensus-based dual decomposition with approximate primal recovery.

Sample Applications Problems as (2) appear in many contexts: The first example we cite is the network utility maximization (NUM) problem, where a group of communication nodes try to maximize their utility subject to a resource allocation constraint [23,24]. NUM problems are very relevant in communication systems. Generalizations of NUM problems, where the cost function is separable and the variables are constrained by linear inequalities, can also be handled by (2), and have been considered, e.g., in model predictive controller design [25] (which is one of the workhorse of nowadays control theory). Another sample application is sensor selection, where a set of nodes try to decide which one of them should be activated to perform a certain task based on a given metric. This is in general a combinatorial problem, yet it can be relaxed to a semi-definite program, which is a generalization of (2), [26,27]. In the latter example, the constraint (2c) plays an important role.

Multi-agent/Multiuser/Networked Problems If the constraints (2b) and (2c) involve only local functions, that is, the sum is only over the neighbors of a particular i , then we have what is known as multi-agent (or multiuser, or networked) problem. These problems can be further complicated by the presence of global decision variables. In all these cases, due to the presence of neighborhood constraint functions only, the dual variables associated with them can be computed locally in the neighborhood (we can refer to them as link dual variables). Therefore, by a proper use of dual decomposition, we can devise distributed algorithms that can be implemented on nodes and connecting links. Relevant recent work on these problems is reported in [28–35]. In our case, the constraints (2b)–(2c) involve constraint functions from all the nodes, in all the decision variables together; therefore, the proposed methods for multi-agent problems cannot be directly applied in our case. In general, the link dual variables become a network-wide dual variable in our case, and we retrieve the standard dual decomposition scheme with the need for a master node to compute such a global network-wide dual variable.

3 Dual Decomposition

The Lagrangian function $L(\mathbf{x}, \mu, \mathbf{G}) : \mathbb{R}^n \times \mathbb{R}_+ \times \mathbb{S}_+^d \rightarrow \mathbb{R}$ is formed, as a first step of dual decomposition,

$$L(\mathbf{x}, \mu, \mathbf{G}) := \sum_{i \in V} f_i(x_i) + \mu \left(\sum_{i \in V} g_i(x_i) \right) - \text{tr} \left[\left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i x_i \right) \mathbf{G} \right], \tag{3}$$

where $\mu \in \mathbb{R}_+$ is the dual variable associated with the constraint (2b), and $\mathbf{G} \in \mathbb{S}_+^d$ is the dual variable associated with (2c). Further, the dual function $q(\mu, \mathbf{G}) : \mathbb{R}_+ \times \mathbb{S}_+^d \rightarrow \mathbb{R}$ can be defined as

$$q(\mu, \mathbf{G}) := \min_{\mathbf{x} \in X} \{L(\mathbf{x}, \mu, \mathbf{G})\}. \tag{4}$$

The set X is compact, which means that the function $q(\mu, \mathbf{G})$ is continuous on $\mathbb{R}_+ \times \mathbb{S}_+^d$. Furthermore, the function $q(\mu, \mathbf{G})$ is concave. For any pair of dual variables (μ, \mathbf{G}) , we can compute the value of the primal minimizers and their set:

$$\tilde{\mathbf{x}} := \underset{\mathbf{x} \in X}{\text{argmin}} \{L(\mathbf{x}, \mu, \mathbf{G})\}, \quad \tilde{X} := \{\mathbf{x} \in X \mid q(\mu, \mathbf{G}) = L(\mathbf{x}, \mu, \mathbf{G})\}. \tag{5}$$

Given the compactness of X and the form of the dual function (4), we can define the subdifferential of $q(\mu, \mathbf{G})$ at μ and \mathbf{G} as the following sets

$$\partial q_\mu(\mu, \mathbf{G}) := \text{conv} \left[\sum_{i \in V} g_i(\tilde{x}_i) \mid \tilde{\mathbf{x}} \in \tilde{X} \right], \tag{6a}$$

$$\partial q_{\mathbf{G}}(\mu, \mathbf{G}) := \text{conv} \left[- \left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \tilde{x}_i \right) \mid \tilde{\mathbf{x}} \in \tilde{X} \right], \tag{6b}$$

Subgradient choices for $q(\mu, \mathbf{G})$ are therefore

$$h(\tilde{\mathbf{x}}) := \sum_{i \in V} g_i(\tilde{x}_i) \in \partial q_\mu(\mu, \mathbf{G}), \quad \mathbf{Q}(\tilde{\mathbf{x}}) := -\mathbf{A}_0 - \sum_{i \in V} \mathbf{A}_i \tilde{x}_i \in \partial q_{\mathbf{G}}(\mu, \mathbf{G}), \tag{7}$$

for any choice of $\tilde{\mathbf{x}} \in \tilde{X}$. In addition, since X is compact and the constraints (2b)–(2c) are represented by continuous functions, the subgradients are bounded, and we set, for all $i \in V$

$$\|h_i(\mathbf{x})\|_2 \leq \max_{x_i \in X_i} \|g_i(x_i)\|_2 =: L, \quad \|\mathbf{Q}_i(\mathbf{x})\|_F \leq \max_{x_i \in X_i} \left\| -\mathbf{A}_0/n - \mathbf{A}_i x_i \right\|_F =: Q, \tag{8}$$

where we have defined $h_i(\mathbf{x}) := g_i(x_i)$, and $\mathbf{Q}_i(\mathbf{x}) := -\mathbf{A}_0/n - \mathbf{A}_i x_i$. Finally, the Lagrangian dual problem can be written as

$$q^* := \sup_{\mu \in \mathbb{R}, \mathbf{G} \in \mathbb{S}_+^d} \{q(\mu, \mathbf{G})\}, \tag{9}$$

and by Slater condition (Assumption 2.3), strong duality holds: $q^* = f^*$.

Since the original convex optimization problem (2) is decomposable, the Lagrangian function is separable as

$$L(\mathbf{x}, \mu, \mathbf{G}) = \sum_{i \in V} \left(f_i(x_i) + \mu g_i(x_i) - \text{tr} \left[\left(\mathbf{A}_0/n + \mathbf{A}_i x_i \right) \mathbf{G} \right] \right) =: \sum_{i \in V} L_i(x_i, \mu, \mathbf{G}), \tag{10}$$

and so is the dual function

$$q(\mu, \mathbf{G}) := \sum_{i \in V} \min_{x_i \in X_i} \{L_i(x_i, \mu, \mathbf{G})\} := \sum_{i \in V} q_i(\mu, \mathbf{G}), \tag{11}$$

and its subgradients.

Dual decomposition with approximate primal recovery as defined in [4] is summarized in the following algorithm.

Dual decomposition with primal recovery

1. Initialize $\mu^0 \in \mathbb{R}_+$, $\mathbf{G}^0 \in \mathbb{S}_+^d$, choose a constant stepsize α ;
2. Local dual optimization: compute in parallel the local dual functions and their primal optimizers

$$q_i(\mu^k, \mathbf{G}^k) = \min_{x_i \in X_i} \left\{ L_i(x_i, \mu^k, \mathbf{G}^k) \right\}, \quad \tilde{x}_i^k = \underset{x_i \in X_i}{\operatorname{argmin}} \left\{ L_i(x_i, \mu^k, \mathbf{G}^k) \right\}, \tag{12a}$$
 as well as their subgradients $g_i(\tilde{x}_i^k)$ and $-\mathbf{A}_0/n - \mathbf{A}_i \tilde{x}_i^k$;
3. Primal recovery step: compute in parallel the ergodic sum, for $k \geq 1$

$$x_i^k = \frac{1}{k} \sum_{t=1}^k \tilde{x}_i^t; \tag{12b}$$
4. Dual update: update the variables μ^k, \mathbf{G}^k as

$$\mu^{k+1} = \mathbb{P}_{\mathbb{R}_+} \left[\mu^k + \alpha \sum_{i \in V} g_i(\tilde{x}_i^k) \right] \tag{12c}$$

$$\mathbf{G}^{k+1} = \mathbb{P}_{\mathbb{S}_+^d} \left[\mathbf{G}^k - \alpha \left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \tilde{x}_i^k \right) \right]. \tag{12d}$$

This algorithm generates a converging sequence $\{x_i^k\}$ as detailed in the following theorem.

Theorem 3.1 *Let the sequence $\{\mu^k, \mathbf{G}^k, \mathbf{x}^k\}$ be generated by the iterations (12). Let L and Q be defined as in (8). Under Assumptions 2.1 till 2.3,*

- (a) *the dual variables are bounded, i.e., $\|\mu^k\|_2 \leq \Lambda_0 < \infty$, $\|\mathbf{G}^k\|_{\mathbb{F}} \leq \Gamma_0 < \infty$, for all $k \geq 1$;*
- (b) *an upper bound on the primal cost of the vector \mathbf{x}^k , $k \geq 1$, is given by*

$$f(\mathbf{x}^k) \leq f^* + \frac{\Lambda_0^2 + \Gamma_0^2}{2\alpha k} + \frac{\alpha n^2(L^2 + Q^2)}{2};$$

- (c) *a lower bound on the primal cost of the vector \mathbf{x}^k , $k \geq 1$, is given by*

$$f(\mathbf{x}^k) \geq f^* - \frac{\Lambda_0^2 + \Gamma_0^2}{\alpha k}.$$

Proof The proof follows from [4, Lemma 3 and Proposition 1]. Since our optimization problem involves also a linear matrix inequality, some extra steps are needed in the proof of part (c). To be more specific, by following the same steps in the proof of [4, Proposition 1(c)], we arrive at the following inequality

$$f(\mathbf{x}^k) \geq f^* - \mu^* h(\mathbf{x}^k) - \text{tr}[\mathbf{Q}(\mathbf{x}^k)\mathbf{G}^*]. \tag{13}$$

where $\mu^* \geq 0$ and $\mathbf{G}^* \succeq 0$ are the optimal dual variables. We now need to find a lower bound for the rightmost term of (13). By similar arguments of the proof of [4, Proposition 1(a)], we obtain for all $k \geq 1$

$$h(\mathbf{x}^k) \leq \frac{\mu^k}{\alpha k}, \quad \frac{\mathbf{G}^k}{\alpha k} - \mathbf{Q}(\mathbf{x}^k) \succeq 0. \tag{14}$$

Given the two positive semi-definite matrices \mathbf{X} and \mathbf{Y} of dimension $n \times n$, we know that $\text{tr}[\mathbf{X}\mathbf{Y}] \geq \lambda_{\min}(\mathbf{X})\text{tr}[\mathbf{Y}] \geq 0$, [36, Lemma 1], which means

$$\text{tr}\left[\left(\frac{\mathbf{G}^k}{\alpha k} - \mathbf{Q}(\mathbf{x}^k)\right)\mathbf{G}^*\right] \geq 0, \quad \text{thus} \quad \text{tr}\left[\left(\frac{\mathbf{G}^k}{\alpha k}\right)\mathbf{G}^*\right] \geq \text{tr}[\mathbf{Q}(\mathbf{x}^k)\mathbf{G}^*].$$

This implies that for $k \geq 1$

$$\text{tr}[\mathbf{Q}(\mathbf{x}^k)\mathbf{G}^*] \leq \text{tr}\left[\frac{\mathbf{G}^k}{\alpha k}\mathbf{G}^*\right] = \left|\text{tr}\left[\frac{\mathbf{G}^k}{\alpha k}\mathbf{G}^*\right]\right| \leq \frac{1}{\alpha k}\|\mathbf{G}^k\|_{\text{F}}\|\mathbf{G}^*\|_{\text{F}} \leq \frac{\Gamma_0^2}{\alpha k}, \tag{15}$$

where we have used Cauchy–Schwarz inequality [37]. By combining (15) and (14) with (13), we obtain the lower bound

$$f(\mathbf{x}^k) \geq f^* - \mu^* h(\mathbf{x}^k) - \text{tr}[\mathbf{Q}(\mathbf{x}^k)\mathbf{G}^*] \geq f^* - \frac{\Lambda_0^2}{\alpha k} - \frac{\Gamma_0^2}{\alpha k},$$

and the claim is proven. □

Although the dual decomposition method of [4] presents several advantages, in practice, the nodes will need to sum the subgradients coming from the whole network in Step 4 in order to maintain common dual variables. This is often not practical in large networks, because it would call for a significant communication overhead.

In the following sections, (i) we propose a consensus-based dual decomposition with primal recovery mechanism to modify Step 4 in order to make it suitable for limited information exchange (i.e., communication only with neighboring nodes); (ii) we prove dual and primal objective convergence of the proposed method up to a bounded error floor which depends (among other things) on the number of communication exchange with the neighboring nodes for each iteration k .

4 Basic Relations

Lemma 4.1 *Suppose Assumption 2.1 till 2.3 hold. Let $\bar{\mu} \geq 0, \bar{\mathbf{G}} \geq 0$ be a pair of dual variables for which the set $\bar{D} := \{(\mu \geq 0, \mathbf{G} \geq 0) | q(\mu, \mathbf{G}) \geq q(\bar{\mu}, \bar{\mathbf{G}})\}$ is nonempty. Then, the set \bar{D} is bounded and we have*

$$\max_{(\mu, \mathbf{G}) \in \bar{D}} \|\mu\|_2 + \|\mathbf{G}\|_F \leq \frac{1}{\gamma} (f(\bar{\mathbf{x}}) - q(\bar{\mu}, \bar{\mathbf{G}})),$$

where $\gamma := \min \left\{ \sum_{i \in V} -g_i(\bar{x}_i), \lambda_{\min}(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \bar{x}_i) \right\}$, $\lambda_{\min}(\cdot)$ is the smallest eigenvalue and $\bar{\mathbf{x}}$ is a vector satisfying the Slater condition.

Proof The lemma follows from [4, Lemma 1] with minor modifications. In particular, we use [36, Lemma 1] to bound the inner product

$$\text{tr} \left[\left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \bar{x}_i \right) \mathbf{G}^k \right] \geq \lambda_{\min} \left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \bar{x}_i \right) \text{tr}[\mathbf{G}^k],$$

and the fact that $\|\mathbf{G}\|_F \leq \text{tr}[\mathbf{G}]$, [37]. The remaining steps are omitted since similar to [4, Lemma 1]. □

It follows from the result of the preceding lemma that under Slater, the dual optimal set D^* is nonempty. Since $D^* := \{(\mu \geq 0, \mathbf{G} \geq 0) | q(\mu, \mathbf{G}) \geq q^*\}$, by using Lemma 4.1, we obtain

$$\max_{(\mu^*, \mathbf{G}^*) \in D^*} \|\mu^*\|_2 + \|\mathbf{G}^*\|_F \leq \frac{1}{\gamma} (f(\bar{\mathbf{x}}) - q^*).$$

Furthermore, although the dual optimal value q^* is not a priori available, one can compute a looser bound by computing the dual function for some couple $(\tilde{\mu} \geq 0, \tilde{\mathbf{G}} \geq 0)$. Owing to optimality, $q^* \geq q(\tilde{\mu}, \tilde{\mathbf{G}})$, thus

$$\max_{(\mu^*, \mathbf{G}^*) \in D^*} \|\mu^*\|_2 + \|\mathbf{G}^*\|_F \leq \frac{1}{\gamma} (f(\bar{\mathbf{x}}) - q(\tilde{\mu}, \tilde{\mathbf{G}})).$$

This result is quite useful to render the dual decomposition method easier to study. In fact, as in [4], we can modify the sets over which we project in Step 4 by considering a bounded superset of the dual optimal solution set. This means that we can substitute Step 4 in (12) with

$$\mu^{k+1} = P_{D_\mu} \left[\mu^k + \alpha \sum_{i \in V} g_i(\bar{x}_i^k) \right], \quad D_\mu := \left\{ \mu \geq 0 \mid \|\mu\|_2 \leq \frac{f(\bar{\mathbf{x}}) - q(\tilde{\mu}, \tilde{\mathbf{G}})}{\gamma} + r \right\} \quad (16a)$$

$$\mathbf{G}^{k+1} = P_{D_G} \left[\mathbf{G}^k - \alpha \left(\mathbf{A}_0 + \sum_{i \in V} \mathbf{A}_i \bar{x}_i^k \right) \right], \quad D_G := \left\{ \mathbf{G} \geq 0 \mid \|\mathbf{G}\|_F \leq \frac{f(\bar{\mathbf{x}}) - q(\tilde{\mu}, \tilde{\mathbf{G}})}{\gamma} + r \right\} \quad (16b)$$

for a given scalar $r > 0$. The nice feature of this modification is that both D_μ and D_G are now compact convex sets. This does not increase computational complexity, and it is a useful modification, for it provides a leverage to derive the convergence rate results. In the following, for convergence purposes, we will use $r \geq \frac{f(\bar{x}) - q(\tilde{\mu}, \tilde{G})}{\gamma}$.

5 Consensus-Based Dual Decomposition

We consider now a consensus-based update to enforce the update rule of dual decomposition in (16) to fit the constraint of a limited communication network. Our approach is inspired by the one of [18] but applied to the dual domain. First of all, we define a consensus matrix $W \in \mathbb{R}^{n \times n}$, with the following properties:

$$[W]_{ij} = 0 \text{ if } j \notin N_i \cup \{i\}, \quad W = W^T, \quad W \mathbf{1}_n = \mathbf{1}_n, \quad \rho \left[W - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n} \right] \leq \nu < 1, \quad (17)$$

where $\rho[\cdot]$ returns the spectral radius and ν is an upper bound on the value of the spectral radius. It is a common practice to generate such consensus matrices; a possible choice is the Metropolis-Hasting weighting matrix [38,39].

A consensus iteration is a linear mapping $C(x) : x \mapsto Wx$ with the property that the result of its repeated application converges to the mean of the initial vector, i.e., for $x \in \mathbb{R}^n$

$$\lim_{\varphi \rightarrow \infty} \underbrace{C \circ C \circ \dots \circ C}_{\varphi \text{ times}}(x) = \lim_{\varphi \rightarrow \infty} W^\varphi x = \frac{\mathbf{1}_n \mathbf{1}_n^T}{n} x.$$

This averaging property is ensured, for example, by conditions as the ones in (17). In addition, given the structure of W in (17), each consensus iteration involves only local communications (only the neighboring nodes will share their local variables), which will be the key point of our modification. In the following, we will study multiple consensus steps, in the sense that the computing nodes will run multiple consensus iterations (each of which involving only local communications) between subsequent iterations k 's. We let the number of consensus steps be $\varphi \in \mathbb{N}$. In this case, the consensus mapping will be of the form $x \mapsto W^\varphi x$. Since we will enable each node to generate its own dual variables on which consensus will be enforced, we start by defining local versions of μ and G as $\mu_i \in \mathbb{R}_+$ and $G_i \in \mathbb{S}_+^d$, respectively. Next, we define our consensus-based dual decomposition as the following algorithm.

We highlight that the proposed algorithm CoBa-DD (or (18)) involves only local communication. The only communication involved is in the φ consensus steps, each of which requiring the nodes to share information with their neighbors. Also, note that computing $(f(\bar{x}) - q(\tilde{\mu}, \tilde{G}))/\gamma$ (for the definition of D_μ and D_G) is not a very difficult task, since a Slater vector is usually easy to find by inspection, and both $f(\bar{x})$ and γ can be computed by a consensus algorithm run in the initialization step of CoBa-DD.

In order to analyze dual and primal convergence of (18), we start by some basic results. First, given that the sets D_μ and D_G are compact, and that μ_i^0 and G_i^0 are picked to be bounded, the dual variables μ_i^k and G_i^k are bounded for each $k \geq 0$.

Consensus-based dual decomposition with primal recovery (CoBa-DD)

1. Initialize $\mu_i^0 \in \mathbb{R}_+$, $\mathbf{G}_i^0 \in \mathbb{S}_+^d$, $i \in V$, choose $\alpha > 0$, determine a Slater vector $\bar{\mathbf{x}}$ and the sets D_μ and D_G of (16) with an arbitrarily picked $\tilde{\mu}$, $\tilde{\mathbf{G}}$ and a scalar $r \geq \frac{f(\bar{\mathbf{x}}) - q(\tilde{\mu}, \tilde{\mathbf{G}})}{\gamma}$; pick a number of consensus steps φ ;

2. Local dual optimization: compute in parallel the local dual functions and their primal optimizers $q_i(\mu_i^k, \mathbf{G}_i^k) = \min_{x_i \in X_i} \{L_i(x_i, \mu_i^k, \mathbf{G}_i^k)\}$, $\tilde{x}_i^k = \operatorname{argmin}_{x_i \in X_i} \{L_i(x_i, \mu_i^k, \mathbf{G}_i^k)\}$, (18a)

as well as their subgradients $g_i(\tilde{x}_i^k)$ and $-A_0/n - A_i \tilde{x}_i^k$;

3. Primal recovery step: compute in parallel the ergodic sum, for $k \geq 1$

$$x_i^k = \frac{1}{k} \sum_{t=1}^k \tilde{x}_i^t; \tag{18b}$$

4. Update the dual variables μ_i^k, \mathbf{G}_i^k as

$$\mu_i^{k+1} = P_{D_\mu} \left[\sum_{j \in V} [\mathbf{W}^\varphi]_{ij} (\mu_j^k + \alpha g_j(\tilde{x}_j^k)) \right] \tag{18c}$$

$$\mathbf{G}_i^{k+1} = P_{D_G} \left[\sum_{j \in V} [\mathbf{W}^\varphi]_{ij} (\mathbf{G}_j^k - \alpha(A_0/n + A_j \tilde{x}_j^k)) \right]. \tag{18d}$$

In particular, we have

$$\|\mu_i^k\|_2 \leq \Lambda < \infty, \quad \|\mathbf{G}_i^k\|_F \leq \Gamma < \infty. \tag{19}$$

Lemma 5.1 *Let $q(\mathbf{x}) : X \rightarrow \mathbb{R}$ be a concave function. Let the set $X \subset \mathbb{R}^n$ be convex and compact, and in particular $\max_{\mathbf{x} \in X} \|\mathbf{x}\|_2 \leq \eta$. There exist two finite scalars $\zeta > 0$ and $\tau > 0$ such that, for all $\mathbf{x} \in X$, for all $g(\mathbf{x}) \in \partial q_{\mathbf{x}}(\mathbf{x})$, and for all vectors $\mathbf{v} \in \mathbb{R}^n$ with $\|\mathbf{v}\|_2 \leq \tau$, the following holds*

$$g(\mathbf{x}) + \mathbf{v} \in \partial_\zeta q_{\mathbf{x}}(\mathbf{x}).$$

Proof The claim is proven by using the definition of subgradient of a concave function (1). Since q is a concave function, for all $\mathbf{x}, \mathbf{y} \in X, \mathbf{v} \in \mathbb{R}^n$,

$$\begin{aligned} q(\mathbf{y}) - q(\mathbf{x}) &\leq \langle g(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle = \langle g(\mathbf{x}) + \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle - \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \\ &\leq \langle g(\mathbf{x}) + \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle + \|\mathbf{v}\|_2 \|\mathbf{y} - \mathbf{x}\|_2 \leq \langle g(\mathbf{x}) + \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle + 2\tau\eta. \end{aligned}$$

For $\tau \leq \zeta/(2\eta)$, the claim follows. □

Lemma 5.2 *Let the initial dual variables in (18), μ_i^0 and \mathbf{G}_i^0 for all $i \in V$, be bounded. Let \mathbf{W} satisfy the conditions (17). Then, the following quantity is bounded by a certain $c_0 \geq 0$,*

$$\begin{aligned} &\left\| \sum_{j \in V} \left[\mathbf{W}^\varphi - \mathbf{1}_n \mathbf{1}_n^\top / n \right]_{ij} \left(\mu_j^0 + \alpha g_j(\tilde{x}_j^0) \right) \right\|_2 \\ &+ \left\| \sum_{j \in V} \left[\mathbf{W}^\varphi - \mathbf{1}_n \mathbf{1}_n^\top / n \right]_{ij} \left(\mathbf{G}_j^0 - \alpha(A_0/n + A_j \tilde{x}_j^0) \right) \right\|_F \leq c_0, \quad \forall i \in V. \end{aligned} \tag{20}$$

Proof The proof follows given the compactness of X and (therefore) the boundedness of the subgradients. □

We now present the main convergence results.

Theorem 5.1 (Dual variable agreement) *Let $\bar{\mu}^k, \bar{\mathbf{G}}^k$ be the mean values of the dual variables generated via the algorithm (18), i.e.,*

$$\bar{\mu}^k = \frac{1}{n} \sum_{i \in V} \mu_i^k, \quad \bar{\mathbf{G}}^k = \frac{1}{n} \sum_{i \in V} \mathbf{G}_i^k.$$

Let Assumptions 2.1 till 2.3 hold and let \mathbf{W} satisfy the conditions (17). Let μ_i^0 and \mathbf{G}_i^0 for $i \in V$ be bounded and let $\beta_0 \geq c_0$, with c_0 defined as in (20). Define L and Q as in (8) and let

$$M := L + Q, \quad p := \frac{v^\delta \beta_0}{\beta_0 + \alpha M}.$$

There exists a number of consensus iterations $\bar{\varphi}$, such that if $\varphi \geq \bar{\varphi} + \delta, \delta \geq 0, k \geq 1$, then the dual variables reach consensus as

$$\begin{aligned} \|\mu_i^{k+1} - \bar{\mu}^{k+1}\|_2 &\leq 2p^{k-1} v^\delta \beta_0 + 2p \alpha M \frac{1 - p^{k-1}}{1 - p}, \quad \forall i \in V, \\ \|\mathbf{G}_i^{k+1} - \bar{\mathbf{G}}^{k+1}\|_F &\leq 2p^{k-1} v^\delta \beta_0 + 2p \alpha M \frac{1 - p^{k-1}}{1 - p}, \quad \forall i \in V. \end{aligned}$$

Furthermore,

$$\bar{\varphi} = \frac{\log(\beta_0) - \log(4n(1 + d^2)(\beta_0 + \alpha M))}{\log(v)}.$$

Corollary 5.1 *Under the same conditions of Theorem 5.1, we obtain*

$$\lim_{k \rightarrow \infty} \|\mu_i^k - \bar{\mu}^k\|_2 \leq \frac{2p \alpha M}{1 - p}, \quad \lim_{k \rightarrow \infty} \|\mathbf{G}_i^k - \bar{\mathbf{G}}^k\|_F \leq \frac{2p \alpha M}{1 - p}, \quad \forall i \in V.$$

Theorem 5.1 and Corollary 5.1 specify how the consensus is reached among the nodes on the value of the dual variables while the algorithm (18) is running. Specifically, the consensus is reached exponentially fast to a steady-state bounded error floor. This bounded error depends on α (which can be tuned), and on p , which can also be tuned by varying φ . In particular, for $\varphi \rightarrow \infty$, due to the fact that $v < 1$ in conditions (17), then $p = 0$ and we obtain back the usual dual decomposition scheme with perfect agreement among the nodes.

Remark 5.1 Computing the lower bound on the number of consensus steps $\bar{\varphi}$ can be done during the initialization of the algorithm. We can always pick β_0 big enough so that $\beta_0 \gg \alpha M$, which means that $\bar{\varphi}$ can be simplified as $\bar{\varphi} = \frac{\log(1/(4n(1+d^2)))}{\log(v)}$, which can be determined in a distributed way [40].

Theorem 5.2 (Dual objective convergence) *Let μ^k, \mathbf{G}^k be the dual variables generated via the algorithm (18). Let μ_i^0 and \mathbf{G}_i^0 for all $i \in V$ be bounded and let β_0 be defined as in Theorem 5.1. Define L and Q as in (8) and let $M := L + Q$. Choose a scalar τ such that $\beta_0/\alpha \leq \tau$. Let ζ be defined as in Lemma 5.1 for the concave function $q(\mu, \mathbf{G})$ and the choice of τ . Let q^* be the optimal value of $q(\mu, \mathbf{G})$. Let Assumptions 2.1 till 2.3 hold and let \mathbf{W} satisfy the conditions (17). Let $\varphi \geq \bar{\varphi} + \delta, \delta \geq 0$ and let $\bar{\varphi}$ be defined as in Theorem 5.1. The following holds true. If $q^* = \infty$, then*

$$\limsup_{k \rightarrow \infty} q(\mu_i^k, \mathbf{G}_i^k) = \infty, \quad \forall i \in V,$$

If $q^ < \infty$, then*

$$\limsup_{k \rightarrow \infty} q(\mu_i^k, \mathbf{G}_i^k) \geq q^* - \alpha n(M + \tau)^2/2 - n(\beta_\infty(9M + 3\tau) + \zeta), \quad \forall i \in V,$$

with $\beta_\infty = \frac{p\alpha M}{1-p}$ and $p = \frac{v^\delta \beta_0}{\beta_0 + \alpha M}$.

Theorem 5.2 implies dual objective convergence up to a bounded error floor. Convergence is even more evident if we remember that owing to optimality, $q(\mu_i^k, \mathbf{G}_i^k) \leq q^*$, and therefore, if we define $q_i^\infty := \limsup_{k \rightarrow \infty} q(\mu_i^k, \mathbf{G}_i^k)$, we obtain

$$0 \geq q_i^\infty - q^* \geq -\alpha n(M + \tau)^2/2 - n(\beta_\infty(9M + 3\tau) + \zeta) =: -\varepsilon^2.$$

Note that the rightmost term $(-\varepsilon^2)$ represents a measure of sub-optimality of the approximate solution.

Theorem 5.3 (Primal objective convergence) *Let $\mu^k, \mathbf{G}^k, \mathbf{x}^k$ be the dual and primal variables generated via the algorithm (18). Let μ_i^0 and \mathbf{G}_i^0 for all $i \in V$ be bounded and let β_0 be defined as in Theorem 5.1. Define L and Q as in (8), Λ and Γ as in (19), and let $M := L + Q$. Choose a scalar τ such that $\beta_0/\alpha \leq \tau$. Let ζ be defined as in Lemma 5.1 for the concave function $q(\mu, \mathbf{G})$ and the choice of τ . Let f^* be the optimal value of $f(\mathbf{x})$. Let Assumptions 2.1 till 2.3 hold and let \mathbf{W} satisfy the conditions (17). Let $\varphi \geq \bar{\varphi} + \delta, \delta \geq 0$ and let $\bar{\varphi}$ be defined as in Theorem 5.1. The following holds true.*

(a) *An upper bound on the primal cost of the vector $\mathbf{x}^k, k \geq 1$, is given by*

$$f(\mathbf{x}^k) \leq f^* + \frac{\Lambda^2 + \Gamma^2}{2k\alpha/n} + e_k;$$

(b) *A lower bound on the primal cost of the vector $\mathbf{x}^k, k \geq 1$, is given by*

$$f(\mathbf{x}^k) \geq f^* - \frac{9(\Lambda^2 + \Gamma^2)}{2k\alpha/n} - e_k;$$

where

$$e_k = \frac{\alpha n(M + \tau)^2}{2} + n\tau(\Lambda + \Gamma) + n(\beta_0(6M + 3\tau) + \zeta).$$

Theorem 5.3 formulates convergence of the primal cost up to an error bound e_k . The rate of convergence is $O(1/k)$. We can also distinguish the error terms that come from the constant stepsize α and the terms that come from the finite number of consensus steps φ . In particular, we can write

$$e_k = \underbrace{\frac{\alpha nM^2}{2}}_{(1)} + \underbrace{\frac{\alpha n(2M\tau + \tau^2)}{2} + n\tau(\Lambda + \Gamma) + n(\beta_0(6M + 3\tau) + \zeta)}_{(2)},$$

and see that the term (1) is due to the constant stepsize, while the term (2) is due to the finite number of consensus steps. Furthermore, if $\varphi \rightarrow \infty$, then $c_0 = 0$, and we can set $\beta_0 = \tau = \zeta = 0$, yielding

$$\lim_{\varphi \rightarrow \infty} e_k = \frac{\alpha nM^2}{2}.$$

This is similar to the error level we obtain for the dual decomposition method in (12), and Theorem 3.1. Theorem 5.3 defines the main trade-offs in designing the algorithm’s parameters α and φ . The larger the stepsize α is, the faster the convergence is, even though the steady-state error becomes larger. If we increase φ , then the communication effort increases and the error e_k decreases.

6 Proof of Theorems 5.1 and 5.2

6.1 Preliminaries

We start our analysis by rewriting Step 4 of (18) in a more compact way. Let $z_i \in \mathbb{R}^{1+d^2}$ be the vector defined as $z_i := (\mu_i, \text{vec}(\mathbf{G}_i)^T)^T$, and let z_{sv} be the stacked vector of all the $z_i, i \in V$. Similarly, let $h_i(x)$ be the vector $h_i(x) := (g_i(x_i), \text{vec}(-\mathbf{A}_0/n - \mathbf{A}_i x_i)^T)^T$, and let $h_{sv}(x)$ the stacked vector of all the $h_i(x), i \in V$. Let Z be the convex set

$$Z := \left\{ z := \left(\mu, \text{vec}(\mathbf{G})^T \right)^T \in \mathbb{R}^{1+d^2} \mid \mu \in D_\mu, \mathbf{G} \in D_G \right\}, \tag{21}$$

and let $Z_{sv} = \prod_{i=1}^n Z$. The iterations in Step 4 of (18) can be rewritten as

$$z_{sv}^{k+1} = P_{Z_{sv}} \left[\mathbf{W}^\varphi \otimes \mathbf{I}_{1+d^2} \left(z_{sv}^k + \alpha h_{sv}(\tilde{x}^k) \right) \right]. \tag{22}$$

The iteration (22) represents a consensus-based subgradient method to maximize the dual function $q(\mu, \mathbf{G})$, i.e, the maximization problem

$$q^* := \max_{\mu \in D_\mu, \mathbf{G} \in D_G} \sum_{i \in V} q_i(\mu, \mathbf{G}) \equiv \max_{z \in Z} \sum_{i \in V} q_i(z), \quad \text{for } z = (\mu, \text{vec}(\mathbf{G})^\top)^\top.$$

In particular, (22) assigns to each node a copy of z , z_i , and enforces consensus among them. Furthermore, by (8), by triangle inequality, and by (19),

$$\|\mathbf{h}_i(\mathbf{x})\|_2 \leq \|h_i(\mathbf{x})\|_2 + \|\mathbf{Q}_i(\mathbf{x})\|_F = L + Q = M, \quad \|\mathbf{h}_{sv}(\mathbf{x})\|_2 \leq nM, \quad (23a)$$

$$\max_{z \in Z} \|z\|_2 \leq \sqrt{\Lambda^2 + \Gamma^2} \leq \Lambda + \Gamma. \quad (23b)$$

Lemma 6.1 ([18, Lemma 1]) *Let $\mathbf{x}_i \in \mathbb{R}^m, i \in V$ be m -dimensional vectors. Let $\bar{\mathbf{x}}$ be the average value of $\mathbf{x}_i, i \in V$, i.e., $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i \in V} \mathbf{x}_i$. The following basic relations hold,*

- (a) *if $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \beta, \forall i, j \in V$, then $\|\mathbf{x}_i - \bar{\mathbf{x}}\|_2 \leq \frac{n-1}{n} \beta$;*
- (b) *if $\|\mathbf{x}_i - \bar{\mathbf{x}}\|_2 \leq \beta, \forall i \in V$, then $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq 2\beta$.*

Lemma 6.2 ([18, Lemma 2]) *Let $\mathbf{x}^k \in \mathbb{R}^n$ be an n -dimensional vector, with components $x_i \in \mathbb{R}, i = 1, \dots, n$. Let $\mathbf{x}^{k+1} = \mathbf{W}^\varphi \mathbf{x}^k$, with $\mathbf{W} \in \mathbb{R}^{n \times n}$ fulfilling conditions (17). Let $\|x_i^k - x_j^k\|_2 \leq \sigma$, for a bounded σ , and for all $i, j = 1, \dots, n$. Then $\|x_i^{k+1} - x_j^{k+1}\|_2 \leq 2v^\varphi n \sigma$ for all $i, j = 1, \dots, n$.*

Lemma 6.3 *Let $\{z_{sv}^k\}$ be generated by (22) under Assumptions 2.1 till 2.3. Let $\mathbf{v}_i^k \in \mathbb{R}^{1+d^2}$, for all $i \in V$ be defined as*

$$\mathbf{v}_i^k = \sum_{j \in V} [\mathbf{W}^\varphi]_{ij} \left(z_j^k + \alpha \mathbf{h}_j(\bar{\mathbf{x}}^k) \right),$$

and let $\bar{\mathbf{v}}^k$ be the average value of $\mathbf{v}_i^k, i \in V$, i.e., $\bar{\mathbf{v}}^k = \frac{1}{n} \sum_{i \in V} \mathbf{v}_i^k$. There exists a $\bar{\varphi} \geq 1$, such that if $\varphi \geq \bar{\varphi} + \delta$ with $\delta \geq 0$, then

$$\|\mathbf{v}_i^k - \bar{\mathbf{v}}^k\|_2 \leq \beta, \forall i \in V \implies \|\mathbf{v}_i^{k+1} - \bar{\mathbf{v}}^{k+1}\|_2 \leq v^\delta \beta, \forall i \in V, k \geq 0.$$

Proof The proof is an adaptation of [18, Lemma 3]. In particular, we can show that for all $i, j \in V$

$$\|\mathbf{v}_i^k - \bar{\mathbf{v}}^k\|_2 \leq \beta \implies \|\mathbf{v}_i^{k+1} - \mathbf{v}_j^{k+1}\|_2 \leq 4v^\varphi n(1 + d^2)(\beta + \alpha M). \quad (24)$$

Therefore, if we choose,

$$\varphi \geq \underbrace{\frac{\log(\beta) - \log(4n(1 + d^2)(\beta + \alpha M))}{\log(v)}}_{=:\bar{\varphi}} + \delta, \quad \delta \geq 0,$$

then

$$\|v_i^k - \bar{v}^k\|_2 \leq \beta, \forall i \in V \implies \|v_i^{k+1} - v_j^{k+1}\|_2 \leq v^\delta \beta, \forall i, j \in V,$$

and the claim follows from Lemma 6.1(a). In order to prove (24), we proceed as follows.

$$\begin{aligned} \|v_i^k - \bar{v}^k\|_2 \leq \beta, \forall i \in V &\stackrel{\text{Lemma 6.1}}{\implies} \|v_i^k - v_j^k\|_2 \leq 2\beta, \forall i, j \in V \\ \implies \|[v_i^k - v_j^k]_\ell\|_2 &\leq 2\beta, \forall i, j \in V, \ell = 1, \dots, 1 + d^2, \end{aligned}$$

where $[\cdot]_\ell$ extracts the ℓ -th component of a vector. Define

$$u_i^{k+1} = P_Z[v_i^k] + \alpha h_i(\bar{x}^{k+1}), \quad \forall i \in V.$$

Prior to the consensus, the distance between the iterates can be bounded as

$$\begin{aligned} \|u_i^{k+1} - u_j^{k+1}\|_2 &= \|P_Z[v_i^k] + \alpha h_i(\bar{x}^{k+1}) - P_Z[v_j^k] - \alpha h_j(\bar{x}^{k+1})\|_2 \\ &\leq \|P_Z[v_i^k] - P_Z[v_j^k]\|_2 + 2\alpha M \leq \|v_i^k - v_j^k\|_2 + 2\alpha M \leq 2(\beta + \alpha M), \end{aligned}$$

which also implies $\|[u_i^k - u_j^k]_\ell\|_2 \leq 2(\beta + \alpha M)$. Given that $z_i^{k+1} = P[v_i^k], \forall i$, after consensus, we have

$$\begin{aligned} \|v_i^{k+1} - v_j^{k+1}\|_2 &= \left\| \sum_{p \in V} [W^\varphi]_{ip} u_p^{k+1} - \sum_{p \in V} [W^\varphi]_{jp} u_p^{k+1} \right\|_2 \\ &\leq \sum_{\ell=1}^{1+d^2} \left\| \left[\sum_{p \in V} [W^\varphi]_{ip} u_p^{k+1} - \sum_{p \in V} [W^\varphi]_{jp} u_p^{k+1} \right]_\ell \right\|_2 \\ &= \sum_{\ell=1}^{1+d^2} \left\| \sum_{p \in V} [W^\varphi]_{ip} [u_p^{k+1}]_\ell - \sum_{p \in V} [W^\varphi]_{jp} [u_p^{k+1}]_\ell \right\|_2 \\ &= \sum_{\ell=1}^{1+d^2} \left\| [W^\varphi \tilde{u}_\ell^{k+1}]_i - [W^\varphi \tilde{u}_\ell^{k+1}]_j \right\|_2, \end{aligned} \tag{25}$$

where $\tilde{u}_\ell^{k+1} = ([u_1^{k+1}]_\ell, \dots, [u_n^{k+1}]_\ell)^\top$. As said $\|[u_i^k - u_j^k]_\ell\|_2 \leq 2(\beta + \alpha M)$ which means $\|[\tilde{u}_\ell^k]_i - [\tilde{u}_\ell^k]_j\|_2 \leq 2(\beta + \alpha M)$. Thus, by using Lemma 6.2, we can bound (25) as

$$\|v_i^{k+1} - v_j^{k+1}\|_2 \leq \sum_{\ell=1}^{1+d^2} \left\| [W^\varphi \tilde{u}_\ell^{k+1}]_i - [W^\varphi \tilde{u}_\ell^{k+1}]_j \right\|_2 \leq 4v^\varphi n(1 + d^2)(\beta + \alpha M),$$

which is the rightmost term in (24) and the claim is proven. □

6.2 Proof of Theorem 5.1

The quantity $\|v_i^0 - \bar{v}^0\|_2$ is upper bounded by $\beta_0 \geq c_0$ by Lemma 5.2 (inequality (20)), thus, $\|v_i^0 - \bar{v}^0\|_2 \leq \beta_0$. Let us choose $\varphi \geq \bar{\varphi} + \delta$, $\delta \geq 0$, with $\bar{\varphi}$ determined as in Theorem 5.1. Then, by Lemma 6.3 and (24), it follows that,

$$\begin{aligned} \|v_i^1 - \bar{v}^1\|_2 &\leq v^\delta \beta_0 \\ \|v_i^2 - \bar{v}^2\|_2 &\leq 4v^\varphi n(1 + d^2)(v^\delta \beta_0 + \alpha M) = v^\delta \beta_0 \frac{v^\delta \beta_0 + \alpha M}{\beta_0 + \alpha M} \\ \|v_i^3 - \bar{v}^3\|_2 &\leq v^\delta \beta_0 \frac{v^\delta \beta_0}{\beta_0 + \alpha M} \left(\frac{v^\delta \beta_0 + \alpha M}{\beta_0 + \alpha M} + \alpha M \right) \\ \|v_i^k - \bar{v}^k\|_2 &\leq v^\delta \beta_0 \left(\frac{v^\delta \beta_0}{\beta_0 + \alpha M} \right)^{k-1} + \alpha M \left(-1 + \sum_{t=0}^{k-1} \left(\frac{v^\delta \beta_0}{\beta_0 + \alpha M} \right)^t \right). \end{aligned}$$

Let $p := \frac{v^\delta \beta_0}{\beta_0 + \alpha M}$, since $p < 1$, then

$$\|v_i^k - \bar{v}^k\|_2 \leq p^{k-1} v^\delta \beta_0 + p \alpha M \frac{1 - p^{k-1}}{1 - p} =: \beta_k, \quad k \geq 1 \tag{26}$$

and by Lemma 6.1(b), we derive $\|v_i^k - v_j^k\|_2 \leq 2\beta_k$.

By using the nonexpansive property of the projection operator, since $z_i^{k+1} = P[v_i^k]$, for all i , we can write

$$\|z_i^{k+1} - z_j^{k+1}\|_2 \leq \|v_i^k - v_j^k\|_2 \leq 2\beta_k, \quad k \geq 1, \tag{27}$$

and by Lemma 6.1(a) the claim follows.

6.3 Proof of Theorem 5.2

We define an average value for z_{sv}^k as $\bar{z}^k = \frac{1}{n} \sum_{i \in V} z_i^k$. For convergence purposes, we need to keep track of the difference $\bar{z}^{k+1} - P_Z[\bar{v}^k]$, and thus we define the vectors $y^k \in \mathbb{R}^{1+d^2}$ and $d^k \in \mathbb{R}^{1+d^2}$ as

$$y^k := P_Z[\bar{v}^{k-1}], \quad d^k := \bar{z}^k - y^k, \quad k \geq 1. \tag{28}$$

The main idea of the proof is to show that y is updated via an approximate ϵ -subgradient method and, then, by using [41, Proposition 4.1], the theorem follows. The first part is formalized in the following lemma.

Lemma 6.4 *Let y^k be defined as in (28). Under the same conditions of Theorem 5.2, for all $k \geq 1$,*

- (a) *The quantity $\|d^k/\alpha\|_2$ is upper bounded by $\beta_{k-1}/\alpha \leq \tau$ (where β_k is defined in (26));*

(b) The following inequalities are true, for all $i \in V$

$$q(\mathbf{y}^k) \leq q(\mathbf{z}_i^k) + 3nM\beta_{k-1} \tag{29}$$

$$q_i(\mathbf{y}) \leq q_i(\mathbf{y}^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}, \mathbf{y} - \mathbf{y}^k \rangle + \epsilon_k/n, \quad \forall \mathbf{y} \in Z. \tag{30}$$

(c) The quantity $\mathbf{g}(\tilde{\mathbf{x}}^k) := \sum_{i \in V} \left(\mathbf{h}_i(\tilde{\mathbf{x}}^k) + \frac{\mathbf{d}^k}{\alpha} \right)$ is an ϵ_k -subgradient of $q(\mathbf{y}^k)$ with respect to \mathbf{y} .

(d) The variable \mathbf{y}^k is updated via an ϵ -subgradient method

$$\mathbf{y}^{k+1} = \mathbf{P}_Z \left[\mathbf{y}^k + \frac{\alpha}{n} \mathbf{g}(\tilde{\mathbf{x}}^k) \right], \quad \mathbf{g}(\tilde{\mathbf{x}}^k) \in \partial_{\epsilon_k} q(\mathbf{y}^k). \tag{31}$$

And $\epsilon_k = n(\beta_{k-1}(6M + 3\tau) + \zeta)$.

Proof (a) We start by bounding $\|\mathbf{d}^k\|_2$,

$$\|\mathbf{d}^k\|_2 = \frac{1}{n} \left\| \sum_{i \in V} \left(\mathbf{P}_Z[\mathbf{v}_i^{k-1}] - \mathbf{P}_Z[\tilde{\mathbf{v}}^{k-1}] \right) \right\|_2 \leq \frac{1}{n} \sum_{i \in V} \|\mathbf{v}_i^{k-1} - \tilde{\mathbf{v}}^{k-1}\|_2 \leq \beta_{k-1},$$

where we have used the inequality (26) to bound the term $\|\mathbf{v}_i^{k-1} - \tilde{\mathbf{v}}^{k-1}\|_2$.

(b) Since $\mathbf{y}^k \in Z$ and $\mathbf{z}_i^k \in Z$, by the concavity of $q_i(\mathbf{z})$ and the definition of subgradient of a concave function (1), we can write for all $i, j \in V$

$$\begin{aligned} q_j(\mathbf{y}^k) &\leq q_j(\mathbf{z}_i^k) + \langle \mathbf{h}, \mathbf{y}^k - \mathbf{z}_i^k \rangle, \quad \text{where } \mathbf{h} \in \partial_{q_j, \mathbf{z}}(\mathbf{z}_i^k) \\ &\leq q_j(\mathbf{z}_i^k) + \|\mathbf{h}\|_2 \|\mathbf{z}_i^k - \mathbf{y}^k\|_2 \leq q_j(\mathbf{z}_i^k) + M(\|\mathbf{z}_i^k - \tilde{\mathbf{z}}^k\|_2 + \|\mathbf{d}^k\|_2) \\ &\leq q_j(\mathbf{z}_i^k) + M(2\beta_{k-1} + \beta_{k-1}) \leq q_j(\mathbf{y}^k) + 3M\beta_{k-1}. \end{aligned}$$

In particular, we have used the fact that any subgradient vector of $q_j(\mathbf{z})$ is bounded by M (23a), and inequality (27). If we sum the last relation over $j \in V$, we obtain (29). In addition for any $\mathbf{y} \in Z$, by using Lemma 5.1

$$\begin{aligned} q_i(\mathbf{y}) &\leq q_i(\mathbf{z}_i^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{z}_i^k \rangle \leq q_i(\mathbf{z}_i^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}, \mathbf{y} - \mathbf{z}_i^k \rangle + \zeta \\ &\leq q_i(\mathbf{y}^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}, \mathbf{y} - \mathbf{z}_i^k \rangle + 3M\beta_{k-1} + \zeta \\ &= q_i(\mathbf{y}^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}, \mathbf{y} - \mathbf{y}^k + \mathbf{y}^k - \mathbf{z}_i^k \rangle + 3M\beta_{k-1} + \zeta \\ &\leq q_i(\mathbf{y}^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}, \mathbf{y} - \mathbf{y}^k \rangle + \|\mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}\|_2 \|\mathbf{y}^k - \mathbf{z}_i^k\|_2 + 3M\beta_{k-1} + \zeta. \end{aligned}$$

We use the fact that $\|\mathbf{v}\|_2 \leq \tau$ by construction in Lemma 5.1, $\|\mathbf{h}_i(\tilde{\mathbf{x}}^k)\|_2 \leq M$ by (23a), $\|\mathbf{z}_i^k - \tilde{\mathbf{z}}^k\|_2 \leq 2\beta_{k-1}$ by (27), and $\|\mathbf{d}^k\|_2 \leq \beta_{k-1}$ by the preceding proof. By using these inequalities, we can bound

$$\|\mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{v}\|_2 \leq M + \tau, \quad \|\mathbf{y}^k - \mathbf{z}_i^k\|_2 = \|\mathbf{z}_i^k - \tilde{\mathbf{z}}^k + \mathbf{d}^k\|_2 \leq 3\beta_{k-1},$$

and we obtain

$$q_i(\mathbf{y}) \leq q_i(\mathbf{y}^k) + \langle \mathbf{h}_i(\tilde{\mathbf{x}}) + \mathbf{v}, \mathbf{y} - \mathbf{y}^k \rangle + (\beta_{k-1}(6M + 3\tau) + \zeta),$$

which is (30).

(c) By using the definition of subdifferential (1), the inequality (30) implies $(\mathbf{h}_i(\tilde{\mathbf{x}}) + \mathbf{v}) \in \partial_{\epsilon_k/n} q_{i,\mathbf{y}}(\mathbf{y})$ with $\epsilon_k/n = (\beta_{k-1}(6M + 3\tau) + \zeta)$. Summation over i yields,

$$q(\mathbf{y}) \leq q(\mathbf{y}^k) + \left\langle \sum_{i \in V} \mathbf{h}_i(\tilde{\mathbf{x}}) + \mathbf{v}, \mathbf{y} - \mathbf{y}^k \right\rangle + n(\beta_{k-1}(6M + 3\tau) + \zeta),$$

for any \mathbf{v} , such that $\|\mathbf{v}\| \leq \tau$. Since $\|\mathbf{d}^k/\alpha\|_2 \leq \tau$ by construction, then we can choose $\mathbf{v} = \mathbf{d}^k/\alpha$, from which the claim follows.

(d) It is sufficient to write explicitly the update rule for \mathbf{y}^k . Starting from the definition of \mathbf{y}^{k+1} in (28) and the definition of \mathbf{v}_i^k in Lemma 6.3, we obtain

$$\begin{aligned} \mathbf{y}^{k+1} &= \mathbf{P}_Z \left[\frac{1}{n} \sum_{i \in V} \mathbf{v}_i^k \right] = \mathbf{P}_Z \left[\frac{1}{n} \sum_{i \in V} \sum_{j \in V} [\mathbf{W}^\varphi]_{ij} \left(\mathbf{z}_j^k + \alpha \mathbf{h}_j(\tilde{\mathbf{x}}^k) \right) \right] \\ &= \mathbf{P}_Z \left[\frac{1}{n} \sum_{i \in V} \left(\mathbf{z}_i^k + \alpha \mathbf{h}_i(\tilde{\mathbf{x}}^k) \right) \right] = \mathbf{P}_Z \left[\mathbf{y}^k + \mathbf{d}^k + \frac{\alpha}{n} \sum_{i \in V} \mathbf{h}_i(\tilde{\mathbf{x}}^k) \right] \\ &= \mathbf{P}_Z \left[\mathbf{y}^k + \frac{\alpha}{n} \left(\sum_{i \in V} \mathbf{h}_i(\tilde{\mathbf{x}}^k) + \frac{\mathbf{d}^k}{\alpha} \right) \right]. \end{aligned}$$

Given part (c) of this Lemma, the claim follows. □

Proof (of Theorem 5.2) By Lemma 6.4, the sequence $\{\mathbf{y}^k\}$ is generated via an ϵ_k subgradient algorithm to maximize $q(\mathbf{y})$. And in particular, $k \geq 1$

$$\mathbf{y}^{k+1} = \mathbf{P}_Z \left[\mathbf{y}^k + \alpha/n \mathbf{g}(\tilde{\mathbf{x}}^k) \right], \quad \|\mathbf{g}(\tilde{\mathbf{x}}^k)\|_2 \leq n(M + \tau).$$

Therefore, we can use any standard result on the convergence of approximate subgradient algorithms. For example, by using [41, Proposition 4.1] (with $m = 1$), the following holds for the sequence $\{\mathbf{y}^k\}$,

If $q^* = \infty$, then

$$\limsup_{k \rightarrow \infty} q(\mathbf{y}^k) = \infty,$$

If $q^* < \infty$, then

$$\limsup_{k \rightarrow \infty} q(\mathbf{y}^k) \geq q^* - \alpha n(M + \tau)^2/2 - n(\beta_\infty(6M + 3\tau) + \zeta),$$

where $\beta_\infty = \lim_{k \rightarrow \infty} \beta_{k-1}$. Then, from the inequality (29), the claim is proven. □

7 Primal Recovery: Proof of Theorem 5.3

7.1 Some Basic Facts

Lemma 7.1 *Let \mathbf{y}^k be defined as (28). Under the same assumptions and notation of Theorem 5.2,*

(a) *For any $\mathbf{y} \in Z$,*

$$\sum_{t=1}^k \langle \mathbf{g}(\tilde{\mathbf{x}}^t), \mathbf{y} - \mathbf{y}^t \rangle \leq \frac{\|\mathbf{y}^1 - \mathbf{y}\|_2^2}{2\alpha/n} + k \frac{\alpha n(M + \tau)^2}{2};$$

(b) *For any $\mathbf{y} \in Z$,*

$$\sum_{t=1}^k \langle \mathbf{g}(\tilde{\mathbf{x}}^t), \mathbf{y} - \mathbf{y}^* \rangle \leq \frac{\|\mathbf{y}^1 - \mathbf{y}\|_2^2}{2\alpha/n} + k \frac{\alpha n(M + \tau)^2}{2} + \sum_{t=1}^k \epsilon_t,$$

where $\epsilon_t = n(\beta_{t-1}(6M + 3\tau) + \zeta)$.

Proof We start from the update rule (31). For any $\mathbf{y} \in Z$,

$$\begin{aligned} \|\mathbf{y}^{k+1} - \mathbf{y}\|_2^2 &= \left\| \mathbf{P}_Z \left[\mathbf{y}^k + \frac{\alpha}{n} \mathbf{g}(\tilde{\mathbf{x}}^k) \right] - \mathbf{P}_Z[\mathbf{y}] \right\|_2^2 \leq \left\| \mathbf{y}^k + \frac{\alpha}{n} \mathbf{g}(\tilde{\mathbf{x}}^k) - \mathbf{y} \right\|_2^2 \\ &\leq \|\mathbf{y}^k - \mathbf{y}\|_2^2 + \frac{2\alpha}{n} \langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y}^k - \mathbf{y} \rangle + \alpha^2(M + \tau)^2. \end{aligned}$$

where we use the fact that $\|\mathbf{g}(\tilde{\mathbf{x}}^k)\|_2 = \|\sum_{i \in V} (\mathbf{h}_i(\tilde{\mathbf{x}}^k) + \mathbf{d}^k/\alpha)\|_2 \leq n(M + \tau)$. Therefore, for any $\mathbf{y} \in Z$

$$\langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{y}^k \rangle \leq \frac{\|\mathbf{y}^k - \mathbf{y}\|_2^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|_2^2}{2\alpha/n} + \frac{\alpha n(M + \tau)^2}{2}, \tag{32}$$

and by summing over k , part (a) follows. Since $\mathbf{g}(\tilde{\mathbf{x}}^k)$ is an ϵ_k -subgradient of the dual function q at \mathbf{y}^k , using the subgradient inequality (1),

$$\langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y}^k - \mathbf{y}^* \rangle \leq q(\mathbf{y}^k) - q(\mathbf{y}^*) + \epsilon_k \leq \epsilon_k,$$

where the last inequality comes from the optimality condition $q(\mathbf{y}^k) \leq q(\mathbf{y}^*)$, which is valid for any $\mathbf{y}^k \in Z$. In particular, ϵ_k is defined in Lemma 6.4(c). We then have

$$\langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{y}^* \rangle = \langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{y}^k \rangle + \langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y}^k - \mathbf{y}^* \rangle \leq \langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{y}^k \rangle + \epsilon_k.$$

From the preceding relation and (32), we obtain

$$\langle \mathbf{g}(\tilde{\mathbf{x}}^k), \mathbf{y} - \mathbf{y}^* \rangle \leq \frac{\|\mathbf{y}^k - \mathbf{y}\|_2^2 - \|\mathbf{y}^{k+1} - \mathbf{y}\|_2^2}{2\alpha/n} + \frac{\alpha n(M + \tau)^2}{2} + \epsilon_k, \quad k \geq 1$$

and summing over k part (b) follows as well. In particular, we remark that $\mathbf{y}^1 = \mathbf{P}_Z[\bar{\mathbf{v}}^0]$, which is bounded, since Z is a compact set. □

7.2 Proof of Theorem 5.3(a)

Proof By convexity of the primal cost $f(\mathbf{x})$ and the definition of \tilde{x}_i^k as a minimizer of the local Lagrangian functions over $x_i \in X_i$, we have,

$$f(\mathbf{x}^k) \leq \frac{1}{k} \sum_{t=1}^k f(\tilde{\mathbf{x}}^t) = \frac{1}{k} \sum_{t=1}^k \sum_{i \in V} \left(q_i(z_i^t) - \langle z_i^t, \mathbf{h}_i(\tilde{\mathbf{x}}^t) \rangle \right), \quad k \geq 1. \tag{33}$$

By Lemma 6.4 inequality (30) with $\mathbf{y} = z_i^t \in Z$,

$$q_i(z_i^t) - q_i(\mathbf{y}^t) \leq \langle \mathbf{h}_i(\tilde{\mathbf{x}}^t), z_i^t \rangle + \langle \mathbf{v}, z_i^t \rangle - \langle \mathbf{h}_i(\tilde{\mathbf{x}}^t) + \mathbf{v}, \mathbf{y}^t \rangle + \epsilon_t/n,$$

with $\epsilon_t/n = \beta_{t-1}(6M + 3\tau) + \zeta$. Summing over $i \in V$,

$$\sum_{i \in V} q_i(z_i^t) \leq q(\mathbf{y}^t) + \sum_{i \in V} \langle \mathbf{h}_i(\tilde{\mathbf{x}}^t), z_i^t \rangle + \sum_{i \in V} \langle \mathbf{v}, z_i^t \rangle - \langle \mathbf{g}(\tilde{\mathbf{x}}^t), \mathbf{y}^t \rangle + \epsilon_t,$$

hence,

$$f(\mathbf{x}^k) \leq \frac{1}{k} \sum_{t=1}^k \left(q(\mathbf{y}^t) + \sum_{i \in V} \langle \mathbf{v}, z_i^t \rangle - \langle \mathbf{g}(\tilde{\mathbf{x}}^t), \mathbf{y}^t \rangle + \epsilon_t \right). \tag{34}$$

We can use Lemma 7.1(a) with $\mathbf{y} = \mathbf{0} \in Z$ to upper bound $-\langle \mathbf{g}(\tilde{\mathbf{x}}^t), \mathbf{y}^t \rangle$, while we bound $\|\langle \mathbf{v}, z_i^t \rangle\|_2$ as $\|\langle \mathbf{v}, z_i^t \rangle\|_2 \leq \tau(\Lambda + \Gamma)$. The latter bound comes from the fact that by construction $\|\mathbf{v}\|_2 \leq \tau$, and $\|z_i^t\|_2 \leq \Lambda + \Gamma$ by (23a). With this in place, we can write (34) as

$$f(\mathbf{x}^k) \leq \frac{1}{k} \sum_{t=1}^k q(\mathbf{y}^t) + n\tau(\Lambda + \Gamma) + \frac{\|\mathbf{y}^1\|_2^2}{2k\alpha/n} + \frac{\alpha n(M + \tau)^2}{2} + \frac{1}{k} \sum_{t=1}^k \epsilon_t.$$

If we now compute

$$\frac{1}{k} \sum_{t=1}^k \epsilon_t = \frac{1}{k} \sum_{t=1}^k n(\beta_{t-1}(6M + 3\tau) + \zeta) \leq n(\beta_0(6M + 3\tau) + \zeta), \tag{35}$$

and remember that by optimality $q(\mathbf{y}^t) \leq q^*$, $q^* = f^*$ by strong duality (Assumption 2.3), and $\|\mathbf{y}^1\|_2^2 \leq \Lambda^2 + \Gamma^2$, then the claim follows. □

7.3 Proof of Theorem 5.3.(b)

Proof Given any dual optimal solution \mathbf{y}^* , we have

$$f(\mathbf{x}^k) = \underbrace{f(\mathbf{x}^k) + \left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \mathbf{g}(\tilde{\mathbf{x}}^t) \right\rangle}_{(a)} - \left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \mathbf{g}(\tilde{\mathbf{x}}^t) \right\rangle. \tag{36}$$

We also know that,

$$\begin{aligned} (a) &= f(\mathbf{x}^k) + \left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \sum_{i \in V} \mathbf{h}_i(\tilde{\mathbf{x}}^t) \right\rangle + n \left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \mathbf{d}^t / \alpha \right\rangle \\ &\geq f(\mathbf{x}^k) + \left\langle \mathbf{y}^*, \sum_{i \in V} \mathbf{h}_i(\mathbf{x}^k) \right\rangle - n(\Lambda + \Gamma)\tau, \end{aligned} \tag{37}$$

where we used the fact that $\mathbf{h}_i(\tilde{\mathbf{x}}^t)$ is a convex function of $\tilde{\mathbf{x}}^t$ and therefore,

$$\frac{1}{k} \sum_{t=1}^k \sum_{i \in V} \mathbf{h}_i(\tilde{\mathbf{x}}^t) \geq \sum_{i \in V} \mathbf{h}_i(\mathbf{x}^k),$$

and the Cauchy–Schwarz inequality to bound

$$\left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \mathbf{d}^t / \alpha \right\rangle \geq -\|\mathbf{y}^*\|_2 \left\| \frac{1}{k} \sum_{t=1}^k \mathbf{d}^t / \alpha \right\|_2 \geq -\tau(\Lambda + \Gamma).$$

Furthermore, by the saddle point property of the Lagrangian function, i.e., for any $\mathbf{x} \in X, \mathbf{y} \in Z$

$$L(\mathbf{x}^*, \mathbf{y}) \leq L(\mathbf{x}^*, \mathbf{y}^*) \leq L(\mathbf{x}, \mathbf{y}^*),$$

and the fact that under strong duality (Assumption 2.3) $L(\mathbf{x}^*, \mathbf{y}^*) = q^* = f^*$, we can write

$$f(\mathbf{x}^k) + \left\langle \mathbf{y}^*, \sum_{i \in V} \mathbf{h}_i(\mathbf{x}^k) \right\rangle - n\tau(\Lambda + \Gamma) = L(\mathbf{x}^k, \mathbf{y}^*) - n\tau(\Lambda + \Gamma) \geq f^* - n\tau(\Lambda + \Gamma). \tag{38}$$

We can now upper bound $\left\langle \mathbf{y}^*, \frac{1}{k} \sum_{t=1}^k \mathbf{g}(\tilde{\mathbf{x}}^t) \right\rangle$ in (36) as in Lemma 7.1(b), with $\mathbf{y} = 2\mathbf{y}^* \in Z$ (by the definition of r). By substituting this bound in (36) and by combining it with (37) and (38), we get

$$f(\mathbf{x}^k) \geq f^* - n\tau(\Lambda + \Gamma) - \frac{\|\mathbf{y}^1 - 2\mathbf{y}^*\|_2^2}{2k\alpha/n} - \frac{\alpha n(M + \tau)^2}{2} - \frac{1}{k} \sum_{t=1}^k \epsilon_t.$$

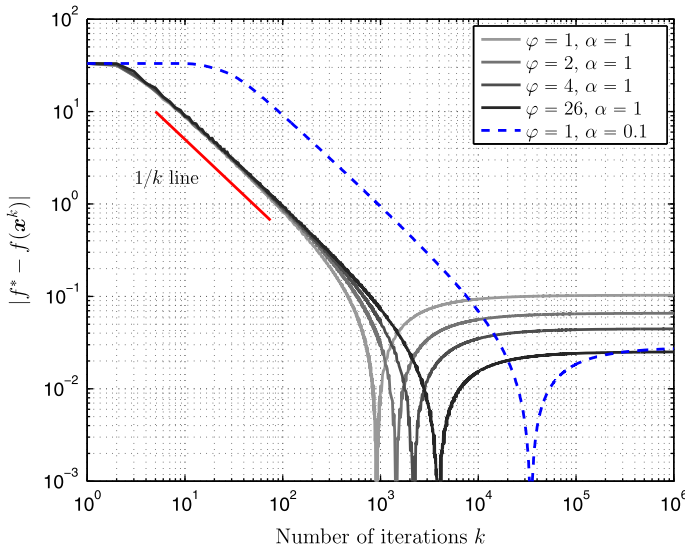


Fig. 1 Convergence of the proposed algorithm for different choices of stepsize α and number of consensus step φ

From the upper bound (35), and $\|y^1 - 2y^*\|_2^2 = \|y^1\|_2^2 + 4\|y^1\|_2\|y^*\|_2 + 4\|y^*\|_2^2$, which can be upper bounded as $9(\Lambda^2 + \Gamma^2)$, the claim follows. \square

8 Numerical Results

In this section, we present some numerical results to assess the proposed algorithm for different φ values in comparison with the standard dual decomposition. We choose the following simple yet representative sample problem,

$$\underset{\substack{x_i \in [0,1] \\ i \in \{1, \dots, 100\}}}{\text{minimize}} f(x) := - \sum_{i=1}^{33} \sigma_i x_i - \sum_{i=34}^{100} \sigma_i \log(1 + x_i), \text{ subject to } \sum_{i=1}^{100} \sigma_i x_i \leq 10,$$

where each $\sigma_i \in [0, 1]$ is drawn from a uniform random distribution. This type of problem has been considered, e.g., in network utility maximization contexts [23]. We solve the problem in Matlab with Yalmip and SDPT3 [42,43], where we also implement the proposed algorithm.¹

For this problem a Slater vector is $x_i = 0$ for all i ; furthermore $\gamma = 10$, while $q(0)$ is solvable by inspection ($x_i = 1$) and gives (for our realization of σ_i) $r = 8.62$. The communication network is randomly selected, and the average number of neighbors is 3.12.

Figure 1 depicts convergence, and it is in line with our theoretical findings: The error decreases as $O(1/k)$ till it reaches a bounded error floor. This bounded error

¹ The code is available at: <http://ens.ewi.tudelft.nl/~asimonetto/NumericalExample.zip>.

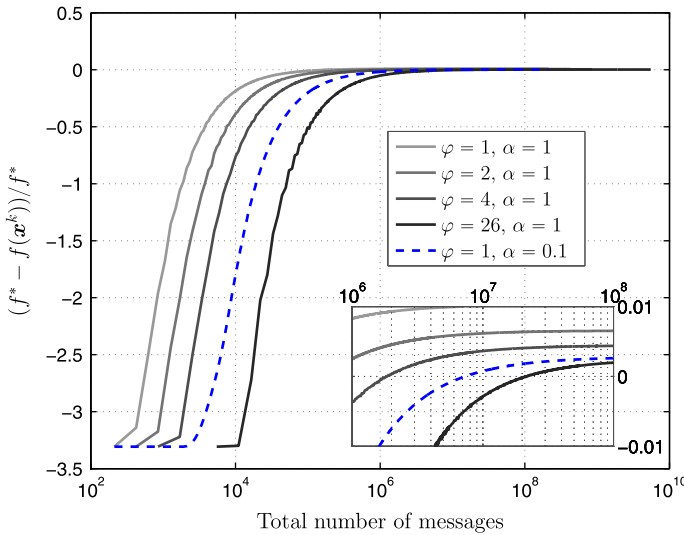


Fig. 2 Relative error and number of exchanged messages for different choices of stepsize α and number of consensus step φ

floor depends on both φ and α as captured in Theorem 5.3. We have also plotted the performance of the standard dual decomposition, which (in the absence of a master node) requires reaching complete consensus at each iteration (in theory $\varphi \rightarrow \infty$, but we have set $\varphi = 26$, which yields a full W^φ).

Figure 2 shows the relative error with respect to the total number of messages the nodes are exchanging. We can see that, in the absence of a master node, the proposed consensus-based algorithm involves significantly fewer number of messages than the standard dual decomposition for the same accuracy level (till up to 1 % error). This is very important in real-life applications.

9 Future Research Questions

Future research encompasses the following points.

First of all, we have used the ergodic mean to recover the primal solution. The reason for it is mainly technical: It helps to derive convergence rate results, via a telescopic cancelation argument. Other convex combinations have been advocated, e.g., in [12], but the results they can offer are typically asymptotical and require vanishing stepsizes. An open question is whether other combinations for primal recovery are possible using constant stepsizes.

Then, in the derivation, we have limited ourselves to objective convergence. It would be relevant to investigate convergence of the ergodic mean to the optimizer set, either in the general convex case or in the strong convex scenario.

Finally, the bound on φ , i.e., $\bar{\varphi}$ has been derived in such a way that we could use ϵ -subgradient arguments in the rest of the convergence proofs. However, it is quite conservative (in fact, in practice, φ can be as small as 1, but this is often not captured

by the bound in Theorem 5.1). This is due to Lemma 6.2 and the use of the spectral radius as an upper bound. A thorough investigation is left for future research.

10 Conclusions

A consensus-based dual decomposition scheme has been proposed to enable a network of collaborative computing nodes to generate approximate dual and primal solutions of a distributed convex optimization problem. We have proven convergence of the scheme both in the dual and the primal objective senses up to a bounded error floor. The proposed scheme is of theoretical and applied importance since it eliminates the need for a centralized entity (i.e., a master node) to collect the local subgradient information, by distributing this task among the nodes. This need has been a major hurdle in the use of dual decomposition for solving certain classes of distributed optimization problems.

Acknowledgments This work was supported in part by STW under the D2S2 project from the ASSYS program (10561) and in part by NWO-STW under the VICI program (10382).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Bertsekas, D.P.: *Nonlinear Programming*, 2nd edn. Athena Scientific (1999)
2. Johansson, B.: *On Distributed Optimization in Networked Systems*. Ph.D. thesis, KTH, Stockholm, Sweden (2008)
3. Boyd, S., Xiao, L., Mutapcic, A., Mattingley, J.: *Notes on Decomposition Methods*. Tech. rep., Stanford University (2008)
4. Nedić, A., Ozdaglar, A.: Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM J. Optim.* **19**(4), 1757–1780 (2009)
5. Polyak, B.T.: *Introduction to Optimization*. Optimization Software Inc (1987)
6. Kiwiel, K.: Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM J. Optim.* **14**(3), 807–840 (2004)
7. Nemirovskii, A.S., Yudin, D.B.: Cesaro convergence of gradient method approximation of saddle points for convex-concave functions. *Doklady Akademii Nauk SSSR* **239**, 1056–1059 (1978)
8. Shor, N.Z.: *Minimization Methods for Nondifferentiable Functions*. Springer, Berlin (1985)
9. Sherali, H.D., Choi, G.: Recovery of primal solutions when using subgradient optimization methods to solve lagrangian duals of linear programs. *Oper. Res. Lett.* **19**, 105–113 (1996)
10. Larsson, T., Patriksson, M., Stömberg, A.B.: Ergodic primal convergence in dual subgradient schemes for convex programming. *Math. Program.* **86**, 283–312 (1999)
11. Ma, J.: *Recovery of Primal Solution in Dual Subgradient Schemes*. Master's thesis, MIT (2007)
12. Gustavsson, E., Patriksson, M., Strömberg, A.B.: Primal Convergence from Dual Subgradient Methods for Convex Optimization. *Math. Program.* **150**(2), 365–390 (2015)
13. Necoara, I., Nedelcu, V.: Rate analysis of inexact dual first-order methods: application to dual decomposition. *IEEE Trans. Autom. Control* **59**(5), 1232–1243 (2014)
14. Nedić, A., Ozdaglar, A.: Subgradient methods for saddle-point problems. *J. Optim. Theory Appl.* **142**(1), 205–228 (2009)
15. Jakubiec, F.Y., Ribeiro, A.: D-MAP: distributed maximum a posteriori probability estimation of dynamic systems. *IEEE Trans. Signal Process.* **61**(2), 450–466 (2013)
16. Simonetto, A., Leus, G.: Distributed asynchronous time-varying constrained optimization. In: *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, USA (2014)

17. Simonetto, A., Kester, L., Leus, G.: Distributed Time-Varying Stochastic Optimization and Utility-based Communication (2014). arxiv.org/abs/1408.5294
18. Johansson, B., Keviczky, T., Johansson, M., Johansson, K.H.: Subgradient Methods and Consensus Algorithms for Solving Convex Optimization Problems. In: Proceedings of the 47th IEEE Conference on Decision and Control, pp. 4185–4190. Cancun, Mexico (2008)
19. Srivastava, K., Nedić, A.: Distributed asynchronous constrained stochastic optimization. *IEEE Trans. Sel. Top. Signal Process.* **5**(4), 772–790 (2011)
20. Nedić, A., Ozdaglar, A., Parrilo, P.A.: Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Control* **55**(4), 922–938 (2010)
21. Duchi, J.C., Agarwal, A., Wainwright, M.: Dual averaging for distributed optimization: convergence analysis and network scaling. *IEEE Trans. Autom. Control* **57**(3), 592–606 (2012)
22. Bürger, M., Notarstefano, G., Allgöwer, F.: A polyhedral approximation framework for convex and robust distributed optimization. *IEEE Trans. Autom. Control* **59**(2), 384–395 (2014)
23. Palomar, D., Chiang, M.: A tutorial on decomposition methods for network utility maximization. *IEEE J. Sel. Areas Commun.* **24**(8), 1439–1451 (2006)
24. Wei, E., Ozdaglar, A., Jadbabaie, A.: A distributed newton method for network utility maximization. In: Proceedings of the 49th IEEE Conference on Decision and Control, pp. 1816–1821. Orlando, FL, USA (2010). doi:[10.1109/CDC.2010.5718026](https://doi.org/10.1109/CDC.2010.5718026)
25. Doan, M.D.: Distributed Model Predictive Controller Design Based on Distributed Optimization. Ph.D. thesis, Delft University of Technology (2012)
26. Jamali-Rad, H., Simonetto, A., Leus, G.: Sparsity-aware sensor selection: centralized and distributed algorithms. *IEEE Signal Process. Lett.* **21**(2), 217–220 (2014)
27. Joshi, S., Boyd, S.: Sensor selection via convex optimization. *IEEE Trans. Signal Process.* **57**(2), 451–462 (2009)
28. Necoara, I., Suykens, J.: Application of a smoothing technique to decomposition in convex optimization. *IEEE Trans. Autom. Control* **53**(11), 2674–2679 (2008)
29. Koshal, J., Nedić, A., Shanbhag, U.Y.: Multiuser optimization: distributed algorithms and error analysis. *SIAM J. Optim.* **21**(3), 1046–1081 (2011)
30. Lu, J., Johansson, M.: A Multi-agent projected dual gradient method with primal convergence guarantees. In: Proceedings of the 51st Allerton Conference, pp. 275–281. Monticello, IL, USA (2013)
31. Lu, J., Johansson, M.: Convergence analysis of approximate primal solutions in dual first-order methods. *Math. Program.* (submitted) (2013). [arXiv:1502.06368](https://arxiv.org/abs/1502.06368)
32. Necoara, I., Nedelcu, V.: Distributed Dual Gradient Methods and Error Bound Conditions (2014). [arXiv:1401.4398](https://arxiv.org/abs/1401.4398)
33. Simonetto, A., Leus, G.: Distributed maximum likelihood sensor network localization. *IEEE Trans. Signal Process.* **62**(6), 1424–1437 (2014)
34. Beck, A., Nedić, A., Ozdaglar, A., Teboulle, M.: An $O(1/k)$ gradient method for network resource allocation problems. *IEEE Trans. Control Netw. Syst.* **1**(1), 64–73 (2014)
35. Simonetto, A., Leus, G.: Double smoothing for time-varying distributed multi-user optimization. In: Proceedings of the IEEE Global Conference on Signal and Information Processing, Atlanta, USA (2014)
36. Wang, S.D., Kuo, T.S., Hsu, C.F.: The bounds on the solution of the algebraic Riccati and Lyapunov equations. *IEEE Trans. Autom. Control* **31**(7), 654–656 (1986)
37. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
38. Xiao, L., Boyd, S.: Fast linear iterations for distributed averaging. *Syst. Control Lett.* **53**(1), 65–78 (2003)
39. Xiao, L., Boyd, S.: Optimal scaling of a gradient method for distributed resource allocation. *J. Optim. Theory Appl.* **129**(3), 469–488 (2006)
40. Kempe, D., McSherry, F.: A decentralized algorithm for spectral analysis. *J. Comput. Syst. Sci.* **74**(1), 70–83 (2008)
41. Nedić, A.: Subgradient Methods for Convex Minimization. Ph.D. thesis, MIT (2002)
42. Löfberg, J.: YALMIP: A toolbox for modeling and optimization in MATLAB. In: Proceedings of the CACSD Conference, Taipei, Taiwan (2004)
43. Toh, K.C., Todd, M.J., Tutuncu, R.H.: SDPT3—a Matlab software package for semidefinite programming. *Optim. Methods Softw.* **11**, 545–581 (1999)