

DISTRIBUTED DETERMINISTIC ASYNCHRONOUS ALGORITHMS
IN TIME-VARYING GRAPHS THROUGH DYKSTRA SPLITTING*C. H. JEFFREY PANG[†]

Abstract. Consider a setting where each vertex of a graph has a function, and communications can only occur between vertices connected by an edge. We wish to minimize the sum of these functions. For the case when each function is the sum of a strongly convex quadratic and a convex function, we propose a distributed version of Dykstra's algorithm. The computations to optimize the dual objective function can run asynchronously without a global clock, and in a distributed manner without a central controller. Convergence to the primal minimizer is deterministic instead of being probabilistic, and is guaranteed as long as, in each cycle, the edges where two-way communications occur connect all vertices. We also look at an accelerated algorithm, and an algorithm for the case when the functions on the nodes are not strongly convex.

Key words. distributed optimization, averaged consensus, Dykstra's algorithm, time-varying graphs

AMS subject classifications. 68W15, 90C25, 90C30, 65K05

DOI. 10.1137/18M1168297

1. Introduction. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected and connected graph defined by the set of nodes (agents) \mathcal{V} and the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. Since \mathcal{G} is undirected, we assume that both (i, j) and (j, i) refer to the same edge when it exists.

Let X be a finite-dimensional Hilbert space. For a closed convex set C , let $\delta_C(\cdot)$ be the indicator function defined as $\delta_C(x) = 0$ if $x \in C$ and equal to ∞ otherwise. For each edge $(i, j) \in \mathcal{E}$, let the hyperplane $H_{(i,j)} \subset X^{|\mathcal{V}|}$ be defined by

$$(1.1) \quad H_{(i,j)} = \{(x_1, \dots, x_n) : x_i = x_j\}.$$

We consider the following problem throughout the rest of this paper.

Problem 1.1. Let $(\mathcal{V}, \mathcal{E})$ be a connected graph. Suppose $H_{(i,j)}$ is defined as in (1.1) for all $(i, j) \in \mathcal{E}$, and let $f_i : X \rightarrow \bar{\mathbb{R}}$ (where $\bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ throughout this paper) be closed convex functions for all $i \in \mathcal{V}$. Let $\mathbf{f}_i : X^{|\mathcal{V}|} \rightarrow \bar{\mathbb{R}}$ be defined by $\mathbf{f}_i(x) = f_i(x_i)$ (i.e., \mathbf{f}_i depends only on the i th variable). The primal problem of interest is

$$(1.2) \quad \min_{x \in X^{|\mathcal{V}|}} \sum_{(i,j) \in \mathcal{E}} \underbrace{\delta_{H_{(i,j)}}(x)}_{h_{(i,j)}(x)} + \sum_{i \in \mathcal{V}} \underbrace{\mathbf{f}_i(x)}_{h_i(x)}.$$

For each $\alpha \in \mathcal{E} \cup \mathcal{V}$, the function $h_\alpha : X^{|\mathcal{V}|} \rightarrow \bar{\mathbb{R}}$ is as marked in (1.2). Since $(\mathcal{V}, \mathcal{E})$ is connected, problem (1.2) is equivalent to

$$(1.3) \quad \min_{x \in X} \sum_{i \in \mathcal{V}} f_i(x),$$

*Received by the editors February 1, 2018; accepted for publication (in revised form) November 26, 2018; published electronically February 14, 2019.

<http://www.siam.org/journals/siop/29-1/M116829.html>

Funding: The author acknowledges grant R-146-000-214-112 from the Faculty of Science, National University of Singapore.

[†]Department of Mathematics, National University of Singapore (matpchj@nus.edu.sg, <http://www.math.nus.edu.sg/~matpchj>).

but we write it in the form (1.2) to emphasize that the only vertex which has knowledge of the function $f_i(\cdot)$ is vertex i .

1.1. Distributed algorithms for (1.2). We give a brief summary of distributed algorithms for minimizing (1.2). Some desirable properties for a distributed algorithm, especially when $|\mathcal{V}|$ is large, are the following.

- (1) The algorithm is applicable to directed graphs, where only one-way communication is allowed between two vertices connected by a directed edge.
- (2) The algorithm has deterministic convergence.
- (3) The algorithm is asynchronous. There is no need for a global clock, and each node can perform calculations at its own pace without being affected by other slower nodes.
- (4) The algorithm is distributed (i.e., in intermediate computations, each node only exchanges data with its neighbors) and decentralized (i.e., there is no central node connected to all other nodes to coordinate computations).
- (5) The algorithm allows for time-varying graphs.

We emphasize that the algorithm that we look at in this paper is only applicable to undirected graphs, and hence does not satisfy property (1). Nevertheless, we give a brief summary of the literature behind distributed algorithms for directed graphs in this paragraph. In the case where only one-way communication is allowed between two vertices connected by a directed edge, the survey [30] records many algorithms derived from the subgradient algorithm for solving (1.2). If the edges in a network are directed, it appears that the subgradient method is the only reasonable method. The subgradient method requires diminishing step sizes for convergence in the general case, which affects its convergence rates. More details of recent developments are given in [31]. A notable paper is [42]. The case of time-varying graphs was first studied in [32] and further extended in [33]. In time-varying graphs, the assumption needed for convergence is for the edge set \mathcal{E} to vary over time. But, if the edges are undirected, then alternative methods may be possible, and would usually be faster than subgradient methods. For strongly convex problems, linear convergence is possible. These algorithms appear to be synchronous, and require the functions involved to be smooth.

Two common methods for minimizing the sum of two convex functions are the alternating direction method of multipliers (ADMM) and Peaceman–Rachford algorithms (with the Douglas–Rachford algorithm being a special case of the latter). The Peaceman–Rachford algorithm is an example of a splitting method, and it is well known that the ADMM is dual to the Douglas–Rachford method [23]. In order to minimize the sum of more than two functions, the product space reformulation is a well-studied option. (See, for example, [11, Chapter 7].) Another strategy, which is a splitting method for the sum of more than two functions without using the product space reformulation, is given in [21, 16]. The latter development in [16] allows for lags in the collection of data for nodes where the computation time is greater, thus allowing for an asynchronous operation. Still, this algorithm requires a central controller, so it is different from the algorithms we consider in this paper.

We now look at asynchronous distributed algorithms with deterministic convergence (rather than probabilistic convergence). In some applications, the guarantees from deterministic convergence can outweigh the other advantages of algorithms with randomized convergence. We mention that the paper [25] and the extension [3] include algorithms that give deterministic convergence for strongly convex problems that are primal in nature, so these algorithms cannot handle more than one constraint set. The

method in [2] may arguably be considered to have these properties. Other than this, we are not aware of a decentralized, asynchronous algorithm that has deterministic convergence for (1.2) and is not a subgradient method.

However, a decentralized asynchronous probabilistic method derived from the ADMM is proposed in [28, 9]. The key idea in [28] is the introduction of randomized Gauss–Seidel iterations of the so-called Douglas–Rachford operator, and [9] extends [28] by incorporating the work of [46, 17]. This concept was generalized in [39]. All the works just mentioned use monotone operator theory (see, for example, the textbook [5]). Such algorithms require computations in the nodes to follow specific probability distributions, so they do not immediately seem to be applicable to the setting of time-varying graphs in [32] mentioned earlier. Another randomized, distributed method for nonconvex functions is [13].

The most similar randomized distributed method to that we discuss in this paper is that in [35]. We discuss this further in subsection 1.2.

As mentioned in the survey [30], the primal problem (1.3) has a dual

$$(1.4) \quad \begin{aligned} & \max_{\{y_i\}_{i \in \mathcal{V}} \subset X} - \sum_{i \in \mathcal{V}} f_i^*(y_i) \\ & \text{s.t.} \quad \sum_{i \in \mathcal{V}} y_i = 0, \end{aligned}$$

which is also known as the resource allocation problem.

1.2. A special case of (1.2) through Dykstra’s algorithm. When some of the functions in the primal problem of the form (1.3) are extended valued, it may be difficult to find a primal feasible point in the first place. We first look at the problem

$$(1.5) \quad \min_{x \in X^{|\mathcal{V}|}} \frac{1}{2} \|x - x_0\|^2 + \sum_{(i,j) \in \mathcal{E}} \delta_{H_{(i,j)}}(x) + \sum_{i \in \mathcal{V}} \mathbf{f}_i(x),$$

where $\|x\|^2 := \langle x, x \rangle = \sum_{i=1}^n \langle x_i, x_i \rangle$ and $x_0 \in X^{|\mathcal{V}|}$. This problem fits the framework of (1.2) since

$$(1.6) \quad \frac{1}{2} \|x - x_0\|^2 + \sum_{i \in \mathcal{V}} \mathbf{f}_i(x) = \sum_{i \in \mathcal{V}} \left[\mathbf{f}_i(x) + \frac{1}{2} \|x_i - [x_0]_i\|^2 \right].$$

In the case when $\mathbf{f}_i \equiv 0$ for all $i \in \mathcal{V}$, the problem reduces to the average consensus problem [10, 19]. The minimizer to (1.5) is the vector $(\bar{x}, \dots, \bar{x}) \in X^{|\mathcal{V}|}$, where $\bar{x} = \frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} [x_0]_i$.

The (Fenchel) dual to (1.5) is

$$(1.7) \quad \max_{z_\alpha \in X^{|\mathcal{V}|}, \alpha \in \mathcal{E} \cup \mathcal{V}} F(\{z_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}),$$

where

$$(1.8) \quad \begin{aligned} & F(\{z_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\ & := -\frac{1}{2} \left\| x_0 - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\|^2 + \frac{1}{2} \|x_0\|^2 - \sum_{(i,j) \in \mathcal{E}} \delta_{H_{(i,j)}}^*(z_{(i,j)}) - \sum_{i \in \mathcal{V}} \mathbf{f}_i^*(z_i). \end{aligned}$$

In the case when $f_i(\cdot) \equiv \delta_{C_i}(\cdot)$ for some closed convex set C_i for all i , Dykstra's algorithm finds the primal minimizer of the problem

$$(1.9) \quad \min_{x \in X} \frac{1}{2} \|x - x_0\|^2 + \sum_{i=1}^n f_i(x),$$

where $x_0 \in X$ (note that, in (1.9), x and x_0 lie in X instead of $X^{|\mathcal{V}|}$ as in (1.5)), by maximizing the dual

$$(1.10) \quad \max_{y \in X^n} -\frac{1}{2} \left\| x_0 - \sum_{i=1}^n y_i \right\| - \sum_{i=1}^n f_i^*(y_i) + \frac{1}{2} \|x_0\|^2$$

through block coordinate minimization. If each $f_i(\cdot)$ is allowed to be any closed convex function, it can now be seen that (1.8) is actually a special case of (1.10).

Dykstra's algorithm was first studied in [20] in the case where $f_i(\cdot) \equiv \delta_{C_i}(\cdot)$ and C_i are closed convex sets for all i . The convergence of the primal iterates to the projection of x_0 onto $\cap_{i=1}^n C_i$ was proved in [12], and is sometimes called the Boyle–Dykstra theorem. In [26] Dykstra's algorithm was independently noted to be a block coordinate minimization on the dual problem. The proof in [12] was adapted in [24] using duality. The Boyle–Dykstra theorem is remarkable because the convergence to the primal minimizer occurs even when there is no dual optimizer. (For example, look at [26, Page 9], where two circles in \mathbb{R}^2 intersect at only one point.) The case when sampling of the sets is noncyclic is addressed in [27] (among other things not directly relevant to this paper). As pointed out in [38], the Boyle–Dykstra theorem holds even if $f_i(\cdot)$ are closed convex functions instead of $\delta_{C_i}(\cdot)$. (We recently became aware that the dual ascent interpretation can be traced to [14, 15, 1], but the connection to distributed optimization was not pointed out there.) For more on the background to Dykstra's algorithm, we refer to [5, 4, 18, 22]. Some recent work on Dykstra's algorithm includes [43].

Dykstra's algorithm was extended to a distributed algorithm by Phade and Borkar [40], and they highlight the works [2, 29, 41, 36] on distributed optimization. The work in [40] is very different than the study of Dykstra's algorithm in [12] and [24].

It turns out that Notarnicola and Notarstefano [35] discuss a similar problem to (1.5). They generalize (1.5) by allowing the functions $x_i \mapsto \frac{1}{2} \|x_i - [x_0]_i\|^2$ to be any strongly convex function, and proceed to calculate that the dual has a similar form to (1.7) and (1.8). Their dual is still a sum of a smooth component and a separable component, which they solve with a randomized dual proximal gradient. We discuss the differences between their paper and ours in subsection 1.3.

1.3. Contributions of this paper. In section 2 of this paper, we look at the formulation (1.5) and show that Dykstra's algorithm applied to this formulation gives an algorithm with properties (2)–(5) from subsection 1.1. As stated in the introduction, we are not aware of any other asynchronous distributed algorithm that has deterministic convergence other than [25, 3, 2], though our assumption that the functions on each vertex have a strongly convex function with known modulus might be a bit strong.

We highlight the differences from [35]. The first difference is that we show that Dykstra's algorithm gives deterministic convergence (property (2)), whereas [35] pointed out probabilistic convergence. A naive application of Dykstra's algorithm to (1.5) would mean that all the edges in the graph have to be used in one cycle, which

would not cover the setting of time-varying graphs as done in [32]. But we show that, as long as the graph is uniformly connected (using the definition in [32]), convergence can be achieved. (See Remark 2.13.) Dykstra's splitting also gives the following two desirable properties that were not noticed in [35].

- (6) The iterates of the algorithm converge to the primal minimizer even when a dual minimizer does not exist.
- (7) Since Dykstra's splitting is a dual ascent algorithm, as many dual variables as possible can be maximized at one time. This is an advantage, as subproblems involving more dual variables lead to a greedier, and possibly greater, increase in the dual objective value.

Next, in section 3, we look at a decentralized dual ascent algorithm for (1.2) (which does not have the quadratic term) through the dual problem (1.4). Once again, the algorithm is asynchronous. In contrast to the adapted Dykstra algorithm, we now optimize dual variables corresponding to a collection of more than one vertex at a time. We show an example where convergence fails. The algorithm works on collections of vertices of the graph at a time, and is thus robust to lost communications in the edges of the graph.

Lastly, in section 4, we discuss ideas for an accelerated proximal gradient method on the dual (1.7). This algorithm runs on a global clock, and it does not work for time-varying graphs (i.e., it does not satisfy properties (4) and (5)). But asynchronous greedy steps satisfying property (7) can be performed to speed up the increase in the dual objective value.

2. Convergence of the distributed Dykstra algorithm. In this section, we state our distributed Dykstra's algorithm, make some remarks that may be helpful in understanding the algorithm, and prove its convergence without constraint qualifications.

2.1. Statement of the distributed Dykstra algorithm. Let $D \subset X^{|\mathcal{V}|}$ be the diagonal set defined by

$$D := \{x \in X^{|\mathcal{V}|} : x_1 = x_2 = \dots = x_{|\mathcal{V}|}\}.$$

With the definition of $H_{(i,j)}$ in (1.1) and $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ being a connected graph, it is obvious that

$$(2.1) \quad \bigcap_{(i,j) \in \mathcal{E}} H_{(i,j)} = D \text{ and } \sum_{(i,j) \in \mathcal{E}} H_{(i,j)}^\perp = D^\perp = \left\{ z \in X^{|\mathcal{V}|} : \sum_{i \in \mathcal{V}} z_i = 0 \right\}.$$

PROPOSITION 2.1. Suppose $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a connected graph. Let $H_{(i,j)}$ be the set (1.1) (defining the linear constraints relating the connection between nodes i and j). Let \mathcal{E}' be a subset of \mathcal{E} . The following conditions are equivalent:

- (1) $\bigcap_{(i,j) \in \mathcal{E}'} H_{(i,j)} = D$;
- (2) $\sum_{(i,j) \in \mathcal{E}'} H_{(i,j)}^\perp = D^\perp$;
- (3) the graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ is connected.

Proof. The equivalence between (1) and (3) is easy, and the equivalence between (1) and (2) is simple linear algebra. \square

DEFINITION 2.2. We say that \mathcal{E}' connects \mathcal{V} if any of the equivalent properties in Proposition 2.1 are satisfied.

We prove the following lemma.

LEMMA 2.3 (expressing v as a sum). *Suppose X is a finite-dimensional Hilbert space. There is a $C_1 > 0$ such that, for all $v \in D^\perp$ and $\mathcal{E}' \subset \mathcal{E}$ such that \mathcal{E}' connects \mathcal{V} , we can find $z_{(i,j)} \in H_{(i,j)}^\perp$ for all $(i,j) \in \mathcal{E}'$ such that $\sum_{(i,j) \in \mathcal{E}'} z_{(i,j)} = v$ and $\|z_{(i,j)}\| \leq C_1 \|v\|$ for all $(i,j) \in \mathcal{E}'$.*

Proof. This is elementary, so we only give an outline. Fix an \mathcal{E}' . We can choose $\tilde{H}_{(i,j)} \subset H_{(i,j)}^\perp$ so that $\sum_{(i,j) \in \mathcal{E}'} H_{(i,j)}^\perp = D^\perp$ is a direct sum of $\{\tilde{H}_{(i,j)}\}_{(i,j) \in \mathcal{E}'}$. So v can be written uniquely as the sum $v = \sum_{(i,j) \in \mathcal{E}'} z_{(i,j)}$, where $z_{(i,j)} \in \tilde{H}_{(i,j)}$. The mapping from v to each $z_{(i,j)} \in \tilde{H}_{(i,j)}$ is linear, and this linear map has a norm bounded by some $C_{(i,j), \mathcal{E}'}$. Letting C_1 be the maximum of these $C_{(i,j), \mathcal{E}'}$ gives us our conclusion. \square

We present our distributed Dykstra algorithm in Algorithm 1.

Algorithm 1 Decentralized Dykstra algorithm.

Consider problem (1.5) along with the associated dual problem (1.7). Let \bar{w} be a positive integer. Let $C_1 > 0$ satisfy Lemma 2.3. Our decentralized Dykstra algorithm is as follows.

01 Let

- $z_i^{1,0} \in X^{|\mathcal{V}|}$ be a starting dual vector for $\mathbf{f}_i(\cdot)$ for each $i \in \mathcal{V}$ so that $[z_i^{1,0}]_j = 0$ for all $j \in \mathcal{V} \setminus \{i\}$,
- $v_H^{1,0} \in D^\perp$ be a starting dual vector for (1.7) (note that $\{z_{(i,j)}^{n,0}\}_{(i,j) \in \mathcal{E}}$ is defined through $v_H^{n,0}$ in (2.2)),
- $x^{1,0} = x_0 - v_H^{1,0} - \sum_{i \in \mathcal{V}} z_i^{1,0}$.

02 For $n = 1, 2, \dots$,

- 03 let $\mathcal{E}_n \subset \mathcal{E}$ be such that \mathcal{E}_n connects \mathcal{V} ,
 04 define $\{z_{(i,j)}^{n,0}\}_{(i,j) \in \mathcal{E}}$ so that

$$(2.2a) \quad z_{(i,j)}^{n,0} = 0 \text{ for all } (i,j) \notin \mathcal{E}_n,$$

$$(2.2b) \quad z_{(i,j)}^{n,0} \in H_{(i,j)}^\perp \text{ for all } (i,j) \in \mathcal{E},$$

$$(2.2c) \quad \|z_{(i,j)}^{n,0}\| \leq C_1 \|v_H^{n,0}\| \text{ for all } (i,j) \in \mathcal{E},$$

$$(2.2d) \quad \text{and } \sum_{(i,j) \in \mathcal{E}} z_{(i,j)}^{n,0} = v_H^{n,0}.$$

(This is possible by Lemma 2.3.)

05 For $w = 1, 2, \dots, \bar{w}$,

06 choose a set $S_{n,w} \subset \mathcal{E}_n \cup \mathcal{V}$ such that $S_{n,w} \neq \emptyset$,

07 define $\{z_\alpha^{n,w}\}_{\alpha \in S_{n,w}}$ by

$$(2.3) \quad \{z_\alpha^{n,w}\}_{\alpha \in S_{n,w}} = \arg \min_{z_\alpha, \alpha \in S_{n,w}} \frac{1}{2} \left\| x_0 - \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w-1} - \sum_{\alpha \in S_{n,w}} z_\alpha \right\|^2 + \sum_{\alpha \in S_{n,w}} h_\alpha^*(z_\alpha).$$

08 Set $z_\alpha^{n,w} := z_\alpha^{n,w-1}$ for all $\alpha \notin S_{n,w}$.

09 End For

10 Let $z_i^{n+1,0} = z_i^{n,\bar{w}}$ for all $i \in \mathcal{V}$ and $v_H^{n+1,0} = v_H^{n,\bar{w}} = \sum_{(i,j) \in \mathcal{E}} z_{(i,j)}^{n,\bar{w}}$.

11 End For

Remark 2.4 (intuition behind Algorithm 1). We now provide some intuition behind Algorithm 1. The classical Dykstra splitting approach is the block coordinate maximization of the dual problem (1.7), (1.8). This is reflected in lines 06–08 of Algorithm 1. In order for Algorithm 1 to handle time-varying graphs, we choose $\mathcal{E}_n \subset \mathcal{E}$ in line 03 so that \mathcal{E}_n connects \mathcal{V} , and the problem

$$\max_{z_\alpha \in X^{|\mathcal{V}|}, \alpha \in \mathcal{E}_n \cup \mathcal{V}} -\frac{1}{2} \left\| x_0 - \sum_{\alpha \in \mathcal{E}_n \cup \mathcal{V}} z_\alpha \right\|^2 + \frac{1}{2} \|x_0\|^2 - \sum_{(i,j) \in \mathcal{E}_n} \delta_{H_{(i,j)}}^*(z_{(i,j)}) - \sum_{i \in \mathcal{V}} \mathbf{f}_i^*(z_i)$$

(note the \mathcal{E}_n in the above formula) would have the same optimal objective value as (1.7), (1.8) since the corresponding primal problems are equivalent and strong duality holds. The subset $\mathcal{E}_n \subset \mathcal{E}$ chosen in line 03 may be such that $z_{(i,j)} \neq 0$, but $(i,j) \notin \mathcal{E}_n$. So we perform line 04 so that (2.2) holds, which implies that $z_{(i,j)} = 0$ for all $(i,j) \notin \mathcal{E}_n$, while preserving $v_H^{n,\bar{w}} = v_H^{n+1,0}$ (see (2.4) later). As we shall see in Remark 2.10 later, the reassignment of $\{z_{(i,j)}\}_{(i,j) \in \mathcal{E}}$ in line 04 is necessary for further analysis, but may be ignored in implementing the algorithm. Algorithm 2 then shows an equivalent formulation of Algorithm 1 where one only keeps track of $x^{n,w}$ and $\{z_i^{n,w}\}_{i \in \mathcal{V}}$. If $f_i(\cdot) \equiv 0$, then $z_i^{n,w}$ is always 0 for all $i \in \mathcal{V}$, so Algorithm 2 reduces to the averaged consensus algorithm [10, 19].

Remark 2.5 (choice of $S_{n,w}$). The choice of $S_{n,w}$ allows for a flexibility in how large one wants the subproblem (2.3) to be. It is easy to see that a small $S_{n,w}$ allows for the subproblems to be small and easy to solve. The larger the size of $S_{n,w}$, the harder the subproblem, but the greater the expected increase in the dual objective value. An issue with choosing large $|S_{n,w}|$ is that we need an extra coordination between the nodes in $V'_{n,w}$. When $S_{n,w} \subset \mathcal{V}$ and $|S_{n,w}| = 1$, there is no coordination needed. Similarly, when $S_{n,w} \subset \mathcal{E}$ and $|S_{n,w}| = 1$, only two nodes need to coordinate with each other, which is okay for an undirected graph. An implementer can, for example, choose a star-like subgraph (i.e., there is a central node in the subgraph connecting to all others) and apply an algorithm suitable for problems with a centralized node. Since the dual objective value acts as a Lyapunov function, one could choose $S_{n,w}$ to be as large as one can reasonably solve, to increase the dual objective value as much as one can. This increase in the dual objective value can be greater if large subproblems are solved partially than if small subproblems are solved fully.

To simplify calculations, we let v_A , v_H , and x be denoted by

$$(2.4a) \quad v_H = \sum_{(i,j) \in \mathcal{E}} z_{(i,j)},$$

$$(2.4b) \quad v_A = v_H + \sum_{i \in \mathcal{V}} z_i,$$

$$(2.4c) \quad x = x_0 - v_A.$$

Intuitively, v_H describes the sum of the dual variables due to $H_{(i,j)}$ for all $(i,j) \in \mathcal{E}$, v_A is the sum of all dual variables, and x is the estimate of the primal variable.

The following inequality describes the duality gap between (1.5) and (1.7):

$$\begin{aligned}
 (2.5) \quad & \frac{1}{2} \|x_0 - x\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x) - F(\{z_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\
 \stackrel{(1.8)}{=} & \frac{1}{2} \|x_0 - x\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} [h_\alpha(x) + h_\alpha^*(z_\alpha)] \\
 & - \left\langle x_0, \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\rangle + \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\|^2 \\
 \stackrel{\text{Fenchel duality}}{\geq} & \frac{1}{2} \|x_0 - x\|^2 + \left\langle x, \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\rangle - \left\langle x_0, \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\rangle + \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\|^2 \\
 = & \frac{1}{2} \left\| x_0 - x - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha \right\|^2 \geq 0.
 \end{aligned}$$

Claim 2.6. In Algorithm 1, for all $\alpha \in S_{n,w}$, we have

- (a) $-x^{n,w} + \partial h_\alpha^*(z_\alpha^{n,w}) \ni 0$,
- (b) $-z_\alpha^{n,w} + \partial h_\alpha(x^{n,w}) \ni 0$, and
- (c) $h_\alpha(x^{n,w}) + h_\alpha^*(z_\alpha^{n,w}) = \langle x^{n,w}, z_\alpha^{n,w} \rangle$.

Proof. By taking the optimality conditions in (2.3) with respect to z_α for $\alpha \in S_{n,w}$ and making use of (2.4) to get $x^{n,w} = x_0 - \sum_{\alpha \in \mathcal{V} \cup \mathcal{E}} z_\alpha^{n,w}$, we deduce (a). The equivalence of (a), (b), and (c) is standard. \square

Even though Algorithm 1 is described so that each node $i \in \mathcal{V}$ and edge $(i,j) \in \mathcal{E}$ contain a variable $z_\alpha \in X^{|\mathcal{V}|}$, the size of the variable z_α that needs to be stored in each node and edge is small due to sparsity.

PROPOSITION 2.7 (sparsity of z_α). *We have $[z_i^{n,w}]_j = 0$ for all $j \in \mathcal{V} \setminus \{i\}$, $n \geq 1$, and $w \in \{0, 1, \dots, \bar{w}\}$. Similarly, $[z_{(i,j)}^{n,w}]_k = 0$ for all $k \in \mathcal{V} \setminus \{i, j\}$, $n \geq 1$, and $w \in \{0, 1, \dots, \bar{w}\}$.*

Proof. The result for $z_i^{n,w}$ holds for $n = 1$ and $w = 0$. Claim 2.6(b) shows that $z_i^{n,w} \in \partial \mathbf{f}_i(x^{n,w})$ for all $i \in S_{n,w}$. Note that, since $[\mathbf{f}_i(x)]_j = 0$ for all $j \in \mathcal{V} \setminus \{i\}$, $[\partial \mathbf{f}_i(x)]_j = 0$ for all $j \in \mathcal{V} \setminus \{i\}$, which easily gives what we need.

For all $(i,j) \in \mathcal{E}$ and $n \geq 1$, (2.2b) implies that $z_{(i,j)}^{n,0} \in H_{(i,j)}^\perp$, and Claim 2.6(b) implies that $z_{(i,j)}^{n,w} \in H_{(i,j)}^\perp$ for all $w \in \{1, \dots, \bar{w}\}$. This implies the result at hand for $z_{(i,j)}^{n,w}$. \square

Dykstra's algorithm is traditionally written in terms of solving for the primal variable x . For completeness, we show the equivalence between (2.3) and the primal minimization problem. The proof is easily extended from [38, Proposition 2.4] (the duality between (2.3) and (2.6) can also be obtained by Fenchel duality).

PROPOSITION 2.8 (on solving (2.3)). *If a minimizer $\{z_\alpha^{n,w}\}_{\alpha \in S_{n,w}}$ for (2.3) exists, then the $x^{n,w}$ in (2.4c) satisfies*

$$(2.6) \quad x^{n,w} = \arg \min_{x \in X^{|\mathcal{V}|}} \sum_{\alpha \in S_{n,w}} h_\alpha(x) + \frac{1}{2} \left\| x - \left(x_0 - \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w} \right) \right\|^2.$$

Conversely, if $x^{n,w}$ solves (2.6) with the dual variables $\{z_\alpha^{n,w}\}_{\alpha \in S_{n,w}}$ satisfying

$$(2.7) \quad \tilde{z}_\alpha^{n,w} \in \partial h_\alpha(x^{n,w}) \text{ and } x^{n,w} - x_0 + \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w} + \sum_{\alpha \in S_{n,w}} \tilde{z}_\alpha^{n,w} = 0,$$

then $\{\tilde{z}_\alpha^{n,w}\}_{\alpha \in S_{n,w}}$ solves (2.3).

2.2. Examples of $S_{n,w}$. In this subsection, we elaborate on how to solve (2.6), and show that Algorithm 1 is an extension of the average consensus algorithm.

For an $S_{n,w}$ such that $S_{n,w} \cap \mathcal{E} \neq \emptyset$, define

$$(2.8) \quad \mathcal{V}'_{n,w} = \{\text{all vertices that are endpoints of some edge in } S_{n,w} \cap \mathcal{E}\}.$$

Suppose $S_{n,w} \cap \mathcal{E}$ is such that the subgraph $(\mathcal{V}'_{n,w}, S_{n,w} \cap \mathcal{E})$ is a connected graph with no cycles, and $S_{n,w} \cap \mathcal{V} \subset \mathcal{V}'_{n,w}$. Let $\tilde{y} \in X^{|\mathcal{V}|}$ be defined by

$$(2.9) \quad \tilde{y} := x_0 - \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w} \stackrel{\text{line}=08}{=} x_0 - \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w-1} \stackrel{(2.4)}{=} x^{n,w-1} + \sum_{\alpha \in S_{n,w}} z_\alpha^{n,w-1}.$$

Then the primal minimization problem (2.6) becomes

$$\begin{aligned} x^{n,w} &= \arg \min_{x \in X^{|\mathcal{V}|}} \sum_{i \in S_{n,w} \cap \mathcal{V}} h_i(x) + \sum_{(i,j) \in S_{n,w} \cap \mathcal{E}} h_{(i,j)}(x) \\ &\quad + \frac{1}{2} \left\| x - \left(x_0 - \sum_{\alpha \notin S_{n,w}} z_\alpha^{n,w} \right) \right\|^2 \\ (2.10) \quad &\stackrel{(2.9)}{=} \arg \min_{x \in X^{|\mathcal{V}|}} \sum_{i \in S_{n,w} \cap \mathcal{V}} h_i(x) + \sum_{(i,j) \in S_{n,w} \cap \mathcal{E}} h_{(i,j)}(x) + \frac{1}{2} \|x - \tilde{y}\|^2. \end{aligned}$$

Recall that $h_i : X^{|\mathcal{V}|} \rightarrow \mathbb{R}$ is a function whose output depends only on the i th coordinate, where $i \in \mathcal{V}$. If $x^{n,w}$ were to solve (2.10), then $h_{(i,j)}(x^{n,w}) = \delta_{H_{(i,j)}}(x^{n,w})$ is finite for all $(i,j) \in S_{n,w} \cap \mathcal{E}$, which shows that $x^{n,w} \in H_{(i,j)}$ for all $(i,j) \in S_{n,w} \cap \mathcal{E}$. This in turn means that all the components of $x^{n,w}$ indexed by $\mathcal{V}'_{n,w}$ would need to have the same value. So problem (2.10) can be reduced to one which optimizes over a variable in X (instead of $X^{|\mathcal{V}|}$), which, for all $i' \in \mathcal{V}'_{n,w}$, takes the form

$$\begin{aligned} (2.11) \quad x_{i'}^{n,w} &\stackrel{(2.10)}{=} \arg \min_{x \in X} \sum_{i \in S_{n,w} \cap \mathcal{V}} f_i(x) + \frac{1}{2} \sum_{i \in \mathcal{V}'_{n,w}} \|x - \tilde{y}_i\|^2 \\ &= \arg \min_{x \in X} \sum_{i \in S_{n,w} \cap \mathcal{V}} f_i(x) + \frac{|\mathcal{V}'_{n,w}|}{2} \left\| x - \frac{1}{|\mathcal{V}'_{n,w}|} \sum_{i \in \mathcal{V}'_{n,w}} \tilde{y}_i \right\|^2, \end{aligned}$$

where $f_i : X \rightarrow \bar{\mathbb{R}}$ is defined as in Problem 1.1. The iterate $x^{n,w} \in X^{|\mathcal{V}|}$ can be expressed in terms of $x^{n,w-1}$ via

$$(2.12) \quad x_{i'}^{n,w} = \begin{cases} x_{i'}^{n,w-1} & \text{if } i' \notin \mathcal{V}'_{n,w}, \\ \text{the formula in (2.11)} & \text{if } i' \in \mathcal{V}'_{n,w}. \end{cases}$$

Remark 2.9 (the $S_{n,w} \cap \mathcal{V} = \emptyset$ case). A notable case is when $S_{n,w} \cap \mathcal{V} = \emptyset$ and $|S_{n,w} \cap \mathcal{E}| = 1$. Let (i,j) be the element in $S_{n,w} \cap \mathcal{E}$. Then one can calculate from (2.12) that $x_i^{n,w} = x_j^{n,w} = \frac{1}{2}(x_i^{n,w-1} + x_j^{n,w-1})$, and the other $|\mathcal{V}|$ components of $x^{n,w}$ remain unchanged from $x^{n,w-1}$. If $f_i(\cdot) \equiv 0$ for all $i \in \mathcal{V}$ and the edges are chosen over the graph $(\mathcal{V}, \mathcal{E})$, the result reduces to the case of averaged consensus studied in [10, 19].

2.3. Simplification of Algorithm 1 and further remarks. We first remark that there is no need to track $\{z_{(i,j)}^{n,w-1}\}_{(i,j) \in \mathcal{E}}$ throughout the algorithm, and we only need to keep track of $\{z_i^{n,w-1}\}_{i \in \mathcal{V}}$ and $x^{n,w-1}$. We make a few more remarks about Algorithm 1.

Remark 2.10 (irrelevance of $z_{(i,j)}^{n,w-1}$). A first observation of the dual objective function is that, as long as $z_{(i,j)} \in H_{(i,j)}^\perp$, we have $\delta_{H_{(i,j)}}^*(z_{(i,j)}) = \delta_{H_{(i,j)}^\perp}(z_{(i,j)}) = 0$. Since

$$-\frac{1}{2} \left\| x_0 - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha^{n,w} \right\|^2 \stackrel{(2.4a)}{=} -\frac{1}{2} \left\| x_0 - v_H^{n,w} - \sum_{\alpha \in \mathcal{V}} z_\alpha^{n,w} \right\|^2,$$

the dual objective function (1.8) thus does not depend directly on each $\{z_{(i,j)}\}_{(i,j) \in \mathcal{E}}$, but rather can be obtained through the sum $v_H := \sum_{(i,j) \in \mathcal{E}} z_{(i,j)}$ that appears in the quadratic term in (2.3). Next, in calculating $\frac{1}{|\mathcal{V}'_{n,w}|} \sum_{i \in \mathcal{V}'_{n,w}} \tilde{y}_i$ in (2.11), we note that, since $z_{(i,j)}^{n,w-1} \in H_{(i,j)}^\perp \subset D^\perp$,

$$\sum_{i' \in \mathcal{V}} [z_{(i,j)}^{n,w-1}]_{i'} \stackrel{(2.1)}{=} 0.$$

Also, by Proposition 2.7, if $\alpha \in S_{n,w} \cap \mathcal{E}$, then $[z_\alpha^{n,w-1}]_i = 0$ if $i \in \mathcal{V} \setminus \mathcal{V}'_{n,w}$. This means that

$$\begin{aligned} \sum_{i \in \mathcal{V}'_{n,w}} \tilde{y}_i &\stackrel{(2.9)}{=} \sum_{i \in \mathcal{V}'_{n,w}} \left[x^{n,w-1} + \sum_{\alpha \in S_{n,w}} z_\alpha^{n,w-1} \right]_i \\ &= \sum_{i \in \mathcal{V}'_{n,w}} x_i^{n,w-1} + \sum_{i \in \mathcal{V}'_{n,w}} \sum_{\alpha \in S_{n,w} \cap \mathcal{V}} [z_\alpha^{n,w-1}]_i + \sum_{i \in \mathcal{V}'_{n,w}} \sum_{\alpha \in S_{n,w} \cap \mathcal{E}} [z_\alpha^{n,w-1}]_i \\ &= \sum_{i \in \mathcal{V}'_{n,w}} x_i^{n,w-1} + \sum_{i \in \mathcal{V}'_{n,w}} \sum_{\alpha \in S_{n,w} \cap \mathcal{V}} [z_\alpha^{n,w-1}]_i + \sum_{i \in \mathcal{V}} \sum_{\alpha \in S_{n,w} \cap \mathcal{E}} [z_\alpha^{n,w-1}]_i \\ &\stackrel{(2.1)}{=} \sum_{i \in \mathcal{V}'_{n,w}} x_i^{n,w-1} + \sum_{i \in \mathcal{V}'_{n,w}} \sum_{\alpha \in S_{n,w} \cap \mathcal{V}} [z_\alpha^{n,w-1}]_i. \end{aligned} \tag{2.13}$$

So one only needs to keep track of $x^{n,w}$ and $\{z_i^{n,w}\}_{i \in \mathcal{V}}$ in Algorithm 1, and there is no need to keep track of $\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E}}$. This justifies the reassignment of $z_{(i,j)}^{n,0}$ in line 04, and Algorithm 1 could have been stated in terms of v_H only, and not $\{z_{(i,j)}\}_{(i,j) \in \mathcal{E}}$. We introduce the variables $\{z_{(i,j)}\}_{(i,j) \in \mathcal{E}}$ so that the analysis in (2.24) can be carried through.

In view of Remark 2.10, Algorithm 1 can thus be simplified to Algorithm 2 without the terms $\{z_{(i,j)}\}_{(i,j) \in \mathcal{E}}$. Furthermore, if $f_i(\cdot) \equiv 0$ for all $i \in V$, the variables $z_i^{n,w}$ would always be zero, and Algorithm 2 would reduce to the well-known averaged consensus problem [10, 19].

Remark 2.11 (distributed asynchronous computation). Proposition 2.7 shows that the storage requirement for each vertex and edge is small. Suppose $S_{n,w}$ and $S_{n,w+1}$ are such that $S_{n,w} \cap \mathcal{V} \subset \mathcal{V}'_{n,w}$, $S_{n,w+1} \cap \mathcal{V} \subset \mathcal{V}'_{n,w+1}$ and $\mathcal{V}'_{n,w} \cap \mathcal{V}'_{n,w+1} = \emptyset$. Then the computations for the iterations (n, w) and $(n, w+1)$ can be conducted in parallel. This is because the calculations for $S_{n,w}$ in (2.3) only use and affect the coordinates of $\{z_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}$ indexed by $\mathcal{V}'_{n,w} \subset \mathcal{V}$ and a similar approach applies for $S_{n,w+1}$. This idea can naturally be extended to the case of $S_{n,w}, S_{n,w+1}, \dots, S_{n,w+j}$ for any $j \geq 1$ to allow for distributed asynchronous computation.

Algorithm 2 Simplified decentralized Dykstra algorithm.

Consider problem (1.5) along with the associated dual problem (1.7). We only keep track of $\{z_i^{n,w}\}_{i \in \mathcal{V}}$ and $x^{n,w}$, and these iterates are equivalent to those of Algorithm 1 by Remark 2.10. Let \bar{w} be a positive integer. Our decentralized Dykstra algorithm is as follows.

01 Let

- $z_i^{1,0} \in X^{|\mathcal{V}|}$ be a starting dual vector for $f_i(\cdot)$ for each $i \in \mathcal{V}$ so that $[z_i^{1,0}]_j = 0$ for all $j \in \mathcal{V} \setminus \{i\}$,
- $v_H^{1,0} \in D^\perp$ be a starting dual vector for (1.7),
- $x^{1,0}$ be $x^{1,0} = x_0 - v_H^{1,0} - \sum_{i \in \mathcal{V}} z_i^{1,0}$.

02 For $n = 1, 2, \dots$,

03 let $\mathcal{E}_n \subset \mathcal{E}$ be such that \mathcal{E}_n connects \mathcal{V} ,

05 for $w = 1, 2, \dots, \bar{w}$,

06 choose a set $S_{n,w} \subset \mathcal{E}_n \cup \mathcal{V}$ such that $S_{n,w} \neq \emptyset$ and $S_{n,w} \cap \mathcal{V} \subset \mathcal{V}'_{n,w}$ for $\mathcal{V}'_{n,w}$ as defined in (2.8),

07 define $x^{n,w} \in X^{|\mathcal{V}|}$ by

$$(2.14) \quad x_{i'}^{n,w} \stackrel{(2.12)}{=} \begin{cases} x_{i'}^{n,w-1} & \text{if } i' \notin \mathcal{V}'_{n,w}, \\ \arg \min_{x \in X} \sum_{i \in S_{n,w} \cap \mathcal{V}} f_i(x) + \frac{|\mathcal{V}'_{n,w}|}{2} \left\| x - \frac{1}{|\mathcal{V}'_{n,w}|} \sum_{i \in \mathcal{V}'_{n,w}} \tilde{y}_i \right\|^2 & \text{if } i' \in \mathcal{V}'_{n,w}, \end{cases}$$

where $\sum_{i \in \mathcal{V}'_{n,w}} \tilde{y}_i$ has the form in (2.13), which does not depend on $\{z_{(i,j)}^{n,w-1}\}_{(i,j) \in \mathcal{E}}$.

Let $\{z_i^{n,w}\}_{i \in S_{n,w} \cap \mathcal{V}}$ be such that $[z_i^{n,w}]_i$ is the subgradient of $f_i(\cdot)$ at $x_{i'}^{n,w}$ that certifies the optimality in (2.14).

08 Set $z_\alpha^{n,w} := z_\alpha^{n,w-1}$ for all $\alpha \notin S_{n,w} \cap \mathcal{V}$.

09 End For

10 Let $z_i^{n+1,0} = z_i^{n,\bar{w}}$ for all $i \in \mathcal{V}$, and $x^{n+1,0} = x^{n,\bar{w}}$.

11 End For

Remark 2.12 (scalability). Algorithm 1 allows for the size of the sets $S_{n,w}$ to be arbitrarily large so that there would be a greedier increase in the dual objective value. One would then expect faster convergence with larger sizes of $S_{n,w}$. Even though for this paper, we only cover the case where $|S_{n,w} \cap \mathcal{V}| \leq 1$, the case where $|S_{n,w} \cap \mathcal{V}| > 1$ can be analyzed using the techniques in [38], where we split vertices in \mathcal{V} according to whether $\text{dom}(f_i) = X$, $f_i(\cdot)$ is an indicator function of a closed convex set, or $f_i(\cdot)$ is a general closed convex function.

Remark 2.13 (time-varying graphs). Note that in line 05 of Algorithm 1 we only need to choose $\mathcal{E}_n \subset \mathcal{E}$ so that \mathcal{E}_n connects \mathcal{V} . As long as $\mathcal{E}_n = [\cup_{w=1}^{\bar{w}} S_{n,w}] \cap \mathcal{E}$, the convergence result in Theorem 2.14 holds. So as long as enough edges are chosen in each cycle to connect the graph, Algorithm 1 will converge. In [32], the terms uniformly strongly connectedness and B -strongly connectedness were used for time-varying directed graphs. Our assumption is equivalent to the way B -connectedness would have been defined for undirected graphs.

2.4. Convergence of Algorithm 1. We state some notation necessary for further discussions. For any $\alpha \in \mathcal{E} \cup \mathcal{V}$ and $n \in \{1, 2, \dots\}$, let

$$p(n, \alpha) = \max\{m : m \leq \bar{w}, \alpha \in S_{n,m}\}.$$

In other words, $p(n, \alpha)$ is the index m such that $\alpha \in S_{n,m}$ but $\alpha \notin S_{n,k}$ for all $k \in \{m+1, \dots, \bar{w}\}$. It follows from line 08 in Algorithm 1 that

$$(2.15) \quad z_\alpha^{n,p(n,\alpha)} = z_\alpha^{n,p(n,\alpha)+1} = \dots = z_\alpha^{n,\bar{w}}.$$

Moreover, $(i, j) \notin \mathcal{E}_n$ implies $(i, j) \notin S_{n,w}$ for all $w \in \{1, \dots, \bar{w}\}$, so

$$(2.16) \quad 0 \stackrel{(2.2a)}{=} z_{(i,j)}^{n,0} = z_{(i,j)}^{n,1} = \dots = z_{(i,j)}^{n,\bar{w}} \text{ for all } (i, j) \notin \mathcal{E}_n.$$

We have the following theorem on the convergence of Algorithm 1.

THEOREM 2.14 (convergence to primal minimizer). *Consider Algorithm 1. Assume that, for all $n \geq 1$, $\mathcal{E}_n = [\cup_{w=1}^{\bar{w}} S_{n,w}] \cap \mathcal{E}$ and $[\cup_{w=1}^{\bar{w}} S_{n,w}] \supset \mathcal{V}$. Assume that there are constants A and B such that*

$$(2.17) \quad \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} \|z_\alpha^{n,\bar{w}}\| \leq A\sqrt{n} + B \text{ for all } n \geq 0.$$

For the sequence $\{z_\alpha^{n,w}\}_{1 \leq n < \infty, 0 \leq w \leq \bar{w}} \subset X^{|\mathcal{V}|}$ for each $\alpha \in \mathcal{E} \cup \mathcal{V}$ generated by Algorithm 1 and the sequences $\{v_H^{n,w}\}_{1 \leq n < \infty, 0 \leq w \leq \bar{w}} \subset X^{|\mathcal{V}|}$ and $\{v_A^{n,w}\}_{1 \leq n < \infty, 0 \leq w \leq \bar{w}} \subset X^{|\mathcal{V}|}$ thus derived, we have the following:

- (i) the sum $\sum_{n=1}^{\infty} \sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\|^2$ is finite and $\{F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}})\}_{n=1}^{\infty}$ is nondecreasing;
- (ii) there is a constant C such that $\|v_A^{n,w}\|^2 \leq C$ for all $n \in \mathbb{N}$ and $w \in \{1, \dots, \bar{w}\}$;
- (iii) there exists a subsequence $\{v_A^{n_k,w}\}_{k=1}^{\infty}$ of $\{v_A^{n,w}\}_{n=1}^{\infty}$ which converges to some $v_A^* \in X^{|\mathcal{V}|}$ and such that

$$\lim_{k \rightarrow \infty} \langle v_A^{n_k,\bar{w}} - v_A^{n_k,p(n_k,\alpha)}, z_\alpha^{n_k,\bar{w}} \rangle = 0 \text{ for all } \alpha \in \mathcal{E} \cup \mathcal{V};$$

- (iv) for the v_A^* in (iii), $x_0 - v_A^*$ is the minimizer of the primal problem (P) and we have $\lim_{k \rightarrow \infty} F(\{z_\alpha^{n_k,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) = \frac{1}{2} \|v_A^*\|^2 + h(x_0 - v_A^*)$, where $h(\cdot) = \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(\cdot)$.

The properties (i) to (iv) in turn imply that $\lim_{n \rightarrow \infty} x^{n,\bar{w}}$ exists and equals $x_0 - v_A^*$, which is the primal minimizer of (1.5).

Proof. We first show that (i) to (iv) implies the final assertion. For all $n \in \mathbb{N}$ we have, from weak duality,

$$(2.18) \quad F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \leq \frac{1}{2} \|x_0 - (x_0 - v_A^*)\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x_0 - v_A^*).$$

Since the values $\{F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}})\}_{n=1}^{\infty}$ are nondecreasing in n , we make use of (iv) to get

$$\lim_{n \rightarrow \infty} F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) = \frac{1}{2} \|x_0 - (x_0 - v_A^*)\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x_0 - v_A^*).$$

Hence, $x_0 - v_A^* = \arg \min_x h(x) + \frac{1}{2} \|x - x_0\|^2$, and (substituting $x = x_0 - v_A^*$ in (2.5))

$$\begin{aligned} & \frac{1}{2} \|x_0 - (x_0 - v_A^*)\|^2 + h(x_0 - v_A^*) - F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\ & \stackrel{(2.5), (2.4a), (2.4b)}{\geq} \frac{1}{2} \|x_0 - (x_0 - v_A^*) - v_A^{n,\bar{w}}\|^2 \\ & \stackrel{(2.4c)}{=} \frac{1}{2} \|x^{n,\bar{w}} - (x_0 - v_A^*)\|^2. \end{aligned}$$

Hence, $\lim_{n \rightarrow \infty} x^{n,\bar{w}}$ is the minimizer in (P).

It remains to prove assertions (i) to (iv).

Proof of (i). From the fact that $\{z_\alpha^{n,w}\}_{\alpha \in S_{n,w}}$ minimize (2.3) (which includes the quadratic regularizer) we have

$$(2.19) \quad F(\{z_\alpha^{n,w-1}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \stackrel{(2.3)}{\leq} F(\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) - \frac{1}{2} \|v_A^{n,w} - v_A^{n,w-1}\|^2.$$

(The last term in (2.19) arises from the quadratic term in (2.3).) By line 10 of Algorithm 1, $z_i^{n+1,0} = z_i^{n,\bar{w}}$ for all $i \in \mathcal{V}$ and $v_H^{n+1,0} = v_H^{n,\bar{w}}$ (even though the decompositions (2.2d) of $v_H^{n+1,0}$ and $v_H^{n,\bar{w}}$ may be different). Combining (2.19) over all $m \in \{1, \dots, n\}$ and $w \in \{1, \dots, \bar{w}\}$, we have

$$F(\{z_\alpha^{1,0}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) + \sum_{m=1}^n \sum_{w=1}^{\bar{w}} \|v_A^{m,w} - v_A^{m,w-1}\|^2 \stackrel{(2.19)}{\leq} F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}).$$

Next, $F(\{z_\alpha^{n,\bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}})$ is bounded from above by weak duality. The proof of the claim is complete.

Proof of (ii). Substituting $\{z_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}$ in (2.5) to be $\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}$ and x to be the primal minimizer x^* , we have

$$\begin{aligned} & \frac{1}{2} \|x_0 - x^*\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x^*) - F(\{z_\alpha^{1,0}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\ & \stackrel{\text{part (i)}}{\geq} \frac{1}{2} \|x_0 - x^*\|^2 + \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x^*) - F(\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\ & \stackrel{(2.5)}{\geq} \frac{1}{2} \left\| x_0 - x^* - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} z_\alpha^{n,w} \right\|^2 \\ & \stackrel{(2.4b)}{=} \frac{1}{2} \|x_0 - x^* - v_A^{n,w}\|^2. \end{aligned}$$

The conclusion is immediate.

Proof of (iii). We first make use of the technique in [5, Lemma 29.1] (which in turn is largely attributed to [12]) to show that

$$(2.20) \quad \liminf_{n \rightarrow \infty} \left[\left(\sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\| \right) \sqrt{n} \right] = 0.$$

Seeking a contradiction, suppose instead that there are an $\epsilon > 0$ and $\bar{n} > 0$ such that if $n > \bar{n}$, then

$$\left(\sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\| \right) \sqrt{n} > \epsilon.$$

By the Cauchy–Schwarz inequality, we have

$$\frac{\epsilon^2}{n} < \left(\sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\| \right)^2 \leq \bar{w} \sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\|^2.$$

This contradicts the earlier claim in (i) that

$$\sum_{n=1}^{\infty} \sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\|^2$$

is finite.

Through (2.20), we find a sequence $\{n_k\}_{k=1}^\infty$ such that

$$(2.21) \quad \lim_{k \rightarrow \infty} \left[\left(\sum_{w=1}^{\bar{w}} \|v_A^{n_k, w} - v_A^{n_k, w-1}\| \right) \sqrt{n_k} \right] = 0.$$

Recalling the assumption (2.17), we get

$$(2.22) \quad \lim_{k \rightarrow \infty} \left[\left(\sum_{w=1}^{\bar{w}} \|v_A^{n_k, w} - v_A^{n_k, w-1}\| \right) \|z_\alpha^{n_k, \bar{w}}\| \right] \stackrel{(2.17),(2.21)}{=} 0 \text{ for all } \alpha \in \mathcal{E} \cup \mathcal{V}.$$

Moreover,

$$(2.23) \quad \begin{aligned} |\langle v_A^{n_k, \bar{w}} - v_A^{n_k, p(n_k, \alpha)}, z_\alpha^{n_k, \bar{w}} \rangle| &\leq \|v_A^{n_k, \bar{w}} - v_A^{n_k, p(n_k, \alpha)}\| \|z_\alpha^{n_k, \bar{w}}\| \\ &\leq \left(\sum_{w=1}^{\bar{w}} \|v_A^{n_k, w} - v_A^{n_k, w-1}\| \right) \|z_\alpha^{n_k, \bar{w}}\|. \end{aligned}$$

By (ii) and the finite dimensionality of X , there exists a further subsequence of $\{v_A^{n_k, \bar{w}}\}_{k=1}^\infty$ which converges to some $v_A^* \in X$. Combining (2.22) and (2.23) gives (iii).

Proof of (iv). From earlier results, we obtain

$$\begin{aligned} (2.24) \quad &-\sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x_0 - v_A^*) \\ &\stackrel{(2.5)}{\leq} \frac{1}{2} \|x_0 - (x_0 - v_A^*)\|^2 - F(\{z_\alpha^{n_k, \bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) \\ &\stackrel{(1.8),(2.15)}{=} \frac{1}{2} \|v_A^*\|^2 + \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} h_\alpha^*(z_\alpha^{n_k, p(n_k, \alpha)}) \\ &\quad + \sum_{(i,j) \notin \mathcal{E}_{n_k}} h_{(i,j)}^*(z_{(i,j)}^{n_k, \bar{w}}) - \langle x_0, v_A^{n_k, \bar{w}} \rangle + \frac{1}{2} \|v_A^{n_k, \bar{w}}\|^2 \\ &\text{Claim 2.6(c), } \alpha \in S_{n_k, p(n_k, \alpha)}, \\ &\stackrel{(2.16)}{=} \frac{1}{2} \|v_A^*\|^2 + \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} \langle x_0 - v_A^{n_k, p(n_k, \alpha)}, z_\alpha^{n_k, p(n_k, \alpha)} \rangle \\ &\quad - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} h_\alpha(x_0 - v_A^{n_k, p(n_k, \alpha)}) - \langle x_0, v_A^{n_k, \bar{w}} \rangle + \frac{1}{2} \|v_A^{n_k, \bar{w}}\|^2 \\ &\stackrel{(2.15)}{=} \frac{1}{2} \|v_A^*\|^2 - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} \langle v_A^{n_k, p(n_k, \alpha)} - v_A^{n_k, \bar{w}}, z_\alpha^{n_k, \bar{w}} \rangle \\ &\quad - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} h_\alpha(x_0 - v_A^{n_k, p(n_k, \alpha)}) - \langle x_0, v_A^{n_k, \bar{w}} \rangle \\ &\quad + \left\langle x_0 - v_A^{n_k, \bar{w}}, \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} z_\alpha^{n_k, p(n_k, \alpha)} \right\rangle + \frac{1}{2} \|v_A^{n_k, \bar{w}}\|^2 \\ &\stackrel{(2.4b),(2.16)}{=} \frac{1}{2} \|v_A^*\|^2 - \frac{1}{2} \|v_A^{n_k, \bar{w}}\|^2 - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} \langle v_A^{n_k, p(n_k, \alpha)} - v_A^{n_k, \bar{w}}, z_\alpha^{n_k, \bar{w}} \rangle \\ &\quad - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} h_\alpha(x_0 - v_A^{n_k, p(n_k, \alpha)}) \end{aligned}$$

Since $\lim_{k \rightarrow \infty} v_A^{n_k, \bar{w}} = v_A^*$, we have $\lim_{k \rightarrow \infty} \frac{1}{2} \|v_A^*\|^2 - \frac{1}{2} \|v_A^{n_k, \bar{w}}\|^2 = 0$. The third term in the last group of formulas (i.e., the sum involving the inner products) converges to 0 by (iii).

Next, recall that if $(i, j) \in \mathcal{E}_n$, by (2.6), we have $h_{(i,j)}(x_0 - v_A^{n,p(n,(i,j))}) = 0$, which gives

$$x_0 - v_A^{n,p(n,(i,j))} \stackrel{(1.2),(1.1)}{\in} H_{(i,j)}.$$

There is a constant $\kappa_{\mathcal{E}_{n_k}} > 0$ such that

$$\begin{aligned} (2.25) \quad d(x_0 - v_A^{n_k, \bar{w}}, \cap_{(i,j) \in \mathcal{E}} H_{(i,j)}) \\ &\stackrel{\mathcal{E}_{n_k} \text{ connects } \mathcal{V}, \text{Prop. 2.1(1)}}{=} d(x_0 - v_A^{n_k, \bar{w}}, \cap_{(i,j) \in \mathcal{E}_{n_k}} H_{(i,j)}) \\ &\leq \kappa_{\mathcal{E}_{n_k}} \max_{(i,j) \in \mathcal{E}_{n_k}} d(x_0 - v_A^{n_k, \bar{w}}, H_{(i,j)}) \\ &\leq \kappa_{\mathcal{E}_{n_k}} \max_{(i,j) \in \mathcal{E}_{n_k}} \|v_A^{n_k, \bar{w}} - v_A^{n_k, p(n_k, (i,j))}\|. \end{aligned}$$

Let $\kappa := \max\{\kappa_{\mathcal{E}'} : \mathcal{E}' \text{ connects } \mathcal{V}\}$. We have $\kappa_{\mathcal{E}_{n_k}} \leq \kappa$. Taking limits of (2.25), the right-hand side converges to zero by (i), so $d(x_0 - v_A^*, \cap_{(i,j) \in \mathcal{E}} H_{(i,j)}) = 0$, or $x_0 - v_A^* \in \cap_{(i,j) \in \mathcal{E}} H_{(i,j)}$. So $\sum_{(i,j) \in \mathcal{E}} h_{(i,j)}(x_0 - v_A^*) = 0$. Together with the fact that $x_0 - v_A^{n_k, p(n_k, (i,j))} \in H_{(i,j)}$, we have

$$(2.26) \quad \sum_{(i,j) \in \mathcal{E}_{n_k}} h_{(i,j)}(x_0 - v_A^{n_k, p(n_k, (i,j))}) = 0 = \sum_{(i,j) \in \mathcal{E}} h_{(i,j)}(x_0 - v_A^*).$$

Lastly, by the lower semicontinuity of $h_i(\cdot)$, we have

$$(2.27) \quad - \lim_{k \rightarrow \infty} \sum_{i \in \mathcal{V}} h_i(x_0 - v_A^{n_k, p(n_k, i)}) \leq - \sum_{i \in \mathcal{V}} h_i(x_0 - v_A^*).$$

As mentioned after (2.24), taking the limits as $k \rightarrow \infty$ would result in the first three terms of the last formula in (2.24) being zero. Hence,

$$\begin{aligned} - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x_0 - v_A^*) &\stackrel{(2.24)}{\leq} \lim_{k \rightarrow \infty} - \sum_{\alpha \in \mathcal{E}_{n_k} \cup \mathcal{V}} h_\alpha(x_0 - v_A^{n_k, p(n_k, \alpha)}) \\ &\stackrel{(2.26),(2.27)}{\leq} - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(x_0 - v_A^*). \end{aligned}$$

So (2.24) becomes an equation in the limit. The first two lines of (2.24) then give

$$\lim_{k \rightarrow \infty} F(\{z_\alpha^{n_k, \bar{w}}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}) = \frac{1}{2} \|v_A^*\|^2 + \sum_{i \in \mathcal{V}} h_i(x_0 - v_A^*),$$

which shows that $x_0 - v_A^*$ is the primal minimizer. \square

A last detail that we need to resolve is to show that (2.17) holds for the choice of $S_{n,w}$ in Algorithm 1.

PROPOSITION 2.15 (growth of $\sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} \|z_\alpha^{n,w}\|$). *If $S_{n,w}$ are such that $|S_{n,w} \cap \mathcal{V}| \leq 1$ for all $n \in \mathbb{N}$ and $w \in \{1, \dots, \bar{w}\}$ as in Algorithm 1, then (2.17) holds.*

Proof. We have either $S_{n,w} \cap \mathcal{V} = \emptyset$ or $|S_{n,w} \cap \mathcal{V}| = 1$. In the second case, let $i^* \in S_{n,w} \cap \mathcal{V}$. Otherwise, in the first case, we let i^* be any

index in \mathcal{V} . We have

$$\begin{aligned}
 (2.28) \quad & \sum_{i \in \mathcal{V}} [v_A^{n,w} - v_A^{n,w-1}]_i \stackrel{\text{line 08}, (2.4)}{=} \sum_{i \in \mathcal{V}} \sum_{\alpha \in S_{n,w}} [z_\alpha^{n,w} - z_\alpha^{n,w-1}]_i \\
 & \stackrel{z_{(i,j)} \in D^\perp, (2.1)}{=} \sum_{i \in \mathcal{V}} [z_{i^*}^{n,w} - z_{i^*}^{n,w-1}]_i \\
 & \stackrel{\text{Prop. 2.7}}{=} [z_{i^*}^{n,w} - z_{i^*}^{n,w-1}]_{i^*}.
 \end{aligned}$$

Recall that the norm $\|\cdot\|$ always refers to the 2-norm unless stated otherwise. By the equivalence of norms in finite dimensions, we can find a constant c_1 such that

$$\begin{aligned}
 (2.29) \quad & \|v_A^{n,w} - v_A^{n,w-1}\| \geq c_1 \sum_{i \in \mathcal{V}} \|[v_A^{n,w} - v_A^{n,w-1}]_i\| \\
 & \geq c_1 \left\| \sum_{i \in \mathcal{V}} [v_A^{n,w} - v_A^{n,w-1}]_i \right\| \\
 & \stackrel{(2.28)}{=} c_1 \|z_{i^*}^{n,w} - z_{i^*}^{n,w-1}\| \stackrel{(2.3)}{=} c_1 \sum_{i \in \mathcal{V}} \|z_i^{n,w} - z_i^{n,w-1}\|.
 \end{aligned}$$

Next, $v_H^{n,w} - v_H^{n,w-1} \stackrel{(2.4b)}{=} v_A^{n,w} - v_A^{n,w-1} - (z_{i^*}^{n,w} - z_{i^*}^{n,w-1})$, so

$$\begin{aligned}
 (2.30) \quad & \|v_H^{n,w} - v_H^{n,w-1}\| \leq \|v_A^{n,w} - v_A^{n,w-1}\| + \|z_{i^*}^{n,w} - z_{i^*}^{n,w-1}\| \\
 & \stackrel{(2.29)}{\leq} \left(1 + \frac{1}{c_1}\right) \|v_A^{n,w} - v_A^{n,w-1}\|.
 \end{aligned}$$

We can choose $\{z_{(i,j)}^{n,w}\}_{(i,j) \in \mathcal{E}}$ such that

$$(2.31) \quad \sum_{(i,j) \in S_{n,w} \cap \mathcal{E}} [z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}] \stackrel{\text{line 08}}{=} \sum_{(i,j) \in \mathcal{E}} [z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}] \stackrel{(2.4a)}{=} v_H^{n,w} - v_H^{n,w-1}.$$

Without loss of generality, we can assume that $S_{n,w} \cap \mathcal{E}$ contains edges that do not form a cycle. This also means that for $v_H^{n,w} - v_H^{n,w-1}$, each $z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}$ can be determined uniquely with a linear map from the relation (2.31). Therefore, there is a constant $\kappa_{(i,j), S_{n,w} \cap \mathcal{E}} > 0$ such that

$$(2.32) \quad \|z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}\| \leq \kappa_{(i,j), S_{n,w} \cap \mathcal{E}} \|v_H^{n,w} - v_H^{n,w-1}\|.$$

Thus there is a constant $\kappa > 0$ such that

$$(2.33) \quad \sum_{(i,j) \in \mathcal{E}} \|z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}\| \stackrel{(2.3)}{=} \sum_{(i,j) \in S_{n,w} \cap \mathcal{E}} \|z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}\| \stackrel{(2.32)}{\leq} \kappa \|v_H^{n,w} - v_H^{n,w-1}\|.$$

Combining (2.29), (2.30), and (2.33) together shows that there is a constant $C_2 > 1$ such that

$$(2.34) \quad \|v_H^{n,w} - v_H^{n,w-1}\| + \sum_{(i,j) \in \mathcal{E}} \|z_{(i,j)}^{n,w} - z_{(i,j)}^{n,w-1}\| + \sum_{i \in \mathcal{V}} \|z_i^{n,w} - z_i^{n,w-1}\| \leq C_2 \|v_A^{n,w} - v_A^{n,w-1}\|.$$

Since $\{z_\alpha^{n,0}\}_{\alpha \in \mathcal{E}}$ was chosen to satisfy (2.2), there is some $M > 1$ such that

$$(2.35) \quad \sum_{\alpha \in \mathcal{E}} \|z_\alpha^{n,0}\| \stackrel{(2.2c)}{\leq} M \|v_H^{n,0}\| \stackrel{(2.2d)}{\leq} M \left(\|v_H^{1,0}\| + \sum_{m=1}^{n-1} \sum_{w=1}^{\bar{w}} \|v_H^{m,w} - v_H^{m,w-1}\| \right).$$

Now, for any $n \geq 1$, we have

$$\begin{aligned} (2.36) \quad \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} \|z_\alpha^{n,\bar{w}}\| &\leq \sum_{\alpha \in \mathcal{E}} \|z_\alpha^{n,0}\| + \sum_{w=1}^{\bar{w}} \sum_{\alpha \in \mathcal{E}} \|z_\alpha^{n,w} - z_\alpha^{n,w-1}\| \\ &\quad + \sum_{m=1}^n \sum_{w=1}^{\bar{w}} \sum_{\alpha \in \mathcal{V}} \|z_\alpha^{m,w} - z_\alpha^{m,w-1}\| + \sum_{\alpha \in \mathcal{V}} \|z_\alpha^{1,0}\| \\ &\stackrel{(2.35)}{\leq} M \|v_H^{1,0}\| + \sum_{\alpha \in \mathcal{V}} \|z_\alpha^{1,0}\| + \sum_{w=1}^{\bar{w}} \left(\sum_{\alpha \in \mathcal{E}} \|z_\alpha^{n,w} - z_\alpha^{n,w-1}\| \right) \\ &\quad + \sum_{m=1}^{n-1} \sum_{w=1}^{\bar{w}} \left(M \|v_H^{m,w} - v_H^{m,w-1}\| + \sum_{\alpha \in \mathcal{V}} \|z_\alpha^{m,w} - z_\alpha^{m,w-1}\| \right) \\ &\stackrel{(2.34)}{\leq} M \|v_H^{1,0}\| + \sum_{\alpha \in \mathcal{V}} \|z_\alpha^{1,0}\| + MC_2 \sum_{m=1}^n \sum_{w=1}^{\bar{w}} \|v_A^{n,w} - v_A^{n,w-1}\|. \end{aligned}$$

By the Cauchy–Schwarz inequality, we have

$$(2.37) \quad \sum_{m=1}^n \sum_{w=1}^{\bar{w}} \|v_A^{m,w} - v_A^{m,w-1}\| \leq \sqrt{n\bar{w}} \sqrt{\sum_{m=1}^n \sum_{w=1}^{\bar{w}} \|v_A^{m,w} - v_A^{m,w-1}\|^2}.$$

Since the second square root of the right-hand side of (2.37) is bounded by Theorem 2.14(i), we make use of (2.36) to obtain the conclusion (2.17) as needed. \square

Remark 2.16 (convergence rate). An aspect of Algorithm 1 that we do not cover in this paper is the convergence rate. In the case where there are no dual minimizers, components of the dual variables $\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}, 1 \leq n < \infty, 0 \leq w \leq \bar{w}}$ need not be bounded. But, in the case where the variables $\{z_\alpha^{n,w}\}_{\alpha \in \mathcal{E} \cup \mathcal{V}, 1 \leq n < \infty, 0 \leq w \leq \bar{w}}$ remain bounded as $n \rightarrow \infty$, an $O(1/n)$ rate was shown for the dual objective function, which leads to an $O(1/\sqrt{n})$ convergence rate of the distance $\|x^{n,w} - x^*\|$ to the optimal solution x^* . The ideas for these results are presented in [38, section 3]. Such ideas were already present in [8, 6], for example.

2.5. Generality of (1.5). Another case of interest is when $\frac{1}{2}\|x - x_0\|^2$ in (1.5) is replaced by $\frac{1}{2}\|x - x_0\|_Q^2$, where $\|x\|_Q^2 = \langle x, Qx \rangle$ and Q is a block diagonal positive definite matrix. In the case where Q is such that $\|x\|_Q^2 = \sum_{i \in \mathcal{V}} \lambda_i \|x_i\|^2$ for some $\lambda \in \mathbb{R}^{|\mathcal{V}|}$ such that $\lambda > 0$ and $\mathbf{f}_i \equiv 0$ for all $i \in \mathcal{V}$, then the minimizer of (1.5) is

$$\frac{1}{\sum_{i \in \mathcal{V}} \lambda_i} \sum_{i \in \mathcal{V}} \lambda_i y_i^0.$$

In other words, (1.5) becomes a weighted average consensus problem.

We show how to transform a problem involving $\frac{1}{2}\|x - x_0\|_Q^2$ to one involving $\frac{1}{2}\|x - x_0\|^2$. Note that

$$\begin{aligned} \min_{x \in X^{|\mathcal{V}|}} \frac{1}{2}\|x - x_0\|_Q^2 + \sum_{(i,j) \in \mathcal{E}} \delta_{H_{(i,j)}}(x) + \sum_{i \in \mathcal{V}} \mathbf{f}_i(x) \\ \equiv \min_{x \in X^{|\mathcal{V}|}} \frac{1}{2}\|Q^{1/2}x - Q^{1/2}x_0\|^2 + \sum_{(i,j) \in \mathcal{E}} \delta_{Q^{1/2}H_{(i,j)}}(Q^{1/2}x) + \sum_{i \in \mathcal{V}} \mathbf{f}_i \circ Q^{-1/2}(Q^{1/2}x). \end{aligned}$$

We can thus let v^0 be $Q^{1/2}x_0$, and seek the variable $v = Q^{1/2}x$. The function $\delta_{Q^{1/2}H_{(i,j)}}(\cdot)$ requires a transformation of the set (1.1), but the transformed problem would fit the framework of Dykstra's algorithm.

3. Distributed algorithm for functions not strongly convex. We saw earlier that the minimization of the sum of strongly convex functions can be minimized over a network. A natural question to ask is whether it is possible to minimize the sum of functions that are not necessarily strongly convex in the same setting.

A technique for minimizing (1.4) is to choose two or more nodes, say $S_k \subset \mathcal{V}$ (which preferably forms a connected subgraph to allow for communications), and then minimize the function varying only the dual variables corresponding to the chosen nodes. This leads to Algorithm 3.

Algorithm 3 Distributed dual ascent algorithm.

Consider problem (1.4). Let $\{y_i^0\}_{i \in \mathcal{V}} \subset X$ be starting variables such that

$$\sum_{i \in \mathcal{V}} y_i^0 = 0.$$

For $n = 1, \dots$,

find a set $S_n \subset \mathcal{V}$ such that $|S_n| \geq 2$,
for all $i \in S_n$, define y_i^n so that

$$\begin{aligned} \{y_i^n\}_{i \in S_n} &\in \arg \max_{\{y_i\}_{i \in S_n}} - \sum_{i \in S_n} f_i^*(y_i) \\ \text{s.t. } \sum_{i \in S_n} y_i &= \sum_{i \in S_n} y_i^{n-1}, \end{aligned}$$

define $y_i^k = y_i^{k-1}$ for all $i \notin S_n$.

End For

Such a method is analogous to the method of alternating minimization, which have stationary points that are not optimal points. We now show an example of such a stationary point for Algorithm 3.

Example 3.1 (Algorithm 3 can get stuck at a nonoptimal value). Consider a graph with $\mathcal{V} = \{1, 2, 3\}$ and $\mathcal{E} = \{(1, 2), (2, 3)\}$. Let $f_i : \mathbb{R} \rightarrow \bar{\mathbb{R}}$, $i \in \{1, 2, 3\}$, be defined by

$$f_1(x) = \frac{1}{2}(x+1)^2, \quad f_2(x) = \delta_{\{0\}}(x), \quad f_3(x) = \frac{1}{2}(x-1)^2,$$

and we have $f_1^*(z) = \frac{1}{2}z^2 - z$, $f_2^*(z) = \delta_{\mathbb{R}}(z)$, and $f_3^*(z) = \frac{1}{2}z^2 + z$. Let $\bar{y} = (0, 0, 0) \in$

\mathbb{R}^3 . For $S \subset \{1, 2, 3\}$, denote the dual problem (DP_S) by

$$(DP_S) \quad \begin{aligned} & \max_{y_i \in \mathbb{R}, i \in S} - \sum_{i \in S} f_i^*(y_i) \\ & \text{s.t. } \sum_{i \in S} y_i = \sum_{i \in S} \bar{y}_i. \end{aligned}$$

The problem $(DP_{\{1,2\}})$ has minimizer $y_1 = y_2 = 0$ with $-1 \in \partial f_1^*(0)$ and $-1 \in \partial f_2^*(0)$, and the problem $(DP_{\{2,3\}})$ minimizer $y_2 = y_3 = 0$ with $1 \in \partial f_2^*(0)$ and $1 \in \partial f_3^*(0)$. Hence, the problem $(DP_{\{1,2,3\}})$ has a stationary point of $\bar{y} = (0, 0, 0)$ for the alternating minimization method. However, the global minimizer to $(DP_{\{1,2,3\}})$ is $(1, 0, -1)$ with $0 \in \partial f_i^*(y_i)$ for $i \in \{1, 2, 3\}$.

Note from this example that the failure can be identified from the fact that $x = -1$ if we use the edge $(1, 2)$, and $x = 1$ if we use the edge $(2, 3)$. Node 2 should be able to figure out that the x values corresponding to edges $(1, 2)$ and $(2, 3)$ are too far apart, and one needs to minimize $f_1^*(\cdot) + f_3^*(\cdot)$ in order to avoid convergence to a nonoptimal value.

We now give a proof for the convergence of Algorithm 3, which is based on the proof in [44] (which in turn cited other references). To shorten notation, for each set $S \subset \mathcal{V}$, we let $\mathcal{D}(S)$ be the set of directions d in $X^{|\mathcal{V}|}$ defined by

$$\mathcal{D}(S) := \left\{ d \in X^{|\mathcal{V}|} : \sum_{i=1}^{|\mathcal{V}|} d_i = 0 \text{ and } \text{supp}(d) \subset S \right\},$$

where $\text{supp}(d)$ is the set $\{i : d_i \neq 0\}$. We define $G : X^{|\mathcal{V}|} \rightarrow \bar{\mathbb{R}}$ by

$$G(y) = \sum_{i \in \mathcal{V}} f_i^*(y_i),$$

and let $G'(y; d)$ be the directional derivative of G at y in the direction d .

THEOREM 3.2 (convergence of Algorithm 3). *Suppose that there is an integer T such that, for every $n \geq 1$, the sets $\{S_{n+i}\}_{i=1}^T$ satisfy the following.*

- (a) *Suppose S' , S'' are elements in $\{S_{n+i}\}_{i=1}^T$ such that $S' \cap S'' \neq \emptyset$ and $y \in X^{|\mathcal{V}|}$. Then, for all $d \in \mathcal{D}(S' \cup S'')$, we can find $d' \in \mathcal{D}(S')$ and $d'' \in \mathcal{D}(S'')$ such that $d = d' + d''$ and*

$$(3.1) \quad G'(y; d) = G'(y; d') + G'(y; d'').$$

- (b) *Suppose $y \in X^{|\mathcal{V}|}$. If, for all $r \in \{1, \dots, T\}$, $G'(y; d_r) \geq 0$ for all $d_r \in \mathcal{D}(S_{n+r})$, then y is a minimizer of $G(\cdot)$.*

Suppose further that the sequence $\{y^n\}_{n=1}^\infty$ is bounded. Then every cluster point of $\{y^n\}_{n=1}^\infty$ is a minimizer of G .

Proof. Our proof is adapted from the ideas in [44, section 4]. Suppose $\bar{y} \in X^{|\mathcal{V}|}$ is a cluster point of $\{y^n\}_{n=1}^\infty$ and that $\mathcal{R} \subset \{1, 2, \dots\}$ is such that $\lim_{r \in \mathcal{R}} y^r = \bar{y}$. We can assume, by taking subsequences if necessary, that $\lim_{r \in \mathcal{R}} y^{r-T+1+j}$ converges to some \bar{y}^j for all $j \in \{1, \dots, T\}$. We have $\bar{y}^{T-1} = \bar{y}$. We also note that $\{G(y^n)\}_{n=1}^\infty$ is a nonincreasing sequence, so

$$(3.2) \quad \lim_{r \in \mathcal{R}} G(y^{r-T+1+j}) \text{ exists for all } j \in \{1, \dots, T\}.$$

Next, we can assume that, for each $j \in \{1, \dots, T\}$, the set $\{S_{r-T+1+j}\}_{r \in \mathcal{R}}$ depends only on j , which we call \bar{S}_j . For each $j \in \{1, \dots, T\}$, since \bar{S}_j is chosen at iteration $r - T + 1 + j$ for $r \in \mathcal{R}$, we have

$$\begin{aligned} G(y^{r-T+1+j}) &\leq G(y^{r-T+1+j} + d_j) \text{ for all } d_j \in \mathcal{D}(\bar{S}_j), \\ y_i^{r-T+j} &= y_i^{r-T+1+j} \text{ for all } i \notin \bar{S}_j. \end{aligned}$$

Then the continuity of $G(\cdot)$ gives us

$$\begin{aligned} (3.3) \quad G(\bar{y}^j) &\leq G(\bar{y}^j + d_j) \text{ for all } d_j \in \mathcal{D}(\bar{S}_j), \\ \bar{y}_i^{j-1} &= \bar{y}_i^j \text{ for all } i \notin \bar{S}_j. \end{aligned}$$

We have $G(\bar{y}^1) = \dots = G(\bar{y}^T)$. The previous line also gives $\bar{y}^j - \bar{y}^{j-1} \in \mathcal{D}(\bar{S}_j)$, so

$$(3.4) \quad G(\bar{y}^{j-1}) = G(\bar{y}^j) \stackrel{(3.3)}{\leq} G(\bar{y}^j + (d_j + \bar{y}^{j-1} - \bar{y}^j)) = G(\bar{y}^{j-1} + d_j) \text{ for all } d_j \in \mathcal{D}(\bar{S}_j).$$

We claim that, for $j = 1, \dots, T-1$,

$$(3.5) \quad G(\bar{y}^j) \leq G(\bar{y}^j + d_k) \text{ for all } d_k \in \mathcal{D}(\bar{S}_1), \mathcal{D}(\bar{S}_2), \dots, \text{ or } \mathcal{D}(\bar{S}_j).$$

By (3.3), (3.5) holds for $j = 1$. Suppose (3.5) holds for $j = 1, \dots, l-1$ for some $l \in \{2, \dots, T-1\}$. We show that (3.5) holds for $j = l$. From (3.4),

$$G(\bar{y}^{l-1}) \leq G(\bar{y}^{l-1} + d_l) \text{ for all } d_l \in \mathcal{D}(\bar{S}_l),$$

implying that

$$(3.6) \quad G'(\bar{y}^{l-1}; \bar{y}^l - \bar{y}^{l-1} + v) \geq 0 \text{ for all } v \in \mathcal{D}(\bar{S}_l).$$

Also, since (3.5) holds for $j = l-1$, we have, for all $k \in \{1, \dots, l-1\}$,

$$G(\bar{y}^{l-1}) \leq G(\bar{y}^{l-1} + d_k - v) \text{ for all } d_k \in \mathcal{D}(\bar{S}_k) \text{ and } v \in \mathcal{D}(\bar{S}_l \cap \bar{S}_k),$$

which in turn implies

$$(3.7) \quad G'(\bar{y}^{l-1}; d_k - v) \geq 0 \text{ for all } d_k \in \mathcal{D}(\bar{S}_k) \text{ and } v \in \mathcal{D}(\bar{S}_l \cap \bar{S}_k).$$

If $\bar{S}_l \cap \bar{S}_k = \emptyset$, then v can be taken to be zero, and we get (3.1). By (3.6) and (3.7) and property (a) for each d_k , we can choose v such that, for all $d_k \in \mathcal{D}(\bar{S}_k)$,

$$(3.8) \quad G'(\bar{y}^{l-1}; \bar{y}^l - \bar{y}^{l-1} + d_k) \stackrel{\text{property (a)}}{=} G'(\bar{y}^{l-1}; \bar{y}^l - \bar{y}^{l-1} + v) + G'(\bar{y}^{l-1}; d_k - v) \stackrel{(3.6),(3.7)}{\geq} 0.$$

Since $G(\cdot)$ is convex,

$$G(\bar{y}^l + d_k) = G(\bar{y}^{l-1} + (\bar{y}^l - \bar{y}^{l-1} + d_k)) \stackrel{(3.8)}{\geq} G(\bar{y}^{l-1}) = G(\bar{y}^l) \text{ for all } d_k \in \mathcal{D}(\bar{S}_k).$$

Since (3.3) holds with $j = l$, (3.5) holds for $j = l$ and thus for all $j \in \{1, \dots, T-1\}$. Taking $j = T-1$ for (3.5) and combining this with property (b) proves that $\bar{y} = \bar{y}^{T-1}$ is a minimizer of $G(\cdot)$. \square

Define $\mathcal{V}_{sm} \subset \mathcal{V}$ to be such that

$$\mathcal{V}_{sm} := \{i \in \mathcal{V} : f_i^*(\cdot) \text{ is smooth}\}.$$

We now give more insight on properties (a) and (b) in Theorem 3.2.

PROPOSITION 3.3. *Theorem 3.2(a) is satisfied if, for any two elements S' , S'' in $\{S_{n+i}\}_{i=1}^T$, either $S' \cap S'' = \emptyset$ or $S' \cap S'' \cap \mathcal{V}_{sm} \neq \emptyset$.*

Proof. We only need to consider the case when $S' \cap S'' \cap \mathcal{V}_{sm} \neq \emptyset$. We want to show that if $d \in \mathcal{D}(S' \cup S'')$, then d can be written as $d = d' + d''$, where $d' \in \mathcal{D}(S')$ and $d'' \in \mathcal{D}(S'')$, so that (3.1) holds. If $S' \cap S'' \cap \mathcal{V}_{sm} \neq \emptyset$, then let $\bar{i} \in S' \cap S'' \cap \mathcal{V}_{sm}$. For a given $d \in \mathcal{D}(S' \cup S'')$, define d' and d'' so that

$$d'_i = \begin{cases} d_i & \text{if } i \in S' \setminus \{\bar{i}\}, \\ -\sum_{i \in S' \setminus \{\bar{i}\}} d_i & \text{if } i = \bar{i}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and } d''_i = \begin{cases} d_i & \text{if } i \in S'' \setminus S', \\ -\sum_{i \in S'' \setminus S'} d_i & \text{if } i = \bar{i}, \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that $d = d' + d''$, $d' \in \mathcal{D}(S')$, and $d'' \in \mathcal{D}(S'')$. From the smoothness of $f_i^*(\cdot)$, we have $[f_{\bar{i}}^*]'(x; d'_i + d''_i) = [f_{\bar{i}}^*]'(x; d'_i) + [f_{\bar{i}}^*]'(x; d''_i)$, which gives (3.1) as needed. \square

PROPOSITION 3.4. *Theorem 3.2(b) is satisfied if the following hold.*

- (1) *For all $n \geq 0$ and y , the condition $G'(y; d) \geq 0$ for all $d \in \mathcal{D}(S_n)$ implies the existence of KKT multipliers of*

$$\begin{aligned} & \max_{y'_i \in X, i \in S_n} - \sum_{i \in S_n} f_i^*(y'_i) \\ & \text{s.t.} \quad \sum_{i \in S_n} y'_i = \sum_{i \in S_n} y_i \end{aligned}$$

at a maximizer \bar{y} . Specifically, there exists $x \in X$ such that $x \in \partial f_i^(\bar{y}_i)$ for all $i \in S_n$.*

- (2) *For every $\bar{i}, \bar{j} \in \mathcal{V}$, we can find a sequence of sets $\{\tilde{S}_k\}_{k=1}^K \subset \{S_{n+r}\}_{r=1}^T$ such that $\bar{i} \in \tilde{S}_1$, $\bar{j} \in \tilde{S}_K$, and $\tilde{S}_k \cap \tilde{S}_{k+1} \cap \mathcal{V}_{sm} \neq \emptyset$ for all $k \in \{1, \dots, K-1\}$.*

Proof. Recall the y in Theorem 3.2(b). By condition (2), it suffices to prove that if \tilde{S}_1 and \tilde{S}_2 are such that $\tilde{S}_1 \cap \tilde{S}_2 \cap \mathcal{V}_{sm} \neq \emptyset$, then there exists x such that $x \in \partial f_i^*(y_i)$ for all $i \in \tilde{S}_1 \cup \tilde{S}_2$, which is in turn easy to prove from condition (1). \square

In Example 3.1, we see that $2 \notin \mathcal{V}_{sm}$, so Theorem 3.2 does not apply.

3.1. Connection between sections 2 and 3. We now give a connection between Algorithms 1–3 in sections 2 and 3. For a graph $(\mathcal{V}, \mathcal{E})$, construct the graph $(\mathcal{V}^+, \mathcal{E}^+)$ via

$$\begin{aligned} \mathcal{V}^+ &= \mathcal{V} \times \{0, 1\}, \\ \mathcal{E}^+ &= \{((i, 0), (j, 0)) : (i, j) \in \mathcal{E} \} \cup \{ ((i, 0), (i, 1)) : i \in \mathcal{V} \}. \end{aligned}$$

One can easily check that $|\mathcal{V}^+| = 2|\mathcal{V}|$ and $|\mathcal{E}^+| = |\mathcal{E}| + |\mathcal{V}|$. Let the function associated with the vertex (i, s) be $f_{i,s} : X \rightarrow \bar{\mathbb{R}}$ defined by

$$f_{(i,s)}(x) = \begin{cases} \frac{1}{2} \|x - [x_0]_i\|^2 & \text{if } s = 0, \\ f_i(x) & \text{if } s = 1. \end{cases}$$

Note that $\min_{x \in X} \sum_{(i,s) \in \mathcal{V}^+} f_{(i,s)}(x)$ is equivalent to problem (1.6) considered in Dykstra's algorithm.

The dual problem

$$\max_{y_{(i,s)} \in X, (i,s) \in \mathcal{V}^+} \left\{ - \sum_{(i,s) \in \mathcal{V}^+} f_{(i,s)}^*(y_{(i,s)}) : \sum_{(i,s) \in \mathcal{V}^+} y_{(i,s)} = 0 \right\}$$

(recall how (1.4) is derived as the dual of (1.3)) can be simplified to

$$(3.9) \quad \begin{aligned} & \max_{y_{i,s} \in X : i \in \mathcal{V}, s \in \{0,1\}} - \sum_{i \in \mathcal{V}} f_i^*(y_{i,1}) - \sum_{i \in \mathcal{V}} \left[\frac{1}{2} \|y_{i,0} + [x_0]_i\|^2 - \frac{1}{2} \|[x_0]_i\|^2 \right] \\ & \text{s.t.} \quad \sum_{i \in \mathcal{V}} (y_{i,0} + y_{i,1}) = 0. \end{aligned}$$

Recall the dual problem in (1.7) and (1.8). Define the variable $z_e \in X^{|\mathcal{V}|}$ to be

$$z_e := \sum_{(i,j) \in \mathcal{E}} z_{(i,j)}.$$

Suppose $z_{(i,j)} \in H_{(i,j)}^\perp$. Then $\delta_{H_{(i,j)}}^*(z_{(i,j)}) = 0$. Also, $z_{(i,j)} \in H_{(i,j)}^\perp \subset D^\perp$, so $z_e \in D^\perp$. The dual problem in (1.7) and (1.8) becomes

$$(3.10) \quad \max_{\{z_e \in D^\perp\} \cup \{z_\alpha \in X^{|\mathcal{V}|} : \alpha \in \mathcal{V}\}} - \sum_{i \in \mathcal{V}} \mathbf{f}_i^*(z_i) - \frac{1}{2} \left\| \sum_{i \in \mathcal{V}} z_i + z_e - x_0 \right\|^2 + \frac{1}{2} \|x_0\|^2.$$

We now show how (3.9) and (3.10) are related.

PROPOSITION 3.5. *Consider the problems (3.9) and (3.10).*

- (1) *The two problems are related through a change of variables. Specifically, if $\{y_{(i,s)}\}_{(i,s) \in \mathcal{V}^+} \subset X$ were obtained from $\{z_i\}_{i \in \mathcal{V}} \subset X^{|\mathcal{V}|}$ and $z_e \in X^{|\mathcal{V}|}$ by*

$$y_{i,1} = [z_i]_i \text{ and } y_{i,0} = -[z_e]_i - [z_i]_i \text{ for all } i \in \mathcal{V},$$

then the objective values in (3.10) and (3.9) coincide. Conversely, if $\{z_i\}_{i \in \mathcal{V}}$ and z_e were obtained from $\{y_{(i,s)}\}_{(i,s) \in \mathcal{V}^+}$ by

$$[z_i]_j = \begin{cases} 0 & \text{if } j \neq i, \\ y_{i,1} & \text{if } j = i, \end{cases} \text{ and } [z_e]_i = -y_{i,0} - y_{i,1} \text{ for all } i, j \in \mathcal{V},$$

then $z_e \in D^\perp$ and the objective values in (3.10) and (3.9) coincide.

- (2) *Let $S^1 \subset \mathcal{V}^+$ be a connected subset of vertices in the graph $(\mathcal{V}^+, \mathcal{E}^+)$ so that $|S^1| > 1$. Define $\Pi_0 S^1 \subset \mathcal{V}$ to be the set*

$$\Pi_0 S^1 := \{i \in \mathcal{V} : (i, 0) \in S^1\}.$$

Let $\Pi_1 S^1$ be similarly defined. With respect to the graph $(\mathcal{V}, \mathcal{E})$, suppose that there is a subset \mathcal{E}' of \mathcal{E} not containing any cycles that connect all the vertices in $\Pi_0 S^1$. Since S^1 is a subset of connected vertices in the graph $(\mathcal{V}^+, \mathcal{E}^+)$, we have $\Pi_1 S^1 \subset \Pi_0 S^1$.

Suppose that, for a fixed $\{y_{(i,s)}\}_{(i,s) \in \mathcal{V}^+} \subset X$, a subproblem of (3.9) is solved with only variables $y_{(i,s)}$ indexed by S^1 that are allowed to vary, while the other variables stay fixed. Then, under the change of variables in (1), this subproblem is equivalent to solving the subproblem in (3.10) where

- (a) $[z_e]_i$ is allowed to vary if and only if i is an endpoint of some edge in \mathcal{E}' , and
- (b) z_i is allowed to vary if and only if $i \in \Pi_1 S^1$.

Proof. Statement (1) is obvious from the constructions.

We now work on statement (2). From $|S^1| > 1$ and the definition of $(\mathcal{V}^+, \mathcal{E}^+)$, if $(i, 1) \in \mathcal{V}^+$, then $(i, 0) \in \mathcal{V}^+$. So, for each $i \in \mathcal{V}$, there are three cases: (1) $(i, 1) \notin \mathcal{V}^+$ and $(i, 0) \notin \mathcal{V}^+$ (in which case there is nothing to do); (2) $(i, 0) \in \mathcal{V}^+$ and $(i, 1) \notin \mathcal{V}^+$; and (3) $(i, 0) \in \mathcal{V}^+$ and $(i, 1) \in \mathcal{V}^+$.

In case (2), if the term $y_{i,0}$ is in S^1 , then $y_{i,0}$ affects only the $\frac{1}{2}\|y_{i,0} + [x_0]_i\|^2$ in (3.9). In turn, $[z_e]_i$ only affects $\frac{1}{2}\|[z_e]_i + [z_i]_i - [x_0]_i\|^2$ in the quadratic term in (3.10).

In case (3), it is clear that the term $f_i(y_{i,1})$ varies through $y_{i,1}$ if and only if $\mathbf{f}_i(z_i)$ varies through z_i . Recall that $\Pi_0 S^1 \subset \Pi_1 S^1$, so if the term $y_{i,1}$ is in S^1 , then $y_{i,0}$ is in S^1 . The terms $y_{i,1}$ and $y_{i,0}$ then combine to affect $f_i^*(y_{i,1}) + \frac{1}{2}\|y_{i,0} + [x_0]_i\|^2$. Correspondingly, $[z_e]_i$ and $[z_i]_i$ combine to affect $\mathbf{f}_i^*(z_i) + \frac{1}{2}\|[z_e]_i + [z_i]_i - [x_0]_i\|^2$.

To wrap up, note that the constraint in (3.9) corresponds to $z_e \in D^\perp$. \square

One can easily figure out that (2a) in Proposition 3.5 corresponds to varying $z_{(i,j)}$ for all $(i, j) \in \mathcal{E}'$ in the original dual problem of (1.7) and (1.8).

4. Accelerated methods for (1.7). In this section, we write down an accelerated proximal gradient (APG) algorithm [34, 7, 45] on the dual problem described through (1.7) and (1.8) that allows for greedy steps that can be performed asynchronously. We showed in [37] that an APG with greedy steps can be performed on the formulation (1.10). However, we point out that the APG derived from (1.7) and (1.8) has a much lower Lipschitz constant and allows for greedy steps of the form (2.3).

We first recall a variant of the accelerated proximal gradient in [45]. In view of the clash of variables, we substitute the variables x , y , and z in [45] by u , v , and w , and then substitute the $f^P(\cdot)$ in [45] for the function $-F(\cdot)$ in (1.8). (The algorithm in [45] allows for the domain of the optimization problem for w^{k+1} to change in each iteration, which we omit.) Let

$$(4.1) \quad l(u, v) = \underbrace{\frac{1}{2} \left\| x_0 - \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} v_\alpha \right\|^2}_{+ \left\langle \left(\sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} v_\alpha \right) - x_0, \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} (u_\alpha - v_\alpha) \right\rangle} - \frac{1}{2} \|x_0\|^2 + \underbrace{\sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} h_\alpha(u_\alpha)}.$$

Remark 4.1. We now explain the formula for $l(u, v)$ above. In [45], the function that Algorithm 4 aims to minimize is $f^P(u) = f(u) + P(u)$, where $f(\cdot)$ is smooth and $P(\cdot)$ admits an easy calculation of its proximal, and has a linearization $f(v) + \langle \nabla f(v), u - v \rangle + P(u)$. The underbraced terms in (4.1) play the role of the terms $f(v)$, $\langle \nabla f(v), u - v \rangle$, and $P(u)$ in the linearization of $-F(u)$ in (1.8).

Remark 4.2. Algorithm 4 requires the condition (4.3). Since $l(\cdot, v)$ is the linearization of $-F(\cdot)$ at v , we have $l(v, v) = -F(v)$. Since the smooth portions of both $l(u, v) + \frac{L}{2}\|u - v\|^2$ and $-F(\cdot)$ are quadratics, showing (4.3) is equivalent to finding $L > 0$ such that the Hessian of the smooth portion of $l(u, v) + \frac{L}{2}\|u - v\|^2$ is greater than that of $-F(\cdot)$, i.e.,

$$(4.2) \quad \frac{L}{2} \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} \|u_\alpha\|^2 \geq \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} u_\alpha \right\|^2.$$

Since the variables $\{u_\alpha\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}$ must satisfy the sparsity pattern in Proposition 2.7, we show that L can be chosen as follows.

Algorithm 4 From [45, Algorithm 1].

We want to find u to minimize $-F(\cdot)$, which is the sum of a convex smooth function and a convex separable function. Choose $\theta_0 \in (0, 1]$, $u^0, w^0 \in \text{dom}(P)$. Let $L > 0$ be such that

$$(4.3) \quad l(u, v) + \frac{L}{2} \|u - v\|^2 \geq -F(u) \text{ for all } u, v \in [X^{|\mathcal{V}|}]^{|\mathcal{V} \cup \mathcal{E}|}.$$

Go to 1.

1. Let

$$\begin{aligned} v^k &= (1 - \theta_k)u^k + \theta_k w^k, \\ w^{k+1} &= \arg \min_{u \in X} \{l(u; v^k) + \theta_k L \frac{1}{2} \|u - w^k\|^2\}, \\ \hat{u}^{k+1} &= (1 - \theta_k)u^k + \theta_k w^{k+1}. \end{aligned}$$

Choose u_{k+1} to be such that

$$(4.4) \quad -F(u^{k+1}) \leq l(\hat{u}^{k+1}; v^k) + \frac{L}{2} \|\hat{u}^{k+1} - v^k\|^2.$$

Choose $\theta_{k+1} \in (0, 1]$ satisfying

$$\frac{1 - \theta_{k+1}}{\theta_{k+1}^2} \leq \frac{1}{\theta_k^2}.$$

$k \leftarrow k + 1$, and go to 1.

PROPOSITION 4.3 (choice of L in Algorithm 4). *In order to satisfy (4.2) while obeying the sparsity pattern in Proposition 2.7, we can choose L to be $\bar{d} + 1$, where \bar{d} is the maximum degree of the vertices in the graph.*

Proof. We look at the i th component $\{[u_\alpha]_i\}_{\alpha \in \mathcal{E} \cup \mathcal{V}}$ of the terms in (4.2) for all $i \in \mathcal{V}$. As long as we can prove that

$$(4.5) \quad \frac{\bar{d} + 1}{2} \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} \| [u_\alpha]_i \|^2 \geq \frac{1}{2} \left\| \sum_{\alpha \in \mathcal{E} \cup \mathcal{V}} [u_\alpha]_i \right\|^2,$$

the conclusion will follow. In view of the sparsity pattern of the u 's in Proposition 2.7, most of the $[u_\alpha]_i$'s are zero. For all $i \in \mathcal{V}$, we define $\tilde{\mathcal{E}}_i$ to be the set of all edges $e \in \mathcal{E}$ such that one of the endpoints is i . Then (4.5) reduces to

$$(4.6) \quad \frac{\bar{d} + 1}{2} \left(\| [u_i]_i \|^2 + \sum_{e \in \tilde{\mathcal{E}}_i} \| [u_e]_i \|^2 \right) \geq \frac{1}{2} \left\| [u_i]_i + \sum_{e \in \tilde{\mathcal{E}}_i} [u_e]_i \right\|^2.$$

We form the vector $\tilde{u} \in X^{|\tilde{\mathcal{E}}_i|+1}$ so that it contains $[u_i]_i$ and $\{[u_e]_i\}_{e \in \tilde{\mathcal{E}}_i}$ as its

components. Formula (4.6) can be seen to be equivalent to

$$\tilde{u}^T \left[\begin{pmatrix} (\bar{d}+1)I & & & \\ & (\bar{d}+1)I & & \\ & & \ddots & \\ & & & (\bar{d}+1)I \end{pmatrix} - \begin{pmatrix} I & I & \cdots & I \\ I & I & \cdots & I \\ \vdots & \vdots & \ddots & \vdots \\ I & I & \cdots & I \end{pmatrix} \right] \tilde{u} \geq 0,$$

which is clearly true. Thus we are done. \square

If we had used the formulation in (2.3) without exploiting the sparsity in Proposition 2.7, then the corresponding L would be $|\mathcal{V} \cup \mathcal{E}|$, which is a much larger number than $\bar{d}+1$ in most large graphs. Recall that L was chosen so that (4.3) holds, and L should be as small as possible subject to this condition so that the step for calculating w^{k+1} in Algorithm 4 would minimize a function closer to $-F(\cdot)$. This lower value of $\bar{d}+1$ is one advantage of applying the APG to the dual problem from (1.7) and (1.8).

We recall the convergence result of Algorithm 4.

THEOREM 4.4 (convergence of Algorithm 4; see [45, Corollary 1]).

Let $\{(u^k, v^k, w^k, \theta_k, X_k)\}_k$ be generated by Algorithm 4 with $\theta_0 = 1$. Suppose $\theta_k \leq \frac{2}{k+2}$. Then, for any $u \in \text{dom}(P)$ with $-F(u) \leq \inf F + \epsilon$, for any $\epsilon > 0$, we have

$$\min_{i=0,1,\dots,k+1} \{-F(u^i)\} \leq -F(u) + \epsilon \text{ whenever } k \geq \sqrt{\frac{4L}{\epsilon}} \|x - x^0\| - 2.$$

In the particular case where there is a minimizer u^* , Theorem 4.4 says that an ϵ -optimal solution for $-F(\cdot)$ is obtained if the number of iterations k is the Nesterov accelerated rate of $O(\sqrt{1/\epsilon})$ [34, 7, 45]. In the case of Dykstra's algorithm (or block coordinate minimization), the number of iterations needed to obtain an ϵ -optimal solution is typically $O(1/\epsilon)$ (see, for example, [8, 6]), which is slower than the $O(\sqrt{1/\epsilon})$ rate.

Remark 4.5 (on Theorem 4.4). The proof of our Theorem 4.4 given in [45] is for the algorithm with a modified (4.4), with the left-hand side instead being $l(u^{k+1}; v^k) + \frac{L}{2} \|u^{k+1} - v^k\|^2$, but it carries over from [45] with no changes at all.

We now elaborate on how the greedy step can be applied to Algorithm 4.

Remark 4.6 (greedy steps in Algorithm 4). We remark that the greedy step can be performed in (4.4). Note that u^{k+1} in (4.4) can be chosen to be \hat{u}^{k+1} . But greedy steps of the form (2.3) can be performed in (4.4) (with u 's in the place of z 's). These greedy steps can be performed asynchronously, as discussed in Remark 2.11.

5. Conclusion. We have done what we set out to do in subsection 1.3. In short, we noticed that a dual ascent algorithm can give us a distributed and asynchronous algorithm with deterministic convergence for time-varying graphs when the function on each vertex is strongly convex with a known modulus. A separate related algorithm is proposed for the case when the function on each vertex is not necessarily strongly convex, but Example 3.1 shows that the algorithm can fail to converge.

Note that in Example 3.1, the failure of Algorithm 3 is identified by the primal variable -1 in $\partial f_1^*(0)$ and $\partial f_2^*(0)$ being obtained when $S = \{1, 2\}$, while the primal variable 1 in $\partial f_2^*(0)$ and $\partial f_3^*(0)$ is obtained when $S = \{2, 3\}$. Is there a primal dual algorithm that has a number of the properties listed in subsections 1.1 and 1.3?

Acknowledgments. We thank Mert Gürbüzbalaban and Necdet Serhat Aybat for the discussions leading to this paper. We also thank the associate editor, the anonymous referees, and the journal staff for the quick review of this paper.

REFERENCES

- [1] F. ABBOUD, E. CHOUZENOUX, J.-C. PESQUET, J.-H. CHENOT, AND L. LABORELLI, *Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences*, J. Math. Imaging Vision, 59 (2017), pp. 415–431.
- [2] N. AYBAT AND E. HAMEDANI, *A primal-dual method for conic constrained distributed optimization problems*, in Adv. Neural Inf. Process. Syst. 29, Curran Associates, Red Hook, NY, 2016, pp. 5049–5057.
- [3] A. AYTEKIN, H. FEYZMAHDAVIAN, AND M. JOHANSSON, *Analysis and Implementation of an Asynchronous Optimization Algorithm for the Parameter Server*, preprint, <https://arxiv.org/abs/1610.05507>, 2016.
- [4] H. BAUSCHKE AND J. BORWEIN, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.
- [5] H. BAUSCHKE AND P. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Berlin, 2011.
- [6] A. BECK, *On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes*, SIAM J. Optim., 25 (2015), pp. 185–209.
- [7] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [8] A. BECK AND L. TETRASHVILI, *On the convergence of block coordinate descent type methods*, SIAM J. Optim., 23 (2013), pp. 2037–2060.
- [9] P. BIANCHI, W. HACHEM, AND F. IUTZELER, *A stochastic coordinate descent primal-dual algorithm and applications*, in 2014 IEEE International Workshop on Machine Learning for Signal Processing, IEEE Press, Piscataway, NJ, 2014, doi: <http://doi.org/10.1109/MLSP.2014.6958866>.
- [10] S. BOYD, A. GHOSH, B. PRABHAKAR, AND D. SHAH, *Randomized gossip algorithms*, IEEE Trans. Inf. Theory, 52 (2006), pp. 2508–2530.
- [11] S. BOYD, N. PARikh, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2010), pp. 1–122.
- [12] J. BOYLE AND R. DYKSTRA, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, in Advances in Order Restricted Statistical Inference, Lect. Notes Stat., Springer, New York, 1985, pp. 28–47.
- [13] L. CANNELLI, F. FACCHINEI, V. KUNGURTSEV, AND G. SCUTARI, *Asynchronous Parallel Algorithms for Nonconvex Big-Data Optimization, Part I: Model and Convergence*, preprint, <https://arxiv.org/abs/1607.04818>, 2016.
- [14] P. COMBETTES, D. DÜNG, AND B. VŨ, *Dualization of signal recovery problems*, Set-Valued Var. Anal., 18 (2010), pp. 373–404.
- [15] P. COMBETTES, D. DÜNG, AND B. VŨ, *Proximity for sums of composite functions*, J. Math. Anal. Appl., 380 (2011), pp. 680–688.
- [16] P. COMBETTES AND J. ECKSTEIN, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions*, Math. Prog., 168 (2018), pp. 645–672.
- [17] L. CONDAT, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479.
- [18] F. DEUTSCH, *Best Approximation in Inner Product Spaces*, CMS Books Math., Springer, Berlin, 2001.
- [19] A. G. DIMAKIS, S. KAR, J. M. F. MOURA, M. G. RABBAT, AND A. SCAGLIONE, *Gossip algorithms for distributed signal processing*, Proc. IEEE, 98 (2010), pp. 1847–1864.
- [20] R. DYKSTRA, *An algorithm for restricted least-squares regression*, J. Amer. Statist. Assoc., 78 (1983), pp. 837–842.
- [21] J. ECKSTEIN AND B. SVAITER, *General projective splitting methods for sums of maximal monotone operators*, SIAM J. Control Optim., 48 (2009), pp. 787–811.
- [22] R. ESCALANTE AND M. RAYDAN, *Alternating Projection Methods*, SIAM, Philadelphia, PA, 2011.
- [23] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrange Methods: Applications to the Solution of Boundary Valued Problems, M. Fortin

- and R. Glowinski, eds., Stud. Math. Appl. 15, North Holland, Amsterdam, 1983, pp. 299–331.
- [24] N. GAFFKE AND R. MATHAR, *A cyclic projection algorithm via duality*, Metrika, 36 (1989), pp. 29–54.
 - [25] M. GÜRBÜZBALABAN, A. OZDAGLAR, AND P. PARRILO, *On the convergence rate of incremental aggregated gradient algorithms*, SIAM J. Optim., 27 (2017), pp. 1035–1048.
 - [26] S. HAN, *A successive projection method*, Math. Program., 40 (1988), pp. 1–14.
 - [27] H. HUNDAL AND F. DEUTSCH, *Two generalizations of Dykstra's cyclic projections algorithm*, Math. Program., 77 (1997), pp. 335–355.
 - [28] F. IUTZELER, P. BIANCHI, P. CIBLAT, AND W. HACHEM, *Asynchronous distributed optimization using a randomized alternating direction method of multipliers*, in Proceedings of the 52nd Conference on Decision Control, IEEE Press, Piscataway, NJ, 2013, pp. 3671–3676.
 - [29] S. LEE AND A. NEDIĆ, *Distributed random projection algorithm for convex optimization*, IEEE J. Sel. Topics Signal Process., 7 (2013), pp. 221–229.
 - [30] A. NEDICH, *Convergence rate of distributed averaging dynamics and optimization in networks*, Found. Trends Systems Control, 2 (2015), pp. 1–100.
 - [31] A. NEDICH, *Fast algorithms for Distributed Optimization over Time-Varying Graphs*, Talk at DIMACS Workshop on Distributed Optimization, Information Processing, and Learning, <https://www.nrel.gov/grid/assets/pdfs/aeg-nedich.pdf>, 2017.
 - [32] A. NEDICH AND A. OLSHEVSKY, *Distributed optimization over time-varying directed graphs*, IEEE Trans. Autom. Control, 60 (2015), pp. 601–615.
 - [33] A. NEDICH, A. OLSHEVSKY, AND W. SHI, *Achieving geometric convergence for distributed optimization over time-varying graphs*, SIAM J. Optim., 27 (2017), pp. 2597–2633.
 - [34] Y. NESTEROV, *A method for solving a convex programming problem with rate of convergence $O(\frac{1}{k^2})$* , Sov. Math. Dokl., 269 (1983), pp. 543–547 (in Russian).
 - [35] I. NOTARNICOLA AND G. NOTARSTEFANO, *Asynchronous distributed optimization via randomized dual proximal gradient*, IEEE Trans. Autom. Control, 62 (2017), pp. 2095–2106.
 - [36] A. OZDAGLAR, A. NEDIĆ, AND P. PARRILO, *Constrained consensus and optimization in multi-agent networks*, IEEE Trans. Autom. Control, 55 (2010), pp. 922–938.
 - [37] C. PANG, *The supporting halfspace: Quadratic programming strategy for the dual of the best approximation problem*, SIAM J. Optim., 26 (2016), pp. 2591–2619.
 - [38] C. PANG, *Dykstra splitting and an approximate proximal point algorithm for minimizing the sum of convex functions*, preprint, <https://arxiv.org/abs/1709.09499>, 2017.
 - [39] Z. PENG, Y. XU, M. YAN, AND W. YIN, *AROCK: An algorithmic framework for asynchronous parallel coordinate updates*, SIAM J. Sci. Comput., 38 (2016), pp. A2851–A2879.
 - [40] S. PHADE AND V. BORKAR, *A distributed Boyle–Dykstra–Han scheme*, SIAM J. Optim., 27 (2017), pp. 1880–1897.
 - [41] S. S. RAM, A. NEDIĆ, AND V. VEERAVALLI, *Distributed stochastic subgradient projection algorithms for convex optimization*, J. Optim. Theory Appl., 147 (2010), pp. 516–545.
 - [42] W. SHI, Q. LING, G. WU, AND W. YIN, *EXTRA: An exact first-order algorithm for decentralized consensus optimization*, SIAM J. Optim., 25 (2015), pp. 944–966.
 - [43] R. TIBSHIRANI, *Dykstra's algorithm, ADMM, and coordinate descent: Connections, insights, and extensions*, in Adv. Neural Inf. Process. Syst. 30, Curran Associates, Red Hook, NY, 2017, pp. 517–528.
 - [44] P. TSENG, *Convergence of a block coordinate descent method for nondifferentiable minimization*, J. Optim. Theory Appl., 109 (2001), pp. 475–494.
 - [45] P. TSENG, *On accelerated proximal gradient methods for convex-concave optimization*, manuscript, <https://www.mit.edu/~dimitrib/PTseng/papers/apgm.pdf>, 2008.
 - [46] B. VŨ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681.