A DUAL ASCENT ALGORITHM FOR ASYNCHRONOUS DISTRIBUTED OPTIMIZATION WITH UNRELIABLE DIRECTED COMMUNICATIONS

C.H. JEFFREY PANG

ABSTRACT. We show that the averaged consensus algorithm on directed graphs with unreliable communications by Bof-Carli-Schenato has a dual interpretation, which could be extended to the case of distributed optimization. We report on our numerical simulations for the distributed optimization algorithm for smooth and nonsmooth functions.

CONTENTS

1. Introduction 1
2. Algorithm derivation and description 2
3. Convergence analysis 5
4. Numerical experiments 9
5. Conclusion 10
References 10

1. INTRODUCTION

Let $G = (V, E)$ be a directed graph. Consider the distributed optimization problem

$$
\min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2 \right].
$$

Here, $f_i(\cdot)$ are closed convex functions. The challenge in distributed optimization is that the communications in the algorithm needs to obey the directed edges in the underlying graph. Note that if $f_i(\cdot)$ are the zero functions and $m = 1$, then the minimizer of (1.1) is exactly $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$, which is precisely the distributed averaging consensus.

A distributed asynchronous algorithm for averaged consensus on a directed graph with unreliable communications was designed in [BCS17]. The paper [BCS17] was inspired by two algorithms for averaged consensus in the literature. In an asynchronous setting, [BBT+10] introduced an algorithm that reaches averaged consensus using the so-called ratio consensus. The paper in [VHDG11] gave the idea of mass transfer used in [BCS17]. (Other papers also mentioned [HVDG16].) The paper...
also proved linear convergence of their algorithm, and pointed out algorithms in [VZC+16, CS10, DGH10] need the averaged consensus algorithm and its linear convergence as a building block. The work in [BCS17] has led to other strategies for distributed asynchronous optimization on directed and unreliable communications [BCN+17, TSDS18].

Ideally, one would want to solve the problem where the quadratic regularizers in (1.1) were removed. The ideas in this paper come from our related work on solving the distributed problem (1.1) on undirected graphs. In [Pan18a, Pan18b, Pan18c, Pan18d], we proposed a distributed asynchronous optimization algorithm. The idea behind those papers is that the problem (1.1) can be written as a product space formulation, and subsequently solved with Dykstra’s algorithm [Dyk83]. Dykstra’s algorithm is identical to block coordinate minimization on the dual [Han88], and is notable because the convergence to its primal minimizer does not rely on the existence of dual optimizers [BD85, GM89]. We were also motivated by these works, as well as [HD97] for the asynchronous operation of the algorithm. Some interesting properties of the algorithm in [Pan18a, Pan18b, Pan18c, Pan18d] include: being able work on time-varying graphs, allow for partial communication of data, allow for more than two of the $f_i(\cdot)$ to be indicator functions of closed convex sets (instead of being smooth functions), has deterministic convergence with rates mostly compatible with well known first order methods, and convergence to the primal solution even when there are no dual optimizers.

The distributed optimization algorithm in [Pan18a] extends the averaged consensus algorithm for undirected graphs. The dual objective value there acts as a potential function (or Lyapunov function) that decreases in each iteration. Another strategy is to use the averaged consensus algorithm as a building block of a distributed optimization algorithm, but one would be faced with the problem of deciding whether enough iterations for the averaged consensus algorithm had been run to achieve the desired result. Getting around such issues means having to make conservative estimates of certain constants or risk nonconvergence.

1.1. Contributions of this paper. In this paper, we propose a distributed algorithm on directed graphs with unreliable communications for the regularized optimization problem (1.1) which generalizes the algorithm in [BCS17]. We show that the dual objective value of (1.1) gives a potential function (or Lyapunov function) similar to that of [Pan18a] whose value is monotonically nonincreasing throughout the algorithm.

2. Algorithm derivation and description

In this section, we derive our algorithm. Those familiar with [BCS17] would recognize operations $A$ and $B$ in Algorithm 2.3, but operation $C$ there requires some preparation in the dual formulation.

Let $\tilde{m} = \frac{1}{|V|} \sum_{i \in V} \tilde{x}_i$. We have

$$\sum_{i \in V} \frac{1}{2} \left\| x - \tilde{x}_i \right\|^2 = \frac{|V|}{2} \left\| x - \tilde{m} \right\|^2 + \sum_{i \in V} \tilde{x}_i^T \tilde{x}_i - \left| V \right| \tilde{m}^T \tilde{m} \quad (2.1)$$
Algorithm 2.3, and Proposition 2.2. (Sparsity) The following results below hold:

\[
\sum_{\alpha \in V \cup E \cup \{r\}} s_{\alpha} = |V|, \quad \text{and} \quad s_{\alpha} \begin{cases} > 0 & \text{for all } \alpha \in V \\ \geq 0 & \text{for all } \alpha \in E \cup \{r\}. \end{cases}
\]

Let \( x \in [\mathbb{R}^m][V \cup E \cup \{r\}] \), and for all \( i \in V \), let \( f_i : [\mathbb{R}^m][V \cup E \cup \{r\}] \rightarrow \mathbb{R} \cup \{\infty\} \) be defined as \( f_i(x) = f_i([x]_i) \). Let the set \( F \) be

\[ F := \{(i, (i, j)) : (i, j) \in E\} \cup \{(j, (i, j)) : (i, j) \in E\} \cup \{(r, \alpha) : \alpha \in V \cup E\}. \]

and let the hyperplane \( H_{(\alpha_1, \alpha_2)} \), where \( \alpha_1, \alpha_2 \in F \), be defined by

\[ H_{(\alpha_1, \alpha_2)} := \{x \in [\mathbb{R}^m][V \cup E \cup \{r\}] : x_{\alpha_1} = x_{\alpha_2}\}. \]

We assume the underlying graph is strongly connected, so

\[ \cap_{\beta \in F} H_{\beta} = \{x \in [\mathbb{R}^m][V \cup E \cup \{r\}] : x_{\alpha_1} = x_{\alpha_2} \text{ for all } \alpha_1, \alpha_2 \in V \cup E \cup \{r\}\}. \]

The primal problem (1.1) can then be equivalently written in the product space formulation as

\[
\min_{x \in [\mathbb{R}^m][V \cup E \cup \{r\}] } \sum_{\alpha \in V \cup E \cup \{r\}} \frac{s_{\alpha}}{2} ||[x]_\alpha - \bar{m}||^2 + \sum_{i \in V} f_i(x) + \sum_{\beta \in F} \delta_{H_{\beta}}(x) + C, \tag{2.2}
\]

where \( C \) is as marked in (2.1). Any component of an optimal solution to (2.2) is an optimal solution to (1.1). The (Fenchel) dual of (2.2) can be calculated to be

\[
\sup_{\alpha \in V \cup E \cup \{r\}} \left[ \frac{|V|}{2} ||\bar{m}||^2 - \sum_{i \in V} \left( f_i^*(z_i) - \sum_{\beta \in F} \delta_{H_{\beta}}(z_{\beta}) - \sum_{\alpha \in V \cup E \cup \{r\}} \frac{s_{\alpha}}{2} \left( \bar{m} - \frac{1}{s_{\alpha}} \left[ \sum_{\alpha_2 \in V \cup E \cup \{r\}} z_{\alpha_2} \right] \right) \right) \right] \tag{2.3}
\]

The case when \( s_{\alpha} = 1 \) for all \( \alpha \in V \) and \( s_{\alpha} = 0 \) for all \( \alpha \in E \) has been discussed in detail in [Pan18a, Pan18b, Pan18c, Pan18d]. The treatment there implies that there is strong duality between (2.2) and (2.3), even if dual optimizers may not exist. For convenience, instead of considering (2.3), we consider

\[
\inf_{z_{\alpha} \in [\mathbb{R}^m][V \cup E \cup \{r\}] } \sum_{i \in V} f_i^*(z_i) + \sum_{\beta \in F} \delta_{H_{\beta}}(z_{\beta}) + \sum_{\alpha \in V \cup E \cup \{r\}} \frac{s_{\alpha}}{2} \left( \bar{m} - \frac{1}{s_{\alpha}} \left[ \sum_{\alpha_2 \in V \cup E \cup \{r\}} z_{\alpha_2} \right] \right)^2. \tag{2.4}
\]

**Remark 2.1.** (On the index \( r \)) Notice that \( s_r \) and \( y_r \) remain as zero throughout Algorithm 2.3 and \( z_{(r, \alpha)} \) also remains as zero for all \( \alpha \in V \cup E \) as well. We introduced this additional index \( r \) in order to simplify the convergence proof in Section 4.

We have the following properties:

**Proposition 2.2.** (Sparsity) The following results below hold:

1. If \( i \in V \), then \( z_i \in [\mathbb{R}^m][V \cup E \cup \{r\}] \) is such that \( [z_i]_{\alpha} = 0 \) for all \( \alpha \in [V \cup E \cup \{r\}] \setminus \{i\} \).
2. If \( \{\alpha_1, \alpha_2\} \in F \), then \( z_{(\alpha_1, \alpha_2)} \in [\mathbb{R}^m][V \cup E \cup \{r\}] \) is such that \( [z_{(\alpha_1, \alpha_2)}]_{\alpha} = 0 \) for all \( \alpha \in [V \cup E \cup \{r\}] \setminus \{\alpha_1, \alpha_2\} \), and \( [z_{(\alpha_1, \alpha_2)}]_{\alpha_1} + [z_{(\alpha_1, \alpha_2)}]_{\alpha_2} = 0 \).
As we have seen in [BCS17], the data that has been received by node $i$ can be rewritten to solve primal problems instead:

$$\delta H_{(\alpha_1, \alpha_2)}(\cdot) = \delta H^+_{(\alpha_1, \alpha_2)}(\cdot),$$

and $\delta H^+_{(\alpha_1, \alpha_2)}(z_{(\alpha_1, \alpha_2)}) < \infty$ implies the conclusions in (2).}

We now describe the Algorithm 2.3 on the following page. In order to link Algorithm 2.3 with the dual objective function (2.4), we define

$$y(i,j) := \sigma_{i,y} - \rho_{(i,j),y} \text{ for all } (i,j) \in E$$

(2.5a)

$$s(i,j) := \sigma_{i,s} - \rho_{(i,j),s} \text{ for all } (i,j) \in E$$

(2.5b)

$$x_\alpha := y_{\alpha}/s_\alpha \text{ for all } \alpha \in V \cup E \cup \{r\}.$$  

(2.5c)

As we have seen in [BCS17], the data $\sigma_{i,y}$ and $\sigma_{i,s}$ represent data transmitted by node $i$, and $\rho_{(i,j),y}$ and $\rho_{(i,j),s}$ represent data from node $i$ that has been received by node $j$ through the edge $(i,j)$. So $\sigma_{i,y} - \rho_{(i,j),y} \text{ and } \sigma_{i,s} - \rho_{(i,j),s}$ represent data that have been transmitted by node $i$ to node $j$ along edge $(i,j)$ that have not yet been received by node $j$. Hence using $y(i,j)$ and $s(i,j)$ to represent these data is natural. It is clear that if $s_\alpha = 0$, then $y_\alpha = 0$. In such a case, the choice of $x_\alpha$ is irrelevant. We want $\{z_\alpha\}_{\alpha \in V \cup F}$ to satisfy

$$m - \frac{1}{s_\alpha} \sum_{z_\alpha \in V \cup F} z_\alpha = x_\alpha \text{ for all } \alpha \in V \cup E \cup \{r\} \text{ such that } s_\alpha > 0.$$  

(2.6)

We now explain (2.6) further. Algorithm 2.3 starts with $s_\alpha^0 = 1$ if $\alpha \in V$, and zero otherwise, and $y^0_\alpha$ is such that $\frac{1}{|V|} \sum_{\alpha \in V} y^0_\alpha = \bar{m}$. So a possible choice of $y^0_\alpha$ is $\bar{x}_i$, as defined in (1.1). Recall $\{z_\alpha^0\}_{\alpha \in F}$ are to be defined to satisfy Proposition 2.2(2), and that as long as $\frac{1}{|V|} \sum_{\alpha \in V} y^0_\alpha = \bar{m}$, $\{z_\alpha^0\}_{\alpha \in F}$ can be chosen to satisfy (2.6). It is clear to see that operations $A$ and $B$ can be written as a composition of operations $D$ and $E$. In Theorem 3.1, we shall prove that throughout Algorithm 2.3, $\{z_\alpha\}_{\alpha \in F}$ can be chosen so that (2.6) is satisfied.

For now, the only new part in Algorithm 2.4 compared to [BCS17] is operation $C$. Let the tuple $T$ be defined as

$$T = \{(s_\alpha)_{\alpha \in V \cup E \cup \{r\}}, \{y_\alpha)_{\alpha \in V \cup E \cup \{r\}}, \{x_\alpha\}_{\alpha \in V \cup E \cup \{r\}}, \{z_\alpha\}_{\alpha \in V \cup F}\}.$$  

(2.7)

and define $T^k$ similarly. Define $\text{Val}(T)$ as

$$\text{Val}(T) := \sum_{i \in V} f_i^*(z_i) + \sum_{\beta \in F} \delta_{H^+}(z_\beta) + \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \|x_\alpha\|^2.$$  

It is quite clear that if $T^{k+1}$ is obtained from $T^k$ using operation $C$, then $\text{Val}(T^{k+1}) \leq \text{Val}(T^k)$. Through duality, the problem of finding new $[z_\alpha^k]$ and $y^k_\alpha$ in lines 12 and 13 can be rewritten to solve primal problems instead:

12. $x^k_j = \arg \min_{x_{\text{temp}}} \frac{s_{\text{temp}}^k}{2} \|x_{\text{temp}} - x\|^2 + f_j(x)$

13. $[z_\alpha^k]_j = s_{\text{temp}}^k (x_{\text{temp}} - x^k_j)$

In Section 3, we shall analyze operations $D$ and $E$ in order to draw conclusions about Algorithm 2.3.
Algorithm 2.3. (Main algorithm) We have the following algorithm.

Start with $V_0$ such that $\sum_{i \in V} y_0^i = |V|\overline{m}$, and $y_0^a = 0$ for all $a \in E \cup \{r\}$.
Start with $s_0^a$ such that $s_i = 1$ for all $i \in V$ and $s_0^a = 0$ for all $a \in E \cup \{r\}$.

Start with $z_a = 0$ for all $a \in V$.

Start with $\sigma_i^a = 0$ and $\sigma_i^a = 0$ for all $i \in V$.

Start with $\rho_i^{(r)} = 0$ and $\rho_i^{(s)} = 0$ for all $(i, j) \in E$.

For $k = 1, \ldots$

% Carry data from last iteration.
$y_k^i = y_{k-1}^i / \text{outdeg}(i) + 1$; $s_k^i := s_{k-1}^i / \text{outdeg}(i) + 1$

$\sigma_{k}^i = \sigma_{k-1}^i + y_k^i$; $\sigma_{k}^i = \sigma_{k-1}^i + s_k^i$.

Algorithm 2.4. (Operations A, B and C) We describe operations A, B and C:

01 A (Node i sends data)
02 Choose a node $i \in V$.
03 $y_k^i = y_{k-1}^i / \text{outdeg}(i) + 1$; $s_k^i := s_{k-1}^i / \text{outdeg}(i) + 1$
04 $\sigma_k^i = \sigma_{k-1}^i + y_k^i$; $\sigma_k^i = \sigma_{k-1}^i + s_k^i$.
05 B (Node j receives data from i)
06 Choose edge $(i, j) \in E$ so that $j$ receives data along $(i, j)$.
07 $y_k^j = y_{k-1}^j + \sigma_{k-1}^i - \rho_{(i, j)}^j$; $s_k^j = s_{k-1}^j + \sigma_{k-1}^i - \rho_{(i, j)}^j$.
08 $\rho_{(i, j)}^j = \sigma_{k-1}^i - \rho_{(i, j)}^j$.
09 C (Update $y_j$ and $|z_j|$ by minimizing dual function)
10 Choose a node $j \in V$.
11 $x_{\text{temp}} = \frac{1}{z_j}(y_k^j + |z_k^j|)$
12 $|z_k^j| := \arg \min_{z} \frac{1}{z} ||x_{\text{temp}} - \frac{1}{z_j}z||^2 + f_j^*(z)$
13 $y_k^j = \frac{s_k^j}{x_{\text{temp}}} - |z_k^j|$.

Algorithm 2.5. (Operations D and E) We describe operations D and E.

14 D (Split with $r$) Suppose $s_r^k = 0$.
15 Choose $\bar{a} \in V \cup E$.
16 Choose $s_{\bar{a}}^+ + s_{\bar{a}}^-$ to be such that $s_{\bar{a}}^+ + s_{\bar{a}}^- = s_r^k$.
17 Let $y_{\bar{a}}^+ = \frac{s_{\bar{a}}^+}{s_{\bar{a}}^+ + s_{\bar{a}}^-} y_{\bar{a}}^+$ and $y_{\bar{a}}^- = \frac{s_{\bar{a}}^-}{s_{\bar{a}}^+ + s_{\bar{a}}^-} y_{\bar{a}}^-$.

$18 s_{\bar{a}}^+ + s_{\bar{a}}^- + y_{\bar{a}}^+ = y_{\bar{a}}^- + y_{\bar{a}}^-$ for all $a \notin \{r, \bar{a}\}$, and $|z_k^a| = |z_k^a|$ for all $i \in V$.
19 E (Combine with $r$) Suppose $s_r^k > 0$.
20 Choose $\bar{a}_2 \in V \cup E$.
21 Let $s_{\bar{a}_2} = s_{\bar{a}_2}^+ + s_{\bar{a}_2}^-$ and $s_{\bar{a}}^+ = 0$.
22 Let $y_{\bar{a}_2}^+ = y_{\bar{a}_2}^+ + y_{\bar{a}}^+$ and $y_{\bar{a}_2}^- = 0$.
23 $s_{\bar{a}}^+ = s_{\bar{a}}^+ + y_{\bar{a}}^+$ for all $a \notin \{r, \bar{a}_2\}$, and $|z_k^a| = |z_k^a|$ for all $i \in V$.

3. Convergence analysis

In this section, we prove the convergence of Algorithm 2.3. We show that operations D and E result in a nonincreasing dual objective value $\text{Val}(\cdot)$, and that they
Proof. We assume throughout that (2.5c) holds throughout. We first look at Algorithm 2.4 and Operations D and E. Then the following hold.

Theorem 3.1. Consider the following conditions.

(A) Suppose the tuple T defined in is such that the relations (2.6) and (2.5c) are satisfied.
(B) \( z^\alpha_{\{r,\alpha\}} = 0 \) for all \( \alpha \in V \cup E \), \( s^\alpha = 0 \) and \( y^\alpha = x^\alpha \).
(C) There is some \( \bar{\alpha}_1 \in V \cup E \) such that \( z^\alpha_{\{r,\alpha\}} = 0 \) for all \( \alpha \neq \bar{\alpha}_1 \).

Then the following hold.

1. Suppose condition (A) is satisfied for the tuple
   \[ T^\circ = (\{s^\alpha\}_{\alpha \in \mathcal{V} \cup \mathcal{E} \cup \mathcal{r}}, \{y^\alpha\}_{\alpha \in \mathcal{V} \cup \mathcal{E} \cup \mathcal{r}}, \{x^\alpha\}_{\alpha \in \mathcal{V} \cup \mathcal{E} \cup \mathcal{r}}, \{z^\alpha\}_{\alpha \in \mathcal{V} \cup \mathcal{E} \cup \mathcal{F}}) \]  
   and condition (B) is satisfied at the start of operation \( D \) in Algorithm 2.4.
   Then we can find \( \{z^\alpha_{+}\}_{\alpha \in \mathcal{V} \cup \mathcal{F}} \) such that the tuple \( T^+ \) defined in a similar manner to (3.1) satisfies conditions (A) and (C). Moreover, \( \text{Val}(T^+) = \text{Val}(T^\circ) \).

2. Suppose condition (A) is satisfied for the tuple \( T^\circ \), condition (C) is satisfied at the start of operation \( E \), and \( \bar{\alpha}_1, \bar{\alpha}_2 \in \mathcal{F} \). Then we can find \( \{z^\alpha_{+}\}_{\alpha \in \mathcal{V} \cup \mathcal{F}} \) such that the tuple \( T^+ \) satisfies conditions (A) and (B). Moreover, \( \text{Val}(T^+) \leq \text{Val}(T^\circ) \).

Proof. We assume throughout that (2.5c) holds throughout. We first look at Operation D. Let \( z^\alpha_{\{r,\alpha\}} \) be such that

\[ [z^\alpha_{\{r,\alpha\}}\}_{\alpha \in \mathcal{V}} = 0 \text{ for all } \alpha \notin \{r, \bar{\alpha}\}, \quad [z^\alpha_{\{r,\alpha\}}\}_{\alpha \in \mathcal{V}} = -[z^\alpha_{\{r,\alpha\}}\}_{\alpha \in \mathcal{F}} \text{ for all } \alpha \notin \{r, \bar{\alpha}\}, \quad \text{and } \quad [z^\alpha_{\{r,\alpha\}}\}_{\alpha \in \mathcal{F}} = 0 \text{ for all } \alpha \notin \{r, \bar{\alpha}\}, \]

and let all other \( z^\alpha_{+} \) be equal to \( z^\alpha_{-} \). So we only need to check that (2.6) holds for \( x^\alpha_{+} \) and \( x^\alpha_{-} \). Note that

\[ \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}} \text{ Line 16, (3.2)} = \frac{1}{s^\alpha} \left( \left[ \sum_{\beta} z^\beta_{-} \right]_{\bar{\alpha}} - \frac{s^\alpha}{s^\beta} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}} \right) \text{ Line 16, (3.3)} = \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}}. \]

Hence

\[ \bar{m} - \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}} \text{ Line 3.3, (3.2)} = \bar{m} - \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{-} \right]_{\bar{\alpha}} \text{ Line 3.2} \]

\[ m - \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}} \text{ Line 4.3, (3.4)} = \bar{m} - \frac{1}{s^\alpha} \left[ \sum_{\beta} z^\beta_{+} \right]_{\bar{\alpha}}. \]

So \( x^\alpha_{+} \text{ Line 17, (2.5c, 2.6, 2.5c)} = \frac{1}{s^\alpha} \sum_{\beta} y^\alpha_{+} \text{ Line 17, (2.5c, 2.6, 2.5c)} = \frac{1}{s^\alpha} \sum_{\beta} y^\alpha_{+} \text{ Line 17, (2.5c, 2.6, 2.5c)} = \frac{1}{s^\alpha} \sum_{\beta} y^\alpha_{+} \text{ Line 17, (2.5c, 2.6, 2.5c)}. \]

which means (2.6) holds for \( x^\alpha_{+} \). Similarly, (2.6) holds for \( x^\alpha_{-} \). In fact, (3.4) also gives \( x^\alpha_{+} = x^\alpha_{-} = x^\alpha_{+} \), which gives

\[ \frac{s^\alpha}{2} \left[ x^\alpha_{+} \right]^2 + \frac{s^\alpha}{2} \left[ x^\alpha_{+} \right]^2 \text{ Line 16, (3.4, 2.5c, 2.5c, 2.5c, 2.5c, 2.5c)} \]

\[ \frac{s^\alpha}{2} \left[ x^\alpha_{+} \right]^2 \text{ Line 16, (3.4, 2.5c, 2.5c, 2.5c, 2.5c, 2.5c)} = \frac{s^\alpha}{2} \left[ x^\alpha_{+} \right]^2. \]
This in turn means $\text{Val}(T^+) = \text{Val}(T^\circ)$.

We now look at operation $E$. Suppose $v \in \mathbb{R}^m$ is such that

$$[z_{\{r, \tilde{a}_1\}}^o]_{\tilde{a}_1} = 0 \text{ for all } \alpha \notin \{r, \tilde{a}_1\}, [z_{\{r, \tilde{a}_1\}}^o]_r = v, \text{ and } [z_{\{r, \tilde{a}_1\}}^o]_{\tilde{a}_1} = -v. \quad (3.5)$$

We then construct $z_{\{\tilde{a}_1, \tilde{a}_2\}}^+$ by

$$[z_{\{\tilde{a}_1, \tilde{a}_2\}}^+]_{\tilde{a}_1} = 0 \text{ for all } \alpha \notin \{\tilde{a}_1, \tilde{a}_2\}, [z_{\{\tilde{a}_1, \tilde{a}_2\}}^+]_{\tilde{a}_2} = -v, \text{ and } [z_{\{\tilde{a}_1, \tilde{a}_2\}}^+]_{\tilde{a}_2} = v, \quad (3.6)$$

and $z_{\{r, \tilde{a}_1\}}^+ = 0$. All other $z_{\alpha}^+$ shall be equal to $z_{\alpha}^o$. Thus we only need to check $\text{(2.6)}$ for $x_{\tilde{a}_2}^\circ$. We have

$$\left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{\tilde{a}_2} + \left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{r} = \left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{\tilde{a}_2} + \left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{\tilde{a}_2} \quad (3.7)$$

Hence

$$y_{\tilde{a}_2}^+ = y_{\tilde{a}_2}^o + y_r^+ \quad \text{Line 22}$$

$$= y_{\tilde{a}_2}^o + y_r^+ \quad \text{Line 21} \quad s_{\tilde{a}_2}^o \left( m - \frac{1}{s_{\tilde{a}_2}^o} \left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{\tilde{a}_2} \right) \quad \text{Line 21} \quad s_{\tilde{a}_2}^o \left( m - \frac{1}{s_{\tilde{a}_2}^o} \left[\sum_{\beta} z_{\beta, \tilde{a}_2}^o \right]_{\tilde{a}_2} \right)$$

which, through $\text{(2.5c)}$, shows that $\text{(2.6)}$ holds for $x_{\tilde{a}_2}^+$. From the convexity of the norm-squared function $\| \cdot \|^2$, we have

$$\frac{s_{\tilde{a}_2}^o}{s_{\tilde{a}_2}^o + s_r^o} \| x_{\tilde{a}_2}^o \|^2 + \frac{s_r^o}{s_{\tilde{a}_2}^o + s_r^o} \| x_r^o \|^2 \geq \frac{s_{\tilde{a}_2}^o s_{\tilde{a}_2}^o s_r^o + s_r^o s_r^o}{s_{\tilde{a}_2}^o + s_r^o} \left\| \frac{y_{\tilde{a}_2}^o}{s_{\tilde{a}_2}^o} \right\|^2 \text{ Lines 21,22} \quad \frac{y_{\tilde{a}_2}^o}{s_{\tilde{a}_2}^o} \| x_{\tilde{a}_2}^+ \|^2 \quad \text{Lines 21,22}$$

The above inequality shows that $\text{Val}(T^+) \leq \text{Val}(T^\circ)$. \hfill \Box

3.2. **Convergence result.** In this subsection, we prove our convergence result.

Let $x^*$ be the optimal solution for $\text{(1.1)}$, and $x^* = \{x^*_\alpha\}_{\alpha \in V \cup E \cup \{r\}}$ be the optimal solution for $\text{(2.2)}$. It is clear that $x^*_\alpha = x^*$ for all $\alpha \in V \cup E \cup \{r\}$. We prove the boundedness of $\{x^k_\alpha\}_k$ for all $\alpha \in V \cup E \cup \{r\}$.

**Theorem 3.2.** (Boundedness of $\{x_\alpha\}$) Let $x^*$ be the optimal solution for $\text{(1.1)}$.

Suppose Algorithm $\text{(2.5)}$ is such that there is some $\bar{\epsilon} > 0$ such that $s_i > \bar{\epsilon}$ for all $i \in V$. Then the iterates $\{x^k_\alpha\}_k$ are bounded for all $\alpha \in V \cup E \cup \{r\}$.

**Proof.** Recall $x^*$ defined just before the statement of this result. From Fenchel duality, we have

$$f_i(x^*) + f^*_i(z_i) \geq \langle x^*, z_i \rangle \text{ and } \delta_{H_\beta}(x^*) + \delta_{H_\beta^+}(z_\beta) \geq 0. \quad (3.8)$$
Let $v_k^i := \frac{1}{\alpha_i} \left[ \sum_{\alpha \in V \cup F} z_{\alpha}^k \right]$. The duality gap (i.e., the difference between the objective values of (2.2) and (2.3)) gives

$$\sum_{\alpha \in V \cup U \cup \{r\}} \frac{s^k_{\alpha}}{2} \| x^* - \bar{m} \|^2 + \sum_{i \in V} f_i(x^*) + \sum_{\beta \in F} \delta_{H_{\beta}}(x^*) - \frac{|V|}{2} \| \bar{m} \|^2$$

$$+ \sum_{\alpha \in V \cup U \cup \{r\}} \frac{s^k_{\alpha}}{2} \| \bar{m} - \frac{1}{s_{\alpha}} v^k_{\alpha} \|^2$$

$$\leq \left< x^*, \sum_{\alpha \in V \cup F} z^k_{\alpha} \right> + \sum_{\alpha \in V \cup U \cup \{r\}} s^k_{\alpha} \left( \frac{1}{2} \| x^* - \bar{m} \|^2 + \frac{1}{2} \| \bar{m} - \frac{1}{s_{\alpha}} v^k_{\alpha} \|^2 - \frac{1}{2} \| \bar{m} \|^2 \right)$$

$$= \sum_{\alpha \in V \cup U \cup \{r\}} s^k_{\alpha} \left( \left< x^*, \frac{1}{s_{\alpha}} v^k_{\alpha} \right> + \frac{1}{2} \| x^* - \bar{m} \|^2 + \frac{1}{2} \| \bar{m} - \frac{1}{s_{\alpha}} v^k_{\alpha} \|^2 \right)$$

$$= \sum_{\alpha \in V \cup U \cup \{r\}} \frac{s^k_{\alpha}}{2} \left< x^* - \left( \bar{m} - \frac{1}{s_{\alpha}} v^k_{\alpha} \right) \right>^2 + \sum_{\alpha \in V \cup U \cup \{r\}} \frac{s^k_{\alpha}}{2} \left< \bar{m} - \frac{1}{s_{\alpha}} v^k_{\alpha} \right>^2. \quad (3.9)$$

The first two lines in (3.9) is nonincreasing due to Theorem 3.1. It is now clear from (3.9) that $\{x^k_i\}_k$ is bounded for all $i \in V$. For all edges $(i, j) \in E$, it is clear that $x^k_{(i,j)}$ lies in the convex hull of $\{x^k_{(i,j)}\}_{0 \leq k \leq K}$, so $\{x^k_{(i,j)}\}_k$ is bounded for all $(i, j) \in E$. A similar argument holds for $\{x^k_{(r)}\}_k$.

**Theorem 3.3.** (Convergence to dual objective value) Suppose there is some number $\bar{\epsilon} > 0$ such that $s^k_{\alpha} > \bar{\epsilon}$ for all $k \geq 0$ and $i \in V$. Assume that the iterates $\{\{z^k_{\alpha}\}_{i \in V}\}_k$ are bounded. Suppose that there is a number $K$ such that in $K$ consecutive iterations, operations $A$, $B$ and $C$ are carried out for all nodes $i \in V$ separately at least once. Then $\{\text{Val}(T^k)\}_k$ is nonincreasing, and its limit is the dual objective value of (2.4).

**Proof.** We consider the tuple $T^k := (\{s^k_{\alpha}\}_{\alpha \in V \cup E}, \{x^k_{\alpha}\}_{\alpha \in V \cup E}, \{z^k_{\alpha}\}_{i \in V})$. Since all these quantities are bounded, there is a subsequence $\{T^{k_i}\}_i$ such that $\lim_{i \to \infty} T^{k_i}$ exists. Taking subsequences if necessary, we can assume that $\lim_{i \to \infty} (T^{k_i}, T^{k_i+1})$ exists, and that the operation (either $A$, $B$ or $C$) to get $T^{k_i+1}$ from $T^{k_i}$ are all the same. Applying this procedure repeatedly shows that we can assume that

$$\lim_{i \to \infty} (T^{k_i}, T^{k_i+1}, \ldots, T^{k_i+K})$$

exists, and that the operations to get $T^{k_i+j}$ from $T^{k_i+j-1}$ are all the same for each $j \in \{1, \ldots, K\}$. For $j \in \{0, \ldots, K\}$, let the limits $\lim_{i \to \infty} T^{k_i+j}$ be

$$\bar{T}_j := (\{s^j_{\alpha}\}_{\alpha \in V \cup E}, \{\bar{x}^j_{\alpha}\}_{\alpha \in V \cup E}, \{\bar{z}^j_{\alpha}\}_{i \in V}).$$

From the continuity of the operations $A$, $B$ and $C$, the operations to get $\bar{T}_j$ from $\bar{T}_{j-1}$ must be the same as that of getting $T^{k_i+j}$ from $T^{k_i+j-1}$. Since $\{\text{Val}(T^k)\}_k$ is a nonincreasing sequence, we conclude that

$$\text{Val}(\bar{T}_0) = \text{Val}(\bar{T}_1) = \cdots = \text{Val}(\bar{T}_K). \quad (3.10)$$

Since $s^k_{\alpha} > \bar{\epsilon}$ for all $k \geq 0$ and $i \in V$, we have $\bar{s}^j_{\alpha} > \bar{\epsilon}$ for all $i \in V$ and $j \in \{0, \ldots, T\}$. From (3.10) and the fact that all operations $A$ and $B$ are continuous, we have that $\bar{x}^j_{\alpha}$ are all equal for $\alpha \in V \cup E$ and $j \in \{0, \ldots, K\}$, say $x^*$. Let $j_i$ be such that getting $\bar{T}_{j_i}$ from $\bar{T}_{j_i-1}$ involves operation $C$ on the index $i$. Such a $j_i$ must exist by
our assumptions. From (3.10), \( \bar{s}_i^j > 0 \), and the fact that operation \( C \) is continuous, we have that
\[
[\bar{z}_i^j]_i = \arg \min_{z_i} \frac{\bar{s}_i^j}{2} \left\| x^* + \frac{1}{\bar{s}_i^j} ([\bar{z}_i^j]_i - z_i) \right\|^2 + f_i^*(z_i) \text{ for all } i \in V,
\]
which means that \( x^* \in \partial f_i^*([\bar{z}_i^j]_i) \) for all \( i \). It is also clear to see that \([\bar{z}_i^j]_i\) is independent of \( j \). By making use of Operations \( D \) and \( E \), all the \( T_j \) can be transformed to some \( \tilde{T} \) where \( \hat{s}_i = 1 \) for all \( i \in V \), and \( \hat{s}_e = 0 \) for all \( e \in E \). From the discussion in Subsection 3.1, we can find some \( \{z_{\beta}\}_{\beta \in V \cup F} \) such that Val(\( \tilde{T} \)) is also equal to \( F(\{\bar{z}_\alpha\}_{\alpha \in V \cup F}) \), where
\[
F(\{\bar{z}_\alpha\}_{\alpha \in V \cup F}) := \sum_{\alpha \in V} f^*_1(z_i) + \sum_{\beta \in F} \delta_{H_{\beta}}(z_{\beta}) + \sum_{\alpha \in V} \frac{1}{\alpha} \left\| \bar{m} - \left[ \sum_{\alpha \in V \cup F} z_{\alpha} \right]_\alpha \right\|^2.
\]
To see that \( \tilde{T} \) is indeed an optimal solution to (2.4), note that our earlier discussions imply that 0 lies in the partial subdifferential of \( F(\cdot) \) with respect to the \( z_{\alpha} \) variable for all \( \alpha \in V \cup E \). Since \( F(\cdot) \) is the sum of a smooth function and a separable function, standard results on block coordinate minimization implies that \( \{\bar{z}_\alpha\}_{\alpha \in V \cup E} \) is a minimizer to the problem of minimizing \( F(\cdot) \).

Moreover, we note that the analysis in Pan18a (which traces back to GM89 and earlier) implies that there is strong duality between the problems (2.2) and (2.3), so the quantity \( \sum_{\alpha \in V \cup E \cup \{r\}} \frac{1}{\alpha} \|x^* - x_{\alpha}\|^2 \) in (3.9) converges to zero. Hence by Theorems 3.2 and 3.3, \( \lim_{k \to \infty} x_k^* \) exists and equals \( x^* \) for all \( i \in V \).

4. Numerical experiments

We conduct some simple experiments by looking at the case where \( m = 6 \) and the graph has 6 nodes and contains two cycles, \( 1 \to 2 \to 3 \to 5 \to 1 \) and \( 2 \to 4 \to 6 \to 2 \). Let \( e \) be ones(\( m,1 \)). First, we find \( \{v_i\}_{i \in V} \) and \( \bar{x} \) such that \( \sum_{i \in V} v_i |V(\bar{x} - e)| = 0 \). We then find closed convex functions \( f_i(\cdot) \) such that \( v_i \in \partial f_i(\bar{x}) \). It is clear from the KKT conditions that \( e \) is the primal optimum solution to (1.1) if \( \bar{x}_i = \bar{x} \) for all \( i \in V \).

We define \( f_i(\cdot) \) as functions of the following type:

(F-S) \( f_i(x) := \frac{1}{2} x^T A_i x + b_i^T x + c_i \), where \( A_i \) is of the form \( vv^T + rI \), where \( v \) is generated by \( \text{rand}(\mathbf{m},1) \), \( r \) is generated by \( \text{rand}(1) \). \( b_i \) is chosen to be such that \( v_i = \nabla f_i(e) \), and \( c_i = 0 \).

(F-NS) \( f_i(x) := \max\{f_{i,1}(x), f_{i,2}(x)\} \), where \( f_{i,j}(x) := \frac{1}{2} x^T A_i x + b_i^T x + c_i \) for \( j \in \{1,2\} \), \( A_i \) is of the form \( vv^T + rI \), where \( v \) is generated by \( \text{rand}(\mathbf{m},1) \), \( r \) is generated by \( \text{rand}(1) \). \( b_{i,1} \) and \( b_{i,2} \) are chosen such that \( v_i = \frac{1}{2} [\nabla f_{i,1}(e)] + [\nabla f_{i,2}(e)] \) but \( v_i \) is neither \( \nabla f_{i,1}(e) \) nor \( \nabla f_{i,2}(e) \), and \( c_{i,1} \) and \( c_{i,2} \) are chosen such that \( f_{i,1}(e) = f_{i,2}(e) \).

Note the algorithms in BCN17, TSDS18 do not handle nonsmooth functions. Also, our algorithm does not require one to choose parameters to be small enough in order to achieve convergence.

We conduct two experiments, one when all functions are of the form (F-S), and another when the functions are all of the form (F-NS). The first and last formulas of (3.9) indicate how fast the primal iterates \( \{x_\alpha\}_{\alpha \in V \cup F} \) are converging.
to the optimal solution $x^*$, and we call these values the “duality gap” and the “norms squared weighted sum.” Figure 4.1 shows a plot of the results obtained by a random experiment where we perform 1000 iterations of the smooth case and 50,000 iterations of the nonsmooth case. The results observed are quite similar to that in [Pan18c]. Specifically, if all the functions $f_i(\cdot)$ are of the form (F-S), then we observe linear convergence (though we have not proved this yet). If all functions $f_i(\cdot)$ are of the form (F-NS), then we observe sublinear convergence. Rather often, this sublinear convergence is seen to be of order $O(1/k)$.

5. Conclusion

To conclude, we make a few observations. The insight that the algorithm in [BCS17] can be written as a dual ascent optimization problem shows that other ideas in algorithm design that were already laid out in the related papers [Pan18a, Pan18b, Pan18c, Pan18d] on a distributed Dykstra’s algorithm, as well as [Pan16], can be incorporated into the algorithm in this paper. It is also straightforward to design an improved algorithm for the case when some of the edges in the graph are undirected while others are directed using Operations $D$ and $E$. We defer the proof of linear convergence of the case when the functions $f_i(\cdot)$ in (1.1) are smooth to a future paper. It would be of interest to incorporate the algorithmic and theoretical properties known for the case of undirected graphs to the case of directed graphs with unreliable communications.

References

[AH16] N.S. Aybat and E.Y. Hamedani, A primal-dual method for conic constrained distributed optimization problems, Advances in Neural Information Processing Systems 29, Curran associates, Red Hook, NY, 2016, pp. 5049–5057.

[BBT+10] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, Weighted gossip: Distributed averaging using non-doubly stochastic matrices, Proceedings ISIT 10, EPFL-CONF-148711. IEEE, 2010.

[BCN+17] N. Bof, R. Carli, G. Notarstefano, L. Schenato, and D. Varagnolo, Newton-Raphson consensus under asynchronous and losey communications for peer-to-peer networks, 2017.

[BCS17] N. Bof, R. Carli, and L. Schenato, Average consensus with asynchronous updates and unreliable communication, Proc. of the IFAC World Congress, 2017, pp. 601–606.
[BD85] J.P. Boyle and R.L. Dykstra, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, Advances in Order Restricted Statistical Inference, Lecture notes in Statistics, Springer, New York, 1985, pp. 28–47.

[CS10] F.S. Cattivelli and A.H. Sayed, *Diffusion strategies for distributed kalman filtering and smoothing*, IEEE Transactions on automatic control 55 (2010), no. 9, 2069–2084.

[DGH10] A.D. Domínguez-Garcia and C.N. Hadjicostis, *Coordination and control of distributed energy resources for provision of ancillary services*, First IEEE International Conference on In Smart Grid Communications (SmartGridComm), 2010, pp. 537–542.

[Dyk83] R.L. Dykstra, *An algorithm for restricted least-squares regression*, J. Amer. Statist. Assoc. 78 (1983), 837–842.

[GM89] N. Gaffke and R. Mathar, *A cyclic projection algorithm via duality*, Metrika 36 (1989), 29–54.

[Han88] S.P. Han, *A successive projection method*, Math. Programming 40 (1988), 1–14.

[HD97] H.S. Hundal and F. Deutsch, *Two generalizations of Dykstra’s cyclic projections algorithm*, Math. Programming 77 (1997), 335–355.

[HVDG16] C.N. Hadjicostis, N.H. Vaidya, and A.D. Domínguez-Garcia, *Robust distributed average consensus via exchange of running sums*, IEEE Trans. Automat. Contr. 31 (2016), no. 6, 1492–1507.

[LN13] S. Lee and A. Nedić, *Distributed random projection algorithm for convex optimization*, IEEE Journal of Selected Topics in Signal Processing 7 (2013), 221–229.

[ONP10] A. Ozdaglar, A. Nedić, and P. Parrilo, *Constrained consensus and optimization in multi-agent networks*, IEEE Trans. Automat. Control 55 (2010), 922–938.

[Pan18a] C.H.J. Pang, *The supporting halfspace - quadratic programming strategy for the dual of the best approximation problem*, SIAM J. Optim. 26 (2016), no. 4, 2591–2619.

[Pan18b] C.H.J. Pang, *Subdifferentiable functions and partial data communication in a distributed deterministic asynchronous Dykstra’s algorithm*, 2018.

[Pan18c] C.H.J. Pang, *Linear and sublinear convergence rates for a subdifferentiable distributed deterministic asynchronous Dykstra’s algorithm*, 2018.

[Pan18d] C.H.J. Pang, *Convergence rate of distributed Dykstra’s algorithm with sets defined as level sets of convex functions*, 2018.

[PB17] S.R. Phade and V.S. Borkar, *A distributed Boyle-Dykstra-Han scheme*, SIAM J. Optim. 27 (2017), no. 3, 1880–1897.

[RNV10] S. Sundar Ram, A. Nedić, and V.V. Veeravalli, *Distributed stochastic subgradient projection algorithms for convex optimization*, J. Optim. Theory and Appl. 147 (2010), 516–545.

[TSDS18] Y. Tian, Y. Sun, B. Du, and G. Scutari, *ASY-SONATA: Achieving geometric convergence for distributed asynchronous optimization*, arxiv preprint 1803.10359v1, 2018.

[VHDG11] N.H. Vaidya, C.N. Hadjicostis, and A.D. Domínguez-Garcia, *Distributed algorithms for consensus and coordination in the presence of packet-dropping communication links-part II: Coefficients of ergodicity analysis approach*, arXiv preprint arXiv:1109.6392, 2011.

[VZC+16] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, *Newton-Raphson consensus for distributed convex optimization*, IEEE Transactions on Automatic Control 61 (2016), no. 4, 994–1009.

Current address: Department of Mathematics, National University of Singapore, Block S17 08-11, 10 Lower Kent Ridge Road, Singapore 119076

E-mail address: matpchj@nus.edu.sg