A distributed, asynchronous, deterministic optimization algorithm

An approach using Dykstra’s splitting/block coordinate dual ascent, extending to graphs with directed unreliable edges, majorizations to increase applicability, and their convergence rates

C.H. Jeffrey PANG

August 1, 2019
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges, and interpretation using optimization
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
Consider the graph $G = (V, E)$. For all $i \in V$, $f_i : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is a closed convex function.
Consider the graph $G = (V, E)$. For all $i \in V$, $f_i : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is a closed convex function.

Problem of interest:

$$
\min_{x \in \mathbb{R}^m} \sum_{i \in V} f_i(x),
$$

where computations are done in a \textbf{distributed} manner:

- Each agent $i \in V$ communicates only with its neighbors.
Consider the graph $G = (V, E)$. For all $i \in V$, $f_i : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is a closed convex function.

Problem of interest:

$$\min_{x \in \mathbb{R}^m} \sum_{i \in V} f_i(x),$$

where computations are done in a **distributed** manner:

- Each agent $i \in V$ communicates only with its neighbors.

Challenges:

- When there can’t exist a central node in $G$.
- When communications are directed and/or unreliable.
- Need to limit communications (size or privacy issues).
- Problems where $|V|$ is large.
  - (wireless network, internet-of-things)

For $\bar{x}_i \in \mathbb{R}^m$ given in advance for all $i \in V$, consider instead the regularized problem

$$\min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2 \right]. \quad \text{(P1)}$$
Consider first averaged consensus problem

- Given $\bar{x}_i$ on each node, find the average $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$ in a distributed manner.

The natural algorithm for averaged consensus is to pick an edge, and then average the values on the end points.
Consider first **averaged consensus** problem

- Given $\bar{x}_i$ on each node, find the average $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$ in a distributed manner.

The natural algorithm for averaged consensus is to pick an edge, and then average the values on the end points.
Consider first \textit{averaged consensus} problem
\begin{itemize}
\item Given $\bar{x}_i$ on each node, find the average $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$ in a distributed manner.
\end{itemize}

The natural algorithm for averaged consensus is to pick an edge, and then average the values on the end points.
Consider first **averaged consensus problem**

- Given $\bar{x}_i$ on each node, find the average $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$ in a distributed manner.

The natural algorithm for averaged consensus is to pick an edge, and then average the values on the end points.

If edges don’t overlap, then averaging operations independent.

- This algorithm is asynchronous.

- Well known linear convergence rate, with rates depending on graph properties. [Boyd- Ghosh- Prabhakar- Shah ’06, Dimakis- Kar- Moura- Rabbat- Scaglione ’10.]
Consider first **averaged consensus problem**

- Given $\bar{x}_i$ on each node, find the average $\frac{1}{|V|} \sum_{i \in V} \bar{x}_i$ in a distributed manner.

The averaged consensus problem can also be written as

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

We will generalize averaged consensus to

$$\min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2 \right].$$
Some distributed optimization algorithms use averaged consensus as a subroutine. [Aybat-Hamedani '16], [Tian-Sun-Du-Scutari '18], SONATA [Scutari-Sun '19].
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges, and interpretation using optimization
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
Some distributed optimization algorithms use averaged consensus as a subroutine. [Aybat-Hamedani ’16], [Tian-Sun-Du-Scutari ’18], SONATA [Scutari-Sun ’19].

Recent work on the averaged consensus problem for graphs with directed unreliable edges [Bof, Carli, Schenato ’17].

Some distributed optimization algorithms use averaged consensus as a subroutine. [Aybat-Hamedani ’16], [Tian-Sun-Du-Scutari ’18], SONATA [Scutari-Sun ’19].

Recent work on the averaged consensus problem for graphs with **directed unreliable edges** [Bof, Carli, Schenato ’17].


We now present algorithm of [Bof, Carli, Schenato ’17]. (We’ll extend it to distributed optimization later)

- Starting conditions: $y_i = x_i$ and $s_i = 1$ for all $i \in V$, $y(i,j) = 0$ and $s(i,j) = 0$ for all $(i,j) \in E$
Directed unreliable edges Averaged consensus:

(Push) Node $i$ transmits:

\[
y^{k+1}_i = y^k_i / [\deg_{out}(i) + 1] \quad \text{and} \quad s^{k+1}_i = s^k_i / [\deg_{out}(i) + 1]
\]

\[y_{(i,j)}^{k+1} = y_{(i,j)}^k + y_{(i,j)}^{k+1}\]

(Sum) Node $j$ receives from $(i, j)$:

\[
y^{k+1}_j = y^k_j + y_{(i,j)}^k \quad \text{and} \quad s^{k+1}_j = s^k_j + s_{(i,j)}^k
\]

\[y_{(i,j)}^k = 0 \quad \text{and} \quad s_{(i,j)}^k = 0\]

- $y_{(i,j)}$ & $s_{(i,j)}$ are data sent from node $i$ not yet received by node $j$.
- Also, $\sum_{\alpha \in V \cup E} y_{\alpha}$ and $\sum_{\alpha \in V \cup E} s_{\alpha}$ remain constant.
[Bof-Carli-Schenato '17] Directed unreliable edges

Averaged consensus:

(Push) Node $i$ transmits:

$$y_{i}^{k+1} = y_{i}^{k} / [\text{deg}_{out}(i) + 1]$$
$$y_{(i,j)}^{k+1} = y_{(i,j)}^{k} + y_{i}^{k+1}$$

and $s_{i}^{k+1} = s_{i}^{k} / [\text{deg}_{out}(i) + 1]$ and $s_{(i,j)}^{k+1} = s_{(i,j)}^{k} + s_{i}^{k+1}$ for all $j \in N_{out}(i)$,

(Sum) Node $j$ receives from $(i, j)$:

$$y_{j}^{k+1} = y_{j}^{k} + y_{(i,j)}^{k}$$
$$y_{(i,j)}^{k+1} = 0$$
$$s_{j}^{k+1} = s_{j}^{k} + s_{(i,j)}^{k}$$
and $s_{(i,j)}^{k+1} = 0$

• **Unreliable edges**: when node $j$ (eg, node 4) doesn’t receive, data is not lost, only delayed.
Finally, [Bof-Carli-Schenato ’17] uses idea of running sums in [Vaidya, Hadjicostis, Dominguez-Garcia ’11].

Recall that $\sum_{\alpha \in V \cup E} y_\alpha$ and $\sum_{\alpha \in V \cup E} s_\alpha$ preserved throughout.

[Bof-Carli-Schenato ’17] proved that if data isn’t stuck,

- The ratios $y_\alpha / s_\alpha$ converges to the average $x^* := \frac{1}{|V|} \sum_{i \in V} x_i$
- Linear convergence to $x^*$. (Proof is by ergodic theory)
  - We’ll generalize with optimization theory later.
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- **Further technical material.**

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges, and interpretation using optimization
- Various majorizations for when \( f_i(\cdot) \) are not proximable
- Convergence rates
- Simple numerical experiments

Primal methods (requires all but one $f_i(\cdot)$ smooth, except †.)
1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17, (ASY-)SONATA Scutari-Sun ’19)
   - (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
† (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.
2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.

Primal dual methods and monotone operators
1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.
2. 3+ Operator splitting (Eckstein and collaborators ’09–’18) violates 2.
3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
   - They may get around 3 by using averaged consensus as subroutine.
1. Distributed  
2. Decentralized  
3. Asynchronous  
4. Deterministic  
5. Multiple proximal terms

---

Primal methods (requires all but one $f_i(\cdot)$ smooth, except †.)

1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17, (ASY-)SONATA Scutari-Sun ’19)

- (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.

† (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.

2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.


Primal dual methods and monotone operators

1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.

2. 3+ Operator splitting (Eckstein and collaborators ’09-’18) violates 2.

3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.

- They may get around 3 by using averaged consensus as subroutine.

Primal methods (requires all but one $f_i(\cdot)$ smooth, except $\dagger$.)

1. Subgradient methods (Nedic-Olshevsky '15, EXTRA Shi-Ling-Wu-Yin '15, DIGing Nedich-Olshevsky-Shi '17, (ASY-)SONATA Scutari-Sun '19)
   - (Duchi-Agarwal-Wainwright '12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
   $\dagger$ (Lee-Nedich '13): some $f_i(\cdot)$ can be indicator functions, but violates 4.

2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo '17, Aytekin-Feyzmahdavian-Johansson '16) violates 2.


Primal dual methods and monotone operators

1. AROCK (Peng-Xu-Yan-Yin '16) (Monotone operators) violates 4.

2. 3+ Operator splitting (Eckstein and collaborators '09-'18) violates 2.

3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
   - They may get around 3 by using averaged consensus as subroutine.

Primal methods (requires all but one $f_i(\cdot)$ smooth, except †.)
1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17, (ASY-)SONATA Scutari-Sun ’19)
   - (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
† (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.
2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.

Primal dual methods and monotone operators
1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.
2. 3+ Operator splitting (Eckstein and collaborators ’09-’18) violates 2.
3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
   - They may get around 3 by using averaged consensus as subroutine.
5. Multiple proximal terms

**Primal methods** (requires all but one $f_i(\cdot)$ smooth, except †.)
1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17, (ASY-)SONATA Scutari-Sun ’19)
   - (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
   † (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.
2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.

**Primal dual methods and monotone operators**
1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.
2. 3+ Operator splitting (Eckstein and collaborators ’09-’18) violates 2.
3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
   - They may get around 3 by using averaged consensus as subroutine.
1. Distributed  
2. Decentralized  
3. Asynchronous  
4. Deterministic  
5. Multiple proximal terms

Primal methods (requires all but one $f_i(\cdot)$ smooth, except †.)
1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA
Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17,
(ASY-)SONATA Scutari-Sun ’19)
- (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
† (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.
2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.

Primal dual methods and monotone operators
1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.
2. 3+ Operator splitting (Eckstein and collaborators ’09-’18) violates 2.
3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
- They may get around 3 by using averaged consensus as subroutine.

Primal methods (requires all but one $f_i(\cdot)$ smooth, except †.)

1. Subgradient methods (Nedic-Olshevsky ’15, EXTRA Shi-Ling-Wu-Yin ’15, DIGing Nedich-Olshevsky-Shi ’17, (ASY-) SONATA Scutari-Sun ’19)
   - (Duchi-Agarwal-Wainwright ’12) has all properties, but slow with $O(1/k^{1/2})$ convergence, even for strongly convex problems.
   † (Lee-Nedich ’13): some $f_i(\cdot)$ can be indicator functions, but violates 4.

2. Incremental gradients and variants (Gurbuzbalaban-Ozdaglar-Parillo ’17, Aytekin-Feyzmahdavian-Johansson ’16) violates 2.


Primal dual methods and monotone operators

1. AROCK (Peng-Xu-Yan-Yin ’16) (Monotone operators) violates 4.

2. 3+ Operator splitting (Eckstein and collaborators ’09–’18) violates 2.

3. Algorithms based on Chambolle-Pock primal-dual algorithm and ADMM usually violate 3.
   - They may get around 3 by using averaged consensus as subroutine.

Dual/ Dykstra’s splitting (Notarnicola-Notarstefano ’15, P. ’18)

\[ \min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \| x - \bar{x}_i \|^2]. \quad (P1) \]

- Regularization term \( \frac{1}{2} \| x - \bar{x}_i \|^2 \) needed in our proofs.
  - Compared to other algorithms, not such a big drawback
  - Many distributed algorithms need strong convexity somewhere
- (Notarnicola-Notarstefano ’15) has more general formulation, but didn’t notice 4.
- We show our work on the dual formulation:
  - In short, show that it has many favourable properties.
  - We call this algorithm the distributed Dykstra’s algorithm.
To minimize

\[
\begin{align*}
\text{smooth and convex:} & \quad f(x_1, \ldots, x_n) + \\
\text{separable and convex:} & \quad \sum_{i=1}^{n} g_i(x_i)
\end{align*}
\]

a well known method is block coordinate minimization:

- Fix \( i \), hold \( x_j \) for all \( j \neq i \) fixed, and minimize w.r.t. only \( x_i \).

Example:

\[
\begin{align*}
x_1^* &= \arg\min_x \left[ f(x, x_2, x_3) + g_1(x) + g_2(x_2) + g_3(x_3) \right] \\
x_3^* &= \arg\min_x \left[ f(x_1^*, x_2, x) + g_1(x_1^*) + g_2(x_2) + g_3(x) \right] \\
x_2^* &= \arg\min_x \left[ f(x_1^*, x, x_3^*) + g_1(x_1^*) + g_2(x) + g_3(x_3^*) \right]
\end{align*}
\]
For averaged consensus

\[ \min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ \frac{1}{2} \| x - \bar{x}_i \|^2 \right], \quad (\text{Avr-con}) \]

a (Fenchel) dual interpretation shows that an averaging operation reduces the dual objective function

\[
\min_{z_{\alpha} \in \mathbb{R}^m | \forall \alpha} F(\{z_{\alpha}\}_{\alpha \in E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in E} z_{\alpha} \right\|^2_i + \sum_{(i,j) \in E} \delta_{H(i,j)}^* (z_{(i,j)}) = [x]_i
\]
For averaged consensus

$$\min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ \frac{1}{2} \| x - \bar{x}_i \|_2^2 \right],$$  \hspace{1cm} (Avr-con)

a (Fenchel) dual interpretation shows that an averaging operation reduces the dual objective function

$$\min_{\alpha \in E} F(\{ z_{\alpha} \}_{\alpha \in E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in E} z_{\alpha} \right\|_2^2 + \sum_{(i,j) \in E} \delta^*_H(i,j) (z_{(i,j)})$$

=\[x\]_i

Specifically, minimizing over one (implicit) variable \(z_{(i,j)}\) corresponds to averaging \([x]_i\) & \([x]_j\), reducing \(F(\{ z_{\alpha} \}_{\alpha \in E})\).
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges, and interpretation using optimization
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
For the distributed optimization problem

$$\min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2],$$  \hspace{1cm} (P1)$$

we, in [Pang, 18], propose to minimize the dual function

$$\min_{z_{\alpha} \in [\mathbb{R}^m]_{\alpha \in V \cup E}} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha} \right\|_i^2 + \sum_{(i,j) \in E} \delta_{H(i,j)}^* (z_{i,j}) + \sum_{i \in V} f_i^*(z_i)$$

$$= [x]_i$$
For the distributed optimization problem

\[ \min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2}\|x - \bar{x}_i\|^2], \quad (P1) \]

we, in [Pang, 18], propose to minimize the dual function

\[
\min_{z_\alpha \in \mathbb{R}^m | V} \sum_{\alpha \in V \cup E} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|_i^2 + \sum_{(i,j) \in E} \delta_{H(i,j)}(z_{i,j}) + \sum_{i \in V} f_i^*(z_i)
\]

\[= [x]_i \]

1. Like before, minimizing over one (implicit) variable \(z_{i,j}\) still corresponds to averaging \([x]_i\) & \([x]_j\), reducing \(F(\{z_\alpha\}_{\alpha \in E})\).

![Graph with nodes and edges representing the optimization problem.](image)
For the distributed optimization problem
\[
\min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2], \quad (P1)
\]
we, in [Pang, 18], propose to minimize the dual function
\[
\min_{z_{\alpha} \in [\mathbb{R}^m]_{|V|}} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha} \right\|^2 + \sum_{(i,j) \in E} \delta_{H(i,j)}(z(i,j)) + \sum_{i \in V} f_i^*(z_i),
\]

2. But now, minimizing over a variable $z_i$ is a proximal operation
\[
\min_{[z_i] \in \mathbb{R}^m} \frac{1}{2} \| [x]_i^{\text{old}} + [z_i]_i^{\text{old}} - [z_i]_i \|^2 + f_i^*([z_i]_i), \text{ reducing } F(\{z_{\alpha}\}_{\alpha \in V \cup E}).
\]
Features of algorithm (page 1 of 2):

- **Reasonable Storage**: Each node stores \([x]_i, [z]_i \in \mathbb{R}^m\).
Features of algorithm (page 1 of 2):

- **Reasonable Storage**: Each node stores \([x]_i, [z_i]_i \in \mathbb{R}^m\).

\[
\min_{\{z_\alpha\}_{\alpha \in V \cup E}} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x}_i - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z_{i,j}) + \sum_{i \in V} f_i^*(z_i)
\]

- **Scalable**: For greedier decrease, minimize more variables.
Features of algorithm (page 1 of 2):

- **Reasonable Storage**: Each node stores $[x]_i, [z_i]_i \in \mathbb{R}^m$.

\[
\min_\substack{z_\alpha \in \mathbb{R}^m | V \alpha \in V \cup E} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|_2^2 + \sum_{(i,j) \in E} \delta^*_H(z_{i,j}) + \sum_{i \in V} f_i^*(z_i)
\]

- **Scalable**: For greedier decrease, minimize more variables.

- **Potential function**/ Lyapunov function for progress.
Features of algorithm (page 1 of 2):

- **Reasonable Storage**: Each node stores $[x]_i, [z]_i \in \mathbb{R}^m$.

$$
\min_{z_{\alpha} \in [\mathbb{R}^m]_{|V|}} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha} \right\|^2_i + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z_{i,j}) + \sum_{i \in V} f^*_i(z_i)
$$

- **Scalable**: For greedier decrease, minimize more variables.
- **Potential function**: Lyapunov function for progress.
- **Distributed, Asynchronous, Decentralized**.
  - Computations in one component independent of others
  - Asynchronous: ideas of proof from [Hundal-Deutsch ’97].
Features of algorithm (page 1 of 2):

- **Reasonable Storage**: Each node stores \([x]_i, [z_i]_i \in \mathbb{R}^m\).

\[
\min_{z_\alpha \in [\mathbb{R}^m]^V \mid \alpha \in V \cup E} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|_2^2 + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z_{(i,j)}) + \sum_{i \in V} f_i^*(z_i)
\]

- **Scalable**: For greedier decrease, minimize more variables.

- **Potential function/Lyapunov function for progress**.

- **Distributed, Asynchronous, Decentralized**.
  - Computations in one component independent of others
  - Asynchronous: ideas of proof from [Hundal-Deutsch '97].

- **Dual method can treat sum of many proximable functions**.
  - Primal only methods require all \(f_i(\cdot)\) to be differentiable.
Features of algorithm (page 2 of 2):

- **Deterministic convergence** of all $[x]_i$ to minimizer $x^*$, and **constraint qualification free** convergence
  - Proof in [Pang, 18] adapted from that of Dykstra’s algorithm of [Boyle-Dykstra ’86], [Gaffke-Mathar ’88].
Features of algorithm (page 2 of 2):

- **Deterministic convergence** of all $[x]_i$ to minimizer $x^*$, and **constraint qualification free** convergence.
  - Proof in [Pang, 18] adapted from that of Dykstra’s algorithm of [Boyle-Dykstra ’86], [Gaffke-Mathar ’88].

- **Time-varying graphs** [Pang, 18]
  - Convergence if there is $K > 0$ s.t. in $K$ consecutive iterations, enough of edge variables $\{z_e\}_{e \in E}$ chosen connect the graph.
  - Adjusted from [Gaffke-Mathar ’88]’s proof.
Features of algorithm (page 2 of 2):

- **Deterministic convergence** of all $[x]_i$ to minimizer $x^*$, and **constraint qualification free** convergence
  - Proof in [Pang, 18] adapted from that of Dykstra’s algorithm of [Boyle-Dykstra ’86], [Gaffke-Mathar ’88].
  
  $$[x]_1, [z_1]_1 \in \mathbb{R}^m \quad [x]_5, [z_5]_5 \in \mathbb{R}^m$$

  $$[x]_2, [z_2]_2 \in \mathbb{R}^m \quad [x]_6, [z_6]_6 \in \mathbb{R}^m$$

  $$[x]_3, [z_3]_3 \in \mathbb{R}^m \quad [x]_7, [z_7]_7 \in \mathbb{R}^m$$

  $$[x]_4, [z_4]_4 \in \mathbb{R}^m \quad [x]_8, [z_8]_8 \in \mathbb{R}^m$$

- **Time-varying graphs** [Pang, 18]
  - Convergence if there is $K > 0$ s.t. in $K$ consecutive iterations, enough of edge variables $\{z_e\}_{e \in E}$ chosen connect the graph.
  - Adjusted from [Gaffke-Mathar ’88]’s proof.

- **Partial communication** of data
  - For $m$ big, need only average $\ll m$ components of data.
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- **Extension for directed unreliable edges**
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
To extend [Bof-Carli-Schenato ’17]'s distributed optimization on directed unreliable graphs on \( \min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2] \),

- Let \( x_\alpha = \frac{y_\alpha}{s_\alpha} \), and minimize the dual objective function

\[
\min_{z_\alpha \in [\mathbb{R}^m]|V \cup E|, \alpha \in V \cup F} \sum_{\alpha \in V \cup E} s_\alpha \left\| \bar{x} - \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right\|^2 + \sum_{\beta \in F} \delta_{H_\beta}(z_\beta) + \sum_{i \in V} f_i^*(z_i) \\
= F(\{z_\alpha\}_{\alpha \in V \cup F};\{s_\alpha\}_{\alpha \in V \cup E})
\]

- Changes: \( \{s_\alpha\}_{\alpha \in V \cup E}, z_\alpha \in [\mathbb{R}^m]|V \cup E|, \alpha \in [V \cup E] \cup F \)
To extend [Bof-Carli-Schenato ’17]’s distributed optimization on directed unreliable graphs on \( \min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2] \),

- Let \( x_\alpha = \frac{y_\alpha}{s_\alpha} \), and minimize the dual objective function

\[
\min_{z_\alpha \in [\mathbb{R}^m]_{V \cup E}} \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| \bar{x} - \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right\|^2 + \sum_{\beta \in F} \delta^*_H(z_\beta) + \sum_{i \in V} f_i^*(z_i)
\]

=: \( F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E}) \)

- Changes: \( \{s_\alpha\}_{\alpha \in V \cup E}, z_\alpha \in [\mathbb{R}^m]_{V \cup E} \), \( \alpha \in [V \cup E] \cup F \)

Both push and sum decrease \( F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E}) \).
To extend [Bof-Carli-Schenato ’17]’s distributed optimization on directed unreliable graphs on $\min_{x \in \mathbb{R}^m} \sum_{i \in V} [f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2],$

- Let $x_\alpha = \frac{y_\alpha}{s_\alpha},$ and minimize the dual objective function

$$\min_{z_\alpha \in [\mathbb{R}^m]_{V \cup E}} \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| \bar{x} - \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right\|_2^2 + \sum_{\beta \in F} \delta_{H_\beta}(z_\beta) + \sum_{i \in V} f_i^*(z_i)$$

The new term

$$=: F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E})$$

- Changes: $\{s_\alpha\}_{\alpha \in V \cup E}, z_\alpha \in [\mathbb{R}^m]_{V \cup V \cup E}, \alpha \in [V \cup E] \cup F$

New operation to adjust one $z_i$ (by solving similar prox pblm):

$$[z_i^{new}]_i := \arg \min_z \frac{s_i}{2} \left\| [x]_i + \frac{1}{s_i} ([z_i^{old}]_i - z) \right\|^2 + f_i^*(z)$$

Adjust $[x]_i$ and $y_i$ accordingly.
\[
\min_{z_{\alpha} \in \mathbb{R}^m | V \cup E} \sum_{\alpha \in V \cup E} \frac{s_{\alpha}}{2} \left[ x_{\alpha} - \frac{1}{s_{\alpha}} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right]_{\alpha}^2 + \sum_{\beta \in F} \delta_{H_{\beta}}(z_{\beta}) + \sum_{i \in V} f_i^*(z_i)
\]

\[=: F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E})\]

Elaboration of how push and sum decrease \( F(\{z_\alpha\}; \{s_\alpha\}) \):

Consider this suboperation.

(\(y_2, s_2\))  \(\rightarrow\)  (\(y_{2,6}, s_{2,6}\))  \(\rightarrow\)  (\(\alpha(y_2, s_2)\))  \(\rightarrow\)  (\(y_{2,6}, s_{2,6}\))

Push can be written as combination of above suboperations.

Sum is the similar suboperation below

Node 2  \(\rightarrow\)  Node 6

Nodes 1 & 6  \(\rightarrow\)  (\(y, s\))
\[
\min_{z_\alpha \in [\mathbb{R}^m]|V \cup E|} \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right\|_2^2 + \sum_{\beta \in F} \delta^*_\beta(z_\beta) + \sum_{i \in V} f^*_i(z_i)
\]

\[= : F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E})\]

**Elaboration of how push and sum decrease** \(F(\{z_\alpha\}; \{s_\alpha\})\):

Consider this suboperation.

\((y_2, s_2), (y_{2,6}, s_{2,6})\) → \(\alpha (y_2, s_2)\) + \((1 - \alpha) (y_{2,6}, s_{2,6})\)

Push can be written as combination of above suboperations.

Nodes 1 & 6 sum

Nodes 2 pushes
Elaboration of how push and sum decrease $F(\{z_\alpha\}; \{s_\alpha\})$:

Consider this suboperation.

Push can be written as combination of above suboperations.
Sum is the similar suboperation below
Elaboration of how push and sum decrease $F(\{z_\alpha\}; \{s_\alpha\})$:

Consider this suboperation.

Decrease in quadratic portion after a suboperation is

$$
\begin{align*}
&= \left[ \frac{s_2^{\text{old}}}{2} \left\| y_2^{\text{old}} \right\|^2 + \frac{s_{2,6}^{\text{old}}}{2} \left\| y_{2,6}^{\text{old}} \right\|^2 \right] - \left[ \frac{s_2^{\text{new}}}{2} \left\| y_2^{\text{new}} \right\|^2 + \frac{s_{2,6}^{\text{new}}}{2} \left\| y_{2,6}^{\text{new}} \right\|^2 \right] \\
&= \left[ \cdots \right] - \left[ \frac{\alpha s_2^{\text{old}}}{2} \left\| \frac{\alpha y_2^{\text{old}}}{s_2^{\text{old}}} \right\|^2 + \frac{(1 - \alpha)s_2^{\text{old}} + s_{2,6}^{\text{old}}}{2} \left\| (1 - \alpha)y_2^{\text{old}} + y_{2,6}^{\text{old}} \right\|^2 \right] \\
&= \left(1 - \alpha\right)s_2^{\text{old}} \left\| \frac{y_2^{\text{old}}}{s_2^{\text{old}}} \right\|^2 + \frac{s_{2,6}^{\text{old}}}{2} \left\| y_{2,6}^{\text{old}} \right\|^2 - \left(1 - \alpha\right)s_2^{\text{old}} + s_{2,6}^{\text{old}} \left\| \frac{(1 - \alpha)y_2^{\text{old}} + y_{2,6}^{\text{old}}}{1 - \alpha}s_2^{\text{old}} + s_{2,6}^{\text{old}} \right\|^2 \geq 0
\end{align*}
$$
\[
\min_{\mathbf{z}_\alpha \in [\mathbb{R}^m]|V \cup E|} \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| \mathbf{x} - \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} \mathbf{z}_{\alpha_2} \right\|^2 + \sum_{\beta \in F} \delta_{H_\beta}(\mathbf{z}_\beta) + \sum_{i \in V} f_i^*(\mathbf{z}_i)
\]

\[
=: F(\{\mathbf{z}_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E})
\]

**Elaboration of how push and sum decrease** \(F(\{\mathbf{z}_\alpha\}; \{s_\alpha\})\):

Consider this suboperation.

\[
(\mathbf{y}_2, s_2) \xrightarrow{2} (\mathbf{y}_{2,6}, s_{2,6}) \to \alpha(\mathbf{y}_2, s_2) \xrightarrow{2} (\mathbf{y}_{2,6}, s_{2,6}) \to \mathbf{y}_S
\]

Decrease in **quadratic portion** after a suboperation is

\[
\begin{align*}
&\left[\begin{array}{c}
\frac{s^\text{old}_2}{2} \left\| \mathbf{y}^\text{old}_2 \right\|^2 \\
\frac{s^\text{old}_{2,6}}{2} \left\| \mathbf{y}^\text{old}_{2,6} \right\|^2
\end{array}\right] - \left[\begin{array}{c}
\frac{s^\text{new}_2}{2} \left\| \mathbf{y}^\text{new}_2 \right\|^2 \\
\frac{s^\text{new}_{2,6}}{2} \left\| \mathbf{y}^\text{new}_{2,6} \right\|^2
\end{array}\right] \\
= &\left[\begin{array}{c}
\alpha s^\text{old}_2 \left\| \frac{\mathbf{y}^\text{old}_2}{\alpha s^\text{old}_2} \right\|^2 \\
\frac{(1 - \alpha)s^\text{old}_2 + s^\text{old}_{2,6}}{2} \left\| \frac{(1 - \alpha)y^\text{old}_2 + y^\text{old}_{2,6}}{(1 - \alpha)s^\text{old}_2 + s^\text{old}_{2,6}} \right\|^2
\end{array}\right]
\end{align*}
\]

\[
= \frac{(1 - \alpha)s^\text{old}_2}{2} \left\| \mathbf{y}^\text{old}_2 \right\|^2 + \frac{s^\text{old}_{2,6}}{2} \left\| \mathbf{y}^\text{old}_{2,6} \right\|^2 - \frac{(1 - \alpha)s^\text{old}_2 + s^\text{old}_{2,6}}{2} \left\| \frac{(1 - \alpha)y^\text{old}_2 + y^\text{old}_{2,6}}{(1 - \alpha)s^\text{old}_2 + s^\text{old}_{2,6}} \right\|^2 \geq 0
\]

- Can generalize for mix of directed/undirected edges.

\[\alpha \in [0, 1]\]
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
\[
\min_{\{z_\alpha\}_{\alpha \in V \cup E}} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \bar{x} - \left[ \sum_{\alpha \in V \cup E} z_\alpha \right] \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z(i,j))
\]

**Handling smooth/subdifferentiable functions.**

As it stands, subproblem when minimizing over one \(z_i\) is

\[
\min_{z_i \in [\mathbb{R}^m]_{V_i}} \frac{1}{2} \left\| \bar{x} - z_i - \left[ \sum_{\alpha \neq i} z^{old}_\alpha \right] \right\|^2 + f_i^*(z_i) \quad \text{(Prox-sub)}
\]
\[
F(\{z_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \bar{x} - \left[ \sum_{\alpha \in V \cup E} z_\alpha \right] \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z(i,j))
\]

**Handling smooth/ subdifferentiable functions.**

As it stands, subproblem when minimizing over one \(z_i\) is

\[
\min_{z_i \in [\mathbb{R}^m]} \frac{1}{2} \left\| \bar{x} - z_i - \left[ \sum_{\alpha \neq i} z^\text{old}_\alpha \right] \right\|^2 + f_i^*(z_i) \quad \text{(Prox-sub)}
\]

So all \(f_i(\cdot)/ f_i^*(\cdot)\) assumed *proximable* (i.e., (Prox-sub) are easy).

- Even if \(f_i(\cdot)\) is smooth, (Prox-sub) can still be hard.
- Many distributed algorithms need only gradient \(\nabla f_i(\cdot)\) info.
$$F(\{z_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \bar{x} - \left[ \sum_{\alpha \in V \cup E} z_\alpha \right] \right\|^2 + \sum_{i \in V} f^*_i(z_i) + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z(i,j))$$

(D)

**Handling smooth/ subdifferentiable functions.**

As it stands, subproblem when minimizing over one $z_i$ is

$$\min_{z_i \in \mathbb{R}^m | V} \frac{1}{2} \left\| \bar{x} - z_i - \left[ \sum_{\alpha \neq i} z^{old}_\alpha \right] \right\|^2 + f^*_i(z_i) \quad \text{(Prox-sub)}$$

So all $f_i(\cdot)$/ $f^*_i(\cdot)$ assumed *proximable* (i.e., (Prox-sub) are easy).

- Even if $f_i(\cdot)$ is smooth, (Prox-sub) can still be hard.
- Many distributed algorithms need only gradient $\nabla f_i(\cdot)$ info.

**Majorization:**

1. **Current iterate at** $x_1$
2. Find easily solvable majorization.
3. Solve it to get $x_2$. 

**Diagram:**

- **X1**
- **Current iterate at X1**
- **Find easily solvable majorization**
- **Solve it to get x2**
\[
\min_{\left\{ z_{\alpha} \right\}_{\alpha \in V \cup E}} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \bar{x} - \left[ \sum_{\alpha \in V \cup E} z_{\alpha} \right] \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z_{i,j})
\]

Handling smooth/subdifferentiable functions.
As it stands, subproblem when minimizing over one \( z_i \) is

\[
\min_{z_i \in \mathbb{R}^m \setminus V} \frac{1}{2} \left\| \bar{x} - z_i - \left[ \sum_{\alpha \neq i} z_{\alpha}^{old} \right] \right\|^2 + f_i^*(z_i) \quad \text{(Prox-sub)}
\]

So all \( f_i(\cdot) / f_i^*(\cdot) \) assumed proximable (i.e., (Prox-sub) are easy).

- Even if \( f_i(\cdot) \) is smooth, (Prox-sub) can still be hard.
- Many distributed algorithms need only gradient \( \nabla f_i(\cdot) \) info.

Majorization:

1. Current iterate at \( x_1 \)
2. **Find easily solvable majorization.**
3. Solve it to get \( x_2 \).
Handling smooth/subdifferentiable functions.

As it stands, subproblem when minimizing over one $z_i$ is

$$\min_{z_i \in \mathbb{R}^m \mid \forall \alpha \neq i} \frac{1}{2} \left\| \bar{x} - z_i - \left[ \sum_{\alpha \neq i} z_\alpha \right] \right\|^2 + f_i^*(z_i)$$

(Prox-sub)

So all $f_i(\cdot)$/$f_i^*(\cdot)$ assumed proximable (i.e., (Prox-sub) are easy).

- Even if $f_i(\cdot)$ is smooth, (Prox-sub) can still be hard.
- Many distributed algorithms need only gradient $\nabla f_i(\cdot)$ info.

Majorization:

1. Current iterate at $x_1$
2. Find easily solvable majorization.
3. Solve it to get $x_2$. 
\[
\min_{\mathbf{z}_\alpha \in \mathbb{R}^m \mid \alpha \in V \cup E} F(\{\mathbf{z}_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \mathbf{x} - \left[ \sum_{\alpha \in V \cup E} \mathbf{z}_\alpha \right] \right\|^2 + \sum_{i \in V} f^*_i(\mathbf{z}_i) + \sum_{(i,j) \in E} \delta^*_H(\mathbf{z}(i,j))
\]

**Handling smooth/ subdifferentiable functions.**

As it stands, subproblem when minimizing over one \(\mathbf{z}_i\) is

\[
\min_{\mathbf{z}_i \in \mathbb{R}^m \mid \mathbf{V}} \frac{1}{2} \left\| \mathbf{x} - \mathbf{z}_i - \left[ \sum_{\alpha \neq i} \mathbf{z}_{\alpha}^{old} \right] \right\|^2 + f^*_i(\mathbf{z}_i) \quad \text{(Prox-sub)}
\]

So all \(f_i(\cdot)/ f^*_i(\cdot)\) assumed *proximable* (i.e., (Prox-sub) are easy).

- Even if \(f_i(\cdot)\) is smooth, (Prox-sub) can still be hard.
- Many distributed algorithms need only gradient \(\nabla f_i(\cdot)\) info.

Can majorize (D) for subdifferentiable \(f_i(\cdot)\)

- Suppose \(\tilde{f}_i(\cdot) \leq f_i(\cdot)\).

\((\text{Simple } \tilde{f}_i(\cdot) \text{ can be obtained with function/ subgradient evaluations})\)
\[
\min_{\{z_\alpha\}_{\alpha \in \mathcal{V} \cup \mathcal{E}}} F := \frac{1}{2} \left\| \bar{\mathbf{x}} - \left[ \sum_{\alpha \in \mathcal{V} \cup \mathcal{E}} z_\alpha \right] \right\|^2 + \sum_{i \in \mathcal{V}} f_i^*(z_i) + \sum_{(i,j) \in \mathcal{E}} \delta^*_{H(i,j)}(z(i,j))
\]

(D)

Handling smooth/subdifferentiable functions.
As it stands, subproblem when minimizing over one \(z_i\) is

\[
\min_{z_i \in [\mathbb{R}^m] \setminus \mathcal{V}} \frac{1}{2} \left\| \bar{\mathbf{x}} - z_i - \left[ \sum_{\alpha \neq i} z^{old}_\alpha \right] \right\|^2 + f_i^*(z_i)
\]

(Prox-sub)

So all \(f_i(\cdot)/ f_i^*(\cdot)\) assumed proximable (i.e., (Prox-sub) are easy).

▶ Even if \(f_i(\cdot)\) is smooth, (Prox-sub) can still be hard.

▶ Many distributed algorithms need only gradient \(\nabla f_i(\cdot)\) info.

Can majorize (D) for subdifferentiable \(f_i(\cdot)\)

▶ Suppose \(\tilde{f}_i(\cdot) \leq f_i(\cdot)\).

Then \(\tilde{f}_i^*(\cdot) \geq f_i^*(\cdot)\).

▶ Substituting \(\tilde{f}_i^*(\cdot)\) into (D) gives a majorization of (D).

▶ Subproblem with \(\tilde{f}_i(\cdot)\) instead easier to solve.

▶ Solving sequence of majorizations finds min value of (D).
\[
\min_{z_\alpha \in [R^m] \forall \alpha \in V \cup E} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \tilde{x} - \left[ \sum_{\alpha \in V \cup E} z_\alpha \right] \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z_{i,j})
\]

(D)

**Second majorization**

\[ f_i(x) = \delta_{\{x:g_i(x) \leq 0\}}(x), \text{ where } g_i(\cdot) \text{ convex.} \]

- It might be difficult to solve \( \min_x \frac{1}{2} \| \tilde{x} - x \|^2 + \frac{1}{2} f_i(x) \).
- Consider \( \tilde{f}_i(x) = \delta_{\{x:g_i(\tilde{x}) + \langle s_i, x - \tilde{x} \rangle \leq 0\}}(x), \) where \( s_i \in \partial f_i(\tilde{x}) \):
  - \( \tilde{f}_i(\cdot) \leq f_i(\cdot), \) \( \tilde{f}_i(\cdot) \) easy to calculate & gives easy subproblem.
  - Subproblem is easy as \( \tilde{f}_i(\cdot) \) is indicator of halfspace.
\[
\min_{\{z_{\alpha}\}_{\alpha \in V \cup E}} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \bar{x} - \left[ \sum_{\alpha \in V \cup E} z_{\alpha} \right] \right\|^2 + \sum_{i \in V} f^*_i(z_i) + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z(i,j))
\]

**Second majorization**
\[
f_i(x) = \delta_{\{x: g_i(x) \leq 0\}}(x), \text{ where } g_i(\cdot) \text{ convex.}
\]

▶ It might be difficult to solve \(\min_x \frac{1}{2} \| \bar{x} - x \|^2 + \frac{1}{2} f_i(x)\).

▶ Consider \(\tilde{f}_i(x) = \delta_{\{x: g_i(\bar{x}) + \langle s_i, x - \bar{x} \rangle \leq 0\}}(x), \text{ where } s_i \in \partial f_i(\bar{x})\):
  
  ▶ \(\tilde{f}_i(\cdot) \leq f_i(\cdot), \tilde{f}_i(\cdot) \text{ easy to calculate & gives easy subproblem.}\)
  
  ▶ Subproblem is easy as \(\tilde{f}_i(\cdot) \text{ is indicator of halfspace.}\)

▶ This majorization was studied in [Combettes '00] & [Bregman- Censor- Reich- Zepkowitz-Malachi '03].
Another majorization (which I still have to write down) is motivated from Combettes & collaborators, 2009 or earlier.

For

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

(P2)

with linear \( L_i : \mathbb{R}^m \to \mathbb{R}^{m_i} \) & convex \( f_i : \mathbb{R}^{m_i} \to \mathbb{R} \).
Third majorization

- Another majorization (which I still have to write down) is motivated from Combettes & collaborators, 2009 or earlier.

- For
  \[
  \min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ f_i \circ L_i(x) + \frac{1}{2} \| x - \bar{x}_i \|^2 \right]
  \]  
  \[
  (P2)
  \]
  with linear $L_i : \mathbb{R}^m \to \mathbb{R}^m_i$ & convex $f_i : \mathbb{R}^m_i \to \mathbb{R}$,

- Details: Dual is

  \[
  \min_{z_\alpha \in [\mathbb{R}^m]_{\forall \alpha}} \frac{1}{2} \left\| \bar{x} - \sum_{i \in V} L_i^* z_i - \sum_{\alpha \in \bar{E}} z_\alpha \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z_{(i,j)})
  \]

- To minimize one variable $z_i$ with $L_i^T L_i \preceq \frac{1}{\rho}$, subproblem is

  \[
  \min_{z \in \mathbb{R}^m} \frac{1}{2} \left\| x_i + L_i[z_i] - L_i z \right\|^2 + f_i^*(z)
  \]

- Majorize red term:

  \[
  \frac{1}{2} \| \tilde{x}_i - L_i z \|^2 = \frac{1}{2} z^T L_i^T L_i z + \tilde{x}_i^T L_i z + \frac{1}{2} \| \tilde{x}_i \|^2
  \]

  \[
  L_i^T L_i \preceq \frac{1}{\rho} \leq \frac{1}{2\rho} \| z \|^2 + \tilde{x}_i^T L_i z + \frac{1}{2} \| \tilde{x}_i \|^2.
  \]
Fourth majorization:

Supporting Halfspace- Quadratic programming (SHQP)

We illustrate the SHQP heuristic [Pang, ’16] first.

Projection onto $C_1$ gives a halfspace $H_1$ s.t. $H_1 \supset C_1$.
Similarly, $H_2 \supset C_2$.
Note $H_1 \cap H_2 \supset C_1 \cap C_2$.
QP Step: Projecting $x_0$ onto $H_1 \cap H_2$ gives new dual variables.
Fourth majorization: Supporting Halfspace-Quadratic programming (SHQP)

We illustrate the SHQP heuristic [Pang, '16] first.

Projection onto $C_1$ gives a halfspace $H_1$ s.t. $H_1 \supset C_1$.

Similarly, $H_2 \supset C_2$.

Note $H_1 \cap H_2 \supset C_1 \cap C_2$.

QP Step: Projecting $x_0$ onto $H_1 \cap H_2$ gives new dual variables.
Fourth majorization:

**Supporting Halfspace- Quadratic programming (SHQP)**

We illustrate the SHQP heuristic [Pang, ’16] first.

Projection onto $C_1$ gives a halfspace $H_1$ s.t. $H_1 \supset C_1$.

Similarly, $H_2 \supset C_2$.

**Note** $H_1 \cap H_2 \supset C_1 \cap C_2$.

QP Step: Projecting $x_0$ onto $H_1 \cap H_2$ gives new dual variables.
Fourth majorization: 

**Supporting Halfspace- Quadratic programming (SHQP)**

We illustrate the SHQP heuristic [Pang, '16] first.

Projection onto $C_1$ gives a halfspace $H_1$ s.t. $H_1 \supset C_1$.

Similarly, $H_2 \supset C_2$.

Note $H_1 \cap H_2 \supset C_1 \cap C_2$.

**QP Step:** Projecting $x_0$ onto $H_1 \cap H_2$ gives new dual variables.
Consider
\[ \min_x \frac{1}{2} \| x - x_0 \|^2 + \sum_{i=1}^{r} \delta_{C_i}(x). \]

with dual
\[
(D) \quad \min_{z_1, z_2, \ldots, z_r} \frac{1}{2} \left\| x_0 - \sum_{i=1}^{r} z_i \right\|^2 + \sum_{i=1}^{r} \delta^*_C(z_i),
\]

Dykstra’s algorithm is block coordinate minimization on (D).
Consider
\[ \min_x \frac{1}{2} \| x - x_0 \|^2 + \sum_{i=1}^r \delta_{C_i}(x). \]
with dual
\[
(D) \quad \min_{z_1, z_2, \ldots, z_r} \frac{1}{2} \left\| x_0 - \sum_{i=1}^r z_i \right\|^2 + \sum_{i=1}^r \delta^*_C(z_i),
\]
Dykstra’s algorithm is block coordinate minimization on (D).

From a starting \( z^{old} = (z_1^{old}, \ldots, z_r^{old}) \), the SHQP heuristic gives

\[
\frac{1}{2} \left\| x_0 - \sum_{i=1}^r z_i^{old} \right\|^2 + \sum_{i=1}^r \delta^*_C(z_i^{old})
\]
\[
= \frac{1}{2} \left\| x_0 - \sum_{i=1}^r z_i^{old} \right\|^2 + \sum_{i=1}^r \delta^*_{H_i}(z_i^{old}) \quad \text{[Halfspaces } H_i \text{ generated by projection]}
\]

QP step
\[
\geq \frac{1}{2} \left\| x_0 - \sum_{i=1}^r z_i^{new} \right\|^2 + \sum_{i=1}^r \delta^*_{H_i}(z_i^{new}) \quad \text{[QP subproblem easy, many } z_i \text{'s at once]}
\]
\[
\geq \frac{1}{2} \left\| x_0 - \sum_{i=1}^r z_i^{new} \right\|^2 + \sum_{i=1}^r \delta^*_C(z_i^{new}). \quad \text{[(Majorization) } H_i \supset C_i, \text{ so } \delta^*_{H_i}(\cdot) \geq \delta^*_C(\cdot)]
\]

The SHQP heuristic applies to the distributed Dykstra’s algorithm too.
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges
- Various majorizations for when $f_i(\cdot)$ are not proximable
- **Convergence rates**
- Simple numerical experiments
Recall dual objective function

\[
\min_{z_\alpha \in \mathbb{R}^m, \alpha \in V \cup E} \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| \begin{bmatrix} x \end{bmatrix}_\alpha - \frac{1}{s_\alpha} \sum_{\alpha_2 \in V \cup F} z_{\alpha_2} \right\|^2 + \sum_{\beta \in F} \delta^*_H(z_\beta) + \sum_{i \in V} f^*_i(z_i)
\]

\[=: F(\{z_\alpha\}_{\alpha \in V \cup F}; \{s_\alpha\}_{\alpha \in V \cup E})\]

Strong duality and Fenchel duality gives

\[F_k = F(\{z^k_\alpha\}_{\alpha \in V \cup F}; \{s^k_\alpha\}_{\alpha \in V \cup E}) - F^* \geq \sum_{\alpha \in V \cup E} \frac{s_\alpha}{2} \left\| x^k_\alpha - x^* \right\|^2 \]

dual gap (nonincreasing, limit zero)

What we actually want to measure
Recall dual objective function

\[
\min_{\mathbf{z}_\alpha \in \mathbb{R}^m|\mathbf{VUE}|, \alpha \in \mathbf{VUF}} \left\{ \frac{1}{2} \sum_{\alpha \in \mathbf{VUE}} \frac{s_\alpha}{S_\alpha} \left\| \mathbf{x} - \frac{1}{s_\alpha} \sum_{\alpha_2 \in \mathbf{VUF}} \mathbf{z}_{\alpha_2} \right\|^2 + \sum_{\beta \in \mathbf{F}} \delta_{H_\beta}^*(\mathbf{z}_\beta) + \sum_{i \in \mathbf{V}} f_i^*(\mathbf{z}_i) \right\}
\]

\[=: F(\{\mathbf{z}_\alpha\}_{\alpha \in \mathbf{VUF}}; \{s_\alpha\}_{\alpha \in \mathbf{VUE}})\]

Strong duality and Fenchel duality gives

\[F_k = F(\{\mathbf{z}_\alpha^k\}_{\alpha \in \mathbf{VUF}}; \{s_\alpha^k\}_{\alpha \in \mathbf{VUE}}) - F^* \geq \sum_{\alpha \in \mathbf{VUE}} \frac{s_\alpha}{2} \left\| \mathbf{x}_\alpha^k - \mathbf{x}^* \right\|^2 \]

dual gap (nonincreasing, limit zero)

What we actually want to measure

\[\text{Undirected case: } F(\{\mathbf{z}_\alpha\}_{\alpha \in \mathbf{VUE}}) - F^* \geq \sum_{\alpha \in \mathbf{V}} \frac{1}{2} \left\| x_i - \mathbf{x}^* \right\|^2.\]

Shall study convergence rates of the quantity on the left.

(Note: Other than linear convergence in the smooth case, the convergence rates are proved only for the undirected case for now.)
\[
\min_{\{z_\alpha\}_{\alpha \in V \cup E}} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{i \in V} f^*_i(z_i) + \sum_{(i,j) \in E} \delta^*_{H(i,j)}(z(i,j))
\]

**[Rates]** When constraint qualifications guarantee a dual optimizer:

1. Majorizations for one block \( \min_z f^*_i(z) + \frac{1}{2} \|z - x_i - [z_i]_i\|^2:\)
   
   (a) \(O(1/k)\) rate for subgradient approximation on \(f_i(\cdot)\).
   
   (b) \(O(1/k)\) rate for \(f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}}\) & \(g_i(\cdot)\) subdifferentiable.
   
   (c) Linear rate for \(f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot)\), where \(h_j(\cdot)\) all smooth & we take subgradient approximation of each \(h_j(\cdot)\)

   ▶ Special case: \(k = 1\)
\[ F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| [\bar{x} - \sum_{\alpha \in V \cup E} z_\alpha]_i \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z(i,j)) \]

[Rates] When constraint qualifications guarantee a dual optimizer:

1. Majorizations for one block \( \min_z f_i^*(z) + \frac{1}{2} \| z - x_i - [z_i]_i \|^2 \):
   \( \begin{align*}
   (a) & \quad O(1/k) \text{ rate for subgradient approximation on } f_i(\cdot). \\
   (b) & \quad O(1/k) \text{ rate for } f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}} \text{ & } g_i(\cdot) \text{ subdifferentiable.} \\
   (c) & \quad \text{Linear rate for } f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot), \text{ where } h_j(\cdot) \text{ all smooth & we take subgradient approximation of each } h_j(\cdot) \\
   & \quad \quad \text{▷ Special case: } k = 1
   \end{align*} \)

2. Multiple block case (distributed Dykstra’s algorithm):
   \( \begin{align*}
   (a) & \quad \text{No majorization (i.e., using only proximal subproblems) is just block coordinate minimization, which has } O(1/k) \text{ rate.}
   \end{align*} \)
\[
F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H_{(i,j)}}(z(i,j))
\]

[Rates] When constraint qualifications guarantee a dual optimizer:

1. Majorizations for one block \( \min_z f_i^*(z) + \frac{1}{2} \|z - x_i - [z_i]_i\|^2 \):
   (a) \( O(1/k) \) rate for subgradient approximation on \( f_i(\cdot) \).
   (b) \( O(1/k) \) rate for \( f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}} \) & \( g_i(\cdot) \) subdifferentiable.
   (c) Linear rate for \( f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot) \), where \( h_j(\cdot) \) all smooth & we take subgradient approximation of each \( h_j(\cdot) \)
      ▶ Special case: \( k = 1 \)

2. Multiple block case (distributed Dykstra’s algorithm):
   (a) No majorization (i.e., using only proximal subproblems) is just block coordinate minimization, which has \( O(1/k) \) rate.
      ▶ \( O(1/k) \) rate preserved if we use majorization 1(c).
      ▶ \( O(1/k^{1/3}) \) rate if we use majorization 1(a) and 1(b).
\[
\min_{z_{\alpha} \in [R^m]|V|, \alpha \in V \cup E} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\|\bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha}\right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z(i,j))
\]

[Rates] When constraint qualifications guarantee a dual optimizer:

1. Majorizations for one block \(\min_z f_i^*(z) + \frac{1}{2}\|z - x_i - [z_i]_i\|^2\):
   (a) \(O(1/k)\) rate for subgradient approximation on \(f_i(\cdot)\).
   (b) \(O(1/k)\) rate for \(f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}}\) & \(g_i(\cdot)\) subdifferentiable.
   (c) Linear rate for \(f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot)\), where \(h_j(\cdot)\) all smooth & we take subgradient approximation of each \(h_j(\cdot)\)
      ▶ Special case: \(k = 1\)

2. Multiple block case (distributed Dykstra’s algorithm):
   (a) (General case)
   (b) Linear convergence if the two block problem

\[
\min_{z_{\alpha} \in [R^m]|V|, \alpha \in V \cup E} \frac{1}{2} \left\|\bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha}\right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{e \in E} \delta_e^*(z_e) \tag{†}
\]

has linear convergence
\[
\min_{\{z_{\alpha}\}_{\alpha \in V \cup E}, \alpha \in V \cup E} F(\{z_{\alpha}\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha} \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H_{(i,j)}}^*(z_{(i,j)})
\]

**[Rates]** When constraint qualifications guarantee a dual optimizer:

1. **Majorizations for one block** \( \min_{z} f_i^*(z) + \frac{1}{2} \| z - x_i - [z_i]_i \|^2 \):
   - (a) \( O(1/k) \) rate for subgradient approximation on \( f_i(\cdot) \).
   - (b) \( O(1/k) \) rate for \( f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}} \) & \( g_i(\cdot) \) subdifferentiable.
   - (c) Linear rate for \( f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot) \), where \( h_j(\cdot) \) all smooth & we take subgradient approximation of each \( h_j(\cdot) \)
     - Special case: \( k = 1 \)

2. **Multiple block case (distributed Dykstra’s algorithm):**
   - (a) (General case)
   - (b) Linear convergence if the two block problem
   \[
   \min_{\{z_{\alpha}\}_{\alpha \in V \cup E}, \alpha \in V \cup E} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_{\alpha} \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{e \in E} \delta_{H_e}^*(z_e) \tag{\dagger}
   \]
     has linear convergence
   Note 1: Linear convergence of (\dagger) well studied using error bounds.
   Note 2: Condition holds if \( f_i(\cdot) \) all smooth (or \( f_i^*(\cdot) \) strongly convex).
\[
F(\{z_\alpha\}_{\alpha \in V \cup E}) := \sum_{i \in V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H_{i,j}}^*(z_{i,j})
\]

**[Rates]** When constraint qualifications guarantee a dual optimizer:

1. Majorizations for one block \(\min_z f_i^*(z) + \frac{1}{2} \left\| z - x_i - [z_i]_i \right\|^2:\)
   (a) \(O(1/k)\) rate for subgradient approximation on \(f_i(\cdot)\).
   (b) \(O(1/k)\) rate for \(f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}} \& g_i(\cdot)\) subdifferentiable.
   (c) Linear rate for \(f_i(\cdot) = \max_{1 \leq j \leq k} h_j(\cdot)\), where \(h_j(\cdot)\) all smooth & we take subgradient approximation of each \(h_j(\cdot)\)
      ▶ Special case: \(k = 1\)

2. Multiple block case (distributed Dykstra’s algorithm):
   (a) (General case)
   (b) Linear convergence if the two block problem

\[
\min_{z_\alpha \in [\mathbb{R}^m]_V} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{e \in E} \delta_{H_e}^*(z_e) \quad (\dagger)
\]

has linear convergence

Note 1: Linear convergence of (\dagger) well studied using error bounds.

Note 2: Condition holds if \(f_i(\cdot)\) all smooth (or \(f_i^*(\cdot)\) strongly convex).
   ▶ Linear rate still holds if we use majorization 1(c).
   ▶ \(O(1/k)\) rate if we use majorization 1(a).
Review

- Averaged consensus algorithm and distributed optimization.
- Averaged consensus for directed unreliable edges.
- Further technical material.

We now talk about new results.

- Basic algorithm for the undirected case
- Extension for directed unreliable edges
- Various majorizations for when $f_i(\cdot)$ are not proximable
- Convergence rates
- Simple numerical experiments
\[
\min_{\alpha \in \mathbb{V} \cup \mathbb{E}} F(\{z_{\alpha}\}_{\alpha \in \mathbb{V} \cup \mathbb{E}}) := \frac{1}{2} \left\| \sum_{\alpha \in \mathbb{V} \cup \mathbb{E}} z_{\alpha} - \bar{x} \right\|^2 + \sum_{i \in \mathbb{V}} f_i^*(z_i) + \sum_{(i,j) \in \mathbb{E}} \delta^*_{H(i,j)}(z(i,j))
\]

\[
F_k = F(\{z_{\alpha}^k\}_{\alpha \in \mathbb{V} \cup \mathbb{F}}; \{s_{\alpha}^k\}_{\alpha \in \mathbb{V} \cup \mathbb{E}}) - F^* \geq \sum_{\alpha \in \mathbb{V} \cup \mathbb{E}} \frac{s_{\alpha}^k}{2} \|x_{\alpha}^k - x^*\|^2
\]

dual gap (nonincreasing, limit zero)

What we actually want to measure

Numerical experiments:
- We measure dual gap and sum of square norms above.
- We use
  - **Smooth functions:** \( f_i(x) := \frac{1}{2} x^T [\nu \nu^T + rI] x + b_i^T x \), where \( \nu \in \mathbb{R}^m \) and \( r \in \mathbb{R}_+ \) are generated randomly.
  - **Nonsmooth functions:** Max of 2 quadratics.
  - **Indicator functions of lower level sets of** \( g_i(\cdot) \): \( f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}}(\cdot), \) where
    - \( g_i(\cdot) \) is a random convex quadratic, or
    - \( g_i(\cdot) \) is a max of 2 random quadratics.
\[
\min_{\alpha \in V \cup E} F(\{z_\alpha\}_{\alpha \in V \cup E}) := \frac{1}{2} \left\| \sum_{\alpha \in V \cup E} z_\alpha - \bar{x} \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{(i,j) \in E} \delta_{H(i,j)}^*(z(i,j))
\]

\[
F_k = F(\{z^{k}_\alpha\}_{\alpha \in V \cup F}; \{s^{k}_\alpha\}_{\alpha \in V \cup E}) - F^* \geq \sum_{\alpha \in V \cup E} \frac{s^{k}_\alpha}{2} \left\| x^{k}_\alpha - x^* \right\|^2
\]

dual gap (nonincreasing, limit zero)

What we actually want to measure

---

**Smooth case (usual)**

- Red dashed line: duality gap (subdifferentiable)
- Blue line: duality gap (proximable)
- Red circle: \(\frac{1}{2}\|x - x^*\|^2\) (subdifferentiable)
- Blue plus: \(\frac{1}{2}\|x - x^*\|^2\) (proximable)

**Nonsmooth subdifferentiable case**

- Blue line: duality gap
- Red plus: \(\frac{1}{2}\|x - x^*\|^2\)
Summary of numerical experiments 1:

- Smooth functions $f_i(\cdot)$, either treat all $f_i(\cdot)$ as proximable, or treat all $f_i(\cdot)$ as subdifferentiable.
- Linear convergence (of $\{F_k\}_{k=1}^{\infty}$) as predicted by theory.

- Treating $f_i(\cdot)$ as proximable gives greedy (and usually better long term) decrease in the dual value $F(\{z_\alpha\}_{\alpha \in V \cup E})$.
- But interestingly, slower 12.5% of the time.
Summary of numerical experiments 2:
- Nonsmooth functions $f_i(\cdot)$
  - If all $f_i(\cdot)$ proximable, then may get linear convergence.
    - deduced by linear convergence conditions
  - If all $f_i(\cdot)$ had subgradient approximations, then $O(1/k)$ rate
- If some $f_i(\cdot)$ are indicator sets of convex functions:
  - Linear convergence observed, but not yet proved.
  - If some $f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}}$ for some closed convex $g_i(\cdot)$:
    - $O(1/k)$ convergence observed, better than $O(1/k^{1/3})$ that we proved.
Summary of numerical experiments 2:

- Nonsmooth functions $f_i(\cdot)$
  - If all $f_i(\cdot)$ proximable, then may get linear convergence.
    - deduced by linear convergence conditions
  - If all $f_i(\cdot)$ had subgradient approximations, then $O(1/k)$ rate

- If some $f_i(\cdot)$ are indicator sets of convex functions:
  - Linear convergence observed, but not yet proved.
- If some $f_i(\cdot) = \delta_{\{x: g_i(x) \leq 0\}}(\cdot)$ for some closed convex $g_i(\cdot)$:
  - $O(1/k)$ convergence observed,
    - Better than $O(1/k^{1/3})$ that we proved.
Summary of features (for both undirected and directed case):

- Each node only needs to store \([x]_i, [z]_i \in \mathbb{R}^m\) (or \(s_i \in \mathbb{R}\)).
- Scalable, Distributed, Asynchronous, Decentralized.
- Deterministic convergence.
- Dual method can treat sum of many proximable functions.

Furthermore, for undirected case:

- Constraint qualification free convergence
- Time-varying graphs, Partial communication of data
- Proved convergence rates.

Drawback: Need \(\frac{1}{2}\|x - \bar{x}\|^2\) regularization term.
Special thanks to Ting Kei Pong and Mert Gurbuzbalaban for discussions leading to the distributed Dykstra’s algorithm.

Thanks to students Kevin Lam Kai Bin and Zhu Yiyao for programming parts of the algorithm.

References:


and other preprints listed on my website


The end. Thanks for your attention!