Optimization in Networkland:
Challenges and Opportunities of Distributed Methods

Giuseppe Notarstefano

Department of Electrical, Electronic, and Information Engineering
Alma Mater Studiorum Università di Bologna
giuseppe.notarstefano@unibo.it

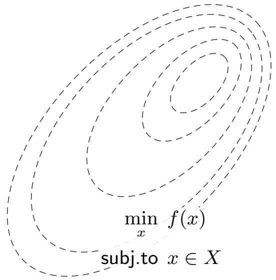
June 28, 2019, Naples (Italy)

RESEARCH FUNDED BY

Why Distributed Optimization?

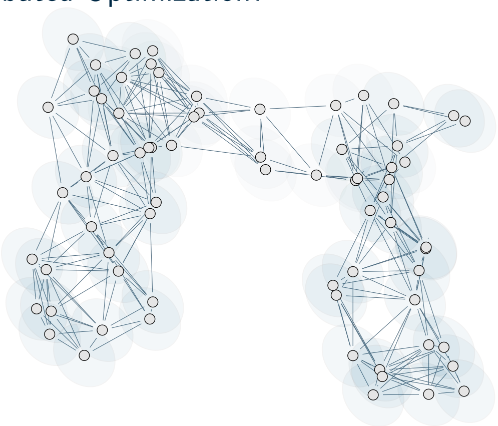
Mathematical problems appearing in several fields
(Engineering, Economics, Biology, Social Sciences, ...)

Well-established numerical schemes



$$\min_x \ f(x)$$
$$\text{subj.to } x \in X$$

Why Distributed Optimization?



Massive computation
and communication

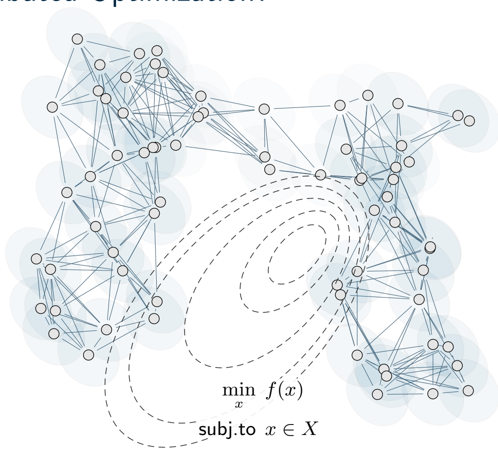
Local private data

Distributed algorithms
(e.g., average consensus)

⇓

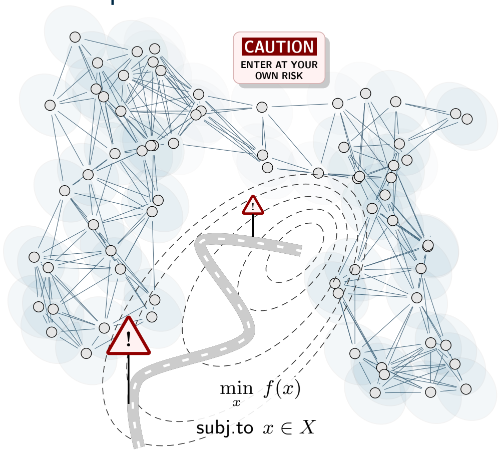
Share computation
instead of data

Why Distributed Optimization?



$$\min_x \ f(x)$$
$$\text{subj.to } x \in X$$

Why Distributed Optimization?

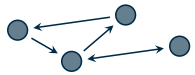


Network Challenges

Optimization Challenges

Network Challenges

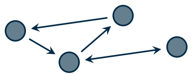
Distributed algorithm: compute locally & communicate with neighbors



Network Challenges

Distributed algorithm: compute locally & communicate with neighbors

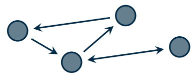
- Undirected/Directed



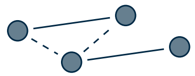
Network Challenges

Distributed algorithm: compute locally & communicate with neighbors

- Undirected/Directed



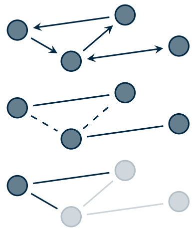
- Fixed/Time-Varying



Network Challenges

Distributed algorithm: compute locally & communicate with neighbors

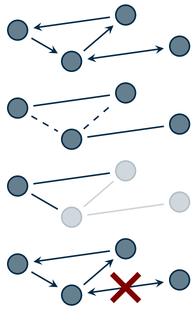
- Undirected/Directed
- Fixed/Time-Varying
- Asynchronous



Network Challenges

Distributed algorithm: compute locally & communicate with neighbors

- Undirected/Directed
- Fixed/Time-Varying
- Asynchronous
- Unreliable



Topology and communication NOT a design parameter

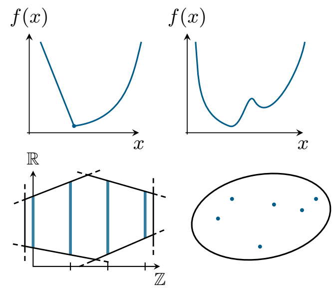
Optimization Challenges

$$\min_{x} \; f(x)$$
$$\text{subj.to } x \in X$$

Optimization Challenges

$$\min_{x} \; f(x)$$
$$\text{subj.to } x \in X$$

Nonsmooth (convex), Nonconvex, Mixed-Integer, Combinatorial

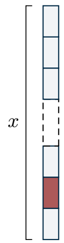


Optimization Challenges

$$\min_x \; f(x)$$
$$\text{subj.to } x \in X$$

Nonsmooth (convex), Nonconvex, Mixed-Integer, Combinatorial

Big-Data (high dim. dec. var.)



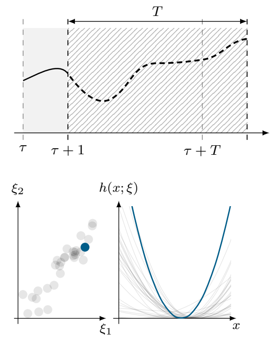
Optimization Challenges

$$\min_x \; f(x)$$
$$\text{subj.to } x \in X$$

Nonsmooth (convex), Nonconvex, Mixed-Integer, Combinatorial

Big-Data (high dim. dec. var.)

Dynamic, Online, Stochastic, Uncertain

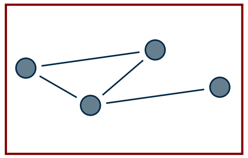


Distributed Optimization Paradigm

Optimization

$$\min_x \; f(x)$$
$$\text{subj.to } x \in X$$

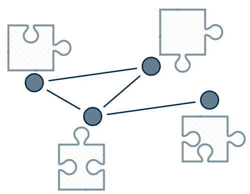
Network



Distributed Optimization Paradigm

Optimization

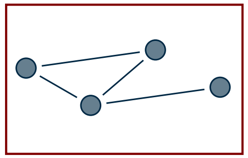
$$\min_x \; f(x)$$
$$\text{subj.to } x \in X$$



Agents

- know only part of optimization problem
- cooperate to compute a solution

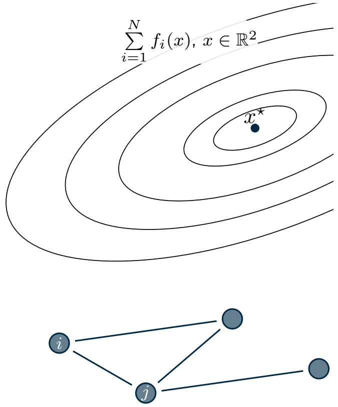
Network



Cost-Coupled Set-up

$$\min_{x \in X} \sum_{i=1}^{N} f_i(x)$$

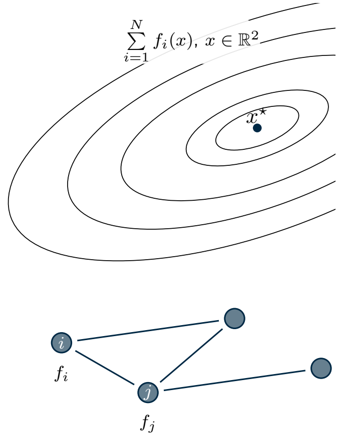
- N agents communicate over graph $\mathcal{G}$



Cost-Coupled Set-up

$$\min_{x \in X} \sum_{i=1}^{N} f_i(x)$$

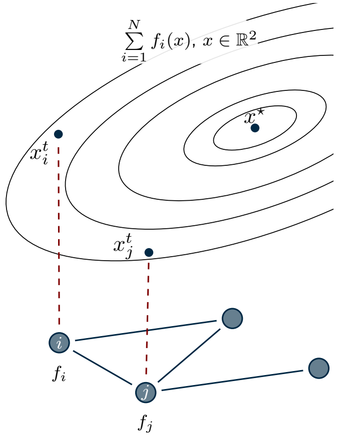
- N agents communicate over graph $\mathcal{G}$
- agent i knows f_i only



Cost-Coupled Set-up

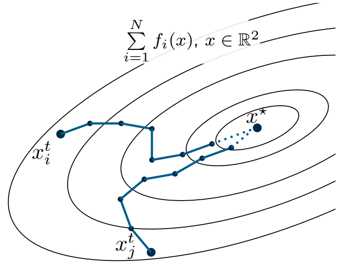
$$\min_{x \in X} \sum_{i=1}^{N} f_i(x)$$

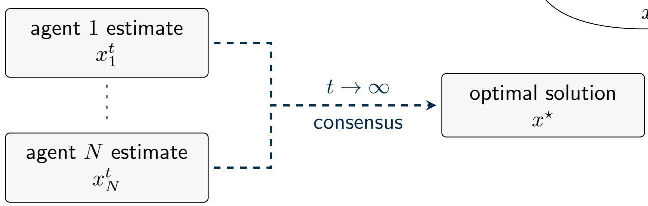
- N agents communicate over graph $\mathcal{G}$
- agent i knows f_i only
- x_i^t solution estimate of i



Cost-Coupled Set-up

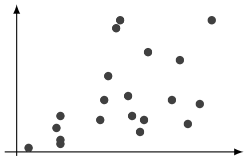
$$\min_{x \in X} \sum_{i=1}^{N} f_i(x)$$





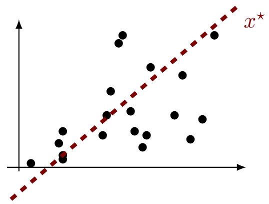
Data Analytics

$$\min_x \sum_{i=1}^{N} \underbrace{\|b_i - D_i x\|^2}_{f_i(x)} + r(x)$$



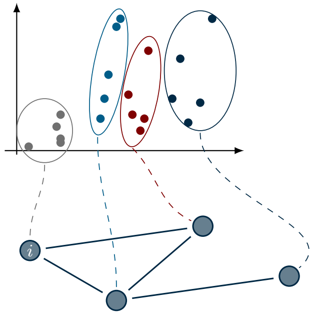
Data Analytics

$$\min_x \sum_{i=1}^{N} \underbrace{\|b_i - D_i x\|^2}_{f_i(x)} + r(x)$$



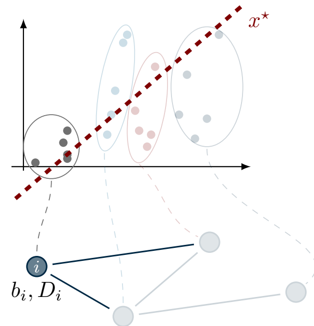
Data Analytics

$$\min_x \sum_{i=1}^{N} \underbrace{\|b_i - D_i x\|^2}_{f_i(x)} + r(x)$$



Data Analytics

$$\min_x \sum_{i=1}^{N} \underbrace{\|b_i - D_i x\|^2}_{f_i(x)} + r(x)$$

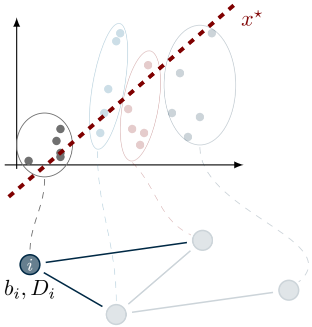


Paradigm

- local private data
- cooperate to learn from all data

Data Analytics

$$\min_x \sum_{i=1}^{N} \underbrace{\|b_i - D_i x\|^2}_{f_i(x)} + r(x)$$



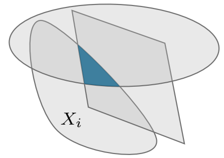
Paradigm

- local private data
- cooperate to learn from all data

Share computation instead of data

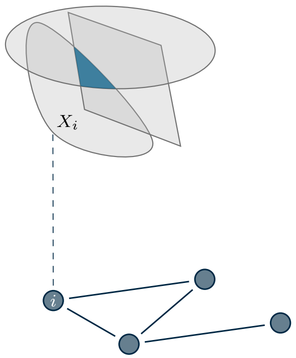
Common Cost Set-up

$$\min_x \; f(x)$$
$$\text{subj.to } x \in \bigcap_{i=1}^{N} X_i$$



Common Cost Set-up

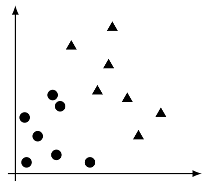
$$\min_x \; f(x)$$
$$\text{subj.to } x \in \bigcap_{i=1}^{N} X_i$$



- agent i knows common cost f and local constraint X_i
- algorithms suited for this set-up

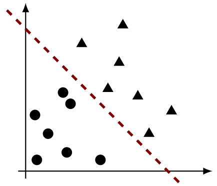
Support Vector Machine

$$\min_{w,b,\xi} \; \frac{1}{2}\|w\|^2 + C\xi$$
$$\text{subj.to } \ell_i(w^\top p_i + b) \geq 1 - \xi, \forall i$$
$$\xi \geq 0$$



Support Vector Machine

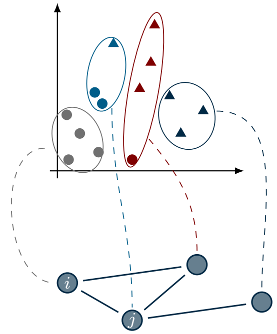
$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\xi$$
$$\text{subj.to} \ \ \ell_i(w^\top p_i + b) \geq 1 - \xi, \forall i$$
$$\xi \geq 0$$



Support Vector Machine

$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\xi$$
$$\text{subj.to} \ \ \ell_i(w^\top p_i + b) \geq 1 - \xi, \forall i$$
$$\xi \geq 0$$

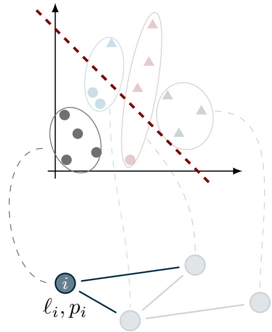
Local constraints and common cost



Support Vector Machine

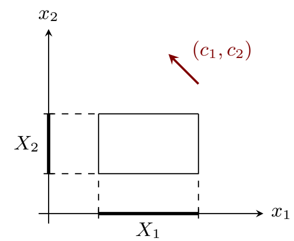
$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|^2 + C\xi$$
$$\text{subj.to} \ \ \ell_i(w^\top p_i + b) \geq 1 - \xi, \forall i$$
$$\xi \geq 0$$

Local constraints and common cost



Constraint-Coupled Set-up

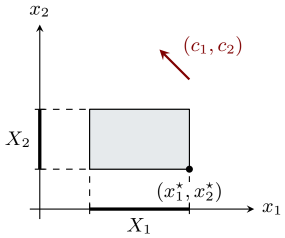
$$\min_{x_1,x_2} \ c_1x_1 + c_2x_2$$
$$\text{subj.to} \ \ x_1 \in X_1, \ x_2 \in X_2$$



Constraint-Coupled Set-up

$$\min_{x_1,x_2} \; c_1 x_1 + c_2 x_2$$

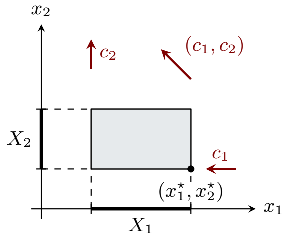
subj.to $x_1 \in X_1, \; x_2 \in X_2$



Constraint-Coupled Set-up

$$\min_{x_1,x_2} \; c_1 x_1 + c_2 x_2$$

subj.to $x_1 \in X_1, \; x_2 \in X_2$

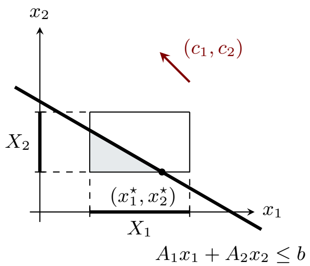


Constraint-Coupled Set-up

$$\min_{x_1,x_2} \; c_1 x_1 + c_2 x_2$$

subj.to $x_1 \in X_1, \; x_2 \in X_2$

$$A_1 x_1 + A_2 x_2 \leq b$$

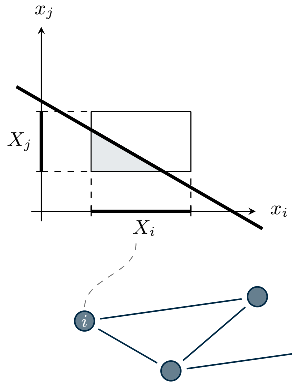


Constraint-Coupled Set-up

$$\min_{x_1,\ldots,x_N} \; \sum_{i=1}^{N} f_i(x_i)$$

subj.to $x_i \in X_i, \; \forall i$

$$\sum_{i=1}^{N} g_i(x_i) \leq 0$$



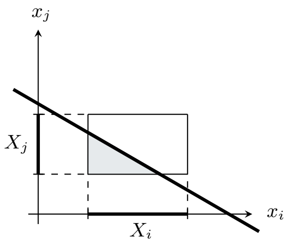
- N agents communicate over graph $\mathcal{G}$
- $(x_1, \ldots, x_N)$ dec. var. – size grows with N
- agent i knows f_i, g_i and X_i

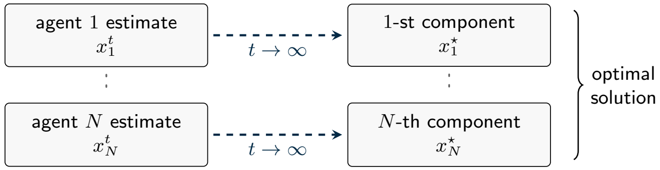
Constraint-Coupled Set-up

$$\min_{x_1,\ldots,x_N} \sum_{i=1}^{N} f_i(x_i)$$

$$\text{subj.to } x_i \in X_i,\ \forall i$$

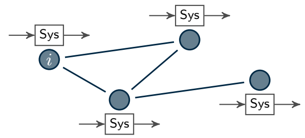
$$\sum_{i=1}^{N} g_i(x_i) \leq 0$$

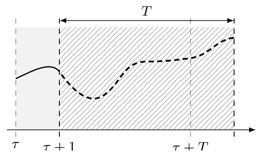




Distributed Optimal Control

$$\min_{\substack{z_1,\ldots,z_N\\u_1,\ldots,u_N}} \sum_{i=1}^{N} \Big(\sum_{\tau=0}^{T} \ell_i(z_i(\tau), u_i(\tau)) + m_i(z_i(T)) \Big)$$



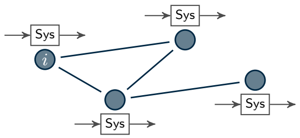


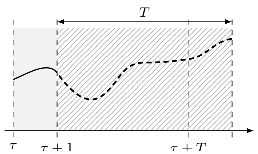
Distributed Optimal Control

$$\min_{\substack{z_1,\ldots,z_N\\u_1,\ldots,u_N}} \sum_{i=1}^{N} \Big(\sum_{\tau=0}^{T} \ell_i(z_i(\tau), u_i(\tau)) + m_i(z_i(T)) \Big)$$

$$\text{subj.to } z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$$

$$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$$





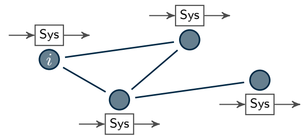
Distributed Optimal Control

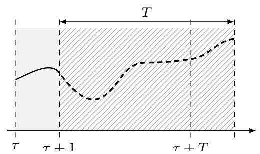
$$\min_{\substack{z_1,\ldots,z_N\\u_1,\ldots,u_N}} \sum_{i=1}^{N} \Big(\sum_{\tau=0}^{T} \ell_i(z_i(\tau), u_i(\tau)) + m_i(z_i(T)) \Big)$$

$$\text{subj.to } z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$$

$$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$$

$$\sum_{i=1}^{N} H_i z_i(\tau) \leq h, \qquad \tau \in [0,T]$$





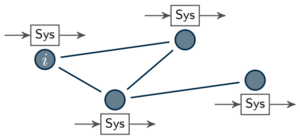
Distributed Optimal Control

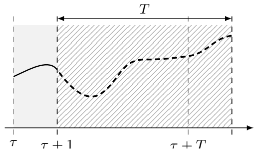
$$\min_{\substack{z_1,\dots,z_N\\u_1,\dots,u_N}} \sum_{i=1}^{N}\Big(\sum_{\tau=0}^{T}\ell_i(z_i(\tau),u_i(\tau)) + m_i(z_i(T))\Big) \qquad \sum_{i=1}^{N} f_i(x_i)$$

subj.to $z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$

$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$

$$\sum_{i=1}^{N} H_i z_i(\tau) \le h, \qquad \tau \in [0,T]$$





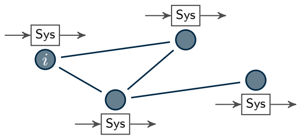
Distributed Optimal Control

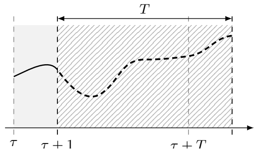
$$\min_{\substack{z_1,\dots,z_N\\u_1,\dots,u_N}} \sum_{i=1}^{N}\Big(\sum_{\tau=0}^{T}\ell_i(z_i(\tau),u_i(\tau)) + m_i(z_i(T))\Big) \qquad \sum_{i=1}^{N} f_i(x_i)$$

subj.to $z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$

$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$ $\qquad x_i \in X_i$

$$\sum_{i=1}^{N} H_i z_i(\tau) \le h, \qquad \tau \in [0,T]$$





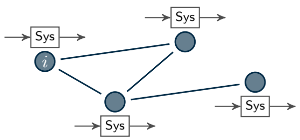
Distributed Optimal Control

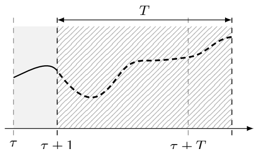
$$\min_{\substack{z_1,\dots,z_N\\u_1,\dots,u_N}} \sum_{i=1}^{N}\Big(\sum_{\tau=0}^{T}\ell_i(z_i(\tau),u_i(\tau)) + m_i(z_i(T))\Big) \qquad \sum_{i=1}^{N} f_i(x_i)$$

subj.to $z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$

$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$ $\qquad x_i \in X_i$

$$\sum_{i=1}^{N} H_i z_i(\tau) \le h, \qquad \tau \in [0,T] \qquad \sum_{i=1}^{N} g_i(x_i) \le 0$$





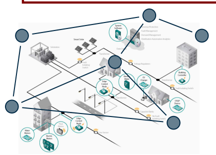
Distributed Optimal Control

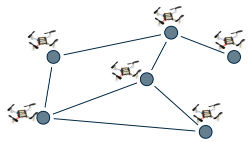
$$\min_{\substack{z_1,\dots,z_N\\u_1,\dots,u_N}} \sum_{i=1}^{N}\Big(\sum_{\tau=0}^{T}\ell_i(z_i(\tau),u_i(\tau)) + m_i(z_i(T))\Big) \qquad \sum_{i=1}^{N} f_i(x_i)$$

subj.to $z_i(\tau+1) = A_i z_i(\tau) + B_i u_i(\tau), \forall i,\ \tau \in [0,T]$

$z_i(\tau) \in Z_i,\ \ u_i(\tau) \in U_i, \qquad \forall i,\ \tau \in [0,T]$ $\qquad x_i \in X_i$

$$\sum_{i=1}^{N} H_i z_i(\tau) \le h, \qquad \tau \in [0,T] \qquad \sum_{i=1}^{N} g_i(x_i) \le 0$$





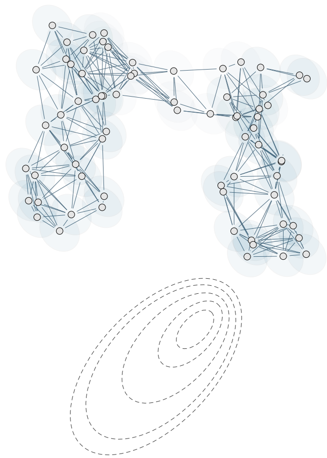
Challenge Wrap-up

Network

- Undirected/Directed
- Fixed/Time-Varying
- Asynchronous, Unreliable

Optimization

- Nonsmooth (convex), Nonconvex, Mixed-Integer, Combinatorial
- Big-Data (high dim. dec. var.)
- Dynamic, Online, Stochastic, Uncertain



Selected Topics

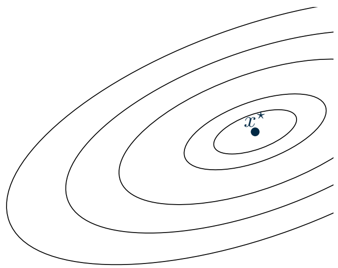
Cost coupled - Distributed Big-Data Optimization

Common cost - Constraint Exchange

Constraint coupled - Distributed Primal Decomposition

Consensus-based Method

$$\min_x \sum_{i=1}^{N} f_i(x)$$



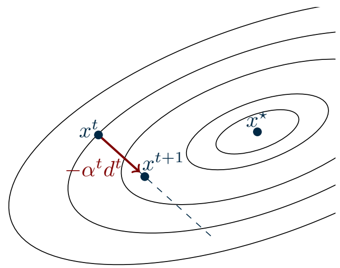
Consensus-based Method

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Centralized (sub)gradient

$$x^{t+1} = x^t - \alpha^t \boxed{d^t}$$

update direction



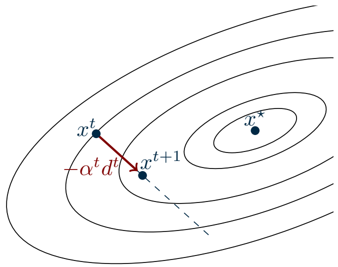
Consensus-based Method

$$\min_x \sum_{i=1}^N f_i(x)$$

Centralized (sub)gradient

$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^N \nabla f_h(x^t)}$$

update direction



Consensus-based Method

$$\min_x \sum_{i=1}^N f_i(x)$$

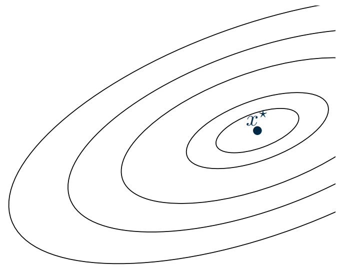
Centralized (sub)gradient

$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^N \nabla f_h(x^t)}$$

update direction

Distributed update

$$x_i^{t+1} =$$



Consensus-based Method

$$\min_x \sum_{i=1}^N f_i(x)$$

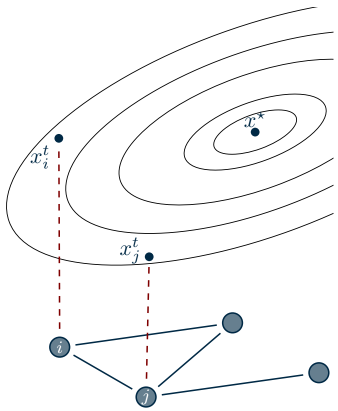
Centralized (sub)gradient

$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^N \nabla f_h(x^t)}$$

update direction

Distributed update

$$x_i^{t+1} =$$



Consensus-based Method

$$\min_x \sum_{i=1}^N f_i(x)$$

Centralized (sub)gradient

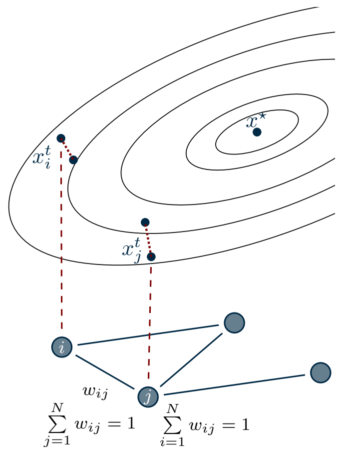
$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^N \nabla f_h(x^t)}$$

update direction

Distributed update

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t}$$

averaging



Consensus-based Method

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Centralized (sub)gradient

$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^{N} \nabla f_h(x^t)}$$

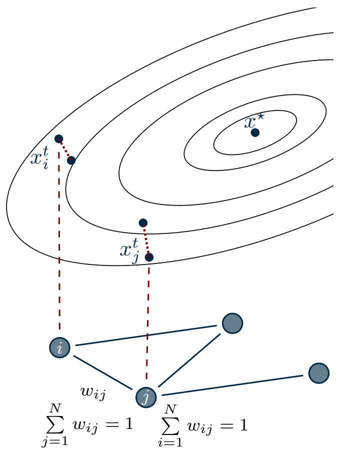
update direction

Distributed update

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t}$$

averaging

(push-sum for digraphs)



Consensus-based Method

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Centralized (sub)gradient

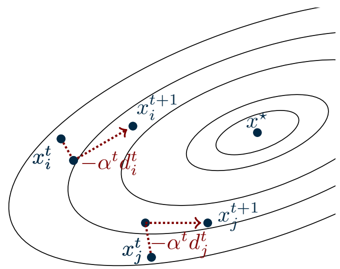
$$x^{t+1} = x^t - \alpha^t \boxed{\sum_{h=1}^{N} \nabla f_h(x^t)}$$

global

Distributed update

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha^t \boxed{d_i^t}$$

averaging local



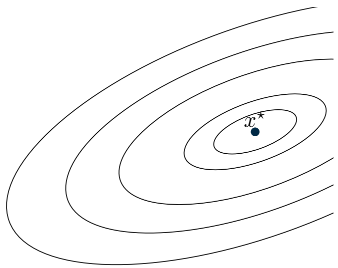
Distributed (Sub)gradient

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Distributed (sub)gradient

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha^t \boxed{\nabla f_i\Big(\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t\Big)}$$

d_i^t



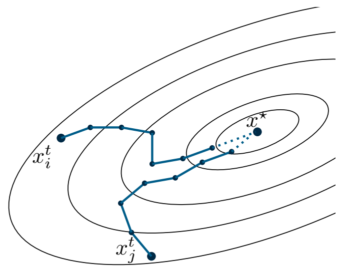
Nedić & Ozdaglar, "Distributed subgradient methods for multi-agent optimization." TAC 2009

Distributed (Sub)gradient

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Distributed (sub)gradient

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha^t \boxed{\nabla f_i\Big(\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t\Big)}$$



Theorem

Convex costs f_i, ...
Diminishing stepsize ($\alpha^t \to 0$ and ...)
Doubly-stochastic weights

$\Longrightarrow$

Consensus: $\lim_{t\to\infty} \|x_i^t - \bar{x}^t\| = 0$

Optimality: $\lim_{t\to\infty} \|\bar{x}^t - x^\star\| = 0$

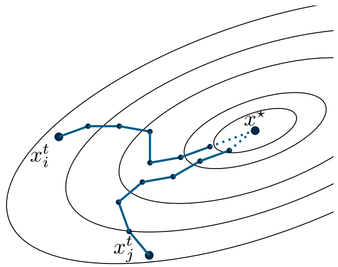
Nedić & Ozdaglar, "Distributed subgradient methods for multi-agent optimization." TAC 2009

Distributed (Sub)gradient

$$\min_x \sum_{i=1}^{N} f_i(x)$$

Distributed (sub)gradient

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha^t \boxed{\nabla f_i\Big(\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t\Big)}$$



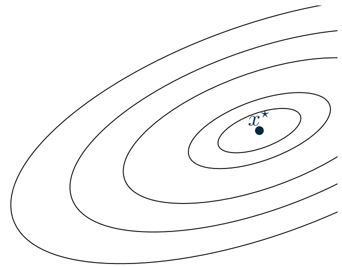
Some early literature

[Johansson,CDC'08], [Nedic,TAC'09], [Nedic,TAC'10], [Cattivelli,TSP'10], [Ram,JOTA'10], [Lobel,TAC'11], [Wang,CDC'11], [Chen,TSP'12], [Lu,TAC'12], [Bianchi,TAC'13], [Lee,STSP'13], [Jakovetic,TAC'14], [Shi,SJO'15], [Shi,TSP'15], [Nedic,TAC'15], ...

Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

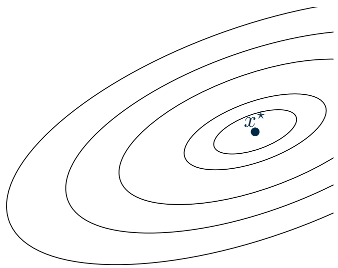
$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{\quad d_i^t \quad}$$



Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{\quad d_i^t \quad}$$



Main idea:

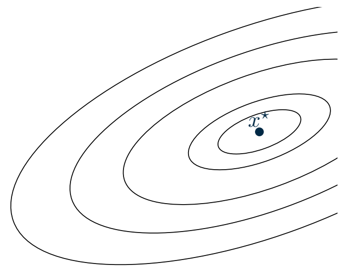
$$d_i^t \xrightarrow[t \to \infty]{} \boxed{\frac{1}{N} \sum_{h=1}^{N} \nabla f_h(x_h^t)}$$

average of gradients

Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

$$x_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{\quad d_i^t \quad}$$



Dynamic average consensus

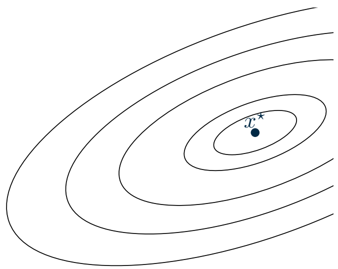
$$d_i^{t+1} = \boxed{\sum_{j \in \mathcal{N}_i} \widetilde{w}_{ij}\, d_j^t} + \boxed{\Big(\nabla f_i(x_i^{t+1}) - \nabla f_i(x_i^t)\Big)}$$

Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

$$x_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{d_i^t}$$

$$d_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} \widetilde{w}_{ij}\, d_j^t} + \boxed{\Big(\nabla f_i(x_i^{t+1}) - \nabla f_i(x_i^t)\Big)}$$

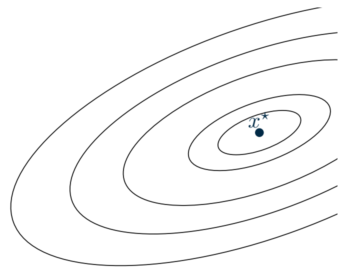


Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

$$x_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{d_i^t}$$

$$d_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} \widetilde{w}_{ij}\, d_j^t} + \boxed{\Big(\nabla f_i(x_i^{t+1}) - \nabla f_i(x_i^t)\Big)}$$



Theorem

Smooth (nonconvex) costs f_i, ...
Constant stepsize α
Row (w_{ij}) Clmn $(\tilde{w}_{ij})$ stoch. weights

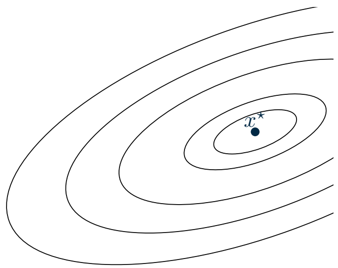
$\Rightarrow$ Consensus: $\lim_{t\to\infty} \|x_i^t - \bar{x}^t\| = 0$; Optimality: $\lim_{t\to\infty} \|\bar{x}^t - x^\star\| = 0$ } linear rate

Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

$$x_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} w_{ij}\, x_j^t} - \alpha \boxed{d_i^t}$$

$$d_i^{t+1} = \boxed{\sum_{j\in\mathcal{N}_i} \widetilde{w}_{ij}\, d_j^t} + \boxed{\Big(\nabla f_i(x_i^{t+1}) - \nabla f_i(x_i^t)\Big)}$$



Some history of gradient tracking

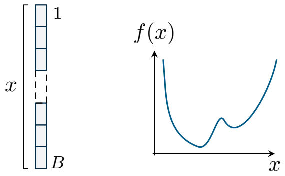
[Zanella,CDC-ECC'11], [DiLorenzo,CAMSAP'15], [Xu,CDC'15], [Nedic,CDC'16], [Qu,CDC'16], [Sun,Asilomar'16], [Qu,CDC'17], [Xi,TAC'18], [Xu,TAC'18], [Xin,L-CSS'18], ...

Other Approaches for Cost Coupled

- Dual decomposition
- Primal-Dual
- Alternating Direction Method of Multipliers (ADMM)

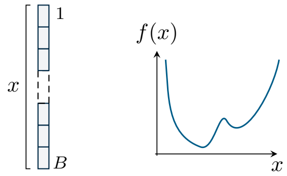
Block-wise Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$



Block-wise Gradient Tracking

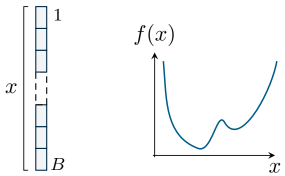
$$\min_x \sum_{i=1}^{N} f_i(x)$$



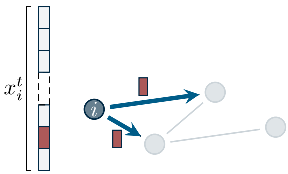
collaboration with G. Scutari and Y. Sun (Purdue University)

Block-wise Gradient Tracking

$$\min_x \sum_{i=1}^{N} f_i(x)$$

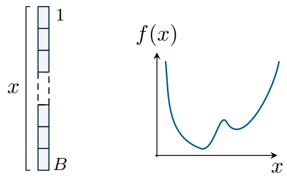


minimizing w.r.t. x too costly
transmitting x unaffordable
→ block-wise optimization/communication



Block-wise Gradient Tracking

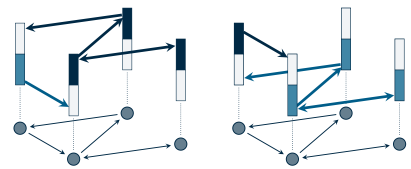
$$\min_x \sum_{i=1}^{N} f_i(x)$$



Block selection rule:
ℓ_i^t block selected at time t by agent i

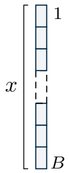
Block-dependent neighbor set:
neighbors of i sending block ℓ at time t

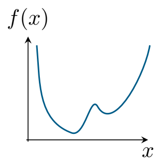
$$\mathcal{N}_{i,\ell}^t \triangleq \{j \in \mathcal{N}_i \mid \ell_j^t = \ell\} \cup \{i\} \subseteq \mathcal{N}_i$$



Block-wise Gradient Tracking

$$\min_x \sum_{i=1}^N f_i(x)$$





$$\Delta x^t_{(i,\ell)} = \begin{cases} -d^t_{(i,\ell)}, & \text{if } \ell = \ell_i^t \\ 0, & \text{otherwise} \end{cases}$$

$$x^{t+1}_{(i,\ell)} = \boxed{\sum_{j \in \mathcal{N}^t_{i,\ell}} \frac{w^t_{ij\ell}\phi^t_{(j,\ell)}}{\phi^{t+1}_{(i,\ell)}} x^t_{(j,\ell)}} + \alpha^t \boxed{\phi^t_{(i,\ell)} \Delta x^t_{(i,\ell)}}$$

$$d^{t+1}_{(i,\ell)} = \boxed{\sum_{j \in \mathcal{N}^t_{i,\ell}} \frac{w^t_{ij\ell}\phi^t_{(j,\ell)}}{\phi^{t+1}_{(i,\ell)}} d^t_{(j,\ell)}} + \boxed{\frac{\nabla_\ell f_i(x^{t+1}_{(i,:)}) - \nabla_\ell f_i(x^t_{(i,:)})}{\phi^{t+1}_{(i,\ell)}}}$$

Block-wise Gradient Tracking Convergence

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by block-wise gradient tracking with

- strongly connected graph + ess. cyclic block-selection rule
- diminishing step-size α^t
- smooth (nonconvex) costs f_i, ...

Let $\bar{x}^t \triangleq \frac{1}{N} \sum_{i=1}^N \phi^t_{(i,:)} x^t_{(i,:)}$, then

(i) consensus: $\|x^t_{(i,:)} - \bar{x}^t\| \to 0$ as $t \to \infty$, for all $i \in \{1, \ldots, N\}$;

(ii) convergence: every limit point $\{\bar{x}^t\}_{t \geq 0}$ is a stationary solution.

Notarnicola *et al.*, "Distributed Big-Data Optimization via Block-wise Gradient Tracking." arXiv 1808.07252

Block-wise Gradient Tracking Convergence

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by block-wise gradient tracking with

- strongly connected graph + ess. cyclic block-selection rule
- diminishing step-size α^t
- smooth (nonconvex) costs f_i, ...

Let $\bar{x}^t \triangleq \frac{1}{N} \sum_{i=1}^N \phi^t_{(i,:)} x^t_{(i,:)}$, then

(i) consensus: $\|x^t_{(i,:)} - \bar{x}^t\| \to 0$ as $t \to \infty$, for all $i \in \{1, \ldots, N\}$;

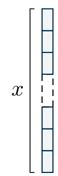
(ii) convergence: every limit point $\{\bar{x}^t\}_{t \geq 0}$ is a stationary solution.

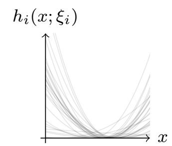
Extensions: regularization and constraints, single-block gradient update

Notarnicola *et al.*, "Distributed Big-Data Optimization via Block-wise Gradient Tracking." arXiv 1808.07252

Stochastic Big-Data Optimization

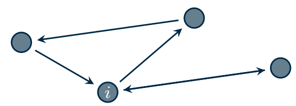
$$\min_x \sum_{i=1}^N \mathbb{E}_{\xi_i}\left[h_i(x;\xi_i)\right]$$





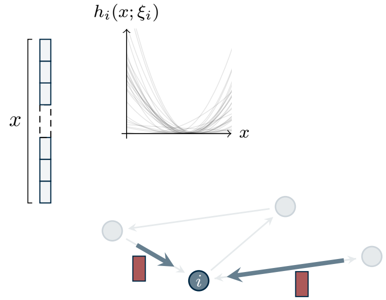
Problem features:

- Big-data
- Stochastic
- Nonsmooth (Convex)



Stochastic Big-Data Optimization

$$\min_x \sum_{i=1}^{N} \mathbb{E}_{\xi_i}\left[h_i(x;\xi_i)\right]$$

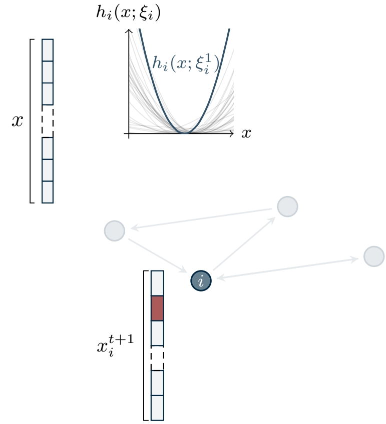


Distributed Block Subgradient

$$y_i^t = \sum_{j\in\mathcal{N}_i} w_{ij} x_j^t$$

Stochastic Big-Data Optimization

$$\min_x \sum_{i=1}^{N} \mathbb{E}_{\xi_i}\left[h_i(x;\xi_i)\right]$$



Distributed Block Subgradient

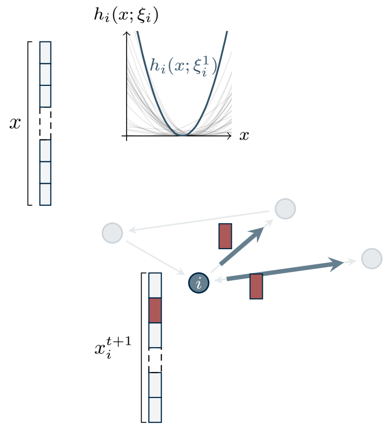
$$y_i^t = \sum_{j\in\mathcal{N}_i} w_{ij} x_j^t$$

DRAW ℓ_i^t and ξ_i^t

$$x_{i,\ell}^{t+1} = \begin{cases} y_{i,\ell}^t - \alpha_i^t \left[\nabla h_i(y_i^t;\xi_i^t)\right]_\ell, & \text{if } \ell = \ell_i^t \\ x_{i,\ell}^t, & \text{otherwise} \end{cases}$$

Stochastic Big-Data Optimization

$$\min_x \sum_{i=1}^{N} \mathbb{E}_{\xi_i}\left[h_i(x;\xi_i)\right]$$



Distributed Block Subgradient

$$y_i^t = \sum_{j\in\mathcal{N}_i} w_{ij} x_j^t$$

DRAW ℓ_i^t and ξ_i^t

$$x_{i,\ell}^{t+1} = \begin{cases} y_{i,\ell}^t - \alpha_i^t \left[\nabla h_i(y_i^t;\xi_i^t)\right]_\ell, & \text{if } \ell = \ell_i^t \\ x_{i,\ell}^t, & \text{otherwise} \end{cases}$$

Distributed Block Subgradient Convergence

Theorem

Let $\{x_1^t,\dots,x_N^t\}_{t\geq 0}$ be generated by distributed block subgradient with

- α_i^t diminishing
- doubly stochastic weights
- unbiased stochastic subgradients

Then, for all $i \in \{1,\dots,N\}$

$$f_{\text{best}}(x_i^t) \xrightarrow[t\to\infty]{} f^\star$$

Farina & Notarstefano, "Randomized Block Proximal Methods for Distributed Stochastic Big-Data Optimization" arXiv 2019

Distributed Block Subgradient Convergence

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by distributed block subgradient with

- α_i^t diminishing
- doubly stochastic weights
- unbiased stochastic subgradients

Then, for all $i \in \{1, \ldots, N\}$

$$f_{\text{best}}(x_i^t) \xrightarrow[t \to \infty]{} f^\star$$

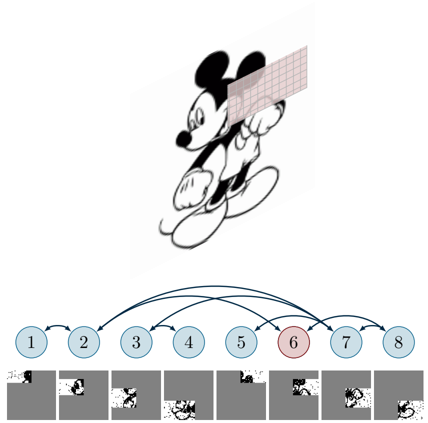
Extensions

- constraint set $X \subseteq \mathbb{R}^d$ and proximal mapping
- awake/idle nodes

Farina & Notarstefano, "Randomized Block Proximal Methods for Distributed Stochastic Big-Data Optimization" arXiv 2019

Distributed Submodular Minimization

$$\min_{X \subseteq V} \sum_{i=1}^N F_i(X)$$



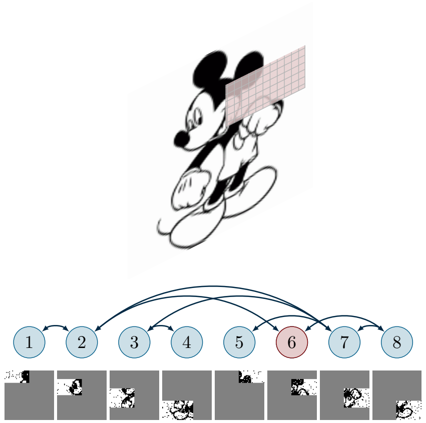
Farina, Testa & Notarstefano, "Distributed Submodular Minimization via Block-Wise Updates and Communications" arXiv 2019

Distributed Submodular Minimization

$$\min_{X \subseteq V} \sum_{i=1}^N F_i(X)$$

Lovàsz extension

$$\min_{x \in [0,1]^{|V|}} \sum_{i=1}^N f_i(x)$$



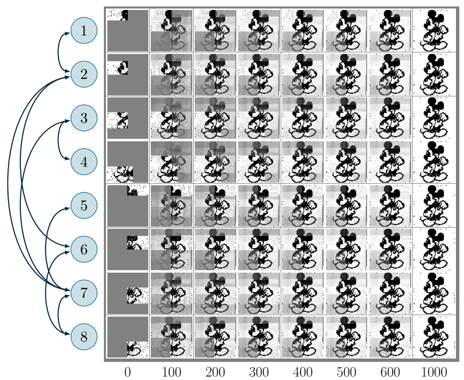
Farina, Testa & Notarstefano, "Distributed Submodular Minimization via Block-Wise Updates and Communications" arXiv 2019

Distributed Submodular Minimization

$$\min_{X \subseteq V} \sum_{i=1}^N F_i(X)$$

Lovàsz extension

$$\min_{x \in [0,1]^{|V|}} \sum_{i=1}^N f_i(x)$$



Farina, Testa & Notarstefano, "Distributed Submodular Minimization via Block-Wise Updates and Communications" arXiv 2019

Selected Topics

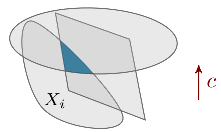
Cost coupled - Distributed Big-Data Optimization

Common cost - Constraint Exchange

Constraint coupled - Distributed Primal Decomposition

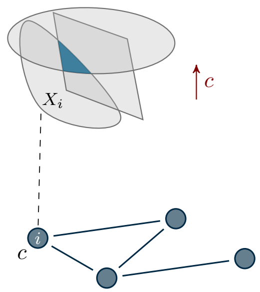
Distributed Abstract Optimization

$$\min_x \; c^\top x$$
$$\text{subj.to } x \in \bigcap_{i=1}^{N} X_i$$



Distributed Abstract Optimization

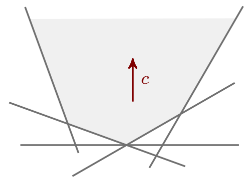
$$\min_x \; c^\top x$$
$$\text{subj.to } x \in \bigcap_{i=1}^{N} X_i$$



- agent i knows c and X_i
- convex optimization problem (abstract programs)
- asynchronous, unreliable, directed communication

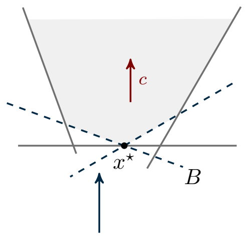
Intuition on Linear Programming

$$\min_x \; c^\top x$$
$$\text{subj.to } \underbrace{a_i^\top x \le b_i}_{X_i} \quad i \in \{1, \dots, N\}$$



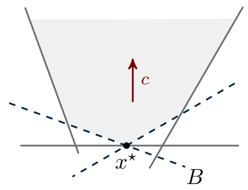
Intuition on Linear Programming

$$\min_x \; c^\top x$$
$$\text{subj.to} \;\; \underbrace{a_i^\top x \le b_i}_{X_i} \quad i \in \{1, \dots, N\}$$



Intuition on Linear Programming

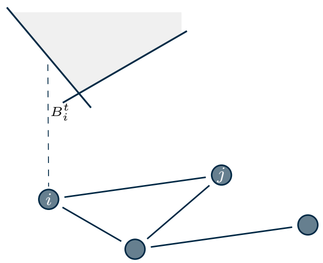
$$\min_x \; c^\top x$$
$$\text{subj.to} \;\; \underbrace{a_i^\top x \le b_i}_{X_i} \quad i \in \{1, \dots, N\}$$



Key idea: each agent stores and exchanges a candidate solution basis B_i^t

Constraints Consensus Algorithm

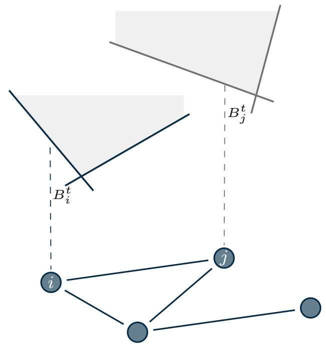
Initial constraint: X_i



Constraints Consensus Algorithm

Initial constraint: X_i

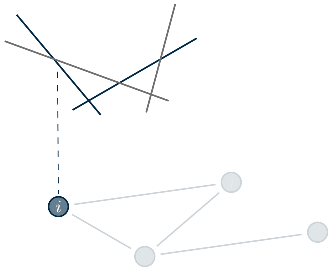
Gather B_j^t from $j \in \mathcal{N}_i^t$



Constraints Consensus Algorithm

Initial constraint: X_i

Gather B_j^t from $j \in \mathcal{N}_i^t$



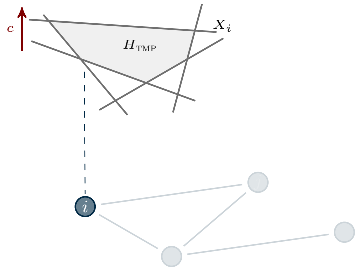
Constraints Consensus Algorithm

Initial constraint: X_i

Gather B_j^t from $j \in \mathcal{N}_i^t$

Build H_{TMP} and compute x_i^{t+1} as

$$\operatorname*{argmin}_x \; c^\top x \quad \text{subj.to } x \in H_{\text{TMP}}$$



Constraints Consensus Algorithm

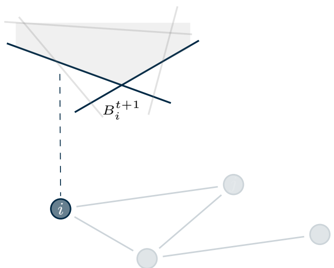
Initial constraint: X_i

Gather B_j^t from $j \in \mathcal{N}_i^t$

Build H_{TMP} and compute x_i^{t+1} as

$$\operatorname*{argmin}_x \; c^\top x \quad \text{subj.to } x \in H_{\text{TMP}}$$

Update B_i^{t+1} as a basis of x_i^{t+1}



Constraints Consensus Algorithm

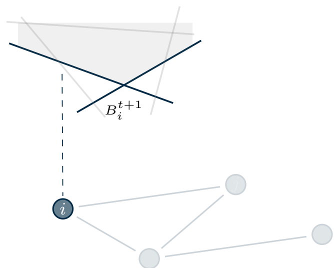
Initial constraint: X_i

Gather B_j^t from $j \in \mathcal{N}_i^t$

Build H_{TMP} and compute x_i^{t+1} as

$$\operatorname*{argmin}_x \; c^\top x \quad \text{subj.to } x \in H_{\text{TMP}}$$

Update B_i^{t+1} as a basis of x_i^{t+1}

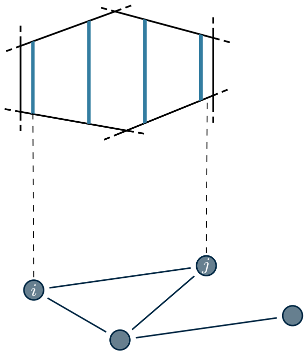


Theorem
Jointly strongly connected digraph
Abstract programs (e.g., convex programs) $\implies$ There exists $T > 0$ s.t. $x_i^t = x^\star, \ \forall i, \ \forall t > T$

Notarstefano & Bullo, "Distributed Abstract Optimization via Constraints Consensus: Theory and Applications." TAC 2011

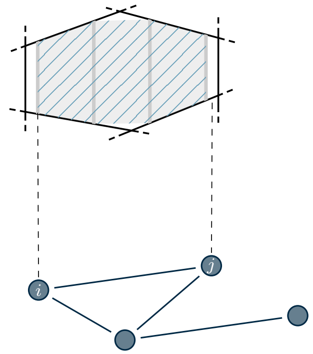
Mixed Integer Linear Programming

$$\min_x \; c^\top x$$
$$\text{subj.to} \; a_i^\top x \le b_i \quad i \in \{1, \dots, N\}$$
$$x \in \mathbb{Z}^{d_Z} \times \mathbb{R}^{d_R}$$



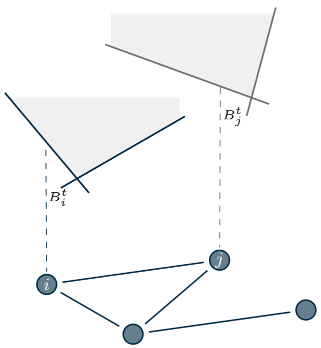
Mixed Integer Linear Programming

$$\min_x \; c^\top x$$
$$\text{subj.to} \; a_i^\top x \le b_i \quad i \in \{1, \dots, N\}$$
$$x \in \mathbb{Z}^{d_Z} \times \mathbb{R}^{d_R}$$



Cut Generation and Constraint Exchange

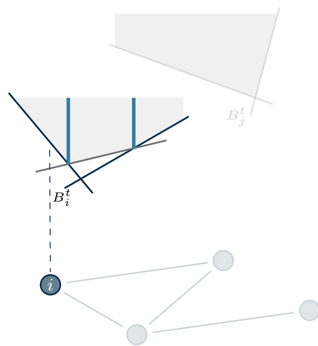
Initial constraint: X_i



Cut Generation and Constraint Exchange

Initial constraint: X_i

Generate cutting planes
(Gomory and cost-based cuts)

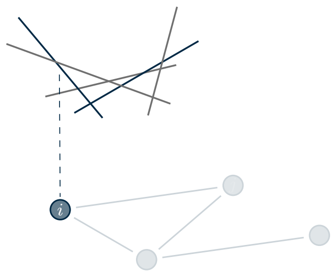


Cut Generation and Constraint Exchange

Initial constraint: X_i

Generate cutting planes
(Gomory and cost-based cuts)

Gather B_j^t from $j \in \mathcal{N}_i^t$



Cut Generation and Constraint Exchange

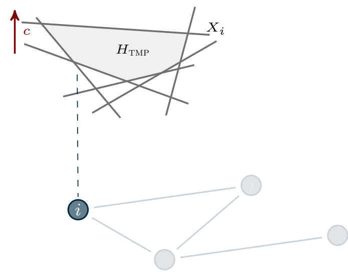
Initial constraint: X_i

Generate cutting planes
(Gomory and cost-based cuts)

Gather B_j^t from $j \in \mathcal{N}_i^t$

Build H_{TMP} and compute x_i^{t+1} as

$$\begin{aligned} &\underset{x}{\text{argmin}} \ c^\top x \\ &\text{subj.to} \ x \in H_{\text{TMP}} \end{aligned}$$



Cut Generation and Constraint Exchange

Initial constraint: X_i

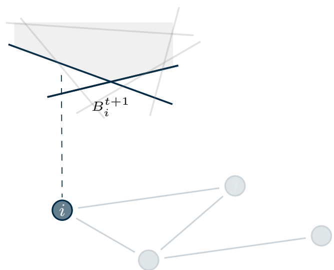
Generate cutting planes
(Gomory and cost-based cuts)

Gather B_j^t from $j \in \mathcal{N}_i^t$

Build H_{TMP} and compute x_i^{t+1} as

$$\begin{aligned} &\underset{x}{\text{argmin}} \ c^\top x \\ &\text{subj.to} \ x \in H_{\text{TMP}} \end{aligned}$$

Update B_i^{t+1} as a basis of x_i^{t+1}



Distributed Algorithm for Integer-Valued MILPs

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by cut generation and constraint exchange algorithm with

- jointly strongly connected digraph
- bounded feasible set
- integer optimal cost

Then, there exists $T > 0$ such that $\forall i = 1, \ldots, N$

$$x_t^i = x^\star, \quad \text{for all } t \geq T \quad \text{(optimal solution)}$$

Testa *et al.*, "Distributed Mixed-Integer Linear Programming via Cut Generation and Constraint Exchange." TAC 2020 (in press)

Distributed Algorithm for Integer-Valued MILPs

OPT4SMART

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by cut generation and constraint exchange algorithm with

- jointly strongly connected digraph
- bounded feasible set
- integer optimal cost

Then, there exists $T > 0$ such that $\forall i = 1, \ldots, N$

$$x_t^i = x^\star, \quad \text{for all } t \geq T \quad \text{(optimal solution)}$$

Extensions

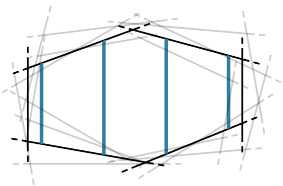
- general MILPs
- ϵ-suboptimal solution by approximate epigraph reformulation

Testa *et al.*, "Distributed Mixed-Integer Linear Programming via Cut Generation and Constraint Exchange." TAC 2020 (in press)

Distributed Robust Optimization

OPT4SMART

$$\min_x \ c^\top x$$
$$\text{subj.to } x \in \bigcap_{i=1}^{N} X_i(q), \forall q \in \mathbb{Q}$$
$$x \in \mathbb{Z}^{d_Z} \times \mathbb{R}^{d_R}$$



- convex (mixed-integer) robust programs
- agents know cost c and $X_i(q)$ ($q \in \mathbb{Q}$ uncertain parameters)
- local verification and re-optimization

collaboration with
M. Chamanbaz and R. Bouffanais (SUTD)

Chamanbaz *et al.*, "Randomized Constraints Consensus for Distributed Robust Mixed-Integer Programming." (subm)

Selected Topics

OPT4SMART

Cost coupled - Distributed Big-Data Optimization

Common cost - Constraint Exchange

Constraint coupled - Distributed Primal Decomposition

Constraint-Coupled Optimization

OPT4SMART

$$\min_{x_1,\ldots,x_N} \ \sum_{i=1}^{N} f_i(x_i)$$
$$\text{subj.to } x_i \in X_i \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \leq b$$

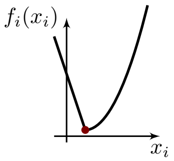
Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$

subj.to $x_i \in X_i \quad \forall i$

$$\sum_{i=1}^{N} A_i x_i \le b$$

convex functions



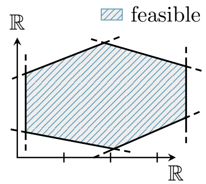
Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$

subj.to $x_i \in X_i \quad \forall i$

$$\sum_{i=1}^{N} A_i x_i \le b$$

convex, compact local sets



Giuseppe Notarstefano – **Optimization in Networkland: Challenges and Opportunities** – ECC '19, Naples – 31

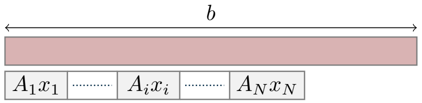
Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$

subj.to $x_i \in X_i \quad \forall i$

$$\sum_{i=1}^{N} A_i x_i \le b$$

linear coupling constraints



Giuseppe Notarstefano – **Optimization in Networkland: Challenges and Opportunities** – ECC '19, Naples – 31

Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$

subj.to $x_i \in X_i \quad \forall i$

$$\sum_{i=1}^{N} A_i x_i \le b$$

Some (incomplete) literature

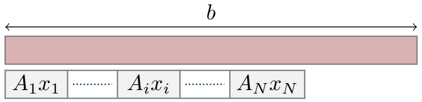
Duality-based: [Bürger,TAC'14], [Chang,TAC'14], [Simonetto,JOTA'16], [Falsone,Aut'17], ...

Resource allocation: [Lakshmanan,SJO'08], [Necoara,TAC'13], [Cherukuri,TCNS'15], ...

Giuseppe Notarstefano – **Optimization in Networkland: Challenges and Opportunities** – ECC '19, Naples – 31

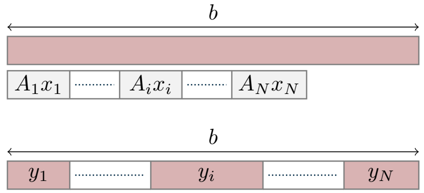
Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$
$$\text{subj.to } x_i \in X_i \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



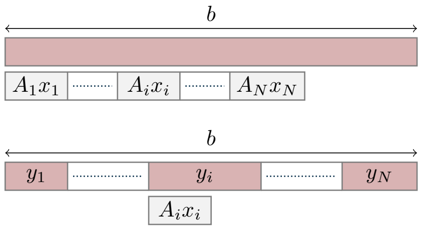
Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$
$$\text{subj.to } x_i \in X_i \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



Constraint-Coupled Optimization

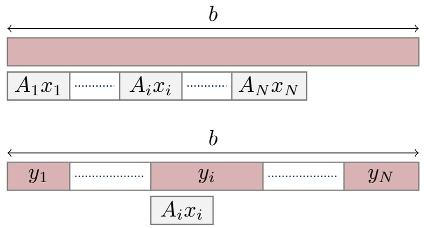
$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$
$$\text{subj.to } x_i \in X_i \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



$$\min_{x_i} \ f_i(x_i)$$
$$\text{subj.to } x_i \in X_i$$
$$A_i x_i \le y_i$$

Constraint-Coupled Optimization

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} f_i(x_i)$$
$$\text{subj.to } x_i \in X_i \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



$$p_i(y_i) = \begin{array}{|l|} \min_{x_i} \ f_i(x_i) \\ \text{subj.to } x_i \in X_i \\ \quad A_i x_i \le y_i \end{array}$$

Primal Decomposition

$p_1(y_1)$ ⋯ $p_i(y_i)$ ⋯ $p_N(y_N)$

$$\min_{x_i} \; f_i(x_i)$$
$$\text{subj.to } \; x_i \in X_i$$
$$A_i x_i \le y_i$$

Primal Decomposition

$$\min_{y_1,\dots,y_N} \; \sum_{i=1}^{N} p_i(y_i)$$
$$\text{subj.to } \; \sum_{i=1}^{N} y_i = b$$
$$y_i \in Y_i \quad \forall\, i$$

$p_1(y_1)$ ⋯ $p_i(y_i)$ ⋯ $p_N(y_N)$

$$\min_{x_i} \; f_i(x_i)$$
$$\text{subj.to } \; x_i \in X_i$$
$$A_i x_i \le y_i$$

Primal Decomposition

master problem

$$\min_{y_1,\dots,y_N} \; \sum_{i=1}^{N} p_i(y_i)$$
$$\text{subj.to } \; \sum_{i=1}^{N} y_i = b$$
$$y_i \in Y_i \quad \forall\, i$$

$p_1(y_1)$ ⋯ $p_i(y_i)$ ⋯ $p_N(y_N)$

subprob 1

$$\min_{x_i} \; f_i(x_i)$$
$$\text{subj.to } \; x_i \in X_i$$
$$A_i x_i \le y_i$$

subprob N

Primal Decomposition

master problem

$$\min_{y_1,\dots,y_N} \; \sum_{i=1}^{N} p_i(y_i)$$
$$\text{subj.to } \; \sum_{i=1}^{N} y_i = b$$
$$y_i \in Y_i \quad \forall\, i$$

$p_1(y_1)$ ⋯ $p_i(y_i)$ ⋯ $p_N(y_N)$

subprob 1

$$\min_{x_i} \; f_i(x_i)$$
$$\text{subj.to } \; x_i \in X_i$$
$$A_i x_i \le y_i$$

subprob N

Subgradient Method on Master Problem

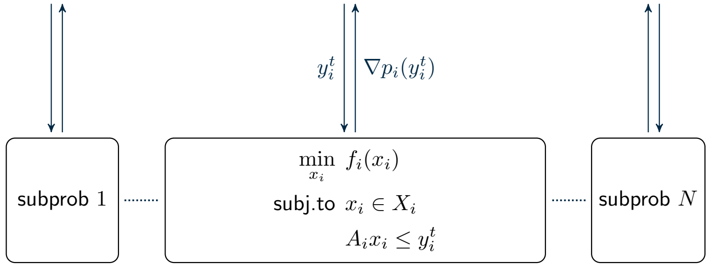
centralized update

$$y_i^{t+1} = y_i^t - \alpha^t \sum_{j=1}^{N} \Big(\nabla p_i(y_i^t) - \nabla p_j(y_j^t) \Big) \qquad \forall\, i$$

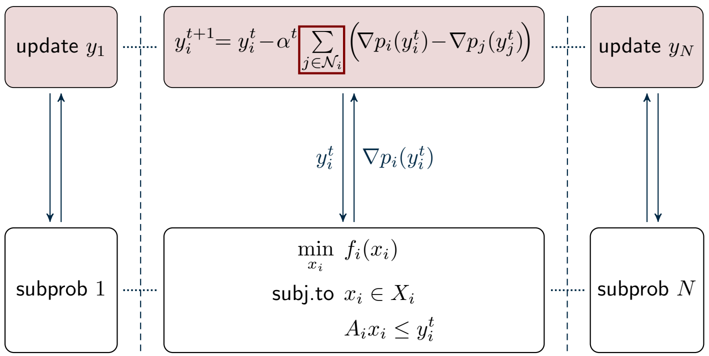
Subgradient Method on Master Problem

centralized update

$$y_i^{t+1} = y_i^t - \alpha^t \sum_{j=1}^{N} \Big(\nabla p_i(y_i^t) - \nabla p_j(y_j^t) \Big) \qquad \forall\, i$$



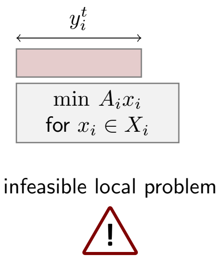
Subgradient Method on Master Problem



More on Subproblems

Q1: What if subproblem is infeasible?

$$\begin{aligned} \min_{x_i} \quad & f_i(x_i) \\ \text{subj.to} \quad & x_i \in X_i \\ & A_i x_i \leq y_i^t \end{aligned}$$

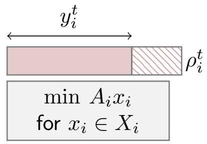


Q2: How to compute $\nabla p_i(y_i^t)$?

More on Subproblems

Q1: What if subproblem is infeasible?

$$\min_{x_i, \rho_i} \quad f_i(x_i) + M\rho_i$$
$$\text{subj.to} \quad x_i \in X_i, \;\; \rho_i \geq 0$$
$$A_i x_i \leq y_i^t + \rho_i \mathbf{1}$$



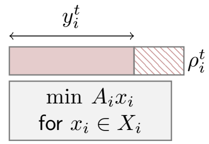
transient violation

Q2: How to compute $\nabla p_i(y_i^t)$?

More on Subproblems

Q1: What if subproblem is infeasible?

$$\min_{x_i, \rho_i} \quad f_i(x_i) + M\rho_i$$
$$\text{subj.to} \quad x_i \in X_i, \;\; \rho_i \geq 0$$
$$\mu_i : A_i x_i \leq y_i^t + \rho_i \mathbf{1}$$



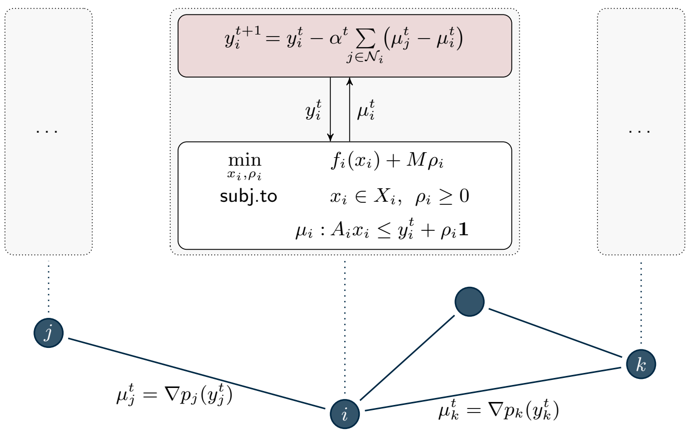
transient violation

Q2: How to compute $\nabla p_i(y_i^t)$?

Use multiplier μ_i^t of $A_i x_i \leq y_i^t + \rho_i \mathbf{1}$:

$$\nabla p_i(y_i^t) = -\mu_i^t$$

Distributed Primal Decomposition Algorithm



Distributed Primal Decomposition Convergence

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by distributed algorithm with

- α^t diminishing step-size
- $M > 0$ sufficiently large

Then:

- $\lim_{t \to \infty} \sum_{i=1}^{N} f_i(x_i^t) = f^\star$
- $\{x_1^t, \ldots, x_N^t\}_{t \geq 0} \xrightarrow[\text{every limit point}]{} (x_1^\star, \ldots, x_N^\star)$ optimal solution

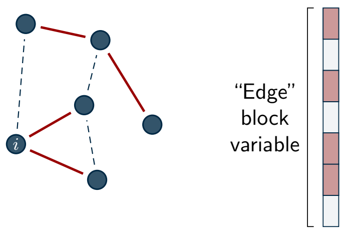
Notarnicola & Notarstefano, "Constraint Coupled Distributed Optimization: a Relaxation and Duality Approach." TCNS 2020 (to appear)

Extension to Time-varying Graphs

OPT4SMART

$$y_i^{t+1} = y_i^t - \alpha^t \boxed{\sum_{j \in \mathcal{N}_i^t}} (\mu_j^{t+1} - \mu_i^{t+1})$$

random edge activation



Analysis approach

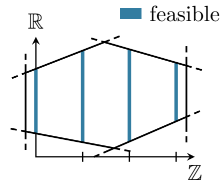
- Block subgradient method
- Random block (edge) selection

Camisa *et al.*, "Distributed Constraint-Coupled Optimization over Random Time-Varying Graphs via Primal Decomposition and Block Subgradient Approaches." (subm. to conf.)

Mixed-Integer Linear Programs

OPT4SMART

$$\min_{x_1,\ldots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in X_i^{\text{MILP}} \subset \mathbb{Z}^{z_i} \times \mathbb{R}^{r_i} \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$

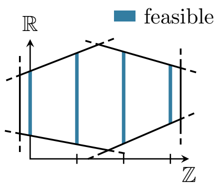


local sets X_i^{MILP}

Mixed-Integer Linear Programs

OPT4SMART

$$\min_{x_1,\ldots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in X_i^{\text{MILP}} \subset \mathbb{Z}^{z_i} \times \mathbb{R}^{r_i} \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



local sets X_i^{MILP}

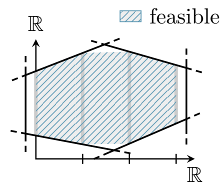
Challenge: large-scale and NP-hard problem to be solved in short time

Goal: fast computation of "high-quality" suboptimal solutions

Mixed-Integer Linear Programs

OPT4SMART

$$\min_{x_1,\ldots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in \text{conv}(X_i^{\text{MILP}}) \quad \forall i$$
$$\sum_{i=1}^{N} A_i x_i \le b$$



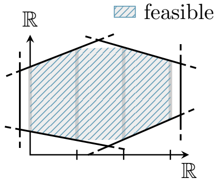
local sets $\text{conv}(X_i^{\text{MILP}})$

Mixed-Integer Linear Programs

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$

subj.to $x_i \in \text{conv}(X_i^{\text{MILP}}) \quad \forall\, i$

$$\sum_{i=1}^{N} A_i x_i \le b \longleftarrow S \text{ constraints}$$



local sets $\text{conv}(X_i^{\text{MILP}})$

Theorem (Shapley-Folkman): Let $(x_1^{\text{conv}},\dots,x_N^{\text{conv}})$ be unique optimal solution

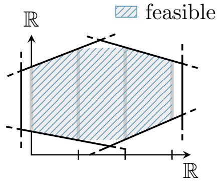
Then x_i^{conv} already mixed integer for at least $N-S$ agents

Mixed-Integer Linear Programs

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$

subj.to $x_i \in \text{conv}(X_i^{\text{MILP}}) \quad \forall\, i$

$$\sum_{i=1}^{N} A_i x_i \le b \longleftarrow S \text{ constraints}$$



local sets $\text{conv}(X_i^{\text{MILP}})$

Theorem (Shapley-Folkman): Let $(x_1^{\text{conv}},\dots,x_N^{\text{conv}})$ be unique optimal solution

Then x_i^{conv} already mixed integer for at least $N-S$ agents

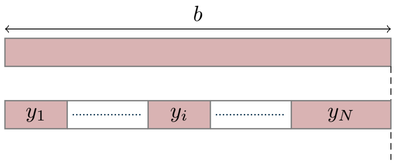
State of the art: dual decomposition (Vujanic,Aut'16, Falsone,Aut'19)

Restriction of Coupling Constraints

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$

subj.to $x_i \in \text{conv}(X_i^{\text{MILP}}) \quad \forall\, i$

$$\sum_{i=1}^{N} A_i x_i \le b$$

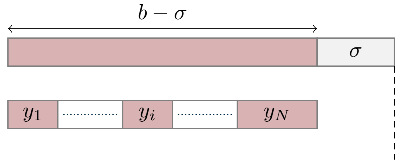


Restriction of Coupling Constraints

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$

subj.to $x_i \in \text{conv}(X_i^{\text{MILP}}) \quad \forall\, i$

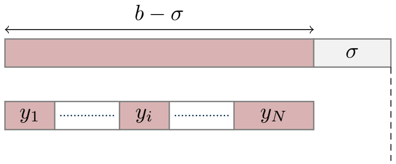
$$\sum_{i=1}^{N} A_i x_i \le b - \boldsymbol{\sigma}$$



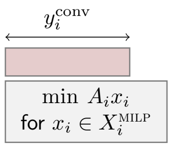
Restriction of Coupling Constraints

OPT4SMART

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in \mathrm{conv}(X_i^{\text{MILP}}) \quad \forall\, i$$
$$\sum_{i=1}^{N} A_i x_i \le b - \boldsymbol{\sigma}$$



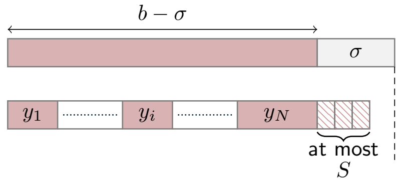
For x_i^{conv} not mixed integer



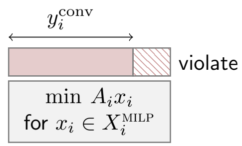
Restriction of Coupling Constraints

OPT4SMART

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in \mathrm{conv}(X_i^{\text{MILP}}) \quad \forall\, i$$
$$\sum_{i=1}^{N} A_i x_i \le b - \boldsymbol{\sigma}$$



For x_i^{conv} not mixed integer



Distributed Primal Decomposition for MILP

OPT4SMART

Convexified problem

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in \mathrm{conv}(X_i^{\text{MILP}}) \quad \forall\, i$$
$$\sum_{i=1}^{N} A_i x_i \le b - \boldsymbol{\sigma}$$

Solve with Distributed Primal Decomposition

$$y_i^t \longrightarrow y_i^{\text{conv}}$$

Distributed Primal Decomposition for MILP

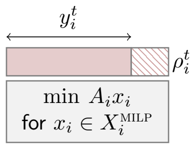
OPT4SMART

Convexified problem

$$\min_{x_1,\dots,x_N} \sum_{i=1}^{N} c_i^\top x_i$$
$$\text{subj.to } x_i \in \mathrm{conv}(X_i^{\text{MILP}}) \quad \forall\, i$$
$$\sum_{i=1}^{N} A_i x_i \le b - \boldsymbol{\sigma}$$

Solve with Distributed Primal Decomposition

$$y_i^t \longrightarrow y_i^{\text{conv}}$$



Compute $x_i^t \in X_i^{\text{MILP}}$ with minimal violation of $A_i x_i \le y_i^t$

Finite-time Theoretical Results

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by distributed algorithm for MILP.

Camisa & Notarnicola & Notarstefano, "A Primal Decomposition Method with Suboptimality Bounds for Distributed Mixed-Integer Linear Programming." CDC 2018

Finite-time Theoretical Results

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by distributed algorithm for MILP.

Then, for given "extra restriction", $\exists\, T > 0$ such that

- MILP feasibility:

$$x_i^t \in X_i^{\text{MILP}} \quad \forall\, i, \qquad \sum_{i=1}^{N} A_i x_i^t \leq b, \qquad \forall\, t \geq T$$

Camisa & Notarnicola & Notarstefano, "A Primal Decomposition Method with Suboptimality Bounds for Distributed Mixed-Integer Linear Programming." CDC 2018

Finite-time Theoretical Results

Theorem

Let $\{x_1^t, \ldots, x_N^t\}_{t \geq 0}$ be generated by distributed algorithm for MILP.

Then, for given "extra restriction", $\exists\, T > 0$ such that

- MILP feasibility:

$$x_i^t \in X_i^{\text{MILP}} \quad \forall\, i, \qquad \sum_{i=1}^{N} A_i x_i^t \leq b, \qquad \forall\, t \geq T$$

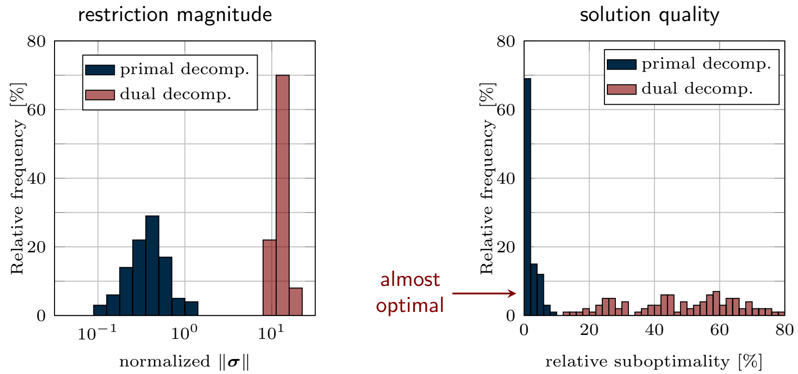
- MILP suboptimality bound:

$$\sum_{i=1}^{N} c_i^\top x_i^t - J^{\text{MILP}} \leq \text{convexification suboptimality} + \text{restriction suboptimality} + \text{distance to convergence} \qquad \forall\, t \geq T$$

Camisa & Notarnicola & Notarstefano, "A Primal Decomposition Method with Suboptimality Bounds for Distributed Mixed-Integer Linear Programming." CDC 2018

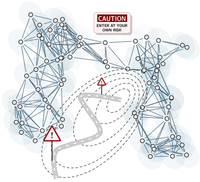
Numerical Computations

Montecarlo simulations with $X_i \subset \mathbb{Z} \times \mathbb{R}$ $(100$ instances, $N = 50$, $S = 10)$

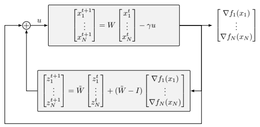


What Next?

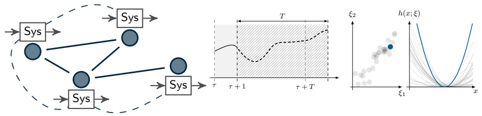
Network/optimization challenges



System theoretical approach
to distributed optimization



Optimal control of complex systems
model-based vs data-driven

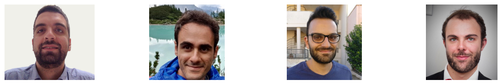


More info @ ...

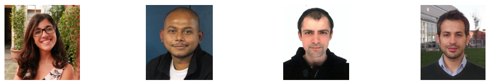
opt4smart.dei.unibo.it

Acknowledgment

OPT4SMART group



Ivano Notarnicola Postdoc @unibo | Francesco Farina Postdoc @unibo | Andera Camisa PhD student @unibo | Andrea Testa PhD student @unile



Sara Spedicato Postdoc @unibo | Sarnavi Mahesh PhD student @unile | Francesco Sasso PhD student @unile | Alessandro Rucco Former Postdoc @unile

Angelo Coluccia | Lorenzo Marconi | Alessandro Macchelli | Andrea Tilli | Nicola Mimmo

Acknowledgment

Ivano Notarnicola Postdoc @unibo | Francesco Farina Postdoc @unibo | Andera Camisa PhD student @unibo | Andrea Testa PhD student @unile

Sara Spedicato Postdoc @unibo | Sarnavi Mahesh PhD student @unile | Francesco Sasso PhD student @unile | Alessandro Rucco Former Postdoc @unile

Angelo Coluccia | Lorenzo Marconi | Alessandro Macchelli | Andrea Tilli | Nicola Mimmo

Acknowledgment

Early collaborators / Mentors


F. Allgöwer


F. Bullo


M. Egerstedt


J. Hauser

Other collaborators

- F. Bayer, M. Bürger
- G. Scutari, Y. Sun
- R. Bouffanais, M. Chamanbaz
- M. Franceschelli
- A. Franchi
- N. Bof, R. Carli, L. Schenato, D. Varagnolo
- S. Chopra
- A. Garulli, A. Giannitrapani
- A. Falsone, M. Prandini
- K. Margellos, A. Papachristodoulou, L. Romao
- M. Bin, L. Marconi

Take-home

Opportunities

- optimization is a building block in many estimation, learning, decision and control problems
- new powerful technology with massive computation and communication capability available

Challenges

- optimization (nonconvex, mixed-integer, combinatorial, big-data, stochastic, uncertain)
- network (asynchronous, unreliable, directed)

RESEARCH FUNDED BY

Take-home

Opportunities

- optimization is a building block in many estimation, learning, decision and control problems
- new powerful technology with massive computation and communication capability available

Distributed Optimization for Smart Cyber-Physical Networks
G. Notarstefano, I. Notarnicola, A. Camisa
Foundations and Trends® in Systems and Control (subm)
arxiv.org/abs/1906.10760

RESEARCH FUNDED BY