Distributed Optimization Algorithms for Networked Systems

Michael M. Zavlanos
Mechanical Engineering & Materials Science
Electrical & Computer Engineering
Computer Science
Duke University

DIMACS Workshop on Distributed Optimization, Information Processing, and Learning
Rutgers University
August 21, 2017

Distributed Optimization

Distributed ≈ Parallel

Distributed (or Decentralized)

Divide problem into smaller sub-problems (nodes)

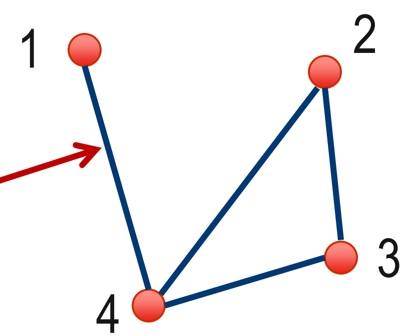
Each node solves only its assigned sub-problem (more manageable)

Only local communications between nodes (no supervisor, more privacy)

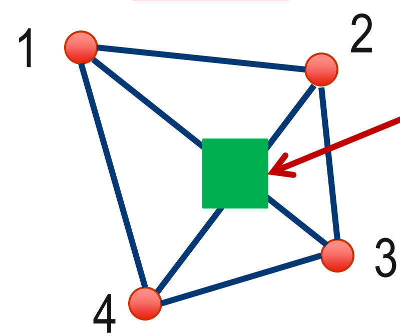
Iterative procedure until convergence

Distributed

Nodes 1&4 can communicate their decisions



Parallel

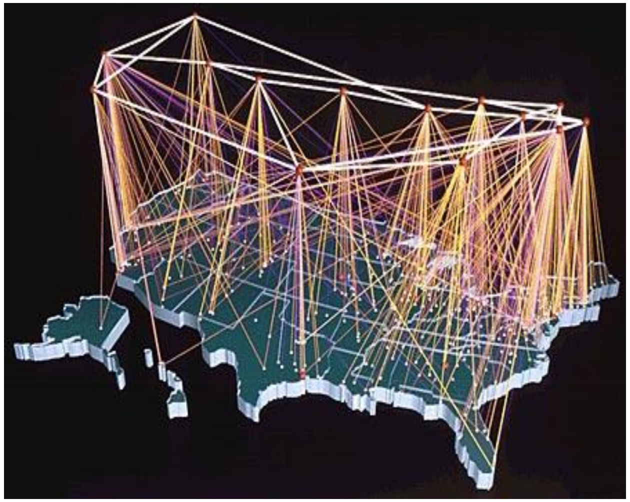


Shared memory may exist.

Why Distributed?

Centralized computation suffers from:

Poor Scalability (curse of dimensionality)
Requires supervising unit
Large communication costs
Significant Delays
Vulnerable to Changes
Security/Privacy Issues



Question to answer in Distributed methods:

Convergence to centralized solution (optimality, speed)?

Duke UNIVERSITY

Distributed Optimization Methods

Primal Decomposition

Dual Decomposition (Ordinary Lagrangians)
[Everett, 1963]

Augmented Lagrangians
Alternating Directions Method of Multipliers (ADMM) [Glowinski et al., 1970], [Eckstein and Bertsekas, 1989]
Diagonal Quadratic Approximation (DQA) [Mulvey and Ruszczyński, 1995]

Optimal Wireless Networking

Newton's Methods
Accelerated Dual Descent (ADD) [Zargham et al., 2011]
Distributed Newton Method [Wei et al., 2011]

Random Projections
[Lee and Nedic, 2013]

Distributed State Estimation

Coordinate Descent
[Mukherjee et al. , 2013], [Liu et al., 2015], [Richtarik and Takac, 2015]

Nesterov-like methods
[Nesterov, 2014], [Jakovetic et al., 2014]

Continuous-time methods
[Mateos and Cortes, 2014], [Kia et al., Arxiv], [Richert and Cortes, Arxiv]

Duke UNIVERSITY

Outline

Accelerated Distributed Augmented Lagrangians (ADAL) method
for optimal wireless networking

Fast convergence
Noisy communication

Accelerated Distributed Augmented Lagrangians (ADAL) method under noise
for optimal wireless networking

Random Approximate Projections (RAP) method with inexact data
for distributed state estimation

Light-weight
Distributed LMI constraints
Mobile networks

Duke
UNIVERSITY

Outline

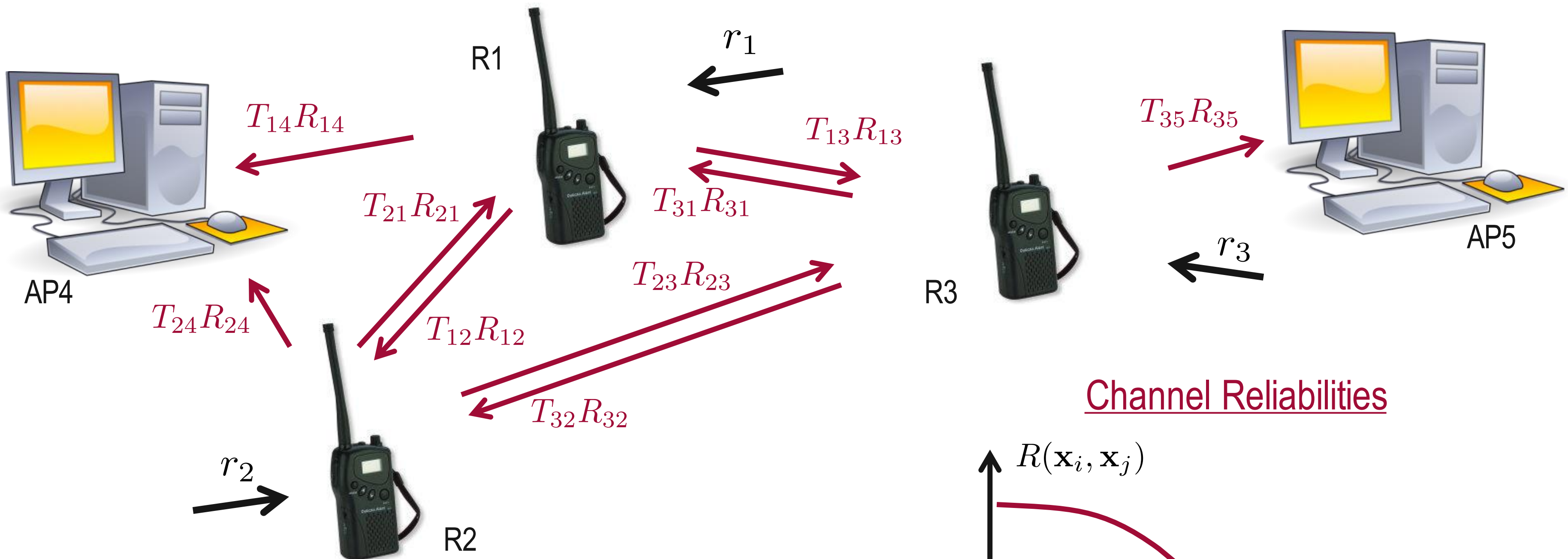
Accelerated Distributed Augmented Lagrangians (ADAL) method
for optimal wireless networking

Accelerated Distributed Augmented Lagrangians (ADAL) method under noise
for optimal wireless networking

Random Approximate Projections (RAP) method with inexact data
for distributed state estimation

Duke UNIVERSITY

Wireless Communication Networks



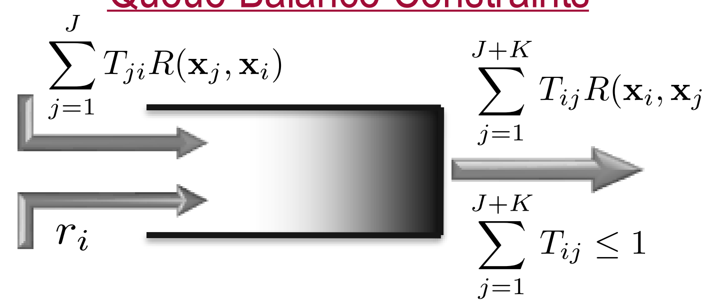
- J source nodes, K access points (APs)
- r_i: the rate of information generated at node i
- R_{ij}: the rate of information correctly transmitted from node i to node j
- T_{ij}: the fraction of time node i selects node j as its destination

Channel Reliabilities

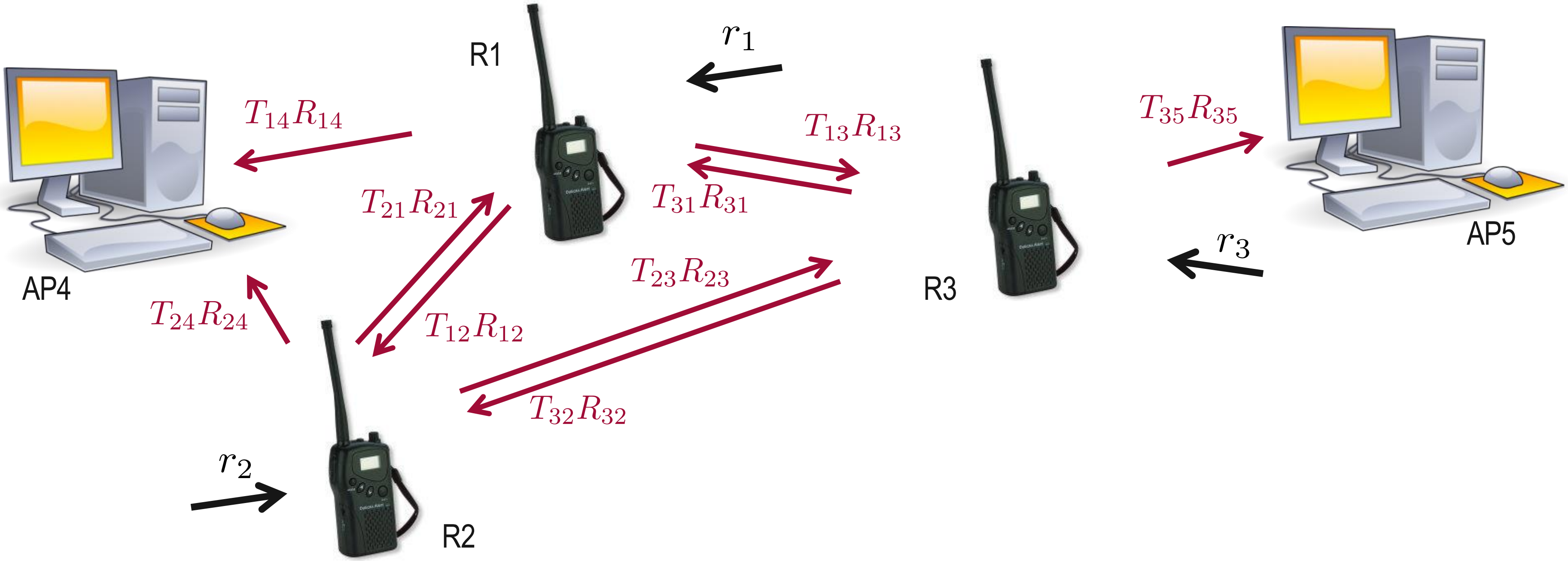
$R(\mathbf{x}_i, \mathbf{x}_j)$

$\|\mathbf{x}_i - \mathbf{x}_j\|_2$

Queue Balance Constraints



Optimal Wireless Networking



Find the routes T that maximize a utility of the rates generated at the sources, while respecting the queue constraints at the radio terminals.

Duke
UNIVERSITY

Mathematical Formulation

Optimal network flow:

Assume a static network

Network cost function

Rate constraint

$$\max_{T\in[0,1]^{J(J+K)}} \boxed{f(r(T))}$$

s.t

$$\boxed{r_{i,\min} \le r_i(T) = \sum_{j\in\mathcal{J}\cup\mathcal{K}} T_{ij}R_{ij} - \sum_{j\in\mathcal{J}} T_{ji}R_{ji},}$$

$$\boxed{\sum_{j\in\mathcal{J}\cup\mathcal{K}} T_{ij} = 1, \quad \forall\, i \in \mathcal{K}}$$

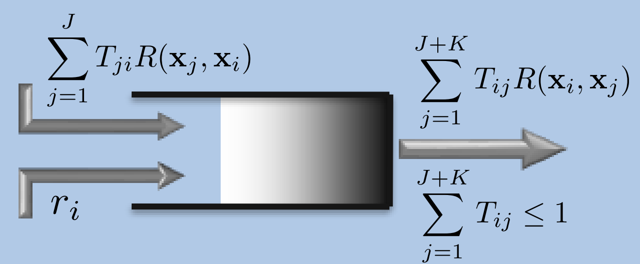
Time slot share

Linear: $$f(r(T)) = \sum_{i\in\mathcal{J}} w_i r_i(T)$$

Logarithmic: $$f(r(T)) = -\sum_{i\in\mathcal{J}} \log(r_i(T))$$

Min-Rate: $$f(r(T)) = \min_{i\in\mathcal{J}}\{r_i(T)\}$$

Rate constraint:



Dual Decomposition

Lagrangian: $$\mathcal{L}(\boldsymbol{\lambda}, \mathbf{T}) = \sum_{i=1}^{J} f_i(r_i(\mathbf{T})) + \sum_{i=1}^{J} \lambda_i \left[\sum_{j=1}^{J+K} T_{ij} R_{ij} - \sum_{j=1}^{J} T_{ji} R_{ji} - r_{i,\min} \right]$$

Local Lagrangian:

$$\mathcal{L}_i(\boldsymbol{\lambda}, \mathbf{T}) = f_i(r_i(\mathbf{T})) - \lambda_i r_{i,\min} + \sum_{j=1}^{J} T_{ij} R_{ij} (\lambda_i - \lambda_j) + \sum_{j=J+1}^{J+K} \lambda_i T_{ij} R_{ij}$$

so that $$\mathcal{L}(\boldsymbol{\lambda}, \mathbf{T}) = \sum_{i=1}^{J} \mathcal{L}_i(\boldsymbol{\lambda}, \mathbf{T})$$

Involves only primal variables $\{T_{ij}\}_{j=1}^{J+K}$ and for a given i .

Therefore, to find the variables $\{T_{ij}(\lambda)\}_{j=1}^{J+K}$ that maximize the global Lagrangian, it suffices to find the arguments that maximize the local Lagrangians.

Duke UNIVERSITY

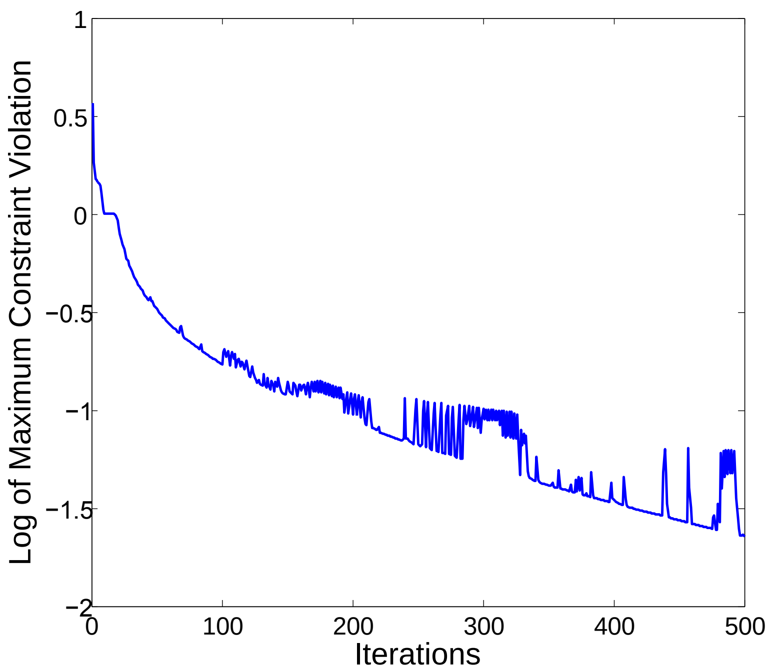
Primal-Dual Method

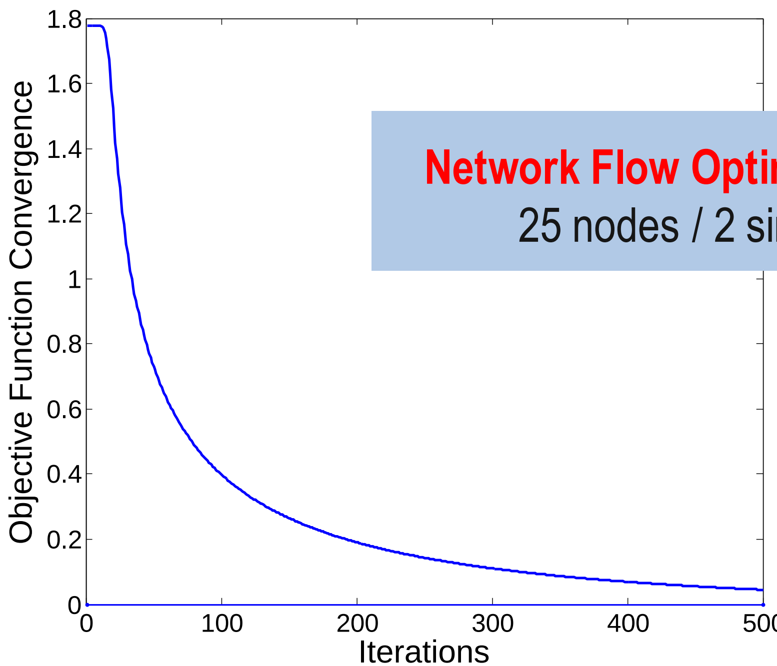
Primal Iteration:

$$\mathbf{T}^k = \operatorname*{argmax}_{\mathbf{T}\in[0,1]^{J(J+K)}} \left\{ \mathcal{L}_i(\boldsymbol{\lambda}^{\boldsymbol{k}}, \mathbf{T}) \mid \sum_{j=1}^{J+K} T_{ij} = 1 \right\}$$

Dual Iteration:

$$\lambda_i^{k+1} = \left[\lambda_i^k - \epsilon \left(\sum_{j=1}^{J+K} T_{ij}^k R_{ij} - \sum_{j=1}^{J} T_{ji}^k R_{ji} - r_{i,\min} \right) \right]^+$$





Network Flow Optimization
25 nodes / 2 sinks

Duke UNIVERSITY

Accelerated Network Optimization

Ordinary Lagrangian methods are attractive because of their simplicity, however, they converge slow. Thus, we opt for regularized methods.

Augmented Lagrangian:

Ordinary Lagrangian

$$\Lambda(\boldsymbol{\lambda}, \mathbf{T}) = \boxed{\sum_{i=1}^{J} f_i(r_i(\mathbf{T})) + \sum_{i=1}^{J} \lambda_i \left(\sum_{j=1}^{J+K} T_{ij} R_{ij} - \sum_{j=1}^{J} T_{ji} R_{ji} - r_{i,\min} \right)}$$

$$\boxed{+ \frac{\rho}{2} \sum_{i=1}^{J} \left(\sum_{j=1}^{J+K} T_{ij} R_{ij} - \sum_{j=1}^{J} T_{ji} R_{ji} - r_{i,\min} \right)}^{2}$$

Regularization term

Non-separable !!

Duke
UNIVERSITY

In Matrix Form

Local variables: $\mathbf{z}_i = [T_{i1},\ T_{i2},\ \ldots, T_{i(J+K)}] \in [0,1]^{J+K}$

Primal problem: $\min_{\mathbf{z}_i \in \mathcal{Z}_i} \quad \sum_{i=1}^{J} f_i(\mathbf{z}_i)$

s.t. $\quad \sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i = \mathbf{r}_{\min}$

Augmented Lagrangian:

$$\Lambda(\mathbf{z}, \boldsymbol{\lambda}) = \sum_{i=1}^{J} f_i(\mathbf{z}_i) + \boldsymbol{\lambda}^T \Big(\sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i - \mathbf{r}_{\min} \Big) + \frac{\rho}{2} \Big\| \sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i - \mathbf{r}_{\min} \Big\|^2$$

Duke UNIVERSITY

Method of Multipliers

Augmented Lagrangian:

$$\Lambda(\mathbf{z},\boldsymbol{\lambda}) = \sum_{i=1}^{J} f_i(\mathbf{z}_i) + \boldsymbol{\lambda}^T\Big(\sum_{i=1}^{J}\mathbf{R}_i\mathbf{z}_i - \mathbf{r}_{\min}\Big) + \frac{\rho}{2}\Big\|\sum_{i=1}^{J}\mathbf{R}_i\mathbf{z}_i - \mathbf{r}_{\min}\Big\|^2$$

Method of Multipliers (Hestenes, Powell 1969):

Step 0: Set k=1 and define initial Lagrange multipliers $\boldsymbol{\lambda}^1$
Step 1: For fixed Lagrange multipliers $\boldsymbol{\lambda}^k$, determine $\mathbf{z}^k$ as the solution of

$$\boxed{\min_{\mathbf{z}} \Lambda(\mathbf{z},\boldsymbol{\lambda}^k) \quad \text{such that} \quad \mathbf{z} \in \mathcal{Z}}$$

Step 2: If the constraints $\sum_{i=1}^{J}\mathbf{R}_i\mathbf{z}_i^k = \mathbf{r}_{\min}$ are satisfied, then stop (optimal solution found). Otherwise, set:

$$\boxed{\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho\Big(\sum_{i=1}^{J}\mathbf{R}_i\mathbf{z}_i^k - \mathbf{r}_{\min}\Big)}$$

Centralized

increase k by one and return to Step 1.

An Accelerated Distributed AL Method

Local Augmented Lagrangian:

$$\Lambda_i(\mathbf{z}_i, \mathbf{z}^k, \boldsymbol{\lambda}) = f_i(\mathbf{z}_i) + \boldsymbol{\lambda}^T \mathbf{R}_i \mathbf{z}_i + \frac{\rho}{2} \left\| \mathbf{R}_i \mathbf{z}_i + \boxed{\sum_{j \neq i} \mathbf{R}_j \mathbf{z}_j^k} - \mathbf{r}_{\min} \right\|^2$$

Step 0: Set k=1 and define initial Lagrange multipliers $\boldsymbol{\lambda}^1$ and initial primal variables $\mathbf{z}^1$

Step 1: For fixed Lagrange multipliers $\boldsymbol{\lambda}^k$, determine $\hat{\mathbf{z}}_i^k$ for every i as the solution of

$$\boxed{\min_{\mathbf{z}_i} \Lambda_i(\mathbf{z}_i, \mathbf{z}^k, \boldsymbol{\lambda}^k) \quad \text{such that} \quad \mathbf{z}_i \in \mathcal{Z}_i}$$

Step 2: Set for every i :

$$\boxed{\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \tau(\hat{\mathbf{z}}_i^k - \mathbf{z}_i^k)}$$

Step 3: If the constraints $\sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i^{k+1} = \mathbf{r}_{\min}$ are satisfied and $\mathbf{R}_i \hat{\mathbf{z}}_i^k = \mathbf{R}_i \mathbf{z}_i^k$,

then stop (optimal solution found). Otherwise, set:

$$\boxed{\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho\tau\left(\sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i^{k+1} - \mathbf{r}_{\min}\right)}$$

Increase k by one and return to Step 1.

Convergence

Assume that:

1) The functions $f_i(\mathbf{z}_i)$ are convex and the sets $\mathcal{Z}_i$ are convex and compact.
2) The Lagrange function has a saddle point $(\mathbf{z}^*, \boldsymbol{\lambda}^*)$ so that:

$$\mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}) \leq \mathcal{L}(\mathbf{z}^*, \boldsymbol{\lambda}^*) \leq \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}^*)$$

Theorem:

1) If $0 < \tau < \frac{1}{q}$ then the sequence

$$\phi(\mathbf{z}^k, \boldsymbol{\lambda}^k) = \sum_{i=1}^{N} \rho \|\mathbf{R}_i(\mathbf{z}_i^k - \mathbf{z}_i^*)\|^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}^k + \rho(1-\tau)\mathbf{r}(\mathbf{z}^k) - \boldsymbol{\lambda}^*\|^2$$

is strictly decreasing.

2) The ADAL method stops at an optimal solution of the problem or generates a sequence of $\boldsymbol{\lambda}^k$ converging to an optimal solution of it. Moreover, any sequence $\{\mathbf{z}^k\}$ generated by the ADAL algorithm has an accumulation point and any such point is an optimal solution.

Residual: $\mathbf{r}(\mathbf{z}) = \sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i - \mathbf{r}_{\min}$

Duke UNIVERSITY

Rate of Convergence

Theorem: Let $F(\mathbf{z}) = \sum_{i=1}^{N} f_i(\mathbf{z}_i)$ and denote by $\tilde{\mathbf{z}}^k = \frac{1}{k}\sum_{p=0}^{k-1} \hat{\mathbf{z}}^p$ the ergodic average of the primal variable sequence generated by ADAL at iteration k. Then,

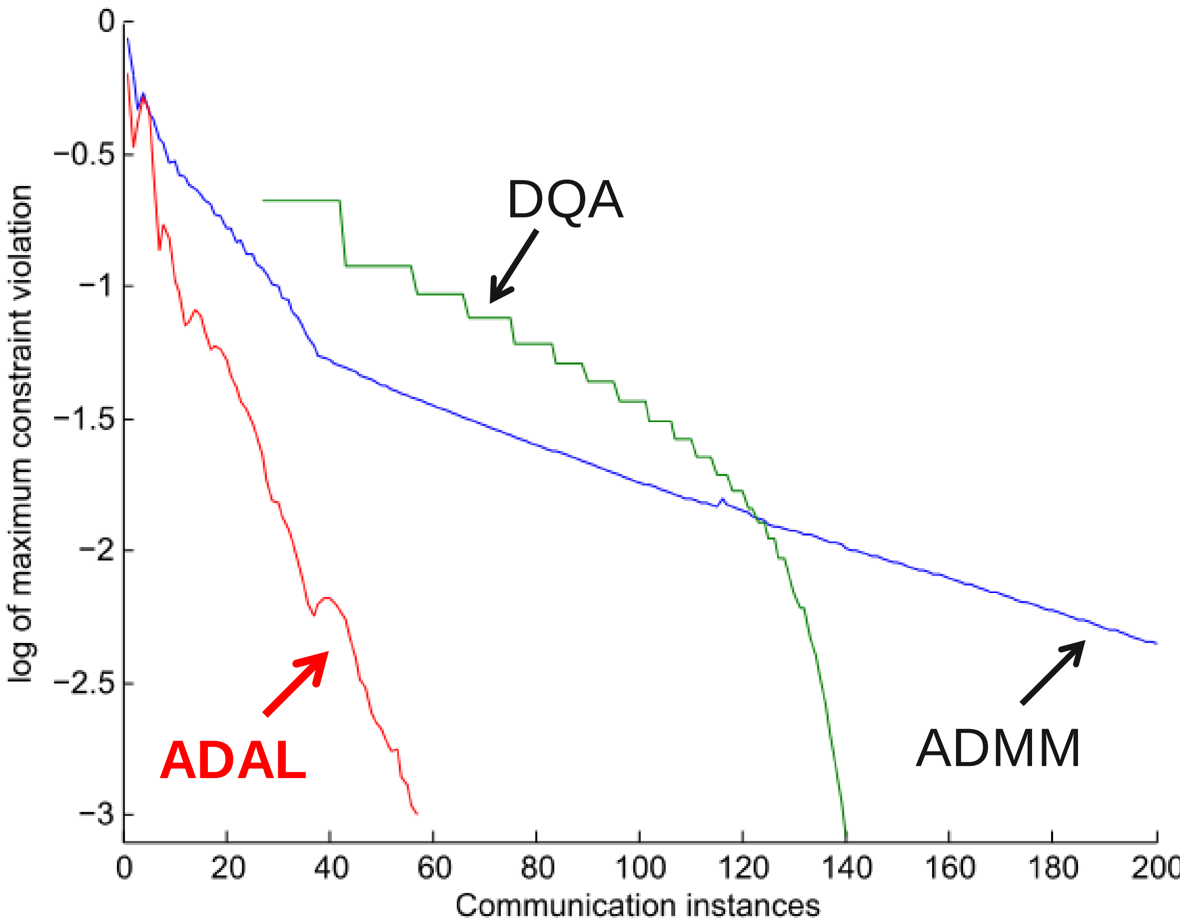
(a) $$|F(\tilde{\mathbf{z}}^k) - F(\mathbf{z}^*)| \leq \boxed{\frac{1}{2k\tau}} \max\{\phi^0(\mathbf{z}^0, \mathbf{0}), \phi^0(\mathbf{z}^0, 2\boldsymbol{\lambda}^*)\}$$

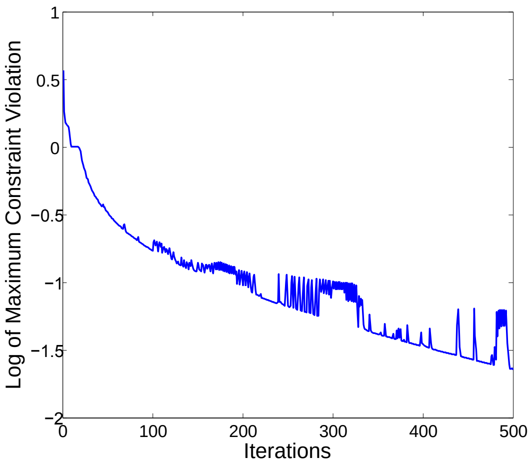
where $$\phi^0(\mathbf{z}^0, \boldsymbol{\lambda}) = \sum_{i=1}^{N} \rho \|\mathbf{R}_i(\mathbf{z}_i^0 - \mathbf{z}_i^*)\|^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}^0 + \rho(1-\tau)\mathbf{r}(\mathbf{z}^0) - \boldsymbol{\lambda}\|^2$$

(b) $$\|\mathbf{r}(\tilde{\mathbf{z}}^k)\| \leq \boxed{\frac{1}{2k\tau}} \left[\sum_{i=1}^{N} \rho \|\mathbf{R}_i(\mathbf{z}_i^0 - \mathbf{z}_i^*)\|^2 + \frac{2}{\rho}\left(\|\boldsymbol{\lambda}^0 + \rho(1-\tau)\mathbf{r}(\mathbf{z}^0) - \boldsymbol{\lambda}^*\|^2 + 1\right)\right]$$

Duke UNIVERSITY

Numerical Experiments





Dual Decomposition

Promising for real-time implementation

Duke UNIVERSITY

Outline

Accelerated Distributed Augmented Lagrangians (ADAL) method
for optimal wireless networking

Accelerated Distributed Augmented Lagrangians (ADAL) method under noise
for optimal wireless networking

Random Approximate Projections (RAP) method with inexact data
for distributed state estimation

Network Optimization under Noise

Noise corruption/Inexact solution of the local optimization steps due to:

i) An exact expression for the objective function is not available (only approximations)
ii) The objective function is updated online via measurements
iii) Local optimization calculations need to terminate at inexact/approximate solutions to save time/resources.

Noise corrupted message exchanges between nodes due to:

i) Inter-node communications suffering from disturbances and/or delays
ii) Nodes can only exchange quantized information.

The noise is modeled as sequences of random variables that are added to the various steps of the iterative algorithm. The convergence of the distributed algorithm is now proved in a stochastic sense (with probability 1).

Duke UNIVERSITY

Deterministic vs Noisy Network Optimization

Where the noise corruption terms appear compared to the deterministic case

Noise in the communicated dual variables

Step 1:

$$\min_{\mathbf{z}_i \in \mathcal{Z}_i} \ \Lambda_i(\mathbf{z}_i, \mathbf{z}^k, \boldsymbol{\lambda}) = f_i(\mathbf{z}_i) + \boldsymbol{\lambda}^T \mathbf{R}_i \mathbf{z}_i + \frac{\rho}{2} \left\| \mathbf{R}_i \mathbf{z}_i + \sum_{j \neq i} \mathbf{R}_j \mathbf{z}_j^k - \mathbf{r}_{\min} \right\|^2$$

Noise in the objective function

Noise in the communicated primal variables

Step 2:

$$\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \tau(\hat{\mathbf{z}}_i^k - \mathbf{z}_i^k)$$

(Trivial local computation = no noise)

Noise in the communicated primal Variables for the dual updates

Step 3:

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho\tau \left(\sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i^{k+1} - \mathbf{r}_{\min} \right)$$

The Stochastic ADAL Algorithm

Step 0: Set k=1 and define initial Lagrange multipliers $\boldsymbol{\lambda}^1$ and initial primal variables $\mathbf{z}^1$

Step 1: For fixed Lagrange multipliers $\boldsymbol{\lambda}^k$, determine $\hat{\mathbf{z}}_i^k$ for every i as the solution of

$$\boxed{\min_{\mathbf{z}_i} \quad \Lambda_i(\mathbf{z}_i, \tilde{\mathbf{z}}_i^k, \tilde{\boldsymbol{\lambda}}_i^k, \xi_i^k) \quad \text{such that} \quad \mathbf{z}_i \in \mathcal{Z}_i}$$

Step 2: Set for every i : $\boxed{\mathbf{z}_i^{k+1} = \mathbf{z}_i^k + \tau_k(\hat{\mathbf{z}}_i^k - \mathbf{z}_i^k)}$ $\boxed{\mathbf{y}_i^{k+1} = \mathbf{z}_i^k + \frac{1}{q}(\hat{\mathbf{z}}_i^k - \mathbf{z}_i^k)}$

Step 3: If the constraints $\sum_{i=1}^{J} \mathbf{R}_i \mathbf{z}_i^{k+1} = \mathbf{r}_{\min}$ are satisfied and $\mathbf{R}_i \hat{\mathbf{z}}_i^k = \mathbf{R}_i \mathbf{z}_i^k$,

then stop (optimal solution found). Otherwise, set:

$$\boxed{\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho\tau_k\left(\sum_{i=1}^{J} \mathbf{R}_i \tilde{\mathbf{y}}_i^{k+1} - \mathbf{r}_{\min}\right)}$$

Noise terms

Increase k by one and return to Step 1.

Convergence

Assumptions (Additional to those of ADAL)

i. Decreasing stepsize (square summable, but not summable)
ii. The noise terms have zero mean, bounded variance, and decrease appropriately as iterations grow

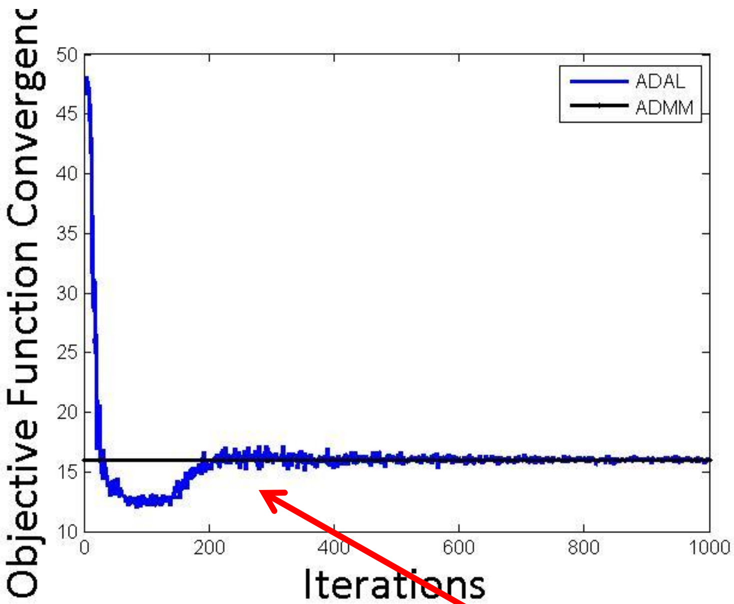
Theorem: The sequence

$$\phi(\mathbf{z}^k, \boldsymbol{\lambda}^k) = \sum_{i=1}^{N} \rho \|\mathbf{R}_i(\mathbf{z}_i^k - \mathbf{z}_i^*)\|^2 + \frac{1}{\rho}\|\boldsymbol{\lambda}^k + \rho(1-\tau)\mathbf{r}(\mathbf{z}^k) - \boldsymbol{\lambda}^*\|^2$$

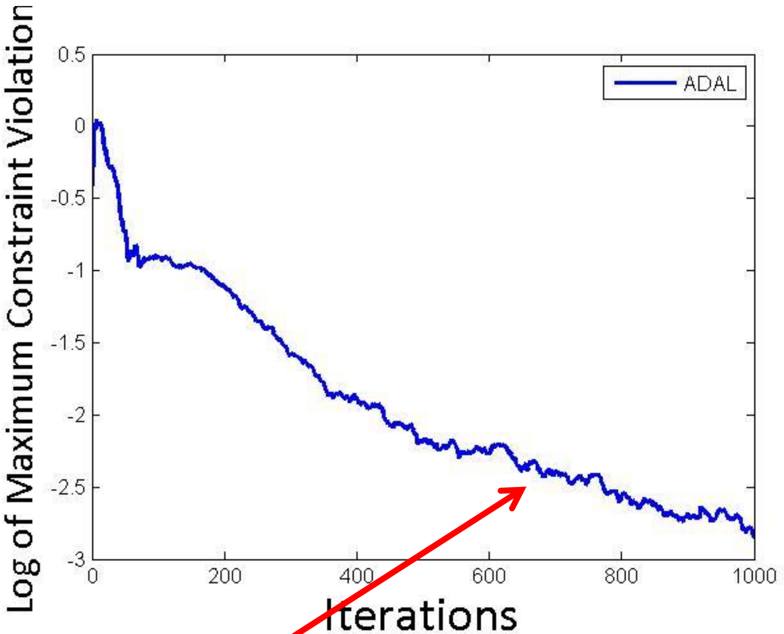
generated by SADAL converges almost surely to zero. Moreover, the residuals $\mathbf{r}(\mathbf{z}^k)$ and the terms $\mathbf{R}_i\hat{\mathbf{z}}_i^k - \mathbf{R}_i\mathbf{z}_i^k$ converge to zero almost surely. This further implies that the SADAL method generates sequences of primal $\{\mathbf{z}^k\}$ and dual variables $\boldsymbol{\lambda}^k$ that converge to their respective optimal sets almost surely.

Numerical Experiments

Objective function convergence



Constraint violation convergence



Oscillatory behavior due to
the presence of noise

Outline

Accelerated Distributed Augmented Lagrangians (ADAL) method
for optimal wireless networking

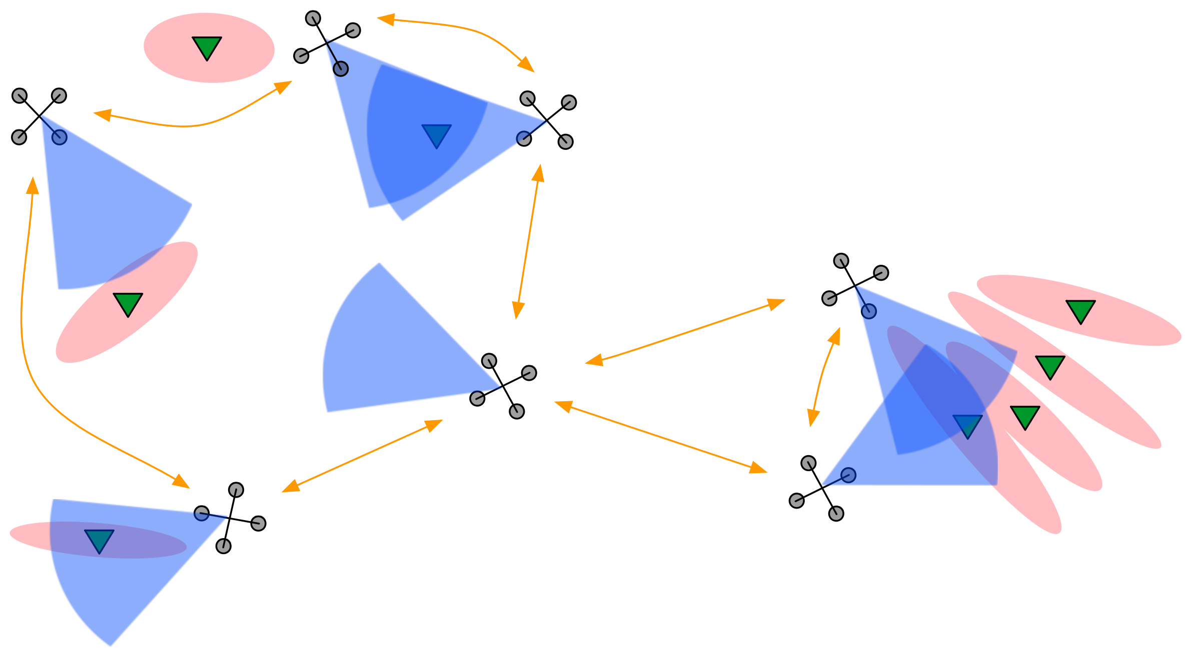
Accelerated Distributed Augmented Lagrangians (ADAL) method under noise
for optimal wireless networking

Random Approximate Projections (RAP) method with inexact data
for distributed state estimation

Duke
UNIVERSITY

Distributed State Estimation

Control a decentralized robotic sensor network to estimate large collections of hidden states with user-specified worst case error.



- Every state can be observed by multiple robots at each time
- Every robot can observe multiple states at each time

Observation Model

Stationary hidden vectors: $\{\mathbf{x}_i \in \mathbb{R}^p\}_{i \in \mathcal{I}}$

Noisy observations form sensors located at $\{\mathbf{r}_s(t) \in \mathbb{R}^q\}_{s \in \mathcal{S}}$ given by:

$$\mathbf{y}_{i,s}(t) = \mathbf{x}_i + \zeta_{i,s}(t) \quad \text{with} \quad \zeta_{i,s}(t) \sim N(\mathbf{0}, \underbrace{Q(\mathbf{r}_s(t), \mathbf{x}_i)}_{\text{information matrix}})$$

Instantaneous observations: $\mathcal{O}(t) = \{(\mathbf{y}_{i,s}(t), Q(\mathbf{r}_s(t), \mathbf{x}_i))\}_{i \in \mathcal{I}, s \in \mathcal{S}}$

Filtered data at time t: $\mathcal{D}(t) = \{\hat{\mathbf{x}}(t), S(t)\}$

where $\hat{\mathbf{x}}(t) \in \mathbb{R}^{pm}$ is the state estimate and $S(t) \in (\mathrm{Sym}_+(p, \mathbb{R}))^m$ is the filtered information matrix

Duke UNIVERSITY

Minimizing Worst-Case Error

- S(t) defines an ellipsoid, related to confidence regions
- Worst case error is the length of the semi-principal axis of the ellipsoid, given by the largest eigenvalue of $S^{-1}(t)$, equivalently, the smallest eigenvalue of S(t)
- Uncertainty thresholds τ_i

$S_i(t+1)$ instead of $\mathbf{x}_i$

$$\{\boldsymbol{\gamma}(t+1), \mathbf{u}(t-1)\} = \operatorname*{argmax}_{(\boldsymbol{\gamma},\mathbf{u})\in\Gamma\times\mathcal{U}} \sum_{i\in\mathcal{I}} \gamma_i$$

$$\text{s.t.} \quad \gamma_i I \preceq \boxed{\underbrace{S_i(t)}_{\text{from } \mathcal{D}(t)} + \underbrace{\sum_{s\in\mathcal{S}} Q(\overbrace{\mathbf{r}_s(t-1)+\mathbf{u}_s}^{\mathbf{r}_s(t)}, \boxed{\overbrace{\hat{\mathbf{x}}_i(t)}^{\text{from } \mathcal{D}(t)}})}_{\text{from } \mathcal{O}(t)}}$$

where $\Gamma = \prod_{i\in\mathcal{I}} [0, \tau_i]$ and $\mathcal{U} \subset B(0,\delta),\ \delta > 0$

Duke UNIVERSITY

Problem Reformulation

Define the state variables $\mathbf{z}(t-1) = \left[\boldsymbol{\gamma}(\boldsymbol{t}+\mathbf{1}), \mathbf{u}(t-1)\right]$

Define local copies $\mathbf{z}_s$ of of the state $\mathbf{z}$

Define local objective functions $f_s(\mathbf{z}_s) = -\sum_{i \in \mathcal{I}} \gamma_{i,s}$

Define by $h(\mathbf{z}; \mathcal{D}, i)$ the linearization of the constraints around $\mathcal{D}(t-1)$

Distributed Optimization with LMI Constraints

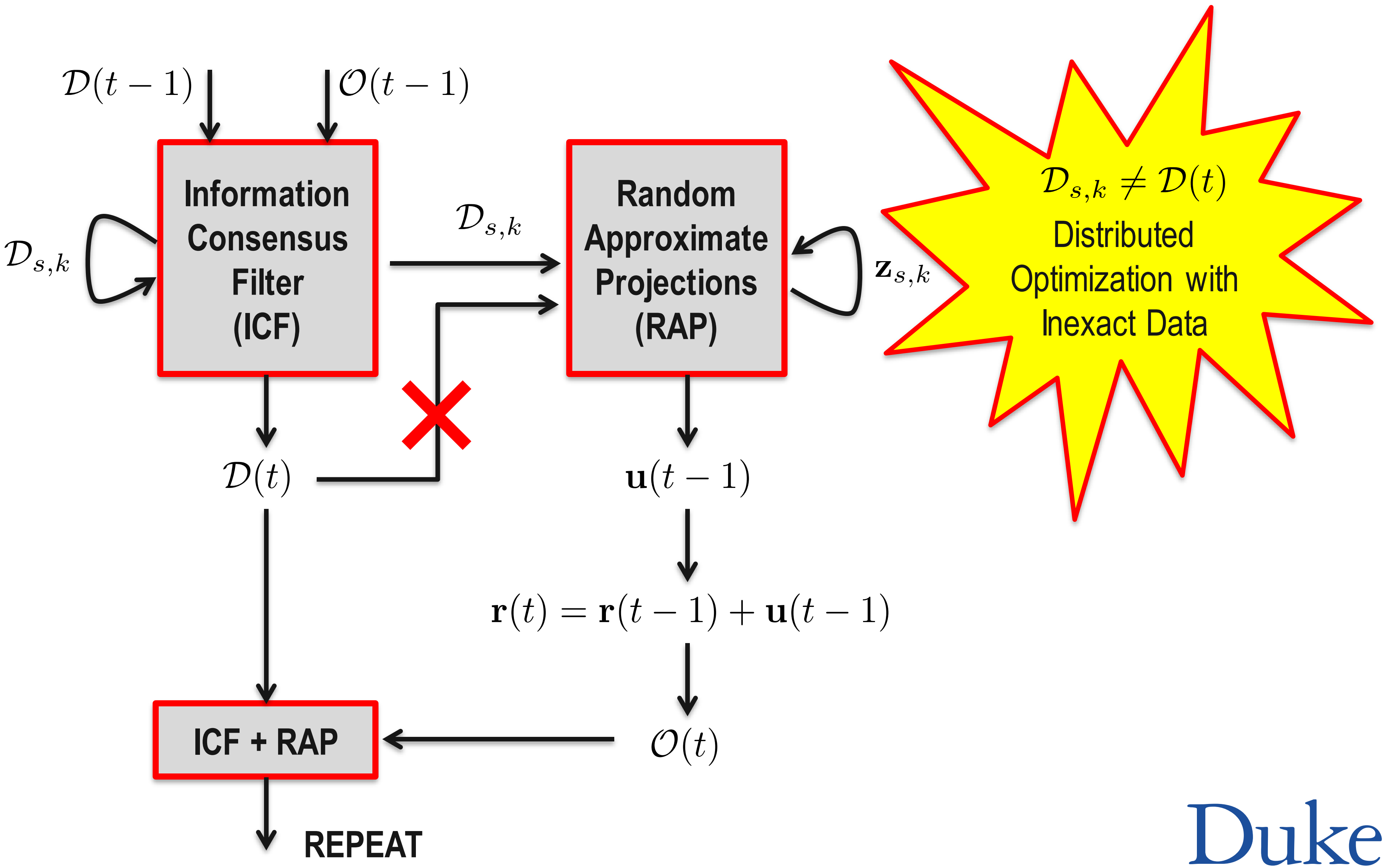
$$
\begin{aligned}
\mathbf{z}(t-1) = \underset{\mathbf{z}_s \in \Gamma \times \mathcal{U}}{\operatorname{argmin}} \quad & \textstyle\sum_{s \in \mathcal{S}} f_s(\mathbf{z}_s) \\
\text{s.t.} \quad & h(\mathbf{z}_s; \mathcal{D}(t), i) \preceq 0, \ \ \forall\, i \in \mathcal{I}, \ \ \forall\, s \in \mathcal{S}
\end{aligned}
$$

Challenges:

- The global parameters $\mathcal{D}(t)$ are unknown to the sensors.
- Agreement on the local state variables $\mathbf{z}_s$

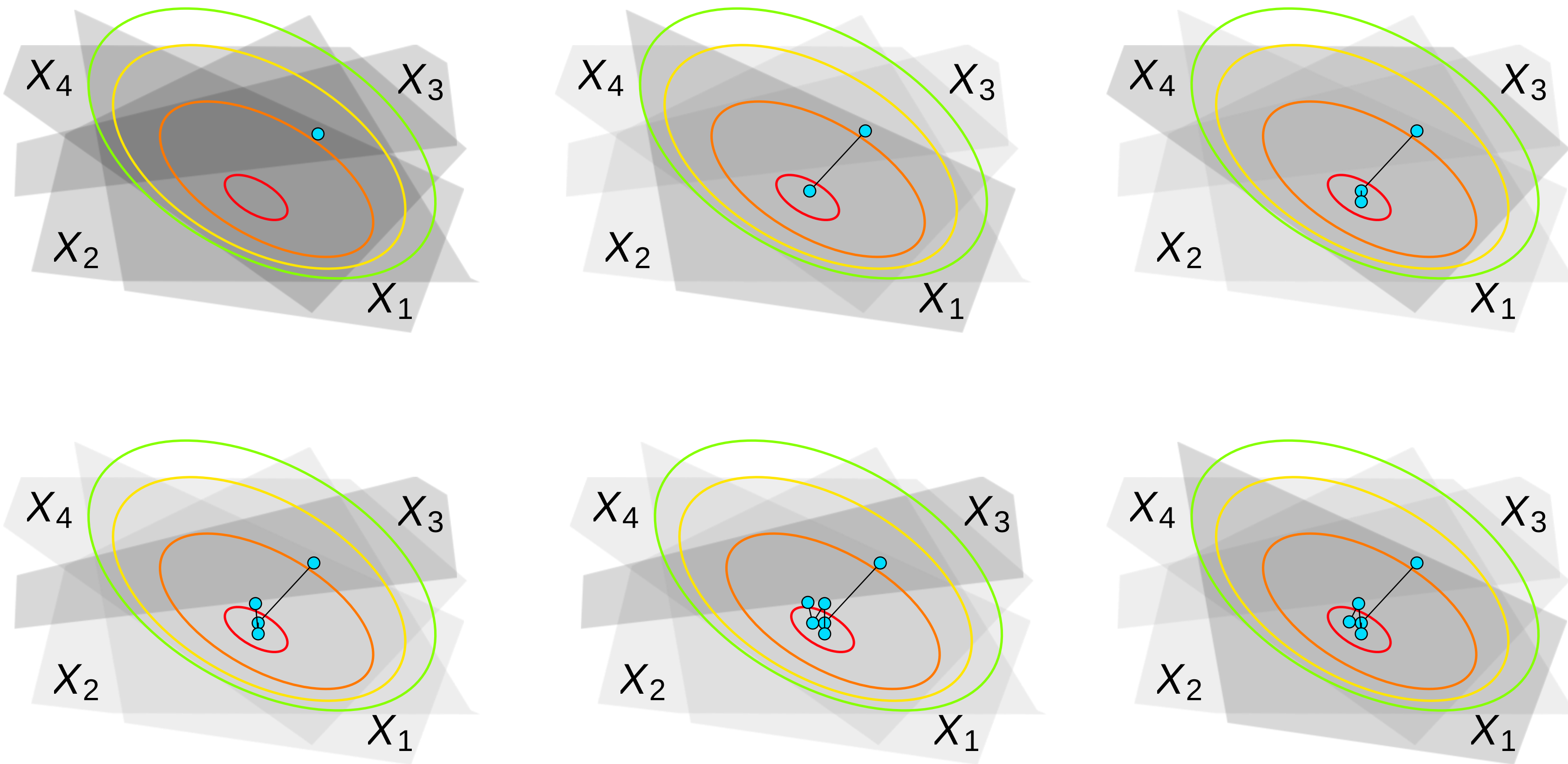
Consensus

Distributed Estimation and Control



Random Projections

Divide the complicated problem into simpler ones



Duke UNIVERSITY

Approximate Projections

Constraint sets $\mathcal{X}_i = \{\mathbf{z} \in \Gamma \times \mathcal{U} \mid h(\mathbf{z}; \mathcal{D}, i) \preceq 0\}$

Exact projection on LMI constraints is computationally expensive.

Define $h_+ = \|\Pi_+ h\|_F$

Define approximate projection onto $\mathcal{X}_i$ by $\mathbf{z} - \beta\, h'_+(\mathbf{z}; \mathcal{D}, i)$

Polyak step size

Projection onto the positive Semidefinite Cone $\Pi_+ X = E(\Lambda)_+ E^T$

E orthogonal matrix of eigenvectors
Λ diagonal matrix of eigenvalues
$(\cdot)_+$ element-wise maximum operator

The RAP Algorithm

Consensus $\mathbf{p}_{s,k} = \sum_{j \in \mathcal{N}_{s,k}} [W_k]_{s,j} \mathbf{z}_{s,k-1}$ W_k row stochastic

Minimization $\mathbf{v}_{s,k} = \Pi_{\mathcal{X}_0} \left(\mathbf{p}_{s,k} - \alpha_k f_s'(\mathbf{p}_{s,k}) \right)$ α_k square summable, non-summable

Polyak step size $\beta_{s,k} = \dfrac{h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k})}{\| h_+'(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \|^2}$ $\mathcal{D}_{s,k}$ from ICF

where $h_+'(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) = d\mathbf{1} > \mathbf{0}$ if $h_+(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) = 0$

Approximate projection $\mathbf{z}_{s,k} = \Pi_{\mathcal{X}_0} \left(\mathbf{v}_{s,k} - \beta_{s,k} h_+'(\mathbf{v}_{s,k}; \mathcal{D}_{s,k}, \omega_{s,k}) \right)$

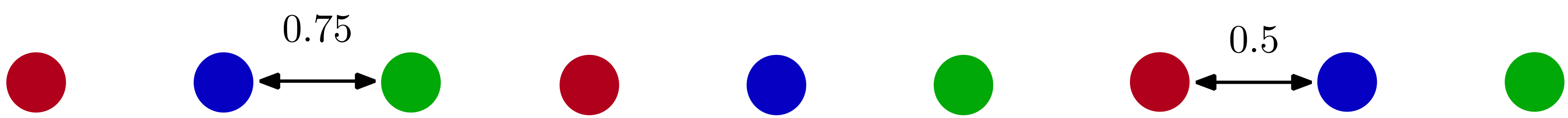
Duke UNIVERSITY

Assumptions

- Information: The information function Q cannot be infinite or change infinitely quickly. Relatively few critical points
- Optimization: Convexity, metric regularity
- RAP: Constraints selected with nonzero probability
- Network: Can have link failures. Require only B-connectivity.

$B = 2, \quad \eta = 0.24$

$k-1$ $\qquad\qquad$ k $\qquad\qquad$ $k+1$



$$\begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0.75 & 0 & 0.25 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Preliminary Results

For a.e. bounded sequence $\mathbf{z}_{s,k}$, the following two sequences are absolutely summable:

Constraint Violation Errors

$$\left\{h_+\left(\mathbf{z}_{s,k};\mathcal{D}_{s,k},\omega_{s,k}\right)-h_+\left(\mathbf{z}_{s,k};\mathcal{D}(t),\omega_{s,k}\right)\right\}_{k\in\mathbb{N}}$$

Constraint Violation Gradient Errors

$$\left\{h'_+\left(\mathbf{z}_{s,k};\mathcal{D}_{s,k},\omega_{s,k}\right)-h'_+\left(\mathbf{z}_{s,k};\mathcal{D}(t),\omega_{s,k}\right)\right\}_{k\in\mathbb{N}}$$

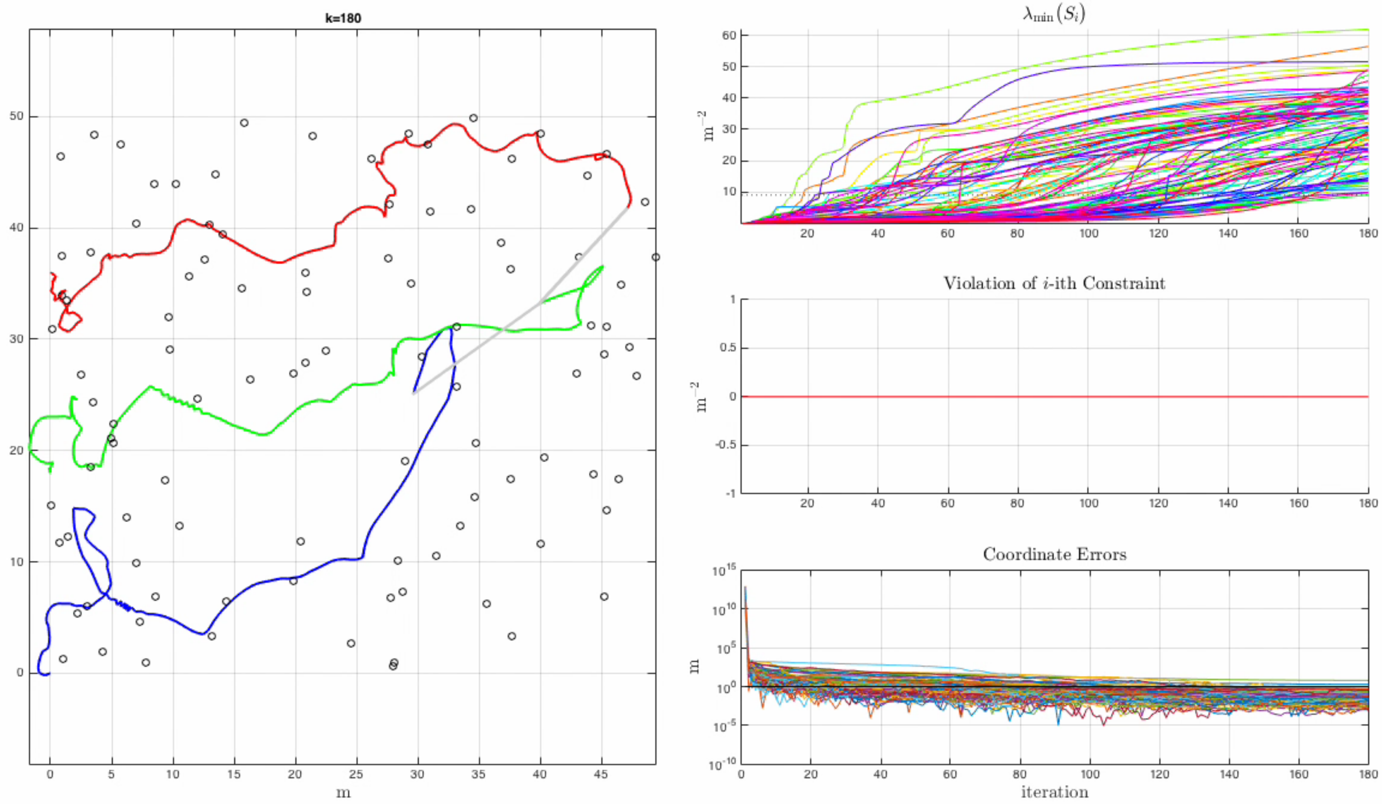
Duke UNIVERSITY

Main Results

Theorem: Let all assumptions be satisfied. Then,

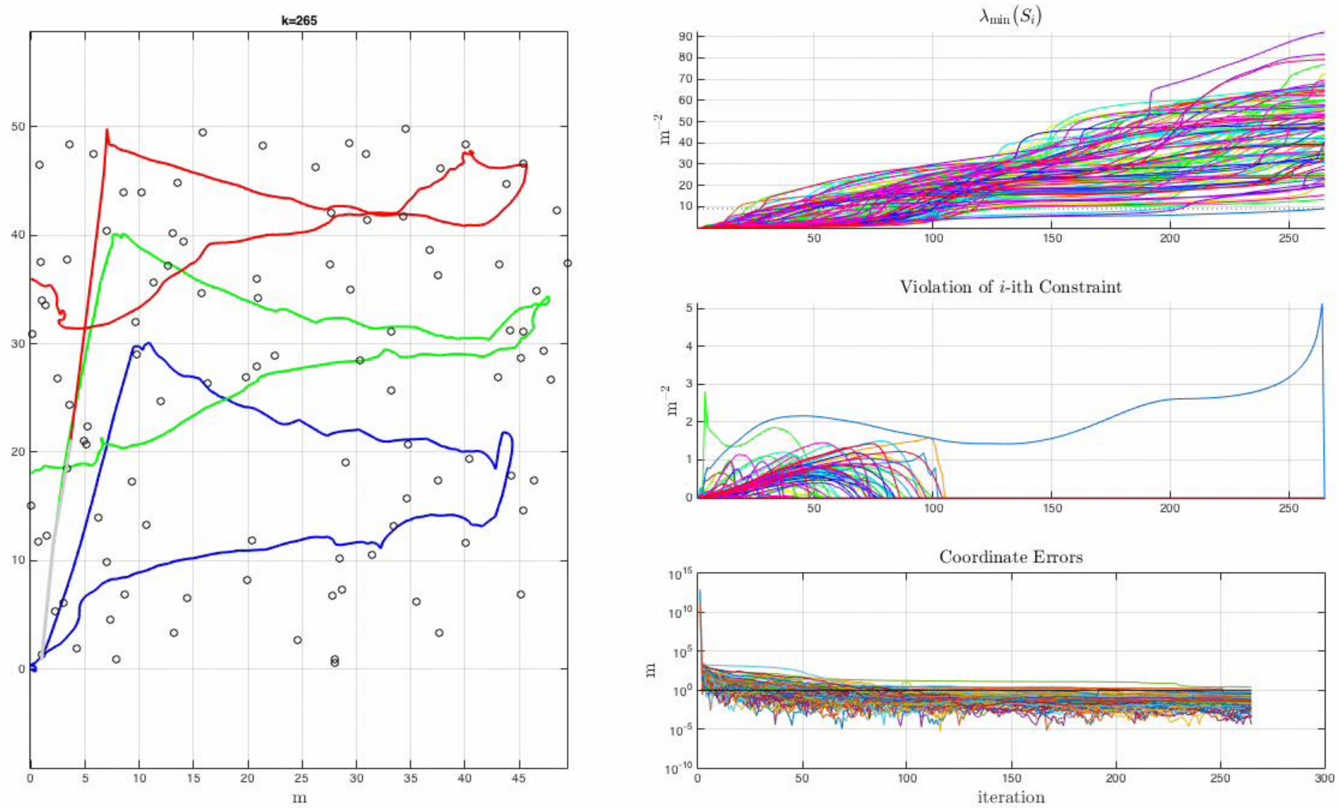
$$\lim_{k \to \infty} \mathbf{z}_{s,k} = \mathbf{z}^*, \quad \forall s \in \mathcal{S} \quad a.s.$$

Simulation Experiments



Minimization of worst-case estimation uncertainty

Simulation Experiments



Minimization of the trace of the estimation uncertainty

Summary

Accelerated Distributed Augmented Lagrangians (ADAL) method
for optimal wireless networking

Fast convergence
Noisy communication

Accelerated Distributed Augmented Lagrangians (ADAL) method under noise
for optimal wireless networking

Random Approximate Projections (RAP) method with inexact data
for distributed state estimation

Light-weight
Distributed LMI constraints
Mobile networks

Acknowledgements

Research Group

Luke Calkins
Yan Zhang
Yiannis Kantaros
Reza Khodayi-mehr
Xusheng Luo

Alumni

Wann-Jiun Ma
Meng Guo
Charlie Freundlich
Soomin Lee
Nikolaos Chatzipangiotis

Thank You

Accelerated Distributed Augmented Lagrangians (ADAL) method

- N. Chatzipanagiotis, D. Dentcheva, and M. M. Zavlanos, “An Augmented Lagrangian Method for Distributed Optimization,” Mathematical Programming , vol. 152, no. 1-2, pp. 405-434, Aug. 2015.
- N. Chatzipanagiotis, S. Lee, and M. M. Zavlanos, “Complexity Certification of a Distributed Augmented Lagrangian Method,” IEEE Transactions on Automatic Control, accepted.

Accelerated Distributed Augmented Lagrangians (ADAL) method under noise

- N. Chatzipanagiotis and M. M. Zavlanos, “A Distributed Algorithm for Convex Constrained Optimization under Noise,” IEEE Transactions on Automatic Control, vol. 61, no. 9, pp. 2496-2511, Sep. 2016.

Random Approximate Projections (RAP) method with inexact data

- C. Freundlich, S. Lee, and M. M. Zavlanos, “Distributed Active State Estimation with User-Specified Accuracy,” IEEE Transactions on Automatic Control, in press.
- S. Lee and M. M. Zavlanos, “Approximate Projections for Decentralized Optimization with SDP Constraints,” IEEE Transactions on Automatic Control, accepted.

Duke UNIVERSITY