Decentralized Power System State Estimation via Non-convex Multi-agent Optimization

Aritra Konar,

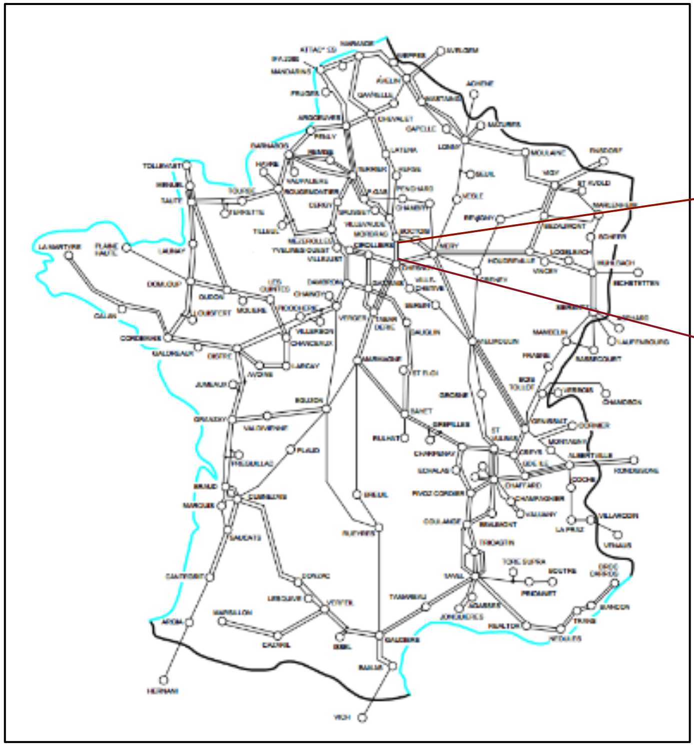
Ahmed Zamzam & Nikos Sidiropoulos

Dept. of ECE, University of Virginia

Dept. of ECE, University of Minnesota

Power System State Estimation

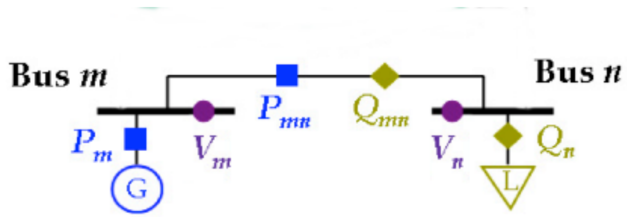
Power Transmission Network:



Undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$

Set of buses (nodes) $\mathcal{N} := \{1, \cdots, N\}$

Set of transmission lines (edges) $\mathcal{E}$



Available SCADA measurements $\mathcal{L}_n$ at bus n

- Active and reactive power injections: $\{P_n, Q_n\}$
- Active and reactive power flows: $\{P_{nm}, Q_{nm}\}$
- Voltage magnitude: $|V_n|$

- Problem Statement:
  - Given grid parameters and SCADA meter readings, estimate complex voltages at all buses

Power System State Estimation

- Weighted LS formulation [Schweppe *et al.* 1970]

$$\min_{\mathbf{v}\in\mathcal{K}} \left\{ F(\mathbf{v}) := \sum_{n=1}^{N} f_n(\mathbf{v}) := \sum_{n=1}^{N} \sum_{l\in\mathcal{L}_n} \frac{(\mathbf{v}^T \mathbf{H}_l \mathbf{v} - z_l)^2}{\sigma_l^2} \right\}$$

  - Set $\mathcal{K} \subset \mathbb{R}^{2N}$ represents constraints on $\mathbf{v}$ known *apriori*; simple, convex, and compact
  - Minimize non-convex cost function subject to convex constraints
  - NP-hard [Lehmann *et al.* 2016]

- Objective:
  - Compute high-quality approximate solutions in polynomial-time via simple decentralized algorithms
  - Adopt multi-agent optimization approach

Assumptions

- Network:
  - Underlying voltage profile is fixed
  - Graph of the network $\mathcal{G}$ is connected
  - Each bus $n \in \mathcal{N}$ only has knowledge of $f_n(.)$ and $\mathcal{K}$
  - Each bus is only aware of its interconnections to its immediate neighbors

- Formulation:
  - The set of minima of $F(.)$ over $\mathcal{K}$ is non-empty and compact
  - Each $f_n(.)$ is continuously differentiable on an open set containing $\mathcal{K}$
  - $\nabla f_n(.)$ is locally Lipschitz continuous on $\mathcal{K}$
  - $F(.)$ is locally Lipschitz continuous on $\mathcal{K}$

Prior Art

- Local state methods:
  - Lagrangian Relaxation [Caro-Conejo-Minguez 2011]
  - Alternating Direction Method of Multipliers [Kekatos-Giannakis 2013]
  - Semidefinite Relaxation [Zhu-Giannakis 2014]
  - Gauss-Newton [Minot-Lu-Li 2016]

- Global state methods:
  - Belief Propagation [Hu-Kuh-Yang-Kavcic 2011]
  - Network Gossiping [Xie-Choi-Kar-Poor 2012]

- Lack theoretical convergence guarantees

Prior Art (contd….)

- Global state methods:
  - Gossip-based Gauss-Newton [Li-Scaglione 2013]
    - Convergence to stationary points
    - Communication model different from grid topology
    - High per-iteration communication overhead
  - This talk: In-network Non-convex Optimization (NEXT) [Lorenzo-Scutari 2016]
    - Communication model defined by grid topology
    - Gradient-consensus
    - Low per-iteration complexity
    - Reduced per-iteration communication overhead
    - Convergence to stationary points
  - Decentralized Gradient Descent (DGD) [Bianchi-Jakubowicz 2013]

In-network Non-convex Optimization (NEXT)

- Algorithm:
  - Local successive convex approximation + dynamic consensus
    - Each bus constructs local convex surrogate of global cost function
    - In this talk: use gradient-based quadratic surrogate
    - Consensus for enforcing agreement and propagating information across network
  - At each iteration $k \in \mathbb{N}$, given local estimates $\{\mathbf{v}_n[k]\}_{n=1}^N$
    - Step 1: Perform local GD updates of the form

$$\tilde{\mathbf{v}}_n[k] = \mathbf{\Pi}_{\mathcal{K}}\left(\mathbf{v}_n[k] - \frac{1}{\tau}(\nabla f_n(\mathbf{v}_n[k]) + \tilde{\boldsymbol{\pi}}_n[k])\right)$$

$$\mathbf{z}_n[k] = \mathbf{v}_n[k] + \beta[k](\tilde{\mathbf{v}}_n[k] - \mathbf{v}_n[k]), \forall\, n \in \mathcal{N}$$

where $\tau > 0$, $\tilde{\boldsymbol{\pi}}_n[k]$ approximates $\boldsymbol{\pi}_n[k] := \sum_{m \neq n} \nabla f_m(\mathbf{v}_n[k])$

- Requires local auxiliary variable $\mathbf{y}_n[k]$ for tracking

$$\overline{\nabla F}(\mathbf{v}_n[k]) := (1/N)\sum_{n=1}^N \nabla f_n(\mathbf{v}_n[k])$$

In-network Non-convex Optimization (NEXT)

- Algorithm: (contd.)
  - Step 2: Perform local exchanges followed by updates

$$\mathbf{v}_n[k+1] = \sum_{m \in \bar{\mathcal{N}}_n} W_{nm} \mathbf{z}_m[k]$$

$$\mathbf{y}_n[k+1] = \sum_{m \in \bar{\mathcal{N}}_n} W_{nm} \mathbf{y}_m[k] + (\nabla f_n(\mathbf{v}_n[k+1]) - \nabla f_n(\mathbf{v}_n[k]))$$

$$\tilde{\boldsymbol{\pi}}_n[k+1] = N\mathbf{y}_n[k+1] - \nabla f_n(\mathbf{v}_n[k+1]), \ \forall\, n \in \mathcal{N}, \forall\, k \in \mathbb{N}$$

where $\overline{\mathcal{N}}_n := \{m | (n,m) \in \mathcal{E}\} \cup \{n\}, \mathbf{W} \in \mathbb{R}^{N\times N}$ is doubly stochastic

- Features:
  - If $\beta[k] \in (0,1], \sum_{k\in\mathbb{N}} \beta[k] = \infty, \ \sum_{k\in\mathbb{N}} \beta^2[k] < \infty$, then
    - We have

$$\|\overline{\nabla F}(\mathbf{v}_n[k]) - \mathbf{y}_n[k+1]\| \underset{k\to\infty}{\longrightarrow} 0, \forall\, n \in \mathcal{N}$$

    - All sequences $\{\mathbf{v}_n[k]\}_{k\in\mathbb{N}}$ asymptotically attain consensus
    - Sequence $\{\mathbf{v}_{\text{avg}}[k]\}_{k\in\mathbb{N}}$ converges to set of stationary points

In-network Non-convex Optimization (NEXT)

- Features:
  - Low per-iteration complexity at each bus
    - Computing local gradient requires sparse-matrix vector multiplications
    - Projection operation is simple to compute
  - Communication overhead per-iteration: $O(N)$
    - Average degree of each bus is small
    - Higher than local state methods, but comes with the benefit of providing each bus with global information
  - Sparsity of power network facilitates synchronous implementation

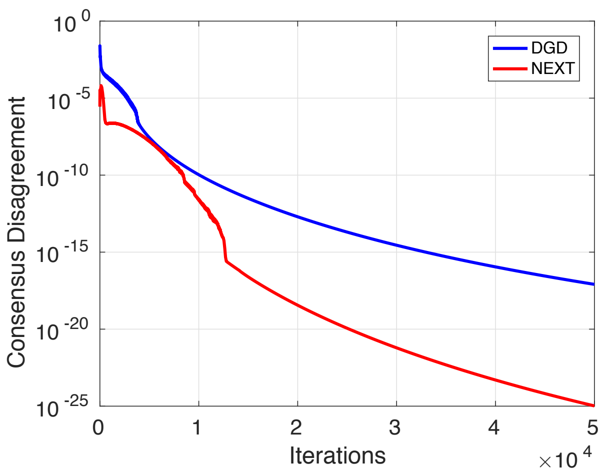
Experiments

- Setup
  - Voltage profile with magnitude $\sim \mathcal{U}[0.9, 1.1]$ and phase $\sim \mathcal{U}[-0.1\pi, 0.1\pi]$
  - Generate SCADA measurements using MATPOWER
  - Phase of reference bus set to zero
  - Initialize algorithms from flat start $\mathbf{v} = [\mathbf{1}^T; \mathbf{0}^T]^T$
  - Feasible set $\mathcal{K} = \{\mathbf{v} \in \mathbb{R}^{2N} \,|\, v_n^2 + v_{n+N}^2 \leq (1.1)^2, \forall\, n \in \mathcal{N}\}$
  - Generate $\mathbf{W}$ using Metropolis-Hastings Rule
  - Step-size rule: $\beta[k+1] = \beta[k](1 - \gamma\beta[k]), \forall\, k \in \mathbb{N}, \beta[0] = (0, 1], \gamma \in (0, 1)$
  - Performance Metrics
    - Consensus disagreement: $C[k] := \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{v}_n[k] - \mathbf{v}_{\text{avg}}[k]\|^2$
    - NMSE: $E[k] := \frac{\|\mathbf{v} - \mathbf{v}_{\text{avg}}[k]\|_2}{\|\mathbf{v}\|_2}$
    - Progress towards stationarity:

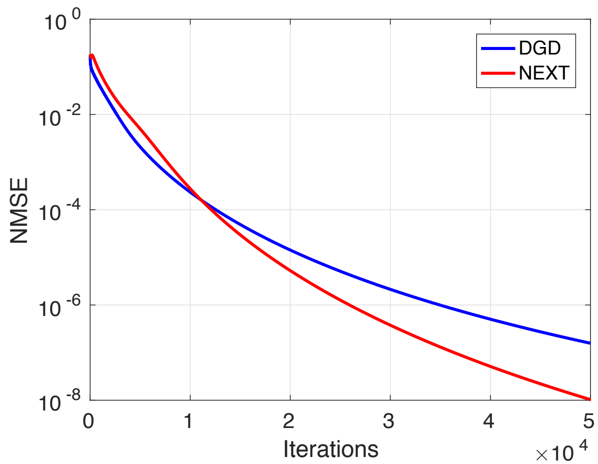
$$S[k] := \|\mathbf{v}_{\text{avg}}[k] - \mathbf{\Pi}_{\mathcal{K}}(\mathbf{v}_{\text{avg}}[k] - \nabla F(\mathbf{v}_{\text{avg}}[k]))\|_\infty, \forall\, k \in \mathbb{N}$$

Illustrative Example

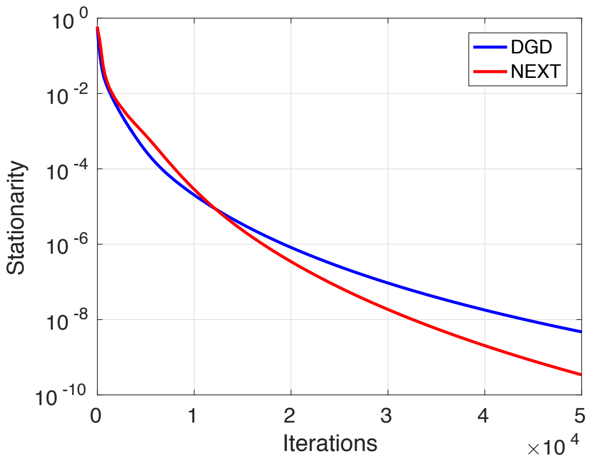
IEEE-14 bus network, noiseless case



Consensus Disagreement vs iterations



NMSE vs iterations



Progress towards stationarity vs iterations

NEXT demonstrates better performance across all metrics

Preliminary Experiments

- Noisy case:
  - Gaussian noise with variances 10 dBm and 13 dBm added to voltage and power measurements respectively

- IEEE-57 bus network:

	$C[k]$	$E[k]$
DGD	8.71×10^{-7}	0.0346
NEXT	6.62×10^{-8}	0.0020

- IEEE-118 bus network:

	$C[k]$	$E[k]$
DGD	4.86×10^{-6}	0.1247
NEXT	9.16×10^{-6}	0.0352

Conclusions and Future Work

- Multi-agent optimization for state estimation
  - Synchronous gradient consensus algorithms
  - NEXT outperforms DGD on tests

- Future work
  - More sophisticated variants of NEXT
  - Hybrid state estimation
  - Robust estimation approaches
  - Tests on larger networks

Thank you!

Decentralized Gradient Descent (DGD)

- **Algorithm:** [Bianchi-Jakubowicz 2013]
  - At each iteration $k \in \mathbb{N}$, given local estimates $\{\bar{\mathbf{v}}_n[k]\}_{n=1}^N$
    - Step 1: Perform local GD updates of the form

$$\bar{\mathbf{z}}_n[k] = \mathbf{\Pi}_{\bar{\mathcal{K}}}(\bar{\mathbf{v}}_n[k] - \alpha[k]\nabla f_n(\bar{\mathbf{v}}_n[k])), \ \forall\, n \in \mathcal{N}$$

    - Step 2: Perform local exchanges followed by weighted average

$$\bar{\mathbf{v}}_n[k+1] = \sum_{m \in \overline{\mathcal{N}}_n} W_{nm} \bar{\mathbf{z}}_m[k], \ \forall\, n \in \mathcal{N}$$

where $\overline{\mathcal{N}}_n := \{m | (n,m) \in \mathcal{E}\} \cup \{n\}, \mathbf{W} \in \mathbb{R}^{N \times N}$ is doubly stochastic

- **Features:**
  - If $\alpha[k] \geq 0, \sum_{k \in \mathbb{N}} \alpha[k] = \infty, \ \sum_{k \in \mathbb{N}} \alpha^2[k] < \infty$, then
    - Asymptotic consensus; i.e., $\|\bar{\mathbf{v}}_n[k] - \bar{\mathbf{v}}_{\text{avg}}[k]\| \underset{k \to \infty}{\longrightarrow} 0, \forall\, n \in \mathcal{N}$
    - Sequence $\{\bar{\mathbf{v}}_{\text{avg}}[k]\}_{k \in \mathbb{N}}$ converges to set of stationary points