Fast Nonconvex SDP Solver for Large-scale Power System State Estimation

Yu Lan[†⋆], Hao Zhu[⋆], and Xiaohong Guan[†]

[†] Institute of Systems Engineering, Xi'an Jiaotong University
[⋆] Department of ECE, The University of Texas at Austin

November 5, 2018

Context

- Power System State Estimation (SE): nonlinear estimation
  - iteratively solved by Gauss-Newton (GN) [Monticelli'00]
  - convergence and numerical stability of GN not guaranteed
- Semidefinite programming (SDP) and convex relaxations
  - near optimality achieved [Zhu et al'14][Mardani et al'16]
  - high-order complexity for generic convex solvers
- Large-scale SDP arises in a wide variety of applications
  - matrix sensing[Jain et al.'13], phase retrieval[Netrapalli et al.'13], quantum state tomography [Kyrillidis et al.'18]
  - simple gradient descent method for a nonconvex reformulation of SDP [Bhojanapalli et al.'16]

- Goal: accelerating SE using fast nonconvex SDP solver

Modeling

- Estimate nodal voltages $\mathbf{v} := [V_1, ..., V_N]^{\mathcal{T}} \in \mathbb{C}^N$ using:
    - $P_n(Q_n)$: the active (reactive) power injection at bus n;
    - $P_{nn'}(Q_{nn'})$: the active (reactive) power flow from bus n to n';
    - $|V_n|$: the voltage magnitude at bus n.
- Nonlinear (quadratic) measurement model:

$$z_\ell = h_\ell(\mathbf{v}) + \epsilon_\ell, \ \forall \ell = 1, \ldots, L \tag{1}$$

- **Weighted Least-Squares (WLS)** error objective:

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v} \in \mathbb{C}^N} \sum_{\ell=1}^{L} w_\ell \left[z_\ell - h_\ell(\mathbf{v})\right]^2 \tag{2}$$

- GN minimizes (2) through iterative linearization

Semidefinite Programming (SDP) Formulation

- Form $\mathbf{V} := \mathbf{v}\mathbf{v}^{\mathcal{H}} \in \mathbb{C}^{N \times N}$ to obtain linear model

$$z_\ell = \mathrm{Tr}(\mathbf{H}_\ell \mathbf{V}) + \epsilon_\ell, \ \forall \ell = 1, \ldots, L. \tag{3}$$

- Rank constraint relaxed for a convex **SDP** formulation

$$\hat{\mathbf{V}} = \arg \min_{\mathbf{V} \in \mathbb{C}^{N \times N}} f(\mathbf{V}) := \sum_{\ell=1}^{L} w_\ell \left[z_\ell - \mathrm{Tr}(\mathbf{H}_\ell \mathbf{V}) \right]^2 \tag{4a}$$

$$\text{s.t. } \mathbf{V} \succeq \mathbf{0}, \ \text{and } \xcancel{rank(\mathbf{V}) = 1} \tag{4b}$$

  - recover vector $\hat{\mathbf{v}}$ from the best rank-one approximation of $\hat{\mathbf{V}}$, followed by GN improvements
  - $\hat{\mathbf{V}}$ typically of very low rank (≤ 2)
  - generic solvers not suitable for real-time implementations

SDP Using Gradient Descent

- Recent SDP approaches advocate nonconvex reformulation using $\mathbf{V} = \mathbf{U}\mathbf{U}^{\mathcal{H}}$ with $\mathbf{U} \in \mathbb{C}^{N\times r}$

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}\in\mathbb{C}^{N\times r}} g(\mathbf{U}) := f(\mathbf{U}\mathbf{U}^{\mathcal{H}}) \qquad (5)$$

  - unconstrained (drop the PSD constraint)
  - low-rank solution ($r \ll n$ with $r = 1$ equivalent to WLS)
  - computational gains (convenient gradient descent updates)

- Factored Gradient Descent (FGD) for the nonconvex SDP

$$\mathbf{U}_{k+1} = \mathbf{U}_k - \eta\nabla g(\mathbf{U}_k)$$

  using the gradient

$$\nabla g(\mathbf{U}) = 2\nabla f(\mathbf{U}\mathbf{U}^{\mathcal{H}})\mathbf{U} = \sum_{\ell=1}^{L} 4w_\ell\left[\mathrm{Tr}(\mathbf{U}^{\mathcal{H}}\mathbf{H}_\ell\mathbf{U}) - z_\ell\right]\mathbf{H}_\ell\mathbf{U}$$

Properties of Function f

- **(C1) M-smooth**
  $$f(\mathbf{V}) \le f(\mathbf{V}') + \langle \nabla f(\mathbf{V}'), \mathbf{V} - \mathbf{V}' \rangle + \frac{M}{2} \|\mathbf{V} - \mathbf{V}'\|_F^2$$
- **(C2) m-strongly convex**
  $$f(\mathbf{V}) \ge f(\mathbf{V}') + \langle \nabla f(\mathbf{V}'), \mathbf{V} - \mathbf{V}' \rangle + \frac{m}{2} \|\mathbf{V} - \mathbf{V}'\|_F^2$$
- For convex f satisfying **(C1)+(C2)**, vanilla GD enjoys linear convergence; Smaller $\kappa = M/m$ for faster convergence rate
- Relaxed condition for SDP objectives
  (C2') (m, r)-restricted strongly convex
  $$f(\mathbf{V}^r) \ge f(\mathbf{V}_o^r) + \langle \nabla f(\mathbf{V}_o^r), \mathbf{V}^r - \mathbf{V}_o^r \rangle + \frac{m}{2} \|\mathbf{V}^r - \mathbf{V}_o^r\|_F^2$$
  for any rank-r matrices $(\mathbf{V}^r, \mathbf{V}^r)$

SDP-SE Analysis

- Compact form of matrix sensing objective using the linear map $\mathcal{H}$

$$f(\mathbf{V}) = \|\mathbf{z} - \mathcal{H}(\mathbf{V})\|_{\mathbf{W}}^2 \tag{6}$$

- **(C1)+(C2')** lead to

$$m \cdot \|\mathbf{V}^r\|_F^2 \leq 2\|\mathcal{H}(\mathbf{V}^r)\|_{\mathbf{W}}^2 \leq M \cdot \|\mathbf{V}^r\|_F^2. \tag{7}$$

- Under the power flow model, we obtain upper/lower bounds for every $\mathbf{V}^1 \in \mathcal{V} := \{\mathbf{V} | \text{rank}(\mathbf{V}) = 1, \underline{V}^2 \leq \mathbf{V}_{nn} \leq \bar{V}^2\}$

$$2\|\mathcal{H}(\mathbf{V}^1)\|_{\mathbf{W}}^2 = \sum_{n \in \mathcal{N}_V} 2w_n^v |V_n|^4 + \sum_{n \in \mathcal{N}_P} 2w_n^p P_n^2 + \sum_{n \in \mathcal{N}_Q} 2w_n^q Q_n^2 + \sum_{(n,n') \in \mathcal{E}_P} 2w_{nn'}^p P_{nn'}^2 + \sum_{(n,n') \in \mathcal{E}_Q} 2w_{nn'}^q Q_{nn'}^2$$

Upper/Lower Bounds

- Let $\bar{V} = \max_n |V_n|$ and $\underline{V} = \min_n |V_n|$.
  - $\underline{V}^4 \leq |V_n|^4 \leq \bar{V}^4$;
  - $P_{nn'}^2 + Q_{nn'}^2 = |S_{nn'}|^2 = |V_n(y_{nn'}(V_n - V_{n'}))^{\mathcal{H}}|^2 \leq 4|y_{nn'}|^2\bar{V}^4$;
  - $P_n^2 + Q_n^2 = |S_n|^2 = |V_n(\sum_\nu y_{n\nu}V_\nu)^{\mathcal{H}}|^2 \leq (\sum_\nu |y_{n\nu}|)^2\bar{V}^4$.
- A good set of weight coefficients:

$$\begin{aligned} w_n^v &= 1/2 \\ w_{nn'}^p = w_{nn'}^q &= 1/(8|y_{nn'}|^2) \\ w_n^p = w_n^q &= 1/[2(\textstyle\sum_\nu |y_{n\nu}|)^2] \end{aligned}$$

- For N_S ($N_{|V|}$) power (voltage) meters

$$m = \frac{N_{|V|}\underline{V}^4}{\|\mathbf{V}\|_F^2}, \; M = \frac{\bar{V}^4}{\|\mathbf{V}\|_F^2}(N_S + N_{|V|})$$

FGD Convergence

- **(as1)** Sufficiently good initialization point
  $\mathrm{Dist}(\mathbf{U}_0, \hat{\mathbf{U}}) \leq \frac{\rho_u}{\kappa}\sigma_r(\hat{\mathbf{U}})$
- **(as2)** Optimal solution approximately rank-r
  $\|\hat{\mathbf{V}} - \hat{\mathbf{V}}^r\|_F \leq \frac{\rho_v}{\kappa^{1.5}}\sigma_r(\hat{\mathbf{V}}^r)$

- **Main results:** [Bhojanapalli et al.'16]
  Under (**as1**)-(**as2**), with $\eta = 1/(16(M\|\mathbf{V}_0\|_2 + \|\nabla f(\mathbf{V}_0)\|_2))$,
  linear convergence achieved by FGD, as

$$\mathrm{Dist}(\mathbf{U}_{k+1}, \hat{\mathbf{U}})^2 \leq \alpha \cdot \mathrm{Dist}(\mathbf{U}_k, \hat{\mathbf{U}})^2 + \beta \cdot \|\hat{\mathbf{V}} - \hat{\mathbf{V}}^r\|_F^2$$

for $0 < \alpha < 1$.

Accelerated Gradient Descent

- Nesterov's acceleration for better convergence rate

$$\mathbf{U}^{+} = \mathbf{U}_k + \left(\frac{k-2}{k+1}\right)(\mathbf{U}_k - \mathbf{U}_{k-1}) \tag{9a}$$

$$\mathbf{U}_{k+1} = \mathbf{U}^{+} - \eta \nabla g(\mathbf{U}^{+}) \tag{9b}$$

  - Require the data from the past two iterations
  - Same computation complexity per iteration as FGD
  - Under resctricted isometry property (RIP) of f, AGD shown to converge linearly for nonconvex SDP formulation [Kyrillidis et al.'18]

AGD and FGD Comparison

- Iterative error with respect to the actual $\mathbf{V}_o$ averaged over 100 Monte-Carlo tests

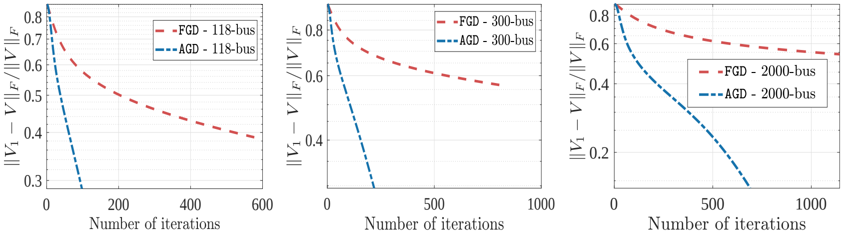

Table 1: Average Run Time of FGD and AGD

	118-bus	300-bus	2000-bus
FGD	0.1584s	1.903s	72.01s
AGD	0.0372s	0.5853s	42.39s

SE Error Performance

- RMSE error criterion: $||\hat{\mathbf{v}} - \mathbf{v}||_2 / ||\mathbf{v}||_2$

Table 2: SE Error and GN Convergence Rate

SE Error	118-bus	300-bus	2000-bus
GN	0.0962 (98%)	0.2604 (79%)	0.4784 (21%)
SDP-GN	0.0100 (100%)	0.0717 (100%)	N/A (N/A)
FGD-GN	0.0100 (100%)	0.0717 (100%)	0.0078 (100%)
AGD-GN	0.0100 (100%)	0.0717 (100%)	0.0078 (100%)

Table 3: Average Run Time of SDP, FGD, and AGD

Time	118-bus	300-bus	2000-bus
SDP	4.887s	50.82s	N/A
FGD	0.1584s	1.903s	72.01s
FGD	0.0372s	0.5853s	42.39s

Concluding Remarks

- Fast SDP-SE solver using recent approaches of local search for nonconvex problems
  - verify the FGD convergence conditions from power flow analysis
  - improve the numerical convergence using AGD
- Ongoing work
  - rigorous analysis of the AGD updates
  - constrained SDP extensions for optimal power flow problem

Acknowledgments: NSF ECCS-1802319 and the Key R&D Project of China 2016YFB0901904.