Improving Power Grid Reliability and Efficiency via Advanced Signal Processing and Machine Learning

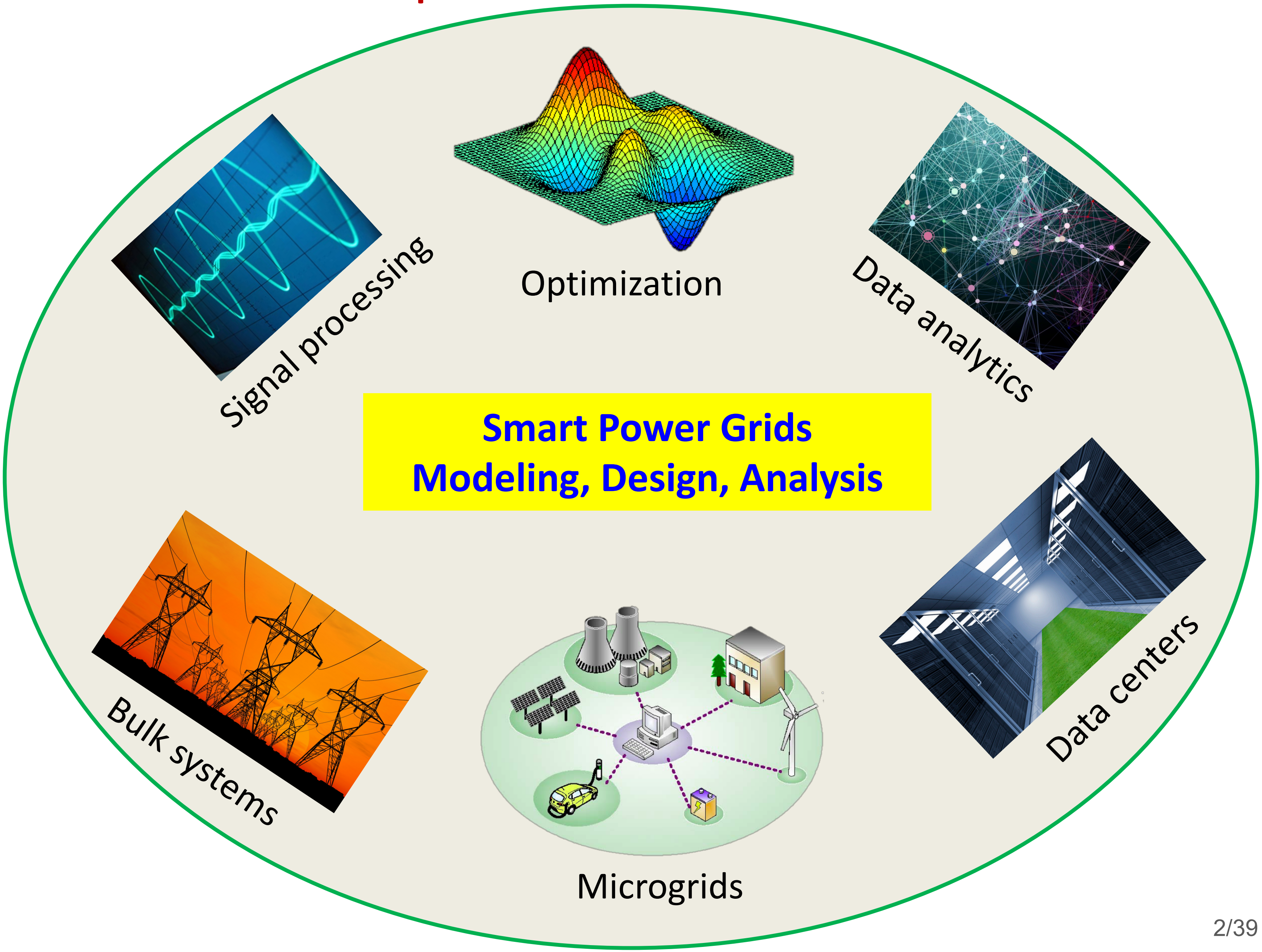
Yu Zhang, Assistant Professor

ECE Department, UC Santa Cruz

ACK: Ramtin Madani, Javad Lavaei, and Vassilis Kekatos

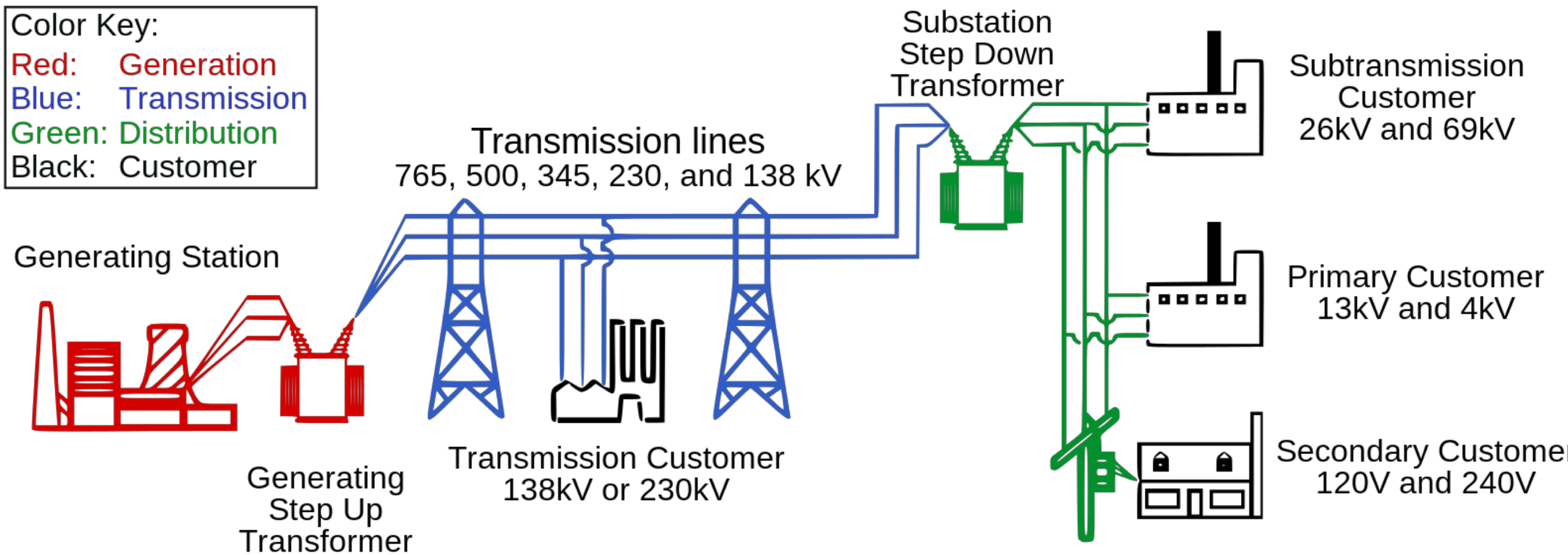
MADDD Seminar, UC Davis
October 29, 2019

Research Landscape



The Electrical Grid Then and Now

"The most significant engineering achievement of the 20th century." [NAE Report'10]

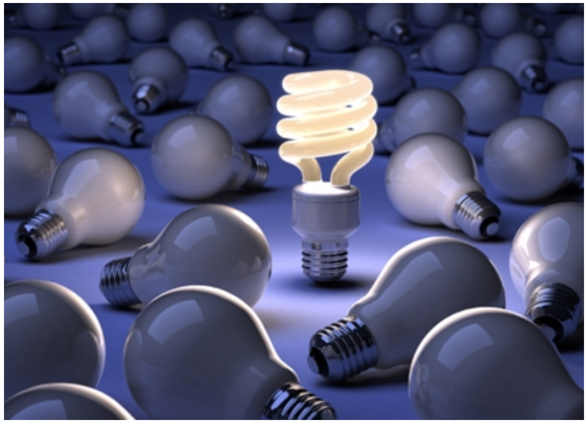


- Several challenges ahead
  - Customer engagement and environmental concerns.
  - 99.97% reliable, but power outages cost $150 billion per year in the US.

Figure courtesy: FERC. "Final Report on the Aug. 14, 2003 Blackout in the US and Canada."

Features of Smart Grids

controllable



efficient

resilient

green/sustainable

self-healing

situational awareness

US Dept. of Energy, "The smart grid: An introduction."

Enabling Technology Advances

renewables

power electronics

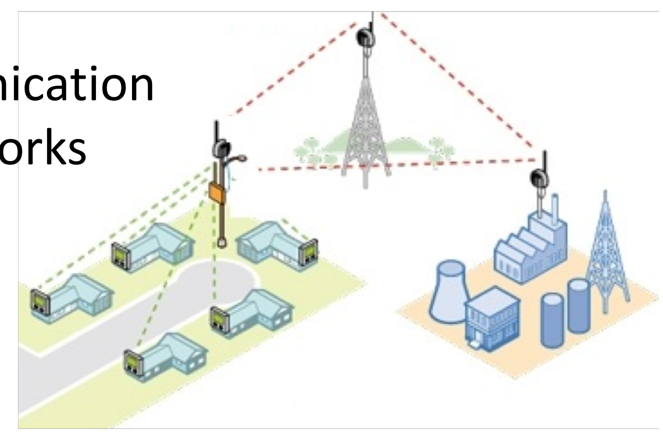
energy storage

Optimization, signal processing, and machine learning toolbox

electric vehicles

metering

communication networks



demand response

Outline

- System modeling and problem formulations
- Convex relaxations for power flow (PF) analysis (noiseless)
  - Design of the objective function
  - Exact recovery of the PF solution
- Convex relaxations for power system state estimation (noisy)
  - Penalized semidefinite relaxation
  - Theoretically guaranteed performance bound
- Electricity Market Inference
  - Low-rank multi-kernel learning approach
- Summary and future directions

Northeast Blackout of 2003

Time: August 14, 2003 starting at 16:05 EDT

Location: Midwest/Northeast US and Ontario

Effects: 50 million people & 61,800 MWs of load lost. No power for 4 days in some parts of the US, and for more than a week in Ontario.

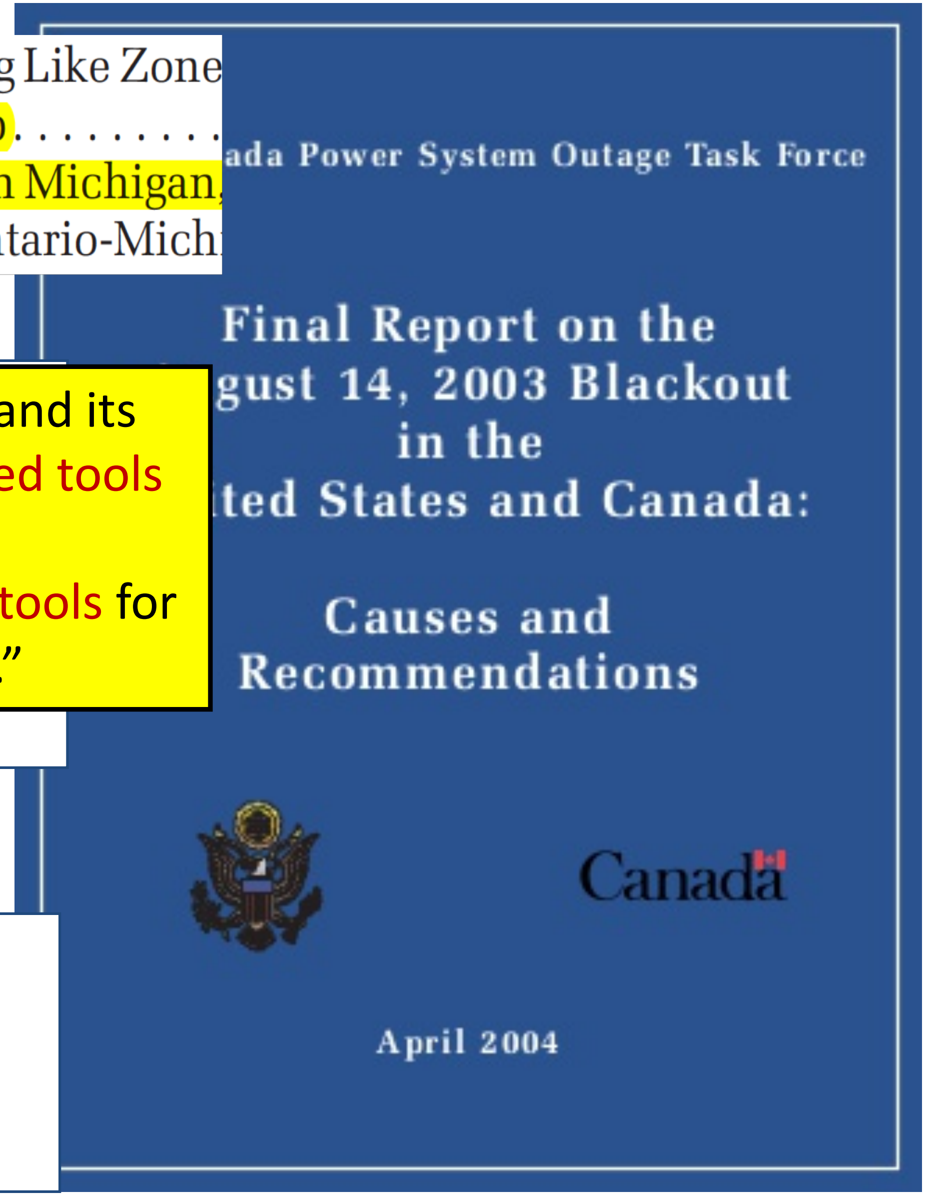
Costs: $4~10 billion in the US and Canada's GDP was down 0.7% in Aug. In Ontario, net loss of 18.9 million work hours, and manufacturing shipments were down $2.3 billion.

Failure of State Estimator

Recommendations: “DOE should expand its research programs on reliability-related tools and technologies.”
“Evaluate and adopt better real-time tools for operators and reliability coordinators.”

awareness”

“**Violation 3:** FE’s state estimation /contingency analysis tools were not used to assess system conditions, violating NERC Operating Policy 5…”


ada Power System Outage Task Force

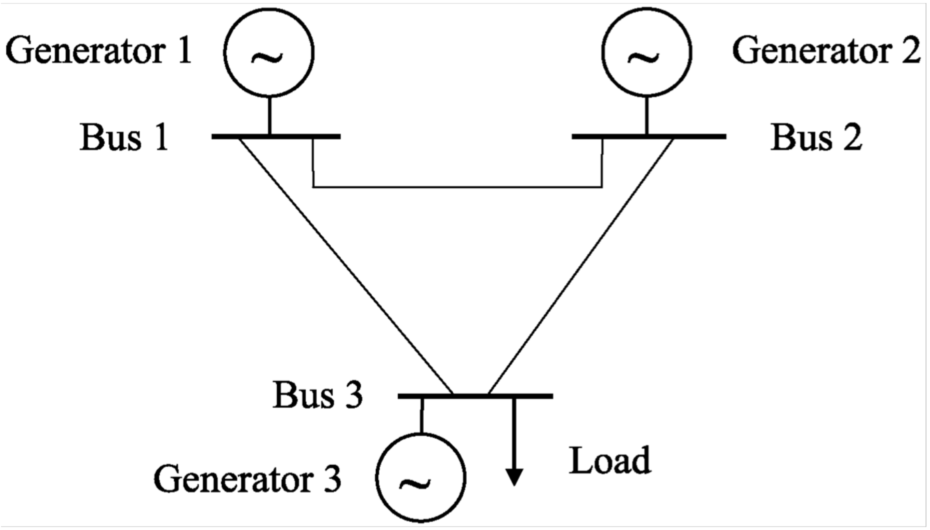
Final Report on the
gust 14, 2003 Blackout
in the
ted States and Canada:

Causes and
Recommendations

Canada

April 2004

System Modeling

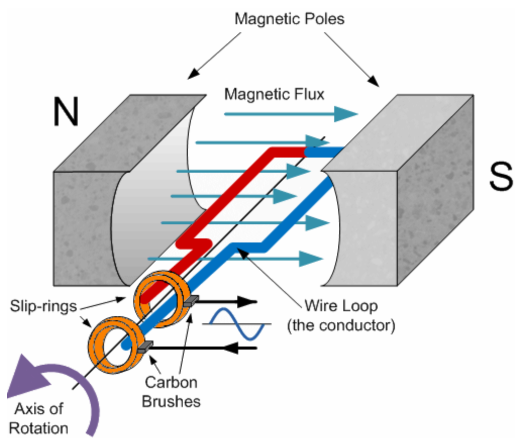


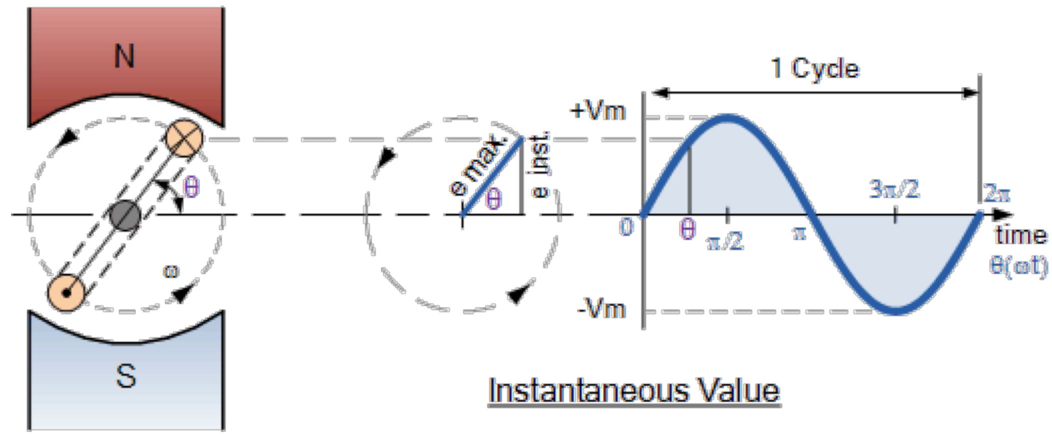
- Represent power grid by a connected graph

$$\mathcal{G} = (\mathcal{N}, \mathcal{L})$$

set of buses — set of power lines

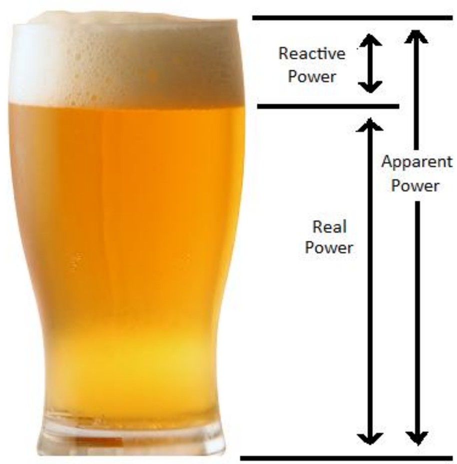
System Modeling (cont’d)

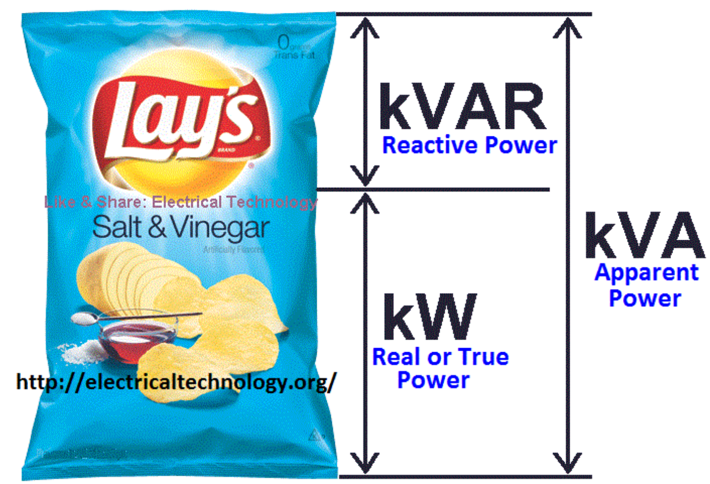




Tin

$v(t$

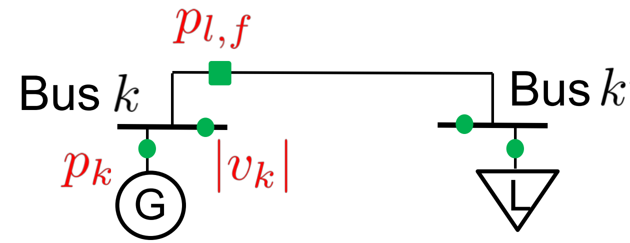




- Net injected complex power: $\mathbf{p} + \mathbf{q}j = \mathrm{diag}(\mathbf{v}\mathbf{i}^*)$

Figure courtesy: www.electronics-tutorials.ws

Nodal and Line Quantities



- Voltage magnitude and nodal power injections

$$|v_k|^2 = \mathrm{Tr}(\mathbf{E}_k \mathbf{v}\mathbf{v}^*), \ p_k = \mathrm{Tr}(\mathbf{Y}_{k,p} \mathbf{v}\mathbf{v}^*), \ q_k = \mathrm{Tr}(\mathbf{Y}_{k,q} \mathbf{v}\mathbf{v}^*)$$

where $\mathbf{E}_k := \mathbf{e}_k \mathbf{e}_k^\top, \ \mathbf{Y}_{k,p} := \frac{1}{2}(\mathbf{Y}^* \mathbf{E}_k + \mathbf{E}_k \mathbf{Y}), \ \mathbf{Y}_{k,q} := \frac{\mathrm{j}}{2}(\mathbf{E}_k \mathbf{Y} - \mathbf{Y}^* \mathbf{E}_k)$

- Branch active and reactive powers

$$\begin{aligned} p_{l,f} &= \mathrm{Tr}(\mathbf{Y}_{l,p_f} \mathbf{v}\mathbf{v}^*), \quad & p_{l,t} &= \mathrm{Tr}(\mathbf{Y}_{l,p_t} \mathbf{v}\mathbf{v}^*) \\ q_{l,f} &= \mathrm{Tr}(\mathbf{Y}_{l,q_f} \mathbf{v}\mathbf{v}^*), \quad & q_{l,t} &= \mathrm{Tr}(\mathbf{Y}_{l,q_t} \mathbf{v}\mathbf{v}^*) \end{aligned}$$

- All quantities are quadratic functions of complex voltage $\mathbf{v}$

$$\boxed{\mathbf{v} = \text{state of the system}}$$

Problem Statement

Power system state estimation (PSSE):

Given noisy measurements $z_j = \mathrm{Tr}(\mathbf{M}_j \mathbf{v}\mathbf{v}^*) + \eta_j,\ j = 1, 2, \ldots, m,$ *estimate the complex voltage* $\mathbf{v}$.

❑ Functionality of PSSE:

- Provides real-time power system conditions
- Constitutes the core of online security analysis
- Provides diagnostics for modeling and maintenance

❑ Measurements:

- From the supervisory control and data acquisition (SCADA) system and phasor measurement units (PMUs)
- Corrupted by noise; missing or grossly inaccurate (outliers/bad data)

Prior Work

- *PF analysis*
  - NP-hard for both T&D networks [Bienstock-Verma'15], [Lehmann et al'16]
  - Newton-Raphson method and fast decoupled load flow (FDLF)
  - Other techniques: Holomorphic embedding LF and numerical polynomial homotopy continuation (NPHC) [Trias'12], [Li'03], [Mehta et al'15]
  - Semidefinite relaxations (SDR) [Madani-Lavaei-Baldick'15]
- *PSSE*
  - Modeling and implementation: [Schweppe et al'70]
  - Gauss-Newton methods [Abur-Gomez'04] [Caro-Conejo-Minguez'09]
  - SDR: [Zhu-Giannakis'11,14], [Weng-Ilic, et al'12,13,15]
- *Our Contributions:*
  - Conditions to guarantee an exact SDR for the PF problem
  - Theoretically quantify the SDR optimal solution for the PSSE problem

Y. Zhang, R. Madani, and J. Lavaei, "Conic Relaxations for Power System State Estimation with Line Measurements," *IEEE Trans. on Control of Network Systems,* Sept. 2018

Power Flow (PF) Problem

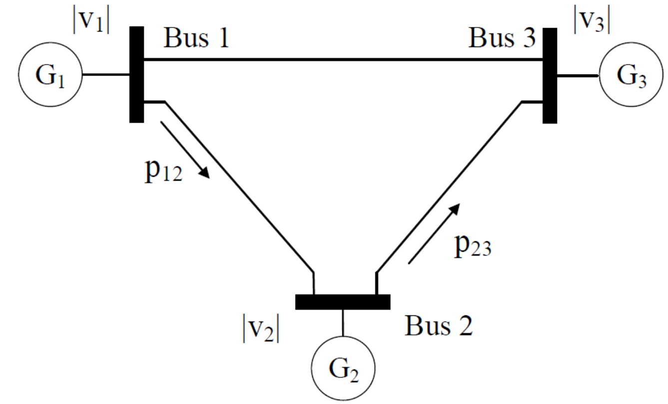
- PF problem: PSSE with noiseless measurements $(\eta_j = 0, \forall j)$

$$\begin{aligned} \text{find} \quad & \mathbf{v} \in \mathbb{C}^n \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{M}_j \mathbf{v}\mathbf{v}^*) = z_j, \ \forall j \in \mathcal{M} \end{aligned}$$

Standard PF	p_k	q_k	$\lvert v_k \rvert$	$\measuredangle v_k$
PV bus	✅	?	✅	?
PQ bus	✅	✅	?	?
Ref bus	?	?	✅	✅

- All measurements are nodal quantities
- Number of equations = number of unknowns $(2n-1)$

Motivating Example



- ❑ Direct calculation of power flows

$$p_{st} = \mathrm{Re}(v_s(v_s - v_t)^* y_{st}^*) \Rightarrow$$

$$\begin{cases} \measuredangle v_1 - \measuredangle v_2 = \arccos\left(\dfrac{p_{12} - |v_1|^2 \mathrm{Re}(y_{12})}{|v_1||v_2||y_{12}|}\right) + \measuredangle y_{12} \\ \measuredangle v_2 - \measuredangle v_3 = \arccos\left(\dfrac{p_{23} - |v_2|^2 \mathrm{Re}(y_{23})}{|v_2||v_3||y_{23}|}\right) + \measuredangle y_{23} \end{cases} \Rightarrow$$

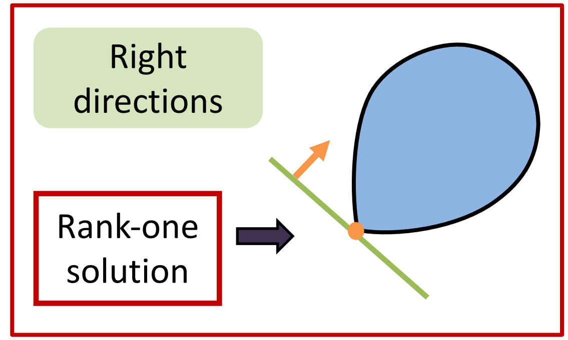
Find unique $\measuredangle v_1, \measuredangle v_2, \measuredangle v_3$ $\quad [0 < (\measuredangle v_s - \measuredangle v_t) - \measuredangle y_{st} < 180^\circ \text{ and } \measuredangle v_{\text{ref}} = 0]$

- ❑ Direct calculation is NOT applicable for noisy measurements
  - Optimization framework to estimate the voltage $\mathbf{v}$?
  - Convexity of the formulation?
  - How good is the performance of the convexification?

Semidefinite Relaxation

- Our approach: Design a linear objective $\mathrm{Tr}(\mathbf{M}_0\mathbf{X})$ with $\mathbf{X} := \mathbf{v}\mathbf{v}^*$

$$\begin{aligned} \underset{\mathbf{X}\in\mathbb{H}^n}{\text{minimize}} \quad & \mathrm{Tr}(\mathbf{M}_0\mathbf{X}) \\ \text{subject to} \quad & \mathrm{Tr}(\mathbf{M}_j\mathbf{X}) = z_j,\ j \in \mathcal{M} \\ & \mathbf{X} \succeq \mathbf{0},\ \cancel{\operatorname{rank}(\mathbf{X}) = 1} \end{aligned}$$



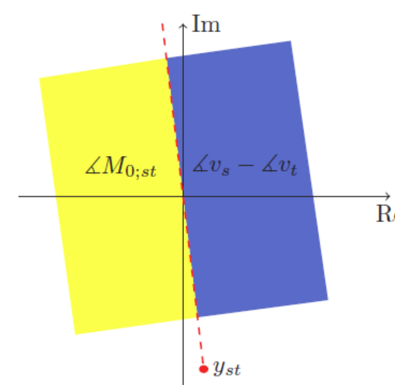
Question: *When is the SDP relaxation exact to recover* $\mathbf{v}$*?*

- Assumptions

A1) Available measurements:
$$\begin{cases} |v_k|^2, \forall k \in \mathcal{N} \\ p_{l,f}\ \ (p_{l,t}), \forall l \in \mathcal{L}_{\mathrm{ST}} \end{cases}$$

A2) Angle conditions:

$$\begin{aligned} -180^\circ < \measuredangle M_{0;st} - \measuredangle y_{st} < 0, \quad & \forall (s,t) \in \mathcal{L}_{\mathrm{ST}}, \\ 0 < (\measuredangle v_s - \measuredangle v_t) - \measuredangle y_{st} < 180^\circ, \quad & \forall (s,t) \in \mathcal{L}_{\mathrm{ST}}, \\ (\measuredangle v_s - \measuredangle v_t) - \measuredangle M_{0;st} \neq 0 \ \text{ or } \ 180^\circ, \quad & \forall (s,t) \in \mathcal{L}_{\mathrm{ST}} \end{aligned}$$



Exact Recovery

Theorem 1

Under assumptions A1-A2, the SDP relaxation recovers the voltage vector $\mathbf{v}$.

- Proof sketch:
  - Dual SDP

$$\begin{aligned} \underset{\boldsymbol{\mu}\in\mathbb{R}^m}{\text{maximize}} \quad & -\mathbf{z}^\top \boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{H}(\boldsymbol{\mu}) := \mathbf{M}_0 + \sum_{j=1}^{m} \mu_j \mathbf{M}_j \succeq \mathbf{0} \end{aligned}$$

  - To show the existence of a dual certificate $\boldsymbol{\mu}$ satisfying

$$\mathbf{H}(\boldsymbol{\mu}) \succeq 0, \quad \mathbf{H}(\boldsymbol{\mu})\mathbf{v} = 0, \quad \operatorname{rank}(\mathbf{H}(\boldsymbol{\mu})) = n-1$$

SOCP Relaxation

$$\begin{aligned} \underset{\mathbf{X}\in\mathbb{H}^N}{\text{minimize}} \quad & \mathrm{Tr}(\mathbf{M}_0\mathbf{X}) \\ \text{subject to} \quad & X_{k,k} = |v_k|^2, \quad \forall k \in \mathcal{N} \\ & \mathrm{Tr}(\mathbf{Y}_{l,p_f}\mathbf{X}) = p_{l,f}, \quad \forall l \in \mathcal{L}_{\mathrm{ST}} \\ & \begin{bmatrix} X_{s,s} & X_{s,t} \\ X_{t,s} & X_{t,t} \end{bmatrix} \succeq \mathbf{0}, \quad \forall (s,t) \in \mathcal{L}_{\mathrm{ST}} \end{aligned}$$

Theorem 2

Under assumptions A1-A2, the SOCP relaxation recovers the voltage $\mathbf{v}$.

- Additional measurements

Corollary 1

Under assumptions A1-A2, the SDP and SOCP relaxations with additional constraints of power injection and line measurements both recover the voltage $\mathbf{v}$.

Power System State Estimation

- Penalized SDP:

$$\begin{aligned} \underset{\mathbf{X}\in\mathbb{H}^n,\boldsymbol{\nu}\in\mathbb{R}^m}{\text{minimize}} \quad & \rho f(\boldsymbol{\nu}) + \mathrm{Tr}(\mathbf{M}_0\mathbf{X}) \\ \text{subject to} \quad & \mathrm{Tr}(\mathbf{M}_j\mathbf{X}) + \nu_j = z_j, \quad \forall j \in \mathcal{M} \\ & \mathbf{X} \succeq \mathbf{0} \end{aligned}$$

- For example: $f_{\mathrm{WLAV}}(\boldsymbol{\nu}) = \sum_{j=1}^{m} |\nu_j|/\sigma_j, \quad f_{\mathrm{WLS}}(\boldsymbol{\nu}) = \sum_{j=1}^{m} \nu_j^2/\sigma_j^2$

Theorem 3

If $\rho \geq \max_{j\in\mathcal{M}} |\sigma_j \hat{\mu}_j|$, *then there exists a scalar* $\beta > 0$ *such that*

$$\zeta := \frac{\|\mathbf{X}^{\mathrm{opt}} - \beta\mathbf{v}\mathbf{v}^*\|_F}{\sqrt{n \times \mathrm{Tr}(\mathbf{X}^{\mathrm{opt}})}} \leq 2\sqrt{\frac{\rho \times f_{\mathrm{WLAV}}(\boldsymbol{\eta})}{n\lambda}},$$

where λ *is the second smallest eigenvalue of* $\mathbf{H}(\hat{\boldsymbol{\mu}})$.

Estimation Error

- Define the root-mean-square error:

$$\zeta := \frac{\|\mathbf{X}^{\mathrm{opt}} - \beta \mathbf{v}\mathbf{v}^*\|_F}{\sqrt{n \times \mathrm{Tr}(\mathbf{X}^{\mathrm{opt}})}} \leq 2 \overbrace{\sqrt{\frac{\rho}{n\lambda}}}^{\omega(\mathcal{G}')} \sqrt{f_{\mathrm{WLAV}}(\boldsymbol{\eta})}$$

> ***Corollary 2***
>
> *Under assumptions of Theorem 3, the tail probability of the estimation error* ζ *is upper bounded as* $\mathbb{P}(\zeta > t) \leq \mathrm{e}^{-\gamma m}$ *for every* $t > 0$, *where* $\gamma = \frac{t^4 \lambda^2}{32 \kappa^2 \rho^2} - \ln 2.$

- Effect of more measurements

> ***Theorem 4***
>
> *Consider two choices of the graph* $\mathcal{G}'$*, denoted as* $\mathcal{G}'_1$ *and* $\mathcal{G}'_2$*, such that* $\mathcal{G}'_1$ *is a subgraph of* $\mathcal{G}'_2$. *Then, the relation* $\omega(\mathcal{G}'_2) \leq \omega(\mathcal{G}'_1)$ *holds.*

Rank-1 Approximation

- Given $\mathbf{X}^{\text{opt}}$ of the penalized SDP , we can recover $\mathbf{v} \in \mathbb{C}^n$ as follows:

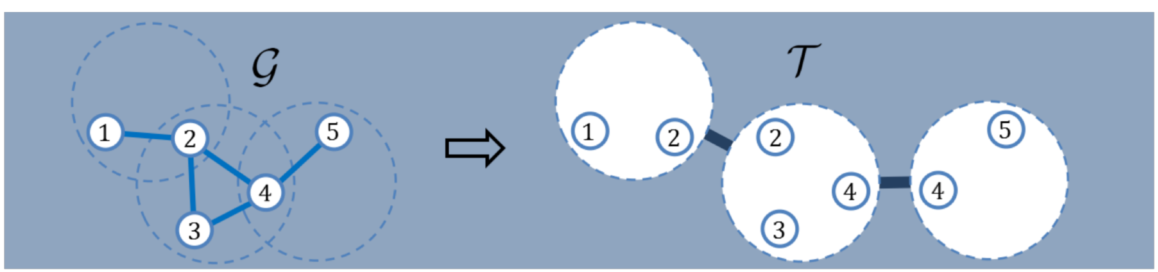
S1) set the voltage magnitude $|\hat{v}_k| = \sqrt{\mathbf{X}^{\text{opt}}_{k,k}}, \text{ for } k = 1, 2, \ldots, n$

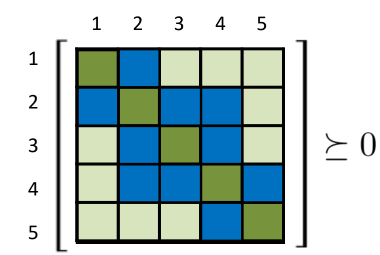
S2) set the voltage angles by solving an LP:

$$
\begin{aligned}
\measuredangle \hat{\mathbf{v}} = \ & \underset{\measuredangle \mathbf{v} \in [-\pi,\pi]^N}{\arg\min} \sum_{(s,t)\in\mathcal{L}} |\measuredangle \mathbf{X}^{\text{opt}}_{s,t} - \measuredangle v_s + \measuredangle v_t| \\
\text{s. t.} \quad & \measuredangle v_{\text{ref}} = 0
\end{aligned}
$$

- Complexity reduction: tree decomposition

Full-scale to decomposed SDP

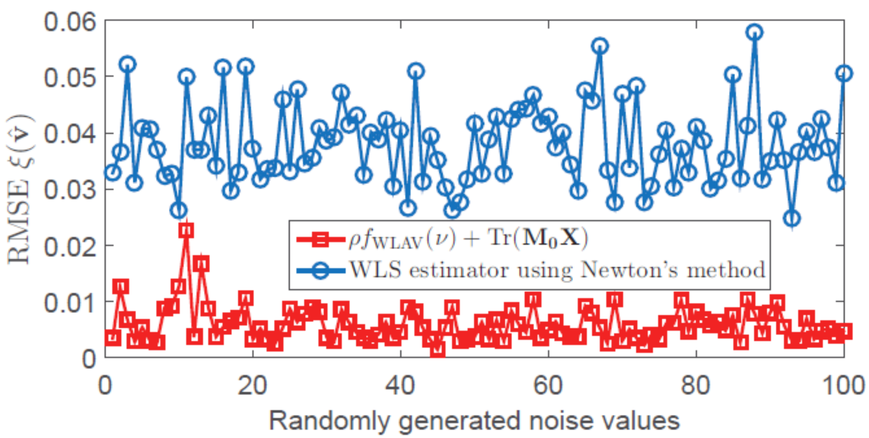




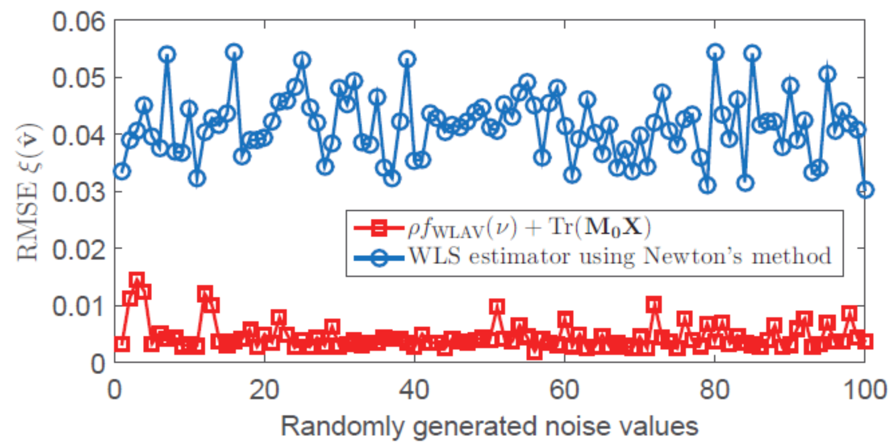
R. Madani, M. Ashraphijuo, and J. Lavaei, "Promises of Conic Relaxation for Contingency-Constrained Optimal Power Flow Problem," *IEEE Trans. Power Systems*, 2016.

Penalized SDP vis-à-vis Newton's Method

- Performance metric: $\mathrm{RMSE} = \|\hat{\mathbf{v}} - \mathbf{v}_o\|/\sqrt{n}$
  - Available measurements: $\{|v_k|^2\}_{k\in\mathcal{N}}$ and $\{p_{l,f}, p_{l,t}\}_{l\in\mathcal{L}}$
  - Bad data: 20% of the line measurements



(a) IEEE 57-bus system



(b) IEEE 118-bus system

Comparison of Different Regularizers

- IEEE benchmark systems
  - Various noise levels: $\sigma_j = c \times |\bar{v}_k|^2,\ 1.5c \times \bar{p}_k,\ \text{or}\ 2c \times \bar{p}_{l,f}$
  - Bad data: 10% of the measurements
  - 50 Monte-Carlo simulations
- Performance measure: $\xi(\hat{\mathbf{v}}) = \|\hat{\mathbf{v}} - \mathbf{v}_o\| / \sqrt{n}$

Methods	$\rho f(\boldsymbol{\nu}) + \mathrm{Tr}(\mathbf{M}_0\mathbf{X})$		$\rho f(\boldsymbol{\nu}) + \|\mathbf{X}\|_*$		$\rho f(\boldsymbol{\nu})$	
	WLAV	WLS	WLAV	WLS	WLAV	WLS
9-bus	0.0648	0.1293	1.2744	1.1483	1.1619	1.1633
14-bus	0.1307	0.1784	1.1320	1.3871	1.4233	1.4215
30-bus	0.2055	0.2543	1.4236	1.4306	1.4269	1.4268
39-bus	0.1324	0.1239	1.1317	1.3135	1.2764	1.2757
57-bus	0.2343	0.2809	1.2981	1.3004	1.3235	1.3098
118-bus	0.1136	0.1641	1.3620	1.3272	1.3445	1.3577

- Proposed SDP approach has the best performance

Error Bound and Scaling Factor

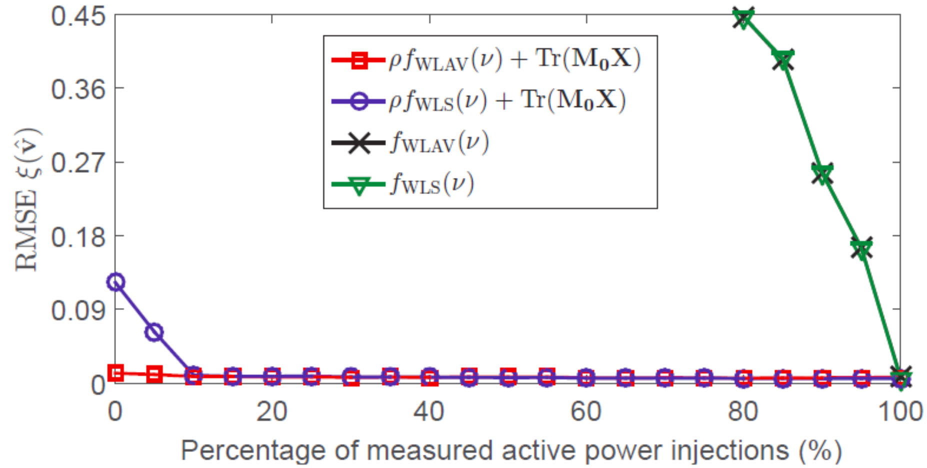
- Error bound $\zeta := \frac{\|\mathbf{X}^{\text{opt}} - \beta \mathbf{v}\mathbf{v}^*\|_F}{\sqrt{n \times \text{Tr}(\mathbf{X}^{\text{opt}})}} \leq 2\sqrt{\frac{\rho}{n\lambda}}\sqrt{f_{\text{WLAV}}(\boldsymbol{\eta})} =: \zeta^{\max}$

Cases	$\xi(\hat{\mathbf{v}})$	ζ	$\zeta^{\max}$	β	λ	f_{WLAV}	$\rho^{\min}$
9-bus	0.0111	0.0145	0.1535	0.9972	1.3417	14.768	0.0048
14-bus	0.0057	0.0078	0.2859	1.0005	0.3812	20.509	0.0053
30-bus	0.0060	0.0084	0.3728	0.9997	0.1094	51.479	0.0022
39-bus	0.0077	0.0083	0.8397	1.0009	0.7438	62.558	0.0817
57-bus	0.0092	0.0102	0.8364	1.0013	0.0912	88.434	0.0103
118-bus	0.0057	0.0079	1.2585	0.9992	0.0878	179.509	0.0228

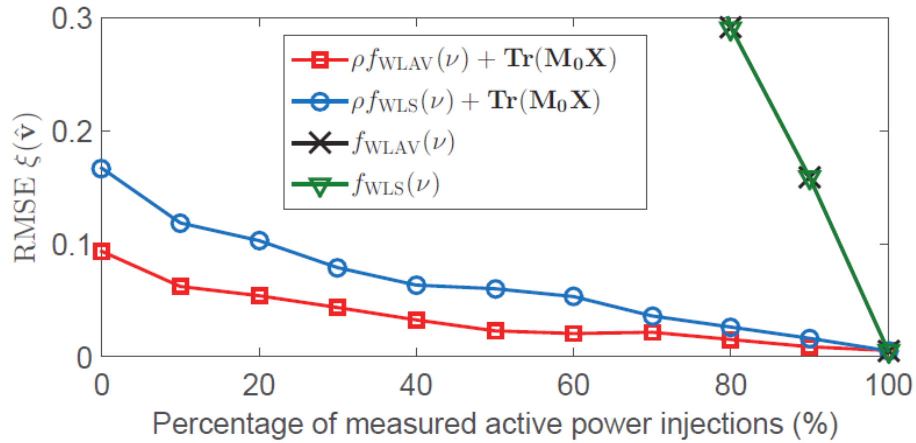
- The RMSEs ζ and $\xi(\hat{\mathbf{v}})$: the same order of the noise level $c = 0.01$
- The scaling factor $\beta \approx 1$: the optimal solution $\mathbf{X}^{\text{opt}}$ is close to the true lifted state $\mathbf{v}\mathbf{v}^*$

Effect of Additional Active Power Injections

- PEGASE 1354-bus: SDP relaxation with tree decomposition



- PEGASE 9241-bus: SOCP relaxation



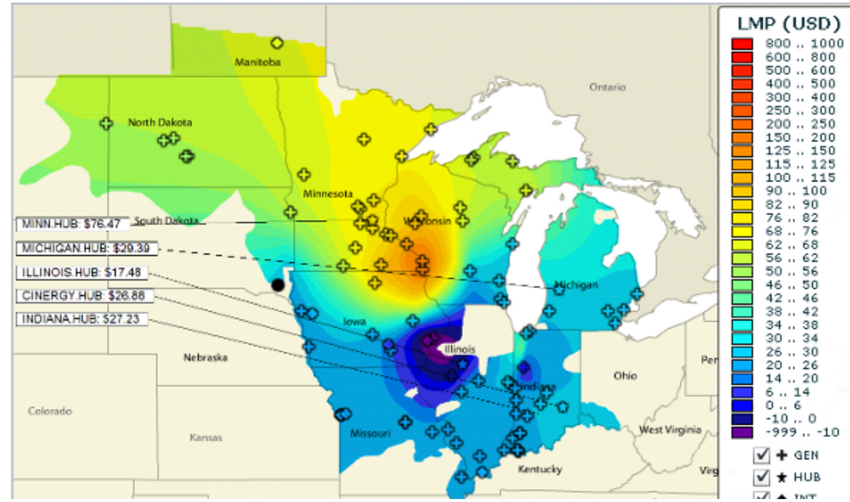
Simulation Time

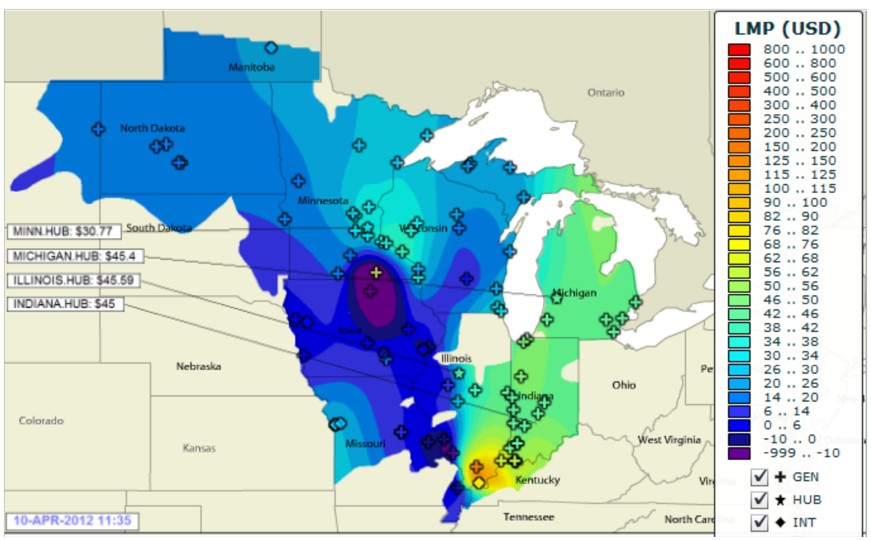
- Modeling tool and solver: CVX+SDPT3

Cases	Solver time	Total time
9-bus	0.89s	1.58s
14-bus	1.23s	2.54s
30-bus	1.33s	3.21s
39-bus	1.56s	3.28s
57-bus	1.97s	4.09s
118-bus	2.38s	5.63s
1354-bus	4.55s	9.48s
2869-bus	13.17s	24.44s
9241-bus	58.00s	109.14s

Market Inference Motivation

- ❑ Interest from
  - market participants
  - independent system operators (ISO)
  - congestion corridors





Locational marginal prices (LMPs) vary *spatiotemporally*

- ❑ Challenges
  - uncertainty: load, renewables
  - bidding and hedging
  - outages and security

Problem Statement

Market inference:
Assume that price $p(n,t)$ *at node n and hour t, depends on nodal features* $\mathbf{x}_n$ *and time features* $\mathbf{y}_t$.
Given historical feature-price pairs $\{(\mathbf{x}_n, \mathbf{y}_t), p(n,t)\}_{n\in\mathcal{N}, t\in\mathcal{T}}$,
infer prices $p(n', t')$ *at given* $\{(\mathbf{x}_{n'}, \mathbf{y}_{t'})\}_{n'\in\mathcal{N}', t'\in\mathcal{T}'}$.

- Interpolation and extrapolation (forecasting)

- Nodal features
  - location
  - bus type: generator/load/interface

- Time features
  - yesterday's prices
  - temperature
  - day, hour, holiday

Prior Work

- Time-series modeling (ARIMA) [Contreras et al‘03], [Conejo’05]
- Artificial intelligence: fuzzy systems, neural networks
  [Gonzalez et al’05], [Li et al’07], [Wu-Shahidehpour’10]
- Nearest-neighbor approach [Lora-Exposito’07]
- Quadratic program with outage combinations [Zhou-Tesfatsion-Liu’11]

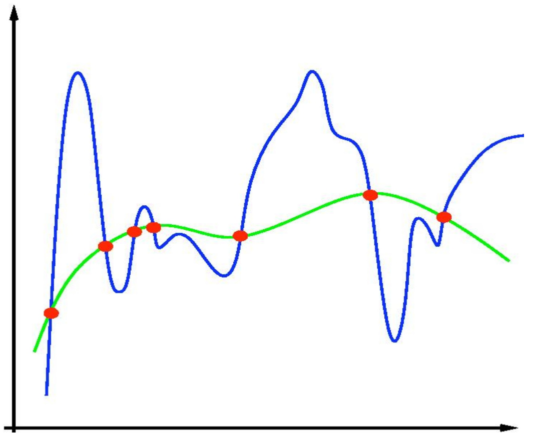
- Existing works: predictors trained per node, no spatial correlation
- Contribution: **kernel-based** network-wide forecasting

Kernel-based Learning

- Given data $\{(x_n, z_n)\}_{n=1}^{N}$, and kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

- Find $$\hat{f} := \arg\min_{f \in \mathcal{H}_{\mathcal{K}}} \sum_{n=1}^{N} (z_n - f(x_n))^2 + \mu \|f\|_{\mathcal{K}}$$

  where $\mathcal{H}_{\mathcal{K}} := \{f(x) = \sum_{n=1}^{\infty} K(x, x_n) a_n\}$,



- Representer theorem: $$\hat{f}(x) = \sum_{n=1}^{N} K(x, x_n) \hat{a}_n$$

$$\hat{\mathbf{a}} := \arg\min_{\mathbf{a} \in \mathbb{R}^N} \|\mathbf{z} - \mathbf{K}\mathbf{a}\|_2^2 + \mu \sqrt{\mathbf{a}^\top \mathbf{K} \mathbf{a}}$$

where $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$

Gaussian kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right)$$

Low-rank Inference

- Inferring $p(n,t)$ posed as learning function $p : \mathcal{N} \times \mathcal{T} \to \mathbb{R}$

- Functional space $\mathcal{P} := \left\{ p(n,t) = \sum_{r=1}^{R} f_r(n) g_r(t),\; f_r \in \mathcal{H}_\mathcal{K},\; g_r \in \mathcal{H}_\mathcal{G} \right\}$
  - tensor product kernel $K_\otimes((n,t),(n',t')) := K(n,n') \cdot G(t,t')$

- *Presumption:* R is small
  - few congested lines yield low-dimensional price differentiation

- Low-rank model via *trace-norm* penalty $\min_{p \in \mathcal{P}} \|\mathbf{Z} - \mathbf{P}\|_F^2 + \mu \|p\|_*$
  - collaborative filtering [Abernethy-Bach-Evgeniou-Vert'09]
  - nuclear-norm optimization [Recht-Fazel-Parrilo'07]

Jacob Abernethy et al, "Low-rank matrix factorization with attributes," Ecole des Mines de Paris, Tech. Rep. 2006.

Multi-Kernel Electricity Price Forecasting

- Joint kernel selection and functional regularization

$$\min_{\mathcal{K},\mathcal{G}} \min_{p \in \mathcal{P}} \|\mathbf{Z} - \mathbf{P}\|_F^2 + \mu \sqrt{\|p\|_*}$$

  - over kernel pools $\mathcal{K} := \text{conv}\left(\{\mathcal{K}_l\}_{l=1}^L\right)$ $\mathcal{G} := \text{conv}\left(\{\mathcal{G}_m\}_{m=1}^M\right)$

- Functional to matrix optimization

$$\min_{\mathbf{P},\{\mathbf{B}_l\},\{\mathbf{\Gamma}_m\}} \|\mathbf{Z} - \mathbf{P}\|_F^2 + \mu \sum_{l=1}^{L} \|\mathbf{B}_l\|_{\mathbf{K}_l} + \mu \sum_{m=1}^{M} \|\mathbf{\Gamma}_m\|_{\mathbf{G}_m}$$

$$\text{s.to} \quad \mathbf{P} = \sum_{l=1}^{L} \sum_{m=1}^{M} \mathbf{K}_l \mathbf{B}_l \mathbf{\Gamma}_m^\top \mathbf{G}_m.$$

- Block coordinate descent (BCD) solver: Guaranteed convergence

Vassilis Kekatos, Yu Zhang, and Georgios Giannakis, "Electricity Market Forecasting via Low-Rank Multi-Kernel Learning," IEEE J. Selected Topics on Signal Proc., 2014.

Block Coordinate Descent (BCD)

$$\min_{\{\mathbf{B}_l\},\{\mathbf{\Gamma}_m\}} \|\mathbf{Z} - \sum_{l=1}^{L}\sum_{m=1}^{M} \mathbf{K}_l\mathbf{B}_l\mathbf{\Gamma}_m^\top\mathbf{G}_m\|_F^2 + \mu\sum_{l=1}^{L}\|\mathbf{B}_l\|_{\mathbf{K}_l} + \mu\sum_{m=1}^{M}\|\mathbf{\Gamma}_m\|_{\mathbf{G}_m}$$

- Guaranteed convergence [Tseng'01]

❑ Canonical problem

$$\hat{\mathbf{X}} := \arg\min_{\mathbf{X}} \ \|\mathbf{A} - \mathbf{B}\mathbf{X}\mathbf{C}^\top\|_F^2 + \mu\|\mathbf{X}\|_{\mathbf{B}}$$

- Unique minimizer found efficiently as

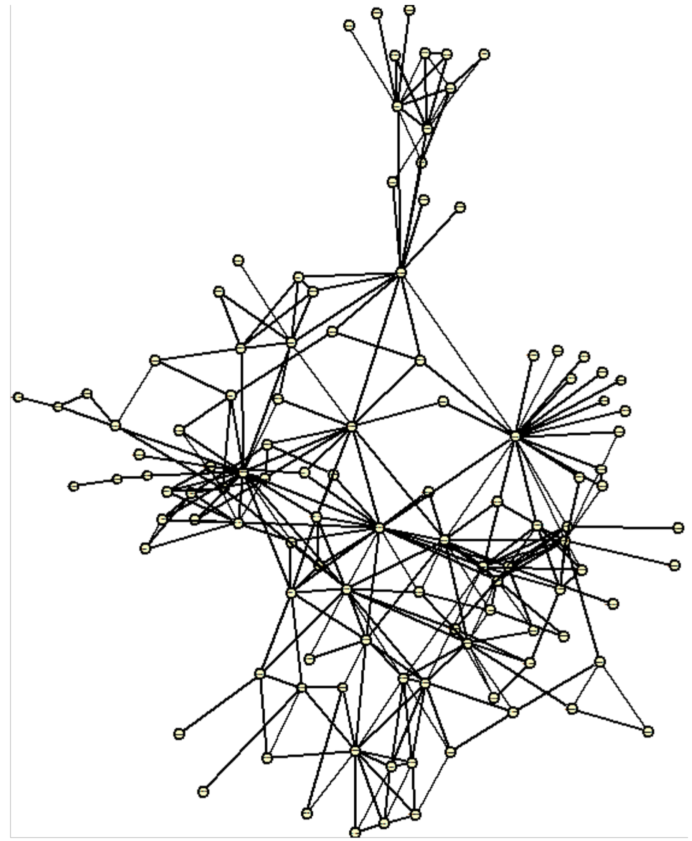
✓ If $\|\mathbf{B}^{1/2}\mathbf{A}\mathbf{C}\|_F \leq \mu/2, \quad \hat{\mathbf{X}} = \mathbf{0};$

✓ else, solve Sylvester equation $\mathbf{B}\hat{\mathbf{X}}\mathbf{C}^\top\mathbf{C} + \frac{\mu^2}{4\hat{w}}\hat{\mathbf{X}} = \mathbf{A}\mathbf{C}$

❑ Once $\{\hat{\mathbf{B}}_l\}$, $\{\hat{\mathbf{\Gamma}}_m\}$ found, forecasting by $\hat{\mathbf{P}}' = \sum_{l=1}^{L}\sum_{m=1}^{M}\mathbf{K}'_l\hat{\mathbf{B}}_l\hat{\mathbf{\Gamma}}_m^\top\mathbf{G}'_m$

MISO Market Prediction

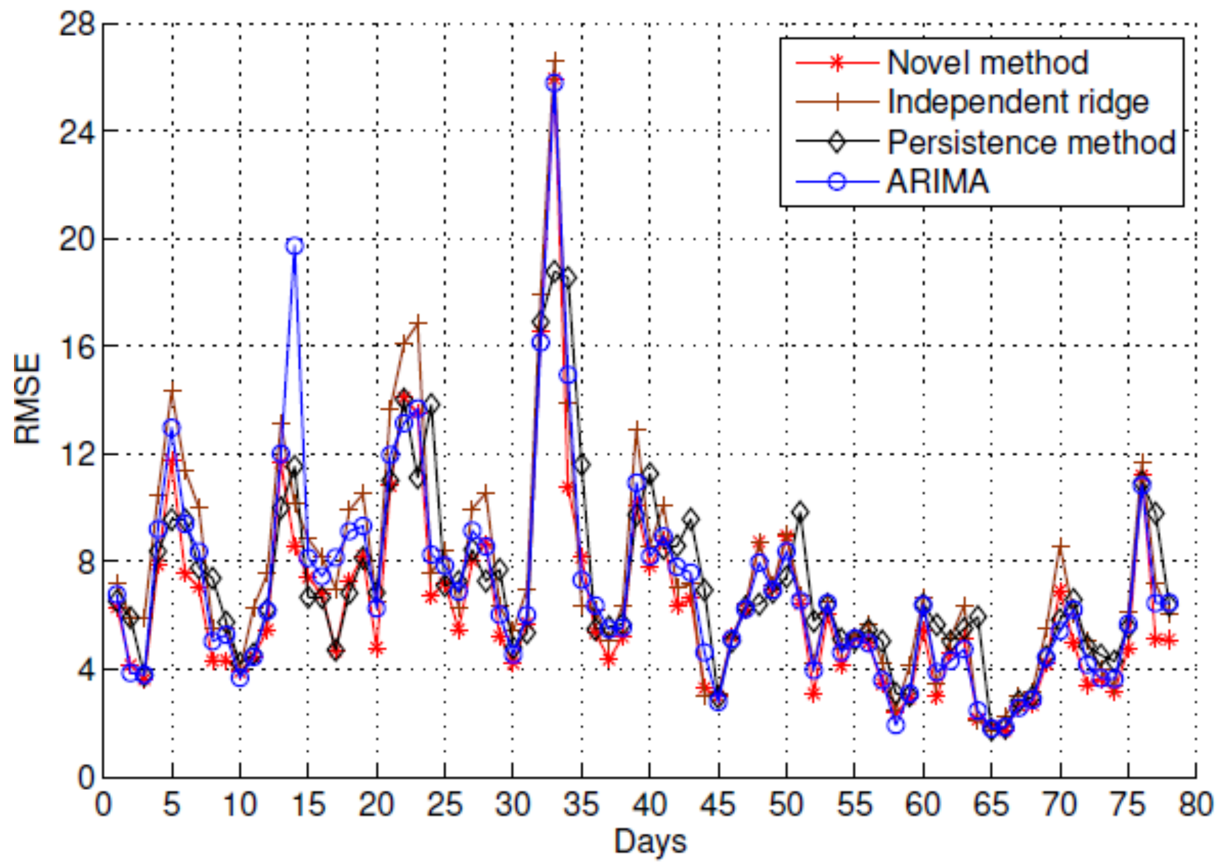
- MISO summer 2012
  - $N = 1,732$ nodes
  - $L = 5, M = 5$ kernels
  - Predict next 24 hrs using last week's data



Connectivity graph of the local balancing authority (LBA)

- Nodal kernel
  - regularized Laplacian
  - diffusion Laplacian
  - Gaussian
  - identity
  - covariance
- Time kernel
  - Gaussian
  - linear

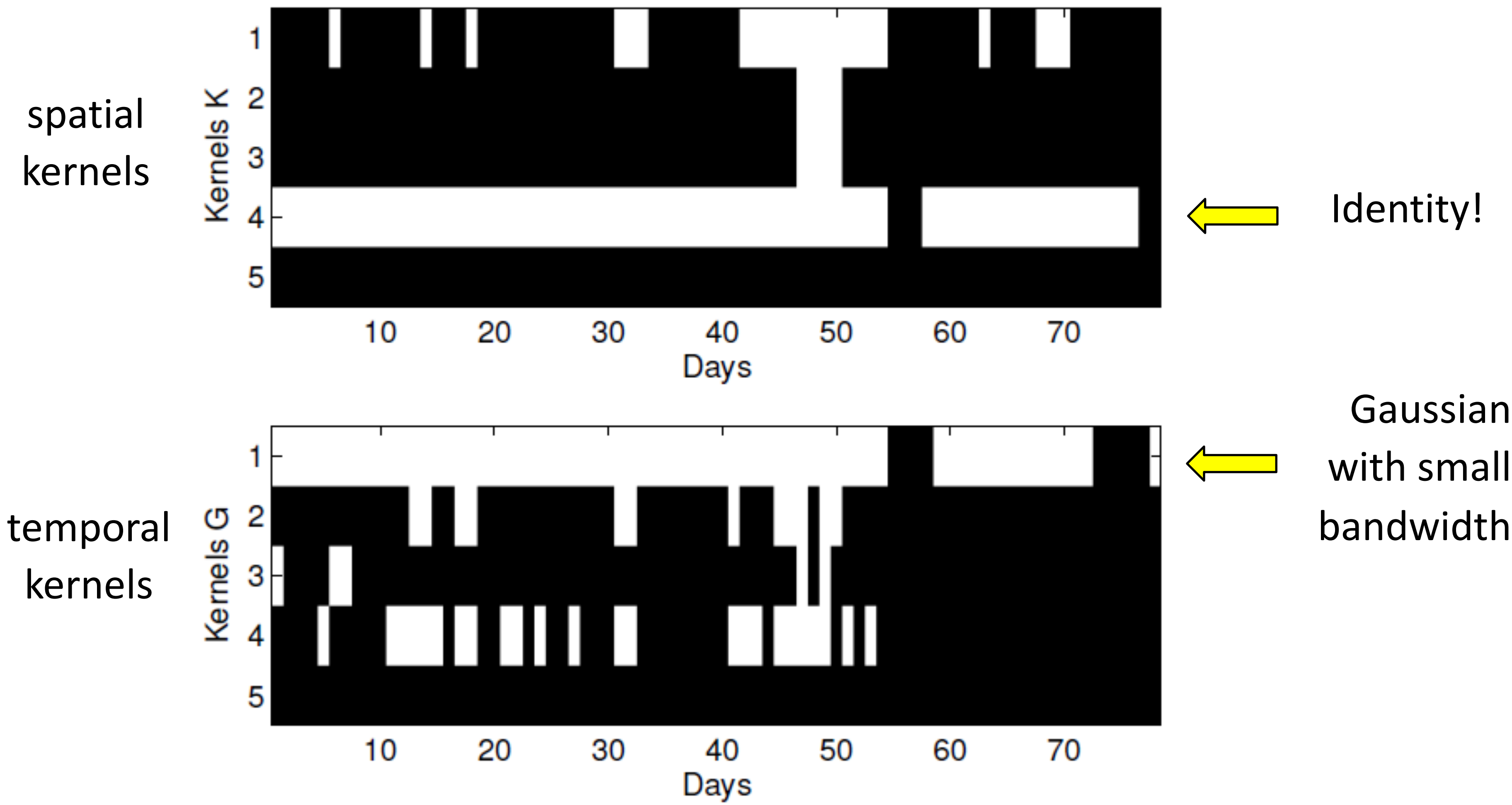
Day-ahead Prediction



- Average forecast error of the MISO market:

	Novel	Ridge	Persistence	ARIMA
RMSE	6.40	7.55	7.20	7.06
MAE	3.51	4.40	3.81	3.80

Kernel Selection



Summary

❑ Takeaways

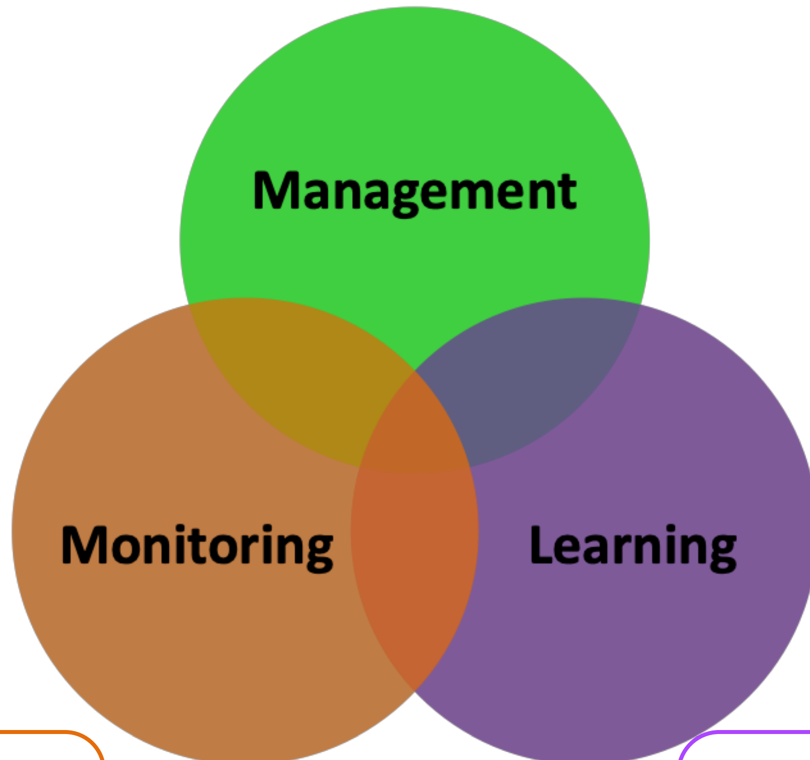
- Framework of conic relaxations for PSSE
- Conditions of exact recovery of the PF solution via SDP/SOCP relaxations
- Estimation error bound of the SDR optimal solution for PSSE
- Low-rank multi-kernel learning to capture spatio-temporal correlations

❑ Ongoing and future works

- Joint PSSE and topology identification: transmission switching
- Robust PSSE: uncertain system parameters; e.g., admittance values
- Dynamic and online PSSE
- Deep neural networks for market inference and energy disaggregation

More Opportunities

- Optimal power flow, unit commitment
- Voltage/frequency regulation
- Transmission switching



- State estimation
- Sensor placement
- Topology identification

- Energy disaggregation
- Anomaly detection
- Cascading failure analysis

Q & A

Thank You!

Proof of Theorem 1

$$\mathbf{H} = \mathbf{M}_0 + \sum_{j=1}^{m} \mu_j \mathbf{M}_j = \sum_{j=n+1}^{m} \overbrace{\left(\mathbf{M}_{0,j} + \frac{1}{m-n} \sum_{k=1}^{n} \mu_k \mathbf{E}_k + \mu_j \mathbf{M}_j\right)}^{\mathbf{H}_j}$$

- Explore the special structure of matrix $\mathbf{Y}_{l,pf}$

$$\mathbf{Y}_{l,pf}(s,t) = \mathbf{Y}_{l,pf}^{*}(t,s) = -\frac{y_{st}}{2}, \quad \mathbf{Y}_{l,pf}(s,s) = \mathrm{Re}(y_{st})$$

Define $\tilde{\mu}_s := \frac{1}{m-n}\mu_s + \mu_s \mathrm{Re}(y_{st})$, $\tilde{\mu}_t := \frac{1}{m-n}\mu_t$

$$\tilde{\mathbf{H}}_j := \begin{bmatrix} \tilde{\mu}_s & m_{st,j} - \frac{\mu_j}{2} y_{st} \\ m_{st,j}^{*} - \frac{\mu_j}{2} y_{st}^{*} & \tilde{\mu}_t \end{bmatrix}$$

$$\mathbf{H}_j \mathbf{v} = \mathbf{0} \Rightarrow \tilde{\mathbf{H}}_j \check{\mathbf{v}} = \mathbf{0} \Rightarrow$$

$$\boxed{\begin{aligned} m_{st,j}^{*} - \frac{\mu_j}{2} y_{st}^{*} &= -\frac{\tilde{\mu}_t v_t}{v_s} \\ \tilde{\mu}_s &= \tilde{\mu}_t \frac{|v_t|^2}{|v_s|^2} \end{aligned}} \Rightarrow \tilde{\mathbf{H}}_j = \tilde{\mu}_t \begin{bmatrix} \frac{|v_t|^2}{|v_s|^2} & -\frac{v_t^*}{v_s^*} \\ -\frac{v_t}{v_s} & 1 \end{bmatrix}$$

rank-1

Proof of Theorem 1 (cont'd)

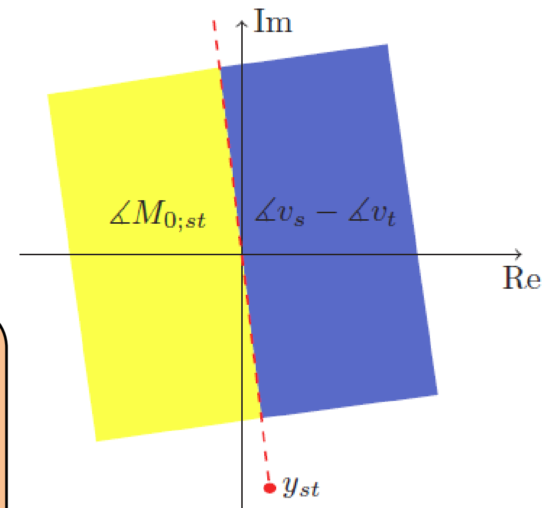
- To prove $\quad \mathrm{rank}(\mathbf{H}) = n-1 \Leftarrow \dim(\mathrm{null}(\mathbf{H})) = 1 \Leftarrow \mathbf{x} = r\mathbf{v},\ \forall \mathbf{x} \in \mathrm{null}(\mathbf{H})$

$$\left.\begin{array}{l} \mathbf{H}_j\mathbf{x} = \mathbf{0} \Rightarrow \frac{x_s}{x_t} = \frac{v_s}{v_t} \text{ for } j\text{-th measurement over line } (s,t) \\ \mathbf{H}_{j'}\mathbf{x} = \mathbf{0} \Rightarrow \frac{x_t}{x_a} = \frac{v_t}{v_a} \text{ for } j'\text{-th measurement over line } (t,a) \end{array}\right\} \Rightarrow$$

$$\frac{x_s}{v_s} = \frac{x_t}{v_t} = \frac{x_a}{v_a} = r$$

- Line measurements over the connected subgraph $\mathcal{G}_s$, repeat the argument for all the buses.

- Angle conditions: $\quad \tilde{\mu}_t = -(m^*_{st,j} - \frac{\mu_j}{2} y^*_{st})\frac{v_s}{v_t} > 0$



$$\Rightarrow \boxed{\begin{array}{rl} -180^\circ < \measuredangle M_{0;st} - \measuredangle y_{st} < 0, & \forall (s,t) \in \mathcal{L}_\mathcal{T}, \\ 0 < (\measuredangle v_s - \measuredangle v_t) - \measuredangle y_{st} < 180^\circ, & \forall (s,t) \in \mathcal{L}_\mathcal{T}, \\ (\measuredangle v_s - \measuredangle v_t) - \measuredangle M_{0;st} \neq 0 \text{ or } 180^\circ, & \forall (s,t) \in \mathcal{L}_\mathcal{T} \end{array}}$$

Proof of Theorem 2

- Penalized SDP: $$\boxed{\min_{\mathbf{X}\succeq\mathbf{0}} \operatorname{Tr}(\mathbf{M}_0\mathbf{X})+\rho\sum_{j=1}^{m}\sigma_j^{-1}\left|\operatorname{Tr}\left(\mathbf{M}_j(\mathbf{X}-\mathbf{v}\mathbf{v}^*)\right)-\eta_j\right|}$$

$$\begin{cases} \operatorname{Tr}\left(\mathbf{M}_0(\mathbf{X}^{\text{opt}}-\mathbf{v}\mathbf{v}^*)\right)+\rho\sum_{j=1}^{m}\sigma_j^{-1}\left|\operatorname{Tr}\left(\mathbf{M}_j(\mathbf{X}^{\text{opt}}-\mathbf{v}\mathbf{v}^*)\right)\right| \le 2\rho f_{\text{WLAV}}(\boldsymbol{\eta}) \\ \mathbf{M}_0=\mathbf{H}(\hat{\boldsymbol{\mu}})-\sum_{j=1}^{m}\hat{\mu}_j\mathbf{M}_j,\ \ \mathbf{H}(\hat{\boldsymbol{\mu}})\mathbf{v}=\mathbf{0} \end{cases}$$

$$\Rightarrow\quad \operatorname{Tr}(\mathbf{H}(\hat{\boldsymbol{\mu}})\mathbf{X}^{\text{opt}})\le 2\rho f_{\text{WLAV}}(\boldsymbol{\eta}),\ \text{if}\ \rho\ge\max_{j\in\mathcal{M}}|\sigma_j\hat{\mu}_j|$$

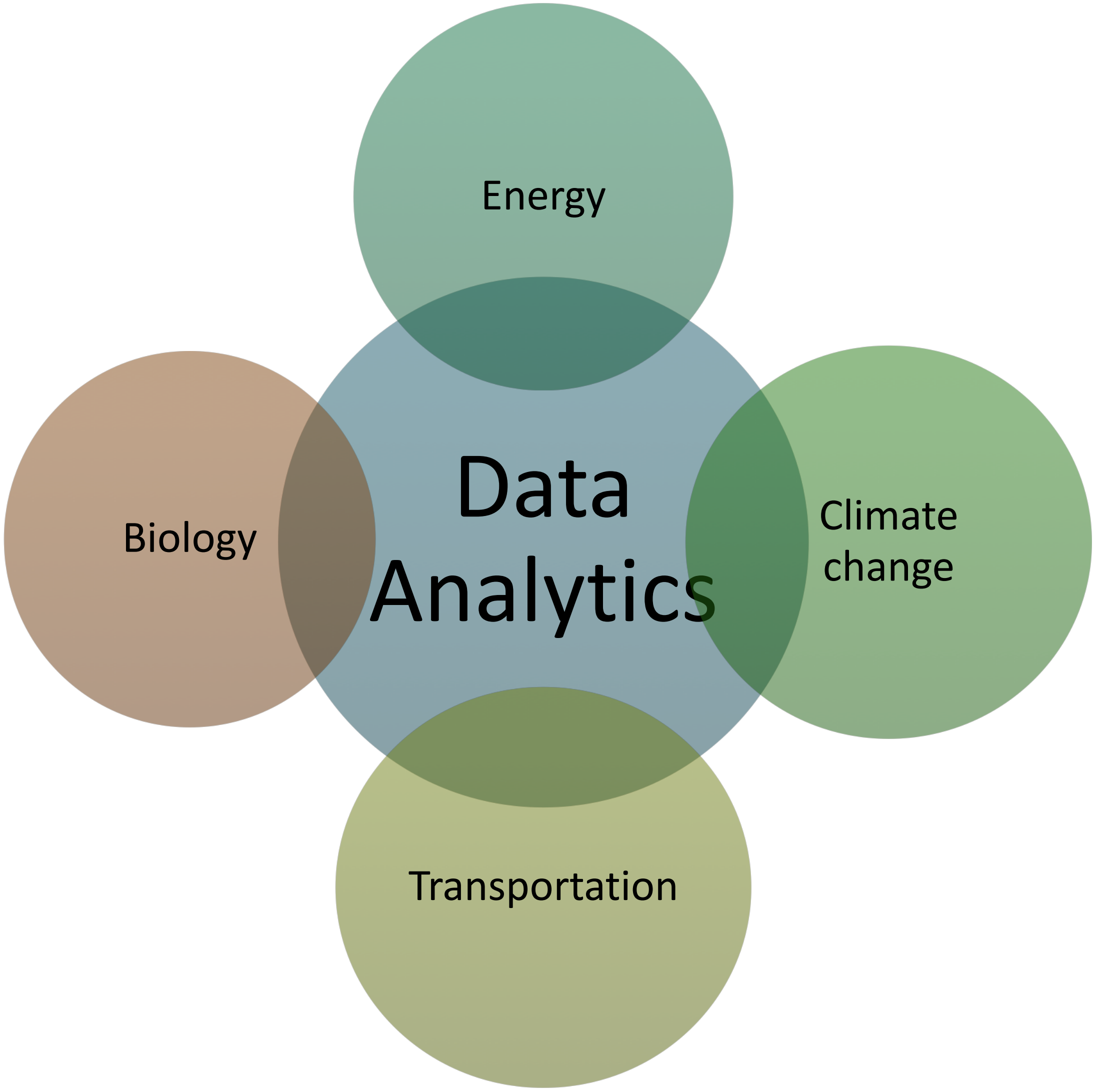
$$\begin{cases} \check{\mathbf{X}}:=\begin{bmatrix}\tilde{\mathbf{X}} & \tilde{\mathbf{x}}\\ \tilde{\mathbf{x}}^* & \alpha\end{bmatrix}=\mathbf{U}^*\mathbf{X}^{\text{opt}}\mathbf{U}\ \Rightarrow\ 2\rho f_{\text{WLAV}}(\boldsymbol{\eta})\ge\operatorname{Tr}(\boldsymbol{\Lambda}\check{\mathbf{X}})\ge\lambda\operatorname{Tr}(\tilde{\mathbf{X}}) \\ \tilde{\mathbf{X}}-\alpha^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}^*\succeq\mathbf{0}\ \Rightarrow\ \|\tilde{\mathbf{x}}\|_2^2\le\alpha\operatorname{Tr}(\tilde{\mathbf{X}})=\operatorname{Tr}(\mathbf{X}^{\text{opt}})\operatorname{Tr}(\tilde{\mathbf{X}})-\operatorname{Tr}^2(\tilde{\mathbf{X}}) \end{cases}$$

$$\mathbf{X}^{\text{opt}}=\left[\tilde{\mathbf{U}}\ \ \tilde{\mathbf{v}}\right]\begin{bmatrix}\tilde{\mathbf{X}} & \tilde{\mathbf{x}}\\ \tilde{\mathbf{x}}^* & \alpha\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{U}}^*\\ \tilde{\mathbf{v}}^*\end{bmatrix}=\tilde{\mathbf{U}}\tilde{\mathbf{X}}\tilde{\mathbf{U}}^*+\tilde{\mathbf{v}}\tilde{\mathbf{x}}^*\tilde{\mathbf{U}}^*+\tilde{\mathbf{U}}\tilde{\mathbf{x}}\tilde{\mathbf{v}}^*+\alpha\tilde{\mathbf{v}}\tilde{\mathbf{v}}^*$$

$$\Rightarrow\quad \|\mathbf{X}^{\text{opt}}-\alpha\tilde{\mathbf{v}}\tilde{\mathbf{v}}^*\|_F^2=\|\tilde{\mathbf{X}}\|_F^2+2\|\tilde{\mathbf{x}}\|_2^2\le\frac{4\rho f_{\text{WLAV}}(\boldsymbol{\eta})}{\lambda}\operatorname{Tr}(\mathbf{X}^{\text{opt}})$$

Roundtable Discussion: NSF-TRIPODS Grant

Data science and machine learning activities at UCSC (just a sample):

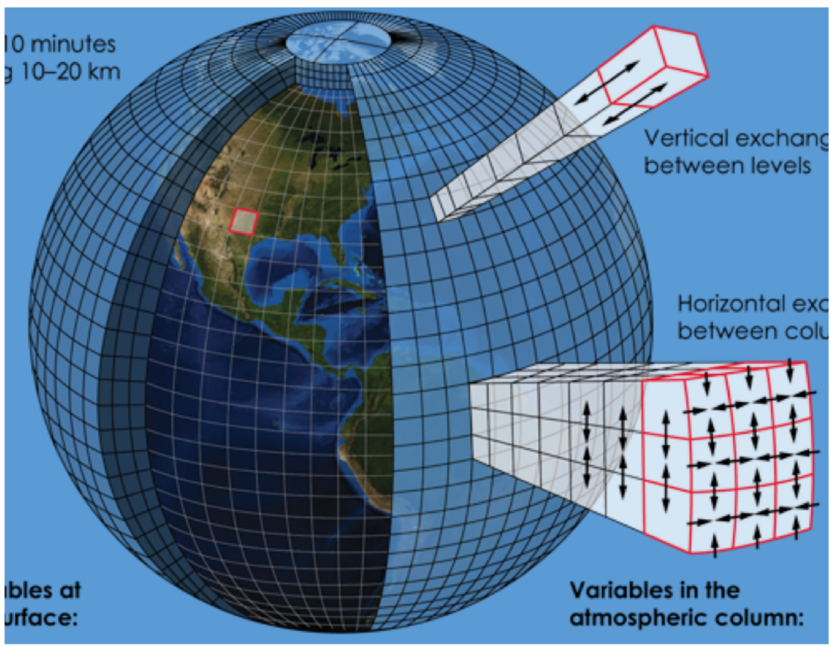


Energy Data Analytics

Figure source: B. P. Bhattarai *et al*., "Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions," in *IET Smart Grid*, 2019.

Climate Change

Elliott Campbell, Dept. Environmental Studies, UCSC
Climate-carbon feedbacks using atmospheric computer simulations



- Biosphere feedback on climate (Campbell et al., *Nature*, 2017)
- High performance computing at DOE/NERSC & NASA/JPL

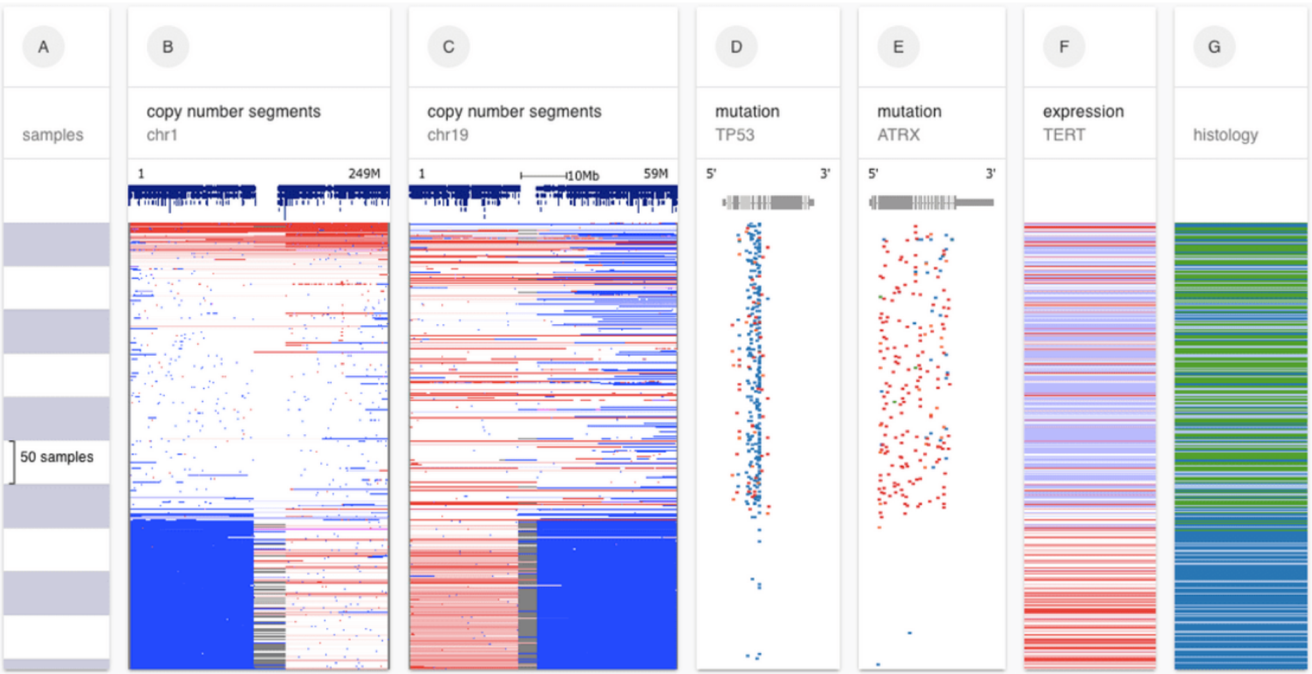
Climate Change (cont’d)

Biology

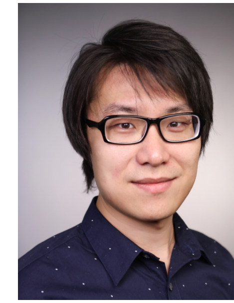
UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

Welcome to the next generation UCSC Cancer Browser: UCSC Xena!

UCSC Xena has all the same functionality of the UCSC Cancer Browser plus new tools, such as the ability to see multiple types/modes of genomic data side-by-side, and plenty of new data, such as the latest from the GDC, GTEx and more.



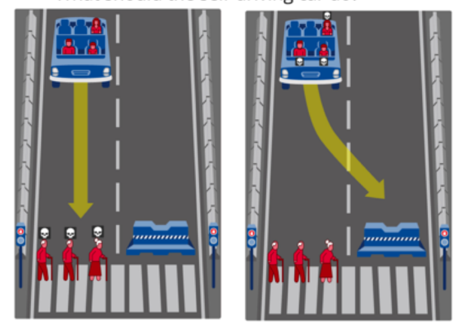
Ethical AI

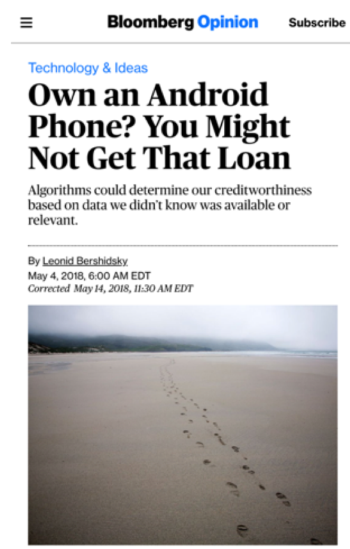


Fair Algorithms – Yang Liu at CSE of UCSC

- **Ethical AI**
  - How do fairness definition fare [AIES'19]
- **Fairness in Machine Learning**
  - Fairness definition [FATML'17, AAAI'19]
  - Actionable Recourse [FAT*'19]
  - Preference in ML [ICML'19]
- **Fairness in sequential decision making**
  - Fair stopping rule [UAI'17]
  - Fairness in bandit [FATML'17, Delayed Impact of Actions in MAB]

What should the self-driving car do?



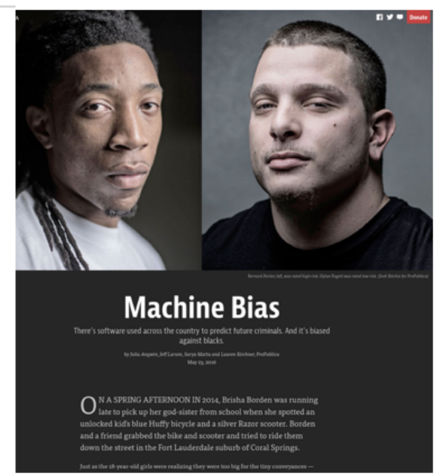


Technology & Ideas

Own an Android Phone? You Might Not Get That Loan

Algorithms could determine our creditworthiness based on data we didn't know was available or relevant.

By Leonid Bershidsky
May 4, 2018, 6:00 AM EDT
Corrected May 14, 2018, 11:30 AM EDT


Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances —

BUSINESS NEWS   OCTOBER 9, 2018 / 11:12 PM / 2 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Papers available here: http://www.yliuu.com/