Monitoring and Learning Algorithms for Future Power Systems

Georgios B. Giannakis

Acknowledgments: IREE grant; Y. Zhang; and Drs. N. Gatsis, E. Dall’Anese, V. Kekatos, S.-J. Kim, G. Mateos, and H. Zhu
Smart Grid: Advanced infrastructure and information technologies to enhance the current electrical power network

- controllable
- resilient
- efficient
- participation
- sustainable
- self-healing
- green
- situational awareness

Enabling technology advances

- Optimization, learning, and signal processing toolbox
- Sensing/metering
- Distributed generation
- Micro-grids
- Renewables
- Electric vehicles
- Power electronics
- Communication networks
- Demand response
Roadmap

- **Power system state estimation (PSSE)**
  - Semi-definite relaxation (SDR) - nonlinear PSSE
  - Decentralized SDR – linear(ized) PSSE
  - Bad data processing and breaker verification

- **Learning loads and consumer elasticity**
  - Load forecasting via low-rank/matrix factorization
  - Learning dynamic consumer elasticity

- **Other directions and the road ahead**
Grid monitoring for FPS

- PSSE for interconnected power systems
e.g., [Gomez-Exposito et al’11]
  - “Workhorse” of grid monitoring;
    - (non)linear estimation task
  - Typically centralized; unreliable telemetry
  - Sequential/hierarchical [Schweppe’70], [Zhao’05]
  - Distributed [Falcao’95], [Baldick’00], [Conejo’07]

Desiderata: efficient (non) linear (D-)SEs to cope with the large-scale and fast-varying states
which challenge the approximate iteratively linear(ized) state estimators
Centralized PSSE

- Power grid with $K$ areas $\mathcal{N} = \bigcup_{k=1}^{K} \mathcal{N}_k$
  
  wish to find bus voltages $v := [V_1, \ldots, V_N]^T \in \mathbb{C}^N$

- Model (quadratic in $v$)
  
  $z_k^\ell = h_k^\ell(v) + \epsilon_k^\ell, \quad \forall k, \ell$

- Optimal least-squares (LS) estimator

$$\hat{v} := \text{arg min}_v \sum_{k=1}^{K} \sum_{\ell=1}^{M_k} [z_k^\ell - h_k^\ell(v)]^2 \quad \text{(C-SE)}$$

- Gauss-Newton iterative solvers via linearization, e.g., [Abur-Exposito’04]
  
  sensitive to initialization (esp. w/ fast-varying states): convergence?
**Convexification via SDR**

- **Trick**: make $z_k^\ell$ linear in $V := vv^H$

\[
z_k^\ell = h_k^\ell(v) + \epsilon_k^\ell = \text{Tr}(H_k^\ell V) + \epsilon_k^\ell
\]

\[
\hat{V} := \arg \min_V \sum_{k=1}^K \sum_{\ell=1}^{M_k} [z_k^\ell - \text{Tr}(H_k^\ell V)]^2
\]

s.to $V \succeq 0$, and $\text{rank}(V) \leq 1$ \hspace{1cm} (C-SDP)

- SDR for SE [Zhu-GG’11] for SE; SDR for OPF [Bai etal’08], [Lavaei-Low’11]
  - Polynomial complexity; (near-)optimal regardless of initialization

**Desiderata:** Decentralized SDR scalable with control area size, privacy-preserving, and solvable at affordable communication cost

Cost decomposition

- Include tie-line buses to split local LS cost per $\mathcal{N}_{(k)}$

$$f_k(\mathbf{V}_{(k)}) := \sum_{\ell=1}^{M_k} \left[ z_k^\ell - \text{Tr}(\mathbf{H}_{(k)}^\ell \mathbf{V}_{(k)}) \right]^2$$

$$\hat{\mathbf{V}} := \arg \min_{\mathbf{V}} \sum_{k=1}^{K} \sum_{\ell=1}^{M_k} \left[ z_k^\ell - \text{Tr}(\mathbf{H}_{k}^\ell \mathbf{V}) \right]^2 \quad \text{s.to} \quad \mathbf{V} \succeq 0$$

$$\hat{\mathbf{V}} := \arg \min_{\mathbf{V}} \sum_{k=1}^{K} f_k(\mathbf{V}_{(k)}) \quad \text{s.to} \quad \mathbf{V} \succeq 0$$

**Challenge:** as $\{\mathcal{N}_{(k)}\}$ overlap partially, PSD constraint couples $\{\mathbf{V}_{(k)}\}$

**Blessing:** overlap $\rightarrow$ global; no overlap: $\mathbf{V} \succeq 0 \iff \mathbf{V}_{(k)} \succeq 0$, $\forall k$
Decentralized SDR for SE

- Under conditions (graph of per-area-state-overlaps is a tree)

\[ \hat{V} := \arg \min_V \sum_{k=1}^{K} f_k(V_k) \quad \text{(C-SDP)} \]

\[ \{\hat{V}_k\} := \arg \min_{\{V_k\}} \sum_{k=1}^{K} f_k(V_k) \]

s.t. \[ V_k \succeq 0, \quad V_j^k = V^k_j \]

- ADMM [Glowinski-Marrocco’75]; for D-Estimation [Schizas-Giannakis’07]
  - Iterates between local variables and multipliers per equality constraint

\[ \{\Lambda_{k,j}(i)\}_{j \in A_k} \rightarrow V_k(i) \rightarrow \{V_j^k(i)\}_{j \in A_k} \rightarrow \{\Lambda_{k,j}(i+1)\}_{j \in A_k} \]

- Converges \[ V_k(i) \rightarrow \hat{V}_k \] even for noisy-async. links [Schizas-GG’08], [Zhu-GG’09]
ADMM convergence in action

- IEEE 14-bus grid with 4 areas; 5 meters on tie-lines

- Errors $\| V_{(k)}^{(i)} - \hat{V}_{(k)} \|_F$ vanish asymptotically
118-bus test case

- Triangular configuration [Min-Abur’06]
- Power flow meters on all tie lines except for (23,24) tree communication graph

Local norms

\[ \| v^{(k)}_i - v^{(k)} \|_2 \]

converge in only 20 iterations!
Decentralized PSSE for linear models

- Local linear(ized) model  \( z_k = H_k x_k + n_k \)

- Regional PSSEs

\[
\min_{x_k \in \mathcal{X}_k} f_k(x_k)
\]

- Coupled local problems

\[
\min_{\{x_k\}} \sum_{k=1}^{K} f_k(x_k) \\
\text{s.t. } x_k[l] = x_l[k]
\]

**S1.** \( x_k^{t+1} \leftarrow \arg \min_{x_k \in \mathcal{X}_k} f_k(x_k) + \frac{c}{2} \sum_{i \in \mathcal{N}_k} (x_k(i) - p_k^t(i))^2 \)

**S2.** \( p_k^{t+1}(i) \leftarrow p_k^t(i) + \left( x_i^{t+1}[i] - \frac{x_k^t(i) + x_l^t[i]}{2} \right) \)

- ADMM solver: convergent with minimal exchanges and privacy-preserving
Simulated test

S1.  \( x_{k}^{t+1} \leftarrow (H_{k}^{T}H_{k} + c \cdot D_{k})^{-1} (H_{k}^{T}z_{k} + c \cdot D_{k}p_{k}^{t}) \)

S2.  \( p_{k}^{t+1}(i) \leftarrow p_{k}^{t}(i) + \left( x_{l}^{t+1}[i] - \frac{x_{k}^{t}(i) + x_{l}^{t}[i]}{2} \right) \)

Decentralized bad data processing

\[ z = Hx + n + o \]

- Reveal *single* and *block* outliers via

\[
f(x) := \min_\theta \frac{1}{2}\|z - Hx - \theta\|^2_2 + \lambda \|	heta\|_1 = \sum_{m=1}^{M} h(z_m - h_m^T x)\]

**S1.** \( x_k^{t+1} \leftarrow (H_k^T H_k + c \cdot D_k)^{-1} (H_k^T (z_k - o_k^{t}) + c \cdot D_k p_k^{t}) \)

**S2.** \( o_k^{t+1} \leftarrow [z_k - H_k x_k^{t+1}]^+ \)

**S3.** \( p_k^{t+1}(i) \leftarrow p_k^t(i) + \left( x_l^{t+1}[i] - \frac{x_k^t(i) + x_l^t[i]}{2} \right) \)
D-PSSE on a 4,200-bus grid
Football and circuit breakers ...
**Generalized PSSE**

- **Usage:** protection and reconfiguration

- Account for CB status (C for instrumented and B_m for un-instrumented)

\[
\min_x \|z - Hx\|_2^2 + \lambda \sum_{m \in S} \|B_m x\|_2
\]

s.t.  \( Cx = 0 \)

Numerical tests

Generalized state estimator
Novel method

CB status errors

Execution time [secs]

MSE

No. of un-instrumented CBs
Take-home messages...

- SDR-based PSSE for interconnected regional operators
  - ADMM decentralized solver with privacy and minimal overhead
  - Chordal graph renders D-SDR equivalent to near-optimal C-SDR

- Comprehensive FPS monitoring
  - Provably convergent for linear(ized) models at affordable complexity
  - Exploit sparsity for bad data processing and breaker verification

- Fruitful research directions
  - Efficient SDR implementation for large-scale power systems; online?
  - D-SDR framework for additional power system optimization tasks
  - Cyber-attacks; nonlinear model observability; outage and topology ID?
  - Big data analytics; and grid clustering
Roadmap

- **Power system state estimation (PSSE)**
  - Semi-definite relaxation (SDR) - nonlinear PSSE
  - Distributed SDR – linear/linearized PSSE
  - Bad data processing and breaker verification

- **Learning loads and consumer elasticity**
  - Load forecasting via low-rank/matrix factorization
  - Learning dynamic consumer elasticity

- Other directions and the road ahead …
Load forecasting

- Essential for economic operation of power systems
  - Economic dispatch, OPF, unit commitment (~hour)
  - Reliability assurance and hydrothermal coordination (~week)
  - Strategic generation and transmission planning (~year)

- Prior art: (non)linear time-series models (ARMA, ARIMA, ARIMAX) [Shahidehpour et al’02]

- Challenges: spatio-temporal structures; load volatility due to e.g., PHEVs

- Problem: given load measurements at $M$ sites and $N$ time slots
  - Predict load at sites/times that data are not available; impute past; forecast future demand
Low-rank plus sparse non-negative factors

- Load matrix comprises low-rank and sparse non-negative components

\[ X \sim L + AB^T \]

- \( L \), \( A \geq 0, B \geq 0 \)

\[
\min_{L,A \geq 0, B \geq 0} \frac{1}{2} \| P_\Omega (X - L - AB^T) \|_F^2 + \lambda \| L \|_* + \mu_1 (\| A \|_1 + \| B \|_1)
\]

- Low-rank \( L \) due to periodicities (daily, weekly, monthly), and a few latent factors (user preference, temperature)
- Non-negative matrix factorization (NMF) \( AB^T \) captures load clusters

- Identifiability issues
  - \( X \sim L + S \) (\( L \): low-rank; \( S \): sparse) [Candes et al’11], [Wright’13]
  - \( X \sim L + CS \) (\( L \): low-rank; \( S \): sparse; \( C \): given) [Mardani et al’13]
  - Identifiability of our model is plausible but yet to be established
Load inference algorithm

- Kernelized formulation allows extrapolation [Bazerque-Giannakis’13]
  - Equivalent formulation \( P \in \mathbb{R}^{M \times r}, Q \in \mathbb{R}^{N \times r}, \text{rank}(L) \leq r \)
    \[
    \min_{P,Q,A \geq 0,B \geq 0} \frac{1}{2} \| \mathcal{P}_\Omega(X - PQ^T - AB^T) \|_F^2 + \frac{\lambda}{2}(\|P\|_F^2 + \|Q\|_F^2) + \mu_1(\|A\|_1 + \|B\|_1),
    \]
  - Solved via block coordinate descent (BCD); closed form per iteration
    \[
    \min_{P,Q,A \geq 0,B \geq 0} \frac{1}{2} \| \mathcal{P}_\Omega(X - PQ^T - AB^T) \|_F^2 \\
    + \frac{\lambda}{2} \left[ \text{tr}(P^TR_p^{-1}P) + \text{tr}(Q^TR_q^{-1}Q) \right] \\
    + \mu_1(\|A\|_1 + \|B\|_1) + \frac{\mu_2}{2} \left[ \text{tr}(A^TR_a^{-1}A) + \text{tr}(B^TR_b^{-1}B) \right]
    \]
  - \( R_p, R_q, R_a, R_b \) : positive-definite sample-averaged covariances (kernels)

Test with real data

- $M = 17$ sites,
- $N = 4 \times 7 \times 24$ (4 weeks)
- $r = \rho = 5$
- Forecast the last 24 hours (RMSE = 0.14)

---

Real-time pricing for DR

- Adapt load schedules based on load prices

**Issues:** Privacy, robustness, real-time, consumer participation

**Goal:** smart real-time pricing by learning consumer preferences

- Adjust energy price in real-time to shape load
- Set prices differently for individual customers
- Load-price elasticity changes across consumers and time

**Challenge:** Learn elasticity with minimal “modeling”
Problem formulation

- **Model**
  - \( p_k^t \): price adjustment for customer \( k \) at time slot \( t \)
  - \( l^t \): load level at slot \( t \) without price adjustment
  - \( \theta_k^t \): elasticity of consumer \( k \) at slot \( t \)
  - \( d_k^t \): load adjustment of customer \( k \) due to price adjustment \( p_k \)

\[
\begin{align*}
  d_k^t &= -\theta_k^t p_k^t \\
  \theta^t &= [\theta_1^t, \ldots, \theta_K^t]^T \\
  l^t &= l^t - \sum_{k} d_k^t = l^t - \theta^T p^t
\end{align*}
\]

- **Goal:** minimize load variance

\[
\frac{1}{2} \sum_{t=1}^{T} \left( l^t - \theta^T p^t - m^t \right)^2
\]

- **Promote sparsity and fairness**

\[
\begin{align*}
  c^t(p^t) &= \frac{1}{2} \left( l^t - \theta^T p^t - m^t \right)^2 + \lambda \| p^t \|_1 + \frac{\mu}{2} \| p^t \|_2^2 \\
  &= \phi^t(p^t) + r(p^t)
\end{align*}
\]
Online convex optimization

**OCO framework:** game between a player and an adversary

- At each time slot $t = 1, 2, \ldots, T$
- Utility (player) chooses $p^t$
- Customers ("adversaries") choose $l^t$ and $\theta^t \rightarrow c^t(\cdot)$
- Player suffers loss $c^t(p^t)$ and receives feedback $F^t$

**OCO goal:** achieve sublinear regret

$$R_c(T) := \sum_{t=1}^{T} c^t(p^t) - \min_{p \in \mathcal{P}} \sum_{t=1}^{T} c^t(p)$$

with $R_c(T)/T \rightarrow 0$ as $T \rightarrow \infty$
Two types of feedback

- Full feedback
  - $F^t = c^t(\cdot)$
  - Utility obtains $l^t$ and $\{d^t_k\}_{k=1}^K$ at the end of slot $t$ ($\hat{\theta}^t_k = -d^t_k/(p^t_k + \varepsilon)$)

- Partial (bandit) feedback
  - $F^t = c^t(p^t)$
  - Utility observes only $l^t_a$ at the end of slot $t$
Algorithms

- Full feedback case
  - Composite objective mirror descent (COMID) [Duchi et al.’10]
    \[
    p^{t+1} = \arg\min_{p \in P} \left[ -\eta (l^t - \theta^t p - m^t)^T \theta^t p + \frac{1}{2} \| p - p^t \|_2^2 + \eta \left( \lambda \| p \|_1 + \frac{\mu}{2} \| p \|_2^2 \right) \right]
    \]
    $\nabla \phi^t(p^t)$
    \(\eta\): step size parameter
  - Provably achieves $O(\sqrt{T})$ regret bound

- Partial feedback case
  - Need random sampling to estimate gradient of $c^t(\cdot)$
  - Our algorithm enjoys $O(T^{3/4})$ regret bound [Kim-Giannakis’ISGT-14]
Numerical tests

Load before and after real-time DR

Price adjustment (full information)

Load curves (bandit information)
Take-home messages...

- Load forecasting via low rank plus sparse non-negative factorizations
  - Joint prediction and imputation across space and time
  - Machine learning tools for regular patterns and constituent load signatures

- Real-time pricing for demand response
  - Online learning of consumer preferences
  - Online convex optimization enjoys sub-linear regrets

- Fruitful research directions
  - OCO framework for additional power system optimization tasks?
  - Big data analytics (anomalies, classification and clustering)

Thank you!