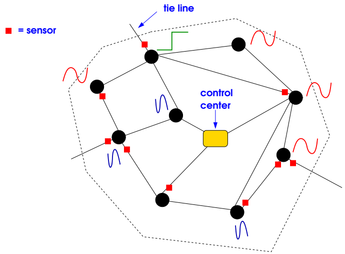
Real-time control of network physical structures to bypass complexity: Optimization, Stochastics and Structure Recognition

D. Bienstock, J. Blanchet, V. Goyal and G. Iyengar

Columbia University, Stanford University

July 2018

Real-time control of networked structures governed by physics



- Today: control enforces separation by time domain
  e.g. in power grids: governor reaction (10^{-3} sec), AGC (sec), OPF (mins)
- Opportunity: fast sensors, algorithms
  Challenges: "smart" loads, complex noise
- **Research Goals:**
  - Avoid separation
  - Quickly recognize system structure. **Time frame: seconds or less**
  - Quickly detect intrusion. **Time frame: seconds or very few minutes**

Challenges:

- How do we solve in near-real time hard problems that we cannot solve offline? E.g. nonconvex, polynomial optimization problems

  $\rightarrow$ To do: warm restart of ADMM-like methods for bilinear optimization.

- How do we handle noise/structure that we do not really understand?

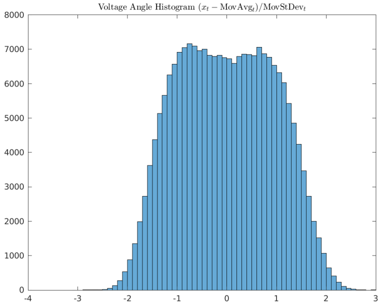
  $\rightarrow$ Now doing: learning real-time correlation (or covariance) from noisy inputs. (NIPS Time series workshop 2017+)

- How do we combine first-order optimization with poorly understood "noise"?

  $\rightarrow$ Now doing: **Variance-aware** first-order optimization. (PSCC 2018)

Noise is not just noise

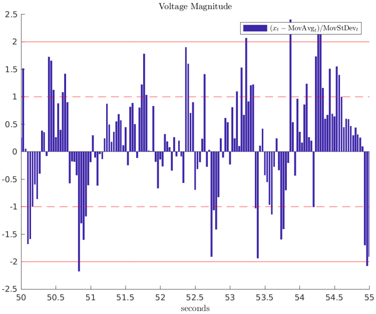
(We have **28 TB** of real data)
Voltage angle deviation histogram



Kolmogorov-Smirnoff gaussianity test strongly rejected, always

Noise is not just noise

From real time series, voltage magnitude deviations



Strong and nontrivial autocorrelation structure

Concrete problem: learning covariance in real time

- **PCA:** principal component analysis.
  - Covariance of real-world data usually has low rank.
  - Fast PCA methods approximately capture the leading modes.
- **Streaming** PCA:
  - Old data gets stale
  - Cannot hold a lot of data
- **Non-stationary** regime for streaming PCA
  - Goal: detect change
  - Research question: what are fundamental computational limits?

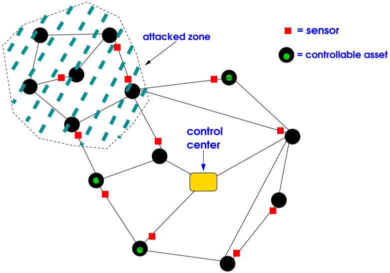
Tecnical! NIPS TSW 2017

with PhD student Apurv Shukla plus S. Kim (ex-LANL)

A streaming algorithm to identify PCA structure within time window

- Generative Model
  - Non-Stationary Spiked Covariance Model
- Sample Complexity:
  - Lower bound relating recovery error to number of samples
- Algorithm:
  - Two-phase iterative algorithm
    - ★ Phase-I: Iterative Eigenvector Computation
    - ★ Phase-II: Matrix Sketching
- **Theorems:** See NIPS paper and forthcoming paper

Application! Detecting intrusion through random physics



- Atttacked zone is unknown to control center
- Attacker causes physical damage and alters sensor signals
- Defense:
  - Use controllable assets to alter **covariance** of physics
  - Changes unpredictable to attacker
  - Attacker (if aware of defense) can learn variance
  - But that takes time and sensor stream is continuous
  - So defender can **learn** the true covariance matrix

Distributionally Robust Optimization

Data-Driven Distributionally Robust Optimization

$$\min_{\theta} \sup_{P:D_c(P,P_n)\leq\delta} E_P\left(h\left(\theta, X\right)\right) \leftarrow \text{game formulation}$$

$$P_n\left(dx\right) = \frac{1}{n}\sum_{k=1}^{n}\delta_{\{X_k\}}\left(dx\right) = \text{Empirical Data}$$

$$E_{P_n}\left[h\left(\theta, X\right)\right] = \frac{1}{n}\sum_{k=1}^{n} h\left(\theta, X_k\right).$$

In words: Select the best response to model perturbations around the data (need to specify D_c).

Distributionally Robust Optimization (DRO)

- Extensive literature on DRO (Scarf (1958), Dupuis, James, Petersen (2000), Hansen and Sargent (2001), Ben-Tal, El Ghaoui, Nemirovski (2009), Delange & Ye (2010),...).
- Typical choices of $D(P, P_n)$

$$D(P, P_n) = E_P\left(\log\left(\frac{dP}{dP_n}\right)\right).$$

- Problem in data-driven setting: must preserve absolute continuity with respect to P_n.
- Choose $D(\cdot)$ based on optimal transport instead of divergence.
  $\longrightarrow$ Works in practice and recovers exactly many machine learning estimators (e.g. Lasso, SVMs, adaptive ridge etc.)

Optimal Transport Metric and Wasserstein Distances

- Definition of Optimal Transport Discrepancy:

$$D_c\left(P, Q\right) = \min\{E_\pi\left(c\left(X, Y\right)\right) : \pi_X = P, \pi_Y = Q\}.$$

- Algorithmically $D_c\left(\cdot\right)$ is obtained by solving an LP

$$\min \sum_{x,y} c\left(x, y\right) \pi\left(x, y\right) \quad \text{subject to}$$

$$\sum_{y} \pi\left(x, y\right) = P\left(x\right), \sum_{x} \pi\left(x, y\right) = Q\left(y\right)$$

$$\pi\left(x, y\right) \geq 0 \text{ for all } x, y .$$

- Formulation includes Wasserstein and earth-mover's distance.

Applications of Data-Driven DRO

- We briefly explain fundamental connections to machine learning.
- Consider linear regression: Estimate $\beta_* \in R^m$ for model

$$Y_i = \beta_* X_i + e_i,$$

  where $\{(Y_i, X_i)\}_{i=1}^n$ is a set of data points.
- Optimal Least Squares approach: Estimate β_* via

$$\min_\beta MSE(\beta) := \min_\beta n^{-1} \sum_{k=1}^{n} \left(Y_k - \beta^T X_k \right)^2$$

- We now apply the DRO formulation via optimal transport...

Connection to Sqrt-Lasso

Theorem (B., Kang, Murthy (2017)) Suppose that

$$c\left((x,y),(x',y')\right) = \begin{cases} \|x-x'\|_q^2 & \text{if} \quad y = y' \\ \infty & \text{if} \quad y \neq y' \end{cases}.$$

Then,

$$\max_{P:D_c(P,P_n)\leq\delta} E_P^{1/2}\left(\left(Y-\beta^T X\right)^2\right) = \sqrt{MSE\left(\beta\right)} + \sqrt{\delta}\,\|\beta\|_p.$$

Remark: This form of Lasso is called sqrt-Lasso (Belloni et al. (2011)).

Enhance Out-of-Sample Performance

- More general adversaries $\rightarrow$ better decisions!
- Intuitive choice:

$$\textbf{Generalized Mahalanobis}: c\left(x,y\right) = \left(y-x\right)^{T} A\left(x\right)\left(y-x\right),$$
$$A\left(\cdot\right) \text{ positive definite.}$$

- We assume $\ell\left(\cdot\right)$ is a convex loss function to be minimized.
- Affine decision rules $\beta^{T}x$ (so empirical risk minimization insolves minimizing $E_{P_n}\left[\ell(\beta^{T}X)\right]$ over β).
- We explore distributionally robust version of this problem.

A Class of DRO Problems with Affine Policies

Theorem (B., Fan, Murthy '18)

Under natural assumptions

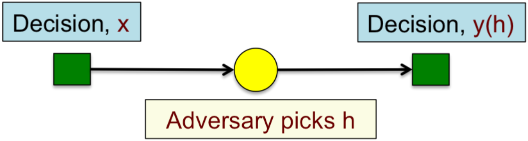
$$\inf_{\beta\in\mathcal{R}^d} \sup_{P:D_c(P,P_n)\leq\delta} E_{P_n}\left[\ell(\beta^T X)\right] = \inf_{\beta,\lambda\geq 0} E_{P_n}[f\,(\beta,\lambda,X)],$$

for a tractable $f\,(\cdot)$ *which is strongly convex in* β,λ *for* $\delta\in[0,\delta_0]$ *and some* $\delta_0>0$. *So, robust problem is not harder to solve than non-robust counterpart. Also, the worst case adversary can be computed.*

Consequence: Should robustify. Challenge: Calibrate the function $c(x, y)$.

Two-stage Adjustable Robust problem

$$
\begin{aligned}
z_{\mathrm{AR}}(\mathcal{U}) = \min\ \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{h}\in\mathcal{U}}\ \min_{\boldsymbol{y}(\boldsymbol{h})}\ & \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{h}) \\
\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{h}) \ & \geq \ \boldsymbol{h} \\
\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{h}) \ & \in \ \mathbb{R}^n_+
\end{aligned}
$$



- Demand uncertainty in unit commitment, facility location ...
- Hard to approximate within a factor better than $O(\log n/\log\log n)$ (Feige et al. 2007).

Approximate Solution Policies

- **Static robust solution**:
  - Single solution $(\boldsymbol{x}, \boldsymbol{y})$ feasible for all scenarios.
  - Easy to compute.
  - Good approximation for symmetric sets (Bertsimas, G and Sun (2011)).
  - Worst case performance bound is $\Omega(m)$.
- **Piecewise static policies**:
  - Also known as K-adaptibility policies.
  - Divide uncertainty set into pieces and a static solution for each piece.
  - Optimal pieces may be exponentially many (El Housni and G (2017))
  - Even designing small number of optimal pieces is NP-hard. (Bertsimas and Caramanis (2012)).

Approximate Solution Policies

- **Affine policy**:
  - $\boldsymbol{y}(\boldsymbol{h}) = \boldsymbol{P}\boldsymbol{h} + \boldsymbol{q}$.
  - Introduced by Ben-Tal et al. (2004)
  - Can be computed efficiently and have good empirical performance.
  - Optimal for simplex uncertainty sets
  - Tight $O(\sqrt{m})$-approximation for general sets (Bertsimas and G (2010)).

- **Piecewise affine policies**
  - Chen and Zhang (2009), Bertsimas and Georghiou (2014), Bertsimas and Dunning (2014), Postek and Den Hertog (2016), Ben-Tal, El Housni and G (2016).
  - Optimal for convex uncertainty sets (Zhen et al. (2016))
  - Hard to compute the optimal pieces that may be exponentially many.

Current Work

- Characterization of performance of affine policies for important classes of uncertainty sets
- **Budget uncertainty sets**:

$$\mathcal{U} = \left\{ \boldsymbol{h} \in [0,1]^m \;\middle|\; \sum_{i=1}^{m} w_i h_i \leq \Gamma \right\}$$

  - Very commonly used class of uncertainty sets.
  - Captures confidence interval sets and CLT sets.
  - Adjustable problem is $\Omega(\log n / \log\log n)$-hard even for these sets

- **Intersection of Budget uncertainty sets**