Data-Driven Optimization under Distributional Uncertainty

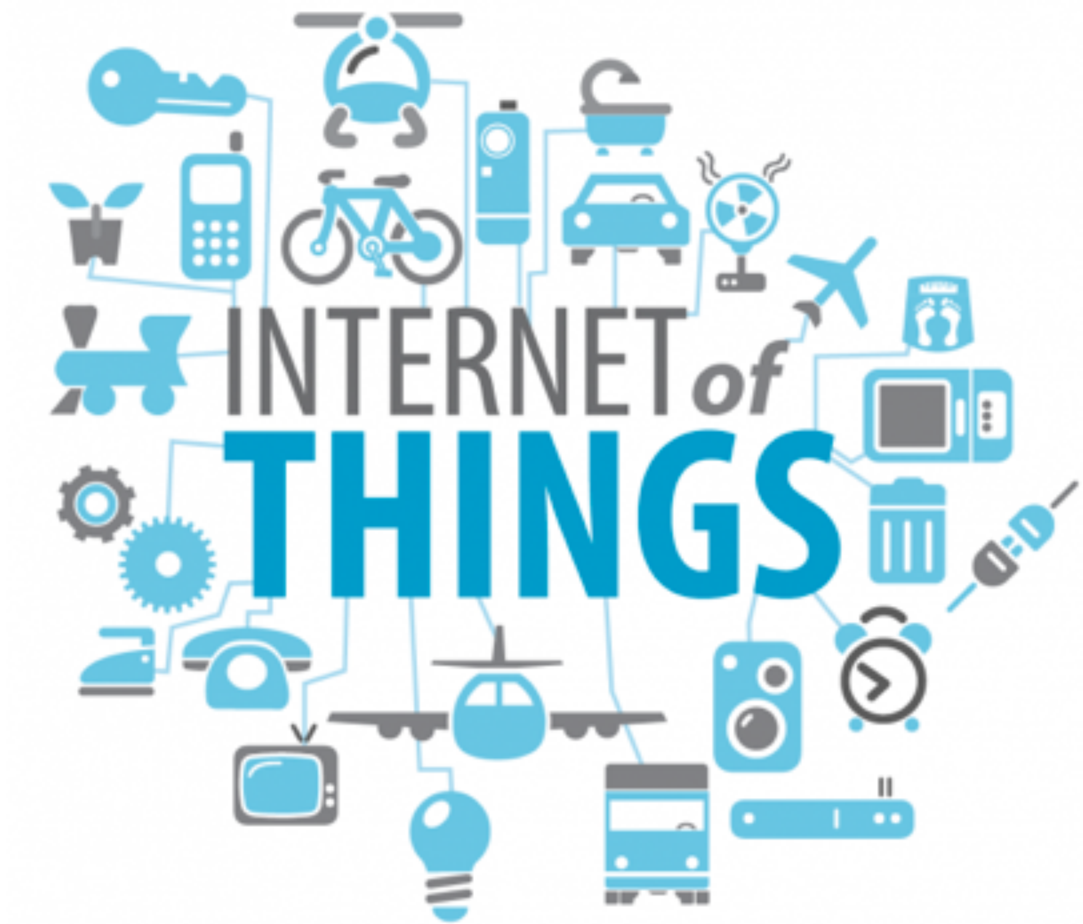
Shuo Han

Postdoctoral Researcher
Electrical and Systems Engineering
University of Pennsylvania

Penn
University *of* Pennsylvania

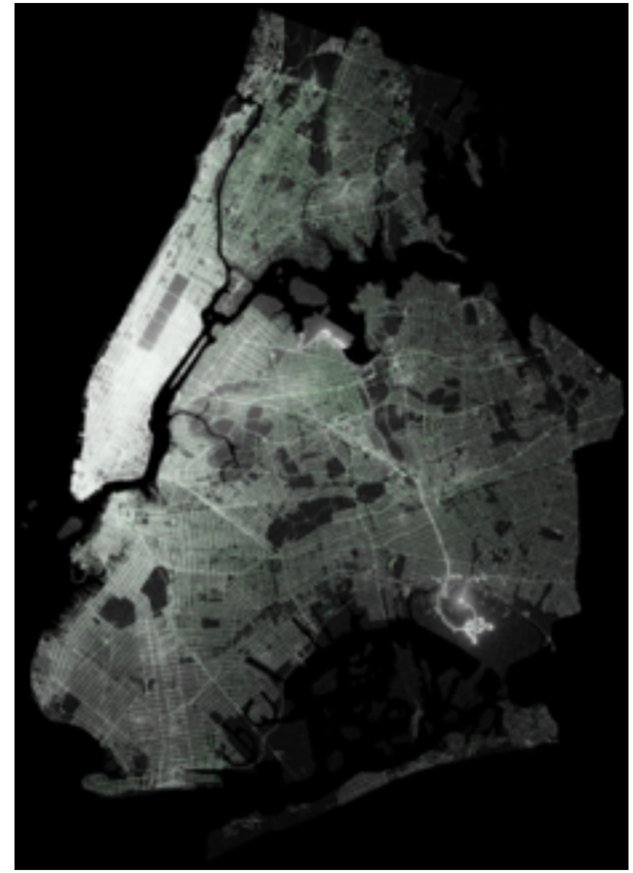
Internet of Things (IoT)

- A network of physical objects
  - Devices
  - Vehicles
  - Buildings
- Allows objects to be
  - Sensed and controlled
  - Remotely across the network
- Growing rapidly, by 2020:
  - 50 billion devices
  - 6.58 devices per person



What IoT Brings: More Sensing (and More Data)

[Source: nyc.gov]



- Total size of dataset: **267 GB**
- **1.1 billion** taxi and Uber trips (2009 - 2015)
- Pick-up and drop-off dates/times, locations, distances...

What IoT Brings: More Control

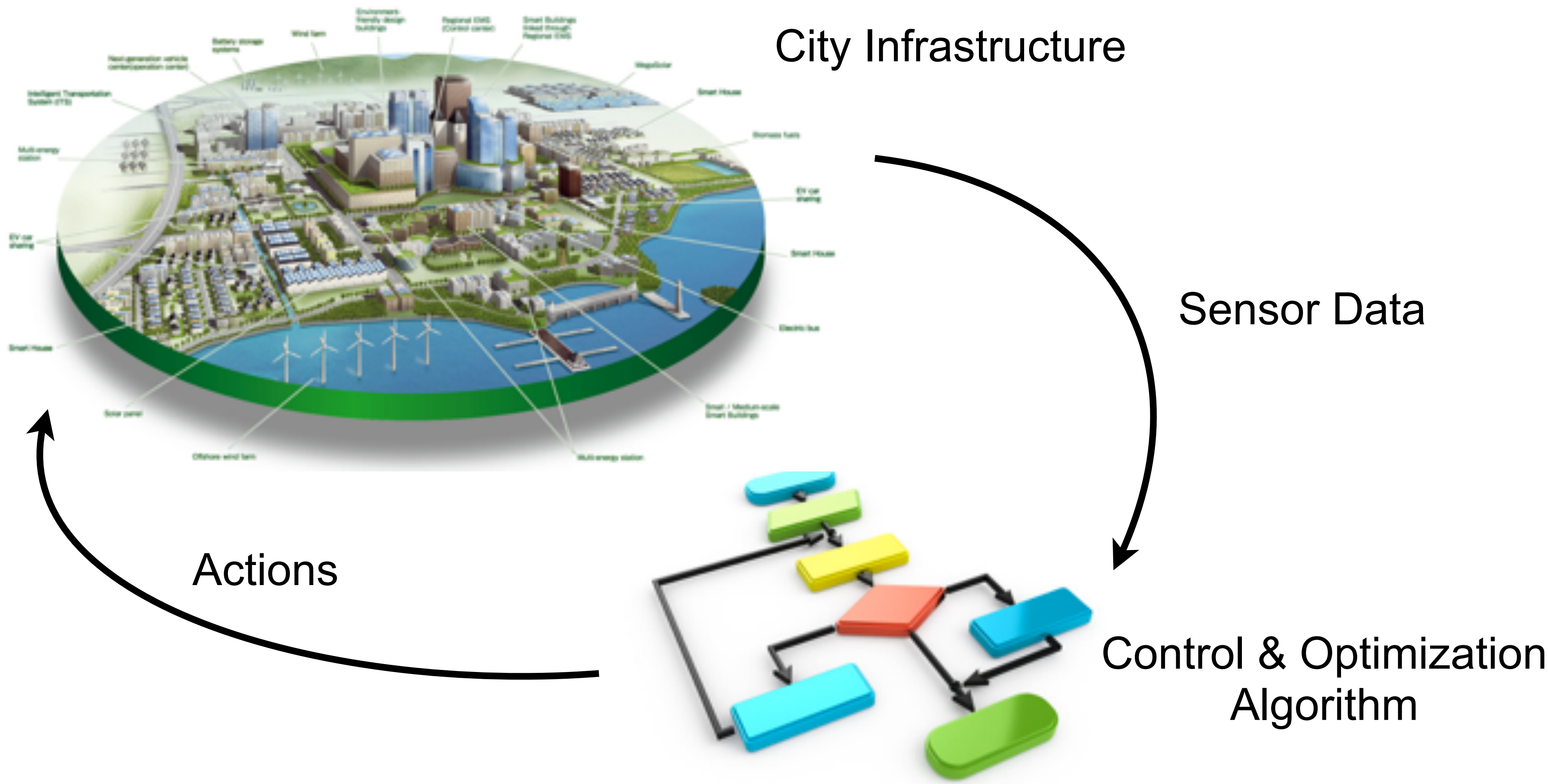
Smart Home Appliances

Connected and Autonomous Vehicles

Wireless Traffic Light Control

Smart Buildings

Smart Cities: IoT + Decision Support



Investment in Smart Cities

The Smart Cities Initiative from the White House (Sep 2015)

“... an infrastructure to continuously improve the collection, aggregation, and use of data to improve the life of their residents – by harnessing the growing data revolution, low-cost sensors, and research collaborations, and doing so securely to protect safety and privacy.”

TerraSwarm: Swarm at The Edge of The Cloud

TerraSwarm

“How should we make use of data?”

[J. Rabaey, ASPDAC’08]

“How should we send data?”

“How should we collect data?”

Research Interests

Research Topics

Stochastic Systems

Theory

Convex Optimization
Control Theory
Statistics

Applications

Energy

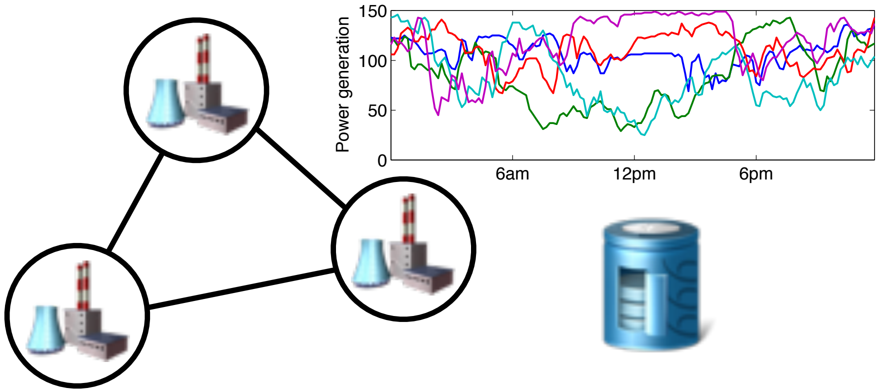
Multi-Agent Systems

Network Dynamics

Transportation

Research Overview

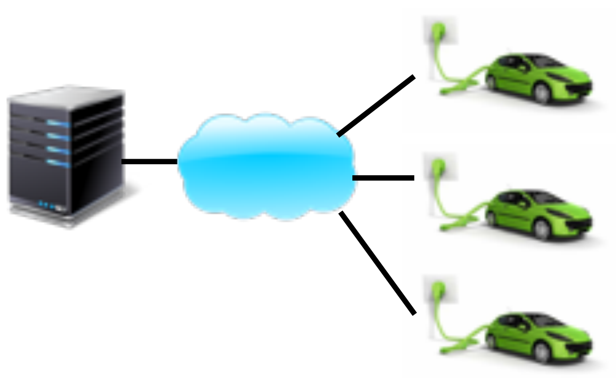
Data-Driven Optimization



[ACC13], [SIOPT15]

[CDC15], [TASE16], [ICCPS17]

Privacy Solutions for Cyber-Physical Systems



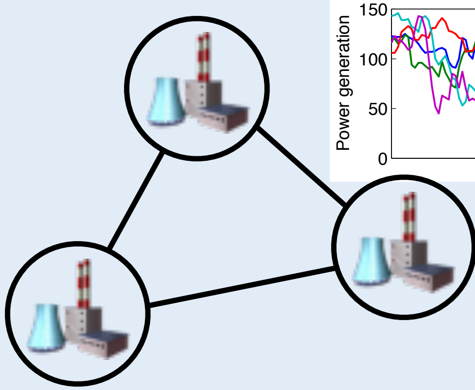
[Allerton14], [TAC16]

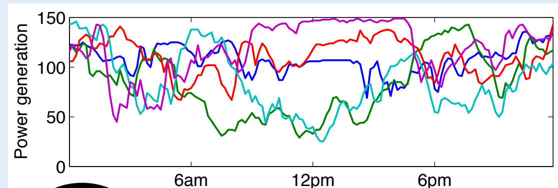
Pricing for Ridesharing

[ACC17]

Research Overview

Data-Driven Optimization

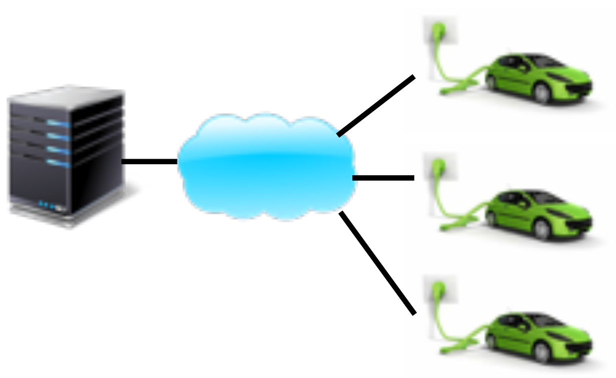




[ACC13], [SIOPT15]

[CDC15], [TASE16], [ICCPS17]

Privacy Solutions for Cyber-Physical Systems

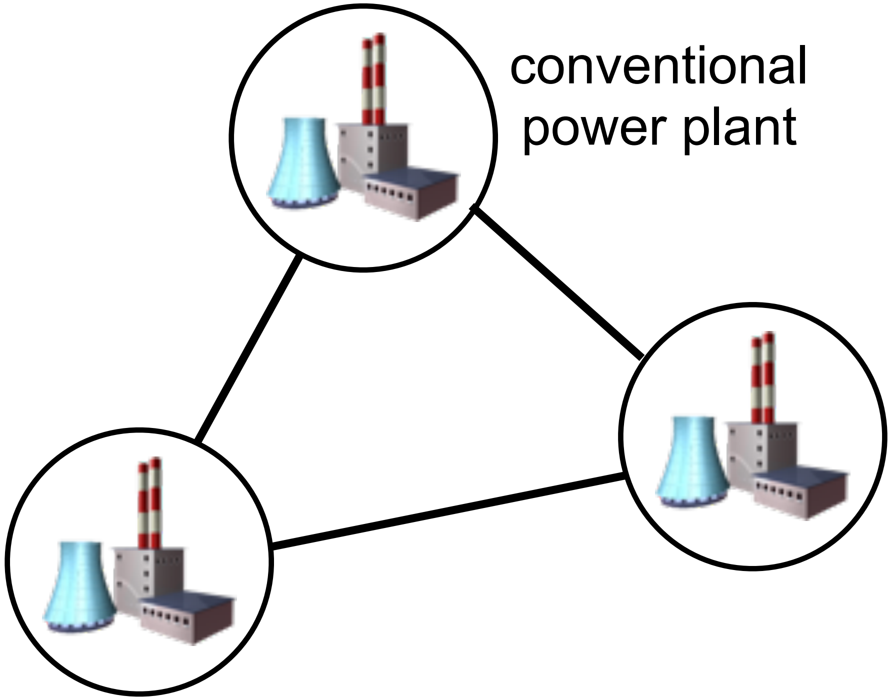


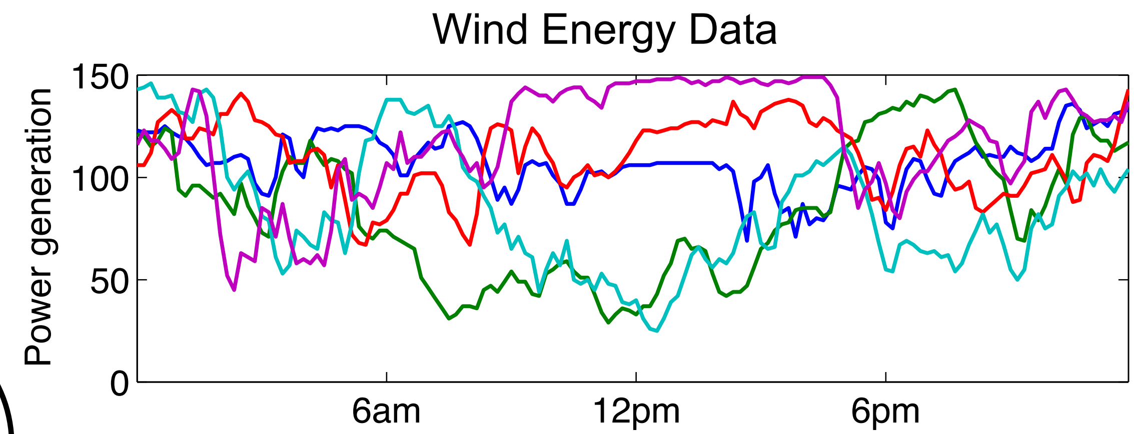
[Allerton14], [TAC16]

Pricing for Ridesharing

[ACC17]

Motivation: Wind Energy Integration





[Source]: AESO

Control Action: Allocation of energy storage

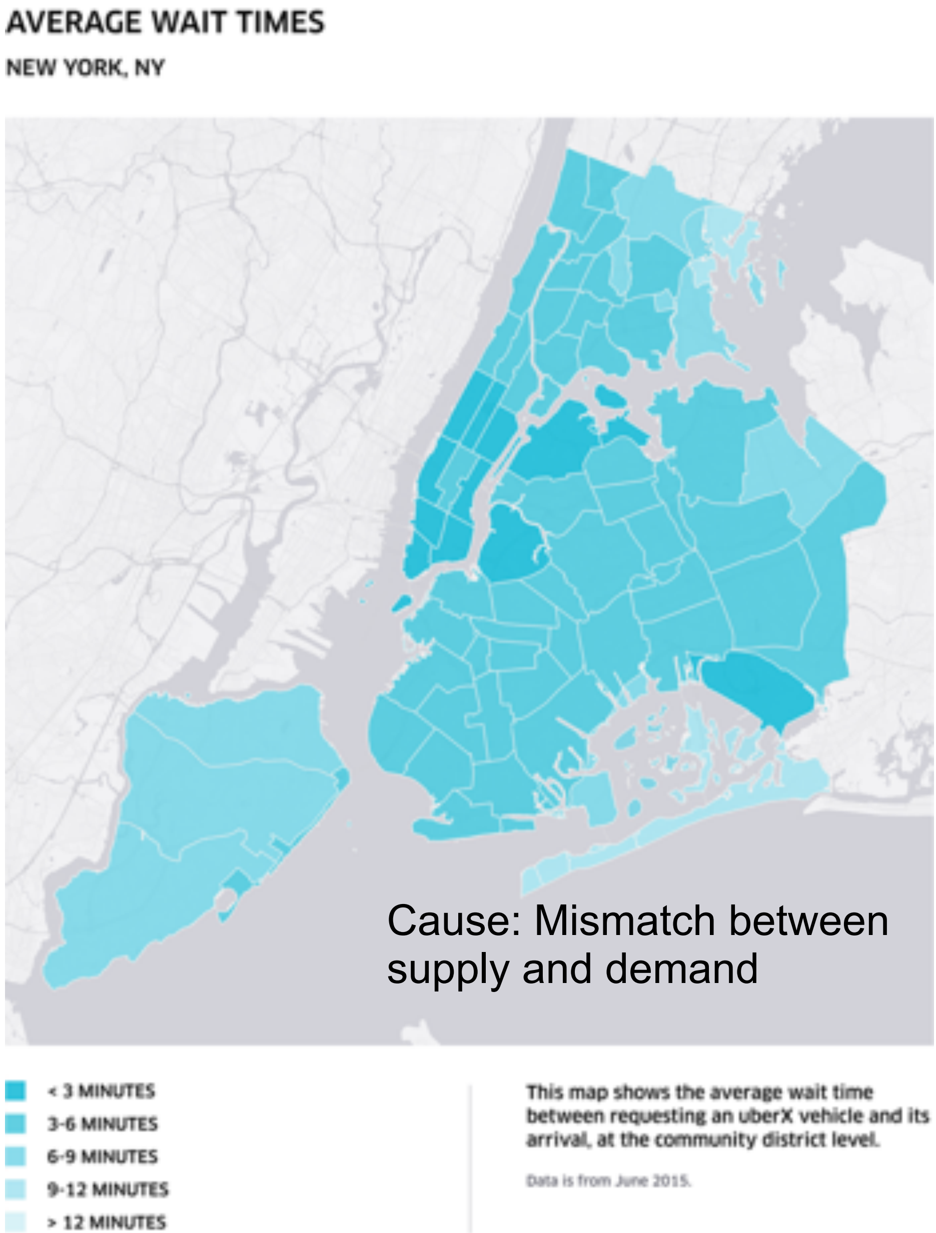
How can we make use of the wind power generation data to maximally utilize wind power?

\+

wind power

storage devices

Motivation: On-Demand Ridesharing in Cities



- Pick-up and drop-off times
- Pick-up/drop-off locations
- Travel distances

Control Action: Redistribution of empty vehicles

How can we make use of the trip data to reduce the average wait time for passengers?

Background: Stochastic Programming

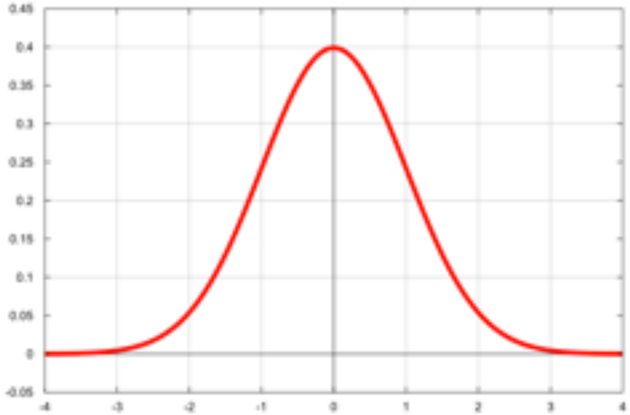
$$\underset{x}{\text{minimize}} \quad \mathbb{E}_{\theta \sim d}\left[f(x,\theta)\right]$$

Probability distribution that models the stochastic phenomenon

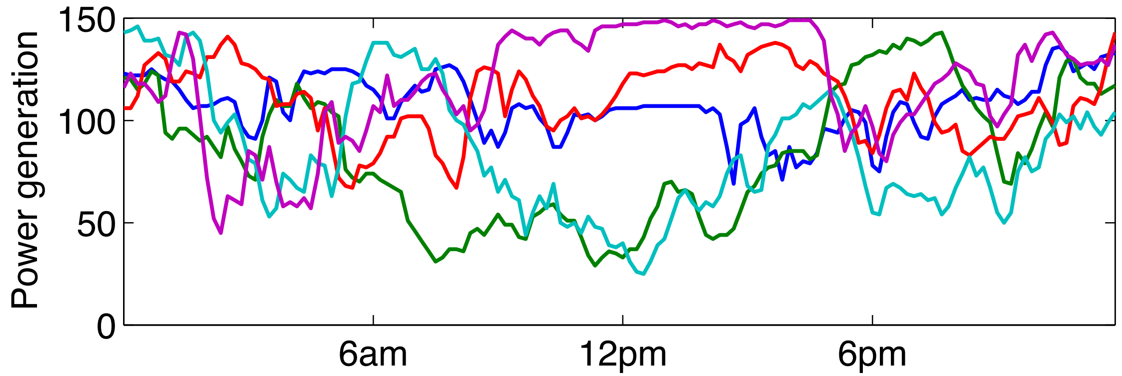
- f : Objective function
- x : Decision variable
  - Ridesharing: Redirection of empty vehicles
  - Wind power integration: Allocation of storage
- θ: Stochastic phenomenon
  - Ridesharing: Future passenger demand
  - Wind power integration: Wind power generation
- d: Probability distribution of θ

Distribution is Not Always Available

We often do not have:



Instead, we have:

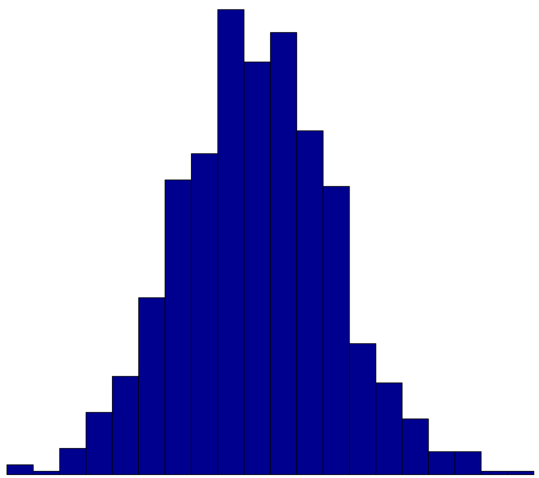


Question: How should these samples be used in a computationally tractable way with performance guarantees?

Using Sampled Data: Previous Methods

Sample average approximation

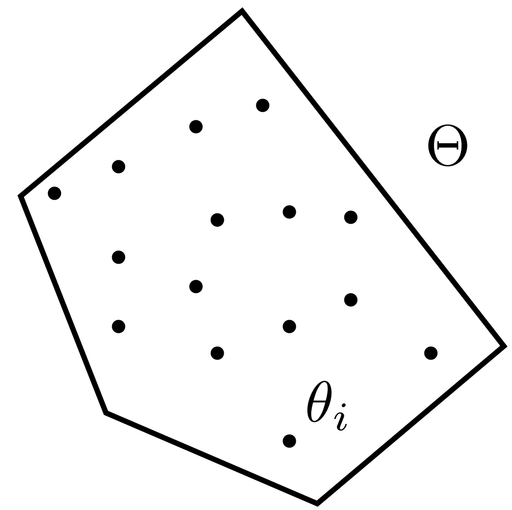
$$\underset{x}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^{n} f(x, \theta_i)$$



- Weak guarantee on performance

Robust optimization

$$\underset{x}{\text{minimize}} \quad \max_{\theta \in \Theta} f(x, \theta)$$

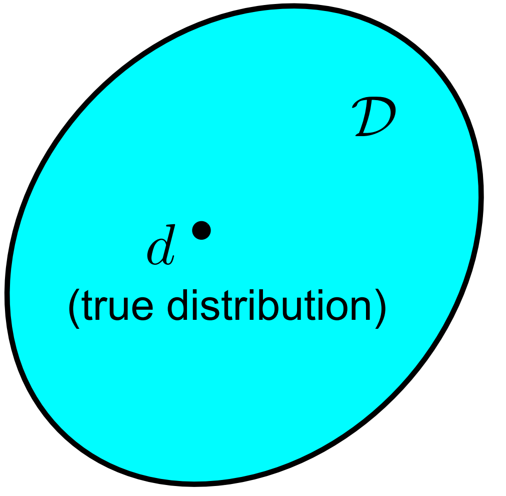


- Can be extremely conservative

Distributional Information + Uncertainty ?

Using Sampled Data: Distributional Uncertainty

- Distributional uncertainty
  - An ambiguity set in the space of probability distributions
  - **No assumption** on the type (continuous vs discrete, Gaussian, uniform, ...) of distributions
  - Contains the true distribution with high probability
  - Informally: “Uncertainty of uncertainty”



- Decision making problem: Distributionally robust optimization

$$\underset{x}{\text{minimize}} \quad \max_{d \in \mathcal{D}} \mathbb{E}_{\theta \sim d}\left[f(x, \theta)\right] \qquad \text{(vs. } \underset{x}{\text{minimize}} \quad \mathbb{E}_{\theta \sim d}\left[f(x, \theta)\right]\text{)}$$

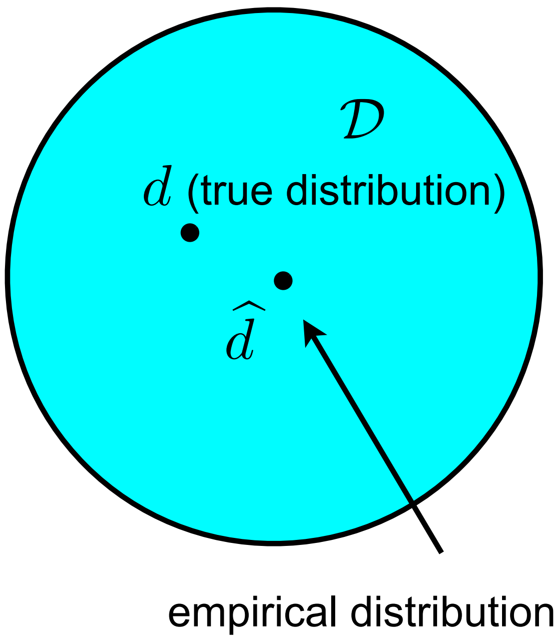
  - Strong worst-case guarantees
  - Subsumes conventional robust optimization

Distributional Uncertainty

- Method 1: Based on certain (pseudo)metric $\mathcal{M}$
  - KL divergence
  - Wasserstein metric (earth mover's distance)
- Metric ball centered at the empirical distribution

$$\mathcal{D}(\epsilon) = \left\{ d \colon \mathcal{M}(d, \widehat{d}) \leq \epsilon \right\}$$

- The ball contains d with high probability



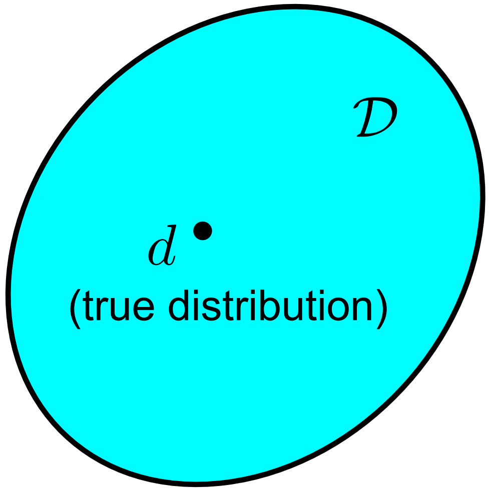
- Advantage: "Nonparametric" characterization
- Disadvantage: Complexity of decision making against $\mathcal{D}$ grows quickly with the number of samples

Distributional Uncertainty (cont'd)

- Method 2: Based on generalized moments (this talk)

$$\mathcal{D} = \{d \colon \mathbb{E}_{\theta \sim d}[g(\theta)] \preceq 0\}$$

- Assume: g is easily bounded



- Examples
  - Moments: $\mathbb{E}[\theta] = \hat{\theta},\ \mathrm{cov}[\theta] = \widehat{\Sigma}$
  - Tail probability: $\mathbb{P}(\theta \geq \bar{\theta}) \leq \epsilon$

- Classical concentration inequalities can be used to compute the probability that $\mathcal{D}$ contains d

Example (Hoeffding's inequality): $$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\theta_i \geq \mathbb{E}\theta + t\right) \leq \exp(-2nt^2)$$

Challenges and My Contribution

$$\underset{x}{\text{minimize}} \quad \max_{d \in \mathcal{D}} \mathbb{E}_{\theta \sim d} \left[f(x, \theta) \right]$$

- **Challenge**: Finding the worst-case distribution
  - Infinite-dimensional optimization problem
  - Not numerically tractable

- Previous work on special instances
  - [Scarf, 1958]: Analytical solution for a special case
  - [Bertsimas, Popescu, 2005]: Optimal probability inequalities
  - [Vandenberghe, Boyd, Comanor, 2007]: Optimal Chebyshev bounds
  - [Delage, Ye, 2010]: Piecewise affine functions

- **My contribution**
  - Formulate equivalent convex optimization problem (under certain conditions)
  - Tractable numerical solutions
  - Conditions apply to many resource allocation and scheduling problems

Main Result: Equivalent Convex Optimization Problem

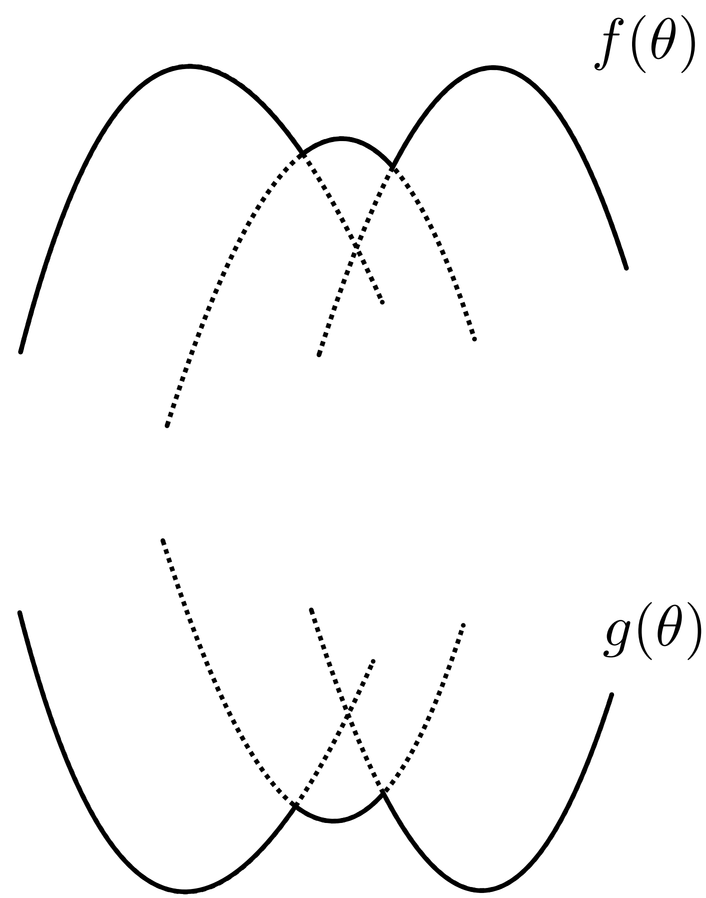
<u>Theorem</u>: There exists an equivalent **convex optimization problem** for computing the worst-case distribution if

- The objective f is *piecewise concave*

$$f(\theta) = \max_k f^{(k)}(\theta) \qquad f^{(k)} \text{ concave}$$

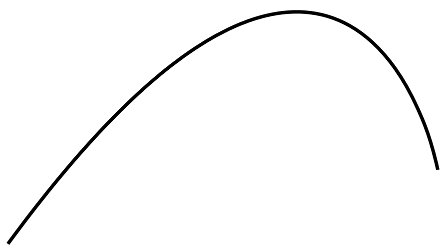
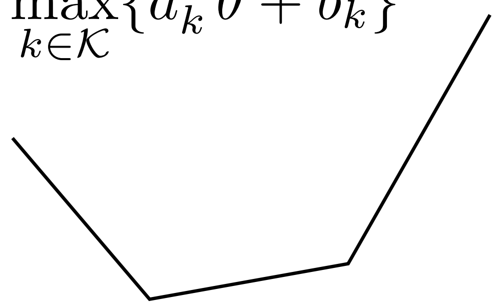
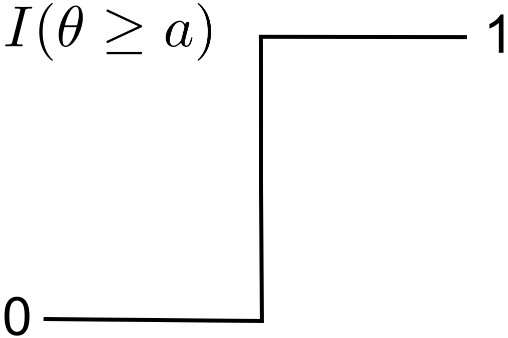
- The constraint g is *piecewise convex*

$$g(\theta) = \min_l g^{(l)}(\theta) \qquad g^{(l)} \text{ convex}$$

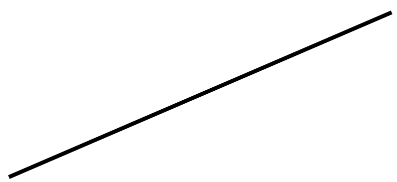
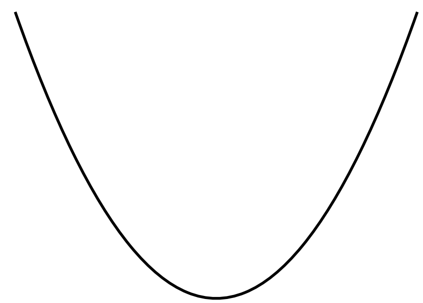
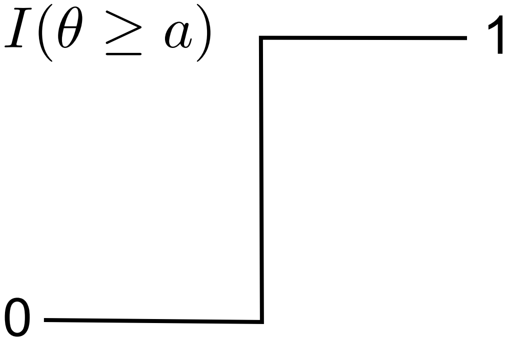


Shuo Han, Molei Tao, Ufuk Topcu, Houman Owhadi, Richard M. Murray, "Convex optimal uncertainty quantification," SIAM Journal on Optimization, 25(3), 1368–1387, 2015.

Piecewise Concave Functions

concave	piecewise affine	0-1 indicator
	$\max_{k \in \mathcal{K}} \{a_k^T \theta + b_k\}$ 	$I(\theta \geq a)$ 
resource allocation/scheduling		failure rate

Piecewise Convex Functions

linear	convex	0-1 indicator
		 $I(\theta \geq a)$
mean	covariance & higher moments	tail probability

The Convex Optimization Problem

$$\begin{aligned} \underset{\{p_{kl},\gamma_{kl}\}_{k,l}}{\text{maximize}} \quad & \sum_{k,l} p_{kl} f^{(k)}(\gamma_{kl}/p_{kl}) \\ \text{subject to} \quad & \sum_{k,l} p_{kl} = 1 \\ & p_{kl} \geq 0, \qquad \forall k, l \\ & \sum_{k,l} p_{kl} g^{(l)}(\gamma_{kl}/p_{kl}) \leq 0 \end{aligned}$$

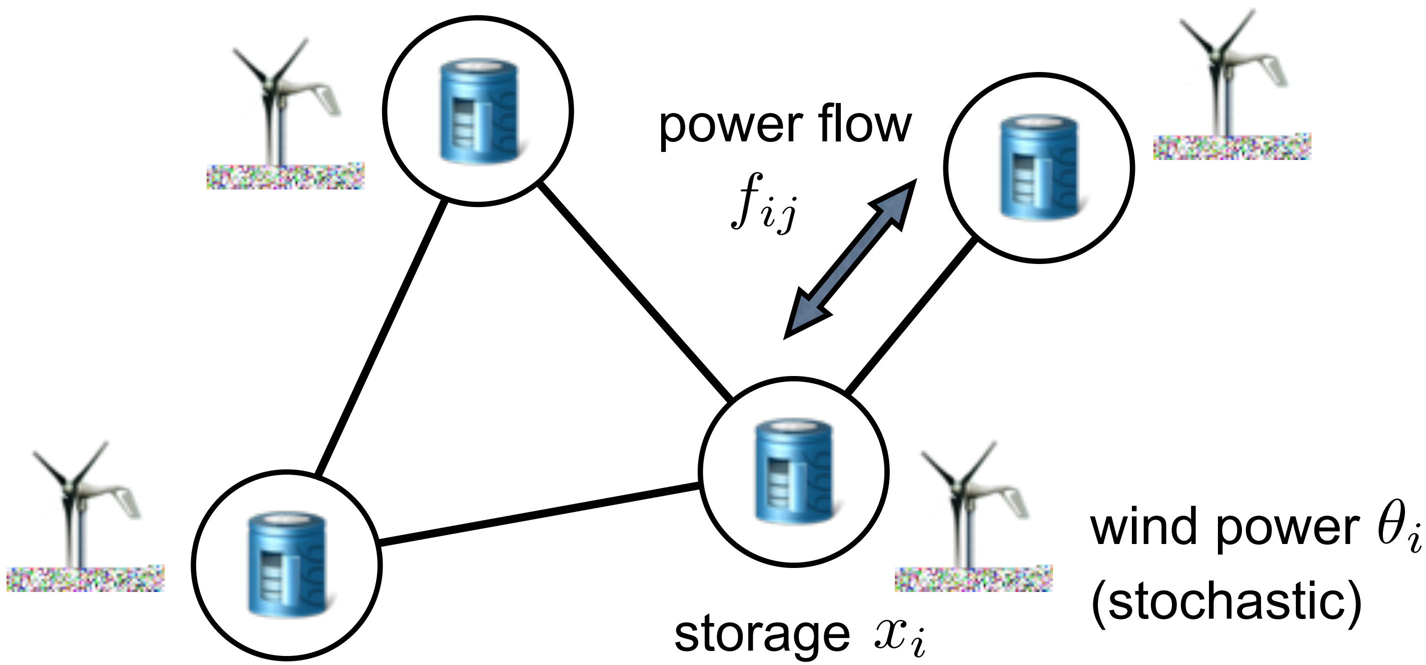
For:

$$f(\theta) = \max_{k \in \{1,2,\cdots,K\}} f^{(k)}(\theta)$$

$$g(\theta) = \min_{l \in \{1,2,\cdots,L\}} g^{(l)}(\theta)$$

- The worst case is always attained by a discrete distribution
- Total number of Dirac masses in the distribution: $K \cdot L$

Storage Allocation for Power Grid



Storage Allocation Problem

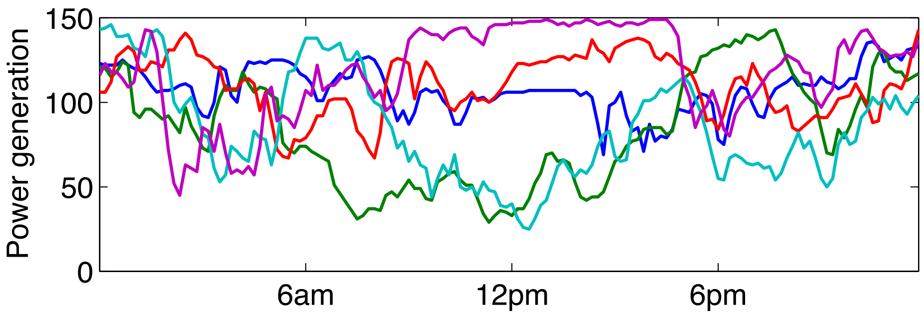
optimal power flow

$$\min_{x}. \quad \max_{d} \mathbb{E}_{\theta \sim d} \left[\min_{f} \text{Wind\_Energy\_Wasted}(x, \theta, f) \right]$$

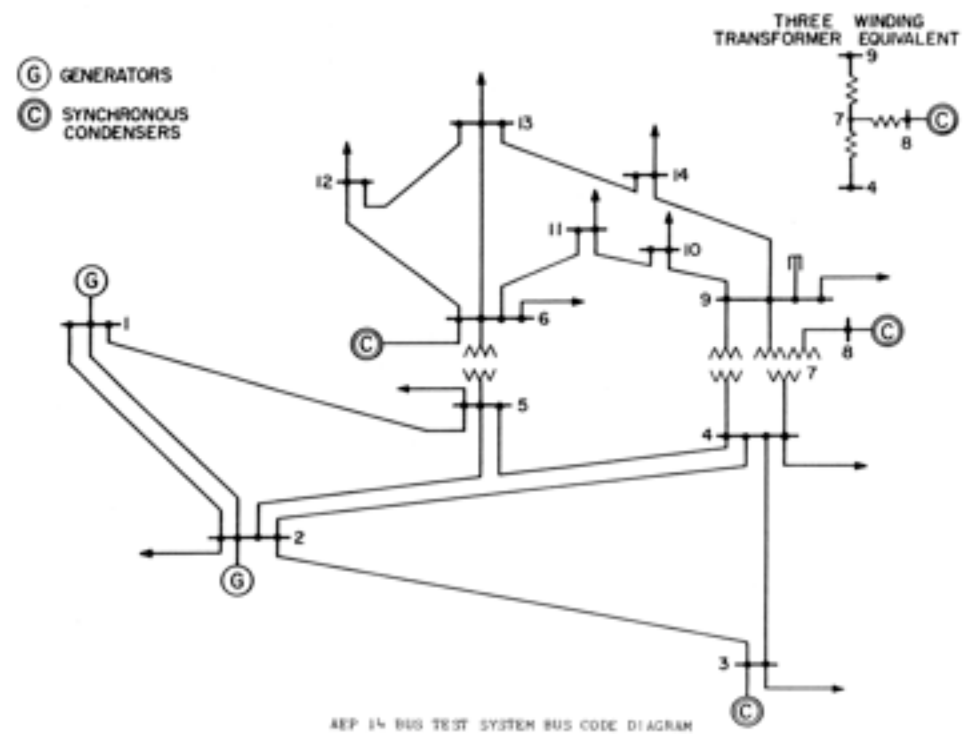
piecewise concave in θ

Numerical Example: IEEE 14-Bus Test Case

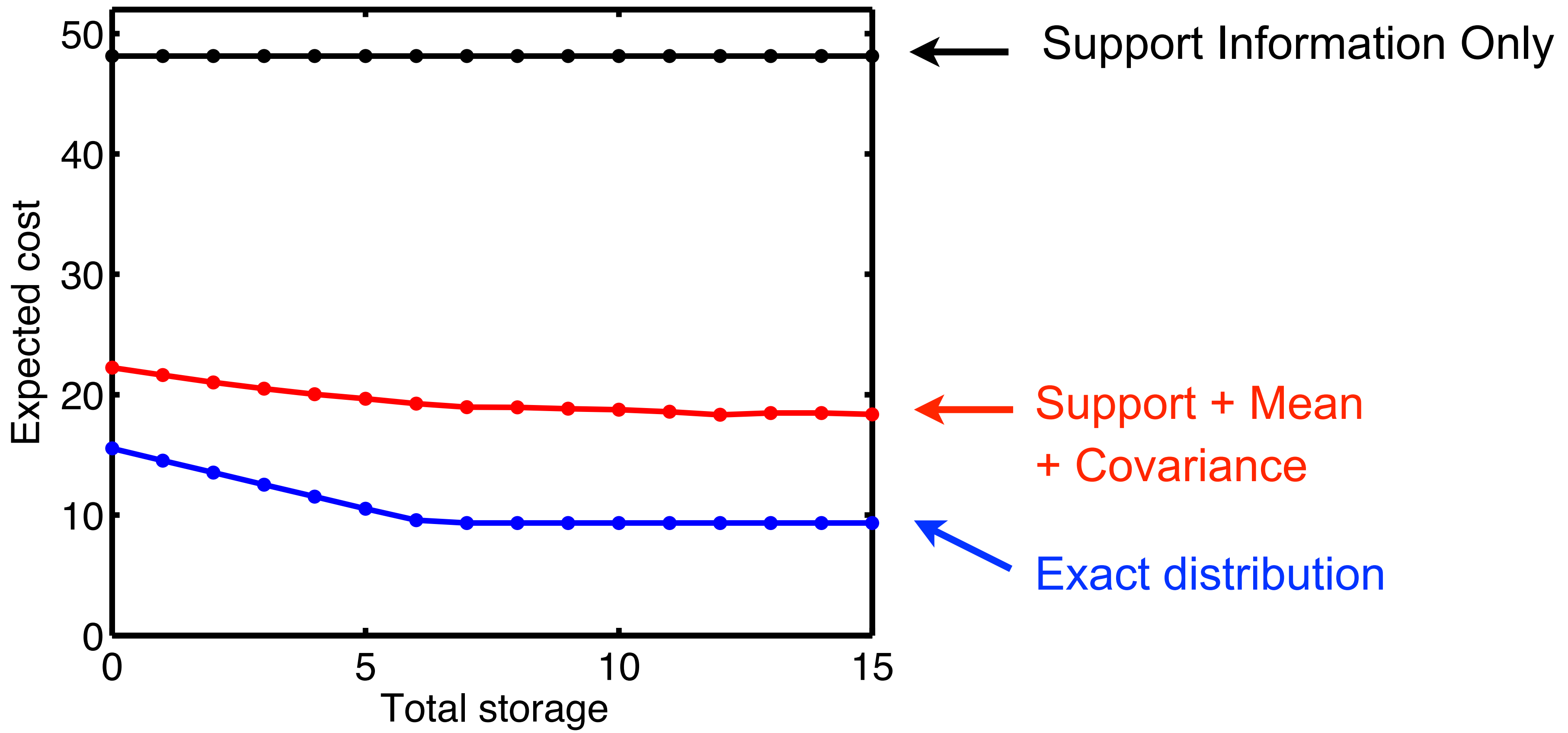
- Network with 5 generators
- Time: one day, 3-hour interval
- Mean and covariance obtained from real wind generation data



[Source]: AESO

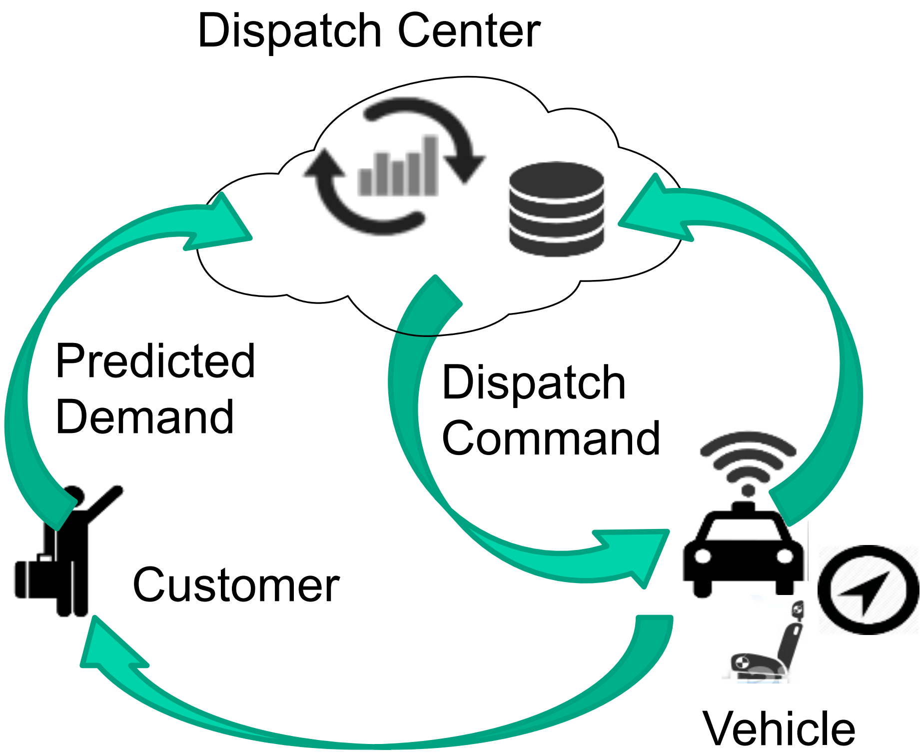


The Influence of Information Constraints

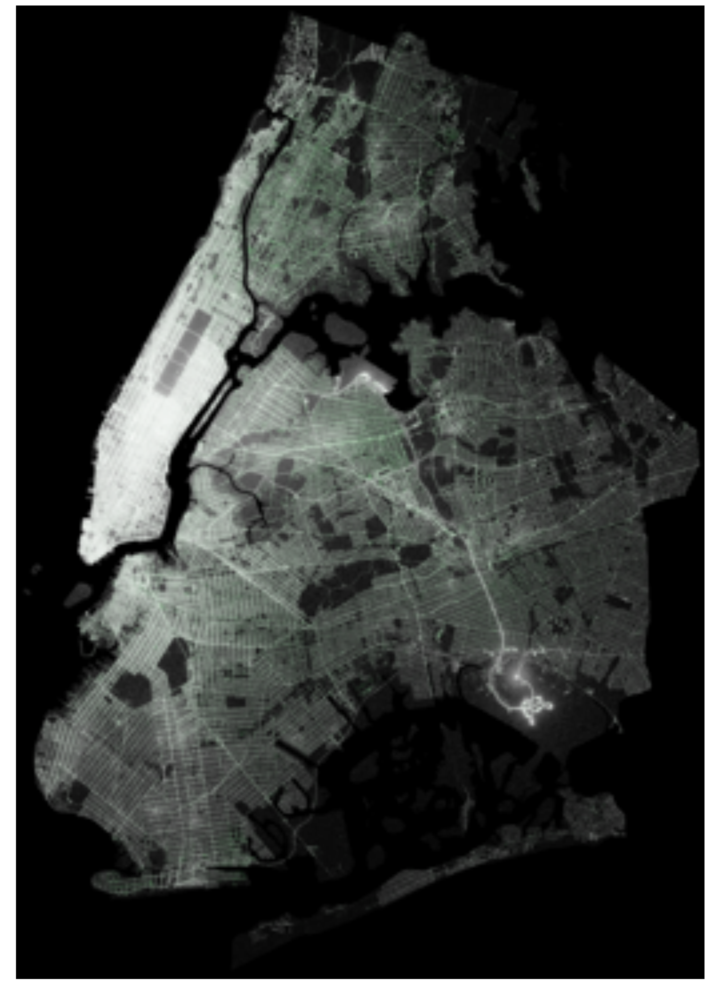


Shuo Han, Ufuk Topcu, Molei Tao, Houman Owhadi, Richard M. Murray, “Convex optimal uncertainty quantification: Algorithms and a case study in energy storage placement for power grids study,” American Control Conference, 2013.

On-Demand Ridesharing



Distribution of Customer Demand



$$\min_{X_{1:T}} \sum_{t=1}^{T} [J_D(X_t) + J_E(X_t, r_t)]$$

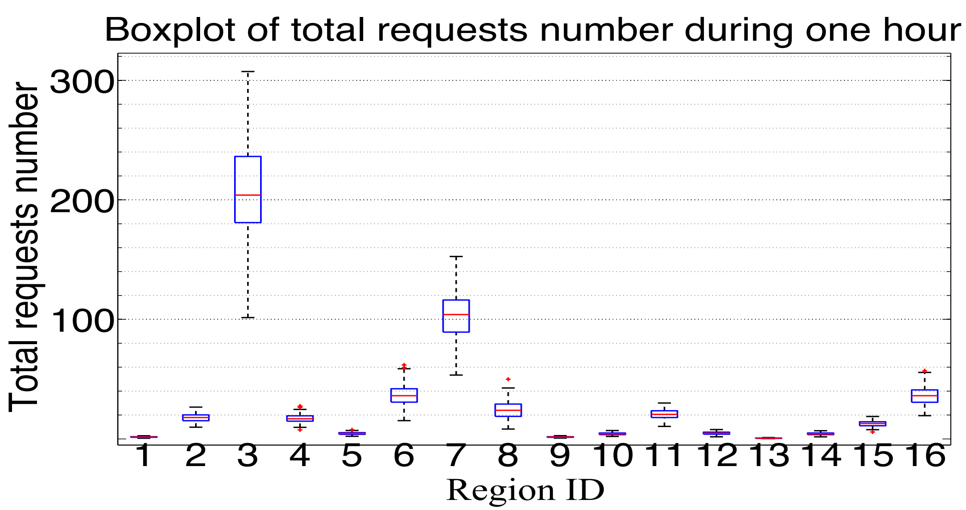
Demand (r_t)

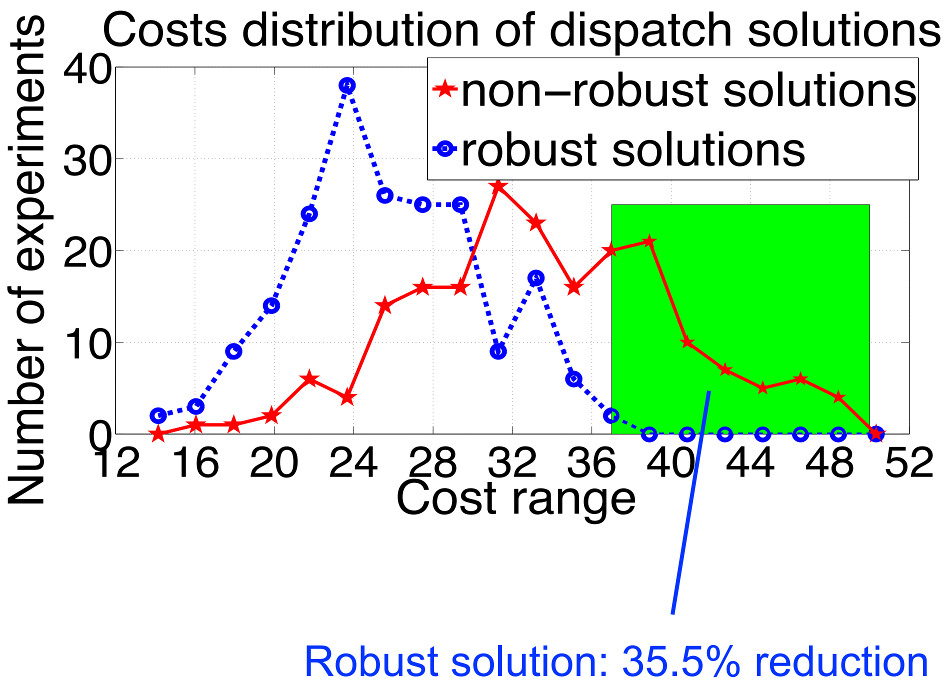
Vehicle Flows ($X_{1:T}$) — Cost of Rebalancing (J_D) — Wait Time (J_E)

Robust vs. Non-Robust

- Robust optimization against demand uncertainty

$$\min_{X_{1:T}} \max_{r_{1:T}\in\Delta} \sum_{t=1}^{T} [J_D(X_t) + J_E(X_t, r_t)]$$





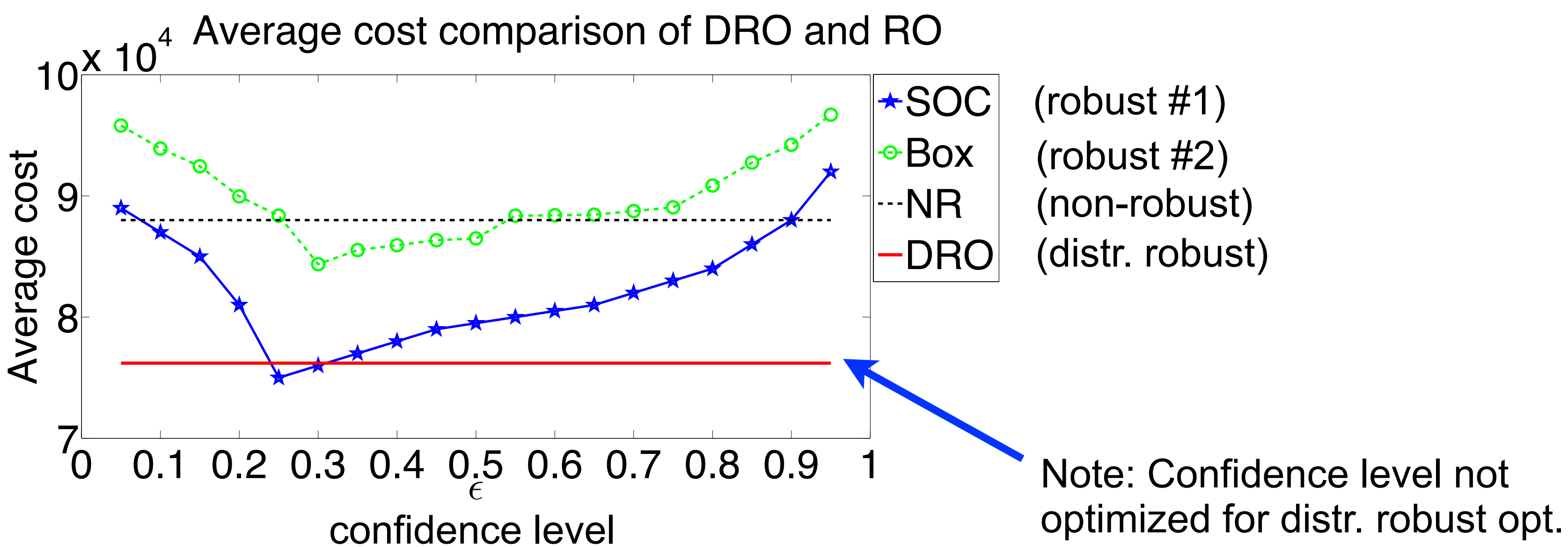
NYC Dataset: 4 years, 100 GB

Fei Miao, Shuo Han, Shan Lin, George J. Pappas, "Taxi dispatch under model uncertainties," IEEE Conference on Decision and Control, 2015.

Conventional vs. Distributionally Robust

- Distributionally robust formulation

$$\min_{X_{1:T}} \ \max_{d \in \mathcal{D}} \ \mathbb{E}_{r \sim d} \left\{ \sum_{t=1}^{T} \left[J_D(X_t) + J_E(X_t, r_t) \right] \right\}$$



Confidence level: Probability the true parameter/distribution lies outside ambiguity set

Fei Miao, Shuo Han, Abdeltawab Hendawi, et al., "Data-driven distributionally robust vehicle balancing with dynamic region partition," ACM/IEEE Intl. Conf. on Cyber-Physical Systems, 2017.

Future Directions

Large Datasets

- Distributed computation
- Approximate algorithms

Structured Models

- Markov properties
- Prior knowledge

Online Optimization

- When to discard old data
- When to re-learn

Summary

- **Distributional uncertainty:** A new approach to data-driven optimization
- **A rigorous way to make use of sampled data**
  - Probabilistic guarantees
  - Worst-case analysis/design: Often required for engineering applications
- **Computationally efficient**
  - Convex formulation available for a large class of problems
  - Examples: Resource allocation and scheduling