Optimal Treatment Assignment to Evaluate Demand Response

Baosen Zhang
Electrical Engineering
University of Washington

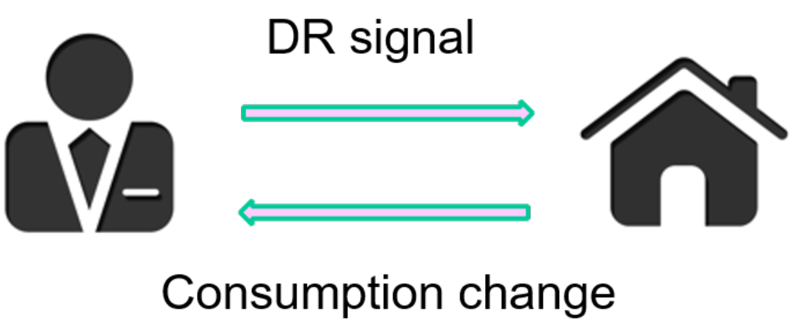
Grid Science Winter School & Conference

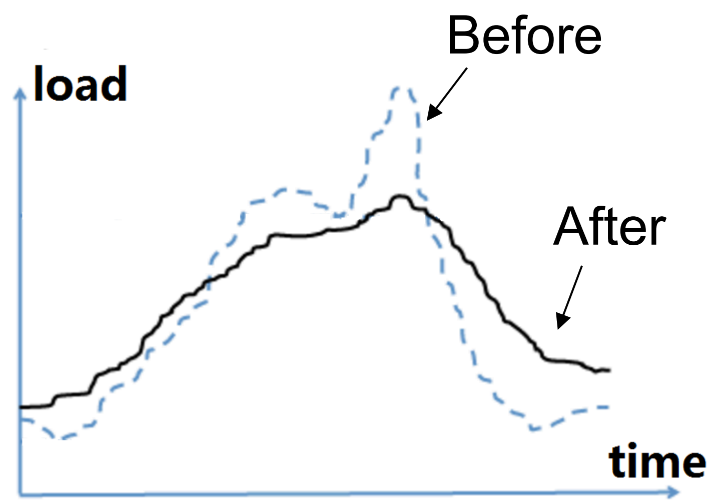
Jan 12, 2017

Joint work with
Pan Li, Yize Chen

Introduction

- Demand Response (DR): send a signal to elicit a change in customer demand
- Change in price, text message, etc.





Introduction

Standard setup for demand response (DR):

1. Direct load control
2. Indirect control:
   - Each user has some utility function (public or private)
   - Maximize the social welfare

Our setting: no direct control and no detailed information

This talk:

How to learn the impact of demand response

Problem Setup

Stylized setup:

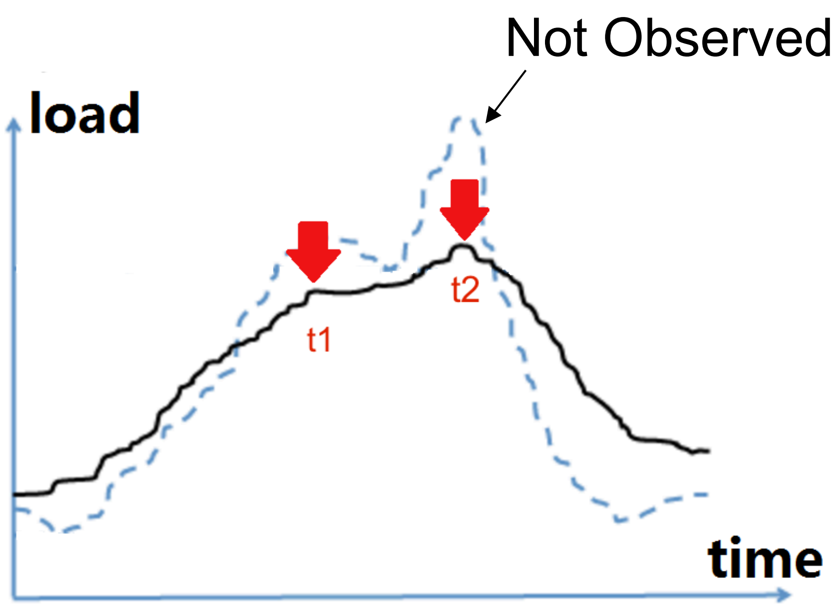
- Utility sends a signal, 0 or 1, to a user
  - 1: perform demand response
  - 0: do nothing (or no signal to the user)

- Quantity of interest: causal impact of DR

$$Consumption|DR - Consumption|No\ DR$$

Challenges

No counterfactual observation



High Dimensionality of Covariates

Covariates
- Temperature
- Type of day
- Size of house
- EVs
- PVs
- Appliances
- Interaction terms

Most of time a user is not called for DR

- E.g., a user can be called no more than 5 times in one month

Overcoming the Challenges

Estimating an effect under infrequent signaling with a large number of covariates is a hard problem

- Existing estimation techniques performs poorly

Our approach: strategically signaling

- Carefully choose DR signals based on the covariates

Result: We show an optimal estimation strategy with high dimensional covariates

Outline

- Linear model
- Signaling strategy
- Theoretical Analysis
- Simulation with real building data
- Online problem

Additive Linear Model

$$y_i = \beta x_i + \gamma^T z_i + \varepsilon_i ,$$

Consumption — DR signal {0,1} — Covariates — Noise

β: the causal impact of DR signal ← Learn This

γ: impact of other covariates, vector of dimension d

Observe n samples:

$y_1, y_2, \ldots, y_n$

$x_1, x_2, \ldots, x_n$

$\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n$

Estimation Problem

Estimate β (impact of DR)

- Given $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_n$
- Limited signaling: design $x_1, x_2, \ldots, x_n$, at most k of x_i can be 1 ($k \ll n$)
- Observe $y_1, y_2, \ldots, y_n$

$\hat{\beta}$: estimate of β

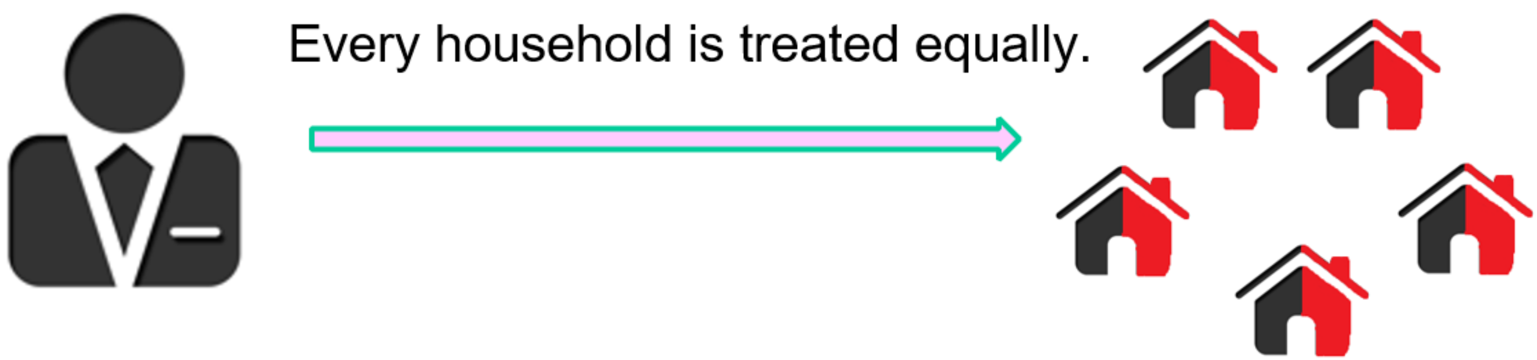
- Unbiased
- Minimize $Var(\hat{\beta})$

High dimensional setting: $d \approx n$

A designer can optimize of the signaling strategy

Standard Practice

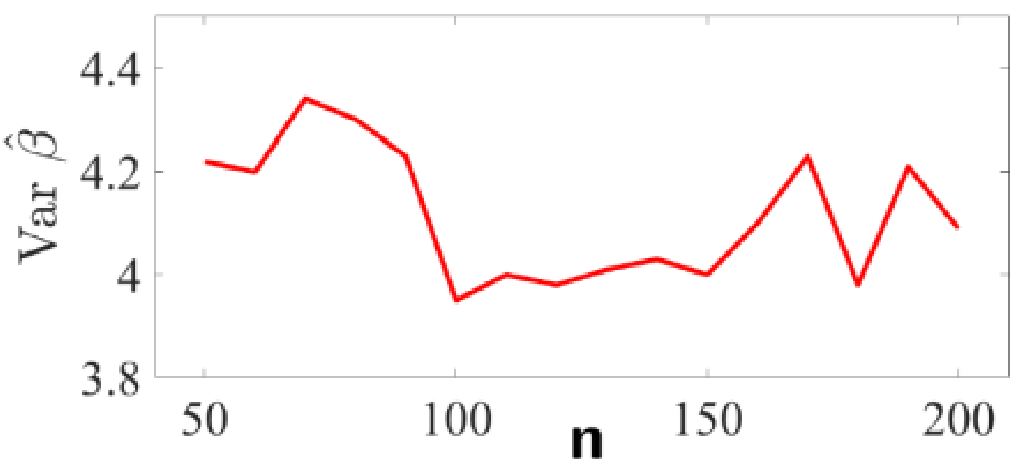
- Signals are randomly assigned
  - E.g., $k/n = 1/3$, $x_i = 1$ with probability $1/3$
- Metric: variance of the estimate, $Var(\beta)$
- High dimension: $d = n - 1$



Standard Practice

- Signals are randomly assigned
  - E.g., $k/n = 1/3$, $x_i = 1$ with probability $1/3$
- Metric: variance of the estimate, $Var(\beta)$
- High dimension: $d = n - 1$

Run a linear regression: $y_i = \beta x_i + \gamma^T z_i + \varepsilon_i,$



Variance does not decrease!

Standard Practice

$$y_i = \beta x_i + \gamma^T z_i + \varepsilon_i,$$

Method 1: Predict then subtract

- Fit the best predictive model, then subtract out the prediction to find the impact of DR

Estimating $\boldsymbol{\gamma}$ is hard!

Method 2: Difference-in-Means

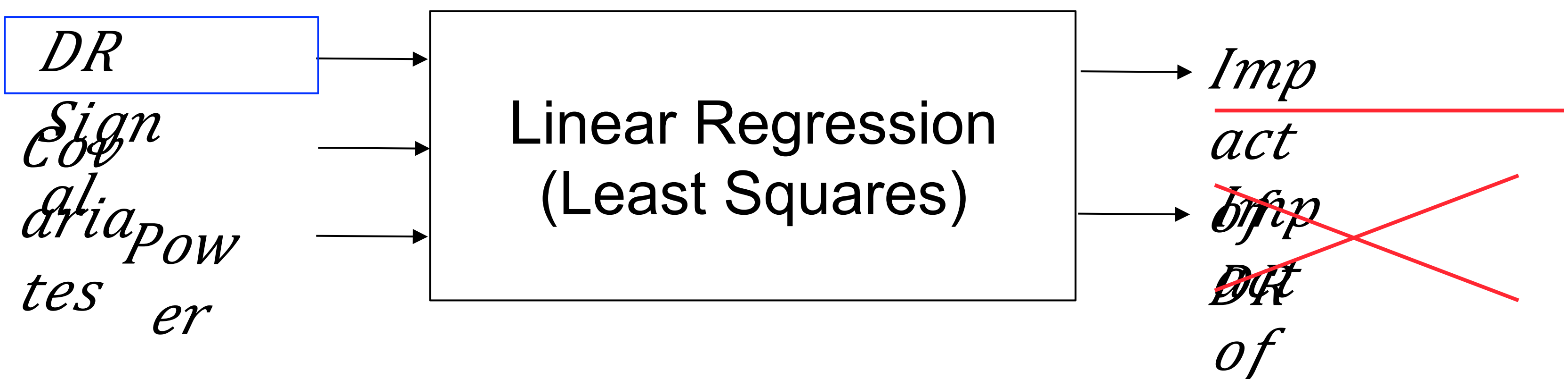
- Ignore covariates, pretend the model is

$$y_i = \beta x_i + \epsilon_i$$

Throwing information away as noise!

Our approach

- Use information in the covariates
- Don't try to do prediction
- Strategically assign signals



Assigning DR signals allows us to tradeoff between the estimation of the two types of parameters

Variance of Estimator

$$y_i = \beta x_i + \gamma^T z_i + \varepsilon_i ,$$

- Running linear regression, the variance of the estimator of beta is given by

$$\mathrm{Var}\left(\hat{\beta}\right) = \frac{\sigma^2}{\mathbf{x}^T \mathbf{P} \mathbf{x}}$$

Where

$$\mathbf{P} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}$$

$\mathbf{x}$: vector of DR signals

$\mathbf{Z}$: matrix of covariates

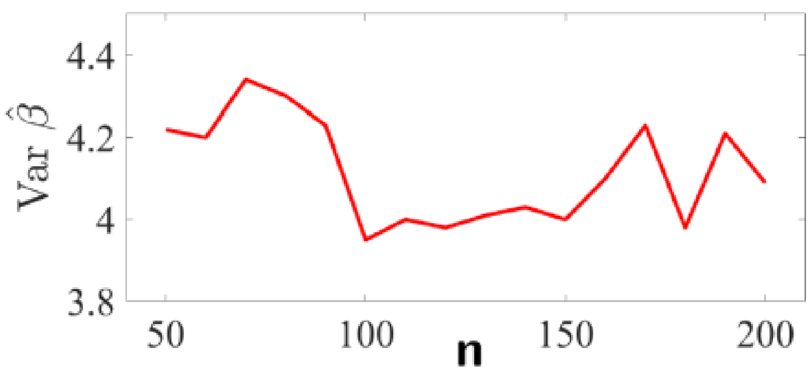
Optimization Problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \operatorname{Var}\hat{\beta} = \frac{\sigma^2}{\boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x}} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

Limited signals ⟷

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

- Non-trivial problem:
  - Non-convex, binary variables
- Is it worth solving? How to solve it?



Random assignment

Optimal Assignment

- A lower bound: No strategy can achieve a better reduction in variance than $1/n$

- Two questions:
  - Can we achieve this rate? Yes, there exist an assignment
  - Can we solve the problem efficiently? Yes, relaxation

Look at rate first

Best Rate

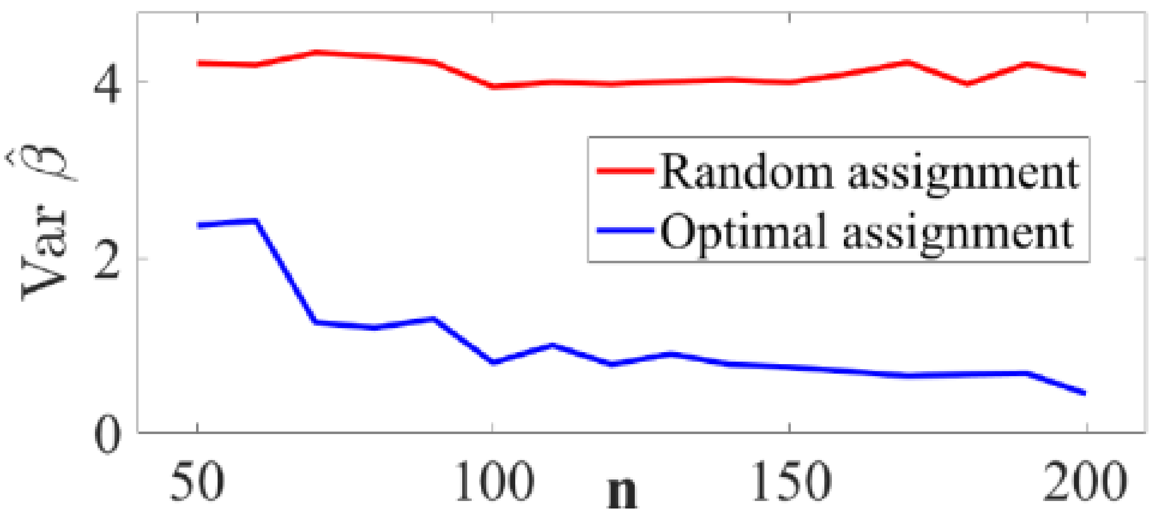
$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \operatorname{Var}\hat{\beta} = \frac{\sigma^2}{\boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x}} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

- Result: There exist a solution such that $Var(\beta)$ scales as $1/n$, as long as $d<n$ and $k/n>\epsilon$, for some fixed ϵ

- Contrast: If $x\downarrow i$ are randomly assigned, then then $Var(\beta)$ stays constant if d is close to n, for all values of k

Example

- Synthetic data:



Achieving Optimal Rate

$$y_i = \beta x_i + \gamma^T \mathbf{z}_i + \varepsilon_i,$$

Dimension d

- Look at the extreme case where $d=n-1$, hardest case to learn β
- Quantity of interest: $x\uparrow T\, Px$
- P is a projection matrix:

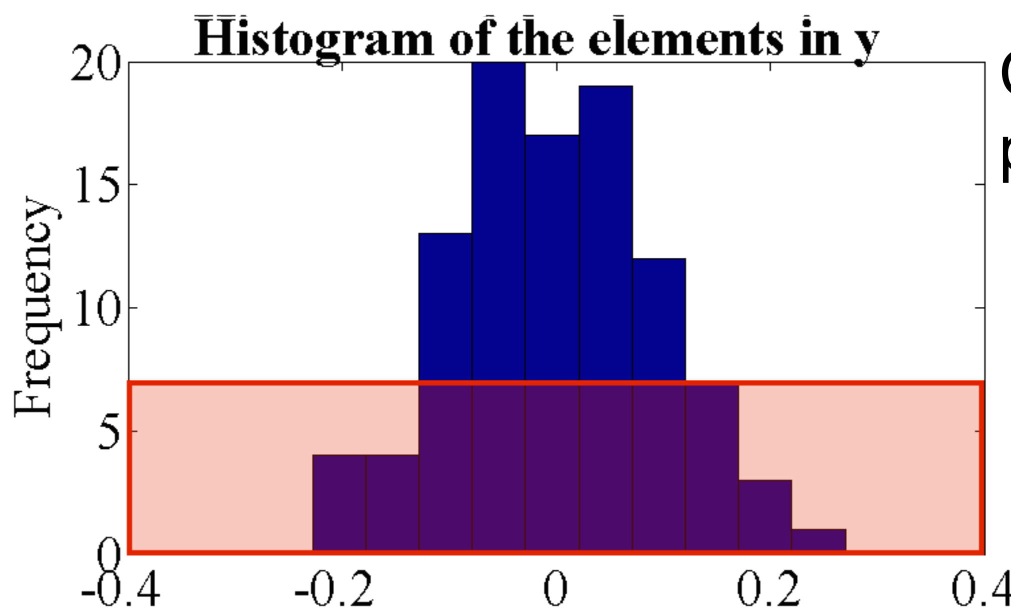
$$\mathbf{P} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^\mathbf{T}\mathbf{Z})^{-1}\mathbf{Z}^\mathbf{T} = \mathbf{y}\mathbf{y}^\mathbf{T}, \;\; \mathbf{Z}^T\mathbf{y} = \mathbf{0}, \;\; ||\mathbf{y}||_2 = 1$$

Goal: maximize

$$(\mathbf{x}^\mathbf{T}\mathbf{y})^2$$

Null Space

- Assume $\boldsymbol{Z}$ has random Gaussian entries, is n by d-1
  - $null(\boldsymbol{Z}^T)$: has a basis with i.i.d. Gaussian entries
  - y: normalized version
- Maximize $(x^T y)^2 = (\sum x_i y_i)^2$



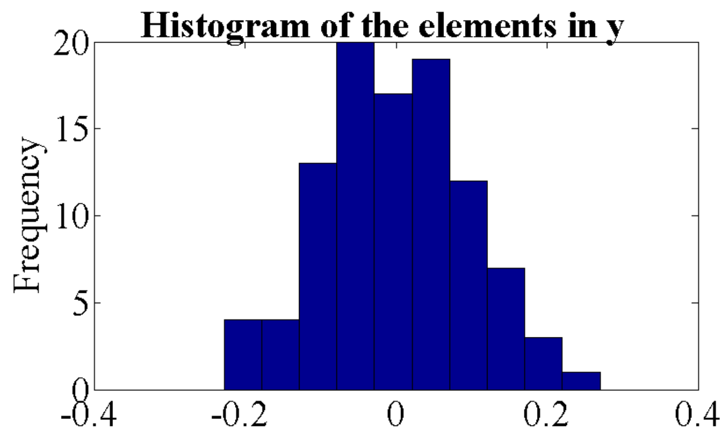
Choose positive values

Random Selection

Extreme Case

Max $(x^T y)^2$

- The information from each signal is not equal
- Strategically assign to get the maximum information



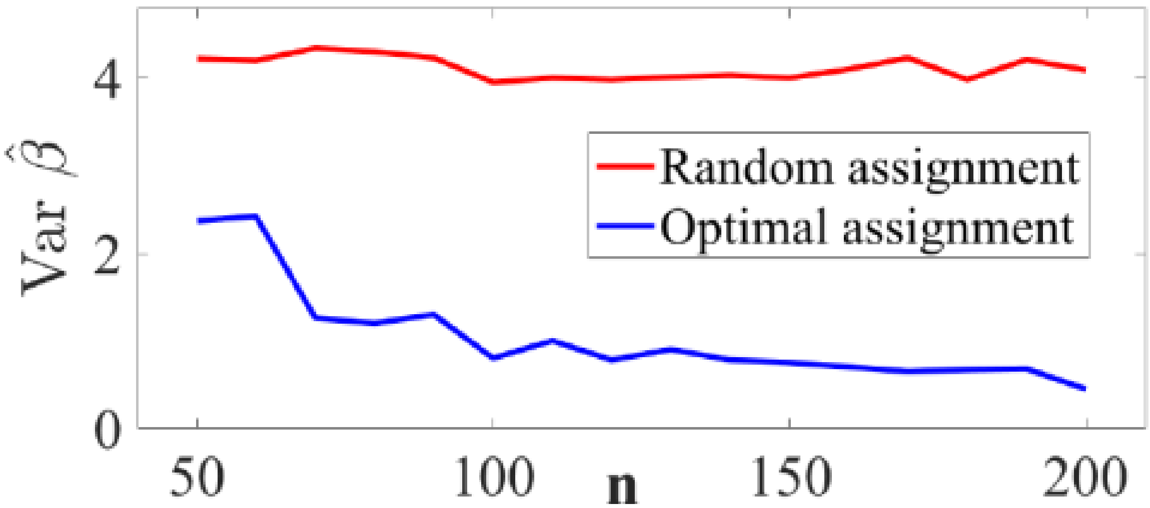
The optimal algorithm is easy

- Find y
- Sort: $y^{(1)} > y^{(2)} > ... > y^{(n)}$
- Assign $x=1$ to the largest k elements

Rate is n as long as $k/n > \epsilon$

Example

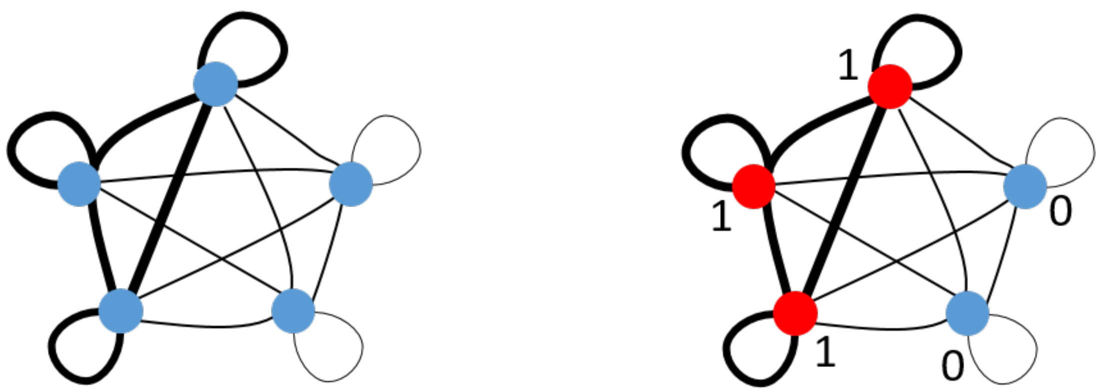
- Synthetic data: d=n-1, k/n=1/3



General Settings

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

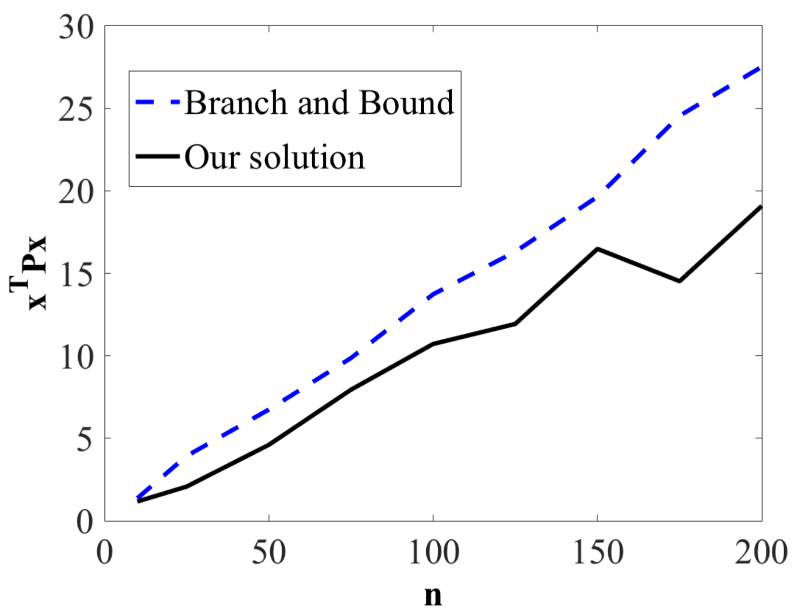
This is actually a graph partition problem:



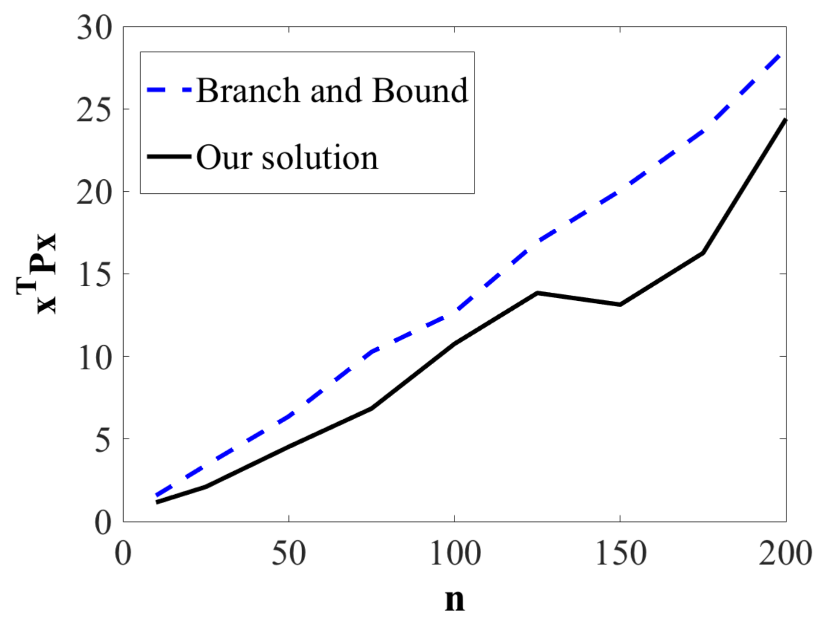
There is a SDP relaxation with provable gaps

Quality of Solution

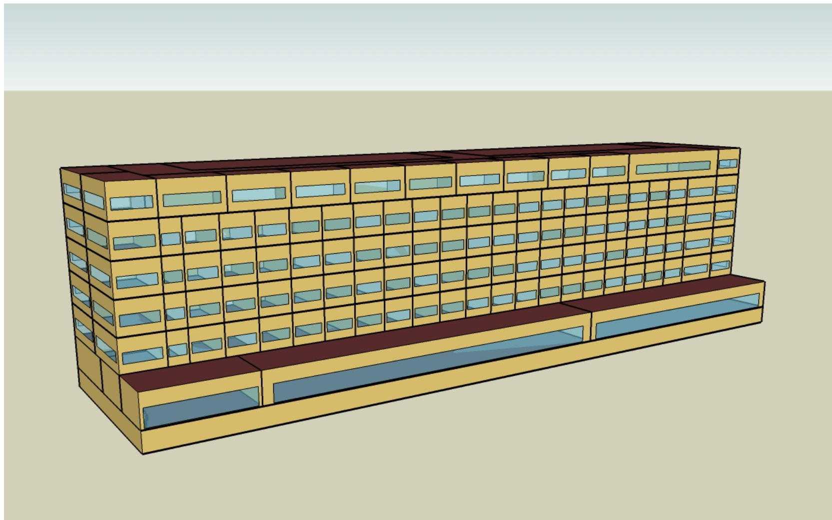
Gaussian Entries



Uniform Entries



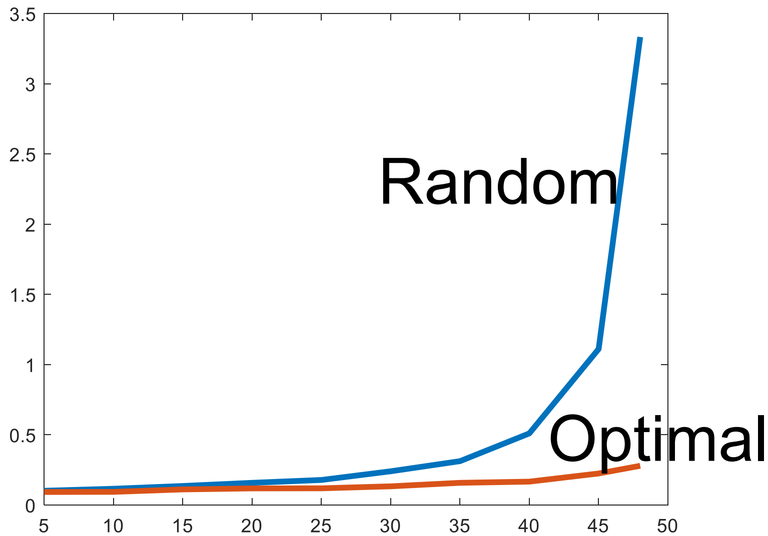
Some Real Data



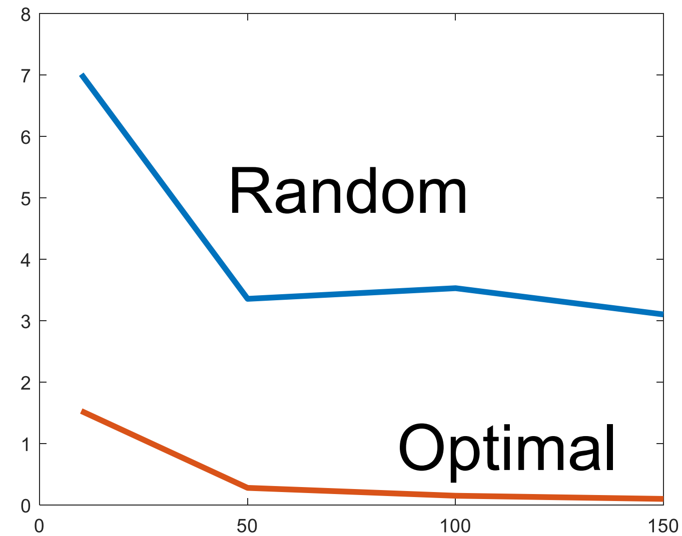
- A hotel in Seattle, with at most 48 covariates including outside temperature, zonal temperature, heating, appliance, etc…
- Train a regression model based on all the data, then simulate DR
  - We can test the impact of covariate dimensions

Estimation Error

- Fix n=50, varying d



Fix d, varying n



- Trying to conduct some trials

Online Setting

- We have considered the offline problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

- Online Setting: approximate P in an online fashion

- Some preliminary results

Conclusion

- An optimal treatment assignment strategy in the context of demand response
  – It is possible to learn under unfavorable conditions

- Future work:
  – Online algorithm
  – Other response models
  – Learning and optimizing

SDP Relaxation

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \boldsymbol{x}^{\mathrm{T}} P \boldsymbol{x} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = k \\ & x_i \in \{1, 0\}. \end{aligned}$$

$$\xrightarrow[X \succcurlyeq 0]{x_i = 2x_i - 1}$$

$$\begin{aligned} \underset{\boldsymbol{X}, \hat{\boldsymbol{x}}}{\text{maximize}} \quad & \frac{1}{4} \sum_i \sum_j P_{ij}(1 + \hat{x}_i + \hat{x}_j + X_{ij}) \\ \text{subject to} \quad & \sum_i \hat{x}_i = 2k - n \\ & X_{ii} = 1 \\ & \sum_i \sum_j X_{i,j} = (2k - n)^2 \\ & \begin{bmatrix} 1 & \hat{\boldsymbol{x}}^{\mathrm{T}} \\ \hat{\boldsymbol{x}} & \boldsymbol{X} \end{bmatrix} \succeq 0. \end{aligned}$$

- There is a randomized algorithm to recover a feasible solution x

Can show

$$E[recovered\ solution]/SDP \geq const$$

Challenges

Three challenges in estimating the impact of DR:

1. The counterfactual is not observed: what would have happened if the opposite was done?

2. There are many other exogenous factors