Formal Privacy Guarantees for Optimization Datasets in Power Systems

Vladimir Dvorkin

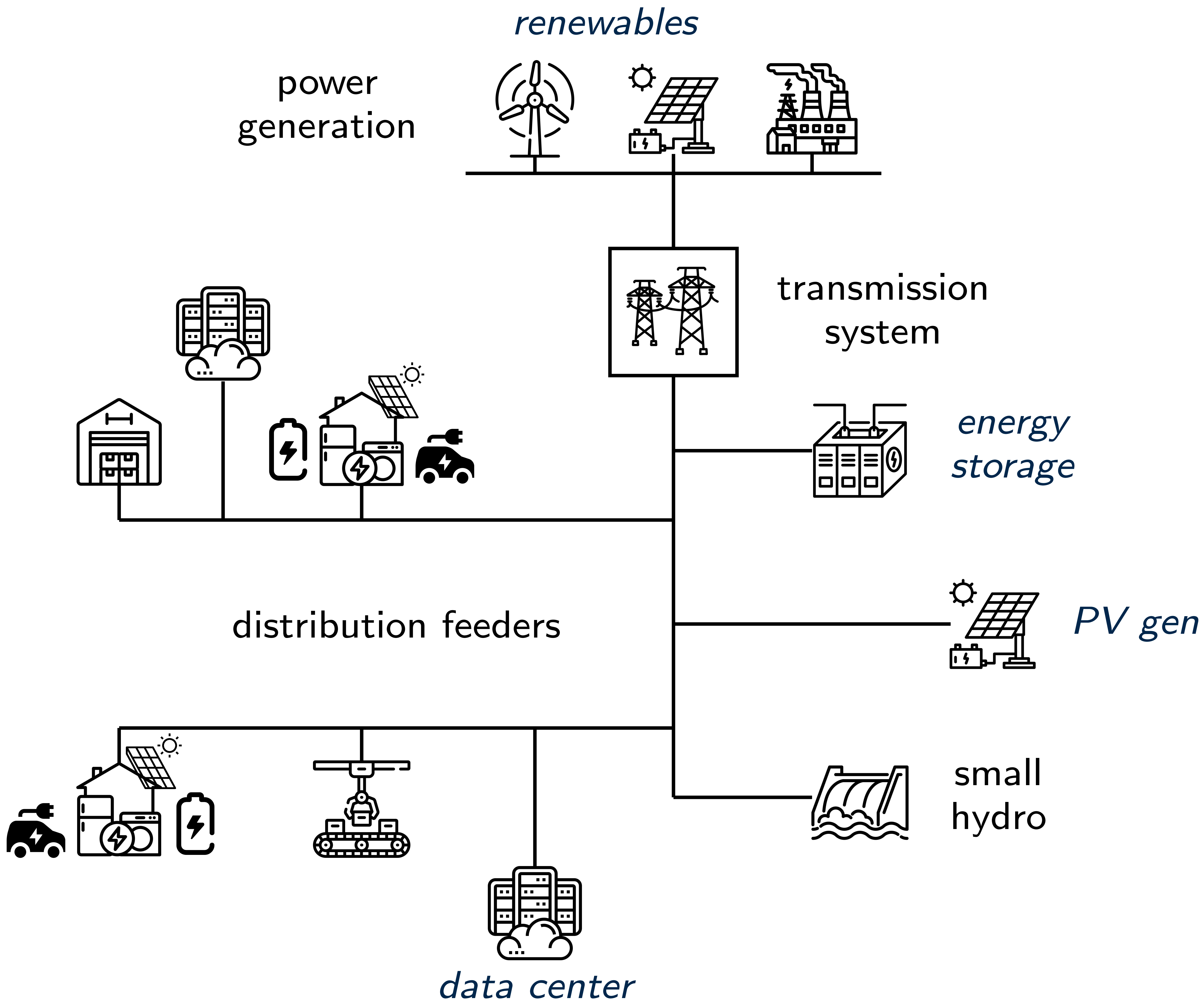
Department of Electrical Engineering and Computer Science

University of Michigan – Ann Arbor

March 7, 2024

UNIVERSITY OF MICHIGAN

Power systems of the future as a collection of optimization problems



Power resource allocation

Data: cost, tech limits, topology
Type: LP, QP, MIQP, NLP
Result: cost-optimal and feas. allocation

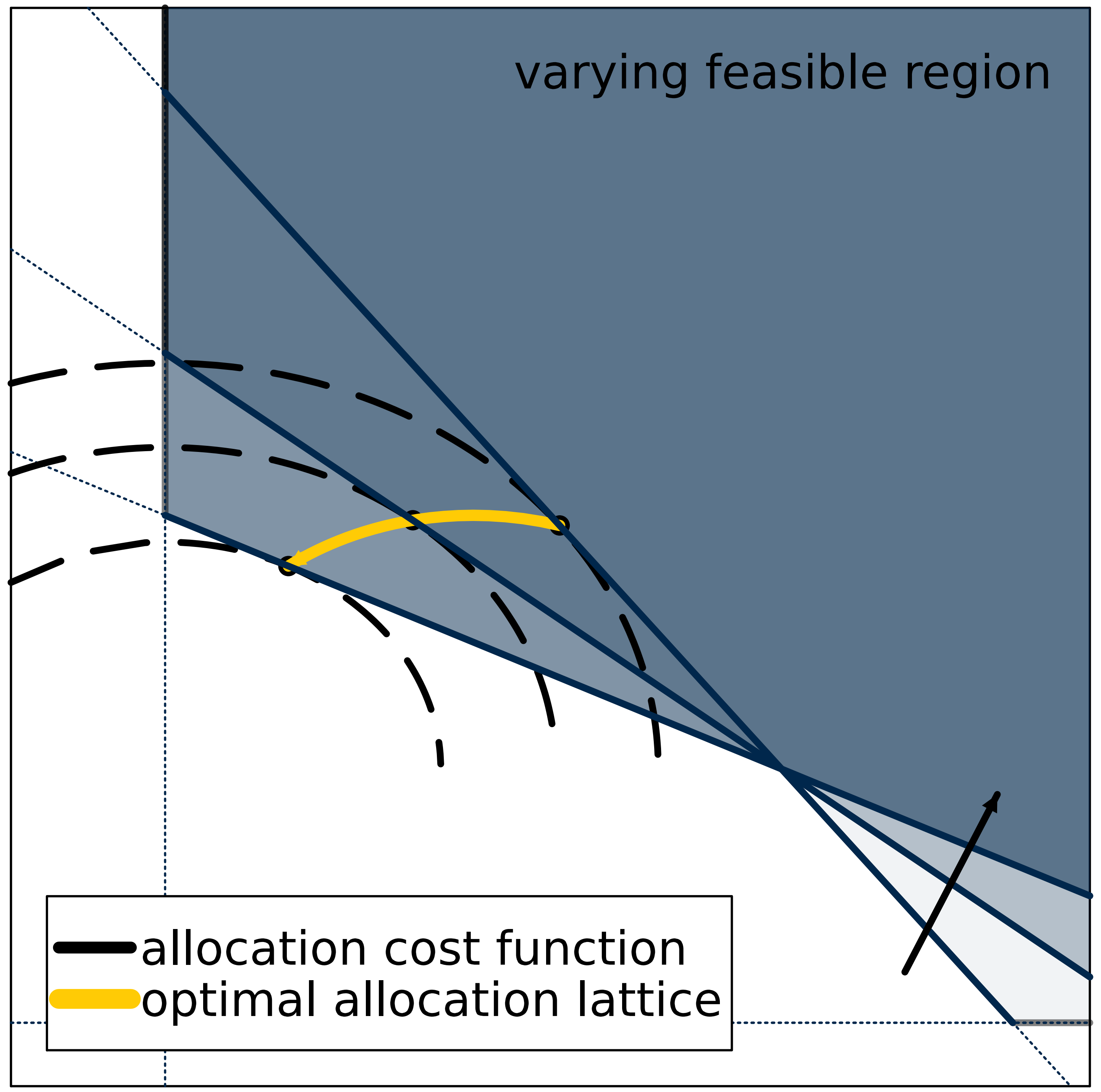
Renewable power forecasting

Data: historical records, weather forecast
Type: QP, convex or NLP
Result: forecast w/ varying leading times

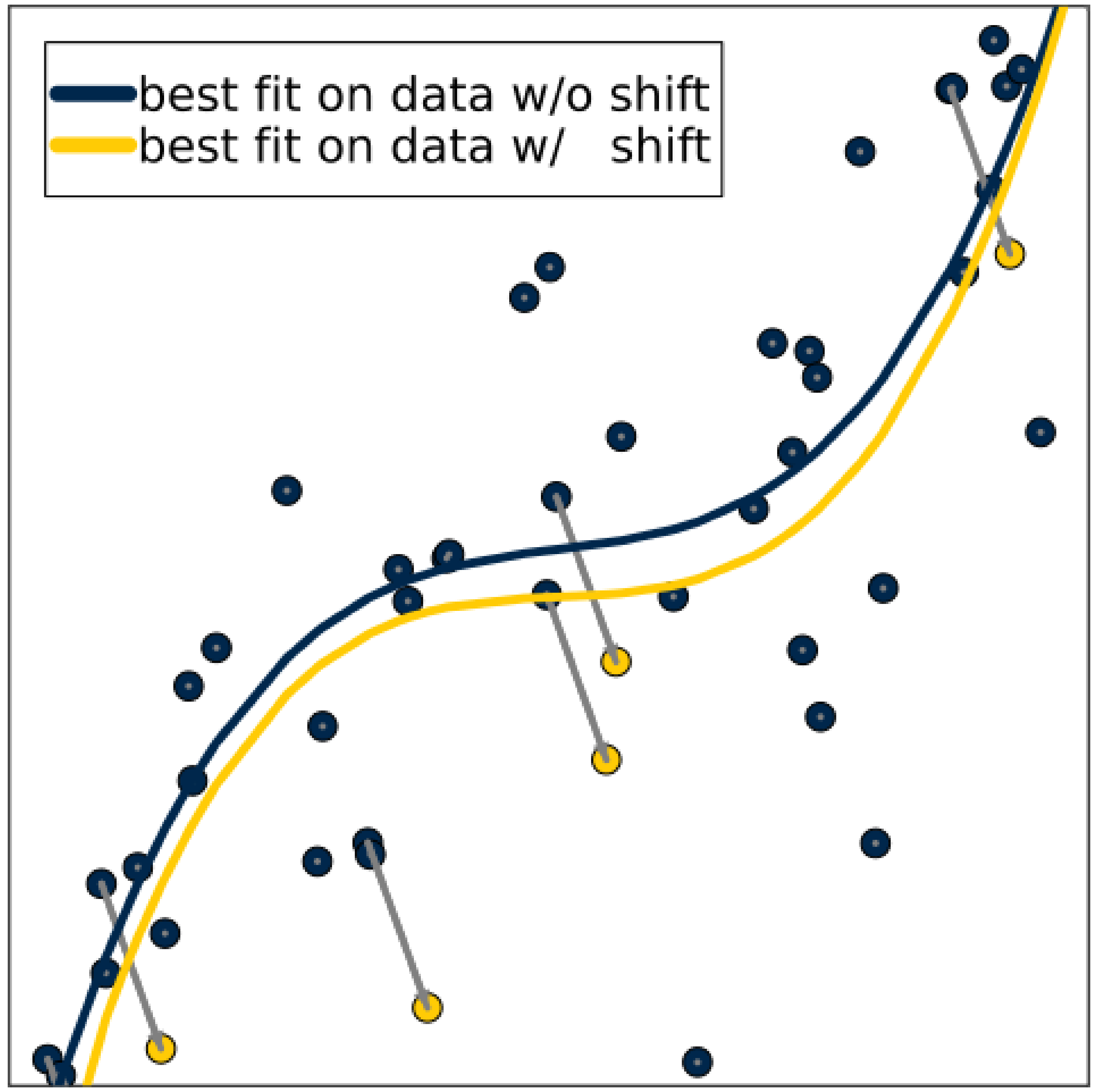
Demand response

Data: loads and tech limits
Type: LP, QP, MIQP, etc.
Result: load timing and geo allocation

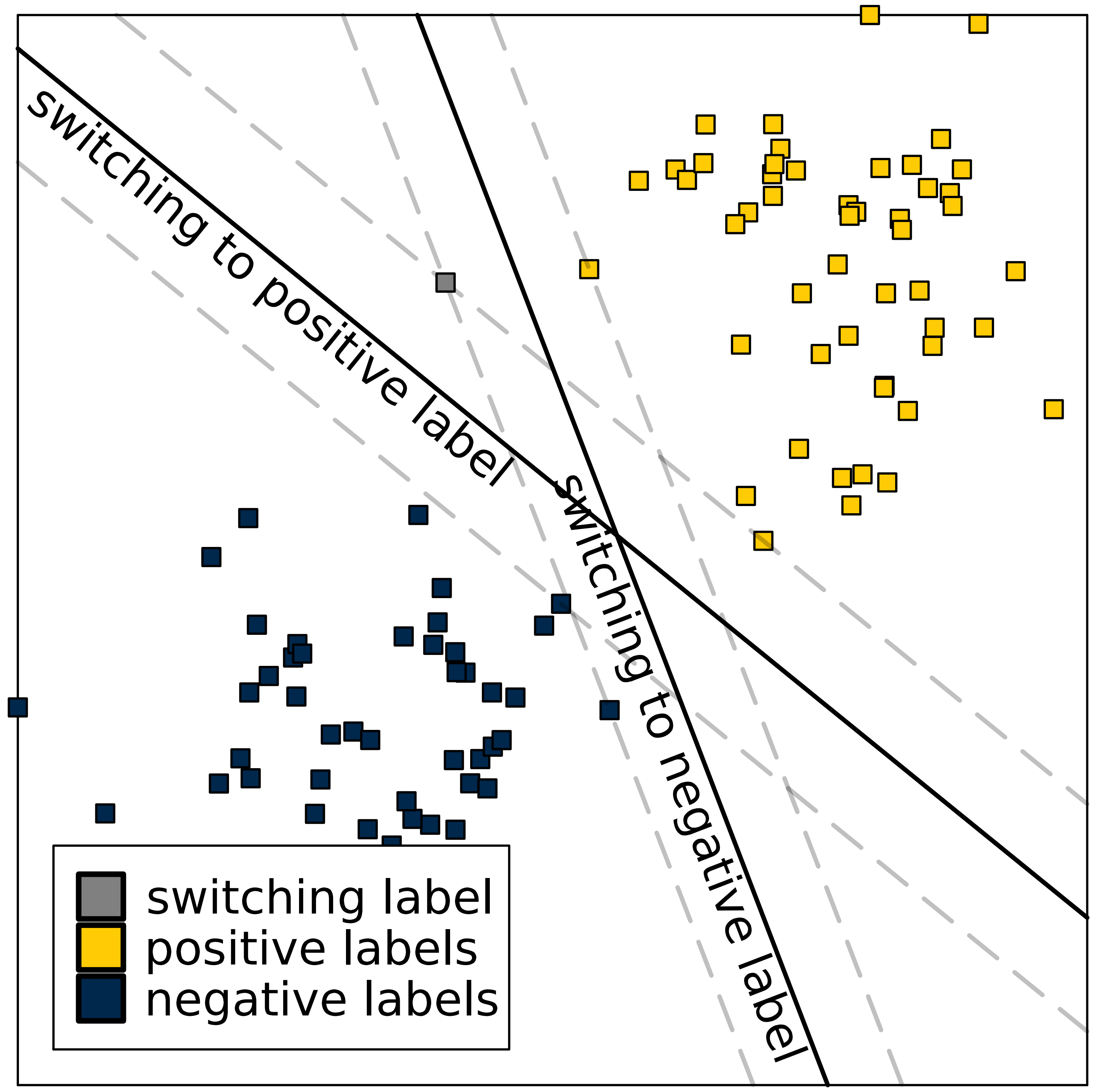
Resource allocation



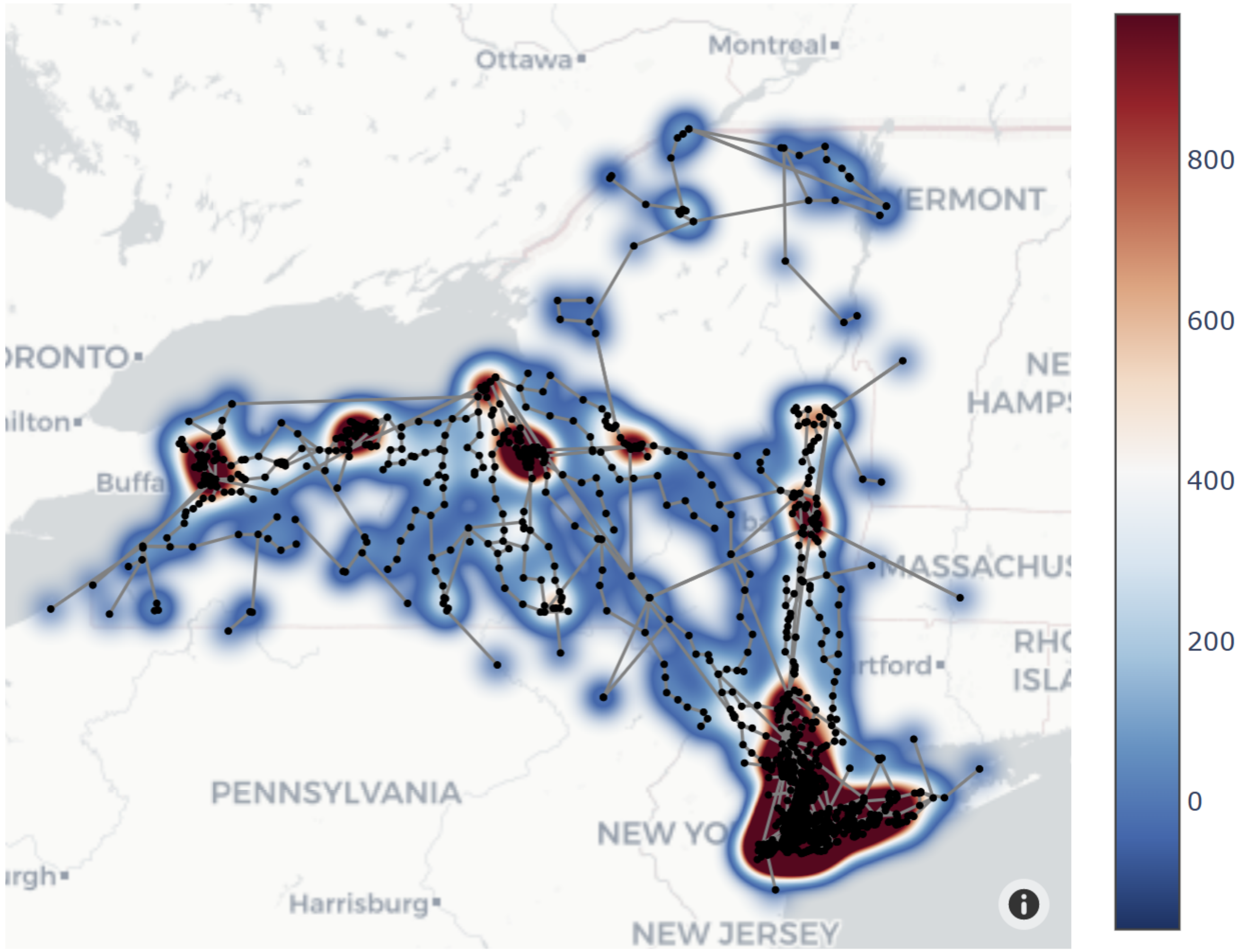
Regression analysis



Classification

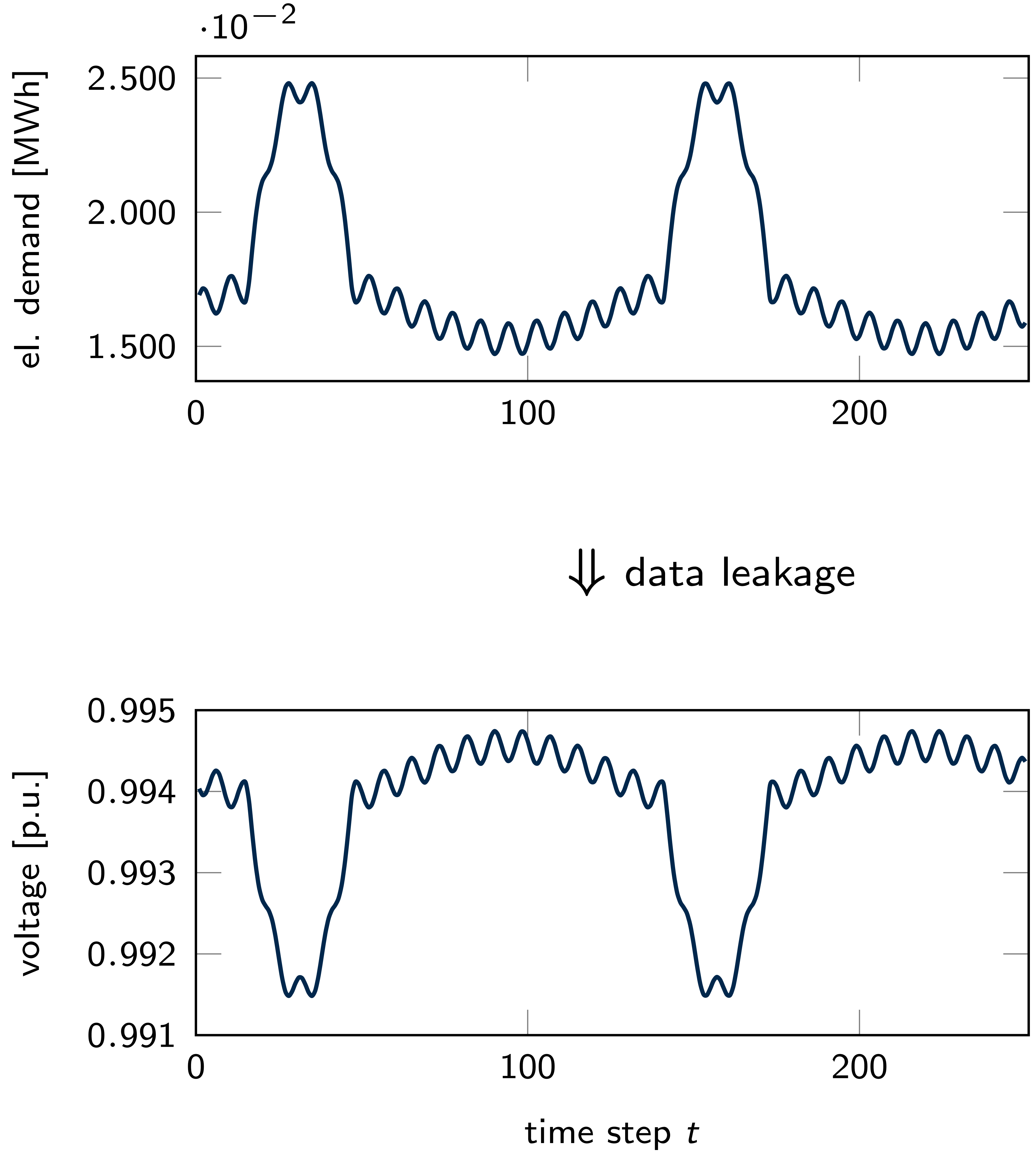


High-voltage systems



Electricity prices [$/MWh] at New York ISO, August 28, 2018

Medium- and low-voltage systems

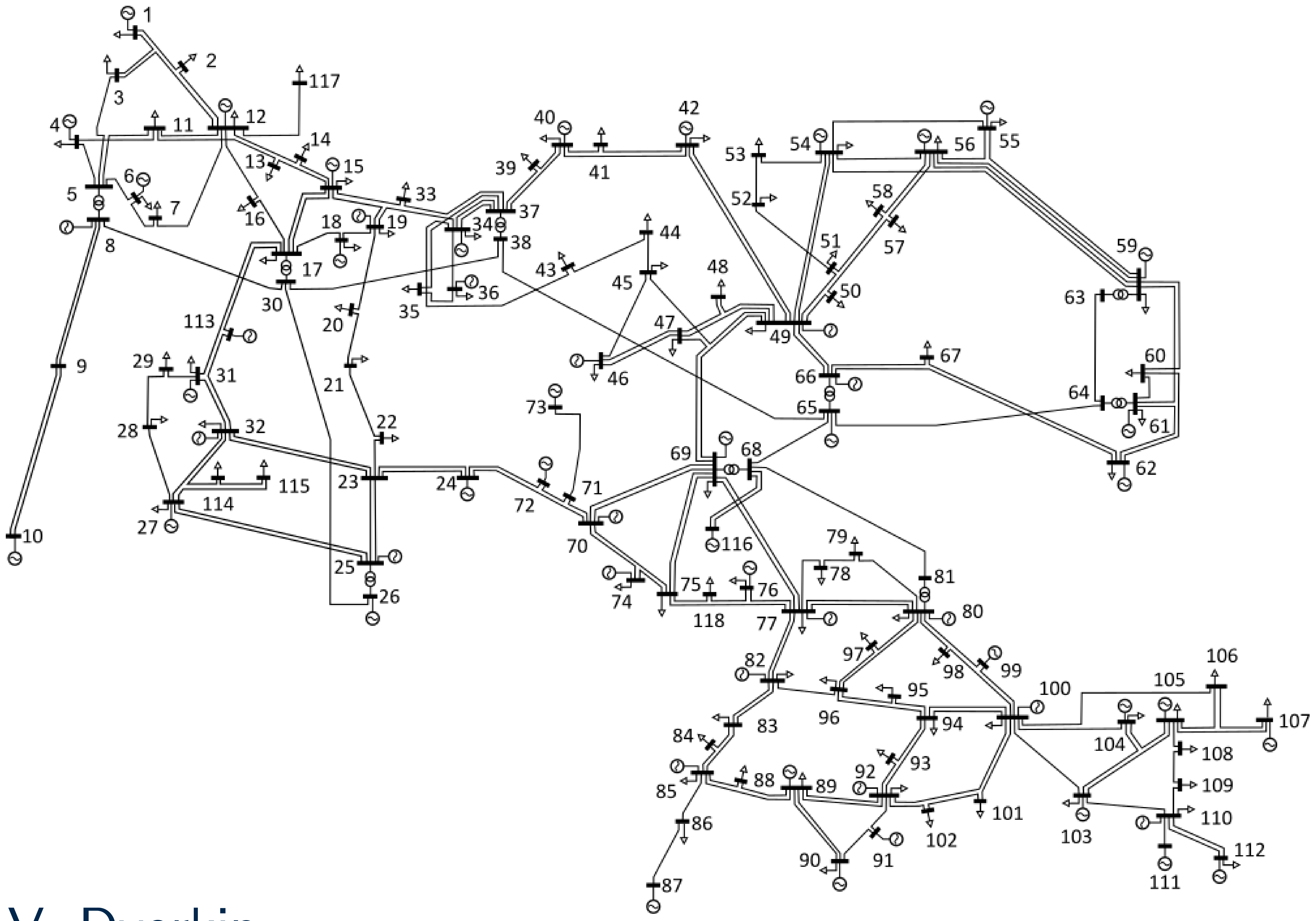


Power systems as a stroll in the fog

- Power systems are critical infrastructures with most of data being classified
- We have only a limited observably, e.g., MISO data disclosure portal
- Market participants hence act on a limited set of system data

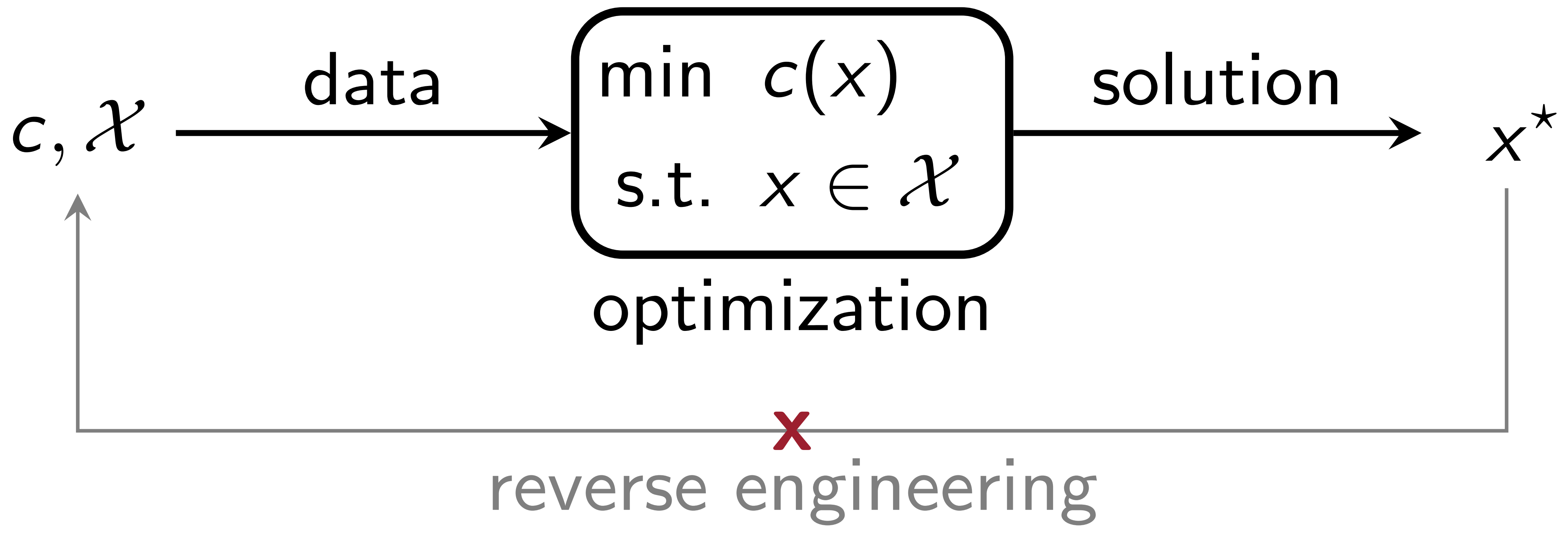
Hedgehog in the Fog
Yuri Norstein (1975)

Example: market clearing in the (small) IEEE 118-Bus system

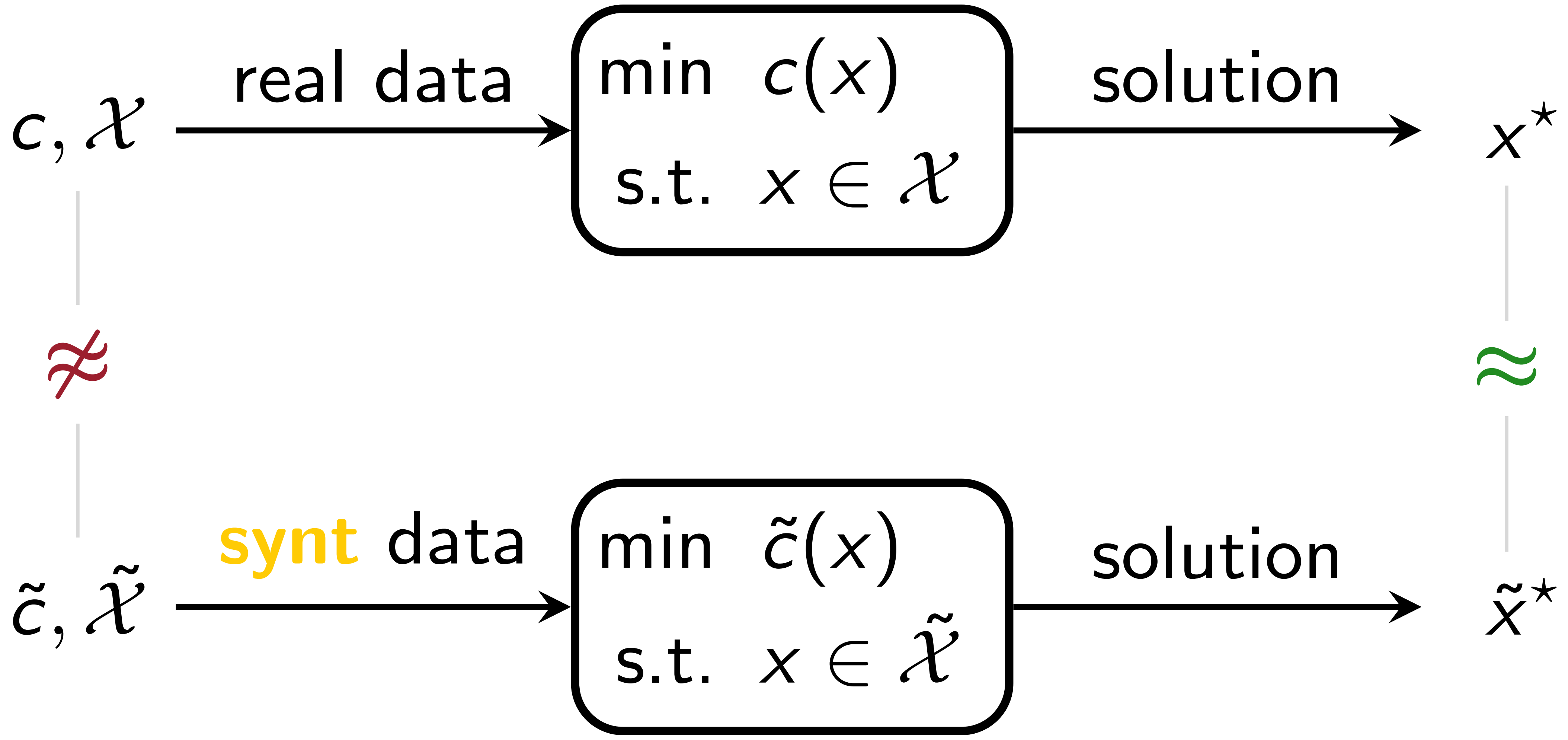


- 1079 rows of element-specific data
- Each generator owns only 2 rows
- The rest of the system remains unknown

In today's talk



Privacy-preserving optimization that does not leak data



Privacy-preserving synthetic datasets with consistent performance

Agenda

1. Introduction

2. Differential privacy basics

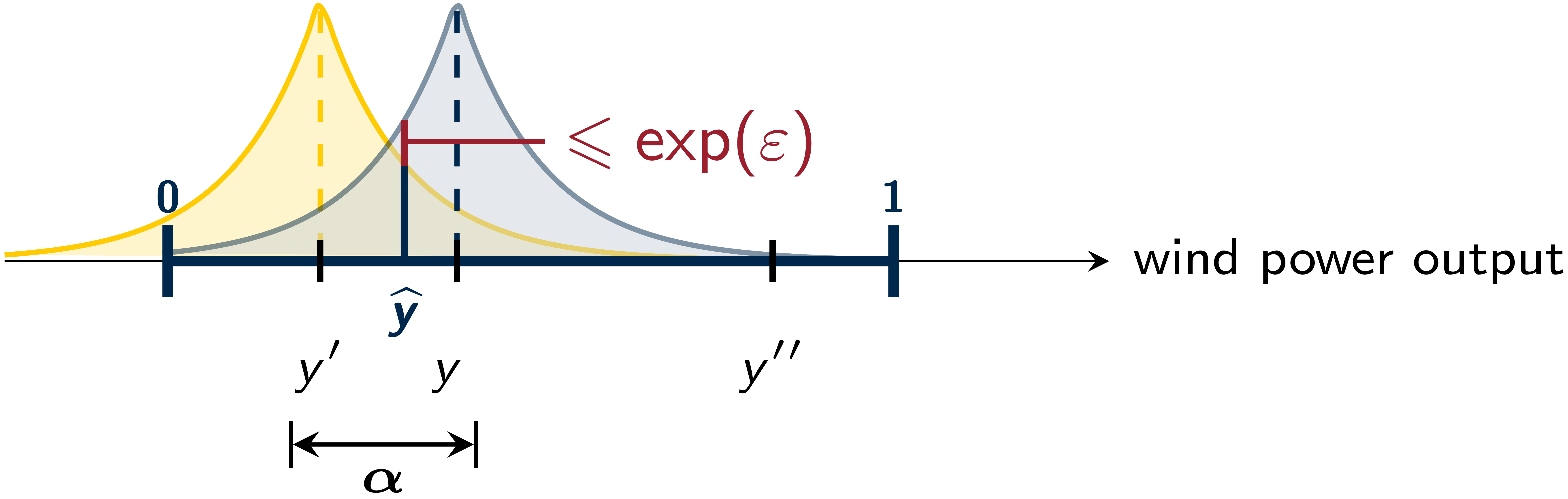
3. Privacy-preserving optimization via stochastic programming

4. Privacy-preserving synthetic dataset generation

5. Outlook

Agenda

1. Introduction
2. Differential privacy basics
3. Privacy-preserving optimization via stochastic programming
4. Privacy-preserving synthetic dataset generation
5. Outlook



- Wind power records $y, y', y'', \ldots \in [0, 1]$
- For given $\boldsymbol{\alpha} > 0$, records y and y' are $\boldsymbol{\alpha}-$adjacent if $\|y - y'\| \leqslant \boldsymbol{\alpha}$

- Let $\text{Lap}(\boldsymbol{\alpha}/\varepsilon)$ be a zero-mean random Laplacian noise
- For some parameter $\varepsilon > 0$, the release is $\varepsilon-$DP if

$$\frac{\Pr[\, y \;+ \text{Lap}(\boldsymbol{\alpha}/\varepsilon) \in \widehat{\boldsymbol{y}}\,]}{\Pr[\, y' + \text{Lap}(\boldsymbol{\alpha}/\varepsilon) \in \widehat{\boldsymbol{y}}\,]} \leqslant \exp(\varepsilon)$$

for any $\alpha-$adjacent pair (y, y') and any outcome $\widehat{\boldsymbol{y}}$

DP basics: Laplace mechanism

- Function $f : \mathbb{D} \mapsto \mathbb{R}$ mapping datasets from data universe $\mathbb{D}$ to reals
- Worst-case sensitivity Δf of function f to datasets, i.e., $\Delta f = \max_{\mathcal{D} \sim_\alpha \mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_1$

Laplace mechanism of DP

Perturbed function $\tilde{f}(\cdot) = f(\cdot) + \text{Lap}\left(\mu = 0, b = \frac{\Delta f}{\varepsilon}\right)$ is $\varepsilon-$DP for datasets, i.e.,

$$\frac{\Pr\left[\tilde{f}(\mathcal{D}) \in \widehat{\boldsymbol{y}}\right]}{\Pr\left[\tilde{f}(\mathcal{D}') \in \widehat{\boldsymbol{y}}\right]} \leqslant \exp(\varepsilon)$$

for any pair $\mathcal{D}, \mathcal{D}' \in \mathbb{D}$ and any outcome $\widehat{y}$.

- **Composition**: a series of $\tilde{f}_1(\mathcal{D}), \dots, \tilde{f}_k(\mathcal{D})$ of $\varepsilon-$DP computations ensures $k\varepsilon-$DP
- **Immunity to post-processing**: if $\tilde{f}(\mathcal{D})$ is $\varepsilon-$DP, then $g \circ \tilde{f}(\mathcal{D})$ is also $\varepsilon-$DP

- Finite population of functions $f_1, \ldots, f_k : \mathbb{D} \mapsto \mathbb{R}$ mapping datasets from universe $\mathbb{D}$ to reals
- Worst-case sensitivity $\Delta = \max_i \Delta f_i$ of functions to datasets

Report noisy max

What function takes the maximum value on a private dataset $\mathcal{D}$?

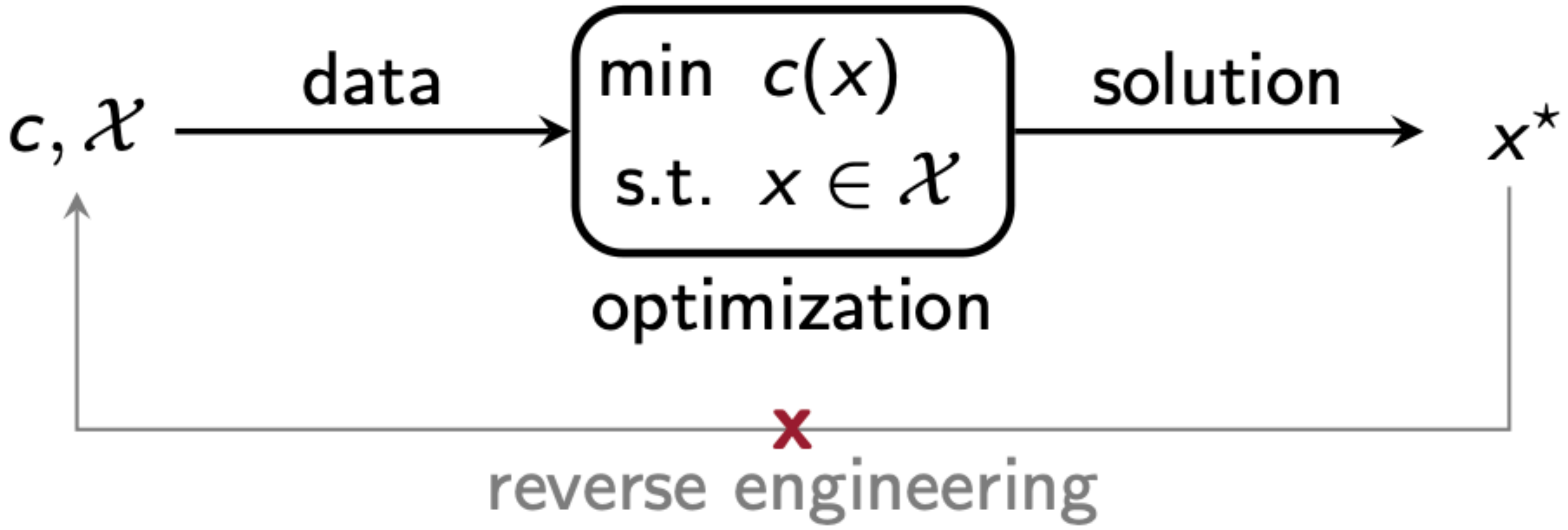
for $i = 1, \ldots, k$
  | $\tilde{f}_i(\mathcal{D}) = f_i(\mathcal{D}) + \text{Lap}\,(\Delta/\varepsilon)$
end
return: $i^\star \in \underset{i}{\text{argmax}}\ \tilde{f}_i(\mathcal{D})$

Releasing index $i^\star$ satisfies $\varepsilon-$DP (despite k computations on private data!)

- **Use case**: what is the worst-case optimization model for a given (private) dataset?

Agenda

1. Introduction
2. Differential privacy basics
3. Privacy-preserving optimization via stochastic programming
4. Privacy-preserving synthetic dataset generation
5. Outlook



Limits of DP applications to convex optimization

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & b - Ax \in \mathcal{K} \end{aligned}$$

- Conic optimization program
- Optimization dataset $\mathcal{D} = \{c, b, A\}$
- Optimal solution $x^\star$ is dataset-specific
- Often, $x^\star(\mathcal{D}) \neq x^\star(\mathcal{D}')$ for different datasets $\mathcal{D}$ and $\mathcal{D}'$

Input perturbation

1. Optimization dataset perturbation

$$\tilde{\mathcal{D}} = \mathcal{D} + \zeta, \quad \zeta \sim \text{Lap}(\alpha/\varepsilon)$$

2. Optimization on perturbed data $x^\star(\tilde{\mathcal{D}})$

Output perturbation

1. Worst-case sensitivity computation

$$\Delta_\alpha = \max_{\mathcal{D},\mathcal{D}' \in \mathbb{D}} \left\| \tilde{x}^\star(\mathcal{D}) - \tilde{x}^\star(\mathcal{D}') \right\|_1$$

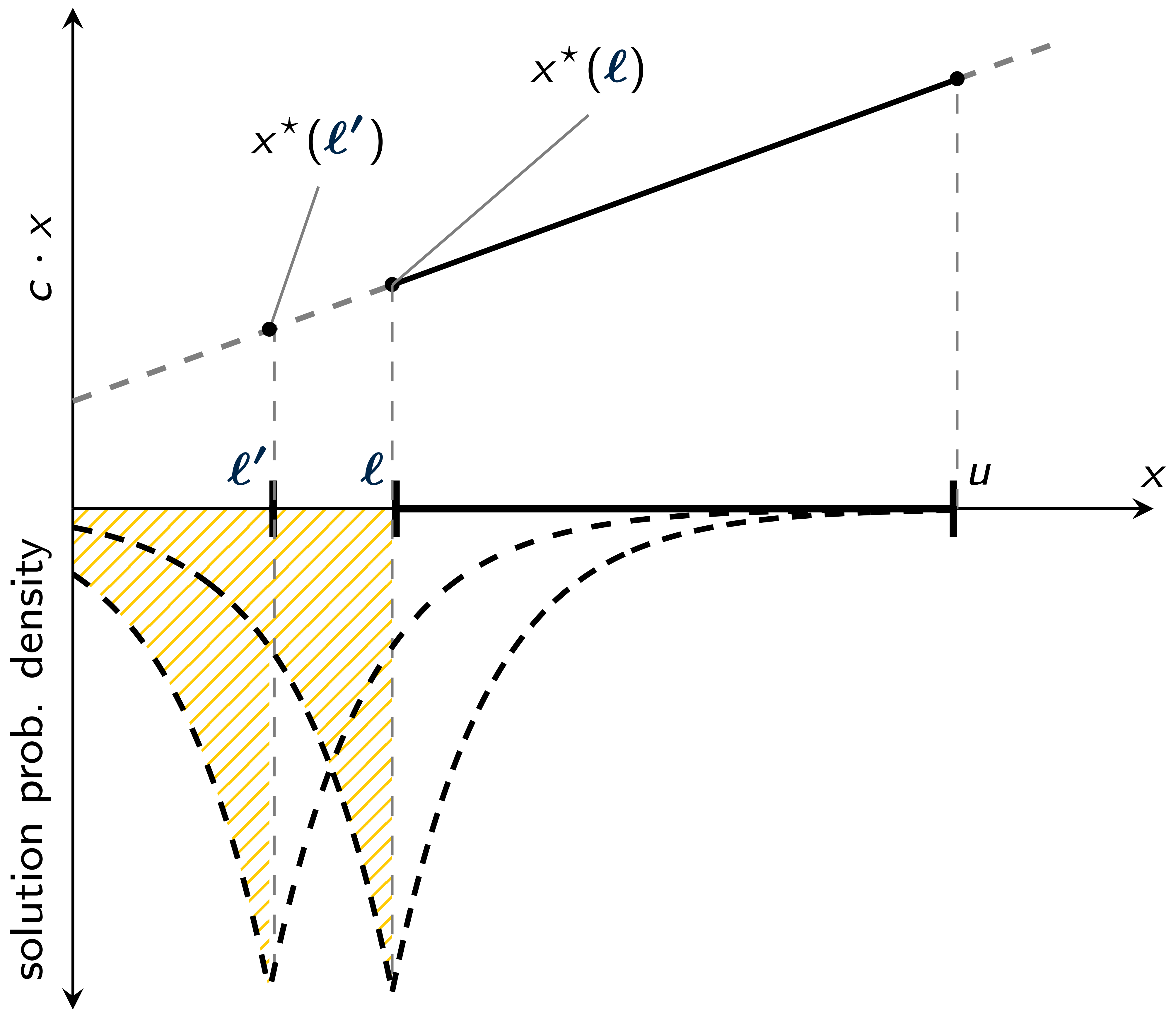
2. Perturbation of optimization results

$$\tilde{x}^\star(\mathcal{D}) = x^\star(\mathcal{D}) + \zeta, \quad \zeta \sim \text{Lap}(\Delta_\alpha/\varepsilon)$$

Linear programming example

$$\underset{x}{\text{minimize}} \quad c \cdot x$$
$$\text{subject to} \quad \ell \leqslant x \leqslant u,$$

- ℓ and ℓ' must be made indistinguishable
- input and output perturbation strategies are equivalent and yield **infeasible** results



Stochastic programming for private optimization queries

For a deterministic conic program, we device a chance-constrained (stochastic) counterpart

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & b - Ax \in \mathcal{K} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \underset{\overline{x}, X \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[c^\top(\overline{x} + X\zeta)\right] && \textbf{exp. cost} \\ \text{subject to} \quad & \Pr\left[b - A(\overline{x} + X\zeta) \in \mathcal{K}\right] \geqslant 1 - \eta && \textbf{feas. guarantee} \end{aligned}$$

Solution vector $x(\mathcal{D})$ is modeled as a **linear decision rule** of the form:

$$\tilde{x}(\mathcal{D}) = \overline{x}(\mathcal{D}) + X(\mathcal{D})\zeta$$

$\overline{x}$ – nominal solution vector

X – recourse matrix

ζ – zero-mean perturbation

Identity query $(\mathcal{X} = \{X | X = I\})$:

$$\tilde{x}(\mathcal{D}) = \overline{x}^\star(\mathcal{D}) + X^\star(\mathcal{D})\zeta = \overline{x}^\star(\mathcal{D}) + \zeta$$

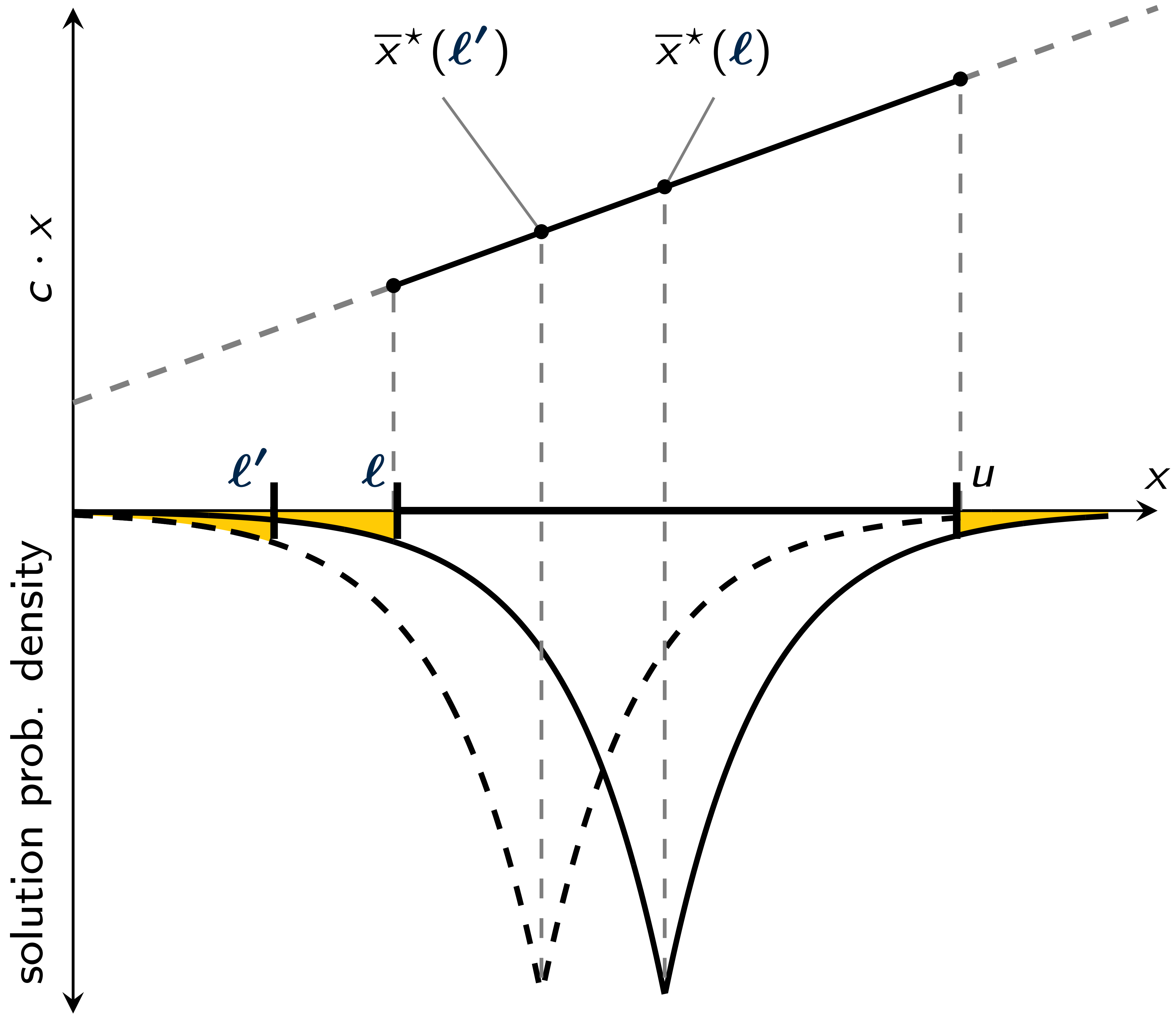
Sum query $(\mathcal{X} = \{X | 1^\top X = 1\})$:

$$1^\top \tilde{x}(\mathcal{D}) = 1^\top \overline{x}^\star(\mathcal{D}) + 1^\top X^\star(\mathcal{D})\zeta = 1^\top \overline{x}^\star(\mathcal{D}) + \zeta$$

Linear programming example

$$\underset{\overline{x}}{\text{minimize}} \quad \mathbb{E}\left[c \cdot (\overline{x} + \zeta)\right]$$
$$\text{subject to} \quad \Pr\left[\boldsymbol{\ell} \leqslant \overline{x} + \zeta \leqslant u\right] \geqslant 1 - \eta,$$

- Perturb. of $\overline{x}^\star$ is feasible with a high prob.
- Sub-optimal due to feasibility requirement



Main result: $\varepsilon-$differential privacy of the optimized linear decision rules

Deterministic program

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & b - Ax \in \mathcal{K} \end{aligned}$$

$\Longrightarrow$

Stochastic program

$$\begin{aligned} \underset{\overline{x}, X \in \mathcal{X}}{\text{minimize}} \quad & \mathbb{E}\left[c^\top(\overline{x} + X\zeta)\right] \\ \text{subject to} \quad & \Pr\left[b - A(\overline{x} + X\zeta) \in \mathcal{K}\right] \geqslant 1 - \eta \end{aligned}$$

Differential privacy of identity optimization queries

- dataset adjacency $\boldsymbol{\alpha}$
- solution sensitivity $\boldsymbol{\Delta}_{\alpha}$
- privacy budget $\boldsymbol{\varepsilon}$

$\Longrightarrow$ $\zeta \sim \text{Lap}(\boldsymbol{\Delta}_{\alpha}/\boldsymbol{\varepsilon})$ $\Longrightarrow$

$$\frac{\Pr[\overline{x}^\star(\mathcal{D}) + X^\star(\mathcal{D})\zeta = \widehat{x}]}{\Pr[\overline{x}^\star(\mathcal{D}') + X^\star(\mathcal{D}')\zeta = \widehat{x}]} \leqslant \exp(\varepsilon)$$

- for any optimization outcome $\widehat{x}$
- and $\alpha-$adjacent dataset pair $(\mathcal{D}, \mathcal{D}')$

Private optimal power flow (OPF) problem

$$\min_{\overline{x},X\in\mathcal{X}} \quad \mathbb{E}[c^{\top}(\overline{x}+X\boldsymbol{\zeta})] \qquad \textit{expected generation cost}$$

$$\text{s.t.} \quad 1^{\top}(\overline{x}+X\boldsymbol{\zeta}-\boldsymbol{d})=0 \qquad \textit{perturbed power balance}$$

$$\Pr\begin{bmatrix}|F(\overline{x}+X\boldsymbol{\zeta}-\boldsymbol{d})|\leqslant f^{\max} \\ x^{\min}\leqslant\overline{x}+X\boldsymbol{\zeta}\leqslant x^{\max}\end{bmatrix}\geqslant 1-\eta \qquad \textit{stochastic network limits}$$

- Load vector $\boldsymbol{d}$ (in MWh) is private information
- Must be stat. similar to any $\alpha-$adjacent load $\boldsymbol{d}'$

- Queries in electricity markets
  - System costs (objective function)
  - Generation by a particular technology

1−DP system cost query on the IEEE 24-Bus RTS

perturbation strategy	OPF infeasibility (%)			OPF sub-optimality (%)		
	$\alpha=1$	$\alpha=3$	$\alpha=10$	$\alpha=1$	$\alpha=3$	$\alpha=10$
input	51.5	49.9	50.3	0.0	0.1	0.0
output	52.7	51.5	48.8	0.0	0.0	0.1
program	0.1	0.1	0.1	1.7	5.1	17.1

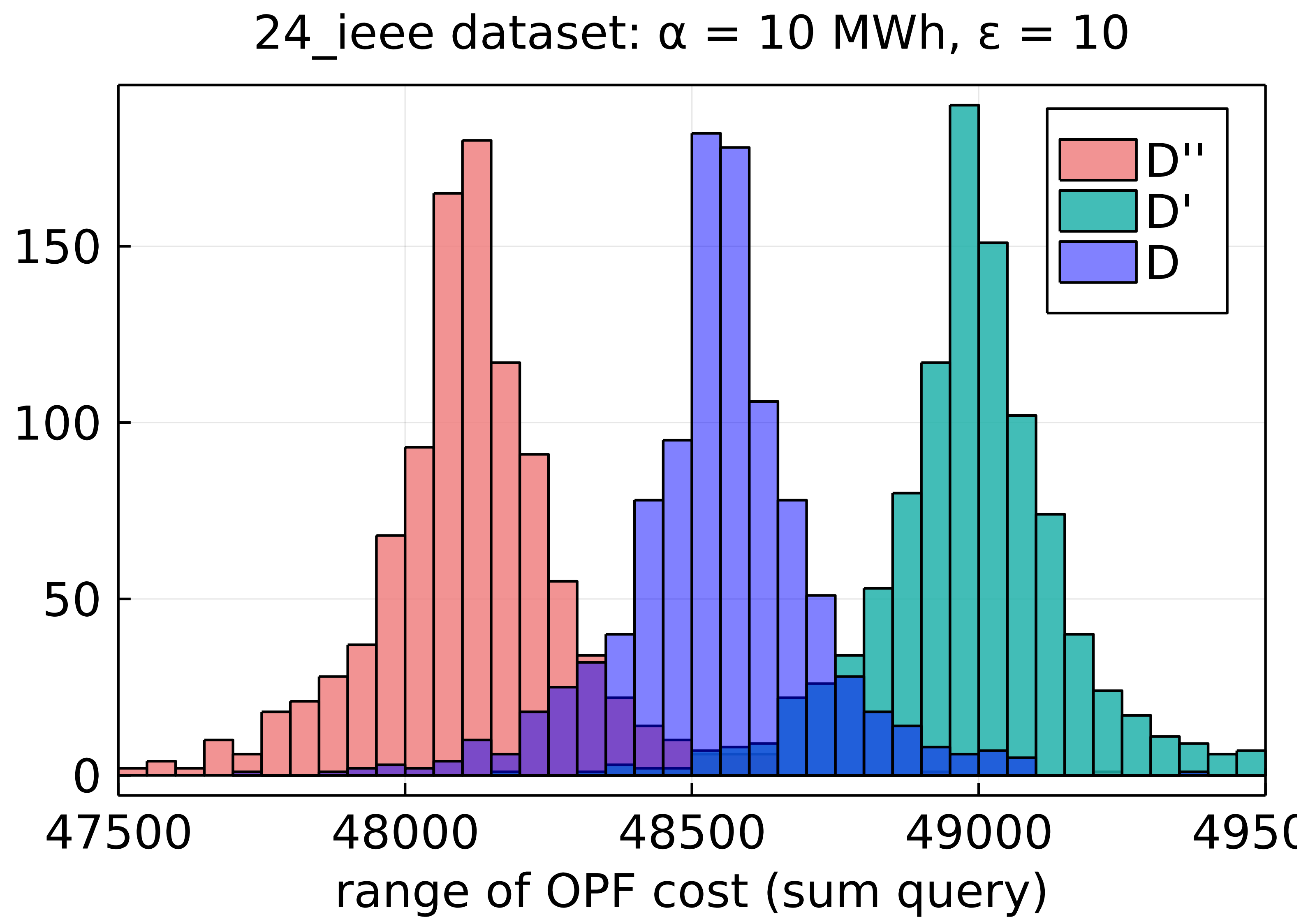
$$\min_{\overline{x}, X \in \mathcal{X}} \quad \mathbb{E}[c^\top(\overline{x} + X\zeta)] \qquad \textit{expected generation cost}$$

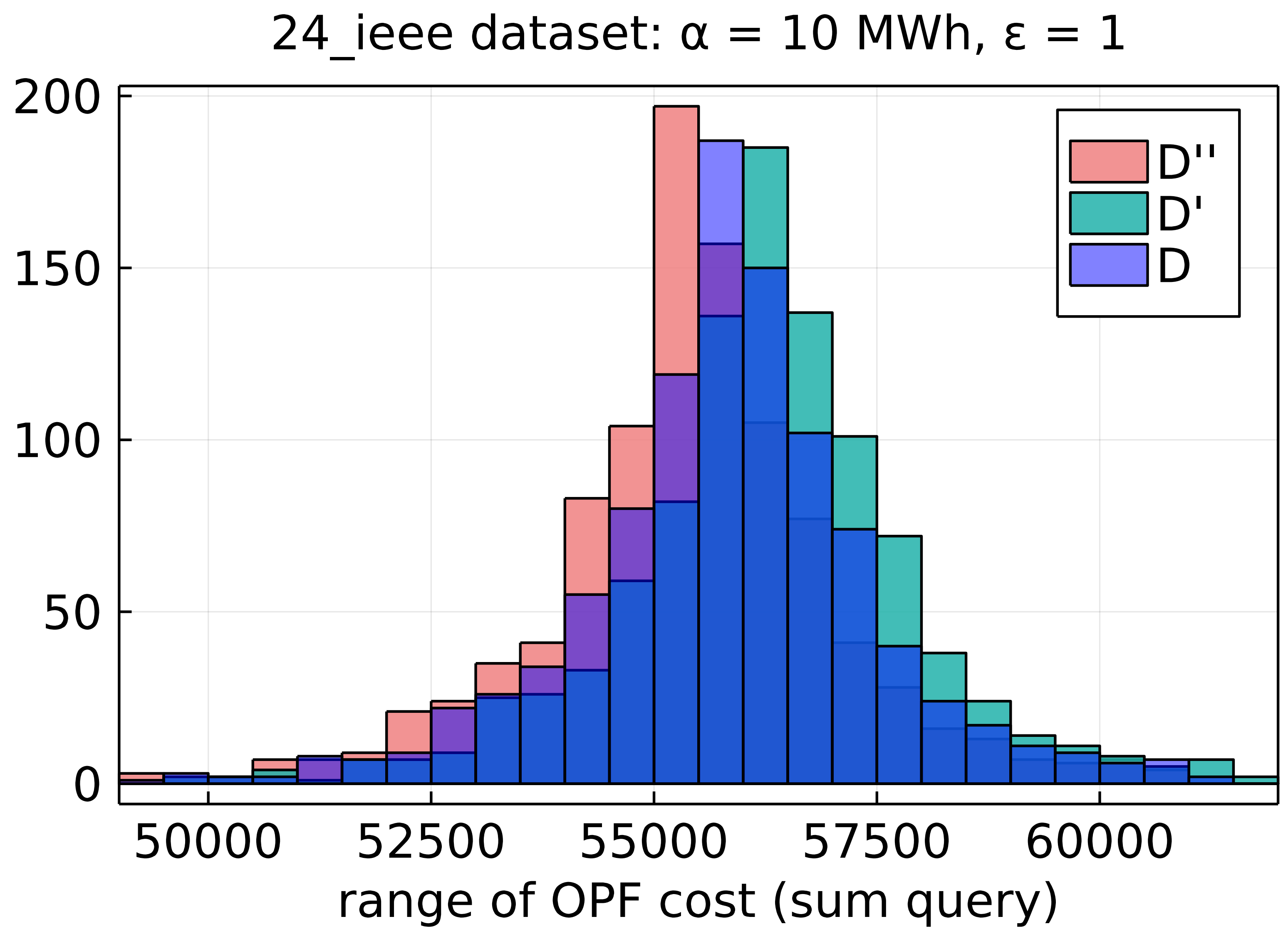
$$\text{s.t.} \quad 1^\top(\overline{x} + X\zeta - \boldsymbol{d}) = 0 \qquad \textit{perturbed power balance}$$

$$\Pr\begin{bmatrix} |F(\overline{x} + X\zeta - \boldsymbol{d})| \leqslant f^{\max} \\ x^{\min} \leqslant \overline{x} + X\zeta \leqslant x^{\max} \end{bmatrix} \geqslant 1 - \eta \qquad \textit{stochastic network limits}$$

- Load vector $\boldsymbol{d}$ (in MWh) is private information
- Must be stat. similar to any $\alpha-$adjacent load $\boldsymbol{d}'$

- Queries in electricity markets
  - System costs (objective function)
  - Generation by a particular technology



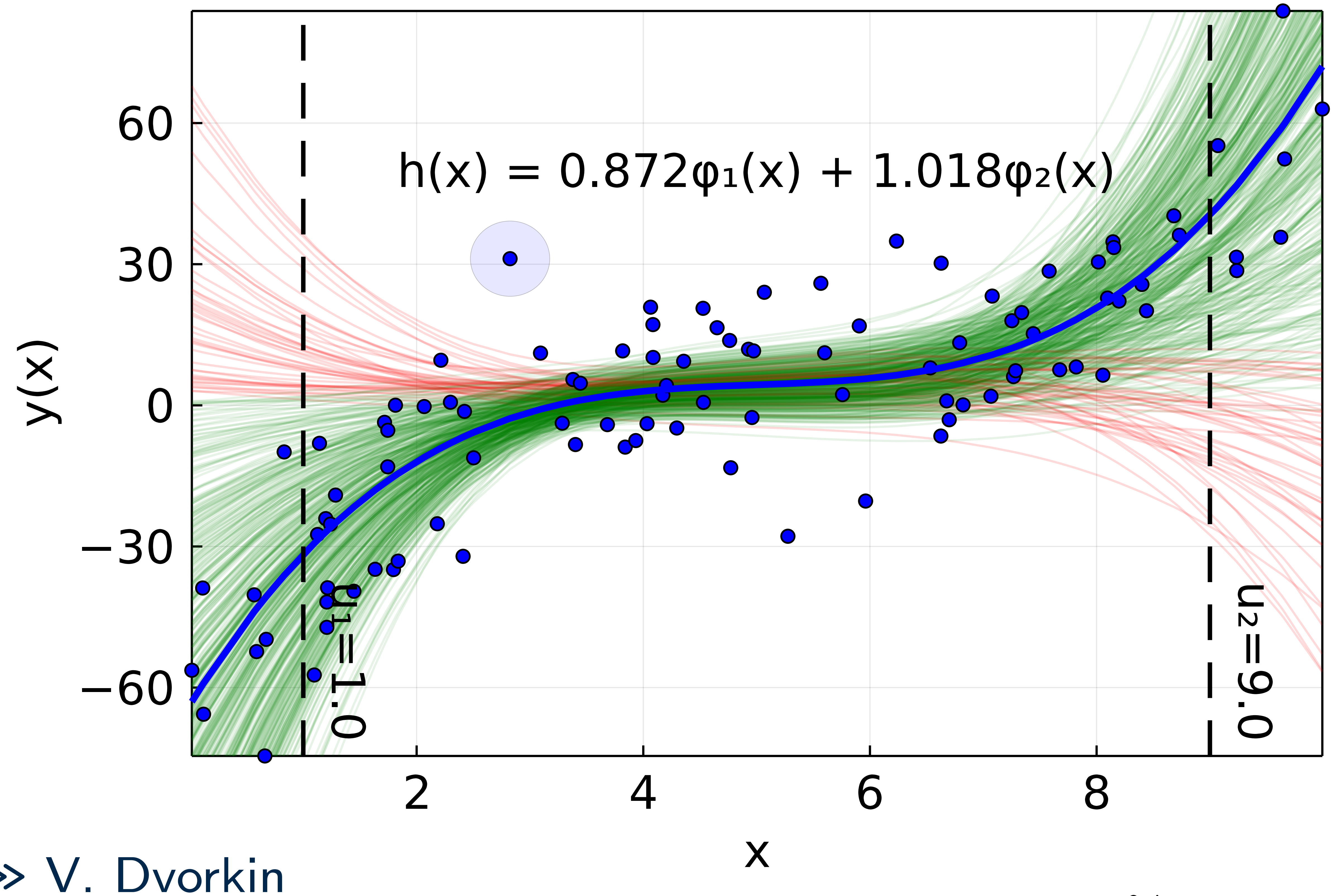


$$\min_{\beta} \quad \mathbb{E}\left[\sum_{i=1}^{n}\left(\underbrace{y_i - \varphi(x_i)^\top\beta}_{\text{business as usual}} \underbrace{-\varphi(x_i)^\top\zeta}_{\text{perturbation}}\right)^2\right]$$

$$\text{s.t.} \quad \mathbb{P}\left[C(\beta+\zeta) \geqslant 0\right] \geqslant 1-\eta,$$

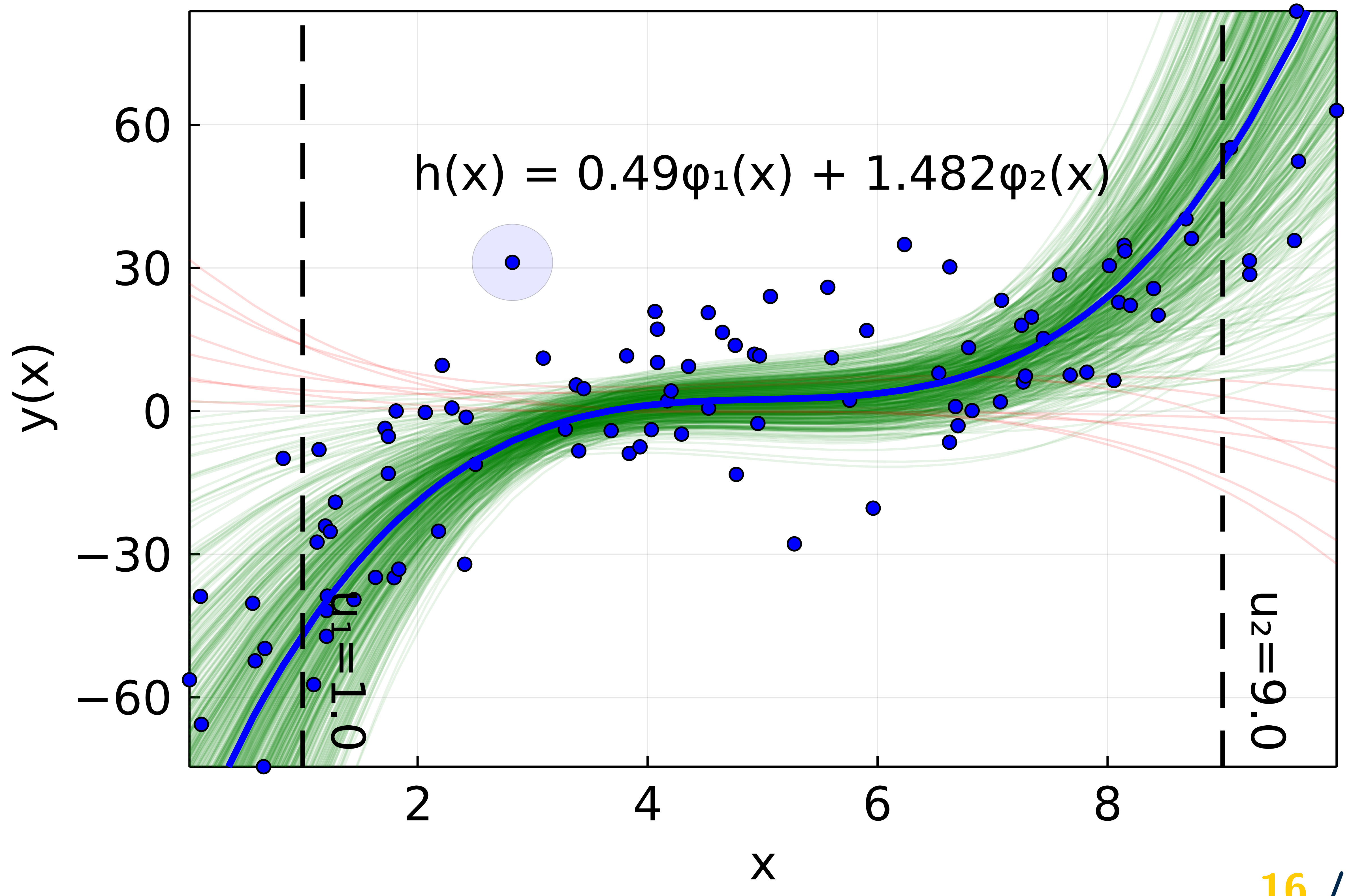
- Dataset $\{(y_1, x_1), \ldots, (y_n, x_n)\}$
- Minimize regression loss function
- By finding optimal weights $\beta^\star$ ...
- ... of basis functions in vector $\varphi(x)$

output perturbation strategy



infeasible curve with probability 0.8%

program perturbation strategy



infeasible curve with probability 1.2%

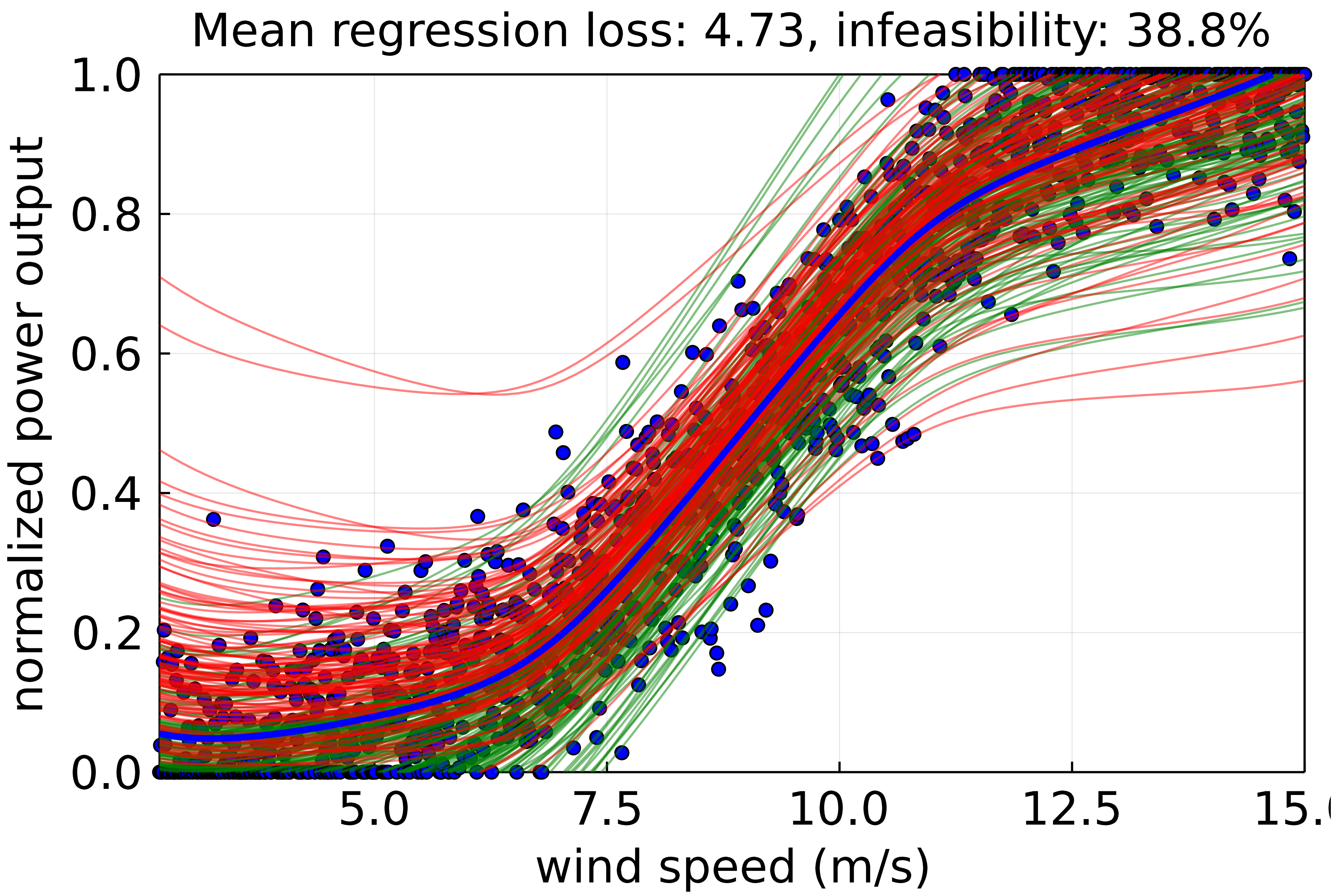
Private monotone curve fitting

$$\min_{\beta} \quad \mathbb{E}\left[\sum_{i=1}^{n}\left(\underbrace{y_i - \varphi(x_i)^\top\beta}_{\text{business as usual}} \underbrace{-\varphi(x_i)^\top\zeta}_{\text{perturbation}}\right)^2\right]$$

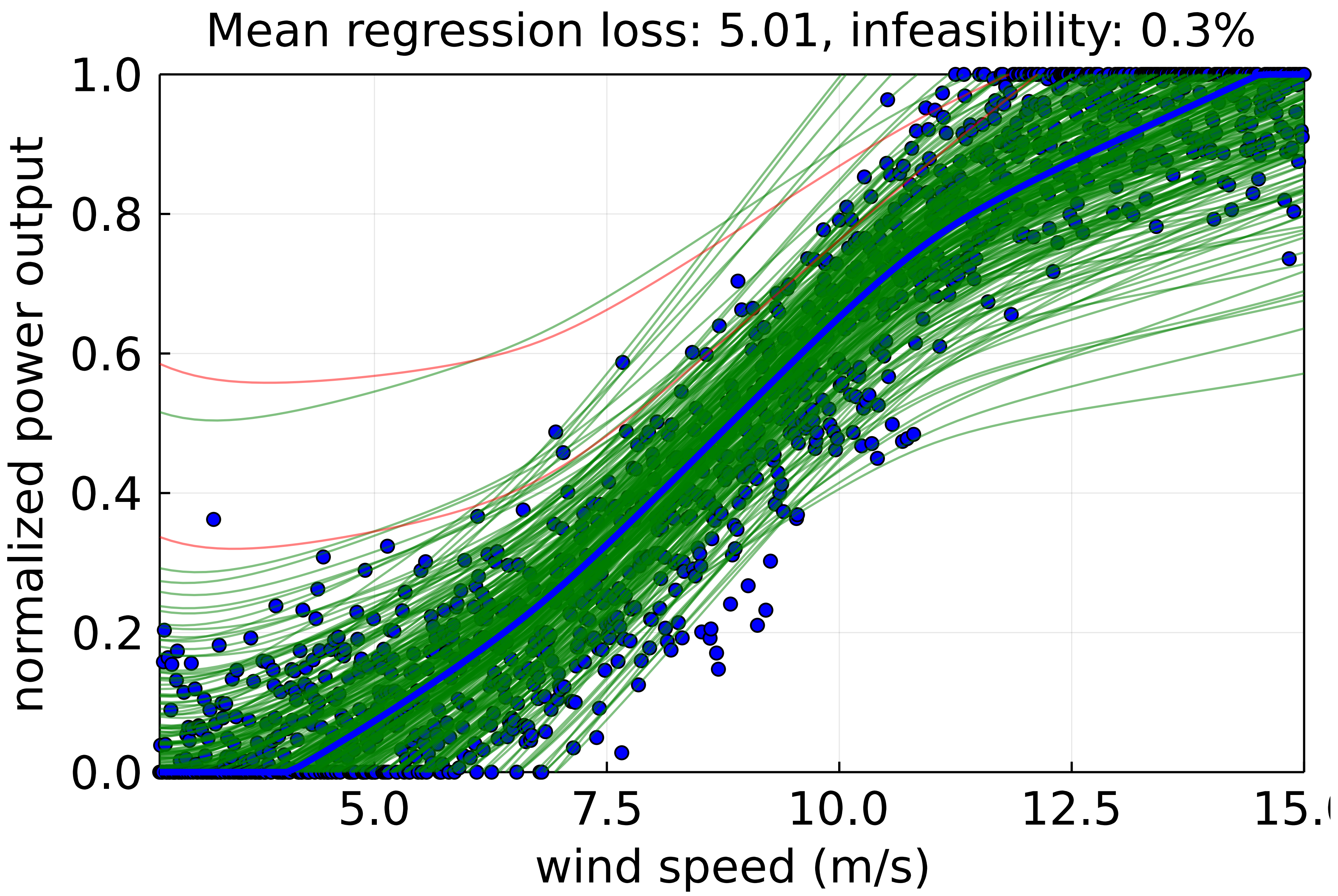
$$\text{s.t.} \quad \mathbb{P}\left[C(\beta+\zeta) \geqslant 0\right] \geqslant 1-\eta,$$

- Dataset $\{(y_1, x_1), \ldots, (y_n, x_n)\}$
- Minimize regression loss function
- By finding optimal weights $\beta^\star$ ...
- ... of basis functions in vector $\varphi(x)$

Alstom.Eco.80



output perturbation



program perturbation

Private support vector machine (SVM) for classification

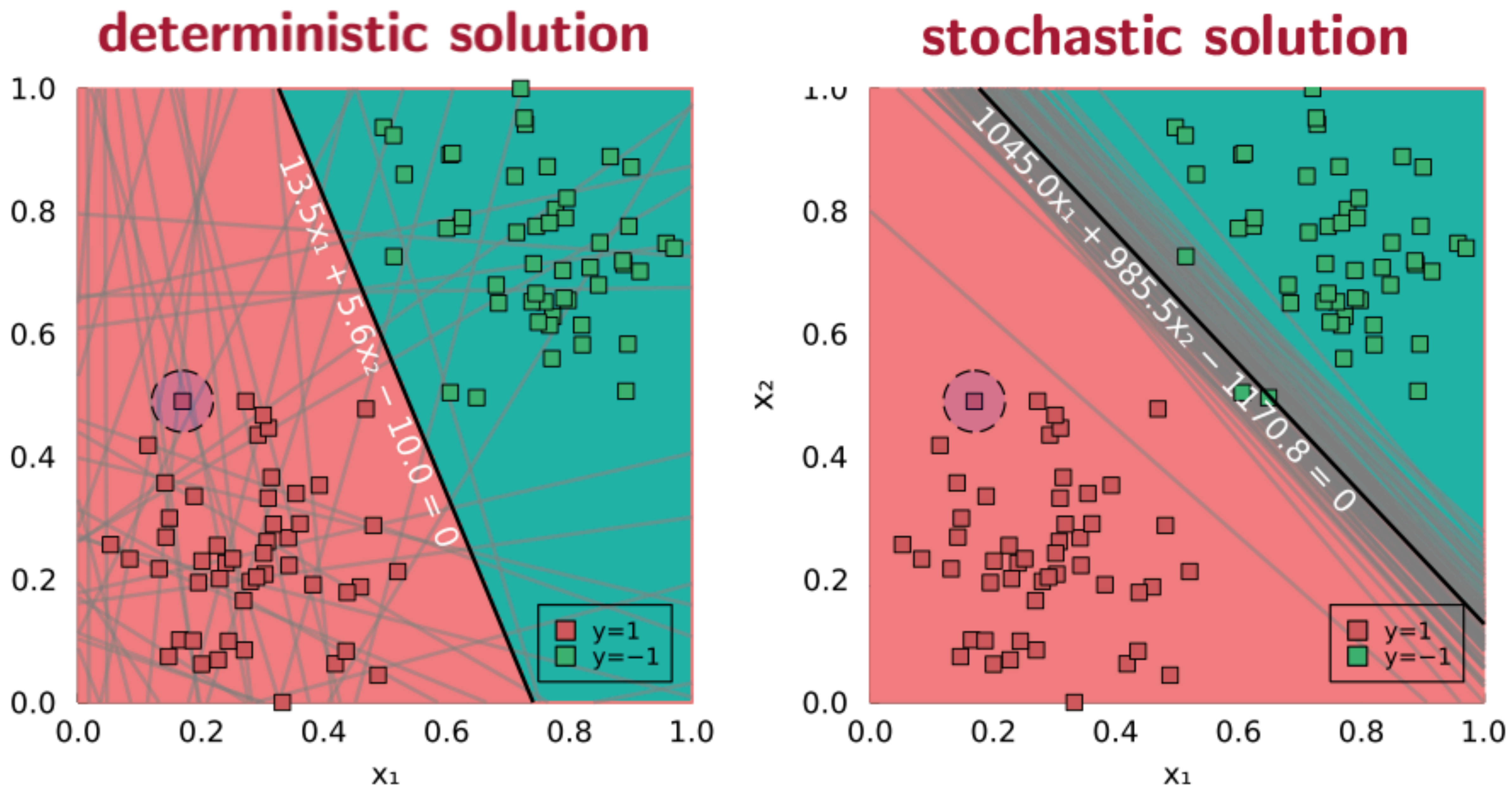
- Dataset $(x_1, y_1), \ldots, (x_m, y_m)$
- Feature $x_i \in \mathbb{R}^n$, label $y_i \in \{-1, 1\}$

- Computes a hyperplane $w^\top x_i - b$
- Classification rule $\text{sign}[w^{\star\top}\hat{x} - b^\star]$

$$\min_{\tilde{b}(\zeta),\tilde{w}(\zeta),z} \quad \mathbb{E}\left[\lambda \|\overline{w}\|^2 + \tfrac{1}{m}\mathbf{1}^\top z + \lambda \|W\zeta\|^2\right]$$

$$\text{s.t. Pr}\left[\begin{array}{l} y_i(\overline{w}^\top x_i - \overline{b}) \geqslant 1 - z_i - y_i((W\zeta)^\top x_i - B\zeta), \\ z_i \geqslant 0, \quad \forall i = 1, \ldots, m \end{array}\right] \geqslant 1 - \eta, \qquad \begin{bmatrix} W \\ B \end{bmatrix} = I$$

- Quering hyperplane parameters
- Deterministic hyperplane is very sensitive to perturbation
- Stochastic hyperplane, in contrast, is very robust to perturbation



mean accuracy 51.2% mean accuracy 97.6%

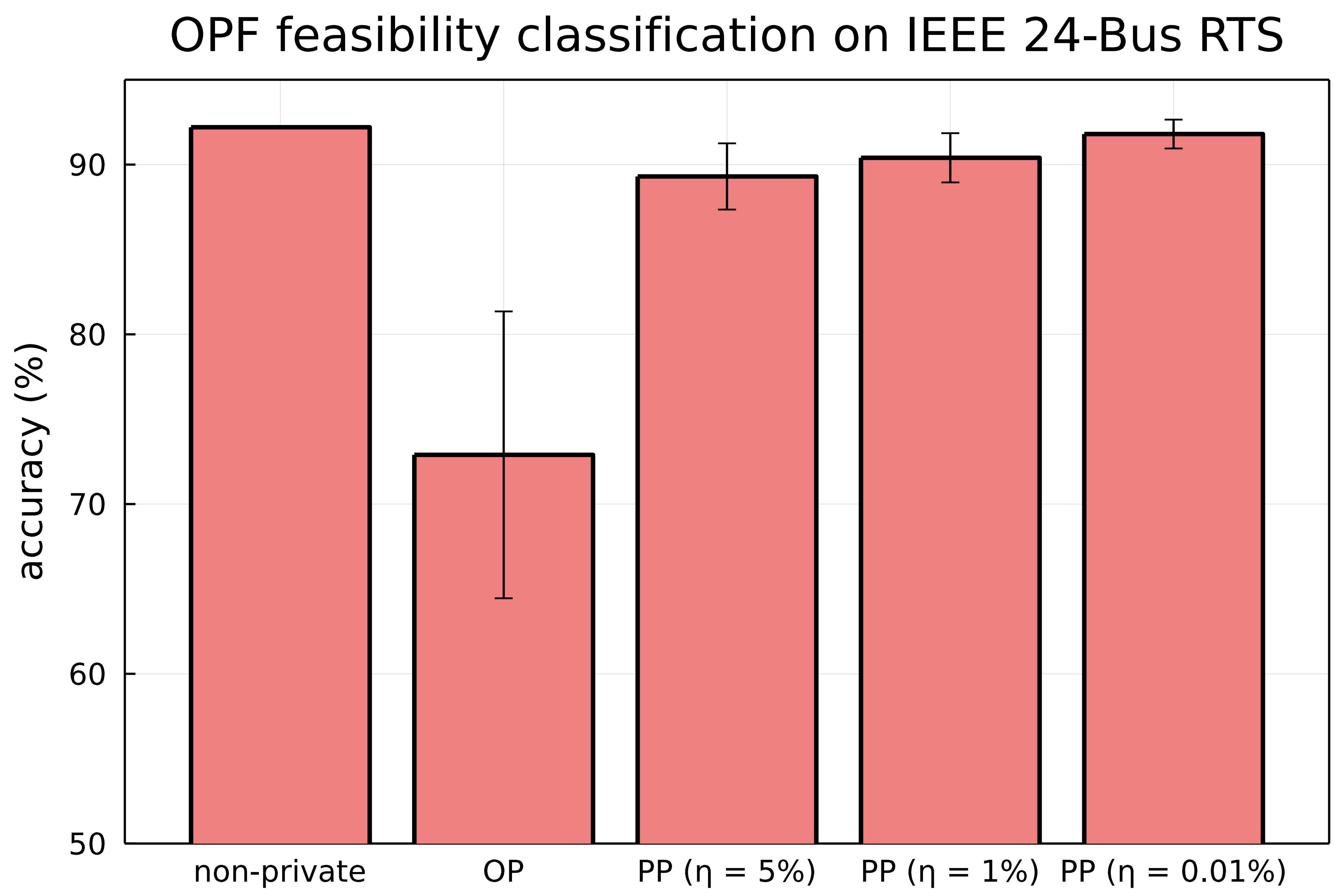
Private support vector machine (SVM) for classification

- Dataset $(x_1, y_1), \ldots, (x_m, y_m)$
- Feature $x_i \in \mathbb{R}^n$, label $y_i \in \{-1, 1\}$
- Computes a hyperplane $w^\top x_i - b$
- Classification rule $\text{sign}[w^{\star\top}\hat{x} - b^\star]$

$$\min_{\tilde{b}(\zeta), \tilde{w}(\zeta), z} \quad \mathbb{E}\left[\lambda \|\overline{w}\|^2 + \tfrac{1}{m}\mathbf{1}^\top z + \lambda \|W\zeta\|^2\right]$$

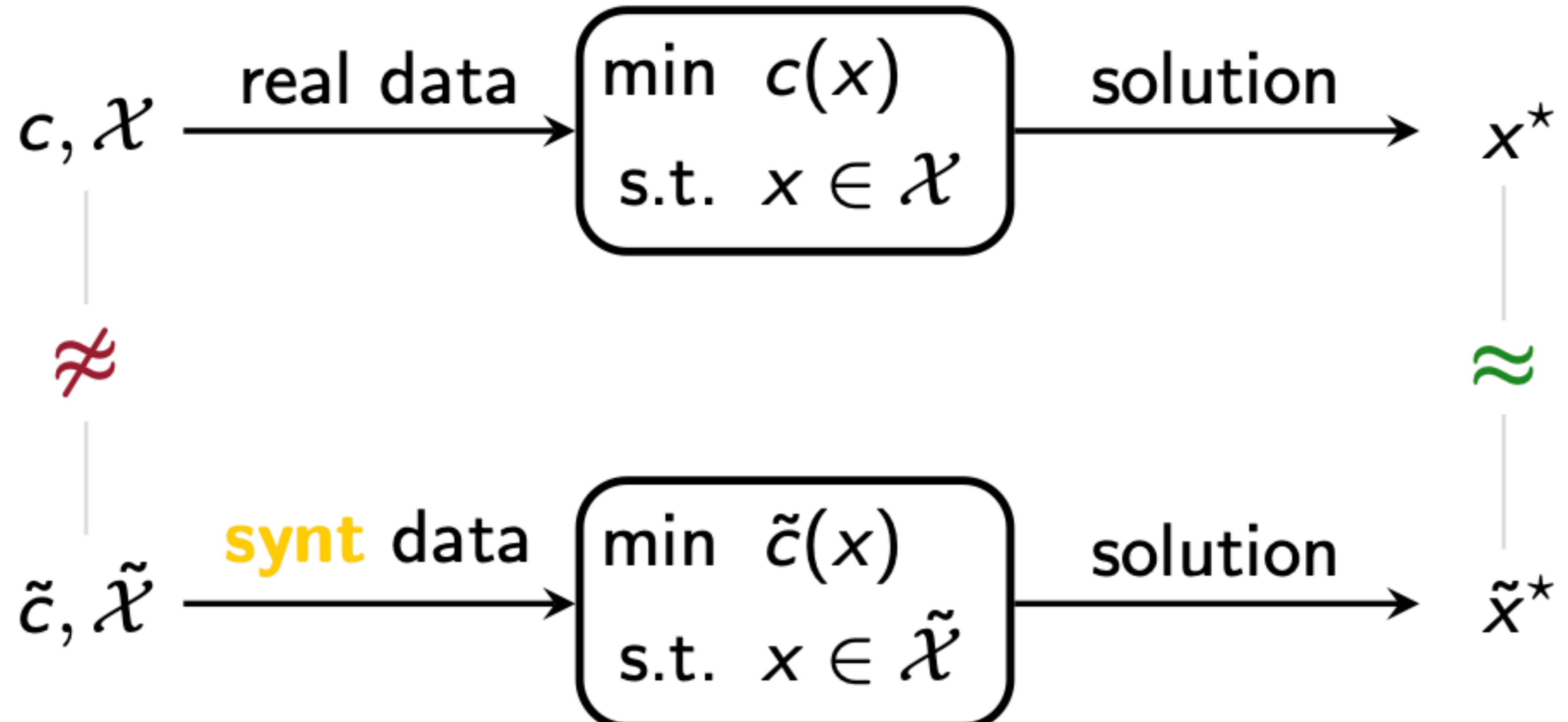
$$\text{s.t. Pr}\begin{bmatrix} y_i(\overline{w}^\top x_i - \overline{b}) \geqslant 1 - z_i - y_i((W\zeta)^\top x_i - B\zeta), \\ z_i \geqslant 0, \quad \forall i = 1, \ldots, m \end{bmatrix} \geqslant 1 - \eta, \quad \begin{bmatrix} W \\ B \end{bmatrix} = I$$

- Load data to classify OPF feasibility
- Output perturbation (OP) accuracy is small
- Program perturbation (PP) accuracy high and improves with a smaller constraint violation prob. η

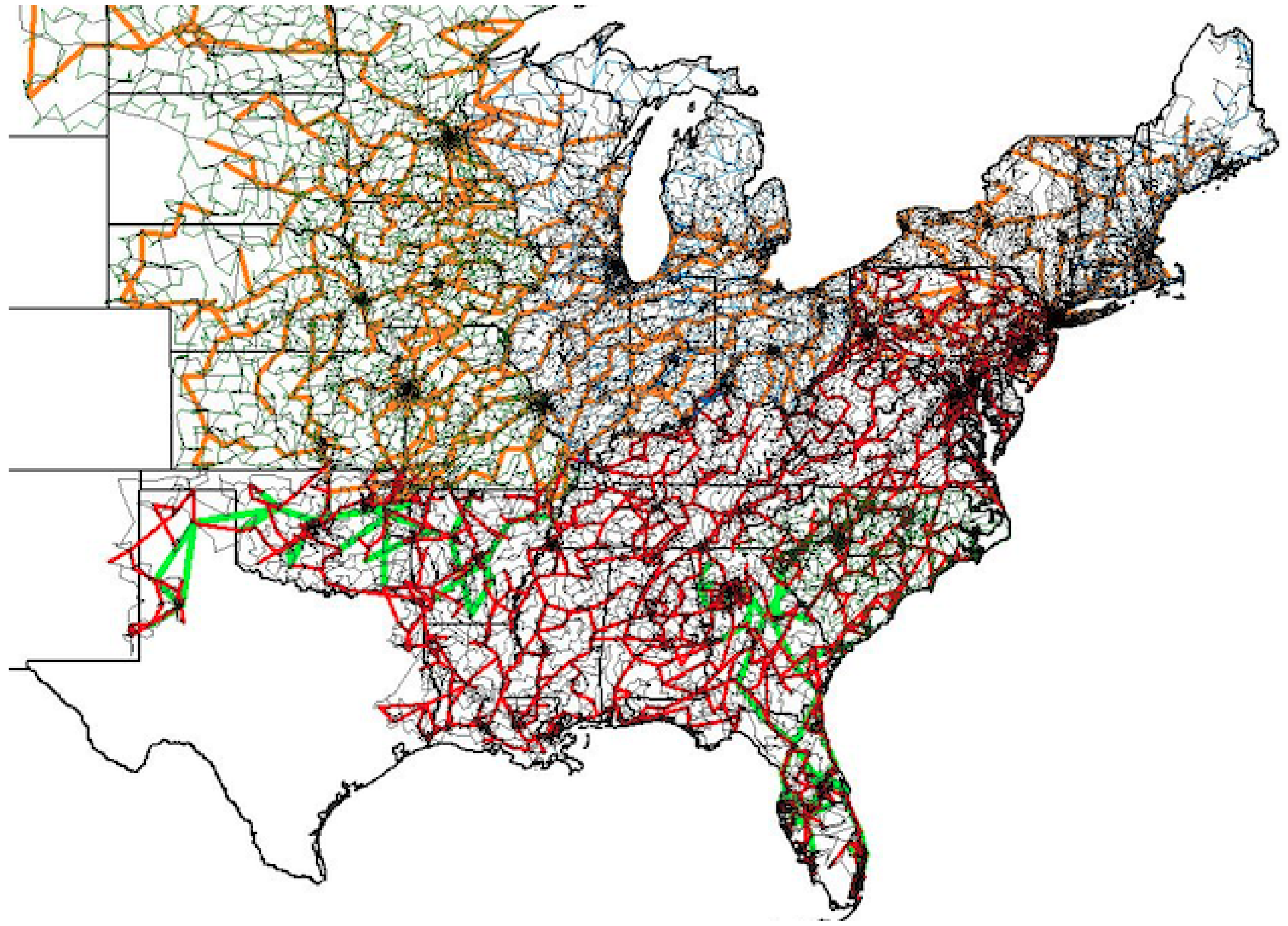


Agenda

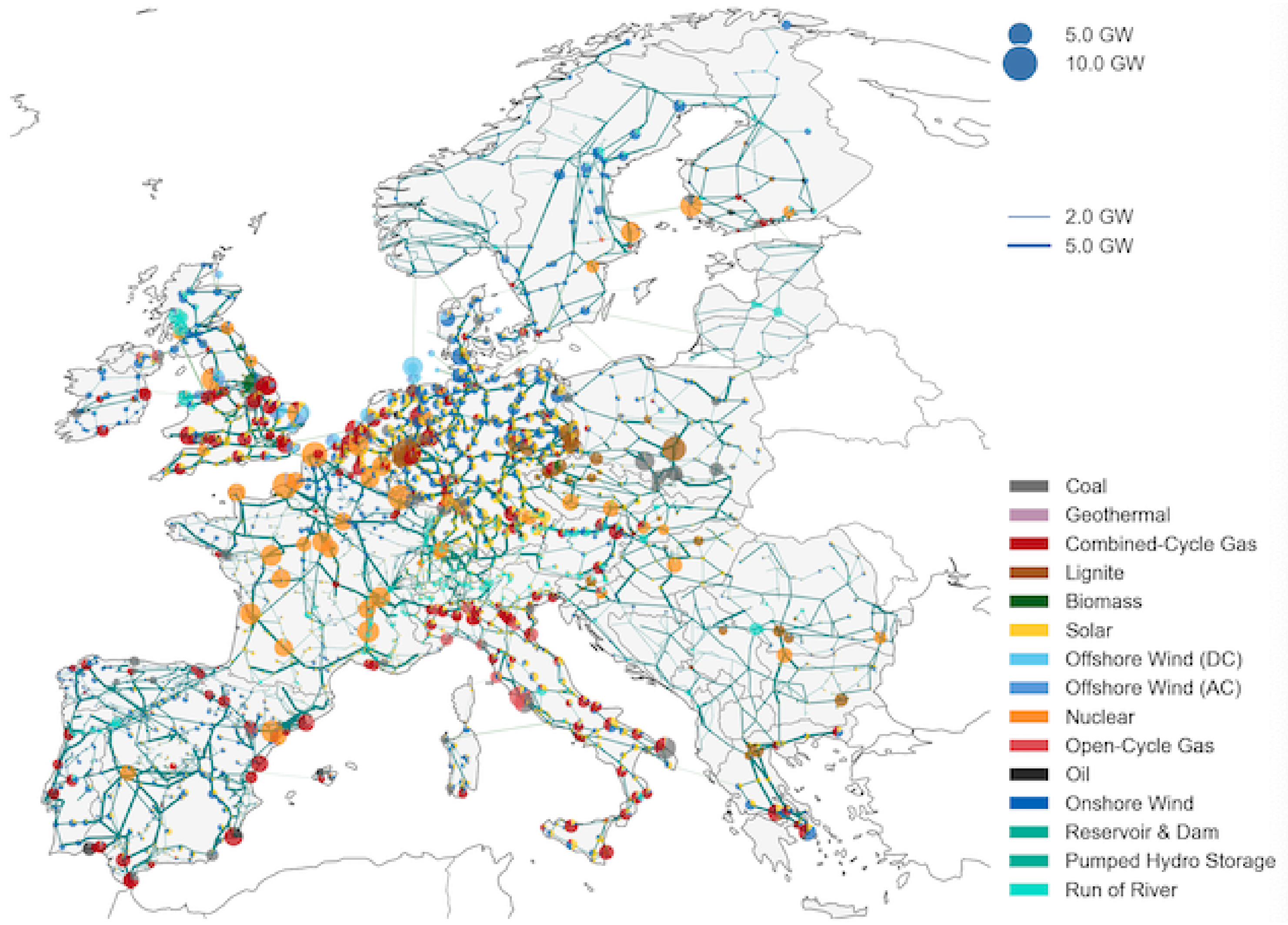
1. Introduction
2. Differential privacy basics
3. Privacy-preserving optimization via stochastic programming
4. Privacy-preserving synthetic dataset generation
5. Outlook



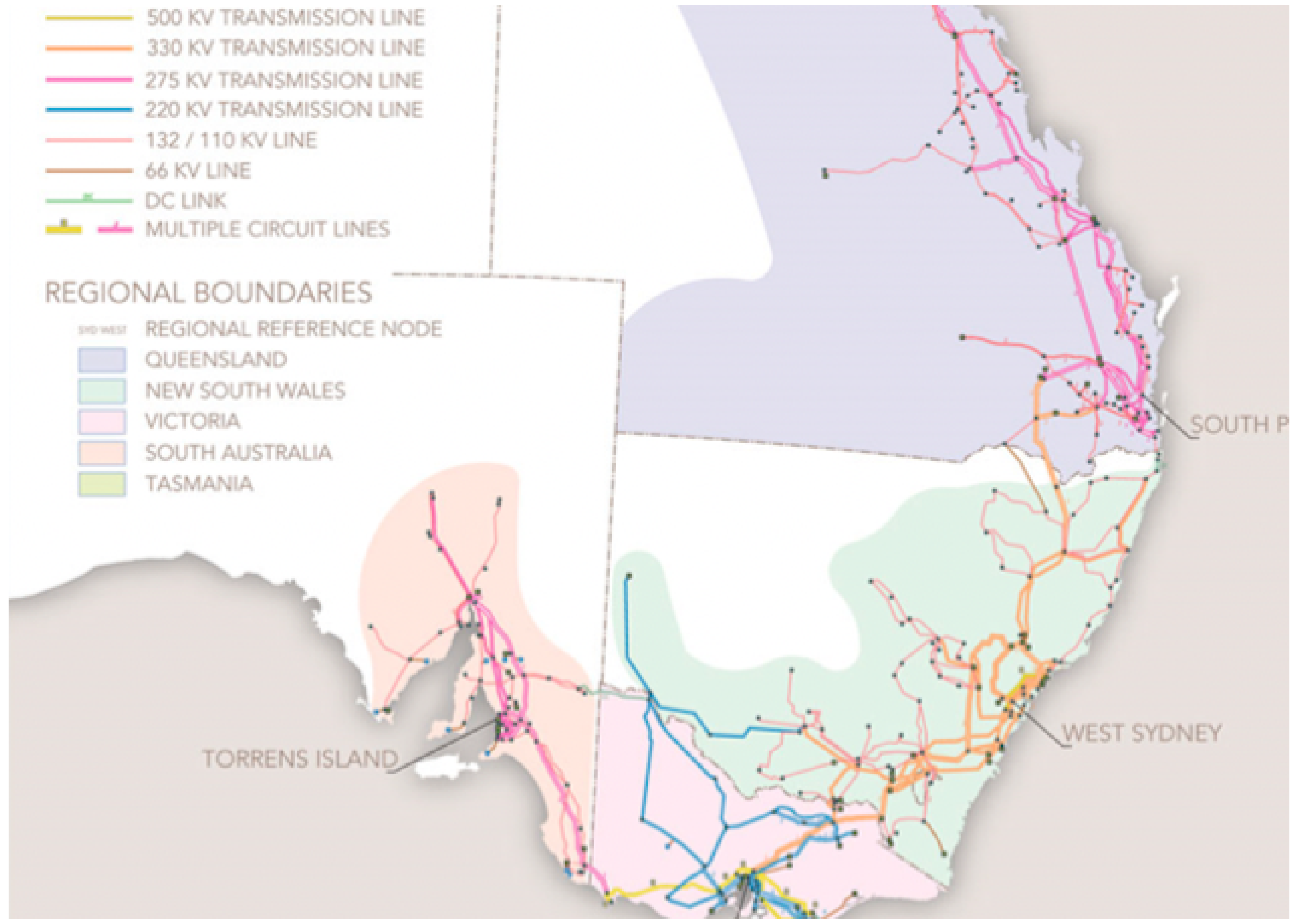
Synthetic datasets are not new to power systems



Texas A&M University Grid Datasets



PyPSA-Eur: European synthetic data

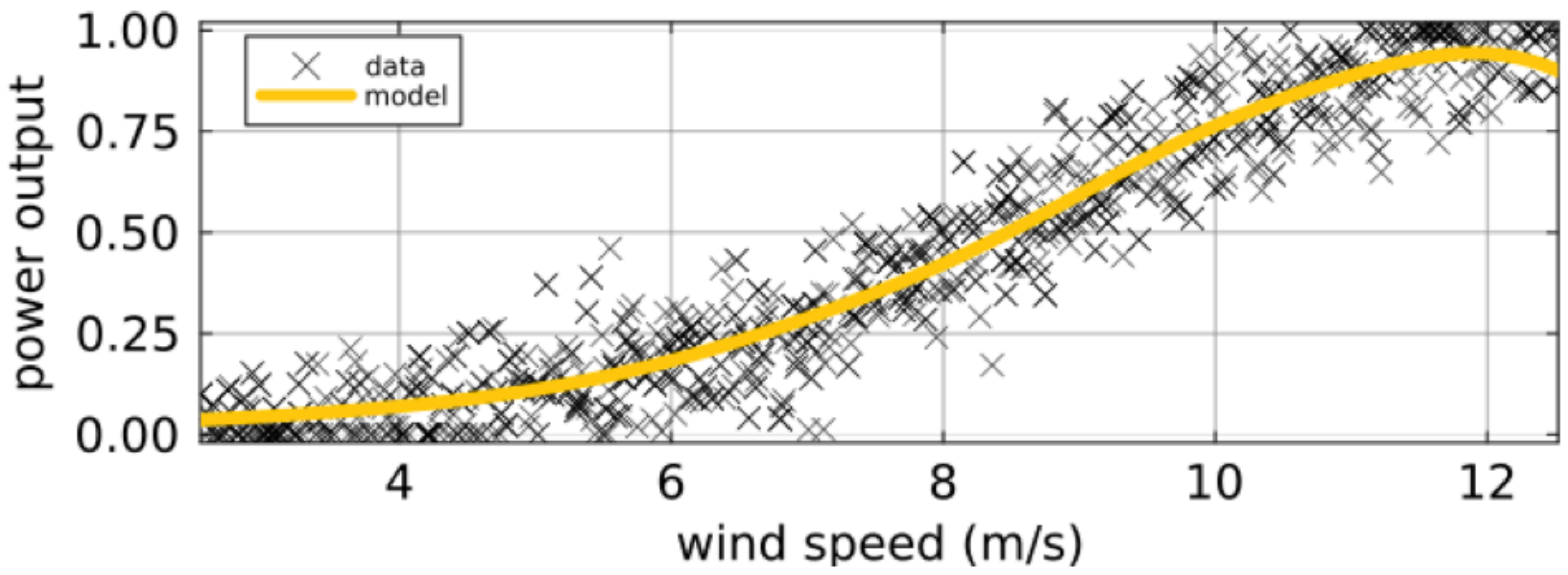


Australian synthetic market data

Why these datasets may not satisfy our needs?

- "[...] data bears **no relation** to the actual grid [...] except that generation and load profiles are similar, based on public data"
- "This test case represents a synthetic (**fictitious**) transmission"
- "This case is synthetic and **does not** model the actual grid"

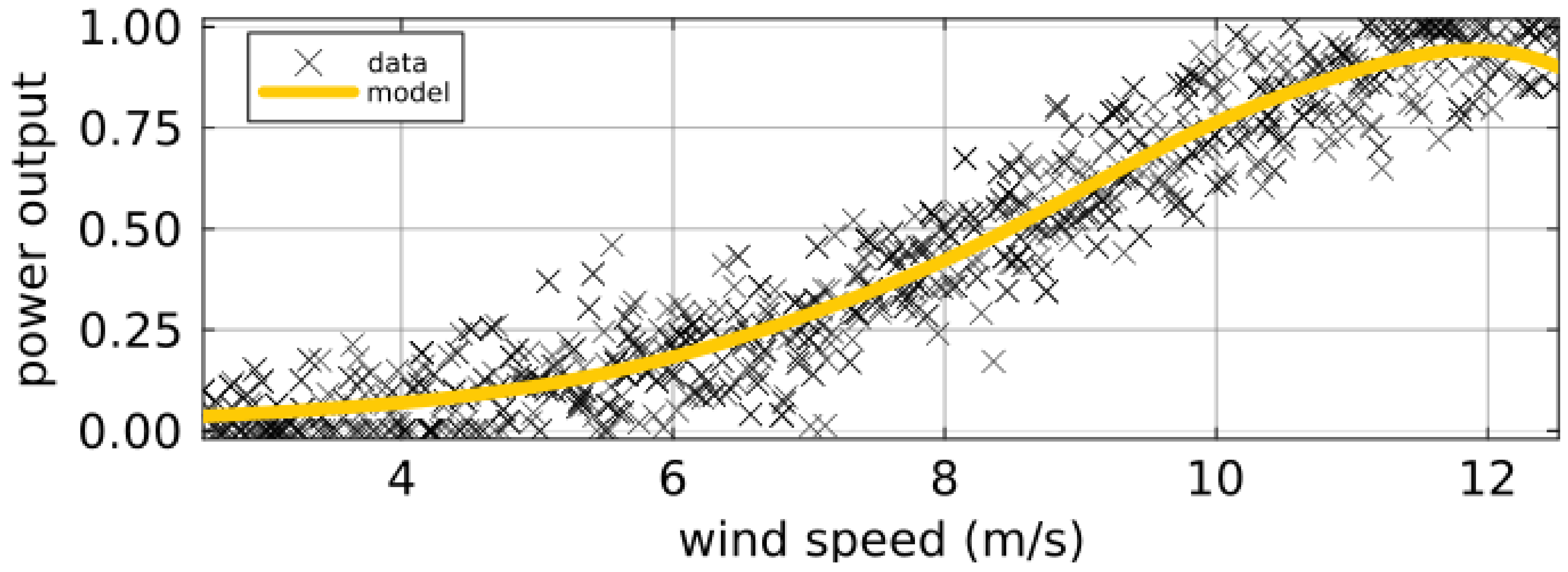
Wind power obfuscation (WPO) algorithm



$$\underset{\beta}{\text{minimum}} \;\; \|X\beta - y\| + \lambda\|\beta\|$$

real dataset: $\mathcal{D} = \{(y_1, x_1), \ldots, (y_n, x_n)\}$

synthetic dataset: $\tilde{\mathcal{D}} = \{(\tilde{y}_1, x_1), \ldots, (\tilde{y}_n, x_n)\}$



- Regression on synthetic data $\tilde{y}$ must match the regression on real data y
- We use regression loss and weights as a measure of accuracy
- Private estimation of regression parameters:

$$\textbf{loss}: \quad \overline{\ell} = \ell(y) + \text{Lap}\left(\frac{\delta_\ell}{\varepsilon}\right), \quad \textbf{weights}: \quad \overline{\beta} = \beta(y) + \text{Lap}\left(\frac{\delta_\beta}{\varepsilon}\right)$$

where $\delta_{(\cdot)}$ is the sensitivity of $(\cdot)$ to data $\boldsymbol{\alpha}-$adjacent datasets

- Lemma (**global sensitivity bounds**):

$$\delta_\ell \leqslant \underset{i=1,\ldots,n}{\text{maximum}} \left\|(X(X^\top X + \lambda I)^{-1}X^\top - I)(e_i \circ \boldsymbol{\alpha})\right\| \qquad \delta_\beta \leqslant \left\|(X^\top X + \lambda I)^{-1}X^\top\right\|_1 \boldsymbol{\alpha}$$

Step 1 Synthetic wind power measurements:

$$\tilde{y}^0 = y + \text{Lap}\left(\alpha/\varepsilon_1\right)$$

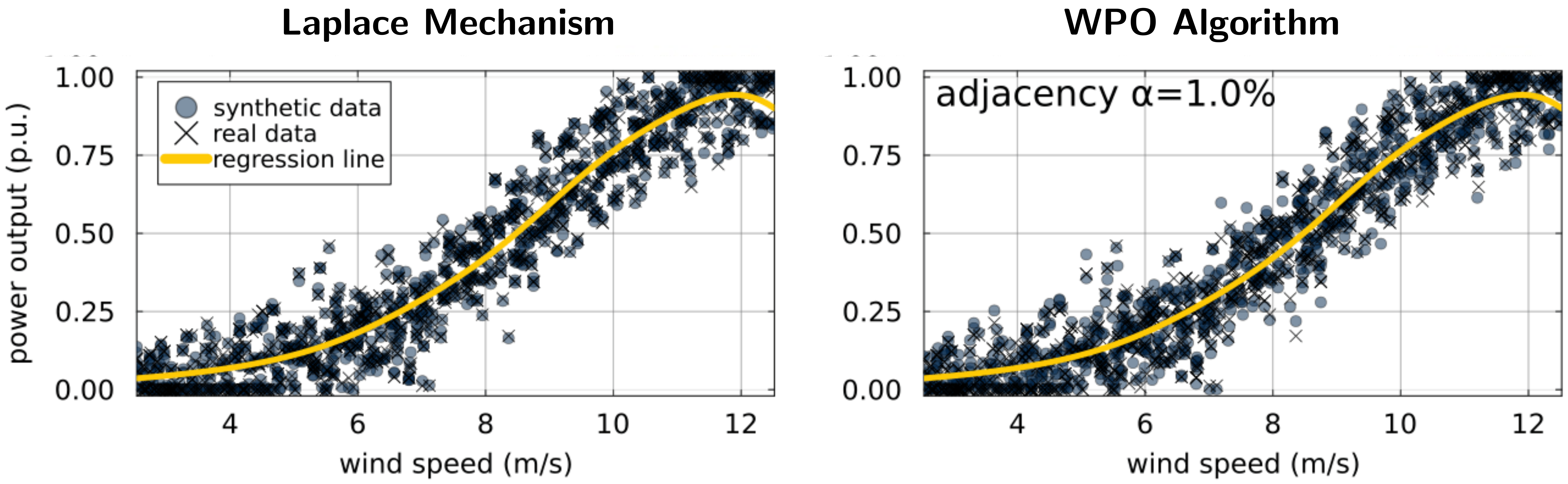
Step 2 Private regression parameters estimation:

$$\overline{\ell} = \ell(y) + \text{Lap}\left(\delta_\ell/\varepsilon_2\right) \quad \overline{\beta} = \beta(y) + \text{Lap}\left(\delta_\beta/\varepsilon_2\right)$$

Step 3 Synthetic dataset post-processing:

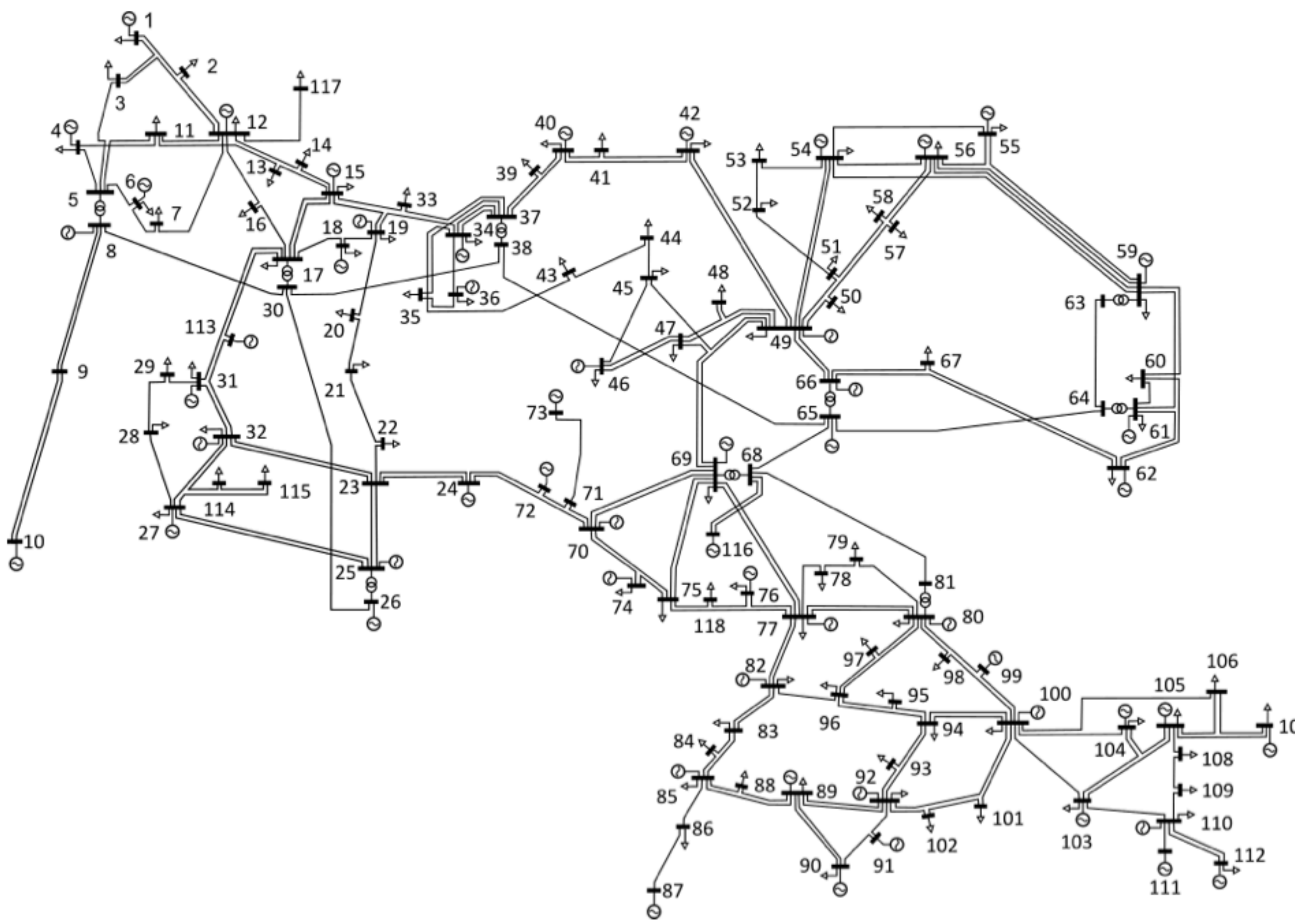
$$\begin{aligned}\tilde{y} \in \underset{\tilde{y}}{\text{argmin}} \quad & \underbrace{\left\|\overline{\ell} - \ell(\tilde{y})\right\|}_{\text{loss accuracy}} + \gamma_\beta \underbrace{\left\|\overline{\beta} - \beta(\tilde{y})\right\|}_{\text{weight accuracy}} + \gamma_y \underbrace{\left\|\tilde{y}^0 - \tilde{y}\right\|}_{\text{regularization}} \\ \text{s.t.} \quad & \mathbf{0} \leqslant \tilde{y} \leqslant \mathbf{1} \\ & \beta(\tilde{y}), \ell(\tilde{y}) \in \underset{\beta}{\text{argmin}} \ \underbrace{\left\|X\beta - \tilde{y}\right\|}_{\ell} + \lambda \left\|\beta\right\|\end{aligned}$$

Theorem: $\varepsilon_1 = \varepsilon/2$ and $\varepsilon_2 = \varepsilon/4$ renders WPO $\varepsilon-$DP for $\alpha-$adjacent wind power datasets.



Accuracy of the WPO Algorithm remains high with a growing privacy requirement α

Transmission capacity obfuscation (TCO) algorithm



Differentially private release of network parameters

Optimal Power Flow (OPF) problem

$$\begin{aligned} \mathcal{C}(\overline{f}) = \min_{p \in \mathcal{P}} \quad & c^\top p && \textit{dispatch costs} \\ \text{s.t.} \quad & \mathbf{1}^\top (p - d) = 0 && \textit{power balance} \\ & |F(p-d)| \leqslant \overline{\boldsymbol{f}} && \textit{power flow limit} \end{aligned}$$

How to release vector of transmission capacities $\overline{\boldsymbol{f}}$ privately?

Laplace mechanism:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon)$$

Almost never feasible

Laplace + Bilevel optimization:

$$\begin{aligned} \min_{\hat{\varphi}} \quad & \left\| \overline{\varphi}^0 - \hat{\varphi} \right\| \\ \text{s.t.} \quad & |\mathcal{C}(\hat{\varphi}) - \mathcal{C}^\star| \leqslant \beta \mathcal{C}^\star \end{aligned}$$

Embedded OPF

Feasible and cost-consistent with respect to a **single** OPF model

Laplace & Exponential mechanisms + Bilevel optimization:

- LM for obfuscation
- EM for worst-case OPF models
- Bilevel opt. on worst-case models

Feasible and cost-consistent with respect to a **population** of OPF models repeat $\boldsymbol{T}$ times

Step 1 Initialize synthetic data using LM:

$$\overline{\varphi}^0 = \overline{f} + \text{Lap}(\alpha/\varepsilon_1)$$

Step 2 Find the worst-case OPF model using EM:

$$\Delta\mathcal{C}_i = \left\|\mathcal{C}_i(\overline{f}) - \mathcal{C}_i^R(\overline{\varphi}^{\boldsymbol{t-1}})\right\|_1 + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right), \forall i = 1, \dots, m$$

return index $k^{\boldsymbol{t}}$ of the worst-case model

Step 3 Compute the worst-case cost using LM:

$$\overline{\mathcal{C}}_{\boldsymbol{t}} = \mathcal{C}_{k^{\boldsymbol{t}}}(\overline{f}) + \text{Lap}\left(\overline{c}\alpha/\varepsilon_2\right)$$

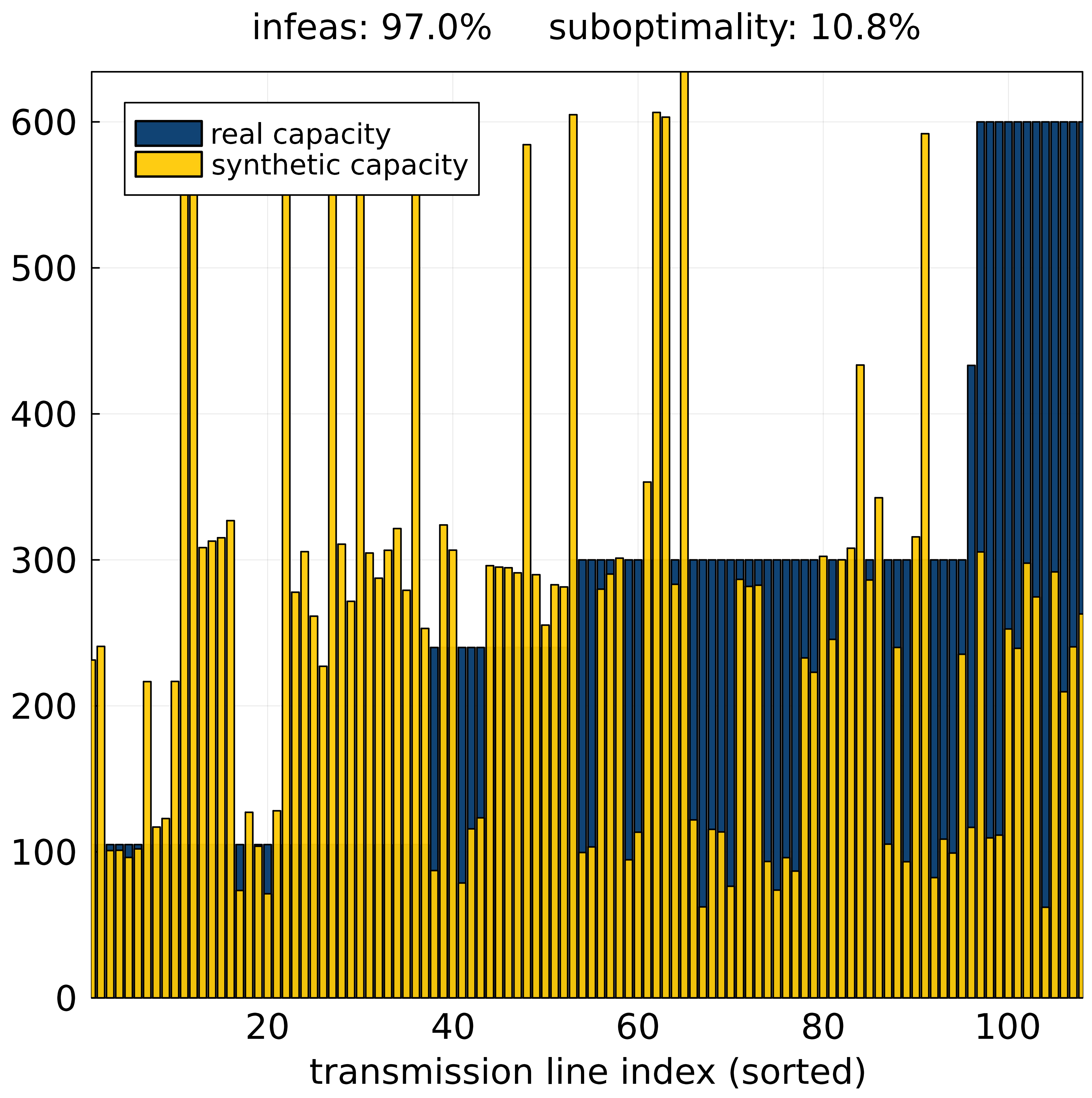
Step 4 Post-processing bilevel optimization:

$$\overline{\varphi}^t \in \underset{\overline{\varphi}}{\text{argmin}} \sum_{\boldsymbol{\tau}=1}^{\boldsymbol{t}} \left\|\overline{\mathcal{C}}_{\boldsymbol{\tau}} - \mathcal{C}_{k^{\boldsymbol{\tau}}}(\overline{\varphi})\right\| + \left\|\overline{\varphi} - \overline{\varphi}^{\boldsymbol{t-1}}\right\|$$

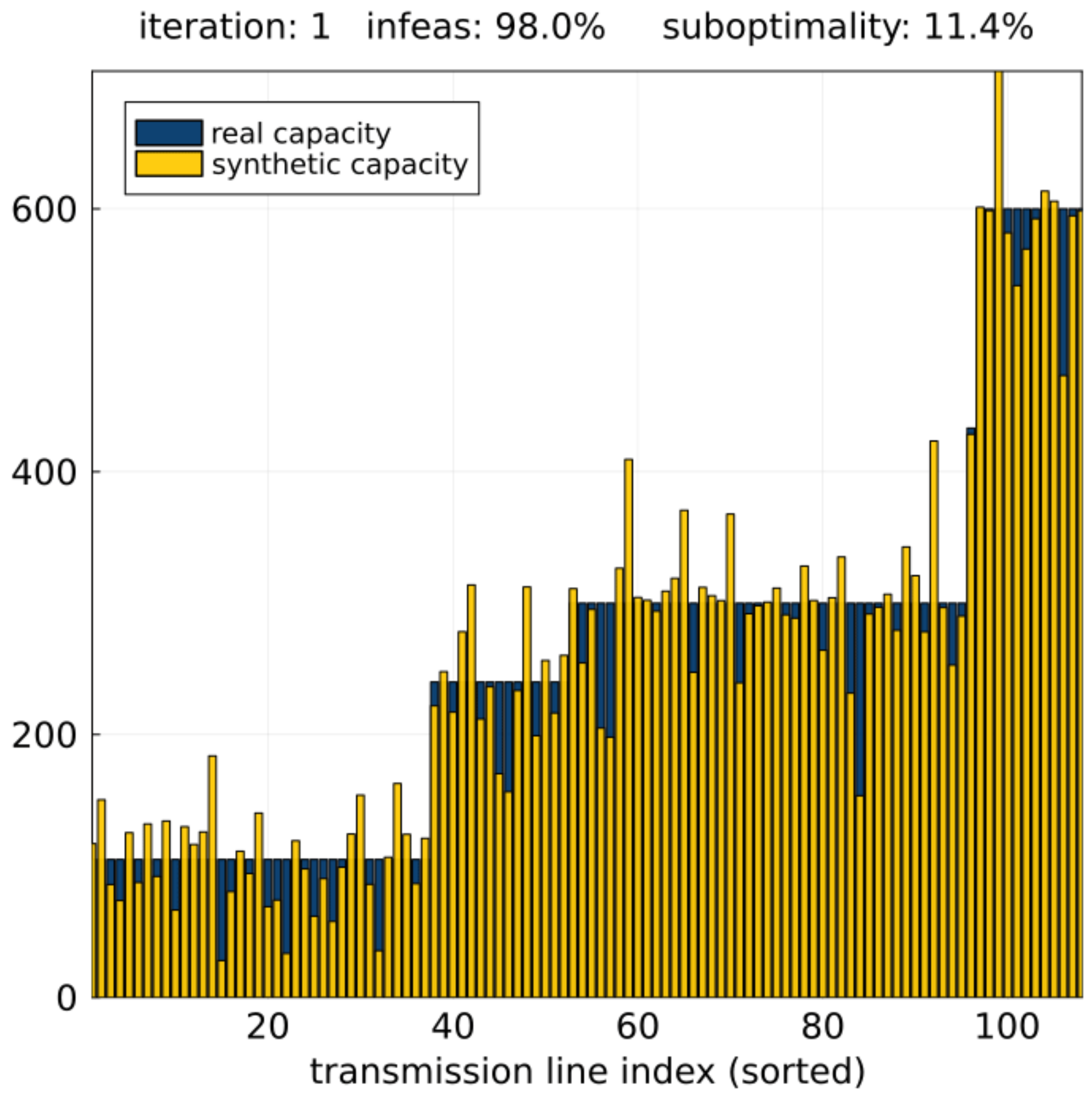
Theorem: $\boldsymbol{\varepsilon_1 = \varepsilon/2}$ and $\boldsymbol{\varepsilon_2 = \varepsilon/(4T)}$ achieve $\varepsilon-$differential privacy

IEEE 73-RTS benchmark: Laplace versus TCO Algorithm

Laplace mechanism

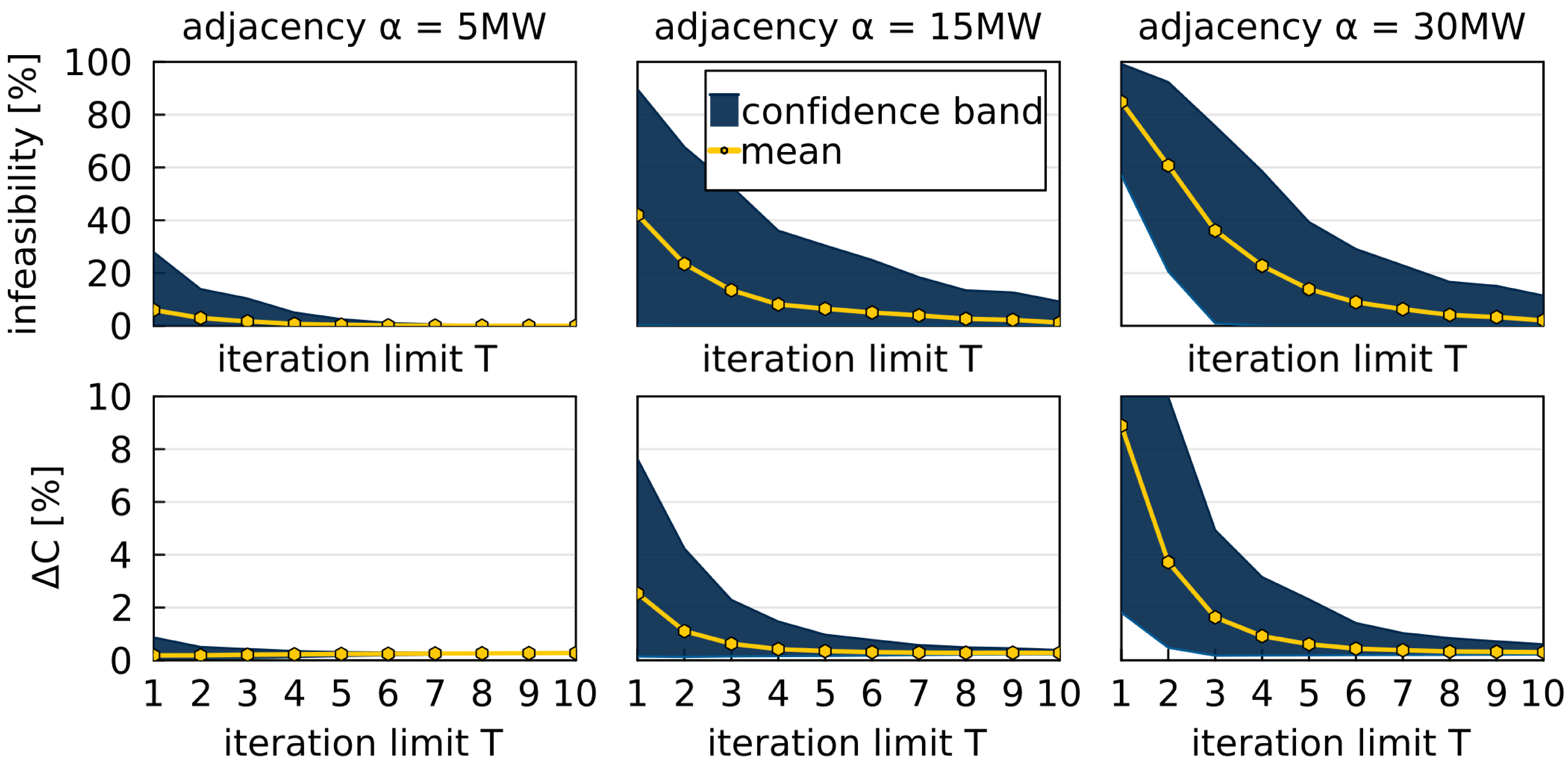


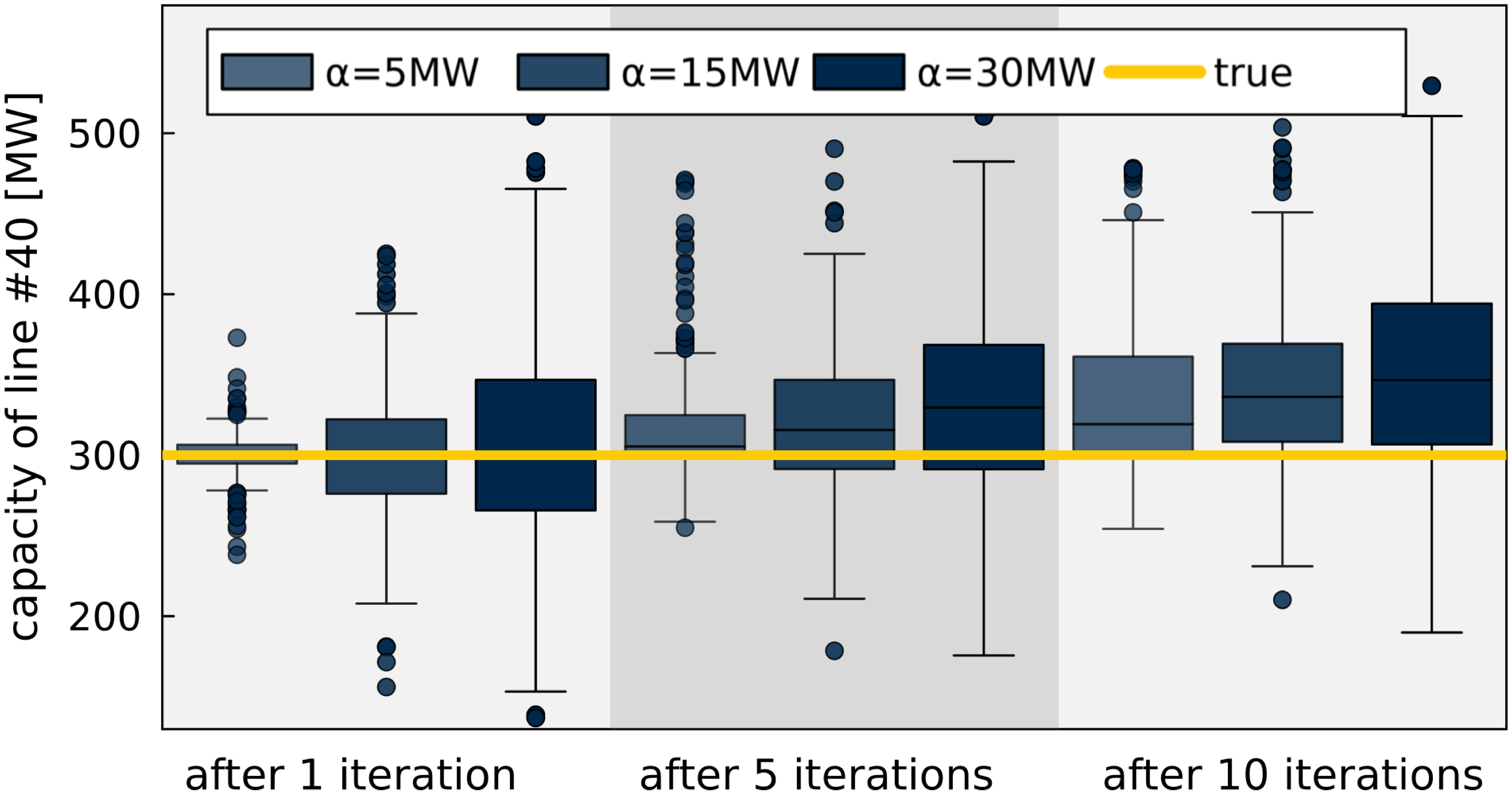
TCO Algorithm



IEEE 73-RTS benchmark: TCO feasibility and sup-optimality

more privacy (more noise)




α=5MW
α=15MW
α=30MW
true
capacity of line #40 [MW]
500
400
300
200
after 1 iteration
after 5 iterations
after 10 iterations

Concluding remarks

What we used to say about synthetic datasets:

- "[...] data bears **no relation** to the actual grid [...]"
- "This test case represents [...] **fictitious** transmission"
- "This case is synthetic and **does not** model the actual grid"

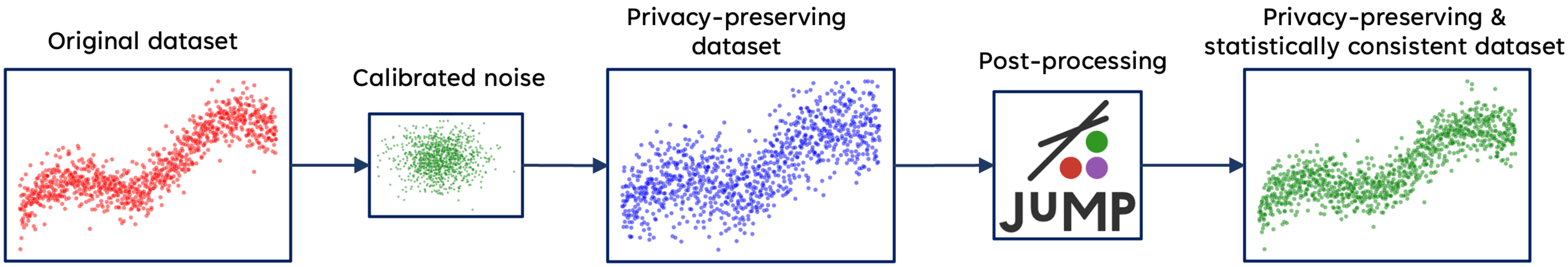
What we will say about synthetic datasets:

- "This synthetic dataset is produced based on the data from a real-world power grid"
- "It is not possible to infer the real data from this synthetic dataset"
- "Computational results on this data are consistent with the real data"

What does it mean for electricity market/system operators?

- New algorithms for **controllable** market transparency:
  - infrastructure data (grid topology, network parameters, generation, loads, etc.)
  - market participation data (bidding quantities, prices, etc.)
- No need for aggregation:
  - system cost/load $\Longrightarrow$ nodal cost/load
  - aggregated generation $\Longrightarrow$ highly granular generation records
- Rigorous privacy quantification $\Longrightarrow$ legal compliance

What is next?



What has been done so far

- Noise addition to obfuscate private data
- Post-processing optimization to improve utility

Diffusion models are the next step

- Privacy-preserving perturbation in the forward process
- Optimization in the reverse process to ensure utility

Thank you for your attention!

References:

1 Dvorkin, V., Botterud A.
Differentially private algorithms for synthetic power system datasets
IEEE Control Systems Letters, 2023

2 Dvorkin, V., Fioretto, F., Van Hentenryck, P., Kazempour, J. and Pinson, P.
Privacy-preserving convex optimization: When differential privacy meets stochastic programming
Priprint, arXiv preprint arXiv:2006.12338, 2022

3 Dvorkin, V., Fioretto, F., Van Hentenryck, P., Pinson, P. and Kazempour J.
Differentially private optimal power flow for distribution grids
IEEE Transactions on Power Systems, 2021
Best 2019–2021 Paper Award

4 Dvorkin, V., Van Hentenryck, P., Kazempour, J. and Pinson P.
Differentially private distributed optimal power flow
2020 Conference on Decision and Control

Let's stay in touch:

DvorkinVladimir   Vladimir-Dvorkin   dvorkin@umich.edu