Privacy-Preserving Convex Optimization: When Differential Privacy Meets Stochastic Programming

Vladimir Dvorkin[1], Ferdinando Fioretto[2], Pascal Van Hentenryck[3], Pierre Pinson[4], Jalal Kazempour[5]

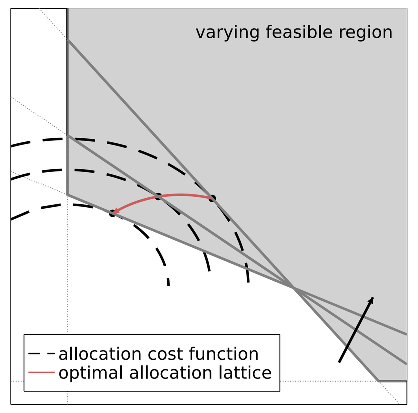
[1]Massachusetts Institute of Technology, [2]Syracuse University, [3]Georgia Institute of Technology, [4]Imperial College London, [5]Technical University of Denmark

Privacy leakages in convex optimization

$$\min_x \; c^\top x \quad \text{s.t.} \quad b - Ax \in \mathcal{K}$$

- 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 $\mathcal{D}'$

Resource allocation

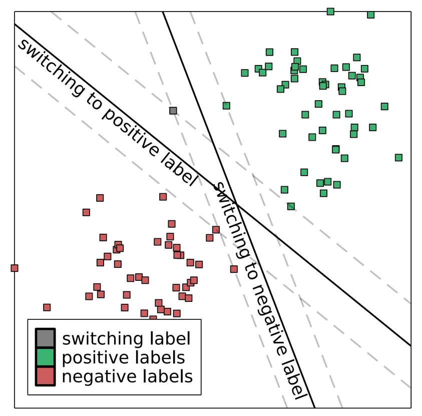


Changes in the feasible region are directly exposed in the optimal allocation and allocation cost

Regression analysis

Changes in training data are exposed in the parameters of regression models (solution uniqueness w.r.t. dataset)

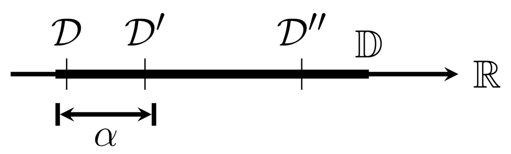
SVM classification



Label of the marginal data point is exposed in the parameters of the support vector machines (SVM) hyperplane

Need for rigorous privacy-preserving methods to formally guarantee data privacy

Formalization of privacy



- Optimization as a mapping $x^\star : \mathbb{D} \mapsto \mathbb{X}$
- Privacy adversary mapping $\mathcal{A} : \mathbb{X} \mapsto \mathbb{D}$
- **Privacy goal** is to make $\alpha-$adjacent dataset indistinguishable (mislead the adversary)

- Let $\tilde{x}^\star$ be a random counterpart of $x^\star$
- For any two datasets $\mathcal{D}, \mathcal{D}' \in \mathbb{D}$:

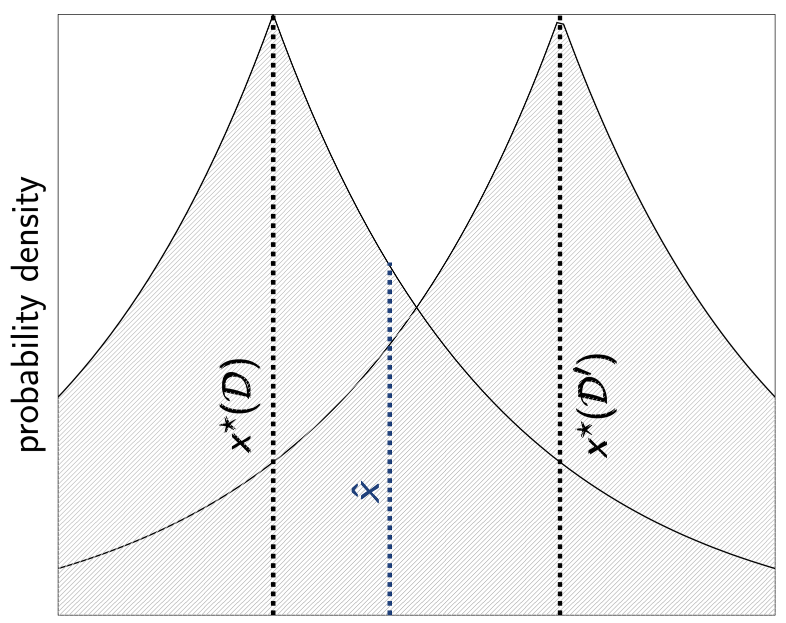
  deterministic mapping: $x^\star(\mathcal{D}) \neq x^\star(\mathcal{D}')$

  randomized mapping: $\tilde{x}^\star(\mathcal{D}) \approx \tilde{x}^\star(\mathcal{D}')$
- **$\varepsilon-$differential privacy** ($\varepsilon-$DP):

$$\frac{\Pr[\tilde{x}^\star(\mathcal{D}) = \hat{x}]}{\Pr[\tilde{x}^\star(\mathcal{D}') = \hat{x}]} \leqslant \exp(\varepsilon)$$

- Smaller ε implies stronger privacy

$$\exp(\varepsilon) \approx 1 + \varepsilon$$



Limits of the standard **input** and **output** DP perturbation strategies

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

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

2. Perturbation of optimization results

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

Both strategies can not guarantee **feasibility** nor **optimality**

Stochastic programming for private optimization queries

- For any deterministic program, we develop a stochastic counterpart to enable DP guarantees
- We model an optimization vector as the **linear decision rule** of the form:

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

$\overline{x}$ – nominal solution vector (function of dataset)

X – solution recourse matrix (function of dataset)

ζ – perturbation calibrated to solution sensitivity Δ_α

- Vector $\overline{x}$ and matrix X are subject to stochastic optimization:

$$\min_{\overline{x}, X \in \mathcal{X}} \; \mathbb{E}\left[c^\top(\overline{x} + X\zeta)\right] \quad \text{s.t.} \quad \Pr\left[b - A(\overline{x} + X\zeta) \in \mathcal{K}\right] \geqslant 1 - \eta$$

- Minimize expected cost (to guarantee optimality)
- Chance constraint (to guarantee feasibility)
- $\mathcal{X}$ for data-independent query perturbation (to guarantee privacy)

- For example, for **identity query**, the recourse is data-independent when X is identity, i.e.,

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

Main result (differential privacy guarantee)

Let Δ_α be the worst-case ℓ_1-sensitivity of optimization results to $\alpha-$adjacent datasets $\mathcal{D}, \mathcal{D}' \in \mathbb{D}$. If for $\zeta \sim \text{Lap}(\Delta_\alpha/\varepsilon)$ the chance-constrained program returns the optimal solution, the optimization result on any adjacent dataset is $\varepsilon-$differentially private. That is, for any dataset pair, we have

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

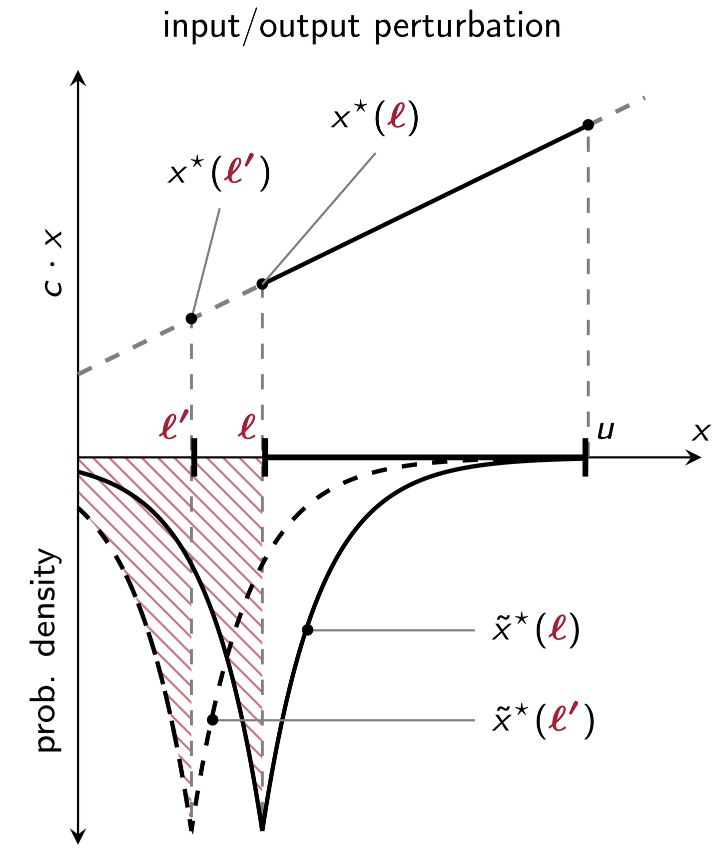
for some arbitrary optimization outcome $\hat{x}$.

Linear programming illustrative example

$$\min_x \; c \cdot x \quad \text{s.t.} \quad \ell \leqslant x \leqslant u$$

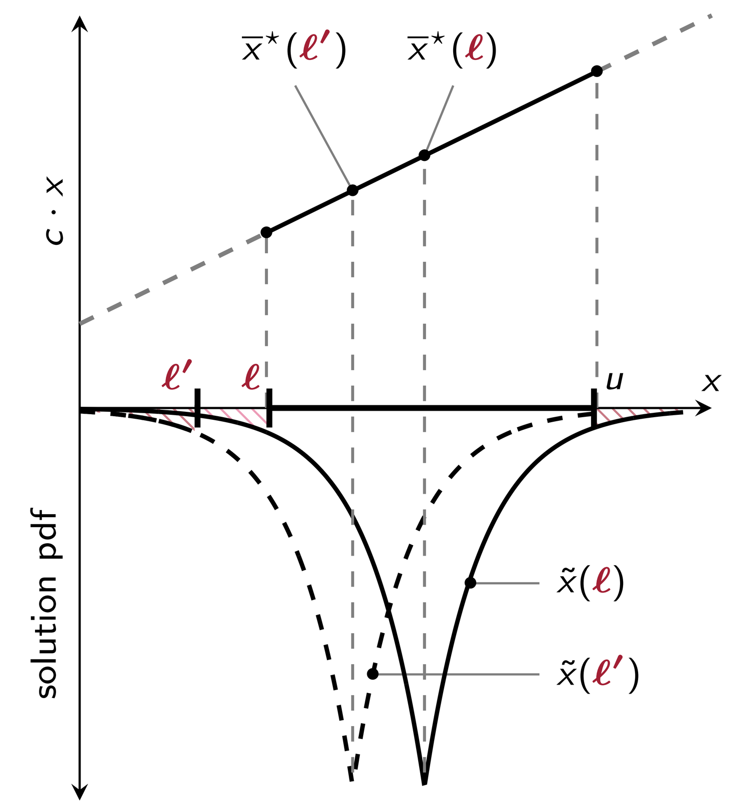
- We make parameters ℓ and ℓ' statistically indistinguishable in optimization result $x^\star$
- Input and output perturbations are equivalent

input/output perturbation



While data (input) or solution (output) perturbations make randomized results statistically similar, there is a 50% chance of an infeasible outcome

program perturbation



Program perturbation finds such a nominal solution $\overline{x}^\star$, whose perturbations is feasible with a high probability (up to chance constraint tolerance)

Private optimal power flow (OPF) problem

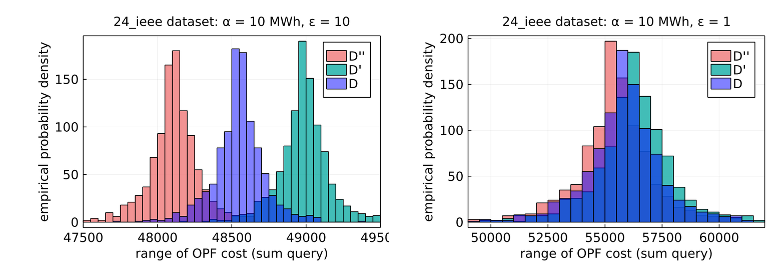
$$\min_{\overline{x}, X \in \mathcal{X}} \; \mathbb{E}[c^\top(\overline{x} + X\zeta)]$$
$$\text{s.t.} \quad 1^\top(\overline{x} + X\zeta - \boldsymbol{d}) = 0$$
$$\Pr\begin{bmatrix} |F(\overline{x} + X\zeta - \boldsymbol{d})| \leqslant f^{\max} \\ x^{\min} \leqslant \overline{x} + X\zeta \leqslant x^{\max} \end{bmatrix} \geqslant 1 - \eta$$

- Given load vector $\boldsymbol{d}$, the OPF problem computes the cost-optimal generator dispatch
- With chance-constrained OPF, we privately release dispatch costs (objective) while ensuring feasibility

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

Unlike input or output perturbation, the program perturbation is feasible with a high probability.



With ↑ privacy requirements (↓ ε), the distance between distributions on adjacent datasets reduces.

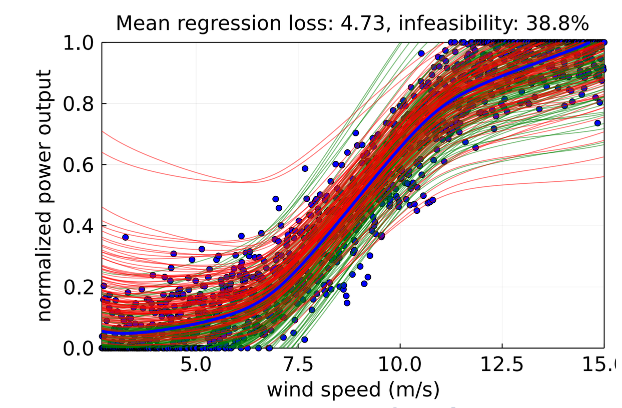
Private wind power curve fitting

$$\min_\beta \; \mathbb{E}\left[\sum_{i=1}^n \Big(\underbrace{y_i - \varphi(x_i)^\top \beta}_{\text{business as usual}} \underbrace{- \varphi(x_i)^\top \zeta}_{\text{perturbation}}\Big)^2\right]$$
$$\text{s.t.} \quad \mathbb{P}\big[\underbrace{C(\beta + \zeta) \geqslant 0}_{\text{monotonic con.}}\big] \geqslant 1 - \eta$$

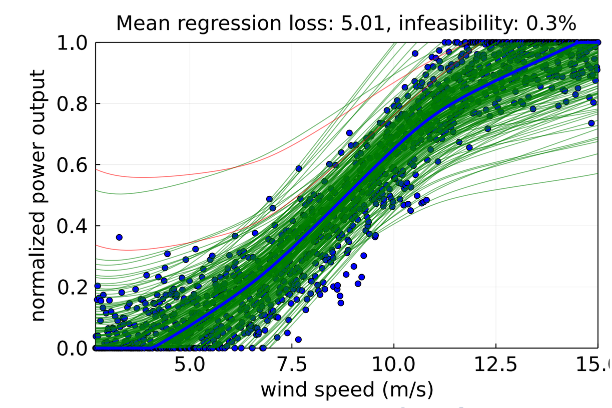
- Dataset $\{(y_1, x_1), \dots, (y_n, x_n)\}$
- Minimize regression loss function
- By finding optimal weights $\beta^\star$ ...
- ... of basis functions in vector $\varphi(x)$

- We want to make training datasets indistinguishable in model weights $\beta^\star$ ...
- ... while preserving monotonic properties of the curve under perturbation

Alstom.Eco.80



output perturbation

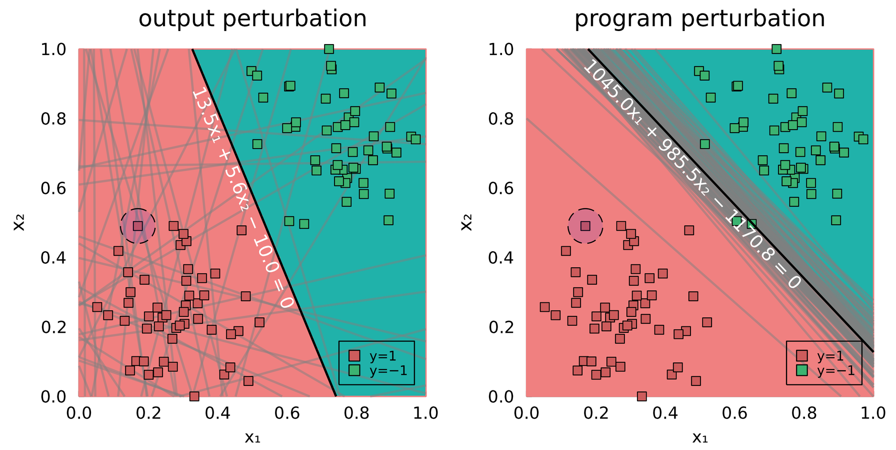


program perturbation

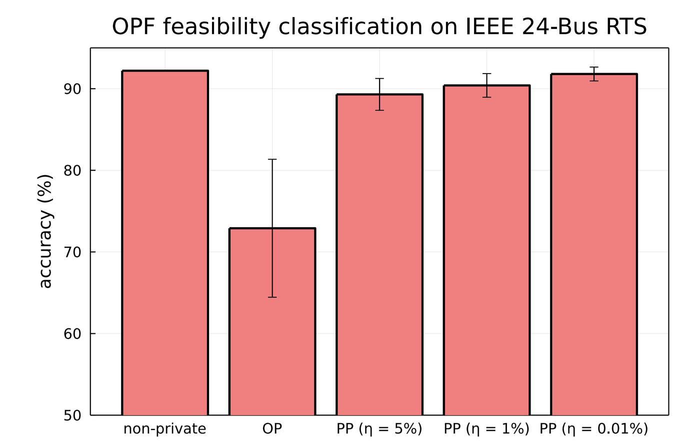
Private OPF feasibility classification (SVM)

$$\min_{\overline{b}, \overline{w}, z} \; \lambda\|\overline{w}\|^2 + \tfrac{1}{m}1^\top z$$
$$\text{s.t.} \quad y_i(\overline{w}^\top x_i - \overline{b}) \geqslant 1 - z_i$$
$$z_i \geqslant 0, \quad \forall i = 1, \dots, m$$

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



While the deterministic hyperplane is sensitive to perturbations (left), the stochastic hyperplane is very robust (right)



After tuning violation tolerance (η), the private classifier is almost as good as non-private one