Private optimization without constraint violations

Andrés Muñoz Medina, Umar Syed, Sergei Vassilvitskii, and **Ellen Vitercik**

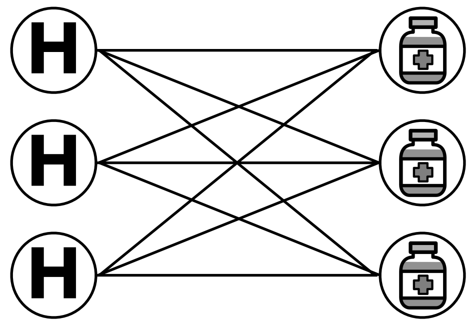
Private, linearly-constrained optimization

Goal: Privately find $\vec{x}$ maximizing $g(\vec{x})$ such that $A\vec{x} \leq \vec{b}(D)$

Private database

Example:

- Hospital branches need drug to treat patients with specific disease
- **Goal:** Determine which pharmacies supply which branches
  - Minimize total transportation cost
- Linear program, but…
  Constraints reveal number of sick patients at each branch
- **Constraints are critical:** ensure hospitals receive enough drugs



How to privately find nearly-optimal point satisfying constraints?

Differential privacy

Each dataset D consists of individuals' records

D and D' **neighboring** ($D \sim D'$) if differ on single record

Sensitivity: $\Delta = \max_{D \sim D'} \left\| \vec{b}(D) - \vec{b}(D') \right\|_1$

$\vec{x}(D)$ is output given:

$g : \mathbb{R}^n \to \mathbb{R}$ (L-Lipschitz), $A \in \mathbb{R}^{m \times n}$, $\vec{b}(D) \in \mathbb{R}^m$

Algorithm is (ϵ, δ)-**differentially private (DP)** if:

For all $D \sim D'$ and all $V \subseteq \mathbb{R}^n$, $\mathbb{P}\left[\vec{x}(D) \in V\right] \leq e^{\epsilon} \cdot \mathbb{P}\left[\vec{x}(D') \in V\right] + \delta$

Theorem: There is no non-trivial $(\epsilon, 0)$-DP algorithm for this problem

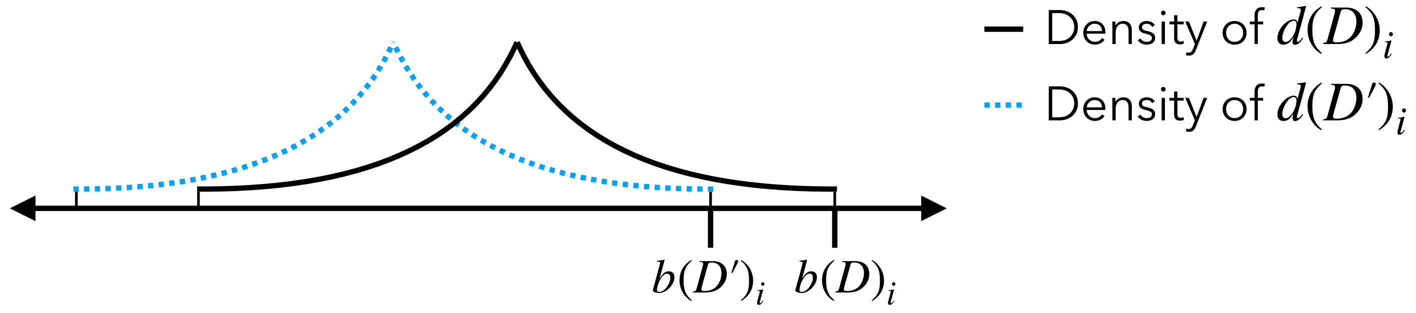
Algorithm

Challenge: Can't satisfy constraints if feasible regions vary too much

Assumption: $\bigcap_D \left\{\vec{x} : A\vec{x} \leq \vec{b}(D)\right\} \neq \emptyset$ (e.g., contains origin)

Algorithm:

1. Map $\vec{b}(D)$ to another vector $\vec{d}(D)$ such that $\vec{d}(D) \leq \vec{b}(D)$ w.p. 1 using the Truncated Laplace Mechanism



2. Return $\vec{x}(D) = \text{argmax}\left\{g(\vec{x}) : A\vec{x} \leq \vec{d}(D)\right\}$

Fact: $A\vec{x}(D) \leq \vec{b}(D)$ with probability 1

Theorem [privacy]: Preserves (ϵ, δ)-DP

Theorem [quality]: Let $\vec{x}^*$ be an optimal solution to original problem

$$g\left(\vec{x}(D)\right) \geq g\left(\vec{x}^*\right) - \frac{2L\Delta}{\epsilon} \cdot \alpha(A) \cdot \ln\left(\frac{m(e^{\epsilon} - 1)}{\delta} + 1\right)$$

Linear system's *condition number* [Li '93]

Specifically,

$$\alpha(A) = \inf_{p \geq 1} \sqrt[p]{m} \cdot \sup_{\vec{u} \geq 0} \left\{ \|\vec{u}\|_{p^*} : \begin{array}{c} \left\| A^{\top}\vec{u} \right\|_{q^*} = 1 \text{ and the rows of } A \\ \text{corresponding to the nonzero entries} \\ \text{of } \vec{u} \text{ are linearly independent} \end{array} \right\}$$

where $\|\cdot\|_q$ is the norm under which g is L-Lipschitz

Ex.: If A is invertible and g is L-Lipschitz under the ℓ_2-norm, $\alpha(A) \leq \frac{\sqrt{m}}{\sigma_{\min}(A)}$

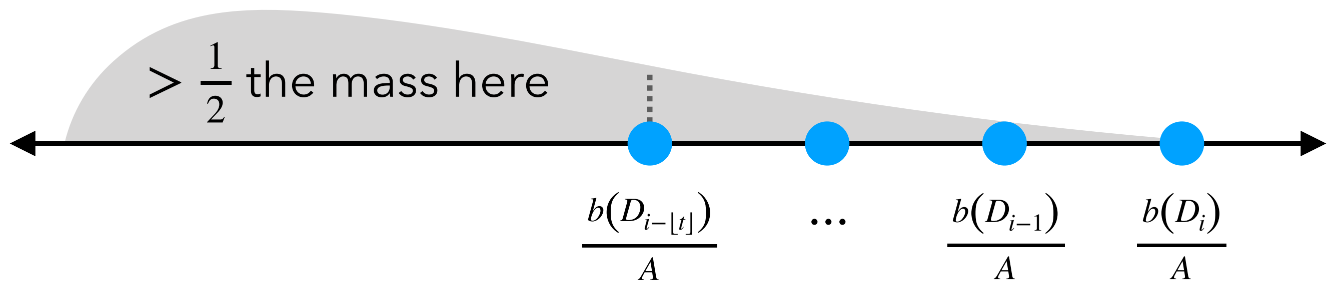
Matching lower bound (up to log factors)

Theorem: There exists an infinite family of matrices $A \in \mathbb{R}^{n \times n}$, a 1-Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$, and a mapping from databases D to vectors $\vec{b}(D)$ for any $\Delta > 0$ s.t.:

1. The sensitivity of $\vec{b}(D)$ equals Δ, and
2. For any (ϵ, δ)-DP mech. returning $\vec{\mu}(D)$ s.t. $A\vec{\mu}(D) \leq \vec{b}(D)$ w.p. 1,
$$\mathbb{E}\left[g\left(\vec{\mu}(D)\right)\right] \leq g\left(\vec{x}^*\right) - \frac{\Delta}{4\epsilon} \cdot \alpha(A) \cdot \ln\left(\frac{e^{\epsilon} - 1}{2\delta} + 1\right)$$

Proof idea when $n = 1$:

- Let $t = \frac{1}{\epsilon} \ln\left(\frac{e^{\epsilon} - 1}{2\delta} + 1\right)$ and $g(x) = x$
- $\forall i \in \mathbb{Z}$, let D_i be a database where $D_i \sim D_{i+1}$ and $b\left(D_i\right) = \Delta i$



Experiments

- Individuals pool money to invest
  - Total amount $b(D)$ private except to investment manager
- Stock returns have mean $\vec{p}$ and covariance Σ
- **Goal:** Minimize variance subject to minimum expected return r
  $\min\left\{\vec{x}^{\top}\Sigma\vec{x} : \vec{p} \cdot \vec{x} \geq r, \sum_{i=1}^{n} x_i \leq b(D)\right\}$
- Data from Dow Jones Industrial Average stocks