Isotonic Regression

- Imposing ordering constraints on estimation problems
- Signal processing, machine learning and statistics
  - Probability calibration for classification
  - Interpretability

Constrained optimization problem

\[
\min_{x \in \mathbb{R}^n} H(x) \text{ such that } \forall (i, j) \in E, \; x_i \geq x_j,
\]
where \( E \subset \{1, \ldots, n\}^2 \) is a directed acyclic graph.

- Most classical example: separable quadratic cost

\[
\min_{x \in \mathbb{R}^n} \frac{1}{2} \|x - z\|^2 \quad \text{such that } x_1 \leq \cdots \leq x_n
\]

- Convex function \( H \) \( \Rightarrow \) convex optimization problem
  - Efficient first-order algorithms
- Non-convex function \( H \)
  - No efficient algorithms general

Submodular Functions

- Definition ([1]): \( \forall x \in \mathbb{R}^n, \forall i \neq j, \frac{\partial^2 H}{\partial x_i \partial x_j}(x) \leq 0 \)
- Example 1 (Laplacian): \( H(x) = \sum_{i,j=1}^n w_{ij}(x_i - x_j)^2 \)
- Example 2 (Separable functions): \( H(x) = \sum_{i=1}^n h_i(x_i) \)
  - For any (e.g., non-convex) functions \( h_i : \mathbb{R} \rightarrow \mathbb{R} \)

Convex Extension with Measures

- \( \mathcal{P}(\mathbb{R}) \) = set of Radon probability measures on \( \mathbb{R} \)
- \( (\text{reverse}) \) cumulative distribution function for \( \mu \in \mathcal{P}(\mathbb{R}) \): \( F_{\mu_i}(x) = \mu_i([x_i, +\infty)) \)
- “Inverse” CDF \( F_{\mu_i}^{-1} \)

- Extension to \( \mu = (\mu_1, \ldots, \mu_n) \)-see [2]:

\[
h(\mu_1, \ldots, \mu_n) = \int_0^1 H[F_{\mu_1}^{-1}(t), \ldots, F_{\mu_n}^{-1}(t)] dt
\]

- Property 1: \( H \) submodular \( \iff \) \( h \) convex
- Property 2: minimizers of \( H \) can be recovered from minimizers of \( h \)
- Consequence: Submodular function minimization is a convex optimization problem

Discretization Algorithms

- Discretize each interval with \( k \) values
  - Approximation \( O(n/k) \)
  - Number of queries of \( H \) is \( O(tnk) \) for \( t \) operations of projected subgradient descent
  - Projection based on (efficient) quadratic cost isotonic regression \([3]\)
  - Extra approximation factor \( O(n/\sqrt{t}) \)
- With \( k = n/\varepsilon \) and \( t = n^2/\varepsilon^2 \)
  - Error \( \varepsilon \) and number of queries \( O(n^4/\varepsilon^5) \)
- For improved behavior, see paper

Experiments

- Robust isotonic regression: \( H(x) = \frac{1}{n} \sum_{i=1}^n G(x_i - z_i) \)
  - \( G(t) = |t| \) or \( |t|^2 \Rightarrow \) convex problem
  - \( G(t) = \kappa^2 \log(1 + |t/\kappa|) \Rightarrow \) submodular problem
  - Vanishing gradients \( \Rightarrow \) Robustness
- Data with adversarial perturbations
  - Goal: learn decreasing functions
  - Perturbation by noise and increasing function (observations in pink)

Extension for Isotonic Constraints

- First-order stochastic dominance between \( \mu, \nu \in \mathcal{P}(\mathbb{R}) \):

\[
\mu \succ \nu \text{ if and only if } \forall x \in \mathbb{R}, \quad F_\mu(x) \geq F_\nu(x)
\]

- Equivalent optimization problem

\[
\min_{\mu \in \mathcal{P}(\mathbb{R})^n} h(\mu) \text{ such that } \forall (i, j) \in E, \; \mu_i \geq \mu_j
\]

References