Efficient Algorithms

Lecture notes up-to-date through Lecture 11

P set 2 Errata: "K-approximate oracle" → "K-approximate relaxation"

Recap last time.

- Showed how to bound truncated moments via Ledoux-Talagrand
- Prove Ledoux-Talagrand via stochastic dominance
- Khinchine + Rosenthal to bound k^{th} moments of ℓ_2-norm

This time.

- Present efficient algo for robust estimation given bounded covariance
- Idea: Write non-convex problem that approximates MD functional
  - ↳ Show "almost" convex

Efficient algo

$\mathcal{G}_{cov}(\sigma)$ = dist's w/ bounded covariance

$$\|\Sigma_p\| \leq \sigma^2$$

MD functional on $\mathcal{G}_{cov}(\sigma)$ gives robust mean estimation w/ error $\mathcal{O}(\sigma\sqrt{\varepsilon})$ for TV corruptions of size ε.

not efficient

→ Find q s.t. $TV(\tilde{p}, q) \approx \varepsilon$

$\|Cov_q\| \approx \sigma$

Representing $\tilde{p}$ and p^*.

$\tilde{p}$ empirical dist^n over n samples $x_1, \ldots, x_n \in \mathbb{R}^d$

Not same as $TV(\tilde{p}, p^*) \le \varepsilon$

p^* empirical dist^n over subset of S of $\{x_1, \dots, x_n\}$, w/ $|S| = (1-\varepsilon)\cdot n$

Suppose $\tilde{p}, p^*$ w/ $TV(p^*, \tilde{p}) \le \varepsilon$

$\Rightarrow \exists\ p'$ s.t. $p' \le \frac{\tilde{p}}{1-\varepsilon}$, $p' \le \frac{p^*}{1-\varepsilon}$

$\| Cov_{p'} \| \le \frac{1}{1-\varepsilon} \| Cov_{p^*} \| \le 2\sigma^2$

$(\varepsilon \le \frac{1}{2})$

Problem statement.

$X_1, \dots, X_n \in \mathbb{R}^d$

$\exists\ (1-\varepsilon)\cdot n$ points $\Rightarrow$ call dist^n p^*, s.t. $\| Cov_{p^*} \| \le \sigma^2$

$\boxed{\text{Goal}}$ Output $\hat{\mu}$ s.t. $\| \mu_{p^*} - \hat{\mu} \|_2 = O(\sigma\sqrt{\varepsilon})$.

$\Rightarrow$ Generally by finding q s.t. $\| \Sigma_q \| \le \sigma^2$ where q dist^n over $(1-\varepsilon)\cdot n$ of points.

Optimization q: $q_1, \dots, q_n$ on $X_1, \dots, X_n$

$$(*) \quad \min_{q_1, \dots, q_n} \sup_{\|v\|_2 \le 1} \sum_{i=1}^n q_i \langle X_i - \mu_q, v \rangle^2$$

s.t. $\mu_q = \sum_{i=1}^n q_i X_i$, $q_i \ge 0$, $\sum_i q_i = 1$

$q_i \le \frac{1}{(1-\varepsilon)n}$

q is a dist^n over $(1-\varepsilon)\cdot n$ distinct points

$\|\Sigma'_q\|$ should be small

q ε-deletion at $\tilde{p}$

Problem (*) is non-convex in q

$\Rightarrow \mu_q$ is linear in q

$q_i \langle x_i - \mu_q, v\rangle^3 \Rightarrow$ non-convex cubic in q

$$\sup_{\|v\|_2 \le 1} \underline{\sum_{i=1}^{n} q_i \langle x_i - \mu_q, v\rangle^3}$$

$$\mathbb{E}_{X\sim q}\left[\langle X - \mu_q, v\rangle^2\right]$$

$$= \mathbb{E}_{X\sim q}\left[v^T (X-\mu_q)(X-\mu_q)^T v\right]$$

$$= v^T \mathbb{E}_{X\sim q}\left[(X-\mu_q)(X-\mu_q)^T\right] v$$

$$= v^T \Sigma_q v$$

$$\sup_{\|v\|_2 \le 1} v^T \Sigma_q v = \|\Sigma_q\|$$

Proposition) Suppose $\tilde{p}$, p^* are as above and $\varepsilon < 1/3$ and $\|\mathrm{Cov}_{p^*}\| \leq \sigma^2$. Then, any stationary point q of $(\star)$ satisfies:

(A) • $TV(q, p^*) \leq \frac{\varepsilon}{1-\varepsilon}$

(B) • $\|\mathrm{Cov}_q\| \leq \left(\frac{1-\varepsilon}{\sqrt{1-3\varepsilon}}\right)^2 \sigma^2$

(C) • $\|\mu_q - \mu_{p^*}\|_2 \leq \frac{\sqrt{4\varepsilon(1-2\varepsilon)}}{1-3\varepsilon} \cdot \sigma$

Note. (C) follows directly from (A), (B).

Lemma) If $TV(p,q) \leq \varepsilon$, then

$$\|\mu_p - \mu_q\|_2 \leq \sqrt{\frac{\varepsilon}{1-\varepsilon}} \left(\sqrt{\|\Sigma_p\|} + \sqrt{\|\Sigma_q\|}\right).$$

Proof. Midpoint lemma + Chebyshev

Optimization q: $q_1, \dots, q_n$ on $x_1, \dots, x_n$

$$(*) \quad \min_{q_1,\dots,q_n} \sup_{\|v\|_2 \le 1} \underline{\sum_{i=1}^n q_i \langle x_i - \overset{\mu_{p^*}}{\cancel{\mu_q}}, v \rangle^2}$$

$$\text{s.t.} \quad \mu_q = \sum_{i=1}^n q_i x_i, \quad \underline{q_i \ge 0, \ \sum_i q_i = 1}$$

$$\underline{q_i \le \frac{1}{(1-\varepsilon)n}}$$

Issue. non-convex in q b/c $\mu_q \Rightarrow$ cubic

Idea. Suppose μ_q w/ $\mu_{p^*} \Rightarrow$ convex

Any "reasonable" q will have $\mu_q \approx \mu_{p^*}$

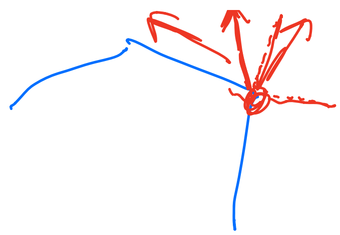
Pay a bit for $\|\mu_q - \mu_{p^*}\|$, but not enough to create problems.

Analyzing stationary points of $(*)$

Let $F(q) = \sup_{\|v\|_2 \le 1} \underbrace{\sum_{i=1}^n q_i \langle x_i - \mu_q, V \rangle^2}_{F_v(q)}$

- $\nabla F(q) = 0$
- q is at boundary of feasible set and $\nabla F(q)$ points outside feasible set

$\nabla F(q)$



$$\langle \nabla F(q), p - q \rangle \geq 0$$

for all feasible p.

general algebraic form

$$\boxed{\nabla F(q) = \nabla F_{v^*}(q)},$$

where v^* maximizes

$$v^{*T} \Sigma_q v^*$$

Danskin's theorem

$$\langle \nabla F_{v^*}(q), p - q \rangle \geq 0$$

$$\nabla_q F_{v^*}(q)$$

$$\frac{\partial}{\partial q_i} F_{v^*}(q) = \frac{\partial}{\partial q_i}\left(\sum_{j=1}^{n} q_j \langle x_j - \mu_q, v\rangle^2\right)$$

$$= \langle x_i - \mu_q, v\rangle^2 + \cancel{\left(2\sum_j q_j \langle x_j - \mu_q, v\rangle\right)\frac{\partial \mu_q}{\partial x_i}}$$

$$= 0$$

$$= \langle x_i - \mu_q, v\rangle^2$$

$$\sum_{i=1}^{n} (p_i - q_i)\langle x_i - \mu_q, v\rangle^2 \geq 0 \quad \forall\, p \in \Delta$$

feasible set

$$\boxed{\mathbb{E}_q\left[\langle X - \mu_q, v^*\rangle^2\right] \leq \mathbb{E}_p\left[\langle X - \mu_q, v^*\rangle^2\right]}$$

For some maximizing v^*

$\to q$ concentrated on $(1-\varepsilon)\cdot n$ smallest values of $\langle X_i - \mu_q, v^*\rangle^2$

Danskin's theorem / Envelope theorem

$$\nabla_\theta \sup_v F(\theta, v) = \nabla_\theta F(\theta, v^*), \text{ where}$$

$$v^* = \arg\max_v F(\theta, v)$$

↳ Issues if v^* not unique

$\nabla F(q)$ may not exist

$\frac{\partial}{\partial x_1} \max(x_1, x_2)$ doesn't exist if $x_1 = x_2$

Rescue w/ "Clarke sub-differential"

Lemma. If q is a stationary point of $(*)$, then for all $p \in \Delta$, there exists v^* maximizing $v^* \Sigma_q v^*$ s.t.

$\|\Sigma_q\|$ $\qquad\qquad\qquad\qquad$ $\approx \|\Sigma_p\|$

$$\mathbb{E}_q\left[\langle X-\mu_q, v^*\rangle^2\right] \leq \mathbb{E}_p\left[\langle X-\mu_q, v^*\rangle^2\right]$$

μ_p

If μ_q on RHS replaced w/ μ_p, this would say that q global minimum

---

$$\|\Sigma_q\| = v^{*T}\Sigma_q v^*$$

$$= \mathbb{E}_q\left[\langle X-\mu_q, v^*\rangle^2\right]$$

$$\leq \mathbb{E}_p\left[\langle X-\mu_q, v^*\rangle^2\right]$$

$(X-\mu_p) + (\mu_p-\mu_q)$

$$= \mathbb{E}_p\left[\langle X-\mu_p, v^*\rangle^2 + 2\langle X-\mu_p, v^*\rangle\langle \mu_p-\mu_q, v^*\rangle + \langle \mu_p-\mu_q, v^*\rangle^2\right]$$

0

$$= \mathbb{E}_p\left[\langle X-\mu_p, v^*\rangle^2\right] + \langle \mu_p-\mu_q, v^*\rangle^2$$

$$\leq \|\Sigma_p\| \qquad + \|\mu_p-\mu_q\|_2^2$$

$$\|\Sigma_q\| \leq \|\Sigma_p\| + \|\mu_p - \mu_q\|_2^2$$

Need to bound this

Bound using median of result

1) $TV(p, q) \leq \frac{\varepsilon}{1-\varepsilon}$ for any feasible $p, q \in \Delta$.

2) Therefore,

$$\|\mu_p - \mu_q\|_2 \leq \sqrt{\frac{\varepsilon/(1-\varepsilon)}{1-\varepsilon/(1-\varepsilon)}} \cdot \left(\sqrt{\|\Sigma_p\|} + \sqrt{\|\Sigma_q\|}\right)$$

$$= \sqrt{\frac{\varepsilon}{1-2\varepsilon}} \left(\sqrt{\|\Sigma_p\|} + \sqrt{\|\Sigma_q\|}\right)$$

$$\|\Sigma_q\| \leq \|\Sigma_p\| + \frac{\varepsilon}{1-2\varepsilon}\left(\sqrt{\|\Sigma_p\|} + \sqrt{\|\Sigma_q\|}\right)^2$$

$a = \sqrt{\|\Sigma_q\|}$ quadratic in a

$$\frac{\varepsilon}{1-2\varepsilon}\|\Sigma_q\|$$

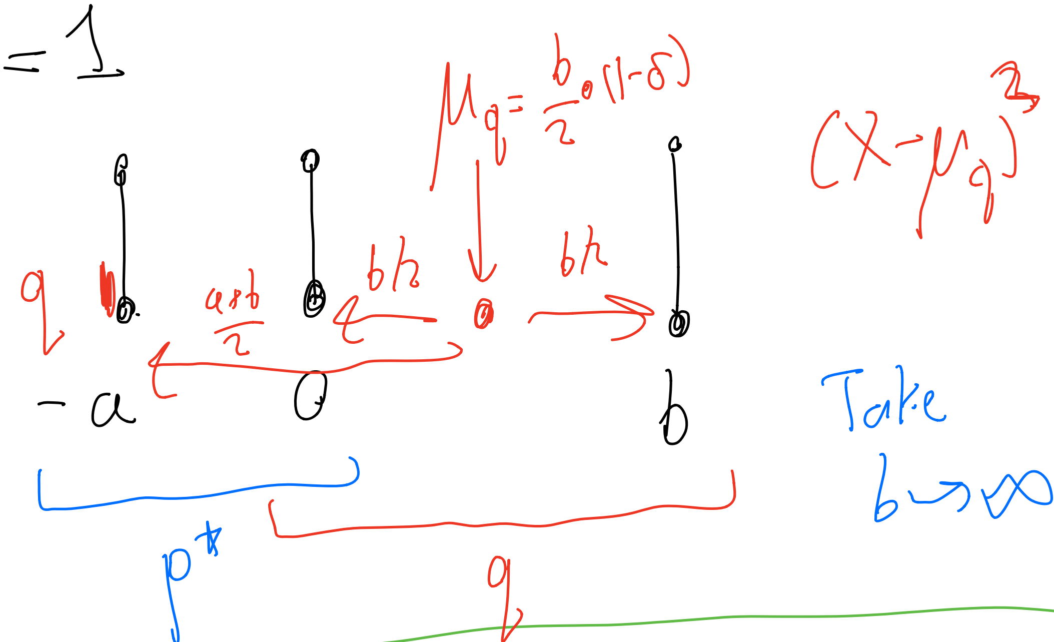
$\frac{\varepsilon}{1-2\varepsilon} < 1$ $\Rightarrow$ should be fine

$\varepsilon < 1/3$

$$\|\Sigma_q\| \leq \left(\frac{1-\varepsilon}{1-3\varepsilon}\right)^2 \|\Sigma_p\| \Rightarrow \|\mu_q - \mu_{p^*}\|_2 \leq \frac{\sqrt{4\varepsilon(1-2\varepsilon)}}{1-3\varepsilon} \cdot \sigma$$

Lower bound for $\varepsilon = \frac{1}{3}$.

$d = 1$



$$\tilde{p} = \frac{1}{3}\left(\delta_{-a} + \delta_{0} + \delta_{b}\right)$$

$$p^* = \frac{1}{2}\left(\delta_{-a} + \delta_{0}\right)$$