Efficient Algorithms for Robust Linear Regression

Logistics: Pset 3 now posted (due March 9th)
Lecture notes updated through Lecture 14

Continuing resilience beyond TV distance

Last time:

- Showed hypercontractivity + bounded noise $\Rightarrow$ resilience for lin reg.

$$\mathbb{E}_p[\langle x,v\rangle^4] \le \kappa\, \mathbb{E}_p[\langle x,v\rangle^2]^2 \quad \forall\, v \in \mathbb{R}^d$$

$$\mathbb{E}_p[x z^2 x^\top] \preceq \sigma^2 \cdot \mathbb{E}[x x^\top]$$

$\forall\, v \in \mathbb{R}^d$

$\hookrightarrow \mathbb{E}_p[\langle x,v\rangle^2 \cdot z^2] \le \sigma^2 \cdot \mathbb{E}[\langle x,v\rangle^2]$

$\hookrightarrow z = y - \langle \theta^*(p), x\rangle^2.$

Main result: If $\epsilon \le \frac{1}{8}$ and $\kappa\epsilon \le \frac{1}{6}$, p is $(\rho, 5\rho, \epsilon)$-resilient with $\rho \le 2\sigma^2\epsilon$. $\Rightarrow$ can handle ϵ-corruptions in TV w/ error $5\rho \le 10\sigma^2\epsilon$.

This time, Construct efficient algo that achieves same result (w/o constants).

Assume finite sample set $x_1, \dots, x_n$, subset S of "good" points, $|S| = (1-\epsilon)\cdot n$.

Idea. Similar to mean estimation.

Recall: $F(q) = \|\Sigma_q\| = \sup_{\|v\|_2 \le 1} \mathbb{E}_q[\langle x,v\rangle^2]$

$$\min_q F(q) \quad \text{s.t.} \quad \boxed{q \ge 0,\ \sum_i q_i = 1,\ q_i \le \frac{1}{(1-\epsilon)n}.}$$

$q \in \Delta_{n,\epsilon}$

$\rightarrow$ Showed: $\cancel{\nabla F(q) = 0} \Rightarrow q$ is approximate global minimum / $\tilde{p}$: uniform over any stationary point

Linear regression

$x_1, \dots, x_n$

$\tilde{p}_i = \frac{1}{n} \ \forall i$

$q \le \frac{p}{1-\varepsilon} \Rightarrow q_i \le \frac{\tilde{p}_i}{1-\varepsilon} = \frac{1}{(1-\varepsilon)n}$.

$$F_1(q) = \sup_v \frac{\mathbb{E}_q[\langle x, v\rangle^4]}{\mathbb{E}_q[\langle x, v\rangle^2]^2}$$

$$F_2(q) = \sup_v \frac{\mathbb{E}_q[\langle x, v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2]}{\mathbb{E}_q[\langle x, v\rangle^2]}$$

Find q s.t. $\boxed{\begin{array}{l} F_1(q) \le K \\ F_2(q) \le \sigma^2 \end{array}}$ $\boxed{q \in \Delta_{n,\varepsilon}}$.

$$\min_q \max\left(\frac{F_1(q)}{K}, \frac{F_2(q)}{\sigma^2}\right) \text{ s.t. } q \in \Delta_{n,\varepsilon}$$

Wrinkles

1. F_1 and F_2 instead of just F
2. Sup over v for F_1 intractable (b/c of 4^{th} moment)
   - ↳ SDP relaxation, but no Grothendieck (small set expansion problem)
   - ↳ assume not just hypercontractive, but "certifiably" so
3. $\nabla F_1(q)$ and $\nabla F_2(q)$ are uglier than for mean estimation
   $\Rightarrow$ may not be the case that stationary points are good

$$F_1(q) = \sup_v \frac{\mathbb{E}_q[\langle x, v\rangle^4]}{\mathbb{E}_q[\langle x, v\rangle^2]^2}$$

$$\nabla F_1(q)_i = \frac{\langle x_i, v\rangle^4}{\mathbb{E}_q[\langle x, v\rangle^2]^2} - 2\frac{\mathbb{E}_q[\langle x, v\rangle^4]\langle x_i, v\rangle^2}{\mathbb{E}_q[\langle x, v\rangle^2]^3}$$

"Quasigradient descent"

$$g_1(x_i, q) = \langle x_i, v\rangle^4, \quad \text{where } v \in \operatorname*{argmax}_{\|v\|_2 = 1} \frac{\mathbb{E}_q[\langle x, v\rangle^4]}{\mathbb{E}_q[\langle x, v\rangle^2]^2}$$

$$g_2(x_i, q) = \langle x_i, v\rangle^2 (y_i - \langle \theta^*(q), x_i\rangle)^2,$$

$$\text{where } v \in \operatorname*{argmax}_{\|v\|_2 = 1} \frac{\mathbb{E}_q[\langle x, v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2]}{\mathbb{E}_q[\langle x, v\rangle^2]}$$

Looking back at mean estimation:

Key lemma:

$$\mathbb{E}_q[\underbrace{\overset{g(x,q)}{\langle x - \mu_q, v\rangle^2}}_{\text{gradient}}] \leq \mathbb{E}_{p_S}[\overset{g(x,q)}{\langle x - \mu_q, v\rangle^2}] \text{ (by stationarity)}$$

$$\Longrightarrow \underset{\|\Sigma_q\|}{F(q)} \leq 3\sigma^2$$

nothing special about this being the gradient

Proof strategy:

1. Show can achieve stationarity for any quasigradient
2. stationarity $\Rightarrow F_1, F_2$ small

Algorithm: QuasigradientDescentLinReg

Input: $(X_1,y_1),\dots,(X_n,y_n)$, K, σ^2

Initialize $q \in \Delta_{n,\varepsilon}$ arbitrarily

while $F_1(q) \geq 2K$ or $F_2(q) \geq 4\sigma^2$

  if $F_1(q) \geq 2K$

    Let $g_{1,i} = \langle X_i, v\rangle^4$, where $v \in \operatorname{argmax}_{\|v\|_2=1} \dfrac{\mathbb{E}_q[\langle X,v\rangle^4]}{\mathbb{E}_q[\langle X,v\rangle^2]^2}$

    Take projected gradient step in g_1 direction.

  else

    Let $\theta^*(q) = \left(\sum_{i=1}^n q_i X_i X_i^T\right)^{-1}\left(\sum_{i=1}^n q_i X_i Y_i\right)$

    Let $g_{2,i} = \langle X_i, v\rangle^2 (Y_i - \langle \theta^*(q), X_i\rangle)^2$, where $v \in \operatorname{argmax} \dots$

    Take projected gradient step in g_2 direction.

  end

end

Output $\theta^*(q)$.

<u>Proposition</u>. Suppose p_S w/ $|S| = (1-\varepsilon)n$, s.t.

$$\mathbb{E}_{p_S}[\langle X,v\rangle^4] \leq K\, \mathbb{E}_{p_S}[\langle X,v\rangle^2]^2 \quad \forall v \in \mathbb{R}^d$$

$$\mathbb{E}_{p_S}[\cdots] \leq \sigma^2\, \mathbb{E}_{p_S}[\cdots] \quad \forall v \in \mathbb{R}^d$$

$\varepsilon \leq \frac{1}{8}$, $K\varepsilon \leq \frac{1}{6}$ $\qquad\qquad$ $10\sigma^2\varepsilon$

Then assuming $K\varepsilon \le \frac{1}{8e}$, Alg. terminates and its output $\theta^*(q)$ satisfies $L(p_s, \theta^*(q)) \le 4e\sigma^2\varepsilon$.

① Stationarity for quasigradients.

Lemma (Informal)

Asymptotically, Alg. generates q s.t.

$$\mathbb{E}_q[g_j(X, q)] \le \mathbb{E}_{p_s}[g_j(X, q)]$$

for $j = 1, 2$.

② Use stationarity to show F_1, F_2 small.

$\Rightarrow L$ small

- Stationarity for $g_1 \Rightarrow F_1$ small.

Lemma. Suppose $\mathbb{E}_{X\sim q}[g_1(X; q)] \le \mathbb{E}_{X\sim p_s}[g_1(X; q)]$ and $K\varepsilon \le \frac{1}{8e}$. Then q is hypercontractive w/ parameter $K' \le 1.5K$.

Pf. $g_1(x_i, q) = \langle x_i, v\rangle^4$, where v maximizes $K(q)$.

WANT: $\mathbb{E}_q[\langle x,v\rangle^2]^2$ large

$\mathbb{E}_q[\langle x,v\rangle^4]$ small

$\hookrightarrow \mathbb{E}_q[\langle x,v\rangle^4] \le \mathbb{E}_{p_3}[\langle x,v\rangle^4]$ (stationarity)

this ratio $\le K$

Goal: $\mathbb{E}_q[\langle x,v\rangle^2]^2 \ge \boxed{\frac{2}{3}} \mathbb{E}_{p_3}[\langle x,v\rangle^2]^2$

$$\left| \mathbb{E}_q[\langle x,v\rangle^2] - \mathbb{E}_{p_3}[\langle x,v\rangle^2] \right| \le \sqrt{\frac{\varepsilon}{(1-2\varepsilon)^2}\left(\mathbb{E}_q[\langle x,v\rangle^4] + \mathbb{E}_{p_3}[\langle x,v\rangle^4]\right)}$$

apply resilience Chebyshev to $\langle x,v\rangle^2$

Chebyshev + $TV(q, p_3) \le \frac{\varepsilon}{1-\varepsilon}$.

$$\le \sqrt{\frac{2\varepsilon}{(1-2\varepsilon)^2} \mathbb{E}_{p_3}[\langle x,v\rangle^4]}$$

$$\le \sqrt{\frac{2\varepsilon K}{(1-2\varepsilon^2)}} \mathbb{E}_{p_3}[\langle x,v\rangle^2]$$

$\le \frac{1}{6}$ if $K\varepsilon \le \frac{1}{80}$.

$$\Rightarrow \left| \mathbb{E}_q[\langle x, v\rangle^2] - \mathbb{E}_{p_3}[\langle x, v\rangle^2] \right| \le \frac{1}{6} \mathbb{E}_{p_3}[\langle x, v\rangle^2]$$

$$\Rightarrow \mathbb{E}_q[\langle x, v\rangle^2]^2 \ge \left(\frac{5}{6}\right)^2 \mathbb{E}_{p_3}[\langle x, v\rangle^2]^2$$

$$\ge \frac{2}{3}$$

$$\Rightarrow \mathbb{E}_q[\langle x, v\rangle^2]^2 \ge \boxed{\frac{2}{3}} \mathbb{E}_{p_3}[\langle x, v\rangle^2]^2.$$

__Lemma__ Suppose that $F_1(q) \le 2K$ and $\mathbb{E}_{X \sim q}[g_2(X, q)] \le \mathbb{E}_{p_3}[g_2(X, q)]$, and $K\varepsilon \le \frac{1}{80}$.

Then q has bounded noise w/ $(\sigma')^2 \le 4\sigma^2$, and also

$$L(p_3, \theta^*(q)) \le 40\sigma^2 \varepsilon.$$

Pf. $g_2(x_i, q) = \langle x_i, v\rangle^2 (y_i - \langle \theta^*(q), x_i\rangle)^2$

Bounded noise for q:

$\mathbb{E}_q[\langle x,v\rangle^2 (y - \langle\theta^*(q), x\rangle)^2]$ small relative to $\mathbb{E}_q[\langle x,v\rangle^2]$

$\le \mathbb{E}_{p_s}[\langle x,v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2]$

$\theta^*(p_s)$

Apply same strategy as in mean estimation

- Take hit depending on $\theta^*(q) - \theta^*(p_s)$
- Bound using resilience

$y - \langle\theta^*(p_s), x\rangle$ $\quad$ $\langle \theta^*(q) - \theta^*(p_s), x\rangle$

$$\mathbb{E}_{p_s}[\langle x,v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2]$$

$$\le 2\Big(\underbrace{\mathbb{E}_{p_s}[\langle x,v\rangle^2 (y - \langle\theta^*(p_s), x\rangle)^2]}_{(a)} + \underbrace{\mathbb{E}_{p_s}[\langle x,v\rangle^2 \langle\theta^*(p_s) - \theta^*(q), x\rangle^2]}_{(b)}\Big)$$

(a): By bounded noise, $(a) \le \sigma^2 \, \mathbb{E}_{p_s}[\langle x,v\rangle^2]$

(b): $\mathbb{E}_{p_s}[\langle x,v\rangle^2 \langle \theta^*(p_s) - \theta^*(q), x\rangle^2]$

$$\le \mathbb{E}_{p_s}[\langle x,v\rangle^4]^{1/2} \, \mathbb{E}_{p_s}[\langle \underbrace{\theta^*(p_s) - \theta^*(q)}_{\text{"}v\text{"}}, x\rangle^4]^{1/2}$$

$$\le K\, \mathbb{E}_{p_s}[\langle x,v\rangle^2] \, \mathbb{E}_{p_s}[\langle \theta^*(p_s) - \theta^*(q), x\rangle^2]$$

$$\downarrow$$

$$(\theta^*(p_3) - \theta^*(q))^T S_{p_3} (\theta^*(p_3) - \hat{\theta}(q))$$

$$= L(p_3, \theta^*(q)) \stackrel{\triangle}{=} R$$

$$(b) \le KR\, \mathbb{E}_{p_3}[\langle x, v\rangle^2]$$

$$\mathbb{E}_{p_3}[\langle x,v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2] \le 2(\sigma^2 + KR)\mathbb{E}_{p_3}[\langle x,v\rangle^2]$$

$$\mathbb{E}_q[\langle x,v\rangle^2 (y - \langle \theta^*(q), x\rangle)^2]$$

$$\le 2(\sigma^2 + KR)\,\mathbb{E}_{p_3}[\langle x,v\rangle^2]$$

$$\le 2.5(\sigma^2 + KR)\,\mathbb{E}_q[\langle x,v\rangle^2]$$

$\Rightarrow$ Shown bounded noise w/

$$(\sigma')^2 \le 2.5(\sigma^2 + KR).$$

Last lect. $(p, 5\rho, \varepsilon)$-resilient w/ $\rho = 2\sigma^2\varepsilon$

Resilience: $R \le 5\rho(K', \sigma')$

$$\leq 10(\sigma')^2 \varepsilon$$

$$= 25(\sigma^2 + K\underline{R})\varepsilon$$

$$R(1 - 25K\varepsilon) \leq 25\sigma^2\varepsilon$$

$$R \leq \frac{25\sigma^2\varepsilon}{1 - 25K\varepsilon} \leq 40\sigma^2\varepsilon. \quad \blacksquare$$