Problem Set 4 Solutions

Samuel B. Hopkins

December 5, 2024

Problem 1. In this problem, we will solve a sparse version of robust mean estimation. Let $\mu \in \mathbb{R}^d$ be an unknown k-sparse vector, in that only k of its entries are non-zero. First $n = \widetilde{\Omega}(k^2(\log d)/\varepsilon^2)$ samples $v_1, \dots, v_n \in \mathbb{R}^d$ are drawn from $\mathcal{N}(\mu, \mathrm{Id})$. Then an adversary alters εn of the samples and reorders them arbitrarily. We observe the resulting dataset $v'_1, \dots, v'_n$. Our goal will be to give an algorithm for estimating μ from these samples.

(a) Let $\overline{v} = \frac{1}{n}\sum_{i=1}^n v_i$. Prove that with 0.99 probability, for all k-sparse vectors $u \in \mathbb{R}^d$ with $\|u\| = 1$,

$$\langle u, \overline{v} - \mu \rangle^2 \leq \varepsilon^2 \,.$$

(b) Define $\Sigma = \frac{1}{n}\sum_{i=1}^n (v_i - \overline{v})(v_i - \overline{v})^T$. Prove that with 0.99 probability, $|\Sigma_{ij}| \leq 1/k$ for $i \neq j$ and $|\Sigma_{ii} - 1| \leq 1/k$ for all $i, j \in [d]$.

(c) Consider the following system, which we call $\mathcal{S}$, with scalar variables $w_1, \dots, w_n$ and d-dimensional variables $z, z_1, \dots, z_n$

$$\begin{aligned}
& w_i^2 = w_i \\
& \sum_{i=1}^n w_i \geq (1-\varepsilon)n \\
& w_i(z_i - v'_i) = 0 \\
& \overline{z} = \frac{1}{n}\sum_{i=1}^n z_i \,,\ \Sigma = \frac{1}{n}\sum_{i=1}^n (z_i - \overline{z})(z_i - \overline{z})^T \\
& -\frac{1}{k} \leq \Sigma_{ij} \leq \frac{1}{k} \quad \text{for all } i \neq j \\
& -\frac{1}{k} \leq \Sigma_{ii} - 1 \leq \frac{1}{k} \quad \text{for all } i
\end{aligned}$$

Prove that with 0.99 probability, there is a feasible solution to this system where the w_i are indicators of the clean samples and the z_i are the actual clean samples.

From now on, assume that the events in (a), (b), (c) hold.

(d) Now we consider the SoS relaxation of the system $\mathcal{S}$. Let $u \in \mathbb{R}^d$ be an arbitrary k-sparse vector with $\|u\| = 1$. Prove that

$$\mathcal{S} \vdash_2 \sum_{i=1}^n \langle u, z_i - v_i \rangle^2 \leq 10n(1 + \langle u, \overline{z} - \mu \rangle^2)$$

where recall v_i are the clean samples drawn from $N(\mu, I)$.

(e) Let $u \in \mathbb{R}^d$ be an arbitrary k-sparse vector with $\|u\| = 1$. Use part (c) to prove that

$$\mathcal{S} \vdash_4 \langle u, \overline{z} - \overline{v} \rangle^2 \leq 100\varepsilon(1 + \langle u, \overline{z} - \mu \rangle^2)$$

(f) Use part (e) to deduce that

$$\mathcal{S} \vdash_4 \langle u, \overline{z} - \mu \rangle^2 \leq O(\varepsilon) .$$

Put everything together to show that there is a polynomial time algorithm that takes the samples $v'_1, \ldots, v'_n$ and with probability 0.9, outputs a k-sparse $\widehat{\mu}$ such that $\|\mu - \widehat{\mu}\| \leq O(\sqrt{\varepsilon})$.

Solution

(a) Note that $\overline{v} - \mu$ is distributed as $\mathcal{N}(0, (1/n)\mathrm{Id})$. For ease of notation, let Z be distributed as $\mathcal{N}(0, (1/n)\mathrm{Id})$, so the goal is to show that with probability 0.99, for all unit k-sparse vectors u,

$$\langle u, Z \rangle^2 \leq \varepsilon.$$

By Cauchy-Schwarz,

$$\langle u, Z \rangle^2 \leq \sum_{i : u_i \neq 0} Z_i^2.$$

By the definition of sparsity, $\{i : u_i \neq 0\}$ is some set of size at most k. Fix some subset $S \subseteq [d]$ of size k. Then,

$$\Pr\left[\sum_{i \in S} Z_i^2 \geq \varepsilon^2\right] = \Pr_{Y \sim \mathcal{N}(0, \mathrm{Id}_k)}\left[\|Y\|^2 \geq n\varepsilon^2\right] .$$

Standard concentration bounds (e.g. Bernstein's inequality used with the subexponentiality of χ^2 random variables) imply that this quantity is bounded by $\exp\left(-\Omega\left(\frac{\varepsilon^2 n}{k}\right)\right) = \exp(-\Omega(k \log d))$. We can now take a union bound over all subsets S of size at most k – there are $\exp(O(k \log(d)))$ such subsets, completing the proof if we take the constant factor in n sufficiently large.

(b) We have

$$\begin{aligned}
\Sigma_{ij} &= \frac{1}{n} \sum_{r=1}^{n} (v_r - \overline{v})_i (v_r - \overline{v})_j \\
&= \frac{1}{n} \sum_{r=1}^{n} (v_r - \mu)_i (v_r - \mu)_j + (\mu - \overline{v})_i (\mu - \overline{v})_j + (v_r - \overline{v})_i (\mu - \overline{v})_j + (\mu - \overline{v})_i (v_r - \overline{v})_j \\
&= (\overline{v} - \mu)_i (\overline{v} - \mu)_j + \frac{1}{n} \sum_{r=1}^{n} (v_r - \mu)_i (v_r - \mu)_j
\end{aligned}$$

Since the distinct coordinates of $(\overline{v} - \mu)$ are distributed as independent copies of $\mathcal{N}(0, 1/n)$, with probability at least 0.95 (say), all their absolute values are less than $1/2k$. Indeed,

$$\Pr_{X \sim \mathcal{N}(0, 1/n)}\left[|X| \geq \frac{1}{2k}\right] \leq \exp\left(-\Omega(n/4k^2)\right) = O\left(\frac{1}{d}\right),$$

and we can take a union bound over the d coordinates (X above is one of the coordinates of $\overline{v}-\mu$). Given that this event happens, note that the first term of Σ_{ij} can be bounded using Cauchy-Schwarz by $\frac{1}{2}\left((\overline{v}-\mu)_i^2+(\overline{v}-\mu)_j^2\right) \le 1/2k$, so it now suffices to bound the second term. That is, shifting the points v_r back by μ to get w_r, we wish to show that given $w_1,\dots,w_n \sim \mathcal{N}(0,\mathrm{Id})$,

$$\left\|\frac{1}{n}\sum_{r=1}^{n} w_r w_r^\top - \mathrm{Id}\right\|_\infty \le \frac{1}{2k}$$

with high probability. For the diagonal entries, this probability is

$$\Pr\left[\frac{1}{n}\sum_{r=1}^{n}(w_r)_i^2 - 1 \ge \frac{1}{2k}\right] = \Pr_{X\sim\mathcal{N}(0,\mathrm{Id}_n)}\left[\|X\|^2 - n \ge \frac{n}{2k}\right].$$

Applying standard concentration bounds again, this is $\exp(-\Omega(n/k^2)) = O(1/d^2)$. For off-diagonal entries ij, this probability is

$$\Pr\left[\frac{1}{n}\sum_{r=1}^{n}(w_r)_i(w_r)_j \ge \frac{1}{2k}\right] = \Pr_{X,X'\sim\mathcal{N}(0,\mathrm{Id}_n)}\left[\langle X, X'\rangle \ge \frac{n}{2k}\right] = \Pr_{\substack{X\sim\mathcal{N}(0,\mathrm{Id}_n)\\ X'\sim\mathcal{N}(0,1)}}\left[\|X\|X' \ge \frac{n}{2k}\right],$$

where the final inequality is because conditioned on X, the projection of X' on the direction of X is distributed as a standard one-dimensional Gaussian. Standard Gaussian tail bounds imply that this is bounded as

$$\begin{aligned}
\Pr_{\substack{X\sim\mathcal{N}(0,\mathrm{Id}_n)\\ X'\sim\mathcal{N}(0,1)}}\left[X' \ge \frac{n}{2k\|X\|}\right] &\le \mathbb{E}_{X\sim\mathcal{N}(0,\mathrm{Id}_n)}\left[\exp\left(-\Omega\left(\frac{n^2}{k^2\|X\|^2}\right)\right)\right] \\
&\le \Pr_{X\sim\mathcal{N}(0,\mathrm{Id}_n)}[\|X\|^2 \ge 2n] + \exp\left(-\Omega\left(\frac{n^2}{k^2 n}\right)\right) \\
&\le \Pr_{X\sim\mathcal{N}(0,\mathrm{Id}_n)}[\|X\|^2 \ge 2n] + \exp\left(-\Omega\left(\frac{n}{k^2}\right)\right) = O(1/d^2),
\end{aligned}$$

where the second inequality is because $\exp(-z)$ is at most 1 for $z \ge 0$, and the final inequality follows exactly like the earlier concentration bound and the fact that $n = \Omega(k^2 \log d)$.
Taking a union bound over all d^2 entries of the matrix completes the proof.

(c) This part is immediate from (a) and (b). First condition on these events (since both happen with probability 0.99). For simplicity, suppose that the adversary does not reorder the sample, so if a sample is not corrupted, then $v_i = v_i'$. Then, one can choose $w_i = \mathbf{1}[v_i = v_i']$, and $z_i = v_i$ for all i. The first constraint is trivially satisfied since each $w_i \in \{0,1\}$, the second because at most εn corruptions are introduced, the third by the definition of w_i and z_i, and the last two by (b).

(d) By Cauchy-Schwarz, we have

$$\mathcal{S} \vdash_2 \frac{1}{n}\sum_{i=1}^{n}\langle u, z_i - v_i\rangle^2 \le 4\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n}\langle u, z_i - \overline{z}\rangle^2}_{\text{(I)}} + \langle u, \overline{z}-\mu\rangle^2 + \underbrace{\langle u, \mu - \overline{v}\rangle^2}_{\text{(II)}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\langle u, \overline{v} - v_i\rangle^2}_{\text{(III)}}\right).$$

First, note that (a) implies that (II) is bounded by ε^2. To bound (III), set $\Sigma' = (1/n)\sum_{i=1}^n (v_i - \overline{v})(v_i - \overline{v})^\top$. By (b), $|\Sigma' - \mathrm{Id}|_\infty \le 1/k$. Then, setting $S = \{i : u_i \neq 0\}$ (with $|S| \le k$),

$$\begin{aligned}
\text{(III)} &= u^\top \Sigma' u \\
&= \sum_{i \in S} u_i^2 \Sigma'_{ii} + \sum_{\substack{i,j \in S \\ i \neq j}} u_i u_j \Sigma'_{ij} \\
&\le \left(1 + \frac{1}{k}\right) + \sum_{\substack{i,j \in S \\ i \neq j}} \frac{u_i^2 + u_j^2}{2} \cdot \frac{1}{k} \\
&\le \left(1 + \frac{1}{k}\right) + \sum_{i \in S} \frac{1}{k} = 2 + \frac{1}{k}.
\end{aligned}$$

Similarly, by the SoS constraints, setting $\Sigma'' = (1/n)\sum_{i=1}^n (z_i - \overline{z})(z_i - \overline{z})^\top$, we have $\mathcal{S} \vdash_2 \|\Sigma'' - \mathrm{Id}\|_\infty \le 1/k$, so

$$\mathcal{S} \vdash_2 \text{(I)} = u^\top \Sigma'' u \le 2 + \frac{1}{k}.$$

Therefore, putting the pieces together, we get that

$$\mathcal{S} \vdash_2 \frac{1}{n}\sum_{i=1}^n \langle u, z_i - v_i \rangle^2 \le 4\left(4 + \frac{2}{k} + \varepsilon^2 + \langle u, z - \mu \rangle^2\right) = O(1) \cdot (1 + \langle u, z - \mu \rangle^2),$$

as desired.

(e) Note that $\mathcal{S} \vdash_2 (z_i - v_i) w_i \mathbf{1}_{v_i = v'_i} = 0$. Consequently

$$\begin{aligned}
\mathcal{S} \vdash_4 \langle u, \overline{z} - \overline{v} \rangle^2 &= \left(\frac{1}{n}\sum_{i=1}^n \langle u, z_i - v_i \rangle\right)^2 \\
&= \left(\frac{1}{n}\sum_{i=1}^n \langle u, (z_i - v_i)(1 - \mathbf{1}_{v_i = v'_i} w_i) \rangle\right)^2 \\
&\le \left(\frac{1}{n}\sum_{i=1}^n \langle u, z_i - v_i \rangle^2\right)\left(\frac{1}{n}\sum_{i=1}^n (1 - \mathbf{1}_{v_i = v'_i} w_i)^2\right),
\end{aligned}$$

where the final inequality follows by Cauchy-Schwarz. Now,

$$\mathcal{S} \vdash_2 \sum_{i=1}^n (1 - \mathbf{1}_{v_i = v'_i} w_i)^2 = \sum_{i=1}^n (1 - \mathbf{1}_{v_i = v'_i} w_i) = \sum_{i=1}^n (1 - w_i) + (1 - \mathbf{1}_{v_i = v'_i}) \underbrace{w_i}_{\mathcal{S} \vdash_2 w_i \le 1} \le 2\varepsilon n,$$

where we used the second constraint in the sum-of-squares system to bound the first sum by εn, and that at most εn of the v_i were corrupted to bound the second term by εn. Substituting this back in the earlier string of equations, and using (d), we get that

$$\langle u, \overline{z} - \overline{v} \rangle^2 \le O(\varepsilon)(1 + \langle u, \overline{z} - \mu \rangle^2)$$

as desired.

(f) Indeed, we have

$$\begin{aligned}
\mathcal{S} \vdash_4 \langle u, \overline{z} - \mu \rangle^2 &\le 2\langle u, \overline{z} - \overline{v} \rangle^2 + 2\langle u, \overline{v} - \mu \rangle^2 \\
&\le O(\varepsilon)(1 + \langle u, \overline{z} - \mu \rangle^2) + O(\varepsilon^2) \\
\langle u, \overline{z} - \mu \rangle^2 &\le O(\varepsilon)
\end{aligned}$$

as desired. Here, the first inequality is Cauchy-Schwarz, and the second by (a) and (e).

The final algorithm is as follows. We find a pseudoexpectation $\widetilde{\mathbb{E}}$ that is feasible for the sum-of-squares relaxation $\mathcal{S}$, and output $\hat{\mu}$, such that $\hat{\mu}_i = (\widetilde{\mathbb{E}}\,\overline{z})_i$ for the k coordinates with largest $(\widetilde{\mathbb{E}}\,\overline{z})_i$. Just as in usual robust mean estimation, we have that for any k-sparse unit vector u,

$$\varepsilon \ge \widetilde{\mathbb{E}}\langle u, \overline{z} - \mu \rangle^2 \ge \langle u, \widetilde{\mathbb{E}}\,\overline{z} - \mu \rangle^2,$$

so $\langle u, \widetilde{\mathbb{E}}\,\overline{z} - \mu \rangle = O(\sqrt{\varepsilon})$. Let $S = \{i : \hat{\mu}_i \ne 0\}$ and $T = \{i : \mu_i \ne 0\}$. For simplicity, assume that $|S| = |T| = k$ (in general one can add in arbitrary coordinates where $\hat{\mu}_i$ or μ_i is 0). Set $z' = (\widetilde{\mathbb{E}}\,z)|_{S \setminus T}$, $z'' = (\widetilde{\mathbb{E}}\,z)|_{S \cap T}$, and $z''' = (\widetilde{\mathbb{E}}\,z)|_{T \setminus S}$ – observe that $\hat{\mu} = z' + z''$. Similarly, set $\mu'' = \mu|_{S \cap T}$ and $\mu''' = \mu|_{T \setminus S}$. Choosing u as the unit vectors in the directions of z', $z'' - \mu''$, and $z''' - \mu'''$, we get that each of $\|z'\|, \|z'' - \mu''\|, \|z''' - \mu'''\|$ is $O(\sqrt{\varepsilon})$. Note that since z consists of the k largest coordinates of $\widetilde{\mathbb{E}}\,z$, and $|S \setminus T| = |T \setminus S|$, $\|z'''\| \le \|z'\| = O(\sqrt{\varepsilon})$. Therefore,

$$\|\hat{\mu} - \mu\| = \|z' + z'' - \mu'' - \mu'''\| \le \|z'\| + \|z'' - \mu''\| + \|z''' - \mu'''\| + \|z'''\| \le O(\sqrt{\varepsilon})$$

as desired. $\square$

Problem 2. Recall the *planted clique* problem, with the "null distribution" $\mathcal{N} = G(n, 1/2)$, and the "planted distribution" $\mathcal{P}$ obtained by drawing G from $G(n, 1/2)$, and adding a uniformly random k-clique. It is believed that for k significantly smaller than $O(\sqrt{n})$ (say $O(n^{1/2-\varepsilon})$), it is computationally hard to distinguish these two distributions. In this question, we will establish this computational hardness for the restricted class of algorithms based on low-degree polynomials.

Concretely, set $k = O(n^{1/2-\varepsilon})$ for some (small) constant $\varepsilon > 0$, and $D \le C \log n$ for some (large) constant $C > 0$. Recall the degree-D χ^2-divergence, defined by

$$\sqrt{\chi^2_{\le D}(\mathcal{P}\|\mathcal{N})} = \max_{\substack{F:\{\text{set of graphs on } n \text{ vertices}\}\to\mathbb{R} \\ F \text{ degree} \le D \text{ polynomial} \\ F \text{ not identically } 0}} \frac{\mathbb{E}_{\mathcal{P}}[F] - \mathbb{E}_{\mathcal{N}}[F]}{\sqrt{\mathbf{Var}_{\mathcal{N}}[F]}}.$$

Further recall that this maximum is attained by the function $\left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D}$, where $\frac{\mathcal{P}}{\mathcal{N}}$ is the likelihood ratio $\frac{\mathcal{P}}{\mathcal{N}}(G) = \frac{\mathcal{P}(G)}{\mathcal{N}(G)}$ and the notation $f^{\le D}$ denotes the projection of f to the space of degree D polynomials. This resulting maximum is equal to

$$\chi^2_{\le D}(\mathcal{P}\|\mathcal{N}) = \left\| \left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D} - 1 \right\|_2^2,$$

with the notation $\|f\|_2^2 = \mathbb{E}_{\mathcal{N}} f^2$.

(a) Let $g = \left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D}$ be a polynomial of degree D in the variables $(x_e)_{e\in\binom{[n]}{2}}$, where $x_e = 1$ if e is an edge in the graph, and -1 otherwise. Express g in terms of its Fourier coefficients as $g = \sum_{\alpha:|\alpha|\le D} \widehat{g}_\alpha x^\alpha$. Determine $\widehat{g}_\alpha$.

(b) Show that in the given parameter regime of k, D, $\chi^2_{\le D}(\mathcal{P}\|\mathcal{N}) = \|g - 1\|^2 = o(1)$.

Solution

(a) By the definition of the Fourier coefficients (and the orthogonality of the associated polynomials),

$$\begin{aligned}
\widehat{g}_\alpha &= \left\langle \left(\frac{\mathcal{P}}{\mathcal{N}}\right)^{\le D}, x^\alpha \right\rangle_{\mathcal{N}} \\
&= \left\langle \left(\frac{\mathcal{P}}{\mathcal{N}}\right), x^\alpha \right\rangle_{\mathcal{N}} \\
&= \mathbb{E}_{\mathcal{P}}\, x^\alpha.
\end{aligned}$$

Note that if some edge in α does not have both its endpoints in the planted clique, the expectation becomes 0 as x^α is ± 1 with probability $1/2$. Thus, this Fourier coefficient is

$$\begin{aligned}
\widehat{g}_\alpha &= \Pr\left[\text{all of } V(\alpha) \text{ is touched by the clique}\right] \cdot \mathbb{E}_{\mathcal{P}}\left[\prod_{e\in\alpha} x_e \mid \text{all of } V(\alpha) \text{ is touched by the clique}\right] \\
&= \frac{\binom{n-|V(\alpha)|}{k-|V(\alpha)|}}{\binom{n}{k}}.
\end{aligned}$$

(b) We have

$$\begin{aligned}
\chi^2_{\leq D}(\mathcal{P} \| \mathcal{N}) &= \|g - 1\|^2 \\
&= \sum_{\substack{|\alpha| \leq D \\ \alpha \neq \emptyset}} \widehat{g}_\alpha^2 \\
&= \sum_{\substack{|\alpha| \leq D \\ \alpha \neq \emptyset}} \frac{\binom{n-|V(\alpha)|}{k-|V(\alpha)|}^2}{\binom{n}{k}^2} \\
&\leq \sum_{\substack{|\alpha| \leq D \\ \alpha \neq \emptyset}} \left(\frac{k}{n}\right)^{2|V(\alpha)|} \\
&\leq \sum_{1 \leq t \leq 2D} n^t \cdot \left(\sum_{r \leq D} \binom{\binom{t}{2}}{r}\right) \cdot \left(\frac{k}{n}\right)^{2t} \\
&\leq \sum_{1 \leq t \leq 2D} \left(\sum_{r \leq D} \binom{\binom{t}{2}}{r}\right) \cdot \left(\frac{k^2}{n}\right)^{t}.
\end{aligned}$$

Here, the second-to-last line follows because there are at most n^t ways to pick $V(\alpha)$ of size t, and at most $\left(\sum_{r \leq D} \binom{\binom{t}{2}}{r}\right)$ ways to choose the edges of α (of size at most D) from within these vertices. This may be bounded by $\min\left\{2^{t^2}, t^{2D}\right\}$, which are good bounds for small t and large t respectively. That is, our goal is now to bound

$$\sum_{1 \leq t \leq 2D} \min\left\{2^{t^2}, t^{2D}\right\} \cdot \left(\frac{k^2}{n}\right)^t.$$

Let us split this summation into two parts, using different bounds in each case. Setting $k = O(n^{1/2 - \varepsilon/2})$, we have

$$\chi^2_{\leq D}(\mathcal{P} \| \mathcal{N}) \leq \sum_{1 \leq t \leq \frac{\varepsilon}{2} \log n} 2^{t^2} \cdot n^{-\varepsilon t} + \sum_{\frac{\varepsilon}{2} \log n \leq t \leq C \log n} t^{2D} \cdot n^{-\varepsilon t}.$$

Each summand of the first sum can be bounded by

$$2^{t^2} \cdot n^{-\varepsilon t} \leq 2^{t \cdot (\varepsilon/2) \cdot \log n} \cdot n^{-\varepsilon t} \leq n^{-\varepsilon t/2} = o(1/\log n),$$

so the first sum is $o(1)$. For each term of the second sum, we have

$$t^{2D} \cdot n^{-\varepsilon t} = e^{2D \log t - \varepsilon t \log n}.$$

Note that because $\frac{\varepsilon}{2} \log n \leq t \leq C \log n$, $D \log t = O(\log n \cdot \log \log n)$, while $t \log n = \Omega(\log^2 n)$. As a result, each term of this sum is also $o(1/\log n)$, and the second sum is $o(1)$, completing the proof.