ORIE 6334 Bridging Continuous and Discrete Optimization Dec 4, 2019

Lecture 22

Lecturer: David P. Williamson *Scribe: Renee Mirka*

1 Discrepancy Minimization Problem and History

Given a collection of sets $S_1, \ldots, S_m \subseteq \{1, 2, \ldots, n\}$, the goal of the minimum discrepancy problem is to find $\chi : \{1, \ldots, n\} \to \{\pm 1\}$ to minimize

$$\max_{i=1,\ldots,m} |\chi(S_i)| \equiv \left| \sum_{j \in S_i} \chi(j) \right|.$$

In 1985, Spencer gave a nonconstructive proof that there exists χ such that $\max_{i=1,\ldots,m} |\chi(S_i)| \leq O(\sqrt{n \log(2m/n)})$ when $m \geq n$. This was followed by an SDP-based randomized polytime algorithm achieving Spencer's bound by Bansal in 2010. Our focus of today's lecture will be the following result from Lovett and Meka in 2012:

Theorem 1 *Let $v_1, \ldots, v_m \in \mathbb{R}^n$ be vectors with $\|v_i\| \leq 1$ for all i, $x_0 \in [-1, 1]^n$ and $\lambda_1, \ldots, \lambda_m \geq 0$ be parameters such that $\sum_{i=1}^m \exp(-\lambda_i^2/16) \leq n/32$. Then we can compute $x \in [-1, 1]^n$ such that $\langle v_i, x - x_0 \rangle \leq 11\lambda_i$ for all i and $|\{j : x(j) = \pm 1\}| \geq n/2$.*

We will show how this result implies Spencer. But first, consider a polytope $P = \{x \in [-1, 1]^n : |\langle v_i, x - x_0 \rangle| \leq \lambda_i\}$. Lovett and Meka's procedure was to start at x_0 and take a random walk in P. Once the walk hits a face (i.e. $\chi(j) = 1, \chi(j) = -1$, or $|\langle v_i, x - x_0 \rangle| = \lambda_i$) the walk 'sticks' to it, and will eventually reach the desired point. Rothvoss modified this idea slightly in 2014. Instead of choosing a random walk, start at x_0, go in a random direction, and find where this direction intersects P. This intersection point is the desired point with some constant probability.

Today: we will see a deterministic, multiplicative weight style algorithm by Levy, Ramadas, Rothvoss (2017).

2 Theorem 1 Implies Spencer's result

First we show how the theorem implies Spencer's result. Consider the following algorithm.

Observations:

Algorithm 1: Spencer Bound

$x_0 \leftarrow \overrightarrow{0}$;
for $s \leftarrow 1, \ldots, \log_2 n$ **do**
  $A_s \leftarrow \{j \in [n] : -1 < x_s(j) < 1\}$;
  $v_i \leftarrow \frac{1}{\sqrt{|A_s|}} \mathbb{1}_{S_i \cap A_s}$ (i.e. 1s in j s.t. $j \in S_i \cap A_s$, 0 otherwise);
  $\lambda_i \leftarrow c\sqrt{\ln \frac{2m}{|A_s|}}, i = 1, \ldots, m$;
  Run alg to get $x_s \in [-1, 1]^n$ s.t.
   $x_s(j) = x_{s-1}(j) \forall j \notin A_s, \langle v_i, x_s - x_{s-1} \rangle \leq 11\lambda_i$;
end
return $\overline{x} = x_{\log_2 n}$

- $\overline{x} \in \{\pm 1\}^n$.
- $\|v_i\| \leq 1$ for all i.
- The theorem applies since $\sum_{k=1}^m e^{-\lambda_i^2/16} = \sum_{i=1}^m e^{-\frac{c^2}{16} \ln \frac{2m}{|A_s|}} \leqslant \frac{|A_s|}{32}$ for good choice of c.
- $|A_s| \leq n/2^{s-1}$ for all s.
- If we have that $\langle v_i, x_s - x_{s-1} \rangle \leq 11\lambda_i$, then

$$\sum_{j \in S_i} [x_s(j) - x_{s-1}(j)] \leqslant O\left(\sqrt{|A_s| \ln \left(\frac{2m}{|A_s|}\right)}\right)$$

  for all s. By summing over all s, we get that

$$\sum_{j \in S_i} \bar{x}(j) \leqslant \sum_{s=1}^{\log_2 n} O\left(\sqrt{2^{-(s-1)} n \ln \left(\frac{2m}{2^{-(s-1)} n}\right)}\right) = O\left(\sqrt{n \ln \left(\frac{2m}{n}\right)}\right),$$

  since the first term dominates.

3 Algorithm to Prove Theorem 1

Assume WLOG $\lambda_i \leq 2\sqrt{n}$, since if $\lambda_i > 2\sqrt{n}, \langle v_i, x - x_0 \rangle \leq \lambda_i$ does not intersect $[-1, 1]^n$. (In other words, the constraint doesn't do anything.)

Definition 1 *Let* $\delta \equiv 1/\lambda_1$ *be the* step size *and* $\rho_i \equiv \exp(-\frac{4\delta^2 \lambda_i^2}{n}) \leq 1$ *be the* discount factor.

The algorithm will run for $O(n/\delta^2)$ iterations $\implies O(n^2)$ iterations.

Algorithm 2: Lovett-Meka

$w_0(i) \leftarrow e^{-\lambda_i^2}, \forall i (\implies \sum_{i=1}^m w_0(i) \le n/32)$;
for $t \leftarrow 1$ **to** ∞ **do**
  Pick unit vector z_t in the span of $\{e_j : -1 < x_t(j) < 1\}$,
  of eigenvectors of $\frac{15}{16}n$ largest eigenvalues of $M_t \equiv \sum_{i=1}^m w_t(i) v_i v_i^T$
  and $\perp$ to x_t,
  to v_i for $\frac{n}{16} i$ that have largest weights $w_t(i)$,
  to v_i for i with $\lambda_i \le 1$, to $\sum_{i=1}^m \lambda_i w_t(i) \rho_i v_i$;
  Choose max $\alpha_t \in (0,1]$ s.t. $x_{t+1} = x_t + \alpha_t \delta z_t \in [-1,1]^n$;
  $w_{t+1}(i) \leftarrow w_t(i) \exp(\lambda_i \delta \langle v_i, \alpha_t z_t \rangle) \rho_i$;
  if $|\{j : -1 < x_{t+1}(j) < 1\}| < \frac{n}{2}$ **then**
    stop;
  end
end

Note: $w_t(i) = \exp\left(\lambda_i \langle v_i, x_t - x_0 \rangle\right) \cdot \rho_i^t \cdot e^{-\lambda_i^2}$ so the weights are exponentially large in the amount by which a constraint is violated, but with discount factor.

Lemma 2 *For all iterations t, we can always pick z_t.*

Proof: We are picking z_t from a space of dimension $\ge n/2 - n/16$ and orthogonal to a space of dimension $\le 1 + n/16 + n/16 + 1$ since $\sum_{i=1}^n e^{-\lambda_i^2/16} \le \frac{n}{32}$ and $e^{-1/16} \ge 1/2$ implies that $|\{i : \lambda_i \le 1\}| \le n/16$. Therefore, we pick z_t from a space of dimension $\ge n/2 - n/16 - n/16 - n/16 - 2 \ge \frac{5}{16}n - 2 \ge 1$ for $n \ge 10$. □

Lemma 3 *The algorithm terminates after $O(n/\delta^2)$ iterations.*

Proof: $\|x_{t+1}\|^2 = \|x_t + \delta \alpha_t z_t\|^2 = \|x_t\|^2 + 2\delta\alpha_t \langle x_t, z_t \rangle + \delta^2 \alpha_t^2 ||z_t||^2 = ||x_t||^2 + \delta^2 \alpha_t^2$, since z_t is orthogonal to z_t. If $\alpha_t = 1$, $||x_{t+1}||^2 = ||x_t||^2 + \delta^2$. We can have $\alpha_t < 1$ at most n times, since each such time $x_{t+1}(j) \in \{\pm 1\}$ for some new index j. Since $x_t \in [-1,1]^n$, $||x_t||^2 \le n$. Therefore, the total number of iterations is at most $n + n/\delta^2$. □

Let $W_t \equiv \sum_{i=1}^m w_t(i)$. The following is the Main Lemma.

Lemma 4 *$W_{t+1} \le W_t$ for all t.*

Note this lemma all says $W_t \le n/32$ for all t. Let T denote the final iteration.

Lemma 5 *$w_T(i) \le 2$ for all i.*

Proof: Suppose otherwise. Let t^* be the last iteration for which i is not among the $n/16$ highest weights. After t^*,

$$w_{t+1}(i) = w_t(i)\exp(\lambda_i \delta \langle v_i, \alpha_t z_t \rangle)\rho_i = w_t(i)\rho_i,$$

since z_t will be chosen orthogonal to v_i when $w_t(i)$ is among the $n/16$ highest weights. This shows that $w_{t+1}(i) \leq w_t(i)$ for $t > t^*$. So,

$$2 < w_T(i) \leq w_{t^*+1}(i) = w_{t^*}(i)\exp(\lambda_i \delta \langle v_i, \alpha_t z_t \rangle)\rho_i \leq w_{t^*}(i)e,$$

since $\|v_i\| \leq 1$, $\alpha_t \leq 1$, $\|z_t\| = 1$, and $\lambda_i \delta \leq 1$. Therefore $w_{t^*}(i) > 2/e$; since i isn't among the $n/16$ highest weights, there exist $n/16$ j such that $w_{t^*}(j) > 2/e$. But this means $W_{t^*} > (n/16)(2/e) \geq n/32$, which contradicts the main lemma. □

Theorem 6 $\langle v_i, \overline{x} - x_0 \rangle \leq 11\lambda_i$.

Proof: If $\lambda_i \leq 1$, then by construction $z_t \perp v_i$ for all t, so that $\langle v_i, \overline{x} - x_0 \rangle = 0 \leq \lambda_i$. Otherwise,

$$w_T(i) = \exp(\lambda_i \langle v_i, \overline{x} - x_0 \rangle)\rho_i^T e^{-\lambda_i^2} \leq 2.$$

Taking the log of both sides,

$$\lambda_i \langle v_I, \overline{x} - x_0 \rangle + T\ln\left(\exp(-\frac{4\delta^2\lambda_i^2}{n})\right) - \lambda_i^2 \leq \ln 2.$$

From here we see

$$\langle v_i, \overline{x} - x_0 \rangle \leq \frac{\ln 2}{\lambda_i} + \lambda_i\left(1 + 4T\frac{\delta^2}{n}\right) \leq 2 + \lambda_i(1+8) \leq 11\lambda_i.$$

For the penultimate inequality, we recall that $\lambda_i > 1$, $T \leq n + n/\delta^2$, and $\delta \leq 1$. □

The next lemma will help prove the main lemma.

Lemma 7 *For any possible z_t, $z_t^T M_t z_t \leq \frac{16}{n}\sum_{i=1}^m w_t(i)\lambda_i^2$.*

Proof: $\mathrm{tr}(M_t) = \sum_{i=1}^m w_t(i)\lambda_i^2 \,\mathrm{tr}\left(v_i v_i^T\right) = \sum_{i=1}^m w_t(i)\lambda_i^2$. Since $M_t \succeq 0$, at most $n/16$ eigenvalues can have value at least $\frac{16}{n}\mathrm{tr}(M_t)$. Therefore, z_t is in the span of eigenvectors of M_t of eigenvalue at most $\frac{16}{n}\mathrm{tr}(M_t)$, so $z_t^T M_t z_t \leq \frac{16}{n}\sum_{i=1}^m w_t(i)\lambda_i^2$. □

Lastly, we provide the proof of the main lemma (Lemma 4).

Proof:

$$
\begin{aligned}
W_{t+1} &= \sum_{i=1}^{m} w_{t+1}(i) = \sum_{i=1}^{m} w_t(i) \exp\left(\lambda_i \delta \langle v_i, \alpha_t z_t\rangle\right) \rho_i \\
&\leq \sum_{i=1}^{m} w_t(i) \left(1 + \lambda_i \delta \left\langle v_i, \alpha_t z_t\right\rangle + \lambda_i^2 \delta^2 \left\langle v_i, \alpha_t z_t\right\rangle^2\right) \cdot \rho_i, \quad \text{using } e^x \leq 1 + x + x^2 \text{ for } |x| \leq 1 \\
&= \sum_{i=1}^{m} w_t(i)\rho_i + \delta \langle \sum_{i=1}^{m} \lambda_i w_t(i) \rho_i v_i, \alpha_t z_t\rangle + \delta^2 \sum_{i=1}^{m} w_t(i) \lambda_i^2 \rho_i \langle v_i, \alpha_t z_t\rangle^2 \\
&= \sum_{i=1}^{m} w_t(i) \cdot \rho_i + \delta^2 \alpha_t^2 z_t^T M_t z_t, \quad \text{using } z_t \perp \sum_{i=1}^{m} \lambda_i w_t(i) \rho_i v_i \\
&\leq \sum_{i=1}^{m} w_t(i)\rho_i + \delta^2 \frac{16}{n} \sum_{i=1}^{m} w_t(i) \lambda_i^2 \\
&\leq \sum_{i=1}^{m} w_t(i) = W_t, \quad \text{using } \rho_i = \exp(-\frac{4\delta^2\lambda_i^2}{n}), \text{ since } e^{-x} \leq 1 - x/2 \text{ for } 0 \leq x \leq 1.
\end{aligned}
$$

□