CSE 525: Randomized Algorithms **Spring 2025**

Lecture 17: Generic Chaining

Lecturer: Shayan Oveis Gharan *06/02/25*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

17.1 A Concrete Setting

Suppose that we have a subset $T \subseteq \mathbb{R}^n$, we pick a random Gaussian vector g from $N(\mathbf{0}, I)$, and we are interested in the random variable

$$\sup_{t \in T} \langle g, t \rangle \tag{17.1}$$

In theoretical computer science, for example, a random variable like (17.1) comes up often in the study of rounding algorithms for semidefinite programming, but this is a problem of much broader interest.

We will be interested both in bounds on the expectations of (17.1) and on its tail, but in this lecture we will mostly reason about its expectation.

The first observation is that by rotational invariance property of Gaussians $\langle g, t \rangle$ is distribution as a (centered) normal $\mathcal{N}(0, \|t\|^2)$-random variable. So,

$$\mathbb{P}\left[\langle g, t \rangle > \ell\right] \leq e^{-\ell^2/(2\|t\|^2)}$$

So, if T is finite, by a simple union bound we can write

$$\Pr\left[\sup_{t \in T} \langle g, t \rangle > \ell\right] \leq |T| \cdot e^{-\ell^2/(2 \sup_{t \in T} \|t\|^2)}$$

It thus follows that

$$\mathbb{E}_{g \sim N(\mathbf{0}, I)}\left[\sup_{t \in T} \langle g, t \rangle\right] \leq O\left(\sqrt{\log |T|} \cdot \sup_{t \in T} \|t\|\right) \tag{17.2}$$

The above bound can be tight, but it is poor if the points of T are densely clustered, and it is useless if T is infinite.

It is useful to note that, if we fix, arbitrarily, an element $t_0 \in T$, then we have

$$\mathbb{E}_{g \sim N(\mathbf{0}, I)}\left[\sup_{t \in T} \langle g, t \rangle\right] = \mathbb{E}_{g \sim N(\mathbf{0}, I)}\left[\sup_{t \in T} \langle g, t - t_0 \rangle\right] \tag{17.3}$$

because $\mathbb{E}\langle g, t_0 \rangle = 0$, as g is a mean zero random variable. The latter expression is nicer to work with because it makes it more explicit that what we are trying to compute is invariant under shifts of T, and only depends on pairwise distances of the elements of T, rather than their norm!

In the cases in which (17.2) gives a poor bound, a natural approach is to reason about an ϵ-net T' of T, that is, a subset $T' \subseteq T$ such that for every $t \in T$ there is an element $\pi(t) \in T'$ such that $\|t - \pi(t)\| \leq \epsilon$. Then we can say that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in T}\langle t-t_0,g\rangle\right] &= \mathbb{E}\left[\sup_{t\in T}\langle t-\pi(t),g\rangle + \langle \pi(t)-t_0,g\rangle\right] \\
&\le \mathbb{E}\left[\sup_{t\in T}\langle t-\pi(t),g\rangle\right] + \mathbb{E}\left[\sup_{t\in T}\langle \pi(t)-t_0,g\rangle\right] \\
&\le O(\sqrt{\log|T|})\cdot \sup_{t\in T}\|t-\pi(t)\| + \mathbb{E}\left[\sup_{t'\in T'}\langle t'-t_0,g\rangle\right] \\
&\le O(\sqrt{\log|T|})\cdot \epsilon + \sqrt{\log|T'|}\cdot \sup_{t'\in T'}\|t'-t_0\|. \qquad (17.4)
\end{aligned}$$

This potentially give a much tighter bound than (17.2). Now, similarly to bound $\mathbb{E}\sup_{t'\in T'}\langle t'-t_0,g\rangle$ it might be better to find an ϵ'-net T'' of T' and so on.

17.2 Generic Chaining

This kind of idea can be applied recursively: we construct a sequence of sets

$$T_0 \subseteq T_1 \subseteq T_2 \subseteq \cdots \subseteq T$$

where $T_0 = \{t_0\}$. We will also assume that for some large enough n, we have $T_n = T$. For every $i \ge 0$, let $\pi_i : T \to T_i$ denote a "closest point map" which sends $t \in T$ to the closest point in T_i.

We simplify notation and for $t \in T$ we write X_t to denote $\langle g,t\rangle$. The main point of "chaining" is that we can now write, for any $t \in T$,

$$X_t - X_{t_0} = \sum_{i\ge 1} X_{\pi_i(t)} - X_{\pi_{i-1}(t)}.$$

Having this we can rewrite (17.1) as

Lemma 17.1. *For every $s,t \in T$ and $\lambda > 0$,*

$$\mathbb{P}\left[X_s - X_t > \lambda\right] \le 2\exp\left(-\frac{\lambda^2}{2\,d(s,t)^2}\right),$$

where $d(s,t) = \|s-t\|$.

Instead of insisting that T_i has a certain level of accuracy, we'll insist that T_i is at most a certain size. Now, the question is should we require $|T_i| \le i$ or $|T_i| \le 2^i$, or use some other function? To figure out the right bound, we look at (17.4). If we want the facts $\sqrt{\log|T_i|}$ telescope, we need that $|T_i| > |T_{i-1}|^2$. Or to put differently, $|T_i| \le 2^{2^i}$.

This leads us to the generic chaining bound, due to Fernique (though the formulation we state here is from Talagrand).

Theorem 17.2. *Let $\{X_t\}_{t\in T}$ be a Gaussian process (as defined above), and let $T_0 \subseteq T_1 \subseteq \cdots \subseteq T$ be a sequence of subsets such that $|T_0| = 1$ and $|T_i| \le 2^{2^i}$ for $i \ge 1$. Then,*

$$\mathbb{E}\sup_{t\in T} X_t \le O(1)\sup_{t\in T}\sum_{i\ge 0} 2^{i/2} d(t,T_i).$$

Proof. By Lemma 17.1 we have for any $i \geq 1, t \in T$ and $\ell > 0$,

$$\mathbb{P}\left[|X_{\pi_i(t)} - X_{\pi_{i-1}(t)}| > \ell \cdot 2^{i/2} d(\pi_i(t), \pi_{i-1}(t))\right] \leq \exp\left(-\frac{\ell^2}{2} 2^i\right).$$

Now, the number of pairs $(\pi_i(t), \pi_{i-1}(t))$ can be bounded by $|T_i| \cdot |T_{i-1}| \leq 2^{2^{i+1}}$, so we have

$$\mathbb{P}\left[\exists t : X_{\pi_i(t)} - X_{\pi_{i-1}(t)} > \ell \cdot 2^{i/2} d(\pi_i(t), \pi_{i-1}(t))\right] \leq 2^{2^{i+1}} \exp\left(-\frac{\ell^2}{2} 2^i\right). \tag{17.5}$$

Let us now define the event

$$\mathcal{E}_\ell = \left\{\forall i \geq 1, t \in T : X_{\pi_i(t)} - X_{\pi_{i-1}(t)} \leq \ell \cdot 2^{i/2} d(\pi_i(t), \pi_{i-1}(t))\right\},$$

then, for $\ell \geq 4$, summing (17.5) yields,

$$\mathbb{P}\left[\overline{\mathcal{E}_\ell}\right] \leq \sum_{i \geq 1} 2^{2^{i+1}} \exp\left(-\frac{\ell^2}{2} 2^i\right) \leq \sum_i e^{-\ell^2 \cdot 2^{i-1} \cdot 0.72^{i+1}} = O(1) e^{-\ell^2} \tag{17.6}$$

Now, we write,

$$S = \sup_{t \in T} \sum_{i \geq 1} 2^{i/2} d(\pi_i(t), \pi_{i-1}(t)).$$

Note that if $\mathcal{E}_\ell$ occurs, then $\sup_{t \in T}(X_t - X_{t_0}) \leq \ell S$. Thus (17.6) implies that for $\ell \geq 4$,

$$\mathbb{P}\left[\sup_{t \in T} X_t - X_{t_0} > \ell S\right] \leq O(1)\, e^{-\ell^2},$$

which implies that

$$\mathbb{E}\left[\sup_{t \in T} X_t\right] \leq O(S) \leq O(1) \sup_{t \in T} \sum_{i \geq 1} 2^{i/2} d(\pi_i(t), \pi_{i-1}(t)).$$

Finally, by the triangle inequality,

$$d(\pi_i(t), \pi_{i-1}(t)) \leq d(t, T_i) + d(t, T_{i-1}) \leq 2\, d(t, T_{i-1}).$$

$\square$

17.3 Optimality

Theorem 17.2 gives us a fairly natural way to upper bound the expected supremum using a hierarchical clustering of T. Rather amazingly, this upper bound is tight. Talagrand's majorizing measure theorem states that if we take the best choice of $\{T_i\}$ in Theorem 17.2, then the upper bound in Theorem 17.2 is within a constant factor of $\mathbb{E} \sup_{t \in T} X_t$

Talagrand's theory allows one to capture the minimum possible such bound over all possible such sequences of nested sets. Specifically, Talagrand introduces the so-called γ_2 *functional*, defined as:

$$\gamma_2(T, d) = \inf \sup_{t \in T} \sum_{i=0}^{\infty} 2^{i/2} d(t, T_i)$$

where the infimum is taken over all sequences of sets $T_0 \subseteq T_1 \subseteq \cdots \subseteq T$ such that $|T_i| \leq 2^{2^i}$.

Then Talagrand proves the deep and beautiful theorem:

Theorem (Talagrand's Majorizing Measure Theorem). Let T be a set equipped with a metric d that is the metric induced by the standard Gaussian process (i.e., $d(s,t)^2 = \mathbb{E}[(\langle g, s\rangle - \langle g, t\rangle)^2] = \|s-t\|^2$). Then:

$$\mathbb{E}_{g\sim N(0,I)}\left[\sup_{t\in T}\langle g, t\rangle\right] = \Theta(\gamma_2(T,d))$$

17.4 Subgaussian Processes

Suppose we have a set of random variables $\{X_t : t \in T\}$; for simplicity we assume these random variables are centered, i.e., $\mathbb{E}\left[X\right]_t = 0$. We call such an object a random process. This random process is called sub-gaussian if there is some constant $c > 0$ such that

$$\mathbb{P}\left[X_s - X_t > \lambda\right] \leq \exp(-c\lambda^2/d(s,t)^2)$$

In particular, Gaussian processes are sub-gaussian with $c = 1/2$. In the above,

$$d(s,t) = \sqrt{\mathbb{E}\left[(\right] X_s - X_t)^2}.$$

If X is a Gaussian process then

$$d(s,t)^2 = \mathbb{E}\left[\langle(g,s\rangle - \langle g,t\rangle)^2\right] = \mathbb{E}\left[\langle g, s-t\rangle^2\right] = \|s-t\|^2.$$

It turns out that the same chaining argument applies to any sub-guassian process. Basically the constant c appears in the RHS of Theorem 17.2.

Another example is a Bernoulli process, i.e. a family of random variables defined thusly: For some subset $T \subseteq \mathbb{R}^n$,

$$\{\epsilon_1 t_1 + \epsilon_2 t_2 + \cdots + \epsilon_n t_n : t = (t_1, \ldots, t_n) \in T\},$$

where $\epsilon_1, \ldots, \epsilon_n \in \{-1, +1\}$ are i.i.d. uniformly random signs.

17.5 Hypergraph Sparsification

Consider a weighted hypergraph $H = (V, E, w)$ where $w_e : e \in E$ are nonnegative edge weights. We associated to H the quadratic expression

$$Q_H(x) = \sum_{e\in E} w_e \max_{u,v\in e}(x_u - x_v)^2.$$

The main observation is that If H were a graph i.e., for every edge e we had $|e| = 2$ for, this would correspond to the quadratic form of the graph Laplacian.

As our main application of chaining we will explain algorithms to sparsify hypergraphs: That is we want to construct another hypergraph $\tilde{H} = (V, \tilde{E}, \tilde{w})$ such that $\tilde{E} \subseteq E$ and such that

$$|Q_H(x) - Q_{\tilde{H}}(x)| \leq \epsilon \cdot QH(x), \qquad \forall x \in \mathbb{R}^n$$

where as usual $\epsilon > 0$ is the accuracy parameter of our sparsifier. Furthermore, similar to the graph case we would like to make $|\tilde{E}|$ as small as possible, ideally near-linear in $|V|$ (while —E— could be as large as $2^{|V|}$ in this case).