Chaining introduction

Jelani Nelson
Harvard

June 22, 2016
What’s this talk about?

Given a collection of random variables $X_1, X_2, \ldots$, we would like to say that $\max_i X_i$ is small with high probability. (Happens all over computer science, e.g. Chernoff+Union bound)

Today’s topic: Beating the Union Bound

Disclaimer: This is a tutorial talk, about ideas which aren’t mine.
What’s this talk about?

Given a collection of random variables $X_1, X_2, \ldots$, we would like to say that $\max_i X_i$ is small with high probability. (Happens all over computer science, e.g. Chernoff+Union bound)
What’s this talk about?

Given a collection of random variables $X_1, X_2, \ldots$, we would like to say that $\max_i X_i$ is small with high probability. (Happens all over computer science, e.g. Chernoff+Union bound)

**Today’s topic: Beating the Union Bound**
What’s this talk about?

Given a collection of random variables $X_1, X_2, \ldots$, we would like to say that $\max_i X_i$ is small with high probability. (Happens all over computer science, e.g. Chernoff+Union bound)

**Today’s topic: Beating the Union Bound**

Disclaimer: This is a tutorial talk, about ideas which aren’t mine.
Chaining applications in computer science

- Analyzing structured RIP matrices for compressed sensing (Candès, Tao’06), (Rudelson, Vershynin’06), (Cheragchi, Guruswami, Velingker’13), (N., Price, Wootters’14), (Bourgain’14), (Haviv, Regev’15)
- Fast Johnson-Lindenstrauss (JL) transforms (Ailon, Liberty’11), (Krahmer, Ward’11), (Bourgain, Dirksen, N.’15), (Oymak, Recht, Soltanolkotabi’15)
- Instance-wise JL bounds (Gordon’88), (Klartag, Mendelson’05), (Mendelson, Pajor, Tomczak-Jaegermann’07), (Dirksen’14)
- Approximate nearest neighbor (Indyk, Naor’07)
- Deterministic alg. to estimate graph cover time (Ding, Lee, Peres’11), (Zhai’14)
- List-decodability of random codes (Wootters’13), (Rudra, Wootters’14)
- Streaming heavy hitters (Braverman et al.’16), (Braverman et al.’17?)
- Dictionary learning (Luh, Vu’15), (Błasiok, Nelson’16)
- ...
Case study for this talk:

gaussian processes
Gaussian process setup

- $T \subset B_{\ell_2}$
Gaussian process setup

- $T \subset B_{\ell^2}$
- Random variables $(Z_x)_{x \in T}$
  
  $Z_x = \langle g, x \rangle$ for a vector $g$ with i.i.d. $\mathcal{N}(0, 1)$ entries
Gaussian process setup

- $T \subset B_{\ell^2}$
- Random variables $(Z_x)_{x \in T}$
  \[ Z_x = \langle g, x \rangle \] for a vector $g$ with i.i.d. $\mathcal{N}(0, 1)$ entries
  so $Z_x = \sum_i g_i x_i$ is a gaussian with variance $\sum_i x_i^2 \leq 1$
Gaussian process setup

- $T \subset B_{\ell^n_2}$
- Random variables $(Z_x)_{x \in T}$
  
  $Z_x = \langle g, x \rangle$ for a vector $g$ with i.i.d. $\mathcal{N}(0, 1)$ entries

  so $Z_x = \sum_i g_i x_i$ is a gaussian with variance $\sum_i x_i^2 \leq 1$

- Define \textit{gaussian mean width} $g(T) = \mathbb{E}_g \sup_{x \in T} Z_x$
Gaussian process setup

- $T \subset B_{\ell^2}$
- Random variables $(Z_x)_{x \in T}$
  
  $Z_x = \langle g, x \rangle$ for a vector $g$ with i.i.d. $\mathcal{N}(0, 1)$ entries

  so $Z_x = \sum_i g_i x_i$ is a gaussian with variance $\sum_i x_i^2 \leq 1$

- Define *gaussian mean width* $g(T) = \mathbb{E}_g \sup_{x \in T} Z_x$

- How can we bound $g(T)$?
Gaussian process setup

- $T \subset B_{\ell_2^n}$
- Random variables $(Z_x)_{x \in T}$
  
  $Z_x = \langle g, x \rangle$ for a vector $g$ with i.i.d. $\mathcal{N}(0, 1)$ entries
  
  so $Z_x = \sum_i g_i x_i$ is a gaussian with variance $\sum_i x_i^2 \leq 1$
- Define *gaussian mean width* $g(T) = \mathbb{E}_g \sup_{x \in T} Z_x$
- How can we bound $g(T)$?

**This talk**: four progressively tighter ways to bound $g(T)$, then discussion of lower bounds.
Gaussian mean width bound 1: union bound

- \( g(\mathcal{T}) = \mathbb{E} \sup_{x \in \mathcal{T}} Z_x = \mathbb{E} \sup_{x \in \mathcal{T}} \langle g, x \rangle \)
Gaussian mean width bound 1: union bound

- $g(T) = \mathbb{E} \sup_{x \in T} Z_x = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- $Z_x$ is a gaussian with variance at most one
Gaussian mean width bound 1: union bound

- \( g(T) = \mathbb{E} \sup_{x \in T} Z_x = \mathbb{E} \sup_{x \in T} \langle g, x \rangle \)
- \( Z_x \) is a gaussian with variance at most one

\[
\mathbb{E} \sup_{x \in T} Z_x \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) du
\]

\( \lesssim \sqrt{\log |T|} \)

(union bound)
Gaussian mean width bound 1: union bound

- $g(T) = \mathbb{E} \sup_{x \in T} Z_x = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- $Z_x$ is a gaussian with variance at most one

\[
\mathbb{E} \sup_{x \in T} Z_x \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) \, du = \int_0^{u^*} \mathbb{P}(\sup_{x \in T} Z_x > u) \, du + \int_{u^*}^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) \, du \\
\leq 1 + \int_{u^*}^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) \, du \\
\leq |T| \cdot e^{-u^2/2} \text{ (union bound)}
\]
Gaussian mean width bound 1: union bound

- \( g(T) = \mathbb{E} \sup_{x \in T} Z_x = \mathbb{E} \sup_{x \in T} \langle g, x \rangle \)
- \( Z_x \) is a gaussian with variance at most one

\[
\mathbb{E} \sup_{x \in T} Z_x \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) du
\]
\[
= \int_0^{u_*} \mathbb{P}(\sup_{x \in T} Z_x > u) du + \int_{u_*}^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) du
\]
\[
\leq 1 + \int_{u_*}^\infty \mathbb{P}(\sup_{x \in T} Z_x > u) du
\]
\[
\leq |T| \cdot e^{-u_*^2/2} \quad \text{(union bound)}
\]
\[
\leq u_* + |T| \cdot e^{-u_*^2/2}
\]
\[
\lesssim \sqrt{\log |T|} \quad \text{(set } u_* = \sqrt{2 \log |T|})
\]
Gaussian mean width bound 2: $\varepsilon$-net

- $g(T) = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- Let $S_{\varepsilon}$ be $\varepsilon$-net of $(T, \ell_2)$ (every $x \in T$ is $\varepsilon$-close to some $x' \in S_{\varepsilon}$ in $\ell_2$)
Gaussian mean width bound 2: $\varepsilon$-net

- $g(T) = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- Let $S_\varepsilon$ be $\varepsilon$-net of $(T, \ell_2)$ (every $x \in T$ is $\varepsilon$-close to some $x' \in S_\varepsilon$ in $\ell_2$)
- $\langle g, x \rangle = \langle g, x' \rangle + \langle g, x - x' \rangle$ ($x' = \operatorname{argmin}_{y \in S_\varepsilon} \|x - y\|_2$)
  
  $g(T) \leq g(S_\varepsilon) + \mathbb{E} g \sup_{x \in T} \langle g, x - x' \rangle$
  
  $\leq \varepsilon \cdot \|g\|_2$
Gaussian mean width bound 2: $\varepsilon$-net

- $g(T) = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- Let $S_{\varepsilon}$ be $\varepsilon$-net of $(T, l_2)$ (every $x \in T$ is $\varepsilon$-close to some $x' \in S_{\varepsilon}$ in $l_2$)
- $\langle g, x \rangle = \langle g, x' \rangle + \langle g, x - x' \rangle$ ($x' = \arg\min_{y \in S_{\varepsilon}} \|x - y\|_2$)
  $g(T) \leq g(S_{\varepsilon}) + \mathbb{E}_g \sup_{x \in T} \langle g, x - x' \rangle \leq \varepsilon \cdot \|g\|_2$
- $\lesssim \sqrt{\log |S_{\varepsilon}|} + \varepsilon (\mathbb{E}_g \|g\|_2^2)^{1/2}$
- $\lesssim \log^{1/2} \mathcal{N}(T, l_2, \varepsilon) + \varepsilon \sqrt{n}$ (smallest $\varepsilon$-net size)

Choose $\varepsilon$ to optimize bound; can never be worse than last slide (which amounts to choosing $\varepsilon = 0$)
Gaussian mean width bound 2: $\varepsilon$-net

- $g(T) = \mathbb{E} \sup_{x \in T} \langle g, x \rangle$
- Let $S_\varepsilon$ be $\varepsilon$-net of $(T, \ell_2)$ (every $x \in T$ is $\varepsilon$-close to some $x' \in S_\varepsilon$ in $\ell_2$)
- $\langle g, x \rangle = \langle g, x' \rangle + \langle g, x - x' \rangle$ ($x' = \arg\min_{y \in S_\varepsilon} \|x - y\|_2$)
  \[ g(T) \leq g(S_\varepsilon) + \mathbb{E}_g \sup_{x \in T} \langle g, x - x' \rangle \]
  \[ \leq \varepsilon \cdot \|g\|_2 \]
- \[ \lesssim \sqrt{\log |S_\varepsilon|} + \varepsilon (\mathbb{E}_g \|g\|_2^2)^{1/2} \]
- \[ \lesssim \log^{1/2} N(T, \ell_2, \varepsilon) + \varepsilon \sqrt{n} \]
  smallest $\varepsilon$–net size
- Choose $\varepsilon$ to optimize bound; can never be worse than last slide (which amounts to choosing $\varepsilon = 0$)
Gaussian mean width bound 3: $\varepsilon$-net sequence

- $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  - $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$
Gaussian mean width bound 3: $\varepsilon$-net sequence

- $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  - $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$
- wlog $|T| < \infty$ (else apply this slide to $\varepsilon$-net of $T$ for $\varepsilon$ small)
- $x = \pi_0 x + (\pi_1 x - \pi_0 x) + (\pi_2 x - \pi_1 x) + \ldots = \pi_0 x + \sum_{r=1}^{\infty} \Delta_r x$
Gaussian mean width bound 3: $\varepsilon$-net sequence

- $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$
- wlog $|T| < \infty$ (else apply this slide to $\varepsilon$-net of $T$ for $\varepsilon$ small)
- $x = \pi_0 x + (\pi_1 x - \pi_0 x) + (\pi_2 x - \pi_1 x) + \ldots = \pi_0 x + \sum_{r=1}^{\infty} \Delta_r x$
  $\Rightarrow \langle g, x \rangle = \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \langle g, \Delta_r x \rangle$
Gaussian mean width bound 3: $\varepsilon$-net sequence

- $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  - $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$
- wlog $|T| < \infty$ (else apply this slide to $\varepsilon$-net of $T$ for $\varepsilon$ small)
- $x = \pi_0 x + (\pi_1 x - \pi_0 x) + (\pi_2 x - \pi_1 x) + \ldots = \pi_0 x + \sum_{r=1}^{\infty} \Delta_r x$
  - $\Rightarrow \langle g, x \rangle = \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \langle g, \Delta_r x \rangle$
- $g(T) \leq \mathbb{E} \sup_{g} \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \mathbb{E}_g \sup_{x \in T} \langle g, \Delta_r x \rangle$
  - $= 0$
Gaussian mean width bound 3: $\varepsilon$-net sequence

• $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$

• wlog $|T| < \infty$ (else apply this slide to $\varepsilon$-net of $T$ for $\varepsilon$ small)

• $x = \pi_0 x + (\pi_1 x - \pi_0 x) + (\pi_2 x - \pi_1 x) + \ldots = \pi_0 x + \sum_{r=1}^{\infty} \Delta_r x$
  $\Rightarrow \langle g, x \rangle = \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \langle g, \Delta_r x \rangle$

• $g(T) \leq \mathbb{E} \sup_{g \in T} \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \mathbb{E} g \sup_{x \in T} \langle g, \Delta_r x \rangle$

• $|\{\Delta_r x : x \in T\}| \leq \mathcal{N}(T, \ell_2, 1/2^r) \cdot \mathcal{N}(T, \ell_2, 1/2^{r-1})$
  $\leq (\mathcal{N}(T, \ell_2, 1/2^r))^2$

• $\sigma_r = \sup_{x \in T} ||\Delta_r x||_2 = \sup ||\pi_r x - x + x - \pi_{r-1} x|| \leq \frac{1}{2^r} + \frac{1}{2^{r-1}}$
Gaussian mean width bound 3: $\varepsilon$-net sequence

- $S_r$ is a $(1/2^r)$-net of $T$, $r \geq 0$
  $\pi_r x$ is closest point in $S_r$ to $x \in T$, $\Delta_r x = \pi_r x - \pi_{r-1} x$
- wlog $|T| < \infty$ (else apply this slide to $\varepsilon$-net of $T$ for $\varepsilon$ small)
- $x = \pi_0 x + (\pi_1 x - \pi_0 x) + (\pi_2 x - \pi_1 x) + \ldots = \pi_0 x + \sum_{r=1}^{\infty} \Delta_r x$
  $\Rightarrow \langle g, x \rangle = \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \langle g, \Delta_r x \rangle$
- $g(T) \leq \mathbb{E} \sup_{g, x \in T} \langle g, \pi_0 x \rangle + \sum_{r=1}^{\infty} \mathbb{E} \sup_{x \in T} \langle g, \Delta_r x \rangle$
  $\leq \mathcal{N}(T, \ell_2, 1/2^r) \cdot \mathcal{N}(T, \ell_2, 1/2^{r-1})$
  $\leq (\mathcal{N}(T, \ell_2, 1/2^r))^2$
- $\sigma_r = \sup_{x \in T} \|\Delta_r x\|_2 = \sup \|\pi_r x - x + x - \pi_{r-1} x\| \leq \frac{1}{2^r} + \frac{1}{2^{r-1}}$
- $g(T) \lesssim \sum_{r=1}^{\infty} (1/2^r) \cdot \log^{1/2} \mathcal{N}(T, \ell_2, 1/2^r)$
  $\lesssim \int_0^{\infty} \log^{1/2} \mathcal{N}(T, \ell_2, u) du$ (Dudley’s theorem)
Again, wlog $|T| < \infty$. Define $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_{k^*} = T$

$|T_0| = 1$, $|T_k| \leq 2^{2^k}$ (call such a sequence “admissible”)

**Gaussian mean width bound 4: generic chaining**
Gaussian mean width bound 4: generic chaining

• Again, wlog $|T| < \infty$. Define $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_{k^*} = T$
  $|T_0| = 1, |T_k| \leq 2^{2^k}$ (call such a sequence “admissible”)

• Exercise: show Dudley’s theorem is equivalent to
  $g(T) \lesssim \inf \{T_k\} \text{ admissible } \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell^2}(x, T_k)$
  (should pick $T_k$ to be the best $\varepsilon = \varepsilon(k)$ net of size $2^{2^k}$)
Gaussian mean width bound 4: generic chaining

- Again, wlog $|T| < \infty$. Define $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_{k*} = T$
  $|T_0| = 1, |T_k| \leq 2^k$ (call such a sequence “admissible”)
- Exercise: show Dudley’s theorem is equivalent to
  $g(T) \lesssim \inf \{T_k\}_{\text{admissible}} \sum_{k=1}^{\infty} 2^{k/2} \sup_{x \in T} d_{\ell^2}(x, T_k)$
  (should pick $T_k$ to be the best $\varepsilon = \varepsilon(k)$ net of size $2^{2^k}$)
- Fernique’76*: can pull the $\sup_x$ outside the sum
- $g(T) \lesssim \inf \{T_k\} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell^2}(x, T_k) \overset{\text{def}}{=} \gamma_2(T, \ell^2)$
Gaussian mean width bound 4: generic chaining

- Again, wlog $|T| < \infty$. Define $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_{k^*} = T$
  $|T_0| = 1, |T_k| \leq 2^k$ (call such a sequence “admissible”)
- Exercise: show Dudley’s theorem is equivalent to
  \[ g(T) \lesssim \inf \{ T_k \text{ admissible} \} \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell^2}(x, T_k) \]
  (should pick $T_k$ to be the best $\varepsilon = \varepsilon(k)$ net of size $2^{2^k}$)
- Fernique’76*: can pull the sup$_x$ outside the sum
- \[ g(T) \lesssim \inf \{ T_k \} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell^2}(x, T_k) \overset{\text{def}}{=} \gamma_2(T, \ell^2) \]
Gaussian mean width bound 4: generic chaining

- Again, wlog $|T| < \infty$. Define $T_0 \subseteq T_1 \subseteq \cdots \subseteq T_{k^*} = T$
  - $|T_0| = 1$, $|T_k| \leq 2^k$ (call such a sequence “admissible”)
- Exercise: show Dudley’s theorem is equivalent to
  $$g(T) \lesssim \inf \{T_k\} \text{ admissible } \sum_{k=1}^{\infty} 2^{k/2} \sup_{x \in T} d_{\ell_2}(x, T_k)$$
  (should pick $T_k$ to be the best $\varepsilon = \varepsilon(k)$ net of size $2^{2^k}$)
- Fernique’76*: can pull the $\sup_x$ outside the sum
- $$g(T) \lesssim \inf \{T_k\} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell_2}(x, T_k) \overset{\text{def}}{=} \gamma_2(T, \ell_2)$$
  * equivalent upper bound proven by Fernique (who minimized some integral over all measures over $T$), but reformulated in terms of admissible sequences by Talgarand
Gaussian mean width bound 4: generic chaining

Proof of Fernique’s bound

\[ g(T) \leq \mathbb{E} \sup_{x \in T} \langle g, \pi_0 x \rangle + \mathbb{E} \sup_{x \in T} \sum_{k=1}^{\infty} \langle g, \Delta_k x \rangle \] (from before)

\[ \forall t, \mathbb{P}(Y_k(x) > t2^{k/2} \| \Delta_k x \|_2) \leq e^{-t2^{2k}/2} \] (gaussian decay)
Gaussian mean width bound 4: generic chaining

Proof of Fernique’s bound

\[ g(T) \leq \mathbb{E} \sup_{x \in T} \langle g, \pi_0 x \rangle + \mathbb{E} \sup_{x \in T} \sum_{k=1}^{\infty} \langle g, \Delta_k x \rangle \quad (\text{from before}) \]

- \( \forall t, \mathbb{P}(Y_k(x) > t2^{k/2}\|\Delta_k x\|_2) \leq e^{-t^22^k/2} \) (gaussian decay)
- \( \mathbb{P}(\exists x, k \ Y_k(x) > t2^{k/2}\|\Delta_k x\|_2) \leq \sum_k (2^k)^2 e^{-t^22^k/2} \)
Proof of Fernique’s bound

\[ g(T) \leq \mathbb{E} \sup_{g, x \in T} \langle g, \pi_0 x \rangle + \mathbb{E} \sup_{g, x \in T} \sum_{k=1}^{\infty} \langle g, \Delta_k x \rangle \text{ (from before)} \]

- \( \forall t, \mathbb{P}(Y_k(x) > t^{2k/2} \| \Delta_k x \|_2) \leq e^{-t^{2k}/2} \) (gaussian decay)
- \( \mathbb{P}(\exists x, k Y_k(x) > t^{2k/2} \| \Delta_k x \|_2) \leq \sum_k (2^{2k})^2 e^{-t^{2k}/2} \)

\[ \mathbb{E} \sup_{g, x \in T} \sum_k Y_k(x) \leq \int_0^{\infty} \mathbb{P}(\sup_{x \in T} \sum_k Y_k(x) > u) \, du \]
Gaussian mean width bound 4: generic chaining

\[ \mathbb{E} \sup_{g \in T} \sum_{k} Y_k(x) \leq \int_{0}^{\infty} \mathbb{P}(\sup_{x \in T} \sum_{k} Y_k(x) > u) du \]
Gaussian mean width bound 4: generic chaining

\[ \mathbb{E} \sup_{g \in T} \sum_{k} Y_k(x) \leq \int_{0}^{\infty} \mathbb{P}(\sup_{x \in T} \sum_{k} Y_k(x) > u) du \]

\[ = \gamma_2(T, \ell_2) \cdot \int_{0}^{\infty} \mathbb{P}(\sup_{x \in T} \sum_{k} Y_k(x) > t \sup_{x \in T} \sum_{k} 2^{k/2} \| \Delta_k x \|_2) dt \]

(change of variables: \( u = t \sup_{x \in T} \sum_{k} 2^{k/2} \| \Delta_k x \|_2 \approx t \gamma_2(T, \ell_2) \))
Gaussian mean width bound 4: generic chaining

\[
\mathbb{E} \sup_{g \in T} \sum_k Y_k(x) \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_k Y_k(x) > u) du
\]

\[
= \gamma_2(T, \ell_2) \cdot \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_k Y_k(x) > t \sup_{x \in T} \sum_k 2^{k/2} \|\Delta_k x\|_2) dt
\]

(change of variables: \( u = t \sup_{x \in T} \sum_k 2^{k/2} \|\Delta_k x\|_2 \approx t \gamma_2(T, \ell_2) \))

\[
\leq \gamma_2(T, \ell_2) \cdot \left[ 2 + \int_2^\infty \left( \sum_{k=1}^\infty (2^2)^2 e^{-t^2 2^k/2} \right) dt \right]
\]

\[
= \gamma_2(T, \ell_2) \cdot \left[ 2 + \sum_{k=1}^\infty \left( \int_2^\infty (2^2)^2 e^{-t^2 2^k/2} dt \right) \right] \approx \gamma_2(T, \ell_2)
\]
Gaussian mean width bound 4: generic chaining

\[ \mathbb{E} \sup_{g \in \mathcal{G}} \sum_{k} Y_k(x) \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_{k} Y_k(x) > u) du \]

\[ = \gamma_2(T, \ell_2) \cdot \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_{k} Y_k(x) > t \sup_{x \in T} \sum_{k} 2^{k/2} \|\Delta_k x\|_2) dt \]

(change of variables: \( u = t \sup_{x \in T} \sum_{k} 2^{k/2} \|\Delta_k x\|_2 \approx t \gamma_2(T, \ell_2) \))

\[ \leq \gamma_2(T, \ell_2) \cdot [2 + \int_2^\infty \left( \sum_{k=1}^\infty (2^{2k})^2 e^{-t^{2k}} \right) dt] \]

\[ = \gamma_2(T, \ell_2) \cdot [2 + \sum_{k=1}^\infty \left( \int_2^\infty (2^{2k})^2 e^{-t^{2k}} dt \right)] \approx \gamma_2(T, \ell_2) \]

- Conclusion: \( g(T) \lesssim \gamma_2(T, \ell_2) \)
Gaussian mean width bound 4: generic chaining

\[ \mathbb{E} \sup_{g \in T} \sum_k Y_k(x) \leq \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_k Y_k(x) > u) du \]

\[ = \gamma_2(T, \ell_2) \cdot \int_0^\infty \mathbb{P}(\sup_{x \in T} \sum_k Y_k(x) > t \sup_{x \in T} \sum_k 2^{k/2} \| \Delta_k x \|_2) dt \]

(change of variables: \( u = t \sup_{x \in T} \sum_k 2^{k/2} \| \Delta_k x \|_2 \) \( \simeq t \gamma_2(T, \ell_2) \))

\[ \leq \gamma_2(T, \ell_2) \cdot [2 + \int_2^\infty \left( \sum_{k=1}^\infty (2^{2^k})^2 e^{-t^2 2^k/2} \right) dt] \]

\[ = \gamma_2(T, \ell_2) \cdot [2 + \sum_{k=1}^\infty \left( \int_2^\infty (2^{2^k})^2 e^{-t^2 2^k/2} dt \right)] \simeq \gamma_2(T, \ell_2) \]

- Conclusion: \( g(T) \lesssim \gamma_2(T, \ell_2) \)
- Talagrand: \( g(T) \simeq \gamma_2(T, \ell_2) \) (will show soon)
  ("Majorizing measures theorem")
Are these bounds really different?

- $\gamma_2(T, \ell_2)$: $\inf\{T_k\} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell_2}(x, T_k)$
- Dudley: $\inf\{T_k\} \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_k)$
  \[\simeq \int_0^{\infty} \log^{1/2} \mathcal{N}(T, \ell_2, u) du\]
Are these bounds really different?

- $\gamma_2(T, \ell_2): \inf_{\{T_k\}} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell_2}(x, T_k)$
- **Dudley:** $\inf_{\{T_k\}} \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_k)$
- $\approx \int_0^\infty \log^{1/2} \mathcal{N}(T, \ell_2, u) du$
- Dudley not optimal: $T = B_{\ell_1^n}$
Are these bounds really different?

- $\gamma_2(T, \ell_2): \inf\{T_k\} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell_2}(x, T_k)$
- **Dudley:** $\inf\{T_k\} \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_k)$
  \[ \simeq \int_0^{\infty} \log^{1/2} \mathcal{N}(T, \ell_2, u) du \]

- Dudley not optimal: $T = B_{\ell_1}$
- $\sup_{x \in B_{\ell_1}} \langle g, x \rangle = \|g\|_{\infty}$, so $g(T) \simeq \sqrt{\log n}$

- **Exercise:** Come up with admissible $\{T_k\}$ yielding $\gamma_2 \lesssim \sqrt{\log n}$ (must exist by majorizing measures)
Are these bounds really different?

- $\gamma_2(T, \ell_2): \inf \{ T_k \} \sup_{x \in T} \sum_{k=1}^\infty 2^{k/2} \cdot d_{\ell_2}(x, T_k)$
- **Dudley:** $\inf \{ T_k \} \sum_{k=1}^\infty 2^{k/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_k)$
  \[ \simeq \int_0^\infty \log^{1/2} \mathcal{N}(T, \ell_2, u) \, du \]
- Dudley not optimal: $T = B_{\ell_1}$
- $\sup_{x \in B_{\ell_1}} \langle g, x \rangle = \| g \|_\infty$, so $g(T) \simeq \sqrt{\log n}$

**Exercise:** Come up with admissible $\{ T_k \}$ yielding $\gamma_2 \lesssim \sqrt{\log n}$ (must exist by majorizing measures)

- Dudley: $\log \mathcal{N}(B_{\ell_1}, \ell_2, u) \simeq (1/u^2) \log n$ for $u$ not too small (consider just covering $(1/(Cu)^2)$-sparse vectors with $(Cu)^2$ in each coordinate). So Dudley stuck at $g(B_{\ell_1}) \lesssim \log^{3/2} n$. 
Are these bounds really different?

• $\gamma_2(T, \ell_2): \inf \{ T_k \} \sup_{x \in T} \sum_{k=1}^{\infty} 2^{k/2} \cdot d_{\ell_2}(x, T_k)$

• Dudley: $\inf \{ T_k \} \sum_{k=1}^{\infty} 2^{k/2} \cdot \sup_{x \in T} d_{\ell_2}(x, T_k) \approx \int_0^\infty \log^{1/2} N(T, \ell_2, u) du$

• Dudley not optimal: $T = B_{\ell_1^n}$

• $\sup_{x \in B_{\ell_1^n}} \langle g, x \rangle = \|g\|_\infty$, so $g(T) \approx \sqrt{\log n}$

• Exercise: Come up with admissible $\{ T_k \}$ yielding $\gamma_2 \lesssim \sqrt{\log n}$ (must exist by majorizing measures)

• Dudley: $\log N(B_{\ell_1^n}, \ell_2, u) \approx (1/u^2) \log n$ for $u$ not too small (consider just covering $(1/(Cu)^2)$-sparse vectors with $(Cu)^2$ in each coordinate). So Dudley stuck at $g(B_{\ell_1^n}) \lesssim \log^{3/2} n$.

• Simple vanilla $\varepsilon$-net argument gives $g(B_{\ell_1^n}) \lesssim \text{poly}(n)$. 

High probability

- So far just talked about $g(T) = \mathbb{E}_g \sup_{x \in T} Z_x$
  
  But what if we want to know $\sup_{x \in T} Z_x$ is small whp, not just in expectation?
High probability

• So far just talked about $g(T) = \mathbb{E}_g \sup_{x \in T} Z_x$
  But what if we want to know $\sup_{x \in T} Z_x$ is small whp, not just in expectation?

• Moment method: bound $\mathbb{E}_g \sup_{x \in T} Z_x^p$ for large $p$ and do Markov

Can bound moments using chaining too; see talk by Dirksen

(also see theorem of Borell, Ibragimov, Sudakov, Tsirelson [Theorem 5.8 of “Concentration Inequalities” by Boucheron, Lugosi, Massart])
Lower Bounds
Showed \( g(T) \lesssim \gamma_2(T, \ell_2) \). What about lower bounds on \( g(T) \)?
Showed $g(T) \preceq \gamma_2(T, \ell_2)$. What about lower bounds on $g(T)$?

**Lemma (Slepian)**

Let $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_n \in \mathbb{R}^N$ be such that for all $t, t'$

$$\|z_t - z_{t'}\|_2 \geq \|z'_t - z'_{t'}\|_2$$

Let $X_t = \langle g, z_t \rangle$, $Y_t = \langle g, z'_t \rangle$. Then

$$\forall u_1, \ldots, u_n \in \mathbb{R}, \mathbb{P}(\bigwedge_{t=1}^n (X_t \leq u_t)) \leq \mathbb{P}(\bigwedge_{t=1}^n (Y_t \leq u_t))$$

In particular, $\mathbb{E} \sup_t X_t \geq \mathbb{E} \sup_t Y_t$
Showed $g(T) \lesssim \gamma_2(T, \ell_2)$. What about lower bounds on $g(T)$?

**Lemma (Slepian)**

Let $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_n \in \mathbb{R}^N$ be such that for all $t, t'$

$$\|z_t - z'_t\|_2 \geq \|z'_t - z'_t\|_2$$

Let $X_t = \langle g, z_t \rangle$, $Y_t = \langle g, z'_t \rangle$. Then

$$\forall u_1, \ldots, u_n \in \mathbb{R}, \ P(\bigwedge_{t=1}^n (X_t \leq u_t)) \leq P(\bigwedge_{t=1}^n (Y_t \leq u_t))$$

*In particular, $\mathbb{E} \sup_t X_t \geq \mathbb{E} \sup_t Y_t$*

Won’t prove today. Can find three different proofs in:

- Ledoux-Talagrand book, “Probability in Banach Spaces” (Corollary 3.14 — loses a factor of 2 in the conclusion)
- Ramon van Handel book, “Probability in High Dimensions” (Thm 6.8)
- Mikahil Gromov’s paper, “Monotonicity of the volume of intersection of balls”
Sudakov minoration

Showed $g(T) \lesssim \gamma_2(T, \ell_2)$. What about lower bounds on $g(T)$?

**Lemma (Slepian)**

Let $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_n \in \mathbb{R}^N$ be such that pairwise Euclidean distances are larger for the $z$ than $z'$. Then $\mathbb{E} \sup_t X_t \geq \mathbb{E} \sup_t Y_t$. 
Sudakov minoration

Showed $g(T) \lesssim \gamma_2(T, \ell_2)$. What about lower bounds on $g(T)$?

Lemma (Slepian)

Let $z_1, \ldots, z_n$ and $z'_1, \ldots, z'_n \in \mathbb{R}^N$ be such that pairwise Euclidean distances are larger for the $z$ than $z'$. Then $\mathbb{E} \sup_t X_t \geq \mathbb{E} \sup_t Y_t$.

Theorem (Sudakov minoration)

Suppose $T \subset \mathbb{R}^N$ contains points $z_1, \ldots, z_n$ such that $\|z_i - z_j\|_2 \geq \alpha$ for all $i \neq j$. Then $g(T) \gtrsim \alpha \sqrt{\log n}$.
Sudakov minoration

Showed \( g(T) \lesssim \gamma_2(T, \ell_2) \). What about lower bounds on \( g(T) \)?

**Lemma (Slepian)**

Let \( z_1, \ldots, z_n \) and \( z'_1, \ldots, z'_n \in \mathbb{R}^N \) be such that pairwise Euclidean distances are larger for the \( z \) than \( z' \). Then \( \mathbb{E}\sup_t X_t \geq \mathbb{E}\sup_t Y_t \).

**Theorem (Sudakov minoration)**

Suppose \( T \subset \mathbb{R}^N \) contains points \( z_1, \ldots, z_n \) such that 
\[
\|z_i - z_j\|_2 \geq \alpha \text{ for all } i \neq j.
\]
Then \( g(T) \gtrsim \alpha \sqrt{\log n} \).

**Proof.**

Use Slepian with \( z'_1, \ldots, z'_n = (\alpha/\sqrt{2})e_1, \ldots, (\alpha/\sqrt{2})e_n \). Then
\[
g(T) \geq \frac{\alpha}{\sqrt{2}} \cdot \mathbb{E}\max\{g_1, \ldots, g_n\} \gtrsim \alpha \sqrt{\log n}.
\]
Sudakov minoration

Compare with Dudley:

\[
g(T) \lesssim \sum_{r=1}^{\infty} \frac{1}{2^r} \log \left(1/2 D(T, \ell_2, 1/2^r)\right)
\]

\[
g(T) \gtrsim \sup_{r \geq 1} \frac{1}{2^r} \log \left(1/2 D(T, \ell_2, 1/2^r)\right)
\]

where \(D(T, d, \varepsilon)\) is the packing number (largest number of radius-\(\varepsilon\) balls under \(d\) that can be disjointly packed in \(T\)).

Fact

For every \(\varepsilon > 0\),

\[
D(T, d, 2\varepsilon) \leq N(T, d, \varepsilon) \leq D(T, d, \varepsilon)
\]

Thus, essentially Sudakov is the largest term in a sum and Dudley is the sum of all terms. Can show only \(k = O(\log N)\) matters for dimension \(N\) (the rest of the sum is \(O(1)\)), so \(O(\log N)\) gap (similar gap between Sudakov and \(\gamma_2(T, \ell_2)\)).
Sudakov minoration

Compare with Dudley:

- **Dudley:** \( g(T) \lesssim \sum_{r=1}^{\infty} \frac{1}{2^r} \log^{1/2} \mathcal{N}(T, \ell_2, \frac{1}{2^r}) \)
- **Sudakov:** \( g(T) \gtrsim \sup_{r \geq 1} \frac{1}{2^r} \log^{1/2} \mathcal{D}(T, \ell_2, \frac{1}{2^r}) \)

where \( \mathcal{D}(T, d, \varepsilon) \) is the *packing number* (largest number of radius-\( \varepsilon \) balls under \( d \) that can be disjointly packed in \( T \))
Sudakov minoration

Compare with Dudley:

- **Dudley**: $g(T) \lesssim \sum_{r=1}^{\infty} \frac{1}{2^r} \log^{1/2} \mathcal{N}(T, \ell_2, \frac{1}{2^r})$
- **Sudakov**: $g(T) \gtrsim \sup_{r \geq 1} \frac{1}{2^r} \log^{1/2} \mathcal{D}(T, \ell_2, \frac{1}{2^r})$

where $\mathcal{D}(T, d, \varepsilon)$ is the *packing number* (largest number of radius-$\varepsilon$ balls under $d$ that can be disjointly packed in $T$).

**Fact**

*For every $\varepsilon > 0$, $\mathcal{D}(T, d, 2\varepsilon) \leq \mathcal{N}(T, d, \varepsilon) \leq \mathcal{D}(T, d, \varepsilon)$.*
Sudakov minoration

Compare with Dudley:

- **Dudley:** \( g(T) \lesssim \sum_{r=1}^{\infty} \frac{1}{2^r} \log^{1/2} \mathcal{N}(T, \ell_2, \frac{1}{2^r}) \)
- **Sudakov:** \( g(T) \gtrsim \sup_{r \geq 1} \frac{1}{2^r} \log^{1/2} \mathcal{D}(T, \ell_2, \frac{1}{2^r}) \)

where \( \mathcal{D}(T, d, \epsilon) \) is the packing number (largest number of radius-\( \epsilon \) balls under \( d \) that can be disjointly packed in \( T \))

**Fact**

*For every \( \epsilon > 0 \), \( \mathcal{D}(T, d, 2\epsilon) \leq \mathcal{N}(T, d, \epsilon) \leq \mathcal{D}(T, d, \epsilon) \).

Thus, essentially Sudakov is the largest term in a sum and Dudley is the sum of all terms. Can show only \( k = O(\log N) \) matters for dimension \( N \) (the rest of the sum is \( O(1) \)), so \( O(\log N) \) gap (similar gap between Sudakov and \( \gamma_2(T, \ell_2) \)).
Strengthening Sudakov

Theorem (Sudakov++)

Suppose $T \subset \mathbb{R}^N$ contains points $x_1, \ldots, x_n$ such that 
$\|x_i - x_j\|_2 \geq \alpha$ for all $i \neq j$. Then for some constants $r \geq 4, \kappa > 0$
$g(T) \geq \kappa \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g(B_{\ell_2}(x_i, \alpha/r) \cap T)$. 

Proof. Let $Z_x = \langle g, x \rangle$. Define $Q_i = \sup_{x \in B_{\ell_2}(x_i, \alpha/r) \cap T} Z_x - Z_{x_i}$. 
Then $E \sup_{x \in T} Z_x \geq E \max_{i=1,\ldots,n} \{ Z_x \} + \min_{i=1,\ldots,n} E Q_i - E \max_{i=1,\ldots,n} |Q_i - E Q_i| \geq c \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g(B_{\ell_2}(x_i, \alpha/r) \cap T)$. 

Pick $r > 2c' / c$. (Last line: gaussian concentration of Lipschitz functions.)
Strengthening Sudakov

Theorem (Sudakov++)

Suppose \( T \subset \mathbb{R}^N \) contains points \( x_1, \ldots, x_n \) such that
\[
\|x_i - x_j\|_2 \geq \alpha \text{ for all } i \neq j.
\]
Then for some constants \( r \geq 4, \kappa > 0 \)
\[
g(T) \geq \kappa \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g\left(B_{\ell_2}(x_i, \alpha/r) \cap T\right).
\]

Proof.
Let \( Z_x = \langle g, x \rangle \). Define
\[
Q_i = \sup_{x \in B_{\ell_2}(x_i, \alpha/r) \cap T} Z_x - Z_{x_i}
\]
Strengthening Sudakov

Theorem (Sudakov++)

Suppose \( T \subset \mathbb{R}^N \) contains points \( x_1, \ldots, x_n \) such that \( \|x_i - x_j\|_2 \geq \alpha \) for all \( i \neq j \). Then for some constants \( r \geq 4, \kappa > 0 \)
\[ g(T) \geq \kappa \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g(B_{\ell_2}(x_i, \alpha/r) \cap T). \]

Proof.

Let \( Z_x = \langle g, x \rangle \). Define

\[ Q_i = \sup_{x \in B_{\ell_2}(x_i, \alpha/r) \cap T} Z_x - Z_{x_i} \]

\[ \mathbb{E} \sup_{x \in T} Z_x \geq \mathbb{E} \sup_{i=1,\ldots,n} \left[ (Q_i - \mathbb{E} Q_i) + \mathbb{E} Q_i + Z_{x_i} \right] \]
Strengthening Sudakov

Theorem (Sudakov++)

Suppose $T \subset \mathbb{R}^N$ contains points $x_1, \ldots, x_n$ such that $\|x_i - x_j\|_2 \geq \alpha$ for all $i \neq j$. Then for some constants $r \geq 4, \kappa > 0$ 

$$g(T) \geq \kappa \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g(B_{\ell_2}(x_i, \alpha/r) \cap T).$$

Proof.

Let $Z_x = \langle g, x \rangle$. Define 

$$Q_i = \sup_{x \in B_{\ell_2}(x_i, \alpha/r) \cap T} Z_x - Z_{x_i}$$

$$\mathbb{E} \sup_{x \in T} Z_x \geq \mathbb{E} \sup_{i=1,\ldots,n} [(Q_i - \mathbb{E} Q_i) + \mathbb{E} Q_i + Z_{x_i}]$$

$$\geq \mathbb{E} \max_{i=1,\ldots,n} \{Z_{x_i}\} + \min_{i=1,\ldots,n} \mathbb{E} Q_i - \mathbb{E} \max_{i=1,\ldots,n} |Q_i - \mathbb{E} Q_i|$$
Strengthening Sudakov

Theorem (Sudakov++)

Suppose $T \subset \mathbb{R}^N$ contains points $x_1, \ldots, x_n$ such that $\|x_i - x_j\|_2 \geq \alpha$ for all $i \neq j$. Then for some constants $r \geq 4, \kappa > 0$

$$g(T) \geq \kappa \alpha \sqrt{\log n} + \min_{i=1,\ldots,n} g(B_{\ell_2}(x_i, \alpha/r) \cap T).$$

Proof.

Let $Z_x = \langle g, x \rangle$. Define

$$Q_i = \sup_{x \in B_{\ell_2}(x_i, \alpha/r) \cap T} Z_x - Z_{x_i}.$$

$$\mathbb{E} \sup_{x \in T} Z_x \geq \mathbb{E} \sup_{i=1,\ldots,n} \left[ (Q_i - \mathbb{E} Q_i) + \mathbb{E} Q_i + Z_{x_i} \right]$$

$$\geq \mathbb{E} \max_{i=1,\ldots,n} \{ Z_{x_i} \} + \min_{i=1,\ldots,n} \mathbb{E} Q_i - \mathbb{E} \max_{i=1,\ldots,n} |Q_i - \mathbb{E} Q_i|$$

$$\geq c \alpha \sqrt{\log n} + \min_i g(B(x_i, \alpha/r) \cap T) - c' \frac{\alpha}{r} \sqrt{\log n}$$

Pick $r > 2c'/c$. (Last line: gaussian concentration of Lipschitz functions.)
Function $f : \mathbb{R}^n \to \mathbb{R}$ is $L$-Lipschitz if for all $x, y \in \mathbb{R}^n$,

$$|f(x) - f(y)| \leq L \cdot \|x - y\|_2.$$ 

**Theorem**

*If $f$ is $L$-Lipschitz, then for all $\lambda > 0$*

$$\mathbb{P}(|f(g) - \mathbb{E} f(g)| > \lambda) < 2e^{-\frac{\lambda^2}{2L}}.$$
Gaussian concentration of Lipschitz functions

Function $f : \mathbb{R}^n \to \mathbb{R}$ is $L$-Lipschitz if for all $x, y \in \mathbb{R}^n$,

$$|f(x) - f(y)| \leq L \cdot \|x - y\|_2.$$

**Theorem**

If $f$ is $L$-Lipschitz, then for all $\lambda > 0$

$$\mathbb{P}_g(|f(g) - \mathbb{E} f(g)| > \lambda) < 2e^{-\frac{\lambda^2}{2L}}.$$

- Recall $Q_i = \sup_{x \in B(x_i, \alpha/r) \cap T} Z_x - Z_{x_i} = \sup_{x \in B(x_i, \alpha/r) \cap T} \langle g, x - x_i \rangle = f(g)$
Gaussian concentration of Lipschitz functions

Function \( f : \mathbb{R}^n \to \mathbb{R} \) is \( L \)-Lipschitz if for all \( x, y \in \mathbb{R}^n \),
\[
|f(x) - f(y)| \leq L \cdot \|x - y\|_2.
\]

Theorem
If \( f \) is \( L \)-Lipschitz, then for all \( \lambda > 0 \)
\[
\mathbb{P}_{g}(|f(g) - \mathbb{E} f(g)| > \lambda) < 2e^{-\frac{\lambda^2}{2L}}.
\]

- Recall \( Q_i = \sup_{x \in B(x_i, \alpha/r) \cap T} Z_x - Z_{x_i} = \sup_{x \in B(x_i, \alpha/r) \cap T} \langle g, x - x_i \rangle = f(g) \)
- \( \|x - x_i\|_2 \leq \alpha/r \), so \( f(\alpha/r) \)-Lipschitz by Cauchy-Schwarz
Gaussian concentration of Lipschitz functions

Function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is $L$-Lipschitz if for all $x, y \in \mathbb{R}^n$,

$$|f(x) - f(y)| \leq L \cdot \|x - y\|_2.$$

**Theorem**

If $f$ is $L$-Lipschitz, then for all $\lambda > 0$

$$\mathbb{P}_g(|f(g) - \mathbb{E} f(g)| > \lambda) < 2e^{-\frac{\lambda^2}{2L}}.

- Recall $Q_i = \sup_{x \in B(x_i, \alpha/r) \cap T} Z_x - Z_{x_i} = \sup_{x \in B(x_i, \alpha/r) \cap T} \langle g, x - x_i \rangle = f(g)$
- $\|x - x_i\|_2 \leq \alpha/r$, so $f(\alpha/r)$-Lipschitz by Cauchy-Schwarz
- Hence $\mathbb{E} \max_{i=1,\ldots,n} |Q_i - \mathbb{E} Q_i| \leq c'(\alpha/r) \sqrt{\log n}$. 
Using Sudakov++ to prove majorizing measures

(based on notes by James Lee, itself based on book by Talagrand)

\[ \gamma_2(T, \ell_2) \leq \inf_{\{T_k\}} \sup_{x \in T} \sum_{k=0}^{\infty} 2^{k/2} \cdot \text{diam}(P_k(x)) \]

where \( \text{diam} \) is \( \ell_2 \) diameter, and \( P_k(x) \) is the set of all points in \( T \) whose nearest neighbor in \( T_k \) is the same as \( x \)'s nearest neighbor.
Using Sudakov++ to prove majorizing measures

(based on notes by James Lee, itself based on book by Talagrand)

\[ \gamma_2(T, \ell_2) \leq \inf_{\{T_k\}_{\text{admissible}}} \sup_{x \in T} \sum_{k=0}^{\infty} 2^{k/2} \cdot \text{diam}(P_k(x)) \]

where \( \text{diam} \) is \( \ell_2 \) diameter, and \( P_k(x) \) is the set of all points in \( T \) whose nearest neighbor in \( T_k \) is the same as \( x \)'s nearest neighbor.

- Want to prove \( g(T) \gtrsim \) (right hand side above)
- Proof by picture: repeated application of Sudakov++
Using Sudakov++ to prove majorizing measures

(based on notes by James Lee, itself based on book by Talagrand)

\[
\gamma_2(T, \ell_2) \leq \inf_{\{T_k\} \text{ admissible}} \sup_{x \in T} \sum_{k=0}^{\infty} 2^{k/2} \cdot \text{diam}(P_k(x))
\]

where \( \text{diam} \) is \( \ell_2 \) diameter, and \( P_k(x) \) is the set of all points in \( T \) whose nearest neighbor in \( T_k \) is the same as \( x \)'s nearest neighbor.

- Want to prove \( g(T) \gtrsim \) (right hand side above)
- Proof by picture: repeated application of Sudakov++
- Earlier defined \( \gamma_2 \) by picking nets \( T_k \) of size \( \leq 2^{2^k} \). Right hand side above equivalent to refining partitions of \( T \) recursively in a tree, where each node has \( \leq 2^{2^k} \) children, and root is entire set \( T \). Will henceforth think in terms of building this tree.
Using Sudakov++ to prove majorizing measures

(based on notes by James Lee, itself based on book by Talagrand)

\[
\gamma_2(T, \ell_2) \leq \inf_{\{T_k\}} \sup_{x \in T} \sum_{k=0}^{\infty} 2^{k/2} \cdot \text{diam}(P_k(x))
\]

where \text{diam} is \ell_2 diameter, and \(P_k(x)\) is the set of all points in \(T\) whose nearest neighbor in \(T_k\) is the same as \(x\)'s nearest neighbor.

- Want to prove \(g(T) \gtrsim \) (right hand side above)
- Proof by picture: repeated application of Sudakov++
- Earlier defined \(\gamma_2\) by picking nets \(T_k\) of size \(\leq 2^{2^k}\). Right hand side above equivalent to refining partitions of \(T\) recursively in a tree, where each node has \(\leq 2^{2^k}\) children, and root is entire set \(T\). Will henceforth think in terms of building this tree.
- cost of tree \(T\) is \(\text{cost}(T) = \sup_{x \in T} \sum_{k=0}^{\infty} 2^{k/2} \cdot \text{diam}(P_k(x))\)
  will show \(\forall T \exists \tilde{T}, g(T) \gtrsim \text{cost}(\tilde{T})\)
Building the tree

- Root partition (level 0) is $P_0(\cdot) = T$
- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children

• Greedy: To pick the $i$th child partition, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$).
  - carve out $g(B(x_i, \Delta/r^2) \cap D_i)$ as the $i$th child

  • Might finish process early and have fewer children (call these all “left children”)
  • Might also not exhaust all of $A$ after $2^{2^k} - 1$ children then dump leftovers into last $2^{2^k}$th child (called “right child”)

• Label each node $A$ with an upper bound $\text{rad}(A)$ on radius. Root has $\text{rad}(A) = \text{diam}(T)$
  - If $\text{rad}(A) = \Delta$, left children have $\text{rad} = \Delta/r$, right child has $\Delta$

• Label node $A$ with value $g(A)$; each edge to child with value $\kappa \cdot \text{rad}(A) / r \cdot 2^{2^k} / 2^k$.

• Note if $P$ is root-to-leaf path for $x \in T$, $\sum 2^{2^k} / 2^k \text{diam}(P_k(x)) \lesssim \text{value}(P)$
Building the tree

- Root partition (level 0) is $\mathcal{P}_0(\cdot) = T$
- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children
- **Greedy:** To pick $i$th child partition, $i \geq 0$, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$). carve out $g(B(x_i, \Delta/r) \cap D_i)$ as $i$th child
Building the tree

- Root partition (level 0) is $P_0(\cdot) = T$
- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children
- **Greedy:** To pick $i$th child partition, $i \geq 0$, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$). Carve out $g(B(x_i, \Delta/r) \cap D_i)$ as $i$th child
- Might finish process early and have fewer children (call these all “left children”)
- Might also not exhaust all of $A$ after $2^{2^k} - 1$ children
  then dump leftovers into last $2^{2^k}$ th child (called “right child”)
Building the tree

- Root partition (level 0) is $\mathcal{P}_0(\cdot) = T$
- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children
- **Greedy:** To pick $i$th child partition, $i \geq 0$, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$). carve out $g(B(x_i, \Delta/r) \cap D_i)$ as $i$th child
- Might finish process early and have fewer children (call these all “left children”)

  Might also not exhaust all of $A$ after $2^{2^k} - 1$ children

  then dump leftovers into last $2^{2^k}$ th child (called “right child”)

- Label each node $A$ with an upper bound $\text{rad}(A)$ on radius. Root has $\text{rad}(A) = \text{diam}(T)$

  If $\text{rad}(A) = \Delta$, left children have $\text{rad} = \Delta/r$, right child has $\Delta$
Building the tree

- Root partition (level 0) is $\mathcal{P}_0(\cdot) = T$
- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children
- **Greedy:** To pick $i$th child partition, $i \geq 0$, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$). carve out $g(B(x_i, \Delta/r) \cap D_i)$ as $i$th child
- Might finish process early and have fewer children
  (call these all “left children”)
  Might also not exhaust all of $A$ after $2^{2^k} - 1$ children
  then dump leftovers into last $2^{2^k}$ th child (called “right child”)
- Label each node $A$ with an upper bound $\text{rad}(A)$ on radius. Root has $\text{rad}(A) = \text{diam}(T)$
  If $\text{rad}(A) = \Delta$, left children have $\text{rad} = \Delta/r$, right child has $\Delta$
- Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$. 
Building the tree

- Root partition (level 0) is $\mathcal{P}_0(\cdot) = T$

- For partition $A$ at level $k$ and radius $\leq \Delta$, will refine into $\leq 2^{2^k}$ children

- **Greedy:** To pick $i$th child partition, $i \geq 0$, let $x_i \in D_i$ maximize $g(B(x_i, \Delta/r^2) \cap D_i)$ ($D_i$ are the points in $A$ not carved out yet; $D_0 = A$). carve out $g(B(x_i, \Delta/r) \cap D_i)$ as $i$th child

- Might finish process early and have fewer children
  (call these all “left children”)

  Might also not exhaust all of $A$ after $2^{2^k} - 1$ children
  then dump leftovers into last $2^{2^k}$ th child (called “right child”)

- Label each node $A$ with an upper bound $\text{rad}(A)$ on radius. Root has $\text{rad}(A) = \text{diam}(T)$
  If $\text{rad}(A) = \Delta$, left children have $\text{rad} = \Delta/r$, right child has $\Delta$

- Label node $A$ with value $g(A)$; each edge to child with value $\kappa \cdot \text{rad}(A) \cdot 2^k/r^2$.

- Note if $P$ is root-to-leaf path for $x \in T$, $\sum_k 2^{k/2} \text{diam}(\mathcal{P}_k(x)) \lesssim \text{value}(P)$
Lower bound from greedy tree

- Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$.
- Note if $P$ is root-to-leaf path for $x \in T$, $\sum_k 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P)$.

Observation 1: R followed by sequence of L’s; R dominates sum of values ($\sum_j \geq 0 2^{(k+j)/2} \Delta r \geq 0$ geometrically decays since $r \geq 4$).

Observation 2: sequence of R’s; last R dominates the sum (rad stays the same but $2^{k/2}$ is geometrically increasing).

Thus for any $P$, its value is $\approx$ same only considering last right turns.
Lower bound from greedy tree

- Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$.
- Note if $P$ is root-to-leaf path for $x \in T$, $\sum_k 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P)$
- suffices to show for all root-to-leaf paths $P$, $g(T) \gtrsim \text{value}(P)$
Lower bound from greedy tree

- Label node \( A \) with value \( g(A) \); each edge to child with value \( \kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2} \).
- Note if \( P \) is root-to-leaf path for \( x \in T \), \( \sum_k 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P) \).
- Suffices to show for all root-to-leaf paths \( P \), \( g(T) \gtrsim \text{value}(P) \).
- View path as sequence of L’s and R’s (Left or Right child).
Lower bound from greedy tree

- Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$.
- Note if $P$ is root-to-leaf path for $x \in T$, $\sum_k 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P)$
- suffices to show for all root-to-leaf paths $P$, $g(T) \gtrsim \text{value}(P)$
- View path as sequence of L’s and R’s (Left or Right child)
  
  **Observation 1:** R followed by sequence of L’s; R dominates sum of values
  \[ (\sum_{j \geq 0} 2^{(k+j)/2} \frac{\Delta}{r^j} \text{ geometrically decays since } r \geq 4) \]
Lower bound from greedy tree

• Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$.

• Note if $P$ is root-to-leaf path for $x \in T$, $\sum k \cdot 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P)$

• suffices to show for all root-to-leaf paths $P$, $g(T) \gtrsim \text{value}(P)$

• View path as sequence of L’s and R’s (Left or Right child)

  Observation 1: R followed by sequence of L’s; R dominates sum of values
  $(\sum j \geq 0 2^{(k+j)/2} \frac{\Delta}{r^j}$ geometrically decays since $r \geq 4$)

  Observation 2: sequence of R’s; last R dominates the sum
  ($\text{rad}$ stays the same but $2^{k/2}$ is geometrically increasing)
Label node $A$ with value $g(A)$; each edge to child with value $\kappa \frac{\text{rad}(A)}{r} \cdot 2^{k/2}$.

Note if $P$ is root-to-leaf path for $x \in T$, $\sum_k 2^{k/2} \text{diam}(P_k(x)) \lesssim \text{value}(P)$ suffices to show for all root-to-leaf paths $P$, $g(T) \gtrsim \text{value}(P)$.

View path as sequence of L’s and R’s (Left or Right child).

**Observation 1:** R followed by sequence of L’s; R dominates sum of values

$(\sum_{j \geq 0} 2^{(k+j)/2} \frac{\Delta}{r^j}$ geometrically decays since $r \geq 4)$

**Observation 2:** sequence of R’s; last R dominates the sum

$(\text{rad stays the same but } 2^{k/2} \text{ is geometrically increasing})$

Thus for any $P$, its value is $\approx$ same only considering last right turns.
Desired inequality

- Inequality is over values in red, so want $g(T) \geq value(P)$
  (only counting last right edges)

*Figures and presentation adapted from notes of James Lee*
Proof of tree lower bound

Lemma 1:

Let $T$, $A$, and $B$ be trees such that $T \supseteq A \supseteq B$. Then \[ g(T) \geq \frac{1}{2} (g(A) + g(B)) \]
Proof of tree lower bound

Lemma 1:

- $B \subseteq A \subseteq T$, so $g(T) \geq g(A) \geq g(B)$

  $\Rightarrow g(T) \geq \frac{1}{2}(g(A) + g(B))$
Proof of tree lower bound

Lemma 2:

\[ g(T) \geq \frac{1}{2} \] follows by applying Lemma 1 once then iterating Lemma 2 down the path.
Lemma 2:

\[ g(T) \geq \frac{1}{2} \text{(sum of last right turns)} \]

follows by applying Lemma 1 once then iterating Lemma 2 down the path.
Proof of tree lower bound

Lemma 2:

- $g(B) \geq g(C)$ since $C \subseteq B$. Now must show $value(A) \geq value(e) + value(D) = \kappa \frac{\Delta}{r} 2^{k/2} + g(D)$. 
Proof of tree lower bound

Lemma 2:

- $g(B) \geq g(C)$ since $C \subseteq B$. Now must show $value(A) \geq value(e) + value(D) = \kappa \frac{\Delta}{r} 2^{k/2} + g(D)$.

- $A$ refined into $m = 2^{2^k}$ pieces. Since carved out balls of radius $\Delta/r$, ball centers $x_i$ are $\frac{\Delta}{r}$-far apart and can apply Sudakov++.

$$g(A) \geq \kappa \frac{\Delta}{r} \sqrt{\log m} + \min_i g(B(x_i, \Delta/r^2) \cap A) = \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap A)$$
Proof of tree lower bound

Lemma 2:

\[ g(A) \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap A) \]
Proof of tree lower bound

Lemma 2:

\[ g(A) \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap A) \]

\[ \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap D_i) (D_i \subseteq A) \]
Proof of tree lower bound

Lemma 2:

\[ g(A) \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap A) \]

\[ \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap D_i) \quad (D_i \subseteq A) \]

\[ = \kappa \frac{\Delta}{r} 2^{k/2} + g(B(x_m, \Delta/r^2) \cap D_m) \quad (\text{by greedy construction}) \]
Proof of tree lower bound

Lemma 2:

\[ g(A) \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap A) \]

\[ \geq \kappa \frac{\Delta}{r} 2^{k/2} + \min_i g(B(x_i, \Delta/r^2) \cap D_i) \quad (D_i \subseteq A) \]

\[ = \kappa \frac{\Delta}{r} 2^{k/2} + g(B(x_m, \Delta/r^2) \cap D_m) \quad \text{(by greedy construction)} \]

\[ = \kappa \frac{\Delta}{r} 2^{k/2} + g(D) \quad (D \subseteq D_m \text{ of radius } \leq \frac{\Delta}{r^2}, \text{ and } x_m \text{ is maximizer}) \]
Issues at the boundary

What if the first turn in $P$ is a left turn?

\[ T \geq c \cdot \text{diam}(T) \]
Issues at the boundary

What if the first turn in $P$ is a left turn?

- $g(T) = \frac{1}{2}(g(T) + g(T))$
Issues at the boundary

What if the first turn in $P$ is a left turn?

- $g(T) = \frac{1}{2}(g(T) + g(T))$
  $\geq \frac{1}{2}(g(T) + g(A))$ ($A \subseteq T$)
Issues at the boundary

What if the first turn in $P$ is a left turn?

- $g(T) = \frac{1}{2}(g(T) + g(T))$
- $\geq \frac{1}{2}(g(T) + g(A)) \quad (A \subseteq T)$
- $\succeq \frac{1}{2}(\text{diam}(T) + g(A)) \quad (g(T) \succeq \text{diam}(T) \text{ always})$
Issues at the boundary

What if the last turn in $P$ is a right turn?

\[ A \geq B \geq c \cdot e \cdot e' \geq g(A) \geq g(B) \geq g(C) \quad (C \subseteq B) \]

then $g(C) \geq \text{value}(e')$ (vanilla Sudakov again)
What if the last turn in $P$ is a right turn?

- $g(A) \geq value(e)$ (vanilla Sudakov)
Issues at the boundary

What if the last turn in $P$ is a right turn?

\[ A \geq B \geq \text{c} \cdot e \cdot e' \]

- $g(A) \gtrsim \text{value}(e)$ (vanilla Sudakov)
- $g(B) \geq g(C)$ ($C \subseteq B$)
  then $g(C) \gtrsim \text{value}(e')$ (vanilla Sudakov again)
The End
Some items to read

- “Notes on Gaussian processes and majorizing measures”, by James Lee (see his website and tcsmath.org blog)
- “Probability in High Dimensions”, by Ramon van Handel
- “Upper and lower bounds for stochastic processes: modern methods and classical problems”, by Michel Talagrand