Define the following processes:

$$Z(x) = \sup_{f \in \mathcal{F}} \left(\mathbb{E}f - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right)$$

and

$$R(x) = \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i).$$

Assume $a \le f(x) \le b$ for all f, x. In the last lecture we proved Z is concentrated around its expectation: with probability at least $1 - e^{-t}$,

$$Z \ < \mathbb{E}Z + (b-a)\sqrt{\frac{2t}{n}}.$$

Furthermore,

$$\begin{aligned}
\mathbb{E}Z(x) &= \mathbb{E} \sup_{f \in \mathcal{F}} \left(\mathbb{E}f - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right) \\
&= \mathbb{E} \sup_{f \in \mathcal{F}} \left(\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^{n} f(x_i') \right] - \frac{1}{n} \sum_{i=1}^{n} f(x_i) \right) \\
&\le \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} (f(x_i') - f(x_i)) \\
&= \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i (f(x_i') - f(x_i)) \\
&\le \mathbb{E} \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i') + \sup_{f \in \mathcal{F}} \left(-\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i) \right) \\
&\le 2\mathbb{E}R(x).
\end{aligned}$$

Hence, with probability at least $1 - e^{-t}$,

$$Z \ < 2\mathbb{E}R + (b-a)\sqrt{\frac{2t}{n}}.$$

It can be shown that R is also concentrated around its expectation: if $-M \le f(x) \le M$ for all f, x, then with probability at least $1 - e^{-t}$,

$$\mathbb{E}R \le R + M\sqrt{\frac{2t}{n}}.$$

Hence, with high probability,

$$Z(x) \leq 2R(x) + 4M\sqrt{\frac{2t}{n}}.$$

Theorem 23.1. *If $-1 \leq f \leq 1$, then*

$$\mathbb{P}\left(Z(x) \leq 2\mathbb{E}R(x) + 2\sqrt{\frac{2t}{n}}\right) \geq 1 - e^{-t}.$$

If $0 \leq f \leq 1$, then

$$\mathbb{P}\left(Z(x) \leq 2\mathbb{E}R(x) + \sqrt{\frac{2t}{n}}\right) \geq 1 - e^{-t}.$$

Consider $\mathbb{E}_\varepsilon R(x) = \mathbb{E}_\varepsilon \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \varepsilon_i f(x_i)$. Since x_i are fixed, $f(x_i)$ are just vectors. Let $F \subseteq \mathbb{R}^n$, $f \in F$, where $f = (f_1, \ldots, f_n)$.

Define *contraction* $\varphi_i : \mathbb{R} \mapsto \mathbb{R}$ for $i = 1, \ldots, n$ such that $\varphi_i(0) = 0$ and $|\varphi_i(s) - \varphi_i(t)| \leq |s - t|$.

Let $G : \mathbb{R} \mapsto \mathbb{R}$ be convex and non-decreasing.

The following theorem is called *Comparison inequality for Rademacher process.*

Theorem 23.2.

$$\mathbb{E}_\varepsilon G\left(\sup_{f\in F}\sum \varepsilon_i\varphi_i(f_i)\right) \leq \mathbb{E}_\varepsilon G\left(\sup_{f\in F}\sum \varepsilon_i f_i\right).$$

Proof. It is enough to show that for $T \subseteq \mathbb{R}^2$, $t = (t_1, t_2) \in T$

$$\mathbb{E}_\varepsilon G\left(\sup_{t\in T} t_1 + \varepsilon\varphi(t_2)\right) \leq \mathbb{E}_\varepsilon G\left(\sup_{t\in T} t_1 + \varepsilon t_2\right),$$

i.e. enough to show that we can erase contraction for 1 coordinate while fixing all others.

Since $\mathbb{P}(\varepsilon = \pm 1) = 1/2$, we need to prove

$$\frac{1}{2}G\left(\sup_{t\in T} t_1 + \varphi(t_2)\right) + \frac{1}{2}G\left(\sup_{t\in T} t_1 - \varphi(t_2)\right) \leq \frac{1}{2}G\left(\sup_{t\in T} t_1 + t_2\right) + \frac{1}{2}G\left(\sup_{t\in T} t_1 - t_2\right).$$

Assume $\sup_{t\in T} t_1 + \varphi(t_2)$ is attained on (t_1, t_2) and $\sup_{t\in T} t_1 - \varphi(t_2)$ is attained on (s_1, s_2). Then

$$t_1 + \varphi(t_2) \geq s_1 + \varphi(s_2)$$

and

$$s_1 - \varphi(s_2) \geq t_1 - \varphi(t_2).$$

Again, we want to show

$$\Sigma = G(t_1 + \varphi(t_2)) + G(s_1 - \varphi(s_2)) \le G(t_1 + t_2) + G(t_1 - t_2).$$

Case 1: $t_2 \le 0, s_2 \ge 0$

Since φ is a contraction, $\varphi(t_2) \le |t_2| \le -t_2$, $-\varphi(s_2) \le s_2$.

$$\begin{aligned}\Sigma = G(t_1 + \varphi(t_2)) + G(s_1 - \varphi(s_2)) &\le G(t_1 - t_2) + G(s_1 + s_2) \\ &\le G\left(\sup_{t\in T} t_1 - t_2\right) + G\left(\sup_{t\in T} t_1 + t_2\right).\end{aligned}$$

Case 2: $t_2 \ge 0, s_2 \le 0$

Then $\varphi(t_2) \le t_2$ and $-\varphi(s_2) \le -s_2$. Hence

$$\Sigma \le G(t_1 + t_2) + G(s_1 - s_2) \le G\left(\sup_{t\in T} t_1 + t_2\right) + G\left(\sup_{t\in T} t_1 - t_2\right).$$

Case 3: $t_2 \ge 0, s_2 \ge 0$

Case 3a: $s_2 \le t_2$

It is enough to prove

$$G(t_1 + \varphi(t_2)) + G(s_1 - \varphi(s_2)) \le G(t_1 + t_2) + G(s_1 - s_2).$$

Note that $s_2 - \varphi(s_2) \ge 0$ since $s_2 \ge 0$ and φ – contraction. Since $|\varphi(s)| \le |s|$,

$$s_1 - s_2 \le s_1 + \varphi(s_2) \le t_1 + \varphi(t_2),$$

where we use the fact that t_1, t_2 attain maximum.

Furthermore,

$$G\Big(\underbrace{(s_1 - s_2)}_{u} + \underbrace{(s_2 - \varphi(s_2))}_{x}\Big) - G\Big(s_1 - s_2\Big) \le G\Big((t_1 + \varphi(t_2)) + (s_2 - \varphi(s_2))\Big) - G\Big(t_1 + \varphi(t_2)\Big)$$

Indeed, $\Psi(u) = G(u+x) - G(u)$ is non-decreasing for $x \ge 0$ since $\Psi'(u) = G'(u+x) - G'(u) > 0$ by convexity of G.

Now,

$$(t_1 + \varphi(t_2)) + (s_2 - \varphi(s_2)) \le t_1 + t_2$$

since

$$\varphi(t_2) - \varphi(s_2) \le |t_2 - s_2| = t_2 - s_2.$$

Hence,

$$G\Big(s_1 - \varphi(s_2)\Big) - G\Big(s_1 - s_2\Big) = G\Big((s_1 - s_2) + (s_2 - \varphi(s_2))\Big) - G\Big(s_1 - s_2\Big)$$
$$\leq G\Big(t_1 + t_2\Big) - G\Big(t_1 + \varphi(t_2)\Big).$$

Case 3a: $t_2 \leq s_2$

$$\Sigma \leq G(s_1 + s_2) + G(t_1 - t_2)$$

Again, it's enough to show

$$G(t_1 + \varphi(t_2)) - G(t_1 - t_2) \leq G(s_1 + s_2) - G(s_1 - \varphi(s_2))$$

We have

$$t_1 - t_2 \leq t_1 - \varphi(t_2) \leq s_1 - \varphi(s_2)$$

since s_1, s_2 achieves maximum and since $t_2 + \varphi(t_2) \geq 0$ (φ is a contraction and $t_2 \geq 0$).
Hence,

$$G\Big(\underbrace{(t_1 - t_2)}_{u} + \underbrace{(t_2 + \varphi(t_2))}_{x}\Big) - G\Big(t_1 - t_2\Big) \leq G\Big((s_1 - \varphi(s_2)) + (t_2 + \varphi(t_2))\Big) - G\Big(s_1 - \varphi(s_2)\Big)$$

Since

$$\varphi(t_2) - \varphi(s_2) \leq |t_2 - s_2| = s_2 - t_2,$$

we get

$$\varphi(t_2) - \varphi(s_2) \leq s_2 - t_2.$$

Therefore,

$$s_1 - \varphi(s_2) + (t_2 + \varphi(t_2) \leq s_1 + s_2$$

and so

$$G(t_1 + \varphi(t_2)) - G(t_1 - t_2) \leq G(s_1 + s_2) - G(s_1 - \varphi(s_2))$$

Case 4: $t_2 \leq 0, s_2 \leq 0$

Proved in the same way as Case 3.

□

We now apply the theorem with $G(s) = (s)^+$.

Lemma 23.1.

$$\mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i\varphi_i(t_i)\right| \le 2\mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i t_i\right|$$

Proof. Note that

$$|x| = (x)^+ + (x)^- = (x)^+ + (-x)^+.$$

We apply the Contraction Inequality for Rademacher processes with $G(s) = (s)^+$.

$$\begin{aligned}
\mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i\varphi_i(t_i)\right| &= \mathbb{E}\sup_{t\in T}\left(\left(\sum_{i=1}^{n}\varepsilon_i\varphi_i(t_i)\right)^+ + \left(\sum_{i=1}^{n}(-\varepsilon_i)\varphi_i(t_i)\right)^+\right) \\
&\le 2\mathbb{E}\sup_{t\in T}\left(\sum_{i=1}^{n}\varepsilon_i\varphi_i(t_i)\right)^+ \\
&\le 2\mathbb{E}\sup_{t\in T}\left(\sum_{i=1}^{n}\varepsilon_i t_i\right)^+ \le 2\mathbb{E}\sup_{t\in T}\left|\sum_{i=1}^{n}\varepsilon_i t_i\right|.
\end{aligned}$$

□