In this lecture, we show that although the VC-hull classes might be considerably larger than the VC-classes, they are small enough to have finite uniform entropy integral.

**Theorem 18.1.** *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measurable space, $\mathcal{F} \subset \{f | f : \mathcal{X} \to \mathbb{R}\}$ be a class of measurable functions with measurable square integrable envelope $F$ (i.e., $\forall x \in \mathcal{X}, \forall f \in \mathcal{F}, |f(x)| < F(x)$, and $\|F\|_2 = (\int F^2 d\mu)^{1/2} < \infty$), and the $\epsilon$-net of $\mathcal{F}$ satisfies $N(\mathcal{F}, \epsilon\|F\|_2, \|\cdot\|) \le C\left(\frac{1}{\epsilon}\right)^V$ for $0 < \epsilon < 1$. Then there exists a constant $K$ that depends only on $C$ and $V$ such that $\log N(conv\mathcal{F}, \epsilon\|F\|_2, \|\cdot\|) \le K\left(\frac{1}{\epsilon}\right)^{\frac{2\cdot V}{V+2}}$.*

*Proof.* Let $N(\mathcal{F}, \epsilon\|F\|_2, \|\cdot\|_2) \le C\left(\frac{1}{\epsilon}\right)^V \triangleq n$. Then $\epsilon = C^{1/v} n^{-1/V}$, and $\epsilon\|F\|_2 = C^{1/V}\|F\|_2 \cdot n^{-1/V}$. Let $L = C^{1/V}\|F\|_2$. Then $N(\mathcal{F}, Ln^{-1/V}, \|\cdot\|_2) \le n$ (i.e., the $L \cdot n^{-1/V}$-net of $\mathcal{F}$ contains at most $n$ elements). Construct $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_n \subset \cdots$ such that each $\mathcal{F}_n$ is a $L \cdot n^{-1/V}$-net, and contains at most $n$ elements. Let $W = \frac{1}{2} + \frac{1}{V}$. We proceed to show that there exists constants $C_k$ and $D_k$ that depend only on $C$ and $V$ and are upper bounded ($\sup_k C_k \vee D_k < \infty$), such that

$$\log N(\text{conv}\mathcal{F}_{n\cdot k^q}, C_k L \cdot n^{-W}, \|\cdot\|_2) \quad \le \quad D_k \cdot n \tag{18.1}$$

for $n, k \ge 1$, and $q \ge 3 + V$. This implies the theorem, since if we let $k \to \infty$, we have $\log N(\text{conv}\mathcal{F}, C_\infty L \cdot n^{-W}, \|\cdot\|_2) \le D_\infty \cdot n$. Let $\epsilon = C_\infty C^{1/V} n^{-W}$, and $K = D_\infty C_\infty^{\frac{2\cdot V}{V+2}} C^{\frac{2}{V+2}}$, we get $C_\infty L \cdot n^{-W} = C_\infty C^{1/V}\|F\|_2 n^{-W} = \epsilon\|F\|_2$, $n = \left(\frac{C_\infty C^{1/V}}{\epsilon}\right)^{1/W}$ and $\log N(\text{conv}\mathcal{F}, \epsilon\|F\|_2, \|\cdot\|_2) \le K \cdot \left(\frac{1}{\epsilon}\right)^{\frac{2\cdot V}{V+2}}$. Inequality 18.1 will proved in two steps: (1)

$$\log N(\text{conv}\mathcal{F}_n, C_1 L \cdot n^{-W}, \|\cdot\|_2) \quad \le \quad D_1 \cdot n \tag{18.2}$$

by induction on $n$, using Kolmogorov's chaining technique, and (2) for fixed $n$,

$$\log N(\text{conv}\mathcal{F}_{n\cdot k^q}, C_k L \cdot n^{-W}, \|\cdot\|_2) \quad \le \quad D_k \cdot n \tag{18.3}$$

by induction on $k$, using the results of (1) and Kolmogorov's chaining technique.

For any fixed $n_0$ and any $n \le n_0$, we can choose large enough $C_1$ such that $C_1 L n_0^{-W} \ge \|F\|_2$. Thus $N(\text{conv}\mathcal{F}_n, C_1 L \cdot n^{-W}, \|\cdot\|_2) = 1$ and 18.2 holds trivially. For general $n$, fix $m = n/d$ for large enough $d > 1$. For any $f \in \mathcal{F}_n$, there exists a projection $\pi_m f \in \mathcal{F}_m$ such that $\|f - \pi_m f\| \le C^{\frac{1}{V}} m^{-\frac{1}{V}} \|F\| = Lm^{-\frac{1}{V}}$ by definition of $\mathcal{F}_m$. Since $\sum_{f\in\mathcal{F}_n} \lambda_f \cdot f = \sum_{f\in\mathcal{F}_m} \mu_f \cdot f + \sum_{f\in\mathcal{F}_n} \lambda_f \cdot (f - \pi_m f)$, we have $\text{conv}\mathcal{F}_n \subset \text{conv}\mathcal{F}_m + \text{conv}\mathcal{G}_n$, and the number of elements $|\mathcal{G}_n| \le |\mathcal{F}_n| \le n$, where $\mathcal{G}_n = \{f - \pi_m f : f \in \mathcal{F}_n\}$. We will find $\frac{1}{2}C_1 L n^{-\frac{1}{W}}$-nets for both $\mathcal{F}_m$ and $\mathcal{G}_n$, and bound the number of elements for them to finish to induction step. We need the following lemma to bound the number of elements for the $\frac{1}{2}C_1 L n^{-\frac{1}{W}}$-net of $\mathcal{G}_n$.

**Lemma 18.2.** *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measurable space and $\mathcal{F}$ be an arbitrary set of $n$ measurable functions $f : \mathcal{X} \to \mathbb{R}$ of finite $L_2(\mu)$- diameter $diam\mathcal{F}$ ($\forall f, g \in \mathcal{F}, \int (f-g)^2 d\mu < \infty$). Then $\forall \epsilon > 0$, $N(conv\mathcal{F}, \epsilon diam\mathcal{F}, \|\cdot\|_2) \le \left(e + en\epsilon^2\right)^{2/\epsilon^2}$.*

*Proof.* Let $\mathcal{F} = \{f_1, \cdots, f_n\}$. $\forall \sum_{i=1}^n \lambda_i f_i$, let $Y_1, \cdots, Y_k$ be i.i.d. random variables such that $P(Y_i = f_j) = \lambda_j$ for all $j = 1, \cdots, n$. It follows that $\mathbb{E}Y_i = \sum \lambda_j f_j$ for all $i = 1, \cdots, k$, and

$$\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^k Y_i - \sum_{j=1}^n \lambda_j f_j\right) \le \frac{1}{k}\mathbb{E}\left(Y_1 - \sum_{j=1}^n \lambda_j f_j\right) \le \frac{1}{k}\left(\text{diam}\mathcal{F}\right)^2.$$

Thus at least one realization of $\frac{1}{k}\sum_{i=1}^k Y_i$ has a distance at most $k^{-1/2}\text{diam}\mathcal{F}$ to $\sum \lambda_i f_i$. Since all realizations of $\frac{1}{k}\sum_{i=1}^k Y_i$ has the form $\frac{1}{k}\sum_{i=1}^k f_{j_k}$, there are at most $\binom{n+k-1}{k}$ of such forms. Thus

$$\begin{aligned} N(k^{-1/2}\text{diam}\mathcal{F}, \text{conv}\mathcal{F}, \|\cdot\|_2) &\le \binom{n+k-1}{k} \\ \le \frac{(k+n)^{k+n}}{k^k n^n} &= \left(\frac{k+n}{k}\right)^k \left(\frac{k+n}{n}\right)^n \\ \le e^k \left(\frac{k+n}{k}\right)^k &= (e + en\epsilon^2)^{2/\epsilon^2} \end{aligned}$$

□

By triangle inequality and definition of $\mathcal{G}_n$, $\text{diam}\mathcal{G}_n = \sup_{g_1, g_2 \in \mathcal{G}_n} \|g_1 - g_2\|_2 \le 2 \cdot Lm^{-1/V}$. Let $\epsilon \cdot \text{diam}\mathcal{G}_n = \epsilon \cdot 2Lm^{-1/V} = \frac{1}{2}C_1 Ln^{-W}$. It follows that $\epsilon = \frac{1}{4}C_1 m^{1/V} \cdot n^{-W}$, and

$$\begin{aligned} N(\text{conv}\mathcal{G}_n, \epsilon\text{diam}\mathcal{G}_n, \|\cdot\|_2) &\le \left(e + en \cdot \frac{1}{16}C_1^2 m^{2/V} \cdot n^{-2W}\right)^{32 \cdot C_1^{-2} m^{2/V} n^{2 \cdot W}} \\ &= \left(e + \frac{e}{16}C_1^2 d^{-2/V}\right)^{32 \cdot C_1^{-2} d^{2/V} n} \end{aligned}$$

By definition of $\mathcal{F}_m$ and and induction assumption, $\log N(\text{conv}\mathcal{F}_m, C_1 L \cdot m^{-W}, \|\cdot\|_2) \le D_1 \cdot m$. In other words, the $C_1 L \cdot m^{-W}$-net of $\text{conv}\mathcal{F}_m$ contains at most $e^{D_1 m}$ elements. This defines a partition of $\text{conv}\mathcal{F}_m$ into at most $e^{D_1 m}$ elements. Each element is isometric to a subset of a ball of radius $C_1 Lm^{-W}$. Thus each set can be partitioned into $\left(\frac{3C_1 Lm^{-W}}{\frac{1}{2}C_1 Ln^{-W}}\right)^m = \left(6d^W\right)^{n/d}$ sets of diameter at most $\frac{1}{2}C_1 Ln^{-W}$ according to the following lemma.

**Lemma 18.3.** *The packing number of a ball of radius $R$ in $\mathbb{R}^d$ satisfies $D(B(0,r), \epsilon, \|\cdot\|) \le \left(\frac{3R}{\epsilon}\right)^d$ for the usual norm, where $0 < \epsilon \le R$.*

As a result, the $C_1 Ln^{-W}$-net of $\text{conv}\mathcal{F}_n$ has at most $e^{D_1 n/d}\left(6d^W\right)^{n/d}\left(e + eC_1^2 d^{-2/V}\right)^{8d^{2/V} C_1^{-2} n}$ elements. This can be upper-bounded by $e^n$ by choosing $C_1$ and $d$ depending only on $V$, and $D_1 = 1$.

For $k > 1$, construct $\mathcal{G}_{n,k}$ such that $\text{conv}\mathcal{F}_{nk^q} \subset \text{conv}\mathcal{F}_{n(k-1)^q} + \text{conv}\mathcal{G}_{n,k}$ in a similar way as before. $\mathcal{G}_{n,k}$ contains at most $nk^q$ elements, and each has a norm smaller than $L\left(n\left(k-1\right)^q\right)^{-1/V}$. To bound the cardinality of a $Lk^{-2}n^{-W}$-net, we set $\epsilon \cdot 2L\left(n\left(k-1\right)^q\right)^{-1/V} = Lk^{-2}n^{-W}$, get $\epsilon = \frac{1}{2}n^{-1/2}\left(k-1\right)^{q/V}k^{-2}$, and

$$\begin{aligned} N(\mathrm{conv}\mathcal{G}_{n,k}, \epsilon \mathrm{diam}\mathcal{G}_{n,k}, \|\cdot\|_2) &\leq \left(e + enk^q\epsilon^2\right)^{2/\epsilon^2} \Rightarrow \\ N(\mathrm{conv}\mathcal{G}_{n,k}, \epsilon \mathrm{diam}\mathcal{G}_{n,k}, \|\cdot\|_2) &\leq \left(e + \frac{e}{4}k^{-4+q+2q/V}\right)^{8\cdot n\cdot k^4(k-1)^{-2q/V}} \end{aligned}$$

. As a result, we get

$$\begin{aligned} C_k &= C_{k-1} + \frac{1}{k^2} \\ D_k &= D_{k-1} + 8k^4(k-1)^{-2q/V}\log(e + \frac{e}{4}k^{-4+q+2q/V}). \end{aligned}$$

For $2q/V - 4 \geq 2$, the resulting sequences $C_k$ and $D_k$ are bounded. □