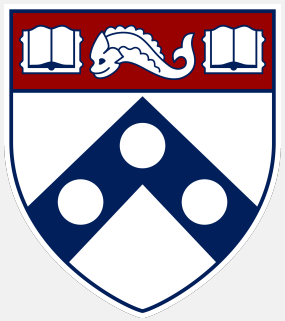
# High-Arity Rademacher Complexity

## joint work with Henry Towsner

Aaron Anderson

Rademacher Instructor, UPenn

UIC, October 29, 2026



# Function Classes from Formulas

We will work with a class $\mathcal{H}$ of functions $X \to [0,1]$.

# Function Classes from Formulas

We will work with a class $\mathcal{H}$ of functions $X \to [0,1]$.

## Example

- Fix a structure $\mathcal{M}$ and a formula $\phi(x;y)$ ($x, y$ can be tuples).
- Given $b \in M^y$, let $h_b : M^x \to \{0,1\}$ be the characteristic function for $\phi(M;b)$.
- Let $X = M^x$, $\mathcal{H}_\phi = \{h_b : b \in M^y\}$.

# More Function Classes from Continuous Logic

In continuous logic, a formula $\phi(x)$ is interpreted over a *metric structure* $\mathcal{M}$ as a function $M^x \to [0, 1]$.

# More Function Classes from Continuous Logic

In continuous logic, a formula $\phi(x)$ is interpreted over a *metric structure* $\mathcal{M}$ as a function $M^x \to [0,1]$.

**Example**

In a metric structure, we can let $\mathcal{H}_\phi = \{\phi^{\mathcal{M}}(x;b) : b \in M^y\}$, no extra steps needed.

# Example: Randomization

- Let $\Omega$ be a probability space, and let $\mathcal{M}$ be an $\mathcal{L}$-structure.
- Let $M^R$ be the set of random variables $\Omega \to M$.

# Example: Randomization

- Let $\Omega$ be a probability space, and let $\mathcal{M}$ be an $\mathcal{L}$-structure.
- Let $M^R$ be the set of random variables $\Omega \to M$.
- (Skipping measurability details.)

# Example: Randomization

- Let $\Omega$ be a probability space, and let $\mathcal{M}$ be an $\mathcal{L}$-structure.
- Let $M^R$ be the set of random variables $\Omega \to M$.
- (Skipping measurability details.)

**Definition**

Make $\mathcal{M}^R$ a metric structure, the *randomization* of $\mathcal{M}$, by equipping it with relations $\mathbb{E}[\phi(\bar{x})]$ for all $\mathcal{L}$-formulas $\phi$.

# Example: Randomization

- Let $\Omega$ be a probability space, and let $\mathcal{M}$ be an $\mathcal{L}$-structure.
- Let $M^R$ be the set of random variables $\Omega \to M$.
- (Skipping measurability details.)

**Definition**

Make $\mathcal{M}^R$ a metric structure, the *randomization* of $\mathcal{M}$, by equipping it with relations $\mathbb{E}[\phi(\bar{x})]$ for all $\mathcal{L}$-formulas $\phi$.

**Theorem (Ben Yaacov, Keisler)**

*$\mathcal{M}^R$ has QE in this language.*

# Uniform Law of Large Numbers

**Theorem (Vapnik-Chervonenkis)**

*The class $\mathcal{H}$ has finite VC-dimension if and only if*

- *for every suitable measure $\mu$ on $X$,*
- *for all $\varepsilon > 0, \delta > 0$, there is $n \in \mathbb{N}$ such that*
- *if $\bar{x} = (x_1, \ldots, x_n)$ consists of $n$ i.i.d. random samples from $\mu$,*

$$\mathbb{P}_{\bar{x}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n} h(x_i) - \mathbb{E}_{\mu}[h]\right| > \varepsilon\right] \leq \delta.$$

# Uniform Law of Large Numbers

**Theorem (Vapnik-Chervonenkis)**

*The class $\mathcal{H}$ has finite VC-dimension if and only if*

- *for every suitable measure $\mu$ on $X$,*
- *for all $\varepsilon > 0, \delta > 0$, there is $n \in \mathbb{N}$ such that*
- *if $\bar{x} = (x_1, \ldots, x_n)$ consists of $n$ i.i.d. random samples from $\mu$,*

$$\mathbb{P}_{\bar{x}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n} h(x_i) - \mathbb{E}_\mu[h]\right| > \varepsilon\right] \leq \delta.$$

The original result was for sets/functions $X \to \{0,1\}$, but the real-valued version is well-known by now.

# Symmetrization

Bound the average error by

- introducing another sample,
- swapping pairs $x_i, y_i$ at random,
- decoupling with the triangle inequality.

$$\mathbb{E}_{\bar{x}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}h(x_i)-\mathbb{E}_{\mu}[h]\right|\right]$$

$$\leq \mathbb{E}_{\bar{x},\bar{y}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}(h(x_i)-h(y_i))\right|\right]$$

$$= \mathbb{E}_{\bar{x},\bar{y},\sigma}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i(h(x_i)-h(y_i))\right|\right]$$

$$\leq 2\mathbb{E}_{\bar{x},\sigma}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right|\right]$$

# Symmetrization

Bound the average error by

- introducing another sample,
- swapping pairs $x_i, y_i$ at random,
- decoupling with the triangle inequality.

$$\mathbb{E}_{\bar{x}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}h(x_i)-\mathbb{E}_{\mu}[h]\right|\right]$$

$$\leq \mathbb{E}_{\bar{x},\bar{y}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}(h(x_i)-h(y_i))\right|\right]$$

$$= \mathbb{E}_{\bar{x},\bar{y},\sigma}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i(h(x_i)-h(y_i))\right|\right]$$

$$\leq 2\,\mathbb{E}_{\bar{x},\sigma}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i h(x_i)\right|\right] \leftarrow \mathcal{R}_{\mathcal{H},\mu}(n)$$

# Rademacher Complexity

**Definition**

Given $\bar{x} \in X^n$, define the *empirical Rademacher complexity*:

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) = \mathbb{E}_\sigma \left[ \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i h(x_i) \right| \right].$$

- Let $\bar{x} \in X^n$ be a specific, worst-case, or random sample.

# Rademacher Complexity

**Definition**

Given $\bar{x} \in X^n$, define the *empirical Rademacher complexity*:

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) = \mathbb{E}_\sigma \left[ \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i h(x_i) \right| \right].$$

- Let $\bar{x} \in X^n$ be a specific, worst-case, or random sample.
- Flip a coin $n$ times to split $\bar{x}$ into two random subsamples.

# Rademacher Complexity

**Definition**

Given $\bar{x} \in X^n$, define the *empirical Rademacher complexity*:

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) = \mathbb{E}_\sigma \left[ \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i h(x_i) \right| \right].$$

- Let $\bar{x} \in X^n$ be a specific, worst-case, or random sample.
- Flip a coin $n$ times to split $\bar{x}$ into two random subsamples.
- How much (in expectation) does $\frac{1}{n} \sum_{i \in E} h(x_i)$ differ between the two samples, for worst-case $h \in \mathcal{H}$?

# Rademacher Complexity

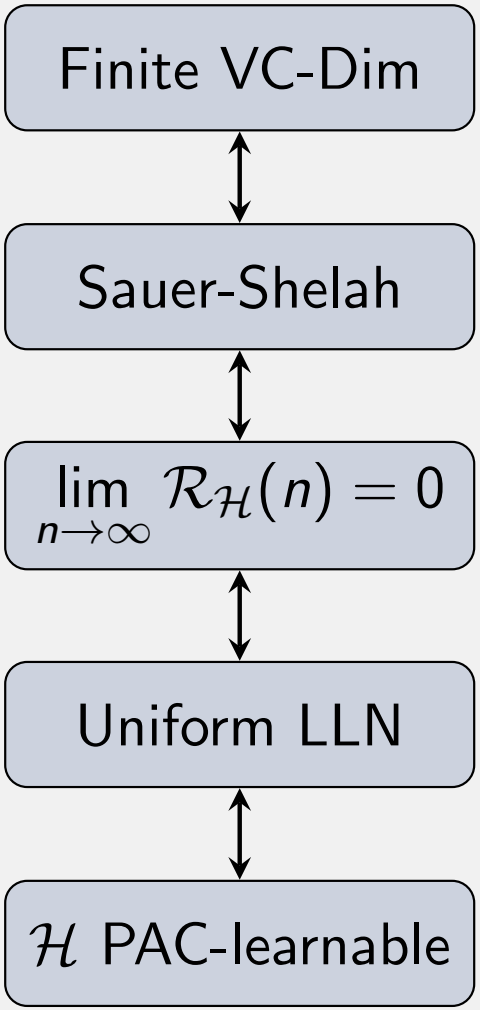
Taking a sup, we get the *worst-case Rademacher complexity*:

$$\mathcal{R}_{\mathcal{H}}(n) = \sup_{\bar{x}} \mathcal{R}_{\mathcal{H}}(\bar{x}),$$

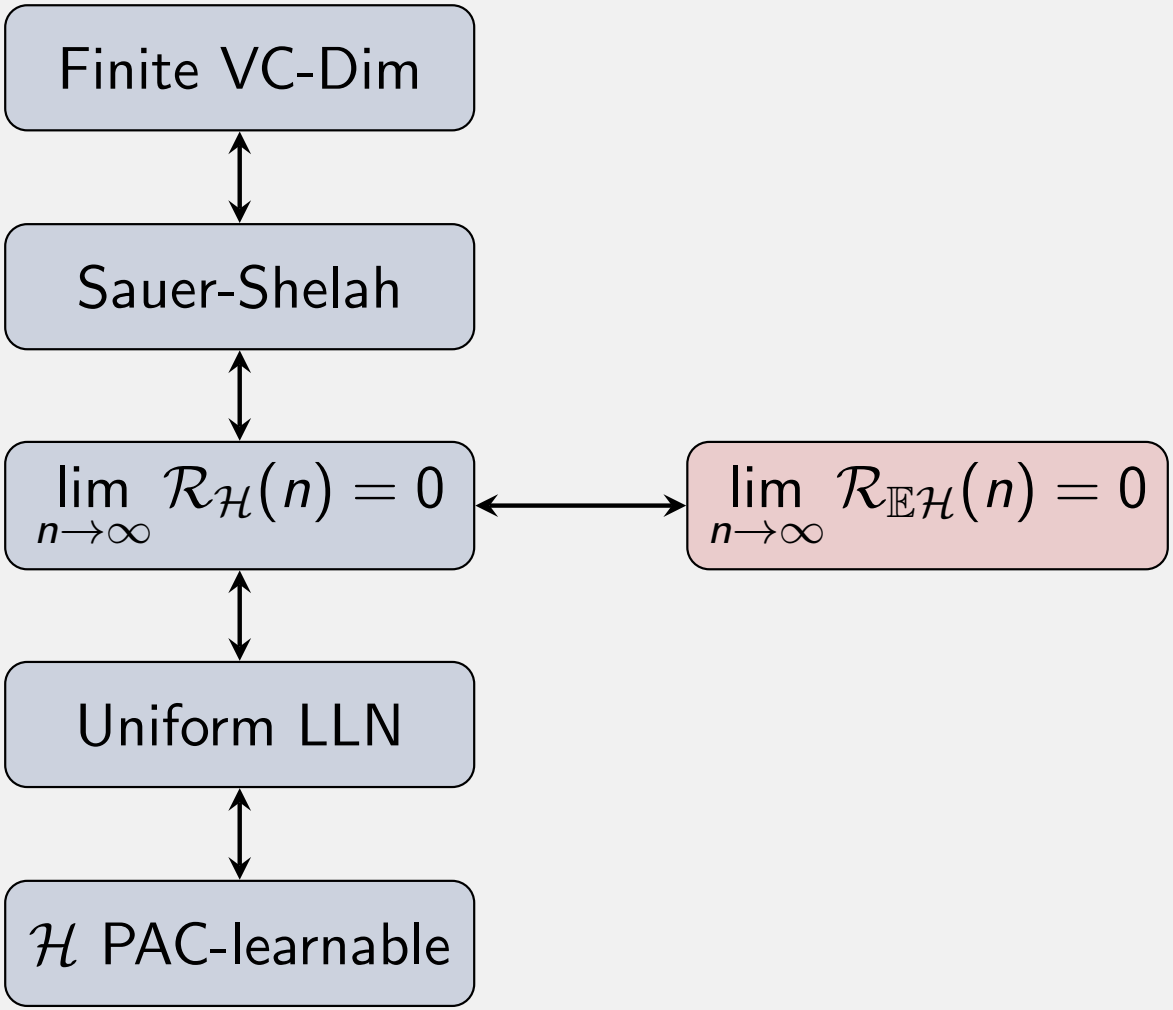
or taking an expectation, just *Rademacher complexity*:

$$\mathcal{R}_{\mathcal{H},\mu}(n) = \mathbb{E}_{\bar{x}} \left[ \mathcal{R}_{\mathcal{H}}(\bar{x}) \right].$$

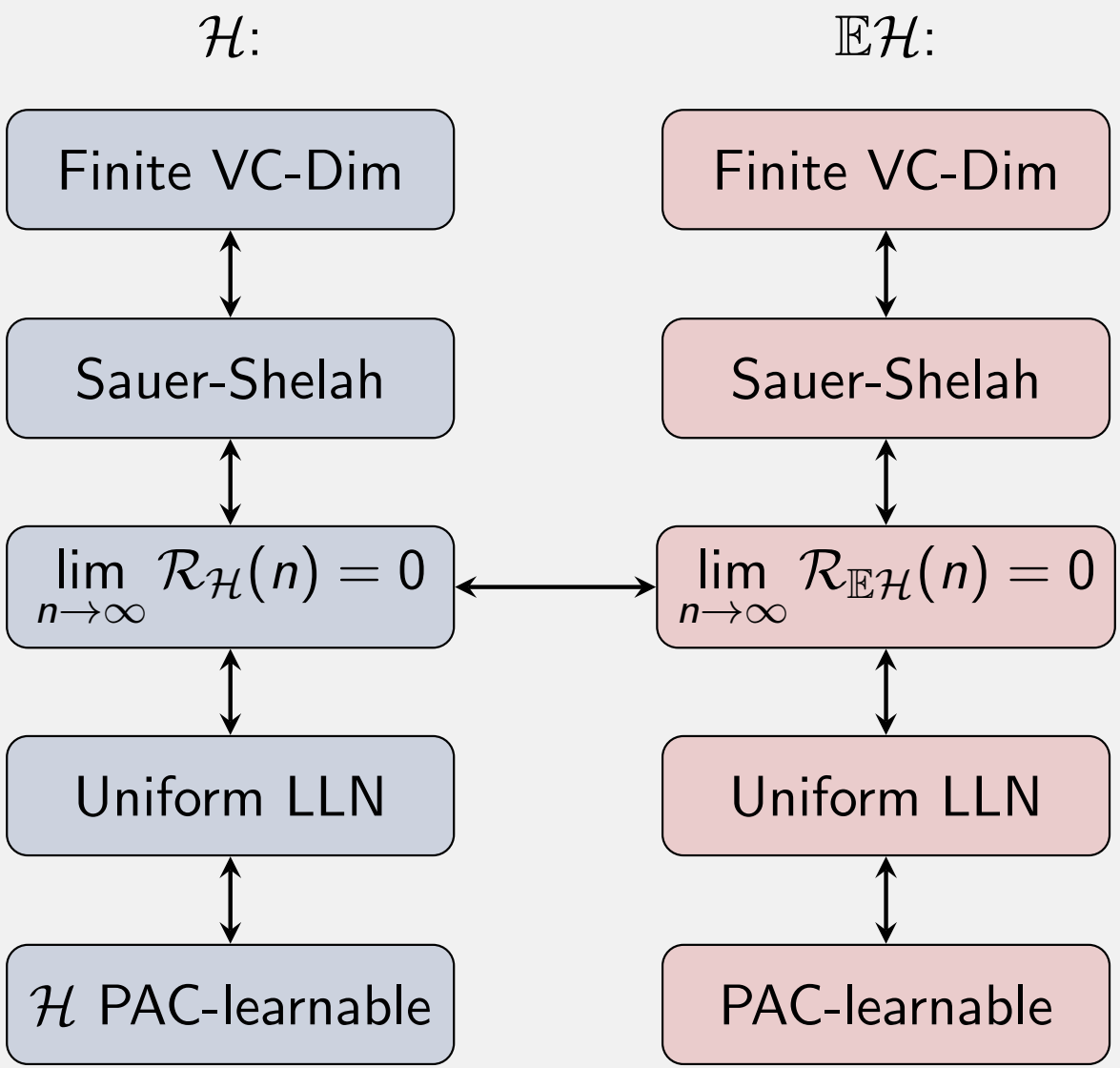
# Bridge to Learnability



# Bridge to Learnability and Randomization



# Bridge to Learnability and Randomization



# Strategy (including results and work in progress)

- Link combinatorial dimensions such as VC-dimension to (variants of) Rademacher complexity.

# Strategy (including results and work in progress)

- Link combinatorial dimensions such as VC-dimension to (variants of) Rademacher complexity.
- Derive some statistical results.

# Strategy (including results and work in progress)

- Link combinatorial dimensions such as VC-dimension to (variants of) Rademacher complexity.
- Derive some statistical results.
- Show that these nice properties are preserved under randomization.

# Real-Valued NIP

## Definition

For $\gamma > 0$, say $\mathcal{H}$ $\gamma$-*shatters* $\bar{x} \in X^n$ when there is a witness $v \in [0,1]^n$ such that for all $\sigma \in \{-1,1\}^n$, there is $h_\sigma \in \mathcal{H}$ with

$$\forall i \in [n], \sigma_i(h_\sigma(x_i) - v_i) \geq \frac{\gamma}{2}.$$

# Real-Valued NIP

**Definition**

For $\gamma > 0$, say $\mathcal{H}$ *$\gamma$-shatters* $\bar{x} \in X^n$ when there is a witness $v \in [0,1]^n$ such that for all $\sigma \in \{-1,1\}^n$, there is $h_\sigma \in \mathcal{H}$ with

$$\forall i \in [n], \sigma_i(h_\sigma(x_i) - v_i) \geq \frac{\gamma}{2}.$$

**Definition**

Say $\mathcal{H}$ has

- *$\gamma$-VC-dimension* $\geq d$ when it shatters $\bar{x} \in X^d$.
- *finite VC-dimension* when $\forall \gamma > 0$, $\gamma$-VC-dim is finite.

Say $\mathcal{M}$ is *NIP* when each $\mathcal{H}_\phi$ has finite VC-dimension.

# High-Arity NIP

## Notation

Suppose $X = \prod_{i=1}^{k} X_i$.
(Think $\phi(x_1, \ldots, x_k, y)$, and $X = \prod_{i=1}^{k} M^{x_i}$.)
Given tuples $\bar{x}^i \in X_i^n$ for $1 \leq i \leq k$, define $\vec{x} = \bigotimes_{i=1}^{k} \bar{x}^i \in X^n$ by

$$\vec{x_i} = (x_{i_1}^1, \ldots, x_{i_k}^k).$$

# High-Arity NIP

## Notation

Suppose $X = \prod_{i=1}^{k} X_i$.
(Think $\phi(x_1, \ldots, x_k, y)$, and $X = \prod_{i=1}^{k} M^{x_i}$.)
Given tuples $\bar{x}^i \in X_i^n$ for $1 \leq i \leq k$, define $\vec{x} = \bigotimes_{i=1}^{k} \bar{x}^i \in X^n$ by

$$\vec{x_\eta} = (x_{i_1}^1, \ldots, x_{i_k}^k).$$

## Definition

Say $\mathcal{H}$ has

- *$\gamma$-$VC_k$-dimension* $\geq d$ when it shatters $\bigotimes_{i=1}^{k} \bar{x}^i$ with $\bar{x}^i \in X^d$.
- *finite $VC_k$-dimension* when $\forall \gamma > 0$, $\gamma$-VC-dim is finite.

Say $\mathcal{M}$ is *$NIP_k$* when for each $\phi(x_1, \ldots, x_k, y)$, $\mathcal{H}_\phi$ has finite $VC_k$-dimension.

# Sauer-Shelah

To bound Rademacher complexity, start with Sauer-Shelah.

**Definition**

Given $\bar{x} \in X^n$, let $\mathcal{H}(\bar{x}) = \{(h(x_1), \ldots, h(x_n)) : h \in \mathcal{H}\}$.

# Sauer-Shelah

To bound Rademacher complexity, start with Sauer-Shelah.

**Definition**

Given $\bar{x} \in X^n$, let $\mathcal{H}(\bar{x}) = \{(h(x_1), \ldots, h(x_n)) : h \in \mathcal{H}\}$.

**Lemma (Sauer-Shelah)**

*If $\mathcal{H}$ is $\{0,1\}$-valued and has VC-dim $\leq d$, $|\mathcal{H}(\bar{x})| = O(n^d)$.*

# Sauer-Shelah

To bound Rademacher complexity, start with Sauer-Shelah.

**Definition**

Given $\bar{x} \in X^n$, let $\mathcal{H}(\bar{x}) = \{(h(x_1), \ldots, h(x_n)) : h \in \mathcal{H}\}$.

**Lemma (Sauer-Shelah)**

*If $\mathcal{H}$ is $\{0,1\}$-valued and has VC-dim $\leq d$, $|\mathcal{H}(\bar{x})| = O(n^d)$.*

**Lemma (Chernikov-Palacin-Takeuchi)**

*If $\mathcal{H}$ is $\{0,1\}$-valued and has $VC_k$-dim $\leq d$, there is $\varepsilon = \varepsilon(k,d)$ with*

$$\left|\mathcal{H}\left(\bigotimes_{i=1}^{k} \bar{x}^i\right)\right| = O(2^{n^{k-\varepsilon}}).$$

# Sauer-Shelah as Covering Number Bound

If $\mathcal{H}$ is $[0,1]$-valued, $\mathcal{H}(\bar{x})$ is probably infinite. Bound covering numbers instead.

**Definition**

Let $\mathcal{N}_\gamma(\mathcal{H}(\bar{x}))$ be the minimum size of a $\gamma$-*cover* $C$:

$$\forall a \in \mathcal{H}(\bar{x}), \exists c \in C, |a - c|_\infty \leq \gamma.$$

**Lemma (Alon, Ben-David, Cesa-Bianchi, Haussler)**

*If* $\mathcal{H}$ *has* $\frac{\gamma}{4}$*-VC-dimension* $\leq d$*,*

$$\log\left(\mathcal{N}_\gamma(\mathcal{H}(\bar{x}))\right) = O\left(d(\log n)^2\right)$$

# High-Arity Covering Number Bound

**Theorem (Erdős)**

*If* $I \subseteq [n]^k$ *has* $|I| \geq z_k(n, d+1)$, *where* $z_k(n, d+1) = O(n^{k-\varepsilon})$ *for some* $\varepsilon = \varepsilon(k, d)$, *then* $I$ *contains some* $\prod_{i=1}^{k} B_i$ *with* $|B_i| \geq d$.

# High-Arity Covering Number Bound

**Theorem (Erdős)**

*If $I \subseteq [n]^k$ has $|I| \geq z_k(n, d+1)$, where $z_k(n, d+1) = O(n^{k-\varepsilon})$ for some $\varepsilon = \varepsilon(k, d)$, then $I$ contains some $\prod_{i=1}^{k} B_i$ with $|B_i| \geq d$.*

If $\mathcal{N}_\gamma(\mathcal{H}(\vec{x}))$ is large, $\mathcal{H}$ shatters so many subtuples of $\vec{x}$ that one must have size $z_k(n, d+1)$, and contain a *grid* of size $d^k$.

# High-Arity Covering Number Bound

**Theorem (Erdős)**

*If $I \subseteq [n]^k$ has $|I| \geq z_k(n, d+1)$, where $z_k(n, d+1) = O(n^{k-\varepsilon})$ for some $\varepsilon = \varepsilon(k, d)$, then $I$ contains some $\prod_{i=1}^{k} B_i$ with $|B_i| \geq d$.*

If $\mathcal{N}_\gamma(\mathcal{H}(\vec{x}))$ is large, $\mathcal{H}$ shatters so many subtuples of $\vec{x}$ that one must have size $z_k(n, d+1)$, and contain a *grid* of size $d^k$.

**Lemma (A., Towsner)**

*If $\mathcal{H}$ has $\gamma$-$VC_k$-dimension $\leq d$, then for some $\varepsilon = \varepsilon(k, \gamma, d)$,*

$$\log\left(\mathcal{N}_\gamma(\mathcal{H}(\otimes_{i=1}^{k} \bar{x}^i))\right) = O\left(n^{k-\varepsilon}\right).$$

# High-Arity Rademacher Complexity

## Definition

If $X = \prod_{i=1}^{k} X_i$, for each $1 \leq i \leq k$, take a sup over $\bar{x}^i \in X_i^n$:

$$\mathcal{R}_{\mathcal{H}}^{k}(n) = \sup_{\bar{x}^i} \mathcal{R}_{\mathcal{H}} \left( \bigotimes_{i=1}^{k} \bar{x}^i \right).$$

# Rademacher Upper Bounds

**Theorem (variation on Massart's Lemma)**

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) \leq \inf_{\gamma \geq 0} \left( \gamma n + \sqrt{\frac{n}{2} \log(\mathcal{N}_{\gamma}(\mathcal{H}(\bar{x})))} \right).$$

# Rademacher Upper Bounds

**Theorem (variation on Massart's Lemma)**

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) \leq \inf_{\gamma \geq 0} \left( \gamma n + \sqrt{\frac{n}{2} \log(\mathcal{N}_{\gamma}(\mathcal{H}(\bar{x})))} \right).$$

**Corollary**

*If $\mathcal{H}$ has VC-dim $\leq d$, then for all $\gamma > 0$,*

$$\mathcal{R}_{\mathcal{H}}(n) \leq \gamma + O(n^{-1/2} \log n) \to \gamma.$$

# Rademacher Upper Bounds

**Theorem (variation on Massart's Lemma)**

$$\mathcal{R}_{\mathcal{H}}(\bar{x}) \leq \inf_{\gamma \geq 0} \left( \gamma n + \sqrt{\frac{n}{2} \log(\mathcal{N}_{\gamma}(\mathcal{H}(\bar{x})))} \right).$$

**Corollary**

*If $\mathcal{H}$ has VC-dim $\leq d$, then for all $\gamma > 0$,*

$$\mathcal{R}_{\mathcal{H}}(n) \leq \gamma + O(n^{-1/2} \log n) \to \gamma.$$

**Corollary (A., Towsner)**

*If $\mathcal{H}$ has $VC_k$-dim $\leq d$, then for all $\gamma > 0$,*

$$\mathcal{R}^k_{\mathcal{H}}(n) \leq \gamma + O\left(n^{-\varepsilon} \log n\right) \to \gamma.$$

# Rademacher Lower Bound

**Lemma**

*If $\mathcal{H}$ $\gamma$-shatters $\bar{x}$, then $\mathcal{R}_{\mathcal{H}}(\bar{x}) \geq \gamma$.*

**Corollary (A., Towsner)**

*If $\mathcal{H}$ has $VC_k$-dim $\geq d$, then $\mathcal{R}^k_{\mathcal{H}}(d) \geq \gamma$.*

**Theorem (A., Towsner)**

*$\mathcal{H}$ has finite $VC_k$-dim* $\iff \lim_{n\to\infty} \mathcal{R}^k_{\mathcal{H}}(n) = 0.$

# High-Arity Uniform Laws of Large Numbers

Using $\mathcal{R}^k_{\mathcal{H}}(n)$, we can reprove two uniform laws of large numbers, by generalizing the classic symmetrization proof.

# High-Arity Uniform Laws of Large Numbers

Using $\mathcal{R}^k_{\mathcal{H}}(n)$, we can reprove two uniform laws of large numbers, by generalizing the classic symmetrization proof.

**Theorem (A., Towsner after Coregliano, Malliaris)**

*If $\mathcal{H}$ has bounded $VC_k$-dimension, for large $n$, samples $\bar{x}^i \in X_i^n$, and uniformly random $E \subseteq [n]^k$,*

$$\mathbb{P}_E \left[ \sup_{h \in \mathcal{H}} \left| \frac{1}{n^k} \sum_{\bar{i} \in [n]^k} h(\vec{x_{\bar{i}}}) - \frac{1}{|E|} \sum_{\bar{i} \in E} h(\vec{x_{\bar{i}}}) \right| > \varepsilon \right] \leq \delta.$$

The average over a subsample approximates the average over the original sample.

# High-Arity Uniform Laws of Large Numbers

The average is stated in terms of conditional expectation, following the packing lemma of Chernikov, Towsner.

**Theorem (A., Towsner after Chernikov, Towsner)**

*If $\mathcal{H}$ has bounded $VC_k$-dimension, for large $n$ and any measures $\mu_i$,*

$$\mathbb{P}_{\mu_{\bar{n}^k}}\left[\sup_{h\in\mathcal{H}}\left|\frac{1}{n^k}\sum_{\bar{i}\in[n]^k} h(\vec{x}_{\bar{i}}) - \mathbb{E}[h(\vec{x}_{\bar{i}})|\mathcal{B}_{k-1}]\right| > \varepsilon\right] \leq \delta.$$

Idea: $\mathbb{E}[h(\vec{x}_{\bar{i}})|\mathcal{B}_{k-1}]$ depends only on $(k-1)$-ary information.

# General Mean Widths

## Definition

Given $A \subseteq [0,1]^n$, a random variable $\sigma \in \mathbb{R}^n$, let

$$w(A, \sigma) = \mathbb{E}\left[\sup_{a \in A} \sum_{i=1}^{n} \sigma_i a_i\right].$$

Can define $\mathcal{R}_{\mathcal{H}}(n), \mathcal{R}^k_{\mathcal{H}}(n), \mathcal{R}^{\text{seq}}_{\mathcal{H}}(n)$ with various $A$ and $\sigma$.

# General Mean Widths

## Definition

Given $A \subseteq [0,1]^n$, a random variable $\sigma \in \mathbb{R}^n$, let

$$w(A, \sigma) = \mathbb{E}\left[\sup_{a \in A} \sum_{i=1}^{n} \sigma_i a_i\right].$$

Can define $\mathcal{R}_{\mathcal{H}}(n), \mathcal{R}^k_{\mathcal{H}}(n), \mathcal{R}^{\text{seq}}_{\mathcal{H}}(n)$ with various $A$ and $\sigma$.

## Theorem (Rakhlin, Sridharan, Tewari)

*$\mathcal{R}^{\text{seq}}_{\mathcal{H}}(n) \to 0$ if and only if $\mathcal{H}$ has finite Littlestone dimension ($\mathcal{H}$ is stable).*

# Slicewise NIP/Stability

## Definition

If $X = \prod_{i=1}^{k} X_i$, say $\mathcal{H}$ has *slicewise VC-dimension* $\leq d$ when for all $1 \leq i \leq k$, $\bar{y} \in \prod_{j \neq i} X_i$, $\mathcal{H}(x_i, \bar{y})$ has VC-dimension $\leq d$.

Similar for Littlestone dimension and real-valued versions.

## Theorem (A., Towsner)

*There are sequences of random variables $\sigma$ defining $\mathcal{R}^{\mathrm{sl}}_{\mathcal{H}}(n), \mathcal{R}^{\mathrm{seq,sl}}_{\mathcal{H}}(n)$ which characterize finite slicewise VC-dimension, slicewise Littlestone dimension respectively.*

# Slicewise NIP/Stability

**Corollary (Originally Coregliano, Malliaris)**

*$\mathcal{H}$ having slicewise VC-dimension $\leq d$ implies a uniform law of large numbers over product measures.*

**Theorem (A., Towsner)**

*$\mathcal{H}$ having slicewise Littlestone dimension $\leq d$ implies an adversarial uniform law of large numbers.*

# Preservation Under Randomization

**Theorem (Ben Yaacov, Keisler)**

*If $\mathcal{M}$ is stable, so is $\mathcal{M}^R$.*

**Theorem (Ben Yaacov)**

*If $\mathcal{M}$ is NIP, so is $\mathcal{M}^R$.*

**Theorem (Chernikov, Towsner)**

*If $\mathcal{M}$ is $NIP_k$, so is $\mathcal{M}^R$.*

Three totally different proofs!

# Preservation Under Randomization

**Theorem (Ben Yaacov, Keisler)**

*If $\mathcal{M}$ is stable, so is $\mathcal{M}^R$.*

**Theorem (Ben Yaacov)**

*If $\mathcal{M}$ is NIP, so is $\mathcal{M}^R$.*
*Uses Gaussian complexity, similar to Rademacher complexity.*

**Theorem (Chernikov, Towsner)**

*If $\mathcal{M}$ is $NIP_k$, so is $\mathcal{M}^R$.*

Three totally different proofs!

# Preservation Under Randomization

**Lemma (A., Benedikt)**

*Let $\bar{\mathbf{x}}$ be a tuple of random variables in $X$, and $\sigma$ random in $[-1,1]^n$.*

$$w(\mathcal{H}_{\mathbb{E}[\phi]}(\bar{\mathbf{x}}), \sigma) \leq \sup_{\omega \in \Omega} w(\mathcal{H}(\bar{x}_\omega), \sigma).$$

# Preservation Under Randomization

## Lemma (A., Benedikt)

*Let $\bar{\mathbf{x}}$ be a tuple of random variables in $X$, and $\sigma$ random in $[-1,1]^n$.*

$$w(\mathcal{H}_{\mathbb{E}[\phi]}(\bar{\mathbf{x}}),\sigma) \leq \sup_{\omega\in\Omega} w(\mathcal{H}(\bar{x}_\omega),\sigma).$$

## Corollary

- $\mathcal{R}_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}_{\mathcal{H}_\phi}(n)$
- $\mathcal{R}^k_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}^k_{\mathcal{H}_\phi}(n)$
- $\mathcal{R}^{\mathrm{seq}}_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}^{\mathrm{seq}}_{\mathcal{H}_\phi}(n)$

# Preservation Under Randomization

**Lemma (A., Benedikt)**

*Let $\bar{\mathbf{x}}$ be a tuple of random variables in $X$, and $\sigma$ random in $[-1,1]^n$.*

$$w(\mathcal{H}_{\mathbb{E}[\phi]}(\bar{\mathbf{x}}), \sigma) \leq \sup_{\omega \in \Omega} w(\mathcal{H}(\bar{x}_\omega), \sigma).$$

**Corollary**

- $\mathcal{R}_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}_{\mathcal{H}_\phi}(n)$
- $\mathcal{R}^k_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}^k_{\mathcal{H}_\phi}(n)$
- $\mathcal{R}^{\text{seq}}_{\mathcal{H}_{\mathbb{E}[\phi]}}(n) \leq \mathcal{R}^{\text{seq}}_{\mathcal{H}_\phi}(n)$

**Corollary**

- *NIP*
- *NIP$_k$*
- *Stability*

*preserved under randomization.*

# Future Questions

When can trace-definability dividing lines be characterized in terms of Rademacher complexity variants?

# Future Questions

When can trace-definability dividing lines be characterized in terms of Rademacher complexity variants?
(Main missing step seems to be Sauer-Shelah.)

# Future Questions

When can trace-definability dividing lines be characterized in terms of Rademacher complexity variants?
(Main missing step seems to be Sauer-Shelah.)

Connect $\mathrm{VC}_k$-dimension to something closer to standard PAC learning.

These complexities are all about the *expected discrepancy* between two random subsamples. Distality/strong Erdös-Hajnal seems to be connected to other measures of discrepancy.

Thank you, UIC!