CHAPTER 6

Epsilon-Nets: Combinatorial Bounds

The initial study of ϵ-nets in the field of computational geometry started in the 1980s with the work of Clarkson who showed the existence of ϵ-nets of size $O\left(\frac{1}{\epsilon}\log\frac{1}{\epsilon}\right)$ for specific geometric set systems. He was mainly interested in their algorithmic applications, such as nearest-neighbor queries for a set of points in Euclidean space. Chapter 2 is largely based on his work.

Independently, Haussler and Welzl showed similar bounds in a purely abstract setting, needing just that the given set system has bounded VC-dimension. In fact, what was needed, given a set system $(X, \mathcal{F})$, was the property that there exist an absolute constant d such that

$$\text{for all } Y \subseteq X,\ |\mathcal{F}|_Y| = O\left(|Y|^d\right) \quad \text{(see Lemma 4.3).}$$

They showed, surprisingly, that this is already a sufficient condition for the existence of small ϵ-nets; the following will be the first theorem of this chapter.

THEOREM 6.1. *Let $(X, \mathcal{F})$ be a finite set system, $d \geq 1$ an integer such that* VC-dim$(\mathcal{F}) \leq d$*, and $\epsilon \in \left(0, \frac{1}{2}\right)$ a given parameter. Let N be a uniform random sample of X of size $t = \left\lceil \frac{56\,d}{\epsilon}\ln\frac{1}{\epsilon}\right\rceil$. Then N is an ϵ-net of $\mathcal{F}$ with probability at least $\frac{1}{2}$.*

It was later shown that the above bound is optimal within constant factors; that is, for every positive integer n, integer $d \geq 2$ and small-enough parameter $\epsilon > 0$, there is a set system $(X, \mathcal{F})$ with $|X| = n$ and VC-dim$(\mathcal{F}) \leq d$, such that any ϵ-net of $\mathcal{F}$ has size $\Omega\left(\frac{d}{\epsilon}\log\frac{1}{\epsilon}\right)$. This lower bound will be presented in Chapters 10 and 11.

Over the past thirty years, it has been observed that improvements to the Clarkson and Haussler–Welzl bounds are possible for a variety of geometric set systems. We have already seen an example in Chapter 3: $O\left(\frac{1}{\epsilon}\right)$-sized ϵ-nets exist for set systems induced by disks in the plane. Early work towards $o\left(\frac{1}{\epsilon}\log\frac{1}{\epsilon}\right)$ upper bounds was fundamentally geometric, involving spatial partitioning along the ideas seen in Chapter 3.

Over the next twenty years, through the work of Aronov, Chan, Clarkson, Ezra, Ray, Sharir, Varadarajan and others, it was realized that geometry is not really needed. In fact, somewhat surprisingly, an entire suite of optimal bounds can be obtained entirely combinatorially, with the shallow-cell complexity being a key parameter of a set system that dictates the size of ϵ-nets.

> The reason that the shallow-cell complexity of a set system comes into play is the following. Given a set system $(X, \mathcal{F})$, the probability that a set $F \in \mathcal{F}$ is not hit by a uniform random sample $N \subseteq X$ decreases exponentially with the size of F. On the other hand, the number of sets

of $\mathcal{F}$ of size at most k is an increasing function of k whose growth is upper bounded by the shallow-cell complexity of the set system. It turns out that the interplay between these two dictates the sizes of ϵ-nets.

The second main theorem of this chapter will be the following.

THEOREM 6.2. *Let $(X, \mathcal{F})$ be a finite set system with shallow-cell complexity $\varphi_{\mathcal{F}}(\cdot, \cdot)$ and with* VC-dim$(\mathcal{F}) \leq d$. *Then for any $\epsilon \in \left(0, \frac{1}{2}\right)$ there exists an ϵ-net of $\mathcal{F}$ of size*

$$O\left(\frac{d}{\epsilon} + \frac{1}{\epsilon}\log \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon}, 48d\right)\right).$$

Consider the primal set system $\mathcal{R}$ induced by disks in the plane: as $\varphi_{\mathcal{R}}(m, k) = O\left(k^2\right)$ (Lemma 1.2) and VC-dim$(\mathcal{R}) \leq 3$, Theorem 6.2 implies the existence of ϵ-nets of size $O\left(\frac{1}{\epsilon}\right)$. Thus we recover Theorem 3.3 from just the shallow-cell complexity of $\mathcal{R}$!

1. A First Bound using Ghost Sampling

Mathematics is one of a few fields in which one can do top-level work without a lot of life experience, something that might be key in the arts or humanities. One does not have to have experience raising children through school, dealing with family tragedies, and so forth, to be able to find three numbers whose fourth powers add up to another one.

Noam Elkies

We prove the following.

THEOREM 6.1. *Let $(X, \mathcal{F})$ be a finite set system, $d \geq 1$ an integer such that* VC-dim$(\mathcal{F}) \leq d$, *and $\epsilon \in \left(0, \frac{1}{2}\right)$ a given parameter. Let N be a uniform random sample of X of size $t = \left\lceil \frac{56\,d}{\epsilon} \ln \frac{1}{\epsilon} \right\rceil$. Then N is an ϵ-net of $\mathcal{F}$ with probability at least $\frac{1}{2}$.*

Set $n = |X|$. We can assume that each set in $\mathcal{F}$ has size at least ϵn.

For ease of calculations, we will allow an element to be picked into N multiple times. That is, N will be a *sequence* of size $t = \left\lceil \frac{56d}{\epsilon} \ln \frac{1}{\epsilon} \right\rceil$, where each element in this sequence is chosen uniformly at random from X. In a natural way, a sequence Y is an ϵ-net of $\mathcal{F}$ if Y contains at least one element from each set of $\mathcal{F}$.

Throughout this section we will work with sequences instead of sets. Moreover, the *size* of the intersection of any set $R \in \mathcal{F}$ with a sequence will count multiplicities.

Overview of ideas. Let $\mathcal{T} = X^t$ denote the set of all t-sized sequences of elements of X, and let $\mathcal{T}_b \subseteq \mathcal{T}$ be the sequences which are *not* an ϵ-net of $\mathcal{F}$. Our goal is to upper bound $|\mathcal{T}_b|$. In particular, we will show that $|\mathcal{T}_b| < \frac{|\mathcal{T}|}{2}$, implying that N—constructed by picking t elements uniformly at random, with replacement—is an ϵ-net of $\mathcal{F}$ with probability at least $\frac{1}{2}$. This implies the same property for a uniform random sample of X of size t, proving Theorem 6.1.

Fix an ordering of the sets of $\mathcal{F}$ and

for each $Y \in \mathcal{T}_b$, let $R_Y \in \mathcal{F}$ be the *first* set in the ordering for which Y fails—that is, for which $R_Y \cap Y = \emptyset$.

We will use the probabilistic averaging technique from Chapter 1. That is, we take a uniform random sequence S of X of a certain size s—with replacement, so $S \in X^s$—and examine the relationship between $\mathcal{T}_b$ and S. Specifically,

we count the expected number of sequences $Y \in \mathcal{T}_b$ s.t. $|R_Y \cap S| \geq \frac{\epsilon s}{2}$.

Lower bound: On one hand, for any $Y \in \mathcal{T}_b$,

$$\mathrm{E}\left[\,|R_Y \cap S|\,\right] = \sum_{i=1}^{s} \frac{|R_Y|}{n} \geq \sum_{i=1}^{s} \epsilon = \epsilon s, \tag{6.3}$$

keeping in mind that each element of R_Y is counted with multiplicity in $|R_Y \cap S|$. A tail bound will then imply that the expected number of sets $Y \in \mathcal{T}_b$ for which $|R_Y \cap S| \geq \frac{\epsilon s}{2}$ is at least $\frac{|\mathcal{T}_b|}{2}$.

Upper bound: On the other hand, this expectation can be calculated exactly:

$$\frac{1}{n^s}\sum_{Z\in X^s}\left|\left\{Y\in\mathcal{T}_b\ :\ |R_Y\cap Z|\ \geq\frac{\epsilon s}{2}\right\}\right|$$

$$= \frac{1}{n^s}\ \left|\left\{(Z,Y):\ Z\in X^s,\ Y\in\mathcal{T}_b\subseteq X^t,\ |R_Y\cap Z|\geq\frac{\epsilon s}{2}\right\}\right|. \tag{6.4}$$

Call each of the above (Z, Y) a *satisfying pair*. That is,

$$Z\in X^s,\quad Y\in\mathcal{T}_b\subseteq X^t,\qquad |R_Y\cap Y|=0\quad\text{while}\quad |R_Y\cap Z|\geq\tfrac{\epsilon s}{2}.$$

We will show that these constraints together force an upper bound on the total number of satisfying pairs.

Combining the lower and upper bounds will then give the desired bound on $|\mathcal{T}_b|$.

Proof of Theorem 6.1. For $Z\in X^s$, define

$$\mathcal{T}_Z=\left\{Y\in\mathcal{T}_b\ :\ |R_Y\cap Z|\ \geq\frac{\epsilon s}{2}\right\}.$$

Set $s=t$ and let S be an element chosen uniformly at random from X^s.

We count the expected size of $\mathcal{T}_S$ in two ways.

Lemma 6.5 (Lower bound).

$$\mathrm{E}\left[\,|\mathcal{T}_S|\,\right]=\sum_{Y\in\mathcal{T}_b}\Pr\left[|R_Y\cap S|\geq\frac{\epsilon s}{2}\right]\geq\frac{1}{2}\cdot|\mathcal{T}_b|.$$

Proof. The proof follows from linearity of expectation and the next claim.

Claim 6.6. For any $R\in\mathcal{F}$, $\Pr\left[|R\cap S|\leq\frac{\epsilon s}{2}\right]<\frac{1}{2}$.

Proof. For $i=1,\ldots,s$, let Y_i be an indicator random variable that is 1 if and only if the i-th element of S is in R. Then

$$|R\cap S|=\sum_{i=1}^{s}Y_i,\quad\text{where}\quad\Pr\left[Y_i=1\right]=\frac{|R|}{n}\geq\epsilon.$$

Then Chernoff's bound (Theorem 1.20) applied to the s variables $\{Y_1,\ldots,Y_s\}$ with $\delta=\frac{1}{2}$ implies that

$$\Pr\left[|R\cap S|\leq\left(1-\frac{1}{2}\right)\epsilon s\right]\leq\exp\left(-\frac{\epsilon s}{8}\right)\leq\exp\left(-\frac{56d\ln\frac{1}{\epsilon}}{8}\right)<\frac{1}{2},$$

recalling that $s=t=\left\lceil\frac{56d}{\epsilon}\ln\frac{1}{\epsilon}\right\rceil$. ☐

The upper bound of $\frac{1}{2}$ in Claim 6.6 can be replaced by $o(1)$ but it doesn't matter for us as this only changes the multiplicative constant factor in the final bound (see discussion). ☐

The upper bound on $\mathrm{E}\left[|\mathcal{T}_S|\right]$ is implied by the following combinatorial statement.

Lemma 6.7 (Upper bound).

$$\sum_{Z\in X^s}|\mathcal{T}_Z|=\left|\left\{(Z,Y):\quad Z\in X^s,\ Y\in\mathcal{T}_b\subseteq X^t,\ |R_Y\cap Z|\geq\frac{\epsilon s}{2}\right\}\right|<\frac{n^{2t}}{4}. \tag{6.8}$$

PROOF. The trick to showing Equation (6.8) is to use averaging again.

For each $U \in X^{s+t}$ and $R \in \mathcal{F}$, let $\Psi(U,R)$ be the number of ways of partitioning U into two subsequences $Z \in X^s$ and $Y \in X^t$ such that (Z,Y) is a satisfying pair, with $R_Y = R$.

As each satisfying pair (Z,Y) can be combined into a sequence of size $s+t$ in $\binom{s+t}{t}$ ways, we have

$$\binom{s+t}{t} \sum_{Z \in X^s} |\mathcal{T}_Z| = \sum_{U \in X^{s+t}} \sum_{R \in \mathcal{F}} \Psi(U,R). \tag{6.9}$$

We make two observations:

(1) For each fixed $U \in X^{s+t}$, let $\mathcal{F}' \subseteq \mathcal{F}$ be such that all the sets in $\mathcal{F}'$ have the same intersection with U. Then all $R \in \mathcal{F}'$ have the same intersection with any (Z,Y) derived from U, implying that $\Psi(U,R)$ is (possibly) non-zero only for the first set, according to our initial ordering, of $\mathcal{F}'$. As there are $|\mathcal{F}|_U|$ distinct intersections of sets of $\mathcal{F}$ with U, there are at most $|\mathcal{F}|_U|$ sets $R \in \mathcal{F}$ for which $\Psi(U,R)$ is non-zero.

(2) For a fixed $R \in \mathcal{F}$ with $|R \cap U| \geq \frac{\epsilon s}{2}$, there are at most $\binom{s+t-\frac{\epsilon s}{2}}{t}$ ways to select Y from U such that Y does not contain any element of R. This is an upper bound on $\Psi(U,R)$ for any fixed U and R.

The above two observations together with Equation (6.9) imply that $\sum_{Z \in X^s} |\mathcal{T}_Z|$ can be upper bounded by

$$\frac{1}{\binom{s+t}{t}} \cdot \sum_{U \in X^{s+t}} |\mathcal{F}|_U| \cdot \binom{s+t-\frac{\epsilon s}{2}}{t} \leq \sum_{U \in X^{s+t}} \cdot \left(\frac{e(s+t)}{d}\right)^d \cdot \frac{\binom{s+t-\frac{\epsilon s}{2}}{t}}{\binom{s+t}{t}},$$

where the second step follows from Lemma 4.3. It remains to simplify this upper bound:

$$\begin{aligned}
&= n^{2t} \cdot \left(\frac{2et}{d}\right)^d \cdot \frac{2t-\frac{\epsilon t}{2}}{2t} \, \frac{2t-\frac{\epsilon t}{2}-1}{2t-1} \cdots \frac{t-\frac{\epsilon t}{2}+1}{t+1} \quad \Big(\text{recalling that } s=t\Big) \\
&= n^{2t} \cdot \left(\frac{2et}{d}\right)^d \cdot \left(1-\frac{\frac{\epsilon t}{2}}{2t}\right) \cdots \left(1-\frac{\frac{\epsilon t}{2}}{t+1}\right) \;\leq\; n^{2t} \cdot \left(\frac{2et}{d}\right)^d \cdot \left(1-\frac{\frac{\epsilon t}{2}}{2t}\right)^t \\
&\leq n^{2t} \cdot \left(\frac{2et}{d}\right)^d \cdot e^{-\frac{\epsilon t}{4}} \leq n^{2t} \cdot \left(\frac{112\,e}{\epsilon}\ \ln\frac{1}{\epsilon}\right)^d \cdot e^{-14d\ln\frac{1}{\epsilon}} \\
&= n^{2t} \cdot \left(\frac{112\,e}{\epsilon}\ \ln\frac{1}{\epsilon}\right)^d \cdot \epsilon^{14d} < n^{2t} \cdot \left(\frac{112\,e}{\epsilon^2}\right)^d \cdot \epsilon^{14d} \\
&< n^{2t} \cdot (112\,e)^d \cdot \epsilon^{12d} < \frac{n^{2t}}{4},
\end{aligned}$$

as $(112e)^d \cdot \epsilon^{12d} \leq (112e)^d \cdot \left(\frac{1}{2}\right)^{12d} < \frac{1}{4}$. ☐

Combining the upper and lower bounds,

$$\frac{1}{2} \cdot |\mathcal{T}_b| \quad \leq \quad \mathrm{E}\left[|\mathcal{T}_S|\right] \quad = \quad \frac{1}{n^s} \sum_{Z \in X^s} |\mathcal{T}_Z| \quad < \quad \frac{n^t}{4},$$

gives $|\mathcal{T}_b| < \frac{n^t}{2} = \frac{|\mathcal{T}|}{2}$, as required. ☐

A remark: the upper bound used in the proof, that

$$|\mathcal{F}|_U| \leq \left(\frac{e|U|}{d}\right)^d,$$

is a consequence of the fact that VC-dim $(\mathcal{F}) \leq d$. If instead we only used $|\mathcal{F}|_U| = O\left(|U|\right)^d$, the final bound would come out to be $O\left(\frac{d}{\epsilon}\log\frac{d}{\epsilon}\right)$.

We conclude with two remarks.

Ghost sampling: A clever double-counting trick in the proof is to upper bound the number of satisfying pairs (Z, Y)—where $Z \in X^s$ and $Y \in X^t$—by enumerating over all $(s+t)$-sized subsets of X. This trick was also used in the proof of Theorem 5.1, though in that case we had $t = 1$.

In statistics and learning theory literature, this instance of double-counting is called *ghost sampling*, and it is a useful technique to avoid discretization when the base set X is infinite. For the case when the set system $(X, \mathcal{F})$ is finite, there are simpler proofs (see Chapter 12).

In the proof that we presented, we only wanted the probability that the random sample N succeeds to be an ϵ-net of $\mathcal{F}$ to be non-zero, which is sufficient to guarantee the existence of an ϵ-net of the required size. By introducing this probability as a parameter in the sample size and re-working the above proof, we arrive at the following statement (a proof is presented in Chapter 12).

THEOREM 6.10. *Let $(X, \mathcal{F})$ be a finite set system, $d \in \mathbb{N}$ a positive integer such that* VC-dim$(\mathcal{F}) \leq d$, *and $\epsilon \in \left(0, \frac{1}{2}\right)$ a given parameter. Then there exists an absolute constant $C_6 > 0$ such that a random sample N constructed by picking each point of X independently with probability $\frac{C_6}{\epsilon|X|}\ln\frac{1}{\epsilon^d\gamma}$ is an ϵ-net of $\mathcal{F}$ with probability at least $1-\gamma$.*

Iterative View: As is often the case with double-counting proofs, one can 'unroll' the proof of Theorem 6.1 to an iterative version, as follows[1]. Let $(X_0, \mathcal{F}_0)$ be a set system with $|X_0| = n$ and $\epsilon > 0$ a parameter such that each set of $\mathcal{F}_0$ has size at least ϵn. A straightforward application of Chernoff's bound implies that there exists a $X_1 \subseteq X_0$ such that each $S \in \mathcal{F}_0$ contains at least $\frac{\epsilon n}{2}$ elements of X_1 and further

$$|X_1| \leq |X_0| \cdot \left(\frac{1}{2} + \sqrt{\frac{10\log|\mathcal{F}_0|}{\epsilon n}}\right).$$

Now one can repeat this step for the set system $(X_1, \mathcal{F}_1 = \mathcal{F}|_{X_1})$ to get a set $X_2 \subseteq X_1$ and so on. After the i-th iteration we have a set X_i such that each $S \in \mathcal{F}_0$ contains at least $\frac{\epsilon n}{2^i}$ elements of X_i and furthermore,

$$|X_i| \leq |X_{i-1}| \cdot \left(\frac{1}{2} + \sqrt{\frac{10\log|\mathcal{F}_{i-1}|}{\frac{\epsilon n}{2^{i-1}}}}\right) \leq \quad \cdots \quad \leq n \cdot \prod_{j=0}^{i-1}\left(\frac{1}{2} + \sqrt{\frac{10\log|\mathcal{F}_j|}{\frac{\epsilon n}{2^j}}}\right).$$

[1] Also this can be done by an inductive argument, though that somewhat obscures the ideas.

We continue the iterations as long as iteration i satisfies $\sqrt{\frac{10\log|\mathcal{F}_i|}{\frac{\epsilon n}{2^i}}} \leq \frac{1}{2}$. Say the above procedure runs for t iterations. Now a calculation shows that for all $i \leq t$,

$$|X_i| \leq n \cdot \prod_{j=0}^{i-1}\left(\frac{1}{2} + \sqrt{\frac{10\log|\mathcal{F}_j|}{\frac{\epsilon n}{2^j}}}\right) \leq c_1 \cdot \frac{n}{2^i}, \tag{6.11}$$

where c_1 is a sufficiently large constant. We set $t = \left\lfloor \log \frac{\epsilon n}{c'\, d\, \log\frac{1}{\epsilon}} \right\rfloor$ for a sufficiently large constant c' depending only on c_1. Then it can be verified that for all $i \leq t$,

$$\sqrt{\frac{10\log|\mathcal{F}_i|}{\frac{\epsilon n}{2^i}}} \leq \sqrt{\frac{10\log|\mathcal{F}_t|}{\frac{\epsilon n}{2^t}}} \leq \sqrt{\frac{10\log\left(\frac{ec_1 n}{2^t d}\right)^d}{\frac{\epsilon n}{2^t}}} \leq \sqrt{\frac{10\log\left(\frac{ec_1 c' \log\frac{1}{\epsilon}}{\epsilon}\right)}{c'\log\frac{1}{\epsilon}}} \leq \frac{1}{2},$$

where the second step follows from Equation (6.11) and Lemma 4.3.
Finally, each $S \in \mathcal{F}_0$ contains at least $\frac{\epsilon n}{2^t} \geq c' d \log\frac{1}{\epsilon} \geq 1$ points of X_t. Thus X_t is an ϵ-net of $\mathcal{F}_0$, with

$$|X_t| \leq c_1 \cdot \frac{n}{2^t} = c_1 \cdot \frac{n}{\frac{\epsilon n}{c'\, d\, \log\frac{1}{\epsilon}}} = O\left(\frac{d}{\epsilon}\log\frac{1}{\epsilon}\right).$$

An elegant strengthened form of this idea is the basis of Chapter 8.

Bibliography and discussion. A slightly weaker bound than the main theorem was first shown in [**HW876a**], and built upon the work in [**VC716a**]. The bound of this section is from [**Blu+89**]. The proof is usually stated entirely in probabilistic language while we have chosen a more combinatorial exposition that makes clear its basis in the probabilistic averaging technique of Chapter 1. The proof also follows immediately from the more general notion of ϵ-approximations (see Chapter 13).
The constant '56' in Theorem 6.1 can be replaced, with more precise calculations, by $1 + o(1)$! In particular, it was shown in [**KPW926a**] that a uniform random sample obtained by $\frac{d}{\epsilon}\left(\log\frac{1}{\epsilon} + 2\log\log\frac{1}{\epsilon} + 3\right)$ independent draws is an ϵ-net with probability at least $1 - e^{-d}$.

[Blu+89] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, *Learnability and the Vapnik-Chervonenkis dimension*, J. Assoc. Comput. Mach. **36** (1989), no. 4, 929–965, DOI 10.1145/76359.76371. MR1072253

[HW876a] D. Haussler and E. Welzl, *ϵ-nets and simplex range queries*, Discrete Comput. Geom. **2** (1987), no. 2, 127–151, DOI 10.1007/BF02187876. MR884223

[KPW926a] J. Komlós, J. Pach, and G. Woeginger, *Almost tight bounds for ϵ-nets*, Discrete Comput. Geom. **7** (1992), no. 2, 163–173, DOI 10.1007/BF02187833. MR1139078

[VC716a] V. N. Vapnik and A. Ja. Červonenkis, *The uniform convergence of frequencies of the appearance of events to their probabilities* (Russian, with English summary), Teor. Verojatnost. i Primenen. **16** (1971), 264–279. MR0288823

2. Optimal ϵ-Nets using Packings

The question you raise, 'how can such a formulation lead to computations?' doesn't bother me in the least! Throughout my whole life as a mathematician, the possibility of making explicit, elegant computations has always come out by itself, as a byproduct of a thorough conceptual understanding of what was going on. Thus I never bothered about whether what would come out would be suitable for this or that, but just tried to understand—and it always turned out that understanding was all that mattered.

Alexandre Grothendieck

Given a set system $(X, \mathcal{F})$, we now show the existence of small ϵ-nets of $\mathcal{F}$ as a function of its shallow-cell complexity, which we first recall.

Definition 4.4. A set system $(X, \mathcal{F})$ has shallow-cell complexity $\varphi_{\mathcal{F}}(\cdot, \cdot)$ if for any positive integer k and any finite $Y \subseteq X$, the number of sets in $\mathcal{F}|_Y$ of size at most k is upper bounded by $|Y| \cdot \varphi_{\mathcal{F}}\big(|Y|, k\big)$.

For a family $\mathcal{R}$ of geometric objects in $\mathbb{R}^d$—e.g., the family of all half-spaces—the shallow-cell complexity of $\mathcal{R}$ is defined to be the shallow-cell complexity of the primal set system $\big(\mathbb{R}^d, \mathcal{R}\big)$.

The main theorem we will prove in this section is the following.

Theorem 6.2. *Let $(X, \mathcal{F})$ be a finite set system with shallow-cell complexity $\varphi_{\mathcal{F}}(\cdot, \cdot)$ and with* VC-dim$(\mathcal{F}) \leq d$. *Then for any $\epsilon \in \big(0, \frac{1}{2}\big)$ there exists an ϵ-net of $\mathcal{F}$ of size*

$$O\left(\frac{d}{\epsilon} + \frac{1}{\epsilon}\log \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon}, 48d\right)\right).$$

Overview of ideas. The proof requires three ideas. For the moment assume that each set in $\mathcal{F}$ has size exactly ϵn.

(1) Fix any *maximal* subset $\mathcal{P} \subseteq \mathcal{F}$ such that every pair of sets of $\mathcal{P}$ have symmetric difference at least $\frac{\epsilon n}{2}$. In other words, $\mathcal{P}$ is a maximal $\big(\epsilon n, \frac{\epsilon n}{2}\big)$-packing of $\mathcal{F}$ (see Definition 5.9). Setting $\delta = \frac{\epsilon n}{2}$, $k = \epsilon n$ and applying Theorem 5.10, we have

$$|\mathcal{P}| \leq \frac{48dn}{\delta} \cdot \varphi_{\mathcal{F}}\left(\frac{8dn}{\delta}, \frac{24dk}{\delta}\right) = O\left(\frac{d}{\epsilon} \cdot \varphi_{\mathcal{F}}\left(O\left(\frac{d}{\epsilon}\right), O(d)\right)\right).$$

The key point here is that for each $F \in \mathcal{F} \setminus \mathcal{P}$, the maximality of $\mathcal{P}$ implies that there exists a set $S \in \mathcal{P}$ such that the size of the symmetric difference between F and S is at most $\frac{\epsilon n}{2}$. In other words, F contains at least $\frac{\epsilon n}{2}$ points from S, where $|S| = \epsilon n$ by assumption. Thus a $\frac{1}{2}$-net N_S for the set system $(S, \mathcal{F}|_S)$ must hit F and consequently $\bigcup_{S \in \mathcal{P}} N_S$ is an ϵ-net of $\mathcal{F}$. Here the sets of $\mathcal{P}$ play the role of canonical objects of Chapter 2.

Simply picking a $\frac{1}{2}$-net N_S *separately* for each $(S, \mathcal{F}|_S)$, $S \in \mathcal{P}$, where each N_S is of constant size by Theorem 6.10, will give an ϵ-net of $\mathcal{F}$ of total size

$$\sum_{S \in \mathcal{P}} |N_S| = O\big(|\mathcal{P}|\big) = O\left(\frac{d}{\epsilon} \cdot \varphi_{\mathcal{F}}\left(O\left(\frac{d}{\epsilon}\right), O(d)\right)\right).$$

This is too big.

(2) The next idea is to 'amortize' the size of the $\frac{1}{2}$-nets by first picking a random sample $R \subseteq X$.

For a fixed $S \in \mathcal{P}$ and $(S, \mathcal{F}|_S)$, Theorem 6.10 states that a random sample of S constructed by picking each point of S independently with probability

$$p = \Theta\left(\frac{1}{(1/2)\,|S|}\log\frac{1}{\gamma} + \frac{d}{(1/2)\,|S|}\log\frac{1}{(1/2)}\right)$$
$$= \Theta\left(\frac{1}{\epsilon n}\log\frac{1}{\gamma} + \frac{d}{\epsilon n}\right).$$

is a $\frac{1}{2}$-net of $\mathcal{F}|_S$ with probability at least $1-\gamma$.

Instead of sampling points separately from each $S \in \mathcal{P}$, we will construct a 'global' random sample R by picking each point of X independently with the above probability p.

For a fixed $S \in \mathcal{P}$, R fails to be a $\frac{1}{2}$-net of $\mathcal{F}|_S$ with probability at most γ. By the union bound, the probability that there exists a $S \in \mathcal{P}$ such that R fails to be a $\frac{1}{2}$-net of $\mathcal{F}|_S$ is at most $|\mathcal{P}| \cdot \gamma$. Setting $\gamma = \frac{1}{|\mathcal{P}|+1}$ implies that, with non-zero probability, R is a $\frac{1}{2}$-net for all $\mathcal{F}|_S$, $S \in \mathcal{P}$. Then

$$\begin{aligned}
\mathrm{E}\left[|R|\right] = np &= O\left(\frac{1}{\epsilon}\log\frac{1}{\gamma} + \frac{d}{\epsilon}\right) \\
&= O\left(\frac{1}{\epsilon}\log\left(\frac{d}{\epsilon}\cdot\varphi_{\mathcal{F}}\left(O\left(\frac{d}{\epsilon}\right), O\left(d\right)\right)\right) + \frac{d}{\epsilon}\right) \\
&= O\left(\frac{1}{\epsilon}\log\frac{d}{\epsilon}\right) + O\left(\frac{1}{\epsilon}\log\varphi_{\mathcal{F}}\left(O\left(\frac{d}{\epsilon}\right), O\left(d\right)\right)\right) + O\left(\frac{d}{\epsilon}\right).
\end{aligned}$$

The large additional term $O\left(\frac{1}{\epsilon}\log\frac{d}{\epsilon}\right)$ still remains.

(3) Since the expected number of sets in $\mathcal{P}$ for which R fails is at most $|\mathcal{P}| \cdot \gamma$, we had set $\gamma < \frac{1}{|\mathcal{P}|}$ to ensure the existence of a set R with no failures. Instead, we will set γ such that the expected number of $\mathcal{F}|_S$, $S \in \mathcal{P}$, for which R fails to be a $\frac{1}{2}$-net is $O\left(\frac{1}{\epsilon}\right)$. The key point is that for each of these failed sets of $\mathcal{P}$, we can afford to separately add a $O\left(d\right)$-sized $\frac{1}{2}$-net for a total of $O\left(\frac{d}{\epsilon}\right)$ additional points. In other words, we set γ so that $|\mathcal{P}| \cdot \gamma = \Theta\left(\frac{1}{\epsilon}\right)$. Then the size of the initial random sample R is

$$\begin{aligned}
O\left(\frac{1}{\epsilon}\log\frac{1}{\gamma} + \frac{d}{\epsilon}\right) &= O\left(\frac{1}{\epsilon}\log\left(\epsilon\,|\mathcal{P}|\right) + \frac{d}{\epsilon}\right) \\
&= O\left(\frac{1}{\epsilon}\log\varphi_{\mathcal{F}}\left(O\left(\frac{d}{\epsilon}\right), O\left(d\right)\right) + \frac{d}{\epsilon}\right),
\end{aligned}$$

as desired.

We now turn to the formal proof with complete calculations.

Proof of Theorem 6.2. For an integer $j \geq 0$, set

$$\epsilon_j = 2^j \cdot \epsilon \qquad \text{and} \qquad \delta_j = \frac{\epsilon_j n}{2}.$$

For each $j = 1, \ldots, \lceil \log \frac{1}{\epsilon} \rceil$, define

$$\mathcal{F}_j = \left\{ S \in \mathcal{F} \ : \ \epsilon_{j-1} n \ \leq \ |S| \ < \ \epsilon_j n \right\}.$$

Further let $\mathcal{P}_j$ be a *maximal* subset of $\mathcal{F}_j$ satisfying the property that

$$\text{for all } S, S' \in \mathcal{P}_j, \qquad |\Delta(S, S')| \geq \delta_j.$$

As each set in $\mathcal{F}_j$ has size less than $\epsilon_j n$, Theorem 5.10 implies that

$$|\mathcal{P}_j| \leq \frac{48\, dn}{\delta_j} \cdot \varphi_{\mathcal{F}} \left(\frac{8\, dn}{\delta_j}, \frac{24\, d \epsilon_j n}{\delta_j} \right) = \frac{96\, dn}{\epsilon_j n} \cdot \varphi_{\mathcal{F}} \left(\frac{16\, dn}{\epsilon_j n}, \frac{48\, d \epsilon_j n}{\epsilon_j n} \right)$$

$$= \frac{96\, d}{\epsilon_j} \cdot \varphi_{\mathcal{F}} \left(\frac{16\, d}{\epsilon_j}, 48\, d \right). \tag{6.12}$$

Claim 6.13. Let $j \in \left\{1, \ldots, \lceil \log \frac{1}{\epsilon} \rceil \right\}$. Suppose the set $N_j \subseteq X$ is a $\frac{1}{2}$-net *simultaneously* for all these set systems:

$$\left\{ \left(S, \mathcal{F}_j|_S\right) \ : \ S \in \mathcal{P}_j \right\}.$$

Then N_j hits all the sets of $\mathcal{F}_j$.

Proof. Let $F \in \mathcal{F}_j$. If $F \in \mathcal{P}_j$, then clearly it is hit by the $\frac{1}{2}$-net of $\mathcal{F}_j|_F$. Otherwise by the maximality of $\mathcal{P}_j$, there exists a $S \in \mathcal{P}_j$ such that

$$|\Delta(F, S)| = |F \setminus S| + |S \setminus F| < \delta_j \qquad \text{and hence} \qquad |F \setminus S| < \delta_j - |S \setminus F|.$$

As $|F| \geq 2^{j-1} \epsilon n = \delta_j$, it follows that

$$|F \cap S| = |F| - |F \setminus S| \geq \delta_j - \left(\delta_j - |S \setminus F|\right) = |S \setminus F|.$$

This implies that $|F \cap S| \geq \frac{|S|}{2}$ and as $F \cap S \in \mathcal{F}_j|_S$, F is hit by the $\frac{1}{2}$-net of $\mathcal{F}_j|_S$. $\square$

Thus it suffices to compute, for each $j = 1, \ldots, \lceil \log \frac{1}{\epsilon} \rceil$, a set N_j such that

$$N_j \ \text{ is a } \tfrac{1}{2}\text{-net of all the } |\mathcal{P}_j| \text{ set systems } (S, \mathcal{F}_j|_S),\ S \in \mathcal{P}_j.$$

We can then return $\bigcup_j N_j$ as an ϵ-net of $\mathcal{F}$. We will construct each N_j separately for each index $j \in \left\{1, \ldots, \lceil \log \frac{1}{\epsilon} \rceil \right\}$ by computing the following two sets R_j and M_j, and setting $N_j = R_j \cup M_j$.

Constructing R_j: Let R_j be a sample constructed by picking each point of X independently with probability

$$C_6 \cdot \left(\frac{d}{(1/2) \cdot \epsilon_{j-1} n} \ln \frac{1}{(1/2)} + \frac{1}{(1/2) \cdot \epsilon_{j-1} n} \ln \left(d \cdot \varphi_{\mathcal{F}} \left(\frac{16d}{\epsilon_j}, 48d \right) \right) \right).$$

where C_6 is the constant from Theorem 6.10. For each $S \in \mathcal{P}_j$, we have $|S| \in [\epsilon_{j-1} n, \epsilon_j n)$ and so Theorem 6.10 applied to

$$\text{the set system } \ (S, \mathcal{F}|_S) \quad \text{with} \quad \epsilon = \frac{1}{2}, \quad \gamma = \frac{1}{d \cdot \varphi_{\mathcal{F}} \left(\frac{16d}{\epsilon_j}, 48d \right)},$$

implies that R_j is a $\frac{1}{2}$-net of $\mathcal{F}|_S$ with probability at least $1 - \gamma$. Furthermore,

$$\mathrm{E}\left[|R_j|\right] = O \left(\frac{d}{\epsilon_j} + \frac{1}{\epsilon_j} \log \left(d \cdot \varphi_{\mathcal{F}} \left(\frac{16d}{\epsilon_j}, 48d \right) \right) \right).$$

Constructing M_j: Initialize $M_j = \emptyset$. For each $S \in \mathcal{P}_j$ for which R_j *fails* to be a $\frac{1}{2}$-net, construct a $\frac{1}{2}$-net of $(S, \mathcal{F}_j|_S)$ of size $O(d)$ (by Theorem 6.10) and add it to M_j. Then

$$\begin{aligned}
\mathrm{E}\left[|M_j|\right] &= \sum_{S \in \mathcal{P}_j} \Pr\left[R_j \text{ is not a } \frac{1}{2}\text{-net of } \mathcal{F}_j|_S\right] \cdot \left(\text{size of } \frac{1}{2}\text{-net of } \mathcal{F}_j|_S\right) \\
&\leq \sum_{S \in \mathcal{P}_j} \frac{1}{d \cdot \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon_j}, 48d\right)} \cdot O(d) = |\mathcal{P}_j| \cdot \frac{1}{d \cdot \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon_j}, 48d\right)} \cdot O(d) \\
&\leq \frac{96d}{\epsilon_j} \cdot \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon_j}, 48d\right) \cdot \frac{1}{d \cdot \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon_j}, 48d\right)} \cdot O(d) = O\left(\frac{d}{\epsilon_j}\right),
\end{aligned}$$

where the second-to-last step uses the upper bound on $|\mathcal{P}_j|$ given in Equation (6.12).

Thus we can conclude with the expected size of the final ϵ-net N of $\mathcal{F}$:

$$\begin{aligned}
\mathrm{E}[|N|] &= \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \mathrm{E}\left[|R_j| + |M_j|\right] = \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \mathrm{E}\left[|R_j|\right] + \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \mathrm{E}\left[|M_j|\right] \\
&= \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} O\left(\frac{d}{\epsilon_j} + \frac{1}{\epsilon_j} \log\left(d \cdot \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon_j}, 48d\right)\right)\right) + \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} O\left(\frac{d}{\epsilon_j}\right) \\
&= \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} O\left(\frac{d}{2^j \epsilon} + \frac{1}{2^j \epsilon} \log d + \frac{1}{2^j \epsilon} \log \varphi_{\mathcal{F}}\left(\frac{16d}{2^j \epsilon}, 48d\right)\right) + \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} O\left(\frac{d}{2^j \epsilon}\right) \\
&= O\left(\frac{d}{\epsilon}\right) + O\left(\frac{1}{\epsilon}\right) \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \frac{1}{2^j} \cdot \log \varphi_{\mathcal{F}}\left(\frac{16d}{2^j \epsilon}, 48d\right) \\
&\leq O\left(\frac{d}{\epsilon}\right) + O\left(\frac{1}{\epsilon}\right) \sum_{j=1}^{\lceil \log \frac{1}{\epsilon} \rceil} \frac{1}{2^j} \cdot \log \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon}, 48d\right) \\
&= O\left(\frac{d}{\epsilon} + \frac{1}{\epsilon} \log \varphi_{\mathcal{F}}\left(\frac{16d}{\epsilon}, 48d\right)\right).
\end{aligned}$$

□

Bibliography and discussion. The proof in this chapter is from [**MDG18**], building on ideas from several earlier works [**AES10**, **CF90**, **CV07**, **PR08**]. In particular, [**PR08**, **AES10**] are important papers that made progress on ϵ-nets for several basic geometric set systems. It was shown in [**KMP17**] that the bound on sizes of ϵ-nets as a function of their shallow-cell complexity is tight.

[AES10] B. Aronov, E. Ezra, and M. Sharir, *Small-size ϵ-nets for axis-parallel rectangles and boxes*, SIAM J. Comput. **39** (2010), no. 7, 3248–3282, DOI 10.1137/090762968. MR2678074

[CF90] B. Chazelle and J. Friedman, *A deterministic view of random sampling and its use in geometry*, Combinatorica **10** (1990), no. 3, 229–249, DOI 10.1007/BF02122778. MR1092541

[CV07] K. L. Clarkson and K. Varadarajan, *Improved approximation algorithms for geometric set cover*, Discrete Comput. Geom. **37** (2007), no. 1, 43–58, DOI 10.1007/s00454-006-1273-8. MR2279863

[KMP17] A. Kupavskii, N. H. Mustafa, and J. Pach, *Near-optimal lower bounds for ϵ-nets for half-spaces and low complexity set systems*, A journey through discrete mathematics, Springer, Cham, 2017, pp. 527–541. MR3726612

[MDG18] N. H. Mustafa, K. Dutta, and A. Ghosh, *A simple proof of optimal epsilon nets*, Combinatorica **38** (2018), no. 5, 1269–1277, DOI 10.1007/s00493-017-3564-5. MR3884789

[PR08] E. Pyrga and S. Ray, *New existence proofs for ϵ-nets*, Computational geometry (SCG'08), ACM, New York, 2008, pp. 199–207, DOI 10.1145/1377676.1377708. MR2504287