3.1 VC dimension

> *"Very little [of mathematics] is easily accessible. But I think a lot more of it can be explained so that a lot more people understand it. On the level we're talking about. I like to try to make mathematics easy, not to make it hard. I think there is a tendency among mathematicians to try to make it hard. I try to combat that when I see people wrap up their mathematics in formal fancy theories that make it less accessible."*
>
> William Thurston

Let P be a set of n points in the plane, and let $\mathcal{D}$ be the primal set system on P induced by disks; namely $S \in \mathcal{D}$ if and only if there exists a disk D such that $S = D \cap P$. A key fact regarding $\mathcal{D}$ is that $|\mathcal{D}| = O(n^3)$. This is a consequence of the fact that disks are 'fixed' by three points, namely that for any set Y of three points in the plane, there exists a disk D containing exactly the points of Y on its boundary. In fact, this is true in an 'hereditary' manner: for any set $P' \subseteq P$, the number of subsets of P' induced by disks is $O\left(|P'|^3\right)$. The key fact here is that this bound depends *polynomially* and *only* on the size of P'.

We now recast this property in purely combinatorial terms. Let $(X, \mathcal{F})$ be an abstract set system. The property stated above, of the number of sets induced on any subset of P' by disks, can be stated combinatorially as the number of subsets of X that can be obtained by intersection with sets of $\mathcal{F}$. Formally, define the *projection* of $\mathcal{F}$ onto any $Y \subseteq X$ as the set system

$$\mathcal{F}|_Y = \{Y \cap S \colon S \in \mathcal{F}\}.$$

Consider again the set system $\mathcal{D}$ induced by disks, and let Y be any subset of P. $\mathcal{D}|_Y$ consists of all subsets $Y' \subseteq Y$ which can be gotten by intersection of Y with a disk, regardless of the remaining points of $P \setminus Y$.

Just as the bound on the number of induced subsets by disks is a by-product of the fact that a disk is 'fixed' by three points, one can derive a bound on the size of the projection $\mathcal{F}|_Y$ for any $Y \subseteq X$ by assuming that the set system is of limited 'expressiveness' with respect to constant-sized subsets. It is not clear how to generalize the property of a disk being 'fixed' by three points to abstract set systems. The idea here is to note that a disk D passing through a set $Q = \{p, q, r\}$ of three points implies that, by slightly shifting D, one can obtain all subsets—$\{p\}, \{q\}, \{r\}, \{p, q\}, \{p, r\}, \{q, r\}$—via intersection with disks in the plane.

This is the property that will be abstracted:

> The *VC dimension* of a set system $(X, \mathcal{F})$, denoted by $\text{VC-dim}(\mathcal{F})$, is the size of the largest subset Y of X for which we have $|\mathcal{F}|_Y| = 2^{|Y|}$. In such case we say that Y is *shattered* by $\mathcal{F}$.

Here is the key statement, whose proof is given a bit later, that makes VC dimension such a useful parameter of set system complexity, and justifies the earlier analogy of such set systems with those induced by geometric objects.

Lemma 3.1 (Primal shatter lemma). *Given a set system $\mathcal{F}$ on X with* $\text{VC-dim}(\mathcal{F}) \le d$*, and any* $Y \subseteq X$*,*

$$\left|\mathcal{F}|_Y\right| \le \sum_{i=0}^{d} \binom{|Y|}{i} \le \left(\frac{e|Y|}{d}\right)^d = O\left(|Y|^d\right).$$

Note that the other direction is true as well—if for each $Y \subseteq X$ we have $|\mathcal{F}|_Y| = O\left(|Y|^d\right)$, then the VC dimension of $\mathcal{F}$ cannot be too large.

Lemma 3.2. *Given a set system* $(X, \mathcal{F})$ *and a constant* c *such that*

$$\textit{for any } Y \subseteq X, \textit{ we have} \quad |\mathcal{F}|_Y| \le \left(\frac{|Y|}{c}\right)^d,$$

then $\text{VC-dim}(\mathcal{F}) \le 2d \cdot \log \frac{d}{c}$.

Proof. Let $t = \text{VC-dim}(\mathcal{F})$ and $Y \subseteq X$ be any set realizing the VC-dimension of $\mathcal{F}$—i.e., $|Y| = t$ and $|\mathcal{F}|_Y| = 2^{|Y|}$. Then we have

$$2^t = \ |\mathcal{F}|_Y| \ \le \left(\frac{|Y|}{c}\right)^d = \left(\frac{t}{c}\right)^d \qquad \Longrightarrow \qquad t \le d \cdot \log \frac{t}{c}. \tag{3.1}$$

Our goal is to upper-bound $t = \text{VC-dim}(\mathcal{F})$; however there does not exist a closed-form bound for t in the above expression. Applying inequality (3.1) repeatedly, one gets

$$\begin{aligned} t &\le d \cdot \log \frac{t}{c} \\ &\le d \cdot \log \left(\frac{d \log \frac{t}{c}}{c}\right) = d \log \frac{d}{c} + d \log \log \frac{t}{c} \\ &\le d \log \frac{d}{c} + d \log \log \left(\frac{d \log \frac{t}{c}}{c}\right) = d \log \frac{d}{c} + d \log \log \frac{d}{c} + d \log \log \log \frac{t}{c}. \end{aligned}$$

Now we use the fact that the expression

$$\log x + \log \log x + \log \log \log x + \cdots$$

can be upper-bounded by a geometric series, and so is at most $2 \log x$. This implies that

$$t \le d \left(\log \frac{d}{c} + \log \log \frac{d}{c} + \log \log \log \frac{d}{c} + \cdots\right) \le 2d \cdot \log \frac{d}{c}.$$

□

It is not hard to see that most geometric set systems have small VC dimension. For example, the primal set system induced by half-spaces in $\mathbb{R}^d$ has VC dimension $d + 1$.

Lemma 3.3. *Let $\mathcal{H}$ be the family of all half-spaces in $\mathbb{R}^d$. Then we have* $\text{VC-dim}(\mathcal{H}) = d + 1$.

Proof. Clearly $\text{VC-dim}(\mathcal{H}) \geq d + 1$, as $(d + 1)$ points at the vertices of a simplex can be shattered by $\mathcal{H}$. On the other hand, apply Radon's lemma to any set P of $d + 2$ points in $\mathbb{R}^d$ to get a partition of P into P_1 and P_2 such that $\text{conv}(P_1)$ intersects $\text{conv}(P_2)$. Then P cannot be shattered, as P_1 and P_2 cannot be separated by half-spaces. □

Lemma 3.4. *Let $\mathcal{B}$ be the family of all balls in $\mathbb{R}^d$. Then we have* $\text{VC-dim}(\mathcal{B}) = d + 1$.

Proof. Assume that a set of points P in $\mathbb{R}^d$ is shattered by the primal set system induced by balls. Then for any $Q \subseteq P$, there exists a ball B with $Q = B \cap P$, and a ball B' with $P \setminus Q = B' \cap P$. Then the hyperplane passing through $B \cap B'$ (or simply separating B and B' if the two balls are disjoint) separates Q from $P \setminus Q$. Thus if a set of points are shattered by the primal set system induced by balls in $\mathbb{R}^d$, then they are shattered by the primal set system induced by half-spaces in $\mathbb{R}^d$, and we're done by the bound on VC-dimension for half-spaces. □

More generally, primal set systems induced by polynomial inequalities can be realized, using Veronese maps, by primal set systems induced by half-spaces in some higher dimension. Formally, identify each d-variate polynomial $f(x_1, \ldots, x_d)$ with its induced set $S_f = \left\{p \in \mathbb{R}^d : \; f(p) \geq 0\right\}$. Then Veronese maps imply the following.

Lemma 3.5. *Let $\mathcal{R}_{d,D}$ be a primal set system induced by all d-variable polynomials over $\mathbb{R}^d$ of degree at most D. Then* $\text{VC-dim}\left(\mathcal{R}_{d,D}\right) \leq \binom{d+D}{d}$.

We next present two basic theorems on set systems with bounded VC dimension.

Primal Shatter Lemma

We return to the proof of the lemma stated earlier:

Lemma 3.1 (Primal shatter lemma)**.** *Given a set system $\mathcal{F}$ on X with* $\text{VC-dim}(\mathcal{F}) \leq d$*, and any $Y \subseteq X$,*

$$\left|\mathcal{F}|_Y\right| \leq \sum_{i=0}^{d} \binom{|Y|}{i} \leq \left(\frac{e|Y|}{d}\right)^d = O\left(|Y|^d\right).$$

The proof that we present uses an operation on set systems, called *shifting*, which is applied repeatedly to a given set system to get a 'simpler' set system.

> Given a set system $\mathcal{F} = \{S_1, \ldots, S_m\}$ on the set X, and any element $a \in X$, one can derive another system $\mathcal{F}_a = \{S'_1, \ldots, S'_m\}$ from $\mathcal{F}$ by *shifting $\mathcal{F}$ with a*—by removing the element a from all sets as long as that does not create duplicate sets.

For each set $S_i \in \mathcal{F}$, the set S_i' derived by shifting with a is

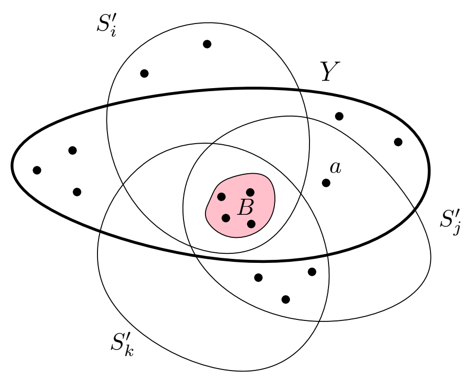
$$S_i' = \begin{cases} S_i & \text{if } S_i \setminus \{a\} \in \mathcal{F}, \\ S_i \setminus \{a\} & \text{otherwise.} \end{cases}$$

Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be the original set system and $\mathcal{F}_a = \{S_1', \ldots, S_m'\}$ be the set system shifted with $a \in X$. There are two key features of this operation. First, that $|\mathcal{F}| = |\mathcal{F}_a|$; this follows immediately from the definition of shifting. Second, shifting does not increase the VC dimension of the set system, as we prove now.

Lemma 3.6. *Given a set system* $(X, \mathcal{F})$ *and any* $a \in X$*, we have* $\text{VC-dim}(\mathcal{F}_a) \le \text{VC-dim}(\mathcal{F})$.

Proof. Fix a set $Y \subseteq X$ that is shattered by $\mathcal{F}_a$. We must show that then it is also shattered by $\mathcal{F}$. If $a \notin Y$, then the intersections $S_i \cap Y$ and $S_i' \cap Y$ are identical, and the statement follows immediately. Thus assume $a \in Y$.

Fix any set $B \subseteq Y$. We show that if B can be realized using a set of $\mathcal{F}_a$, then it can also be realized using a set of $\mathcal{F}$. So let $S_i' \in \mathcal{F}_a$ be such that $B = Y \cap S_i'$. We now exhibit a set $S_k \in \mathcal{F}$ such that $B = Y \cap S_k$.



There are two cases.

- $\mathbf{a} \in \mathbf{B}$. As $B = Y \cap S_i'$, $a \in S_i'$ and so the set S_i must not have been shifted. Then $S_i = S_i'$ and $B = Y \cap S_i$.

- $\mathbf{a} \notin \mathbf{B}$. Then $a \notin S_i'$, but now S_i need not be equal to S_i', as a could have been in S_i, in which case $Y \cap S_i$ would contain the additional element a. See the figure. Crucially, as Y is shattered by $\mathcal{F}_a$, there exists some other set S_j' such that $B \cup \{a\} = Y \cap S_j'$. Furthermore, the set $S_k = S_j' \setminus \{a\}$ *must* be in $\mathcal{F}$—otherwise we would have shifted S_j'. And so $B = Y \cap S_k$.

We should remark here that in the above argument, to show that any fixed B that can be realized using a set of $\mathcal{F}_a$ can also be realized using a set of $\mathcal{F}$, we needed to use the fact that Y was shattered by $\mathcal{F}_a$. Simply the fact that an individual B can be realized by a set of $\mathcal{F}_a$ is not sufficient. □

Proof of Primal Shatter Lemma. Repeatedly apply shifting on $\mathcal{F}|_Y$ with any element of Y; let $\mathcal{F}'|_Y$ be the resulting set system where shifting does not change any set. By Lemma 3.6, the VC dimension of $\mathcal{F}'|_Y$ is at most d.

Observe that $\mathcal{F}'|_Y$ is downwards closed: if $A \in \mathcal{F}'|_Y$ and $B \subseteq A$, then $B \in \mathcal{F}'|_Y$. Thus the largest cardinality of a set in $\mathcal{F}'|_Y$ is d. Now the proof follows by summing up the sets in $\mathcal{F}'|_Y$ by their sizes:

$$|\mathcal{F}|_Y| = |\mathcal{F}'|_Y| \leq \sum_{i=0}^{d} \binom{|Y|}{i}.$$

□

Unit distance graphs

We give another example of the shifting technique applied to set systems of bounded VC dimension. Define the symmetric difference of two sets S, S' to be

$$\Delta(S, S') = (S \setminus S') \bigcup (S' \setminus S).$$

Also define the *unit distance* graph on a set system as

> Given a set system $\mathcal{F} = \{S_1, \ldots, S_m\}$ on X, the unit distance graph $G_U(\mathcal{F}) = (\mathcal{F}, E_{\mathcal{F}})$ on the vertex set $\mathcal{F}$ has the set of edges $E_{\mathcal{F}}$ such that $\{S_i, S_j\} \in E_{\mathcal{F}}$ if and only if $|\Delta(S_i, S_j)| = 1$.

Lemma 3.7. *Given a set system $\mathcal{F} = \{S_1, \ldots, S_m\}$ on X, let $G_U(\mathcal{F}) = (\mathcal{F}, E_{\mathcal{F}})$ be its unit distance graph. If* $\text{VC-dim}(\mathcal{F}) \leq d$*, then* $|E_{\mathcal{F}}| \leq d \cdot |\mathcal{F}|$.

Proof. Repeatedly shift $\mathcal{F}$ with elements of X, and let $\mathcal{F}'$ be the resulting set system where shifting does not change any set. We will show that shifting does not decrease the number of edges in the unit distance graph. In particular,

1. $\mathcal{F}'$ is downwards closed,
2. $|\mathcal{F}| = |\mathcal{F}'|$,
3. $\text{VC-dim}(\mathcal{F}') \leq \text{VC-dim}(\mathcal{F})$, and
4. $|E_{\mathcal{F}'}| \geq |E_{\mathcal{F}}|$.

Assuming these four facts (the first three follow from the earlier proof), we can finish the proof. Charge each edge $e \in E_{\mathcal{F}'}$ to the bigger of the two sets of $\mathcal{F}'$ representing the two vertices of e. As $\mathcal{F}'$ is downwards closed, each set $S' \in \mathcal{F}'$ has precisely $|S'|$ edges in $E_{\mathcal{F}'}$ charged to it, namely all the edges between S' and each subset of S' of size $|S'| - 1$. As before, the largest set in $\mathcal{F}'$ has size at most d, and thus each set of $\mathcal{F}'$ gets charged at most d edges. Thus we have

$$|E_{\mathcal{F}'}| \leq d|\mathcal{F}'| = d|\mathcal{F}|.$$

Assuming fact 4., we can conclude that $|E_{\mathcal{F}}| \leq |E_{\mathcal{F}'}| \leq d|\mathcal{F}|$.

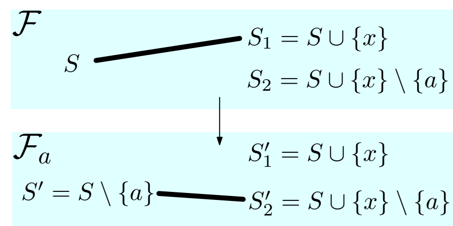
It remains to show $|E_{\mathcal{F}}| \leq |E_{\mathcal{F}_a}|$ for any $a \in X$, where $\mathcal{F}_a$ is the set system obtained by shifting $\mathcal{F}$ with a.

> Consider an edge $\{S, S_1\} \in E_{\mathcal{F}}$, defined by a set $S \in \mathcal{F}$ and another set $S_1 = S \cup \{x\}$, where $x \in X$.

How can this edge not be present in the unit distance graph of $\mathcal{F}_a$? If $a \notin S$, then clearly both sets S, S_1 are unaffected—either $x \neq a$ and then both S, S_1 are unaffected by shifting, or $x = a$, in which case S_1 will not be shifted. Thus the edge remains in $E_{\mathcal{F}_a}$.

Thus we can assume that $a \in S$. If the sets S and S_1 both get shifted by a, or both do not get shifted by a, the edge remains in $E_{\mathcal{F}_a}$. Thus precisely one of the sets gets shifted.

First assume S gets shifted to S' by a, while the shifted set S_1' of S_1 remains unchanged. This implies that the set $S_2 = S_1 \setminus \{a\}$ is already in $\mathcal{F}$; see figure. The edge between S' and S_1' is no longer present in $E_{\mathcal{F}_a}$; however, it is replaced by the new edge between S' and S_2'. Since for each pair $S \in \mathcal{F}$ and $x \in X$, we have replaced the edge $\{S, S \cup \{x\}\}$ in $E_{\mathcal{F}}$ with the edge $\{S', S \cup \{x\} \setminus \{a\}\}$ in $E_{\mathcal{F}_a}$, the replacements are all distinct.



The remaining case—when S does not get shifted while $S \cup \{x\}$ does—is similar. Then $S \setminus \{a\}$ already exists in $\mathcal{F}$ and thus in $\mathcal{F}_a$. Therefore the edge $\{S, S \cup \{x\}\}$ in $E_{\mathcal{F}}$ gets replaced by the edge $\{S \setminus \{a\}, S \cup \{x\} \setminus \{a\}\}$ in $E_{\mathcal{F}_a}$.

This concludes the proof. □

Bibliography and discussion. The use of the shifting technique to prove these theorems is folklore. See [1] for a survey on this technique.

[1] P. Frankl. The shifting technique in extremal set theory. *London Math. Soc. Lecture Note Ser.*, 123:81–110, 1987.

4.3 Combinatorial Set Systems

> *"[T]he main object of physical science is not the provision of pictures, but is the formulation of laws governing phenomena and the application of these laws to the discovery of new phenomena. If a picture exists, so much the better; but whether a picture exists or not is a matter of only secondary importance. In the case of atomic phenomena no picture can be expected to exist in the usual sense of the word "picture," by which is meant a model functioning essentially on classical lines. One may however extend the meaning of the word "picture" to include any way of looking at the fundamental laws which makes their self-consistency obvious. With this extension, one may gradually acquire a picture of atomic phenomena by becoming familiar with the laws of quantum theory."*
>
> Paul Dirac

Given a set P of points lying in the cube $[0,n]^d$ with the property that $d(p_i,p_j) \geq \delta$ for every pair of points $p_i, p_j \in P$, it is not hard to show that then $|P| = O\big((\frac{n}{\delta})^d\big)$: as the $|P|$ balls of radius $\frac{\delta}{2}$ centered at each $p \in P$ must be pairwise disjoint, a volume argument implies that

$$|P| \cdot \left(\frac{\delta}{2}\right)^d \leq (n+\delta)^d .$$

This is an example of a geometric packing argument, and this section deals with an important lemma which generalizes packing properties of geometric objects to that of abstract set systems.

The geometric notion of packing relies on an underlying notion of 'distance', and broadly the packing question concerns the number of geometric objects that can exist together while being pair-wise 'distant' from each other. In moving to purely combinatorial set systems, the notion of distance between points is replaced, in a natural way, by the cardinality of the set symmetric difference between sets.

Given two finite sets X, Y, the set symmetric difference of X and Y is defined to be

$$\Delta\left(X,Y\right) = \left(X \setminus Y\right) \cup \left(Y \setminus X\right).$$

The main result of this section is the following.

Lemma 4.3 (Packing Lemma)**.** *Let $\mathcal{F} = \{S_1,\ldots,S_m\}$ be a set system on a set X of n elements, and let $d, \delta > 0$ be integers such that* $\text{VC-dim}(\mathcal{F}) = d$ *and for every $1 \leq i < j \leq m$, we have $|\Delta(S_i,S_j)| \geq \delta$. Then*

$$|\mathcal{F}| \leq 2 \cdot \mathrm{E}\left[|\mathcal{F}|_A|\right],$$

where A is a subset of size $s = \frac{4dn}{\delta} - 1$ picked uniformly at random from X.

We first give an intuition for the proof. Consider a set system $\mathcal{F}$ on a set X of n elements, such that for every $S, S' \in \mathcal{F}$, we have $|\Delta(S, S')| \geq \delta$. Now construct a random sample A by picking each element of X independently with probability $\frac{1}{\delta}$, and consider the set system $\mathcal{F}|_A$. We would like to argue that each set of $\mathcal{F}$ maps to a distinct set of $\mathcal{F}|_A$, and so $|\mathcal{F}| = \mathrm{E}\left[|\mathcal{F}|_A|\right]$, as desired.

However, it could be that two sets of $\mathcal{F}$ map to the same set of $\mathcal{F}|_A$. This happens precisely when no element in the symmetric difference of the two sets is picked in A. As we picked each element independently with probability $\frac{1}{\delta}$, and the symmetric difference of every pair of sets is at least δ, in expectation we will pick at least one element from the symmetric difference of these two sets, and thus they will not be 'merged' in $\mathcal{F}|_A$.

This entire line of thinking was 'in expectation', and the task now is to convert this to a formal proof.

⋆ ⋆ ⋆

Proof of the Packing Lemma

Proof of Lemma 4.3. The proof is an application of the Clarkson-Shor technique.

For a random sample R of size $s = \Theta\left(\frac{n}{\delta}\right)$, we will count, in two ways, the number of pairs of sets in $\mathcal{F}$ that end up at unit symmetric difference in $\mathcal{F}|_R$. On one hand, an upper-bound will be provided by the fact that there are only $O\left(|\mathcal{F}|_R|\right)$ pairs of sets at unit distance symmetric difference in $\mathcal{F}|_R$[†]. On the other hand, for every pair of sets $S_i, S_j \in \mathcal{F}$, there is some positive probability of the pair $\{S_i \cap R, S_j \cap R\}$ ending up as a unit symmetric difference pair in $G_U\left(\mathcal{F}|_R\right)$, and so, in expectation, a large proportion of such pairs end up in $\mathcal{F}|_R$. Putting these bounds together implies an upper-bound on $|\mathcal{F}|$.

However, there are two technical issues that must be overcome:

- We intend to count the number of pairs of sets in $\mathcal{F}$ that are at unit symmetric difference. However, while a set $S \in \mathcal{F}$ maps to the set $S \cap R \in \mathcal{F}|_R$, each set of $\mathcal{F}|_R$ potentially corresponds to many sets of $\mathcal{F}$. Therefore one has to look at the *weighted* unit-distance graph on the sets of $\mathcal{F}|_R$. First define

  for $S' \in \mathcal{F}|_R$, the weight $w(S')$ is the number of sets of $\mathcal{F}$ mapping to S'.

  Then instead of counting the number of edges in the unit distance graph $G_U(\mathcal{F}|_R)$, we would like to add up the weights of edges, where each $\{S'_i, S'_j\} \in E$ is assigned

[†]Recall the unit-distance graph, denoted by $G_U(\mathcal{F}|_R)$, on the vertex set $\mathcal{F}|_R$: two sets of $\mathcal{F}|_R$ are connected if and only if their symmetric difference has size precisely one. Lemma 3.7 gives a linear bound on the number of edges in such a graph.

the weight $w(\{S'_i, S'_j\}) = w(S'_i) \cdot w(S'_j)$. In fact, a technical trick we will use here is to instead define

$$w\left(\left\{S'_i, S'_j\right\}\right) = \min\left\{w\left(S'_i\right), w\left(S'_j\right)\right\}.$$

This will simplify the calculations without significant change—in any case, for any $a, b > 0$, we have[†]

$$\min\{a, b\} \cdot \frac{(a+b)}{2} \leq ab \leq \min\{a, b\} \cdot (a+b). \tag{4.3}$$

- We need to compute, for any two sets $S_i, S_j \in \mathcal{F}$, the probability that $\left\{S_i \cap R, S_j \cap R\right\}$ ends up as a unit-distance pair in $G_U(\mathcal{F}|_R)$. This happens if and only if exactly one element of $\Delta(S_i, S_j)$ is picked into R; in other words, this depends only on $|\Delta(S_i, S_j)|$. This is a difficult computation.

  However, given that we are only interested in the sum of these $\binom{|\mathcal{F}|}{2}$ probabilities—namely the expected number of pairs at unit-distance in $G_U(\mathcal{F}|_R)$—a clever idea is to further use double-counting to count this sum in a uniform way. Rather than summing up over pairs of sets in $\mathcal{F}$, we count, for the i-th element of R, the expected number of pairs at unit distance due to *that* element. By symmetry, this value is the same for all elements of R. So conditioned on the first $|R| - 1$ elements of R, we need to compute the expected number of pairs of sets that are put at unit distance by the $|R|$-th random element. This is an easier computation.

Having covered the main ideas, it remains to do the precise calculations.

⋆ ⋆ ⋆

Pick a random set R of size $s = \frac{4dn}{\delta}$ from X (without repetitions). Let $G_U(\mathcal{F}|_R) = (\mathcal{F}|_R, E_R)$ be the unit symmetric distance graph on $\mathcal{F}|_R$. For each $S' \in \mathcal{F}|_R$, define $w(S')$ to be the number of sets of $\mathcal{F}$ mapping to S':

$$w(S') = \left|\{S \in \mathcal{F} \colon S \cap R = S'\}\right|.$$

Define the weight of an edge $\left\{S'_i, S'_j\right\} \in E_R$ as

$$w\left(\left\{S'_i, S'_j\right\}\right) = \min\left\{w(S'_i), w(S'_j)\right\}.$$

Let $W = \sum_{e \in E_R} w(e)$ be the total weight of all the edges. As outlined earlier, we will count $\mathrm{E}[W]$ in two ways. Recall that $m = |\mathcal{F}|$.

First, the upper-bound on W.

[†]Assume $a \leq b$. Then it is equivalent to the fact that $\frac{a}{2} + \frac{b}{2} \leq b \leq a + b$.

Claim 4.4.

$$W \leq 2d \cdot m.$$

Proof. By Lemma 3.7, $|E_R| \leq d \cdot |\mathcal{F}|_R|$. Thus there exists a vertex $S' \in \mathcal{F}|_R$ of degree at most $2d$[†]. By the definition of the edge weights, the weight of each edge incident to S' is at most $w(S')$. Thus the sum of the weight of all the edges adjacent to S' is at most $2d \cdot w(S')$.

Remove S' from $G_U(\mathcal{F}|_R)$. The remaining graph is still a unit distance graph on $|\mathcal{F}|_R| - 1$ vertices; thus by applying Lemma 3.7 again, it has at most $d \cdot (|\mathcal{F}|_R| - 1)$ edges and so we can inductively bound the weight of edges in the remaining graph. Thus the total weight of edges can be upper-bounded as

$$2d \sum_{S' \in F|_R} w(S') = 2d \cdot m.$$

□

Next, the lower-bound on W.

Claim 4.5.

$$\mathrm{E}\left[W\right] \geq 4dm - 4d\,\mathrm{E}\left[|\mathcal{F}|_R|\right].$$

Proof. Imagine picking R by first choosing randomly a set A of $s - 1 = \frac{4dn}{\delta} - 1$ elements, and then choosing the last element uniformly from $X \setminus R$. Let W_1 be the weight of the edges in $G_U(\mathcal{F}|_R)$ where the element is the symmetric difference. By symmetry, we have $\mathrm{E}[W] = s \cdot \mathrm{E}[W_1]$.

To compute $\mathrm{E}[W_1]$, assume we have picked the first $s - 1$ vertices, say the set Y. In fact, we will show something even stronger: *regardless* of the choice of the first $s - 1$ vertices, we will lower-bound the expected weight due to the symmetric difference being the last random element picked. So, conditioned on any fixed choice of the first $s - 1$ vertices, we show the following.

Claim 4.6.

$$\mathrm{E}\left[W_1 \mid A = Y\right] \geq \frac{\delta}{n}\Big(m - |\mathcal{F}|_Y|\Big).$$

Proof. Note that the expectation here is only over the choice of the last element.

$\mathcal{F}|_Y$ contains the projected sets of $\mathcal{F}$ after having picked the first $s - 1$ elements as the set Y and projected our set system $\mathcal{F}$ into Y.

Next we pick a random element a from $X \setminus Y$ and set $R = Y \cup \{a\}$. Then each set in $\mathcal{F}|_Y$ could be split into two sets—projections of those sets of $\mathcal{F}$ that contained a, and those that

[†] $\sum_{S' \in \mathcal{F}|_R} \deg(S') = 2|E_R| \leq 2d|\mathcal{F}|_R|$, which implies that there exists a $S' \in \mathcal{F}|_R$ with $\deg(S') \leq \frac{2d|\mathcal{F}|_R|}{|\mathcal{F}|_R|}$.

did not. These two projected sets are now at unit distance apart in $\mathcal{F}|_R$ (with the element a being their symmetric difference), and it is their total weight W_1 that we have to bound.

Consider a set $Q \in \mathcal{F}|_Y$, and let $\mathcal{F}_Q$ be the sets of $\mathcal{F}$ mapping to Q—namely their projection onto A is Q. Let $b = |\mathcal{F}_Q|$.

Once the choice of a has been made, Q will be split into two sets, those sets containing that choice of a—say there are b_1 of these, and those sets not containing a, say $b_2 = b - b_1$ in number. The weight of the edge between these two sets will be $\min\{b_1, b_2\}$. We next compute this expected weight.

For each pair of fixed sets in $\mathcal{F}_Q$, the probability that the randomly chosen last element a will cause their symmetric difference in R to be 1 is at least $\frac{\delta}{n-(s-1)} \geq \frac{\delta}{n}$. Therefore the expected contribution of each pair of sets in $\mathcal{F}_Q$ to $b_1 b_2$ is at least $\frac{\delta}{n}$. Noting that $b = b_1 + b_2$ is fixed independent of the choice of a, summing up over all pairs of sets in $\mathcal{F}_Q$, we can lower-bound the expected contribution of the sets in $\mathcal{F}_Q$ to W_1 by

$$
\begin{aligned}
\mathrm{E}\left[\min\{b_1, b_2\}\right] &\geq \mathrm{E}\left[\frac{b_1 b_2}{b_1 + b_2}\right] \qquad \text{(by inequality 4.3)} \\
&= \frac{\mathrm{E}[b_1 b_2]}{b_1 + b_2} \qquad \text{(as } b = b_1 + b_2 \text{ is a constant)} \\
&= \frac{\sum_{S,S' \in \mathcal{F}_Q} \Pr\left[S \text{ and } S' \text{ differ on } a\right]}{b} \\
&\geq \frac{\sum_{S,S' \in \mathcal{F}_Q} \delta/n}{b} \\
&= \frac{|\mathcal{F}_Q|\left(|\mathcal{F}_Q| - 1\right) \cdot \delta/n}{|\mathcal{F}_Q|} = \frac{\delta}{n} \cdot \left(|\mathcal{F}_Q| - 1\right).
\end{aligned}
$$

Summing up over all sets of $\mathcal{F}|_Y$,

$$
\mathrm{E}\left[W_1 \mid A = Y\right] \geq \sum_{Q \in \mathcal{F}|_Y} \frac{\delta}{n}\Big(|\mathcal{F}_Q| - 1\Big) = \frac{\delta}{n}\left(\sum_{Q \in \mathcal{F}|_Y} |\mathcal{F}_Q| - \sum_{Q \in \mathcal{F}|_Y} 1\right) = \frac{\delta}{n}\Big(m - |\mathcal{F}|_Y|\Big).
$$

□

Finally we compute a lower-bound for $\mathrm{E}[W]$:

$$
\begin{aligned}
\mathrm{E}[W] = s \cdot \mathrm{E}[W_1] &= s \cdot \sum_{\substack{Y \subseteq X \\ |Y| = s-1}} \mathrm{E}\left[W_1 | A = Y\right] \cdot \Pr[A = Y] \\
&\geq s \cdot \sum_{\substack{Y \subseteq X \\ |Y| = s-1}} \frac{\delta}{n}\Big(m - |\mathcal{F}|_Y|\Big) \cdot \Pr\left[A = Y\right]
\end{aligned}
$$

$$= \frac{s\delta}{n} \left(m \sum_{\substack{Y \subseteq X \\ |Y| = s-1}} \Pr\left[A = Y\right] - \sum_{\substack{Y \subseteq X \\ |Y| = s-1}} |\mathcal{F}|_Y| \cdot \Pr\left[A = Y\right] \right)$$

$$= \frac{s\delta}{n} \left(m - \mathrm{E}\left[|\mathcal{F}|_A| \right] \right) = 4dm - 4d \, \mathrm{E}\left[|\mathcal{F}|_A| \right],$$

where the last equality follows from $s = \frac{4dn}{\delta}$. □

Putting the upper- and lower- bounds on $\mathrm{E}[W]$, we get

$$2dm \geq 4dm - 4d \, \mathrm{E}\left[|\mathcal{F}|_A| \right], \quad \text{implying that} \quad m \leq 2 \, \mathrm{E}\left[|\mathcal{F}|_A| \right].$$

This finishes the proof of Lemma 4.3. □

Bibliography and discussion. The packing lemma is from Haussler [1], who gave the proof for the specific case of set systems with bounded VC dimension. The more general bound in terms of projection sizes is from Mustafa [2].

[1] D. Haussler. Sphere packing numbers for subsets of the boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69(2):217–232, 1995.

[2] N. H. Mustafa. A simple proof of the shallow packing lemma. *Discrete & Computational Geometry*, 55(3):739–743, 2016.