

SECTION 4

Packing and Covering in Euclidean Spaces

The maximal inequality from Theorem 3.6 will be useful only if we have suitable bounds for the packing numbers of the set \mathcal{F} . This section presents a method for finding such bounds, based on a geometric property that transforms calculation of packing numbers into a combinatorial exercise.

The combinatorial approach generalizes the concept of a Vapnik-Červonenkis class of sets. It identifies certain subsets of \mathbb{R}^n that behave somewhat like compact sets of lower dimension; the bounds on the packing numbers grow geometrically, at a rate determined by the lower dimension. For comparison's sake, let us first establish the bound for genuinely lower dimensional sets.

(4.1) LEMMA. *Let \mathcal{F} be a subset of a V dimensional affine subspace of \mathbb{R}^n . If \mathcal{F} has finite diameter R , then*

$$D(\epsilon, \mathcal{F}) \leq \left(\frac{3R}{\epsilon}\right)^V \quad \text{for } 0 < \epsilon \leq R.$$

PROOF. Because Euclidean distances are invariant under rotation, we may identify \mathcal{F} with a subset of \mathbb{R}^V for the purposes of calculating the packing number $D(\epsilon, \mathcal{F})$. Let $\mathbf{f}_1, \dots, \mathbf{f}_m$ be points in \mathcal{F} with $|\mathbf{f}_i - \mathbf{f}_j| > \epsilon$ for $i \neq j$. Let B_i be the (V -dimensional) ball of radius $\epsilon/2$ and center \mathbf{f}_i . These m balls are disjoint; they occupy a total volume of $m(\epsilon/2)^V \Gamma$, where Γ denotes the volume of a unit ball in \mathbb{R}^V . Each \mathbf{f}_i lies within a distance R of \mathbf{f}_1 ; each B_i lies inside a ball of radius $3/2R$ and center \mathbf{f}_1 , a ball of volume $(3/2R)^V \Gamma$. It follows that $m \leq (3R/\epsilon)^V$. \square

A set of dimension V looks thin in \mathbb{R}^n . Even if projected down onto a subspace of \mathbb{R}^n it will still look thin, if the subspace has dimension greater than V . One way to capture this idea, and thereby create a more general notion of a set being thin, is to think of how much of the space around any particular point can be occupied by

the set. The formal concept involves the collection of 2^k orthants about each point \mathbf{t} in \mathbb{R}^k defined by means of all possible combinations of coordinatewise inequalities.

(4.2) DEFINITION. For each \mathbf{t} in \mathbb{R}^k and each subset J of $\{1, \dots, k\}$, define the J^{th} orthant about \mathbf{t} to consist of all those \mathbf{x} in \mathbb{R}^k for which

$$\begin{aligned} x_i &> t_i && \text{if } i \in J, \\ x_i &< t_i && \text{if } i \in J^c. \end{aligned}$$

A subset of \mathbb{R}^k will be said to *occupy* the J^{th} orthant of \mathbf{t} if it contains at least one point in that orthant. A subset will be said to *surround* \mathbf{t} if it occupies all 2^k of the orthants defined by \mathbf{t} .

There is a surprising connection between the packing numbers of a set in \mathbb{R}^n and the maximum number of orthants its lower dimensional projections can occupy. The projections that we use will differ slightly from the usual notion. For each k -tuple $I = (i(1), \dots, i(k))$ of integers from the range $1, \dots, n$, call $(x_{i(1)}, \dots, x_{i(k)})$ the I -*projection* of the point (x_1, \dots, x_n) in \mathbb{R}^n , even if the integers $i(1), \dots, i(k)$ are not all distinct. Call such a map into \mathbb{R}^k a k -*dimensional coordinate projection*. If all the integers are distinct, call it a *proper coordinate projection*.

(4.3) DEFINITION. Say that a subset \mathcal{F} of \mathbb{R}^n has a *pseudodimension* of at most V if, for every point \mathbf{t} in \mathbb{R}^{V+1} , no proper coordinate projection of \mathcal{F} can surround \mathbf{t} .

The concept of pseudodimension bears careful examination. It requires a property for all possible choices of $I = (i(1), \dots, i(V+1))$ from the range $1, \dots, n$. For each such choice and for each \mathbf{t} in \mathbb{R}^{V+1} , one must extract a J from I such that no \mathbf{f} in \mathcal{F} can satisfy the inequalities

$$\begin{aligned} f_i &> t_i && \text{for } i \in J, \\ f_i &< t_i && \text{for } i \in I \setminus J. \end{aligned}$$

Clearly any duplication amongst the elements of I will make this task a triviality. Only for distinct integers $i(1), \dots, i(V+1)$ must one expend energy to establish impossibility. That is why only proper projections need be considered.

If a set \mathcal{F} actually sits within an affine space of dimension V then it has pseudodimension at most V . To see this, notice that a $(V+1)$ -dimensional projection of such an \mathcal{F} must be a subset of an affine subspace \mathcal{A} with dimension strictly less than $V+1$. There exists a nontrivial vector $\boldsymbol{\beta}$ in \mathbb{R}^{V+1} and a constant γ such that $\boldsymbol{\beta} \cdot \boldsymbol{\alpha} = \gamma$ for every $\boldsymbol{\alpha}$ in \mathcal{A} . We may assume that $\beta_i > 0$ for at least one i . If \mathbf{t} has $\boldsymbol{\beta} \cdot \mathbf{t} \leq \gamma$ it is impossible to find an $\boldsymbol{\alpha}$ in \mathcal{A} such that

$$\begin{aligned} \alpha_i &< t_i && \text{when } \beta_i > 0, \\ \alpha_i &\geq t_i && \text{when } \beta_i \leq 0, \end{aligned}$$

for these inequalities would lead to the contradiction $\gamma = \sum_i \beta_i \alpha_i < \sum_i \beta_i t_i \leq \gamma$. If $\boldsymbol{\beta} \cdot \mathbf{t} > \gamma$ we would interchange the roles of “ $\beta_i > 0$ ” and “ $\beta_i \leq 0$ ” to reach a similar

contradiction. For the pseudodimension calculation we need the contradiction only for $\alpha_i > t_i$, but to establish the next result we need it for $\alpha_i \geq t_i$ as well.

(4.4) LEMMA. *Suppose the coordinates of the points in \mathcal{F} can take only two values, c_0 and c_1 . Suppose also that there is a V -dimensional vector subspace Λ of \mathbb{R}^n with the property: for each $f \in \mathcal{F}$ there is a $\lambda \in \Lambda$ such that $f_i = c_1$ if and only if $\lambda_i \geq 0$. Then \mathcal{F} has pseudodimension at most V .*

PROOF. We may assume that $c_0 = 0$ and $c_1 = 1$. Suppose that some proper I -projection of \mathcal{F} surrounds a point \mathbf{t} in \mathbb{R}^{V+1} . Each coordinate t_i must lie strictly between 0 and 1. The inequalities required for the projection of \mathbf{f} to occupy the orthant corresponding to a subset J of I are

$$\begin{aligned} f_i &= 1 && \text{for } i \in J, \\ f_i &= 0 && \text{for } i \in I \setminus J. \end{aligned}$$

That is,

$$\begin{aligned} \lambda_i &\geq 0 && \text{for } i \in J, \\ \lambda_i &< 0 && \text{for } i \in I \setminus J. \end{aligned}$$

As shown above, there exists a J such that this system of inequalities cannot be satisfied. \square

The connection between pseudodimension and packing numbers is most easily expressed if we calculate the packing numbers not for the usual Euclidean, or ℓ_2 , distance on \mathbb{R}^n , but rather for the ℓ_1 distance that corresponds to the norm

$$\|\mathbf{x}\|_1 = \sum_{i \leq n} |x_i|.$$

To distinguish between the two metrics on \mathbb{R}^n let us add subscripts to our notation, writing $D_1(\epsilon, \mathcal{F})$ for the ℓ_1 packing number of the set \mathcal{F} , and so on. [Notice that the ℓ_1 norm is not invariant under rotation. The invariance argument used in the proof of Lemma 4.1 would be invalid for ℓ_1 packing numbers.]

A set in \mathbb{R}^n of the form $\prod_i [\alpha_i, \beta_i]$ is called a *box*. It has ℓ_1 diameter $\sum_i (\beta_i - \alpha_i)$. The smallest integer greater than a real number x is denoted by $\lceil x \rceil$.

(4.5) LEMMA. *Let \mathcal{F} lie within a box of ℓ_1 diameter one in \mathbb{R}^n . If \mathcal{F} contains m points, each pair separated by an ℓ_1 distance of at least ϵ , then: for $k = \lceil 2\epsilon^{-1} \log m \rceil$, there exists a point \mathbf{t} in \mathbb{R}^k and a k -dimensional coordinate projection of \mathcal{F} that occupies at least m orthants of \mathbf{t} .*

PROOF. We may assume that the box has the form $\prod_i [0, p_i]$, where the p_i are nonnegative numbers summing to one. Partition $[0, 1]$ into subintervals I_1, \dots, I_n of lengths p_1, \dots, p_n . Generate $i(1), \dots, i(k)$ and $\mathbf{t} = (t_1, \dots, t_k)$ randomly, from a set of independent Uniform $[0, 1]$ random variables U_1, \dots, U_k , in the following way. If U_α lands in the subinterval I_j , put $i(\alpha)$ equal to j and t_α equal to the

distance of U_α from the left endpoint of I_j . [That is, the method chooses edge i with probability p_i , then chooses t_i uniformly from the $[0, p_i]$ interval.]

Let \mathcal{F}_0 be the subset of \mathcal{F} consisting of the m points with the stated separation property. To each \mathbf{f} in \mathcal{F}_0 there corresponds a set of n points in $[0, 1]$: the j^{th} lies in I_j , at a distance f_j from the left endpoint of that interval. The $2n$ points defined in this way by each pair \mathbf{f}, \mathbf{g} from \mathcal{F}_0 form n subintervals of $[0, 1]$, one in each I_j . The total length of the subintervals equals $|\mathbf{f} - \mathbf{g}|_1$, which is greater than ϵ , by assumption. If U_α lands within the interior of the subintervals, the coordinates $f_{i(\alpha)}$ and $g_{i(\alpha)}$ will be on opposite sides of t_α ; the projections of \mathbf{f} and \mathbf{g} will then lie in different orthants of \mathbf{t} . Each U_α has probability at most $1 - \epsilon$ of failing to separate \mathbf{f} and \mathbf{g} in this way. Therefore the projections have probability at most $(1 - \epsilon)^k$ of lying in the same orthant of \mathbf{t} .

Amongst the $\binom{m}{2}$ possible pairs from \mathcal{F}_0 , the probability that at least one pair of projections will occupy the same orthant of \mathbf{t} is less than

$$\binom{m}{2}(1 - \epsilon)^k < \frac{1}{2} \exp(2 \log m - k\epsilon).$$

The value of k was chosen to make this probability strictly less than one. With positive probability the procedure will generate $i(1), \dots, i(k)$ and \mathbf{t} with the desired properties. \square

Notice that the value of k does not depend on n , the dimension of the space \mathbb{R}^n in which the set \mathcal{F} is embedded.

The next result relates the occupation of a large number of orthants in \mathbb{R}^k to the property that some lower-dimensional projection of the set completely surrounds some point. This will lead to a checkable criterion for an \mathcal{F} in \mathbb{R}^n to have packing numbers that increase at the same sort of geometric rate as for the low-dimensional set in Lemma 4.1. The result is a thinly disguised form of the celebrated Vapnik-Červonenkis lemma.

(4.6) LEMMA. *A coordinate projection into \mathbb{R}^k of a set with pseudodimension at most V can occupy at most*

$$\binom{k}{0} + \binom{k}{1} + \dots + \binom{k}{V}$$

orthants about any point of \mathbb{R}^k .

PROOF. Let \mathcal{H} be a set with pseudodimension at most V . Its projection into \mathbb{R}^k also has pseudodimension at most V . So without loss of generality we may assume that $\mathcal{H} \subseteq \mathbb{R}^k$. Let \mathcal{S} denote the set of all k -tuples $\sigma = (\sigma_1, \dots, \sigma_k)$ with $\sigma_i = \pm 1$ for each i . Identify the 2^k orthants of \mathbf{t} with the 2^k vectors in \mathcal{S} . The orthants of \mathbf{t} that are occupied by \mathcal{H} correspond to a subset \mathcal{A} of \mathcal{S} . Suppose $\#\mathcal{A}$ is strictly greater than the asserted bound, and then argue for a contradiction.

The vectors in \mathcal{S} also index the proper coordinate projections on \mathbb{R}^k . Let us denote by π_σ the projection that discards all those coordinates for which $\sigma_i = -1$. The orthants of $\pi_\sigma \mathbf{t}$ correspond to vectors η in \mathcal{S} with $\eta \leq \sigma$: we merely ignore those

coordinates η_i for which $\sigma_i = -1$, then identify the orthants by means of the signs of the remaining η_i . For the projection $\pi_\sigma \mathcal{H}$ to occupy the orthant corresponding to η , there must exist an α in \mathcal{A} such that $\alpha_i = \eta_i$ whenever $\sigma_i = +1$; that is, $\alpha \wedge \sigma = \eta$.

Let \mathcal{S}_V denote the set of all vectors σ in \mathcal{S} with $\sigma_i = +1$ for at least $V + 1$ coordinates. The assumption of pseudodimension at most V means that $\pi_\sigma \mathcal{H}$ does not surround $\pi_\sigma t$, for every σ in \mathcal{S}_V . Thus for each σ in \mathcal{S}_V there exists an $\eta(\sigma) \leq \sigma$ such that $\alpha \wedge \sigma \neq \eta(\sigma)$ for every α in \mathcal{A} . For definiteness define $\eta(\sigma) = \sigma$ for $\sigma \notin \mathcal{S}_V$.

Invoke the Basic Combinatorial Lemma from Section 1 to obtain a one-to-one map θ from \mathcal{S} onto itself such that $\theta(\sigma) \wedge \sigma = \eta(\sigma)$ for every σ . The assumption about the size of \mathcal{A} ensures that

$$\#\{\theta^{-1}(\alpha) : \alpha \in \mathcal{A}\} + \#\mathcal{S}_V > 2^k,$$

which implies that there exists an α in \mathcal{A} for which $\theta^{-1}(\alpha) \in \mathcal{S}_V$. But then, for that α , we have

$$\alpha \wedge \theta^{-1}(\alpha) \neq \eta(\theta^{-1}(\alpha)) = \theta(\theta^{-1}(\alpha)) \wedge \theta^{-1}(\alpha),$$

a contradiction that establishes the assertion of the lemma. \square

The V in the statement of the last lemma plays almost the same role as the dimension V in Lemma 4.1, which gave the $O(\epsilon^{-V})$ bound on packing numbers. By combining the assertions of the last two lemmas we obtain the corresponding bound in terms of the pseudodimension.

(4.7) THEOREM. *Let \mathcal{F} have pseudodimension at most V and lie within a box of ℓ_1 diameter one in \mathbb{R}^n . Then there exist constants A and W , depending only on V , such that*

$$D_1(\epsilon, \mathcal{F}) \leq A(1/\epsilon)^W \quad \text{for } 0 < \epsilon \leq 1.$$

PROOF. Fix $0 < \epsilon \leq 1$. Let $m = D_1(\epsilon, \mathcal{F})$. Choose $k = \lceil 2\epsilon^{-1} \log m \rceil$ as in Lemma 4.5. From Lemma 4.6,

$$\binom{k}{0} + \cdots + \binom{k}{V} \geq m.$$

The left-hand side of this inequality is a polynomial of degree V in k ; it is smaller than $(1+V)k^V$. [There is not much to be gained at this stage by a more precise upper bound.] Thus

$$(1+V) \left(\frac{1+2\log m}{\epsilon} \right)^V \geq m,$$

whence

$$\frac{(1+V)}{\epsilon^V} \geq \frac{m}{(1+2\log m)^V}.$$

For some positive constant C depending on V , the right-hand side is greater than $C\sqrt{m}$, for all positive integers m . The asserted inequality holds if we take $A = (1+V)^2/C^2$ and $W = 2V$. \square

For the sake of comparison with Lemma 4.1, let us see what sort of bound is given by Theorem 4.7 when \mathcal{F} is contained within a V -dimensional affine subspace of \mathbb{R}^n . If \mathcal{F} also lies within an ℓ_1 box of diameter one, the argument from the proof of Theorem 4.7 gives packing numbers that grow as $O(\epsilon^{-W})$, for $W = 2V$. We could reduce W to any constant slightly larger than V . [Use $Cm^{1-\delta}$, for some tiny positive δ , instead of $C\sqrt{m}$, in the proof.] This falls just slightly short of the $O(\epsilon^{-V})$ bound from Lemma 4.1.

Theorem 4.7 has a slightly more general version that exploits an invariance property of orthants. For each vector $\alpha = (\alpha_1, \dots, \alpha_n)$ of nonnegative constants, and each \mathbf{f} in \mathbb{R}^n , define the pointwise product $\alpha \odot \mathbf{f}$ to be the vector in \mathbb{R}^n with i^{th} coordinate $\alpha_i f_i$. Write $\alpha \odot \mathcal{F}$ to denote the set of all vectors $\alpha \odot \mathbf{f}$ with \mathbf{f} in \mathcal{F} . At least when $\alpha_i > 0$ for every i , a trivial, but significant, property of orthants is: \mathcal{F} occupies orthant J of \mathbf{t} if and only if $\alpha \odot \mathcal{F}$ occupies orthant J of $\alpha \odot \mathbf{t}$. Similarly, if some coordinate projection of \mathcal{F} cannot surround a point \mathbf{t} then the corresponding coordinate projection of $\alpha \odot \mathcal{F}$ cannot surround $\alpha \odot \mathbf{t}$. The key requirement of the theorem is unaffected by such coordinate rescalings. We can rescale any bounded set \mathcal{F} with an envelope \mathbf{F} —that is, a vector such that $|f_i| \leq F_i$ for each $\mathbf{f} \in \mathcal{F}$ and each i —to lie within a box of ℓ_1 diameter one, and then invoke the theorem.

(4.8) THEOREM. *Let \mathcal{F} be a bounded subset of \mathbb{R}^n with envelope \mathbf{F} and pseudo-dimension at most V . Then there exist constants A and W , depending only on V , such that*

$$D_1(\epsilon|\alpha \odot \mathbf{F}|_1, \alpha \odot \mathcal{F}) \leq A(1/\epsilon)^W \quad \text{for } 0 < \epsilon \leq 1,$$

for every rescaling vector α of non-negative constants.

PROOF. We may assume $\alpha_i > 0$ for every i . (The cases where some α_i are zero correspond to an initial projection of \mathcal{F} into a lower dimensional coordinate subspace.) Apply Theorem 4.7 to the rescaled set \mathcal{F}^* consisting of vectors \mathbf{f}^* with coordinates

$$f_i^* = \frac{\alpha_i f_i}{2 \sum_j \alpha_j F_j}.$$

Then observe that, for vectors in \mathcal{F}^* ,

$$|\mathbf{f}^* - \mathbf{g}^*|_1 > \epsilon/2 \quad \text{if and only if} \quad |\alpha \odot \mathbf{f} - \alpha \odot \mathbf{g}|_1 > \epsilon|\alpha \odot \mathbf{F}|_1.$$

Absorb the extra factor of 2^W into the constant A . \square

Sets with an $O(\epsilon^{-W})$ bound on packing numbers arise in many problems, as will become apparent in the sections on applications. The main role of the pseudo-dimension of a set \mathcal{F} will be to provide such a geometric rate of growth for packing numbers of \mathcal{F} . It also applies to *any subclass of \mathcal{F} under its natural envelope*. For subclasses with small natural envelopes, this method sometimes leads to bounds unattainable by other methods.

The added generality of an inequality that holds uniformly over all rescaling vectors allows us to move back and forth between ℓ_1 and ℓ_2 packing numbers. The bounds from Theorem 4.8 will translate into bounds on ℓ_2 packing numbers suitable for the chaining arguments in the Section 3.

(4.9) LEMMA. For each bounded \mathcal{F} with envelope \mathbf{F} , and each $\epsilon > 0$,

$$D_2(\epsilon, \mathcal{F}) \leq D_1(\frac{1}{2}\epsilon^2, \mathbf{F} \odot \mathcal{F}) \leq D_2(\frac{1}{2}\epsilon^2, |\mathbf{F}|_2, \mathcal{F}).$$

PROOF. For each pair of vectors \mathbf{f}, \mathbf{g} in \mathcal{F} ,

$$|\mathbf{f} - \mathbf{g}|_2^2 \leq 2|\mathbf{F} \odot \mathbf{f} - \mathbf{F} \odot \mathbf{g}|_1 \leq 2|\mathbf{F}|_2|\mathbf{f} - \mathbf{g}|_2.$$

The first inequality follows from the bound $(f_i - g_i)^2 \leq 2F_i|f_i - g_i|$; the second follows from the Cauchy-Schwarz inequality. \square

(4.10) COROLLARY. If \mathcal{F} is a bounded subset of \mathbb{R}^n with envelope \mathbf{F} and pseudo-dimension at most V , then there exist constants A_2 and W_2 , depending only on V , such that

$$D_2(\epsilon|\alpha \odot \mathbf{F}|_2, \alpha \odot \mathcal{F}) \leq A_2 (1/\epsilon)^{W_2} \quad \text{for } 0 < \epsilon \leq 1$$

and every rescaling vector α of non-negative constants.

PROOF. The set $\alpha \odot \mathcal{F}$ has envelope $\beta = \alpha \odot \mathbf{F}$. Because $\beta \odot \alpha \odot \mathcal{F}$ has envelope $\beta \odot \beta$ and $|\beta|_2^2 = |\beta \odot \beta|_1$, the ℓ_2 packing number is bounded by

$$D_1(\frac{1}{2}\epsilon^2|\beta \odot \beta|_1, \beta \odot \alpha \odot \mathcal{F}) \leq A(\frac{1}{2}\epsilon^2)^{-W},$$

with A and W from Theorem 4.8. \square

The presence of an arbitrary rescaling vector in the bound also gives us added flexibility when we deal with sets that are constructed from simpler pieces, as will be explained in the next section.

REMARKS. My definition of pseudodimension abstracts the concept of a Vapnik-Červonenkis subgraph class of functions, in the sense of Dudley (1987). Most of the results in the section are reformulations or straightforward extensions of known theory for Vapnik-Červonenkis classes, as expounded in Chapter II of Pollard (1984), for example. See that book for a listing of who first did what when.

The nuisance of improper coordinate projections was made necessary by my desire to break the standard argument into several steps. The arguments could be rewritten using only proper projections, by recombining Lemma 4.5 and Theorem 4.7. The proof of Lemma 4.6 is a novel rearrangement of old ideas: see the comments at the end of Section 1 regarding the Basic Combinatorial Lemma.