Notes for seminar on covering numbers

These are notes for a few seminar talks on covering numbers delivered in winter 2010, on March 8, 15, 22, and April 18.

1 Covering numbers

Given sets $A, B \subseteq \mathbb{R}^n$, we define the *covering number of* A *by* B as

$$N(A, B) = \inf \{\operatorname{card}(T) \colon A \subseteq T + B\} ,$$

where card denotes the counting measure. Recalling that $T + B = \bigcup_{t \in T}(t + B)$, we see that $N(A, B)$ is the smallest number of translates of B that suffices to cover A. (By the well-ordering principle, if it is possible to cover A with finitely many translates of B, then the infimum in the definition is attained.)

Covering numbers satisfy a multiplicative triangle inequality:

$$N(A, C) \leq N(A, B)N(B, C) . \tag{1}$$

The idea is to construct a covering of A by translates of C as follows: first cover A by translates of B, then cover each translate of B by translates of C. The formal proof: Suppose $A \subseteq S + B$ and $B \subseteq T + C$. Then $A \subseteq S + (T + C) = (S + T) + C$, and so $N(A, C) \leq \operatorname{card}(S + T) \leq \operatorname{card}(S \times T) = \operatorname{card}(S) \operatorname{card}(T)$. Optimizing S and T yields (1). □

Adding the sets of translation vectors also yields

$$N(A + B, C + D) \leq N(A, C)N(B, D) . \tag{2}$$

Similar methods yield the following obvious statements:

- If $A \subseteq A'$ then $N(A, B) \leq N(A', B)$. (Larger sets are harder to cover.)
- If $B \subseteq B'$ then $N(A, B) \geq N(A, B')$. (Larger sets are better at covering.)
- $\frac{1}{2}(N(A, B) + N(A', B)) \leq N(A, B) \vee N(A', B) \leq N(A \cup A', B) \leq N(A, B) + N(A', B)$. (Here $\vee$ denotes maximum.)
- $N(A + t, B) = N(A, B)$ for any $t \in \mathbb{R}^n$.
- $N(A, B + t) = N(A, B)$ for any $t \in \mathbb{R}^n$.
- $N(TA, TB) \leq N(A, B)$ for any linear map $T \in L(\mathbb{R}^n)$, with equality when T is invertible. (A case of some interest to us is when T is a projection onto a lower-dimensional subspace.)

2 Covering from the inside

For some purposes it is desirable to have a cover in which the reference points of the translates lie inside the set being covered. We accordingly define the covering number *from inside* of A by B by

$$\overline{N}(A, B) = \inf \{\text{card}(T) \colon T \subseteq A \subseteq T + B\} \,.$$

This number does not satisfy the multiplicative triangle inequality, but it does satisfy an inequality like (2):

$$\overline{N}(A + B, C + D) \leq \overline{N}(A, C)\overline{N}(B, D) \,, \tag{3}$$

and using the same proof.

Obviously $N(A, B) \leq \overline{N}(A, B)$. For an inequality of the reverse type, we want to take a cover of A by translates of B and construct a cover from points inside A.

Suppose $A \subseteq T + B$. For each $t \in T$, the translate $t + B$ covers a certain portion of A, namely $A \cap (t + B)$. We will choose some point s in this portion of A and make s responsible for covering this portion of A. We cannot be sure that $s + B$ will cover $A \cap (t + B)$, so we should cover by some other set instead of B. If $a \in A \cap (t+B)$. we will need $a - s$ to be in the set being used to cover A; since $a \in t+B$ and $s \in t+B$, we have $a - s \in B - B$; thus $A \cap (t+B) \subseteq s+B-B$. Choosing one point s for each point t yields

$$\overline{N}(A, B - B) \leq N(A, B) \,.$$

In short, we can assume the translation vectors are in A, at the cost of passing from covering by B to covering by $B - B$.

In fact, this argument gives the slightly stronger result

$$\overline{N}(A, B - B) \leq \overline{N}(A, (A - A) \cap (B - B)) \leq N(A, B) \,. \tag{4}$$

For example, we can use this fact to pass from coverings to coverings in sections. Assume for simplicity that K and L are symmetric and convex. If E is a subspace of $\mathbb{R}^n$, then (4) yields

$$\overline{N}(K \cap E, 2(K \cap L \cap E)) \leq N(K \cap E, L) \leq N(K, L) \,.$$

Sasha pointed out that if B is (a multiple of) the Euclidean ball B_2^n, and A is convex, then in fact we lose nothing by moving the centres into A:

$$\overline{N}(K, B_2^n) = N(K, B_2^n) \qquad \text{if } K \text{ convex.} \tag{5}$$

The idea is to move the centre of each translate $t + B_2^n$ to the closest point in K; writing $p(t)$ for that closest point, one can show that $K \cap (t + B_2^n) \subseteq K \cap (p(t) + B_2^n)$.

3 Separated points

Let $A, B \subseteq \mathbb{R}^n$. Let us define

$$A \text{ is separated by } B \iff (\forall x, y \in A\colon x \neq y \implies (x + B) \cap (y + B) = \varnothing) . \quad (6)$$

Thus, such a set A consists of translation vectors for a pairwise disjoint collection of translates of B.

The condition of translates being disjoint is closely related to the Minkowski difference of two sets, since

$$(x + A) \cap (y + B) \neq \varnothing \iff x - y \in B - A . \quad (7)$$

Proof ($\Rightarrow$) If $(x+A)\cap(y+B) \neq \varnothing$ then there exists $z \in \mathbb{R}^n$ such that $z \in x+A$ and $z \in y+B$, that is, $z-x \in A$ and $z-y \in B$, so $x-y = (z-y)-(z-x) \in B-A$. ($\Leftarrow$) If $x - y \in B - A$ then there exist $a \in A$ and $b \in B$ such that $x - y = b - a$, that is, $x + a = y + b$. Since $x + a \in x + A$ and $y + b \in y + B$, we are done. □

The condition of one set being separated by another can thus be restated in terms of difference sets:

$$A \text{ is separated by } B \iff (A - A) \cap (B - B) \subseteq \{0\} \quad (8)$$

(If A and B are non-empty, we actually have $=$, not just $\subseteq$.)

Proof A is separated by B

$$\begin{aligned} &\iff (\forall x, y \in A\colon x \neq y \implies (x + B) \cap (y + B) = \varnothing) \\ &\iff (\forall x, y \in A\colon (x + B) \cap (y + B) \neq \varnothing \implies x = y) \\ &\iff (\forall x, y \in A\colon x - y \in B - B \implies x = y) && \text{(by (7))} \\ &\iff (\forall z \in A - A\colon z \in B - B \implies z = 0) && (z := x - y) \\ &\iff (A - A) \cap (B - B) \subseteq \{0\} \end{aligned}$$

□

From (8) we see that the condition of being separated by B depends not on B but only on $B - B$. Thus, for example, A is separated by B iff A is separated by $-B$, showing that

$$M(A, B) = M(A, -B) . \quad (9)$$

A curious consequence of (8) is that A is separated by B iff B is separated by A. We don't usually use this because we usually take A to be a discrete, finite set of translation vectors and B to be the unit ball of a norm (or at least, a convex body), so the roles of A and B are quite different.

It is also interesting (but perhaps not important) that the set $(A-A)\cap(B-B)$ occurs both in (8) and in (4).

We can pass between coverings and separated points by the following inequalities for $M(A, B)$, which resemble (4):

$$\overline{N}(A, B - B) \leq M(A, B) \leq N(A, B) . \tag{10}$$

The left-hand inequality requires that $B \neq \varnothing$.

Proof Upper inequality: By (9), it suffices to show that $M(A, B) \leq N(A, -B)$. For this, note the obvious fact that

$$x \in t + B \iff t \in x - B .$$

Consequently,

$$x \in (s + B) \cap (t + B) \iff \{s, t\} \subseteq x - B .$$

That is, the intersection of two translates of B is exactly the set of reference points for which a translate of $-B$ covers the two given reference points. So if $T \subseteq A$ and T is separated by B, so that the translates $t + B$ (for $t \in T$) are pairwise disjoint, then no two points in T can be covered by the same translate of $-B$; therefore any cover of T (and thus any cover of A) by translates of $-B$ needs at least $\operatorname{card}(T)$ points. Thus $M(A, B) \leq N(A, -B)$.

Lower inequality: Let $T \subseteq A$ be a set separated by B and maximal for these conditions. (If no such set exists, then $M(A, B) = \infty$ and we are done.) By maximality, we have

$$\begin{aligned}
&(\forall a \in A \setminus T \colon \exists t \in T \colon (a + B) \cap (t + B) \neq \varnothing) \\
\iff &(\forall a \in A \setminus T \colon \exists t \in T \colon a - t \in B - B) && \text{(by (7))} \\
\iff &(\forall a \in A \setminus T \colon \exists t \in T \colon a \in t + B - B) \\
\iff &A \setminus T \subseteq T + B - B
\end{aligned}$$

Since $B \neq \varnothing$, we have $0 \in B - B$, whence $T \subseteq T + B - B$. Thus $A = (A \setminus T) \cup T \subseteq T + B - B$, showing that $\overline{N}(A, B - B) \leq \operatorname{card}(T)$. □

We can get the difference set $B - B$ on the right-hand side of (10) as well, if B is convex. For this, first note that

$$\begin{aligned}
\tfrac{1}{2}(B - B) - \tfrac{1}{2}(B - B) &= \tfrac{1}{2}B - \tfrac{1}{2}B - \tfrac{1}{2}B + \tfrac{1}{2}B \\
&= \tfrac{1}{2}(B + B) - \tfrac{1}{2}(B + B) \\
&\supseteq B - B
\end{aligned}$$

with equality when B is convex. Thus

$$(A - A) \cap (\tfrac{1}{2}(B - B) - \tfrac{1}{2}(B - B)) \supseteq (A - A) \cap (B - B) ,$$

with equality when B is convex. By (8), it follows that

$$A \text{ is separated by } \tfrac{1}{2}(B - B) \implies A \text{ is separated by } B\ ,$$

and the converse holds if B is convex. Thus

$$M(A, \tfrac{1}{2}(B - B)) \leq M(A, B)\ , \quad \text{with equality when B is convex.} \tag{11}$$

Combining the equality situation with (10) yields

$$N(A, B - B) \leq M(A, B) \leq N(A, \tfrac{1}{2}(B - B)) \quad \text{if B is convex.} \tag{12}$$

4 Volume estimates

In this section, assume all relevant sets are measurable and of positive, finite measure.

The most basic volume estimate for covering numbers is given by noting that if $A \subseteq T + B$, then

$$\operatorname{vol}(A) \leq \operatorname{vol}(T + B) = \operatorname{vol}\Big(\bigcup_{t \in T}(t + B)\Big) \leq \sum_{t \in T} \operatorname{vol}(t + B) = \operatorname{card}(T)\operatorname{vol}(B)\ .$$

This shows that

$$\frac{\operatorname{vol}(A)}{\operatorname{vol}(B)} \leq N(A, B)\ . \tag{13}$$

For $M(A, B)$, we have

$$\frac{\operatorname{vol}(A)}{\operatorname{vol}(B - B)} \leq M(A, B) \leq \frac{\operatorname{vol}(A + B)}{\operatorname{vol}(B)}\ . \tag{14}$$

Proof The upper estimate follows by noting that if $T \subseteq A$ and T is separated by B, then the translates $t + B$ for $t \in T$ are disjoint, so that

$$\operatorname{card}(T)\operatorname{vol}(B) = \operatorname{vol}(T + B) \leq \operatorname{vol}(A + B)\ .$$

The lower estimate follows by (13) and (10). □

Examples Conspicuously missing in (13) is an upper estimate for $N(A, B)$ in terms of volumes. To illustrate some techniques for obtaining such estimates, we consider $N(K, \epsilon K)$, where K is a convex body and $0 < \epsilon \leq 1$.

1. First, the lower bound. By (13), we have

$$N(K, \epsilon K) \geq \frac{\operatorname{vol}(K)}{\operatorname{vol}(\epsilon K)} = \left(\frac{1}{\epsilon}\right)^n\ .$$

2. If K is symmetric, then $\epsilon K = \frac{\epsilon}{2}K - \frac{\epsilon}{2}K$, so we can use the upper estimate on M:

$$\begin{aligned} N(K, \epsilon K) &= N(K, \tfrac{\epsilon}{2}K - \tfrac{\epsilon}{2}K) \\ &\le M(K, \tfrac{\epsilon}{2}K) && \text{(by (10))} \\ &\le \frac{\operatorname{vol}(K + \frac{\epsilon}{2}K)}{\operatorname{vol}(\frac{\epsilon}{2}K)} && \text{(by (14))} \\ &= \frac{(1+\frac{\epsilon}{2})^n}{(\frac{\epsilon}{2})^n} \\ &= \left(1 + \frac{2}{\epsilon}\right)^n \\ &\le \left(\frac{3}{\epsilon}\right)^n && \text{(since } \epsilon \le 1\text{)} \end{aligned}$$

3. If K is not symmetric, we can replace ϵK with something symmetric and smaller, and then use the same technique as in the previous item. Thus

$$\begin{aligned} N(K, \epsilon K) \le N(K, \epsilon(K \cap -K)) &\le \frac{\operatorname{vol}(K + \frac{\epsilon}{2}(K \cap -K))}{\operatorname{vol}(\frac{\epsilon}{2}(K \cap -K))} \\ &\le \frac{\operatorname{vol}(K + \frac{\epsilon}{2}K)}{\operatorname{vol}(\frac{\epsilon}{2}(K \cap -K))} \le \left(1 + \frac{2}{\epsilon}\right)^n \frac{\operatorname{vol}(K)}{\operatorname{vol}(K \cap -K)} \,. \end{aligned}$$

Nothing in this argument so far uses the choice of origin; so now we choose the origin to minimize $\operatorname{vol}(K)/\operatorname{vol}(K \cap -K)$. From [10] we know that if we choose the origin randomly inside K then this ratio has expected value 2^n; from [5] we know that if we take the centroid of K as the origin then this value is at most 2^n. Thus we get

$$N(K, \epsilon K) \le \left(2 + \frac{4}{\epsilon}\right)^n \le \left(\frac{6}{\epsilon}\right)^n \,.$$

4. We can also use the Rogers–Zong lemma [9], which asserts that if K and L are convex bodies then

$$N(K, L) \le \theta(L) \frac{\operatorname{vol}(K - L)}{\operatorname{vol}(L)} \,,$$

where $\theta(L)$ is the "covering density" of L. (If we cover $\mathbb{R}^n$ by translates of L in the most economical way, the density of the resulting covering is $\theta(L)$; if L tiles $\mathbb{R}^n$ then $\theta(L) = 1$; if the covering has some overlap between translates of L then $\theta(L) > 1$.) Furthermore, from [7] we have

$$\theta(L) \le 7n \log n \,.$$

(See the next section for a little more information about these results.) Finally, the Rogers–Shephard inequality [8] asserts that

$$\operatorname{vol}(K - L) \operatorname{vol}(K \cap L) \leq \binom{2n}{n} \operatorname{vol}(K) \operatorname{vol}(L) .$$

Putting all this together yields

$$\begin{aligned} N(K, \epsilon K) \leq \theta(\epsilon K) \frac{\operatorname{vol}(K - \epsilon K)}{\operatorname{vol}(\epsilon K)} &\leq \theta(K) \binom{2n}{n} \frac{\operatorname{vol}(K) \operatorname{vol}(\epsilon K)}{\operatorname{vol}(\epsilon K) \operatorname{vol}(K \cap \epsilon K)} \\ &\leq \theta(K) \binom{2n}{n} \left(\frac{1}{\epsilon}\right)^n \leq 7n \log n \left(\frac{4}{\epsilon}\right)^n \end{aligned}$$

5. The Rogers–Zong lemma also applies when K is symmetric; a computation similar to the above (but slightly simpler) yields

$$N(K, \epsilon K) \leq 7n \log n \left(\frac{2}{\epsilon}\right)^n$$

in this case.

Volume estimates can also be obtained using measures other than Lebesgue measure. An argument by Talagrand is of this type, and shows

$$M(B_2^n, K) \leq 2e^{2\ell(K)^2} \quad \text{if } K \text{ is a symmetric convex body,} \tag{15}$$

where $\ell(K)$ is the gaussian average of the K-norm, that is,

$$\ell(K) = \int_{\mathbb{R}^n} \|x\|_K \, d\gamma_n(x) .$$

Now, $\ell(K) \sim \sqrt{n} M(K)$ (see [11], p.24), so combining (15) with (10) yields

$$\overline{N}(B_2^n, K) \leq 2e^{cnM(K)^2} . \tag{16}$$

This is known as the dual Sudakov inequality. (See [6]. The proof given below is a later one by Talagrand; see [3], pp. 82–83. For a slightly more general version, see [4], Lemma 4.)

The argument for (15) (given below) requires that K be symmetric, but symmetry is not essential in (16); indeed,

$$\overline{N}(B_2^n, K) \leq \overline{N}(B_2^n, K \cap -K) \leq 2e^{cnM(K \cap -K)^2} ,$$

and

$$\begin{aligned} M(K \cap -K) = \int_{S^{n-1}} \|\theta\|_{K \cap -K} \, d\sigma(\theta) &= \int_{S^{n-1}} (\|\theta\|_K \vee \|\theta\|_{-K}) \, d\sigma(\theta) \\ &\leq \int_{S^{n-1}} (\|\theta\|_K + \|\theta\|_{-K}) \, d\sigma(\theta) = 2M(K) , \end{aligned}$$

so for asymmetric bodies we obtain (16) with a slightly worse constant.

Now to prove (15).

Proof Let $r > 0$ be chosen later. Let $T \subseteq rB_2^n$ be such that $\{t + rK \colon t \in T\}$ are pairwise disjoint and $\operatorname{card}(T) \le M(rB_2^n, rK) = M(B_2^n, K)$. Then

$$\begin{aligned}
1 &\ge \gamma_n(T + rK) \\
&= \sum_{t\in T} \gamma_n(t + rK) \\
&= \sum_{t\in T} \int_{t+rK} e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}} \\
&= \sum_{t\in T} \int_{rK} e^{-|x+t|^2/2} \frac{dx}{(2\pi)^{n/2}} \\
&= \sum_{t\in T} e^{-|t|^2/2} \int_{rK} e^{-\langle x,t\rangle} e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}} \\
&= \sum_{t\in T} e^{-|t|^2/2} \int_{rK} e^{-\langle x,t\rangle}\, d\gamma_n(x) \\
&= \sum_{t\in T} e^{-|t|^2/2} \gamma_n(rK) \mathbb{E} e^{-\langle X,t\rangle}
\end{aligned}$$

(where X is a standard gaussian random variable truncated to rK)

$$\begin{aligned}
&\ge \sum_{t\in T} e^{-|t|^2/2} \gamma_n(rK) e^{-\langle \mathbb{E}X,t\rangle} && \text{(Jensen's inequality)} \\
&= \sum_{t\in T} e^{-|t|^2/2} \gamma_n(rK) e^{-\langle 0,t\rangle} && \text{(X is symmetric)} \\
&= \sum_{t\in T} e^{-|t|^2/2} \gamma_n(rK) \\
&\ge \operatorname{card}(T) e^{-r^2/2} \gamma_n(rK) && \text{(since } T \subseteq rB_2^n\text{)}
\end{aligned}$$

Finally, taking $r = 2\ell(K)$, we have $\gamma_n(rK) \ge \frac{1}{2}$ by Markov's inequality, and so

$$\operatorname{card}(T) \le \frac{e^{r^2/2}}{\gamma_n(rK)} \le 2e^{2\ell(K)^2} .$$

□

5 Covering density

Here we sketch the proofs of the Rogers–Zong lemma [9] (which estimates covering numbers using covering density) and the 1957 result of Rogers [7] (which estimates covering density). I omit all details, and even some crucial definitions, such as that of covering density itself. Intuitively, $\theta(K)$ is the density of the most economical covering of $\mathbb{R}^n$ by translates of K. (For precise definitions, see [7], or [1], §1.1.) What matters for us is that if T is the set of reference points for the most economical covering $\mathbb{R}^n = T + K$, then for any set A, the translates of A contain on average (in some sense) $\theta(K)\operatorname{vol}(A)/\operatorname{vol}(K)$ points of T.

If K and L are convex bodies in $\mathbb{R}^n$ then

$$N(K, L) \le \theta(L)\frac{\operatorname{vol}(K - L)}{\operatorname{vol}(L)} . \tag{17}$$

Proof sketch Let $\mathbb{R}^n = S + L$ be a covering of $\mathbb{R}^n$ by translates of L, with density $\theta(L)$. Then, for any $t \in \mathbb{R}^n$, $t + K \subseteq S + L$. But we do not need all the translates given by S to cover a particular $t + K$; the translates that matter are:

$$\begin{aligned} s + L \text{ matters} &\iff (s + L) \cap (t + K) \ne \varnothing \\ &\iff s - t \in K - L \\ &\iff s \in t + K - L \end{aligned}$$

Thus we can cover $t + K$ with $\operatorname{card}(S \cap (t + K - L))$ translates of L. We wish to choose t to make $\operatorname{card}(S \cap (t + K - L))$ small; as discussed a moment ago, on average (in some sense) this quantity is $\theta(L)\operatorname{vol}(K - L)/\operatorname{vol}(L)$. □

If K is a convex body in $\mathbb{R}^n$, $n \ge 3$, then

$$\theta(K) \le n \log n + n \log\log n + 5n . \tag{18}$$

Proof sketch Tile $\mathbb{R}^n$ with large cubes, say of side length R, and cover each cube as follows. First choose N translates of K randomly (N large, chosen later), obtaining $T_1 + K$, $\operatorname{card}(T_1) = N$. (These translates cover at least $1 - e^{-N/R^n}$ of the cube, if we assume wlog that $\operatorname{vol}(K) = 1$.) Then pack the uncovered space with translates of $-\delta K$ (δ small, chosen later), obtaining $T_2 - \delta K$, $\operatorname{card}(T) = M$. (Bound M by a volume argument.) By the maximality of the packing, all points in the cube are either close to a copy of K from T_1 or a copy of $-\delta K$ from T_2. It can then be shown that $(T_1 \cup T_2) + (1 + \delta)K$ covers the cube, with density $(1 + \delta)^n(N + M)/R^n$. A suitable choice of parameters δ, N, M, R yields the desired estimate. □

See chapter 1 of [1] for comprehensive information about covering density.

6 Small sections from covering by the ball

Here we show that if a convex body is small in the sense that it can be covered by not too many copies of B_2^n, then it is small in the sense that it has

k-dimensional sections that are contained in balls that are not too large. More precisely:

$$\begin{aligned}&\text{If } K \text{ is a convex body in } \mathbb{R}^n\\&\text{and } A \in \mathbb{R} \text{ is such that } N(K, B_2^n)^{1/n} \le A,\\&\text{then for any } k \in [1..n),\\&\text{there exists } E \in G_{n,k}\\&\text{such that } K \cap E \subseteq (24A)^{1/(1-\frac{k}{n})}(B_2^n \cap E).\end{aligned} \tag{19}$$

(Here $G_{n,k}$ denotes the set of k-dimensional subspaces of $\mathbb{R}^n$, and $[1..n) = \mathbb{Z} \cap [1, n)$.) Note that the size of the sections in the conclusion depends only on the covering parameter A and on the proportion $\frac{k}{n}$ between the dimension of the section and the dimension of the whole space.

The statement (19) closely resembles the volume ratio theorem (see [11], §3); the constants are different, and the hypothesis involves covering numbers instead of volumes, but otherwise they are the same. So we suspect that we could prove (19) by transforming the hypothesis $N(K, B_2^n)^{1/n} \le A$ into a statement comparing $\text{vol}(K)$ and $\text{vol}(B_2^n)$ (presumably using the estimates of §4) and then applying the volume ratio theorem. The only difficulty with this plan is that the volume ratio theorem requires $K \supseteq B_2^n$ (so that $\|\cdot\|_K$ is Lipschitz, a fact which plays an essential role in the proof), and we do not have this in (19). The maneuver which avoids this problem is to replace K with $K + B_2^n$.

Proof Let K and A be as in (19). Then

$$\frac{\text{vol}(K + B_2^n)}{2^n \,\text{vol}(B_2^n)} \le N(K + B_2^n, 2B_2^n) \le N(K, B_2^n) N(B_2^n, B_2^n) \le A^n .$$

Thus $B_2^n \subseteq K + B_2^n$ and $(\text{vol}(K + B_2^n)/\text{vol}(B_2^n))^{1/n} \le 2A$. The volume ratio theorem then yields sections such as described in (19). □

In fact we can deduce the volume ratio theorem from (19) as well (again, except for the exact constants): if $K \supseteq B_2^n$ and $(\text{vol}(K)/\text{vol}(B_2^n))^{1/n} \le A$ then

$$\begin{aligned}N(K, B_2^n) &\le M(K, \tfrac{1}{2}B_2^n) && (B_2^n \text{ is symmetric and convex})\\&\le \frac{\text{vol}(K + \frac{1}{2}B_2^n)}{\text{vol}(\frac{1}{2}B_2^n)}\\&\le \frac{\text{vol}(K + \frac{1}{2}K)}{\text{vol}(\frac{1}{2}B_2^n)} && (\text{since } B_2^n \subseteq K)\\&= 3^n \frac{\text{vol}(K)}{\text{vol}(B_2^n)}\\&\le (3A)^n\end{aligned}$$

Applying (19) then yields sections as desired.

We now give a direct proof of (19), using ideas similar to those in the proof of the volume ratio theorem. We will in fact show that subspaces such as described not merely exist, but that a random subspace has this property with high probability.

Proof Let $r > 0$ be chosen later.

(The idea: If r is small enough, then $rK \cap S^{n-1}$ has small measure, since convex bodies are spiky. A random subspace E will miss $rK \cap S^{n-1}$, and so $rK \cap E$ does not meet S^{n-1}, whence $rK \cap E \subseteq B_2^n \cap E$.)

Fix $E_0 \in G_{n,k}$. Choose $Q \in O(n)$ randomly according to the Haar probability measure. Then

$$\begin{aligned}
&\mathbb{P}(QE_0 \text{ meets } rK \cap S^{n-1}) \\
&= \mathbb{P}(\exists \theta \in S^{n-1} \cap E_0 \colon Q\theta \in rK) \\
&\le \mathbb{P}(\exists \theta \in \Lambda\colon \operatorname{dist}(Q\theta, rK) < r) && (\text{let } \Lambda \text{ be an } r\text{-net for } S^{n-1} \cap E_0) \\
&= \mathbb{P}(\exists \theta \in \Lambda\colon Q\theta \in \widetilde{K}) && (\text{set } \widetilde{K} = (rK + rB_2^n) \cap S^{n-1}) \\
&\le \operatorname{card}(\Lambda)\sigma(\widetilde{K})
\end{aligned}$$

where, as usual, σ denotes the uniform probability measure on S^{n-1}.

To show existence, then, we want $\operatorname{card}(\Lambda)\sigma(\widetilde{K}) < 1$, while to show high probability, we want $\operatorname{card}(\Lambda)\sigma(\widetilde{K}) \ll 1$.

We can choose Λ such that

$$\operatorname{card}(\Lambda) \le \frac{1}{\sigma_k(\text{cap of Euclidean radius } \frac{r}{2})} \le \frac{1}{(cr)^{k-1}}\,,$$

where σ_k denotes the uniform probability measure on S^{k-1} (see [11], p.20). To estimate $\sigma(\widetilde{K})$, first note that

$$\begin{aligned}
\widetilde{K} &= (rK + rB_2^n) \cap S^{n-1} \\
&\subseteq (T + rB_2^n + rB_2^n) \cap S^{n-1} \\
&= (T + 2rB_2^n) \cap S^{n-1}
\end{aligned}$$

This last set is the union of $\operatorname{card}(T)$ caps, each of radius at most cr. Indeed, if $t \in T$ and $\theta \in (t + 2rB_2^n) \cap S^{n-1}$ then

$$\begin{aligned}
\left|\theta - \frac{t}{|t|}\right| &= 2\sin(\tfrac{1}{2}\arcsin \operatorname{dist}(\theta, \operatorname{span}\{t\})) \\
&\le 2\sin(\tfrac{1}{2}\arcsin 2r) && (\text{since } \theta \in t + 2rB_2^n) \\
&\le \arcsin 2r && (\text{since } \sin x \le x \text{ for } x \ge 0) \\
&\le \pi r && (\text{since arcsin is convex on } [0,1])
\end{aligned}$$

Therefore

$$\sigma(\widetilde{K}) \le \operatorname{card}(T)\sigma_n(\text{cap of radius } cr) \le A^n (cr)^{n-1}\,.$$

(At this point we see why this strategy works: on the one hand, making r smaller means we need more points in Λ so we have more chances to hit $\widetilde{K}$; on the other, making r smaller means $\widetilde{K}$ gets smaller and so easier to miss; but Λ is getting bigger at a rate like r^{k-1} while $\widetilde{K}$ is getting smaller at a rate like $1/r^{n-1}$, so the latter phenomenon wins.)

Putting it altogether, we have

$$\mathbb{P}(QE_0 \text{ meets } rK \cap S^{n-1}) \leq (cA)^n r^{n-k} .$$

For existence, then, we require $(cA)^n r^{n-k} < 1$, that is,

$$(cA)^{1/(1-\frac{k}{n})} < \frac{1}{r} ,$$

so we take such r. For high probability, we may require, say, $(cA)^n r^{n-k} < \frac{1}{2^n}$, that is, $(2cA)^n r^{n-k} < 1$, which is the same condition as for existence, except for the constant. □

7 Large sections from covering of the ball

Here we prove a statement which is, in spirit at least, dual to (19). We will show, more or less, that if a convex body is large in the sense that not too many copies of it are needed to cover B_2^n, then it is large in the sense that it has k-dimensional projections that contain large balls. The statement will, however, not be as strong as this suggests, since our hypothesis will require more information about covering, and the conclusion is subject to a few conditions. The precise statement:

$$\begin{aligned}
&\text{There is a constant } c > 0 \text{ such that}\\
&\text{if } K \text{ is a convex body in } \mathbb{R}^n\\
&\text{and } A \in \mathbb{R} \text{ is such that } \forall r > 0\colon \overline{N}(B_2^n, rK) \leq 2e^{An/r^2},\\
&\text{then for any } k \in \mathbb{N} \text{ such that } k \leq n - c\sqrt{n},\\
&\text{there exists } E \in G_{n,k}\\
&\text{such that } \operatorname{proj}_E K \supseteq \frac{c}{\sqrt{A}}\left(1-\sqrt{\frac{k}{n}}\right)^2 \operatorname{proj}_E B_2^n.
\end{aligned} \tag{20}$$

(My presentation here is based on [4], which gives generalizations of this type of statement for quasi-convex bodies; see that paper for references to previous results.)

The main hypothesis in (20) is a control on how fast the covering number grows as the covering body shrinks. This hypothesis is not too strange, in view of (16), which implies such an inequality with $A = cM(K)^2$.

The main condition on the conclusion in (20) is that the dimension of the subspace onto which we project is not entirely arbitrary, e.g., we cannot have

$k = n - 1$. Still, if we are only concerned with the proportion $\frac{k}{n}$, the condition is not severe, since we can make any desired proportion allowable by taking n large enough. (More precisely: for any $\lambda \in (0, 1)$, if n is sufficiently large, then for any $k \le \lambda n$ there exist projections as in (20).)

To prove (20) we need two lemmas. First, a statement from [11], p.18: Suppose $A \subseteq \mathbb{R}^n$ is bounded, $K \subseteq \mathbb{R}^n$ is closed and convex, and $0 \le \lambda < 1$. Then

$$A \subseteq (1-\lambda)K + \lambda A \implies A \subseteq K\,.$$

An immediate corollary is that, with the same assumptions on A, K, and λ,

$$A \subseteq K + \lambda A \implies (1-\lambda)A \subseteq K\,. \tag{21}$$

Second, we need the following result of Johnson and Lindenstrauss ([2], Lemma 1), which asserts if you have a collection of (not too many) points in $\mathbb{R}^n$, then for most projections, the images of these points are all about the same length. To be precise:

There is a constant $c > 0$ such that
for any $\epsilon > 0$, $n \in \mathbb{N}$, and $k \in [1..n]$,
if $N \le e^{ck\epsilon^2}$,
then for any points $(x_i)_1^N$ in $\mathbb{R}^n$,
and any orthogonal projection P of rank k, (22)
we have $\mu(Q \in O(n)\colon (\forall i \in [1..N]\colon |PQx_i| \overset{\epsilon}{\approx} D_{n,k}|x_i|)) \ge 1 - 2e^{-ck\epsilon^2}$,
where $D_{n,k} = \Gamma(\frac{n}{2})\Gamma(\frac{k+1}{2})/\Gamma(\frac{n+1}{2})\Gamma(\frac{k}{2})$,
μ is the Haar probability measure on $O(n)$,
and $x \overset{\epsilon}{\approx} y$ means that $(1-\epsilon)y \le x \le (1+\epsilon)y$.

First note that $|PQx_i| = |Q^{-1}PQx_i|$, and if Q ranges over $O(n)$, then $Q^{-1}PQ$ ranges over all projections of rank k; thus $Q^{-1}PQ$ is a natural implementation of a "random projection".

Next, note that $D_{n,k} = \mathbb{E}|P\theta|$, where θ is uniformly distributed on S^{n-1} and P is any orthogonal projection of rank k. (This expected value can be computed by changing the integral over the sphere into a gaussian integral; see [11], p.23) By Stirling's approximation, $D_{n,k} \sim \sqrt{\frac{k}{n}}$ as $k, n \to \infty$, and in fact,

$$\frac{1}{2}\sqrt{\frac{k}{n}} < D_{n,k} < \sqrt{\frac{k}{n}}\,.$$

Proof sketch Fix P. The function $S^{n-1} \to \mathbb{R}$, $\theta \mapsto |P\theta|$ is Lipschitz, so it concentrates around its average:

$$\sigma\big(\theta \in S^{n-1}\colon \big||P\theta| - D_{n,k}\big| > t\big) \le ce^{-cnt^2}\,.$$

Taking $t = \epsilon D_{n,k}$ yields

$$\sigma(\theta \in S^{n-1} \colon |P\theta| > (1+\epsilon)D_{n,k}) \le ce^{-cnD_{n,k}^2\epsilon^2} \le ce^{-ck\epsilon^2} .$$

Thus, for fixed $\theta \in S^{n-1}$,

$$\mu(Q \in O(n) \colon |PQ\theta| > (1+\epsilon)D_{n,k}) \le ce^{-ck\epsilon^2} ,$$

and so for fixed $(\theta_i)_1^N \subset S^{n-1}$,

$$\mu(Q \in O(n) \colon (\exists i \colon |PQ\theta_i| > (1+\epsilon)D_{n,k}) \le cNe^{-ck\epsilon^2} ,$$

so that if $N \le e^{ck\epsilon^2/2}$, then we obtain the desired estimate. □

Now we can prove (20).

Proof Let c' be the constant from (22). Let K and A be as in (20). Assume

$$\sqrt{k} \le \sqrt{n} - \sqrt{\frac{8\ln 2}{c'}} .$$

Set

$$\epsilon = \frac{1}{2}\left(\frac{1}{D_{n,k}} - 1\right) .$$

Note that $\epsilon > 0$ and that $c'\epsilon^2 k > 2\ln 2$. Let r be such that $2e^{An/r^2} = e^{c'\epsilon^2 k}$; note that

$$r = \sqrt{\frac{An}{c'\epsilon^2 k - \ln 2}} < \sqrt{\frac{2An}{c'\epsilon^2 k}} = \frac{1}{\epsilon}\sqrt{\frac{2A}{c'}}\sqrt{\frac{n}{k}} .$$

Let $T \subseteq \mathbb{R}^n$ be such that $T \subseteq B_2^n \subseteq T + rK$ and $\operatorname{card}(T) \le e^{c'\epsilon^2 k}$. Note that $2e^{-c'\epsilon^2 k} < 1$. Thus (22) yields an orthogonal projection P of rank k such that, for all $t \in T$, $|Pt| \le (1+\epsilon)D_{n,k}|t|$, and so $PT \subseteq (1+\epsilon)D_{n,k}B_2^n$. Therefore

$$PB_2^n \subseteq PT + rPK \subseteq (1+\epsilon)D_{n,k}PB_2^n + rPK .$$

Since $(1+\epsilon)D_{n,k} < 1$, it follows by (21) that

$$\begin{aligned} PK &\supseteq \frac{1}{r}(1-(1+\epsilon)D_{n,k})PB_2^n \\ &\supseteq \sqrt{\frac{c'}{2A}}\sqrt{\frac{k}{n}}\,\epsilon(1-(1+\epsilon)D_{n,k})PB_2^n \end{aligned}$$

The value of ϵ was chosen to minimize this expression; plugging it in yields (for suitable constant c)

$$PK \supseteq \frac{c}{\sqrt{A}}\left(1 - \sqrt{\frac{k}{n}}\right)^2 PB_2^n ,$$

as desired. □

References

[1] Peter Brass, William Moser, and János Pach. *Research Problems in Discrete Geometry*. Springer, New York, 2005. (Cited on page 9.)

[2] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math.*, 26:189–206, 1984. (Cited on page 13.)

[3] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer-Verlag, 1991. (Cited on page 7.)

[4] Alexander E. Litvak, Vitali D. Milman, and Alain Pajor. The covering numbers and "low M^*-estimate" for quasi-convex bodies. *P. Am. Math. Soc.*, 127:1499–1507, 1999. doi:10.1090/S0002-9939-99-04593-1. (Cited on pages 7 and 12.)

[5] Vitali Milman and Alain Pajor. Entropy and asymptotic geometry of non-symmetric convex bodies. *Adv. Math.*, 152(2):314–335, 2000. (Cited on page 6.)

[6] Alain Pajor and Nicole Tomczak-Jaegermann. Subspaces of small codimension of finite-dimensional Banach spaces. *P. Am. Math. Soc.*, 97:637–642, 1986. doi:10.1090/S0002-9939-1986-0845980-8. (Cited on page 7.)

[7] C. A. Rogers. A note on coverings. *Mathematika*, 4:1–6, 1957. (Cited on pages 6, 8, and 9.)

[8] C. A. Rogers and G. C. Shephard. The difference body of a convex body. *Arch. Math.*, 8:220–231, 1957. (Cited on page 7.)

[9] C. A. Rogers and C. Zong. Covering convex bodies by translates of convex bodies. *Mathematika*, 44:215–218, 1997. (Cited on pages 6 and 8.)

[10] S. Stein. The symmetry function in a convex body. *Pac. J. Math.*, 6:145–148, 1956. (Cited on page 6.)

[11] Steven Taschuk. Notes for seminar on volume ratio, February 2010. http://www.amotlpaa.org/math/sem200911.pdf. (Cited on pages 7, 10, 11, and 13.)