CHAPTER 19

Dimension Reduction – The Johnson-Lindenstrauss (JL) Lemma

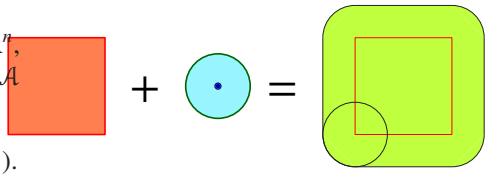
In this chapter, we will prove that given a set P of n points in $\mathbb{R}^{\mathsf{d}}$, one can reduce the dimension of the points to $k = O(\varepsilon^{-2} \log n)$ such that distances are $1 \pm \varepsilon$ preserved. Surprisingly, this reduction is done by randomly picking a subspace of k dimensions and projecting the points into this random subspace. One way of thinking about this result is that we are "compressing" the input of size $n\mathsf{d}$ (i.e., n points with d coordinates) into size $O(n\varepsilon^{-2} \log n)$, while (approximately) preserving distances.

19.1. The Brunn-Minkowski inequality

For a set $\mathcal{A} \subseteq \mathbb{R}^{\mathsf{d}}$ and a point $\mathsf{p} \in \mathbb{R}^{\mathsf{d}}$, let $\mathcal{A} + \mathsf{p}$ denote the translation of $\mathcal{A}$ by p. Formally, $\mathcal{A} + \mathsf{p} = \left\{ \mathsf{q} + \mathsf{p} \;\middle|\; \mathsf{q} \in \mathcal{A} \right\}$.

Definition 19.1. For two sets $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{R}^n$, let $\mathcal{A} + \mathcal{B}$ denote the ***Minkowski sum*** of $\mathcal{A}$ and $\mathcal{B}$. Formally,

$$\mathcal{A}+\mathcal{B} = \left\{ a + b \;\middle|\; a \in \mathcal{A}, b \in \mathcal{B} \right\} = \bigcup_{\mathsf{p} \in \mathcal{A}} (\mathsf{p}+\mathcal{B}).$$



Remark 19.2. It is easy to verify that if $\mathcal{A}', \mathcal{B}'$ are translated copies of $\mathcal{A}, \mathcal{B}$ (that is, $\mathcal{A}' = \mathcal{A} + \mathsf{p}$ and $\mathcal{B}' = \mathcal{B} + \mathsf{q}$, for some points $\mathsf{p}, \mathsf{q} \in \mathbb{R}^{\mathsf{d}}$) respectively, then $\mathcal{A}' + \mathcal{B}'$ is a translated copy of $\mathcal{A} + \mathcal{B}$. In particular, since volume is preserved under translation, we have that $\mathrm{Vol}(\mathcal{A}' + \mathcal{B}') = \mathrm{Vol}((\mathcal{A} + \mathcal{B}) + \mathsf{p} + \mathsf{q}) = \mathrm{Vol}(\mathcal{A} + \mathcal{B})$.

Our purpose here is to prove the following theorem.

Theorem 19.3 (Brunn-Minkowski inequality). *Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty compact sets in $\mathbb{R}^n$. Then*

$$\mathrm{Vol}(\mathcal{A} + \mathcal{B})^{1/n} \geq \mathrm{Vol}(\mathcal{A})^{1/n} + \mathrm{Vol}(\mathcal{B})^{1/n}.$$

Definition 19.4. A set $\mathcal{A} \subseteq \mathbb{R}^n$ is a ***brick set*** if it is the union of finitely many (close) axis parallel boxes with disjoint interiors.

It is intuitively clear, by limit arguments, that proving Theorem 19.3 for brick sets will imply it for the general case.

Lemma 19.5 (Brunn-Minkowski inequality for brick sets). *Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty brick sets in $\mathbb{R}^n$. Then*

$$\mathrm{Vol}(\mathcal{A} + \mathcal{B})^{1/n} \geq \mathrm{Vol}(\mathcal{A})^{1/n} + \mathrm{Vol}(\mathcal{B})^{1/n}.$$

PROOF. We prove by induction on the number k of bricks in $\mathcal{A}$ and $\mathcal{B}$. If $k = 2$, then $\mathcal{A}$ and $\mathcal{B}$ are just bricks, with dimensions $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_n$, respectively. In this case, the dimensions of $\mathcal{A} + \mathcal{B}$ are $\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n$, as can be easily verified. Thus, we need to prove that $\left(\prod_{i=1}^n \alpha_i\right)^{1/n} + \left(\prod_{i=1}^n \beta_i\right)^{1/n} \leq \left(\prod_{i=1}^n (\alpha_i + \beta_i)\right)^{1/n}$. Dividing the left side by the right side, we have

$$\left(\prod_{i=1}^n \frac{\alpha_i}{\alpha_i + \beta_i}\right)^{1/n} + \left(\prod_{i=1}^n \frac{\beta_i}{\alpha_i + \beta_i}\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n \frac{\alpha_i}{\alpha_i + \beta_i} + \frac{1}{n}\sum_{i=1}^n \frac{\beta_i}{\alpha_i + \beta_i} = 1,$$

by the generalized arithmetic-geometric mean inequality[①], and the claim follows for this case.

Now let $k > 2$ and suppose that the Brunn-Minkowski inequality holds for any pair of brick sets with fewer than k bricks (together). Let $\mathcal{A}$ and $\mathcal{B}$ be a pair of sets having k bricks together. Assume that $\mathcal{A}$ has at least two (disjoint) bricks. However, this implies that there is an axis parallel hyperplane h that separates the interior of one brick of $\mathcal{A}$ from the interior of another brick of $\mathcal{A}$ (the hyperplane h might intersect other bricks of $\mathcal{A}$). Assume that h is the hyperplane $x_1 = 0$ (this can be achieved by translation and renaming of coordinates).

Let $\overline{\mathcal{A}^+} = \mathcal{A} \cap h^+$ and $\overline{\mathcal{A}^-} = \mathcal{A} \cap h^-$, where h^+ and h^- are the two open halfspaces induced by h. Let $\mathcal{A}^+$ and $\mathcal{A}^-$ be the closure of $\overline{\mathcal{A}^+}$ and $\overline{\mathcal{A}^-}$, respectively. Clearly, $\mathcal{A}^+$ and $\mathcal{A}^-$ are both brick sets with (at least) one fewer brick than $\mathcal{A}$.

Next, observe that the claim is translation invariant (see Remark 19.2), and as such, let us translate $\mathcal{B}$ so that its volume is split by h in the same ratio $\mathcal{A}$'s volume is being split. Denote the two parts of $\mathcal{B}$ by $\mathcal{B}^+$ and $\mathcal{B}^-$, respectively. Let $\rho = \operatorname{Vol}(\mathcal{A}^+)/\operatorname{Vol}(\mathcal{A}) = \operatorname{Vol}(\mathcal{B}^+)/\operatorname{Vol}(\mathcal{B})$ (if $\operatorname{Vol}(\mathcal{A}) = 0$ or $\operatorname{Vol}(\mathcal{B}) = 0$, the claim trivially holds).

Observe that $\mathcal{X}^+ = \mathcal{A}^+ + \mathcal{B}^+ \subseteq \mathcal{A} + \mathcal{B}$, and $\mathcal{X}^+$ lies on one side of h (since $h \equiv (x_1 = 0)$), and similarly $\mathcal{X}^- = \mathcal{A}^- + \mathcal{B}^- \subseteq \mathcal{A} + \mathcal{B}$ and $\mathcal{X}^-$ lies on the other side of h. Thus, by induction and since $\mathcal{A}^+ + \mathcal{B}^+$ and $\mathcal{A}^- + \mathcal{B}^-$ are interior disjoint, we have

$$\begin{aligned}
\operatorname{Vol}(\mathcal{A} + \mathcal{B}) &\geq \operatorname{Vol}(\mathcal{A}^+ + \mathcal{B}^+) + \operatorname{Vol}(\mathcal{A}^- + \mathcal{B}^-) \\
&\geq \left(\operatorname{Vol}(\mathcal{A}^+)^{1/n} + \operatorname{Vol}(\mathcal{B}^+)^{1/n}\right)^n + \left(\operatorname{Vol}(\mathcal{A}^-)^{1/n} + \operatorname{Vol}(\mathcal{B}^-)^{1/n}\right)^n \\
&= \left[\rho^{1/n} \operatorname{Vol}(\mathcal{A})^{1/n} + \rho^{1/n} \operatorname{Vol}(\mathcal{B})^{1/n}\right]^n \\
&\quad + \left[(1-\rho)^{1/n} \operatorname{Vol}(\mathcal{A})^{1/n} + (1-\rho)^{1/n} \operatorname{Vol}(\mathcal{B})^{1/n}\right]^n \\
&= \Big(\rho + (1-\rho)\Big)\left[\operatorname{Vol}(\mathcal{A})^{1/n} + \operatorname{Vol}(\mathcal{B})^{1/n}\right]^n \\
&= \left[\operatorname{Vol}(\mathcal{A})^{1/n} + \operatorname{Vol}(\mathcal{B})^{1/n}\right]^n,
\end{aligned}$$

establishing the claim. ∎

PROOF OF THEOREM 19.3. Let $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \cdots \subseteq \mathcal{A}_i \subseteq \cdots$ be a sequence of finite brick sets, such that $\bigcup_i \mathcal{A}_i = \mathcal{A}$, and similarly let $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \cdots \subseteq \mathcal{B}_i \subseteq \cdots$ be a sequence of finite brick sets, such that $\bigcup_i \mathcal{B}_i = \mathcal{B}$. By the definition of volume[②], we have that $\lim_{i\to\infty} \operatorname{Vol}(\mathcal{A}_i) = \operatorname{Vol}(\mathcal{A})$ and $\lim_{i\to\infty} \operatorname{Vol}(\mathcal{B}_i) = \operatorname{Vol}(\mathcal{B})$.

[①] Here is a proof of the generalized form: Let $x_1, \ldots, x_n$ be n positive real numbers. Consider the quantity $R = x_1 x_2 \cdots x_n$. If we fix the sum of the n numbers to be equal to γ, then R is maximized when all the x_is are equal, as can be easily verified. Thus, $\sqrt[n]{x_1 x_2 \cdots x_n} \leq \sqrt[n]{(\gamma/n)^n} = \gamma/n = (x_1 + \cdots + x_n)/n$.

[②] This is the standard definition of volume. The reader unfamiliar with this fanfare can either consult a standard text on calculus (or measure theory) or take it for granted as this is intuitively clear.

We claim that $\lim_{i\to\infty} \mathrm{Vol}(\mathcal{A}_i+\mathcal{B}_i) = \mathrm{Vol}(\mathcal{A}+\mathcal{B})$. Indeed, consider any point $z \in \mathcal{A}+\mathcal{B}$, and let $u \in \mathcal{A}$ and $v \in \mathcal{B}$ be such that $u + v = z$. By definition, there exists an i, such that for all $j > i$ we have $u \in \mathcal{A}_j$, $v \in \mathcal{B}_j$, and as such $z \in \mathcal{A}_j + \mathcal{B}_j$. Thus, $\mathcal{A} + \mathcal{B} \subseteq \bigcup_j (\mathcal{A}_j + \mathcal{B}_j)$ and $\bigcup_j (\mathcal{A}_j + \mathcal{B}_j) \subseteq \bigcup_j (\mathcal{A} + \mathcal{B}) \subseteq \mathcal{A} + \mathcal{B}$; namely, $\bigcup_j (\mathcal{A}_j + \mathcal{B}_j) = \mathcal{A} + \mathcal{B}$.

Furthermore, for any $i > 0$, since $\mathcal{A}_i$ and $\mathcal{B}_i$ are brick sets, we have

$$\mathrm{Vol}(\mathcal{A}_i + \mathcal{B}_i)^{1/n} \geq \mathrm{Vol}(\mathcal{A}_i)^{1/n} + \mathrm{Vol}(\mathcal{B}_i)^{1/n},$$

by Lemma 19.5. Thus,

$$\begin{aligned}\mathrm{Vol}(\mathcal{A} + \mathcal{B})^{1/n} &= \lim_{i\to\infty} \mathrm{Vol}(\mathcal{A}_i + \mathcal{B}_i)^{1/n} \geq \lim_{i\to\infty} \Big(\mathrm{Vol}(\mathcal{A}_i)^{1/n} + \mathrm{Vol}(\mathcal{B}_i)^{1/n}\Big) \\ &= \mathrm{Vol}(\mathcal{A})^{1/n} + \mathrm{Vol}(\mathcal{B})^{1/n}.\end{aligned}$$

■

THEOREM 19.6 (Brunn-Minkowski theorem for slice volumes). *Let $\mathcal{P}$ be a convex set in $\mathbb{R}^{n+1}$, and let $\mathcal{A} = \mathcal{P} \cap (x_1 = a)$, $\mathcal{B} = \mathcal{P} \cap (x_1 = b)$, and $\mathcal{C} = \mathcal{P} \cap (x_1 = c)$ be three slices of $\mathcal{A}$, for $a < b < c$. We have* $\mathrm{Vol}(\mathcal{B}) \geq \min(\mathrm{Vol}(\mathcal{A}), \mathrm{Vol}(\mathcal{C}))$.

In fact, consider the function

$$v(t) = (\mathrm{Vol}(\mathcal{P} \cap (x_1 = t)))^{1/n},$$

and let $\mathcal{I} = [t_{\min}, t_{\max}]$ be the interval where the hyperplane $x_1 = t$ intersects $\mathcal{P}$. Then, $v(t)$ is concave on $\mathcal{I}$.

PROOF. If a or c is outside $\mathcal{I}$, then $\mathrm{Vol}(\mathcal{A}) = 0$ or $\mathrm{Vol}(\mathcal{C}) = 0$, respectively, and then the claim trivially holds.

Otherwise, let $\alpha = (b - a)/(c - a)$. We have that $b = (1 - \alpha) \cdot a + \alpha \cdot c$, and by the convexity of $\mathcal{P}$, we have $(1 - \alpha)\mathcal{A} + \alpha\mathcal{C} \subseteq \mathcal{B}$. Thus, by Theorem 19.3 we have

$$\begin{aligned}v(b) = \mathrm{Vol}(\mathcal{B})^{1/n} &\geq \mathrm{Vol}((1 - \alpha)\mathcal{A} + \alpha\mathcal{C})^{1/n} \geq \mathrm{Vol}((1 - \alpha)\mathcal{A})^{1/n} + \mathrm{Vol}(\alpha\mathcal{C})^{1/n} \\ &= \Big((1 - \alpha)^n \mathrm{Vol}(\mathcal{A})\Big)^{1/n} + \Big(\alpha^n \mathrm{Vol}(\mathcal{C})\Big)^{1/n} \\ &= (1 - \alpha)\,\mathrm{Vol}(\mathcal{A})^{1/n} + \alpha\,\mathrm{Vol}(\mathcal{C})^{1/n} \\ &= (1 - \alpha)v(a) + \alpha v(c).\end{aligned}$$

Namely, $v(\cdot)$ is concave on $\mathcal{I}$, and in particular $v(b) \geq \min(v(a), v(c))$, which in turn implies that $\mathrm{Vol}(\mathcal{B}) = v(b)^n \geq (\min(v(a), v(c)))^n = \min(\mathrm{Vol}(\mathcal{A}), \mathrm{Vol}(\mathcal{C}))$, as claimed. ■

COROLLARY 19.7. *For $\mathcal{A}$ and $\mathcal{B}$ compact sets in $\mathbb{R}^n$,* $\mathrm{Vol}((\mathcal{A}+\mathcal{B})/2) \geq \sqrt{\mathrm{Vol}(\mathcal{A})\,\mathrm{Vol}(\mathcal{B})}$.

PROOF. We have that

$$\begin{aligned}\mathrm{Vol}((\mathcal{A} + \mathcal{B})/2)^{1/n} &= \mathrm{Vol}(\mathcal{A}/2 + \mathcal{B}/2)^{1/n} \geq \mathrm{Vol}(\mathcal{A}/2)^{1/n} + \mathrm{Vol}(\mathcal{B}/2)^{1/n} \\ &= (\mathrm{Vol}(\mathcal{A})^{1/n} + \mathrm{Vol}(\mathcal{B})^{1/n})/2 \geq \sqrt{\mathrm{Vol}(\mathcal{A})^{1/n}\,\mathrm{Vol}(\mathcal{B})^{1/n}}\end{aligned}$$

by Theorem 19.3 and since $(a + b)/2 \geq \sqrt{ab}$ for any $a, b \geq 0$. The claim now follows by raising this inequality to the power n. ■

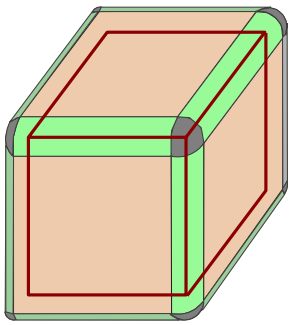
19.1.1. The isoperimetric inequality. The following is not used anywhere else and is provided because of its mathematical elegance. The skip-able reader can thus employ their special gift and move on to Section 19.2.

The ***isoperimetric inequality*** states that among all convex bodies of a fixed surface area, the ball has the largest volume (in particular, the unit circle is the largest area planar region with perimeter 2π). This problem can be traced back to antiquity; in particular Zenodorus (200–140 BC) wrote a monograph (which was lost) that seemed to have proved

the claim in the plane for some special cases. The first formal proof for the planar case was done by Steiner in 1841. Interestingly, the more general claim is an easy consequence of the Brunn-Minkowski inequality.

Let K be a convex body in $\mathbb{R}^n$ and let $\mathbf{b}$ be the n-dimensional ball of radius 1 centered at the origin. Let $\mathsf{S}(X)$ denote the surface area of a compact set $X \subseteq \mathbb{R}^n$. The ***isoperimetric inequality*** states that

$$\left(\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(\mathbf{b})}\right)^{1/n} \leq \left(\frac{\mathsf{S}(K)}{\mathsf{S}(\mathbf{b})}\right)^{1/(n-1)}. \tag{19.1}$$



Namely, the left side is the radius of a ball having the same volume as K, and the right side is the radius of a sphere having the same surface area as K. In particular, if we scale K so that its surface area is the same as $\mathbf{b}$, then the above inequality implies that $\mathrm{Vol}(K) \leq \mathrm{Vol}(\mathbf{b})$.

To prove (19.1), observe that $\mathrm{Vol}(\mathbf{b}) = \mathsf{S}(\mathbf{b})/n$[③]. Also, observe that $K + \varepsilon\mathbf{b}$ is the body K together with a small "atmosphere" around it of thickness ε. In particular, the volume of this "atmosphere" is (roughly) $\varepsilon\mathsf{S}(K)$ (in fact, Minkowski defined the surface area of a convex body to be the limit stated next).

Formally, we have

$$\mathsf{S}(K) = \lim_{\varepsilon\to 0+} \frac{\mathrm{Vol}(K+\varepsilon\mathbf{b}) - \mathrm{Vol}(K)}{\varepsilon} \geq \lim_{\varepsilon\to 0+} \frac{\left(\mathrm{Vol}(K)^{1/n} + \mathrm{Vol}(\varepsilon\mathbf{b})^{1/n}\right)^n - \mathrm{Vol}(K)}{\varepsilon},$$

by the Brunn-Minkowski inequality. Now $\mathrm{Vol}(\varepsilon\mathbf{b})^{1/n} = \varepsilon\,\mathrm{Vol}(\mathbf{b})^{1/n}$, and as such

$$\begin{aligned}\mathsf{S}(K) &\geq \lim_{\varepsilon\to 0+} \frac{\mathrm{Vol}(K) + \binom{n}{1}\varepsilon\,\mathrm{Vol}(K)^{\frac{n-1}{n}}\,\mathrm{Vol}(\mathbf{b})^{\frac{1}{n}} + \binom{n}{2}\varepsilon^2\langle\cdots\rangle + \cdots + \varepsilon^n\,\mathrm{Vol}(\mathbf{b}) - \mathrm{Vol}(K)}{\varepsilon} \\ &= \lim_{\varepsilon\to 0+} \frac{n\varepsilon\,\mathrm{Vol}(K)^{\frac{n-1}{n}}\,\mathrm{Vol}(\mathbf{b})^{\frac{1}{n}}}{\varepsilon} = n\,\mathrm{Vol}(K)^{\frac{n-1}{n}}\,\mathrm{Vol}(\mathbf{b})^{\frac{1}{n}}.\end{aligned}$$

Dividing both sides by $\mathsf{S}(\mathbf{b}) = n\,\mathrm{Vol}(\mathbf{b})$, we have

$$\frac{\mathsf{S}(K)}{\mathsf{S}(\mathbf{b})} \geq \frac{\mathrm{Vol}(K)^{(n-1)/n}}{\mathrm{Vol}(\mathbf{b})^{(n-1)/n}} \quad\Rightarrow\quad \left(\frac{\mathsf{S}(K)}{\mathsf{S}(\mathbf{b})}\right)^{1/(n-1)} \geq \left(\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(\mathbf{b})}\right)^{1/n},$$

establishing the isoperimetric inequality.

19.2. Measure concentration on the sphere

Let $\mathbb{S}^{(n-1)}$ be the unit sphere in $\mathbb{R}^n$. We assume there is a uniform probability measure defined over $\mathbb{S}^{(n-1)}$, such that its total measure is 1. Surprisingly, most of the mass of this measure is near the equator. Indeed, consider an arbitrary equator π on $\mathbb{S}^{(n-1)}$ (that it, it is the intersection of the sphere with a hyperplane passing through the center of the ball inducing the sphere). Next, consider all the points that are within distance $\approx \ell(n) = c/n^{1/3}$ from π. The question we are interested in is what fraction of the sphere is covered by this strip T (depicted in Figure 19.1).

[③]Indeed, $\mathrm{Vol}(\mathbf{b}) = \int_{r=0}^{1} \mathsf{S}(\mathbf{b})\, r^{n-1} dr = \mathsf{S}(\mathbf{b})/n$.

Notice that as the dimension increases, the width $\ell(n)$ of this strip decreases. But surprisingly, despite its width becoming smaller, as the dimension increases, this strip contains a larger and larger fraction of the sphere. In particular, the total fraction of the sphere not covered by this (shrinking!) strip converges to zero.

Furthermore, counterintuitively, this is true for *any* equator. We are going to show that even a stronger result holds: The mass of the sphere is concentrated close to the boundary of any set $A \subseteq \mathbb{S}^{(n-1)}$ such that $\mathbf{Pr}[A] = 1/2$.

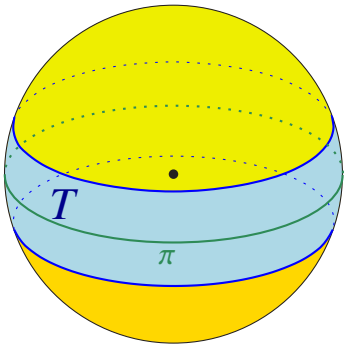


FIGURE 19.1

Before proving this somewhat surprising theorem, we will first try to get an intuition about the behavior of the hypersphere in high dimensions.

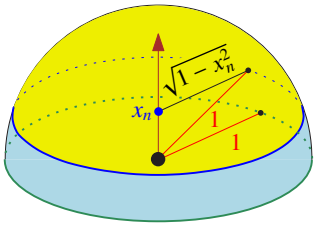
19.2.1. The strange and curious life of the hypersphere. Consider the ball of radius r in $\mathbb{R}^n$ (denoted by $r\,\mathbf{b}^n$), where $\mathbf{b}^n$ is the unit radius ball centered at the origin. Clearly, $\mathrm{Vol}(r\,\mathbf{b}^n) = r^n \mathrm{Vol}(\mathbf{b}^n)$. Now, even if r is very close to 1, the quantity r^n might be very close to zero if n is sufficiently large. Indeed, if $r = 1 - \delta$, then $r^n = (1-\delta)^n \leq \exp(-\delta n)$, which is very small if $\delta \gg 1/n$. (Here, we used $1 - x \leq e^x$, for $x \geq 0$.) Namely, for the ball in high dimensions, its mass is concentrated in a very thin shell close to its surface.

The volume of a ball and the surface area of a hypersphere. Let $\mathrm{Vol}(r\mathbf{b}^n)$ denote the volume of the ball of radius r in $\mathbb{R}^n$, and let $\mathsf{S}(r\mathbf{b}^n)$ denote the surface area of its bounding sphere (i.e., the surface area of $r\mathbb{S}^{(n-1)}$). It is known that

$$\mathrm{Vol}(r\mathbf{b}^n) = \frac{\pi^{n/2} r^n}{\Gamma(n/2+1)} \quad \text{and} \quad \mathsf{S}(r\mathbf{b}^n) = \frac{2\pi^{n/2} r^{n-1}}{\Gamma(n/2)},$$

where the gamma function, $\Gamma(\cdot)$, is an extension of the factorial function. Specifically, if n is even, then $\Gamma(n/2+1) = (n/2)!$, and for n odd $\Gamma(n/2+1) = \sqrt{\pi}(n!!)/2^{(n+1)/2}$, where $n!! = 1 \cdot 3 \cdot 5 \cdots n$ is the ***double factorial***. The most surprising implication of these two formulas is that, as n increases, the volume of the unit ball first increases (till dimension 5) and then starts decreasing to zero.

Similarly, the surface area of the unit sphere $\mathbb{S}^{(n-1)}$ in $\mathbb{R}^n$ tends to zero as the dimension increases. To see the above explicitly, compute the volume of the unit ball using an integral of its slice volume, when it is being sliced by a hyperplane perpendicular to the nth coordinate.



We have (see figure on the right) that

$$\mathrm{Vol}(\mathbf{b}^n) = \int_{x_n=-1}^{1} \mathrm{Vol}\Big(\sqrt{1-x_n^2}\,\mathbf{b}^{n-1}\Big)\,dx_n = \mathrm{Vol}\Big(\mathbf{b}^{n-1}\Big) \int_{x_n=-1}^{1} \Big(1-x_n^2\Big)^{(n-1)/2}\,dx_n.$$

Now, the integral on the right side tends to zero as n increases. In fact, for n very large, the term $\Big(1-x_n^2\Big)^{(n-1)/2}$ is very close to 0 everywhere except for a small interval around 0. This implies that the main contribution of the volume of the ball happens when we consider slices of the ball created by hyperplanes of the form $x_n = \delta$, where δ is small (roughly, for $|\delta| \leq 1/\sqrt{n}$).

If one has to visualize what such a ball in high dimensions looks like, it might be best to think about it as a star-like creature: It has very little mass close to the tips of any set of orthogonal directions we pick, and most of its mass (somehow) lies on hyperplanes passing close to its center.[④]

19.2.2. Measure concentration on the sphere.

THEOREM 19.8 (Measure concentration on the sphere). *Let $A \subseteq \mathbb{S}^{(n-1)}$ be a measurable set with $\mathbf{Pr}[A] \geq 1/2$, and let A_t denote the set of points of $\mathbb{S}^{(n-1)}$ within a distance at most t from A, where $t \leq 2$. Then $1 - \mathbf{Pr}[A_t] \leq 2\exp\left(-nt^2/2\right)$.*

PROOF. We will prove a slightly weaker bound, with $-nt^2/4$ in the exponent. Let $\widehat{A} = T(A)$, where

$$T(X) = \left\{ \alpha x \;\middle|\; x \in X, \alpha \in [0,1] \right\} \subseteq \mathbf{b}^n$$

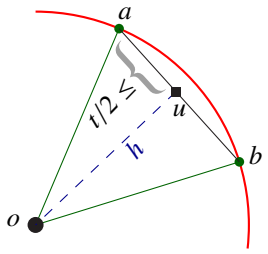
and $\mathbf{b}^n$ is the unit ball in $\mathbb{R}^n$. We have that $\mathbf{Pr}[A] = \mu\left(\widehat{A}\right)$, where $\mu\left(\widehat{A}\right) = \mathrm{Vol}\left(\widehat{A}\right) / \mathrm{Vol}(\mathbf{b}^n)$[⑤].

Let $B = \mathbb{S}^{(n-1)} \setminus A_t$ and $\widehat{B} = T(B)$. We have that $\|a - b\| \geq t$ for all $a \in A$ and $b \in B$. By Lemma 19.9 below, the set $\left(\widehat{A} + \widehat{B}\right)/2$ is contained in the ball $r\mathbf{b}^n$ centered at the origin, where $r = 1 - t^2/8$. Observe that $\mu(r\mathbf{b}^n) = \mathrm{Vol}(r\mathbf{b}^n)/\mathrm{Vol}(\mathbf{b}^n) = r^n = \left(1 - t^2/8\right)^n$. As such, applying the Brunn-Minkowski inequality in the form of Corollary 19.7, we have

$$\left(1 - \frac{t^2}{8}\right)^n = \mu(r\mathbf{b}^n) \geq \mu\left(\frac{\widehat{A} + \widehat{B}}{2}\right) \geq \sqrt{\mu\left(\widehat{A}\right)\mu\left(\widehat{B}\right)} = \sqrt{\mathbf{Pr}[A]\,\mathbf{Pr}[B]} \geq \sqrt{\mathbf{Pr}[B]\,/2}.$$

Thus, $\mathbf{Pr}[B] \leq 2(1 - t^2/8)^{2n} \leq 2\exp(-2nt^2/8)$, since $1 - x \leq \exp(-x)$, for $x \geq 0$. ■

LEMMA 19.9. *For any $\widehat{a} \in \widehat{A}$ and $\widehat{b} \in \widehat{B}$, we have* $\left\| \dfrac{\widehat{a} + \widehat{b}}{2} \right\| \leq 1 - \dfrac{t^2}{8}$.



PROOF. Let $\widehat{a} = \alpha a$ and $\widehat{b} = \beta b$, where $a \in A$ and $b \in B$. We have

(19.2)

$$\|u\| = \left\| \frac{a+b}{2} \right\| = \sqrt{1^2 - \left\| \frac{a-b}{2} \right\|^2} \leq \sqrt{1 - \frac{t^2}{4}} \leq 1 - \frac{t^2}{8},$$

since $\|a - b\| \geq t$. As for $\widehat{a}$ and $\widehat{b}$, assume that $\alpha \leq \beta$, and observe that the quantity $\left\| \widehat{a} + \widehat{b} \right\|$ is maximized when $\beta = 1$. As such, by the triangle inequality, we have

$$\left\| \frac{\widehat{a} + \widehat{b}}{2} \right\| = \left\| \frac{\alpha a + b}{2} \right\| \leq \left\| \frac{\alpha(a+b)}{2} \right\| + \left\| (1-\alpha)\frac{b}{2} \right\| \leq \alpha\left(1 - \frac{t^2}{8}\right) + (1-\alpha)\frac{1}{2} = \tau,$$

by (19.2) and since $\|b\| = 1$. Now, τ is a convex combination of the two numbers $1/2$ and $1 - t^2/8$. In particular, we conclude that $\tau \leq \max(1/2, 1 - t^2/8) \leq 1 - t^2/8$, since $t \leq 2$. ■

[④] In short, it looks like a Boojum [**Car76**].

[⑤] This is one of these "trivial" claims that might give the reader a pause, so here is a formal proof. Pick a random point p uniformly inside the ball $\mathbf{b}^n$. Let ψ be the probability that $\mathsf{p} \in \widehat{A}$. Clearly, $\mathrm{Vol}\left(\widehat{A}\right) = \psi\,\mathrm{Vol}(\mathbf{b}^n)$. So, consider the normalized point $\mathsf{q} = \mathsf{p}/\|\mathsf{p}\|$. Clearly, $\mathsf{p} \in \widehat{A}$ if and only if $\mathsf{q} \in A$, by the definition of $\widehat{A}$. Thus, $\mu\left(\widehat{A}\right) = \mathrm{Vol}\left(\widehat{A}\right)/\mathrm{Vol}(\mathbf{b}^n) = \psi = \mathbf{Pr}\left[\mathsf{p} \in \widehat{A}\right] = \mathbf{Pr}\left[\mathsf{q} \in A\right] = \mathbf{Pr}[A]$, since q has a uniform distribution on the hypersphere by assumption.

19.3. Concentration of Lipschitz functions

Consider a function $f : \mathbb{S}^{(n-1)} \to \mathbb{R}$, and imagine that we have a probability density function defined over the sphere. Let $\mathbf{Pr}[f \leq t] = \mathbf{Pr}\left[\left\{x \in S^{n-1} \;\middle|\; f(x) \leq t\right\}\right]$. We define the ***median*** of f, denoted by $\mathrm{med}(f)$, to be $\sup t$, such that $\mathbf{Pr}[f \leq t] \leq 1/2$.

We define $\mathbf{Pr}[f < \mathrm{med}(f)] = \sup_{x<\mathrm{med}(f)} \mathbf{Pr}[f \leq x]$. The following is obvious but requires a formal proof.

LEMMA 19.10. *We have* $\mathbf{Pr}\Big[f < \mathrm{med}(f)\Big] \leq 1/2$ *and* $\mathbf{Pr}\Big[f > \mathrm{med}(f)\Big] \leq 1/2$.

PROOF. Since $\bigcup_{k\geq 1}(-\infty, \mathrm{med}(f) - 1/k] = (-\infty, \mathrm{med}(f))$, we have

$$\mathbf{Pr}\Big[f < \mathrm{med}(f)\Big] = \sup_{k\geq 1} \mathbf{Pr}\left[f \leq \mathrm{med}(f) - \frac{1}{k}\right] \leq \sup_{k\geq 1} \frac{1}{2} = \frac{1}{2}.$$

The second claim follows by a symmetric argument. ■

DEFINITION 19.11 (***c*-Lipschitz**). A function $f : A \to B$ is ***c*-Lipschitz** if, for any $x, y \in A$, we have $\|f(x) - f(y)\| \leq c\,\|x - y\|$.

THEOREM 19.12 (Lévy's lemma). *Let* $f : \mathbb{S}^{(n-1)} \to \mathbb{R}$ *be* 1*-Lipschitz. Then for all* $t \in [0, 1]$*, we have*

$$\mathbf{Pr}\Big[f > \mathrm{med}(f) + t\Big] \leq 2\exp\left(-t^2 n/2\right) \quad \textit{and} \quad \mathbf{Pr}\Big[f < \mathrm{med}(f) - t\Big] \leq 2\exp\left(-t^2 n/2\right).$$

PROOF. We prove only the first inequality; the second follows by symmetry. Let

$$A = \left\{x \in \mathbb{S}^{(n-1)} \;\middle|\; f(x) \leq \mathrm{med}(f)\right\}.$$

By Lemma 19.10, we have $\mathbf{Pr}[A] \geq 1/2$. Consider a point $x \in A_t$, where A_t is as defined in Theorem 19.8. Let $\mathsf{nn}(x)$ be the nearest point in A to x. We have by definition that $\|x - \mathsf{nn}(x)\| \leq t$. As such, since f is 1-Lipschitz and $\mathsf{nn}(x) \in A$, we have that

$$f(x) \leq f(\mathsf{nn}(x)) + \|\mathsf{nn}(x) - x\| \leq \mathrm{med}(f) + t.$$

Thus, by Theorem 19.8, we get $\mathbf{Pr}[f > \mathrm{med}(f) + t] \leq 1 - \mathbf{Pr}[A_t] \leq 2\exp\left(-t^2 n/2\right)$. ■

19.4. The Johnson-Lindenstrauss lemma

LEMMA 19.13. *For a unit vector* $x \in \mathbb{S}^{(n-1)}$*, let*

$$f(x) = \sqrt{x_1^2 + x_2^2 + \cdots + x_k^2}$$

be the length of the projection of x into the subspace formed by the first k coordinates. Let x be a vector randomly chosen with uniform distribution from $\mathbb{S}^{(n-1)}$*. Then* $f(x)$ *is sharply concentrated. Namely, there exists* $m = m(n, k)$ *such that*

$$\mathbf{Pr}\Big[f(x) \geq m + t\Big] \leq 2\exp(-t^2 n/2) \quad \textit{and} \quad \mathbf{Pr}\Big[f(x) \leq m - t\Big] \leq 2\exp(-t^2 n/2),$$

for any $t \in [0, 1]$*. Furthermore, for* $k \geq 10 \ln n$*, we have* $m \geq \frac{1}{2}\sqrt{k/n}$.

PROOF. The orthogonal projection $p : \mathbb{R}^n \to \mathbb{R}^k$ given by $p(x_1, \ldots, x_n) = (x_1, \ldots, x_k)$ is 1-Lipschitz (since projections can only shrink distances; see Exercise 19.3). As such, $f(x) = \|p(x)\|$ is 1-Lipschitz, since for any x, y we have

$$|f(x) - f(y)| = \Big|\,\|p(x)\| - \|p(y)\|\Big| \leq \|p(x) - p(y)\| \leq \|x - y\|,$$

by the triangle inequality and since p is 1-Lipschitz. Theorem 19.12 (i.e., Lévy's lemma) gives the required tail estimate with $m = \mathrm{med}(f)$.

Thus, we only need to prove the lower bound on m. For a random $x = (x_1, \ldots, x_n) \in \mathbb{S}^{(n-1)}$, we have $\mathbf{E}\left[\|x\|^2\right] = 1$. By linearity of expectations and by symmetry, we have $1 = \mathbf{E}\left[\|x\|^2\right] = \mathbf{E}\left[\sum_{i=1}^n x_i^2\right] = \sum_{i=1}^n \mathbf{E}\left[x_i^2\right] = n\,\mathbf{E}\left[x_j^2\right]$, for any $1 \le j \le n$. Thus, $\mathbf{E}\left[x_j^2\right] = 1/n$, for $j = 1, \ldots, n$. Thus,

$$\mathbf{E}\left[(f(x))^2\right] = \mathbf{E}\left[\sum_{i=1}^k x_i^2\right] = \sum_{i=1}^k \mathbf{E}[x_i] = \frac{k}{n},$$

by linearity of expectation.

We next use that f is concentrated to show that f^2 is also relatively concentrated. For any $t \ge 0$, we have

$$\frac{k}{n} = \mathbf{E}\left[f^2\right] \le \mathbf{Pr}\left[f \le m + t\right](m+t)^2 + \mathbf{Pr}\left[f \ge m + t\right] \cdot 1 \le 1 \cdot (m+t)^2 + 2\exp(-t^2 n/2),$$

since $f(x) \le 1$, for any $x \in \mathbb{S}^{(n-1)}$. Let $t = \sqrt{k/5n}$. Since $k \ge 10 \ln n$, we have that $2\exp(-t^2 n/2) \le 2/n$. We get that

$$\frac{k}{n} \le \left(m + \sqrt{k/5n}\right)^2 + 2/n,$$

implying that $\sqrt{(k-2)/n} \le m + \sqrt{k/5n}$, which in turn implies that $m \ge \sqrt{(k-2)/n} - \sqrt{k/5n} \ge \frac{1}{2}\sqrt{k/n}$. ∎

Next, we would like to argue that given a fixed vector, projecting it down into a random k-dimensional subspace results in a random vector such that its length is highly concentrated. This would imply that we can do dimension reduction and still preserve distances between points that we care about.

To this end, we would like to flip Lemma 19.13 around. Instead of randomly picking a point and projecting it down to the first k-dimensional space, we would like x to be fixed and randomly pick the k-dimensional subspace we project into. However, we need to pick this random k-dimensional space carefully. Indeed, if we rotate this random subspace, by a transformation T, so that it occupies the first k dimensions, then the point $T(x)$ needs to be uniformly distributed on the hypersphere if we want to use Lemma 19.13.

As such, we would like to randomly pick a rotation of $\mathbb{R}^n$. This maps the standard orthonormal basis into a randomly rotated orthonormal space. Taking the subspace spanned by the first k vectors of the rotated basis results in a k-dimensional random subspace. Such a rotation is an orthonormal matrix with determinant 1. We can generate such a matrix, by randomly picking a vector $e_1 \in \mathbb{S}^{(n-1)}$. Next, we set e_1 as the first column of our rotation matrix and generate the other $n-1$ columns, by generating recursively $n-1$ orthonormal vectors in the space orthogonal to e_1.

Remark 19.14 (Generating a random point on the sphere). At this point, the reader might wonder how we pick a point uniformly from the unit hypersphere. The idea is to pick a point from the multi-dimensional normal distribution $N^n(0,1)$ and normalize it to have length 1. Since the multi-dimensional normal distribution has the density function

$$(2\pi)^{-n/2} \exp\left(-(x_1^2 + x_2^2 + \cdots + x_n^2)/2\right),$$

which is symmetric (i.e., all the points at distance r from the origin have the same distribution), it follows that this indeed generates a point randomly and uniformly on $\mathbb{S}^{(n-1)}$.

Generating a vector with multi-dimensional normal distribution is no more than picking each coordinate according to the normal distribution; see Lemma 27.11$_{\text{p338}}$. Given a source of random numbers according to the uniform distribution, this can be done using $O(1)$ computations per coordinate, using the Box-Muller transformation [**BM58**]. Overall, each random vector can be generated in $O(n)$ time.

Since projecting down the n-dimensional normal distribution to the lower-dimensional space yields a normal distribution, it follows that generating a random projection is no more than randomly picking n vectors according to the multi-dimensional normal distribution $v_1, \ldots, v_n$. Then, we orthonormalize them, using Graham-Schmidt, where $\widehat{v_1} = v_1 / \|v_1\|$ and $\widehat{v_i}$ is the normalized vector of $v_i - w_i$, where w_i is the projection of v_i to the space spanned by $v_1, \ldots, v_{i-1}$.

Taking those vectors as columns of a matrix generates a matrix A, with determinant either 1 or -1. We multiply one of the vectors by -1 if the determinant is -1. The resulting matrix is a random rotation matrix.

We can now restate Lemma 19.13 in the setting where the vector is fixed and the projection is into a random subspace.

LEMMA 19.15. *Let $x \in \mathbb{S}^{(n-1)}$ be an arbitrary unit vector. Now, consider a random k-dimensional subspace $\mathcal{F}$, and let $f(x)$ be the length of the projection of x into $\mathcal{F}$. Then, there exists $m = m(n,k)$ such that*

$$\mathbf{Pr}\Big[f(x) \geq m + t\Big] \leq 2\exp\Big(-t^2 n/2\Big) \quad \text{and} \quad \mathbf{Pr}\Big[f(x) \leq m - t\Big] \leq 2\exp\Big(-t^2 n/2\Big),$$

for any $t \in [0,1]$. Furthermore, for $k \geq 10 \ln n$, we have $m \geq \frac{1}{2}\sqrt{k/n}$.

PROOF. Let v_i be the ith orthonormal vector having 1 at the ith coordinate and 0 everywhere else. Let $\mathbf{M}$ be a random translation of space generated as described above. Clearly, for arbitrary fixed unit vector x, the vector $\mathbf{M}x$ is distributed uniformly on the sphere. Now, the ith column of the matrix $\mathbf{M}$ is the random vector e_i, and $e_i = \mathbf{M}^T v_i$. As such, we have

$$\langle \mathbf{M}x, v_i \rangle = (\mathbf{M}x)^T v_i = x^T \mathbf{M}^T v_i = x^T e_i = \langle x, e_i \rangle .$$

In particular, treating $\mathbf{M}x$ as a random vector and projecting it on the first k coordinates, we have that

$$f(x) = \sqrt{\sum_{i=1}^{k} \langle \mathbf{M}x, v_i \rangle^2} = \sqrt{\sum_{i=1}^{k} \langle x, e_i \rangle^2}.$$

But $e_1, \ldots, e_k$ is just an orthonormal basis of a random k-dimensional subspace. As such, the expression on the right is the length of the projection of x into a k-dimensional random subspace. As such, the length of the projection of x into a random k-dimensional subspace has exactly the same distribution as the length of the projection of a random vector into the first k coordinates. The claim now follows by Lemma 19.13. ■

DEFINITION 19.16. The mapping $f : \mathbb{R}^n \to \mathbb{R}^k$ is called ***K*-bi-Lipschitz** for a subset $X \subseteq \mathbb{R}^n$ if there exists a constant $c > 0$ such that

$$cK^{-1} \cdot \|\mathsf{p} - \mathsf{q}\| \leq \|f(\mathsf{p}) - f(\mathsf{q})\| \leq c \cdot \|\mathsf{p} - \mathsf{q}\| ,$$

for all $\mathsf{p}, \mathsf{q} \in X$.

The least K for which f is K-bi-Lipschitz is called the ***distortion*** of f and is denoted dist(f). We will refer to f as a ***K*-embedding** of X.

REMARK 19.17. Let $X \subseteq \mathbb{R}^m$ be a set of n points, where m potentially might be much larger than n. Observe that in this case, since we only care about the inter-point distances of points in X, we can consider X to be a set of points lying in the affine subspace $\mathcal{F}$ spanned by the points of X. Note that this subspace has dimension $n-1$. As such, each point of X can be interpreted as an $(n-1)$-dimensional point in $\mathcal{F}$. Namely, we can assume, for our purposes, that the set of n points in Euclidean space we care about lies in $\mathbb{R}^n$ (i.e., $\mathbb{R}^{n-1}$).

Note that if $m < n$, we can always pad all the coordinates of the points of X by zeros, such that the resulting point set lies in $\mathbb{R}^n$.

THEOREM 19.18 (Johnson-Lindenstrauss lemma / JL lemma). *Let X be an n-point set in a Euclidean space, and let $\varepsilon \in (0,1]$ be given. Then there exists a $(1+\varepsilon)$-embedding of X into $\mathbb{R}^k$, where $k = O(\varepsilon^{-2}\log n)$.*

PROOF. By Remark 19.17, we can assume that $X \subseteq \mathbb{R}^n$. Let $k = 200\varepsilon^{-2}\ln n$. Assume $k < n$, and let $\mathcal{F}$ be a random k-dimensional linear subspace of $\mathbb{R}^n$. Let $P_{\mathcal{F}} : \mathbb{R}^n \to \mathcal{F}$ be the orthogonal projection operator of $\mathbb{R}^n$ into $\mathcal{F}$. Let m be the number around which $\|P_{\mathcal{F}}(x)\|$ is concentrated, for $x \in \mathbb{S}^{(n-1)}$, as in Lemma 19.15.

Fix two points $x, y \in \mathbb{R}^n$. We prove that

$$\left(1-\frac{\varepsilon}{3}\right) m \|x-y\| \le \|P_{\mathcal{F}}(x) - P_{\mathcal{F}}(y)\| \le \left(1+\frac{\varepsilon}{3}\right) m \|x-y\|$$

holds with probability $\ge 1 - n^{-2}$. Since there are $\binom{n}{2}$ pairs of points in X, it follows that with constant probability (say $> 1/3$) this holds for all pairs of points of X. In such a case, the mapping p is a D-embedding of X into $\mathbb{R}^k$ with $D \le \frac{1+\varepsilon/3}{1-\varepsilon/3} \le 1+\varepsilon$, for $\varepsilon \le 1$.

Let $u = x - y$. We have $P_{\mathcal{F}}(u) = P_{\mathcal{F}}(x) - P_{\mathcal{F}}(y)$ since $P_{\mathcal{F}}(\cdot)$ is a linear operator. Thus, the condition becomes $\left(1-\frac{\varepsilon}{3}\right) m \|u\| \le \|P_{\mathcal{F}}(u)\| \le \left(1+\frac{\varepsilon}{3}\right) m \|u\|$. Again, since projection is a linear operator, for any $\alpha > 0$, the condition is equivalent to

$$\left(1-\tfrac{\varepsilon}{3}\right) m \|\alpha u\| \le \|P_{\mathcal{F}}(\alpha u)\| \le \left(1+\tfrac{\varepsilon}{3}\right) m \|\alpha u\|.$$

As such, we can assume that $\|u\| = 1$ by picking $\alpha = 1/\|u\|$. Namely, we need to show that

$$\left| \|P_{\mathcal{F}}(u)\| - m \right| \le \frac{\varepsilon}{3} m.$$

Let $f(u) = \|P_{\mathcal{F}}(u)\|$. By Lemma 19.13 (exchanging the random space with the random vector), for $t = \varepsilon m/3$, we have that the probability that this does not hold is bounded by

$$\mathbf{Pr}[|f(u) - m| \ge t] \le 4\exp\left(-\frac{t^2 n}{2}\right) = 4\exp\left(\frac{-\varepsilon^2 m^2 n}{18}\right) \le 4\exp\left(-\frac{\varepsilon^2 k}{72}\right) < n^{-2},$$

since $m \ge \frac{1}{2}\sqrt{k/n}$ and $k = 200\varepsilon^{-2}\ln n$. ∎

19.5. Bibliographical notes

Our presentation follows Matoušek [**Mat02**]. The Brunn-Minkowski inequality is a powerful inequality which is widely used in mathematics. A nice survey of this inequality and its applications is provided by Gardner [**Gar02**]. Gardner says, "In a sea of mathematics, the Brunn-Minkowski inequality appears like an octopus, tentacles reaching far and wide, its shape and color changing as it roams from one area to the next." However, Gardner is careful in claiming that the Brunn-Minkowski inequality is one of the most powerful inequalities in mathematics since, as a wit put it, "The most powerful inequality is $x^2 \ge 0$, since all inequalities are in some sense equivalent to it."

A striking application of the Brunn-Minkowski inequality is the proof that in any partial ordering of n elements, there is a single comparison that, knowing its result, reduces

the number of linear extensions that are consistent with the partial ordering, by a constant fraction. This immediately implies (the uninteresting result) that one can sort n elements in $O(n \log n)$ comparisons. More interestingly, it implies that if there are m linear extensions of the current partial ordering, we can *always* sort it using $O(\log m)$ comparisons. A nice exposition of this surprising result is provided by Matoušek [**Mat02**, Section 12.3].

There are several alternative proofs of the Johnson-Lindenstrauss lemma (i.e., JL lemma); see [**IM98**] and [**DG03**]. Interestingly, it is enough to pick each entry in the dimension-reducing matrix randomly out of $-1, 0, 1$. This requires a more involved proof [**Ach01**]. This is useful when one cares about storing this dimension reduction transformation efficiently. In particular, recently, there was a flurry of work on making the JL lemma both faster to compute (per point) and sparser. For example, Kane and Nelson [**KN10**] show that one can compute a matrix for dimension reduction with $O(\varepsilon^{-1} \log n)$ non-zero entries per column (the target dimension is still $O(\varepsilon^{-2} \log n)$). See [**KN10**] and references therein for more details.

Magen [**Mag07**] observed that the JL lemma preserves angles, and in fact can be used to preserve any "k-dimensional angle", by projecting down to dimension $O(k\varepsilon^{-2} \log n)$. In particular, Exercise 19.4 is taken from there.

Surprisingly, the random embedding preserves much more structure than distances between points. It preserves the structure and distances of surfaces as long as they are low dimensional and "well behaved"; see [**AHY07**] for some results in this direction.

Dimension reduction is crucial in computational learning, AI, databases, etc. One common technique that is being used in practice is to do PCA (i.e., principal component analysis) and take the first few main axes. Other techniques include independent component analysis and MDS (multi-dimensional scaling). MDS tries to embed points from high dimensions into low dimension ($\mathsf{d} = 2$ or 3), while preserving some properties. Theoretically, dimension reduction into really low dimensions is hopeless, as the distortion in the worst case is $\Omega(n^{1/(k-1)})$, if k is the target dimension [**Mat90**].

19.6. Exercises

Exercise 19.1 (Boxes can be separated). (Easy) Let A and B be two axis parallel boxes that are interior disjoint. Prove that there is always an axis parallel hyperplane that separates the interior of the two boxes.

Exercise 19.2 (Brunn-Minkowski inequality, slight extension). Prove the following claim.

> Corollary 19.19. *For A and B compact sets in $\mathbb{R}^n$, we have for any $\lambda \in [0, 1]$ that* $\mathrm{Vol}(\lambda A + (1 - \lambda)B) \geq \mathrm{Vol}(A)^{\lambda} \mathrm{Vol}(B)^{1-\lambda}$.

Exercise 19.3 (Projections are contractions). (Easy) Let $\mathcal{F}$ be a k-dimensional affine subspace, and let $P_{\mathcal{F}} : \mathbb{R}^{\mathsf{d}} \to \mathcal{F}$ be the projection that maps every point $x \in \mathbb{R}^{\mathsf{d}}$ to its nearest neighbor on $\mathcal{F}$. Prove that $P_{\mathcal{F}}$ is a contraction (i.e., 1-Lipschitz). Namely, for any $\mathsf{p}, \mathsf{q} \in \mathbb{R}^{\mathsf{d}}$, we have that $\|P_{\mathcal{F}}(\mathsf{p}) - P_{\mathcal{F}}(\mathsf{q})\| \leq \|\mathsf{p} - \mathsf{q}\|$.

Exercise 19.4 (JL lemma works for angles). Show that the Johnson-Lindenstrauss lemma also $(1 \pm \varepsilon)$-preserves angles among triples of points of P (you might need to increase the target dimension however by a constant factor). [**Hint:** For every angle, construct an equilateral triangle whose edges are being preserved by the projection (add the vertices of those triangles (conceptually) to the point set being embedded). Argue that this implies that the angle is being preserved.]