CHAPTER 5

Metric Entropy and Concentration of Measure in Classical Spaces

This chapter presents two fundamental concepts which will be applied in later chapters: the metric entropy (a.k.a. packing and covering) and the concentration of measure. Their conjunction leads to the Dvoretzky theorem, which will be presented in Chapter 7.

5.1. Nets and packings

We will introduce now the complementary concepts of covering numbers (also called metric entropy) and packing numbers, which quantify the complexity of a given compact metric set. It will turn out that these parameters are closely related to the volume and the mean width considered in the preceding chapter.

We first analyze the special but fundamental cases of the sphere and the discrete cube. We subsequently discuss classical groups and manifolds, and general convex bodies.

5.1.1. Definitions. If K is a compact subset of a metric space (M, d), a finite subset $\mathcal{N} \subset K$ is called an *ε-net* of K if, for every $x \in K$, $\operatorname{dist}(x, \mathcal{N}) \leqslant \varepsilon$. Since this is equivalent to the union of the corresponding balls containing K, an alternative terminology is that of a *covering*, see Figure 5.1. We denote by $N(K, \varepsilon)$ (or by $N(K, d, \varepsilon)$, if there is an ambiguity as to the choice of the metric) the minimal cardinality of an ε-net in K.

A subset $\mathcal{P} \subset K$ is called *ε-separated* if any pair (x, y) of distinct elements from $\mathcal{P}$ satisfies $d(x, y) > \varepsilon$. This property implies that the balls of radius $\varepsilon/2$ centered at elements of $\mathcal{P}$ are disjoint (a configuration usually referred to as *packing*, whence the usage of the letter P; see Figure 5.1), and in most contexts the two properties are essentially equivalent. We denote by $P(K, \varepsilon)$ or $P(K, d, \varepsilon)$ the largest cardinality of an ε-separated set in K. The quantities $N(K, \varepsilon)$ and $P(K, \varepsilon)$ are called, respectively, *covering numbers* and *packing numbers*. The function $\varepsilon \mapsto N(K, d, \varepsilon)$, and its various generalizations, is also often referred to as the *metric entropy* of (K, d).

For any compact metric space K, the following two relations between nets and packings are fundamental. First, if $\mathcal{P}$ is a 2ε-separated set and $\mathcal{N}$ is an ε-net, then the open balls of radius ε centered at elements from $\mathcal{N}$ cover K, and each ball contains at most one element of $\mathcal{P}$. Second, an ε-separated set which is maximal (with respect to inclusion) is an ε-net (the reader not familiar with this circle of ideas is encouraged to check these elementary facts). It follows that we have the inequalities

$$P(K, 2\varepsilon) \leqslant N(K, \varepsilon) \leqslant P(K, \varepsilon). \tag{5.1}$$

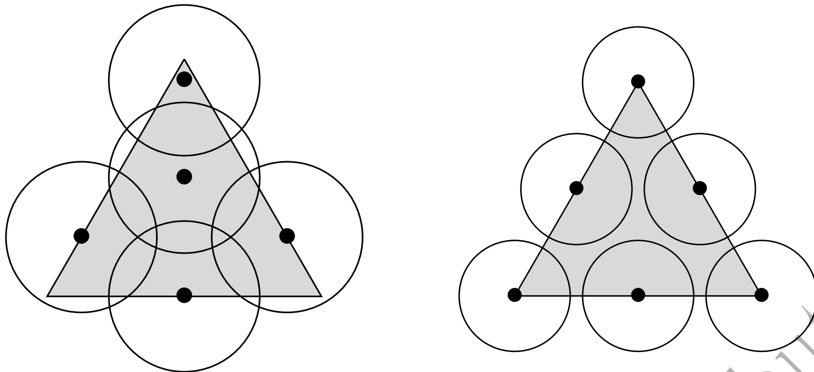

FIGURE 5.1. A net (left) and a packing (right) for an equilateral triangle (with the Euclidean metric in $\mathbb{R}^2$). For optimal packings or covering with few "classical" convex bodies in the plane (squares, circles or triangles), see the website **[@1]**.

Packings and coverings have been extensively studied, particularly for "standard" metric spaces. In various applications it is useful to know that there exist "large" packings and/or "small" nets, and often to be able to exhibit them in a constructive manner. By (5.1), both notions are equivalent whenever the resolution parameter ε is specified only up to a multiplicative constant. On the other hand, for some applications, such as coding theory, very precise results are in high demand.

In many situations the isometry group of K acts transitively and preserves a natural probability measure μ. In particular, all balls of radius ε have then the same measure, denoted by $V(\varepsilon)$, and we have the simple inequalities

$$\frac{1}{V(\varepsilon)} \leqslant N(K,\varepsilon) \leqslant P(K,\varepsilon) \leqslant \frac{1}{V(\varepsilon/2)}. \tag{5.2}$$

EXERCISE 5.1. Here, we introduce variations on the definitions and check their equivalence. Let M be a metric space and K a compact subset. Denote by $N'(K,\varepsilon)$ the smallest cardinality of a family of closed balls of radius ε in M whose union contains K (the difference with the definition of $N(K,\varepsilon)$ is that the centers are not required to be in K). It is sometimes more convenient to allow sets of *diameter* $\leqslant 2\varepsilon$ in place of balls of *radius* ε; call the resulting the quantity $N''(K,\varepsilon)$. Let also $P'(K,\varepsilon)$ be the largest cardinality of a family of disjoint open balls of radius $\varepsilon/2$ with centers in K. Check the inequalities

$$N''(K,\varepsilon) \leqslant N'(K,\varepsilon) \leqslant N(K,\varepsilon) \leqslant P(K,\varepsilon) \leqslant N''(K,\varepsilon/2)$$

and

$$P(K,\varepsilon) \leqslant P'(K,\varepsilon) \leqslant N(K,\varepsilon/2).$$

Give examples showing that inequalities may be strict (see also Exercise 5.16).

5.1.2. Nets and packings on the Euclidean sphere. We first consider the specific case of the sphere S^{n-1} for $n \geqslant 2$; denote by g the geodesic distance and by σ the normalized Haar measure. In some cases, it is more appropriate to consider the extrinsic distance inherited from $\mathbb{R}^n$. However, any result about one distance transfers automatically to the other distance (see Appendix B.1 for details). We give a brief overview of known estimates for packing and covering numbers for the sphere. The first point of business will be a discussion of volumes of spherical caps, which enter the subject via (5.2).

5.1.2.1. *Estimates on volumes of spherical caps.* Given $x_0 \in S^{n-1}$, let $C(x_0, \varepsilon)$ be the cap of center x_0 and geodesic radius ε, and denote $V(\varepsilon) = \sigma(C(x_0, \varepsilon))$ ($\varepsilon \in [0, \pi]$ is tacitly assumed). We have

$$V(\varepsilon) = \frac{\int_0^\varepsilon \sin^{n-2}\theta \, \mathrm{d}\theta}{\int_0^\pi \sin^{n-2}\theta \, \mathrm{d}\theta}. \tag{5.3}$$

The denominator at the right-hand side of (5.3) (Wallis integral) equals $\sqrt{2\pi}/\kappa_{n-1}$. Note that $V(\pi - \varepsilon) = 1 - V(\varepsilon)$, in particular $V(\pi/2) = 1/2$. For fixed $0 < \varepsilon < \pi/2$, $V(\varepsilon)$ tends to 0 exponentially fast in the dimension: one has $V(\varepsilon)^{1/n} \sim \sin(\varepsilon)$. The following proposition gives elementary but reasonably precise bounds. The first one is sharp when the radius is small, and the second one for a radius slightly smaller than $\pi/2$.

PROPOSITION 5.1. *If $0 \leqslant t \leqslant \pi/2$, then $V(t) \leqslant \frac{1}{2}\sin^{n-1}(t)$. More precisely*

$$(\sqrt{2\pi}\kappa_n)^{-1}(\sin t)^{n-1} \leqslant V(t) \leqslant (\sqrt{2\pi}\kappa_n \cos t)^{-1}(\sin t)^{n-1}, \tag{5.4}$$

where $\kappa_n \sim \sqrt{n}$ is given by (A.8). *Moreover, if $n > 2$, then*

$$V(\pi/2 - t) \leqslant \frac{1}{2}\exp(-nt^2/2). \tag{5.5}$$

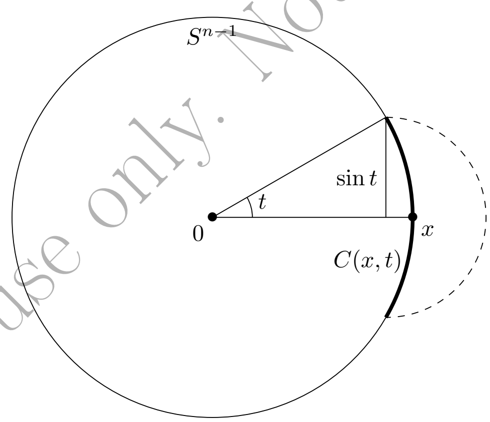

FIGURE 5.2. Proof that $V(t) \leqslant \frac{1}{2}\sin^{n-1}(t)$. The surface area of $C(x,t)$ (bold) does not exceed the surface area of a half-sphere of radius $\sin t$ (dashed).

A proof of (5.4) is sketched in Exercise 5.4. It is based on the fact that, for convex sets, surface area is monotone with respect to inclusion (Exercise 5.2). The inequality (5.5) is from **[Jen13]** (see also **[JS]**); a version with $n-1$ instead of n in the exponent is proved in Exercise 5.3.

The following fact is only marginally used in what follows, but we include it since we did not encounter it in the convexity/functional analysis literature.

PROPOSITION 5.2 (Convavity properties of $V(\cdot)$, see Exercise 5.5). *If $V(r)$ is the measure of a spherical cap of radius r, then the function $t \mapsto \log V(e^t)$ is concave. A fortiori, the function $r \mapsto \log V(r)$ is strictly concave on $[0, \pi]$.*

A consequence of Proposition 5.2 is that, for $0 \leqslant s \leqslant t \leqslant \pi$,

$$V(t) \leqslant \left(\frac{t}{s}\right)^{n-1} V(s). \tag{5.6}$$

Inequality (5.6) is a well-known fact in differential geometry; for example, it constitutes the trivial case of the Gromov–Bishop comparison theorem. It is very likely that Proposition 5.2 also follows from similar general results.

Exercise 5.2 (Surface area is monotone with respect to inclusion). Show that if $K \subset L$ are convex bodies, then $\operatorname{area}(K) \leqslant \operatorname{area}(L)$.

Exercise 5.3. Using Exercise 5.2, show that for $t \in [0, \pi/2]$, we have $V(t) \leqslant \frac{1}{2}\sin^{n-1}(t)$. Conclude that

$$V(\pi/2 - t) \leqslant \frac{1}{2}(\cos t)^{n-1} \leqslant \frac{1}{2}\exp(-(n-1)t^2/2).$$

This is only slightly weaker than the bound (5.5) and sharper than the estimates typically cited in the literature.

Exercise 5.4 (Sharp bounds for volumes of caps). Using Exercise 5.2, show the inequalities (5.4). Then strengthen the lower bound to $(\sqrt{2\pi}\,\kappa_n \cos(t/2))^{-1} \sin^{n-1} t$.

Exercise 5.5 (Convavity properties of $V(\cdot)$). Prove Proposition 5.2 and derive the inequality (5.6).

5.1.2.2. *Nets in the sphere.* If $\varepsilon \in [\pi/2, \pi)$, we clearly have $N(S^{n-1}, g, \varepsilon) = 2$. The interesting case is when $\varepsilon \in (0, \pi/2)$. In that range, the proportion $V(\varepsilon)$ of the sphere covered by a cap of geodesic radius ε decays exponentially with n. It follows that the cardinality of ε-nets grows also exponentially fast. For example, the first estimate from Proposition 5.1 implies that, for $\varepsilon \in (0, \pi/2)$,

$$N(S^{n-1}, g, \varepsilon) \geqslant V(\varepsilon)^{-1} \geqslant \frac{2}{\sin^{n-1}\varepsilon}. \tag{5.7}$$

A basic and extremely useful bound for ε-nets (formulated in the extrinsic distance) is the following

Lemma 5.3. *For every dimension n and every $\varepsilon \leqslant 1$, there is an ε-net in $(S^{n-1}, |\cdot|)$ with less than $(2/\varepsilon)^n$ elements. In other words, $N(S^{n-1}, |\cdot|, \varepsilon) \leqslant (2/\varepsilon)^n$.*

The standard and often quoted volumetric argument (which is a special case of Lemma 5.8 below) gives a slightly worse bound $(1 + 2/\varepsilon)^n$. The improved bound $(2/\varepsilon)^n$ can be achieved by a finer analysis combining a version (based on [**Dum07**]) of Proposition 5.4 below with the use of explicit nets in lower dimensions, see [**Swe**]. We also note that there exist simple *explicit* ε-nets in S^{n-1} with cardinality at most $(C/\varepsilon)^n$ (see Exercise 5.22).

To discuss finer results it is more convenient to switch to the geodesic distance. We know from the volume argument (5.2) that $N(S^{n-1}, g, \varepsilon) \geqslant V(\varepsilon)^{-1}$. It turns out that this trivial estimate is remarkably sharp: an almost-matching upper estimate is provided by an elegant random covering argument due to Rogers.

Proposition 5.4 (Random covering bound). *For every $0 < \eta < \theta$, we have*

$$N(S^{n-1}, g, \theta + \eta) \leqslant \left\lceil \frac{1}{V(\theta)} \log\left(\frac{V(\theta)}{V(\eta)}\right) \right\rceil + \frac{1}{V(\theta)}.$$

PROOF. Let $N = \lceil \frac{1}{V(\theta)} \log(V(\theta)/V(\eta)) \rceil$. Choose $(x_i)_{1 \leqslant i \leqslant N}$ randomly, independently according to σ, and denote $A = \bigcup\{C(x_i, \theta) \,:\, 1 \leqslant i \leqslant N\}$. The expected proportion of the sphere missed by A can be computed using the Fubini–Tonelli theorem

$$\mathbf{E}\sigma(S^{n-1} \backslash A) = (1 - V(\theta))^N \leqslant \exp(-NV(\theta)) \leqslant \frac{V(\eta)}{V(\theta)}. \tag{5.8}$$

In particular, there exist (x_i) such that $\sigma(S^{n-1} \backslash A) \leqslant V(\eta)/V(\theta)$. Let $\{C(y_j, \eta) \,:\, 1 \leqslant j \leqslant M\}$ be a maximal family of disjoint balls of radius η contained in $S^{n-1} \backslash A$. It follows from (5.8) that $M \leqslant 1/V(\theta)$. By construction, S^{n-1} is covered by the family

$$\big\{B(x_i, \theta + \eta) \,:\, 1 \leqslant i \leqslant N\big\} \cup \big\{B(y_j, 2\eta) \,:\, 1 \leqslant j \leqslant M\big\}. \qquad \square$$

COROLLARY 5.5 (Neat random covering bound, see Exercise 5.8). *For every $0 < \varepsilon < \pi/2$, we have*

$$N(S^{n-1}, g, \varepsilon) \leqslant Cn \log n \, V(\varepsilon)^{-1} \tag{5.9}$$

for some absolute constant C.

It follows from (5.7), (5.9) and (5.4) that, for a fixed $\varepsilon \in (0, \pi/2)$, we have

$$\lim_{n \to \infty} \frac{1}{n} \log N(S^{n-1}, g, \varepsilon) = -\log(\sin \varepsilon). \tag{5.10}$$

We note for future reference the following fact.

PROPOSITION 5.6. *Let $P \subset \mathbb{R}^n$ be a polytope such that $d_{BM}(P, B_2^n) \leqslant \lambda$. Then P has at least $2\exp((n-1)/2\lambda^2)$ vertices and at least $2\exp((n-1)/2\lambda^2)$ facets.*

PROOF. Consider first the statement about vertices. Without loss of generality we may assume that $\lambda^{-1} B_2^n \subset P \subset B_2^n$, and that the vertices of P are unit vectors. Let V be the set of vertices of P. The hypothesis is equivalent to saying that V is a θ-net in (S^{n-1}, g) for $\cos\theta = 1/\lambda$ (see Exercise 5.7). Using (5.7), it follows that $\operatorname{card} V \geqslant 2(\sin\theta)^{-(n-1)} \geqslant 2\exp((n-1)/2\lambda^2)$, where we used the inequality $\sin \arccos t \leqslant \exp(-t^2/2)$ for $0 \leqslant t \leqslant 1$. Since $d_{BM}(P, B_2^n) = d_{BM}(P^\circ, B_2^n)$, and since vertices of P° are in bijection with facets of P, the statement about facets follows. $\square$

We also point out that it is possible to approximate the sphere by polytopes with at most exponentially many vertices and, *simultaneously*, at most exponentially many facets (see Exercise 7.22).

EXERCISE 5.6. Check that the constant 2 cannot be replaced by a smaller number in the statement of Lemma 5.3.

EXERCISE 5.7 (Nets and convex hulls). Let $\mathcal{N} \subset S^{n-1}$ and $\theta \in (0, \pi/2)$. Prove that $\mathcal{N}$ is a θ-net in (S^{n-1}, g) if and only if $(\cos\theta) B_2^n \subset \operatorname{conv} \mathcal{N}$.

EXERCISE 5.8 (Proof of the neat random covering bound). Deduce Corollary 5.5 from Proposition 5.4.

EXERCISE 5.9 (On the optimality of Corollary 5.5). Let C_n be the smallest number such that the inequality $N(S^{n-1}, g, \varepsilon) \leqslant C_n V(\varepsilon)^{-1}$ holds for any $\varepsilon > 0$. By considering ε slightly smaller than $\pi/2$, show that $C_n \geqslant \frac{n+1}{2}$. A less trivial fact is that $C_n = \Omega(n)$ is also witnessed by taking ε very close to 0, see **[CFR59]** and Notes and Remarks.

Exercise 5.10 (Nets in the projective space). Prove the following result, which will be useful in Sections 8.1 and 9.4. Let $\varepsilon \in (0, \pi/2)$. If $\mathcal{N}$ is an ε-net in the projective space $\mathbf{P}(\mathbb{C}^d)$ (equipped with the Fubini-Study metric (B.5)), then $\operatorname{card} \mathcal{N} \geqslant (c/\varepsilon)^{2d-2}$ for some absolute positive constant c. In the opposite direction, there exists an ε-net of cardinality not exceeding $(C/\varepsilon)^{2d-2}$.

Exercise 5.11 (Volume of balls in $\mathbf{P}(\mathbb{C}^d)$). Consider the projective space $\mathbf{P}(\mathbb{C}^d)$ equipped with the Fubini-Study metric (B.5) and the invariant probability measure. If $\varepsilon \in (0, \pi/2]$, then the measure of any ball of radius ε in $\mathbf{P}(\mathbb{C}^d)$ is $\sin^{2d-2} \varepsilon$.

5.1.2.3. *Packing on the sphere.* Recall that $P(S^{n-1}, g, \varepsilon)$ is the maximal number of disjoint caps of geodesic radius $\varepsilon/2$. The exact value is known for $\pi/2 \leqslant \varepsilon < \pi$ (we have $P(S^{n-1}, g, \pi/2) = 2n$, see Exercise 5.12) and so we restrict our discussion to the range $0 < \varepsilon < \pi/2$.

Packing problems are usually harder than covering problems. For example, as opposed to (5.10), the exponential rate at which packing numbers increase, i.e., the value of

$$p(\varepsilon) = \limsup_{n \to \infty} \frac{1}{n} \log P(S^{n-1}, g, \varepsilon)$$

is not known for $\varepsilon \in (0, \pi/2)$. We know from (5.2) that $V(\varepsilon)^{-1} \leqslant P(S^{n-1}, g, \varepsilon) \leqslant V(\varepsilon/2)^{-1}$, and therefore

$$-\log \sin(\varepsilon) \leqslant p(\varepsilon) \leqslant -\log \sin(\varepsilon/2). \tag{5.11}$$

In this context the lower bound is known as the Chabauty–Shannon–Wyner bound and actually corresponds to using the trivial algorithm to produce packings: pick separated points, no matter how, as long as you can. It is an amazing fact that the lower bound $p(\varepsilon) \geqslant -\log \sin \varepsilon$ has never been improved: nobody knows how to substantially beat the worst possible choices!

On the other hand, the upper bound in (5.11) has received various improvements. It has been shown by Rankin that for $\varepsilon \in (0, \pi/2)$

$$p(\varepsilon) \leqslant -\log(\sqrt{2} \sin(\varepsilon/2))$$

which matches the lower bound from (5.11) as ε increases to $\pi/2$. For small ε, further improvements due to Kabatjanskiĭ–Levenšteĭn are based on the so-called linear programming bound (see Notes and Remarks).

Exercise 5.12 (Packing large caps on the sphere). Suppose that (x_i) are N points in S^{n-1} such that $\langle x_i, x_j \rangle \leqslant t$ for $i \neq j$.
(i) Show that $N \leqslant 1 - 1/t$ if $t < 0$,
(ii) Show that $N \leqslant 2n$ if $t = 0$
If $t > 0$ is fixed, we know from (5.11) that exponentially many points in the sphere may have pairwise inner products at most t. The situation when t tends to zero with n is investigated in the following exercise.

Exercise 5.13 (Coarse approximation of B_2^n by polytopes with few vertices). Suppose that (x_i) are N points in S^{n-1} such that $|\langle x_i, x_j \rangle| \leqslant t$ whenever $i \neq j$, for some $t > 0$.
(i) If $t < 1/\sqrt{n}$, show that $N \leqslant n/(1 - nt^2)$.
(ii) By considering the family $(x_i^{\otimes k})_{1 \leqslant i \leqslant N}$ for a suitable large k, show that if $t \leqslant 1/2$, then $N \leqslant (C/t)^{Ct^2 n}$ for some absolute constant C.

(iii) Deduce that, for $r \geqslant 2$, there is a polytope P with at most $(Cr)^{Cn/r^2}$ vertices such that $d_g(P, B_2^n) \leqslant r$.

5.1.3. Nets and packings in the discrete cube. Although the discussion from the previous sections dealt specifically with spheres, some ideas carry over directly to other settings. As an illustration we consider the case of the *discrete cube* $\{0,1\}^n$ (a.k.a. *Boolean cube*) equipped with the normalized *Hamming distance*

$$d_H(x,y) = \frac{1}{n}\,\mathrm{card}\{i \ : \ x_i \neq y_i\}. \tag{5.12}$$

We denote by $V(t)$ the volume (i.e., the cardinality) of a ball of radius $t \in (0,1)$. We have

$$V(t) = \mathrm{card}\left\{y \in \{0,1\}^n \ : \ d_H(x,y) \leqslant t\right\} = \sum_{k=0}^{\lfloor tn \rfloor} \binom{n}{k}.$$

The quantity $V(t)$ is governed by the binary entropy function H defined for $x \in (0,1)$ by $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$. For $t \leqslant 1/2$ such that tn is an integer, we have (see Exercise 5.15)

$$\frac{1}{n+1}\,2^{nH(t)} \leqslant V(t) \leqslant 2^{nH(t)}. \tag{5.13}$$

Related estimates will be used when discussing concentration of measure, see (5.59).

As in the case of the sphere, the covering problem is simpler than the packing problem (at least in some asymptotic regimes). In particular (see Exercise 5.14), a random covering argument similar to Proposition 5.4—in combination with (5.13)—implies that, for $0 < \varepsilon < 1/2$,

$$\lim_{n\to\infty} \frac{1}{n}\log_2 N(\{0,1\}^n, d_H, \varepsilon) = 1 - H(\varepsilon). \tag{5.14}$$

On the other hand, the corresponding limit for packing is unknown; we only get from (5.2) the asymptotic bounds

$$1 - H(\varepsilon) \leqslant \limsup_{n\to\infty} \frac{1}{n}\log_2 P(\{0,1\}^n, d_H, \varepsilon) \leqslant 1 - H(\varepsilon/2) \tag{5.15}$$

for $0 < \varepsilon < 1/2$. As in the case of the sphere, the lower bound from (5.15) (known in this context as the Gilbert–Varshamov bound) has not been improved, while the upper bound has been subject to various enhancements.

For the q-ary version of the cube, i.e., the space $\{0,\dots,q-1\}^n$ (also equipped with normalized Hamming distance), the entropy function has to be replaced by

$$H_q(x) := -x\log_q x - (1-x)\log_q(1-x) + x\log_q(q-1).$$

Indeed, if $V_q(t)$ denotes the cardinality of a ball of radius t in $\{1,\dots,q-1\}^n$, for $t \in (0, 1-1/q)$ such that tn is an integer, then

$$\frac{1}{n+1}\,q^{nH_q(t)} \leqslant V_q(t) \leqslant q^{nH_q(t)}. \tag{5.16}$$

Estimates about the q-ary cube are useful when one wants to construct nets or separated sets in products of metric spaces. The following specific fact, which is an easy consequence of (5.16) and (5.1), will be used later.

PROPOSITION 5.7. *Let (K,d) be a metric space such that $P(K,d,\varepsilon) \geqslant q$. Given integer $n \in \mathbb{N}$, equip K^n with the distance*

$$d_n((x_1,\dots,x_n),(y_1,\dots,y_n)) = d(x_1,y_1) + \cdots + d(x_n,y_n).$$

Then, for $t \in (0, 1-1/q)$,

$$P(K^n, d^n, t\varepsilon n) \geqslant P(\{0,\dots,q-1\}^n, d_H, t) \geqslant \frac{q^n}{V_q(t)} \geqslant q^{n(1-H_q(t))}. \tag{5.17}$$

EXERCISE 5.14 (Efficient random nets of the Boolean cube). Show (5.14) by adapting the random covering argument from Proposition 5.4.

EXERCISE 5.15 (Volume of balls in the q-ary discrete cube). Show (5.16) (which specified to $q = 2$ gives (5.13)).

5.1.4. Metric entropy for convex bodies. If the metric space (M,d) is actually a normed space with a unit ball B, we write $N(K,B,\varepsilon)$ or $N(K,\varepsilon B)$ instead of $N(K,d,\varepsilon)$. It is possible to come up with an alternative definition which does not refer to the norm, by saying that $N(K,B,\varepsilon)$ is the minimum number N such that there exist $x_1,\dots,x_N$ in K with

$$K \subset \bigcup_{i=1}^{N} (x_i + \varepsilon B). \tag{5.18}$$

This alternative definition does not require the set B to be symmetric, or even convex, or to have nonempty interior, even though that is usually the case. In our context, the minimal reasonable hypothesis appears to be asking that B be *star-shaped* with respect to the origin, i.e., that $tB \subset B$ for $t \in [0,1]$.

The technology for estimating covering/packing numbers of subsets (particularly convex subsets) of normed spaces is quite well-developed and frequently rather sophisticated. We quote here a simple well-known result that expresses $N(\cdot,\cdot)$ in terms of a "volume ratio."

LEMMA 5.8. *Let L be a symmetric convex body in $\mathbb{R}^n$ and let $K \subset \mathbb{R}^n$ be a Borel set. Then, for any $\varepsilon > 0$,*

$$\left(\frac{1}{\varepsilon}\right)^n \frac{\mathrm{vol}(K)}{\mathrm{vol}(L)} \leqslant N(K,L,\varepsilon) \leqslant \left(\frac{2}{\varepsilon}\right)^n \frac{\mathrm{vol}(K+\frac{\varepsilon}{2}L)}{\mathrm{vol}(L)}. \tag{5.19}$$

PROOF. If (x_i) is an ε-net in K with respect to $\|\cdot\|_L$, then the union of the sets $x_i + \varepsilon L$ contains K, and the left-hand side inequality in (5.19) follows from volume comparison. Consider now a family (x_i) of N elements of K which is ε-separated for $\|\cdot\|_L$. This means that the sets $x_i + \frac{\varepsilon}{2}L$ have disjoint interiors. Since they are all included in $K + \frac{\varepsilon}{2}L$, we have $N\,\mathrm{vol}(\frac{\varepsilon}{2}L) \leqslant \mathrm{vol}(K + \frac{\varepsilon}{2}L)$. Together with (5.1), this implies the right-hand side inequality in (5.19) □

When K is convex and the "regularizing" trick implicit in Exercise 5.17 below is applied, the lower and upper bounds are often as close as one can expect provided K and L are is the M-position (see Notes and Remarks). The case $K = L$ in Lemma 5.8 is related to the approximation of convex bodies by polytopes.

LEMMA 5.9. *Let $0 < \varepsilon < 1$, $K \subset \mathbb{R}^n$ be a symmetric convex body and $\mathcal{N}$ be an ε-net in K with respect to $\|\cdot\|_K$. Then $\mathrm{conv}\,\mathcal{N} \supset (1-\varepsilon)K$.*

Proof. Let $P = \operatorname{conv} \mathcal{N}$ and denote $A = \sup\{\|y\|_P \ : \ y \in K\}$. One checks that P contains 0 in the interior, so that $A < \infty$. Given $x \in K$, there is $x' \in \mathcal{N}$ such that $\|x - x'\|_K \leqslant \varepsilon$, and therefore $\|x\|_P \leqslant \|x'\|_P + \|x - x'\|_P \leqslant 1 + \varepsilon A$. Taking supremum over x gives $A \leqslant 1 + \varepsilon A$, so that $A \leqslant (1-\varepsilon)^{-1}$, which is equivalent to the inclusion $P \supset (1-\varepsilon)K$. □

The following is an immediate consequence of Lemmas 5.8 and 5.9.

Corollary 5.10. *Let $\varepsilon \in (0,1)$. Any symmetric convex body in $\mathbb{R}^n$ is $(1-\varepsilon)^{-1}$-close, in the Banach–Mazur distance, to a polytope with at most $(1+2/\varepsilon)^n$ vertices.*

For an extension of Lemma 5.9 and 5.10 to not-necessarily-symmetric convex bodies, see Exercises 5.18–5.20. Note that the dependence on ε in Corollary 5.10 is not sharp (see Notes and Remarks). For the special case $K = B_2^n$, the conclusion of Lemma 5.9 can be easily improved to $\operatorname{conv} \mathcal{N} \supset (1 - \varepsilon^2/2)K$, see Exercise 5.7.

Exercise 5.16 (Covering with balls whose centers lie outside of the set). For convex bodies K, L in $\mathbb{R}^n$, let $N'(K, L)$ be the smallest number N such that there exist $x_1, \dots, x_N$ in $\mathbb{R}^n$ with $K \subset \bigcup_{1 \leqslant i \leqslant N}(x_i + L)$ (the difference with $N(K, L)$ is that x_i are not required to belong to K). Give an example with L symmetric for which $N'(K, L) < N(K, L)$. Can we have such an example with also K symmetric?

Exercise 5.17 (A regularizing trick). Let K, L be convex bodies in $\mathbb{R}^n$, with $0 \in L$. Show that $N(K, \varepsilon L) = N(K, (K-K) \cap \varepsilon L)$.

Exercise 5.18 (Approximating by polytopes with few vertices). Let $K \subset \mathbb{R}^n$ be a convex body with centroid at the origin (K is not assumed to be symmetric). Using Lemma 5.8 and Proposition 4.18, show that for every $\varepsilon \in (0,1)$ we have $N(K, \varepsilon K) \leqslant (2 + 4/\varepsilon)^n$, where $N(K, \varepsilon K) = N(K, K, \varepsilon)$ is defined as in (5.18). By arguing as in the proof of Lemma 5.9, conclude that there exists a polytope P with at most $(2+4/\varepsilon)^n$ vertices such that $(1-\varepsilon)K \subset P \subset K$.

Exercise 5.19 (Approximating by polytopes with few facets). Let $\varepsilon \in (0,1)$ and $K \subset \mathbb{R}^n$ be a convex body with centroid at the origin. Show that there exists a polytope Q with at most $(2+4/\varepsilon)^n$ facets such that $(1-\varepsilon)Q \subset K \subset Q$.

Exercise 5.20 (Approximating by polytopes and the Santaló inequality). Let K be a convex body in $\mathbb{R}^n$ and let $\kappa = \operatorname{vrad}(K)\operatorname{vrad}(K^\circ) < \infty$ (i.e., K satisfies approximately the Santaló inequality, see Theorem 4.17 and the comments following it). If $\varepsilon \in (0,1)$, then K can be approximated up to ε (in the sense of Exercises 5.18 and 5.19) by a polytope P with at most $(C\kappa/\varepsilon)^n$ vertices (resp., facets).

Exercise 5.21 (Duality of metric entropy for ellipsoids). Let $\mathscr{E}$ and $\mathscr{F}$ be 0-symmetric ellipsoids in $\mathbb{R}^n$. Check that for every $\varepsilon > 0$, $N(\mathscr{E}, \mathscr{F}, \varepsilon) = N(\mathscr{F}^\circ, \mathscr{E}^\circ, \varepsilon)$.

Exercise 5.22 (Explicit nets in S^{n-1}). Here is an explicit construction of an ε-net in S^{n-1} with at most $(C/\varepsilon)^n$ elements, for some (suboptimal) constant C.
(i) Show that, if $\mathcal{N}$ is an ε-net in B_2^n (with $0 < \varepsilon < 1$), then the set $\{x/|x| \ : \ x \in \mathcal{N}\}$ is an η-net in $(S^{n-1}, |\cdot|)$ for $\eta = \sqrt{2 - 2\sqrt{1-\varepsilon^2}}$.
(ii) Let $\mathcal{N} = B_2^n \cap \frac{\varepsilon}{\sqrt{n}}\mathbb{Z}^n$. Show that $\mathcal{N}$ is an ε-net in B_2^n and that $\operatorname{card} \mathcal{N} \leqslant (C/\varepsilon)^n$.

5.1.5. Nets in Grassmann manifolds, orthogonal and unitary group. We now extend the results given for the sphere to other classical manifolds, including unitary and orthogonal groups and Grassmann manifolds (which are introduced in Appendix B). Metric structures on such manifolds are induced by unitarily invariant norms on the corresponding matrix spaces, with Schatten p-norms being the most popular choices. While there are several natural ways (also discussed in detail in Appendix B) to define a metric on a manifold starting from a given Schatten norm, all such metrics—for a fixed p—differ by at most by a multiplicative factor of $\pi/2$. Accordingly, the behavior of covering numbers in all such situations can be subsumed in the following single statement.

THEOREM 5.11 (not proved here, but see Exercise 5.23). *Let M be either $\mathsf{SO}(n)$, $\mathsf{U}(n)$, $\mathsf{SU}(n)$, $\mathsf{Gr}(k,\mathbb{R}^n)$ or $\mathsf{Gr}(k,\mathbb{C}^n)$, equipped with a metric generated by the Schatten norm $\|\cdot\|_p$ for some $1 \leqslant p \leqslant \infty$. Then for any $\varepsilon \in (0, \operatorname{diam} M]$,*

$$\left(\frac{c \operatorname{diam} M}{\varepsilon}\right)^{\dim M} \leqslant N(M,\varepsilon) \leqslant \left(\frac{C \operatorname{diam} M}{\varepsilon}\right)^{\dim M}, \tag{5.20}$$

where $C, c > 0$ are universal constants (independent of n, k, p and ε), $\dim M$ is the real dimension of M, and $\operatorname{diam} M$ the diameter of M with respect to the corresponding metric.

For easy reference, we list in Table 5.1 some of the values of the parameters (dimensions, diameters) that appear in (5.20).

TABLE 5.1. Real dimensions and diameters from the bounds (5.20) for covering numbers of a selection of classical manifolds. The distances used on $\mathsf{SO}(n)$ and $\mathsf{U}(n)$ are the extrinsic metrics obtained from the Schatten p-norm on M_n, and the distances on Grassmann manifolds are the corresponding quotient metrics. The restriction $k \leqslant n/2$ is imposed to reduce clutter (note that $\mathsf{Gr}(k,\mathbb{R}^n)$ and $\mathsf{Gr}(n-k,\mathbb{R}^n)$ are isometric).

M	$\dim M$	$\operatorname{diam} M$	comments
$\mathsf{SO}(n)$	$n(n-1)/2$	$2n^{1/p}$	
$\mathsf{U}(n)$	n^2	$2n^{1/p}$	
$\mathsf{Gr}(k,\mathbb{R}^n)$	$k(n-k)$	$2^{1/2}(2k)^{1/p}$	$k \leqslant n/2$
$\mathsf{Gr}(k,\mathbb{C}^n)$	$2k(n-k)$	$2^{1/2}(2k)^{1/p}$	$k \leqslant n/2$

EXERCISE 5.23 (Metric entropy of classical groups and manifolds). Prove Theorem 5.11 for $M = \mathsf{U}(n)$, $M = \mathsf{SU}(n)$ or $M = \mathsf{SO}(n)$ and for $p = \infty$, by appealing to Lipschitz properties of the exponential map with matrix argument (Exercise B.8).

EXERCISE 5.24. Derive the formula for diameter of $\mathsf{Gr}(k,\mathbb{R}^n)$ in Table 5.1.

EXERCISE 5.25 (Volume of balls in classical groups and manifolds). Let M be either $\mathsf{SO}(n), \mathsf{U}(n)$ or $\mathsf{Gr}(k,\mathbb{R}^n)$, equipped with a metric as in Theorem 5.11. Denoting by σ the Haar probability measure on M, deduce from Theorem 5.11 a two-sided estimate for $\sigma(B(x,\varepsilon))$, where $B(x,\varepsilon)$ denotes the ball of radius ε centered at $x \in M$.

5.2. Concentration of measure

The classical isoperimetric inequality in $\mathbb{R}^n$ (Eq. (4.27), also known as Dido's problem) states that among all sets of given volume, the Euclidean balls have the smallest surface area. As we already noticed in the setting of $\mathbb{R}^n$ in Section 4.3.1, an alternative methodology is to consider, instead of the surface area, the family of ε-enlargements of a given set. The latter approach makes sense in any metric space X equipped with a measure μ (a *metric measure space*, or a *metric probability space* if $\mu(X) = 1$, which will be assumed as a default): for a subset $A \subset X$ and $\varepsilon > 0$, we define

$$A_\varepsilon = \{x \in X \ : \ \mathrm{dist}(x, A) \leqslant \varepsilon\}.$$

The two viewpoints are roughly equivalent since the "surface area" relative to μ can be retrieved (when that makes sense) as the first-order variation of $\mu(A_\varepsilon)$ when ε goes to 0, cf. (4.23) and, conversely, the growth of the function $\varepsilon \mapsto \mu(A_\varepsilon)$ on the macroscopic scale can be recovered from the knowledge of its derivative. However, the enlargement-based approach seems simpler (a more flexible definition) and is often more fruitful since some otherwise useful bounds on $\mu(A_\varepsilon)$ may be meaningless for small ε, and/or may be available in absence of any clue with regard to the nature of extremal sets.

Lower bounds for $\mu(A_\varepsilon)$ can be rephrased as deviation inequalities for Lipschitz functions. This leads, in some settings, to a remarkable phenomenon: every Lipschitz function concentrates strongly around some "central value." Statements to such and similar effect will be the focus of our presentation. Specifically, we will look for estimates of the form

$$\mu(f > M_f + t) \leqslant Ce^{-\lambda t^2} \tag{5.21}$$

and

$$\mu(f > \mathbf{E}f + t) \leqslant Ce^{-\lambda t^2}, \tag{5.22}$$

to be valid for any real-valued 1-Lipschitz function on X and all $t > 0$, where M_f and $\mathbf{E}f$ are the *median* and the *expected value* of f calculated with respect to μ. (A number M is said to be a median for a random variable X if $\mathbf{P}(X \geqslant M) \geqslant 1/2$ and $\mathbf{P}(X \leqslant M) \geqslant 1/2$.) Clearly, (5.21) and (5.22) formally imply then similar *two-sided* estimates for $\mu(|f - M_f| > t)$ and $\mu(|f - \mathbf{E}f| > t)$ with C replaced by $2C$. Concentration of this type is referred to as *subgaussian* (more on this terminology in Section 5.2.6). For the convenience of a casual reader—and for easy reference—we list in Table 5.2 the constants and the exponents that appear in subgaussian concentration inequalities for a selection of classical objects.

REMARK 5.12. We point out that if a function f is such that one of the inequalities (5.21) or (5.22) holds (for all $t > 0$) with constants C, λ, then the other inequality similarly holds (for the same function) with some other constants. For example, if (5.22) holds with $C \geqslant \frac{1}{2}$ and λ, then (5.21) holds with $2C^2$ and $\lambda/2$; if (5.21) holds with $C \geqslant e^{-1/3} \approx 0.717$ and λ, then (5.21) holds with eC^2 and $\lambda/2$ (see Proposition 5.29 and Remarks 5.30, 5.31.) Sharper results of this nature (i.e., with better dependence on C, λ) can sometimes be obtained if we assume that (5.21) (or (5.22)) holds for *all* real-valued 1-Lipschitz functions on X; some questions in that spirit are considered in **[Led01]** (see, e.g., Exercise 5.48).

Table 5.2. Constants and exponents in subgaussian concentration inequalities for a selection of classical objects. When applicable, the reference measure is the canonical invariant measure on the object in question. We made an effort to come up with reasonable values of constants/exponents, and some of them are optimal. Unless indicated otherwise, the metric used for manifolds is the Riemannian geodesic distance. d_H stands for the normalized Hamming distance (5.12). References: (a) Theorem 5.24. (b) Log-Sobolev inequality (LSI), see Table 5.4. (c) Corollary 5.17. (d) Proposition 5.20; what follows from the LSI is $\lambda = \frac{n-1}{2}$. (e) Ricci curvature, see Table 5.3. (f) Remark 5.12. (g) Corollary 5.52. (h) LSI on the discrete cube, see Theorem 5.1 and Exercise 5.5 in [**BLM13**]. (i) Theorem 5.54; convex or concave functions only. (j) The constant in the exponent is $\frac{1}{8}$ and not $\frac{1}{2}$ due to rescaling ($\{-1,1\}$ vs. $\{0,1\}$) (k) Theorem 5.56; convex functions only. (l) Theorem 5.38. (m) Theorem 5.39. (n) $(C,\lambda) = (2,\frac{1}{4})$ if $n = 2$; Remark 5.12. (o) Exercise 5.54. (p) Remark 5.19. (q) If we use instead the non-Riemannian metric (B.11), the parameter λ needs to be multiplied by 2 in view of (B.12). (r) Remark 5.53.

Object	C,λ in (5.21)–median	C,λ in (5.22)–mean	Comments
Gauss space $(\mathbb{R}^n, \lvert\cdot\rvert, \gamma_n)$	$\frac{1}{2}, \frac{1}{2}$ (a)	$1, \frac{1}{2}$ (b)	
Gauss space $(\mathbb{C}^n, \lvert\cdot\rvert, \gamma_n^{\mathbb{C}})$	$\frac{1}{2}, 1$ (a)	$1, 1$ (b)	
(S^{n-1}, g) or $(S^{n-1}, \lvert\cdot\rvert)$	$\frac{1}{2}, \frac{n}{2}$ (c)	$1, \frac{n}{2}$ (d)	$n > 2$ for (S^{n-1}, g) (p)
$\mathsf{SO}(n)$	$\frac{1}{2}, \frac{n-1}{8}$ (e)	$1, \frac{n-1}{8}$ (b)	metric (B.8)
$\mathsf{SU}(n)$	$\frac{1}{2}, \frac{n}{4}$ (e)	$1, \frac{n}{4}$ (b)	metric (B.8)
$\mathsf{U}(n)$	$2, \frac{n}{24}$ (f)	$1, \frac{n}{12}$ (b)	metric (B.8)
$\mathsf{Gr}(k, \mathbb{R}^n)$	$\frac{1}{2}, \frac{n-2}{4}$ (e)	$1, \frac{n-2}{4}$ (b)	metric (B.10) (q)
$\mathsf{Gr}(k, \mathbb{C}^n)$	$\frac{1}{2}, \frac{n}{2}$ (e)	$1, \frac{n}{2}$ (b)	metric (B.10) (q)
$(\{-1,1\}^n, d_H)$	$1, 2n$ (g)	$1, 2n$ (h)	$n \geqslant 3$ (r)
$(\{-1,1\}^n, \lvert\cdot\rvert)$	$2, \frac{1}{8}$ (i)(j)	$1, \frac{1}{8}$ (k)(j)	appropriate convexity hypotheses
Ricci curvature $\geqslant c$	$\frac{1}{2}, \frac{c}{2}$ (l)	$1, \frac{c}{2}$ (b)	
LSI with constant $\leqslant \alpha$	$2, \frac{1}{4\alpha}$ (f)	$1, \frac{1}{2\alpha}$ (m)	
$(S^{n-1})^k$	$\frac{1}{2}, \frac{n-2}{2}$ (e)(n)	$1, \frac{n-1}{2}$ (b)	ℓ_2 product metric
$[0,1]^k$	$\frac{1}{2}, \pi$ (o)	$1, \frac{\pi^2}{2}$ (b)	ℓ_2 product metric

In the next two subsections we will exemplify the concentration phenomenon and related techniques in the case of the Euclidean sphere and the Gaussian space. In subsequent subsections we will survey some general methods for proving isoperimetric/concentration results and present a selection of examples, in particular those listed in Table 5.2. We will concentrate on the objects that exhibit subgaussian concentration; more general settings will be addressed briefly in exercises and in Notes and Remarks (an exception is Section 5.2.6 which treats sums of independent subexponential random variables). A comprehensive presentation of diverse aspects and manifestations of the concentration phenomenon is beyond the scope of this work; we refer the interested reader to the monographs [**Led01, BLM13**] and/or to other sources listed in Notes and Remarks. Here we restrict our attention to highlighting several central techniques and, subsequently, to going over examples that appear to be of relevance to the quantum theory.

5.2.1. A prime example: concentration on the sphere. The settings of the Euclidean sphere and of the projective space are directly relevant to quantum information theory since the latter identifies canonically with the set of pure states. In the language of enlargements, the isoperimetric inequality on the sphere can be stated as follows.

THEOREM 5.13 (Spherical isoperimetric inequality, not proved here). *Equip the unit sphere $S^{n-1} \subset \mathbb{R}^n$ with the geodesic distance g and the uniform probability measure σ. If $A \subset S^{n-1}$ and if $C \subset S^{n-1}$ is a spherical cap such that $\sigma(A) = \sigma(C)$, then, for any $\varepsilon > 0$,*

$$\sigma(A_\varepsilon) \geqslant \sigma(C_\varepsilon). \tag{5.23}$$

Recall that the spherical cap with center $x \in S^{n-1}$ and radius ε is the set

$$C(x,\varepsilon) = \{y \in S^{n-1} \ : \ g(x,y) \leqslant \varepsilon\}.$$

Note that the class of spherical caps is stable under enlargements and that we have

$$C(x,\varepsilon)_\delta = C(x,\varepsilon+\delta) \quad \text{for any} \quad \delta,\varepsilon > 0. \tag{5.24}$$

In view of the simple relationship between g and the extrinsic (or chordal) distance inherited from the ambient Euclidean space (see Appendix B.1), Theorem 5.13 is valid also for the latter. However, it is traditionally stated for the geodesic distance. Also, the formula (5.24) for $C(x,\varepsilon)_\delta$ stated above would be more complicated if we used $|\cdot|$ to define caps.

The usefulness of Theorem 5.13 comes from the fact that there are explicit integral formulas and sharp bounds for the measure of spherical caps, which were explored in Section 5.1.2. However, while in the study of packing and covering small caps seemed most interesting, in the present context of concentration the radii close to $\pi/2$ are most relevant. This is because arguably the most useful instance of Theorem 5.13 is $\sigma(A) = \frac{1}{2}$, in which case the radius of the corresponding cap C is $\pi/2$ and the radius of its ε-enlargement, C_ε, is $\pi/2 + \varepsilon$. Taking into account the bound (5.5) leads then to

COROLLARY 5.14. *If $n > 2$ and if $A \subset S^{n-1}$ with $\sigma(A) \geqslant \frac{1}{2}$ and $\varepsilon > 0$, then*

$$\sigma(A_\varepsilon) \geqslant \sigma\left(C\left(x, \frac{\pi}{2} + \varepsilon\right)\right) \geqslant 1 - \frac{1}{2}e^{-n\varepsilon^2/2}. \tag{5.25}$$

There is no simple proof of the isoperimetric inequality on the sphere (Theorem 5.13) that we know of. However, a result just slightly weaker than Corollary 5.14 follows easily from the Brunn–Minkowski inequality (4.21). We have the following

PROPOSITION 5.15. *If* $\varepsilon \in (0, \pi/2]$ *and* $K, L \subset S^{n-1}$ *are such that* $\operatorname{dist}(K, L) \geqslant \varepsilon$ *(in the geodesic distance), then* $\sigma(K)\sigma(L) \leqslant e^{-n\varepsilon^2/4}$. *In particular, if* $\sigma(K) \geqslant 1/2$, *then* $\sigma(K_\varepsilon) \geqslant 1 - 2e^{-n\varepsilon^2/4}$.

PROOF. The second statement follows by applying the first one with $L = K_\varepsilon^c$. It thus remains to prove the first statement.

Define $K' \subset B_2^n$ via $K' := \{tx : x \in K, t \in [0,1]\}$ and similarly for L'. Then $\operatorname{vol}(K') = \sigma(K)\operatorname{vol}(B_2^n)$ and $\operatorname{vol}(L') = \sigma(L)\operatorname{vol}(B_2^n)$. Consequently, by the Brunn–Minkowski inequality in the form (4.21),

$$\operatorname{vol}\left(\frac{K' + L'}{2}\right) \geqslant \sqrt{\operatorname{vol}(K')\operatorname{vol}(L')} = \sqrt{\sigma(K)\sigma(L)}\ \operatorname{vol}(B_2^n).$$

On the other hand, if $x, y \in S^{n-1}$ and the angle between x and y is at least ε, then $|(x+y)/2| \leqslant \cos(\varepsilon/2)$. If $\varepsilon \leqslant \pi/2$ (and so $\langle x, y\rangle \geqslant 0$), a simple calculation shows that the same is true if we replace x and y by $x' = sx$ and $y' = ty$, where $s, t \in [0,1]$ (in fact this is even true if $\varepsilon \leqslant 2\pi/3$). This means that we have then $\frac{K'+L'}{2} \subset \cos(\varepsilon/2)B_2^n$ and so $\sqrt{\sigma(K)\sigma(L)} \leqslant (\cos(\varepsilon/2))^n$. It remains to appeal to the (subtle but elementary) inequality $\cos u \leqslant e^{-u^2/2}$ (see Exercise 5.3). □

REMARK 5.16. (1) Proposition 5.15 holds actually for the entire nontrivial range of ε, which is $[0, \pi]$; this follows *a posteriori* from the estimate in Lévy's lemma (see Exercise 5.26). The above proof fails for large ε; however, only the range $[0, \pi/2]$ is relevant to the second statement and to Corollary 5.14: if $\mu(K) \geqslant 1/2$, then no point x can verify $\operatorname{dist}(x, K) > \pi/2$.
(2) The estimate in the Proposition is pretty tight: if K, L are opposite (i.e., $K = -L$) caps with $\operatorname{dist}(K, L) = 2\varepsilon$, we conclude from the Proposition that $\mu(K) \leqslant e^{-n\varepsilon^2/2}$. This compares fairly well with the bound $\frac{1}{2}e^{-n\varepsilon^2/2}$ implicit in (5.25).

Corollary 5.14 readily implies a concentration result for Lipschitz functions, which is often referred to in quantum information circles as Lévy's lemma.

COROLLARY 5.17 (Lévy's lemma). *Let* $n > 2$. *If* $f : (S^{n-1}, g) \to \mathbb{R}$ *is a* L*-Lipschitz function and if* M_f *is a median for* f, *then, for any* $t > 0$,

$$\sigma(f > M_f + t) \leqslant \frac{1}{2}\exp(-nt^2/2L^2), \tag{5.26}$$

and therefore

$$\sigma(|f - M_f| > t) \leqslant \exp(-nt^2/2L^2). \tag{5.27}$$

PROOF. Let $A = \{x \in S^{n-1} : f(x) \leqslant M_f\}$ and set $\varepsilon = t/L$. Since $f \leqslant M_f$ on A and since f is L-Lipschitz (i.e., $|f(x) - f(y)| \leqslant Lg(x,y)$ for $x, y \in S^{n-1}$), it follows that for any $y \in S^{n-1}$ we have $f(y) \leqslant M_f + Lg(y, A)$. In particular, if $y \in A_\varepsilon$, then $g(y, A) \leqslant \varepsilon$ and so $f(y) \leqslant M_f + L\varepsilon = M_f + t$. In other words, we proved that $A_\varepsilon \subset \{f \leqslant M_f + t\} = \{f > M_f + t\}^c$. The first inequality in Corollary 5.17 follows now by observing that, by the definition of the median, $\sigma(A) \geqslant \frac{1}{2}$ and by appealing to Corollary 5.14.

The second inequality follows from the first one combined with an identical bound on $\sigma(f < M_f - t)$, which is shown either by the same argument applied to

$A = \{x \in S^{n-1} : f(x) \geqslant M_f\}$, or by appealing to the first inequality with f replaced by $-f$. □

REMARK 5.18. Both parts of the above proof are quite general. First, any lower bounds on measures of enlargements of sets of measure $\frac{1}{2}$ imply (in fact are equivalent to, see Exercise 5.27) bounds for deviation of Lipschitz function from their medians. Second, any one-sided bound for deviation from the median (or the expected value, or any other "symmetric" parameter) implies a two-sided bound, at the cost of a factor of 2.

REMARK 5.19. In Corollaries 5.14 and 5.17 we have to assume that $n > 2$ because the bound (5.5) is not valid in the entire nontrivial range $0 \leqslant t \leqslant \pi/2$. If $n = 2$, one needs to replace the function $\frac{1}{2}e^{-nt^2/2}$ by $\max\{\frac{1}{2} - \frac{t}{\pi}, 0\}$. However, no modifications are needed if the enlargements or the Lipschitz constants are calculated with respect to the ambient space metric, or if only small values of ε or t are of interest, say, $\varepsilon \leqslant 1$ or $t \leqslant L$.

Concentration around the median follows naturally from the isoperimetric inequality. As we mentioned in Remark 5.12, this implies formally concentration around the expectation with altered constants. In some situations, it is possible to obtain good constants with extra work.

PROPOSITION 5.20 (Lévy's lemma for the mean, not proved here). *Let $n > 2$. If $f : (S^{n-1}, g) \to \mathbb{R}$ is a 1-Lipschitz function, then for any $t > 0$,*

$$\sigma(f > \mathbf{E}f + t) \leqslant \exp(-nt^2/2). \tag{5.28}$$

As mentioned in Remark 5.18, the inequality $\sigma(|f - \mathbf{E}f| > t) \leqslant 2\exp(-nt^2/2)$ follows formally, but is probably not optimal. See Problem 5.26 for questions about possible better bounds in this and similar settings.

EXERCISE 5.26 (Proposition 5.15 holds for the full range of ε). Show that it follows *a posteriori* from Theorem 5.13 and the bound (5.5) that, for $n > 2$, in the notation and under the hypotheses of Proposition 5.15, we have $\sigma(K)\,\sigma(L) \leqslant \frac{1}{4}e^{-n\varepsilon^2/4}$. For $n = 2$, the optimal inequality is $\sigma(K)\,\sigma(L) \leqslant \frac{1}{4}\big(1 - \frac{\varepsilon}{\pi}\big)^2$ (cf. Remark 5.19).

EXERCISE 5.27 (Concentration implies isoperimetry). Show that, for a metric probability space (X, μ), concentration implies isoperimetry in the following sense: if $\mu(f > M_f + t) \leqslant \alpha$ for any 1-Lipschitz function f, then $\mu(A_t) \geqslant 1 - \alpha$ for any $A \subset X$ with $\mu(A) = \frac{1}{2}$.

EXERCISE 5.28 (A finer bound tor the mean width of a union). Let K, L be two bounded sets in $\mathbb{R}^n$, and R the outradius of $K \cup L$. Show that $w(\mathrm{conv}(K \cup L)) \leqslant \max(w(K), w(L)) + \sqrt{\frac{2\pi}{n}}\, R$.

5.2.2. Gaussian concentration. Another classical setting where isoperimetry and concentration have been widely studied is the Gaussian space $\big(\mathbb{R}^n, |\cdot|, \gamma_n\big)$, where γ_n is the standard Gaussian measure on $\mathbb{R}^n$ (see Appendix A.2 for the notation, basic properties and relevant facts). It turns out that the extremal sets for the isoperimetric problem are then half-spaces, and since their enlargements are also half-spaces, the solution to the problem can be expressed simply in terms of the cumulative distribution function of an $N(0,1)$ variable, i.e., in terms of $\Phi(x) := \gamma_1((-\infty, x])$. We have

Theorem 5.21 (Gaussian isoperimetric inequality, see Exercise 5.30). *Let $A \subset \mathbb{R}^n$, and let $a \in \mathbb{R}$ be defined by $\gamma_1\big((-\infty, a]\big) = \gamma_n(A)$. Then, for any $\varepsilon > 0$,*

$$\gamma_n(A_\varepsilon) \geqslant \gamma_1\big((-\infty, a+\varepsilon]\big) \tag{5.29}$$

or, equivalently,

$$\Phi^{-1}(\gamma_n(A_\varepsilon)) \geqslant \Phi^{-1}(\gamma_n(A)) + \varepsilon. \tag{5.30}$$

The solution to the Gaussian isoperimetric problem (Theorem 5.21) was originally derived from the spherical isoperimetric inequality (Theorem 5.13) via the following classical fact.

Theorem 5.22 (Poincaré's lemma, see Exercise 5.29). *For $n, N \in \mathbb{N}$ with $N \geqslant n$, we consider $\mathbb{R}^n$ to be a subspace of $\mathbb{R}^N$. Next, fix n and let ν_N be the pushforward to $\mathbb{R}^n$, via the orthogonal projection, of the normalized uniform measure on $\sqrt{N}S^{N-1}$. Then, as $N \to \infty$, (ν_N) converges to γ_n, the standard Gaussian measure on $\mathbb{R}^n$.*

The convergence in Theorem 5.22 holds in a very strong sense, e.g., in total variation, or in uniform convergence of densities.

Another derivation of the Gaussian isoperimetric inequality is based on the following analogue of the Brunn–Minkowski inequality in the Gaussian setting.

Theorem 5.23 (Ehrhard's inequality, not proved here). *Let A, B be Borel subsets of $\mathbb{R}^n$ and let $\lambda \in [0,1]$. Then*

$$\Phi^{-1}(\gamma_n((1-\lambda)A + \lambda B)) \geqslant (1-\lambda)\Phi^{-1}(\gamma_n(A)) + \lambda\Phi^{-1}(\gamma_n(B)). \tag{5.31}$$

Ehrhard's inequality is stronger than log-concavity of the Gaussian measure (Section 4.3.2), see Exercise 5.31. Assuming Ehrhard's inequality, the derivation of the Gaussian isoperimetric inequality goes as follows. Fix A, ε and let $\lambda \in (0,1)$. Since $A_\varepsilon = A + \varepsilon B_2^n = (1-\lambda)(1-\lambda)^{-1}A + \lambda\varepsilon\lambda^{-1}B_2^n$, we have, by (5.31),

$$\Phi^{-1}(\gamma_n(A_\varepsilon)) \geqslant (1-\lambda)\Phi^{-1}(\gamma_n((1-\lambda)^{-1}A)) + \lambda\Phi^{-1}(\gamma_n(\varepsilon\lambda^{-1}B_2^n)). \tag{5.32}$$

We now let $\lambda \to 0^+$. The first term on the right-hand side of (5.32) converges clearly to $\Phi^{-1}(\gamma_n(A))$, while the second term converges to ε (this is a little harder, but elementary, see Exercise 5.32), and so we proved the Gaussian isoperimetric inequality in the form (5.30).

The next theorem follows from Theorem 5.21 according to the general scheme indicated in Remark 5.18, with the explicit exponential bound being a consequence of Exercise A.1.

Theorem 5.24. *If $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz and M_f denotes its median (with respect to γ_n), then for any $t > 0$*

$$\gamma_n(f > M_f + t) \leqslant \gamma_1\big((t/L, \infty)\big) \leqslant \frac{1}{2}e^{-t^2/2L^2}, \tag{5.33}$$
$$\gamma_n(|f - M_f| > t) \leqslant e^{-t^2/2L^2}.$$

As we already noted in the setting of the sphere, concentration around the median formally implies similar concentration around the mean (see Remark 5.12). However, this approach leads to suboptimal constants. A more precise technique relies on the log-Sobolev inequality from Section 5.2.4.2, which specified to the Gaussian setting yields the following.

THEOREM 5.25 (see Theorem 5.39 and Proposition 5.42). *If $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz and $\mathbf{E}f$ is the mean of f (with respect to γ_n), then for any $t > 0$*

$$\max\{\gamma_n(f > \mathbf{E}f + t), \gamma_n(f < \mathbf{E}f - t)\} \leqslant e^{-t^2/2L^2}. \tag{5.34}$$

There is some numerical evidence that the assertion of Theorem 5.25 can be further strengthened. We pose

PROBLEM 5.26. *If $f : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz and $\mathbf{E}f$ denotes its average with respect to γ_n, is it true that $\gamma_n(|f - \mathbf{E}f| > t) \leqslant e^{-t^2/2}$? The case $n = 1$ implies the general case and is probably not that hard to settle. Similarly, is it true that $\sigma(|f - \mathbf{E}f| > t) \leqslant \exp(-nt^2/2)$ if $f : (S^{n-1}, g) \to \mathbb{R}$ is a 1-Lipschitz function (and $n > 2$; see Remark 5.19 for comments on peculiarities of the case $n = 2$)?*

An example of a function for which Theorem 5.24 is meaningful is the Euclidean norm, which is trivially 1-Lipschitz. This gives the following (see also Exercise 5.37).

COROLLARY 5.27. *Let G be a standard Gaussian vector in $\mathbb{R}^n$. Then, for any $t > 0$,*

$$\mathbf{P}\big(|G| \geqslant \sqrt{n} + t\big) \leqslant \frac{1}{2}e^{-t^2/2} \quad \text{and} \quad \mathbf{P}\Big(|G| \leqslant \sqrt{n - \frac{2}{3}} - t\Big) \leqslant \frac{1}{2}e^{-t^2/2}.$$

The distribution of $|G|^2$ is commonly known as $\chi^2(n)$, the chi-squared distribution with n degrees of freedom. Denoting by m_n the median of $|G|$, what is required to deduce Corollary 5.27 from Theorem 5.24 are the inequalities $\sqrt{n - \frac{2}{3}} \leqslant m_n \leqslant \sqrt{n}$. The lower bound is proved in Exercise 5.34 and the upper bound follows from Proposition 5.34): we have $m_n \leqslant \kappa_n \leqslant \sqrt{n}$.

EXERCISE 5.29 (Weak convergence in Poincaré's lemma). In the context of Poincaré's lemma (Theorem 5.22), show without any computation that the sequence (ν_N) converges weakly towards γ_n.

EXERCISE 5.30 (Gaussian isoperimetric inequality via Poincaré lemma). Derive the Gaussian isoperimetric inequality (5.29) from the Poincaré lemma (Theorem 5.22) and the spherical isoperimetric inequality (Theorem 5.13).

EXERCISE 5.31 (Ehrhard's inequality implies log-concavity). Show that Theorem 5.23 (Ehrhard's inequality) formally implies that the Gaussian measure γ_n satisfies the log-concavity inequality (4.28).

EXERCISE 5.32 (Gaussian measure of large balls). Show that

$$\lim_{r \to +\infty} \frac{\Phi^{-1}\big(\gamma_n(rB_2^n)\big)}{r} = 1.$$

EXERCISE 5.33 (Ehrhard-like (a-)symmetrization). Show that the following statement is equivalent to the validity of Ehrhard's inequality for convex bodies. *Let $K \subset \mathbb{R}^n$ be a convex body and let $E \subset \mathbb{R}^n$ be a k-dimensional subspace with $0 < k < n$. Identify E and $E^\perp$ with, respectively, $\mathbb{R}^k$ and $\mathbb{R}^{n-k}$ and define a set $L \subset \mathbb{R}^{k+1}$ by*

$$(x, s) \in L \iff s \leqslant \Phi^{-1}(\gamma_{n-k}(\{y \in E^\perp : (x, y) \in K\})),$$

where $x \in E, s \in \mathbb{R}$. Then L is convex.
In the case when $E = u^\perp$ is a hyperplane (i.e., $k = n - 1$) the transformation $K \mapsto L$ is called Ehrhard (a-)symmetrization in direction u.

Exercise 5.34 (Median of the chi-squared distribution, based on [**CR86**]). Let X be a random variable with distribution $\chi^2(n)$, and $V = \left(\frac{X}{n-2/3}\right)^{1/3}$. Show that the density h of V satisfies the inequality $h(1-t) \leqslant h(1+t)$ for $t \in [0,1]$, and conclude that the median of V is greater than 1, therefore the median of X is larger than $n-2/3$. Higher order two-sided bounds for the median can be found in [**BS**].

5.2.3. Concentration tricks and treats. This section contains a selection of largely elementary facts related to the concentration phenomenon. It supplies a set of tools allowing for flexible applications of concentration results. As a rule, the facts are well known to experts in the area and are included here for future reference. Proofs are relegated to exercises.

5.2.3.1. *Laplace transform.* We mostly restrict ourselves to settings where concentration exhibits a subgaussian behaviour as in (5.21) or (5.22). Such behaviour can be proved via estimating the bilateral Laplace transform, using the *exponential Markov inequality* $\mathbf{P}(X > t) \leqslant e^{-st}\,\mathbf{E}\exp(sX)$ for $s > 0$.

Lemma 5.28 (Laplace transform method). *Let X be a random variable such that $\mathbf{E}\exp(sX) \leqslant A\exp(\beta s^2)$ for every $s \in \mathbb{R}$. Then, for every $t > 0$,*

$$\max(\mathbf{P}(X > t), \mathbf{P}(-X > t)) \leqslant A\exp(-t^2/4\beta).$$

Exercise 5.35. Prove Lemma 5.28 about the Laplace transform method.

Exercise 5.36. Prove Hoeffding's lemma: if X is a mean zero random variable taking values in an interval $[a,b]$, then $\mathbf{E}\exp(sX) \leqslant \exp(\frac{1}{8}s^2(b-a)^2)$ for any $s \in \mathbb{R}$.

Exercise 5.37 (A large deviation bound for chi-squared variable, based on [**Vem04**]). Let X be a random variable with distribution $\chi^2(n)$, for example $X = |G|^2$ where G is a standard Gaussian vector in $\mathbb{R}^n$. Show that $\mathbf{E}\exp(sX) = (1-2s)^{-n/2}$ for any $s < 1/2$. Conclude that $\mathbf{P}(X \geqslant (1+\varepsilon)n) \leqslant \big((1+\varepsilon)\exp(-\varepsilon)\big)^{n/2}$ for any $\varepsilon > 0$ and that $\mathbf{P}(X \leqslant (1-\varepsilon)n) \leqslant \big((1-\varepsilon)\exp(\varepsilon)\big)^{n/2}$ for $\varepsilon \in (0,1]$. (We known from Cramér's large deviations theorem that this bounds are sharp.) Conclude that

$$\mathbf{P}(|X-n| \geqslant \varepsilon n) \leqslant 2\exp\left(-\frac{n\varepsilon^2}{4+8\varepsilon/3}\right). \tag{5.35}$$

5.2.3.2. *Central values.* Once we know that a function is concentrated around some value, we can *a posteriori* infer that it also concentrates around the mean or the median, or any other particular quantile. This can be formalized by the concept of a central value. If Y is a real random variable, we will say that M is a *central value* of Y if M is either the mean of Y, or any number between the 1st and the 3rd quartile of Y (i.e., if $\min\{\mathbf{P}(Y \geqslant M), \mathbf{P}(Y \leqslant M)\} \geqslant \frac{1}{4}$; this happens in particular if M is the median of Y). The numbers $\frac{1}{4}$ and $\frac{3}{4}$ play no special role and can be changed to other numbers from $(0,1)$ at the cost of deteriorating (or improving) the constants in the statements that follow (see, e.g., Remark 5.31).

Proposition 5.29 (see Exercises 5.38–5.40). *Let Y be a real random variable and let M be any central value for Y. Let $a \in \mathbb{R}$ and let constants $A \geqslant \frac{1}{2}, \lambda > 0$ be such that, for any $t > 0$,*

$$\max\{\mathbf{P}(Y > a+t), \mathbf{P}(Y < a-t)\} \leqslant A\exp(-\lambda t^2). \tag{5.36}$$

Then $|M - a| \leqslant \sqrt{\log(4A)}\,\lambda^{-1/2}$*. Consequently, for any* $t \geqslant \sqrt{\log(4A)}\,\lambda^{-1/2}$*,*

$$\max\{\mathbf{P}(Y > M + t), \mathbf{P}(Y < M - t)\} \leqslant 4A^2 \exp(-\lambda t^2/2). \tag{5.37}$$

Remark 5.30 (Improvements to Proposition 5.29). The expressions $\sqrt{\log(4A)}$ and $4A^2$ in the assertion of Proposition 5.29 can be replaced by $\sqrt{\log(\kappa A)}$ and κA^2, where $\kappa = 2$ when M is the median of Y and $\kappa = e$ when M is the expectation of Y; see Exercises 5.38, 5.39 and 5.40.

Remark 5.31 (On the necessity of restrictions on t in Proposition 5.29). We point out that the bound on the first (resp., the second) probability appearing in (5.37) is valid under the formally weaker restriction $t > (M - a)^+$ (resp., $t > (M - a)^-$). The restriction $t \geqslant \sqrt{\log(4A)}\,\lambda^{-1/2}$, while annoying, cannot be completely avoided if we want to keep full generality because the hypothesis (5.36) does not necessarily supply any information about the probabilities appearing in the assertion if t is small. However, this is only a minor inconvenience since for such t the upper bound in (5.37) is never small and often holds for trivial reasons. In particular, (5.37) holds for all $t > 0$ if M is the mean or any quantile between the 27th and 73rd percentile, or if $A \geqslant 3^{2/3}/4 \approx 0.52$, and always if we replace the factor $4A^2$ by $3\sqrt{2}A^2$. If M is the median, we can go even further: no restrictions on t are needed even if we replace $4A^2$ by $2A^2$ on the right hand side of (5.37); if M is the mean, similar improvement (i.e., eA^2 on the right hand side) is possible when $A \geqslant e^{-1/3} \approx 0.717$ (these last observations were used in Remark 5.12).

Corollary 5.32 (Lévy's lemma for central values). *Let* $f : (S^{n-1}, g) \to \mathbb{R}$ *be an L-Lipschitz function and let M be any central value for f. Then* $|M - M_f| \leqslant \sqrt{2\log 2}\, n^{-1/2}$ *and, for any* $\varepsilon > 0$*,*

$$\mathbf{P}(f \geqslant M + \varepsilon) \leqslant \exp\Big(-\frac{n\varepsilon^2}{4L^2}\Big). \tag{5.38}$$

We sketch proofs and give more precise bounds and/or variations on the above results in Exercises 5.38–5.48. Note that while (5.38) follows from Proposition 5.29 and Corollary 5.17 for $n > 2$ and for ε not-too-small, a separate argument is needed to cover the remaining cases (cf. Remark 5.31). We also point out that while Proposition 5.29 is meant to give reasonably good estimates valid in the most general setting when concentration is present, better bounds are available in specific instances. For example, Corollary 5.32 can be improved when M is the mean (see Table 5.2 and Exercise 5.44), and similarly in the Gaussian case.

The heuristics behind Corollary 5.32 is as follows: *if we know that all sets of measure at least* $\frac{1}{2}$ *have large enlargements, then approximately the same is true for all sets of measure at least* $\frac{1}{4}$. Actually, almost the same is true for much smaller sets; here is a sample result.

Proposition 5.33 (see Exercise 5.49). *Let* (X, d, μ) *be a metric probability space and let* $\varepsilon > 0$*. Suppose that any set* $A \subset X$ *with* $\mu(A) \geqslant \frac{1}{2}$ *verifies* $\mu(A_\varepsilon) \geqslant 1 - Ce^{-\lambda\varepsilon^2}$*. Then* $\mu(B_{2\varepsilon}) \geqslant 1 - Ce^{-\lambda\varepsilon^2}$ *for any set* $B \subset X$ *with* $\mu(B) \geqslant Ce^{-\lambda\varepsilon^2}$*.*

A common feature of concentration inequalities presented up to now is that in order to translate them to concrete bounds for concrete functions, we need to calculate—or at least reasonably estimate—the medians or expected values, or similar parameters of the functions under consideration. A selection of tools, some of them quite sharp, to handle expected values will be described in Section 6.1. The preceding three results tell us that it doesn't really matter which central value we employ, as long as we are willing to pay a small penalty in the form of an additional multiplicative constant in the exponent and in front of the exponential. The following observation shows that, in the Gaussian context, sometimes no penalty is needed at all.

PROPOSITION 5.34 (see Exercise 5.50). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Denote by M_f (resp., $\mathbf{E}f$) the median (resp., the expectation) of f with respect to the standard Gaussian measure γ_n. Then $M_f \leqslant \mathbf{E}f$.*

EXERCISE 5.38. Show that a random variable Y_0 such that $P(Y_0 > t) \leqslant A\exp(-t^2)$ for $t > 0$ must verify $\mathbf{E}\,Y_0 \leqslant \mathbf{E}\,Y_0^+ \leqslant \min\{A\sqrt{\pi}/2, \sqrt{1+\log^+ A}\}$. Deduce the first assertion of Proposition 5.29 and the corresponding improvement from Remark 5.30 if M is the mean of Y.

EXERCISE 5.39. Show that if Y_0 is a random variable such that $P(Y_0 > t) \leqslant A\exp(-t^2)$ for $t > 0$ and if $M_{3/4}$ is its 3rd quartile, then $M_{3/4} \leqslant \sqrt{\log^+(4A)}$. Deduce the first assertion of Proposition 5.29 if M is between the 1st or the 3rd quartile of Y, and the strengthening from Remark 5.30: $|M-a| \leqslant \sqrt{\log^+(2A)}\,\lambda^{-1/2}$ *if M is the median of Y.*

EXERCISE 5.40. Prove the inequality $e^{-s^2} \leqslant e^{\delta^2} e^{-(s+\delta)^2/2}$ for $s, \delta \in \mathbb{R}$. Use it and the last two exercises to show the second assertion of Proposition 5.29, and its strengthenings stated in Remark 5.30 when M is the median or the mean of Y.

EXERCISE 5.41. Verify the assertions in the last two sentences of Remark 5.31.

EXERCISE 5.42. Given $\alpha \in (0,1)$, prove a version of (5.37) with the right-hand side of the form $B\exp(-\alpha\lambda t^2)$, where B depends only on A and α (and on κ from Remark 5.30, if applicable).

EXERCISE 5.43 (Lévy's lemma for central values). Let $n > 2$. Use Exercise 5.26 to derive Corollary 5.32 for any quantile between the 1st and the 3rd quartile.

EXERCISE 5.44 (The median and the mean on the sphere). Let f be a 1-Lipschitz function on (S^{n-1}, g) with $n > 2$. Show that the median and the mean of f differ at most by $\sqrt{\pi/8n}$ and describe the extremal function.

EXERCISE 5.45 (Variance of a Lipschitz function on the sphere). Let f be a 1-Lipschitz function on (S^{n-1}, g) with $n > 1$. Show that $\mathbf{Var}(f) \leqslant \frac{2}{n}$ and give an example with $\mathbf{Var}(f) \geqslant \frac{1}{n}$. What function gives the maximal variance?

EXERCISE 5.46 (Concentration around L_2 average). Let f be a 1-Lipschitz and positive function on (S^{n-1}, g) with $n > 1$. Set $q = (\mathbf{E}f^2)^{1/2}$. Show that for any $t > 0$, $\mathbf{P}(f \geqslant q + t) \leqslant \exp(-nt^2/2)$ and $\mathbf{P}(f \leqslant q - t) \leqslant e\exp(-nt^2/2)$.

EXERCISE 5.47 (The case of S^1). Using directly the solution to the isoperimetric problem on S^1, show that Corollary 5.32 holds also for $n = 2$.

EXERCISE 5.48. Let (X, d, μ) be a metric probability space and let $\alpha : [0,\infty) \to [0,\infty)$ be such that $\mu(f \geqslant \mathbf{E}f + t) \leqslant \alpha(t)$ for any bounded 1-Lipschitz function $f : X \to \mathbb{R}$ and for all $t > 0$. Then, for any such function f and for any $t > 0$, $\mu(f \geqslant M_f + t) \leqslant \alpha(t/2)$. Equivalently, $\mu(A_\varepsilon) \geqslant 1 - \alpha(\varepsilon/2)$ for any $A \subset X$ with $\mu(A) \geqslant 1/2$ and any $\varepsilon > 0$. The preceding argument can be iterated, see (1.18) in [**Led01**].

Exercise 5.49. Prove Proposition 5.33 about enlargements of fairly small sets.

Exercise 5.50 (Median vs. mean for convex functions of Gaussian variables). Prove Proposition 5.34 by showing first that the function $g : t \mapsto \Phi^{-1}(\gamma_n(\{f \leqslant t\}))$ is concave.

Exercise 5.51. Show that the following statement is a consequence of Proposition 5.34. If $(X_1, \ldots, X_N)$ are jointly Gaussian random variables and $f : \mathbb{R}^N \to \mathbb{R}$ is a convex function, then the median of the random variable $f(X_1, \ldots, X_N)$ does not exceed its expectation.

5.2.3.3. *Local versions.* It sometimes happens that a function defined on the sphere S^{n-1} has a poor global Lipschitz behaviour, while its restriction to a subset of large measure is much more regular. To take advantage of such situation, we formulate a "local" version of Lévy's lemma.

Corollary 5.35 (Lévy's lemma, local version). *Let $\Omega \subset S^{n-1}$ be a subset of measure larger than 3/4. Let $f : (S^{n-1}, g) \to \mathbb{R}$ be a function such that the restriction of f to Ω is L-Lipschitz. Then, for every $\varepsilon > 0$,*

$$\mathbf{P}(\{|f(x) - M_f| > \varepsilon\}) \leqslant \mathbf{P}(S^{n-1} \backslash \Omega) + 2\exp(-n\varepsilon^2/4L^2),$$

where M_f is the median of f.

One scenario under which the hypotheses of Corollary 5.35 may be satisfied is when we have an upper bound on some Sobolev norm of f (a "global" parameter, which suggests that "restricted version of Lévy's lemma" could have been better terminology). However, our applications of the Corollary will be rather straightforward and will not require any advanced notions.

Exercise 5.52. Prove Corollary 5.35, the local version of Lévy's lemma.

5.2.3.4. *Pushforward.* The following elementary result is very useful for establishing concentration phenomenon for many classical spaces. In a nutshell, it says that concentration results can be "pushed forward" by surjective contractions.

Proposition 5.36 (Contraction principle). *Let (X, μ) and (Y, ν) be metric probability spaces. Assume that there exists a surjective contraction $\phi : X \to Y$ which pushes forward μ to ν (i.e., $\nu(B) = \mu(\phi^{-1}(B))$) and let $a \in (0, 1)$ and $\varepsilon > 0$. Then*

$$\inf_{B \subset Y,\, \nu(B) \geqslant a} \nu(B_\varepsilon) \geqslant \inf_{A \subset X,\, \mu(A) \geqslant a} \mu(A_\varepsilon). \tag{5.39}$$

Similarly, for any $t > 0$,

$$\sup_{g : Y \to \mathbb{R},\, g \text{ 1-Lipschitz}} \nu(g - \mathbf{E}g > t) \leqslant \sup_{f : X \to \mathbb{R},\, f \text{ 1-Lipschitz}} \mu(f - \mathbf{E}f > t). \tag{5.40}$$

Moreover, (5.40) *holds if expectation is replaced by median on both sides.*

Exercise 5.53. Prove Proposition 5.36, the contraction principle. State a more general version with $\phi : X \to Y$ assumed to be L-Lipschitz rather than a contraction.

Exercise 5.54 (Concentration on the solid cube via Gaussian pusforward). Let Y be the solid cube $[0, 1]^n$ endowed with the Lebesgue measure and the Euclidean metric inherited from $\mathbb{R}^n$. Use Proposition 5.36 to show that Y verifies (5.21) with $(C, \lambda) = \left(\frac{1}{2}, \pi\right)$ and (5.22) with $(C, \lambda) = (1, \pi)$.

5.2.3.5. *Direct products.* It is easy to see that the concentration phenomenon passes to direct products of metric probability spaces. Indeed, let X and Y be two such spaces that exhibit the concentration phenomenon and let $X \times Y$ be endowed with the product measure and some reasonable product metric, such as the ℓ_p *product metric* defined for (x_1, y_1) and (x_2, y_2) in $X \times Y$ as

$$d((x_1, y_1), (x_2, y_2)) = (d_X(x_1, x_2)^p + d_Y(x_1, x_2)^p)^{1/p}, \tag{5.41}$$

the limit case $p = \infty$ being interpreted as a maximum. If f is a 1-Lipschitz function on $X \times Y$, then $\phi(x) = M_{f(x,\cdot)}$ is 1-Lipschitz on X and hence concentrated around its median M_ϕ. Since, for each $x \in X$, $f(x, \cdot)$ is concentrated around $\phi(x)$, it follows that f is concentrated around M_ϕ. (See Exercise 5.55 for precise statements.) The above argument can be clearly iterated. Here is another elementary result involving product measures.

Proposition 5.37 (Concentration on product spaces, see Exercise 5.55). *Let (X_i, d_i, μ_i), $1 \leqslant i \leqslant n$, be bounded metric probability spaces and denote $D_i = \operatorname{diam} X_i$. Let $X = X_1 \times \ldots \times X_n$ be endowed with the product measure μ and the ℓ_1 product metric d. Then, for every 1-Lipschitz function $f : X \to \mathbb{R}$ and for any $t \geqslant 0$,*

$$\mu(f \geqslant \mathbf{E}f + t) \leqslant e^{-2t^2/D^2}, \tag{5.42}$$

where $D = \left(\sum_{i=1}^n D_i^2\right)^{1/2}$.

Both approaches to products of metric probability spaces that are sketched above share an unsatisfactory feature: the constants deteriorate as the number of factors increases. In complete generality, this feature is unavoidable (see Section 5.2.5). However, in some natural settings (e.g., the Gaussian space) dimension-free results are possible.

Exercise 5.55 (Concentration on product spaces, a naive approach). For the purpose of this exercise the median of a random variable F is defined as $M_F = \frac{1}{2}(\sup\{t : \mathbf{P}(F \geqslant t) \geqslant 1/2\} + \inf\{t : \mathbf{P}(F \leqslant t) \geqslant 1/2\})$, but most other definitions would work if applied consistently and with sufficient care. Let (X, d_1, μ) and (Y, d_2, ν) be metric probability spaces. Consider the space $(X \times Y, d, \pi)$, where $\pi = \mu \otimes \nu$ and d is any metric verifying

$$d((x_1, y), (x_2, y)) = d_1(x_1, x_2) \quad \text{and} \quad d((x, y_1), (x, y_2)) = d_2(y_1, y_2)$$

for all $x, x_1, x_2 \in X$ and $y, y_1, y_2 \in Y$ and let $f : X \times Y \to \mathbb{R}$ be a 1-Lipschitz function with respect to d.
(i) Show that the function $\phi(x) = M_{f(x,\cdot)}$ is 1-Lipschitz on X.
(ii) If X and Y exhibit the concentration phenomenon in the sense of (5.21) for some C and λ, then $\pi(f > M_\phi + t) \leqslant 2Ce^{-\lambda t^2/4}$ for all $t > 0$, and similarly for $\pi(f < M_\phi - t)$.
(iii) Show that M_ϕ is a central value in the sense of Section 5.2.3.
(iv) Same as (ii) with (5.21) replaced by (5.22) and M_ϕ by $\mathbf{E}f$.

Exercise 5.56 (Concentration on product spaces, Laplace transform method). The Laplace functional of a probability metric space (X, d, μ) is defined for $\lambda \in \mathbb{R}$ as $E_{(X,d,\mu)}(\lambda) = \sup \int e^{\lambda f} \, d\mu$, where the supremum is taken over all 1-Lipschitz functions $f : X \to \mathbb{R}$ with mean 0.
(i) Show that if X has diameter D, then $E_{(X,d,\mu)}(\lambda) \leqslant \exp(\lambda^2 D^2/8)$ (use Exercise

5.36).

(ii) Show that if (X_1, d_1, μ_1) and (X_2, d_2, μ_2) are two metric probability spaces, if d denotes the ℓ_1 product metric on $X_1 \times X_2$ as defined in (5.41), then

$$E_{(X_1\times X_2, d, \mu_1\otimes\mu_2)}(\lambda) \leqslant E_{(X_1,d_1,\mu_1)}(\lambda) E_{(X_2,d_2,\mu_2)}(\lambda).$$

(iii) Show that in the context of Proposition 5.37, we have

$$E_{(X,d,\mu)}(\lambda) \leqslant \exp(\lambda^2 D^2/8).$$

(iv) Prove Proposition 5.37 using Lemma 5.28.

Exercise 5.57 (Hoeffding's inequality). Show that Proposition 5.37 implies Hoeffding's inequality: *if $X_1, \ldots, X_n$ are independent random variables such that X_i takes values in an interval of length l_i, then for any $t > 0$,*

$$\mathbf{P}(S \geqslant \mathbf{E}S + t) \leqslant e^{-2t^2/L^2}, \tag{5.43}$$

where $S = X_1 + \cdots + X_n$ and $L^2 = l_1^2 + \cdots + l_n^2$.

5.2.4. Geometric and analytic methods. Classical examples. In Sections 5.2.1 and 5.2.2 we sketched isoperimetric/concentration results on the Euclidean sphere and for the Gaussian measure. While these are admittedly very special situations, the fact of the matter is that, in high-dimensional settings, some form of concentration phenomenon is the rule rather than the exception.

5.2.4.1. *Gromov's comparison theorem.* The first result asserts that isoperimetric and concentration inequalities hold under geometric assumptions which significantly generalize the spherical case. The invariant that can be related to sphere-like behavior is the *Ricci curvature*, which describes the rate of growth of volume under geodesic flow on the manifold with the similar rate in the Euclidean space. For example (see Figure 5.3), the circumference of a circle of geodesic radius θ $(< \pi)$ on the sphere S^2 is $2\pi \sin\theta$, and hence the length of the arc of the circle corresponding to an angle α (measured on the plane tangent at the center of the circle) is $\alpha \sin\theta \approx \alpha\big(\theta - \frac{\theta^3}{6}\big) = \alpha\theta\big(1 - \frac{\theta^2}{6}\big)$ compared to $\alpha\theta$ for the Euclidean plane. (Here and in the next paragraph $\approx$ means equality up to higher order terms.)

Repeating this calculation *mutatis mutandis* for an m-dimensional sphere (in $\mathbb{R}^{m+1}$) of radius R and a solid m-dimensional angle α we get $\alpha\big(R\sin\frac{\theta}{R}\big)^{m-1} \approx \alpha\big(\theta - \frac{\theta^3}{6R^2}\big)^{m-1} \approx \alpha\theta^{m-1}\big(1 - \frac{m-1}{R^2}\frac{\theta^2}{6}\big)$ compared to $\alpha\theta^{m-1}$ in the Euclidean setting (i.e., in $\mathbb{R}^m$). This is subsumed by saying that the Ricci curvature of RS^m, the m-dimensional sphere of radius R, *at every point and in each direction* is $\frac{m-1}{R^2}$. The notion is generalized to an arbitrary point p on a Riemannian manifold X of dimension greater than or equal to 2 and to an arbitrary unit vector u in the tangent space at p by considering infinitesimal (solid) angles in the direction of u and finding the coefficient of $\frac{\theta^2}{6}$ in the corresponding expression for the volume on the geodesic sphere or radius θ centered at p; this coefficient is denoted by $\mathrm{Ric}_p(u)$. The minimum of $\mathrm{Ric}_p(u)$ over $p \in X$ and over directions u is denoted by $c(X)$.

Such straightforward calculation may be difficult to perform for more complicated manifolds. On a less elementary level, the Ricci curvature can be computed using the following formula expressed in the language of Riemannian geometry: whenever $(u_1, \ldots, u_m)$ is an orthonormal basis in the tangent space at p (thought

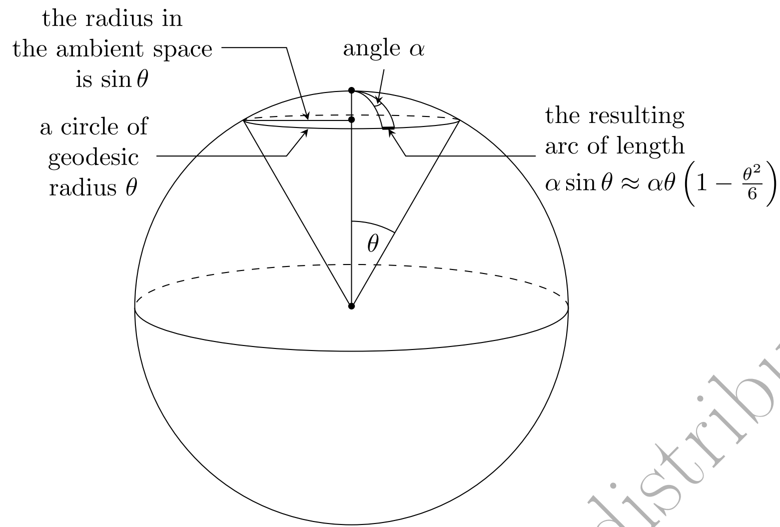

FIGURE 5.3. Volume growth on the sphere S^2 as a function of geodesic distance.

of as a real inner product space), we have

$$\mathrm{Ric}_p(u_1) = \sum_{i=2}^{m} \sec(u_1, u_i), \tag{5.44}$$

where sec denotes the sectional curvature. This leads to an alternative explanation of the value of the Ricci curvature for the sphere, for other manifolds of constant sectional curvature such as the Euclidean space or the hyperbolic space, or for their quotients by discrete groups of symmetries (e.g., for tori or for the real projective space). In the case of Lie groups, sectional curvature can be expressed via Lie brackets. For examples of computations, see Exercises 5.58 and 5.59.

We are now ready to state the main result of this section. By RS^m we denote the sphere of radius R in $\mathbb{R}^{m+1}$.

THEOREM 5.38 (Gromov's comparison theorem, not proved here). *Let $m \geqslant 2$ and let X be an m-dimensional connected Riemannian manifold such that $c(X) \geqslant \frac{m-1}{R^2} = c(RS^m)$. Let $A \subset X$ and let $C \subset RS^m$ be a cap such that $\mu_X(A) = \mu_{RS^m}(C)$, where μ_X and μ_{RS^m} are normalized Riemannian volumes on, respectively, X and RS^m. Then, for every $\varepsilon > 0$, $\mu_X(A_\varepsilon) \geqslant \mu_{RS^m}(C_\varepsilon)$.*

It follows then (same proof as Corollary 5.17) that any 1-Lipschitz function $f : X \to \mathbb{R}$ with median M_f satisfies, for any $t > 0$,

$$\mu_X(\{f > M_f + t\}) \leqslant \frac{1}{2}\exp(-(m+1)t^2/2R^2).$$

As it turns out, the hypotheses of Theorem 5.38 are verified for many (but not all) manifolds that naturally appear in mathematics and that play a role in physics, notably for most classical Lie groups and their homogeneous spaces, see Table 5.3.

EXERCISE 5.58 (Ricci curvature of Grassmannians). For $\mathsf{Gr}(k, \mathbb{R}^n)$ or $\mathsf{Gr}(k, \mathbb{C}^n)$, the tangent space at any point can be identified with $\mathsf{M}_{k,n-k}$. If $X, Y \in \mathsf{M}_{k,n-k}$ are

TABLE 5.3. Optimal bounds on Ricci curvature for a selection of classical manifolds. We restrict our attention to manifolds for which that curvature is nonnegative, which in particular excludes the hyperbolic space and its quotients. All the bounds concerning specific objects can be derived via formula (5.44) involving the (more standard) sectional curvatures. This is straightforward for spaces, for which the sectional curvatures are constant (R^n, S^{n-1}, and $\mathsf{P}(\mathbb{R}^n)$); the remaining cases are covered by Exercises 5.58 and 5.59. Note that the values for the projective spaces $\mathsf{P}(V)$ and the corresponding $\mathsf{Gr}(1, V)$ do not coincide due to different normalization of the metric (an additional $\sqrt{2}$ factor in (B.10) when compared to (B.5)).

X	metric	$c(X)$	comments
$\mathbb{R}^n$	Euclidean	0	
S^{n-1}	geodesic	$n-2$	$n \geqslant 2$
$\mathsf{SO}(n)$	standard (B.8)	$\frac{n-2}{4}$	$n \geqslant 2$
$\mathsf{SU}(n)$	standard (B.8)	$\frac{n}{2}$	
$\mathsf{U}(n)$	standard (B.8)	0	
$\mathsf{Gr}(k, \mathbb{R}^n)$	quotient from $\mathsf{O}(n)$ (B.10)	$\frac{n-2}{2}$	$1 \leqslant k \leqslant n-1$
$\mathsf{Gr}(k, \mathbb{C}^n)$	quotient from $\mathsf{U}(n)$ (B.10)	n	$1 \leqslant k \leqslant n-1$
$\mathsf{P}(\mathbb{R}^n)$	Fubini–Study (B.5)	$n-2$	$n \geqslant 2$
$\mathsf{P}(\mathbb{C}^n)$	Fubini–Study (B.5)	$2n$	$n \geqslant 2$
$X_1 \times X_2$	ℓ_2 product metric (5.41)	$\min\{c(X_1), c(X_2)\}$	

orthogonal, one can show (see Section 8.2.1 in [**Pet06**]) that

$$\sec(X, Y) = \frac{1}{4}\left(\|XY^\dagger - YX^\dagger\|_{\mathrm{HS}}^2 + \|X^\dagger Y - Y^\dagger X\|_{\mathrm{HS}}^2\right). \tag{5.45}$$

Use this formula and (5.44) to compute the corresponding values from Table 5.3. In some references we find the coefficient $\frac{1}{2}$ instead of $\frac{1}{4}$ because of a different normalization of the metric.

EXERCISE 5.59 (Ricci curvature of classical groups). For $\mathsf{G} = \mathsf{SO}(n)$, $\mathsf{SU}(n)$ or $\mathsf{U}(n)$, the tangent space at I (or at any point) can be identified with the corresponding Lie algebra $\mathfrak{g}$ ($= \mathfrak{so}_n$, $\mathfrak{su}_n$ or $\mathfrak{u}_n$). If $X, Y \in \mathfrak{g}$ are orthonormal, one can show (see Exercise 2.19 in [**Pet06**]) that $\sec(X, Y) = \frac{1}{4}\|XY - YX\|_{\mathrm{HS}}^2$. Use this formula and (5.44) to compute the corresponding values from Table 5.3.

5.2.4.2. *Log-Sobolev inequalities* (LSI). The next technique that we present is of analytic nature. It is based on a class of inequalities which at the first sight seem irrelevant to the subject at hand. Let (X, μ) be a measure space and let f be a non-negative function on X. The (continuous Shannon) *entropy* is defined by

$$\mathrm{Ent}_\mu(f) := \int f \log f \, \mathrm{d}\mu \tag{5.46}$$

if $\int f \, \mathrm{d}\mu = 1$, where we used the convention $0 \log 0 = 0$, and then extended to non-negative integrable functions by 1-homogeneity. An explicit formula that implements the extension is

$$\mathrm{Ent}_\mu(f) := \int f \log f \, \mathrm{d}\mu - \int f \, \mathrm{d}\mu \, \log\Big(\int f \, \mathrm{d}\mu\Big). \tag{5.47}$$

By Jensen's inequality, $\mathrm{Ent}_\mu(f) \geqslant 0$, with $+\infty$ being a possibility.

We now assume that X is a Riemannian manifold and that μ is a Borel measure on X. We say that (X, μ) verifies a logarithmic Sobolev inequality with parameter α if for every (sufficiently smooth) function $f : X \to \mathbb{R}$ we have

$$\mathrm{Ent}_\mu(f^2) \leqslant 2\alpha \int |\nabla f|^2 \, \mathrm{d}\mu. \tag{5.48}$$

The smallest constant α that works in (5.48) is called the *log-Sobolev constant* of (X, μ) and denoted by $\mathrm{LS}(X, \mu)$.

The relevance of this circle of ideas to the concentration phenomenon is explained by the following result.

Theorem 5.39 (Herbst's argument). *Let X be a Riemannian manifold and let μ be a Borel probability measure on X such that $\mathrm{LS}(X, \mu) \leqslant \alpha$. Then every 1-Lipschitz function $F : X \to \mathbb{R}$ is integrable and satisfies, for every $t > 0$,*

$$\mu\Big(F > \int F \mathrm{d}\mu \; + \; t\Big) \leqslant e^{-t^2/2\alpha}. \tag{5.49}$$

Remark 5.40. The above Theorem can be extended to the setting of general metric spaces, with essentially the same proof, once $|\nabla f|$ is properly defined. For example, we may use $|\nabla f|(x) = \limsup_{y \to x} \frac{|f(y) - f(x)|}{\mathrm{dist}(y,x)}$ if X has no isolated points; discrete spaces may also be handled with some care. However, for clarity of the exposition, we will assume for the rest of this subsection that the underlying spaces are (connected) Riemannian manifolds.

Proof of Theorem 5.39. First, we may assume that F is smooth and that $\int F \, \mathrm{d}\mu = 0$; this may be achieved by replacing F by an appropriate approximation and subtracting a constant. The strategy is to show that the (bilateral) Laplace transform of F verifies

$$\int e^{\lambda F} \, \mathrm{d}\mu \leqslant e^{\alpha \lambda^2/2} \quad \text{for all} \quad \lambda \in \mathbb{R}, \tag{5.50}$$

which by Lemma 5.28 implies that $\mu(F > t) \leqslant e^{-t^2/2\alpha}$, as needed. To establish (5.50), we introduce an auxiliary function $f = f_\lambda > 0$ defined via $f^2 = e^{\lambda F - \alpha\lambda^2/2}$. In other words, $f = e^{\lambda F/2 - \alpha\lambda^2/4}$ and it is readily checked that $\nabla f = \frac{\lambda}{2} f \nabla F$. Since $|\nabla F| \leqslant 1$ (because F is 1-Lipschitz), it follows that $|\nabla f|^2 \leqslant \frac{\lambda^2}{4} f^2$. Consequently, by (5.48) (cf. (5.47)),

$$\mathrm{Ent}_\mu(f^2) = \int f^2 \Big(\lambda F - \frac{\alpha\lambda^2}{2}\Big) \, \mathrm{d}\mu - \int f^2 \, \mathrm{d}\mu \, \log\Big(\int f^2 \, \mathrm{d}\mu\Big) \leqslant \frac{\alpha\lambda^2}{2} \int f^2 \, \mathrm{d}\mu. \tag{5.51}$$

We now set $\phi(\lambda) = \int f^2 \, \mathrm{d}\mu$ and note that differentiating under the integral sign gives

$$\phi'(\lambda) = \int f^2 (F - \alpha\lambda) \, \mathrm{d}\mu.$$

This allows to rewrite (5.51) as

$$\lambda\phi'(\lambda) - \phi(\lambda)\,\log\big(\phi(\lambda)\big) \leqslant 0,$$

which, for $\lambda \neq 0$, is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big(\frac{\log\big(\phi(\lambda)\big)}{\lambda}\Big) \leqslant 0. \tag{5.52}$$

On the other hand, given that $\phi(0) = 1$, l'Hôpital's rule yields

$$\lim_{\lambda\to 0}\frac{\log\big(\phi(\lambda)\big)}{\lambda} = \lim_{\lambda\to 0}\frac{\phi'(\lambda)}{\phi(\lambda)} = \frac{\phi'(0)}{\phi(0)} = \frac{\int F\,\mathrm{d}\mu}{1} = 0. \tag{5.53}$$

Combining (5.52) and (5.53) we conclude that $\log\big(\phi(\lambda)\big)/\lambda \leqslant 0$ for $\lambda > 0$ and $\log\big(\phi(\lambda)\big)/\lambda \geqslant 0$ for $\lambda < 0$, which just means that $\phi(\lambda) \leqslant 1$ for all $\lambda \in \mathbb{R}$. In other words, $\int e^{\lambda F - \alpha\lambda^2/2}\,\mathrm{d}\mu \leqslant 1$ for $\lambda \in \mathbb{R}$, which is just a restatement of (5.50) and concludes the argument. □

Apart from the median being replaced by the expected value (which is largely a matter of convenience or elegance, see Proposition 5.29 in Section 5.2.3), the assertion of Theorem 5.39 closely resembles (5.26) and (5.33), which quantified the concentration phenomenon for Lipschitz functions in the spherical and Gaussian settings. However, its usefulness depends on availability of spaces (X, μ) verifying logarithmic Sobolev inequalities. The next few results ensure that the supply is indeed quite ample. For easy reference, the spaces and estimates on their log-Sobolev constants are cataloged in Table 5.4.

PROPOSITION 5.41 (not proved here). *Let X be an m-dimensional Riemannian manifold such that $c(X) > 0$ and let μ be the normalized Riemannian volume. Then* $\mathrm{LS}(X,\mu) \leqslant \frac{m-1}{mc(X)}$.

PROPOSITION 5.42 (not proved here). *Let μ be a measure on $\mathbb{R}^n$ whose density with respect to the Lebesgue measure is of the form e^{-U}, where U verifies* $\mathrm{Hess}(U) \geqslant \beta\,\mathrm{I}$ *for some $\beta > 0$. Then* $\mathrm{LS}(\mathbb{R}^n,\mu) \leqslant \beta^{-1}$. *In particular,* $\mathrm{LS}(\mathbb{R}^n,\gamma_n) \leqslant 1$ *and* $\mathrm{LS}(\mathbb{C}^n,\gamma_n^{\mathbb{C}}) \leqslant \frac{1}{2}$.

PROPOSITION 5.43 (not proved here, but see Exercise 5.61). *We have*

$$\mathrm{LS}(S^1,\sigma) = 1 \quad \text{and} \quad \mathrm{LS}([0,1],\mathrm{vol}_1) = \pi^{-2}.$$

PROPOSITION 5.44 (Tensorization property of LSI, not proved here). *Given (X_i,μ_i), $i = 1,\dots,k$, let $X = X_1 \times \cdots \times X_k$ be endowed with the ℓ_2 product metric as defined in* (5.41) *and the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_k$. Then* $\mathrm{LS}(X,\mu) = \max_{1\leqslant i\leqslant k}\mathrm{LS}(X_i,\mu_i)$.

REMARK 5.45 (Poincaré's inequality). Another related famous functional inequality is the Poincaré inequality, which reads as follows: for every smooth function $f : X \to \mathbb{R}$

$$\mathbf{Var}_\mu\, f \leqslant \alpha \int |\nabla f|^2\,\mathrm{d}\mu, \tag{5.54}$$

where $\mathbf{Var}_\mu\, f$ denotes the quantity $\int f^2\,\mathrm{d}\mu - \big(\int f\,\mathrm{d}\mu\big)^2$. The smallest α is called the Poincaré constant of (X,μ) and denoted $\mathrm{P}(X,\mu)$. Inequality (5.54) is implied by the LSI (5.48) (with the same constant α); it implies sub-exponential instead of subgaussian concentration. A list of Poincaré constants for common spaces can be found in Table 5.4. An example of a probability measure satisfying the Poincaré inequality but not the LSI is the (symmetric) exponential distribution on $\mathbb{R}$.

REMARK 5.46 (Contraction principle for LSI and Poincaré's inequality). If $\phi : (X, \mu) \to (Y, \nu)$ is a surjective contraction which pushes forward μ onto ν, then $\mathrm{LS}(Y, \nu) \leqslant \mathrm{LS}(X, \mu)$ and $\mathrm{P}(Y, \nu) \leqslant \mathrm{P}(X, \mu)$. This can be proved as in Exercise 5.53 and is especially transparent if we define $|\nabla f|$ as in Remark 5.40.

TABLE 5.4. Bounds on log-Sobolev and Poincaré constants for a selection of classical manifolds. We use the same metrics as in Table 5.3. Except as indicated, the estimates on log-Sobolev constants follow from estimates on the Ricci curvature (see Proposition 5.41). Most of the time we use the bound $\mathrm{LS}(X, \mu) < c(X)^{-1}$; the more precise expressions involving the dimension of X lead to slightly better but often cumbersome formulas. The upper bounds on the Poincaré constants of Grassmann manifolds follow from Remark 5.46. For more comments and references about Poincaré constants, see Notes and Remarks.

X or (X, μ)	$\mathrm{LS}(X, \mu)$	$\mathrm{P}(X, \mu)$	Comments
$\left([a, b], \frac{vol_1}{b-a}\right)$	$\frac{(b-a)^2}{\pi^2}$	$\frac{(b-a)^2}{\pi^2}$	Prop. 5.43
S^{n-1}	$\frac{1}{n-1}$	$\frac{1}{n-1}$	Prop. 5.43 for S^1
$\mathsf{P}(\mathbb{R}^n)$	$\leqslant \frac{1}{n-1}$	$\frac{1}{2n}$	
$\mathsf{P}(\mathbb{C}^n)$	$< \frac{1}{2n}$	$\frac{1}{4n}$	
$(\mathbb{R}^n, \gamma_n)$	1	1	Exercise 5.60
$\mathsf{SO}(n)$	$< \frac{4}{n-2}$	$\frac{2}{n-1}$	
$\mathsf{SU}(n)$	$< \frac{2}{n}$	$\frac{n}{n^2-1}$	
$\mathsf{U}(n)$	$\leqslant \frac{6}{n}$	$\frac{1}{n}$	**[MM13]**
$\mathsf{Gr}(k, \mathbb{R}^n)$	$< \frac{2}{n-2}$	$\leqslant \frac{2}{n-1}$	$1 \leqslant k \leqslant n-1$
$\mathsf{Gr}(k, \mathbb{C}^n)$	$< \frac{1}{n}$	$\leqslant \frac{1}{n}$	$1 \leqslant k \leqslant n-1$
$(X \times Y, \mu_X \otimes \mu_Y)$	$\max\{\mathrm{LS}(X), \mathrm{LS}(Y)\}$	$\max\{\mathrm{P}(X), \mathrm{P}(Y)\}$	ℓ_2 product metric

EXERCISE 5.60 (Log-Sobolev constant for the Gaussian space). Show that $\mathrm{LS}(\mathbb{R}^n, \gamma_n) \geqslant 1$ (we have actually equality, see Proposition 5.42).

EXERCISE 5.61 (Log-Sobolev constants for segments and circles). (i) Use the contraction principle from Remark 5.46 to show that $\mathrm{LS}([0, 1], \mathrm{vol}_1) \leqslant \pi^{-2}\mathrm{LS}(S^1, \sigma)$ and $\mathrm{P}([0, 1], \mathrm{vol}_1) \leqslant \pi^{-2}\mathrm{P}(S^1, \sigma)$. (ii) Verify that $\mathrm{P}(S^1, \sigma) = 1$. (iii) Verify that $\mathrm{P}([0, 1], \mathrm{vol}_1) \geqslant \pi^{-2}$ (see Notes and Remarks for the reasons why there is actually an equality).

5.2.4.3. *Hypercontractivity, Gaussian polynomials.* We give a brief introduction to the concept of hypercontractivity and illustrate it to give an example of a concentration inequality for Gaussian polynomials.

We work on the probability space $(\mathbb{R}^n, \gamma_n)$. We define the Ornstein–Uhlenbeck semigroup of operators $(P_t)_{t \geqslant 0}$ as follows. For $f : \mathbb{R}^n \to \mathbb{R}$ a bounded measurable function, and $x \in \mathbb{R}^n$, let

$$(P_t f)(x) = \mathbf{E}\, f\Big(e^{-t}x + \sqrt{1-e^{-2t}}G\Big), \tag{5.55}$$

where G is a standard Gaussian vector in $\mathbb{R}^n$. These operators satisfy the semigroup property $P_sP_t = P_{s+t}$. Moreover it is easily checked (Exercise 5.62) that for every $p \geqslant 1$ and $t \geqslant 0$,

$$\|P_t f\|_{L_p(\gamma_n)} \leqslant \|f\|_{L_p(\gamma_n)},$$

and therefore P_t extends to a bounded (contractive) operator on $L_p(\gamma_n)$. Remarkably, a stronger statement is true: provided $p > 1$ and $t > 0$, P_t is a contraction from $L_p(\gamma_n)$ to $L_q(\gamma_n)$ for some $q = q(t) > p$. This phenomenon is called hypercontractivity.

Proposition 5.47 (not proved here, but see Exercise 5.63). *Let $1 \leqslant p \leqslant q < \infty$ and $t > 0$ such that $q \leqslant 1 + e^{2t}(p-1)$. Then*

$$\|P_t f\|_{L_q(\gamma_n)} \leqslant \|f\|_{L_p(\gamma_n)}.$$

The eigenvectors of P_t are the Hermite polynomials. In the one-dimensional case, denote by $(h_k)_{k\in\mathbb{N}}$ the sequence of polynomials obtained by orthonormalizing the sequence $(1, x, x^2, \dots)$ in the space $\mathscr{H}_1 := L_2(\mathbb{R}, \gamma_1)$. (In this context, we exceptionally mean $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.) Given a multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, let h_α be the multivariate polynomial

$$h_\alpha(x_1, \dots, x_n) = h_{\alpha_1}(x_1) \cdots h_{\alpha_n}(x_n). \tag{5.56}$$

The family $(h_\alpha)_{\alpha\in\mathbb{N}^n}$ is an orthonormal basis in $\mathscr{H}_n := L_2(\mathbb{R}^n, \gamma_n)$, and we have

$$P_t h_\alpha = e^{-t|\alpha|} h_\alpha, \tag{5.57}$$

where $|\alpha| = \sum_{i=1}^n \alpha_i$ is the weight of the multi-index α, or the total degree of the polynomial h_α. Note that formula (5.57) allows to define P_tQ for any polynomial Q even when t is negative.

Proposition 5.48. *Let Q be a polynomial in n variables of (total) degree at most k. Then, for every $q \geqslant 2$,*

$$\|Q\|_{L_q(\gamma_n)} \leqslant (q-1)^{k/2} \|Q\|_{L_2(\gamma_n)}.$$

Proof. For any $t \geqslant 0$, we have $P_tP_{-t}Q = Q$ (see the remark following (5.57)). Choosing $t > 0$ such that $q - 1 = e^{2t}$, we may apply Proposition 5.47 to conclude that $\|Q\|_{L_q(\gamma_n)} \leqslant \|P_{-t}Q\|_{L_2(\gamma_n)}$. We may write the decomposition of Q in the basis of Hermite polynomials

$$Q = \sum_{|\alpha|\leqslant k} c_\alpha h_\alpha$$

for some coefficients (c_α). It follows that $\|Q\|^2_{L_2(\gamma_n)} = \sum c_\alpha^2$, while

$$\|P_{-t}Q\|^2_{L_2(\gamma_n)} = \sum_{|\alpha|\leqslant k} e^{2t|\alpha|} c_\alpha^2 \leqslant e^{2tk} \|Q\|^2_{L_2(\gamma_n)},$$

whence the result follows. □

Corollary 5.49 (Concentration inequality for Gaussian polynomials). *Let $Z_1, \dots, Z_n$ be independent $N(0,1)$ variables and let $X = Q(Z_1, \dots, Z_n)$, where Q is a polynomial of (total) degree at most k. Then, for any $t \geqslant (2e)^{k/2}$,*

$$\mathbf{P}\left(|X - \mathbf{E}X| \geqslant t\sqrt{\mathbf{Var}\,X}\right) \leqslant \exp\left(-\frac{k}{2e}t^{2/k}\right).$$

Proof. There is no loss of generality in assuming that $Z_1, \dots, Z_n$ are defined as the coordinate functions on $(\mathbb{R}^n, \gamma_n)$, so that Proposition 5.48 applies. We may assume $\mathbf{E}X = 0$, $\mathbf{Var}\,X = 1$ and write by Markov's inequality, for any $q \geqslant 2$,

$$\mathbf{P}(|X| \geqslant t) \leqslant t^{-q}\,\mathbf{E}\,|X|^q \leqslant t^{-q}(q-1)^{kq/2} \leqslant (q^{k/2}/t)^q$$

where we used Proposition 5.48. The choice $q = t^{2/k}/e$ (which is larger than 2 provided $t \geqslant (2e)^{k/2}$) yields the result. □

Remark 5.50. The phenomenon of hypercontractivity is not specific to the Gaussian case and is essentially equivalent to a log-Sobolev inequality (see Theorem 5.2.3 in [**BGL14**]). Similar concentration results are true for polynomials in binary random variables (see Theorem 9.21 in [**O'D14**]) and for polynomials on the sphere (cf. [**Mon12**]). Here is a precise statement of the latter. *If Q be a polynomial with total degree at most k in $n_1 + \cdots + n_d$ variables and $X = (X_1, \dots, X_d)$ with X_i independent and uniformly distributed on S^{n_i-1}, then for every $q \geqslant 2$, $\|Q(X)\|_{L_q} \leqslant (q-1)^{k/2}\|Q(X)\|_{L_2}$.* (This is slightly more general than Corollary 12 in [**Mon12**] which assumes that $n_1 = \cdots = n_d$ and that the partial degrees in each variable are equal.) The argument is similar to the Gaussian case, using spherical harmonics instead of Hermite polynomials. Concentration estimates similar to Corollary 5.49 follow.

Exercise 5.62 (Ornstein–Uhlenbeck semigroup is contractive). Show that P_t is a contraction on $L_p(\gamma_n)$ for any $t \geqslant 0$ and $p \geqslant 1$.

Exercise 5.63 (Sharpness of the hypercontractive inequality). When $n = 1$, compute $P_t f_\lambda$ when $f_\lambda(x) = e^{\lambda x}$. Conclude that Proposition 5.47 is sharp in the following sense: when $q > 1 + e^{2t}(p-1)$, there is no constant C such that the inequality $\|P_t f\|_{L_q(\gamma_1)} \leqslant C\|f\|_{L_p(\gamma_1)}$ holds.

5.2.5. Some discrete settings. All the *specific* instances of concentration we identified thus far involved manifolds. However, the phenomenon also occurs in the discrete case. We will exemplify it (and the issues that may arise) on the fundamental example of the *Boolean cube* $\{0,1\}^n$, or $\{-1,1\}^n$, endowed with the normalized counting measure μ and the normalized Hamming distance $d_H(x,y) := \frac{1}{n}\,\mathrm{card}\{i : x_i \neq y_i\}$, which up to normalization coincides with the ℓ_1 metric in the ambient space $\mathbb{R}^n$. (This setting was already studied in Section 5.1.3; other product measures, or metrics induced by ℓ_p-norms for other p are also frequently considered, more about that later.)

A nearly optimal concentration result for the Boolean cube follows already from Proposition 5.37. However, we can do better: the exact solution to the isoperimetric problem on the cube is known. To describe it, we introduce a total order $<$ on $\{0,1\}^n$ (called the *simplicial order*) as follows: for $x = (x_i)$ and $y = (y_i)$ in $\{0,1\}^n$, declare that $x < y$ if either $x_1 + \cdots + x_n < y_1 + \cdots + y_n$ or $x_1 + \cdots + x_n = y_1 + \cdots + y_n$ and x precedes y in the lexicographic order. Then the initial segments for this order are isoperimetric sets. As opposed to the Gaussian and spherical case, the extremal sets are not unique in any reasonable sense (see Exercise 5.66)

THEOREM 5.51 (Harper's isoperimetric inequality, not proved here). *For any integer N with $1 \leqslant N < 2^n$, let $A \subset \{0,1\}^n$ be the set of N smallest elements with respect to the simplicial order. Then A has the smallest ε-enlargements (for all $\varepsilon > 0$) among all sets of the same cardinality. The set A verifies*

$$B(x, k/2^n) \subset A \subset B(x, (k+1)/2^n) \tag{5.58}$$

for some $k \in \{0, \ldots, n-1\}$.

If we define the boundary of A as $\partial A := \{y \in \{0,1\}^n : \operatorname{dist}(y, A) = 1/n\}$, the sets from Theorem 5.51 also have the "smallest boundary" among subsets of $\{0,1\}^n$ of the same measure. In this language, the condition (5.58) says that A consists of a ball and a part of its boundary. If $N = \sum_{j=1}^{k} \binom{n}{j}$ for some k, the situation becomes simple: the optimal sets are balls, and so are their enlargements.

For example, if $n = 2m+1$ is odd, an example of an optimal set of measure $\frac{1}{2}$ is

$$A = \{y \in \{0,1\}^n : Y \leqslant m\},$$

where $Y = \sum_{j=1}^{n} y_j$. The enlargements of A are then clearly of the form $A_{s/n} = \big\{Y \leqslant m+s\big\}$ and, consequently,

$$\mu(A_{s/n}) = \frac{\sum_{j=1}^{m+s} \binom{n}{j}}{2^n} = 1 - \frac{\sum_{j>m+s} \binom{n}{j}}{2^n} \geqslant 1 - e^{-2s^2/n}, \tag{5.59}$$

where the inequality follows from Hoeffding's inequality (5.43). A similar analysis can be performed when n is even (see Exercise 5.64 for details). To summarize, we have

COROLLARY 5.52. *If $A \subset \{0,1\}^n$ with $\mu(A) \geqslant \frac{1}{2}$, $s \in \mathbb{N}$ and $\varepsilon = s/n$, then $\mu(A_\varepsilon) \geqslant 1 - e^{-2n\varepsilon^2}$. Consequently, if $f : \{0,1\}^n \to \mathbb{R}$ is a 1-Lipschitz function and M is its median, then $\mu(f > M + \varepsilon) \leqslant e^{-2n\varepsilon^2}$.*

REMARK 5.53. Some authors assert that the bound $\mu(A_\varepsilon) \geqslant 1 - e^{-2n\varepsilon^2}$ (for A satisfying $\mu(A) \geqslant \frac{1}{2}$) holds for all $\varepsilon > 0$. However, this may be false, but only if $n = 1$ or 2 and only for certain values of $\varepsilon \in (0, 1/n)$, see Exercise 5.65.

The setting of Corollary 5.52 is a special case of that of Proposition 5.37. (The differences include the mean being replaced by the median, and the numerical constants being better in the former, which is not surprising since it is a more specialized result.) The Corollary is an elegant and sharp result, but it exhibits the following unsatisfactory feature: if we use the standard Euclidean metric to define the 1-Lipschitz property of f or the expansions A_t, the exponential term in the estimates becomes $e^{-2t^2/n}$. This should be compared to the dimension-free (and differently scaled) term $\frac{1}{2}e^{-t^2/2}$ in Theorem 5.24, the Gaussian isoperimetric inequality. However, there is a fix to this difficulty due to Talagrand: if the function f is convex, its restriction to $\{0,1\}^n$ exhibits dimension-free subgaussian concentration. We have

THEOREM 5.54 (Talagrand's convex concentration inequality for the Boolean cube, not proved here). *Let A be a non-empty subset of $\{0,1\}^n \subset \mathbb{R}^n$ and set*

$\phi_A(x) := \operatorname{dist}(x, \operatorname{conv} A)$, *where the distance is calculated with respect to the Euclidean metric. Then*

$$\mathbf{E}\, e^{\frac{1}{2}\phi_A^2} \leqslant 1/\mu(A) \tag{5.60}$$

and so $\mu(\phi_A > t) \leqslant e^{-t^2/2}/\mu(A)$ *for* $t > 0$. *Consequently, if* $f : [0,1]^n \to \mathbb{R}$ *is a convex (or concave)* 1*-Lipschitz function and* M *is its median with respect to* μ, *then* $\mu(f > M + t) \leqslant 2e^{-t^2/2}$ *for* $t > 0$.

In the statement of Theorem 5.54 we tacitly assume that μ is a measure on $\mathbb{R}^n$ supported on $\{0,1\}^n$. The second assertion of the Theorem follows from (5.60) by Markov's inequality. Some finer issues related to the derivation of the last assertion are addressed in Exercise 5.67. See also Exercise 5.68.

Theorem 5.54 turned out to be very useful (for example in the context of random matrices) and has been generalized in various ways. Here is one possible statement.

Theorem 5.55 (not proved here). *Let* $V_1, V_2, \ldots, V_N$ *be finite-dimensional normed spaces and let* $V = \bigoplus_{j=1}^N V_j$ *be their sum in the* ℓ_q*-sense (for some* $q \geqslant 2$*). For* $j = 1, 2, \ldots, N$, *let* μ_j *be a measure on* V_j *supported on a set of diameter at most* 1 *and let* $\mu = \otimes_{j=1}^N \mu_j$. *Further, assume that* $F : V \to \mathbb{R}$ *is* 1*-Lipschitz and quasiconvex (i.e.,* $F^{-1}\big((-\infty, a]\big)$ *is convex for all* $a \in \mathbb{R}$*) or quasiconcave. Then*

$$\mu(F > M + t) \leqslant 2e^{-\frac{1}{4}t^q} \quad \text{for all} \quad t > 0, \tag{5.61}$$

where M *is the median of* F *with respect to* μ.

We conclude this section with a result that is the counterpart of Theorem 5.54 with the median replaced by the mean, whose degree of generality is intermediate between those of Theorem 5.54 and Theorem 5.55.

Theorem 5.56 (Convex concentration inequality for the mean, not proved here). *Let* $\mu = \mu_1 \otimes \cdots \otimes \mu_k$ *be a product measure on* $[0,1]^n \subset \mathbb{R}^n$ *and let* $f : [0,1]^n \to \mathbb{R}$ *be a function which is* 1*-Lipschitz with respect to the Euclidean distance and convex with respect to each variable. Then, for any* $t \geqslant 0$,

$$\mu(f > \mathbf{E}f + t) \leqslant e^{-t^2/2}. \tag{5.62}$$

While, by Remark 5.12 (which was based on the very general results from Section 5.2.3.2), statements about concentration around the median formally imply similar statements about the mean, we state Theorem 5.56 separately since it combines good constants with a different set of hypotheses.

Exercise 5.64 (Concentration on even-dimensional Boolean cube). If $n = 2m$ is even, an example of a set $A \subset \{0,1\}^n$ with $\mu(A) = \frac{1}{2}$ that is optimal in the sense of Theorem 5.51 is $A = \big\{\sum_{j=1}^n y_j < m\big\} \cup \big\{\sum_{j=1}^n y_j = m \text{ and } y_1 = 1\big\}$. Show that also in this case $\mu(A_{s/n}) \geqslant 1 - e^{-2s^2/n}$ for $s \in \mathbb{N}$.

Exercise 5.65. Show that the bound $\mu(A_\varepsilon) \geqslant 1 - e^{-2n\varepsilon^2}$ from Corollary 5.52 may fail for some $\varepsilon > 0$ if $n = 1$ or 2, but that it always holds if $n > 2$ or if $\varepsilon \geqslant 1/n$.

Exercise 5.66 (Non uniqueness in Harper's theorem). Give an example of a value N and two sets of N elements in $\{0,1\}^4$ with smallest ε-enlargements (for all values of ε) among sets with N elements, which are distinct up to symmetries of the hypercube. **Note**: it appears to be unknown whether uniqueness can be assured by insisting that both A and its complement are isoperimetric sets for all sizes of enlargement.

Exercise 5.67 (Talagrand's concentration inequality for concave functions). Derive the bound $\mu(f > M + t) \leqslant 2e^{-t^2/2}$ for concave f in Theorem 5.54 (or, equivalently, $\mu(f < M - t) \leqslant 2e^{-t^2/2}$ for convex f) from the inequalities preceding it.

Exercise 5.68 (Existence of convex Lipschitz extensions). Let $K \subset \mathbb{R}^n$ be a convex set and let $f : K \to \mathbb{R}$ be a convex 1-Lipschitz function. Then f admits a convex 1-Lipschitz extension to $\mathbb{R}^n$. Consequently, in Theorem 5.54 it doesn't matter whether we assume f to be convex and 1-Lipschitz on $\mathbb{R}^n$ or just on $[0, 1]^n$.

Exercise 5.69 (No dimension-free subgaussian bound in absence of convexity). Here is an example showing that convexity is crucial in Theorem 5.54. Define $f : \{-1, 1\}^n \to \mathbb{R}$ by $f(x_1, \dots, x_n) = \max(0, x_1 + \cdots + x_n)^{1/2}$. Show that f has median 0 and is $\frac{1}{\sqrt{2}}$-Lipschitz with respect to the Euclidean metric, while $\mu\left(f > cn^{1/4}\right) \geqslant c$ for some absolute constant $c > 0$.

5.2.6. Deviation inequalities for sums of independent random variables. In this section we gather some simple but useful facts about deviation inequalities for sum of independent mean zero random variables. We mostly focus on two families of random variables: subgaussian and subexponential variables.

In a probabilistic setting, the L_p-norm (for $p \geqslant 1$) of a random variable X is $\|X\|_p = (\mathbf{E}\,|X|^p)^{1/p}$. As a preliminary step, consider two prototypical examples: let Z be an $N(0, 1)$ random variable and T be a symmetric exponential variable with parameter 1 (i.e., $\mathbf{P}(T > t) = \mathbf{P}(-T > t) = \frac{1}{2}e^{-t}$ for $t > 0$). A simple computation (cf. (A.1)) shows that

$$\|Z\|_p = \frac{\sqrt{2}}{\pi^{1/2p}}\Gamma\left(\frac{p+1}{2}\right)^{1/p} \sim \sqrt{\frac{p}{e}}, \tag{5.63}$$

$$\|T\|_p = \Gamma(p+1)^{1/p} \sim \frac{p}{e} \tag{5.64}$$

as p tends to infinity.

The growth of the L_p-norms motivates the following definitions: a random variable X is said to be *subgaussian* (or ψ_2) when

$$\|X\|_{\psi_2} := \sup_{p\geqslant 1} p^{-1/2}\|X\|_p < \infty. \tag{5.65}$$

This terminology is consistent with that introduced in the preamble to Section 5.2 and based on the tail behavior (cf. (5.21), (5.22); see Exercise 5.70 and Lemma 5.57 below). Similarly, X is said to be *subexponential* (or ψ_1) when

$$\|X\|_{\psi_1} := \sup_{p\geqslant 2} \frac{\|X\|_p}{\|T\|_p} < \infty. \tag{5.66}$$

The reader may be familiar with the arguably less *ad hoc* forms of ψ_r conditions, based on either the rate of growth of the (bilateral) Laplace transform or the appropriate Orlicz norms, or on the tail behavior of the type

$$\mathbf{P}(|X| > t) \leqslant Ce^{-\lambda t^r} \quad \text{for} \quad t \geqslant 0$$

(cf. (5.21) and (5.22)). There is no need to be alarmed, though: while not identical, all these approaches lead to quantities that are equivalent up to universal constants. The definitions (5.65)–(5.66) were chosen out of convenience in view of the sample applications we present. See Notes and Remarks for more details and references.

If follows from (5.63) and (5.64) that $\|T\|_{\psi_1} = 1$, $\|Z\|_{\psi_2} = \sqrt{2/\pi}$ and that $\|\cdot\|_{\psi_1} \leqslant \|\cdot\|_{\psi_2}$ (see Exercise 5.75). We have obviously $\|\cdot\|_{\psi_2} \leqslant \|\cdot\|_\infty$ and $\|\cdot\|_{\psi_1} \leqslant \|\cdot\|_\infty$, so the present discussion also applies to bounded variables. Another important example of subgaussian variables is obtained by taking the inner product with a fixed vector of a randomly chosen unit vector in $\mathbb{R}^d$ or $\mathbb{C}^d$. This has to be compared with Poincaré's lemma (Theorem 5.22) which says that the Gaussian measure appears at the limit $d \to \infty$.

LEMMA 5.57. *If X is uniformly distributed on S^{d-1} (resp., $S_{\mathbb{C}^d}$), then for every $u \in \mathbb{R}^d$ (resp., $u \in \mathbb{C}^d$), we have $\|\langle X, u\rangle\|_{\psi_2} \leqslant |u|/\sqrt{d}$.*

PROOF. We may assume by homogeneity that $|u| = 1$. Let G be a standard Gaussian vector in $\mathbb{R}^d$. The variable uniformly distributed on S^{d-1} can be then represented as $X = G/|G|$. Moreover, $|G|$ is independent of X and hence, for $p \geqslant 1$,

$$\|\langle G, u\rangle\|_p = \||G|\|_p \|\langle X, u\rangle\|_p.$$

We have $\||G|\|_p \geqslant \||G|\|_1 = \kappa_d$ (see Section 4.3.3). Since $\langle G, u\rangle$ has distribution $N(0,1)$, we know from (5.63) that $\|\langle X, u\rangle\|_{\psi_2} = \sqrt{2/\pi} = \kappa_1$. Therefore, using Proposition A.1(ii), we obtain $\|\langle X, u\rangle\|_{\psi_2} \leqslant \frac{\kappa_1}{\kappa_d} \leqslant \frac{1}{\sqrt{d}}$. The complex case is similar. □

We also note that the square of a subgaussian variable is subexponential, as follows easily from the definitions. We now consider the case of a sum of either subgaussian or subexponential mean zero random variables. If the random variables are bounded, we can apply Hoeffding's inequality (5.43). It turns our that essentially the same result holds for subgaussian variables.

PROPOSITION 5.58 (see Exercise 5.73). *Let $X_1, \dots, X_n$ be independent subgaussian real random variables with mean zero, and $S = X_1 + \dots + X_n$. Define $K > 0$ by $K^2 = \|X_1\|_{\psi_2}^2 + \dots + \|X_n\|_{\psi_2}^2$. Then for every $t > 0$,*

$$\mathbf{P}(|S| > t) \leqslant 2\exp\left(-\frac{t^2}{8eK^2}\right).$$

The proof actually yields a better bound $2\exp(-\frac{t^2}{2eK^2})$ when (X_i) are symmetric random variables (i.e., such that X_i and $-X_i$ have the same distribution for any fixed i).

In the case of ψ_1 variables, the situation is slightly more complicated since two tails enter the picture: subgaussian tails for moderate deviations (which are reminiscent of the central limit phenomenon) and subexponential tails for large deviations (which come from the tails of individual variables)

PROPOSITION 5.59 (Bernstein's inequalities, see Exercise 5.76). *Let $X_1, \dots, X_n$ be independent real random variables with mean zero, and assume that $\|X_i\|_{\psi_1} \leqslant K$ for every index i. Then, for every vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and every $t \geqslant 0$,*

$$\mathbf{P}\left(\left|\sum_{i=1}^n a_i X_i\right| > t\right) \leqslant 2\exp\left(-\min\left(\frac{t^2}{8K^2\|a\|_2^2}, \frac{t}{4K\|a\|_\infty}\right)\right).$$

Remark 5.60. Propositions 5.58 and 5.59 readily generalize to the complex case (with possibly different numerical constants).

Exercise 5.70 (Lipschitz function on a Gaussian space is subgaussian). Let G be a standard Gaussian vector on $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ a 1-Lipschitz function such that $f(G)$ has mean zero. Deduce from the results of Section 5.2.2 that $\|f(G)\|_{\psi_2} \leqslant C$ for some absolute constant C. (Except for the value of the constant C, this is a generalization of Lemma 5.57.)

Exercise 5.71 (Khintchine inequalities). Let $X = \sum_{i=1}^n \varepsilon_i a_i$, where $a_1, \dots, a_n$ are real numbers and (ε_i) is a sequence of independent random variables with $\mathbf{P}(\varepsilon_i = 1) = \mathbf{P}(\varepsilon_i = -1) = 1/2$. Show that, for any $p \geqslant 1$,

$$A_p \|X\|_{L_2} \leqslant \|X\|_{L_p} \leqslant B_p \|X\|_{L_2}$$

where $A_p > 0$ and B_p are constants depending only on p. Show that $B_p = O(\sqrt{p})$ as $p \to \infty$.

Exercise 5.72 (Khintchine–Kahane inequalities). Khintchine inequalities have a vector-valued generalization which is due to Kahane: *If $x_1, \dots, x_n$ belong to some normed space Y and X' denotes the random variable $\|\sum_{i=1}^n \varepsilon_i x_i\|_Y$, then*

$$A'_p \|X'\|_{L_2} \leqslant \|X'\|_{L_p} \leqslant B'_p \|X'\|_{L_2}$$

where $A'_p > 0$ and B'_p are constants depending only on p. Prove this. Moreover, we have $A_1 = A'_1 = 1/\sqrt{2}$ and $B'_p = \Theta(\sqrt{p})$ as $p \to \infty$.

Exercise 5.73. Prove Proposition 5.58 by following the outline given below.
(i) If X is symmetric, show that $\mathbf{E}\exp(\lambda X) \leqslant \exp(\frac{e}{2}\|X\|_{\psi_2}^2 \lambda^2)$ for any $\lambda > 0$.
(ii) Let Y be an independent copy of a mean zero random variable X. Show that $\mathbf{E}\exp(\lambda X) \leqslant \mathbf{E}\exp(\lambda(X - Y))$. Using this symmetrization trick, deduce from (i) that the inequality $\mathbf{E}\exp(\lambda X) \leqslant \exp(2e\|X\|_{\psi_2}^2 \lambda^2)$ holds for any mean zero random variable X.
(iii) Deduce Proposition 5.58 using Lemma 5.28.

Exercise 5.74 (Linear combinations of subgaussian random variables are subgaussian). Show the following variant of Proposition 5.58: if $X_1, \dots, X_n$ are independent and mean zero, then $\|X_1 + \dots + X_n\|_{\psi_2} \leqslant C(\|X_1\|_{\psi_2}^2 + \dots + \|X_n\|_{\psi_2}^2)$ for some absolute constant C.

Exercise 5.75. Verify that $\|Z\|_{\psi_2} = \sqrt{2/\pi}$ and that, for any variable X, $\|X\|_{\psi_1} \leqslant \|X\|_{\psi_2}$.

Exercise 5.76 (Bernstein's inequalities). (i) Show that if $\mathbf{E}X = 0$ and $\|X\|_{\psi_1} \leqslant 1$, then $\mathbf{E}\exp(\lambda X) \leqslant 1 + 2\lambda^2 \leqslant \exp(2\lambda^2)$ for $|\lambda| < 1/2$ (cf. Lemma 5.28).
(ii) Under the hypotheses of Proposition 5.59, assuming $K = 1$ and denoting $S = a_1 X_1 + \dots + a_n X_n$, prove that $\mathbf{E}\exp(\lambda S) \leqslant \exp(2\lambda^2 \sum a_i^2)$ for $|\lambda| \leqslant 1/(2\|a\|_\infty)$.
(iii) Prove Proposition 5.59.

Notes and Remarks

Section 5.1. An encyclopedic reference for sphere packings is the book [**CS99**]. Other valuable and historically significant references are [**Rog64, Bör04, FT97**].

Packing and covering on the Euclidean sphere and the discrete cube. To complement Proposition 5.1, it has been proved in [**BGK⁺01**] that for $0 \leqslant t \leqslant$

arccos $\sqrt{2/n}$, we have $V(t) \geqslant (6\sqrt{n}\cos t)^{-1}(\sin t)^{n-1}$ (similar estimates appear in [**Bör04**], Lemma 6.8.6). For some values of n, t (roughly for $t > 1.14$ and for large n), this is better than the lower bound from (5.4), and similarly superior to the improved bound from Exercise 5.4 if $t > 1.221$.

The random covering argument from Proposition 5.4 is due to Rogers [**Rog57, Rog63**]. The factor $Cn \log n$ from Corollary 5.5 is usually referred to as the *density of the covering*, even though calling it "the overlap" or "the redundancy" would seem more logical. Both the original Rogers's argument, and the one presented here, allow achieving $C = 1$ at the expense of additional lower order terms (see Exercise 5.8 and its hint). Recent advances by Dumer [**Dum07**] improve the bound on the density to $(\frac{1}{2} + o(1))n \log n$. The paper [**Dum07**] establishes also a density bound $\frac{1}{2}n \log n + 2n \log\log n + 5n$, valid for all $\varepsilon \in (0,1)$ and all $n \geqslant 4$. It should be noted, however, that the latter result deals with a slightly easier problem, covering the sphere $S^{n-1} \subset \mathbb{R}^n$ by balls whose centers are not required to belong to S^{n-1} (i.e., with the parameter N' from Exercise 5.1). Finally, at the price of increasing the constant C, the result from Corollary 5.5 can be strengthened as follows: for any dimension n and angle ε, there is a covering of S^{n-1} by caps of radius ε such that any point belongs to at most $400n \log n$ caps [**BW03**].

Since the sphere looks locally like a Euclidean space, as the radii of the caps tend to 0, the packing/covering problems for S^{n-1} converge to the corresponding problems for $\mathbb{R}^{n-1}$. (The original random covering argument of Rogers [**Rog57**] considered an even more general question, economical coverings of $\mathbb{R}^n$ by translates of an arbitrary convex body—the spherical variant being an afterthought—and led to an upper bound of $n \log n + n \log\log n + 5n$ for the appropriately defined asymptotic density.) In that setting, a lower bound on density of optimal coverings by Euclidean balls is $\Omega(n)$ [**CFR59**] and this estimate can be transferred back to S^{n-1} if the radius is small enough; see Example 6.3 in [**BW03**] for an argument that works if $\varepsilon \leqslant \arcsin(1/\sqrt{n})$.

References for the results mentioned about packing are [**Ran55**] (Rankin) and [**KL78**] (Kabatjanskiĭ–Levenšteĭn), we refer to [**CS99**] for more information (see also [**BN06a**]). Again, when the radius of the cap tends to 0, the problem becomes the classical sphere packing problem in $\mathbb{R}^n$. In this context, a classical result due to Minkowski–Hlawka shows the existence of lattice packings of Euclidean balls (or actually, of any symmetric convex body) in $\mathbb{R}^n$ which cover a proportion $1/2^{n-1}$ of the space (a.k.a. *packing density*). Remarkably, this result has been only marginally improved in the past century [**Rog47, DR47, Bal92b**] and is exponentially far from Kabatjanskiĭ–Levenšteĭn upper bound—which is approximately of order 0.66^n—for the proportion covered by a (non-necessarily) lattice packing (see [**Gru07**] for more on this topic).

Covering and particularly packing in the Hamming cube is of fundamental importance in coding theory, see, e.g., [**Rot06, CHLL97**]. The case of (very small) balls of radius $1/n$ in $\{0, \ldots, q-1\}^n$ is treated in [**KP88**].

The Gilbert–Varshamov bound has been improved in the q-ary cube for certain large values of q in [**TVZ82**], using a link with modular curves.

Packing and covering for convex bodies. For early references on metric entropy of convex bodies see [**CS90**], [**Pis89b**].

The arguments from [**Bar14**] imply the following improvement on the volumetric bound from Corollary 5.10: for $\varepsilon \in (0,1)$, any symmetric convex body in

$\mathbb{R}^n$ is $(1+\varepsilon)$-close in Banach–Mazur distance to a polytope with $(C/\sqrt{\varepsilon})^n$ vertices. (This is sharp: consider the case of the sphere.) To the best of our knowledge, it is not known whether analogous statement holds for not-necessarily symmetric bodies and the affine version (4.2) of the Banach–Mazur distance. Similar questions can be considered for large ε, or even ε growing with the dimension. In the case of the sphere, this is essentially the problem considered in Exercise 5.13. Again, [**Bar14**] contains good estimates in the general case. However, the bounds from [**Bar14**] deteriorate as the *asymmetry* of the body (defined, for example, as the minimal distance d_{BM} to a symmetric body) increases. Estimates that are superior for some ranges of parameters can be found in [**Sza**].

Let us also mention an important open problem, known as the duality conjecture: *do there exist absolute constants $c, C > 0$ such that for every two symmetric convex bodies $K, L \subset \mathbb{R}^n$ we have*

$$\log N(L^\circ, K^\circ) \leqslant C \log N(K, cL)? \tag{5.67}$$

This was proved when K or L is the Euclidean ball [**AMS04**] and extended to the case when a bound on the K-convexity constant (as defined in Section 7.1.2) is present in [**AMSTJ04**]. Another possible generalization to the setting of non-symmetric convex bodies is more tricky; in that case, even the proper formulation of (5.67) is not entirely clear.

A deep fact about covering numbers is the following ([**Mil86**], see also the discussion in [**Pis89b**]): there is an absolute constant C such that, for every symmetric convex body $K \subset \mathbb{R}^n$ there is an 0-symmetric ellipsoid $\mathscr{E}$ such that

$$\max\left(N(K, \mathscr{E}), N(\mathscr{E}, K)\right) \leqslant C^n. \tag{5.68}$$

Note that since metric entropy duality (5.67) is known to hold when one of the bodies is an ellipsoid, it follows then that similar bounds automatically hold also for $N(K^\circ, \mathscr{E}^\circ)$ and $N(\mathscr{E}^\circ, K^\circ)$. (In the original definitions, all four quantities were included explicitly or implicitly.) Such an ellipsoid $\mathscr{E}$ is called an *M-ellipsoid for* K, and K is said to be in the *M-position* when B_2^n is an M-ellipsoid for K. The M-ellipsoids are discussed in detail in [**AAGM15**].

Metric entropy of classical manifolds. Theorem 5.11 is from [**Sza82**], which covers the case of all metrics induced by unitarily invariant norms (see also [**Sza83, Sza98**] and [**Paj99**]). Examples of packings in some Grassmannians (mostly low-dimensional), some of them optimal, can be found in [**CHS96, SS98**]. More recent references, motivated by information transmission issues and concentrated on different asymptotics (k fixed and n tending to infinity), are [**BN02, BN05, BN06b**]. It appears that the theoretical computer science community is not aware that questions of that nature were considered in AGA already in 1980s.

Section 5.2. Classical general references about concentration of measure are [**Led01**] and [**Sch03**]. We particularly recommend the recent monograph [**BLM13**]. For a presentation directed towards applications to data science, see [**Ver**].

Isoperimetry and concentration. A geometry-oriented reference about isoperimetric inequalities is [**BZ88**]. The paternity of the isoperimetric inequality on the sphere (Theorem 5.13) is usually attributed to Lévy [**Lév22, Lév51**] although the arguments he presented were not fully rigorous; [**Sch48**] is usually cited as the first rigorous proof. Remarkably, the functional version (Lévy's lemma, in the language of our Corollary 5.17) appears explicitly in [**Lév22**] (see p. 279) and is therefore almost one century old!

A self-contained proof of the isoperimetric inequality on S^{n-1}, based on the concept of spherical symmetrization, appears in [**FLM77**]. Another symmetrization procedure (the two-point symmetrization) is applied in [**Ben84**]. The simple proof of the non-sharp inequality from Proposition 5.15 is based on [**AdRBV98**]. Proposition 5.20 is from [**JS**].

The Gaussian isoperimetric inequality was proved independently by Borell [**Bor75b**] and Sudakov–Tsireslon [**SC74**]. For a proof of Poincaré's lemma (Theorem 5.22) going beyond the weak convergence version from Exercise 5.29, we refer to [**DF87**] (which also advocates that the statement was first formulated by Borel and not by Poincaré). See also [**Led96**] and references therein. For a direct proof of concentration of measure on Gauss space, see [**Pis86**].

Ehrhard's inequality (5.31) was proved in [**Ehr83**] for convex sets, then extended in [**Lat96**] to the case where only one of the sets set is convex, with the general case being treated in [**Bor03**]. *A priori*, deriving an isoperimetric inequality such as (5.29) requires validity of (5.31) for an arbitrary Borel set and a ball; the paper [**Ehr83**], however, contains a direct application of the technique to prove (5.29). A general reference for this circle of ideas is [**Lat02**].

The concept of central values was formalized and applied in the context of QIT in [**ASW11**], which also contains versions of Corollaries 5.32 and 5.35. However, instances of the arguments can be found in [**Has09**] and in AGA literature dating to (at least) 1980s.

Proposition 5.34 appears in [**Dmi90, Kwa94, Fer97**]. Exercise 5.48 appears as Proposition 1.7 in [**Led01**]. Proposition 5.37 is Corollary 1.17 from [**Led01**].

There are various generalizations of Hoeffding's inequality appearing in Exercise 5.57, notably due to Azuma [**Azu67**] and McDiarmid [**McD89**] in the context of martingales.

Geometric and analytical methods. General references for Section 5.2.4 are [**MS86, Sch03, DS01, GM00, BLM13, BGL14, GZ03**].

Gromov's comparison theorem (Theorem 5.38) appeared first in the preprint [**Gro80**]. A proof can be found in an appendix in [**MS86**]. A new proof and an extension to non-Riemannian spaces was proposed recently in [**CM15**]. While the theorem is sharp as stated, there is a reason to suspect that a more precise result should be available: the proof proceeds via a local/variational argument and the *globally* normalized volume appears only *a posteriori*. A more satisfactory variant appears in [**Mil15**]. In addition to the curvature, it takes into account the actual diameter of the manifold in question, which may be strictly smaller than the bound following indirectly from the curvature. However, since the results in [**Mil15**] necessarily involve model manifolds more complicated than spheres, their statements are somewhat technical.

The case of manifolds of dimension 1 is a little special. First, while the definition of Ricci curvature in dimension 1 needs to be properly construed, the only sensible value is 0 since every such manifold looks locally like a segment. Accordingly, Proposition 5.41 is then vacuously true. Next, the solution to the isoperimetric problem in S^1 (resp., in $\mathbb{R}$) is very simple: among sets of any (positive, but not full) measure, the boundary is the smallest if it consists of exactly two points.

Consequently, the solutions, both for the "smallest boundary" and the "smallest enlargement" problems, are arcs (resp., segments). However, finer analytic statements (including but not limited to LSI) are interesting and highly nontrivial already in dimension 1. For example, in view of Proposition 5.44, the validity of (5.48) for the 1-dimensional Gaussian measure implies the same inequality in any dimension (with the same constant α, which, in view of Proposition 5.42, can be taken to be 1, which is optimal). Indeed, even statements about spaces consisting of only two points can be deep as for example in the elementary proof of the Gaussian isoperimetric inequality presented in [**Bob97**]. We will return to the same theme further when reporting on developments directly related to LSI and hypercontractivity.

Log-Sobolev inequalities (LSI) were introduced in a seminal paper by Gross [**Gro75**]. Again, the case of manifolds of dimension 1 (segments, circles) is a little special; see [**GMW14**] for an elementary overview of this aspect of the subject and for references. The link with concentration of measure (the Herbst argument) originates in an unpublished letter from Herbst to Gross. The connection between LSI, Ricci curvature, and the Hessian of the density was put forward in [**BÉ85, Bak94**]. For a comprehensive treatment of functional inequalities (including complete references), see [**BGL14**]. Another fruitful approach is the connection between LSI and the quadratic transportation cost inequalities; see Chapter 6 in [**Led01**].

As exemplified in Table 5.4, the values of the Poincaré constants can often be computed exactly. Indeed, the Poincaré inequality (5.54) can be rewritten as $\mathbf{Var}_\mu f \leqslant \alpha \int (-\Delta f) f \, d\mu$, where Δ is the Laplace–Beltrami operator on $L_2(X,\mu)$. It follows that the optimal α is equal to the reciprocal of the "spectral gap," i.e., the smallest nonzero eigenvalue of $-\Delta$. In some examples the eigenfunctions of the Laplace–Beltrami operator can be explicitly described: for the Gauss space they are the Hermite polynomials, for the sphere they are the spherical harmonics (see the elementary [**See66**], or [**BGM71**] which covers also the case of the projective spaces). On S^{n-1}, equality in (5.54) is achieved for functions of the form $x \mapsto \langle x, y\rangle$ with $y \in \mathbb{R}^n$. For Lie groups there is a connection with the spectrum of the Casimir operator and representations of the associated Lie algebra (see Proposition 10.6 in [**Hal15**]), which allows to derive the entire spectrum of $-\Delta$. The case of $\mathsf{SO}(n)$ and $\mathsf{SU}(n)$ appears in [**SC94**] (for $\mathsf{U}(n)$, see [**Voi91**]). Note that in these examples there is equality in (5.54) when f is a function of the form $M \mapsto \mathrm{Tr}(AM)$ for $A \in \mathsf{M}_n$. For a complete list of semisimple Lie algebras, see [**Rot86**]. The spectrum of Grassmann manifolds is considered in [**Tsu81, EC04, TK04, Hal07**], which allows in principle to retrieve the value of the Poincaré constant for specific dimensions if needed.

Hypercontractivity for the Ornstein–Uhlenbeck semigroup (Proposition 5.47) has been first established by Nelson [**Nel73**]. The connection with log-Sobolev inequalities was put forward by Gross [**Gro75**].

In many situations, the Gaussian case can be treated as a limit case from the case of the hypercube via the central limit theorem. By the tensorization property (Proposition 5.44), this amounts ultimately to verifying statements about the two-point space $\{-1, 1\}$ (see [**Gro75**] for a proof of the Gaussian LSI along these lines). The hypercontractivity inequality on the discrete cube is known as the Bonami–Beckner inequality [**Bon70, Bec75**]. Some variants of Proposition 5.48 appear in [**Jan97**]. For a more sophisticated technology giving sharp estimations on the moments of Gaussian polynomials (or Gaussian chaoses) see [**Lat06**]. The statement about concentration on polynomials on products of spheres appearing in Remark 5.50 follows from the proof of Corollary 12 in [**Mon12**].

Discrete settings. A reference focusing on the case of the hypercube is [**O'D14**] (it contains in particular the versions of Proposition 5.48 and Corollary 5.49 for the hypercube alluded to in Remark 5.50). In addition to [**O'D14**], general references for Section 5.2.5 are [**Mat02, McD98**]. The main statement of Theorem 5.51 was proved in [**Har66**] and rediscovered in [**Kat75**]. A short proof may be found in [**FF81**]; we also recommend the reference [**Lea91**]. Theorem 5.51 deals with vertex-isoperimetry. If we consider instead edge-isoperimetry (minimizing the number of edges joining A to A^c), the optimal sets are no longer Hamming balls but subcubes.

Theorem 5.54 is taken from [**Tal88**] (Note that [**Tal88**] states the result for the cube $\{-1,1\}^n$ and so the coefficient in the exponent in the estimate corresponding to (5.60) is there $\frac{1}{8}$.) Theorem 5.55 appears in [**JS91**] and [**Mec04**]. The latter paper addresses general unconditional direct sums and not only ℓ_q-sums; see also [**Mec03**]. Similar results, but with quite different proofs were presented in [**Mau91**] and [**Dem97**]. The most abstract (and most flexible) statements are arguably in [**Tal95, Tal96b, Tal96a**]. The arguments addressing settings more general than that of Theorem 5.54 usually led to a coefficient $\frac{1}{4}$ in the exponent as in (5.61), except for [**Tal95**], which includes a statement (Theorem 4.2.4) featuring coefficient $\frac{1}{2}$, but at the cost of introducing additional factors of lower order and restricting the range of t. A clean proof of Theorem 5.56 (which also has coefficient $\frac{1}{2}$ in the exponent) can be found in [**BLM13**]; the argument is attributed to [**Led97**] and the result itself to [**Tal96b**].

Deviation inequalities. Some references for Section 5.2.6 are [**Ver12**] and [**CGLP12**] (the latter treats also the case of intermediate growth between subgaussian and subexponential). As pointed out in the main text, there are several possible forms of ψ_r conditions and of definitions of the ψ_r-norms. The original ones were (presumably) in terms of Orlicz/Young functions: given an increasing convex function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\psi(0) = 0$ and $\psi(x) \to \infty$ as $x \to \infty$, we may define a the ψ-norm of a random variable X as (for example)

$$\|X\|_\psi = \inf\{c > 0 \ : \ \mathbf{E}\,\psi(|X|/c) \leqslant \psi(1)\}.$$

If one considers $\psi_r(x) = \exp(x^r) - 1$ ($r \geqslant 1$), then, for $r = 1, 2$, one gets norms which are equivalent (although not equal) to the ones defined in (5.66) and (5.65). For precise statements and proofs, see Theorem 1.1.5 in [**CGLP12**], which also covers the link to (the rate of growth of) the Laplace transforms mentioned in the main text; cf. Lemma 5.28 and Exercise 5.76. Overall, Section 1.1 of [**CGLP12**] is an excellent reference for ψ_r conditions/norms, which are otherwise difficult to extract from books/surveys on the more general Orlicz spaces.

For a historical account of Bernstein's contributions, we refer to pp. 126–128 in [**AAGM15**]. For more precise results about moments of sums of independent variables, see [**Lat97**]. For non-commutative analogues of these inequalities (i.e., for sums of random matrices), see [**Tro12**].

Finally, among other techniques to prove concentration of measure, we mention the so-called martingale method which implies for example concentration on permutation groups (see [**Sch82, Mau79, MS86**]): *If we equip the symmetric group $\mathfrak{S}_n$ with the uniform probability measure and the distance $d(\sigma,\tau) =$*

$\frac{1}{n}\,\mathrm{card}\{i \;:\; \sigma(i) \neq \tau(i)\}$, *then any* 1*-Lipschitz function* f *on* $(\mathfrak{S}_n, d)$ *satisfies* $\mathbf{P}(f \geqslant \mathbf{E}f + t) \leqslant \exp(-nt^2/8)$ *for any* $t \geqslant 0$.

The best constants in Khintchine inequalities (see Exercise 5.72) have been found in [**Sza76**] (who proved $A_1 = 1/\sqrt{2}$) and in [**Haa81**] (for $p > 1$). The Khintchine–Kahane inequalities from Exercise 5.72 were first proved in [**Kah85**]. The correct asymptotic order of the constants as $p \to \infty$ was found in [**Kwa76**], while the value $A'_1 = 1/\sqrt{2}$ is from [**LO94**]. A complete proof of the Khintchine–Kahane inequalities can be found by consulting Theorem 3.5.2 of [**AAGM15**].