Isoperimetry and concentration of measure

D.J.H. Garling

Contents

Introduction

The course comes in three parts. In the first, we prove isoperimetric theorems in four classical settings:

- d-dimensional Euclidean space, with its usual measure;
- the d-dimensional hypercube.
- the surface of a d-dimensional sphere;
- d-dimensional space, with Gaussian measure;

Each of these theorems requires its own technique, and we shall establish all the results that we shall need on the way (the Prékopa-Leindler inequality, the Brunn-Minkowski inequality, Haar measure, Poincaré's Lemma,...).

In the second part, we investigate how the isoperimetric theorems are used in studying the geometry of Banach spaces, and prove Dvoretzky's Theorem on spherical sections.

A typical application of the isoperimetric theorems is that in high dimensions, a Lipschitz function takes values near its median with high probability, and the probability of large deviations is small. This is known as the *concentration of measure phenomenon*, or *the theory of large deviations*. In the third part of the course, we shall study this, even in settings where an isoperimetric theorem does not exist, such as in Σ_n.

In spite of the geometric setting, this will be a course on Analysis and Probability. Attendance at the Part II courses on Probability and Measure, and Linear Analysis, or their equivalents, will be an advantage, as will be attendance at Michaelmas term Analysis courses. In the third part, we shall study probabilities on metric spaces: the book by Dudley is an excellent reference for this.

Reading list

(i) K.M. Ball, An elementary introduction to modern convex geometry, in *Flavors of Geometry*, edited by Silvio Levy, CUP 1997.

(ii) Y. Benyamini, Two point symmetrization, the isoperimetric inequality on the sphere and some applications, *Longhorn notes*, Texas functional analysis seminar, 1983-1984, pp53-76.

(iii) R.M. Dudley, *Real Analysis and Probability*, second edition, CUP 2002.

(iv) W.B. Johnson and J. Lindenstrauss (Editors), *Handbook of the geometry of Banach spaces, Volumes I and 2*, North Holland 2001, 2003. (Chapters 1, 4, 8, 17, 27, 28, 37).

(v) I.B. Leader, Discrete isoperimetric inequalities, in *Probabilistic combinatorics and its applications* (Proceedings of Symposia in Applied Mathematics 44) AMS 1992.

(vi) M. Ledoux, *The Concentration of measure phenomenon*, AMS 2001.

(vii) M. Ledoux, *Isoperimetry and Gaussian analysis*, Saint-Flour Summer School, 1994, Springer Lecture Notes, Volume 1648.

(viii) V.D. Milman, G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Springer Lecture Notes, Volume 1200.

(ix) G. Pisier, *The volume of convex bodies and Banach space geometry*, CUP 1989.

(x) N. Tomczak-Jaegermann, *Banach space distances and finite-dimensional operator ideals*, Pitman 1989.

(xi) Robert J. Zimmer, *Essential results of functional analysis*, Chicago Lectures in Mathematics, University of Chicago Press 1990.

1

Isoperimetry

1.1 Isoperimetry in $\mathbf{R}^d$

Suppose that $1 \leq p < \infty$ and that $1/p + 1/p' = 1$.

We have Hölder's inequality:

if $f \in L^p(\Omega, \Sigma, \mu)$ and $g \in L^{p'}(\Omega, \Sigma, \mu)$ then $fg \in L^1(\Omega, \Sigma, \mu)$ and

$$\|fg\|_1 = \int |fg| \, d\mu \leq (\int |f|^p \, d\mu)^{1/p} (\int |g|^{p'} \, d\mu)^{1/p'} = \|f\|_p \, \|g\|_{p'} \, .$$

Our first inequality goes in the opposite direction, for functions on $\mathbf{R}^d$ with Lebesgue measure λ.

Proposition 1.1.1 *Suppose that A and B are sets of finite positive measure in $\mathbf{R}$. Then*

$$\lambda(A/p + B/p') \geq \lambda(A/p) + \lambda(B/p') = \lambda(A)/p + \lambda(B)/p'.$$

Proof First observe that we can translate sets. Let $A_k = A - k$, $B_l = B - l$. Then $A_k/p + B_l/p' = A_{k/p+l/p'}$; if we prove the result for A_k and B_l, then we get the result for A and B.

First suppose that A and B are compact subsets of $\mathbf{R}$. By translating, we can suppose that $\sup(A) = \inf(B) = 0$ so that $A \cap B = \{0\}$. Then $A/p + B/p'$ contains the (almost disjoint) union of A/p and B/p', so that $\lambda(A/p + B/p') \geq \lambda(A/p) + \lambda(B/p')$. If A is measurable, then

$$\lambda(A) = \sup\{\lambda(K) : K \text{ compact } K \subseteq A\},$$

so that an easy approximation argument shows that the result holds for general measurable A, B. □

Proposition 1.1.2 *Suppose that* $u \in L^p(\mathbf{R})$, $v \in L^{p'}(\mathbf{R})$ *and that*

$$|w(\frac{x}{p} + \frac{y}{p'})| \geq |u(x)||v(y)| \text{ for all } x, y \in \mathbf{R}.$$

Then $\|w\|_1 \geq \|u\|_p \|v\|_{p'}$.

Proof We can clearly suppose that $u, v, w \geq 0$ and that u and v are non-zero. Since

$$\|u\|_p = \sup\{\|f\|_p : f \text{ bounded }, |f| \leq u\},$$

we can also suppose that u and v are bounded and non-zero. By scaling, we can suppose that $\|u\|_\infty = \|v\|_\infty = 1$. Recall that $\int w\,d\lambda = \int_0^\infty \lambda(w > t)\,dt$. Let $A_t = (u > t^{1/p}) = (u^p > t)$, $B_t = (v > t^{1/p'}) = (v^{p'} > t)$. Then A_t and B_t are sets of finite positive measure for $0 < t < 1$, and $\lambda(A_t) = \lambda(B_t) = 0$ for $t \geq 1$. If $0 \leq t < 1$ and $x \in A_t$ and $y \in B_t$ then $w(x/p+y/p') > t^{1/p}t^{1/p'} = t$, so that $(w > t) \supset A_t/p + B_t/p'$ and $\lambda(w > t) \geq \lambda(A_t)/p + \lambda(B_t)/p'$. Integrating, and using Jensen's inequality,

$$\begin{aligned}
\int w\,d\lambda &= \int_0^\infty \lambda(w > t)\,dt \geq \int_0^1 \lambda(w > t)\,dt \\
&\geq \int_0^1 \lambda(A_t)/p + \lambda(B_t)/p'\,dt \\
&= \|u\|_p^p/p + \|v\|_{p'}^{p'}/p' \geq \|u\|_p \|v\|_{p'}.
\end{aligned}$$

□

Theorem 1.1.1 (The Prékopa-Leindler inequality) *Suppose that* $u \in L^p(\mathbf{R}^d)$,$v \in L^{p'}(\mathbf{R}^d)$ *and that*

$$|w(\frac{x}{p} + \frac{y}{p'})| \geq |u(x)||v(y)| \text{ for all } x, y \in \mathbf{R}^d.$$

Then $\|w\|_1 \geq \|u\|_p \|v\|_{p'}$.

Proof We prove this by induction on d. It is true for $d = 1$: suppose that it is true for $d-1$. Let $H_r = \{x : x_d = r\}$, and let us identify $\mathbf{R}^{d-1}$ with H_0, so that we can write a point of $\mathbf{R}^d$ as (x, t), with $x \in \mathbf{R}^d$ and $t \in \mathbf{R}$. Let us write $w_t(x) = w(x, t)$, etc. If $t = r/p+s/p'$ then $w_t(x/p+y/p') \geq u_r(x)v_s(y)$, and so by the $d-1$-dimensional result $\|w_t\|_1 \geq \|u_r\|_p . \|v_s\|_{p'}$. Thus if we set $W(t) = \|w_t\|_1$, $U(r) = \|u_r\|_p$ and $V(s) = \|v_s\|_{p'}$ then $W(r/p + s/p') \geq U(r)V(s)$. Applying, the one-dimensional result,

$$\|w\|_1 = \|W\|_1 \geq \|U\|_p \|V\|_{p'} = \|u\|_p \|v\|_{p'}.$$

$\square$

Corollary 1.1.1 (The Brunn-Minkowski inequality) *If A and C are sets of finite positive measure in $\mathbf{R}^d$ then*

$$(\lambda(A+C))^{1/d} \geq (\lambda(A))^{1/d} + (\lambda(C))^{1/d}.$$

Proof Let $\alpha = (\lambda(A))^{1/d}$ and $\gamma = (\lambda(C))^{1/d}$. Let $\tilde{A} = A/\alpha$, $\tilde{C} = C/\gamma$, so that $\lambda(\tilde{A}) = \lambda(\tilde{C}) = 1$. Let $u = I_{\tilde{A}}$, $v = I_{\tilde{C}}$. Then $\|u\|_p = \|v\|_{p'} = 1$. Let $1/p = \alpha/(\alpha+\gamma)$, so that $1/p' = \gamma/(\alpha+\gamma)$. Then $\tilde{A}/p = A/(\alpha+\gamma)$ and $\tilde{C}/p' = C/(\alpha+\gamma)$, so that

$$\tilde{A}/p + \tilde{C}/p' = (A+C)/(\alpha+\gamma)$$

and we can take $w = I_{(A+C)/(\alpha+\gamma)}$. Thus

$$\lambda((A+C)/(\alpha+\gamma)) = \|w\|_1 \geq \|u\|_p \|v\|_{p'} = 1,$$

and so $\lambda(A+C) \geq (\alpha+\gamma)^d$; taking d-th roots, we get the result. $\square$

We now obtain the isoperimetric inequality in $\mathbf{R}^d$. We avoid measuring surface areas in the following way. If A is a closed subset of a metric space (X,ρ) and $\epsilon > 0$, we set $A_\epsilon = \{x : d(x,A) \leq \epsilon\}$. In $\mathbf{R}^d$, $A_\epsilon = A + \epsilon B$, where B is the closed unit ball in $\mathbf{R}^d$.

Corollary 1.1.2 (The isoperimetric inequality in $\mathbf{R}^d$) *If A is a closed subset in $\mathbf{R}^n$ and $\lambda(A) = \lambda(B)$ then $\lambda(A_\epsilon) \geq \lambda(B_\epsilon)$.*

Proof By Brunn-Minkowski,

$$\begin{aligned}(\lambda(A_\epsilon))^{1/d} &= (\lambda(A+\epsilon B)^{1/d} \\ &\geq (\lambda(A))^{1/d} + \epsilon(\lambda(B))^{1/d} \\ &= (1+\epsilon)(\lambda(B))^{1/d} = (\lambda(B_\epsilon))^{1/d}.\end{aligned}$$

$\square$

Exercises

(i) Provide the 'easy approximation argument' required in Theorem 1.1.1.
(ii) Calculate $I_d = \int_0^\pi \sin^d\theta\, d\theta$.
(iii) Let B_p^d denote the unit ball of $\mathbf{R}^d$, equipped with the norm $\|x\|_p = (\sum_{i=1}^d |x_i|^p)^{1/p}$, and with $\|x\|_\infty = \max_i |x_i|$. Calculate the volumes of B_p^d for $p = 1, 2, \infty$, and their $(d-1)$-dimensional surface areas. How do these quantities behave as $d \to \infty$?

(iv) Let $r_{d,p}B_p^d$ be the multiple of B_p^d with the same volume as the euclidean ball B_2^d. How does $r_{d,p}$ behave as $d \to \infty$, for $p = 1, \infty$? How does the surface area of $r_{d,p}B_p^d$ compare with that of B_2^d, for $p = 1, \infty$?

(v) Suppose that A and B are non-empty compact sets in $\mathbf{R}^d$, and that $0 < t < 1$. Show that

$$\lambda((1-t)A + tB)^{1/d} \geq (1-t)(\lambda(A))^{1/d} + t(\lambda(B))^{1/d},$$

and deduce that

$$\lambda((1-t)A + tB) \geq (\lambda(A))^{1-t}(\lambda(B))^t.$$

Show that the Brunn-Minkowski inequality can be deduced from this last inequality.

1.2 Isoperimetry in the hypercube

Let $\mathbf{Q}^d = \{0,1\}^d = \{x = (x_1, \ldots, x_d) : x_i = 0 \text{ or } 1\}$, be the set of vertices of the unit hypercube in $\mathbf{R}^d$. The map $A \to I_A$ is a bijection from the set $P(\{1, \ldots, d\})$ of subsets of $\{1, \ldots, d\}$ onto $\mathbf{Q}^d$, so that we can identify $\mathbf{Q}^d$ with $P(\{1, \ldots, d\})$.

Let $d(x, y) = \sum_{i=1}^d |x_i - y_i| = \|x - y\|_1$. d is the *Hamming metric* on $\mathbf{Q}^d$. Let $l(x) = d(x, 0)$, and let $B_k = \{x : l(x) \leq k\}$, $S_k = \{x : l(x) = k\}$.

If A is a non-empty subset of $\mathbf{Q}^d$, let $N(A) = \{x : d(x, A) \leq 1\}$. What is $\min\{|N(A)| : |A| = k\}$? For what sets is the minimum attained?

We define a total order on $\mathbf{Q}^d$. If $x \neq y$, we set $x < y$ if either $l(x) < l(y)$ or $l(x) = l(y)$ and x comes before y in the (reverse?) lexicographic order: if $j = \inf\{i : x_i \neq y_i\}$ then $x_j = 1$ and $y_j = 0$. Thus I is an initial segment if and only if there exists r such that $B_r \subseteq I \leq B_{r+1}$ and $I \setminus B_r$ is an initial segment of S_{r+1} in the lexicographic order.

Theorem 1.2.1 (Harper's theorem) *If I is an initial segment of length k in $\mathbf{Q}^d$ and $|A| = k$ then $|N(A)| \geq |N(I)|$.*

Proof The proof is by induction on d. The case $d = 1$ is trivial. Suppose that the result is true for $d - 1$. For the moment, fix $1 \leq i \leq d$. Let $[i] = \{1, \ldots, d\} \setminus \{i\}$, and let $Q[i]$ be the corresponding hypercube. If $x \in \mathbf{Q}^d$ let $P(x)_j = x_j$ for $j \in [i]$. P is a $2 - 1$ mapping of $\mathbf{Q}^d$ onto $\mathbf{Q}[i]$.

For $\eta = 0, 1$, let $A_\eta = \{x \in A : x_i = \eta\}$, let $B_\eta = P(A_\eta)$, let $k_\eta = |A_\eta| = |B_\eta|$, and let I_η be the initial segment of $\mathbf{Q}[i]$ of length k_η. Let

$C_\eta = \{x \in \mathbf{Q}^d : x_i = \eta, P(x) \in I_\eta\}$ and let $C = C_0 \cup C_1$. Note that $|C| = |A|$.

We claim that $|N(A) \geq |N(C)|$. $P((N(A_0))_0) = N(B_0)$ and $P((N(A_1))_0) = B_1$, and $(N(A))_0 = (N(A_0))_0 \cup (N(A_1))_0$, so that $P((N(A))_0) = N(B_0) \cup B_1$, and $|(N(A))_0| = |N((B_0) \cup B_1|$. Similarly, $|(N(C))_0| = |N(I_0) \cup I_1|$. But $N(I_0)$ is also an initial segment, and so $N(I_0) \cup I_1$ is either $N(I_0)$ or I_1. Thus, using the inductive hypothesis,

$$\begin{aligned} |(N(A))_0| &= |N(B_0) \cup B_1| \geq \max(|N(B_0)|, |B_1|) \\ &\geq \max(|N((I_0)|, |I_1|) = |N(I_0) \cup I_1| = |(N(C))_0|. \end{aligned}$$

Similarly, $|(N(A))_1| \geq |(N(C))_1|$, and so $|N(A)| \geq |N(C)|$.

We call $C = C_i(A)$ the *compression* of A in the i direction. Note that $|C_i(A)| = |A|$ and that if $A \neq C_i(A)$ then

$$\sum\{l(x) : x \in C_i(A)\} < \sum\{l(x) : x \in A\}. \qquad (*).$$

Thus starting with A, either $A = C_i(A)$ for all i, or we compress in a certain direction. We iterate this. By $(*)$, the process must stop, and we obtain a set D such that $|D| = |A|$, $|N(D)| \leq |N(A)|$, and $D = C_i(D)$ for each i.

Unfortunately, this does not imply that D is an initial segment. Suppose that D is not an initial segment. There exist $x \notin D$ and $y \in D$ with $x < y$. Since $D = C_i(D)$, $x_i \neq y_i$. This holds for each i. Thus x is uniquely determined by y, and y is also uniquely determined by x. This means that $D = I \setminus \{x\}$, where I is an initial segment with largest element y, and x is the predecessor of y. This can happen uniquely, but in a different way, depending on the parity of d.

If $d = 2r + 1$ then $l(x) = r$ and $l(y) = r + 1$. Thus

$$x = (1, \dots, 1, 0, \dots, 0) \text{ (r ones), and } y = (0, \dots, 0, 1, \dots, 1) \text{ ($r+1$ ones).}$$

If $d = 2r$ then $l(x) = l(y) = r$. Thus

$$x = (0, 1, \dots, 1, 0, \dots, 0) \text{ (r ones), and } y = (1, 0, \dots, 0, 1, \dots, 1) \text{ (r ones).}$$

In either case, $|N(D)| \geq |N(J)|$, where J is the initial segment with $|D|$ elements. □

[This account of the theorem is based on the notes on 'Extremal combinatorics' on Paul Russell's DPMMS home page, and on the paper by Professor Leader listed in the introduction.]

Exercises

(i) Calculate $|B_k|$ and $|S_k|$

(ii) It is as natural to work with the probability measure $\mathbf{P}(A) = |A|/|Q^d|$ as it is with counting measure $|A|$. Suppose that $d = 2r$ is even. Find an expression for $\mathbf{P}((B_r)_\epsilon)$. How does it behave as d and ϵ grow?

1.3 The Hausdorff metric

Suppose that (X, d) is a compact metric space. Let $K^*(X)$ denote the set of closed non-empty subsets of X. We set

$$\rho(K, L) = \inf\{\epsilon > 0 : K \subseteq L_\epsilon, L \subseteq K_\epsilon\}.$$

Since $K_{\epsilon+\eta} \supseteq (K_\epsilon)_\eta$, it follows easily that ρ is a metric on $K^*(X)$. Note that the set of one-point subsets of X is a closed subset of $K^*(X)$, naturally isometric to X.

Proposition 1.3.1 *The finite sets are dense in $K^*(X)$, and $K^*(X)$ is precompact.*

Proof An ϵ-net in a metric space is a subset N such that $X \subseteq \cup\{B_\epsilon(n) : n \in N\}$. A set is precompact if and only if for each $\epsilon > 0$ there exists a finite ϵ-net.

Suppose that $\epsilon > 0$. Let N be a finite ϵ-net in X. If $K \in K^*(X)$, let $M = N \cap K_\epsilon$. Then M is a non-empty subset of K_ϵ. If $x \in K$ then there exists $y \in N$ with $d(x, y) \leq \epsilon$. Then $y \in M$ and $x \in M_\epsilon$. Thus $X \subseteq M_\epsilon$, and $\rho(X, M) \leq \epsilon$. The set $P^*(N)$ of non-empty subsets of N is therefore a finite ϵ-net in $K^*(X)$. □

Proposition 1.3.2 *$(K^*(X), \rho)$ is complete.*

Proof First we show that if (K^n) is a decreasing sequence in $K^*(X)$ then $K^n \to K = \cap_n K^n$ as $n \to \infty$. Suppose that $\epsilon > 0$. Certainly $K \subseteq K^n \subseteq K^n_\epsilon$ for all n. We claim that there exists n_0 such that $K^n \subseteq K^{n_0} \subseteq K_\epsilon$ for $n \geq n_0$. If not, for each n there exists $x_n \in K^n$ with $d(x_n, K) > \epsilon$. Since X is compact there exists a convergent sequence (x_{n_k}), convergent to x, say. Then $x \in K$, but $d(x, K) \geq \epsilon$, giving a contradiction.

Secondly, suppose that (K^n) is a Cauchy sequence, and suppose that $\epsilon > 0$. Let $L^n = \overline{\cup_{m \geq n} K^n}$. Then $K^n \subseteq L^n \subseteq L^n_\epsilon$ for all n. There exists n_0 such that $\rho(K^m, K^n) \leq \epsilon$ for $m \geq n \geq n_0$. Thus if $m \geq n \geq n_0$ then $K^m \subseteq K^n_\epsilon$, and so $L^n \subseteq K^n_\epsilon$. Thus $\rho(K^n, L^n) \leq \epsilon$ for $n \geq n_0$. But (L^n) decreases to $L = \cap_n L^n$, and so $L^n \to L$. Consequently, $K^n \to L$. □

Corollary 1.3.1 $(K^*(X), \rho)$ *is compact.*

As examples, take X to be the unit ball in l_2^d. (In fact, any finite-dimensional normed space will do.) Then the set of closed non-empty compact subsets of X is closed in $K^*(X)$. So also is the set of sets of the form $E \cap X$, where E is a k-dimensional subspace of l_2^d. In this way the *Grassmann manifold* $G_{n,k}$ of k-dimensional subspaces of l_2^d becomes a compact metric space.

Exercises

(i) Verify that ρ is a metric on $K^*(X)$.
(ii) Let $X = [0, 1]$. Is the set of two-point sets closed in $K^*(X)$? What about the set of non-empty sets with at most k points?
(iii) Suppose that (X, d) is a locally compact metric space. Show that the set $K^*(X)$ of non-empty compact subsets of X is a locally compact metric space under the Hausdorff metric.
(iv) Show that the set of non-empty convex compact subsets of $\mathbf{R}^d$ is closed in $K^*(\mathbf{R}^d)$.
(v) Show that the set of sets of the form $E \cap X$, where E is a k-dimensional subspace of l_2^n, is closed in $K^*(X)$.

1.4 Haar measure

Proposition 1.4.1 *Suppose that g is an isometry of a compact metric space (X, d) into itself. Then g is surjective.*

Proof $g(X)$ is compact, and therefore closed. If $X \neq g(X)$, there exists $x \in X \setminus g(X)$. Let $\delta = d(x, g(X))$. Then $d(x, g^j x) \geq \delta$ for all j. Since g is an isometry, $d(g^k(x), g^l(x)) \geq \delta$ for all k, l with $k < l$. Thus $(g^k(x))$ has no convergent subsequence, giving a contradiction. □

Theorem 1.4.1 *Suppose that (X, d) is a compact metric space. Let I_X be the group of isometries of X onto itself. If $g, h \in I_X$, let $\rho(g, h) = \sup\{d(g(x), h(x)) : x \in X\}$. Then ρ is a translation invariant metric on I_X under which I_X is compact, and the mapping $(g, x) \to g(x) : I_X \times X \to X$ is jointly continuous.*

Proof It is easy to see that ρ is a translation invariant metric on I_X. Suppose that (g_n) is a sequence in I_X. (X, d) is separable: let (x_m) be a dense sequence in X. By a standard diagonal argument, there exists a subsequence $(h_k) = (g_{n_k})$ of (g_n) such that $g_{n_k}(x_m)$ converges as $k \to \infty$, for each m. Suppose that $\epsilon > 0$. There exists M such that $\{x_1, \ldots, x_M\}$ is an $\epsilon/3$-net in X, and there exists K such that $d(h_j(x_m), h_k(x_m)) < \epsilon/3$ for $j, k \geq K$ and $1 \leq m \leq M$. If $x \in X$, there exists $1 \leq m \leq M$ such that $d(x, x_m) \leq \epsilon/3$. Then

$$d(h_j(x), h_k(x)) \leq d(h_j(x), h_j(x_m)) + d(h_j(x_m), h_k(x_m)) + d(h_k(x_m), h_k(x)) < \epsilon.$$

Thus (h_j) converges uniformly to h, say, on X. h is an isometry of X, and $\rho(h_k, h) \to 0$ as $k \to \infty$ Thus (I_X, ρ) is sequentially compact, and so is compact.

If $\rho(g, h) < \epsilon/2$ and $d(x, y) < \epsilon/2$ then

$$d(g(x), h(y)) \leq d(g(x), g(y)) + d(g(y), h(y)) \leq d(x, y) + \rho(g, h) < \epsilon,$$

and we have joint continuity. □

Theorem 1.4.2 *Suppose that (X, d) is a compact metric space, Then there exists a probability measure μ on the Borel sets of X such that $\int f(x)\, d\mu(x) = \int f(g(x))\, d\mu(x)$ for all $f \in C(X)$, $g \in I_X$. If I_X acts transitively on X (given $x, y \in X$ there exists an isometry of X such that $g(x) = y$) then μ is unique.*

Proof $C(X)$ is a Banach space under the supremum norm, and by the Riesz representation theorem its dual can be identified with the signed Borel measures on X. $C(X)$ is separable, so that the unit ball of $M(X)$ is compact and metrizable under the weak*-topology. The set $\mathbf{P}(X) = \{\mu : \|\mu\| = 1 = \mu(X)\}$ of probability measures is weak*-closed, and so is also compact and metrizable under the weak*-topology.

Let (ϵ_k) be a decreasing sequence of positive numbers tending to 0. For each k, let N_k be an ϵ_k-net in X with a minimal number n_k of terms. For $f \in C(X)$, let $\mu_k(f) = (1/n_k) \sum_{x \in N_k} f(x)$. Then $\mu_k \in P(X)$, and there exists a weak*-convergent subsequence (which we denote again by (μ_k)) which is weak*- convergent to μ, say.

We now show that μ does not depend upon the choice of net (of minimal size). For this we need Hall's marriage theorem:

Theorem 1.4.3 *Suppose that A is a finite set, and that j is a mapping from A into the set $P(B)$ of subsets of a set B. Then there exists a marriage mapping - a one-one mapping $\phi : A \to B$ such that $\phi(a) \in j(a)$ for all $a \in A$ - if and only if whenever $C \subseteq A$ then $\#(\cup\{j(a) : a \in C\} \geq \#(C)$.*

Proof The condition is trivially necessary. Sufficiency is proved by induction on $\#(A)$. The result is trivially true if $\#(A) = 1$. Suppose that it is true for sets of cardinality less than d, and that $\#(A) = d$. There are two possibilities; first $\#(\cup\{j(a) : a \in C\}) > \#(C)$ for each non-empty proper subset C of A. Then pick $a \in A$ and $\phi(a) \in j(A)$. Let $A' = A \setminus \{a\}$, and if $a' \in A'$ let $j'(a') = j(a') \setminus \{\phi(a)\}$. Then j' satisfies the conditions, and, by the inductive hypothesis, we can define a marriage mapping $\phi' : A' \to B$. For $a' \in A'$, we define $\phi(a') = \phi'(a')$.
Secondly, there exists a non-empty proper subset C of A such that $\#(\cup\{j(a) : a \in C\} = \#(C)$. By the inductive hypothesis, we can find a marriage mapping $\psi : C \to B$. Let $D = A \setminus C$. If $d \in D$, let $j'(d) = j(d) \setminus \psi(C)$. Then it is easy to see that j' satisfies the conditions of the theorem. By the inductive hypothesis, we can find a marriage mapping $\chi : D \to B$. If we set $\phi(a) = \psi(a)$ if $a \in C$ and $\phi(a) = \chi(a)$ if $a \in D$, then ϕ is a marriage mapping $\psi : A \to B$. □

Let's return to the proof of Theorem 1.4.2. Suppose that N'_k is another ϵ_k-net with a minimal number of elements. If $n \in N_k$, let

$$j(n) = \{n' \in N'_k : B(n, \epsilon_k) \cap B(n', \epsilon_k) \neq \emptyset\}.$$

Note that $B(n, \epsilon_k) \subseteq \cup_{n' \in j(n)} B(n', \epsilon_k)$. Suppose that $C \subseteq N_k$, and let $E = \cup\{j(n) : n \in C\}$. Then

$$\cup\{B(n', \epsilon_k) : n' \in E\} \supseteq \cup\{B(n, \epsilon_k) : n \in C\},$$

so that $(N_k \setminus C) \cup E$ is an ϵ_k-net. By minimality, $\#(E) \geq \#(C)$. Thus the conditions of the Hall marriage theorem are satisfied, and so there exists a marriage mapping $\phi : N_k \to N'_k$. Note that $d(n, \phi(n)) \leq 2\epsilon_k$. If $f \in C(X)$ then

$$\begin{aligned}
|\mu_k(f) - \mu'_k(f)| &= |\frac{1}{n_k} \sum_{n \in N_k} f(n) - f(\phi(n))| \\
&\leq \frac{1}{n_k} \sum_{n \in N_k} |f(n) - f(\phi(n))| \\
&\leq \sup\{|f(x) - f(y)| : d(x, y) \leq 2\epsilon_k\} \to 0
\end{aligned}$$

as $k \to \infty$, by the uniform continuity of f.

Suppose now that $g \in I_X$. Then, for each k, $g(N_k)$ is an ϵ_k-net with a minimal number of terms, and so

$$\int f(x)\, d\mu(x) = \lim_{k \to \infty} \left(\frac{1}{n_k} \sum_{n \in N_k} f(n) \right)$$

$$= \lim_{k\to\infty}\left(\frac{1}{n_k}\sum_{n\in N_k} f(g(n))\right)$$
$$= \int f(g(x))\,d\mu(x).$$

We now turn to uniqueness, in the case where I_X acts transitively. If $g, h, k \in I_X$ then $\rho(hg, kg) = \rho(h, k)$, and so there exists a measure ν on the Borel sets of I_X such that if $f \in C(I_X)$ then $\int f(h)\,d\nu(h) = \int f(hg)\,d\nu(h)$ for all $g \in I_X$. Suppose that $f \in C(X)$. Then

$$\begin{aligned}\int_X f(x)\,d\mu(x) &= \int_X f(h(x))\,d\mu(x) \text{ for } h \in I_X \\ &= \int_{I_X}(\int_X f(h(x))\,d\mu(x))\,d\nu(h) \\ &= \int_X(\int_{I_X} f(h(x))\,d\nu(h))\,d\mu(x).\end{aligned}$$

But if $x, y \in X$ there exists $g \in I_X$ such that $y = g(x)$. Then

$$\int_{I_X} f(h(y))\,d\nu(h) = \int_{I_X} f(h(g(x)))\,d\nu(h) = \int_{I_X} f(h(x))\,d\nu(h),$$

so that $\int_{I_X} f(h(x))\,d\nu(h)$ takes a constant value C on X. Thus $\int_X f(x)\,d\mu(x) = C$. But the same argument shows that if μ' satisfies the conclusion of the theorem then $\int_X f(x)\,d\mu'(x) = C$, and so $\mu = \mu'$. □

In the case where I_X acts transitively on X, the unique invariant measure is called *Haar measure*. The standard example is the case where (G, ρ) is a compact metric group whose topology is defined by a translation invariant metric ρ. If $f \in C(G)$, then

$$\int_G f(h)\,d\mu(h) = \int_G f(gh)\,d\mu(h) = \int_G f(hg)\,d\mu(h)$$

for all $g \in G$: these equations extend to functions in $L^1(G, \mu)$.

More generally we consider the case where the compact group (G, ρ) acts continuously as a transitive group of isometries of a compact metric space (X, d). For example, SO_n acts transitively on S^{n-1}, and also acts transitively on the Grassmann manifold $G_{n,k}$.

Proposition 1.4.2 *Suppose that a compact group (G, ρ) acts continuously as a transitive group of isometries of a compact metric space (X, d). Let μ be Haar measure on X, and let ν be Haar measure on G. If $x \in X$ and A is a Borel set in X then $\nu(\{g : g(x) \in A\}) = \mu(A)$.*

Proof

$$\mu(A) = \int_X I_A(y)\, d\mu(y) = \int_X I_A(g(y))\, d\mu(y) \text{ for } g \in G,$$

and so

$$\mu(A) = \int_G \Big(\int_X I_A(g(y))\, d\mu(y)\Big)\, d\nu(g) = \int_X \Big(\int_G I_A(g(y))\, d\nu(g)\Big)\, d\mu(y).$$

But $\int_G I_A(g(y))\, d\nu(g)$ is independent of y, by transitivity, and so

$$\mu(A) = \int_G I_A(g(x))\, d\nu(g) = \nu(\{g : g(x) \in A\}).$$

□

Similar results hold when (G, τ) is a compact topological group. Haar measure is then a regular measure on the Borel sets of G. A good account is given in the book by Zimmer listed above.

The situation is more complicated for locally compact groups. There is a measure, invariant under left translations, and unique up to scaling, and a measure, invariant under right translations, and unique up to scaling, but these need not be the same.

Exercises

(i) Let (I_X, ρ) be the group of isometries of a compact metric space (X, d), and let e be the identity map on X. Show that $\rho(gh, e) \leq \rho(g, e) + \rho(h, e)$ and that $\rho(g^{-1}, e) = \rho(g, e)$. Deduce that (I_X, ρ) is a compact topological group.

(ii) Show that dx/x is Haar measure on $(\mathbf{R}^+, \times)$.

(iii) Let $G = \mathbf{R}^+ \times \mathbf{R}^+ \times \mathbf{R}$, with composition

$$(x, y, z)(u, v, w) = (xu, yv, xw + zv).$$

Identify G with a group of upper triangular matrices. Show that the left-invariant Haar measure is $dxdydz/x^2y$ and the right-invariant Haar measure is $dxdydz/xy^2$.

1.5 Isoperimetry in S^d

We consider S^d with geodesic metric d and Haar measure μ. A *cap* in is a set of the form $B_r(x)$. Our aim is to show the following.

Theorem 1.5.1 *If A is a measurable subset of S^d and C is a cap of equal measure, then $\mu(C_\epsilon) \leq \mu(A_\epsilon)$ for each $\epsilon > 0$.*

Let $N = (0, \dots, 0, 1)$ be the north pole of S^d. If $\phi \in S^d$ and $\langle \phi, N \rangle > 0$ let

$$\begin{aligned} K_\phi^+ &= K^+ = \{x \in S^d : \langle \phi, x \rangle \geq 0\} \\ K_\phi^- &= K^- = \{x \in S^d : \langle \phi, x \rangle < 0\}. \\ E_\phi &= \{x \in S^d : \langle \phi, x \rangle = 0\}\text{: } E_\phi \text{ is the } \phi\text{-equator.} \end{aligned}$$

Let $P_\phi(x) = x - 2\langle \phi, x \rangle \phi$: P_ϕ is the reflection in the hyperplane $\{y : \langle \phi, y \rangle = 0\}$.

Suppose that A is a closed subset of S^d. Let

$$\begin{aligned} A^b &= \{x \in A : P_\phi(x) \in A\}, \\ A^+ &= \{x \in A \cap K^+ : P_\phi(x) \notin A\}, \\ A^- &= \{x \in A \cap K^- : P_\phi(x) \notin A\}, \\ A_\phi^* &= A^* = A^b \cup A^+ \cup P_\phi(A^-). \end{aligned}$$

Proposition 1.5.1 *(i) A^* is closed.*
(ii) $\mu(A) = \mu(A^*)$.
(iii) $(A^*)_\epsilon \subseteq (A_\epsilon)^*$.

Proof (i) $A^b \cup A^+$ is closed and $\overline{P_\phi(A^-)} = P_\phi(A^-) \cup (\overline{A^-} \cap E_\phi) \subseteq A^*$, so that A^* is closed.

(ii) Trivial.

(iii) Suppose that $x \in A^*$ and that $d(x, y) \leq \epsilon$. We consider cases.

(a) $x \in A$, $x \in K^+$, $y \in K^+$.
Then $y \in A_\epsilon$ so $y \in (A_\epsilon)^*$.

(b) $x \in A$, $x \in K^+$, $y \in K^-$.
Then $d(x, P_\phi(y)) \leq d(x, y) \leq \epsilon$, so that $y \in (A_\epsilon)^b \subseteq (A_\epsilon)^*$.

(c) $x \in A$, $x \in K^-$.
Then $x \in A^b$, so that $y \in (A_\epsilon)^b \subseteq (A_\epsilon)^*$.

(d) $x \notin A$.
Then $P_\phi(x) \in A^-$ and $P_\phi(y) \in A_\epsilon$. Thus $y = P_\phi(P_\phi(y)) \in (A_\epsilon)^*$. □

Corollary 1.5.1 $\mu((A^*)_\epsilon) \leq \mu(A_\epsilon)$.

We need a couple of lemmas.

Lemma 1.5.1 *The set*

$$T = \{B : \mu(B) = \mu(A) \text{ and } \mu(B_\epsilon) \leq \mu(A_\epsilon) \text{ for all } \epsilon > 0\}$$

is closed in $K^*(S^d)$.

Proof Suppose that $D \in \bar{T}$. Then given $\eta > 0$ there exists $B \in T$ with $\rho(D, B) < \eta$. Then $D_\epsilon \subseteq (B_\eta)_\epsilon \subseteq B_{\eta+\epsilon}$, so that $\mu(D_\epsilon) \leq \mu(B_{\eta+\epsilon}) \leq \mu(A_{\eta+\epsilon})$. Let $\eta \to 0$: $\mu(B_\epsilon) \leq \mu(A_\epsilon)$. Similarly, $\mu(D) \leq \mu(A)$. Further, $B \subseteq D_\eta$, so that $\mu(A) = \mu(B) \leq \mu(D_\eta)$. Letting $\eta \to 0$ we see that $\mu(A) \leq \mu(D)$. Thus $D \in T$ and T is closed. □

Let S be the smallest closed subset of $K^*(S^d)$ such that $A \in S$ and if $B \in S$ then $B^*_\phi \in S$ for all ϕ with $\langle \phi, N \rangle > 0$.

Corollary 1.5.2 *If* $B \in S$ *then* $\mu(B) = \mu(A)$ *and* $\mu(B_\epsilon) \leq \mu(A_\epsilon)$ *for all* $\epsilon > 0$.

Now let C be the cap with centre at the north pole N and with $\mu(C) = \mu(A)$.

Lemma 1.5.2 *Suppose that* $\alpha > 0$. *The set*

$$J_\alpha = \{B \in S : \mu(B \cap C) \geq \alpha\}$$

is closed in $K^*(S^d)$.

Proof Suppose that $D \in \overline{J_\alpha}$. Given $\eta > 0$ there exists $B \in J_\alpha$ with $\rho(D, B) \leq \eta$. If $x \in B \cap C$ then there exists $y \in D$ with $d(x, y) \leq \eta$, so that $y \in D \cap C_\eta$ and $x \in (D \cap C_\eta)_\eta \subseteq D_\eta \cap C_{2\eta}$. Thus $B \cap C \subseteq D_\eta \cap C_{2\eta}$, and $\mu(D_\eta \cap C_{2\eta}) \geq \alpha$. But $\cap_{\eta>0}(D_\eta \cap C_{2\eta}) = D \cap C$, and so $\mu(D \cap C) \geq \alpha$. □

Corollary 1.5.3 $\mu(B \cap C)$ *attains its maximum on* S *at a set* B_0 *in* S.

Proof Compactness. □

Proof of Theorem 1.5.1. We show that $B_0 \supseteq C$: this clearly suffices.

If not, then there exist x and $\epsilon > 0$ such that $B_\epsilon(x) \subseteq C$ and $B_\epsilon(x) \cap B_0 = \emptyset$. Then $\mu(C \setminus B_0) > 0$ and so $\mu(B_0 \setminus C) > 0$. Since $B_0 \setminus C$ is precompact, it can be covered by finitely many balls of radius $\epsilon/3$, and one of these, say $B_{\epsilon/3}(y)$, must intersect $B_0 \setminus C$ in a set of positive measure. Note that $d(x, y) \geq 2\epsilon/3$. Let $\phi = (x - y)/|x - y|$, so that $N \in K^+_\phi$ and $P_\phi(y) = x$.

Then P_ϕ moves $B_{\epsilon/3}(y)\cap(B_0\setminus C)$ into C, contradicting the maximal property of B_0. □

This proof comes from Yoav Benyamini's lecture notes, listed above.

1.6 The Beta and Gamma distributions

Suppose that $k > 0$. A non-negative random variable has a $\Gamma(k)$ distribution if it has density

$$(1/\Gamma(k))t^{k-1}e^{-t}, \text{ where } \Gamma(k) = \int_0^\infty t^{k-1}e^{-t}\,dt = (k-1)!.$$

For example if $X = Y^2/2$, where Y has an $N(0,1)$ distribution, then

$$\begin{aligned}\mathbf{P}(X \leq t) &= \frac{1}{\sqrt{2\pi}}\int_{-\sqrt{2t}}^{\sqrt{2t}} e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int_0^{\sqrt{2t}} e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{\pi}}\int_0^t s^{-1/2}e^{-s}\,ds,\end{aligned}$$

so that X has a $\Gamma(1/2)$ distribution.

Let γ denote canonical Gaussian measure on $\mathbf{R}$ (with density $e^{-x^2/2}/\sqrt{2\pi}$). Making the substitution $t = s^2/2$, we see that

$$\Gamma_k = \int_0^\infty \frac{s^{2k-1}}{2^{k-1}}e^{-s^2/2}\,ds = \frac{\sqrt{2\pi}}{2^k}\mathbf{E}_\gamma(|s|^{2k-1}) : \quad \int_0^\infty s^{2k-1}e^{-s^2/2}\,ds = 2^{k-1}\Gamma_k.$$

Suppose that $m, n > 0$. A random variable taking values in $[0,1]$ has a $B(m,n)$ distribution if it has density

$$\frac{t^{m-1}(1-t)^{n-1}}{B(m,n)}, \text{ where } B(m,n) = \int_0^1 t^{m-1}(1-t)^{n-1}\,dt.$$

Theorem 1.6.1 *If X has a $\Gamma(m)$ distribution and Y has a $\Gamma(n)$ distribution, and if X and Y are independent, then $U = X+Y$ and $V = X/(X+Y)$ are independent; U has a $\Gamma(m+n)$ distribution, and V has a $B(m,n)$ distribution.*

Proof (X,Y) has density $s^{m-1}t^{n-1}e^{-(s+t)}/\Gamma(m)\Gamma(n)$. If $u = s+t$ and $v = s/(s+t)$ then $s = uv$ and $t = u(1-v)$. The Jacobian is $-u$, so that (U,V) has density

$$\begin{aligned}&\frac{u^{m+n-1}v^{m-1}(1-v)^{n-1}e^{-u}}{\Gamma(m)\Gamma(n)} = \\ &\quad = \frac{\Gamma(m+n)B(m,n)}{\Gamma(m)\Gamma(n)}.\frac{u^{m+n-1}e^{-u}}{\Gamma(m+n)}.\frac{v^{m-1}(1-v)^{n-1}}{B(m,n)}.\end{aligned}$$

□

Corollary 1.6.1 $\Gamma(m+n)B(m,n) = \Gamma(m)\Gamma(n)$.

Corollary 1.6.2 *If $X_1, \ldots, X_n$ are independent $N(0,1)$ random variables then $\frac{1}{2}(X_1^2 + \cdots + X_n^2)$ has a $\Gamma(n/2)$ distribution.*

Corollary 1.6.3 *If $X_1, \ldots, X_N$ are independent $N(0,1)$ random variables and $1 \leq n \leq N$ then $(X_1^2+\cdots+X_n^2)/(X_1^2+\cdots+X_N^2)$ has a $B(n/2, (N-n)/2)$ distribution.*

The Gamma function has its place in Gaussian calculus, when we use spherical polar co-ordinates. Let γ_d denote canonical Gaussian measure on $\mathbf{R}^d$ (with density $(2\pi)^{-d/2}e^{-|x|^2/2}$). Let μ_{d-1} be rotation-invariant Haar measure on S^{d-1}, and let A_{d-1} be the $d-1$-dimensional volume of S^{d-1}, so that $\sigma_{d-1} = A_{d-1}\mu_{d-1}$ is $d-1$-dimensional volume measure on S^{d-1}. Then using spherical polar co-ordinates,

$$\begin{aligned}
(2\pi)^{d/2} &= \int_{\mathbf{R}^d} e^{-|x|^2/2}\,dx \\
&= A_{d-1}\int_0^\infty u^{d-1}e^{-u^2/2}\,du = 2^{d/2-1}A_{d-1}\Gamma(d/2).
\end{aligned}$$

Thus $A_{d-1} = 2\pi^{d/2}/\Gamma(d/2)$. Further,

$$V_d = \text{ vol }(B_1(l_2^d)) = A_{d-1}/d = \pi^{d/2}/\Gamma(d/2+1).$$

If $f \in L^1(\gamma_d)$ then

$$\begin{aligned}
\int_{\mathbf{R}^d} f\,d\gamma_d &= \frac{1}{(2\pi)^{d/2}}\int_{\mathbf{R}^d} f(x)e^{-|x|^2/2}\,dx \\
&= \frac{A_{d-1}}{(2\pi)^{d/2}}\int_0^\infty \left(\int_{S^{d-1}} f(u\theta)u^{d-1}e^{-u^2/2}\,d\mu_{d-1}(\theta)\right)du \\
&= \frac{2}{2^{d/2}\Gamma(d/2)}\int_0^\infty \left(\int_{S^{d-1}} f(u\theta)\,d\mu_{d-1}(\theta)\right)u^{d-1}e^{-u^2/2}\,du \\
&= \frac{2}{2^{d/2}\Gamma(d/2)}\int_{S^{d-1}} \left(\int_0^\infty f(u\theta)u^{d-1}e^{-u^2/2}\,du\right)d\mu_{d-1}(\theta)
\end{aligned}$$

Exercises

(i) Let $I_n = \int_0^\pi sin^n\theta\,d\theta$. Show that if $n = 2k$ then

$$I_n = \frac{(2k)!\pi}{2^{2k}(k!)^2} \sim \sqrt{\frac{2\pi}{n}}.$$

Find a similar formula for I_n when n is odd.

(ii) Show that $B(n/2, 1/2) = I_{n-1}$.

(iii) Show that $\Gamma(1/2) = \sqrt{\pi}$.

(iv) Show that $\Gamma(n+1/2)/\Gamma(n/2) = \Gamma(1/2)/I_{n-1} \sim \sqrt{n/2}$.

(v) Show that if $\alpha > 0$ then

$$\int_{\mathbf{R}^d} |x|^\alpha \, d\gamma_d = 2^{\alpha/2}\Gamma((d+\alpha)/2)/\Gamma(d/2).$$

Show that if $2k$ is an even integer then

$$\int_{\mathbf{R}^d} |x|^{2k} \, d\gamma_d = (d+k-1)(d+k-2)\dots d.$$

Show that

$$\int_{\mathbf{R}^d} |x| \, d\gamma_d \sim \sqrt{d} \text{ as } d \to \infty.$$

(vi) Suppose that $\|.\|$ is a semi-norm on $\mathbf{R}^d$. Show that

$$\mathbf{E}_{\gamma_d}(\|x\|) = \frac{\sqrt{2}\pi}{I_{d-1}} \int_{S^{d-1}} \|x\| \; d\mu_{d-1}(x).$$

[Check the constants that occur in these identities!]

1.7 Poincaré's lemma

We want to approximate canonical Gaussian measure γ_d (with density $(2\pi)^{-d/2}e^{-|x|^2/2}$). One standard way is given by the central limit theorem, but there is another way.

Suppose that $N > d$. Let T_N be the sphere with centre 0 and radius $\sqrt{N}$ in $\mathbf{R}^{N+1}$, equipped with rotation-invariant Haar measure τ_N. Let P_{N+1} be the orthogonal projection of $\mathbf{R}^{N+1}$ onto $\mathbf{R}^d$, and let π_N be its restriction to T_N.

Theorem 1.7.1 (Poincaré's Lemma) *If A is a Borel subset of $\mathbf{R}^d$ then $\tau_N(\pi_N^{-1}(A)) \to \gamma_d(A)$ as $N \to \infty$.*

Proof Let (g_i) be an independent sequence of $N(0,1)$ random variables. Let $R_{N+1} = (g_1^2 + \cdots + g_{N+1}^2)^{1/2}$. Then

$$V_N = \frac{\sqrt{N}}{R_{N+1}}(g_1, \dots, g_{N+1})$$

has distribution τ_N, so that if A is a Borel subset of $\mathbf{R}^d$ then

$$\begin{aligned}\tau_N(\pi_N^{-1}(A)) &= \mathbf{P}(V_N \in \pi_N^{-1}(A)) \\ &= \mathbf{P}(\frac{\sqrt{N}}{R_{N+1}}(g_1,\dots,g_d) \in A) \\ &= \mathbf{P}\left(\left(\frac{NR_d^2}{R_{N+1}^2}\right)^{1/2}\left(\frac{(g_1,\dots,g_d)}{R_d}\right) \in A\right).\end{aligned}$$

Now the random variables $R_{N+1}^2 - R_d^2$, R_d^2 and $(g_1,\dots,g_d)/R_d$ are independent, so that R_d^2/R_{N+1}^2 and $(g_1,\dots,g_d)/R_d$ are independent. The first has distribution $B(d/2,(N+1-d)/2^)$ and the second is uniformly distributed over the sphere S^{d-1}. Thus if μ_{d-1} is rotation-invariant Haar measure on S^{d-1} and $B = B(d/2,(N+1-d)/2)$, making the substitution $u = \sqrt{Nt}$,

$$\begin{aligned}&\tau_N(\pi_N^{-1}(A)) = \\ &= \int_{S^{d-1}}\left(\frac{1}{B}\int_0^1 I_A(\sqrt{Nt}\theta)t^{d/2-1}(1-t)^{(N-d-1)/2}\,dt\right)d\mu_{d-1}(\theta) \\ &= \int_{S^{d-1}}\left(\frac{2}{N^{d/2}B}\int_0^{\sqrt{N}} I_A(u\theta)u^{d-1}(1-\frac{u^2}{N})^{(N-d-1)/2}\,du\right)d\mu_{d-1}(\theta).\end{aligned}$$

Now as $N \to \infty$,

$$\begin{aligned}\frac{2}{N^{d/2}B(d/2,(N+1-d)/2)} &= \frac{2\Gamma((N+1/2)}{\Gamma(d/2)\Gamma((N+1-d)/2)N^{d/2}} \\ &\to \frac{2}{\Gamma(d/2)2^{d/2}},\end{aligned}$$

and

$$(1-u^2/N)^{(N-d-1)/2} \to e^{-u^2/2}.$$

Further, $(1-u^2/N)^{(N-d-1)/2} \le e^{-u^2/4}$, for $N > 2d+2$.

We can therefore apply the theorem of dominated convergence:

$$\begin{aligned}\tau_N(\pi_N^{-1}(A)) &\to \frac{2}{\Gamma(d/2)2^{d/2}}\int_{S^{d-1}}\left(\int_0^\infty I_A(u\theta)u^{d-1}e^{-u^2/2}\,du\right)d\mu_{d-1}(\theta) \\ &= \int_{\mathbf{R}^d} I_A\,d\gamma_d,\end{aligned}$$

as $N \to \infty$. $\square$

1.8 Gaussian isoperimetry

Let γ_d denote canonical Gaussian measure on $\mathbf{R}^d$ (with density $(2\pi)^{-d/2}e^{-|x|^2/2}$). Let H_s be the half-space $\{x : x_d \leq s\}$. Note that

$$\gamma_d(H_s) = \Phi(s) = (2\pi)^{-d/2} \int_{-\infty}^{s} e^{-t^2/2}\, dt.$$

Theorem 1.8.1 *Suppose that A is a measurable subset of $\mathbf{R}^d$ and that $\gamma_d(A) = \Phi(s) = \gamma_d(H_s)$. Then $\gamma_d(A_\epsilon) \geq \Phi(s+\epsilon) = \gamma_d((H_s)_\epsilon)$ for each $\epsilon > 0$.*

Thus half-spaces solve the isoperimetric problem.

Proof Suppose that $r < s$. Then there exists N_0 such that $\tau_N(\pi_N^{-1}(A)) > \frac{1}{2}(\Phi(s)+\Phi(r))$ and $\tau_N(\pi_N^{-1}(H_r)) < \frac{1}{2}(\Phi(s)+\Phi(r))$ for $N \geq N_0$. Now π_N is distance-decreasing, so that $\pi_N^{-1}(A_\epsilon) \supseteq (\pi_N^{-1}(A))_\epsilon$. Also, $\pi_N^{-1}(H_r)$ is a cap in S^N with measure less than $\tau_N(\pi_N^{-1}(A))$. Thus

$$\tau_N(\pi_N^{-1}(A_\epsilon)) \geq \tau_N((\pi_N^{-1}(A))_\epsilon) \geq \tau_N((\pi_N^{-1}(H_r))_\epsilon).$$

Now $(\pi_N^{-1}(H_r))_\epsilon$ is a cap in S^N of the form $\pi_N^{-1}(H_{r+\eta_N})$. Let $r/\sqrt{N} = \cos\theta_N$, and let $\phi_N = \theta_N - \epsilon/\sqrt{N}$. Then

$$\begin{aligned} r + \eta_N &= \sqrt{N}\cos\phi_N \\ &= \sqrt{N}(\cos\theta_N \cos(\epsilon/\sqrt{N}) + \sin\theta_N \sin(\epsilon/\sqrt{N})) \\ &= r\cos(\epsilon/\sqrt{N}) + \sqrt{N}\sin\theta_N \sin(\epsilon/\sqrt{N}) \to r + \epsilon \end{aligned}$$

as $N \to \infty$, since $\theta_N \to \pi/2$ as $N \to \infty$. Thus if $0 < \epsilon' < \epsilon$ there exists $N_1 \geq N_0$ such that $r + \eta_N > r + \epsilon'$, so that $\tau_N(\pi_N^{-1}(A)_\epsilon) \geq \tau_N(\pi_N^{-1}(H_{r+\epsilon'}))$. Finally,

$$\begin{aligned} \gamma_d(A_\epsilon) &= \lim_{N\to\infty} \tau_N(\pi_N^{-1}(A_\epsilon)) \geq \lim_{N\to\infty} \tau_N(\pi_N^{-1}(A)_\epsilon) \\ &\geq \lim_{N\to\infty} \tau_N(\pi_N^{-1}(H_r + \epsilon')) = \Phi(r+\epsilon'). \end{aligned}$$

Since this holds for all $r < s$ and $0 < \epsilon' < \epsilon$, the result follows. □

1.9 Some function spaces

Let us introduce some function spaces that we shall work with. Suppose that (X, d) is a metric space.

The space $C(X)$ is the vector space of all continuous real-valued functions on X, and $C_b(X)$ is the space of all bounded continuous functions on X.

$C_b(X)$ is a Banach space under the norm $\|f\|_\infty = \sup\{|f(x)| : x \in X\}$. If X is compact, then $C(X) = C_b(X)$.

The space Lip (X) is the space of all real-valued Lipschitz functions on X: that is, functions f for which there is a constant L such that $|f(x) - f(y)| \leq Ld(x, y)$ for all $x, y \in X$. Lip (X) is a lattice under the natural ordering, contains the constants, and separates points. Thus if X is compact, then Lip (X) is dense in $C(X)$. The quantity

$$\|f\|_L = \sup\{\frac{|f(x) - f(y)|}{d(x,y)} : x \neq y\}$$

is a seminorm on Lip (X). [$\|f\|_L = 0$ if and only if f is constant.] A 1-Lipschitz function is a function for which $\|f\|_L \leq 1$.

Theorem 1.9.1 *Suppose that g is a Lipschitz function on a subset A of X. Then g can be extended to a Lipschitz function f on X with $\|f\|_L = \|g\|_L$.*

Proof By Zorn's lemma, there is a maximal extension h of g with $\|h\|_L = \|g\|_L = L$, say, to $B \subseteq X$. We must show that $B = X$. Suppose that $x \notin B$. If $b, c \in B$ then

$$h(b) - h(c) \leq Ld(b, c) \leq Ld(b, x) + Ld(x, c),$$

so that $h(b) - Ld(b, x) \leq h(c) + Ld(x, c)$. Let $h(x) = \sup\{h(b) - Ld(b, x) : b \in B\}$. Then $h(b) - h(x) \leq Ld(b, x)$ for all $b \in B$. Further, $h(x) - h(c) \leq Ld(x, c)$ for all $c \in B$, and so the extension to $B \cup \{x\}$ is Lipschitz, without increase of seminorm. □

The space $BL(X)$ is the space of bounded real-valued Lipschitz functions on X. It is a Banach space under the norm $\|f\|_{BL} = \|f\|_\infty + \|f\|_L$. $BL(f)$ is a lattice, and a Banach algebra under pointwise multiplication.

Theorem 1.9.2 *Suppose that g is a bounded Lipschitz function on a subset A of X. Then g can be extended to a bounded Lipschitz function f on X with $\|f\|_{BL} = \|g\|_{BL}$.*

Proof By Theorem 1.9.1, g can be extended without increase of Lipschitz norm to h on X. Let $f = (h \wedge \|g\|_\infty) \vee (-\|g\|_\infty)$. □

Exercises

(i) Show that $BL(X)$ is a Banach space, a lattice and a Banach algebra under the norm $\|f\|_{BL} = \|f\|_L + \|f\|_\infty$.

(ii) Let $\mathrm{Lip}_0(\mathbf{R}) = \{f \in \mathrm{Lip}\,(\mathbf{R}) \cap L^1(\mathbf{R}) : \int_{-\infty}^{\infty} f\, d\lambda = 0\}$. Show that $\|.\|_L$ is a norm on $\mathrm{Lip}_0(\mathbf{R})$, and that $\mathrm{Lip}_0(\mathbf{R})$ is not complete under this norm. Let $\|f\|_{L_0} = \|f\|_L + |f(0)|$. Is $\mathrm{Lip}_0(\mathbf{R})$ complete under this norm? Is $\mathrm{Lip}_0(\mathbf{R})$ a Banach algebra (under pointwise multiplication)?

1.10 Isoperimetry and concentration of measure

The isoperimetric inequalities that we have established allow us to obtain accurate estimates of the concentration of measure. Suppose that μ is a probability measure on a compact metric space (X, d). We consider how $\mu(C(A_\epsilon))$ decays for sets A of measure at least 1/2. We define the *concentration function* as

$$\alpha_\mu(\epsilon) = \sup\{\mu(C(A_\epsilon)) : \mu(A) \geq 1/2\}.$$

This is a sensible thing to do: if f is a 1-Lipschitz function on X with constant 1 then m is a *median*, or *Lévy mean* if $\mu(f \leq m) \geq 1/2$ and $\mu(f \geq m) \geq 1/2$. Then $\mu(f > m + \epsilon) \leq \alpha_\mu(\epsilon)$ and $\mu(f < m - \epsilon) \leq \alpha_\mu(\epsilon)$, so that $\mu(|f - m| > \epsilon) \leq 2\alpha_\mu(\epsilon)$.

Proposition 1.10.1 $\alpha_{\gamma_d}(\epsilon) = 1 - \Phi(\epsilon) \leq e^{-\epsilon^2/2}/\epsilon\sqrt{2\pi}$.

Proof The first equation is an immediate consequence of Gaussian isoperimetry. As for the inequality:

$$1 - \Phi(\epsilon) = \frac{1}{\sqrt{2\pi}} \int_\epsilon^\infty e^{-x^2/2}\, dx \leq \frac{1}{\epsilon\sqrt{2\pi}} \int_\epsilon^\infty x e^{-x^2/2}\, dx = e^{-\epsilon^2/2}/\epsilon\sqrt{2\pi}.$$

□

The important feature here is that the concentration function does not depend on the dimension d.

Proposition 1.10.2 *Suppose that μ is Haar measure on S^d. Then*

$$\alpha_\mu(\theta) = I_{d-1}(\theta)/2I_{d-1}(0), \text{ where } I_d(\theta) = \int_\theta^{\pi/2} \cos^d t\, dt.$$

Further, $\alpha_\mu(\theta) \leq (\sqrt{\pi/8})e^{-(d-1)\theta^2/2}$.

Proof Let A_d denote the d-dimensional measure of S^d. Then

$$A_d = \int_{-\pi/2}^{\pi/2} A_{d-1} \cos^{d-1}\theta\, d\theta = 2I_{d-1}(0)A_{d-1},$$

and in the same way if H is a hemisphere then the d-dimensional measure of $C(H_\theta)$ is $I_{d-1}(\theta)A_{d-1}$. Thus isoperimetry on S^d gives the formula.

Now if $0 < t < \pi/2$ then

$$\cos t \leq 1 - \frac{t^2}{2} + \frac{t^4}{24} \leq 1 - \frac{t^2}{2} + \frac{t^4}{8} - \frac{t^6}{48} \leq e^{-t^2/2},$$

so that

$$\begin{aligned} I_d(\theta) &\leq \int_\theta^{\pi/2} e^{-dt^2/2}\,dt \\ &= \int_0^{\pi/2-\theta} e^{-d(t+\theta)^2/2}\,dt \\ &\leq e^{-d\theta^2/2}\int_0^\infty e^{-dt^2/2}\,dt = \sqrt{\frac{\pi}{2d}}e^{-d\theta^2/2}. \end{aligned}$$

Integrating by parts, $dI_d(0) = (d-1)I_{d-2}(0)$. Since $(d-1)/d \geq \sqrt{(d-2)/d}$, $\sqrt{d}I_d(0) \geq \sqrt{d-2}I_{d-2}(0)$. Thus $\sqrt{d}I_d(0) \geq \min(I_1(0), \sqrt{2}I_2(0)) = 1$. Thus

$$\alpha_\mu(\theta) \leq \sqrt{\frac{\pi}{2(d-1)}}e^{-(d-1)\theta^2/2}/2I_{d-1}(0) \leq \sqrt{\frac{\pi}{8}}e^{-(d-1)\theta^2/2}.$$

□

Here the essential feature is that, as $d \to \infty$, the total mass is concentrated closer and closer to the equator.

1.11 Sub-Gaussian random variables

Suppose that $(\Omega, \Sigma, \mathbf{P})$ is a probability space. We define $L_{\exp^2}$ to be the space of measurable functions f for which there exists an $\alpha > 0$ such that $\mathbf{E}(e^{\alpha|f|^2}) \leq 2$. This is a vector space, which becomes a Banach space (an Orlicz function space) when we take $\{f : \mathbf{E}(e^{|f|^2}) \leq 2\}$ as its unit ball. We denote the norm by $\|.\|_{\exp^2}$.

If X is a random variable with a Gaussian distribution with mean 0 and variance $\mathbf{E}(X^2) = \sigma^2$, its moment generating function $E(e^{tX})$ is $e^{\sigma^2 t^2/2}$. This led Kahane to make the following definition. A random variable X is *sub-Gaussian, with exponent* b, if $\mathbf{E}(e^{tX}) \leq e^{b^2t^2/2}$ for $-\infty < t < \infty$.

Theorem 1.11.1 *Suppose that X is a sub-Gaussian random variable with exponent b. Then*

(i) $P(X > R) \leq e^{-R^2/2b^2}$ and $P(X < -R) \leq e^{-R^2/2b^2}$ for each $R > 0$;

(ii) $X \in L_{\exp^2}$ and $\|X\|_{\exp^2} \leq 2b$;

(iii) X is integrable, $\mathbf{E}(X) = 0$, and $\mathbf{E}(X^{2k}) \leq 2^{k+1}k!b^{2k}$ for each positive integer k.

Conversely if X is a real random variable which satisfies (iii) then X is sub-Gaussian with exponent $2b$.

Proof (i) By Markov's inequality, if $t > 0$ then

$$e^{tR}\mathbf{P}(X > R) \leq \mathbf{E}(e^{tX}) \leq e^{b^2t^2/2}.$$

Setting $t = R/b^2$, we see that $\mathbf{P}(X > R) \leq e^{-R^2/2b^2}$. Since $-X$ is also sub-Gaussian with exponent b, $\mathbf{P}(X < -R) \leq e^{-R^2/2b^2}$ as well.

(ii)

$$\begin{aligned} \mathbf{E}(e^{X^2/4b^2}) &= \frac{1}{2b^2}\int_0^\infty te^{t^2/4b^2}\mathbf{P}(|X| > t)\,dt \\ &\leq \frac{1}{b^2}\int_0^\infty te^{-t^2/4b^2}\,dt = 2. \end{aligned}$$

(iii) Since $X \in L_{\exp^2}$, X is integrable. Since $tx \leq e^{tx} - 1$, $t\mathbf{E}(X) \leq e^{b^2t^2/2} - 1$, from which it follows that $\mathbf{E}(X) \leq 0$. Since $-X$ is also sub-Gaussian, $\mathbf{E}(X) \geq 0$ as well. Thus $\mathbf{E}(X) = 0$. Further,

$$\begin{aligned} E(X^{2k}) &= 2k\int_0^\infty t^{2k-1}\mathbf{P}(|X| > t)\,dt \\ &\leq 2.2k\int_0^\infty t^{2k-1}e^{-t^2/2b^2}\,dt \\ &= (2b^2)^k 2k\int_0^\infty s^{k-1}e^{-s}\,ds = 2^{k+1}k!b^{2k}. \end{aligned}$$

Note that $\|X\|_{2k} \leq b\sqrt{2k}$ for $k \geq 2$.

Finally, suppose that X is a real random variable which satisfies (iii). If $y > 0$ and $k \geq 1$ then

$$\frac{y^{2k+1}}{(2k+1)!} \leq \frac{y^{2k}}{(2k)!} + \frac{y^{2k+2}}{(2k+2)!},$$

so that

$$\begin{aligned} \mathbf{E}(e^{tX}) &\leq 1 + \sum_{n=2}^\infty \mathbf{E}(\frac{|tX|^n}{n!}) \leq 1 + 2\sum_{k=1}^\infty \mathbf{E}(\frac{|tX|^{2k}}{(2k)!}) \\ &\leq 1 + 4\sum_{k=1}^\infty \frac{k!(2b^2t^2)^k}{(2k)!} \leq 1 + \sum_{k=1}^\infty \frac{(4b^2t^2)^k}{k!} = e^{4b^2t^2}, \end{aligned}$$

since $2(k!)^2 \leq (2k)!$. □

Note that this theorem shows that if X is a bounded random variable with zero expectation then X is sub-Gaussian.

If $X_1, \dots, X_N$ are independent sub-Gaussian random variables with exponents $b_1, \dots, b_N$ respectively, and $a_1, \dots, a_N$ are real numbers, then

$$\mathbf{E}(e^{t(a_1X_1+\cdots+a_NX_N)}) = \prod_{n=1}^{N} \mathbf{E}(e^{ta_nX_n}) \le \prod_{n=1}^{N} e^{a_n^2 b_n^2/2},$$

so that $a_1X_1 + \cdots + a_NX_N$ is sub-Gaussian, with exponent $(a_1^2b_1^2 + \cdots + a_N^2b_N^2)^{\frac{1}{2}}$.

Suppose that ϵ is a Bernoulli random variable: $\mathbf{P}(\epsilon = 1 = \mathbf{P}(\epsilon = -1) = 1/2$. Then $\mathbf{E}(e^{\lambda\epsilon}) = \cosh\lambda \le e^{\lambda^2/2}$, so that ϵ is sub-Gaussian with index 1.

Proposition 1.11.1 *Let d be the Hamming metric on $\mathbf{Q}^d$, and let μ be Haar measure. Then $\alpha_\mu(\epsilon) \le e^{-2\epsilon^2/d}$.*

Proof Let $s_d = \sum_{j=1}^d \epsilon_j$. Then s_d is sub-Gaussian, with exponent $\sqrt{d}$. By isoperimetry,

$$\alpha_\mu(\epsilon) = \mu(l > d/2 + \epsilon) = \mu(s_d > 2\epsilon) \le e^{-(2\epsilon)^2/2d} = e^{-2\epsilon^2/d}.$$

$\square$

1.12 Khintchine's inequality

Suppose that (ϵ_i) is a sequence of Bernoulli random variables and that $a \in l_2^d$, and that $\|a\|_2 = \sigma$. Then $S_d = \sum_{i=1}^n a_i\epsilon_i$ is sub-Gaussian with index σ. We have the following:

Theorem 1.12.1 (Khintchine's inequality) *Suppose that $(\epsilon_1, \dots, \epsilon_d)$ are Bernoulli random variables and that $a = (a_1, \dots, a_d) \in l_2^d$. Let $S_d = \sum_{i=1}^d a_i\epsilon_i$. Then $\|S_d\|_{\exp^2} \le 2\sigma$, and there exist universal constants C_p, with $C_p \sim \sqrt{p}$ such that*

$$\sigma = \|S_d\|_2 \le \|S_d\|_p \le C_p\sigma.$$

For $0 < p < 2$ there exist universal constants A_p such that

$$A_p\sigma \le (\mathbf{E}|S_d|^p)^{1/p} \le \sigma.$$

Proof If $2k-2 \le p \le 2k$, then $\sigma \le \|S_d\|_p \le \|S_d\|_{2k}$, so that by Theorem 1.11.1 we can take $C_p = C_{2k} = (2^{k+1}k!)^{1/2k}$, and $C_p \sim \sqrt{p}$, by Stirling's formula. In fact we can do a bit better:

$$\begin{aligned}
\left\|\sum_{i=1}^{d}\epsilon_i a_i\right\|_{2k}^{2k} &= \mathbf{E}(\sum_{i=1}^{d}\epsilon_i a_i)^{2k} \\
&= \sum_{j_1+\cdots+j_d=2k}\frac{(2k)!}{j_1!\dots j_d!}a_1^{j_1}\dots a_d^{j_d}\mathbf{E}(\epsilon_1^{j_1}\dots\epsilon_d^{j_d}) \\
&= \sum_{j_1+\cdots+j_d=2k}\frac{(2k)!}{j_1!\dots j_d!}a_1^{j_1}\dots a_d^{j_d}\mathbf{E}(\epsilon_1^{j_1})\dots\mathbf{E}(\epsilon_d^{j_d}),
\end{aligned}$$

by independence. Now $\mathbf{E}(\epsilon_n^{j_n}) = \mathbf{E}(1) = 1$ if j_n is even, and $\mathbf{E}(\epsilon_n^{j_n}) = \mathbf{E}(\epsilon_n) = 0$ if j_n is odd. Thus many of the terms in the sum are 0, and

$$\left\|\sum_{j=1}^{d}\epsilon_j a_j\right\|_{2k}^{2k} = \sum_{k_1+\cdots+k_d=k}\frac{(2k)!}{(2k_1)!\dots(2k_d!)}a_1^{2k_1}\dots a_d^{2k_d}.$$

But $(2k_1)!\dots(2k_n)! \geq 2^{k_1}k_1!\dots 2^{k_d}k_d! = 2^k k_1!\dots k_d!$, and so

$$\begin{aligned}
\left\|\sum_{j=1}^{d}\epsilon_j a_j\right\|_{2k}^{2k} &\leq \frac{(2k)!}{2^k k!}\sum_{k_1+\cdots+k_d=k}\frac{k!}{(k_1)!\dots(k_d!)}a_1^{2k_1}\dots a_d^{2k_d} \\
&= \frac{(2k)!}{2^k k!}\sigma^{2k}
\end{aligned}$$

Thus we can take $C_{2k} = ((2k)!/2^k k!)^{1/2k}$. In particular, we can take $C_4 = 3^{1/4}$.

For the second part, we need Littlewood's inequality:

Proposition 1.12.1 (Littlewood's inequality) *Suppose that $0 < p_0 < p_1 < \infty$ and that $0 < \theta < 1$. Define p by $1/p = (1-\theta)/p_0 + \theta/p_1$. If $f \in L^{p_0} \cap L^{p_1}$ then $f \in L^p$ and $\|f\|_p \leq \|f\|_{p_0}^{1-\theta}\|f\|_{p_1}^{\theta}$.*

Proof Let $1-\gamma = (1-\theta)p/p_0$, so that $\gamma = \theta p/p_1$. We apply Hölder's inequality with exponents $1/(1-\gamma)$ and $1/\gamma$:

$$\begin{aligned}
\|f\|_p &= \left(\int |f|^p\,d\mu\right)^{1/p} = \left(\int |f|^{(1-\theta)p}|f|^{\theta p}\,d\mu\right)^{1/p} \\
&\leq \left(\int |f|^{(1-\theta)p/(1-\gamma)}\,d\mu\right)^{(1-\gamma)/p}\left(\int |f|^{\theta p/\gamma}\,d\mu\right)^{\gamma/p} \\
&= \left(\int |f|^{p_0}\,d\mu\right)^{(1-\theta)/p_0}\left(\int |f|^{p_1}\,d\mu\right)^{\theta/p_1} = \|f\|_{p_0}^{1-\theta}\|f\|_{p_1}^{\theta}.
\end{aligned}$$

□

Thus if we choose θ such that $1/2 = (1-\theta)/p + \theta/4$ then

$$\|S_d\|_2 \leq \|S_d\|_p^{1-\theta} \|S_d\|_4^{\theta} \leq C_4^{\theta} \|S_d\|_p^{1-\theta} \|S_d\|_2^{\theta},$$

so that $\|S_d\|_2 \leq C_4^{\theta/(1-\theta)} \|S_d\|_p = 3^{1/p-1/2} \|S_d\|_p$. □

This theorem gives $A_1 = \sqrt{3}$. When Littlewood proved the second part of Khintchine's inequality, he made a mistake, and obtained $A_1 = \sqrt{2}$. For many years it was an open problem to obtain this constant: this was done by Szarek, and by Haagerup. Finally, a beautiful proof was given, that also works for vector-valued sums.

Theorem 1.12.2 (Latala-Oleszkiewicz) *Let $S_d = \sum_{i=1}^d \epsilon_i a_i$, where $\epsilon_1, \ldots, \epsilon_d$ are Bernoulli random variables and $a_1, \ldots, a_d$ are vectors in a normed space E. Then $\|S_d\|_{L^2(E)} \leq \sqrt{2} \|S_d\|_{L^1(E)}$.*

Proof Take $\Omega = D_2^d$, where D_2 is the multiplicative group $\{1, -1\}$, and $\epsilon_i(\omega) = \omega_i$. If $A \subseteq \{1, \ldots, d\}$, let the Walsh function $w_A = \prod_{i \in A} \epsilon_i$. The Walsh functions form an orthonormal basis for $L^2(D_2^d)$, so that if $f \in C_{\mathbf{R}}(D_2^d)$ then

$$f = \sum_A \hat{f}_A w_A = \mathbf{E}(f) + \sum_{i=1}^d \hat{f}_i \epsilon_i + \sum_{|A|>1} \hat{f}_A w_A,$$

and $\|f\|_2^2 = \langle f, f \rangle = \sum_A \hat{f}_A^2$.

We now consider a graph with vertices the elements of D_2^d and edges the set of pairs

$$\{(\omega, \eta) : \omega_i \neq \eta_i \text{ for exactly one } i\}.$$

If (ω, η) is an edge, we write $\omega \sim \eta$. We use this to define the *Graph Laplacian* of f as

$$L(f)(\omega) = \tfrac{1}{2} \sum_{\{\eta : \eta \sim \omega\}} (f(\eta) - f(\omega)),$$

and the *energy* $\mathcal{E}(f)$ of f as $\mathcal{E}(f) = -\langle f, L(f) \rangle$. Let us calculate the Laplacian for the Walsh functions. If $\omega \sim \eta$ and $\omega_i \neq \eta_i$, then

$$\begin{aligned} w_A(\omega) &= w_A(\eta) \text{ if } i \notin A, \\ w_A(\omega) &= -w_A(\eta) \text{ if } i \in A, \end{aligned}$$

so that $L(w_A) = -|A| w_A$. Thus the Walsh functions are the eigenvectors of

L, and L corresponds to differentiation. Further,

$$-L(f) = \sum_{i=1}^{d} \hat{f}_i \epsilon_i + \sum_{|A|>1} |A| \hat{f}_A w_A,$$

so that

$$\mathcal{E}(f) = \sum_{i=1}^{d} \hat{f}_i^2 + \sum_{|A|>1} |A| \hat{f}_A^2.$$

Thus

$$2\left\| f \right\|_2^2 = \langle f, f \rangle \leq \mathcal{E}(f) + 2(\mathbf{E}(f))^2 + \sum_{i=1}^{d} \hat{f}_i^2.$$

We now embed D_2^d as the vertices of the unit cube of l_∞^d. Let $f(x) = \|x_1 a_1 + \cdots + x_d a_d\|$, so that $f(\omega) = \|S_d(\omega)\|$, $\langle f, f \rangle = \|S_d\|_{L^2(E)}^2$, and $\mathbf{E}(f) = \|S_d\|_{L^1(E)}$. Since f is an even function, $\hat{f}_i = 0$ for $1 \leq i \leq d$, and since f is convex and positive homogeneous,

$$\frac{1}{d} \sum_{\{\eta:\eta\sim\omega\}} f(\eta) \geq f\left(\frac{1}{d} \sum_{\{\eta:\eta\sim\omega\}} \eta\right) = f\left(\frac{d-2}{d}\omega\right) = \frac{d-2}{d} f(\omega),$$

by Jensen's inequality. Consequently

$$-Lf(\omega) \leq \tfrac{1}{2}(df(\omega) - (d-2)f(\omega)) = f(\omega)$$

so that $\mathcal{E}(f) \leq \|f\|_2^2$ and $2\|f\|_2^2 \leq \|f\|_2^2 + 2(\mathbf{E}(f))^2$. Thus $\|S_d\|_{L^2(E)} \leq \sqrt{2}\,\|S_d\|_{L^1(E)}$. $\square$

2

Finite-dimensional normed spaces

2.1 The Banach-Mazur distance

Suppose that $(E, \|.\|_E)$ and $(F, \|.\|_F)$ are two d-dimensional normed spaces. The *Banach-Mazur* distance $d(E, F)$ is defined as

$$d(E, F) = \inf\{\|T\| \, . \left\|T^{-1}\right\| : T \text{ invertible in } L(E, F)\}.$$

Proposition 2.1.1 *$d(l_1^d, l_2^d) = \sqrt{d}$.*

Proof Let $I : l_1^d \to l_2^d$ be the identity map. Then $\|I\| = 1$, and, since, by Cauchy-Schwarz, $\sum_{i=1}^d |a_i| \le \sqrt{d}(\sum_{i=1}^d |a_i|^2)^{1/2}$ with equality when $a_i = 1$ for all i, $\|I^{-1}\| = \sqrt{d}$. Thus $d(l_1^d, l_2^d) \le \sqrt{d}$.

Suppose that $T : l_1^d \to l_2^d$ be invertible: without loss of generality, suppose that $\|T\| = 1$. By the parallelogram law,

$$\frac{1}{2^d} \sum \{ \left\| \sum_{i=1}^d \epsilon_i T(e_i) \right\|^2 : \epsilon_i = \pm 1\} = \sum_{i=1}^d \|T(e_i)\|^2 \le d,$$

so there exists (ϵ_i) such that $\left\|\sum_{i=1}^d \epsilon_i T(e_i)\right\| \le \sqrt{d}$. But $\left\|\sum_{i=1}^d \epsilon_i e_i\right\| = d$, and so $\|T^{-1}\| \ge \sqrt{d}$. Thus $d(l_1^d, l_2^d) \ge \sqrt{d}$. □

Corollary 2.1.1 *$d(l_2^d, l_\infty^d) = \sqrt{d}$.*

Proof Duality. □

What about $d(l_1^d, l_\infty^d)$? By the above, it's bounded by d. Is this the right order?

Define matrices W_k recursively by $W_0 = [1]$,

$$W_k = \frac{1}{\sqrt{2}} \begin{bmatrix} W_{k-1} & -W_{k-1} \\ W_{k-1} & W_{k-1} \end{bmatrix}.$$

Then each W_k is a $2^k \times 2^k$ orthogonal matrix, and each entry in W_k has modulus $1/2^{k/2}$. Thus $\left\| W_k : l_1^{2^k} \to l_\infty^{2^k} \right\| = 1/2^{k/2}$. On the other hand, we can write $W_k^{-1} = IW_k^{-1}J$, where $J : l_\infty^{2^k} \to l_2^{2^k}$ and $I : l_2^{2^k} \to l_1^{2^k}$ are the identity maps, each of norm $2^{k/2}$, and $W_k^{-1} : l_2^{2^k} \to l_2^{2^k}$ is an isometry. Thus $d(l_1^{2^k}, l_\infty^{2^k}) \le 2^{k/2}$.

Exercises

2.1 Show that there exists C such that $d(l_1^d, l_\infty^d) \le C\sqrt{d}$.

2.2 Show that $d(l_2^d, l_p^d) \le 2^{|1/p-1/2|}$.

2.3 Suppose that $(E, \|.\|_E)$ and $(F, \|.\|_F)$ are two d-dimensional normed spaces. Show that there exists T invertible in $L(E,F)$ with $\|T\| . \|T^{-1}\| = d(E,F)$.

2.2 Caratheodory's theorem

Theorem 2.2.1 *Suppose that E is d-dimensional and that A is a non-empty bounded closed subset of E. If $x \in conv(A)$, x can be written as*

$$x = \sum_{i=0}^{d} \lambda_i a_i, \text{ where } a_i \in A,\ \lambda_i \ge 0 \text{ and } \sum_{i=0}^{d} \lambda_i = 1.$$

Proof We can write

$$x = \sum_{i=0}^{k} \lambda_i a_i, \text{ where } a_i \in A,\ \lambda_i > 0 \text{ and } \sum_{i=0}^{k} \lambda_i = 1,$$

with k as small as possible. Let $y = x - a_0$ and $b_i = a_i - a_0$, for $0 \le i \le k$, so that $b_0 = 0$ and $y = \sum_{i=0}^{k} \lambda_i b_i$. We show that $b_1, \ldots, b_k$ are linearly independent, so that $k \le d$.

Suppose not. Then we can write $\mu_1 b_1 + \cdots + \mu_k b_k = 0$ for some $\mu_1, \ldots, \mu_k$, with not all μ_k zero. Let $\mu_0 = -(\mu_1 + \cdots + \mu_k)$, so that $\sum_{i=0}^{k} \mu_i = 0$. There exists j such that λ_j/μ_j is minimal positive. Then

$$y = \sum_{i=0}^{k} (\lambda_i - \frac{\lambda_j}{\mu_j}\mu_i) b_i = \sum_{i \ne j} \nu_i b_i.$$

But $\nu_i \geq 0$ for all i, and $\sum_i \nu_i = 1$, and $x = \sum_{i \neq j} \nu_i a_i$, contradicting the minimality of k.

Let $\Delta_{d+1} = \{(\lambda_1, \ldots, \lambda_{d+1}) : \lambda_i \geq 0, \sum_i \lambda_i = 1\}$. The map

$$(\lambda, (a_1, \ldots, a_d)) \to \sum_{i=1}^{d+1} \lambda_i a_i : \Delta_{d+1} \times \prod_{i=1}^{d+1} (A)_i \to E$$

maps the compact set $\Delta_{d+1} \times \prod_{i=1}^{d+1} (A)_i$ continuously onto $\mathrm{conv}(A)$, and so $\mathrm{conv}(A)$ is compact, and therefore closed. $\square$

2.3 Operator norms

Suppose that $(E, \|.\|_E)$ and $(F, \|.\|_F)$ are finite-dimensional normed spaces of dimensions m and n respectively. We can give $L(E, F)$ the operator norm $\|.\|$, but there are other interesting possibilities. The first is defined in terms of trace duality.

If $T \in L(E, F)$ has rank 1 then we can write $T(x) = \phi(x)y$, where $\phi \in E^*$ and $T(E) =$ span (y). We write T as $\phi \otimes y$.

Suppose that $(e_i, \ldots e_m)$ is a basis for E, with dual basis $(\phi_1, \ldots, \phi_m)$. If $T \in L(E, F)$ then

$$T(x) = T(\sum_{i=1}^{m} \phi_i(x) e_i) = \sum_{i=1}^{m} \phi_i(x) T(e_i),$$

so that $T = \sum_{i=1}^{m} \phi_i \otimes T(e_i)$. Thus if $S \in L(F, E)$ then $ST = \sum_{i=1}^{m} \phi_i \otimes ST(e_i)$. Thus

$$tr(ST) = \sum_{i=1}^{m} \phi_i(ST(e_i) = \sum_{i=1}^{m} S^*(\phi_i)(T(e_i)).$$

This is non-singular bilinear form, which does not depend on the choice of basis of E, on $L(E, F) \times L(F, E)$, which we denote by $\langle T, S \rangle$. in this way, we identify $L(F, E)$ with the dual of $L(E, F)$. Thus if α is a norm on $L(E, F)$ that there is a dual norm α^* on $L(F, E)$:

$$\alpha^*(S) = \sup\{|\langle T, S \rangle| : \alpha(T) \leq 1\}.$$

We denote the norm dual to the operator norm by n: n is the *nuclear* norm.

Theorem 2.3.1

$$n(S) = \inf\{\sum_{j=1}^{mn+1} \|\phi_j\| \|x_j\| : S = \sum_{j=1}^{mn+1} \phi_j \otimes x_j\}.$$

Proof If $T \in L(E,F)$ then there exists $x \in E$ with $\|x\| = 1$ such that $\|T(x)\| = \|T\|$, and there exists $\phi \in F^*$ with $\|\phi\| = 1$ such that $\phi(T(x)) = \|T(x)\| = \|T\|$. Thus $\langle T, \phi \otimes x\rangle = \|T\|$. Let $A = \{\phi \otimes x : \|\phi\| = \|x\| = 1\}$. Then $\|T\| = \sup\{\langle T, S\rangle : S \in A\}$, and by the theorem of bipolars, the n-unit ball of $L(F,E)$ is

$$\begin{aligned}
\overline{\text{conv}}(A) &= \text{conv}(A) \\
&= \{S : S = \sum_{j=1}^{mn+1} \lambda_j \phi_j \otimes x_j : \|\phi_j\| = \|x_j\| = 1, \lambda_j \geq 0, \sum_{j=1}^{mn+1} \lambda_j = 1\} \\
&= \{S : S = \sum_{j=1}^{mn+1} \phi_j \otimes x_j : \sum_{j=1}^{mn+1} \|\phi_j\|\,\|x_j\| \leq 1\}.
\end{aligned}$$

Thus $n(S) \leq 1$ if and only if $S = \sum_{j=1}^{mn+1} \phi_j \otimes x_j$ with $\sum_{j=1}^{mn+1} \|\phi_j\|\,\|x_j\| \leq 1$, from which the result follows. □

Here is another example. Suppose that $E = l_2^m$, $F = l_2^n$. If $S, T \in L(E,F)$ then S and T are represented by matrices, and

$$tr(S^*T) = \langle T, S^*\rangle = \sum_{i,j} \bar{s}_{ij} t_{ij}.$$

This is an inner product on $L(E,F)$ which we denote by $\langle T, S\rangle_{HS}$: the corresponding norm $\|T\|_{HS} = (\sum_{ij} |t_{ij}|^2)^{1/2} = (\sum_j \|T(e_j)\|^2)^{1/2}$ is called the *Hilbert-Schmidt* norm.

Can we define this in a co-ordinate-free way?

Proposition 2.3.1 *If $E = l_2^m$, $F = l_2^n$ and $T \in L(E,F)$ then $\|T\|_{HS}^2 = \pi_2(T)$, where*

$$(\pi_2(T))^2 = \sup\{\sum_{j=1}^{k} \|T(x_j)\|^2 : \sum_{j=1}^{k} |\langle x_i, y\rangle|^2 \leq 1 \textit{ for } \|y\| \leq 1\}.$$

Proof First take $x_j = e_j$ for $1 \leq i \leq m$. If $\|y\| \leq 1$ then $\sum_{j=1}^m |\langle e_j, y\rangle|^2 = \|y\|^2 \leq 1$, and so $\|T\|_{HS}^2 = \sum_{j=1}^m \|T(e_j)\|^2 \leq (\pi_2(T))^2$.

Conversely, suppose that $\sum_{i=1}^k |\langle x_i, y\rangle|^2 \leq 1$ for $\|y\| \leq 1$. Let $f_1, \dots, f_n$ be the unit vectors in $F = l_2^n$. Then

$$\begin{aligned}
\sum_{j=1}^{m} \|T(x_j)\|^2 &= \sum_{i,j} |\langle T(x_j), f_i\rangle|^2 = \sum_{i,j} |\langle x_j, T^*(f_i)\rangle|^2 \\
&\leq \sum_{i=1}^{n} \|T^*(f_j)\|^2 = \|T^*\|_{HS}^2 = \|T\|_{HS}^2 .
\end{aligned}$$

□

We can define $\pi_2(T)$ for general E and F:

$$\pi_2(T) = \sup\{(\sum_{j=1}^{k} \|T(x_j)\|^2)^{1/2} : \sum_{j=1}^{k} |\phi(x_j)|^2 \leq 1 \text{ for } \|\phi\| \leq 1\}.$$

π_2 is the 2-*summing* norm. Note that $\pi_2(RST) \leq \|R\|\,\pi_2(S)\,\|T\|$.

We have the following factorization result.

Theorem 2.3.2 *Suppose that dim $E = m$ and that F is an inner-product space. Let H_m be an inner-product space of dimension m. If $T \in L(E,F)$, we can write $T = AB$, where $B \in L(E,H_m)$ and $A \in L(H_m,F)$ and $\pi_2(A) \leq \sqrt{n(T)}$, $\pi_2(B) \leq \sqrt{n(T)}$.*

Proof By Theorem 2.3.1, we can write $T = \sum_{j=1}^{k} \phi_j \otimes y_j$, with $k = m$ dim $(T(E)) + 1$ and $\sum_{j=1}^{k} \|\phi_j\|^* \|y_j\| = n(T)$. We can scale so that $\|\phi_j\|^* = \|y_j\|$, so that $\sum_{j=1}^{k}(\|\phi_j\|^*)^2 = \sum_{j=1}^{k} \|y_j\|^2 = n(T)$. Now let

$$\begin{aligned} S(x) &= (\phi_j(x)), \text{ so that } S : E \to l_2^k \\ R(\alpha) &= \sum_{j=1}^{k} \alpha_j y_j, \text{ so that } R : l_2^k \to F; \end{aligned}$$

$T = RS$. dim $(S(E)) \leq m$; let H_m be an m-dimensional subspace of l_2^k containing $S(E)$, and let $B : E \to H_m$ be defined by $B(x) = S(x)$. Let $J : H_m \to l_2^k$ be the inclusion mapping, and let A be the restriction of R to H_m. Then $T = AB$.

Suppose that $z_1, \dots z_m \in E$, with $\sup\{\sum_i |\phi(z_i)|^2 : \|\phi\|^* \leq 1\} \leq 1$. Then

$$\begin{aligned} \sum_{i=1}^{l} \|B(z_i)\|^2 &= \sum_{i=1}^{l}\sum_{j=1}^{k} |\phi_j(z_i)|^2 \\ &= \sum_{j=1}^{k}\sum_{i=1}^{l} |\phi_j(z_i)|^2 \leq \sum_{j=1}^{k} \|\phi_j\|^2 = n(T), \end{aligned}$$

so that $\pi_2(B) \leq \sqrt{n(T)}$.

Also $\sum_{j=1}^{k} \|R(e_j)\|^2 = \sum_{j=1}^{k} \|y_j\|^2 = n(T)$, and so $\pi_2(R) = \sqrt{n(T)}$. Then $\pi_2(A) = \pi_2(JR) \leq \pi_2(R) = \sqrt{n(T)}$. □

Exercises

2.1 Suppose that $T \in L(E,F)$. Show that $n(T) = n(T^*)$.

2.2 Suppose that $T \in L(l_\infty^d, E)$, so that $T^* \in L(E^*, l_1^d)$. Show that

$$\begin{aligned} n(T) = n(T^*) &= \sup\{\sum_{i=1}^d \langle T^*(\phi_i), e_i\rangle : \|\phi_i\| \leq 1\} \\ &= \sup\{\sum_{i=1}^d \phi_i(T(e_i) : \|\phi_i\| \leq 1\} \\ &= \sum_{i=1}^d \|T(e_i)\|. \end{aligned}$$

Compare this with the results in the proof of Auerbach's theorem, below.

2.4 Lewis' theorem

Theorem 2.4.1 *Suppose that dim $E =$ dim $F = n$ and that α is a norm on $L(E,F)$. Then there exists $T \in L(E,F)$ with $\alpha(T) = 1$ and $\alpha^*(T^{-1}) = n$.*

Proof Note that for invertible $S \in L(E,F)$,

$$n = tr(S^{-1}S) \leq \alpha(S)\alpha^*(S^{-1}).$$

Choosing bases for E and F, every S can be represented by a matrix (s_{ij}); we define $\det(S) = \det(s_{ij})$. $|\det(S)|$ is a continuous function on $L(E,F)$, and so it attains its supremum on the α unit sphere at a point T. Certainly this supremum is positive, and so T is invertible. Note that $\alpha^*(T^{-1}) \geq \sqrt{n}$. If $T+S$ is invertible, then $|\det(T+S)/\alpha(T+S)| \leq |\det T|$, so that $|\det(T+S)| \leq (\alpha(T+S))^n |\det T|$. If $S \in L(E,F)$ then $T+\epsilon S$ is invertible for small enough ϵ and then

$$|\det T||\det(I+\epsilon T^{-1}S)| = |\det(T+\epsilon S)| \leq |\det T|(\alpha(T+\epsilon S))^n,$$

so that

$$|\det(I+\epsilon T^{-1}S)| \leq (\alpha(T+\epsilon S))^n \leq (1+\epsilon\alpha(S))^n.$$

But $(\det(I+\epsilon T^{-1}S)-1)/\epsilon \to tr(T^{-1}S)$ as $\epsilon \to 0$, so that $|tr(T^{-1}S| \leq n\alpha(S)$, and so $\alpha^*(T^{-1}) \leq n$. □

Corollary 2.4.1 (Auerbach's Theorem) *Suppose that $(F, \|.\|_F)$ has dimension n. Then there is a basis $(f_1, \ldots, f_n)$ such that*

$$\sup_{1\leq i\leq n} |\alpha_i| \leq \left\| \sum_{i=1}^n \alpha_i f_i \right\|_F \leq \sum_{i=1}^n |\alpha_i|.$$

This says that the identity map $l_1^n \to l_\infty^n$ factors through F, without increase of norm.

Proof Let $E = l_1^n$. For $T \in L(E,F)$ let $\alpha(T) = \|T\| = \sup_i \|T(e_i)\|$. If $S \in L(F,E)$, and $S(y) = (s_1(y),\dots,s_n(y))$, where $s_i \in F^*$, let $\beta(S) = \sum_{i=1}^n \|s_i\|^*$. Then

$$|tr(ST)| = |\sum_{i=1}^n s_i(T(e_i))| \leq \alpha(T)\sum_{i=1}^n \|s_i\|^* = \alpha(T)\beta(S),$$

so that $\alpha^*(S) \leq \beta(S)$ and $\beta^*(T) \leq \alpha(T)$. On the other hand, given T there exists i such that $\alpha(T) = \|T(e_i)\|$, and there exists $\phi \in F^*$ with $\|\phi\|^* = 1$ such that $\phi(T(e_i)) = \|T(e_i)\|$. Let $S = \phi \otimes e_i$. Then $tr(ST) = \|T(e_i)\| = \alpha(T)$ and $\beta(S) = 1$, so that $\beta^*(T) \geq \alpha(T)$ and $\alpha^*(T) \geq \beta(T)$. Thus $\beta = \alpha^*$.

By Lewis' theorem, there exists T with $\alpha(T) = 1$ and $\beta(T^{-1}) = n$. Let $f_i = T(e_i)$. Then $\|f_i\|_F \leq 1$ for each i, so that $\|\sum_{i=1}^n \alpha_i f_i\|_F \leq \sum_{i=1}^n |\alpha_i|$. On the other hand, $T^{-1}(y) = (\phi_1(y),\dots\phi_n(y))$, where $(\phi_1,\dots\phi_n)$ is the dual basis of $(f_1,\dots,f_n)$, so that $\sum_{i=1}^n \|\phi_i\|^* = n$. But $\phi_i(f_i) = 1$, so that $\|\phi_i\|^* \geq 1$ for each i. Thus $\|\phi_i\|^* = 1$ for each i, and so $|\alpha_i| = |\phi_i(\sum_{j=1}^n \alpha_j f_j)| \leq \left\|\sum_{j=1}^n \alpha_j f_j\right\|$ for each i. □

2.5 The ellipsoid of maximal volume

Theorem 2.5.1 *Suppose that dim $(E) = n$ and that α is a norm on $L(l_2^n, E)$ with the property that $\alpha(TR) \leq \|R\|\,\alpha(T)$ for all $R \in L(l_2^n)$. If S,T are invertible elements of $L(l_2^n,E)$ with $\alpha(S) = \alpha(T) = 1$, $\alpha^*(S^{-1}) = \alpha^*(T^{-1}) = n$ then there exists a unitary (orthogonal) U such that $T = SU$.*

Proof Note that if U is unitary (orthogonal) then $\alpha(JU) \leq \alpha(J) = \alpha(JUU^*) \leq \alpha(JU)$, so that $\alpha(JU) = \alpha(J)$. It is enough to prove the result for a particular S: by Lewis' Theorem, we can suppose that $\det(S) = \sup\{\det(J) : \alpha(J) \leq 1\}$.

Let $V = S^{-1}T$. Then V^*V is positive, so there exists an orthonormal basis (f_i) of l_2^n and $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \lambda_n^2 > 0$ such that $V^*V(f_i) = \lambda_i^2 f_i$. Note that $\|V(f_i)\|^2 = \langle V^*V(f_i), f_i\rangle = \lambda_i^2$ and that $\langle V(f_i), V(f_j)\rangle = \langle V^*V(f_i), f_j\rangle = 0$ for $i \neq j$. Let $W(f_i) = V(f_i)/\lambda_i$ and $R(f_i) = \lambda_i f_i$, and extend by linearity to define W, R in $L(l_2^n)$. $V = WR$, and W is unitary, R positive, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n > 0$.

Now $T = SWR$, so that $|\det T| = |\det S \det W \det R| = (\prod_i \lambda_i)|\det S|$. Thus $\prod_i \lambda_i \leq 1$.

Also $|tr(R^{-1})| = |tr(T^{-1}SW)| \leq \alpha^*(T^{-1})\alpha(SW) = n$, so that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i} \leq 1 \leq (\prod_{i=1}^{n}\frac{1}{\lambda_i})^{1/n}.$$

Thus we have equality in the AM-GM inequality, which happens only when all the terms are equal. Thus $\lambda_i = 1$ for all i, $R = I$ and $T = SW$. □

Let us apply this to $\alpha(T) = \|T\|$. There exists T such that $|\det(T)| = \sup\{|\det J| : \|J\| \leq 1\}$. This says that $T(B_{l_2^n})$ is an ellipsoid of maximal volume within B_E. If $S(B_{l_2^n})$ is another such, then $T = SU$, so that $T(B_{l_2^n}) = SU(B_{l_2^n}) = S(B_{l_2^n})$. Thus the ellipsoid of maximal volume is unique.

Theorem 2.5.2 *Let $T \in L(l_2^n, E)$ be such that $T(B_{l_2^n})$ is the ellipsoid of maximal volume within B_E. Then $\pi_2(T^{-1}) = \sqrt{n}$.*

Proof
First, $\sqrt{n} = \pi_2(I_{l_2^n}) = \pi_2(T^{-1}T) \leq \pi_2(T^{-1})\,\|T\| = \pi_2(T^{-1})$.

By Lewis' Theorem, $n(T^{-1}) = n$, and by Theorem 2.3.2 we can write $T^{-1} = AB$, where $B \in L(E, H_n)$ and $A \in L(H_n, l_2^n)$ and $\pi_2(A) \leq \sqrt{n}$, $\pi_2(B) \leq \sqrt{n}$. But

$$\langle BT, A^*\rangle_{HS} = tr(ABT) = tr(I_n) = n \leq \pi_2(A)\pi_2(BT) \leq \pi_2(A)\pi_2(B) \leq n,$$

and so we have equality throughout. Thus, since we have equality in the Cauchy-Schwarz inequality, $A^* = \alpha BT = \alpha A^{-1}$, for some $\alpha > 0$, so that $A = \sqrt{\alpha}J$, where J is an isometry. Then $\pi_2(A) = \sqrt{\alpha n}$, and so $\alpha \leq 1$. Thus $\pi_2(T^{-1}) = \pi_2(AB) \leq \|A\|\,\pi_2(B) \leq \sqrt{n}$. □

Corollary 2.5.1 $\pi_2(I_E) \leq \sqrt{n}$.

Proof $\pi_2(I_E) = \pi_2(TT^{-1}) \leq \|T\|\,\pi_2(T^{-1}) \leq \sqrt{n}$. □

In fact, we have equality.

Corollary 2.5.2 (John) $\|T^{-1}\| \leq \sqrt{n}$, *and* $d(E, l_2^n) \leq \sqrt{n}$.

Thus, if $\mathcal{E}$ is the ellipsoid of maximum volume in B_E, $\mathcal{E} \subseteq B_E \subseteq \sqrt{n}\mathcal{E}$.

$\mathcal{E}$ can be taken as the unit ball of an inner-product norm on E.

Proposition 2.5.1 *There exists an orthonormal basis (e_i) of E with respect to the inner-product norm defined by $\mathcal{E}$ such that*

$$1 = \|e_1\|_E \geq \|e_2\|_E \cdots \geq \|e_n\|_E \text{ and } \|e_j\|_E \geq 2^{-n/(n-j)} \text{ for each } j.$$

Proof There exists $e_1 \in \mathcal{E}$ with $\|e_1\|_E = 1$. Having defined $e_1, \ldots, e_{k-1}$, let $F_k = \{e_1, \ldots, e_{k-1}\}^\perp$. Choose $e_k \in \mathcal{E} \cap F_k$ with $\|e_k\|_E$ as large as possible. Note that $\mathcal{E} \cap F_k \subseteq \|e_k\|_E B_E$. Consider the ellipsoid

$$\mathcal{D}_j = \{\sum_{i=1}^n a_i e_i : \sum_{i=1}^{j-1} |a_i|^2 + \|e_j\|_E^2 (\sum_{i=j}^n |a_i|^2) \leq 1/4\}.$$

If $\sum_{i=1}^n a_j e_j \in D_j$ then $\sum_{i=1}^{j-1} |a_i|^2 \leq 1/4$, so that $\sum_{i=1}^{j-1} a_i e_i \in \mathcal{E}/2 \subseteq B_E/2$. Similarly $\sum_{i=j}^n |a_i|^2 \leq 1/(4\|e_j\|_E^2)$, so that $\sum_{i=j}^n a_i e_i \in (\mathcal{E} \cap F_j)/(2\|e_j\|_E) \subseteq B_E/2$. Thus $\mathcal{D}_j \subset B_E$, and so vol $(\mathcal{D}_j) \leq$ vol $(\mathcal{E})$. But

$$\text{vol}\,(\mathcal{D}_j) = 2^{-n} \|e_j\|_E^{-(n-j)} \ \text{vol}\,(\mathcal{E}),$$

and so the result follows. □

Corollary 2.5.3 $\|e_j\|_E \geq 1/4$ *for* $1 \leq j \leq n/2$.

Exercises

(i) Suppose that α is a norm on $L(l_2^d, F)$. Show that the following are equivalent:
 (a) $\alpha(TS) \leq \alpha(T)\|S\|$ for all $S \in L(l_2^d)$, $T \in L(l_2^d, F)$;
 (b) $\alpha(TU) \leq \alpha(T)$ for all orthogonal $U \in L(l_2^d)$, $T \in L(l_2^d, F)$;
 (c) $\alpha(TU) = \alpha(T)$ for all orthogonal $U \in L(l_2^d)$, $T \in L(l_2^d, F)$.

 In the following exercises, a d-dimensional normed space $(E, \|.\|_E)$ is identified with $\mathbf{R}^d$ in such a way that the unit ball in the Euclidean norm $|.|$ is the ellipsoid of maximum volume. We also use the inner product to identify E^* with $\mathbf{R}^d$.

(ii) Show that $\|x\|_E \leq |x| \leq \|x\|_{E^*}$.

(iii) Show that there are non-zero vectors $y_1, \ldots y_k$ and vectors $x_1, \ldots x_k$, with $|x_j| = 1$ for all j such that $z = \sum_{j=1}^k \langle z, y_j \rangle x_j$ for all z, and $n = \sum_{j=1}^k \|y_j\|_{E^*}$.

(iv) By considering traces, show that

$$n = \sum_{j=1}^k \langle x_j, y_j \rangle \leq \sum_{j=1}^k |y_j| \leq \sum_{j=1}^k \|y_j\|_{E^*} = n.$$

(v) Deduce that $y_j = c_j x_j$, where $c_j > 0$, and that $\|x_j\|_{E^*} = 1$.

(vi) By considering $\langle x_j, x_j \rangle$, deduce that $\|x_j\|_E = 1$.

(vii) Conclude that $z = \sum_{j=1}^k c_j \langle z, x_j \rangle x_j$ for all z, where $\sum_{j=1}^k c_j = n$ and the points x_j are contact points: points with $\|x_j\|_E = |x_j| = 1$, and where the hyperplane $\{z : \langle z, x_j \rangle = 1\}$ is tangent to both unit balls.

2.6 Estimating the median

We consider S^{d-1}, with Haar measure $\mu = \mu_{d-1}$, as a subset of $\mathbf{R}^d$ with Gaussian measure γ_d.

Proposition 2.6.1 *Suppose that f is a 1-Lipschitz function on S^{d-1}, with median M_f and mean $\mathbf{E}(f) = A_f$. Then $|A_f - M_f| \leq \pi/4\sqrt{d-2}$.*

Proof

$$\mathbf{E}((f-M_f)^+) = \int_0^\infty \mu_{d-1}(f > M_f+t)\,dt \leq \sqrt{\frac{\pi}{8}} \int_0^\infty e^{-(d-2)t^2/2}\,dt = \frac{\pi}{4\sqrt{d-2}}.$$

Similarly $\mathbf{E}((f - M_f)^-) \leq \pi/4\sqrt{d-2}$. □

Proposition 2.6.2 *Suppose that $4 \leq k \leq d$, and let $m_k(x) = \max_{i=1}^k |x_i|$. Then*

$$\int_{S^{d-1}} m_k \, d\mu_{d-1} \geq c\sqrt{\frac{\log k}{d}},$$

where c is an absolute constant.

Proof Since $\int_{S^{d-1}} m_k \, d\mu_{d-1} = (I_{d-1}/\sqrt{2\pi}) \int_{R^d} m_k \, d\gamma_d$ (Exercise 1.6 (vi)), it is enough to show that $\int_{R^d} m_k \, d\gamma_d \geq c\sqrt{\log k}$. Now

$$\gamma(|x| > \alpha) \geq 2\gamma(\alpha < x < \alpha + 1) \geq \frac{2}{\sqrt{2\pi}} e^{-(\alpha+1)^2/2},$$

so that

$$\gamma_d(m_k \leq \alpha) \leq (1 - \frac{2}{\sqrt{2\pi}} e^{-(\alpha+1)^2/2})^k.$$

Now put $\alpha = \sqrt{2\log k} - 1$, so that $k = e^{(\alpha+1)^2/2}$ and

$$\gamma_d(m_k \leq \alpha) \leq (1 - \sqrt{\frac{2}{\pi}}\frac{1}{k})^k \leq e^{-\sqrt{2/\pi}} < 1/2,$$

and

$$\int_{R^d} m_k \, d\gamma_d \geq \frac{\alpha}{2} = \frac{1}{2}(\sqrt{2\log k} - 1) \geq \frac{\sqrt{\log k}}{4}.$$

□

Suppose that $(E, \|.\|_E)$ is an d-dimensional normed space. We consider the ellipsoid of maximal volume contained in B_E, use this to define an inner product on E, take an orthonormal basis $(e_1, \ldots, e_d)$ which satisfies Proposition 2.5.1, and use this to identify E with $\mathbf{R}^d$. We consider $r(x) =$

$\|x\|_E$ on S^{d-1}. It is a non-negative Lipschitz function on S^{d-1}, $1/\sqrt{d} \leq r(x) \leq 1$, and $r(e_j) \geq 1/4$ for $1 \leq j \leq d/2$.

Proposition 2.6.3 *If $d \geq 3$ then $M_r/2 \leq A_r \leq 2M_r$.*

Proof Since $r \geq 0$, $A_r \geq \int_{(r \geq M_r)} r\, d\mu_{d-1} \geq M_r/2$. Also $M_r \geq 1/\sqrt{d}$, so that

$$\frac{\pi M_r}{4} \geq \frac{\pi}{4\sqrt{d}} \geq \sqrt{\frac{(d-2)}{d}}(A_r - M_r),$$

so that $A_r \leq (\pi/4 + 1)M_r \leq 2M_r$. □

Proposition 2.6.4 *There exists an absolute constant c such that $M_r \geq c\sqrt{\log d/d}$.*

Proof We can suppose that $d \geq 8$. It is enough to establish the corresponding result for A_r. Let $(\epsilon_1, \ldots, \epsilon_d)$ be Bernoulli random variables. Then

$$\begin{aligned} \int_{S^{d-1}} \left\| \sum_{i=1}^d a_i e_i \right\| \, d\mu_{d-1}(a) &= \int_{S^{d-1}} \left\| \sum_{i=1}^d \epsilon_i(\omega) a_i e_i \right\| \, d\mu_{d-1} \\ &= \int_{S^{d-1}} \mathbf{E}(\left\| \sum_{i=1}^d \epsilon_i a_i e_i \right\|) \, d\mu_{d-1}. \end{aligned}$$

Now $\mathbf{E}(\left\|\sum_{i=1}^d \epsilon_i a_i e_i\right\|) \geq \|a_j e_j\|$ for each j, so that

$$\mathbf{E}(\left\| \sum_{i-1}^d \epsilon_i a_i e_i \right\|) \geq \max_{j \leq d/2} \|a_j e_j\| \geq \frac{1}{4} \max_{j \leq d/2} |a_j|,$$

and so

$$A_r = \int_{S^{d-1}} \left\| \sum_{i=1}^d a_i e_i \right\| \, d\mu_{d-1} \geq \frac{1}{4} \int_{S^{d-1}} m_{d/2}(a) \, d\mu_{d-1}(a) \geq c\sqrt{\frac{\log d}{d}}.$$

□

2.7 Dvoretzky's theorem

Proposition 2.7.1 *Suppose that $(F, \|.\|_F)$ is a k-dimensional normed space and that $1 > \theta > 0$. Then there exists a θ-net in $S_F = \{x : \|x\|_F = 1\}$ with*

$$|N| \leq (1 + 2/\theta)^k \leq (3/\theta)^k.$$

Proof Let N be a maximal subset of S_F with $\|x-y\| > \theta$ for $x, y \in N$. Then N is a θ-net. The sets $\{x + (\theta/2)B_F\}_{x \in N}$ are disjoint, and contained in $(1+\theta/2)B_F$, and so $\sum_{x \in N}$ vol $(x + (\theta/2)B_F) \leq$ vol $(1+\theta/2)B_F$, which gives the result. □

Suppose that $(E, \|.\|_E)$ is an d-dimensional normed space. As before, we consider the ellipsoid of maximal volume contained in B_E, use this to define an inner product on E, take an orthonormal basis $(e_1, \ldots, e_d)$ which satisfies Proposition 2.5.1, and use this to identify E with $\mathbf{R}^d$. We consider $r(x) = \|x\|_E$ on S^{d-1}. If F is a k-dimensional subspace of E (an element of the Grassmannian $G_{d,k}$), we denote the Euclidean sphere in F by S_F.

Proposition 2.7.2 *Suppose that* $1 > \epsilon > 0$ *and that* $1 > \theta > 0$*. Suppose that* $k log(3/\theta) \leq (d-2)\epsilon^2/4$*. Let*

$$C_k = \{F \in G_{d,k} : \textit{ there exists a } \theta\textit{-net } N \subseteq S_F \textit{ with } \sup_{x \in N} |r(x) - M_r| < \epsilon\}.$$

Then $\mathbf{P}(C_k) \geq 1 - \sqrt{\pi/2}e^{-\epsilon^2(d-2)/4}$.

Proof Let $A = \{x \in S^{d-1} : r(x) = M_r\}$, and let $B = \{x \in S^{d-1} : |r(x) - M_r| > \epsilon\}$. By Proposition 1.10.2,

$$\mu_{d-1}(B) \leq \sqrt{\frac{\pi}{2}}e^{-\epsilon^2(d-2)/2}.$$

Let F be any k-dimensional subspace of E, and let N be a θ-net in S_F with $|N| \leq (3/\theta)^k \leq e^{(d-2)\epsilon^2/4}$. For each $x \in N$, $\mathbf{P}(U \in SO_d : U(x) \in B) = \mu(B)$, by Proposition 1.4.2. Thus if $G = \{U \in SO_d : U(x) \in B \text{ for some } x \in N\}$ then

$$\mathbf{P}(G) \leq |N|\mu(B) \leq \sqrt{\frac{\pi}{2}}e^{-\epsilon^2(d-2)/4}.$$

But if $U \notin G$ then $U(N)$ is a net in $U(F)$ with the required properties. Applying Proposition 1.4.2, we see that $\mathbf{P}(C_k) \geq 1 - \mathbf{P}(G)$. □

Theorem 2.7.1 (Dvoretzky's theorem) *Suppose that* $d \geq 4$*. Suppose that* $1 > \delta > 0$*. There exists a constant* $c = c(\delta)$ *such that if* $(E, \|.\|_E)$ *is a d-dimensional normed space, and* $|.|$ *is the norm defined by the ellipsoid of maximal volume contained in* B_E*, and if* $k \leq cdM_r^2$ *and*

$$D_k = \{F \in G_{d,k} : (1-\delta)M_r|x| \leq \|x\|_E \leq (1+\delta)M_r|x| \textit{ for } x \in E\}$$

then $P(D_k) \geq 1 - \sqrt{\pi/2}e^{-cdM_r^2}$.

Proof Let $\theta = \delta/3$, let $c = c_\delta = \theta^2/(8\log(3/\theta))$ and let $\epsilon = \theta M_r$. Then

$$\frac{\epsilon^2(d-2)}{4\log 3/\theta} = 2cM_r^2(d-2) \geq cdM_r^2 \geq k \text{ and } cdM_r^2 \leq \frac{\epsilon^2(d-2)}{4}.$$

Thus, defining C_k as in Proposition 2.7.2,

$$\mathbf{P}(C_k) \geq 1 - \sqrt{\frac{\pi}{2}} e^{-(d-2)\epsilon^2/4} \geq 1 - \sqrt{\frac{\pi}{2}} e^{-cdM_r^2}.$$

Suppose that $F \in C_k$, and that N is a suitable θ-net in F. Suppose that $x \in S_F$. Then there exists $n_0 \in N$ with $|x - n_0| = \alpha_1\theta$ with $0 \leq \alpha_1 \leq 1$. If $\alpha_1 \neq 0$, there exists $n_1 \in N$ with $|(x-n_0)/\alpha_1\theta - n_1| = \alpha_2\theta$ with $0 \leq \alpha_2 \leq 1$, so that $|x - n_0 - \alpha_1 n_1| \leq \alpha_1\alpha_2\theta^2$. Continuing in this way, we can write

$$x = n_0 + \sum_{j=1}^{\infty} \beta_j \theta^j n_j, \text{ with } 0 \leq \beta_j \leq 1.$$

Thus

$$\|x\| \leq \frac{M_r + \epsilon}{1-\theta} = \frac{1+\theta}{1-\theta} M_r \leq (1+\delta)M_r$$

and

$$\begin{aligned} \|x\| &\geq \|n_0\| - \|x - n_0\| \geq (M_r - \epsilon) - \frac{\theta}{1-\theta}(M_r + \epsilon) \\ &= (1 - \theta - \frac{\theta(1+\theta)}{1-\theta})M_r \geq (1-\delta)M_r, \end{aligned}$$

and so $C_k \subseteq D_k$. □

Recall that $dM_r^2 \geq c\log d$ (Proposition 2.6.4). Thus we have the following general result.

Corollary 2.7.1 *Given* $0 < \delta < 1$ *and* $0 < \eta < 1$ *there exists* $c = c(\delta, \eta) > 0$ *such that if* E *is a* d*-dimensional normed space and* $k \leq c\log d$ *then if*

$$E_k = \{F \in G_{d,k} : d(F, l_2^k) < 1 + \delta\}$$

then $\mathbf{P}(E_k) \geq 1 - 1/d^c$.

2.8 Type and cotype

In certain circumstances, we can improve on Proposition 2.6.4.

In order to do this, we introduce the notions of type and cotype. These involve Bernoulli sequence of random variables: for the rest of this chapter, (ϵ_n) will denote such a sequence.

Let us begin, by considering the parallelogram law. This says that if $x_1, \ldots, x_n$ are vectors in a Hilbert space H then

$$\mathbf{E}\left(\left\|\sum_{j=1}^{n} \epsilon_j x_j\right\|^2\right) = \sum_{j=1}^{n} \|x_j\|_H^2 .$$

We deconstruct this equation; we split it into two inequalities, we change an index, and we introduce constants.

Suppose that $(E, \|.\|_E)$is a Banach spaces and that $1 \leq p < \infty$. We say that E is of *type* p if there is a constant C such that if $x_1, \ldots, x_n$ are vectors in E then

$$\left(\mathbf{E}\left(\left\|\sum_{j=1}^{n} \epsilon_j x_j\right\|_E^2\right)\right)^{1/2} \leq C(\sum_{j=1}^{n} \|x_j\|_E^p)^{1/p}.$$

The smallest possible constant C is denoted by $T_p(E)$, and is called the *type* p constant of E. Similarly, we say that E is of *cotype* p if there is a constant C such that if $x_1, \ldots, x_n$ are vectors in E then

$$(\sum_{j=1}^{n} \|x_j\|_E^p)^{1/p} \leq C\left(\mathbf{E}\left(\left\|\sum_{j=1}^{n} \epsilon_j (x_j)\right\|_E^2\right)\right)^{1/2}.$$

The smallest possible constant C is denoted by $C_p(E)$, and is called the *cotype* p constant of E.

Thus the parallelogram law states that a Hilbert space H is of type 2 and cotype 2, and $T_2(H) = C_2(H) = 1$.

It follows from the parallelogram law that if E is of type p, for $p > 2$, or cotype p, for $p < 2$, then $E = \{0\}$. If E is of type p then E is of type q, for $1 \leq q < p$, and $T_q(E) \leq T_p(E)$; if E is of cotype p then E is of cotype q, for $p < q < \infty$, and $C_q(E) \leq C_p(E)$. Every Banach space is of type 1.

Proposition 2.8.1 *If E is of type p, then E^* is of cotype p', and $C_{p'}(E^*) \leq T_p(E)$.*

Proof Suppose that $\phi_1, \ldots, \phi_n$ are vectors in F^* and $x_1, \ldots, x_n$ are vectors in E. Then

$$|\sum_{j=1}^{n} \phi_j(x_j)| \quad = \quad |\mathbf{E}((\sum_{j=1}^{n} \epsilon_j \phi_j)(\sum_{j=1}^{n} \epsilon_j x_j))|$$

$$
\begin{aligned}
&\leq \mathbf{E}(|(\sum_{j=1}^n \epsilon_j \phi_j)(\sum_{j=1}^n \epsilon_j x_j)|) \\
&\leq \mathbf{E}(\left\|\sum_{j=1}^n \epsilon_j \phi_j\right\| \left\|\sum_{j=1}^n \epsilon_j x_j\right\|) \\
&\leq \left(\mathbf{E}\left\|\sum_{j=1}^n \epsilon_j \phi_j\right\|^2\right)^{1/2} \left(\mathbf{E}\left\|\sum_{j=1}^n \epsilon_j x_j\right\|^2\right)^{1/2} \\
&\leq \left(\mathbf{E}\left\|\sum_{j=1}^n \epsilon_j \phi_j\right\|^2\right)^{1/2} T_p(E)(\sum_{j=1}^n \|x_j\|^p)^{1/p}.
\end{aligned}
$$

But

$$
\begin{aligned}
(\sum_{j=1}^n \|\phi_j\|^{p'})^{1/p'} &= \sup\{|\sum_{j=1}^n |c_j|\,\|\phi_j\| : (\sum_{j=1}^n |c_j|^p)^{1/p} \leq 1\}, \\
&= \sup\{|\sum_{j=1}^n \phi_j(x_j)| : (\sum_{j=1}^n \|x_j\|^p)^{1/p} \leq 1\},
\end{aligned}
$$

and so

$$
(\sum_{j=1}^n \|\phi_j\|^{p'})^{1/p'} \leq T_p(E)\left(\mathbf{E}\left\|\sum_{j=1}^n \epsilon_j \phi_j\right\|^2\right)^{1/2}.
$$

□

As we shall see, the converse of this proposition is not true.

If in the definitions of type and cotype we replace the Bernoulli sequence (ϵ_n) by (g_n), where the g_i are independent $N(0,1)$ random variables, we obtain the definitions of Gaussian type and cotype. We denote the corresponding constants by T_p^γ and C_p^γ.

Proposition 2.8.2 *If E is of type p (cotype p) then it is of Gaussian type p (Gaussian cotype p).*

Proof Let us prove this for cotype: the proof for type is similar. Let $m_p(\gamma) = \|g\|_p$, where g is an $N(0,1)$ random variable. Let $x_1, \ldots, x_n$ be vectors in E. Suppose that the sequence (g_n) is defined on Ω and the sequence (ϵ_n) on Ω'. Then for fixed $\omega \in \Omega$,

$$\begin{aligned}\sum_{j=1}^{n} |(g_j(\omega)|^p \|x_j\|_F^p &\leq \left((C_p(E))^2 \mathbf{E}_{\Omega'}\left(\left\|\sum_{j=1}^{n} \epsilon_j g_j(\omega) x_j\right\|_E^2\right)\right)^{p/2} \\ &\leq (C_p(E))^p \mathbf{E}_{\Omega'}\left(\left\|\sum_{j=1}^{n} \epsilon_j g_j(\omega) x_j\right\|_E^p\right).\end{aligned}$$

Taking expectations over Ω, applying Fubini's Theorem, and using the symmetry of the Gaussian sequence, we find that

$$\begin{aligned}(m_p(\gamma))^p \sum_{j=1}^{n} \|x_j\|_F^p &\leq (C_p(E))^p \mathbf{E}_{\Omega}\left(\mathbf{E}_{\Omega'}\left(\left\|\sum_{j=1}^{n} \epsilon_j g_j x_j\right\|_E^p\right)\right) \\ &\leq (C_p(E))^p \mathbf{E}'_{\Omega}\left(\mathbf{E}_{\Omega}\left(\left\|\sum_{j=1}^{n} \epsilon_j g_j x_j\right\|_E^p\right)\right) \\ &= (C_p(E))^p \mathbf{E}_{\Omega}\left(\left\|\sum_{j=1}^{n} g_j x_j\right\|_E^p\right).\end{aligned}$$

□

In fact, the converse is also true.

We state the following theorem (which we shall not use) without proof.

Theorem 2.8.1 *A Banach space $(E, \|.\|_E)$ is isomorphic to a Hilbert space if and only if it is of type 2 and cotype 2.*

Let us give some examples. We need the following standard result.

Proposition 2.8.3 *Suppose that f is a non-negative measurable function on $(\Omega_1, \Sigma_1, \mu_1) \times (\Omega_2, \Sigma_2, \mu_2)$and that $0 < p \leq q < \infty$. Then*

$$\begin{aligned}&\left(\int_{X_1} (\int_{X_2} f(x,y)^p \, d\mu_2(y))^{q/p} \, d\mu_1(x)\right)^{1/q} \leq \\ &\leq \left(\int_{X_2} (\int_{X_1} f(x,y)^q \, d\mu_1(x))^{p/q} \, d\mu_2(y)\right)^{1/p}.\end{aligned}$$

Proof Let $r = q/p$. Then

$$\left(\int_{X_1} (\int_{X_2} f(x,y)^p \, d\mu_2(y))^{q/p} \, d\mu_1(x)\right)^{1/q} =$$

$$
\begin{aligned}
&= \left(\int_{X_1}(\int_{X_2} f(x,y)^p\,d\mu_2(y))^r\,d\mu_1(x)\right)^{1/rp} \\
&= \left(\int_{X_1}(\int_{X_2} f(x,y)^p\,d\mu_2(y))g(x)\,d\mu_1(x)\right)^{1/p} \\
&\qquad \text{for some } g \text{ with } \|g\|_{r'} = 1 \\
&= \left(\int_{X_2}(\int_{X_1} f(x,y)^p g(x)\,d\mu_1(x))\,d\mu_2(y)\right)^{1/p} \\
&\qquad \text{(by Fubini's theorem)} \\
&\leq \left(\int_{X_2}(\int_{X_1} f(x,y)^{pr}\,d\mu_1(x))^{1/r}\,d\mu_2(y)\right)^{1/p} \\
&= \left(\int_{X_2}(\int_{X_1} f(x,y)^q\,d\mu_1(x))^{p/q}\,d\mu_2(y)\right)^{1/p}.
\end{aligned}
$$

□

We can consider f as a vector-valued function $f(y)$ on Ω_2, taking values in $L^q(\Omega_1)$, and with $\int_{\Omega_2}\|f(y)\|_q^p\,d\mu_2 < \infty$: thus $f \in L^p_{\Omega_2}(L^q_{\Omega_1})$. The proposition then says that $f \in L^q_{\Omega_1}(L^p_{\Omega_2})$ and $\|f\|_{L^q_{\Omega_1}(L^p_{\Omega_2})} \leq \|f\|_{L^p_{\Omega_2}(L^q_{\Omega_1})}$.

Theorem 2.8.2 *Suppose that (Ω, Σ, μ) is a measure space.*

(i) If $1 \leq p \leq 2$ then $L^p(\Omega, \Sigma, \mu)$ is of type p and cotype 2.

(ii) If $2 \leq p < \infty$ then $L^p(\Omega, \Sigma, \mu)$ is of type 2 and cotype p.

Proof (i) Suppose that $f_1, \ldots, f_n$ are in $L^p(\Omega, \Sigma, \mu)$. To prove the cotype inequality, we use the fact that the inclusion $L^p \to L^2$ is norm-decreasing, Khintchine's inequality and Proposition 2.8.3.

$$
\begin{aligned}
\left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^2\right)\right)^{1/2} &\geq \left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^p\right)\right)^{1/p} \\
&= \left(\mathbf{E}\left(\int_\Omega |\sum_{j=1}^n \epsilon_j f_j(\omega)|^p\,d\mu(\omega)\right)\right)^{1/p} \\
&= \left(\int_\Omega \mathbf{E}\left(|\sum_{j=1}^n \epsilon_j f_j(\omega)|^p\right)\,d\mu(\omega)\right)^{1/p} \\
&\geq A_p^{-1}\left(\int_\Omega (\sum_{j=1}^n |f_j(\omega)|^2)^{p/2}\,d\mu(\omega)\right)^{1/p}
\end{aligned}
$$

$$
\begin{aligned}
&\geq A_p^{-1}\left(\sum_{j=1}^n \left(\int_\Omega (|f_j(\omega)|^p)\, d\mu(\omega)\right)^{2/p}\right)^{1/2} \\
&= (A_p^{-1}(\sum_{j=1}^n \|f_j\|_p^2)^{1/2}.
\end{aligned}
$$

Thus $L^p(\Omega, \Sigma, \mu)$ is of cotype 2.

To prove the type inequality, we use Theorem 1.12.2, the fact that the inclusions $L^2 \to L^p \to L^1$ are norm-decreasing, and the fact that the inclusion $l_p \to l_2$ is norm-decreasing.

$$
\begin{aligned}
\left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^2\right)\right)^{1/2} &\leq \sqrt{2}\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p\right) \leq \sqrt{2}\left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^p\right)\right)^{1/p} \\
&= \sqrt{2}\left(\mathbf{E}\left(\int_\Omega |\sum_{j=1}^n \epsilon_j f_j(\omega)|^p\, d\mu(\omega)\right)\right))^{1/p} \\
&= \sqrt{2}\left(\int_\Omega \mathbf{E}\left(|\sum_{j=1}^n \epsilon_j f_j(\omega)|^p\right)\, d\mu(\omega)\right)^{1/p} \\
&\leq \sqrt{2}\left(\int_\Omega (\sum_{j=1}^n |f_j(\omega)|^2)^{p/2}\, d\mu(\omega)\right)^{1/p} \\
&\leq \sqrt{2}\left(\sum_{j=1}^n \left(\int_\Omega |f_j(\omega)|^p\, d\mu(\omega)\right)\right)^{1/p} \\
&= \sqrt{2}(\sum_{j=1}^n \|f_j\|^p)_p^{1/p}.
\end{aligned}
$$

Thus $L^p(\Omega, \Sigma, \mu)$ is of type p.

(ii) Since $L^{p'}(\Omega, \Sigma, \mu)$ is of type p', $L^p(\Omega, \Sigma, \mu)$ is of cotype p, by Proposition 2.8.1. Suppose that $f_1, \ldots, f_n$ are in $L^p(\Omega, \Sigma, \mu)$. To prove the type inequality, we use the fact that the inclusion $L^p \to L^2$ is norm-decreasing, Khintchine's inequality and Corollary 2.8.3.

$$
\left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^2\right)\right)^{1/2} \leq \left(\mathbf{E}\left(\left\|\sum_{j=1}^n \epsilon_j f_j\right\|_p^p\right)\right)^{1/p}
$$

$$
\begin{aligned}
&= \left(\mathbf{E}\left(\int_\Omega |\sum_{j=1}^n \epsilon_j f_j(\omega)|^p \, d\mu(\omega)\right)\right)^{1/p} \\
&= \left(\int_\Omega \mathbf{E}\left(|\sum_{j=1}^n \epsilon_j f_j(\omega)|^p\right) \, d\mu(\omega)\right)^{1/p} \\
&\leq C_p \left(\int_\Omega (\sum_{j=1}^n |f_j(\omega)|^2)^{p/2} \, d\mu(\omega)\right)^{1/p} \\
&\leq C_p \left(\sum_{j=1}^n \left(\int_\Omega |f_j(\omega)|^p \, d\mu(\omega)\right)^{2/p}\right)^{1/2} \\
&= C_p (\sum_{j=1}^n \|f_j\|_p^2)^{1/2}.
\end{aligned}
$$

Thus $L^p(\Omega, \Sigma, \mu)$ is of type 2. □

We now return to the situation of Proposition 2.6.4.

Proposition 2.8.4 *If* $2 \leq p < \infty$*, there exists a constant* c_p *such that*

$$M_r \geq \frac{c_p}{C_p^\gamma(E)} d^{1/p-1/2}.$$

Proof Once again, it is enough to establish the result for A_r. If we knew that

$$\mathbf{E}(\left\|\sum_{i=1}^d \gamma_i e_i\right\|_E) \geq \sqrt{2}(\mathbf{E}(\left\|\sum_{i=1}^d \gamma_i e_i\right\|_E^2))^{1/2}$$

(which follows from Theorem 1.12.2 and de Moivre's central limit theorem), we could use the following argument:

$$
\begin{aligned}
C_p^\gamma(E) A_r &\geq C_p^\gamma(E) c'_p d^{-1/2} \int_{\mathbf{R}^d} \left\|\sum_{i=1}^d a_i e_i\right\|_E d\gamma_d(a) \\
&= C_p^\gamma(E) c'_p d^{-1/2} \mathbf{E}(\left\|\sum_{i=1}^d \gamma_i e_i\right\|_E) \\
&\geq c''_p d^{-1/2} (\sum_{i=1}^d \|e_i\|_E^p)^{1/p} \geq c''_p d^{-1/2} \frac{1}{4} \left(\frac{d}{2}\right)^{1/p} = c_p d^{1/p-1/2}.
\end{aligned}
$$

Instead, we argue as follows, replacing the Gaussian cotype constant by the cotype constant.

$$\begin{aligned}
A_r &= \int_{S^{d-1}} \left\| \sum_{i=1}^d a_i e_i \right\|_E \, d\mu_{d-1}(a) \\
&= \int_{S^{d-1}} \left\| \sum_{i=1}^d \epsilon_i(\omega) a_i e_i \right\|_E \, d\mu_{d-1}(a) \\
&= \int_{S^{d-1}} \mathbf{E}(\left\| \sum_{i=1}^d \epsilon_i a_i e_i \right\|_E) \, d\mu_{d-1}(a) \\
&\geq \frac{1}{\sqrt{2}} \int_{S^{d-1}} (\mathbf{E}(\left\| \sum_{i=1}^d \epsilon_i a_i e_i \right\|_E^2))^{1/2} \, d\mu_{d-1}(a) \\
&\geq \frac{1}{\sqrt{2} C_p(E)} \int_{S^{d-1}} (\sum_{i=1}^d \|a_i e_i\|_E^p)^{1/p} \, d\mu_{d-1}(a) \\
&\geq \frac{1}{4\sqrt{2} C_p(E)} \int_{S^{d-1}} (\sum_{i=1}^{\lfloor d/2 \rfloor} |a_i|^p)^{1/p} \, d\mu_{d-1}(a) \\
&\geq \frac{d^{1/p}}{4\sqrt{2} C_p(E) d^{1/2}} \int_{S^{d-1}} (\sum_{i=1}^{\lfloor d/2 \rfloor} |a_i|^2)^{1/2} \, d\mu_{d-1}(a) \\
&\geq \frac{c_p' d^{1/p}}{C_p(E) d} \int_{\mathbf{R}^d} (\sum_{i=1}^{\lfloor d/2 \rfloor} |x_i|^2)^{1/2} \, d\gamma_d(x) \\
&\geq \frac{c_p d^{1/p}}{C_p(E) d^{1/2}}.
\end{aligned}$$

Here we use the fact that $(\sum_{j=1}^k |a_i|^p)^{1/p} \geq k^{1/p-1/2}(\sum_{j=1}^k |a_i|^2)^{1/2}$, and that if $X_1, \ldots, X_k$ are independent normalized Gaussian random variables and $Z_k = \frac{1}{2}(X_1^2 + \cdots + X_k^2)$ then

$$\int_{\mathbf{R}^d} (\sum_{i=1}^k |x_i|^2)^{1/2} \, d\gamma_d(x) = \sqrt{2}\mathbf{E}(Z_k^{1/2}) = \sqrt{2}\Gamma((k+1)/2)/\Gamma(k/2) \sim \sqrt{k}.$$

□

Corollary 2.8.1 *There exists c_1 such that if $E = l_p^d$ and $1 \leq p < 2$ then $M_r \geq c_1$. If $2 < p < \infty$ there exists c_p such that if $E = l_p^d$ then $M_r \geq c_p d^{1/p-1/2}$.*

Exercises

(i) Show that if E is of type p, for $p > 2$, or cotype p, for $p < 2$, then $E = \{0\}$. Show that if E is of type p then E is of type q, for $1 \leq q < p$, and $T_q(E) \leq T_p(E)$; show that if E is of cotype p then E is of cotype q, for $p < q < \infty$, and $C_q(E) \leq C_p(E)$. Every Banach space is of type 1.

(ii) Is a subspace of type p (or of cotype p) of type p (cotype p)? What about quotient spaces?

(iii) Let $\Omega = \mathbf{Z}_2^{\mathbf{N}}$. Show that there is an isometric embedding of l_1 into $L^\infty(\Omega)$, and show that $L^\infty(\Omega)$ is not of cotype p for $2 \leq p < \infty$, and is not of type p, for $p > 1$.

(iv) Complete the proof of Proposition 2.8.2.

(v) Show that if $p > q > 0$ then $(\sum_{j=1}^k |a_i|^p)^{1/p} \geq k^{1/p-1/q}(\sum_{j=1}^k |a_i|^q)^{1/q}$.

(vi) Let X be a random variable taking values in a finite-dimensional normed space $(E, \|.\|_E)$, with $\mathbf{E}(\|X\|) < \infty$. By considering a norming functional, show that $\|\mathbf{E}(X)\| \leq \mathbf{E}(\|X\|)$.

(vii) Suppose that $\epsilon_1, \ldots, \epsilon_n$ are Bernoulli random variables on a probability space Ω and $g_1, \ldots, g_n$ are independent normalized Gaussian random variables on Ω'. Show that (g_i), $(\epsilon_i g_i)$ and $(\epsilon_i |g_i|)$ have the same distribution.

Suppose that $1 \leq p < \infty$. Show that

$$\begin{aligned}
\left\|\sum_{i=1}^n \epsilon_i(\omega) x_i\right\| &= \sqrt{\frac{\pi}{2}} \left\|\sum_{i=1}^n \epsilon_i(\omega)\mathbf{E}'(|g_i|)x_i\right\| \\
&\leq \sqrt{\frac{\pi}{2}}\mathbf{E}'(\left\|\sum_{i=1}^n \epsilon_i(\omega)|g_i|x_i\right\|) \\
&\leq \sqrt{\frac{\pi}{2}}\mathbf{E}'(\left\|\sum_{i=1}^n \epsilon_i(\omega)|g_i|x_i\right\|^p)^{1/p}.
\end{aligned}$$

Show that

$$\left\|\sum_{i=1}^n \epsilon_i x_i\right\|_p \leq \sqrt{\frac{\pi}{2}} \left\|\sum_{i=1}^n g_i x_i\right\|_p .$$

2.9 Dvoretzky's Theorem, revisited

Theorem 2.9.1 *Suppose that $p \geq 2$. Then there exists a constant $k_p(\delta)$ such that if $k \leq k_p(\delta) d^{2/p}/(C_p^\gamma(E))^2$ then $\mathbf{P}(D_k) \geq 1 - \sqrt{\pi/2}e^{-k_p(\delta)d^{2/p}/(C_p^\gamma(E))^2}$.*

In particular, if $E = l_1^d$ and $k \leq k_2(\delta)d/2$ then $\mathbf{P}(D_k) \geq 1 - \sqrt{\pi/2}e^{-k_2(\delta)d/2}$, and if $p \geq 2$ there exists $k'_p(\delta)$ such that if $k \leq k'_p(\delta)d^{2/p}$ then $\mathbf{P}(D_k) \geq 1 - \sqrt{\pi/2}e^{-k'_p(\delta)d^{2/p}}$.

These results are sharp.

Proposition 2.9.1 *Suppose that $2 < p < \infty$ and E is a k-dimensional subspace of $\mathbf{R}^d$ on which $\|x\|_p \leq |x| \leq L\|x\|_p$. Then $k \leq L^2 C_p^2 d^{2/p}$, where C_p is the constant in Khintchine's inequality.*

Proof Let $(u^{(1)}, \ldots, u^{(k)})$ be an orthonormal basis of E, and let $(\epsilon_1, \ldots, \epsilon_n)$ be Bernoulli random variables. Then

$$k^{1/2} = \left\| \sum_{j=1}^k \epsilon_j(\omega) u^{(j)} \right\|_2 \leq L \left\| \sum_{j=1}^k \epsilon_j(\omega) u^{(j)} \right\|_p,$$

so that

$$k^{p/2} \leq L^p \sum_{i=1}^d |\sum_{j=1}^k \epsilon_j(\omega) u_i^{(j)}|^p,$$

and

$$\begin{aligned} k^{p/2} &\leq L^p \mathbf{E}\left(\sum_{i=1}^d |\sum_{j=1}^k \epsilon_j(\omega) u_i^{(j)}|^p \right) \\ &= L^p \sum_{i=1}^d \mathbf{E}\left(|\sum_{j=1}^k \epsilon_j(\omega) u_i^{(j)}|^p \right) \\ &\leq L^p C_p^p \sum_{i=1}^d \left(\sum_{j=1}^k |u_i^{(j)}|^2 \right)^{p/2} \\ &= L^p C_p^p \sum_{i=1}^d \left(\sum_{j=1}^k |\left\langle u^{(j)}, e_i \right\rangle|^2 \right)^{p/2} \\ &\leq L^p C_p^p d. \end{aligned}$$

Thus $k \leq L^2 C_p^2 d^{2/p}$. □

We deduce a corresponding result for l_∞^d, which shows that we cannot improve the general result in Dvoretzky's theorem. Recall that $C_p \leq 2\sqrt{p}$.

Corollary 2.9.1 *Suppose that E is a k-dimensional subspace of $\mathbf{R}^d$ on which $\|x\|_\infty \leq \|x\|_2 \leq L\|x\|_\infty$. Then $k \leq 4L^2 e^2 \log d$.*

Proof $\|x\|_{\log d} \le \|x\|_2 \le L\,\|x\|_\infty \le L\,\|x\|_{\log d}$, so that

$$k \le L^2 C_{\log d}^2 d^{2/\log d} = L^2 e^2 C_{\log d}^2 \le 4L^2 e^2 \log d.$$

□

2.10 The Kashin decomposition

It follows from Dvoretzky's theorem that for $\delta > 0$ there exists a constant c_δ such that if $(E, \|.\|_E)$ is a d-dimensional normed space then for $k \le c_\delta C_2(E) d$ there are many k-dimensional subspaces F of E with $d(F, l_2^k) \le 1 + \delta$. For small δ, c_δ must be small. But what if we want k to be big, and are prepared for δ to be quite big?

We consider l_1^{2k}, with unit ball B. Then the ellipsoid of maximal volume contained in B is $D = \{x \in R^{2k}; \|x\|_2 \le \sqrt{2k}\}$. Now the volume of B is $2^{2k}/(2k)!$ (it is the union of 2^{2k} hyper-quadrants) and, as we have seen, the volume of D is $\pi^k/(2k)^k k!$, so that

$$\left(\frac{\text{vol } B}{\text{vol } D}\right)^{1/2k} = \left(\frac{8}{\pi}\right)^{1/2} \left(\frac{k^k k!}{(2k)!}\right)^{1/2k} \le \left(\frac{2e}{\pi}\right)^{1/2},$$

a bound that does not depend on the dimension $2k$.

Generally, if $(E, \|.\|_E)$ is a d-dimensional normed space, and D is the ellipsoid of maximal volume contained in B_E, we define the *volume ratio* $\text{vr}(E)$ as

$$\text{vr}(E) = \left(\frac{\text{vol } (B_E)}{\text{vol } (D)}\right)^{1/d}.$$

Theorem 2.10.1 *Suppose that $(E, \|.\|_E)$ is a $2k$-dimensional normed space, with volume ratio R. Let $|.|$ be the inner-product norm defined by the ellipsoid of maximum volume contained in B_E. If*

$$F_k = \{F \in G_{2k,k} : |x| \le 108R^2 \|x\| \text{ for } x \in F\},$$

then $\mathbf{P}(F_k) \ge 1 - 1/2^{2k}$.

Proof We can suppose that D is the unit ball of l_2^{2k}. Then

$$\text{vol } B_E = A_{2k-1} \int_{S^{2k-1}} \left(\int_0^{1/\|\theta\|} u^{2k-1}\, du \right) d\mu(\theta) = V_{2k} \int_{S^{2k-1}} \|\theta\|^{-2k}\, d\mu(\theta).$$

Let $L_k = \{F \in G_{2k,k} : \int_{S_F} \|\theta\|^{-2k}\, d\mu(\theta) > (2R)^{2k}\}$. Then

$$R^{2k} = \int_{S^{2k-1}} \|\theta\|^{-2k}\, d\mu(\theta) = \int_{G_{2k,k}} \left(\int_{S_F} \|\theta\|^{-2k}\, d\mu(\theta) \right) d\mathbf{P}(F) \geq (2R)^{2k}\mathbf{P}(L_k).$$

so that $\mathbf{P}(L_k) \leq 2^{-2k}$.

Let $t = 1/36R^2$, so that $t/9 = (2tR)^2$. Let

$$b_F = \{\theta \in S_F : \|\theta\| < t\} = \{\theta \in S_F : \|\theta\|^{-2k} > 1/t^{2k}\}.$$

Suppose that $F \notin L_k$. Then

$$\frac{\mu(b_F)}{t^{2k}} \leq \int_{S_F} \|\theta\|^{-2k}\, d\mu(\theta) \leq (2R)^{2k},$$

so that $\mu(b_F) \leq (2Rt)^{2k} \leq (t/9)^k$. Now by Proposition 2.7.1 there exists a $t/3$-net N in S_F with less than $(9/t)^k$ points. Then since the balls $\{B_{t/3}(n) : n \in N\}$ cover S_F, $\mu(B_{t/3}(n) \cap S_F) > (t/9)^k$. Thus if $n \in N$ then $B_{t/3}(n) \cap S_F$ is not contained in b_F, and so there exists $y \in B_{t/3}(n) \cap S_F$ with $\|y\| \geq t$. Consequently, for each $x \in S_F$ there exists $y \in S_F$ with $\|y\| \geq t$ and $\|x - y\| \leq |x - y| \leq 2t/3$. Thus $\|x\| \geq t/3 = 1/108R^2$ for all $x \in S_F$. □

Corollary 2.10.1 *Let $J_k = \{F \in G_{2k,k} : F \in F_k$ and $F^{\perp} \in F_k\}$. Then $\mathbf{P}(J_k) \geq 1 - 1/2^{2k-1}$.*

Exercise

Investigate the extent to which you can reduce the constant 108 in Theorem 2.10.1 if you only require that $\mathbf{P}(F_k) > 1/2$ and $\mathbf{P}(J_k) > 0$.

3
Concentration of measure

We now consider the problem of determining how a probability measure on a metric space is concentrated, when we do not have an isoperimetric theorem to help us.

Throughout, we shall suppose that $\mathbf{P}$ is a probability measure on the Borel sets of a complete separable metric space (X, d).

3.1 Martingales

Many of the results for independent random variables, such as Khintchine's inequality, can be extended to martingales. Suppose that $(\Omega, \Sigma, \mathbf{P})$ is a probability space and that Σ_0 is a sub-σ-field of Σ. Then if f is non-negative and Σ-measurable, there exists a unique non-negative Σ_0-measurable function g (possibly taking the value ∞) such that $\int_A g\, d\mathbf{P} = \int_A f\, d\mathbf{P}$ for all $A \in \Sigma_0$. g is the *conditional expectation* of f, and is denoted $\mathbf{E}(f|\Sigma_0)$. In particular, $\mathbf{E}(\mathbf{E}(f|\Sigma_0)) = \mathbf{E}(f)$. If h is a non-negative Σ_0-measurable function then $\mathbf{E}(hf|\Sigma_0) = h\mathbf{E}(f|\Sigma_0)$. If $\Sigma_0 \subseteq \Sigma_1 \subseteq \Sigma$ then $\mathbf{E}(f|\Sigma_0) = \mathbf{E}(\mathbf{E}(f|\Sigma_1)|\Sigma_0)$. If $f \in L^1(\Omega, \Sigma, \mathbf{P})$, we can define $\mathbf{E}(f|\Sigma_0) = \mathbf{E}(f^+|\Sigma_0) - \mathbf{E}(f^-|\Sigma_0)$, and $\mathbf{E}(\cdot|\Sigma_0)$ is a norm-decreasing projection: $L^1(\Sigma) \to L^1(\Sigma_0)$. Its restriction to $L^2(\Sigma)$ is the orthogonal projection of $L^2(\Sigma)$ onto $L^2(\Sigma_0)$.

Suppose that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_k$ is a filtration of σ-fields of Borel sets of a metric probability space $(X, d, \mathbf{P})$. We assume that $\mathcal{F}_0$ is the trivial field $\{X, \emptyset\}$. If $f \in L^1(\mathcal{F}_k)$, let $f_j = \mathbf{E}(f|\mathcal{F}_j)$. We suppose that $\mathbf{E}(f) = f_0 = 0$. Then $(f_j)_{j=0}^k$ is a martingale. Let $d_j = f_j - f_{j-1}$, for $1 \le j \le k$: (d_j) is a *martingale difference sequence.* Note that $\mathbf{E}(d_j|\mathcal{F}_{j-1}) = 0$.

Recall that a random variable X is *sub-Gaussian, with exponent* b, if $\mathbf{E}(e^{tX}) \le e^{b^2t^2/2}$ for $-\infty < t < \infty$.

Theorem 3.1.1 (Azuma's theorem) *Suppose that $(f_j)_{j=0}^k$ is a martingale* *with $f_0 = 0$, and that $\mathbf{E}(e^{td_j}|\mathcal{F}_{j-1}) \leq e^{b_j^2t^2/2}$ for each t and j, where b_j is a constant. Then $f = f_k$ is sub-Gaussian, with exponent $(\sum_{j=1}^k b_j^2)^{1/2}$.*

Proof

$$\begin{aligned}\mathbf{E}(e^{tf_k}) &= \mathbf{E}(\mathbf{E}(e^{tf_{k-1}}e^{td_k}|\mathcal{F}_{k-1})) \\ &= \mathbf{E}(e^{tf_{k-1}}\mathbf{E}(e^{td_k}|\mathcal{F}_{k-1})) \\ &\leq \mathbf{E}(e^{tf_{k-1}})e^{b_k^2t^2/2},\end{aligned}$$

and so

$$\mathbf{E}(e^{tf}) \leq \mathbf{E}(e^{tf_0})\prod_{j=1}^k e^{b_j^2t^2/2} = e^{(\sum_{j=1}^k b_j^2)t^2/2}.$$

[Note that all the functions that we consider are integrable.] □

This theorem holds in particular when each d_j is bounded. Then $f = f_k$ is sub-Gaussian, with exponent $(\sum_{j=1}^k \|d_j\|_\infty^2)^{1/2}$.

We shall apply Azuma's theorem to finite sets, equipped with the uniform probability. The problem is then to find a good metric and good filtration.

Suppose that (X, d) is a finite metric space, and that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_k$ is a filtration, that $\mathcal{F}_0$ is the trivial field $\{X, \emptyset\}$ and that $\mathcal{F}_k = P(X)$. Suppose that for each j there exists a_j such that if A_j and B_j are two $\mathcal{F}_j$ atoms inside the same $\mathcal{F}_{j-1}$ atom C_{j-1} then there is a bijection $\phi : A_j \to B_j$ such that $d(x, \phi(x)) \leq a_j$. Then we say that (X, d) has length at most $(\sum_{j=1}^k a_j^2)^{1/2}$.

Example 1. The hypercube $\mathbf{Q}^d$, with the Hamming metric. Let $\mathcal{F}_j$ be the partition defined by fixing the first j terms. If C_{j-1} is an $\mathcal{F}_{j-1}$ atom, then $C_{j-1} = A_j \cup B_j$, where A_j and B_j are $\mathcal{F}_j$ atoms. Let ϕ be the mapping $A_j \to B_j$ such that $(\phi(x))_i = x_i$, for $i \neq j$. Then $d(x, \phi(x)) = 1$, and so $\mathbf{Q}^d$ has length at most $\sqrt{d}$.

Example 2. Σ_n, the group of permutations of $\{1, \ldots, n\}$. We define a metric by setting $d(\sigma, \rho) = 1$ if $\sigma\rho^{-1}$ is a transposition, and define d to be the path length. Thus Σ_n has diameter $n - 1$, since every permutation can be written as the product of at most $n - 1$ transpositions. Let $\mathcal{F}_j$ be the partition defined by fixing $\sigma(1), \ldots, \sigma(j)$. If C_{j-1} is an $\mathcal{F}_{j-1}$ atom, then C_{j-1} is the union of $n - j + 1$ $\mathcal{F}_j$ atoms, each determined by the value of $\sigma(j)$. If

$$\begin{aligned} A_j &= \{\sigma : \sigma(1) = i_1, \ldots, \sigma(j-1] = i_{j-1}, \sigma_j = k\}, \\ \text{and } B_j &= \{\sigma : \sigma(1) = i_1, \ldots, \sigma(j-1] = i_{j-1}, \sigma_j = l\},\end{aligned}$$

then $B_j = (k,l)A_j$: set $\phi(\sigma) = (k,l)\sigma$. Then $d(\sigma, \phi(\sigma)) = 1$, and so the length of Σ_n is at most $\sqrt{n-1}$.

Note that both of these examples concern a finite group G and a set of generators of G.

Theorem 3.1.2 *Suppose that (X, d) is a finite metric space, equipped with the uniform probability distribution, and of length l. If f is a 1-Lipschitz function on X with $\mathbf{E}(f) = 0$ then f is sub-Gaussian, with exponent l.*

Proof Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_k$ be a filtration which realizes the length. Let C_{j-1} be a $\mathcal{F}_{j-1}$ atom, let A_j and B_j be atoms contained in C_{j-1} and let $\phi : A_j \to B_j$ be a suitable mapping. f_j is constant on each of A_j and B_j: let $f_j(A_j)$ and $f_j(B_j)$ be the values. Then

$$f_j(A_j) - f_j(B_j) = \frac{1}{|A_j|} \sum_{x \in A_j} f(x) - \frac{1}{|B_j|} \sum_{x \in b_j} f(x) = \frac{1}{|A_j|} \sum_{x \in A_j} (f(x) - \phi(x)) \leq a_j.$$

Thus $|f_j(A_j) - f_j(B_j)| \leq a_j$. Since $f_{j-1}(C_{j-1})$ is the average of the $f_j(A_j)$s, it follows that $|d_j| \leq a_j$. We now apply Azuma's theorem to obtain the result. □

Corollary 3.1.1 $\mathbf{P}(f > \epsilon) \leq e^{-\epsilon^2/2l^2}$.

3.2 Sub-Gaussian measures

Let

$$\text{Lip}_0(X) = \{f \in \text{Lip}\,(X) : f \text{ integrable and } \mathbf{E}(f) = 0\}.$$

We define the *Laplace functional* $E_{\mathbf{P}}$ of $\mathbf{P}$ as

$$E_{\mathbf{P}}(t) = \sup\{\mathbf{E}(e^{tf}) : f \in \text{Lip}_0(X), \|f\|_{\text{Lip}} \leq 1\}, \text{ for } t \in \mathbf{R}.$$

We say that $\mathbf{P}$ is sub-Gaussian, with exponent b, if $E_{\mathbf{P}}(t) \leq e^{b^2t^2/2}$, for all $t \in \mathbf{R}$.

Proposition 3.2.1 *If $\mathbf{P}$ is sub-Gaussian with exponent b then Lip $_0(X) \subseteq L_{\exp^2}$, and $\|f\|_{L_{\exp^2}} \leq 2b \|f\|_{Lip}$, for $f \in$ Lip $_0(X)$.*

Conversely, if Lip $_0(X) \subseteq L_{\exp^2}$, and $\|f\|_{L_{\exp^2}} \leq 2b \|f\|_{Lip}$, for $f \in$ Lip $_0(X)$ then $\mathbf{P}$ is sub-Gaussian, with exponent $4b$.

Proof The first statement is an immediate consequence of Theorem 1.11.1.

For the converse, if $f \in \mathrm{Lip}_0(X)$ and $\|f\|_{\mathrm{Lip}} \le 1/2b$ then

$$\mathbf{E}(f^{2k}) \le k!\mathbf{E}(e^{f^2}) \le 2(k!).$$

Thus f is sub-Gaussian with exponent 2, by Theorem 1.11.1. Thus $\mathbf{P}$ is sub-Gaussian, with exponent $4b$. □

Proposition 3.2.2 *If* $\mathbf{P}$ *is sub-Gaussian with exponent* b *then the concentration function* α_P *satisfies* $\alpha_{\mathbf{P}}(\epsilon) \le e^{-\epsilon^2/8b^2}$.

Corollary 3.2.1 *If* (X, d) *is a finite metric space, equipped with the uniform probability distribution, and of length* l *then* $\alpha_{\mathbf{P}}(\epsilon) \le e^{-\epsilon^2/8l^2}$.

Proof Suppose that $\mathbf{P}(A) \ge 1/2$. Let $f_A(x) = \min(d(x, A), \epsilon)$. Then $\mathbf{E}(f_A) \le \epsilon/2$, so that

$$1 - \mathbf{P}(A_\epsilon) \le \mathbf{P}(f_A = \epsilon) \le \mathbf{P}(f_A \ge \mathbf{E}(f_A) + \epsilon/2) \le e^{-\epsilon^2/8b^2},$$

since $f_A - \mathbf{E}(f_A)$ is a 1-Lipschitz function in $\mathrm{Lip}_0(X)$. □

Proposition 3.2.3 *Suppose that* $\mathbf{P}$ *is sub-Gaussian on* (X, d), *with exponent* b, *and that* $\mathbf{Q}$ *is sub-Gaussian on* (Y, ρ), *with exponent* c. *Give* $X \times Y$ *the metric* $\tau((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + \rho(y_1, y_2)$. *Then* $\mathbf{P} \times \mathbf{Q}$ *is sub-Gaussian, with exponent* $(b^2 + c^2)^{1/2}$.

Proof Suppose that f is an integrable 1-Lipschitz function in $\mathrm{Lip}_0(X \times Y)$. Let $g(y) = \int_X f(x, y)\, d\mathbf{P}(x)$. Note that g is a 1-Lipschitz function on Y and $\mathbf{E}(g) = 0$. Then

$$\begin{aligned}\int_{X\times Y} e^{tf(x,y)}\, d\mathbf{P}(x)d\mathbf{Q}(y) &= \int_Y e^{tg(y)} \left(\int_X e^{t(f(x,y)-g(y))}\, d\mathbf{P}(x)\right) d\mathbf{Q}(y) \\ &\le \int_Y e^{tg(y)} e^{t^2b^2/2}\, d\mathbf{Q}(y) \le e^{t^2b^2/2} e^{t^2c^2/2}.\end{aligned}$$

□

When is $\mathbf{P}$ sub-Gaussian? We don't have to consider all Lipschitz functions to find out. We consider $X \times X$ with the metric

$$d((x_1, y_1), (x_2, y_2)) = d(x_1, x_2) + d(y_1, y_2).$$

Let

$$\mu_{\mathbf{P}}(d) = \int_{X\times X} d(x, y)\, d\mathbf{P}(x)d\mathbf{P}(y).$$

$\mu_{\mathbf{P}}(d)$ is the mean distance apart of two points of X.

Proposition 3.2.4 *The following are equivalent:*

(i) $\mu_{\mathbf{P}}(d) < \infty$.

(ii) $\int_X d(x, x_0)\, d\mathbf{P}(x) < \infty$ *for some* $x_0 \in X$;

(iii) $\int_X d(x, x_0)\, d\mathbf{P}(x) < \infty$ *for all* $x_0 \in X$;

Proof If (i) holds, then by Fubini's theorem, $\int_X d(x,y)\, d\mathbf{P}(x) < \infty$ for almost all y, and so for some y.

$\int_X d(x,y)\, d\mathbf{P}(x) \leq d(y,z) + \int_X d(x,z)\, d\mathbf{P}(x)$, so that (ii) implies (iii).

Since $d(x,y) \leq d(x,z) + d(y,z)$,

$$\mu_{\mathbf{P}}(d) \leq 2 \int_X d(x,z)\, d\mathbf{P}(x), \text{ for all } z \in X.$$

Thus (iii) implies (i). □

Theorem 3.2.1 $\mathbf{P}$ *is sub-Gaussian if and only if* $\mu_{\mathbf{P}}(d) < \infty$ *and* $d(x,y) - \mu_{\mathbf{P}}(d)$ *is sub-Gaussian on* $(X \times X, \mathbf{P} \times \mathbf{P})$.

Proof The condition is necessary, by Proposition 3.2.3.

Conversely, suppose that $d(x,y) - \mu_{\mathbf{P}}(d)$ is sub-Gaussian on $(X \times X, \mathbf{P} \times \mathbf{P})$, with exponent b. Suppose that f is an integrable 1-Lipschitz function on X with $\mathbf{E}(f) = 0$. Note that $\mathbf{E}(e^{tf}) \geq 1$ for all t, by Jensen's inequality. If $k > 0$ then

$$\|d(x,y)\|_{2k} \leq \|d(x,y) - \mu_{\mathbf{P}}(d)\|_{2k} + \mu_{\mathbf{P}}(d) \leq 2^{k+1} k! b + \mu_{\mathbf{P}}(d)$$

so that

$$\mathbf{E}(d(x,y)^{2k}) \leq (2^{k+1} k! b + \mu_{\mathbf{P}}(d))^{2k} \leq 2^{k+1} k! (b + \mu_{\mathbf{P}}(d)/4)^{2k}.$$

Thus

$$\int_{X \times X} (f(x) - f(y))^{2k}\, d\mathbf{P}(x) d\mathbf{P}(y) \leq 2^{k+1} k! (b + \mu_{\mathbf{P}}(d)/4)^{2k}.$$

Thus it follows from Theorem 1.11.1 that $f(x) - f(y)$ is sub-Gaussian, with exponent $2\sqrt{2}(b + \mu_{\mathbf{P}}(d)/4)$. Thus

$$\begin{aligned} \mathbf{E}(e^{tf}) &\leq \int_X e^{tf(x)}\, d\mathbf{P}(x) \int_X e^{-tf(y)}\, d\mathbf{P}(y) \\ &\leq e^{4t^2 (b + \mu_{\mathbf{P}}(d)/4)^2}. \end{aligned}$$

□

3.3 Convergence of measures

Suppose that Σ is a σ-field of subsets of X. A *signed measure* is a bounded countably additive real-valued function on Σ. The set of signed measures forms a vector space $M(X)$. The Hahn-Jordan decomposition theorem says that if $\mu \in M(X)$ then there exists a partition $X = P \cup N$, with $P, N \in \Sigma$ such that if $A \in \Sigma$ then $\mu(A \cap P) \geq 0$ and $\mu(A \cap N) \leq 0$. Then, setting $\mu^+(A) = \mu(A \cap P)$ and $\mu^-(A) = -\mu(A \cap N)$, μ^+ and μ^- are (non-negative) measures on Σ, and $\mu = \mu^+ - \mu^-$. Set $|\mu| = \mu^+ + \mu^-$, and set $\|\mu\| = \|\mu\|_{TV} = |\mu|(X)$. $\|.\|_{TV}$ is a norm on $M(X)$, the *total variation* norm, under which $M(X)$ is a Banach space, and a Banach lattice.

We are interested in measures on topological spaces, and in particular on compact Hausdorff spaces and metric spaces. We require that (X, τ) is Hausdorff and normal: it then follows from Urysohn's lemma that if A and B are closed and disjoint then there exists a continuous $f : X \to [0, 1]$ with $f(a) = 0$ for $a \in A$ and $f(b) = 1$ for $b \in B$. Note that if $f \in C(X)$ then $(f \leq t) = \int_{n=1}^{\infty}(f < t = 1/n)$ is a closed G_δ set - a countable intersection of open sets. It is therefore natural to consider the *Baire* σ-field - the σ-field generated by the closed G_δ sets: this is the smallest σ- field for which the continuous real-valued functions on X are measurable.

Theorem 3.3.1 *Suppose that μ is a finite Baire measure on a normal Hausdorff space (X, τ). Then μ is* closed G_δ regular*: if A is a Baire set then*

$$\begin{aligned} \mu(A) &= \sup\{\mu(F) : F \textit{ a closed } G_\delta \textit{ set}, F \subseteq A\} \\ &= \inf\{\mu(U) : U \textit{ an open } F_\sigma \textit{ set}, U \supseteq A\}. \end{aligned}$$

[An F_σ set is a countable union of closed sets.]

Proof Let T be the collection of Baire sets for which the result holds. Suppose that $A = \cap_n U_n$ is a closed G_δ set. Then there exist $f_n : X \to [0, 1]$ with $f_n(a) = 0$ for $a \in A$ and $f_n(x) = 1$ for $x \notin U_n$. Then $A \subseteq (f_n < 1) \subseteq U_n$ so that $A = \cap_n(f_n < 1)$; since $(f_n < 1)$ is an open F_σ set, $A \in T$.

It is therefore enough to show that T is a σ-field. Since $A \in T$ if and only if $C(A) \in T$, it is enough to show that if (A_n) is a sequence in T then $A = \cup_n A_n \in T$. Suppose that $\epsilon > 0$. Then for each n there exist $F_n \subseteq A_n \subseteq U_n$ (F_n a closed G_δ set, U_n an open F_σ set) with $\mu(A_n \setminus F_n) < \epsilon/2^{n+1}$ and $\mu(U_n \setminus A_n) < \epsilon/2^n$. Then $U = \cup_n U_n$ is an open F_σ set, and $U \setminus A \subset \cup_n(U_n \setminus A_n)$, so that $\mu(U \setminus A) \leq \sum_n \mu(U_n \setminus A_n) < \epsilon$. Let $B_n = \cup_{i=1}^n A_i$. Then $B_n \nearrow A$, and so there exists N such that $\mu(A \setminus B_N) < \epsilon/2$. Then

$G_N = \cup_{i=1}^N F_j$ is a closed G_δ set, and $B_N \setminus G_N \subseteq \cup_{i=1}^N (A_i \setminus F_i)$ so that $\mu(B_N \setminus G_N) \leq \sum_{i=1}^N \mu(A_i \setminus F_i), \epsilon/2$. Thus $\mu(A \setminus G_N) < \epsilon$. □

We shall be concerned with probability measures on metric spaces (X, d). Note that any closed set A in a metric space (X, d) is a G_δ set, since $A = \{x : d(x, A) < 1/n\}$, and so the Baire sets and the Borel sets are the same. We shall restrict attention to complete separable metric spaces, or *Polish* spaces. These need not be compact, nor even locally compact. Nevertheless, compactness plays an essential role. A probability measure $\mathbf{P}$ is *regular* if

$$\mathbf{P}(A) = \sup\{\mathbf{P}(K) : K \text{ compact }, K \subseteq A\}$$

for each Borel set A. By Theorem 3.3.1, this happens if and only if $\mathbf{P}$ is *tight* - that is $\sup\{\mathbf{P}(K) : K \text{ compact }\} = 1$.

Theorem 3.3.2 *A probability measure* $\mathbf{P}$ *on a Polish space is tight.*

Proof Let (x_n) be a dense sequence in (X, d), and, for $\delta > 0$, let $B_\delta(x_n) = \{x : d(x, x_n) \leq \delta\}$. Suppose that $\epsilon > 0$. The balls $B_{1/m}x_n$ cover X, and so there exists N_m such that, setting $C_m = \cup_{n=1}^{N_m} B_{1/m}(x_n)$, $\mathbf{P}(C_m) > 1 - \epsilon/2^m$. Let $K = \cap_{m=1}^\infty C_m$. Then K is closed and precompact, and so it is compact. Since $\mathbf{P}(X \setminus K) \leq \sum_{m=1}^\infty \epsilon/2^m = \epsilon$, $\mathbf{P}$ is tight. □

Suppose that X is a normal Hausdorff space and that $M(X)$ is the space of signed Baire measures on X. If $f \in C_b(X)$ and $\mu \in M(X)$, let $\phi_\mu(f) = \int_X f\, d\mu$. ϕ is an order preserving isometry of $(M(X), \|.\|_{TV})$ into $C_b(X)^*$:

$$\|\mu\|_{TV} = \sup\{|\int_X f\, d\mu| : f \in C_b(X), \|f\|_\infty = 1\}.$$

Things become easier when (X, τ) is a compact Hausdorff space.

Theorem 3.3.3 (The Riesz representation theorem) *If* (X, τ) *is a compact Hausdorff space then the mapping* ϕ *is surjective.*

The norm topology $\|.\|_{TV}$ is too strong for most purposes: instead, we consider the weak topology, induced by the weak*-topology on $C_b(X)^*$.

Corollary 3.3.1 *If* (X, τ) *is a compact Hausdorff space then the closed unit ball* $M_1(X)$ *is weakly compact, and* $P(X)$ *is a weakly closed, and therefore weakly compact, subset of* $M_1(X)$.

Proof The first statement follows from the Banach-Alaoglu theorem, and the second from the fact that $P(X) = \{\mu : \mu \in M_1(X) : \mu(X) = \int_X 1\,d\mu = 1\}$. $\square$

Suppose now that (X, d) is a Polish space. We show that weakly convergent sequences in $P(X)$ can be defined in terms of a norm. [We do not show that the norm defines the weak topology on $P(X)$.]

If $\mathbf{P}, \mathbf{Q} \in P(X)$, let

$$\beta(\mathbf{P}, \mathbf{Q}) = \sup\{|\int_X f\,d\mathbf{P} - \int_X f\,d\mathbf{Q}| : \|f\|_{BL} \le 1\} = \|\mathbf{P} - \mathbf{Q}\|^*_{BL}.$$

Then β is a metric on $P(X)$.

Theorem 3.3.4 *Suppose that (X, d) is a Polish space. If $C \subset P(X)$ is β-precompact then A is uniformly tight: given $\epsilon > 0$ there exists a compact subset K of X such that $\mathbf{P}(K) > 1 - \epsilon$ for all $\mathbf{P} \in C$.*

We need some lemmas, the first of interest in its own right.

Lemma 3.3.1 *Suppose that $0 < \epsilon < 1$. If $\beta(\mathbf{P}, \mathbf{Q}) \le \epsilon^2/2$ then $\mathbf{Q}(A_\epsilon) \ge \mathbf{P}(A) - \epsilon$, for all closed sets A.*

Proof Let $f(x) = (1 - d(x, A)/\epsilon)^+$. Then $\|f\|_{BL} \le 1 + 1/\epsilon$, so that

$$|\int_X f\,d\mathbf{P} - \int_X f\,d\mathbf{Q}| \le \epsilon^2(1 + 1/\epsilon)/2 < \epsilon.$$

Thus

$$Q(A_\epsilon) \ge \int_X f\,d\mathbf{Q} \ge \int_X f\,d\mathbf{P} - \epsilon \ge \mathbf{P}(A) - \epsilon.$$

$\square$

Lemma 3.3.2 *If $0 < \epsilon < 1$ there exists a finite set F in X such that $\mathbf{Q}(F_\epsilon) > 1 - \epsilon$ for all $\mathbf{Q} \in C$.*

Proof There exists a finite set D in C such that $C \subseteq D_{\epsilon^2/8}$. There exists a compact subset K_D of X such that $\mathbf{P}(K_D) > 1 - \epsilon/2$ for $\mathbf{P} \in D$. There exists a finite subset F of X such that $K_D \subseteq F_{\epsilon/2}$, so that $(K_D)_{\epsilon/2} \subseteq F_\epsilon$. If $\mathbf{Q} \in C$, there exists $\mathbf{P} \in D$ with $\beta(\mathbf{Q}, \mathbf{P}) \le \epsilon^2/8$, and so

$$\mathbf{Q}(F_\epsilon) \ge \mathbf{Q}((K_D)_{\epsilon/2}) \ge \mathbf{P}(K_D) - \epsilon/2 \ge 1 - \epsilon.$$

$\square$

Proof of the Theorem. Suppose that $0 < \epsilon < 1$. Let $\epsilon_n = \epsilon/2^n$. For each n, there exists a finite set F_n such that $\mathbf{Q}((F_n)_{\epsilon_n}) \geq 1 - \epsilon_n$, for $\mathbf{Q} \in C$. Let $K = \cap_{n=1}^{\infty}((F_n)_{\epsilon_n})$. K is compact, and $\mathbf{Q}(K) \geq 1 - \epsilon$ for all $\mathbf{Q} \in C$. □

Theorem 3.3.5 $\mathbf{P}_n \to \mathbf{P}$ *weakly if and only if* $\beta(\mathbf{P}_n, \mathbf{P}) \to 0$ *as* $n \to \infty$.

Proof Suppose first that $\mathbf{P}_n \to \mathbf{P}$ weakly. Suppose that $0 < \epsilon < 1$. By tightness,there exists a compact subset K of X such that $\mathbf{P}(K) > 1 - \epsilon/11$. By the Arzelà-Ascoli theorem, $A = \{f_{|K}; \|f\|_{BL} \leq 1\}$ is precompact in $C(K)$, and so there exist $g_1, \dots g_k \in A$ such that $A \subseteq \cup_{i=1}^{k} B_{\epsilon/11}(g_i)$. By Theorem 1.9.1, each g_i can be extended without increasing the BL norm to $f_i \in BL(X)$. Also, let $h(x) = (1 - 11d(x, K)/\epsilon)^+$. Then there exists N such that

$$|\int_X h\, d\mathbf{P}_n - \int_X h(x)\, d\mathbf{P}| < \epsilon/11 \text{ and } |\int_X f_i\, d\mathbf{P}_n - \int_X f_i(x)\, d\mathbf{P}| < \epsilon/11$$

for $1 \leq i \leq k$ and $n \geq N$. Note that

$$P_n(K_{\epsilon/11}) \geq \int_X h\, d\mathbf{P}_n \geq \int_X h\, d\mathbf{P} - \epsilon/11 \geq 1 - 2\epsilon/11$$

for $n \geq N$. If $\|f\|_{BL} \leq 1$ there exists f_i such that $|f_i(x) - f(x)| \leq \epsilon/11$ for $x \in K$. Using the Lipschitz condition, $|f_i(y) - f(y)| \leq 3\epsilon/11$ for $y \in K_{\epsilon/11}$.

Now

$$\int_X |f - f_i|\, d\mathbf{P}_n \leq \int_{K_{\epsilon/11}} |f - f_i|\, d\mathbf{P}_n + \int_{X \setminus K_{\epsilon/11}} (|f| + |f_i|)\, d\mathbf{P}_n \leq 3\epsilon/11 + 4\epsilon/11$$

and

$$\int_X |f - f_i|\, d\mathbf{P} = \int_K |f - f_i|\, d\mathbf{P} + \int_{X \setminus K} (|f| + |f_i|)\, d\mathbf{P} \leq \epsilon/11 + 2\epsilon/11$$

Thus if $n \geq N$ then

$$\begin{aligned} &|\int_X f\, d\mathbf{P}_n - \int_X f\, d\mathbf{P}| \leq \\ &\leq \; |\int_X f_i\, d\mathbf{P}_n - \int_X f_i\, d\mathbf{P}| + \int_X |f - f_i|\, d\mathbf{P}_n + \int_X |f - f_i|\, d\mathbf{P} \\ &\leq \; \epsilon/11 + 7\epsilon/11 + 3\epsilon/11 = \epsilon, \end{aligned}$$

and so $\beta(\mathbf{P}_n, \mathbf{P}) \leq \epsilon$ for $n \geq N$.

Conversely, suppose that $\beta(\mathbf{P}_n, \mathbf{P}) \to 0$. Then $\{P_n : n \in \mathbf{N}\}$ is β-precompact, and so is uniformly tight. Thus given $\epsilon > 0$ there exists a compact subset K of X such that $\mathbf{P}(K) \geq 1 - \epsilon/7$ and $\mathbf{P}_n(K) \geq 1 - \epsilon/7$ for all n. Suppose that $f \in C_b(X)$ and that $\|f\|_\infty \leq 1$. By the Stone-Weierstrass theorem, there exists $g \in BL(K)$ such that $\|g\|_{C(K)} \leq 1$ and

$\|f-g\|_{C(K)} \le \epsilon/7$. We can extend g to $h \in BL(X)$, with $\|h\|_{C(X)} \le 1$ Then

$$|\int_X f\,d\mathbf{P}_n - \int_X f\,d\mathbf{P}| \le |\int_X h\,d\mathbf{P}_n - \int_X h\,d\mathbf{P}| + \int_X |f-h|\,d\mathbf{P}_n + \int_X |f-h|\,d\mathbf{P}.$$

Now

$$\int_X |f-h|\,d\mathbf{P}_n \le \int_K |f-h|\,d\mathbf{P}_n + \int_{C(K)} |f|\,d\mathbf{P}_n + \int_{C(K)} |h|\,d\mathbf{P}_n \le 3\epsilon/7,$$

and similarly $\int_X |f-h|\,d\mathbf{P} \le 3\epsilon/7$, and so

$$|\int_X f\,d\mathbf{P}_n - \int_X f\,d\mathbf{P}| \le |\int_X h\,d\mathbf{P}_n - \int_X h\,d\mathbf{P}| + 6\epsilon/7 < \epsilon$$

for large enough n. □

Theorem 3.3.6 $(P(X), \beta)$ *is complete.*

Proof Suppose that $(\mathbf{P}_n)$ is a β-Cauchy sequence. Then $\{\mathbf{P}_n : n \in \mathbf{N}\}$ is β-precompact, and so is uniformly tight. Thus there exists an increasing sequence (K_j) of compact subsets of X such that $P_n(K_j) > 1/2^j$ for all n and j. Let $\mathbf{P}_n^{(j)}$ be the restriction of $\mathbf{P}_n$ to the Borel sets of K_j. Then, taking a subsequence if you must (but it isn't necessary), $\mathbf{P}_n^{(j)}$ converges weakly to some $\mu_j \in M(K_j)$. Then if A is a Borel set contained in K_j, $\mu_k(A) = \mu_j(A)$ for $k \ge j$. If A is any Borel set in X, let $\mathbf{P}(A) = \lim_{j\to\infty} \mu_j(A)$. Then it is straightforward to verify that $\mathbf{P}$ is a measure, that $\mathbf{P}(X) = 1$ and that $\mathbf{P}_n \to \mathbf{P}$. □

Exercise

(i) Suppose that $A \subseteq P(X)$ is uniformly tight. Show that A is β-precompact.
(ii) Give the full details of Theorem 3.3.6.

3.4 The transportation problem

Suppose that (X, τ) and (Y, ρ) are compact Hausdorff spaces, that $\mathbf{P}$ and $\mathbf{Q}$ are Baire probability measures on X and Y respectively, and that $c \in C^+(X \times Y)$. $c(x, y)$ is the cost of transportation from x to y. We want to transport a mass distributed as $\mathbf{P}$ to a mass distributed as $\mathbf{Q}$ at minimal cost. This is the *Monge transportation problem.*

Let π be a Baire probability measure on $X \times Y$. π defines a transport plan: $\pi(A, B)$ denotes the amount of matter transported from A to B. π must have marginal distributions $\mathbf{P}$ and $\mathbf{Q}$: $\pi(A \times Y) = \mathbf{P}(A)$ and $\pi(X \times B) = \mathbf{Q}(B)$. The cost is then $\int_{X \times Y} c\, d\pi$, and we want to choose π to minimize this. Thus we have a constrained optimization problem. When X and Y are finite, this is a classical linear programming problem. But we, like Monge, want to consider the more general problem.

Theorem 3.4.1 (The Kantorovitch-Rubinstein theorem I) *Suppose that (X, τ) and $(Y \rho)$ are compact Hausdorff spaces, that $\mathbf{P}$ and $\mathbf{Q}$ are Baire probability measures on X and Y respectively, and that $c \in C^+(X \times Y)$. Then there is a Baire probability measure π on $X \times Y$ with marginals $\mathbf{P}$ and $\mathbf{Q}$ which minimises $\int_{X \times Y} c\, d\pi$ under these constraints. Let*

$$m_c(\mathbf{P}, \mathbf{Q}) = \sup\{\int_X f\, d\mathbf{P} + \int_Y g\, d\mathbf{Q} : f \in C(X),\, g \in C(Y),\, f(x) + g(y) \leq c(x, y)\}.$$

Then $\int_{X \times Y} c\, d\pi = m_c(\mathbf{P}, \mathbf{Q})$.

Proof If π has marginals $\mathbf{P}$ and $\mathbf{Q}$, and $f(x) + g(y) \leq c(x, y)$ then

$$\int_X f\, d\mathbf{P} + \int_Y g\, d\mathbf{Q} = \int_{X \times Y} f(x) + g(y)\, d\pi(x, y) \leq \int_{X \times Y} c\, d\pi,$$

so that $m_c(\mathbf{P}, \mathbf{Q}) \leq \int_{X \times Y} c\, d\pi$. We need to find π for which the reverse inequality holds.

Let $L = \{f(x) + g(y) : f \in C(X), g \in C(Y)\} \subseteq C(X \times Y)$. L is a linear subspace of $C(X \times Y)$. If $f + g \in L$ let $\phi(f + g) = \int_X f\, d\mathbf{P} + \int_Y\, d\mathbf{Q}$. This is a well-defined linear functional on L, since if $f + g = f' + g'$ then $f - f' = g' - g$ is a constant k, so that $\int_X f'\, d\mathbf{P} = \int_X f\, d\mathbf{P} - k$ and $\int_Y g'\, d\mathbf{Q} = \int_Y g\, d\mathbf{Q} + k$. Further, $\phi(1) = 1$, so that ϕ is non-zero.

Now let

$$U = \{h \in C(X \times Y) : h(x, y) < c(x, y) \text{ for all } (x, y) \in X \times Y\}.$$

U is a non-empty convex open subset of $C(X \times Y)$. $U \cap L$ is also non-empty.

If $f+g \in U \cap L$ then

$$\begin{aligned}\phi(f+g) &\leq \sup\{f(x): x \in X\} + \sup\{g(y): y \in Y\} \\ &< \sup\{c(x,y): (x,y) \in X \times Y\},\end{aligned}$$

so that ϕ is bounded above on $U \cap L$. Let $M = \sup\{\phi(h): h \in U \cap L\}$, and let $B = \{l \in L : \phi(l) \geq M\}$. Then B is a non-empty convex set disjoint from U. By the Hahn-Banach theorem, there exists a non-zero continuous linear functional ψ on $C(X \times Y)$ such that if $h \in U$ then $\psi(h) < K = \inf\{\psi(b) : b \in B\}$. If $h > 0$ then $-\alpha h \in U$ for all sufficiently large α, and so $\psi(h) \geq 0$. In particular, $\psi(1) \geq 0$. But since $\psi \neq 0$, $\psi(1) > 0$. Let $\theta = \psi/\psi(1)$. θ is a non-negative linear functional on $C(X \times Y)$, and $\theta(1) = 1$; by the Riesz representation theorem θ is represented by a Baire probability measure π on $X \times Y$. We shall show that π has the required properties.

Note that if $h \in U$, $\theta(h) < \Lambda = \inf\{\theta(b) : b \in B\}$. If $l_0 \in L$ and $\phi(l_0) = 0$, then $\phi(M.1 + \alpha l_0) = M$, so that $M.1 + \alpha l_0 \in B$ for all α, and so $\theta(M.1 + \alpha l_0) = M + \alpha\theta(l_0) \geq \Lambda$ for all α, and so $\theta(l) = 0$. If $l \in L$ then $l = \phi(l)1 + l_0$, where $\phi(l_0) = 0$, and so $\theta(l) = \phi(l)$: θ extends ϕ. In particular, this means that $\Lambda = M$. If $f \in C(X)$ then $\int_{X\times Y} f(x)\,d\pi(x,y) = \phi(f) = \int_X f\,d\mathbf{P}$, and a similar result holds for $g \in C(Y)$; thus π has marginals $\mathbf{P}$ and $\mathbf{Q}$. Finally,

$$\begin{aligned}m_c(\mathbf{P},\mathbf{Q}) &= \sup\{\int_X f\,d\mathbf{P} + \int_Y g\,d\mathbf{Q} : f(x) + g(y) \leq c(x,y)\} \\ &= \sup\{\phi(h) : h \in U \cap L\} = M,\end{aligned}$$

and $\int c\,d\pi = \sup\{\int h\,d\pi : h \in U\} \leq M$.

□

3.5 The Wasserstein metric

Recall that $\mu_{\mathbf{P}}(d) = \int_{X\times X} d(x,y)\,d\mathbf{P}(x)\,d\mathbf{P}(y)$; $\mu_{\mathbf{P}}(d)$ is the *mean distance* on X. We restrict attention to $P_1(x) = \{\mathbf{P} \in P(X) : \mu_{\mathbf{P}}(d) < \infty\}$. If $\mathbf{P}, \mathbf{Q} \in P_1(X)$, let

$$\gamma(\mathbf{P},\mathbf{Q}) = \sup\{|\int_X f\,d\mathbf{P} - \int_X f\,d\mathbf{Q}| : f \in \text{Lip}_{\ X}, \|f\|_L \leq 1\}.$$

γ is a metric on $P_1(X)$, and $\gamma \geq \beta_{|P_1(X)}$.

If $\mathbf{P}, \mathbf{Q} \in P_1(X)$ let

$$m_L(\mathbf{P},\mathbf{Q}) = \sup\{\int_X f\,d\mathbf{P} + \int_X g\,d\mathbf{P} : f, g \in \text{Lip }(X) : f(x) + g(y) \leq d(x,y)\}.$$

Note that if (X, d) is compact, then $m_L = m_d$, the metric obtained by taking d as the cost function.

Proposition 3.5.1 *If* $\mathbf{P}, \mathbf{Q} \in P_1(X)$ *then* $\gamma(\mathbf{P}, \mathbf{Q}) = m_L(\mathbf{P}, \mathbf{Q})$.

Proof Setting $g = -f$, we see that $\gamma(\mathbf{P}, \mathbf{Q}) \leq m_L(\mathbf{P}, \mathbf{Q})$.

Conversely, suppose that $f, g \in \mathrm{Lip}\ (X)$ and $f(x) + g(y) \leq d(x, y)$. Let

$$h(x) = \inf\{d(x, y) - g(y) : y \in X\}.$$

Then $h(x) - h(z) \leq d(x, z)$, so that $\|h\|_L \leq 1$. Also $f(x) \leq h(x) \leq -g(x)$, so that

$$\int_X f\, d\mathbf{P} + \int_X g\, d\mathbf{Q} \leq \int_X h\, d\mathbf{P} - \int_X h\, d\mathbf{Q}.$$

Thus $\gamma(\mathbf{P}, \mathbf{Q}) \geq m_L(\mathbf{P}, \mathbf{Q})$. □

Corollary 3.5.1 *If* (X, d) *is compact then* $\gamma(\mathbf{P}, \mathbf{Q}) = m_d(\mathbf{P}, \mathbf{Q})$.

We now define, for $\mathbf{P}, \mathbf{Q} \in P_1(X)$,

$$W(\mathbf{P}, \mathbf{Q}) = \inf\{\int_{X \times X} d(x, y)\, d\pi(x, y) : \pi \in P_1(X, Y) \text{ with marginals } \mathbf{P} \text{ and } \mathbf{Q}\}.$$

Theorem 3.5.1 (The Kantorovitch-Rubinstein theorem II) $W = \gamma$.

Proof If π has marginals $\mathbf{P}$ and $\mathbf{Q}$ and if $\|f\|_L \leq 1$ then

$$\int_X f\, d\mathbf{P} - \int_X f\, d\mathbf{Q} = \int_{X \times X} f(x) - f(y)\, d\pi(x, y) \leq \int_{X \times X} d(x, y)\, d\pi(x, y),$$

so that $\gamma(\mathbf{P}, \mathbf{Q}) \leq W(\mathbf{P}, \mathbf{Q})$.

Conversely, pick a base point $x_0 \in X$, and let $r(x) = d(x, x_0)$.

Now $rd\mathbf{P}$ and $rd\mathbf{Q}$ are bounded measures on (X, d) and are therefore tight. Thus there exists an increasing sequence (K_n) of compact subsets of X containing x_0 such that

$$\mathbf{P}(K_n) > 1/n,\ \mathbf{Q}(K_n) > 1/n,\ \int_{X \setminus K_n} r\, d\mathbf{P} < 1/n \text{ and } \int_{X \setminus K_n} r\, d\mathbf{Q} < 1/n.$$

If A is a Borel subset of K_n, let

$$\begin{aligned} \mathbf{P}_n(A) &= \mathbf{P}(A) + (1 - \mu(K_n))\delta_{x_0}(A) \\ \mathbf{Q}_n(A) &= \mathbf{Q}(A) + (1 - \mu(K_n))\delta_{x_0}(A). \end{aligned}$$

Then by the compact Kantorovitch-Rubinstein theorem there exists a measure π_n on $K_n \times K_n$ with marginals $\mathbf{P}_n$ and $\mathbf{Q}_n$ such that $m_d(\mathbf{P}_n, \mathbf{Q}_n) =$

$W(\mathbf{P}_n, \mathbf{Q}_n)$. We can extend $\mathbf{P}_n$, $\mathbf{Q}_n$ to probability measures on X and π_n to a probability measure on $X \times X$ in the obvious way. Then the sequence (π_n) is uniformly tight on $X \times X$ and so it is weakly relatively compact. Thus there is a subsequence (π_{n_k}) which converges weakly to a probability π. The marginals $\mathbf{P}_{n_k}$ and $\mathbf{Q}_{n_k}$ then converge weakly to $\mathbf{P}$ and $\mathbf{Q}$ respectively.

Now let $I = \int_{X\times X} d(x,y)\, d\pi(x,y)$ and suppose that $\epsilon > 0$. Let $d_j(x,y) = \min(d(x,y), j)$. d_j is a bounded metric on X equivalent to d, and $I_j = \int_{X\times X} d_j\, d\pi \nearrow I$. Thus there exists J such that $I_j \geq I - \epsilon/4$ for $j \geq J$. But $\int_{X\times X} d_J\, d\pi_{n_k} \to I_J$ as $k \to \infty$, and so there exists K such that

$$\int_{X\times X} d(x,y)\, d\pi_{n_k} \geq \int_{X\times X} d_J(x,y)\, d\pi_{n_k} \geq I - \epsilon/2$$

for $k \geq K$. Now suppose that $k \geq K$. There exists $f \in \mathrm{Lip}\,(K_{n_k})$ with $\|f\|_L \leq 1$ and $f(x_0) = 0$ such that

$$\int_{K_{n_k}} f\, d\mathbf{P}_{n_k} - \int_{K_{n_k}} f\, d\mathbf{Q}_{n_k} > \int_{X\times X} d(x,y)\, d\pi_{n_k} - \epsilon/4 \geq I - 3\epsilon/4.$$

By Theorem 1.9.1, we can extend $f \to X$ without increasing the Lipschitz norm. Then $|f(x)| \leq r(x)$ for all $x \in X$. Thus

$$\begin{aligned} |(\int_X f\, d\mathbf{P} - \int_X f\, d\mathbf{Q}) - (\int_X f\, d\mathbf{P}_{n_k} - \int_X f\, d\mathbf{Q}_{n_k})| &= \\ = \;& |\int_{X\setminus K_{n_k}} f\, d\mathbf{P} - \int_{X\setminus K_{n_k}} f\, d\mathbf{Q}| \\ \leq \;& \int_{X\setminus K_{n_k}} r\, d\mathbf{P} + \int_{X\setminus K_{n_k}} r\, d\mathbf{Q} \\ \leq \;& 2/n_k < \epsilon/4 \end{aligned}$$

for large enough k. Thus

$$\int_X f\, d\mathbf{P} - \int_X f\, d\mathbf{Q} \geq I - \epsilon,$$

and so $W(\mathbf{P}, \mathbf{Q}) \leq \gamma(\mathbf{P}, \mathbf{Q})$. □

The metric W is called the Wasserstein metric.

Exercises

3.1 Suppose that $x, y \in (X, d)$. Calculate $\|\delta_x - \delta_y\|_{TV}$, $\beta(\delta_x, \delta_y)$ and $W(\delta_x, \delta_y)$.

3.2 On $\mathbf{R}$, let $\mathbf{P}_n = (1-1/n)\delta_0 + \delta_n/n$. Calculate $\|\mathbf{P}_n - \delta_0\|_{TV}$, $\beta(\mathbf{P}_n, \delta_0)$ and $W(\mathbf{P}_n, \delta_0)$.

3.6 Entropy

We now turn to entropy, which provides a measure of the spread of values of a function. If $f \geq 0$ and $f \in L^1$ we define the *entropy* $\mathrm{Ent}(f)$ as

$$\mathrm{Ent}(f) = \mathbf{E}(f \log f) - \|f\|_1 \log \|f\|_1 = \mathbf{E}(f \log(f/\|f\|_1)).$$

Note the following:

- We take $0 \log 0 = 0$, since $x \log x \to 0$ as $x \to 0$.
- If $\|f\|_1 = 1$ then $\mathrm{Ent}(f) = \mathbf{E}(f \log f)$.
- $\mathrm{Ent}(\alpha f) = \alpha \mathrm{Ent}(f)$ for $\alpha > 0$.
- By Jensen's inequality, $\mathrm{Ent} f \geq 0$, and $\mathrm{Ent} f = 0$ if and only if f is constant.

We have the following inequality, known as the *entropy inequality*.

Proposition 3.6.1 *Suppose that $f \geq 0$, $f \in L^1(\mu)$ and $fg^- \in L^1(\mu)$. Then*

$$\int fg \, d\mu \leq \|f\|_1 \log(\int e^g \, d\mu) + \mathit{Ent}(f).$$

Proof The condition on g ensures that $\int fg \, d\mu$ exists, taking values in $(-\infty, \infty]$. By homogeneity, we can suppose that $\|f\|_1 = 1$. Then

$$\begin{aligned} \int fg \, d\mu &= \int_{f>0} f \log f + f \log \left(\frac{e^g}{f} \right) d\mu \\ &= \mathrm{Ent}(f) + \int_{f>0} \log \left(\frac{e^g}{f} \right) f d\mu. \end{aligned}$$

Since $f d\mu$ is a probability measure and log is concave, the result follows from Jensen's inequality. □

Corollary 3.6.1 *If $f \geq 0$ and $f \in L^1(\mu)$ then*

$$\mathit{Ent}(f) = \sup\{\int fg \, d\mu : fg^- \in L^1(\mu), \int e^g \, d\mu = 1\}.$$

Proof The proposition implies that

$$\mathrm{Ent}(f) \geq \sup\{\int fg \, d\mu : fg^- \in L^1(\mu), \int e^g \, d\mu = 1\}.$$

But setting $g = \log(f/\|f\|_1)\chi_{(f>0)}$,

$$\int e^g = \int (f/\|f\|_1) \, d\mu \text{ and } \int fg \, d\mu = \mathrm{Ent} f,$$

and so we obtain equality. □

Where does this definition come from? In information theory, one considers a probability $\mathbf{P}$ on a finite set I of size n, and if $\mathbf{P}(\{i\}) = p_i$, defines the information entropy as $\mathrm{ent}(\mathbf{P}) = \sum_I p_i \log(1/p_i) = -\sum_I p_i \log(p_i)$. Let π be the uniform probability on I, so that $\pi(\{i\}) = 1/n$, and let $f_i = np_i$. Then

$$\begin{aligned}
\mathrm{Ent}_\pi(f) &= \sum_{i\in I}(np_i \log(np_i)/n \\
&= \sum_{i\in I} p_i \log p_i + \log n = \mathrm{ent}(\pi) - \mathrm{ent}(\mathbf{P})
\end{aligned}$$

Thus our entropy is a relative entropy, measuring the extent to which $\mathrm{ent}(\mathbf{P})$ is less than the maximum entropy $\log n$. This accounts for the change in sign.

Theorem 3.6.1 (The Csiszár-Kullback-Pinsker inequality) *Suppose that $f \in L^1(\mathbf{P})^+$ and that $\mathbf{E}(f) = 1$. Then $\|\mathbf{P} - fd\mathbf{P}\|_{TV} = \mathbf{E}(|f - \mathbf{E}(f)|) \leq \sqrt{2Ent_{\mathbf{P}}(f)}$.*

Proof We need the following inequality, due to Pinsker:

$$g(x) = 2(x+2)(x\log x - x + 1) - 3(x-1)^2 \geq 0 \text{ for } x > 0.$$

For $g(1) = 0$ and

$$g'(x) = 4((x+1)\log x - 2(x-1)) \geq 0 \text{ for } x \geq 1$$

since $\log x = \int_1^x dt/t \geq 2(x-1)/(x+1)$, by Jensen's inequality. If $0 < x < 1$ and $y = 1/x$ then

$$g'(x) = -4((y+1)\log y - 2(y-1))/y \leq 0.$$

Thus, using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbf{E})(|f - \mathbf{E}(f)|) &\leq \sqrt{2/3}\mathbf{E}(\sqrt{(f+2)(f\log f - f + 1)}) \\
&\leq \sqrt{2/3}(\mathbf{E}(f+2))^{1/2}(\mathbf{E}(f\log f - f + 1)^{1/2} \\
&= \sqrt{2}(\mathbf{E}(f\log f))^{1/2}.
\end{aligned}$$

□

The total variation norm does not take account of the metric d, but the Wasserstein metric does. We say that a probability measure $\mathbf{P}$ on a Polish metric space (X, d) satisfies a *transport inequality* with constant c if $W(f\mathbf{P}, \mathbf{P}) \leq \sqrt{2c\mathrm{Ent}_{\mathbf{P}}(f)}$ for all $f \in L^1(\mathbf{P})^+$ with $\mathbf{E}(f) = 1$.

Theorem 3.6.2 (Marton) *If that a probability measure* $\mathbf{P}$ *on a Polish metric space* (X,d) *satisfies a transport inequality then the concentration function* $\alpha_P(r)$ *satisfies* $\alpha_{\mathbf{P}}(r) \leq e^{-r^2/8c}$ *for* $r^2 \geq 8c\log 2$.

Proof Suppose that $\mathbf{P}(A) > 0$ and $\mathbf{P}(B) > 0$. Let $f_A = I_A/\mathbf{P}(A)$, $f_B = I_B/\mathbf{P}(B)$. Then by the triangle inequality,

$$\begin{aligned} W(f_A d\mathbf{P}, f_B d\mathbf{P}) &\leq W(f_A d\mathbf{P}, \mathbf{P}) + W(f_A d\mathbf{P}, \mathbf{P}) \\ &\leq \sqrt{2c\mathrm{Ent}_{\mathbf{P}}(f_A)} + \sqrt{2c\mathrm{Ent}_{\mathbf{P}}(f_B)} \\ &= \sqrt{2c\log(1/\mathbf{P}(A))} + \sqrt{2c\log(1/\mathbf{P}(B))} \end{aligned}$$

On the other hand, if π has marginals $f_A d\mathbf{P}$ and $f_B d\mathbf{P}$ then π must be supported on $A \times B$, so that

$$W(f_A d\mathbf{P}, f_B d\mathbf{P}) \geq d(A,B) = \inf\{d(a,b) : a \in A, b \in B\}.$$

Applying this to a set A with $\mathbf{P}(A) \geq 1/2$, and to $B = C(A_r)$, we see that

$$r \leq \sqrt{2c\log 2} + \sqrt{2c\log(1/\mathbf{P}(C(A_r)))},$$

from which the result follows. □

Proposition 3.6.2 *Suppose that* $\mathbf{P} \in P_1(X)$*, where* (X,d) *is a Polish metric space. Then* $\mathbf{P}$ *is sub-Gaussian, with exponent* b*, if and only if, whenever* $\psi \in \mathit{Lip}_{\,0}(X)$ *with* $\|\psi\|_L \leq 1$*, and* $f \in (L^1)^+$ *with* $\mathbf{E}(f) = 1$*, then*

$$\mathbf{E}((t\psi - b^2t^2/2)f) \leq \mathit{Ent}_{\mathbf{P}}(f) \text{ for all } t \in \mathbf{R}. \qquad (\dagger)$$

Proof If $\mathbf{P}$ is sub-Gaussian, with exponent b, then $\mathbf{E}(e^{t\psi - b^2t^2/2}) \leq 1$. If $f \in (L^1)^+$ and $\mathbf{E}(f) = 1$ then, by Corollary 3.6.1,

$$\begin{aligned} \mathrm{Ent}_P(f) &= \sup\{\int fg\,d\mu : fg^- \in L^1(\mu), \int e^g\,d\mu = 1\} \\ &\geq \mathbf{E}((t\psi - b^2t^2/2)f). \end{aligned}$$

Conversely, suppose that $(\dagger)$ holds, and suppose that $\psi \in \mathrm{Lip}_{\,0}$, with $\|\psi\|_L \leq 1$. Let $\psi_N = (\psi \wedge N) - \mathbf{E}(\psi \wedge N)$. Then $e^{t\psi_N}$ is bounded; let $f_N = e^{t\psi_N}/\mathbf{E}(e^{t\psi_N})$. Then

$$\begin{aligned} \mathrm{Ent}_{\mathbf{P}}(f_N) &= \mathbf{E}\left(\frac{e^{t\psi_N}}{\mathbf{E}(e^{t\psi_N})}(t\psi_N - \log\mathbf{E}(e^{t\psi_N}))\right) \\ &\geq \mathbf{E}\left(\frac{e^{t\psi_N}}{\mathbf{E}(e^{t\psi_N})}(t\psi_N - b^2t^2/2)\right) \end{aligned}$$

so that $\log(\mathbf{E}(e^{t\psi_N})) \le b^2t^2/2$, and $\mathbf{E}(e^{t\psi_N}) \le e^{b^2t^2/2}$. Suppose that $t \ge 0$. Since $\mathbf{E}(\psi_N) \le \mathbf{E}(\psi) = 0$, $\psi \wedge N \le \psi_N$, and so $\mathbf{E}(e^{t(\psi\wedge N)}) \le e^{b^2t^2/2}$. Thus $\mathbf{E}(e^{t\psi}) \le e^{b^2t^2/2}$, by monotone convergence. Since $-\psi$ also satisfies (†), $\mathbf{E}(e^{t\psi}) \le e^{b^2t^2/2}$ for $t < 0$, as well. □

Theorem 3.6.3 (Bobkov-Götze) *Suppose that $\mathbf{P} \in P_1(X)$, where (X, d) is a Polish metric space. Then $\mathbf{P}$ is sub-Gaussian, with exponent b, if and only if it satisfies a transport inequality with constant b^2.*

Proof Suppose first that P is sub-Gaussian, with exponent b, and that $f \in (L^1)^+$, with $\mathbf{E}(f) = 1$. Suppose that $\psi \in \mathrm{Lip}\,(X)$ and that $\|\psi\|_L \le 1$. Then $\psi \in L^1(\mathbf{P})$, since $\mathbf{P} \in P_1(X)$. Let $\psi_0 = \psi - \mathbf{E}(\psi)$. Then, by Proposition 3.6.2,

$$t\mathbf{E}(\psi f - \psi) = \mathbf{E}(t\psi_0 f) \le \frac{b^2t^2}{2} + \mathrm{Ent}_{\mathbf{P}}(f),$$

for all $t \in \mathbf{R}$. Setting $t = \sqrt{2\mathrm{Ent}_P(f)}/b$,

$$\mathbf{E}(\psi f - \psi) \le \sqrt{2b^2\mathrm{Ent}_P(f)},$$

so that $W(fd\mathbf{P}, \mathbf{P}) = \gamma(fd\mathbf{P}, \mathbf{P}) \le \sqrt{2b^2\mathrm{Ent}_P(f)}$, by the Kantorovitch-Rubinstein theorem.

Conversely, suppose that $W(fd\mathbf{P}, \mathbf{P}) \le \sqrt{2b^2\mathrm{Ent}_P(f)}$. Then if $f \in (L^1)^+$, with $\mathbf{E}(f) = 1$, and $\psi \in \mathrm{Lip}\,_0(X)$, with $\|\psi\|_L \le 1$, and if $t > 0$

$$\mathbf{E}(\psi f) = \mathbf{E}(\psi f - \psi) \le \sqrt{2b^2\mathrm{Ent}_P(f)} \le \frac{b^2t}{2} + \frac{1}{t}\mathrm{Ent}_{\mathbf{P}}(f),$$

so that $\mathbf{E}((t\psi - b^2t^2/2)f) \le \mathrm{Ent}_{\mathbf{P}}(f)$. Replacing ψ by $-\psi$, we see that the same inequality holds for $t < 0$, and so $\mathbf{P}$ is sub-Gaussian, with exponent b, by Proposition 3.6.2. □

4
Appendix

4.1 The bipolar theorem

Suppose first that E is a finite-dimensional vector space with dual E^*. We can take a basis, and use it to give E and E^* Euclidean norms. If A is a non-empty subset of E we define its polar as

$$A^\circ = \{\phi \in E^* : \sup_{a \in A} |\phi(a)| \leq 1\}.$$

A° is a closed convex symmetric subset of E^*. Similarly if B is a non-empty subset of E^*, we define $B^{\circ\circ} \subseteq E$.

Theorem 4.1.1 *If A is a non-empty closed convex symmetric subset of E, then $A = A^{\circ\circ}$.*

Proof Certainly $A \subseteq A^{\circ\circ}$. Suppose that $y \notin A$. There exists $z \in A$ such that $\|y - z\| = d(y, A)$. Let $w = y - z$, and let $\phi(x) = \langle x, w \rangle$, so that $\phi(y) - \phi(z) = \|w\|^2 > 0$. If $a \in A$ and $0 \leq \lambda \leq 1$ then $(1 - \lambda)z + \lambda a \in A$, and $y - ((1 - \lambda)z + \lambda a) = w + \lambda(z - a)$, so that

$$\langle w + \lambda(z - a), w + \lambda(z - a) \rangle \geq \langle w, w \rangle \,;$$

that is,

$$2\lambda \langle z - a, w \rangle + \lambda^2 \|z - a\|^2 \geq 0.$$

Thus $\phi(z) \geq \phi(a)$. Since A is symmetric, $\phi(z) \geq |\phi(a)|$. Choose $\phi(z) < r < \phi(y)$, and let $\psi = \phi/r$. Then $\psi \in A^\circ$ and so $y \notin A^{\circ\circ}$. □

Corollary 4.1.1 *If A is non-empty and symmetric, then $\overline{conv(A)} = A^{\circ\circ}$.*

What about the infinite-dimensional case? Here we use the Hahn-Banach Theorem.

Theorem 4.1.2 *Suppose that C is a non-empty closed convex subset of a real normed space $(E, \|.\|_E)$, and that $x \notin E$. Then there exists a continuous linear functional ϕ on E such that $\phi(x) > \sup\{\phi(c) : c \in C\}$.*

Proof Without loss of generality, we can suppose that $0 \in C$. Let $0 < d < d(x, C)/2$, and let $A = C + dB$, where B is the unit ball of E. Then A is a convex absorbing set. Let $p(y) = \inf\{\lambda : y \in \lambda A\}$. p is a sublinear functional on E. Since $x - dx/\|x\| \notin A$, $p(x) > 1$. Let $\psi(\alpha x) = \alpha p(x)$ Then ψ is a linear functional on span (x) and $\psi(\alpha x) \leq p(\alpha x)$. By the Hahn-Banach theorem, ψ extends to a linear functional ϕ on E with $\phi(y) \leq p(y)$ for all $y \in E$. Then $\phi(x) > 1$ and $\phi(y) \leq 1$ for $y \in A$, so that $\phi(c) \leq 1$ for $c \in C$. Finally, if $\|y\| \leq 1$ then $dy \in A$ and so $\phi(y) = \phi(dy)/d \leq 1/d$: thus ϕ is continuous. □

Corollary 4.1.2 *Suppose that C is a symmetric closed convex subset of a real normed space $(E, \|.\|_E)$. Then $C = C^{\circ\circ}$.*

Proof As before, $C \subseteq C^{\circ\circ}$. If $x \notin C$, there exists a continuous linear functional ϕ on E with $\phi(x) > 1$, $\phi(c) \leq 1$ for $c \in C$. Since C is symmetric, $|\phi(c)| \leq 1$ for $c \in C$, so that $\phi \in C^\circ$ and $x \notin C^{\circ\circ}$. □