Introduction

The purpose of this book is to introduce readers to certain topics in random matrix theory that specifically involve the phenomenon of concentration of measure in high dimension. Partly this work was motivated by researches in the EC network *Phenomena in High Dimension*, which applied results from functional analysis to problems in statistical physics. Pisier described this as the transfer of technology, and this book develops this philosophy by discussing applications to random matrix theory of:

(i) optimal transportation theory;
(ii) logarithmic Sobolev inequalities;
(iii) exponential concentration inequalities;
(iv) Hankel operators.

Recently some approaches to functional inequalities have emerged that make a unified treatment possible; in particular, optimal transportation links together seemingly disparate ideas about convergence to equilibrium. Furthermore, optimal transportation connects familiar results from the calculus of variations with the modern theory of diffusions and gradient flows.

I hope that postgraduate students will find this book useful and, with them in mind, have selected topics with potential for further development. Prerequisites for this book are linear algebra, calculus, complex analysis, Lebesgue integration, metric spaces and basic Hilbert space theory. The book does not use stochastic calculus or the theory of integrable systems, so as to widen the possible readership.

In their survey of random matrices and Banach spaces, Davidson and Szarek present results on Gaussian random matrices and then indicate that some of the results should extend to a wider context by the theory of concentration of measure [152]. This book follows this programme in the context of generalized orthogonal ensembles and compact Lie groups. While the Gaussian unitary ensemble and Wishart ensembles have special properties, they provide a helpful model for other cases. The book covers the main examples of the subject, such as Gaussian random matrices, within the general context of invariant ensembles.

The coverage of material is deliberately uneven, in that some topics are treated more thoroughly than others and some results from other areas of analysis are recalled with minimal discussion. There are detailed accounts of familiar topics such as the equilibrium measure of the quartic potential, since these illustrate techniques that are useful in many problems. The book develops classical and free probability in parallel, in the hope that the analogy makes free probability more accessible.

The presentation is mainly rigorous, although some important proofs are omitted. In order to understand the standard ensembles of random matrix theory, the reader must have some knowledge of Lie groups, so the book contains an abbreviated treatment which covers the main cases that are required and emphasizes the classical compact linear groups. Likewise, the presentations of Gaussian measures in Chapter 11 and the Ornstein–Uhlenbeck process in Chapters 12 and 13 are self-contained, but do not give a complete perspective on the theory. Similarly, the treatment of free probability describes only one aspect of the topic.

Some of the results and proofs are new, although the lack of a specific reference does not imply originality. In preparing the Sections 2.3, 2.4 and 2.6 on Lie groups, I have used unpublished notes from lectures given by Brian Steer in Oxford between 1987 and 1991. Chapter 5 features results originally published by the author in [17] and [19], with technical improvements due to ideas from Bolley's thesis [29]. The material in Chapter 6 on gradient flows was originally written for an instructional lecture to postgraduate students attending the North British Functional Analysis Seminar at Lancaster in 2006; likewise, Sections 8.1, 8.2, and 7.3 are drawn from postgraduate lectures at Lancaster. Conversely, Sections 7.2, 12.2 and 2.5 are based upon dissertations written by my former students Katherine Peet, Stefan Olphert and James Groves. Substantial portions of Chapter 9 and Section 10.3 are taken from Andrew McCafferty's PhD thesis [113], which the author supervised.

In his authoritative guide to lakeland hillwalking [168], Wainwright offers the general advice that one should keep moving, and he discusses 6 possible ascents of Scafell Pike, the optimal route depending upon the starting point, the time available and so on. Similarly, the author proposes 6 routes through the book, in addition to the obvious progression 1.1–14.3 which goes over the local maxima.

(1) Compact groups feature in the first half of the book, especially in Sections 1.2, 2.3–2.9, 3.8, 3.9, 5.1, 7.4, 10.5.
(2) Generalized orthogonal ensembles feature in the middle of the book, particularly in 1.5, 2.5, 3.4–3.7, 4.4–4.7, 6.3.
(3) Convergence to equilibrium distributions is the topic in 1.1, 3.4–3.9, 5.2–5.5, 10.6, 11.4.
(4) Free probability features in 4.3, 4.5, 4.8, 6.5, 6.6, 13.5, 14.1–3.
(5) Semicircular and similar special distributions appear in 4.4–4.7, 5.5, 7.3, 13.5, 14.3.
(6) Integrable operators appear in 9.1–9.7 and 11.2.

To summarize the contents of sections or the conclusions of examples, we sometimes give lists of results or definitions with bullet points. These should be considered in context, as they generally require elaboration. There are exercises that the reader should be able to solve in a few hours. There are also problems, which are generally very difficult and for which the answer is unknown at the time of writing.

There are many important topics in random matrix theory that this book does not cover, and for which we refer the reader elsewhere:

(i) the orthogonal polynomial technique and Riemann–Hilbert theory, as considered by Deift in [56];
(ii) connections with analytic number theory as in [98, 50];
(iii) applications to von Neumann algebras, as developed by Voiculescu and others [163, 164, 165, 166, 83, 84, 85, 77];
(iv) applications to physics as in [89];
(v) joint distributions of pairs of random matrices as in [76];
(vi) random growth models, and similar applications.

Jessica Churchman first suggested this topic as the subject for a book. I am most grateful to Graham Jameson, François Bolley, Alex Belton, Stefan Olphert, Martin Cook and especially James Groves for reading sections and suggesting improvements. Finally, I express thanks to Roger Astley of Cambridge University Press for bringing the project to fruition.

1

Metric measure spaces

Abstract

The contents of this chapter are introductory and covered in many standard books on probability theory, but perhaps not all conveniently in one place. In Section 1.1 we give a summary of results concerning probability measures on compact metric spaces. Section 1.2 concerns the existence of invariant measure on a compact metric group, which we later use to construct random matrix ensembles. In Section 1.3, we resume the general theory with a discussion of weak convergence of probability measures on (noncompact) Polish spaces; the results here are technical and may be omitted on a first reading. Section 1.4 contains the Brunn–Minkowski inequality, which is our main technical tool for proving isoperimetric and concentration inequalities in subsequent chapters. The fundamental example of Gaussian measure and the Gaussian orthogonal ensemble appear in Section 1.5, then in Section 1.6 Gaussian measure is realised as the limit of surface area measure on the spheres of high dimension. In Section 1.7 we state results from the general theory of metric measure spaces. Some of the proofs are deferred until later chapters, where they emerge as important special cases of general results. A recurrent theme of the chapter is weak convergence, as defined in Sections 1.1 and 1.3, and which is used throughout the book. Section 1.8 shows how weak convergence gives convergence for characteristic functions, cumulative distribution functions and Cauchy transforms.

1.1 Weak convergence on compact metric spaces

Definition (*Polish spaces*). Let (Ω, d) be a metric space. Then (Ω, d) is said to be complete if every Cauchy sequence converges; that is, whenever a sequence (x_n) in Ω satisfies $d(x_n, x_m) \to 0$ as $n, m \to \infty$, there exists $x \in \Omega$ such that $d(x_n, x) \to 0$ as $n \to \infty$.

A metric space (Ω, d) is said to be separable if there exists a sequence $(x_n)_{n=1}^{\infty}$ in Ω such that for all $\varepsilon > 0$ and all $x \in \Omega$, there exists x_n such that $d(x, x_n) < \varepsilon$. Such a sequence (x_n) is said to be dense.

A complete and separable metric space (Ω, d) is called a Polish space. A map $\varphi : (\Omega_1, d_1) \to (\Omega_2, d_2)$ between metric spaces is an *isometry* if $d_2(\varphi(x), \varphi(y)) = d_1(x, y)$ for all $x, y \in \Omega$.

Let $C_b(\Omega; \mathbf{R})$ be the space of bounded and continuous functions $f : \Omega \to \mathbf{R}$ with the supremum norm $\|f\|_\infty = \sup\{|f(x)| : x \in \Omega\}$.

Definition (*Compact metric spaces*). A metric space is said to be (sequentially) compact if for any sequence (x_n) in Ω there exist $x \in \Omega$ and a subsequence (x_{n_k}) such that $d(x_{n_k}, x) \to 0$ as $n_k \to \infty$. The reader may be familiar with the equivalent formulation in terms of open covers. See [150].

Definition (*Total boundedness*). Let (Ω, d) be a metric space. An ε-net is a finite subset S of Ω such that for all $x \in \Omega$, there exists $s \in \Omega$ such that $d(x, s) < \varepsilon$. If (Ω, d) has an ε-net for each $\varepsilon > 0$, then (Ω, d) is totally bounded.

A metric space is compact if and only if it is complete and totally bounded. See [150].

Proposition 1.1.1 *Suppose that (K, d) is a compact metric space. Then $C(K; \mathbf{R})$ is a separable Banach space for the supremum norm.*

Proof. Let (x_n) be a dense sequence in K and let $f_n : K \to \mathbf{R}$ be the continuous function $f_n(x) = d(x, x_n)$. Then for any pair of distinct points $x, y \in K$ there exists n such that $f_n(x) \neq f_n(y)$. Now the algebra

$$A = \Big\{\beta(0)\mathbf{I} + \sum_{S:S\subset \mathbf{N}} \beta_S \prod_{j:j\in S} f_j(x) : \beta_S \in \mathbf{Q} \quad \text{for all} \quad S; \beta_S = 0$$
$$\text{for all but finitely many} \quad S; \quad S \quad \text{finite}\Big\} \tag{1.1.1}$$

that is generated by the f_n and the rationals is countable and dense in $C(K; \mathbf{R})$ by the Stone–Weierstrass theorem; hence $C(K; \mathbf{R})$ is separable. See [141]. □

Definition (*Dual space*). Let $(E, \|\,.\,\|)$ be a real Banach space. A bounded linear functional is a map $\varphi : E \to \mathbf{R}$ such that:

(i) $\varphi(sx+ty) = s\varphi(x) + t\varphi(y)$ for all $x, y \in E$ and $s, t \in \mathbf{R}$;
(ii) $\|\varphi\| = \sup\{|\varphi(x)| : x \in E; \|x\| \leq 1\} < \infty$.

Let E^* be the space of all bounded linear functionals. Let $B = \{x \in E : \|x\| \leq 1\}$; then the product topology on $[-1,1]^B$ is generated by the open sets

$$\{(x_b)_{b\in B} : |x_{b_j} - y_{b_j}| < \varepsilon_j; j = 1, \dots, n\} \tag{1.1.2}$$

given by $b_j \in B$, $y_{b_j} \in [-1,1]$ and $\varepsilon > 0$ for $j = 1, \dots, n$. Further, $B^* = \{\phi \in E^* : \|\phi\| \leq 1\}$ may be identified with a closed subspace of $[-1,1]^B$ via the map $\phi \mapsto (\phi(x))_{x\in B}$. This is the weak* or $\sigma(E^*, E)$ topology on B^*. See [63, 141].

Theorem 1.1.2 (*Mazur*). *Let E be a separable Banach space. Then B^* is a compact metric space for the weak* topology. Further, E is linearly isometric to a closed linear subspace of $C(B^*; \mathbf{R})$.*

Proof. By Tychonov's theorem [141], $[-1,1]^B$ is a compact topological space, and hence the closed subspace $\{(\phi(x))_{x\in B} : \phi \in B^*\}$ is also compact; this is known as Alaoglu's theorem. Now we show that B^* has a metric that gives an equivalent topology; that is, gives the same collection of open sets.

Let $(x_n)_{n=1}^\infty$ be a dense sequence in B and let

$$d(\psi, \varphi) = \sum_{n=1}^\infty 2^{-n} |\varphi(x_n) - \psi(x_n)| \qquad (\varphi, \psi \in B^*), \tag{1.1.3}$$

so that d defines a metric on B^*. Now we check that d induces a compact Hausdorff topology on B^*, which must coincide with the weak* topology. Let (φ_j) be a sequence in B^*. We extract a subsequence $(\varphi_{j_1(k)})_{k=1}^\infty$ such that $\varphi_{j_1(k)}(x_1)$ converges as $j_1(k) \to \infty$; from this we extract a further subsequence $(\varphi_{j_2(k)})_{k=1}^\infty$ such that $\varphi_{j_2(k)}(x_2)$ converges as $j_2(k) \to \infty$; and so on. Generally we have $j_k : \mathbf{N} \to \mathbf{N}$ strictly increasing and $j_k(n) = j_{k-1}(m)$ for some $m \geq n$. Then we introduce the diagonal subsequence $(\varphi_{j_k(k)})$. By Alaoglu's theorem there exists $\phi \in B^*$ that is a weak* cluster point of the diagonal subsequence, and one checks that

$$d(\phi, \varphi_{j_k(k)}) = \sum_{n=1}^\infty 2^{-n} |\phi(x_n) - \varphi_{j_k(k)}(x_n)| \to 0 \tag{1.1.4}$$

as $j_k(k) \to \infty$ since $\varphi_{j_k(k)}(x_n) \to \phi(x_n)$ as $j_k(k) \to \infty$ for each n.

Let $f \in E$. Then f gives a function $\hat{f} : B^* \to \mathbf{R}$ by $\varphi \mapsto \varphi(f)$ which is continuous by the definition of the weak* topology. Further, by the Hahn–Banach Theorem [141] we have

$$\|\hat{f}\|_\infty = \sup\{|\varphi(f)| : \varphi \in B^*\} = \|f\|, \tag{1.1.5}$$

so $f \mapsto \hat{f}$ is a linear isometry $E \to C(B^*; \mathbf{R})$. The range of a linear isometry on a Banach space is complete and hence closed. □

Definition (*Borel measures*). Let (Ω, d) be a Polish space. A σ-algebra $\mathcal{A}$ on Ω is a collection of subsets of Ω such that:

(σ1) $\Omega \in \mathcal{A}$;
(σ2) if $A \in \mathcal{A}$, then $\Omega \setminus A \in \mathcal{A}$;
(σ3) if $(A_n)_{n=1}^\infty$ satisfies $A_n \in \mathcal{A}$ for all n, then $A = \bigcup_{n=1}^\infty A_n$ has $A \in \mathcal{A}$.

The sets A in a σ algebra $\mathcal{A}$ are called events.

The open subsets of Ω generate the Borel σ-algebra $\mathcal{B}(\Omega)$ and $M_b(\Omega)$ is the space of bounded Borel measures $\mu : \mathcal{B}(\Omega) \to \mathbf{R}$ such that

(i) $\|\mu\|_{var} = \sup\{\sum_j |\mu(E_j)| : E_j \in \mathcal{B}(\Omega)$ mutually disjoint, $j = 1, \dots, N\} < \infty$;
(ii) $\mu(\bigcup_{j=1}^\infty E_j) = \sum_{j=1}^\infty \mu(E_j)$ for all $(E_j)_{j=1}^\infty$ mutually disjoint $E_j \in \mathcal{B}(\Omega)$.

We write $M_b^+(\Omega)$ for the subspace of $\mu \in M_b(\Omega)$ such that $\mu(E) \geq 0$ for all $E \in \mathcal{B}(\Omega)$ and $Prob(\Omega)$ for the subspace $\{\mu \in M_b^+(\Omega) : \mu(\Omega) = 1\}$ of probability measures. Further, we write $M_1(\Omega) = \{\mu \in M_b(\Omega) : \|\mu\|_{var} \leq 1\}$. An event is a Borel-measurable subset of Ω. See [88]. For any Borel set A, $\mathbf{I}_A$ denotes the indicator function of A which is one on A and zero elsewhere, so $\mu(A) = \int_\Omega \mathbf{I}_A \, d\mu$.

A probability space $(\Omega, \mathbf{P})$ consists of a σ algebra $\mathcal{A}$ on Ω, and a probability measure $\mathbf{P} : \mathcal{A} \to \mathbf{R}$.

Theorem 1.1.3 (*Riesz representation theorem*). *Let (Ω, d) be a compact metric space and $\varphi : C_b(\Omega; \mathbf{R}) \to \mathbf{R}$ a bounded linear functional. Then there exists a unique $\mu \in M_b(\Omega)$ such that*

(iii) $\varphi(f) = \int f(x)\mu(dx)$ *for all* $f \in C_b(\Omega; \mathbf{R})$.

Conversely, each $\mu \in M_b(\Omega)$ defines a bounded linear functional φ via (iii) such that $\|\varphi\| = \|\mu\|_{var}$. Further, μ is a probability measure if and only if

(iv) $\|\varphi\| = 1 = \varphi(\mathbf{I})$.

Proof. See [88]. ☐

Definition (*Weak convergence*). Let $(\mu_j)_{j=1}^{\infty}$ be a sequence in $M_b(\Omega)$, and let $\mu \in M_b(\Omega)$. If

$$\lim_{j\to\infty} \int f\, d\mu_j = \int f d\mu \qquad (f \in C_b(\Omega)), \tag{1.1.6}$$

then (μ_j) converges weakly to μ. The term weak convergence is traditional in analysis; whereas the term weak* convergence would be more suggestive, since we have convergence in the $\sigma(M_b(\Omega); C_b(\Omega))$ topology.

Proposition 1.1.4 *Let $E = C_b(\Omega)$ and let $J : Prob(\Omega) \to B^* \subset [-1,1]^B$ be the map $J(\mu) = (\int f d\mu)_{f\in B}$. For a sequence (μ_j) in $Prob(\Omega)$ and $\mu \in Prob(\Omega)$,*

$$\mu_j \to \mu \quad \text{weakly} \Leftrightarrow J(\mu_j) \to J(\mu) \quad \text{in} \quad [-1,1]^B \qquad (j \to \infty). \tag{1.1.7}$$

Proof. This is immediate from the definitions. ☐

Proposition 1.1.5 *Let K be a compact metric space. Then $Prob\,(K)$ with the weak topology is a compact metric space.*

Proof. This follows immediately from Theorems 1.1.2 and Theorem 1.1.3 since $Prob(K)$ is linearly isometric to a compact subset of $C(K;\mathbf{R})^*$. ☐

Theorem 1.1.2 thus gives a metric for weak convergence on a compact metric space so that *Prob* becomes a compact metric space. The definition of the metric in Theorem 1.1.2 is rather contrived, so in Section 3.3 we shall introduce a more natural and useful metric for the weak topology, called the Wasserstein metric.

Examples. (i) The metric space $(M_b(\Omega), \|\,.\,\|_{var})$ is nonseparable when Ω is uncountable. Indeed $\|\delta_y - \delta_x\|_{var} = 2$ for all distinct pairs $x, y \in \Omega$.

(ii) Whereas B^* is compact as a subspace of $[-1,1]^B$, $J(Prob(\Omega))$ is not necessarily compact when Ω is noncompact. For example, when $\Omega = \mathbf{N}$ and δ_n is the Dirac unit mass at $n \in \mathbf{N}$, (δ_n) does not have any subsequence that converges to any $\mu \in Prob(\mathbf{N})$. In Proposition 1.2.5

we shall introduce *tightness* criteria that ensure that measures do not leak away to infinity in this way.

In applications, one frequently wishes to move measures forward from one space to another by a continuous function. The following result defines the *push forward* or *induced* measure $\nu = \varphi\sharp\mu$ on a compact metric space. A more general version appears in Theorem 1.3.5.

Proposition 1.1.6 (*Induced measure*). *Let $\varphi : (\Omega_0, d_0) \to (\Omega_1, d_1)$ be a Borel map between metric spaces where (Ω_1, d_1) is compact. Then for each $\mu \in M_b(\Omega_0)$ there exists a unique $\nu \in M_b(\Omega_1)$ such that*

$$\int_{\Omega_1} f(y)\nu(dy) = \int_{\Omega_0} f(\varphi(x))\mu(dx) \qquad (f \in C_b(\Omega_1)). \tag{1.1.8}$$

Proof. For $f \in C_b(\Omega_1)$, the function $f \circ \varphi$ is also bounded and Borel, hence integrable with respect to μ. The right-hand side clearly defines a bounded linear functional on $C_b(\Omega_1)$, and hence by Theorem 1.1.3 there exists a unique measure ν that realizes this functional. □

The following result is very useful when dealing with convergence of events on probability space. See [73, 88].

Theorem 1.1.7 (*First Borel–Cantelli lemma*). *Let $(A_n)_{n=1}^{\infty}$ be events in a probability space $(\Omega; \mathbf{P})$, and let C be the event with elements given by: $\omega \in C$ if and only if $\omega \in A_k$ for infinitely many values of k.*

If $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(C) = 0$.

Proof. We shall begin by checking that

$$C = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k. \tag{1.1.9}$$

By axiom $(\sigma 3)$, $C_n = \bigcup_{k=n}^{\infty} A_k$ is an event for each integer $n \geq 1$; consequently, $\bigcap_{n=1}^{\infty} C_n$ is also an event. If x belongs to C, then for each n, there exists $k_n \geq n$ with $x \in A_{k_n}$, so $x \in C_n$. Consequently x belongs to $\bigcap_{n=1}^{\infty} C_n$. Conversely, if $x \in \bigcap_{n=1}^{\infty} C_n$, then for each n, x belongs to C_n; so there exists $k_n \geq n$ with $x \in A_{k_n}$. But then x belongs to infinitely many A_j, and hence x is an element of C.

We can estimate the probability of $C_n = \bigcup_{k=n}^{\infty} A_k$ by

$$\mathbf{P}(C_n) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k), \tag{1.1.10}$$

using the axioms of measure. By hypothesis, the right-hand side is the tail sum of a convergent series, and hence $\sum_{k=n}^{\infty} \mathbf{P}(A_k) \to 0$ as $n \to \infty$. Further, $C \subseteq C_n$, so we can form a sandwich

$$0 \leq \mathbf{P}(C) \leq \mathbf{P}(C_n) \leq \sum_{k=n}^{\infty} \mathbf{P}(A_k) \to 0 \qquad (n \to \infty). \qquad (1.1.11)$$

Hence $\mathbf{P}(C) = 0$. □

Exercise 1.1.8 Let $\mu, \nu \in Prob(\Omega)$ be mutually absolutely continuous.

(i) Show that

$$\rho(\mu, \nu) = \int_\Omega \left(\frac{d\mu}{d\nu}\right)^{1/2} d\nu$$

satisfies $\rho(\mu, \nu) \leq 1$.

(ii) Now let $\delta(\mu, \nu) = -\log \rho(\mu, \nu)$. Show that:

(a) $\delta(\mu, \nu) \geq 0$;
(b) $\delta(\mu, \nu) = 0$ if and only if $\mu = \nu$ as measures;
(c) $\delta(\mu, \nu) = \delta(\nu, \mu)$.

(The triangle inequality does not hold for δ.)

1.2 Invariant measure on a compact metric group

- A compact metric group has a unique Haar probability measure.

Definition (*Compact metric group*). A topological group is a topological space G that is a group with neutral element e such that multiplication $G \times G \to G : (x, y) \mapsto xy$ and inversion $G \to G : x \mapsto x^{-1}$ are continuous. Furthermore, if the topology on G is induced by a metric d, then (G, d) is a metric group. Finally, if (G, d) is a metric group that is compact as a metric space, then G is a *compact metric group*.

Metric groups can be characterized amongst topological groups by their neighbourhoods of the identity as in [87, page 49], and we mainly use metric groups for convenience. Our first application of Theorem 1.2.1 is to show that a compact metric group has a unique probability measure that is invariant under left and right translation. The proof given here is due to von Neumann and Pontrjagin [130]. Compactness is essential to several stages in the proof; the result is actually valid for compact Hausdorff topological groups [87, 130].