Combinatorial Excursions to High-Dimensional Convex Geometry (text for the Spring School 2001)

Jiří Matoušek
Department of Applied Mathematics
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Rev. Mar/3/2001

This is a draft version. If you find errors or have other comments I'll be very grateful if you let me know.
J.M.

Introduction

This text has been prepared for the participants of the Spring School of Combinatorics 2001, organized by the Department of Applied Mathematics of the Charles University in Prague and its partners from abroad. Up to small changes, it is a part of a book in preparation on discrete geometry. At least until the Spring School, it is possible to obtain other chapters of this book project from the author's web page at
`http://www.ms.mff.cuni.cz/acad/kam/matousek/lectnotes.html`

The text is divided into 13 sections. There are a few short ones but most of the sections will probably require at least 1 hour or more for presentation. There will hardly be enough time at the Spring School to present everything, and I leave the choice of the material to the team preparing the lectures. For some sections, they might choose to explain the result but to omit the proof. My general advice is to leave enough time for explanations and discussions, rather than trying to cover as much as possible. I also do not recommend to give only sketches of proofs; preferably, either omit a proof or present it in full. Of course, some of the sections require results from preceding parts. Sections whose results are not required elsewhere and which could thus be omitted include 1.1, 1.3, 2.2, 2.3, 3.2, 3.5, and 3.6.

The present text is compiled from many sources (there is hardly anything new), but all references to the literature are omitted. Those interested in the history of the presented results, bibliography, and further information on related problems, can look into the corresponding three chapters available at the above address, which contain bibliographic notes to each section.

The topics treated here are mainly geometric and not combinatorial. However, I believe that they can be very useful for the education of a combinatorialist. As time progresses, more and more among the first-rate combinatorial results are proved by methods drawn from seemingly distant areas of mathematics, and geometric methods are among the most prominent. Some direct combinatorial applications are included in this text, but many more exist and probably still more are going to be found in the future.

Many of the proofs presented here cannot be considered easy; in fact, some of them are traditionally regarded as quite difficult. I've tried to make everything as concrete and elementary as reasonably possible and I have included numerous pictures and intuitive explanations. Still, the reader should not expect to understand everything painlessly. The text deals mostly with geometric questions in high-dimensional spaces, where the usual planar or 3-dimensional geometric intuition does not work too well, and one certainly needs some time and effort to get used to thinking about high-dimensional objects.

Solving at least some of the exercises should be extremely helpful in this training (easier parts of proofs are often presented as exercises). The framed number at the end of each exercise should indicate its expected difficulty. Exercised marked with $\boxed{1}$ should be more or less routine, while those with $\boxed{5}$, say, can be very challenging. Unfortunately, with no empirical data available, this marking follows just a subjective guess.

Preliminaries

For most of the text, one should suffice with the background from introductory mathematical courses (linear algebra, calculus, elementary probability, some graph theory). Here we recall some geometric terminology. For a more detailed introduction we refer e.g. to the initial chapters of the book project mentioned in the previous section (mainly the first chapter and the chapter on convex polytopes). In the lectures, these notions can be recalled when needed.

We work in the n-dimensional Euclidean space $\mathbf{R}^n$. The *linear subspaces* of $\mathbf{R}^n$ are subspaces of $\mathbf{R}^n$ considered as a real vector space. For example, in the plane $\mathbf{R}^2$, they are $\{0\}$, the whole $\mathbf{R}^2$, and all lines passing through 0. An *affine subspace* of $\mathbf{R}^n$ is a translation of a linear subspace. Affine subspaces in $\mathbf{R}^n$ of dimension $n-1$ are called *hyperplanes*. Any hyperplane can be written in the form $\{x \in \mathbf{R}^n\colon \langle a, x\rangle = b\}$ for some $a \in \mathbf{R}^n \setminus \{0\}$ and some $b \in \mathbf{R}$; here $\langle a, x\rangle = a_1x_1 + \cdots + a_nx_n$ is the usual scalar product. A (closed) *halfspace* is a subset of $\mathbf{R}^n$ bounded by a hyperplane, of the form $\{x \in \mathbf{R}^n\colon \langle a, x\rangle \geq b\}$ for some $a \in \mathbf{R}^n \setminus \{0\}$.

We mainly consider convex sets. These are sets $C \subseteq \mathbf{R}^n$ such that if $x, y \in C$ then $tx + (1-t)y \in C$ for all $t \in [0, 1]$, i.e. the line segment xy is contained in C. The *convex hull* of a set X is the set of all *convex combinations* of points in X; that is, $\mathrm{conv}(X) = \{t_1x_1+t_2x_2+\cdots+t_kx_k\colon k \geq 1, x_1, \ldots, x_k \in X, t_1, \ldots, t_k \geq 0, t_1 + \cdots + t_k = 1\}$. *Carathéodory's Theorem* states that for $X \subseteq \mathbf{R}^n$, any $x \in \mathrm{conv}(X)$ is the convex combination of at most $n + 1$ points of X. The *Separation Theorem* asserts that any two disjoint convex sets can be separated by a hyperplane: if $C, D \subset \mathbf{R}^n$ are convex then there are $a \in \mathbf{R}^n \setminus \{0\}$ and $b \in \mathbf{R}$ such that $\langle a, x\rangle \leq b$ for all $x \in C$ and $\langle a, y\rangle \geq b$ for all $y \in D$. In particular, if C is closed and D is a single point then there is a strictly separating hyperplane (with strict inequalities).

A *convex polytope* is the convex hull of a finite point set in $\mathbf{R}^n$, or equivalently, a bounded intersection of finitely many closed halfspaces in $\mathbf{R}^n$. A *face* of a convex polytope P is P itself or a set of the form $P \cap h$, where h is a hyperplane that does not separate P (all of P lies in only one of the two closed halfspaces defined by h). A face is again a convex polytope. Zero-dimensional faces are called vertices, and $(n-1)$-dimensional faces of an n-dimensional convex polytope are called *facets*. A k-dimensional *simplex* is the convex hull of $k+1$ points in $\mathbf{R}^n$ that do not lie in a common $(k-1)$-dimensional affine subspace. Simplices of dimensions 0,1,2,3 are points, segments, triangles, and tetrahedra, respectively.

If $K \subseteq \mathbf{R}^n$ is a closed convex set containing 0 in the interior then $K^* = \{y \in \mathbf{R}^n\colon \langle x, y\rangle \leq 1 \text{ for all } x \in K\}$ is the *dual* (or *polar*) set of K. If K is an n-dimensional convex polytope then K^* is a convex polytope too, and we have $f_k(K) = f_{n-k-1}(K^*)$ for all $k = 0, 1, \ldots, n-1$, where $f_k(P)$ denotes the number of k-dimensional faces of P.

For a Lebesgue-measurable $A \subseteq \mathbf{R}^n$, we let $\mathrm{vol}(A)$ denote the n-dimensional measure of A. We will mostly be dealing with simple geometric figures, where $\mathrm{vol}(A)$ is the usual volume. For $A \subset \mathbf{R}^n$ being a set and $x \in \mathbf{R}^n$ being a point, the *cone* with base A and apex x is the union of all segments connecting x to a point of A. If A lies in a hyperplane h, has $(n-1)$-dimensional volume v, and x has distance t from h, then the cone with base A and apex x has volume $\frac{1}{n}tv$ (this is the n-dimensional generalization of the usual formula for the area of a triangle: half of the height times the base).

1

Two Applications of High-Dimensional Polytopes

In the first and third sections, we touch upon *polyhedral combinatorics*. Let E be a finite set, for example the edge set of a graph G, and let $\mathcal{S}$ be some interesting system of subsets of E, such as the set of all matchings in G or the set of all Hamiltonian circuits of G. In polyhedral combinatorics, one usually considers the convex hull of the characteristic vectors of the sets of $\mathcal{S}$; the characteristic vectors are points of $\{0,1\}^E \subset \mathbf{R}^E$. For the two examples above, we thus obtain the *matching polytope* of G and the *traveling salesman polytope* of G. The basic problem of polyhedral combinatorics is to find, for a given $\mathcal{S}$, inequalities describing the facets of the resulting polytope. Sometimes one succeeds in describing all the facets, as is the case for the matching polytope. This may give insights into the combinatorial structure of $\mathcal{S}$, and often it has algorithmic consequences. If we know all the facets and they have sufficiently nice structure, we can optimize of any linear function over the polytope in polynomial time. This means that, given some real weights of the elements of E, we can find in polynomial time the set of maximum weight in $\mathcal{S}$ (e.g. the maximum-weight matching). In other cases, such as for the traveling salesman polytope, describing all facets is beyond reach. The knowledge of some facets may still yield interesting consequences, and on the practical side, it can provide a good approximation algorithm for maximum-weight set. Indeed, the largest traveling salesman problems solved in practice, with thousands of vertices, have been attacked by these methods.

We do not treat polyhedral combinatorics in any systematic manner; rather we focus on two gems (partially) belonging to this area. The first one is the celebrated Weak Perfect Graph Conjecture, stating that the complement of any perfect graph is perfect, which is proved by combining combinatorial and polyhedral arguments. The second one is an algorithmically motivated problem of sorting with partial information, discussed in Section 1.3. We associate a polytope with every finite partially ordered set, and we reduce the question to slicing such a polytope into two parts of roughly equal volume by a suitable hyperplane. A key role in this proof is played by the Brunn–Minkowski inequality. This fundamental geometric inequality is explained and proved in Section 1.2.

1.1 The Weak Perfect Graph Conjecture

First we recall a few notions from graph theory. Let $G = (V, E)$ be a finite undirected graph on n vertices. By $\overline{G}$ we denote the *complement* of G, i.e. the graph $(V, \binom{V}{2} \setminus E)$. An *induced subgraph* of G is any graph that can be obtained from G by deleting some vertices and all edges incident to the deleted vertices (but an edge must not be deleted if both of its vertices remain in the graph). Let $\omega(G)$denote the *clique number* of G, which is the maximum size of a complete subgraph of G, and let $\alpha(G) = \omega(\overline{G})$be *independence number* of G, which is the maximum size of an independent set in G. The symbol $\chi(G)$stands for the chromatic number of G; so $\chi(G)$ is the smallest number of independent sets covering all vertices of G.

Both the problems of finding $\omega(G)$ and finding $\chi(G)$ are computationally hard. It is NP-complete to decide whether $\omega(G) \leq k$, where k is a part of input, and it is NP-complete to decide whether $\chi(G) = 3$. Even approximating $\chi(G)$ or $\omega(G)$ is hard. So classes of graphs where the clique number and/or the chromatic number are computationally tractable are of great interest.

The perfect graphs are one of the most important such classes, and they include many other classes found earlier. A graph $G = (V, E)$ is called *perfect* if $\omega(G') = \chi(G')$ for every induced subgraph G' of G (including $G' = G$).

For every graph G, we have $\chi(G) \geq \omega(G)$, so a high clique number is a "reason" for a high chromatic number. But in general it is not the only possible reason, as there are graphs with $\omega(G) = 2$ but $\chi(G)$ arbitrarily large. Perfect graphs can be regarded as graphs where the chromatic number is exclusively controlled by the cliques, and this is true for G and also for all of its subgraphs.

For perfect graphs, the clique number ω (and hence also χ) can be computed in polynomial time by a sophisticated algorithm . It is not known how hard is the algorithmic problem of deciding perfectness: no polynomial-time algorithm has been found, but neither a proof of NP-completeness. But for graphs arising in many applications we know in advance that they are perfect.

Typical non-perfect graphs are the odd cycles C_{2k+1} of length 5 and larger, since $\omega(C_{2k+1}) = 2$ for $k \geq 2$ while $\chi(C_{2k+1}) = 3$.

The following two conjectures were formulated by Berge at early stages of research in perfect graphs. One is the so-called

> **Strong Perfect Graph Conjecture:** *A graph G is perfect if and only if neither G nor its complement contain an odd cycle of length ≥ 5 as an induced subgraph.*

This is still open, in spite of a considerable effort. The second conjecture is the

> **Weak Perfect Graph Conjecture:** *A graph is perfect if and only if its complement is perfect.*

This was first proved by Lovász. We reproduce a proof using convex polytopes.

1.1.1 Definition. *Let $G = (V, E)$ be a graph on n vertices. We assign a convex polytope $P(G) \subset \mathbf{R}^n$ to G. Let the coordinates in $\mathbf{R}^n$ be indexed by the vertices of G, i.e. if $V = \{v_1, \ldots, v_n\}$ then the points of $P(G)$ are of the form $x = (x_{v_1}, \ldots, x_{v_n})$. For an $x \in \mathbf{R}^n$ and a subset $U \subseteq V$ put $x(U) = \sum_{v \in U} x_v$.*

The polytope $P(G)$ is defined by the following inequalities:

(i) *$x_v \geq 0$ for each vertex $v \in V$, and*

(ii) *$x(K) \leq 1$ for each clique (complete subgraph) K in the graph G.*

Observations.

- $P(G) \subseteq [0, 1]^n$: the inequality $x_v \leq 1$ is obtained from (ii) by choosing $K = \{v\}$.
- The characteristic vector of each independent set lies in $P(G)$.
- If the vector $x \in P(G)$ is integral (i.e. it is a 0/1 vector) then it is the characteristic vector of an independent set.

Before we start proving the Weak Perfect Graph Conjecture, let us introduce some more notation. Let $w: V \to \mathbf{Z}_0^+$ be a function assigning non-negative integer weights to the vertices of G. We define the *weighted clique number* $\omega(G, w)$

as the maximum possible weight of a clique, where the weight of a clique is the sum of the weights of its vertices. We also define the *weighted chromatic number* $\chi(G, w)$as the minimum number of independent sets such that each vertex $v \in V$ is covered by $w(v)$ of these independent sets.

Now we can formulate the main theorem.

1.1.2 Theorem. *The following conditions are equivalent for a graph G:*

(i) *G is perfect.*

(ii) $\omega(G, w) = \chi(G, w)$ *for any nonnegative integral weight function w.*

(iii) *All vertices of the polytope $P(G)$ are integral and they correspond to the independent sets in G.*

(iv) *The graph $\overline{G}$ is perfect.*

Proof of (i) $\Rightarrow$ (ii). This part is purely graph-theoretical. For every weight function $w: V \to \{0, 1, 2, \ldots\}$, we need to exhibit a covering of V by independent sets witnessing $\chi(G, w) = \omega(G, w)$. If w attains only values 0 and 1 then we can use (i) directly, since selecting an induced subgraph of G is the same as specifying an 0/1 weight function on the vertices.

For the other w we proceed by induction on $w(V)$. Let w be given and let v_0 be a vertex with $w(v_0) > 1$. We define a new weight function w':

$$w'(v) = \begin{cases} w(v) - 1 & \text{for } v = v_0 \\ w(v) & \text{for } v \neq v_0. \end{cases}$$

Since $w'(V) < w(V)$, by the inductive hypothesis we assume that we have independent sets $I_1, I_2, \ldots, I_N$ covering each v exactly $w'(v)$-times, where $N = \omega(G, w')$. If $\omega(G, w) > N$ then we can obtain the appropriate covering for w by adding the independent set $\{v_0\}$, so let us suppose $\omega(G, w) = N$.

Let the notation be chosen so that $v_0 \in I_1$. We define another weight function w'':

$$w''(v) = \begin{cases} w(v) - 1 & \text{for } v \in I_1 \\ w(v) & \text{for } v \notin I_1. \end{cases}$$

We claim that $\omega(G, w'') < N$. If not then there exists a clique K with $w''(K) = N = \omega(G, w')$. By the choice of the I_i, we have $N \leq w'(K) = \sum_{i=1}^{N} |I_i \cap K|$. Since a clique intersects an independent set in at most one vertex, K has to intersect each I_i. In particular, it intersects I_1 and so $w(K) > w''(K) = N$, contradicting $\omega(G, w) = N$.

We thus have $\omega(G, w'') < N$. By the inductive hypothesis, we can produce a covering by independent sets witnessing $\chi(G, w'') < N$. By adding I_1 to it we obtain a covering witnessing $\chi(G, w) = N$.

Proof of (ii) $\Rightarrow$ (iii). Let $x = (x_{v_1}, \ldots, x_{v_n})$ be a vertex of the convex polytope $P(G)$. Since all the inequalities defining $P(G)$ have rational coefficients, x has rational coordinates and we can find a natural number q such that qx is an integral vector. We interpret the coordinates of qx as weights of the vertices of G. Let K be a clique with weight $N = \omega(G, qx)$. One of the inequalities defining $P(G)$ is $x(K) \leq 1$, and hence $N = qx(K) \leq q$.

By (ii) we have $\chi(G, qx) = \omega(G, qx) \leq q$, and so there are independent sets $I_1, \ldots, I_q$ (some of them may be empty) covering each vertex $v \in V$ precisely (qx_v)-times. Let c_i be the characteristic vector of I_i; then this property of the sets I_i can be written as $x = \sum_{i=1}^{q} \frac{1}{q} c_i$. Thus x is a convex combination of the c_i and since it is a vertex of $P(G)$, it must be equal to some c_i, which is a characteristic vector of an independent set in G.

Proof of (iii) $\Rightarrow$ (iv). It suffices to prove $\chi(\overline{G}) = \omega(\overline{G})$ for every G satisfying (iii) since (iii) is preserved by passing to an induced subgraph.

We prove that a graph G fulfilling (iii) has a clique K intersecting all independent sets of the maximum size $\alpha(G)$. Then the graph $G \setminus K$ has independence number $\alpha(G) - 1$, and by repeating the same procedure we can cover G by $\alpha(G)$ cliques.

To find the required K, let us consider all the independent sets of size $\alpha = \alpha(G)$ in G and let $M \subseteq P(G)$ be the convex hull of their characteristic vectors. We note that M lies in the hyperplane $h = \{x\colon x(V) = \alpha\}$. This h defines a (proper) face of $P(G)$, for otherwise we would have vertices of $P(G)$ on both sides of h and, in particular, there would be a vertex z with $z(V) > \alpha$, which is impossible since by (iii), z would correspond to an independent set bigger than α.

Each facet of $P(G)$ corresponds to an equality in some of the inequalities defining $P(G)$, i.e. either to an equality of the form $x_v = 0$ or of the form $x(K) = 1$. The face $F = P(G) \cap h$ is the intersection of some of the facets. Not all of these facets can be of the type $x_v = 0$ since then their intersection would contain 0, while $0 \notin h$. Hence all $x \in M$ satisfy $x(K) = 1$ for a certain clique K, and this means exactly that $K \cap I \neq \emptyset$ for each independent set I of size α.

Proof of (iv) $\Rightarrow$ (i). This is the implication (i) $\Rightarrow$ (iv) for the graph $\overline{G}$. $\square$

Exercises

1. Find some non-integral vertex of the polytope $P(C_5)$. Is there a nonzero integral vertex? $\boxed{1}$

2. (On König's Edge-Covering Theorem) Explain why bipartite graphs are perfect, and why the perfectness of the complements of bipartite graphs is equivalent to König's Edge-Covering Theorem asserting that the maximum number of vertex-disjoint edges in a bipartite graph equals the minimum number of vertices needed to intersect all edges. $\boxed{2}$

3. (Comparability graphs and Dilworth's theorem) For a finite partially ordered set $(X, \leq)$ (see Section 1.3 for the definition), let $G = (X, E)$ be the graph with $E = \{\{u, v\} \in \binom{X}{2} \colon u < v \text{ or } v < u\}$, i.e. edges correspond to pairs of comparable elements. Any graph isomorphic to such a G is called a *comparability graph*. We also need the notions of a *chain* (a subset of X linearly ordered by $\leq$) and an *antichain* (a subset of X with no two elements comparable under $\leq$).

   (a) Prove that any finite $(X, \leq)$ is the union of at most c antichains, where c is the length of the longest chain, and check that this implies the perfectness of comparability graphs. $\boxed{3}$

   (b) Derive from (a) the *Erdős–Szekeres Lemma*: if $a_1, a_2, \ldots, a_n$ are arbitrary real numbers then there exist indices $i_1, i_2, \ldots, i_k$ with $k^2 \geq n$ and such that the subsequence $a_{i_1}, a_{i_2}, \ldots, a_{i_k}$ is monotone (nondecreasing or decreasing). $\boxed{2}$

   (c) Check that the perfectness of the complements of comparability graphs is equivalent to the following theorem of Dilworth: any finite $(X, \leq)$ is the union of at most a chains, where a is the maximum number of elements of an antichain. $\boxed{1}$

4. (Hoffman's characterization of polytope integrality) Let P be a (bounded) convex polytope in $\mathbf{R}^n$ such that for every $a \in \mathbf{Z}^n$, the minimum of the function $x \mapsto \langle a, x \rangle$ over all $x \in P$ is an integer. Prove that all vertices of P are integral (i.e. they belong to $\mathbf{Z}^n$). $\boxed{5}$

5. (Kruskal–Hoffman theorem)

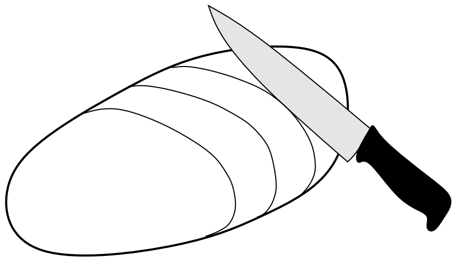
   (a) Show that if A is a nonsingular $n \times n$ totally unimodular matrix (all square submatrices have determinant 0 or ± 1), then the mapping $x \mapsto Ax$ maps $\mathbf{Z}^n$ bijectively onto $\mathbf{Z}^n$. $\boxed{3}$

(b) Show that if A is an $m \times n$ totally unimodular matrix and b is an m-dimensional integer vector such that the system $Ax = b$ has a real solution x, then it has an integral solution as well. $\boxed{3}$

(c) Let A be an $m \times n$ totally unimodular matrix and let $u, v \in \mathbf{Z}^n$ and $w, z \in \mathbf{Z}^m$ be integer vectors. Show that all vertices of the convex polyhedron given by the inequalities $u \le x \le v$ and $w \le Ax \le z$ are integral. $\boxed{1}$

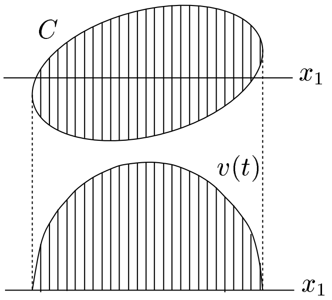
1.2 The Brunn–Minkowski Inequality

Let us consider a 3-dimensional convex loaf of bread and slice it by three parallel planar cuts.



As we will derive below, the middle cut cannot have area smaller than both of the other two cuts. Let us choose the coordinate system so that the cuts are perpendicular to the x_1-axis and denote by $v(t)$ the area of the cut by the plane $x_1 = t$. Then the claim can be stated as follows: for any $t_1 < t < t_2$ we have $v(t) \ge \min(v(t_1), v(t_2))$. Thus, there is some t_0 such that the function $t \mapsto v(t)$ is nondecreasing on $(-\infty, t_0]$ and nonincreasing on $[t_0, \infty)$. Such a function is called *unimodal*. A similar result is true for any convex body C in $\mathbf{R}^{n+1}$ if $v(t)$ denotes the n-dimensional volume of the intersection of C with the hyperplane $\{x_1 = t\}$.

How can one prove such a statement? In the planar case, with $n = 1$, it is easy to see that $v(t)$ is a concave function on the interval obtained by projecting C on the x_1-axis.



This might tempt one to think that $v(t)$ is concave on the appropriate interval in higher dimension too, but this is false in general! (See Exercise 1.) There is concavity in the game but the right function to look at in $\mathbf{R}^{n+1}$ is $v(t)^{1/n}$. Perhaps a little more intuitively, we can define $r(t)$ as the radius of the n-dimensional ball whose volume equals $v(t)$. We have $r(t) = R_n v(t)^{1/n}$, where R_n is the radius of a unit-volume ball in $\mathbf{R}^n$; let us call $r(t)$ the *equivalent radius* of C at t.

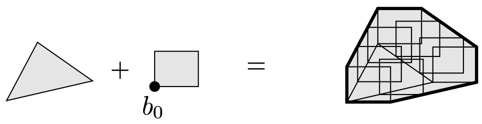
1.2.1 Theorem (Brunn's inequality for slice volumes). *Let $C \subset \mathbf{R}^{n+1}$ be a compact convex body and let the interval $[t_{min}, t_{max}]$ be the projection of C on the x_1-axis. Then the equivalent radius function $r(t)$ (or, equivalently, the function $v(t)^{1/n}$) is concave on $[t_{min}, t_{max}]$. Consequently, for any $t_1 < t < t_2$ we have $v(t) \geq \min(v(t_1), v(t_2))$.*

Brunn's inequality is a consequence of the following more general and more widely applicable statement dealing with two arbitrary compact sets.

1.2.2 Theorem (Brunn–Minkowski inequality). *Let A and B be nonempty compact sets in $\mathbf{R}^n$. Then*

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

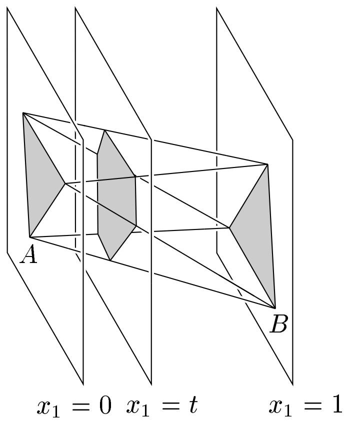
Here $A + B = \{a + b \colon a \in A, b \in B\}$denotes the *Minkowski sum* of A and B. If A' is a translated copy of A and B' a translated copy of B, then $A' + B'$ is a translated copy of $A + B$. So the position of $A + B$ with respect to A and B depends on the choice of coordinate system, but the shape of $A + B$ does not. One way of interpreting the Minkowski sum is as follows: keep A fixed, pick a point $b_0 \in B$, and translate B into all possible positions for which b_0 lies in A. Then $A + B$ is the union of all such translates. Here is a planar example:



Sometimes it is also useful to express the Minkowski sum $A+B$ as a projection of the Cartesian product $A \times B \subset \mathbf{R}^{2n}$ by the mapping $(x,y) \mapsto x+y$, $x, y \in \mathbf{R}^n$.

Proof of Brunn's inequality for slice volumes from the Brunn–Minkowski inequality. First we consider "convex combinations" of sets $A, B \subset \mathbf{R}^n$ of the form $(1-t)A + tB$, where $t \in [0,1]$ and where tA stands for $\{ta\colon a \in A\}$. As t goes from 0 to 1, $(1-t)A + t)B$ changes shape continuously from A to B.

Now if A and B are both convex and we place them into $\mathbf{R}^{n+1}$ so that A lies in the hyperplane $\{x_1 = 0\}$ and B in the hyperplane $\{x_1 = 1\}$, it is not difficult to check that $(1-t)A + tB$ is the slice of the convex body $\operatorname{conv}(A \cup B)$ by the hyperplane $\{x_1 = t\}$; see Exercise 2:



Let us consider the situation as in Brunn's inequality, where $C \subset \mathbf{R}^{n+1}$ is a convex body. Let A and B be the slices of C by the hyperplanes $\{x_1 = t_1\}$ and $\{x = t_2\}$, respectively, where $t_1 < t_2$ are such that $A, B \neq \emptyset$. For convenient notation, we change the coordinate system so that $t_1 = 0$ and $t_2 = 1$. To prove the concavity of the function $v(t)^{1/n}$ in Brunn's inequality, we need to show that for all $t \in (0,1)$,

$$(1-t)\operatorname{vol}(A)^{1/n} + t\operatorname{vol}(B)^{1/n} \leq \operatorname{vol}(M)^{1/n}, \tag{1.1}$$

where M is the slice M of C by the hyperplane $h_t = \{x_1 = t\}$. Let $C' = \mathrm{conv}(A \cup B)$ and $M' = C' \cap h_t$. We have $C' \subseteq C$ and $M' \subseteq M$. By the remark above, $M' = (1-t)A + tB$, and so the Brunn–Minkowski inequality applied to the sets $(1-t)A$ and tB yields

$$\begin{aligned}\mathrm{vol}(M)^{1/n} &\geq \mathrm{vol}(M')^{1/n} = \mathrm{vol}((1-t)A + tB)^{1/n}\\ &\geq \mathrm{vol}((1-t)A)^{1/n} + \mathrm{vol}(tB)^{1/n}\\ &= (1-t)\,\mathrm{vol}(A)^{1/n} + t\,\mathrm{vol}(B)^{1/n}.\end{aligned}$$

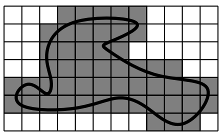
This verifies (1.1). □

Proof of the Brunn–Minkowski inequality. The idea of this proof is simple but perhaps surprising in this context. Call a set $A \subseteq \mathbf{R}^d$ a *brick set* if it is a union of finitely many closed axis-parallel boxes with disjoint interiors. First we show that it suffices to prove the inequality for brick sets (which is easy but a little technical), and then for brick sets the proof goes by induction on the number of bricks.

1.2.3 Lemma. *If the Brunn–Minkowski inequality fails for some two compact sets $A, B \subset \mathbf{R}^n$ then it fails for some brick sets $A, B \subset \mathbf{R}^n$ as well.*

Proof. We use a basic fact from measure theory, namely that if $X_1 \supseteq X_2 \supseteq X_3 \supseteq \ldots$ is a sequence of measurable sets in $\mathbf{R}^n$ such that $X = \bigcap_{i=1}^{\infty} X_i$ then the numbers $\mathrm{vol}(X_i)$ converge to $\mathrm{vol}(X)$.

In our situation, we suppose $\mathrm{vol}(A+B)^{1/n} < \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}$ for compact $A, B \subset \mathbf{R}^n$. For $k = 1, 2, \ldots$, consider the closed axis-parallel cubes with side length 2^{-k} centered at the points of the scaled grid $2^{-k}\mathbf{Z}^n$ (these cubes cover $\mathbf{R}^n$ and have disjoint interiors). Let A_k be the union of all such cubes intersecting the set A, and similarly for B_k.



We have $A_1 \supseteq A_2 \supseteq \ldots$ and $\bigcap_k A_k = A$ (since any point not belonging to A has a positive distance from it, and the distance of any point of A_k from A is at most $2^{-k}\sqrt{n}$). Therefore $\mathrm{vol}(A_k) \to \mathrm{vol}(A)$, $\mathrm{vol}(B_k) \to \mathrm{vol}(B)$.

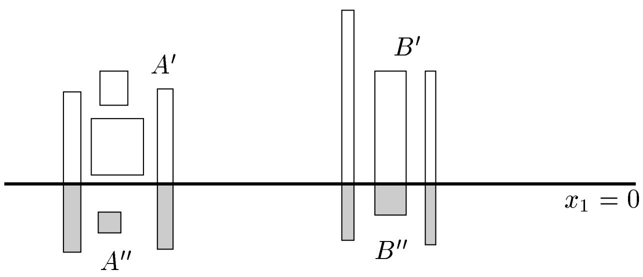
We claim that $A+B \supseteq \bigcap_k (A_k + B_k)$. Indeed, if $x \in A_k + B_k$ for all k, we pick $y_k \in A_k$ and $z_k \in B_k$ with $x = y_k + z_k$, and by passing to convergent subsequences we may assume that $y_k \to y \in A$ and $z_k \to z \in B$. Then we obtain $x = y + z \in A + B$. Thus $\lim_{k\to\infty} \mathrm{vol}(A_k + B_k) \le \mathrm{vol}(A + B)$ and we can find k sufficiently large so that $\mathrm{vol}(A_k + B_k)^{1/n} < \mathrm{vol}(A_k)^{1/n} + \mathrm{vol}(B_k)^{1/n}$. So A_k and B_k are brick sets violating the Brunn–Minkowski inequality. □

Proof of the Brunn–Minkowski inequality for brick sets. Let A and B be brick sets consisting of k bricks in total. If $k = 2$ then both A and B, and $A + B$ too, are bricks. Then if $x_1, \ldots, x_n$ are the sides of A and $y_1, \ldots, y_n$ are the sides of B, it suffices to establish the inequality in Exercise 3; we omit this part.

Now let $k > 2$ and suppose that the Brunn–Minkowski inequality holds for all pairs A, B of brick sets together consisting of fewer than k bricks. Let A and B have k bricks, and let the notation be chosen so that A has at least two bricks. Then it is easily seen that there exists hyperplane h parallel to some of the coordinate hyperplanes and with at least one full brick of A on one side and at least one full brick of A on the other side (Exercise 4). By a suitable choice of the coordinate system, we may assume that h is the hyperplane $\{x_1 = 0\}$.

Let A' be the part of A on one side of h and A'' the part on the other side. More precisely, A' is the closure of the intersection of A with one of the open halfspaces defined by h, and similarly for A''. Hence both A' and A'' have at least one brick fewer than A.

Next, we translate the set B in the x_1-direction in such a way that the hyperplane h divides its volume in the same ratio as A is divided (translation does not influence the validity of the Brunn–Minkowski inequality). Let B' and B'' be the respective parts of B.



Putting $\rho = \mathrm{vol}(A')/\mathrm{vol}(A)$, we also have $\rho = \mathrm{vol}(B')/\mathrm{vol}(B)$. (If $\mathrm{vol}(A) = 0$ or $\mathrm{vol}(B) = 0$ then the Brunn–Minkowski inequality is obvious.)

The sets A' and B' together have fewer than k bricks, so we can use the inductive assumption for them, and similarly for A'', B''.

The set $A' + B'$ is contained in one of the closed halfspaces defined by h and $A'' + B''$ lies in the other closed halfspace. Therefore, crucially, $\mathrm{vol}(A + B) \geq \mathrm{vol}(A' + B') + \mathrm{vol}(A'' + B'')$. We calculate

$$\begin{aligned}
\mathrm{vol}(A+B) &\geq \mathrm{vol}(A'+B') + \mathrm{vol}(A''+B'') \\
\text{(induction)} &\geq \left[\mathrm{vol}(A')^{1/n} + \mathrm{vol}(B')^{1/n}\right]^n + \left[\mathrm{vol}(A'')^{1/n} + \mathrm{vol}(B'')^{1/n}\right]^n \\
&= \left[\rho^{1/n}\,\mathrm{vol}(A)^{1/n} + \rho^{1/n}\,\mathrm{vol}(B)^{1/n}\right]^n \\
&\qquad + \left[(1-\rho)^{1/n}\,\mathrm{vol}(A)^{1/n} + (1-\rho)^{1/n}\,\mathrm{vol}(B)^{1/n}\right]^n \\
&= \left[\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}\right]^n.
\end{aligned}$$

This concludes the proof of the Brunn–Minkowski inequality. □

Exercises

1. Let A be a single point and B the n-dimensional unit cube. What is the function $v(t) = \mathrm{vol}(tA+(1-t)B)$? Show that $v(t)^\beta$ is not concave on $[0, 1]$ for any $\beta > \frac{1}{n}$. $\boxed{1}$

2. Let $A, B \subseteq \mathbf{R}^n$ be convex sets. Show that the sets $\mathrm{conv}\big((\{0\} \times A) \cup (\{1\} \times B)\big)$ and $\bigcup_{t\in[0,1]}\big[\{t\} \times ((1-)A + tB)\big]$ (in $\mathbf{R}^{n+1}$) are equal. $\boxed{2}$

3. Prove
$$\left(\prod_{i=1}^n x_i\right)^{1/n} + \left(\prod_{i=1}^n y_i\right)^{1/n} \leq \left(\prod_{i=1}^n (x_i + y_i)\right)^{1/n}$$
for arbitrary positive reals x_i, y_i. $\boxed{3}$

4. Show that for any brick set A with at least two bricks, there exists a hyperplane h parallel to one of the coordinate hyperplanes which has at least one full brick of A on each side. $\boxed{2}$

5. (Dimension-free form of Brunn–Minkowski) Consider the following two statements:

   (i) Theorem 1.2.2, i.e. $\mathrm{vol}(A + B)^{1/n} \geq \mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}$ for every nonempty compact $A, B \subset \mathbf{R}^n$.

(ii) For all compact $C, D \subset \mathbf{R}^n$ and all $t \in [0,1]$, $\operatorname{vol}(tC + (1-t)D) \geq \operatorname{vol}(C)^t \operatorname{vol}(D)^{1-t}$.

(a) Derive (ii) from (i); prove and use the inequality $tx + (1-t)y \geq x^t y^{1-t}$ (x, y positive reals, $t \in [0,1]$). $\boxed{3}$

(b) Prove (i) from (ii). $\boxed{3}$

6. Give a short proof of the 1-dimensional Brunn–Minkowski: $\operatorname{vol}(A + B) \geq \operatorname{vol}(A) + \operatorname{vol}(B)$ for any nonempty measurable $A, B \subset \mathbf{R}$. $\boxed{2}$

7. (Brunn–Minkowski via Prékopa–Leindler)The goal is to establish statement (ii) in Exercise 5.

   (a) Let $f, g, h\colon \mathbf{R} \to \mathbf{R}$ be bounded nonnegative measurable functions such that $h(tx + (1-t)y) \geq f(x)^t g(y)^{1-t}$ for all $x, y \in \mathbf{R}$ and all $t \in [0,1]$. Use the one-dimensional Brunn–Minkowski (Exercise 6) to prove $\int h \geq t \left(\int f\right) + (1-t)\left(\int g\right)$ (all integrals over $\mathbf{R}$); by the inequality in Exercise 5(a), the latter expression is at least $\left(\int f\right)^t \left(\int g\right)^{1-t}$. First show that we may assume $\sup f = \sup g = 1$. $\boxed{4}$

   (b) Prove the statement (ii) in Exercise 5 by induction on the dimension, using (a) in the induction step. $\boxed{3}$

1.3 Sorting Partially Ordered Sets

Here we present an amazing application of polyhedral combinatorics and of the Brunn–Minkowski inequality for a problem in theoretical computer science: sorting of partially ordered sets. Recall that a *partially ordered set*, or *poset* for short, is a pair $(X, \preceq)$, where X is a set and $\preceq$ is a binary relation on X (called an *ordering*) satisfying three axioms: reflexivity ($x \preceq x$ for all x), transitivity ($x \preceq y$ and $y \preceq z$ implies $x \preceq z$), and weak antisymmetry (if $x \preceq y$ and $y \preceq x$ then $x = y$). The ordering $\preceq$ is *linear* if every two elements of $x, y \in X$ are comparable; that is, $x \preceq y$ or $y \preceq x$.

Let X be a given finite set with some linear ordering $\leq$. For example, the elements of X could be identical-looking golden coins ordered by their weights (assuming that no two weights exactly coincide). We want to sort X according to $\leq$; that is, to list the elements of X in increasing order. We can get information about $\leq$ by *pairwise comparisons*: we can choose two elements $a, b \in X$ and ask an oracle whether $a \leq b$ or $a \geq b$ holds. In our example, we have precise scales such that only one coin fits to each scale, which allows us to make pairwise comparisons. Our sorting procedure may be adaptive; that is, the elements to be compared may be selected depending on the outcome of previous comparisons. We want to make as few comparisons as possible.

In the usual sorting problem, we begin with no information about the ordering $\leq$ whatsoever. As is well-known, in this case $\Theta(n \log n)$ comparisons are sufficient and also necessary in the worst case. Here we consider a different setting, when we start with some information already given. Namely, we obtain (explicitly) some partial ordering $\preceq$ on X, and we are guaranteed that $x \preceq y$ implies $x \leq y$; that is, $\leq$ is a *linear extension* of $\preceq$. In the example with coins, some weighings have already been made for us before we start. How many comparisons do we need to sort?

Let $E(\preceq)$denote the set of all linear extensions of a partial ordering $\preceq$ and let $e(\preceq) = |E(\preceq)|$

be the number of linear extensions. By sorting, we have to select just one element of $E(\preceq)$, which is one possibility out of $e(\preceq)$ many. Since any comparison of distinct elements a and b can have two outcomes, we need at least $\log_2 e(\preceq)$ comparisons in the worst case to distinguish the appropriate linear extension, i.e. to sort. Is this lower bound always asymptotically tight? Can one always sort using $O(\log_2 e(\preceq))$ comparisons, for any $\preceq$? An affirmative answer is implied by the following theorem:

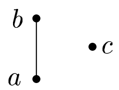
1.3.1 Theorem (Efficient Comparison Theorem).
Let $(X, \preceq)$ be a poset, and suppose that $\preceq$ is not linear. Then there exist elements $a, b \in X$ such that

$$\delta \leq \frac{e(\preceq + (a,b))}{e(\preceq)} \leq 1 - \delta,$$

where $\delta > 0$ is an absolute constant and $\preceq + (a,b)$ stands for the transitive closure of the relation $\preceq \cup \{(a,b)\}$; that is, the partial ordering we obtain from $\preceq$ if we are told that a precedes b.

How do we use this for sorting $\preceq$? For the first comparison, we choose the two elements a, b as in the theorem. Depending on the outcome of this comparison, we pass either to the partial ordering $\preceq +(a,b)$ or to $\preceq +(b,a)$. In both cases, the number of linear extensions has been reduced by the factor $1-\delta$: for $a \leq b$ this is clear by the theorem, and for $a \geq b$ this follows from the equality $e(\preceq + (a,b)) + e(\preceq + (b,a)) = e(\preceq)$. Hence, proceeding by induction, we can sort any partial ordering $\preceq$ using at most $\lceil \log_{1/(1-\delta)} e(\preceq) \rceil$ comparisons.

The conjectured "right" value of δ in Theorem 1.3.1 is $\frac{1}{3} \approx 0.33$; obviously, one cannot do any better for the poset



(meaning that $a \preceq b$ is the only pair of distinct elements in the relation $\preceq$). The proof below gives $\delta = \frac{1}{2e} \approx 0.184$, and more complicated proofs yield better values, although $\frac{1}{3}$ seems still elusive.

Order polytopes. For the proof, we need to assign certain polytopes to partial orderings and develop some of their simple properties.

1.3.2 Definition (Order polytope). *Let $(X, \preceq)$ be an n-element poset. Let the coordinates in $\mathbf{R}^n$ be indexed by the elements of X. We define a polytope $P(\preceq)$, the* order polytope *of $\preceq$, as the set of all $x \in [0,1]^n$ satisfying the following inequalities:*

$$x_a \leq x_b \quad \text{for every } a, b \in X \text{ with } a \preceq b.$$

Here is an alternative description of the order polytope:

1.3.3 Observation. *The vertices of the order polytope $P(\preceq)$ are precisely the characteristic vectors of all* up-sets *in $(X, \preceq)$, where an up-set is a subset $U \subseteq X$ such that if $a \in U$ and $a \preceq b$ then $b \in U$ too.*

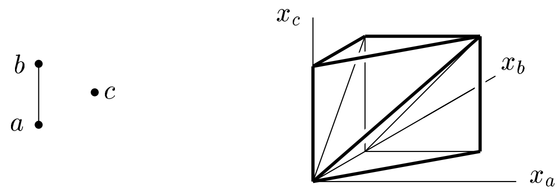
Proof. It is easy to see that the characteristic vector of an up-set is in $P(\preceq)$, and that any 0/1 vector in $P(\preceq)$ determines an up-set. It remains to check that all vertices of $P(\preceq)$ are integral. Any vertex is an intersection of some n facet hyperplanes. Since all potential facet hyperplanes have the form $x_a = x_b$, or $x_a = 0$, or $x_a = 1$, the integrality is obvious. $\square$

1.3.4 Observation. *Let X be an n-element set.*

(i) *If $\leq$ is a linear ordering on X then $P(\leq)$ is a simplex of volume $1/n!$.*

(ii) *For any partial ordering $\preceq$ on X, the simplices of the form $P(\leq)$, where $\leq$ is a linear extension of $\preceq$, cover $P(\preceq)$ and have disjoint interiors. Hence $\mathrm{vol}(P(\preceq)) = \frac{1}{n!}e(\preceq)$.*

Here is the order polytope and the subdivision into tetrahedra corresponding to the linear extensions, for a 3-element poset:



Proof of Observation 1.3.4. In (i), consider the ordering $1 \leq 2 \leq \ldots \leq n$. Then the characteristic vectors of up-sets have the form

$$(0, 0, \ldots, 0, 1, 1, \ldots, 1).$$

There are $n + 1$ of them and they are affinely independent, so $P(\leq)$ is a simplex. Other linear orderings differ just by a permutation of coordinates, so we get congruent simplices. The volume could be calculated directly but it follows easily from considerations below.

As for (ii), any point $(x_1, \ldots, x_n) \in P(\preceq)$ with pairwise distinct coordinates determines a unique linear extension of $\preceq$, namely the one given by the natural ordering of its coordinates as real numbers. Conversely, for any linear extension $\leq \in E(\preceq)$, we have $P(\leq) \subseteq P(\preceq)$ by definition. Hence the congruent simplices corresponding to linear extensions subdivide $P(\preceq)$.

To see that the simplices have volume $1/n!$, take the discrete ordering (no two distinct elements are comparable) for $\preceq$. The order polytope is the unit cube $[0, 1]^n$, and it is subdivided into $n!$ congruent simplices corresponding to the $n!$ possible linear orderings. □

Height and center of gravity. Let X be a finite set and $\leq$ a linear ordering on it. For $a \in X$, we define the *height of* a in $\leq$, denoted by $h_{\leq}(a)$, as $|\{x \in X\colon x \leq a\}|$. For a poset $(X, \preceq)$, the height of an element is defined as the average height over all linear extensions:

$$h_{\preceq}(a) = \frac{1}{e(\preceq)} \sum_{\leq \in E(\preceq)} h_{\leq}(a).$$

If $\preceq$ is clear from context we omit it in the subscript and we write just $h(a)$.

The "good" elements a, b in the Efficient Comparison Theorem 1.3.1 can be selected using the height. Namely, we show that any two distinct a, b

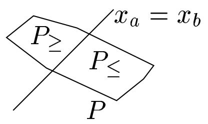
with $|h(a)-h(b)|<1$ will do. (It is simple to check that if $\preceq$ is not a linear ordering then such a and b always exist; see Exercise 1.)

We now relate the height to the order polytope.

1.3.5 Lemma. *For any n-element poset $(X,\preceq)$, the center of gravity of the order polytope $P(\preceq)$ is $c=(c_a\colon a\in X)$, where $c_a=\frac{1}{n+1}\,h_{\preceq}(a)$.*

Proof. The center of gravity of $P(\preceq)$ is the arithmetic average of centers of gravity of the simplices $P(\leq)$ with $\leq\in E(\preceq)$. Hence it suffices to prove the lemma for a linear ordering $\leq$. By permuting coordinates, it suffices to calculate that for the simplex with vertices of the form $(0,\ldots,0,1,\ldots,1)$, the center of gravity is $\frac{1}{n+1}(1,2,\ldots,n)$. This is left as Exercise 2. □

Proof of Theorem 1.3.1. Given the poset $(X,\preceq)$, we consider two elements $a,b\in X$ with $|h(a)-h(b)|<1$. We want to show that the number of linear extensions of both $\preceq+(a,b)$ and $\preceq+(b,a)$ is at least a constant fraction of $e(\preceq)$. Consider the order polytopes $P=P(\preceq)$, $P_{\leq}=P(\preceq+(a,b))$, and $P_{\geq}=P(\preceq+(b,a))$. Geometrically, P is sliced into $P_{\leq}$ and $P_{\geq}$ by the hyperplane $h=\{x_a=x_b\}$.



By Observation 1.3.4(ii), it suffices to show that the volumes of both $P_{\leq}$ and $P_{\geq}$ are at least a constant fraction of $\mathrm{vol}(P)$.

For convenience, let us introduce a new coordinate system in $\mathbf{R}^n$, where the first coordinate y_1 is x_b-x_a and the others complete it to an orthonormal coordinate system $(y_1,\ldots,y_n)$. Hence h is the hyperplane $y_1=0$. Let $c(P)$ denote the center of gravity of P, and let $c_1=c_1(P)$ be its y_1-coordinate.

What geometric information do we have about P? It is a convex body with the following properties:

- The projection of P onto the y_1-axis is the interval $[-1,1]$. This is because there is an up-set of $\preceq$ containing a and not b, and also an up-set containing b but not a, and thus P has a vertex with $x_a=1$, $x_b=0$ and a vertex with $x_a=0$, $x_b=1$.

- We have $-\frac{1}{n+1}\leq c_1\leq\frac{1}{n+1}$, since $c_1=\frac{1}{n+1}(h(a)-h(b))$.

The proof of Theorem 1.3.1 is finished by showing that any compact convex body $P \subset \mathbf{R}^n$ with the just mentioned two properties satisfies

$$\operatorname{vol}(P_{\le}) \ge \frac{1}{2e}\operatorname{vol}(P) \quad \text{and} \quad \operatorname{vol}(P_{\ge}) \ge \frac{1}{2e}\operatorname{vol}(P),$$

where $P_{\le}$ is the part of P in the halfspace $\{y_1 \le 0\}$ and similarly for $P_{\ge}$.

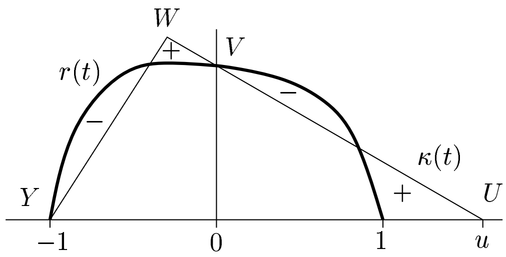
For $t \in [-1, 1]$, let P_t be the $(n-1)$-dimensional slice of P by the hyperplane $\{y_1 = t\}$, and let $r(t)$ be the equivalent radius of P_t, i.e. the radius of the $(n-1)$-dimensional ball of volume $\operatorname{vol}_{n-1}(P_t)$. By Brunn's inequality for slice volumes (Theorem 1.2.1), $r(t)$ is concave on $[-1, 1]$.

The y_1-coordinate of the center of gravity of P can be expressed as

$$c_1(P) = \tfrac{1}{\operatorname{vol}(P)} \int_{-1}^{1} t \operatorname{vol}_{n-1}(P_t)\, \mathrm{d}t$$

(imagine P composed of thin plates perpendicular to the y_1-axis). Hence c_1 is fully determined by the function $r(t)$. In other words, the shape of the slices of P does not really matter, only their volumes do, and so we may imagine that P is a rotational body whose slice P_t is an $(n-1)$-dimensional ball of radius $r(t)$ centered at $(t, 0, \ldots, 0)$.

We want to show that if $c_1(P) \ge -\frac{1}{n+1}$ then $\operatorname{vol}(P_{\ge}) \ge \frac{1}{2e}\operatorname{vol}(P)$. The inequality for $\operatorname{vol}(P_{\le})$ follows by symmetry. The key step is to pass to another, especially simple rotational convex body K. The slice K_t of K has radius $\kappa(t)$; the functions $\kappa(t)$ and $r(t)$ are schematically plotted below:

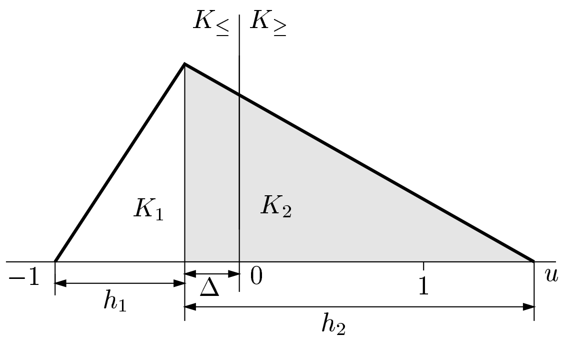


The graph of the function $\kappa(t)$ consists of two linear segments, and so K is a double cone. First we construct the function $\kappa(t)$ for t positive. Here the graph is a segment starting at the point $V = (0, r(0))$ and ending at the point $U = (u, 0)$. The number u is chosen in such a way that $\operatorname{vol}(K_{\ge}) = \operatorname{vol}(P_{\ge})$. Since $r(t)$ is concave and $\kappa(t)$ is linear on $[0, u]$, we have $u \ge 1$.

Moreover, as t grows from 0 to 1, we first have $r(t) \geq \kappa(t)$, and then from some point on $r(t) \leq \kappa(t)$. This ensures that the center of gravity of $K_{\geq}$ is to the right of the center of gravity of $P_{\geq}$ (we can imagine that $P_{\geq}$ is transformed into $K_{\geq}$ by peeling off some mass in the region labeled "−" and moving it right, to the region labeled "+").

Next, we define $\kappa(t)$ for $t < 0$. We extend the segment UV to the left until the (unique) point W such that, when YWV is the graph of $\kappa(t)$ for negative t, we have $\mathrm{vol}(K_{\leq}) = \mathrm{vol}(P_{\leq})$. As t goes from 0 down to -1, $\kappa(t)$ is first above $r(t)$ and then below it. This is because at V, the segment WU decreases more steeply than the function $r(t)$. Therefore, we also have $c_1(K_{\leq}) \geq c_1(P_{\leq})$, and hence $c_1(K) \geq c_1(P) \geq -\frac{1}{n+1}$. So, as was noted above, it remains to show $\mathrm{vol}(K_{\geq}) \geq \frac{1}{2e}\,\mathrm{vol}(K)$, which is a more or less routine calculation.

We fix the notation as in the following picture:



We note that $c_1(K)$ is a weighted average of $c_1(K_1)$ and $c_1(K_2)$; the weights are the volumes of K_1 and K_2, whose ratio is $h_1 : h_2$. The center of gravity of an n-dimensional cone is at $\frac{1}{n+1}$ of its height, and hence $c_1(K_1) = -\frac{h_1}{n+1} - \Delta$ and $c_1(K_2) = \frac{h_2}{n+1} - \Delta$. Therefore

$$c_1(K) = \frac{h_1\left(-\frac{h_1}{n+1}\right) + h_2\left(\frac{h_2}{n+1}\right)}{h_1 + h_2} - \Delta = \frac{h_2 - h_1}{n+1} - \Delta.$$

We have $\Delta = 1 - h_1$, and so from the condition $c_1(K) \geq -\frac{1}{n+1}$ we obtain $h_2 + nh_1 \geq n$. If we substitute $h_1 = u - h_2 + 1$ and rearrange, we get

$$\frac{u}{h_2} \geq 1 - \frac{1}{n}. \tag{1.2}$$

We are interested in bounding $\mathrm{vol}(K_{\geq})$ from below. The cone $K_{\geq}$ is similar to K_2, with ratio u/h_2. So

$$\begin{aligned}\mathrm{vol}(K_{\geq}) &= \left(\frac{u}{h_2}\right)^n \mathrm{vol}(K_2) = \left(\frac{u}{h_2}\right)^n \frac{h_2}{h_1+h_2}\,\mathrm{vol}(K)\\ &= \frac{u}{u+1}\left(\frac{u}{h_2}\right)^{n-1} \mathrm{vol}(K).\end{aligned}$$

Now we substitute for u/h_2 from (1.2), obtaining

$$\mathrm{vol}(K_{\geq}) \geq \frac{u}{u+1}\left(1-\frac{1}{n}\right)^{n-1} \mathrm{vol}(K).$$

Finally $\frac{u}{u+1} \geq \frac{1}{2}$ (as $u \geq 1$), and $(1-\frac{1}{n})^{n-1} > e^{-1}$ for all n, so $\mathrm{vol}(K_{\geq}) \geq \frac{1}{2e}\,\mathrm{vol}(K)$ follows. $\square$

Exercises

1. Let $(X, \preceq)$ be a finite poset. Prove that if $\preceq$ is not a linear ordering then there always exist $a, b \in X$ with $|h(a) - h(b)| < 1$. $\boxed{1}$

2. Show that the center of gravity of a simplex with vertices $a_0, a_1, \ldots, a_d$ is the same as the center of gravity of its vertex set. $\boxed{2}$

3. Let K be a bounded convex body in $\mathbf{R}^n$, h a halfplane passing through the center of gravity of K, and let K_1, K_2 be the parts into which K is divided by h.

   (a) Prove $\mathrm{vol}(K_1), \mathrm{vol}(K_2) \geq (\frac{n}{n+1})^n\,\mathrm{vol}(K)$. $\boxed{3}$

   (b) Show that the bound in (a) cannot be improved in general. $\boxed{2}$

2

Volumes in High Dimension

We begin with comparing the volume of the n-dimensional cube with the volume of the unit ball inscribed in it, in order to realize that volumes of "familiar" bodies behave quite differently in high dimensions from what the 3-dimensional intutition suggests. Then we calculate that any convex polytope in $\mathbf{R}^n$, whose number of vertices is at most polynomial in n and which is contained in the unit ball, is quite small compared to the ball. This has interesting consequences for deterministic algorithms for approximating the volume of a given convex body: if they only look at polynomially many points of the considered body then they are unable to distinguish a gigantic ball from a tiny polytope. Finally, we prove a classical result, John's Lemma, which states that for every n-dimensional symmetric convex body K there are two similar ellipsoids with ratio $\sqrt{n}$ such that the smaller ellipsoid lies inside K and the larger one contains K. So, in a very crude scale where the ratio $\sqrt{n}$ can be ignored, each symmetric convex body looks like an ellipsoid.

Besides presenting nice and important results, this chapter could help the reader in acquiring proficiency and good intuition in geometric computations, which are skills obtainable mainly by practice. Several calculations of nontrivial length are presented in detail, and while some parts do not require any great ideas, they still contain useful small tricks.

2.1 Volumes, Paradoxes of High Dimension, and Nets

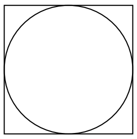
In the next section, we are going to estimate the volumes of various convex polytopes. Here we start, more modestly, with the volumes of the simplest bodies.

The ball in the cube. Let V_n denote the volume of the n-dimensional ball B^n of unit radius. A neat way of calculating V_n is indicated in Exercise 2; the result, which can be verified in various other ways and found in many books of formulas, is

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} = \frac{\pi^{\lfloor n/2 \rfloor} 2^{\lceil n/2 \rceil}}{\prod_{i:\, 0 \le 2i < n}(n-2i)}.$$

Here $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\,\mathrm{d}t$ is the usual Gamma-function, with $\Gamma(k+) = k!$ for natural numbers k.

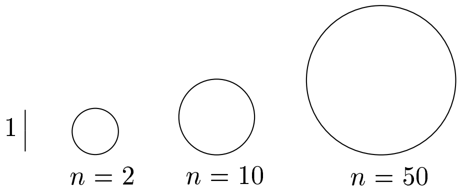
Let us compare the volume of the unit cube $[0,1]^n$ with that of the inscribed ball (of radius $\frac{1}{2}$).



(Using Exercise 1, the reader may want to add the crosspolytope inscribed in both bodies to the comparison.) For dimension $n = 3$, the volume of the ball is about 0.52, but for $n = 11$ it is already less than 10^{-3}. Using Stirling's formula, we find that it behaves roughly like $(\frac{2\pi e}{n})^{n/2}$. For large n, the inscribed ball is thus like a negligible dust particle in the cube, as far as the volume is concerned.

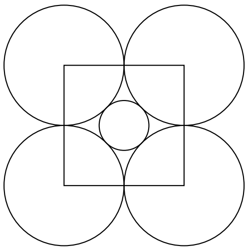
This can be experienced if one tries to generate random points uniformly distributed in the unit ball B^n. A straightforward metod is first to generate a random point x in the cube $[-1,1]^n$, by producing n independent random numbers $x_1, x_2, \ldots, x_n \in [-1,1]^n$. If $\|x\| > 1$ then x is discarded and the experiment is repeated, and if $\|x\| \le 1$ then x is the desired random point in the unit ball. This works reasonably in dimensions below 10, say, but in dimension 20, we expect about 40 million points to be discarded for each accepted point and the method is rather useless.

Another way of comparing the ball and the cube is to picture the sizes of the n-dimensional ball having the same volume as the unit cube:



For large n, the radius grows approximately like $0.24\sqrt{n}$. This indicates that the n-dimensional unit cube is actually quite a huge body; for example, its diameter (the length of the longest diagonal) is $\sqrt{n}$. Here is another example illustrating the largeness of the unit cube quite vividly.

Balls enclosing a ball. Place balls of radius $\frac{1}{2}$ into each of the 2^n vertices of the unit cube $[0,1]^n$, so that they touch along the edges of the cube, and consider the ball concentric with the cube and just touching the other balls:



Obviously, this ball is quite small and it is fully contained in the cube, right? No: already for $n = 5$ it starts protruding out through the facets.

Nets in a sphere. We conclude this section by introducing a generally useful tool. Let $S^{n-1} = \{x \in \mathbf{R}^n \colon \|x\| = 1\}$ denote the unit sphere in $\mathbf{R}^n$ (note that S^2 is the 2-dimensional sphere living in $\mathbf{R}^3$). We are given a number $\eta > 0$ and we want to place a reasonably small finite set N of points on S^{n-1} in such a way that each $x \in S^{n-1}$ has some point of N at distance no larger than η. (Numerous corporations are trying to do this, if not with the whole surface of the globe then with large parts of it, with η quite small.) Such an N is called *η-dense* in S^{n-1}. For example, the set $N = \{e_1, -e_1, \ldots, e_n, -e_n\}$ of the $2n$ orthonormal unit vectors of the standard basis is $\sqrt{2}$-dense. But it is generally difficult to find good explicit constructions for arbitrary η and n. The following simple but clever existential argument yields an η-dense set whose size has essentially the best possible order of magnitude.

Let us call a subset $N \subseteq S^{n-1}$ *η-separated* if every two distinct points of N have (Euclidean) distance greater than η. In a sense, this is opposite to being η-dense.

In order to construct a small η-dense set, we start with the empty set and keep adding points one by one. The trick is that we do not worry about

η-density along the way but we always keep the current set η-separated. Clearly, if no more points can be added, the current set must be η-dense.

The result of this algorithm is called an *η-net.*That is, $N \subseteq S^{n-1}$ is an η-net if it is an inclusion-maximal η-separated subset of S^{n-1}, i.e. if N is η-separated but $N \cup \{x\}$ is not η-separated for any $x \in S^{n-1} \setminus N$. (These definitions apply to an arbitrary metric space in place of S^{n-1}.) A volume argument bounds the maximum size of an η-net.

2.1.1 Lemma (Size of η-nets in the sphere). *For each $\eta \in (0, 1]$, any η-net $N \subseteq S^{n-1}$ satisfies*

$$|N| \leq \left(\frac{4}{\eta}\right)^n .$$

Later on, we will check that for η small, no η-dense set can be much smaller (Exercise 3.1.1).

Proof. For each $x \in N$, consider the ball of radius $\frac{\eta}{2}$ centered at x. These balls are all disjoint and they are contained in the ball $B(0, 1+\eta) \subseteq B(0, 2)$. Therefore $\mathrm{vol}(B(0, 2)) \geq |N| \,\mathrm{vol}(B(0, \frac{\eta}{2}))$, and since $\mathrm{vol}(B(0, r))$ in $\mathbf{R}^n$ is proportional to r^n, the lemma follows. $\square$

Exercises

1. Calculate the volume of the n-dimensional crosspolytope, i.e. the convex hull of $\{e_1, -e_1, \ldots, e_n, -e_n\}$, where e_i is the ith vector in the standard basis of $\mathbf{R}^n$. $\boxed{2}$

2. (Ball volume via the Gaussian distribution)

   (a) Let $I_n = \int_{\mathbf{R}^d} e^{-\|x\|^2} \,\mathrm{d}x$, where $\|x\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ is the Euclidean norm. Express I_n using I_1. $\boxed{1}$

   (b) Express I_n using $V_n = \mathrm{vol}(B^n)$ and a suitable one-dimensional integral, by considering the contribution to I_n of the spherical shell with inner radius r and outer radius $r + \mathrm{d}r$. $\boxed{2}$

   (c) Calculate I_n by using (b) for $n = 2$ and (a). $\boxed{1}$

   (d) Integrating by parts, set up a recurrence and calculate the integral appearing in (b). Compute V_n. $\boxed{2}$

2.2 Hardness of Volume Approximation

The theorem in this section can be regarded as a variation on one of the "paradoxes of high dimension" mentioned in the previous section, namely that the volume of the ball inscribed in the unit cube gets negligible as dimension grows. The theorem addresses a dual situation: the volume of a convex polytope inscribed in the unit ball.

2.2.1 Theorem. *Let B^n denote the unit ball in $\mathbf{R}^n$, and let P be a convex polytope contained in B^n and having at most N vertices. Then*

$$\frac{\mathrm{vol}(P)}{\mathrm{vol}(B^n)} \leq \left(\frac{C \ln N}{n}\right)^{n/2}$$

with an absolute constant C.

Thus, unless the number of vertices is exponential in n, the polytope is very tiny compared to the ball.

The proof shown below is neat but quite elementary and it makes seemingly very rough estimates. Nevertheless, it turns out that the bound is tight, up to the value of the constant C. A construction of polytopes witnessing this is discussed in the next section.

Application to hardness of volume approximation. Computing or estimating the volume of a given convex body in $\mathbf{R}^n$, with n large, is a fundamental algorithmic problem. Many combinatorial counting problems can be reduced to it, such as counting the number of linear extension of a given poset, as we saw in Section 1.3. Since many of these counting problems are computationally intractable, one cannot expect to compute the volume precisely, and so approximation up to some multiplicative factor is sought.

It turns out that no polynomial-time deterministic algorithm can generally achieve approximation factor better than exponential in the dimension. A concrete lower bound, derived with help of Theorem 2.2.1, is $(cn/\log n)^n$. This can also be almost achieved: deterministic algorithms are known with factor $O(n^{3n/2})$, resp. $O(n^n)$ for centrally symmetric bodies.

In striking contrast to this, there are *randomized* polynomial-time algorithms which can approximate the volume within a factor of $(1+\varepsilon)$ for each fixed $\varepsilon > 0$ with high probability. Here "randomized" means that the algorithm makes random decisions (like coin tosses) during its computation; it does not imply any randomness of the input. These are marvelous developments but they are not treated in this book. We only briefly explain the relation of Theorem 2.2.1 to the deterministic volume approximation.

To understand this connection, one needs to know how the input convex body is presented to an algorithm. A general convex body cannot be exactly described by finitely many parameters, so caution is certainly necessary. One way of specifying certain convex bodies, namely convex polytopes, is to give them as convex hulls of finite point set (V-presentation) or as intersections of finite sets of halfspaces (H-presentation). But there are many other computationally important convex bodies which are not polytopes, or have no polynomial-size V-presentation or H-presentation.

In order to abstract the considerations from the details of the presentation of the input body, the *oracle model* was introduced for computation with convex bodies. If $K \subset \mathbf{R}^n$ is a convex body, a *membership oracle* for K is, roughly speaking, an algorithm (subroutine, black box) that, for any given input point $x \in \mathbf{R}^n$, outputs YES if $x \in K$ and NO if $x \notin K$.

This is simplified because, in order to be able to compute with the body, one needs to assume more. Namely, K should contain a ball $B(0,r)$ and be contained in a ball $B(0,R)$ where R and $r > 0$ are written using at most polynomially many digits. On the other hand, the oracle need not (and often cannot) be exact, so a wrong answer is allowed for points very close to the boundary. These are important but rather technical issues and we will ignore them. Let us note that a polynomial-time membership oracle can be constructed for both V-presented and H-presented polytopes, as well as for many other bodies.

Let us now assume that a deterministic algorithm approximates the volume of each convex body given by a suitable membership oracle. First we call the algorithm with $K = B^n$ being the unit ball. The algorithm asks the oracle about some points $\{x_1, x_2, \ldots, x_N\}$, gets the correct answers, and outputs an estimate for $\mathrm{vol}(B^n)$. Next, we call the algorithm with the body $K = \mathrm{conv}(\{x_1, x_2, \ldots, x_N\} \cap B^n)$. The answers of the oracle are exactly the same, and since the algorithm has no other information about the body K and it is deterministic, it has to output the same volume estimate as it did for B^n. But by Theorem 2.2.1, $\mathrm{vol}(B^n)/\,\mathrm{vol}(K) \geq (cn/\ln N)^{n/2}$, and so the error of the approximation must be at least this factor. If N, the number of oracle calls, is polynomial in n, it follows that the error is at least $(c'n/\log n)^{n/2}$.

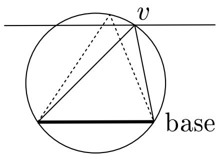
By more refined consideration, one can improve the lower bound to approximately the square of the just given quantity. The idea is to input the dual body K^* to the algorithm too, for which it gets the same answers, and then use a deep result stating that $\mathrm{vol}(K)\,\mathrm{vol}(K^*) \geq c^n/n!$ for any centrally symmetric n-dimensional convex body K, with an absolute constant $c > 0$

(some technical steps are omitted here). This improvement is interesting because, as was remarked above, for symmetric convex bodies it almost matches the performance of the best known algorithm.

Idea of the proof of Theorem 2.2.1. Let V be the set of vertices of the polytope $P \subset B^n$, $|V| = N$. We choose a suitable parameter $k = \lceil \frac{n}{\ln N} \rceil$ and prove that for every $x \in P$, there is a k-tuple J of points of V such that x is close to $\mathrm{conv}(J)$. Then $\mathrm{vol}(P)$ is simply estimated as $\binom{N}{k}$ times the maximum possible volume of the appropriate neighborhood of the convex hull of k points in B^n. Here is the first step towards realizing this program.

2.2.2 Lemma. *Let S in $\mathbf{R}^n$ be an n-dimensional simplex , i.e. the convex hull of $n+1$ affinely independent points, and let $R = R(S)$ and $\rho = \rho(S)$ be the circumradius and inradius of S, respectively, i.e. the radius of the smallest enclosing ball and of the largest inscribed ball. Then $\frac{R}{\rho} \geq n$.*

Proof. First sketch the proof of an auxiliary claim: *among all simplices contained in B^n, the regular simplex inscribed in B^n has the largest volume.* The volume of a simplex is proportional to the $(n-1)$-dimensional volume of its base times the corresponding height. It follows that in a maximum-volume simplex S inscribed in B^n, the hyperplane passing through a vertex v of S and parallel to the facet of S not containing v is tangent to B^n, for otherwise v could be moved to increase the height:



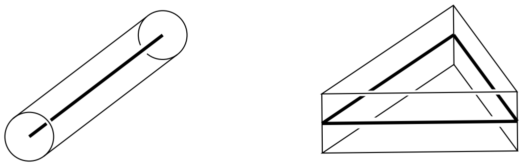
It can be easily shown that this property characterizes the regular simplex (so the regular simplex is even the unique maximum).

Another, slightly more difficult argument shows that if S is a simplex of minimum volume circumscribed to B^n then each facet of S touches B^n at its center of gravity (Exercise 2), and it follows that the volume is minimized by the regular simplex circumscribed to B^n.

Let S_0 be a simplex contained in B^n. We consider two auxiliary regular simplices S_1 and S_2, where S_1 is inscribed in B^n and S_2 satisfies $\mathrm{vol}(S_2) = \mathrm{vol}(S_0)$. Since $\mathrm{vol}(S_1) \geq \mathrm{vol}(S_0) = \mathrm{vol}(S_2)$, S_1 is at least as big as S_2, and so $\rho(S_0) \leq \rho(S_2) \leq \rho(S_1)$. Calculation shows that $\rho(S_1) = \frac{1}{n}$ (Exercise 1(a)).

□

Let F be a j-dimensional simplex in $\mathbf{R}^n$. Define the *orthogonal ρ-neighborhood* F_ρ of F as the set of all points $x \in \mathbf{R}^n$ for which there is a point $y \in F$ such that the segment xy is orthogonal to F and $\|x - y\| \leq \rho$. The next drawing shows orthogonal neighborhoods in $\mathbf{R}^3$ of a 1-simplex and of a 2-simplex:



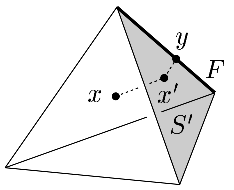
The orthogonal ρ-neighborhood of F can be expressed as the Cartesian product of F with a ρ-ball of dimension $n - j$, and so $\mathrm{vol}_n(F_\rho) = \mathrm{vol}_j(F) \cdot \rho^{n-j} \cdot \mathrm{vol}_{n-j}(B^{n-j})$.

2.2.3 Lemma. *Let S be an n-dimensional simplex contained in B^n, let $x \in S$, and let k be an integer parameter, $1 \leq k \leq n$. Then there is a k-tuple J of affinely independent vertices of S such that x lies in the orthogonal ρ-neighborhood of* $\mathrm{conv}(J)$, *where*

$$\rho = \rho(n, k) = \left(\sum_{i=k}^{n} \frac{1}{i^2} \right)^{1/2}.$$

Proof. We proceed by induction on $n - k$. For $n = k$, this is Lemma 2.2.2: consider the largest ball centered at x and contained in S; it has radius at most $\frac{1}{n}$, it touches some facet F of S at a point y, and the segment xy is perpendicular to F, witnessing $x \in F_{1/n}$.

For $k < n$, using the case $k = n$, let S' be a facet of S and $x' \in S'$ a point at distance at most $\frac{1}{n}$ from S' such that $xx' \perp S'$. By the inductive assumption for S', we find a $(k-1)$-face F of S' and a point $y \in F$ with $\|x' - y\| \leq \rho(n-1, k)$ and $x'y \perp F$. Here is an illustration for $n = 3$ and $k = 2$:



Then $xx' \perp x'y$ (because the whole S' is perpendicular to xx'), and so $\|x-y\|^2 = \|x-x'\|^2 + \|x'-y\|^2 \le \rho(n,k)^2$. Finally, $xy \perp F$ since both the vectors $x'-y$ and $x-x'$ lie in the orthogonal complement of the linear subspace generated by $F-y$. □

Proof of Theorem 2.2.1. By Carathéodory's Theorem and by Lemma 2.2.3, $P = \mathrm{conv}(V)$ is covered by the union of all the orthogonal ρ-neighborhoods $\mathrm{conv}(J)_\rho$, $J \in \binom{V}{k}$, where $\rho = \rho(n,k)$ is as in the lemma. The maximum $(k-1)$-dimensional volume of $\mathrm{conv}(J)$ is no larger than the $(k-1)$-dimensional volume of the regular $(k-1)$-simplex inscribed in B^{k-1}, which is

$$M(k-1) = \left(\frac{k}{k-1}\right)^{(k-1)/2} \frac{\sqrt{k}}{(k-1)!};$$

see Exercise 1(b). (If we do not care about the value of C in the theorem then $M(k-1)$ can also be trivially estimated by $\mathrm{vol}_{k-1}(B^{k-1})$ or even by 2^{k-1}.)

What remains is calculation. We have

$$\frac{\mathrm{vol}(P)}{\mathrm{vol}(B^n)} \le \binom{N}{k} \cdot M(k-1) \cdot \rho(n,k)^{n-k+1} \cdot \frac{\mathrm{vol}_{n-k+1}(B^{n-k+1})}{\mathrm{vol}(B^n)}. \tag{2.1}$$

We first estimate

$$\rho(n,k)^2 = \sum_{i=k}^{n} \frac{1}{i^2} \le \sum_{i=k}^{n} \frac{1}{i(i-1)} = \sum_{i=k}^{n} \left(\frac{1}{i-1} - \frac{1}{i}\right) = \frac{1}{k-1} - \frac{1}{n} \le \frac{1}{k-1}.$$

Before plunging into further calculations, let us see the orders of magnitude of various terms involved. Recall that $k = \lceil \frac{n}{\ln N} \rceil$. On the one hand, $\mathrm{vol}(P) = 0$ for $N \le n$ and so $k \le n/\ln n$. On the other hand, if $C \ln N \ge n$ then the bound asserted in the theorem is trivially valid, and so we may assume that k is larger than any suitable constant.

It is convenient take (natural) logarithms on both sides of (2.1). The logarithm of the bound we are heading for is $\frac{n}{2}(\ln\ln N - \ln n + O(1))$, and so we can neglect terms up to the order $O(n)$. The main term in our estimate (2.1) turns out to be the one with $\rho(n,k)$. For its logarithm, we find $\ln \rho(n,k)^{n-k+1} \le -(n+o(n)) \ln \sqrt{k} = (1+o(1))\frac{n}{2}(\ln\ln N - \ln n + O(1))$. All the remaining terms are negligible: $\ln \binom{N}{k} \le \ln N^k = n + o(n)$, $\ln M(k-1) \le 0$ for k large, and

$$\ln \frac{\mathrm{vol}_{n-k+1}(B^{n-k+1})}{\mathrm{vol}(B^n)} = \tfrac{n-k+1}{2} \ln \pi - \ln \Gamma(\tfrac{n-k+1}{2} + 1) - \tfrac{n}{2} \ln \pi + \ln \Gamma(\frac{n}{2} + 1)$$

$$= \ln \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n-k+1}{2}+1)} + O(n) \le \ln n^{k/2} + O(n) = O(n).$$

The proof of Theorem 2.2.1 is complete. □

Exercises

1. (a) Calculate the inradius and circumradius of a regular n-dimensional simplex.

   (b) Calculate the volume of the regular n-dimensional simplex inscribed in the unit ball B^n.

2. Let $S \subset \mathbf{R}^n$ be a simplex circumscribed to B^n and let F be a facet of S touching B^n at a point c. Show that if c is not the center of gravity of F then there is another simplex S' (arising by slightly moving the hyperplane that determines the facet F) that contains B^n and has volume smaller than $\mathrm{vol}(S)$.

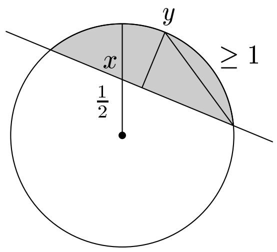
3. The *width* of a convex body K is the minimum distance of two parallel hyperplanes such that K lies between them. Prove that the convex hull of N points in B^n has width at most $O\big(\sqrt{(\ln N)/n}\big)$.

4. (A weaker but simpler estimate) Let $V \subset \mathbf{R}^n$ be a finite set. Prove that $\mathrm{conv}(V) \subseteq \bigcup_{v\in V} B(\frac{1}{2}v, \frac{1}{2}\|v\|)$, where $B(x,r)$ is the ball of radius r centered at x. Deduce that the convex hull of n points contained in B^n has volume at most $\frac{m}{2^n}\,\mathrm{vol}(B^n)$. $\boxed{4}$

2.3 Constructing Polytopes of Large Volume

For all N in the range $2n \le N \le 4^n$, we construct a polytope $P \subset B^n$ with N vertices containing a ball of radius $r = \Omega\big(\sqrt{(\ln \frac{N}{n})/n}\big)$. This shows that Theorem 2.2.1 is tight for $n^2 \le N \le 4^n$, since $\mathrm{vol}(P)/\mathrm{vol}(B^n) \ge r^n$. We begin with two extreme cases.

First we construct a k-dimensional polytope $P_0 \subset B^k$ with 4^k vertices containing the ball $\frac{1}{2}B^k$. There are several possible ways; the simplest is based on η-nets. We choose a 1-net $V \subset S^{k-1}$ and set $P_0 = \mathrm{conv}(V)$. According to Lemma 2.1.1, we have $N = |V| \le 4^k$. If there were an x with $\|x\| = \frac{1}{2}$ not lying in P_0,



then the separating hyperplane passing through x and avoiding P_0 would define a cap (shaded) whose center y would be at distance at least 1 from V.

Another extreme case is with $N = 2q$ vertices in dimension $n = q$. Then we can take the crosspolytope, i.e. the convex hull of the vectors e_1, $-e_1$,..., e_q, $-e_q$, where $(e_1, \ldots, e_q)$ is an orthonormal basis. The radius of the inscribed ball is $r = \frac{1}{\sqrt{q}}$ which matches the asserted formula.

Next, suppose that $n = qk$ for integers q and k, where $k \geq \ln n$, and set $N = q4^k$. We have $k \approx \ln N$ and $q \approx \frac{n}{\ln N}$, and so it suffices to construct an N-vertex polytope $P \subset B^n$ containing the ball rB^n with $r = \frac{1}{2\sqrt{q}}$.

The construction of P is, in a sense, a combination of the two constructions above. We interpret $\mathbf{R}^n$ as the product $\mathbf{R}^k \times \mathbf{R}^k \times \cdots \times \mathbf{R}^k$ (q factors). In each of the copies of $\mathbf{R}^k$, we choose a polytope P_0 with 4^k vertices as above, and we let P be the convex hull of their union. More formally,

$$P = \operatorname{conv}\Big\{(\underbrace{0, 0, \ldots, 0}_{(i-1)k\times}, x_1, x_2, \ldots, x_k, 0, 0, \ldots, 0)\colon (x_1, \ldots, x_k) \in V,\\ i \in \{1, 2, \ldots, q\}\Big\}$$

where V is the vertex set of P_0.

We want to show that P contains the ball rB^n, $r = \frac{1}{2\sqrt{q}}$. Let x be a point of norm $\|x\| \leq r$ and let x_i be the vector obtained from x by retaining the coordinates in the ith block, i.e. in positions $(i-1)k+1, \ldots, ik$, and setting all the other coordinates to 0. These x_i are pairwise orthogonal and x lies in the q-dimensional subspace spanned by them. Let $y_i = \frac{x_i}{2\|x_i\|}$ be the vector of length $\frac{1}{2}$ in the direction of x_i. Each y_i is contained in P since P_0 contains the ball of radius $\frac{1}{2}$. The convex hull of the y_i is a q-dimensional crosspolytope of circumradius $\frac{1}{2}$ and so it contains all vectors of norm $\frac{1}{2\sqrt{q}}$ in the subspace spanned by the x_i, including x.

This construction assumes that n and N are of a special form, but it is not difficult to extend the bounds to all $n \geq 2$ and all N in the range

$n^2 \leq N \leq 4^n$ by monotonicity considerations; we omit the details. This proves that the bound in Theorem 2.2.1 is tight up to the value of the constant C for $n^2 \leq N \leq 4^n$. $\square$

Exercises

1. (Polytopes with polynomially many facets inscribed in a ball)

   (a) Show that the cube inscribed in the unit ball B^n, which is a convex polytope with $2n$ facets, has volume of a larger order of magnitude than any convex polytope in B^n with polynomially many vertices (and so, concerning volume, "facets are better than vertices"). $\boxed{2}$

   (b) Prove that the inradius of any convex polytope with N facets contained in B^n is at most $O\big(\sqrt{(\ln N)/n}\,\big)$ (and so, in this respect, facets are not better than vertices). $\boxed{3}$

2.4 Approximating Convex Bodies by Ellipsoids

One of the most important issues in the life of convex bodies is their approximation by ellipsoids, as ellipsoids are in many respects the simplest imaginable compact convex bodies. The following result tells us how well they can generally be approximated (or how badly, depending on the point of view).

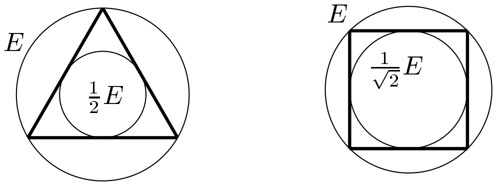
2.4.1 Theorem (John's Lemma). *Let $K \subset \mathbf{R}^n$ be a bounded closed convex body with nonempty interior. Then there exists an ellipsoid E such that*

$$E' \subseteq K \subseteq E,$$

where E' is E shrunk from its center by the factor n. If K is symmetric *about the origin, then we have the improved approximation*

$$\tfrac{1}{\sqrt{n}} E \subseteq K \subseteq E.$$

Thus, K can be approximated from outside and from inside by similar ellipsoids with ratio $1 : n$, or $1 : \sqrt{n}$ for the centrally symmetric case. Both these ratios are the best possible in general: this is shown by K being the regular simplex in the general case and the cube in the centrally symmetric case.



In order to work with ellipsoids, we need a rigorous definition. A suitable one is to consider ellipsoids as *affine images of the unit ball*: if B^n denotes the unit ball in $\mathbf{R}^n$, an ellipsoid E is a set $E = f(B^n)$, where $f\colon \mathbf{R}^n \to \mathbf{R}^n$ is an affine map of the form $f\colon x \mapsto Ax + c$. Here x is regarded as a column vector, $c \in \mathbf{R}^n$ is a translation vector, and A is a nonsingular $n \times n$ matrix. A very simple case is when $c = 0$ and A is a diagonal matrix with positive entries $a_1, a_2, \ldots, a_n$ on the diagonal. Then

$$E = \left\{ x \in \mathbf{R}^n \colon \frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \cdots + \frac{x_n^2}{a_n^2} \leq 1 \right\} \tag{2.2}$$

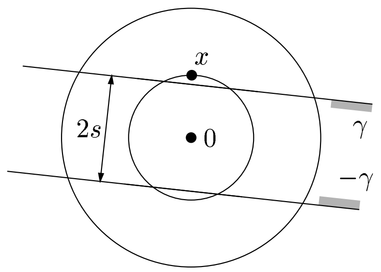
as is easy to check; this is an ellipsoid with center at 0 and with semiaxes $a_1, a_2, \ldots, a_n$. In this case, we have $\mathrm{vol}(E) = a_1 a_2 \cdots a_n \cdot \mathrm{vol}(B^n)$. An arbitrary ellipsoid E can be brought to this form by a suitable translation and rotation around the origin. In the language of linear algebra, this corresponds to diagonalizing a positive definite matrix using an orthonormal basis consisting of its eigenvectors; see Exercise 1.

Proof of Theorem 2.4.1. In both cases in the theorem, E is chosen as an ellipsoid of the smallest possible volume with $K \subseteq E$ (such an ellipsoid is often called the *Löwner–Johnellipsoid* of K). In Lemma 2.4.2 below, we show that such an E always exists (the minimum is attained).

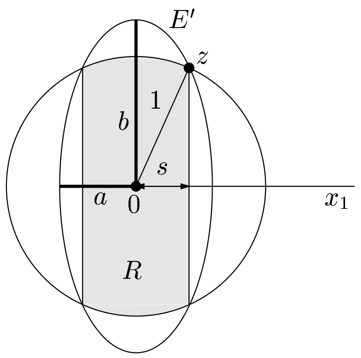
We prove only the centrally symmetric case. The non-symmetric case follows the same idea but the calculations are different and more complicated, and we leave them to Exercise 2.

So we suppose that K is symmetric about 0, we fix E, and make a linear transformation so that E becomes the unit ball B^n. Assuming that the smaller ball $n^{-1/2}B^n$ is not contained in K, we derive a contradiction by exhibiting a smaller enclosing ellipsoid for K.

We know that there is a point $x \notin K$ with $\|x\| \leq n^{-1/2}$. By the Separation Theorem, K can be strictly separated from x, and so there is a closed halfspace γ containing K but not x. Since K is symmetric, K is contained in the strip $\gamma \cap (-\gamma)$ of width $2s$, where $s < n^{-1/2}$:



For convenient notation, suppose that γ is perpendicular to the x_1-axis. To finish the proof, we show that the region $R = \{x \in \mathbf{R}^n \colon -s \leq x_1 \leq s,\ \|x\| \leq 1\}$, and therefore K too, are contained in an ellipsoid E' of volume smaller than $\mathrm{vol}(B^n)$:



Since R is symmetric under all rotations fixing the x_1-axis, we naturally choose E' symmetric too. Let the semiaxis in the x_1-direction have length a and the remaining $n-1$ semiaxes length $b \geq a$. As the picture suggests, and as is not difficult to verify formally, if E' contains a point $z \in \partial\gamma \cap \partial B^n$ then it contains all of R. The coordinates of z can be chosen as $(s, \sqrt{1-s^2}, 0, 0, \ldots, 0)$. Substituting this into the defining inequality of E' of the form (2.2), we obtain the requirement

$$\frac{s^2}{a^2} + \frac{1-s^2}{b^2} \leq 1. \tag{2.3}$$

We now want to choose $b \geq a > 0$ so that (2.3) holds and ab^{n-1} is minimum. This is an exercise in calculus. For convenience, we minimize the square, i.e. $a^2 b^{2(n-1)}$. It turns out that the minimum is attained for $a^2 = ns^2$

and $b^2 = \frac{n}{n-1}(1-s^2)$. Since $s^2 < \frac{1}{n}$, we get that $a^2 b^{2(n-1)} < 1$, and thus $\mathrm{vol}(E') < \mathrm{vol}(B^n)$. This contradicts our assumption that B^n was a minimum-volume enclosing ellipsoid for K, and Theorem 2.4.1 is proved, up to the proof of the following lemma. □

2.4.2 Lemma. *Let X be a bounded set in $\mathbf{R}^n$ that is not contained in a hyperplane, and let $\mathcal{E}$ be the set of all ellipsoids that contain X. Then there is an ellipsoid $E_0 \in \mathcal{E}$ such that $\mathrm{vol}(E_0) = \min_{E\in\mathcal{E}} \mathrm{vol}(E)$.*

Proof. We want to consider $\mathrm{vol}(\cdot)$ as a continuous function on a suitable compact space of ellipsoids; there are several possibilities of choosing an appropriate space.

Let $K = \mathrm{conv}(X)$. After a translation, we may assume that K is sandwiched between suitable balls: $B(0,r) \subseteq K \subseteq B(0,R)$, $r > 0$. We can restrict ourselves to enclosing ellipsoids of volume at most $\mathrm{vol}(B(0,R))$. There is a number M such that if $\|x\| \geq M$ then $\mathrm{vol}(\mathrm{conv}(\{x\}\cup B(0,r))) > \mathrm{vol}(B(0,R))$, and so we can consider only the ellipsoids contained in $B(0,M)$.

An ellipsoid E is given as the image $f(B^n)$, where $f\colon x \mapsto Ax + c$. So E can be identified with the pair (A,c), which can be regarded as a point in $\mathbf{R}^{n^2+n}$. Taking the standard Euclidean metric on the latter space, this induces a metric on the space of ellipsoids.

We check that for $E \subseteq B(0,M)$, the entries of c and A cannot be too large. Clearly, we have $|c_i| \leq M$ for all i. If $a_{ij} > 2M$, say, we let $x = e_j$ be the jth vector of the standard orthonormal basis, and we find that $\|f(e_j)\| \geq f(e_j)_i > a_{ij} - M \geq M$. Therefore $|a_{ij}| \leq 2M$.

Define $\mathcal{E}_M$ as the space of all ellipsoids with $|c_i| \leq M$ and $|a_{ij}| \leq 2M$ for all i,j. This space is compact and contains all ellipsoids of interest to us. Since $\mathrm{vol}(E) = |\det(A)| \cdot \mathrm{vol}(B^n)$, the volume is a continuous function on $\mathcal{E}_M$.

For a point $x \in X$, let $\mathcal{E}_M(x) = \{E \in \mathcal{E}_M\colon x \in E\}$. This $\mathcal{E}_M(x)$ can be described by polynomial inequalities, and so it is a closed subspace of $\mathcal{E}_M$. Therefore $\mathcal{E}_M(X) = \bigcap_{x\in X} \mathcal{E}_M(x)$ is compact and the continuous function $\mathrm{vol}(\cdot)$ attains its minimum on it. □

Exercises

1. Let E be the ellipsoid $f(B^n)$, where $f\colon x \mapsto Ax$ for an $n\times n$ nonsingular matrix A.

(a) Show that $E = \{x \in \mathbf{R}^n \colon x^T B x \le 1\}$. What is the matrix B? $\boxed{2}$

(b) Recall or look up appropriate theorems in linear algebra, showing that there is an orthonormal matrix R such that $B' = R^{-1}BR$ is a diagonal matrix with the eigenvalues of B on the diagonal (check and use the fact that B is positive definite in our case). $\boxed{2}$

(c) What is the geometric meaning of R, and what is the relation of the entries of $R^{-1}BR$ to the semiaxes of the ellipsoid E? $\boxed{2}$

2. Prove the part of Theorem 2.4.1 dealing with not necessarily symmetric convex bodies. $\boxed{4}$

3

Measure Concentration and Almost Spherical Sections

In the first two sections, we are going to discuss measure concentration on a high-dimensional unit sphere. Roughly speaking, measure concentration says that if $A \subseteq S^{n-1}$ is a set occupying at least half of the sphere then almost all points of S^{n-1} are quite close to A, at distance about $O(n^{-1/2})$. Measure concentration is an extremely useful technical tool in high-dimensional geometry. From the point of view of probability theory, it provides tail estimates for random variables defined on S^{n-1}, and in this respect it resembles Chernoff-type tail estimates for the sums independent random variables. But it is of a more general nature, more like tail estimates for Lipschitz functions on discrete spaces obtained using martingales.

The second main theme of this chapter are almost-spherical sections of convex bodies. Given a convex body $K \subset \mathbf{R}^n$, we want to find a k-dimensional subspace L of $\mathbf{R}^n$ such that $K \cap L$ is almost spherical, i.e. it contains a ball of some radius r and it is contained in the concentric ball of radius $(1+\varepsilon)r$. A remarkable Ramsey-type result, Dvoretzky's Theorem, shows that with k being about $\varepsilon^{-2}\log n$, such a k-dimensional almost-spherical section exists for every K. We also include an application concerning convex polytopes, showing that a high-dimensional centrally symmetric convex polytope cannot have both a small number of vertices and a small number of facets.

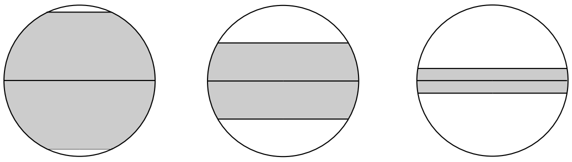
Both measure concentration and the existence of almost-spherical sections are truly high-dimensional phenomena, practically meaningless in the familiar dimensions 2 and 3. The low-dimensional intuition is of little use here, but perhaps by studying some number of results and examples, one can develop intuition on what to expect in high dimensions.

We present only a few selected results from an extensive and well-developed theory of high-dimensional convexity. Most of it was built in the so-called *local theory of Banach spaces*, which deals with the geometry of finite-dimensional subspaces of various Banach spaces. In the literature, the theorems are usually formulated in the language of Banach spaces, so instead of symmetric convex bodies, one speaks about norms, and so on. Here we introduce some rudimentary terminology concerning normed spaces but we express most of the notions in geometric language, in the hope of making it more accessible to non-specialists in Banach spaces. So, for example, in the formulation of Dvoretzky's Theorem, we do not speak about the Banach–Mazur distance to an inner product norm but rather about almost spherical convex bodies. On the other hand, for a more serious study of this theory, the language of normed spaces seems most appropriate.

3.1 Measure Concentration on the Sphere

Let P denote the usual surface measure on the unit Euclidean sphere S^{n-1}, scaled so that the whole S^{n-1} has measure 1 (a rigorous definition will be mentioned later). This P is a probability measure, and we often think of S^{n-1} as a probability space. For a set $A \subseteq S^{n-1}$, $\mathrm{P}[A]$ is the P-measure of A and also the probability that a random point of S^{n-1} falls into A. The letter P should suggest "probability of", and the notation $\mathrm{P}[A]$ is analogous to $\mathrm{Prob}[A]$ used elsewhere in the book.

Measure concentration on the sphere can be approached in two steps. The first step is the observation, interesting but rather easy to prove, that for large n, most of S^{n-1} lies quite close to the "equator". For example, the following diagram shows the width of the band around the equator that contains 90% of the measure, for various dimensions n:



$n = 3$ $n = 11$ $n = 101$

That is, if the width of the gray stripe is $2w$, it means that

$$\mathrm{P}\left[\{x \in S^{n-1} \colon -w \leq x_n \leq w\}\right] = 0.9.$$

As we will see later, w is of the order $n^{-1/2}$ for large n. (Of course, one might ask, why the measure is concentrated just around the "equator" $x_n = 0$. But, counterintuitive as it may sound, it is concentrated around *any* equator, i.e. near any hyperplane containing the origin.)

The second, considerably deeper step shows that the measure on S^{n-1} is concentrated not only around the equator, but near the boundary of any (measurable) subset $A \subset S^{n-1}$ covering half of the sphere. Here is a precise quantitative formulation.

3.1.1 Theorem (Measure concentration for the sphere). *Let $A \subseteq S^{n-1}$ be a measurable set with $\mathrm{P}[A] \geq \frac{1}{2}$, and let A_t denote the t-neighborhood of A; that is, the set of all $x \in S^{n-1}$ whose Euclidean distance to A is at most t. Then*

$$1 - \mathrm{P}[A_t] \leq 2e^{-t^2 n/2}.$$

Thus, if A occupies half of the sphere, almost all points of the sphere lie at distance at most $O(n^{-1/2})$ from A; only extremely small reserves can vegetate undisturbed by the nearness of A. To recover the concentration around the equator, it suffices to choose A as a hemisphere.

We present a simple and direct geometric proof of a slightly weaker version of Theorem 3.1.1, with $-t^2 n/4$ in the exponent instead of $-t^2 n/2$. It deals with both the steps mentioned above in one stroke.

It is based on the Brunn–Minkowski inequality (Theorem 1.2.2): $\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n} \leq \mathrm{vol}(A+B)^{1/n}$ for any nonempty compact sets $A, B \subset \mathbf{R}^n$. We will actually use a slightly different version of the inequality, which resembles the well-known inequality between the arithmetic and geometric means, at least optically:

$$\mathrm{vol}(\tfrac{1}{2}(A+B)) \geq \sqrt{\mathrm{vol}(A)\,\mathrm{vol}(B)}. \tag{3.1}$$

This is easily derived from the usual version: we have $\mathrm{vol}(\frac{1}{2}(A+B))^{1/n} \geq \mathrm{vol}(\frac{1}{2}A)^{1/n} + \mathrm{vol}(\frac{1}{2}B)^{1/n} = \frac{1}{2}(\mathrm{vol}(A)^{1/n} + \mathrm{vol}(B)^{1/n}) \geq (\mathrm{vol}(A)\,\mathrm{vol}(B))^{1/2n}$ by the inequality $\frac{1}{2}(a+b) \geq \sqrt{ab}$.

Proof of a weaker version of Theorem 3.1.1. For a set $A \subseteq S^{n-1}$, we define $\tilde{A}$ as the union of all the segments connecting the points of A to 0: $\tilde{A} = \{\alpha x \colon x \in A,\ \alpha \in [0,1]\} \subseteq B^n$. Then we have

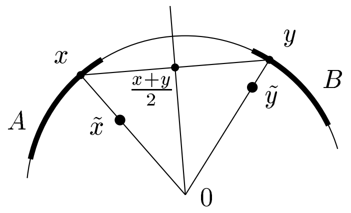
$$\mathrm{P}[A] = \mu(\tilde{A}),$$

where $\mu(\tilde{A}) = \mathrm{vol}(\tilde{A}) / \mathrm{vol}(B^n)$ is the normalized volume of $\tilde{A}$; in fact, this can be taken as the definition of $\mathrm{P}[A]$.

Suppose that $t \in [0,1]$. Let $\mathrm{P}[A] \geq \frac{1}{2}$ and let $B = S^{n-1} \setminus A_t$. Therefore, $\|a - b\| \geq t$ for all $a \in A$, $b \in B$.

3.1.2 Lemma. *For any $\tilde{x} \in \tilde{A}$ and $\tilde{y} \in \tilde{B}$, we have $\|\frac{\tilde{x}+\tilde{y}}{2}\| \leq 1 - t^2/8$.*

Proof of the lemma. Let $\tilde{x} = \alpha x$, $\tilde{y} = \beta y$, $x \in A$, $y \in B$:



First we calculate, by the Pythagoras Theorem and by elementary calculus,

$$\left\|\frac{x+y}{2}\right\| \leq \sqrt{1-\tfrac{t^2}{4}} \leq 1-\tfrac{t^2}{8}.$$

For passing to $\tilde{x}$ and $\tilde{y}$, we may assume that $\beta = 1$. Then

$$\begin{aligned}\left\|\frac{\tilde{x}+\tilde{y}}{2}\right\| &= \left\|\frac{\alpha x+y}{2}\right\| \leq \alpha\left\|\frac{x+y}{2}\right\| + (1-\alpha)\left\|\frac{y}{2}\right\| \\ &= \alpha(1-\tfrac{t^2}{8}) + (1-\alpha)(1-\tfrac{1}{2}) \leq 1-\tfrac{t^2}{8}.\end{aligned}$$

The lemma is proved.

By the lemma, the set $\frac{1}{2}(\tilde{A}+\tilde{B})$ is contained in the ball of radius $1-t^2/8$ around the origin. Applying Brunn–Minkowski in the form (3.1) to $\tilde{A}$ and $\tilde{B}$, we have

$$\left(1-\tfrac{t^2}{8}\right)^n \geq \mu(\tfrac{1}{2}(\tilde{A}+\tilde{B})) \geq \sqrt{\mu(\tilde{A})\mu(\tilde{B})} = \sqrt{\mathrm{P}[A]\,\mathrm{P}[B]} \geq \sqrt{\tfrac{1}{2}\mathrm{P}[B]}.$$

So

$$\mathrm{P}[B] \leq 2\left(1-\tfrac{t^2}{8}\right)^{2n} \leq 2e^{-t^2 n/4}.$$

□

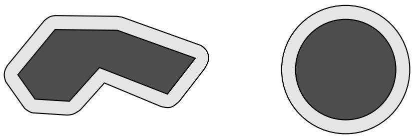
Exercises

1. Use Theorem 3.1.1 to show that any 1-dense set in the unit sphere S^{n-1} has at least $\frac{1}{2}e^{-n/8}$ points. $\boxed{2}$

3.2 Isoperimetric Inequalities and More on Concentration

The usual proof of Theorem 3.1.1 (measure concentration) has two steps: first, $\mathrm{P}[A_t]$ is bounded for A being the hemisphere (which is elementary calculus), and second, it is shown that among all sets A of measure $\frac{1}{2}$, the hemisphere has the smallest $\mathrm{P}[A_t]$. The latter result is an example of an *isoperimetric inequality.*

Before we formulate that this inequality, let us begin with the mother of all isoperimetric inequalities, the one for planar geometric figures. It states that among all planar geometric figures with a given perimeter, the circular disc has the largest possible area. (This is well-known but so easy to prove rigorously.) More general isoperimetric inequalities are usually formulated using the volume of a neighborhood instead of "perimeter". They claim that among all sets of a given volume in some metric space under consideration, a ball of that volume has the smallest volume of the t-neighborhood:



(In the picture, assuming that the dark areas are the same, then the light gray area is the smallest for the disc.) Letting $t \to 0$, one can get a statement involving the perimeter or surface area. But the formulation with t-neighborhood makes sense even in spaces where "surface area" is not defined; it suffices to have a metric and a measure on the considered space.

For the Euclidean space $\mathbf{R}^n$ with the Lebesgue measure, this "neighborhood" form of isoperimetric inequality claims that for any compact set $A \subset \mathbf{R}^d$ and any $t \geq 0$, we have $\mathrm{vol}(A_t) \geq \mathrm{vol}(B_t)$, where B is a ball of the same volume as A. Although we do not need this particular result in the further development, let us digress and mention a nice proof using the Brunn–Minkowski inequality (Theorem 1.2.2). By re-scaling, we may assume that B is the ball of unit radius. Then $A_t = A + tB$, and so

$$\begin{aligned}\mathrm{vol}(A_t) &= \mathrm{vol}(A + tB) \geq \left(\mathrm{vol}(A)^{1/n} + t\,\mathrm{vol}(B)^{1/n}\right)^n \\ &= (1+t)^n\,\mathrm{vol}(B) = \mathrm{vol}(B_t).\end{aligned}$$

For the sphere with the usual Euclidean metric inherited from $\mathbf{R}^n$, an r-ball is a spherical cap, i.e. an intersection of S^{n-1} with a halfspace. The isoperimetric inequality states that for all measurable sets $A \subseteq S^{n-1}$ and all $t \geq 0$, we have $\mathrm{P}[A_t] \geq \mathrm{P}[C_t]$, where C is a spherical cap with $\mathrm{P}[C] = \mathrm{P}[A]$. We are not going to prove this; no really simple proof seems to be known.

The measure concentration on the sphere (Theorem 3.1.1) is a rather direct consequence of this isoperimetric inequality, by the argument already indicated above. If $\mathrm{P}[A] = \frac{1}{2}$, then $\mathrm{P}[A_t] \geq \mathrm{P}[C_t]$, where C is a cap with $\mathrm{P}[C] = \frac{1}{2}$, i.e. a hemisphere. Thus, it suffices to estimate the measure of the complementary cap $S^{n-1} \setminus C_t$.

Gaussian concentration. There are many other metric probability spaces with measure concentration phenomena analogous to Theorem 3.1.1. Perhaps the most important one is $\mathbf{R}^n$ with the Euclidean metric and with the n-dimensional Gaussian measure γ given by

$$\gamma(A) = (2\pi)^{-n/2} \int_A e^{-\|x\|^2/2}\,\mathrm{d}x.$$

This is a probability measure on $\mathbf{R}^n$ corresponding to the n-dimensional normal distribution. Let $Z_1, Z_2, \ldots, Z_n$ be independent real random variables, each of them with the standard normal distribution $N(0,1)$, i.e. such that

$$\mathrm{Prob}[Z_i \leq z] = \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\,\mathrm{d}t.$$

Then the vector $(Z_1, Z_2, \ldots, Z_n) \in \mathbf{R}^n$ is distributed according to the measure γ. This γ is spherically symmetric: the density function $(2\pi)^{-n/2} e^{-\|x\|^2/2}$ only depends on the distance of x from the origin. The distance of a point chosen at random according to this distribution is sharply concentrated around $\sqrt{n}$, and in many respects, choosing a random point according to γ is similar to choosing a random point from the uniform distribution on the sphere $\sqrt{n}\, S^{n-1}$.

The isoperimetric inequality for the Gaussian measure claims that among all sets A with given $\gamma(A)$, a halfspace has the smallest possible measure of the t-neighborhood. By simple calculation, this yields the corresponding theorem about measure concentration for the Gaussian measure:

3.2.1 Theorem (Gaussian measure concentration). *Let a measurable set $A \subseteq \mathbf{R}^n$ satisfy $\gamma(A) \geq \frac{1}{2}$. Then $\gamma(A_t) \geq 1 - e^{-t^2/2}$.*

Note that the dimension does not appear in this inequality, and indeed the Gaussian concentration has infinite-dimensional versions as well.

Measure concentration on S^{n-1}, with slightly suboptimal constants, can be proved as an easy consequence of the Gaussian concentration.

Most of the results in the sequel obtained using measure concentration on the sphere can be derived from the Gaussian concentration as well. In more advanced applications, the Gaussian concentration is often technically preferable, but here we will stick to the perhaps more intuitive measure concentration on the sphere.

Other important "continuous" spaces with concentration results similar to Theorem 3.1.1 include the n-dimensional torus (the n-fold Cartesian product $S^1 \times \cdots \times S^1 \subset \mathbf{R}^{2n}$) and the group $SO(n)$ of all rotations around the origin in $\mathbf{R}^n$ (see Section 3.4 for more about $SO(n)$).

Discrete metric spaces. Similar concentration inequalities also hold in many discrete metric spaces encountered in combinatorics. One of the simplest examples is the n-dimensional Hamming cube $C_n = \{0,1\}^n$. The points are n-component vectors of 0s and 1s, and their Hamming distance is the number of positions where they differ. The "volume" of a set $A \subseteq \{0,1\}^n$ is defined as $\mathrm{P}[A] = \frac{1}{2^n}|A|$. An r-ball $B(r)$ is the set of all 0/1 vectors that differ from a given vector in at most r coordinates, and so its volume is $\mathrm{P}[B(r)] = 2^{-n}\left(1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{r}\right)$. The isoperimetric inequality for the Hamming cube, due to Harper, is exactly of the form announced above:

If $A \subseteq C_n$ *is any set with* $\mathrm{P}[A] \geq \mathrm{P}[B(r)]$ *then* $\mathrm{P}[A_t] \geq \mathrm{P}[B(r+t)]$.

Of course, if A is an r-ball, then A_t is an $(r+t)$-ball, and we have an equality. Suitable estimates (tail estimates for the binomial distribution in probability theory) then give an analogue of Theorem 3.1.1:

3.2.2 Theorem (Measure concentration for the cube). *Let* $A \subseteq C_n$ *satisfy* $\mathrm{P}[A] \geq \frac{1}{2}$. *Then* $1 - \mathrm{P}[A_t] \leq e^{-t^2/2n}$.

This is very similar to the situation for S^{n-1}, only the scaling is different: while the Hamming cube C_n has diameter n, and the interesting range of t is from about $\sqrt{n}$ to n, the sphere S^{n-1} has diameter 2, and the interesting t are in range from about $\frac{1}{\sqrt{n}}$ to 2.

Another significant discrete metric space with similar measure concentration is the space S_n of all permutations of $\{1,2,\ldots,n\}$ (i.e. bijective mappings $\{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$). The distance of two permutations p_1 and p_2 is $|\{i\colon p_1(i) \neq p_2(i)\}|$, and the measure is the usual uniform probability measure on S_n, where every single permutation has measure $\frac{1}{n!}$.

3.3 Concentration of Lipschitz Functions

Here we derive a form of the measure concentration which is very suitable for applications. It says that any Lipschitz function on a higher-dimensional sphere is tightly concentrated around its expectation. (Any measurable real function $f: S^{n-1} \to \mathbf{R}$ can be regarded as a random variable, and its expectation is given by $\mathbf{E}[f] = \int_{S^{n-1}} f(x)\, \mathrm{dP}(x)$.)

We recall that a mapping f between metric spaces is *C-Lipschitz* , where $C > 0$ is a real number, if the distance of $f(x)$ and $f(y)$ is never larger than the distance of x and y multiplied by C. We first show that a 1-Lipschitz function $f: S^{n-1} \to \mathbf{R}$ is concentrated around its median. The *median* of a real-valued function f is defined as

$$\operatorname{med}(f) = \sup\{t \in \mathbf{R}\colon \mathrm{P}[f \le t] \le \tfrac{1}{2}\}.$$

Here P is the considered probability measure on the domain of f; in our case, it is the normalized surface measure on S^{n-1}. The notation $\mathrm{P}[f \le t]$ is the usual probability-theory shorthand for $\mathrm{P}\left[\{x \in S^{n-1}\colon f(x) \le t\}\right]$. The following lemma looks obvious, but an actual proof is perhaps not so obvious:

3.3.1 Lemma. *Let $f: \Omega \to \mathbf{R}$ be a measurable function on a space Ω with a probability measure* P. *Then*

$$\mathrm{P}[f < \operatorname{med}(f)] \le \tfrac{1}{2} \text{ and } \mathrm{P}[f > \operatorname{med}(f)] \le \tfrac{1}{2}.$$

Proof. The first inequality can be derived from the σ-additivity of the measure P:

$$\begin{aligned}\mathrm{P}[f < \operatorname{med}(f)] &= \sum_{k=1}^{\infty} \mathrm{P}\left[\operatorname{med}(f) - \tfrac{1}{k-1} < f \le \operatorname{med}(f) - \tfrac{1}{k}\right] \\ &= \sup_{k \ge 1} \mathrm{P}\left[f \le \operatorname{med}(f) - \tfrac{1}{k}\right] \le \tfrac{1}{2}.\end{aligned}$$

The second inequality follows similarly. □

We are ready to prove that any 1-Lipschitz function $S^{n-1} \to \mathbf{R}$ is concentrated around its median:

3.3.2 Theorem (Lévy's Lemma). *Let $f: S^{n-1} \to \mathbf{R}$ be 1-Lipschitz. Then for all $t \in [0, 1]$,*

$$\mathrm{P}[f \ge \operatorname{med}(f) + t] \le 2e^{-t^2 n/2} \quad \textit{and} \quad \mathrm{P}[f \le \operatorname{med}(f) - t] \le 2e^{-t^2 n/2}.$$

For example, on 99% of S^{n-1}, the function f attains values deviating from $\mathrm{med}(f)$ by at most $3.5n^{-1/2}$.

Proof. We prove only the first inequality. Let $A = \{x \in S^{n-1}\colon f(x) \leq \mathrm{med}(f)\}$. By Lemma 3.3.1, $\mathrm{P}[A] \geq \frac{1}{2}$. Since f is 1-Lipschitz, we have $f(x) \leq \mathrm{med}(f) + t$ for all $x \in A_t$. Therefore, by Theorem 3.1.1, we get $\mathrm{P}[f \geq \mathrm{med}(f) + t] \leq \mathrm{P}\left[S^{n-1} \setminus A_t\right] \leq 2e^{-t^2n/2}$. □

The median is generally difficult to compute. But for a 1-Lipschitz function, it cannot be too far from the expectation, which is usually easier to estimate:

3.3.3 Proposition. *Let $f\colon S^{n-1} \to \mathbf{R}$ be 1-Lipschitz. Then*

$$|\,\mathrm{med}(f) - \mathbf{E}[f]\,| \leq 12n^{-1/2}.$$

Proof.

$$\begin{aligned}|\,\mathrm{med}(f) - \mathbf{E}[f]\,| &\leq \mathbf{E}[|f - \mathrm{med}(f)|] \leq \sum_{k=0}^{\infty} \tfrac{k+1}{\sqrt{n}} \mathrm{P}\left[|f - \mathrm{med}(f)| \geq \tfrac{k}{\sqrt{n}}\right] \\ &\leq n^{-1/2} \sum_{k=0}^{\infty} (k+1) \cdot 4e^{-k^2/2} \leq 12n^{-1/2}\end{aligned}$$

(the numerical estimate of the last sum is not important; it is important that it converges to some constant, which is obvious). □

We derive a consequence of Lévy's Lemma on finding k-dimensional subspaces where a given Lipschitz function is almost constant. But first we need some notions and results.

Random rotations and random subspaces. In subsequent considerations, we will need to speak about a random k-dimensional (linear) subspace of $\mathbf{R}^n$. We thus need to specify a probability measure on the set of all k-dimensional linear subspaces of $\mathbf{R}^n$ (the so-called *Grassman manifold* or *Grassmanian*). An elegant way of doing this is via random rotations.

A rotation ρ is an isometry of $\mathbf{R}^n$ fixing the origin and preserving the orientation. In algebraic terms, ρ is a linear mapping $x \mapsto Ax$ given by an orthonormal matrix A with determinant 1. The result of performing the rotation ρ on the standard orthonormal basis $(e_1, \ldots, e_n)$ in $\mathbf{R}^n$ is an n-tuple of orthonormal vectors, and these vectors are the columns of A.

The group of all rotations in $\mathbf{R}^n$ around the origin with the operation of composition (corresponding to multiplication of the matrices) is denoted by $SO(n)$, which stands for the *s*pecial *o*rthogonal group. With the natural topology (obtained by regarding the corresponding matrices as points in $\mathbf{R}^{n^2}$), it is a compact group. By a general theorem in the theory of topological groups, there is a unique Borel probability measureon $SO(n)$ (the *Haar measure*) that is invariant under the action of the elements of $SO(n)$. Here is a more concrete description of this probability measure. To obtain a random rotation ρ, we first choose a vector $a_1 \in S^{n-1}$ uniformly at random. Then we pick a_2 orthogonal to a_1; this a_2 is drawn from the uniform distribution on the $(n-2)$-dimensional sphere that is the intersection of S^{n-1} with the hyperplane perpendicular to a_1 and passing through 0. Then a_3 is chosen from the unit sphere within the $(n-2)$-dimensional subspace perpendicular to a_1 and a_2, and so on.

In the sequel, we will need only the following intuitively obvious fact about a random rotation $\rho \in SO(n)$: for any fixed $u \in S^{n-1}$, $\rho(u)$ is a random vector of S^{n-1}. Therefore, if $u \in S^{n-1}$ is fixed, $A \subseteq S^{n-1}$ is measurable, and $\rho \in SO(n)$ is random, then the probability of $\rho(u) \in A$ equals $\mathrm{P}[A]$.

Let L_0 be the k-dimensional subspace spanned by the first k coordinate vectors. A random k-dimensional linear subspace $L \subset \mathbf{R}^n$ can be defined as $\rho(L_0)$, where $\rho \in SO(n)$ is a random rotation.

By Lévy's Lemma, a 1-Lipschitz function on S^{n-1} is "almost constant" on a subset A occupying almost all of S^{n-1}. Generally we do not know anything about the shape of such an A. But the next proposition shows that the almost-constant behavior can be guaranteed on the intersection of S^{n-1} with a linear subspace of $\mathbf{R}^n$ of relatively large dimension.

3.3.4 Proposition (Subspace where a Lipschitz function is almost constant). *Let $f\colon S^{n-1} \to \mathbf{R}$ be a 1-Lipschitz function and let $\delta \in (0,1]$. Then there is a linear subspace $L \subseteq \mathbf{R}^n$ such that all values of f restricted to $S^{n-1} \cap L$ are in the interval* $[\mathrm{med}(f) - \delta, \mathrm{med}(f) + \delta]$ *and*

$$\dim L \geq \frac{\delta^2}{8\log(8/\delta)} \cdot n - 1.$$

Proof. Let L_0 be the subspace spanned by the first $k = \lceil n\delta^2/8 \log \frac{8}{\delta} - 1 \rceil$ coordinate vectors. Fix a $\frac{\delta}{2}$-net N (as defined above Lemma 2.1.1) in $S^{n-1} \cap$

L_0. Let $\rho \in SO(n)$ be a random rotation. For $x \in N$, $\rho(x)$ is a random point, and so by Lévy's Lemma, the probability that $|f(\rho(x)) - \mathrm{med}(f)| > \frac{\delta}{2}$ for at least one point $x \in N$ is no more than $|N| \cdot 4e^{-\delta^2 n/8}$. Using the bound $|N| \leq (\frac{8}{\delta})^k$ from Lemma 2.1.1, we calculate that with a positive probability, $|f(y) - \mathrm{med}(f)| \leq \frac{\delta}{2}$ for all $y \in \rho(N)$.

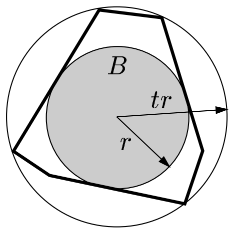
We choose a ρ with this property and let $L = \rho(L_0)$. For each $x \in S^{n-1} \cap L$, there is some $y \in \rho(N)$ with $\|x - y\| \leq \frac{\delta}{2}$, and since f is 1-Lipschitz, we obtain $|f(x) - \mathrm{med}(f)| \leq |f(x) - f(y)| + |f(y) - \mathrm{med}(f)| \leq \delta$. □

Exercises

1. Derive the measure concentration on the sphere (Theorem 3.1.1) from Lévy's Lemma 3.3.2. $\boxed{2}$

3.4 Almost Spherical Sections: the First Steps

For a real number $t \geq 1$, we call a convex body K *t-almost spherical* if it contains a (Euclidean) ball B of some radius r and it is contained in the concentric ball of radius tr.



Given a centrally symmetric convex body $K \subset \mathbf{R}^n$ and $\varepsilon > 0$, we are interested in finding a k-dimensional (linear) subspace L, with k as large as possible, such that the "section" $K \cap L$ is $(1+\varepsilon)$-almost spherical.

Ellipsoids. First we deal with ellipsoids, where the existence of large spherical sections is not very surprising. But in the sequel it gives us additional freedom: instead of looking for a $(1+\varepsilon)$-spherical section of a given convex body, we can as well look for a $(1+\varepsilon)$-ellipsoidal section, while losing only a factor of at most 2 in the dimension. This means that we are free to transform a given body by any (nonsingular) affine map, which is often convenient. Let us remark that in the local theory of Banach spaces, almost-ellipsoidal sections are usually as good as almost-spherical ones, and so the following lemma is often not even mentioned.

3.4.1 Lemma (Ellipsoids have large spherical sections). *For any $(2k-1)$-dimensional ellipsoid E, there is a k-flat L passing through the center of E such that $E \cap L$ is a Euclidean ball.*

Proof. Let $E = \{x \in \mathbf{R}^{2k-1}\colon \sum_{i=1}^{2k-1} \frac{x_i^2}{a_i^2} \leq 1\}$ with $0 < a_1 \leq a_2 \leq \cdots \leq a_{2k-1}$. We define the k-dimensional linear subspace L by a system of $k-1$ linear equations. The ith equation is $\alpha_i x_i = \sqrt{1-\alpha_i^2}\, x_{2k-i}$, $i = 1, 2, \ldots, k-1$, where $\alpha_i \in [0,1]$ is a root of the equation

$$\frac{\alpha_i^2}{a_i^2} + \frac{1-\alpha_i^2}{a_{2k-i}^2} = \frac{1}{a_k^2}$$

(a root exists because $a_i \leq a_k \leq a_{2k-i}$). All of this is chosen so that for $x \in L$, we have $\frac{x_i^2}{a_i^2} + \frac{x_{2k-i}^2}{a_{2k-i}^2} = \frac{1}{a_k^2}(x_i^2 + x_{2k-i}^2)$. It follows that for $x \in L$, we have $x \in E$ if and only if $\|x\| \leq a_k$, and so $E \cap L$ is a ball of radius a_k. The reader is invited to find a geometric meaning of this proof and/or express it in the language of eigenvalues. $\square$

To make formulas simpler, we consider only the case $\varepsilon = 1$ in the rest of this section. An arbitrary $\varepsilon > 0$ can always be handled very similarly.

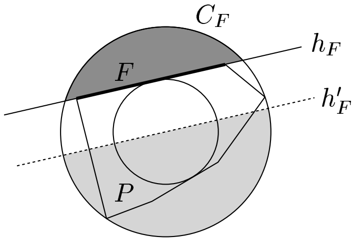
The cube. The cube $[-1,1]^n$ is a good test case for finding almost-spherical sections; it seems hard to imagine how a cube could have very round slices. In some sense, this intuition is not totally wrong, since the almost-spherical sections of a cube can have only logarithmic dimension, as we verify next. (But the n-dimensional crosspolytope has $(1+\varepsilon)$-spherical sections of dimension as high as $c(\varepsilon)n$ and yet it does not look any rounder than the cube; so much for the intuition.)

The intersection of the cube with a k-dimensional linear subspace of $\mathbf{R}^n$ is a k-dimensional convex polytope with at most $2n$ facets. But we have

3.4.2 Lemma. *Let P be a k-dimensional 2-almost spherical convex polytope. Then P has at least $\frac{1}{2}\,e^{k/8}$ facets.*

Therefore, any 2-almost spherical section of the cube has dimension at most $O(\log n)$.

Proof of Lemma 3.4.2. After a suitable affine transform, we may assume $\frac{1}{2}B^k \subseteq P \subseteq B^k$. Each point $x \in S^{k-1}$ is separated from P by one of the facet hyperplanes. For each facet F of P, the facet hyperplane h_F cuts off a cap C_F of S^{k-1}, and these caps together cover all of S^{k-1}. The cap C_F is at distance at least $\frac{1}{2}$ from the hemisphere defined by the hyperplane h'_F parallel to h_F and passing through 0.



By Theorem 3.1.1 (measure concentration), we have $\mathrm{P}[C_F] \leq 2e^{-k/8}$. $\square$

Next, we show that the n-dimensional cube actually does have 2-almost spherical sections of dimension $\Omega(\log n)$. It is not difficult to construct a k-dimensional 2-almost spherical polytope with 4^k facets. First, we note that if P is a convex polytope with $B^k \subset P \subset tB^k$ then the dual polytope P^* satisfies $\frac{1}{t}B^k \subset P^* \subset B^k$ (Exercise 1). So it suffices to construct a k-dimensional 2-almost spherical polytope with 4^k *vertices*, and this has been done in Section 2.3: we can take any 1-net in S^{k-1} as the vertex set. (Let us remark that an exponential lower bound for the number of vertices also follows from Theorem 2.2.1.)

By at most doubling the number of facets, we may assume that our k-dimensional 2-almost spherical polytope is centrally symmetric. It remains to observe that *every bounded k-dimensional centrally symmetric convex polytope P with $2n$ facets is the affine image of the section $[-1,1]^n \cap L$ for a suitable k-dimensional linear subspace $L \subseteq \mathbf{R}^n$.* Indeed, a such a P can be expressed as the intersection $\bigcap_{i=1}^n \{x \in \mathbf{R}^k : |\langle a_i, x\rangle| \leq 1\}$, where $\pm a_1, \ldots, \pm a_n$ are suitably normalized normal vectors of the facets of P. Let $f: \mathbf{R}^k \to \mathbf{R}^n$ be the linear map given by

$$f(x) = (\langle a_1, x\rangle, \langle a_2, x\rangle, \ldots, \langle a_n, x\rangle).$$

Since P is bounded, the a_i span the whole $\mathbf{R}^k$ and so f has rank k. Consequently, its image $L = f(\mathbf{R}^k)$ is a k-dimensional subspace of $\mathbf{R}^n$. We have

$P = f^{-1}([-1,1]^n)$, and so the intersection $[-1,1]^n \cap L$ is the affine image of P.

Together with the existence of 2-almost spherical centrally symmetric polytopes with exponentially many facets, this implies that the n-dimensional cube has 2-almost elliptical sections of dimension $\Omega(\log n)$ (and by Lemma 3.4.1, 2-almost spherical sections exist as well).

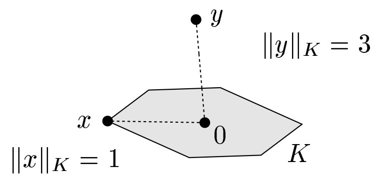
Next, we make preparatory steps for finding almost-spherical sections of arbitrary centrally symmetric convex bodies. These considerations are most conveniently formulated in the language of norms.

Reminder on norms. We recall that a *norm* on a real vector space Z is a mapping that assigns a nonnegative real number $\|x\|_Z$ to each $x \in Z$, such that $\|x\|_Z = 0$ implies $x = 0$, $\|\alpha x\|_Z = |\alpha| \cdot \|x\|_Z$ for all $\alpha \in \mathbf{R}$, and the triangle inequality holds: $\|x+y\|_Z \leq \|x\|_Z + \|y\|_Z$. (Since we have reserved $\|\cdot\|$ for the Euclidean norm, we write other norms with various subscripts, or occasionally we use the symbol $|\cdot|$.)

Norms are in one-to-one correspondence with closed bounded convex bodies symmetric about 0 and containing 0 in their interior. Here we only need one direction of this correspondence: given a convex body K with the listed properties, we assign to it the norm $\|\cdot\|_K$ given by

$$\|x\|_K = \min\{t > 0\colon \tfrac{x}{t} \in K\}.$$

Here is an illustration:



It is easy to verify the axioms of the norm (the convexity of K is needed for the triangle inequality). The body K is the unit ball of the norm $\|\cdot\|_K$. The norm of points *decreases* by blowing up the body K.

General body: the first attempt. Let $K \subset \mathbf{R}^n$ be a convex body defining a norm (i.e. closed, bounded, symmetric, 0 in the interior). Let us define the function $f_K\colon S^{n-1} \to \mathbf{R}$ as the restriction of the norm $\|\cdot\|_K$ on S^{n-1}; that is, $f_K(x) = \|x\|_K$. We note that K is t-almost spherical if (and only if) there is a number $a > 0$ such that $a \leq f(x) \leq ta$ for all $x \in S^{n-1}$. So for finding a large almost-spherical section of K, we need a linear subspace L such that f does not vary too much on $S^{n-1} \cap L$, and this is where Proposition 3.3.4, about subspaces where a Lipschitz function is almost constant, comes in.

Of course, that proposition has its assumptions, and one of them is that f_K be 1-Lipschitz. A sufficient condition for that is that should K contain the unit ball:

3.4.3 Observation. *Suppose that the convex body K contains the R-ball $B(0, R)$. Then $\|x\|_K \leq \frac{1}{R}\|x\|$ for all x and the function $x \mapsto \|x\|_K$ is $\frac{1}{R}$-Lipschitz with respect to the Euclidean metric.* □

Then we can easily prove

3.4.4 Proposition. *Let $K \subset \mathbf{R}^n$ be a convex body defining a norm and such that $B^n \subseteq K$, and let $m = \mathrm{med}(f_K)$, where f_K is as above. Then there exists a 2-almost-spherical section of K of dimension at least*

$$\Omega\left(\frac{nm^2}{\log(24/m)}\right).$$

Proof. By Observation 3.4.3, f_K is 1-Lipschitz. Let us set $\delta = \frac{m}{3}$ (note that $B^n \subseteq K$ also implies $m \leq 1$). Proposition 3.3.4 shows that there is a subspace L such that $f_K \in [\frac{2}{3}m, \frac{4}{3}m]$ on $S^{n-1} \cap L$, where

$$\dim L = \Omega\left(\frac{n\delta^2}{\log(8/\delta)}\right) = \Omega\left(\frac{nm^2}{\log(24/m)}\right). \tag{3.2}$$

The section $K \cap L$ is 2-almost spherical. □

A slight improvement. It turns out that the factor $\log(24/m)$ in the just proved result can be eliminated by a refined argument, which uses the fact that f_K comes from a norm.

3.4.5 Theorem. *With the assumptions as in Proposition 3.4.4, a 2-almost spherical section exists of dimension at least βnm^2, where $\beta > 0$ is an absolute constant.*

Proof. The main new observation is that for our f_K, we can afford a much less dense net N in the proof of Proposition 3.3.4. Namely, it suffices to let N be a $\frac{1}{5}$-net in S^{k-1}, where $k = \lceil \beta m^2 n \rceil$.

If $\beta > 0$ is sufficiently small, Lévy's Lemma gives the existence of a rotation ρ such that $\frac{14}{15}m \le f_K(y) \le \frac{16}{15}m$ for all $y \in \rho(N)$; this is exactly as in the proof of Proposition 3.3.4. It remains to verify $\frac{2}{3}m \le f_K(x) \le \frac{4}{3}m$ for all $x \in S^{n-1} \cap L$, where $L = \rho(L_0)$. This is implied by the following claim with $a = \frac{16}{15}m$ and $|\cdot| = \|\cdot\|_K$:

Claim. *Let N be a $\frac{1}{5}$-net in S^{k-1} with respect to the Euclidean metric, and let $|\cdot|$ be a norm on $\mathbf{R}^k$ satisfying $\frac{7}{8}a \le |y| \le a$ for all $y \in N$ and for some number $a > 0$. Then $\frac{5}{8}a \le |x| \le \frac{5}{4}a$ for all $x \in S^{k-1}$.*

To prove the claim, we begin with the upper bound (this is where the new trick lies). Let $M = \max\{|x|: x \in S^{k-1}\}$ and let $x_0 \in S^{k-1}$ be a point where M is attained. Choose a $y_0 \in N$ at distance at most $\frac{1}{5}$ from x_0, and let $z = (x_0 - y_0)/\|x_0 - y_0\|$ be the unit vector in the direction of $x_0 - y_0$. Then $M = |x_0| \le |y_0| + |x_0 - y_0| \le a + \|x_0 - y_0\| \cdot |z| \le a + \frac{1}{5}M$. The resulting inequality $M \le a + \frac{1}{5}M$ yields $M \le \frac{5}{4}a$.

The lower bound is now routine: if $x \in S^{k-1}$ and $y \in N$ is at distance at most $\frac{1}{5}$ from it, $|x| \le |y| - |x - y| \ge \frac{7}{8}a - \frac{1}{5} \cdot \frac{5}{4}a \ge \frac{5}{8}a$. The claim, as well as Theorem 3.4.5, are proved. $\square$

Theorem 3.4.5 yields almost-spherical sections of K *provided that* we can estimate $\mathrm{med}(f_K)$ (after re-scaling K so that $B^n \subseteq K$). We must warn that this in itself does not yet give almost spherical sections for every K (Dvoretzky's Theorem), and another twist is needed, shown in Section 3.6. But in order to reap some benefits from the hard work done up until now, we first explain an application to convex polytopes.

Exercises

1. Let K be a convex body containing 0 in its interior. Check that $K \subseteq B^n$ if and only if $B^n \subseteq K^*$ (recall that $K^* = \{x \in \mathbf{R}^k: \langle x, y\rangle \le 1 \text{ for all } y \in K\}$). Derive that if $B^k \subset K \subset tB^k$ then $\frac{1}{t}B^k \subset K^* \subset B^k$. $\boxed{2}$

3.5 Many Faces of Symmetric Polytopes

Can an n-dimensional convex polytope have both few vertices and few facets? Yes, an n-simplex has $n+1$ vertices and $n+1$ facets. What about a *centrally symmetric polytope*? The n-dimensional cube has only $2n$ facets but 2^n vertices. Its dual, the crosspolytope (regular octahedron for $n = 3$)

has few vertices but many facets. It turns out that every centrally symmetric polytope has many facets or many vertices.

3.5.1 Theorem. *There is a constant $\alpha > 0$ such that for any centrally symmetric n-dimensional convex polytope P, we have $\log f_0(P) \cdot \log f_{n-1}(P) \geq \alpha n$ (recall that $f_0(P)$ denotes the number of vertices and $f_{n-1}(P)$ the number of facets).*

For the cube, the expression $\log f_0(P) \cdot \log f_{n-1}(P)$ is about $n \log n$, which is even slightly larger than the lower bound in the theorem. However, polytopes can be constructed with both $\log f_0(P)$ and $\log f_{n-1}(P)$ bounded by $O(\sqrt{n})$ (Exercise 1).

Proof of Theorem 3.5.1. We use the dual polytope P^* with $f_0(P) = f_{n-1}(P^*)$, and we prove the theorem in the equivalent form $\log f_{n-1}(P^*) \cdot \log f_{n-1}(P) \geq \alpha n$.

John's Lemma (Theorem 2.4.1) claims that for any symmetric convex body K, there exists a (non-singular) linear map that transforms K into a $\sqrt{n}$-almost spherical body. We can thus assume that the considered n-dimensional polytope P is $\sqrt{n}$-almost spherical (this is crucial for the proof).

After re-scaling, we may suppose $B^n \subset P \subset \sqrt{n}\, B^n$. Letting $m = \mathrm{med}(f_P)$, where f_P is the restriction of $\|\cdot\|_P$ on S^{n-1} as usual, Theorem 3.4.5 tells us that there is a linear subspace L of $\mathbf{R}^n$ with $P \cap L$ being 2-almost spherical and with $\dim(L) = \Omega(nm^2)$. Thus, since any k-dimensional 2-almost spherical polytope has $e^{\Omega(k)}$ facets, we have $\log f_{n-1}(P) = \Omega(nm^2)$.

Now we look at P^*. Since $B^n \subset P \subset \sqrt{n}\, B^n$, by Exercise 3.4.1 we have $n^{-1/2} B^n \subset P^* \subset B^n$. In order to apply Theorem 3.4.5, we set $\tilde{P} = \sqrt{n}\, P^*$, and obtain a 2-almost spherical section $\tilde{L}$ of $\tilde{P}$ of dimension $\Omega(n\tilde{m}^2)$, where $\tilde{m} = \mathrm{med}(f_{\tilde{P}})$. This implies $\log f_{n-1}(P^*) = \Omega(n\tilde{m}^2)$.

It remains to observe the following inequality:

3.5.2 Lemma. *Let P be a polytope in $\mathbf{R}^n$ defining a norm and let P^* be the dual polytope. Then we have $\mathrm{med}(f_P)\,\mathrm{med}(f_{P^*}) \geq 1$.*

We leave the easy proof as Exercise 2. Since $\tilde{m} = \mathrm{med}(f_{P^*})/\sqrt{n}$, we finally obtain

$$\log f_{n-1}(P) \cdot \log f_{n-1}(P^*) = \Omega(n^2 m^2 \tilde{m}^2) = \Omega(n\,\mathrm{med}(f_P)^2\,\mathrm{med}(f_{P^*})^2) = \Omega(n).$$

This concludes the proof of Theorem 3.5.1. $\square$

Exercises

1. Construct an n-dimensional convex polytope P with $f_0(P) = \Omega(\sqrt{n})$ and $f_{n-1}(P) = \Omega(\sqrt{n})$, thereby demonstrating that Theorem 3.5.1 is asymptotically optimal. Start with the interval $[0,1] \subset \mathbf{R}^1$, and alternate the operations $(\cdot)^*$ (passing to the dual polytope) and $\times$ (Cartesian product) suitably. $\boxed{3}$

   The polytopes obtained from $[0,1]$ by a sequence of these operations are called *Hammer polytopes*, and they form an important class of examples.

2. Let K be a bounded centrally symmetric convex body in $\mathbf{R}^n$ containing 0 in its interior and let K^* be the dual body.

   (a) Show that $\|x\|_K \cdot \|x\|_{K^*} \geq 1$ for all $x \in S^{n-1}$. $\boxed{1}$

   (b) Let $f, g\colon S^{n-1} \to \mathbf{R}$ be (measurable) functions with $f(x)g(x) \geq 1$ for all $x \in S^{n-1}$. Show that $\operatorname{med}(f)\operatorname{med}(g) \geq 1$. $\boxed{2}$

3.6 Dvoretzky's Theorem

Here is the remarkable Ramsey-type result in high-dimensional convexity promised at the beginning of this chapter.

3.6.1 Theorem (Dvoretzky's Theorem). *For any natural number k and any real $\varepsilon > 0$, there exists an integer $n = n(k, \varepsilon)$ with the following property. For any n-dimensional centrally symmetric convex body $K \subseteq \mathbf{R}^n$, there exists a k-dimensional linear subspace $L \subseteq \mathbf{R}^n$, such that the section $K \cap L$ is $(1+\varepsilon)$-almost spherical.*

The best known estimates give $n(k, \varepsilon) = e^{O(k/\varepsilon^2)}$.

Thus, no matter how "edgy" may a high-dimensional K be, there is always a slice of not too small dimension that is almost a Euclidean ball. Another way of expressing the statement is that any normed space of a sufficiently large dimension contains a large subspace on which the norm is very close to the Euclidean norm (with a suitable choice of a coordinate system in the subspace). Note that the Euclidean norm is the *only* norm with this universal property, since all sections of the Euclidean ball are again Euclidean balls.

As we saw in Section 3.4, the n-dimensional cube shows that the largest dimension of a 2-almost spherical section is only $O(\log n)$ in the worst case.

The assumption that K be symmetric can in fact be omitted; it suffices to require that 0 is an interior point of K. The proof of this more general version is not much more difficult than the one shown below.

We prove Dvoretzky's Theorem only for $\varepsilon = 1$, since in Section 3.4 we prepared the tools for this particular setting. But the general case is not very different.

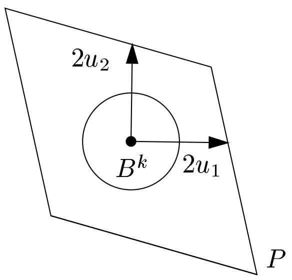
Preliminary considerations. Since affine transforms of K are practically for free in view of Lemma 3.4.1, we may assume that $B^n \subseteq K \subseteq \sqrt{n}\, B^n$ by John's Lemma (Theorem 2.4.1). So the norm induced by K satisfies $n^{-1/2}\|x\| \le \|x\|_K \le \|x\|$ for all x. If f_K is the restriction of $\|\cdot\|_K$ to S^{n-1}, we have the obvious bound $\mathrm{med}(f_K) \ge n^{-1/2}$. Immediate application of Theorem 3.4.5 shows the existence of a 2-almost spherical section of K of dimension $\Omega(n\,\mathrm{med}(f_K)^2) = \Omega(1)$—so this approach gives nothing at all! On the other hand, it *just* fails, and a small improvement in the order of magnitude of the lower bound for $\mathrm{med}(f_K)$ already yields Dvoretzky's Theorem.

We will not try to improve the estimate for $\mathrm{med}(f_K)$ directly. Instead, we find a relatively large subspace $Z \subset \mathbf{R}^n$ such that the section $K \cap Z$ can be enclosed in a not too large parallelotope P. Then we estimate, by direct computation, $\mathrm{med}(f_P)$ (over the unit sphere in Z).

The selection of the subspace Z is known as the *Dvoretzky–Rogers Lemma*. We present a version with a particularly simple proof, where we need a rescaling of K and where $\dim Z \approx n/\log n$. (For our purposes, we would suffice with even much weaker estimates, say $\dim Z \ge n^{\delta}$ for some fixed $\delta > 0$.)

3.6.2 Lemma (A version of the Dvoretzky–Rogers Lemma). *Let $K \subset \mathbf{R}^n$ be a centrally symmetric convex body. Then there exist a linear subspace $Z \subset \mathbf{R}^n$ of dimension $k = \lfloor \frac{n}{\log_2 n} \rfloor$, an orthonormal basis $u_1, u_2, \ldots, u_k$ of Z, and a nonsingular linear transform T of $\mathbf{R}^n$ such that if we let $\tilde{K} = T(K) \cap Z$ then $\|x\|_{\tilde{K}} \le \|x\|$ for all $x \in Z$ and $\|u_i\|_{\tilde{K}} \ge \frac{1}{2}$ for all $i = 1, 2, \ldots, k$.*

Geometrically, the lemma asserts that $\tilde{K}$ is sandwiched between the unit ball B^k and a parallelotope P as in the picture:



(The lemma claims that the points $2u_i$ are outside of K or on its boundary, and P is obtained by separating these points from K by hyperplanes.)

Proof. By John's Lemma (Theorem 2.4.1), we may assume $B^n \subseteq K \subseteq tB^n$, where $t = \sqrt{n}$. Interestingly, the full power of John's Lemma is not needed here; the same proof works with, say, $t = n$ or $t = n^{10}$, only the bound for k would become worse by a constant factor.

Let $X_0 = \mathbf{R}^n$ and $K_0 = K$. Here is the rough idea. The current body K_i is enclosed between an inner ball and an outer ball. Either K_i approaches the inner ball sufficiently closely at "many" places, and in this case we can construct the desired $u_1, \ldots, u_k$, or it stays away from the inner ball on a "large" subspace. In the latter case, we can restrict to that subspace and inflate the inner ball. But since the outer ball remains the same the inflation of the inner ball cannot continue indefinitely. A precise argument follows; for notational reasons, instead of inflating the inner ball, we will shrink the body and the outer ball.

We consider the following condition:

> (*) Each linear subspace $Y \subseteq X_0$ with $\dim(X_0) - \dim(Y) < k$ contains a vector u with $\|u\| = 1$ and $\|u\|_{K_0} \geq \frac{1}{2}$.

This condition may or may not be satisfied. If it holds, we construct the orthonormal basis $u_1, u_2, \ldots, u_k$ by an obvious induction. If it is not satisfied, we obtain a subspace X_1 of dimension greater than $n - k$ such that $\|x\|_{K_0} \leq \frac{1}{2}\|x\|$ for all $x \in X_1$. Thus, $K_0 \cap X_1$ is twice "more spherical" than K_0. Setting $K_1 = \frac{1}{2}(K_0 \cap X_1)$, we have

$$\tfrac{2}{t}\, \|\cdot\| \leq \|\cdot\|_{K_1} \leq \|\cdot\|.$$

We again check the condition (*) with X_1 and K_1 instead of X_0 and K_0. If it holds we find the u_i within X_1, and if it does not, we obtain a subspace

X_2 of dimension greater than $n-2k$, etc. After the ith step, we have

$$\tfrac{2^i}{t}\,\|\cdot\| \le \|\cdot\|_{K_i} \le \|\cdot\|.$$

This construction cannot proceed all the way to step $i = i_0 = \lfloor \log_2 n \rfloor$, since $2^{i_0} > t = \sqrt{n}$. Thus, the condition (*) must hold for X_{i_0-1} at the latest. We have $\dim X_{i_0-1} > n-(i_0-1)k \ge k$, and so the required basis $u_1,\ldots,u_k$ can be constructed. □

The parallelotope is no worse than the cube. From now on, we work within the subspace Z as in Lemma 3.6.2. For convenient notation, we assume that Z is the whole $\mathbf{R}^n$ and K is as $\tilde{K}$ in the above lemma, i.e. $B^n \subseteq K$ and $\|u_i\|_K \ge \frac{1}{2}$, $i = 1,2,\ldots,n$, where $u_1,\ldots,u_n$ is an orthonormal basis of $\mathbf{R}^n$. (Note that the reduction of the dimension from n to $n/\log n$ is nearly insignificant for the estimate of $n(k,\varepsilon)$ in Dvoretzky's Theorem.)

The goal is to show $\mathrm{med}(f_K) = \Omega\bigl(\sqrt{(\log n)/n}\,\bigr)$, where f_K is $\|\cdot\|_K$ restricted to S^{n-1}. Instead of estimating $\mathrm{med}(f_K)$, we will bound the expectation $\mathbf{E}[f_K]$. Since f_K is 1-Lipschitz (as $B^n \subseteq K$), the difference $|\,\mathrm{med}(f_K) - \mathbf{E}[f_K]\,|$ is $O(n^{-1/2})$ by Proposition 3.3.3, which is negligible compared to the lower bound we are heading for.

We have $\|\cdot\|_K \ge \|\cdot\|_P$, where P is the parallelotope as in the illustration to the Lemma 3.6.2. So we actually bound $\mathbf{E}[f_P]$ from below.

First we show, by an averaging trick, that $\mathbf{E}[f_P] \ge \mathbf{E}[f_C]$, where $f_C(x) = \frac{1}{2}\|x\|_\infty = \frac{1}{2}\max_i |x_i|$ is the norm induced by the cube C of side 4. The idea of the averaging is to consider, together with a point $x = \sum_{i=1}^n \alpha_i u_i \in S^{n-1}$, the 2^n points of the form $\sum_{i=1}^n \varepsilon_i\alpha_i u_i$, where $\varepsilon \in \{-1,1\}^n$ is a vector of signs. For any measurable function $f_P\colon S^{n-1} \to \mathbf{R}$, we have

$$\begin{aligned}\int_{S^{n-1}} \sum_{\varepsilon\in\{-1,1\}^n} f_P\Bigl(\sum_{i=1}^n \varepsilon_i\alpha_i u_i\Bigr)\,\mathrm{dP}(\alpha) &= \sum_{\varepsilon}\int_{S^{n-1}} f_P\Bigl(\sum_{i=1}^n \varepsilon_i\alpha_i u_i\Bigr)\,\mathrm{dP}(\alpha)\\ &= 2^n\int_{S^{n-1}} f_P(x)\,\mathrm{dP}(x) = 2^n\,\mathbf{E}[f_P]\,.\end{aligned}$$

The following lemma with $v_i = \alpha_i u_i$ and with $|\cdot| = \|\cdot\|_P$ then implies that the integrand on the left-hand side is always at least $2^n \max_i \|\alpha_i u_i\|_P \ge 2^n \cdot \frac{1}{2}\max_i |\alpha_i|$, and so indeed $\mathbf{E}[f_P] \ge \mathbf{E}[f_C]$.

3.6.3 Lemma. *Let $v_1, v_2, \ldots, v_n$ be arbitrary vectors in a normed space* *with norm* $|\cdot|$. *Then*

$$\sum_{\varepsilon\in\{-1,1\}^n}\left|\sum_{i=1}^n \varepsilon_i v_i\right| \ge 2^n \max_i |v_i|.$$

The proof is left as Exercise 1. It remains to estimate $\mathbf{E}[f_C]$ from below.

3.6.4 Lemma. *For a suitable positive constant* c *and for all* n *we have*

$$\mathbf{E}[f_C] = \tfrac{1}{2}\int_{S^{n-1}} \|x\|_\infty \,\mathrm{dP}(x) \ge c\sqrt{\tfrac{\log n}{n}},$$

where $\|x\|_\infty = \max_i |x_i|$ *is the* ℓ_∞ *(or maximum) norm.*

Note that once this lemma is proved, Dvoretzky's Theorem (with $\varepsilon = 1$) follows from what we have already done and from Theorem 3.4.5.

Proof of Lemma 3.6.4. There are various proofs; a neat way is based on the generally useful fact that the n-dimensional normal distribution is spherically symmetric around the origin. We use a probabilistic terminology. Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables, each of them with the standard normal distribution $N(0,1)$. As was mentioned in Section 3.1, the random vector $Z = (Z_1, Z_2, \ldots, Z_n)$ has a spherically symmetric (Gaussian) distribution, and consequently, the random variable $\frac{Z}{\|Z\|}$ is uniformly distributed in S^{n-1}. Thus

$$\mathbf{E}[f_C] = \tfrac{1}{2}\,\mathbf{E}\left[\frac{\|Z\|_\infty}{\|Z\|}\right].$$

We show that, first, we have $\|Z\| \le \sqrt{3n}$ with probability at least $\frac{2}{3}$, and second, for a suitable constant $c_1 > 1$, $\|Z\|_\infty \ge c_1\sqrt{\log n}$ holds with probability at least $\frac{2}{3}$. It follows that both these events occur simultaneously with probability at least $\frac{1}{3}$, and so $\mathbf{E}[f_C] \ge c\sqrt{\log n/n}$ as claimed.

As for the Euclidean norm $\|Z\|$, we find $\mathbf{E}\left[\|Z\|^2\right] = n\mathbf{E}\left[Z_1^2\right] = n$, since an $N(0,1)$ random variable has variance 1. By Markov's inequality, we obtain $\mathrm{Prob}\left[\|Z\| \ge \sqrt{3n}\,\right] = \mathrm{Prob}\left[\|Z\|^2 \ge 3\mathbf{E}\left[\|Z\|^2\right]\right] \le \frac{1}{3}$.

Further, by the independence of the Z_i we have

$$\begin{aligned}\mathrm{Prob}[\|Z\|_\infty \le z] &= \mathrm{Prob}[|Z_i| \le z \text{ for all } i = 1, 2, \ldots, n]\\ &= \mathrm{Prob}[|Z_1| \le z]^n = \left(1 - \tfrac{2}{\sqrt{2\pi}}\int_z^\infty e^{-t^2/2}\,\mathrm{d}t\right)^n.\end{aligned}$$

We can estimate $\int_z^\infty e^{-t^2/2}\,\mathrm{d}t \geq \int_z^{z+1} e^{-t^2/2}\,\mathrm{d}t \geq e^{-(z+1)^2/2}$. Thus, setting $z = \sqrt{\ln n} - 1$, we have $\mathrm{Prob}[\|Z\|_\infty \leq z] \leq (1 - \frac{2}{\sqrt{2\pi}} n^{-1/2})^n$, which is below $\frac{1}{3}$ for sufficiently large n. Lemma 3.6.4 is proved. □

Exercises

1. Prove Lemma 3.6.3. $\boxed{4}$
2. (Large almost spherical sections of the crosspolytope) Use Theorem 3.4.5 and the method of the proof of Lemma 3.6.4 for proving that the n-dimensional crosspolytope (i.e. the unit ball of the ℓ_1-norm $\|\cdot\|_1$) has a 2-almost spherical section of dimension at least cn, for a suitable constant $c > 0$. $\boxed{3}$