Chapter 6

Volume of a convex body

We arrive at one of the most important applications of the Markov chain Monte Carlo method: the estimation of the volume of a convex body. For a convex body K in low dimensional Euclidean space, say two or three dimensions, it is not too difficult to estimate the volume of K within reasonable relative error using a direct Monte Carlo approach. Depending on how K is presented, it may even be possible to find the volume exactly without too much difficulty. In this chapter, therefore, we imagine the dimension n of the space to be large, and certainly greater than 3.

There are two related problems:

- sample uniformly at random a point from the convex body K;
- estimate the volume $\mathrm{vol}_n K$ of K.

We will first look at the problem of random sampling in K. Since volume is the limit of a sum, it is not surprising, in the light of examples contained in previous chapters, that the second problem can be reduced to the first. We shall look first at the problem of random sampling in K; the reduction of volume estimation to sampling will be covered at the end of the chapter.

The convex body is given as an oracle which, for a point $x \in \mathbb{R}^n$, tells whether or not $x \in K$ (see Figure 6.1). This oracle model subsumes several possible conventions for describing inputs. For example, in the case of a convex polytope defined by a set of linear inequalities it is of course easy to implement the oracle. A convex polytope presented as the convex hull of its vertices it is a little harder, but it can still be done, by linear programming. In some applications, the assumption of an *exact* oracle that accurately decides whether $x \in K$ may be unrealistic. In an implementation we would almost certainly be using arithmetic with bounded precision, and we could not always know for sure whether were in or out. In fact, it is possible to relax the definition of oracle to incorporate some fuzziness at the boundary of K without loosing much algorithmically. One of the many simplifications we shall make in this chapter is to assume exact arithmetic and an exact oracle. For a much fuller picture, refer to Kannan, Lovász and Simonovits [50].

The first thing to be noticed in this endeavour is that some intuitively appealing approaches do not work very well. Let us consider a conventional application of the Monte Carlo method to the problem. Say we shrink a box C around K as tightly as possible (see Figure 6.2), sample a point x uniformly at random from C, and return x

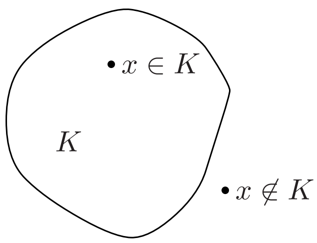

Figure 6.1: Oracle for K.

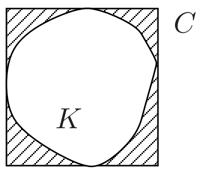

Figure 6.2: Sampling by “direct” Monte Carlo.

if $x \in K$; otherwise repeat the sampling if $x \notin K$. This simple idea works well in low dimension, but not in high dimension, where the volume ratio $\mathrm{vol}_n K/\mathrm{vol}_n C$ can be exponentially small. This phenomenon may be illustrated by a very simple example. Let $K = B_n(0,1)$ be the unit ball, and $C = [-1,1]^n$ the smallest enclosing cube. In this instance the ratio in question may be calculated exactly, and is $\mathrm{vol}_n K/\mathrm{vol}_n C = 2\pi^{n/2}/(2^n n\,\Gamma(n/2))$, which decays rapidly with n.[1] In the light of this observation, it seems that a random walk through K may provide a better alternative.

Dyer, Frieze and Kannan [28] were the first to propose a suitable random walk for sampling random points in a convex body K and prove that its mixing time scales as a polynomial in the dimension n. As a consequence, they obtained the first FPRAS for the volume of a convex body. Needless to say, this result was a major breakthrough in the field of randomised algorithms. Their approach was to divide K into a n-dimensional grid of small cubes, with transitions available between cubes sharing a facet (i.e., an $(n-1)$-dimensional face). This proposal imposes a preferred coordinate system on K leading to some technical complications. Here, instead, we use the coordinate-free “ball walk” of Lovász and Simonovits [55].

Given a point $X_t \in K$, which is the position of the random walk at time t, we choose X_{t+1} uniformly at random from $B(X_t,\delta)\cap K$, where $B(x,r)$ denotes the ball or radius r centred at x, and δ is a small appropriately chosen constant.[2] (Refer to Figure 6.3.) We will show that this Markov chain has a stationary distribution that is nearly uniform over K, and that its mixing time is polynomial in the dimension n, provided step size δ is chosen judiciously, and that K satisfies certain reasonable conditions. The stochastic process (X_t) is Markovian — the distribution of X_{t+1} depends only on X_t and not on the prior history $(X_0,\dots,X_{t-1})$ — but unlike the Markov chains so far encountered has

[1] The Gamma function extends the factorial function to non-integer values. When n is even, $\Gamma(n/2) = (n/2-1)!$, so it is easy to see that the ratio $\mathrm{vol}_n K/\mathrm{vol}_n C$ tends to 0 exponentially fast.

[2] What is described here is a “heat-bath” version of the ball wall, which has been termed the “speedy walk” in the literature. There is also a slower “Metropolis” version that we shall encounter presently.

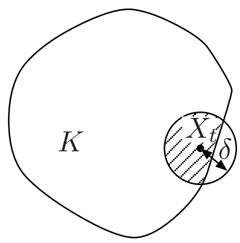

Figure 6.3: One step of the Ball Walk

infinite, even uncountable state space. We therefore pause to look briefly into the basic theory of Markov chains on $\mathbb{R}^n$.

6.1 A few remarks on Markov chains with continuous state space

Our object of study in this chapter is an MC whose state space, namely K, is a subset of $\mathbb{R}^n$. We cannot usefully speak directly of the probability of making a transition from $x \in K$ to $y \in K$, since this probability is generally 0. The solution is to speak instead of the probability $P(x, A) := \Pr[X_1 \in A \mid X_0 = x]$ of being in a (measurable) set $A \subseteq K$ at time 1 conditioned on being at x at time 0. The t step transition probabilities can then be defined inductively by $P^1 := P$ and

$$P^t(x, A) := \int_K P^{t-1}(x, dy)\, P(y, A) \tag{6.1}$$

for $t > 1$. In the case of the ball walk,

$$P(x, A) = \frac{\mathrm{vol}_n(B(x, \delta) \cap A)}{\mathrm{vol}_n(B(x, \delta) \cap K)},$$

for any (measurable) $A \subseteq K$, and

$$P(x, dy) = \frac{dy}{\mathrm{vol}_n(B(x, \delta) \cap K)}, \tag{6.2}$$

provided $y \in B(x, \delta) \cap K$.

A MC with continuous state space may have one or more invariant measures μ, which by analogy with the finite case means that μ satisfies

$$\mu(A) = \int_K P(x, A)\, \mu(dx),$$

for all measurable sets $A \subseteq K$. As in the finite case, the MC may converge to a unique invariant measure μ in the sense that $P^t(x, A) \to \mu(A)$ as $t \to \infty$ for all $x \in K$ and all measurable $A \subseteq K$.

For compactness, we shall sometimes drop explicit reference to the variable of integration in situations where no ambiguity arises, and write, e.g., $\int_K f\, d\mu$ in place of $\int_K f(x)\, \mu(dx)$.

6.2 Invariant measure of the ball walk

If we were to choose δ, the step-size of the ball walk, to be greater than the diameter $D := \sup\{\|x - y\| : x, y \in K\}$ of K, then the the ball walk would converge in one step to the uniform measure on K. (For convenience, we'll drop the subscript in the Euclidean norm $\|\cdot\|_2$.) There must be a catch! A moment's reflection reveals that the problem is one of implementability: to perform one step of the ball walk when $\delta \geq D$ we must sample a point uniformly at random from K, which is exactly the problem we set ourselves at the outset. However, provided we choose δ small enough, specifically so the ratio $\mathrm{vol}_n\big(B(X_t,\delta) \cap K\big)/\,\mathrm{vol}_n\, B(X_t,\delta)$ is not too small, we may obtain a random sample from $B(X_t,\delta) \cap K$ by repeatedly sampling from $B(X_t,\delta)$ until we obtain a point in $B(X_t,\delta) \cap K$. This is the so-called "rejection sampling" method, which is efficient provided that the probability of a successful trial is not too small.

This foregoing observation leads us to introduce a "Metropolis" version of the ball walk (which should be compared with the heat-bath version specified earlier): select a point y u.a.r. from $B(X_t,\delta)$; if $y \in K$ then set $X_{t+1} \leftarrow y$, else set $X_{t+1} \leftarrow X_t$. The Metropolis version of the ball walk has the advantage of implementability over the heat-bath version. However, it has the disadvantage that it can get stuck in sharp corners. Consider what would happen, for example, if the Metropolis walk ended up very close (in relation to the step size δ) to the corner of an n-dimensional cube. To make progress, the point y would have to move in the correct direction in each of the coordinate axes, an event that occurs with probability close to 2^{-n}. So the Metropolis walk cannot be rapidly mixing in the usual sense. We could try to loosen the definition of mixing time by somehow excluding sharp corners as possible initial states, and excluding them also from the metric employed to measure distance from stationarity. But it is cleaner to argue about the mixing time of the heat-bath version of the ball walk, and then separately argue about the relationship of the heat-bath and Metropolis walks.

The primary aim of this chapter is to convey the key ideas underlying the analysis of the ball walk, and not to obtain the most general theorems. We therefore simplify our analysis by imposing a "curvature condition" on K that rules out sharp corners. This condition radically simplifies certain technical aspects of the proof, while leaving intact all the main insights. One immediate effect of this simplification is that the Metropolis walk becomes only a constant factor slower than the heat-bath walk, so we have an easy job relating the two. Towards the end of the section, we shall review the proof and see what extra work needs to be done to eliminate the curvature condition. Provided we are prepared to accept a bound on mixing time that is wrong by a factor of n, the curvature condition may be dropped with little effort. Obtaining the correct mixing time in the absence of the curvature condition requires an analysis of substantial additional technical complexity, but requiring no further significant insights. This improvement will therefore be sketched only.

In the light of the preceeding discussion, we cannot expect the mixing time of the Metropolis version of the ball walk to be short if K is very long and thin. The small "width" of K would dictate a small δ, but then very many steps would be required to get from one end of K to the other. In the full strength version of the bound on mixing time of the ball walk, this issue is resolved by expressing the mixing time in terms of some measure of the "aspect ratio" of K. More precisely, it is supposed that K contains the unit ball $B(0,1)$ centred at the origin, and then the mixing time is expressed as a function of the diameter of K.[3] In fact, as already indicated, we simplify our presentation by making a stronger assumption, namely that the curvature of K should not be too large. We embody this simplifying assumption in the *curvature condition*:

$$\text{For all points } x \in K \text{ there is some point } y \in K \text{ such that } x \in B(y,1) \text{ and } B(y,1) \subseteq K. \tag{6.3}$$

By definition, all balls will be closed. Note that the curvature assumption is much stronger that the "official" one, which merely asserts that $B(0,1) \subseteq K$ and, in particular, rules out the interesting case of K a polytope. For the main body of this chapter, and until further notice, "ball walk" will implicitly mean the heat-bath version, and the curvature condition will be assumed.

Remark 6.1. What if we *are* presented with a body that is "thin"? It turns out that it is always possible to apply a linear transformation to K to yield a new convex body which contains a unit ball and whose diameter is quite reasonable. But this is another long story, and we do not embark on it here. Refer to Kannan, Lovász and Simonovits [50].

The stationary measure of the ball walk — we shall see presently that the ball walk is ergodic — is not uniform over K, but is close to uniform provided the step size δ is not too large. To describe the stationary measure, we introduce a function $\ell : K \to \mathbb{R}$ (called *local conductance* by Lovász and Simonovits) defined as

$$\ell(x) := \frac{\mathrm{vol}_n(B(x,\delta) \cap K)}{\mathrm{vol}_n B(x,\delta)}, \tag{6.4}$$

which may be interpreted as the probability of staying in K when choosing a random point in a δ-ball around x. Note that $\ell(x)^{-1}$ is the expected number of repetitions of this trial in order produce a point lying in $B(x,\delta) \cap K$ using rejection sampling. We want to normalise $\ell(x)$ in order to get a density which will turn out to be the density of the stationary measure of the ball walk:

$$\mu(A) := \frac{\int_A \ell(x)\,dx}{L} \quad \text{where} \quad L = \int_K \ell(x)\,dx. \tag{6.5}$$

Our first task is to verify that μ is an invariant measure for the ball walk. That it is unique follows as a weak consequence of our rapid mixing proof.

Lemma 6.2. *If X_0 has distribution μ, then X_1 does also.*

[3] Note, as a by-product, we know that K contains the origin, so we have a suitable starting point for the random walk.

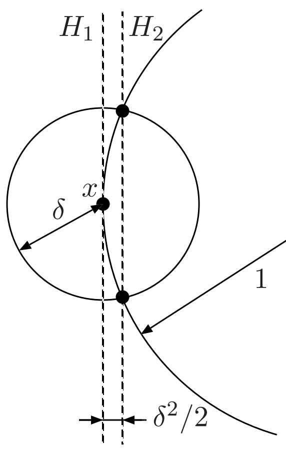

Figure 6.4: Bounding the volume of intersection

Proof. Let μ_1 denote the distribution of X_1. Then

$$\begin{aligned}
\mu_1(A) = \int_A \mu_1(dy) &= \int_A \int_K P(x, dy)\, \mu(dx) \\
&= \int_A dy \int_{B(y,\delta)\cap K} \frac{\mu(dx)}{\mathrm{vol}_n(B(x,\delta)\cap K)} && \text{by (6.2)} \\
&= \frac{1}{L}\int_A dy \int_{B(y,\delta)\cap K} \frac{\ell(x)\, dx}{\mathrm{vol}_n(B(x,\delta)\cap K)} && \text{by (6.5)} \\
&= \frac{1}{L}\int_A dy \int_{B(y,\delta)\cap K} \frac{dx}{\mathrm{vol}_n B(x,\delta)} && \text{by (6.4)} \\
&= \frac{1}{L}\int_A \ell(y)\, dy = \mu(A) && \text{by (6.4, 6.5).}
\end{aligned}$$

Hence μ is an invariant measure for the ball walk. □

Exercise 6.3. Show that the uniform distribution on K is an invariant measure for the *Metropolis* version of the ball walk.

It is clear that the distribution μ is not uniform over K, but for a suitable choice of δ it is close to it.

Lemma 6.4. *Assume the curvature condition (6.3), and suppose that $\delta \le c_1/\sqrt{n}$ (where c_1 is a dimension-independent constant). Then $0.4 \le \ell(x) \le 1$ for all $x \in K$.*

Proof. The upper bound on $\ell(x)$ is trivial from the definition of ℓ. For the lower bound we need an argument.

Recall that we assume that every $x \in K$ lies in a 1-ball $B(y,1) \subseteq K$. The inequality above will follow from

$$\frac{\mathrm{vol}_n(B(x,\delta)\cap B(y,1))}{\mathrm{vol}_n B(x,\delta)} \ge 0.4.$$

It is enough to show the relation for a point x on the boundary of $B(y,1)$. Consider the tangent plane H_1 to $B(y,1)$ through x and its parallel plane H_2 through the intersection of the boundaries of the two balls. (Refer to Figure 6.4.) Orient them such that their positive side H_i^+ $(i = 1, 2)$ contains the point y. Notice that

$$B(x, \delta) \cap H_2^+ \subset B(y, 1)$$

(δ is assumed to be smaller than 1). Therefore it is enough to show that the set $B(x, \delta) \cap H_2^+$ has volume at least $0.4 \operatorname{vol}_n B(x, \delta)$. We will do this by showing that $B(x, \delta) \cap H_2^- \cap H_1^+$ has very small volume, i.e., at most a 0.1 fraction of the volume of $B(x, \delta)$. The set in question is contained in the cylinder with ground face $B(x, \delta) \cap H_1$ (which is an $(n-1)$-dimensional ball with radius δ) whose height is the distance apart of H_1 and H_2. A simple computation reveals that this distance is exactly $\delta^2/2$. From the volume formula of balls of dimensions $n - 1$ and n, and Stirling's approximation for the Γ-function, we obtain the following relation

$$\frac{\operatorname{vol}_{n-1}(B(x, \delta) \cap H_1)}{\operatorname{vol}_n B(x, \delta)} \leq \frac{c\sqrt{n}}{\delta},$$

for some universal constant c. Hence the volume of the cylinder is at most a $\frac{1}{2}c\delta\sqrt{n}$ fraction of the volume of $B(x, \delta)$. Setting $c_1 = 1/5c$ gives the desired bound. □

What this lemma also says is that we can implement one transition of the ball walk efficiently: going from a point $x \in K$ to a random point in $B(x, \delta)$ we have a probability of at least 0.4 of ending up in K immediately; in other words, the Metropolis version of the ball walk is only a factor 2.5 slower than the heat-bath version.

6.3 Mixing rate of the ball walk

We will show now that the ball walk mixes rapidly. The next lemma is a powerful weapon and forms the basis of one of our standard techniques.

Lemma 6.5. *Let f be a measurable function over a measurable set S. Partition S into measurable sets $S_0, \dots, S_{m-1}$. Then*

$$\int_S f^2 \, d\mu = \sum_{i=0}^{m-1} \int_{S_i} (f - \bar{f}_i)^2 \, d\mu + \sum_{i=0}^{m-1} \mu(S_i) \bar{f}_i^2, \tag{6.6}$$

where

$$\bar{f}_i := \frac{1}{\mu(S_i)} \int_{S_i} f \, d\mu.$$

Remark 6.6. Suppose that $\mathbb{E}_\mu f := \int_K f \, d\mu = 0$. Then on the l.h.s. of the equality we have simply $\operatorname{Var}_\mu f$. The two terms on the r.h.s. of the equality may be interpreted as (i) the sum of the variances of f *within* each of the regions S_i, and (ii) the variance of f *between* the regions, respectively.

Proof of Lemma 6.5.

$$\begin{aligned}\int_{S_i} (f - \bar{f}_i)^2 \, d\mu + \mu(S_i)\bar{f}_i^2 &= \int_{S_i} f^2 \, d\mu + \int_{S_i} \bar{f}_i^2 \, d\mu - 2\int_{S_i} \bar{f}_i f \, d\mu + \mu(S_i)\bar{f}_i^2 \\ &= \int_{S_i} f^2 \, d\mu + \mu(S_i)\bar{f}_i^2 - 2\mu(S_i)\bar{f}_i^2 + \mu(S_i)\bar{f}_i^2 \\ &= \int_{S_i} f^2 \, d\mu.\end{aligned}$$

□

As in the analysis of the matchings MC, our approach to bounding the mixing time involves taking a (measurable) test function $f : K \to \mathbb{R}$ (with $\mathbb{E} f = 0$ for convenience) and examining how the variance of f decays as a result of the averaging effect of the ball-wall. To this end, introduce a function $h : K \to \mathbb{R}$ given by

$$\begin{aligned}h(x) &:= \frac{1}{2}\int_K P(x, dy)\,(f(x) - f(y))^2 \\ &= \frac{1}{2\operatorname{vol}_n(B(x,\delta)\cap K)} \int_{B(x,\delta)\cap K} (f(x) - f(y))^2 \, dy,\end{aligned} \tag{6.7}$$

and define

$$\operatorname{Var}_\mu f := \int_K f^2 \, d\mu \quad \text{and} \quad \mathcal{E}_P(f,f) := \int_K h \, d\mu;$$

these are the now-familiar variance (global variation of f over K) and Dirichlet form (local variation of f at the scale of the step size δ of the ball walk). As with the matching MC, the key to the analysis of the ball walk lies in obtaining a sharp Poincaré inequality linking $\operatorname{Var}_\mu f$ and $\mathcal{E}_P(f,f)$. Our eventual goal is to show:

Theorem 6.7 (Poincaré inequality)**.** *Let $K \subset \mathbb{R}^n$ be a convex body of diameter D satisfying the curvature condition (6.3), and suppose that δ is as in Lemma 6.4. For any (measurable) function $f : K \to \mathbb{R}$,*

$$\mathcal{E}_P(f,f) \geq \lambda \operatorname{Var}_\mu f \tag{6.8}$$

where

$$\lambda := \frac{c_2\delta^2}{D^2 n}$$

for some universal constant c_2.

We apply the technique by Mihail (as we did with matchings in §5.2) and obtain from λ a bound on mixing time. As before, we deal with periodicity by considering either a continuised or lazy walk.

Corollary 6.8. *For any $\varepsilon > 0$ let $\tau(\varepsilon)$ denote the time at which the ball walk (in either its continuised or lazy variants) reaches within total variation distance ε of the stationary distribution μ. Then, under the curvature condition (6.3),*

$$\tau(\varepsilon) \leq O\left(\lambda^{-1}\left(\log \varepsilon^{-1} + i(\mu_0)\right)\right),$$

where λ is as in Theorem 6.7 and $i(\mu_0)$ expresses the dependence on the initial distribution μ_0.

Remark 6.9. The expression $i(\mu_0)$ is closely related to the term $\ln \pi(x_0)^{-1}$ familiar from the discrete case. But if we now start from a fixed point (in other words our initial distribution μ_0 is a single point mass at $x_0 \in K$) no meaning can be attached to $\ln \pi(x_0)^{-1}$. To escape from this, imagine that we start at time -1 from a point x_0 such that $B(x_0, \delta) \subseteq K$, and consider the situation at time 0. Thus the initial distribution μ_0 is uniform over some ball of radius δ. In this case, we may take $i(\mu_0) = n \ln(D/2\delta)$.

Exercise 6.10. Verify Corollary 6.8. Doing this essentially involves translating Theorem 5.6 to the setting of continuous state space. In case you skip this exercise, a full derivation may be found in §6.8.

At an intuitive level, Theorem 6.7 seems to be close to the truth. With a step size of δ, the distance travelled parallel to any axis fixed in advance (in particular, one parallel to a diameter of K) is of order $\delta/\sqrt{n}$. The time taken for the walk to "diffuse" along a diameter is the square of the ratio of D to the typical distance moved along the diameter in one step, namely $(D\sqrt{n}/\delta)^2$, which is of order λ^{-1}. To minimise mixing time we clearly wish to take δ as small as possible consistent with implementability, which by Lemma 6.4 is of order $n^{-1/2}$. With that step size, the Poincaré constant scales as $(nD)^{-2}$.

The next section is devoted to the proof of what is essentially the main result of this chapter.

6.4 Proof of the Poincaré inequality (Theorem 6.7)

Assume the converse to (6.8), namely that there exists a function $f : K \to \mathbb{R}$ with

$$\mathcal{E}_P(f, f) < \lambda \operatorname{Var}_\mu f; \tag{6.9}$$

informally, f sustains high global variation simultaneously with low local variation.

We will define smaller and smaller *violating sets* S such that the ratio

$$\int_S h \, d\mu \Big/ \int_S (f - \bar{f})^2 \, d\mu \tag{6.10}$$

is small, where $\bar{f} = \int_S f \, d\mu$. Our starting point is of course $S = K$, where we know that this ratio is less than λ. Eventually, S will be small even with respect to δ. Then the function f will have to be almost constant in S since the local variation (as measured by the numerator) is small; however the global variation (as measured by the denominator) is large. Here we reach a contradiction. This in outline is our proof.

First we will shrink the violating set to a set K_1 which is very small in all but one dimension, a so-called "needle-like" body. It transpires that we can do this while keeping ratio (6.10) bounded throughout by λ. It is only when we attempt to shrink along the final dimension that we have to give something away. Before embarking on the process of shrinking K to a needle-like body, we need a pair of geometrical lemmas, whose proofs we defer to §6.5.

Lemma 6.11. *Let R be a convex set in $\mathbb{R}^2$. There is a point $x \in R$ such that every line through x partitions R into pieces of area at least $\frac{1}{3}$ of the area of R.*

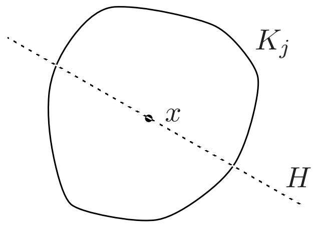

Figure 6.5: The expectation of f is zero on both $K_j \cap H^+$ and $K_j \cap H^-$.

Remark 6.12. The bound $\frac{1}{3}$ can in fact be replaced by $\frac{4}{9}$, which is tight as can been seen by considering an equilateral triangle; see Egglestone [31, §6.4]. However, any strictly positive bound is adequate for our purposes.

The *width* of a convex set R in $\mathbb{R}^2$ is the minimum, over all pairs of parallel supporting lines sandwiching R, of the distance between those lines.[4]

Lemma 6.13. *Let R be a convex set in $\mathbb{R}^2$ of area A. Then the width of R is at most $\sqrt{2A}$.*

Remark 6.14. Again, the bound is not the best possible, but is adequate for our purposes. The extremal set (i.e., the one of given area that maximises width) is again an equilateral triangle.

To resume: With the aim of establishing a contradiction we are assuming the existence of a function $f : K \to \mathbb{R}$ satisfying (6.9). We may further assume (by adding an appropriate constant function to f) that $\mathbb{E}_\mu f = 0$. This additional assumption will be convenient on the first leg of our journey towards the contradiction.

Claim 6.15. *Assume $f : K \to \mathbb{R}$ satisfies inequality (6.9), and $\mathbb{E}_\mu f = 0$. Then, for every $\varepsilon > 0$, there is a convex subset $K_1 \subseteq K$ satisfying*

$$\int_{K_1} h \, d\mu < \lambda \int_{K_1} f^2 \, d\mu \quad \textit{as well as} \quad \int_{K_1} f \, d\mu = 0,$$

and such that K_1 lies in the box $[0, D] \times [0, \varepsilon]^{n-1}$ in some Cartesian coordinate system.

Proof. Suppose, for some $j \geq 2$, that K_j is a violating set which lies in $[0, D]^j \times [0, \varepsilon]^{n-j}$, and that $\int_{K_j} f \, d\mu = 0$; i.e., we have already shrunk our violating set down on $n - j$ coordinates. (The base case $K_n = K$ is of course covered by (6.9).) To shrink along a further coordinate we use a beautiful divide-and-conquer argument due to Payne and Weinberger: see Bandle [4, Thm 3.24].

Let R be the projection of K_j onto the first two (i.e., "fat") axes. Let x be a point satisfying the conditions of Lemma 6.11. Consider all $(n-1)$-dimensional planes through x whose normals lie in the 2-dimensional plane spanned by the first two axes.

[4] In some sense, width it is the opposite of diameter, which may be defined as the maximum such distance. This was not how we defined diameter in §6.2, but the two definitions are equivalent.

These planes project to lines through x in the plane of R. Among these planes there is at least one, say H, such that

$$\int_{K_j \cap H^+} f\, d\mu = \int_{K_j \cap H^-} f\, d\mu = 0.$$

To see this, choose any $(n-1)$-dimensional plane G through x whose normal lies within the plane of R. If G does not already have the desired property, then, since $\int_{K_j \cap G^+} f\, d\mu + \int_{K_j \cap G^-} f\, d\mu = 0$, one integral or the other has to be positive. By rotating G about x by an angle of π, the signs exchange. So by continuity and the mean value theorem we have to have hit the sought-for H at some point.

It is easy to convince oneself that K_j intersected with one side of H (i.e., either $K_j \cap H^+$ or $K_j \cap H^-$) is also a violating set, in the sense that the ratio (6.10) is bounded by λ when $S = K_j \cap H^+$ (or $S = K_j \cap H^-$, as appropriate). Now iterate this procedure. By Lemma 6.11, the area of the projection R of the convex body drops by a constant factor at each iteration, and must eventually drop below $\frac{1}{2}\varepsilon^2$. At this point the width of R, by Lemma 6.13, is at most ε. Then, rotating the fat axes as appropriate, the projection of the convex body onto (say) the first of these axes is a line segment of length at most ε. The convex set now has exactly the properties we require of the set K_{j-1}, i.e., the same properties as K_j, but with $j-1$ replacing j. Hence by induction we can find our set K_1. □

The above line of argument requires at least two fat dimensions in order to provide enough freedom in selecting the plane H. We need a new approach in order to shrink the needle-line set along the remaining fat dimension.

Claim 6.16. *Let K_1 and f be as in the conclusion of Claim 6.15, δ be as in Lemma 6.4, and let $\eta := c_3\delta/\sqrt{n}$ where $c_3 > 0$ is any constant. Then, under the curvature condition (6.3), there is a convex subset $K_0 \subseteq K_1$ satisfying*

$$\int_{K_0} h\, d\mu < \frac{1}{10} \int_{K_0} (f - \bar{f}\,)^2\, d\mu \tag{6.11}$$

where

$$\bar{f} = \frac{1}{\mu(K_0)} \int_{K_0} f\, d\mu, \tag{6.12}$$

and such that K_0 lies in the box $[-\eta, \eta] \times [0, \varepsilon]^{n-1}$ in some Cartesian coordinate system.

Remark 6.17. We will choose the constant c_3 later; in order to obtain an eventual contradiction, it will need to be small enough. The choice of c_3 will then determine the universal constant c_2 of Theorem 6.7: the smaller c_3, the smaller c_2.

Our strategy for proving Claim 6.16 is to chop K_1 into short sections and show that at least one of these sections (or perhaps the union of two adjacent ones) satisfies the inequality (6.11). (Refer to Figure 6.6.) Before embarking on the proof proper, we need another geometric lemma, which is a consequence of the Brunn-Minkowski Theorem; the proof is again deferred to §6.5.

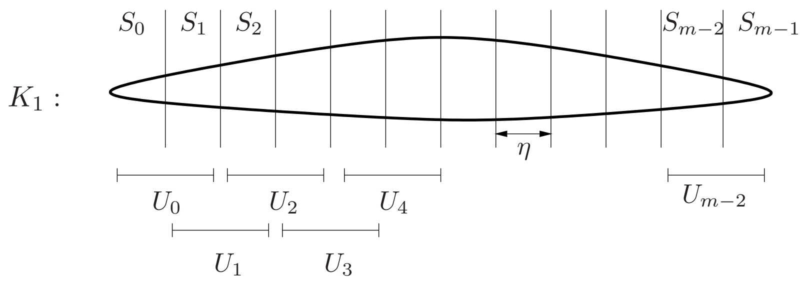

Figure 6.6: Partitioning of K_1

Lemma 6.18. *Let convex body K_1 be partitioned into m pieces $S_0 \dots S_{m-1}$ of equal width by planes orthogonal to a fixed axis. Then the sequence*

$$\frac{1}{\operatorname{vol}_n S_0}, \frac{1}{\operatorname{vol}_n S_1}, \dots, \frac{1}{\operatorname{vol}_n S_{m-1}}$$

is convex.

We are ready to resume the chopping argument.

Proof of Claim 6.16. Let convex body K_1 be partitioned into m pieces by planes orthogonal to the fat axis, as specified in Lemma 6.18, so that each piece S_i has width $\eta = c_3\delta/\sqrt{n}$. Additionally, define $U_i := S_i \cup S_{i+1}$ for $i = 0, 1, \dots, m-2$. Note that $m = O(D\sqrt{n}/\delta)$. Using Lemma 6.5, we find

$$\int_{K_1} f^2 \, d\mu = \underbrace{\sum_{i=0}^{m-1} \int_{S_i} (f - \bar{f}_i)^2 \, d\mu}_{A} + \underbrace{\sum_{i=0}^{m-1} \mu(S_i)\bar{f}_i^2}_{B}, \tag{6.13}$$

where for convenience we define

$$\bar{f}_i := \frac{1}{\mu(S_i)} \int_{S_i} f \, d\mu.$$

In the case that sum A is greater or equal to sum B, we readily find a piece S_i that serves as a violating set. We start with

$$\begin{aligned}
\sum_{i=0}^{m-1} \int_{S_i} h \, d\mu &= \int_{K_1} h \, d\mu && (6.14)\\
&< \lambda \int_{K_1} f^2 \, d\mu && \text{by assumption}\\
&\le 2\lambda \sum_{i=0}^{m-1} \int_{S_i} (f - \bar{f}_i)^2 \, d\mu && \text{by (6.13) and } A \ge B. \quad (6.15)
\end{aligned}$$

Comparing sums (6.14) and (6.15) we see there must be an S_i such that

$$\int_{S_i} h \, d\mu \le 2\lambda \int_{S_i} (f - \bar{f}_i)^2 \, d\mu.$$

Setting $K_0 = S_i$ satisfies the conclusion of the claim with plenty to spare. (Note in this context that $\lambda = O(n^{-2})$.)

The case $B > A$ is a little more difficult. Using the alternative expression for variance which we have seen before, and recalling that the expectation of f with respect to μ on K_1 is 0, we have

$$\begin{aligned}
\mu(K_1) \int_{K_1} f^2 \, d\mu &< 2\,\mu(K_1) \sum_{i=0}^{m-1} \mu(S_i) \bar{f}_i^2 && \text{since } B > A \\
&= 2 \sum_{0 \le i < j < m} \mu(S_i)\mu(S_j)(\bar{f}_i - \bar{f}_j)^2 && \text{using (5.5).}
\end{aligned} \tag{6.16}$$

Our aim is to replace the r.h.s. of (6.16) by a sum with similar terms, but restricted to *adjacent* pairs i, j. This will enable us to finish with an argument similar to the $A \ge B$ case.

For convenience, we introduce the abbreviation $w_i = \mu(S_i)$, and set

$$a_{i,j} := w_i w_j \sum_{k=i}^{j-1} \frac{w_k + w_{k+1}}{w_k w_{k+1}} \le 2 w_i w_j \sum_{k=i}^{j} \frac{1}{w_k}. \tag{6.17}$$

Inequality (6.16) may be massaged as follows:

$$\begin{aligned}
\mu(K_1) \int_{K_1} f^2 \, d\mu &< 2 \sum_{i<j} w_i w_j (\bar{f}_i - \bar{f}_j)^2 \\
&= 2 \sum_{i<j} w_i w_j \left[\sum_{k=i}^{j-1} (\bar{f}_k - \bar{f}_{k+1}) \right]^2 \\
&= 2 \sum_{i<j} w_i w_j \left[\sum_{k=i}^{j-1} \sqrt{\frac{w_k + w_{k+1}}{w_k w_{k+1}}} \cdot \sqrt{\frac{w_k w_{k+1}}{w_k + w_{k+1}}} \, (\bar{f}_k - \bar{f}_{k+1}) \right]^2 \\
&\le 2 \sum_{i<j} a_{i,j} \sum_{k=i}^{j-1} \frac{w_k w_{k+1}}{w_k + w_{k+1}} (\bar{f}_k - \bar{f}_{k+1})^2,
\end{aligned} \tag{6.18}$$

where the final inequality is Cauchy-Schwarz combined with (6.17). Define $\hat{f}_k$ to be the expectation of f over $U_k = S_k \cup S_{k+1}$:

$$\hat{f}_k := \frac{1}{\mu(U_k)} \int_{U_k} f \, d\mu = \frac{w_k \bar{f}_k + w_{k+1} \bar{f}_{k+1}}{w_k + w_{k+1}}.$$

Then, by Lemma 6.5,

$$\begin{aligned}
\frac{w_k w_{k+1}}{w_k + w_{k+1}} (\bar{f}_k - \bar{f}_{k+1})^2 &= w_k (\bar{f}_k - \hat{f}_k)^2 + w_{k+1} (\bar{f}_{k+1} - \hat{f}_k)^2 \\
&\le \int_{U_k} (f - \hat{f}_k)^2 \, d\mu
\end{aligned} \tag{6.19}$$

(The first line may be viewed as the special case $|\Omega| = 2$ of (5.5), or may be verified by elementary algebraic manipulation. Inequality (6.19) comes from Lemma 6.5, noting that the first sum on the r.h.s. of (6.6) is clearly positive.) Applying bound (6.19) to the terms in (6.18) yields

$$\mu(K_1)\int_{K_1} f^2\,d\mu < 2\sum_{i<j} a_{i,j}\sum_{k=i}^{j-1}\int_{U_k}(f-\hat{f}_k)^2\,d\mu. \tag{6.20}$$

Taking stock momentarily: inequality (6.20) appears to be telling us that if the variance of f is large on K_1 then it must be large on *some* U_k; but there is still some work to be done on the way to quantifying this effect.

Recall that

$$w_i = \mu(S_i) = L^{-1}\int_{S_i}\ell(x)\,dx,$$

where $L = \int_K \ell(x)\,dx$. Thus, by Lemma 6.4,

$$0.4\,L^{-1}\operatorname{vol}_n S_i \le w_i \le L^{-1}\operatorname{vol}_n S_i, \tag{6.21}$$

leading to the following upper bound on $a_{i,j}$:

$$\begin{aligned}
a_{i,j} &\le 2\,w_iw_j\sum_{k=i}^{j}\frac{L}{0.4\operatorname{vol}_n S_k} && \text{by (6.17) and (6.21)}\\
&\le 2.5\,w_iw_jL\,(j-i+1)\Big(\frac{1}{\operatorname{vol}_n S_i}+\frac{1}{\operatorname{vol}_n S_j}\Big) && \text{by Lemma 6.18}\\
&\le 2.5(j-i+1)(w_i+w_j) && \text{by (6.21).} \qquad (6.22)
\end{aligned}$$

Since $j-i+1$ never exceeds m, we have the following crude bound on the sum of the $a_{i,j}$:

$$\begin{aligned}
\sum_{i<j} a_{i,j} &\le 2.5\sum_{i<j}(j-i+1)(w_i+w_j)\\
&\le 2.5\,m\sum_{i<j}(w_i+w_j)\\
&\le 2.5\,m^2\sum_i w_i \qquad (6.23)\\
&= 2.5\,m^2\mu(K_1). \qquad (6.24)
\end{aligned}$$

To see (6.23), fix attention on a particular index k and note that w_k occurs exactly $m-1$ times in the double sum.

Returning now to (6.20),

$$\begin{aligned}
\mu(K_1)\int_{K_1} f^2\,d\mu &< 2\sum_{i<j} a_{i,j}\sum_{k=i}^{j-1}\int_{U_k}(f-\hat{f}_k)^2\,d\mu\\
&\le 2\sum_{i<j} a_{i,j}\sum_{k=0}^{m-2}\int_{U_k}(f-\hat{f}_k)^2\,d\mu\\
&\le 5m^2\mu(K_1)\sum_{k=0}^{m-2}\int_{U_k}(f-\hat{f}_k)^2\,d\mu && \text{by (6.24),}
\end{aligned}$$

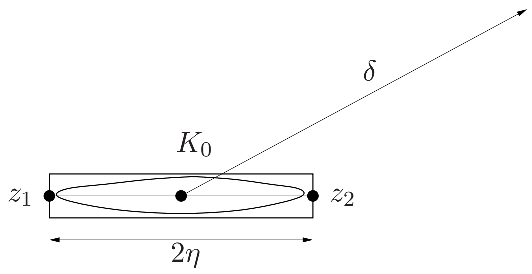

Figure 6.7: "Needle like" body K_0

from which

$$\int_{K_1} f^2\, d\mu \le 5m^2 \sum_{k=0}^{m-2} \int_{U_k} (f - \hat{f}_k)^2\, d\mu. \tag{6.25}$$

Inequality (6.25) is the one we sought, expressing the fact that if the variance of f is large on the whole of K_1 then it must be fairly large on some piece U_k. Proceeding now by analogy with the $A \le B$ case, using (6.25) and the conclusion of Claim 6.15,

$$\sum_{k=0}^{m-2} \int_{U_k} h\, d\mu \le 2 \int_{K_1} h\, d\mu < 2\lambda \int_{K_1} f^2\, d\mu \le 10m^2\lambda \sum_{k=0}^{m-2} \int_{U_k} (f - \hat{f}_k)^2\, d\mu.$$

Therefore there must exist a k such that

$$\int_{U_k} h\, d\mu < 10m^2\lambda \int_{U_k} (f - \hat{f}_k)^2\, d\mu. \tag{6.26}$$

By setting c_2 sufficiently small, specifically $c_2 < c_3^2/100$, we obtain

$$10m^2\lambda = 10\left(\frac{D\sqrt{n}}{c_3\delta}\right)^2 \frac{c_2\delta^2}{D^2 n} < \frac{1}{10}.$$

Setting $K_0 := U_k$, we recognise (6.26) as the inequality promised in the statement of the claim. This concludes the case $B > A$ and hence the proof. □

We pick up the proof of Theorem 6.7. At the outset we assumed, with a view to obtaining a contradiction, the converse of (6.8). Now, from Claims 6.15 and 6.16, we deduce the existence of a convex set $K_0 \subset K$ satisfying inequality (6.11) such that K_0 is contained in a prism of height 2η whose cross section is an $(n-1)$-dimensional cube of side ε. We are close to obtaining the desired contradiction.

Let C be the centre axis of the prism, and let z_1 and z_2 be the points at which C intersects the end facets of the prism. (Refer to Figure 6.7.) Let $\delta' := \delta - \varepsilon\sqrt{n}$, and choose ε sufficiently small that

$$\operatorname{vol}_n B(0, \delta') \ge 0.9 \operatorname{vol}_n B(0, \delta). \tag{6.27}$$

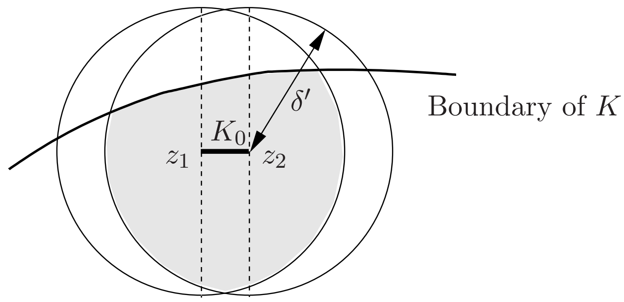

Figure 6.8: Construction of the set I (shown shaded)

(Recall that we are free to choose ε as small as we like.) Set $I := B(z_1, \delta') \cap B(z_2, \delta') \cap K$. (Refer to Figure 6.8.) We shall argue that by choosing c_3 (and hence η) sufficiently small we can ensure

$$\mathrm{vol}_n \big(B(z_1, \delta') \cap B(z_2, \delta') \big) \geq 0.8 \, \mathrm{vol}_n \, B(0, \delta), \tag{6.28}$$

and hence

$$\mathrm{vol}_n \, I = \mathrm{vol}_n \big(B(z_1, \delta') \cap B(z_2, \delta') \cap K \big) \geq 0.2 \, \mathrm{vol}_n \, B(0, \delta). \tag{6.29}$$

The calculation supporting (6.28) proceeds exactly as in the proof of Lemma 6.4. Divide $B(z_1, \delta') \cap B(z_2, \delta')$ into two congruent pieces by the plane bisecting the line (z_1, z_2) and orthogonal to it. Each piece can be viewed as a half-ball less a segment that can be contained in a cylinder of height η $(= c_3\delta/\sqrt{n}\,)$ and radius $\delta' \leq \delta$. By setting c_3 small enough — refer to the calculation in the proof of Lemma 6.4 — we may ensure that the volume of this cylinder is less than $0.05 \, \mathrm{vol}_n \, B(0, \delta)$. Now, by (6.27), the combined volume of the two half balls is at least $0.9 \, \mathrm{vol}_n \, B(0, \delta)$, so after removing the two segments we are still left with a set of volume $0.8 \, \mathrm{vol}_n \, B(0, \delta)$, as claimed in (6.28). Inequality (6.29) is now immediate: just observe that the piece of $B(z_1, \delta') \cap B(z_2, \delta')$ that we loose when we intersect with K is contained in $B(z_1, \delta) \setminus K$, which by Lemma 6.4 has volume at most $0.6 \, \mathrm{vol}_n \, B(0, \delta)$.

Inequality (6.29) expresses one key property of I, namely that its volume is not too small. The other key property is that every point in I may be reached from any point in K_0 in one step of the ball walk. For by construction,

$$\sup \big\{ \|x - y\| : x \in C \text{ and } y \in I \big\} \leq \delta',$$

from which, by the triangle inequality,

$$\sup \big\{ \|x - y\| : x \in K_0 \text{ and } y \in I \big\} \leq \delta' + \varepsilon\sqrt{n} = \delta.$$

Since $I \subseteq K$, we may conveniently reformulate this fact as

$$I \subseteq B(x, \delta) \cap K, \quad \text{for all } x \in K_0. \tag{6.30}$$

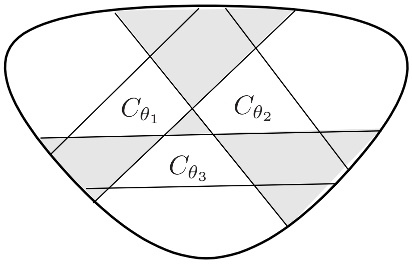

Figure 6.9: A paradoxical subset of R.

So,

$$\begin{aligned}
\int_{K_0} h\,d\mu &\geq \frac{1}{2}\int_{K_0} \frac{\mu(dx)}{\mathrm{vol}_n(B(x,\delta)\cap K)}\int_I \big(f(x)-f(y)\big)^2\,dy && \text{by (6.7, 6.30)}\\
&\geq \frac{1}{2\,\mathrm{vol}_n B(0,\delta)}\int_{K_0}\mu(dx)\int_I \big(f(x)-f(y)\big)^2\,dy\\
&\geq \frac{1}{2\,\mathrm{vol}_n B(0,\delta)}\int_I dy\int_{K_0} \big(f(x)-f(y)\big)^2\,\mu(dx) && \text{(Fubini)}\\
&\geq \frac{1}{2\,\mathrm{vol}_n B(0,\delta)}\int_I dy\int_{K_0} (f-\bar{f}\,)^2\,d\mu && (6.31)\\
&\geq \frac{1}{10}\int_{K_0} (f-\bar{f}\,)^2\,d\mu && \text{by (6.29),}
\end{aligned}$$

where $\bar{f}$, as in (6.12), is the μ-expectation of f over K_0. Inequality (6.31) uses a simple fact about variance, namely that $\int_{K_0}(f-c)^2\,d\mu$ is minimised by setting $c=\bar{f}$. But the combined inequality contradicts (6.11). This completes the proof of Theorem 6.7.

6.5 Proofs of the geometric lemmas

In this section we tie up the loose ends by providing proofs for the three geometric lemmas used in the proof of Theorem 6.7.

Proof of Lemma 6.11. The following proof is due to Alan Riddell; I thank him and also Toby Bailey for communicating it to me.

Consider all possible partitions of R into three regions of equal area by a pair of parallel lines. (There is one partition corresponding to each orientation for the lines.) Let $\{C_\theta : 0 \leq \theta < \pi\}$ be an indexing of the central bands in these partitions, considered as closed sets. Suppose there exist bands C_{θ_1}, C_{θ_2} and C_{θ_3} with no point in common. The set $\mathbb{R}^2\setminus(C_{\theta_1}\cup C_{\theta_2}\cup C_{\theta_3})$ consists of six unbounded regions and one triangle. Consider the partition of R into seven pieces obtained by extending the edges of the triangle to the boundary of R, and in particular the four pieces shown shaded in Figure 6.9. Each of the shaded pieces other than the central triangle has area at least $\frac{1}{3}\,\mathrm{vol}_2\,R$, since it is the intersection of two regions of R of area $\frac{2}{3}\,\mathrm{vol}_2\,R$. The central triangle itself has positive area. Thus the total shaded area exceeds $\mathrm{vol}_2\,R$, a contradiction.

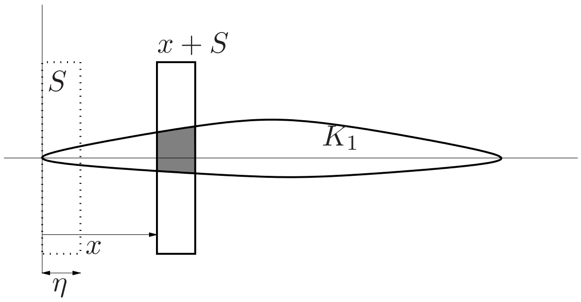

Figure 6.10: Slab S sweeping over K_1

Hence every triple from $\{C_\theta\}$ has a common point and, by Helly's theorem (see Egglestone [31, Thm 17]), the intersection $\bigcap_\theta C_\theta$ of all central bands is non-empty. Any point in this intersection will do as our choice for x. □

Proof of Lemma 6.13. Suppose R is a convex region in $\mathbb{R}^2$ of area A. Let ℓ_1 and ℓ_1' be parallel supporting lines of R, touching R at the points α and α'. We may arrange for lines ℓ_1 and ℓ_1' to be perpendicular to the line segment $[\alpha, \alpha']$, e.g., by choosing $[\alpha, \alpha']$ to be a diameter of R. Now let ℓ_2 and ℓ_2' be supporting lines of R perpendicular to ℓ_1 and ℓ_1', touching R at the points β and β'. The rectangle formed by these supporting lines has area at least w^2, where w is the width of R. It is easy to see that the convex hull of $\{\alpha, \alpha', \beta, \beta'\}$ has area $\frac{1}{2}w^2$. (The fact that $[\alpha, \alpha']$ is parallel with an edge of the rectangle is crucial here.) But the convex hull of $\{\alpha, \alpha', \beta, \beta'\}$ is contained within R. It follows that $A \geq \frac{1}{2}w^2$. □

Proof of Lemma 6.18. For what follows, we abbreviate $\mathrm{vol}_n S_i$ by v_i. In order to prove the lemma, the notation of Minkowski sums is useful: Let A and B be sets of points and λ a real number. A point p is represented by the vector pointing from 0 to p. Then we define the set $A+B$ as the set of points $a+b$ with $a \in A$ and $b \in B$. Furthermore, for a scalar λ, λA is the set of points λa with $a \in A$.

We prove the lemma by showing properties of the function $\mathrm{vol}_n\big((xe_1 + S) \cap K_1\big)$ for $x \in [0, D]$, where S is a "slab" of width η, and e_1 is a unit vector parallel to the fat axis. (The slab is defined as the intersection of two halfspaces orthogonal to the fat axis and distant η apart; assume that the origin is placed at the leftmost point of K_1.) Thus we move the slab S from left to right and observe how the volume of the intersection $K_1 \cap S$ behaves. Note that $v_i := \mathrm{vol}_n S_i = \mathrm{vol}_n\big((i\eta e_1 + S) \cap K_1\big)$. (Refer to Figure 6.10.)

The proof of the lemma relies on a theorem of Brunn and Minkowski (see Egglestone [31, Thm 46]).

Theorem 6.19 (Brunn-Minkowski)**.** *Let K' and K'' be two convex bodies in $\mathbb{R}^n$. Then*

$$\mathrm{vol}_n(K' + K'')^{1/n} \geq \mathrm{vol}_n(K')^{1/n} + \mathrm{vol}_n(K'')^{1/n}.$$

To continue with the proof of Lemma 6.18, observe that

$$(\lambda x + (1-\lambda)y + S) \cap K_1 \supseteq \lambda((x+S) \cap K_1) + (1-\lambda)((y+S) \cap K_1). \tag{6.32}$$

To verify this, assume z is in the set on the right hand side. This means that we can write $z = z' + z''$ with $z' \in \lambda((x+S)\cap K_1)$ and $z'' \in (1-\lambda)((y+S)\cap K_1)$. Therefore, $z' \in \lambda K_1$ and $z'' \in (1-\lambda)K_1$. Thus $z \in K_1$. On the other hand, we have $z' \in \lambda(x+S)$ and $z'' \in (1-\lambda)(y+S)$ which leads to $z \in \lambda x + (1-\lambda)y + S$.

Using the Brunn-Minkowski Theorem in conjunction with (6.32), we find

$$\begin{aligned}
&\operatorname{vol}_n \left[(\lambda x + (1-\lambda)y + S)\cap K_1\right]^{1/n}\\
&\qquad \geq \operatorname{vol}_n \left[\lambda((x+S)\cap K_1) + (1-\lambda)((y+K_1)\cap K_1)\right]^{1/n}\\
&\qquad \geq \operatorname{vol}_n[\lambda((x+S)\cap K_1)]^{1/n} + \operatorname{vol}_n[(1-\lambda)((y+S)\cap K_1)]^{1/n}\\
&\qquad = \lambda \operatorname{vol}_n[(x+S)\cap K_1]^{1/n} + (1-\lambda)\operatorname{vol}_n[(y+S)\cap K_1]^{1/n}.
\end{aligned}$$

In the last step, we used $\operatorname{vol}_n(\lambda K) = \lambda^n \operatorname{vol}_n K$. As a special case of this inequality, we find that the sequence $(v_i^{1/n})$ is concave:

$$2v_i^{1/n} \geq v_{i-1}^{1/n} + v_{i+1}^{1/n} \tag{6.33}$$

Now it is easily checked that if (a_i) is any concave sequence, and g any monotone non-increasing convex function, then the sequence $(g(a_i))$ is convex. The lemma then follows from (6.33) by setting $a_i = v_i^{1/n}$ and $g(x) = x^{-n}$. □

6.6 Relaxing the curvature condition

What happens if we do not have the curvature condition (6.3)? As we shall see, the question is of some importance, not least because the standard reduction from volume estimation to sampling introduces sharp corners, even if these are absent in the given convex body K. The most obvious consequence of dropping (6.3) is that the expected number of Metropolis steps to simulate a single heat-bath step is no longer bounded by a constant. Worse, as we have argued, the expected number steps may be exponential in n for a worst-case choice for the current point $X_t = x$. The most we can hope for is that, in a typical evolution of the ball walk, we are very unlikely to visit this bad region of K. This turns out indeed to be the case, provided $\delta = O(1/\sqrt{n}\,)$, the body K contains the unit ball $B(0,1)$, and we make a reasonable choice of initial state. See Kannan, Lovász and Simonovits [50].

Remark 6.20. To get a feel for what is going on, imagine the Metropolis ball walk in some n-dimensional polytope K. In order to mix, the walk needs potentially to "see all the boundary" of K, otherwise it cannot gain information about the body. In the case of a polytope this means that we would have to treat the case of coming close to *facets* (i.e., $(n-1)$-dimensional faces) of the polytope. There the random walk can "learn" a lot about the shape of K. But it does not necessarily have to come close to smaller-dimensional faces, where the walk might get stuck for long periods. Not surprisingly, the main technical difficulties then arise from showing that close encounters with low-dimensional faces are rare.

A problem arises, however, before we ever reach the comparison of the heat-bath and Metropolis versions of the ball walks. Specifically, our derivation of the key Poincaré inequality contained in Theorem 6.7 made use of the curvature condition at two points: at inequalities (6.21) and (6.29), both of which rely on Lemma 6.4, and both of which fail in the absence of (6.3).

We may avoid the first of these inequalities entirely, thus removing the curvature condition (6.3) from the statement of Claim 6.16. First we make some observations concerning the local conductance ℓ.

Lemma 6.21. *The local conductance ℓ defined in (6.4) satisfies:*

(i) $\ell(x)^{1/n}$ *is concave over K;*

(ii) $\ln \ell$ *is Lipschitz; specifically* $\left|\ln \ell(x) - \ln \ell(y)\right| \le \dfrac{n}{\delta}\,\|x-y\|$, *for all $x, y \in K$.*

Proof (sketch). We are in a similar situation to that already encountered in the proof of Lemma 6.18: a convex body — there a slab defined by parallel $(n-1)$-dimensional planes, here a ball of radius δ — is translated in a straight line and its intersection with K studied with the aid of the Brunn-Minkowski Theorem (Theorem 6.19). The proof of part (i) here is analogous.

For part (ii), observe that the definition of the function ℓ, presented in (6.4), makes sense outside its official domain, namely K. Observe also that part (i) continues to hold over the larger region $K + B(0,\delta)$, the Minkowski sum of K and the ball of radius δ. Given $x, y \in K$, let z be the point colinear with x and y, at distance δ from y, and on the opposite side of y to x. Note that $z \in K + B(0,\delta)$. Thus, by part (i),

$$\delta\,\ell(x)^{1/n} + \|x-y\|\,\ell(z)^{1/n} \le (\delta + \|x-y\|)\,\ell(y)^{1/n},$$

and hence

$$\frac{\ell(x)}{\ell(y)} \le \left(\frac{\delta + \|x-y\|}{\delta}\right)^n.$$

Taking the logarithm of both sides,

$$\ln \ell(x) - \ln \ell(y) \le n \ln\left(\frac{\delta + \|x-y\|}{\delta}\right) \le \frac{n\,\|x-y\|}{\delta}.$$

Since the argument is symmetric in x and y, part (ii) of the lemma follows. $\square$

We may now avoid inequality (6.21) by taking a more direct route, which is opened up by replacing Lemma 6.18 by:

Lemma 6.22. *With $S_0, S_1, \ldots, S_{m-1}$ as in Lemma 6.18, the sequence*

$$\mu(S_0)^{1/2n},\ \mu(S_1)^{1/2n},\ \ldots,\ \mu(S_{m-1})^{1/2n}$$

is concave. Consequently, the sequence

$$\frac{1}{\mu(S_0)},\ \frac{1}{\mu(S_1)},\ \ldots,\ \frac{1}{\mu(S_{m-1})}$$

is convex.

This lemma follows from a functional version of the Brunn-Minkowski Theorem due to Dinghas [24, Satz 1]. We state this theorem in a slightly less general form than it appears in [24].

Theorem 6.23 (Dinghas)**.** *Suppose A_1 and A_2 are non-empty, bounded, measurable sets in $\mathbb{R}^n$, and let $A_0 = A_1 + A_2$ be the Minkowski sum of A_1 and A_2. Suppose further that f_1 and f_2 are measurable functions defined on A_1 and A_2, respectively, and form the function g_0 defined by*

$$g_0(x) = \sup\Big\{\big((f_1(x')^{1/r} + f_2(x'')^{1/r}\big)^r : x' \in A_1, x'' \in A_2 \text{ and } x' + x'' = x\Big\},$$

for all $x \in A_0$. If f_0 is any measurable function on A_0 satisfying $f_0(x) \geq g_0(x)$ for all $x \in A_0$, then

$$\left[\int_{A_0} f_0(x)\,dx\right]^{1/(r+n)} \geq \left[\int_{A_1} f_1(x)\,dx\right]^{1/(r+n)} + \left[\int_{A_2} f_2(x)\,dx\right]^{1/(r+n)}.$$

Proof of Lemma 6.22. In Theorem 6.23 make the following identifications: $r = n$, $A_1 = S_{i-1}$, $A_2 = S_{i+1}$, $f_1 = f_2 = \ell$ and $f_0(x) = 2^r\ell(x/2)$. By part (i) of Lemma 6.21, we then have $f_0 \geq g_0$, as required; also observe that $2S_i \supseteq S_{i-1} + S_{i+1} = A_0$. The first claim in Lemma 6.22 may then be read off from the concluding inequality of Theorem 6.23. The second claim uses the same reasoning as in the final step of the proof of Lemma 6.18. See also [55, Lemma 2.1]. □

Armed with Lemma 6.22, the upper bound on $a_{i,j}$ derived in the sequence of inequalities ending at (6.22) — with improved constant 1 in place of 2.5 — follows directly from the definition (6.17) of $a_{i,j}$. This establishes Claim 6.16 in the absence of the curvature condition (6.3).

The other place at which the curvature condition is used, namely in establishing (6.29), is trickier to handle. (Note that we used it in going from (6.28) to (6.29).) Our use of curvature is more substantial here, and we need to modify the partitioning of the needle-like body K_1 used in the proof of Claim 6.16 (see Figure 6.6) to recover the proof. If we are prepared to settle for a Poincaré constant λ smaller by a factor n (i.e., $\lambda = c_2\delta^2/D^2n^2$) then it is not too difficult to establish Theorem 6.7 in the absence of (6.3), and we shall see presently how this is done. Getting the correct (up to a constant factor) λ in the absence of (6.3) requires a more complicated analysis, which we only sketch here.

What is it we were trying to achieve with inequality (6.29)? Well, the final contradiction required us to find a set $I \subseteq K$ with the properties that: (i) every point of K_0 is within distance δ of every point of I; and (ii) the ratio $\mathrm{vol}_n I/\mathrm{vol}_n(B(x,\delta) \cap K)$ is bounded below by a universal constant for every $x \in K_0$. Without (6.3) there is currently no guarantee that such a set I exists. However, if we chop K_1 more finely, into slabs of width $\eta = c_3\delta/n$ (instead of $\eta = c_3\delta/\sqrt{n}$), then we are assured to find the required set I. This finer partition increases the number of slabs m by a factor $\sqrt{n}$, and hence reduces the Poincaré constant by a factor n. We borrow the following lemma from Kannan, Lovász and Simonovits [50, Lemma 3.5].

Lemma 6.24. *Suppose $\delta' > 0$, and $x, y \in K$ with $\|x - y\| \le \delta'/\sqrt{n}$. Then*

$$\begin{aligned}&\mathrm{vol}_n(B(x,\delta') \cap B(y,\delta') \cap K)\\ &\qquad \ge \frac{1}{1+e}\min\bigl\{\mathrm{vol}_n(B(x,\delta') \cap K),\ \mathrm{vol}_n(B(y,\delta') \cap K)\bigr\}.\end{aligned}$$

Recall that $\mathrm{vol}_n(B(x,\delta') \cap K)$ is proportional to $\ell(x)$. (This is by definition (6.4) of local conductance ℓ.) Now, with η smaller than before, part (ii) of Lemma 6.21 (the Lipschitz inequality for ℓ) ensures that $\mathrm{vol}_n(B(x,\delta') \cap K)$ varies by at most a constant factor as x ranges over K_0. So, choosing δ' a little less than δ, as before, we see that the set $I := B(z_1,\delta') \cap B(z_2,\delta') \cap K$ has the properties we desire: property (i) is by the triangle inequality, and property (ii) is by Lemma 6.24. This establishes Theorem 6.7 without assumption (6.3) but with λ smaller by a factor n.

Exercise 6.25. Flesh out the details of the above proof sketch.

Finally, some inadequate pointers on how to drop assumption (6.3) without losing the factor n in λ. Let's step back and consider what we need to have in order to be able to construct the contradictory set I, using Lemma 6.24. Certainly we need the slabs in the decomposition to have width $O(\delta/\sqrt{n}\,)$; but we also require that the local conductance ℓ varies by at most a constant factor over each slab. As we have seen, these two requirements can be met by using slabs of width $O(\delta/n)$, but then the number of slabs increases, and our estimate of the Poincaré constant worsens.

So it seems that we need to partition K_1 into slabs of unequal thickness, using thinner slabs where ℓ is rapidly varying. We might as well use the coarsest possible partition that will allow us to draw the final contradiction. Starting at the leftmost point of K_1, partition K_1 into slabs $S_0, S_1, \dots, S_{m-1}$ as in Figure 6.6, finishing with slab S_{m-1} at the rightmost point of K_1. Having created $S_0, S_1 \dots, S_{i-1}$, choose the plane defining S_i to be the rightmost plane subject to the conditions:

(i) the distance from the previous plane (i.e., the thickness of slab S_i) is at most $c_3\delta/\sqrt{n}$; and

(ii) the local conductance $\ell(x)$ varies by at most a factor 2 as x ranges over S_i.

Thus the partition of K_1 into slabs S_i is the coarsest possible, subject to conditions (i) and (ii).

Note that conditions (i) and (ii) together allow us to construct, using Lemma 6.24, the set I that leads to the final contradiction. We need of course to fix up the proof of Claim 6.16, which was conducted under the assumption that K_1 is partitioned into slabs of constant width $O(\delta/\sqrt{n}\,)$. Specifically, we need work harder to prove the key inequality (6.24).

Exercise 6.26. Complete the Proof of Theorem 6.7 (the Poincaré inequality) in the absence of the curvature condition (6.3), using the programme outlined above. The main technical challenge lies in reproving Claim 6.16 in the absence of (6.3), specifically in re-establishing (6.24), taking due account of the amended partition of K_1 into slabs. You will find that the partition of Figure 6.6 (using the amended construction just presented) can be divided into three sections: $S_0, \dots, S_{\ell-1}$, then $S_\ell, \dots, S_{r-1}$ and $S_r \dots, S_{m-1}$, where the slabs in the middle section are all of full width η, and the others are all of strictly smaller width. (Either or both of the outer sections may be empty.) The existence of such a division relies on log-concavity of the local conductance ℓ, which is a consequence of Lemma 6.21(i). The middle section is dealt with exactly as before, since the number of slabs contained within it is $r - \ell \leq D/\eta = O(D\sqrt{n}/\delta)$. In the left (right) sections it can be shown that $w_i = \mu(S_i)$ is increasing (decreasing) geometrically; thus the sum (6.17) is determined, up to a constant factor, by its first (last) term. (This step uses log-concavity of ℓ and Brunn-Minkowski.) Thus it doesn't matter so much that the number of terms in the sum (i.e., slabs in the partition) may grow faster than $O(D\sqrt{n}/\delta)$. Note that this is a challenging, verging on speculative, exercise. To keep the technical complexities within bounds, you may want to assume $\delta = O(D/\sqrt{n}\,)$. This is not a restriction in the volume application, where $\delta = \Theta(1/\sqrt{n}\,)$ and $D = \Omega(1)$. However, the assumption is a definite blemish, in that Theorem 6.7 should hold even when δ is of the same order as D.

Remark 6.27. Kannan, Lovász and Simonovits [50] restrict the function f to be an indicator function $f : K \to \{0, 1\}$. The parameter $\varPhi$ corresponding to λ in the inequality

$$\mathcal{E}_P(f, f) \geq \varPhi \operatorname{Var}_\mu f, \quad \text{for all (measurable) } f : K \to \{0, 1\}$$

is called the *conductance* of the ball walk. Since the class of functions f is restricted, the conductance $\varPhi$ is potentially larger than λ. However it is known — a version of Cheeger's inequality — that $\lambda \geq \frac{1}{2}\varPhi^2$. (See Sinclair [71] or Aldous and Fill [2] for relationships between various MC parameters, including these two.) The approach to the ball walk in [50] is to show that the conductance $\varPhi$ is of order $\delta/D\sqrt{n}$, which leads by Cheeger to the required bound on λ. However, the restriction of f to the class of indicator functions unfortunately does not seem to lead to any significant technical simplification in the proof.

6.7 Using samples to estimate volume

In order to estimate the volume of a convex body using our sampling procedure, we follow the basic "product of ratios" approach used in earlier examples. Briefly, the procedure is as follows.

Given our convex body K, we define a series of concentric balls $B_0 \subset B_1 \subset \cdots \subset B_k$ such that $B_0 \subseteq K$ and $K \subseteq B_k$. (Refer to Figure 6.11.) Additionally, we require that the volume of these balls does not grow too quickly, say $\operatorname{vol}_n B_{i+1} \leq 2 \operatorname{vol}_n B_i$. We can estimate the ratios

$$\varrho_i = \frac{\operatorname{vol}_n(B_i \cap K)}{\operatorname{vol}_n(B_{i+1} \cap K)}$$

by repeatedly sampling points from $B_{i+1} \cap K$ and determining the fraction of these points which lie also in $B_i \cap K$. Let Z_i be an estimate for ϱ_i obtained by taking the sample mean. We then get the desired estimate of $\operatorname{vol}_n K$ from

$$\operatorname{vol}_n K \approx \operatorname{vol}_n B_0 \cdot \prod_{i=0}^{k-1} \frac{1}{Z_i}.$$

Of course, we may calculate $\operatorname{vol}_n B_0$ from an explicit formula.

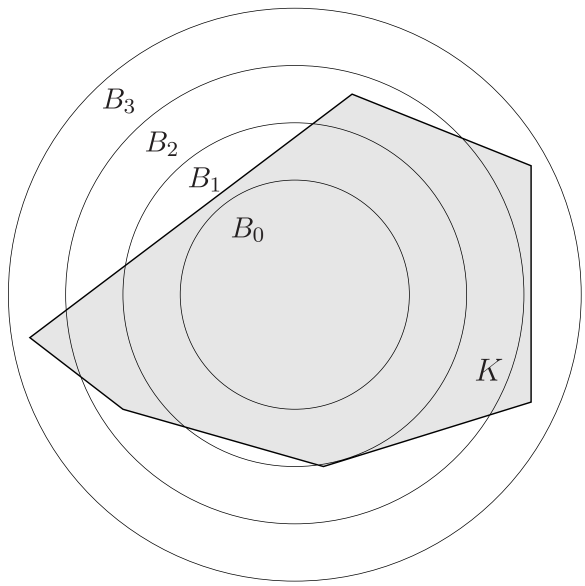

Figure 6.11: A convex body K and concentric balls

We have glossed over important issues here, not least the obvious fact that k must not be too large if we are to control the variance of our product estimator for $\mathrm{vol}_n K$. If K is "well rounded" then, indeed, k need not be very large. But if K is very elongated it will be necessary to apply a linear transformation to K to render it well rounded. For details of this step, and many further refinements, refer to [50].

6.8 Appendix: a proof of Corollary 6.8

We work with the lazy version of the ball walk, which stays put with probability $\frac{1}{2}$. For the first leg, we follow closely the proof of Theorem 5.6, but replacing sums by integrals. Because of the close similarity of the arguments we record only the main steps here:

$$[P_{\mathrm{zz}}f](x) = \frac{1}{2}\int_K P(x,dy)\,\big(f(x)+f(y)\big),$$

$$\mathrm{Var}_\mu(P_{\mathrm{zz}}f) \le \frac{1}{4}\int_K \mu(dx)\int_K P(x,dy)\big(f(x)+f(y)\big)^2,$$

and

$$\mathrm{Var}_\mu f = \frac{1}{2}\int_K \mu(dx)\int_K P(x,dy)\big(f(x)^2+f(y)^2\big).$$

It follows that

$$\begin{aligned}\mathrm{Var}_\mu f - \mathrm{Var}_\mu(P_{\mathrm{zz}}f) &\ge \frac{1}{4}\int_K \mu(dx)\int_K P(x,dy)\big(f(x)-f(y)\big)^2\\ &= \frac{1}{2}\mathcal{E}_P(f,f)\\ &\ge \frac{1}{2}\lambda\,\mathrm{Var}_\mu f,\end{aligned}$$

and hence

$$\mathrm{Var}_\mu(P_{\mathrm{zz}} f) \le \Big(1 - \frac{\lambda}{2}\Big) \mathrm{Var}_\mu f.$$

Iterating the above, we obtain

$$\mathrm{Var}_\mu(P^t_{\mathrm{zz}} f) \le \Big(1 - \frac{\lambda}{2}\Big)^t \mathrm{Var}_\mu f \le \exp(-\tfrac{1}{2}\lambda t). \tag{6.34}$$

Now suppose A is measurable subset of K, and let $f : K \to \mathbb{R}$ be the function that is 1 on A and 0 outside A. Assume that we start our walk from a point X_0 selected uniformly at random from the ball $B = B(x, \delta) \subseteq K$. (This is, of course, equivalent to starting the walk at point x at time -1.) For $\varepsilon > 0$ we want to find a time t such that the variation distance of the t-step distribution from stationarity is at most ε; equivalently, we require

$$|\Pr(X_t \in A) - \mu(A)| = \left| \frac{1}{\mathrm{vol}_n B} \int_B \big\{ [P^t_{\mathrm{zz}} f](y) - \mu(A) \big\}\, dy \right| \le \varepsilon, \tag{6.35}$$

uniformly over the choice of A. (In this context, recall the definition of total variation distance (3.2), and the fact that $[P^t_{\mathrm{zz}} f](y)$ may be interpreted as $\Pr(X_t \in A \mid X_0 = y)$.)

Noting $\mathbb{E}_\mu(P^t_{\mathrm{zz}} f) = \mu(A)$, we find

$$\begin{aligned}
\mathrm{Var}_\mu(P^t_{\mathrm{zz}} f) &\ge \int_B \big\{ [P^t_{\mathrm{zz}} f](y) - \mu(A) \big\}^2 \mu(dy) \\
&\ge \frac{0.4}{\mathrm{vol}_n K} \int_B \big\{ [P^t_{\mathrm{zz}} f](y) - \mu(A) \big\}^2 dy \qquad (6.36) \\
&\ge \frac{0.4 \,\mathrm{vol}_n B}{\mathrm{vol}_n K} \left[\frac{1}{\mathrm{vol}_n B} \int_B \big\{ [P^t_{\mathrm{zz}} f](y) - \mu(A) \big\}\, dy \right]^2, \qquad (6.37)
\end{aligned}$$

where inequality (6.36) follows from the definition (6.5) of μ and Lemma 6.4; and (6.37) from the fact that the expectation of the square of a r.v. is at least as large as the square of its expectation. Thus, to achieve the desired bound (6.35) on variation distance, we require

$$\mathrm{Var}_\mu(P^t_{\mathrm{zz}} f) \le \frac{0.4\,\varepsilon^2 \,\mathrm{vol}_n B}{\mathrm{vol}_n K}.$$

Now, the volume of K is maximised, for specified diameter D, when K is a ball of radius $D/2$. Thus it is enough that we achieve

$$\mathrm{Var}_\mu(P^t_{\mathrm{zz}} f) \le 0.4\,\varepsilon^2 \Big(\frac{2\delta}{D}\Big)^n.$$

According to (6.34), this inequality will hold, provided

$$t \ge \left\lceil \frac{2}{\lambda} \left(\ln\left\{ \frac{5}{2\varepsilon^2} \right\} + n \ln\left\{ \frac{D}{2\delta} \right\} \right) \right\rceil.$$

This is the mixing time claimed in Corollary 6.8, with $i(\mu_0)$ specialised to an initial distribution that is uniform and supported on a ball of radius δ.