Deterministic Volume Approximation

by

Rares Cristian

Faculty Signatures

(PRASAD TETALI)

Santosh Vempala

Contents

Introduction

Computing the volume of a convex body $K \subset \mathbb{R}^n$ has been a long-studied question, important from both a theoretical and practical perspective. A key aspect of the problem lies in the representation of the set. In the most general case, access to the set is provided solely through an oracle. However, given only a well-guaranteed separation oracle, no deterministic polynomial-time algorithm can approximate the volume to within a factor exponential in the dimension [1]. However, access to an explicit description of the set removes this obstacle. As such, we restrict our attention to computing the volume of polytopes, which are expressed as the intersection of halfspaces, $\{Ax \leq b\}$ where $A \in \mathbb{R}^{m \times n}$.

Surprisingly, there do exist efficient randomized algorithms to compute volume given only an oracle. The primary technique, which we will mainly follow, is to construct a series of bodies $K = K_0 \supset K_1 \supset \cdots \supset K_m$ in such a way that the volume of K_m can be easily calculated, and the ratio of volumes of K_i to K_{i+1} can be well approximated. This last step may be done by sampling uniformly at random from the interior of K_i and taking the ratio of points lying in K_{i+1}. The current state-of-the-art methods perform this sampling via random walks inside the body. The volume can be directly found as $\text{vol}(K) = \text{vol}(K_m) \prod_{i=0}^{m-1} \text{vol}(K_i)/\text{vol}(K_{i+1})$.

While there are numerous randomized algorithms that efficiently approximate the volume, no deterministic algorithm is currently known to exist. This is part of a fundamental question in algorithms: when is randomness truly needed? We investigate whether the notion of chaos can serve as a substitute for randomness in this setting. There is a distinction to be made here between chaos and randomness. Given the initial state of a system, we cannot predict the future state of a random process. On the other hand, a chaotic one is fully deterministic. But this process is sensitive to initial conditions: any small change in initial conditions will result in vastly different orbits. In particular, we consider the orbit created by the free motion of a point particle inside the polytope with mirror-like reflections off the boundary. We analyze two different reflection methods.

1) Flat Reflections: These are simple specular reflections off the facets.

2) Ball Reflections: Each facet is tiled by an infinite grid of width δ. At each gridpoint, we place a ball of radius R so that each point on the facet is covered by at least one ball. Reflections now occur against the balls instead of the facets.

Crucially, the ball reflections provide a chaotic attribute to the billiard flow. We are interested in determining if this flow is ergodic, which will imply that any orbit will almost

surely uniformly cover the polytope. Moreover, we work on bounding the system's rate of convergence to its limit distribution.

Finally, we give some attention to possible implementation issues. At each iteration of this algorithm, a small amount of precision is lost due to inherent precision issues in implementation. In a chaotic system, due to sensitivity to initial conditions, these small errors can lead to a vastly different trajectory than the original. The notion of shadowing comes into play here. Every pseudo-trajectory, which can be thought of as a numerically-computed trajectory with rounding errors introduced at each step, stays uniformly close to some true trajectory (note this may have a slightly perturbed initial position). That is, the pseudo-trajectory we generate is *shadowed* by a true one. Therefore, while not producing a true billiard path, we do generate a sequence which closely follows some true path.

Literature Review

2.1 Preliminaries

We first provide an introduction to the current approaches to approximating volume of high-dimensional polytopes $K \subset \mathbb{R}^n$, and afterwards present a connection to chaotic billiards.

2.1.1 Volume

We define the volume of K as the Lebesgue measure over K. This is a standard way of assigning measure to subsets of $\mathbb{R}^n$ and follows the standard definitions of area and volume in 2 and 3 dimensions that we are used to. We denote the volume of K as $\text{vol}(K)$.

2.1.2 Model of Computation

A key aspect of the problem lies in the representation of the set. Perhaps in the most general sense, a description of the body is not directly given, but rather access is provided by an oracle. A *membership oracle* may be given an input $x \in \mathbb{R}^n$ and returns whether or not x is contained in K. Alternatively, a *separation* oracle determines if x lies in K, and if not, provides a hyperplane separating x and K. A well-guaranteed oracle additionally provides an initial point, x_0 guaranteed to be in K, as well as bounds on the size of K: a guarantee that a ball of radius r is contained completely within K, and a guarantee that a ball of radius R fully contains K. That is, $x_0 + rB^n \subset K \subset x_0 + RB^n$.

On the other hand, we may be given an explicit description of the set. The most common example is that of a polytope. Even here, there are various ways to represent it, for example, as the convex hull of a set of points, or as the intersection of halfspaces. The latter is often more expressive, since with only a polynomial number of halfspaces, we potentially require an exponential number of vertices to describe the same polytope. Take the simple example of a cube in n dimensions. It is the intersection $2n$ halfspaces, but has 2^n corners.

2.2 Algorithms

Given only a well-guaranteed separation oracle, no deterministic polynomial-time algorithm can approximate the volume to within a factor exponential in the dimension [1]. For any sequence of n^a points there exist two different convex sets with volume ratio

$$\left(\frac{cn}{a \log n}\right)^n$$

whose oracles produce the same result to each query (for some universal constant c). Therefore, given no further information, there is no way to differentiate between the two bodies. Surprisingly, there do exist efficient randomized algorithms to compute volume given only an oracle. Volume computation is one of a few problems where randomization provably helps.

2.2.1 Overview

One major technique is to construct a series of bodies $K = K_0 \supset K_1 \supset \cdots \supset K_m$ in such a way that the volume of K_m can be easily calculated, and the ratio of volumes of K_i to K_{i+1} can be well approximated. This last step may be done by sampling uniformly at random from the interior of K_i and taking the ratio of points lying in K_{i+1}. The current state-of-the-art methods perform this sampling via random walks inside the body. The volume can be directly found as

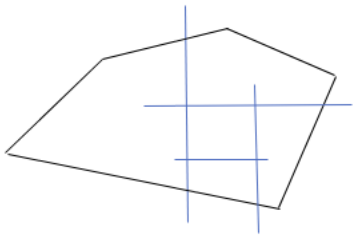
$$\text{vol}(K) = \text{vol}(K_m) \prod_{i=0}^{m-1} \frac{\text{vol}(K_i)}{\text{vol}(K_{i+1})}$$

At this point, one wishes for two things: to reduce the number of bodies K_i needed, and to reduce the number of samples required at each step. Note that the errors in approximating volume ratios are multiplicative; having fewer intermediary bodies allows us to have larger errors in approximating volume ratios, which in turn allows us to sample fewer points. At the same time, we want the ratios to be fairly large, so K_{i+1} needs to have a constant fraction of the volume of K_i. This is so we only require a polynomial number of samples to accurately approximate the ratio.

2.2.2 Constructing K_i

In addition to a convex body $K \subset \mathbb{R}^n$, say we are given that a cube of side length $r \in R$ contained in K, as well as an oracle for computing the centroid. Let e_j be an axis vector so that the width of K_i along e_j is at least r. Let z_i be the center of mass of K_i. Consider the hyperplane through z_i perpendicular to e_j. This hyperplane divides the polytope into two halves, and we let K_{i+1} be the half which contains z_0, the center of mass of K.

Figure 2.1: Centroid Cutting Plane Method



Since we choose these hyperplanes perpendicular to the axes, we terminate this process when K_m is a box, whose volume is simply the product of its side lengths. This is guaranteed to terminate once the width along each direction is at most r. It remains to bound m. But first, we require some more geometric results.

2.2.3 Geometry

Brunn-Minkowski Inequality

We begin with a general question: how does volume of a convex body change as we move along some direction θ? For simplicity, we will assume that θ is the x_1 direction. Let $K(t)$ be the $(n-1)$-dimensional "slices" of K perpendicular to θ. That is,$K(t) = K \cap \{x \in \mathbb{R}^n : x_1 = t\}$. What can we say about $K(t)$ as we vary t? We begin with a useful inequality regarding volumes of arbitrary sets, not only convex bodies.

Theorem 2.2.1 (Brunn-Minkowski Inequality)**.** *Let $A, B \subset \mathbb{R}^n$ be compact measurable sets. Then,*

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

where $A + B$ is the Minkowski sum of two sets, $\{a + b : a \in A, b \in B\}$

Since $\text{vol}(\lambda A) = \lambda^n \text{vol}(A)$, we have the following equivalent version of the inequality, which we use in the next lemma.

$$\begin{aligned}
\text{vol}(\lambda A + (1-\lambda)B)^{1/n} &\geq \text{vol}(\lambda A)^{1/n} + \text{vol}((1-\lambda)B)^{1/n}, \\
&\geq (\lambda^n \text{vol}(A))^{1/n} + ((1-\lambda)^n \text{vol}(B))^{1/n} \\
&\geq \lambda \text{vol}(A)^{1/n} + (1-\lambda)\text{vol}(B)^{1/n}
\end{aligned}$$

Moreover, this implies that the volume function is $1/n$-concave with respect to Minkowski sum.

Lemma 2.2.2. *The function $(\text{vol}_{n-1} K(t))^{1/(n-1)}$ is concave.*

Proof. Let $x, y \in \mathbb{R}, \lambda \in [0,1]$ and $c = \lambda x + (1-\lambda)y$. Let $a \in K(x), b \in K(y)$. By convexity, points of the form $\lambda a + (1-\lambda)b$ all lie in K. Since they also have their first coordinate equal to the first coordinate of c, we find that $\lambda K(x) + (1-\lambda)K(y) \subset K(c)$. By the Brunn-Minkowski inequality (2.2.1),

$$\text{vol}(K(c))^{1/(n-1)} \geq \lambda \text{vol}(K(x))^{1/(n-1)} + (1-\lambda)\text{vol}(K(y))^{1/(n-1)}$$

Note that the above is in $(n-1)$ dimensions, since $K(t)$ is an $(n-1)$-dimensional slice of K.

□

Hyperplanes through the Centroid

Definition 2.2.3 (Centroid)**.** *The center of mass, or centroid, of a body $K \subset \mathbb{R}^n$ is*

$$\frac{1}{\text{vol}(K)} \int_K x dx$$

Let's take a look at a specific example: the cone. Clearly, the centroid will be on the perpendicular drawn from the apex to the base, so it remains to find the height from the base. We can integrate along the height of the cone, C, and each cross-section is an $(n-1)$-dimensional ball of radius tR/h, where R is the radius of the base, and h is the height of the cone. We say the volume of an ball of radius r in n dimensions is $f(n) \cdot r^n$, for some function f. So, the centroid is at height h^* where

$$\begin{aligned} h^* &= \frac{1}{\text{vol}(C)} \int_{t=0}^{h} t \cdot f(n-1) \left(\frac{tR}{h}\right)^{n-1} dt \\ &= \frac{f(n-1)}{\text{vol}(C)} \frac{R^{n-1}}{h^{n-1}} \int_{t=0}^{h} t^n dt \\ &= \frac{f(n-1)}{\text{vol}(C)} \frac{R^{n-1} h^2}{n+1} \end{aligned}$$

Since the volume of C is

$$\text{vol}(C) = \frac{f(n-1)R^{n-1}h}{n}$$

we find that $h^* = \frac{n}{n+1}h$. Equivalently, the centroid is at a distance of $\frac{1}{n+1}h$ from the base.

Theorem 2.2.4 (Grunbaum's Inequality)**.** *Given a convex body $K \subset \mathbb{R}^n$, any halfspace H which contains the center of gravity of K, also contains $1/e$ of the volume of K.*

We have a related result regarding the width. Define the support function of K as $h_K(x) = \max_{y \in K} \langle x, y \rangle$ for $x \in K$.

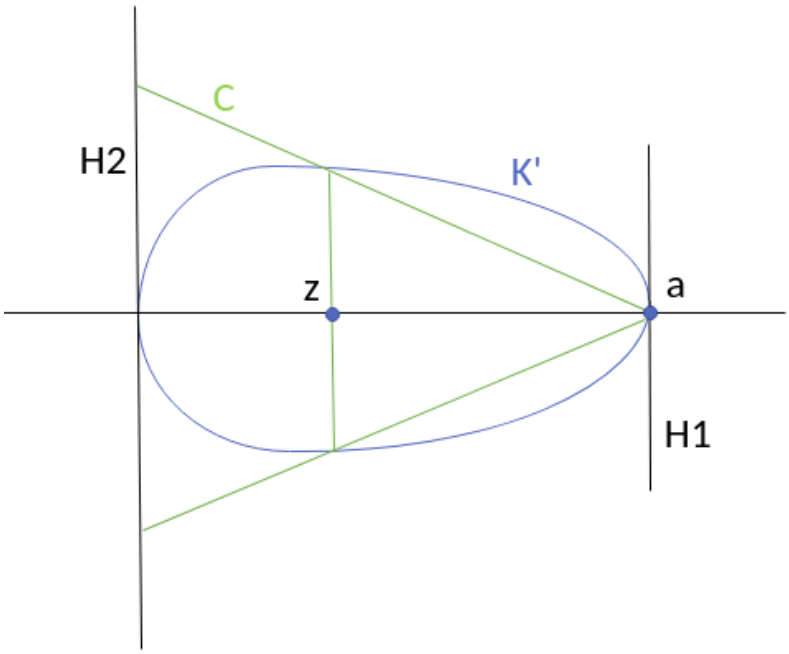
Lemma 2.2.5. *Given body K with centroid at the origin, for any unit vector θ,*

$$\frac{h_K(\theta)}{h_K(-\theta)} \leq n$$

Proof. Again, for simplicity, assume that θ is the positive x_1 direction. Consider replacing each slice $K(t)$ with an $(n-1)$-dimensional ball of equal volume to create a body K'. Let $r(t)$ be the radius of this ball. The volume of $K(t)$ is $f(n-1) \cdot r(t)^{n-1}$. By 2.2.2, $r(c) \geq \lambda r(x) + (1-\lambda) r(y)$. Hence, r is concave.

Let H_1, H_2 be the two supporting hyperplanes of K (and thus of K' as well). Let z be the centroid of K and a the point on H_1 along θ. Consider the cone C' with apex a and base $K(z)$. Extend this to cone C which has base on H_2 as shown below.

Figure 2.2: Cone C and rounded body K' from lemma 2.5



For C, we have precisely $\frac{h_K(\theta)}{h_K(-\theta)} = n$. So, we want to show that z is to the right of c, the centroid of the cone C. This would imply the inequality. Suppose instead that c is to the right of z. Define the following

$$H_z^+ = K' \cap \{x \in \mathbb{R}^n : x_1 \geq z_1\}$$
$$H_z^- = K' \cap \{x \in \mathbb{R}^n : x_1 \leq z_1\}$$
$$H_c^+ = C \ \cap \{x \in \mathbb{R}^n : x_1 \geq c_1\}$$
$$H_c^- = C \ \cap \{x \in \mathbb{R}^n : x_1 \leq c_1\}$$

Since c is the centroid of C, the moments of H_c^+ and H_c^- are equal about the hyperplane through c. Call this value M. By convexity, $H_z^+ \supset H_c^+$, so the momentum, M^+, of H_z^+ about the hyperplane through z is larger than M. Similarly, since the radius function is concave, $H_z^- \subset H_c^-$, hence the momentum, M^-, of H_c^- about the hyperplane through z is less than M. This leads to a contradiction since M^- should equal M^+ as z is the centroid of K'.

□

The proof of 2.2.4 is very similar, so we omit it here. Both of these results are tight, which can be seen for a cone.

Isotropy

Definition 2.2.6 (Isotropic Position)**.** *A set S is in isotropic position if both $\mathbb{E}_S[x] = 0$ and $\mathbb{E}_S[x^T x] = I$.*

Moreover, we for any convex set S, there exists an affine transformation which brings S into isotropic position. We may assume that $\mathbb{E}_S[x] = 0$ since a simple translation can bring the centroid to the origin. Let C be the covariance matrix $\mathbb{E}_S[xx^T]$. Since C is positive definite, there exists a matrix A with $A^2 = C$. So, the body $AS = \{A^{-1}x : x \in S\}$ is isotropic: letting $y = A^{-1}x$, $\mathbb{E}[yy^T] = A^{-1}\mathbb{E}[xx^T]A^{-1} = I$.

Theorem 2.2.7. *For a body $K \subset \mathbb{R}^n$ in isotropic position,*

$$\sqrt{\frac{n+2}{n}}B^n \subseteq K \subseteq \sqrt{n(n+2)}B^n$$

Essentially, a ball of unit radius is contained in K, and a ball of radius $O(n)$ contains K. Isotropy is a useful property that we'll make use of throughout. Often, if we are interested in some affine-invariant property of a body, we can restrict our attention to isotropic bodies. For example, the centroid of a body is affine-invariant. That is, if f is some affine transformation, f maps the centroid of K to the centroid of $f(K)$.

2.2.4 Bounding Iterations

Note that since the center of gravity is invariant under affine transformations, so is the algorithm described. We now focus on bounding m for a body in isotropic position, which will directly imply the general case.

Suppose K contains a cube of side length r, and is contained in a cube of side length R. Recall, the algorithm terminates when the width along each standard direction, e_j is at most r. Thus, for the support $[a, b]$ along e_j we have $a, b \geq r/n$. The final volume of K_m will be $(r/n)^n$. The initial volume is at most R^n. Moreover, by 2.2.4

$$\frac{\text{vol}(K_{i+1})}{\text{vol}(K_i)} \leq 1 - \frac{1}{e}$$

That is, the volume at each step is guaranteed to decrease by a constant fraction. Therefore, $m = O(n \log nR/r)$. For an isotropic body, this comes out to $O(n \log n)$ phases.

However, computing the centroid exactly is #P-hard [2]. On the other hand, the average of $O(\log^2 m)$ sample points provides a good approximation of the centroid if K is a polytope, and $O(n)$ for arbitrary convex bodies [3]. By "good" approximation, we mean that in expectation, any halfspace containing their average will also contain a constant fraction of the volume.

It remains to find these uniform samples. Existing algorithms do this via Markov Chains methods. We now give a brief overview of the main ideas.

2.3 Sampling

A geometric random walk is a sequence of points $x_0, x_1, \ldots$ in $\mathbb{R}^n$ for which x_{i+1} is chosen from a neighborhood of x_i according to some distribution depending only on x_i. In particular, we aim for random walks within K whose stationary distribution is the uniform distribution over K. Additionally, we wish to bound the rate of convergence. Here, we have a brief discussion of random walks, as well as isoperimetry, a key tool used in their analysis.

2.3.1 Examples

- Grid Walk

  The first major breakthrough in a polynomial-time algorithm for approximating volume came from Dyer and Frieze [4] in which they first introduced the ball walk. In it, we define a grid of width δ, and begin at a gridpoint x_0 within the body. With probability $\frac{1}{2}$ we stay in place, and with probability $\frac{1}{2}$ choose point y uniformly at random from the $2n$ neighbors of x_i. If $y \in K$, set $x_{i+1} = y$, otherwise, $x_{i+1} = x_i$.

- Ball Walk

  This is another lazy walk, as with probability $\frac{1}{2}$ we stay in place. Here, we pick a uniform random point y from the ball of radius δ centered at the current point x. If $y \in K$, go to y. Otherwise, stay at x. Some care needs to be taken here. If x is near a corner, it possible that an exponentially small portion of the ball centered at x will lie within K. This would imply that to take a single step, we may require an exponential number of steps.

Intuitively, it seems like this may not be an issue since during the algorithm the probability of reaching such a corner is equally small. However, x_0 itself may be one of these pathological points. One possible solution is to begin from a *warm start*. That is, we have an initial point drawn from a distribution already close to the uniform distribution (closeness here is measured by total variation distance).

We now discuss some main techniques on bounding the rate of convergence of random walks to their stationary distribution. First, we introduce the notion of conductance.

Definition 2.3.1. *The* conductance *of a Markov chain is*

$$\Phi(A) = \min_{0<Q(A)\leq 1/2} \frac{\int_A P_u(K \setminus A) dQ(u)}{Q(A)}$$

where $P_u(X)$ is the probability of taking a step from point u to the set X.

Suppose the conductance of a Markov chain is small, then there must be some subset from which it is hard to escape from. This in turn would require more iterations to converge to the limit distribution. We make this intuition more precise:

Theorem 2.3.2. *We define the "distance" between two distributions P, Q as their total variation distance $\|P - Q\|_{TV}$. Let Q_i be the distribution after i iterations, and Q the limit distribution. Then,*

$$\|Q - Q_i\| \leq \|Q - Q_0\| \cdot \left(1 - \frac{\Phi^2}{2}\right)^i$$

So, it remains to bound the conductance. This notion is closely related to isoperimetry, a purely geometric property of the body.

2.3.2 Isoperimetry

The classical isoperimetric problem is to find a set with minimal surface area for a given volume. The solution to this problem, the sphere, has been known since the ancient Greeks. Here we pose a similar question regarding the boundaries of a surface dividing K into two parts: what is the largest ψ such that

$$\mathrm{vol}_{n-1}(\partial S) \geq \psi \min\{\mathrm{vol}(S), \mathrm{vol}(K \setminus S)\}$$

for any $S \subset K$.

Let's take a closer look at the conductance of the ball walk. To have any chance of taking a step from S to $K \setminus S$, we must initially be near the boundary between the two sets. If the volume of points near the boundary is large then it seems likely that the conductance is high. Indeed,

Theorem 2.3.3 (Isoperimetry [5]). *The ball walk mixes from a warm start in*

$$O^*\left(\frac{n^2}{\psi^2}\right)$$

steps.

This is a fairly broad overview of the main ideas. For a complete survey on geometric random walks, we defer to [6].

2.4 Dynamical Systems

We now change track and discuss dynamical systems which will lead us into dynamical billiards. The next sections present some main ideas that hint toward a different approach to finding samples.

Definition 2.4.1. *A* topological dynamical system *consists of a metric space X and a continuous map $T : X \to X$. T is called* reversible *if it is a homeomorphism. We call X the* phase space *of T.*

Our primary focus will be in describing the trajectory of points under repeated application of the map T. Here, we present the main notions that we will be using in later sections.

Definition 2.4.2. *Given a dynamical system $T : X \to X$, the* orbit *of a point $x \in X$ is the set $\mathcal{O}_x = \{T^k(x) : x \in \mathbb{Z}\}$*

Definition 2.4.3. *A point $x \in X$ is a* fixed *point of T if $T(x) = x$ and therefore, $\mathcal{O}_x = \{x\}$. Similarly, a point $x \in X$ is a* periodic *point of T if there exists some $k \geq 1$ such that $T^k(x) = x$. The minimal such k is the the* period *of x. Moreover, x is a fixed point of T^k.*

Definition 2.4.4. *We say T is a contraction if there exists a constant $0 \leq \tau < 1$ such that*

$$d(T(x), T(y)) \leq \tau d(x, y), \ \forall x, y \in X$$

Theorem 2.4.5 (Banach Fixed-Point Theorem). *Let (X, d) be a complete metric space with contraction mapping $T : X \to X$ with contraction constant τ. Then, T admits a unique fixed point $x^* \in X$. Moreover, for any $x_0 \in X$, the sequence $\{x_n\}$ defined by $x_n = T(x_{n-1})$ converges to x^* at the following rate:*

$$d(x^*, x_n) \leq \frac{\tau^n}{1-\tau} d(x_1, x_0)$$

2.4.1 Hyperbolicity

Hyperbolicity is characterized by a uniform contraction in one direction and a uniform expansion in the other direction. Formally,

Definition 2.4.6. *Let M be a compact C^∞ Riemannian manifold, and let $\phi : M \to M$ be a diffeomorphism. An invariant set $\Lambda \subset X$ is* hyperbolic *if for each $x \in \Lambda$, the tangent space $\mathcal{T}_x M$ splits into a direct sum*

$$\mathcal{T}_x M = E^s(x) \oplus E^u(x)$$

where E^s, E^u are invariant under T in the sense that

$$D\phi(x)E(x) = E(\phi(x)),\ \forall x \in \Lambda$$

for both $E = E^s$ and $E = E^u$. Additionally, there exist constants $C \geq 1$ and $0 < \mu < 1$ so that

$$|D\phi^k(x)v| \leq C\mu^k|v|,\ \ \forall v \in E^s(x), \forall k \geq 0$$
$$|D\phi^{-k}(x)v| \leq C\mu^k|v|,\ \ \forall v \in E^u(x), \forall k \geq 0$$

Hyperbolicity is a particularly useful property of a dynamical system. Many interesting results follow from it, as we see in later sections.

2.5 Dynamical Billiards

A billiard system is defined by a domain $D \subset \mathbb{R}^n$ and the free motion of a point particle inside D with mirror-like reflections off the boundary ∂D. In two dimensions, reflection is most easily described as enforcing that the angle of incidence is equal to the angle of reflection. In higher dimension this still holds true, however there are many directions which would all have the same angle of reflection. Instead, we relate the direction of movement v' after the collision to the direction v before the collision as

$$v' = v - 2\langle v, n\rangle n$$

where n is the unit normal vector to the point of collision. The boundary ∂D is described as a finite union of smooth ($C^k, k \geq 3$) $(n-1)$-dimensional surfaces, $\partial D = \Gamma_1 \cup \cdots \cup \Gamma_r$. For example, if D is a polytope, then Γ_i are its facets. Note that reflection is not properly defined at the "edges" of D, $\partial\Gamma_1 \cup \cdots \cup \partial\Gamma_r$. We distinguish between three types of boundaries.

Definition 2.5.1. Γ_i *is said to be* dispersing *if it is convex inward,* focusing *if it is concave inward, and* flat *otherwise.*

We now present billiards in the setting of dynamical systems. The state of a particle can be described by its position $q \in D$ along with the direction it is moving in, $v \in S^{n-1}$.

Definition 2.5.2 (Phase Space)**.** *The phase space of the system is*

$$\Omega = \{(q, v) : q \in D, v \in S^1\} = D \times S^{n-1}$$

Let $\pi_q(x) = q$ *be the projection of* $x = (q, v) \in \Omega$ *onto* D *and* $\pi_v(x) = v$ *the projection onto* S^{n-1}*.*

We define a *billiard flow* on the phase space as a family of maps $\{\Phi^t\}$. Φ^t advances a point in Ω forward in time by t units. This family forms a group as $\Phi^{s+t} = \Phi^s \circ \Phi^t$ and $\Phi^{-s} \circ \Phi^s = \Phi^0 = \mathrm{Id}$.

Definition 2.5.3. *The* orbit*,* $\mathcal{O}_x$ *of a point* $x \in \Omega$ *is the set* $\{\Phi^t x : t \in \mathbb{R}\}$*. The* billiard trajectory *of* x *is the projection of its orbit,* $\{\pi_q(\Phi^t x)\}$*, onto the table.*

A natural discretization of billiards is to directly map boundary points to boundary points since the trajectory between collisions is a straight line. Define a hypersurface $M \subset \Omega$ which consists of the post-collision vectors after reflection with the boundary: $M = \{(q, v) \in \Omega : q \in \partial Q, \langle v, n\rangle \geq 0\}$. Finally we define a return map $T : M \to M$ which describes successive collisions of the orbit of x. Let $Tx = \Phi^{\tau(x)}x$ where $\tau(x) = \min\{t > 0 : \Phi^t x \in M\}$.

Let's go back to the original volume problem. Suppose that some billiard trajectory will spend time in a subset $A \subset K$ proportional to $\mathrm{vol}(A)/\mathrm{vol}(K)$ for any A. That is, the orbit uniformly covers the polytope. Clearly, we cannot run this process indefinitely, however, perhaps it converges to a near uniform distribution in some polynomial number of reflections which we can use in place of random samples. Billiard tables that have all Γ_i dispersing have some of the strongest properties. Indeed, in two dimensions, if the boundary is additionally smooth, the billiards are hyperbolic, ergodic, mixing [7], and have exponential decay of correlations [8]. We define these in a later section, but roughly they imply a fast rate of convergence to their limit distribution.

2.5.1 Unfolding

For simplicity, we express the following ideas mainly in 2 dimensions, although they easily generalize to higher dimension. Here we present a useful approach to analyzing billiards in polygonal tables. Instead of reflecting the orbit off the edge of the polygon it intersects, we reflect the polygon itself over this edge (see figure 2.3).

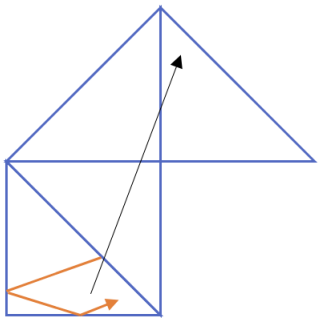


Figure 2.3: Unfolding

If we folded the triangles back, the black line would directly overlap with the true trajectory, the orange line. This allows us to inspect a single straight line. The cleanest unfoldings are for polygons which can edge-tessellate the plane such as the square and the triangle shown above.

2.5.2 The Square

Before taking a look at the general case of billiards, we present some key results for the square as it is simpler to work with. We will look at orbits moving along rational directions and show that these are always periodic. Moreover, orbits of irrational directions are not periodic and uniformly cover the square.

Uniformity

Here, we will assume that v is irrational. We first show that we can reduce this to a one-dimensional problem.

Without loss of generality, assume that the square has side length 1, $v_y = 1$, $v_x \in \mathbb{R} \setminus \mathbb{Q}$, and the initial point q_1 lies on the bottom edge of the square. Let's take a closer look at the tiling created by unfolding the square. Each square can be identified by a row and column as shown below, where $(0, 0)$ corresponds to the initial square.

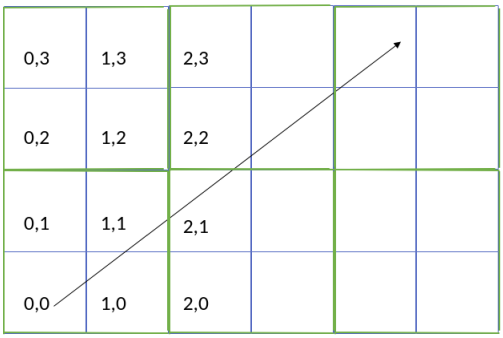


Figure 2.4: Tiling

Squares at coordinates (a_1, b_1), (a_2, b_2) are translations of each other if $a_1 \equiv a_2 \pmod 2$ and $b_1 \equiv b_2 \pmod 2$, since an even number of reflections along the same direction has no change on symmetry. So, the green squares above are translations of each other.

Let q_i denote the i^{th} intersection of the line with the horizontal edges of the green squares. These points map back to points on the bottom edge of our initial green square at $(0, 0)$. Indeed, these will be $((q_i)_x \bmod 2, 0)$. Later, we prove that these points are equidistrbituted along that edge. We first use this to prove the uniformity of the orbit.

For the moment, let's only consider portions of the orbit which travel from the bottom edge to the top edge of the square. These are described by the black arrows in the image below. Since the bottom endpoints are equidistributed, the orbits themselves uniformly cover the space shown.

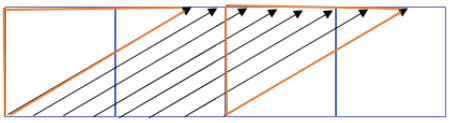


Figure 2.5: Partial Unfolding

The two orange triangles are translations of each other, so indeed the partial orbit we're considering uniformly covers the square. A similar argument holds for the portions of the orbit which begin at the top edge and travel to the bottom edge, so the entire orbit uniformly covers the square.

Lemma 2.5.4. *The sequence* $a_i \pmod 2$ *is equidistributed in the interval* $[0, 2]$.

Proof. The line will intersect the vertical edge after moving 2 units upwards. Therefore, it moves $2v_x$ units to the right before hitting the bottom. So, we are interested in showing that the sequence $a_0 + 2kv_x \pmod 2$, where v_x is irrational, is equidistributed in $[0, 2]$. This is a well known result due to Weyl [9].

□

This proof of uniformity is fairly specific to the square, and is not generalizable to other shapes. In the next section, we explore a more useful criterion for determining uniformity.

2.6 Ergodicity and Mixing

Definition 2.6.1. *Let $T : X \to X$. A function f is T-invariant if $f \circ T = f$. Similarly, a set $A \subset X$ is T-invariant if $T^{-1}(A) = A$.*

Definition 2.6.2. *Let T be a measure-preserving map on a probability space $(X, \mathcal{A}, \mu)$. T is ergodic if every T-invariant set has measure 0 or 1.*

Proposition 2.6.3. *If a measurable function $f : X \to \mathbb{R}$ is invariant under an ergodic map T, then f is constant almost everywhere.*

Proof. Define the level sets $A_c = \{x \in X : f(x) \leq c\}$. We first show that A_c is T-invariant. Suppose $x \in A_c$. Then $f(x) \leq c$, and by invariance, $f(T(x)) \leq c$. Finally $T(x) \in A_c$ and so $A_c \subset T^{-1}(A_c)$. We can similarly show $T^{-1}(A_c) \subset A_c$ and hence $T^{-1}(A_c) = A_c$. By the ergodicity of T, $\mu(A_c) = 0$ or $\mu(A_c) = 1$. Let $p = \inf\{c : \mu(A_c) = 1\}$. Then, since $\mu(A_{p-1/n}) = 0$, $f(x) \geq p$ a.e. and since $\mu(A_p) = 1$, $f(x) \leq p$ a.e. The claim follows.

□

Theorem 2.6.4 (Birkhoff-Khinchin)**.** *Let $(X, \mathcal{A}, \mu)$ be a probability space, and $T : X \to X$ a measure-preserving map. If $f : X \to \mathbb{R}$ is an integrable function, the limit*

$$\tilde{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j(x))$$

exists for almost everywhere $x \in X$, and the function $\tilde{f}$ is T-invariant, integrable, and

$$\int_X \tilde{f} d\mu = \int_X f d\mu$$

The function $\tilde{f}$ is called the time average *of f.*

With this, we investigate how the orbit of a point x under an ergodic map T covers X.

Definition 2.6.5 (Frequency of Return)**.** *Let $A \subset X$, $x \in X$. The limit*

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \text{card}\{0 \leq m < n : T^m(x) \in A\}$$

is defined as the frequency of returns of the point x to the set A.

Let χ_A denote the indicator function of A. Then,

$$\tau(x, A) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_A(T^j(x))$$

As defined in theorem 2.6.4, with χ_A taking the place of f, $\tau(x, A) = \tilde{\chi}_A(x)$. Therefore,

$$\begin{aligned}\int_X \tau(x, A)d\mu(x) &= \int_X \tilde{\chi}_A(x)d\mu(x) \\ &= \int_X \chi_A(x)d\mu(x) \\ &= \mu(A)\end{aligned}$$

with the second equality following from 2.6.4. Note that we have not yet used the fact that T is ergodic. We can easily see that τ is T-invariant as a function of x since $\tau(x, A) = \tau(T(x), A)$. By 2.6.3, it follows that $\tau(x, A)$ is a constant and so

$$\begin{aligned}\mu(A) &= \int_X \tau(x, A)d\mu(x) \\ &= \tau(x, A) \int_X d\mu(x) \\ &= \tau(x, A)\mu(X) \\ &= \tau(x, A)\end{aligned}$$

for almost everywhere $x \in X$. Hence, the orbit of $x \in X$ almost surely spends time in the set A proportional to its measure, $\mu(A)$. We now present a stronger property than ergodicty: mixing.

Definition 2.6.6. *A measure-preserving endomorphism T on a probability space $(X, \mathcal{A}, \mu)$ is said to be* mixing *if for any $A, B \in \mathcal{A}$,*

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$$

Intuitively, one can see this as describing the mixing of two different liquids. Suppose we're pouring red-colored water into a glass of water. Let A be the region originally occupied by the red water and B is any subset of the glass. We expect that eventually the proportion of red water in B will be the same as in the entire glass.

Proposition 2.6.7. *Any mixing map is also ergodic.*

Proof. Suppose $A \subset X$ is a T-invariant set. Since $T^{-n}(A) = A$, $\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A \cap B)$. Since T is mixing, $\mu(A \cap B) = \mu(A)\mu(B)$. Take $A = B$. Then, $\mu(A) = \mu(A)^2$ and hence $\mu(A) = 0$ or $\mu(A) = 1$. □

2.7 Decay of Correlations

Given that a system is ergodic, we would additionally like to describe the rate at which a trajectory approaches its final distribution. Consider the sequence

$$S_n = \sum_{j=0}^{n-1} f(T^j(x))$$

By the Birkhoff Ergodic Theorem 2.6.4, S_n/n converges to $\int_X f d\mu$ almost everywhere. Intuitively, this can be thought of as the law of large numbers in a dynamical systems setting: time average converges to space average. Are there any other relations to probability theory? For instance, does $f(T^n(x))$ effectively become independent of $f(x)$? We make this precise via correlation functions:

Definition 2.7.1. *We denote by C_f the time correlation function of dynamical system T where*

$$C_f(n) = \int_X f(T^n(x))f(x)d\mu - \left(\int_X f(x)d\mu\right)^2$$

Proposition 2.7.2. *Let $T : X \to X$ be a mixing map preserving measure μ. For any $f, g \in \mathcal{L}^2(X)$,*

$$\lim_{n\to\infty} \int_X f(T^n(x))g(x)d\mu = \int_X f(x)d\mu \int_X g(x)d\mu$$

Specifically, if the map T is mixing, C_f converges to 0, and we say that the correlations *decay*. Just as we had an analog of the law of large numbers, we also have one for the central limit theorem.

Definition 2.7.3. *We say that f satisfies the central limit theorem if*

$$\lim_{n\to\infty} \mu\left\{x : \frac{S_n(x) - n\int_X f d\mu}{\sqrt{n}} < z\right\} = \frac{1}{\sqrt{2\pi\sigma_f}} \int_{-\infty}^{z} \exp\frac{-s^2}{2\sigma_f^2} ds$$

for all $z \in \mathbb{R}$, where $\sigma_f \geq 0$ is a constant related to the correlation function as

$$\sigma_f^2 = C_f(0) + 2\sum_{n=1}^{\infty} C_f(n)$$

While the Birkhoff Ergodic theorem assured the existence of the limit, the above provides detail into the rate of convergence. The deviation of S_n/n from $\int_X f d\mu$ (scaled by $1/\sqrt{n}$) is asymptotically Gaussian. However, this only holds if $\sum_{n=1}^{\infty} C_f(n)$ converges. We say this converges *exponentially* if $|C_f(n)| \leq c \cdot e^{-an}$ for some constants $a, c > 0$ and *polynomially* if $|C_f(n)| \leq c \cdot n^{-a}$.

2.8 Shadowing

For the moment, let (X, d) be a metric space, and $T : X \to X$ a dynamical system.

Definition 2.8.1. *A sequence $\xi = \{x_k \in X : k \in \mathbb{N}\}$ is a δ-pseudo-orbit of T if*

$$d(T(x_k), x_{k+1}) \leq \delta, \ \forall k \in \mathbb{N}$$

Definition 2.8.2. *We say that a point $x \in X$ ϵ-shadows a δ-pseudo-orbit $\xi = \{x_k\}$ if*

$$d(T^k(x), x_k) \leq \epsilon, \ \forall k \in \mathbb{N}$$

Additionally, T has the shadowing property *on $Y \subset X$ if given $\epsilon > 0$, there exists $\delta > 0$ such that for any δ-pseudo-orbit $\xi \subset Y$, there is a point x that ϵ-shadows ξ.*

Every pseudo-orbit, which can be thought of as a numerically-computed orbit with rounding errors introduced at each step, stays uniformly close to some true orbit (note this may have a slightly perturbed initial position). That is, the pseudo-trajectory we generate is *shadowed* by a true one.

We are interested in determining which classes of dynamical systems T have the shadowing property, as well as for which sets $Y \subset X$.

Lemma 2.8.3. *Suppose a dynamical system $T : X \to X$ is a contraction with constant τ. Then T has the shadowing property.*

Proof. Let $\mathbf{x} = \{x_n\}_{n\in\mathbb{Z}} \subset X$ be a δ-pseudo-orbit. Define the metric space $(\mathbf{E}, d)$ by

$$\mathbf{E} = \{\mathbf{y} : \mathbf{y} = \{y_n\}_{n\in\mathbb{Z}}, d(x_n, y_n) \leq \epsilon\}$$

with metric

$$D(\mathbf{x}, \mathbf{y}) = \sup\{d(x_n, y_n) : n \in \mathbb{Z}\}$$

That is, $\mathbf{E}$ contains all sequences which are pointwise within ϵ of our δ-pseudo-orbit, $\mathbf{x}$. Define a sequence $\mathbf{T}(\mathbf{y})$ for sequences $\mathbf{y}$ as

$$\mathbf{T}(\mathbf{y})_n = T(y_{n-1})$$

If $\mathbf{E}$ contains an orbit of T, the claim follows. We aim to show the existence of a fixed point of $\mathbf{T}$, since such a point would be an orbit of T. We first prove $\mathbf{E}$ is $\mathbf{T}$-invariant. By the triangle inequality,

$$d(T(y_{n-1}), x_n) \leq d(T(y_{n-1}), T(x_{n-1})) + d(T(x_{n-1}), x_n)$$

Since T is a contraction, and $\{x_n\}$ is a δ-pseudo-orbit,

$$d(T(y_{n-1}), x_n) \leq \tau d(y_{n-1}, x_{n-1}) + \delta$$

We can choose our value of δ for any given ϵ. Note that $d(y_{n-1}, x_{n-1}) \leq \epsilon$ since $\mathbf{y} \in \mathbf{E}$. Letting $\delta = (1-\tau)\epsilon$,

$$\begin{aligned} d(T(y_{n-1}), x_n) &\leq \tau\epsilon + (1-\tau)\epsilon \\ &\leq \epsilon \end{aligned}$$

Hence, for any $\mathbf{y} \in \mathbf{E}$, $\mathbf{T}(\mathbf{y}) \in \mathbf{E}$ and so $\mathbf{E}$ is indeed $\mathbf{T}$-invariant. Finally, we show that $\mathbf{T}$ is a contraction with contraction constant τ. Take $\mathbf{y}, \mathbf{z} \in \mathbf{E}$. Then,

$$\begin{aligned} D(\mathbf{T}(\mathbf{y}), \mathbf{T}(\mathbf{z})) &= \sup\{d(T(y_{n-1}), T(x_{n-1})) : n \in \mathbb{Z}\} \\ &\leq \tau \ \cdot \sup\{d(y_{n-1}, z_{n-1}) : n \in \mathbb{Z}\} \\ &\leq \tau D(\mathbf{y}, \mathbf{z}) \end{aligned}$$

By the Banach Fixed-Point theorem 2.4.5, $\mathbf{T}$ has a fixed point, hence proving the claim. $\square$

Shadowing holds with greater generality. As hyperbolicity was a key property for many of the previous properties, it is for shadowing as well. Note that in the previous lemma, T was indeed hyperbolic as $E^u = 0$.

Lemma 2.8.4. *If Λ is a hyperbolic set for a diffeomorphism ϕ, then there exists a neighborhood W of Λ such that ϕ has the shadowing property on W.*

Bibliography

[1] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discrete & Computational Geometry*, 2(4):319–326, 1987.

[2] Luis A. Rademacher. Approximating the centroid is hard. In *Proceedings of the Twenty-third Annual Symposium on Computational Geometry*, SCG '07, pages 302–305, New York, NY, USA, 2007. ACM.

[3] Dimitris Bertsimas and Santosh Vempala. Solving convex programs by random walks. *J. ACM*, 51(4):540–556, 2004.

[4] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, January 1991.

[5] Ravindran Kannan, László Lovász, and Miklós Simonovits. Random walks and an $o^*(n^5)$ volume algorithm for convex bodies. *Random Structures and Algorithms - RSA*, 11:1–50, 08 1997.

[6] Santosh Vempala. Geometric random walks: a survey. *Combinatorial and computational geometry*, 52(573-612):2, 2005.

[7] N. Chernov, R. Markarian, and American Mathematical Society. *Chaotic Billiards*. Mathematical surveys and monographs. American Mathematical Society, 2006.

[8] N. Chernov. Decay of correlations and dispersing billiards. *Journal of Statistical Physics*, 94(3):513–556, Feb 1999.

[9] Hermann Weyl. Über die gleichverteilung von zahlen mod. eins. *Mathematische Annalen*, 77(3):313–352, Sep 1916.

Rares Cristian 25 January 2019

Research Work Plan

Goal / Questions to Answer	Low Target	Expected Target	High Target
Finish implementing Ball reflection version of sampling algorithm. Write up details of how the algorithm works.	February 8th	February 1st	February 1st
Compare Ball to Facet reflections. How does convergence change as we increase dimension? Number of facets? Analyze difference for each of these:	---------- ---------- ----------	---------- ---------- ----------	---------- ---------- ---------- ---------- ----------
- Distance between average and centroid as we perform more iterations - Convergence of marginal distribution. Is it sufficient to observe the distribution each time we double the number of iterations or is something finer needed?	February 15th	February 15th	February 8th
- Convergence of average signed distance of points to an arbitrary hyperplane containing the centroid.	February 29th	February 22nd - Can we somehow relate this to the marginal distribution?	February 22nd
Find good center for each facet to anchor a large ball - add slight convex curvature to facets. Compare with both ball and facet reflections.	March 22nd	March 8th	March 8th
Create reflection function dependent only on position, not on the incoming direction. Compare as above	---------- ----------	March 22nd	March 15th
How does each sampling method fare for approximating volume? - Which stopping criteria mentioned above is most relevant? - How many iterations are required to achieve an ϵ-approximation depending on dimension?	April 12th	April 5th - Can we say anything about time complexity polynomial in dimension and error?	March 29th
Lines nearly parallel to a facet can drastically increasing running time. How often do these occur? - Analyze average angle with facet. - As we near the facet, what is probability of hitting a ball?	END OF SEM. – First weeks of Fall 2019	END OF SEM. - Prove this indeed occurs rarely.	END OF SEM.
How to deterministically choose initial direction?	Sept. 27th	September 13th	September 6th
Does initial starting position matter significantly?	Oct. 11th	September 27th	September 20th
Given a hyperplane cutting the polytope in a portion of small volume, can we get stuck there for a long time?	Nov. 1st	October 18th	October 4th
Put results together. Additionally prove everything mentioned in the expected and high targets if we were unable to. There may be some we don't know how to prove.	END OF 2019 - We would have less time to prove certain results if we haven't been able to along the way.	END OF 2019	END OF 2019