CS 798 - Convexity and Optimization, Winter 2017, Waterloo

Lecture 23: Random sampling in convex body

We study the method of random sampling in convex bodies by random walks, and discuss one application in approximately computing the volume.

Random sampling by random walks

The problem that we study today is to generate a uniformly random point from a convex set.

For simple objects such as cuboids, balls, ellipsoids and simplices, we know how to generate a random point directly. Try these out and we will use the ball case later.

For polytopes, we do not know how to generate a random point directly.

Suppose we know a bounding ball of the polytope. One simple approach is to generate a random point in the ball and check if it is in the polytope. It should be clear that any point in the polytope is equally likely, but the problem is that we may need exponentially many samples (in the dimension) before sampling a point in the polytope (e.g. a ball of radius $\sqrt{n}$ that contains the cube $[-1,+1]^n$).

From HW3, we know how to find a $O(n^{1.5})$-rounding ellipsoid of any convex body.

These ideas, although not directly applicable, will eventually be incorporated in random sampling and approximating volume.

To sample from a complicated set, a general approach is to do random walks.

We start from an arbitrary point, define a simple neighborhood of the current point, move to a random point in the neighborhood, and repeat for a large enough number of steps and then return the final point.

This approach is very general and works in many settings including combinatorial problems (e.g. random spanning trees).

In the case of sampling random points in convex bodies $K \subseteq \mathbb{R}^n$, the following are some specific algorithms:

- Grid-walk: Fix a small number $\delta > 0$. At the current point $x \in K$, choose a random coordinate $v \in \{\pm e_1, \ldots, \pm e_n\}$ and consider the point $y = x + \delta v$. If $y \in K$, set y to be the current point; otherwise, stay at x.
- Ball-walk: Fix a small number $\delta > 0$. At the current point $x \in K$, choose a random point $v \in B_n$ where B_n is the unit ball. If $y = x + \delta v \in K$, set y to be the current point; otherwise stay at x.
- Hit-and-run: Pick a uniform random line ℓ through the current point x. Go to a uniform random point on the chord $\ell \cap K$.

The first polynomial time algorithm for random sampling and volume computation by Dyer, Frieze and Kannan uses grid-walk, which has many technical issues due to its discrete nature.

Hit-and-run is "fast and fun" (a paper by Lovász and Vempala), but the ideas are quite different and more advanced.

We will focus on the analysis of the ball-walk, with the following assumptions.

Assumption:

- We have a membership oracle for K. Given x, the oracle returns "yes" if $x \in K$ and "no" if $x \notin K$.
- $B_n \subseteq K \subseteq R \cdot B_n$. This is the same as the assumption in the ellipsoid method, that the set K is not super thin in some direction and not super long in some direction.

Stationary distribution and convergence rate

There are two basic questions about the random walks that we need to address.

One is whether the limiting distribution is the uniform distribution, another is the convergence rate to the limiting distribution.

We will quickly review these questions in the discrete setting and then discuss the generalizations to the continuous setting.

Random walks on graphs

In the discrete setting, we are given an undirected graph $G=(V,E)$, and the random walk starts at an arbitrary vertex v_0, and in each step moves to a uniformly random neighbor.

Let $W \in \mathbb{R}^{|V| \times |V|}$ be the random walk matrix with $W_{i,j} = \frac{1}{\deg(i)}$ if $ij \in E$ to denote the transition probability from vertex i to vertex j and zero otherwise. Note that we can have self-loops.

A probability distribution $p \in \mathbb{R}^{|V|}$ is a <u>stationary distribution</u> if $p^T W = p^T$, such that the probability distribution remains the same after one step of random walk.

It is straightforward to check that $p_i = \frac{\deg(i)}{2|E|}$ is a stationary distribution.

In particular, if the graph is regular (every vertex has the same degree), then the uniform distribution $\frac{\vec{1}}{n}$ is a stationary distribution.

If the uniform distribution is our target distribution, we can add self-loops to the graph to make it regular.

Limiting distribution

If we get to a stationary distribution, then it will stay and it is the limiting distribution, but is it unique and must we get there?

The eigenvalues of the random walk matrix hold the key to the answer.

Assume the graph is regular, so $\vec{1}$ is an eigenvector of W with eigenvalue 1.

It is not difficult to check that all eigenvalues of W are between -1 and $+1$.

If $\vec{1}$ is the only eigenvector with eigenvalue of absolute value 1, then a simple spectral analysis will show that the uniform distribution will be the unique limiting distribution regardless of the starting vertex.

What graphs will satisfy this property? Let $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$ be the eigenvalues of W.

If the graph is non-bipartite, then $\lambda_n > -1$.

A standard way to achieve non-bipartiteness is to do a lazy random walk, such that it stays in the same vertex with probability $\frac{1}{2}$ and moves to a random neighbor with the remaining probability.

This modification will slow down the random walks by a factor of 2, but will guarantee that $\lambda_n \geq 0$.

If the graph is connected, then $\lambda_2 < 1$, and so the lazy random walks must converge to the uniform distribution if the graph is regular (perhaps after adding self-loops).

Convergence rate

We measure the convergence rate by how close is the current distribution to the limiting distribution.

Given two probability distribution $p, q \in \mathbb{R}^n$, the total variational distance is defined as $\|p-q\|_{TV} := \sum_{i=1}^{n} |p_i - q_i|$.

For a subset $S \subseteq [n]$, let $p(S) = \sum_{i \in S} p_i$, then an equivalent characterization of $\|p-q\|_{TV} = \frac{1}{2} \max_{S \subseteq [n]} |p(S) - q(S)|$.

Assume the graph is regular, our goal is to bound $\|p_t - \frac{\vec{1}}{n}\|_{TV}$ where p_t is the distribution after t steps.

We say T is the mixing time if $\|p_T - \frac{\vec{1}}{n}\|_{TV} \leq \frac{1}{4}$.

A standard spectral analysis shows that the mixing time of the lazy random walk is $O(\frac{\ln n}{1-\lambda_2})$.

The total variational distance can be decreased to $\|p_t - \frac{\vec{1}}{n}\|_{TV} \leq \varepsilon$ in $O(\frac{\ln n/\varepsilon}{1-\lambda_2})$ steps.

Conductance

What graphs would have $1-\lambda_2$ large so that the mixing time is small?

It turns out that it is closely related to the conductance/expansion of a graph.

Given $S \subseteq V$, the volume of S, denoted by $\mathrm{vol}(S)$, is defined as $\mathrm{vol}(S) := \sum_{i \in S} \deg(i)$.

The conductance of $S \subseteq V$ is defined as $\phi(S) := \frac{|\delta(S)|}{\min\{\mathrm{vol}(S), \mathrm{vol}(V-S)\}}$ where $\delta(S)$ is the set of edges with one endpoint in S and one endpoint in $V-S$.

The conductance of G is defined as $\phi(G) := \min_{S \subseteq V} \phi(S)$.

Informally, $\phi(G)$ is large if every subset has a large fraction of edges going out.

Cheeger's inequality states that $\frac{1-\lambda_2}{2} \leq \phi(G) \leq \sqrt{2(1-\lambda_2)}$.

This implies that the mixing time of the lazy random walk is $O\left(\frac{\ln |V|}{\phi(G)^2}\right)$.

A graph is called an <u>expander graph</u> if $\phi(G)$ is a constant, in which case the mixing time is $O(\ln |V|)$,

which is logarithmic in terms of the graph size.

It can also be shown that $\Omega\left(\frac{1}{\phi(G)}\right)$ is a lower bound on the mixing time, as intuitively a set

of small conductance will imply a slow mixing time.

<u>Summary</u>

If we do lazy random walks on a regular graph, the uniform distribution is the unique limiting distribution.

The convergence rate is qualitatively equivalent to the conductance of the graph, so we need to

establish a lower bound on the conductance to prove that the mixing time is small, which is

required to conclude that the final vertex is a close to uniformly random vertex.

<u>Random walks in continuous space</u>

A Markov chain is defined using a σ-algebra $(K, \mathcal{A})$ where K is the state space and $\mathcal{A}$ is a set of

subsets of K that is closed under complements and countable unions.

For each element u of K, we have a probability measure P_u on $(K, \mathcal{A})$ to describe the distribution

after one-step of random walk from u, which is specified by $P_u(A)$ for each $A \in \mathcal{A}$, the probability being in A.

The triple $(K, \mathcal{A}, \{P_u : u \in K\})$ and a starting distribution Q_0 defines a Markov chain, where w_0 is

chosen from Q_0, and then subsequently w_{i+1} is chosen from P_{w_i} for $i \geq 0$.

Given the current distribution Q_t, Q_{t+1} is defined as $Q_{t+1}(A) = \int_K P_u(A)\, dQ_t(u)$.

A distribution Q is called <u>stationary</u> if for any $A \in \mathcal{A}$, $\int_K P_u(A)\, dQ(u) = Q(A)$.

The <u>ergodic flow</u> of subset A with respect to the distribution Q is defined as $\Phi(A) = \int_A P_u(K \setminus A)\, dQ(u)$,

the probability to leave A after one step of random walk if we start from a random point in A.

The following claim is quite intuitive, as otherwise the distribution is not "stable" yet.

<u>Claim</u> A distribution Q is stationary if and only if $\Phi(A) = \Phi(K \setminus A)\ \forall A \in \mathcal{A}$.

proof $\Phi(A) - \Phi(K\setminus A) = \int_A P_u(K \setminus A)\, dQ(u) - \int_{K\setminus A} P_u(A)\, dQ(u) = \int_A (1 - P_u(A))\, dQ(u) - \int_{K \setminus A} P_u(A)\, dQ(u)$

$$= Q(A) - \int_K P_u(A)\, dQ(u). \quad \square$$

The conductance of a subset A is defined as $\phi(A) := \dfrac{\Phi(A)}{\min\{Q(A), Q(K\setminus A)\}}$, and the conductance of the Markov chain is defined as $\phi(K) := \min_A \phi(A)$, where Q is a stationary distribution.

A stationary distribution Q is atom-free if there is no $x \in K$ with $Q(\{x\}) > 0$.

The following theorem is analogous to the convergence theorem in the discrete setting.

Theorem (Lovász-Simonovits) Suppose the given lazy Markov chain has an atom-free stationary distribution Q.
Let $M = \sup_A Q_0(A)/Q(A)$. Then for every $A \in \mathcal{A}$, $|Q_t(A) - Q(A)| \leq \sqrt{M}\,(1 - \frac{1}{2}\phi^2)^t$.
This implies that the mixing time of the Markov chain is $O(\frac{\log M}{\phi^2})$.

The proof of Lovász-Simonovits directly uses conductance to bound the mixing time, instead of going through eigenvalues.
Under our assumption that $B_n \subseteq K \subseteq RB_n$, we can start from a uniformly random point in B_n, and this implies that $M \leq Q_0(B_n)/Q(B_n) \leq 1 / \frac{1}{R^n} = R^n$, and so the mixing time is $O(\frac{n \log R}{\phi^2})$.
The key to analyze the mixing time is to understand the conductance of the Markov chain.

Ball walk

In the Markov chain for ball walk in a convex body K, the state space is K, $\mathcal{A}$ is the set of all measurable subsets of K, and $P_u(\{u\}) = 1 - \dfrac{\mathrm{vol}(K \cap (u+\delta B_n))}{\mathrm{vol}(\delta B_n)}$,

and $P_u(A) = \dfrac{\mathrm{vol}(A \cap (u+\delta B_n))}{\mathrm{vol}(u+\delta B_n)}$ if $u \notin A \in \mathcal{A}$ and $P_u(A) = P_u(A\setminus\{u\}) + P_u(\{u\})$ if $u \in A \in \mathcal{A}$.

It is not difficult to see that the uniform distribution is a stationary distribution $= Q(A) = \dfrac{\mathrm{vol}(A)}{\mathrm{vol}(K)}$.
One way is to observe that the Markov chain is "time-reversible", or that it is "regular".
One can also do a direct calculation: In the uniform distribution $dQ(u) = \frac{1}{\mathrm{vol}(K)}du$ and the probability into A is

$$\int_{u\in A} P_u(\{u\})\, dQ(u) + \int_{u \in A} \int_{v \in K \cap B_\delta(u) \setminus u} \frac{1}{\mathrm{vol}(\delta B_n)}\, dQ(v)\, du \quad \text{(the second term is from neighbors of } u \text{ to } u)$$

$$= \int_{u\in A} \left(1 - \frac{\mathrm{vol}(K \cap B_\delta(u))}{\mathrm{vol}(\delta B_n)}\right) dQ(u) + \int_{u \in A} \frac{\mathrm{vol}(K \cap B_\delta(u))}{\mathrm{vol}(\delta B_n)}\, dQ(u) = \int_{u\in A} dQ(u) = Q(A).$$

Using the Lovász-Simonovits theorem, we just need to bound the conductance of the ball-walk Markov chain.

Local conductance

Let $\ell(u) := 1 - P_u(\{u\}) = \dfrac{\mathrm{vol}(K \cap (u+\delta B_n))}{\mathrm{vol}(\delta B_n)}$ be the local conductance at $u \in K$.

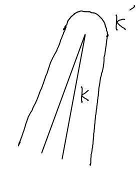
After a moment of thought we realize that the local conductance could be exponentially small

Let $\ell(u) := 1 - P_u(\{u\}) = \dfrac{\mathrm{vol}(K \cap (u+\delta B_n))}{\mathrm{vol}(\delta B_n)}$ be the local conductance at $u \in K$.

After a moment of thought, we realize that the local conductance could be exponentially small,

e.g. a point near the apex of a tall and thin cone.

Since local conductance is a special case of (global) conductance, it shows that the conductance $\phi(K)$ could be exponentially small, and the ball walk in K is not mixing in polynomial (in n) time.

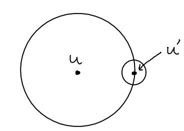


There is a fix to this problem.

We consider $K' := K + \alpha B_n$ for some $\alpha \geq \delta\sqrt{n}$. See the picture.

This improves the local conductance of every point in K'.

<u>Claim</u> $\ell(u') \geq \frac{1}{8}$ for every point $u' \in K'$.



<u>proof</u> For every point $u' \in K'$, there is a point $u \in K$ such that $u' \in u + \alpha B_n \subseteq K'$.

We assume without loss that u is the original and $\alpha = \delta\sqrt{n}$, and $u' = \alpha e_1$.

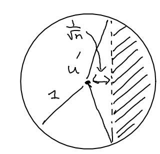
We would like to argue that $\dfrac{\mathrm{vol}(B_\alpha(u) \cap B_\delta(u'))}{\mathrm{vol}(B_\delta(u'))} \geq \frac{1}{8}$ and this clearly implies the claim since $B_\alpha(u) \subseteq K'$.

We claim that all the points in $B_\delta(u')$ with $x_1 \leq \alpha - \frac{\delta}{\sqrt{n}}$ belongs to $B_\alpha(u)$.

To see this, all points in $B_\delta(u')$ satisfy $(x_1 - \alpha)^2 + \sum_{i=2}^n x_i^2 \leq \delta^2$. So, if $x_1 \leq \alpha - \frac{\delta}{\sqrt{n}}$, then $\sum_{i=1}^n x_i^2 \leq \delta^2 - \frac{\delta^2}{n}$

So, every such point will have $\sum_{i=1}^n x_i^2 \leq (\alpha - \frac{\delta}{\sqrt{n}})^2 + \sum_{i=2}^n x_i^2 \leq \alpha^2 - \frac{2\alpha\delta}{\sqrt{n}} + \frac{1}{n} + \delta^2 - \frac{1}{n} \leq \alpha^2$ using $\alpha = \delta\sqrt{n}$.

Think of B_δ as a unit ball, then the above calculation shows that all the slides beyond $x_1 \geq \frac{1}{\sqrt{n}}$ belong to K'.



Recall from L18 that the slice area distributes like a Gaussian random variable with standard derivation $\frac{1}{\sqrt{n}}$.

So, there are at least $\approx 0.16 \geq \frac{1}{8}$ fraction of the δ-ball in K'. $\square$

This fix guarantees that the local conductance of every point in K' is at least $\frac{1}{8}$.

So, the natural idea is to do the ball walk in K' instead, but we need to address two issues.

① To do the ball walk in K', we need a membership oracle for K'.

We just have a membership oracle for K, and we need to determine whether a point is of distance $\leq \alpha$ to K.

If we had a separation oracle for K, then we can determine this using the ellipsoid method (as done in L22).

An important and deep result by Yudin and Nemirovski (see [GLS 4.3]) shows that one can also optimize with only a membership oracle using the ellipsoid method.

So, we can use Yudin-Nemirovski's algorithm to build a membership oracle for K' using that of K and the ellipsoid method. So, this issue can be fixed, with some deep result and some time complexity overhead.

② When we stop the random walk in K', we may stop at a point in $K' \setminus K$ and this is not useful for us. The failure probability that this happens is $1 - \frac{\text{vol}(K)}{\text{vol}(K+\alpha B_n)}$, assuming we have a uniform sample of K'.

Since we assume $B \subseteq K$, we have that $K + \alpha B_n \subseteq (1+\alpha) K$, and so the failure probability is at most

$$1 - \frac{\text{vol}(K)}{\text{vol}(K+\alpha B_n)} \leq 1 - \frac{\text{vol}(K)}{\text{vol}((1+\alpha)K)} = 1 - \frac{1}{(1+\alpha)^n}.$$

So, if we choose $\alpha \leq \frac{1}{n}$, then the failure probability is at most $1 - \frac{1}{e}$, as $(1+\frac{1}{n})^n \leq e$.

Since we need to choose $\alpha \geq \delta\sqrt{n}$ for the local conductance, we are forced to choose $\delta \leq n^{-\frac{3}{2}}$.

This restricts us to do the ball walk with very small steps, and thus increases the running time.

Anyway, if we just care about polynomial time solutions, the above fix and the corresponding modifications would work as it only creates a polynomial factor blowup.

There is a more elegant and efficient fix by [Kannan-Lovász-Simonovits] using the concept of average conductance.

Bounding conductance

By working on K', we know that the local conductance is good, but does it imply that the conductance is good?

This is a nontrivial question, but the answer turns out to be yes.

The heart of the proof is an isoperimetric inequality for any convex body.

Theorem Let S_1, S_2, S_3 be a partition into measurable sets of a convex body K of diameter D.

Then, $\text{vol}(S_3) \geq \frac{d(S_1, S_2)}{D} \min\{\text{vol}(S_1), \text{vol}(S_2)\}$, where $d(S_1, S_2) = \min\{\|x-y\| \mid x \in S_1, y \in S_2\}$.

This is a nontrivial theorem. We will study its proof and discuss some strengthenings next time.

We will assume this theorem and derive a lower bound on the conductance of ball walk in K.

We also need one more technical lemma for the conductance lower bound.

Lemma Let u, v be such that $\|u-v\|_2 \leq \frac{\delta}{\sqrt{n}}$ and $\ell(u), \ell(v) \geq \ell$.

Then $\|P_u - P_v\|_{TV} \leq 1 - \frac{\ell}{e+1} < 1 - \frac{\ell}{4}$

This lemma says that if u, v are close enough, then there is a large overlap in their one-step distribution.

It is easy to prove if both $B_\delta(u) \subseteq K'$ and $B_\delta(v) \subseteq K'$, as this implies that a large cap of each ball is in the intersection as in the proof of the local conductance claim.

We also need to handle the case when not both balls are contained in K; we will do the proof later if time permits.

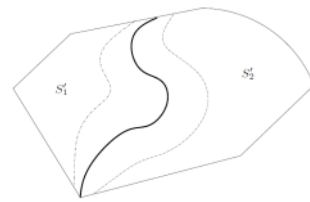
Let's prove a lower bound on the conductance of the ball walk assuming these results.

<u>Theorem</u> Let K be a convex body of diameter D so that for every point u in K, the local conductance of the ball walk with step size δ is at least l. Then $\phi \geq \frac{l\delta}{32\sqrt{n}D}$.

<u>Proof</u> Let $K = S_1 \cup S_2$ be a partition of K into two measurable sets.

Our goal is to prove that $\int_{S_1} P_x(S_2)\,dx \geq \frac{l\delta}{32\sqrt{n}D} \min\{vol(S_1), vol(S_2)\}$.

Consider the points that are deep inside these sets.



Let $S_1' := \{x \in S_1 \mid P_x(S_2) < \frac{l}{8}\}$ and $S_2' := \{x \in S_2 \mid P_x(S_1) < \frac{l}{8}\}$.

Let $S_3' := K \setminus S_1' \setminus S_2'$.

If $vol(S_1') < vol(S_1)/2$, then $\int_{S_1} P_x(S_2)\,dx \geq \frac{l}{8} vol(S_1 \setminus S_1') \geq \frac{l}{16} vol(S_1)$, and we are done.

So, we assume $vol(S_1') \geq vol(S_1)/2$ and $vol(S_2') \geq vol(S_2)/2$.

For any $u \in S_1'$ and $v \in S_2'$, note that $\|P_u - P_v\|_{TV} \geq 1 - P_u(S_2) - P_v(S_1)$ (if both don't cross, then no overlap)

$$> 1 - \frac{l}{4}.$$

By the lemma, we get that $\|u - v\| \geq \frac{\delta}{\sqrt{n}}$, and thus $d(S_1', S_2') \geq \frac{\delta}{\sqrt{n}}$.

By the theorem, $vol(S_3') \geq \frac{\delta}{\sqrt{n}D} \min\{vol(S_1'), vol(S_2')\} \geq \frac{\delta}{2\sqrt{n}D} \min\{vol(S_1), vol(S_2)\}$.

Therefore, $\int_{S_1} P_x(S_2)\,dx = \frac{1}{2}\int_{S_1} P_x(S_2)\,dx + \frac{1}{2}\int_{S_2} P_x(S_1)\,dx$ as the uniform distribution is stationary

$$\geq \frac{1}{2} vol(S_3') \frac{l}{8} \geq \frac{l\delta}{32\sqrt{n}D} \min\{vol(S_1), vol(S_2)\}. \quad \square$$

<u>Time complexity</u>

Recall that we choose $l = \frac{1}{8}$ and $\delta = n^{-\frac{3}{2}}$, so $\phi = \Omega(\frac{1}{n^2 D})$.

By the Lovász-Simonovits theorem, the mixing time is $O(\frac{\ln M}{\phi^2}) = O(\frac{n \ln R}{\phi^2}) = O(n^5 D \ln R)$

By HW3 Q1, we can find in polynomial time an ellipsoid $\mathcal{E} \subseteq K \subseteq n^{\frac{3}{2}}\mathcal{E}$.

We can do a linear transformation A so that $B \subseteq A(K) \subseteq n^{\frac{3}{2}} B$.

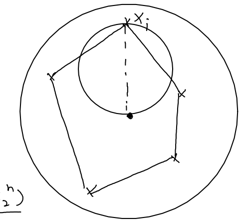
Therefore, we can assume that $D \leq 2n^{\frac{3}{2}}$ and $R \leq n^{\frac{3}{2}}$, and thus the mixing time is $\tilde{O}(n^{6.5})$.

Don't forget the membership oracle for K', which requires the Yudin-Nemirovsky algorithm, which blows up the total time complexity by another polynomial factor.

<u>Approximating volume</u>

A primary motivation for random sampling in convex bodies is to estimate the volume of a convex body. This is a task that cannot be solved by deterministic algorithms.

<u>Theorem</u> If the deterministic algorithm only calls the membership oracle m times, then there are two convex bodies K_1 and K_2 that are consistent with the answers but $vol(K_1)/vol(K_2) \geq 2^n/m$.

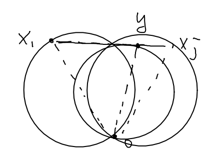


<u>proof</u> An adversary can answer "yes" if the query point $x_i \in B_2^n$.

Suppose $x_1, \dots, x_m \in B_2^n$. Let $K_1 = B_2^n$ and $K_2 = conv\{x_1, \dots, x_m\}$.

For each $x_i \in \mathbb{R}^n$, consider the ball B_i with center $\frac{x_i}{2}$ and radius $\|x_i\|/2$.

We will argue that $conv\{x_1, \dots, x_m\} \subseteq \bigcup_{i=1}^{m} B_i$, and thus $vol(conv\{x_1,\dots,x_m\}) \leq m\,vol(B_i) \leq \frac{m\,vol(B_2^n)}{2^n}$, and this would imply the theorem.



We claim that if any point y in the line segment of x_i and x_j, the ball $B_y \subseteq B_i \cup B_j$. By induction, any point in the convex hull is in $\bigcup_{i=1}^{m} B_i$.

We leave it to the reader to prove the claim. □

In sharp contrast, one can use the random sampling algorithm to approximate the volume of K up to arbitrarily high accuracy. The difference is that the adversary doesn't know the query points beforehand.

<u>Theorem</u> There is a polynomial time randomized algorithm to estimate $vol(K)$ up to $1 \pm \beta$ multiplicative factor.

<u>proof</u> We assume that we can generate uniform random samples of K.

The idea is quite standard in approximate counting, using a multiphase Monte-Carlo algorithm.

Our assumption is that $B_2^n \subseteq K \subseteq R B_2^n$.

Consider $K_i = K \cap (2^{i/n} B_2^n)$ for $1 \leq i \leq n \log R$. So there are $m := n \log R$ phases.

Note that $\frac{vol(B_2^n)}{vol(K)} = \prod_{i=1}^{m} \frac{vol(K_{i-1})}{vol(K_i)}$, and also $vol(K_i) \leq 2\,vol(K_{i-1})$.

We estimate each term $vol(K_{i-1})/vol(K_i)$ accurately, and then multiply them to get an approximation of $vol(K)$.

To estimate $vol(K_{i-1})/vol(K_i)$, we sample random points from K_i and use the fraction of points falling in K_{i-1} as an estimate of $vol(K_{i-1})/vol(K_i)$.

Since $vol(K_i) \leq 2\,vol(K_{i-1})$, by Chernoff's bound, we only need $O(\frac{\ln m}{\varepsilon^2})$ samples to get an $1 \pm \varepsilon$ approximation w.p. $\geq 1 - \frac{1}{poly(m)}$.

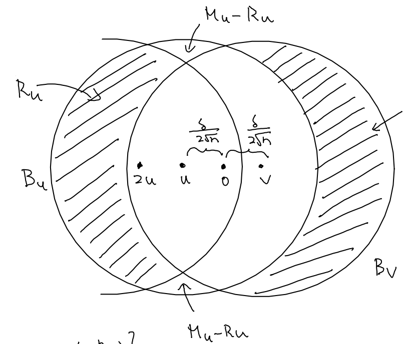
To get an $1 \pm \beta$ approximation of $vol(B_2^n)/vol(K)$, we set $\varepsilon = \frac{\beta}{m}$ so that $\tilde{O}(\frac{m^2}{\beta^2})$ samples are needed in each phase, and the total number of samples is $\tilde{O}(\frac{m^3}{\beta^2}) = \tilde{O}(\frac{n^3 \log^3 R}{\beta^2})$.

The approximation ratio is $(1 \pm \frac{\beta}{m})^m \approx 1 \pm \beta$. □

One-step lemma

We would like to prove that if $\|u-v\|_2 \le \frac{\delta}{\sqrt{n}}$, then $\mathrm{vol}(K \cap (B_\delta(u) \cap B_\delta(v))) \ge \frac{1}{e+1} \min\{\mathrm{vol}(K \cap B_\delta(u)), \mathrm{vol}(K \cap B_\delta(v))\}$.

This will imply that $\|P_u - P_v\|_{TV} = 1 - \frac{\mathrm{vol}(K \cap (B_\delta(u) \cap B_\delta(v)))}{\mathrm{vol}(B_\delta(u))} \ge 1 - \frac{1}{e+1} \frac{\mathrm{vol}(K \cap B_\delta(u))}{\mathrm{vol}(B_\delta(u))} = 1 - \frac{\ell(u)}{e+1}$, proving the lemma.



We assume that $u = -\frac{\delta}{2\sqrt{n}} e_1$ and $v = \frac{\delta}{2\sqrt{n}} e_1$.

Let $C = (u + \delta B_2^n) \cap (v + \delta B_2^n)$ be the intersection of the two δ-balls.

Consider the "moons" $M_u = (u + \delta B_2^n) \setminus (v + \delta B_2^n)$ and $M_v = (v + \delta B_2^n) \setminus (u + \delta B_2^n)$.

Also let $R_u = M_u \cap (2u + C)$ and $R_v = M_v \cap (2v + C)$.

We would like to prove $\mathrm{vol}(K \cap C) \ge \frac{1}{e+1} \min\{\mathrm{vol}(K \cap (u + \delta B_2^n)), \mathrm{vol}(K \cap (v + \delta B_2^n))\}$.

We do this by comparing $\mathrm{vol}(K \cap C)$ to $\mathrm{vol}(K \cap (M_u - R_u))$ and $\mathrm{vol}(K \cap R_u)$.

We will show: ① $\mathrm{vol}(K \cap C) \ge \frac{1}{e-1} \mathrm{vol}(K \cap (M_u - R_u))$ and

② $\mathrm{vol}(K \cap C) \ge \min\{\mathrm{vol}(K \cap R_u), \mathrm{vol}(K \cap R_v)\}$.

Note that
$$\mathrm{vol}(K \cap R_u) = \mathrm{vol}(K \cap M_u) - \mathrm{vol}(K \cap (M_u \setminus R_u)) = \mathrm{vol}(K \cap (u + \delta B_2^n)) - \mathrm{vol}(K \cap C) - \mathrm{vol}(K \cap (M_u \setminus R_u))$$
$$\ge \mathrm{vol}(K \cap (u + \delta B_2^n)) - \mathrm{vol}(K \cap C) - (e-1)\,\mathrm{vol}(K \cap C) \quad \text{by ①.}$$

Therefore, by ②, $\mathrm{vol}(K \cap C) \ge \min\{\mathrm{vol}(K \cap R_u), \mathrm{vol}(K \cap R_v)\}$
$$\ge \min\{\mathrm{vol}(K \cap (u + \delta B_2^n)), \mathrm{vol}(K \cap (v + \delta B_2^n))\} - e\,\mathrm{vol}(K \cap C).$$

This proves the lemma.

It remains to prove ① and ②.

We first consider ② and show that it follows from the Brunn-Minkowski inequality.

Note that $R_u = M_u \cap (2u + C) \subseteq K \cap (2u + C)$ and similarly $R_v \subseteq K \cap (2v + C)$.

Observe that $\frac{1}{2}(K \cap (2u + C)) + \frac{1}{2}(K \cap (2v + C)) \subseteq K \cap C$ as $u = -v$.

Therefore, $\mathrm{vol}(K \cap C) \ge \mathrm{vol}\left(\frac{1}{2}(K \cap (2u + C)) + \frac{1}{2}(K \cap (2v + C))\right) \ge \mathrm{vol}(K \cap (2u + C))^{\frac{1}{2}} \mathrm{vol}(K \cap (2v + C))^{\frac{1}{2}}$ by Brunn-Minkowski
$$\ge \mathrm{vol}(K \cap R_u)^{\frac{1}{2}} \mathrm{vol}(K \cap R_v)^{\frac{1}{2}},$$ and ② follows.

To prove ①, we will show that if we blow up C by a factor of $1 + \frac{1}{4n}$, then it will contain $M_u \setminus R_u$.

This will imply that $\mathrm{vol}(K \cap (M_u \setminus R_u)) \le (1 + \frac{1}{4n})^n \mathrm{vol}(K \cap C) \le e\,\mathrm{vol}(K \cap C)$, proving ①.

Let $p \in M_u \setminus R_u$ and write $p = cu + w$ where $w \perp u$.

Since $p \in u + \delta B_2^n$, we have $\delta^2 \geq \|p-u\|_2^2 = \|(c-1)u + w\|_2^2 = (c-1)^2\|u\|_2^2 + \|w\|_2^2$.

Since $p \notin v + \delta B_2^n$, we have $\delta^2 < \|p-v\|_2^2 = \|p+u\|_2^2 = \|(c+1)u + w\|_2^2 = (c+1)^2\|u\|_2^2 + \|w\|_2^2$.

Since $p \notin R_u$, we have $\delta^2 < \|p-2u\|_2^2 = \|(c-2)u + w\|_2^2 = (c-2)^2\|u\|_2^2 + \|w\|_2^2$

These imply that $(c-2)^2 > (c-1)^2$ and $(c+1)^2 > (c-1)^2$, and thus $\frac{1}{2} < c < \frac{3}{2}$.

We want to show that $\|\alpha p - v\|_2^2 \leq \delta^2$ for some α close to 1, and this would imply blowing up C by $\frac{1}{\alpha}$ would do.

Note that $\|\alpha p - v\|_2^2 = \|\alpha p + u\|_2^2 = \|(c\alpha+1)u + \alpha w\|_2^2 = (c\alpha+1)^2\|u\|_2^2 + \alpha^2\|w\|_2^2$

$$
\begin{aligned}
&\leq (c\alpha+1)^2\|u\|_2^2 + \alpha^2\delta^2 - \alpha^2(c-1)^2\|u\|_2^2 && \text{using } p \in u + \delta B_2^n. \\
&= \big((1+\alpha)(2\alpha c + 1 - \alpha)\big)\|u\|_2^2 + \alpha^2\delta^2 \\
&\leq \big((1+\alpha)(2\alpha - 1)\big)\|u\|_2^2 + \alpha^2\delta^2 && \text{using } c < \tfrac{3}{2} \\
&\leq \big((1+\alpha)(2\alpha-1)\big)\frac{\delta^2}{4n} + \alpha^2\delta^2 && \text{using } \|u\|_2 \leq \tfrac{\delta}{2\sqrt{n}} \\
&= \delta^2\left(\frac{(4n+2)\alpha^2 + \alpha - 1}{4n}\right) \\
&\leq \delta^2 && \text{if we choose } \alpha^2 = 1 - \tfrac{2}{4n+2}
\end{aligned}
$$

Then $\alpha \approx 1 - \frac{1}{4n+2}$ and thus we just need to blow up C by a factor of $\frac{1}{\alpha} \approx 1 + \frac{1}{4n+1}$.

References: [Vempala, chapter 4-6]

- Vempala, Geometric random walks: a survey
- Kannan, Lovász, Simonovits. Random walks and an $\tilde{O}(n^5)$ volume algorithm for convex bodies.
- Lovász, Simonovits, Random walks in a convex body and an improved volume algorithm.