High-Dimensional Sampling Algorithms

Santosh Vempala
Algorithms and Randomness Center
Georgia Tech

Format

- Please ask questions
- Indicate that I should go faster or slower
- Feel free to ask for more examples
- And for more proofs

- Exercises along the way.

High-dimensional problems

Input:

- A set of points S in n-dimensional space R^n or a distribution in R^n
- A function f that maps points to real values (could be the indicator of a set)

Algorithmic Geometry

- What is the complexity of computational problems as the dimension grows?

- Dimension = number of variables

- Typically, size of input is a function of the dimension.

Problem 1: Optimization

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε .

Output: point y such that

$$f(y) \geq \max f - \epsilon$$

Problem 2: Integration

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε .

Output: number A such that:

$$(1-\epsilon)\int f \leq A \leq (1+\epsilon)\int f$$

Problem 3: Sampling

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε.

Output: A point y from a distribution within distance ε of distribution with density proportional to f.

Problem 4: Rounding

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε.

Output: An affine transformation that approximately "sandwiches" f between concentric balls.

Problem 5: Learning

Input: i.i.d. points with labels from an unknown distribution, error parameter ε.

Output: A rule to correctly label 1- ε of the input distribution.

(generalizes integration)

Sampling

- Generate a uniform random point from a set S or with density proportional to function f.
- Numerous applications in diverse areas: statistics, networking, biology, computer vision, privacy, operations research etc.
- This course: mathematical and algorithmic foundations of sampling and its applications.

Course Outline

- Lecture 1. Introduction to Sampling, high-dimensional Geometry and Complexity.
- L2. Algorithms based on Sampling.
- L3. Sampling Algorithms.

Lecture 1: Introduction

- Computational problems in high dimension
- The challenges of high dimensionality
- Convex bodies, Logconcave functions
- Brunn-Minkowski and its variants
- Isotropy
- Summary of applications

Lecture 2: Algorithmic Applications

- Convex Optimization
- Rounding
- Volume Computation
- Integration

Lecture 3: Sampling Algorithms

- Sampling by random walks
- Conductance
- Grid walk, Ball walk, Hit-and-run
- Isoperimetric inequalities
- Rapid mixing

High-dimensional problems are hard

P1. Optimization. Find minimum of f over a set.

P2. Integration. Find the average (or integral) of f.

- These problems are intractable (hard) in general, i.e., for arbitrary sets and general functions
- Intractable for arbitrary sets and linear functions
- Intractable for polytopes and quadratic functions

P1 is NP-hard or worse

- min number of unsatisfied clauses in a 3-SAT formula

P2 is #P-hard or worse

- Count number of satisfying solutions to a 3-SAT formula

High-dimensional Algorithms

P1. Optimization. Find minimum of f over the set S.

Ellipsoid algorithm [Yudin-Nemirovski; Shor;Khachiyan;GLS]

S is a convex set and f is a convex function.

P2. Integration. Find the integral of f.

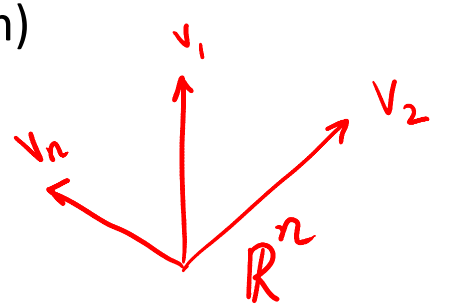
Dyer-Frieze-Kannan algorithm

f is the indicator function of a convex set.

A glimpse of the complexity frontier

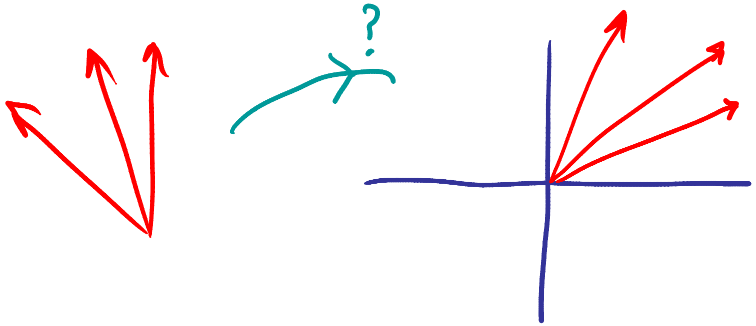
1. Are the entries of a given matrix inner products of a set of vectors?

   $A = BB^T$?   (semidefinite program)



$A_{ij} = v_i \cdot v_j$

2. Are they inner products of a set of nonnegative vectors?



   Is $A = BB^T$, $B \geq 0$?   (completely positive)

Structure

Q. What geometric structure makes algorithmic problems computationally tractable?

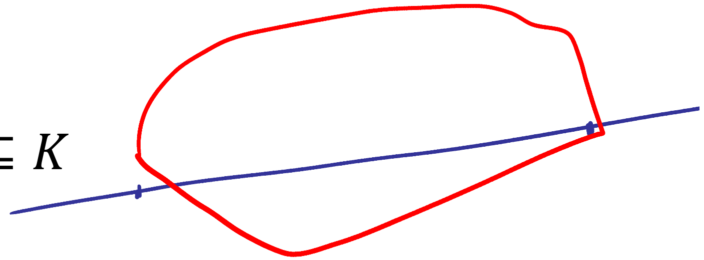
(i.e., solvable with polynomial complexity)

- "Convexity often suffices."
- Is convexity the frontier of polynomial-time solvability?
- Appears to be in many cases of interest

Convexity

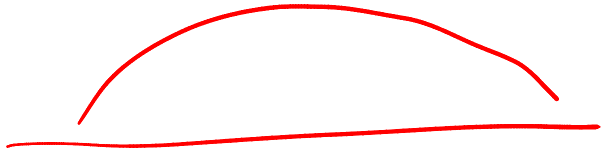
(Indicator functions of) Convex sets:

$\forall x, y \in R^n, \lambda \in [0,1], x, y \in K \Rightarrow \lambda x + (1-\lambda)y \subseteq K$



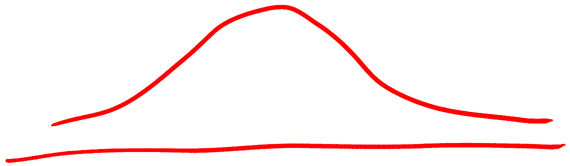
Concave functions:

$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$



Logconcave functions:

$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda}\, f(y)^{1-\lambda}$

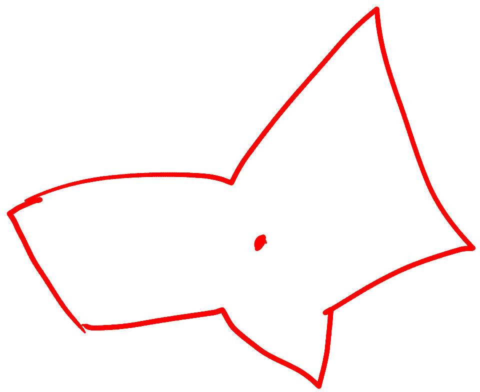


Quasiconcave functions:

$f(\lambda x + (1-\lambda)y) \geq \min\ \ f(x), f(y)$

Star-shaped sets:

$\exists x \in S\ s.t. \forall y \in S, \lambda x + (1-\lambda)y \in S$



How to specify a convex set?

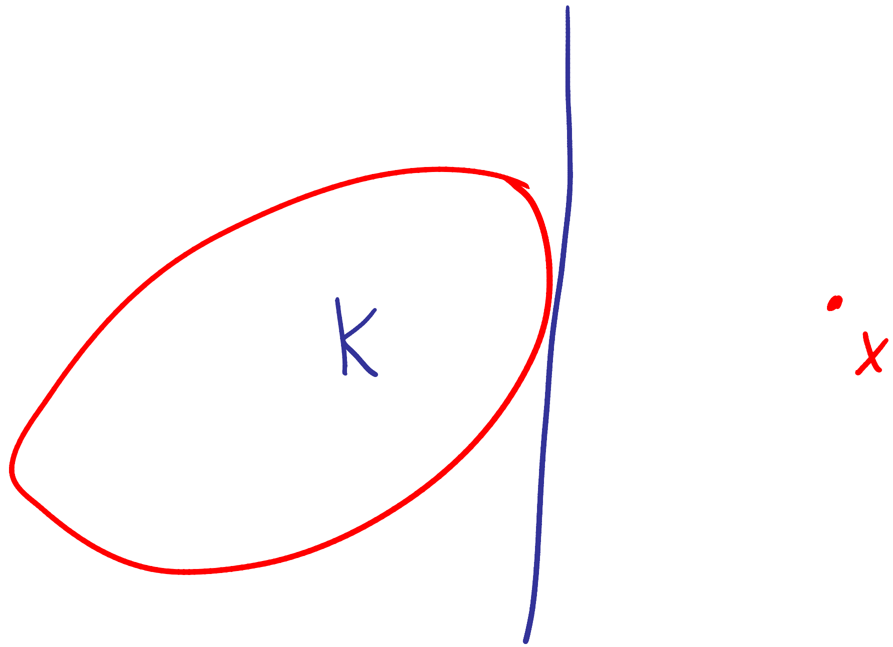
- Explicit list of constraints, e.g., a linear program:

  $Ax \leq \mathrm{b}$

- What about the set of positive semidefinite matrices?

- Or the set of vectors on the edges of a graph that have weight at least one on every cut?

Structure I: Separation Oracle



Either x $\in$ K or there is a halfspace containing K and not x.

Convex sets have separation oracles

- If x is not in K, let y be the point in K that is closest to x.

- y is unique: If y_1, y_2 are both closest, then $(y_1 + y_2)/2$ is closer.

- Take the hyperplane normal to (x-y):

$$\{z : (x - y)^T z \leq (x - y)^T y\}$$

Separation oracles

- For an LP, simply check all the linear constraints
- For a ball or ellipsoid, find the tangent plane
- For the SDP cone, check if the eigenvalues are all nonnegative; if not eigenvector gives a separating hyperplane.
- For cut example, find mincut to check if all cuts are at least 1.

Example: Learning by Sampling

Sequence of points X_1, X_2, …,
Unknown -1/1 function f

We get X_i and have to guess $f(X_i)$

Goal: minimize number of wrong guesses.

Learning Halfspaces

Unknown -1/1 function f

f(X) = 1 if $w^T X > 0$ and f(X) = -1 otherwise

For an unknown vector w, with each component w_i being a b-bit integer.

What is the minimum number of mistakes?

Majority algorithm

After $X_1, X_2, \ldots, X_k$
the set of consistent functions f correspond to
$S_k = \{w : \ w^T(\text{sign}(X_i)X_i) > 0 \ \text{ for } i = 1,2,\ldots, k \}$

Guess $f(X_{k+1})$, to be the majority of the predictions of each w in S_k

Claim. Number of wrong guesses $\leq bn$

But how to compute majority?? $|S_k|$ could be 2^{bn} !

Random algorithm

- Pick random w in S_k
- Guess $w^T X$

Random algorithm

- Pick random w in S_k
- Guess $w^T X_{k+1}$

Lemma 1. E(#wrong guesses) ≤ 2bn.

Proof idea. Every time random guess is wrong, majority algorithm has probability at least ½ of being wrong.

Exercise 1. Prove Lemma 1.

Learning by Sampling

- How to pick random w in $S_{k?}$
- S_k is a convex set!

- It can be efficiently sampled.

Structure of Convex Bodies

- Volume(unit cube) = 1
- Volume(unit ball) $\sim \left(\frac{c}{n}\right)^{\frac{n}{2}}$ u
  - drops exponentially with n
- For any central hyperplane, most of the mass of a ball is within distance $1/\sqrt{n}$.

Structure of Convex Bodies

- Volume(unit cube) = 1
- Volume(unit ball) $\sim \left(\frac{c}{n}\right)^{\frac{\mathrm{n}}{2}}$
  - drops exponentially with n
- Most of the volume is near the boundary:

$$\mathrm{vol}((1-\varepsilon)K) = (1-\varepsilon)^n \mathrm{vol}(K)$$

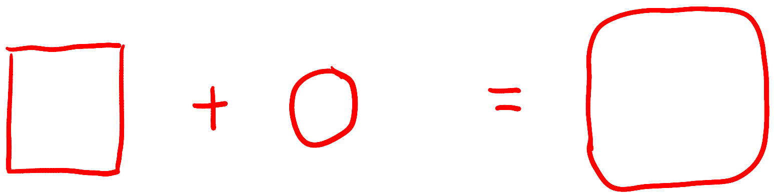
So,

$$\mathrm{vol}(K) - \mathrm{vol}\big((1-\varepsilon)K\big) \geq (1- e^{-\varepsilon n})\mathrm{vol}(K)$$

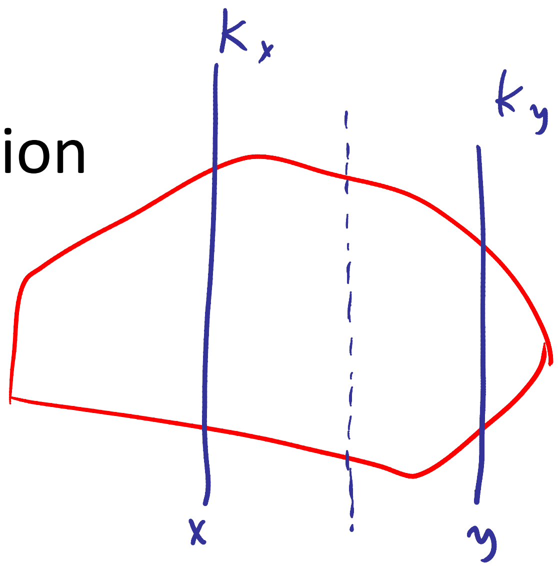
Structure II: Volume Distribution

A,B sets in R^n, their Minkowski sum is:

$$A + B = \{x + y : x \in A, y \in B\}$$



For a convex body, the hyperplane section at (x+y)/2 contains $(A_x + A_y)/2$.



What is the volume distribution?

Brunn-Minkowski inequality

A, B compact sets in R^n

Thm. $\forall\lambda \in [0,1]$,

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

Suffices to prove

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

by taking the sets to be $\lambda A, (1-\lambda)B$

Brunn-Minkowski inequality

Thm. A, B: compact sets in R^n

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

Proof. First take A, B to be cuboids, i.e.,

A = $[0, a_1] \times [0, a_2] \times \ldots \times [0, a_n]$

B = $[0, b_1] \times [0, b_2] \times \ldots \times [0, b_n]$

Then

A+B = $[0, a_1 + b_1] \times [0, a_2 + b_2] \times \ldots \times [0, a_n + b_n]$.

Brunn-Minkowski inequality

Thm. A, B: compact sets in R^n

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

Proof. Next take A,B to be finite unions of disjoint cuboids: A = $\bigcup_i A_i$ and B = $\bigcup_i B_i$

Finally, note that any compact set can be approximated to arbitrary accuracy by the union of a finite set of cuboids.

Logconcave functions

- $f: R^n \to R$ is concave if for any $x, y \in R^n$,

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y)$$

- $f: R^n \to R_+$ is logconcave if for any $x, y \in R^n$,

$$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda}$$

i.e., f is nonnegative and its logarithm is concave.

Logconcave functions

- $f: R^n \to R_+$ is logconcave if for any $x, y \in R^n$,
$$f(\lambda x + (1-\lambda)y) \geq f(x)^{\lambda} f(y)^{1-\lambda}$$

- Examples:
  - Indicator functions of convex sets are logconcave
  - Gaussian density function,
  - exponential function

- Level sets of f, $L_t = \{x : f(x) \geq t\}$, are convex.
- Many other useful geometric properties

Prekopa-Leindler inequality

Prekopa-Leindler: $f, g, h: R^n \to R_+ \ s.t.$

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}$$

then

$$\int h \geq \left(\int f\right)^\lambda \left(\int g\right)^{1-\lambda}.$$

Functional version of [B-M], *equivalent to it*.

Properties of logconcave functions

For two logconcave functions f and g

- Their sum might not be logconcave

- But their product $h(x) = f(x)g(x)$ is logconcave
- And so is their minimum $h(x) = \min f(x), g(x)$.

Properties of logconcave functions

- Convolution is logconcave

$$h(x) = \int_{R^n} f(y)g(x-y)dy$$

- And so is any marginal:

$$h(x_1, x_2, \ldots, x_k) = \int_{R^{n-k}} f(x)dx_{k+1}dx_{k+2} \ldots dx_n$$

Exercise 2. Prove the above properties using the Prekopa-Leindler inequality.

Isotropic position

- Affine transformations preserve convexity and logconcavity.
- What can one use as a canonical position?
- E.g., ellipsoids map to a ball, parallelopipeds map to cubes.
- What about general convex bodies? Logconcave functions?

Isotropic position

- Let x be a random point from a convex body K
- z = E(x) is the center of gravity (or centroid). Shift so that z = 0.
- Now consider the covariance matrix

$$A = E(xx^T), \qquad A_{ij} = E(x_i x_j)$$

- A has bounded entries; it is positive semidefinite; it is full rank unless K lies in a lower-dimensional subspace.

Isotropic position

- $A = E(xx^T)$
- $A = B^2$ for some n x n matrix B.

- Let $K' = B^{-1}K = \{B^{-1}x : x \in K\}$.

- Then a random point y from K' satisfies:

$$E(y) = 0, \qquad E(yy^T) = I_n.$$

- K' is in isotropic position.

Isotropic position: Exercises

- Exercise 3. Find R s.t. the origin-centered cube of side length 2R is isotropic.

- Exercise 4. Show that for a random point x from a set in isotropic position, for any unit vector v, we have

$$E\left((v^T x)^2\right) = 1.$$

Isotropic position and sandwiching

- For any convex body K (in fact any set/distribution with bounded second moments), we can apply an affine transformation so that for a random point x from K :

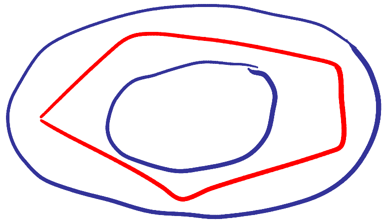
$$E(x) = 0, \qquad E(xx^T) = I_n.$$

- Thus K "looks like a ball" up to second moments.
- How close is it really to a ball? Can it be sandwiched between two balls of similar radii?
- Yes!

Sandwiching

Thm (John). Any convex body K has an ellipsoid E s.t.

$$E \subseteq K \subseteq nE.$$



The minimum volume ellipsoid contained in K can be used.

Thm (KLS). For a convex body K in isotropic position,

$$\sqrt{\frac{n+1}{n}}\, B \subseteq K \subseteq \sqrt{n(n+1)}\, B$$

- Also a factor n sandwiching, but with a different ellipsoid.
- As we will see, isotropic sandwiching (rounding) is algorithmically efficient while the classical approach is not.

Lecture 2: Algorithmic Applications

- Convex Optimization
- Rounding
- Volume Computation
- Integration

Lecture 3: Sampling Algorithms

- Sampling by random walks
- Conductance
- Grid walk, Ball walk, Hit-and-run
- Isoperimetric inequalities
- Rapid mixing

High-Dimensional Sampling Algorithms

Santosh Vempala
Algorithms and Randomness Center
Georgia Tech

Format

- Please ask questions
- Indicate that I should go faster or slower
- Feel free to ask for more examples
- And for more proofs

- Exercises along the way.

High-dimensional problems

Input:

- A set of points S in R^n or a distribution in R^n
- A function f that maps points to real values (could be the indicator of a set)

Algorithmic Geometry

- What is the complexity of computational problems as the dimension grows?

- Dimension = number of variables

- Typically, size of input is a function of the dimension.

Problem 1: Optimization

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε .

Output: point y such that

$$f(y) \geq \max f - \epsilon$$

Problem 2: Integration

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε .

Output: number A such that:

$$(1-\epsilon)\int f \leq A \leq (1+\epsilon)\int f$$

Problem 3: Sampling

Input: function f: $R^n \to R$ specified by an oracle,

point x, error parameter ε.

Output: A point y from a distribution within distance ε of distribution with density proportional to f.

Problem 4: Rounding

Input: function f: $R^n \rightarrow R$ specified by an oracle,

point x, error parameter ε.

Output: An affine transformation that approximately "sandwiches" f between concentric balls.

Problem 5: Learning

Input: i.i.d. points (with labels) from unknown distribution, error parameter ε.

Output: A rule to correctly label 1- ε of the input distribution.

(generalizes integration)

Sampling

- Generate a uniform random point from a set S or with density proportional to function f.

- Numerous applications in diverse areas: statistics, networking, biology, computer vision, privacy, operations research etc.

- This course: mathematical and algorithmic foundations of sampling and its applications.

Lecture 2: Algorithmic Applications

Given a blackbox for sampling, we will study algorithms for:

- Rounding
- Convex Optimization
- Volume Computation
- Integration

High-dimensional Algorithms

P1. Optimization. Find minimum of f over the set S.

Ellipsoid algorithm [Yudin-Nemirovski; Shor] works when
S is a convex set and f is a convex function.

P2. Integration. Find the integral of f.

Dyer-Frieze-Kannan algorithm works when
f is the indicator function of a convex set.

Structure

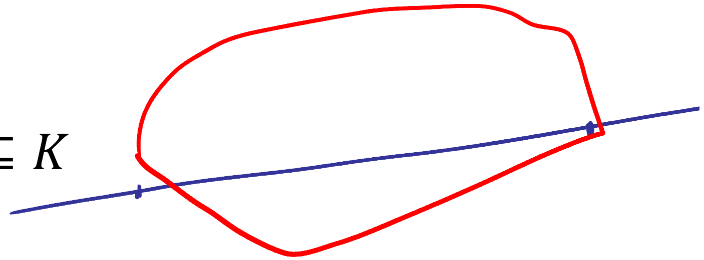
Q. What geometric structure makes algorithmic problems computationally tractable?

(i.e., solvable with polynomial complexity)

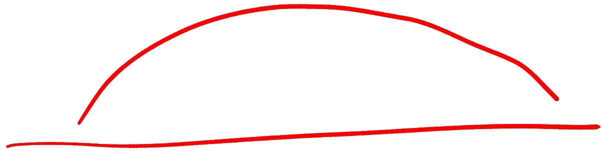
- “Convexity often suffices.”
- Is convexity the frontier of polynomial-time solvability?
- Appears to be in many cases of interest

Convexity

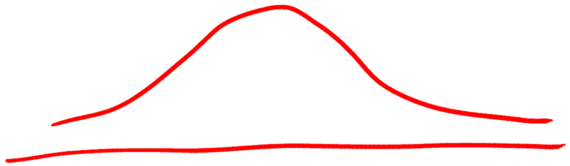
(Indicator functions of) Convex sets:
$\forall x, y \in R^n, \lambda \in [0,1], x, y \in K \Rightarrow \lambda x + (1-\lambda) y \subseteq K$



Concave functions:
$f(\lambda x + (1-\lambda) y) \geq \lambda f(x) + (1-\lambda) f(y)$

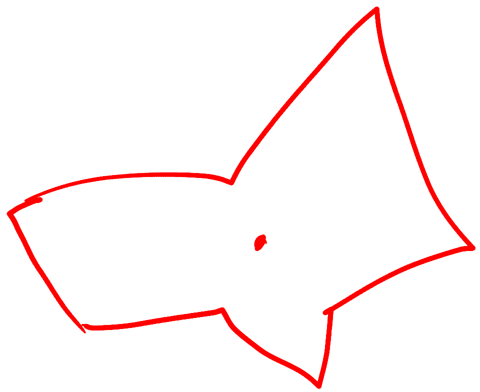


Logconcave functions:
$f(\lambda x + (1-\lambda) y) \geq f(x)^{\lambda} f(y)^{1-\lambda}$



Quasiconcave functions:
$f(\lambda x + (1-\lambda) y) \geq \min\ f(x), f(y)$

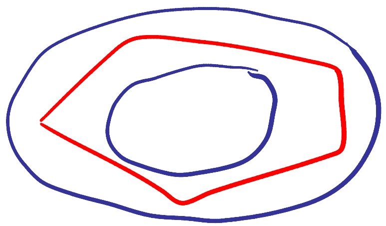
Star-shaped sets:
$\exists x \in S\ s.t. \forall y \in S, \lambda x + (1-\lambda) y \in S$



Sandwiching

Thm (John). Any convex body K has an ellipsoid E s.t.

$$E \subseteq K \subseteq nE.$$



The minimum volume ellipsoid contained in K can be used.

Thm (KLS). For a convex body K in isotropic position,

$$\sqrt{\frac{n+1}{n}}\, B \subseteq K \subseteq \sqrt{n(n+1)}\, B$$

- Also a factor n sandwiching, but with a different ellipsoid.
- As we will see, isotropic sandwiching (rounding) is algorithmically efficient while the classical approach is not.

Rounding via Sampling

1. Sample m random points from K;
2. Compute sample mean z and sample covariance matrix A.
3. Compute $\mathrm{B} = \mathrm{A}^{-\frac{1}{2}}$.

Applying B to K achieves near-isotropic position.

Thm. C(ε).n random points suffice to achieve $E\left(||B-I||_2\right) \leq \epsilon$ for isotropic K.

[Adamczak et al.;Srivastava-Vershynin; improving on Bourgain;Rudelson]

I.e., for any unit vector v, $1+\epsilon \leq E\left((v^T x)^2\right) \leq 1+\epsilon$.

Convex Feasibility

Input: Separation oracle for a convex body K, guarantee that if K is nonempty, it contains a ball of radius r and is contained in the ball of radius R centered the origin.

Output: A point x in K.

Complexity: #oracle calls and #arithmetic operations.

To be efficient, complexity of an algorithm should be bounded by poly(n, log(R/r)).

Convex optimization reduces to feasibility

- To minimize a convex (or even quasiconvex) function f, we can reduce to the feasibility problem via a binary search.

- $K \coloneqq K \cap \{x : f(x) \leq t\}$

- Maintains convexity.

How to choose oracle queries?

Convex feasibility via sampling

[Bertsimas-V. 02]

1. Let z=0, P = $[-R, R]^n$.
2. Does $z \in K$? If yes, output K.
3. If no, let H = $\{x : a^T x \leq a^T z\}$ be a halfspace containing K.
4. Let $P := P \cap H$.
5. Sample $x_1, x_2, \ldots, x_m$ uniformly from P.
6. Let $z := \frac{1}{m} \sum x_i$. Go to Step 2.

Centroid algorithm

- [Levin ‘65]. Use centroid of surviving set as query point in each iteration.

- #iterations = O(nlog(R/r)).
- Best possible.

- Problem: how to find centroid?
- #P-hard! [Rademacher 2007]

Why does centroid work?

Does not cut volume in half.

But it does cut by a constant fraction.

Thm [Grunbaum '60]. For any halfspace H containing the centroid of a convex body K,

$$\mathrm{vol}(K \cap H) \geq \frac{1}{e}\mathrm{vol}(K).$$

Centroid cuts are balanced

K convex. Assume centroid is origin. Fix normal vector of halfspace to be e_1.

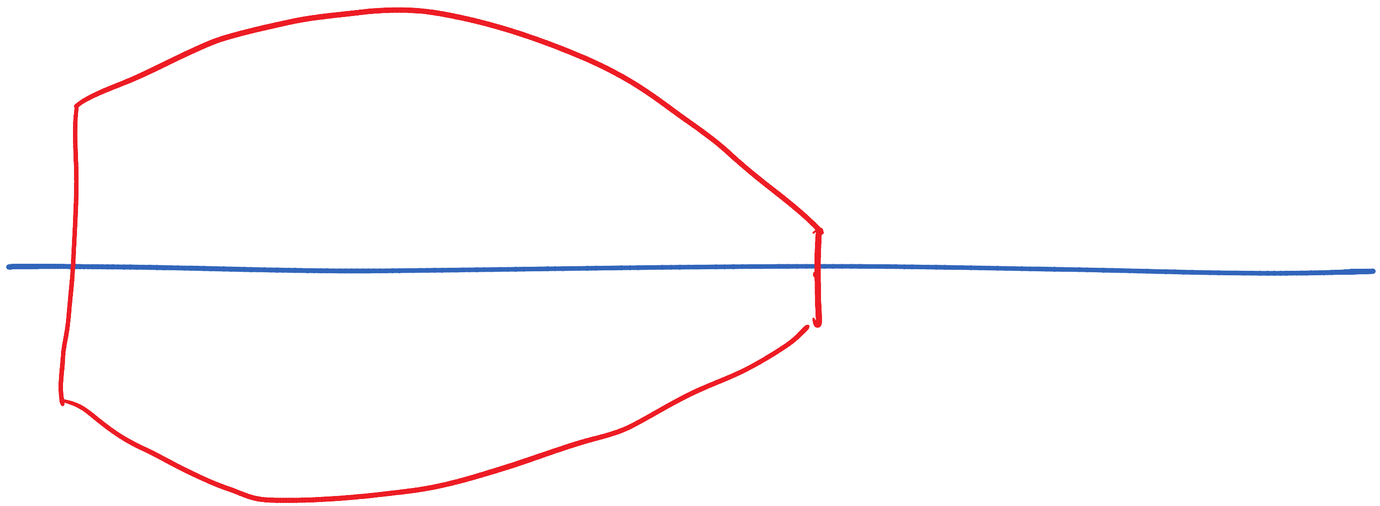
Let $K_t = \{x \in K: x_1 = t\}$ be the slice of K at t.

Symmetrize K: Replace each slice K_t with a ball of the same volume as K_t.

Claim. Resulting set is convex.

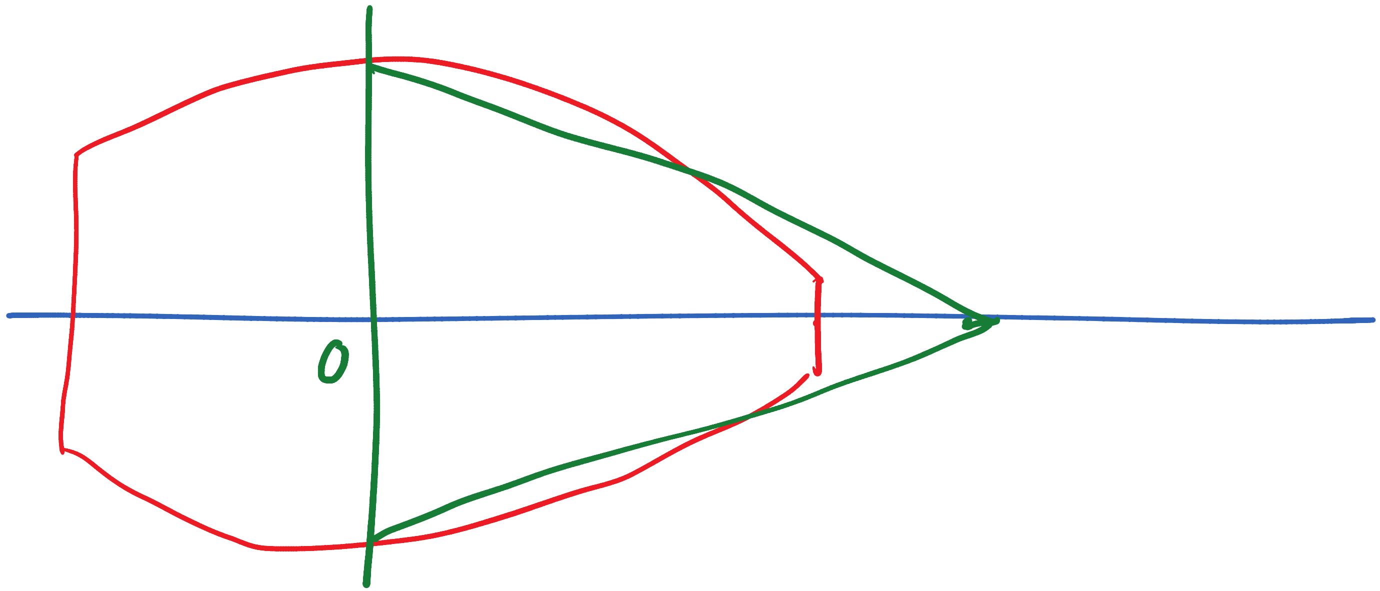
Pf. Use Brunn-Minkowski.

Centroid cuts are balanced



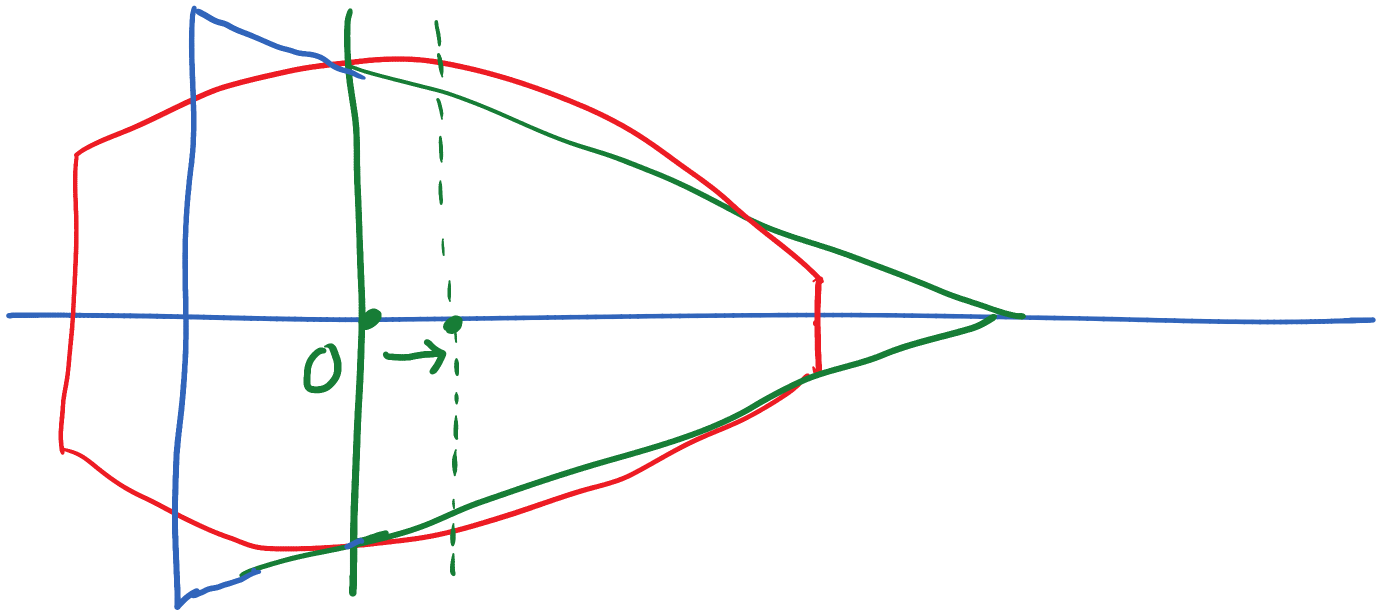
- Transform K to a cone while making the halfspace volume no larger.
- For a cone, the lower bound of the theorem holds.

Centroid cuts are balanced



- Transform K to a cone.
- Maintain volume of right “half”. Centroid moves right, so halfspace through centroid has smaller mass.

Centroid cuts are balanced



- Complete K to a cone. Again centroid moves right.
- So cone has smaller halfspace volume than K.

Cone volume

- Exercise 1. Show that for a cone, the volume of a halfspace containing its centroid can be as small as $\left(\frac{n}{n+1}\right)^n$ times its volume but no smaller.

Convex optimization via Sampling

- How many iterations for the sampling-based algorithm?

- If we use only 1 random sample in each iteration, then the number of iterations could be exponential!

- Do poly(n) samples suffice?

Approximating the centroid

Let $x_1, x_2, \ldots, x_m$ be uniform random from K and y be their average.

Suppose K is isotropic. Then,

E(y)=0, $\quad \mathrm{E}\left(||\mathrm{y}||^2\right) = \frac{1}{\mathrm{m}}\mathrm{E}(||\mathrm{x_i}||^2) = \frac{\mathrm{n}}{\mathrm{m}}$.

So m = O(n) samples give a point y within constant distance of the origin, IF K is isotropic.

Is this good enough? What if K is not isotropic?

Robust Grunbaum: cuts near centroid are also balanced

Lemma [BV02]. For isotropic convex body K and halfspace H containing a point within distance t of the origin,

$$\mathrm{vol}(K \cap H) \geq \left(\frac{1}{e} - t\right) \mathrm{vol}(K).$$

Thm [BV02]. For any convex body K and halfspace H containing the average of m random points from K,

$$E(\mathrm{vol}(K \cap H)) \geq \left(\frac{1}{e} - \sqrt{\frac{n}{m}}\right) \mathrm{vol}(K).$$

Robust Grunbaum: cuts near centroid are also balanced

Lemma. For isotropic convex body K and halfspace H containing a point within distance t of the origin,

$$\mathrm{vol}(K \cap H) \geq \left(\frac{1}{e} - t\right) \mathrm{vol}(K).$$

Proof uses similar ideas as Grunbaum, with more structural properties. In particular,

Lemma. For any 1-dimensional isotropic logconcave function f,

$$\max f < 1.$$

Optimization via Sampling

Thm. For any convex body K and halfspace H containing the average of m random points from K,

$$E(\mathrm{vol}(K \cap H)) \geq \left(\frac{1}{e} - \sqrt{\frac{n}{m}}\right) \mathrm{vol}(K).$$

Proof. We can assume K is isotropic since affine transformations maintain vol(K ∩ H)/vol(K).

Distance of y, the average of random samples, from the centroid is bounded.

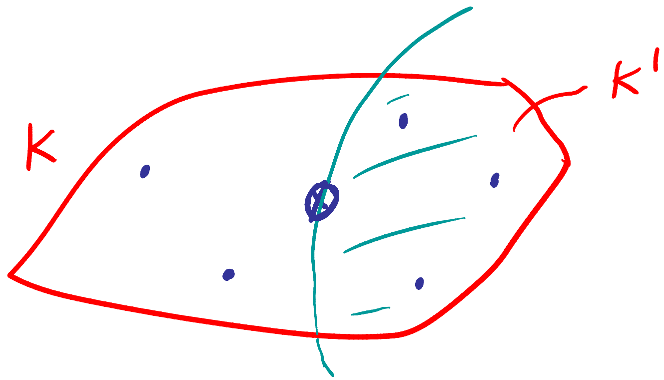
So O(n) samples suffice in each iteration.

Optimization via Sampling

Thm. [BV02] Convex feasibility can be solved using O(n log R/r) oracle calls.

Ellipsoid takes n^2, Vaidya's algorithm also takes O(n log R/r).

With sampling, one can solve convex optimization using only a membership oracle and a starting point in K. We will see this later.



Integration

We begin with the important special case of volume computation: Given convex body K, and parameter ϵ, find a number A s.t.

$$(1-\epsilon)\mathrm{vol}(K) \leq A \leq (1+\epsilon)\mathrm{vol}(K).$$

Volume via Rounding

- Using the John ellipsoid or the Inertial ellipsoid

$$\mathrm{E} \subseteq \mathrm{K} \subseteq n\mathrm{E} \;\Rightarrow\; \mathrm{vol(E)} \leq \mathrm{vol(K)} \leq \mathrm{n^n\,vol(E)}.$$

- Polytime algorithm, $n^{O(n)}$ approximation to volume

- Can we do better?

Complexity of Volume Estimation

Thm [E86, BF87]. For any deterministic algorithm that uses at most n^a membership calls to the oracle for a convex body K and computes two numbers A and B such that $\mathrm{A} \leq \mathrm{vol(K)} \leq \mathrm{B}$, there is some convex body for which the ratio B/A is at least

$$\left(\frac{cn}{a \log n}\right)^{\frac{\mathrm{n}}{2}}$$

where c is an absolute constant.

Complexity of Volume Estimation

Thm [BF]. For deterministic algorithms:

# oracle calls	approximation factor
n^a	$\left(\frac{cn}{a \log n}\right)^{\frac{n}{2}}$
$\left(\frac{1}{\varepsilon}\right)^n$	$(1+\varepsilon)^n$

Thm [DV12].

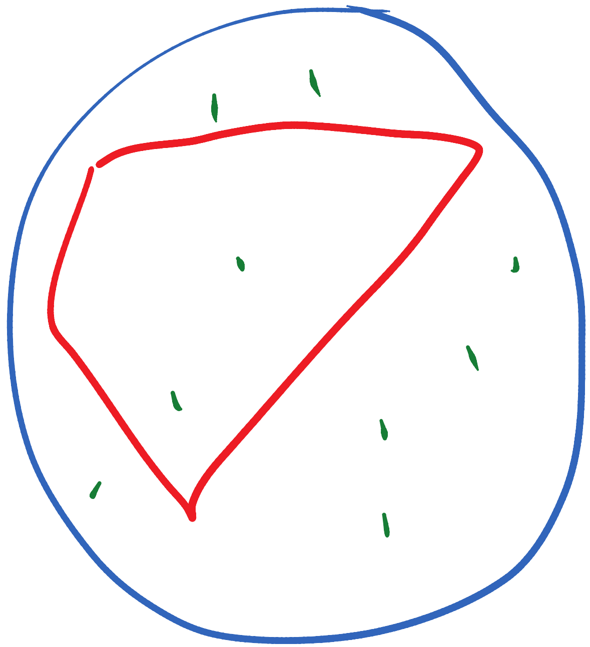
Matching upper bound of $(1+\epsilon)^n$ in time $\left(\frac{1}{\epsilon}\right)^{O(n)}$ poly(n).

Volume computation

[DFK89]. Polynomial-time randomized algorithm that estimates volume with probability at least $1 - \delta$ in time poly(n, $\frac{1}{\epsilon}$, $\log\left(\frac{1}{\delta}\right)$).

Volume by Random Sampling

- Pick random samples from ball/cube containing K.
- Compute fraction c of sample in K.
- Output c.vol(outer ball).

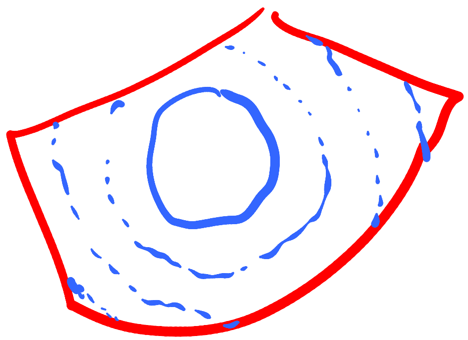


- Need too many samples

Volume via Sampling

$$B \subseteq K \subseteq RB.$$

Let $K_i = K \cap 2^{i/n} B,\ \ i = 0, 1, \ldots, m = n \log R.$



$$\mathrm{vol(K)} = \mathrm{vol(B)}.\frac{\mathrm{vol(K_1)}}{\mathrm{vol(K_0)}}\frac{\mathrm{vol(K_2)}}{\mathrm{vol(K_1)}}\cdots\frac{\mathrm{vol(K_m)}}{\mathrm{vol(K_{m-1})}}.$$

Estimate each ratio with random samples.

Volume via Sampling

$$K_i = K \cap 2^{i/n} B, \qquad i = 0, 1, \dots, m = n \log R.$$

$$\mathrm{vol(K)} = \mathrm{vol(B)}. \frac{\mathrm{vol(K_1)}}{\mathrm{vol(K_0)}} \frac{\mathrm{vol(K_2)}}{\mathrm{vol(K_1)}} \dots \frac{\mathrm{vol(K_m)}}{\mathrm{vol(K_{m-1})}}.$$

Claim. $\mathrm{vol(K_{i+1})} \le 2. \mathrm{vol(K_i)}$.

Total #samples $= m. \frac{m}{\epsilon^2} = O^*(n^2)$.

Variance of product

Exercise 2. Let Y be the product estimator

$$Y = \prod X^i$$

with each X^i, i=1,2,…, m, estimated using k samples

as $X^i = \frac{1}{k}\sum_j X_j^i$ with $E\left(X_j^i\right) = \frac{vol(K_{i-1})}{vol(K_i)}$.

Show that

$$\mathrm{var}(\mathrm{Y}) \le \left(\left(1 + \frac{3}{k}\right)^m - 1\right)\mathrm{E}(\mathrm{Y})^2.$$

Appears to be optimal

- n phases, O*(n) samples in each phase.
- If we only took m < n phases, then the ratio to be estimated in some phase could be as large as $n^{n/m}$ which is superpoly for m = o(n).
- Is $\Omega(n^2)$ total samples the best possible?

Simulated Annealing [Kalai-V.04,Lovasz-V.03]

To estimate $\int f$ consider a sequence

$$f_0, f_1, f_2, \dots, f_m = f$$

with $\int f_0$ being easy, e.g., constant function over ball.

Then, $$\int f = \int f_0 . \frac{\int f_1}{\int f_0} . \frac{\int f_2}{\int f_1} \cdots \frac{\int f_m}{\int f_{m-1}} .$$

Each ratio can be estimated by sampling:

1. Sample X with density proportional to f_i
2. Compute $Y = \frac{f_{i+1}(X)}{f_i(X)}$

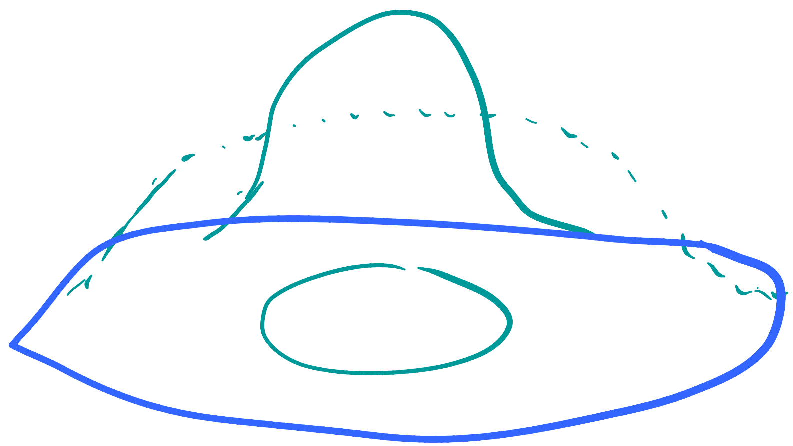
$$E(Y) = \int \frac{f_{i+1}(X)}{f_i(X)} . \frac{f_i(X)}{\int f_i(X)} dX = \frac{\int f_{i+1}}{\int f_i} .$$

A tight reduction [LV03]

Define: $f_i(X) = e^{-a_i||X||}$

$$a_0 = 2R, \quad a_{i+1} = a_i\left(1 - \frac{1}{\sqrt{n}}\right), \quad a_m = \frac{\epsilon}{2R}$$

$m \sim \sqrt{n}\log(2R/\epsilon)$



Volume via Annealing

$$f_i(X) = e^{-a_i||X||} \qquad a_{i+1} = a_i\left(1 - \frac{1}{\sqrt{n}}\right),$$

$$Y = \frac{f_{i+1}(X)}{f_i(X)} \qquad E(Y) = \frac{\int f_{i+1}}{\int f_i}$$

Lemma. $E(Y^2) \leq 4E(Y)^2$ for large enough n.

Although expectation of Y can be large (exponential even), it has small variance!

Proof via logconcavity

Exercise 2. For a logconcave function $f: R^n \to R$,

let $Z(a) = \int f(X)^a dX$ for $a > 0$.

Show that $a^n Z(a)$ is a logconcave function.

[Hint: Define $F(x, a) = f\left(\frac{x}{t}\right)^t$.]

Proof via logconcavity

$Z(a) = \int f(aX)dX$

$a^n Z(a)$ is a logconcave function.

$$E(Y_i) = \frac{Z(a_{i+1})}{Z(a_i)} \quad E\left(Y_i^2\right) = \frac{Z(2a_{i+1}-a_i)}{Z(a_i)}$$

$$\frac{E\left(Y_i^2\right)}{E(Y_i)} = \frac{Z(2a_{i+1}-a_i)Z(a_i)}{Z(a_{i+1})^2} \leq \frac{(a_{i+1})^{2n}}{(2a_{i+1}-a_i)^n (a_i)^n} \leq 4.$$

Progress on volume

	Power	New ideas
Dyer-Frieze-Kannan 91	23	everything
Lovász-Siminovits 90	16	localization
Applegate-K 90	10	logconcave integration
L 90	10	ball walk
DF 91	8	error analysis
LS 93	7	multiple improvements
KLS 97	5	speedy walk, isotropy
LV 03,04	4	annealing, wt. isoper.
LV 06	4	integration, local analysis

Optimization via Annealing

We can minimize quasiconvex function f over convex set S given only by a membership oracle and a starting point in S. [KV04, LV06].

Almost the same algorithm, in reverse: to find max f, define

$f_i(X) = f(X)^{a_i} \quad i = 1, \dots, m. \qquad a_0 = \epsilon, \; a_m =$ M.

sequence of functions starting at nearly uniform and getting more and more concentrated points of near-optimal objective value.

Lecture 3: Sampling Algorithms

- Sampling by random walks
- Conductance
- Grid walk, Ball walk, Hit-and-run
- Isoperimetric inequalities
- Rapid mixing

High-Dimensional Sampling Algorithms

Santosh Vempala
Algorithms and Randomness Center
Georgia Tech

Sampling

- Generate a uniform random point from a set S or with density proportional to function f.
- Numerous applications in diverse areas: statistics, networking, biology, computer vision, privacy, operations research etc.
- This course: mathematical and algorithmic foundations of sampling and its applications.

Structure

Q. What geometric structure makes algorithmic problems computationally tractable?

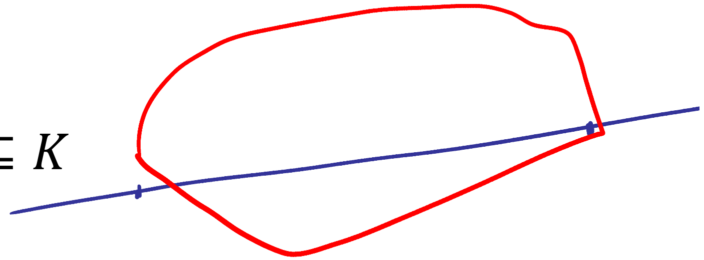
(i.e., solvable with polynomial complexity)

- “Convexity often suffices.”
- Is convexity the frontier of polynomial-time solvability?
- Appears to be in many cases of interest

Convexity

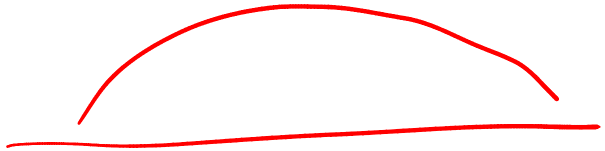
(Indicator functions of) Convex sets:

$\forall x, y \in R^n, \lambda \in [0,1], x, y \in K \Rightarrow \lambda x + (1-\lambda) y \subseteq K$



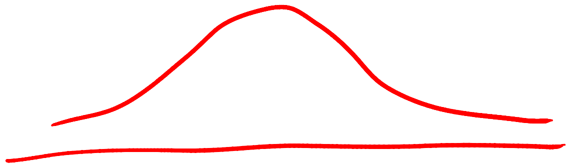
Concave functions:

$f(\lambda x + (1-\lambda) y) \geq \lambda f(x) + (1-\lambda) f(y)$



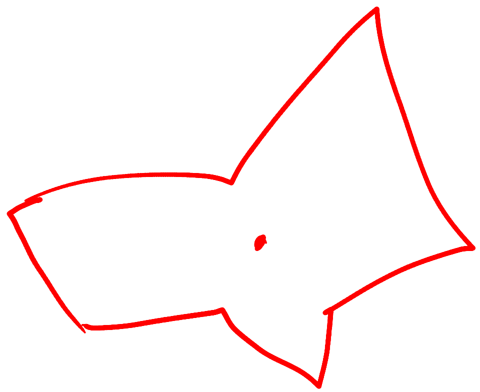
Logconcave functions:

$f(\lambda x + (1-\lambda) y) \geq f(x)^{\lambda} \, f(y)^{1-\lambda}$



Quasiconcave functions:

$f(\lambda x + (1-\lambda) y) \geq \min \;\; f(x), f(y)$



Star-shaped sets:

$\exists x \in S \; s.t. \forall y \in S, \lambda x + (1-\lambda) y \in S$

Annealing

Integration

- $f_i(X) = f(X)^{a_i}, X \in K$
- $a_0 = \frac{\epsilon}{2R}, \quad a_m = 1$
- $a_{i+1} = a_i \left(1 + \frac{1}{\sqrt{n}}\right)$
- Sample with density prop. to $f_i(X)$.
- Estimate $W_i \sim \int f_{i+1}(X) / \int f_i(X)$
- Output $W = W_1 W_2 \ldots W_m$.

Optimization

- $f_i(X) = f(X)^{a_i}, X \in K$
- $a_0 = \frac{\epsilon}{2R}, \quad a_m = \frac{2n}{\epsilon}$
- $a_{i+1} = a_i \left(1 + \frac{1}{\sqrt{n}}\right)$
- Sample with density prop. to $f_i(X)$.
- Output X with max f(X).

How to sample?

Take a random walk in K.

Consider a lattice intersected with K

Grid (lattice) walk:

At grid point x,

pick random y from $\{x \pm \delta e_i\}$

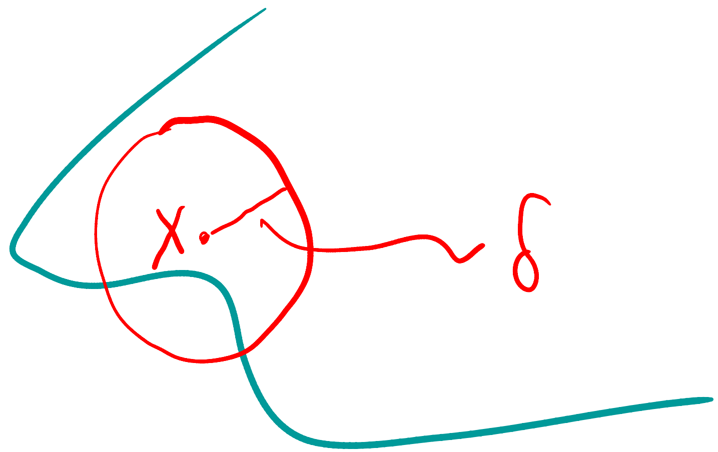
if y is in K, go to y

Ball walk

At x,

pick random y from $x + \delta B_n$

if y is in K, go to y



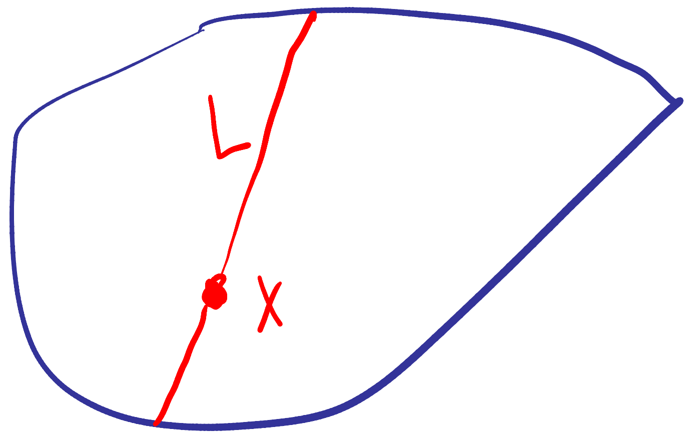
Hit-and-run

[Boneh, Smith]

At x,

-pick a random chord L through x

-go to a uniform random point y on L



Markov chains

- State space K,
- set of measurable subsets that form a σ-algebra, i.e., closed under finite unions and intersections
- A next step distribution $P_u(.)$ associated with each point u in the state space.
- A starting point.

- $w_0, w_1, \ldots, w_k, \ldots$ s.t.

$$P(w_k \in A \mid w_0, w_1, \ldots, w_{k-1}) = P(w_k \in A \mid w_{k-1})$$

Convergence

Stationary distribution Q, ergodic “flow” defined as

$$\Phi(A) = \int_A P_u(K¥\mathrm{A})\,dQ(u)$$

For a stationary distribution Q, we have

$$\Phi(A) = \Phi(K¥\mathrm{A})$$

Random walks in K

- For both walks, the distribution of the current point tends to uniform in K.

- The uniform distribution is stationary, in fact,

$$Q(u)P_u(v) = Q(v)P_v(u).$$

Exercise 1. Show that the uniform distribution is stationary for hit-and-run.

- Question: How many steps are needed?

Rate of convergence?

Ergodic “flow”:

$$\Phi(A) = \int_A P_u(K¥\mathrm{A})\, dQ(u)$$

Conductance:

$$\phi(A) = \frac{\Phi(A)}{\min Q(A),\ Q(K¥\mathrm{A})}$$

$$\phi = \inf\ \phi(A)$$

Conductance

Mixing rate cannot be faster than $1/\phi$

Since it takes this many steps to even escape from some subsets.

Does ϕ give an upper bound? Yes, for discrete Markov chains

Thm. [Jerrum-Sinclair] $\frac{\phi^2}{2} \leq 1 - \lambda \leq 2\phi$

Where λ is the second eigenvalue of the transition matrix.

Thus, mixing rate = $\frac{1}{1-\lambda} \leq \frac{2}{\phi^2}$.

Rate of convergence

THM [LS93]. $M = \sup_A \frac{Q_0(A)}{Q(A)}$ "WARM START"

$$d_{tv}(Q_t, Q) \le \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t.$$

THM [LS93]. $M = \mathrm{Var}_Q\left(\frac{Q_0(u)}{Q(u)}\right)$

$$\forall \varepsilon > 0,\ d_{tv}(Q_t, Q_0) \le \varepsilon + \sqrt{\frac{M}{\varepsilon}}\left(1 - \frac{\phi^2}{2}\right)^t.$$

High conductance => rapid mixing

Proof does not go through eigenvalue gap

How to bound conductance?

- Conductance of ball walk is not bounded!
- Local conductance can be arbitrarily small.

$$\ell(x) = \frac{\mathrm{vol}(x + \delta B_n \cap K)}{\mathrm{vol}(\delta B_n)}$$

- What can we do?
- Modify K slightly
- Or start with a nearly random point in K.

Smoothing a convex body

$$K + \alpha B_n$$

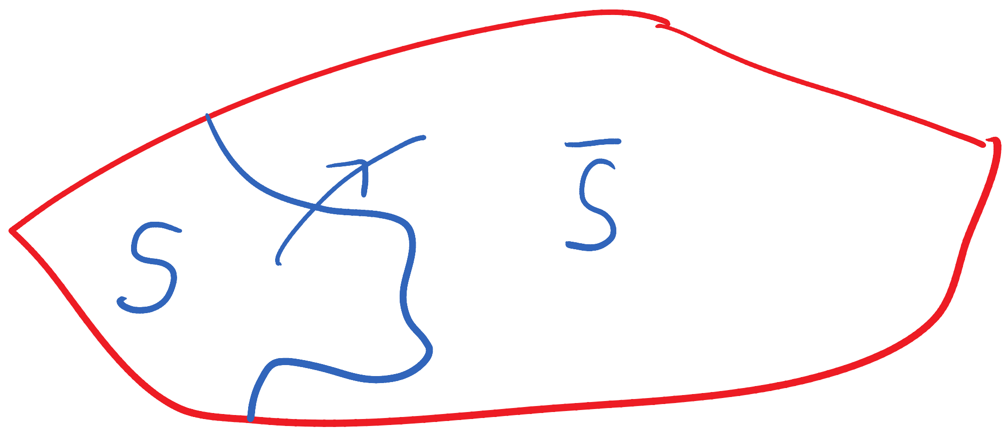
Each point of the original body has a small ball around it.

What about new points? No worse than local conductance of boundary points of a small ball.

Choosing step radius $\delta \leq \alpha/\sqrt{n}$ will ensure that every point has local conductance at least a fixed constant.

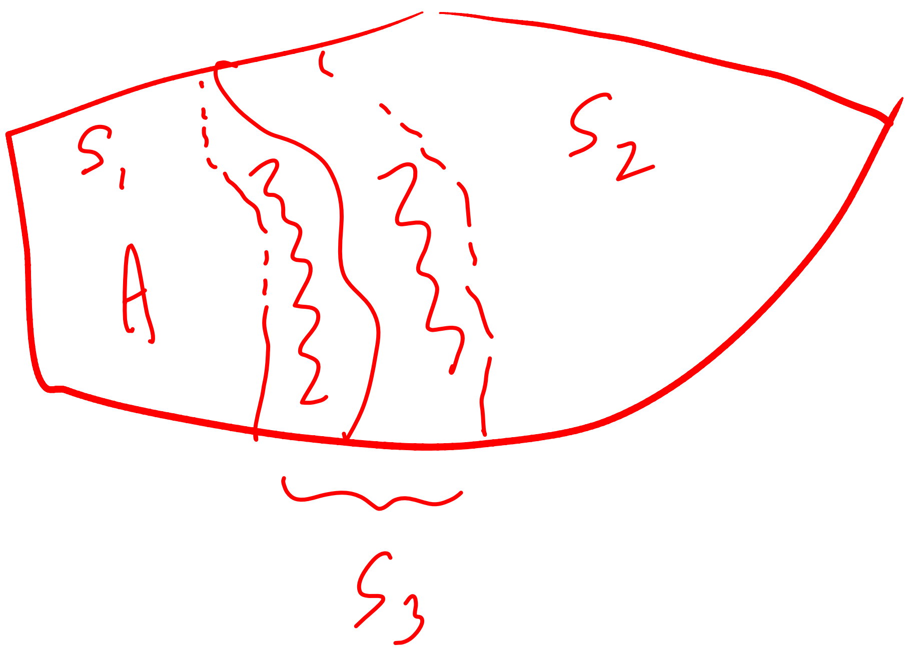
Conductance

Consider an arbitrary measurable subset S.



We need to show that the escape probability from S is large.

Conductance



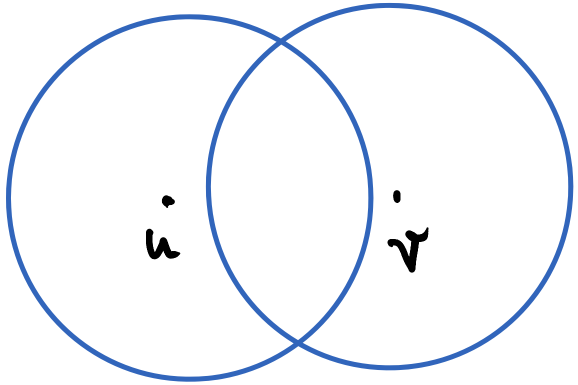
Need:

- Points that do not cross over are far from each other
- If two subsets are far, then the rest of the set is large

One-step distributions

Idea:

$d(P_u, P_v)$ large



$\Rightarrow$ the balls around u,v have small intersection

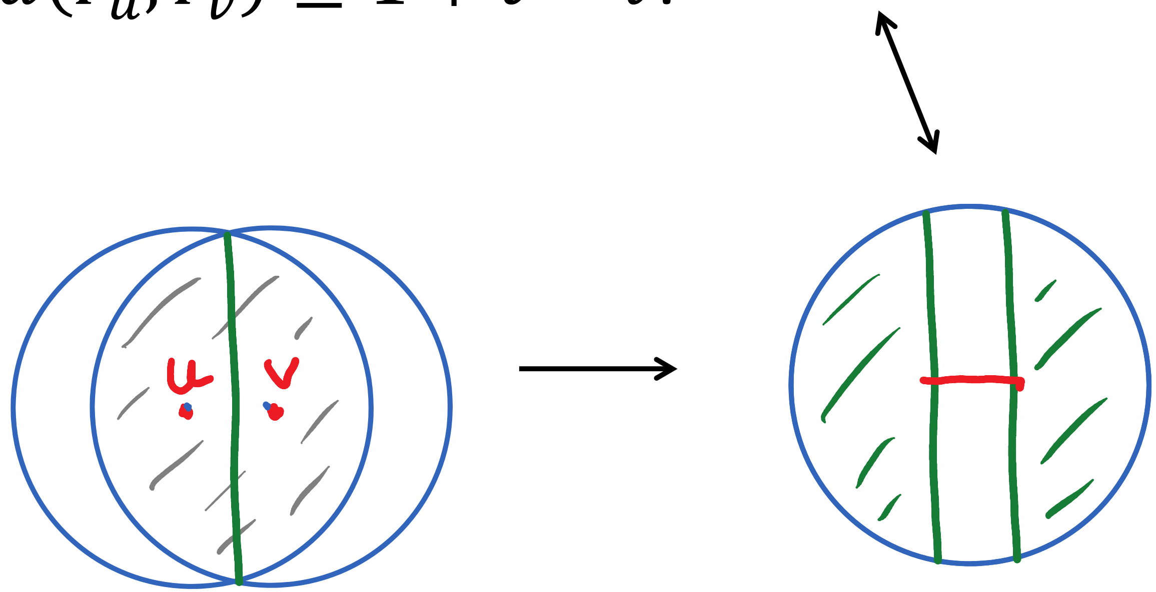
$\Rightarrow$ u,v must be far

Prob. distance ⇒ Geometric distance

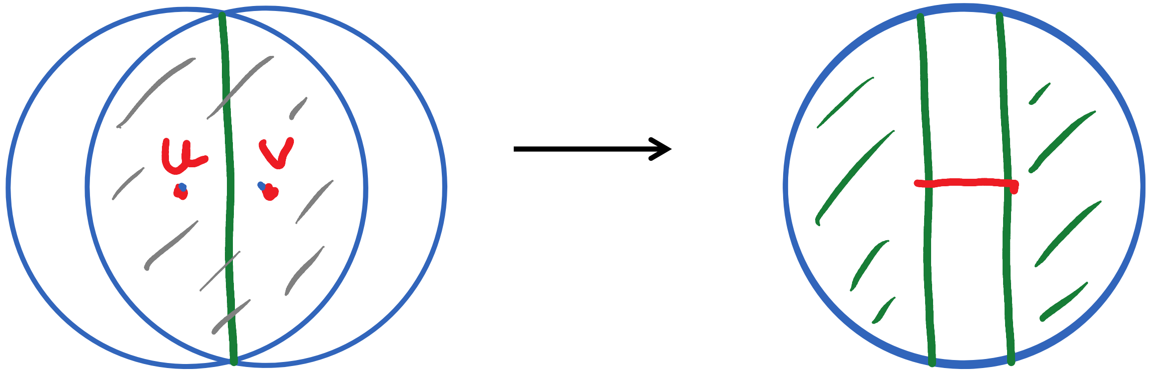
Lemma. $u, v \in K, \ell(u), \ell(v) \geq \ell$ for the ball walk with δ-steps. If

$$d(u, v) \leq \frac{t\delta}{\sqrt{n}},$$

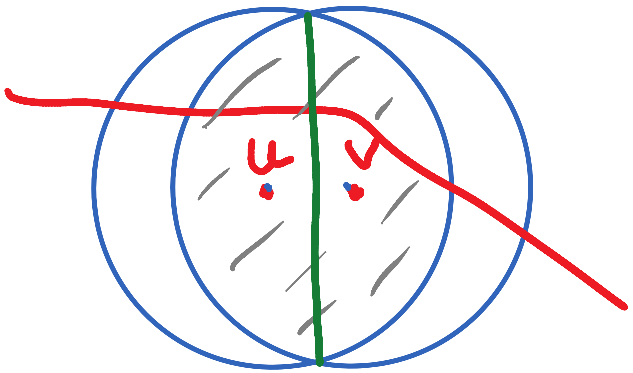
then $d(P_u, P_v) \leq 1 + t - \ell$.



Coupling 1-step distributions



$$d(P_u, P_v) = \mathrm{vol}(\square) \leq t$$

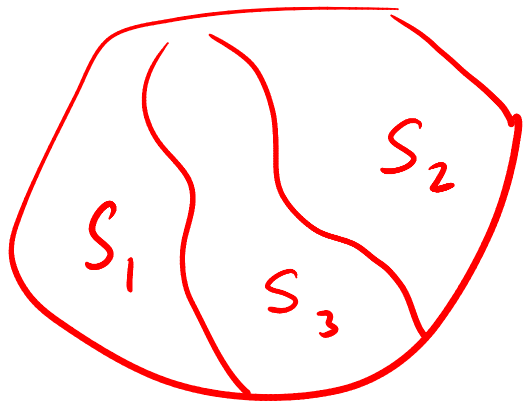


if $\ell < 1, d(P_u, P_v) \leq t + 1 - \ell$

Isoperimetry

THM [LS92, DF91].

$$\mathrm{vol}(S_3) \geq \frac{2d(S_1,S_2)}{D} \min\{\mathrm{vol}(S_1), \mathrm{vol}(S_2)\}$$



Extends to logconcave densities:

$$\pi_f(S_3) \geq \frac{2d(S_1,S_2)}{D} \min\{\pi_f(S_1), \pi_f(S_2)\}$$

π : DISTRIBUTION WITH LOGCONCAVE DENSITY f.

$$f\left(\frac{x+y}{2}\right) \geq \sqrt{f(x)f(y)}$$

Conductance

Thm. Conductance of ball walk is at least

$$\frac{\ell^2 \delta}{16\sqrt{n}D}$$

We can use

$$\ell = \frac{1}{n}, \delta = \frac{\ell}{\sqrt{n}} = \frac{1}{n\sqrt{\mathrm{n}}}$$

So

$$\phi \geq \frac{C}{n^2 D}, \qquad \text{mixing rate} = O(n^4 D^2)$$

Conductance

Thm. Conductance of ball walk is at least $\frac{\ell^2 \delta}{16\sqrt{n}D}$

Pf.

$$S_1 = \left\{x \in S : \ P_x\,(K¥\mathrm{S}) < \frac{\ell}{4}\right\} \quad S_2 = \left\{x \in K¥S : \ P_x\,(\mathrm{S}) < \frac{\ell}{4}\right\}$$

$$S_3 = K¥\mathrm{S}_1¥\mathrm{S}_2$$

$$vol(S_1) \geq \frac{vol(S)}{2}, vol(S_2) \geq \frac{vol(K¥\mathrm{S})}{2}$$

If not,

$$\int_S P_x(K¥\mathrm{S})dx \geq \frac{\ell}{4}.\frac{1}{2}vol(S) \quad \Rightarrow \quad \phi(S) \geq \frac{\ell}{8}.$$

Conductance

Thm. Conductance of ball walk is at least $\frac{\ell^2\delta}{16\sqrt{n}D}$

Pf.

$$S_1 = \left\{x \in S : \ P_x\,(K\text{¥}\mathrm{S}) < \frac{\ell}{4}\right\} \quad S_2 = \left\{x \in K\text{¥}S : \ P_x\,(\mathrm{S}) < \frac{\ell}{4}\right\}$$

For $u \in S_1, v \in S_2$,

$$d(P_u, P_v) \geq 1 - P_u(K\text{¥}\mathrm{S}) - P_v(S) > 1 - \frac{\ell}{2} \ \Rightarrow \ d(u, v) \geq \frac{\ell\delta}{2\sqrt{n}}.$$

$$vol(S_3) \geq \frac{\ell\delta}{\sqrt{n}} \min \ vol(S_1), vol(S_2)$$

$$\geq \frac{\ell\delta}{2\sqrt{n}} \min \ vol(S), vol(K\text{¥}\mathrm{S}).$$

Conductance

Thm. Conductance of ball walk is at least $\frac{\ell^2 \delta}{16\sqrt{n}D}$

Pf.

$$\int_S P_x(K¥S)dx = \frac{1}{2}\int_S P_x(K¥S)dx + \frac{1}{2}\int_{K¥S} P_x(S)dx$$

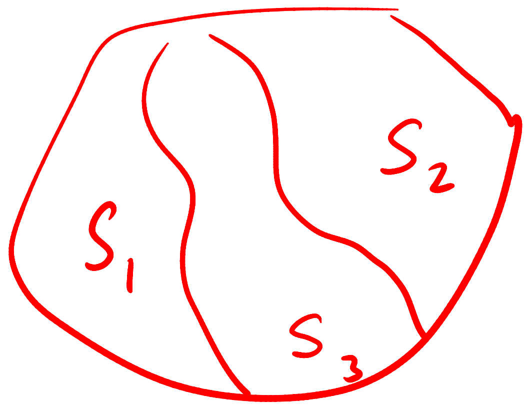
$$\geq \frac{1}{2} . \frac{\ell}{4} . vol(S_3)$$

$$\geq \frac{\ell^2 \delta}{16\sqrt{n}D} \min\ vol(S), vol(K¥S).$$

KLS hyperplane conjecture

A: covariance matrix of stationary distribution

$$E(\|X-\bar{x}\|^2) = \operatorname{tr}(A) = \sum_i \lambda_i(A)$$



THM. $\pi_f(S_3) \geq \dfrac{\ln 2 \, d(S_1,S_2)}{\sqrt{\sum \lambda_i(A)}} \pi_f(S_1)\pi_f(S_2)$

OR $\phi_f \geq \dfrac{1}{\sqrt{\sum_i \lambda_i(A)}}$

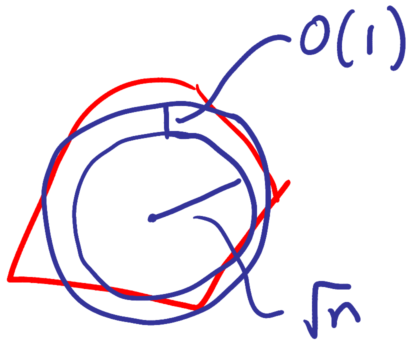
CONJ. $\exists c > 0,$

$$\phi_f \geq \frac{c}{\sqrt{\lambda_1(A)}} = c \text{ for isotropic } f.$$

Thin shell conjecture

Theorem [Bobkov].

$$\phi_f \geq \frac{c}{Var(\|x\|^2)^{1/4}}.$$



Conj. (Thin shell) $VAR(\|x\|^2) = O(n)$

Alternatively: $E((\|x\| - \sqrt{n})^2) = O(1)$

Current best bound [Guedon-E. Milman]: $n^{1/3}$

KLS-Slicing-Thin-shell

	known	conj
thin shell	$n^{1/3}$	$O(1)$
slicing	$n^{1/4}$	$O(1)$
KLS	$n^{5/12}$	$O(1)$

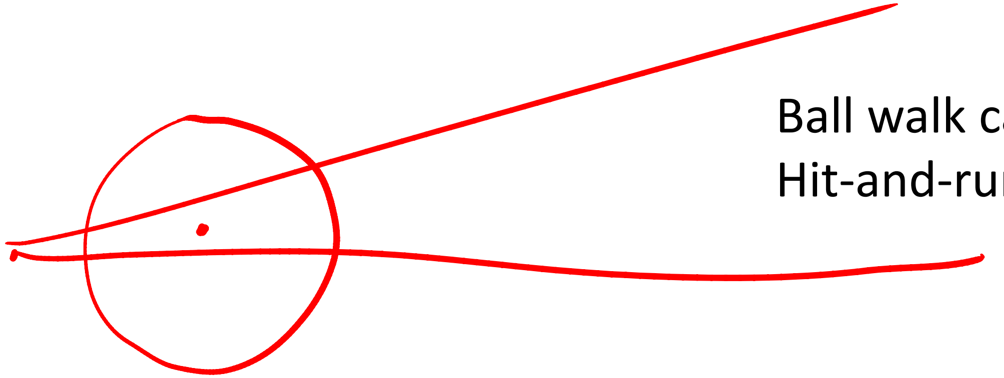
Moreover, KLS implies the others [Ball] and thin-shell implies slicing [Eldan-Klartag10].

Convergence

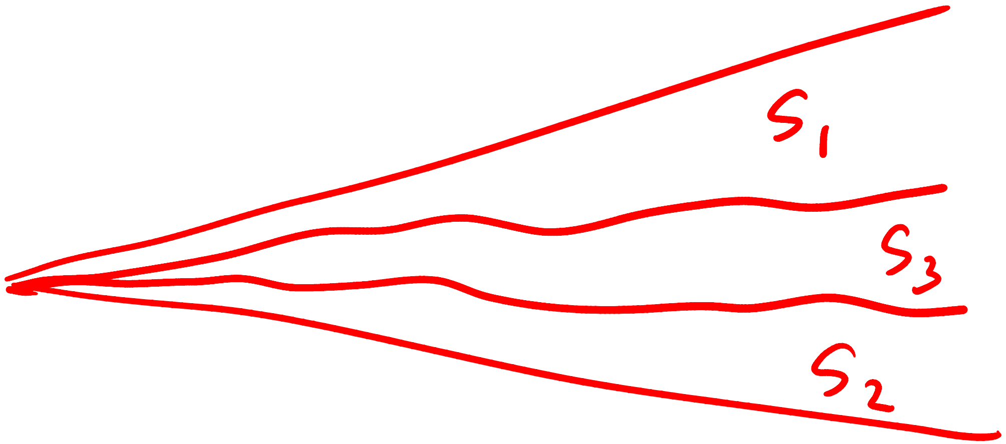
Thm. [LS93, KLS97] If S is convex, then the ball walk with an M-warm start reaches an (independent) nearly random point in poly(n, D, M) steps.

- Strictly speaking, this is not rapid mixing!
- How to get the first random point?
- Better dependence on diameter D?

Is rapid mixing possible?



Ball walk can have bad starts, but
Hit-and-run escapes from corners



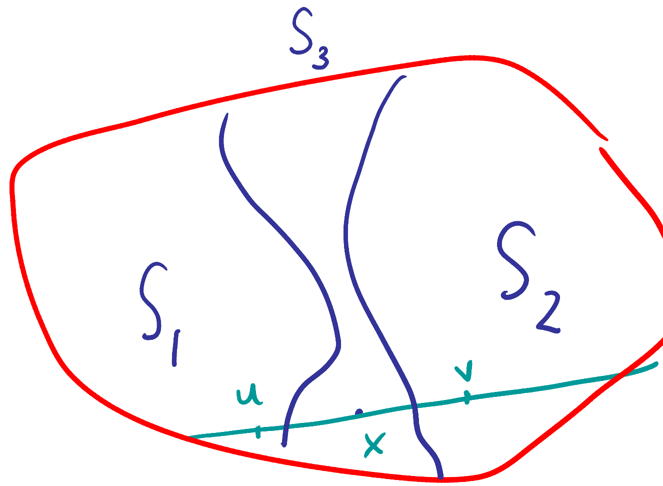
$d(S_1, S_2) \to 0$

Min distance based isoperimetry is too coarse

Average distance isoperimetry

- How to average distance?

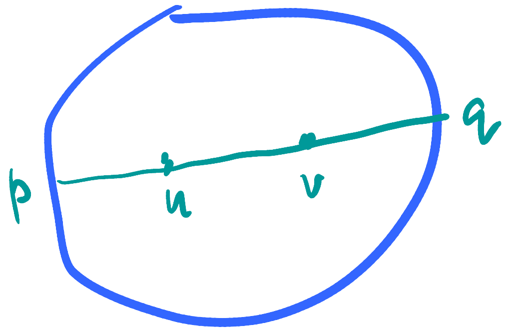
$$h(x) \le \min_{\substack{u \in S_1,\, v \in S_2 \\ x \in \ell(x,y)}} d(u,v)$$



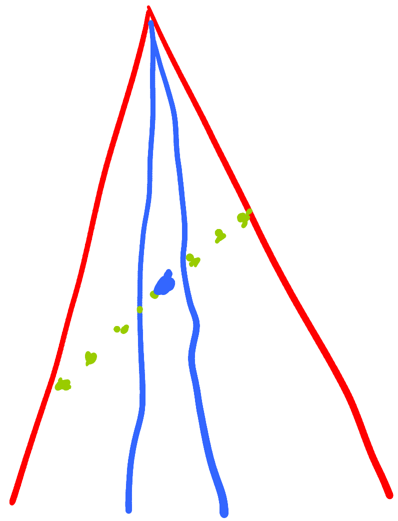
- Theorem.[LV04; Dieker-V.12]

$$\pi(S_3) \ge \frac{E(h(x))}{D} \pi(S_1)\pi(S_2)$$

Average distance Isoperimetry



$$d_k(u,v) = \frac{|u-v|\,|p-q|}{|p-u|\,|v-q|}$$



<u>Thm</u> [LV]. $\pi_f(S_3) \geq d_k(S_1,S_2)\,\pi_f(S_1)\,\pi_f(S_2)$

$$h(x) = \frac{1}{3} \min_{\substack{u\in S_1,\, v\in S_2,\\ x\in[u,v]}} d_k(u,v)$$

<u>Thm</u> [LV04] $\pi_f(S_3) \geq E(h(x))\,\pi_f(S_1)\,\pi_f(S_2)$.

Hit-and-run

- Thm [LV04]. Hit-and-run mixes in polynomial time from *any* starting point inside a convex body.
- Conductance = $\Omega\left(\frac{1}{nD}\right)$
- Gives $O^*(n^3)$ sampling algorithm

Multi-point random walks

- Maintain m points
- For each point X,
  - Pick a random combination of the m points
  - Use this to update X

Stationary distribution: m uniform random points!

Sampling

Q1. Is starting at a nice point faster? E.g., does ball walk mix rapidly starting at a single point, e.g., the centroid?

Q2. How to check convergence to stationarity on the fly? Does it suffice to check that the measures of all halfspaces have converged?

(Note: poly(n) sample can estimate all halfspace measures approximately)

Sampling: current status

Can be sampled efficiently:

- Convex bodies
- Logconcave distributions
- (1/n-1)-harmonic-concave distributions
- Near-logconcave distributions
- Star-shaped bodies
- ??

Cannot be sampled efficiently:

- Quasiconcave distributions

High-dimensional sampling algorithms

- Sampling manifolds
- Random reflections
- Deterministic sampling?
- Other applications…