High-dimensional Sampling Algorithms

Santosh Vempala, Georgia Tech

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)

- 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.

Sampling

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

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

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

Problem 1: Optimization

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

point x, error parameter ε.

Output: An affine transformation that approximately "sandwiches" f between two concentric balls (ellipsoids).

Problem 5: Learning

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

Output: A rule to correctly label $1-\varepsilon$ of the input distribution.

(more general than integration)

Talk Outline

- Part I. Quick intro to high-dimensional geometry.
- Part 2. Algorithms based on sampling.
- Part 3. Sampling algorithms.

High-dimensional Algorithms

These problems are intractable in general, but

P1. Optimization. Minimize f over 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.

Structure

Q. What structure makes high-dimensional problems such as sampling computationally tractable? (i.e., solvable with polynomial complexity)

- Convexity appears to be the frontier of polynomial-time solvability in many settings.

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:

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

Structure I: Separation Oracle

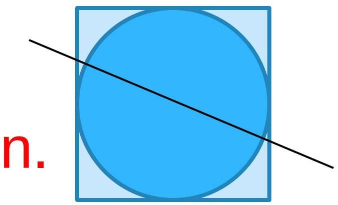
Q. How to specify a convex set?

Either x is in K or some halfspace contains K but not x.

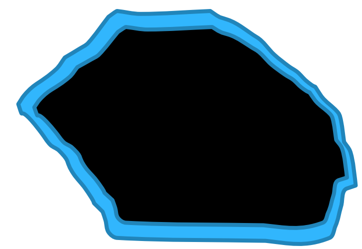
LP, SDP, graph problems…

Structure II: Volume distribution

- Volume(unit cube) = 1
- Volume(unit ball) $\sim \left(\frac{c}{n}\right)^{\frac{n}{2}}$ drops exponentially with n.



- For any central hyperplane, most of the mass of a ball is within distance $1/\sqrt{n}$.



- Most of the volume is near the boundary:

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

So,

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

- “Everything interesting for a convex body happens near its boundary” --- Imre Bárány.

Brunn-Minkowski inequality

A, B compact sets in R^n

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

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

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

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 the following are:

- Product
- Minimum
- Convolution

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

- Any marginal:

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

Isotropic position

- Let x be a random point from a convex body K
- Set E(x)=0. Consider the covariance matrix
$$A = E(xx^T), \qquad A_{ij} = E(x_i x_j)$$
- $A = B^2$ for some n x n matrix B.

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

- For a random point y from K',
$$E(y) = 0, \qquad E(yy^T) = I_n.$$
- K' is in isotropic position.

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 comparable radii?
- Yes!

Structure III: Sandwiching

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

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

The maximum volume ellipsoid contained in K can be used.

Thm (KLS95). For a convex body K in isotropic position,

- 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.

Part 2: Algorithmic Applications

Given a blackbox for sampling, we can get efficient algorithms for:

- Rounding
- Convex Optimization
- Volume Computation/Integration

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. For isotropic K, C(ϵ).n random points suffice to get

$$E\left(\left\|A - I\right\|_2\right) \le \epsilon.$$

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

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

Convex Optimization/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.

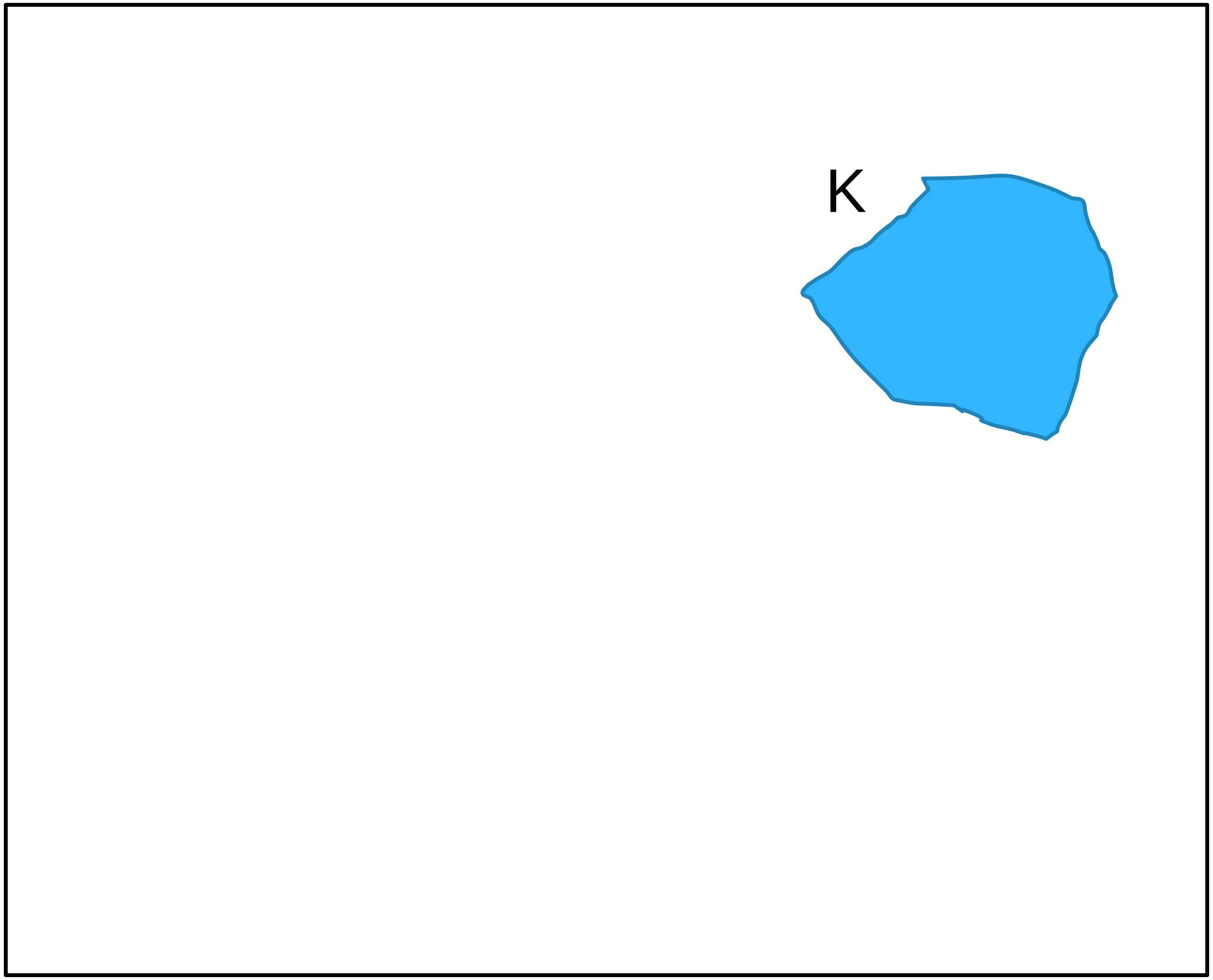
Output: A point x in K.

- We can reduce to the feasibility problem for (quasi-)concave functions via a binary search.
- $K := K \cap \{x : f(x) \leq t\}$

Complexity: #oracle calls and #arithmetic operations.

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

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 $P \coloneqq P \cap H$.
5. Sample $x_1, x_2, \ldots, x_m$ uniformly from P.
6. Let $z \coloneqq \frac{1}{m} \sum x_i$. 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 would 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,

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

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?

Robust Grunbaum: cuts near centroid are also balanced

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

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

Optimization via Sampling

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

Also achieved by Vaidya's algorithm; Ellipsoid takes n^2.

With sampling, one can solve convex optimization using only a membership oracle and a starting point in K.

Volume/Integration

Given convex body K, find a number A s.t.

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

Given function f, find A s.t.

$$(1-\epsilon)\int_{R^n} f \le A \le (1+\epsilon)\int_{R^n} f$$

Volume via Rounding

- Using the John ellipsoid or the Inertial ellipsoid

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

- 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 $\mathrm{A} \leq \mathrm{vol}(\mathrm{K}) \leq \mathrm{B}$, 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

Thm [Dadush-V.12].

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

Volume computation/Integration

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

[Applegate-K91]. Polytime randomized algorithm to estimate integral of any (Lipshitz) logconcave function.

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

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

Simulated Annealing [LV03, Kalai-V.04]

To estimate $\int f$ consider a sequence $f_0, f_1, f_2, \ldots, f = f_m$ with $\int f_0$ being easy, e.g., constant function over ball.

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

Each ratio can be estimated by sampling:

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

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

Annealing [LV06]

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

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

- $Var\left(Y = \frac{f_{i+1}(X)}{f_i(X)}\right) < 8\, E(Y)^2.$

- $m \sim \sqrt{n}\log(2R/\epsilon)$
- #samples = $O^*(n)$.

- Although expectation of Y can be large (exponential even), need only a few samples to estimate it!

Volume Computation: an ongoing adventure

	Power	New aspects
Dyer-Frieze-Kannan 89	23	everything
Lovász-Simonovits 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
Cousins-V. 13	3	Gaussian volume

Optimization via Annealing

- We can minimize a quasiconvex function f over a 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

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

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

- Parallelizable: Cloud implementation by Mahoney et al [2013].

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).

Part 3. Sampling Algorithms

Ball walk:

At x,

-pick random y from $x + \delta B_n$

-if y is in K, go to y

Hit-and-Run:

At x,

-pick a random chord L through x

-go to a 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" is:

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

For any subset A, we have $\Phi(A) = \Phi(K\backslash \mathrm{A})$

Conductance:

$$\phi(A) = \frac{\int_A P_u(K\backslash \mathrm{A})\,dQ(u)}{\min Q(A),\ Q(K\backslash \mathrm{A})} \qquad \phi = \inf_A \phi(A)$$

Rate of convergence is bounded by $\frac{1}{\phi^2}$ [LS93, JS86].

Conductance

Arbitrary measurable subset S.

How large is the conditional escape probability from S?

Local conductance can be arbitrarily small for the ball walk.

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

How to bound conductance?

$$B_n \subseteq K \subseteq DB_n$$

- Average local conductance is high $\sim \frac{1}{\sqrt{n}}$
- Conductance of large subsets is high $\sim \frac{1}{nD}$
- From a warm start, the ball walk mixes in $O^*(n^2D^2)$ steps.
- Q. Can this be improved? Apparently not:

Conductance

Need:

- Nearby points have overlapping one-step distributions
- Large subsets have large boundaries [isoperimetry]

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

Isoperimetry and the KLS conjecture

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

A: covariance matrix of stationary distribution π

$$E_\pi(||x - \bar{x}||^2) = Tr(A) = \sum_i \lambda_i(A)$$

Thm. [KLS95]. $\pi(S_3) \geq \frac{c}{\sqrt{Tr(A)}} d(S_1, S_2) \min \pi(S_1), \pi(S_2)$

Conj. [KLS95]. $\pi(S_3) \geq \frac{c}{\sqrt{\lambda_1(A)}} d(S_1, S_2) \min \pi(S_1), \pi(S_2)$

KLS, Slicing, Thin-shell

	current bound	
thin shell	$n^{1/3}$	[Guedon-Milman]
slicing	$n^{1/4}$	[Bourgain, Klartag]
KLS	$\sim n^{1/3}$	[Bobkov; Eldan-Klartag]

All are conjectured to be O(1).

Conjectures are equivalent! [Ball, Eldan-Klartag].

Convergence

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

Thm. [LV03]. Same holds for arbitary logconcave density functions.

- 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

Min distance isoperimetry
is too coarse

Average distance isoperimetry

- How to average distance?
- $h(x) \leq min\ \{d(u,v) :\ u \in S_1, v \in S_2, x \in \ell(x,y)\}$

Thm.[LV04; Dieker-V.13]

$$\pi(S_3) \geq E\big(h(x)\big)\pi(S_1)\pi(S_2)$$

Hit-and-run

- Thm [LV04]. Hit-and-run mixes in polynomial time from *any* starting point inside a convex body.
- Conductance = $\Omega\left(\frac{1}{nD}\right)$
- Along with isotropic transformation, gives $O^*(n^3)$ sampling, $O^*(n^4)$ volume.
- Is this practical?
- [Deak-Lovász 2012] Implementation of [LV03]
- works for cubes of dimension up to 9 (taking a couple of hours).

Sampling: current status

Can be sampled efficiently:

- Convex bodies
- Logconcave distributions
- 1/(n-1)-harmonic-concave distributions [Chandarasekaran-Deshpande-V.09]
- Near-logconcave distributions [AK91]
- Star-shaped bodies [Dadush-Chandrasekaran-V.10]
- Positive curvature manifolds (e.g., boundary of convex body) [Belkin-Narayanan-Niyogi; Dieker-V.13]
- …

Cannot be sampled efficiently:

- Quasiconcave distributions

Gaussian sampling/volume

- Sample from Gaussian restricted to K
- Compute Gaussian measure of K

- Use annealing:
- Define $f_i(X) = e^{-\frac{||X||^2}{2\sigma_i^2}}$

- Start with σ_0 small $\sim \frac{1}{\sqrt{n}}$, increase in phases till 1.

- To compute Gaussian volume, compute ratios of integrals of consecutive phases.
- $\frac{\int f_{i+1}}{\int f_i}$

Gaussian sampling

- KLS conjecture holds for Gaussian restricted to any convex body (via Brascamp-Lieb inequality).

Thm. $\pi(S_3) \geq \frac{c}{\sigma}\, d(S_1, S_2) \min \pi(S_1), \pi(S_2)$

- Not enough on its own, but can be used to show:

Thm. [Cousins-V. 13]. Ball walk applied to Gaussian restricted to a convex body containing the unit ball mixes in $O^*(n^2)$ time from a warm start.

Speedy walk: a thought experiment

- Take sequence of points visited by ball walk:

$$w_0, {\color{red}w_1}, w_2, {\color{red}w_3}, \ldots, w_i, w_{i+1}, {\color{red}w_{i+3}} \ldots$$

- Subsequence of "proper" attempts that stay inside K

- This subsequence is a Markov chain and is rapidly mixing from any point

- For a warm start, the total number of steps is only a constant factor higher

Gaussian volume

- Theorem [Cousins-V.] The Gaussian volume of a convex body K containing the unit ball can be approximated to within relative error ϵ in time $O^*(n^3)$.

- No need to adjust for isotropy!
- Each step samples a 1-d Gaussian from an interval

- Is this practical?!
- Note: this is number of oracle calls.

Practical volume/integration?

Start with a concentrated Gaussian
Run the algorithm till the Gaussian is nearly flat

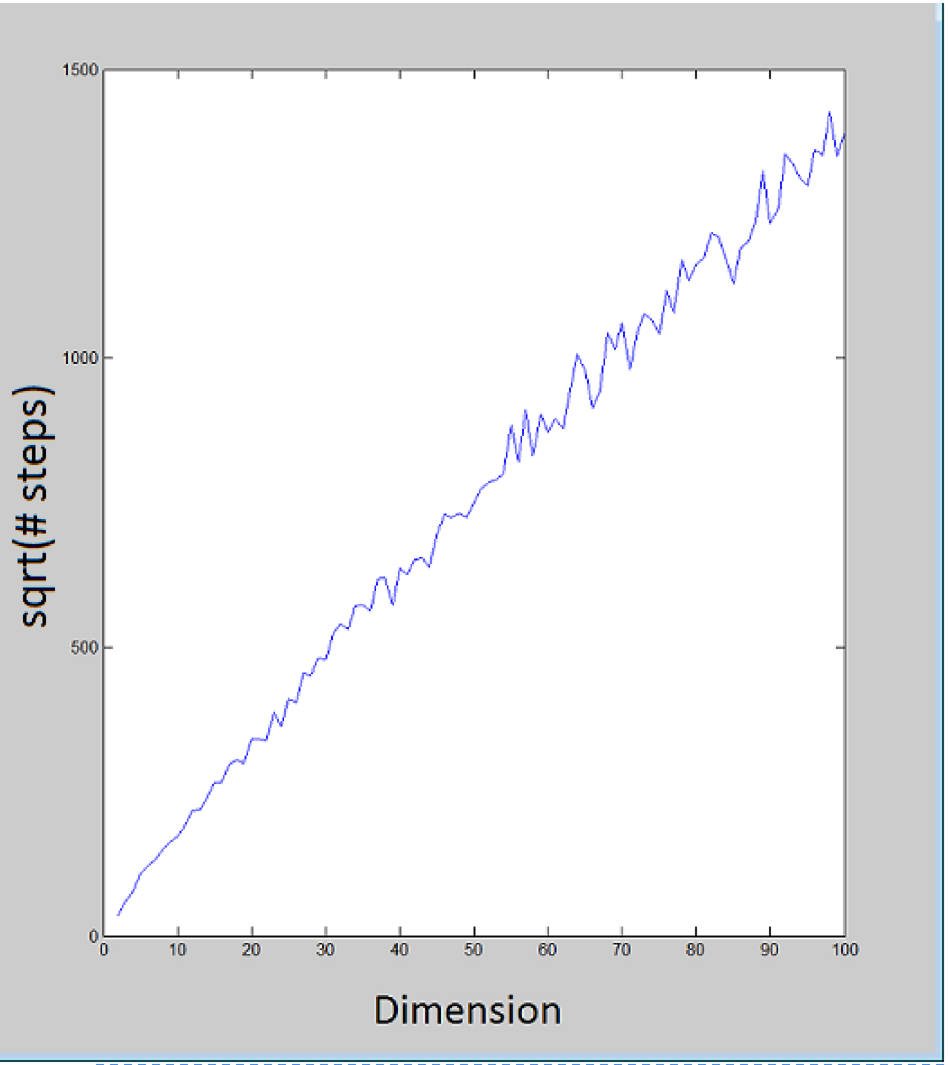
In each phase, flatten Gaussian as much as possible while keeping variance of ratio of integrals bounded

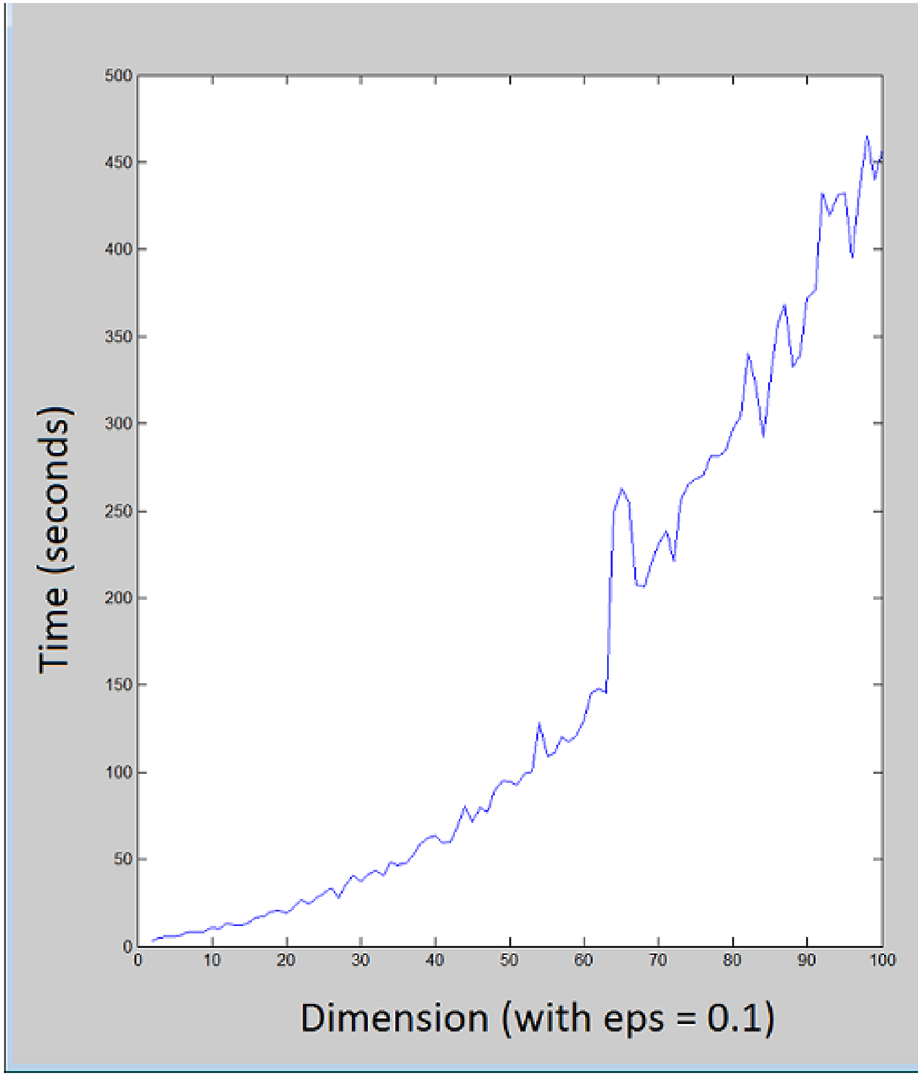
Variance can be estimated with a small constant number of samples

If covariance is skewed (as seen by SVD of O(n) points), scale down high variance subspace

“Adaptive” annealing

Rotated cubes





demo

3 problems to understand better

- How true is the KLS conjecture?
- How efficiently can we learn a polytope from Gaussian or uniform random points?
- Can we compute the volume of a polytope in deterministic polynomial time?

One core algorithmic problem

sampling

Thank you.