The Interplay of Sampling and Optimization in High Dimension

Santosh Vempala, Georgia Tech

Deep Learning

- Neural nets can represent any real-valued function approximately.
- What can they learn efficiently?
- I don't know.

High-dimensional problems

- 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.
- Two fundamental problems: Optimization, Sampling

Problem 1: Optimization

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

point x, error parameter ε .

Output: point y such that

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

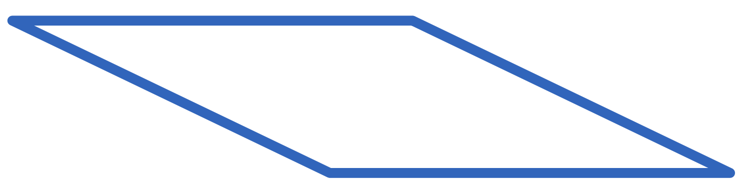
Examples: $\max c \cdot x\ s.t. Ax \geq b$, $\min||x||\ s.t. x \in K$.

Problem 2: Sampling

Input: function f: $R^n \to R_+$, $\int f < \infty$, specified by an oracle, a point x, error parameter ε.

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

Examples: $f(x) = 1_K(x)$, $f(x) = e^{-a||x||} 1_K(x)$



High-dimensional problems

- Optimization
- Sampling

Also:

- Integration (volume)
- Learning
- Rounding
- …

All intractable in general, even to approximate.

Q. What structure makes high-dimensional problems computationally tractable? (polytime solvable)

High-dimensional breatkthroughs

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

Ellipsoid algorithm 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.

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:

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

Tractable cases: cONVEx++

Problem Input type	Optimization $min/max\ f(x): x \in K$	Sampling $x \sim f(x)$
Convex	Linear: Kha79 Convex: GLS83	DFK89
Logconcave	GLS83(implicit)	Lipshitz: AK91 General: LV03
Stochastic/approx. convex	BV03,KV03,LV06, BLNR15, FGV15	KV03, LV06
Star-shaped	X	CDV2010

Efficient sampling gives efficient volume computation, integration for logconcave functions.

Computational model

Well-guaranteed Membership oracle:

Compact set K is given by

- a membership oracle: answers YES/NO to "$x \in K?$"
- a point $x_0 \in K$
- Numbers r, R s.t. $x_0 + rB^n \subseteq K \subseteq RB^n$

Well-guaranteed Function oracle

- An oracle that returns $f(x)$ for any $x \in R^n$
- A point x_0 with $f(x_0) \geq \beta$
- Numbers r, R s.t.

$$x_0 + rB^n \subset L_f\left(\frac{1}{8}\right) \text{ and } R^2 = E_f\left(\left\|X - \bar{X}\right\|^2\right)$$

A closer look: polytime convex optimization

Linear programs	Khachiyan79	Ellipsoid
	Karmarkar83	Interior-point method (IPM)
	Renegar88, Vaidya90	IPM
	Dunagan-V04	Perceptron+rescaling
	Kelner-Spielman05	Simplex+rescaling
	….	
	Lee-Sidford13,15	IPM
Convex programs	GLS83*	Ellipsoid++
	Vaidya89*	Volumetric center
	NN94**	IPM (universal barrier)
	Bertsimas-V.02*	Random walk+cutting plane
	Kalai-V06*	Simulated annealing
	Lee-Sidford-Wong15#	hybrid:IPM+cutting plane

*: need only a membership/function oracle!

**: yes, even this.

#: needs separation oracle: NO answer comes with violating hyperplane

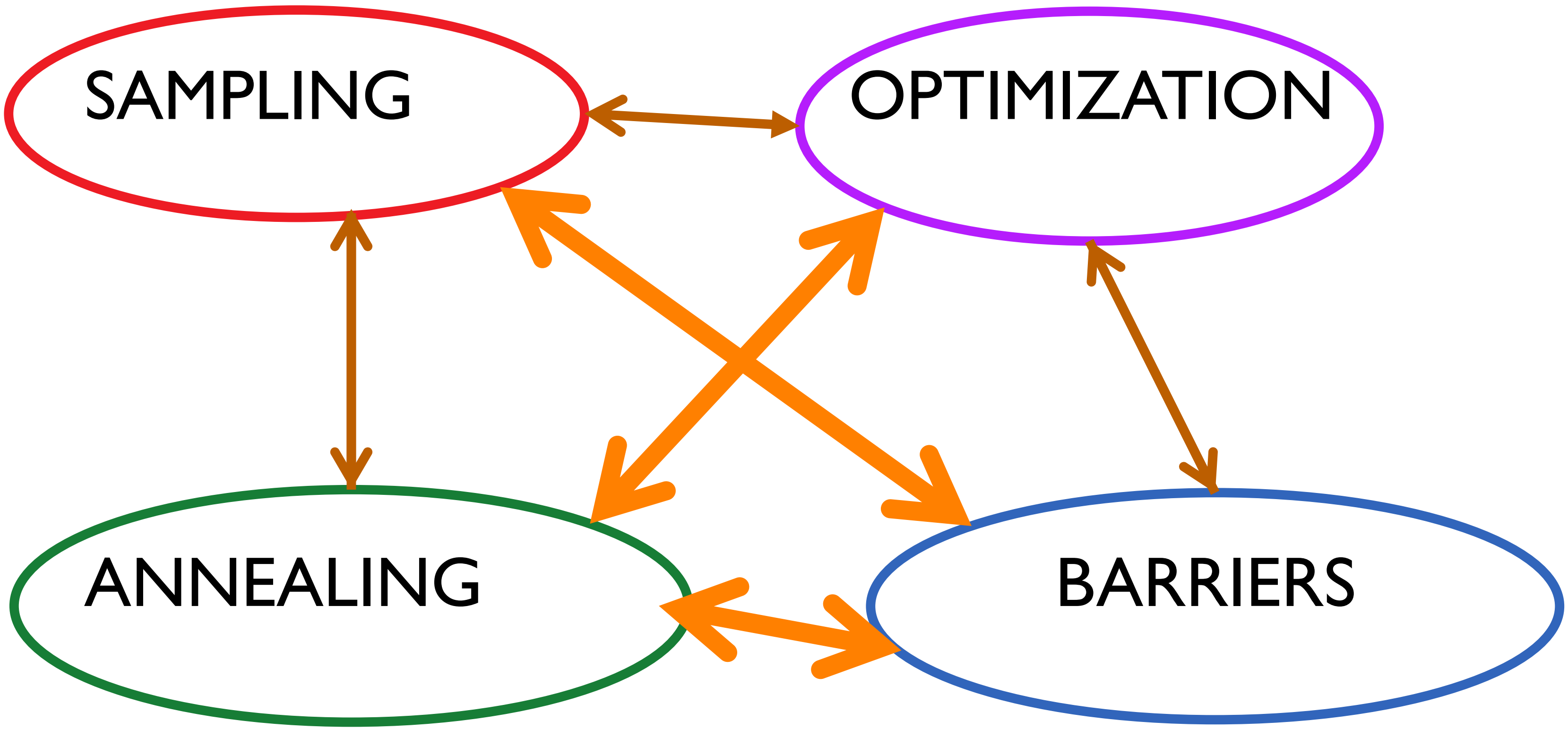
A closer look: polytime sampling

Convex bodies	DFK89	grid walk	n^{20}	
	LS91,93	ball walk	n^7	
	KLS97	ball walk	n^3	
	LV03	hit-and-run		
Logconcave functions	AK91	grid walk	n^8	
	LS93	ball walk	n^7	
	LV03,06	hit-and-run	n^3	
Polytopes	KLS,LV	Ball/hit-and-run	n^3	mn
	KN09	Dikin walk	mn	$mn^{\omega-1}$
	Lee-V16+	Geodesic walk	mn^c	$mn^{\omega-1}$

This talk: interplay of sampling and optimization

- Part I: Optimization via Sampling
  - Cutting Plane Method
  - Simulated Annealing
  - Interior Point Method
- Part II: Sampling via Optimization (ideas)
  - Ball walk, hit-and-run
  - Dikin walk
  - Geodesic walk
  - Faster polytope sampling
- Part III: Open questions

You want interplay?



Convex feasibility from separation

Input: Separation oracle for a convex body K, r, R.

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)).

Q. Which sequence of points to query?

Each query either solves the problem or restricts the remaining set via the violating linear inequality.

(Convex optimization → Convex Feasibility via binary search)

Centroid cutting-plane algorithm

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

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

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

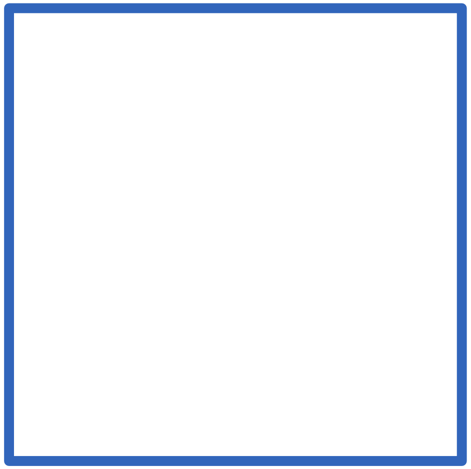
- #queries = O(nlog(R/r)).
- Best possible.

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

Randomized cutting plane

[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.



Optimization via Sampling

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

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

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).

Optimization from membership

Sampling suggests a conceptually very simple algorithm.

Simulated Annealing [Kalai-V.04]

To optimize f consider a sequence $f_0, f_1, f_2, \dots,$
with f_i more and more concentrated near the optimum.

$$f_i(x) = e^{-t_i \langle c,x \rangle}$$

Corresponding distributions:

$$P_{t_i}(x) = \frac{e^{-t_i \langle c,x \rangle}}{\int_K e^{-t_i \langle c,x \rangle} dx}$$

Lemma. $E_{P_t}(c \cdot x) \le \min c \cdot x + \frac{n}{t}.$

So going up to $t = \frac{n}{\epsilon}$ suffices to obtain an ϵ approximation.

Simulated Annealing [Kalai-V.04]

- $P_{t_i}(x) = \frac{e^{-t_i\langle c,x\rangle}}{\int_K e^{-t_i\langle c,x\rangle}dx}$

Lemma. $E_{P_t}(c \cdot x) \le \min c \cdot x + \frac{n}{t}.$

Proof: First reduce to cone, only makes it worse.

For cone, $E(c \cdot x) = \frac{\int y e^{-ty} y^{n-1} dy}{\int e^{-ty} y^{n-1} dy} = \frac{\frac{n!}{t^{n+1}}}{\frac{(n-1)!}{t^n}} = \frac{n}{t}.$

Optimization via Simulated Annealing

- $\min \langle c, x\rangle = \max f(x) = e^{-\langle c,x\rangle}$
- Can replace $\langle c, x\rangle$ with any convex function

For $i = 1, \ldots, m$:

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

Output $\arg\max f(X)$.

Complexity of Sampling

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

$$M = \sup \frac{Q_0(S)}{Q(S)} \quad or \quad M = E_{Q_0}\left(\frac{Q_0(x)}{Q(x)}\right)$$

Thm. [LV03]. Same holds for arbitrary logconcave density functions. Complexity is $O^*(M^2 n^2 R^2)$.

Isotropic transformation: $R = O(\sqrt{n})$; Warm start: $M = O(1)$.

Simulated Annealing

To optimize $f = \langle c, x \rangle$, take a sequence $t_0, t_1, t_2, \ldots$, with corresponding distributions:

$$P_i(x) = \frac{e^{-t_i\langle c,x\rangle}}{\int_K e^{-t_i\langle c,x\rangle} dx}$$

For sampling to be efficient, consecutive distributions must

- have significant overlap in L_2 distance:

$$\|P_i/P_{i+1}\|_2 = E_{P_i}\left(\frac{P_i(x)}{P_{i+1}(x)}\right) = O(1)$$

- Maintain near-isotropy, which is implied by $\|P_{i+1}/P_i\|_2 = O(1)$.

Lemma. For $t_{i+1} = \left(1 + \frac{1}{\sqrt{n}}\right) t_i$, $\|P_i/P_{i+1}\|_2, \|P_{i+1}/P_i\|_2 < 5$.

- Hence, $\sim \sqrt{n}$ phases. Best possible even if we assume only (a) logconcave distributions and (b) overlap in TV distance [KV03].

Interplay of sampling and optimization

- Part I: Optimization via Sampling
  - Cutting Plane Method
  - Simulated Annealing
  - Interior Point Method
- Part II: Sampling via Optimization (ideas)
  - Ball walk, hit-and-run
  - Dikin walk
  - Geodesic walk
  - Faster polytope sampling
- Part III: Open questions

Interior point method

- $\psi_i(x) = \min_K t_i \cdot f(x) + \phi(x)$
- $f(x)$: objective function, $f(x) = \langle c, x \rangle$ for LP.
- $\phi(x)$: is smooth convex function "barrier", blows up on the boundary of K.
- $t_0 = 1, t_{i+1} = \left(1 + \frac{1}{\sqrt{\nu}}\right) t_i$
- Step: $x_i \approx \arg\min \psi_i(x)$

Q1. How to implement a step?
A1. Newton method

Q2. Number of steps?
A2. $\tilde{O}\left(\sqrt{\nu}\right)$ for ν-self-concordant ϕ.

- Barriers mollify boundary effects.

Interior point method

- $\psi_i(x) = \min_K t_i \cdot f(x) + \phi(x)$

- Log barrier for polytopes has $\nu \leq m$

$$\phi(x) = -\sum_{i=1}^{m} \log(A_i \cdot x - b_i)$$

- Universal barrier (Nemirovski-Nesterov94) has $\nu = O(n)$:

$$\phi(x) = \log vol(K - x)^o$$

- Entropic barrier (Bubeck-Eldan) has $\nu = \big(1 + o(1)\big)n$

$$\phi(x) = \sup_{\mathbf{R}^n} -\theta \cdot x - \log \int_K e^{-\theta \cdot x} dx$$

- Canonical barrier (Hildebrand) has $\nu \leq n + 1$

IPM complexity

- $O(\sqrt{\nu}) = O(\sqrt{n})$ for general convex bodies via Universal, Entropic or Canonical barriers.
- The Universal and Entropic barriers can be computed via sampling and integration over convex bodies and logconcave densities.
- But there is an even closer connection!

Annealing | IPM

- Entropic barrier IPM is equivalent to Simulated Annealing

$$E_{P_t}(x) = \arg\min t\langle c, x\rangle + \phi(x)$$

- Desired step of IPM is computed exactly by Simulated Annealing in one phase.

Thm[AH16]. Simulated annealing path = IPM Central path

- Two very different analyses of the same algorithm: one with calculus (self-concordance), the other with probability.

This talk: interplay of sampling and optimization

- Part I: Optimization via Sampling
  - Cutting Plane Method
  - Simulated Annealing
  - Interior Point Method
- Part II: Sampling via Optimization (ideas)
  - Ball walk, hit-and-run
  - Dikin walk
  - Geodesic walk
  - Faster polytope sampling
- Part III: Open questions

How to Sample?
Ball walk

At x, pick random y from $x + \delta B_n$

if y is in K, go to y

Boundary Effect #1:

Step size cannot be too large,

else rejection probability becomes too high!

Hit-and-run

[Boneh, Smith]

At x, pick a random chord L through x

go to a uniform random point y on L

Boundary Effect #1:

Average chord length is small!

Larger steps?

- Step size for both walks, even in isotropic position, is $O\left(\frac{1}{\sqrt{n}}\right)$ leading to mixing in $\theta\left(n^2\right)$ steps.
- This constraint is due to points near the boundary
- Idea: take larger steps for points in the interior
- Boundary Effect #2: Then stationary distribution is not uniform.

Dikin walk with barrier ϕ

- At x, pick next step from Ellipsoid defined by $\phi(x)$:
$$Ell(x) = \left\{y: \|y - x\|_{\nabla^2\phi(x)} = (y-x)^T \nabla^2\phi(x)(y-x) \le 1\right\}.$$
- Alternatively, $y \sim N\left(x, \frac{1}{n}\left(\nabla^2\phi(x)\right)^{-1}\right)$
- Apply Metropolis filter to make steady state uniform

Log barrier: $\phi(x) = -\sum_{i=1}^{m} \log(Ax-b)_i$ and $\nabla^2\phi(x) = A^T S_x^{-2} A$

Thm. [Kannan-Narayanan09] For any polytope in R^n with m facets, the Dikin walk with the log barrier mixes in $O^*(mn)$ steps from a warm start. Each step takes $O(mn^{w-1})$ time.

- Faster than hit-and-run when not too many constraints.

Dikin walk

- Dikin ellipsoid is fully contained in K
- Idea: Pick next step y from a blown-up Dikin ellipsoid. Can afford to blow up by $\sim \sqrt{n/\log m}$. WHP $y \in K$.

- But, rejection probability of filter becomes too high!

Markov chains

- State space K , next step distribution $P_u(.)$ associated with each point u in K.
- Stationary distribution Q, ergodic "flow" defined as

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

- For a stationary distribution, we have $\Phi(A) = \Phi(K\backslash \mathrm{A})$
- Conductance:

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

- Thm. [LS93] Q_t: distribution after t steps

$$M = \sup_{A\subset K} \frac{Q_0(A)}{Q(A)}: \quad d_{TV}(Q_t, Q) \le \sqrt{M}\left(1 - \frac{\phi^2}{2}\right)^t$$

$$M = E_{Q_0}\left(\frac{Q_0(x)}{Q(x)}\right): \ d_{TV}(Q_t, Q) \le \epsilon + \sqrt{\frac{M}{\epsilon}\left(1 - \frac{\phi^2}{2}\right)^t} \ \ \forall \epsilon > 0$$

Conductance

Consider an arbitrary measurable subset S.

Need to show that the escape probability from S is large.

- (Smoothness of 1-step distribution) Points that do not cross over are far from each other i.e., nearby points have large overlap in 1-step distributions
- (Isoperimetry) Large subsets have large boundaries

Isoperimetry

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

$$R^2 = E_\pi\left(\|x - \bar{x}\|^2\right)$$

A $= E((x - \overline{x})(x - \bar{x})^T)$: covariance matrix of π

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

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

Convergence of ball walk

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?

Convergence of hit-and-run

Cross-ratio distance:

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

Thm. [L98;LV04] $\pi_f(S_3) \geq d_K(S_1, S_2)\pi_f(S_1)\pi_f(S_2)$

Conductance = $\Omega\left(\frac{1}{nD}\right)$

Thm [LV04]. Hit-and-run mixes in polynomial time from *any* starting point inside a convex body.

Leads to $O^*(n^3)$ mixing.

KLS hyperplane conjecture

$$A = E(xx^T)$$

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

- Could improve sampling complexity by a factor of n
- Implies well-known conjectures in convex geometry: slicing conjecture and thin-shell conjecture
- [CV13] True for the product of a logconcave function and a Gaussian.
- Best case scenario: $O(n^2)$.

This talk: interplay of sampling and optimization

- Part I: Optimization via Sampling
  - Cutting Plane Method
  - Simulated Annealing
  - Interior Point Method
- Part II: Sampling via Optimization (ideas)
  - Ball walk, hit-and-run
  - Dikin walk
  - Geodesic walk
  - Faster polytope sampling
- Part III: Open questions

Can we sample faster?

- Brownian motion SDE:

$$dx_t = \mu(x_t, t)dt + \sqrt{A(x_t, t)}dW_t$$

- Roughly speaking, next step is from infinitesimal Gaussian $N(\mu, A(x_t, t))$

- Each point $x \in K$ has its own local scaling (metric)

Thm. [Fokker-Planck] Diffusion equation of SDE is $\frac{\partial}{\partial t}p(x,t) = \frac{1}{2}\nabla \cdot \left(A(x,t)\nabla p(x,t)\right)$
i.e.,

$$\frac{\partial}{\partial t}p(x,t) = -\sum_{i}^{n}\frac{\partial}{\partial x_i}[\mu(x,t)p(x,t)] + \frac{1}{2}\sum_{i}^{n}\sum_{j}^{n}\frac{\partial^2}{\partial x_i \partial x_j}[A_{ij}(x,t)p(x,t)]$$

- For any metric, SDE gives diffusion equation.
- When $A = I$, this is the heat equation: $\frac{\partial}{\partial t}p(x,t) = \frac{1}{2}\nabla^2 p(x,t)$.

Which metric to use?

- Natural choice: metric defined by Hessian of a barrier function.
- Why? It was useful in optimization to move quickly in the polytope, by avoiding boundaries.

- For barrier ϕ, diffusion equation is

$$\frac{\partial}{\partial t} p(x,t) = \frac{1}{2} \nabla \cdot \left(\nabla^2 \phi(x)\right)^{-1} \nabla p(x,t)$$

- and the SDE is

$$dx_t = \mu(x_t)dt + \left(\nabla^2 \phi(x_t)\right)^{-1/2} dW_t$$

with $\mu(x_t) = -\frac{1}{2}\left(\nabla^2 \phi(x)\right)^{-1} \nabla \log \det \nabla^2 \phi(x),$ a Newton step!

- But how to simulate Brownian motion? Studied in practice, but not much from a complexity perspective.

Discretization: which coordinate system?

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t$$

- Euler-Maruyama: For small h,

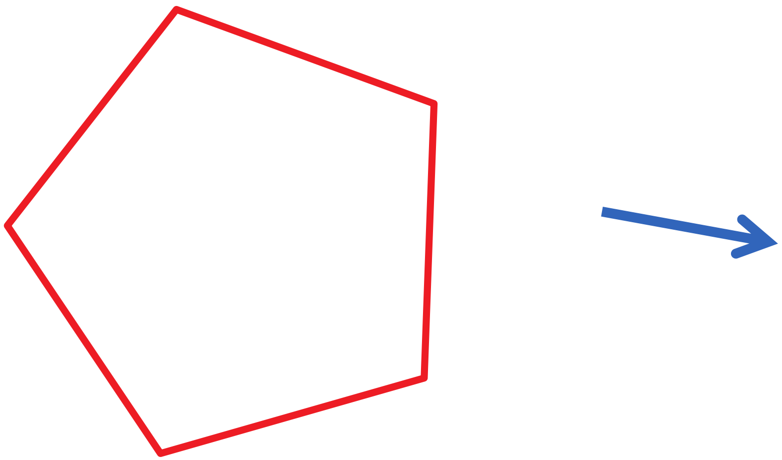
$$x_{t+h} = x_t + \mu(x_t)h + \sigma(x_t)w_t\sqrt{h}$$

- Not so great:
  - Euclidean coordinates are arbitrary
  - Approximation is weaker if local metric $\sigma(x_t)$ varies a lot

- Is there a more natural coordinate system? That effectively keeps local metric constant?

Enter Riemannian manifolds

- n-dimensional manifold M is an n-dimensional surface in R^k for some $k > n$.
- We are going to map the polytope to a manifold.



- Idea: distances will be shortest paths on manifold
- Examples: flight paths on earth, light in relativity

Enter Riemannian manifolds

- n-dimensional manifold M is an n-dimensional surface.
- Each point p has a linear tangent space T_pM of dimension n, the local linear approximation of M at p. Tangents of curves in M lie in T_pM.
- The inner product in T_pM depends on p: $\langle u, v \rangle_p$

Enter Riemannian manifolds

- Each point p has a linear tangent space T_pM.
- The inner product in T_pM depends on p: $\langle u, v \rangle_p$
- Length of a curve $c\colon [0,1] \to M$ is

$$L(c) = \int_0^1 \left\| \frac{d}{dt} c(t) \right\|_{c(t)} dt$$

- Distance between x,y in M is the infimum over all paths in M between x and y. This is the Riemannian metric.

Riemannian manifold/metric

- The inner product in T_pM depends on p: $\langle u, v \rangle_p$
- Length of a curve c: $[0,1] \to M$ is
  - $L(c) = \int_0^1 \left\| \frac{d}{dt} c(t) \right\|_{c(t)} dt$
- Metric: $d_M(x, y) = \inf L(path(x, y))$
- Geodesic: curve γ: $[a, b] \to M$ that has
  - constant speed: $\left\| \frac{d}{dt} \gamma(t) \right\|_{\gamma(t)}$ is a constant
  - is a locally shortest path

Riemannian manifold/metric

- Geodesic: curve $\gamma: [a, b] \to M$ that has
  - constant speed: $\left\| \frac{d}{dt}\gamma(t) \right\|_{\gamma(t)}$ is a constant
  - is a locally shortest path
- Exponential map $\exp_p: T_pM \to M$ is defined as
  - $\exp_p(v) = \gamma_v(1)$,
  - γ_v: unique geodesic from p with initial velocity v.

- Locally, $\exp_p^{-1}(x)$ gives x in "normal" coordinates at p.

Hessian manifold

- Local inner product is defined by Hessian:
  - $\langle u, v\rangle_p = u^T \nabla^2 \phi(p) v$
  - $||v||_p = ||v||_{\nabla^2 \phi(p)}$

Lemma. Let $F = \exp_{x_0}^{-1}$, where $\exp_x : T_x M \to M$. Then the SDE

$$dx_t = \mu(x_t)dt + \sigma(x_t)dW_t$$

with $\sigma(x) = \left(\nabla^2 \phi(x)\right)^{-1/2}$ implies

$$dF(x_0) = \frac{1}{2}\mu(x_0)dt + \sigma(x_0)dW_0.$$

Ergo: Geodesic walk

In tangent plane at x,

1. pick $w \sim N_x(0, I)$, i.e. mean standar Gassian in $\|.\|_x$
2. Compute $y = \exp_x\left(\frac{h}{2}\mu(x) + \sqrt{h}w\right)$
3. Compute w' s.t. $x = \exp_y\left(\frac{h}{2}\mu(y) + \sqrt{h}w'\right)$
4. Accept with probability Min $\left(1, \frac{p(y \overset{w}{\rightarrow} x)}{p(x \overset{w'}{\rightarrow} y)}\right)$

How to compute geodesic and rejection probability?

Implementing one step

- Geodesic equation for $\phi(x) = -\sum_{i=1}^{m} \log(Ax - b)_i$
  - Let $S_x = Diag(Ax - b),\ A_x = S_x^{-1} A$

$$\gamma'' = \left(A_\gamma^T A_\gamma\right)^{-1} A_\gamma^T \left(A_\gamma \gamma'\right)^2$$

- And the probability $p(x \rightarrow y)$
- are both second-order ODEs.
- We show this can be solved efficiently to inverse polynomial accuracy by the Collocation Method, for any Hessian manifold under smoothness assumptions.
- $\tilde{O}(mn^{\omega-1})$ per step for log barrier

Mixing of Geodesic walk

Thm 1. [Lee-V16] For log barrier, Geodesic walk mixes in $\tilde{O}(mn^{0.75})$ steps.

This is a special case of more general theorem:

Thm 2. [Lee-V16] For any Hessian manifold, for small enough step size h, Geodesic walk mixes in $O\left(\frac{G^2}{h}\right)$ steps, where G is the expansion of metric wrt to Hilbert metric.

Q. How large can the step size be?

7-parameter mixing theorem

Convergence for general Hessian manifolds is based on:

- $D_0 = \sup||\mu(x)||_x$: maximum norm of drift
- $D_1 = \sup\frac{d}{dt}\left||\mu(\gamma(t))|\right|^2_{\gamma(t)}$: smoothness of drift norm
- $D_2 = \sup||\nabla_s\mu(x)||_x$: smoothness of drift
- $G_1 = \sup\left|\left(\log\det g(\gamma(t))\right)'''\right|$: smoothness of volume element of local metric g.
- $G_2 = \sup\frac{d(x,y)}{d_H(x,y)}$: smoothness of metric (d_H: Hilbert dist)
- R_1 : stability of Jacobi field, R_2 : smoothness of Ricci curvature

Thm. Suppose $h \leq c\min\frac{1}{(nD_0R_1)^{\frac{2}{3}}},\frac{1}{D_2},\frac{1}{nR_1},\frac{1}{n^{\frac{1}{3}}D_1^{\frac{2}{3}}},\frac{1}{nG_1^{\frac{2}{3}}},\frac{1}{(nR_2)^{\frac{2}{3}}\}}.$

Then, the geodesic walk has conductance $\Omega\left(\frac{\sqrt{h}}{G_2}\right)$ and mixes in $O\left(\frac{G_2^2}{h}\right)$ steps.

7-parameter mixing for log barrier

Convergence for general Hessian manifolds is based on:

- $D_0 = \sup \left|\left|\mu(x)\right|\right|_x$: maximum norm of drift $O(\sqrt{n})$
- $D_1 = \sup \frac{d}{dt} \left|\left|\mu(\gamma(t))\right|\right|_{\gamma(t)}^2$: smoothness of drift norm $O(n\sqrt{h})$
- $D_2 = \sup \left|\left|\nabla_s \mu(x)\right|\right|_x$: smoothness of drift $O(\sqrt{n})$
- $G_1 = \sup \left|\left(\log \det g(\gamma(t))\right)'''\right|$: smoothness of volume element of local metric g $O(\sqrt{h})$
- $G_2 = \sup \frac{d(x,y)}{d_H(x,y)}$: smoothness of metric (d_H: Hilbert dist) $O(\sqrt{\nu}) = O(\sqrt{m})$
- R_1 : stability of Jacobi field $O\left(\frac{1}{\sqrt{n}}\right)$, R_2 : smoothness of Ricci curvature $O(\sqrt{nh})$

Thm (log barrier). Suppose $h \le c \min \frac{1}{(nD_0R_1)^{\frac{2}{3}}}, \frac{1}{D_2}, \frac{1}{nR_1}, \frac{1}{n^{\frac{1}{3}} D_1^{\frac{2}{3}}}, \frac{1}{nG_1^{\frac{2}{3}}}, \frac{1}{(nR_2)^{\frac{2}{3}}\}}$. $O(n^{-0.75})$

Then, the geodesic walk has conductance $\Omega\left(\frac{\sqrt{h}}{G_2}\right)$ and mixes in $O\left(\frac{G_2^2}{h}\right)$ steps. $O(mn^{0.75})$

Proof outline

Need to show:

- Rejection probability is small
- 1-step distributions are smooth
- Isoperimetry is good: $\frac{1}{\sqrt{G_2}}$ $\left(= \frac{1}{\sqrt{m}} \text{ for log barrier}\right)$
  - Follows from isoperimetry for hit-and-run

1-step distribution

Lemma. For any $x \in M, h > 0$, the probability density of the one-step distribution (before rejection sampling) is:

$$p_x(y) = \sum_{\exp_x(v_x)=y} \det(d \exp_x(v_x))^{-1} \sqrt{\frac{\det(g(y))}{(2\pi h)^n}}\, e^{-\frac{1}{2h}\left\|v_x - \frac{h}{2}\mu(x)\right\|_x^2}.$$

- Pick $v_x \sim N_x(0, I)$,
- then apply $\exp_x(v_x)$,
- then account for new metric at y

1-step distribution

$$p_x(y) = \sum_{\exp_x(v_x)=y} \det(d\exp_x(v_x))^{-1} \sqrt{\frac{\det(g(y))}{(2\pi h)^n}}\, e^{-\frac{1}{2h}\left\|v_x - \frac{h}{2}\mu(x)\right\|_x^2}.$$

Under conditions on h,

Lemma 1. $\left|\log\left(\frac{p_x(y)}{p_y(x)}\right)\right| < \frac{1}{4}$.

- (rejection probability is small)

Lemma 2. $d_{TV}(P_x, P_y) < 0.31415926$

- (nearby points have large overlap in one-step distributions)

You can implement this!

Next steps

- Analyze Geodesic walk for Lee-Sidford barrier function
- Improve analysis further to allow larger step size

- Higher order simulation, anyone?

Open questions: Algorithms

- Faster LP/convex optimization?
- Faster optimization with a membership oracle?
- Faster sampling: geodesic walk, reflection walk, coordinate hit-and-run, …

Open questions: Geometry

- How true is the KLS conjecture?

$$\inf_{|S| \le \frac{|K|}{2}} \frac{|\partial S|}{|S|}$$

- A weaker conjecture:

$$\frac{1}{\sqrt{\sum_i \lambda_i(A)}} > \frac{1}{\left(\sum_i \lambda_i(A)^2\right)^{1/4}} = \frac{1}{\|A\|_F^{1/2}} > \frac{1}{\sqrt{\lambda_1(A)}}$$

Open questions: Geometry

$$\psi_d(S) = \inf_{\epsilon>0} \frac{vol(\{x: d(x,y) \leq \epsilon\})}{\epsilon \cdot \min vol(S), vol(K \setminus S)}$$

Generalized KLS.

Question: For any Hessian metric, is ψ_d achieved to within a constant factor by a hyperplane cut?

Open questions: Riemannian manifolds

- Learning: learn metric approximately?
- Testing:
- Modeling:
- Voronoi diagrams?

- Note: General relativity is Einstein-Kahler metric, this is induced by the canonical barrier!

Open questions: Probability

Q1. Does ball walk mix rapidly starting at a single nice point, e.g., the centroid?

Q2. When to stop? 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)

Open questions: Algorithms

- How efficiently can we learn a polytope P given only random points?

- With O(mn) points, cannot "see" structure, but enough information to estimate the polytope! Algorithms?
- For convex bodies:
  - [KOS][GR] need $2^{\Omega(\sqrt{n})}$ points to learn P
  - [Eldan] need 2^{n^c} even to estimate the volume of P

Open questions: Algorithms

- Can we estimate the volume of an explicit polytope in *deterministic* polynomial time?

$$Ax \geq b$$

Thank you!

And to:

Ravi Kannan

Laci Lovasz

Adam Kalai

Ben Cousins

Yin Tat Lee