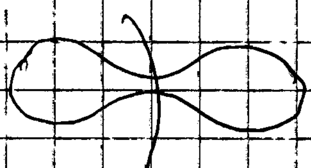
Day 4 Talk 4

Yin Tat Lee

For isotropic log-concave measure

Main result $\mu(\partial S) \gtrsim D^{-\frac{1}{2}} \mu(S) \sqrt{\ln(1/\mu(S))}$ for $\mu(S) \leq 2$

Cheeger constant $C_\mu = \min_{\mu(S) \leq \frac{1}{2}} \frac{\mu(\partial S)}{\mu(S)}$



← small Cheeger Constant

Log-Cheeger / Gaussian isoperimetric constant

$$LC_\mu = \min_{\mu(S) \leq \frac{1}{2}} \frac{\mu(\partial S)}{\mu(S)\sqrt{\ln(1/\mu(S))}}$$

- can be ∞

Correction to slides.

- if $< \infty$ ⇝ Gaussian tail
- relates to Log-Sobolev constant LC_μ

$$LS_\mu \simeq LC_\mu^2$$

Bounds for LC_μ?

- $LC_\mu \gtrsim 1$ Gaussian μ

Correction ⟶ - $LC_\mu \gtrsim \sqrt{c}$ $d\mu = e^{-c\|x\|^2} q(x) dx$

q - log-concave

Recent progress on sampling

Problem: Sample points from the uniform distribution on a given convex set K

$Ax = b$ linear eq — $n^{2.\ldots}$

$\max c^T x$ s.t. $Ax \leq b$ — $n^{2.5}$

sampling

Want to do all in n^2 time.

Sampling problem for polytopes

Best result || L-Vempala '17 Hamiltonian walk

$mn^{0.66}$ $mn^{1.38}$

Experimental by Ben: $\tilde{O}(1)$

Using a modified Brownian motion

Many applications of estimates for log-Cheeger
(see slides)

Application 1: Large deviation inequality

Correction
Formula: $\frac{d\alpha_t}{dt} = \frac{d\mu(\partial E_t)}{d(E_t)} = \frac{\sqrt{\alpha_t}}{\sqrt{\sqrt{n}+t}}$

Application 2: Mixing time

Proof ideas: log-Cheeger constant

why? $p(x)$ = dens of log-concave

$dp_t(x) = (x-\mu_t)^T dW_t \cdot p_t(x)$
$\rightsquigarrow$ gaussian.

- $(1+x)(1-x) \sim 1-x^2$

...

How to bound $\|A_t\|_{op}$?

By Ito's Lemma we know

Correction: $dA_t = \cdots dW_t + A^2 dt$

Open Q:

$LS_\mu = \Theta\left(P_\mu + \frac{1}{\psi_2(\mu)}\right)$

Log-Sobolev ↗ Poincare ↑ isoperimetry ↖

Convergence of Hamiltonian Monte Carlo and Faster Polytope Volume Computation

Yin Tat Lee,
University of Washington

Santosh S. Vempala
Georgia Tech

Improved Log-Sobolev Constant and Recent Progress on Polytope Sampling

Yin Tat Lee,
University of Washington

Santosh S. Vempala
Georgia Tech

New Abstract

- We refined the result (on my website): For isotropic logconcave measure μ with diameter D,

$$\mu(\partial S) \gtrsim D^{-\frac{1}{2}}\mu(S)\sqrt{\ln\left(1/\mu(S)\right)} \quad \text{for } \mu(S) \leq 1/2$$

- Proof: combination of stochastic localization and $\mathrm{Tr}(uI - \mathrm{Cov}(\mu))^{-2}$ potential.

Corollaries:

- Log Sobolev constant is $\Omega(D^{-1})$ and it is tight.
- It recovers the current best bound of KLS, thin shell, slicing constant 😄.
- It implies/improves the current best bound of large deviation inequality (except small ball).

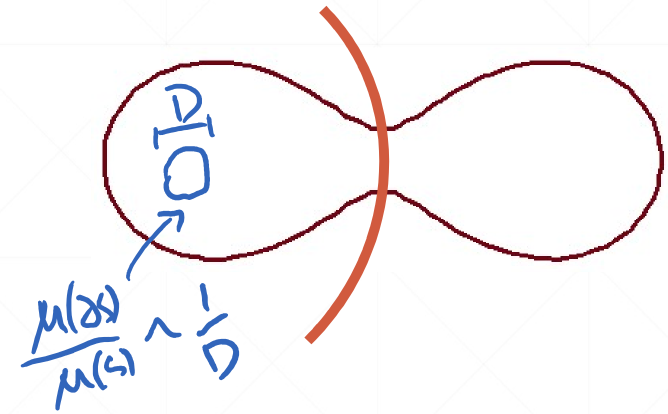
At the end:

- Review the recent progress on sampling.
- Give some related open problems.

Cheeger Constant

- For any measure μ, we define the Cheeger constant

$$\mathcal{C}_\mu = \min_{\mu(S)\leq\frac{1}{2}} \frac{\mu(\partial S)}{\mu(S)}.$$

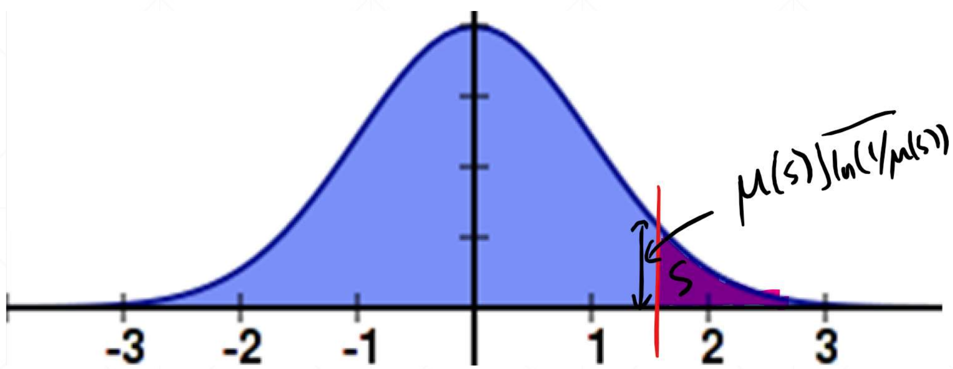


$\mathcal{C}_\mu$ is small if it is easy to cut μ into two.

- For small set S, the ratio should be $\sim 1/\mathrm{Diam}(S)$, instead of a constant.
- For Gaussian measure μ on $\mathbb{R}^n$, we have

$$\mu(\partial S) \gtrsim \mu(S)\sqrt{\ln(1/\mu(S))}$$

Why $\sqrt{\ln(1/\mu(S))}$?



Log-Cheeger Constant

- For any measure μ, we define the Log-Cheeger constant

$$\mathcal{LC}_\mu = \min_{\mu(S)\le\frac{1}{2}} \frac{\mu(\partial S)}{\mu(S)\sqrt{\ln(1/\mu(S))}}.$$

$\mathcal{LC}_\mu < \infty$ means smaller set has larger boundary ratio.

- Logconcave measure can has $\mathcal{LC}_\mu = \infty$, e.g. exponential distribution.
- It implies a Gaussian tail, e.g. $\mathbb{P}_\mu\left(|X| \ge \mathbb{E}|X| + \mathcal{LC}_\mu \cdot t\right) \lesssim e^{-t^2/2}$.
- Cheeger-like inequality **[Ledoux 94]**

$$\mathcal{LC}_\mu^2 \lesssim \mathcal{LS}_\mu \lesssim \mathcal{LC}_\mu$$

where log-Sobolev $\mathcal{LS}_\mu$ is the smallest number such that for any $\int f^2 d\mu = 1$,

$$\int |\nabla f(x)|^2 d\mu \ge \mathcal{LS}_\mu \cdot \int f^2(x) \ln^2 f(x)\, d\mu$$

Bound for Log-Cheeger Constant?

- $\mathcal{L}C_\mu \gtrsim 1$ for Gaussian measure μ.
- $\mathcal{L}C_\mu \gtrsim 1/\sqrt{c}$ for c-strongly logconcave measure μ. **[Bakry, Ledoux 96]**
  i.e. $d\mu(x) = e^{-c\|x\|^2} q(x) dx$ where q is logconcave.
- $\mathcal{L}C_\mu \gtrsim 1/D$ for log-concave measure μ with diameter D. **[Kannan-Lovász-Montenegro 05]**
- $\mathcal{L}C_\mu \gtrsim 1/D$ for non-negative curvature manifold with diameter D. **[Wang 97]**

All the results above are tight.

Our result:

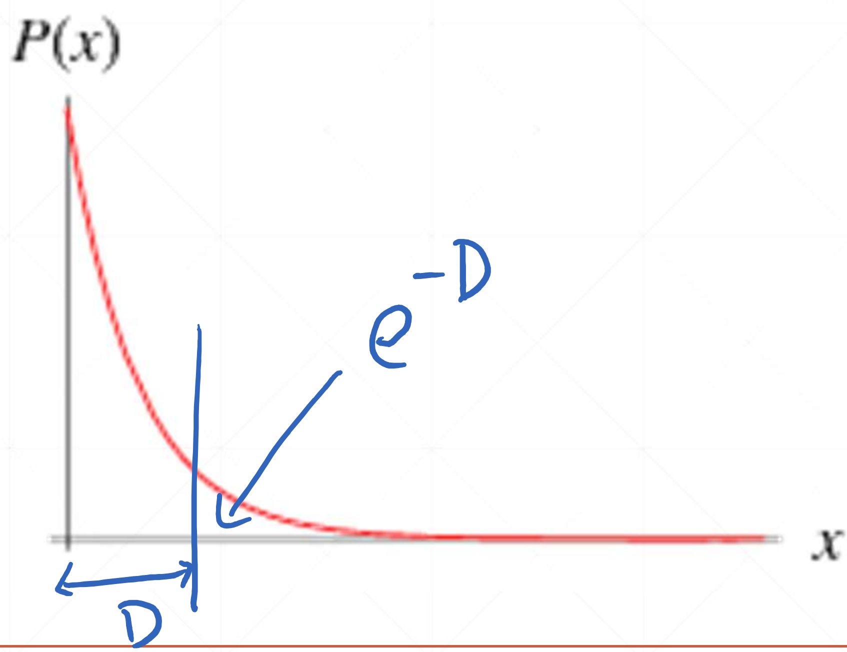
- $\mathcal{L}C_\mu \gtrsim 1/\sqrt{D}$ for **isotropic** log-concave measure μ with diameter D.

This is tight also.

Highlight: this recovers many known results and the proof is just 10-20 pages.

Why diameter?

- Works for any isotropic-logconcave by a simply truncation.
- Natural workaround for exponential distribution.



Originally,

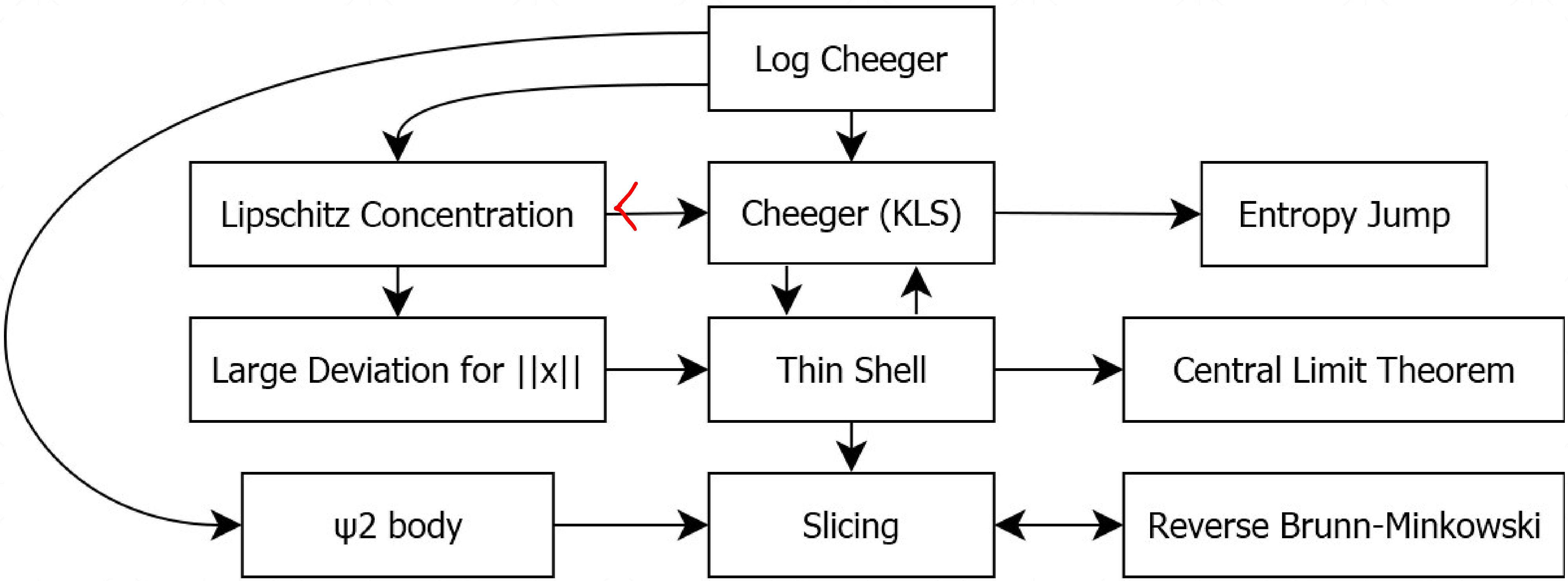
$$\mu(\partial S) \gtrsim \mu(S).$$

Now,

$$\mu(\partial S) \gtrsim \frac{1}{\sqrt{D}} \mu(S) \underbrace{\sqrt{\ln(1/\mu(S))}}_{\sqrt{D}}.$$

Applications

(Tell me if I miss some important part or arrow)



Applications 1: Large deviation inequality

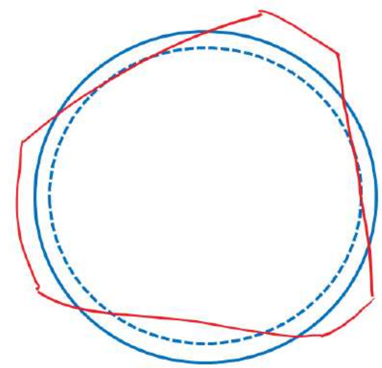
For isotropic logconcave, we have

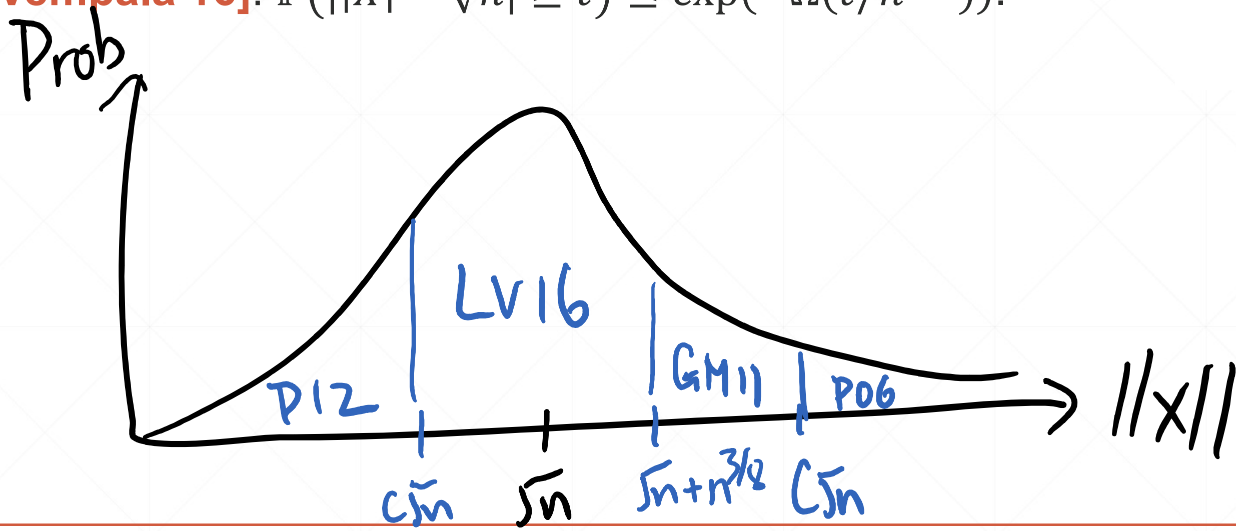
Theorem [Paouris 06]: $\mathbb{P}(|X| \geq t) \leq \exp(-\Omega(t))$ for $t \gg \sqrt{n}$. **(tight!)**

Theorem [Guedon-Milman 11]: $\mathbb{P}\left(\left||X| - \sqrt{n}\right| \geq t\right) \leq \exp\left(-\Omega\left(t^3/n\right)\right)$.

Theorem [Paouris 12]: $\mathbb{P}\left(|X| \leq t\sqrt{n}\right) \leq t^{O(\sqrt{n})}$ for $t \ll 1$.

Theorem [L-Vempala 16]: $\mathbb{P}\left(\left||X| - \sqrt{n}\right| \geq t\right) \leq \exp(-\Omega(t/n^{1/4}))$.





Applications 1: Large deviation inequality

$\mu(\partial S) \gtrsim D^{-\frac{1}{2}}\mu(S)\sqrt{\ln(1/\mu(S))}$

Theorem [L-Vempala 2017]:

For any 1-Lipschitz g and any isotropic logconcave μ, **(no diameter assumption)**

$$\mathbb{P}_\mu(|g(x) - \mathbb{E}g(x)| \geq t) \leq \exp\left(-\Omega\left(\frac{t^2}{t+\sqrt{n}}\right)\right).$$

Implies or improves previous bounds except small ball probability (still trying for small ball).

Proof: Define $E_t = \{x : \|x\| \geq \sqrt{n} + t\}$ and $\alpha_t = \ln(1/\mu(E_t))$.

By the log-Cheeger constant,

$$\frac{d\alpha_t}{dt} = \frac{\mu(\partial E_t)}{\mu(E_t)} \gtrsim \frac{1}{\sqrt{\sqrt{n}+t}}.$$

(By truncating μ outside $2\sqrt{n} + t$.)

Solving it, gives $\alpha_t \gtrsim \frac{t^2}{t+\sqrt{n}}$.

For the Lipschitz case, consider $E_t = \{x : g(x) \geq \mathbb{E}g(x) + t\}$.

Applications 2: Mixing time

Theorem [Jerrum-Sinclair 88]

For any ergodic Markov chain,

$$\Phi^{-1} \lesssim \mathcal{T} \lesssim \Phi^{-2} \ln(1/\pi_0)$$

where $\mathcal{T}$ is the mixing time, π_0 is minimum stationary probability of any state and

$$\text{conductance } \Phi = \min_{\pi(S) \leq 1/2} \frac{Q(S)}{\pi(S)} \stackrel{\text{def}}{=} \min_{\pi(S) \leq 1/2} \frac{\sum_{i \in S} \sum_{i \notin S} \pi_i \, p_{ij}}{\pi(S)}$$

Theorem [Lovász-Kannan 99]

$$\mathcal{T} \lesssim \int_{\pi_0}^{1/2} \frac{1}{x\Phi(x)^2} dx$$

where $\text{conductance } \Phi(x) = \min_{\pi(S)=x} \frac{Q(S)}{\pi(S)}$.

If $\Phi(x) \geq C \cdot \sqrt{\ln(1/x)}$, then $\mathcal{T} \lesssim C^{-2} \ln\ln(1/\pi_0)$.

[Frieze and Kannan 97] asks what is the log-Sobolev constant for isotropic logconcave.

Proof by picture!

Stochastic Localization

Proof of log-Cheeger constant

Let $p(x)$ be the density of the given logconcave distribution. Consider

$$dp_t(x) = (x - \mu_t)^T dW_t \cdot p_t(x).$$

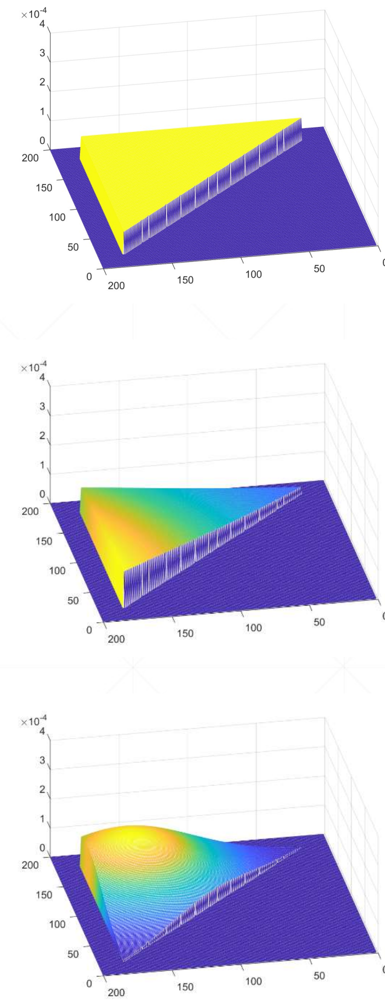
Fix a set E with $p(E) \leq 1/2$.

Goal: $p(\partial E) \gtrsim D^{-1/2} p(E) \sqrt{\ln(1/p(E))}$.

Strategy:

- $p(\partial E) = \mathbb{E}\, p_t(\partial E)$.
- $p_t(\partial E) \gtrsim t^{-1/2} p_t(E) \sqrt{\ln(1/p_t(E))}$ a.e. for any t.
- $\mathbb{E}\, p_t(E) \sqrt{\ln(1/p_t(E))} \gtrsim p(E) \sqrt{\ln(1/p(E))}$ up to $t = 1/D$.

The first one follows from p_t is a martingale.



$p_t(\partial E) \gtrsim t^{-1/2} p_t(E)\sqrt{\ln(1/p_t(E))}$

$$dp_t(x) = (x - \mu_t)^T dW_t \cdot p_t(x).$$

Ito's lemma shows

$$p_t(x) = Z_t^{-1} \exp(c_t^T x - \frac{t}{2}\|x\|^2)$$

Namely, p_t is strongly logconcave.

Theorem [Bakry, Ledoux 96]: For any $p(x) \stackrel{\text{def}}{=} q(x)\exp(-\frac{t}{2}\|x\|^2)$ with logconcave q, we have

$$p(\partial E) \gtrsim t^{-1/2} p(E)\sqrt{\ln(1/p(E))}$$

Alternative Proof:

Apply localization lemma, it suffices to prove the case for 1 dimension.

Ask your student to prove the 1 dim problem.

$$\mathbb{E}\, p_t(E)\sqrt{\ln(1/p_t(E))} \gtrsim p(E)\sqrt{\ln(1/p(\boldsymbol{E}))}$$

$$dp_t(x) = (x-\mu_t)^T dW_t \cdot p_t(x).$$

μ_t is the mean of p_t.

For any stochastic process $dX_t = \mu_t dt + \sigma_t dW_t$, we have that

$$df(x_t) = \frac{df}{dx}(\mu_t + \sigma_t dW_t) + \frac{d^2 f}{dx^2}\frac{\sigma_t^2}{2}dt.$$

Let $g_t = p_t(E)$. Ito's lemma shows

$$dg_t\sqrt{\log\frac{e}{g_t}} = \frac{2\log\frac{e}{g_t}-1}{2\sqrt{\log\frac{e}{g_t}}}dg_t - \frac{2\log\frac{e}{g_t}+1}{8g_t\log^{\frac{3}{2}}\frac{e}{g_t}}d[g_t]_t.$$

For expectation, it suffices to upper bound $d[g_t]_t$.

Note that $dg_t = \left\langle \int_E (x-\mu_t)p_t(x)dx\,, dW_t\right\rangle$, $\quad d[g_t]_t = \left\| \int_E (x-\mu_t)p_t(x)dx \right\|_2^2 dt.$

(Next Slide)

Since $d[g_t]_t \lesssim g_t^2 \ln^2(1/g_t)\, dt$, 2nd term $\sim g_t \ln^{1.5}(1/g_t)\, dt$ (good for $t \leq \frac{1}{\ln(1/g_t)}$)

Since $d[g_t]_t \lesssim g_t^2 D^2 dt$, 2nd term $\sim D^2 g_t/\ln^{0.5}(1/g_t)\, dt$ (good for $t \leq \frac{\ln(1/g_t)}{D^2}$)

Combining both cases, we are good up to $t \leq 1/D$.

QED

μ_t is the mean of p_t.
A_t is the covariance of p_t.

$d[g_t]_t \lesssim g_t^2 \ln^2(1/g_t)\, dt$

Recall $d[g_t]_t = \left\| \int_E (x-\mu_t) p_t(x) dx \right\|_2^2 dt$.

Note that

$$\begin{aligned}\left\| \int_E (x-\mu_t) p_t(x) dx \right\|_2 &= \max_{\|\zeta\|_2=1} \int_E (x-\mu_t)^T \zeta \cdot p_t(x) dx \\ &\le \max_{\|\zeta\|_2=1} \left(\int_E \left|(x-\mu_t)^T \zeta\right|^k \cdot p_t(x) dx \right)^{\frac{1}{k}} \left(\int_E p_t(x) dx \right)^{1-\frac{1}{k}} \\ &\le 2k \left\|A_t\right\|_{\mathrm{op}}^{1/2} \cdot {g_t}^{1-\frac{1}{k}}\end{aligned}$$

(If the body is ψ_2, one may get better bound…)

If $\|A_t\|_{op} = O(1)$, then we can take $k = \ln(1/g_t)$.

How to bound $\|A_t\|_{op}$?

By Ito's Lemma, we know that

$$dA_t = \cdots dW_t.$$

Each step, we are adding a mean 0 random matrix into A_t.

Simplest way: Apply standard matrix concentration result such as matrix Chernoff bound.

This gives $\|A_t\|_{op} = O(1)$ for $t \lesssim \frac{1}{\sqrt{n}\ln(n)}$. But we need $t \sim 1/\sqrt{n}$.

Underlying Chernoff bound, it used the potential $\mathrm{Tr}(e^{\ln(n)\cdot A_t})$.

The extra $\ln(n)$ makes the derivative of the potential too large.

If $\cdots$ is a random diagonal matrix, $\ln(n)$ term is unavoidable.

But, dA_t more like a random Gaussian matrix.

BSS potential

Ideally, one would like to track $\lambda_{\max}(A_t)$ instead. (Again not nice).

[Batson Spielman Srivastava 08] shows one can use the potential u_t where

$$\mathrm{Tr}\,(u_t I - A_t)^{-1} = n.$$

u_t is an approximate maximum eigenvalue of A_t.

They used that for constructing graph sparsifier. Later on, it is used on

- Analyze empirical covariance matrix **[Srivastava 11]**
- Online learning **[Audibert-Bubeck 13]**
- Even faster graph sparsifier **[AllenZhu-Orecchia 15, L-Sun 17]**
- ….

Once we know the potential, the rest follows from calculation on du_t.

Open problem

For isotropic logconcave μ, we know

- $\psi_2(\mu) \lesssim \frac{1}{\mathcal{LS}_\mu}$ where $\psi_2(\mu)$ is the ψ_2 constant of μ.
- $\mathcal{P}_\mu \lesssim \mathcal{LS}_\mu$ where $\mathcal{P}_\mu$ is the Poincare constant of μ.

Question: is it true that

$$\mathcal{LS}_\mu = \Theta\left(\mathcal{P}_\mu + \frac{1}{\psi_2(\mu)}\right)?$$

(Maybe too optimistic?)

Recent Progress on Sampling

Sampling Problem

Problem: sample a point from the uniform distribution on a given convex set K.

- Oracle setting: membership oracle of K.
- Polytope setting: $K = \{Ax \geq b\}$.

Why:

- Lying at the heart of optimization theory
- Understand the model
- Compute volume, center of gravity, covariance matrix, diameter, …
- Robust optimization, Bandit problem, …
- Provide a window to learn about convex sets!

Biased view: Mother of many easy problems

Problem is easy iff it can be written as a "convex" problem.

Here are increasingly difficult convex problems:

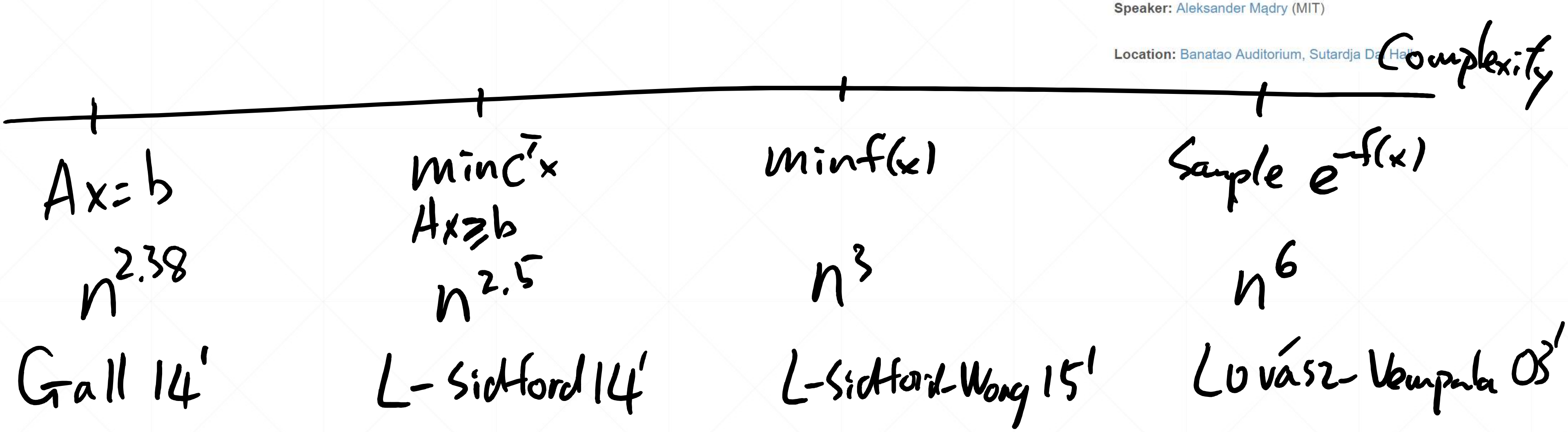
Events | Fall 2017

Gradient Descent: The Mother of All Algorithms?

Dec. 11, 2017 4:00 pm – 5:00 pm

Speaker: Aleksander Mądry (MIT)

Location: Banatao Auditorium, Sutardja Dai Hall

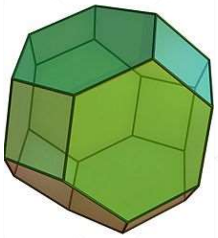


f is convex function.

One ultimate goal of the field: Solve all of them in n^2 time!

Without extra assumption, there is some information lower bound?

Sampling problem on polytopes



Input: a polytope $K = \{Ax \geq b\}$ where m is # of constraints, n is # of variables.

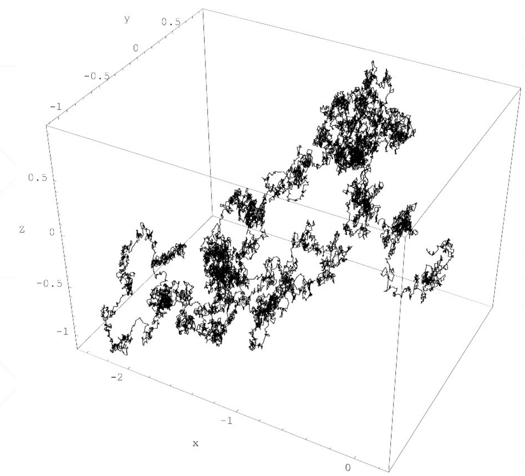
Output: sample a point from the uniform distribution on K.

		Iterations	Time/Iter
[Lovász-Vempala 03]	Ball walk	n^4	mn
[Kannan-Narayanan 09]	Dikin walk	mn	$mn^{1.38}$ (matrix multiplication)
[L-Vempala 16]	Geodesic walk	$mn^{0.75}$	$mn^{1.38}$ (matrix multiplication)
[L-Vempala 17]	Hamiltonian walk	$mn^{0.66}$	$mn^{1.38}$ (matrix multiplication)
Ben's experiment:	Coordinate Hit-and-Run	n^2	m
	Hamiltonian walk	1	Solve $\tilde{O}(1)$ linear system

How does nature mix particles?

Brownian Motion.

It works for sampling on $\mathbb{R}^n$.



However, convex set has boundary ☹.

Option 1) Reflect it when you hit the boundary.

However, it need tiny step for discretization.

How does the nature mixes particle?

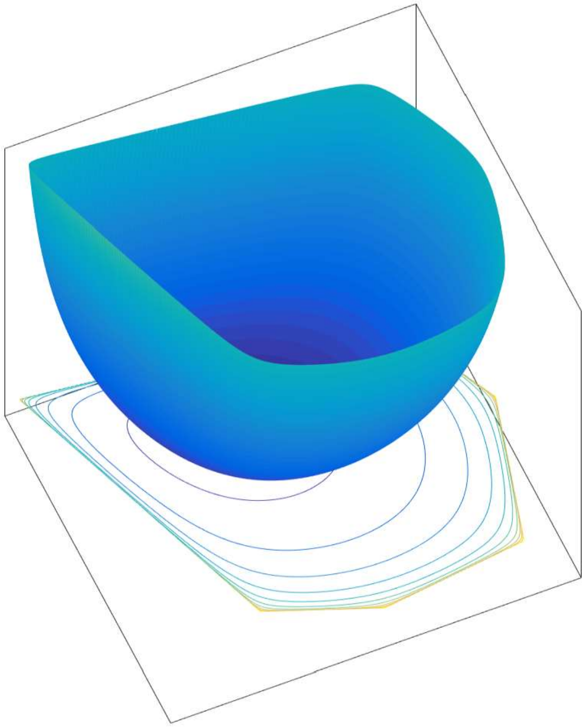
Brownian Motion.

It works for sampling on $\mathbb{R}^n$.

However, convex set has boundary ☹.

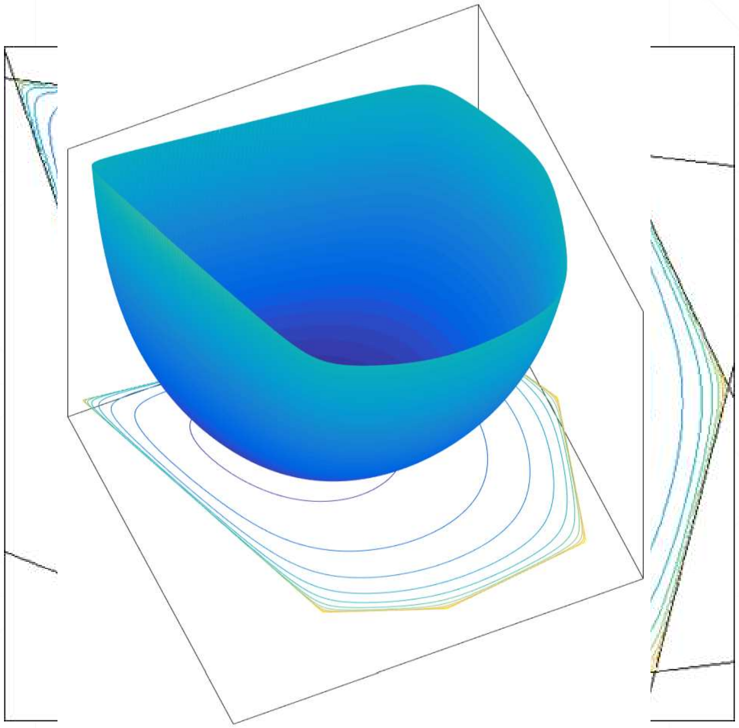
Option 2) Remove the boundary by blowing up.

However, this requires explicit polytopes.



Hessian manifold

Hessian manifold: a subset of $\mathbb{R}^n$ with
inner product defined by $\langle u, v\rangle_p = u^T \nabla^2 \phi(p) v$.



For polytope $\{a_i^T x \geq b_i \ \forall i\}$, we use the log barrier function

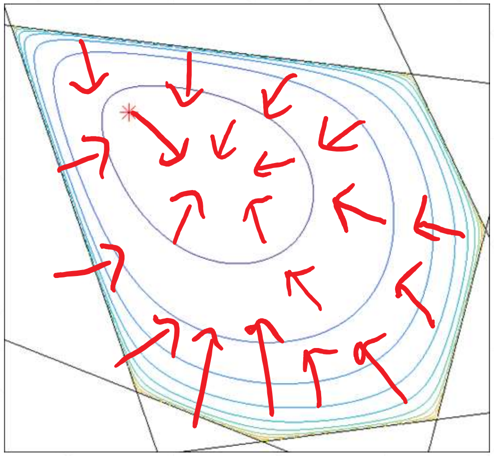
$$\phi(x) = \sum_{i=1}^{m} \log(\frac{1}{s_i(x)})$$

- $s_i(x) = a_i^T x - b_i$ is the distance from x to constraint i
- p blows up when x close to boundary
- Our walk is slower when it is close to boundary.
- To make sure it converges to uniform, we add appropriate amount of drift.

Our algorithm

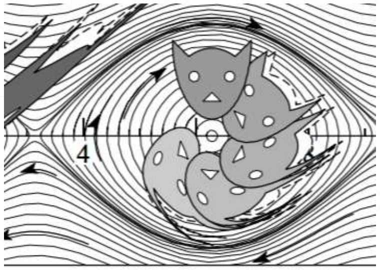
- Run the drifted Brownian Motion on the Hessian manifold.

(The drift is given by $-\frac{1}{2}\nabla \ln \det g\,(x)$ so that the stationary distribution is uniform).



We do a piecewise approximation as follows

- Sample $x'(t)$ from Gaussian vector on $T_{x(t)}\mathcal{M}$.
- Solve the equation $D_t \frac{\partial x}{\partial t} = -\frac{1}{2}\nabla \ln \det g\,(x)$.



The flow $D_t \frac{\partial x}{\partial t} = -\frac{1}{2}\nabla \ln \det g\,(x)$ preserves volume and is symmetric.

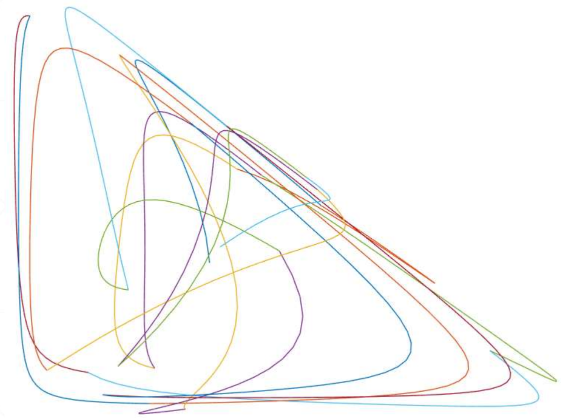
Do not require the Metropolis filter step.

It is like the generalization of hit-and-run on manifold.

(Theoretical) Implementation

To solve $D_t \frac{\partial x}{\partial t} = -\frac{1}{2} \nabla \ln \det g\,(x)$,

we show that the solution can be approximated by $\tilde{O}(1)$ order polynomials.



[L-Vempala 16] Consider the ODE $y' = f(t, y(t))$ with $y(0) = y_0$.

- $Lip(f) \leq 0.001$
- There is a degree d poly p such that $|p' - f'| \leq \varepsilon$.

Then, we can find a y such that $|y - y(1)| = O(\varepsilon)$ in time

$$O(d\log^2(d\varepsilon^{-1})) \text{ with } O(d\log(d\varepsilon^{-1})) \text{ evaluations of } f.$$

Remark: No need to compute f'!

In general, the runtime is $\tilde{O}(nd\mathrm{Lip}^{O(1)}(f))$ instead for n variables ODE.

Convergence Theorem

[L-Vempala 17]: For log barrier, the ham walk mixes in $\tilde{O}(mn^{0.66})$ steps.

[L-Vempala 17]: For log barrier on $[0,1]^n$, it mixes in $\tilde{O}(1)$ steps. ☺

All previous algorithm such as ball-walk, hit-and-run and Dikin walk

takes $\Omega(n)$ steps for $[0,1]^n$.

Open Problem: What is the best metric to use that is still computable?

Open Problem: What is the corresponding KLS conjecture on manifold setting?