The Manifold Joys of Sampling in High Dimension

Santosh S. Vempala

Algorithms and Randomness Center, Georgia Tech

The Sampling Problem

Input: integrable function $f: R^n \rightarrow R_+$, point x s.t. $f(x) \geq \beta$, error parameter ε.

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

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

The Sampling Problem

Problem: sample a point from the uniform distribution on a given convex set K or according to a logconcave density f.

- Oracle setting: membership for K or value of function f.
- Polytope setting: $K = \{Ax \geq b\}$.

Why:

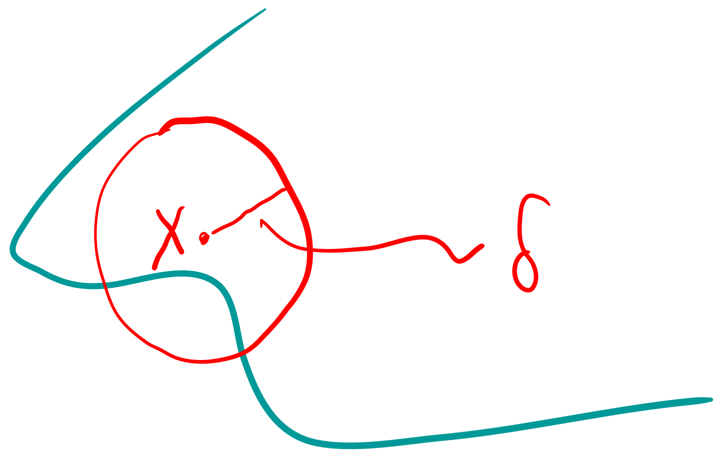
- Compute volume, center of gravity, covariance matrix, …
- Robust/online/private optimization, model exploration, learning
- Provides a lens to understand convexity!
- and optimization, and the model of computation

How to sample?

Ball walk:

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

if y is in K, go to y



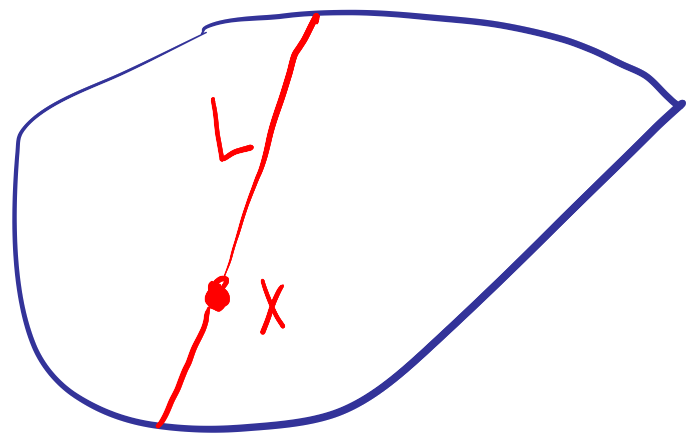
- The process is symmetric
- So the stationary distribution is uniform
- Discrete time version of Brownian motion with reflection.

Hit-and-Run

[Boneh],[Smith]

At x, pick a random chord L through x

go to a uniform random point y on L



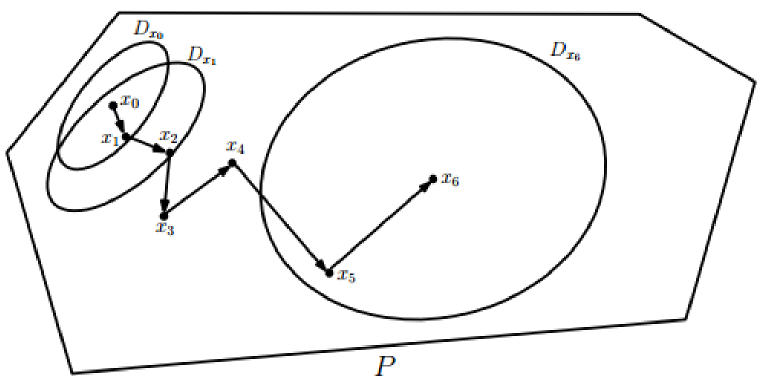
- Random walk is symmetric,
- stationary distribution is uniform
- No need to have a step-size parameter δ
- Coordinate Hit-and-Run: pick random axis direction

Dikin Walk

At x,

pick random y from $E_x = \{y: \|A_x(y - x)\| \leq 1\}$

if $x \in E_y$, go to y with prob. $\min 1, \frac{\mathrm{vol}(E_x)}{\mathrm{vol}(E_y)}$

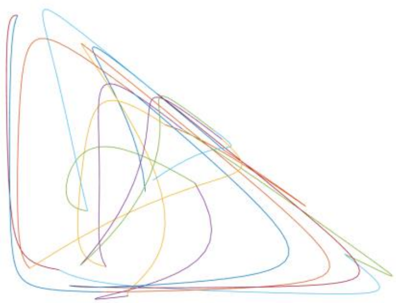


Hamiltonian Monte Carlo

Hamiltonian: function of position and velocity.

Each step is according to an ODE defined by the Hamiltonian:

$$\frac{dx}{dt} = \frac{\partial H(x,v)}{\partial v} \qquad \frac{dv}{dt} = -\frac{\partial H(x,v)}{\partial x}$$



Ham walk: To sample according to $e^{-f(x)}$, set

$$H(x,v) = f(x) + \log\left((2\pi)^n g(x)\right) + v^T g(x)^{-1} v$$

At current point x,

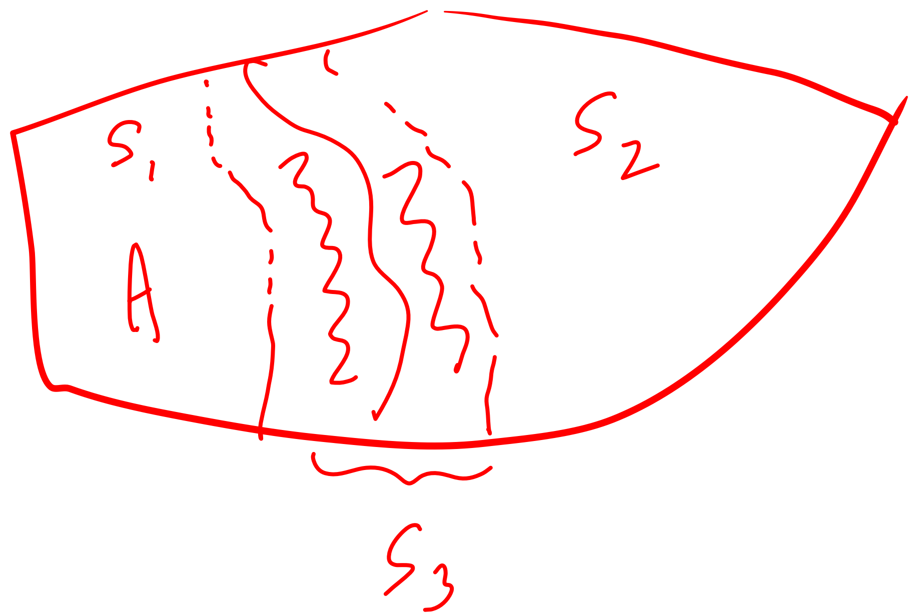
- Pick a random velocity v according to a local distribution $N(0, g(x)^{-1})$ defined by x (in the Euclidean setting, this is a standard Gaussian).
- Move along the curve defined by Hamiltonian dynamics at (x, v) for time δ or $-\delta$, each with probability 0.5.

State of the art, *in theory*

Year/Authors	New ingredients	Steps
1989/Dyer-Frieze-Kannan	Everything	n^{23}
1990/Lovász-Simonovits	Better isoperimetry	n^{16}
1990/Lovász	Ball walk	n^{10}
1991/Applegate-Kannan	Logconcave sampling	n^{10}
1990/Dyer-Frieze	Better error analysis	n^{8}
1993/Lovász-Simonovits	Localization lemma	n^{7}
1997/Kannan-Lovász-Simonovits	Speedy walk, isotropy	n^{5}
2003/Lovász-V.	Annealing, hit-and-run	n^{4}
2015/Cousins-V. (well-rounded)	Gaussian Cooling	n^{3}
2017/Lee-V. (polytopes)	Hamiltonian Walk	$mn^{2/3}$
2021/Jia-Lee-Laddha-V.	Better Rounding	n^{3}

"In Theory today, Ball Walk is Best," i.e., fastest known polynomial-time algorithm.

Convergence depends on isoperimetry

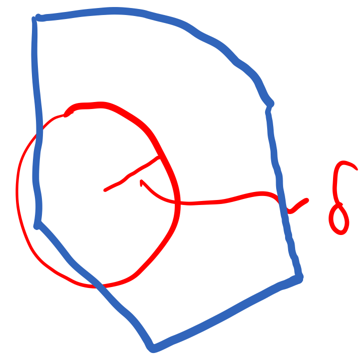


- Technique [LS93]: "conductance" of Markov chain is large.
  - (one-step overlap): Nearby points have overlapping one-step distributions
  - (isoperimetry) Large subsets have large boundaries:

$$\pi(S_3) \geq C \cdot d(S_1, S_2) \min \pi(S_1), \pi(S_2)$$

Convergence of ball walk

Theorem [KLS97]. The ball walk applied to a near-isotropic logconcave density p, from a warm start, converges in $O^*\left(n^2\psi_p^2\right)$ steps.



$$\frac{1}{\psi_p} = \min_S \frac{p(\partial S)}{\min(p(S), p(S^c))}$$

"Cheeger constant of this Markov chain is determined by Cheeger constant of its stationary distribution"

Gaussian Cooling

Thm [Cousins-V'15]. The complexity of sampling/volume computation of any well-rounded convex body is $O^*(n^3)$ membership queries.

- Well-rounded: K contains a unit ball and
$$E\left(\|x-\bar{x}\|^2\right) = \tilde{O}(n)$$
- Most of K lies in a ball of radius $\tilde{O}\left(\sqrt{n}\right)$
- No warm start assumption

- [LV03]: can put K in near-isotropic position in n^4.
- Isotropic position ($E(x) = 0; E(xx^\top) = I$) $\Rightarrow$ well-rounded

- LV rounding + CV algorithm → n^4 sampling for any K.

Rounding and KLS?

- Can we round faster than n^4 ?

Thm [Jia-Laddha-Lee-V'21]. Any convex body can be brought into near-isotropic position using $\tilde{O}(n^3\psi_n^2)$ membership queries.

Cor. Sampling/Volume of any convex body in $O^*(n^3\psi_n^2)$.

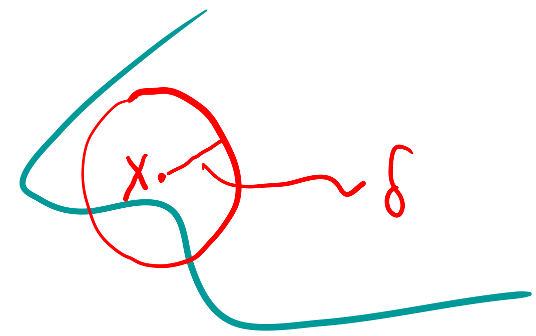
- $n^2\psi_n^2$ for subsequent samples, since we will have a warm start in an isotropic body.

Sampling

Ball Walk, with membership oracle

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

if y is in K, go to y



Thm [KLS97].

n^5 queries for first sample, n^3 queries for later samples.

KLS conjecture ⇒ n^2 for later samples ("warm start" and "isotropic density")

Thm [Jia-Laddha-LV21]

n^3 for first sample.

Thm. [Klartag-Lehec22] KLS true up to polylog.

⇒ n^2 for later samples.

Q. Best possible?

Rounding and Integration (Volume)

Thm. [DFK89]
Volume of a convex body in n^{23} oracle calls.
Thm. [LV06]
Integration of a logconcave function in n^4 oracle calls.

Thm. [Cousins-V.15]
Volume of well-rounded convex body in n^3.

Rounding problem:
Find affine transformation s.t. $y = Ax$ has $E(y) = 0, E(yy^\top) \simeq I$.

Thm. [JLLV21]
Rounding in n^3.

Q. Is quadratic the best possible?

Why "so" slow?

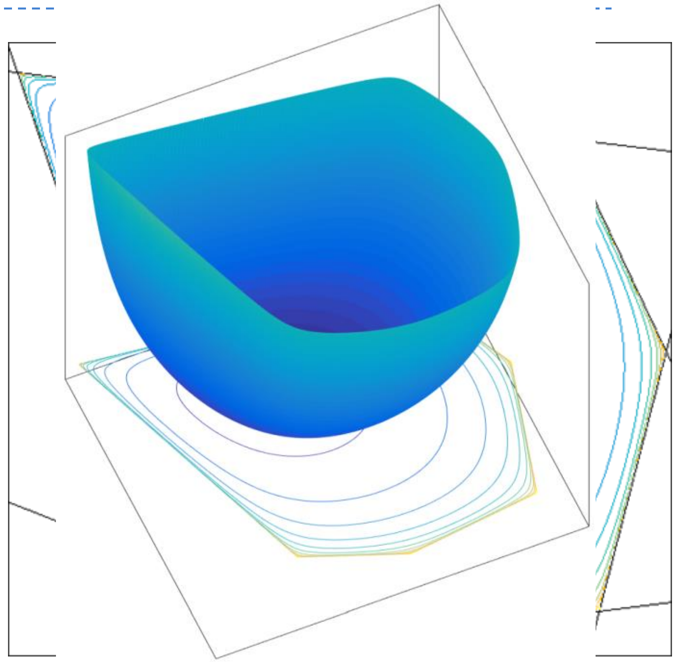
- Bottleneck: Step size, i.e., can only take small steps to maintain polytime, roughly $1/\sqrt{n}$.
- If larger, most steps are wasted, i.e., go outside the body, even in a hypercube.
- How about bigger steps deeper inside, smaller steps near boundary?

- Can we use the "local" geometry?

Polytope → Hessian manifold

Hessian manifold: a subset of $\mathbb{R}^n$ with inner product

$\langle u, v \rangle_x = u^T \left(\nabla^2 \phi(x)\right) v$ for convex ϕ.



For a polytope $\{a_i^T x \geq b_i \ \forall i\}$,

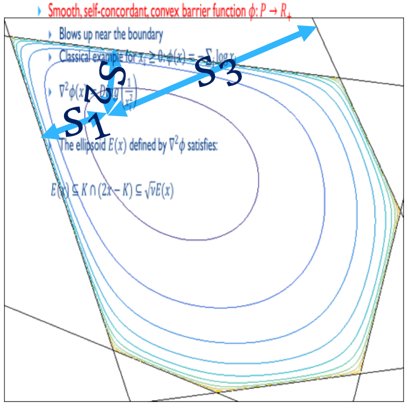
we use the log barrier function:

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

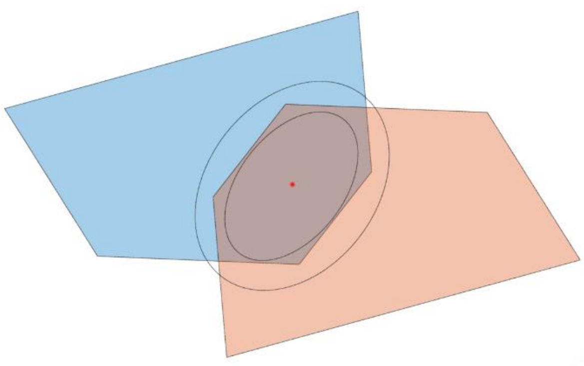
- $s_i(x) = a_i^T x - b_i$ is the distance from x to constraint i
- ϕ blows up when x is close to the boundary
- Distances "stretch" near the boundary

Local geometry from Convex Barriers

- Smooth, self-concordant, convex barrier function $\phi : P \rightarrow R_+$
  - Blows up near the boundary
  - Classical example for $x_i \geq 0$: $\phi(x) = -\sum_i \log x_i$
  - $\nabla^2 \phi(x) = Diag\left(\frac{1}{x_i^2}\right)$
  - The ellipsoid $E(x)$ defined by $\nabla^2 \phi$ satisfies:



$$E(x) \subseteq K \cap (2x - K) \subseteq \sqrt{\nu} E(x)$$



Interior-Point Method

- [Nesterov-Nemirovski94, following Dikin, Karmarkar,…]
- Instead of minimizing $c^\top x$, consider $c^\top x + t \cdot \phi(x)$ where
  - Easier to minimize smooth convex functions (Newton iteration)
  - Gradually reduce t:

$$t \leftarrow t\left(1 - \frac{1}{\sqrt{\nu}}\right)$$

  - where ν is the symmetry parameter
  - #iterations: $\sqrt{\nu}$
- Sequence of optimal points, the central path, is strictly interior

- ϕ needs to be self-concordant, i.e., Hessian $H(x) = \nabla^2\phi\,(x)$ changes slowly:

$$\left\|H(x)^{-1/2} DH(x)[h] H(x)^{-1/2}\right\| \leq 2h^T H(x) h$$

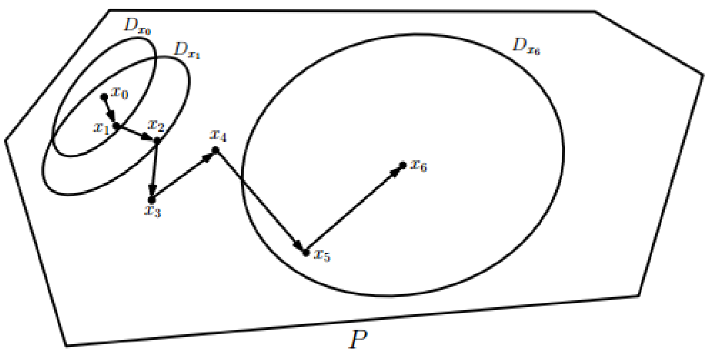
(when $H(x) = I$, then this is $\|DH(x)[h]\| \leq 2\|h\|^2$)

Interior-Point Method 2.0

- Has led to improvements in the past decade for Combinatorial Optimization and Linear Programming!
  - Universal barrier: $\nu = n + 1$, time: $poly(n)$
  - Entropic barrier: $\nu = n$, time: $poly(n)$
  - Log barrier: $\nu = m$, fast
- Thm. [LS14] Weighted log barrier: $\nu = \tilde{O}(n)$, fast!
  - Implies $\sqrt{n}$ iterations to solve a linear program with one linear system per iteration

Sampling with an adaptive step size

- Use the ellipsoid defined by the Hessian of a convex function!
- Hessian $H = \nabla^2 \phi$ defines a local metric: $\|v\|_x^2 = v^\top H(x) v$.
- Dikin walk: At x,
  - pick random y from $E_x = \{y: \|A_x(y-x)\| \le 1\}$
  - if $x \in E_y$, go to y with prob. $\min\left\{1, \frac{\mathrm{vol}(E_x)}{\mathrm{vol}(E_y)}\right\}$
- For log barrier, $A_x = Diag\left(\frac{1}{s_i(x)}\right) A$
  - Each row is scaled by distance to boundary
- $H(x) = A_x^\top A_x$

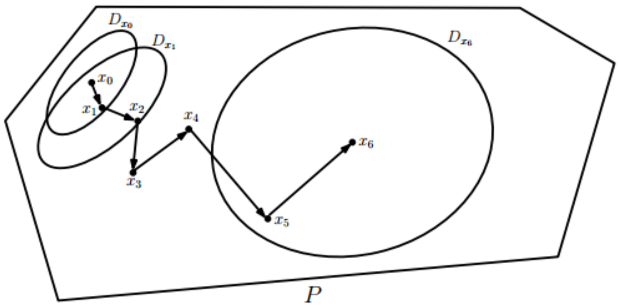


Thm. [K-Narayanan12]

Dikin walk with log barrier mixes in mn steps, $mn^{\omega-1}$ per step.

Weighted Dikin walk

- Dikin walk: At x,
  - pick random y from $E_x = \{y: \|H(x)^{1/2}(y-x)\| \leq 1\}$
  - if $x \in E_y$, go to y with prob. $\min 1, \frac{\mathrm{vol}(E_x)}{\mathrm{vol}(E_y)}$



Thm. [K-Narayanan12]

Mixes in mn steps, $mn^{\omega-1}$ per step.

Thm. [Laddha-LV20]

Mixes in $n\nu$ steps for *any* strongly self-concordant barrier.

- Log barrier: mn steps, $nnz(A) + n^2$ per step.
- Weighted log barrier: n^2 steps, $mn^{\omega-1}$ per step.
- Strongly self-concordant:

$$\left\|H(x)^{-1/2}DH(x)[h]H(x)^{-1/2}\right\|_F = O\left(h^T H(x) h\right)$$

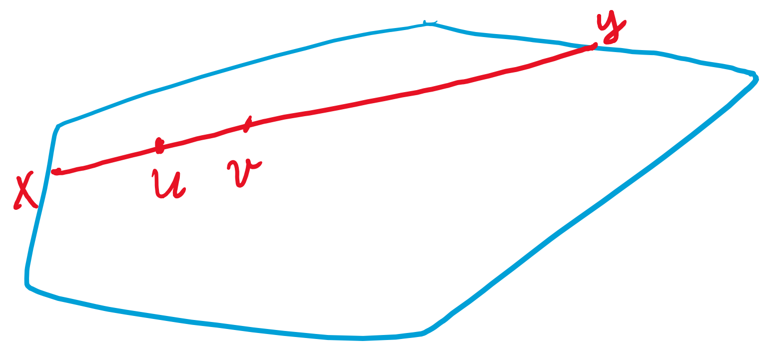
Isoperimetry

- Isoperimetry is in a non-Euclidean metric: For any partition of a convex body K into subsets S_1, S_2, S_3,

$$p(S_3) \geq d_K(S_1, S_2) p(S_1) p(S_2)$$

Cross-ratio distance:

$$d_K(u, v) = \frac{\|u-v\|\|x-y\|}{\|x-u\|\|v-y\|}$$



Hilbert distance:

$$d_H(u, v) = \log(1 + d_K(u, v))$$

is a metric.

Q. Does weighted Dikin mix in n steps? (mn is tight for log barrier)

- Aside: KLS conjecture ⇒ strong self-concordance for Universal and Entropic barriers ☺

The rejection probability bounds step size

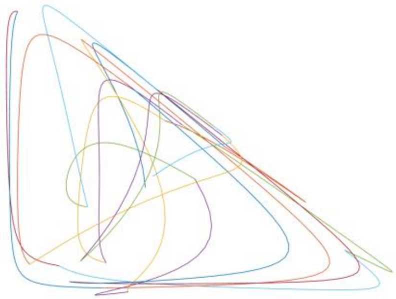
- How to take a larger step?
- Can we avoid the Metropolis filter?

- Let's use a deterministic "drift" instead.

Riemannian Hamiltonian Montian Carlian

Hamiltonian: function of position and velocity.
Each step is according to an ODE defined by the Hamiltonian:

$$\frac{dx}{dt} = \frac{\partial H(x,v)}{\partial v} \qquad \frac{dv}{dt} = -\frac{\partial H(x,v)}{\partial x}$$



Ham walk: To sample according to $e^{-f(x)}$, set

$$H(x,v) = f(x) + \frac{1}{2}\log\left((2\pi)^n \det g(x)\right) + \frac{1}{2} v^T g(x)^{-1} v$$

At current point x,

- Pick a random velocity v according to a local distribution $N(0, g(x)^{-1})$ defined by x (in the Euclidean setting, this is a standard Gaussian).
- Move along the curve defined by Hamiltonian dynamics at (x, v) for time δ or $-\delta$, each with probability 0.5.

Convergence of RHMC

Thm [Lee-V.17]: With log barrier, RHMC mixes in $\tilde{O}(mn^{2/3})$ steps.

- Subquadratic!

Thm [Lee-V.17]: For log barrier on $[0,1]^n$, RHMC mixes in $\tilde{O}(1)$ steps.

- Previous algorithms such as ball walk, hit-and-run and Dikin walk take $\Omega(n)$ steps for $[0,1]^n$.
- Each step is the solution of a linear system, so $mn^{\omega-1}$

Q: Can we use dynamic data structures to reduce the per-step cost?

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

Q: What is the right KLS conjecture in the Hessian manifold setting?

Constrained RHMC

- Typical problems often have equality constraints $Ax = b$.
- Pick the metric g in the subspace:

$$H(x,v) = f(x) + \frac{1}{2}||v||^2_{g(x)^\dagger} + \frac{1}{2}\log \mathrm{pdet} g(x).$$

$$H(x,v) = f(x) + \frac{1}{2}v^\top X(I - XA^\top(AX^2A^\top)^{-1}AX)Xv - \sum_i \log x_i + \frac{1}{2}\log\det(AX^2A^\top)$$

CRHMC Algo:

- Sample $v \sim e^{-H(x,v)}$ (conditional on x)
- $(x,v) \leftarrow T(x,v)$ (T preserves the density $e^{-H(x,v)}$)

The map $T(x,v)$ is given by an ODE (solved at $t = 1$)

$$\frac{dx}{dt} = \frac{dH}{dv}, \qquad \frac{dv}{dt} = -\frac{dH}{dx}, \qquad x(0) = x, v(0) = v.$$

State of the art, *in theory*

General Logconcave	Gaussian in Convex Body	Uniform in Convex Body	Uniform in Polytope
[Lovász-V'06]	[Cousins-V'15]	[Jia-Laddha-Lee-V'21]	[JLLV'21]
$n^4 \cdot n^2$	$n^3 \cdot n^2$	$n^3 \cdot n^2$	$mn^{3.2}$
$n^2 \cdot n^2$	$n^2 \cdot n^2$	$n^2 \cdot n^2$	$mn^{2.3}$ (warm start)
Ball walk /H-and-R	Ball walk	Ball walk	Ball walk

RHMC: $mn^{2/3} \cdot mn^{1.38}$

Weighted Dikin: $n^2 \cdot mn^{1.38}$

"In Theory today, Ball Walk is Best," i.e., fastest known polynomial-time algorithm.

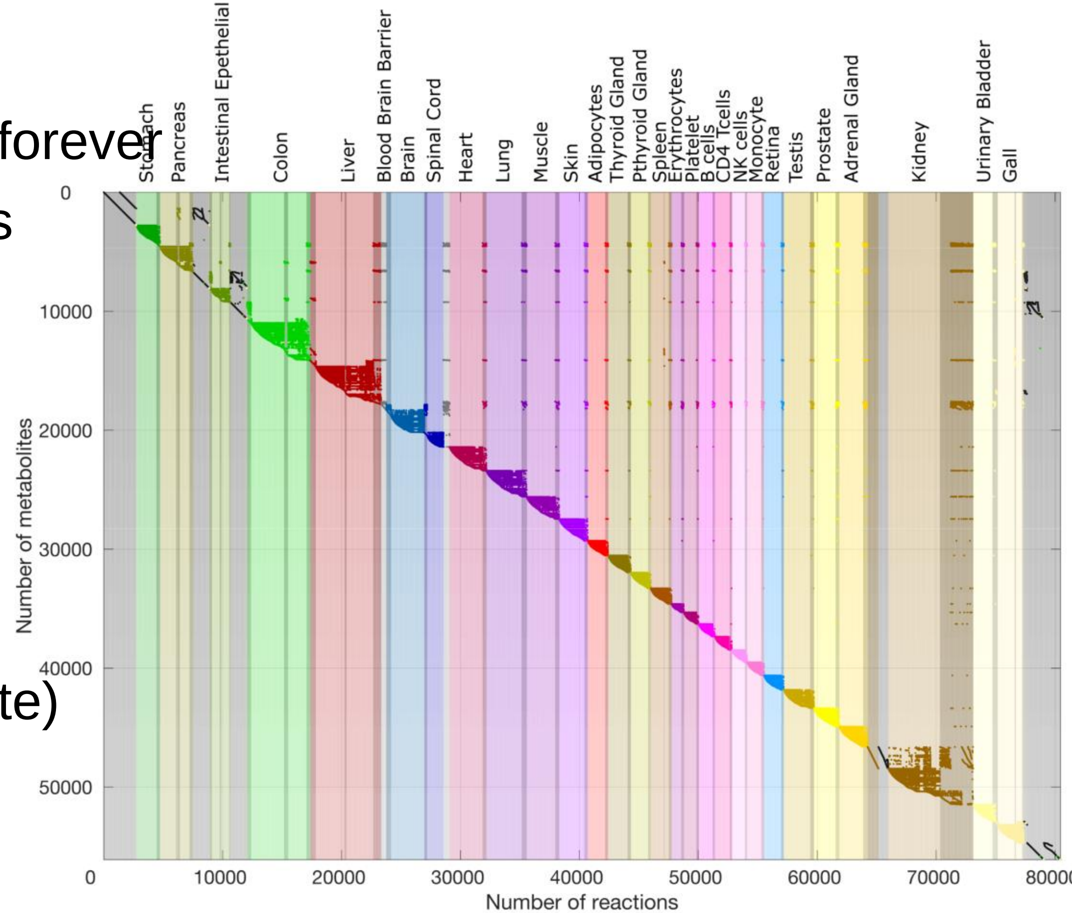
State of the art, *in practice*: CRHMC*

- Ronan Fleming gave us the latest, largest metabolic model.
- 670,114 reactions and 585,662 Metabolites

- Zero'th-order methods take forever
- Existing first-order packages simply can't move at all.

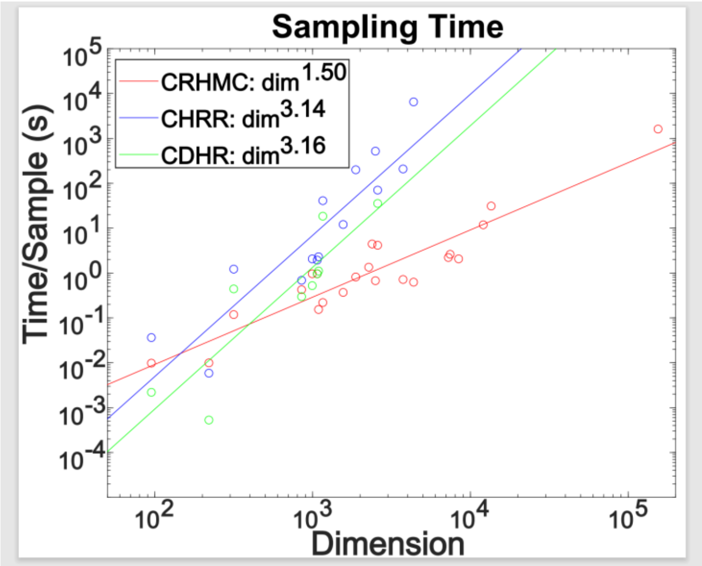
- CRHMC takes <1 hr per sample
- Can also sample polytopes in netlib (notoriously degenerate)

*: pronounced CRuHMCh



You can try it!

- https://github.com/ConstrainedSampler/PolytopeSamplerMatlab
- With Yunbum Kook, Yin Tat Lee, Ruoqi Shen (2022)
- Now in COBRA, the leading system biology analysis tool (Ronan Fleming, Ines Thiele et al.)



Earlier packages for Volume/Sampling

- Cousins-V' (circa 2013)
  - MATLAB ("A Practical Volume Algorithm", Math. Prog. C 2016)
  - incorporated in COBRA (with R. Fleming, H. Haroldsdottir)
  - Computes volume using a membership oracle
  - Goes up to 1000 full-dimensional polytopes on laptop in < 1hr.
  - https://www.mathworks.com/matlabcentral/fileexchange/43596-volume-and-sampling
- VolEsti (Fisikopoulos et al.)
  - C++ (Emris-Fisikopoulos, ACM Trans. on Math. Software 2018)
  - Reported better run times for some benchmarks
  - https://github.com/GeomScale/volume_approximation

Continuous Algorithms

OPT: $dX_t = -\nabla f(X_t)dt$ (GD)

Sampling: $dX_t = -\nabla f(X_t)dt + \sqrt{2}dB_t$ (LD)

- Langevin Diffusion converges to distribution with density proportional to $e^{-f(x)}$

Thm. [Jordan-Kinderlehrer-Otto98; Wibisono18]

Sampling by LD is optimization in the space of measures with Wasserstein metric and objective relative entropy to target e^{-f}.

Can we sample faster?

- Brownian motion SDE:

$$dx_t = \mu(x_t, t)dt + \sqrt{2A(x_t, t)}dW_t$$

- Each point $x \in K$ has its own local scaling (metric) given by $A(x_t, t)$.

Thm. [Fokker-Planck] Diffusion equation of above SDE is

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

- When $\mu = 0, A = I$, this is the heat equation: $\frac{\partial}{\partial t}p(x,t) = \frac{1}{2}\Delta p(x,t)$.

- For any metric, SDE gives diffusion equation.
- Using $\mu(x) = -Df(x)$ gives stationary $p(x) = e^{-f(x)}$.

Sampling by Diffusion: Isoperimetry suffices

- Rate of convergence?
- $dX_t = -\nabla f(X_t)dt + \sqrt{2}dB_t$

Thm. [Bakry-Gentil-Ledoux14] $H_\nu(\rho_t) \leq e^{-2\alpha t} H_\nu(\rho_0)$

Here α is the Log-Sobolev constant of e^{-F} wrt the metric.

$$H_\nu(\rho) = E_\rho\left(\log\frac{\rho}{\nu}\right) \leq \frac{1}{2\alpha} E_\rho\left(\left\|\log\frac{\rho}{\nu}\right\|^2\right) = \frac{1}{2\alpha} I_\nu(\rho)$$

- Proof notes that $\frac{d\rho}{dt} = -\nabla_\rho H_\nu(\rho)$ and LSI is "gradient domination.
- How about an algorithm?

Diffusion→Algorithm: Isoperimetry suffices

- Unadjusted Langevin Algorithm:

$$X_{k+1} = X_k - h\nabla f(X_k) + \sqrt{2h}\, Z \quad \text{where } Z \sim N(0, I)$$

Thm.[V.-Wibisono19] Assuming f is L-smooth ($\|\nabla f\| \leq L$),

$$H_\nu(\rho_k) \leq e^{-h\alpha k} H_\nu(\rho_0) + \frac{8L^2 n}{\alpha} h.$$

So, with $h = \alpha\delta / nL^2$,
after $k = \frac{nL^2}{\delta} \log(\frac{2H_\nu(\rho_0)}{\delta})$ steps, we have $H_\nu(\rho_k) \leq \delta$.

- Note: no convexity assumption; dependence on dimension is linear.
- An active field, with many results based on smoothness parameters for interesting classes of functions.

Manifold Diffusion → Algorithm

- What about using local geometry?

Riemannian Langevin Diffusion

- In Euclidean coordinates:

$$dX_t = \big(D \cdot g(X_t)^{-1} - g(X_t)^{-1} Df(X_t)\big)dt + \sqrt{2g(x)^{-1}}dB_t$$

- In manifold local coordinates:

$$dX_t = \big(\nabla \cdot g(X_t)^{-1} - \nabla F(X_t)\big)dt + \sqrt{2g(x)^{-1}}dB_t$$

- where ∇ is the manifold derivative, $F(x) = f(x) + \frac{1}{2}\log\det g(x)$
- Convergence in KL-divergence under log-Sobolev inequality wrt manifold measure holds

In progress: Riemannian Langevin Algorithm

- discretization of RLD [Erdogdu-Li21, Ahn-Chewi21, Gatmiry-V.22]

The Story of Isoperimetry

KLS conjecture: Cheeger constant (expansion) of isotropic logconcave density is $\Omega(1)$, or

$$\psi = \inf_{\nu(S)\le\frac{1}{2}} \frac{\nu_n(S)}{\nu_{n-1}(\partial S)} = O(1).$$

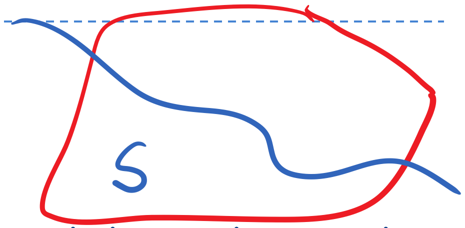
[KLS95]	$\sqrt{n}$
[Guedon-Milman]	$n^{1/3}$
[LV17]	$n^{1/4}$
[Chen20]	$2^{\sqrt{\log n \log\log n}}$
[Klartag-Lehec22]	$\log^5 n$
	$\dots$

Thm.[KLS97]. Sampling in $n^2\psi^2$.

Thm.[JLLV21]. Rounding in $n^3\psi^2$.

Thm.[CV15]. Volume of well-rounded body in n^3.

Isoperimetry: the next decade



- How true is the KLS conjecture? Does it matter?
  - Dimension-independent bound would be so nice
  - Implies dimension-independent bounds for many other well-known, existing conjectures in convex geometry: Slicing, Thin-Shell, Central Limit, Concentration, Entropy Jump etc.
  - But here's a concrete TCS reason:

KLS ⇒ Certifiable sub-Gaussianity [Kothari-Steinhardt17]

  - If KLS is true, then there is an SoS proof of moment inequalities for any logconcave density.
  - This implies results on robustly clustering Gaussians can be generalized to robustly clustering logconcave densities!
  - Getting a constant is critical for polytime, with the SoS approach.

Q. Are they equivalent?!

Almost: certifiable sub-Gaussianity ⇒ thin-shell ⇒ KLS is $O(\log n)$.

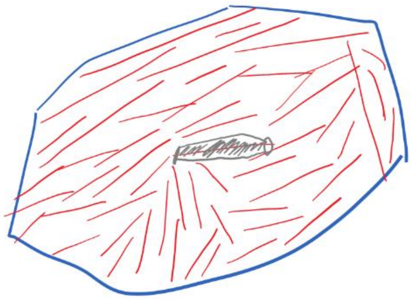
Isoperimetry: the next decade

- Q. What is the right KLS conjecture on Hessian manifolds?

An attempt: there is a subset defined by a hyperplane that is within $O(1)$ of the minimum isoperimetry subset.

- A decomposition conjecture for convex bodies ($\Rightarrow$ KLS).

Conj: For any isotropic convex body,
any decomposition of it into cylinders,
a constant fraction of the cylinders
must be of length $O(1)$.



Cylinder: cross section is convex and has small diameter

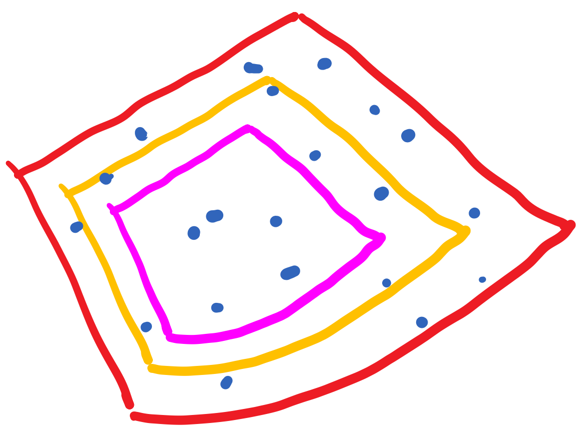
Open Problems: Probability

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

- Ben Cousin's uniformity test:

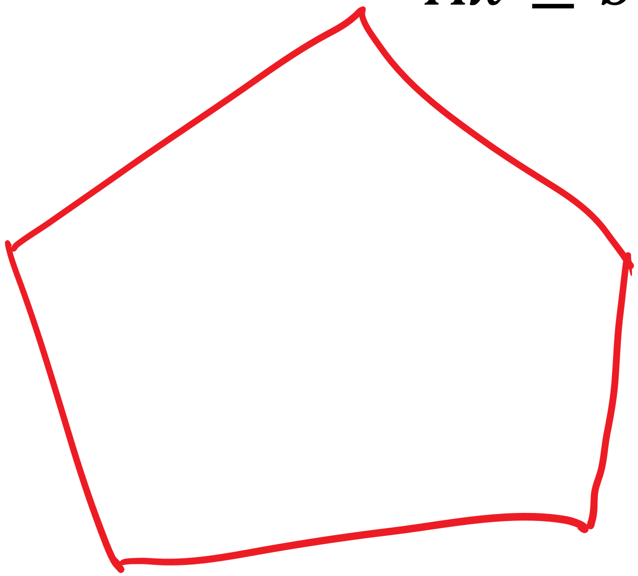
Check if time spent in scaling $(1-\alpha)K$ is $(1-\alpha)^n$.



Randomness

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

$$Ax \leq b$$



Thank you!

and:

Ravi Kannan

Laci Lovász

Adam Kalai

Ronan Fleming

Ben Cousins

Yin Tat Lee

He Jia

Aditi Laddha

Ruoqi Shen

Yunbum Kook

Khashayar Gatmiry