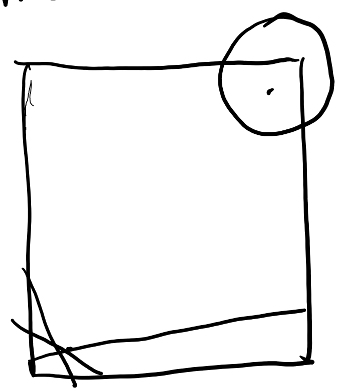
IPM for Sampling

Tuesday, April 15, 2025 4:00 PM

Yuhao

Current best general samplers are
ball walk — $O(n^2\|A\|_{op}^2)$ from an M-warm start
and
Hit-and-run — $O(n^2 \operatorname{tr}(A)) \longrightarrow L_2$-warm start.

The bottleneck for both is the boundary.



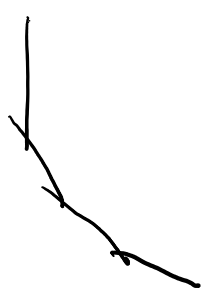
$[-\sqrt{3}, \sqrt{3}]^n$ isotropic

- need to set $\delta = O(\frac{1}{\sqrt{n}})$ to keep rejection prob. small for ball walk.
- Expected chord length in random direction from random point is $O(\frac{1}{\sqrt{n}})$, so typical step-size of H-&-R is $O(\frac{1}{\sqrt{n}})$.

Q. How to deal with boundary? Can we take larger steps?

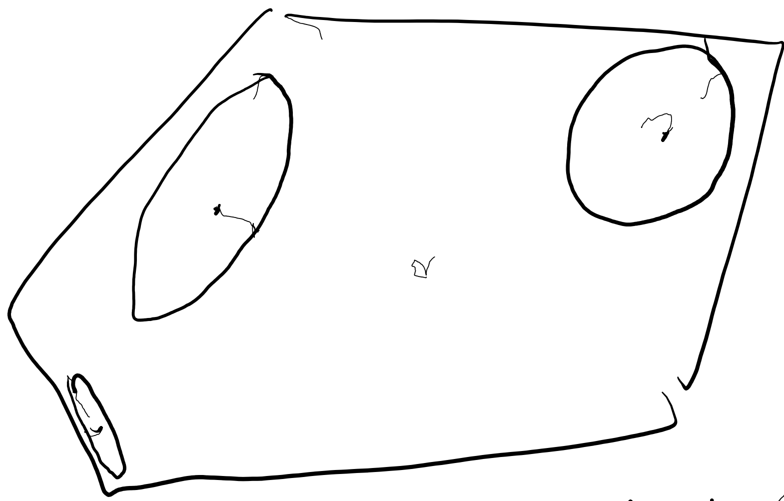
This is not a discrete time issue — Brownian Motion also has to deal with boundary issues.

What about OPT.? GD can be very slow — keep hitting the boundary.



IPM → avoids the boundary

(Later: reflect at boundary!)



Idea Take step based on current position relative to boundary. At x, consider $\nabla^2\phi(x)$ for convex barrier ϕ.

$$E(x) = \{ y : (y-x)^T \nabla^2 \phi(x) (y-x) \le 1 \}$$

DIKIN ellipsoid at x.

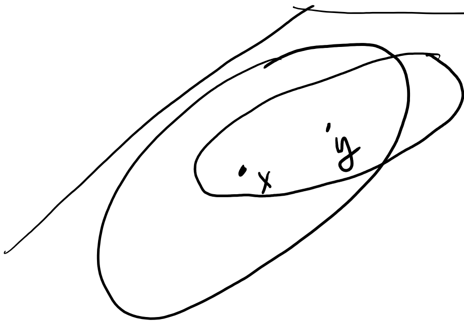
Lemma For a self-concordant barrier ϕ for convex K, $E(x) \subseteq K$.

Proof.....

DIKIN walk. At x:

- sample $y \sim E(x)$
- If $x \in E(y)$, go to y w.p. $\min\left\{1, \frac{vol(E(x))}{vol(E(y))}\right\}$

Alternative



Alternative

$$y \sim N(x, r^2 \cdot \nabla^2 \phi(x)^{-1}).$$

This is affine-invariant!

Q. What is the mixing time?

- First we need to ensure that rejection probability is bounded

For this it suffices to scale E_x by a constant

$$E_x = \{y : \|y-x\|_x \le r\} \quad r = \frac{1}{512}.$$

- One-step coupling:

if x, y are "close" then $d_{TV}(P_x, P_y) \le 0.9$

- Isoperimetry: Large subsets have large boundaries.

Q. What is the right notion of distance?

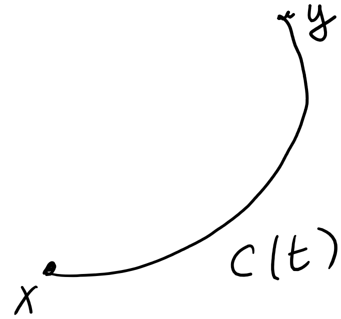
Convex barrier ϕ induces a metric via its Hessian:

$$\|v\|_x = \sqrt{v^T \nabla^2 \phi(x) v}$$

This is a Riemanian Manifold, in fact a special case called a Hessian Manifold.

For a curve $c : [0,1] \to M$

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



$$d_\phi(x,y) = \min_{c \text{ from } x \text{ to } y} \text{length}(c).$$

Lemma 1. $x, y \in P$ $\quad \|x-y\|_x < \frac{1}{512\sqrt{n}} \Rightarrow d_{TV}(P_x, P_y) < \frac{3}{4}$.

ϕ: log barrier

Lemma 2. $\quad E_x \subseteq K \cap 2x - K \subseteq \sqrt{m}\, E_x$

for any self-concordant barrier $\qquad$ for log barrier

Lemma 3. S_1, S_2, S_3 partition of K.

$\forall x \in S_1,\ y \in S_2,\ \|x-y\|_x \geq d$.

Then $\quad \text{vol}(S_3) \gtrsim \frac{d}{\sqrt{n}} \min\{\text{vol}(S_1), \text{vol}(S_2)\}$

$\sqrt{\bar{\nu}}$

<u>Lemma</u> 4. $d_K(x,y) \geq \frac{\|x-y\|_x}{\sqrt{\bar{\nu}}}$ $\boxed{\bar{\nu}: \ E_x \subseteq K \cap 2x-K \subseteq \sqrt{\bar{\nu}}\, E_x}$

<u>Thm</u> (1) $\forall x,y \in M$ $\|x-y\|_x \leq \Delta < 1 \Rightarrow d_{TV}(P_x, P_y) < 0.9$.

(2) Isoperimetry ψ_M wrt to $\nabla^2\phi$ metric.

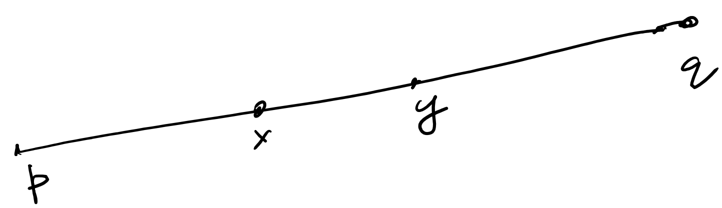
Then conductance of DIKIN $\gtrsim \psi_M \Delta$.

<u>Cor.</u> For log barrier in polytope, $\Delta = \frac{1}{512\sqrt{n}}$, $\psi_M \geq \frac{1}{2\sqrt{m}}$.

mixing rate $= O(mn)$.

Recall $\mathrm{Vol}(S_3) \geq d_K(S_1,S_2) \frac{\mathrm{Vol}(S_1)\,\mathrm{Vol}(S_2)}{\mathrm{Vol}(K)}$

$d_K(x,y) = \frac{\|x-y\|\,\|p-q\|}{\|p-x\|\,\|y-q\|}$.



<u>Pf L4</u>. Assume $\|p-x\|_2 \leq \|y-q\|_2$.

$p \in K \cap 2x-K$. $\Rightarrow \|p-x\|_x \leq \sqrt{\bar{\nu}}$. (def of $\bar{\nu}$).

So $d_K(x,y) \geq \frac{\|x-y\|_2}{\|p-x\|_2} = \frac{\|x-y\|_x}{\|p-x\|_x} \geq \frac{\|x-y\|_x}{\sqrt{\bar{\nu}}}$.

Pf. L2: $\nabla^2\phi(y) \succsim \dfrac{\nabla^2\phi(x)}{(1-\|x-y\|_x)^2}$ is bounded if $\|x-y\|_x < 1$

$\Rightarrow y \in K$.

Let $S_x = \begin{pmatrix} b_i - a_i^T x & & 0 \\ & \ddots & \\ 0 & & \ddots \end{pmatrix}$

$|a_i^T y - a_i^T x| \le b_i - a_i^T x$

Then $\| S_x^{-1} A (y-x) \|_\infty \le 1$

all constraints are satisfied

$\nabla^2\phi(x) = A^T S_x^{-2} A$

$\therefore \|x-y\|^2_{\nabla^2\phi(x)} = (x-y)^T A^T S_x^{-2} A(x-y)$

$= \sum_{i=1}^{m} \left(S_x^{-1} A(x-y)\right)_i^2 \le m.$

Each step is a bit more complicated.

For log barrier it can be implemented in $O(\mathrm{nnz}(A) + n^2)$ amortized time.

This is still hitting the quadratic bound.

This is still hitting the quadratic norm.

Faster?

Let's go truly non-Euclidean.

Move along <u>curves</u> rather than straight lines.

shortest path curves.

<u>geodesic</u>: locally length-minimizing curve.

Manifold M. $\forall p \in M$ $\quad \|v\|_p$

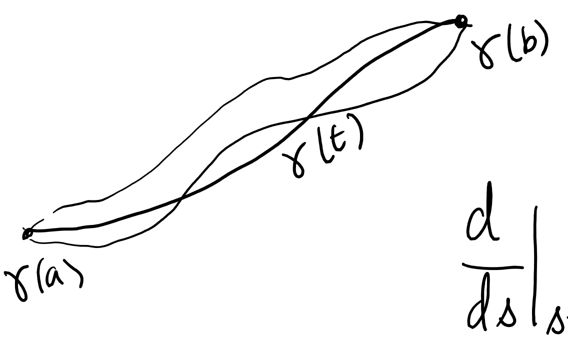
$\gamma(t) : [a,b] \to M$ is a <u>geodesic</u> if

- $\left\| \frac{d}{dt} c(t) \right\|_{c(t)}$ is constant

- For family of curves $c(t,s)$, $\quad c(t,0) = \gamma(t)$

  $c(a,s) = \gamma(a)$

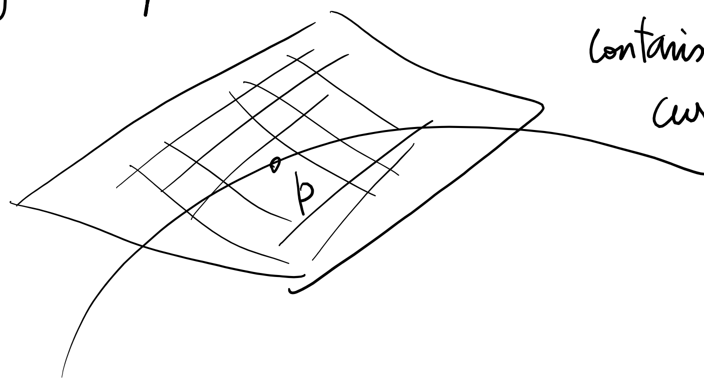
  $c(b,s) = \gamma(b)$



$$\frac{d}{ds}\bigg|_{s=0} \int_a^b \left\| \frac{d}{dt} c(t,s) \right\|_{c(t,s)} = 0 .$$

Tangent space. $\forall p \in M$ $\quad T_pM \equiv \mathbb{R}^n$.

contains gradients of all curves through p.

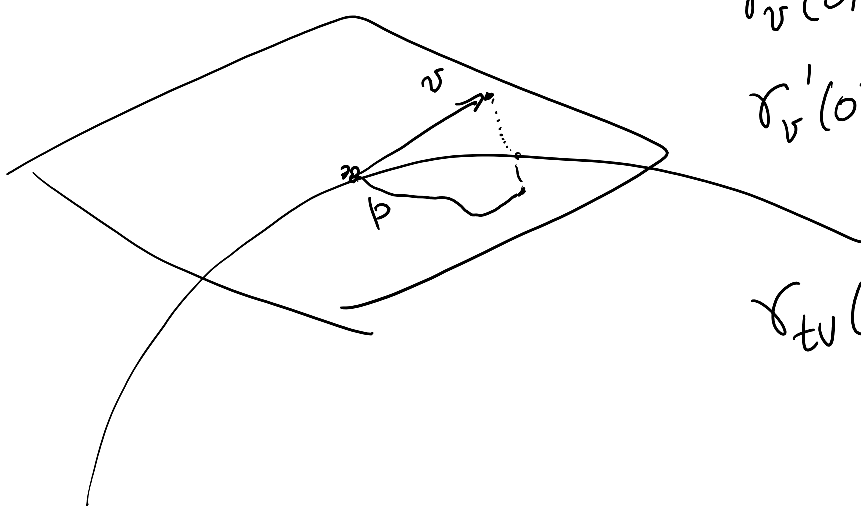


Exponential map. $\quad \exp_p(v) = \gamma_v(1)$

$p \in M$

$\gamma_v(0) = p$ starts at p

$\gamma_v'(0) = v$ initial velocity v.



$\gamma_{tv}(1) = \gamma_v(t)$.

Diffusion $\quad dX_t = \mu(X_t)\,dt + \nabla^2\phi(x)^{-\frac{1}{2}} dB_t$

By F-P.

$$\frac{\partial}{\partial t} p(x,t) = -\sum_i \frac{\partial}{\partial x_i}\left(\mu(x)\, p(x,t)\right) + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}\left(\nabla^2\phi(x)^{-1} p(x,t)\right)$$

∂t

$\frac{\partial p}{\partial t} = 0.\ p(x,t) = 1.$ $\qquad \mu(x) = \frac{1}{2} \nabla \cdot \left(\nabla^2 \phi(x)^{-1} \right)$

How to discretize?

$z \sim N(0, I)$

$$X_{t+h} = X_t + \mu_{X_t} \cdot h + \sqrt{h}\ \nabla^2 \phi(x)^{-1/2} z$$

is in arbitrary Euclidean coordinates.

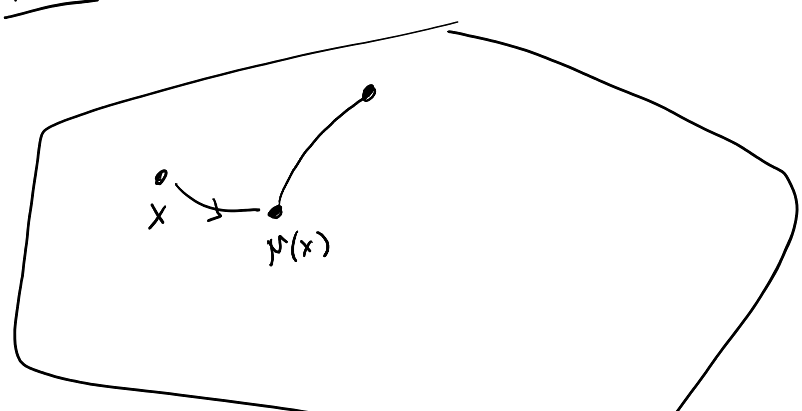
Rather in manifold coordinates, $F = \exp_{x_0}^{-1}$

$$dF(x_0) = \frac{1}{2} \mu(x_0)\, dt + \nabla^2 \phi(x_0)^{-1/2} dB_t$$

$$X_{t+h} = \exp_{X_t} \left(\frac{1}{2} \mu(X_t) h + \sqrt{h}\ z \right).$$

Thm. Geodesic walk mixes in $mn^{3/4}$ steps for Log barrier metric. $\left(h = n^{-3/8}. \right)$.

Thm. Mixes in $n^{1/3}$ for $[0,1]^n$.



natural ?

Is this the most natural?

No!

Adjust drift continuosly.

$H(x,v)$: "Hamiltonian".

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

$T_\delta(x,y)$ apply these equations for the δ.

$$H(x,v) = f(x) + \frac{1}{2}\|v\|^2$$

$$\frac{dx}{dt} = v \qquad \frac{dv}{dt} = -\nabla f(x).$$

$$\frac{d}{dt} H(x,v) = \frac{\partial H(x,v)}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial H(x,v)}{\partial v} \cdot \frac{dv}{dt} = 0.$$

Hamiltonian is preserved.

Hamiltonian Monte-Carlo

At (x,v)

pick $\bar{v} \propto e^{-H(x,\cdot)}$

For the δ go to $T_\delta(x,\bar{v})$ or $T_{-\delta}(x,\bar{v})$

w.p. $\frac{1}{2}$ each.

w.p. $\frac{1}{2}$ each.

$\pi(x) \propto e^{-f(x)}$. $\quad H(x,v) = f(x) + \frac{1}{2}\|v\|^2$.

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

g: local metric. $\longrightarrow e^{-f}$.

<u>Thm</u>. Mixes in $mn^{2/3}$ for log barrier on polytope.

<u>Thm</u>. ——— $\log n$ for $[0,1]^n$.