Sampling

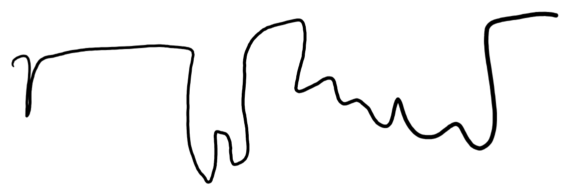
Wednesday, February 19, 2025 9:41 AM

goal: Sample $x \sim \Omega$ $Q(x) \propto e^{-f(x)}$ e.g. uniform.

Ω could be discrete or continuous.

Q. For what target distributions Q is this efficiently solvable?

$\boxed{\text{Hard in general.}}$



access to x via f.

we saw Langevin when ∇f is accessible.

What are "nice" targets?

Spanning trees of a graph?

(perfect) Matchings of a graph?

random solutions of an integer program?

Linear program?

Uniform in a convex K

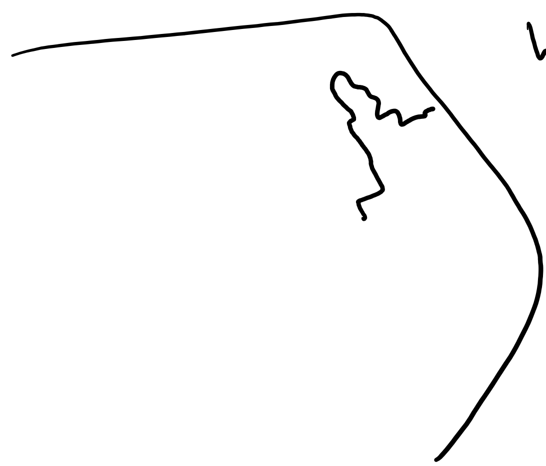
Sample $\sim e^{-f}$ f is convex

$\vdots$

Nature's solution: diffusion

$dX_t = dW_t$ $\qquad$ $dX_t = -\nabla f(X_t)\,dt + \sqrt{2}\,dW_t$.

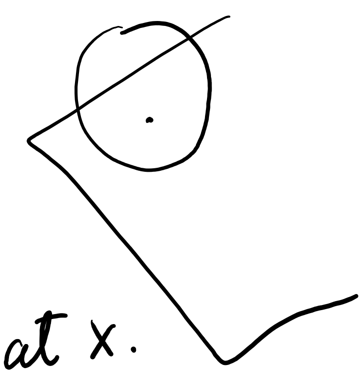
no boundary $\qquad$ need gradient.



What to do at (near) boundary?

reflect? ... ?

the <u>Ball Walk</u> (δ)



At $x \in K$:

| Sample $y \sim \text{unif}(B(x,\delta))$
| if $y \in K$ go to y else stay at x.

Stationary distribution?

<u>Thm</u>. Q uniform is stationary for ball walk in closed, bdd set.

<u>Pf</u>. Suffices to check

$$Q(u)\,P(u \to v) = Q(v)\,P(v \to u)$$

because $\quad \int Q(v)\,P(v \to u)\,dv = \int Q(u)\,P(u \to v)\,dv$

because $Q_1(u) = \int_V Q(v)\, P(v \to u)\, dv = \int_V Q(u)\, P(u \to v)\, dv$

$= Q(u).$

For ball walk

$$P(u \to v) = \begin{cases} \frac{1}{\mathrm{vol}(\delta B)} & \text{if } v \in K \\ 0 & \text{if } v \notin K \\ 1 - \frac{\mathrm{vol}(u + \delta B \cap K)}{\mathrm{vol}(\delta B)} & v = u \end{cases}$$

in fact

$P(u \to v) = P(v \to u).$

and $Q(u) \min\left\{1, \frac{Q(v)}{Q(u)}\right\} = Q(v) \min\left\{1, \frac{Q(u)}{Q(v)}\right\}$

We can view ball walk as i.e. $P(u \to v) = \frac{1}{\mathrm{vol}(\delta B)} \min\left\{1, \frac{Q(v)}{Q(u)}\right\}.$

How quickly does Q_k after k steps converge?

<u>Markov Scheme</u> state space Ω

σ-algebra $\mathcal{A}$: subsets of Ω that are closed under countable unions and complement.

$\forall u \in \Omega, \quad P_u(\cdot) : \mathcal{A} \to [0, 1]$

a prob. measure for each point in Ω.

"next step distribution".

"next-step distribution".

$(\Omega, \mathcal{A}, \{P_u\}, Q_0)$ define a Markov chain

$w_0 \sim Q_0,\ w_1 \sim P_{w_0} \ \dots \ w_i \sim P_{w_{i-1}}$.

<u>Stationary distribution</u>

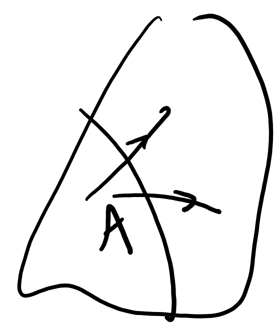
Q is stationary iff $\forall A \in \mathcal{A}$

$$Q(A) = \int_u P_u(A)\, dQ(u)$$

<u>Examples</u> : grid walk, colorings, Coordinate H&R (Gibbs).

<u>Ergodic flow</u>

$$\Phi(A) = \int_{u \in A} P_u(\Omega \setminus A)\, dQ(u)$$



Q is stationary $\equiv \forall A \in \mathcal{A}\ \ \Phi(A) = \Phi(\Omega \setminus A)$

<u>Lazy</u> M.C. : w.p. $\frac{1}{2}$ do nothing. w.p. apply M.C.

<u>Thm</u> (Exercise) If Q is stationary for lazy, <u>ergodic</u> M.C. (all measurable subsets are reachable). then it is the unique stationary dist.

Conductance

"conditional escape probability".

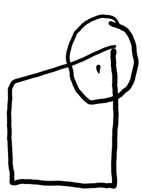
$$\phi(A) = \frac{\Phi(A)}{\min Q(A), Q(\Omega \setminus A)}$$

$$\phi = \inf_{A} \phi(A).$$

mixing rate $\geq \frac{1}{\phi}$.

"local" conductance $\quad \ell(u) = 1 - P_u(\{u\})$.

For the ball walk, $\ell(u) \longrightarrow 0$



so $\phi \longrightarrow 0$, Mixing rate $\to \infty$.

Can we ensure $\ell(u) \geq \ell$? $\quad B \subseteq K \subseteq RB$

$$K' = K + \alpha B^n$$

$$\mathrm{vol}(K + \alpha B^n) \leq \mathrm{vol}((1+\alpha)K)$$
$$\leq (1+\alpha)^n \mathrm{vol}(K)$$
$$\alpha = \frac{\varepsilon}{2n} \longrightarrow (1+\varepsilon)\,\mathrm{vol}(K)$$

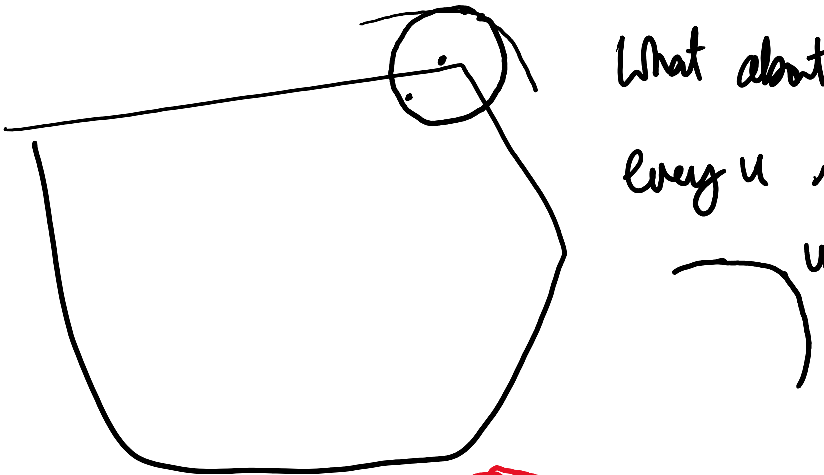
So sampling from K' suffices (+ one rejection step).

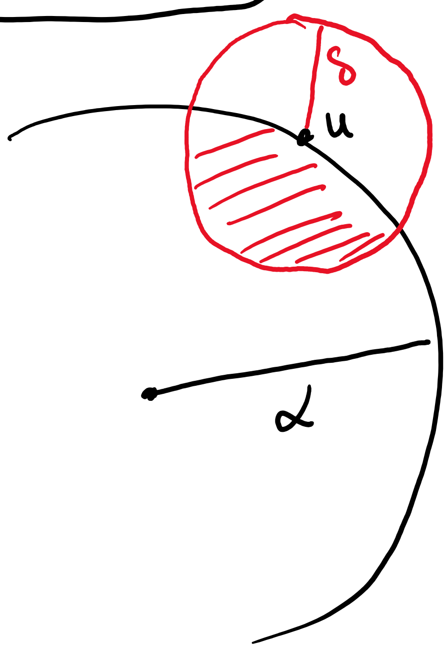
What about $\ell(u)$ for $u \in K'$? $\quad \forall u \in K$, $u + \alpha B^n \subseteq K'$.

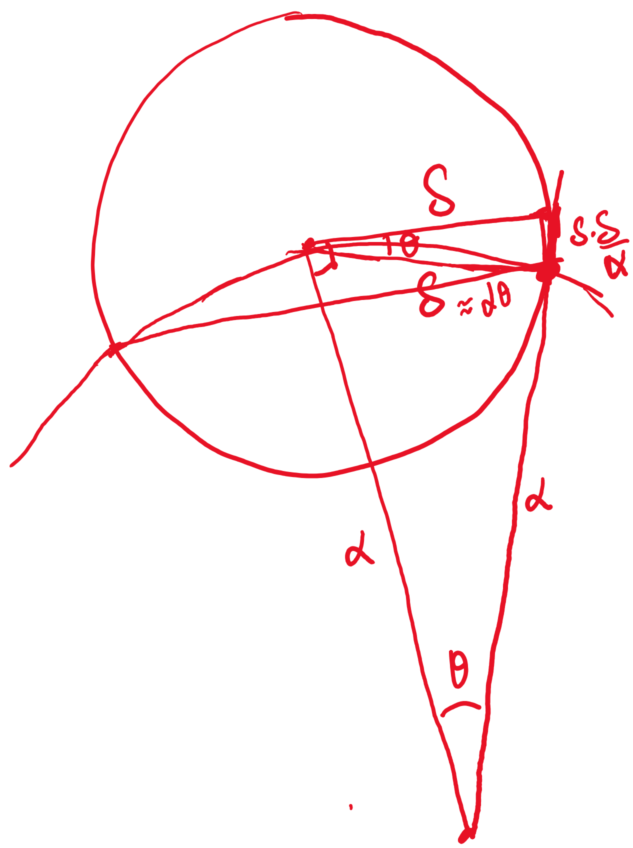
$\ell(u) = 1$ for $\delta \le \alpha$.

What about $u \in K' \setminus K$?

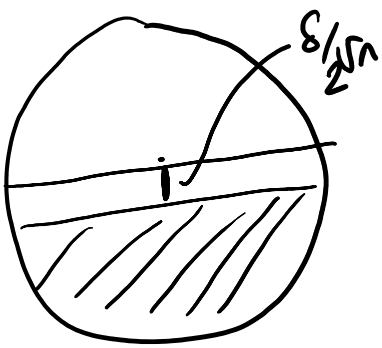
Every u is in ball of radius α contained in K.







$$\alpha = \frac{\delta}{2\sqrt{n}}$$



$$\ell(u) \geq \frac{\mathrm{vol}(\ )}{\mathrm{vol}(\bigcirc)} \geq \frac{1}{8}.$$ (Exercise).

How to bound rate of convergence?

$$\chi^2(P,Q) = \mathbb{E}_Q\left(\left(\frac{dP}{dQ}-1\right)^2\right)$$

$$d_{TV}(P,Q) = \sup_A |P(A) - Q(A)|$$

$$d_{KL}(P,Q) = \mathbb{E}_P\left(\log\frac{dP}{dQ}\right).$$

$$\vdots$$

χ^2 decreases monotonically for a lazy chain!

d_{TV} does not.

"Data processing" inequality

$\forall$ P, Q prob. measures, Markov transition operator M

$$D_f(MP \| MQ) \leq D_f(P \| Q).$$

D_f: f-divergence captures both χ^2 and d_{KL}

$$D_f(P \| Q) = \int f\left(\frac{dP}{dQ}\right) dP$$

$f: \mathbb{R} \to \mathbb{R}$ convex. $f(1) = 0$.

$f: \Omega \to \mathbb{R}$ convex. $f(1) = 0$.

We now bond d_{TV} using a general approach.

Consider $\forall x \in [0,1]$ $\quad \sup_{A:\, Q(A) = x} Q_t(A) - Q(A)$

$$\mathcal{G}_x = \left\{ g: \Omega \to [0,1] : \int_\Omega g(u)\, dQ(u) = x \right\}$$

$$h_t(x) = \sup_{g \in \mathcal{G}_x} \int g(u)\, (dQ_t(u) - dQ(u)) = \int g(u)\, dQ_t(u) - x.$$

Lemma. 1 $\quad h_t(x)$ is concave.

If Q is atom-free, $h_t(x)$ is achieved by $g = \mathbb{1}_A$

Lemma. 2 $\forall t \geq 1$, $\quad y = \min\{x, 1-x\}$, conductance ϕ.

$$h_t(x) \leq \frac{1}{2} h_t(x - 2\phi y) + \frac{1}{2} h_t(x + 2\phi y)$$

Lemma 3 $\quad h_t(x) \leq C \min\{\sqrt{x}, \sqrt{1-x}\} \left(1 - \frac{\phi^2}{2}\right)^t$.

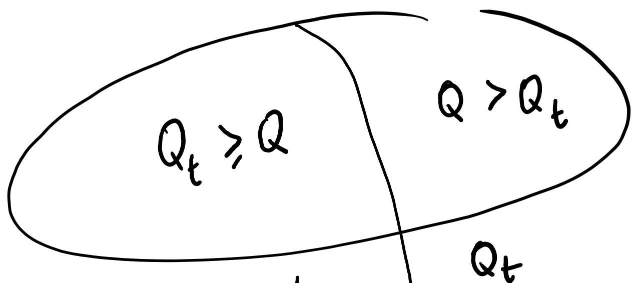
assuming $h_0(x) \leq C \min\{\sqrt{x}, \sqrt{1-x}\}$.

Thm $\quad \ldots \left(1 - \phi^2\right)^t d_{\ldots}(Q_0, Q)$

Thm. $d_{TV}(Q_t, Q) \le \sqrt{M}\left(1-\frac{\phi^2}{2}\right)^t d_{TV}(Q_0, Q)$

$M = \sup \frac{dQ_0}{dQ}$

Pf. (L1). h_t is supremum over linear functions.



Sort by $\frac{Q_t}{Q}$.

(L2). Assume $0 \le x \le \frac{1}{2}$

goal: $h_t(x) \le \frac{1}{2} h_t(x - 2\phi x) + \frac{1}{2} h_t(x + 2\phi x)$

Fix A: $Q_t(A) - Q(A) = h_t(x)$. $Q(A) = x$

$$g_1(u) = \begin{cases} 2P_u(A) - 1 & u \in A \\ 0 & u \notin A \end{cases} \qquad g_2(u) = \begin{cases} 1 & u \in A \\ 2P_u(A) & u \notin A \end{cases}$$

$g_1, g_2 \in \mathcal{G}_x$ (note: lazy $\Rightarrow 2P_u(A) - 1 \ge 0$).

$\frac{1}{2}(g_1 + g_2) = P_u(A)$

$\frac{1}{2}(g_1 + g_2) = P_u(A)$

$$\therefore \; Q_t(A) = \int P_u(A)\, dQ_{t-1}(u) = \frac{1}{2}\int g_1(u)\, dQ_{t-1} + \frac{1}{2}\int g_2(u)\, dQ_{t-1}$$

$$x_1 = \int g_1(u)\, dQ(u) \qquad \frac{1}{2}(x_1 + x_2) = \int P_u(A)\, dQ(u) = Q(A) = x.$$

$$\frac{1}{2}(g_1 + g_2) \in \mathcal{G}_x.$$

$$h_t(x) = Q_t(A) - Q(A)$$

$$= \frac{1}{2}\int g_1(u)\, dQ_{t-1}(A) + \frac{1}{2}\int g_2(u)\, dQ_{t-1}(A) - x$$

$$= \frac{1}{2}\int g_1(u)\left(dQ_{t-1}(A) - dQ(A)\right) + \frac{1}{2}\int g_2(u)\left(dQ_{t-1}(A) - dQ(A)\right)$$

$$\leq \frac{1}{2} h_{t-1}(x_1) + \frac{1}{2} h_{t-1}(x_2).$$

$$x_1 = \int_\Omega g_1(u)\, dQ(u) = 2\int_A P_u(A)\, dQ(u) - x$$

$$= 2\int_A \left(1 - P_u(\Omega \setminus A)\right) dQ(u) - x$$

$$= x - 2\int_A P_u(\Omega \setminus A)\, dQ(u)$$

$$\le x - 2\phi x .$$

So $x_1 \le x(1-2\phi) \le x \le x(1+2\phi) \le x_2$.

and by concavity $h_t(x) \le \frac{1}{2} h_{t-1}(x(1-2\phi)) + \frac{1}{2} h_{t-1}(x(1+2\phi))$

Pf. (L3) $h_0 \le C\sqrt{x}$

induction:

$$h_t(x) \le \frac{1}{2} h_{t-1}(x(1-2\phi)) + \frac{1}{2} h_{t-1}(x(1+2\phi))$$

$$\le C\left(\frac{1}{2}\sqrt{x(1-2\phi)} + \frac{1}{2}\sqrt{x(1+2\phi)}\right)\left(1-\frac{\phi^2}{2}\right)^{t-1}$$

$$\le C\sqrt{x}\left(\frac{1}{2}\sqrt{1-2\phi} + \frac{1}{2}\sqrt{1+2\phi}\right)\left(1-\frac{\phi^2}{2}\right)^{t-1}$$

$$\le C\sqrt{x}\left(1-\frac{\phi^2}{2}\right)^{t} .$$

Pf (Thm). $\frac{dQ_0}{dQ} \le M$ $h_0(x) = \sup Q_t(A) - Q$

$\le (M-1)x \le M$ ..

$\Rightarrow Q_0(A) \le M\,Q(A)$
