CPSC 661: Sampling Algorithms in ML

Andre Wibisono

March 3, 2021

Yale University

References

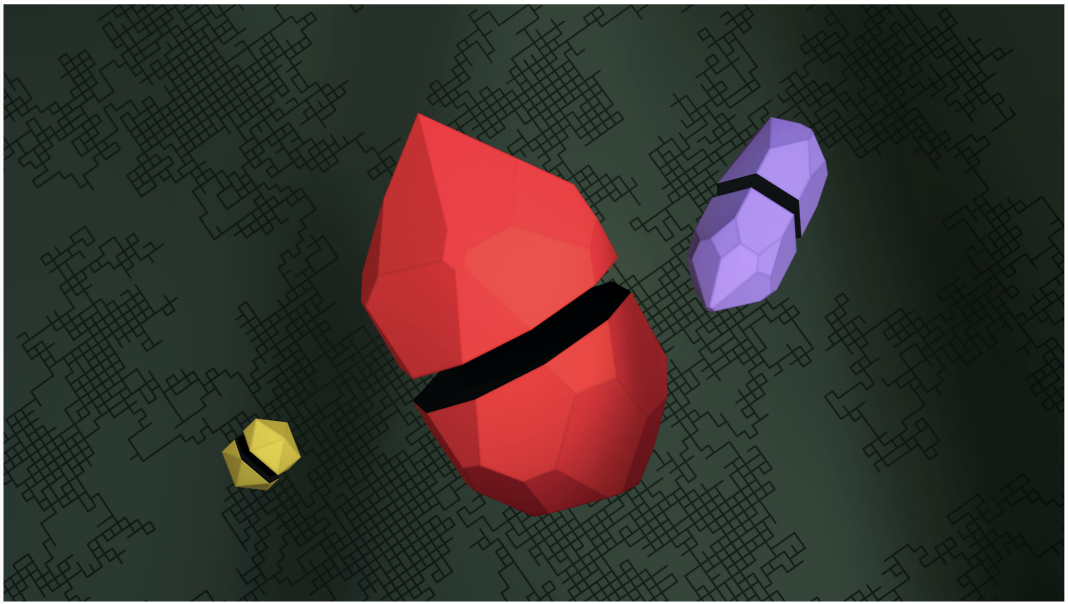
- Dwivedi, Chen, Wainwright, and Yu, *Log-Concave Sampling: Metropolis-Hastings Algorithms are Fast*, Journal of Machine Learning Research, 2019

Aside: KLS Conjecture

GEOMETRY

Statistics Postdoc Tames Decades-Old Geometry Problem

To the surprise of experts in the field, a postdoctoral statistician has solved one of the most important problems in high-dimensional convex geometry.



Olena Shmahalo/Quanta Magazine

https://www.quantamagazine.org/
statistics-postdoc-tames-decades-old-geometry-problem-20210301

Recap

To sample from ν on $\mathbb{R}^n$:

1. Start from any Markov chain P
2. Apply Metropolis-Hastings filter to get $\tilde{P}$ reversible wrt ν
3. Assume ν is α-SLC, so isoperimetric with $\psi = \Omega(\sqrt{\alpha})$

   Let $\mathcal{R}_s = \mathbb{B}(x^*, r(s)\sqrt{\frac{n}{\alpha}})$, so $\nu(\mathcal{R}_s) \geq 1 - s$
4. Show $\tilde{P}$ satisfies one-step overlap property:

$$x, y \in \mathcal{R}_s,\ \|x - y\|_2 \leq \Delta_s \quad \Rightarrow \quad \mathrm{TV}(\tilde{P}_x, \tilde{P}_y) \leq \frac{3}{4}$$

$\Rightarrow$ $\tilde{P}$ has s-conductance

$$\phi_s \geq \min\left\{\frac{1}{16}, \frac{\sqrt{\alpha}\,\Delta_s}{128}\right\} = \Omega(\sqrt{\alpha}\,\Delta_s).$$

$\Rightarrow$ mixing time in TV distance: $TV(\rho_k, \nu) \leq \varepsilon$

$$\tau(\epsilon) = \frac{2}{\phi_s^2}\log\frac{2M}{\epsilon} = O\left(\frac{1}{\alpha\,\Delta_s^2}\log\frac{2M}{\epsilon}\right)$$

where $s = \dfrac{\epsilon}{2M}$ and $M = M_\nu^\infty(\rho_0)$ is warm-start

What random walk?

1. $P =$ Brownian motion (Gaussian walk)
   $\Rightarrow \tilde{P} =$ Metropolis Random Walk (MRW)
   (Today)

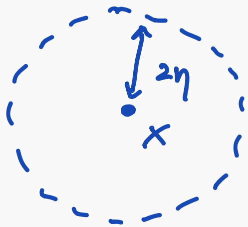
2. $P =$ Unadjusted Langevin Algorithm (ULA)
   $\Rightarrow \tilde{P} =$ Metropolis-Adjusted Langevin Algorithm (MALA)
   (Next time)

Metropolis Random Walk (MRW)

① P = Brownian motion (Gaussian walk) on $\mathbb{R}^n$

with step size $\eta > 0$.

- Random walk: $x' = x + \sqrt{2\eta}\, Z$, $\quad Z \sim \mathcal{N}(0, I)$ independent
- Markov chain: $P_x = \mathcal{N}(x, 2\eta I)$



$$P_x(y) = \frac{1}{(4\pi\eta)^{n/2}}\, e^{-\frac{\|x-y\|^2}{4\eta}}$$

- symmetric: $P_x(y) = P_y(x)$
- stationary distribution = Lebesgue measure

+ Metropolis-Hastings filter for ν

(2) $\tilde{P}$ = Metropolis Random Walk (MRW)

- Random walk:
  - from x, draw $y = x + \sqrt{2\eta}\, Z$, $Z \sim \mathcal{N}(0, I)$ independent
  - Compute acceptance probability

$$a_x(y) = \min\left\{1, \frac{\nu(y)\cdot \cancel{P_y(x)}}{\nu(x)\cdot \cancel{P_x(y)}}\right\}$$

$$= \min\left\{1, \frac{\nu(y)}{\nu(x)}\right\}$$

  - move to $x' = \begin{cases} y & \text{with probability } a_x(y) \\ x & \text{with probability } 1 - a_x(y) \end{cases}$

- Markov chain:

$$\tilde{P}_x(y) = a_x(y)\cdot P_x(y) + A(x)\cdot \delta_x(y)$$

where $A(x) = 1 - \int_{\mathbb{R}^n} a_x(y)\cdot P_x(y)\, dy$

- Reversible wrt ν $\Rightarrow$ stationary distribution is ν.

- Zero-order: only depends on ν (up to a constant)
  (not on gradient)

Analysis of MRW

P = Brownian motion, $\tilde{P}$ = MRW

TV is a distance metric

$\Rightarrow$ satisfies triangle inequality:

$$TV(\tilde{P}_x, \tilde{P}_y) \leq \underbrace{TV(\tilde{P}_x, P_x)}_{1} + \underbrace{TV(P_x, P_y)}_{2} + \underbrace{TV(P_y, \tilde{P}_y)}_{3}$$

Plan:

- for 1 and 3 : bound acceptance probability of MH
- for 2 : bound via KL divergence

Kullback-Leibler (KL) Divergence

Let ρ and ν be probability distributions on $\mathcal{X}$.

The **Kullback-Leibler (KL) divergence** of ρ with respect to ν is

$$H_\nu(\rho) = \int_{\mathcal{X}} \rho(x) \log \frac{\rho(x)}{\nu(x)}\, dx$$

- Also called *relative entropy*
- Relative form of *Shannon entropy*: $H(\rho) = -\int_{\mathcal{X}} \rho(x) \log \rho(x)\, dx$
- Non-negative: $H_\nu(\rho) \geq 0$, and $H_\nu(\rho) = 0$ if and only if $\rho = \nu$
- *Not* a metric: Not symmetric, does not satisfy triangle inequality
- **Pinsker's Inequality**:

$$\boxed{\mathrm{TV}(\rho, \nu) \leq \sqrt{\frac{1}{2} H_\nu(\rho)}}$$

Lemma

If $\rho = \mathcal{N}(\mu_1, \Sigma)$ and $\nu = \mathcal{N}(\mu_2, \Sigma)$, then

$$H_\nu(\rho) = \frac{1}{2}(\mu_1 - \mu_2)^\top \Sigma^{-1}(\mu_1 - \mu_2)$$

Proof:

$$\log \frac{\rho(x)}{\nu(x)} = -\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) - \frac{1}{2}\log \cancel{\det}(2\pi\Sigma)$$
$$+ \frac{1}{2}(x-\mu_2)^T \Sigma^{-1}(x-\mu_2) + \frac{1}{2}\log\cancel{\det}(2\pi\Sigma)$$

$$= x^T \Sigma^{-1}(\mu_1-\mu_2) - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

then $H_\nu(\rho) = \mathbb{E}_\rho\left[\log \frac{\rho(X)}{\nu(X)}\right]$

$$= \underbrace{\mathbb{E}_\rho[X]}_{\mu_1}{}^T \Sigma^{-1}(\mu_1-\mu_2) - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

$$= \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}\mu_2 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

$$= \frac{1}{2}(\mu_1-\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)$$

$\square$

One-step overlap of Brownian motion

Let P = Brownian motion with step size η

$$P_x = \mathcal{N}(x, 2\eta I)$$

<u>Lemma:</u> If $\|x-y\|_2 \le \sqrt{2\eta}$, then $TV(P_x, P_y) \le \frac{1}{2}$.

<u>Proof:</u> By Pinsker's inequality:

$$TV(P_x, P_y) \le \sqrt{\frac{1}{2} H_{P_y}(P_x)}$$

$$= \sqrt{\frac{1}{2} \cdot \frac{1}{2}(x-y)^T \left(\frac{1}{2\eta} I\right)(x-y)}$$

$$= \sqrt{\frac{\|x-y\|_2^2}{8\eta}}$$

$$= \frac{\|x-y\|_2}{2\sqrt{2\eta}}$$

$$\le \frac{1}{2}. \qquad \square$$

Strong log-concavity and log-smoothness

Let $\nu \propto e^{-f}$ be a probability distribution on $\mathbb{R}^n$

Recall ν is α-**strongly log-concave** if f is α-strongly convex:

$$\nabla^2 f(x) \succeq \alpha I$$

Recall ν is L-**log-smooth** if f is L-smooth:

$$\nabla^2 f(x) \preceq L I$$

If ν is α-SLC and L-log-smooth, then the **condition number** is

$$\kappa = \frac{L}{\alpha} \geq 1$$

Bounding acceptance probability of MRW

Let P = Brownian motion with step size η

$\tilde{P}$ = MRW = P + Metropolis-Hastings for ν

Lemma: Assume ν is α-SLC and L-log-smooth on $\mathbb{R}^n$.

If $x \in R_s = B\left(x^*, r(s)\sqrt{\frac{n}{\alpha}}\right)$

where $r(s) = 2 + 2\max\left\{\left(\frac{1}{n}\log\frac{1}{s}\right)^{\frac{1}{4}}, \left(\frac{1}{n}\log\frac{1}{s}\right)^{\frac{1}{2}}\right\}$

and $\eta \le \dfrac{\alpha}{10^5 \cdot n \cdot L^2 \cdot r(s)^2}$

then $TV(\tilde{P}_x, P_x) \le \frac{1}{8}$.

Proof:

Recall $\tilde{P}_x(y) = a_x(y) \cdot P_x(y) + A(x) \cdot \delta_x(y)$

where $A(x) = 1 - \int_{\mathbb{R}^n} a_x(y) \cdot P_x(y)\, dy$

TV distance is

$$TV(\tilde{P}_x, P_x) = \frac{1}{2} \int_{\mathbb{R}^n} |\tilde{P}_x(y) - P_x(y)|\, dy$$

$$= \frac{1}{2}\Big(A(x) + \underbrace{\int_{\mathbb{R}^n} (1 - a_x(y)) P_x(y)\, dy}_{= 1 - \int_{\mathbb{R}^n} a_x(y) P_x(y)\, dy \; = A(x)} \Big)$$

$$= A(x)$$

$$= 1 - \mathbb{E}_{Y \sim P_x} [a_x(Y)]$$

$$= 1 - \mathbb{E}_{Y \sim P_x} \left[\min\left\{1, \frac{\nu(Y)}{\nu(x)}\right\} \right]$$

$$= 1 - \mathbb{E}_{Y \sim P_x} \left[\min\left\{1, e^{f(x) - f(Y)}\right\} \right]$$

By Markov inequality, $\forall\ 0 < t \leq 1$:

$$\mathbb{E}\left[\min\left\{1, e^{f(x)-f(Y)}\right\}\right] \geq t \cdot \mathbb{P}\left(\min\left\{1, e^{f(x)-f(Y)}\right\} \geq t\right)$$

$$\geq t \cdot \mathbb{P}\left(e^{f(x)-f(Y)} \geq t\right)$$

$$= t \cdot \mathbb{P}\left(f(x) - f(Y) \geq \log t\right)$$

We will prove high-probability bound on $f(x) - f(Y)$

(x fixed, $Y \sim P_x$, which means $Y = x + \sqrt{2\eta}\, Z$)

We have:

$$f(x) - f(Y) \geq \nabla f(Y)^T (x - Y) \qquad \text{because } f \text{ is convex}$$

$$= \nabla f(x)^T (x - Y) - (\nabla f(x) - \nabla f(Y))^T (x - Y)$$

$$\geq \nabla f(x)^T (x - Y) - L \| x - Y \|^2 \qquad \text{since } f \text{ is } L\text{-smooth}$$

$$\Big(f \ L\text{-smooth} \iff \nabla^2 f(x) \preceq L \cdot I$$

$$\iff \| \nabla f(x) - \nabla f(y) \|_2 \leq L \cdot \| x - y \|_2$$

$$\iff (\nabla f(x) - \nabla f(y))^T (x - y) \leq L \cdot \| x - y \|_2^2 \Big)$$

$Y \sim P_x$ which means $Y = x + \sqrt{2\eta}\, Z$, $Z \sim \mathcal{N}(0, I)$

$$f(x) - f(Y) \geq \sqrt{2\eta}\, \underbrace{\left(-\nabla f(x)^T Z \right)}_{\mathrm{I}} - L \cdot 2\eta \cdot \underbrace{\| Z \|^2}_{\mathrm{II}}$$

* for I : $Z \sim \mathcal{N}(0, I)$ on $\mathbb{R}^n$

$$-\nabla f(x)^T Z \sim \mathcal{N}(0, \| \nabla f(x) \|^2) \quad \text{on } \mathbb{R}^1$$

since $x \in R_s = \mathbb{B}\left(x^*, r(s) \sqrt{\frac{n}{\alpha}} \right)$

$$\Rightarrow \| \nabla f(x) \| = \| \nabla f(x) - \nabla f(x^*) \| \qquad \text{since } \nabla f(x^*) = 0$$

$$\leq L \cdot \| x - x^* \| \qquad \text{by smoothness}$$

$$\leq L \cdot r(s) \cdot \sqrt{\frac{n}{\alpha}} := D_s$$

by tail bound for 1-dimensional Gaussian,

$$\mathbb{P}\left(-\nabla f(x)^T Z \geq -2 D_s \sqrt{\log \frac{1}{\varepsilon}} \right) \geq 1 - \varepsilon \qquad \forall\, \varepsilon > 0$$

* for II : $z \sim \mathcal{N}(0, I)$

$\|z\|_2^2 = z_1^2 + \dots + z_n^2 \sim \chi^2$ distribution with n degrees of freedom

sub-exponential random variable $\Rightarrow$ tail bound

$\mathbb{E}[\|z\|^2] = n$

$\Rightarrow$ concentration around n

Tail bound for χ^2-random variable

Lemma
Let W be a χ^2-random variable with n degrees of freedom. For all $\epsilon > 0$: " $\|z\|^2$

$$\Pr(W \leq n\beta_\epsilon) \geq 1 - \epsilon$$

where $\beta_\epsilon = 1 + 2\sqrt{\log(1/\epsilon)} + 2\log(1/\epsilon)$

- Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge University Press, 2019

  (Draft of Chapter 2 available at: `https://www.stat.berkeley.edu/~mjwain/stat210b/Chap2_TailBounds_Jan22_2015.pdf`)

Bounding acceptance probability of MRW (continued)

From above,

$$f(x) - f(Y) \geq \sqrt{2\eta}\left(-\nabla f(x)^T z\right) - 2\eta \cdot L \cdot \|z\|_2^2$$

$$\geq -2\sqrt{2\eta} \cdot L \cdot r(s)\sqrt{\frac{n}{\alpha}} \cdot \sqrt{\log \frac{1}{\varepsilon}} - 2\eta \cdot L \cdot n \cdot \beta_\varepsilon \quad \Bigg\} \text{ with probability } \geq 1 - \varepsilon - \varepsilon = 1 - 2\varepsilon$$

want $\geq -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} = -\varepsilon$

this happens if

$$\boxed{\eta \leq \frac{\varepsilon^2 \cdot \alpha}{32\, L^2\, r(s)^2\, n \cdot \log \frac{1}{\varepsilon}} \quad \text{and} \quad \eta \leq \frac{\varepsilon}{4\, n\, L\, \beta_\varepsilon}} \qquad (*)$$

then $f(x) - f(Y) \geq -\varepsilon$ with probability $\geq 1 - 2\varepsilon$.

so that with $t = e^{-\varepsilon}$ ($\log t = -\varepsilon$)

$$
\begin{aligned}
\mathbb{E}[\min\{1, e^{f(x)-f(Y)}\}] &\geq t \cdot \mathbb{P}(f(x) - f(Y) \geq \log t) \\
&= e^{-\varepsilon} \cdot \mathbb{P}(f(x) - f(Y) \geq -\varepsilon) \\
&\geq e^{-\varepsilon} \cdot (1 - 2\varepsilon) \\
&\geq (1-\varepsilon) \cdot (1 - 2\varepsilon) \\
&\geq (1 - 3\varepsilon) \\
\text{want} &= \frac{7}{8} \iff \boxed{\varepsilon = \frac{1}{24}}
\end{aligned}
$$

so that

$$
\begin{aligned}
TV(\tilde{P}_x, P_x) &= 1 - \mathbb{E}[\min\{1, e^{f(x)-f(Y)}\}] \\
&\leq 1 - \frac{7}{8} = \frac{1}{8}.
\end{aligned}
$$

which is what we want.

for $\varepsilon = \frac{1}{24}$:

$$\log \frac{1}{\varepsilon} = \log 24 \approx 3.2 < 4$$

$$\beta_\varepsilon = 1 + 2\sqrt{\log \frac{1}{\varepsilon}} + 2 \log \frac{1}{\varepsilon} = 10.9 < 11$$

so $\eta \leq \dfrac{1}{1056 \cdot nL} \leq \dfrac{\varepsilon}{4nL\beta_\varepsilon}$

and $\eta \leq \dfrac{\alpha}{10^5\, n L^2 r(s)^2} \leq \dfrac{\varepsilon^2 \alpha}{32 L^2 r(s)^2 n \log \frac{1}{\varepsilon}}$

} which is (*)

□

One-step overlap of MRW

Combining the steps above, we have the following:

Assume $\nu \propto e^{-f}$ on $\mathbb{R}^n$ is α-SLC and L-log-smooth.

Lemma
Let the step size be

$$\eta \leq \frac{\alpha}{10^5 \cdot n \cdot L \cdot r(s)^2} .$$

If $x, y \in \mathcal{R}_s$ and $\|x - y\|_2 \leq \sqrt{2\eta}$, then

$$TV(\tilde{P}_x, \tilde{P}_y) \leq \frac{3}{4} .$$

Proof: By the previous two lemmas, we have

$$TV(\tilde{P}_x, P_x) \leq \frac{1}{8}$$

$$TV(P_x, P_y) \leq \frac{1}{2}$$

$$TV(\tilde{P}_y, P_y) \leq \frac{1}{8}$$

Then $$\mathrm{TV}(\tilde{P}_x, \tilde{P}_y) \le \mathrm{TV}(\tilde{P}_x, P_x) + \mathrm{TV}(P_x, P_y) + \mathrm{TV}(P_y, \tilde{P}_y)$$
$$\le \frac{1}{8} + \frac{1}{2} + \frac{1}{8}$$
$$= \frac{3}{4}.$$

$\square$

Note: We can choose $\eta = \dfrac{c \cdot \alpha}{n \cdot L^2 \cdot r(s)^2} = \Theta\left(\dfrac{\alpha}{n \cdot L^2 \cdot r(s)^2}\right)$ for small enough constant c $(\le 10^{-5})$.

Then we can choose the distance threshold Δ_s in the one-step overlap property to be

$$\Delta_s = \sqrt{2\eta} = \Theta\left(\frac{\sqrt{\alpha}}{\sqrt{n} \cdot L \cdot r(s)}\right).$$

We can now plug this in to our mixing time bound.

Mixing time of MRW

Assume $\nu \propto e^{-f}$ on $\mathbb{R}^n$ is α-SLC and L-log-smooth.

Theorem
Choose step size

$$\eta = \Theta\left(\frac{\alpha}{n \cdot L^2 \cdot r(s)^2}\right)$$

Starting from ρ_0 with $M = M_\nu^\infty(\rho_0) < \infty$, the mixing time of MRW is

$$\tau(\varepsilon) = O\left(\frac{1}{\alpha \cdot \Delta_s^2} \cdot \log \frac{2M}{\varepsilon}\right)$$

$$= O\left(\frac{n \cdot L^2 \cdot r(s)^2}{\alpha^2} \cdot \log \frac{2M}{\varepsilon}\right)$$

$$= O\left(n \cdot \kappa^2 \cdot r(s)^2 \cdot \log \frac{2M}{\varepsilon}\right)$$

where $\kappa = \frac{L}{\alpha}$ is condition number, and $s = \frac{\varepsilon}{2M}$.

Warm start

Now let us derive a bound on the warmness parameter $M = M_\nu^\infty(\rho_0)$.

Assume $\nu \propto e^{-f}$ on $\mathbb{R}^n$ is α-SLC and L-log-smooth

Let $\rho_0 = \mathcal{N}(x^*, \frac{1}{L}I)$ where $x^* =$ mode of $\nu =$ minimizer of f

Lemma

$$M_\nu^\infty(\rho_0) \le \kappa^{n/2}$$

(Note: Also with approximate mode x^*, see [DCWY'19, Sec. 3.2.1])

Proof of Lemma:

By strong convexity and smoothness, for all $x \in \mathbb{R}^n$ we have

$$\frac{\alpha}{2}\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{L}{2}\|x - x^*\|^2$$

Then we can bound the normalizing constant for ν:

$$\int_{\mathbb{R}^n} e^{-f(x)}\,dx \le e^{-f(x^*)} \int_{\mathbb{R}^n} e^{-\frac{\alpha}{2}\|x-x^*\|^2}\,dx$$

$$= e^{-f(x^*)} \cdot \left(\frac{2\pi}{\alpha}\right)^{n/2} .$$

We can also bound the density of ν:

$$\nu(x) = \frac{e^{-f(x)}}{\int_{\mathbb{R}^n} e^{-f(x)}\,dx} \ge \frac{\cancel{e^{-f(x^*)}} \cdot e^{-\frac{L}{2}\|x-x^*\|^2}}{\cancel{e^{-f(x^*)}} \cdot \left(\frac{2\pi}{\alpha}\right)^{n/2}} .$$

Then we can bound the ratio of the densities:

$$\frac{\rho(x)}{\nu(x)} = \frac{e^{-\frac{L}{2}\|x-x^*\|^2}}{\left(\frac{2\pi}{L}\right)^{n/2}} \cdot \frac{1}{\nu(x)}$$

$$\le \frac{\cancel{e^{-\frac{L}{2}\|x-x^*\|^2}}}{\left(\frac{2\pi}{L}\right)^{n/2}} \cdot \frac{\left(\frac{2\pi}{\alpha}\right)^{n/2}}{\cancel{e^{-\frac{L}{2}\|x-x^*\|^2}}}$$

$$= \left(\frac{L}{\alpha}\right)^{n/2} = K^{n/2} .$$

So $M_\nu^\infty(\rho_0) = \sup_{x\in\mathbb{R}^n} \left| \frac{\rho(x)}{\nu(x)} - 1 \right| \le K^{n/2} - 1 \le K^{n/2}$.

$\square$

Now with $M = K^{n/2}$, we can bound (ignoring constants):

$$s = \frac{\varepsilon}{2M} = \frac{\varepsilon}{2K^{n/2}}$$

$$\Rightarrow \log \frac{2M}{\varepsilon} = \log \frac{1}{s} \sim \frac{n}{2} \log K + \log \frac{1}{\varepsilon} \sim n \log\left(\frac{K}{\varepsilon^{1/n}}\right)$$

And $$r(s) = 2 + 2 \max\left\{ \left(\frac{1}{n} \log \frac{1}{s}\right)^{\frac{1}{4}}, \left(\frac{1}{n} \log \frac{1}{s}\right)^{\frac{1}{2}} \right\}$$

$$\sim \left(\frac{1}{n} \log \frac{1}{s}\right)^{\frac{1}{2}}$$

$$\sim \sqrt{\log\left(\frac{K}{\varepsilon^{1/n}}\right)}$$

Then the mixing time bound becomes:

$$\tau(\varepsilon) = O\left(n \cdot K^2 \cdot r(s)^2 \cdot \log \frac{2m}{\varepsilon}\right)$$

$$= O\left(n^2 \cdot K^2 \cdot \log^2\left(\frac{K}{\varepsilon^{1/n}}\right)\right).$$

Recap for MRW

To sample from $\nu \propto e^{-f}$ on $\mathbb{R}^n$ which is α-SLC, L-log-smooth

MRW algorithm:

1. Start from $x_0 \sim \rho_0 = \mathcal{N}(x^*, \frac{1}{L} I)$
2. Set step size $\eta = c \dfrac{\alpha}{nL^2 \log(\kappa/\epsilon^{1/n})} = \tilde{O}\left(\dfrac{\alpha}{nL^2}\right)$ for small enough c
3. For $k = 0, 1, 2, \ldots$:
   - Draw $y_k = x_k + \sqrt{2\eta}\, z_k$, $z_k \sim \mathcal{N}(0, I)$ independent
   - Set $x_{k+1} = y_k$ with prob $\min\{1, e^{f(x_k) - f(y_k)}\}$, else $x_{k+1} = x_k$

Guarantee: $x_k \sim \rho_k$ satisfies $\mathrm{TV}(\rho_k, \nu) \le \epsilon$ for

$$k \ge c' n^2 \kappa^2 \log^{\cancel{1.5}\,2}\left(\frac{\kappa}{\epsilon^{1/n}}\right) = \tilde{O}\left(n^2 \kappa^2\right)$$

for some constant c'.