CPSC 661: Sampling Algorithms in Machine Learning	March 3, 2021
Lecture 9	
Lecturer: Andre Wibisono	*Scribe: Eli Levin*

The primary reference for this material is Log-Concave Sampling: Metropolis-Hastings Algorithms are Fast (**JMLR**).

Aside: Of independent interest, check out Postdoc Tames Decades Old Geometry Problem regarding the KLS Conjecture for some work by Yuansi Chen, one of the authors of the main reference, when he was a postdoc.)

Recap

The problem is that we want to sample from a distribution (ν) on the Euclidean space ($\mathbb{R}^n$) but cannot easily access ν directly. We can start from any Markov chain (P) - doesn't have to have anything to do with ν - and then we force it to be reversible by applying the Metropolis-Hastings filter (i.e. the accept/reject step).

Then, we want to analyze the convergence of this new MC ($\tilde{P}$). By assuming our target distribution (ν) has some nice properties - cardinally, that it is α-strongly log-concave (α-SLC) - we can infer the (dimensionless) isoperimetry constant ($\psi = \Omega\left(\sqrt{\alpha}\right)$).

By defining a ball ($\mathcal{R}_s$), centered at the mode of the target distribution, with radius a function of the *effective radius of the distribution* ($\sqrt{n/\alpha}$) and a volume-preserving function ($r(s)$), we can demonstrate that for sufficiently small steps within the ball (i.e. $\leq \Delta_s$), the new MC satisfies the *one-step overlap* property.

Having demonstrated that $\tilde{P}$ abides by the one-step overlap property, we may bound the *s-conductance* (ϕ_s) as a function of the isoperimetry constant and, in turn, bound the mixing time ($\tau(\cdot)$) in Total Variational (TV) distance. Figure 1. restates the above in a procedural fashion.

Today's Goal

Having established the general procedure for generating $\tilde{P}$ such that it converges to within TV distance ε of the target distribution ν in a number of iterations on the order of $\tau(\varepsilon)$, today we will make this procedure concrete by discussing how to appropriately set the threshold Δ_s and the warm-start parameter M.

We'll do this for two random walks ($P \to \tilde{P}$): Today is Brownian motion (Gaussian Walk) $\to$ Metropolis Random Walk (MRW); and next time is Unadjusted Langevin Algorithm (ULA) $\to$ Metropolis-Adjusted Langevin Algorithm (MALA).

Want 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 with respect to ν
3. Assume ν is α–SLC $\Rightarrow$ isoperimetric with $\psi = \Omega\left(\sqrt{\alpha}\right)$. Let (the ball) $\mathcal{R}_s$ be defined as:

$$\mathcal{R}_s = \mathbb{B}\left(x^*, r(s)\sqrt{\frac{n}{\alpha}}\right) \implies \nu\left(\mathcal{R}_s\right) \geq 1 - s$$

4. Show $\tilde{P}$ satisfies one-step overlap property:

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

NB: The value of the rhs fraction is not important so long as we can bound away from 1.

---

one-step overlap $\Rightarrow$ $\tilde{P}$ has s-conductance:

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

lower bound on s-conductance $\Rightarrow$ mixing time in TV distance: *NB:* $\mathrm{TV}\left(\rho_k, \nu\right) \leq \epsilon$

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

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

Figure 1: Recap

Metropolis Random Walk (MRW)

Brownian Motion (Gaussian walk)

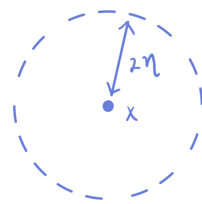
For P = Brownian Motion (Gaussian walk) on $\mathbb{R}^n$ with step size $\eta > 0$, the **random walk** is defined by

$$x' = x + \sqrt{2\eta} Z, \quad Z \sim \mathcal{N}(0, I) \text{ independent}$$

and the corresponding **Markov chain** by

$$\mathcal{P}_x = \mathcal{N}\left(x, 2\eta I\right)$$

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



This is **symmetric** in that

$$\mathcal{P}_x(y) = \mathcal{P}_y(x)$$

with **stationary distribution** = Lebesgue measure.

NB: We get the following section by adding the Metropolis-Hastings filter for ν

Metropolis Random Walk (MRW)

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

Random Walk + Metropolis-Hastings filter

1. from x, draw $y = x + \sqrt{2\eta} Z$, $Z \sim \mathcal{N}\left(0, I\right)$ independent

2. compute acceptance probability

$$\begin{aligned} a_x(y) &= \min\left\{1, \frac{\nu(y) \cdot \cancel{\mathcal{P}_y(x)}}{\nu(x) \cdot \cancel{\mathcal{P}_x(y)}}\right\} \\ &= \min\left\{1, \frac{\nu(y)}{\nu(x)}\right\} \end{aligned}$$

3. move to

$$x' = \begin{cases} y & w.p. \ a_x(y) \\ x & w.p. \ 1 - a_x(y) \end{cases}$$

Applying the above procedure results in a Markov chain $\tilde{P}$ such that

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

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

This MC is ***reversible*** with respect to $\nu (\Rightarrow$ stationary distribution is ν). It is also ***zero-order***, meaning it only depends on ν (up to a constant) and does *not* depend on the gradient.

Analysis

Let P and $\tilde{P}$ be defined as the MCs corresponding to Brownian motion and MRW, respectively. Leveraging the fact that TV is a metric (and therefore satisfies the triangle inequality), we can observe the following bound:

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

Plan:

- for ① and ③: bound acceptance probability of Metropolis-Hastings filter
- 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 known as ***relative entropy***, KL-divergence is a relative form of ***Shannon entropy***

$$H(\rho) = -\int_{\mathcal{X}} \rho(x) \log \rho(x) dx.$$

It's **non-negative** in that $H_\nu(\rho) \geq 0$, and $H_\nu(\rho) = 0$ if and only if $\rho = \nu$. And is **not a metric** as it is not symmetric and does not satisfy the triangle inequality.

A useful property based on KL-divergence is **Pinsker's inequality**:

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

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

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

Proof.

$$\begin{aligned}
\log \frac{\rho(x)}{\nu(x)} &= -\frac{1}{2}\left(x - \mu_1\right)^\top \Sigma^{-1}\left(x - \mu_1\right) - \cancel{\frac{1}{2}\log\det\left(2\pi\Sigma\right)} \\
&\quad + \frac{1}{2}\left(x - \mu_1\right)^\top \Sigma^{-1}\left(x - \mu_1\right) + \cancel{\frac{1}{2}\log\det\left(2\pi\Sigma\right)} \\
&= x^\top \Sigma^{-1}\left(\mu_1 - \mu_2\right) - \frac{1}{2}\mu_1^\top \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^\top \Sigma^{-1}\mu_2
\end{aligned}$$

then

$$\begin{aligned}
H_\nu(\rho) &= \mathbb{E}_\rho\left[\log \frac{\rho(x)}{\nu(x)}\right] \\
&= \underbrace{\mathbb{E}_\rho\left[X\right]^\top}_{\mu_1} \Sigma^{-1}\left(\mu_1 - \mu_2\right) - \frac{1}{2}\mu_1^\top \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^\top \Sigma^{-1}\mu_2 \\
&= \frac{1}{2}\mu_1^\top \Sigma^{-1}\mu_1 - \mu_1^\top \Sigma^{-1}\mu_2 + \frac{1}{2}\mu_2^\top \Sigma^{-1}\mu_2 \\
&= \frac{1}{2}\left(\mu_1 - \mu_2\right)^\top \Sigma^{-1}\left(\mu_1 - \mu_2\right)
\end{aligned}$$

$\square$

One-step overlap of Brownian motion

Let P = Brownian motion with step size η such that

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

Lemma. *If* $\|x - y\|_2 \leq \sqrt{2\eta}$, *then*

$$\mathrm{TV}\left(P_x, P_y\right) \leq \frac{1}{2}.$$

Proof. By Pinsker's inequality:

$$\begin{aligned}
\mathrm{TV}\left(P_x, P_y\right) &\leq \sqrt{\frac{1}{2} H_{P_y}\left(P_x\right)} \\
&= \sqrt{\frac{1}{2} \cdot \frac{1}{2}\left(x-y\right)^{\top}\left(\frac{1}{2\eta} I\right)\left(x-y\right)} \\
&= \sqrt{\frac{\|x-y\|_2^2}{8\eta}} \\
&= \frac{\|x-y\|_2}{2\sqrt{2\eta}} \\
&\leq \frac{1}{2}.
\end{aligned}$$

□

Strong log-concavity and log-smoothness

Let $\nu \propto e^{-f}$ be a probability distribution on $\mathbb{R}^n$; and recall that ν is α-**strongly log-concave** if f is α -strongly convex:

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

and that ν is L-**log-smooth** if f is L-smooth:

$$\nabla^2 f(x) \preceq LI$$

Pulling these notions together, if ν is α-SLC and L-log-smooth, then the **condition number** is

$$\kappa = \frac{L}{\alpha}.$$

NB: By construction $\kappa \leq 1$

Bounding acceptance probability of MRW

Let $P =$ Brownian motion with step size η

$$\tilde{P} = \mathrm{MRW} = P + \text{ Metropolist-Hastings for } \nu$$

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

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

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

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

then

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

Proof. Recall

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

where

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

The TV distance is

$$\begin{aligned}
\mathrm{TV}\left(\tilde{P}_x, P_x\right) &= \frac{1}{2}\int_{\mathbb{R}^n} |\tilde{P}_x(y) - P_x(y)| dy \\
&= \frac{1}{2}(A(x) + \underbrace{\int_{\mathbb{R}^n} \left(1 - a_x(y)\right) P_x(y) dy}_{\substack{= 1 - \int_{\mathbb{R}^n} a_x(y) P_x(y) dy \\ = A(x)}}) \\
&= A(x) \\
&= 1 - \mathbb{E}_{Y \sim P_x}\left[a_x(Y)\right] \\
&= 1 - \mathbb{E}_{Y \sim P_x}\left[\min\left\{1, \frac{\nu(Y)}{\nu(x)}\right\}\right] \\
&= 1 - \mathbb{E}_{Y \sim P_x}\left[\min\left\{1, e^{f(x) - f(Y)}\right\}\right]
\end{aligned}$$

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

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

We will prove high-probability bound on $f(x) - f(Y)$ (*NB*: x fixed, $Y \sim P_x$, which means $Y = x + \sqrt{2\eta} Z$).
We have:

$$\begin{aligned}f(x) - f(Y) &\geq \nabla f(Y)^\top (x - Y) \qquad \text{because } f \text{ is convex}\\ &= \nabla f(x)^\top (x - Y) - (\nabla f(x) - \nabla f(Y))^\top (x - Y)\\ (*)\ &\geq \nabla f(x)^\top (x - Y) - L\, \|x - Y\|^2 \qquad \text{since } f \text{ is } L\text{-smooth}\end{aligned}$$

$$\boxed{\begin{aligned}\textit{NB}:\quad f\ L\text{-smooth} &\iff \nabla^2 f(x) \preccurlyeq L \cdot I\\ &\iff \|\nabla f(x) - \nabla f(g)\|_2 \leq L \cdot \|x - y\|_2\\ &\iff (\nabla f(x) - \nabla f(y))^\top (x - y) \leq L \cdot \|x - y\|_2^2\end{aligned}}$$

$Y \sim P_x$ which means $Y = x + \sqrt{2\eta} Z,\ Z \sim \mathcal{N}(0, 1)$

$$((*) \implies)\quad f(x) - f(y) \geq \sqrt{2\eta}\underbrace{\left(-\nabla f(x)^\top Z\right)}_{\mathrm{I}} - L \cdot 2\eta \cdot \underbrace{\|Z\|^2}_{\mathrm{II}}$$

- **for** I: $Z \sim \mathcal{N}(0, 1)$ on $\mathbb{R}^n$

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

since $x \in \mathbb{R}_s = \mathbb{B}\left(x^*, r(s)\sqrt{\frac{n}{2}}\right)$

$$\begin{aligned}\implies\quad \|\nabla f(x)\| &= \|\nabla f(x) - \nabla f(x^*)\| \text{ since } \nabla f(x^* = 0)\\ &\leq L \cdot \|x - x^*\| \text{ by smoothness}\\ &\leq L \cdot r(s) \cdot \sqrt{\frac{n}{2}} \equiv \mathcal{D}_s\end{aligned}$$

by tail bound for 1-dimensional Gaussian,

$$\mathbb{P}\left(-\nabla f(x)^\top Z \geq -2\mathcal{D}_s \sqrt{\log \frac{1}{\varepsilon}}\right) \geq 1 - \varepsilon \quad \forall\, \varepsilon > 0$$

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

$$\|Z\|_2^2 = Z_1^2 + \cdots + Z_n^2 \;\sim\; \mathcal{X}^2 \text{ distribution with } n \text{ degrees of freedom}$$

sub-exponential random variable $\implies$ tail bound

$$\mathbb{E}\left[\|Z\|^2\right] = n$$

$\implies$ concentration around n.

$\square$

Tail bound for $\mathcal{X}^2$ random variable

Lemma. *Let W $(= \|Z\|^2)$ be a $\mathcal{X}^2$-random variable with n degrees of freedom. For all $\varepsilon > 0$:*

$$\mathbb{P}\left(W \leq n\beta_\varepsilon\right) \geq 1 - \varepsilon$$

where $\beta_\varepsilon = 1 + 2\sqrt{\log\left(1/\varepsilon\right) + 2\log\left(1/\varepsilon\right)}$.

- (**highDimStats**) Draft of chapter 2 available here.

Bounding acceptance probability of MRW (continued)

Proof. From above,

$$\begin{aligned}
f(x) - f(Y) &\geq \sqrt{2\eta}\left(-\nabla f(x)^\top Z\right) - 2\eta \cdot L \cdot \|Z\|_2^2 \\
* &\geq -2\sqrt{2\eta} \cdot L \cdot r(s)\sqrt{\frac{n}{\alpha}} \cdot \sqrt{\log\frac{1}{\varepsilon}} \\
&\quad - 2\eta \cdot L \cdot n \cdot \beta_\varepsilon \\
\text{want } &\geq -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} \\
&= -\varepsilon
\end{aligned}$$

Where * happens with probability $\geq 1 - \varepsilon - \varepsilon = 1 - 2\varepsilon$. This happens if

$$\left\{\begin{matrix} \eta & \leq & \frac{\eta^2 \cdot \alpha}{32n \cdot L^2 \cdot r(s)^2 \cdot \log\frac{1}{\varepsilon}} \\ \eta & \leq & \frac{\varepsilon}{4n \cdot L \cdot \beta_\varepsilon} \end{matrix}\right\} *$$

then $f(x) - f(Y) \geq -\varepsilon$ with probability $\geq 1 - 2\varepsilon$, so that with $t = e^{-\varepsilon}$ (*NB:* $\log t = -\varepsilon$)

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

so that

$$\begin{aligned}
\mathrm{TV}\left(\tilde{P}_x, P_x\right) &= 1 - \mathbb{E}\left[\min\left\{1, e^{f(x)-f(Y)}\right\}\right] \\
&\leq 1 - \frac{7}{8} \\
&= \frac{1}{8} \quad \text{which is what we want}
\end{aligned}$$

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

$$\left\{\begin{array}{lllll} \eta & \leq & \frac{1}{1056 n \cdot L} & \leq & \frac{\varepsilon}{4n \cdot L \cdot \beta_\varepsilon} \\ \eta & \leq & \frac{\alpha}{10^5 n \cdot L^2 \cdot r(s)^2} & \leq & \frac{\varepsilon^2 \alpha}{32 L^2 \cdot r(s)^2 \cdot n \cdot \log \frac{1}{\varepsilon}} \end{array}\right\} \text{ which is } *.$$

□

One-step overlap of MRW

Assume $\nu \propto e^{-f}$ on $\mathbb{R}^n$ is α-SLC and L-smooth. Combining with the step above, we get the following:

Lemma. *Let the step size be*

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

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

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

Proof. By the previous two lemmas, we have

$$\begin{aligned} \mathrm{TV}\left(\tilde{P}_x, P_x\right) &\leq \tfrac{1}{8} \\ \mathrm{TV}\left(P_x, P_y\right) &\leq \tfrac{1}{2} \\ \mathrm{TV}\left(\tilde{P}_y, P_y\right) &\leq \tfrac{1}{8} \end{aligned}$$

Then

$$\begin{aligned} \mathrm{TV}\left(\tilde{P}_x, \tilde{P}_y\right) &\leq \mathrm{TV}\left(\tilde{P}_x, P_x\right) + \mathrm{TV}\left(P_x, P_y\right) + \mathrm{TV}\left(P_y, \tilde{P}_y\right) \\ &\leq \frac{1}{8} + \frac{1}{2} + \frac{1}{8} \\ &= \frac{3}{4}. \end{aligned}$$

□

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

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

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

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

Mixing time of MRW

Theorem. *Choose step size*

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

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

$$\begin{aligned} \tau(\varepsilon) &= O\left(\frac{1}{\alpha\cdot\Delta_s^2}\cdot\log\frac{2m}{\varepsilon}\right) \\ &= O\left(\frac{n\cdot L^2\cdot r(s)^2}{\alpha^2}\cdot\log\frac{2m}{\varepsilon}\right) \\ &= O\left(n\cdot k^2\cdot r(s)^2\cdot\log\frac{2m}{\varepsilon}\right) \end{aligned}$$

where $k = \frac{L}{\alpha}$ *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-smooth. Let $\rho_0 = \mathcal{N}\left(x^*, \frac{1}{L} I\right)$ where x^* = mode of ν = minimizer of f.

> **Lemma.**
> $$M_\nu^\infty\left(\rho_0\right) \le \kappa^{n/2}$$
> *NB: Also with approximate mode x^*, see (**JMLR**)[Sec. 3.2.1]*

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

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

Then we can bound the normalizing constant for ν:

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-f(x)} dx &\le e^{-f(x^*)} \int_{\mathbb{R}^n} e^{-\frac{\alpha}{2}\|x-x^*\|^2} dx \\
&= e^{-f(x^*)} \cdot \left(\frac{2\pi}{\alpha}\right)^{n/2}
\end{aligned}$$

Then 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:

$$\begin{aligned}
\frac{\rho(x)}{\nu(x)} &= \frac{e^{-\frac{L}{2}\|x-x^*\|^2}}{\left(\frac{2\pi}{L}\right)^{n/2}} \cdot \frac{1}{\nu(x)} \\
&\le \frac{\cancel{e^{-\frac{L}{2}\|x-x^*\|^2}}}{\left(\frac{2\pi}{L}\right)^{n/2}} \cdot \frac{\left(\frac{2\pi}{\alpha}\right)^{n/2}}{\cancel{e^{-\frac{L}{2}\|x-x^\pi\|^L}}} \\
&= \left(\frac{L}{\alpha}\right)^{n/2} \\
&= k^{n/2}.
\end{aligned}$$

So

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

$\square$

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

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

$$\implies \log\frac{2m}{\varepsilon} = \log\frac{1}{s} \sim \frac{n}{2}\log k + \log\frac{1}{\varepsilon} \sim n\log\left(\frac{k}{\varepsilon^{1/n}}\right)$$

And

$$\begin{aligned} r(s) &= 2 + 2\max\left\{\left(\frac{1}{n}\log\frac{1}{s}\right)^{\frac{1}{4}}, \left(\frac{1}{n}\log\frac{1}{s}\right)^{\frac{1}{2}}\right\} \\ &\sim \left(\frac{1}{n}\log\frac{1}{s}\right)^{\frac{1}{2}} \\ &\sim \sqrt{\log\left(\frac{k}{\varepsilon^{1/n}}\right)} \end{aligned}$$

Then the mixing time bound becomes:

$$\begin{aligned} \tau(\varepsilon) &= O\left(n \cdot k^2 \cdot r(s)^2 \cdot \log\frac{2m}{\varepsilon}\right) \\ &= O\left(n^2 \cdot k^2 \cdot \log^2\left(\frac{k}{\varepsilon^{1/n}}\right)\right). \end{aligned}$$

Recap for MRW

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

MRW algorithm

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

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

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

for some constant c'.