Gradient Flow and Diffusion

Optimization for Machine Learning

Lecture 14

08 April, 2021

Background and Definitions

- Let $\mathcal{P}$ denote the space of probability densities over $\mathbb{R}^d$
- We will briefly use the 2-Wasserstein distance, which is a metric over $\mathcal{P}$. It can be defined as

$$W_2(\mu_1, \mu_2) = \inf_{v \in \Gamma(\mu_1,\mu_2)} \sqrt{\mathbb{E}_{x \sim \mu_1}[\|v(x)\|_2^2]}$$

where $\Gamma(\mu_1, \mu_2)$ is a collection of maps from $\mathbb{R}^d \to \mathbb{R}^d$, defined as follows:

A map $v(x)$ belongs to $\Gamma(\mu_1, \mu_2)$ if and only if **"$\boldsymbol{\mu_2}$ is the push-forward of $\boldsymbol{\mu}$ under $\boldsymbol{v}$"**

In other words, **"if $\boldsymbol{x} \sim \boldsymbol{\mu_1}$, then $\left(\boldsymbol{x} + \boldsymbol{v}(\boldsymbol{x})\right) \sim \boldsymbol{\mu_2}$"**

Given any μ_1, μ_2, let $V^{\mu_1 \to \mu_2} := \operatorname{argmin}_{v \in \Gamma(\mu_1,\mu_2)} \sqrt{\mathbb{E}_{x \sim \mu_1}[\|v(x)\|_2^2]}$ (always attained)

**note: the above definition of W_2 is equivalent to the standard definition only when μ_1, μ_2 are densities (i.e. no atoms)*

Background and Definitions

- Given some target density $\pi \in \mathbb{R}^d$, we define the Relative Entropy functional $\mathcal{H}_\pi : \mathcal{P} \to \mathbb{R}^+$ as

$$\mathcal{H}_\pi(\mu) \coloneqq \int \mu(x) \log\left(\frac{\mu(x)}{\pi(x)}\right) dx$$

fact: $\mathcal{H}_\pi(\mu)$ is minimized at $\mu = \pi$, with $\mathcal{H}_\pi(\pi) = 0$.

Throughout this lecture, $\mathcal{H}_\pi(\cdot)$ plays the role of an "objective which we seek to minimize"

Motivating Problem

We would like to sample from the distribution $\pi^*(x) \propto \exp\big(-U(x)\big)$,

where $U(x)$ is $\boldsymbol{m}$**-strongly convex** and has $\boldsymbol{L}$**-Lipschitz gradients**.

A basic algorithm is the Langevin MCMC algorithm:

$$x_{k+1} = x_k - \eta \nabla U(x_k) + \sqrt{2\eta}\,\xi_k$$

where $\xi_k \sim \mathcal{N}(0, I)$ i.i.d

Let μ_k denote the distribution of x_k. Then

$$\mathcal{H}_{\pi^*}(\mu_{k+1}) \leq (1 - m\eta)\mathcal{H}_{\pi^*}(\mu_k) + \eta^2 d L^2$$

which implies

$$\mathcal{H}_{\pi^*}(\mu_k) \leq e^{-mk\eta} + \frac{\eta d L^2}{m}$$

Taking the continuous-time limit

The Langevin MCMC Algorithm:

$$\tilde{\boldsymbol{x}}_{(k+1)\eta} = \tilde{\boldsymbol{x}}_{k\eta} - \delta \nabla U\left(\tilde{\boldsymbol{x}}_{k\eta}\right) + \sqrt{2\delta}\,\xi_k$$

where $\xi_k \sim \mathcal{N}(0, I)$.

Guarantee:

$$\mathcal{H}_{\pi^*}\left(\tilde{\mu}_{(k+1)\eta}\right) \leq (1 - m\eta)\mathcal{H}_{\pi^*}\left(\tilde{\mu}_{k\eta}\right) + \eta^2 d L^2$$

By taking the limit of $\delta \to 0$, we obtain the Langevin Diffusion:

$$d\boldsymbol{x}_t = -\nabla U(\boldsymbol{x}_t)dt + \sqrt{2}dB_t$$

Guarantee:

$$\frac{d}{dt}\mathcal{H}_{\pi^*}(\mu_t) \leq -m\mathcal{H}_{\pi^*}(\mu_t)$$

A familiar algorithm : Gradient Descent in Euclidean Space

Objective: $\min_{y} f(y)$

If $f(y)$ is m-strongly convex, it satisfies, for all y_1, y_2

$$f(y_2) \geq f(y_1) + \langle \nabla f(y_1), y_2 - y_1 \rangle + \frac{m}{2} \|y_2 - y_1\|_2^2$$

the above implies "gradient domination":

$$f(y) - f(y^*) \leq \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Gradient descent is given by

$$\tilde{y}^{k+1} = \tilde{y}^k - \eta \nabla f\left(\tilde{y}^k\right)$$

it satisfies (assuming L-Lipschitz gradients)

$$f\left(\tilde{y}^{k+1}\right) - f\left(\tilde{y}^k\right)$$
$$\leq -2m\eta \left(f\left(\tilde{y}^k\right) - f(y^*)\right) + \frac{\eta^2 L}{2} \left\|f\left(\tilde{y}^k\right) - f(y^*)\right\|_2^2$$

Continuous-time limit is gradient flow:

$$d\, y_t = -\nabla f(y_t) dt$$

it satisfies

$$\frac{d}{dt} f(y_t) - f(y^*) \leq -2m \left(f(y_t) - f(y^*)\right)$$

Comparison of the two problems

Objective: $\min_{\mu\in\mathcal{P}} \mathcal{H}(\mu) := \min_{\mu\in\mathcal{P}} \int \mu(x)\log\left(\frac{\mu(x)}{\pi^*(x)}\right)dx$

Objective: $\min_x f(x)$

If $U(x) = -\log\pi^*(x)$ is m-strongly convex, then

$\mathcal{H}(\mu)$ is m-geodesically-strongly convex (HWI Inequallity)

$$\mathcal{H}(\mu_1) \geq \mathcal{H}(\mu_2) + \int \left\langle \nabla\log\frac{\mu_1(x)}{\pi^*(x)}, V^{1\to 2}(x)\right\rangle \mu_1(x)dx + \frac{m}{2}\int \|V^{1\to 2}(x)\|_2^2\mu_1(x)dx$$

for all $\mu_1, \mu_2 \in \mathcal{P}$

$f(y)$ is m-strongly convex

$$f(y_2) \geq f(y_1) + \langle \nabla f(y_1), y_2 - y_1\rangle + \frac{m}{2}\|y_2 - y_1\|_2^2$$

for all $y_1, y_2 \in \mathbb{R}^d$

$\mathcal{H}(\mu)$ is m-gradient dominant (Log-Sobolev Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \leq \frac{1}{2m}\int \left\|\nabla\log\frac{\mu(x)}{\pi^*(x)}\right\|_2^2 \mu_1(x)dx$$

$f(y)$ is m-gradient dominant

$$f(y) - f(y^*) \leq \frac{1}{2m}\|\nabla f(y)\|_2^2$$

$\mathcal{H}(\mu)$ is lower bounded by W_2 distance (Talagrand Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \geq \frac{m}{2}W_2^2(\mu_1, \mu_2)$$

$f(y)$ is lower bounded by $\|y - y^*\|_2^2$

$$f(y) - f(y^*) \geq \frac{m}{2}\|y - y^*\|_2^2$$

Langevin Diffusion vs (Euclidean) Gradient Flow

$\mathcal{H}(\mu)$ is m-gradient dominant (Log-Sobolev Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \leq \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$

$f(y)$ is m-gradient dominant

$$f(y) - f(y^*) \leq \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Gradient Flow wrt $\mathcal{H}$

$$\frac{\partial}{\partial t} \mu_t(x) = div\left(\mu_t(x) \nabla \log \frac{\mu_t(x)}{\pi^*(x)}\right)$$

Sample Space Dynamics:

$$d\, x_t = -\nabla U(x_t) dt + \sqrt{2} dB_t$$

Gradient Flow wrt f

$$d\, y_t = -\nabla f(y_t) dt$$

Gradient Flow Convergence

$$\frac{d}{dt}\, \mathcal{H}(\mu_t) - \mathcal{H}(\pi^*) = -\int \left\| \nabla \log \frac{\mu_1(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$
$$\leq -2m\big(\mathcal{H}(\mu_t) - \mathcal{H}(\pi^*)\big)$$

Gradient Flow Convergence

$$\frac{d}{dt}\, f(y_t) - f(y^*) = -\|\nabla f(y_t)\|_2^2$$
$$\leq -2m\, \big(f(y_t) - f(y^*)\big)$$

Part 1: Strong Convexity implies Gradient Dominance

$$\mathcal{H}(\mu_2) \geq \mathcal{H}(\mu_1) + \int \left\langle \nabla \log \frac{\mu_1(x)}{\pi^*(x)}, V^{\mu_1 \to \mu_2}(x) \right\rangle \mu_1(x) dx + \frac{m}{2} \int \|V^{\mu_1 \to \mu_2}(x)\|_2^2 \mu_1(x) dx$$

$$\Rightarrow \qquad \mathcal{H}(\mu) - \mathcal{H}(\pi^*) \leq \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$

$$f(y_2) - f(y_1) \geq \langle \nabla f(y_1), y_2 - y_1 \rangle + \frac{m}{2} \|y_2 - y_1\|_2^2$$

$$\Rightarrow \qquad f(y) - f(y^*) \leq \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Proof: let $\mu_1 = \mu$, $y_2 = \pi^*$, then

$$\begin{aligned}
&\mathcal{H}(\mu) - \mathcal{H}(\mu^*) \\
&\leq -\int \left\langle \nabla \log \frac{\mu(x)}{\pi^*(x)}, V^{\mu \to \pi^*}(x) \right\rangle \mu(x) dx - \frac{m}{2} \int \left\|V^{\mu \to \pi^*}(x)\right\|_2^2 \mu(x) dx \\
&\leq \int \sup_v \left(\left\langle \nabla \log \frac{\mu(x)}{\pi^*(x)}, v \right\rangle - \frac{m}{2} \|v\|_2^2 \right) \mu(x) dx \\
&\leq \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx
\end{aligned}$$

Proof: let $y_1 = y$, $y_2 = y^*$, then

$$\begin{aligned}
&f(y) - f(y^*) \\
&\leq \langle \nabla f(y), y - y^* \rangle - \frac{m}{2} \|y - y^*\|_2^2 \\
&\leq \sup_v \langle \nabla f(y), v \rangle - \frac{m}{2} \|v\|_2^2 \\
&\leq \frac{1}{2m} \|\nabla f(y)\|_2^2
\end{aligned}$$

Langevin Diffusion vs (Euclidean) Gradient Flow

$\mathcal{H}(\mu)$ is m-gradient dominant (Log-Sobolev Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \le \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$

$f(y)$ is m-gradient dominant

$$f(y) - f(y^*) \le \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Gradient Flow wrt $\mathcal{H}$

$$\frac{\partial}{\partial t} \mu_t(x) = div\left(\mu_t(x) \nabla \log \frac{\mu_t(x)}{\pi^*(x)} \right)$$

Sample Space Dynamics:

$$d\, x_t = -\nabla U(x_t) dt + \sqrt{2} dB_t$$

Gradient Flow wrt f

$$d\, y_t = -\nabla U(y_t) dt$$

Gradient Flow Convergence

$$\frac{d}{dt}\, \mathcal{H}(\mu_t) - \mathcal{H}(\pi^*) = -\int \left\| \nabla \log \frac{\mu_1(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$
$$\le -2m\big(\mathcal{H}(\mu_t) - \mathcal{H}(\pi^*)\big)$$

Gradient Flow Convergence

$$\frac{d}{dt}\, f(y_t) - f(y^*) = -\|\nabla f(y_t)\|_2^2$$
$$\le -2m\,\big(f(y_t) - f(y^*)\big)$$

Part 2: The Dynamics in Probability Space

Sample-space Dynamics

$$d\,x_t = -\nabla U(x_t)dt + \sqrt{2}dB_t$$

Gradient Flow wrt f

$$d\,y_t = -\nabla U(y_t)dt$$

Fokker Planck Equation:

$$\begin{aligned}\frac{\partial}{\partial t}\mu_t(x) &= div\big(\mu_t(x)\nabla U(x_t)\big) + Tr\big(\nabla^2 \mu_t(x)\big)\\ &= -div(\mu_t(x)\nabla \log \pi^*(x)) + div(\mu_t(x)\,\nabla\log \mu_t(x))\end{aligned}$$

Probability-space Dynamics

$$\frac{\partial}{\partial t}\mu_t(x) = div\left(\mu_t(x)\nabla \log \frac{\mu_t(x)}{\pi^*(x)}\right)$$

Langevin Diffusion vs (Euclidean) Gradient Flow

$\mathcal{H}(\mu)$ is m-gradient dominant (Log-Sobolev Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \leq \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$

$f(y)$ is m-gradient dominant

$$f(y) - f(y^*) \leq \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Gradient Flow wrt $\mathcal{H}$

$$\frac{\partial}{\partial t} \mu_t(x) = div\left(\mu_t(x) \nabla \log \frac{\mu_t(x)}{\pi^*(x)}\right)$$

Sample Space Dynamics:

$$d\, x_t = -\nabla U(x_t) dt + \sqrt{2} dB_t$$

Gradient Flow wrt f

$$d\, y_t = -\nabla U(y_t) dt$$

Gradient Flow Convergence

$$\frac{d}{dt} \mathcal{H}(\mu_t) - \mathcal{H}(\pi^*) = -\int \left\| \nabla \log \frac{\mu_1(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$
$$\leq -2m\big(\mathcal{H}(\mu_t) - \mathcal{H}(\pi^*)\big)$$

Gradient Flow Convergence

$$\frac{d}{dt} f(y_t) - f(y^*) = -\|\nabla f(y_t)\|_2^2$$
$$\leq -2m\left(f(y_t) - f(y^*)\right)$$

Part 3: Evolution of the Objective

$$
\begin{aligned}
&\frac{\partial}{\partial t}\mathcal{H}(\mu_t) \\
&= \frac{\partial}{\partial t}\int \mu_t(x)\log\frac{\mu_t(x)}{\pi^*(x)}dx \\
&= \int \frac{\partial \mu_t(x)}{\partial t}\log\frac{\mu_t(x)}{\pi^*(x)}dx + \mu_t(x)\frac{\partial}{\partial t}\left(\log\frac{\mu_t(x)}{\pi^*(x)}\right)dx \\
&= \int \frac{\partial \mu_t}{\partial t}(x)\log\frac{\mu_t(x)}{\pi^*(x)}dx + \int \frac{\partial \mu_t}{\partial t}(x)dx \\
&= \int \frac{\partial \mu_t}{\partial t}(x)\log\frac{\mu_t(x)}{\pi^*(x)}dx \\
&= \int \left(div\left(\mu_t(x)\nabla\log\frac{\mu_t(x)}{\pi^*(x)}\right)\right)\cdot\log\frac{\mu_t(x)}{\pi^*(x)}dx \\
&= -\int \left\langle\left(\mu_t(x)\nabla\log\frac{\mu_t(x)}{\pi^*(x)}\right), \nabla\log\frac{\mu_t(x)}{\pi^*(x)}\right\rangle dx \\
&= -\int \left\|\nabla\log\frac{\mu_t(x)}{\pi^*(x)}\right\|_2^2 \mu_t(x)dx \\
&\leq -2m\left(\mathcal{H}(\mu_t) - \mathcal{H}(\pi^*)\right)
\end{aligned}
$$

By Log-Sobolev Inequality

$$
\begin{aligned}
&\frac{d}{dt}f(y_t) - f(y^*) \\
&= \left\langle \nabla f(y_t), \frac{dy_t}{dt}\right\rangle \\
&= \langle \nabla f(y_t), -\nabla f(y_t)\rangle \\
&= -\|\nabla f(y_t)\|_2^2 \\
&\leq -2m\left(f(y_t) - f(y^*)\right)
\end{aligned}
$$

From Part 2:

$$\frac{\partial \mu_t(x)}{\partial t} = div\left(\mu_t(x)\nabla\log\frac{\mu_t(x)}{\pi^*(x)}\right)$$

Integration by parts: for any $g: \mathbb{R}^d \to \mathbb{R}$ and $v: \mathbb{R}^d \to \mathbb{R}^d$ such that $\|v(x)\|_2$ decays sufficiently fast as $\|x\|_2 \to \infty$

$$\int g(x)div\big(v(x)\big)dx = -\int \langle \nabla g(x), v(x)\rangle dx$$

Langevin Diffusion vs (Euclidean) Gradient Flow

$\mathcal{H}(\mu)$ is m-gradient dominant (Log-Sobolev Inequality)

$$\mathcal{H}(\mu) - \mathcal{H}(\pi^*) \leq \frac{1}{2m} \int \left\| \nabla \log \frac{\mu(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$

$f(y)$ is m-gradient dominant

$$f(y) - f(y^*) \leq \frac{1}{2m} \|\nabla f(y)\|_2^2$$

Gradient Flow wrt $\mathcal{H}$

$$\frac{\partial}{\partial t} \mu_t(x) = div\left(\mu_t(x) \nabla \log \frac{\mu_t(x)}{\pi^*(x)}\right)$$

Sample Space Dynamics:

$$d\, x_t = -\nabla U(x_t) dt + \sqrt{2} dB_t$$

Gradient Flow wrt f

$$d\, y_t = -\nabla U(y_t) dt$$

Gradient Flow Convergence

$$\frac{d}{dt} \mathcal{H}(\mu_t) - \mathcal{H}(\pi^*) = -\int \left\| \nabla \log \frac{\mu_1(x)}{\pi^*(x)} \right\|_2^2 \mu(x) dx$$
$$\leq -2m\big(\mathcal{H}(\mu_t) - \mathcal{H}(\pi^*)\big)$$

Gradient Flow Convergence

$$\frac{d}{dt} f(y_t) - f(y^*) = -\|\nabla f(y_t)\|_2^2$$
$$\leq -2m\,\big(f(y_t) - f(y^*)\big)$$

Part 4: Discretization and Iteration Complexity

Langevin MCMC

$$\tilde{x}^{k+1} = \tilde{x}^k - \eta \nabla U(\tilde{x}^k) + \sqrt{2\delta \xi^k}$$

Can be re-written as a continuous-time stochastic process:

$$d\tilde{x}_t = -\nabla U(\tilde{x}_{k\eta})dt + \sqrt{2}dB_t \quad \text{for } t \in [k\eta, (k+1)\eta)$$

Again by Fokker Planck, for $t \in [k\eta, (k+1)\eta)$:

$$\frac{\partial}{\partial t}\tilde{\mu}_t(x) = div\left(\tilde{\mu}_t(x)\left(\nabla \log \tilde{\mu}_t(x) + \nabla U(x_{k\eta})\right)\right)$$
$$= div\left(\tilde{\mu}_t(x)\left(\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right)\right) + div\left(\tilde{\mu}_t(x)\left(\nabla U(\tilde{x}_{k\eta}) - \nabla U(\tilde{x})\right)\right)$$

$$\mathcal{H}\left(\tilde{\mu}_{(k+1)\eta}\right) - \mathcal{H}\left(\tilde{\mu}_{k\eta}\right)$$
$$= \int_{k\eta}^{(k+1)\eta} div\left(\tilde{\mu}_t(x)\left(\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right)\right) \cdot \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\, dt$$
$$+ \int_{k\eta}^{(k+1)\eta} div\left(\tilde{\mu}_t(x)\left(\nabla U(\tilde{x}_{k\eta}) - \nabla U(\tilde{x})\right)\right) \cdot \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\, dt$$
$$\leq \int_{k\eta}^{(k+1)\eta} -\left\|\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right\|_2^2 + L\left\|\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right\|_2 \left\|\tilde{x}_t - \tilde{x}_{k\eta}\right\|_2\, dt$$

Gradient Descent

$$\tilde{y}^{k+1} = \tilde{y}^k - \eta \nabla f(\tilde{y}^k)$$

Can be re-written as a continuous-time process:

$$d\tilde{y}_t = -\nabla f(\tilde{y}_{k\eta})dt \qquad \text{for } t \in [k\eta, (k+1)\eta)$$
$$= -\nabla f(\tilde{y}_t) + \left(\nabla f(\tilde{y}_t) - \nabla f(\tilde{y}_{k\eta})\right)$$

Then

$$f\left(\tilde{y}_{(k+1)\eta}\right) - f\left(\tilde{y}_{k\eta}\right)$$
$$= \int_{k\eta}^{(k+1)\eta} \left\langle \nabla f(\tilde{y}_t), \frac{d}{dt}\tilde{y}_t \right\rangle dt$$
$$= \int_{k\eta}^{(k+1)\eta} \langle \nabla f(\tilde{y}_t), -\nabla f(\tilde{y}_t)\rangle\, dt$$
$$+ \int_{k\eta}^{(k+1)\eta} \left\langle \nabla f(\tilde{y}_t), \nabla f(\tilde{y}_t) - \nabla f(\tilde{y}_{k\eta})\right\rangle dt$$
$$\leq \int_{k\eta}^{(k+1)\eta} -\|\nabla f(\tilde{y}_t)\|_2^2 + L\|\nabla f(\tilde{y}_t)\|_2 \left\|\tilde{y}_t - \tilde{y}_{k\eta}\right\|_2\, dt$$
$$\leq -\eta\|\nabla f(\tilde{y})\|_2^2 + \frac{L\eta^2}{2}\|\nabla f(\tilde{y})\|_2^2$$

Part 4: Discretization and Iteration Complexity

$$\mathcal{H}(\tilde{\mu}_{(k+1)\eta}) - \mathcal{H}(\tilde{\mu}_{k\eta})$$
$$= \int_{k\eta}^{(k+1)\eta} \int div\left(\tilde{\mu}_t(x)\left(\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right)\right) \cdot \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\, dx\ dt$$
$$+ \int_{k\eta}^{(k+1)\eta} \int div\left(\tilde{\mu}_t(x)\left(\nabla U(\tilde{x}_{k\eta}) - \nabla U(\tilde{x})\right)\right) \cdot \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)} dx\, dt$$
$$\leq \int_{k\eta}^{(k+1)\eta} \int -\left\|\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right\|_2^2 \tilde{\mu}_t(x) dx\, dt$$
$$+L \int_{k\eta}^{(k+1)\eta} \int \left\|\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right\|_2 \left\|x - \tilde{x}_{k\eta}\right\|_2\ \tilde{\mu}_t(x) dx\, dt$$
$$\leq -\frac{1}{2}\int_{k\eta}^{(k+1)\eta} \int \left\|\nabla \log \frac{\tilde{\mu}_t(x)}{\pi^*(x)}\right\|_2^2 dx\, dt$$
$$+\frac{L^2}{2}\int_{k\eta}^{(k+1)\eta} \int \left\|x - \tilde{x}_{k\eta}\right\|_2^2\ \tilde{\mu}_t(x) dx\, dt$$

$$\frac{L^2}{2}\int_{k\eta}^{(k+1)\eta} \int \left\|x - \tilde{x}_{k\eta}\right\|_2^2\ \tilde{\mu}_t(x) dx\, dt = \frac{L^2}{2}\int_{k\eta}^{(k+1)\eta} \mathbb{E}\left[\left\|\tilde{x}_t - \tilde{x}_{k\eta}\right\|_2^2\right] dt$$
$$= o(L^2\eta^2 d)$$

Gradient Descent

$$\tilde{y}^{k+1} = \tilde{y}^k - \eta \nabla f(\tilde{y}^k)$$

Can be re-written as a continuous-time process:

$$d\tilde{y}_t = -\nabla f(\tilde{y}_{k\eta}) dt \qquad \text{for } t \in [k\eta, (k+1)\eta)$$
$$= -\nabla f(\tilde{y}_t) + \left(\nabla f(\tilde{y}_t) - \nabla f(\tilde{y}_{k\eta})\right)$$

Then

$$f(\tilde{y}_{(k+1)\eta}) - f(\tilde{y}_{k\eta})$$
$$= \int_{k\eta}^{(k+1)\eta} \left\langle \nabla f(\tilde{y}_t), \frac{d}{dt}\tilde{y}_t \right\rangle dt$$
$$= \int_{k\eta}^{(k+1)\eta} \langle \nabla f(\tilde{y}_t), -\nabla f(\tilde{y}_t) \rangle\, dt$$
$$+ \int_{k\eta}^{(k+1)\eta} \left\langle \nabla f(\tilde{y}_t), \nabla f(\tilde{y}_t) - \nabla f(\tilde{y}_{k\eta}) \right\rangle dt$$
$$\leq \int_{k\eta}^{(k+1)\eta} -\|\nabla f(\tilde{y}_t)\|_2^2 + L\|\nabla f(\tilde{y}_t)\|_2 \left\|\tilde{y}_t - \tilde{y}_{k\eta}\right\|_2\ dt$$
$$\leq -\eta\|\nabla f(\tilde{y})\|_2^2 + \frac{L\eta^2}{2}\|\nabla f(\tilde{y})\|_2^2$$

References

“Rapid convergence of the unadjusted langevin algorithm: Isoperimetry suffices” -SS Vempala, A Wibisono 2019

“Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality” - F Otto, C Villani 2000