WK1 Notes.

① posterior Π_n prior π. Θ : parameter space.

$$\Pi_n(\theta) = \frac{\pi(\theta)\cdot L(\theta|X^n)}{\int_\Theta \pi(\theta)\cdot L(\theta|X^n)\,d\theta}.$$

$$\min_{q\in\Gamma} D_{KL}(q\,\|\,\Pi_n) \qquad \Gamma = \{q_\alpha : \alpha\in\mathbb{R}^D\}.$$

Generative modeling (implicit method).

$$\underbrace{\mu}_{\text{reference.}} = N(0, I_d). \quad X\sim\mu, \quad \underbrace{T(X)}_{T\#\mu\ \leftarrow\ \text{transport map.}} \qquad T \text{ is one-to-one.}$$

$$\min_{T\in\mathcal{T}} D_{KL}(T\#\mu\,\|\,\Pi_n).$$

$\mathcal{T}$ → class of maps from $\mathbb{R}^d\to\mathbb{R}^d$.

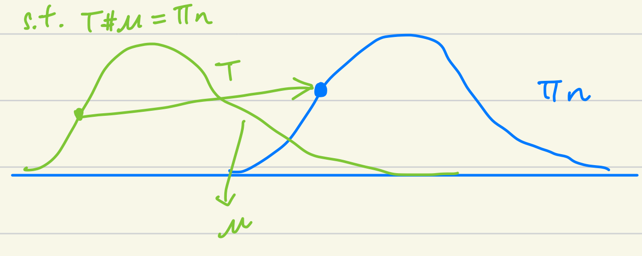
②. Given μ and Π_n. There are infinitely many maps s.t. $T\#\mu = \Pi_n$.

Monge Pbm.

Optimal transport map. (MP) $\displaystyle\min_T \int \underbrace{\|T(X)-X\|^2}_{L^2\ \text{cost}}\,d\mu = \mathbb{E}_\mu[\|\underbrace{T(X)}_{\sim\Pi_n}-X\|^2].$

s.t. $T\#\mu = \Pi_n$

arg min T : optimal transport map.



Goal : find the optimal transport map from μ to Π_n.

③ If μ is absolutely continuous w.r.t. $\mathbb{R}^d$ Lebsgue measure, then T is unique.

$T = \nabla\phi$. ϕ: convex function.

- lemma: If $T \# \mu = \pi_n$ and $T = \nabla\phi$, ϕ convex function, then T is the unique optimal transport map.

(MP). $\min\limits_{T} \int \| T(x) - x \|^2 d\mu(x)$ s.t. $T \# \mu = \pi_n$.

④.

As long as we can find ϕ convex s.t. $(\nabla\phi)\#\mu = \pi_n$, then $T = \nabla\phi$ is the optimal transport map. $\min\limits_{\phi \text{ convex}} D_{KL}((\nabla\phi)\#\mu \,\|\, \pi_n)$.

① model convex function $\mathbb{R}^d$:

Input convex NN (ICNN). $f_\alpha : \mathbb{R}^d \to \mathbb{R}$ convex.

② $\min\limits_{q} D_{KL}(q \,\|\, \pi_n)$.

⑤. Wasserstein space. Dist over $\mathbb{R}^d$ metric; W_2 2-Wassertein space.

- Discretized Wassertein gradient flow.

$$\min_q F(q) = \underline{D_{KL}(q \| \pi_n)}.$$

JKO scheme. step k.

$$q^{k+1} = \arg\min_q \left(F(q) + \frac{1}{2\tau} W_2^2(q, q^k) \right)$$

step size. output from last iterate.

Euclidean case. $\min_{x \in \mathbb{R}^d} f(x)$.

⑥.

$$x^{k+1} = x^k - \tau \nabla f(x^k).$$ explicit Gradient Descent.

(solve for x^{k+1}).

$$x^{k+1} = x^k - \tau \nabla f(x^{k+1}).$$ implicit G.D.

(JKO): $x^{k+1} = \min_x \left\{ f(x) + \frac{1}{2\tau} \underbrace{\|x - x^k\|^2}_{W_2} \right\}$ proximal method.

First order optimality condition. $\nabla f(x^{k+1}) + \frac{1}{\tau}(x^{k+1} - x^k) = 0$

$\Leftrightarrow x^{k+1} = x^k - \tau \nabla f(x^{k+1})$ implicit G.D.