

• eg. $F(p) = \int p \log p \cdot dx$ (negative entropy).

want to $\min F(p) \Leftrightarrow \max$ entropy.

$$\begin{aligned} \frac{d}{d\varepsilon} F(p + \varepsilon X) \Big|_{\varepsilon=0} &= \frac{d}{d\varepsilon} \int (p + \varepsilon X) \log(p + \varepsilon X) dx \Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon} \int p \log(p + \varepsilon X) + \varepsilon X \log(p + \varepsilon X) dx \Big|_{\varepsilon=0} \\ &= \int \frac{p \cdot X}{p + \varepsilon X} + X \log(p + \varepsilon X) + \varepsilon X \cdot \frac{X}{p + \varepsilon X} dx \Big|_{\varepsilon=0} \\ &= \int X + X \log(p) dx \end{aligned}$$

$$\Rightarrow \frac{\delta F}{\delta p} = \log p + 1.$$

$$\nabla \left(\frac{\delta F}{\delta p} \right) = \nabla (\log p + 1) = \frac{1}{p} \cdot \nabla p.$$

\Rightarrow The Gradient flow is $p_t - \nabla \cdot (p \cdot \frac{1}{p} \cdot \nabla p)$

$$= p_t - \nabla \cdot (\nabla p)$$

$$0 = p_t - \Delta p.$$

$$\Rightarrow p_t = \Delta p.$$

$$KL(q \parallel \pi_n) = - \int q(\underline{\theta}) \cdot \log \frac{\pi_n(\underline{\theta})}{q(\underline{\theta})} d\underline{\theta}.$$

$$= - \int q(\underline{\theta}) \log \pi_n(\underline{\theta}) d\underline{\theta} + \int q(\underline{\theta}) \cdot \log q(\underline{\theta}) d\underline{\theta}.$$

$$= \int p(\underline{\theta}) \cdot \log q(\underline{\theta}) d\underline{\theta} + \int V(\underline{\theta}) d q(\underline{\theta}). \quad V(\underline{\theta}) = - \log \pi_n(\underline{\theta}).$$

$$F(p) = \int p \log p dx + \int V \cdot p dx. \quad \text{want to min } F(p).$$

$$\begin{aligned} \left. \frac{d}{d\varepsilon} F(p + \varepsilon X) \right|_{\varepsilon=0} &= \left. \frac{d}{d\varepsilon} \int (p + \varepsilon X) \log (p + \varepsilon X) dx \right|_{\varepsilon=0} + \left. \frac{d}{d\varepsilon} \int V \cdot (p + \varepsilon X) dx \right|_{\varepsilon=0} \\ &= \left. \frac{d}{d\varepsilon} \int p \log (p + \varepsilon X) + \varepsilon X \log (p + \varepsilon X) dx \right|_{\varepsilon=0} + \left. \frac{d}{d\varepsilon} \int V \cdot p + V \varepsilon X dx \right|_{\varepsilon=0}. \end{aligned}$$

$$\begin{aligned} &= \int \frac{pX}{p + \varepsilon X} + X \log (p + \varepsilon X) + \varepsilon X \cdot \frac{X}{p + \varepsilon X} dx \Big|_{\varepsilon=0} \\ &\quad + \int V X dx \Big|_{\varepsilon=0} \end{aligned}$$

$$= \int X + X \log (p) + V X dx.$$

$$\Rightarrow \frac{\delta F}{\delta p} = \log p + 1 + V.$$

$$\nabla \left(\frac{\delta F}{\delta p} \right) = \nabla (\log p + 1 + V) = \frac{1}{p} \cdot \nabla p + \nabla V$$

$$0 = p_t - \nabla \cdot (p \cdot (\frac{1}{p} \nabla p + \nabla V))$$

$$= p_t - \nabla \cdot (\nabla p + p \nabla V)$$

$$= p_t - \nabla \cdot \nabla p + \nabla (p \nabla V)$$

$$= p_t - \Delta p + \nabla (p \nabla V)$$

$$KL(p \parallel \pi_n) = - \int p \log \left[\frac{\pi_n}{p} \right] dx$$

$$= \int p \log p \, dx - \int p \log \pi_n \, dx$$

$$= \mathbb{E}_p [\log p] - \mathbb{E}_p [\log \pi_n]$$

$$= \mathbb{E}_p [\log p] - \mathbb{E}_p \left[\log \left(\frac{Z_n \cdot \pi}{p(x)} \right) \right]$$

$$= \mathbb{E}_p [\log p] - \mathbb{E}_p [\ln + \log \pi] + \mathbb{E}_p [\log p(x)]$$

$$= \underbrace{\mathbb{E}_p [\log p] - \mathbb{E}_p [\ln + \log \pi] + \log p(x)}_{= -\text{ELBO. Evidence Lower Bound.}}$$

$$\text{ELBO}(p) = \mathbb{E}_p [\ln + \log \pi] - \mathbb{E}_p [\log p]$$

$$\text{To max ELBO}(p) \iff \text{min } F(p), \text{ where } F(p) = -\text{ELBO}(p).$$

$$F(p) = \int p \log p \, dx - \int p (\ln + \log \pi) \, dx.$$

$$= \int p \log p \, dx + \int p V \, dx \quad \text{where } V = -(\ln + \log \pi_n).$$

Similar to the KL case, The gradient flow is

$$0 = p_t - \Delta p + \nabla(p \nabla V) \quad (\text{Fokker-Planck equation}).$$