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

want to min $F(\rho)$ $\Longleftrightarrow$ max entropy.

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon X)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} \int (\rho + \varepsilon X) \log(\rho + \varepsilon X)\, dx \Big|_{\varepsilon=0}.$$

$$= \frac{d}{d\varepsilon} \int \rho \log(\rho + \varepsilon X) + \varepsilon X \log(\rho + \varepsilon X)\, dx \Big|_{\varepsilon=0}$$

$$= \int \frac{\rho \cdot X}{\rho + \varepsilon X} + X \log(\rho + \varepsilon X) + \varepsilon X \cdot \frac{X}{\rho + \varepsilon X}\, dx \Big|_{\varepsilon=0}$$

$$= \int X + X \log(\rho)\, dx$$

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

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

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

$$= \rho_t - \nabla \cdot (\nabla \rho)$$

$$0 = \rho_t - \Delta \rho.$$

$$\Rightarrow \rho_t = \Delta \rho.$$

$$KL(q \| \pi_n) = -\int q(\underset{\sim}{\theta}) \cdot \log \frac{\pi_n(\underset{\sim}{\theta})}{q(\underset{\sim}{\theta})} \, d\underset{\sim}{\theta}.$$

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

$$= \int \rho(\underset{\sim}{\theta}) \cdot \log q(\underset{\sim}{\theta}) \, d\underset{\sim}{\theta} + \int V(\underset{\sim}{\theta}) \, dq(\underset{\sim}{\theta}). \quad V(\underset{\sim}{\theta}) = -\log \pi_n(\underset{\sim}{\theta}).$$

$F(\rho) = \int \rho \log \rho \, dx + \int V \cdot \rho \, dx.$ want to min $F(\rho)$.

$$\frac{d}{d\varepsilon} F(\rho + \varepsilon X)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon} \int (\rho + \varepsilon X) \log(\rho + \varepsilon X) \, dx \Big|_{\varepsilon=0} + \frac{d}{d\varepsilon} \int V \cdot (\rho + \varepsilon X) \, dx \Big|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon} \int \rho \log(\rho + \varepsilon X) + \varepsilon X \log(\rho + \varepsilon X) \, dx \Big|_{\varepsilon=0} + \frac{d}{d\varepsilon} \int V \cdot \rho + V \varepsilon X \, dx \Big|_{\varepsilon=0}.$$

$$= \int \frac{\rho X}{\rho + \varepsilon X} + X \log(\rho + \varepsilon X) + \varepsilon X \cdot \frac{X}{\rho + \varepsilon X} \, dx \Big|_{\varepsilon=0}$$

$$+ \int V X \, dx \Big|_{\varepsilon=0}$$

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

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

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

$$0 = \rho_t - \nabla \cdot \left(\rho \cdot \left(\frac{1}{\rho} \nabla \rho + \nabla V \right) \right)$$

$$= \rho_t - \nabla \cdot (\nabla \rho + \rho \nabla V)$$

$$= \rho_t - \nabla \cdot \nabla \rho + \nabla (\rho \nabla V)$$

$$= \rho_t - \Delta \rho + \nabla (\rho \nabla V)$$

$$KL(\rho \,\|\, \pi_n) = -\int \rho \log\left[\frac{\pi_n}{\rho}\right] dx$$

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

$$= \mathbb{E}_\rho [\log \rho] - \mathbb{E}_\rho [\log \pi_n]$$

$$= \mathbb{E}_\rho [\log \rho] - \mathbb{E}_\rho \left[\log\left(\frac{L_n \cdot \pi}{P(x)}\right)\right]$$

$$= \mathbb{E}_\rho [\log \rho] - \mathbb{E}_\rho [l_n + \log \pi] + \mathbb{E}_q [\log p(x)]$$

$$= \underbrace{\mathbb{E}_\rho [\log \rho] - \mathbb{E}_\rho [l_n + \log \pi]}_{= -\text{ELBO}.} + \log p(x)$$

Evidence Lower Bound.

$$\text{ELBO}(\rho) = \mathbb{E}_\rho [l_n + \log \pi] - \mathbb{E}_\rho [\log \rho]$$

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

$$F(\rho) = \int \rho \log \rho \, dx - \int \rho (l_n + \log \pi) \, dx.$$

$$= \int \rho \log \rho \, dx + \int \rho V \, dx \qquad \text{where } V = -(l_n + \log \pi_n).$$

Similar to the KL case, The gradient flow is

$$0 = \rho_t - \Delta \rho + \nabla(\rho \nabla V) \qquad \text{(Fokker-Planck equation).}$$