

Smooth Transform

1° let P, Q be two smooth density on X

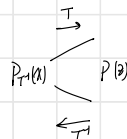
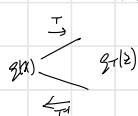
and $T = T(x)$ be an one-to-one transform on X differentiable w.r.t x

q be the density w.r.t $z = T(x)$ n.d.

$$\nabla_{\theta} KL(Q \| P) = -E_{x \sim Q} [\langle DT(x)^T, \nabla_{\theta} T(x) \rangle + \langle \nabla_{\theta} \log p(x) |_{z=T(x)}, \nabla_{\theta} T(x) \rangle]$$

$$\begin{aligned} KL(Q \| P) &= \int_{\mathcal{Z}} q(z) \log \frac{q(z)}{p(z)} dz \\ &= \int_{\mathcal{Z}} q(T(x)) \log \frac{q(T(x))}{p(T(x))} dx \\ &= \int_{\mathcal{Z}} \frac{q(x)}{|DT(x)|} \log \frac{q(x)}{p(x)} |DT(x)| dx \\ &= \int_{\mathcal{Z}} \frac{q(x)}{|DT(x)|} \log \frac{q(x)}{p(x)} |DT(x)| dx \\ &= KL(Q \| P_T) \end{aligned}$$

$$\text{let } z = T(x) \quad x = T^{-1}(z)$$



$$\begin{aligned} \nabla_{\theta} KL(Q \| P) &= \nabla_{\theta} \int q(x) \log \frac{q(x)}{p(x)} dx \\ &= -E_{x \sim Q} [\nabla_{\theta} \log p_T(x)] \\ &= -E_{x \sim Q} [\nabla_{\theta} \log (p(T(x)) \cdot |DT(x)|)] \\ &= -E_{x \sim Q} [\nabla_{\theta} \log p(T(x)) + \nabla_{\theta} \log |DT(x)|] \\ &= -E_{x \sim Q} [\langle DT(x)^T, \nabla_{\theta} DT(x) \rangle + \langle \nabla_{\theta} \log p(z) |_{z=T(x)}, \nabla_{\theta} T(x) \rangle] \end{aligned}$$

2° Consider the mapping $T(x) = x + \phi(x)$ where ϕ is a smooth function

① When $|\phi|$ is sufficiently small, T is a one-to-one mapping

$$\nabla_{\theta} KL(Q \| P)|_{\phi=0} = -E_{x \sim Q} [\underbrace{\text{trace}(\nabla_{\theta} \log p(x) \phi(x))}_{\text{Score Operator } (\phi^T \nabla_{\theta} \phi)(x)} + D\phi(x)^T]$$

① When $|\phi|$ is sufficiently small

$DT(x) = I + D\phi(x)$ is full rank

By inverse function theorem, a map is locally invertible if its linearization is invertible

$$D_x T(x)|_{\phi=0} = I + D\phi(x)|_{\phi=0} = I$$

$$\nabla_{\theta} DT(x)|_{\phi=0} = D\phi(x)$$

$$\nabla_{\theta} T(x)|_{\phi=0} = \phi(x)$$

$$\begin{aligned} \nabla_{\theta} KL(Q \| P) &= -E_{x \sim Q} [\langle DT(x)^T, \nabla_{\theta} DT(x) \rangle + \langle \nabla_{\theta} \log p(z) |_{z=T(x)}, \nabla_{\theta} T(x) \rangle] \\ &= -E_{x \sim Q} [\text{trace}(D\phi(x)) + \langle \nabla_{\theta} \log p(z) |_{z=T(x)}, \phi(x) \rangle] \end{aligned}$$

$$\text{let } p(x) = \mathbb{E}_{x \sim q}[(\nabla \phi(x))x],$$

$$\text{then } \|p\|_{q,p} = \sqrt{S_K(q,p)},$$

$$\langle \phi, p \rangle_{q,p} = \mathbb{E}_{x \sim q}[\text{trace}(\phi \nabla \phi(x))] = - \nabla_{\phi} \text{KL}(q \| p) \Big|_{\phi=0}$$

consider all perturbations $\| \phi \|_{q,p} \leq \sqrt{S_K(q,p)} = \|p\|_{q,p}$

to make a "Steepest descent".

$$\max_{\phi} \langle \phi, p \rangle_{q,p}$$

$$\text{S.t. } \| \phi \|_{q,p} \leq \|p\|_{q,p}$$

$$\text{then } \phi^* = p, \nabla_{\phi} \text{KL}(q \| p) \Big|_{\phi=0} = - \langle \phi^*, p \rangle_{q,p} = - \|p\|_{q,p}^2 = - S_K(q,p)$$

↓

suggests an iterative update

$$x := x + \epsilon \cdot \phi^*(x)$$

$$= x + \epsilon p(x)$$

$$= x + \epsilon \cdot \mathbb{E}_{x \sim q}[(\nabla \phi(x))x]$$

$$= x + \epsilon \mathbb{E}_{x \sim q}[\nabla \log p(x) \text{Ker}(x) + \nabla_{\phi} \text{Ker}(x)]$$

next section shows that

this is also a functional gradient

Stein Variational Gradient Descent

consider the transformation $T(x) = x + f(x)$

at each iteration, apply $x := x + f(x)$ to minimize $\text{KL}(q \| p)$

1' let $T(x) = x + f(x)$, where $f \in \mathbb{R}^d$ and q be the density of $z = T(x)$ now

$$\text{then } \text{DF} \text{KL}(q \| p) \Big|_{f=0} = - \mathbb{E}_{x \sim q}[\nabla_{\phi} \log p(x) \text{Ker}(x) + \nabla_{\phi} \text{Ker}(x)]$$

$$\text{let } T(x) = x + f(x), \tilde{T}(x) = x + f(x) + \epsilon g(x)$$

$$\text{let } F(f) = \text{KL}(q \| p) = \text{KL}(q \| p_{f,1}) \quad F(f+g) = F(f) + \epsilon \langle \text{DF}(f), g \rangle_{\mathbb{R}^d} + o(\epsilon^2)$$

$$F(f+g) = \text{KL}(q \| p_{f+g,1})$$

$$= \int x q(x) \log \frac{q(x)}{p_{f+g,1}(x)} dx$$

$$= \mathbb{E}_{x \sim q}[\log q(x) - \log p(\tilde{T}(x)) - \log \det(D\tilde{T}(x))]$$

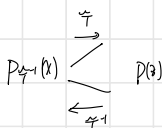
$$= \mathbb{E}_{x \sim q}[\log q(x) - \log p(T(x)) - \log \det(DT(x))$$

$$+ \log p(\tilde{T}(x)) + \log \det(D\tilde{T}(x))$$

$$- \log p(\tilde{T}(x)) - \log \det(D\tilde{T}(x))]$$

$$= \text{KL}(q \| p_{f,1}) + \mathbb{E}_{x \sim q}[\log p(T(x)) - \log p(\tilde{T}(x))]$$

$$+ \mathbb{E}_{x \sim q}[\log \det(DT(x)) - \log \det(D\tilde{T}(x))]$$



$$\begin{aligned}
 0 &= E_{x,y} [\log p(x+y) - \log p(x+y+eDg)] \\
 &= -\epsilon E_{x,y} [\nabla_x \log p(x)|_{x=x+y}^T g] + o(\epsilon^2) \\
 &= -\epsilon E_{x,y} [\langle \nabla_x \log p(x)|_{x=x+y}, [\langle g_1, k(x_1) \rangle, \dots, \langle g_n, k(x_n) \rangle]^T \rangle] + o(\epsilon^2) \\
 &= -\epsilon \cdot \langle E_{x,y} [\nabla_x \log p(x)|_{x=x+y}, k(x)], g \rangle_{\mathcal{H}} + o(\epsilon^2)
 \end{aligned}$$

$$\begin{aligned}
 ② &= E_{x,y} [\log \det(I + Df(x)) - \log \det(I + Df(x) + \epsilon Dg)] + o(\epsilon^4) \\
 &= -\epsilon E_{x,y} [\langle (I + Df(x))^T, Dg(x) \rangle] + o(\epsilon^2) \\
 &= -\epsilon E_{x,y} [\sum_{i,j} (I + Df(x))_{i,j}^T \frac{\partial}{\partial g_j} g(x)] + o(\epsilon^2) \\
 &= -\epsilon E_{x,y} [\sum_{i,j} (I + Df(x))_{i,j}^T \langle \frac{\partial}{\partial g_j} k(x_i), g \rangle_{\mathcal{H}}] + o(\epsilon^2) \\
 &= -\epsilon E_{x,y} [\sum_i \langle \sum_j (I + Df(x))_{i,j}^T \frac{\partial}{\partial g_j} k(x_i), g \rangle_{\mathcal{H}}] + o(\epsilon^2) \\
 &= -\epsilon E_{x,y} [\sum_i \langle (I + Df(x))^T_{i,:} \nabla_x k(x_i), g \rangle_{\mathcal{H}}] + o(\epsilon^2) \\
 &= -\epsilon \langle E_{x,y} [(I + Df(x))^T \nabla_x k(x)], g \rangle_{\mathcal{H}} + o(\epsilon^2)
 \end{aligned}$$

$$\begin{aligned}
 F(f+eg) &= F(f) - \epsilon \langle E_{x,y} [\nabla_x \log p(x)|_{x=x+y}, k(x)], g \rangle \\
 &\quad - \epsilon \langle E_{x,y} [(I + Df(x))^T \nabla_x k(x)], g \rangle + o(\epsilon^2)
 \end{aligned}$$

$$DF(f) = -E_{x,y} [\nabla_x \log p(x)|_{x=x+y}, k(x)] + (I + Df(x))^T \nabla_x k(x)]$$

$$\text{at } f=0, \quad DF(f)|_{f=0} = -E_{x,y} [\nabla_x \log p(x), k(x)] + \nabla_x k(x)]$$

Algorithm:

$$D_f KL(\mathcal{Z} \| P) \big|_{f=0} = -E_{x,y} [\nabla_x \log p(x), k(x)] + \nabla_x k(x)]$$

$$x^{(t)} := x^{(t-1)} - \epsilon \cdot D_f KL(\mathcal{Z} \| P) \big|_{f=0}$$

$$= x^{(t-1)} + \epsilon E_{x,y} [\nabla_x \log p(x), k(x, x^{(t-1)}) + \nabla_x k(x, x^{(t-1)})]$$

$$\approx x^{(t-1)} + \epsilon \frac{1}{m} \sum_{j=1}^m [\nabla_x \log p(x^{(j)}), k(x^{(j)}, x^{(t-1)}) + \nabla_x k(x^{(j)}, x^{(t-1)})]$$

$$\text{Eg. } \mathcal{N}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

$$\nabla_x \mathcal{N}(x; \mu, \Sigma) = -\frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \Sigma^{-1}(x-\mu)$$

$$k(x, x_0) = \exp(-b \|x - x_0\|_2^2)$$

$$\nabla_x k(x, x_0) = \exp(-b \|x - x_0\|_2^2) \cdot -2b(x - x_0)$$