· Smooth Transform

1° let P q be two smooth density on $\mathcal{X}$

and $T = T_\epsilon(\cdot)$ be an one-to-one transform on $\mathcal{X}$ differentiable wrt x, ϵ

q_T be the density of $z = T_\epsilon(x)$ $x \sim q$

$$\nabla_\epsilon KL(q_T \| p) = -E_{x\sim q}\left[\langle DT(x)^{-T}, \nabla_\epsilon \nabla_x T(x) \rangle + \langle \nabla_z \log p(z)|_{z=T(x)}, \nabla_\epsilon T(x) \rangle \right]$$

$$KL(q_T \| p) = \int_z q_T(z) \log \frac{q_T(z)}{p(z)} dz$$

$$= \int_z q_T(T(x)) \log \frac{q_T(T(x))}{p(T(x))} dz$$

$$= \int_z \frac{q(x)}{|DT(x)|} \log \frac{q(x)}{p_{T^{-1}}(x)} d\,T(x)$$

$$= \int_z \frac{q(x)}{|DT(x)|} \log \frac{q(x)}{p_{T^{-1}}(x)} |DT(x)|\, dx$$

$$= KL(q \| p_{T^{-1}})$$

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

$$q(x) \underset{T^{-1}}{\overset{T}{\rightleftarrows}} q_T(z) \qquad p_{T^{-1}}(x) \underset{T^{-1}}{\overset{T}{\rightleftarrows}} p(z)$$

$$\nabla_\epsilon KL(q \| p) = \nabla_\epsilon \int q(x) \log \frac{q(x)}{p_{T^{-1}}(x)} dx$$

$$= -E_{x\sim q}\left[\nabla_\epsilon \log p_{T^{-1}}(x) \right]$$

$$= -E_{x\sim q}\left[\nabla_\epsilon \log \left(p(T(x)) \cdot |DT(x)| \right) \right]$$

$$= -E_{x\sim q}\left[\nabla_\epsilon \log \det(DT(x)) + \nabla_\epsilon \log p(T(x)) \right]$$

$$= -E_{x\sim q}\left[\langle DT(x)^{-T}, \nabla_\epsilon DT(x) \rangle + \langle \nabla_z \log p(z)|_{z=T(x)}, \nabla_\epsilon T(x) \rangle \right]$$

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

① when $|\epsilon|$ is sufficiently small, T is a one-to-one mapping

② $\nabla_\epsilon KL(q_T \| p)|_{\epsilon=0} = -E_{x\sim q}\left[\underbrace{\operatorname{trace}(\nabla_x \log p(x) \phi(x) + D\phi(x)^T)}_{\text{Stein Operator } (\mathcal{A}_p \phi)(x)} \right]$

① when $|\epsilon|$ is sufficiently small

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

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

② $D_x T(x)|_{\epsilon=0} = I + \epsilon D\phi(x)|_{\epsilon=0} = I$

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

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

$$\nabla_x KL(q_T \| p) = -E_{x\sim q}\left[\langle DT(x)^{-T}, \nabla_\epsilon DT(x) \rangle + \langle \nabla_z \log p(z)|_{z=T(x)}, \nabla_\epsilon T(x) \rangle \right]$$

$$= -E_{x\sim q}\left[\operatorname{trace}(D\phi(x)) + \langle \nabla_z \log p(z)|_{z=T(x)}, \phi(x) \rangle \right]$$

let $\beta(x') = E_{x\sim q}[(\mathcal{A}_p k_{x'})(x)]$,

then $\|\beta\|_{\mathcal{H}^d} = \sqrt{S_k(q,p)}$,

$\langle \phi, \beta \rangle_{\mathcal{H}^d} = E_{x\sim q}[\mathrm{trace}((\mathcal{A}_p \phi)(x))] = -\nabla_\epsilon KL(q_T \| p)|_{\epsilon=0}$

consider all perturbations $\|\phi\|_{\mathcal{H}^d} \le \sqrt{S_k(q,p)} = \|\beta\|_{\mathcal{H}^d}$

to make a "Steepest descent",

$$\max_{\phi} \ \langle \phi, \beta \rangle_{\mathcal{H}^d}$$

$$\text{s.t.} \ \ \|\phi\|_{\mathcal{H}^d} \le \|\beta\|_{\mathcal{H}^d}$$

then $\phi^* = \beta$. $\nabla_\epsilon KL(q_T \| p)|_{\epsilon=0} = -\langle \phi^*, \beta \rangle_{\mathcal{H}^d} = -\|\beta\|^2_{\mathcal{H}^d} = -S_k(q,p)$

$\Downarrow$

Suggests an iterative update

$$\left.\begin{aligned} x &:= x + \epsilon \cdot \phi^*(x) \\ &= x + \epsilon \beta(x) \\ &= x + \epsilon \cdot E_{\tilde{x}\sim q}[(\mathcal{A}_p k_x)(\tilde{x})] \\ &= x + \epsilon E_{\tilde{x}\sim q}[\nabla \log p(\tilde{x}) k(x,\tilde{x}) + \nabla_{\tilde{x}} k(\tilde{x}, x)] \end{aligned}\right\} \text{next section shows that this is also a functional gradient}$$

- Stein Variational Gradient Descent

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

at each iteration, apply $x := x + f(x)$ to minimize $KL(q_T \| p)$

1° let $T(x) = x + f(x)$, where $f \in \mathcal{H}^d$ and q_T be the density of $z = T(x)$ $x \sim q$

then $D_f KL(q_T \| p)|_{f=0} = -E_{x\sim q}[\nabla_x \log p(x) k(x,\cdot) + \nabla_x k(x,\cdot)]$

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

let $F(f) = KL(q_T \| p) = KL(q \| p_{T^{-1}})$ $\qquad F(f+\epsilon g) = F(f) + \epsilon \langle DF(f), g \rangle_{\mathcal{H}^d} + O(\epsilon^2)$

$$F(f+\epsilon g) = KL(q \| p_{\tilde{T}^{-1}})$$

$$p_{\tilde{T}^{-1}}(x) \underset{\tilde{T}^{-1}}{\overset{\tilde{T}}{\rightleftarrows}} p(z)$$

$$\begin{aligned} &= \int_x q(x) \log \frac{q(x)}{p_{\tilde{T}^{-1}}(x)} dx \\ &= E_{x\sim q}[\log q(x) - \log p(\tilde{T}(x)) - \log\det(D\tilde{T}(x))] \\ &= E_{x\sim q}[\log q(x) - \log p(T(x)) - \log\det(DT(x)) \\ &\qquad + \log p(T(x)) + \log\det(DT(x)) \\ &\qquad - \log p(\tilde{T}(x)) - \log\det(D\tilde{T}(x))] \\ &= KL(q \| p_{T^{-1}}) + \underbrace{E_{x\sim q}[\log p(T(x)) - \log p(\tilde{T}(x))]}_{①} \\ &\qquad + \underbrace{E_{x\sim q}[\log\det(DT(x)) - \log\det(D\tilde{T}(x))]}_{②} \end{aligned}$$

$$① = E_{x\sim q}\left[\log p(x+f(x)) - \log p(x+f(x)+\epsilon g(x))\right]$$

$$= -\epsilon\, E_{x\sim q}\left[\nabla_z \log p(z)\big|_{z=x+f(x)}^{T}\, g(x)\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\left\langle \nabla_z \log p(z)\big|_{z=x+f(x)},\ \begin{bmatrix}\langle g_1, k(x,\cdot)\rangle \\ \vdots\end{bmatrix}\right\rangle\right] + o(\epsilon^2)$$

$$= -\epsilon \cdot \left\langle E_{x\sim q}\left[\nabla_z \log p(z)\big|_{z=x+f(x)}\, k(x,\cdot)\right], g\right\rangle_{\mathcal{H}^d} + o(\epsilon^2)$$

$$② = E_{x\sim q}\left[\log\det(I+Df(x)) - \log\det(I+Df(x)+\epsilon Dg(x))\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\left\langle (I+Df(x))^{-T}, Dg(x)\right\rangle\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\sum_{i,j} (I+Df(x))^{-T}_{ij}\ \frac{d}{dx_j} g_i(x)\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\sum_{i,j} (I+Df(x))^{-T}_{ij}\ \left\langle \frac{d}{dx_j} k(x,\cdot), g_i\right\rangle_{\mathcal{H}}\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\sum_i \left\langle \sum_j (I+Df(x))^{-T}_{ij}\ \frac{d}{dx_j} k(x,\cdot),\ g_i\right\rangle_{\mathcal{H}}\right] + o(\epsilon^2)$$

$$= -\epsilon\, E_{x\sim q}\left[\sum_i \left\langle (I+Df(x))^{-T}_{i,:}\ \nabla_x k(x,\cdot),\ g_i\right\rangle_{\mathcal{H}}\right] + o(\epsilon^2)$$

$$= -\epsilon \left\langle E_{x\sim q}\left[(I+Df(x))^{-T}\nabla_x k(x,\cdot)\right], g\right\rangle_{\mathcal{H}^d} + o(\epsilon^2)$$

$$F(f+\epsilon g) = F(f) - \epsilon\left\langle E_{x\sim q}\left[\nabla_z \log p(z)\big|_{z=x+f(x)}\, k(x,\cdot)\right], g\right\rangle$$
$$- \epsilon\left\langle E_{x\sim q}\left[(I+Df(x))^{-T}\nabla_x k(x,\cdot)\right], g\right\rangle + o(\epsilon^2)$$

$$DF(f) = -E_{x\sim q}\left[\nabla_z \log p(z)\big|_{z=x+f(x)}\, k(x,\cdot) + (I+Df(x))^{-T}\nabla_x k(x,\cdot)\right]$$

at $f=0$, $$DF(f)\big|_{f=0} = -E_{x\sim q}\left[\nabla_x \log p(x)\, k(x,\cdot) + \nabla_x k(x,\cdot)\right]$$

· Algorithm :

$$D_f KL(q_f\|p)\big|_{f=0} = -E_{x\sim q}\left[\nabla_x \log p(x)\, k(x,\cdot) + \nabla_x k(x,\cdot)\right]$$

$$x^{(i)} := x^{(i)} - t\cdot D_f KL(q_f\|p)\big|_{f=0}$$

$$= x^{(i)} + t\, E_{x\sim q}\left[\nabla \log p(x)\, k(x, x^{(i)}) + \nabla_x k(x, x^{(i)})\right]$$

$$\approx x^{(i)} + t\ \frac{1}{m}\sum_{j=1}^{m}\left[\nabla \log p(x^{(j)})\, k(x^{(j)}, x^{(i)}) + \nabla_{x^{(j)}} k(x^{(j)}, x^{(i)})\right]$$

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

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

$$k(x_1,x_2) = \exp(-b\,\|x_1-x_2\|_2^2)$$

$$\nabla_{x_1} k(x_1,x_2) = \exp(-b\,\|x_1-x_2\|_2^2)\cdot -2b(x_1-x_2)$$