Section 4.1: Particle-Based Update Rule

I. The Optimization Objective

$$\mathcal{D}(q,p) = \max_{\substack{\phi \in \mathcal{H}^d \\ \|\phi\|_{\mathcal{H}^d} \leq 1}} \mathbb{E}_{x\sim q}\left[\operatorname{trace}\left(A_p\phi(x)\right)\right]$$

We have to find the optimal direction ϕ in RKHS that gives maximum descent of KL divergence.

II. Use the Stein Operator

$$A_p\phi(x) = \nabla_x \log p(x) \cdot \phi(x)^\top + \nabla_x \phi(x)$$

We have to find:

$$\mathbb{E}_{x\sim q}\left[\operatorname{trace}\left(A_p\phi(x)\right)\right]$$

We can rewrite this as:

$$\mathbb{E}_{x\sim q}\left[\operatorname{trace}\left(A_p\phi(x)\right)\right] = \mathbb{E}_{x\sim q}\left[\operatorname{trace}\left(\nabla_x \log p(x) \cdot \phi(x)^\top + \nabla_x \phi(x)\right)\right]$$

Using linearity property of trace:

$$= \mathbb{E}_{x\sim q}\left[\nabla_x \log p(x)^\top \phi(x) + \operatorname{trace}\left(\nabla_x \phi(x)\right)\right]$$

III. RKHS Ball Optimization

Maximize the inner product in RKHS:

$$\langle \phi, \phi^\star_{q,p} \rangle_{\mathcal{H}^d}$$

$$\phi^\star_{q,p} = \arg \max_{\|\phi\|_{\mathcal{H}^d} \leq 1} \mathbb{E}_{x\sim q}\left[\operatorname{trace}\left(A_p\phi(x)\right)\right]$$

$$= \nabla_f \mathrm{KL}[q[f]\|p]\Big|_{f=0}$$

This means that optimal direction is basically the gradient of KL divergence in RKHS. We can also show it using the kernel trick.

IV. RKHS Form

In RKHS, represent any function $\phi(x) \in \mathcal{H}^d$ as:

$$\phi(x) = \sum_{i=1}^{n} a_i \, k(x_i, x)$$

Rewriting functional gradient of KL as:

$$\phi^{\star}_{q,p}(x) = \mathbb{E}_{x' \sim q} \left[A_p \, k(x', x) \right]$$

Here, we apply the Stein operator to the kernel function $k(x', x)$.

V. Apply Stein Operator to Kernel

$$A_p \, k(x', x) = \nabla_{x'} \log p(x') \cdot k(x', x) + \nabla_{x'} k(x', x)$$

Take expectation over $x' \sim q$:

$$\boxed{\phi^{\star}_{q,p}(x) = \mathbb{E}_{x' \sim q} \left[\nabla_{x'} \log p(x') \cdot k(x', x) + \nabla_{x'} k(x', x) \right]}$$

$\Rightarrow$ This is Equation 6. It helps to update balances attraction to the posterior and diversity across particles.