Reference measure $\mu_x \in \mathbb{P}(\mathbb{R}^d)$, target measure $\mu_y \in \mathbb{P}(\mathcal{M}_y)$. For densities/measures(abuse of notation) $\mu, \alpha \in \mathbb{P}(\mathcal{X})$ we define $\alpha * \mu$ to be the measure induced by re-sampling random variable $x \sim \mu$ according to α. The density of $\alpha * \mu$ is just given by product of densities, ie $(\alpha * \mu)(x) = \alpha(x)\mu(x)$. Given RKHS $\mathcal{H}_\Lambda$ associated to PDS kernel $\Lambda : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we define $\mathbb{P}(\mathcal{H}_\Lambda) = \{\alpha(\tilde{x}) := e^{\beta_\alpha(\tilde{x})} : \beta_\alpha \in \mathcal{H}_\Lambda \text{ s.t } \int_{\mathcal{X}} e^{\beta_\alpha(x)}\, dx = 1\}$. Then for $\alpha \in \mathbb{P}(\mathcal{H}_\Lambda)$ we define the enduced 'norm' of α as $\|\alpha\|_{\mathbb{P}(\mathcal{H}_\Lambda)} := \|\beta_\alpha\|_{\mathcal{H}_\Lambda}$.
Let $T \in \mathcal{H}_{\Lambda^D}$, $\alpha_y, \alpha_x \in \mathbb{P}(\mathcal{H}_\Lambda)$. We want to solve for the 'optimal' latent re-sampling of μ_y. By optimal we mean
I. There exists a good(in MMD sense) and regular(in $\|\cdot\|_{\mathcal{H}_{\Lambda^D}}$ sense) transport map T from μ_x to $\alpha_y * \mu_y$.
II. There exists a good(in MMD sense) and regular(in $\|\cdot\|_{\mathcal{H}_\Lambda}$ sense) resampling map α_x from $\alpha_y * \mu_y$ to μ_y.
We obtain this map as the solution to the following optimization problem.

$$\alpha_y^* = \underset{\alpha_y, \alpha_x, T}{\operatorname{argmin}}\ \mathrm{MMD}(\alpha_x * T^{\#}\mu_x, \mu_y) + \mathrm{MMD}(T^{\#}\mu_x, \alpha_y * \mu_y) + \lambda_1 \|T\|_{H_T} + \lambda_2 \|\alpha_y\|_{\mathbb{P}(\mathcal{H}_\Lambda)}$$

This optimization problem admits representor form in terms of $Z \in \mathbb{R}^{n \times D}$, $\beta_Y, \beta_X \in \{\beta \in \mathbb{R}^n : \|\exp(\beta)\|_1 = 1\}$ and $\alpha_Y, \alpha_X := \exp(\beta_Y), \exp(\beta_X)$.
$\alpha_Y^* = \underset{\alpha_Y, \alpha_X, Z}{\operatorname{argmin}}\ \hat{\mathrm{MMD}}(\alpha_X, Z) + \hat{\mathrm{MMD}}(\alpha_Y, Z) + \|\hat{Z}\|_{\mathcal{H}_\Lambda} + \|\hat{\alpha_Y}\|_{\mathbb{P}(\mathcal{H}_\Lambda)}$ for
I. $\hat{\mathrm{MMD}}(\alpha_X, Z) = \alpha_X^T K(Z, Z) \alpha_X^T - \frac{2}{n} \alpha_X^T K(Z, Y) \mathbb{1}$
II. $\hat{\mathrm{MMD}}(\alpha_Y, Z) = \alpha_Y^T K(Y, Y) \alpha_Y - \frac{2}{n} \mathbb{1}^T K(Z, Y) \alpha_Y$
III. $\|\hat{Z}\|_{H_\Lambda} = \text{ trace}(Z^T \Lambda(X, X)^{-1} Z)$, IV. $\|\hat{\alpha_Y}\|_{\mathbb{P}(\mathcal{H}_\Lambda)} = \beta_Y^T \Lambda(Y, Y)^{-1} \beta_Y$
In practice can use correction factors and add penalty losses to ensure $\|\exp(\beta_X)\|_1, \|\exp(\beta_Y)\|_1 \approx 1$. I've found emperically this works fairly well. The idea here is that this optimization yields a $T^{\#}\mu_x$ which is a good a proxy for some latent resampling of μ_y, ie $T^{\#}\mu_x \sim \alpha_y * \mu_y$ and $\alpha_x = \alpha_y^{-1}$. Further by including RKHS norms we ensure $T^{\#}, \alpha_y$ are regular.
We cannot regularize α_x directly as it's domain is the support of $\alpha_y * \mu_y$ and not or μ_x or μ_y. However, note $\alpha_x = \alpha_y^{-1}$, and resampling maps have the special property that $\beta_{\alpha_y^{-1}} = -\beta_{\alpha_y}$. I have strong intuition that in this case regularizing α_y will be equivalent to regularzing $\alpha_y^{-1} = \alpha_x$. The numerical experiments I've so far agree with this, though a proof is beyond me so far

Regularity of α_Y is $\mathop{\mathbb{E}}_{\mu_y^N}[\|\hat{\alpha_Y}\|_{\mathbb{P}(\mathcal{H}_\Lambda)}] = \mathop{\mathbb{E}}_{Y \sim \mu_y^N}[\beta_{\alpha_y}(Y)^T \Lambda(Y,Y)^{-1} \beta_{\alpha_y}(Y)]$

Regularity of α_Y^{-1} is $\mathop{\mathbb{E}}_{\alpha_Y * \mu_y}[\|\hat{\alpha_Y^{-1}}\|_{\mathbb{P}(\mathcal{H}_\Lambda)}] = \mathop{\mathbb{E}}_{\tilde{Y} \sim (\alpha_Y * \mu_y)^N}[\beta_{\alpha_y^{-1}}(Y)^T \Lambda(\tilde{Y},\tilde{Y})^{-1} \beta_{\alpha_y^{-1}}(Y)]$

Some numerical tests I did to this effect: