ICNN - JKO.

- outer loop learning rate η_1; inner loop learning rate η_2; JKO discretization steps T_2. $\varphi_\theta \leftarrow$ basic ICNN model with pretrained parameters.; h is the JKO step size.

- for $t_1 = 1, 2, \dots, T_1$

  set $\varphi'_\theta \leftarrow \varphi_\theta$ (previous ICNN result) (with $\theta + \delta$) ↗ small difference.

  for $t_2 = 1, 2, \dots, T_2$

    Sample $Z \sim \mu$ (μ is the reference Dist) $\bar{\mu}(\cdot)$ is the p.d.f of ref Dist.

    W_2^2 or L_2^2

    $W_2^2 \leftarrow W_2^2(\nabla\varphi'_\theta(z), \nabla\varphi_\theta(z))$.

    $\pi_n(\cdot)$ p.d.f of posterior.

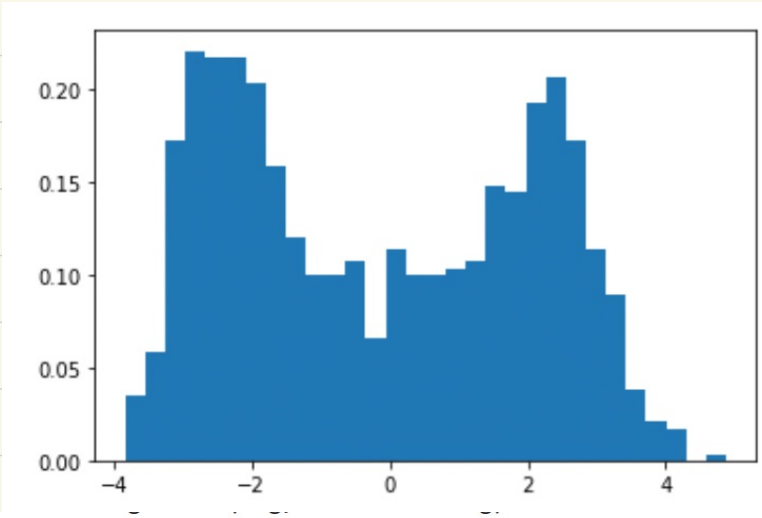
    $$F \leftarrow \frac{1}{N} \sum_{z \in Z} \left[\log\det \nabla^2 \varphi'_\theta(z) \cdot \bar{\mu}(z) - \log\left(\pi_n\left(\nabla \varphi'_\theta(z)\right)\right) \right].$$

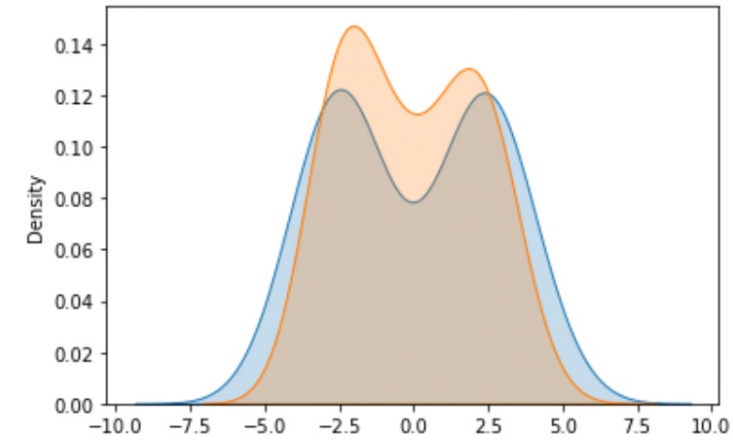
    $$\mathcal{L} \leftarrow F + \frac{1}{2h} W_2^2$$

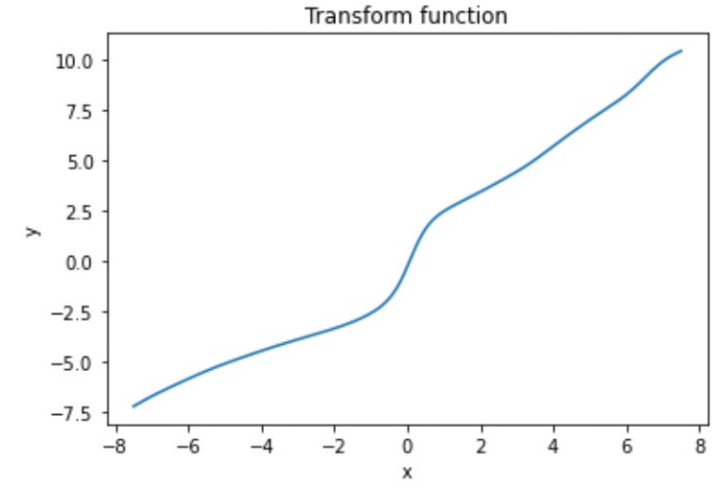
    Perform a gradient step over θ by using $\frac{\partial \hat{\mathcal{L}}}{\partial \theta}$ to update φ'_θ

  $\varphi_\theta \leftarrow \varphi'_\theta$

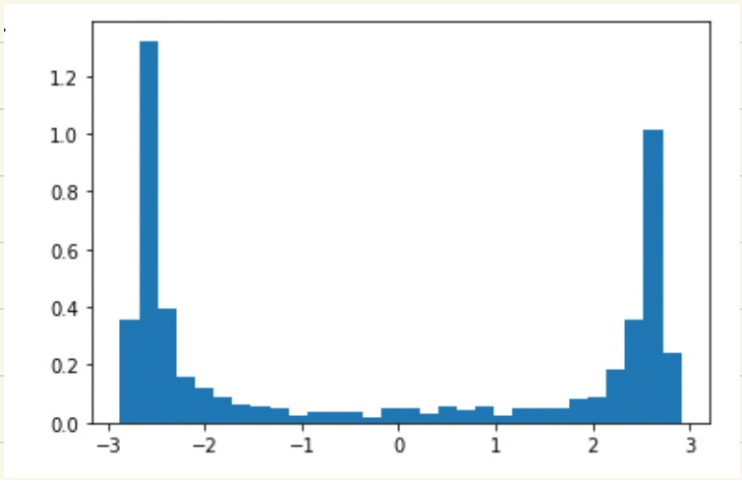
W_2^2 with SGD.

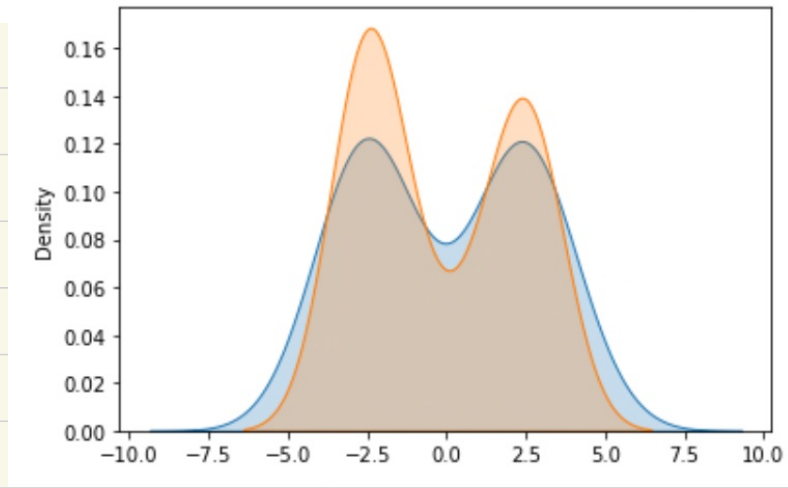


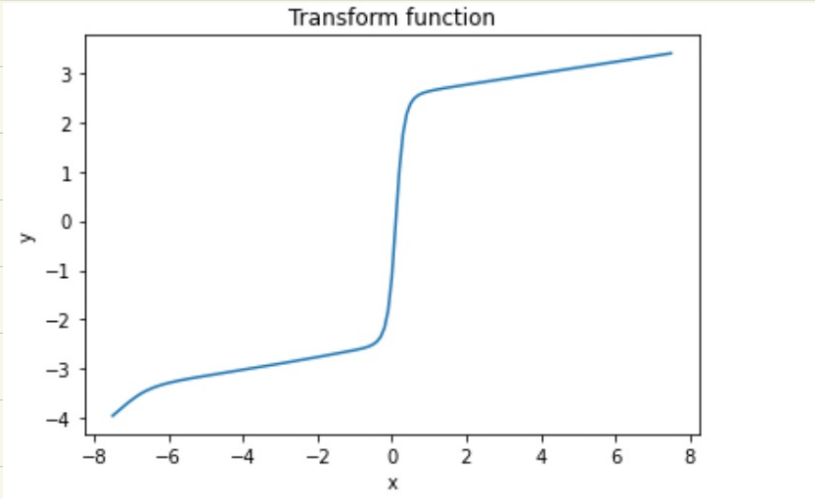




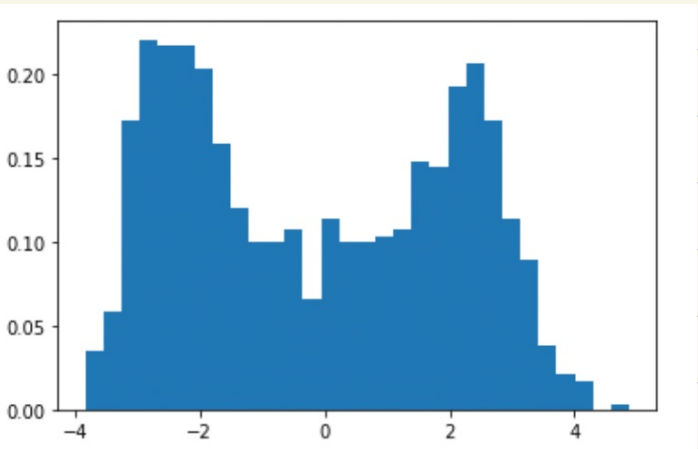
W_2^2 with Adam.

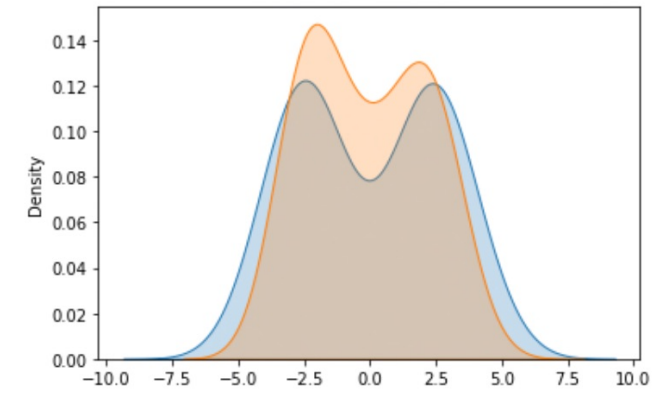


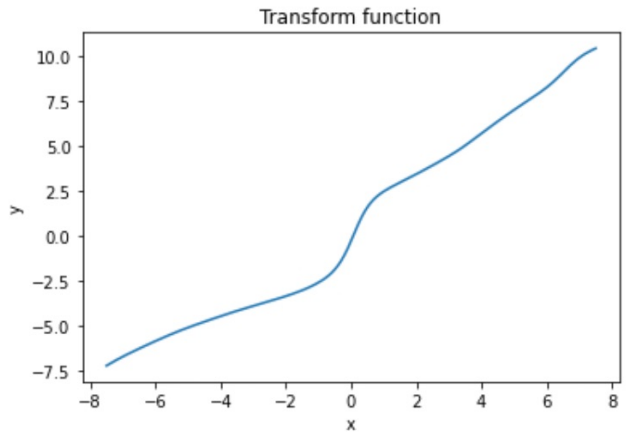




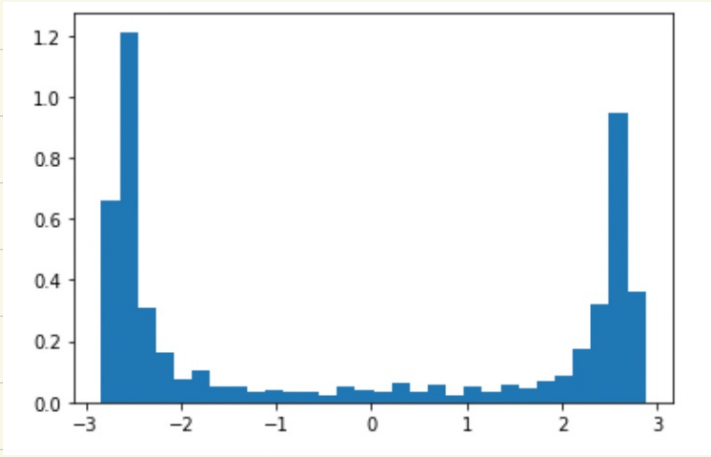
L_2 with SGD

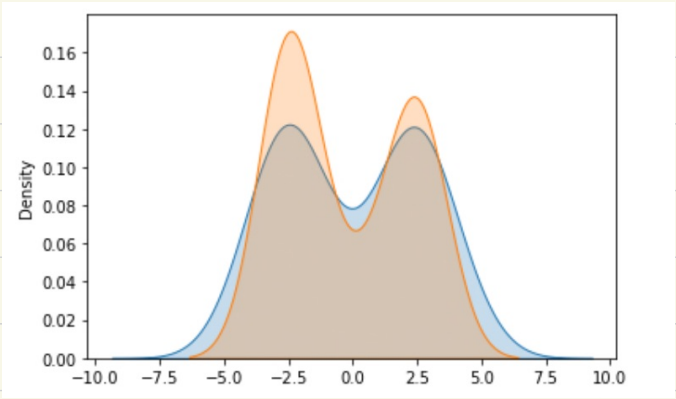


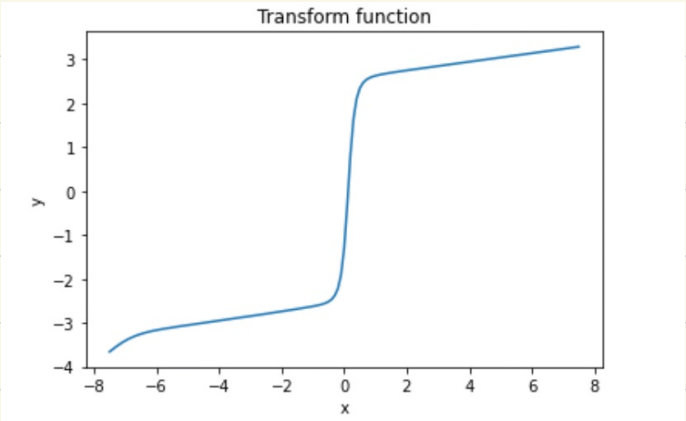




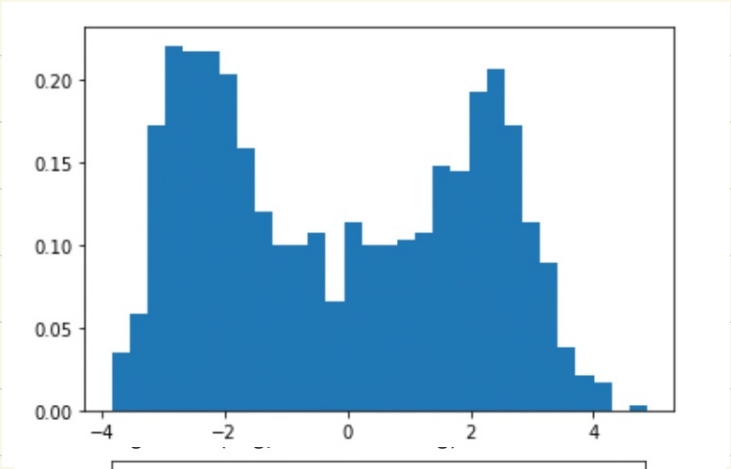
L_2 with Adam.

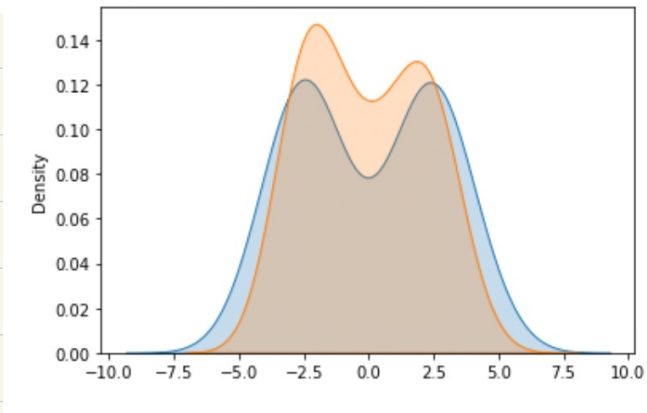


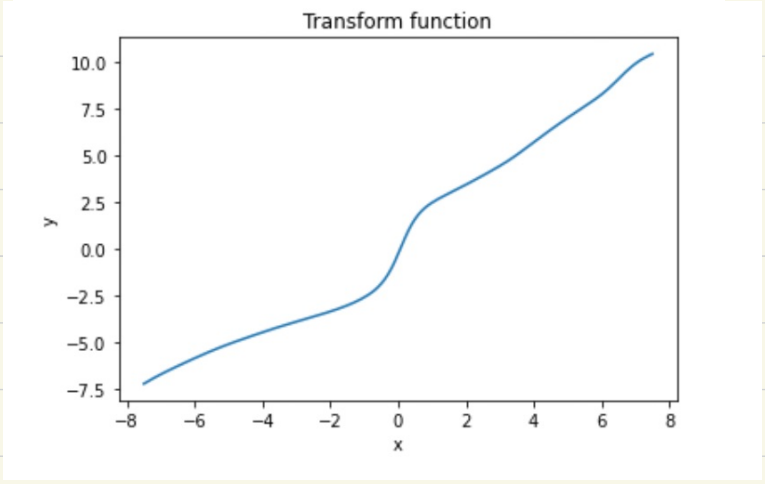




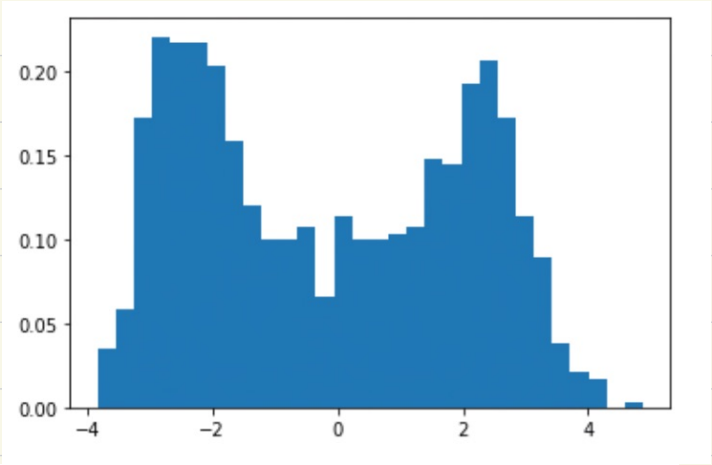
W_2^2 with SGD

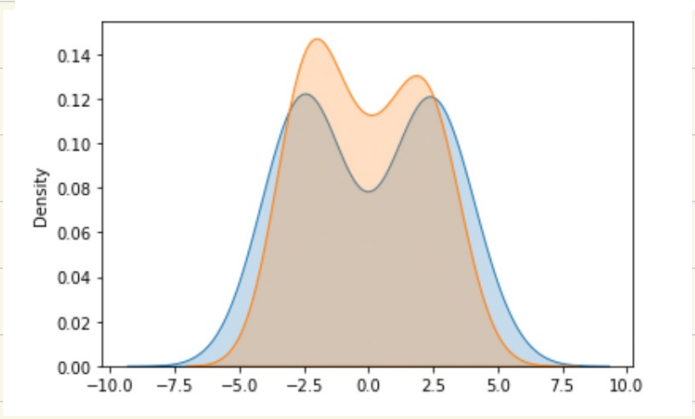


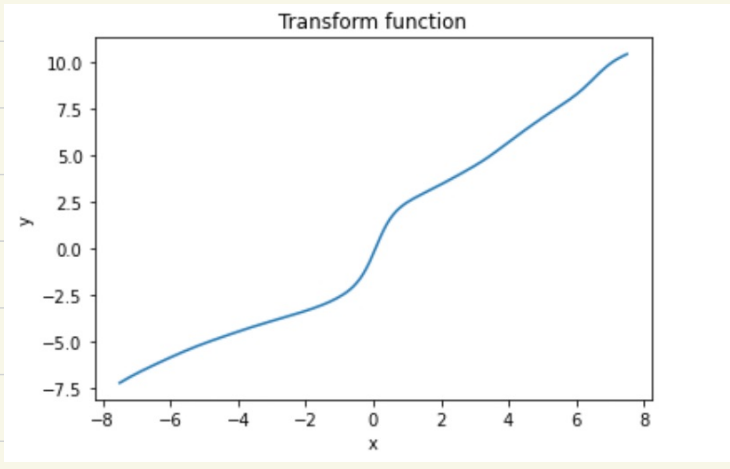




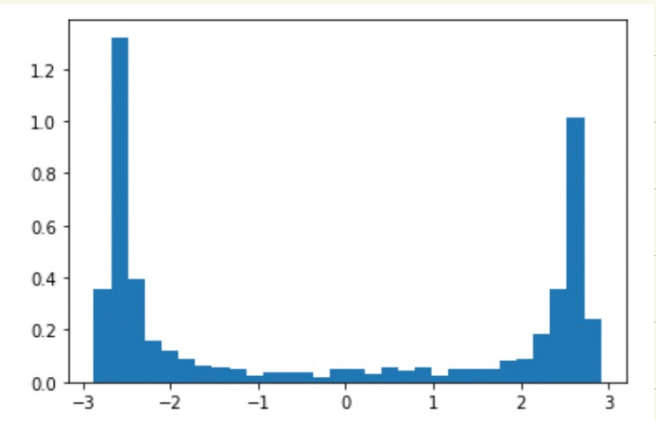
L_2^2 with SGD.

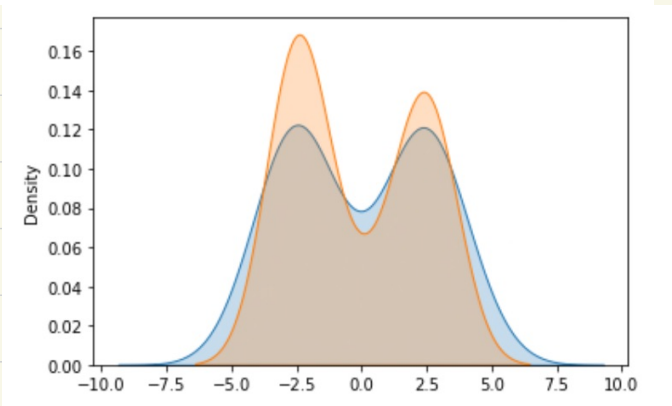


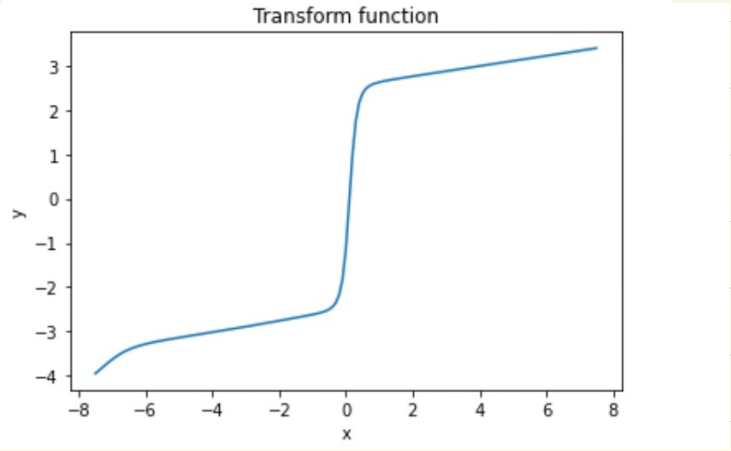




W_2^2 with Adam







L_2^2 with Adam.