So following James' suggestion to modify the loss function to use a Gaussian

$$\eta_\lambda(\Lambda; \Lambda_0, \sigma_\Lambda) = \frac{\lambda}{\sqrt{2\pi\sigma_\Lambda^2}} \exp\left[\frac{(\Lambda - \Lambda_0)^2}{2\sigma_\Lambda}\right], \tag{1}$$

where

$$\Lambda \equiv \Lambda(\xi, \xi') = \delta(\xi, \xi') A(\xi'|\xi), \tag{2}$$

and

$$\delta(\xi, \xi') = \|x - x'\|_2^2 \tag{3}$$

The total loss is then computed by averaging $\eta_\lambda(\Lambda; \Lambda_0, \sigma_\Lambda)$ over all chains.

TODO:

- ☐ Monitor tunneling rate during training.
- ☐ Try increasing the batch size.
- ☐ Try changing the number of leapfrog steps.
- ☐ Try different annealing schedule.
- ☐ Modify the **loss function** as follows:
  1. Let $\Lambda \equiv \Lambda(\xi, \xi') = \delta(\xi', \xi) A(\xi'|\xi)$, where $\delta(\xi', \xi) = \|x - x'\|_2^2$
  2. In terms of Λ, we can write the $\ell_\lambda(\xi, \xi', A(\xi'|\xi))$ (Eq. 7 from the L2HMC paper) as
  $$\ell_\lambda(\xi, \xi', A(\xi'|\xi)) = \frac{\lambda^2}{\Lambda(\xi, \xi')} - \frac{\Lambda(\xi, \xi')}{\lambda^2}. \tag{4}$$
  3. We are interested to see how the model behaves if we replace ℓ_λ with a quadratic gaussian, η_λ:
  $$\ell_\lambda(\xi, \xi') \longrightarrow \eta_\lambda(\xi, \xi') \equiv \exp\left[\frac{(\Lambda - \Lambda_0)^T (\Lambda - \Lambda_0)}{2\,\sigma_\Lambda^2}\right] \tag{5}$$
  where Λ_0 is a parameter related to the "average distance" between the initial and proposed configuration and σ_Λ^2 controls the width of the Gaussian.
- ☐ Get a reasonable estimate of the **integrated autocorrelation length**.

- ☐ Look at Neal's proof and see if there's anything in the L2HMC algorithm that might violate reversibility.
- ☐ Try reducing the step size ε during inference by some multiplicative factor $\varepsilon \to \alpha \cdot \varepsilon$ so that the acceptance probability $p_x \to 1$.
  - In doing so we need to increase the number of run steps by $1/\alpha$.
  - See if the bias in the average plaquette scales with ε^2.
- ☐ Try and figure out why the Jacobian isn't (*exactly*) the same for the forward and backward updates.
- ☐ Try removing the forward/backward masks during inference (i.e. run strictly forward or run strictly backward)
- ☐ Try eliminating either the first x update (i.e. $x \to x'$) or second x update ($x' \to x''$) when running inference by setting either $m^t = [1, 1, \ldots, 1]$ and $\bar{m}^t = [0, 0, \ldots, 0]$ or vice versa (i.e. removing the mask when running inference).

Completed:

- ☑ Implement reversibility checker that ensures that the reversibility condition holds (for both position and momentum)
- ☑ Try with 'float64'. (**Error still present.**)
- ☑ Try as few as 10 hidden nodes and see if the error persists. (**It does.**)
- ☑ Try anti-symmetric Gaussian Mixture Model and see if the trained model is an accurate representation of the target distribution (e.g. by looking at the locations of the means) (link to post).
- ☑ Loop over `net_weights = [scale_w, translation_w, transformation_w]` from $[0, 0, 0]$ up to $[2, 0, 0]$, $[0, 2, 0]$, $[0, 0, 2]$, and see how the error in the average plaquette changes (link to post).