AMATH 563 - HOMEWORK 5 (COMPUTATIONAL REPORT)

NATE WHYBRA

Applied Mathematics Department, University of Washington, Seattle, WA
nwhybra@uw.edu

1. Introduction

Given two probability distributions on $\mathbb{R}^n$, a source η and a target ν, the problem of finding a map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that "T pushes η to ν", ie $T(\eta) = \nu$ is called a transport problem. In this paper, we use the Reproducing Kernel Hilbert Space (RKHS) formulation of transport to learn maps $T : \mathbb{R}^2 \to \mathbb{R}^2$ which push $\eta \sim N(0, I) \to \nu$, where our target distributions ν are the "moons", "swissroll", and "pinwheel" benchmark images from **[1]**. More specifically our process of learning these map is to use Kernel Ridge Regression (KRR) and experimenting with various Kernel functions (RBF, Laplacian, Polynomial) with different sets of hyper-parameters. The data sets used in this study are shown below.

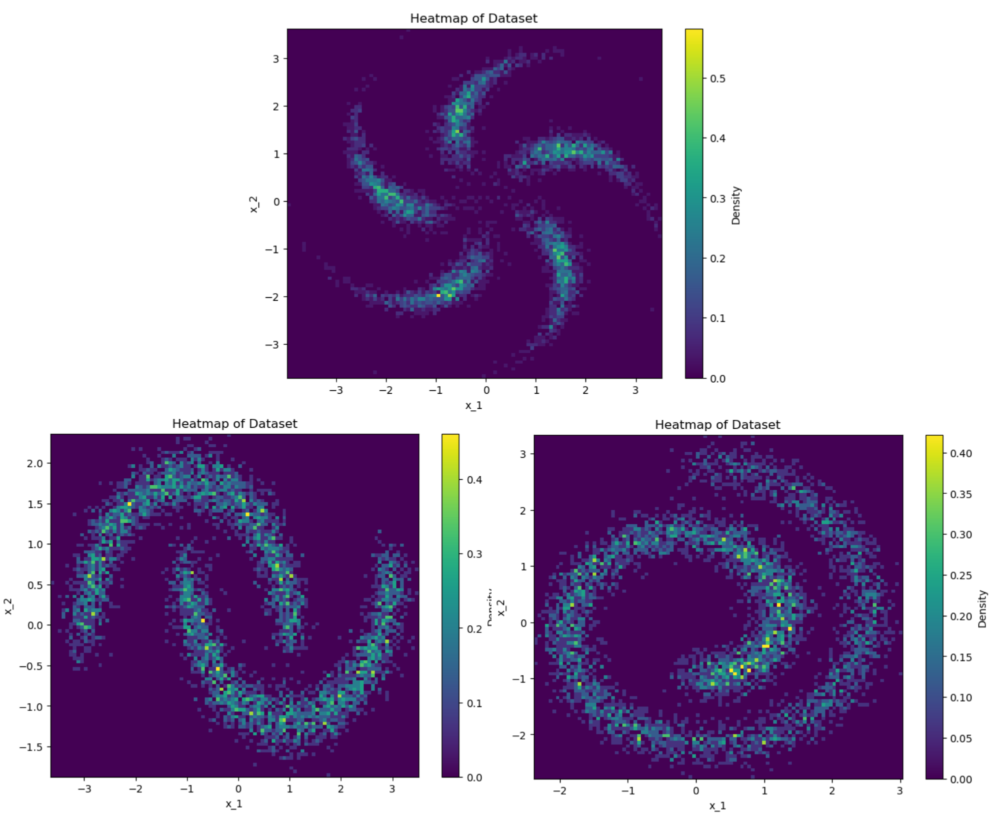


Figure 1. Target distributions ν: Pinwheel (Top Middle), Moons (Bottom Left), Swiss Roll (Bottom Right).

Date: 5/26/2025.

2. Methods

Given a statistical divergence function $D : P(\mathbb{R}^2) \times P(\mathbb{R}^2) \to \mathbb{R}$ and a regularizer R, the problem of finding an optimal transport $T^* : \mathbb{R}^2$ function that pushes $Z \to X$ can be formulated as:

$$T^* = \text{argmin}_{T \in H} \left[D(T(Z), X) + R(T) \right]$$

Where H is an RKHS of functions with associated kernel function K. For us, D is the "maximum mean discrepancy" (MMD):

$$D(u, u') = MMD(X, X') = \frac{1}{N^2} \left(\sum_{i,j=1}^{N} K(x_i, x_j) + K(x'_i, x'_j) - 2K(x_i, x'_j) \right)$$

Where X and X' are distributions, but in our application are really just $N \times 2$ matrices where each column is a coordinate in $\mathbb{R}^2$. In correspondence with RKHS theory we should really take $R(T) = (\lambda/2) \|T\|_H^2$ to guarantee a unique solution, however we found that using $R(T) = \lambda \|\alpha\|_2^2$ where $\alpha \in \mathbb{R}^{n \times 2}$ is the corresponding weight coefficient matrix to provide stable computation and satisfactory solutions.

$$K(x, x') = \exp\left(-\frac{(x - x')^2}{2\sigma^2} \right) \quad \text{RBF Kernel}$$

$$K(x, x') = \exp\left(-\frac{|x - x'|}{2\sigma^2} \right) \quad \text{Laplacian Kernel}$$

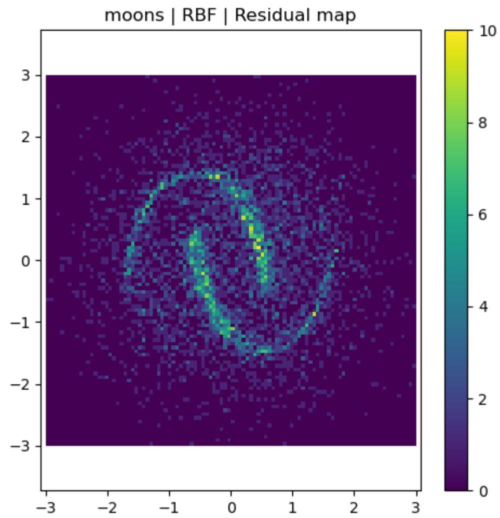
$$K(x, x') = (1 + x^T x')^d \quad \text{Polynomial Kernel } (d = 1, 2, 4)$$

To find suitable values for the length scale parameter σ we use the so-called median heuristic as a starting point for hand-tuning. To actually train our models, we used the Adam optimization/automatic differentiation algorithm built into PyTorch. As a stopping criterion for training, we set an arbitrary tolerance of 0.004 for the loss combined with the norm of the gradient having a tolerance of 0.003. Our results are summarized in the next sections.

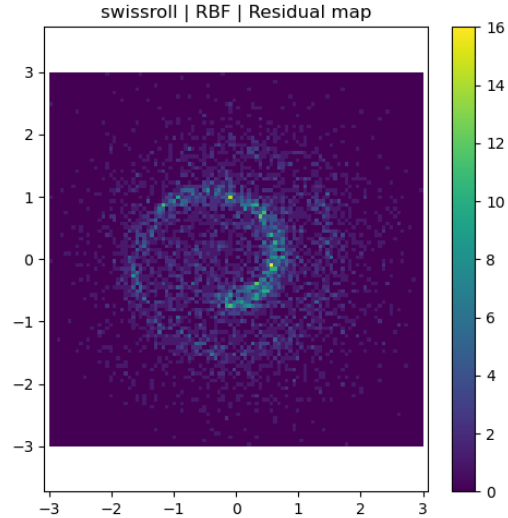
3. Results

Distribution/Kernel	RBF	Laplacian	Polynomial (d=2)
moons	0.01	0.019	0.37
swissroll	0.0022	0.45	0.022
pinwheel	0.0017	0.013	0.0010

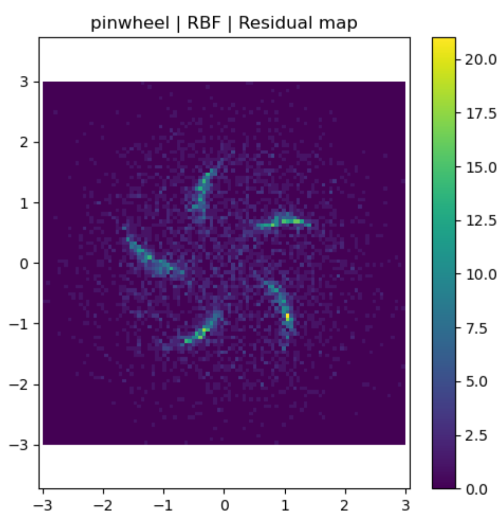
Table 1. MDD of Target Distributions vs. Kernel Function (2 decimals places). Note that the Target Distributions were first standard normalized before this computation.



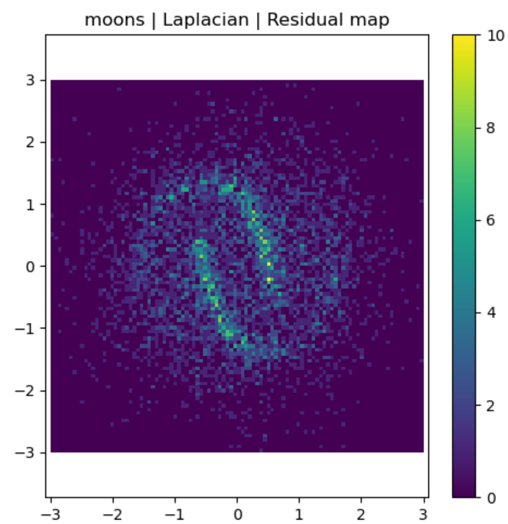
(A) moons RBF Kernel, $\sigma = 0.1$, $MMD = 0.0034$



(B) swissroll RBF Kernel, $\sigma = 0.1$, $MMD = 0.0031$

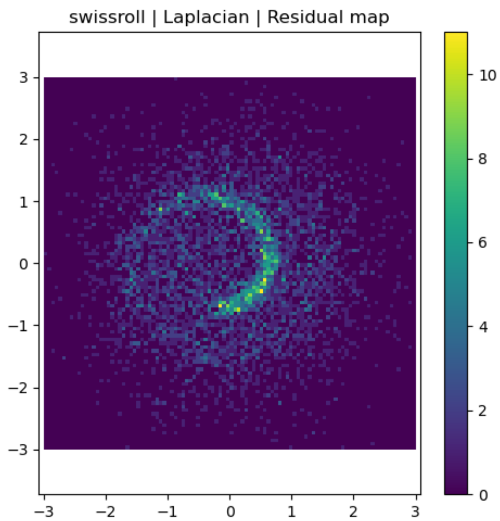


(C) pinwheel RBF Kernel, $\sigma = 0.1$, $MMD = 0.0062$

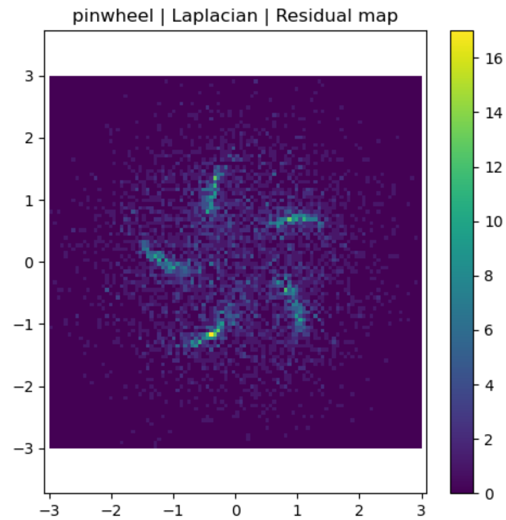


(D) moons Laplacian Kernel, $\sigma = 0.1$, $MMD = 0.0045$

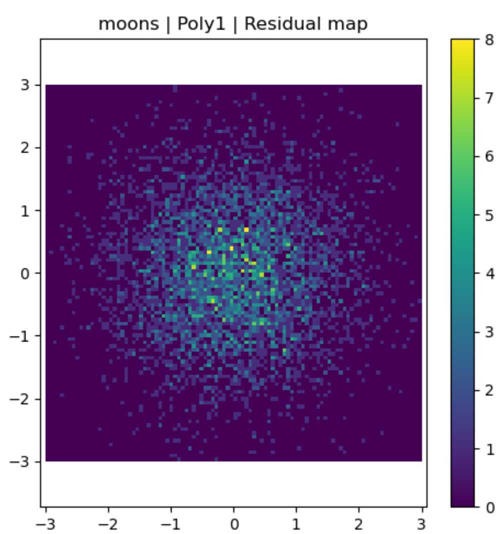
FIGURE 2. Histograms of trained transport models, Part 1.



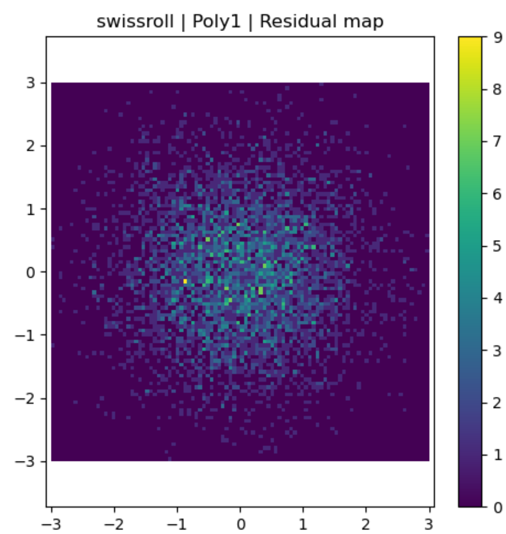
(A) swissroll Laplacian Kernel, $\sigma = 0.1$, $MMD = 0.0031$



(B) pinwheel Laplacian Kernel, $\sigma = 0.1$, $MMD = 0.0061$

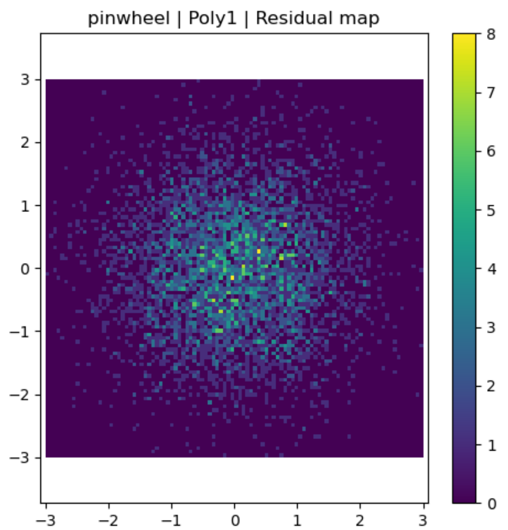


(C) moons Polynomial Kernel, $d = 1$, $MMD = N/A$

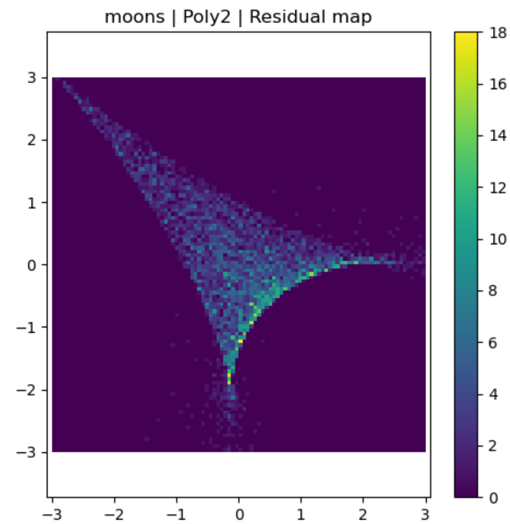


(D) swissroll Polynomial Kernel, $d = 1$, $MMD = N/A$

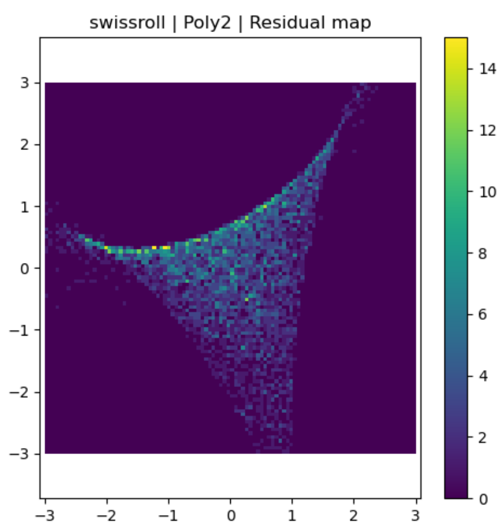
FIGURE 3. Histograms of trained transport models, Part 2.



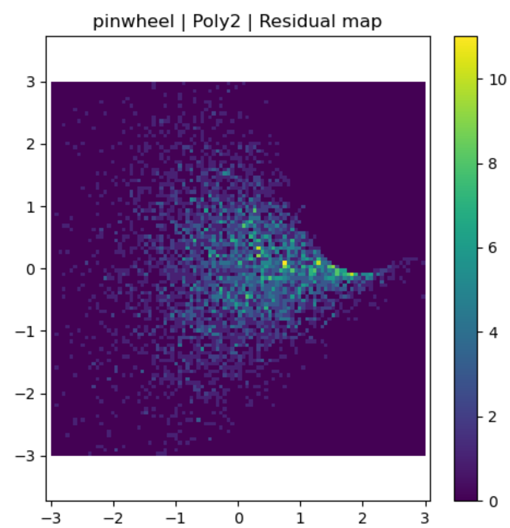
(A) pinwheel Polynomial Kernel, $d = 1$, $MMD = N/A$



(B) moons Polynomial Kernel, $d = 2$, $MMD = 0.00022$

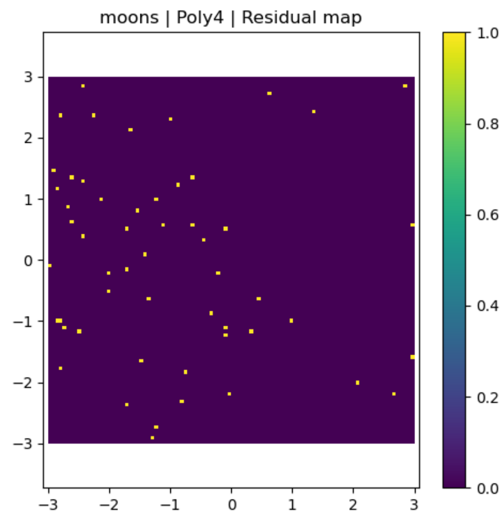


(C) swissroll Polynomial Kernel, $d = 2$, $MMD = 0.000018$

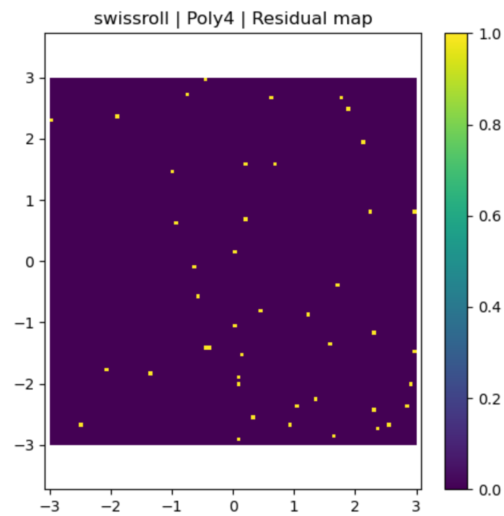


(D) pinwheel Polynomial Kernel, $d = 2$, $MMD = 0.0013$

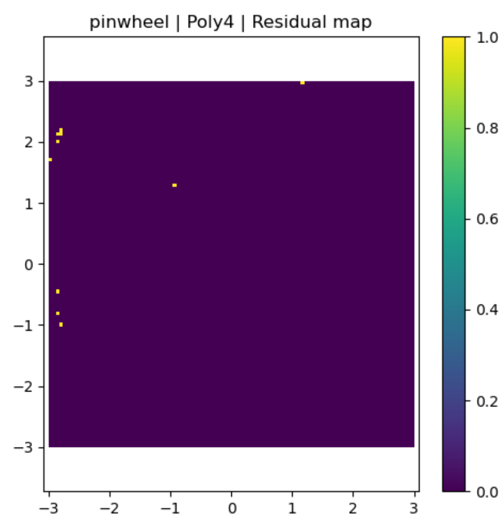
FIGURE 4. Histograms of trained transport models, Part 3.



(A) moons Polynomial Kernel, $d = 4$, $MMD =$ very large



(B) swissroll Polynomial Kernel, $d = 4$, $MMD =$ very large



(C) pinwheel Polynomial Kernel, $d = 4$, $MMD =$very large

FIGURE 5. Histograms of trained transport models, Part 4.

4. Summary and Conclusions

After training many models, we are able to find qualitatively successful transport maps using the RBF and Laplacian Kernels, with the same set of hyper-parameters. Here qualitatively means that they look similar when you look at their heat maps. Quantitatively however, although we have small MMD values for these models, this seems to not be the best indicator of model performance here. Although the the models found using the Polynomial Kernels do not resemble the target distribution whatsover, they have either very small or very large MMD values. This likely coincides with numerical averaging effects and potentially large or small differences when either of the distributions has 0 pixels. Although not shown here, changing the value of N had significant effects on the quality of the models, but similar patterns highlighting the more obvious features seemed to be found by the RBF and Laplacian kernels as well in this case.

Acknowledgments

This time I really have nobody to thank! This course was pretty cool! This is the last homework I have to turn in for my masters program, which is exciting. On to the next chapter I suppose. Anyways, see you next time!

References

[1] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations (ICLR)*, 2019.

[2] B. Hosseni. Lecture notes for AMATH 563: Lecture Notes. Lecture notes distributed in class, 2025. University of Washington, unpublished lecture notes.

[2] [1]