CS 726 : Course Project

200050125, 200050153, 200050160

February 2023

Contents

1 Results for Helix

1.1 Number of training steps

We varied the number of training steps and batch size keeping other hyperparameters fixed. We considered 500, 1000, 2000 epochs and batch sizes of 512, 1024 and performed six experiments. We used EMD and NLL as performance metrics. The results are summarised as follows. Other hyperparameters : Number of steps : 200, Complexity of model : 4 linear and 4 ReLU layers.

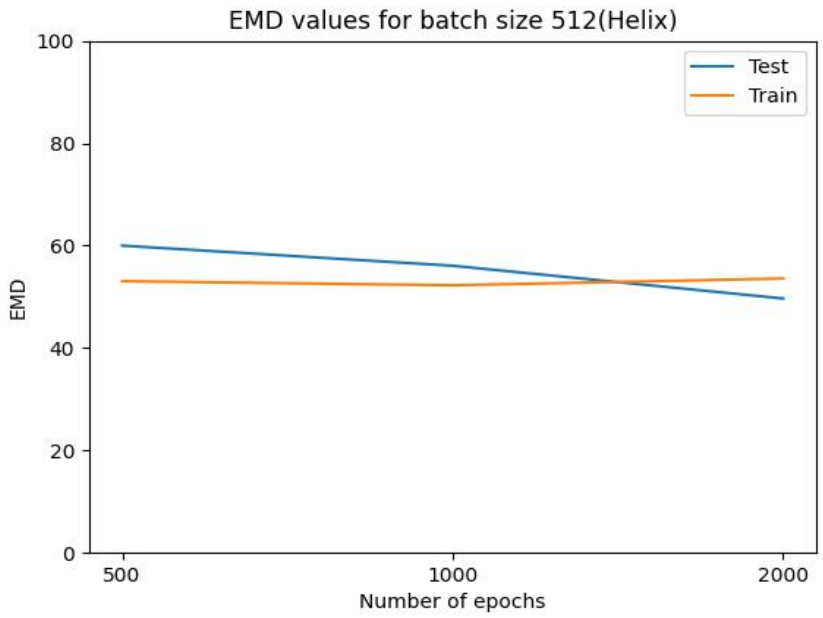

Figure 1: EMD values for batch size 512

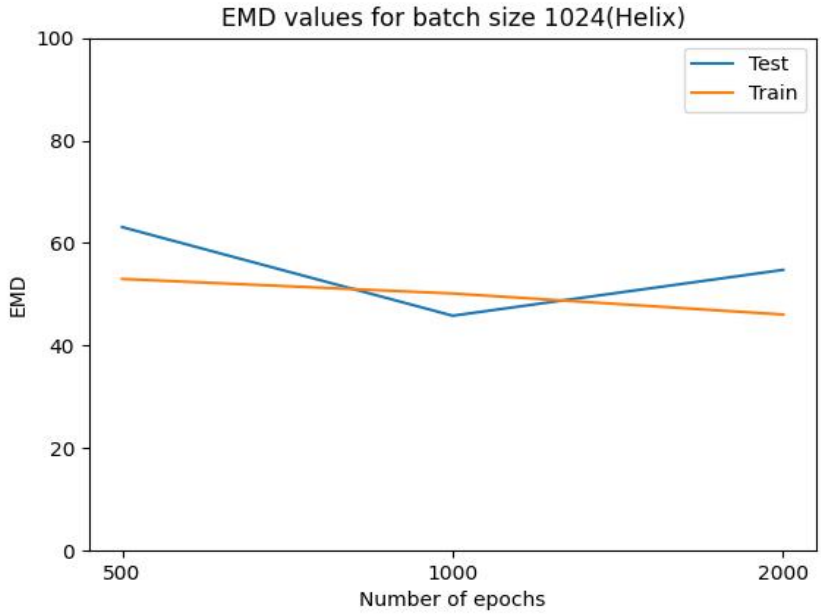

Figure 2: EMD values for batch size 1024

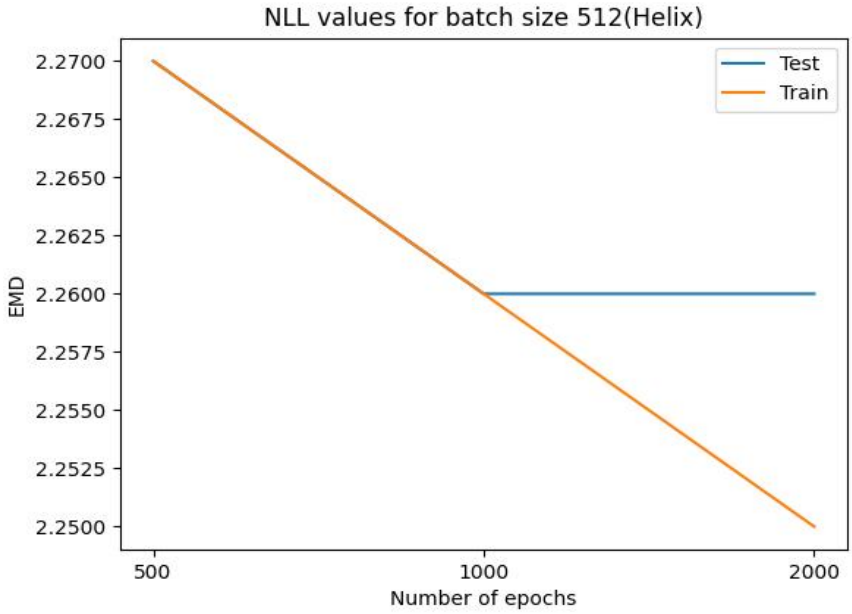

Figure 3: NLL values for batch size 512

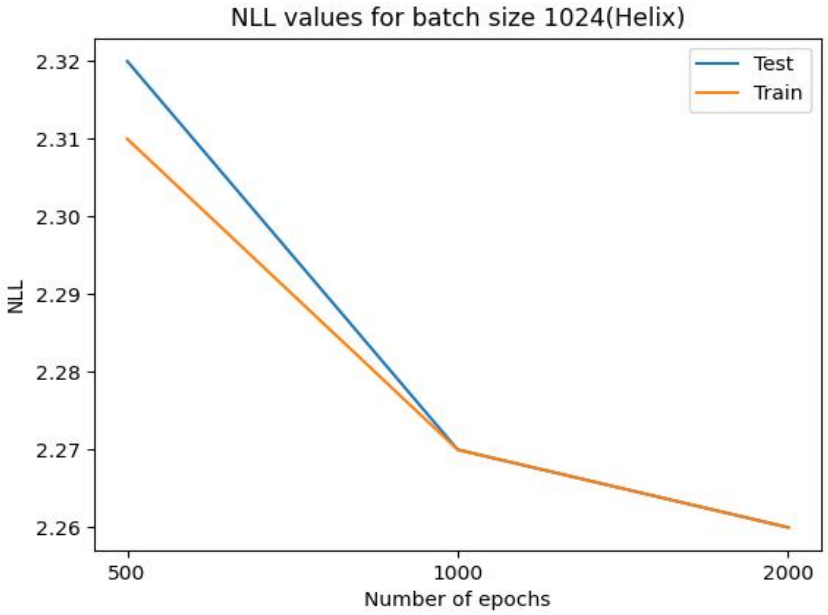

Figure 4: NLL values for batch size 1024

We concluded that training for 1000 epochs with a batch size 1024 gives us the best results for Helix dataset. We will use these values for all subsequent experiments.

1.2 Model Complexity

Here, we analysed three different models, named model 1, 2, 3 respectively. Model 1 contains four linear layers and four RelU layers. Model 2 contains six linear layers and six ReLU layers. Model 3 contains eight linear layers and eight ReLU layers. The number of layers is the same in time_embed as well as model in each of the above cases. We observe that model 1 with 4 layers gives us the best performance. We shall use this model for all subsequent experiments.

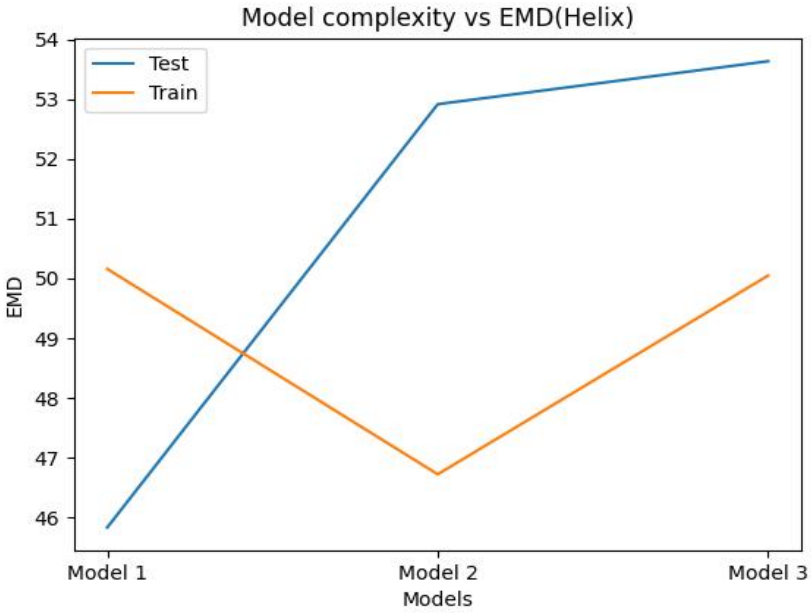

Figure 5: Comparison of EMD values for the three models

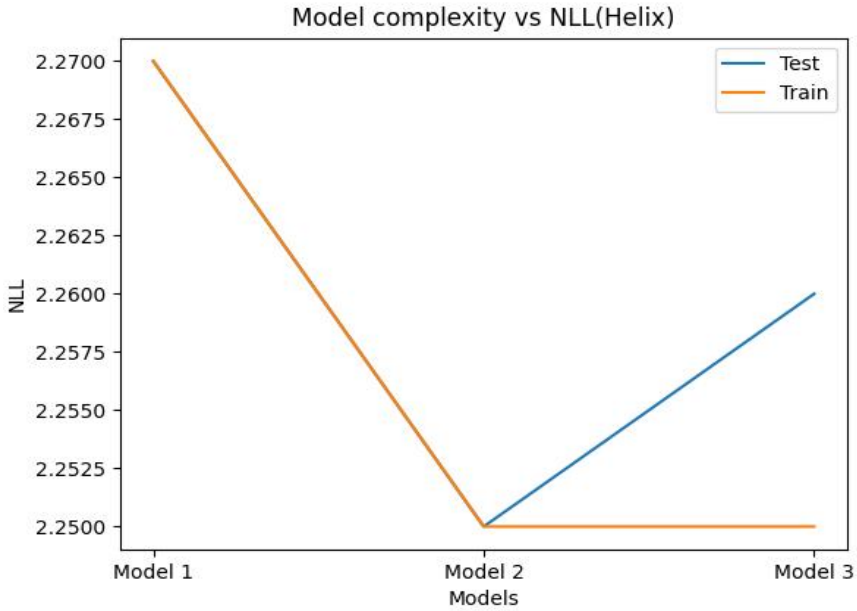

Figure 6: Comparison of NLL values for the three models

1.3 Number of training steps

We varied the number of training steps and experimented with 10, 50, 100, 150, 200 steps keeping the other hyperparameters fixed. Here we considered only EMD for performance evaluation as NLL values turned out to be pretty close for all cases(specifically for test and train). The results are summarized as follows :

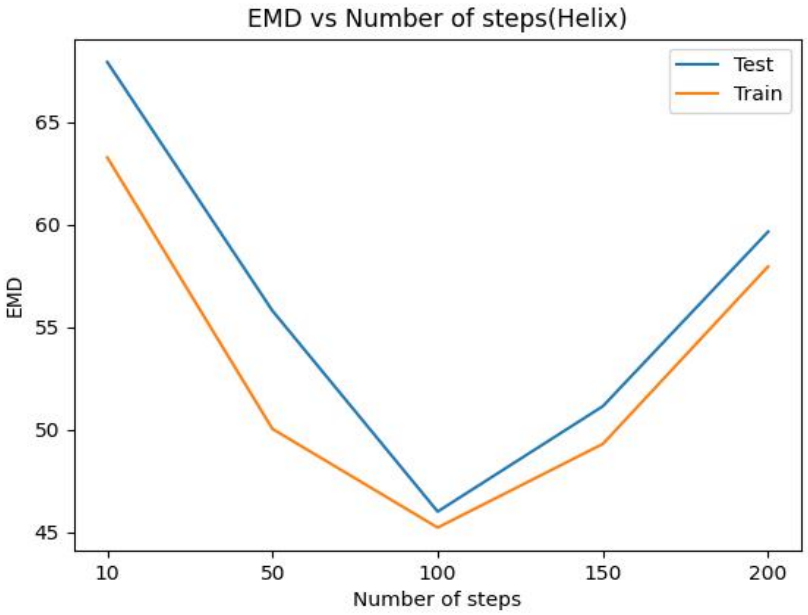

Figure 7: Comparison of EMD values for different number of training steps

We can observe that training with 100 steps gives us the best EMD. These are the final graphs obtained for each of the number of training steps mentioned above.

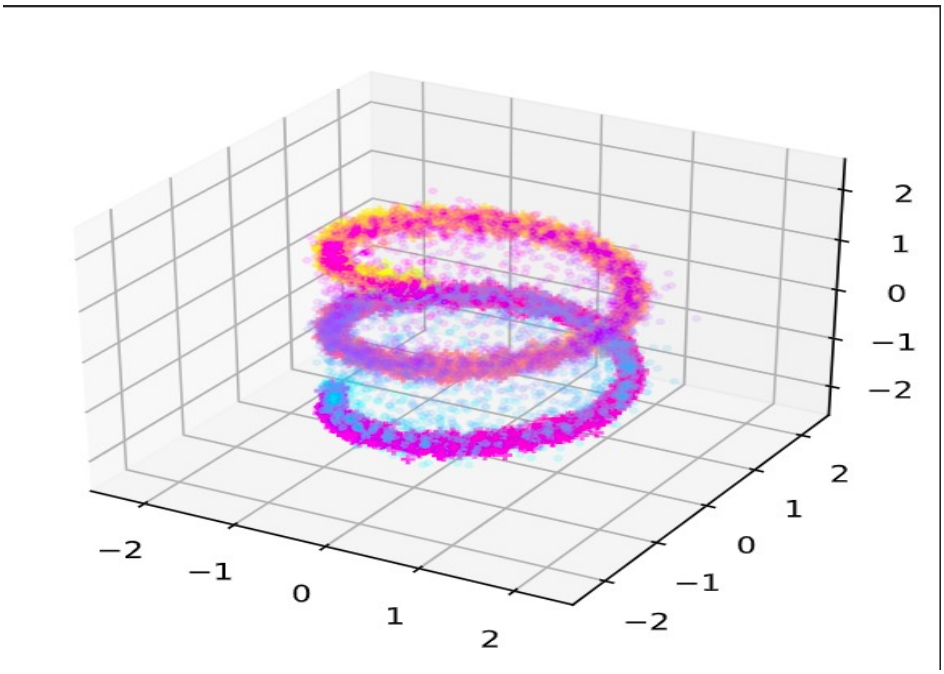

Figure 8: 10 steps

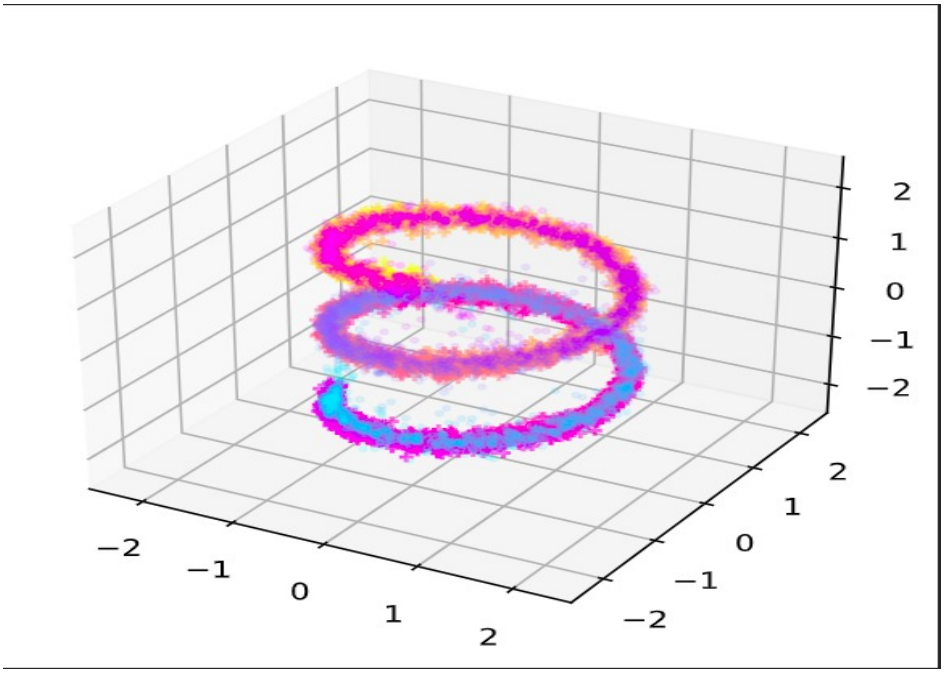

Figure 9: 50 steps

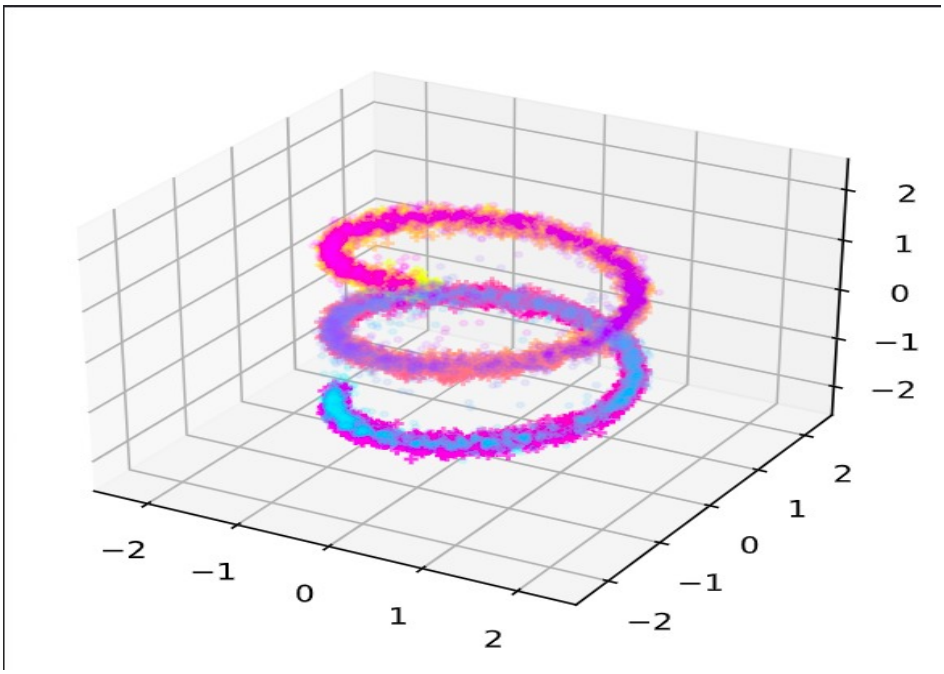

Figure 10: 100 steps

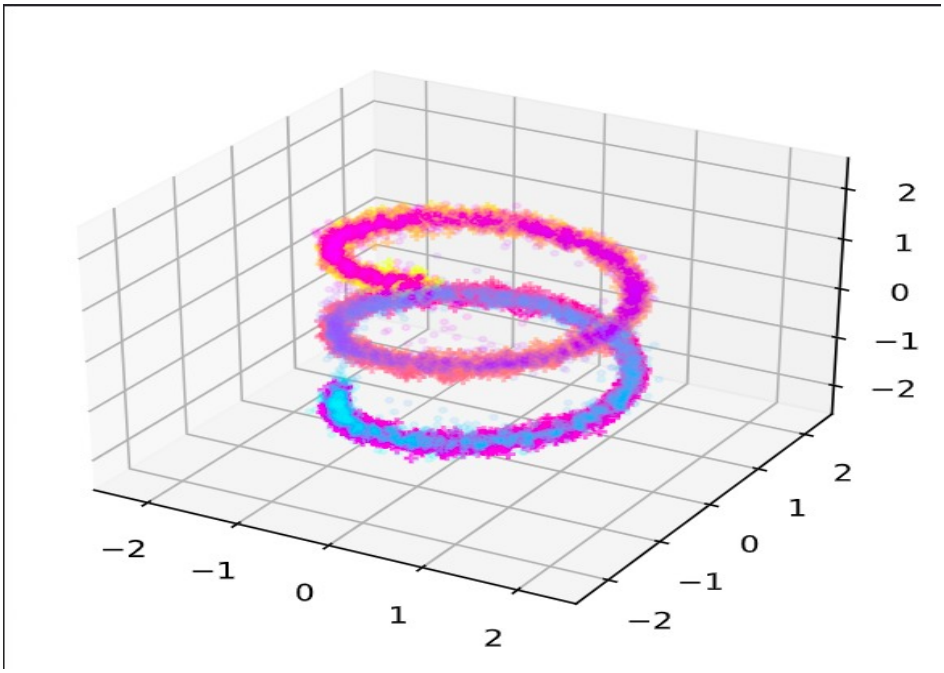

Figure 11: 150 steps

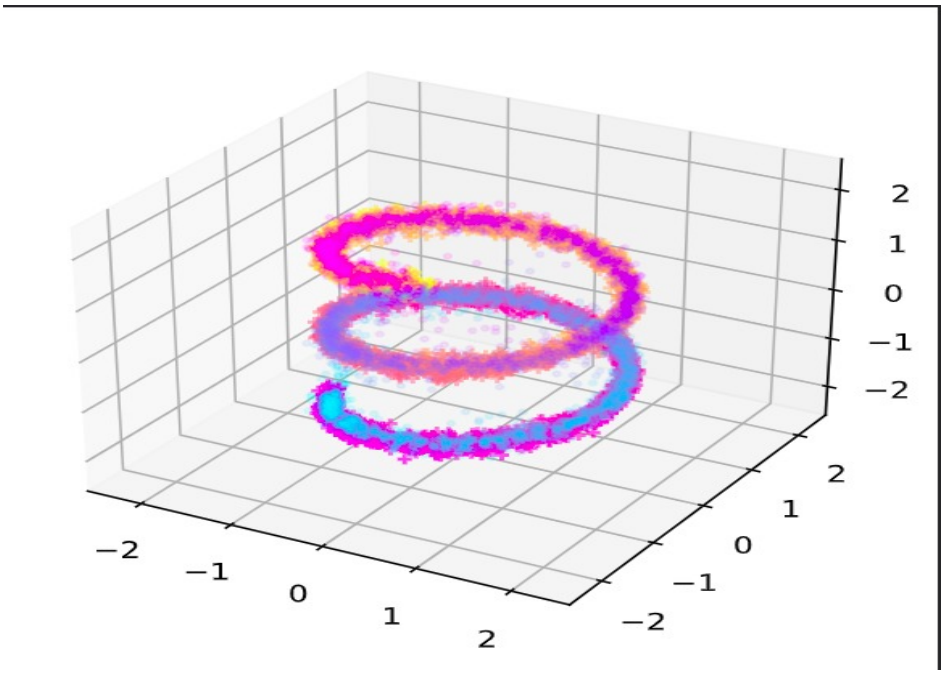

Figure 12: 200 steps

1.4 Noise scheduling

We implemented 5 different types of noise scheduling :

- Linear
- Quadratic
- Bi-Quadratic
- Sigmoid
- Cosine

We obtained the EMD values for each scheduling method and plotted the results as follows :

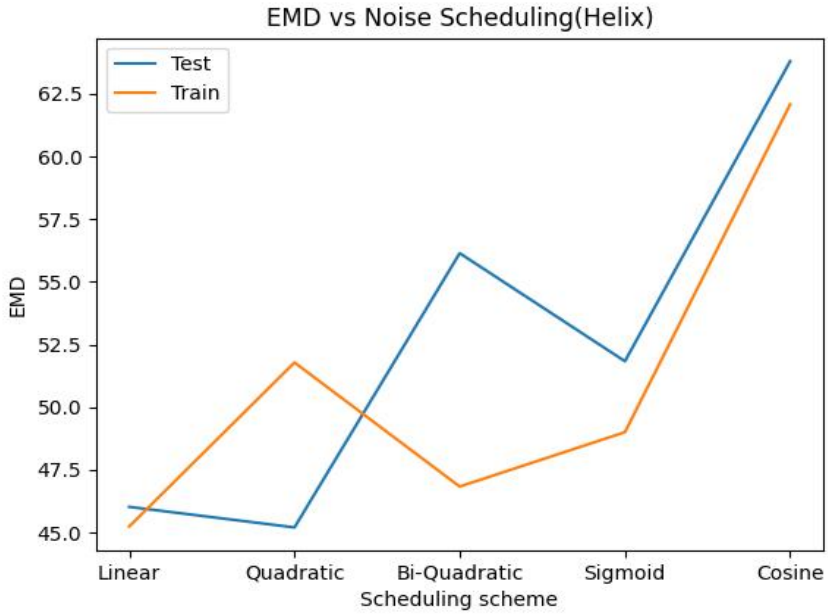

Figure 13: EMD values for noise scheduling schemes

We can see that quadratic scheduling gives the best EMD.

2 Results for Sine

2.1 Number of training steps

We varied the number of training steps and batch size keeping other hyperparameters fixed. We considered 500, 1000, 2000 epochs and batch sizes of 512, 1024 and performed six experiments. We used EMD and NLL as performance metrics. The results are summarised as follows. Other hyperparameters : Number of steps : 100, Complexity of model : 4 linear and 4 ReLU layers.

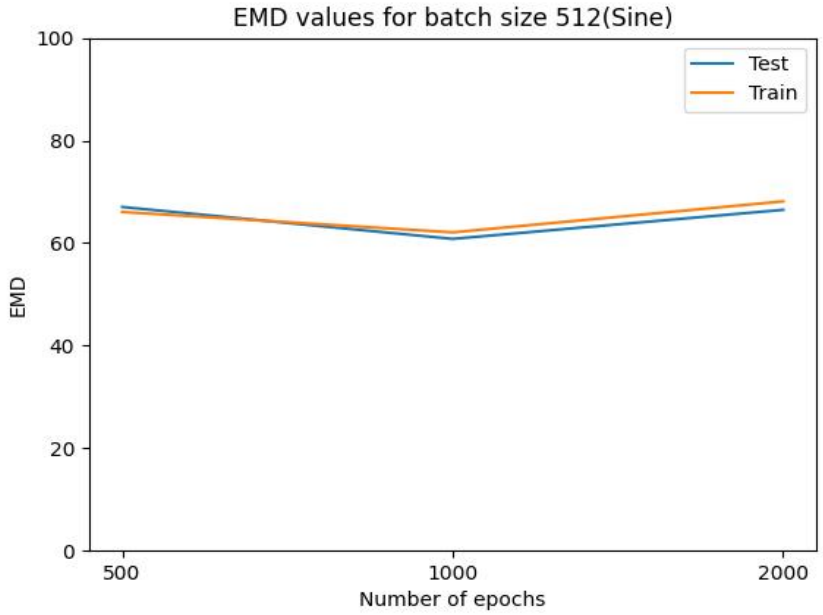

Figure 14: EMD values for batch size 512

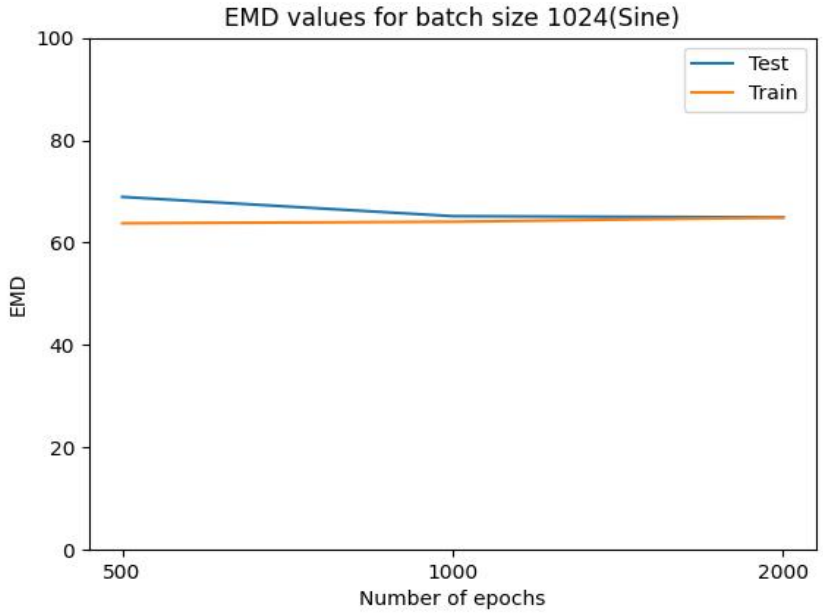

Figure 15: EMD values for batch size 1024

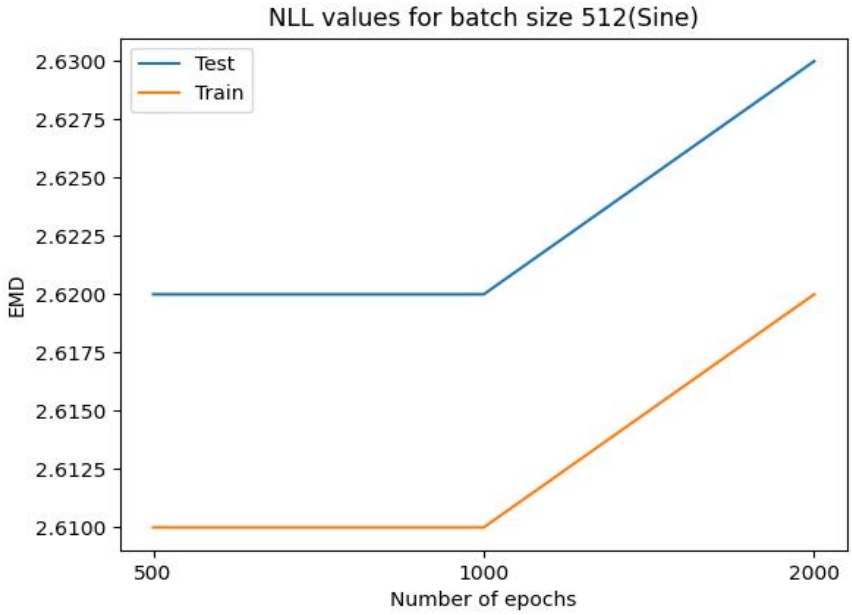

Figure 16: NLL values for batch size 512

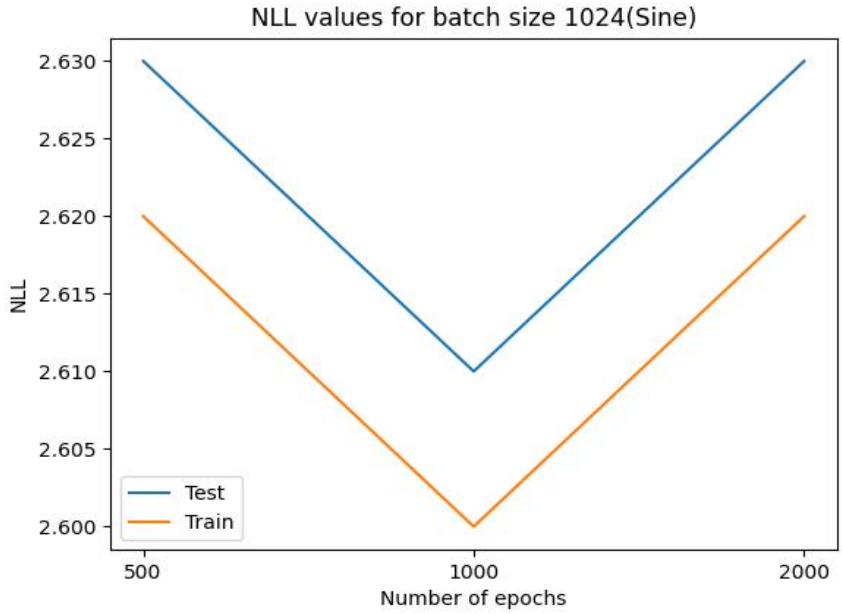

Figure 17: NLL values for batch size 1024

We concluded that training for 1000 epochs with a batch size 512 gives us the best results for Sine dataset. We will use these values for all subsequent experiments.

2.2 Model Complexity

Here, we analysed three different models, named model 1, 2, 3 respectively. Model 1 contains four linear layers and four RelU layers. Model 2 contains six linear layers and six ReLU layers. Model 3 contains eight linear layers and eight ReLU layers. The number of layers is the same in time_embed as well as model in each of the above cases. We observe that model 1 with 4 layers gives us the best performance. We shall use this model for all subsequent experiments.

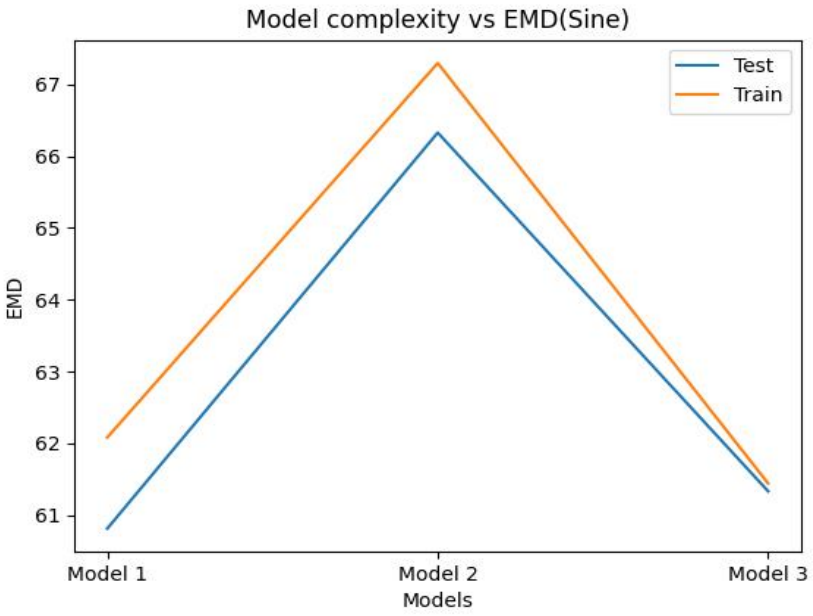

Figure 18: Comparison of EMD values for the three models

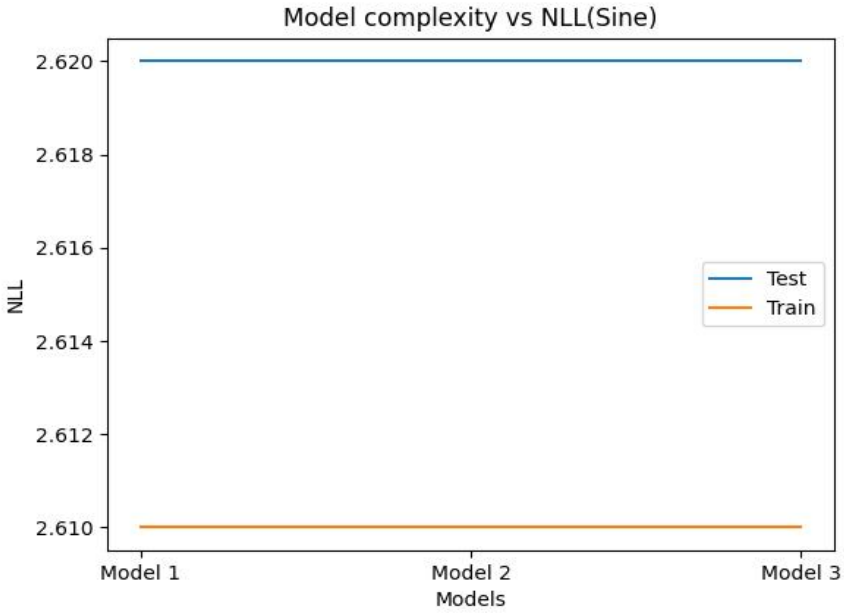

Figure 19: Comparison of NLL values for the three models

2.3 Number of training steps

We varied the number of training steps and experimented with 10, 50, 100, 150, 200 steps keeping the other hyperparameters fixed. Here we considered only EMD for performance evaluation as NLL values turned out to be pretty close for all cases(specifically for test and train). The results are summarized as follows :

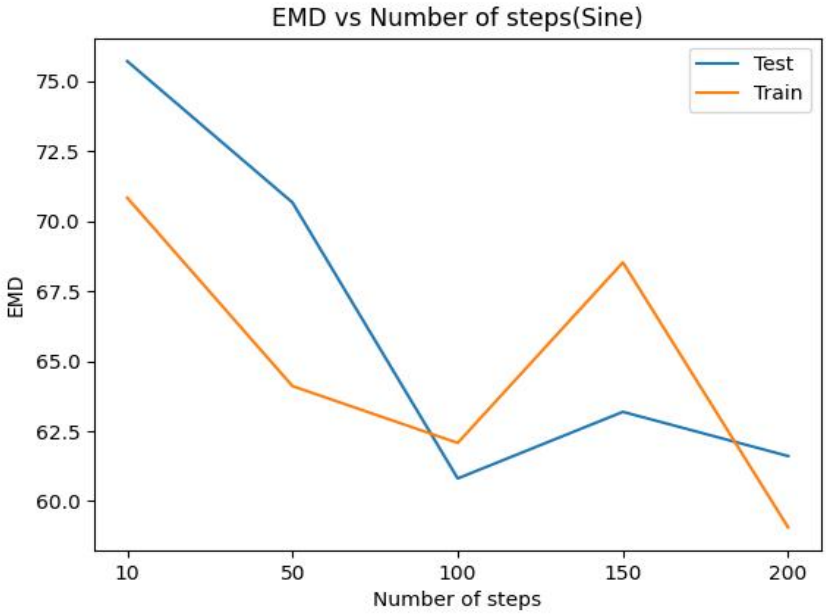

Figure 20: Comparison of EMD values for different number of training steps

We can observe that training with 100 steps gives us the best EMD. These are the final graphs obtained for each of the number of training steps mentioned above.

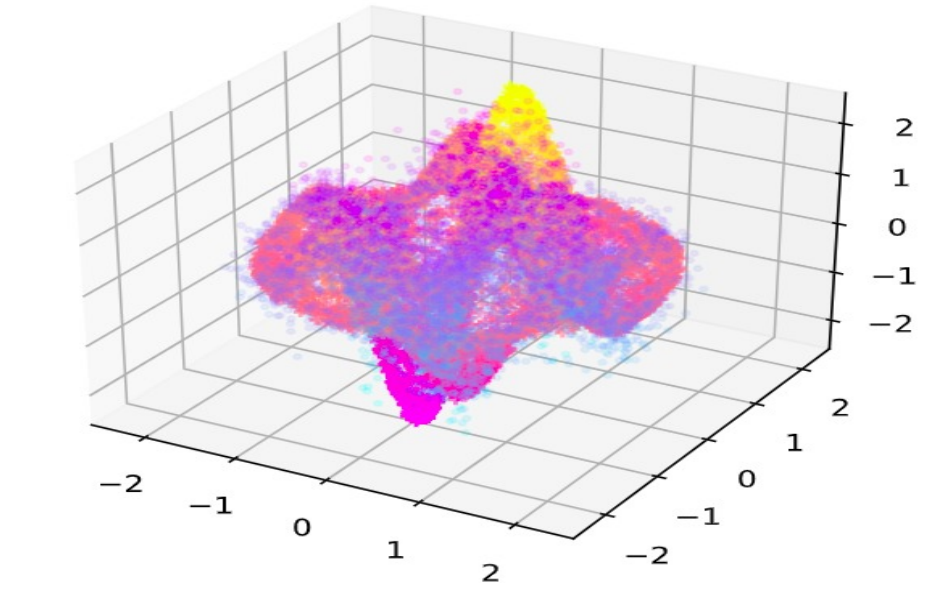

Figure 21: 10 steps

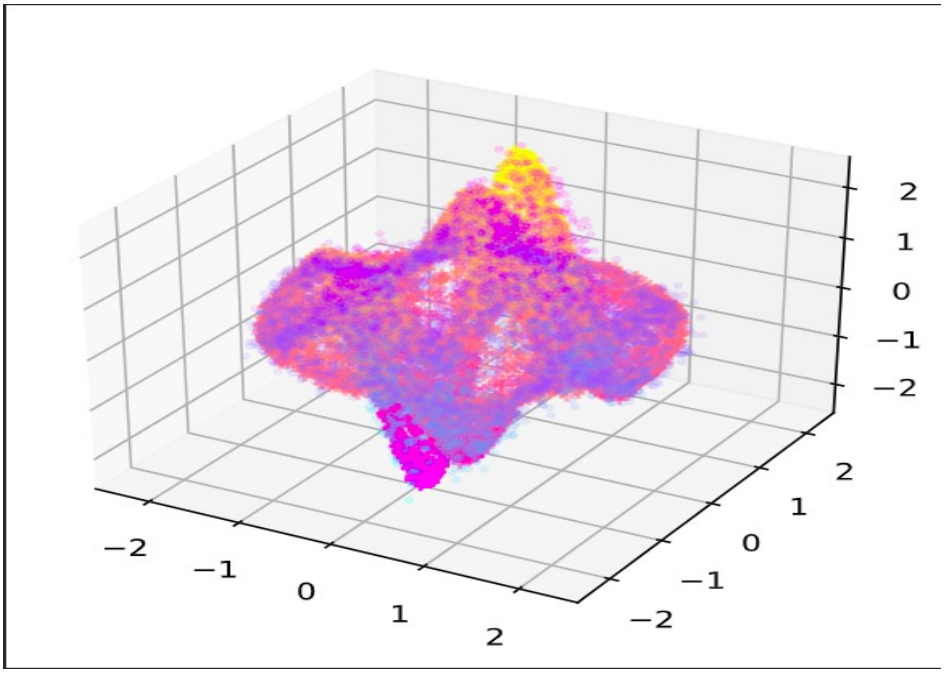

Figure 22: 50 steps

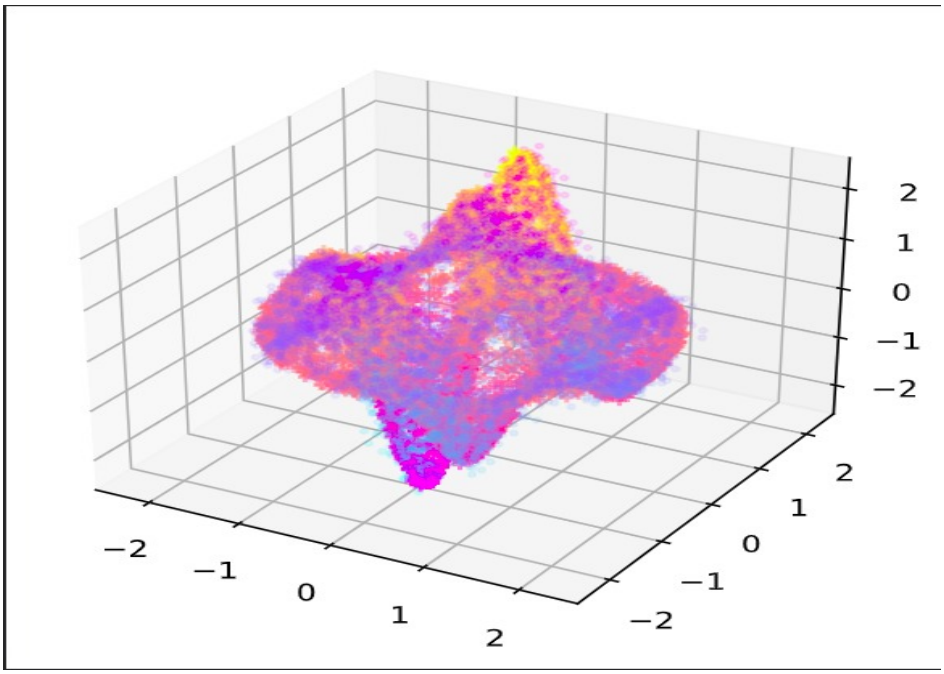

Figure 23: 100 steps

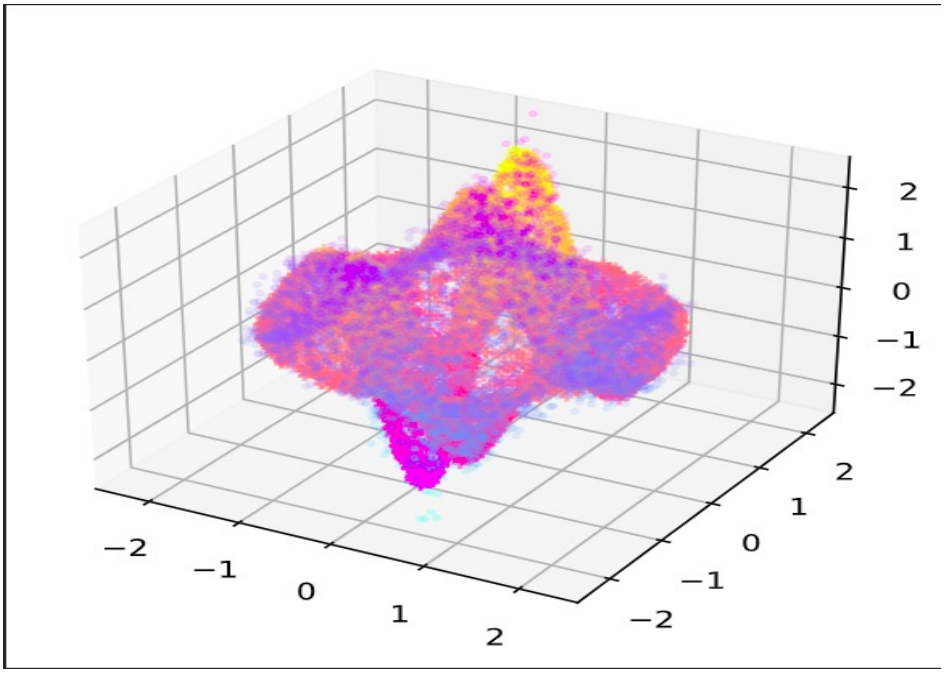

Figure 24: 150 steps

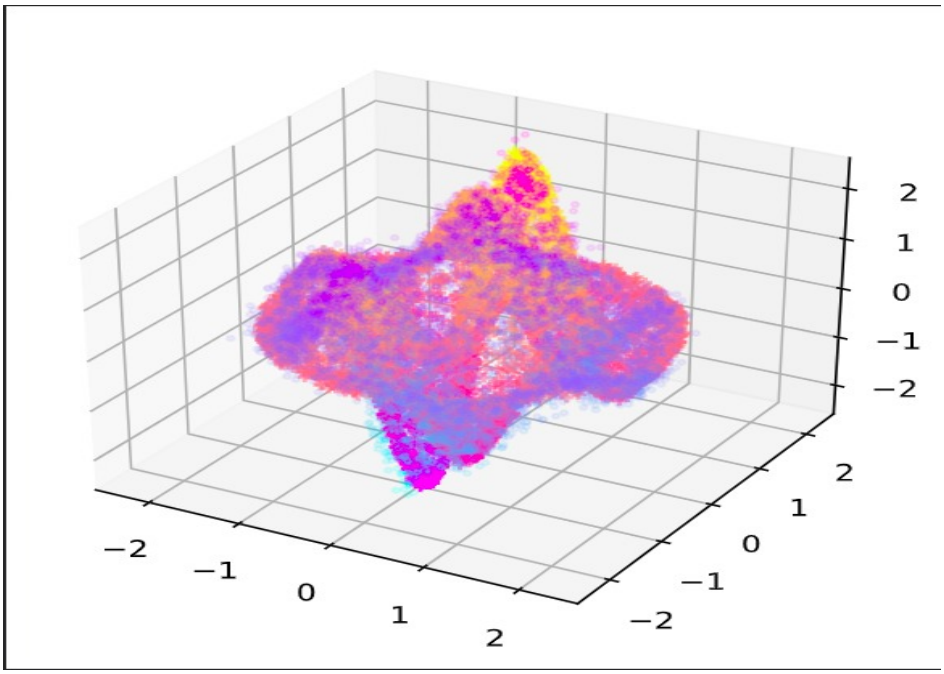

Figure 25: 200 steps

2.4 Noise scheduling

We used the same types of scheduling as mentioned in the helix section. We found out that linear schdeuling gives us best EMD. We obtained the EMD values for each scheduling method and plotted the results as follows :

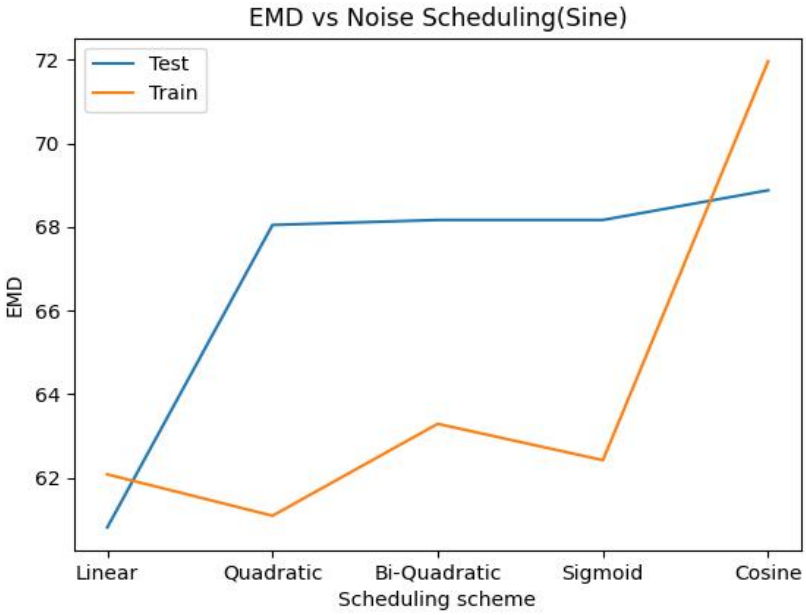

Figure 26: EMD values for noise scheduling schemes

3 Additional comments and comparisons

In all the above experiments, we could observe that the NLL values remain approximately the same for train and test sets and even across various experiments, the values don't differ by much. Hence, we shall focus our attention to EMD metric.

3.1 On optimal number of epochs for training

It makes sense that with the increase in the number of epochs, loss decreases and model performs better. But in our case, the loss initially decreased, but then after some number of epochs, the loss remained approximately constant. Hence, 1000 epochs performs the best. This holds for both sine and helix.

3.2 On model complexity

In both cases, the simplest model was giving the best results. This might probably due to the fact that we are dealing with low dimensional points here and not some large images where the number of parameters to be learned would be very large.

3.3 On the number of diffusion steps

We generally expect the performance to increase with the number of diffusion steps. We can see that this is mostly the case in both the datasets expect for the case with 100 diffusion steps. In both the datasets, we observe the EMD Is least when the number of diffusion steps is 100.

3.4 On noise schedules

We explored 5 different types of noise schedules. We observed that for sine, linear schedule was giving the best results, whereas for helix, quadratic schedule was giving the best results.