Exploring Denoising Diffusion Probabilistic Models

CS-726 Assignment - 2

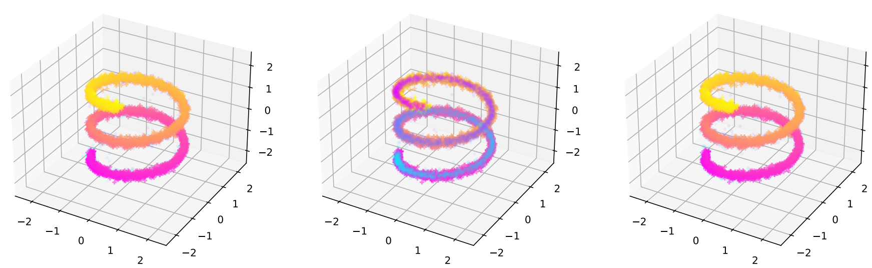
Jimut Bahan Pal (22D1594)
Sachin Kumar Giroh (22M2159)
IIT Bombay

February 18, 2023

1 Effect of number of training steps

We kept the following hyperparameters fixed while training, n_dim=3, n_steps=100, lbeta=1.000e-05, ubeta=1.280e-02, batch_size=1024 and changed the epochs to 500, 1000 and 2000 for training, saved the related models and evaluated on the same. We did this for the 3d_sin_5_5 dataset and the helix_3D dataset. The results for the helix_3D dataset is shown in Figure 1 and the results for the 3d_sin_5_5 dataset is shown in Figure 2. Here, in this report we sometime named the 3d_sin_5_5 dataset as sin 3D dataset or sin dataset, and the helix_3D dataset as the helix dataset respectively. The experiments are conducted by studying this paper [1].

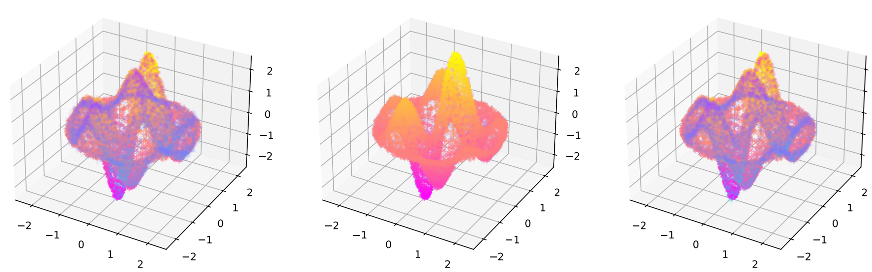
1.1 helix_3D dataset



(a) 500 epochs (b) 1000 epochs (c) 2000 epochs

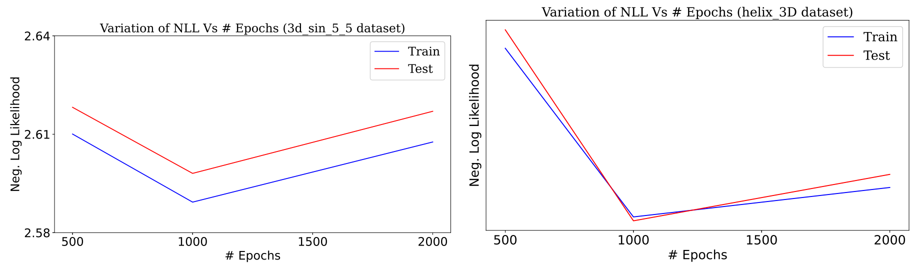
Figure 1: Vis overlay plot for different epochs on helix_3D dataset

1.2 3d_sin_5_5 dataset



(a) 500 epochs (b) 1000 epochs (c) 2000 epochs

Figure 2: Vis overlay plot for different epochs on 3d_sin_5_5 dataset



(a) 3d_sin_5_5 dataset (b) helix_3D dataset

Figure 3: Variation of Negative Log Likelihood Vs. Number of Epochs for both the datasets.

2 Effect of model complexity

We considered 3 models, of varying shape and sizes of the feed-forward layers. The architectures of the models are as follows:

First model: This comprise of linear layers of 5-32-64-64-3 with ReLU activation in between, except the last layer.

Second model: This comprise of linear layers of 5-64-128-256-64-3 with ReLU activation in between, except the last layer.

Third model: This comprise of linear layers of 5-16-32-64-32-16-3 with ReLU activation in between, except the last layer.

The second model performs the best in all the cases. The variation of negative log likelihood with respect to the model complexity is shown in Figure 4.

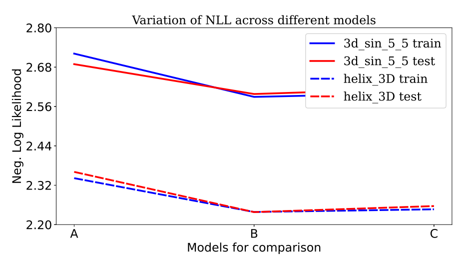
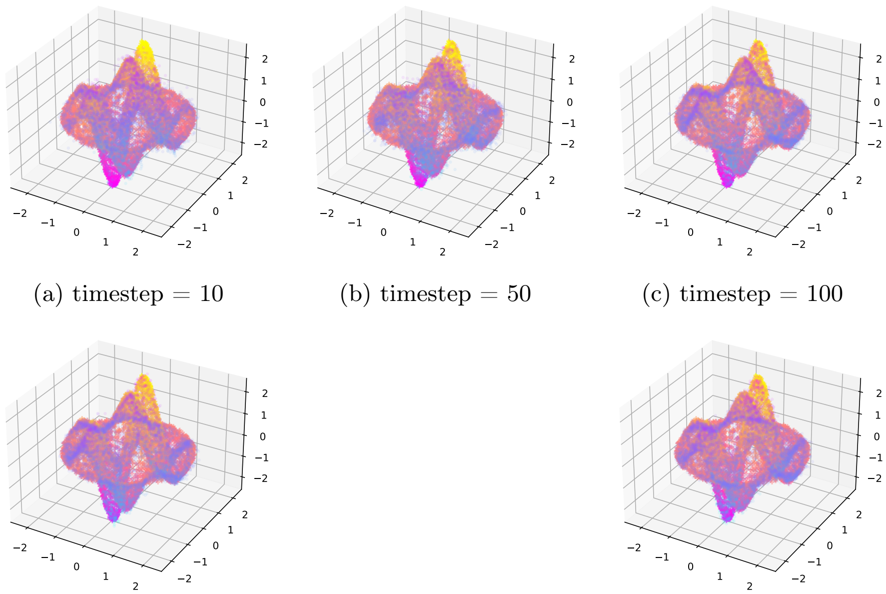


Figure 4: The variation of Negative Log Likelihood with respect to model complexity.

3 Effect of number of diffusion steps

We have considered 5 different timesteps i.e., 10, 50, 100, 150, 200 and 100 is giving the best results for both the datasets. We can see the visualizations generated by the sin dataset in Figure 5 and the same for helix dataset in Figure 6. The variation for both the datasets can be shown in Figure 7. This is because



(a) timestep = 10 (b) timestep = 50 (c) timestep = 100

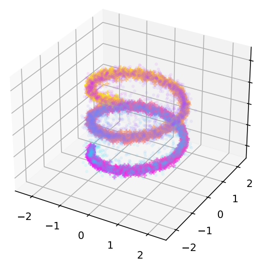
(d) timestep = 150 (e) timestep = 200

Figure 5: Vis overlay plot for different epochs on 3d_sin_5_5 dataset

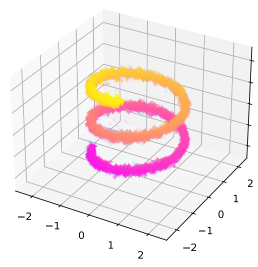
if we have a very less timestep then it adds noise at a rapid rate, and if we have a very high time-step it adds noise very slowly. The timestep of 100 performs best for all the datasets in the given model, henceforth, all experiments are done with a timestep of 100.

4 Effect of changing the noise schedule

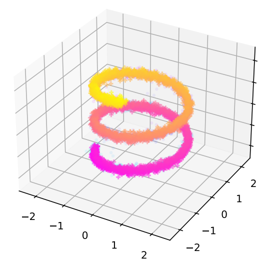
Here we have considered three types of schedulers, namely, linear, quadratic and sigmoidal. Also, we have considered different values of lbetas and ubetas. Three set of values considered are l_beta=1e-05,u_beta=1e-02, l_beta=1e-07,u_beta=1e-01, and l_beta=1e-07,u_beta=1e-03. In the case of sigmoid, as we decrease the value of lbeta and ubeta, the loss increases. Almost same effect is observed in case of quadratic scheduler. These experiments are conducted on the 3D sin dataset as shown in Figure 8 and the variation of plot is shown in Figure 9. We get the best result with the Quadratic scheduler and the u_beta and l_beta values are 1e-01 and 1e-07 respectively.



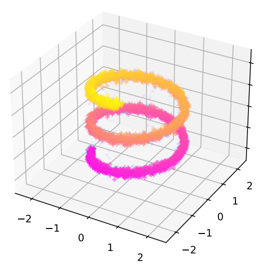
(a) timestep = 10



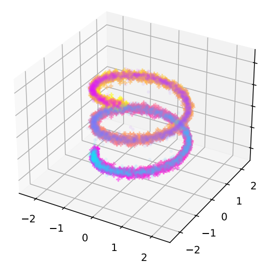
(b) timestep = 50



(c) timestep = 100

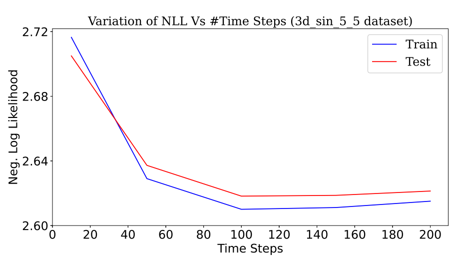


(d) timestep = 150

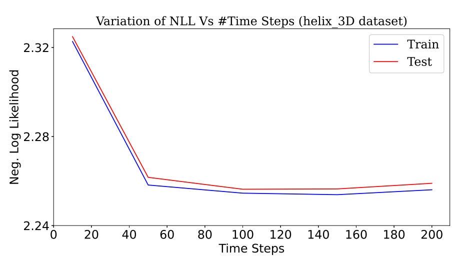


(e) timestep = 200

Figure 6: Vis overlay plot for different epochs on 3d_sin_5_5 dataset

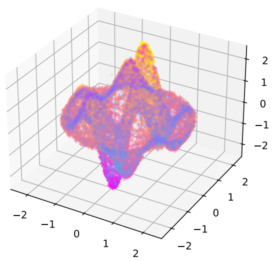


(a) Sin dataset

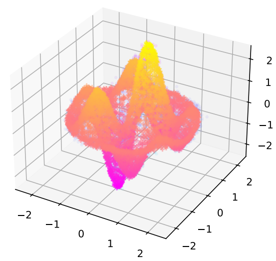


(b) Helix dataset

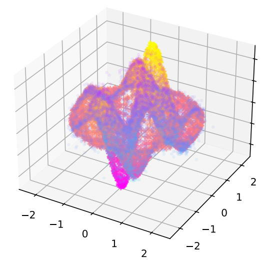
Figure 7: The variation for both the datasets in number of timesteps.



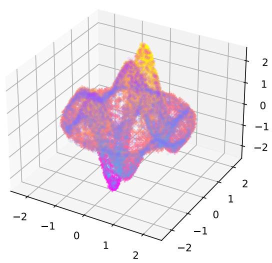
(a) Scheduler = Sigmoid l_beta=1e-05,u_beta=1e-02



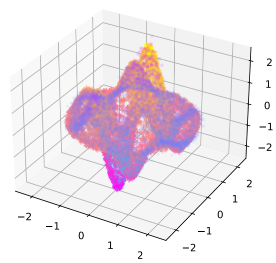
(b) Scheduler = Sigmoid l_beta=1e-07,u_beta=1e-01



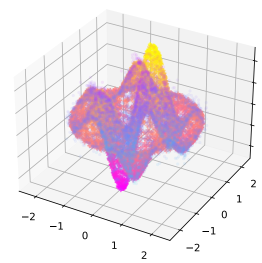
(c) Scheduler = Sigmoid l_beta=1e-07,u_beta=1e-03



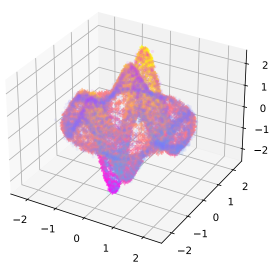
(d) Scheduler = Quad l_beta=1e-05,u_beta=1e-02



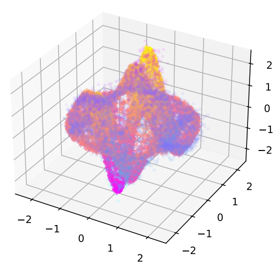
(e) Scheduler = Quad l_beta=1e-07,u_beta=1e-01



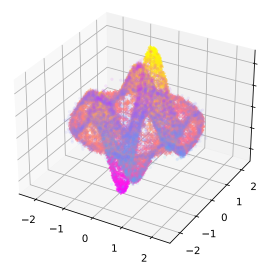
(f) Scheduler = Quad l_beta=1e-07,u_beta=1e-03



(g) Scheduler = Linear l_beta=1e-05,u_beta=1e-02



(h) Scheduler = Linear l_beta=1e-07,u_beta=1e-01



(i) Scheduler = Linear l_beta=1e-07,u_beta=1e-03

Figure 8: Variation of noise scheduler with different values of u_betas and l_betas.

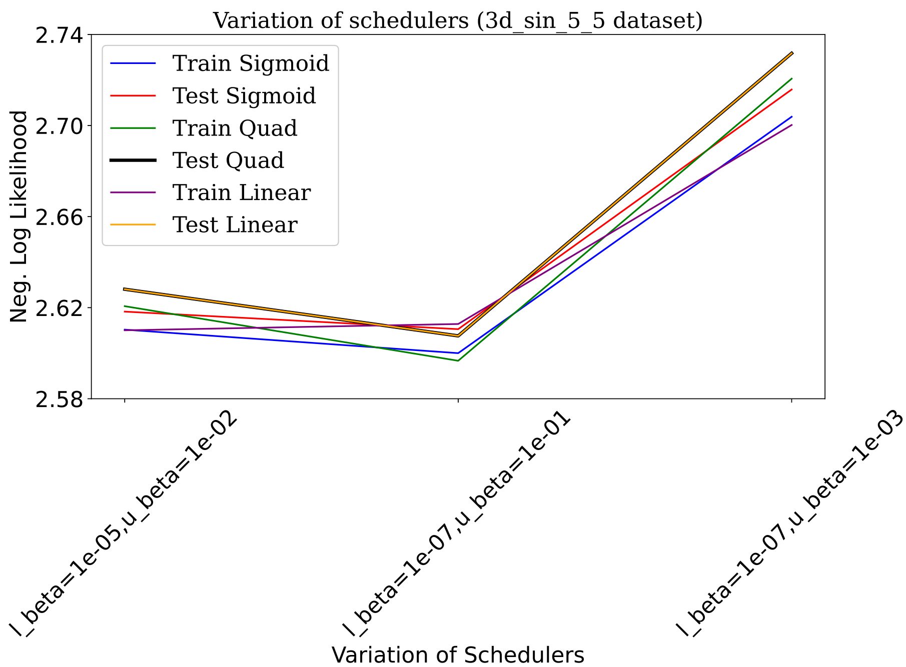


Figure 9: Variation of schedulers with different values of u_betas and l_betas.

Bibliography

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.