

APPENDIX FOR VISUAL ANALYSIS OF THE BUMPINESS AND RUGGEDNESS OF RESIDUAL NEURAL NETWORK LANDSCAPES

Anonymous authors

Paper under double-blind review

A APPENDIX

A.1 LOSS-GRADIENT CLOUDS WITH BUMPINESS

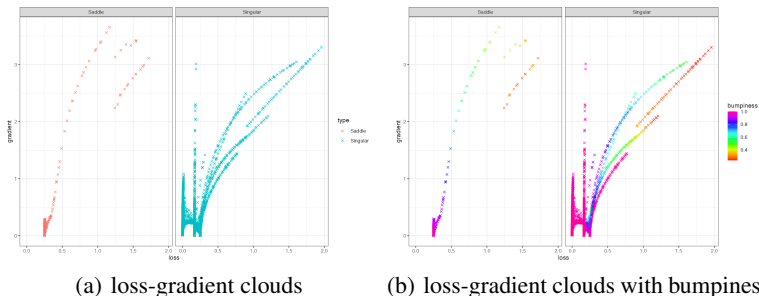


Figure 1: Figure (a) is the loss-gradient cloud. Figure (b) shows that as the optimization algorithm progresses, the curvature of the landscapes, which are predominantly singular, approaches convexity.

A.2 CONVEX, CONCAVE, SADDLE, SINGULAR POINTS IN THE XOR, IRIS, DIABETES DATASETS

Table 1: Statistics of different types of sampling points

Datasets	Number of hidden layers	Residual	Number of convex points	Number of concave points	Number of saddle points	Number of singular points
XOR	2	with	0	0	103	9897
		without	0	0	1014	8986
	3	with	0	0	0	10000
		without	0	0	0	10000
	4	with	0	0	0	10000
		without	0	0	0	10000
Iris	2	with	0	0	10	29990
		without	0	0	957	29043
	3	with	0	0	0	30000
		without	0	0	0	30000
	4	with	0	0	0	30000
		without	0	0	0	30000
Diabetes	2	with	0	0	0	10000
		without	0	0	0	10000
	3	with	0	0	0	10000
		without	0	0	0	10000
	4	with	0	0	0	10000
		without	0	0	0	10000

A.3 EXAMPLES OF HESSIAN CLOUDS

Below is a comprehensive illustration of the Hessian clouds of the MLPs with two hidden layers utilized in our experiments. To illustrate the trajectory of the sampling, we change the scatters' color to the corresponding step.

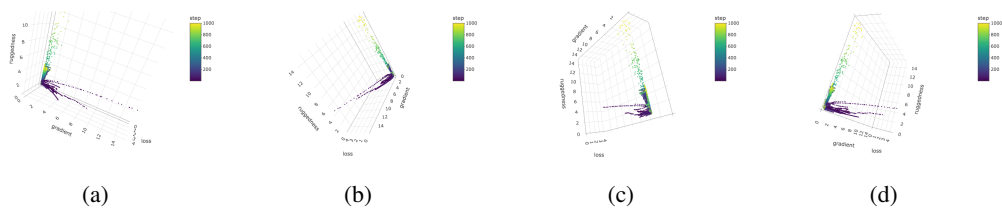


Figure 2: Hessian clouds, XOR, 2 hidden layers

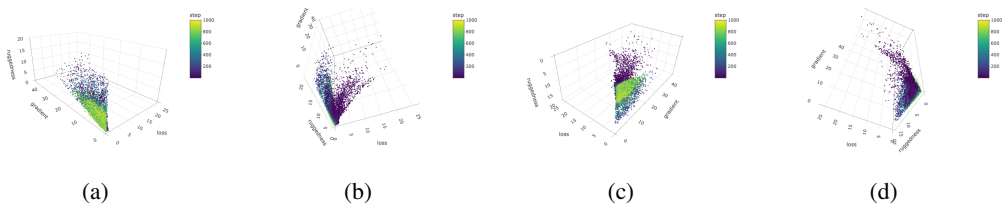


Figure 3: Hessian clouds, Iris, 2 hidden layers

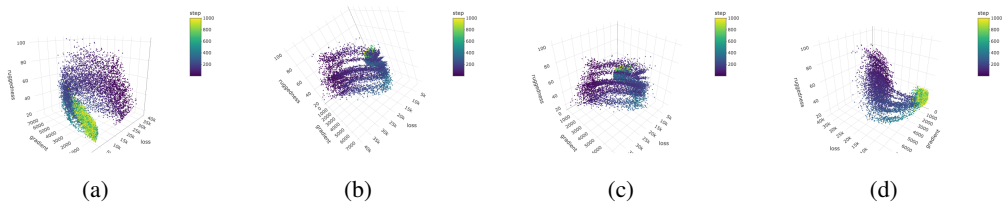


Figure 4: Hessian clouds, Diabetes, 2 hidden layers

A.4 V-SHAPED OSCILLATION AREAS IN THE HESSIAN CLOUDS OF THE XOR DATASET

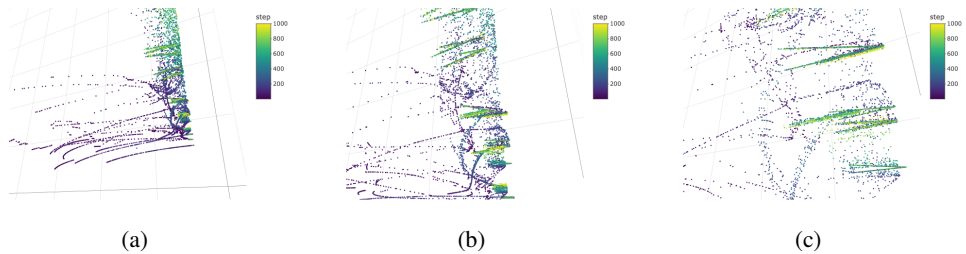


Figure 5: V-shaped oscillation areas, XOR, 2 hidden layers

A.5 ANOTHER BASIN IN NON-RESIDUAL NN LANDSCAPES OF THE IRIS DATASET

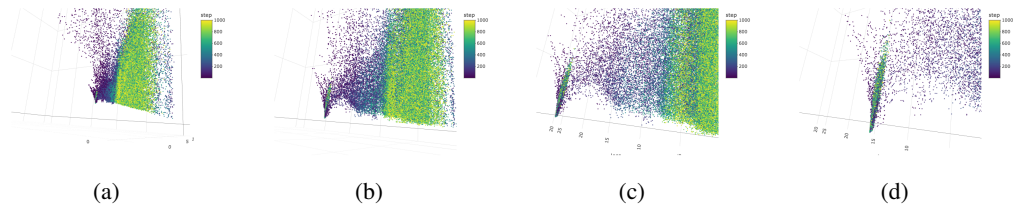


Figure 6: another basin, Iris, 2 hidden layers