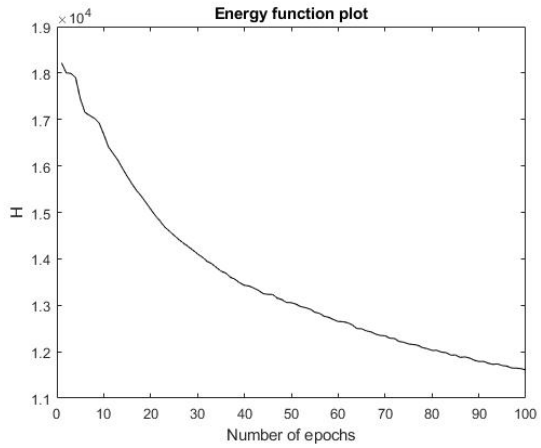

Figure 1: The plot shows H evaluated for all training data as a function of epoch where H is the energy function.

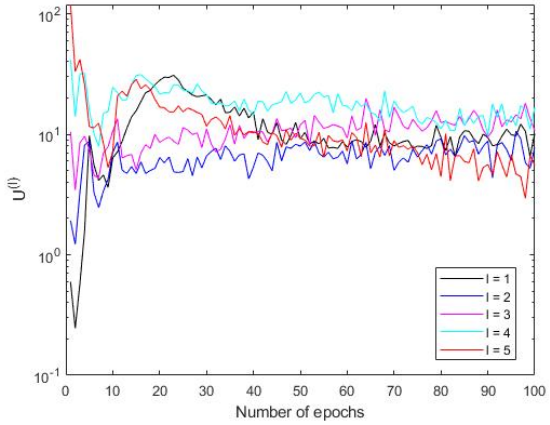

Figure 2: The plot shows $U^{(l)}$ as a function of epoch for l=1 to l=5 where $U^{(l)}$ is the learning speed of layer l.

1 Results and Discussion

The results and discussion are the following:

1. The learning speed shows a lognormal distribution with time, and it is different in the different layers.
2. In the initial phase of training that is, phase I, we observe the learning slowdown due to severity of the vanishing gradient problem.