

CSE849 HW1

Javen W. Zamojcin
 zamojci1@msu.edu

2025, February 21

1 Question 1

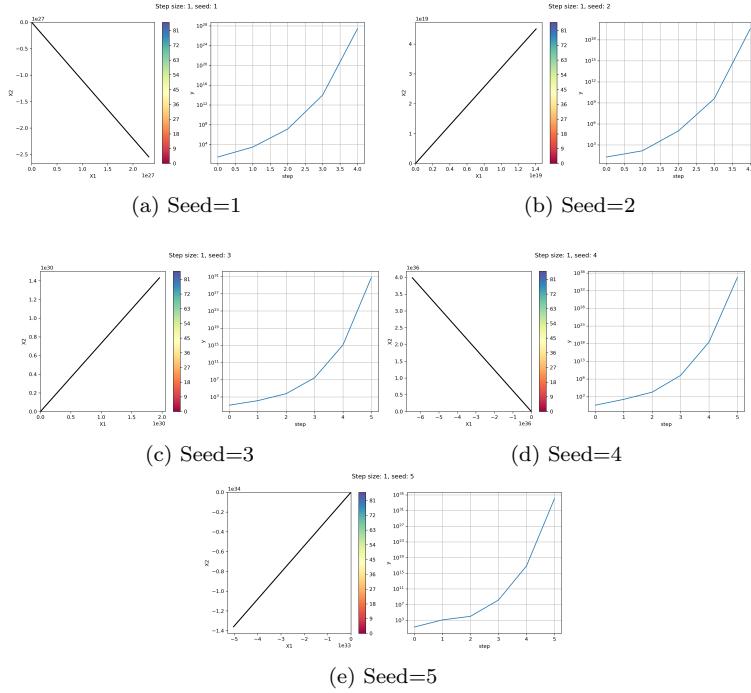


Figure 1: Q1 Trajectories $\lambda = 1$

Figure 1 displays the trajectory plots for a learning rate parameter of 1.0. All of these plots not only failed to converge anywhere near the global optima, but ended in divergent behavior with exponentially increasing loss values. This learning rate is far too large for any productive training,

Figure 2 displays the trajectory plots for a learning rate parameter of 0.1. Of all the learning rates, these plots display the most erratic behavior, often

making large leaps that overshoot the global optima. This suggests the learning rate is too large for this problem.

Figure 3 displays the trajectory plots for a learning rate parameter of 0.01. Of all the learning rates, these plots consistently display the most stable behavior, steadily approaching the global optima in relatively few steps.

Figure 4 displays the trajectory plots for a learning rate parameter of 0.001. These plots also display very stable behavior and convergence but in a slower manner, taking far more steps than the 0.01 plots.

Figure 5 displays the trajectory plots for a learning rate parameter of 0.0001. All of these plots failed to converge anywhere near the global optima, even after 2000 steps each. Their behavior isn't erratic, but are just too slow. This learning rate definitely is too small.

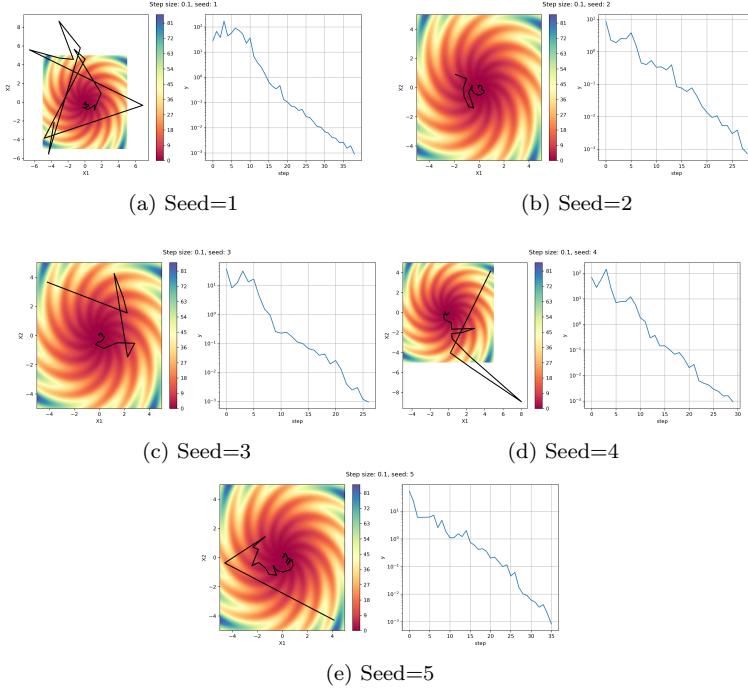


Figure 2: Q1 Trajectories $\lambda = 0.1$

2 Question 2

See Figure 6 for the training vs. validation loss plot of question 2.

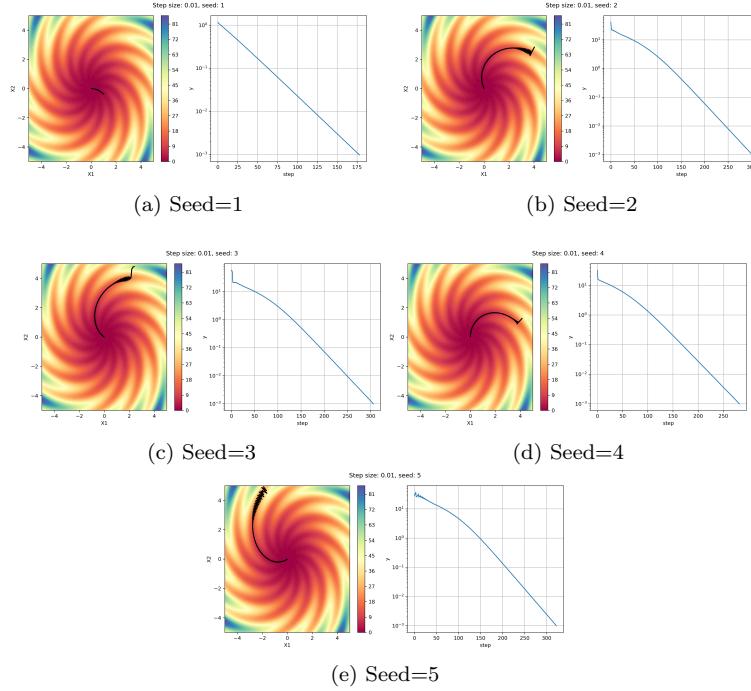


Figure 3: Q1 Trajectories $\lambda = 0.01$

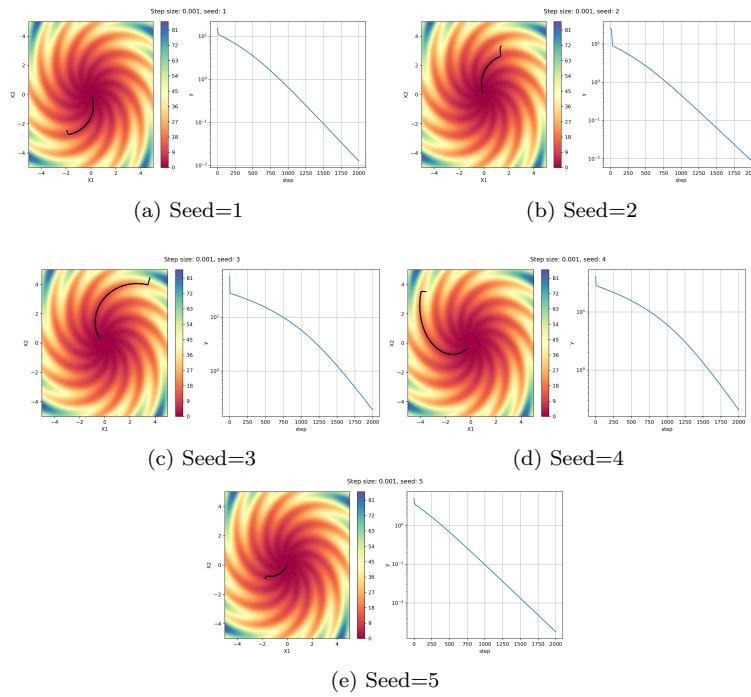


Figure 4: Q1 Trajectories $\lambda = 0.001$

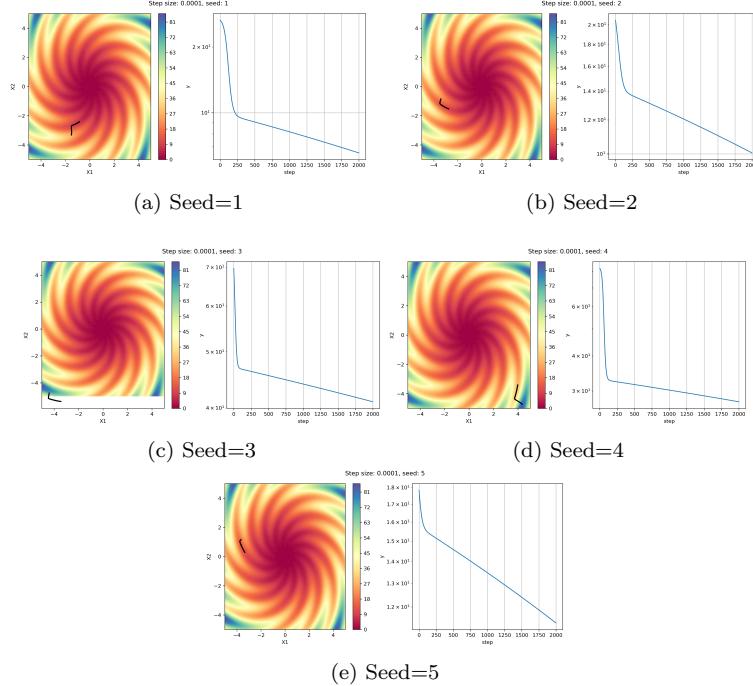


Figure 5: Q1 Trajectories $\lambda = 0.0001$

Q2 Training vs. Validation Loss

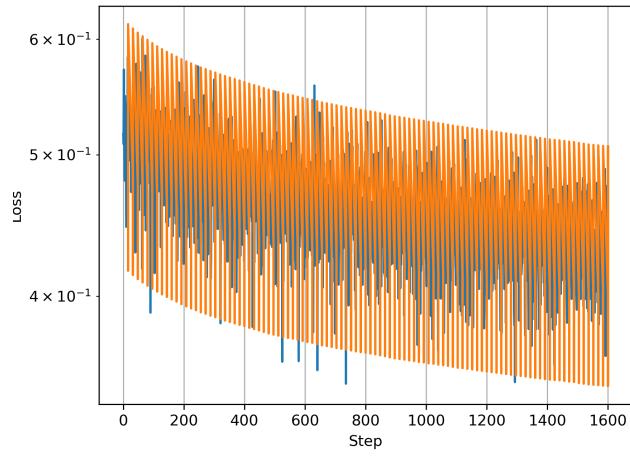


Figure 6: Q2 Training vs. Validation Loss