CS 237B: Principles of Robot Autonomy II
Problem Set X

Name:
SUID:

Problem 1

(i) Initializing all weights as zero may result in finding local optima, as opposed to using randomized weight selection. Conversely, the results may be more reproducible if using zero weights.

(ii) Xavier initialization attempts to keep the variance across the weights of each layer to be the same. The purpose is to prevent what is termed "exploding or vanishing gradients". Exploding gradients can be understood as when weights are initialized too large, the output of each layer increases exponentially. This leads to poor accuracy of the network, as the target oscillates around a minima (the cost oscillating around the minimum value). Similarly, when the weights are initialized to small, the network may converge too quickly, as the cost gets exponentially lower ("vanishing").

Problem 2

(i) Using the associative property of matrix multiplication (for rearranging the torque m

Problem 3

(i) See code.