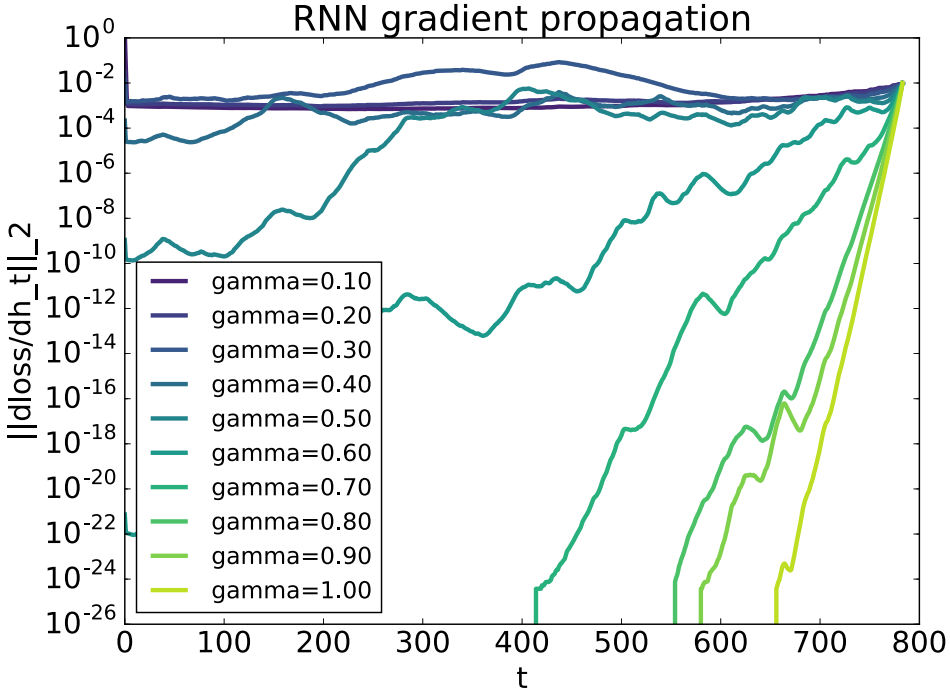
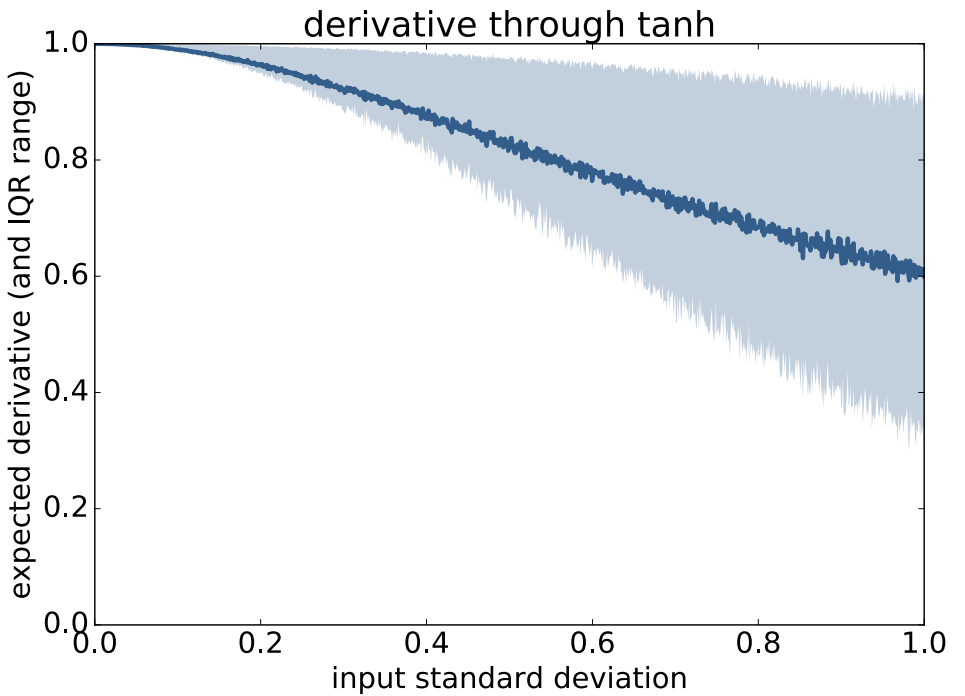


(a) We visualize the gradient flow through a batch-normalized tanh RNN as a function of γ. High variance causes vanishing gradient.



(b) We show the empirical expected derivative and interquartile range of tanh nonlinearity as a function of input variance. High variance causes saturation, which decreases the expected derivative.