MNIST - Gradient Norms vs. Loss Loss  $||g||_{1}$  $\|g_{sam}\|_1$ 0.09  $||g||_2$  $\|g_{sam}\|_2$ 10<sup>0</sup> 0.08 **Gradient Norm** 0.07 0.06  $10^{-1}$ 0.05 0.04 80 20 40 60 **Iterations**