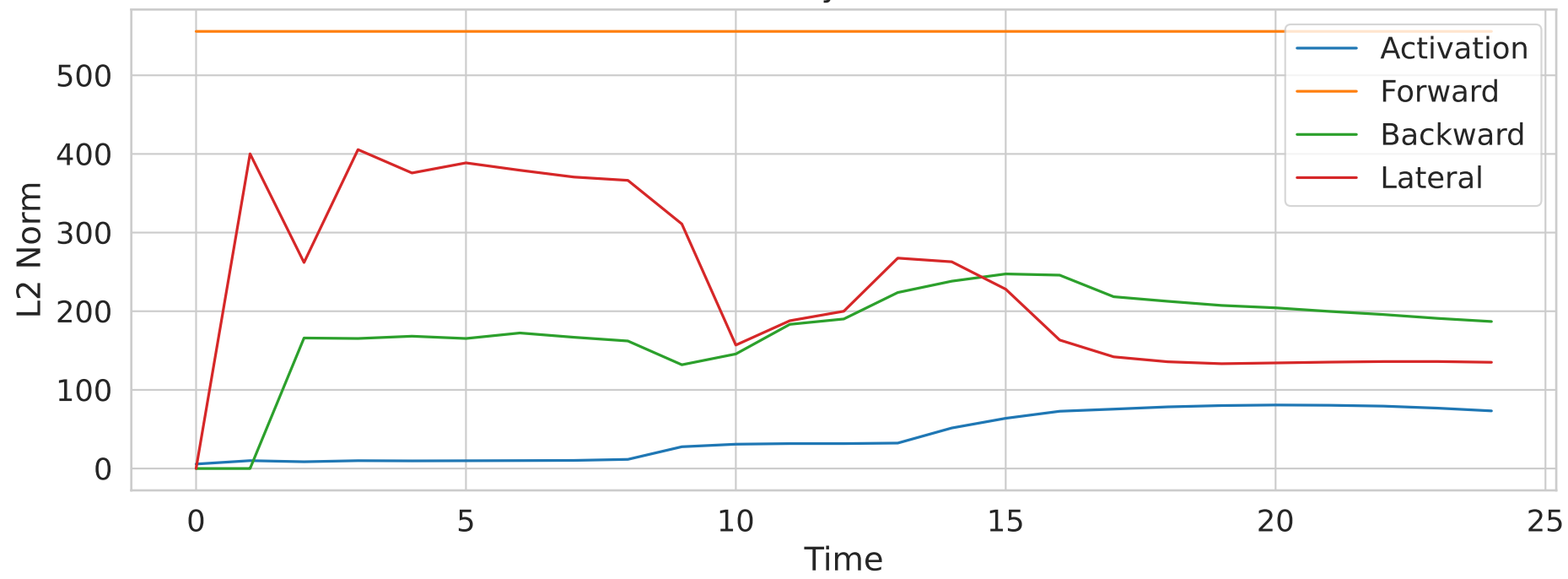
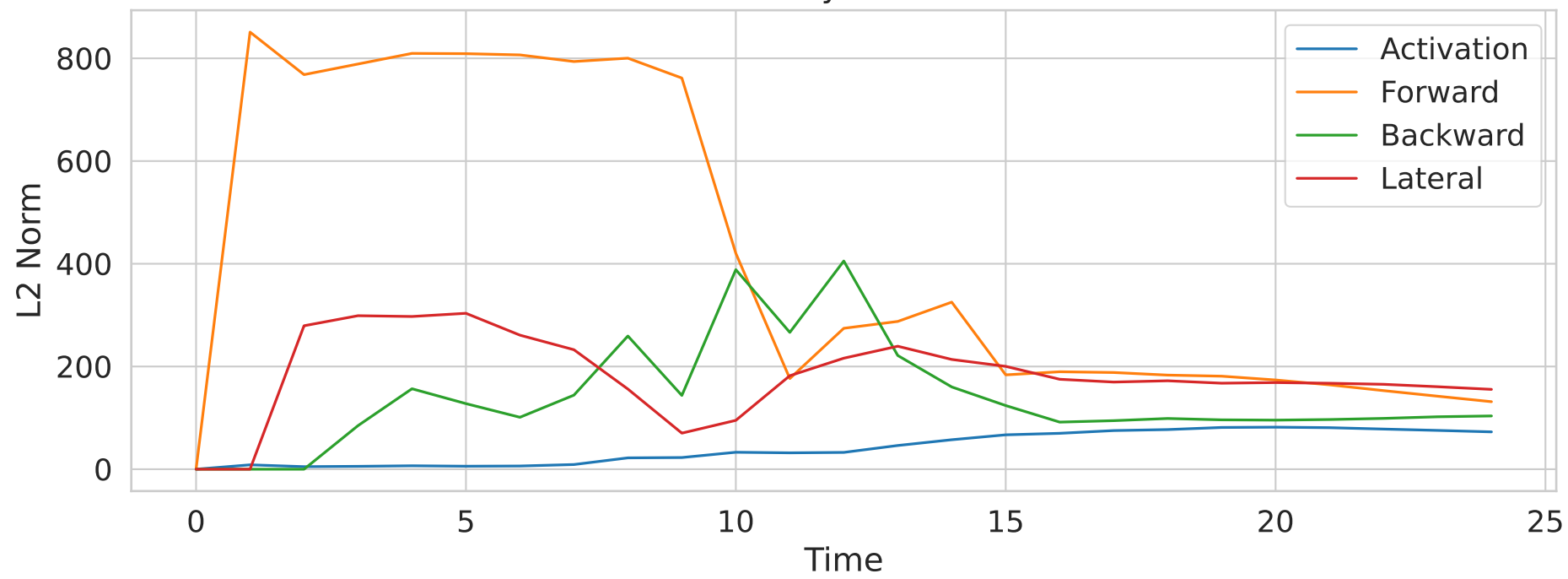


L2 Norm of Activations Over Time (Negative Data)

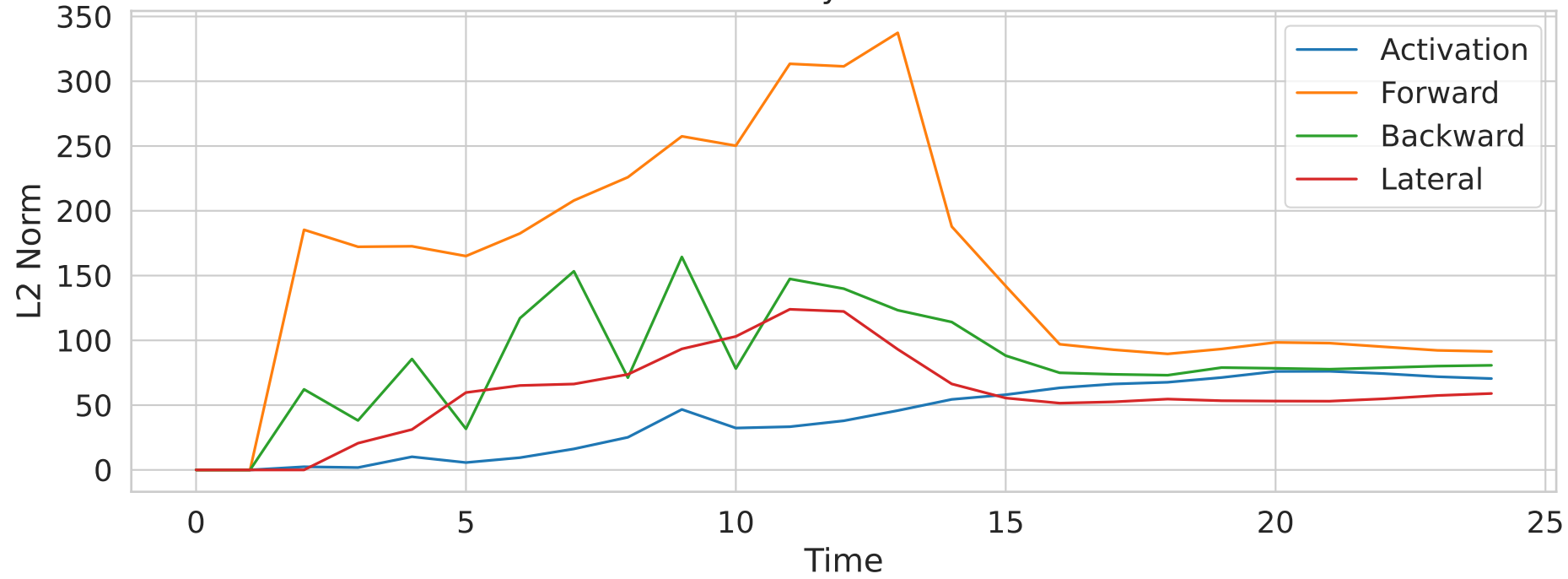
Layer 1



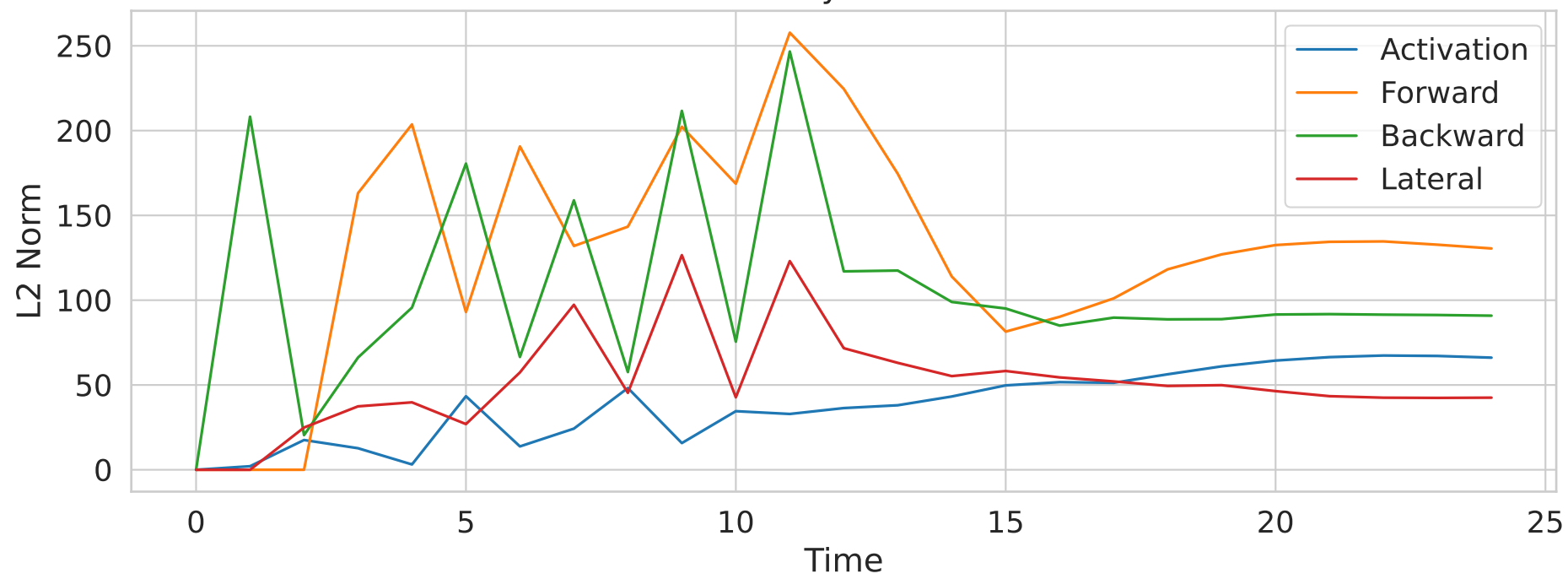
Layer 2



Layer 3



Layer 4



Layer 5

