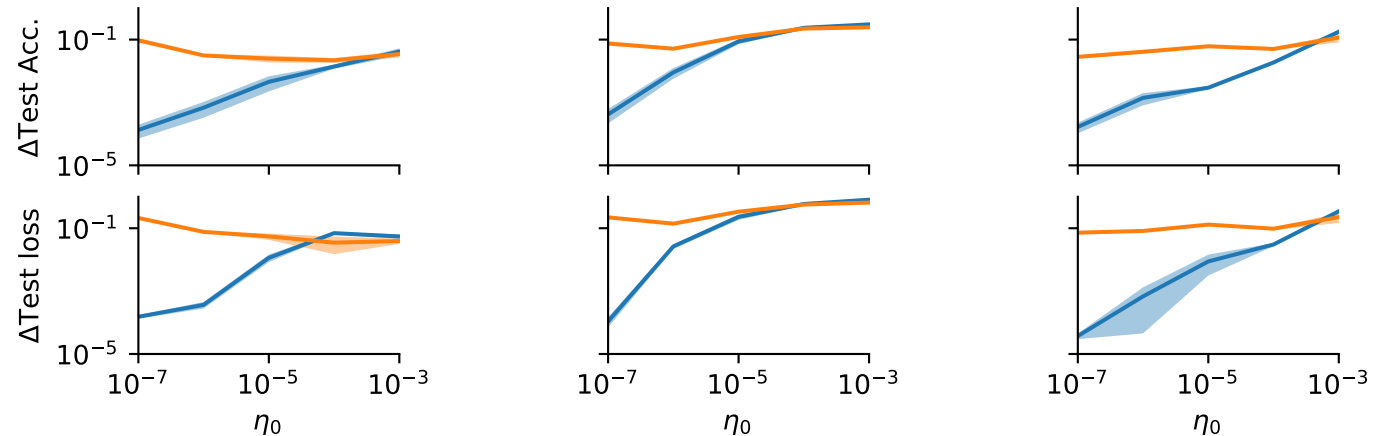


ResNet w/ BatchNorm

ResNet w/ LayerNorm

ViT



Batch size invariant Adam      Standard Adam w/ square-root scaling