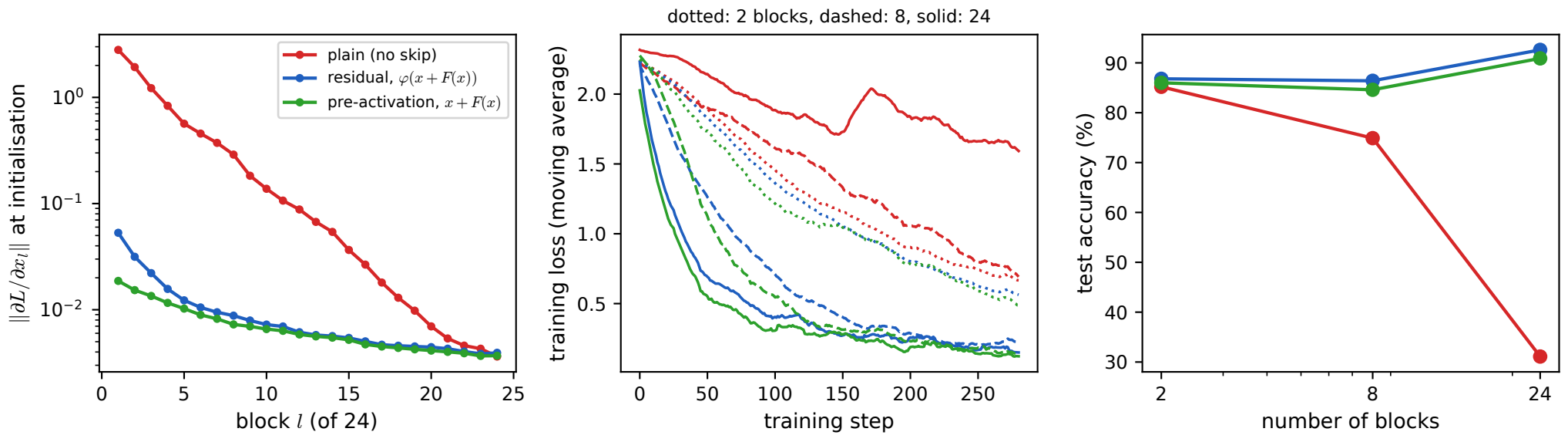

plain (no skip)
residual, $\varphi(x+F(x))$
pre-activation, $x+F(x)$
$\|\partial L/\partial x_l\|$ at initialisation
block l (of 24)
dotted: 2 blocks, dashed: 8, solid: 24
training loss (moving average)
training step
test accuracy (%)
number of blocks