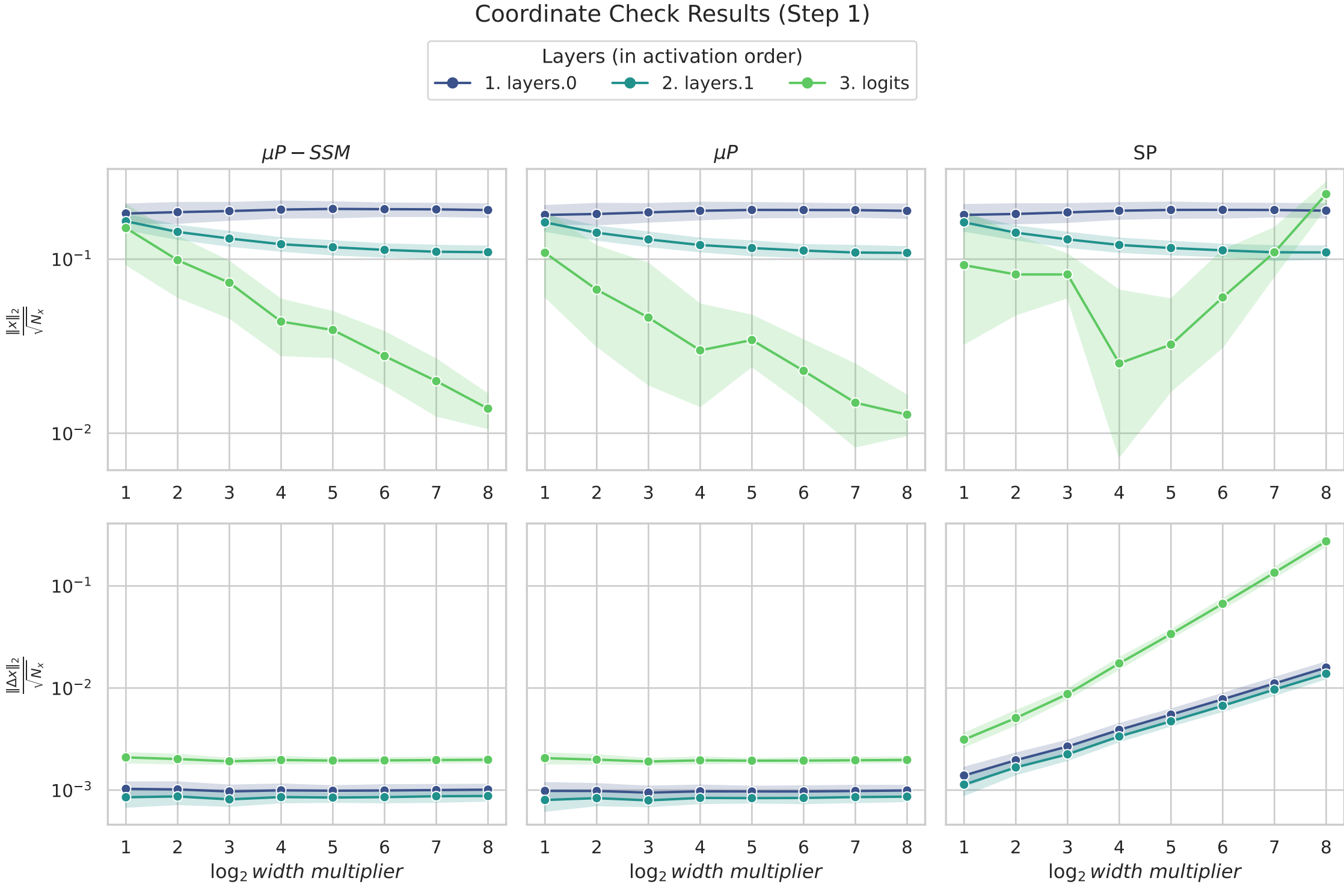
Coordinate Check Results (Step 1)
Layers (in activation order)
1. layers.0
2. layers.1
3. logits
μP − SSM
μP
SP
$\frac{\|x\|_2}{\sqrt{N_x}}$
$\frac{\|\Delta x\|_2}{\sqrt{N_x}}$
10^{-1}
10^{-2}
10^{-3}
1 2 3 4 5 6 7 8
$\log_2$ width multiplier