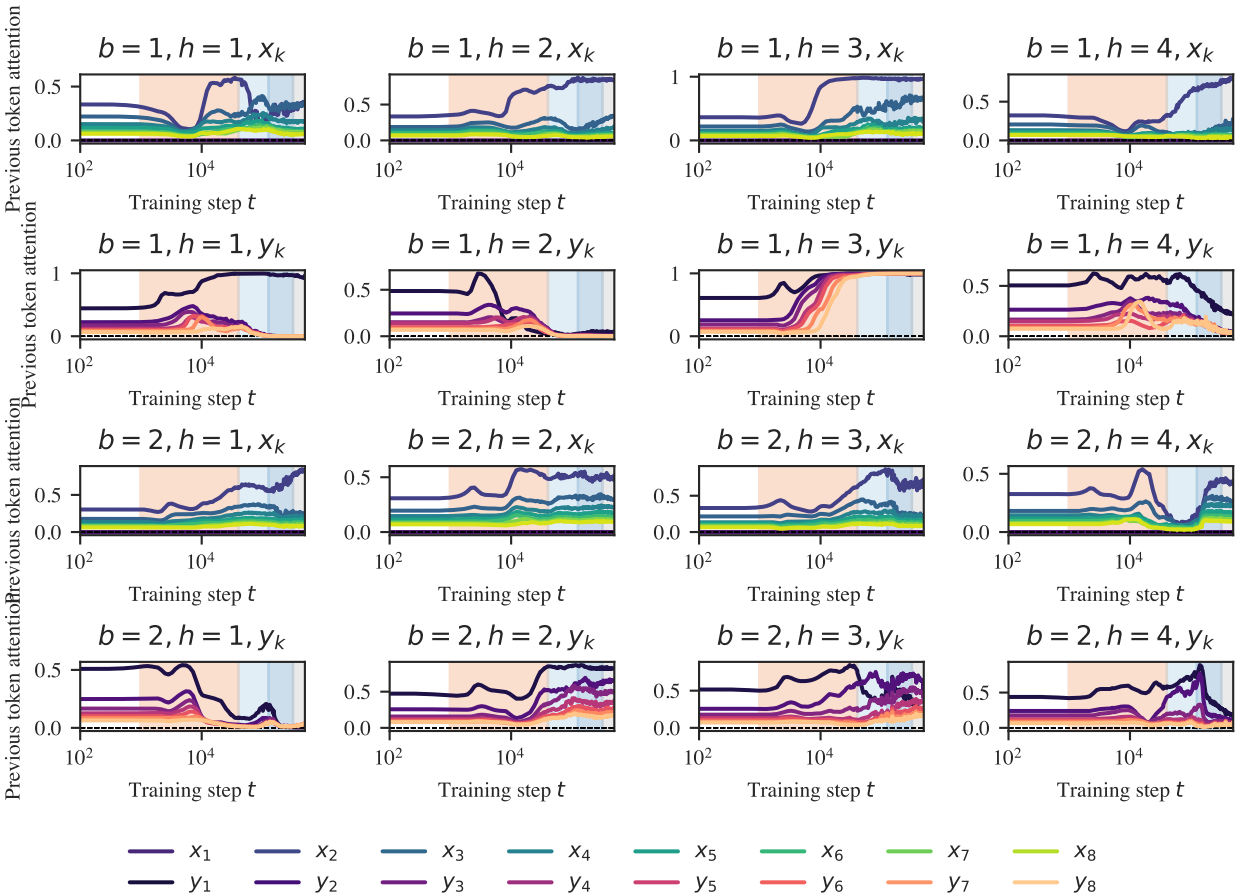
$b = 1, h = 1, x_k$
$b = 1, h = 2, x_k$
$b = 1, h = 3, x_k$
$b = 1, h = 4, x_k$
$b = 1, h = 1, y_k$
$b = 1, h = 2, y_k$
$b = 1, h = 3, y_k$
$b = 1, h = 4, y_k$
$b = 2, h = 1, x_k$
$b = 2, h = 2, x_k$
$b = 2, h = 3, x_k$
$b = 2, h = 4, x_k$
$b = 2, h = 1, y_k$
$b = 2, h = 2, y_k$
$b = 2, h = 3, y_k$
$b = 2, h = 4, y_k$
Previous token attention
Training step t
10^2
10^4
0.0
0.5
0
1
x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8
y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8