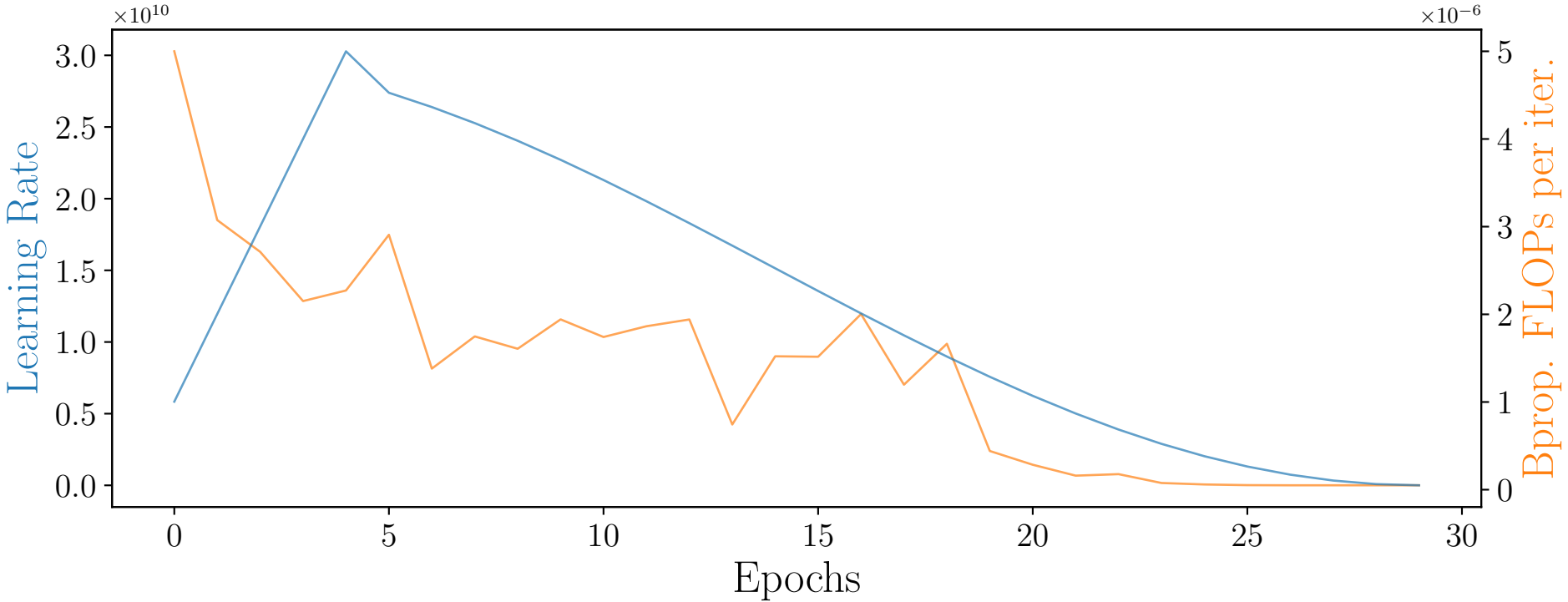
×10^10
×10^−6
Learning Rate
Bprop. FLOPs per iter.
Epochs
3.0
2.5
2.0
1.5
1.0
0.5
0.0
5
4
3
2
1
0
0
5
10
15
20
25
30