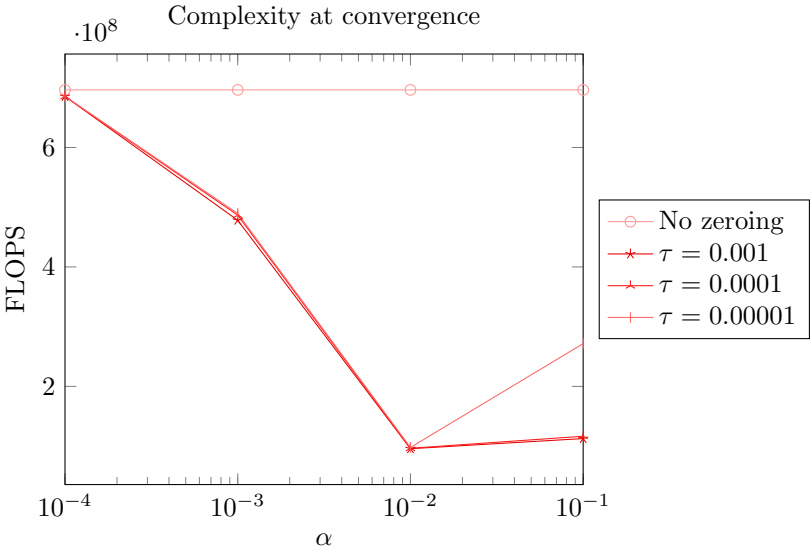
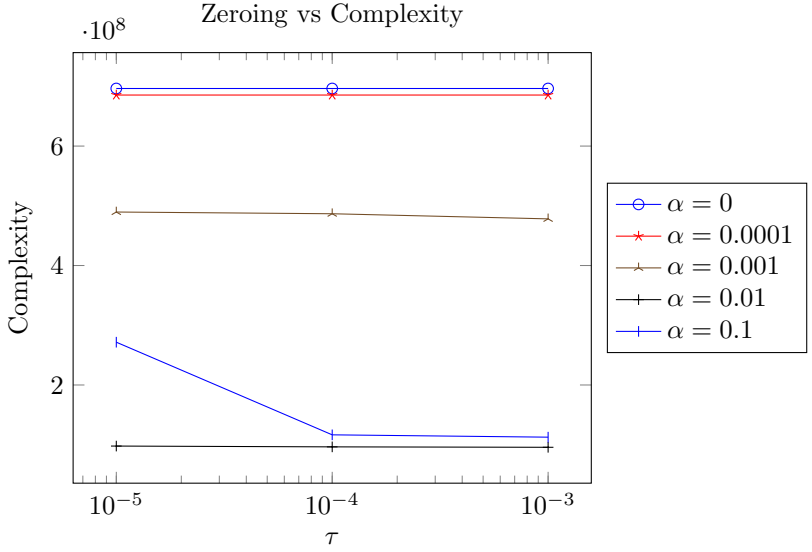
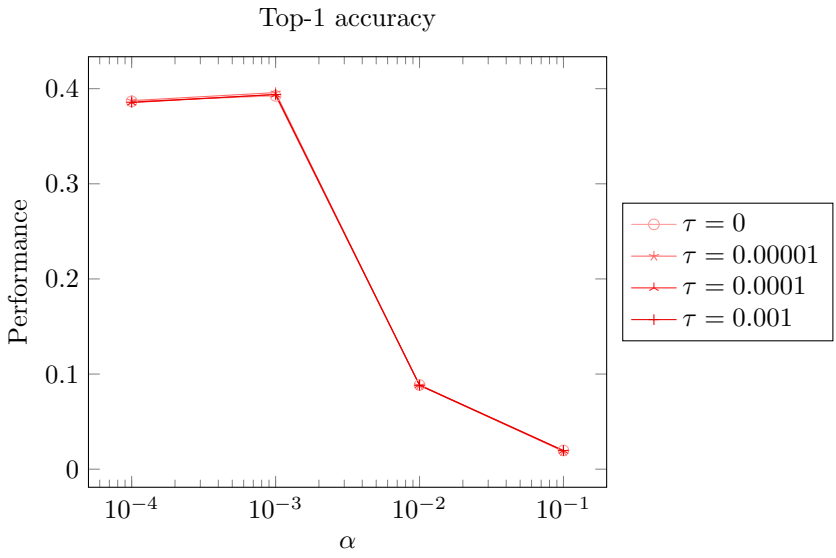
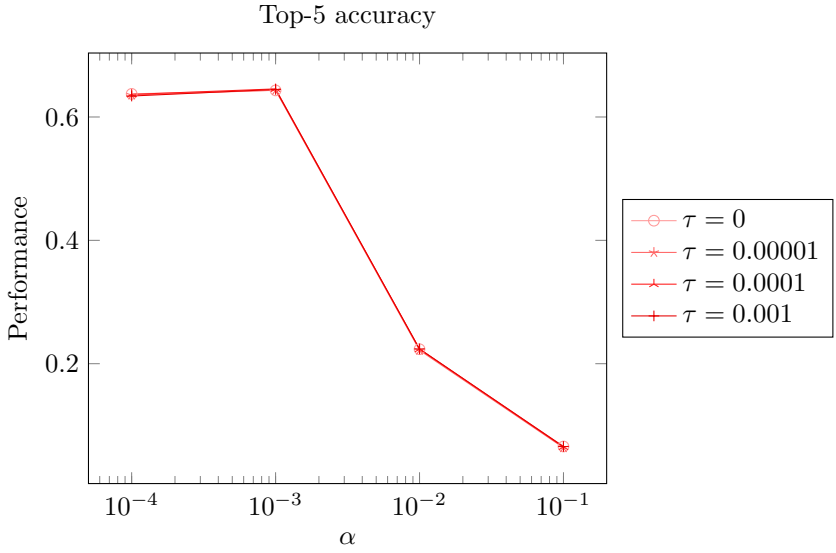
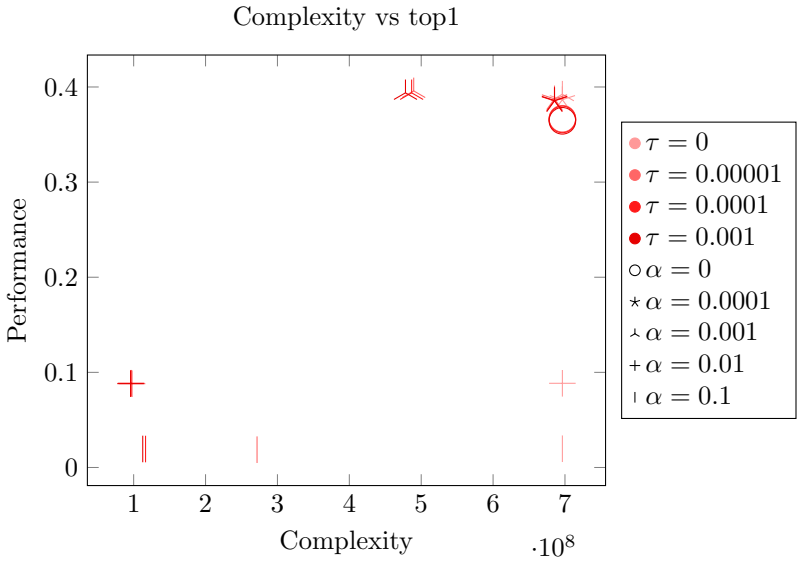
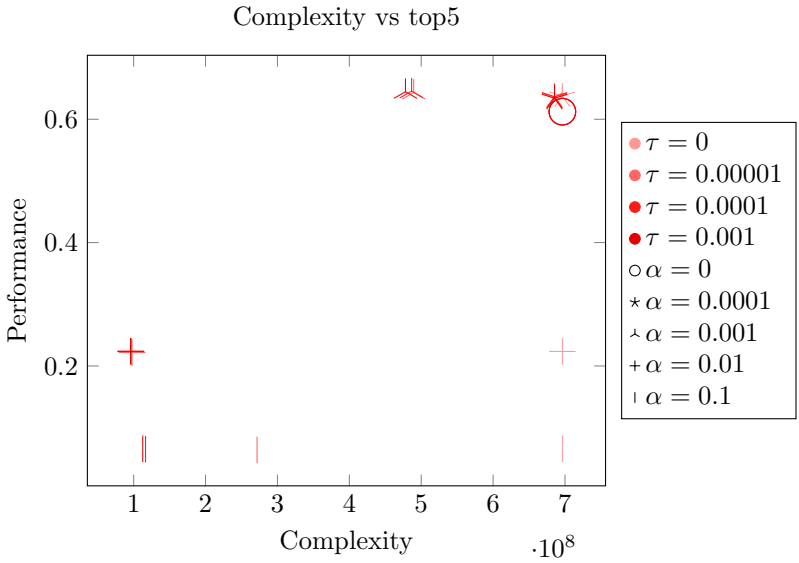

Complexity at convergence
$\cdot 10^8$
FLOPS
6
4
2
10^{-4}
10^{-3}
10^{-2}
10^{-1}
α
No zeroing
$\tau = 0.001$
$\tau = 0.0001$
$\tau = 0.00001$


Zeroing vs Complexity
$\cdot 10^8$
Complexity
6
4
2
10^{-5}
10^{-4}
10^{-3}
τ
$\alpha = 0$
$\alpha = 0.0001$
$\alpha = 0.001$
$\alpha = 0.01$
$\alpha = 0.1$

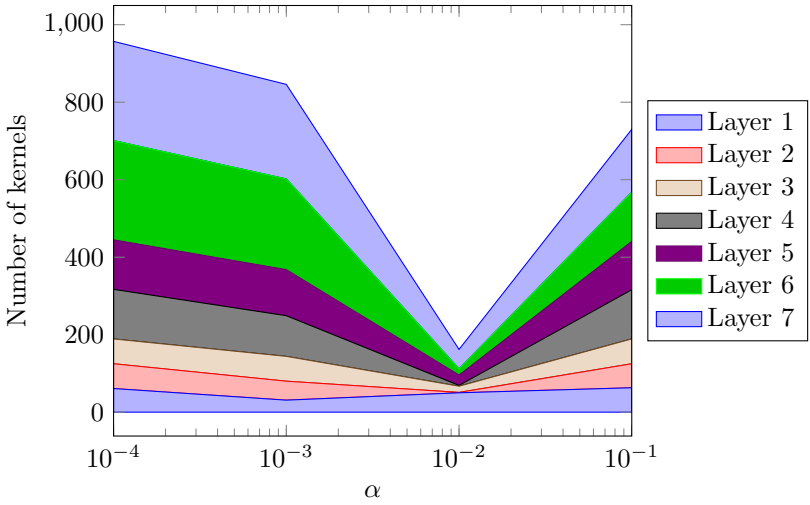

Top-1 accuracy
Performance
0.4
0.3
0.2
0.1
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
α
$\tau = 0$
$\tau = 0.00001$
$\tau = 0.0001$
$\tau = 0.001$


Top-5 accuracy
Performance
0.6
0.4
0.2
10^{-4}
10^{-3}
10^{-2}
10^{-1}
α
$\tau = 0$
$\tau = 0.00001$
$\tau = 0.0001$
$\tau = 0.001$

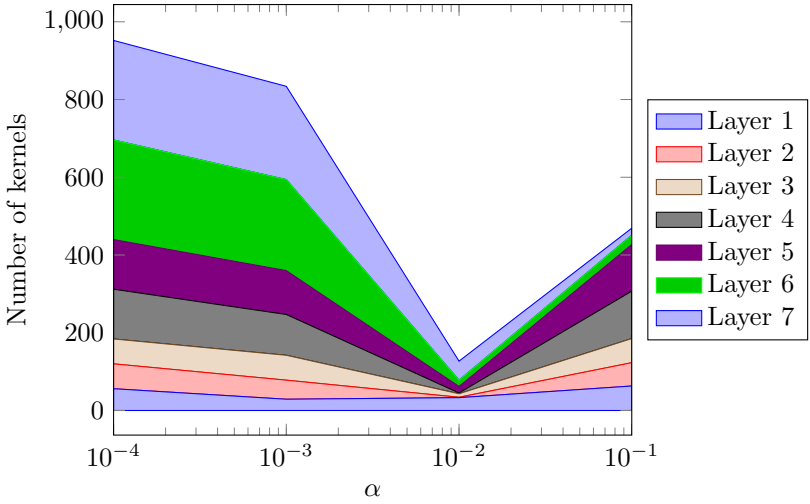

Complexity vs top1
Performance
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
Complexity
$\cdot 10^8$
$\tau = 0$
$\tau = 0.00001$
$\tau = 0.0001$
$\tau = 0.001$
$\alpha = 0$
$\alpha = 0.0001$
$\alpha = 0.001$
$\alpha = 0.01$
$\alpha = 0.1$


Complexity vs top5
Performance
0.6
0.4
0.2
1
2
3
4
5
6
7
Complexity
$\cdot 10^8$
$\tau = 0$
$\tau = 0.00001$
$\tau = 0.0001$
$\tau = 0.001$
$\alpha = 0$
$\alpha = 0.0001$
$\alpha = 0.001$
$\alpha = 0.01$
$\alpha = 0.1$

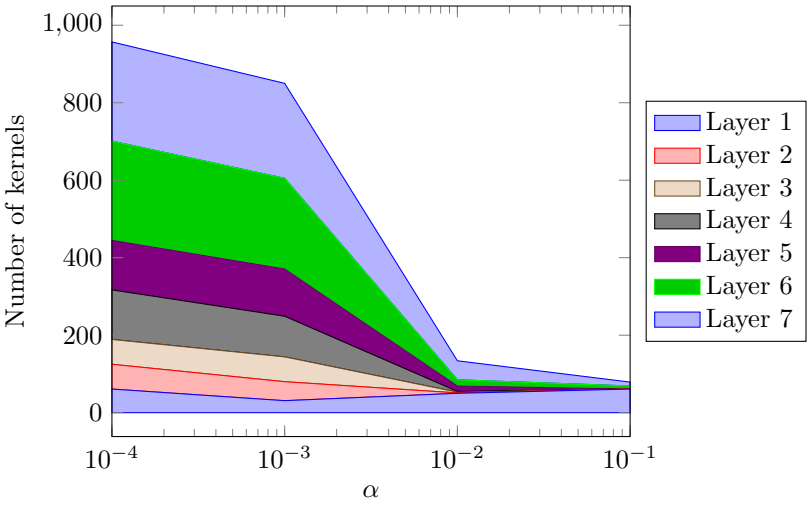
Number of kernels at convergence - L2 losses $\leq 1\%$ mean L2 loss



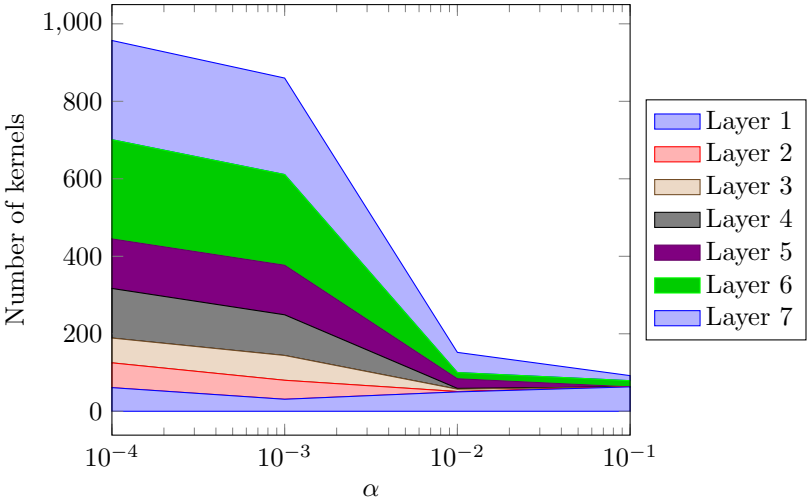
Number of kernels at convergence - L2 losses $\leq 5\%$ mean L2 loss

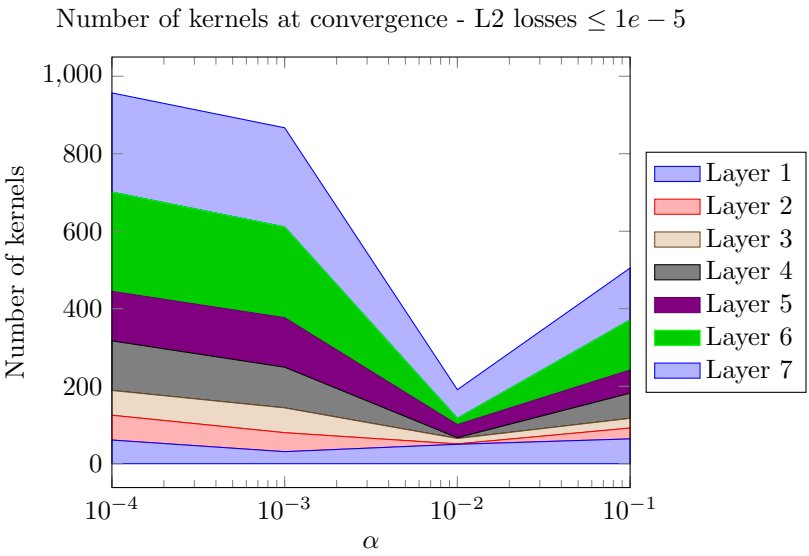


Number of kernels at convergence - L2 losses $\leq 1e-3$

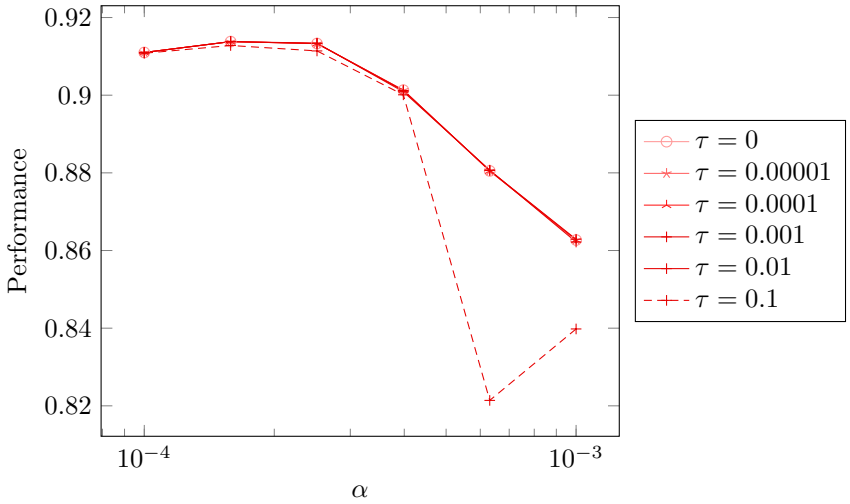


Number of kernels at convergence - L2 losses $\leq 1e-4$

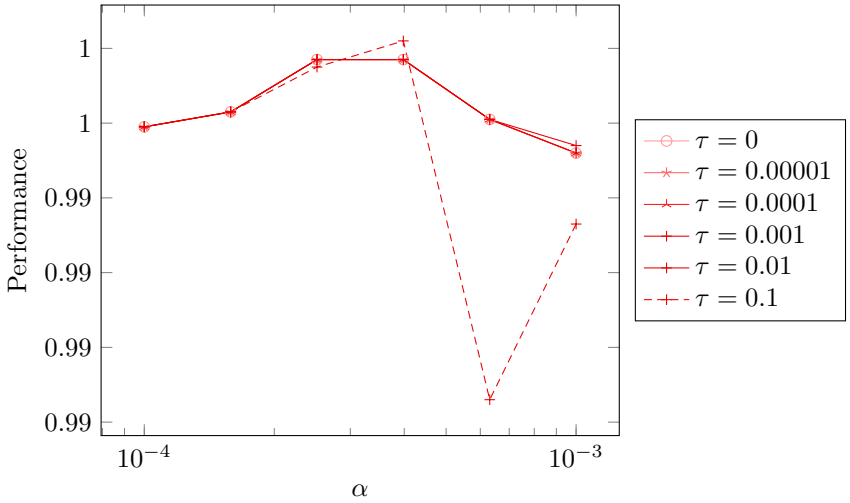


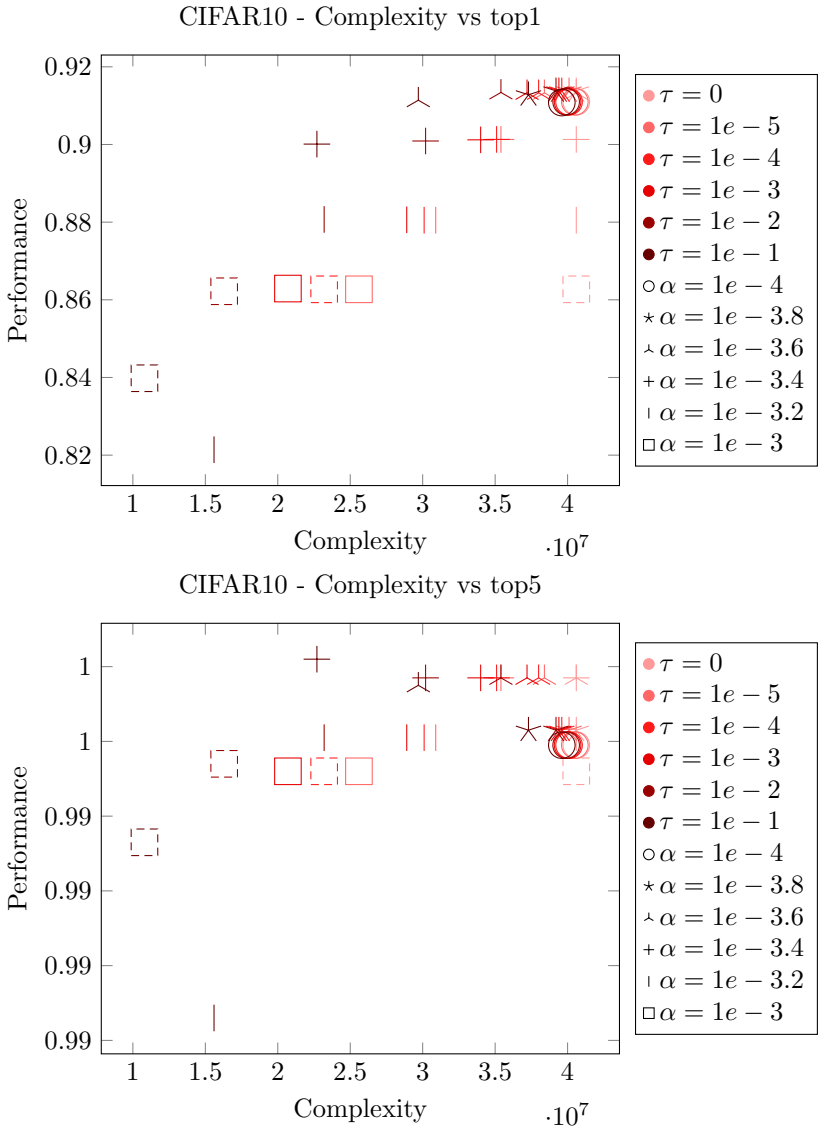

Number of kernels at convergence - L2 losses $\leq 1e-5$
Number of kernels
1,000
800
600
400
200
0
10^{-4}
10^{-3}
10^{-2}
10^{-1}
α
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
Layer 7

CIFAR-10 - Top-1 accuracy



CIFAR10 - Top-5 accuracy




CIFAR10 - Complexity vs top1
Performance
0.92
0.9
0.88
0.86
0.84
0.82
1
1.5
2
2.5
3
3.5
4
Complexity
$\cdot 10^7$
$\tau = 0$
$\tau = 1e-5$
$\tau = 1e-4$
$\tau = 1e-3$
$\tau = 1e-2$
$\tau = 1e-1$
$\alpha = 1e-4$
$\alpha = 1e-3.8$
$\alpha = 1e-3.6$
$\alpha = 1e-3.4$
$\alpha = 1e-3.2$
$\alpha = 1e-3$
CIFAR10 - Complexity vs top5
Performance
1
1
0.99
0.99
0.99
0.99
1
1.5
2
2.5
3
3.5
4
Complexity
$\cdot 10^7$
$\tau = 0$
$\tau = 1e-5$
$\tau = 1e-4$
$\tau = 1e-3$
$\tau = 1e-2$
$\tau = 1e-1$
$\alpha = 1e-4$
$\alpha = 1e-3.8$
$\alpha = 1e-3.6$
$\alpha = 1e-3.4$
$\alpha = 1e-3.2$
$\alpha = 1e-3$