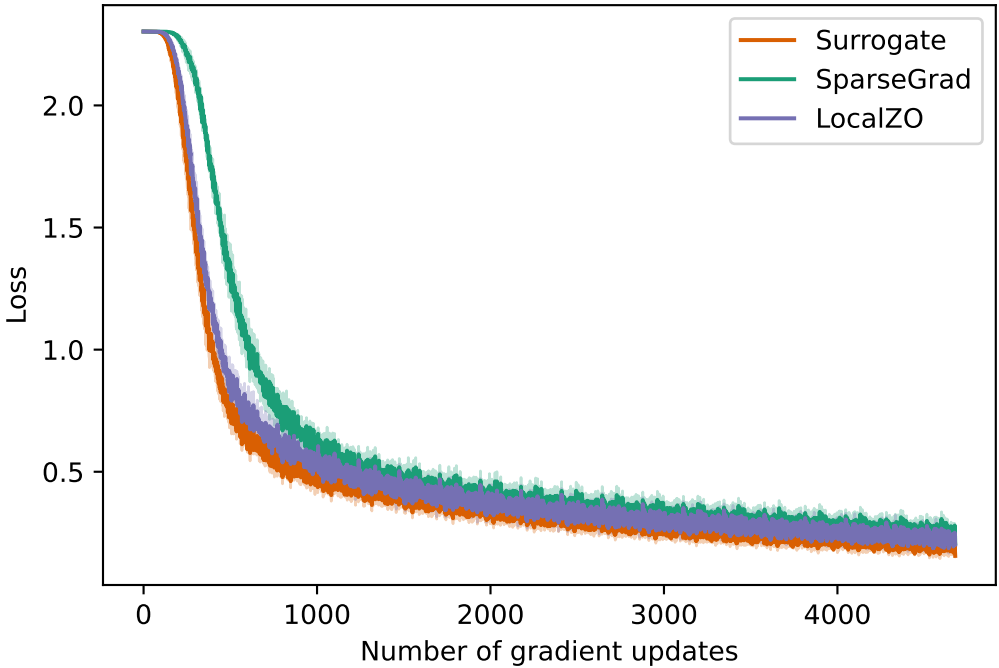
Surrogate
SparseGrad
LocalZO
Loss
2.0
1.5
1.0
0.5
0
1000
2000
3000
4000
Number of gradient updates