Learning by Turning: Neural Architecture Aware Optimisation

Yang Liu* Jeremy Bernstein* Markus Meister Yisong Yue

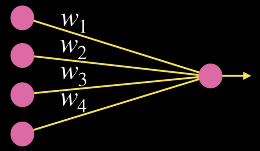
Introduction

Deep learning optimisers like Adam and SGD require lots of hyperparameter tuning.

We propose Nero—a new optimiser that we found usually works well *out-of-the-box*.

Nero: the neuronal rotator

The Nero update works per-neuron:



Nero involves two ingredients:

1. Projected gradient descent under two per-neuron constraints:

$$\sum_i w_i = 0, \quad \sum_i w_i^2 = 1.$$

2. Per-neuron relative updates:

$$\|\Delta w\|_2 \leq \eta \|w\|_2.$$

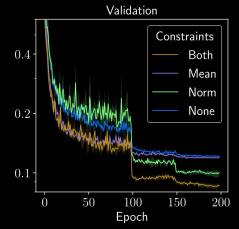
Putting this all together, Nero rotates each neuron by an angle $\approx \eta$ each iteration.

Main benchmarks

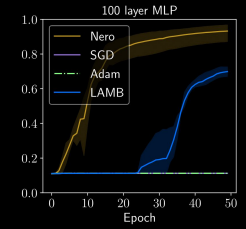
Task	Dataset	Model	Metric (↑↓)	Nero	SGD	Adam	LAMB	Nero η	SGD η	Adam η	LAMB η
cGAN	CIFAR-10	BigGAN-like	FID (↓)	**15.43 ± 0.37**	33.06 ± 0.42	23.42 ± 0.85	16.32 ± 0.23	0.01	0.01	0.0001	0.01
Classification	CIFAR-10	VGG11	Top-1 Error (↓)	**11.16% ± 0.17**	12.61% ± 0.21	12.86% ± 0.34	13.66% ± 0.05	0.01	0.1	0.001	0.01
Classification	CIFAR-10	ResNet-18	Top-1 Error (↓)	**5.75% ± 0.07**	7.75% ± 0.17	5.93% ± 0.19	6.46% ± 0.12	0.01	0.1	0.01	0.1
Language Model	Wikitext-2	Transformer	Perplexity (↓)	**172.99 ± 0.51**	181.76 ± 0.49	178.05 ± 0.96	200.54 ± 0.53	0.01	1.0	0.0001	0.01
Translation	WMT16 En–De	Transformer	Perplexity (↓)	**11.35 ± 1.20**	92.40 ± 89.48	12.63 ± 0.34	16.36 ± 0.29	0.001	0.0001	0.0001	0.01
PPO	Atari Pong	vanilla CNN	Reward (↑)	**20.62 ± 0.05**	11.99 ± 8.65	15.92 ± 3.40	−19.46 ± 0.10	0.01	0.1	0.0001	0.001

Out-of-the-box Nero outperformed Adam, SGD and LAMB in 5 out of 6 experiments.

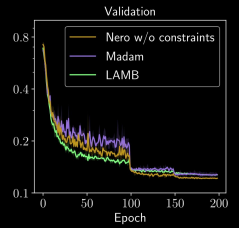
Additional experiments

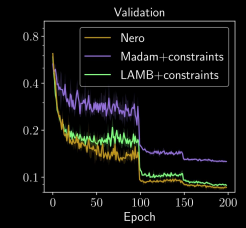


Constraints help.



Training a 100 layer MLP.

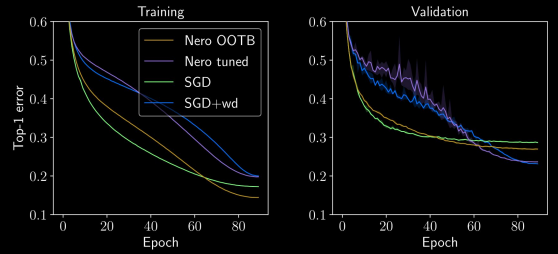




Testing per-neuron/synapse/layer updates.

Future work

To match fully tuned SGD with weight decay on ImageNet, Nero needed a regulariser that pushes batch norm gains towards one.



Future work could further explore this idea of *neural architecture aware regularisation.*

github.com/jxbz/nero