How to train your MAML

Antreas Antoniou[1], Harrison Edwards[1,2] and Amos Storkey[1]

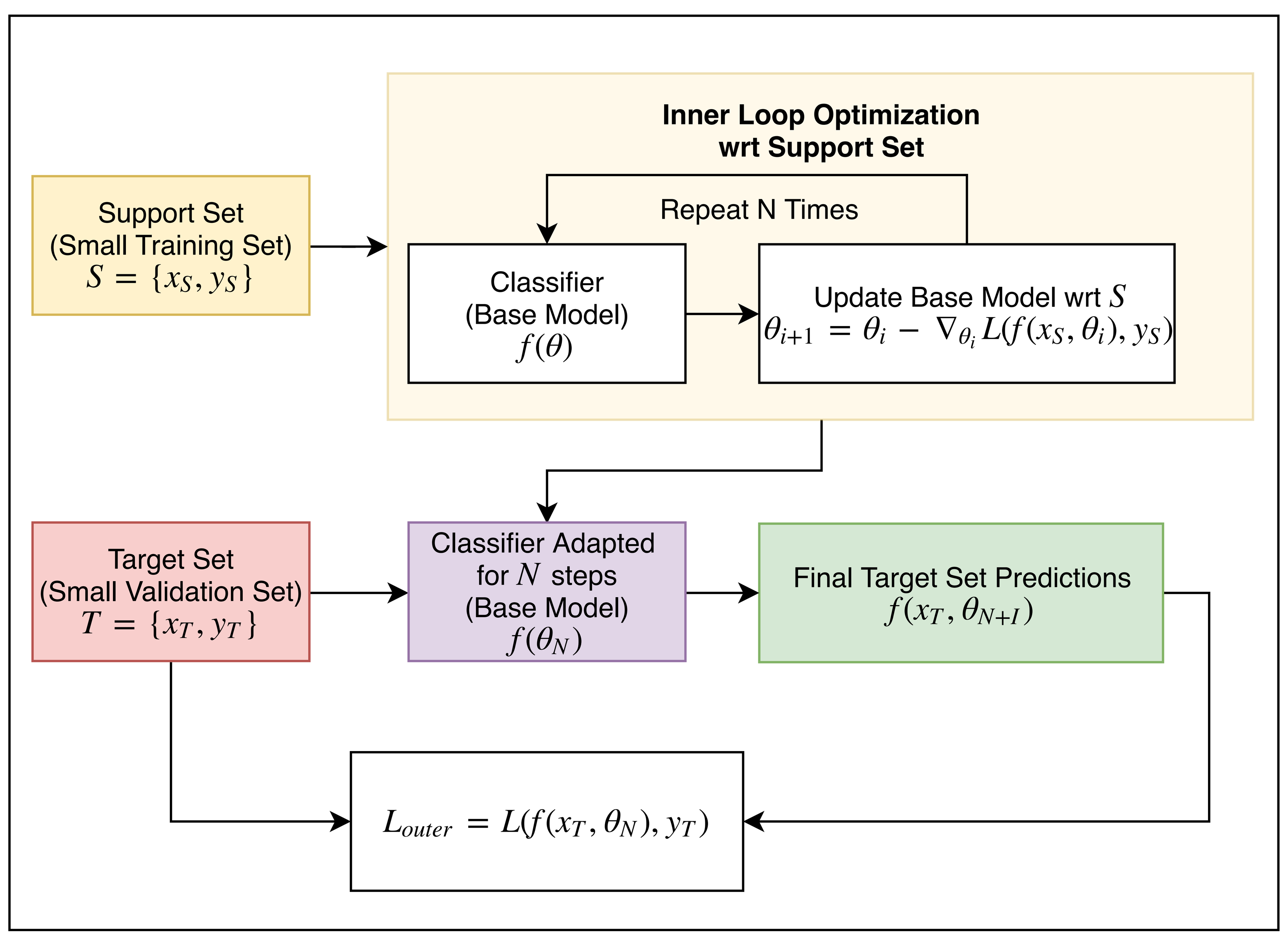
[1]University of Edinburgh, [2]OpenAI

Overview

1. *Model Agnostic Meta Learning* is currently one of the best meta-learning frameworks.
2. However, it suffers from a variety of issues, such as training instability, which can currently only be alleviated by arduous architecture and hyperparameter searches. In addition, MAML is very computationally expensive, mainly due to the second order derivatives computation that forms an integral part of the model.
3. We propose an improved variant of MAML, called *MAML++*, that addresses these problems, producing much improved training stability and decoupling the dependence of training stability on the model architecture, thus allowing MAML to be used in conjuction with more complicated neural networks, which in turn can facilitate learning of more complicated functions, such as loss functions, optimizers or even gradient computation functions.

Pre-requisities: Model Agnostic Meta Learning

Model Agnostic Meta-Learning (MAML) [1] is a meta-learning framework for few-shot learning. In a sentence, MAML learns good initialization parameters for a network, such that after a few steps of standard training on a few-shot dataset, the network will perform well on that few shot task.

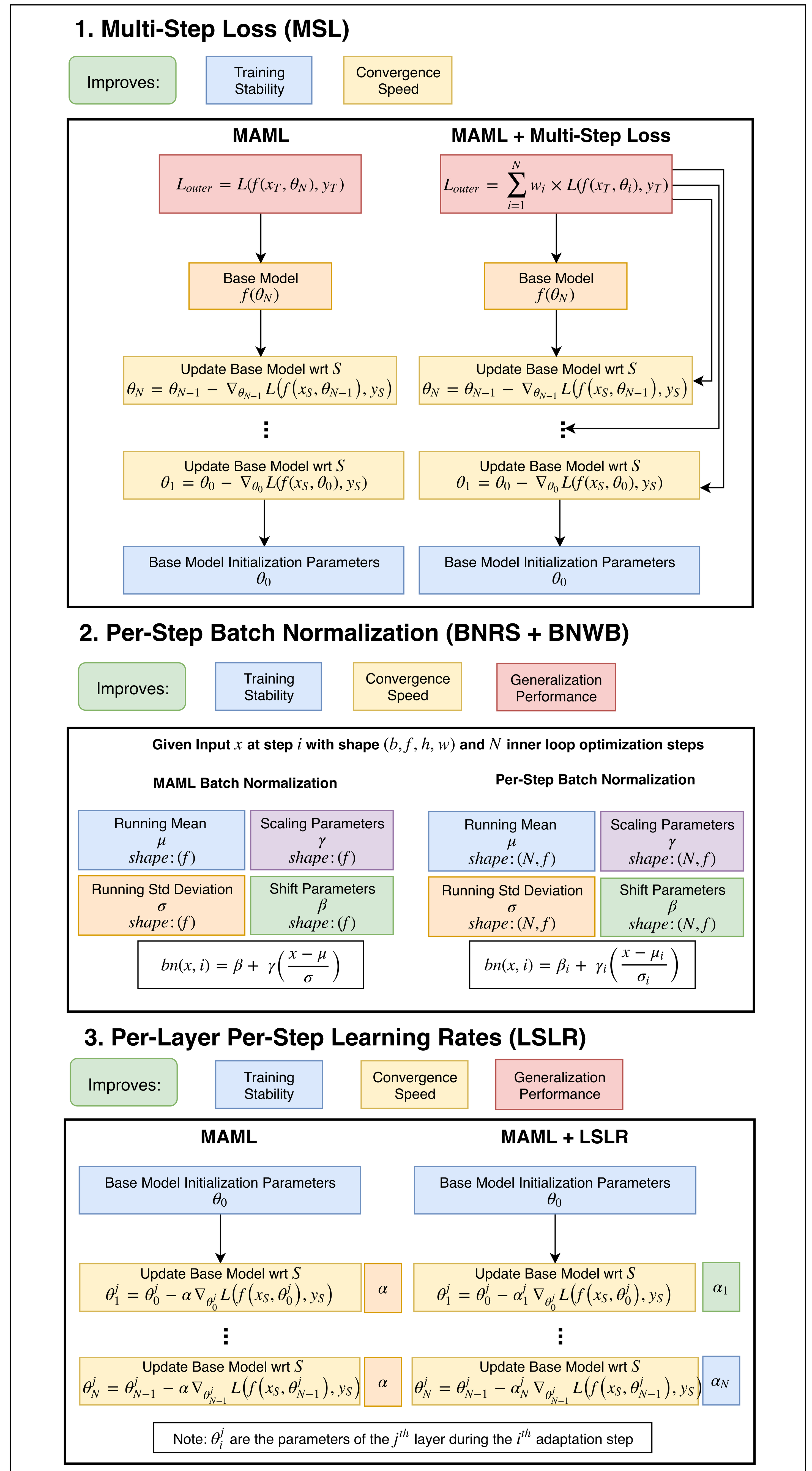


The resulting update for the meta-parameters θ_0 can be expressed as:

$$\theta_0 = \theta_0 - \beta \nabla_\theta \sum_{b=1}^{B} \mathcal{L}_{outer_b}(f_{\theta^b_N(\theta_0)}) \quad (1)$$

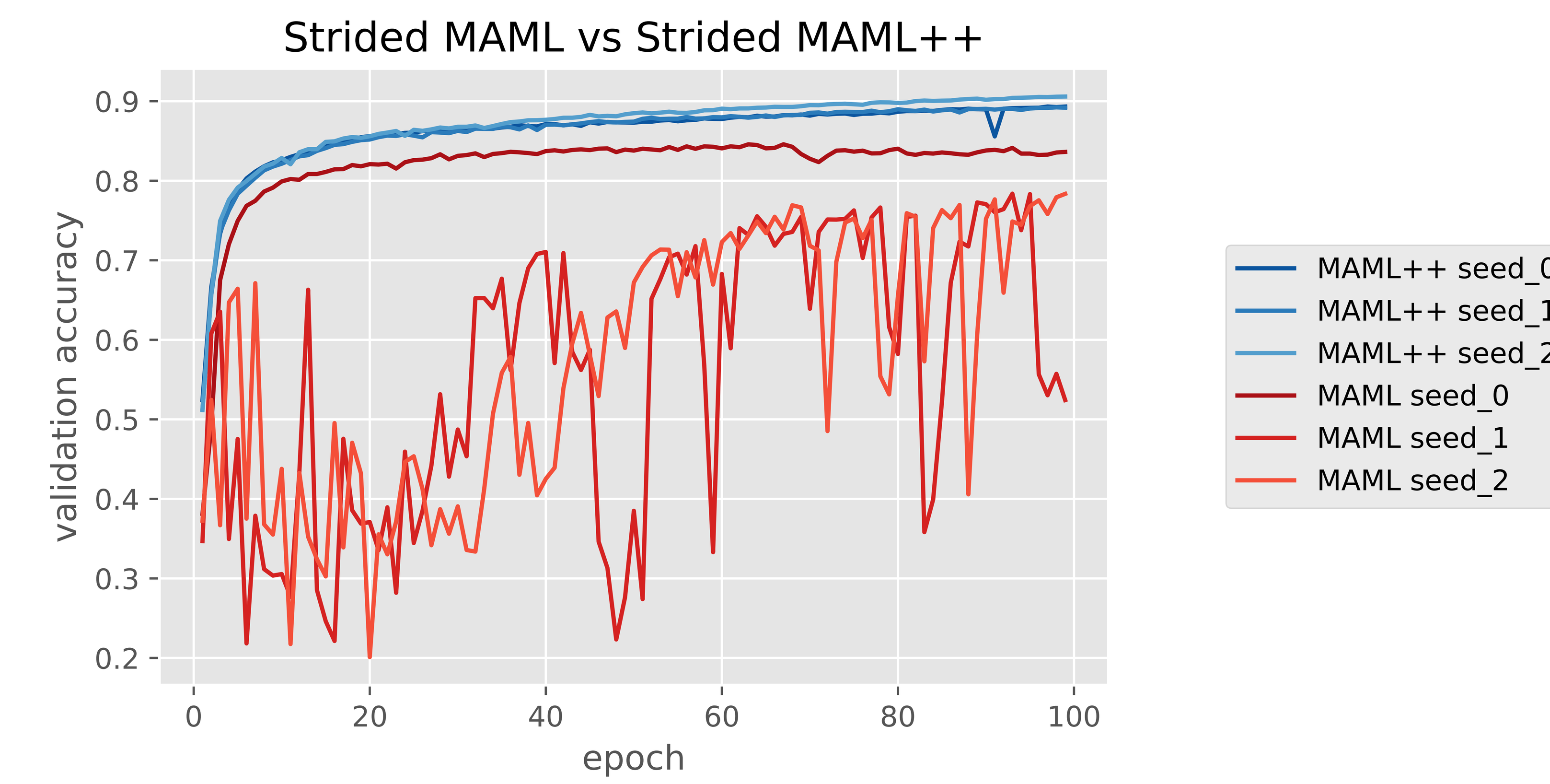
where β is a learning rate and $\mathcal{L}_{T_b}$ denotes the loss on the target set for task b. In this paper we use the cross-entropy [2] loss, denoted as $\mathcal{L}$ throughout.

Proposed Methods



4. Second Order Derivative Cost → Derivative-Order Annealing (DA): Annealing from first-order to second order gradients, at a given training epoch greatly speeds up training, while retaining the generalization performance of models trained with only second order gradients.

Results



Stabilizing MAML

	Accuracy			
	Omniglot 20-way		Mini-ImageNet 5-way	
Approach	**1-shot**	**5-shot**	**1-shot**	**5-shot**
Siamese Nets	88.2%	97.0%	-	-
Matching Nets	93.8%	98.5%	43.56%	55.31%
Neural Statistician	93.2%	98.1%	-	-
Memory Mod.	95.0%	98.6%	-	-
Meta-SGD	95.93±0.38%	98.97±0.19%	50.47±1.87%	64.03±0.94%
Meta-Networks	97.00%	–	49.21%	-
MAML (original)	95.8±0.3%	98.9±0.2%	48.70±1.84%	63.11±0.92%
MAML (local replication)	91.27±1.07%	98.78%	48.25±0.62%	64.39±0.31%
MAML++	**97.65±0.05%**	**99.33±0.03%**	**52.15±0.26%**	**68.32±0.44%**

MAML++ Few-Shot Results

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *In Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.

[2] Reuven Rubinstein. The cross-entropy method for combinatorial and continuous optimization. *Methodology and computing in applied probability*, 1(2):127–190, 1999.

[3] Geoffrey E Hinton and David C Plaut. Using fast weights to deblur old memories. 1987.

[4] Yoshua Bengio, Samy Bengio, and Jocelyn Cloutier. *Learning a synaptic learning rule.* Université de Montréal, Département d'informatique et de recherche opérationnelle, 1990.

[5] Jürgen Schmidhuber. Learning to control fast-weight memories: An alternative to dynamic recurrent networks. *Neural Computation*, 4(1):131–139, 1992.

[6] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pages 3630–3638, 2016.

[7] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. *Proceedings of the 4th International Conference on Learning Representations*, 2016.

[8] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few shot learning. *arXiv preprint arXiv:1707.09835*, 2017.

[9] Tsendsuren Munkhdalai and Hong Yu. Meta networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2554–2563, 2017.