This paper aims to develop the implicit MAML (iMAML) algorithm, an approach for optimization-based meta-learning with deep neural networks that removes the need for differentiating through the optimization path. The algorithm provided aims to learn a set of parameters
such that an optimization algorithm that is initialized at and regularized to this parameter vector
leads to good generalization for a variety of learning tasks. By leveraging the implicit differentiation approach, they derive an analytical expression for the meta (or outer level) gradient that depends only on the solution to the inner optimization and not the path is taken by the inner optimization algorithm. This decoupling of meta-gradient computation and choice of inner-level optimizer has several appealing properties. The algorithms provided in the paper are:

Algorithm 1 Implicit Model-Agnostic Meta-Learning (iMAML)

Require: Distribution over tasks $P(\mathcal{T})$, outer step size η, regularization strength λ,
while not converged **do**
  Sample mini-batch of tasks $\{\mathcal{T}_i\}_{i=1}^{B} \sim P(\mathcal{T})$
  for Each task $\mathcal{T}_i$ **do**
    Compute task meta-gradient $\boldsymbol{g}_i = \texttt{Implicit-Meta-Gradient}(\mathcal{T}_i, \boldsymbol{\theta}, \lambda)$
  end for
  Average above gradients to get $\hat{\nabla} F(\boldsymbol{\theta}) = (1/B) \sum_{i=1}^{B} \boldsymbol{g}_i$
  Update meta-parameters with gradient descent: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \hat{\nabla} F(\boldsymbol{\theta})$ *// (or Adam)*
end while

Algorithm 2 Implicit Meta-Gradient Computation

Input: Task $\mathcal{T}_i$, meta-parameters $\boldsymbol{\theta}$, regularization strength λ
Hyperparameters: Optimization accuracy thresholds δ and δ'
Obtain task parameters $\boldsymbol{\phi}_i$ using iterative optimization solver such that: $\|\boldsymbol{\phi}_i - \mathcal{A}lg_i^{\star}(\boldsymbol{\theta})\| \leq \delta$
Compute partial outer-level gradient $\boldsymbol{v}_i = \nabla_{\boldsymbol{\phi}} \mathcal{L}_{\mathcal{T}}(\boldsymbol{\phi}_i)$
Use an iterative solver (e.g. CG) along with reverse mode differentiation (to compute Hessian vector products) to compute $\boldsymbol{g}_i$ such that: $\|\boldsymbol{g}_i - \left(\boldsymbol{I} + \frac{1}{\lambda}\nabla^2 \hat{\mathcal{L}}_i(\boldsymbol{\phi}_i)\right)^{-1} \boldsymbol{v}_i\| \leq \delta'$
Return: $\boldsymbol{g}_i$

The paper also provides us with a detailed comparison between existing MAML algorithms vs theirs in terms of time complexity and space used and the consequent error achieved.

Algorithm	Compute	Memory	Error
MAML (GD + full back-prop)	$\kappa \log\left(\frac{D}{\delta}\right)$	$\text{Mem}(\nabla \hat{\mathcal{L}}_i) \cdot \kappa \log\left(\frac{D}{\delta}\right)$	0
MAML (Nesterov's AGD + full back-prop)	$\sqrt{\kappa} \log\left(\frac{D}{\delta}\right)$	$\text{Mem}(\nabla \hat{\mathcal{L}}_i) \cdot \sqrt{\kappa} \log\left(\frac{D}{\delta}\right)$	0
Truncated back-prop [53] (GD)	$\kappa \log\left(\frac{D}{\delta}\right)$	$\text{Mem}(\nabla \hat{\mathcal{L}}_i) \cdot \kappa \log\left(\frac{1}{\epsilon}\right)$	ϵ
Implicit MAML (this work)	$\sqrt{\kappa} \log\left(\frac{D}{\delta}\right)$	$\text{Mem}(\nabla \hat{\mathcal{L}}_i)$	δ

The authors also provide an experimental analysis to prove that the theory works. They primarily work on two datasets. One is a simple synthetic example for which the exact meta-gradient can be computed and compared against it. Second, the few-shot image recognition domains of Omniglot and Mini-ImageNet. For the synthetic example, they find that both iMAML and MAML asymptotically match the exact meta-gradient, but iMAML computes a better approximation in finite iterations. With 2 CG iterations, iMAML incurs a small terminal error which is expected. The computational cost, as well as memory of iMAML with 100 inner GD steps, is significantly smaller than MAML with 100 GD steps. On the Omniglot domain, we find that the GD version of iMAML is competitive with the full MAML algorithm, and substantially better than its approximations (i.e., first-order MAML and Reptile), especially for the harder 20-way tasks. We also find that iMAML with Hessian-free optimization performs substantially better than the other methods, suggesting that powerful optimizers in the inner loop can offer benefits to meta-learning. In the Mini-ImageNet domain, we find that iMAML performs better than MAML and FOMAML. To conclude, In this paper, a method for optimization-based meta-learning that removes the need for differentiating through the inner optimization path is developed, allowing us to decouple the outer meta-gradient computation from the choice of the inner optimization algorithm. We showed how this gives us significant gains in computing and memory efficiency, and also conceptually allows us to use a variety of inner optimization methods.

References: *Rajeswaran, et al. Meta-learning with implicit gradients. NeurIPS 2019.*