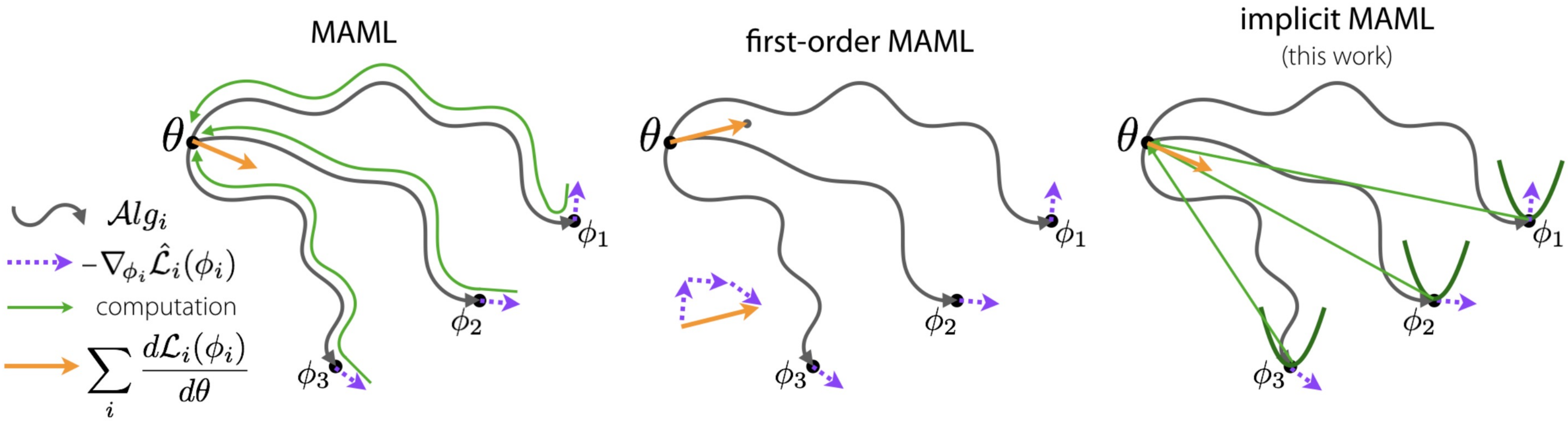
LUH|AI Leibniz Universität Hannover

iMAML: Meta-Learning with Implicit Gradient

[Rajeswaran et al. (2019)](https://arxiv.org/abs/1909.04630)



iMAML: Meta-Learning with Implicit Gradient

Rajeswaran et al. (2019)

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

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

Algorithm 2 Implicit Meta-Gradient Computation

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

iMAML: Meta-Learning with Implicit Gradient

Rajeswaran et al. (2019)

- Inner-loop regularization

$$\mathcal{A}lg^{\star}(\boldsymbol{\theta}, \mathcal{D}_i^{\mathrm{tr}}) = \underset{\boldsymbol{\phi}' \in \Phi}{\mathrm{argmin}} \; \mathcal{L}(\boldsymbol{\phi}', \mathcal{D}_i^{\mathrm{tr}}) + \frac{\lambda}{2} \, ||\boldsymbol{\phi}' - \boldsymbol{\theta}||^2$$

- Derivative of implicit Jacobian

$$\frac{d\mathcal{A}lg_i^{\star}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \left(\boldsymbol{I} + \frac{1}{\lambda} \nabla_{\boldsymbol{\phi}}^2 \hat{\mathcal{L}}_i(\boldsymbol{\phi}_i) \right)^{-1}$$