A Base for Any Order Gradient Estimation in Stochastic Computation Graphs

ICML 2019

Jingkai Mao[*1], Jakob N. Foerster[*2], Tim Rocktäschel[3], Maruan Al-Shedivat[4], Gregory Farquhar[2], Shimon Whiteson[2]

[1]Man AHL (Work done at the University of Oxford)

[2]University of Oxford

[3]University College London

[4]Carnegie Mellon University

June, 2019

[*]Equal Contribution

Authors

A Baseline for Any Order Gradient Estimation in Stochastic Computation Graphs

Jingkai Mao[*1]

Jakob N. Foerster[*2]

Tim Rocktäschel[3]

Maruan Al-Shedivat[4]

Gregory Farquhar[2]

Shimon Whiteson[2]

[1]Man AHL (Work done at the University of Oxford)

[2]Department of Computer Science, University of Oxford

[3]Department of Computer Science, University College London

[4]School of Computer Science, Carnegie Mellon University

[*] Equal Contribution

- **Applications of higher order gradients in reinforcement learning and meta-learning**
- **Estimating higher order gradients accurately and efficiently**

Computer Science > Artificial Intelligence

Learning with Opponent-Learning Awareness

Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, Igor Mordatch

(Submitted on 13 Sep 2017 (v1), last revised 19 Sep 2018 (this version, v4))

Computer Science > Machine Learning

Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks

Chelsea Finn, Pieter Abbeel, Sergey Levine

(Submitted on 9 Mar 2017 (v1), last revised 18 Jul 2017 (this version, v3))

Computer Science > Machine Learning

DiCE: The Infinitely Differentiable Monte-Carlo Estimator

Jakob Foerster, Gregory Farquhar, Maruan Al-Shedivat, Tim Rocktäschel, Eric P. Xing, Shimon Whiteson

(Submitted on 14 Feb 2018 (v1), last revised 19 Sep 2018 (this version, v3))

Computer Science > Machine Learning

Continuous Adaptation via Meta-Learning in Nonstationary and Competitive Environments

Maruan Al-Shedivat, Trapit Bansal, Yuri Burda, Ilya Sutskever, Igor Mordatch, Pieter Abbeel

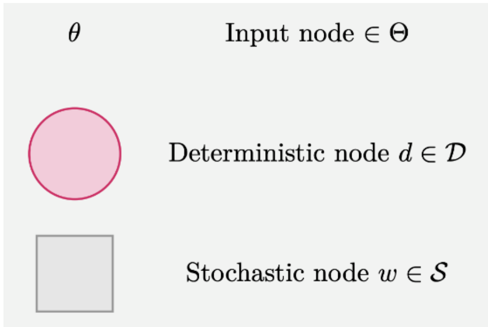
(Submitted on 10 Oct 2017 (v1), last revised 23 Feb 2018 (this version, v2))

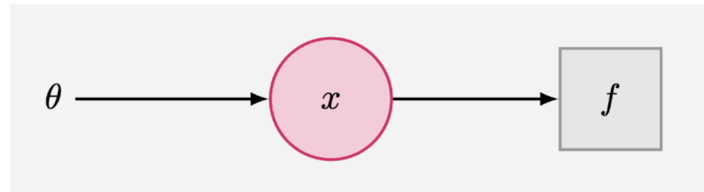
Background

A Baseline for Any Order Gradient Estimation in Stochastic Computation Graphs

Stochastic Computation Graphs (Schulman et al., 2015)

- **Nodes:**





- **Objective function:**

$$\mathcal{L} = \mathbb{E}\left[\sum_{c \in \mathcal{C}} c\right]$$

- **First Order Gradient**
  - **Score function estimator**
- **Higher Order Gradient?**

DiCE: The Infinitely Differentiable Monte Carlo Estimator (Foerster et al., 2018b)

- **The magic box operator**: for a set of stochastic nodes

$$1.\ \boxdot(\mathcal{W}) \rightarrowtail 1,$$

$$2.\ \nabla_\theta \boxdot(\mathcal{W}) = \boxdot(\mathcal{W}) \sum_{w \in \mathcal{W}} \nabla_\theta \log p(w; \theta).$$

- **DiCE objective function**

$$\mathcal{L} = \mathbb{E}\left[\sum_{c \in \mathcal{C}} c\right] \Rightarrow \mathcal{L}_{\boxdot} = \sum_{c \in \mathcal{C}} \boxdot(\mathcal{S}_c) c \Rightarrow \mathbb{E}[\nabla_\theta^n \mathcal{L}_{\boxdot}] \rightarrowtail \nabla_\theta^n \mathcal{L}$$

$$\mathcal{S}_c = \{s | s \in \mathcal{S}, s \prec c, \theta \prec s\}$$

- **Variance Reduction: baseline term** $\mathcal{B}_{\boxdot}$

$$\mathcal{L}_{\boxdot}^b = \mathcal{L}_{\boxdot} + \mathcal{B}_{\boxdot} \Rightarrow \nabla_\theta^n \mathcal{L}^b\ ?$$

- **First Order Baseline (from original DiCE)**

$$\mathcal{B}_{\boxdot}^{(1)} = \sum_{w \in \mathcal{S}} (1 - \boxdot(\{w\}))b_w$$

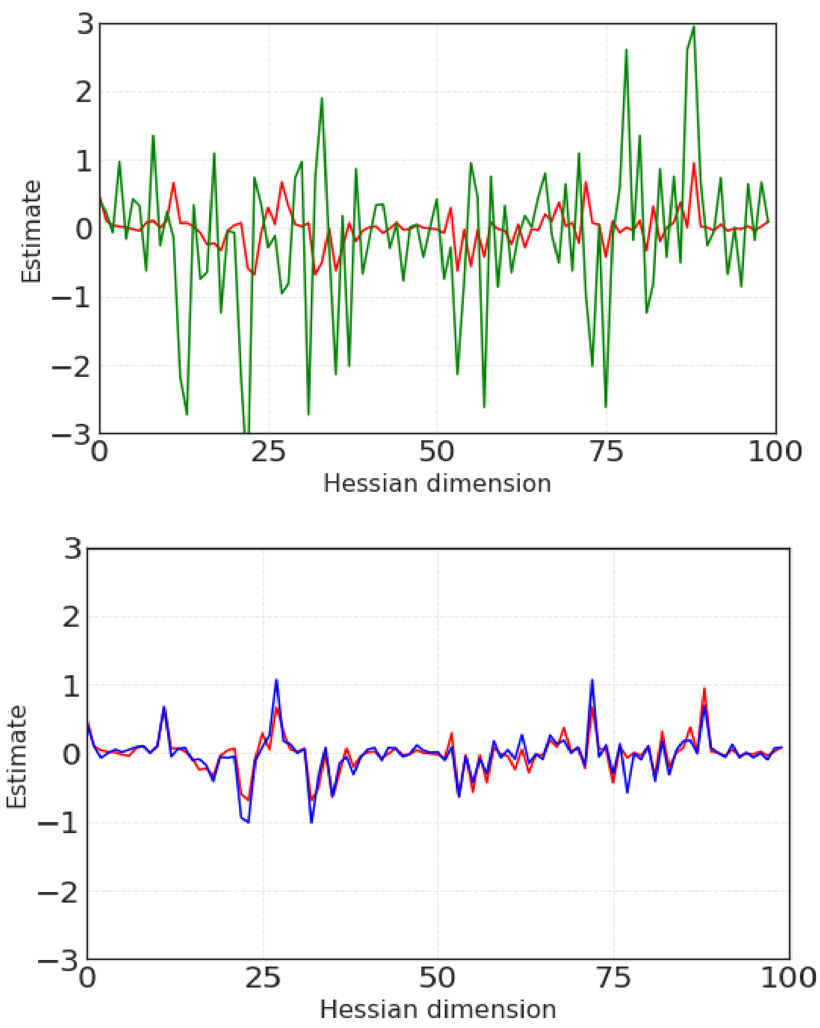
- **Second Order Baseline**

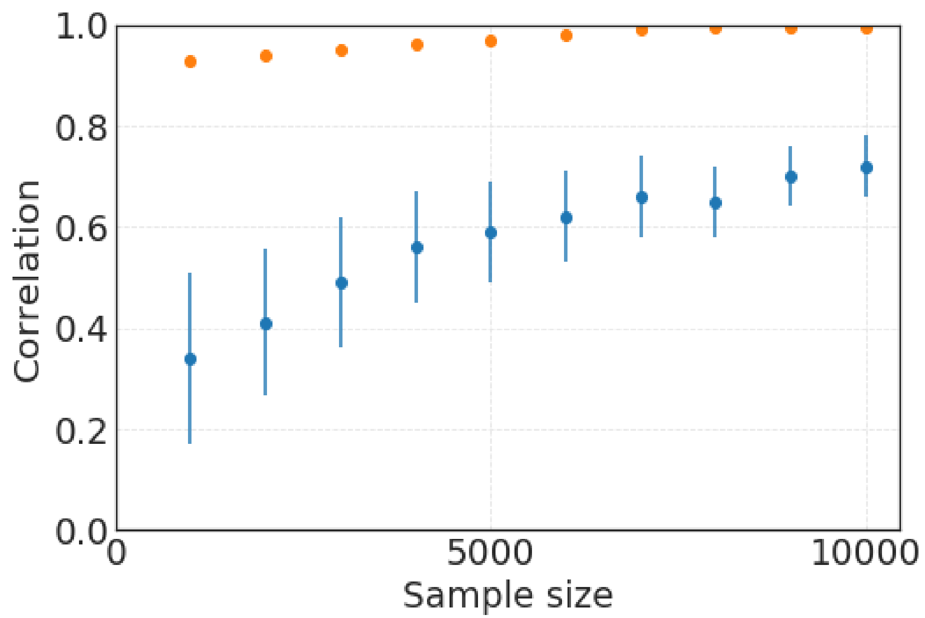
$$\mathcal{B}_{\boxdot}^{(2)} = -\sum_{w \in \mathcal{S}} \big(1 - \boxdot(\{w\})\big)\big(1 - \boxdot(\mathcal{S}_w)\big)b_w$$

- **Higher Order Baseline**

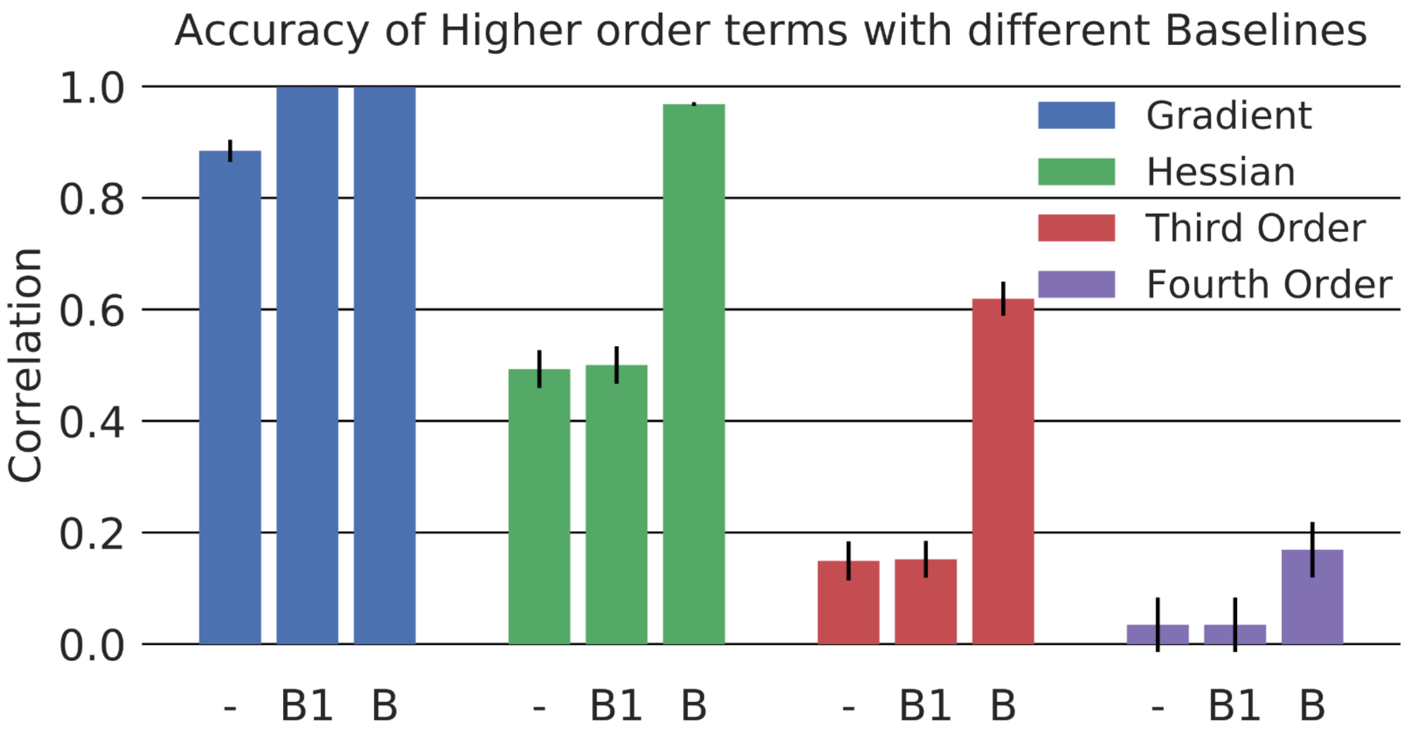
$$\begin{aligned}\mathcal{B}_{\boxdot} &= \mathcal{B}_{\boxdot}^{(1)} + \mathcal{B}_{\boxdot}^{(2)} \\ &= \sum_{w \in \mathcal{S}} b_w \big(1 - \boxdot(\{w\})\big) \cdot \boxdot(\mathcal{S}_w)\end{aligned} \quad \Longrightarrow \quad \mathbb{E}\left[\nabla_\theta^2 \mathcal{L}_{\boxdot}^{b}\right] \rightarrowtail \nabla_\theta^2 \, \mathbb{E}\left[\mathcal{L}\right]$$

First and Second Order Baselines





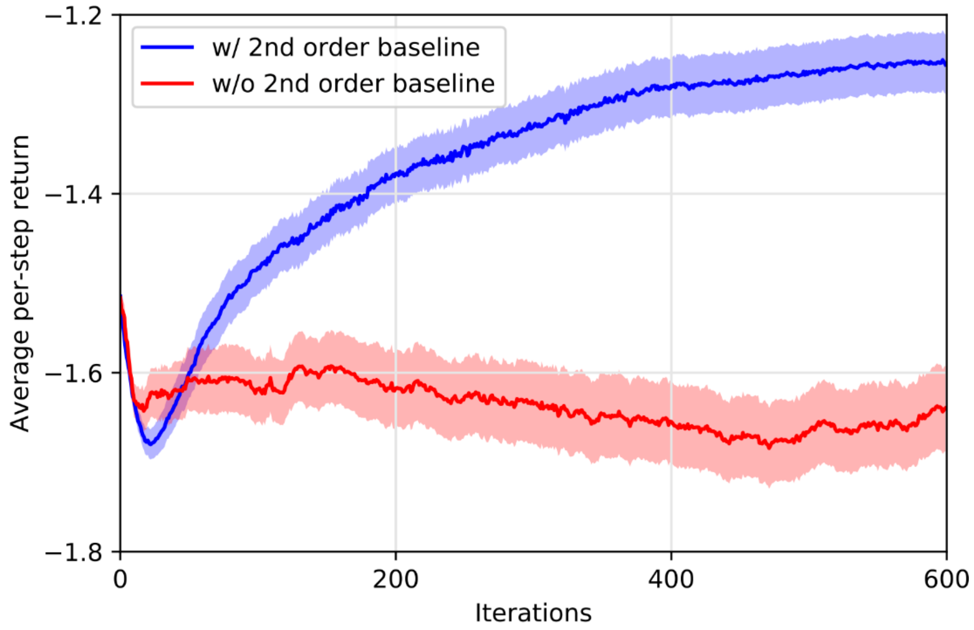
Higher Order Baselines



Learning with Opponent Learning Awareness – DiCE (Foerster et al., 2018b)

$$\mathcal{L}^1(\theta_1, \theta_2)_{\mathrm{LOLA}} = \mathbb{E}_{\pi_{\theta_1}, \pi_{\theta_2 + \Delta\theta_2(\theta_1, \theta_2)}}[\sum_{t=0}^{T} \gamma^t r_t^1]$$

$$\Delta\theta_2(\theta_1, \theta_2) = \alpha_2 \nabla_{\theta_2} \mathbb{E}_{\pi_{\theta_1}, \pi_{\theta_2}}[\sum_{t=0}^{T} \gamma^t r_t^2]$$



Man

- **A baseline for efficient estimations of higher order gradients using DiCE formalism**
- **Examples:**
  - **Gradient estimations up to 4th order**
  - **LOLA-DiCE with 2nd order baseline**
- **Future work**
  - **Extending the framework to a base-generating term for any order gradient estimators**
  - **Further applications in RL and meta-learning**

- Al-Shedivat, M., Bansal, T., Burda, Y., Sutskever, I., Mordatch, I., and Abbeel, P. Continuous adaptation via meta-learning in nonstationary and competitive environments. *CoRR*, abs/1710.03641, 2017.
- Finn, C., Abbeel, P., and Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017*, pp. 1126–1135, 2017.
- Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 122–130. International Foundation for Autonomous Agents and Multiagent Systems, 2018a.
- Foerster, J., Farquhar, G., Al-Shedivat, M., Rocktäschel, T., Xing, E., and Whiteson, S. Dice: The infinitely differentiable monte carlo estimator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1524–1533, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018b. PMLR. URL http://proceedings.mlr. press/v80/foerster18a.html.
- Schulman, J., Heess, N., Weber, T., and Abbeel, P. Gradient estimation using stochastic computation graphs. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems*, pp. 3528–3536, 2015.

Important Information

This information has, unless otherwise stated, been prepared by the relevant AHL or Man entity identified below (collectively the "Company") subject to the following conditions and restriction in their respective jurisdictions.

Opinions expressed are those of the author and may not be shared by all personnel of Man Group plc ('Man'). These opinions are subject to change without notice, are for information purposes only and do not constitute an offer or invitation to make an investment in any financial instrument or in any product to which the Company and/or its affiliates provides investment advisory or any other financial services. Any organisations, financial instrument or products described in this material are mentioned for reference purposes only which should not be considered a recommendation for their purchase or sale. Neither the Company nor the authors shall be liable to any person for any action taken on the basis of the information provided. Some statements contained in this material concerning goals, strategies, outlook or other non-historical matters may be forward-looking statements and are based on current indicators and expectations. These forward-looking statements speak only as of the date on which they are made, and the Company undertakes no obligation to update or revise any forward-looking statements. This material is proprietary information of the Company and its affiliates and may not be reproduced or otherwise disseminated in whole or in part without prior written consent from the Company. The Company believes the content to be accurate. However accuracy is not warranted or guaranteed. The Company does not assume any liability in the case of incorrectly reported or incomplete information. Unless stated otherwise all information is provided by the Company.

Unless stated otherwise this information has been prepared by and is communicated by AHL Partners LLP which is registered in England and Wales at Riverbank House, 2 Swan Lane, London, EC4R 3AD. Authorised and regulated in the UK by the Financial Conduct Authority.