# A framework for bilevel optimization that enables stochastic and global variance reduction algorithms

Mathieu Dagréou
Université Paris-Saclay, Inria, CEA, Palaiseau, 91120, France
mathieu.dagreou@inria.fr

Bilevel optimization, the problem of minimizing a *value function* which involves the arg-minimum of another function, appears in many areas of machine learning such as hyperparameter selection [4], neural architecture search [3] or Deep Equilibrium Networks [1]. In a large scale setting where the number of samples is huge, it is crucial to develop stochastic methods, which only use a few samples at a time to progress.

However, computing the gradient of the value function involves solving a linear system, which makes it difficult to derive unbiased stochastic estimates. To overcome this problem we introduce a novel framework, in which the solution of the inner problem, the solution of the linear system, and the main variable evolve at the same time. These directions are written as a sum, making it straightforward to derive unbiased estimates. The simplicity of our approach allows us to develop global variance reduction algorithms, where the dynamics of all variables is subject to variance reduction.

In this framework, we propose SOBA, a natural extension of stochastic gradient descent, and SABA, a natural adaptation of the variance reduction algorithm SAGA [2]. We demonstrate that SABA has $O(\frac{1}{T})$ convergence rate, and that it achieves linear convergence under Polyak-Łojasciewicz assumption. This is the first stochastic algorithm for bilevel optimization that verifies either of these properties. Numerical experiments on hyperparameter selection for $\ell^2$-regularized logistic regression (Figure 1) validate the usefulness of our method.

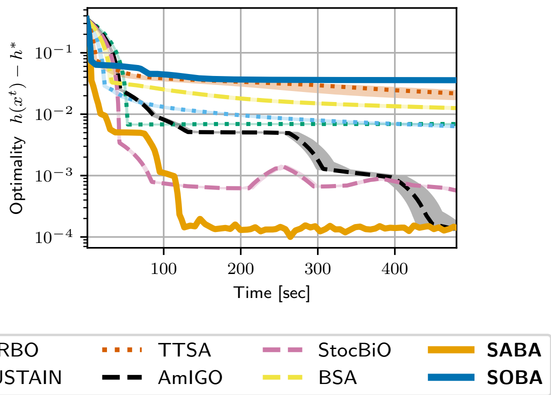


Figure 1: Suboptimality gap for hyperparameter selection for $\ell^2$ penalized logistic regression on IJCNN1 dataset

**Joint work with:** Pierre Ablin, Samuel Vaiter, Thomas Moreau.

## References

[1] S. Bai, J. Z. Kolter, V. Koltun. Deep Equilibrium Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[2] A. Defazio, F. Bach, S. Lacoste-Julien. SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[3] H. Liu, K. Simonyan, Y. Yang. Darts: Differentiable architecture search. In *International Conference on Learn-ing Representations (ICLR)*, 2018.

[4] F. Pedregosa. Hyperparameter optimization with approximate gradient. In *International Conference on MachineLearning (ICML)*, 2016.