Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization

Fabian Pedregosa†, Rémi Leblond†, Simon Lacoste–Julien⋆

† INRIA and École Normale Supérieure, Paris, France. ⋆ MILA and DIRO, Université de Montréal, Canada

Summary

Optimization methods need to be adapted to the **parallel** setting to leverage modern computer architectures.

Highly efficient variants of stochastic gradient descent have been recently proposed, such as Hogwild [1], Kromagnon [2], ASAGA [3].

They assume that the objective function is smooth, so are inapplicable to problems such as Lasso, optimization with constraints, etc.

Contributions:

1. **Sparse Proximal SAGA**, a sparse variant of the linearly-convergent proximal SAGA algorithm.
2. **ProxASAGA**, the first parallel asynchronous variance-reduced method that supports composite objective functions.

Problem setting

We consider optimization problems of the form

$$\underset{\boldsymbol{x}\in\mathbb{R}^p}{\text{minimize}}\, f(\boldsymbol{x}) + h(\boldsymbol{x})\,, \quad \text{with } f(x) \stackrel{\text{def}}{=} \tfrac{1}{n}\textstyle\sum_{i=1}^n f_i(\boldsymbol{x})$$

where f is differentiable with L-Lipschitz gradient. We also assume that h is block-separable and "simple", in the sense that we have access to its proximal operator, defined as

$$\mathbf{prox}_h \stackrel{\text{def}}{=} \arg\min_{\boldsymbol{x}} h(\boldsymbol{x}) + \tfrac{1}{2}\|\boldsymbol{x}-\boldsymbol{z}\|^2\,.$$

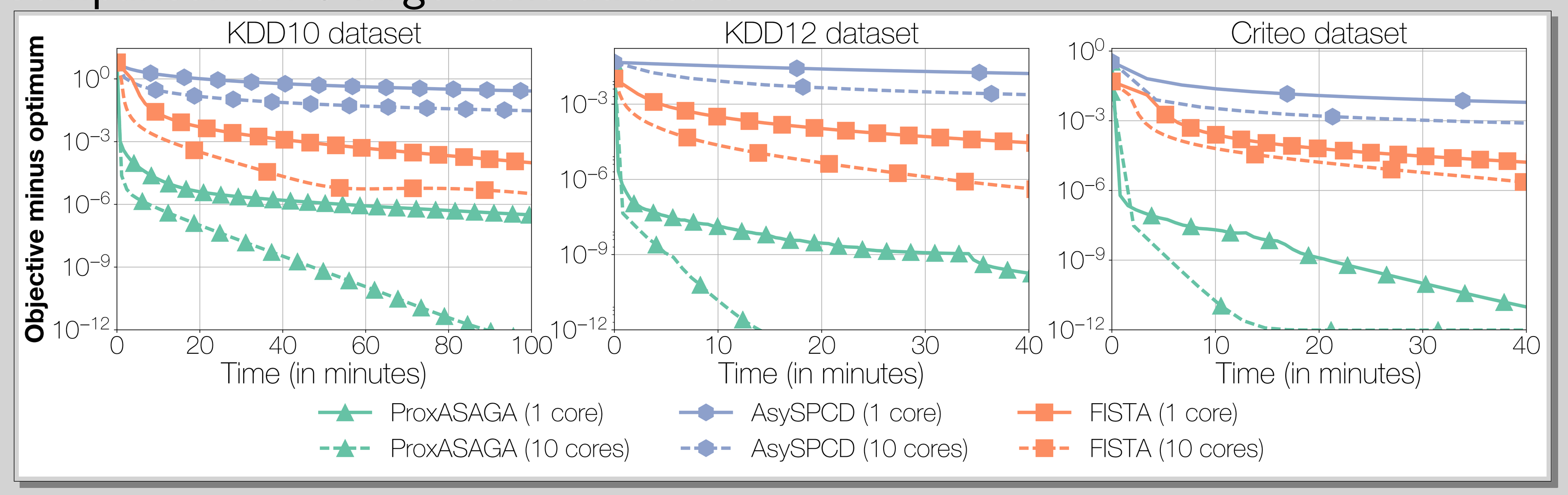
A new algorithm: Sparse Proximal SAGA

$$\boldsymbol{v}_i = \nabla f_i(\boldsymbol{x}) - \boldsymbol{\alpha}_i + \boldsymbol{D}_i\,,\ \boldsymbol{x}^+ = \mathbf{prox}_{\gamma\varphi_i}(\boldsymbol{x}-\gamma\boldsymbol{v}_i)$$

Proximal Asynchronous SAGA (ProxASAGA)

Experimental results

Comparison on 3 large-scale datasets



References

1. Niu, F., Recht, B., Re, C. & Wright, S. *Hogwild: A lock-free approach to parallelizing stochastic gradient descent.* in *NIPS* (2011).
2. Mania, H. *et al.* Perturbed iterate analysis for asynchronous stochastic optimization. *SIAM Journal on Optimization* (2017).
3. Leblond, R., Pedregosa, F. & Lacoste-Julien, S. ASAGA: asynchronous parallel SAGA. *AISTATS* (2017).