SCAFFLSA: Taming Heterogeneity in Federated Linear Stochastic Approximation

Paul Mangold, Sergey Samsonov, Safwan Labbi, Ilya Levin, Reda Alami, Alexey Naumov, Eric Moulines

CAp Conference

July 2nd, 2024

Linear Stochastic Approximation

Find $\theta_\star^c$ such that

$$A^c \theta_\star^c = b^c$$

Linear Stochastic Approximation

Find $\theta_\star^c$ such that

$$A^c \theta_\star^c = b^c$$

... but we only have unbiased estimators $A^c(Z), b^c(Z)$

Applications: TD learning, linear regression

Federated Linear Stochastic Approximation

Find $\theta_\star$ such that

$$\left(\frac{1}{N}\sum_{c=1}^{N} A^c\right)\theta_\star = \frac{1}{N}\sum_{c=1}^{N} b^c$$

... but we only have unbiased estimators $A^c(Z), b^c(Z)$

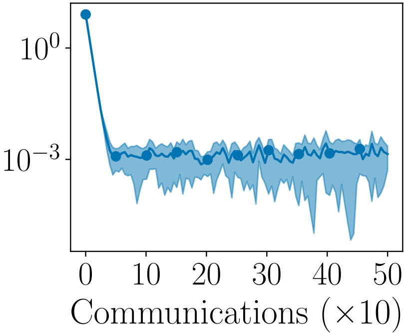
The FedLSA algorithm

* Initialize θ_0
* For $t = 0$ to $T - 1$:
    * Set $\theta^c_{t+1,0} = \theta_t$
    * For each agent c, for $h = 0$ to $H - 1$:

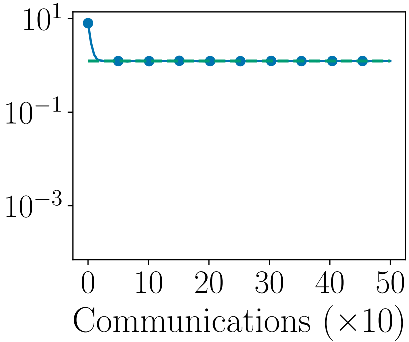
$$\theta^c_{t+1,h+1} = \theta^c_{t+1,h} - \eta(A^c\theta^c_{t+1,h} - b^c)$$

* Aggregate $\theta_{t+1} = \frac{1}{N}\sum_{c=1}^{N} \theta^c_{t+1,H}$

Works if agents are homogeneous ($H = 1000$)



Biased if agents are heterogeneous ($H = 1000$)



$\rightarrow$ and we can give a formal expression of this bias: if η and H are small, then bias is also small!

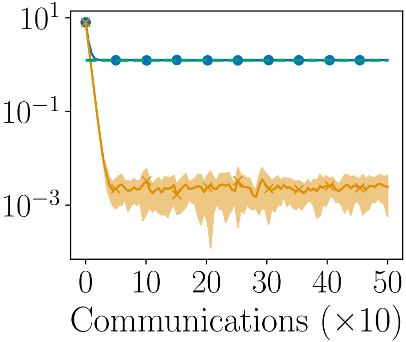
SCAFFLSA: Use Control Variates!

* Initialize $\theta_0, \xi_0^1, \ldots, \xi_0^N$
* For $t = 0$ to $T-1$:
  * Set $\theta_{t+1,0}^c = \theta_t$
  * For each agent c, for $h = 0$ to $H-1$:

$$\theta_{t+1,h+1}^c = \theta_{t+1,h}^c - \eta(A^c\theta_{t+1,h}^c - b^c - \xi_t^c)$$

* Aggregate $\theta_{t+1} = \frac{1}{N}\sum_{c=1}^N \theta_{t+1,H}^c$
* Update $\xi_{t+1}^c = \xi_t^c + \frac{1}{\eta H}(\theta_{t+1} - \theta_{t+1,H}^c)$

Works even if agents are heterogeneous ($H = 1000$)



Algorithm	Communication T	Local updates H	Total samples
FedLSA	$\mathcal{O}\left(\frac{1}{a^2\epsilon}\log\frac{1}{\epsilon}\right)$	$\mathcal{O}\left(\frac{1}{N\epsilon}\right)$	$\mathcal{O}\left(\frac{1}{Na^2\epsilon^2}\log\frac{1}{\epsilon}\right)$
Scafflsa	$\mathcal{O}\left(\frac{1}{a^2}\log\frac{1}{\epsilon}\right)$	$\mathcal{O}\left(\frac{1}{N\epsilon^2}\right)$	$\mathcal{O}\left(\frac{1}{Na^2\epsilon^2}\log\frac{1}{\epsilon}\right)$

Come to the poster for theoretical results:

* linear speed-up
* acceleration in the setting where noise dominates