Intra-iteration Approximation for Large Scale Parallel Gradient Descent Optimization

Chengjie Qin & Florin Rusu

EECS, University of California, Merced

Motivation & Goal

- Per iteration time of large-scale training can be long even if parallel training is deployed.
- A user can do nothing while an time consuming iteration is running.
- We introduce intra-iteration techniques to speed up large scale parallel training and to allow interactive parameter tuning.

(Stochastic) Gradient Descent

- Problem definition:

$$min_{w\in\mathbb{R}^d} \Lambda(\vec{w}) \stackrel{def}{=} \sum_{i=1}^{N} f\left(\vec{w}, \vec{x_i}; y_i\right) + \mu R(\vec{w}) \quad (1)$$

$\vec{w}$ is the parameter to be learned, $\vec{x_i}$ is feature vector, y_i is label

- Gradient Descent

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla\Lambda\left(\vec{w}^{(k)}\right)$$

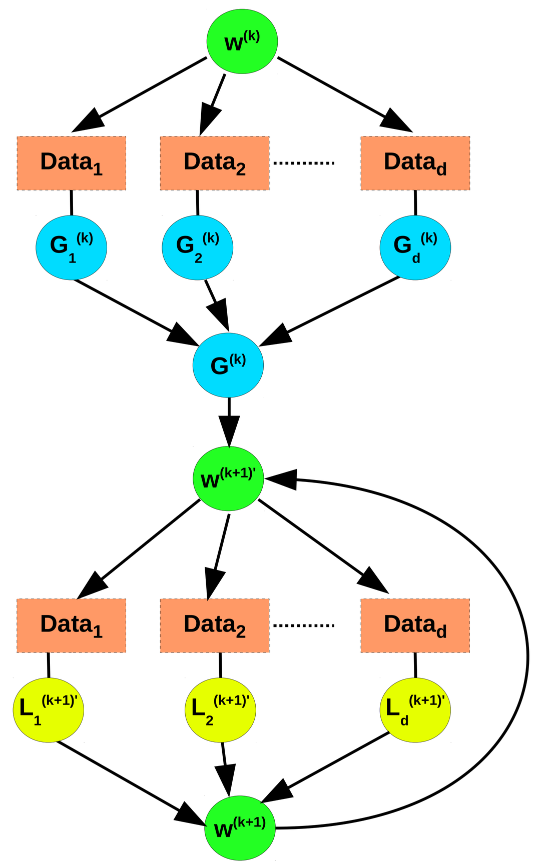
- Stochastic Gradient Descent

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla f_i\left(\vec{w}^{(k)}\right)$$

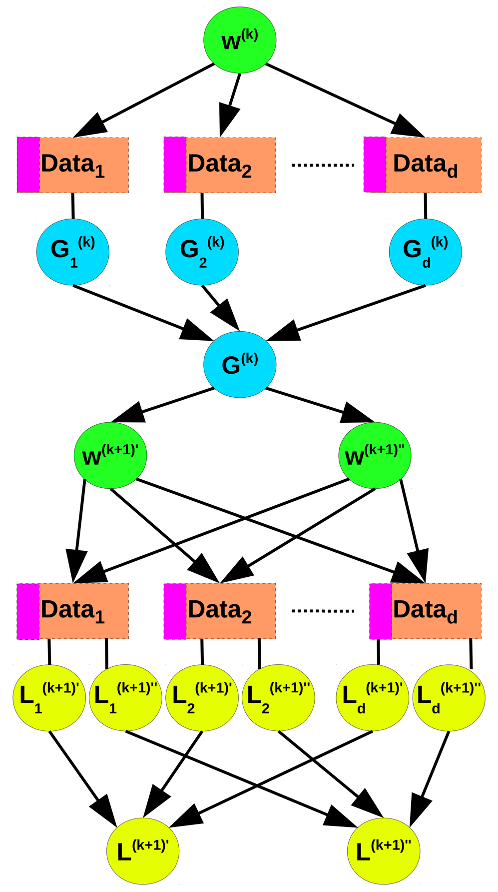
Datasets & Tasks

Dataset	Dimension	# Examples	Size
`splice`	13M	50M	3TB

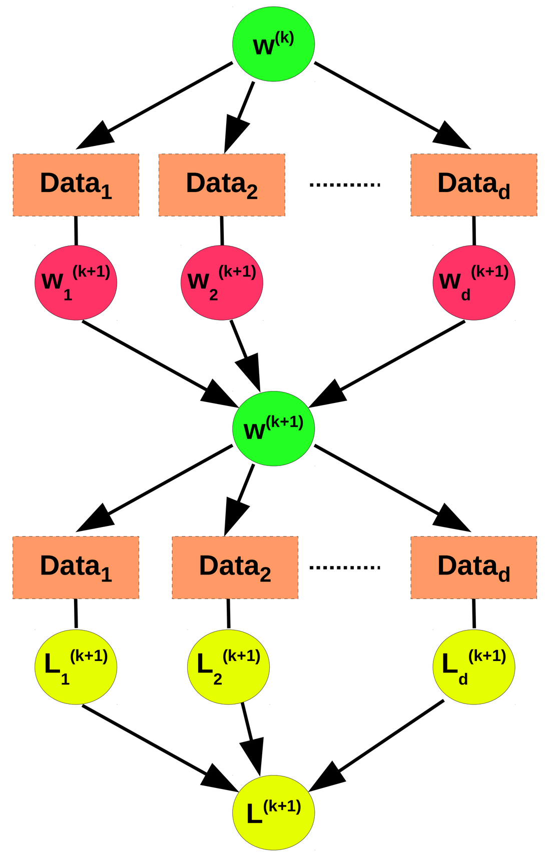
Query Overlapping & Intra-iteration Approximation Parallel (Stochastic) Gradient Descent in GLADE



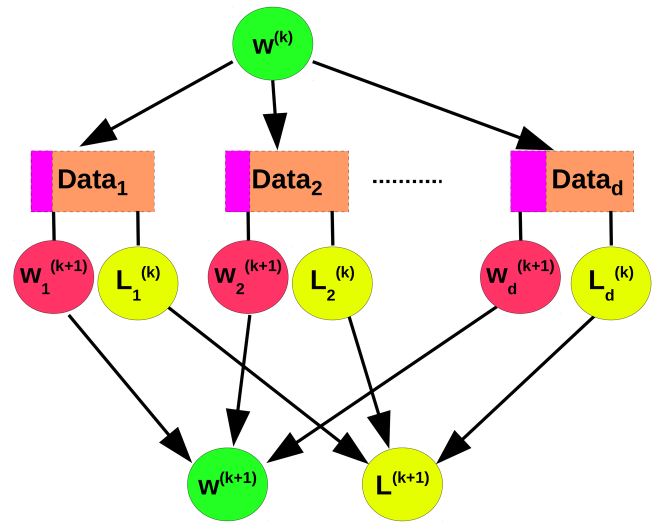
Parallel Gradient Descent (PGD)



Online Aggregation + Query Overlapping Parallel Gradient Descent (OLA PGD)

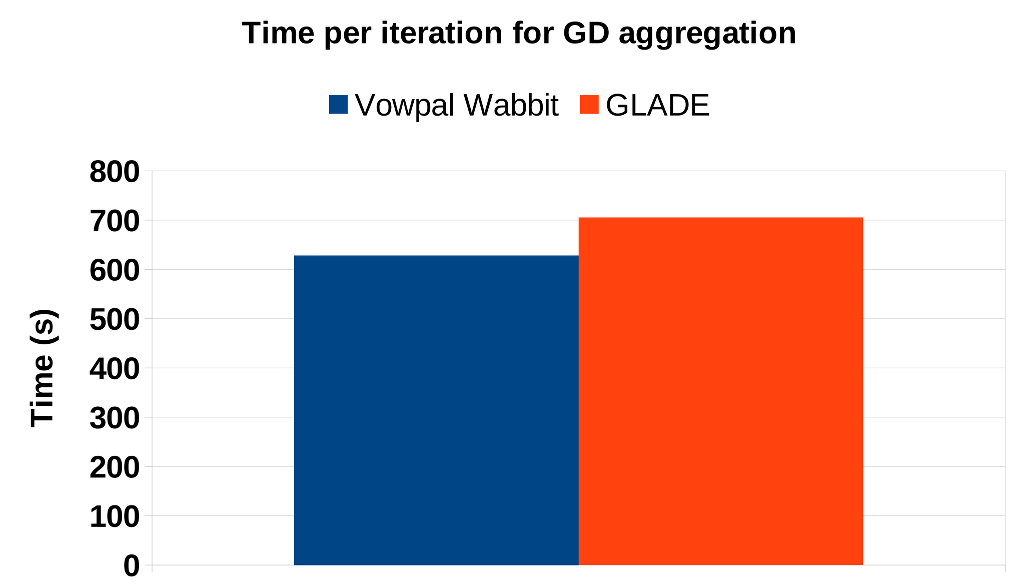


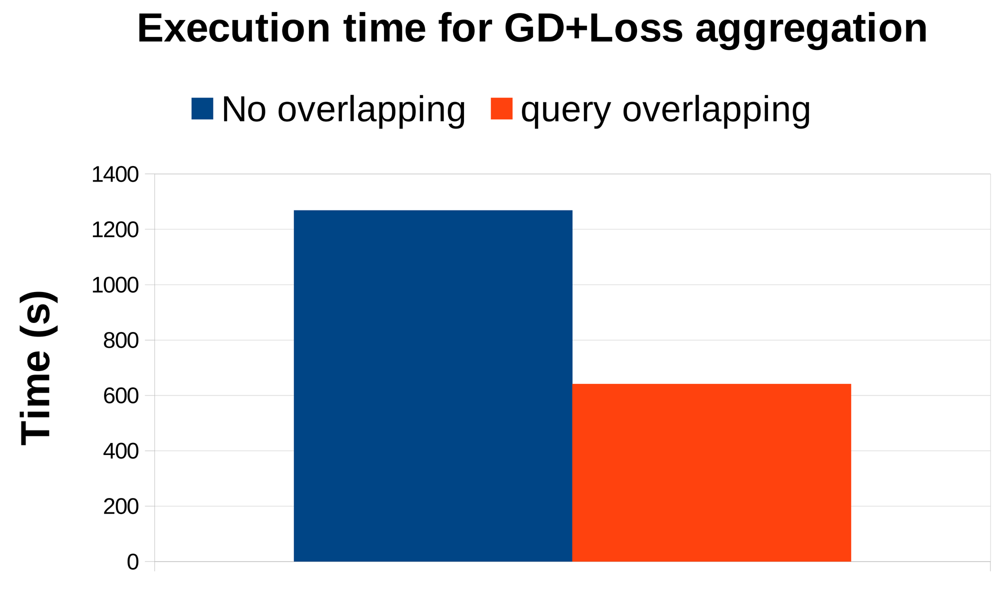
Parallel Stochastic Gradient Descent (PSGD)



Online Aggregation + Query Overlapping Parallel Stochastic Gradient Descent (OLA PSGD)

Iteration Time Comparison





- SVM on `splice` 50m examples, 13m features
- Single nodes: 16 cores @ 2GHz; 20GB RAM; 4 disks @ 110MB/s throughput/disk

OLA PGD Loss Comparison

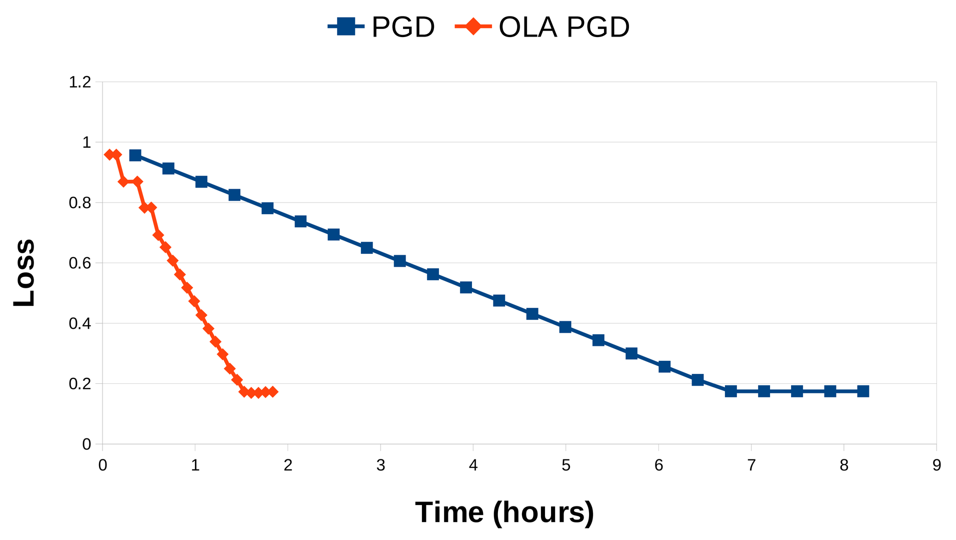


Figure: Loss comparison between Query Overlapping + PGD and OLA + Query Overlapping + PGD

- SVM on `splice`, 9 nodes

OLA PSGD Loss Comparison

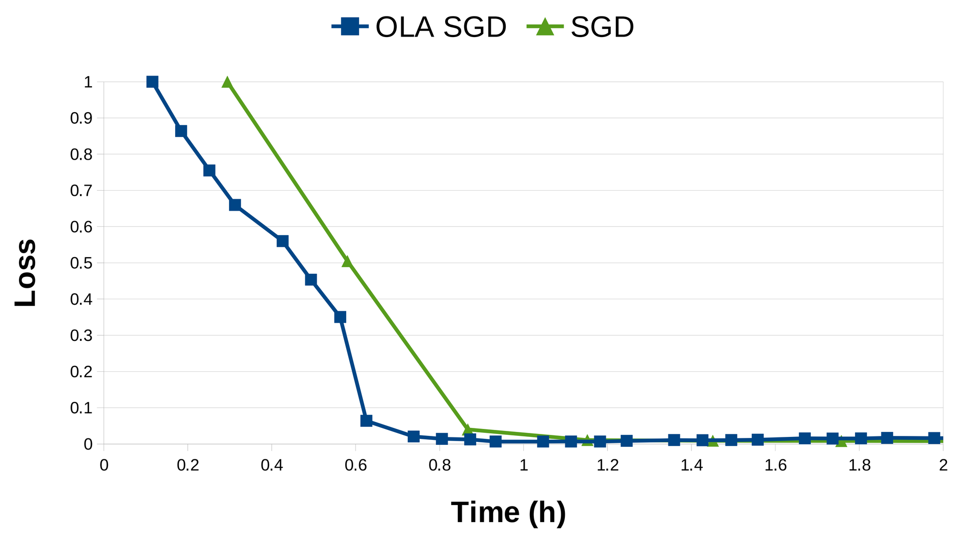


Figure: Loss comparison between Query Overlapping + PSGD and OLA + Query Overlapping + PSGD

- SVM on `splice`, 9 nodes

University of California, Merced

Mail: Chengjie Qin `cqin3@ucmerced.edu`, Florin Rusu `frusu@ucmerced.edu`