Effective Model Calibration for Terascale Analytics

Chengjie Qin & Florin Rusu

EECS, University of California, Merced

Motivation & Goal

- We present our gradient descent solution in GLADE with three distinct features from other data analytics infrastructures:
- I/O-bound execution with linear speed-up
- Data-model-mixed parallelism
- On-the-fly concurrent hyper-parameter testing and convergence detection by utilizing online aggregation (OLA)

(Incremental) Gradient Descent

- Problem definition:

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

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

- Batch Gradient Descent (BGD)

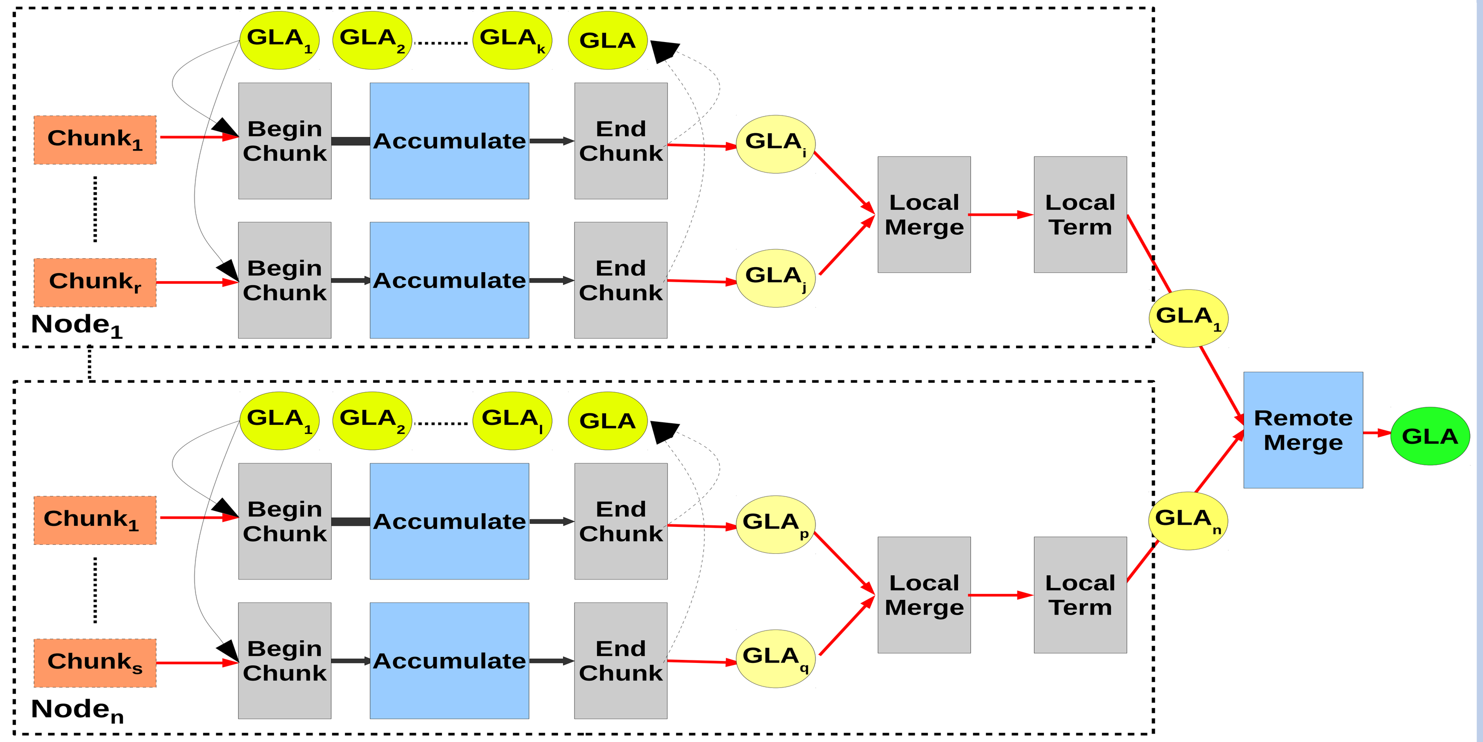
$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla\Lambda\left(\vec{w}^{(k)}\right)$$

- Incremental Gradient Descent (IGD)

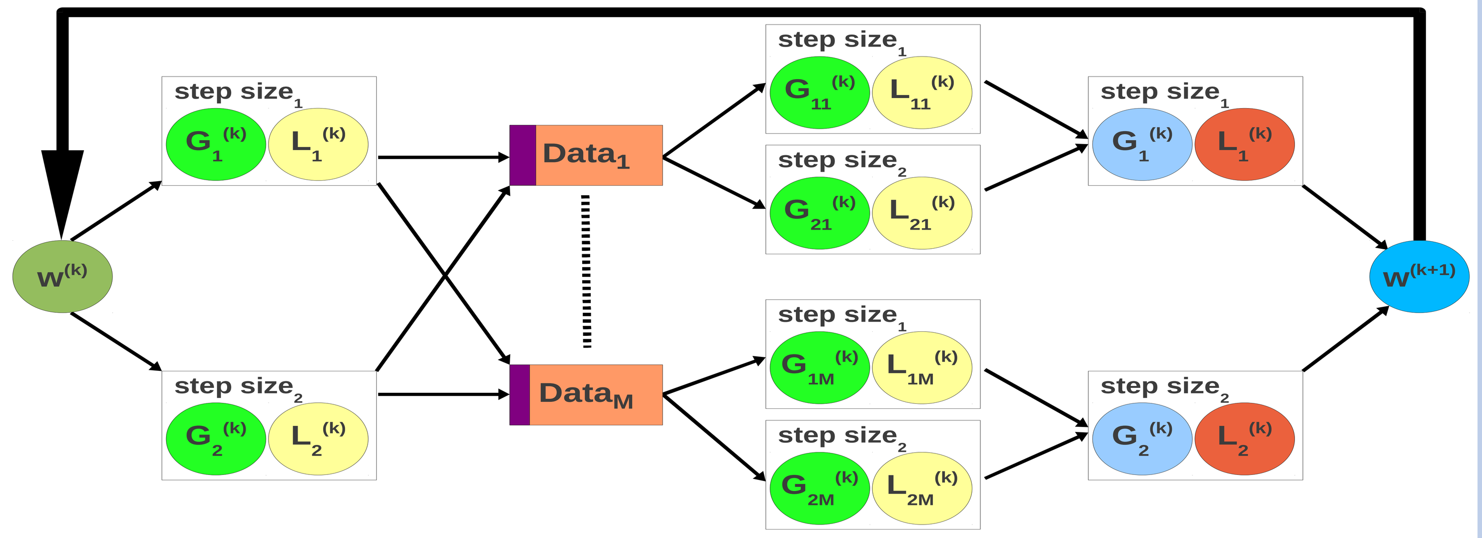
$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \alpha^{(k)} \nabla f_i\left(\vec{w}^{(k)}\right)$$

- The hyper-parameter step-size α is critical to convergence rate. In all the systems α has to be hand-tuned by the user for different training tasks and datasets combinations.

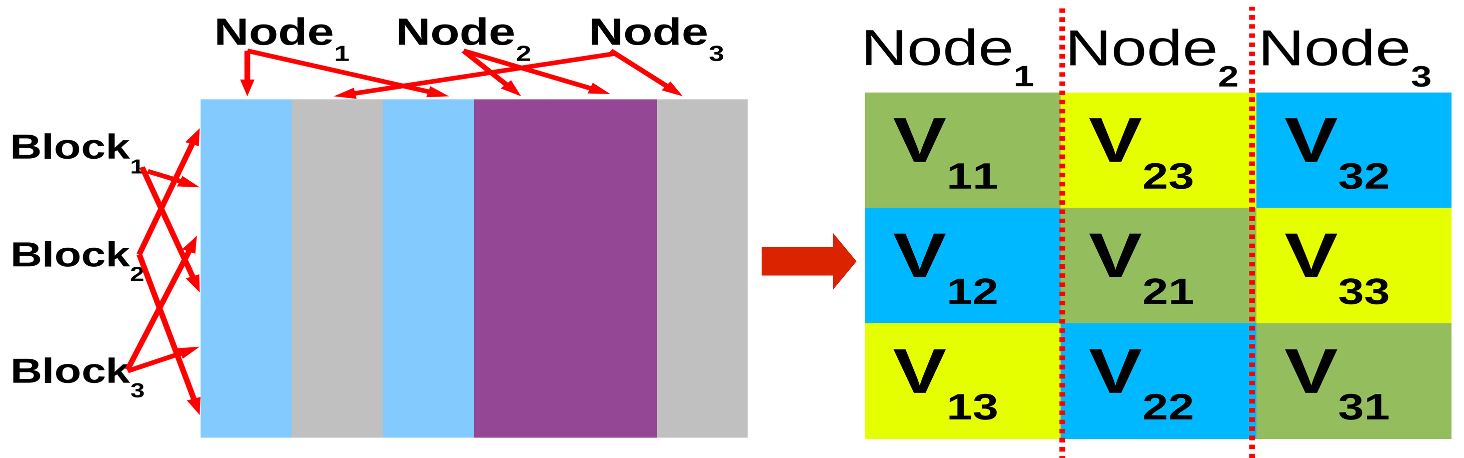
GLADE Execution Flow



On-the-fly Concurrent Parameter Testing



Data-model-mixed Parallel Low-rank Matrix Factorization (LMF)



Data Partition and Sub-iterations (V_{11}, V_{12}, V_{13} are sub-iterations of $Node_1$.)

Datasets

Dataset	Dimension	# Examples	Size
`splice`	13M	50M	3TB
`matrix`	1M × 1M	1.7B	36GB
`classify`	200	50M	136 GB

Linear Speed-up Experiment

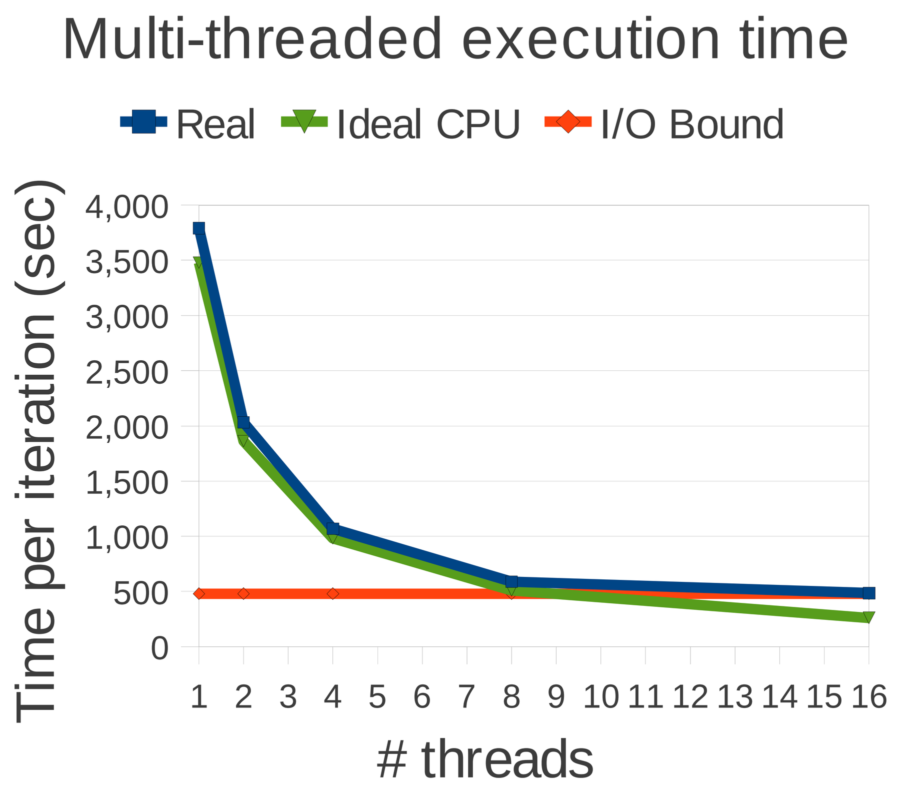


Figure: Multi-threaded LMF on `matrix`

Data-model-mixed Parallel LMF

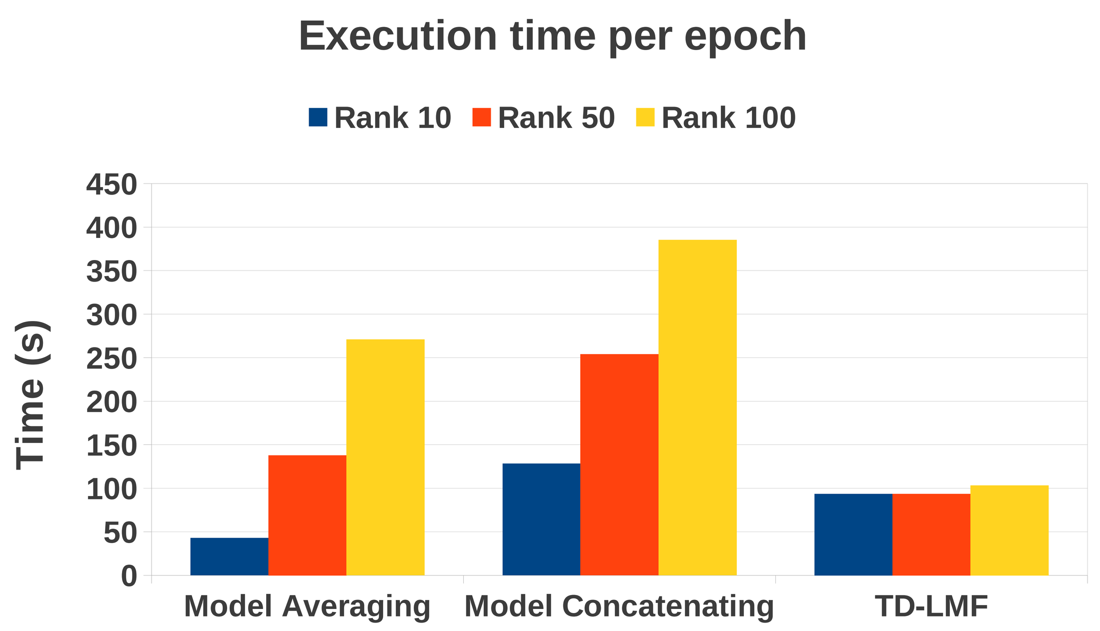


Figure: Time per iteration comparison on `matrix`

Results For On-the-fly Parameter Testing

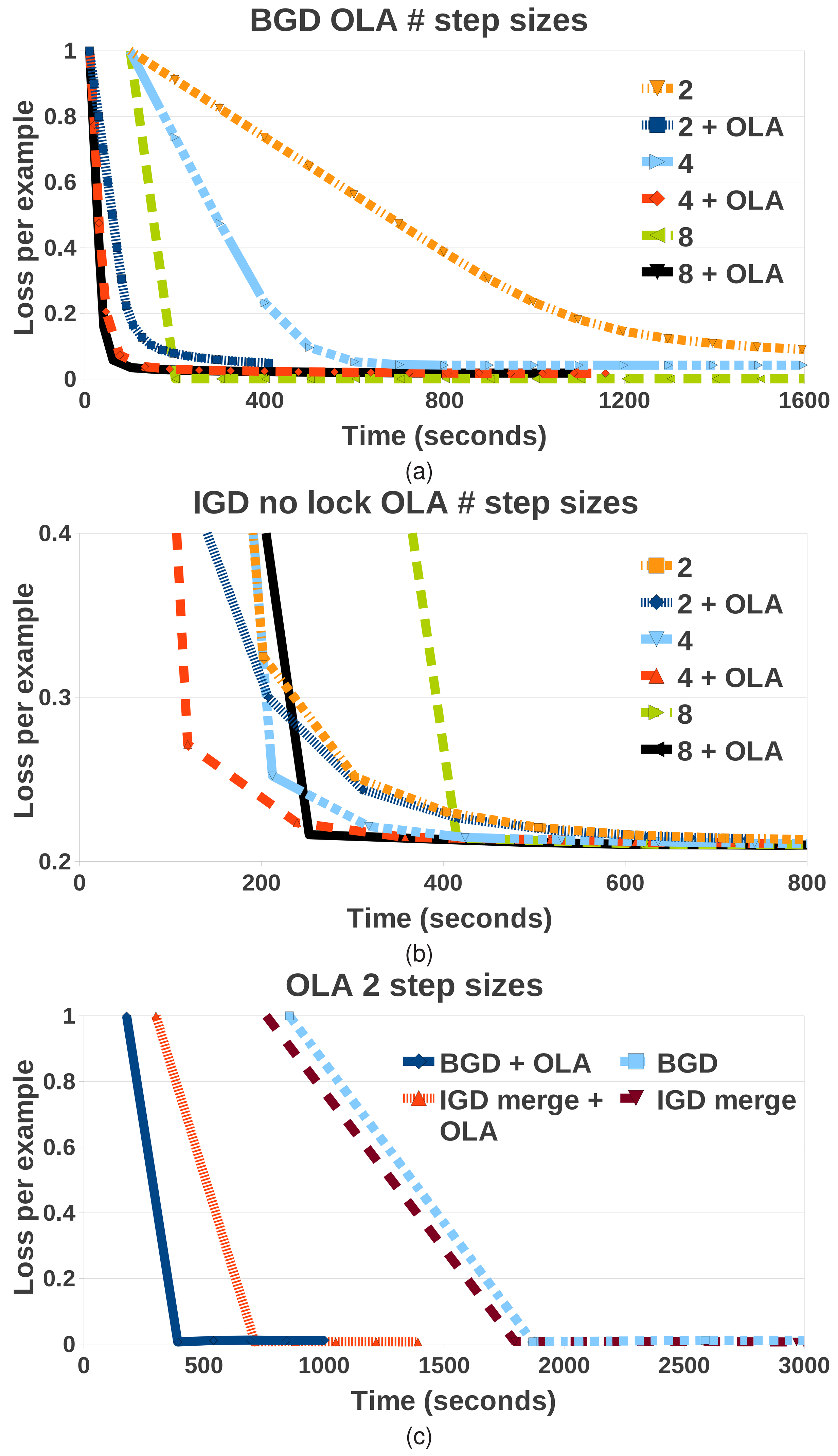


Figure: (a) Support Vector Machine (SVM) with BGD (`classify`). (b) Logistic Regression (LR) with IGD (`classify`). (c) SVM (`splice`).

University of California, Merced

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