SF2568 - Project Proposal

Dan Vicente (danvi@kth.se) & Erik Lindé (elinde2@kth.se)

February 2024

Project Summary

- **Title:** Deep learning with Elastic Averaging Stochastic Gradient Descent for image classification
- **Description:** Implementation of the Elastic Averaging Stochastic Gradient Descent algorithm (EASGD) for stochastic optimization of a deep convolutional neural network trained on labelled image data. The description of the algorithm, some previous results etc are found in the original paper at: `https://arxiv.org/abs/1412.6651`.
- **Applied Methods:** We will apply the EASGD algorithm to train a deep CNN for image classification. The intention is to leverage the frontend from existing libraries for creating the neural network architecture in an object oriented manner (we define a class for the CNN in `c++`). We then use the methods from the course (communication of workers with `MPI`) for applying parallelised training of the neural network. The learning rate $\eta > 0$, moving rate $\alpha > 0$ and communication period $\tau \in \mathbb{N}$ are hyperparameters, which we intend to tune empirically (e.g using cross-validation). The parallelisation is done by defining a central variable which each process updates. The communication period τ controls how often the communication between the master process and the workers is done. We will also implement a sequential stochastic gradient descent and compare the timings of the parallel training contra the sequential training. Moreover, we will try to get empirical results on speedup and efficiency.
- **Software tools used:** `pytorch` for `c++` (`https://pytorch.org/tutorials/advanced/cpp_frontend.html`) for defining the convolutional network class. `MPI` for communication of workers. Mathematical packages in `c++`. We will not use any parallel libraries (other than `MPI.H`).
- **Data used:** `CIFAR-10` image dataset. `https://www.cs.toronto.edu/~kriz/cifar.html`.