Overview

Goal: Fixing the communication bottleneck for distributed optimization in supervised ML.

Contributions

COCOA+ is a primal-dual framework for distributed optimization
- efficient additive aggregation of local updates
- strong convergence guarantees
- framework: guarantees for arbitrary local solvers
- significant practical speedup

Main Idea

Propose a local subproblem to allow additive, data dependent aggregation

\[
A_1\theta(x) + A_2\theta(x) + \cdots + A_K\theta(x) = w + \sum_{k=1}^{K} \Delta w_k
\]

\[
\Delta w_k = \frac{1}{n} \sum_{i=1}^{n} \ell_i(x_i)
\]

Convergence results

Local $\Theta$-Approximation

For $\Theta \in (0, 1)$, we assume the local solver finds a (possibly) randomized approximate solution satisfying:

\[
E[\Theta(\Delta w_k) - \Theta(\Delta w_k) \leq \Theta(\Delta w_k) - \Theta(0)]
\]

Theorem. Let $\ell_i(\cdot)$ be $L$-Lipschitz

Obtain suboptimality $\epsilon$, after $T$ iterations, with:

- CoCoA, averaging $\gamma = 1/K$
  \[
  T \geq \Theta \left( \frac{K}{1 - \Theta} \left( \frac{SL^2}{\lambda \mu} + \delta \right) \right)
  \]

- CoCoA+, adding $\gamma = 1$
  \[
  T \geq \Theta \left( \frac{1}{1 - \Theta} \left( \frac{SL^2}{\lambda \mu} + \delta \right) \right)
  \]

Local Subproblem

\[
\max_{\Delta w_k} \Theta(\Delta w_k) \leq \frac{1}{n} \sum_{i=1}^{n} \ell_i(x_i)
\]

Flexible: can use arbitrary local solver

Setup

Primal problem formulation

\[
\min_{w \in \mathbb{R}^d} \frac{1}{2} \|w\|^2 + \frac{1}{n} \sum_{i=1}^{n} \ell_i(w^T x_i)
\]

Data partitioned by examples $A_i = x_i$

Dual problem

\[
\max_{\alpha \in \mathbb{R}^n} -\frac{1}{2} \|\Delta w\|^2 - \frac{1}{n} \sum_{i=1}^{n} \ell_i(x_i)
\]

Information: local shared

CoCoA+ Framework

Input: aggregation parameter $\gamma \in (0, 1]$
subproblem parameter $\gamma$
Objective: $\frac{1}{n} \sum_{i=1}^{n} \ell_i(x_i)$

Theorem. Let $\ell_i(\cdot)$ be $1/\mu$-smooth

Obtain suboptimality $\epsilon$, after $T$ iterations, with:

- CoCoA, averaging $\gamma = 1/K$
  \[
  T \geq \Theta \left( \frac{K}{1 - \Theta} \left( \frac{SL^2}{\lambda \mu} + \delta \right) \right)
  \]

- CoCoA+, adding $\gamma = 1$
  \[
  T \geq \Theta \left( \frac{1}{1 - \Theta} \left( \frac{SL^2}{\lambda \mu} + \delta \right) \right)
  \]

Information: local shared

References