3. Goal

SGD type of methods are often seen as sampling rows of a data matrix. Conversely, CD methods usually sample columns of the data matrix.

The aim of this work is to develop a hybrid of S2GD and CD, which efficiently samples both rows and columns of data. The method computes a stochastic estimate of the partial gradient $\nabla_{ij} f(x)$ with variance diminishing property and updates only one coordinate at each iteration.

4. The S2CD algorithm [1]

**S2CD Algorithm — Semi-Stochastic Coordinate Descent**

**parameters:** m (max # of stochastic steps per epoch); h > 0 (stepsize parameter); $x_0 \in \mathbb{R}^d$ (starting point); $\beta = \sum_{m=1}^{n} (1 - \mu h)^{m-1}$.

for $k = 0, 1, 2, \ldots$ do

1. Compute and store $\nabla f(x_k)$ where $f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x_i)$.
2. Update: $x_{k+1} = x_k - h \nabla f(x_k)$

end for

The total work of S2CD can be written as

$$\mathcal{O} \left( \left( \frac{n \mathcal{C}_{grad} + \kappa \mathcal{C}_{pd}}{\epsilon} \right) \log \left(\frac{1}{\epsilon}\right) \right),$$

where $\kappa = \frac{L}{\mu}$ and either $L = L_{max}$ ([4, 2, 3, 6]), or $L = \mathcal{L}_{avg} = \frac{L_{max}}{n \sum_{i=1}^{n} \mathcal{L}_{ij} (\mathcal{S})}$ [5].

The difference between our result and existing results is in the term $\kappa_{avg}$ — previous results have $\kappa_{avg}$ in that place. This difference constitutes a trade-off: while $\kappa \geq \kappa_{avg}$, we clearly have $\mathcal{C}_{pd} \leq \mathcal{C}_{avg}$. The comparison of the quantities $\kappa_{avg}$ and $\kappa_{pd}$ is not straightforward and is problem dependent.

Conclusion

S2CD can be both better or worse than S2GD/SVRG/SAGA, depending on whether the increase of the condition number from $\kappa$ to $\kappa_{avg}$ can or can not be compensated by the decrease of the derivative evaluation from $\mathcal{C}_{grad}$ to $\mathcal{C}_{pd}$.

6. References