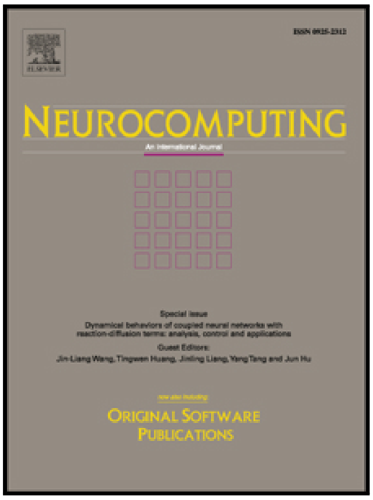
Communicated by Dr Yiming Ying

Accepted Manuscript



On the Flexibility of Block Coordinate Descent for Large-Scale Optimization

Xiangfeng Wang, Wenjie Zhang, Junchi Yan, Xiaoming Yuan, Hongyuan Zha

PII: S0925-2312(17)31270-5
DOI: 10.1016/j.neucom.2017.07.024
Reference: NEUCOM 18712

To appear in: *Neurocomputing*

Received date: 19 January 2017
Revised date: 30 April 2017
Accepted date: 10 July 2017

Please cite this article as: Xiangfeng Wang, Wenjie Zhang, Junchi Yan, Xiaoming Yuan, Hongyuan Zha, On the Flexibility of Block Coordinate Descent for Large-Scale Optimization, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.07.024

On the Flexibility of Block Coordinate Descent for Large-Scale Optimization

Xiangfeng Wang[1], Wenjie Zhang[1], Junchi Yan[1,2], Xiaoming Yuan[3] and Hongyuan Zha[1,4]

[1] *Shanghai Key Lab for Trustworthy Computing, School of Computer Science and Software Engineering, East China Normal University, Shanghai, China*
[2] *IBM Research -- China*
[3] *Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*
[4] *School of Computing, Georgia Institute of Technology, Atlante, USA*

Abstract

We consider a large-scale minimization problem (not necessarily convex) with non-smooth separable convex penalty. Problems in this form widely arise in many modern large-scale machine learning and signal processing applications. In this paper, we present a new perspective towards the parallel Block Coordinate Descent (BCD) methods. Specifically we explicitly give a concept of so-called two-layered block variable updating loop for parallel BCD methods in modern computing environment comprised of multiple distributed computing nodes. The outer loop refers to the block variable updating assigned to distributed nodes, and the inner loop involves the updating step inside each node. Each loop allows to adopt either Jacobi or Gauss-Seidel update rule. In particular, we give detailed theoretical convergence analysis to two practical schemes: Jacobi/Gauss-Seidel and Gauss-Seidel/Jacobi that embodies two algorithms respectively. Our new perspective and behind theoretical results help devise parallel BCD algorithms in a principled fashion, which in turn lend them a flexible implementation for BCD methods suited to the parallel computing environment. The effectiveness of the algorithm framework is verified on the benchmark tasks of large-scale ℓ_1 regularized sparse logistic regression and non-negative matrix factorization.

Keywords: Block Coordinate Descent, Large-Scale Optimization, Jacobi, Gauss-Seidel