Numerical Study of Sparse Optimization Methods for PageRank Problem

Anton Anikin

Matrosov Institute for System Dynamics and Control Theory, SB RAS, Russia
anton.anikin@htower.ru

Keywords: huge-scale optimization, sparsity, Frank–Wolfe conditional gradient method, coordinate descent method, direct gradient method in L1–norm.

PargeRank vector search is one of the most famous modern huge-scale problem. Initially, it is the task of finding of eigenvector corresponding to eigenvalue 1 of the (column) stochastic matrix: $Px = x$, where P is web-graph sparse matrix which element P_{ij} is non-zero only if page j (graph node) cites page i. The original PageRank problem can be reduced to the optimization form in different ways, in the paper we discuss the following version:

$$f(x) = \frac{1}{2}\|Ax\|_2^2 \to \min_{\langle x,e\rangle=1}, \tag{1}$$

where $A = P^T - I$, I is identity matrix, $e = (1, \ldots, 1)^T$, $x \in \mathbb{R}_+^n$.

The most interesting and important feature of PageRank problem is the sparsity of solution. For "close-to-reality" web-graphs (i.e. non-synthetic cases) we can find almost sparse vector x which will provide solution with required accuracy. This great feature allows us to use sparse optimization methods which based on ideas of Yuri Nesterov [**?**]. Such methods provides ideology of "lightweight iterations" which modifies only 1-2 variables of x and then performs *update* of function value instead it's traditional full re-calculation. Such update technique explicitly uses problem's internal structure and allows significantly reduce the cost of iteration [**?**].

The paper presents results of numerical study of problem (1) with number of sparse optimization methods: direct gradient method in L1–norm, Frank–Wolfe conditional gradient method, deterministic and randomized variants of coordinate descent methods. The methods behavior and properties was studied with different simulated and real data sets. The results of numerical experiments confirmed the efficiency of proposed approaches.

The research was partly supported by the RFBR, project no. 16-07-00664.