Research Highlight: A highly efficient method for solving Lasso problems

Work of Dr LI Xudong, Professor SUN Defeng and Professor TOH Kim Chuan

The Big Data era brings new challenges in analyzing massive data having a large number of samples and/or high dimensionality. In order to respond to these challenges, l_1-regularized least squares regression models (Lasso) are intensively studied. Despite the fact that there exist many solvers in the literature for the Lasso regularized regression problems, there are currently no robust solvers that can efficiently handle difficult large scale regression problems with real data. Professors Sun and Toh and their former PhD student, Dr. Li Xudong, have developed an extremely fast and robust semismooth Newton augmented Lagrangian algorithm (SSNAL) to solve the Lasso problems. While the semismooth Newton method would normally be considered as computationally expensive, they showed that for Lasso problems with sparse optimal solutions, these computational costs can be extremely cheap. They achieve such a counter-intuitive success by carefully analyzing and exploiting the second order sparsity structure present in the underlying problems. For challenging Lasso problems with large scale real datasets, their algorithm can be over 100 times faster than the state-of-the-art solvers. For example, for a problem with over 4 million features and 16000 samples, SSNAL can solve it in 20 seconds, while the best alternative solver took 2400 seconds.

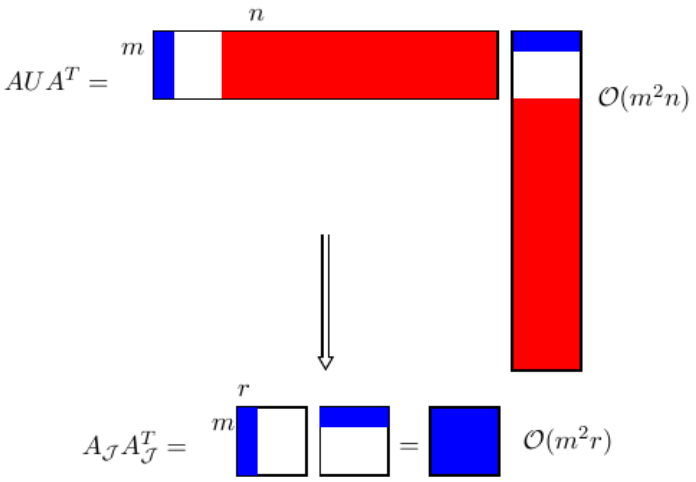

Figure 1: Reducing the computational costs from $\mathcal{O}(m^2n)$ to $\mathcal{O}(m^2r)$

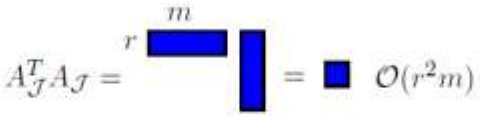

Figure 2: Further reducing the computational costs to $\mathcal{O}(r^2m)$

Reference:

X. D. Li, D. F. Sun, K.-C. Toh, “A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems”, arXiv:1607.05428, 2016.