FaLRR: A Fast Low Rank Representation Solver

Shijie Xiao†, Wen Li†, Dong Xu†, Dacheng Tao‡
†School of Computer Engineering, Nanyang Technological University, Singapore. ‡Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia.

In this paper, we develop a fast solver of low rank representation (LRR) [3] called FaLRR, which achieves order-of-magnitude speedup over existing LRR solvers, and is theoretically guaranteed to obtain a global optimum.

LRR [3] has shown promising performance for various computer vision applications such as face clustering. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ be a set of data samples drawn from a union of several subspaces, where d is the feature dimension and n is the total number of data samples. LRR seeks a low-rank data representation matrix $\mathbf{Z} \in \mathbb{R}^{n \times n}$ such that $\mathbf{X}$ can be self-expressed (*i.e.*, $\mathbf{X} = \mathbf{XZ}$) when the data is clean. Considering that input data may contain outliers (*i.e.*, some columns of $\mathbf{X}$ are corrupted), the LRR problem can be formulated as,

$$\min_{\mathbf{Z},\mathbf{E}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \quad \text{s.t.} \ \ \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \tag{1}$$

where λ is a tradeoff parameter and $\mathbf{E} \in \mathbb{R}^{d \times n}$ denotes the representation error. The nuclear norm based term $\|\mathbf{Z}\|_*$ acts as an approximation of the rank regularizer, and the $\ell_{2,1}$ norm based term $\|\mathbf{E}\|_{2,1}$ encourages $\mathbf{E}$ to be column-sparse.

Regarding optimization, several algorithms [2, 3, 4] were proposed to exactly solve LRR. Moreover, to efficiently obtain an approximated solution of LRR, a distributed framework [5] was developed. However, the existing algorithms are usually based on the original formulation in (1) or a similar variant [4], which are two-variable problems with regard to the original data matrix. In this paper, we develop a fast LRR solver named FaLRR, which is based on a new reformulation of LRR as an optimization problem with regard to factorized data (which is obtained by skinny SVD on the original data matrix).

Reformulation. Specifically, we study a more general formulation of LRR as follows,

$$\min_{\mathbf{Z} \in \mathbb{R}^{n \times m}, \mathbf{E} \in \mathbb{R}^{d \times m}} \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \quad \text{s.t.} \ \ \mathbf{XD} = \mathbf{XZ} + \mathbf{E} \tag{2}$$

which includes (1) as a special case. Let r denote the rank of $\mathbf{X}$. Moreover, let us factorize $\mathbf{X}$ via the *skinny* singular value decomposition (SVD): $\mathbf{X} = \mathbf{U}_r \mathbf{S}_r \mathbf{V}_r'$, where $\mathbf{U}_r \in \mathbb{R}^{d \times r}$ and $\mathbf{V}_r \in \mathbb{R}^{n \times r}$ are two column-wise orthogonal matrices that satisfy $\mathbf{U}_r'\mathbf{U}_r = \mathbf{V}_r'\mathbf{V}_r = \mathbf{I}_r$, $\mathbf{S}_r \in \mathbb{R}^{r \times r}$ is a diagonal matrix defined as $\mathbf{S}_r = diag([\sigma_1, \ldots, \sigma_r]')$, in which $\{\sigma_i\}_{i=1}^r$ are the r positive singular values of $\mathbf{X}$ sorted in descending order. Based on the definitions above, we present the reformulation by the following theorem:

Theorem 1 *Let $\mathbf{W}^*$ denote an optimal solution of the following problem,*

$$\min_{\mathbf{W} \in \mathbb{R}^{r \times m}} \|\mathbf{W}\|_* + \lambda \|\mathbf{S}_r(\mathbf{V}_r'\mathbf{D} - \mathbf{W})\|_{2,1} \ . \tag{3}$$

Then, $\{\mathbf{Z}^*, \mathbf{E}^*\}$, defined as $\mathbf{Z}^* = \mathbf{V}_r\mathbf{W}^*$ and $\mathbf{E}^* = \mathbf{XD} - \mathbf{XV}_r\mathbf{W}^*$, is an optimal solution of the problem in (2). In particular, $\|\mathbf{Z}^*\|_* = \|\mathbf{W}^*\|_*$ and $\|\mathbf{E}^*\|_{2,1} = \|\mathbf{S}_r(\mathbf{V}_r'\mathbf{D} - \mathbf{W}^*)\|_{2,1}$ always hold, implying that the two problems in (2) and (3) have equal optimal objective values.

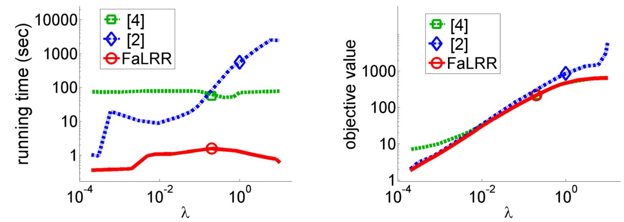
Optimization. In terms of optimization, we rewrite the problem in (3) as follows by introducing another variable $\mathbf{Q} \in \mathbb{R}^{r \times m}$:

$$\min_{\mathbf{W},\mathbf{Q} \in \mathbb{R}^{r \times m}} \ \|\mathbf{W}\|_* + \lambda \|\mathbf{S}_r\mathbf{Q}\|_{2,1} \quad \text{s.t.} \ \ \mathbf{W} + \mathbf{Q} = \mathbf{V}_r'\mathbf{D}, \tag{4}$$

and develop an efficient algorithm based on the alternating direction method (ADM) [1, 2], in which both resultant subproblems can be solved exactly. The corresponding augmented Lagrangian [1] *w.r.t.* (4) is

$$\begin{aligned} & \mathcal{L}_\rho(\mathbf{W}, \mathbf{Q}, \mathbf{L}) \\ = \ & \|\mathbf{W}\|_* + \lambda \|\mathbf{S}_r\mathbf{Q}\|_{2,1} + \langle \mathbf{L}, \mathbf{V}_r'\mathbf{D} - \mathbf{W} - \mathbf{Q} \rangle + \frac{\rho}{2} \|\mathbf{V}_r'\mathbf{D} - \mathbf{W} - \mathbf{Q}\|_F^2, \end{aligned}$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) (b)

Figure 1: (a) the running time *w.r.t.* λ and (b) the resultant objective value *w.r.t.* λ, for solving LRR on the ExtYaleB dataset. The positions of markers indicate the optimal parameters for the three LRR solvers, respectively.

where $\mathbf{L} \in \mathbb{R}^{r \times m}$ is the Lagrangian multiplier and $\rho > 0$ is the penalty parameter. By employing ADM, we iteratively update the variables $\{\mathbf{W}, \mathbf{Q}\}$, the Lagrange multiplier $\mathbf{L}$ and the penalty parameter ρ until convergence.

In particular, the subproblem for updating $\mathbf{W}$ is in the form of

$$\min_{\mathbf{W} \in \mathbb{R}^{r \times m}} \|\mathbf{W}\|_* + \frac{\rho}{2} \|\mathbf{W} - \mathbf{G}\|_F^2 \, ,$$

where $\mathbf{G} \in \mathbb{R}^{r \times m}$ is constant *w.r.t.* $\mathbf{W}$. To efficiently solve this subproblem, we propose a *tentative* strategy, which is motivated by our experimental observations and theoretical analysis.

On the other hand, the subproblem for updating $\mathbf{Q}$ is in the form of

$$\min_{\mathbf{Q} \in \mathbb{R}^{r \times m}} \lambda \|\mathbf{S}_r\mathbf{Q}\|_{2,1} + \frac{\rho}{2} \|\mathbf{Q} - \mathbf{C}\|_F^2 \, ,$$

where $\mathbf{C} \in \mathbb{R}^{r \times m}$ is constant *w.r.t.* $\mathbf{Q}$. We show that such problem can be efficiently solved with $\mathcal{O}(rm)$ complexity.

Overall, the total time complexity of each iteration for our algorithm is $\mathcal{O}(rm\min(r,m) + rm)$. Particularly, for solving the LRR problem in (1) where $m = n$ and $r \leq n$, our time complexity per iteration is $\mathcal{O}(nr^2 + nr)$. Moreover, we observe that the total number of iterations of our algorithm is often relatively small in our experiments.

Incorporation into a distributed framework Our algorithm can be readily incorporated in the distributed framework [5] called DFC-LRR, to further improve the efficiency.

Experiments Extensive experiments on synthetic and real-world datasets demonstrate that our FaLRR achieves order-of-magnitude speedup over existing LRR solvers. In Figure 1, we take the ExtYaleB dataset as an example to compare our FaLRR with the LRR solvers in [4] and [2] in terms of the running time and the resultant objective value. Moreover, the efficiency can be further improved by incorporating our algorithm into DFC-LRR.

[1] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[2] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *NIPS*, pages 612–620, 2011.

[3] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010.

[4] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *T-PAMI*, 35(1):171–184, 2013.

[5] Ameet Talwalkar, Lester Mackey, Yadong Mu, Shih-Fu Chang, and Michael I Jordan. Distributed low-rank subspace segmentation. In *ICCV*, pages 3543–3550, 2013.