Elastic-Net Regularization of Singular Values for Robust Subspace Learning

Eunwoo Kim[1], Minsik Lee[2], and Songhwai Oh[1]
[1]Department of ECE, ASRI, Seoul National University, Seoul, Korea. [2]Division of EE, Hanyang University, Korea.

Low-rank matrix approximation has attracted much attention in the areas of data reconstruction, image denoising, collaborative filtering, background modeling, structure from motion, and photometric stereo [1, 2, 3, 4, 6], to name a few. It is sometimes assumed that the rank of a matrix is fixed or known beforehand and it is also known as a subspace learning problem.

We consider the low-rank matrix and sparse matrix separation problem [1] based on convex envelopes of rank and sparsity functions:

$$\min_{P,X} \|W \odot (Y - PX)\|_1 + \lambda \|PX\|_*, \tag{1}$$

where Y is an observation matrix, and λ is a pre-defined weighting parameter. $\|\cdot\|_1$ and $\|\cdot\|_*$ denote the entry-wise l_1-norm and the nuclear-norm, which are convex relaxation of the l_0-norm and the rank function, respectively. Here, $\odot$ is the component-wise multiplication or the Hadamard product and W is an indicator matrix, Generally, (1) is a non-convex and non-smooth problem, making it difficult to find a solution efficiently and exactly. To solve the problem efficiently, a common strategy is to use an alternating minimization approach which solves for one variable while other variables are fixed [2].

Note that the regularization term in (1), $\|PX\|_*$, can be interpreted as a sum of singular values, $\sum_i^r |\sigma_i|$, where σ_i is the ith singular value of a low-rank matrix PX and r is the rank of PX. It leads to a lasso problem [5], which has a thresholding effect over singular values. But, lasso-based approaches lack a shrinkage effect due to their weak convexity, which makes the algorithm unstable when highly corrupted data are presented. To improve the stability of the algorithm, we introduce a strong convex regularizer over singular values with the l_2-norm penalty of singular values, $\lambda_1 \sum_i^r |\sigma_i| + \frac{\lambda_2}{2} \sum_i^r |\sigma_i|^2$. Based on the fact that $\|D\|_F^2 = \mathrm{tr}(V\Sigma U^T U \Sigma V^T) = \mathrm{tr}(\Sigma^2) = \sum_i |\sigma_i|^2$, where $D = U\Sigma V^T$ is SVD of D, we introduce a new penalized optimization problem as follows:

$$\min_{P,X} \|W \odot (Y - PX)\|_1 + \lambda_1 \|PX\|_* + \frac{\lambda_2}{2} \|PX\|_F^2. \tag{2}$$

In (2), we have elastic-net regularization of singular values, which has shown its superiority compared to lasso [5]. Our method is a holistic approach which utilizes both nuclear-norm minimization and bilinear factorization. It is capable of stabilizing a lasso-type method due to its strong convexity, owing to the Frobenius norm [7]. In addition, we have both a thresholding effect over singular values from the l_1 regularizer and a shrinkage effect from the l_2 regularizer to make a parsimonious and stable model.

Unfortunately, (2) can suffer from heavy computational complexity for large-scale problems because of an SVD operation at each iteration to solve a nuclear-norm based cost function. To solve (2) in practice, the alternative form of the nuclear-norm, $\|D\|_* = \min_{D=PX} \frac{1}{2}\left(\|P\|_F^2 + \|X\|_F^2\right)$, is utilized. Using the alternate form, we make an equivalent form of (2) by introducing auxiliary variable $\widehat{D}$ and solve the following problem:

$$\begin{aligned} \min_{P,X,D,\widehat{D}} \; & \|W \odot (Y - \widehat{D})\|_1 + \frac{\lambda_1}{2}\left(\|P\|_F^2 + \|X\|_F^2\right) + \frac{\lambda_2}{2}\|D\|_F^2 \\ \text{s.t.} \quad & D = PX, \;\; \widehat{D} = D, \end{aligned} \tag{3}$$

and the problem (3) is solved under the augmented Lagrangian framework. Based on the previous analysis, we derive a robust matrix factorization algorithm called *factEN*. It is a fast algorithm which has $O(mnr)$ complexity and is not sensitive to initial values.

Figure 1 shows results of the proposed method compared to Unifying [1], a lasso-based method, and ground-truth on a simple example (100×100) with 20% outliers. From the figure, the proposed method gives a stable

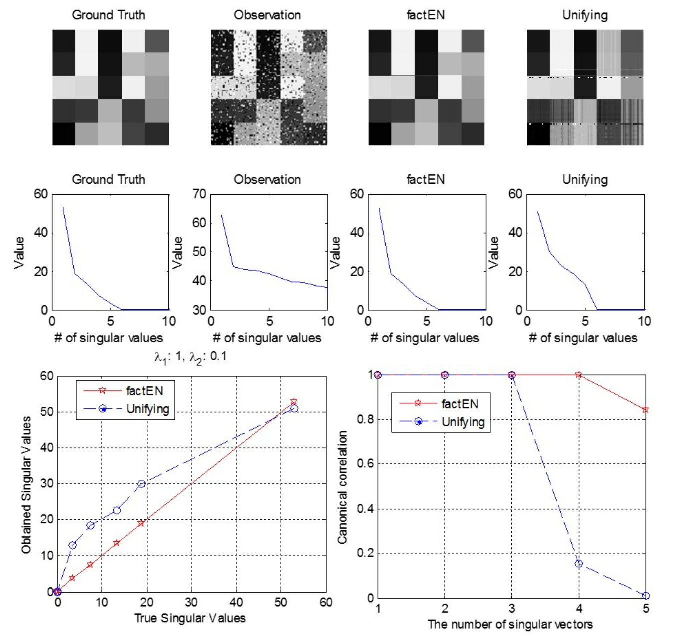

Figure 1: Evaluation of the proposed method (factEN) and a lasso-based method (Unifying [1]) for a toy example. Bottom left shows ordered singular values recovered by two algorithms against true singular values. Bottom right shows canonical correlations between singular vectors from applied algorithms and true singular vectors.

result against outliers and eliminates noises by suppressing the singular values, whereas Unifying finds relatively inaccurate and higher singular values and shows a poor reconstruction result compared to the proposed method.

We also analyze the convergence property of the proposed method. We provide a proof of weak convergence of factEN by showing that under mild conditions any limit point of the iteration sequence generated by the algorithm is a stationary point that satisfies the Karush-Kuhn-Tucker (KKT) conditions. It is worth proving that any converging point must be a point that satisfies the KKT conditions because they are necessary conditions to be a local optimal solution. This result provides an assurance about the behavior of the proposed algorithm.

In experiments, the proposed method is applied to a number of low-rank matrix approximation problems including synthetic examples, non-rigid motion estimation, photometric stereo, and background subtraction to demonstrate its efficiency in the presence of heavy corruptions. The experimental results show that the proposed method outperforms other existing methods in terms of the approximation error and execution time and is stable against missing data, outliers, and different parameter values.

[1] Ricardo Cabral, Fernando De la Torre, Joao P. Costeira, and Alexandre Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *Proc. ICCV*, 2013.

[2] Q. Ke and T. Kanade. Robust l_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *Proc. CVPR*, 2005.

[3] E. Kim, M. Lee, C.-H. Choi, N. Kwak, and S. Oh. Efficient l_1-norm-based low-rank matrix approximations for large-scale problems using alternating rectified gradient method. *IEEE T-NNLS*, 26(2):237–251, 2015.

[4] N. Srebro. Weighted low-rank approximations. In *Proc. ICML*, 2003.

[5] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[6] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust l_1-norm. In *CVPR*, 2012.

[7] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.

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