Elastic-Net Regularization of Singular Values for Robust Subspace Learning

Eunwoo Kim¹, Minsik Lee², and Songhwa Oh¹
¹Department of ECE, ASRI, Seoul National University, Seoul, Korea. ²Division 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:

$$\min_{P_X} \|W \odot (Y - PX)\|_1 + \lambda_1 \|PX\|_r, \quad (1)$$

where $Y$ is an observation matrix, and $\lambda$ is a pre-defined weighting parameter. $\| \cdot \|_1$ and $\| \cdot \|_2$ 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), $\|PX\|_r$, can be interpreted as a sum of singular values, $\sum_i |\sigma_i|$, where $\sigma_i$ is the $i$th singular value of a low-rank matrix $P_X$ and $r$ is the rank of $P_X$. 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.

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.

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, $\|D\|_F = \text{tr}(V \Sigma U^T U \Sigma V^T) = \text{tr}(\Sigma^2) = \sum_i |\sigma_i|^2$, where $D = U \Sigma V^T$ is SVD of $D$, we introduce a new penalized optimization problem as follows:

$$\min_{P_X} \|W \odot (Y - PX)\|_1 + \lambda_1 \|PX\|_r + \frac{\lambda_2}{2} \|PX\|_F. \quad (2)$$

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.

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 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.


