OptShrink Algorithm and Its Applications to Image Denoising by Low-Rank Approximation

Szu-Yun Lin
Dept. of Civil Engineering
University of Michigan
Ann Arbor, MI 48105
sylin@umich.edu

Yizhi Liu
Dept. of Civil Engineering
University of Michigan
Ann Arbor, MI 48105
liuyiz@umich.edu

Xi Wang
Dept. of Civil Engineering
University of Michigan
Ann Arbor, MI 48105
wangix@umich.edu

Abstract

Approaches for signal matrix denoising from a signal-plus-noise matrix are applied to many fields including statistical signal processing, computer vision, data mining, and bioinformatics. In this project, the authors aim to show how the OptShrink algorithm can be used to improve the denoising performance. Two applications are presented. Firstly, the performance of the OptShrink algorithm and the truncated singular value decomposition algorithm in low-rank image recovery is compared. Secondly, the authors show the potential of the OptShrink method for improving the performance of existing methods by plugging the OptShrink algorithm into the Nearest Subspace classification method, and the results show that the advantage of the improved algorithm is significant.

1 Problem statement

Nowadays, image pattern recognition and classification are playing an increasingly important role in various fields including remote sensing, criminal investigation and bioinformatics [1] [2]. Images acquired in these practical settings usually contain noises, which greatly reduce the accuracy of the application of image pattern recognition and classification. Therefore, it is indispensable to implement image denoising before applying it to pattern recognition and classification in practice.

Meanwhile, the images being analyzed often contains a large number of features, attributes or characteristics. Processing these data require a high volume of space and computational resources. As a consequence, an approach to save computational resource by selecting or extracting important information out of the image data is in need.

Fortunately, it is found that by reducing the rank of the image, the important information in the images could be retained as the useful message for future analysis and trivial information such as noise could be excluded. This causes the rise of low-rank image modeling researches in the data science field.

Mathematically, the assumption of low dimensionality is: $\operatorname{rank}(\mathbf{X}) \ll \min(n, m)$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ denotes the entire dataset and vectors $\mathbf{x}_i \in \mathbb{R}^m$. A typical example is an application in signal processing [3], in that $\mathbf{x}_i$ represents a vector of signal intensities received by an antenna array at time point i. However, the raw experiential data is barely perfectly translated into the low-rank approximation in reality. Thus, it is often expressed as $\mathbf{X} = \mathbf{S} + \mathbf{N}$ in the models to be close to the real situations, where $\mathbf{S}$ represents the low-rank signal component and $\mathbf{N}$ denotes a corresponding component of the noise or error in measurements. The techniques for low-rank signal matrix extraction from a signal-plus-noise matrix are necessary and appear prominently in many fields. In this article, we focus on the applications in image denoising.

2 Related work

A conventional approach to form an estimate of the low-rank signal matrix is the truncated singular value decomposition (tSVD) which is based on the famous Eckart-Young-Mirsky matrix approximation theorem [4]-[6]. It is to solve the following optimization problem:

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{S}\|_F, \quad \mathrm{s.t.}\,\mathrm{rank}(\mathbf{S}) \leq \mathrm{rank}(\mathbf{X}) = \mathrm{r}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and the optimal solution (or said EYM estimator) is given by $\sum_{i=1}^{k} \sigma'_i \mathbf{u}'_i \mathbf{v}'^{\mathrm{H}}_i$, where $\sigma'_i, \mathbf{u}'_i, \mathbf{v}'_i, \forall\, i = 1, \cdots, k$ are the singular values, left singular vectors, and right singular vectors of X. This procedure also corresponds to Principal Component Analysis (PCA) in statistics.

In the past decades, many extended algorithms and new methods for low-rank matrix recovery have been proposed. Most of the extended studies related to the EYM estimator were committed to improving estimation performance by adding additional exploitable structure in the low-rank signal matrix, such as the structured sparse PCA [7]-[9], Toeplitz structured [10], Hankel structured [11], and nonnegative structured [12]. On the other hand, some studies applied a weighted approximation to improve EYM estimator instead of exploiting structure in the signal matrix like [13] and [14].

The OptShrink algorithm proposed by Nadakuditi (2014) [14] is the improved method based on the truncated singular value decomposition method which is focused on the weighted approximation to improve tSVD method. The main different part between these two methods is that tSVD method uses top k singular values and their left and right singular vectors $\sum_{i=1}^{k} \sigma'_i \mathbf{u}'_i \mathbf{v}'^{\mathrm{H}}_i$ to obtain the approximation of the image. OptShrink algorithm uses top k left and right singular vectors as well, but it replaces the singular value with ω which is the optimal weights in form of shrinkage-and-thresholding operator. And it can perform better in denoising image than the original method. More details about the algorithm will be explained in Section 4.

3 Motivation

As stated previously, images acquired in the real world usually contain noises, which greatly reduce the accuracy of the application of image pattern recognition and classification. Therefore, implementing image denoising before applying it to pattern recognition and classification is important in practice. The classical tSVD method can reduce the rank of the original dataset so that the noise part of the matrix can be partially eliminated, further, the computational complexity can be reduced as well. Hence, this method appears prominently in denoising the noise-plus image to low-rank image. And based on Nadakuditi's paper (2014) [14], the OptShrink algorithm inherits the advantage part of tSVD method in reducing the rank of the matrix and the optimal weight in the OptShrink method can perform better in denoising the matrix.

Also, Zhou et al. (2014) [15] reviewed several representative low-rank algorithms and applications in image analysis, the modeling methods categorized the modeling methods into two types: convex programming-based methods [16] [17] and factorization-based methods [18] [19]. The convex programming-based methods normally attain a stable performance for both noiseless and noisy case due to the global optimality in optimization and this type of methods also inspired us to investigate the OptShrink algorithm which brings into sharp focus the shrinkage-and-thresholding with convex penalty functions to form the optimal weights.

Thus, all these real-world problems and these notable features of OptShrink algorithm motivated us to investigate what's the performance of this algorithm and how its performance.

4 Methodology

In this section, we first showed the objective and procedures of implementing the OptShrink algorithm, which we used for low-rank matrix restoration and denoising. Next, we explained the rationale behind it and how the algorithm is achieved.

The original signal-plus-noise input matrix $\mathbf{X} \in \mathbb{R}^{\boldsymbol{n}\times\boldsymbol{m}}$ we observed could be modeled as

$$\mathbf{X} = \mathbf{S} + \mathbf{N}$$

where **S** is the signal matrix and **N** is the random noise-only matrix. Here we denote the SVD of matrix **X** as $\mathbf{X} = \sum_{i=1}^{q} \sigma_i u_i v_i^H$ where u_i and v_i represent the left and right singular value of **X** with the singular value σ_i, and $q = \min(n, m)$. The SVD of the signal matrix **S** was denoted as $\mathbf{S} = \sum_{i=1}^{r} \theta_i u'_i v'^H_i$ where u'_i and v'_i represent the left and right singular value of **S** with the singular value σ_i and r is the estimate of the rank of the latent low-rank matrix which can effectively reflect the information contained.

In order to get the best estimate of the signal matrix, we define the square error as

$$SE(\omega) = \left\| \mathrm{S} - \sum_{i=1}^{l} \omega_i \mathbf{u}_i \mathbf{v}_i^H \right\|_F^2$$

Therefore, we have the denoising optimization problem

$$\omega := \underset{\omega=[\omega_1 \ldots \omega_r]^T \epsilon\, \mathbb{R}_+^r}{\operatorname{argmin}} \quad SE(\omega),$$

which could also be written as

$$\omega := \underset{\omega=[\omega_1 \ldots \omega_r]^T \epsilon\, \mathbb{R}_+^r}{\operatorname{argmin}} \left\| \sum_{i=1}^{r} \theta_i \mathbf{u}'_i \mathbf{v}'^H_i - \sum_{i=1}^{r} \omega_i \mathbf{u}_i \mathbf{v}_i^H \right\|_F^2 \qquad (1)$$

The above optimization problem is unobservable because it depends on the unknown matrix, however, the solution to the problem is computable as described below.

The solution is given by

$$\omega_i = \left(\Re \left\{ \sum_{j=1}^{r} \theta_j (\mathbf{u}_i^H \mathbf{u}'_j)(\mathbf{v}_j^H \mathbf{v}'_i) \right\} \right)_+ \qquad (2)$$

with almost sure convergence to

$$-2 \frac{D_{\mu_X}(\rho_i)}{D'_{\mu_X}(\rho_i)}$$

where $x_+ = \max(0, x)$, $\rho_i = D_{\mu X}^{-1}\left(\frac{1}{\theta_i^2}\right)$, μ_X is the empirical singular value distribution, $D_{\mu X}(\cdot)$ is the *D*-Transform of μ_X defined as

$$D_{\mu_X}(z) := \left[\int \frac{z}{z^2 - t^2} d\mu_X(t) \right] \times \left[c \int \frac{z}{z^2 - t^2} d\mu_X(t) + \frac{1-c}{z} \right]$$

for $z \notin \operatorname{supp} \mu_X$, and $D_{\mu X}^{-1}(\cdot)$ denotes its functional inverse.

For a matrix X with the dimension $n \times m$, by defining

$$D(z; X) := \frac{1}{n} Tr(z(z^2 I - XX^H)^{-1}) \cdot \frac{1}{m} Tr(z(z^2 I - X^H X)^{-1})$$

and

$$D'(z; X) := \frac{1}{n} Tr[z(z^2 I - XX^H)^{-1}] \cdot \frac{1}{m} Tr[-2z^2(z^2 I - X^H X)^{-2} + (z^2 I - X^H X)^{-1}]$$

$$+\frac{1}{m}Tr[z(z^2I - X^HX)^{-1}] \cdot \frac{1}{n}Tr[-2z^2(z^2I - XX^H)^{-2} + (z^2I - XX^H)^{-1}]$$

and the definition of the D-transform, we could reach the conclusion that $D(z; X)$ almost sure converge to $D_{\mu_X}(z)$,and $D'(z; X)$ almost sure converge to $D'_{\mu_X}(z)$ for z outside the support of μ_X. Since ρ_i is the large matrix limit of the i-th largest singular value which has the biased but asymptotically consistent estimator $\rho_i = \sigma_i$, the weight ω could be written as

$$\omega_i = -2\frac{D(\sigma_i; \mathbf{\Sigma}_r)}{D'(\sigma_i; \mathbf{\Sigma}_r)}$$

The final denoised signal matrix is

$$\mathbf{S} = \sum_{i=1}^{r} \omega_i \, \mathbf{u}_i \mathbf{v}_i^H$$

The solution to (1) is expressed by (2). The related proof can be found in [14]. The procedures of implementing the OptShrink low-rank matrix algorithm are summarized below:

OptShrink Algorithm ($\mathbf{X}$, k):

(a) Input: Signal-plus-noise matrix observed $\mathbf{X}$ ($\mathbf{X} \in \mathbb{R}^{\boldsymbol{n \times m}}$)

Input: Estimate of the effective rank of the latent low-rank matrix $\mathbf{k}$

(b) Compute $\mathbf{X} = \sum_{i=1}^{q} \sigma_i \mathbf{u}_i \mathbf{v}_i^H$ where $q = \min(n, m)$;

(c) Compute $\mathbf{\Sigma}_r = \mathrm{diag}(\sigma_{r+1}, \dots, \sigma_q) \in \mathbb{R}^{\boldsymbol{(n-r)\times (m-r)}}$

(d) For $i = 1, \dots, r$

Compute $D(\sigma_i; \Sigma_r)$ and $D'(\sigma_i; \Sigma_r)$

Compute $\omega_{i,r} = -2\frac{D(\sigma_i; \mathbf{\Sigma}_r)}{D'(\sigma_i; \mathbf{\Sigma}_r)}$

End

(e) Return $\mathbf{S} = \sum_{i=1}^{r} \omega_{i,r} \mathbf{u}_i \mathbf{v}_i^H$ as the denoised estimate of the rank r signal matrix

where

$$D(\sigma_i; \mathbf{\Sigma}_r) := \frac{1}{n}Tr\left(\sigma_i\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r\mathbf{\Sigma}_r{}^H\right)^{-1}\right) \cdot \frac{1}{m}Tr\left(\sigma_i\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r{}^H\mathbf{\Sigma}_r\right)^{-1}\right)$$

and

$$D'(\sigma_i; \mathbf{\Sigma}_r) := \frac{1}{n}Tr\left[\sigma_i\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r\mathbf{\Sigma}_r{}^H\right)^{-1}\right] \cdot \frac{1}{m}Tr\left[-2\sigma_i{}^2\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r{}^H\mathbf{\Sigma}_r\right)^{-2} + \left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r{}^H\mathbf{\Sigma}_r\right)^{-1}\right]$$

$$+\frac{1}{m}Tr\left[\sigma_i\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r{}^H\mathbf{\Sigma}_r\right)^{-1}\right] \cdot \frac{1}{n}Tr\left[-2\sigma_i{}^2\left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r\mathbf{\Sigma}_r{}^H\right)^{-2} + \left(\sigma_i{}^2\mathbf{I} - \mathbf{\Sigma}_r\mathbf{\Sigma}_r{}^H\right)^{-1}\right]$$

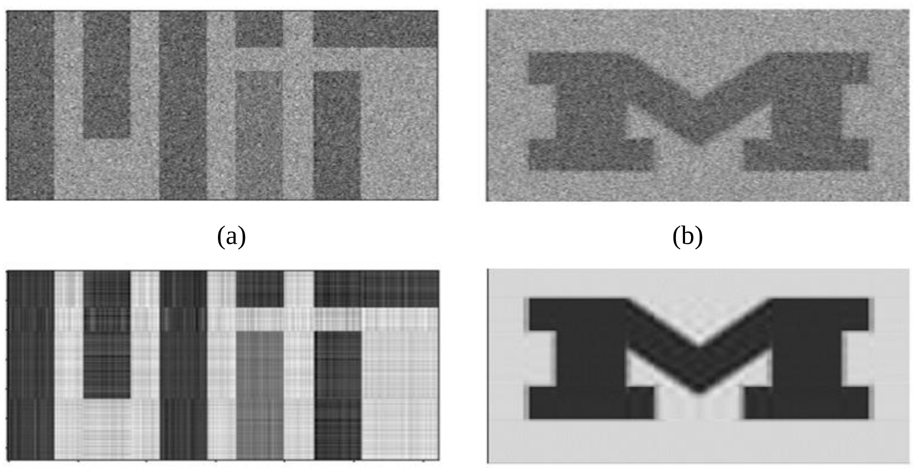
5 Applications

In this section, two applications of the OptShrink algorithm are presented and the test results are compared with the ones by truncated singular value decomposition method. The difference between these two methods is in the way of performing the approximation of the image. In the tSVD method, the top k singular values and corresponding left and right singular vectors are chosen and $\mathrm{U}[:,1\!:\!k] * \Sigma[1\!:\!k, 1\!:\!k] * \mathrm{V}[:,1\!:\!k]'$ is used to form the approximation of the image. However, in the OptShrink method, top k left and right singular vectors are used as well, but the singular values are replaced with the weights ω. As explained in the methodology section, this is the reason why the OptShrink algorithm may have better performance in image denoising. To show the versatility of the OptShrink algorithm in denoising image, two different datasets: the MNIST Dataset [22] and the RiteKit Company logo dataset [23] were used for the tests.

5.1 Application 1: low-rank image recovery by OptShrink algorithm

The purpose of the first application is to verify that the OptShrink method may perform better than the tSVD method in denoising the low-rank image, as claimed in the previous section. Firstly, the i.i.d. Gaussian noise was added into the original test images, i.e., MIT and Umich logo, as shown in Figure 1(a) and 1(b). Then, the OptShrink algorithm was conducted for image recovery, and the denoised image are plotted in Figure 1(c) and 1(d).

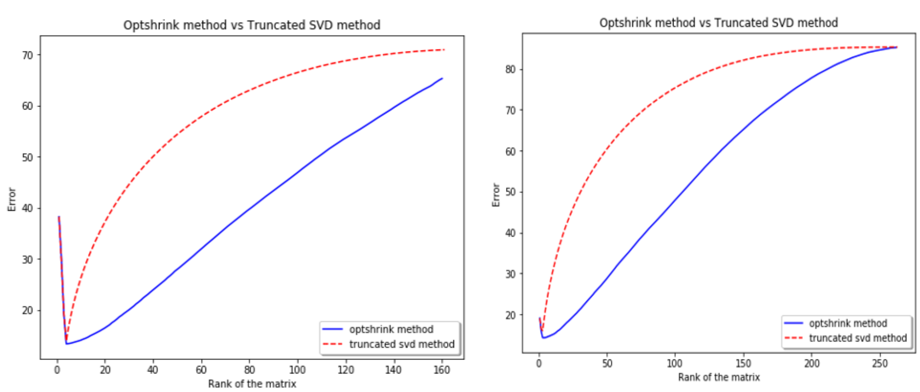
The Frobenius norm of errors, i.e., the difference between the original image (an image without noise) and denoised image, are compared for two testing images in Figure 2. From Figure 2, it can be found that the OptShrink denoise method had smaller errors than the tSVD method in low-rank image. This is consistent with our assumption that the OptShrink method performances better than tSVD method for every rank k approximation.



(a) (b)

(c) (d)

Figure 1: Testing images: (a) MIT logo with noise; (b) Umich logo with noise; (c) denoised MIT logo; (d) denoised Umich logo.



(a) (b)

Figure 2: The Frobenius norm of errors for (a) MIT logo case (b) Umich logo case

5.2 Application 2: noise-plus handwritten digit classification

The purpose of the second application is to demonstrate the potential of the OptShrink method for improving the performance of existing methods. The Classification of handwritten digits is a classic problem in machine learning which is to construct a mapping that can correctly predict the digits, given training example of old handwritten digits with truth labels. One widely-used algorism is the Nearest Subspace classification (NSS) method [24]. The NSS method assumes data lie on multiple affine subspaces and finds an estimate for these subspaces and assigns each instance to the nearest subspace. The optimal solution of the NSS method is the Singular Value Decomposition of the data matrix for the k^{th} class. The other detailed introduction of NSS method can be found in [24]. The NSS method has been used successfully in face recognition, handwritten digit recognition, and speech recognition. However, when the training data was infected by noise, the original NSS often cannot have a good performance. Thus, in this application, we want to improve the NSS method by plugging the OptShrink algorithm into it and apply to the classification problem of the noise-plus digit dataset.

5.2.1 Algorithm of Nearest Subspace and OptShrink Mixed Method

(a) Input: Signal-plus-noise matrix $\mathbf{X}$ ($\mathbf{X} \in \boldsymbol{R}^{n\times m}$)
(b) Parameter selection:
  For rank = 1 ,…, 50
    For i = 1 …, min(m, n)
      Prediction accuracy = MSS method*** (OptShrink algorithm (X, rank), i)
  rank = k, in which prediction accuracy will achieve max value.
(c) Fix rank k of OptShrink method
(d) X_{opt} = OptShrink ($\mathbf{X}$, k)
(e) For r = 1 ,…, min(m, n)
    Perform *Nearest Subspace method**** ($\mathbf{X}_{opt}$, r)
(f) Find the best prediction accuracy

****Nearest Subspace Method*: input parameters ($\mathbf{X}$, r)

(a) Input: Matrix $\mathbf{X}$ ($\mathbf{X} \in \boldsymbol{R}^{n\times m\times d}$); n, m, d = size($\mathbf{X}$); $\mathbf{U}_j$ = zeros (n, r, d)
(b) For i = 1 ,…, d (class)
    Compute the top r left singular vectors of the $X[:,:,i]$
    Store these r left singular vectors in $U[:,:,i]$.
(c) Error for each digit of d classes = sum ((test matrix – $U[:,:,j]$. * ($U[:,:,j]^T$ * test data matrix) ^{2}, axis=0)
(d) Label of each digit = minimum error of different classes

* r is the parameter decided by you but need to be less than the rank of training dataset

5.2.2 Data exploration and Model selection

Firstly, the handwritten digits from MNIST dataset were used in this application: 8,000 handwritten digits as training samples and 2,000 digits as testing samples. Each digit (0 – 9) has 800 training samples and 200 testing samples. Secondly, the i.i.d. Gaussian noise was added to the training data as the noise-plus handwritten digits. The original and noise-plus digits (from 0 to 9) are shown in Figure 3. Next, the parameter selection was done by using the leave-one-out cross-validation method. It was found that when parameter k in the OptShrink algorithm was 10, the classification error was minimum, and the best predict accuracy for n-fold cross validation was 95.2%.

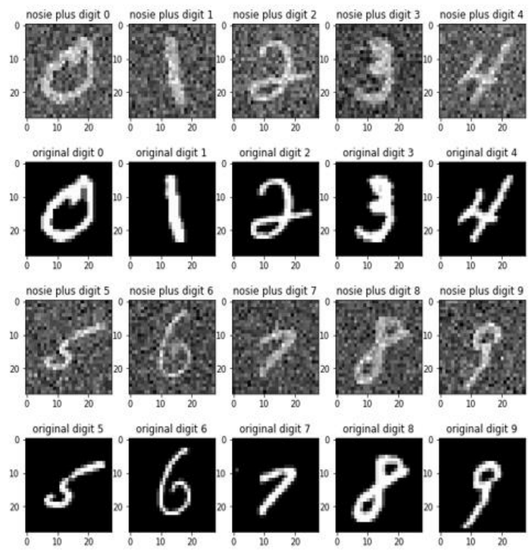


Figure 3: Original and noise-plus images

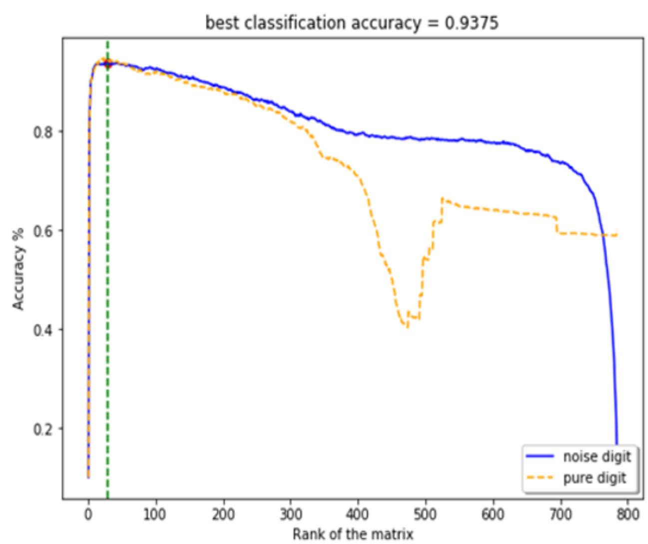


Figure 4: Accuracy of NSS and OptShrink mixed method

5.2.3 Evaluation and comparison

As shown in Table 1, in this application, the original NSS method has been used successfully in pure handwritten digit recognition (without noise) and the classification accuracy can achieve 94.75%. However, when the training data was infected by noise, the prediction accuracy of original NSS method decreased to 86.46%. By using NSS and OptShrink mixed method, the best prediction accuracy was improved to 93.75%. From Figure 4, the best prediction accuracy of NSS and OptShrink mixed algorithm is very close to the best prediction accuracy of the pure training data.

Specifically, in Figure 4, the blue line shows the classification result of the NSS and OptShrink mixed algorithm on noise-plus digit. It can be found that, when rank is 28, the best prediction accuracy is about 94% which means the OptShrink algorithm can almost restore the main information of the pure digit even if it was infected by noise. From the quantitative comparison, the advantage of the improved algorithm is significant, with performance improvement of 8.43% compared with the original method when we use noise-plus training data. Therefore, the mixed algorithm successfully classifies the noise-plus digit and improves the original NSS method.

Table 1: Comparison of different methods

Classification method	Data type	Best predict accuracy (test data)
Original Nearest Subspace Method	Pure	94.75%
Original Nearest Subspace Method	Add Noise	86.46%
Nearest Subspace and OptShrink Mixed Method	Add Noise	93.75%

6 Conclusions

In this project, we studied and implemented OptShrink algorithm to denoise university logo to the low-rank image in Application 1. Moreover, we also extended this algorithm by combining it with the existing Nearest Subspace method to classify noise-plus handwritten digits from digit 0 – 9 in Application 2. In Application 1, the results of the OptShrink algorithm in denoising University logo achieves better performance than the original truncated singular value decomposition method, especially when the rank of the approximation matrix is less than 50. It means the OptShrink algorithm has a better denoise performance when the approximation rank is low. In other words, when the original data matrix is infected by noise data, we can use OptShrink algorithm to denoise and have a better low rank denoised matrix than tSVD method, and also the influence of noise part can be reduced. In Application 2, the improved algorithm, i.e., Algorithm of Nearest Subspace and OptShrink Mixed Method, can overcome the noise-plus handwritten digit. It is found that the best prediction accuracy of the mixed algorithm is about 94% which is very similar to the best prediction accuracy when we used the pure training data. This shows that the improved algorithm successfully classifies the noise-plus digit and outperform the original Nearest Subspace method when the training data was infected by the noise. Moreover, from the results of Application 2, i.e., Figure 4, it can be observed that the mixed algorithm also has much stronger robustness compared with the original NSS method. However, the cost time of the mixed algorithm is about twice of the original NSS algorithm. This is due to the extra denoise process caused by the OptShrink algorithm. Even though we can improve the classification accuracy by 8.5%, the compute time increase by 100%. Thus, in practice, we need consider the tradeoff between time and performance deliberately. Overall, the advantage of the OptShrink algorithm is understandable in these two applications. For the future works, the mixed algorithm should be further tested by other data with different noise distribution, and the idea of mixed algorithm may be extended and applied to the other denoising methods.

7. Description of individual effort

Szu-Yun Lin focuses on the background research and the reviews of previous related works. Xi Wang works on the description and explanation of the focused algorithm. Yizhi Liu focuses on the implementation of the algorithms in Python. The evaluation of performance and discussions part were done by the whole group.

The related codes are available on Github:

https://github.com/wangix3/Group-18_EECS-545-Project.git

References

[1] Jolliffe, I. T. (2002) *Principal Component Analysis*, vol. 2. New York, NY, USA: Wiley.

[2] Hu, Y., Liu, X. & Jacob, M. (2018) Adaptive structured low rank algorithm for MR image recovery. *2018 IEEE 15th International Symposium on Biomedical Imaging*, Washington, DC, 1260-1263.

[3] Markovsky, I. (2012) *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer.

[4] Eckart, C. & Young, G. (1936) The approximation of one matrix by another of lower rank. *Psychometrika* **1**(3):211–218.

[5] Mirsky, L. (1960) Symmetric gauge functions and unitarily invariant norms. *Quart. J. Math.* **11**(1): 50–59.

[6] Golub, G. H., Hoffman, A. & Stewart, G. W. (1987) A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and its Applications* **88**: 317–327.

[7] d'Aspremont, A., Bach, F. & El Ghaoui, L. (2008) Optimal solutions for sparse principal component analysis. *Journal of Machine Learning Research* **9**:1269–1294.

[8] Johnstone, I. M. & Lu, A. (2009) On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104**(486):682–693.

[9] Jennaton, R., Obozinksi, G. & Bach, F. (2010) Structured sparse principal component analysis. *International Conference on Artificial Intelligence and Statistics* **9**:366–373.

[10] Wikes, D. M. & Hayes, M. H. (1988) Iterated toeplitz approximation of covariance matrices. *Proc. ICASSP*: 1663–1666.

[11] Li, Y., Liu, K. J. R. & Razavilar, J. (1997) A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation. *IEEE Trans. Signal Process.* **45**(2):481–486.

[12] Langville, A. N., Meyer, C. D., Albright, R., Cox, J. & Duling, D. (2006) Initializations for the nonnegative matrix factorization. *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 23–26.

[13] Srebro, N. & Jaakkola, T. (2003) Weighted low-rank approximations. *Proc. 20th Int. Conf. Mach. Learn.* **20**:720–727.

[14] Nadakuditi, R. R. (2014) OptShrink: An Algorithm for Improved Low-Rank Signal Matrix Denoising by Optimal, Data-Driven Singular Value Shrinkage. *IEEE Transactions on Information Theory*, **60**(5): 3002-3018.

[15] Zhou, X., Yang, C., Zhao, H., & Yu, W. (2014) Low-rank modeling and its applications in image analysis. *ACM Comput. Surv.* **47**(2):36:1-36:33.

[16] Cai, J. F., Candès, E. J. & Shen, Z. (2010) A singular value thresholding algorithm for matrix completion. *SIAM Journal of Optimization* **20**(4):1956–1982.

[17] Ma, S., Goldfarb, D. & Chen, L. (2011) Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming* **128**(1–2):321–353.

[18] Rennie, J. & Srebro, N. (2005) Fast maximum margin matrix factorization for collaborative prediction. *Proceedings of the International Conference on Machine Learning*.

[19] Wang, N. & Yeung, D.-Y. (2013) Bayesian robust matrix factorization for image and video processing. *Proceedings of the International Conference on Computer Vision*

[20] Drineas, P., Kannan, R. & Mahoney, M. W. (2006) Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix, *SIAM Journal on Computing* **36**(1):158–183.

[21] Koltchinskii, V., Lounici, K. & Tsybakov, A. B. (2011) Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Annals of Statistics* **39**(5):2302–2329.

[22] MNIST Dataset. http://yann.lecun.com/exdb/mnist/

[23] RiteKit Company logo dataset. https://ritekit.com/api-demo/company-logo

[24] Wang, Yi. (2015). Consistency Analysis of Nearest Subspace Classifier. arXiv:1501.06060.