# Robust $\ell_p$-norm Singular Value Decomposition

**Kha Gia Quach [1], Khoa Luu [2], Chi Nhan Duong [1], Tien D. Bui [1]**
[1] Concordia University, Computer Science and Software Engineering, Montréal, Québec, Canada
[2] Carnegie Mellon University, CyLab Biometrics Center, Pittsburgh, PA, USA
[1] {k_q, c_duon, bui}@encs.concordia.ca, [2] kluu@andrew.cmu.edu

## Abstract

This paper presents a novel Robust $\ell_p$-norm Singular Value Decomposition (RP-SVD) method for analyzing two-way functional data. The proposed RP-SVD is formulated as a $\ell_p$-norm based penalized loss minimization problem where a robust loss function is employed to measure the reconstruction error of a low-rank matrix approximation of the data. An appropriately defined $\ell_p$-norm penalty function is used to ensure smoothness along each of the two functional domains. The Alternating Direction Method of Multipliers is then used to find appropriate solutions to this problem. The method achieves higher accuracy in face image reconstruction compared to the state-of-the-art SVD and its extensions, i.e. Robust SVD, Regularized SVD and Robust Regularized SVD, in various scenarios.

## 1 Introduction

The Singular Value Decomposition (SVD) has become one of the basic and most important tools of modern numerical analysis, particularly numerical linear algebra. The SVD problem [1] can be simply solved in a regular closed form using a $\ell_2$-norm cost function. However, the $\ell_2$-norm process treats all input data equally and doesn't have ability to detect outliers or sparse components. Therefore, SVD subspaces are sensitive to outliers and noisy values from given input data. Fig. 1 shows an example of the limitations in SVD and other previous SVD extensions. When input data is free of noise or outliers, SVD can generate a good subspace to represent the data distribution. However, when the data contains some noise or outliers, this subspace contains a structure distortion; hence it doesn't represent well the data distribution. In addition, there is no mechanism to deal with missing values in the regular SVD representation. The decomposed matrix $\mathbf{X}$ must be completely filled with values for all $d \times n$ elements; otherwise the problem is unsolvable.

This paper presents a novel Robust $\ell_p$-norm ($0 < p < 1$) Singular Value Decomposition (RP-SVD) approach to solve the SVD problem approximately using $\ell_p$-norm solution. Far apart from the traditional SVD approaches, our proposed RP-SVD method is able to deal with input matrices containing missing values and can find optimal solutions for the matrix completion problems. In addition, it can also find optimal subspaces that are robust to noise and outliers. Compared to previous $\ell_1$-norm methods (see section 2) , our RP-SVD approach can find higher sparsity solutions thank to the $\ell_p$-norm in the objective function. The Alternating Direction Method of Multipliers (ADMM) is employed to find appropriate solutions.

## 2 Related Work

The SVD was established for real square matrices in the 1870's by Beltrami and Jordan and for general rectangular matrices by Eckart and Young (reviewed in [1]). In this section, we review recent SVD studies. Huang et al. [3] proposed a regularized SVD (RSVD) for dimension reduction and feature extraction. RSVD was posed as a low-rank matrix approximation problem with a squared loss function on reconstruction errors and a quadratic penalty on the factorized solutions. However,

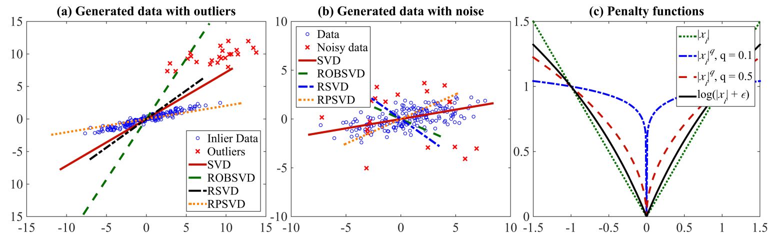


Figure 1: (a) and (b) show principal directions obtained by using SVD, ROBSVD [2], RSVD [3], and our proposed RP-SVD on the toy data set with outliers and noise. (c) Illustration of common convex and non-convex regularized functions.

RSVD is also sensitive to outliers as showed in Fig. 1. Liu et al [2] presented a robust SVD (ROBSVD) that can cope with outliers and impute missing values for microarray data. Bai et al. [4] proposed a supervised SVD (SSVD), less sensitive to outliers, to improve the robustness of analyzing functional Magnetic Resonance Imaging (fMRI) brain images. Zhang et al. [5] developed a robust regularized SVD (ROBRSVD) method to lessen the effects of outliers. ROBRSVD is a robustified RSVD method using a robust loss function instead of the non-robust squared-error loss as in [3]. It can also be considered as smoothing of a robust SVD [2] method with the penalty term. Zhang et al. suggested to iteratively impute the missing values by replacing them with values from the previous iteration, then apply the iterative reweight least square (IRLS) algorithm to solve it.

## 3 Our Proposed Robust $\ell_p$-norm SVD (RP-SVD)

Given a matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$ that contains missing values, noise and outliers, this work aims to introduce a novel RP-SVD approach using $\ell_p$-norm, where $0 < p < 1$, to further enhance the robustness of SVD to deal with outliers and noise. The SVD problem can be then formulated as follows:

$$\min_{\mathbf{U},\mathbf{\Sigma},\mathbf{V}} \|\mathbf{M} \odot (\mathbf{X} - \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top)\|_p \text{ s.t. } \mathbf{U}^\top\mathbf{U} = \mathbf{I}, \mathbf{V}^\top\mathbf{V} = \mathbf{I} \tag{1}$$

Far apart from the traditional SVD method, our proposed RP-SVD approach presented in Eqn. (1) allows to decompose an input matrix $\mathbf{X}$ containing missing values and outliers denoted by the weight matrix $\mathbf{M}$, where $\mathbf{M}(i,j) > 0$ if the data point $\mathbf{X}_{i,j}$ exists, otherwise $\mathbf{M}(i,j) = 0$. $\odot$ denotes the component-wise multiplication. Generally, Eqn. (1) is a non-convex problem. Let $\mathbf{E} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{L}\mathbf{R}$, where $\mathbf{L}$ is an orthogonal matrix, i.e. $\mathbf{L}^\top\mathbf{L} = \mathbf{I}$, then $\|\mathbf{E}\|_* = \|\mathbf{L}\mathbf{R}\|_* = \|\mathbf{R}\|_*$. The objective function is reformulated as follows:

$$\min_{\mathbf{L},\mathbf{R},\mathbf{E}} \|\mathbf{M} \odot (\mathbf{X} - \mathbf{E})\|_p + \lambda\|\sigma(\mathbf{R})\|_p \quad \text{s.t. } \mathbf{E} = \mathbf{L}\mathbf{R}, \mathbf{L}^\top\mathbf{L} = \mathbf{I} \tag{2}$$

where the parameter $\lambda$ controls the trade-off between trace norm regularization and reconstruction fidelity. To solve the problem in Eqn. (2), we first linearize it by using first order Taylor expansion to form an equivalent form as follows:

$$\min_{\mathbf{L},\mathbf{R},\mathbf{E}} \textstyle\sum_{i,j} g\left(\mathbf{M}_{i,j}(\mathbf{X}_{i,j} - \mathbf{E}_{i,j})\right) + \lambda \sum_j g\left(\sigma_j(\mathbf{R})\right) \quad \text{s.t. } \mathbf{E} = \mathbf{L}\mathbf{R}, \mathbf{L}^\top\mathbf{L} = \mathbf{I} \tag{3}$$

where $g(\cdot) = |\cdot|^p$. Then, the corresponding augmented Lagrangian function is derived as follows:

$$\begin{aligned}\mathcal{L}_\beta(\mathbf{L},\mathbf{R},\mathbf{E},\mathbf{Y}) &\triangleq \textstyle\sum_{i,j} g\left(\mathbf{M}_{i,j}(\mathbf{X}_{i,j} - \mathbf{E}^k_{i,j})\right) + \left\langle \nabla g\left(\mathbf{M}_{i,j}(\mathbf{X}_{i,j} - \mathbf{E}^k_{i,j})\right), \mathbf{M}_{i,j}\left(\mathbf{E}^k_{i,j} - \mathbf{E}_{i,j}\right)\right\rangle \\ &+\lambda \textstyle\sum_j g\left(\sigma_j(\mathbf{R}^k)\right) + \left\langle \nabla g\left(\sigma_j(\mathbf{R}^k)\right), \sigma_j(\mathbf{R}) - \sigma_j(\mathbf{R}^k)\right\rangle + < \mathbf{Y}, \mathbf{E} - \mathbf{L}\mathbf{R} > + \frac{\beta}{2}\|\mathbf{E} - \mathbf{L}\mathbf{R}\|_F^2\end{aligned} \tag{4}$$

where $\mathbf{Y}$ is the Lagrange multipliers ensuring the linear constraints, $\beta > 0$ is the penalty parameter for the violation of the constraints. The problem defined in Eqn. (3) can be solved using ADMM by iteratively solving the following convex optimization sub-problems.

**Step 1 - Given $\mathbf{R}^k$ and $\mathbf{E}^k$, find $\mathbf{L}^{k+1}$:** By fixing $\mathbf{R}^k$ and $\mathbf{E}^k$ in the iteration $k$, $\mathbf{L}^{k+1}$ can be updated by solving the sub-problem as follows:

$$\min_{\mathbf{L}} \frac{\beta}{2}\|(\mathbf{E}^k + \beta^{-1}\mathbf{Y}^k) - \mathbf{L}\mathbf{R}^k)\|_F^2 \quad s.t. \;\; \mathbf{L}^\top\mathbf{L} = \mathbf{I} \tag{5}$$

The global optimal solution of this optimization problem can be found by first applying SVD as $[\mathbf{U}', \mathbf{S}', \mathbf{V}'] = svd((\mathbf{E}^k + \beta^{-1}\mathbf{Y}^k)\mathbf{R}^{k\top})$. Then, $\mathbf{L}^{k+1}$ can be updated as [6], $\mathbf{L}^{k+1} \leftarrow \mathbf{U}'\mathbf{V}'^{\top}$.

**Step 2 – Given $\mathbf{L}^{k+1}$ and $\mathbf{E}^k$, find $\mathbf{R}^{k+1}$:** In the second step, given $\mathbf{L}^{k+1}$ and $\mathbf{E}^k$, $\mathbf{R}^{k+1}$ can be found using the following formula,

$$\min_{\mathbf{R}} \lambda \textstyle\sum_j v_j^k \sigma_j + < \mathbf{Y}^k, \mathbf{E}^k - \mathbf{L}^{k+1}\mathbf{R} > + \frac{\beta}{2}\|\mathbf{E}^k - \mathbf{L}^{k+1}\mathbf{R}\|_F^2 \tag{6}$$

where $v_j^k = \nabla g(\sigma_j(\mathbf{R}^k))$ and $\sigma_j$ is the $j$-th singular values of the matrix $\mathbf{R}^k$. Since $\mathbf{L}^{k+1}$ is orthogonal, Eqn. (6) can be rewritten as,

$$\min_{\mathbf{R}} \lambda\beta^{-1} \textstyle\sum_j v_j^k \sigma_j + \frac{1}{2}\|\mathbf{R} - \mathbf{L}^{k+1\top}(\mathbf{E}^k + \beta^{-1}\mathbf{Y}^k)\|_F^2 \tag{7}$$

Based on Theorem 1 in [7], the solution of (7) is given by the weighted singular value thresholding (WSVT). In WSVT, the SVD is first employed, $[\mathbf{U}', \mathbf{S}', \mathbf{V}'] = svd(\mathbf{L}^{k+1\top}(\mathbf{E}^k + \beta^{-1}\mathbf{Y}^k))$, the optimal values of $\mathbf{R}^{k+1}$ can be then updated by shrinking the diagonal matrix $\mathbf{S}'$ via the soft-thresholding (shrinkage) operator $\mathbf{T}_\tau[x]$ as, $\mathbf{R}^{k+1} \leftarrow \mathbf{U}'\mathbf{T}_{\lambda\beta^{-1}v_j^k}[\mathbf{S}']\mathbf{V}'^{\top}$, where the weights $v_j^k$ are updated at each iteration as $v_j^k = p(\sigma_j^k + \epsilon)^{p-1}(0 < \epsilon \ll 1)$. The shrinkage operator is defined as $\mathbf{T}_\tau[x] = max(|x| - \tau, 0)sgn(x)$, where $sgn(x)$ is the sign function.

**Step 3 – Given $\mathbf{L}^{k+1}$ and $\mathbf{R}^{k+1}$, find $\mathbf{E}^{k+1}$:** Given $\mathbf{L}^{k+1}$ and $\mathbf{R}^{k+1}$, $\mathbf{E}^{k+1}$ can be updated using the shrinkage technique in [6],

$$\min_{\mathbf{E}} \textstyle\sum_{i,j} \mathbf{W}_{i,j}^k(\mathbf{M}_{i,j}(\mathbf{X}_{i,j} - \mathbf{E}_{i,j})) + \frac{\beta}{2}\|\mathbf{E} - (\mathbf{L}^{k+1}\mathbf{R}^{k+1} - \beta^{-1}\mathbf{Y}^k)\|_F^2 \tag{8}$$

where $\mathbf{W}_{i,j}^k = \nabla g\left(\mathbf{M}_{i,j}^k(\mathbf{X}_{i,j} - \mathbf{E}_{i,j}^k)\right) = p(\mathbf{M}_{i,j}^k(\mathbf{X}_{i,j} - \mathbf{E}_{i,j}^k) + \epsilon)^{p-1}$. Therefore, the observed $\mathbf{M} \odot \mathbf{E}$ and missing values $\overline{\mathbf{M}} \odot \mathbf{E}$ ($\overline{\mathbf{M}}$ is the complement of $\mathbf{M}$) in $\mathbf{E}$ can be updated as follows,

$$\begin{cases} \mathbf{M} \odot \mathbf{E} \leftarrow \mathbf{M} \odot (\mathbf{X} - \mathbf{T}_{\beta^{-1}\mathbf{W}}[\mathbf{X} - \left(\mathbf{L}^{k+1}\mathbf{R}^{k+1} - \beta^{-1}\mathbf{Y}^k\right)]) \\ \overline{\mathbf{M}} \odot \mathbf{E} \leftarrow \overline{\mathbf{M}} \odot (\mathbf{L}^{k+1}\mathbf{R}^{k+1} - \beta^{-1}\mathbf{Y}^k) \end{cases} \tag{9}$$

**Step 4: Update Lagrange multiplier:** $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \beta(\mathbf{E}^{k+1} - \mathbf{L}^{k+1}\mathbf{R}^{k+1})$

# 4 Experimental Evaluation

## 4.1 Synthetic Data

In this experiment, an input matrix $\mathbf{X}_0 \in R^{400\times500}$ is randomly generated. The elements $\mathbf{X}_{0i,j}$ are drawn from an uniform distribution between $[-1, 1]$ independently. 20% of the number of elements are then randomly selected as missing values by setting the corresponding entries in the mask matrix $\mathbf{M}$ to zeros. In addition, uniformly distributed noise over $[-5, 5]$ are added to 10% of the observed elements in $\mathbf{X}_0$ as outliers and Gaussian noise with $\sigma = 0.01$ are also added to all elements to form a new noisy matrix $\mathbf{X}$. The comparison algorithms, i.e. SVD (Matlab), ROBSVD [2], RSVD [3], ROBRSVD [5] and our proposed RP-SVD, factorize the noisy/outlier matrix $\mathbf{X}$ into subspaces. Then, the reconstructed matrices $\hat{\mathbf{X}}$ are computed. The reconstruction errors are measured as $OER_{\ell_1} = \|\mathbf{X}_0 - \hat{\mathbf{X}}\|_1/(m \times n)$. Table 1 shows the average errors and processing time (s) on 500 different matrices $\mathbf{X}$. We also perform two experiments with various missing data and outlier ratios. First, the missing data ratios are set from 10% to 90%. The average $\ell_1$-norm errors ($\mathrm{OER}_{\ell_1}$) over observed entries are recorded with the outlier ratios fixed at 20%. The first experiment is repeated 100 times for each level of missing data. Then, the missing data ratios are fixed to be 30%, and the outlier ratio is varied from 10% to 25%. Similarly, we repeat 100 times for each outlier ratio level. The results (the average $\ell_1$-norm errors in log scale) of the experiments are shown in Fig. 2 (a).

## 4.2 Eigenfaces

One of the classical applications of SVD is facial image analysis using eigenfaces. The eigenface displays the underlying low $K$-dimensional subspace best describing the training data. In this experiment, we aim at showing the robustness of our RP-SVD method in reconstructing eigenface decomposition in the presence of outliers. A set of 30 randomly selected $64 \times 64$ face images from

Table 1: Evaluation Results on Synthetic Data.

| **Methods** | $\mathbf{X}_0$ | | $\mathbf{X}_0$ + noise | | $\mathbf{X}_0$ + outlier | | $\mathbf{X}_0$ + noise + outlier | |
|---|---|---|---|---|---|---|---|---|
| | $OER_{\ell_1}$ | Time | $OER_{\ell_1}$ | Time | $OER_{\ell_1}$ | Time | $OER_{\ell_1}$ | Time |
| SVD | **1.6e-15** (±1.3e-16) | **0.022** | 0.005 (±6.4e-6) | **0.041** | 0.5 (±3e-3) | **0.04** | 0.5 (±3.3e-4) | **0.04** |
| ROBSVD | 0.088 (±3e-3) | 0.34 | 0.088 (±3e-3) | 0.36 | 0.137 (±2e-3) | 0.23 | 0.14 (±2e-3) | 0.25 |
| RSVD | 0.305 (±8e-3) | 667.3 | 0.305 (±9e-3) | 694 | 0.55 (±6e-3) | 728 | 0.55 (±4e-3) | 613 |
| ROBRSVD | 0.3 (±8e-3) | 884 | 0.302 (±9e-3) | 677 | 0.33 (±8e-3) | 795.4 | 0.33 (±8e-3) | 771.4 |
| **RP-SVD** | 9e-9 (±1.9e-9) | 1.53 | **0.005** (±3.4e-5) | 3.11 | **8.11e-8** (±1.6e-8) | 2.37 | **0.005** (±2.8e-5) | 3.13 |

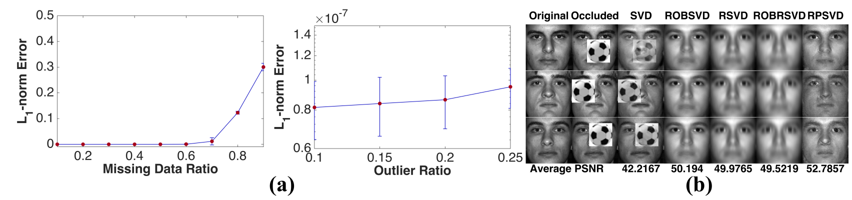


Figure 2: Experiments with outlier and missing data. (a) the average errors on synthetic data with varying missing data and outlier ratios. (b) An experiment on Extended Yale-B face database.

the Extended Yale B face database [8] are used as training set (i.e. a $4096 \times 30$ training data matrix). A $32 \times 32$ outlier image (i.e. an image of a football) is added to a random training image at a random location. The comparison methods, i.e. SVD, ROBSVD, RSVD and RP-SVD, are then applied to reconstruct the occluded facial image with $K = 10$. We repeat this procedure for 100 times. Fig. 2 (b) shows the resulting reconstructed facial images using those methods and the average PSNR also reported in this figure. Our method achieves the best PSNR value (52.79).

### 4.3 Structure from Motion

This experiment evaluates the proposed method in a real-world application named Structure from Motion. The standard Dinosaur sequence [1] containing projections of 195 points tracked over 36 views, was used in this experiment. Each tracked point is located in at least 16 views while it is occluded in other views. Thus, the measurement matrix has 74.26% of its elements missing and the originally measured tracks are illustrated in Fig. 3 (a). Fig. 3 (b), (c) and (d) shows the results obtained by Damped Newton [9] method, Damped Wiberg method [10] and our RP-SVD method, respectively. We should have closed and circular tracks from the sequence since the views of Dinosaur was captured while rotating it. Our method achieves the best reconstructions with more closed circular tracks.

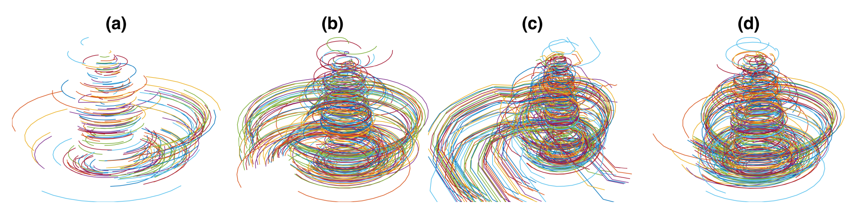


Figure 3: The experiment on the Dinosaur sequence reconstruction (a) shows the original tracks in the measurement matrix. (b) (c) and (d) show the recovered tracks using the Damped Newton [9], Damped Wiberg [10] and our RP-SVD method.

[1] available from `http://www.robots.ox.ac.uk/~vgg/data/data-mview.html`

## 5 Conclusions

This paper has presented a RP-SVD method for analyzing two-way functional data. Our proposed RP-SVD method is evaluated in various applications, i.e. noise and outlier removal, estimation of missing values, structure from motion reconstruction and facial image reconstruction. We show that RP-SVD method can achieve better results compared to the state-of-the-art SVD and its extensions.

## References


[1] V. C. Klema and A. J. Laub, "Singular Value Decomposition: Its Computation and Some Applications," *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980.

[2] L. Liu, D. M. Hawkins, S. Ghosh, and S. S. Young, "Robust singular value decomposition analysis of microarray data," *Proceedings of the NAS*, vol. 100, no. 23, pp. 13 167–13 172, 2003.

[3] J. Z. Huang, H. Shen, and A. Buja, "The analysis of two-way functional data using two-way regularized singular value decompositions," *Journal of the American Statistical Association*, vol. 104, no. 488, 2009.

[4] P. Bai, H. Shen, X. Huang, and Y. Truong, "A supervised singular value decomposition for independent component analysis of fmri," *Statistica Sinica*, vol. 18, pp. 1233–1252, 2008.

[5] L. Zhang, H. Shen, J. Z. Huang *et al.*, "Robust regularized singular value decomposition with application to mortality data," *The Annals of Applied Statistics*, vol. 7, no. 3, pp. 1540–1561, 2013.

[6] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi, "Practical low-rank matrix approximation under robust l 1-norm," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012*, 2012, pp. 1410–1417.

[7] W. Dong, G. Shi, X. Li, Y. Ma, and F. Huang, "Compressive sensing via nonlocal low-rank regularization," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3618–3632, 2014.

[8] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[9] A. M. Buchanan and A. W. Fitzgibbon, "Damped newton algorithms for matrix factorization with missing data," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005*, vol. 2, 2005, pp. 316–322.

[10] T. Okatani, T. Yoshida, and K. Deguchi, "Efficient algorithm for low-rank matrix factorization with missing components and performance comparison of latest algorithms," in *IEEE International Conference on Computer Vision (ICCV), 2011*, 2011, pp. 842–849.