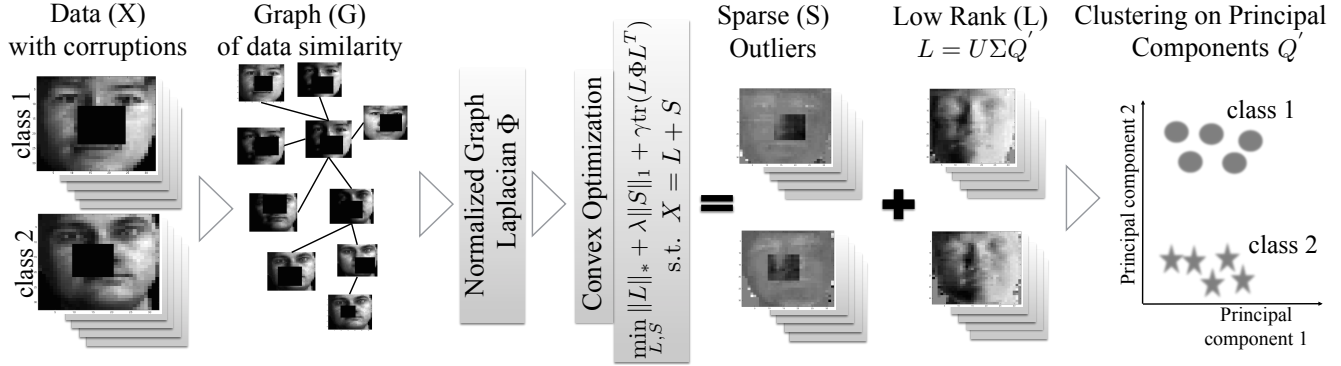# Robust Principal Component Analysis on Graphs

Nauman Shahid*, Vassilis Kalofolias*, Xavier Bresson*, Michael Bronstein†, Pierre Vandergheynst*

*Signal Processing Laboratory (LTS2), EPFL, Switzerland, `firstname.lastname@epfl.ch`

† Universita della Svizzera Italiana, Switzerland, `michael.bronstein@usi.ch`

Figure: Data (X) with corruptions → Graph (G) of data similarity → Normalized Graph Laplacian Φ → Convex Optimization $\min_{L,S} \|L\|_* + \lambda\|S\|_1 + \gamma \mathrm{tr}(L\Phi L^T)$ s.t. $X = L + S$ = Sparse (S) Outliers + Low Rank (L) $L = U\Sigma Q'$ → Clustering on Principal Components Q' (class 1, class 2)

Abstract

Principal Component Analysis (PCA) is the most widely used tool for linear dimensionality reduction and clustering. Still it is highly sensitive to outliers and does not scale well with respect to the number of data samples. Robust PCA solves the first issue with a sparse penalty term. The second issue can be handled with the matrix factorization model, which is however non-convex. Besides, PCA based clustering can also be enhanced by using a graph of data similarity. In this article, we introduce a new model called 'Robust PCA on Graphs' which incorporates spectral graph regularization into the Robust PCA framework. Our proposed model benefits from 1) the robustness of principal components to occlusions and missing values, 2) enhanced low-rank recovery, 3) improved clustering property due to the graph smoothness assumption on the low-rank matrix, and 4) convexity of the resulting optimization problem. Extensive experiments on 8 benchmark, 3 video and 2 artificial datasets with corruptions clearly reveal that our model outperforms 10 other state-of-the-art models in its clustering and low-rank recovery tasks.

1. Introduction

What is PCA? Given a data matrix $X \in \mathbb{R}^{p \times n}$ with n p-dimensional data vectors, the classical PCA can be formulated as learning the projection $Q \in \mathbb{R}^{d \times n}$ of X in a d-dimensional linear space characterized by an orthonormal basis $U \in \mathbb{R}^{p \times d}$ (1st model in Table 1). Traditionally, U and Q are termed as *principal directions* and *principal*

components and the product UQ is known as the *low-rank approximation* $L \in \mathbb{R}^{p \times n}$ of X. .

Applications of PCA: PCA has been widely used for two important applications.

1. Dimensionality reduction or low-rank recovery.
2. Data clustering using the principal components Q in the low dimensional space.

However, classical PCA formulation is susceptible to errors in the data because of the quadratic term. Thus, an outlier in the data might result in erratic principal components Q which can effect both the dimensionality reduction and clustering.

Robust dimensionality reduction using PCA: Candes *et al.* [6] demonstrated that PCA can be made robust to outliers by exactly recovering the low-rank representation L even from grossly corrupted data X by solving a simple convex problem, named Robust PCA (RPCA, 2nd model in Table 1). In this model, the data corruptions are represented by the sparse matrix $S \in \mathbb{R}^{p \times n}$. Extensions of this model for inexact recovery [25] and outlier pursuit [22] have been proposed.

The clustering quality of PCA can be significantly improved by incorporating the data manifold information in the form of a graph G [10, 24, 11, 20]. The underlying assumption is that the low-dimensional embedding of data X lies on a smooth manifold [2]. Let $G = (V, E, A)$ be the graph with vertex set V as the samples of X, E

Table 1: A comparison of various PCA models and their properties. $X \in \mathbb{R}^{p \times n}$ is the data matrix, $U \in \mathbb{R}^{p \times d}$ and $Q \in \mathbb{R}^{d \times n}$ are the principal directions and principal components in a d dimensional linear space (rank = d). $L = UQ \in \mathbb{R}^{p \times n}$ is the low-rank representation and $S \in \mathbb{R}^{p \times n}$ is the sparse matrix. Φ, Φ^g and $\Phi_h^g \in \mathbb{S}^{n \times n}$ characterize a simple graph or a hypergraph G between the samples of X. $\| \cdot \|_F$, $\| \cdot \|_*$ and $\| \cdot \|_1$ denote the Frobenius, nuclear and l_1 norms respectively.

	Model	Objective	Constraints	Parameters	Graph?	Factors?	Convex?
1	PCA	$\min_{U,Q} \|X - UQ\|_F^2$	$U^T U = I$	d	no	yes	no
2	RPCA [6]	$\min_{L,S} \|L\|_* + \lambda \|S\|_1$	$X = L + S$	λ	no	no	yes
3	**PROPOSED**	$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 + \gamma \operatorname{tr}(\mathbf{L\Phi L^T})$	$\mathbf{X = L + S}$	λ, γ	**YES**	**NO**	**YES**
4	GLPCA [10]	$\min_{U,Q} \|X - UQ\|_F^2 + \gamma \operatorname{tr}(Q\Phi Q^T)$	$QQ^T = I$				
5	RGLPCA [10]	$\min_{U,Q} \|X - UQ\|_{2,1} + \gamma \operatorname{tr}(Q\Phi Q^T)$		d, γ			
6	MMF [24]	$\min_{U,Q} \|X - UQ\|_F^2 + \gamma \operatorname{tr}(Q\Phi Q^T)$	$U^T U = I$		yes	yes	no
7	MMMF [20]	$\min_{U,Q,\boldsymbol{\alpha}} \|X - UQ\|_F^2 + \gamma \operatorname{tr}(Q(\sum_g \alpha_g \Phi^g)Q^T) + \beta\|\boldsymbol{\alpha}\|^2$	$U^T U = I$	d, γ, β			
8	MHMF [11]	$\min_{U,Q,\boldsymbol{\alpha}} \|X - UQ\|_F^2 + \gamma \operatorname{tr}(Q(\sum_g \alpha_g \Phi_h^g)Q^T) + \beta\|\boldsymbol{\alpha}\|^2$	$\mathbf{1}^T \boldsymbol{\alpha} = \mathbf{1}$				

the set of pairwise edges between the vertices V and A the adjacency matrix which encodes the weights of the edges E. Then, the *normalized graph Laplacian* matrix $\Phi \in \mathbb{R}^{n \times n}$ which characterizes the graph G is defined as $\Phi = D^{-1/2}(D - A)D^{-1/2}$, where D is the degree matrix defined as $D = \operatorname{diag}(d_i)$ and $d_i = \sum_j A_{ij}$. Assuming that a p-nearest neighbors graph is available, several methods to construct A have been proposed in the literature. The three major weighting methods are: 1) binary, 2) heat kernel and 3) correlation distance [11].

In this paper we propose a novel convex method to improve the clustering and low-rank recovery applications of PCA by incorporating spectral graph regularization to the Robust PCA framework. Extensive experiments reveal that our proposed model is more robust to occlusions and missing values as compared to 10 state-of-the-art dimensionality reduction models.

2. Main Idea & Proposed Model

The main contributions of our work are:

1. Exact recovery of the low-rank representation L from grossly corrupted data X.

2. Recovery of the low-rank representation L that also reveals the underlying class separation.

3. High cluster purity in the low dimensional space even when no clean data is available.

4. A simple convex optimization problem with minimal parameters to achieve these objectives.

The figure on page 1 illustrates the main idea of our work.

Without any doubt the contributions 1 and 4 are given by the work of Candes *et al.* [6], namely Robust PCA (RPCA, 2nd model in Table 1). Thus, as a first step, we propose that instead of utilizing a classical PCA-like model so as to explicitly learn principal directions U and principal components Q of data X, one can directly recover the low-rank matrix L itself ($L = UQ$).

Secondly and more importantly, to achieve contributions 2 and 3 we propose: *The low-rank matrix L itself lies on*

a smooth manifold and it can be recovered directly on this manifold. **Our proposed model** is as follows:

$$\min_{L,S} \|L\|_* + \lambda \|S\|_1 + \gamma \operatorname{tr}(L\Phi L^T), \qquad (1)$$

$$\text{s.t. } X = L + S,$$

where the sparse errors in the data are modeled by S and L is the low-rank approximation of X. Parameters λ and γ control the amount of sparsity of S and smoothness of L on the graph Φ respectively. We will define our graph in Section 6.3.

Generalization of Robust PCA: We call our proposed model (1) Robust PCA on Graphs. It is a direct extension of the Robust PCA proposed by Candes *et al.* [6] with smoothness manifold regularization. That is, setting $\gamma = 0$ in our model (1) we obtain the standard model of [6].

3. Related Works: Factorized PCA Models

Both manifold regularization and robustness techniques for PCA have been proposed in the literature, either separately or combined [10, 24, 11, 20]. Following the classical PCA model they explicitly learn two factors U and Q such that $X \approx UQ$. We will, therefore refer to these models as *factorized PCA models*. Furthermore, unlike our model, some of these works [10] assume the graph smoothness of principal components Q (instead of $L = UQ$).

Jiang *et al.* proposed Graph Laplacian PCA (GLPCA) [10] (4th model in Table 1) which leverages the graph regularization of principal components Q using the term $\operatorname{tr}(Q\Phi Q^T)$ for clustering in the low dimensional space (see also [1]). They also proposed a robust version of their model (5th model in Table 1) and demonstrated the robustness of principal components Q to occlusions in the data.

Zhang and Zhao [24] proposed Manifold Regularized Matrix Factorization (MMF, 6th model in Table 1) which exploits the orthonormality constraint on the principal directions U (contrary to [10]) to acquire a unique low-rank matrix $L = UQ$ for any optimal pair U, Q. In this case we have $\operatorname{tr}(Q\Phi Q^T) = \operatorname{tr}(UQ\Phi(UQ)^T) = \operatorname{tr}(L\Phi L^T)$, therefore this model implicitly assumes the graph smoothness of

L. The extensions of this model with an ensemble of graph and hypergraph regularization terms have been proposed by Tao *et al.* [20] and Jin *et al.* [11] respectively (7th and 8th models in Table 1).

Shortcomings of state-of-the-art: Although the models proposed in [10, 24, 11, 20] leverage the graph to learn enhanced class structures, they still suffer from numerous problems. Most of these models are not robust to data corruptions [24, 11, 20]. Those which leverage the robustness suffer from non-convexity [10]. An ensemble of graphs or hypergraphs leverages the non-linearity of data in an effective manner [11, 20]. However, it makes the models non-convex and the resulting alternating direction methods can get stuck in local minima.

Notation & Terminology: Throughout this article $\|\cdot\|_F$, $\|\cdot\|_*$ and $\|\cdot\|_1$ denote the Frobenius, nuclear and l_1 norms respectively. We will refer to the regularization term $\mathrm{tr}(Q\Phi Q^T)$ as *principal components graph regularization*. Note that the graph regularization involves principal components Q, not principal directions U. We will also refer to the regularization term $\mathrm{tr}(L\Phi L^T)$ as *low-rank graph regularization*. RPCA and our proposed models (2nd and 3rd models in Table 1) which perform exact low-rank recovery by splitting $X = L + S$ will be referred to as *non-factorized PCA models*. A comparison of various PCA models introduced so far is presented in Table 1. Note that only RPCA and our proposed model leverage convexity and enjoy a unique global optimum with guaranteed convergence.

4. Comparison with Related Works

The main differences between our model (1) and the various state-of-the-art factorized PCA models [10, 24, 17, 11, 20] are, as summarized in Table 1, the following.

Non-factorized model: Instead of explicitly learning the principal directions U and principal components Q, it learns their product, *i.e.* the low-rank matrix L. Hence, (1) is a non-factorized PCA model.

Exact low-rank recovery: Unlike factorized models we target the exact low-rank recovery by modeling the data matrix as the sum of low-rank L and a sparse matrix S.

Different graph regularization term: Model (1) is based on the assumption that it is the low-rank matrix L that is smooth on the graph, and not just the principal components matrix Q. Therefore we replace the principal components graph term $\mathrm{tr}(Q\Phi Q^\top)$ with the low-rank graph term $\mathrm{tr}(L\Phi L^\top)$. Note that as explained in Section 3, the two terms are only equivalent if orthogonality of U is further assumed, as in [24, 17, 11, 20] and not in [10].

4.1. Advantages over Factorized PCA Models

Robustness to gross corruptions for clustering & low-rank recovery: The low-rank graph $\mathrm{tr}(L\Phi L^T)$ can be more realistic than the principal components graph $\mathrm{tr}(Q\Phi Q^T)$. It

allows the principal directions U to benefit from the graph regularization as well (recall that $L = UQ$). Thus, our model enjoys an enhanced low-rank recovery and class separation even from grossly corrupted data. For details, please refer to Sections 7 & 8 (also see Fig. 12 in supplementary material).

Convexity: It is a strictly convex problem and a unique global optimum can be obtained by using standard methods like an Alternating Direction Method of Multipliers (ADMM) [4].

One model parameter only: Our model does not require the rank of L to be specified up-front. The nuclear norm relaxation enables the automatic selection of an appropriate rank based on the parameters λ and γ. Furthermore, as illustrated in our experiments, the value $\lambda = 1/\sqrt{\max(n, p)}$ proposed in [6] gives very good results. As a result, the only unknown parameter to be selected is γ, and for this we can use methods such as cross validation (For additional details please see Figs. 15 & 16 in the supplementary material).

5. Optimization Solution

We use an ADMM [4] to rewrite Problem (1):

$$\min_{L,S,W} \|L\|_* + \lambda\|S\|_1 + \gamma\,\mathrm{tr}(W\Phi W^T)$$
$$\text{s.t. } X = L + S,\ L = W.$$

Thus, the augmented Lagrangian and iterative scheme are:

$$(L, S, W)^{k+1} = \operatorname*{argmin}_{L,S,W} \|L\|_* + \lambda\|S\|_1 + \gamma\,\mathrm{tr}(W\Phi W^T)$$
$$+ \langle Z_1^k, X - L - S\rangle + \frac{r_1}{2}\|X - L - S\|_F^2$$
$$+ \langle Z_2^k, W - L\rangle + \frac{r_2}{2}\|W - L\|_F^2,$$
$$Z_1^{k+1} = Z_1^k + r_1(X - L^{k+1} - S^{k+1}),$$
$$Z_2^{k+1} = Z_2^k + r_2(W^{k+1} - L^{k+1}),$$

where $Z_1 \in \mathbb{R}^{p\times n}$ and $Z_2 \in \mathbb{R}^{p\times n}$ are the lagrange multipliers and k is the iteration index. Let $H_1^k = X - S^k + Z_1^k/r_1$ and $H_2^k = W^k + Z_2^k/r_2$, then this reduces to the following updates for L, S and W as:

$$L^{k+1} = \mathrm{prox}_{\frac{1}{(r_1+r_2)}\|L\|_*}\left(\frac{r_1 H_1^k + r_2 H_2^k}{r_1 + r_2}\right),$$

$$S^{k+1} = \mathrm{prox}_{\frac{\lambda}{r_1}\|S\|_1}\left(X - L^{k+1} + \frac{Z_1^k}{r_1}\right),$$

$$W^{k+1} = r_2(\gamma\Phi + r_2 I)^{-1}\left(L^{k+1} - \frac{Z_2^k}{r_2}\right),$$

where prox_f is the proximity operator of the convex function f as defined in [7]. The details of this solution, algorithm, convergence and computational complexity are given in the supplementary material Sections A.1, A.2 & A.3.

6. Experimental Setup

We use the model (1) to solve two major PCA-based problems.

1. Data clustering in low-dimensional space with corrupted and uncorrupted data (**Section 7**).

2. Low-rank recovery from corrupted data (**Section 8**).

Our extensive experimental setup is designed to test the robustness and generalization capability of our model to a wide variety of datasets and corruptions for the above two applications. Precisely, we perform our experiments on 8 benchmark, 3 video and 2 artificial datasets with 10 different types of corruptions and compare with 10 state-of-the-art dimensionality reduction models as explained in sections 6.1 & 6.2.

6.1. Setup for Clustering

6.1.1 Datasets

All the datasets are well-known benchmarks. **The 6 image databases** include CMU PIE[1], ORL[2], YALE[3], COIL20[4], MNIST[5] and USPS data sets. **MFeat database** [6] consists of features extracted from handwritten numerals and the **BCI database**[7] comprises of features extracted from a Brain Computer Interface setup. Our choice of datasets is based on their various properties such as pose changes, rotation (for digits), data type and non-negativity, as presented in Table 2 in the supplementary material.

6.1.2 Comparison with 10 models

We compare the clustering performance of our model with k-means on the original data X and 9 other dimensionality reduction models. These models can be divided into two categories.

1. Models without graph: 1) classical Principal Component Analysis (PCA) 2) Non-negative Matrix Factorization (NMF) [13] and 4) Robust PCA (RPCA) [6].

2. Models with graph: These models can be further divided into two categories based on the graph type. **a. Principal components graph**: 1) Normalized Cuts (NCuts) [19], 2) Laplacian Eigenmaps (LE) [2], 3) Graph Laplacian PCA (GLPCA) [10], 4) Robust Graph Laplacian PCA (RGLPCA) [10], 5) Manifold Regularized Matrix Factorization (MMF) [24], 6) Graph Regularized Non-negative Matrix Factorization (GNMF) [5], **b. Low-rank graph**: Our proposed model.

Table 2 and Fig. 8 in the supplementary material give a summary of all the models.

---

[1] http://vasc.ri.cmu.edu/idb/html/face/
[2] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[3] http://vision.ucsd.edu/content/yale-face-database
[4] http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[5] http://yann.lecun.com/exdb/mnist/
[6] https://archive.ics.uci.edu/ml/datasets/Multiple+Features
[7] http://olivier.chapelle.cc/ssl-book/benchmarks.html

6.1.3 Corruptions in datasets

Corruptions in image databases: We introduce three types of corruptions in each of the 6 image databases:

1. No corruptions.

2. Fully corrupted data. Two types of corruptions are introduced in all the images of each database: a) Block occlusions ranging from 10 to 40% of the image size. b) Missing pixels ranging from 10% to 40% of the total pixels in the image. These corruptions are modeled by placing zeros uniformly randomly in the images.

3. Sample specific corruptions. The above two types of corruptions (occlusions and missing pixels) are introduced in only 25% of the images of each database.

Corruptions in non-image databases: We introduce only full and sample specific missing values in the non-image databases because the block occlusions in non-image databases correspond to an unrealistic assumption. Example missing pixels and block occlusions in the image are shown in Fig. 9 in the supplementary material.

Pre-processing: For NCuts, LE, PCA, GLPCA, RGLPCA, MMF, RPCA and our model we pre-process the datasets to zero mean and unit standard deviation along the features. Additionally for MMF all the samples are made unit norm as suggested in [24]. For NMF and GNMF we only pre-process the non-negative datasets to unit norm along the samples. We perform pre-processing after introducing the corruptions.

6.1.4 Clustering Metric

We use *clustering error* to evaluate the clustering performance of all the models considered in this work. The clustering error is $E = \left(\frac{1}{n}\sum_{r=1}^{k} n_r\right) \times 100$, where n_r is the number of misclassified samples in cluster r. We report the minimum clustering error from 10 runs of k-means (k = number of classes) on the principal components Q. This procedure reduces the bias introduced by the non-convex nature of k-means. For RPCA and our model we obtain principal components Q' via SVD of the low-rank matrix $L = U\Sigma Q'$ during the nuclear proximal update in every iteration. For more details of the clustering error evaluation and parameter selection scheme for each of the models, please refer to Fig. 10 and Table 3 of the supplementary material.

6.2. Setup for Low-Rank Recovery

Since the low-rank ground truth for the 8 benchmark datasets used for clustering is not available, we perform the following two types of experiments.

1. **Quantitative evaluation** of the normalized low-rank reconstruction error using corrupted artificial datasets.

2. **Visualization** of the recovered low-rank representations for 1) occluded images of the CMU PIE dataset and 2) static background of 3 different videos[8].

---

[8] https://sites.google.com/site/backgroundsubtraction/test-sequences

6.3. Normalized Graph Laplacian

In order to construct the graph Laplacian Φ, the pairwise Euclidean distance is computed between each pair of the vectorized data samples (x_i, x_j). Let Ω be the matrix which contains all the pairwise distances, then Ω_{ij}, the Euclidean distance between x_i and x_j is given as:

Case I: Block Occlusions

$$\Omega_{ij} = \sqrt{\frac{\|M_{ij} \circ (x_i - x_j)\|_2^2}{\|M_{ij}\|_1}},$$

where $M_{ij} \in \{0,1\}^p$ is the vector mask corresponding to the intersection of uncorrupted values in x_i and x_j. Thus

$$M_{ij}^l = \begin{cases} 1 & \text{if features } x_i^l \ \& \ x_j^l \text{ observed,} \\ 0 & \text{otherwise} \end{cases}$$

Thus, we detect the block occlusions and consider only the observed pixels.

Case II: Random Missing Values: We use a Total Variation denoising procedure and calculate Ω_{ij} using the cleaned images.

Let ω_{min} be the minimum of all the pairwise distances in Ω. Then the adjacency matrix A for the graph G is constructed by using

$$A_{ij} = \exp\left(-\frac{(\Omega_{ij} - \omega_{min})^2}{\sigma^2}\right)$$

Finally, the normalized graph Laplacian $\Phi = I - D^{-1/2} A D^{-1/2}$ is calculated, where D is the degree matrix. Note that different types of data might call for different distance metrics and values of σ^2, however for all the experiments and datasets used in this work the Euclidean distance metric and $\sigma^2 = 0.05$ work well. All the models are evaluated using the same graph construction technique. A detailed example with corrupted and uncorrupted images is given in Fig. 11 of supplementary material.

7. Clustering in Low Dimensional Space

Fig. 1 presents experimental results for the first important application of our model, *i.e.* clustering. Fig. 2 illustrates the principal components Q for three classes of occluded CMU PIE data set in 3-dimensional space. Our model outperforms others in most of the cases with different types and levels of occlusions (please refer to Tables 4, 5 & 6 in the supplementary material for additional results). In the next few paragraphs we elaborate further on 1) the advantage of graph over non-graph models, 2) the advantage of low-rank graph over principal components graph. Throughout the description of our results, the comparison with RPCA is of specific interest because it is a special case of our proposed model.

7.1. Is the Graph Useful?

Our model performs better than k-means, standard PCA, NMF and RPCA which do not leverage the graph.

Example case I: CMU PIE database with no pose variation: Consider the case of CMU PIE dataset in Fig. 1. This dataset does not suffer from pose changes and we observe that our model attains as low as 0% error when there are no corruptions in the data. Furthermore, we attain lowest error even with the increasing levels of occlusions and missing pixels. This can also be observed visually from Fig. 2 where the principal components for our model are better separated than others.

Example case II: COIL20 database with pose variation: Our model outperforms RPCA and other non-graph models also for the COIL20 database which suffers from significant pose changes. Thus, even a graph constructed using the simple scheme of Section 6.3 enhances the robustness of our model to gross data corruptions and pose changes. Similar conclusions can be drawn for all the other databases as well. Please note that the results on YALE dataset are worse due to changes in the facial expressions in this dataset.

7.2. Low-Rank or Principal Components Graph?

We compare the performance of our model with NCuts, LE, GLPCA, RGLPCA, MMF & GNMF which use principal components graph. It is obvious from Fig. 1 that our model outperforms the others even for the datasets with pose changes. Similar conclusions can be drawn for all the other databases with corruptions and by visually comparing the principal components of these models in Fig. 2 as well.

Unlike factorized models, the principal directions U in the low-rank graph $\mathrm{tr}(L\Phi L^T)$ benefit from the graph regularization as well and show robustness to corruptions. This leads to a better clustering in the low-dimensional space even when the graph is constructed from corrupted data (please refer to Fig. 12 of the supplementary material for further experimental explanation).

7.3. Robustness to Graph Quality

We perform an experiment on the YALE dataset with 35% random missing pixels. Fig. 3 shows the variation in clustering error of different PCA models by using a graph constructed with decreasing information about the mask. Our model still outperforms others even though the quality of graph deteriorates with corrupted data. It is essential to note that in the worst case scenario when the mask is not known at all, our model performs equal to RPCA but still better than those which use a principal components graph.

8. Low-Rank Recovery
8.1. Low-Rank Recovery from Artificial Datasets

To perform a quantitative comparison of exact low-rank recovery between RPCA and our model, we perform experiments on 2 artificial datasets as in [6]. Note that only RPCA and our model perform exact low-rank recovery so we do not present results for other PCA models. We generate low-rank square matrices $L = A^T B \in \mathbb{R}^{n \times n}$ (rank $d =$
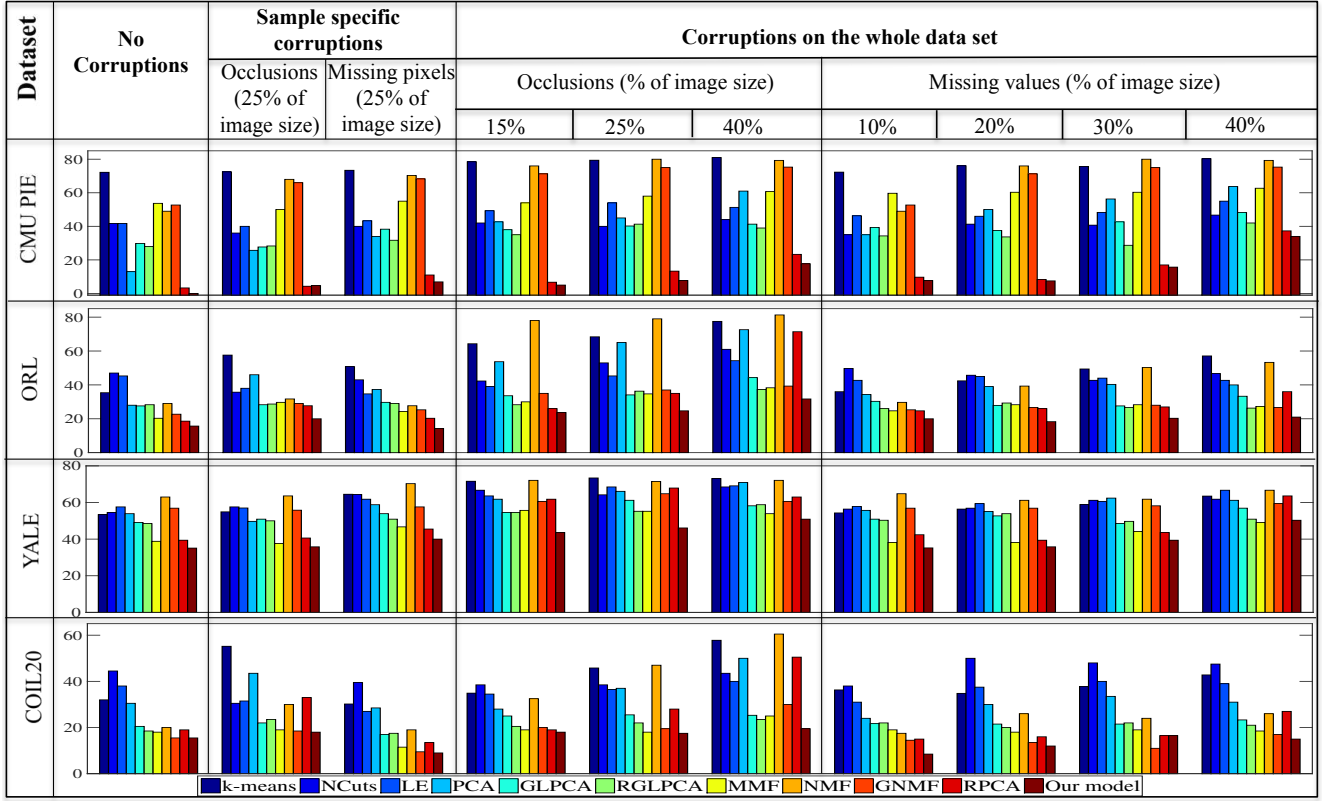
Figure 1: Clustering error of various dimensionality reduction models for CMU PIE, ORL, YALE & COIL20 datasets. The compared models are: 1) k-means without dimensionality reduction 2) Normalized Cuts (NCuts) [19] 3) Laplacian Eigenmaps (LE) [2] 4) classical Principal Component Analysis (PCA) 5) Graph Laplacian PCA (GLPCA) [10] 6) Robust Graph Laplacian PCA (RGLPCA) [10] 7) Manifold Regularized Matrix Factorization (MMF) [24] 8) Non-negative Matrix Factorization [13] 9) Graph Regularized Non-negative Matrix Factorization (GNMF) [5], 10) Robust PCA (RPCA) [6] and 11) Robust PCA on Graphs (proposed model). Two types of full and partial corruptions are introduced in the data: 1) Block occlusions and 2) Random missing values. The numerical errors corresponding to this figure along with additional results on MNIST, USPS, MFeat and BCI datasets are presented in Tables 4, 5 & 6 of the supplementary material.
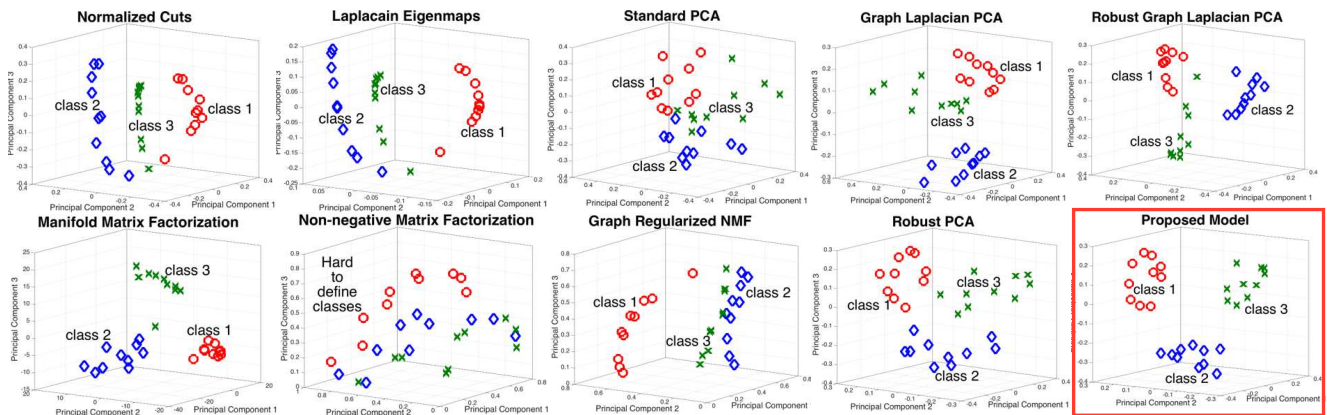

Figure 2: Principal components Q of three classes of CMU PIE data set in 3 dimensional space. Only ten instances of each class are used with block occlusions occupying 20% of the entire image. Our proposed model (on the lower right corner) gives well-separated principal components without any clustering error.
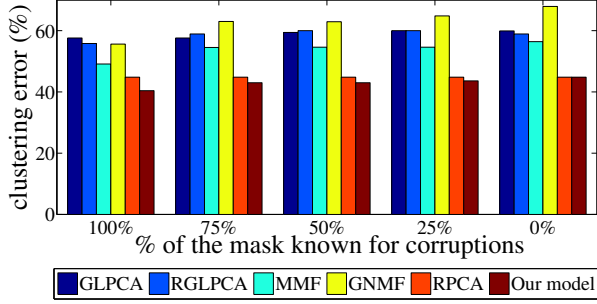
Figure 3: Effect of lack of information about the corruptions mask on the clustering error. YALE dataset with 35% random missing pixels is used for this experiment. Our model (last bar in each group) outperforms the others.

$0.02n$ to $0.3n$) with A and B independently chosen $d \times n$ matrices with i.i.d. Gaussian entries of zero mean and variance $1/n$. We introduce 6% to 30% errors $\rho = \|S\|_0/n^2$ in these matrices from an i.i.d. Bernoulli distribution for support of S with two sign schemes.

1. Random signs: Each corrupted entry takes a value ± 1 with a probability $\rho/2$.

2. Coherent signs: The sign for each corrupted entry is coherent with the sign of the corresponding entry in L.

Fig. 4 compares the variation of log normalized low-rank reconstruction error $\log(\|L - \hat{L}\|_F/\|L\|_F)$ with rank (x-axis) and error (y-axis) between RPCA (b and d) and our model (a and c). The larger and darker the blue region, the better is the reconstruction. The first two plots correspond to the random sign scheme and the next ones to coherent sign scheme. It can be seen that for each (rank,error) pair the reconstruction error for our model is less than that for RPCA. Hence, the low-rank graph helps in an enhanced low-rank recovery.

8.2. Low-Rank Recovery from Corrupted Faces

We use the PCA models to recover the clean low-rank representations from a set of occluded images of CMU PIE dataset as shown in Fig. 5. None of the factorized models using principal components graph is able to perfectly separate the block occlusion from the actual image. This is not surprising because these models (except RGLPCA) are not robust to gross errors. Our model is able to separate the occlusion better than all the models. Even though it inherits the robustness from RPCA, we observe that the robust recovery of the low-rank representation L is greatly enhanced by using the low-rank graph. Please refer to Fig. 13 in the supplementary material for more results.

8.3. Low-Rank Background Extraction from Videos

Static background separation from the dynamic foreground is an interesting application of low-rank recovery. We use 3 videos [8] (1000 frames each) to recover the static

low-rank background. The graph Laplacian Φ is constructed between the frames of videos without utilizing any prior information about sparse errors. The low-rank ground truth is not available for these videos so we present a visual comparison between RPCA and our model for one of the frames in Fig. 6. The RPCA model (middle) is unable to completely remove the person in the middle of the frame and his shadow from the low-rank frame. However, the presence of graph in our model (right) helps in a better recovery of the static background, which in this case is the empty shopping mall lobby. The pictures are best viewed on a computer screen on the electronic version of this article. **Complete videos for our model can be found here** [9] [10] [11]. For more results please refer to Fig. 14 of supplementary material.

9. Parameter Selection Scheme for Our Model

There are two parameters for our model: 1) Sparsity penalty λ and 2) Graph regularization penalty γ. The parameter λ can be set approximately equal to $1/\sqrt{\max(n, p)}$ where n is the number of data samples and p is the dimension. This simple rule, as suggested by Candes *et al.* [6] works reasonably well for our clustering & low-rank recovery error experiments. After fixing λ, the parameter γ can be easily selected by cross-validation. The minimum clustering error always occurs around this value of λ, irrespective of the size, number of classes and % of corruptions in the datasets. Due to lack of space we present the detailed results on the variation of clustering & low-rank recovery error over the (λ, γ) grid in Figs. 15 & 16 of the supplementary material.

10. Comparison of Computation Times

We corrupt different sizes of the CMU PIE dataset (n = 300, 600 and 1200) with 20% occlusions and compute the time for one run of each model which gives the minimum clustering error (Fig. 7). Clearly, RGLPCA has the highest computation time, followed by our proposed model. However, the trade-off between the clustering error and computational time is worth observing from Figs. 1 and 7 (more details in the supplementary material Table 7). The large computation time of our model is dominated by the expensive SVD step in every iteration.

11. Conclusion

We present 'Robust PCA on Graphs', a generalization of the Robust PCA framework which leverages spectral graph regularization on the low-rank representation. The proposed model targets exact low-rank recovery and enhanced clustering in the low-dimensional space from grossly corrupted datasets by solving a convex optimization problem with minimal parameters. Experiments on several

(a) Our model, random signs   (b) RPCA, random signs   (c) Our model, coherent signs   (d) RPCA, coherent signs
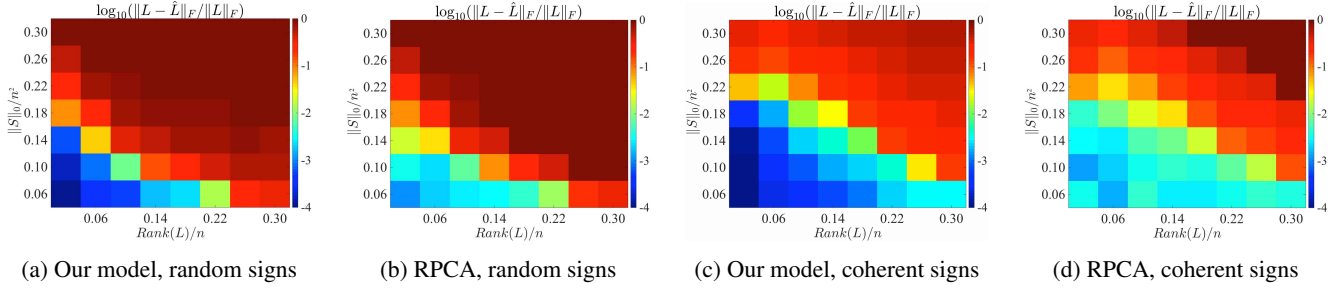
Figure 4: Variation of Log normalized low-rank reconstruction error on artificial data (Section 8.1) with (rank, error). The larger and darker the blue area, the lower reconstruction error of the model.

original   whitened (occluded, un-occluded)   GLPCA   RGLPCA   MMF   RPCA   Proposed
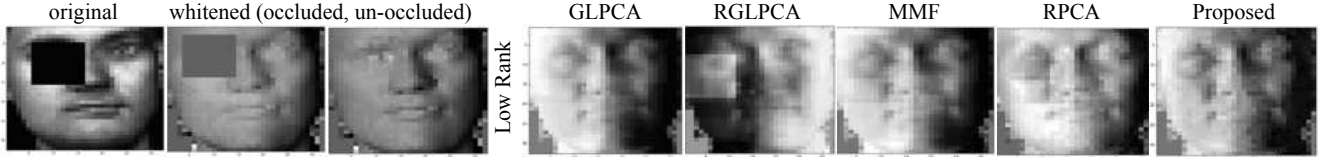
Figure 5: Clean low-rank recovery of one image of the CMU PIE data set corresponding to each of the PCA models. The images of one person are corrupted by 10% block occlusions. 1st figure shows the actual occluded image, 2nd and 3rd figures show the whitened occluded and un-occluded images. Since PCA requires whitening, the recovered low-rank images in figures 4 to 8 should resemble the un-occluded whitened image.

Figure 6: One frame of the recovered background from a video of a shopping mall lobby. left) actual frame, middle) recovered low-rank background using Robust PCA and right) recovered background using our proposed model. The presence of graph in our model enables better recovery and removes all the walking people from the frame.
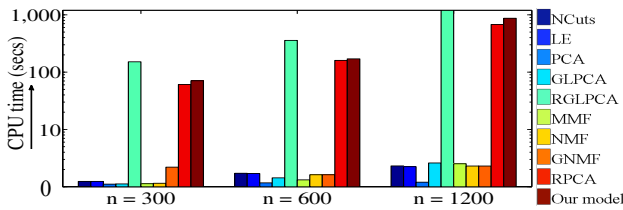

Figure 7: Computation time of different models for CMU PIE dataset with 20% occlusions.

benchmark datasets reveal that clustering using our low-rank graph regularization scheme outperforms various other state-of-the-art factorized PCA based models which use principal components graph regularization. Experiments for exact low-rank recovery from artificial datasets and static background separation from videos demonstrate its ability of to perform an enhanced low-rank recovery. Moreover, it is also robust w.r.t. the graph construction strategy and performs well even when the graph is constructed from corrupted data. Our future work will involve the theoretical investigation of the model and cost reduction by using randomized algorithms for SVD [21], parallel processing [18] or other methods [15, 16]. We will also work on foreground extraction and present quantitative results as in [3].

References

[1] Y. Aflalo, H. Brezis, and R. Kimmel. On the optimality of shape and data representation in the spectral domain. *arxiv.org*, September 2014. 2

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003. 1, 4, 6, 14, 16, 17

[3] T. Bouwmans and E. H. Zahzah. Robust pca via principal component pursuit: a review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014. 8

[4] S. Boyd, N. Parikhl, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers, found. trends mach. learn. 3 (1)(2011) 1–122, ht tp. *dx. doi. org/10.1561/2200000016*, 2010. 3, 10, 11

[5] D. Cai, X. He, J. Han, and T. S. Huang. Graph regularized nonnegative matrix factorization for data representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(8):1548–1560, 2011. 4, 6, 14, 16, 17

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 1, 2, 3, 4, 5, 6, 7, 11, 14, 16, 17, 20

[7] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011. 3, 10

[8] Z. Gao, L.-F. Cheong, and M. Shan. Block-sparse rpca for consistent foreground detection. In *Computer Vision–ECCV 2012*, pages 690–703. Springer, 2012. 20

[9] D. Goldfarb and S. Ma. Convergence of fixed-point continuation algorithms for matrix rank minimization. *Foundations of Computational Mathematics*, 11(2):183–210, 2011. 11

[10] B. Jiang, C. Ding, and J. Tang. Graph-laplacian pca: Closed-form solution and robustness. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3492–3498. IEEE, 2013. 1, 2, 3, 4, 6, 14, 16, 17

[11] T. Jin, J. Yu, J. You, K. Zeng, C. Li, and Z. Yu. Low-rank matrix factorization with multiple hypergraph regularizers. *Pattern Recognition*, 2014. 1, 2, 3

[12] S. Kontogiorgis and R. R. Meyer. A variable-penalty alternating directions method for convex optimization. *Mathematical Programming*, 83(1-3):29–53, 1998. 11

[13] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. 4, 6, 14, 16, 17

[14] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979. 11

[15] X. Liu, Z. Wen, and Y. Zhang. Limited memory block krylov subspace optimization for computing dominant singular value decompositions. *SIAM Journal on Scientific Computing*, 35(3):A1641–A1668, 2013. 8

[16] X. Liu, Z. Wen, and Y. Zhang. An efficient gauss-newton algorithm for symmetric low-rank product matrix approximations. 2014. 8

[17] X. Lu, Y. Wang, and Y. Yuan. Graph-regularized low-rank representation for destriping of hyperspectral images. *IEEE transactions on geoscience and remote sensing*, 51(7):4009–4018, 2013. 3

[18] A. Lucas, M. Stalzer, and J. Feo. Parallel implementation of fast randomized algorithms for low rank matrix decomposition. *Parallel Processing Letters*, 24(01), 2014. 8, 11

[19] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000. 4, 6, 14, 17

[20] L. Tao, H. H. Ip, Y. Wang, and X. Shu. Low rank approximation with sparse integration of multiple manifolds for data representation. *Applied Intelligence*, pages 1–17, 2014. 1, 2, 3

[21] R. Witten and E. Candes. Randomized algorithms for low-rank matrix factorizations: sharp performance bounds. *Algorithmica*, pages 1–18, 2013. 8, 11

[22] H. Xu, C. Caramanis, and S. Sanghavi. Robust pca via outlier pursuit. In *Advances in Neural Information Processing Systems*, pages 2496–2504, 2010. 1

[23] X. Yuan and J. Yang. Sparse and low-rank matrix decomposition via alternating direction methods. *preprint*, 2009. 11

[24] Z. Zhang and K. Zhao. Low-rank matrix approximation with manifold regularization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(7):1717–1729, 2013. 1, 2, 3, 4, 6, 14, 16, 17

[25] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma. Stable principal component pursuit. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1518–1522, June 2010. 1