Manifold Denoising by Nonlinear Robust Principal Component Analysis

He Lyu, Ningyu Sha, Shuyang Qin, Ming Yan, Yuying Xie, Rongrong Wang
Michigan State University

Summary and Contributions

This work extends Robust PCA [1] to manifold setting, where the observed data is the sum of a sparse component and a component drawn from some low dimensional manifold.
We propose an optimization framework that separates the sparse component from the manifold under noisy data.

- A theoretical guarantee for the method
- A curvature estimation method that may be of independent interest

Problem Formulation

Consider the following data model

$$\tilde{X} = X + S + E \tag{1}$$

- $\tilde{X} = [\tilde{X}_1, \dots, \tilde{X}_n] \in \mathbb{R}^{p\times n}$: noisy data
- X: clean data matrix lying on a manifold $M \subseteq \mathbb{R}^p$ with an intrinsic dimension $d \ll p$
- S: the matrix of the sparse noise
- E: the matrix of Gaussian noise

Key idea: use and integrate the local information.
We find the sparse noise S by solving

$$\min_{S, L^{(i)}} \sum_{i=1}^{n} (\lambda_i \|\tilde{X}^{(i)} - L^{(i)} - S^{(i)}\|_F^2 + \|\mathcal{C}(L^{(i)})\|_* + \beta \|S^{(i)}\|_1) \tag{2}$$

subject to $S^{(i)} = \mathcal{P}_i(S)$.

- $\mathcal{P}_i$ restricts the input to a neighbourhood around $\tilde{X}_i$
- $\tilde{X}^{(i)} = \mathcal{P}_i(\tilde{X})$, local patches
- $\mathcal{C}$: the centering operator

Remark: the constraints $S^{(i)} = \mathcal{P}_i(S)$ ensure that local sparse noises $S^{(i)}$ are restrictions of a global noise matrix, thus reducing the degree of freedom of $\{S^{(i)}\}_{i=1}^n$ to np, while the degree of freedom of $\{L^{(i)}\}_{i=1}^n$ is still knp.

For a subspace T, its *coherence* is defined as

$$\mu(V) = \frac{m}{r} \max_{k\in\{1,\dots,m\}} \|V^* \mathbf{e}_k\|_2^2.$$

where V is is an orthonormal basis of T.

Theoretical Error Bound

Theorem Suppose the support of the noise matrix $S^{(i)}$ is uniformly distributed among all sets of cardinality m_i, and $\bar{\mu}$ is the maximal coherence over all tangent spaces of M. Then as long as $d < \rho_r \min\{k,p\}\bar{\mu}^{-1}\log^{-2}\max\{k,p\}$, and $m_i \le 0.4\rho_s pk$ (ρ_r and ρ_s are positive constants), with probability over $1 - c_1 n \max\{k,p\}^{-10} - e^{-c_2 k}$, the minimizer $\hat{S}$ to (2) with weights

$$\lambda_i = \frac{\min\{k,p\}^{1/2}}{\epsilon_i}, \quad \beta = \max\{k,p\}^{-1/2} \tag{3}$$

has the error bound

$$\sum_i \|\mathcal{P}_i(\hat{S}) - S^{(i)}\|_{2,1} \le C\sqrt{pn}k\|\epsilon\|_2.$$

Here ϵ_i is the linear approximation error of $\tilde{X}^{(i)} - S^{(i)}$.

Algorithm

Once $\hat{S}$ is found, we use the following denoised local patches $\hat{L}^{(i)}_{\tau^*}$

$$\hat{L}^{(i)}_{\tau^*} = H_{\tau^*}(\mathcal{C}(\tilde{X}^{(i)} - \mathcal{P}_i(\hat{S}))) + (I - \mathcal{C})(\tilde{X}^{(i)} - \mathcal{P}_i(\hat{S})), \tag{4}$$

where H_{τ^*} is the Singular Value Hard Thresholding Operator with the optimal threshold as defined in [2]. We use the resulting $\hat{L}^{(i)}_{\tau^*}$ to construct a final estimate $\hat{X}$ of X via least squares fitting

$$\hat{X} = \arg\min_{Z\in\mathbb{R}^{p\times n}} \sum_{i=1}^{n} \lambda_i \|\mathcal{P}_i(Z) - \hat{L}^{(i)}_{\tau^*}\|_F^2 = (\sum_{i=1}^{n} \lambda_i \hat{L}^{(i)}_{\tau^*} P_i^T)(\sum_{i=1}^{n} \lambda_i P_i P_i^T)^{-1} \tag{5}$$

Algorithm 1: Nonlinear Robust PCA

Input: Noisy data matrix $\tilde{X}$, k (number of neighbors in each local patch), T (number of neighborhood updates)
Output: the denoised data $\hat{X}$, the estimated sparse noise $\hat{S}$

```
1 Estimate the curvature and the linear approximation error ε_i;
2 Estimate λ_i, i = 1, ..., n and β as in (3);
3 Ŝ ← 0;
4 for iter = 1: T do
5     Construct the restriction operators {P_i}_{i=1}^n using the kNN of X̃ − Ŝ;
6     Construct the local data matrices X̃^(i) = P_i(X̃)
7     Ŝ ← minimizer of (2);
8 end
9 Compute each L̂^(i)_τ* from (4) and assign X̂ from (5).
```

Numerical Experiments

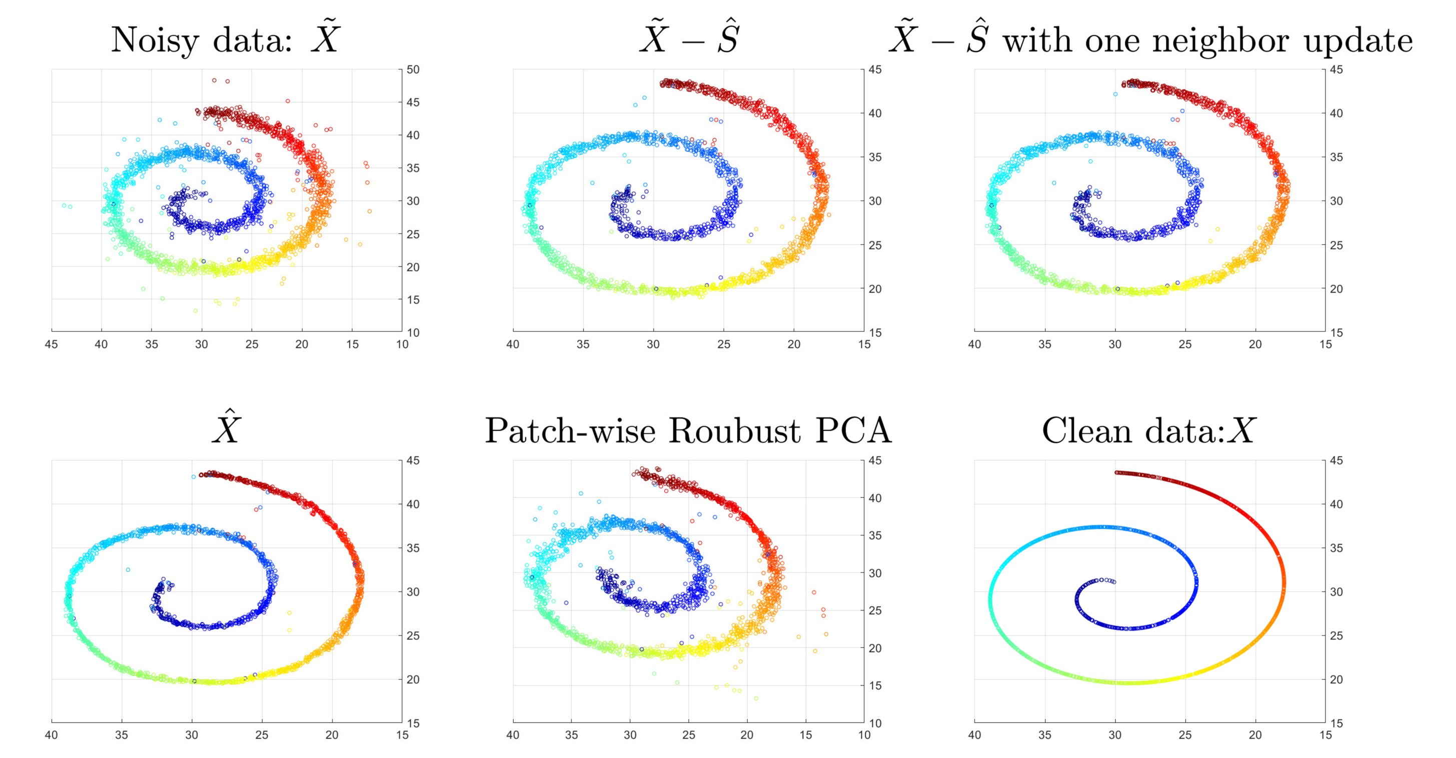

Figure 1: NRPCA applied to the noisy 3D Swiss roll dataset. $\tilde{X} - \hat{S}$ is the result after subtracting the sparse noise estimated by setting $T = 1$ in NRPCA, i.e., no neighbour update; "$\tilde{X} - \hat{S}$ with one neighbor update" used the $\hat{S}$ obtained by setting $T = 2$ in NRPCA; $\hat{X}$ is the data obtained via fitting the denoised tangent spaces as in (5). Compared to"$\tilde{X} - \hat{S}$ with one neighbor update", it further removed the Gaussian noise from the data; "Patch-wise Robust PCA" refers to the ad-hoc application of the vanilla Robust PCA to each local patch independently.

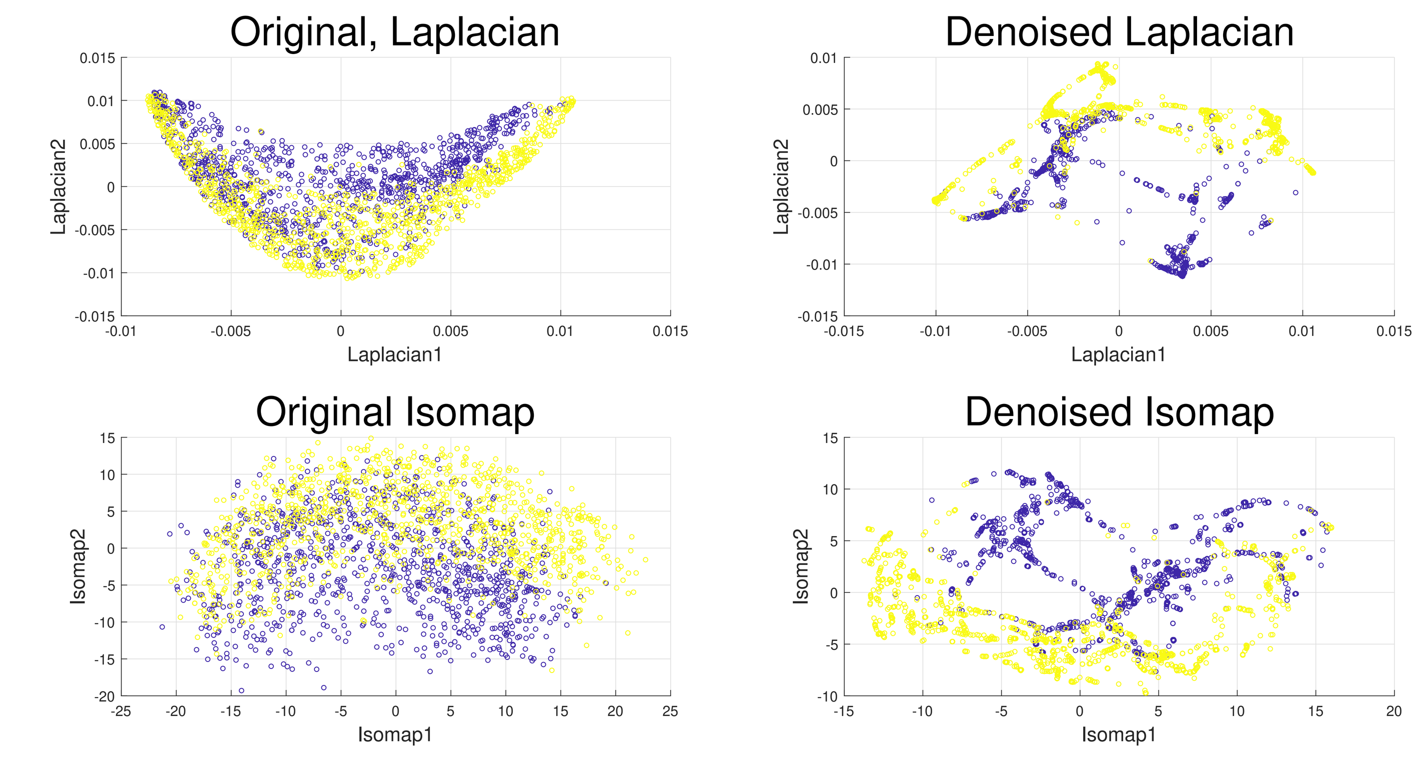

Figure 2: Laplacian eigenmaps and Isomap results for the original and the NRPCA denoised digits 4 and 9 from the MNIST dataset.

References

[1] Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis?. Journal of the ACM (JACM). 2011 May 1;58(3):11.

[2] Gavish M, Donoho DL. The optimal hard threshold for singular values is $4/\sqrt{3}$. IEEE Transactions on Information Theory. 2014 Jun 30;60(8):5040-53.