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

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

- $\tilde{X} = [\tilde{X}_1, \ldots, \tilde{X}_n] \in \mathbb{R}^{p \times n}$: 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

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

$$\text{st. } S^{(i)} = \mathcal{P}_i(S)$$

- $\mathcal{P}_i$ 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

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

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

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

has the error bound

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

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

Algorithm

Input: Noisy data matrix $\tilde{X}$, patch size k, number of iterations T
Output: The denoised data $\hat{X}$, the estimated sparse noise $\hat{S}$

- **Step 1.** For each $\tilde{X}_i$, randomly pick m points q_j lying within a proper distance to $\tilde{X}_i$, compute the corresponding radius R_{γ_j}, which is the radius of circular approximation to the geodesic joining $\tilde{X}_i$ and q_j. Estimate the average curvature

$$\bar{\Gamma}(\tilde{X}_i) \equiv \mathbb{E}(R_{\gamma_j}^{-2})^{1/2} \leftarrow \left(\frac{1}{m} \sum_{j=1}^{m} R_{\gamma_j}^{-2}\right)^{1/2}$$

Set ϵ_i as following, λ_i, $i = 1, \ldots, n$ and β as in (3), $\hat{S} \leftarrow 0$

$$\hat{\epsilon}_i := \left((k+1)p\sigma^2 + \sum_{j=1}^{k} \frac{\|\tilde{X}_i - \tilde{X}_{i_j}\|_2^4}{4} \bar{\Gamma}^2(\tilde{X}_i)\right)^{1/2}$$

- **Step 2.** Remove sparse noise
  for *iter* = 1: T **do**
  - Construct the restriction operators $\{\mathcal{P}_i\}_{i=1}^n$ using the kNN of $\tilde{X} - \hat{S}$;
  - Construct the local data matrices $\tilde{X}^{(i)} = \mathcal{P}_i(\tilde{X})$;
  - $\hat{S} \leftarrow$ minimizer of (2)

  end
- **Step 3.** Remove Gaussian noise using SVHT as in [2]

Numerical Experiments

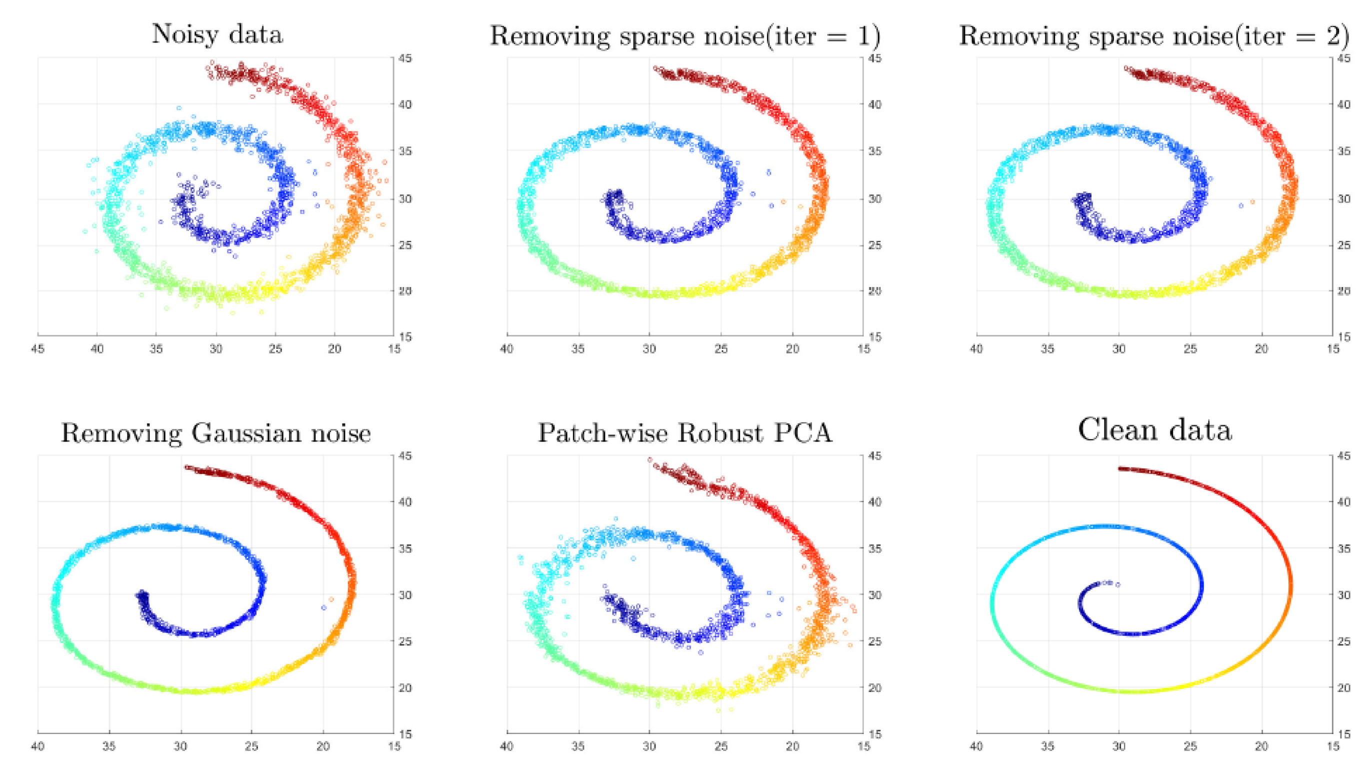

Figure 1: NRPCA applied to the noisy 3D Swiss roll dataset.

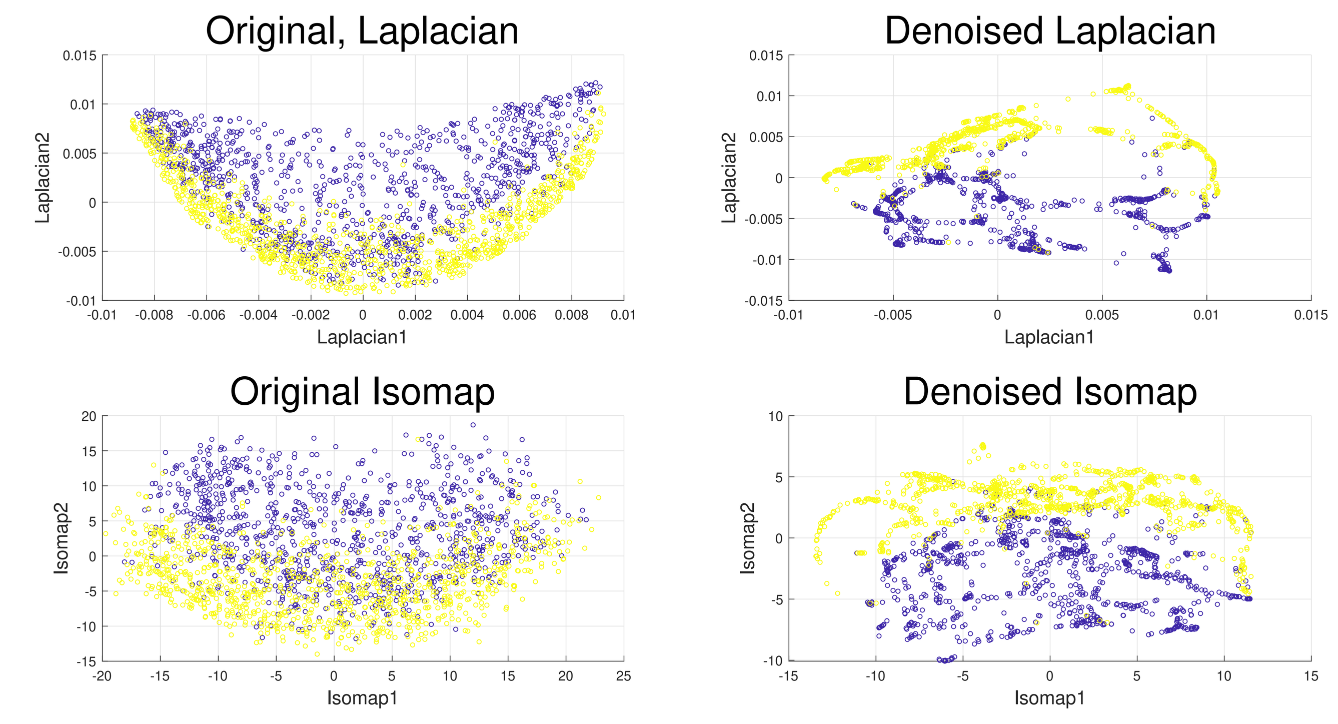

Figure 2: Laplacian eigenmaps and Isomap results for the original and the NRPCA denoised digits 4 and 9 from the MNIST dataset.

Figure 3: MNIST NRPCA denoising results

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.