Robust Tensor PCA

Fan Wu
School of Information Science and Technology
ShanghaiTech University
{wufan1}@shanghaitech.edu.cn

1. Introduction

Tensor can be used in many useful applacations, An array of numbers arranged on a regular grid with a variable number of axes is described as a tensor. When the dimensionality reaches beyond 3D, some widely-used data representations such as vector-based and matrix-based can be insufficient, Tensors are used as input formats for many computer visions. In some deep learning networks, tensor convolution is a very popular operation for extracting independent features. Until now, there have been many types of convolution operations in addition to conventional convolution. Derived the tensor SVD is another popular tool to get the Principal Component of a tensor. In the tensor SVD method, the rank of tensor can have many different define-tion, here we use the definition in [3], which characterizes the intrinsic "dimension" or "degree of freedom" for highly dimensional data sets. Robust Principal Component Analysis (RPCA) [7], aiming to recover the low-rank and sparse matrices both accurately and efficiently, has been widely studied in data compressed sensing and computer vision. Tensor Robust Principal Component Analysis (TRPCA)[1] extends the RPCA from matrix to the tensor case, that is, to recover a low-rank tensor and a sparse (noise entries) tensor.

2. Problem statement

Higher order tensor problems are far more than extension of matrices, tensor problems include more structures with high-dimensionality and computational difficulty. an N-way or Nth-order tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system.

2.1. Notation and preliminaries about tensors

Fibers are the higher order analogue of matrix rows and columns. Slices are two-dimensional sections of a tensor,

mode.png

Figure 1: Fibers of a 3rd-order tensor

defined by fixing all but two indices. The norm of a tensor

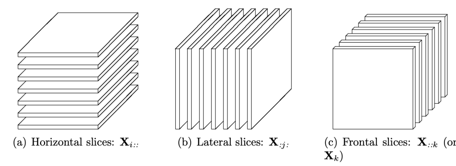

Figure 2: Slices of a 3rd-order tensor

$\tilde{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the square root of the sum of the squares of all its elements, i.e.,

$$\|\tilde{X}\| = \sqrt{\sum_{i=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}$$

which is analogous to the matrix Frobenius norm. The inner product of two same-sized tensor $\tilde{X}, \tilde{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the sum of the products of their entries, i.e.,

$$\langle \tilde{X}, \tilde{Y} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N} y_{i_1 i_2 \cdots i_N}$$

A mode-k matrix product is a special contraction that involves a matrix and a tensor. A new tensor of the same order is obtained by applying the matrix to each mode-k fiber of the tensor. For example, a tenser$\tilde{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$with a matrix $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ is denoted by $\tilde{X} \times_n \mathbf{U}$ and is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$,

$$(\tilde{X} \times_n \mathbf{U})_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{j_n}$$

2.2. Robust PCA

Robust PCA is analogous to traditional PCA but instead of recovering a low rank approximation of the matrix under some Gaussian noise assumption, it decomposes it as the sum of of a low rank matrix and a sparse one. Tensor Robust PCA further generalises this notion to tensors.

To minimise the Frobenius norm of the reconstruction error under orthogonality constraint over the projection weights, the problem can be modeled as a optimaztion problem following:

$$\begin{aligned} &\min_{W} \left\| X - XWW^T \right\|_F^2 \\ &\text{subject to } W^T W = I_d, \end{aligned} \tag{1}$$

where $X \in \mathbb{R}^{M\times N}$ be an N by M matrix, $W \in \mathbb{R}^{M\times K}, K \leq M$ is a orthonormal projection matrix the algorithm want to obtain. Rewrite $D = XWW^T$ as the reconstructed matrix, the problem can be equivalently write as:

$$\begin{aligned} &\min_{W} \|X - D\|_F^2 \\ &\text{subject to } \operatorname{rank}(D) \leq K \end{aligned} \tag{2}$$

Robust PCA aims at minimise the error under the assumption that the noise is grass but sparse. Instead of minimising an ℓ_2, minimise an ℓ_0 norm of the reconstruction error. In addition, instead of setting an upper bound on the rank of the reconstruction,

$$\min_{D} \operatorname{rank}(D) + \|X - D\|_0 \tag{3}$$

However, this problem is NP-hard, which means it cannot be solved in Polynomial time. Relax the problem using the convex surrogate of the ℓ_0 norm and the rank operator as proposed by Cantes in his paper[2], This turns the problem into a tractable one:

$$\min_{D} \|D\|_* + \lambda \|X - D\|_1, \tag{4}$$

λ controls how sparse the error matrix will be and needs to be tuned to each particular data.
Robust Tensor PCA generalises this problem to tensors. Instead of minimizing the nuclear norm of the data tensor $\tilde{X}$, minimise the nuclear norm of its unfoldings, $\tilde{X}_{[i]}$ along each mode i,$i \in [1, 2, 3, \ldots, N]$

$$\min_{\tilde{D}} \sum_{i=1}^{N} \left\| \tilde{D}_{[i]} \right\|_* + \lambda \|\tilde{X} - \tilde{D}\|_1 \tag{5}$$

3. Generalized Robust Tensor PCA

The original PCA[4] is proposed for reducing the dimensions of vector data, which exploits low-dimensional structure in high-dimensional data, One major shortcoming of RPCA is that it can only handle 2-way (matrix) data. However, real data is usually multi- dimensional in nature-the information is stored in multi-way arrays known as tensors.[6] proposed a model based on the recently proposed tensor-tensor product (or t-product) [5]. There are already many algorithms trying to construct on the data structure of tensor. Although the amount of calculation is one of the main reasons that limit its development at present, the accuracy and speed of three-dimensional data are not as good as vectors and matrices, the tensor algebra structure is also very complicated, but tensor can better explain the data relationship in the high-dimensional nonlinear space.

Robust Tensor PCA via ADMM can deal with case that support for missing values and outlier noises, which decomposes a tensor $\mathcal{X}$ into the sum of a low-rank component $\mathcal{L}$ and a sparse component $\mathcal{E}$.

$$\min_{\mathcal{L},\mathcal{E}} \|\mathcal{L}\|_* + \lambda \|\mathcal{E}\|_1, \text{ s.t. } \mathcal{X} = \mathcal{L} + \mathcal{E} \tag{6}$$

Algorithm 1 Solve (6) by ADMM

1: Initialize:

$$\mathcal{L}_0 = \mathcal{S}_0 = \mathcal{Y}_0 = 0, \rho = 1.1, \mu_0 = 1\mathrm{e}-3,$$

$$\mu_{\max} = 1\mathrm{e}10, \epsilon = 1\mathrm{e}-8$$

while

2: **while** not converged **do**

3: Update$\mathcal{L}_{k+1}$by

$$\mathcal{L}_{k+1} = \underset{\mathcal{L}}{\operatorname{argmin}} \lambda \|\mathcal{L}\|_* + \frac{\mu_k}{2} \left\| \mathcal{L} + \mathcal{E}_k - \mathcal{E} - \mathcal{X} + \frac{\mathcal{Y}_k}{\mu_k} \right\|_F^2$$

4: Update$\mathcal{E}_{k+1}$by

$$\mathcal{E}_{k+1} = \underset{\mathcal{E}}{\operatorname{argmin}} \lambda \|\mathcal{E}\|_1 + \frac{\mu_k}{2} \left\| \mathcal{L} + \mathcal{E}_{k+1} - \mathcal{E} - \mathcal{X} + \frac{\mathcal{Y}_k}{\mu_k} \right\|_F^2$$

5: $\mathcal{Y}_{k+1} = \mathcal{Y}_k + \mu_k(\mathcal{L}_{k+1} + \mathcal{E}_{k+1} - \mathcal{X})$

6: Update μ_{k+1} by $\mu_{k+1} = \min(\rho\mu_k, \mu_{\max})$

7: Check the convergence conditions

$$\|\mathcal{L}_{k+1} - \mathcal{L}_{k+1} - \mathcal{X}\| \, \mathcal{E}_{k+1} - \mathcal{E}_k\|_\infty \leq \epsilon$$

$$\|\mathcal{L}_{k+1} + \mathcal{E}_{k+1} - \mathcal{X}\|_\infty \leq \epsilon$$

8: **end while**

4. Case study

4.1. apply Robust Tensor PCA on the YaleB dataset

The YaleB dataset consists images of the faces of several people taken from the same angle but with different illumniations. The robustPCA functions returns two tensors with identical dimensions to the tensor decomposed: the low rank part and the sparse part, so the experiment design two kind noise add to the original data. first kind noise is the salt-pepper noise, second kind is combination of salt-pepper noise and patch-noise. Salt-and-pepper noise refers to two kinds of noise, one is salt noise and the other is pepper noise. Salt = white (0), pepper = black (255). The former is high gray noise and the latter is low gray noise. Generally, two kinds of noise appear at the same time, and appear as black and white noise on the image. The patching noise means there exist a small region where the pixel values are missing. The results show that tensor RPCA still work well on such task under the assumption the noise are sparse.

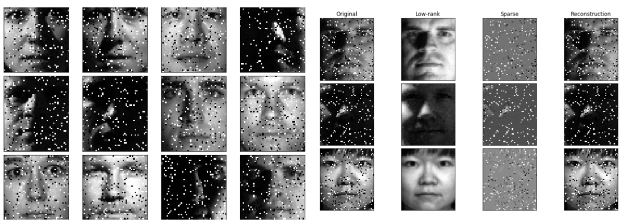

Figure 3: Add salt-pepper noise, apply tensor PCA

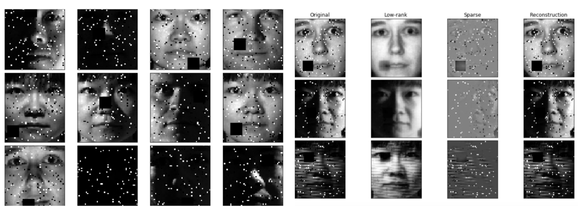

Figure 4: Add salt-pepper-mask, apply tensor PCA

4.2. Apply Robust Tensor PCA on a video

Use a video download from website, then apply the robust tensor PCA to the frame obtained from the video to distinguish the background and the cars

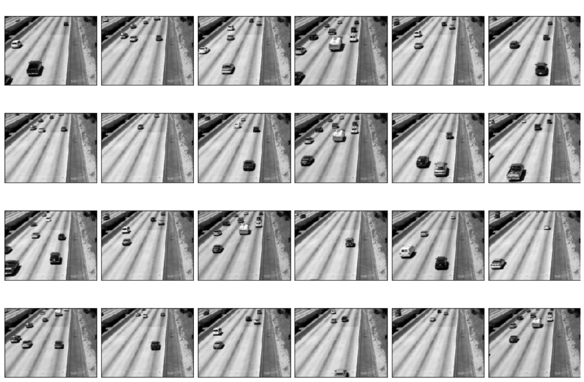

Figure 5: highway video

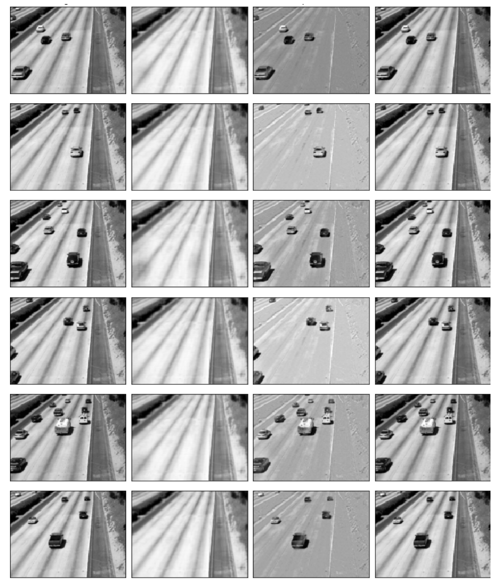

Figure 6: Apply tensor-PCA

References

[1] S. Cai, Q. Luo, M. Yang, W. Li, and M. Xiao. Tensor robust principal component analysis via non-convex low rank approximation. *Applied Sciences*, 9:1411, 04 2019.

[2] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[3] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, M. Sugiyama, and D. P. Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. *Foundations Trends® in Machine Learning*, 9(6):249–429.

[4] I. T. Jolliffe and J. Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[5] M. E. Kilmer and C. D. Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.

[6] C. Lu, J. Feng, W. Liu, Z. Lin, S. Yan, et al. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[7] M. Partridge and M. Jabri. Robust principal component analysis. In *Neural Networks for Signal Processing X. Proceedings of the 2000 IEEE Signal Processing Society Workshop (Cat. No.00TH8501)*, 2002.