TENSOR ERROR CORRECTION FOR CORRUPTED VALUES IN VISUAL DATA

Yin Li, Yue Zhou, Junchi Yan, Jie Yang

Institute of Image Processing & Pattern Recognition
Shanghai Jiaotong University

Xiangjian He

Faculty of Engineering and Information Technology
University of Technology Sydney

ABSTRACT

The multi-channel image or the video clip has the natural form of tensor. The values of the tensor can be corrupted due to noise in the acquisition process. We consider the problem of recovering a tensor L of visual data from its corrupted observations $X = L + S$, where the corrupted entries S are unknown and unbounded, but are assumed to be sparse. Our work is built on the recent studies about the recovery of corrupted low-rank matrix via trace norm minimization. We extend the matrix case to the tensor case by the definition of tensor trace norm in [6]. Furthermore, the problem of tensor is formulated as a convex optimization, which is much harder than its matrix form. Thus, we develop a high quality algorithm to efficiently solve the problem. Our experiments show potential applications of our method and indicate a robust and reliable solution.

Index Terms— tensor decomposition, trace norm minimization, sparse coding, convex optimization

1. INTRODUCTION

Error correction for the visual data has long been an active topic in the field of image processing and computer vision. Typically, the visual data, either an image or a video, has the natural form of a multi-dimensional matrix, namely the tensor. For example, the (color) channels of an image, or the frames of a video make up a 3-mode tensor. On one hand, one may seek to recover the noise-free or even the latent low dimensional version of the data, such as image denoising or image compression [1]. On the other hand, one may seek to detect the noisy or the salient part of the data, such as outlier detection [2] or foreground segmentation [3]. Therefore, it is necessary to develop tools to capture both the global structure and the salient part of the tensor-like visual data.

In the two-dimensional case, i.e. the matrix case, the "rank" plays an important part in capturing the global information of visual data. One simple and useful assumption is that the data lie near certain low-dimensional subspace, which is closely related to the minimization of rank. Although the "rank" itself is nonconvex, it can be approximated by its convex envelop, namely the trace norm. The validation of this approximation is justified in theory [4]. Among all the trace norm minimization problems, matrix completion may be a well-known one [4, 5]. Recently, [6] extends the famous problem of matrix completion to the tensor case and develop a efficient solution.

The "sparsity" is also a useful tool for visual data analysis, since the noisy or the salient part usually occupies a small portion of the data. The sparse prior has demonstrated a wide range of applications including image denoising [7], saliency detection [8] and face recognition [9]. It was not until very recently that had much attention been focused on the rank-sparsity problem for matrix [10, 11], namely the Principal Component Pursuit (PCP) or the Robust Principal Component Analysis (RPCA). These work seek to directly decompose a matrix into a low-rank part plus a sparse part. Theoretic analysis [11] shows that under rather weak assumptions, the problem can be solved by trace norm and l_1 norm minimization.

Although the PCP problem has been studied for matrix, there is no much work on tensors, which can be treated as higher order matrix. Mainly inspired by the work of [6, 10], we extend the matrix case to the tensor case by the definition of the tensor trace norm in [6]. Through relaxation technique, the problem of tensor decomposition is formulated as a convex optimization, which is much harder than its matrix form. Thus, we develop a block coordinate descent (BCD) based high quality algorithm to efficiently solve this problem. Furthermore, we explore preliminary applications of our method.

The paper is organized as follows. In Section 2, we briefly review the problem of Principal Component Pursuit in the matrix case. In Section 3, we extend the problem into the tensor case and present our proposed algorithm. In Section 4, we explore several potential applications of our algorithm. Finally, Section 5 concludes the paper.

2. PRINCIPAL COMPONENT PURSUIT

Our derivation is inspired by the recently emerging Principal Component Pursuit (PCP) problem [10]. The PCP assumes that the observed data matrix $D \in R^{l \times m}$ was generated by corrupting some of the entries of a low-rank matrix $A \in R^{l \times m}$. The corruption is produced by an additive error

The work is supported by National Science Foundation of China (Grant No.60772097); Open Projects Program of National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science.

$E \in R^{l \times m}$, so that $D = A + E$. The error is unknown and unbounded, but affects only a portion of the entries of D, so that E is a sparse matrix. Then the idealized PCP problem can be formulated as follows:

Problem: *Given $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$, where $\boldsymbol{A}$ is an unknown low-rank matrix and $\boldsymbol{E}$ is an unknown sparse matrix, recover $\boldsymbol{A}$ and $\boldsymbol{E}$.*

This problem formulation suggests a conceptual solution: seek the lowest rank A that could have generated the data, subject to the constraint that the errors are sparse: $\|E\|_0 \leq k$. The Lagrangian reformulation of the optimization problem is:

$$\min \ \ rank(\boldsymbol{A}) + \lambda \|\boldsymbol{E}\|_0 \ \ s.t \ \ \boldsymbol{A} + \boldsymbol{E} = \boldsymbol{D} \tag{1}$$

λ balances between rank and sparsity [11]. For an appropriate λ, the solution of the (1) recover the pair (A_0, E_0) that generated the data D. However, (1) is a highly nonconvex optimization and no efficient solution is guaranteed.

Given the fact that the trace norm $\|A\|_{tr}$ and l_1 norm $\|E\|_1$ are the convex envelop of $rank(A)$ and $\|E\|_0$ respectively, one can relax $rank(A) + \lambda\|E\|_0$ by $\|A\|_{tr} + \lambda\|E\|_1$ and obtain a tractable optimization problem:

$$\min \ \ \|\boldsymbol{A}\|_{tr} + \lambda \|\boldsymbol{E}\|_1 \ \ s.t \ \ \boldsymbol{A} + \boldsymbol{E} = \boldsymbol{D} \tag{2}$$

where the trace norm, or the nuclear norm of matrix A is defined as the sum of its singular values σ_j, i.e. $\|\boldsymbol{A}\|_{tr} = \sum_j \sigma_j(\boldsymbol{A})$. Moreover, recent theoretic analysis [11] indicates that under rather weak assumptions, (2) is able to exactly recover the low-rank matrix A_0 and sparse matrix E_0 in high probability. Various algorithms exist [10, 11, 12].

3. TENSOR DECOMPOSITION BY PRINCIPAL COMPONENT PURSUIT

In this section, we present our extension of the PCP problem into the tensor case. Although the new problem is much more difficult than the matrix case, we rely on the relaxation technique and propose an efficient solution.

3.1. Tensor Error Correction

A n-mode tensor is defined as $X \in R^{I_1 \times I_2 \times \ldots \times I_n}$, with its elements $x_{i_1} \ldots x_{i_k} \ldots x_{i_n} \in R$. Therefore, a vector can be seen as a 1-mode tensor and a matrix can be seen as a 2-mode tensor. It is often convenient to convert a tensor into a matrix, also called matricizing or unfolding. The "unfold" operation along the k-th mode on a tensor X is defined as $unfold(X, k) := X_{(k)} \in R^{I_k \times (I_1 \ldots I_{k-1} I_{k+1} \ldots I_n)}$. Accordingly, its inverse operator $fold$ can be defined as $fold(X_{(k)}, k) := X$. Moreover, denote $\|X\|_F := (\sum_{i_1, i_2, \ldots i_n} |x_{i_1 i_2 \ldots i_n}|^2)^{\frac{1}{2}}$ and $\|X\|_1 := \sum_{i_1, i_2, \ldots i_n} |x_{i_1 i_2 \ldots i_n}|$ as the Frobenius norm and l_1 norm the of a tensor. Then, we have $\|X\|_F = \|X_{(k)}\|_F$ and $\|X\|_1 = \|X_{(k)}\|_1$ for any $1 \leq k \leq n$.

Since the tensor is a higher dimensional extension of matrix, the PCP algorithm can be extended to the tensor case by solving the following optimization:

$$\begin{aligned} \min_{L,S} \quad & \|L\|_{tr} + \lambda \|S\|_1 \\ s.t. \quad & L + S = X \end{aligned} \tag{3}$$

where X, L and S are n-mode tensors with identical size in each mode. X is the observed data tensor. L and S represent the correspondent low rank part and sparse part, respectively. The remaining issue is the definition of trace norm for general tensor case.

However, the notion of trace norm for tensors of order greater than two is subtle. For example, there are alternative approaches for the higher order SVD [13, 14], leading to different definition of the trace norm. We propose the following definition for the tensor trace norm as in [6]

$$\|X\|_{tr} := \frac{1}{n} \sum_{i=1}^{n} \|X_{(i)}\|_{tr} \tag{4}$$

Essentially, the trace norm of a tensor is the average of the trace norms of all matrix unfolded along each mode. In particular, when the mode $n = 2$ (i.e. the matrix case), this definition is consistent with the matrix trace norm, since $X_{(1)}^T = X_{(2)}$. Under this definition, the optimization in (3) can be rewritten as:

$$\begin{aligned} \min_{L,S} \quad & \frac{1}{n} \sum_{i=1}^{n} \|L_{(i)}\|_{tr} + \lambda \|S\|_1 \\ s.t. \quad & L + S = X \\ & unfold(L_{(i)}, i) = L \qquad \forall i \end{aligned} \tag{5}$$

3.2. Simplified Formulation

Problem (5) is hard to solve due to the interdependent trace norm and l_1 norm constraint. To simplify the problem, we introduce additional auxiliary matrix $M_i = L_{(i)}$ and $N_i = S_{(i)}$. Thus, we obtain the equivalent formulation:

$$\begin{aligned} \min_{L,S,M_i,N_i} \quad & \frac{1}{n} \sum_{i=1}^{n} \|M_i\|_{tr} + \frac{\lambda}{n} \sum_{i=1}^{n} \|N_i\|_1 \\ s.t. \quad & L + S = X \\ & M_i = L_{(i)} \qquad \forall i \\ & N_i = S_{(i)} \qquad \forall i \end{aligned} \tag{6}$$

In (6), the constrains $M_i = L_{(i)}$ and $N_i = S_{(i)}$ still enforce the consistency of all M_i and N_i. Thus, we further relax the equality constrains $M_i = L_{(i)}$ and $N_i = S_{(i)}$ by $\|M_i - L_{(i)}\|_F \leq \varepsilon_1$ and $\|N_i - S_{(i)}\|_F \leq \varepsilon_2$. Furthermore, if we allow a dense noise term over X, we can relax $L + S = X$ by $\|L + S - X\|_F \leq \varepsilon_3$, corresponding to the stable Principle

Component Pursuit (SPCP) in the matrix case [15]. Therefore, we get the relaxed form:

$$
\begin{aligned}
\min_{L,S,M_i,N_i} \quad & \frac{1}{n}\sum_{i=1}^{n}\|M_i\|_{tr} + \frac{\lambda}{n}\sum_{i=1}^{n}\|N_i\|_1 \\
s.t. \quad & \|M_i - L_{(i)}\|_F^2 \le \varepsilon_1 \qquad \forall i \\
& \|N_i - S_{(i)}\|_F^2 \le \varepsilon_2 \qquad \forall i \\
& \|L + S - X\|_F^2 \le \varepsilon_3
\end{aligned}
\tag{7}
$$

For certain α_i, β_i and γ_i, (7) can be converted to its dual form by Lagrange multiplier.

$$
\begin{aligned}
\min_{L,S,M_i,N_i} \quad & \frac{1}{2n}\sum_{i=1}^{n}\alpha_i\|M_i - L_{(i)}\|_F^2 \\
& + \frac{1}{2n}\sum_{i=1}^{n}\beta_i\|N_i - S_{(i)}\|_F^2 \\
& + \frac{1}{2n}\sum_{i=1}^{n}\gamma_i\|M_i + N_i - X_{(i)}\|_F^2 \\
& + \frac{1}{n}\sum_{i=1}^{n}\|M_i\|_{tr} + \frac{\lambda}{n}\sum_{i=1}^{n}\|N_i\|_1
\end{aligned}
\tag{8}
$$

Intuitively, the weights α_i, β_i and γ_i indicate the preference towards different 'unfold' operation, i.e. the configuration of the tensor. For example, we would prefer to explain the tensor representation of a video as the collection of frames. The optimization problem in (8) is convex but nondifferentiable. Next, we show how to solve this problem.

3.3. The Proposed Algorithm

We propose to employ the alternative direction strategy for the optimization (8), leading to a block coordinate descent (BCD) algorithm. The core idea of the BCD is to optimize a group of variables while fixing the other groups. The variables in the optimization are N_1,...,N_n, M_1,...,M_n,L,S, which can be divided into $2n+2$ blocks. To achieve the optimum solution, we estimate N_i, M_i, L and S sequentially in each iteration. For clarity, we first define the "shrinkage" operator $D_\tau(x)$ by

$$
D_\tau(x) = \begin{cases} x - \tau & \text{if } x > \tau \\ x + \tau & \text{if } x < -\tau \\ 0 & \text{otherwise} \end{cases}
\tag{9}
$$

The operator can be extended to the matrix or tensor case by performing the shrinkage operator towards each element.

Computing N_i: The optimum N_i with all other variables fixed is the solution to the following subproblem

$$
\begin{aligned}
\min_{N_i} \quad & \frac{\beta_i}{2}\|N_i - S_{(i)}\|_F^2 + \frac{\gamma_i}{2}\|M_i + N_i - X_{(i)}\|_F^2 \\
& + \lambda\|N_i\|_1
\end{aligned}
\tag{10}
$$

$Algorithm 1$ (RSTD:Rank Sparsity Tensor Decomposition)

$\boldsymbol{Input:}$ X

$\boldsymbol{Output:}$ L, S, M_i, N_i from 1 to n

1. Set $L = X$, $S = 0$, $M_i = L_{(i)}$, $N_i = 0$
2. $\boldsymbol{while}$ no convergence
3. $\quad \boldsymbol{for}$ $i = 1$ to n
4. $\qquad N_i^* = D_{\frac{\lambda}{\beta_i+\gamma_i}}(\frac{\beta_i S_{(i)} + \gamma_i(X_{(i)} - M_i)}{\beta_i + \gamma_i})$
5. $\qquad M_i^* = UD_{\frac{1}{\alpha_i+\gamma_i}}(\Lambda)V^T$
6. $\quad \boldsymbol{end\ for}$
7. $\quad S^* = \frac{\sum_{i=1}^{n}\beta_i fold(N_i,i)}{\sum_{i=1}^{n}\beta_i}$
8. $\quad L^* = \frac{\sum_{i=1}^{n}\alpha_i fold(M_i,i)}{\sum_{i=1}^{n}\alpha_i}$
9. $\boldsymbol{end\ while}$

By the well-known l_1 minimization [16], the global minimum of the optimization problem in (10) is given by

$$
N_i^* = D_{\frac{\lambda}{\beta_i+\gamma_i}}(\frac{\beta_i S_{(i)} + \gamma_i(X_{(i)} - M_i)}{\beta_i + \gamma_i})
\tag{11}
$$

where D_τ is the "shrinkag" operation.

Computing M_i: The optimum M_i with all other variables fixed is the solution to the following subproblem:

$$
\begin{aligned}
\min_{M_i} \quad & \frac{\alpha_i}{2}\|M_i - L_{(i)}\|_F^2 + \frac{\gamma_i}{2}\|M_i + N_i - X_{(i)}\|_F^2 \\
& + \|M_i\|_{tr}
\end{aligned}
\tag{12}
$$

From current trace norm minimization literature [5], the global minimum of the optimization problem in (12) is given by

$$
M_i^* = UD_{\frac{1}{\alpha_i+\gamma_i}}(\Lambda)V^T
\tag{13}
$$

where $U\Lambda V^T$ is the singular value decomposition given by

$$
U\Lambda V^T = \frac{\alpha_i L_{(i)} + \gamma_i(X_{(i)} - N_i)}{\alpha_i + \gamma_i}
\tag{14}
$$

Computing S_i: The optimum S with all other variables fixed is the solution to the following subproblem

$$
\min_{S} \quad \frac{1}{2}\sum_{i=1}^{n}\beta_i\|N_i - S_{(i)}\|_F^2
\tag{15}
$$

It is easy to show that the solution to (15) is given by

$$
S^* = \frac{\sum_{i=1}^{n}\beta_i fold(N_i, i)}{\sum_{i=1}^{n}\beta_i}
\tag{16}
$$

Computing L_i: The optimum L with all other variables fixed is the solution to the following subproblem

$$
\min_{L} \quad \frac{1}{2}\sum_{i=1}^{n}\alpha_i\|M_i - L_{(i)}\|_F^2
\tag{17}
$$

Similar to (15), the solution to (17) is given by

$$L^* = \frac{\sum_{i=1}^{n} \alpha_i fold(M_i, i)}{\sum_{i=1}^{n} \alpha_i} \quad (18)$$

We call the proposed algorithm Rank Sparsity Tensor Decomposition (RSTD). The pseudo-code of RSTD is summarized in Algorithm.1. We choose the difference of L and S in consecutive iterations as the stopping criterion. We can further show that BCD for RSTD is guaranteed to reach the global optimum, since the first three terms in (8) are differentiable and the last two terms are separable [17].

4. EXPERIMENTS

It is easy to check that with appropriate weights α_i,β_i,γ_i, our method can approximate the solution in [10], leading to applications such as **background modeling** for video or **shadow removal** for faces. Besides the aforementioned applications, we also present preliminary results for image denoising.

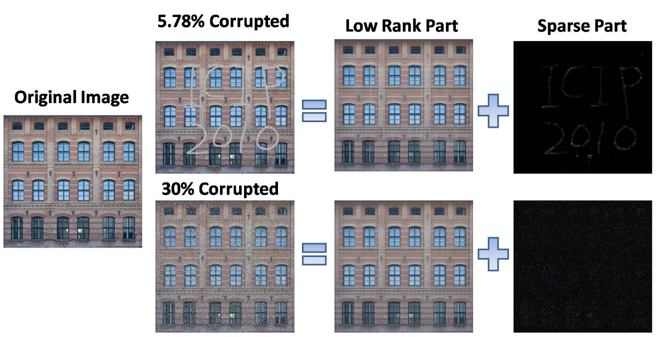

Fig. 1. RSTD on noisy images (Zoom in for better details).

Image Denoising: One straightforward application of our method is the image denoising task. We tested our algorithm on different images and the results are surprising. Fig.1 demonstrates some of the results. Our algorithm is able to blindly separate a reasonable low-rank image from the noise, even if the data is grossly corrupted.

5. CONCLUSION AND FUTURE WORK

In this paper, we extend the Principal Component Pursuit to the tensor case and design a highly efficient algorithm for the problem. To our best knowledge, we are the first to propose a solution for PCP problem in the tensor case. We are currently working on the efficiency of our algorithm, e.g., numerical approximations of a few largest singular values, since large-scale full SVD is slow and unnecessary. We would like to further explore additional applications and to investigate the theoretical side of our method in the future work.

References

[1] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing, 3nd ed*, Prentice Hall, 2008.

[2] O. Boiman and M. Irani, "Detecting irregularities in images and in video," in *ICCV*, Oct. 2005.

[3] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.

[4] Emmanuel J. Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Found. of Comput. Math*, vol. 9, pp. 717–772, 2008.

[5] Jian-Feng Cai, Emmanuel J. Candes, and Zuowei Shen, "A singular value thresholding algorithm for matrix completion," *preprint*, 2008.

[6] Ji Liu, Przemyslaw Musialski, Peter Wonka, and Jieping Ye, "Tensor completion for estimating missing values in visual data," *ICCV*, Oct. 2009.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[8] Yin Li, Yue Zhou, Lei Xu, and Jie Yang, "Incremental sparse saliency detection," *ICIP*, Sep. 2009.

[9] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[10] John Wright, Arvind Ganesh, Shankar Rao, Yigang Peng, and Yi Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," *Neural Information Processing Systems (NIPS)*, Dec. 2009.

[11] Emmanuel Cands, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *preprint*, Dec. 2009.

[12] Zhouchen Lin, Minming Chen, Leqin Wu, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep., Dep of Computer Science, UIUC, Dec. 2009.

[13] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[14] M. A. O. Vasilescu and Demetri Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *ECCV '02*, London, UK, 2002, pp. 447–460, Springer-Verlag.

[15] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel Candes, and Yi Ma, "Stable principal component pursuit," *preprint*, 2010.

[16] Elaine T. Hale, Wotao Yin, and Yin Zhang, "Abstract. fixed-point continuation for l1-minimization: Methodology and convergence," *SIAM Journal on Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.

[17] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.