SURE based truncated tensor nuclear norm regularization for low rank tensor completion
Morison, Gordon

Published in:
Proceedings of the 28th European Signal Processing Conference

Publication date:
2020

Document Version
Peer reviewed version

Link to publication in ResearchOnline

Citation for published version (Harvard):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please view our takedown policy at https://edshare.gcu.ac.uk/id/eprint/5179 for details of how to contact us.
SURE Based Truncated Tensor Nuclear Norm Regularization for Low Rank Tensor Completion
Morison, Gordon

Publication date:
2020

Document Version
Peer reviewed version

Link to publication in ResearchOnline

Citation for published version (Harvard):
Morison, G 2020, ‘SURE Based Truncated Tensor Nuclear Norm Regularization for Low Rank Tensor Completion’.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please view our takedown policy at https://edshare.gcu.ac.uk/id/eprint/5179 for details of how to contact us.
SURE Based Truncated Tensor Nuclear Norm Regularization for Low Rank Tensor Completion

Gordon Morison
Department of Computing
Glasgow Caledonian University
Glasgow, Scotland, UK
gordon.morison@gcu.ac.uk

Abstract—Low rank tensor completion aims to recover the underlying low rank tensor obtained from its partial observations, this has a wide range of applications in Signal Processing and Machine Learning. A number of recent low rank tensor methods have successfully utilised the tensor singular value decomposition method with tensor nuclear norm minimisation via tensor singular value thresholding. This approach while proving to be effective has the potential issue that it may over or under shrink the singular values which will effect the overall performance. A truncated nuclear norm based method has been introduced which explicitly exploits the low rank assumption within the optimization in combination with tensor singular value thresholding. In this work the truncated nuclear norm approach is extended to incorporate a data driven approach based on Stein’s unbiased risk estimation method which efficiently thresholds the singular values. Experimental results in a colour image denoising problem demonstrate the efficiency and accuracy of the method.

Index Terms—Truncated Tensor Nuclear Norm, Singular Value Shrinkage, SURE

I. INTRODUCTION

In recent years data models have played a central role in Machine Learning, Computer Vision and Signal Processing [1]. To incorporate fully the multidimensional nature of these signals has meant that a natural way of modelling this type of data is utilising tensors (multidimensional arrays). Thus the estimation of data modelled as tensors from their incomplete or corrupted observations has become increasingly important due to the enormous amount of data generated by systems such as low cost high resolution sensors, bio-informatics, imaging/video, mobile devices, sensor and social networks [2]. The key approach so far to tackle this problem has been based on exploring the inherent redundancy of the underlying modelling tensor, which has been formulated for the completion problem as a low rank estimation problem.

Beginning first with the matrix case [3], this low rank estimation can be formulated for a low rank matrix \( X \in \mathbb{R}^{m \times n} \) as follows:

\[
\min_{X} \text{rank}(X) \quad \text{s.t. } P_{\Omega}(X) = P_{\Omega}(Y)
\]

(1)

where \( X \in \mathbb{R}^{m \times n} \) and \( P_{\Omega} \) is an orthogonal projection operator onto the observed set \( \Omega \) such that:

\[
[P_{\Omega}(B)]_{ij} = \begin{cases} B_{ij} & (i,j) \in \Omega \\ 0 & (i,j) \in \Omega^{c} \end{cases}
\]

(2)

Yet this formulation is unfortunately generally NP-hard due to the non-convex and potentially discontinuous nature of the rank function. Thus a wide range of empirical studies have demonstrated successful that a tightly convex relaxation of the rank minimisation problem is the Nuclear Norm [3]. Thus the optimization problem is reformulated as:

\[
\min_{X} ||X||_* \quad \text{s.t. } P_{\Omega}(X) = P_{\Omega}(Y)
\]

(3)

where the nuclear norm \( ||X||_* = \sum_{k} \sigma_k \) is defined as the the sum of the singular values \( \sigma_k \). The aim of this method generally is given a matrix \( Y \in \mathbb{R}^{m \times n} \) find a low rank matrix \( X \in \mathbb{R}^{m \times n} \) which satisfies the following objective function formulated as a regularized least squares problem:

\[
\hat{X} = \arg \min_{X} \frac{1}{2} ||P_{\Omega}(X) - P_{\Omega}(Y)||^2_F + \lambda ||X||_*
\]

(4)

the parameter \( \lambda \) represents the trade off between the loss function and the low rank regularization induced by the nuclear norm. This was solved by utilizing proximal gradient descent, which is formulated as singular value thresholding [12].

Taking into account that the largest singular values contribute less to the rank of the matrix, this idea was extended by Hu et al. [4] to incorporate the truncated nuclear norm, which achieves a better approximation of the rank function by only considering the smallest \( r \) singular values in the optimization as follows:

\[
\min_{X} ||X||_r \quad \text{s.t. } P_{\Omega}(X) = P_{\Omega}(Y)
\]

(5)

where the truncated nuclear norm \( ||X||_r = \sum_{k=r+1}^{\min(m,n)} \sigma_k \) is defined as the the sum of the \( \min(m,n) - r \) minimum singular values \( \sigma_k \). This has been found to achieve a better approximation of the underlying rank function than the nuclear norm.

The success of the nuclear norm minimization methods has recently extended to the tensor case by utilizing a tensor decomposition paradigm proposed by Kilmer et al [5], the tensor singular value decomposition (t-SVD), and a promising approximation of the tensor rank which is derived from the t-SVD, defined as the tubal nuclear norm [7], which takes the sum of the nuclear norm of all front tensor slices in the Fourier
Thus, following the work introduced by Kilmer et al. [5], [14] the for a tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $B \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ then the tensor product $A \ast B \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ is given as:

$$A \ast B \triangleq \text{fold(bcirc}(A),\text{unfold}(B)) \quad (10)$$

Thus exploiting the above tensor product led to the development of the tensor SVD (t-SVD) defined in the following section.

B. t-SVD

Utilizing the developments for the tensor product detailed above Kilmer et al. [5] proposed the Tensor Singular Value Decomposition (t-SVD) for a tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as follows:

$$A \triangleq U \ast S \ast V^T \quad (11)$$

where $U \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $V \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ and $S \in \mathbb{R}^{n_3 \times n_3}$, where each frontal slice of $S$ is diagonal. Taking into account the property that block circulant matrices can be transformed to block diagonal matrices using normalized Fourier transform matrices then we have the following block diagonal representation:

$$\text{bdiag}(\vec{A}^{(1)}, \ldots, \vec{A}^{(n_3)}) = (F_{n_3} \otimes I_{n_1}).\text{bcirc}(A).(F_{n_3}^T \otimes I_{n_2}) \quad (12)$$

where the function $\text{bdiag}$ creates a block diagonal matrix, $F_{n_3}$ denotes the $n_3 \times n_3$ DFT matrix and $\otimes$ is the Kronecker product. Yet it was shown that it is not necessary to create the block diagonal matrix, the SVD can be performed on the frontal slices of the tensor in the Fourier domain. This is detailed in Algorithm 1. This development led to the

Algorithm 1 t-SVD

Input: $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$

$\mathcal{X} \leftarrow \text{fft}([\mathcal{X}, \mathcal{X}, \mathcal{X}])$

for $k = 1$ to $n_3$ do

$[U(k), S(k), V(k)] = \text{svd}(\text{fft}([\mathcal{X}, \mathcal{X}, \mathcal{X}]))$

$U(k) = U; S(k) = S; V(k) = V;$

end for

Output: $U \leftarrow \text{ifft}(U, [\mathcal{X}, \mathcal{X}]) \quad S \leftarrow \text{ifft}(S, [\mathcal{X}, \mathcal{X}]) \quad V \leftarrow \text{ifft}(V, [\mathcal{X}, \mathcal{X}])$

development of a new type of rank definition by taking the t-SVD of a tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ as $U \ast S \ast V^T$, the tensor tubal rank is defined as the maximum rank among all diagonal slices of $S$, $\max_k(\text{rank}(S(k)))$, where $S(k)$ is the frontal slices of the tensor $S$. Using the t-SVD leads to a definition of the tensor nuclear norm as the sum of the trace of the singular values in the frontal slices of $S$ defines as follows:

$$||A||_* \triangleq \text{tr}(S) = \sum_{k=1}^{n_3} \text{tr}(S(k)) \quad (13)$$

Xue et al. [8], [9] exploited and proved the symmetry properties of the trace of the tensor product to demonstrate as follows:

$$\text{tr}(A \ast B) = \text{tr}(\vec{A}^{(1)} B^{(1)}) \quad (14)$$
This was then subsequently utilised to developed a simplified version of the tensor nuclear norm:

\[ \|A\|_\tau \triangleq \text{tr}(S) = \text{tr}(S^{(1)}) = \|\bar{A}^{(1)}\|_\tau \]  

(15)

This approach represents a significant reduction in computation in comparison to the t-SVD based method requiring only a FFT and a SVD. This is implemented within the truncated tensor nuclear norm method detailed in the following section.

III. Tensor Truncated Nuclear Norm

Extending the Hu et al. [4] truncated nuclear norm from the matrix case to the tensor case Xue et al. [8], [9] adapted the approach within the t-SVD case by aiming to minimize the smallest \( \min (m, n) - r \) singular values. Thus incorporating the above reduced complexity Fourier based tensor nuclear norm Xue et al. defined the tensor truncated nuclear norm as follows:

\[ \|X\|_r \triangleq \|\bar{X}\|_r = \sum_{k=r+1}^{\min(n_1, n_2)} \sigma_k(\bar{X}^{(1)}) \]

\[ = \sum_{k=1}^{\min(n_1, n_2)} \sigma_k(\bar{X}^{(1)}) - \sum_{k=1}^{r} \sigma_k(\bar{X}^{(1)}) \]  

(16)

where \( \bar{X}^{(k)} \) represents the Fourier transform of the \( k \)th frontal slices of the tensor \( X \). Extending the Von Neumann trace inequality based analysis defined in Hu et al. [4] the following truncated tensor nuclear norm function is developed:

\[ \|X\|_r \triangleq \|\bar{X}\|_r - \max_{A^{(1)} A^{(1)^T} = I, B^{(1)} B^{(1)^T} = I} \text{tr}(A^{(1)} \bar{X}^{(1)} B^{(1)^T}) \]

\[ = \|X\|_r - \max_{A^{2} A^{2^T} = I, B^{2} B^{2^T} = I} \text{tr}(A * X * B^T) \]  

(17)

where \( A \) and \( B \) are derived from the t-SVD. The final optimization problem proposed in the Xue et al. [8], [9] paper extending from the matrix case in Hu et al. [4] was given as:

\[ \min_{X} \|X\|_r - \text{tr}(A_{t} * X * B_{t}^T) \]

\[ \text{s.t. } \mathcal{P}_{\Omega}(X) = \mathcal{P}_{\Omega}(M) \]  

(18)

with the full optimization procedure detailed in Algorithm 2. The tensor based version of the Alternating Direction Method of Multipliers Optimization (ADMM) based optimization method extended from Hu et al. [4] is detailed in Algorithm 3. Within Algorithm 3 Step 1 is solved using the singular value soft thresholding algorithm [12] which is defined as follows for a threshold \( \lambda \geq 0 \):

\[ \text{SVT}_\lambda(X) = \text{diag}((\sigma_k - \lambda)_+) \]  

(19)

where the operator \((t)_+ = \max(0, t)\). This operator shrinks the singular values of towards zero. This is a mathematically appealing approach as the singular value thresholding approach represents the proximal gradient of the nuclear norm [12]. Yet this approach still requires the value of \( \lambda \) to be chosen in advance. The key focus of this paper is to suggest an improved method for estimating \( \lambda \) for application within the truncated tensor nuclear norm framework for tensor completion. This is detailed within the following section.

Algorithm 2 Low Rank Tensor Completion via Truncated Tensor Nuclear Norm

\textbf{Input: } \( M_t \) incomplete data, \( \Omega \) index of known elements, \( \Omega^c \) set of unknown elements

\textbf{Initialize: } \( X_1 = M_t, \epsilon = 10^{-3}, l = 1, L = 50 \).

1: \textbf{repeat}

2: \hspace{1em} \textbf{Step 1: } \( [U_t, S_t, V_t] = \text{t-SVD}(X_t) \)

3: \hspace{1em} \text{Compute } A_t \text{ and } B_t \text{ for } r \leq \min(n_1, n_2) \text{ as: }

4: \hspace{2em} A_t = U(:, 1 : r : :)^T, B_t = V(:, 1 : r : :)^T

4: \hspace{1em} \textbf{Step 2: } \text{Optimize: }

5: \hspace{2em} X_{t+1} = \arg \min_{X} \|X\|_r - \text{tr}(A_t * X * B_t^T)

6: \hspace{2em} \text{s.t. } \mathcal{P}_{\Omega}(X) = \mathcal{P}_{\Omega}(M_t)

7: \hspace{2em} l = l + 1

8: \textbf{until } \|X_{t+1} - X_t\|_F \leq \epsilon \text{ or } l > L

\textbf{Output: } Completed tensor \( X_{\text{rec}} \)

Algorithm 3 Truncated Tensor Nuclear Norm

\textbf{Input: } \( A_t, B_t, M, \mu = 5 \times 10^{-4}, \xi = 10^{-4}, K = 200 \)

\textbf{Initialize: } \( X_1 = M_t, W_1 = Y_1 = X_1, k = 1 \).

1: \textbf{repeat}

2: \hspace{1em} \textbf{Step 1: } \( X_{k+1} = \text{SVT}_\lambda \left( W_k - \frac{1}{\mu} Y_k \right) \)

3: \hspace{1em} \textbf{Step 2: } \( W_{k+1} = X_{k+1} + \frac{1}{\mu} (A_{t} * X_{k+1} * B_{t}^T) \)

4: \hspace{1em} \textbf{Step 3: } \( Y_{k+1} = Y_k + \mu (X_{k+1} - W_{k+1}) \)

5: \hspace{2em} \text{Keep the values of known elements constant}

6: \hspace{2em} W_{k+1} = (W_{k+1} + Y_{k+1} + M_t)

7: \hspace{2em} k = k + 1

8: \textbf{until } \|X_{k+1} - X_k\|_F \leq \xi \text{ or } k > K

\textbf{Output: } Completed tensor

A. Optimization via SURE Singular Value Shrinkage

Within the development of the Xue et al. [8], [9] algorithm the classic singular value thresholding algorithm [13] is applied with a fixed threshold value \( \lambda \). This provides the question how much shrinkage should be applied to the singular values based on the data, too much will result in a large bias, whereas too little will result in a large variance. The ideal scenario would be to minimise the mean square error or risk of the estimator, but as this would require the true low rank tensor which is unknown. Within [11] a solution to this problem was offered based on the Stein Unbiased Risk Estimator (SURE)
The norm minimization method as detailed in Algorithm 4. The utilised to extend upon the Xue et al. \[8\], \[9\] truncated nuclear provides a method to determine a suitable threshold value for \(\lambda\), which is normally valid in practise. Equation 20

\[
\text{Initialize: } X
\]

\[
\text{Input: } \lambda, \mu, A_l, B_l, M
\]

\[
\text{Algorithm 4: SURE Optimized Truncated Tensor Nuclear Norm}
\]

\[
\text{Fig. 1: Experimental Colour Image Examples}
\]
<table>
<thead>
<tr>
<th>No.</th>
<th>TNNR</th>
<th>SMNN</th>
<th>ATNN</th>
<th>Tubal-NN</th>
<th>T-TNN</th>
<th>ST-TNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.56</td>
<td>22.19</td>
<td>25.04</td>
<td>28.91</td>
<td>29.65</td>
<td><strong>30.79</strong></td>
</tr>
<tr>
<td></td>
<td>(665)</td>
<td>(343)</td>
<td>(285)</td>
<td>(264)</td>
<td>(165)</td>
<td>(105)</td>
</tr>
<tr>
<td>2</td>
<td>28.79</td>
<td>23.19</td>
<td>27.43</td>
<td>31.45</td>
<td>32.55</td>
<td><strong>32.95</strong></td>
</tr>
<tr>
<td></td>
<td>(541)</td>
<td>(346)</td>
<td>(287)</td>
<td>(258)</td>
<td>(148)</td>
<td>(107)</td>
</tr>
<tr>
<td>3</td>
<td>24.33</td>
<td>22.23</td>
<td>24.42</td>
<td>26.17</td>
<td>27.67</td>
<td><strong>27.25</strong></td>
</tr>
<tr>
<td></td>
<td>(878)</td>
<td>(341)</td>
<td>(291)</td>
<td>(262)</td>
<td>(178)</td>
<td>(155)</td>
</tr>
<tr>
<td>4</td>
<td>30.82</td>
<td>26.40</td>
<td>28.97</td>
<td>34.19</td>
<td>35.32</td>
<td><strong>35.22</strong></td>
</tr>
<tr>
<td></td>
<td>(614)</td>
<td>(351)</td>
<td>(276)</td>
<td>(244)</td>
<td>(181)</td>
<td>(106)</td>
</tr>
<tr>
<td>5</td>
<td>28.96</td>
<td>24.95</td>
<td>26.34</td>
<td>30.29</td>
<td>31.20</td>
<td><strong>31.43</strong></td>
</tr>
<tr>
<td></td>
<td>(656)</td>
<td>(346)</td>
<td>(286)</td>
<td>(246)</td>
<td>(180)</td>
<td>(158)</td>
</tr>
<tr>
<td>6</td>
<td>23.23</td>
<td>22.95</td>
<td>23.60</td>
<td>25.48</td>
<td>26.24</td>
<td><strong>27.25</strong></td>
</tr>
<tr>
<td></td>
<td>(813)</td>
<td>(339)</td>
<td>(277)</td>
<td>(241)</td>
<td>(212)</td>
<td>(161)</td>
</tr>
<tr>
<td>7</td>
<td>30.55</td>
<td>27.85</td>
<td>29.05</td>
<td>33.55</td>
<td>34.45</td>
<td><strong>34.60</strong></td>
</tr>
<tr>
<td></td>
<td>(624)</td>
<td>(345)</td>
<td>(283)</td>
<td>(247)</td>
<td>(188)</td>
<td>(164)</td>
</tr>
<tr>
<td>8</td>
<td>26.04</td>
<td>22.80</td>
<td>25.68</td>
<td>29.02</td>
<td>29.93</td>
<td><strong>29.93</strong></td>
</tr>
<tr>
<td></td>
<td>(573)</td>
<td>(344)</td>
<td>(278)</td>
<td>(249)</td>
<td>(173)</td>
<td>(152)</td>
</tr>
<tr>
<td>9</td>
<td>24.17</td>
<td>22.53</td>
<td>24.12</td>
<td>27.92</td>
<td>28.98</td>
<td><strong>29.86</strong></td>
</tr>
<tr>
<td></td>
<td>(639)</td>
<td>(347)</td>
<td>(283)</td>
<td>(260)</td>
<td><strong>(140)</strong></td>
<td>(159)</td>
</tr>
<tr>
<td>10</td>
<td>26.60</td>
<td>22.40</td>
<td>26.28</td>
<td>31.59</td>
<td>32.60</td>
<td><strong>33.57</strong></td>
</tr>
<tr>
<td></td>
<td>(745)</td>
<td>(349)</td>
<td>(284)</td>
<td>(254)</td>
<td>(184)</td>
<td>(109)</td>
</tr>
</tbody>
</table>

Based singular value thresholding method. The experimental results demonstrate that the method improves upon the previously developed truncated tensor nuclear norm method for the case of colour image recovery. A limitation of the method is it is constrained to third order tensors. Future developments will include further optimization of the singular value shrinkage functions and extensions of the approach to N-way tensors ($N > 3$).

**REFERENCES**


