

ONLINE LOW-RANK TENSOR SUBSPACE TRACKING FROM INCOMPLETE DATA BY CP DECOMPOSITION USING RECURSIVE LEAST SQUARES

HiroYuki Kasai

The University of Electro-Communications, Tokyo, JAPAN
kasai@is.uec.ac.jp

ABSTRACT

We propose an online tensor subspace tracking algorithm based on the CP decomposition exploiting the recursive least squares (RLS), dubbed OnLine Low-rank Subspace tracking by TEnsor CP Decomposition (OLSTEC). Numerical evaluations show that the proposed OLSTEC algorithm gives faster convergence per iteration comparing with the state-of-the-art online algorithms.

Index Terms— Online subspace tracking, Tensor completion, CP decomposition, Recursive least squares

1. INTRODUCTION

The problem of tensor subspace tracking of multidimensional data, which are naturally represented by a tensor, has been studied intensively in recent years. The usual structural assumption on a tensor is that the tensor has *low-rank* in every mode. The popular *convex relaxation* [1, 2, 3] approach minimizes the sum of the nuclear norms of the unfolding matrices of the tensor by extending the successful results in matrix completion problem [4] under theoretical performance guarantees. However, due to the limited scalability towards large-scale data of convex relaxations, the *fixed-rank* non-convex approach with tensor decomposition [5, 6] has gained big attentions recently because of superior performance in practice in despite of local minima. This also comes from the success of matrix cases [7, 8, 9]. Considering that the data are sequentially acquired, or the underlying low-rank structure changes over time, online subspace tracking and estimation is essential to avoid expensive repetitive computations of batch-based algorithms.

With regard to matrix-based online tracking, a representative research is the projection approximation subspace tracking (PAST) [10]. GROUSE [11] recently proposes an incremental gradient descent algorithm on the Grassmannian $\mathcal{G}(d, n)$, the space of all d -dimensional subspace of \mathbb{R}^n [12, 13]. The algorithm minimizes ℓ_2 -norm cost function. GRASTA[14] enhances robustness against outliers by exploiting ℓ_1 -norm cost function. PETRELS [15] calculates the underlying subspace via a discounted recursive process for each row of the subspace matrix in parallel.

On the other hand, as for tensor-based tracking, Nion and Sidiropoulos propose an adaptive algorithm to obtain the CP (CANDECOMP/PARAFAC) decompositions [16]. Yu et al. also propose an accelerated online tensor learning algorithm (ALTO) based on the Tucker decomposition [17]. However, they do not deal with missing data presence. Mardani et al. propose an online imputation algorithm based on the CP decomposition under the presence of missing data [18]. This considers the stochastic gradient descent (SGD) for large-scale data. However, considering the situations where the subspace changes dramatically and the processing speed is enough faster than data acquiring speed, a faster convergence algorithm per iteration to track this change is crucial.

This paper presents a new online tensor tracking algorithm, dubbed OLSTEC, for the partially observed high-dimensional data stream corrupted by noise. We focus on the fixed-rank tensor completion algorithm with a second-order stochastic gradient descent based on the CP decomposition exploiting the recursive least squares (RLS). The rest of paper is organized as follows. Section 2 formulates the problem of online subspace tracking and Section 3 proposes the new algorithm. Numerical evaluations are performed in Section 4, after which we conclude in Section 5.

2. PROBLEM FORMULATION

This paper addresses the problem of low-rank tensor completion in an online manner when the rank is a priori known or estimated. Without loss of generality, we focus on 3-order tensors of which one order increases over time. In other words, we address $\mathcal{Y} \in \mathbb{R}^{L \times W \times T}$ of which 3-rd order increases infinitely. Assuming $\mathcal{Y}_{i_1, i_2, i_3}$ are only known for some indices $(i_1, i_2, i_3) \in \Omega$, where Ω is a subset of the complete set of indices (i_1, i_2, i_3) , a general *batch-based* fixed-rank tensor completion problem is formulated as

$$\begin{aligned} \min_{\mathcal{X} \in \mathbb{R}^{L \times W \times T}} \quad & \frac{1}{2} \|\mathcal{P}_\Omega(\mathcal{X}) - \mathcal{P}_\Omega(\mathcal{Y})\|_F^2 \\ \text{subject to} \quad & \text{rank}(\mathcal{X}) = R, \end{aligned} \quad (1)$$

where the operator $\mathcal{P}_\Omega(\mathcal{X})_{i_1, i_2, i_3} = \mathcal{X}_{i_1, i_2, i_3}$ if $(i_1, i_2, i_3) \in \Omega$ and $\mathcal{P}_\Omega(\mathcal{X})_{i_1, i_2, i_3} = 0$ otherwise and (with a slight abuse of notation) $\|\cdot\|_F$ is the Frobenius norm. $\text{rank}(\mathcal{X})$ is the

rank of \mathcal{X} (see [19] for a detailed discussion on tensor rank). $R \ll \{L, W, T\}$ enforces a low-rank structure. Hereafter, the t -th slice in the third mode of \mathcal{Y} , i.e. $\mathcal{Y}_{:, :, t}$ and its value at (l, w) , i.e., $\mathcal{Y}_{l, w, t}$, are denoted as \mathbf{Y}_t and $[\mathbf{Y}_t]_{l, w}$, respectively.

The CP decomposition that we address in this paper decomposes a tensor into a sum of component rank-one tensors [19], as $\mathcal{X} \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{c}_r \circ \mathbf{b}_r$, where $\mathbf{a}_r \in \mathbb{R}^L$, $\mathbf{b}_r \in \mathbb{R}^T$, and $\mathbf{c}_r \in \mathbb{R}^W$. The symbol \circ represents the vector outer product. The *factor matrices* refer to the combination of the vectors from the rank-one components, i.e., $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}^{L \times R}$ and likewise for \mathbf{B} and \mathbf{C} . It should be noted that \mathbf{A} , \mathbf{B} and \mathbf{C} can be also represented by row vectors, i.e., horizontal vectors, for example, $\mathbf{A} = [(\mathbf{a}^1)^T, \dots, (\mathbf{a}^R)^T]^T$, where $\mathbf{a}^l \in \mathbb{R}^R$. Thus, $\mathbf{Y}_t = \mathbf{A} \text{diag}(\mathbf{b}^t) \mathbf{C}^T = \sum_{r=1}^R \mathbf{b}^t(r) \mathbf{a}_r \mathbf{c}_r^T$. Then, the problem (1) is reformulated with ℓ_2 regularizers as [18]

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \|\mathcal{P}_\Omega(\mathcal{Y}) - \mathcal{P}_\Omega(\mathcal{X})\|_F^2 + \mu(\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2 + \|\mathbf{C}\|_F^2)$$

$$\text{subject to } \mathbf{X}_\tau = \mathbf{A} \text{diag}(\mathbf{b}^\tau) \mathbf{C}^T \text{ for } \tau = 1, \dots, t. \quad (2)$$

where μ is a regularizer parameter. Consequently, considering the situation where the partially observed tensor slice $\Omega_\tau \otimes \mathbf{Y}_\tau$ is acquired sequentially over time, we estimate $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\}$ by minimizing the exponentially weighted least squares;

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \left[\|\Omega_\tau \otimes [\mathbf{Y}_\tau - \mathbf{A} \text{diag}(\mathbf{b}^\tau) \mathbf{C}^T]\|_F^2 + \bar{\mu}(\|\mathbf{A}\|_F^2 + \|\mathbf{C}\|_F^2) + \mu[\tau] \|\mathbf{b}^\tau\|_2^2 \right], \quad (3)$$

where $\bar{\mu} = \mu[\tau] / \sum_{\tau=1}^t \lambda^{t-\tau}$, and $0 < \lambda \leq 1$ is the so-called forgetting parameter. $\lambda = 1$ case is equivalent to the batch-based problem (2). The symbol \otimes denotes the Hadamard Product, which is the element-wise product.

3. PROPOSED TENSOR TRACKING: OLSTEC

The unknown variables in (3) are \mathbf{A} , \mathbf{C} , and \mathbf{b} . Since \mathbf{A} and \mathbf{C} are non-convex set, this function is non-convex. The proposed OLSTEC algorithm, as summarized by Algorithm 1, alternates between a least-square estimation of $\mathbf{b}[t]$ for fixed $\mathbf{A}[t-1]$ and $\mathbf{C}[t-1]$, and a second order stochastic gradient step using the RLS on $\mathbf{A}[t]$ and $\mathbf{C}[t]$ for fixed $\mathbf{b}[t]$. It should be noted that $\mathbf{W}[t]$ with the square bracket indicates the calculated \mathbf{W} after performing t -times updates.

3.1. Calculation of $\mathbf{b}[t]$

The estimate $\mathbf{b}[t]$ of \mathbf{b}^t is obtained in a closed form by least-squares by denoting $\mathbf{g}_{l, w}[t] = \mathbf{a}^l[t-1] \otimes \mathbf{c}^w[t-1] \in \mathbb{R}^R$ as

$$\min_{\mathbf{b}^t \in \mathbb{R}^R} \frac{1}{2} \left[\sum_{l=1}^L \sum_{w=1}^W ([\Omega_t]_{l, w} ([\mathbf{Y}_t]_{l, w} - (\mathbf{g}_{l, w}[t])^T \mathbf{b}^t))^2 + \mu[t] \|\mathbf{b}^t\|_2^2 \right]$$

Algorithm 1 OLSTEC algorithm

Require: $\{\mathbf{Y}_t \text{ and } \Omega_t\}_{t=1}^\infty, \lambda, \mu$

- 1: Initialize $\{\mathbf{A}[0], \mathbf{b}[0], \mathbf{C}[0], \mathbf{Y}[0] = \mathbf{0}, (\mathbf{R}\mathbf{A}_l[0])^{-1} = (\mathbf{R}\mathbf{C}_w[0])^{-1} = \gamma \mathbf{I}_R, \gamma > 0.$
- 2: **for** $t = 1, 2, \dots$ **do**
- 3: Calculate $\mathbf{b}[t]$ Equation (4)
- 4: $\mathbf{X}_t = \mathbf{A}[t-1] \text{diag}(\mathbf{b}_t) (\mathbf{C}[t-1])^T$
- 5: **for** $l = 1, 2, \dots, L$ **do**
- 6: Calculate $\mathbf{R}\mathbf{A}_l[t]$ Equation (7)
- 7: Calculate $\mathbf{a}^l[t]$ Equation (8)
- 8: **end for**
- 9: **for** $w = 1, 2, \dots, W$ **do**
- 10: Calculate $\mathbf{R}\mathbf{C}_l[t]$ Equation (10)
- 11: Calculate $\mathbf{c}^w[t]$ Equation (9)
- 12: **end for**
- 13: **end for**
- 14: **return** $\mathbf{X}_t = \mathbf{A}[t] \text{diag}(\mathbf{b}[t]) (\mathbf{C}[t])^T$

Defining $F[t]$ as the inner objective to be minimized, we obtain $\mathbf{b}[t]$ since $\mathbf{b}[t]$ satisfies $\partial F[t] / \partial \mathbf{b}[t] = 0$ as

$$\mathbf{b}[t] = \left[\mu[t] \mathbf{I}_R + \sum_{l=1}^L \sum_{w=1}^W \Omega[t]_{l, w} \mathbf{g}_{l, w}[t] (\mathbf{g}_{l, w}[t])^T \right]^{-1} \left[\sum_{l=1}^L \sum_{w=1}^W \Omega[t]_{l, w} \mathbf{Y}[t]_{l, w} \mathbf{g}_{l, w}[t] \right]. \quad (4)$$

3.2. Calculation of $\mathbf{A}[t]$ and $\mathbf{C}[t]$ based on RLS

The calculation of $\mathbf{C}[t]$ uses $\mathbf{A}[t-1]$, and the calculation of $\mathbf{A}[t]$ uses $\mathbf{C}[t-1]$. This paper addresses a second-order stochastic gradient based on the RLS with forgetting parameters, which has been widely used in tracking of time varying parameters in many fields. Its computation is efficient since we update the estimates recursively every time new data becomes available.

As for $\mathbf{A}[t]$, the problem (3) is reformulated as

$$\min_{\mathbf{A} \in \mathbb{R}^{L \times R}} \frac{1}{2} \sum_{\tau} \lambda^{t-\tau} \left[\|\Omega_\tau \otimes [\mathbf{Y}_\tau - \mathbf{A} \text{diag}(\mathbf{b}[\tau]) \mathbf{C}[\tau-1]^T]\|_F^2 \right] + \frac{\mu[t]}{2} \|\mathbf{A}\|_F^2. \quad (5)$$

The objective function in (5) decomposes into a parallel set of smaller problems, one for each row of \mathbf{A} , as

$$\mathbf{a}^l[t] = \arg \min_{\mathbf{a}^l \in \mathbb{R}^R} \frac{1}{2} \sum_{\tau=1}^t \left[\lambda^{t-\tau} \sum_{w=1}^W [\Omega_\tau]_{l, w} ([\mathbf{Y}_\tau]_{l, w} - (\mathbf{a}^l)^T \text{diag}(\mathbf{b}[\tau]) \mathbf{c}^w[\tau-1])^2 \right] + \frac{\mu[t]}{2} \|\mathbf{a}^l\|_2^2.$$

Here, denoting $\text{diag}(\mathbf{b}[\tau]) \mathbf{c}^w[\tau-1]$ as $\boldsymbol{\alpha}_w[\tau] \in \mathbb{R}^R$, $\mathbf{a}^l[t]$ is obtained by setting the derivative to zero as

$$\mathbf{R}\mathbf{A}_l[t] \mathbf{a}^l[t] = \mathbf{s}_l[t], \quad (6)$$

where $\mathbf{RA}_l[t] \in \mathbb{R}^{R \times R}$ and $\mathbf{s}_l[t] \in \mathbb{R}^R$ are defined as

$$\begin{aligned}\mathbf{RA}_l[t] &= \sum_{\tau=1}^t \left[\sum_{w=1}^W \lambda^{t-\tau} [\mathbf{\Omega}_\tau]_{l,w} \boldsymbol{\alpha}_w[\tau] \boldsymbol{\alpha}_w[\tau]^T \right] + \mu[t] \mathbf{I}_R \\ \mathbf{s}_l[t] &= \sum_{\tau=1}^t \left[\sum_{w=1}^W \lambda^{t-\tau} [\mathbf{\Omega}_\tau]_{l,w} [\mathbf{Y}_\tau]_{l,w} \boldsymbol{\alpha}_w[\tau] \right].\end{aligned}$$

Here, $\mathbf{RA}_l[t]$ is transformed by separating t -th term as

$$\begin{aligned}\mathbf{RA}_l[t] &= \lambda \mathbf{RA}_l[t-1] + \sum_{w=1}^W [\mathbf{\Omega}_t]_{l,w} \boldsymbol{\alpha}_w[t] (\boldsymbol{\alpha}_w[t])^T \\ &\quad + (\mu[t] - \lambda \mu[t-1]) \mathbf{I}_R.\end{aligned}\quad (7)$$

Likewise, $\mathbf{s}_l[t]$ is obtained as $\mathbf{s}_l[t] = \lambda \mathbf{s}_l[t-1] + \sum_{w=1}^W [\mathbf{\Omega}_t]_{l,w} [\mathbf{Y}_t]_{l,w} \boldsymbol{\alpha}_w[t]$. Thus, from (6), we reformulate $\mathbf{RA}_l[t]$ as

$$\begin{aligned}\mathbf{RA}_l[t] \mathbf{a}^l[t] &= \mathbf{RA}_l[t] \mathbf{a}^l[t-1] - (\mu[t] - \lambda \mu[t-1]) \mathbf{a}^l[t-1] \\ &\quad + \sum_{w=1}^W [\mathbf{\Omega}_t]_{l,w} ([\mathbf{Y}_t]_{l,w} - \boldsymbol{\alpha}_w[t]^T \mathbf{a}^l[t-1]) \boldsymbol{\alpha}_w[t].\end{aligned}$$

Finally, $\mathbf{a}^l[t]$ is obtained as

$$\begin{aligned}\mathbf{a}^l[t] &= \mathbf{a}^l[t-1] - (\mu[t] - \lambda \mu[t-1]) (\mathbf{RA}_l[t])^{-1} \mathbf{a}^l[t-1] + \\ &\quad \sum_{w=1}^W [\mathbf{\Omega}_t]_{l,w} ([\mathbf{Y}_t]_{l,w} - (\boldsymbol{\alpha}_w[t])^T \mathbf{a}^l[t-1]) (\mathbf{RA}_l[t])^{-1} \boldsymbol{\alpha}_w[t].\end{aligned}\quad (8)$$

Similarly, $\mathbf{c}^w[t]$ for $\mathbf{C}[t]$ can be obtained as

$$\begin{aligned}\mathbf{c}^w[t] &= \mathbf{c}^w[t-1] - (\mu[t] - \lambda \mu[t-1]) (\mathbf{RC}_w[t])^{-1} \mathbf{c}^w[t-1] + \\ &\quad \sum_{l=1}^L [\mathbf{\Omega}_t]_{l,w} ([\mathbf{Y}_t]_{l,w} - \boldsymbol{\beta}_w[t]^T \mathbf{c}^w[t-1]) (\mathbf{RC}_w[t])^{-1} \boldsymbol{\beta}_w[t],\end{aligned}\quad (9)$$

where $\boldsymbol{\beta}_w[\tau] \in \mathbb{R}^{1 \times R}$ is $(\mathbf{a}^l[\tau])^T \text{diag}(\mathbf{b}[\tau])$, and $\mathbf{RC}_w[t]$ is defined as

$$\begin{aligned}\mathbf{RC}_w[t] &= \lambda \mathbf{RC}_w[t-1] + \sum_{l=1}^L [\mathbf{\Omega}_t]_{l,w} \boldsymbol{\beta}_w[t]^T \boldsymbol{\beta}_w[t] \\ &\quad + (\mu[t] - \lambda \mu[t-1]) \mathbf{I}_R.\end{aligned}\quad (10)$$

3.3. Complexity and memory consumption

With respect to computational complexity per iteration, OLSTEC requires $\mathcal{O}(|\mathbf{\Omega}_t| R^2 + L R^3)$ because of $\mathcal{O}(|\mathbf{\Omega}_t| R^2)$ for $\mathbf{b}[t]$ in (4) and $\mathcal{O}(L R^3)$ for the inversion of \mathbf{RA}_l and \mathbf{RC}_w in (8) and (9), respectively. As for memory consumption, $\mathcal{O}((L+W) R^2)$ is required for $\mathbf{RA}[t]$ and $\mathbf{RC}[t]$, respectively.

4. NUMERICAL EVALUATIONS

We show numerical comparisons of the OLSTEC algorithm¹ with state-of-the-art algorithms for synthetic and real-world

¹Matlab source code is available at <http://www.kasailab.com/research/olstec>.

datasets. All the following experiments are done on a PC with 3.0 GHz Intel Core i7 CPU and 16 GB RAM. We first evaluate the performance of our proposed algorithm using synthetic dataset with the state-of-the-art online algorithm proposed in [18], termed as “TeCPSGD” algorithm in this paper. We first generate a low R -rank tensor $\mathcal{Y} \in \mathbb{R}^{L \times W \times T}$ where its factor matrices are generated with i.i.d standard Gaussian $\mathcal{N}(0, 1)$ entries, and Gaussian noise with i.i.d $\mathcal{N}(0, \epsilon^2)$ entries are added. We set $L = W = \{100, 200, 300\}$, $T = 1000$, $R = \{5, 10, 15\}$, and the noise level $\epsilon = 10^{-3}$. The observation ratio, ρ , is $\{0.1, 0.05\}$. $\mu[t] = 10^{-9}$ and $\lambda = 0.88$ are configured in the proposed algorithm. It should be noted that we implement TeCPSGD with our configured parameters because the source code of TeCPSGD is not available. Figure 1 shows the *running-averaging estimation error* $\frac{1}{T} \sum_{\tau=1}^T \|\mathbf{X}_\tau - \mathbf{Y}_\tau\|_F^2 / \|\mathbf{Y}_\tau\|_F^2$ for each observation ratio ρ , where five runs are performed independently, and the results show the average with standard deviations. From these results, the proposed OLSTEC algorithm shows much lower estimation error, especially when observation ratios are lower. In addition, the standard derivations are also smaller, thus, the convergence property of the proposed algorithm is stabler than that of TeCPSGD. Figure 2 (a) and (b) show the *normalized residual error* $\|\mathbf{X}_t - \mathbf{Y}_t\|_F^2 / \|\mathbf{Y}_t\|_F^2$ when the observation ratios are 0.1 and 0.05, respectively. Additionally, we show, as reference, the result of CP-WOPT [20], the state-of-the-art batch algorithm. The relative change in function value tolerance is set to 10^{-9} and the maximum iterations is 300 for CP-WOPT. Our proposed algorithm gives superior convergence performances than those of TeCPSGD.

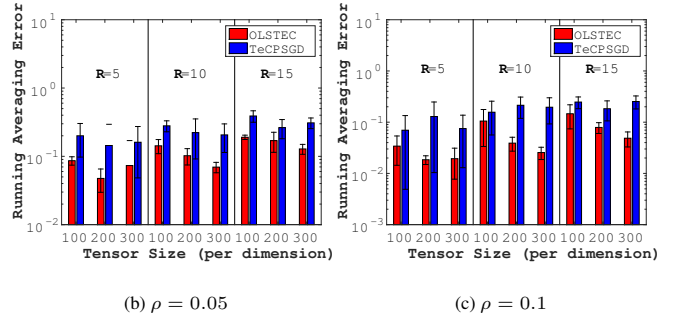


Fig. 1. Running-averaging error in synthetic dataset.

We also evaluate a scenario where a subspace of rank changes abruptly periodically. Four rank-5 tensors of $100 \times 100 \times 250$ are concatenated in series at the 3-rd order direction. Figure 2(c) shows the normalized residual error at each iteration. This shows that the subspace tracking behavior of the OLSTEC algorithm gives a superior performance than that of TeCPSGD which cannot recover correct subspaces after abrupt changes.

Next, we evaluate the tracking performances using surveillance video as a real-world dataset. Although each video frame does not have low-rank structure and a tensor-based

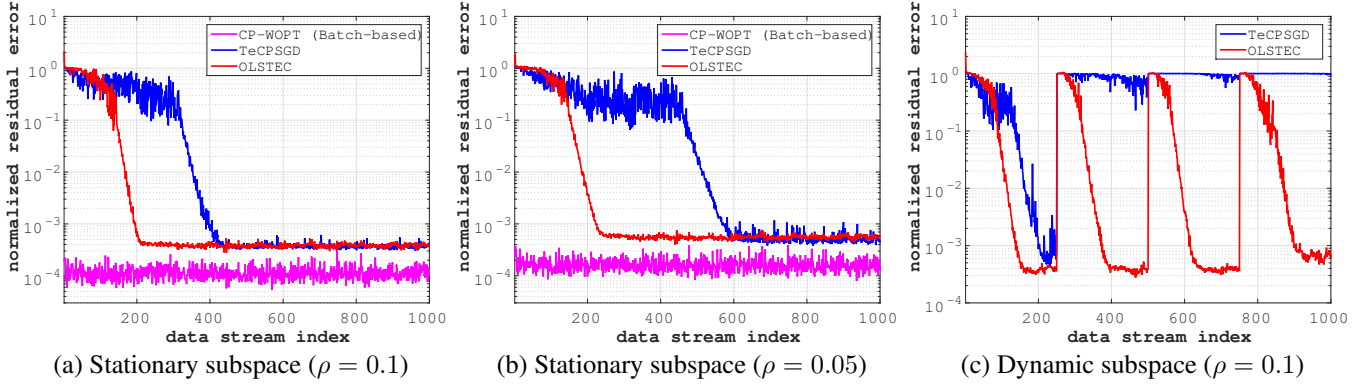


Fig. 2. The normalized estimation error in synthetic dataset.

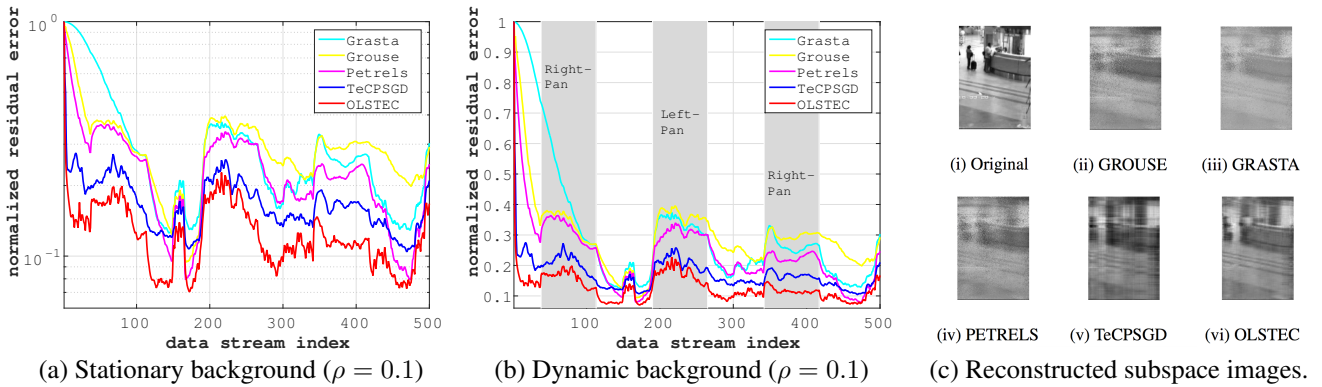


Fig. 3. The normalized estimation error in real-world dataset.

approach basically has a disadvantage for the approximation of its underlying subspace, this experiment demonstrates the superior tracking performance of OLSTEC. We compare OLSTEC with TeCPSGD as well as the matrix-based algorithms including GROUSE [11], GRASTA [14], and PETRELS [15]. We use Matlab codes provided by the respective authors except for TeCPSGD with our configured parameters. “Airport Hall” dataset of size 288×352 with 500 frames is used. Moreover, for fair comparison between tensor and matrix-based algorithms, the rank is set to 20 and 10 for the former, i.e., OLSTEC and TeCPSGD, and for the latter, respectively. Still, the tensor-based algorithms has much less free parameters than those of the matrix-based algorithms. This experiment also considers two scenarios. The first separates foreground objects with static background and moving objects in the foreground. Figure 3 (a) shows the superior performance of OLSTEC against other algorithms. Furthermore, we examine the performances against a dynamic moving background as the second scenario. The input video is created virtually by moving cropped partial image from its original entire frame image of video. The cropping window with 288×200 moves from the leftmost partial image to the rightmost, then returns to the leftmost image after stopping

a certain period of time. The generated video includes right-panning video from 38-th to 113-th frame and from 342-th to 417-th frame, and left-panning video from 190-th to 265-th frame. Figure 3(b) shows how OLSTEC can quickly adapt to the changed background. Figure 3(c) shows the reconstructed (i.e., completed) image at 110-th frame of OLSTEC gives better quality than those of others.

5. CONCLUSION AND FUTURE WORK

We have proposed a new online tensor subspace tracking algorithm, dubbed OLSTEC, for the partially observed high-dimensional data stream corrupted by noise. Especially, we addressed a second-order stochastic gradient descent based on the recursive least squares to achieve faster convergence of subspace tracking. Numerical comparisons suggest that our proposed algorithm has superior performances on synthetic as well as real-world datasets. As a future research direction, we will investigate the ways of the Tucker decomposition.

Acknowledgments I thank Prof. Wolfgang Kellerer and Prof. Martin Kleinsteuber for useful discussions on the paper. H. Kasai is (partly) supported by the Ministry of Internal Affairs and Communications, Japan, as the SCOPE Project.

6. REFERENCES

- [1] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, 2013.
- [2] R. Tomioka, K. Hayashi, and H. Kashima, “Estimation of low-rank tensors via convex optimization,” *arXiv:1010.0789*, 2011.
- [3] M. Signoretto, Q. T. Dinh, L. D. Lathauwer, and J. A.K. Suykens, “Learning with tensors: a framework based on convex optimization and spectral regularization,” *Mach. Learn.*, vol. 94, no. 3, pp. 303–351, 2014.
- [4] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.
- [5] M. Filipović and A. Jukić, “Tucker factorization with missing data with application to low- n -rank tensor completion,” *Multidim. Syst. Sign. P.*, 2013.
- [6] D. Kressner, M. Steinlechner, and B. Vandereycken, “Low-rank tensor completion by Riemannian optimization,” *BIT Numer. Math.*, vol. 54, no. 2, pp. 447–468, 2014.
- [7] N. Boumal and P.-A. Absil, “RTRMC : A Riemannian trust-region method for low-rank matrix completion,” in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 2011.
- [8] B. Mishra, G. Meyer, F. Bach, and R. Sepulchre, “Low-rank optimization with trace norm penalty,” *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2124–2149, 2013.
- [9] T. Ngo and Y. Saad, “Scaled gradients on Grassmann manifolds for matrix completion,” in *NIPS*, 2012, pp. 1421–1429.
- [10] B. Yang, “Projection approximation subspace tracking,” *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.
- [11] L. Balzano, R. Nowak, and B. Recht, “Online identification and tracking of subspaces from highly incomplete information,” *arXiv:1006.4046*, 2010.
- [12] A. Edelman, T.A. Arias, and S.T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.
- [13] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.
- [14] J. He, L. Balzano, and A. Szlam, “Incremental gradient on the grassmannian for online foreground and background separation in subsampled video,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [15] Y. Chi, Y. C. Eldar, and R. Calderbank, “Petrels: Parallel subspace estimation and tracking using recursive least squares from partial observations,” *IEEE Trans. on Signal Processing*, vol. 61, no. 23, pp. 5947–5959, 2013.
- [16] D. Nion and N.D. Sidiropoulos, “Adaptive algorithms to track the parafac decomposition of a third-order tensor,” *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2299–2310, 2009.
- [17] R. Yu, D. Cheng, and Y. Liu, “Accelerated online low-rank tensor learning for multivariate spatio-temporal streams,” *International Conference on Machine Learning (ICML)*, 2015.
- [18] M. Mardani, G. Mateos, and G.B. Giannakis, “Subspace learning and imputation for streaming big data matrices and tensors,” *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 266–2677, 2015.
- [19] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [20] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, “Scalable tensor factorizations with missing data,” in *Proceedings of the 2010 SIAM International Conference on Data Mining (SDM10)*, 2010, pp. 701–712.