Detection of change by L1-norm principal-component analysis

Gallone, Ginevra, Varma, Kavita, Pados, Dimitris, Colonnese, Stefania

Ginevra Gallone, Kavita Varma, Dimitris A. Pados, Stefania Colonnese, "Detection of change by L1-norm principal-component analysis," Proc. SPIE 11395, Big Data II: Learning, Analytics, and Applications, 113950D (19 May 2020); doi: 10.1117/12.2559976

SPIE.

Event: SPIE Defense + Commercial Sensing, 2020, Online Only

Detection of Change by L1-norm Principal-Component Analysis

Ginevra Gallone[ab], Kavita Varma[a], Dimitris A. Pados[a], and Stefania Colonnese[b]

[a]Department of Computer and Electrical Engineering and Computer Science & I-SENSE, Florida Atlantic University, Boca Raton, FL 33431, USA

[b]Department of Information Engineering, Electronics and Telecommunications , La Sapienza University of Rome, Rome, Italy 00185

ABSTRACT

We consider the problem of detecting a change in an arbitrary vector process by examining the evolution of calculated data subspaces. In our developments, both the data subspaces and the change identification criterion are novel and founded in the theory of $L1$-norm principal-component analysis (PCA). The outcome is highly accurate, rapid detection of change in streaming data that vastly outperforms conventional eigenvector subspace methods ($L2$-norm PCA). In this paper, illustrations are offered in the context of artificial data and real electroencephalography (EEG) and electromyography (EMG) data sequences.

Keywords: Detection of change, L_1-norm, principal-component analysis, streaming data, time series.

1. INTRODUCTION

Time series analysis describes the behavior of natural or man-made systems over time. In particular, the behavior of a system can change over time due to external or internal causes.[1] Change detection is the process of identifying changes in the state of a system by observing the system at different times.[2] Depending on the nature of the system, image differencing, principal-component analysis,[3–5] and post-classification comparison are common methods used in detection of change.[6]

In this paper, we propose a new method to identify change in arbitrary vector processes by novel $L1$-norm principal-component analysis and comparison of $L1$-norm subspaces. In particular, we introduce a new way to measure the difference between subspaces that we call the "maximum left-right projection" (MLRP) criterion. The broad purpose of conventional principal-component analysis (PCA) is to reduce the dimensionality of a dataset consisting of a number of interrelated variables (coordinates), while retaining as much as possible of the variation present in the dataset. This is achieved by summarizing the dataset in the form of a set of vectors, principal components (PCs), which are orthogonal and ordered so that the first components retain most of the variation present in all of the original dataset. In this context, conventionally, PCA uses the $L2$-norm based singular-value decomposition of the data matrix. It is well understood by now that $L2$-norm PCA is sensitive to outliers/changed data, so recently there has been significant interest in $L1$-norm based approaches.[7,8] The $L1$-norm gives one exponent lower weight than the $L2$-norm to data that are far away from the majority and offers enhanced immunity against outliers and more accurate representation of the nominal data. In this work, we use $L1$-PCA to identify where the point of change in a vector time series is. To do so, we split the sequence into a "left" and "right" part, calculate their $L1$-norm subspaces and compare. Extensive experimentation with both artificial and real time-series data shows extraordinary success and improvement against $L2$-norm detection of change methods.

Ginevra Gallone: E-mail: ggallone@fau.edu
Kavita Varma: E-mail: kvarma2018@fau.edu
Dimitris A. Pados: E-mail: dpados@fau.edu
Stefania Colonnese: E-mail: stefania.colonnese@uniroma1.it

Big Data II: Learning, Analytics, and Applications, edited by Fauzia Ahmad,
Proc. of SPIE Vol. 11395, 113950D · © 2020 SPIE · CCC
code: 0277-786X/20/$21 · doi: 10.1117/12.2559976

The remainder of this paper is organized as follow. In Section 2 the proposed $L1$-norm principal-component analysis method and the "maximum left-right projections" criterion are developed and presented. In Section 3 we extend our detection of change method to streaming data. In Section 4, the effectiveness of the proposed algorithm is demonstrated through three experiments: Artificial multivariate Gaussian data, Electroencephalography (EEG) eye-state detection data and robotic-control data. In Section 5 we draw some conclusion.

Notation Throughout this paper we denote by $\mathbb{R}$ and $\mathbb{C}$ the set of real and complex numbers, respectively. Bold lowercase letters represent vectors and bold uppercase letters represent matrices. $\Re\{(\cdot)\}$, $\Im\{(\cdot)\}$, $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ denote real part, imaginary part, complex conjugate, transpose, and conjugate transpose (Hermitian) of the argument, respectively. $\mathbf{0}_{m\times n}$, $\mathbf{1}_{m\times n}$, and $\mathbf{I}_m$ are the $m \times n$ all-zero, $m \times n$ all-one, and size-m identity matrices; $\mathbf{diag}(\cdot)$ is the diagonal matrix formed by the entries of the vector argument. For any $\mathbf{A} \in \mathbb{C}^{m\times n}$, $[\mathbf{A}]_{i,j}$ denotes its (i, j)th entry, $\|\mathbf{A}\|_p \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |[\mathbf{A}]_{i,j}|^p\right)^{\frac{1}{p}}$ is the pth entry-wise norm of $\mathbf{A}$, span($\mathbf{A}$) represents the vector subspace spanned by the columns of $\mathbf{A}$, rank($\mathbf{A}$) is the dimension of span($\mathbf{A}$), and null($\mathbf{A}^T$) is the kernel of span($\mathbf{A}$) (i.e., the nullspace of $\mathbf{A}^T$).

2. DETECTION OF CHANGE BY L1-NORM PCA

We are given a sequence of N points in a D dimensional real space, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_N$, where $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, 2, \ldots,$ N. We assume that there is one point of change at $i = p$ over the given data sequence where there is a shift in the subspace manifestation of the data. Our objective is to detect the point of change $p \in \{2, 3, \ldots,$ N-1$\}$. We define a running variable $l \in \{2, 3, \ldots,$ N-1$\}$ to split the data into two subsets, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{l-1}$ and $\mathbf{x}_l, \ldots, \mathbf{x}_N$, as illustrated in Figure 1.

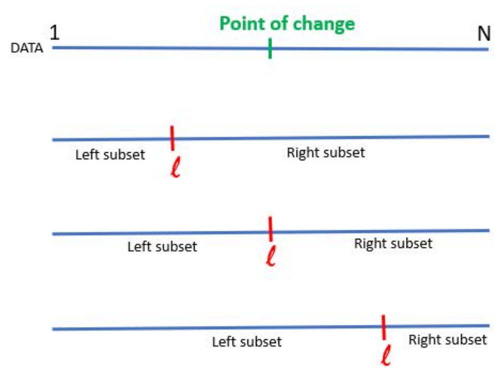


Figure 1: Visualization of left-right splitting operation by the running variable $l \in \{2, 3, ..., N-1\}$ for a data sequence of length N.

We organize the data subset to the left of l in the form of a matrix $\mathbf{X}_L(l) \in \mathbb{R}^{D\times(l-1)}$, that is $\mathbf{X}_L(l) = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{l-1}]$, and we define its low dimensional subspace $\mathbf{Q}_L(l) \in \mathbb{R}^{D\times r_1}$ of rank r_1 calculated by

$$\mathbf{Q}_L(l) = \underset{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}_{r_1}}{\arg\max} \; ||\mathbf{X}_L(l)^T\mathbf{Q}||_1. \tag{1}$$

where $||\mathbf{A}||_1 = \sum_{i,j} |A_{i,j}|$ for any matrix A.

Similarly, for the subset of data to the right of l, that is $\mathbf{x}_l$, $\mathbf{x}_{l+1} \ldots, \mathbf{x}_N$, we form the matrix $\mathbf{X}_R(l) \in \mathbb{R}^{D\times(N-(l-1))}$ and define its low dimensional subspace $\mathbf{Q}_R(l) \in \mathbb{R}^{D\times r_2}$ of rank r_2 by

$$\mathbf{Q}_R(l) = \underset{\mathbf{Q}^T\mathbf{Q}=\mathbf{I}_{r_2}}{\arg\max} \; ||\mathbf{X}_R(l)^T\mathbf{Q}||_1. \tag{2}$$

The r_1 columns of $\mathbf{Q}_L$ and the r_2 columns of $\mathbf{Q}_R$ in (1) and (2) are the so-called L1 principal components of the data subsets. Exact calculation of the L1 principal components can be carried out optimally using exhaustive search in exponential time,[9] or polynomial time[9] or suboptimally via the single-bit-flipping (SBF) fast algorithm.[10] In this work, in view of the computational requirements of our time-series analysis problem, we use the bit-flipping algorithm because it is computationally faster. In particular, we set $r_1 = r_2 = 1$ in (1) and (2) and use the bit-flipping algorithm to identify the first principal component only of the left and right data sets. Subsequent components are calculated conditionally, orthogonal to the previous components. That is, by *successive orthogonal L1-PCA* we calculate the first principal component $\mathbf{q}_L(l)$ by (1) and $\mathbf{q}_R(l)$ by (2) and then we project the left and right datasets onto their corresponding orthogonal subspaces by

$$\mathbf{X}^{\perp}_{L/R}(l) = \left(\mathbf{I_D} - \mathbf{q}_{L/R}(l)\mathbf{q}^T_{L/R}(l)\right)\mathbf{X}_{L/R}(l).$$

Then, we proceed recursively to calculate $r_1 - 1$ additional $L1$-norm principal components for $\mathbf{X}_L(l)$ and $r_2 - 1$ additional $L1$-norm principal components for $\mathbf{X}_R(l)$, for desired r_1 and r_2 values. The calculated PCs are organized in final matrix form as $\mathbf{Q}_L(l) \in \mathbb{R}^{D\times r_1}$ and $\mathbf{Q}_R(l) \in \mathbb{R}^{D\times r_2}$, respectively.

Subsequently, upon projection of the left and right subsets of the split data onto their respective bases, we calculate $v_L(l)$ and $v_R(l)$, respectively, by

$$v_L(l) = \max_{1<j<r_1} \sum_{i=1}^{l-1} \left| \left(\mathbf{X}^T_L(l)\mathbf{Q}_L(l)\right)_{i,j} \right|, \tag{3}$$

$$v_R(l) = \max_{1<j<r_2} \sum_{i=1}^{N-(l-1)} \left| \left(\mathbf{X}^T_R(l)\mathbf{Q}_R(l)\right)_{i,j} \right| \tag{4}$$

where $|\cdot|$ indicates absolute value. After computation of the data projection metrics $v_L(l)$, $v_R(l)$, we declare the point of change decision $\hat{p}$ by our proposed maximum left-right projection criterion

$$\hat{p} = \underset{2\leq l\leq N-1}{\arg\max} \left[v_L(l) + v_R(l)\right]. \tag{5}$$

The overall algorithm is summarized in Fig. 2. The process can be easily generalized from detection of change over of a block of data of size N to streaming data as we see in the following section.

3. STREAMING DATA

So far, we have seen the proposed L1-PCA detection of change method operating oven a static dataset. The method can be directly generalized to operate on streaming data. In particular, there are two approaches that we may follow.

3.1 Growing-Block model

Following the notation of Section II, for any given point l under investigation, we keep the left side of data $\mathbf{X}_L(l) = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{l-1}]$ constant in (1) and grow the right side of data $\mathbf{X}_R(l) = [\mathbf{x}_l, \mathbf{x}_{l+1}, \ldots, \mathbf{x}_n]$ in (2) as data come in. As $n \to \infty$, the quality of the estimate of $\mathbf{Q}_R(l)$ improves and so does our final decision $\hat{p}$ by (5).

To computationally improve the operation of the growing data-block model, we can introduce subspace tracking to update $\mathbf{Q}_R(l)$ as new data come in, instead of calculating a new subspace from scratch everytime. There are different ways to carry out subspace tracking and the one that we propose is the method in,[11] with the only difference that herein we just append data without any removal. We update the L1-norm subspace by

$$\mathbf{P}^{(k)}_t = \underset{\mathbf{P}\in R^{DXr}, \mathbf{P^T P}=\mathbf{I}_r}{\arg\max} ||(\mathbf{X}_t\mathbf{W}^{(k)}_t)^T\mathbf{P}||_1, k = 1, 2, ..., \tag{6}$$

where k is the iteration index, r is the rank (number of PCs), $\mathbf{X}_t$ is the data set of interest, and $\mathbf{W}^{(K)}_t$ is the data weighting operator at time t and iteration k.

Algorithm: L1-PCA and Maximum Left-Right Projection criterion

Input: $\mathbf{X}_{DxN}$ data matrix, l step size
$\mathbf{X}_{left} \leftarrow X[1 : l-1]$
$\mathbf{X}_{right} \leftarrow X[l : N]$
for $n = 2 : N/l$
    $\mathbf{Q}_L \leftarrow L1PCA\_SUCC(\mathbf{X}_{left})$
    $\mathbf{Q}_R \leftarrow L1PCA\_SUCC(\mathbf{X}_{right})$
    $v_L = \max\limits_{1<j<k} \sum_{i=0}^{D} \left| \left(\mathbf{X}_L^T \mathbf{Q}_L \right)_{ij} \right|$
    $v_R = \max\limits_{1<j<k} \sum_{i=0}^{D} \left| \left(\mathbf{X}_R^T \mathbf{Q}_R \right)_{ij} \right|$
$\hat{l} = \arg\max \left[v_L(l) + v_R(l) \right]$
Output: $\hat{l}$

Function *L1PCA_SUCC*

Input: X data matrix, n number of orthogonal PCs
$\mathbf{Q} \leftarrow L1PCA[X, 1]$
    for $i = 2 : n$
        $\mathbf{X}_{new} = [\mathbf{I}_D - \mathbf{Q}\mathbf{Q}^T]\mathbf{X}$
        $\mathbf{Q}_{new} \leftarrow L1PCA[X_{new}, 1]$
        $\mathbf{Q}(:, i) \leftarrow [\mathbf{Q}_{new}]$
end for
Output: $\mathbf{Q}$

Figure 2: $L1$-PCA algorithm for detecting a change in a vector process by $L1$-norm subspace comparisons.

3.2 Sliding window model

For simplicity in operation and to maintain a constant computational load, we may consider instead a sliding window model. In this case, as we add new data to right data set $\mathbf{X}_R$, we remove the same amount of data from the beginning of the left data set $\mathbf{X}_L$ to maintain a total data set size of N, which makes the operation a sliding window of length N that moves along the data stream. Certainly, the quality of decision making by (5) is formally a function of the window length N and so is the overall computational complexity. In any case, the point of change has to be inside the window.

4. EXPERIMENTAL STUDIES

Using the algorithmic procedures described in Sections 2 and 3, success has been achieved in detecting the point of change in *synthetic data* and *real-time based data* experiments that we carried out. Below, we give a description of the experiments and their respective results.

4.1 Multivariate Gaussian Data

We generate a sequence of N =200, $(D = 4)$-dimensional Gaussian vectors $\mathbf{x}_1$, $\mathbf{x}_2$, $\ldots$, $\mathbf{x}_{100}$, $\ldots$, $\mathbf{x}_{200}$, $\mathbf{x}_i \in \mathbb{R}^D$, $i = 1, \ldots, 200$. In particular,

$$\mathbf{x}_i \sim \mathscr{N}(\mathbf{0}, \mathbf{R}_1), i = 1, \ldots, 100,$$

and

$$\mathbf{x}_i \sim \mathscr{N}(\mathbf{0}, \mathbf{R}_2), i = 101, \ldots, 200,$$

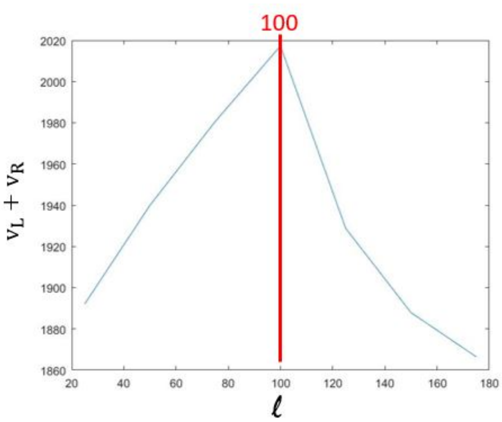
that is, there is a change in the autocorrelation matrix from

$$\mathbf{R}_1 = \begin{bmatrix} 16 & -12 & -12 & -16 \\ -12 & 25 & 1 & -4 \\ -12 & 1 & 17 & 14 \\ -16 & -4 & 14 & 57 \end{bmatrix}$$

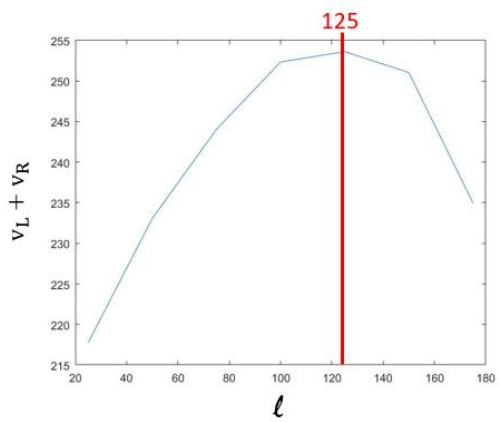
to

$$\mathbf{R}_2 = \begin{bmatrix} 30 & 40 & 50 & 60 \\ 40 & 54 & 68 & 82 \\ 50 & 68 & 86 & 104 \\ 60 & 82 & 104 & 126 \end{bmatrix}$$

at time $i = 101$. Our objective is to identify the time of change using the proposed $L1$-norm method and the proposed maximum left-right projection criterion. In Fig. 3, we plot the decision matrix $v_L(l) + v_R(l)$ in (5) with step size $\Delta l = 25$. Fig. 3(a) shows our findings when $L1$-PCA is deployed, while Fig. 3(b) deploys $L2$-PCA. It is pleasing to see not only the accuracy of the proposed method in Fig. 3(a), but also the sharpness of the proposed decision criterion.



(a) L1-PCA



(b) L2-PCA

Figure 3: Gaussian data sequence example of $N = 200$ points with time of change of autocorrelation matrix at $l = 100$ (Part (a) proposed $L1$-PCA subspace calculation; Part (b) conventional $L2$-PCA subspace calculation)

4.2 EEG Eye-State Detection

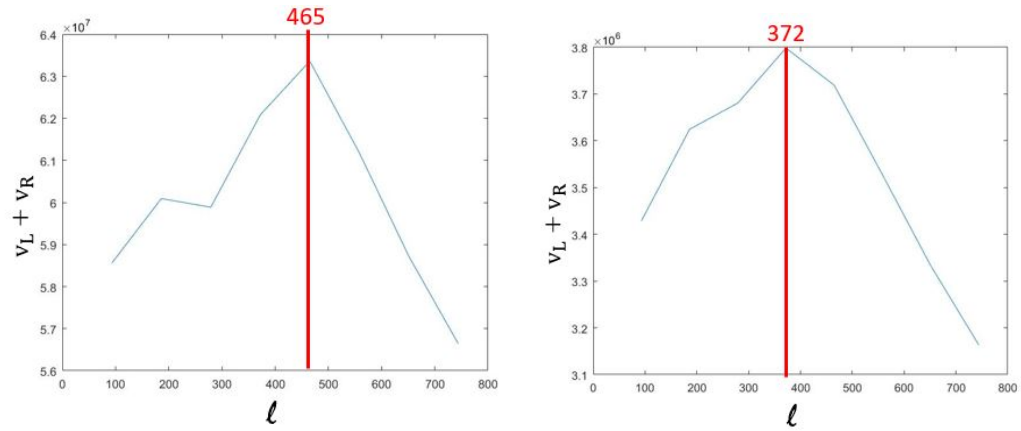
Next, we apply our algorithm to a real biomedical dataset. EEG is a recording of the electrical activity of the brain from the scalp. The dataset[14] consists of 14 EEG values (14 electrodes) of a subject with closed or open eyes. The duration of the measurements is 117 seconds. The ground truth of the eye state is detected via a camera during the EEG measurement and added to the file.

Our goal is to detect the point where the person changes state from closed eyes to open or vice versa without using the camera information*. We take a part of the dataset ($N = 767$) where there is one change and we use it as our data matrix. The true change of state is at $l = 465$, so we expect to see a maximum there. As we can see in Fig. 4, we have a clear maximum at exactly $l = 465$ under $L1$-PCA (Fig. 4(a)) but not under $L2$-PCA (Fig. 4(b)).

4.3 Robotic Control Data

We consider an electromyography (EMG) lower limb dataset in.[14] EMG is the study of muscle function through analysis of the electrical signals emanated during muscular contractions. The subjects in this dataset undergo three movements to analyze the behavior associated with the knee, gait, leg extension from a sitting position, and

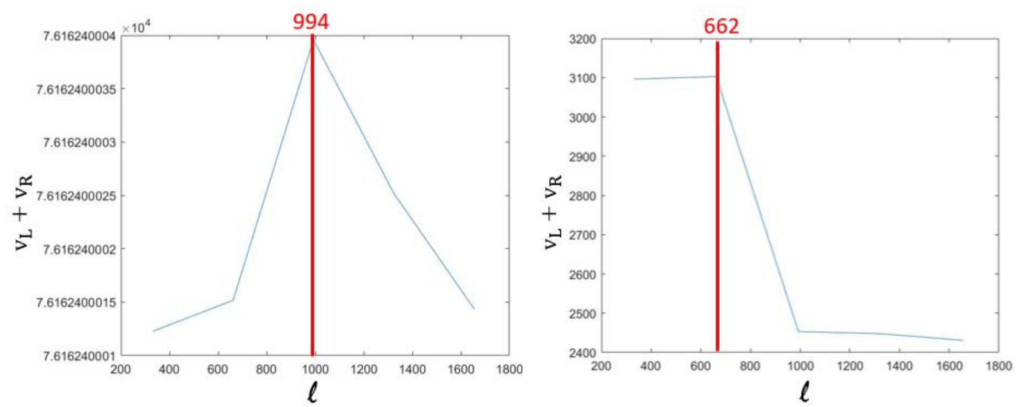
*Before applying our algorithm, we pre-process the raw EEG signals using EEGlab.[15]



(a) L1-PCA (b) L2-PCA

Figure 4: EEG eye-state dataset: Fourteen ($D = 14$) EEG channels to detect a change in eye-state of a subject (eyes open) that happens at time $l = 465$ over a data record of $N = 767$ points.

flexion of the leg up. The acquisition process was conducted with four electrodes, vastus medialis, semitendinosus, biceps femoris and rectus femoris, plus a goniometer at the knee ($D = 5$). In our analysis, two exercises were considered in our detection of change method: sitting and standing. We attempt to find when a person changes from sitting to standing or vice versa. Our data record has size $N = 1986$ and the time change of state happens at $l = 994$. In Fig. 5(a), we observe the sharp success of the proposed $L1$-PCA scheme (while $L2$-PCA fails, see Fig. 5(b)).



(a) L1-PCA (b) L2-PCA

Figure 5: EMG lower-limb dataset: $D = 5$ data channels to detect change of state (sitting/standing) that occurs at time $l = 994$ over $N = 1986$ time points.

5. CONCLUSIONS

We presented a novel algorithm to carry out detection of change in vector processes using successive orthogonal L1-PCA subspace calculation and a new subspace comparison metric that we called maximum left-right projection criterion. The algorithm was applied to synthetic and real datasets and returned at all times values of high discrimination value and accuracy.

REFERENCES

[1] G. D. Montanez, "Inertial Hidden Markov Models: Modeling Change in Multivariate Time Series," in Proc. *AAAI Conference on Artificial Intelligence*, Austin, 2015, pp 1819–1825.

[2] A. Singh, "Review article digital change detection techniques using remotely-sensed data", in *International Journal of Remote Sensing*, 1989, pp 989-1003.

[3] K. Pearson,"On lines and planes of closest fit to systems of points in space," *Philosophical magazine*, vol. 2, pp. 559–572, 1901.

[4] M. E. Wall, A. Rechtsteiner, and L. M. Rocha," Singular value decomposition and principal component analysis," in *A Practical Approach to Microarray Data Analysis*, Norwell: Springer, 2003, pp. 91–109.

[5] I. T. Jolliffe, *Principal Component Analysis, Second Edition*. Springer, 2002.

[6] M. Basseville, I.V. Nikiforov, *Detection of Abrupt Changes - Theory and Application*. Prentice-Hall, Inc., 1993.

[7] Q. Keand, T. Kanade,“Robust L1 factorization in the presence of outliers and missing data by alternative convex programming,” in *Proc. IEEE CVPR*, 2005, pp. 739–746.

[8] J. P. Brooks, J. H. Dulá, and E. L. Boone, “A pure L1-norm principal component analysis,” *Elsevier Comput. Stat. Data Anal.*, vol. 61, pp. 83–98, 2013.

[9] P. P. Markopoulos, G. N. Karystinos, and D. A. Pados, "Optimal algorithms for L1-subspace signal processing," *IEEE Transactions on Signal Processing*, vol. 62, Oct. 2014, pp. 5046-5058.

[10] P. P. Markopoulos, S. Kundu, S. Chamadia, and D. Pados, "Efficient L1-norm principal-component analysis via bit flipping," *IEEE Transactions on Signal Processing*, vol. 65, Aug. 2017, pp. 4252–4264.

[11] Y. Liu, K. Tountas, D. A. Pados, S. N. Batalama, M. J. Medley, "L1-Subspace Tracking for Streaming Data," *Pattern Recognition*, vol. 97, 2019.

[12] S. Kundu, P.P. Markopoulus, D. A. Pados, "Fast computation of the L1-principal component of real-valued data," in. *Proc. IEEE ICASSP*, 2014, pp.5722-5726.

[13] Y. Liu, D. A. Pados, S. N. Batalama, and M. J. Medley, "Iterative Re-weighted L1-Norm Principal Component Analysis," in. *Asilomar Conference on Signals, Systems, and Computers*, Oct. 2017, pp 425-429.

[14] D. Dua, C. Graff, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2019.

[15] A. Delorme, S. Makeig "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics." *Journal of Neuroscience Methods*, vol. 134, pp 9-21, 2004.