Change-Point Estimation of High-Dimensional Streaming Data via Sketching

Yuejie Chi[1] and Yihong Wu[2]
[1]Electrical and Computer Engineering, The Ohio State University, Columbus, OH
[2]Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL
Emails: chi.97@osu.edu, yihongwu@illinois.edu

***Abstract*—Change-point detection is of great interest in applications such as target tracking, anomaly detection and trend filtering. In many cases, it is also desirable to localize the change-point, if it exists. Motivated by the unprecedented scale and rate of modern high-dimensional streaming data, we propose a change-point detection and estimation procedure based on data sketching, which only requires a single sketch per high-dimensional data vector, by cyclically applying a small set of Gaussian sketching vectors. We demonstrate that when the underlying changes exhibit certain low-dimensional structures, such as sparsity, and the signal-to-noise ratio is not too small, the change-points can be reliably detected and located with a small number of sketching vectors based on filtering via convex optimization. Our procedure can be implemented in an online fashion to handle multiple change-points, since it sequentially operates on small windows of observations.**

***Index Terms*—streaming data, change-point detection, sketching, atomic norm**

I. Introduction

High-dimensional streaming data arises in many applications from video surveillance, social networks, to medical care. The volume and velocity of modern data generation make it prohibitive to either observe and store the complete data stream within the budget and time constraints of the sensing platform. In recent years, sketching has been advocated [1], [2] as an attractive manner to reduce the dimensionality of streaming data *on the fly* while still being able to preserve some key properties of the stream, such as graph cuts [3], principal subspace [4], and covariance structures [5]. Often, the data stream exhibits certain low-dimensional structures, which allow the number of required sketches to be much smaller than its ambient dimension, reducing the complexity in sensing and storage.

Extracting useful information from high-dimensional streaming data in an online fashion is of great interest to allow real-time situation awareness. In particular, change-point detection and estimation are greatly desirable for a number of applications such as target tracking [4], anomaly detection [6] and trend filtering [7]. Unfortunately, traditional change-point detection algorithms often require sampling of the complete data stream, and fail to recognize the inherit low dimensionality of the change-points, making them not scaling well to handle high-dimensional data [8].

This work is supported in part by NSF under grants CCF-1422966, CCF-1423088 and AFOSR under grant FA9550-15-1-0205.

In this paper, we're interested in designing data sketching schemes from which the change-points can be reliably detected and localized. We assume the change possesses certain low-dimensional structures, such as sparsity or low-rankness, which is well captured as a signal with small atomic norms [9]. For each of the data vector in the stream, a single sketch is obtained by applying a sketching vector cyclically chosen from a small set. The sketches are then used to construct an atomic norm minimization algorithm that detects and estimates the change-points. We further characterize the performance trade-off between the number of sketches, the low-dimensionality of the change-point, and the signal-to-noise ratio. Numerical examples are provided to demonstrate the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section II presents the problem formulation and review the backgrounds. Section III describes the proposed approach and its performance analysis. Section IV provides numerical simulations, and we conclude the paper in Section V.

II. Problem Formulation and Backgrounds

A. Data stream model

Consider a high-dimensional data stream, where a data vector $\boldsymbol{\beta}_t \in \mathbb{R}^p$ is generated at each time t, $1 \leq t \leq n$. For simplicity, we assume there exists only one change-point in the stream, at time $1 \leq t^\star \leq n-1$, and that

$$\boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_1, & 1 \leq t \leq t^\star \\ \boldsymbol{\beta}_2, & t^\star + 1 \leq t \leq n \end{cases}.$$

Note that the data stream is assumed to take a constant value before and after the change-point. Furthermore, we assume the data stream does not change significantly; in other words, the change, defined as

$$\boldsymbol{\Delta} \triangleq \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1,$$

has certain low-dimensional structures such as sparsity, which may be determined by a much smaller number of parameters than its ambient dimension p, but not the data vectors $\boldsymbol{\beta}_t$ themselves. This assumption is in the same spirit to those in sparse linear discriminant analysis cf. [**?**] where the mean vectors are assumed to differ by a small number of coordinates. Furthermore, in many applications assuming the sparsity of the change is far more practical than that of the data vectors. For example, in hyperspectral imaging [6], where $\boldsymbol{\beta}_t$ corresponds to the image frames which are typically of high entropy, the differences between consecutive frames only contain a few salient features due to changes in the scene. As another example, in sensor networks where $\boldsymbol{\beta}_t$ is the readings from p sensors, if a small number of sensors break down or are maliciously attacked, the resulting difference in measurements at the change-point is a sparse vector whose nonzero entries correspond to the affected sensors.

B. Sketching framework

Our goal is to develop a low-complexity data sketching scheme that can reliably detect and localize change-points. Sketching, when it is linear, reduces the dimensionality of the data stream by means of linear transformations, possibly noisy. Motivated by the recent work [5] where a single sketch per data vector is demonstrated sufficient for retrieving covariance structures of high-dimensional streaming data, we propose the following single-sketch observation model via random projections:

$$y_t = \boldsymbol{x}_t^T \boldsymbol{\beta}_t + \epsilon_t, \quad t = 1, \ldots, n. \tag{1}$$

where y_t is referred to as the sketch, $\boldsymbol{x}_t \in \mathbb{R}^p$ is the sketching vector, and the noise ϵ_t's are independent and distributed as $\mathcal{N}(0, \sigma^2/2)$. The sketching scheme (1) can be accomplished with a single pass of the data, since each entry of the data stream only needs to be accessed once in order to acquire the observation (1). This allows dimensionality reduction *on the fly*, where the complete data stream need not be stored.

C. Atomic norm for low-complexity signals

As mentioned earlier, the change $\boldsymbol{\Delta} = \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$ is assumed to possess certain low-dimensional structures, such that it may be determined by a much smaller number of parameters than its ambient dimension p. A general framework to model such parsimonious structures is the *atomic norm* [9]. Consider the representation of $\boldsymbol{\Delta}$ as

$$\boldsymbol{\Delta} = \sum_{i=1}^{k} c_i \boldsymbol{a}_i, \quad \boldsymbol{a}_i \in \mathcal{A}, \; c_i \geq 0, \tag{2}$$

where $\mathcal{A}$ is a set of atoms that constitutes simple building blocks of $\boldsymbol{\Delta}$, and k is the number of employed atoms, with $k \ll p$. The atomic norm of $\boldsymbol{\Delta}$ is defined as

$$\|\boldsymbol{\Delta}\|_{\mathcal{A}} = \inf_t \{t > 0 : \boldsymbol{\Delta} \in t \, \mathrm{conv}(\mathcal{A})\} \tag{3}$$

where $\mathrm{conv}(\mathcal{A})$ is the convex hull of $\mathcal{A}$. Many of the well-known structural-motivating norms are special cases of (3), such as the ℓ_1 norm for promoting sparsity [10], the nuclear norm for low-rankness [11], etc. For most parts of this paper, we will focus on the case when $\boldsymbol{\Delta}$ is k-sparse, but it is straightforward to generalize to other low-dimensional structures under the atomic norm framework.

III. Change-point Detection and Estimation

Under the single-sketch observation model, we propose to employ a set of m sketching vectors in a cyclic order for (1), where $\boldsymbol{x}_t = \boldsymbol{x}_{t+m}$, $\forall t$, and $\{\boldsymbol{x}_t\}_{t=1}^m$ consists of i.i.d. Gaussian vectors drawn from $\mathcal{N}(0, \frac{1}{p} I_p)$.

For $m \leq t \leq n-1$, we further define the sketch difference between two consecutive sketches exploiting the same sketching vector as

$$\begin{aligned} z_{t+1} &= y_{t+1} - y_{t-m+1} \\ &= \boldsymbol{x}_{t+1}^T (\boldsymbol{\beta}_{t+1} - \boldsymbol{\beta}_{t-m+1}) + \epsilon_{t+1} - \epsilon_{t-m+1} \\ &= \boldsymbol{x}_{t+1}^T \boldsymbol{\eta}_{t+1} + \tilde{\epsilon}_{t+1}, \end{aligned} \tag{4}$$

where $\tilde{\epsilon}_{t+1} = \epsilon_{t+1} - \epsilon_{t-m+1} \sim \mathcal{N}(0, \sigma^2)$ and

$$\boldsymbol{\eta}_{t+1} = \begin{cases} \mathbf{0}, & m \leq t \leq t^\star - 1 \\ \boldsymbol{\Delta}, & t^\star \leq t \leq t^\star + m - 1 \\ \mathbf{0}, & t^\star + m \leq t \leq n - 1 \end{cases}.$$

The proposed change-point detection algorithm (for a single sparse change-point) can be described below as:

- **Input:** the sketches $\{y_t\}_{t=1}^n$, the sketching vectors $\{\boldsymbol{x}_t\}_{t=1}^m$, the regularization parameter λ, the threshold γ;
- **Filtering by LASSO:** for $m \leq t \leq n-m$, compute the solution of the following problem:

$$\begin{aligned} \hat{\boldsymbol{\eta}}_t &= \operatorname*{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^p} \sum_{i=t+1}^{t+m} (z_i - \boldsymbol{x}_i^T \boldsymbol{\eta})^2 + \lambda \|\boldsymbol{\eta}\|_1 \\ &= \operatorname*{argmin}_{\boldsymbol{\eta} \in \mathbb{R}^p} \|\boldsymbol{z}_t - \boldsymbol{X}_t \boldsymbol{\eta}\|_2^2 + \lambda \|\boldsymbol{\eta}\|_1, \end{aligned} \tag{5}$$

  where $\boldsymbol{z}_t = [z_{t+1}, z_{t+2}, \ldots, z_{t+m}]^T$, $\boldsymbol{X}_t = [\boldsymbol{x}_{t+1}, \boldsymbol{x}_{t+2}, \ldots, \boldsymbol{x}_{t+m}]^T$. Though it appears that we need to solve a different convex program at each time, at consecutive times the convex programs only differ by two measurements (the measurement matrix $\boldsymbol{X}_t$ is actually fixed across time up to a row shift) and thus can be solved efficiently via, e.g., homotopy-based approaches [12].
- **Thresholding:** Set $\xi_t = \|\hat{\boldsymbol{\eta}}_t\|_2$. If $\max_{m \leq t \leq n-m} \xi_t < \gamma$, declare there are no change-points; otherwise declare there is a change-point and determine its location as

$$\hat{\tau} = \operatorname*{argmax}_{m \leq t \leq n-m} \xi_t.$$

The key idea of the approach is explained by Fig. 1. At each candidate change-point t, as in Fig. 1 (a), we examine the sketch difference between the windows of length m before and after t, which can be written as (4). The difference is maximized at the change-point, when the sketch difference $\boldsymbol{z}_{t^\star}$ compressively measures the sparse difference $\boldsymbol{\Delta}$, as shown in Fig. 1 (b). The LASSO algorithm (5) allows denoising and identification of the embedded vector $\boldsymbol{\Delta}$ from a number of sketches much smaller than p by exploiting its sparsity. Moreover, a window of m sketches needs to be stored instead of all sketches.

A. Performance Analysis

For appropriately selected window size, the proposed approach is guaranteed to detect and localize the change-point. as shown by the following theorem.

Theorem 1 (Single change-point). *Set* $\lambda = 4\sigma\sqrt{\log p}$. *The change-point can be detected with probability at least* $1 -$

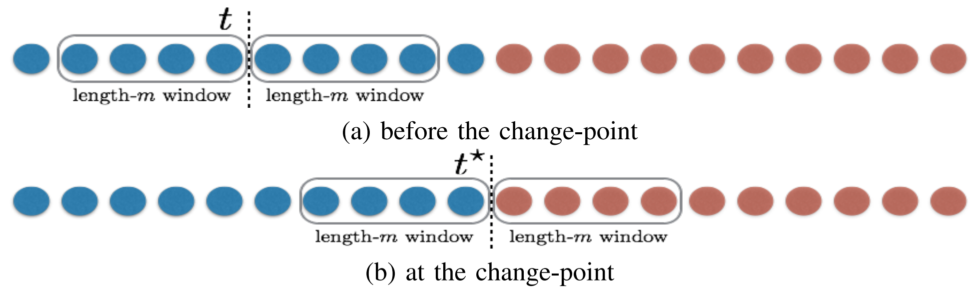

Fig. 1. Illustration of the proposed data sketching and change-point detection procedure.

$c_1 np^{-c_2}$, *if the window size* m *satisfies*

$$\|\mathbf{\Delta}\|_2 > C\sigma\sqrt{\frac{k\log p}{m}}, \tag{6}$$

where c_1, c_2, C *are some universal constants. The threshold can be selected as* $\gamma = \|\mathbf{\Delta}\|_2/2$. *Moreover, the detected change-point satisfies* $|\hat{t} - t^\star| < m$.

Proof. When $m \le t \le t^\star - m$ or $t^\star + m \le t \le n - m$, the signal underlying the LASSO problem (5) can be regarded as $\boldsymbol{\eta}_t = 0$, where the noise vector $\boldsymbol{\epsilon} = [\tilde{\epsilon}_{t+1}, \ldots, \tilde{\epsilon}_{t+m}]^T$ contains i.i.d. Gaussian entries $\mathcal{N}(0, \sigma^2)$. From [13, Corollary 1] and the union bound, with probability at least $1 - c_1(n - 3m + 1)p^{-c_2}$, we have

$$\xi_t = \|\hat{\boldsymbol{\eta}}_t - \mathbf{0}\|_2 \le C\sigma\sqrt{\frac{k\log p}{m}},$$

for some constants c_1, c_2, C. On the other hand, at the change-point $t = t^\star$, the signal underlying the LASSO problem (5) can be regarded as $\boldsymbol{\eta}_{t^\star} = \mathbf{\Delta}$, with the noise similarly contains i.i.d. Gaussian entries $\mathcal{N}(0, \sigma^2)$. With probability at least $1 - c_1 p^{-c_2}$, we have

$$\begin{aligned} \xi_{t^\star} &= \|\hat{\boldsymbol{\eta}}_{t^\star} - \mathbf{\Delta} + \mathbf{\Delta}\|_2 \\ &\ge \|\mathbf{\Delta}\|_2 - \|\hat{\boldsymbol{\eta}}_{t^\star} - \mathbf{\Delta}\|_2 \\ &\ge \|\mathbf{\Delta}\|_2 - C\sigma\sqrt{\frac{k\log p}{m}}. \end{aligned}$$

Therefore, as long as $\|\mathbf{\Delta}\|_2 \ge 2C\sigma\sqrt{\frac{k\log p}{m}}$, by setting $\gamma = \|\mathbf{\Delta}\|_2/2$, we can detect the existence of the change-point with high probability. Moreover, ξ_t that is a window-length away from the actual change-point $t^\star$ is below the threshold, therefore making sure the localization accuracy is within m. □

A few discussions are in order. First, from (6) the choice of the window length m depends on the SNR $\triangleq \|\mathbf{\Delta}\|_2^2/\sigma_2^2$, so that

$$m \gtrsim \frac{k\log p}{\text{SNR}},$$

which implies the localization accuracy improves with the increase of the SNR, and we can use fewer sketching vectors if the sparsity level of the change-point is small.

To extend to *other low-dimensional structures*, one simply replace the ℓ_1 norm by the atomic norm $\|\cdot\|_{\mathcal{A}}$, and the window length m will be similarly constrained [9] by the following metric:

$$m \gtrsim \frac{w^2(T_{\mathcal{A}}(\mathbf{\Delta}) \cap \mathbb{S}^{p-1})}{\text{SNR}}$$

where $T_{\mathcal{A}}(\mathbf{\Delta}) = \text{cone}\{\boldsymbol{z} - \mathbf{\Delta} : \|\boldsymbol{z}\|_{\mathcal{A}} \le \|\mathbf{\Delta}\|_{\mathcal{A}}\}$, and $w(K) \triangleq \mathbb{E}[\sup\{y^T g : y \in K\}]$ for $g \sim \mathcal{N}(0, I_p)$ denotes the *Gaussian width* of a set K.

B. Comparisons to Related Work

Our approach is different from the line of work on quickest change-point detection [14], which focuses on detecting the onset of the change-point as quickest as possible. Our approach is inspired by the interesting work of Soh and Chandrasekaran [8] considered an approach that combines filtering and proximal denoising with by assuming the data vectors possess low-dimensional structures rather than the change. Moreover, no sketching is applied to the data to reduce dimensionality. Xie et.al. considered change-point detection for partially observed data stream when the data vectors lie in a low-dimensional subspace [15]. Multiple change-point detection has been studied as a variable selection problem by [16] using LASSO for scalar processes and by Angelosante and Giannakis [17] using sparse group LASSO [18] for vector processes, by assuming the number of change-points is small and the corresponding changes are sparse. Zhang et. al. [19] extended the approach in [17] to linear regression models which appears in a similar form as our sketching model. However, these approaches are offline and require batch processing.

IV. Numerical Examples

In this section, we provide a few numerical experiments to demonstrate the effectiveness of the proposed approach. Our approach can be applied even with multiple change-points, as long as the spacing between change points exceeds $2m$. In this case, we will first threshold to find neighborhoods that contain a single change-point, and then search locally for peaks to determine each change-point.

We set $p = 200$, and $n = 500$. Let the change-points occur at $t^\star \in \{100, 200, 300, 400\}$ with the change $\mathbf{\Delta}$ at each change-point being a k-sparse vector with uniformly selected support and non-zeros drawn uniformly at random from $\{\pm 1\}$. We randomly generate m sketching vectors whose entries are drawn i.i.d. from $\mathcal{N}(0, 1/p)$, and the noise in the sketch is drawn i.i.d. from $\mathcal{N}(0, \sigma^2)$. The regularization parameter for LASSO is set as $\lambda = \sigma\sqrt{2\log p}$. We first compare the proposed scheme against a baseline change-point detector, which is to directly compute the energy of the sketch difference $\|\boldsymbol{z}_t\|_2$ in each sketch window without performing LASSO. Fig. 2 compares $\|\boldsymbol{z}_t\|_2$ with $\xi_t = \|\hat{\boldsymbol{\eta}}_t\|_2$ from the proposed approach when $\sigma = 0.05$, $k = 3$ and $m = 20$. It is evident that LASSO helps filtering the noise in the sketches by motivating the sparsity of the change vector, therefore the *relative* difference in the computed statistics is much more significant after filtering by the LASSO.

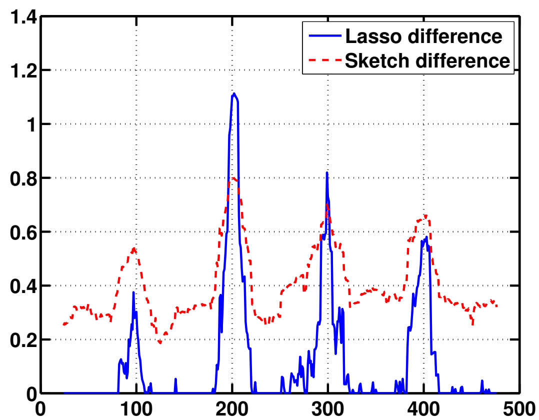

Fig. 2. Comparison of the statistics from the proposed change-point detection algorithm (marked as Lasso difference) with a baseline approach (marked as Sketch difference) without filtering by LASSO.

Next we examine the effects of the noise level σ, and the number of sketches m on the performance of the proposed approach when $k = 4$. Fig. 3 shows the value of ξ_t of the proposed approach against the time index under different parameter settings. The calculated statistic ξ_t exhibits more fluctuations in its value with the increase of the noise level and the decrease of the window size. Also, the peaks of the ξ_t are less localized with the increase of the window size.

V. Conclusions

This paper proposes a low-complexity sketching scheme for detecting and estimating change-points from only a single pass of high-dimensional streaming data. A set of sketching vectors is applied cyclically to the streaming data to obtain a single sketch per data vector. The scheme exploits parsimonious representations of the change-point, and uses convex optimization to improve the detection performance and recover the change-point. Furthermore, it can be implemented in an online fashion, and can handle multiple change-points as long as they are not too close. In the future work, it is interesting to further study the trade-offs between localization accuracy, sketching complexity, and the signal-to-noise ratio.

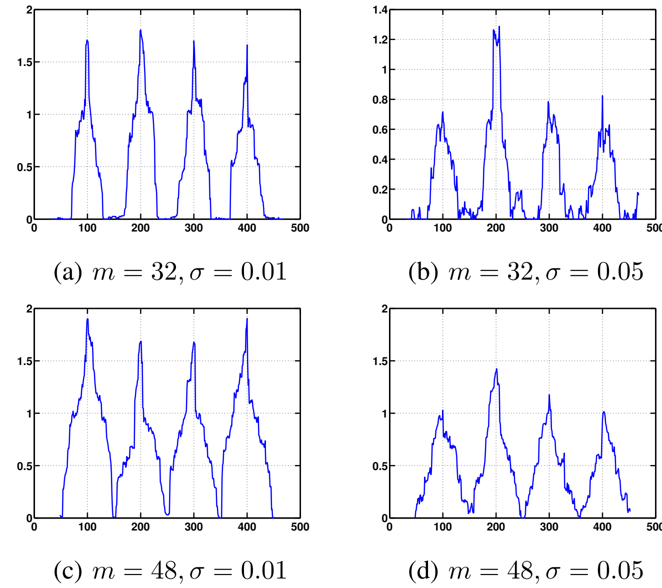

Fig. 3. The calculated statistic ξ_t of the proposed approach for various m and σ when $p = 200$ and $k = 4$.

References

[1] S. Muthukrishnan, *Data streams: Algorithms and applications*. Now Publishers Inc, 2005.
[2] K. Slavakis, G. Giannakis, and G. Mateos, "Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge," *Signal Processing Magazine, IEEE*, vol. 31, no. 5, pp. 18–31, 2014.
[3] K. J. Ahn, S. Guha, and A. McGregor, "Graph sketches: sparsification, spanners, and subgraphs," in *Proceedings of the 31st symposium on Principles of Database Systems*. ACM, 2012, pp. 5–14.
[4] Y. Chi, Y. C. Eldar, and R. Calderbank, "Petrels: Parallel subspace estimation and tracking by recursive least squares from partial observations," *Signal Processing, IEEE Transactions on*, vol. 61, no. 23, pp. 5947–5959, 2013.
[5] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *Information Theory, IEEE Transactions on*, vol. 61, no. 7, pp. 4034–4059, July 2015.
[6] J. Meola, M. T. Eismann, R. L. Moses, and J. N. Ash, "Detecting changes in hyperspectral imagery using a model-based approach," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 7, pp. 2647–2661, 2011.
[7] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "ℓ_1 trend filtering," *SIAM review*, vol. 51, no. 2, pp. 339–360, 2009.
[8] Y. S. Soh and V. Chandrasekaran, "High-dimensional change-point estimation: Combining filtering with convex optimization," *arXiv preprint arXiv:1412.3731*, 2014.
[9] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex algebraic geometry of linear inverse problems," *48th Annual Allerton Conference on Communication, Control, and Computing*, pp. 699–703, 2010.
[10] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, April 2006.
[11] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
[12] M. Asif and J. Romberg, "Sparse recovery of streaming signals using ℓ_1-homotopy," *Signal Processing, IEEE Transactions on*, vol. 62, no. 16, pp. 4209–4223, Aug 2014.
[13] S. Negahban, B. Yu, M. J. Wainwright, and P. K. Ravikumar, "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers," in *Advances in Neural Information Processing Systems*, 2009, pp. 1348–1356.
[14] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press Cambridge, 2009, vol. 40.
[15] Y. Xie, J. Huang, and R. Willett, "Change-point detection for high-dimensional time series with missing data," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 1, pp. 12–27, 2013.
[16] C. Levy-leduc and Z. Harchaoui, "Catching change-points with lasso," in *Advances in Neural Information Processing Systems*, 2008, pp. 617–624.
[17] D. Angelosante and G. B. Giannakis, "Group lassoing change-points in piecewise-constant ar processes," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–16, 2012.

[18] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A sparse-group lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[19] B. Zhang, J. Geng, and L. Lai, "Multiple change-points estimation in linear regression models via sparse group lasso," *Signal Processing, IEEE Transactions on*, vol. 63, no. 9, pp. 2209–2224, May 2015.