ARTICLE IN PRESS
YACHA:1099

Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

High-dimensional change-point estimation: Combining filtering with convex optimization

Yong Sheng Soh [a,*], Venkat Chandrasekaran [a,b]

[a] Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, United States
[b] Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, United States

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form 3 November 2015
Accepted 6 November 2015
Available online xxxx
Communicated by Guy Nason

Keywords:
High-dimensional time series
Convex geometry
Atomic norm thresholding
Filtered derivative

ABSTRACT

We consider change-point estimation in a sequence of high-dimensional signals given noisy observations. Classical approaches to this problem such as the filtered derivative method are useful for sequences of scalar-valued signals, but they have undesirable scaling behavior in the high-dimensional setting. However, many high-dimensional signals encountered in practice frequently possess latent low-dimensional structure. Motivated by this observation, we propose a technique for high-dimensional change-point estimation that combines the filtered derivative approach from previous work with convex optimization methods based on atomic norm regularization, which are useful for exploiting structure in high-dimensional data. Our algorithm is applicable in online settings as it operates on small portions of the sequence of observations at a time, and it is well-suited to the high-dimensional setting both in terms of computational scalability and of statistical efficiency. The main result of this paper shows that our method performs change-point estimation reliably as long as the product of the smallest-sized change (the Euclidean-norm-squared of the difference between signals at a change-point) and the smallest distance between change-points (number of time instances) is larger than a Gaussian width parameter that characterizes the low-dimensional complexity of the underlying signal sequence.

1. Introduction

Change-point estimation is the identification of abrupt changes or anomalies in a sequence of observations. Such problems arise in numerous applications such as product quality control, data segmentation, network analysis, and financial modeling; an overview of the change-point estimation literature can be found in [1–4]. As in other inferential tasks encountered in contemporary settings, a key challenge underlying many modern change-point estimation problems is the increasingly large dimensionality of the underlying sequence

* Corresponding author.
E-mail addresses: ysoh@caltech.edu (Y.S. Soh), venkatc@caltech.edu (V. Chandrasekaran).

http://dx.doi.org/10.1016/j.acha.2015.11.003
1063-5203/

Please cite this article in press as: Y.S. Soh, V. Chandrasekaran, High-dimensional change-point estimation: Combining filtering with convex optimization, Appl. Comput. Harmon. Anal. (2015), http://dx.doi.org/10.1016/j.acha.2015.11.003

of signals – that is, the signal at each location in the sequence is not scalar-valued but rather lies in a high-dimensional space. This challenge leads both to computational difficulties as well as to complications with obtaining statistical consistency in settings in which one has access to a small number of observations (relative to the dimensionality of the space in which these observations live).

A prominent family of methods for estimating the locations of change-points in a sequence of noisy scalar-valued observations is based on the *filtered derivative* approach [1,5–8]. Broadly speaking, these procedures begin with an application of a low-pass filter to the sequence, followed by a computation of pairwise differences between successive elements, and finally the implementation of a thresholding step to estimate change-points. A large body of prior literature has analyzed the performance of this family of algorithms and its variants [5,7,8]. Unfortunately, as we describe in Section 3, the natural extension of this procedure to the high-dimensional setting leads to performance guarantees for reliable change-point estimation that require the underlying signal to remain unchanged for long portions of the sequence. Such requirements tend to be unrealistic in applications such as financial modeling and network analysis in which rapid transitions in the underlying phenomena trigger frequent changes in the associated signal sequences.

1.1. Our contributions

To alleviate these difficulties, modern signal processing methods for high-dimensional data – in a range of statistical inference tasks such as denoising [9–11], model selection [12–14], the estimation of large covariance matrices [15,16], and others [17–21] – recognize and exploit the observation that signals lying in high-dimensional spaces typically possess low-dimensional structure. For example, images frequently admit sparse representations in an appropriately transformed domain [17,22] (e.g., the wavelet domain), while covariance matrices are well-approximated as low-rank matrices in many settings (e.g., correlations between financial assets). The exploitation of low-dimensional structure in solving problems such as denoising leads to consistency guarantees that depend on the intrinsic low-dimensional "complexity" of the data rather than on the ambient (large) dimension of the space in which they live. A notable feature of several of these structure-exploiting procedures is that they are based on convex optimization methods, which can lead to tractable numerical algorithms for large-scale problems as well as to insightful statistical performance analyses. Motivated by these ideas, we propose a new approach for change-point estimation in high dimensions by integrating a convex optimization step into the filtered derivative framework (see Section 3). We prove that the resulting method provides reliable change-point estimation performance in high-dimensional settings, with guarantees that depend on the underlying low-dimensional structure in the sequence of observations rather than on their ambient dimension.

To illustrate our ideas and arguments concretely, we consider a setup in which we are given a sequence $\mathbf{Y}[t] \in \mathbb{R}^p$ for $t = 1, \ldots, n$ of observations of the form:

$$\mathbf{Y}[t] = \mathbf{X}^{\star}[t] + \boldsymbol{\varepsilon}[t]. \tag{1}$$

Here $\mathbf{X}^{\star}[t] \in \mathbb{R}^p$ is the underlying signal and the noise is independent and identically distributed across time as $\boldsymbol{\varepsilon}[t] \sim \mathcal{N}(0, \sigma^2 I_{p\times p})$. The signal sequence $\mathcal{X} := \{\mathbf{X}^{\star}[t]\}_{t=1}^{n}$ is assumed to be piecewise constant with respect to t. The set of change-points is denoted by $\tau^{\star} \subset \{1, \ldots, n\}$, i.e., $t \in \tau^{\star} \Leftrightarrow \mathbf{X}^{\star}[t] \neq \mathbf{X}^{\star}[t+1]$, and the objective is to estimate the set $\tau^{\star}$. A central aspect of our setup is that each $\mathbf{X}^{\star}[t]$ is modeled as having an efficient representation as a linear combination of a small number of elements from a known set $\mathcal{A}$ of building blocks or atoms [9,10,17–19,21,23–26]. This notion of signal structure includes widely studied models in which signals are specified by sparse vectors and low-rank matrices. It also encompasses several others such as low-rank tensors, orthogonal matrices, and permutation matrices. The convex optimization step in our approach exploits knowledge of the atomic set $\mathcal{A}$; specifically, the algorithm described in Section 3.2 consists of a denoising operation in which the underlying signal is estimated from local averages of the