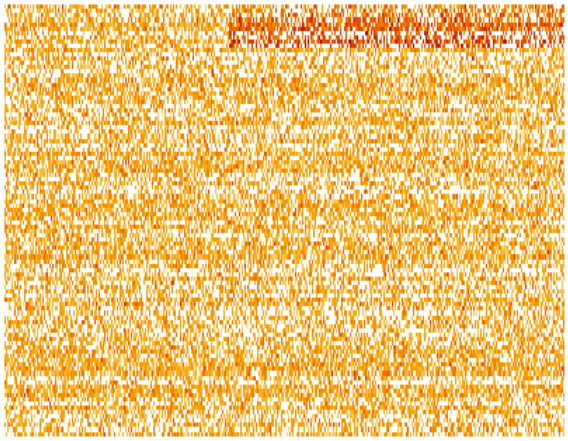
High dimensional changepoint estimation with missing data via sparse projection

MAP594 Internship report

Bertille Follain
under the supervision of Richard J. Samworth and Tengyao Wang

DPMMS, University of Cambridge

April-August 2020



Déclaration d'intégrité relative au plagiat

Je soussignée Bertille Follain, X2017

- déclare que ce rappot de stage est un document original fruit d'un travail personnel
- suis au fait que la loi sanctionne sévèrement la pratique qui consiste à prétendre être l'auteur d'un travail écrit par une autre personne
- atteste que le travail que je n'ai pas réalisé moi-même se distingue clairement dans le corps du rapport et renvoie bien à son ou ses auteurs

Summary of scientific work

Changepoints are often observed in 'big data' that is collected in the shape of a data stream. We study high-dimensional time series which are composed of multiple segments with their own mean. The change in mean structure occurs in a sparse subset of the coordinates. On top of that, measurement devices can be faulty, creating missing data in different manners across different coordinates. The goal is to use the different coordinates to detect smaller changes than what could be done through only one coordinate time series. Our two steps procedure, called Inspect Miss, for the estimation of the changepoint consists in the following: first, we argue that a good projection direction can be obtained as the leading left singular vector of the matrix that solves an optimisation problem derived from the cusum transformation of the time series (adapted for the missingness). We then use an existing univariate changepoint estimation algorithm on the projected cumulative transformation matrix. Our theory provides strong guarantees on the rate of convergence of the changepoint location, and our numerical studies validate its competitive empirical performance for a wide range of data-generating mechanisms and missingness-generating mechanisms. Software implementing the methodology is available in the **R** package **InspectChangepointMissingness**.

Key words: changepoint estimation; Optimisation; Dimension reduction; Piecewise stationary; Segmentation; Sparsity; High-dimensional, Missing data

Team presentation

(a) Pr. Richard Samworth

(b) Dr. Tengyao Wang

(c) Bertille Follain, myself

Figure 2: High-dimensional changepoint detection with missing data team members

I carried out my 3rd year internship at the DPMMS (Department of Pure Mathematics and Mathematical Statistics), University of Cambridge, which is where I will also study for my 4A (Master of Advanced Study in Mathematical Statistics as a member of Trinity College and the recipient of the Knox Studentship). My main supervisor was Pr. Richard Samworth, director of the Statistical Laboratory (sub-department of the DPMMS) whose research focuses on non-parametric and high-dimensional statistics, while my other supervisor was Dr. Tengyao Wang, Lecturer at UCL and former postdoc of Pr. Samworth. We met weekly for an hour to discuss my ideas and further directions I should look into. I chose this internship because I was interested in high-dimensional problems and really appreciated Pr. Samworth's work, and I had the desire to look into the UK's academic scene. I carried out the internship working remotely from France and will continue working on the project during the next academic year to (hopefully) bring it to publication. I intend to continue my studies by pursuing a PhD either in France or in the UK, with the area of focus still a lingering question.

I would like to thank Pr. Richard Samworth and Dr. Tengyao Wang for their help and support during this internship. Their supervision was excellent and is responsible for an important part of the happiness I felt doing research at Cambridge.

Contents

List of Figures

List of Algorithms

List of Tables

1 Introduction

In the era of Big Data, most interesting and challenging collected data does not follow the trope of the independent and identically distributed (IID) setting. This causes new issues and challenges as the classical statistical procedures may not easily adapt to these different data-generating mechanisms. There are many ways in which heterogeneity may appear in data: missing data, data from different sources, correlations, etc...

Collecting data overtime, through physical measures for example, can result in non-stationarity and missingness. The most simple case of such non-stationarity may be when the population mean changes at precise points in time, those being rather rare. These 'changepoints' can help divide the time series data in different segments, later studied through methods designed for stationary time series. The missingness in the data can appear through many different mechanisms, but a reasonable assumption is that the missingness is not dependent on the value that should have been measured.

In this work, we study the problem of high dimensional time series with heterogeneous missingness and, for the most part, a single changepoint. We assume that the changepoint occurs on a sparse subset of the coordinates and that the missingness does not depend on the data. However, the missingness can be of many forms, be it random or not. We shall make no further assumptions for the creation and use of the algorithm.

This problem, without considering the important factor of missingness has been studied in the Inspect paper [Wang and Samworth, 2018]. Even without missingness, this problem is particularly interesting in terms of applications, as was mentioned in the Inspect paper. For instance, in the case of stock price data, it may well be the case that stocks in related industry sectors experience virtually simultaneous 'shocks' [Jandhyala et al., 2002]. In internet security monitoring, a sudden change in traffic at multiple routers may be an indication of a distributed denial of service attack [Peng et al., 2004]. In functional Magnetic Resonance Imaging (fMRI) studies, a rapid change in blood oxygen level dependent (BOLD) contrast in a subset of voxels may suggest neurological activity of interest [Aston and Kirch, 2013].

Adding missingness to this problem statement enhances the already wide range of possible applications. Measuring device can be faulty, either at random or be broken and then repaired, leaving multiple days without measures. If different coordinates are gathered by different organisations, their frequency of measure may be different, or some started to be measured way before others. There may be dependencies between the missingness in different coordinates and/or at the different measuring times. Our methodology assumes nothing more than the independence of the missingness (or observation matrix) and the actual data. As for the theoretical results, they will be provided in a specific case, that of homogeneous per row missingness (see **Section 4**).

The main contribution of this work is to provide a methodology for the sparse multiple changepoint setting with heterogeneous missingness, with theoretical results in the single changepoint setting. The sparsity of the change is challenge-inducing, as the coordinates which undergo the change are unknown. In brief, the method consists in finding the best projection direction possible to transform the high dimensional CUSUM transformation of the data to a one dimensional problem. The CUSUM transformation must be first adapted to the missingness in the data, as it is otherwise not calculable. The best projection is then chosen as the sparse leading left singular vector of the CUSUM matrix. Then, an equivalent of a simple univariate procedure is applied. Our algorithm is therefore named Inspect Miss, short for informative sparse projection for estimation of changepoint with missingness; it is implemented in the **R** package **InspectChangepointMissingness** (see **Appendix 2**).

To be more specific, in the single changepoint setting with heterogeneous missingness that we con-

sider, we first look at the adapted CUSUM transformation for missingness. On each row, we calculate this CUSUM transformation, which is different on different rows as it depends on the missingness of each row. To transform this quantity to a one-dimensional vector, we therefore look for the leading left singular vector of the adapted CUSUM matrix, and then project it. This makes sense as the sparsity of the left leading singular vector of the adapted CUSUM transformation of the mean matrix is the same as the sparsity of the vector of mean change. Finding the sparse leading left singular vector of a matrix is an NP-hard problem. We therefore find an approximate solution through a relaxation of the optimisation that describes such vectors. The changepoint is then defined as the argmax of the absolute value of the projected quantity. For the multiple changepoint setting, the single changepoint method is combined with the method of Wild Binary Segmentation of [Fryzlewicz, 2014].

Illustration of the Inspect Miss algorithm is given in **Figure 3**. First we simulate a 1000 by 2000 data matrix with independent normal columns with identity covariance and with one changepoint at location 800. The change occurs in 50 coordinates. The vector of mean change had l_2 norm of 1.5, uniformly shared across the coordinates of change. The missingness is homogeneous, with probability of missing 0.5 (or probability of presence 0.5), independently over all measures. **Figure 3 (a)** represents the data matrix, white representing missingness. **Figure 3 (b)** represents the missingness-adapted CUSUM transformation while **Figure 3 (c)** shows the full data without missingness. **Figure 3 (d)** shows the projected CUSUM transformation and therefore the estimated changepoint. We wish to remind the reader that we solely focus on the offline problem, where the entire dataset is available to us before we aim to detect the changepoints. The on-line problem, where the aim is to detect the changepoint as soon as it occurs is truly interesting [Chen et al., 2020], but is not our concern here.

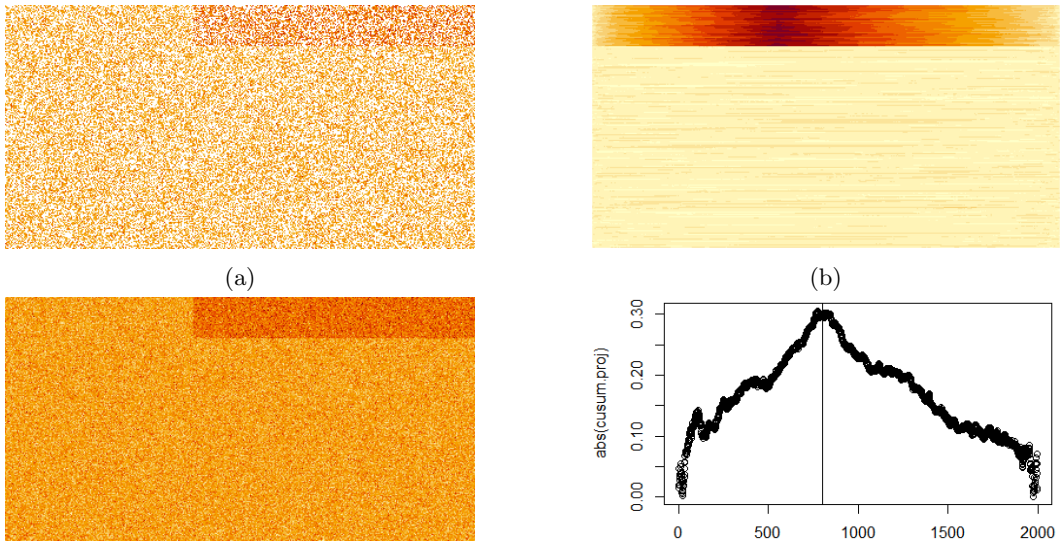


Figure 3: Example of the Inspect Miss algorithm in action: (a) visualisation of the data matrix with missingness, (b) adapted CUSUM transformation of the data in (a), visualisation of the full data matrix (c), projected CUSUM transformation (d).

Our theoretical results show that the estimated projection direction does not deviate too much in angle from the ideal left-leading singular vector of the adapted CUSUM transformation of the mean matrix. We then focus on the homogeneous per row case, which shows that under appropriate conditions, the proven rate of convergence in the single changepoint estimation case is $\rho_n = \sqrt{\{\log n\}/n}$. The minimax rate is $1/n$, but the numerical studies show the great performance of our algorithm in

many missingness and data-generating mechanisms.

Changepoints problems have been studied at least since the 90's by [Page, 1955] and has shown to have applications in many different areas, such as genetics [Olshen et al., 2004], disease outbreak watch [Sparks et al., 2010] and aerospace engineering [Henry et al., 2010], as well as those previously mentioned. Literature on changepoint detection and localisation is on the rise, especially in the univariate case. Surveys of such methods can be found in [Csörgö and Horváth, 1997] and [Horváth and Rice, 2014]. For the univariate problem, we can cite the pruned exact linear time method [Killick et al., 2012], wild binary segmentation [Fryzlewicz, 2014] and simultaneous multiscale changepoint estimator [Frick et al., 2013] as state of the arts methods.

Some of the univariate methods have been adapted for multiple changepoints detection, such as [Ombao et al., 2005], [Aue et al., 2009] and [Kirch et al., 2014]. In the high-dimensional setting, fewer methods are available, particularly if we suppose some sparsity structure of the mean change. [Enikeeva and Harchaoui, 2013] have considered a scan statistic that takes sparsity into account. [Cho and Fryzlewicz, 2014] propose Sparse Binary Segmentation, which also takes sparsity into account and can be viewed as a hard-thresholding of the CUSUM matrix followed by an l_1 aggregation. [Cho, 2016] proposes a double-CUSUM algorithm that performs a CUSUM transformation along the location axis on the column-wise sorted CUSUM matrix. As for the methods compatible with missingness, [Xie et al., 2013], developed a method based on the detection of abrupt change in the underlying manifold generated by the data stream. [Bardwell et al., 2016] use a likelihood like quantity to detect possible changepoints on each coordinates and then aggregate the information.

The outline of the paper is as follows: first we discuss the mathematical statement of the problem and the data and missingness generating mechanisms that we work under in **Section 2**. We then discuss how the immediate adaption of the Inspect algorithm (for the problem without missing data) does not yield satisfactory results. We therefore develop our own methodology for the single changepoint estimation in **Section 3**. Theoretical results for the homogeneous per row missingness case are developed in **Section 4**, first through **Proposition 1** showing the closeness between the optimal projection direction and the estimator. Then through **Theorem 1** which guarantees the mentioned rate of convergence. The numerical studies showing the great performance of our algorithm can be found in **Section 5**. **Section 6** discusses the extension of the algorithm to multiple changepoint detection. **Section 7** presents some ideas and results on extension to dependence. Some ideas about further work to be undertaken are discussed in **Section 8**.

We conclude this section by introducing some notations that are used throughout the paper. For a vector $u = (u_1, \ldots, u_M)^T \in \mathbb{R}^M$, a matrix $A = (A_{ij}) \in \mathbb{R}^{M \times N}$ and for $q \in [1, \infty[$, we write $||u||_q := (\sum_{i=1}^M |u_i|^q)^{1/q}$ and $||A||_q := (\sum_{i=1}^M \sum_{j=1}^N |A_{ij}|^q)^{1/q}$ for their (entry-wise) l_q-norms, as well as $||u||_\infty := \max_{i=1,\ldots,M} |u_i|$ and $||A||_\infty := \max_{i=1,\ldots,M, j=1,\ldots,N} |A_{ij}|$. We write $||A||_* := \sum_{i=1}^{\min(M,N)} \sigma_i(A)$ and $||A||_{op} := \max \sigma_i(A)$ respectively for the nuclear norm and operator norm of matrix A, where $\sigma_1(A), \ldots, \sigma_{\min(M,N)}(A)$ are its singular values. We also write $||u||_0 := \sum_{i=1}^M \mathbb{1}_{u_i \neq 0}$. For $u \in \mathbb{R}^p$, we denote by $\mathrm{diag}(u)$ the p by p matrix with u as its diagonal and zeros elsewhere. For $S \subseteq \{1, \ldots, M\}$ and $T \subseteq \{1, \ldots, N\}$, we write $u_S := (u_i : i \in S)^T$ and write $M_{S,T}$ for the $|S| \times |T|$ sub-matrix of A obtained by extracting the rows and columns with indices in S and T respectively. For two matrices $A, B \in \mathbb{R}^{M \times N}$, we denote their trace inner product as $\langle A, B \rangle = tr(A^T B)$. We also denote their Hadamard product as $A \; o \; B$. For two non-zero vectors $u, v \in \mathbb{R}^p$, we write

$$\angle(u, v) := \cos^{-1}\left(\frac{|\langle u, v \rangle|}{||u||_2 ||v||_2}\right)$$

for the acute angle bounded between them. We let $\mathbb{S}^{p-1} := \{x \in \mathbb{R}^p : ||x||_2 = 1\}$ be the unit Euclidean sphere in $\mathbb{R}^p$, and let $\mathbb{S}^{p-1}(k) := \{x \in \mathbb{S}^{p-1} : ||x||_0 \leq k\}$.

2 Problem Description

We first study the following basic independent time series model: let $X_1, \ldots, X_n$ be independent p-dimensional random vectors sampled from

$$X_t \sim N_p\left(\mu_t, \sigma^2 I_p\right) \qquad 1 \leq t \leq n \tag{1}$$

and combine the observations into a matrix $X = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$.

We first assume the existence of a single changepoint in mean at time z, $1 \leq z \leq n-1$ i.e.

$$\mu_1, \ldots, \mu_z = \mu^{(1)} \text{ and } \mu_{z+1}, \ldots, \mu_n = \mu^{(2)} \tag{2}$$

In **Section 6**, we discuss the multiple changepoints settings, where we assume that there are ν changepoints at time $1 \leq z_1 \leq \ldots \leq z_\nu \leq n-1$ such that:

$$\mu_{z_i} = \ldots = \mu_{z_{i+1}} = \mu^{(i)} \tag{3}$$

with $z_0 = 1$ and $z_{\nu+1} = n$ implicitly.

Back to the single changepoint: we write $\theta = \mu^{(2)} - \mu^{(1)}$ for the (non-zero) difference in means between the two segments. We shall later assume that the change in mean is sparse in the sense that there exist $k \in \{1, .., p\}$ (typically k is much smaller than p) such that:

$$||\theta||_0 \leq k \tag{4}$$

However, we remark that our methodology does not require knowledge of the level of sparsity and can be applied in non-sparse settings as well.

We only observe X partially through $X_\Omega \in \mathbb{R}_\Omega^{p \times n}$ (defined as the set of p by n matrices with observation matrix Ω) where

$$\left(X_\Omega\right)_{j,t} = \begin{cases} X_{j,t} \text{ if } \omega_{j,t} = 1 \\ \text{NA if } \omega_{j,t} = 0 \end{cases} \tag{5}$$

and $\Omega = (\omega_{j,t}) \in \mathbb{R}^{p \times n}$ is the observation matrix. (i.e, $\omega_{j,t} = \mathbb{1}_{O_{j,t}}$ with $O_{j,t}$ the event that the j,t entry of X is observed). The $O_{j,t}$ are independent of X but do not have to be independent of one another or even random.

Our goal is to estimate the changepoint z in the high dimensional regime, where p may be comparable with, or even larger than, the length n of the series. The signal strength of the estimation problem is determined by the magnitude of the mean change θ, the lengths of the stationary segments $z, n-z$ and the observation matrix Ω whereas the noise is related to the variance σ^2, the dimensionality p of the observed data points and the observation matrix Ω. For our theoretical results, we shall assume that the changepoint location satisfies

$$n^{-1} min\left(z, n-z\right) \geq \tau \tag{6}$$

Suppose that an estimation procedure outputs the changepoint at $0 \leq \hat{z} \leq n-1$. Our finite sample bounds will imply a rate of convergence for Inspect Miss in an asymptotic setting where the problem parameters are allowed to depend on n. Suppose that $\mathcal{P}_n$ is a class of distributions of $X \in \mathbb{R}^{p \times n}$ and Ω with sample size n. In this context, we follow the convention in the literature [Venkatraman et al., 1992] and say that the procedure is consistent for $\mathcal{P}_n$ with rate of convergence ρ_n if

$$\inf_{P \in \mathcal{P}_n} \mathbb{P}_P\left(|z - \hat{z}| \leq n\rho_n\right) \to 1 \tag{7}$$

as $n \to +\infty$.

This setting is the same as in [Wang and Samworth, 2018] except for the specification on the missingness-generating mechanisms.

3 Single changepoint estimation

In the original Inspect paper [Wang and Samworth, 2018], the procedure to compute a single changepoint is as follow: first compute the classic CUSUM of each row, take the sparse leading left singular vector of this matrix (obtained through the relaxation of an optimisation problem) which corresponds to an estimate of the vector of mean change θ, project the CUSUM matrix and take the argmax of the absolute value of the resulting one-dimensional quantity.

The method and limit theory of CUSUM statistics in the univariate case can be traced back to [Darling and Erdös, 1956]. To recall the classic CUSUM transformation in the full data case for $p \in \mathbb{N}^*$ and $n \geq 2$ $\mathcal{T} : M \in \mathbb{R}^{p \times n} \to \mathbb{R}^{p \times n-1}$:

$$[\mathcal{T}(M)]_{j,t} = \sqrt{\frac{t(n-t)}{n}} \left(\frac{\sum_{r=t+1}^{n} M_{j,r}}{n-t} - \frac{\sum_{r=1}^{t} M_{j,r}}{t} \right) \tag{8}$$

The first problem is to adapt the CUSUM transformation to this new case where there is missingness, as the usual CUSUM is now not calculable. For $p \in \mathbb{N}^*$ and $n \geq 2$ we define the Ω-CUSUM transformation $\mathcal{T}_\Omega : M \in \mathbb{R}_\Omega^{p \times n} \to \mathbb{R}^{p \times n-1}$ by:

$$[\mathcal{T}_\Omega(M)]_{j,t} = \sqrt{\frac{\left(\sum_{r=1}^{t} w_{j,r}\right)\left(\sum_{r=t+1}^{n} w_{j,r}\right)}{\sum_{r=1}^{n} w_{j,r}}} \left(\frac{\sum_{r=t+1}^{n} w_{j,r} M_{j,r}}{\sum_{r=t+1}^{n} w_{j,r}} - \frac{\sum_{r=1}^{t} w_{j,r} M_{j,r}}{\sum_{r=1}^{t} w_{j,r}} \right) \tag{9}$$

when this quantity is computable (i.e. $\sum_{r=1}^{t} w_{j,r}, \sum_{r=t+1}^{n} w_{j,r} > 0$), or 0 otherwise.

Each row of $\mathcal{T}_\Omega(M)$ is actually a spread out version of the classic one-dimensional CUSUM transformation for the data M with its missing data removed (of length $\tilde{n} = \sum_{r=1}^{n} w_{j,r}$). Each row is therefore piecewise constant as long as no new data is observed: if the t-th data is missing in row j then $(\mathcal{T}_\Omega(M))_{j,t} = (\mathcal{T}_\Omega(M))_{j,(t-1)}$

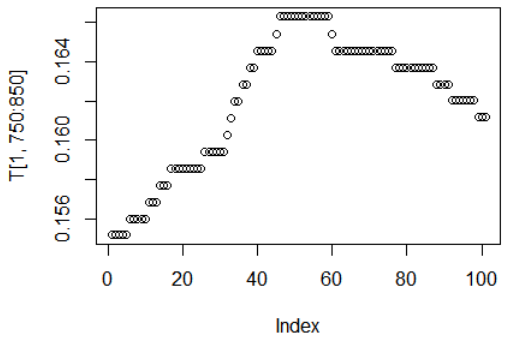


Figure 4: Spreading of the CUSUM transformation in case of missingness on a one-dimensional time series

An interesting fact about this adapted Ω-CUSUM transformation is that it takes strength across different coordinates depending on the amount of data in each row. Let us look at a simple univariate

case. With full data in the noiseless setting, the peak of the CUSUM transformation is $\theta\sqrt{\tau(1-\tau)}\sqrt{n}$. With missing data in the noiseless setting, the peak is of value $\theta\sqrt{\tilde{\tau}(1-\tilde{\tau})}\sqrt{\tilde{n}}$, where $\tilde{\tau}$ is the minimum of the percentage of data on the left and on the right of the changepoint and $\tilde{n}$ is the amount of available data. In the multivariate case, a row is therefore more significant than another if the mean change is greater and/or if there is more data in this row.

Another interesting fact is that the multivariate setting allows to compensate for some of the missingness. Let us take an example, with two covariates and no noise. Let us suppose we have data on both sides of the changepoint in both coordinates, but there is missingness right before the changepoint on the first coordinates and right after on the second coordinate. With only one coordinate, even in this noiseless case, it would be impossible to exactly find the changepoint. With no data on the five measurements before for example, there is no way of finding which one of the five timestamps it is. By using multiple coordinates, one can compensate those badly placed missing data.

Let us now decompose the Ω-CUSUM transformation in noise and signal components. We write:

$$X_\Omega = \mu_\Omega + W_\Omega \tag{10}$$

where $(\mu_\Omega)_{j,t} = \mu_{j,t}$ if $w_{j,t} = 1$ and NA otherwise, and W_Ω is a $p \times n$ random matrix with observation matrix Ω and independent $N\left(\mu_{j,t}, \sigma^2\right)$ entries. Let $T_\Omega := \mathcal{T}_\Omega\left(X_\Omega\right), A_\Omega := \mathcal{T}_\Omega\left(\mu_\Omega\right), E_\Omega := \mathcal{T}_\Omega\left(W_\Omega\right)$. By the linearity of the Ω-CUSUM transformation:

$$T_\Omega = A_\Omega + E_\Omega \tag{11}$$

In this case, the entries of the matrix A_Ω can be computed explicitly:

$$(A_\Omega)_{j,t} = \begin{cases} \sqrt{\frac{\left(\sum_{r=1}^{t} w_{j,r}\right)\left(\sum_{r=t+1}^{n} w_{j,r}\right)}{\sum_{r=1}^{n} w_{j,r}}}\frac{\sum_{r=z+1}^{n} w_{j,r}}{\sum_{r=t+1}^{n} w_{j,r}}\theta_j & \text{if } t \leq z,\ \sum_{r=1}^{t} w_{j,r}, \sum_{r=t+1}^{n} w_{j,r} > 0 \\ \sqrt{\frac{\left(\sum_{r=1}^{t} w_{j,r}\right)\left(\sum_{r=t+1}^{n} w_{j,r}\right)}{\sum_{r=1}^{n} w_{j,r}}}\frac{\sum_{r=1}^{z} w_{j,r}}{\sum_{r=1}^{t} w_{j,r}}\theta_j, & \text{if } t > z,\ \sum_{r=1}^{t} w_{j,r}, \sum_{r=t+1}^{n} w_{j,r} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

In the single dimension case, we would take the argmax of the Ω-CUSUM transformation T_Ω (which would be a vector) to get the best estimate of z. This is because A_Ω is maximised at z (even though not strictly). In the multi-dimensional case, each row of A_Ω is maximised at z, (still not strictly if there is missingness to the right of the changepoint). Therefore for any projection direction v, $v^T A_\Omega$ is maximised in absolute value at z.

However, to improve the signal-to-noise-ratio, we would prefer to select a better projection direction. In the full data case, the best projection direction is θ, which is also the leading left singular vector of A. As the original Inspect algorithm can be seen as a method taking the CUSUM transformation as input, we could actually use the end of the Inspect algorithm with the new adapted Ω-CUSUM. Let us take a look at the optimisation scheme used in [Wang and Samworth, 2018]. The k-sparse leading left singular vector of A, which is θ could be approximated by $\arg\max_{\tilde{v}\in\mathbb{S}^{p-1}(k)} ||T^T\tilde{v}||_2$. However, one can rewrite this as:

$$\max_{u\in\mathbb{S}^{p-1}(k)} ||u^T T||_2 = \max_{M\in\mathcal{M}} \langle M, T\rangle \tag{13}$$

where $\mathcal{M} = \{M \in \mathbb{R}^{p\times n} : ||M||_* = 1, \text{rank}(M) = 1, M \text{ has at most k non-zero rows}\}$. We then consider a convex relaxation of the problem with an l_1 penalty. This yields (for tuning parameter $\lambda \geq 0$):

$$\hat{M} \in \arg\max_{M\in\mathcal{S}_i} \{\langle T, M\rangle - \lambda||M||_1\} \tag{14}$$

where $S_1 = \{M \in \mathbb{R}^{p\times n} : ||M||_* \leq 1\}$ and $S_2 = \{M \in \mathbb{R}^{p\times n} : ||M||_2 \leq 1\}$. The S_2 version being an approximation of S_1 but with an easily computable closed-form solution.

The problem here is that our new A_Ω is not of rank 1 as A was, making the S_2 convex relaxation too coarse. The theoretical guarantees do not apply in that case. For the S_1 relaxation, the theoretical guarantees hold. The only problem being that the optimisation scheme used in [Wang and Samworth, 2018] is too slow as n grows. Therefore finding the left leading singular vector of A_Ω is not possible using the methods of the Inspect paper. Moreover, θ is no longer the left leading singular vector of A_Ω and is not necessarily the best projection direction anymore, as each row is impacted by its own missingness.

One possibility for the projection direction is to still use an estimate of u, the leading left singular vector of A_Ω. This is not a multiple of θ anymore, but this vector has the same sparsity pattern as θ, meaning that we would only use coordinates with change. However, we have seen that we cannot use the optimisation scheme of the Inspect paper. We therefore need to create a new relaxation of the problem of finding the sparse leading left singular vector.

Let us begin again: in the ideal case where k is known, we could in principle let $\hat{v}_{max,k}$ be a k-sparse leading left singular vector of T_Ω, defined by:

$$\hat{v}_{max,k} = \arg \max_{\tilde{v}\in\mathbb{S}^{p-1}(k)} ||T_\Omega^T \tilde{v}||_2 \tag{15}$$

However, the optimisation problem (15) is non-convex and hard to solve. In fact, computing the k-sparse leading left singular vector of a matrix is known to be 'NP hard' ([Tillmann and Pfetsch, 2014]). The naive algorithm that scans through all possible k-subsets of the rows of T_Ω has running time exponential in k, which quickly becomes impractical to run for even moderate sizes of k. A natural approach to remedy this computational issue is to work with a relaxation of the optimisation problem (15) instead. In fact, we can write

$$\max_{u\in\mathbb{S}^{p-1}(k)} ||u^T T_\Omega||_2 = \max_{u\in\mathbb{S}^{p-1}(k), w\in\mathbb{S}^{n-2}} u^T T_\Omega w = \max_{u\in\mathbb{S}^{p-1}, w\in\mathbb{S}^{n-2}, ||u||_0\leq k} \langle uw^T, T_\Omega\rangle \tag{16}$$

As the sparsity constraint on u is not convex, we relax it to an entrywise l_1-norm penalty. The optimisation problem of finding:

$$(u_{max}, w_{max}) \in \arg \max_{u\in\mathbb{S}^{p-1}, w\in\mathbb{S}^{n-2}} \{\langle T_\Omega, uw^T\rangle - \lambda\sqrt{n}||u||_1\} \tag{17}$$

where $\lambda > 0$ is a tuning parameter to be chosen later, is therefore a relaxation of problem (15). This non-convex problem may be solved using alternating optimisation of u and w through Lagrangian multipliers.

Using **Lemma 1**, this yields that for u_{max} and w_{max} the maximisers of problem (17), we have:

$$\begin{aligned} w_{max} &= \frac{T_\Omega^T u_{max}}{||T_\Omega^T u_{max}||_2} \\ u_{max} &= T_\Omega w_{max} \; o\max\left(0, 1 - \frac{\lambda\sqrt{n}}{|T_\Omega w_{max}|}\right) / ||T_\Omega w_{max} \; o\max\left(0, 1 - \frac{\lambda\sqrt{n}}{|T_\Omega w_{max}|}\right)||_2 \end{aligned} \tag{18}$$

We therefore alternate between the optimisation of u and w until convergence.

Algorithm 1: Optimisation scheme

Result: $\hat{u}, \hat{w}$ estimates of $u_{max}, w_{max} \in \arg\max_{u \in \mathbb{S}^{p-1}, w \in \mathbb{S}^{n-2}} \{< T_\Omega, uw^T > -\lambda\sqrt{n}||u||_1\}$
Input: $\lambda > 0, T_\Omega \in \mathbb{R}^{p \times (n-1)}$;
$u \leftarrow$ leading left singular vector of T_Ω;
while *u has not converged* **do**
  $w \leftarrow T_\Omega^T u / ||T_\Omega^T u||_2$;
  $u \leftarrow T_\Omega w \; o\max\left(0, 1 - \frac{\lambda\sqrt{n}}{|T_\Omega w|}\right) / ||T_\Omega w \; o\max\left(0, 1 - \frac{\lambda\sqrt{n}}{|T_\Omega w|}\right)||_2$;
end
Output: $\hat{u}, \hat{w}$

However theoretical guarantees for this kind of alternating method are not readily available. Nevertheless, numerical convergence is reached rapidly even for significant values of n.

The full algorithm consists in obtaining the projection direction through **Algorithm 1**, projecting the adapted CUSUM transformation on it and then taking the argmax of the absolute value of the resulting vector. This is presented in the pseudocode of the algorithm below:

Algorithm 2: Inspect Miss

Result: $\hat{z} \in [1, n-1]$
Input: $\lambda > 0, X_\Omega \in \mathbb{R}_\Omega^{p \times n}$;
$T_\Omega \leftarrow \mathcal{T}_\Omega(X_\Omega) \in \mathbb{R}^{p \times (n-1)}$;
$\hat{u} \leftarrow$ **Algorithm1** (λ, T_Ω);
$\hat{z} \leftarrow \arg\max\left(|\hat{u}^T T_\Omega|\right)$;
Output: $\hat{z}, |\hat{u}^T T_\Omega|_{\hat{z}}$

This closes the section on the presentation of the algorithm. Theoretical results are presented in **Section 5** in the homogeneous per row setting. This does not mean that the procedure should only be used in this specific setting. As we have seen in this section, there are more empirical ideas that justify the good behaviour of the algorithm in other missingness-generating mechanisms.

Choosing an estimate of the left-leading singular vectors of A_Ω allows to project on the direction that provides the best rank 1 approximation of the CUSUM transformation matrix (in terms of l_2-norm). In the noiseless case, the changepoint is exactly found if there is at least observed data right before and right after the changepoint in some coordinates used by the projection. It does not even have to be the same coordinate! If this condition is not verified, it is easy to see that the problem is intractable.

We will see in the following section that the theoretical results hold for $\lambda = 2\sigma\sqrt{\log\{p \log n\}}$. In practice, we recommend using $\lambda = \sigma\sqrt{\log\{p \log n\}/2}$ as the other proposition is a little too restrictive. However, this still requires the knowledge of σ. If σ is unknown, we can estimate it robustly using the median absolute deviation [Hampel, 1974] of the marginal one dimensional time series for each row and then normalise them and use $\sigma = 1$.

The projected adapted Ω-CUSUM transformation of the mean matrix still presents the same 'mountain' shape, peaking at the changepoint. The projected noise is still a one-dimensional vector with marginal $N(0, \sigma^2)$ distribution. If the signal-to-noise ratio is high enough, this should suffice to obtain a good estimate of the changepoint location, no matter the shape of the observation matrix Ω.

4 Homogeneous per row setting

Although our algorithm is usable with all observation matrices Ω (as long as the missingness is independent of X, even the entries of Ω do not have to be independent), we will provide theoretical properties in the case of homogeneous missingness per row, i.e. $\mathbb{P}(O_{j,t}) = p_j,\ \forall j \in [\![1,p]\!]$, with the $O_{j,t}$ being independent from one another. The probabilities of presence per row are denoted as $\mathbf{p} = (p_1, \ldots, p_p)$. This setting makes sense in some real life applications. For example, the case of different measuring devices for different coordinates, each device having its own probability of failing to correctly function each and every day.

Let us recall that in the full data case, we have for the CUSUM transformation of the mean matrix:

$$A_{j,t} = \begin{cases} \sqrt{\frac{t}{n(n-t)}}\,(n-z)\,\theta_j \ for\ t \leq z \\ \sqrt{\frac{(n-t)}{nt}}\,z\theta_j \ for\ t > z \end{cases} \tag{19}$$

Therefore we can write:

$$A = \theta\gamma^T \tag{20}$$

where $\gamma := \frac{1}{\sqrt{n}}\left(\sqrt{\frac{1}{n-1}}\,(n-z), \ldots, \sqrt{z(n-z)}, \ldots, \sqrt{\frac{1}{n-1}}z\right)$.

In the homogeneous per row case that we described earlier, we have $\mathbb{E}\left(\sum_{r=1}^{t} w_{j,r}\right) = p_j t$ for any $1 \leq t \leq n$.

Hence, $\mathrm{diag}\left(\sqrt{\mathbf{p}}\right) A = \left(\theta\ o\ \sqrt{\mathbf{p}}\right)\gamma^T$ can be expected to be close to the actual A_Ω. Its (unique strictly positive) singular value is: $||\theta\ o\ \sqrt{\mathbf{p}}||_2||\gamma||_2$, with associated leading left singular vector: $\theta\ o\ \sqrt{\mathbf{p}}$ (normalized) and associated leading right singular vector γ (normalized). Let us remark that once again, rows with higher probability of presence of data, all other things being equal, have higher peaks, and are therefore more important in the 'eyes' of the algorithm. This is what we want as well, as we believe rows with more data are more trustworthy.

To further describe our data in this setting, let us suppose we have the following property:

$$||\theta_{\mathbf{p}}||_2 \geq \vartheta_{\mathbf{p}} \text{ where } \theta_{\mathbf{p}} = \theta\ o\ \sqrt{\mathbf{p}} \tag{21}$$

To describe the theoretical properties of **Algorithm 2**, we introduce the following class of distributions: let $\mathcal{P}\left(n, p, k, 1, \vartheta_{\mathbf{p}}, \tau, \sigma^2, \mathbf{p}\right)$ denote the class of distribution of $X \in \mathbb{R}^{p\times n}$ drawn as in (1), the observation matrix Ω drawn in an independent homogeneous per row setting with probabilities $\mathbf{p}$, the changepoint location verifies (2) and (6) and the vector of mean change verifies (4) and (21). For ease of redaction, let us supposed that the coordinates of change are the first k ones. This does not change the results in any manner, and we remind the reader that k is generally unknown and that the coordinates with change are not known either. However, when we now write some conditions on $p_1, \ldots, p_k$, it signifies conditions on the probabilities of presence of the rows with mean change.

Our first theoretical result presents the closeness in angle between the estimated projection direction and the true leading left singular vector of A_Ω (for which we have no closed-form solution). We are still here only considering the single changepoint setting (as shows the 1 in the considered family of distribution).

Proposition 1. *Suppose that* $(\hat{u}, \hat{w})$ *verifies* $(\hat{u}, \hat{w}) \in \arg\max_{u \in \mathbb{S}^{p-1}, w \in \mathbb{S}^{n-1}} \{\langle T_\Omega, uw^T \rangle - \lambda\sqrt{n}||u||_1\}$. *Then for* $\lambda \geq 2\sigma\sqrt{\log\{p \log n\}}$, $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/2\}}{\log kn}} \leq 1$ *and* $\forall\ j \in [\![1, k]\!]$, $\frac{24 \log\{k \log n\}}{p_j \log n} \leq 1$, $3 \leq k \log n \leq k\tau n$, $n \geq 6$ *:*

$$\sup_{P \in \mathcal{P}(n,p,k,1,\vartheta_{\mathbf{p}},\tau,\sigma^2,\mathbf{p})} \mathbb{P}_P\left(\sin \angle(u, \hat{u}) > \frac{32\lambda\sqrt{k}}{\tau\vartheta_{\mathbf{p}}\sqrt{n}}\right) \leq \frac{4}{\{p \log n\}^{1/2}} + \frac{4}{k \log n - k} \tag{22}$$

The following corollary restates the rate of convergence of the projection estimator in a simple asymptotic regime.

Corollary 1. *Consider an asymptotic regime where* $\log p = O\{\log n\}$, σ *and* $\mathbf{p}$ *are constants,* $\vartheta_p \asymp n^{-a}$, $\tau \asymp n^{-b}$ *and* $k \asymp n^c$ *for some* $a \in \mathbb{R}, b \in [0, 1]$ *and* $c \geq 0$. *Then, setting* $\lambda = 2\sigma\sqrt{\log\{p \log n\}}$ *and provided that* $a + b + c/2 < 1/2$, *we have for every* $\delta > 0$ *that*

$$\sup_{P \in \mathcal{P}(n,p,k,1,\vartheta_{\mathbf{p}},\tau,\sigma^2,\mathbf{p})} \mathbb{P}_P\left(\angle(u, \hat{u}) > n^{-(1-2a-2b-c)/2+\delta}\right) \to 0 \tag{23}$$

This is an equivalent of **Proposition 1** of the original Inspect paper [Wang and Samworth, 2018]. The main differences are that u is not known in closed form anymore as it depends on θ and on the observation matrix Ω, and that there are more conditions to apply the proposition. Some of those are very easily verified, therefore not limiting the use of said proposition (such as $3 \leq k \log n$). There is however a new condition on the probabilities of observation for the rows with mean change. Those should not be too small. The bound inside the probability is the same, with the difference that $\vartheta_{\mathbf{p}}$ now depends on the probabilities per row, as missing data lessens the strength of the signal. The rate outside the probability is of the same order in n.

The algorithm then consists in projecting the adapted Ω-CUSUM transformation onto the vector returned by **Algorithm 1**. From a theoretical point of view, the fact that $\hat{u}$ is estimated using the entire available data set X_Ω makes it difficult to analyse the post-projection noise structure. For this reason, in the analysis below, we work with a slight variant of **Algorithm 2**. We assume for convenience that $n = 2n_1$ is even, and define $X_{\Omega^{(1)}} \in \mathbb{R}^{p \times n_1}_{\Omega^{(1)}}$, $X_{\Omega^{(2)}} \in \mathbb{R}^{p \times n_1}_{\Omega^{(2)}}$, by

$$(X_{\Omega^{(1)}})_{j,t} := (X_\Omega)_{j,2t-1} \text{ and } (X_{\Omega^{(2)}})_{j,t} := (X_\Omega)_{j,2t} \qquad \forall\ 1 \leq j \leq p,\ 1 \leq t \leq n_1 \tag{24}$$

where $\Omega^{(1)}$ and $\Omega^{(2)}$ are the uneven and even columns of Ω respectively.

Algorithm 3: Inspect Miss (sample-splitting variant, for homogeneous per row setting)

Result: $\hat{z} \in [1, n-1]$
Input: $\lambda > 0, X_\Omega \in \mathbb{R}^{p \times n}_\Omega$;
$T_{\Omega^{(1)}} \leftarrow \mathcal{T}_{\Omega^{(1)}}(X_{\Omega^{(1)}}) \in \mathbb{R}^{p \times (n_1-1)}_{\Omega^{(1)}}$;
$T_{\Omega^{(2)}} \leftarrow \mathcal{T}_{\Omega^{(2)}}(X_{\Omega^{(2)}}) \in \mathbb{R}^{p \times (n_1-1)}_{\Omega^{(2)}}$;
$\hat{u} \leftarrow$ **Algorithm1** $(\lambda, T_{\Omega^{(1)}})$;
$\hat{z} \leftarrow 2 \arg\max\left(\hat{u}^T T_{\Omega^{(2)}}\right)$;
Output: $\hat{z}$

We then use $X_{\Omega^{(1)}}$ to estimate the oracle projection direction and use $X_{\Omega^{(2)}}$ to estimate the change-point location after projection (see **Algorithm 3**). However, we recommend using **Algorithm 2** in practice to exploit the full signal strength in the data, especially in missingness-generating mechanism that are not homogeneous per row, where crucial data could be missed! (Imagine the case of data with two coordinates, where all of the even-numbered data is missing in the first coordinate, and all the

uneven for the second coordinate). We summarise the overall estimation performance of **Algorithm 3** in the following theorem:

Theorem 1. *Suppose that $\sigma > 0$ is known. Let $\hat{z}$ be the output **Algorithm 3** with input $X_\Omega \sim P \in \mathcal{P}(n, p, k, 1, \vartheta_{\mathbf{p}}, \tau, \sigma^2, \mathbf{p})$ and $\lambda = 2\sigma\sqrt{\log\{p \log n\}}$.*

If $\frac{128\sigma}{\vartheta_{\mathbf{p}}\tau}\sqrt{\frac{k\log\{p\log n/2\}}{n/2}} \leq 1$, $216\sqrt{\frac{\log kn/2}{n\tau/2}} + \frac{12\sqrt{6}\sigma}{\vartheta_{\mathbf{p}}}\sqrt{\frac{\log n/2}{n\tau/2}} \leq 1$, $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/4\}}{\log kn/2}} \leq 1$ and $\forall\ j \in [\![1, k]\!]$, $\frac{24\log\{k\log n/2\}}{p_j \log n/2} \leq 1$, $3 \leq k\log n/2 \leq k\tau n/2$, n is even and $n \geq 12$, z is even, then:

$$\mathbb{P}\left(\frac{|z-\hat{z}|}{n} \leq 108\sqrt{2}\tau^{1/2}\sqrt{\frac{\log kn}{n}} + \frac{12\sqrt{3}\sigma\tau^{1/2}}{\vartheta_{\mathbf{p}}}\sqrt{\frac{\log n}{n}}\right) \geq 1 - \frac{4}{\{p\log n/2\}^{1/2}} - \frac{1}{\{\log n/2\}^{1/2}} - \frac{8}{k\log n/2 - k} \tag{25}$$

Corollary 2. *Suppose that σ and $\mathbf{p}$ are constants, $\log p = O\{\log n\}$, $\vartheta_p \asymp n^{-a}$, $\tau \asymp n^{-b}$ and $k \asymp n^c$ for some $a \in \mathbb{R}$ and $b \in [0, 1]$ and $c \geq 0$. If $a + b + c/2 < 1/2$, then the output $\hat{z}$ of **Algorithm 3** with $\lambda := 2\sigma\sqrt{\log\{p\log n\}}$ is a consistent estimator of the true changepoint z with rate of convergence $\rho_n = o\left(n^{-1/2-b/2+a+\delta}\right)$ for any $\delta > 0$.*

We remark here that compared to the original Inspect paper [Wang and Samworth, 2018], we have a lesser rate of convergence, as we have replaced some $\log\{\log n\}/n$ by $\sqrt{\{\log n\}/n}$. We believe this could be improved through further work on the end of the proof of **Theorem 1**, as dependency of the projected noise vector has not yet been exploited.

5 Numerical studies

For our numerical studies section, we are interested in the performance of our method in simulated settings as well as on a real-life dataset example.

miss	proba	oncp	$\angle(u, \hat{u}_{naive})$	$\angle(u, \hat{u})$	$\angle(u, \hat{u}_{S_1})$	$\angle(u, \hat{u}_{S_2})$
0	0.2	0	14.89	9.74	9.16	9.74
0	0.4	0	10.31	6.30	5.72	6.30
0	0.6	0	8.59	5.15	5.15	5.15
0	0.8	0	7.44	4.58	4.58	4.58
1	0.2	-1	16.61	13.17	12.03	12.03
1	0.2	0	17.76	13.75	12.60	12.60
1	0.2	1	16.61	13.17	12.03	12.03
1	0.4	-1	11.45	8.02	7.44	8.02
1	0.4	0	12.60	8.02	7.44	8.02
1	0.4	1	11.45	8.02	7.44	7.44
1	0.6	-1	9.74	6.30	6.30	6.30
1	0.6	0	9.74	6.30	5.72	6.30
1	0.6	1	9.74	6.30	6.30	6.30
1	0.8	-1	8.59	5.15	5.15	5.15
1	0.8	0	8.594	5.15	5.15	5.15
1	0.8	1	8.59	5.15	5.15	5.15

Table 1: Comparison between different estimates of the projection direction

We first decide to look into the comparison between the angle between the true leading left singular vector of μ_Ω: u (that we can obtain as we consider data that we generated ourselves) and different estimates. We therefore compare an estimate obtained using our optimisation scheme: $\hat{u}$, an estimate obtained using the S_1 approximation: $\hat{u}_{S_1}$ and another using the S_2 approximation: $\hat{u}_{S_2}$ as in the original Inspect paper [Wang and Samworth, 2018], and lastly simply the leading left singular vector of A_Ω: $\hat{u}_{naive}$ (obtained through the power method). The angle is given in degrees and averaged over 100 iterations. The angle is considered in different settings of missingness (hence the values of miss, proba and oncp). Other parameters for the function are n = 500, p = 10, k = 3, z = 200, vartheta = 1.5, sigma = 1, shape = 1, noise = 0, corr = 0. See the package documentation of **single.change** in **Appendix 2**. The results are compiled in **Table 1**.

We observe that the best estimate is always the S_1 one, but that we do not lose to much by using S_2 or our method. As the S_1 estimate soon becomes impossible to compute for larger values of n and p, we only now compare the other three estimates. The angle is given in degrees and average over 100 repetitions. The other parameters are: n = 500, p = 1000, k = 30, z = 200, vartheta = 1.5, sigma = 1, shape = 1, noise = 0, corr = 0. The results are compiled into **Table 2**.

We observe that our estimate is always better than the one given by S_2 which itself is better than the naive estimate. This shows how our projection direction is close to the optimal solution of the leading left singular vector of μ_Ω in this sparse (k =30, p =1000) high-dimensional setting.

We now look into the influence of the choice of λ both on the angle and on the mean squared error of our method in the single changepoint estimation. The angle is given in degrees and both the error and angle are averaged over 100 repetitions. Other parameters are n = 1000, p = 500, k = 10, z = 400, vartheta = 1, sigma = 1, shape = 1, noise = 0, corr = 0. The Missingness on the figure is the sum of the arguments miss and proba. Missingness 2.2 signifies miss = 2, proba = 0.2 . The results are displayed in **Figure 5**.

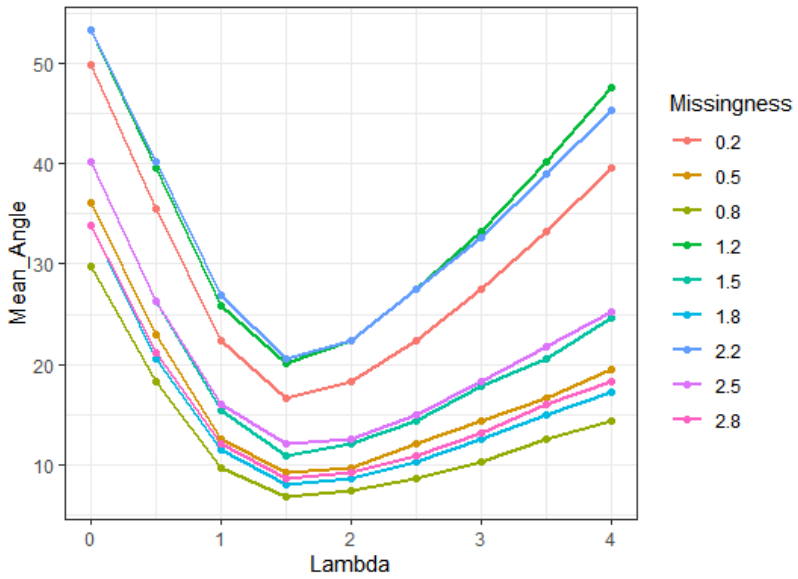
We observe that the choice of λ has a lot of influence on the mean angle but not so much on the mean squared error. It is interesting to notice that the best λ is the same for all missingness settings presented, if all the other parameters are equal. In practice we therefore recommend to use $\sqrt{\log\{p \log n\}/2}$.

Technically, Inspect Miss can be used on data with no missingness. The CUSUM transformation will therefore be the same as in the Inspect method, but the projection directions based on this CUSUM matrix are different. We therefore look at how the results of both methods compare in different settings in terms of n, p, k, z. Other parameters of **single.change** are sigma = 1, shape = 0, noise = 0, corr = 0, miss = -1. The results are compiled in **Table 3**.

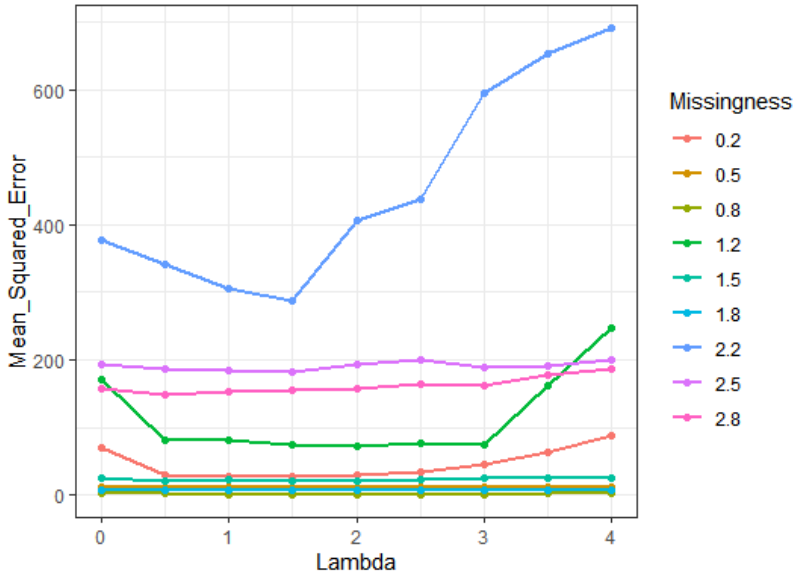
We notice that on average, Inspect performs better than Inspect Miss on the full data case, except on very sparse setting. This is due to the different l_1 penalty in the optimisation schemes. In Inspect, the penalty applies to the whole matrix, whereas in Inspect Miss, it focuses on the leading left-singular vector of the matrix. If this behaviour is not desirable, nothing stops us from actually changing this in the optimisation scheme of Inspect Miss. It is not complicated algorithmically and does not invalidate the theory.

miss	proba	oncp	$\angle(u, \hat{u}_{naive})$	$\angle(u, \hat{u})$	$\angle(u, \hat{u}_{S_2})$
0	0.2	0	41.82	16.04	18.33
0	0.4	0	32.08	10.31	12.03
0	0.6	0	27.50	8.02	9.74
0	0.8	0	24.06	6.87	8.02
1	0.2	-1	45.83	18.33	21.77
1	0.2	0	46.40	18.90	21.77
1	0.2	1	46.40	18.90	21.77
1	0.4	-1	36.09	12.03	14.32
1	0.4	0	36.09	12.03	14.32
1	0.4	1	36.09	12.03	14.32
1	0.6	-1	30.36	9.16	11.45
1	0.6	0	30.93	9.16	11.45
1	0.6	1	30.93	9.74	11.45
1	0.8	-1	27.50	8.02	9.74
1	0.8	0	26.92	8.02	9.74
1	0.8	1	27.50	8.02	9.74
2	0.2	-1	14.32	5.72	6.30
2	0.2	0	45.83	18.90	21.77
2	0.2	1	70.47	52.71	59.58
2	0.4	-1	17.18	5.72	6.87
2	0.4	0	36.09	12.60	14.89
2	0.4	1	63.02	35.52	43.54
2	0.6	-1	18.33	5.72	6.87
2	0.6	0	30.36	9.74	11.45
2	0.6	1	58.44	28.07	35.52
2	0.8	-1	18.90	5.72	6.87
2	0.8	0	26.92	8.02	9.74
2	0.8	1	54.43	24.06	30.36
3	0.2	-1	32.085	16.04	17.18
3	0.2	0	69.32	55.57	57.86
3	0.2	1	81.93	83.07	81.93
3	0.4	-1	29.22	12.03	13.17
3	0.4	0	53.28	26.35	29.79
3	0.4	1	74.48	65.31	69.90
3	0.6	-1	26.35	9.16	10.88
3	0.6	0	41.25	16.04	18.33
3	0.6	1	67.60	45.83	53.28
3	0.8	-1	24.06	8.02	9.16
3	0.8	0	33.23	10.88	13.17
3	0.8	1	61.30	33.23	40.10

Table 2: Comparison between different estimates of the projection direction for bigger p



(a) Impact of λ on the angle between u and $\hat{u}$ for different missingness settings



(b) Impact of λ on the mean squared error for different missingness settings

Figure 5: Dependence of projection direction and mean squared error on λ

n	p	k	z	vartheta	rmse inspect miss	rmse inspect
500	500	3	200	0.8/sqrt(k)	17.61	17.61
500	500	50	200	0.8/sqrt(k)	77.19	46.53
500	500	500	200	0.8/sqrt(k)	101.27	55.29
500	1000	3	200	0.8/sqrt(k)	21.35	16.45
500	1000	100	200	0.8/sqrt(k)	104.39	60.91
500	1000	1000	200	0.8/sqrt(k)	97.65	66.39
500	2000	3	200	0.8/sqrt(k)	10.8	14.87
500	2000	200	200	0.8/sqrt(k)	105.03	61.02
500	2000	2000	200	0.8/sqrt(k)	111.19	62.41
1000	500	3	400	0.8/sqrt(k)	8.42	12.94
1000	500	50	400	0.8/sqrt(k)	133.18	57.14
1000	500	500	400	0.8/sqrt(k)	197.64	110.5
1000	1000	3	400	0.8/sqrt(k)	9.97	10.13
1000	1000	100	400	0.8/sqrt(k)	163.55	90.6
1000	1000	1000	400	0.8/sqrt(k)	202.82	113.71
1000	2000	3	400	0.8/sqrt(k)	11.7	12.13

Table 3: Comparison between Inspect and Inspect Miss in full data setting

We now consider a real-life dataset example from [De Vito et al., 2008]. The AirQuality dataset contains 15 different types of measures on the quality of air in an Italian city from 2004 to 2005 (one year of data with a frequency of one measure per hour, therefore n=9357). The dataset contains missingness, especially for the 3rd, 6th and 8th measures, reaching approximately 16000 NAs. We therefore use our method on this dataset and obtain the following result (blue lines represent the detected changepoints, white squares the missing data):

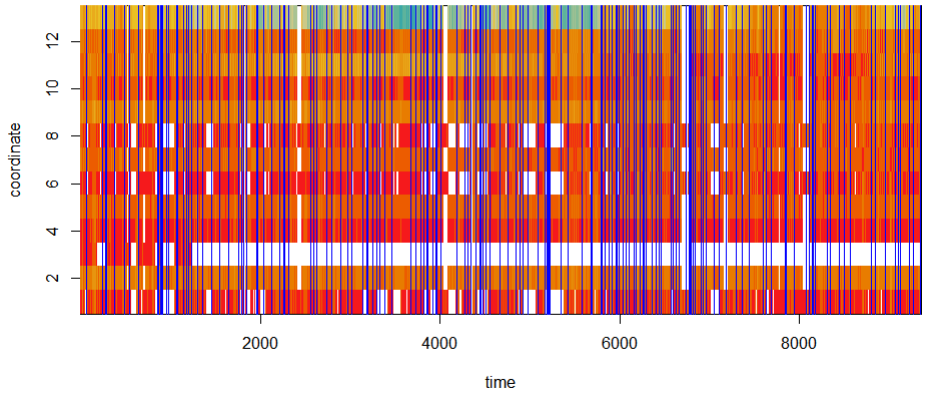


Figure 6: Inspect Miss on AirQuality dataset

We detect a high number of changepoints, not concentrated in a particular area. The frequency of apparition of the changepoints is high, meaning that the rotation of week days and weekend or daytime and nighttime could explain the difference in air quality in the city, yielding the detection of a changepoint.

6 Multiple changepoints estimation

In the preceding sections, we have discussed the issue of single changepoint estimation. However, the data generating mechanism can include multiple changepoints, with the question now being, how many changepoints are there and where?

We now assume the following data-generating mechanism: $X_t \sim N(\mu_t, \sigma^2 I_p)$, with μ being divided into $\nu + 1$ segments where ν is the number of changepoints, as in (3). The missingness-generating mechanism is the same as in the single changepoint case, i.e. the observation matrix Ω can be generated in any-way as long as it does not depend on X.

In this section, we do not present theoretical results, but **Algorithm 4** (below), which is the combination of the single changepoint estimation algorithm (**Algorithm 2**) and the wild binary segmentation scheme of [Fryzlewicz, 2014].

Algorithm 4: Multiple changepoints Inspect Miss

Result: $\hat{\nu} \in [\![1, n-1]\!], \hat{z}_1 \leq \ldots \leq \hat{z}_{\hat{\nu}} \in [\![1, n-1]\!]$
Input: $\lambda > 0, \xi > 0, \beta > 0, Q \in \mathbb{N}, X_\Omega \in \mathbb{R}_\Omega^{p \times n}$;
$\hat{Z} \leftarrow \emptyset$;
draw Q pairs or integers $(s_1, e_1), \ldots, (s_Q, e_Q)$ uniformly at random from the set $\{(l, r) \in \mathbb{Z}^2 : 0 \leq l \leq r \leq n\}$;
run $wbs(0, n)$ where wbs is defined below;
$\hat{\nu} \leftarrow |\hat{Z}|$ and sort the elements of $\hat{Z}$ in increasing order to obtain $\hat{z}_1 \leq \ldots \leq \hat{z}_{\hat{\nu}}$;
Definition of $wbs(s, e)$;
Set $Q_{s,e} \leftarrow \{q : s + n\beta \leq s_q < e_q \leq e - n\beta\}$;
for $q \in Q_{s,e}$ **do**
  run **Algorithm 2** with $X_\Omega^{[q]}$, λ as input, and let $\hat{z}^{[q]}$, $(\bar{T}_\Omega^{[q]})_{\max}$ be the output ;
end
Find $q0 \in \arg\max_{q \in Q_{(s,e)}} (\bar{T}_\Omega^{[q]})_{\max}$ and set $b \leftarrow s_{q_0} + \hat{z}^{[q_0]}$;
if $(\bar{T}_\Omega^{[q_0]})_{\max} \geq \xi$ **then**
  $\hat{Z} \leftarrow \hat{Z} \cup \{b\}$
end
$wbs(s, b)$;
$wbs(b, e)$;

This modification was already the one used in [Wang and Samworth, 2018] to upgrade from the single changepoint detection presented earlier to the multiple changepoints detection setting we are now interested in. The modifications are the following: the single changepoint detection method now used is the one presented in this work, compatible with missingness and the threshold ξ used (if it is not provided by the user) is computed by applying the single changepoint detection algorithm on data with no changepoint and with the same observation matrix Ω.

Here below is an example of the multiple changepoint detection algorithm in action on data with three changepoints and a high signal-to-noise ratio ($n = 500, p = 100$). The blue lines represent the position of the detected changepoints, while the white squares represent missing data. We can see the great performance of the algorithm in this particular setting.

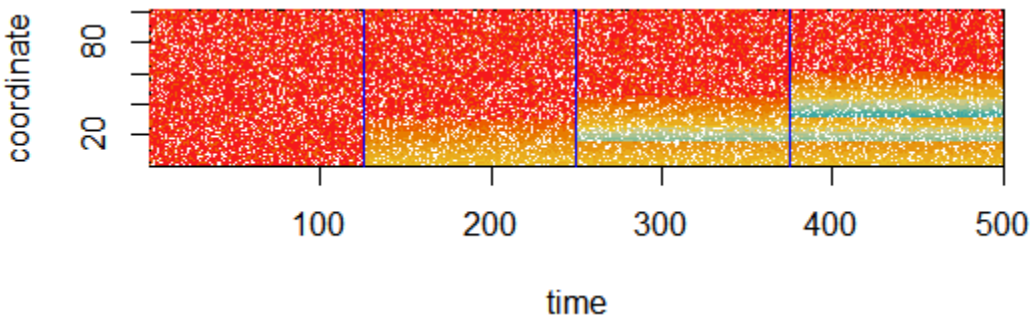


Figure 7: Representation of multiple changepoints detection with Inspect Miss

7 Extension: temporal or spatial dependence

Here we discuss how our algorithm adapts to more realistic settings where our stream of data has temporal or spatial dependence. We assume that we are in the single changepoint setting with the same mean structure as presented earlier and in the homogeneous per row setting for the missingness.

7.1 Temporal dependence

A first relaxation of the assumption of independence on the noise of our data is to consider that the noise vectors $W_1, \dots, W_n$ are stationary. Defining $K(u) := cov(W_t, W_{t+u})$, we assume here that $W = (W_1, \dots, W_n)$ forms a centred, stationary Gaussian process with covariance function K. As we are not concerned with spatial dependence in this subsection, we consider that each coordinate evolves independently from the others, yielding that $K(u)$ is a diagonal matrix for every u. We now write $\sigma^2 := ||K(0)||_{op}$, and further assume that the dependence is short ranged, meaning that

$$||\sum_{u=0}^{n-1} K(u)||_{op} \leq B\sigma^2 \tag{26}$$

for some universal constant B. This was also the setting considered in [Wang and Samworth, 2018]. Our method Inspect Miss does not require any change in this setting, and the oracle projection direction is still u, the left-leading singular vector of A_Ω. As for theoretical performance in this setting, the following theorem holds:

Theorem 2. *Suppose that $\sigma, B \geq 0$ are known. Let $\hat{z}$ be the output of **Algorithm 3** with input X_Ω and $\lambda := \sigma\sqrt{8B\log\{np\}}$. If $216\sqrt{\frac{\log kn/2}{n\tau/2}} + \frac{12\sqrt{6}\sigma}{\vartheta_{\mathbf{p}}}\sqrt{\frac{\log n/2}{n\tau/2}} \leq 1$, $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/4\}}{\log kn/2}} \leq 1$, $\forall\ j \in [\![1,k]\!]$, $\frac{24\log\{k\log n/2\}}{p_j \log n/2} \leq 1$ and $3 \leq k\log n/2 \leq k\tau n/2$, $n \geq 12$ and n is even and z is even. Then there is a universal constant C such that if:*

$$\frac{C\sigma\sqrt{kB\log\{pn/2\}}}{\tau\vartheta_{\mathbf{p}}\sqrt{n/2}} \leq 1 \tag{27}$$

then

$$\mathbb{P}\left(\frac{1}{n}|z-\hat{z}| \leq 108\sqrt{2}\tau^{1/2}\sqrt{\frac{\log kn}{n}} + \frac{12\sqrt{3}\sigma\tau^{1/2}}{\vartheta_{\mathbf{p}}}\sqrt{\frac{\log n}{n}}\right) \geq 1 - \leq \frac{1}{pn/2} - \frac{1}{\{\log n/2\}^{1/2}} - \frac{8}{k\log n/2 - k} \tag{28}$$

7.2 Spatial dependence

Now let us consider the case where there is dependence between the different coordinates of the data stream. To be more specific, we suppose that $W_1, \dots, W_n \sim^{IID} N_p(0, \Sigma)$ where Σ is a positive definite matrix of size p by p. In this setting, our algorithm requires some modification.

When there is no missingness, we can see that the optimal projection direction changes, as what we are really trying to achieve is to maximise the signal-to-noise ratio. For $a \in \mathbb{S}^{p-1}$, we have $a^T X \sim N(a^T\mu, a^T\Sigma a)$. The maximiser of $\frac{|a^T\theta|}{\sqrt{a^T\Sigma a}}$ in $\mathbb{S}^{p-1}$ is now $\Sigma^{-1}\theta$ (properly normalised) [Wang and Samworth, 2018].

We can therefore apply the same idea in our missingness setting: if we have an estimate of Σ^{-1}, $\hat{\Sigma}^{-1}$, and an estimate of the leading left-singular vector of A_Ω, u, let us say $\hat{u}$, we should use as projection direction $\hat{\Sigma}^{-1}\hat{u}$. Estimating Σ^{-1} or Σ in itself can be an impossible challenge depending on the shape of the missingness. For example, if there are two coordinates with one missing all of the even data, and the other missing all of the uneven data, it is not possible to estimate Σ fully but it does not matter! The practitioner should therefore be cautious of how the estimate of $\hat{\Sigma}^{-1}$ is obtained.

8 Further work and Conclusion

There are multiple angles of this work that have been left untouched and that would have to be considered in order to write a paper and provide a thorough methodology. Below are provided axes of work to be developed later.

The cases of multiple changepoint has not been studied, but the algorithm has been modified. A more recent approach of binary segmentation provides better results and could be used. As for the theory, combining the arguments of the original Inspect paper as well as those of **Theorem 1** of this work should do the trick.

In the case of single changepoint detection, I have not given theoretical results or insights in the case of temporal dependence.

The numerical studies have probably not been pushed to the deepest they could have been. There are no tests results for the multiple changepoints settings or for the temporal and spatial dependence single changepoint setting. It would be particularly useful to compare the performance of our method to that of [Bardwell et al., 2016] and [Xie et al., 2013]. The numerical results might make us consider to change the optimisation scheme in terms of the l_1-norm penalty.

The convergence rate of **Theorem 1** could be improved. The minimax convergence rate being $1/n$ and the current rate of **Theorem 1** is $\sqrt{\{\log n\}/n}$. As much as the logarithmic factor is not a problem, the half-power is not optimal and the end of the proof could be improved through the adaption of **Lemma 6** of the original paper.

The conditions of **Proposition 1** and **Theorem 1** on $\mathbf{p}$ are not ideal. One can see that the conditions depend on the probability of presence of every row with mean change, no matter the actual importance in the change of said row. Moreover, those conditions are very hard to verify as $\{\log\log n\}/\log n$ decreases quite slowly.

Providing theoretical results for any observation matrix Ω would, in my opinion, be difficult, as the projection of the Ω-CUSUM transformation does not have the same clean representation as before.

The performance of the algorithm is still on point in many missingness generating mechanisms as seen in the Numerical Studies section. The practitioner should therefore not fear to use our methodology in any type of missingness for lack of theoretical results.

To conclude, we have provided a very-user friendly tool and method for single and multiple sparse changepoints detection in high-dimensional data with heterogeneous missingness. Some theoretical results have been developed and are encouraging, as well as numerical results that show the adaptability of the **Inspect Miss** method. We can therefore only hope that this will be of use to practitioners and of interests to fellow researchers.

References

[Aston and Kirch, 2013] Aston, J. and Kirch, C. (2013). Evaluating stationarity via change-point alternatives with applications to fmri data. *The Annals of Applied Statistics*, 6.

[Aue et al., 2009] Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.

[Bardwell et al., 2016] Bardwell, L., Eckley, I., Fearnhead, P., Smith, S., and Spott, M. (2016). Most recent changepoint detection in panel data.

[Chen et al., 2020] Chen, Y., Wang, T., and Samworth, R. J. (2020). High-dimensional, multiscale online changepoint detection.

[Cho, 2016] Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, 10(2):2000–2038.

[Cho and Fryzlewicz, 2014] Cho, H. and Fryzlewicz, P. (2014). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):475–507.

[Csörgö and Horváth, 1997] Csörgö, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*.

[Darling and Erdös, 1956] Darling, D. A. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.*, 23(1):143–155.

[De Vito et al., 2008] De Vito, S., Massera, E., Piga, M., Martinotto, L., and Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B Chemical*, 129:750–757.

[Enikeeva and Harchaoui, 2013] Enikeeva, F. and Harchaoui, Z. (2013). High-dimensional change-point detection with sparse alternatives.

[Frick et al., 2013] Frick, K., Munk, A., and Sieling, H. (2013). Multiscale change-point inference.

[Fryzlewicz, 2014] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection.

[Hampel, 1974] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393.

[Henry et al., 2010] Henry, D., Simani, S., and Patton, R. (2010). *Fault Detection and Diagnosis for Aeronautic and Aerospace Missions*, volume 399, pages 91–128.

[Horváth and Rice, 2014] Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *TEST*, 23:219–255.

[Jandhyala et al., 2002] Jandhyala, V., Hawkins, D., and Fotopoulos, S. B. (2002). Testing and locating variance changepoints with applications to stock prices (vol 92, pg 738). *Journal of the American Statistical Association*, 97:661–661.

[Killick et al., 2012] Killick, R., Fearnhead, P., and Eckley, I. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:1590–1598.

[Kirch et al., 2014] Kirch, C., Muhsal, B., and Ombao, H. (2014). Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association*, 110:00–00.

[Olshen et al., 2004] Olshen, A., Venkatraman, E., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based dna copy number data. *Biostatistics (Oxford, England)*, 5:557–72.

[Ombao et al., 2005] Ombao, H., von Sachs, R., and Guo, W. (2005). Slex analysis of multivariate nonstationary time series. *Journal of the American Statistical Association*, 100(470):519–531.

[Page, 1955] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3-4):523–527.

[Peng et al., 2004] Peng, T., Leckie, C., and Ramamohanarao, K. (2004). Proactively detecting ddos attack using source ip address monitoring.

[Sparks et al., 2010] Sparks, R., Keighley, T., and Muscatello, D. (2010). Early warning cusum plans for surveillance of negative binomial daily disease counts. *Journal of Applied Statistics*, 37:2010.

[Tillmann and Pfetsch, 2014] Tillmann, A. and Pfetsch, M. (2014). The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *Information Theory, IEEE Transactions on*, 60:1248–1259.

[Venkatraman et al., 1992] Venkatraman, E., of Statistics, S. U. D., and Agency, U. S. N. S. (1992). *Consistency Results in Multiple Change-point Problems*. Stanford University.

[Wang and Samworth, 2018] Wang, T. and Samworth, R. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80.

[Xie et al., 2013] Xie, Y., Huang, J., and Willett, R. (2013). Change-point detection for high-dimensional time series with missing data. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):12–27.

9 Appendix 1: Lemmas and Proofs

Lemma 1. *(Obtaining the optimisation iteration scheme)*

Let T_Ω be an p by $n-1$ real matrix. Let $\arg\max_{u\in\mathbb{S}^{p-1},w\in\mathbb{S}^{n-2}}\{\langle T_\Omega, uw^T\rangle - \lambda\sqrt{n}||u||_1\}$ be our optimisation problem, and (u_{max}, w_{max}) one of its solutions. Then

$$\begin{aligned} w_{max} &= \frac{T_\Omega^T u_{max}}{||T_\Omega^T u_{max}||_2} \\ u_{max} &= T_\Omega w_{max} \; o\max\left(0, 1-\frac{\lambda\sqrt{n}}{|T_\Omega w_{max}|}\right) / ||T_\Omega w_{max} \; o\max\left(0, 1-\frac{\lambda\sqrt{n}}{|T_\Omega w_{max}|}\right)||_2 \end{aligned} \tag{29}$$

Proof. The maximizers exist because the function is continuous on a closed and bounded set.

If we denote by $\mathcal{L}(u, w, \mu_u, \mu_w)$ the quantity:

$$\langle T_\Omega, uw^T\rangle - \lambda\sqrt{n}||u||_1 + \mu_u\left(||u||_2^2 - 1\right) + \mu_w\left(||w||_2^2 - 1\right) \tag{30}$$

then we get:

$$\frac{\partial\mathcal{L}}{\partial w} = T_\Omega^T u + 2\mu_w w \tag{31}$$

and

$$\frac{\partial\mathcal{L}}{\partial u_j} = (T_\Omega w)_j - \lambda\sqrt{n}\,\mathrm{sign}(u_j) + 2\mu_u u_j \text{ if } u_j \neq 0 \tag{32}$$

Therefore

$$w_{max} = \pm\frac{T_\Omega^T u_{max}}{||T_\Omega^T u_{max}||_2} \tag{33}$$

using the condition on the norm of w_{max}. By re-injecting in the quantity to maximize, we observe that the sign is necessarily a $+$. Hence the proposition on w_{max}.

Now for u_{max}:

- If $(u_{max})_j = 0$, $\langle (T_\Omega)_j, aw_{max}^T\rangle - \lambda\sqrt{n}|a| \leq 0 \quad \forall a \in \mathbb{R}$, therefore: $|(T_\Omega w_{max})_j| \leq \lambda\sqrt{n}$ and $1 - \frac{\lambda\sqrt{n}}{|(T_\Omega w_{max})_j|} \leq 0$.
- Else, $(u_{max})_j \neq 0$, $|(T_\Omega w_{max})_j| > \lambda\sqrt{n}$. We have $\mathrm{sign}\left(-(T_\Omega w)_j + \lambda\sqrt{n}\,\mathrm{sign}\left((u_{max})_j\right)\right) = -\mathrm{sign}\left((T_\Omega w)_j\right) = \mathrm{sign}\left(\mu_u (u_{max})_j\right)$. We get $\mathrm{sign}\left((u_{max})_j\right) = -\mathrm{sign}(\mu_u)\,\mathrm{sign}\left((T_\Omega w)_j\right)$. Then $(u_{max})_j = \frac{-(T_\Omega w_{max})_j - \lambda\sqrt{n}\,\mathrm{sign}(\mu_u)\,\mathrm{sign}\left((T_\Omega w_{max})_j\right)}{2\mu_u} = \frac{-(T_\Omega w_{max})_j}{2\mu_u} - \frac{\lambda\sqrt{n}\,\mathrm{sign}\left((T_\Omega w_{max})_j\right)}{|\mu_u|}$

But $|\mu_u|$ is fixed, by using the condition on the l_1 norm of u_{max}. By re-injecting in the quantity to be optimised, it appears that $\mu_u < 0$. Hence the property on u_{max}.

□

Lemma 2. *(Closeness between the estimate and the leading left singular vector of A_Ω):*

Let $u \in \mathbb{S}^{p-1}(k), w \in \mathbb{S}^{n-2}$ be the left leading singular vector and the right leading singular vector of A_Ω. Let $\lambda > 0$ be a tuning parameter. Let $\hat{u}, \hat{w} = \arg\max_{||w||_2=1, ||u||_2=1} \langle T_\Omega, uw^T \rangle - \lambda\sqrt{n}||u||_1$. Let's suppose that $||A_\Omega - T_\Omega||_\infty \leq \lambda$. Let δ be the singular gap of A_Ω. Then:

$$\sin \angle (u, \hat{u}) \leq \frac{4\lambda\sqrt{kn}}{\delta} \tag{34}$$

Proof. Using Lemma 2 of Original paper (p. 37):

$$||uw^T - \hat{u}\hat{w}^T||_2^2 \leq \frac{2}{\delta}\left(\langle T_\Omega, uw^T - \hat{u}\hat{w}^T\rangle + \langle A - T_\Omega, uw^T - \hat{u}\hat{w}^T\rangle\right) \tag{35}$$

Because $\hat{u}$ and $\hat{w}$ are maximizers of the optimisation problem, we have

$$\langle T_\Omega, \hat{u}\hat{w}^T\rangle - \langle T_\Omega, uw^T\rangle \leq \lambda\sqrt{n}\left(||u||_1 - ||\hat{u}||_1\right) \tag{36}$$

Using the preceding remark and the fact that $||A_\Omega - T_\Omega||_\infty \leq \lambda$ yields:

$$||uw^T - \hat{u}\hat{w}^T||_2^2 \leq \frac{2}{\delta}\left(\lambda\sqrt{n}\left(||u||_1 - ||\hat{u}||_1\right) + \lambda||uw^T - \hat{u}\hat{w}^T||_1\right) \tag{37}$$

Let us denote $S_u := \{j : 1 \leq j \leq p, u_j \neq 0\}$.

$$\begin{aligned} ||uw^T - \hat{u}\hat{w}^T||_2^2 \leq & \frac{2}{\delta}\left(\lambda\sqrt{n}\left(||u_{S_u}||_1 - ||\hat{u}_{S_u}||_1\right) + \lambda\left(||u_{S_u}w^T - \hat{u}_{S_u}\hat{w}^T||_1 + ||\hat{u}_{\bar{S_u}}||_1||\hat{w}||_1 - \sqrt{n}\right)\right) \\ \leq & \frac{2}{\delta}\left(\lambda\sqrt{n}\left(||u_{S_u}||_1 - ||\hat{u}_{S_u}||_1\right) + \lambda||u_{S_u}w^T - \hat{u}_{S_u}\hat{w}^T||_1\right) \end{aligned} \tag{38}$$

We remark that: $||u_{S_u}||_1 - ||\hat{u}_{S_u}||_1 \leq \min\left(||u_{S_u} - \hat{u}_{S_u}||_1, ||u_{S_u} + \hat{u}_{S_u}||_1\right)$. We also necessarily have $\min\left(||u_{S_u} - \hat{u}_{S_u}||_1, ||u_{S_u} + \hat{u}_{S_u}||_1\right) \leq \min\left(||u_{S_u} - \hat{u}_{S_u}||_2, ||u_{S_u} + \hat{u}_{S_u}||_2\right)\sqrt{k}$.

But using equation 19 page 39 of the original paper: $\min\left(||u_{S_u} - \hat{u}_{S_u}||_2, ||u_{S_u} + \hat{u}_{S_u}||_2\right) \leq ||u_{S_u}w^T - \hat{u}_{S_u}\hat{w}^T||_2$ as well as $\sin\angle(u, \hat{u}) \leq ||uw^T - \hat{u}\hat{w}^T||_2$.

Hence

$$\begin{aligned} ||uw^T - \hat{u}\hat{w}^T||_2^2 \leq & \frac{2}{\delta}\left(\lambda||uw^T - \hat{u}\hat{w}^T||_2\sqrt{kn} + \lambda||uw^T - \hat{u}\hat{w}^T||_2\sqrt{kn}\right) \\ \leq & \sin\angle(u, \hat{u}) \leq ||uw^T - \hat{u}\hat{w}^T||_2 \leq \frac{4\lambda\sqrt{kn}}{\delta} \end{aligned} \tag{39}$$

□

Lemma 3. *(Show that A_Ω is close to $\text{diag}\left(\sqrt{\mathbf{p}}\right)A$ entry-wise in the Homogeneous per row setting)*

If $\forall\ j \in [\![1, k]\!]$, $\frac{24\log\{k\log n\}}{p_j \log n} \leq 1$ and $3 \leq k\log n \leq k\tau n$, then:

$$\mathbb{P}\left(\forall j, t, |\left(A_\Omega\right)_{j,t} - \left(\text{diag}\left(\sqrt{\mathbf{p}}\right)A\right)_{j,t}| \leq 6\sqrt{6}|\theta_j|\sqrt{\log kn}\right) \geq 1 - \frac{4}{k\log n - k} \tag{40}$$

Proof. Let $\epsilon_{j,t} := \sqrt{6p_j t\log kt}$ and $a := \log n$. For ease, let us write $\tilde{t}_j = \sum_{r=1}^{t} w_{j,r}$ for $1 \leq t \leq n$ and $P = A_\Omega - \text{diag}\left(\sqrt{\mathbf{p}}\right)A$. Let us also define:

$$\Omega_0^c = \cup_{j=1}^{k}\left(\cup_{t=0}^{n-\lceil a\rceil}|\tilde{n}_j - \tilde{t}_j - p_j(n-t)| \geq \epsilon_{j,n-t}\right) \cup \left(\cup_{t=\lceil a\rceil}^{n}|\tilde{t}_j - p_j t| \geq \epsilon_{j,t}\right) \tag{41}$$

.

Using multiplicative Chernoff bounds we get that:

$$\begin{aligned}\mathbb{P}\left(\cup_{j=1}^{k}\left(\cup_{t=0}^{n-\lceil a\rceil}|\tilde{n}_j-\tilde{t}_j-p_j\left(n-t\right)|\geq\epsilon_{j,n-t}\right)\cup\left(\cup_{t=\lceil a\rceil}^{n}|\tilde{t}_j-p_jt|\geq\epsilon_{j,t}\right)\right)&\\ \leq 2\sum_{j=1}^{k}\sum_{t=\lceil a\rceil}^{n}2\exp\left(-2\log kt\right)&\\ \leq 2\sum_{j=1}^{k}\sum_{t=\lceil a\rceil}^{n}\frac{2}{(kt)^2}&\\ \leq\frac{4}{k^2}\sum_{j=1}^{k}\frac{1}{\lceil a\rceil-1}&\\ \leq\frac{4}{(k\log n-k)}&\end{aligned}\tag{42}$$

We now work on the event Ω_0.

For t such that $a\leq t\leq n-a$, since $\frac{\epsilon_{j,t}}{p_jt}$ is decreasing (as a function of t after a):

$$|P_{j,t}|\leq|\theta_j|\sqrt{p_j}\gamma_t\left(\frac{\left(1+\epsilon_{j,\min(t,n-t,\tau n)}/\left(p_j\min(t,n-t,\tau n)\right)\right)^{3/2}}{\left(1-\epsilon_{j,\min(t,n-t,\tau n)}/\left(p_j\min(t,n-t,\tau n)\right)\right)}-1\right)\tag{43}$$

Since $\frac{\epsilon_{j,\min(t,n-t,\tau n)}}{p_j\min(t,n-t,\tau n)}\leq 1/2$:

$$|P_{j,t}|\leq 6|\theta_j|\sqrt{p_j}\gamma_t\frac{\epsilon_{j,\min(t,n-t,\tau n)}}{p_j\min(t,n-t,\tau n)}=|\theta_j|\gamma_t6\sqrt{\frac{6\log k(\min(t,n-t,\tau n))}{\min(t,n-t,\tau n)}}\tag{44}$$

But $\gamma_t\leq\sqrt{\min(t,n-t,\tau n)}$, therefore: $|P_{j,t}|\leq|\theta_j|6\sqrt{6}\sqrt{\log(k\min(t,n-t,\tau n))}\leq|\theta_j|6\sqrt{6}\sqrt{\log kn}$

For $t<a, n-a<t$, since $\frac{\epsilon_{j,a}}{p_ja}\leq 1/2$:

$$|P_{j,t}|\leq|\theta_j|\gamma_t\max(\sqrt{p_j},\frac{1+\epsilon_{j,a}/p_ja}{1-\epsilon_{j,a}/p_ja}-\sqrt{p_j})\leq 3|\theta_j|\gamma_t\tag{45}$$

Therefore: $|P_{j,t}|\leq|\theta_j|3\sqrt{\log n}$.

This yields that

$$\forall\ t\in[\![1,n-1]\!],j\in[\![1,p]\!],|P_{j,t}|\leq|\theta_j|6\sqrt{6}\sqrt{\log kn}\tag{46}$$

□

Proof. **(Proposition 1)**

Using **Lemma 2** of this document, on the event that $||A_\Omega-T_\Omega||_\infty\leq\lambda$:

$$\sin\angle\left(u,\hat{u}\right)\leq\frac{4\lambda\sqrt{kn}}{\delta}\tag{47}$$

We have $A_\Omega=\operatorname{diag}\left(\sqrt{\mathbf{p}}\right)A+A_\Omega-\operatorname{diag}\left(\sqrt{\mathbf{p}}\right)A=\operatorname{diag}\left(\sqrt{\mathbf{p}}\right)A+P$. Let $\sigma_1\geq\ldots\geq\sigma_p$ be the singular values of A_Ω. Using Weyl's theorem:

$$|\ ||\theta_{\mathbf{p}}||_2||\gamma||_2-\sigma_1|\leq\sqrt{2}||P||_{op}\text{ and }\forall i>1,\quad|\sigma_i|\leq\sqrt{2}||P||_{op}\tag{48}$$

Hence the singular gap: $\delta = \sigma_1 - \sigma_2 \geq ||\theta_{\mathbf{p}}||_2||\gamma||_2 - 2\sqrt{2}||P||_{op} \geq ||\theta_{\mathbf{p}}||_2||\gamma||_2 - 2\sqrt{2}||P||_F$

On the event that $||P||_F = ||A_\Omega - \text{diag}(\sqrt{\mathbf{p}})A||_F \leq \frac{||\theta_{\mathbf{p}}||_2||\gamma||_2}{4\sqrt{2}}$, we have $\delta \geq \frac{||\theta_{\mathbf{p}}||_2||\gamma||_2}{2} \geq \frac{1}{8}\vartheta_{\mathbf{p}} n\tau$, where the last inequality uses **Lemma 3** p 39 of the original paper.

Therefore on both of those events:

$$\sin \angle (u, \hat{u}) \leq \frac{32\lambda\sqrt{k}}{\vartheta_{\mathbf{p}}\tau\sqrt{n}} \tag{49}$$

Now for the probability bounds: $||A_\Omega - \text{diag}\left(\sqrt{\mathbf{p}}\right) A||_F \leq \frac{||\theta_{\mathbf{p}}||_2||\gamma||_2}{4\sqrt{2}}$ and $||A_\Omega - T_\Omega||_\infty \leq \lambda$ are independent events. Using **Lemma 3** and the conditions $\forall\ j \in [\![1, k]\!]$, $\frac{24\log\{k\log n\}}{p_j \log n} \leq 1$, $3 \leq k\log n \leq k\tau n$, $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/2\}}{\log kn}} \leq 1$ yields:

$$\sup_{P\in\mathcal{P}(n,p,k,1,\vartheta_{\mathbf{p}},\tau,\sigma^2,\mathbf{p})} \mathbb{P}_P\left(||A_\Omega - \text{diag}\left(\sqrt{\mathbf{p}}\right) A||_F \geq ||\theta||_2 6\sqrt{6}n\sqrt{\log kn} \geq \frac{||\theta_{\mathbf{p}}||_2||\gamma||_2}{4\sqrt{2}}\right) \leq \frac{4}{k\log n - k} \tag{50}$$

Using **Lemma 4** p. 41 of the original paper and the conditions $n \geq 6$ and $\lambda \geq 2\sigma\sqrt{\log\{p\log n\}}$ yields:

$$\sup_{P\in\mathcal{P}(n,p,k,1,\vartheta_{\mathbf{p}},\tau,\sigma^2,\mathbf{p})} \mathbb{P}_P\left(||A_\Omega - T_\Omega||_\infty \geq \lambda\right) \leq \frac{4}{\{p\log n\}^{1/2}} \tag{51}$$

$$\sup_{P\in\mathcal{P}(n,p,k,1,\vartheta_{\mathbf{p}},\tau,\sigma^2,\mathbf{p})} \mathbb{P}_{\mathcal{P}}\left(||A_\Omega - \text{diag}\left(\sqrt{\mathbf{p}}\right) A||_F \leq ||\theta_{\mathbf{p}}||_2||\gamma||_2/(4\sqrt{2}) \cap\ ||A_\Omega - T_\Omega||_\infty \leq \lambda\right) \geq 1 - \frac{4}{\{p\log n\}^{1/2}} + \frac{4}{k\log n - k} \tag{52}$$

□

Proof. **(Theorem 1)**

Recall the definition of $X_{\Omega^{(1)}}$, $X_{\Omega^{(2)}}$, $T_{\Omega^{(1)}}$ and $T_{\Omega^{(2)}}$. Define similarly $\mu^{(1)} = (\mu_1^{(1)}, \ldots, \mu_{n_1}^{(1)})$, $\mu^{(2)} = (\mu_1^{(2)}, \ldots, \mu_{n_1}^{(2)}) \in \mathbb{R}^{p\times n_1}$ by $\mu_t^{(2)} := \mu_{2t}$, $\mu_t^{(1)} := \mu_{2t-1}$. Let $W^{(2)} = (W_1^{(2)}, \ldots, W_{n_1}^{(2)})$ be a random matrix taking values in $\mathbb{R}^{p\times n_1}$ $W_t^{(2)} = W_{2t}$. Now let $A_{\Omega^{(1)}} := \mathcal{T}_{\Omega^{(1)}}(\mu^{(1)}_{\Omega^{(1)}})$, $A_{\Omega^{(2)}} := \mathcal{T}_{\Omega^{(2)}}(\mu^{(2)}_{\Omega^{(2)}})$, $A^{(1)} = \mathcal{T}(\mu^{(1)})$, $A^{(2)} = \mathcal{T}(\mu^{(2)})$ and $E_{\Omega^{(2)}} := \mathcal{T}_{\Omega^{(2)}}(W^{(2)}_{\Omega^{(2)}})$. Furthermore, we write $\bar{T}_\Omega := (\hat{u}^{(1)})^T T_{\Omega^{(2)}}$, $\bar{A}_\Omega := (\hat{u}^{(1)})^T A_{\Omega^{(2)}}$, $\bar{E}_\Omega := (\hat{u}^{(1)})^T E_{\Omega^{(2)}}$, $\overline{\text{diag}(\sqrt{\mathbf{p}})A} = (\hat{u}^{(1)})^T \text{diag}(\sqrt{\mathbf{p}})A^{(2)}$ and $\overline{\theta_{\mathbf{p}}} := (\hat{u}^{(1)})^T\theta_{\mathbf{p}}$ the one-dimensional projected images (as row vectors) of the corresponding p-dimensional quantities. Let $P^{(1)} := A_{\Omega^{(1)}} - \text{diag}(\sqrt{\mathbf{p}})A^{(1)}$, $P^{(2)} := A_{\Omega^{(2)}} - \text{diag}(\sqrt{\mathbf{p}})A^{(2)}$, $\bar{P} := (\hat{u}^{(1)})^T P^{(2)}$, $z^{(2)} := z/2$, $\hat{z}^{(2)} := \hat{z}/2$.

We have that for each $t \in [\![1, n_1 - 1]\!]$ $\bar{E}_{\Omega,t}$ follows a $N(0, \sigma^2)$ distribution but the entries are not independent. Let $\lambda_1 = \sigma\sqrt{\log\{n/2\}}$. Let $Y \sim N(0, 1)$.

$$\mathbb{P}(||\bar{E}_\Omega||_\infty \geq \lambda_1) \leq \sum_{t=1}^{n_1-1} \mathbb{P}(|\bar{E}_{\Omega,t}| \geq \lambda_1) = (n_1 - 1)\mathbb{P}(\sigma|Y| \geq \lambda_1) \leq 2n_1 \frac{1}{\sqrt{2\pi}\lambda_1/2\sigma^2} e^{-\lambda_1^2/2\sigma^2} \leq \frac{1}{\sqrt{\log\{n/2\}}} \tag{53}$$

Therefore we have

$$\mathbb{P}(||\bar{E}_\Omega||_\infty \leq \lambda_1) \geq 1 - \frac{1}{\sqrt{\log n/2}} \tag{54}$$

We define $\Omega_1 := \{\forall\, t \in [\![1, n_1-1]\!], j \in [\![1,p]\!], |P^{(1)}_{j,t}| \leq |\theta_j|6\sqrt{6}\sqrt{\log kn/2}, |P^{(2)}_{j,t}| \leq |\theta_j|6\sqrt{6}\sqrt{\log kn/2}\}$, $\Omega_2 := \{||\bar{E}_\Omega||_\infty \leq \sigma\sqrt{\log n/2}\}$ and $\Omega_3 := \{||A_{\Omega^{(1)}} - T_{\Omega^{(1)}}||_\infty \leq \lambda\}$.

As the conditions to apply **Lemma 3** twice (on $P^{(1)}$ and $P^{(2)}$) are verified (for Ω_1), $\lambda \geq 2\sigma\sqrt{\log\{p\log n/2\}}$ and $n \geq 12$ allowing us to use **Lemma 4** p.41 (for Ω_3), and using (54) (for Ω_2) we have:

$$\mathbb{P}\left(\Omega_1^c \cup \Omega_2^c \cup \Omega_3^c\right) \leq \frac{4}{\{p\log n/2\}^{1/2}} + \frac{1}{\{\log n/2\}^{1/2}} + \frac{8}{k\log n/2 - k} \tag{55}$$

We now work on $\Omega_1 \cap \Omega_2 \cap \Omega_3$.

Using $\frac{128\sigma}{\vartheta_{\mathbf{p}}\tau}\sqrt{\frac{k\log\{p\log n/2\}}{n/2}} \leq 1$ and $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/4\}}{\log kn/2}} \leq 1$ yields $\angle(u,\hat{u}) \leq \pi/6$ (as in the proof of **Proposition 1**).

We also have $\sin\angle(u,\theta_{\mathbf{p}}) \leq ||uw^T - \frac{\mathrm{diag}(\sqrt{\mathbf{p}})A^{(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2}||_2 \leq ||uw^T - \frac{A_{\Omega(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2}||_2 + ||\frac{A_{\Omega(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2} - \frac{\mathrm{diag}(\sqrt{\mathbf{p}})A^{(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2}||_2$ $\leq 2||\frac{A_{\Omega(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2} - \frac{\mathrm{diag}(\sqrt{\mathbf{p}})A^{(1)}}{||\theta_{\mathbf{p}}||_2||\gamma||_2}||_2 \leq \frac{1}{2\sqrt{2}} \leq 1/2$, where u and w are the leading left and right singular vectors of $A_{\Omega^{(1)}}$. Therefore $\angle(\theta_{\mathbf{p}},\hat{u}) \leq \pi/3$.

Hence we have $\angle(\hat{u},\theta_{\mathbf{p}}) \leq \pi/3$ and therefore $|\overline{\theta_{\mathbf{p}}}| \geq \vartheta_{\mathbf{p}}/2$.

We also know that $|\overline{\left(\mathrm{diag}\left(\sqrt{\mathbf{p}}\right)A\right)}_t - \bar{A}_{\Omega,t}| \leq |\sum_{j=1}^p \theta_j\hat{u}_j|6\sqrt{6}\sqrt{\log kn/2} \leq |\overline{\theta_{\mathbf{p}}}|6\sqrt{6}\sqrt{\log kn/2}\ \forall t \in [\![1, n_1-1]\!]$.

Since $(\overline{\left(\mathrm{diag}\left(\sqrt{\mathbf{p}}\right)A\right)}_t)_t$ and $(\bar{T}_{\Omega,t})_t$ are respectively maximised in absolute value at $t = z$ and $t = \hat{z}$, and $\bar{T}_\Omega = \bar{A}_\Omega + \bar{E}_\Omega = \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A} + \bar{P} + \bar{E}_\Omega$, we have:

Case 1: $\overline{\theta_{\mathbf{p}}} \geq 0$

$$\begin{aligned}
0 &\leq \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A_{z^{(2)}}} - \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}_{\hat{z}^{(2)}} \\
&\leq \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A_z} - \bar{A}_{\Omega,z} + \bar{A}_{\Omega,z} - |\bar{T}_{\Omega,z}| + |\bar{T}_{\Omega,z}| - |\bar{T}_{\Omega,\hat{z}}| + \bar{T}_{\Omega,\hat{z}} - \bar{A}_{\Omega,\hat{z}} + \bar{A}_{\Omega,\hat{z}} - \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}_{\hat{z}} \\
&\leq |\overline{\mathrm{diag}(\sqrt{\mathbf{p}})A_z} - \bar{A}_{\Omega,z}| + |\bar{A}_{\Omega,z} - |\bar{T}_{\Omega,z}|\ |\ +|\ |\bar{T}_{\Omega,\hat{z}}| - \bar{A}_{\Omega,\hat{z}}| + |\bar{A}_{\Omega,\hat{z}} - \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}_{\hat{z}}| \\
&\leq |\overline{\mathrm{diag}(\sqrt{\mathbf{p}})A_z} - \bar{A}_{\Omega,z}| + |\ |\bar{A}_{\Omega,z}| - |\bar{T}_{\Omega,z}|\ |\ +|\ |\bar{T}_{\Omega,\hat{z}}| - |\bar{A}_{\Omega,\hat{z}}|\ |\ + |\bar{A}_{\Omega,\hat{z}} - \overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}_{\hat{z}}| \\
&\leq 12\sqrt{6}\ |\overline{\theta_{\mathbf{p}}}|\sqrt{\log kn/2} + 2\lambda_1 \leq 12\sqrt{6}\sqrt{\log kn/2}\ |\overline{\theta_{\mathbf{p}}}| + 2\sigma\sqrt{\log n/2}
\end{aligned} \tag{56}$$

The other case yields the same result.

The row vector $\overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}$ has the explicit form

$$\overline{\mathrm{diag}(\sqrt{\mathbf{p}})A}_t = \begin{cases} \sqrt{\frac{t}{n_1(n_1-t)}}(n_1 - z^{(2)})\overline{\theta_{\mathbf{p}}} & \text{if } t \leq z^{(2)} \\ \sqrt{\frac{n_1-t}{n_1 t}}z^{(2)}\overline{\theta_{\mathbf{p}}} & \text{if } t > z^{(2)} \end{cases} \tag{57}$$

Because $12\sqrt{6}\sqrt{\log kn/2}\ |\overline{\theta_{\mathbf{p}}}| + 2\sigma\sqrt{\log n/2} \leq |\overline{\theta_{\mathbf{p}}}|\frac{\sqrt{n\tau/2}}{3\sqrt{6}}$ and **Lemma 7** p.43, we necessarily have $|\hat{z}^{(2)} - z^{(2)}| \leq n_1\tau/2$, therefore we can apply **Lemma 7** in full strength, yielding:

$$\frac{|\hat{z}-z|}{n} \leq \frac{6\sqrt{3}\tau^{1/2}\left(12\sqrt{6}\sqrt{\log kn}\ |\overline{\theta_{\mathbf{p}}}| + 2\lambda_1\right)}{2|\overline{\theta_{\mathbf{p}}}|\sqrt{n}} \leq 108\sqrt{2}\tau^{1/2}\sqrt{\frac{\log kn}{n}} + \frac{12\sqrt{3}\sigma\tau^{1/2}}{\vartheta_{\mathbf{p}}}\sqrt{\frac{\log n}{n}} \tag{58}$$

□

Proof. **Theorem 2**

We define: $\Omega_1 := \{\forall\, t \in [\![1, n_1-1]\!], j \in [\![1, p]\!], |P^{(1)}_{j,t}| \leq |\theta_j| 6\sqrt{6}\sqrt{\log kn/2}, |P^{(2)}_{j,t}| \leq |\theta_j| 6\sqrt{6}\sqrt{\log kn/2}\}$, $\Omega_2 := \{||\bar{E}_\Omega||_\infty \leq \sigma\sqrt{\log n/2}\}$ and $\Omega_3 := \{||A_{\Omega^{(1)}} - T_{\Omega^{(1)}}||_\infty \leq \lambda\}$.

Writing $E_{\Omega^{(1)}} := \mathcal{T}_{\Omega^{(1)}}(W_{\Omega^{(1)}})$ and $n_1 := n/2$, by **Lemma 10** p 45 of the original Inspect paper (on-line supplement) and a union bound, we have that the event $\{||E_{\Omega^{(1)}}||_\infty \leq \lambda\}$ satisfies

$$\mathbb{P}\left(\{||E_{\Omega^{(1)}}||_\infty \geq \lambda\ \right) \leq (n_1 - 1)p \exp\{-2\log(n_1 p)\} \leq \frac{1}{n_1 p} \tag{59}$$

Using the conditions presented (as in the proof of **Theorem 1**) but using **Lemma 10** p 45 of the original Inspect paper for Ω_3 yields:

$$\mathbb{P}\left(\Omega_1^c \cup \Omega_2^c \cup \Omega_3^c\right) \leq \frac{1}{pn/2} + \frac{1}{\{\log n/2\}^{1/2}} + \frac{8}{k\log n/2 - k} \tag{60}$$

We now work on $\Omega_1 \cap \Omega_2 \cap \Omega_3$.

Moreover, following the proof of **Proposition 1** and given that the constant C is sufficiently large, $\frac{\tau}{48\sqrt{3}}\sqrt{\frac{\log\{en/4\}}{\log kn/2}} \leq 1$ yields $\angle(u, \hat{u}) \leq \pi/6$ and the conditions necessary to apply **Lemma 3**:

$$\sin(\angle(\hat{u}, u)) \leq \frac{32\lambda\sqrt{k}}{\tau\vartheta_{\mathbf{p}}\sqrt{n_1}} \leq \frac{64\sigma\sqrt{kB\log\{n_1 p\}}}{\tau\vartheta_{\mathbf{p}}\sqrt{n_1}} \leq 1/2 \tag{61}$$

We also have $\sin(\angle(u, \theta_{\mathbf{p}}) \leq 1/2$. Therefore $\angle(\hat{u}, \theta_{\mathbf{p}}) \leq \pi/3$.

The proof therefore follows from that of **Theorem 1**. □

10 Appendix 2: InspectChangepointMissingness Package Documentation

The **InspectChangepointMissingness** package (in **R**) is an extension of the **InspectChangepoint** package, which was presented in [Wang and Samworth, 2018].

The two functions **multi.change** and **single.change** have been updated to allow for the simulation of different missingness-generating mechanisms. **locate.change.miss** implements the single changepoint estimation method presented in **Algorithm 2** and uses the adapted Ω-CUSUM transformation implemented in **cusum.transform.miss**. The multiple changepoints detection method is implemented in **inspect.miss** which uses a threshold adapted to the missingness computed by **compute.threshold.miss**.

The documentation of the package can be found right below.

Package 'InspectChangepointMissingness'

August 20, 2020

Type Package

Title High-Dimensional Changepoint Estimation with Missing Data via Sparse Projection

Version 0.1.0

Author Bertille Follain, Tengyao Wang, Richard Samworth

Maintainer Bertille Follain <bertille.follain@polytechnique.edu>

Imports stats, graphics, MASS, Rlab, matrixcalc

Description Provides a data-driven projection-based method for estimating changepoints in highdimensional time series with missing data.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.0

R topics documented:

`compute.threshold` *Computing threshold used in* `inspect`

Description

The threshold level to be used in `inspect` is computed via Monte Carlo simulation of multivariate time series that do not contain any changepoints.

Usage

```
compute.threshold(n, p, nrep = 100)
```

Arguments

`n`	Time length of the observation.
`p`	Dimension of the multivariate time series.
`nrep`	Number of Monte Carlo repetition to be used.

Value

A numeric value indicating the threshold level that should be used based on the Monte Carlo simulation.

Examples

```
compute.threshold(n=200, p=50)
```

`compute.threshold.miss`
Computing threshold used in `inspect.miss`

Description

The threshold level to be used in `inspect.miss` is computed via Monte Carlo simulation of multivariate time series that do not contain any changepoints with the same observation matrix as the argument of inspect miss.

Usage

```
compute.threshold.miss(n, p, omega, nrep = 100)
```

Arguments

n	Time length of the observation.
p	Dimension of the multivariate time series.
omega	Observation matrix of original data
nrep	Number of Monte Carlo repetition to be used.

Value

A numeric value indicating the threshold level that should be used based on the Monte Carlo simulation.

Examples

```
compute.threshold.miss(n=200, p=50, omega = matrix(rbinom(200*50,1,0.5),50,200))
```

---

cum.miss *Cumulative sums for vector with NAs used in* cusum.transform.miss

---

Description

Computes the cummulative sums of a vector considering that NA values are zeroes

Usage

```
cum.miss(x)
```

Arguments

x	Input vector

Value

The vector of the cummulative sums of observed values

Examples

```
n <- 2000; p <- 1000; k <- 10; z <- 400; vartheta <- 0.8; sigma <- 1; shape <- 3
noise <- 0; corr <- 0; miss <- 1; proba <- 0.5; oncp <- 0
obj <- single.change(n,p,k,z,vartheta,sigma,shape,noise,corr,miss,proba,oncp)
x <- obj$x
cum.miss(x)
```

`cum.obs` *Cumulative count of observed values used in* `cusum.transform.miss`

Description

Computes the cumulative counts of observed (not NAs) values

Usage

```
cum.obs(x)
```

Arguments

`x`	Input vector

Value

The vector of the cumulative counts of observed values

Examples

```
n <- 2000; p <- 1000; k <- 10; z <- 400; vartheta <- 0.8; sigma <- 1; shape <- 3
noise <- 0; corr <- 0; miss <- 1; proba <- 0.5; oncp <- 0
obj <- single.change(n,p,k,z,vartheta,sigma,shape,noise,corr,miss,proba,oncp)
x <- obj$x
cum.obs(x)
```

`cusum.transform` *CUSUM transformation*

Description

Performing CUSUM transformation to the input matrix of multivariate time series. If the input is a vector, it is treated as a matrix of one row.

Usage

```
cusum.transform(x)
```

Arguments

`x`	input matrix

Details

For any integers p and n, the CUSUM transformation $T_{p,n} : R^{p\times n} \to R^{p\times(n-1)}$ is defined by

$$[T_{p,n}(M)]_{j,t} := \sqrt{t(n-t)/n}\left(\frac{1}{n-t}\sum_{r=t+1}^{n} M_{j,r} - \frac{1}{t}\sum_{r=1}^{t} M_{j,r}\right).$$

Value

The transformed matrix is returned. Note that the returned matrix has the same number of rows but one fewer columns compared with the input matrix.

Examples

```
x <- matrix(rnorm(20),4,5)
cusum.transform(x)
```

`cusum.transform.miss` *CUSUM transformation adapted to missingness*

Description

Performing CUSUM transformation to the input matrix of multivariate time series with missingness. If the input is a vector, it is treated as a matrix of one row.

Usage

```
cusum.transform.miss(x)
```

Arguments

`x`	input matrix

Details

For any integers p and n and observation matrix Ω, the CUSUM transformation $T_\Omega : R_\Omega^{p\times n} \to R^{p\times(n-1)}$ is defined by

$$[T_\Omega(M)]_{j,t} = \sqrt{\frac{\left(\sum_{r=1}^{t} w_{j,r}\right)\left(\sum_{r=t+1}^{n} w_{j,r}\right)}{\sum_{r=1}^{n} w_{j,r}}}\left(\frac{\sum_{r=t+1}^{n} w_{j,r}M_{j,r}}{\sum_{r=t+1}^{n} w_{j,r}} - \frac{\sum_{r=1}^{t} w_{j,r}M_{j,r}}{\sum_{r=1}^{t} w_{j,r}}\right)$$

If this value is not computable because of a division by 0, it is inputed as 0.

Value

The transformed matrix is returned. Note that the returned matrix has the same number of rows but one fewer columns compared with the input matrix.

Examples

```
n <- 2000; p <- 1000; k <- 10; z <- 400; vartheta <- 0.8; sigma <- 1; shape <- 3
noise <- 0; corr <- 0; miss <- 1; proba <- 0.5; oncp <- 0
obj <- single.change(n,p,k,z,vartheta,sigma,shape,noise,corr,miss,proba,oncp)
x <- obj$x
cusum.transform.miss(x)
```

`inspect` *Informative sparse projection for estimation of changepoints (inspect)*

Description

This is the main function of the package InspectChangepoint. The function `inspect` estimates the locations of multiple changepoints in the mean structure of a multivariate time series. Multiple changepoints are estimated using a (wild) binary segmentation scheme, whereas each segmentation step uses the `locate.change` function.

Usage

```
inspect(x, lambda, threshold, schatten=c(1,2), M)
```

Arguments

`x`	The input data matrix of a high-dimensional time series, with each component time series stored as a row.
`lambda`	Regularisation parameter used in `locate.change`. If no value is supplied, the dafault value is chosen to be log(log(n)*p/2), where p and n are the number of rows and columns of the data matrix x respectively.
`threshold`	Threshold level for testing whether an identified changepoint is a true changepoint. If no value is supplied, the threshold level is computed via Monte Carlo simulation of 100 repetitions from the null model.
`schatten`	The Schatten norm constraint to use in the `locate.change` function. Default is schatten = 2, i.e. a Frobenius norm constraint.
`M`	The Monte Carlo parameter used for wild binary segmentation. Default is M = 0, which means a classical binary segmentation scheme is used.

Details

The input time series is first standardised using the `rescale.variance` function. Recursive calls of the `locate.change` function then segments the multivariate time series using (wild) binary segmentation. A changepoint at time z is defined here to mean that the time series has constant mean structure for time up to and including z and constant mean structure for time from z+1 onwards.

More details about model assumption and theoretical guarantees can be found in Wang and Samworth (2016). Note that Monte Carlo computation of the threshold value can be slow, especially for large p. If `inspect` is to be used multiple times with the same (or similar) data matrix size, it is better to precompute the threshold level via Monte Carlo simulation by calling the `compute.threshold` function.

Value

The return value is an S3 object of class 'inspect'. It contains a list of two objeccts:

- x The input data matrix

- changepoints A matrix with three columns. The first column contains the locations of estimated changepoints sorted in increasing order; the second column contains the maximum CUSUM statistics of the projected univariate time series associated with each estimated changepoint; the third column contains the depth of binary segmentation for each detected changepoint.

References

Wang, T. and Samworth, R. J. (2018) High dimensional changepoint estimation via sparse projection. *J. Roy. Statist. Soc., Ser. B*, **80**, 57–83.

Examples

```
n <- 500; p <- 100; ks <- 30; zs <- c(125,250,375)
varthetas <- c(0.1,0.15,0.2); overlap <- 0.5
obj <- multi.change(n, p, ks, zs, varthetas, overlap)
x <- obj$x
threshold <- compute.threshold(n,p)
ret <- inspect(x, threshold = threshold)
ret
summary(ret)
plot(ret)
```

`inspect.miss` *Informative sparse projection for estimation of changepoints with missing data (inspect miss)*

Description

This is the main function of the package InspectChangepointMissingness. The function `inspect` estimates the locations of multiple changepoints in the mean structure of a multivariate time series with missing data. Multiple changepoints are estimated using a (wild) binary segmentation scheme, whereas each segmentation step uses the `locate.change` function.

Usage

```
inspect.miss(x, lambda, threshold, M)
```

Arguments

`x`	The input data matrix of a high-dimensional time series, with each component time series stored as a row.
`lambda`	Regularisation parameter used in `locate.change`. If no value is supplied, the dafault value is chosen to be sqrt(log(log(n)*p/2), where p and n are the number of rows and columns of the data matrix x respectively.
`threshold`	Threshold level for testing whether an identified changepoint is a true changepoint. If no value is supplied, the threshold level is computed via Monte Carlo simulation of 100 repetitions from the null model with the observation matrix of the data.
`M`	The Monte Carlo parameter used for wild binary segmentation. Default is M = 0, which means a classical binary segmentation scheme is used.

Details

The input time series is first standardised using the rescale.variance function. Recursive calls of the locate.change function then segments the multivariate time series using (wild) binary segmentation. A changepoint at time z is defined here to mean that the time series has constant mean structure for time up to and including z and constant mean structure for time from z+1 onwards.

More details about model assumption and theoretical guarantees can be found in UNPUBLISHED. Note that Monte Carlo computation of the threshold value can be slow, especially for large p. If inspect is to be used multiple times with the same (or similar) data matrix size, it is better to precompute the threshold level via Monte Carlo simulation by calling the compute.threshold function.

Value

The return value is an S3 object of class 'inspect.miss'. It contains a list of two objeccts:

- xThe input data matrix
- changepoints A matrix with three columns. The first column contains the locations of estimated changepoints sorted in increasing order; the second column contains the maximum CUSUM statistics of the projected univariate time series associated with each estimated changepoint; the third column contains the depth of binary segmentation for each detected changepoint.

References

Unpublished yet

Examples

```
n <- 500; p <- 100; ks <- 30; zs <- c(125,250,375)
varthetas <- c(12,15,20); sigma <- 1; overlap <- 0.5
shape <- 1; miss <- 0; proba <- 0.8
obj <- multi.change(n, p, ks, zs, varthetas, sigma, overlap, shape, miss, proba)
x <- obj$x
omega <- obj$omega
threshold <- compute.threshold.miss(n,p,omega)
ret <- inspect.miss(x, threshold = threshold)
ret
summary(ret)
plot(ret)
```

---

locate.change *Single changepoint estimation*

Description

Estimate the location of one changepoint in a multivariate time series. It uses the function sparse.svd to estimate the best projection direction, then using univariate CUSUM statistics of the projected time series to estimate the changepoint location.

Usage

```
locate.change(
  x,
  lambda,
  schatten = 2,
  sample.splitting = FALSE,
  standardize.series = FALSE,
  view.cusum = FALSE
)
```

Arguments

`x`	A (p x n) data matrix of multivariate time series, each column represents a data point
`lambda`	Regularisation parameter. If no value is supplied, the dafault value is chosen to be sqrt(log(log(n)*p/2)) for p and n number of rows and columns of the data matrix x respectively.
`schatten`	The Schatten norm constraint to use in the `sparse.svd` function. Default is schatten = 2, i.e. a Frobenius norm constraint.
`sample.splitting`	Whether the changepoint should be estimated via sample splitting. The theoretical result is proven only for the sample splitted version of the algorithm. However, the default setting in practice is without sample splitting.
`standardize.series`	Whether the given time series should be standardised before estimating the projection direction. Default is FALSE, i.e. the input series is assume to have variance 1 in each coordinate.
`view.cusum`	Whether to show a plot of the projected CUSUM series

Value

A list of two items:

- changepoint - A single integer value estimate of the changepoint location is returned. If the estimated changepoint is z, it means that the multivariate time series is piecewise constant up to z and from z+1 onwards.
- cusum - The maximum absolute CUSUM statistic of the projected univariate time series associated with the estimated changepoint.
- vector.proj - the vector of projection, which is proportional to an estimate of the vector of change.

References

Wang, T., Samworth, R. J. (2016) High-dimensional changepoint estimation via sparse projection. Arxiv preprint: arxiv1606.06246.

Examples

```
n <- 2000; p <- 1000; k <- 32; z <- 400; vartheta <- 0.12; sigma <- 1; shape <- 3
noise <- 0; corr <- 0
obj <- single.change(n,p,k,z,vartheta,sigma,shape,noise,corr)
```

```
x <- obj$x
locate.change(x)
```

`locate.change.miss` *Single changepoint estimation with missing data*

Description

Estimate the location of one changepoint in a multivariate time series with missingness. It uses the function `sparse.svd.miss` to estimate the best projection direction, then using the projected CUSUM transformation (which is adapted to the missingness pattern) to estimate the changepoint location.

Usage

```
locate.change.miss(
  x,
  lambda,
  sample.splitting = FALSE,
  standardize.series = FALSE,
  view.cusum = FALSE
)
```

Arguments

`x`	A (p x n) data matrix of multivariate time series, each column represents a data point, missing values are represented by NAs.
`lambda`	Regularisation parameter. If no value is supplied, the dafault value is chosen to be sqrt(log(log(n)*p/2)) for p and n number of rows and columns of the data matrix x respectively.
`sample.splitting`	Whether the changepoint should be estimated via sample splitting. The theoretical result is proven only for the sample splitted version of the algorithm. However, the default setting in practice is without sample splitting.
`standardize.series`	Whether the given time series should be standardised before estimating the projection direction. Default is FALSE, i.e. the input series is assume to have variance 1 in each coordinate.
`view.cusum`	Whether to show a plot of the projected CUSUM series

Value

A list of three items:

- changepoint - A single integer value estimate of the changepoint location is returned. If the estimated changepoint is z, it means that the multivariate time series is piecewise constant up to z and from z+1 onwards.
- cusum - The maximum absolute CUSUM statistic of the projected univariate time series associated with the estimated changepoint.
- vector.proj - the vector of projection.

References

unpublished work

Examples

```
n <- 2000; p <- 1000; k <- 10; z <- 400; vartheta <- 0.8; sigma <- 1; shape <- 3
noise <- 0; corr <- 0; miss <- 1; proba <- 0.5; oncp <- 0
obj <- single.change(n,p,k,z,vartheta,sigma,shape,noise,corr,miss,proba,oncp)
x <- obj$x
locate.change.miss(x)
```

`multi.change` *Generating a high-dimensional time series with multiple changepoints and missingness*

Description

The data matrix is generated via X = (mu + W)o(Omega), where mu is the mean structure matrix that captures the changepoint locations and sparsity structure, W is a random noise matrix having independent N(0,sigmaˆ2) entries and Omega is the observation matrix.

Usage

```
multi.change(
  n,
  p,
  ks,
  zs,
  varthetas,
  sigma = 1,
  overlap = 0,
  shape = 3,
  miss = -1,
  proba = 0.5
)
```

Arguments

`n`	Time length of the observation
`p`	Dimension of the multivariate time series
`ks`	A vector describing the number of coordinates that undergo a change in each changepoint. If only a scalar is supplied, each changepoint will have the same number of coordinates that undergo a change.
`zs`	A vector describing the locations of the changepoints.
`varthetas`	A vector describing the root mean squared change magnitude in coordinates that undergo a change for each changepoint. If only a scalar is supplied, each changepoint will have the same signal strength value.
`sigma`	noise level
`overlap`	A number between 0 and 1. The proportion of overlap in the signal coordinates for successive changepoints.

shape	How the signal strength is distributed across signal coordinates. When shape = 0, all signal coordinates are changed by the same amount; when shape = 1, their signal strength are proportional to 1, sqrt(2), ..., sqrt(k); when shape = 2, they are proportional to 1, 2, ..., k; when shape = 3, they are proportional to 1, 1/sqrt(2), ..., 1/sqrt(k).
miss	Used to specify missingness structure. When miss = -1, there is no missingness, when miss = 0, the missingness is homogeneous with probability proba, when miss = 1, the missingness is homogeneous per column, with the missing probability on each column drawn from a uniform distribution on [proba/2, proba] when miss = 2, the missingness is homogeneous per row, with the missing probability on each row drawn from a uniform distribution on [proba/2, proba]. When miss = 3, the missingness of each input is the product of the missingness of the row by the missingness of the column, with both of these drawn from independant uniform distributions on [proba/2, proba]
proba	Used to specify missingness importance. See `miss` for more details.

Value

An S3 object of the class 'hdchangeseq' is returned.

- x - The generated data matrix
- mu - The mean structure of the data matrix
- omega - The observation matrix indicating where data is present (1) or missing (0)
- fullx - The generated data matrix without missing data

See Also

`plot.hdchangeseqmiss`

Examples

```
n <- 2000; p <- 200; ks <- 40;
zs <- c(500,1000,1500); varthetas <- c(0.1,0.15,0.2); overlap <- 0.5
sigma <- 0.2; shape <- 1; miss <- 0; proba <- 0.8
obj <- multi.change(n, p, ks, zs, varthetas, sigma, overlap, shape, miss, proba)
plot(obj, noise = TRUE)
```

PiS *Matrix projection onto the nuclear norm unit sphere*

Description

Projection (with respect to the inner product defined by the Frobenius norm) of a matrix onto the unit sphere defined by the nuclear norm.

Usage

```
PiS(M)
```

Arguments

M	Input matrix

Details

This is an auxiliary function used by the InspectChangepoint package. The projection is achieved by first performing a singular value decomposition, then projecting the vector of singular values onto the standard simplex, and finally using singular value decomposition in reverse to build the projected matrix.

Value

A matrix of the same dimension as the input is returned.

Examples

```
M <- matrix(rnorm(20),4,5)
PiS(M)
```

PiW *Projection onto the standard simplex*

Description

The input vector is projected onto the standard simplex, i.e. the set of vectors of the same length as the input vector with non-negative entries that sum to 1.

Usage

```
PiW(v)
```

Arguments

v	Input vector

Details

This is an auxiliary function used by the InspectChangepoint package.

Value

A vector in the standard simplex that is closest to the input vector is returned.

References

Chen, Y. and Ye, X. (2011) Projection onto a simplex. arXiv preprint, arxiv:1101.6081.

Examples

```
v <- rnorm(10)
PiW(v)
```

plot.hdchangeseqmiss *Plot function for 'hdchangeseqmiss' class*

Description

Visualising the high-dimensional time series in an 'hdchangeseqmiss' class object. The data matrix or its mean structure is visualised using a grid of coloured rectangles with colours corresponding to the value contained in corresponding coordinates. A heat-spectrum (red to white for values low to high) is used to convert values to colours.

Usage

```
## S3 method for class 'hdchangeseqmiss'
plot(x, noise = TRUE, shuffle = FALSE, ...)
```

Arguments

x	An object of 'hdchangeseqmiss' class
noise	If noise == TRUE, the data matrix is plotted, otherwise, only the mean structure is plotted.
shuffle	Whether to shuffle the rows of the plotted matrix.
...	Other graphical parameters are not used.

Examples

```
n <- 2000; p <- 100; k <- 10; z <- 800; vartheta <- 1; sigma <- 1
shape <- 3; noise <- 0; corr <- 0; miss <- 2; proba <- 0.5; oncp <- 0
obj <- single.change(n,p,k,z,vartheta,sigma, shape, noise, corr, miss, proba, oncp)
plot(obj, noise = TRUE)
```

plot.inspect *Plot function for 'inspect' class objects*

Description

Plot function for 'inspect' class objects

Usage

```
## S3 method for class 'inspect'
plot(x, ...)
```

Arguments

x	an 'inspect' class object
...	other arguments to be passed to methods are not used

See Also

inspect

`plot.inspect.miss` *Plot function for 'inspect.miss' class objects*

Description

Plot function for 'inspect.miss' class objects

Usage

```
## S3 method for class 'inspect.miss'
plot(x, ...)
```

Arguments

`x`	an 'inspect.miss' class object
`...`	other arguments to be passed to methods are not used

See Also

`inspect`

`power.method` *Power method for finding the leading eigenvector of a symmetric matrix*

Description

Power method for finding the leading eigenvector of a symmetric matrix

Usage

```
power.method(A, eps = 1e-10, maxiter = 10000)
```

Arguments

`A`	a square symmetric matrix
`eps`	tolerance for convergence (in Frobenius norm)
`maxiter`	maximum iteration

Value

a unit-length leading eigenvector of A

print.inspect *Print function for 'inspect' class objects*

Description

Print function for 'inspect' class objects

Usage

```
## S3 method for class 'inspect'
print(x, ...)
```

Arguments

x	an 'inspect' class object
...	other arguments to be passed to methods are not used

See Also

inspect

print.inspect.miss *Print function for 'inspect.miss' class objects*

Description

Print function for 'inspect.miss' class objects

Usage

```
## S3 method for class 'inspect.miss'
print(x, ...)
```

Arguments

x	an 'inspect.miss' class object
...	other arguments to be passed to methods are not used

See Also

inspect

rescale.variance *Noise standardisation for multivariate time series.*

Description

Each row of the input matrix is normalised by the estimated standard deviation computed through the median absolute deviation of increments.

Usage

```
rescale.variance(x)
```

Arguments

x	An input matrix of real values.

Details

This is an auxiliary function used by the `InspectChangepoint` package.

Value

A rescaled matrix of the same size is returned.

Examples

```
x <- matrix(rnorm(40),5,8) * (1:5)
x.rescaled <- rescale.variance(x)
x.rescaled
```

rescale.variance.miss *Noise standardisation for multivariate time series with missingness.*

Description

Each row of the input matrix is normalised by the estimated standard deviation computed through the median absolute deviation of increments.

Usage

```
rescale.variance.miss(x)
```

Arguments

x	An input matrix of real values with missing data.

Details

This is an auxiliary function used by the `InspectChangepointMissingness` package.

Value

A rescaled matrix of the same size is returned.

Examples

```
x <- matrix(rnorm(40),5,8) * (1:5)
omega = matrix(rbinom(40,1,0.5),5,8)
x <- x*omega
x.rescaled <- rescale.variance.miss(x)
x.rescaled
```

`single.change` *Generating high-dimensional time series with exactly one change in the mean structure and with different type of missingness*

Description

The data matrix is generated via X = (mu + W)o(Omega), where mu is the mean structure matrix that captures the changepoint location and sparsity structure, W is a random noise matrix and Omega is the observation matrix

Usage

```
single.change(
  n,
  p,
  k,
  z,
  vartheta,
  sigma = 1,
  shape = 3,
  noise = 0,
  corr = 0,
  miss = -1,
  proba = 0.5,
  oncp = 0
)
```

Arguments

`n`	Time length of the observation
`p`	Dimension of the multivariate time series
`k`	Number of coordinates that undergo a change
`z`	Changepoint location, a number between 1 and n-1.
`vartheta`	The root mean squared change magnitude in coordinates that undergo a change
`sigma`	noise level, see `noise` for more details.

shape — How the signal strength is distributed across signal coordinates. When shape = 0, all signal coordinates are changed by the same amount; when shape = 1, their signal strength are proportional to 1, sqrt(2), ..., sqrt(k); when shape = 2, they are proportional to 1, 2, ..., k; when shape = 3, they are proportional to 1, 1/sqrt(2), ..., 1/sqrt(k).

noise — Noise structure of the multivarite time series. For noise = 0, 0.5, 1, columns of W have independent multivariate normal distribution with covariance matrix Sigma. When noise = 0, Sigma = sigmaˆ2 * I_p; when noise = 0.5, noise has local dependence structure given by Sigma_i,j = sigma*corrˆ—i-j—; when noise = 1, noise has global dependence structure given by matrix(corr,p,p)+diag(p)*(1-corr))) * sigma. When noise = 2, rows of the W are independent and each having an AR(1) structure given by W_j,t = W_j,t-1 * sqrt(corr) + rnorm(sd = sigma) * sqrt(1-corr). For noise = 3, 4, entries of W have i.i.d. uniform distribution and exponential distribution respectively, each centred and rescaled to have zero mean and variance sigmaˆ2.

corr — Used to specify correlation structure in the noise. See noise for more details.

miss — Used to specify missingness structure. When miss = -1, there is no missingness, when miss = 0, the missingness is homogeneous with probability proba, when miss = 1, the missingness is homogeneous per column, with the missing probability on each column drawn from a uniform distribution on [proba/2, proba] when miss = 2, the missingness is homogeneous per row, with the missing probability on each row drawn from a uniform distribution on [proba/2, proba]. When miss = 3, the missingness of each input is the product of the missingness of the row by the missingness of the column, with both of these drawn from independant uniform distributions on [proba/2, proba] Some modification on the cases miss = 1, 2, 3 is induced by the value of oncp.

proba — Used to specify missingness importance. See miss for more details.

oncp — Used to specify missingness location in the cases miss = 1, miss = 2. If oncp = 0, nothing changes. If oncp = 1, missingness is stronger on the coordinates of the changepoint (miss = 2) or on the columns around the changepoint (miss = 1), or both (miss = 3). This is achieved by drawing from a uniform distribution on [proba, 1] instead of [proba/2, proba]. If oncp = -1, the missingness is weaker in the same places, by drawing from [0, proba/2] instead.

Value

An S3 object of the class 'hdchangeseqmiss' is returned.

- x - The generated data matrix with the missing data (NAs)
- mu - The mean structure of the data matrix without missing data
- omega - The observation matrix indicating where data is present (1) or missing (0)
- fullx - The generated data matrix without missing data

See Also

plot.hdchangeseqmiss

Examples

```
n <- 2000; p <- 100; k <- 10; z <- 800; vartheta <- 1; sigma <- 1
shape <- 3; noise <- 0; corr <- 0; miss <- 2; proba <- 0.8; oncp = -1
obj <- single.change(n,p,k,z,vartheta, sigma, shape, noise, corr, miss, proba, oncp)
plot(obj, noise = TRUE)
```

`sparse.svd` *Computing the sparse leading left singular vector of a matrix*

Description

Estimating the sparse left leading singular vector by first computing a maximiser Mhat of the convex problem

$$< Z, M > -\lambda |M|_1$$

subject to the Schatten norm constraint —M—_schatten ¡= 1 using alternating direction method of multipliers (ADMM). Then the leading left singular vector of Mhat is returned.

Usage

```
sparse.svd(Z, lambda, schatten = c(1, 2), tolerance = 1e-05, max.iter = 10000)
```

Arguments

`Z`	Input matrix whose left leading singular vector is to be estimated.
`lambda`	Regularisation parameter
`schatten`	Schatten norm constraint to be used. Default uses Schatten-2-norm, i.e. the Frobenius norm. Also possible to use Schatten-1-norm, the nuclear norm.
`tolerance`	Tolerance criterion for convergence of the ADMM algorithm. Not used when shatten=2.
`max.iter`	Maximum number of iteration in the ADMM algorithm. Not used when shatten=2.

Details

In case of schatten = 2, a closed-form solution for Mhat using matrix soft thresholding is possible. We use the closed-form solution instead of the ADMM algorithm to speed up the computation.

Value

A vector that has the same length as nrow(Z) is returned.

Examples

```
Z <- matrix(rnorm(20),4,5)
lambda <- 0.5
sparse.svd(Z, lambda)
```

`sparse.svd.miss` *Computing the sparse leading left singular vector of a matrix used in* `locate.change.miss`

Description

Estimating the sparse left leading singular vector by solving the following convex optimization problem:

$$u, w = argmax < Z, uw^T > -\lambda * sqrt(n) * ||u||_1 underconstraints ||u||_2 = ||w||_2 = 1$$

Usage

```
sparse.svd.miss(Z, lambda, tolerance = 1e-15, max.iter = 10000)
```

Arguments

`Z`	Input matrix whose left leading singular vector is to be estimated.
`lambda`	Regularisation parameter
`tolerance`	Tolerance criterion for convergence of the optimization algorithm.
`max.iter`	Maximum number of iteration in the optimization algorithm.

Value

A vector that has the same length as nrow(Z) is returned.

Examples

```
Z <- matrix(rnorm(20),4,5)
lambda <- 0.5
sparse.svd.miss(Z, lambda)
```

`summary.inspect` *Summary function for 'inspect' class objects*

Description

Summary function for 'inspect' class objects

Usage

```
## S3 method for class 'inspect'
summary(object, ...)
```

Arguments

`object`	an 'inspect' class object
`...`	other arguments to be passed to methods are not used

See Also

`inspect`

summary.inspect.miss *Summary function for 'inspect.miss' class objects*

Description

Summary function for 'inspect.miss' class objects

Usage

```
## S3 method for class 'inspect.miss'
summary(object, ...)
```

Arguments

object	an 'inspect.miss' class object
...	other arguments to be passed to methods are not used

See Also

inspect

vector.norm *Norm of a vector*

Description

Calculate the entrywise L_q norm of a vector or a matrix

Usage

```
vector.norm(v, q = 2, na.rm = FALSE)
```

Arguments

v	a vector of real numbers
q	a nonnegative real number or Inf
na.rm	boolean, whether to remove NA before calculation

Value

the entrywise L_q norm of a vector or a matrix

`vector.soft.thresh` *Soft thresholding a vector*

Description

entries of v are moved towards 0 by the amount lambda until they hit 0.

Usage

```
vector.soft.thresh(x, lambda)
```

Arguments

`x`	a vector of real numbers
`lambda`	soft thresholding value

Value

a vector of the same length

Index