Submitted to the Annals of Statistics
arXiv: arXiv:1312.1900

HIGH-DIMENSIONAL CHANGE-POINT DETECTION UNDER SPARSE ALTERNATIVES

By Farida Enikeeva[*,†,§] and Zaid Harchaoui[*,‡]

Université de Poitiers[†], *University of Washington*[‡], *and IITP*[§]

We consider the problem of detecting a change in mean in a sequence of high-dimensional Gaussian vectors. The change in mean may be occurring simultaneously in an unknown subset components. We propose a hypothesis test to detect the presence of a change-point and establish the detection boundary in different regimes under the assumption that the dimension tends to infinity and the length of the sequence grows with the dimension. A remarkable feature of the proposed test is that it does not require any knowledge of the subset of components in which the change in mean is occurring and yet automatically adapts to yield optimal rates of convergence over a wide range of statistical regimes.

1. Introduction.

Consider a sequence of n independent d-dimensional Gaussian vectors $\mathbb{X}^n = (X_1, \dots, X_n)$ with a possible change in mean at an unknown location $\tau \in \mathcal{T}$,

$$X_i = \theta + \Delta\theta_\tau \mathbf{1}\{i > \tau\} + \xi_i, \quad i = 1, \dots, n \tag{1}$$

where $(\xi_i)_{1 \le i \le n}$ are i.i.d. random vectors drawn from $\mathcal{N}(0, I_d)$, $\Delta\theta_\tau \in \mathbb{R}^d$, $\theta \in \mathbb{R}^d$, and $\mathcal{T} = \{1, \dots, n-1\}$ is the set of possible change-point locations. Our goal is to propose a hypothesis test for the change-point problem

$$\mathrm{H}_0: \quad \Delta\theta_\tau = \mathbf{0} \quad \text{against} \quad \mathrm{H}_\mathrm{A}: \quad \exists\, \tau \in \mathcal{T} \text{ such that } \Delta\theta_\tau \neq \mathbf{0}.$$

Under the null hypothesis H_0, there is no change in mean, that is $\Delta\theta_\tau = \mathbf{0}$. Under the alternative H_A, a change in mean occurs at a location $\tau \in \mathcal{T}$, that is $\Delta\theta_\tau \neq \mathbf{0}$. The change occurs in exactly p coordinates of the mean vector θ that correspond to the support $\mathrm{supp}(\Delta\theta_\tau)$ of the jump vector $\Delta\theta_\tau \in \mathbb{R}^d$. Neither the change-point τ, nor the set $\mathrm{supp}(\Delta\theta_\tau)$ nor its dimension $p \in \{1, \dots, d\}$ are known. We will test the hypothesis of no change in mean against this composite alternative.

[*]This work was supported by the project Gargantua (CNRS-Mastodons), the LabEx PERSYVAL-Lab (ANR-11-LABX-0025), the project Macaron (ANR-14-CE23-0003-01), the NSF TRIPODS award (CCF-1740551), and the OPTIMYST-II project of Minalogic.

AMS 2000 subject classifications: Primary 62G10, 62H15; secondary 60G35

Keywords and phrases: change-point problem, high-dimensional data, minimax optimality, sparsity

Change-point problems with multivariate Gaussian observations have received a lot of attention for decades. The usual setting assumes that the change occurs in all components. Then the problem is studied in a classical asymptotic regime, that is by letting the number of observations n grow to infinity while the dimension of the vector d remains fixed. We refer the reader to, e.g. [27, 4, 10, 7, 11, 20] for a review.

We consider the problem in a high-dimensional double-asymptotic setting, where both the number of observations and their dimension grow to infinity. Such a setting is particularly appropriate for recent instances of change-point problems that arise in real-world applications where observations are typically high-dimensional – for example, in biostatistics [30, 32], in network traffic data analysis [23], in multimedia indexation [14], and in astro-statistics [6, 24]. In all these applications, one is interested in detecting change-points in relatively short sequences of observations (say, $n = 100$) whose dimension can be high (say, $d = 10^3$). However, in these applications, prior information suggests that the change is likely to occur *only in a small subset of components*. Therefore, the effective dimension of the change-point problem is actually the dimension of $\mathrm{supp}(\Delta\theta_\tau)$, instead of d. Thus the statistical problem is potentially tractable even for observations living in a high-dimensional ambient space [15].

Korostelev and Lepski [21] studied high-dimensional change-point problems in the white noise framework. The change is assumed to occur simultaneously in all components, that is $p = d$. The authors propose an asymptotically minimax estimator of the change-point location, under double-asymptotics with the Euclidean norm $\|\Delta\theta\|_2 \to \infty$ as $d \to \infty$. See also [28] for an early treatment of a related problem. Recently, Xie and Siegmund [31] considered a problem similar to ours from a methodological point of view. In [31], a Bayes-type test statistic is proposed, where the authors introduce a mixture model that hypothesizes an assumed fraction of changing components.

In this work, we establish the detection boundary for the change-point problem under sparse alternatives. The detection boundary is an asymptotic condition on the norm of $\Delta\theta_\tau$ defining the minimax separability of the hypotheses H_0 and H_A. It depends on the location of the change-point and on the number of changing components p. The proposed test is based on two test statistics: a *linear statistic*, that considers all components simultaneously, and a *scan statistic* that searches for a change over all possible combinations of changing components. Although the latter problem has a combinatorial structure that seems challenging at first sight, we show that the scan test statistic can actually be computed efficiently, in almost linear time with respect to the dimension of the problem. We derive the minimax separation rate of our test, prove that it is adaptive to the unknown set of changing components, and establish that it is rate-optimal in the high sparsity regime.

2. Statement of the problem. We first consider the problem of testing the hypothesis of no change against the alternative of a change in mean at a given location τ in exactly p components. We will later describe the test that is adaptive to the case of unknown p and τ.

We use the following notation throughout the text: $\mathcal{M}(d,p)$ stands for the collection of all subsets of $\{1,\dots,d\}$ of cardinality p, and $\mathcal{M}$ stands for the set of all possible subsets of $\{1,\dots,d\}$. We denote by $\Pi_m v$ the projection of a vector $v \in \mathbb{R}^d$ onto a subspace indexed by $m \in \mathcal{M}$. The location of a change is parametrized by $t \in \mathcal{T}$, where $\mathcal{T} = \{1,\dots,n-1\}$. For two sequences x_d and y_d we write $x_d \asymp y_d$ if $x_d/y_d \to c \in \mathbb{R}\setminus\{0\}$ as $d \to \infty$.

Both the null and the alternative hypotheses can be simply formulated in terms of the norm of the jumps of the mean vector $\Delta\theta_\tau$, so that we have no change if the norm of the jumps is zero, $\|\Delta\theta_\tau\| = 0$. In what follows $\|\cdot\|$ denotes the Euclidean norm. We say that the change occurs at location τ, if the norm of the jumps at location τ satisfies $\|\Delta\theta_\tau\| \geq r$ for some $r > 0$.

Define the set $\Theta_p[r] = \{v \in V_p^d : \ \|v\| \geq r\}$, where V_p^d is the subspace of $\mathbb{R}^d$-vectors with exactly p non-zero components

$$V_p^d = \{v = (\varepsilon_1 v_1, \dots, \varepsilon_d v_d) : \ v_j \in \mathbb{R},\ \varepsilon_j \in \{0,1\},\ \sum_{j=1}^d \varepsilon_j = p\}.$$

Recall the notation $\mathcal{T} = \{1,\dots,n-1\}$ and denote by $\mathcal{D}$ the set $\{1,\dots,d\}$. We consider two general problems based on model (1).

(P1) Testing a change in exactly $p \in \mathcal{D}$ unknown components of the mean at a given location $\tau \in \mathcal{T}$:

$$\mathrm{H}_0 : \ \Delta\theta_\tau = 0 \quad \text{against} \quad \mathrm{H}_\mathrm{A} : \ \Delta\theta_\tau \in \Theta_p[r],$$

where $r > 0$ may depend on τ, p, d, and n.

(P2) Testing the presence of a change in an unknown number of components $p \in \mathcal{D}$ that occurs within the time interval $\mathcal{T} = \{1,\dots,n-1\}$:

$$\mathrm{H}_0 : \ \Delta\theta_\tau = 0 \quad \text{against} \quad \mathrm{H}_\mathrm{A} : \ \exists\, \tau \in \mathcal{T} : \text{ such that } \Delta\theta_\tau \in \Theta[r],$$

where $\Theta[r] = \bigcup_{p\in\mathcal{D}} \Theta_p[r]$. Note that r might depend on τ, p, d and n as well, $r = r(\tau,p,d,n)$. To keep the notation as light as possible, we omit the explicit dependence on these parameters and simply write r.

Note that Problem (P1) corresponds to a two-sample testing problem [22], with a difference in mean lying on a subset of changing components and equal variance; see [9] for a review of recent works on this topic.

A test $\psi = \psi(\mathbb{X}^n)$ is a measurable function of observations from model (1) taking values in $\{0, 1\}$. We say that there is a change if $\psi(\mathbb{X}^n) = 1$.

Let $\mathbb{P}_{0,\theta}$ be the measure corresponding to the null hypothesis case of no change in mean and $\mathbb{P}_{\Delta\theta_\tau,\theta}$ be the measure corresponding to the case of a $\Delta\theta_\tau$ change in mean at location τ. The parameter $\theta \in \mathbb{R}^d$ is a nuisance parameter and we will see that the errors of our test do not depend on it. For any test ψ, define the type I error as

$$\alpha(\psi) = \sup_{\theta \in \mathbb{R}^d} \mathbb{P}_{0,\theta}\{\psi = 1\}.$$

The type II errors for Problems (P1) and (P2) are respectively defined as

$$\beta(\psi, \Theta_p[r], \tau) = \sup_{(\theta, \Delta\theta_\tau) \in \mathbb{R}^d \times \Theta_p[r]} \mathbb{P}_{\Delta\theta_\tau,\theta}\{\psi = 0\}$$

and

$$\beta^*(\psi, \Theta[r]) = \sup_{\tau \in \mathcal{T}} \sup_{(\theta, \Delta\theta_\tau) \in \mathbb{R}^d \times \Theta[r]} \mathbb{P}_{\Delta\theta_\tau,\theta}\{\psi = 0\} \equiv \sup_{\tau \in \mathcal{T}} \sup_{p \in \mathcal{D}} \beta(\psi, \Theta_p[r], \tau).$$

Define the global testing errors [19] for the two problems:

$$\gamma(\psi, \Theta_p[r], \tau) := \alpha(\psi) + \beta(\psi, \Theta_p[r], \tau),$$
$$\gamma^*(\psi, \Theta[r]) := \alpha(\psi) + \beta^*(\psi, \Theta[r]).$$

We could also define the global testing error as a linear combination of the two errors $\gamma(\psi, \Theta_p[r], \tau) = \alpha(\psi) + s\beta(\psi, \Theta_p[r], \tau)$ with $s > 0$. The results of the paper will still hold in the case where s does not depend on the dimension d. Choosing the error weights to be dependent on the sparsity index p might be useful in applications. However, the thresholds for the tests will be different from those ones proposed in the present paper.

We make the following Assumptions (A1-A2-A3) on the asymptotic behavior of n, d, p, and τ. *Throughout the paper, the asymptotics of p and n are parametrized by d, where $d \to \infty$.* The asymptotic of the location τ is naturally parametrized by n, which, in turn, depends on d.

(A1) The number of observations $n > 1$ can be fixed or grow with the vector dimension, $n = n(d) \to \infty$ as $d \to \infty$.

(A2) The number of components with a change is sufficiently large

$$p \to \infty \quad \text{and} \quad p/d \to 0 \quad \text{as} \quad d \to \infty.$$

(A3) For Problem (P2), we need an additional assumption, namely that

$$\sup_{p\in\mathcal{D}} \frac{\log n}{p\log(d/p)} \to 0, \quad d\to\infty.$$

Thus, we have $\log(np)/\log\binom{d}{p} \to 0$ and $\log n/p \to 0$ as $d\to\infty$ if (A2) holds. Therefore the number of observations n cannot be too large.

We consider an asymptotic setting where both the number of observations n and the dimension d are growing. Therefore the rates we shall establish depend also on the dimension d. We choose here to parameterize the quantities involved in the asymptotics with respect to d in order to model real-world problems where the dimension of the observations can be large. For example, in the problem of simultaneous segmentation of several genomic profiles [30], the length of a profile n can be of order $n\approx 10^2 - 10^3$ nucleotides, while the number of profiles d can be of order $d\approx 10^4$.

REMARK 1. *If p depends on d via a sparsity coefficient $\beta\in(0,1)$, $p\asymp d^{1-\beta}$, then Assumption (A2) is satisfied. We shall distinguish between the cases of high sparsity, $\beta\in(1/2,1)$ and low sparsity, $\beta\in(0,1/2]$. Later we shall see why $\beta=1/2$ defines a boundary between two sparsity regimes (see Remark 6 after Theorem 2).*

We are interested in the minimax separation conditions for problems (P1) and (P2). The question is how far from the origin the sets $\Theta_p[r]$ and $\Theta[r]$ should be in order to separate the hypotheses H_0 and H_A in problems (P1) and (P2), respectively. For any $\gamma\in(0,1)$, the sequence r_d is called a *minimax separation rate* [16, 20] for problem (P1) if

(i) there exists a constant $C^*>0$ and a test ψ^* such that $\forall C>C^*$

$$\limsup_{d\to\infty} \gamma(\psi^*,\Theta_p[Cr_d],\tau)\le\gamma;$$

(ii) there exists a constant $C_*>0$ such that $\forall C<C_*$ and for any test ψ

$$\liminf_{d\to\infty} \gamma(\psi,\Theta_p[Cr_d],\tau)\ge\gamma.$$

Conditions (i)-(ii) are respectively *upper and lower bounds* on the minimax testing error; we refer to [19, 20, 3, 12] for further discussion about these definitions. A testing procedure ψ^* is *minimax rate-optimal* if conditions (i) and (ii) hold. Note that the constants and the rate-optimal test may depend on the given overall significance level γ. We can similarly define the detection boundary conditions for problem (P2).

By abuse of notation we shall always denote by γ both the global testing error and the overall significance level of the test.

3. Testing procedure. Let us first define the following d-dimensional process describing the change in mean at time $t \in \mathcal{T}$,

$$Z_n(t) = \sqrt{\frac{t(n-t)}{n}} \left(\frac{1}{t} \sum_{i=1}^{t} X_i - \frac{1}{n-t} \sum_{i=t+1}^{n} X_i \right), \quad t \in \mathcal{T}. \tag{2}$$

The tests we propose for problems (P1)–(P2) are based on two χ^2-type test statistics, which we shall refer to as *linear statistic* and *scan statistic*.

The linear statistic is given by

$$L_{\mathrm{lin}}(t) = \frac{\|Z_n(t)\|^2 - d}{\sqrt{2d}}, \quad t \in \mathcal{T}. \tag{3}$$

For each fixed $p \in \mathcal{D}$ the scan statistic is defined as

$$L^p_{\mathrm{scan}}(t) = \max_{m \in \mathcal{M}(d,p)} \left\{ \frac{\|\Pi_m Z_n(t)\|^2 - p}{\sqrt{2p}} \right\}, \quad t \in \mathcal{T}. \tag{4}$$

For any fixed t, the components of the process $Z_n(t)$ are standard Gaussians. Thus under the null hypothesis, the statistic $\|\Pi_m Z_n(t)\|^2$, $m \in \mathcal{M}(d,p)$ has a χ^2_p distribution with mean p and variance $2p$. To make the test statistics in (3) and (4) defined at the same scale, we divide the squared norm of the process $Z_n(t)$ by its standard deviation under the null hypothesis. Note that the normalization is critical in a high-dimensional setting.

REMARK 2. *Assume that $\Delta\theta_\tau \in V^d_p$ and let $m \in \mathcal{M}(d,p)$ be a given subset of p components with a change at the location τ. The choice of the test statistics based on $Z_n(t)$ is motivated by the generalized likelihood ratio test. We have*

$$\log \frac{\max\limits_{(\theta,\Delta\theta_\tau) \in \mathbb{R}^d \times \Pi_m \mathbb{R}^d} \mathcal{L}(\theta, \Delta\theta_\tau; \mathbb{X}^n)}{\max\limits_{\theta \in \mathbb{R}^d} \mathcal{L}(\theta, 0; \mathbb{X}^n)} = \frac{1}{2} \|\Pi_m Z_n(\tau)\|^2, \tag{5}$$

where $\mathcal{L}(\theta, \Delta\theta_\tau; \mathbb{X}^n)$ is the likelihood of the parameters θ and $\Delta\theta_\tau$ given the observations $\mathbb{X}^n$.

3.1. *Known number of components with a change and known τ.* The proposed decision rule for Problem (P1) is based on the combination of two tests

$$\psi^*_p = \psi_{\mathrm{lin}} \vee \psi^p_{\mathrm{scan}},$$

with

$$\psi_{\mathrm{lin}} = \mathbf{1}\big\{L_{\mathrm{lin}}(\tau) > H\big\}, \quad \psi^p_{\mathrm{scan}} = \mathbf{1}\big\{L^p_{\mathrm{scan}}(\tau) > T_p\big\}.$$

The thresholds H and T_p should be set in such a way that the global risk error $\gamma(\psi_p^*, \Theta_p[r], \tau)$ is asymptotically less than a given global significance level γ as $d \to \infty$. Theorem 1 answers this question, and provides a guideline to set the corresponding thresholds depending on d (and henceforth on p and n). For any significance level α we could set the thresholds for the two tests using the quantiles of the corresponding χ^2 distributions:

$$H = \frac{q_{\chi^2_d}(1-\alpha_l) - d}{\sqrt{2d}}, \quad T_p = \frac{q_{\chi^2_p}\left(1-\alpha_s \Big/ \binom{d}{p}\right) - p}{\sqrt{2p}},$$

where $\alpha_l + \alpha_s = \alpha$, $\alpha_l, \alpha_s > 0$. However, such a strategy results in a computational burden for the scan test, since the quantiles of such a high order could be difficult to compute precisely even for moderate values of d. We propose instead the following formulas for the thresholds

$$H = \sqrt{2\log\frac{1}{\alpha_l}} + \sqrt{\frac{2}{d}}\log\frac{1}{\alpha_l}, \tag{6}$$

$$T_p = \sqrt{\frac{2}{p}}\log\left[\binom{d}{p}\frac{1}{\alpha_s}\right] + \sqrt{2\log\left[\binom{d}{p}\frac{1}{\alpha_s}\right]}, \tag{7}$$

obtained using concentration inequalities for the norm of a d-dimensional Gaussian vector [5].

REMARK 3. *The scan statistic ψ_{scan}^p and the linear test statistic ψ_{lin} behave differently depending on the sparsity level. In the case of high sparsity, $\beta \in (1/2, 1)$, the scan test outperforms the linear test, since the scan statistic searches for a change over all possible subsets of components. Basically, as we will see later, the scan test can detect a change with much smaller magnitude than the linear test. In the case of high sparsity, its minimax separation rate is faster than the one of the linear test.*

However, in the case of moderate sparsity, $\beta \in (0, 1/2]$, the linear one has a faster detection rate. In other words, averaging statistical information across all vector components is more effective than a full search over all possible combinations of components. Therefore, the proposed test statistic ψ_p^* gets the "best of both worlds", performing well in both the moderate sparsity and the high sparsity regimes. The case of no sparsity $p = d$ is also covered by these two tests that become identical in this case.

3.2. *Adaptation to unknown number of components with a change.* For Problem (P2), the proposed adaptive decision rule is, again, based on the combination of two tests

$$\psi^* = \psi_{\text{lin}}^* \vee \psi_{\text{scan}}^*.$$

Here, the linear test $\psi^*_{\rm lin}$ maximizes the test statistic with respect to all possible locations of the change-point

$$\psi^*_{\rm lin} = \mathbf{1}\left\{\max_{t\in\mathcal{T}} L_{\rm lin}(t) > H^*\right\}.$$

The corresponding threshold H^* providing the significance level α_l is now given by

$$H^* = \sqrt{\log d} + (1+\varepsilon)\sqrt{2\log\frac{1}{\alpha_\varepsilon}} + (1+\varepsilon)\sqrt{\frac{2}{d}}\log\frac{1}{\alpha_\varepsilon}, \tag{8}$$

where

$$\alpha_\varepsilon = \alpha_l\frac{\log(1+\varepsilon)}{\log n}, \quad \varepsilon = \sqrt{\frac{2\log d}{d}}. \tag{9}$$

For a given significance level α_s, the scan statistic is now defined as

$$\begin{aligned} L_{\rm scan}(\mathcal{T}) &:= \max_{p\in\mathcal{D}}\frac{1}{T_{p,n}}\max_{t\in\mathcal{T}} L^p_{\rm scan}(t) \\ &= \max_{p\in\mathcal{D}}\frac{1}{T_{p,n}}\max_{t\in\mathcal{T}}\max_{m\in\mathcal{M}(d,p)}\left\{\frac{\|\Pi_m Z_n(t)\|^2 - p}{\sqrt{2p}}\right\}, \end{aligned} \tag{10}$$

where

$$T_{p,n} = \sqrt{\frac{2}{p}}\log\left[\binom{d}{p}\frac{2np^2}{\alpha_s}\right] + \sqrt{2\log\left[\binom{d}{p}\frac{2np^2}{\alpha_s}\right]}. \tag{11}$$

Note that the threshold $T_{p,n}$ is built in the scan test statistic. In order to make the test adaptive to unknown p, we have to compare the values of the statistic $\max_{t\in\mathcal{T}} L^p_{\rm scan}(t)$ with $T_{p,n}$ for all $p\in\mathcal{D}$. The threshold $T_{p,n}$ penalizes the number of combinations of indexes in $\mathcal{M}(d,p)$ as well as the length of the sequence. The adaptive scan test is defined as

$$\psi^*_{\rm scan} = \mathbf{1}\left\{\max_{p\in\mathcal{D}}\frac{1}{T_{p,n}}\max_{t\in\mathcal{T}} L^p_{\rm scan}(t) > 1\right\}.$$

Again, for any overall significance level $\alpha = \alpha_l + \alpha_s$, we derive the thresholds (8) and (11) for the two tests using concentration inequalities [5]; see the proof of Lemma 3 for details.

REMARK 4. *At first sight, the scan statistic may seem difficult to compute, since it involves a combinatorial search which could be computationally* *hard. Recent work [1] considered several general classes of problems where scan statistics are computationally hard to compute. However, in our problem, it turns out that the scan statistic can be efficiently computed in almost linear-time with respect to the dimension of the problem. Indeed, we have*

$$\max_{m\in\mathcal{M}(d,p)} \frac{\|\Pi_m Z_n(t)\|^2 - p}{\sqrt{2p}} = \frac{1}{\sqrt{2p}}\left(\sum_{j=1}^{p}[Z_n^{(j)}(t)]^2 - p\right),$$

where $[Z_n^{(j)}(t)]^2$ *are the ordered squared components of the vector* $Z_n(t)$*:*

$$[Z_n^{(1)}(t)]^2 > [Z_n^{(2)}(t)]^2 > \cdots > [Z_n^{(d)}(t)]^2.$$

Thus the computational complexity of the adaptive test statistic is $O(nd\log d)$*. Computation relies on sorting the squared components of vectors* $Z_n(t)$ *for each* $t \in \mathcal{T}$*.*

REMARK 5. *Again, the proposed adaptive test statistic* ψ^* *covers moderate and high sparsity cases. This is reflected by our theoretical results in Theorems 3–4, which show that the proposed adaptive test statistic is minimax-optimal and rate-adaptive with only a logarithmic loss. The simulations in Section 5 corroborate our theoretical results.*

4. Main results.

4.1. *Upper and lower bounds on minimax testing error in problem (P1).* We shall first derive the rate of testing r_d for the test ψ_p^*. In the high sparsity case, we calculate the testing risk constant of our test (detection boundary condition (i)). Next, we shall prove that for the same minimax separation rate r_d yet with a different constant the detection boundary condition (ii) holds. The minimax separation rate for Problems (P1) and (P2) depends on the location of the change-point via the function

$$h(\tau) = \frac{\tau}{n}\left(1 - \frac{\tau}{n}\right), \tag{12}$$

which commonly arises in change-point problems; see [11] for details.

The following theorem gives the upper bound for the test ψ_p^* in problem (P1). The upper boundary conditions (13) and (14) correspond, respectively, to the linear test ψ_{lin} and to the scan test ψ_{scan}^p.

THEOREM 1. *Let* $\gamma \in (0,1)$*. Assume that Assumptions (A1)–(A2) hold and* $r = r(d)$ *satisfies either*

$$\liminf_{d\to\infty} \frac{r^2 n h(\tau)}{\sqrt{d}} \geq 4(\log[2/\gamma])^{1/2} \tag{13}$$

or

$$\liminf_{d\to\infty} \frac{r^2 nh(\tau)}{p\log(d/p)} \geq 2. \tag{14}$$

Let $\alpha \in (0,\gamma)$ be a given significance level, $\alpha_l = \gamma/2$ and $\alpha_s \in (0,\gamma/2)$ be such that $\alpha_l + \alpha_s = \alpha$. Let ψ_p^* be a test with H and T_p be defined as in (6)–(7). Then $\alpha(\psi_p^*) \leq \alpha$ and $\limsup_{d\to\infty} \beta(\psi_p^*, \Theta_p[r], \tau) \leq \gamma - \alpha$. Moreover, for the test ψ_p^* we have $\limsup_{d\to\infty} \gamma(\psi_p^*, \Theta_p[r], \tau) \leq \gamma$.

The following theorem establishes the lower bound on the minimax testing error in problem (P1) and the minimax separation rates that provide the separability conditions between the hypotheses.

THEOREM 2. *Assume that $p \asymp d^{1-\beta}$ as $d \to \infty$ and Assumption (A1) holds. For any $\gamma \in (0,1)$ and for any test ψ*

$$\liminf_{d\to\infty} \gamma(\psi, \Theta_p[r], \tau) \geq \gamma$$

if $r = r(d)$ satisfies one of the following conditions:

$$\limsup_{d\to\infty} \frac{r^2 nh(\tau)}{\sqrt{d}} \leq \sqrt{2\log(1+4(1-\gamma)^2)} \qquad \text{for } \beta \in [0,1/2), \tag{15}$$

$$\limsup_{d\to\infty} \frac{r^2 nh(\tau)}{p\log(d/p)} < 2 - \frac{1}{\beta} \qquad \text{for } \beta \in (1/2,1). \tag{16}$$

Note that the result of Theorem 2 is valid even for the non-sparsity case of $p = d$ components with a change. Since both scan and linear tests coincide in this case, Theorem 1 will be valid under assumption (13) only. We can find the corresponding constant when $\beta = 1/2$ as well; see Section 4.3 and the proof of Lemma 5 in Supplement A.

REMARK 6. *Theorems 1 and 2 give the detection boundaries for two sparsity regimes. In the case of moderate sparsity when $\beta \in (0,1/2)$ (and in the no-sparsity case $p = d$) it is the linear test that can detect a change if (13) holds. On the other hand, there is no test that can detect a change if (15) is satisfied. Thus the minimax separation rate for the moderate sparsity case is of order $(\sqrt{d}/(nh(\tau)))^{1/2}$. In the case of high sparsity, $\beta \in (1/2,1)$, the scan test can detect a change under the condition (14). There is no test that can detect a change if (16) holds. Thus the separation rate is $\big(p\log(d/p)/(nh(\tau)\big)^{1/2}$ for the case of high sparsity. Note that in this case the* *linear test fails if (13) is not satisfied. Roughly speaking, since* $p\log(d/p) \lesssim \sqrt{d}$*, the scan test is able to detect a smaller change in mean than the linear test can detect. On the contrary, in the case of moderate sparsity, the linear test can detect a change of order* $d^{1/4}/\sqrt{nh(\tau)}$*, whereas the scan test cannot. Thus the theorem establishes the boundary between high and moderate sparsity regimes,* $\beta = 1/2$*. We refer to [3] for further discussion.*

REMARK 7. *Consider the problem of comparing the means of two d-dimensional Gaussian vectors. This problem is equivalent to the problem of testing a change at* $\tau = 1$ *in a sequence of* $n = 2$ *vectors* X_1, X_2 *defined in (1). This problem is equivalent to testing p non-zero components in the mean of a Gaussian vector* $Z_2(1)$*. A related problem were previously considered by Baraud [3] and by Ingster and Suslina [18]. For the latter problem, we recover in (15)–(16), a detection boundary which is similar to the one given in [3]. The quantity* $2-1/\beta$ *in (16) coincides with the key quantities arising in the problem of classification of a Gaussian vector with* $p \asymp d^{1-\beta}$ *non-zero components in the mean [17] and in the problem of detection [18, 16]. Butucea and Ingster [8] obtained similar results for the problem of detection of a sparse submatrix of a noisy matrix of growing size.*

4.2. *Adaptation.* The following theorem gives the upper bound for the adaptive test $\psi^* = \psi^*_{\text{lin}} \vee \psi^*_{\text{scan}}$.

THEOREM 3. *Assume that Assumptions (A1–A3) hold and* $r = r(d)$ *satisfies either*

$$\liminf_{d\to\infty} \min_{\tau\in\mathcal{T}} \frac{r^2 nh(\tau)}{\sqrt{d\log(d\log n)}} \geq \sqrt{2} \tag{17}$$

or

$$\liminf_{d\to\infty} \min_{\tau\in\mathcal{T}} \min_{p\in\mathcal{D}} \frac{r^2 nh(\tau)}{p\log(d/p)} \geq 2. \tag{18}$$

Let α *be a given significance level and* $\alpha_l, \alpha_s > 0$ *such that* $\alpha_l + \alpha_s = \alpha$ *and the thresholds* H *and* T_p *be defined as in (8) and (11). Then* $\alpha(\psi^*) \leq \alpha$ *and* $\limsup_{d\to\infty} \beta^*(\psi^*, \Theta[r]) \leq \alpha$*. Moreover, for any* $\gamma \in (0,1)$ *there exists an adaptive test* ψ^* *such that its risk of testing is at most* γ*,* $\limsup_{d\to\infty} \gamma^*(\psi^*, \Theta[r]) \leq \gamma$*.*

The lower bound follows directly from Theorem 2.

THEOREM 4. *Assume that (A1) holds. Let $\gamma \in (0,1)$. If there exist $\tau_0 \in \mathcal{T}$ and $p_0 \in \mathcal{D}$, $p_0 \asymp d^{1-\beta_0}$ as $d \to \infty$ such that*

$$\lim_{d\to\infty} \frac{r^2 nh(\tau_0)}{\sqrt{d}} \le \sqrt{2\log(1+4(1-\gamma)^2)} \quad \text{for} \quad \beta_0 \in [0, 1/2)$$

and

$$\limsup_{d\to\infty} \frac{r^2 nh(\tau_0)}{p_0 \log(d/p_0)} < 2 - \frac{1}{\beta_0} \quad \text{for} \quad \beta_0 \in (1/2, 1),$$

then $\liminf_{d\to\infty} \gamma^*(\psi, \Theta[r]) \ge \gamma$ *for any test* ψ.

REMARK 8. *Theorems 3–4 show that in the case of adaptation to an unknown p the scan test achieves the minimax-optimal rate of convergence with no loss compared to the minimax separation rate. The linear test achieves a rate with a* $\sqrt{\log(d\log n)}$ *loss compared to the minimax separation rate.*

4.3. *Discussion: detection boundary and rates.* Suppose that the change size in each component is constant, $\Delta\theta_j = a_{d,n}\mathbf{1}\{j \in m\}$ for $m \in \mathcal{M}(d,p)$. Thus the squared norm of the jumps of the vector mean $\|\Delta\theta_\tau\|^2$ equals $pa_{d,n}^2$ if we have p components with a change. The detection boundary conditions can now be written in terms of the asymptotic behavior of the jump size $a_{d,n}$. Recall that $p \asymp d^{1-\beta}$ where $\beta \in [0,1)$ is the sparsity coefficient.

In the case of high sparsity we can express the detection boundary condition in the following way. Suppose that $a_{d,n} = C_d\big[\log d/(nh(\tau))\big]^{1/2}$. Then the detection is impossible if $\limsup_{d\to\infty} C_d < (2\beta-1)^{1/2}$. If $\liminf_{d\to\infty} C_d > (2\beta)^{1/2}$, then we can always detect a change by applying the scan test. We see that there is a gap between the constants but they do not depend on γ.

In the case of moderate sparsity, $\beta \in [0, 1/2)$, the detection boundary is of the form $a_{d,n}^2 d^{1/2-\beta} nh(\tau) \asymp 1$. Indeed, in this case $\|\Delta\theta_\tau\|^2 \asymp d^{1-\beta}a_{d,n}^2$, the upper bound of error is attained by the linear test (see Theorem 1) and the lower bound follows from Theorem 2. There is a gap between the upper and the lower bound constants that are equal, respectively, to $2\log(2/\gamma)^{1/4}$ and $(2\log(1+4(1-\gamma)^2))^{1/4}$. We see that in case of the moderate sparsity, the constants at the minimax separation rate depend on the overall testing level γ. If $\beta = 1/2$ and $pd^{-1/2} \to K$ for some constant $K > 0$, the lower bound constant will also depend on K (see the proof of Lemma 5). It will satisfy

$$\frac{[C^*(K)]^4}{2K^2} = 1 + \log\left[\frac{1}{K}\log\Big(1+4(1-\gamma)^2\Big)\right].$$

In the case of unknown p, the scan test is adaptive to β with the same rate $[\log d/(nh(\tau))]^{1/2}$. For the linear test we have a loss in the boundary of order $\big[\log(d \log n)\big]^{1/2}$. We conjecture that the $\log\log n$ loss cannot be improved.

The main results on the detection boundaries are gathered in the table.

	high sparsity, $\beta \in (1/2, 1)$	moderate sparsity, $\beta \in [0, 1/2)$
boundary	$a_{d,n} = C_d \Big(\frac{\log d}{nh(\tau)}\Big)^{1/2}$	$a_{d,n} = C_d \Big(\frac{d^{\beta-1/2}}{nh(\tau)}\Big)^{1/2}$
upper bound	$\liminf_{d\to\infty} C_d > \sqrt{2\beta}$	$\liminf_{d\to\infty} C_d \geq 2(\log(2/\gamma))^{1/4}$
lower bound	$\limsup_{d\to\infty} C_d < \sqrt{2\beta - 1}$	$\limsup_{d\to\infty} C_d \leq (2\log(1 + 4(1-\gamma)^2))^{1/4}$

5. Simulations. We perform a simulation study to evaluate the empirical behavior of the proposed test statistics, depending on the number of observations n, the sparsity index p, the renormalized size of the jumps $\|\Delta\theta_\tau\|/\sqrt{p}$, and the dimension d. We consider several situations depending on the sparsity level: the change in mean in all components ($p = d$), the case of moderate sparsity ($p \gg \sqrt{d}$), and the case of high sparsity ($p \ll \sqrt{d}$). We show the dependence of the power of the tests ψ^*, ψ^*_{lin} and ψ^*_{scan} on the sparsity index β (Figure 1) and on the size of jumps (Figures 2–4). We also propose and evaluate two strategies for calibrating the proposed test statistics.

We perform 500 replications and report results averaged over all replications. The significance level is set to $\alpha = 0.05$. In all simulations we use the properly scaled $1-\alpha/(2n)$ quantile of the χ^2_d-distribution for the calibration of the linear test at level $\alpha/2$. We use the $1-\alpha/(2d)$ empirical quantiles of the statistics $\max_{t\in\mathcal{T}} L^p_{\text{scan}}(t)$, $p = 1, \dots, d$ under H_0 to calibrate the scan test at level $\alpha/2$. More precisely, we use these quantiles instead of $T_{p,n}$ in the adaptive scan test $\psi^*_{\text{scan}} = \mathbf{1}\Big\{\max_{p\in\mathcal{D}} \frac{1}{T_{p,n}} \max_{t\in\mathcal{T}} L^p_{\text{scan}}(t) > 1\Big\}$.

5.1. *Fixed change size, growing dimension.* We generate $n = 100$ independent Gaussian vectors X_i, $i = 1, \dots, n$, with independent components with a change in mean in $p = 5$ components. The dimension d varies from 5 to 200 while $p = 5$ is fixed so that the sparsity index β decreases as $d \to \infty$. The change location is fixed at $\tau = n/2$ or at $\tau = n/4$ (Figure 1 on the left or right). The vector of a change in mean $\Delta\theta_\tau$ has 5 non-zero components with absolute values $|\Delta\theta^j_\tau| = 0.6$, $j \in m$, where the subset of the change support $m \in \mathcal{M}(d, p)$ remains fixed as d is growing. More precisely, we set the first $p = 5$ changing components, $m = \{1, \dots, p\}$, we vary the dimension $d \geq p$ by adding the next $d - p$ components of the mean vector that do not change. Figure 1 compares the power of the linear adaptive test ψ^*_{lin}, the

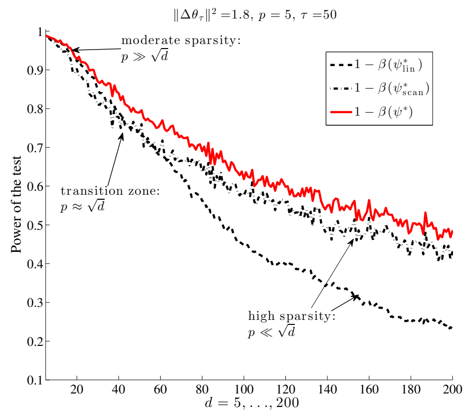

Fig 1. *Power of three tests for $n = 100$, $p = 5$, $\|\Delta\theta_\tau\|^2 = 1.8$. The change point is $\tau = 50$ on the left graph and $\tau = 25$ on the right graph.*

scan adaptive test ψ^*_{scan}, and the final test ψ^* (in red color) for two cases of the change-point location. First, we see that the power is greater if the change is in the middle (left graph). The scan test performs better in the high sparsity case, whereas the power of the linear test is higher in the moderate sparsity case. In the transition zone $p \approx \sqrt{d}$, both tests have similar power. The difference in the power of the two tests for $d = 5, \ldots, 10$ is not very large. This is due to the fact that the chosen norm of the change is quite large for a rather small dimension, so both tests can detect the change easily.

5.2. *Growing change size, fixed dimension.* We also generate $n = 100$ independent d-dimensional Gaussian vectors X_i, $i = 1, \ldots, n$ with independent components. The dimension of X_i is set to $d = 100$ and remains fixed. We run simulations in three sparsity regimes, $p = 3,\ 10,\ 50$ that correspond to the case of the high sparsity ($p = 3$), the transition zone ($p = 10$) and the case of the medium sparsity ($p = 50$). The absolute value of the change in mean at each component was the same for each changing vector component, $\forall j \in m$, $m \in \mathcal{M}(d, p)$, $|\Delta\theta^j_\tau| = \delta > 0$. The norm of the change is then equal to $\|\Delta\theta_\tau\| = \delta\sqrt{p}$. We make δ vary from 0.1 to 0.8. The set m is fixed for each of the three sparsity regimes.

In Figure 2, we report the empirical power of the three tests depending on $\delta = \|\Delta\theta_\tau\|/\sqrt{p}$. We observe that the detection boundary constant, which is proportional to $\|\Delta\theta_\tau\|$, increases as p decreases. In the high sparsity case ($p = 3$), we observe that the test based on the scan statistics outperforms the linear test. On the other hand, in the moderate sparsity case ($p = 50$), the linear test works better than the scan statistic test. In the transition

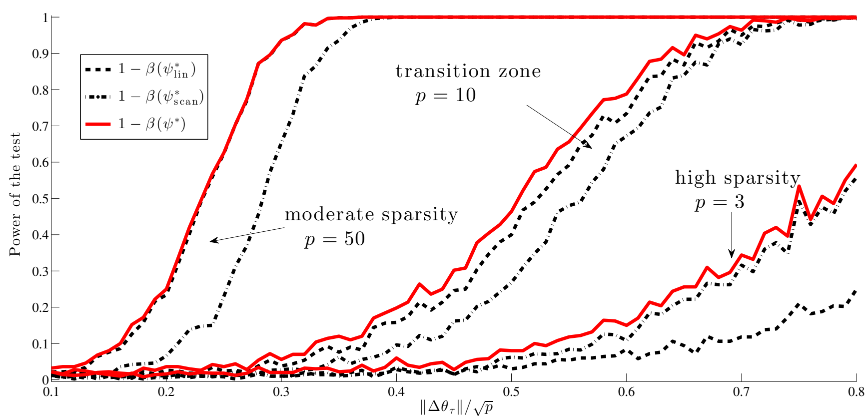

FIG 2. *Power of the tests for $n = 100$, $d = 100$, $\tau = 25$ for different sparsity regimes and fixed change size $\delta = \|\Delta\theta_\tau\|/\sqrt{p}$ in p coordinates.*

zone, $p = 10$, the linear test is slightly better, but we see that both tests contribute to the performance of the final test ψ^*.

5.3. *Two calibration strategies.* We will now compare the powers of the tests for two calibration strategies of the scan test. The first calibration strategy consists in using the simulated quantiles of level $1 - \alpha/(2d)$ of the process $\max_{t\in\mathcal{T}} L^p_{\mathrm{scan}}(t)$ under H_0 instead of the thresholds $T_{p,n}$. These quantiles are the quantiles of the norm of the maximum of a centered normalized discrete d-dimensional Brownian bridge,

$$\max_{t\in\mathcal{T}} L^p_{\mathrm{scan}}(t) \overset{d}{=} \max_{t\in\mathcal{T}} \frac{1}{\sqrt{2p}} \left(\sum_{j=1}^{p} [\xi^j_n(t)]^2 - p \right) \quad \text{under } \mathrm{H}_0,$$

where $\xi^j_n(t)$, $j = 1, \dots, d$ is a normalized discrete Brownian bridge defined in (21). The second calibration strategy consists in calibrating the scan test using the theoretical quantiles. Instead of $T_{p,n}$ given by (11) we used the following approximate version:

$$T^*_{p,n} = \sqrt{2p} \log\left[\frac{de}{p}\right] + \sqrt{\frac{2}{p}} \log\left[\frac{2nd}{\alpha}\right],$$

where we used the bound $\log\binom{d}{p} \leq p\log(de/p)$ and omitted the second-order terms of $T_{p,n}$. The linear test ψ^*_{lin} is always calibrated by using the appropriately normalized quantile of level $1 - \alpha/(2n)$ of the χ^2_d distribution instead of H^*.

In Figures 3 and 4 we present the results of simulations for $n = 1000$, $d = 100$ and for $n = 100$, $d = 1000$, respectively. In both cases, the change occurs in the middle of the observation interval, $\tau = n/2$. We consider the cases of high sparsity ($p = 1$), no sparsity, ($p = d$), and the intermediate case when $p \approx \sqrt{d}$. The absolute value of the change in mean at each component was the same for each changing vector component: $\forall j \in m$, $m \in \mathcal{M}(d, p)$, $|\Delta\theta^j_\tau| = \delta > 0$. The norm of the change is then equal to $\|\Delta\theta_\tau\| = \delta\sqrt{p}$. We make δ vary from 0 to 0.8 in the case of $d = 100$, $n = 1000$ and from 0 to 2 in case of $d = 1000$, $n = 100$. The set m is fixed for each of the three sparsity regimes.

We can see from Figure 3 that the threshold $T^*_{p,n}$ gives satisfactory results. In the cases of no sparsity and medium sparsity, the powers of both scan tests with empirical and theoretical calibration are quite comparable (blue and black dashed lines correspond to the empirical and theoretical thresholds, respectively). The red and magenta colors show the power of the corresponding test $\psi^* = \psi^*_{\text{lin}} \vee \psi^*_{\text{scan}}$.

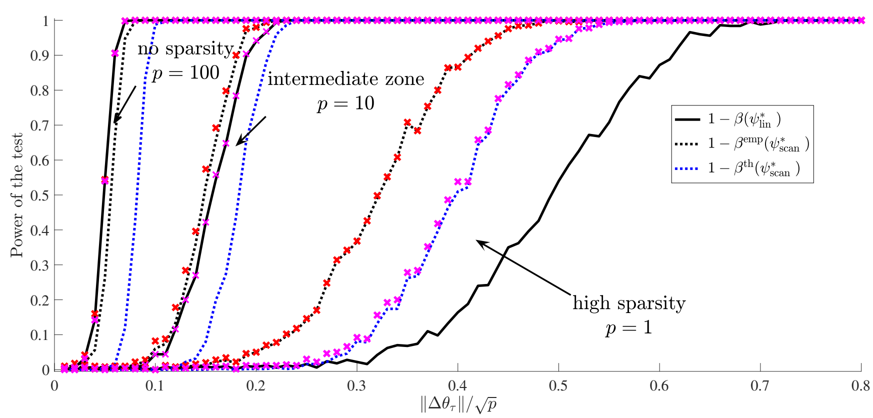

FIG 3. *Power of the test for $n = 1000$, $d = 100$, $\tau = n/2$ for $p = 1, 10, 100$. The red and magenta markers show the power of the final adaptive test $\psi^* = \psi^*_{\text{lin}} \vee \psi^*_{\text{scan}}$ for the empirical and theoretical thresholds, respectively.*

In Figure 4, where the dimension d is large, we can see that the linear test is much less powerful when the sparsity is high, in contrast to the lower-dimensional setting considered in Figure 3. We refer the reader to Supplement A for the results in the case of moderate sparsity where $p = 900$ as well as to the results in the case of known p and τ.

Estimation of thresholds for the scan statistics using Monte–Carlo simu-

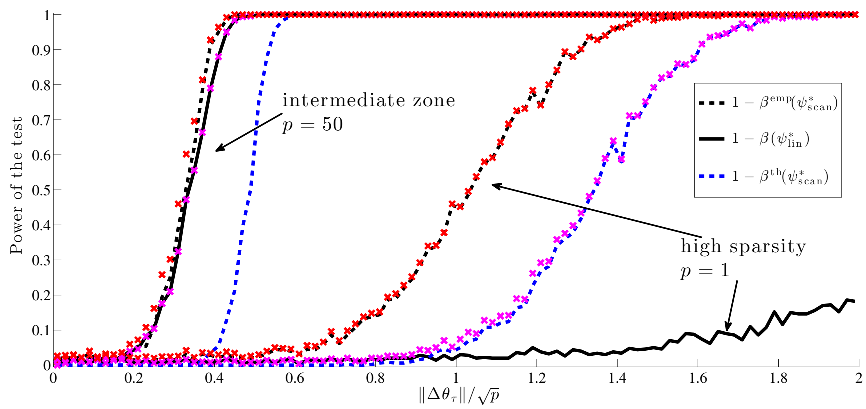

FIG 4. *Power of the test for $n = 100$, $d = 1000$, $\tau = n/2$ for $p = 1, 50$. In red and magenta colors the power of the proposed test ψ^* is presented for the empirical and theoretical thresholds, respectively.*

lations suffers from the curse of dimensionality, since for a fixed accuracy the required number of Monte–Carlo replications will grow with the dimension d. Thus the theoretical thresholds can be attractive options in high-dimensional settings.

5.4. *Type I errors.* We present empirical type I errors in Figure 5 for the case of the adaptive test. The dimension d varies from 100 to 750 and the sequence length is $n = 100$. The thresholds are set empirically, the level of the test is $\alpha = 0.05$. We plug in theoretical quantiles of the χ^2_d distribution in the threshold of the linear test, $H^* = (q_{\chi^2_d}(1 - \frac{\alpha}{2n}) - d)/\sqrt{2d}$. We use the empirical quantiles instead of the thresholds $T_{p,n}$ in the scan test as described in the previous subsections. We can notice that the type I error

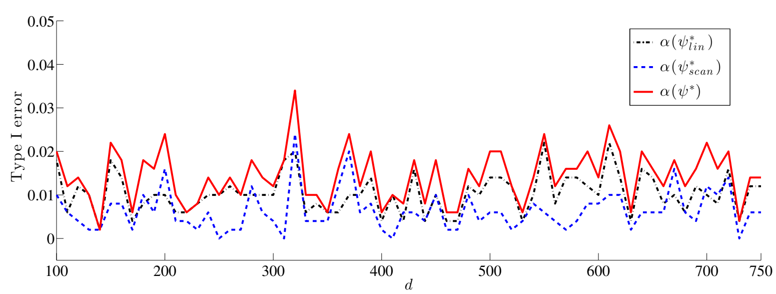

FIG 5. *Empirical type I error of the adaptive tests for $n = 100$, $d = 100, \ldots, 750$*

of the final decision rule ψ is about two times smaller than the declared level of the test α. Indeed, the thresholds are conservative, since we use the Bonferroni procedure. Next, we can see that the empirical type I error of the linear test is almost always greater than the one of the scan test. This is explained by the fact that for any p the threshold $T_{p,n}$ of the scan test is greater than the threshold of the linear test. Thus, the scan test tends to not reject the null hypothesis even if the linear test rejects it. In general, the scan test will be useful to detect a change in a small number of components, since in this case its detection boundary (14) is smaller than the one of the linear test (13).

Finally, we would like to mention that both tests can be also calibrated by the empirical quantiles of the corresponding limiting processes. This can potentially increase the power of the test since the type I error will get closer to the significance level.

6. Application to real data.

We present two illustrations of the method on real data. The first example is classical. We apply the sliding window version of our testing procedure to the problem of simultaneous segmentation of comparative genomic hybridization (CGH) profiles. The second example comes from astrophysics. We apply our method to the detection of distant galaxies.

6.1. *CGH data segmentation.*

Comparative genomic hybridization (CGH) is a technique to obtain the number of copies of genes in a DNA profile. The number of copies of certain genes can vary in a tumor cell with respect to a normal cell. This phenomenon is called copy-number variation. The task consists in finding which regions of DNA contain this elevated or lowered copy number for each particular type of tumor.

We consider a publicly available dataset of bladder tumor profiles [29]. This dataset contains $d = 57$ bladder tumor samples. Each profile is a sequence of 2385 relative quantities of DNA that describe the copy number variation. The problem is to test whether there is a simultaneous change in copy numbers of some of 57 profiles as well as to estimate the location of the change. We remove the observations corresponding to the sex chromosomes as suggested in [30] since the sex mismatch between the patients leads to less reliable results. We also remove the vectors containing outliers for more than 9 patients at the same position on the chromosome. The final dataset contains $N = 2041$ observations recorded from 22 chromosomes for each patient.

The problem clearly fits our framework, since the change in the number of copies does not necessarily occur for all profiles. Moreover, the CGH profiles are independent. A profile is usually modeled by a piecewise constant regression with independent Gaussian errors [25]. Thus we observe a sequence of d-dimensional vectors y_i modeled by

$$y_i = \theta + \sum_{k=0}^{K-1} \Delta\theta_i \mathbf{1}\{\tau_k < i \leq \tau_{k+1}\} + \xi_i, \quad i = 1, \dots, N,$$

where θ and $\Delta\theta_i$ are the unknown means of the relative DNA quantities and their changes, respectively, $\xi_i \sim \mathcal{N}(0, \Sigma)$ are independent with a diagonal covariance matrix Σ and the τ_k are the unknown copy-number change locations, $k = 0, \dots, K-1$.

We apply a sliding window version of the proposed testing procedure to estimate the change-points. The data is normalized before applying our method: we divide each column y_i by its empirical standard deviation calculated for each profile. We suppose that after this normalization the data can be modeled by the model (1) within any window of a fixed size h. For a given significance level α, we test whether there is a change over a sliding window of size h. We use simulated quantiles as the thresholds for both tests for $n = h$ observations. If the change is detected within the interval $(u, u+h]$, we estimate the change-point as

$$\widehat{\tau} = \arg \max_{u+1 \leq t \leq u+h} \Big\{ \max\big(L_{\text{lin}}(t), L_{\text{scan}}(t)\big) \Big\}.$$

Figure 6 summarizes the results of this analysis. The black piecewise con-

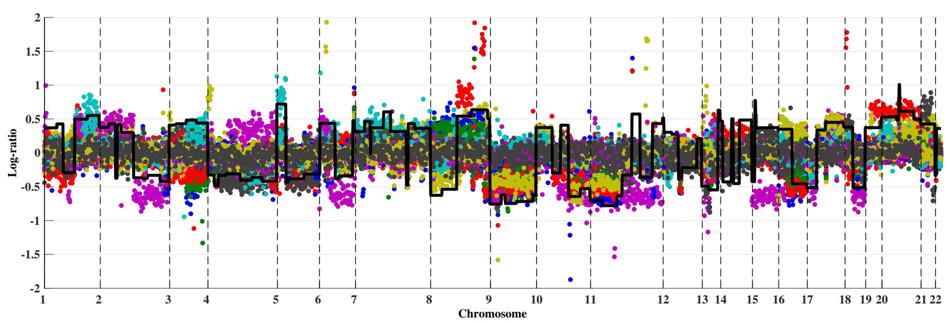

FIG 6. *Simultaneous segmentation of $d = 57$ CGH profiles of length $n = 2041$. The profiles of different patients are shown in different colors. The black line corresponds to the scaled norm of the mean multiplied by the sign of the highest profile mean of the segment.*

stant line shows the scaled estimated norm of the mean $\|\Delta\theta_i + \theta\|_2/\sqrt{d}$ within each segment $(\tau_k, \tau_{k+1}]$, multiplied by the sign of the largest empirical mean within the segment. We have chosen this quantity as a reference value to graphically represent the size and the direction of the change in mean in the profiles since we cannot plot the mean changes in all the profiles.

The experimental results are quite similar to the segmentation results obtained in [30]. The sliding window size is fixed to $h = 40$, the significance level is set to $\alpha = 0.05$ and both tests were calibrated at the level $\alpha/2$. We present only 20 profiles in the figure although the estimation was done using all 57 profiles. Vertical lines show the chromosome borders. The results are satisfactory. The method can quickly detect a change in a small subset of components of a high-dimensional vector. Higher robustness to outliers, which occur in such data, could be an interesting improvement to consider in future work.

6.2. *Detection of distant galaxies.* The data we consider here represents the spectra of galaxies obtained by the Multi-Unit Spectroscopic Explorer (MUSE) [2]. The data collected with the MUSE instrument is massive hyper-spectral data cubes of up to 4000 images of 300×300 pixels, where each image corresponds to a certain wavelength.

The data we consider in this paper was provided by expert astrophysicists. First, the simulations of astronomical scenes are run. Then, the resulting simulations are processed by applying the MUSE Instrument Numerical Model. We therefore have three cubes of data of size $100 \times 100 \times 3600$. Each cube represents the real value of the linear spectra s, the observed value y and the noise variance Σ. More precisely, we observe d-dimensional vectors, $d = 3600$, modeled by

$$y_{ij} = s_{ij} + \varepsilon_{ij}, \quad (i, j) \in A,$$

where $(i, j) \in A$ are spatial coordinates, A is indexed by $\{1, \dots, 100\}^2$, $s_{ij} \in \mathbb{R}^d$ is the spectral column, and $\varepsilon_{ij} \sim \mathcal{N}(0, \Sigma_{ij})$ are independent for all $(i, j) \in A$. The covariance matrices Σ_{ij} are diagonal, with components $[\sigma_{ij}^k]^2$, $k = 1, \dots, d$. The mean of the signal-to-noise ratio y_{ij}/s_{ij} over the spectra is shown in Figure 7. The data is sparse with respect to the spectral values as well as with respect to the spatial coordinates.

The goal is to detect the presence of the galaxies using the information about their spectra; see [26, 24] for some recent work on this topic. The data is renormalized by dividing the coordinates of each observation by the corresponding standard deviation σ_{ij}^k provided by the astrophysicists. We suppose that within a small window the new data $x_{ij} := \Sigma_{ij}^{-1} y_{ij}$ follow model (1) for any fixed i or j. For each fixed spatial coordinate $i \in \{1, \dots, 100\}$ we test the hypotheses of no change in mean in each contiguous couple

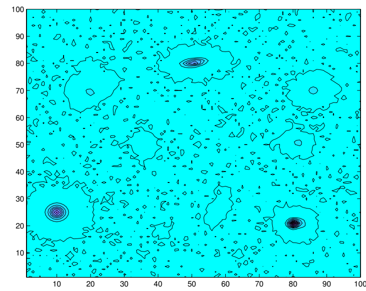

Fig 7. *The mean signal-to-noise ratio over 3600 spectral values.*

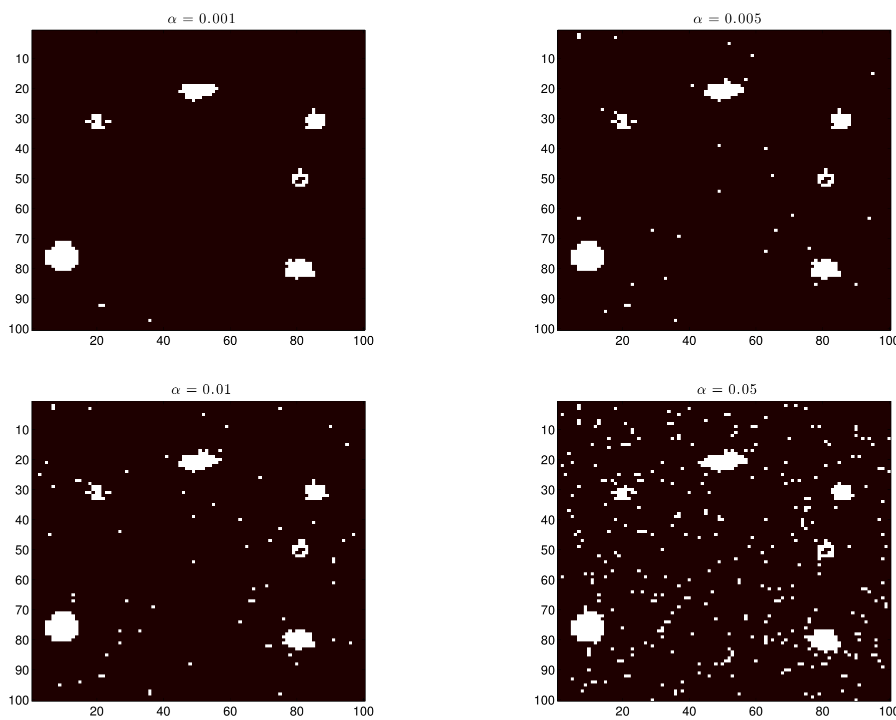

Fig 8. *The results of testing for* $\alpha = 0.001, 0.005, 0.01, 0.05$. *The detected galaxies are shown in white.*

of vectors $x_{i,j}$ and $x_{i,j+1}$, $j = 1, \ldots, 99$. Thus we test a change-point at $(i, j) \in \{1, \ldots, 100\} \times \{1, \ldots, 99\}$. The same tests are performed for each fixed $j \in \{1, \ldots, 100\}$. We say that the galaxy is detected at $(i, j) \in A$ if one of the tests, column-wise or row-wise, has detected the change in mean (we use only one test for the coordinates $(i, 100)$ and $(100, j)$). The results are shown in Figure 8 for the significance levels $\alpha = 0.001, 0.005, 0.01, 0.05$.

To get the overall significance level α, both tests are calibrated at level $\alpha/4$, column-wise and row-wise. We use the theoretical thresholds for the scan test at $\alpha_s = \alpha/4$ and the thresholds based on the chi-squared distribution quantiles $q_{\chi^2_{3600}}(1 - \alpha/4)$ for the linear test. Comparing the results with Figure 7, we can see that the locations of six galaxies are detected. However, the test fails in the presence of faint signals. A possible approach to increase the power of the test is to take into account the simultaneous change in variance. We refer the reader to [26] for some other methods as well as for the details about the MUSE instrument data.

7. Conclusion. We have proposed a test for a change in mean of a sequence of high-dimensional Gaussian vectors under the assumption that the change occurs in a subset of components of unknown size p of the mean vector. The proposed test is based on a combination of two basic tests, the linear test and the scan test. The test is adaptive to the unknown sparsity index p. The obtained detection boundary conditions provide a correct minimax separation rate up to a constant. The constant in the lower bound of the minimax testing error matches the constant that appears in related testing and classification problems. The constant in the upper bound could potentially be improved. In the case of high-sparsity, a method based on higher criticism [12] could be worthwhile investigating. The proposed approach can be readily extended to the case of unknown equal variance across components.

APPENDIX A: UPPER BOUNDS

First, using (2) we reduce the model (1) to the model

$$Z_n(t) = -\Delta\theta_\tau \mu_n(t) + \xi_n(t), \quad t \in \mathcal{T}, \tag{19}$$

where $\mathcal{T} = \{1, \ldots, n-1\}$,

$$\mu_n(t) = \sqrt{\frac{t(n-t)}{n}} \left(\frac{n-\tau}{n-t} \mathbf{1}\{t \le \tau\} + \frac{\tau}{t} \mathbf{1}\{t > \tau\} \right). \tag{20}$$

and $\xi_n(t) = (\xi_n^1(t), \ldots, \xi_n^d(t))^T$ are Gaussian vectors with $\mathcal{N}(0, 1)$ dependent components given by

$$\xi_n^j(t) = \sqrt{\frac{n}{t(n-t)}} \left(\sum_{i=1}^t \xi_i^j - \frac{t}{n} \sum_{i=1}^n \xi_i^j \right), \quad j = 1, \ldots, d. \tag{21}$$

Note that the process $\xi_n^j(t)$, $t \in \mathcal{T}$ is a discrete normalized Brownian bridge.

A.1. Upper bound for Problem (P1). The proof of Theorem 1 is based on the following two lemmas proven in Supplement A.

LEMMA 1. *Let α be a given significance level and the thresholds H and T_p be defined in (6) and (7), where $\alpha_l + \alpha_s = \alpha$, $\alpha_l, \alpha_s > 0$. Then the type I error $\alpha(\psi_p^*)$ is smaller than α.*

In the next lemma we derive the conditions on r that allow us to control the the Type II error.

LEMMA 2. *Let $p \in \mathcal{D}$ and $\tau \in \mathcal{T}$ be given and $\alpha_l, \alpha_s \in (0,1)$ be given significance levels of the linear and the scan test, respectively.*

1. Let $\beta_l \in (0,1)$ be given. Assume that there exists a $\lambda \in (0,1)$ such that

$$(1-\lambda)\frac{r^2 nh(\tau)}{\sqrt{2d}} \geq \left[2\log\frac{1}{\alpha_l}\right]^{1/2} + \left[2\log\frac{1}{\beta_l}\right]^{1/2} + \sqrt{\frac{2}{d}}\left(\log\frac{1}{\alpha_l} + \frac{1}{\lambda}\log\frac{1}{\beta_l}\right). \tag{22}$$

Then $\beta(\psi_{\text{lin}}, \Theta_p[r], \tau) \leq \beta_l$ for the linear test ψ_{lin} of level α_l.

2. Let $\beta_s \in (0,1)$ be given. Assume that there exists a $\lambda \in (0,1)$ such that

$$\begin{aligned}(1-\lambda)\frac{r^2 nh(\tau)}{\sqrt{2p}} \geq{}& \left[2\log\frac{1}{\alpha_s}\right]^{1/2} + \left[2\log\frac{1}{\beta_s}\right]^{1/2} + \sqrt{\frac{2}{p}}\left(\log\frac{1}{\alpha_s} + \frac{1}{\lambda}\log\frac{1}{\beta_s}\right)\\ &+ \sqrt{2p}\left(\log\frac{de}{p} + \left[\log\frac{de}{p}\right]^{1/2}\right).\end{aligned} \tag{23}$$

Then $\beta(\psi_{\text{scan}}^p, \Theta_p[r], \tau) \leq \beta_s$ for the scan test ψ_{scan}^p of level α_s.

Now using these lemmas we can prove the theorem.

Proof of Theorem 1. Lemma 1 implies that $\alpha(\psi_p^*) \leq \alpha$ if H and T_p are chosen as in (6) and (7). On the other hand,

$$\beta(\psi_p^*, \Theta_p[r], \tau) \leq \min\Big(\beta(\psi_{\text{lin}}, \Theta_p[r], \tau), \beta(\psi_{\text{scan}}^p, \Theta_p[r], \tau)\Big).$$

Note that if r satisfies (22) or (23) with the corresponding significance levels α_l and α_s, then the type II error of the test ψ_p^* is smaller than $\min(\beta_l, \beta_s)$. We have to show that (22) and (23) are satisfied if (13) and (14) hold true.

Consider the linear test. Inequality (22) becomes

$$\frac{r^2 nh(\tau)}{\sqrt{d}} \geq \frac{2}{1-\lambda}\left(\left[\log\frac{1}{\alpha_l}\right]^{1/2} + \left[\log\frac{1}{\beta_l}\right]^{1/2}\right) + \frac{2}{\sqrt{d}(1-\lambda)}\left(\log\frac{1}{\alpha_l} + \frac{1}{\lambda}\log\frac{1}{\beta_l}\right).$$

Taking $\lambda = \log d/\sqrt{d}$ we obtain that the right-hand side of this inequality tends to $2(\sqrt{-\log \alpha_l} + \sqrt{-\log \beta_l})$ which is minimal for $\alpha_l = \beta_l = \gamma/2$ under the constraints $0 < \alpha_l + \beta_l \le \gamma$. If condition (13) is satisfied, then the above inequality holds true for sufficiently large d given that $\alpha_l = \beta_l = \gamma/2$. Applying Lemma 2 we obtain $\lim_{d\to\infty} \beta(\psi_{\text{lin}}, \Theta_p[r], \tau) \le \gamma/2$.

For the scan test, set $\lambda = 1/\sqrt{p}$. Then $\lambda \to 0$ as $d \to \infty$, since $p \to \infty$ as $d \to \infty$ by Assumption (A2). We can rewrite (23) as follows,

$$\begin{aligned}\frac{r^2 nh(\tau)}{p\log(d/p)} \ge{}& \frac{2}{1-\lambda}\frac{1}{\sqrt{p}\log(d/p)}\left(\left[\log\frac{1}{\alpha_s}\right]^{1/2} + \left[\log\frac{1}{\beta_s}\right]^{1/2}\right)\\ &+ \frac{2}{(1-\lambda)p\log(d/p)}\left(\log\frac{1}{\alpha_s} + \frac{1}{\lambda}\log\frac{1}{\beta_s}\right)\\ &+ \frac{2}{(1-\lambda)\log(d/p)}\left(1 + \log\frac{d}{p} + \left[\log\frac{de}{p}\right]^{1/2}\right).\end{aligned}$$

Taking into account Assumption (A2), we can also see that $\sqrt{p}\log(d/p) \to \infty$ and $\log(d/p) \to \infty$. Thus $\forall \alpha_s, \beta_s \in (0,1)$ the above inequality holds if

$$\frac{r^2 nh(\tau)}{p\log(d/p)} \ge 2 + o(1), \quad d \to \infty,$$

which follows from condition (14). Take $\beta_s \in (0, \min(\gamma - \alpha, \beta_l))$. Then

$$\lim_{d\to\infty}[\alpha(\psi_p^*) + \beta(\psi_p^*, \Theta_p[r], \tau)] \le \alpha_l + \alpha_s + \min(\beta_l, \beta_s) \le \gamma. \qquad \blacksquare$$

A.2. Upper bound for Problem (P2). The proof of the upper bound in the adaptive case is similar to the proof in the case of known p and θ. The detailed proofs can be found in Supplement A.

LEMMA 3. *Let $0 < \varepsilon < n-1$. The thresholds of the linear and scan test are given by*

$$H_\varepsilon = (1+\varepsilon)\sqrt{2\log\Big[\frac{2\log n}{\alpha_1\log(1+\varepsilon)}\Big]} + (1+\varepsilon)\sqrt{\frac{2}{d}}\log\Big[\frac{2\log n}{\alpha_l\log(1+\varepsilon)}\Big] + \varepsilon\sqrt{\frac{d}{2}}, \tag{24}$$

$$T_{p,n} = \sqrt{2\log\Big[\binom{d}{p}\frac{2np^2}{\alpha_s}\Big]} + \sqrt{\frac{2}{p}}\log\Big[\binom{d}{p}\frac{2np^2}{\alpha_s}\Big], \tag{25}$$

where $\alpha_l + \alpha_s \le \alpha$, $\alpha_l, \alpha_s > 0$. Then $\alpha(\psi^*) \le \alpha$.

LEMMA 4. *Let α_l and α_s be given significance levels for the linear and the scan test, respectively. Recall that τ is the true change-point. Let $\beta_l \in (0,1)$ and $\beta_s \in (0,1)$ be given.*

1. *Assume that* $\forall \tau \in \mathcal{T}$ *there exists a* $\lambda \in (0,1)$ *and an* $\varepsilon > 0$ *such that*

$$(1-\lambda)\frac{r^2 nh(\tau)}{\sqrt{2d}} \geq \sqrt{2\log\frac{1}{\beta_l}} + \frac{1}{\lambda}\sqrt{\frac{2}{d}}\log\frac{1}{\beta_l} + (1+\varepsilon)\sqrt{\frac{2}{d}}\log\frac{1}{\alpha_\varepsilon} + (1+\varepsilon)\sqrt{2\log\frac{1}{\alpha_\varepsilon}} + \varepsilon\sqrt{\frac{d}{2}}, \tag{26}$$

where α_ε *is defined in (9). Then* $\beta(\psi^*_{\text{lin}}, \Theta[r], I) \leq \beta_l$ *for the linear test* ψ^*_{lin} *of level* α_l*.*

2. *Assume that* $\forall p \in \mathcal{D}$ *and* $\forall \tau \in \mathcal{T}$ *there exists a* $\lambda \in (0,1)$ *such that*

$$(1-\lambda)\frac{r^2 nh(\tau)}{\sqrt{2p}} \geq \sqrt{2\log\frac{1}{\alpha_s}} + \sqrt{\frac{2}{p}}\log\frac{1}{\alpha_s} + \sqrt{2\log\frac{1}{\beta_s}} + \frac{1}{\lambda}\sqrt{\frac{2}{p}}\log\frac{1}{\beta_s} + \sqrt{2\log(2np^2)} + \sqrt{\frac{2}{p}}\log(2np^2) + \sqrt{2p}\log\frac{de}{p} + \sqrt{2p\log\frac{de}{p}}. \tag{27}$$

Then $\beta(\psi^*_{\text{scan}}, \Theta[r], I) \leq \beta_s$ *for the adaptive scan test* ψ^*_{scan} *of level* α_s*.*

Proof of Theorem 3. As in the proof of Theorem 1, we can easily see that if the thresholds are chosen as in (8) and (10), then the type II error of the adaptive test is smaller than α,

$$\beta(\psi^*, \Theta[r], \tau) \leq \min\Big(\beta(\psi^*_{\text{lin}}, \Theta[r], I), \beta(\psi^*_{\text{scan}}, \Theta[r], I)\Big) \leq \min(\alpha_l, \alpha_s) \leq \alpha.$$

Thus we need to show that the conditions of Lemma 3 are satisfied if the assumptions of Theorem 3 hold.

Let us start with the scan test. Lemma 3 implies that for $T_{p,n}$ chosen as in (11), the type I error of the test is bounded by α_s, if r satisfies (27) $\forall p \in \mathcal{D}$. Rewrite the latter inequality as

$$\begin{aligned}\frac{r^2 nh(\tau)}{2p\log(d/p)} &\geq (1-\lambda)^{-1}\frac{\sqrt{\log\frac{1}{\alpha_s}} + \sqrt{\log\frac{1}{\beta_s}}}{\sqrt{p}\log(d/p)} + \frac{1}{1-\lambda}\frac{\log\frac{1}{\alpha_s} + \lambda^{-1}\log\frac{1}{\beta_s}}{p\log(d/p)} \\ &\quad + (1-\lambda)^{-1}\left[1 + [\log(d/p)]^{-1}\right]\cdot\left[(\log(d/p))^{-1/2} + (1+\log(d/p))^{-1}\right]^{1/2} \\ &\quad + (1-\lambda)^{-1}\left[\frac{\log(2np^2)}{p\log(d/p)}\right]^{1/2}\cdot\left[(\log(d/p))^{-1/2} + \left(\frac{\log(2np^2)}{p\log(d/p)}\right)^{1/2}\right].\end{aligned} \tag{28}$$

Set $\lambda = 1/\sqrt{p}$. Note that $\lambda \to 0$ as $d \to \infty$. Taking into account Assumption (A2), we can also see that $\sqrt{p}\log(d/p) \to \infty$, $\log(d/p) \to \infty$ and the first three terms of the above inequality tend to 0 for any $\alpha_s, \beta_s \in (0,1)$. Assumptions (A3) and (A2) imply for all $p \in \mathcal{D}$

$$\frac{\log(np^2)}{p\log(d/p)} = \frac{\log n}{p\log(d/p)} + \frac{2\log p}{p\log(d/p)} \to 0, \quad d \to \infty,$$

therefore the last term tends to 1 as $d \to \infty$. If condition (18) is satisfied, then $\forall \tau \in \mathcal{T}$ and $\forall p \in \mathcal{D}$

$$\frac{r^2 nh(\tau)}{2p\log(d/p)} \geq 1 + o(1), \quad d \to \infty.$$

Therefore inequality (28) holds for any $\Delta\theta_\tau \in \Theta[r]$, $\forall \tau \in \mathcal{T}$ and $\forall p \in \mathcal{D}$, and the scan test's type II error is smaller than β_s, $\lim_{d\to\infty} \beta(\psi^*_{\text{scan}}, \Theta[r]) \leq \beta_s$.

Let us now turn to the linear test. Rewrite the inequality (26) as follows

$$\begin{aligned} \frac{r^2 nh(\tau)}{\sqrt{2d}} &\geq \frac{1}{1-\lambda}\sqrt{2\log\frac{1}{\beta_l}} + \frac{1}{\lambda(1-\lambda)}\sqrt{\frac{2}{d}}\log\frac{1}{\beta_l} \\ &+ \frac{1+\varepsilon}{1-\lambda}\sqrt{2\log\frac{1}{\alpha_\varepsilon}}\left[\left(\frac{\log\frac{1}{\alpha_\varepsilon}}{d}\right)^{1/2} + 1\right] + \varepsilon\sqrt{\frac{d}{2}}. \end{aligned} \tag{29}$$

Set $\lambda = 1/\sqrt{d}$ and $\varepsilon = \sqrt{2\log d/d}$. Then the threshold H_ε becomes H^* defined in (8). Note that for $d \geq 2$ we have $0 < \varepsilon < 1$. We have the following inequalities

$$\log\varepsilon - \varepsilon \leq \log\log(1+\varepsilon) \leq \log\varepsilon$$

implying the asymptotic $\log\log(1+\varepsilon) \asymp -\log d$ as $d \to \infty$. Using this inequality we can show that

$$\begin{aligned} \log\frac{1}{\alpha_\varepsilon} &= \log\frac{1}{\alpha_l} + \log\log n - \log\log(1+\varepsilon) \\ &\leq \log\frac{1}{\alpha_l} + \log\log n + \sqrt{\frac{2\log d}{d}} - \frac{1}{2}\log 2 - \frac{1}{2}\log\log d + \frac{1}{2}\log d. \end{aligned}$$

Thus, we will have a loss of rate of order $\sqrt{\log(\sqrt{d}\log n)}$ in the detection boundary. Indeed, the first two terms in (29) are asymptotically constant if $\lambda = 1/\sqrt{d}$ and $d \to \infty$. The third term is of order $\sqrt{\log\log n + \log\sqrt{d}}$ since $n, d \to \infty$. The last one equals $\sqrt{\log d}$. Thus, (26) is satisfied for sufficiently large d, if for any $\tau \in \mathcal{T}$

$$\liminf_{d\to\infty} \frac{r^2 nh(\tau)}{\sqrt{2d\log(d\log n)}} \geq 1.$$

This proves the theorem. ∎

APPENDIX B: PROOF OF THE LOWER BOUND

Proof of Theorem 2. We have to show that under conditions of Theorem 2 for any $\gamma \in (0,1)$ there exist a constant $C_* > 0$ such that $\forall\ C < C_*$

$$\inf_{\psi} \gamma(\psi, \Theta_p[Cr_d], \tau) \geq \gamma, \quad d \to \infty.$$

The classical approach to the construction of lower bounds in testing problems with sparsity goes back to the works of [19]. We also refer to [3] for the non-asymptotic analysis of the related problem of testing non-zero coordinates in a Gaussian vector.

Recall that we observe the sequence $\mathbb{X}^n = (X_1, \dots, X_n)$ of d-dimensional random vectors

$$X_i = \theta + \Delta\theta_\tau \mathbf{1}\{i > \tau\} + \xi_i, \quad i = 1, \dots, n$$

with a possible change at the location $\tau \in \mathcal{T}$. Let us define the following class of mean vectors θ and the changes $\Delta\theta$:

$$\Theta_p^\delta[r,\tau] = \Big\{(\theta, \Delta\theta_\tau) \in \mathbb{R}^{2d} :\ \theta = -\delta\Big(1 - \frac{\tau}{n}\Big),\ \Delta\theta_\tau = \delta,\ \delta \in V_p^d \subset \mathbb{R}^d,\ \|\delta\| = r\Big\}.$$

Note that if $(\theta, \Delta\theta_\tau) \in \Theta_p^\delta[r,\tau]$, then $\Delta\theta_\tau \in \Theta_p[r]$. Recall the definition of type II error for problem (P1):

$$\beta(\psi, \Theta_p[r], \tau) = \sup_{(\theta, \Delta\theta_\tau) \in \mathbb{R}^d \times \Theta_p[r]} \mathbb{P}_{\Delta\theta_\tau, \theta}\{\psi = 0\}.$$

Define the type II error for the alternatives that belong to the class $\Theta_p^\delta[r,\tau]$:

$$\beta(\psi, \Theta_p^\delta[r,\tau]) = \sup_{(\theta, \Delta\theta_\tau) \in \Theta_p^\delta[r,\tau]} \mathbb{P}_{\Delta\theta_\tau, \theta}\{\psi = 0\}.$$

Let $\mathbb{P}_{\pi,r}$ be a prior distribution on the set $\Theta_p^\delta[r,\tau]$ and $\mathbb{P}_0$ be the prior corresponding to the case of no change in mean and zero means before and after the change: $\delta = 0$ in the definition of $\Theta_p^\delta[r,\tau]$. Using the standard lower bound machinery (see, *e.g.*, [19]), we have

$$\begin{aligned}
\inf_{\psi} \gamma(\psi, \Theta_p[r], \tau) &= \inf_{\psi}\Big(\alpha(\psi) + \beta(\psi, \Theta_p[r], \tau)\Big) \\
&\geq \inf_{\psi}\Big(\alpha(\psi) + \beta(\psi, \Theta_p^\delta[r,\tau])\Big) \\
&\geq 1 - \frac{1}{2}\|\mathbb{P}_{\pi,r} - \mathbb{P}_0\|_1 \\
&= 1 - \frac{1}{2}\mathbb{E}_0|\mathcal{L}_{\pi,r}(\mathbb{X}^n) - 1| \\
&\geq 1 - \frac{1}{2}\Big(\mathbb{E}_0\mathcal{L}_{\pi,r}^2(\mathbb{X}^n) - 1\Big)^{1/2},
\end{aligned}$$

where $\|\cdot\|_1$ denotes the total variation norm and $\mathcal{L}_{\pi,r}(\mathbb{X}^n) = \frac{\mathrm{d}\mathbb{P}_{\pi,r}}{\mathrm{d}\mathbb{P}_0}(\mathbb{X}^n)$ is the corresponding likelihood ratio. The second inequality follows from Proposition 2.11 of [19]. The third inequality is true if the measure $\mathbb{P}_{\pi,r}$ is absolutely continuous with respect to $\mathbb{P}_0$, and the last one follows from the Cauchy–Schwarz inequality.

The prior $\mathbb{P}_{\pi,r}$ is the mixture of priors on $\Theta_p^\delta[r,\tau] \subset V_p^d$ and on the set $V_p^d = \bigcup_{m\in\mathcal{M}(p,d)} \Pi_m\mathbb{R}^d$, where m is uniformly distributed over $\mathcal{M}(d,p)$ according to the measure π. Define the prior $\mathbb{P}_{m,r}$ on the set $\Theta_p^\delta[r,\tau]\cap\Pi_m\mathbb{R}^d$ by letting

$$\delta_j = \frac{r}{\sqrt{p}}\varepsilon_j^m, \quad j = 1,\dots,d. \tag{30}$$

where $\varepsilon^m = (\varepsilon_1^m,\dots,\varepsilon_d^m)$ is a random vector with components

$$\varepsilon_j^m = \begin{cases} \pm 1, & j\in m \\ 0, & j\notin m \end{cases}$$

such that $\mathbb{P}\{\varepsilon_j^m = 1\} = \mathbb{P}\{\varepsilon_j^m = -1\} = 1/2$, $j\in m$ and $\mathbb{P}\{\varepsilon_j^m = 0\} = 1$, $j\notin m$. Note that $\|\varepsilon^m\|^2 = p$, therefore $\|\delta\|^2 = r^2$ and the prior is well defined. Thus, the likelihood ratio is given by

$$\mathcal{L}_{\pi,r}(\mathbb{X}^n) = \binom{d}{p}^{-1} \sum_{m\in\mathcal{M}(d,p)} \frac{\mathrm{d}\mathbb{P}_{m,r}}{\mathrm{d}\mathbb{P}_0}(\mathbb{X}^n).$$

To prove the lower detection bound, we need to show that $\liminf_{d\to\infty} \gamma(\psi,\Theta_p[r],\tau) \geq \gamma$ for any test ψ. Using the above lower bound inequality, we can see that it remains to show that $\limsup_{d\to\infty} \mathbb{E}_0[\mathcal{L}_{\pi,r}^2(\mathbb{X}^n)] \leq 1 + 4(1-\gamma)^2$ if (15) and (16) are satisfied. The proof of this statement is given in Lemma 5. ∎

LEMMA 5. *Let conditions (15) and (16) of Theorem 2 be satisfied for some $\gamma \in (0,1)$. Let $\mathcal{L}_{\pi,r}(\mathbb{X}^n)$ be the likelihood ratio that corresponds to the uniform prior on the subset of coordinates $m\in\mathcal{M}(d,p)$ and the prior distribution $\mathbb{P}_{m,r}$ defined in (30). Then*

$$\limsup_{d\to\infty} \mathbb{E}_0[\mathcal{L}_{\pi,r}^2(\mathbb{X}^n)] \leq 1 + 4(1-\gamma)^2.$$

ACKNOWLEDGEMENTS

This work was first presented at the "Conference on Structural Inference in Statistics" in September 2013. We would like to thank Roland Bacon, PI of the MUSE instrument for providing the galaxy spectra data. The authors are grateful to the reviewers and the editor for their valuable remarks and to André Ferrari, David Mary, Alexander Goldenshluger, Yuri Golubev, Anatoli Juditsky, Axel Munk, and David Siegmund for fruitful discussions.

SUPPLEMENTARY MATERIAL

Supplement A: Supplement to "High-dimensional change-point detection under sparse alternatives"
(doi: COMPLETED BY THE TYPESETTER; .pdf). The supplementary material [13] contains omitted proofs and some additional simulation results.

REFERENCES

[1] Addario-Berry, L., Broutin, N., Devroye, L. and Lugosi, G. (2010). On combinatorial testing problems. *The Annals of Statistics* **38** 3063–3092.

[2] Bacon, R., Brinchmann, J., Richard, J., Contini, T., Drake, A., Franx, M., Tacchella, S., Vernet, J., Wisotzki, L., Blaizot, J., Bouché, N., Bouwens, R., Cantalupo, S., Carollo, C. M., Carton, D., Caruana, J., Clément, B., Dreizler, S., Epinat, B., Guiderdoni, B., Herenz, C., Husser, T. O., Kamann, S., Kerutt, J., Kollatschny, W., Krajnovic, D., Lilly, S., Martinsson, T., Michel-Dansac, L., Patricio, V., Schaye, J., Shirazi, M., Soto, K., Soucail, G., Steinmetz, M., Urrutia, T., Weilbacher, P. and de Zeeuw, T. (2015). The MUSE 3D view of the Hubble Deep Field South. *Astronomy and Astrophysics* **575** A75.

[3] Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **5** 577–606.

[4] Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[5] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

[6] Bourguignon, S., Mary, D. and Slezak, E. (2011). Restoration of astrophysical spectra with sparsity constraints: models and algorithms. *IEEE Journal of Selected Topics in Signal Processing* **5** 1002-1013.

[7] Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric methods in change-point problems*. Kluwer Acad. Publ., the Netherlands.

[8] Butucea, C. and Ingster, Y. I. (2013). Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** 2652–2688.

[9] Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B* **76** 349-372.

[10] Carlstein, E., Müller, H. and Siegmund, D., eds. (1994). *Change-point problems. IMS Monograph* **23**.

[11] Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics.

[12] Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics* **32** 962–994.

[13] Enikeeva, F. and Harchaoui, Z. Supplement to "High-dimensional change-point detection under sparse alternatives".

[14] Harchaoui, Z., Vallet, F., Lung-Yut-Fong, A. and Cappé, O. (2009). A regularized kernel-based approach to unsupervised audio segmentation. In *ICASSP* 1665-1668.

[15] Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical learning with sparsity: the Lasso and generalizations*. CRC Press.

[16] Ingster, Y. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** 47–69.

[17] Ingster, Y., Pouet, C. and Tsybakov, A. (2009). Classification of sparse high-dimensional vectors. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys.* **367** 4427-4448.

[18] Ingster, Y. and Suslina, I. (2002). On a detection of a signal of known shape in multichannel system. *Zapiski Nauchn. Sem. POMI* **294** 88–112.

[19] Ingster, Y. and Suslina, I. (2003). *Nonparametric goodness-of-fit testing under Gaussian models*. Springer, New York.

[20] Korostelev, A. and Korosteleva, O. (2011). *Mathematical statistics. Asymptotic minimax theory.* Graduate Studies in Mathematics 119. Providence, RI: American Mathematical Society (AMS).

[21] Korostelev, A. and Lepski, O. (2008). On a multi-channel change point problem. *Math. Meth. Statist.* **17** 187–197.

[22] Lehmann, E. and Romano, J. (2005). *Testing statistical hypotheses (3rd ed.).* Springer.

[23] Levy-Leduc, C. and Roueff, F. (2009). Detection and localization of change-points in high-dimensional network traffic data. *The Annals of Applied Statistics* **3** 637–662.

[24] Meillier, C., Chatelain, F., Michel, O., Bacon, R., Piqueras, L., Bacher, R. and Ayasso, H. (2016). SELFI: an object-based, Bayesian method for faint emission line source detection in MUSE deep field data cubes. *Astronomy and Astrophysics* **588** A140.

[25] Olshen, A. B., Venkatraman, E. S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.

[26] Paris, S., Mary, D. and Ferrari, A. (2013). Detection Tests Using Sparse Models, With Application to Hyperspectral Data. *IEEE Trans. Signal Processing* **61** 1481–1494.

[27] Shiryaev, A. N. (1978). *Optimal Stopping Rules*. Springer-Verlag, New York.

[28] Siegmund, D. (1985). *Sequential analysis: tests and confidence intervals. Springer Series in Statistics*. Springer-Verlag.

[29] Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S. G., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., Graham, A., Southgate, J., Asselain, B., Allory, Y., Abbou, C. C., Albertson, D. G., Thiery, J. P., Chopin, D. K., Pinkel, D. and Radvanyi, F. (2006). Regional copy number-independent deregulation of transcription in cancer. *Nat. Genet.* **38** 1386–1396.

[30] Vert, J. P. and Bleakley, K. (2010). Fast detection of multiple change-points shared by many signals using group LARS. In *Advances in Neural Information Processing Systems 23* (J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta, eds.) 2343–2351. Curran Associates, Inc.

[31] Xie, Y. and Siegmund, D. (2013). Sequential multi-sensor change-point detection. *Ann. Statist.* **41** 670–692.

[32] Zhang, N. R., Siegmund, D. O., Ji, H. and Li, J. (2010). Detecting simultaneous change-points in multiple sequences. *Biometrika* **97** 631–645.

Université de Poitiers, CNRS
Laboratoire de Mathématiques et Applications
Téléport 2, Boulevard Marie et Pierre Curie
86962 Futuroscope Chasseneuil Cedex France
E-mail: farida.enikeeva@math.univ-poitiers.fr

University of Washington
Department of Statistics
Box 354322
Seattle, WA, USA
E-mail: zaid@uw.edu

A.A. Kharkevich Institute for Information Transmission Problems
Russian Academy of Sciences
Bolshoy Karetny per. 19-1, Moscow, 127051, Russia