



HAL
open science

Convex set detection

Victor-Emmanuel Brunel

► **To cite this version:**

| Victor-Emmanuel Brunel. Convex set detection. 2014. hal-00982900

HAL Id: hal-00982900

<https://hal.science/hal-00982900>

Preprint submitted on 24 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Convex Set Detection

Victor-Emmanuel Brunel
CREST, Paris (France) - University of Haifa, Haifa (Israel)
victor.emmanuel.brunel@ensae-paristech.fr

Abstract

We address the problem of one dimensional segment detection and estimation, in a regression setup. At each point of a fixed or random design, one observes whether that point belongs to the unknown segment or not, up to some additional noise. We try to understand what the minimal size of the segment is so it can be accurately seen by some statistical procedure, and how this minimal size depends on some *a priori* knowledge about the location of the unknown segment.

Keywords. change point, detection, hypothesis testing, minimax, separation rate, set estimation

1 Introduction

Consider the statistical model:

$$Y_i = \mathbb{1}(X_i \in G) + \xi_i, i = 1, \dots, n, \quad (1)$$

where $\mathbb{1}(\cdot)$ is the indicator function. The set made of the points X_i is called the design. The unknown set G is a segment on $[0, 1]$, the noise terms ξ_i are i.i.d. random variables, independent of the design. We distinguish two types of design:

- (DD) Deterministic, and regular design: $X_i = i/n, i = 1, \dots, n$;
- (RD) Random, uniform design: the variables $X_i, i = 1, \dots, n$, are i.i.d. uniform on $[0, 1]$.

In the sequel, the design will be denoted by \mathcal{X} . This model can be interpreted as a partial and noisy observation of an image, $[0, 1]$, in which there is an unknown object G . From this observation, one would like to determine

whether it is true that there is an object - the unknown set G might be empty -, and/or to recover that object, i.e. to estimate G . Another framework for this model is the noisy observation of some signal, here $\mathbb{1}(X_i \in G)$, and one would like to determine if the observation comes from a pure noise, or if there actually is some signal, and/or to recover that signal.

As it was mentioned in [Bru13], what makes difficult estimation of G is the complexity of the class of possible candidates, and detectability of G . The unknown set belongs to the class of all segments $[a, b]$ of $[0, 1]$. This is a parametric class, with two parameters $0 \leq a \leq b \leq 1$. Therefore, one may expect to be able to estimate G at the parametric speed $1/n$, up to a positive multiplicative constant. However, it turns out that G may be too hard to detect and thus, the speed of estimation would be deteriorated. For instance, if the segment is too small, and if no point of the design falls inside G , no procedure would see G . We would like to understand under which assumptions detection is not an obstacle for estimating G . In particular, the two following assumptions will be of interest for us :

Assumption 1. *The set G is of the form $[0, \theta]$, for some unknown number $0 \leq \theta \leq 1$.*

Assumption 2. $|G| \geq \mu$.

Here and in the rest of the paper, $\mu \in (0, 1)$ is a given positive number. The brackets $|\cdot|$ stand for the Lebesgue measure.

Let us discuss these two assumptions in order to understand, intuitively, why they facilitate detection of G . The first one gives some information on the location of G . In other words, it tells that the set G starts from the left side of the frame. The second one tells that G is not too small, and thus should not be unnoticed by the statistician. Model (1), together with Assumption 1, is well known under the name of the change point problem. It can be rewritten as :

$$Y_i = \mathbb{1}(X_i \leq \theta) + \xi_i, i = 1, \dots, n, \quad (2)$$

for some number $\theta \in [0, 1]$. The change point problem was studied in [KT93, Sec. 1.9], and a continuous-time version of this model is addressed in [Kor06]. The aim is to estimate the breakpoint θ . In the continuous time version, Korostelev [Kor06] proposed a more general framework. Instead of the indicator function in the regression equation (2), that is a function with a jump at the point θ , and satisfying a Lipschitz condition both on the left and on the right sides of θ . In these two works [Kor06], [KT93], the change point

θ is estimated with a precision, in expectation, of the order of $1/n$, if θ is assumed to be separated from 0 and 1 : $h \leq \theta \leq 1 - h$, for some $h \in (0, 1/2)$. Ibragimov and Khasminskii [IK84], under, among others, the same assumption of separation from 0 and 1, also proposed a consistent estimator of the discontinuity point of a regression function, with precision of order $1/n$ as well. This separation hypothesis is common in this kind of estimation problems. For instance, it is made in [KT92] and [KT93, Chap 3], where Tsybakov and Korostelev propose an estimator for boundary fragments. A boundary fragment, in dimension $d \geq 1$, is a subset G of \mathbb{R}^d which can be described as the subgraph of a positive function $g : [0, 1]^{d-1} \rightarrow [0, 1]$:

$$G = \left\{ (x_1, \dots, x_d) \in [0, 1]^d : 0 \leq x_d \leq g(x_1, \dots, x_{d-1}) \right\}. \quad (3)$$

The authors study a similar model to Model (1), where G is a boundary fragment instead of a segment. This is one possible generalization in higher dimensions of our problem. They assume that the true g belongs to some Hölder class, and is separated from 0 and 1:

$$h \leq g(x) \leq 1 - h, \forall x \in [0, 1]^{d-1},$$

for a given parameter $h \in (0, 1/2)$. In [KT92], the same authors estimate the support of a uniform density, assuming it is a boundary fragment. Again, they make the hypothesis that the underlying function g is separated from 0 and 1. In both models, they build a piecewise polynomial estimator of the function g , and prove that it is optimal in a minimax sense - see more details below -. One goal of the present work is to understand whether this separation hypothesis is necessary, in the simple case of dimension 1.

Two other extensions in higher dimension have been addressed in [Bru13]. A segment of $[0, 1]$ can be interpreted both as a one dimensional convex body, or as a one dimensional convex polytope. If the dimension is greater than 1, then the class of convex bodies is much bigger than that of convex polytopes. This is what explains that the optimal speed of estimation is better for polytopes than for general convex bodies - details in [Bru13] -. In this previous work, we did not study the impact of such assumptions as Assumptions 1 and 2. This question is beyond the scope of the present paper, but we believe that the results would be similar in the case of convex polytopes. Indeed, that case allows to keep the parametric property and, as we mentioned in [Bru13, Section 5], detectability becomes the main obstacle for estimation of the unknown set. In [Gay01], a testing problem is addressed, in a different model: the set of interest, as in [KT92], is the support of a uniform density, assumed to be a boundary fragment (see (3)). Of course, detection is not

pertinent in that model, since, as long as there are observations, there must be a nonempty support. However, the author studied the separation rate for distinguishing hypotheses of the type $H_0 : G = G_0$ and $H_1 : d(G, G_0) \geq h$, where G_0 is a given boundary fragment, d is the Nykodim distance between sets, defined below, and h is a given positive number which may depend on n .

In [CW13], the detection question is addressed in a slightly different framework. In Model (1), we assume that the strength of the signal is given, equal to 1. It is also interesting to deal with the case of a signal of unknown strength, i.e.

$$Y_i = \delta \mathbf{1}(X_i \in G) + \xi_i, i = 1, \dots, n, \quad (4)$$

where δ is a positive number. For the signal to be detectable, there should be a tradeoff between its length $|G|$ and its strength δ . Naturally, if δ is small, then the set G should be big enough and conversely, if δ is large, the set G is allowed to be small, so that the signal can be detected. Testing the presence of a signal, i.e. whether $\delta = 0$ or not, is considered in [CW13]. This work is concerned with the power of two tests: the scan - or maximum - likelihood ratio, and the average likelihood ratio. The two tests are compared in two regimes: signals of small scales, i.e. $|G| \rightarrow 0$, and signals of large scales, i.e. $\liminf_{n \rightarrow \infty} |G| > 0$. The design is (DD), and it is proved that if

$$\delta \sqrt{n|G|} \geq \sqrt{2 \ln \frac{1}{|G|}} + b_n, \text{ for some sequence } b_n \text{ such that } b_n \rightarrow \infty, \text{ then}$$

there is a test whose power is asymptotically 1. Note that here, $\delta \sqrt{|G|}$ is exactly the L^2 -norm of the signal, $\|\delta \mathbf{1}(\cdot \in G)\|_2 = \delta \sqrt{|G|}$. In [LT00], signals of unknown shape but known smoothness were considered. Exact minimax separation rates, in terms of the L^2 -norm of the signal, for distinguishing the null hypothesis, under which observations are pure noise, and the alternative one, under which there is a signal, are given. Detection is harder in that framework, because unlike in Model (1) or (4), where the shape of the signal is known - it is piecewise constant -, only its smoothness is known, and the separation rates are larger than those of models (1) and (4), in the sense that they allow less freedom for the size of the signal. However, this problem is different from ours, since we are concerned with the location of the signal, not the signal itself.

Model 1 deals with change points in the mean of the observations, conditionally to the design. Under Assumption 1, there is only one change point, and under Assumption 2, there are two. A problem of interest in time series analysis is that of detecting change points in the mean, or in the covari-

ance function, or other characteristics of the series. We refer to [SZ10] and the references therein. In [FMSnt], a sample of n independent observations Y_1, \dots, Y_n is given, and one assumes that Y_i admits a density $f(\cdot, \vartheta(\frac{i}{n}))$ with respect to a given measure, for $i = 1, \dots, n$, where f belong to an exponential parametric class of densities. The real valued function ϑ is assumed to be piecewise constant on $[0, 1]$, with a finite number K of jumps, not necessarily known. Under a similar condition on δ and $|G|$ to that of [CW13], it is shown that at least one change point is consistently detectable: the proposed estimator of K is positive with probability that goes to 1, as $n \rightarrow \infty$. When $f(\cdot, \theta)$ is the density of a Gaussian distribution with mean θ and given variance $\sigma^2 > 0$, the problem was addressed in [Leb03]. This model includes models (1) and (4), when the design is (DD). However, Assumption 2 is not considered in that work, and optimality of the estimator of the change point is not treated, in the case when it is assumed to be unique - which corresponds to Assumption 1 - or when it is known that there are only two of them.

1.1 Notation

If G is a segment of $[0, 1]$, we denote, respectively, by \mathbb{P}_G , \mathbb{E}_G and \mathbb{V}_G the joint probability measure of the observations $((X_1, Y_1), \dots, (X_n, Y_n))$ that come from Model (1), and the corresponding expectation and variance operators. If the event inside the probability, expectation or variance sign does not depend on G , we omit the subscript and write only \mathbb{P} , \mathbb{E} and \mathbb{V} .

If G_1 and G_2 are two segments of $[0, 1]$, we denote by $G_1 \Delta G_2$ their symmetric difference. The Lebesgue measure of the symmetric difference of G_1 and G_2 , $|G_1 \Delta G_2|$, is also called the Nykodom distance between G_1 and G_2 .

In the present work, we do two types of inference on the underlying set G : hypothesis testing and estimation. The tests consist of deciding whether G is empty or not. Estimation consists of giving an estimator of G . A subset \hat{G}_n of $[0, 1]$, whose construction depends on the sample is called a set estimator or, more simply, an estimator. Given an estimator \hat{G}_n , we measure its accuracy in a minimax framework. The risk of \hat{G}_n on a class \mathcal{C} of Borel subsets of \mathbb{R} is defined as

$$\mathcal{R}_n(\hat{G}_n; \mathcal{C}) = \sup_{G \in \mathcal{C}} \mathbb{E}_G[|G \Delta \hat{G}_n|]. \quad (*)$$

The rate (a sequence depending on n) of an estimator on a class \mathcal{C} is the speed at which its risk converges to zero when the number n of available

observations tends to infinity. For all the estimators defined in the sequel, we are interested in upper bounds on their risk, in order to get information about their speed of convergence. The minimax risk on a class \mathcal{X} , when n observations are available, is defined as

$$\mathcal{R}_n(\mathcal{C}) = \inf_{\hat{G}_n} \mathcal{R}_n(\hat{G}_n; \mathcal{C}), \quad (**)$$

where the infimum is taken over all set estimators depending on n observations. If $\mathcal{R}_n(\mathcal{C})$ converges to zero, we call the minimax rate of convergence on the class \mathcal{C} the speed at which $\mathcal{R}_n(\mathcal{C})$ tends to zero. It is interesting to provide a lower bound for $\mathcal{R}_n(\mathcal{C})$. By definition, no estimator can achieve a better rate on \mathcal{C} than that of the lower bound. This bound gives also information on how close the risk of a given estimator is to the minimax risk. If the rate of the upper bound on the risk of an estimator matches the rate of the lower bound on the minimax risk on the class \mathcal{C} , then the estimator is said to have the minimax rate of convergence, or to be optimal in the minimax sense - up to constants - on this class.

A test consists of deciding whether to reject or not a given hypothesis, called the null hypothesis, when it is compared to an alternative one. Let $h \in (0, 1)$. In the whole paper, we will consider the following null hypothesis:

$$H_0 : G = \emptyset,$$

and the alternative hypothesis:

$$H_1 : |G| \geq h.$$

Testing H_0 against H_1 is equivalent to deciding whether there is a set G or not. A test is a random variable τ_n , which is built from the data set, and whose possible values are 0 and 1. The decision associated to the test τ_n is to reject H_0 if and only if $\tau_n = 1$. We measure the performance of a test τ_n on a class \mathcal{C} using

$$\gamma_n(\tau_n, \mathcal{C}) = \mathbb{P}_\emptyset[\tau_n = 1] + \sup_{G \in \mathcal{C}, |G| \geq h} \mathbb{P}_G[\tau_n = 0].$$

This quantity is the sum of the errors of the first and the second kinds of the test τ_n . We say that τ_n is consistent on the class \mathcal{C} if and only if $\gamma_n(\tau_n, \mathcal{C}) \rightarrow 0$, when $n \rightarrow \infty$. Let us allow the number h to depend on n . We call the separation rate on the class \mathcal{C} any sequence of positive numbers r_n such that:

- if $\frac{h}{r_n} \xrightarrow{n \rightarrow \infty} \infty$, then there exists a consistent test on \mathcal{C} , and
- if $\frac{h}{r_n} \xrightarrow{n \rightarrow \infty} 0$, then no test is consistent test on \mathcal{C} .

With regard to both Assumptions 1 and 1, we focus on three different classes of sets, which are defined as bellow:

- $\mathcal{S} = \{[a, b] : 0 \leq a \leq b \leq 1\}$ is the class of all segments on $[0, 1]$,
- $\mathcal{S}_0 = \{[0, \theta] : 0 \leq \theta \leq 1\}$ is the class of all segments on $[0, 1]$, satisfying Assumption 1,
- $\mathcal{S}(\mu) = \{G \in \mathcal{S} : |G| \geq \mu\}$ is the class of all segments on $[0, 1]$, satisfying Assumption 2.

In the whole paper, we assume that the noise terms ξ_1, \dots, ξ_n from Model (1) satisfy

$$\mathbb{E} \left[e^{u\xi_i} \right] \leq e^{\sigma^2 u^2/2}, \forall u \in \mathbb{R}, \quad (5)$$

for some positive constant $\sigma > 0$. This constant need not be known. By satisfying this inequality, the noise is said to be subgaussian. An important case of such subgaussian variables is zero mean Gaussian variables, for which σ^2 turns to be the variance.

Since the design and the noise are assumed to be independent, reordering the X_i 's does not modify the model. Indeed, there exists a reordering $\{i_1, \dots, i_n\}$ of $\{1, \dots, n\}$, such that $X_{i_1} \leq \dots \leq X_{i_n}$. The random indexes i_1, \dots, i_n are independent of the noise, and therefore the new noise vector $(\xi_{i_1} \leq \dots \leq \xi_{i_n})$ has the same distribution as $(\xi_1 \leq \dots \leq \xi_n)$. With regard to this remark, we assume from now on that \mathcal{X} is the reordering of a preliminary design, and therefore $X_1 \leq \dots \leq X_n$ almost surely, without loss of generality.

For two real valued sequences A_n and B_n , and a parameter ϑ , which may be multidimensional, we will write $A_n \asymp_{\vartheta} B_n$ to say that there exist positive constants $c(\vartheta)$ and $C(\vartheta)$ which depend on ϑ , such that $c(\vartheta)B_n \leq A_n \leq C(\vartheta)B_n$, for n large enough. If we put no subscript under the sign \asymp , this means that the involved constants are universal, i.e. do not depend on any parameter.

1.2 Contributions of the paper

In this paper, we propose three new results.

1. The first result is a preliminary to the rest of the paper, and has been already mostly covered in the existing literature. Model (1) is a simple change point problem. If Assumption 1 is satisfied, the change point is unique. Otherwise, it is known that there are two change points, and the size of the jump is also known. In a hypothesis testing setup, we give the the minimal size - as a function of n - of the unknown set G , in Model (1), must satisfy, for that set to be detectable, depending on Assumption 1 being satisfied or not. We prove our results under both designs (DD) and (RD).
2. The second result concerns the change point problem is not necessary. We show that the speed $1/n$ of convergence in estimating the break-point θ in (2) can be achieved without assuming $a \leq \theta \leq b$ for some positive numbers $0 < a \leq b < 1$.
3. The last result shows that recovering a set requires that set to be sufficiently big, in order to achieve the parametric rate $1/n$ uniformly. We show that without this assumption, a logarithmic factor appears in the minimax rate.

Our work is made in the minimax setup. We provide exact rates of convergence of the minimax risks on the classes which we study. These rates are obtained by precise deviation inequalities for all the estimators that we propose. In the second and third results, the minimax rates correspond exactly to the separation rates given by the first result, for testing emptiness of the set G .

1.3 Organization of the paper

This paper is organized as follows. Next section deals with the question of deciding whether there is a set or not. We give the separation rates for the corresponding detection test, for the change point problem - i.e. under Assumption 1 - and on the class \mathcal{S} . The third section deals with estimating the unknown set, and we give the minimax rates of estimation on the classes \mathcal{S} , \mathcal{S}_0 and $\mathcal{S}(\mu)$. The fourth section is dedicated to the conclusion, followed by a discussion about possible extensions. The fifth and last section is devoted to the proofs.

2 Detection of a set

We test the null hypothesis $H_0 : G = \emptyset$, against the alternative hypothesis $H_1 : |G| \geq h$. We find the asymptotic minimal value of h , as a function of n , so the two hypotheses are well separated and there exists a consistent test. This minimal value depends on the class of possible sets G . Intuitively, the *a priori* knowledge that the unknown set G belongs to the class \mathcal{S}_0 gives an important piece of information about the location of this set, and therefore it should be easier to detect it. Actually, the following theorem confirms this intuition, by showing that the separation rate is smaller - by a logarithmic factor - for the subclass \mathcal{S}_0 than for the whole class \mathcal{S} .

The idea, for the class \mathcal{S}_0 , is the following. Under H_1 , $[0, h] \subseteq G$. Therefore, we check among those observations (X_i, Y_i) for which $X_i \leq h$ if there is a sufficiently large number of Y_i 's that are large, e.g. larger than $1/2$. Let $N = \max\{i = 1, \dots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$. Let S be the following test statistics:

$$S = \#\{i = 1, \dots, N : Y_i \leq \frac{1}{2}\}.$$

If the alternative hypothesis holds, i.e. if $|G| \geq h$, all the $X_i, i = 1, \dots, N$ fall inside the set G , and the corresponding Y_i should not be too small. The test statistic S counts how many of these Y_i 's are suspiciously small. This is how is built the test T_n^0 :

$$T_n^0 = \mathbb{1}(S \leq cN),$$

where $c \in (\mathbb{P}[\xi_1 \leq -1/2], \mathbb{P}[\xi_1 \leq 1/2])$.

For the class \mathcal{S} , we propose a scan test, i.e., a procedure which scans the whole frame $[0, 1]$ and seeks for a large enough quantity of successive observations for which Y_i is large. If $G \in \mathcal{S}$, let $R(G) = \sum_{i=1}^n Y_i \mathbb{1}(X_i \in G) - \frac{\#(\mathcal{X} \cap G)}{2}$, and $R = \sup_{|G| \geq h} R(G)$. Under the alternative hypothesis, R should be quite large, and we define the test $T_n^1 = \mathbb{1}(R \geq 0)$.

Theorem 1. *Let Model (1) hold.*

1. *Assume that the design is (DD) or (RD), and that the noise satisfies:*

$$\mathbb{P}[\xi_1 \leq -1/2] < \mathbb{P}[\xi_1 \leq 1/2].$$

Then, if $nh \rightarrow \infty$, the test T_n^0 is consistent, i.e. $\gamma_n(T_n^0, \mathcal{S}_0) \rightarrow 0$. If, in addition, the noise is Gaussian, then a separation rate on the class \mathcal{S}_0 is $r_n = 1/n$.

2. Assume that the design is (DD) or (RD). Then, if $nh/\ln n \rightarrow \infty$, the test T_n^1 is consistent, i.e. $\gamma_n(T_n^1, \mathcal{S}) \rightarrow 0$. If, in addition, the noise is Gaussian, then a separation rate on the class \mathcal{S} is $r_n = \ln(n)/n$.

In the next section, we show that the separation rates given in Theorem 1 are the minimax rates of convergence on the corresponding classes.

3 Estimation of a set

3.1 Least square estimators

Let Model (1) hold. For $G' \in \mathcal{S}$, let $A_0(G') = \sum_{i=1}^n (Y_i - \mathbf{1}(X_i \in G'))^2$ be the sum of squared errors. A way to estimate G is to find a random set \hat{G}_n which minimizes $A_0(G')$, among all possible candidates G' . Note that minimizing $A_0(G')$ is equivalent to maximizing

$$A(G') = \sum_{i=1}^n (2Y_i - 1)\mathbf{1}(X_i \in G'). \quad (6)$$

Denote by $S = \{i = 1, \dots, n : X_i \in G\}$ and by $S' = \{i = 1, \dots, n : X_i \in G'\}$, for some $G' \in \mathcal{S}$. Denote by $\#(\cdot)$ the cardinality, for finite sets. The criterion $A(G')$ becomes, if denoted as a function of S' ,

$$\begin{aligned} A(S') &= \sum_{i \in S'} (2Y_i - 1) \\ &= \sum_{i \in S'} (2\mathbf{1}(X_i \in G) + 2\xi_i - 1) \\ &= 2\#(S \cap S') - \#S' + 2 \sum_{i \in S'} \xi_i, \end{aligned}$$

so,

$$A(S') - A(S) = -\#(S \Delta S') + 2 \left(\sum_{i \in S' \setminus S} \xi_i - \sum_{i \in S \setminus S'} \xi_i \right). \quad (7)$$

A subset S' of $\{1, \dots, n\}$ is called convex if and only if it is of the form $\{i, \dots, j\}$, for some $1 \leq i \leq j \leq n$. It is clear that if a convex subset S' of $\{1, \dots, n\}$ maximizes $A(S') - A(S)$, then by defining $G' = [X_{\min S'}, X_{\max S'}]$, the segment G' maximizes $A(G')$ (cf. (6)).

3.2 Estimation of the change point

Let Model (1) hold, with design (DD). Assume that G belongs to \mathcal{S}_0 . This is the change point problem. For some $\theta \in [0, 1]$, G can be written as $G = [0, \theta]$. Let us make one preliminary remark. For any estimator \hat{G}_n of G , the random segment $\tilde{G}_n = [0, \sup \hat{G}_n]$ performs better than \hat{G}_n , since $|\tilde{G}_n \triangle G| \leq |\hat{G}_n \triangle G|$ almost surely. Therefore, it is sufficient to consider only estimators of the form $\hat{G}_n = [0, \hat{\theta}_n]$, where $\hat{\theta}_n$ is a random variable. Then, $|\hat{G}_n \triangle G| = |\hat{\theta}_n - \theta|$, and the performance of the estimator \hat{G}_n of G is that of the estimator $\hat{\theta}_n$ of θ . Let us build a least square estimator (LSE) of θ . For $M = 1, \dots, n$, let

$$\begin{aligned} F(M) &= A(\{1, \dots, M\}) \\ &= \sum_{i=1}^M (2Y_i - 1). \end{aligned}$$

Let $\hat{M}_n \in \text{ArgMax}_{M=1, \dots, n} F(M)$, and $\hat{\theta}_n = X_{\hat{M}_n}$. The following theorem follows.

Theorem 2. *Let $n \geq 1$. Let Model (1) hold, with design (DD). Let $\hat{G}_n = [0, \hat{\theta}_n]$. Then,*

$$\sup_{G \in \mathcal{S}_0} \mathbb{P}_G \left[|\hat{G}_n \triangle G| \geq \frac{x}{n} \right] \leq C_0 e^{-x/(8\sigma^2)}, \forall x > 0,$$

where C_0 is a positive constant which depends on σ only.

A simple application of Fubini's theorem leads to the following result.

Corollary 1. *Let the assumptions of Theorem 2 be satisfied. Then, for all $q > 0$, there exists a positive constant A_q which depends on q and σ only, such that*

$$\sup_{G \in \mathcal{S}_0} \mathbb{E}_G \left[|\hat{G}_n \triangle G|^q \right] \leq \frac{A_q}{n^q}.$$

This corollary shows that the minimax risk on the class \mathcal{S}_0 is bounded from above by $1/n$, up to multiplicative constants. Next theorem proves that up to constants, $1/n$ is also a lower bound on the minimax risk, if the noise is Gaussian.

Theorem 3. *Consider Model (1), with design (DD). Then for all integer $n \geq 1$,*

$$\mathcal{R}_n(\mathcal{S}_0) \geq \frac{1}{2n}.$$

Combining Theorems 2 and 3 yields

Theorem 4. *Consider Model (1), with design (DD). Then, the minimax risk on the class \mathcal{S} satisfies*

$$\mathcal{R}_n(\mathcal{S}_0) \asymp_\sigma \frac{1}{n}.$$

3.3 Recovering any set

Let us now assume that the unknown set G does not necessarily contain 0. We shall prove that whether to assume that G belongs to the class $\mathcal{S}(\mu)$ or not does not lead to the same minimax rate. As we saw in Section 3.1, an estimator of G in Model (1) can be obtained by maximizing the Gaussian process (7) over all segments of $\{1, \dots, n\}$. This is not the track that we will borrow, but it would be interesting to work precisely on this process. This would probably be the first step to extensions of our results in higher dimensions. However, this problem remains open for now. The methods that we develop in this section are quite different. If G is only assumed to belong to the biggest class \mathcal{S} , the proposed estimator is the LSE, which was already detailed in [Bru13] for convex polytopes, in higher dimension. If $|G|$ is *a priori* known to be greater or equal to μ , then we first build a preliminary estimator of G - the LSE -, using one half of the observed sample. This estimator is not optimal, but it is close to G with high probability. We show that the middle point \hat{m}_n of this estimator is in G with high probability. This brings us back to the change point problem, where 0 is now replaced by \hat{m}_n , and we use the second half of the observed sample to estimate two change points.

Let us first state the following theorem, which is, for the design (RD), a particular case of [Bru13, Theorem 1], for $d = 1$.

Theorem 5. *Let $n \geq 2$. Let Model hold, with design (DD) or (RD). Let $\hat{G}_n \in \underset{G' \in \mathcal{S}}{\text{ArgMax}} A(G')$ be a LSE estimator of G . Then, there exist two positive constants C_1 and C_2 which depend on σ only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{P}_G \left[n \left(|\hat{G}_n \triangle G| - \frac{4 \ln n}{C_2 n} \right) \geq x \right] \leq C_1 e^{-C_2 x}, \forall x > 0.$$

The expressions of C_1 and C_2 are given in the proof of [Bru13, Theorem 1], for the design (RD). For the design (DD), we do not give a proof of this theorem here, but it can be easily adapted from that of the case of the design (RD). The next corollary is immediate.

Corollary 2. *Let the assumptions of Theorem 5 be satisfied. Then, for all $q > 0$, there exists a positive constant B_q which depends on q and σ only, such that*

$$\sup_{G \in \mathcal{S}} \mathbb{E}_G \left[|\hat{G}_n \triangle G|^q \right] \leq B_q \left(\frac{\ln n}{n} \right)^q.$$

This corollary shows that the minimax risk on the class \mathcal{S} is bounded from above by $\ln(n)/n$, up to a multiplicative constant. The following theorem establishes a lower bound, if the noise is supposed to be Gaussian.

Theorem 6. *Consider Model (1), with design (DD) or (RD). Assume that the noise terms ξ_i are i.i.d. Gaussian random variables, with variance $\sigma^2 > 0$. For any large enough n ,*

$$\mathcal{R}_n(\mathcal{S}) \geq \frac{\alpha^2 \sigma^2 \ln n}{n},$$

where α is a universal positive constant.

This lower bound comes from [Bru13, Theorem 2] in the case of the design (RD), and the proof is easily adapted for the design (DD). Eventually, the minimax risk on the class \mathcal{S} is of the order $\ln(n)/n$:

Theorem 7. *Consider Model (1), with design (DD) or (RD). Assume that the noise terms ξ_i are i.i.d. Gaussian random variables, with variance $\sigma^2 > 0$. The minimax risk on the class \mathcal{S} satisfies, asymptotically:*

$$\mathcal{R}_n(\mathcal{S}) \asymp_{\sigma} \frac{\ln n}{n}.$$

For the design (DD), we combine both Theorems 2 and 5 to find the minimax rate on the class $\mathcal{S}(\mu)$. Let Model (1) hold, and let $G \in \mathcal{S}(\mu)$. First, we split the sample into two equal parts. Let \mathcal{D}_0 be the set of sample points with even indexes, and \mathcal{D}_1 the set of sample points with odd indexes. Note that $\mathcal{D}_0 \cup \mathcal{D}_1$ is exactly the initial sample, that these two subsample are independent, and that each of them is made of at least $(n-1)/2$ data. Let \hat{G}_n be the LSE estimator of G given in Theorem 5, built from the subsample \mathcal{D}_0 . Let \hat{m}_n be the middle of \hat{G}_n . As it will be shown in the proof of the next theorem, \hat{m}_n satisfies both following properties, with high probability:

1. $\hat{m}_n \in G$,
2. $\mu/2 \leq \hat{m}_n \leq 1 - \mu/2$.

This brings us to estimating two change points - the endpoints of G - , using the second subsample \mathcal{D}_1 . From Theorem 2, we know that this can be done at the speed $1/n$, up to multiplicative constants.

Theorem 8. *Consider Model (1), with design (DD). There exists an estimator \tilde{G}_n of G , such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{P}_G \left[|\tilde{G}_n \Delta G| \geq \frac{x}{n} \right] \leq 2C_0 e^{-\mu x / (256\sigma^2)} + C_1 n^4 e^{-C_2 \mu n / 2}, \forall x > 0,$$

for n large enough. The positive constants C_0 and C_2 appeared in Theorems 2 and 5.

Naturally, Theorem 8 leads to the next corollary.

Corollary 3. *Let the assumptions of Theorem 8 be satisfied. Then, for all $q > 0$, there exists a positive constant B'_q which depends on q , μ and σ only, such that*

$$\sup_{G \in \mathcal{S}(\mu)} \mathbb{E}_G \left[|\tilde{G}_n \Delta G|^q \right] \leq \frac{B'_q}{n^q}.$$

This corollary, for $q = 1$, shows that the minimax risk on the class $\mathcal{S}(\mu)$ is bounded from above by $1/n$, up to a multiplicative constant. A very similar proof to that of Theorem 3 yields a lower bound for this minimax risk, which leads to the next theorem.

Theorem 9. *Consider Model (1), with design (DD). The minimax risk on the class $\mathcal{S}(\mu)$ satisfies, asymptotically,*

$$\mathcal{R}_n(\mathcal{S}(\mu)) \asymp_{\mu, \sigma} \frac{1}{n}.$$

Remark 1. *Note that, in Theorem 8, the upper bound contains one residual term which does not depend on x . This term, in order to be sufficiently small, requires that μ - if allowed to depend on n - is of larger order than $\ln(n)/n$. This is an echo to Theorem 1, in which we showed that the smallest set which can be detected has measure of this order exactly. In addition, if μ is of the order of $\ln(n)/n$, then the proof of the lower bound of Theorem 6 can be applied, and the minimax risk on the class $\mathcal{S}(\mu)$ will satisfy $\mathcal{R}_n(\mathcal{S}(\mu)) \asymp_{\sigma} \frac{\ln n}{n}$.*

4 Conclusion and discussion

We summarize our results in Table 1. The rates that are written in this table hold for Gaussian noise, which is the most important case. In each case we indicate the design for which the rate holds.

	\mathcal{S}	\mathcal{S}_0	$\mathcal{S}(\mu)$
Minimax rate	$\ln(n)/n$ (DD,RD)	$1/n$ (DD)	$1/n$ (DD)
Separation rate	$\ln(n)/n$ (DD,RD)	$1/n$ (DD,RD)	.

Table 1: Minimax risks and separation rates for the classes \mathcal{S} , \mathcal{S}_0 and $\mathcal{S}(\mu)$.

Note that in two cases, only the design (DD) has been considered. This is mainly for technical reasons, and we believe that the rates are still the same for the design (RD).

It comes out that asymptotically, a segment can be estimated infinitely faster when it is *a priori* supposed either to contain a given point (here, 0), or to be large enough. The main question that remains is: does this phenomenon still hold for two - or higher - dimensional sets? An important - if not essential - assumption which has been done all over this paper is the convexity of the unknown set. In dimension 1, the class of convex subsets of $[0, 1]$ is simple, and parametric. In any higher dimension d , the class \mathcal{C}_d of convex subsets of $[0, 1]^d$ is much more complex. In particular, its metric entropy is much larger than that of a parametric family [Bro76], and it seems to us that it is the complexity of this class that makes it harder to estimate a set, than detectability.

In Model (1), if G belongs to \mathcal{C}_d , estimation of G can be done at the minimax rate $n^{-2/(d+1)}$ [Bru13]. However, an adaptation of the proof of [Bru13, Theorem 2] shows the following: on any subclass of \mathcal{C}_d invariant under translations and invertible affine transformations - which keep a set G inside the frame $[0, 1]^d$ -, the minimax rate is at least of order $\ln(n)/n$. In addition, we believe this is the separation rate for the detection problem, on any such subclass of \mathcal{C}_d .

For the whole class \mathcal{C}_d , since $\ln(n)/n$ is much smaller than $n^{-2/(d+1)}$, the minimax rate is of the order of $n^{-2/(d+1)}$. However, for a parametric subclass, such as that of all convex polytopes with a given number of vertices, we believe that the complexity of the class leads to a term of order $1/n$ in the minimax risk, although detectability leads to a term of order $\ln(n)/n$. In our opinion, this is what explains that, as shown in [Bru13], the minimax

risk on the class of all convex polytopes of $[0, 1]^d$, with a given number of vertices, is of order $\ln(n)/n$.

Motivated by Theorem 8, we also conjecture that $1/n$ is the minimax rate on the class of all convex polytopes of $[0, 1]^d$, of volume greater than a given $\mu > 0$. However, it is not possible to extend Theorem 8 to higher dimensions. An adaptation of the proof of [Bru13, Theorem 2], by taking, as the M hypotheses used in the proof, sets which contain the origin and have pairwise zero measure intersections, would show that $(\ln n)/n$ remains a lower bound under that assumption. Yet, we believe that if the unknown set is assumed to contain a given section of positive $d - 1$ dimensional Lebesgue measure, of a given hyperplane in \mathbb{R}^d , then an analog of Theorem 8 should hold, and the minimax rate should be of order $1/n$.

To end this discussion, we note that Model (4) can be rewritten in the vector form:

$$y = \beta + \xi, \tag{8}$$

where $y = (Y_1, \dots, Y_n)^\top$, $\beta = (\delta \mathbf{1}(X_1 \in G), \dots, \delta \mathbf{1}(X_n \in G))^\top \in \{0, \delta\}^n$ and $\xi = (\xi_1, \dots, \xi_n)^\top$. If we denote by A the $n \times n$ lower triangular matrix with coefficients on and under the main diagonal equal to 1, equation (8) can be rewritten as:

$$y = A\gamma + \xi, \tag{9}$$

where $\gamma = (\beta_1, \beta_2 - \beta_1, \dots, \beta_n - \beta_{n-1})^\top$ is the vector of differences of β . The vector γ is sparse: it has exactly two nonzero coefficients, corresponding to the change points locations. This formulation of our initial problem leads to a high dimensional regression setup, under sparsity constraints. If δ is known, equal to 1 -this is Model (1)-, we are not interested in the estimation of the vector γ itself, but only of its support, which indicates the location of G , if the design is (DD). If δ is unknown, the two nonzero coefficients of γ take values δ and $-\delta$, and it is also of interest to estimate γ itself, in addition to its support. Penalized regression methods, with Lasso-type penalizations (see [Tib96] for details about Lasso), have been addressed in [MvdG97], [BKL⁺09], among others. In these two works, the penalization is written in terms of the number of jumps in the vector β in (8), which is equal, for Model (1), to the number of nonzero coefficients of γ in (9). Therefore, the penalty is equivalent to a L_0 or BIC-type penalty, see for example [BTW07]. Estimation of the support is usually a secondary question, which is addressed in terms of consistency, and not of rate of convergence. The main focus is about estimation of γ , or prediction, i.e., estimation of β . We believe that the parameter μ , if both μ and G are unknown, can be estimated at the speed $\sqrt{(\ln n)/n}$. In that case, some results about estimation rates of G , for

the design (DD), are given in [BKL⁺09], but the risks, in expectation, are not computed.

It would be interesting to understand how Lasso or BIC-type estimators could be adapted in order to achieve the rate $1/n$ on the classes \mathcal{S}_0 and $\mathcal{S}(\mu)$. The penalties of the classical estimators do not use the information that is contained in those classes, and we believe that these estimators are suboptimal. However, one should try to modify the penalties, according to that information, in order to improve the accuracy of these estimators.

5 Proofs

5.1 Proof of Theorem 1

On the class \mathcal{S}_0

Upper bound Let us first prove the upper bound, i.e. assume that $nh \rightarrow \infty$, and prove that there exists a consistent test. Recall that $N = \max\{i = 1, \dots, n : X_i \leq h\} = \#(\mathcal{X} \cap [0, h])$. If the design is (DD), then N is just equal to the integer part of nh . If the design (RD), then N is a binomial random variable, with parameters n and h . Let us show first that the error of the first kind of the test T_n^0 goes to zero, when $n \rightarrow \infty$.

$$\begin{aligned} \mathbb{P}_\emptyset [S \leq cN] &= \mathbb{P}_\emptyset \left[\# \left\{ i = 1, \dots, N : Y_i > \frac{1}{2} \right\} \geq (1-c)N \right] \\ &\leq \mathbb{E} \left[\mathbb{P}_\emptyset \left[\# \left\{ i = 1, \dots, N : \xi_i > \frac{1}{2} \right\} \geq (1-c)N \mid \mathcal{X} \right] \right]. \end{aligned}$$

Since the ξ_i 's are independent of \mathcal{X} , the distribution of $\#\{i = 1, \dots, N : \xi_i > \frac{1}{2}\}$ conditionally to \mathcal{X} is binomial, with parameters N and β , where $\beta = \mathbb{P}[\xi_1 > 1/2] \in [0, 1)$. Thus, by Bernstein's inequality for binomial random variables, by defining $\gamma = \frac{(1-c-\beta)^2}{2\beta(1-\beta) + (1-c-\beta)/3} > 0$,

$$\mathbb{P}_\emptyset [S \leq cN] \leq \mathbb{E} [\exp(-\gamma N)].$$

If \mathcal{X} satisfies (DD), then $N \geq nh - 1$ and it is clear that $\mathbb{P}_\emptyset [S \leq cN] \rightarrow 0$. If \mathcal{X} satisfies (RD), then

$$\mathbb{E} [\exp(-\gamma N)] = \exp(-nh(1 - e^{-\gamma})),$$

so $\mathbb{P}_\emptyset [S \leq cN] \rightarrow 0$.

Let us show, now, that the error of the second kind goes to zero as well. Let $G \in \mathcal{S}_0$ satisfying the alternative hypothesis, i.e. $|G| \geq h$. Denote by $\beta' = \mathbb{P}[\xi_1 \leq -1/2]$, and by $\gamma' = \frac{(c - \beta')^2}{2\beta'(1 - \beta') + (c - \beta')/3} > 0$

$$\begin{aligned} \mathbb{P}_G[S > cN] &= \mathbb{P}_G \left[\# \left\{ i = 1, \dots, N : Y_i \leq \frac{1}{2} \right\} > cN \right] \\ &\leq \mathbb{E} \left[\mathbb{P}_\emptyset \left[\# \left\{ i = 1, \dots, N : \xi_i \leq -\frac{1}{2} \right\} > cN \mid \mathcal{X} \right] \right] \\ &\leq \mathbb{E} \left[\exp(-\gamma'N) \right], \end{aligned}$$

by a similar computation to that for the error of the first kind. Since the right-side of the last inequality does not depend on G ,

$$\sup_{|G| \geq h} \mathbb{P}_G[S > cN] \leq \mathbb{E} \left[\exp(-\gamma'N) \right]$$

and therefore, by the same argument as for the error of the first kind, goes to zero when $n \rightarrow \infty$, for both designs (DD) and (RD).

Lower bound Assume, now, that $nh \rightarrow 0$. Let τ_n be any test. Let $G_1 = [0, h]$. We denote by \mathcal{H} the Hellinger distance between probability measures. The following computation uses properties of this distance, which can be found in [Tsy09].

$$\begin{aligned} \gamma_n(\tau_n, \mathcal{C}) &\geq \mathbb{E}_\emptyset[\tau_n] + \mathbb{E}_{G_1}[1 - \tau_n] \\ &\geq \int \min(d\mathbb{P}_\emptyset, d\mathbb{P}_{G_1}) \\ &\geq \frac{1}{2} \left(1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})}{2} \right)^2. \end{aligned} \quad (10)$$

Let $G, G' \in \mathcal{S}$. A simple computation shows that, for the design (DD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_G, \mathbb{P}_{G'})}{2} = \exp \left(-\frac{\#(\mathcal{X} \cap (G \Delta G'))}{8\sigma^2} \right), \quad (11)$$

and for the design (RD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_G, \mathbb{P}_{G'})}{2} = \left(1 - \left(1 - e^{-\frac{1}{8\sigma^2}} \right) |G \Delta G'| \right)^n. \quad (12)$$

In particular, for the design (DD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1})}{2} \geq \exp \left(-\frac{nh}{8\sigma^2} \right),$$

and for the design (RD),

$$1 - \frac{\mathcal{H}(\mathbb{P}_\emptyset, \mathbb{P}_{G_1}^{\otimes n})}{2} = \left(1 - \left(1 - e^{-\frac{1}{8\sigma^2}}\right)h\right)^n.$$

In both cases, we showed that the right side of (10) tends to $1/2$, when $n \rightarrow \infty$. Therefore the test τ_n is not consistent.

On the class \mathcal{S}

Upper bound Assume that $\frac{nh}{\ln n} \rightarrow \infty$. Let us first show that the error of the first kind of T_n^1 goes to zero, when $n \rightarrow \infty$. Recall that $T_n^1 = \mathbf{1}(R \geq 0)$, where $R = \sup_{|G| \geq h} R(G)$ and $R(G) = \sum_{i=1}^n Y_i \mathbf{1}(X_i \in G) - \frac{\#(X \cap G)}{2}$, for all $G \in \mathcal{S}$. Note that $R(G)$ is piecewise constant, and can only take a finite number of values. It is clear that

$$\{R(G) : G \in \mathcal{S}, |G| \geq h\} = \{R([X_k, X_l]) : 1 \leq k < l \leq n, X_l - X_k > h\}.$$

Recall that for $1 \leq k < l \leq n$, $R([X_k, X_l]) = \frac{1}{2} \sum_{i=k}^{l-1} (2Y_i - 1)$. Therefore,

$$\begin{aligned} \mathbb{P}_\emptyset[R \geq 0] &= \mathbb{P}_\emptyset \left[\max_{\substack{1 \leq k < l \leq n \\ X_l - X_k > h}} R([X_k, X_l]) > 0 \right] \\ &\leq \mathbb{P}_\emptyset \left[\bigcup_{1 \leq k < l \leq n} \{R([X_k, X_l]) > 0\} \cap \{X_l - X_k > h\} \right] \\ &\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_\emptyset [R([X_k, X_l]) > 0, X_l - X_k > h] \\ &\leq \sum_{1 \leq k < l \leq n} \mathbb{P}_\emptyset \left[\sum_{i=k}^{l-1} (2\xi_i - 1) > 0 \right] \mathbb{P}[X_l - X_k > h]. \end{aligned} \quad (13)$$

For $1 \leq k < l \leq n$,

$$\mathbb{P}_\emptyset \left[\sum_{i=k}^{l-1} (2\xi_i - 1) > 0 \right] \leq \exp \left(-\frac{(l-k)\sigma^2}{8} \right), \quad (14)$$

using Markov's inequality and (5).

If the design is (DD), then $\mathbb{P}[X_l - X_k > h]$ is 1 if and only if $l - k > nh$, 0 otherwise, so from (13) and (14) we get that

$$\begin{aligned} \mathbb{P}_\emptyset[R \geq 0] &\leq \sum_{l-k > nh} \exp\left(-\frac{(l-k)\sigma^2}{8}\right) \\ &\leq \sum_{l-k > nh} \exp\left(\frac{(-nh)\sigma^2}{8}\right) \\ &\leq \frac{n^2}{2} \exp\left(\frac{(-nh)\sigma^2}{8}\right) \rightarrow 0, \end{aligned}$$

when $n \rightarrow \infty$, which proves that the error of the first kind goes to zero.

If the design is (RD), let us use the following Lemma.

Lemma 1. *Let X_1, \dots, X_n be the (RD) design. Then, for any $1 \leq k < l \leq n$, and $h > 0$,*

$$\mathbb{P}[X_l - X_k > h] \leq n \exp(-nh(1 - e^{-u}) + u(l - k)), \forall u > 0.$$

Proof of Lemma 1 Note that the event $\{X_l - X_k > h\}$ is equivalent to $\{\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k\}$. Let us denote by X'_1, \dots, X'_n the preliminary design, from which X_1, \dots, X_n is the reordered version. The random variables X'_1, \dots, X'_n are then i.i.d., with uniform distribution on $[0, 1]$. Hence,

$$\begin{aligned} \mathbb{P}[X_l - X_k > h] &= \sum_{j=1}^n \mathbb{P}[\#(\mathcal{X} \cap (X_k, X_k + h)) < l - k, X_k = X'_j] \\ &\leq \sum_{j=1}^n \mathbb{P}[\#(\mathcal{X} \cap (X'_j, X'_j + h)) < l - k] \\ &\leq \sum_{j=1}^n \mathbb{E}[\mathbb{P}[\#(\mathcal{X} \cap (X'_j, X'_j + h)) < l - k | X'_j]] \\ &\leq \sum_{j=1}^n \mathbb{E}[\mathbb{P}[n - \#(\mathcal{X} \cap (X'_j, X'_j + h)) \geq n - l + k | X'_j]] \\ &\leq \sum_{j=1}^n \mathbb{E}[f(X'_j)], \end{aligned} \tag{15}$$

where $f(x) = \mathbb{P}[n - \#(\mathcal{X} \cap (x, x + h)) \geq n - l + k]$, for all $x \in [0, 1]$. The random variable $n - \#(\mathcal{X} \cap (x, x + h))$ is binomial with parameters n and $1 - h$, and by Markov's inequality, for all $u > 0$,

$$f(x) \leq e^{nu} (1 - h(1 - e^{-u}))^n e^{-u(n-l+k)}, \tag{16}$$

and (15) and (16) yield the lemma. \square

Therefore, by (13), (14) and (15), and Lemma 1 with $u = \sigma^2/8$,

$$\begin{aligned} \mathbb{P}_\emptyset[R \geq 0] &\leq \sum_{1 \leq k < l \leq n} \exp\left(-\frac{(l-k)\sigma^2}{8} - nh(1 - e^{-u}) + u(l-k)\right) \\ &\leq \frac{n^2}{2} \exp\left(-nh(1 - e^{-\sigma^2/8})\right) \rightarrow 0, \end{aligned}$$

when $n \rightarrow \infty$, which proves that the error of the first kind goes to zero.

Let us bound, now, the error of the second kind. Let $G \in \mathcal{S}$ satisfying $|G| \geq h$. For this G , denote by $N_G = \#(\mathcal{X} \cap G)$. Then,

$$\begin{aligned} \mathbb{P}_G[R < 0] &\leq \mathbb{P}_G[R(G) \leq N_G/2] \\ &\leq \mathbb{P}\left[\sum_{i=1}^n \xi_i \mathbf{1}(X_i \in G) \leq -N_G/2\right]. \end{aligned} \quad (17)$$

For the design (DD), N_G is the integer part of $n|G|$, so $N_G \geq nh$. Therefore, by Markov's inequality, and by (5), (17) becomes

$$\mathbb{P}_G[R < 0] \leq \exp\left(-\frac{N_G}{8\sigma^2}\right) \leq \exp\left(-\frac{nh}{8\sigma^2}\right). \quad (18)$$

For the design (RD), N_G is a random binomial variable, with parameters n and $|G|$. By conditioning to the design and using Markov's inequality, (17) becomes

$$\begin{aligned} \mathbb{P}_G[R < 0] &\leq \mathbb{P}\left[\sum_{i=1}^n -\xi_i \mathbf{1}(X_i \in G) \geq N_G/2\right] \\ &\leq \mathbb{E}\left[\exp\left(-\frac{N_G}{8\sigma^2}\right)\right] \\ &\leq \exp(-Cn|G|) \leq \exp(-Cnh), \end{aligned} \quad (19)$$

where $C = 1 - e^{-\frac{1}{8\sigma^2}}$.

In both cases (18) and (19), the right side does not depend on G , and goes to zero as $n \rightarrow \infty$. We conclude that, for both designs (DD) and (RD),

$$\sup_{|G| \geq h} \mathbb{P}_G[R < 0] \rightarrow 0,$$

which ends the proof of the upper bound.

Lower bound We more or less reproduce the proof of [Gay01], Theorem 3.1. Here, the noise is supposed to be Gaussian, with variance σ^2 . Let us assume that $\frac{nh}{\ln n} \rightarrow 0$. Let $M = 1/h$, assumed to be an integer, without loss of generality. For $q = 0, \dots, M$, let $G_q = [qh, (q+1)h]$. For $q = 1, \dots, M$, let $Z_q = \frac{d\mathbb{P}_{G_q}}{d\mathbb{P}_\emptyset}(X_1, Y_1, \dots, X_n, Y_n)$, and denote by $\bar{Z} = \frac{1}{M} \sum_{q=1}^M Z_q$. Let τ_n be any test. Then,

$$\begin{aligned}
\gamma_n(\tau_n, \mathcal{S}) &\geq \mathbb{P}_\emptyset[\tau_n = 1] + \frac{1}{M} \sum_{q=1}^M \mathbb{P}_{G_q}[\tau_n = 0] \\
&\geq \frac{1}{M} \sum_{q=1}^M (\mathbb{P}_\emptyset[\tau_n = 1] + \mathbb{P}_{G_q}[\tau_n = 0]) \\
&\geq \frac{1}{M} \sum_{q=1}^M (\mathbb{E}_\emptyset[\tau_n] + \mathbb{E}_{G_q}[1 - \tau_n]) \\
&\geq \frac{1}{M} \sum_{q=1}^M \mathbb{E}_\emptyset[\tau_n + (1 - \tau_n)Z_q] \\
&\geq \mathbb{E}_\emptyset[\tau_n + (1 - \tau_n)\bar{Z}] \\
&\geq \mathbb{E}_\emptyset[(\tau_n + (1 - \tau_n)\bar{Z}) \mathbf{1}(\bar{Z} \geq 1/2)] \\
&\geq \frac{1}{2} \mathbb{P}_\emptyset[\bar{Z} \geq 1/2]. \tag{20}
\end{aligned}$$

Let us prove that $\mathbb{E}_\emptyset[\bar{Z}] = 1$, and that $\mathbb{V}_\emptyset[\bar{Z}] \rightarrow 0$. This will imply that the right side term of (20) goes to zero, when $n \rightarrow \infty$.

For $q = 1, \dots, M$, under the null hypothesis,

$$\begin{aligned}
Z_q &= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n ((Y_i - \mathbf{1}(X_i \in G_q))^2 - Y_i^2)\right) \\
&= \exp\left(\frac{1}{2\sigma^2} \sum_{i=1}^n (2\xi_i - 1)\mathbf{1}(X_i \in G_q)\right). \tag{21}
\end{aligned}$$

By its definition, Z_q has expectation 1 under \mathbb{P}_\emptyset :

$$\mathbb{E}_\emptyset[\bar{Z}] = 1. \tag{22}$$

Since, almost surely, no design point falls in two G_q 's at the time, a simple computation shows that the random variables $Z_q, q = 1, \dots, M$, are

not correlated. Thus,

$$\mathbb{V}_\emptyset[\bar{Z}] = \frac{1}{M^2} \sum_{q=1}^M \mathbb{V}_\emptyset[Z_q].$$

Let us bound from above $\mathbb{V}_\emptyset[Z_q]$, for $q = 1, \dots, M$:

$$\begin{aligned} \mathbb{V}_\emptyset[Z_q] &\leq \mathbb{E}_\emptyset[Z_q^2] \\ &= \mathbb{E} \left[\exp \left(-\frac{\#(\mathcal{X} \cap G_q)}{\sigma^2} \right) \mathbb{E}_\emptyset \left[\exp \left(\frac{2}{\sigma^2} \sum_{i=1}^n \xi \mathbf{1}(X_i \in G_q) \right) \mid \mathcal{X} \right] \right] \\ &= \mathbb{E} \left[\exp \left(\frac{\#(\mathcal{X} \cap G_q)}{\sigma^2} \right) \right]. \end{aligned} \quad (23)$$

If the design is (DD), then we get that

$$\mathbb{V}_\emptyset[Z_q] \leq \exp \left(\frac{nh + 1}{\sigma^2} \right),$$

and the variance of \bar{Z} is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp \left(\frac{nh + 1}{\sigma^2} \right). \quad (24)$$

If the design is (RD), then $\#(\mathcal{X} \cap G_q)$ is a binomial random variable with parameters n and h , so from (23), we get that

$$\begin{aligned} \mathbb{V}_\emptyset[Z_q] &\leq \left(1 + \left(e^{1/\sigma^2} - 1 \right) h \right)^n \\ &\leq \exp(Cnh), \end{aligned}$$

where $C = e^{1/\sigma^2} - 1$, and the variance of \bar{Z} is then bounded from above:

$$\mathbb{V}_\emptyset[\bar{Z}] \leq h \exp(Cnh). \quad (25)$$

Since we assumed that $nh/\ln n \rightarrow 0$, the right side terms of (24) and (25) go to zero, and therefore, for both designs (DD) and (RD),

$$\mathbb{V}_\emptyset[\bar{Z}] \rightarrow 0. \quad (26)$$

Finally, we get from (20), (22) and (26), that

$$\liminf_{n \rightarrow \infty} \gamma_n(\tau_n, \mathcal{S}) \geq \frac{1}{2}.$$

This concludes the proof. ■

5.2 Proof of Theorem 2

The beginning of this proof holds for any design $\{X_1, \dots, X_n\}$, independent of the noise $\xi_i, i = 1, \dots, n$. Let $G \in \mathcal{S}_0$. Let $M = \max\{i = 1, \dots, n : X_i \in G\}$ - set $M = 0$ if the set is empty -. Then, $\hat{M}_n \in \underset{M'=1, \dots, n}{\text{ArgMax}} (F(M') - F(M))$, and, by (7),

$$F(M') - F(M) = -|M' - M| + \begin{cases} 2 \sum_{i=M'+1}^M \xi_i & \text{if } M > M', \\ 0 & \text{if } M' = M, \\ -2 \sum_{i=M+1}^{M'} \xi_i & \text{if } M < M'. \end{cases}$$

Let us complete the i.i.d. sequence ξ_1, \dots, ξ_n to obtain an infinite double sided i.i.d. sequence $(\xi_i)_{i \in \mathbb{Z}}$, independent of the design. Let $k \in \mathbb{N}^*$ be any positive integer. Define, for $i \in \mathbb{Z}, \tilde{\xi}_i = \xi_{i+M}$. Since M depends on the design only, it is independent of the $\xi_i, i \in \mathbb{Z}$, and therefore, the $\tilde{\xi}_i, i \in \mathbb{Z}$ are i.i.d., with same distribution as ξ_1 . Let E_k be the event $\{\hat{M}_n \geq M + k\}$. If E_k holds, then :

$$0 \leq F(\hat{M}_n) - F(M) = M - \hat{M}_n - 2 \sum_{i=M+1}^{\hat{M}_n} \xi_i,$$

and it follows that

$$\begin{aligned} 0 &\leq \max_{M+k \leq j \leq n} \left(M - j - 2 \sum_{i=M+1}^j \xi_i \right) \\ &\leq \max_{M+k \leq j \leq n} \left(M - j - 2 \sum_{i=1}^{j-M} \tilde{\xi}_i \right) \\ &\leq \max_{k \leq j \leq n-M} \left(-j - 2 \sum_{i=1}^j \tilde{\xi}_i \right). \end{aligned}$$

Hence, for all $u > 0$,

$$\begin{aligned}
\mathbb{P}_G[E_k] &\leq \mathbb{P}_G \left[\max_{k \leq j} \left(-j - 2 \sum_{i=1}^j \hat{\xi}_i \right) \geq 0 \right] \\
&\leq \mathbb{P} \left[\max_{k \leq j} \left(-j - 2 \sum_{i=1}^j \xi_i \right) \geq 0 \right] \\
&\leq \sum_{j=k}^{\infty} \mathbb{P} \left[-2 \sum_{i=1}^j \xi_i \geq j \right] \\
&\leq \sum_{j=k}^{\infty} \frac{\mathbb{E} \left[e^{-2u \sum_{i=1}^j \xi_i} \right]}{e^{uj}}, \text{ by Markov's inequality} \\
&\leq \sum_{j=k}^{\infty} e^{(-u+2\sigma^2 u^2)j}, \text{ by (5)}
\end{aligned}$$

and, by choosing $u = 1/(4\sigma^2)$,

$$\mathbb{P}_G[E_k] \leq C e^{-k/(8\sigma^2)},$$

where $C = (1 - e^{-1/(8\sigma^2)})^{-1}$ is a positive constant. By symmetry, we obtain that :

$$\mathbb{P}_G[|\hat{M}_n - M| \geq k] \leq 2C e^{-k/(8\sigma^2)}. \quad (27)$$

If the design is (DD), the conclusion is straightforward, since for all $i, j = 1, \dots, n$, $|X_i - X_j| = \frac{|i-j|}{n}$, and Theorem 2 is proved. \blacksquare

If the design is (RD), it is not clear how to go from (27) to an upper bound for the probability $\mathbb{P}_G[|\hat{\theta}_n - \theta| \geq \epsilon]$, for $\epsilon > 0$. This question should be addressed in a future work.

5.3 Proof of Theorem 3

The proof is straightforward. Let $G_1 = [0, 0]$ and $G_2 = [0, 1/(2n)]$. Then $\mathbb{P}_{G_1} = \mathbb{P}_{G_2}$, since no point of the design falls in $G_1 \triangle G_2$, and for any estima-

tor \hat{G}_n ,

$$\begin{aligned}
\sup_{G \in \mathcal{S}_0} \mathbb{E}_G [|\hat{G}_n \triangle G|] &\geq \mathbb{E}_{G_1} [|\hat{G}_n \triangle G_1|] + \mathbb{E}_{G_2} [|\hat{G}_n \triangle G_2|] \\
&\geq \mathbb{E}_{G_1} [|\hat{G}_n \triangle G_1| + |\hat{G}_n \triangle G_2|] \\
&\geq \mathbb{E}_{G_1} [|G_1 \triangle G_2|] \text{ by the triangle inequality} \\
&\geq \frac{1}{2n}.
\end{aligned}$$

■

5.4 Proof of Theorem 8

Let I_0 be the set of even positive integers less or equal to n , and I_1 the set of odd such integers. Note that $\{X_i : i \in I_0\}$ is a deterministic and regular design, with step $2/n$. Let $\hat{G}_n \in \text{ArgMax}_{G' \in \mathcal{S}} \sum_{i \in I_0} (2Y_i - 1) \mathbf{1}(X_i \in G')$ be the LSE estimator given in Theorem 5, built using only the subsample $\{X_i : i \in I_0\}$. Let $x > 0$, whose value will be specified in the course of the proof. Consider the event $E_x = \{|\hat{G}_n \triangle G| < \frac{x \ln n}{n}\}$. By Theorem 5, this event holds with probability at least $1 - C_1 e^{-(C_2 x - 4) \ln n}$. Choose x such that $\frac{x \ln n}{n} \leq \mu/2$. This choice implies that on the event E_x , $|\hat{G}_n \triangle G| < \mu \leq |G|$, so necessary, \hat{G}_n and G must intersect. Thus, still on the event E_x ,

$$|\hat{G}_n \triangle G| = |\hat{b}_n - b| + |\hat{a}_n - a|,$$

where we denoted by $G = [a, b]$ and $\hat{G}_n = [\hat{a}_n, \hat{b}_n]$. Let $m = \frac{a+b}{2}$ and $\hat{m}_n = \frac{\hat{a}_n + \hat{b}_n}{2}$ be, respectively, the middle points of G and \hat{G}_n . From now on, let us assume that E_x holds. Then,

$$\begin{aligned}
|\hat{m}_n - m| &\leq \frac{1}{2} (|\hat{b}_n - b| + |\hat{a}_n - a|) \\
&\leq \frac{1}{2} |\hat{G}_n \triangle G| \\
&\leq \frac{x \ln n}{2n} \\
&\leq \frac{\mu}{4}.
\end{aligned} \tag{28}$$

Therefore, $\hat{m}_n \in G$ and, combining (28) with the fact that $|G| \geq \mu$,

$$\min(\hat{m}_n, 1 - \hat{m}_n) \geq \frac{\mu}{4}. \tag{29}$$

Let us define

$$I_1^+ = \{i \in I_1 : X_i \geq \hat{m}_n\},$$

and

$$I_1^- = \{i \in I_1 : X_i \leq \hat{m}_n\}.$$

By (29), $\#I_1^\epsilon \geq \frac{\mu n}{8} - 1 \geq \frac{\mu n}{16}$ for n large enough, and for $\epsilon \in \{+, -\}$. Note that $\{X_i : i \in I_1^+\}$ (resp. $\{X_i : i \in I_1^-\}$) is a deterministic and regular design of the segment $[\hat{m}_n, 1]$ (resp. $[0, \hat{m}_n]$), of cardinality greater or equal to $\frac{\mu n}{16}$, as we saw just before. Then, since we have both

$$Y_i = \mathbb{1}(X_i \leq b) + \xi_i, \forall i \in I_1^+$$

and

$$Y_i = \mathbb{1}(X_i \geq a) + \xi_i, \forall i \in I_1^-,$$

the change points a and b can be estimated as in Theorem 2, using the subsamples $\{(X_i, Y_i) : i \in I_1^+\}$ and $\{(X_i, Y_i) : i \in I_1^-\}$ respectively, and we get two estimators \tilde{a}_n and \tilde{b}_n which satisfy:

$$\mathbb{P}_G \left[|\tilde{a}_n - a| \geq \frac{16y}{\mu n}, E_x \right] \leq C_0 e^{-y/(8\sigma^2)}$$

and

$$\mathbb{P}_G \left[|\tilde{b}_n - b| \geq \frac{16y}{\mu n}, E_x \right] \leq C_0 e^{-y/(8\sigma^2)},$$

for all $y > 0$. Set $\tilde{G}_n = [\tilde{a}_n, \tilde{b}_n]$, on the event E_x , and $\tilde{G}_n = \emptyset$ on its complementary \bar{E}_x . By setting $x = \frac{\mu n}{2 \ln n}$, which is the maximal value authorized in this proof,

$$\begin{aligned} \mathbb{P}_G \left[|\tilde{G}_n \triangle G| \geq \frac{y}{n} \right] &\leq \mathbb{P}_G \left[|\tilde{G}_n \triangle G| \geq \frac{y}{n}, E_x \right] + \mathbb{P}_G[\bar{E}_x] \\ &\leq 2C_0 e^{-\mu y/(256\sigma^2)} + C_1 n^4 e^{-C_2 \mu n/2}, \end{aligned}$$

which ends the proof of Theorem 8. ■

References

- [BKL⁺09] L. Boysen, A. Kempe, V. Liescher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37:157–183, 2009.
- [Bro76] E. M. Bronshtein. ϵ -entropy of convex sets and functions. *Siberian Mathematical Journal*, 17:303–398, 1976.

- [Bru13] V.-E. Brunel. Adaptive estimation of convex polytopes and convex sets from noisy data. *Electronic Journal of Statistics*, 7:1301–1327, 2013.
- [BTW07] F. Bunea, A. B. Tsybakov, and M. H. Wegkamp. Aggregation for gaussian regression. *The Annals of Statistics*, 35:1674–1697, 2007.
- [CW13] H. P. Chan and G. Walther. Detection with the scan and the average likelihood ratio. *Statistica Sinica*, 23:409–428, 2013.
- [FMSnt] S. Frick, A. Munk, and H. Sieling. Multiscale change-point inference. *arXiv:1301.7212*, preprint.
- [Gay01] G. Gayraud. Minimax hypothesis testing about the density support. *Bernouilli*, 7:507–525, 2001.
- [IK84] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1984.
- [Kor06] A. P. Korostelev. On minimax estimation of a discontinuous signal (translation of a paper in russian, published in 1986). *Theory Probab. Appl.*, 32:727–730, 2006.
- [KT92] A. P. Korostelev and A. B. Tsybakov. Asymptotically minimax image reconstruction problems. in topics in nonparametric estimation. (*R. Z. Khasminskii, ed.*) *Amer. Math. Soc., Providence, RI*, pages 45–86, 1992.
- [KT93] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*. Springer, NY, 1993.
- [Leb03] E. Lebarbier. Detecting multiple change-points in the mean of gaussian process by model selection. *Rapport de recherche de l’INRIA, n. 4740*, pages 1–24, 2003.
- [LT00] O. V. Lepski and A. B. Tsybakov. Asymptotically exact non-parametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117:17–48, 2000.
- [MvdG97] E. Mammen and S. van de Geer. Locally adaptive regression splines. *The Annals of Statistics*, 25:387–413, 1997.

- [SZ10] X. Shao and X. Zhang. Testing for change points in time series. *Journal of the American Statistical Association*, 105:1228–1240, 2010.
- [Tib96] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statistic. Soc. Ser. B*, 58:267–288, 1996.
- [Tsy09] A.B. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009.