OFF-THE-GRID PREDICTION AND TESTING FOR MIXTURES OF TRANSLATED FEATURES

CRISTINA BUTUCEA, JEAN-FRANÇOIS DELMAS, ANNE DUTFOY, AND CLÉMENT HARDY

Abstract. We consider a model where a signal (discrete or continuous) is observed with an additive Gaussian noise process. The signal is issued from a linear combination of a finite but increasing number of translated features. The features are continuously parameterized by their location and depend on some scale parameter. First, we extend previous prediction results for off-the-grid estimators by taking into account here that the scale parameter may vary. The prediction bounds are analogous, but we improve the minimal distance between two consecutive features locations in order to achieve these bounds.

Next, we propose a goodness-of-fit test for the model and give non-asymptotic upper bounds of the testing risk and of the minimax separation rate between two distinguishable signals. In particular, our test encompasses the signal detection framework. We deduce upper bounds on the minimal energy, expressed as the ℓ_2-norm of the linear coefficients, to successfully detect a signal in presence of noise. The general model considered in this paper is a non-linear extension of the classical high-dimensional regression model. It turns out that, in this framework, our upper bound on the minimax separation rate matches (up to a logarithmic factor) the lower bound on the minimax separation rate for signal detection in the high dimensional linear model associated to a fixed dictionary of features. We also propose a procedure to test whether the features of the observed signal belong to a given finite collection under the assumption that the linear coefficients may vary, but do not change to opposite signs under the null hypothesis. A non-asymptotic upper bound on the testing risk is given.

We illustrate our results on the spikes deconvolution model with Gaussian features on the real line and with the Dirichlet kernel, frequently used in the compressed sensing literature, on the torus.

1. Introduction

1.1. **Model.** This paper is motivated by the study of the spikes deconvolution model [10] with applications in spectroscopy ([6]). In this model, a linear combination (or mixture) of spikes continuously parameterized is observed with an additive Gaussian noise process. We assume that the spikes are parameterized by a location parameter, that the noise and the observation space can vary with some parameter T increasing with the quality of the observations. More general non-linear models for the spikes have been discussed in [5], and the particular case of location families has been discussed in Section 8 of that paper. However, we allow here the scale of the spikes to vary with T, which makes the approach very different from the previous one.

We are also interested in goodness-of-fit testing, that is we want to test whether the observations are issued from a given linear combination of spikes. We remark that it includes the case of signal detection. This test problem finds an application in spectroscopy to detect the presence of a chemical compound in a material. In addition, we are interested in testing whether the observed signal is a linear combination of spikes located at a prescribed list of locations with linear coefficients having prescribed signs under the null hypothesis.

Let $T \in \mathbb{N}$. We observe a random element y in the Hilbert space $L^2(\lambda_T)$ of square integrable functions with respect to the measure λ_T on the Borel σ-field of some metric space Θ. The observation is the sum of a deterministic signal and a Gaussian random process w_T in $L^2(\lambda_T)$. We shall assume that the signal is an

Date: December 5, 2022.

unknown finite mixture of s features belonging to a continuously parametrized subfamily $(\varphi_T(\theta),\, \theta \in \Theta)$ of $L^2(\lambda_T)$. We call this family a continuous dictionary, the weights of the mixture - the linear coefficients, and the parameters of the features - the non-linear parameters.

The quality of the information provided by the observations depends on the support of the measure λ_T and on the noise w_T. It increases with the parameter T. For example, we will consider the case where the sequence of measures $(\lambda_T, T \in \mathbb{N})$ converges towards the Lebesgue measure, noted Leb, on Θ, so that in the limit model the observation corresponds to a square integrable random process indexed on Θ. We consider the cases where $\Theta = \mathbb{R}$ and the limit measure is the Lebesgue measure on $\mathbb{R}$ as well as the case where Θ is the torus $\mathbb{R}/\mathbb{Z}$ and the limit measure is the Lebesgue measure on this manifold.

We consider in this paper a dictionary given by the location model:

$$\Big(\varphi_T(\theta) = h(\theta - \cdot, \sigma_T),\, \theta \in \Theta\Big) \tag{1}$$

where h is a real-valued function defined on $\Theta \times \mathfrak{S}$, smooth with respect to its first variable and normalized so that $\|h(\cdot, \sigma_T)\|_{L^2(\mathrm{Leb})} = 1$, and where σ_T is an element of the set $\mathfrak{S}$ of admissible positive scale parameters. See Section 2.2 for examples of functions h including the Gaussian spike and the low-pass filter. Even though the location model considered here is a restriction when compared to general non-linear dictionnaries of features considered by e.g. [5], the scaling σ_T introduced here makes this dictionnary different. Indeed, this scaling is allowed to depend on T and may improve previous results in the sense that the sufficient conditions on the non-linear parameters in the mixture in order to obtain the prediction and estimation bounds are milder. The least separation distance between the location parameters in this model is allowed to be smaller when compared to unscaled dictionnaries, see Remark 2.2.

The Hilbert space $L^2(\lambda_T)$ is endowed with the natural scalar product noted $\langle \cdot, \cdot \rangle_{L^2(\lambda_T)}$ and norm $\|\cdot\|_{L^2(\lambda_T)}$. Let us define the normalized function ϕ_T defined on Θ by:

$$\phi_T(\theta) = \varphi_T(\theta)/\|\varphi_T(\theta)\|_{L^2(\lambda_T)}, \tag{2}$$

and the multivariate function Φ_T on Θ^s by:

$$\Phi_T(\vartheta) = (\phi_T(\theta_1), \dots, \phi_T(\theta_s))^\top \quad \text{for} \quad \vartheta = (\theta_1, \cdots, \theta_s) \in \Theta^s.$$

We assume that the signal contains an unknown number $s \in \mathbb{N}$ of active features. We consider the model with unknown non-zero linear coefficients $\beta^\star$ in $(\mathbb{R}^*)^s$ and unknown distinct parameters $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star) \in \Theta^s$:

$$y = \beta^\star \Phi_T(\vartheta^\star) + w_T \quad \text{in } L^2(\lambda_T), \tag{3}$$

where when $s = 0$, we set by convention that $\beta^\star \Phi_T(\vartheta^\star) = 0$ as well as $A^s = \{0\}$ for any set A. We denote by $\mathcal{Q}^\star = \{\theta_\ell^\star,\, 1 \leq \ell \leq s\}$ the set of the non-linear parameters associated to an active feature. The process y is observed over the support of the measure λ_T. Therefore it is legimate to consider models whose location parameters belong to the smallest interval covering the support of the measure λ_T. Hence, we introduce the set Θ_T, a compact interval of Θ, and we shall assume that $\mathcal{Q}^\star$ is a subset of Θ_T. We denote by $|\Theta_T|$ the Euclidean diameter of the set Θ_T.

We consider a large variety of Gaussian noise processes. Indeed, we only assume the following mild assumption on w_T, where the decay rate $\Delta_T > 0$ controls the noise variance decay as the parameter T grows and $\overline{\sigma} > 0$ is the intrinsic noise level. A wide range of noise processes satisfy our assumptions, see [5]; they can be discrete or continuous, white or coloured under these constraints.

Assumption 1.1 (Admissible noise). *Let $T \in \mathbb{N}$. The Gaussian noise process w_T satisfies $\mathbb{E}\left[\|w_T\|^4_{L^2(\lambda_T)}\right] < +\infty$, and there exist a noise level $\overline{\sigma} > 0$ and a decay rate $\Delta_T > 0$ such that for all $f \in L^2(\lambda_T)$, the random variable $\langle f, w_T \rangle_{L^2(\lambda_T)}$ is a centered Gaussian random variable satisfying:*

$$\mathrm{Var}\left(\langle f, w_T \rangle_{L^2(\lambda_T)}\right) \leq \overline{\sigma}^2\, \Delta_T\, \|f\|^2_{L^2(\lambda_T)}.$$

We assume that the quantity $\mathbb{E}\left[\|w_T\|^2_{L^2(\lambda_T)}\right]$ is known for the considered models. We consider the variance of the squared norm of the noise:

$$\Xi_T = \mathrm{Var}\left(\|w_T\|^2_{L^2(\lambda_T)}\right). \tag{4}$$

1.2. **Examples of Gaussian noise processes.** We consider a large variety of models: discrete models where the process y is observed on a grid or continuous models where the process is observed on an interval.

1.2.1. *Discrete-time process observed on a regular grid.* Consider a real-valued process y observed over a regular grid $t_1 < \cdots < t_T$ of a symmetric interval $[a_T, b_T] \subset \mathbb{R}$, with $T \geq 1$, $a_T = -b_T < 0$, $t_j = a_T + j\Delta_T$ for $j = 1, \ldots, T$ and grid step: $\Delta_T = (b_T - a_T)/T$. We set $\lambda_T = \Delta_T \sum_{j=1}^T \delta_{t_j}$ and see y as an element of $L^2(\lambda_T)$. We assume that $(b_T, T \geq 2)$ is a sequence of positive numbers, such that: $\lim_{T\to\infty} b_T = +\infty$ and $\lim_{T\to\infty} \Delta_T = 0$ so that the sequence of measures $(\lambda_T, T \geq 1)$ converges with respect to the vague topology towards the Lebesgue measure. In this formalism, the noise $w_T \in L^2(\lambda_T)$ is given by:

$$w_T(t) = \sum_{j=1}^T G_j \mathbf{1}_{\{t_j\}}(t), \tag{5}$$

where $\mathbf{1}_A$ denotes the indicator function of an arbitrary set A and $(G_1, \cdots, G_T)$ is a centered Gaussian random vector with independent entries of variance $\overline{\sigma}^2$. In this case, $\mathbb{E}[\|w_T\|^4_{L^2(\lambda_T)}] = \overline{\sigma}^4 \Delta_T^2 T(T+2)$ and Assumption 1.1 holds with an equality. We readily obtain that $\Xi_T = 2\overline{\sigma}^4 \Delta_T^2 T$.

Notice that in this particular example we have for any function $f \in L^2(\lambda_T)$ that $\|f\|_{L^2(\lambda_T)} = \sqrt{\Delta_T}\|f\|_{\ell_2}$, where the right-hand side is understood as the ℓ_2-norm (Euclidean norm) of the vector $(f(t_1), \ldots, f(t_T))$.

1.2.2. *Continuous-time processes.* Assume we observe a process y on an interval. We note λ_T for a σ-finite measure on $\mathbb{R}$ or on $\mathbb{R}/\mathbb{Z}$. In this framework, y is an element of $L^2(\lambda_T)$. Let us assume that the noise is $w_T = \sum_{k\in\mathbb{N}} \sqrt{\xi_k}\, G_k\, \psi_k$, where $(G_k, k \in \mathbb{N})$ are independent centered Gaussian random variables with variance $\overline{\sigma}^2$, $(\psi_k, k \in \mathbb{N})$ is an o.n.b. of $L^2(\lambda_T)$ on $\mathbb{R}$ or on $\mathbb{R}/\mathbb{Z}$, and that $\xi = (\xi_k, k \in \mathbb{N})$ is a square summable sequence of non-negative real numbers. We remark that Assumption 1.1 holds as $\mathbb{E}[\|w_T\|^4_T] = 3\overline{\sigma}^4 \sum_{k\in\mathbb{N}} \xi_k^2 + \overline{\sigma}^4 \sum_{k,\ell\in\mathbb{N}, k\neq\ell} \xi_k \xi_\ell$ is finite and $\mathrm{Var}(\|w_T\|^2_{L^2(\lambda_T)}) = 2\overline{\sigma}^4 \sum_{k\in\mathbb{N}} \xi_k^2$. Moreover, we have:

$$\mathrm{Var}(\langle f, w_T\rangle_{L^2(\lambda_T)}) = \overline{\sigma}^2 \sum_{k\in\mathbb{N}} \xi_k \,\langle f, \psi_k\rangle^2_{L^2(\lambda_T)} \leq \overline{\sigma}^2\, \Delta_T\, \|f\|^2_{L^2(\lambda_T)} \quad \text{with} \quad \Delta_T = \sup_{k\in\mathbb{N}} \xi_k.$$

In this example the noise w_T depends on the parameter T only if ξ, and thus Δ_T, depend on T. We may consider different choices for ξ that lead to different values for Ξ_T, the variance of the squared norm of the noise. For instance, our framework encompasses the truncated white noise by taking for all $k \in \mathbb{N}$, $\xi_k = T^{-1}\mathbf{1}_{\{1\leq k\leq T\}}$. In this case, elementary calculations give $\Delta_T = 1/T$ and $\Xi_T = 2\overline{\sigma}^4/T$.

1.3. **Description of the results.** The aim of this paper is twofold. First, we improve on [5] in the case of linear combination of translated spikes by giving bounds on the prediction error under milder separation constraints between the non-linear parameters in $\mathcal{Q}^\star$. This is achieved by taking the scale parameter of the features σ_T into account. In particular, in the case of Gaussian spikes deconvolution, the separation is of order σ_T.

Then, test problems are studied. We give procedures for the goodness-of-fit of the mixture model in order to determine whether the unknown signal $\beta^\star\Phi_T(\vartheta^\star)$ is equal to a reference signal $\beta^0\Phi_T(\vartheta^0)$ for some known vectors $\beta^0 \in (\mathbb{R}^*)^{s^0}$ and $\vartheta^0 \in \Theta_T^{s^0}$. This setup includes the case of signal detection where the null hypothesis is $\beta^\star \equiv 0$, that is $s = 0$. We propose a combined procedure based on differences between the reference signal $\beta^0\Phi_T(\vartheta^0)$ and either the observation y or a reconstructed signal obtained from estimators of the model parameters. In order to successfully perform the test, we remove from the alternative hypothesis the signals whose proximity with the reference signal $\beta^0\Phi_T(\vartheta^0)$ with respect to the norm $\|\cdot\|_{L^2(\lambda_T)}$ is below some separation parameter. We give a non-asymptotic upper bound of the testing risk and deduce an upper bound on the minimal separation needed to distinguish two different signals. This upper bound yields two regimes depending on whether the observed signal and the reference signal are sparse or not. In the case of signal detection, the separation can be expressed as the ℓ_2-norm of the linear coefficients of the observed mixture. In particular, when the observation y is issued from a non-linear extension of the classical high-dimensional regression model, our upper bound matches (up to logarithmic factors) the asymptotic lower bound of the minimal separation needed to distinguish two signals that are mixture of features from a finite high-dimensional dictionary.

We also test the presence of at most s_0 prescribed features in the mixture with arbitrary linear coefficients of given sign. That is, we test whether for each $\epsilon = \pm 1$ the unknown set $\mathcal{Q}^{\star,\epsilon} = \{\theta_k^\star \in \mathcal{Q}^\star : \epsilon\beta_k^\star > 0\}$ is a subset of $\mathcal{Q}^{0,\epsilon}$, with $\mathcal{Q}^{0,+}$ and $\mathcal{Q}^{0,-}$ being disjoint finite subsets of Θ_T. This setup is issued from an application to spectroscopy (see [6]), where the presence of other chemical components than the prescribed ones are indicating aging or substantial modifications of the analyzed material. To separate the null hypothesis from the alternative hypothesis, we introduce a discrepancy that is 0 under all parameters $(\beta^\star, \vartheta^\star)$ belonging to the null hypothesis. We give an upper bound on the minimal separation to successfully perform our test. The test statistic introduced and studied in this context makes explicit use of the certificates used in [5] for establishing the prediction rates of the estimators of $(\beta^\star, \vartheta^\star)$. We stress the fact that the test statistic is not an estimator of the discrepancy measure separating the null and the alternative hypotheses, as is usually the case in non-parametric tests.

1.4. **Previous work.** Estimating the linear coefficients and the parameters of model (3) from an observation y has attracted a lot of attention over the past decade. A major contribution in this field comes from the formulation of the BLasso problem in [9]. This optimization problem on a space of measures allows to estimate both linear coefficients and non-linear parameters without using a grid on the parameter space. This off-the-grid method has succesfully been used in [8] and [7] in the context of super-resolution as well as in [10] for spikes deconvolution. High probability bounds for the prediction error have been given in [18] and [4] for the specific dictionary of complex exponentials continuously paremetrized by their frequencies and more recently in [5] for a wide range of dictionnaries parametrized over a one-dimensional space. These results are based on certificate functions whose existence have been proven in a very general framework in [17] provided that the non-linear parameters of the mixture are well-separated with respect to a Riemannian metric.

Goodness-of-fit tests are used to check whether observations are indeed derived from a given statistical model. We refer to the monograph [13] for a comprehensive presentation of goodness-of-fit testing. When we consider a finite dictionary of features $(\varphi_T(\theta), \theta \in \mathcal{Q})$ with $\mathcal{Q}$ a known finite subset of Θ, the model (3) can be rewritten as a linear regression model, possibly of high dimension depending on the size of the finite dictionary $p := \mathrm{Card}(\mathcal{Q})$. In this case, testing the goodness-of-fit of the model amounts to testing whether the linear coefficients in the mixture are equal to some given linear coefficients. When the dictionary is known, the testing problem is homogeneous in the linear coefficients β and is therefore equivalent to testing $\beta \equiv 0$, which is a signal detection problem.

Signal detection has raised a lot of interest over the past decades. It is well known that the alternative hypothesis H_1 (presence of signal) must be well separated from the null hypothsesis H_0 (only noise) in order to have tests with small risks. The separation can be seen as a minimal signal intensity allowing the detection. Then, it is a matter of interest to evaluate the minimax separation rate, i.e., the smallest separation that allows to distinguish the tested hypotheses. In [11], asymptotic rates for the minimax separation in the framework of signal detection are derived for the non-parametric Gaussian white noise model. Non-asymptotic rates were then derived in [3] and later in [15] to tackle the case of heterogeneous variances. We refer to the monograph [12] for an overview of non-parametric hypotheses testing. Regarding the high dimensional regression model where the observation is of dimension T and the dictionary is fixed, known and of size p, the work of [14] established the following asymptotic minimax separation rates under coherence assumptions on the dictionary:

$$\frac{1}{T^{\frac{1}{4}}} \wedge \sqrt{\frac{s}{T}\log(p)} \wedge \frac{p^{\frac{1}{4}}}{\sqrt{T}}.$$

The signal intensity is expressed by the ℓ_2-norm of the linear coefficients. Their lower bounds on the asymptotic minimax separation stand for both fixed and random designs whereas their upper bounds stand for random designs. The work of [2] does not tackle the high dimension but provides tests achieving the minimax separation for fixed designs under coherence assumptions on the dictionary.

In this paper we shall consider that our features come from a continuous dictionary and have unknown location parameters. Hence, the existing results do not apply. Furthermore, for the considered non-linear extension of linear regression models, goodness-of-fit testing does not reduce to signal detection. Therefore, we introduce new testing procedures. We stress that one of the test statistics is not derived from estimators of the linear coefficients. In fact, depending on the sparsity of the signal, the dimension of the observation and the size of the dictionary, plug-in methods using sparse estimators might not be the best way to proceed. They do not always lead to the minimal separation. In this sense, testing is a very different statistical problem from estimation.

1.5. **Roadmap of the paper.** In Section 2, we start by presenting the assumptions needed to perform a successful estimation of the linear coefficients and location parameters of our model. After giving a prediction bound in Theorem 2.3, we show in Lemma 2.4 that the required assumptions are sufficient conditions for the identifiability of the model. In Section 3, we test whether the observation derives from a given mixture or from some other mixture sufficiently separated from the latter. We give in Theorems 3.1 and 3.3 bounds of the testing risks associated to two different test procedures. We show in Corollaries 3.2 and 3.5 that these two tests give two regimes for our upper bound on the minimal separation to distinguish two different signals from an observation contaminated by noise. We also provide a discussion on the comparison of our upper bounds with some existing lower bounds. In Section 4, we propose a procedure to test whether the active features in the observed signal belong to a given finite collection with linear coefficients of prescribed signs. Both hypotheses of this test problem are composite and a new measure of the separation between these hypotheses has been introduced. The proposed test makes use of the certificates used in the proof of the prediction bounds in an original way. A bound of the testing risk is given in Theorem 4.3 and in Corollary 4.4, we give an upper bound on the minimax separation rate. The examples of Gaussian scaled spikes deconvolution on $\mathbb{R}$ and low-pass filter on $\mathbb{R}/\mathbb{Z}$ are adressed in Sections 5 and 6.

2. Assumptions and prediction bounds

We recall in this section assumptions and definitions from Sections 3-5 of [5] in a simpler way adapted to our framework. In [5], the authors established high probability bounds for prediction and estimation errors associated to some estimators of $\beta^\star$ and $\vartheta^\star$ tackling a wider range of dictionaries.

2.1. **Regularity of the features.** We gather in this section the hypotheses that will be required on the features defined by (1).

Recall that the parameter space Θ is either $\mathbb{R}$ or the torus $\mathbb{R}/\mathbb{Z}$ endowed with the Lebesgue measure Leb. For convenience, we write $|x-y|$ for the Euclidean distance between x and y either on $\mathbb{R}$ or on the torus. Recall also that $L^2(\lambda_T)$ and $L^2(\text{Leb})$ are the sets of square integrable functions on Θ with respect to the measures λ_T and Leb respectively. We denote $\mathfrak{S}$ the set of scale parameters.

Assumption 2.1 (Smoothness of the features)**.** *Let h be a function defined on $\Theta \times \mathfrak{S}$. Let $T \in \mathbb{N}$ and $\sigma_T \in \mathfrak{S}$. We assume that the function $\theta \mapsto h(\theta, \sigma_T)$ is of class $\mathcal{C}^3$ on Θ. We assume furthermore that $\|h(\cdot, \sigma_T)\|_{L^2(\mathrm{Leb})} = 1$, and that for all $\theta \in \Theta$ $\|h(\theta - \cdot, \sigma_T)\|_{L^2(\lambda_T)} > 0$ and all $i \in \{0, \cdots, 3\}$ $\left\|\partial_\theta^i h(\cdot, \sigma_T)\right\|_{L^2(\mathrm{Leb})} < +\infty$ and $\left\|\partial_\theta^i h(\theta - \cdot, \sigma_T)\right\|_{L^2(\lambda_T)} < +\infty$.*

Recall the function φ_T defined by (1) and notice that Assumption 2.1 implies $\|\varphi_T(\theta)\|_{L^2(\lambda_T)} > 0$ on Θ. We define the function:

$$g_T(\theta) = \|\partial_\theta \phi_T(\theta)\|^2_{L^2(\lambda_T)}, \quad \text{where } \phi_T(\theta) = \varphi_T(\theta)/\|\varphi_T(\theta)\|_{L^2(\lambda_T)}. \tag{6}$$

Assumption 2.2 (Positivity of g_T)**.** *Assumption 2.1 holds and we have $g_T > 0$ on Θ.*

Let us mention that if for all $\theta \in \Theta$, $\varphi_T(\theta)$ and $\partial_\theta \varphi_T(\theta)$ are linearly independent functions of $L^2(\lambda_T)$ and $\|\partial_\theta \varphi_T(\theta)\|_{L^2(\lambda_T)} > 0$, then $g_T(\theta) > 0$ for all $\theta \in \Theta$ (see [5, Lemma 3.1]).

2.2. **Examples of feature functions.** We provide some examples from the literature.

(i) *Spike deconvolution.* The noisy mixture of translated and scaled Gaussian features corresponds to:

$$h(t, \sigma) \mapsto \frac{\exp(-t^2/2\sigma^2)}{\pi^{1/4}\sigma^{1/2}} \quad \text{on } \Theta \times \mathfrak{S} = \mathbb{R} \times \mathbb{R}_+^*. \tag{7}$$

The example of Gaussian spikes deconvolution is analyzed in full details in [5, Section 8] when σ_T does not depend on T. We shall consider here that the scale parameter σ_T may vary with T.

(ii) *Multi-resolution approximation.* We consider the normalized Shannon scaling function:

$$h(t, \sigma) \mapsto \sqrt{\sigma}\, \frac{\sin(\pi t/\sigma)}{\pi t} \quad \text{on } \Theta \times \mathfrak{S} = \mathbb{R} \times \mathbb{R}_+^*.$$

The associated dictionary allows to recover functions whose Fourier transform have their support in $[-\pi/\sigma, \pi/\sigma]$ (see [16, Theorem 3.5]).

(iii) *Low-pass filter.* We consider the normalized Dirichlet kernel on the torus for some cut-off frequency $f_c \in \mathbb{N}^*$ and $T = 2f_c + 1$:

$$h(t, \sigma) = \frac{1}{\sqrt{T}} \sum_{k=-f_c}^{f_c} \mathrm{e}^{2i\pi k t} = \frac{\sin(T\pi t)}{\sqrt{T}\, \sin(\pi t)}, \quad \text{with } \sigma = \frac{1}{T},\ T \in 2\mathbb{N}^* + 1 \text{ and } t \in \Theta = \mathbb{R}/\mathbb{Z}. \tag{8}$$

The example of the low-pass filter is adressed in [10], where exact support recovery results are obtained for the BLasso estimators. This dictionary is also used in [7] in the context of super-resolution. Bounds on some prediction risks (different from those considered in this paper) are established therein for estimators obtained by solving the constrained formulation of the BLasso.

2.3. **Definition of the kernel and its approximation.**

2.3.1. *Measuring the colinearity of the features.* We define the symmetric kernel $\mathcal{K}_T$ on Θ^2 by:

$$\mathcal{K}_T(\theta, \theta') = \langle \phi_T(\theta), \phi_T(\theta') \rangle_{L^2(\lambda_T)}. \tag{9}$$

The kernel $\mathcal{K}_T$ measures the colinearity of two features belonging to the continuous dictionary. It does not *a priori* have a simple form. In the following, we approximate this kernel by another kernel easier to handle.

As mentionned in the introduction, we consider in this paper a setting where the sequence of measures $(\lambda_T, T \geq 1)$ converges towards the Lebesgue measure Leb on Θ. However, since σ_T may drop towards zero, it is often pointless to follow [5] by taking the pointwise limit kernel of the sequence of kernels $(\mathcal{K}_T, T \geq 1)$ as an approximation of the kernel $\mathcal{K}_T$. Indeed, in the next example this pointwise limit kernel is degenerate.

Example 2.1 (Degenerate limit kernel). Consider the discrete-time process presented in Section 1.2.1 and the Gaussian features (7) from Section 2.2 scaled by the sequence $(\sigma_T, T \geq 1)$ that tends towards zero when T grows to infinity so that $\lim_{T\to+\infty} \Delta_T/\sigma_T = 0$. In this case, the sequence of measures $(\lambda_T, T \geq 1)$ converges with respect to the vague topology towards the Lebesgue measure and it is easy to check that the pointwise limit of the kernel $\mathcal{K}_T$ is equal to zero almost everywhere.

In what follows, we shall approximate the kernel $\mathcal{K}_T$ by a kernel $\mathcal{K}_T^{\text{prox}}$ of the form:

$$\mathcal{K}_T^{\text{prox}} : (\theta, \theta') \mapsto F(|\theta - \theta'|/\sigma_T), \tag{10}$$

where F is a real-valued even function defined on $\mathbb{R}$ with $F(0) = 1$. Since F is even, notice that if it is of class $\mathcal{C}^{2\ell}$ then $\mathcal{K}_T^{\text{prox}}$ is of class $\mathcal{C}^{\ell,\ell}$. The choice of the function F follows from the model given by h, so that $\mathcal{K}_T$ and $\mathcal{K}_T^{\text{prox}}$ are close (see *(iii)* of Assumption 2.4). We refer to Sections 5 and 6 for examples with h given by (7) and (8). The introduction of the kernel $\mathcal{K}_T^{\text{prox}}$ is significantly different from the approximation developed in [5].

2.3.2. *Covariant derivatives of the kernel.* Let $\mathcal{K}$ be a symmetric kernel of class $\mathcal{C}^2$ such that the function $g_{\mathcal{K}}$ defined on Θ by:

$$g_{\mathcal{K}}(\theta) = \partial^2_{x,y}\mathcal{K}(\theta, \theta), \tag{11}$$

is positive and locally bounded, where ∂_x (respectively ∂_y) denotes the usual derivative with respect to the first (respectively second) variable. Under Assumptions 2.1 and 2.2, the definitions (6) and (11) coincide so that $g_T = g_{\mathcal{K}_T}$ on Θ.

Similarly to [17], we introduce the covariant derivatives which reduce to elementary expressions since the location parameters are one-dimensional. More precisely following [5, Section 4], we set for a smooth function f defined on Θ, $\tilde{D}_{0;\mathcal{K}}[f] = f$, $\tilde{D}_{1;\mathcal{K}}[f] = g_{\mathcal{K}}^{-1/2} f'$ and for $i \geq 2$:

$$\tilde{D}_{i;\mathcal{K}}[f] = \tilde{D}_{1;\mathcal{K}}[\tilde{D}_{i-1;\mathcal{K}}[f]]. \tag{12}$$

Let us assume that the kernel $\mathcal{K}$ has the form $\mathcal{K}(\theta, \theta') = \langle f(\theta), f(\theta')\rangle_{L^2(\lambda)}$ for some function f of class $\mathcal{C}^3$ and some measure λ on Θ. We then define the covariant derivatives of $\mathcal{K}$ for $i, j \in \{0, \ldots, 3\}$ and $\theta, \theta' \in \Theta$ by:

$$\mathcal{K}^{[i,j]}(\theta, \theta') = \langle \tilde{D}_{i;\mathcal{K}}[f](\theta), \tilde{D}_{j;\mathcal{K}}[f](\theta')\rangle_{L^2(\lambda)}. \tag{13}$$

We also define the function $h_{\mathcal{K}}$ on Θ by:

$$h_{\mathcal{K}}(\theta) = \mathcal{K}^{[3,3]}(\theta, \theta). \tag{14}$$

Before stating technical assumptions on the function F, we set:

$$g_\infty = -F''(0). \tag{15}$$

For a real valued function f defined on a set A, we write $\|f\|_\infty = \sup_{x\in A} |f(x)|$.

Assumption 2.3 (Properties of the function F)**.** *We assume that the function F is of class $\mathcal{C}^6$ and that we have:*

$$g_\infty > 0, \quad L_6 := g_\infty^{-3}|F^{(6)}(0)| < +\infty, \quad \textit{and} \quad L_i := g_\infty^{-i/2} \left\| F^{(i)} \right\|_\infty < +\infty \quad \textit{for all } i \in \{0, \cdots, 4\}. \tag{16}$$

We give the covariant derivatives of the kernel $\mathcal{K}_T^{\text{prox}}$ according to the definition given in [5, (27)]. This definition coincides with (13) when $\mathcal{K}_T^{\text{prox}}(\theta, \theta') = \langle f(\theta), f(\theta')\rangle_{L^2(\lambda)}$ on Θ^2 for some smooth function f and some measure λ on Θ, see [5, Lemma 4.3]. We get for any $\theta, \theta' \in \Theta$ and $i, j \in \{0, \cdots, 3\}$:

$$\mathcal{K}_T^{\text{prox}[i,j]}(\theta, \theta') = \frac{(-1)^j}{g_\infty^{(i+j)/2}} F^{(i+j)}\left(|\theta - \theta'|/\sigma_T\right). \tag{17}$$

We notice that we have for any $\theta \in \Theta$:

$$g_{\mathcal{K}_T^{\mathrm{prox}}}(\theta) = g_\infty/\sigma_T^2. \tag{18}$$

2.3.3. *Measuring the quality of the approximation.* In this section, we quantify the proximity of the kernel $\mathcal{K}_T$ and $\mathcal{K}_T^{\mathrm{prox}}$.

Following [17], we define the one-dimensional Riemannian metric $\mathfrak{d}_T(\theta, \theta')$ between $\theta, \theta' \in \Theta$ by:

$$\mathfrak{d}_T(\theta, \theta') = |G_T(\theta) - G_T(\theta')|, \tag{19}$$

where G_T is a primitive of the function $\sqrt{g_T}$ assumed positive on Θ thanks to Assumption 2.2.

Recall that Θ_T, introduced below the model (3), is a compact sub-interval of Θ. Since Θ_T is compact, under Assumptions 2.2 and 2.3, we deduce that the constant C_T below is positive and finite, where:

$$C_T = \max\left(\sup_{\Theta_T} \sqrt{\frac{g_{\mathcal{K}_T^{\mathrm{prox}}}}{g_T}}, \sup_{\Theta_T} \sqrt{\frac{g_T}{g_{\mathcal{K}_T^{\mathrm{prox}}}}}\right). \tag{20}$$

Elementary calculations show that the metric $\mathfrak{d}_T$ defined in (19) is equivalent, up to a factor σ_T, to the Euclidean metric on Θ_T^2 as for any $\theta, \theta' \in \Theta_T$:

$$\frac{1}{C_T}\, \sqrt{g_\infty}\, \sigma_T^{-1}\, |\theta - \theta'| \leq \mathfrak{d}_T(\theta, \theta') \leq C_T\, \sqrt{g_\infty}\, \sigma_T^{-1}\, |\theta - \theta'|. \tag{21}$$

In order to quantify the approximation of $\mathcal{K}_T$ by $\mathcal{K}_T^{\mathrm{prox}}$, we set:

$$\mathcal{V}_T = \max(\mathcal{V}_T^{(1)}, \mathcal{V}_T^{(2)}) \quad \text{with} \quad \mathcal{V}_T^{(1)} = \max_{i,j \in \{0,1,2\}} \sup_{\Theta_T^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}_T^{\mathrm{prox}[i,j]}| \quad \text{and} \quad \mathcal{V}_T^{(2)} = \sup_{\Theta_T} |h_{\mathcal{K}_T} - h_{\mathcal{K}_T^{\mathrm{prox}}}|. \tag{22}$$

2.4. **Boundedness and local concavity on the diagonal of the approximating kernel.** Recall the definition of the kernel $\mathcal{K}_T^{\mathrm{prox}}$ given by (10) using the even function F. We quantify the boundedness and local concavity on the diagonal of the kernel $\mathcal{K}_T^{\mathrm{prox}}$ using for $r > 0$:

$$\varepsilon(r) = 1 - \sup\{|F(r')|; \quad r' \geq r\}, \tag{23}$$

$$\nu(r) = -\sup\{F''(r')/g_\infty; \quad r' \in [0, r]\}. \tag{24}$$

We also quantify the colinearity between $s \in \mathbb{N}$ features belonging to the continuous dictionary, by setting for $u > 0$:

$$\delta(u, s) = \inf\Big\{\delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k \neq \ell}^{s} g_\infty^{-\frac{i}{2}} |F^{(i)}(x_\ell - x_k)| \leq u, \quad \text{for all } i \in \{0,1,2,3\} \text{ and } (x_1, \cdots, x_s) \in \mathbb{R}^s(\delta)\Big\}, \tag{25}$$

where for any subset A of $\mathbb{R}$ or $\mathbb{R}/\mathbb{Z}$ and for any $\delta \geq 0$,

$$A^s(\delta) = \Big\{(\theta_1, \cdots, \theta_s) \in A^s : |\theta_\ell - \theta_k| > \delta \text{ for all distinct } k, \ell \in \{1, \ldots, s\}\Big\}. \tag{26}$$

with the conventions $\inf \emptyset = +\infty$, and for $s = 0, 1$: $A^0(\delta) = \{0\}$ and $A^1(\delta) = A$.

Following [5], we define quantities which depend only on the function F and on a real parameter $r > 0$:

$$H_\infty^{(1)}(r) = \frac{1}{2} \wedge L_2 \wedge L_3 \wedge L_4 \wedge L_6 \wedge \frac{\nu(2r)}{10} \wedge \frac{\varepsilon(r/2)}{10},$$

$$H_\infty^{(2)}(r) = \frac{1}{6} \wedge \frac{8\varepsilon(r/2)}{10(5 + 2L_1)} \wedge \frac{8\nu(2r)}{9(2L_2 + 2L_3 + 4)},$$

where the constants L_i are defined in (16).

Under Assumption 2.4 defined below, we shall build consistent estimators for $\beta^\star$ and $\vartheta^\star$ of the model (3).

Assumption 2.4. *Let $T \in \mathbb{N}$, $s \in \mathbb{N}$, $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$, $\eta \in (0,1)$ and a subset $\mathcal{Q} \subset \Theta_T$ of cardinal s.*

(i) ***Regularity of the dictionary φ_T:*** *The dictionary function φ_T satisfies the smoothness conditions of Assumption 2.1. The function g_T defined in (6), satisfies the positivity condition of Assumption 2.2.*

(ii) ***Properties of the function F:*** *Assumption 2.3 holds and we have $\varepsilon(r/2) > 0$ and $\nu(2r) > 0$.*

(iii) ***Proximity to the limit setting:*** *The kernel $\mathcal{K}_T$ defined from the dictionary, see (9), is sufficiently close to the kernel $\mathcal{K}_T^{prox}$ in the sense that we have:*

$$C_T \leq 2$$

and if $s \geq 1$, we have in addition:

$$\mathcal{V}_T \leq H_\infty^{(1)}(r) \quad \textit{and} \quad (s-1)\mathcal{V}_T \leq (1-\eta) H_\infty^{(2)}(r).$$

(iv) ***Separation of the non-linear parameters:*** *If $s \geq 1$, we have:*

$$\delta(\eta H_\infty^{(2)}(r), s) < +\infty \qquad \textit{and for any } \theta \neq \theta' \in \mathcal{Q}, \quad |\theta - \theta'| > \sigma_T\, \Sigma(\eta, r, s),$$

where,

$$\Sigma(\eta, r, s) = 4 \max\left(r g_\infty^{-1/2}, 2\, \delta(\eta H_\infty^{(2)}(r), s) \right).$$

Remark 2.2 (On the separation). We shall perfom the estimation of $\beta^\star$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star)$ from model (3) under the separation condition:

$$|\theta_k^\star - \theta_\ell^\star| \geq \sigma_T\, \Sigma(\eta, r, s), \qquad \text{for all } 1 \leq k, \ell \leq s,\ k \neq \ell, \tag{27}$$

with $\Sigma(\eta, r, s)$ given in *(iv)* of Assumption 2.4. Taking into account the separation condition, the number of admissible features which can be used for the prediction is at most of order $|\Theta_T|/\sigma_T$; this provides a natural upper bound on s. As η is usually fixed, we highlight that the least separation bound tends towards zero when the scaling σ_T goes down to zero.

2.5. Prediction error bound.

We define the estimators $\hat{\beta}$ and $\hat{\vartheta}$ of $\beta^\star$ and $\vartheta^\star$ as the solution to the following regularized optimization problem with a real tuning parameter $\kappa > 0$ and a bound K on the unknown number s of active features in the observed mixture:

$$(\hat{\beta}, \hat{\vartheta}) \in \underset{\beta \in \mathbb{R}^K, \vartheta \in \Theta_T^K}{\operatorname{argmin}} \quad \frac{1}{2} \|y - \beta \Phi_T(\vartheta)\|_{L^2(\lambda_T)}^2 + \kappa \|\beta\|_{\ell_1}, \tag{28}$$

where $\|\cdot\|_{\ell_1}$ corresponds to the usual ℓ_1 norm. Since the interval Θ_T on which the optimization of the non-linear parameters is performed is a compact interval and the function Φ_T is continuous, the existence of at least a solution is guaranteed. The bound K on the number s of features in the mixture from model (3) allows to formulate an optimization problem. It can be arbitrarily large. In particular, it is not involved in the bounds on estimation and prediction risks given in [5] with high probability (see Remark 2.4 therein). We stress that the constants in [5] appearing in those bounds may *a priori* depend on T when the features are scaled by σ_T. We show below that, in fact, those bounds still hold with constants free of T. The results in [5] as well as the proof of Theorem 2.3 below rely on the existence of certificate functions. In [5], sufficient conditions for the certificate functions to exist are given, see Proposition 7.4 and 7.5 therein. Those conditions require the non-linear parameters in $\mathcal{Q}^\star$ to satisfy the separation condition (27). In our framework where the scaling σ_T decreases to zero, it turns out that this separation is in general increasing with s and decreasing with T. However, for some dictionary composed of translated spikes that vanish quickly, it converges to zero when both s and T grow to infinity. We refer to Section 5 in this direction.

Recall the definitions of g_∞ and L_2 given by (15) and (16). The following theorem is a variation of [5, Theorem 2.1].

Theorem 2.3. *Let $T \in \mathbb{N}, s \in \mathbb{N}^*$, $K \in \mathbb{N}^*$, $\eta \in (0,1)$, $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$. Assume we observe the random element y of $L^2(\lambda_T)$ under the regression model (3) with unknown parameters $\beta^\star \in (\mathbb{R}^*)^s$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star)$ a vector with distinct entries in Θ_T, a compact interval of Θ, such that Assumption 2.4 holds for $\mathcal{Q}^\star = \{\theta_1^\star, \cdots, \theta_s^\star\} \subset \Theta_T$. Assume that the unknown number of active features s is bounded by K. Suppose also that the noise process w_T satisfies Assumption 1.1 for a noise level $\overline{\sigma} > 0$ and a decay rate for the noise variance $\Delta_T > 0$.*

Then, there exist finite positive constants $\mathcal{C}_i$, for $i = 0, \ldots, 3$, depending on the function F and on r such that for any $\tau > 1$ and a tuning parameter:

$$\kappa \geq \mathcal{C}_1 \overline{\sigma} \sqrt{\Delta_T \log(\tau)}, \tag{29}$$

we have the prediction error bound of the estimators $\hat{\beta}$ and $\hat{\vartheta}$ defined in (28) given by:

$$\left\| \hat{\beta} \Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_{L^2(\lambda_T)} \leq \mathcal{C}_0 \sqrt{s}\, \kappa, \tag{30}$$

with probability larger than $1 - \mathcal{C}_2 \left(\frac{|\Theta_T|}{\sigma_T\, \tau \sqrt{\log(\tau)}} \vee \frac{1}{\tau} \right)$ where $|\Theta_T|$ is the Euclidean length of Θ_T. Moreover, with the same probability, the difference of the ℓ_1-norms of $\hat{\beta}$ and $\beta^\star$ is bounded by:

$$\Big| \|\, \hat{\beta}\|_{\ell_1} - \|\, \beta^\star\|_{\ell_1} \Big| \leq \mathcal{C}_3\, \kappa\, s. \tag{31}$$

Proof. The proof is similar to the proof of [5, Theorem 2.1] where one replaces the limit kernel noted $\mathcal{K}_\infty$ therein by the approximating kernel $\mathcal{K}_T^{\text{prox}}$ defined in (10). The main difference is in checking condition (v) in Theorem 2.1 on the existence of certificate functions. This is done by using Propositions 7.4 and 7.5 therein, and by noticing that the special form of the approximating kernel $\mathcal{K}_T^{\text{prox}}$ implies that the constants involved do not depend on the scale parameter σ_T. Indeed Equation (17) clearly entails that they do not depend on the scale parameter. The details of the proof are left to the interested reader. Details are given in Section A. □

Notice that even if the constants $\mathcal{C}_i$, for $i = 0, \ldots, 3$, depend only the function F and on r, Assumption 2.4 *(iii)* implies that F is chosen according to the function h. The estimation risks on $\beta^\star$ and $\vartheta^\star$ can be further deduced as in [5, Equations (9-10)].

The following lemma gives an identifiability result for the considered model. It relies on the construction of certificates from [5] and is based on ideas developed in [9] for exact reconstruction of measures, see Lemma 1.1 therein. We recall that by convention $\beta^\star \Phi_T(\vartheta^\star) = 0$ when $s = 0$.

Lemma 2.4 (Sufficient conditions for identifiability)**.** *Let $T \in \mathbb{N}$, $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$, $\eta \in (0,1)$. Suppose that Assumption 2.4 holds for the set $\mathcal{Q}^\star = \{\theta_1^\star, \cdots, \theta_s^\star\} \subset \Theta_T$ of cardinal $s \in \mathbb{N}$ and for the set $\mathcal{Q}^0 = \{\theta_1^0, \cdots, \theta_{s^0}^0\} \subset \Theta_T$ of cardinal $s^0 \in \mathbb{N}$. Then, for any vectors $\beta^\star \in (\mathbb{R}^*)^s, \beta^0 \in (\mathbb{R}^*)^{s^0}$, we have that, up to the same permutation on the components of $\beta^\star$ and $\vartheta^\star$:*

$$\beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0) \quad \text{in } L^2(\lambda_T), \quad \text{implies that} \quad s = s^0, \quad \beta^\star = \beta^0 \quad \text{and} \quad \vartheta^\star = \vartheta^0. \tag{32}$$

Remark 2.5. Recall that if $s \geq 1$, then $\beta^\star$ is a s-dimensional vector with non-zero entries. Under the assumptions of Lemma 2.4 we have that:

$$\beta^\star \Phi_T(\vartheta^\star) = 0 \quad \text{if and only if} \quad s = 0.$$

Remark 2.6. Notice that $\beta^\star \Phi_T(\vartheta^\star) = \beta^0 \Phi_T(\vartheta^0)$ can be re-written as $\tilde{\beta} \Phi_T(\tilde{\vartheta}) = 0$ for some $(\tilde{\beta}, \tilde{\vartheta}) \in \mathbb{R}^{\tilde{s}} \times \Theta_T^{\tilde{s}}$ where the components of $\tilde{\vartheta}$ are the elements of $\mathcal{Q}^\star \cup \mathcal{Q}^0$, $\tilde{s} = \text{Card}(\mathcal{Q}^\star \cup \mathcal{Q}^0)$ and the entries of $\tilde{\beta}$ are up to a sign those of $\beta^\star$ or β^0. In fact, one could show Lemma 2.4 by supposing that Assumption 2.4 stands for the set $\mathcal{Q}^\star \cup \mathcal{Q}^0$. However, as Assumption 2.4 requires pairwise separations between the considered location parameters (see *(iv)* of Assumption 2.4), we remark that this condition would be much stronger than requiring that the sets $\mathcal{Q}^\star$ and $\mathcal{Q}^0$ verify Assumption 2.4 separately.

Proof of Lemma 2.4. First, for $s \geq 1$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star)$ such that Assumption 2.4 stands for the set $\mathcal{Q}^\star$, we show that the application $\beta \mapsto \beta\Phi_T(\vartheta^\star)$ defined from $\mathbb{R}^s$ to $L^2(\lambda_T)$ is injective.

We have that $\|\beta\Phi_T(\vartheta^\star)\|_{L^2(\lambda_T)} = \beta\Gamma\beta^\top$, where $\Gamma \in \mathbb{R}^{s\times s}$ is the symmetric matrix defined by $\Gamma_{k,\ell} = \mathcal{K}_T(\theta_k^\star, \theta_\ell^\star)$. Let $\lambda_{\min}$ be the smallest eigenvalue of Γ. Using Gershgorin's theorem and the definition of $\mathcal{V}_T$ given by (22), we have that:

$$\lambda_{\min} \geq 1 - \max_{1\leq\ell\leq s} \sum_{k=1,k\neq\ell}^{s} |\mathcal{K}_T(\theta_\ell^\star, \theta_k^\star)| \geq 1 - \max_{1\leq\ell\leq s} \sum_{k=1,k\neq\ell}^{s} \left| F\left(\frac{|\theta_\ell^\star - \theta_k^\star|}{\sigma_T}\right)\right| - (s-1)\mathcal{V}_T.$$

The separation condition from Point *(iv)* of Assumption 2.4 implies that for all $k, \ell \in \{1, \cdots, s\}$ such that $k \neq \ell$ we have $|\theta_k^\star - \theta_\ell^\star| \geq \sigma_T \Sigma(\eta, r, s) \geq 8\,\sigma_T\,\delta(\eta H_\infty^{(2)}(r), s)$. Recall the definition of $\delta(u, s)$ given by (25). We deduce that:

$$\max_{1\leq\ell\leq s} \sum_{k=1,k\neq\ell}^{s} \left| F\left(\frac{|\theta_\ell^\star - \theta_k^\star|}{\sigma_T}\right)\right| \leq \eta H_\infty^{(2)}(r).$$

By Point *(iii)* of Assumption 2.4, we have $(s-1)\mathcal{V}_T \leq (1-\eta)H_\infty^{(2)}(r)$ and $H_\infty^{(2)}(r) \leq 1/6$. Thus, we get:

$$\lambda_{\min} \geq 5/6. \tag{33}$$

Hence, the symmetric matrix Γ is positive-definite. This proves that the application $\beta \mapsto \beta\Phi_T(\vartheta^\star)$ is injective from $\mathbb{R}^s$ to $L^2(\lambda_T)$. By symmetry, we obtain for $s^0 \geq 1$ that the application $\beta \mapsto \beta\Phi_T(\vartheta^0)$ is injective from $\mathbb{R}^{s^0}$ to $L^2(\lambda_T)$.

If $s = 0$, we have $\beta^\star\Phi_T(\vartheta^\star) = 0$. For $s^0 \geq 1$, we have $\beta^0 \in (\mathbb{R}^*)^{s^0}$ and since $\beta \mapsto \beta\Phi_T(\vartheta^0)$ is injective, we deduce that $\beta^0\Phi_T(\vartheta^0) \neq 0$. Thus, $s = 0$ and $\beta^\star\Phi_T(\vartheta^\star) = \beta^0\Phi_T(\vartheta^0)$ implies that $s^0 = 0$. By symmetry, $s^0 = 0$ and $\beta^\star\Phi_T(\vartheta^\star) = \beta^0\Phi_T(\vartheta^0)$ implies also that $s = 0$.

Assume from now on that $s, s^0 \in \mathbb{N}^*$ and that $\beta^\star\Phi_T(\vartheta^\star) = \beta^0\Phi_T(\vartheta^0)$. Let us consider the application $v : \mathcal{Q}^\star \mapsto \{-1, 1\}$ defined by: $v(\theta_k^\star) = \text{sign}(\beta_k^\star)$ for any $k \in \{1, \cdots, s\}$. According to Lemma 4.2, there exists $p^\star \in L^2(\lambda_T)$ such that:

$$\|\beta^\star\|_{\ell_1} = \sum_{k=1}^{s} \beta_k^\star \left\langle \phi_T(\theta_k^\star), p^\star\right\rangle_{L^2(\lambda_T)} = \left\langle \beta^\star\Phi_T(\vartheta^\star), p^\star\right\rangle_{L^2(\lambda_T)}.$$

Using the fact that $\beta^\star\Phi_T(\vartheta^\star) = \beta^0\Phi_T(\vartheta^0)$ and Properties *(i)* and *(ii)* of $p^\star$ in Lemma 4.2, we get:

$$\|\beta^\star\|_{\ell_1} = \sum_{k=1}^{s^0} \beta_k^0 \left\langle \phi_T(\theta_k^0), p^\star\right\rangle_{L^2(\lambda_T)} \leq \left\|\beta^0\right\|_{\ell_1}. \tag{34}$$

The role of $(\beta^\star, \vartheta^\star)$ and (β^0, ϑ^0) being symmetric, we also get $\left\|\beta^0\right\|_{\ell_1} \leq \|\beta^\star\|_{\ell_1}$. Hence, we have $\left\|\beta^0\right\|_{\ell_1} = \|\beta^\star\|_{\ell_1}$ and $\text{sign}(\beta_k^0) = \left\langle \phi_T(\theta_k^0), p^\star\right\rangle_{L^2(\lambda_T)}$ for $k \in \{1, \cdots, s^0\}$. Using Properties *(i)* and *(ii)* of $p^\star$ in Lemma 4.2, we remark that for any $\theta \notin \mathcal{Q}^\star$

$$\left|\left\langle \phi_T(\theta), p^\star\right\rangle_{L^2(\lambda_T)}\right| < 1.$$

Thus, we deduce from (34) that $\mathcal{Q}^0 \subseteq \mathcal{Q}^\star$ and by symmetry $\mathcal{Q}^0 = \mathcal{Q}^\star$. Hence, we obtain $\vartheta^\star = \vartheta^0$ (up to a permutation on the components of $\vartheta^\star$) and $s = s^0$. Then use the injectivity of the function $\beta \mapsto \beta\Phi_T(\vartheta^\star)$ to get that $\beta^\star = \beta^0$ (up to the same permutation). This finishes the proof of the Lemma. ☐

3. Goodness-of-fit for the mixture model

In this section, we build a test procedure to decide if the observation y derives from a given mixture of translated features. We build a test Ψ, *i.e.* a measurable function of the observation y taking value in $\{0,1\}$, in order to distinguish a null hypothesis H_0 against an alternative $H_1(\rho)$ depending on a nonnegative separation parameter ρ. We recall that the maximal type I and II error probabilities are $\sup_{(\beta^\star,\vartheta^\star)\in H_0} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[\Psi]$ and $\sup_{(\beta^\star,\vartheta^\star)\in H_1(\rho)} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[1-\Psi]$, respectively, where Ψ is a function of y which is equal to $\beta^\star\Phi_T(\vartheta^\star) + w_T$ under $\mathbb{E}_{(\beta^\star,\vartheta^\star)}$. The maximal testing risk is the sum of the former quantities, that is:

$$R_\rho(\Psi) = \sup_{(\beta^\star,\vartheta^\star)\in H_0} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[\Psi] + \sup_{(\beta^\star,\vartheta^\star)\in H_1(\rho)} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[1-\Psi],$$

and the minimax testing risk is:

$$R^\star_\rho = \inf_{\Psi} R_\rho(\Psi), \tag{35}$$

where the infinimum is taken over all the measurable functions from $L^2(\lambda_T)$ to $\{0,1\}$. The minimax separation rate of the test problem is defined for any $\alpha \in (0,1)$ as:

$$\rho^\star(\alpha) = \inf\{\rho > 0 : R^\star_\rho \le \alpha\}. \tag{36}$$

3.1. **Test problem.** Let $s^0 \in \mathbb{N}$ and consider the set $\Theta_T^{s^0}(\delta^0) \subset \Theta_T^{s^0}$ of vectors whose components are pairwise separated by a distance $\delta^0 \ge 0$ (recall the definition (26)). Consider the vectors $\beta^0 \in (\mathbb{R}^*)^{s^0}$ and $\vartheta^0 = (\theta_1^0, \cdots, \theta_{s^0}^0) \in \Theta_T^{s^0}(\delta^0)$. By convention, we have for $s^0 = 0$ that $\beta^0 = 0$, $\vartheta^0 = 0$ and $\beta^0\Phi_T(\vartheta^0) = 0$.

We build a test procedure based on the observation y to decide, for some $\delta^\star \ge 0$, whether:

$$\begin{cases} H_0: & (\beta^\star,\vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star) \quad \text{such that} \quad \beta^\star\Phi_T(\vartheta^\star) = \beta^0\Phi_T(\vartheta^0), \\ H_1(\rho): & (\beta^\star,\vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star) \quad \text{such that} \quad \left\|\beta^\star\Phi_T(\vartheta^\star) - \beta^0\Phi_T(\vartheta^0)\right\|_{L^2(\lambda_T)} \ge \rho, \end{cases} \tag{37}$$

where ρ is a nonnegative separation parameter. When Assumption 2.4 holds for the sets $\mathcal{Q}^\star = \{\theta_1^\star, \cdots, \theta_s^\star\}$ and $\mathcal{Q}^0 = \{\theta_1^0, \cdots, \theta_{s^0}^0\}$, by Lemma 2.4, the null hypothesis implies that $(\beta^\star,\vartheta^\star) = (\beta^0,\vartheta^0)$ (up to the same permutation on the components of $\beta^\star$ and $\vartheta^\star$). We remark that the separation condition from Point *(iv)* of Assumption 2.4 required between the elements of $\mathcal{Q}^\star$ (resp. $\mathcal{Q}^0$) is automatically satisfied when $\delta^\star \ge \sigma_T\, \Sigma(\eta, r, s)$ (resp. $\delta^0 \ge \sigma_T\, \Sigma(\eta, r, s^0)$).

We shall denote the distribution under the null hypothesis as associated to the parameters (β^0,ϑ^0) and see that the maximal type I error probability writes in this case $\mathbb{E}_{(\beta^0,\vartheta^0)}[\Psi]$ for $\mathbb{E}_{(\beta^\star,\vartheta^\star)}[\Psi]$. Furthermore, when $s^0 = 0$, under Assumption 2.4 for the set $\mathcal{Q}^\star$, Lemma 2.4 implies that the null hypothesis reduces to $H_0 : s = 0$.

3.2. **Main results.** We consider the test procedure $\Psi_{\mathcal{T}}(t)$ associated to a real valued statistic $\mathcal{T}$ (measurable function of the observation y) and a threshold $t > 0$ (defining a critical region) given by:

$$\Psi_{\mathcal{T}}(t) = \mathbf{1}_{\{|\mathcal{T}|>t\}}. \tag{38}$$

We recall that for a test Ψ, we accept H_0 when $\Psi = 0$ and reject it when $\Psi = 1$.

Let $s^0 \in \mathbb{N}$ and consider known linear coefficients and location parameters $\beta^0 \in (\mathbb{R}^*)^{s^0}$ and $\vartheta^0 = (\theta_1^0, \cdots, \theta_{s^0}^0) \in \Theta_T^{s^0}$, respectively. We define two statistics $\mathcal{T}_1$ and $\mathcal{T}_2$ by:

$$\mathcal{T}_1 = \left\|y - \beta^0\Phi_T(\vartheta^0)\right\|^2_{L^2(\lambda_T)} - \mathbb{E}\left[\|w_T\|^2_{L^2(\lambda_T)}\right] \quad \text{and } \mathcal{T}_2 = \left\|\hat\beta\Phi_T(\hat\vartheta) - \beta^0\Phi_T(\vartheta^0)\right\|^2_{L^2(\lambda_T)}, \tag{39}$$

where $\hat\beta$ and $\hat\vartheta$ denote the estimators obtained from (28) for a given value of the tuning parameter κ and a bound K on the unknown number $s \in \mathbb{N}$ of active features in the observed signal.

Recall the definition (4) of Ξ_T, the variance of the squared $L^2(\lambda_T)$-norm of the noise w_T. The following theorem gives an upper bound of the maximal testing risk associated to the test $\Psi_{\mathcal{T}_1}(t)$ for some positive threshold t and positive separation ρ.

Theorem 3.1. *Let $T \in \mathbb{N}$ and $s^0 \in \mathbb{N}$. Let $\delta^\star \geq 0$ and $\delta^0 \geq 0$. Assume that we observe the random element y of $L^2(\lambda_T)$ under the regression model (3) with unknown parameters $s \in \mathbb{N}$, $\beta^\star \in (\mathbb{R}^*)^s$ and $\vartheta^\star \in \Theta_T^s(\delta^\star)$. Let $\beta^0 \in (\mathbb{R}^*)^{s^0}$ and $\vartheta^0 \in \Theta_T^{s^0}(\delta^0)$. Suppose that Assumption 2.1 on the smoothness of the features holds. Suppose that Assumption 1.1 holds for a noise level $\overline{\sigma} > 0$ and a decay rate for the noise variance $\Delta_T > 0$.*

Then, the test $\Psi_{\mathcal{T}_1}$ in (38) using $\mathcal{T}_1$ in (39) satisfies:

$$R_\rho\left(\Psi_{\mathcal{T}_1}(t)\right) \leq \frac{\Xi_T}{t^2} + \frac{4\,\Xi_T}{(\rho^2-t)^2} + \mathrm{e}^{-(\rho^2-t)^2/(32\overline{\sigma}^2\Delta_T\rho^2)}, \tag{40}$$

for any threshold t and any separation ρ such that $\rho^2 > t > 0$.

Proof. We give a bound of the type I error probability. Using that under H_0 we have $y = \beta^0\Phi_T(\vartheta^0) + w_T$, we get:

$$\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_1}(t)] = \mathbb{P}\left(\left|\ \|w_T\|^2_{L^2(\lambda_T)} - \mathbb{E}\left[\|w_T\|^2_{L^2(\lambda_T)}\right]\right| > t\right).$$

Using Chebyshev's inequality, we obtain:

$$\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_1}(t)] \leq \frac{\Xi_T}{t^2}. \tag{41}$$

We now give a bound of the type II error probability. We set:

$$R = \left\|\beta^0\Phi_T(\vartheta^0) - \beta^\star\Phi_T(\vartheta^\star)\right\|_{L^2(\lambda_T)},$$

where $(\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star)$. Using the decomposition of y from the model (3) and the triangle inequality, we have:

$$|\mathcal{T}_1| \geq R^2 - \left|\|w_T\|^2_{L^2(\lambda_T)} - \mathbb{E}[\|w_T\|^2_{L^2(\lambda_T)}]\right| - 2\left|\left\langle \beta^0\Phi_T(\vartheta^0) - \beta^\star\Phi_T(\vartheta^\star), w_T\ \right\rangle_{L^2(\lambda_T)}\right|.$$

Notice that by Assumption 1.1, the random variable $\left\langle \beta^0\Phi_T(\vartheta^0) - \beta^\star\Phi_T(\vartheta^\star), w_T\ \right\rangle_{L^2(\lambda_T)}$ is Gaussian with zero mean and variance bounded by $\overline{\sigma}^2\,\Delta_T\,R^2$. Hence, using that under $H_1(\rho)$ we have $R \geq \rho$, we obtain:

$$\begin{aligned}\mathbb{E}_{(\beta^\star,\,\vartheta^\star)}[1 - \Psi_{\mathcal{T}_1}(t)] \leq &\mathbb{P}\left((\rho^2 - t)/2 \leq \left|\|w_T\|^2_{L^2(\lambda_T)} - \mathbb{E}[\|w_T\|^2_{L^2(\lambda_T)}]\right|\right)\\ &+ \mathbb{P}\left((R^2 - t)/2 \leq 2\overline{\sigma}\sqrt{\Delta_T}\,R\,|G|\right),\end{aligned} \tag{42}$$

where G is a standard Gaussian random variable. On the one hand, for $t < \rho^2$, using Chebyshev's inequality we get:

$$\mathbb{P}\left((\rho^2 - t)/2 \leq \left|\|w_T\|^2_{L^2(\lambda_T)} - \mathbb{E}[\|w_T\|^2_{L^2(\lambda_T)}]\right|\right) \leq \frac{4\,\Xi_T}{(\rho^2-t)^2}. \tag{43}$$

On the other hand, we have:

$$\mathbb{P}\left((R^2 - t)/2 \leq 2\overline{\sigma}\sqrt{\Delta_T}\,R\,|G|\right) \leq \mathbb{P}\left(\frac{\rho^2 - t}{4\overline{\sigma}\sqrt{\Delta_T}\rho} \leq |G|\right) \leq \mathrm{e}^{-(\rho^2-t)^2/(32\overline{\sigma}^2\Delta_T\rho^2)}. \tag{44}$$

where we used that $\rho \leq R$ and the tail bound (see [1, Formula 7.1.13]):

$$\frac{1}{\sqrt{2\pi}}\int_u^{+\infty} \mathrm{e}^{-t^2/2}\,\mathrm{d}t \leq \frac{1}{2}\,\mathrm{e}^{-u^2/2}, \qquad \text{for } u > 0. \tag{45}$$

By combining (42) with (43) and (44), we get the following bound on the type II error probability:

$$\mathbb{E}_{(\beta^\star,\vartheta^\star)}[1-\Psi_{\mathcal{T}_1}(t)] \leq \frac{4\,\Xi_T}{(\rho^2-t)^2} + \mathrm{e}^{-(\rho^2-t)^2/(32\overline{\sigma}^2\Delta_T\rho^2)}. \tag{46}$$

Then, by putting together (41) and (46), we obtain (40). □

We deduce from Theorem 3.1 upper bounds on the minimax separation $\rho^\star$ defined in (36) for the goodness-of-fit test problem (37).

Corollary 3.2. *Under the framework and the assumptions of Theorem 3.1, the minimax separation rate for the test problem* (37) *verifies for any* $\alpha\in(0,1)$*:*

$$\rho^\star(\alpha)\leq\rho^{(1)}(\alpha) \quad \textit{with} \quad \rho^{(1)}(\alpha):=\max\left(\left(\frac{40\Xi_T}{\alpha}\right)^{1/4}, 8\,\overline{\sigma}\sqrt{2\Delta_T\log\left(\frac{2}{\alpha}\right)}\right). \tag{47}$$

Proof of Corollary 3.2. This result is a direct consequence of Theorem 3.1 by taking the threshold t of the test therein equal to $\rho^2/2$. Then, we have that for $\rho>0$:

$$R_\rho^\star \leq R_\rho\left(\Psi_{\mathcal{T}_1}(\rho^2/2)\right) \leq \frac{4\Xi_T}{\rho^4}+\frac{16\,\Xi_T}{\rho^4}+\mathrm{e}^{-\rho^2/(128\overline{\sigma}^2\Delta_T)} = \frac{20\,\Xi_T}{\rho^4}+\mathrm{e}^{-\rho^2/(128\overline{\sigma}^2\Delta_T)}.$$

We deduce that $R_\rho^\star\leq\alpha$ for any $\alpha\in(0,1)$ whenever the separation ρ satisfies:

$$\rho\geq\left(\frac{40\Xi_T}{\alpha}\right)^{\frac14}\vee\overline{\sigma}\sqrt{128\,\Delta_T\log\left(\frac{2}{\alpha}\right)}. \tag{48}$$

This implies (47). □

In the following theorem, we give a bound of the maximal testing risk associated to the test $\Psi_{\mathcal{T}_2}(t)$ using $\mathcal{T}_2$ in (39) for solving the test problem (37). The statistic $\mathcal{T}_2$ is defined using estimators of the model parameters $(\beta^\star,\vartheta^\star)$. In view of recovering the latter, we assume that the minimal distance $\delta^\star$ (resp. δ^0) is large enough so that Point *(iv)* of Assumption 2.4 is satisfied for the components of $\vartheta^\star$ (resp. ϑ^0).

Recall the definitions of g_∞ and L_2 given by (15) and (16), that $|\Theta_T|$ denotes the Euclidean length of the compact set Θ_T and Σ defined in *(iv)* of Assumption 2.4.

Theorem 3.3. *Let* $T\in\mathbb{N}$*,* $s^0\in\mathbb{N}$ *and choose* $K\in\mathbb{N}$ *such that* $s_0\leq K$*. Let also* $\eta\in(0,1)$ *and* $r\in\left(0,1/\sqrt{2\,g_\infty\,L_2}\right)$*. Let* $\delta^\star\geq\sigma_T\,\Sigma(\eta,r,s)$ *and* $\delta^0\geq\sigma_T\,\Sigma(\eta,r,s^0)$*. Assume we observe the random element* y *of* $L^2(\lambda_T)$ *under the regression model (3) with unknown parameters* $s\in\mathbb{N}$ *such that* $s\leq K$*,* $\beta^\star\in(\mathbb{R}^*)^s$ *and* $\vartheta^\star=(\theta_1^\star,\cdots,\theta_s^\star)\in\Theta_T^s(\delta^\star)$*. Let* $\beta^0\in(\mathbb{R}^*)^{s^0}$ *and* $\vartheta^0=(\theta_1^0,\cdots,\theta_{s^0}^0)\in\Theta_T^{s^0}(\delta^0)$*. Suppose that Assumption 2.4 holds for the sets* $\mathcal{Q}^\star=\{\theta_1^\star,\cdots,\theta_s^\star\}\subset\Theta_T$ *of cardinal* s *and* $\mathcal{Q}^0=\{\theta_1^0,\cdots,\theta_{s^0}^0\}\subset\Theta_T$ *of cardinal* s^0*. Suppose also that the noise process* w_T *satisfies Assumption 1.1 for a noise level* $\overline{\sigma}>0$ *and a decay rate for the noise variance* $\Delta_T>0$*.*

Then, there exist finite positive constants $\mathcal{C}_0$*,* $\mathcal{C}_1$*,* $\mathcal{C}_2$*, depending on* r *and on the function* F*, such that for the tuning parameter* κ*:*

$$\kappa\geq\mathcal{C}_1\overline{\sigma}\sqrt{\Delta_T\log(\tau)}, \quad \textit{for some } \tau>1, \tag{49}$$

the test $\Psi_{\mathcal{T}_2}$ *using* $\mathcal{T}_2$ *in* (39) *satisfies:*

$$R_\rho\left(\Psi_{\mathcal{T}_2}(t)\right)\leq 2\,\mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\sqrt{\log(\tau)}}\vee\frac{1}{\tau}\right), \tag{50}$$

for any threshold t *and any separation* ρ *satisfying:*

$$0<t, \quad \mathcal{C}_0\sqrt{s^0}\,\kappa\leq\sqrt{t}<\rho \quad \textit{and} \quad \sqrt{t}+\mathcal{C}_0\sqrt{s}\,\kappa\leq\rho. \tag{51}$$

Remark 3.4 (On the bound K). The bound K on s is assumed to be known. It is needed to formulate the optimization problem (28) whose solutions are the estimators of $\beta^\star$ and $\vartheta^\star$. However, we stress that the constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$ and the bound on the maximal testing risk do not depend on K. Thus, K can be taken arbitrarily large.

Proof of Theorem 3.3. **Case** $s > 0$**.** Let $(\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star)$. We consider the estimators $(\hat{\beta}, \hat{\vartheta})$ defined in (28). Notice that the hypotheses of Theorem 2.3 are in force. We use the constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$ defined therein. Under H_0, we have $s = s^0$. Thus, for $\sqrt{t} \geq \mathcal{C}_0 \sqrt{s}\,\kappa$, we get the following bound on the type I error probability:

$$\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_2}(t)] \leq \mathbb{P}\left(\left\|\hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star\Phi_T(\vartheta^\star)\right\|_{L^2(\lambda_T)} > \mathcal{C}_0\sqrt{s}\,\kappa\right) \leq \mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right), \tag{52}$$

where we used that $\beta^0\Phi_T(\vartheta^0) = \beta^\star\Phi_T(\vartheta^\star)$ and that $\sqrt{t} \geq \mathcal{C}_0\sqrt{s}\,\kappa$ for the first inequality and Theorem 2.3 for the second.

We now bound the type II error probability. Under $H_1(\rho)$, since $\left\|\beta^\star\Phi_T(\vartheta^\star) - \beta^0\Phi_T(\vartheta^0)\right\|_{L^2(\lambda_T)} \geq \rho$, we obtain that:

$$\mathbb{E}_{(\beta^\star,\,\vartheta^\star)}[1 - \Psi_{\mathcal{T}_2}(t)] \leq \mathbb{P}\left(\rho - \sqrt{t} \leq \left\|\hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star\Phi_T(\vartheta^\star)\right\|_{L^2(\lambda_T)}\right) \leq \mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right), \tag{53}$$

where we used the triangle inequality for the first inequality and Theorem 2.3 as well as $\rho - \sqrt{t} \geq \mathcal{C}_0\sqrt{s}\,\kappa$ for the second.

Case $s = 0$**.** Since $s = 0$, we have $y = w_T$ according to (3). Let us first bound the type I error probability $\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_2}(t)]$. Assume that the hypothesis H_0 holds so that $s = s^0 = 0$. By definition we have:

$$\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_2}(t)] = \mathbb{P}\left(\left\|\hat{\beta}\Phi_T(\hat{\vartheta})\right\|^2_{L^2(\lambda_T)} > t\right).$$

We get from the definition of the estimators $\hat{\beta}$ and $\hat{\vartheta}$ from (28) that:

$$\frac{1}{2}\left\|w_T - \hat{\beta}\Phi_T(\hat{\vartheta})\right\|^2_{L^2(\lambda_T)} + \kappa\left\|\hat{\beta}\right\|_{\ell_1} \leq \frac{1}{2}\|w_T\|^2_{L^2(\lambda_T)}.$$

By rearranging some terms in the equation above, we get:

$$\frac{1}{2}\left\|\hat{\beta}\Phi_T(\hat{\vartheta})\right\|^2_{L^2(\lambda_T)} \leq \left\langle\hat{\beta}\Phi_T(\hat{\vartheta}), w_T\right\rangle_{L^2(\lambda_T)} - \kappa\left\|\hat{\beta}\right\|_{\ell_1} \leq \left\|\hat{\beta}\right\|_{\ell_1}\left(\sup_{\Theta_T}|\langle\phi_T(\theta), w_T\rangle_{L^2(\lambda_T)}| - \kappa\right). \tag{54}$$

Let us define the event:

$$\mathcal{A} = \{\sup_{\theta\in\Theta_T}|\langle\phi_T(\theta), w_T\rangle_{L^2(\lambda_T)}| < \kappa\}. \tag{55}$$

We deduce from (54) that on the event $\mathcal{A}$ we have $\left\|\hat{\beta}\Phi_T(\hat{\vartheta})\right\|_{L^2(\lambda_T)} = 0$. Therefore we get:

$$\mathbb{E}_{(\beta^0,\,\vartheta^0)}[\Psi_{\mathcal{T}_2}(t)] \leq \mathbb{P}\left(\left\|\hat{\beta}\Phi_T(\hat{\vartheta})\right\|_{L^2(\lambda_T)} > 0\right) \leq \mathbb{P}(\mathcal{A}^c). \tag{56}$$

We shall bound later $\mathbb{P}(\mathcal{A}^c)$, see (58).

We now consider the type II error probability. We asume H_1, that is $\left\|\beta^0\Phi_T(\vartheta^0)\right\|_{L^2(\lambda_T)} \geq \rho$. We obtain:

$$\mathbb{E}_{(\beta^\star,\,\vartheta^\star)}[1 - \Psi_{\mathcal{T}_2}(t)] = \mathbb{P}\left(\left\|\hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^0\Phi_T(\vartheta^0)\right\|_{L^2(\lambda_T)} \leq \sqrt{t}\right) \leq \mathbb{P}\left(\rho - \sqrt{t} \leq \left\|\hat{\beta}\Phi_T(\hat{\vartheta}))\right\|_{L^2(\lambda_T)}\right) \leq \mathbb{P}(\mathcal{A}^c). \tag{57}$$

where we used the definition of $\mathcal{T}_2$ and the triangle inequality for the first inequality, the second inequality of (56) as well as $\rho - \sqrt{t} > 0$ for the second.

We shall apply [5, Lemma A.1] to bound $\mathbb{P}(\mathcal{A}^c)$. It amounts to controling the supremum of the Gaussian process $\theta \mapsto \langle \phi_T(\theta), w_T \rangle_{L^2(\lambda_T)}$. Recall that Assumptions 2.1 and 2.2 hold. The function ϕ_T is of class $\mathcal{C}^1$ from the interval Θ_T to $L^2(\lambda_T)$, with Θ_T a sub-interval of Θ. We have also, with $\phi_T^{[1]} = \tilde{D}_{1;\mathcal{K}_T}[\phi_T]$, that:

$$\|\phi_T(\theta)\|_{L^2(\lambda_T)} = 1 \quad \text{and} \quad \left\| \phi_T^{[1]}(\theta) \right\|^2_{L^2(\lambda_T)} = \mathcal{K}_T^{[1,1]}(\theta, \theta) = 1.$$

Since Assumption 1.1 on the noise w_T holds, the hypotheses of [5, Lemma A.1] hold and we deduce from [5, Lemma A.1] (with $C_1 = C_2 = 1$ therein) that:

$$\mathbb{P}(\mathcal{A}^c) = \mathbb{P}\left(\sup_{\theta \in \Theta_T} |\langle \phi_T(\theta), w_T \rangle_{L^2(\lambda_T)}| \geq \kappa \right) \leq 3 \cdot \left(\frac{2\overline{\sigma}\sqrt{g_\infty}|\Theta_T|\sqrt{\Delta_T}}{\sigma_T \kappa} \vee 1 \right) \mathrm{e}^{-\kappa^2/(4\overline{\sigma}^2\Delta_T)},$$

where the diameter $|\Theta_T|_{\mathfrak{d}_T}$ of the set Θ_T with respect to the metric $\mathfrak{d}_T$ is bounded by $2\sqrt{g_\infty}|\Theta_T|/\sigma_T$ using (21) and the fact that $C_T \leq 2$. By taking $\kappa \geq 2\overline{\sigma}\sqrt{\Delta_T \log(\tau)}$, we get:

$$\mathbb{P}(\mathcal{A}^c) = \mathbb{P}\left(\sup_{\theta \in \Theta_T} |\langle \phi_T(\theta), w_T \rangle_{L^2(\lambda_T)}| \geq \kappa \right) \leq 3 \cdot \left(\frac{\sqrt{g_\infty}|\Theta_T|}{\sigma_T \tau \sqrt{\log(\tau)}} \vee \frac{1}{\tau} \right). \tag{58}$$

Notice that the constant $\mathcal{C}_2$ from Theorem 2.3 is equal to $2\sqrt{g_\infty}\,\mathcal{C}_2'$ where $\mathcal{C}_2'$ is given by [5, $\mathcal{C}_2$ from Eq. (84) therein] and is greater than 3. The constant $\mathcal{C}_2$ depends only on r and the function F. Finally, by putting together (52), (53), (56) and (57), we obtain for $\kappa \geq \mathcal{C}_1\overline{\sigma}\sqrt{\Delta_T \log(\tau)}$ (where the constant $\mathcal{C}_1$ is defined in [5, Proof of Theorem 2.1 (p.32)] and is superior to 4) the bound on the maximal testing risk from Theorem 3.3. This finishes the proof. □

In the next Corollary, we obtain an additionnal upper bound on the minimax separation rate.

Corollary 3.5. *Under the framework and the assumptions of Theorem 3.3 and provided that $|\Theta_T|/\sigma_T \geq 1$, there exist finite positive constants c and C, depending on r and the function F, such that the minimax separation rate for the test problem (37) verifies for any $\alpha \in (0, 1)$:*

$$\rho^\star(\alpha) \leq \rho^{(2)}(\alpha), \quad \rho^{(2)}(\alpha) := C\,\overline{\sigma}\sqrt{(s \vee s^0 \vee 1)\Delta_T \log\left(\frac{c\,|\Theta_T|}{\alpha\,\sigma_T}\right)}. \tag{59}$$

Remark 3.6 (On the condition $|\Theta_T|/\sigma_T \geq 1$). We recall that the set Θ_T is a compact subset of Θ. In the case where Θ is the torus $\mathbb{R}/\mathbb{Z}$, $\Theta_T = \Theta$ and the scale parameter σ_T tends towards 0 when T grows to infinity, the condition $|\Theta_T|/\sigma_T \geq 1$ is satisfied for T large enough. This condition also holds for T large enough in the Gaussian spikes deconvolution example, with the particular choices for Θ_T and σ_T from Section 5, where $\Theta = \mathbb{R}$, $\lim_{T\to+\infty} \Theta_T = \Theta$ and $\lim_{T\to+\infty} \sigma_T = 0$.

Proof of Corollary 3.5. Notice that all the assumptions of Theorem 3.3 are in force. The result is a direct consequence of Theorem 3.3. We fix the tuning parameter $\kappa = \mathcal{C}_1\overline{\sigma}\sqrt{\Delta_T \log(\tau)}$ by taking the equality in (49). Then, for

$$\rho \geq \mathcal{C}_0\sqrt{s \vee 1}\,\kappa + \sqrt{t} \quad \text{and} \quad t = \mathcal{C}_0^2\,(s^0 \vee 1)\,\kappa^2, \tag{60}$$

we have (51) (in particular $0 < t < \rho$) and by Theorem 3.3 for $\tau > 1$:

$$R_\rho^\star \leq R_\rho\left(\Psi_{\mathcal{T}_2}(t)\right) \leq 2\mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau} \right),$$

where the finite positive constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$, from Theorem 3.3 depend on r and F.

Then, taking $\tau = c|\Theta_T|/(\alpha\sigma_T)$ with $c = (2\mathcal{C}_2) \vee \mathrm{e}$ and using that by assumption $|\Theta_T|/\sigma_T \geq 1$, we get for $\rho \geq \sqrt{2}\mathcal{C}_0\mathcal{C}_1\overline{\sigma}\sqrt{(s+s^0) \vee 2}\sqrt{\Delta_T \log(c|\Theta_T|/(\alpha\sigma_T))}$ and $\alpha \in (0,1)$ that $R^\star_\rho \leq \alpha$. We readily deduce (59) with $C = 2\mathcal{C}_0\mathcal{C}_1$. □

Remark 3.7 (Combining the upper bounds of Corollaries 3.2 and 3.5). Let $\alpha \in (0,1)$. Suppose that the assumptions of Corollaries 3.2 and 3.5 hold. Previous results show that each procedure may perform better than the other one in convenient regimes of the parameters, involving the unknown parameter s. In order to aggregate the two procedures into an automatic one, we take the maximum of the two test procedures. This aggregated test procedure rejects as soon as at least one of the procedures rejects, and accepts otherwise.

More precisely, let $\rho^{(1)}(\alpha/2)$ be defined by (47) with α replaced by $\alpha/2$ and set $t^{(1)} = (\rho^{(1)}(\alpha/2))^2/2$; and let $\rho^{(2)}(\alpha/2)$ be defined in (59) and $t^{(2)}$ be given by (60) with α replaced by $\alpha/2$. Then, Corollaries 3.2 and 3.5 imply that $R_{\rho^{(1)}}\left(\Psi_{\mathcal{T}_1}(t^{(1)})\right) \leq \alpha/2$ and $R_{\rho^{(2)}}\left(\Psi_{\mathcal{T}_2}(t^{(2)})\right) \leq \alpha/2$. We define the test:

$$\Psi^{\max} = \max(\Psi_{\mathcal{T}_1}(t^{(1)}), \Psi_{\mathcal{T}_2}(t^{(2)})).$$

It is straightforward to see that the type I error probability satisfies:

$$\sup_{(\beta^\star,\vartheta^\star)\in H_0} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[\Psi^{\max}] \leq \alpha.$$

Moreover, we have for $\rho^{\min}(\alpha) = \rho^{(1)}(\alpha/2) \wedge \rho^{(2)}(\alpha/2)$ the following bound on the type II error probability:

$$\sup_{(\beta^\star,\vartheta^\star)\in H_1(\rho^{\min})} \mathbb{E}_{(\beta^\star,\vartheta^\star)}[1 - \Psi^{\max}] \leq \alpha/2.$$

Therefore, we deduce an upper bound on $\rho^\star(\alpha)$ of order $\rho^{\min}(\alpha)$, that is:

$$\rho^{\min}(\alpha) = \min\left(\left(\frac{80\Xi_T}{\alpha}\right)^{1/4}, C\overline{\sigma}\sqrt{(s \vee s^0 \vee 1)\Delta_T \log\left(\frac{2\,c\,|\Theta_T|}{\alpha\,\sigma_T}\right)}\right), \tag{61}$$

for a positive constant $c \geq 2$. We identify two regimes depending on whether the observed signal is sparse or not. Indeed, we notice that when α is fixed and:

$$s \vee s^0 \vee 1 \ll \left(\frac{\Xi_T}{\alpha}\right)^{1/2} \cdot \left(\overline{\sigma}^2\Delta_T \log\left(\frac{2\,c\,|\Theta_T|}{\alpha\,\sigma_T}\right)\right)^{-1},$$

Corollary 3.5 yields a sharper upper bound on the separation rate than Corollary 3.2.

3.3. **Minimax separation rates for signal detection.** We illustrate our results on a simple model motivated by [14] for sparse linear regression. We consider a discrete-time process y over a regular grid $t_1 < \cdots < t_T$ on $\Theta = \mathbb{R}/\mathbb{Z}$ with grid step $\Delta_T = 1/T$. We set λ_T and w_T as in Section 1.2.1. We recall that $\Xi_T = 2\overline{\sigma}^4\Delta_T^2 T$ where $\overline{\sigma} > 0$ is the noise level. In the following, we assume without any loss of generality that $\overline{\sigma} = 1$.

Let us consider the framework of signal detection when $s^0 = 0$. Under the assumptions of Corollary 3.5, the test problem (37) reduces to:

$$\begin{cases} H_0: & \beta^\star = 0, \\ H_1(\rho): & (\beta^\star,\vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star) \quad \text{such that} \quad \|\beta^\star\Phi_T(\vartheta^\star)\|_{L^2(\lambda_T)} \geq \rho. \end{cases} \tag{62}$$

Moreover, under the assumptions of Corollary 3.5 and with the same arguments used to establish (33), we can show that:

$$5/6 \leq C_{\min} := \min_\beta \frac{\|\beta\Phi_T(\vartheta^\star)\|_{L^2(\lambda_T)}}{\|\beta\|_{\ell_2}} \quad \text{and} \quad C_{\max} := \max_\beta \frac{\|\beta\Phi_T(\vartheta^\star)\|_{L^2(\lambda_T)}}{\|\beta\|_{\ell_2}} \leq 7/6. \tag{63}$$

Therefore, the separation in the alternative hypothesis $H_1(\rho)$ can be formulated as a lower bound on $\|\beta^\star\|_{\ell_2}$ since we have:

$$C_{\min}\|\beta^\star\|_{\ell_2} \leq \|\beta^\star\Phi_T(\vartheta^\star)\|_{L^2(\lambda_T)} \leq C_{\max}\|\beta^\star\|_{\ell_2}.$$

We set $\Theta_T = \Theta$ and thus $|\Theta_T| = 1$. We get from (61) the following upper bound on $\rho^\star(\alpha)$ for any $\alpha \in (0,1)$:

$$\rho(\alpha) = C\min\left(\frac{1}{(\alpha T)^{\frac{1}{4}}}, \sqrt{\frac{s}{T}\log\left(\frac{c}{\alpha\,\sigma_T}\right)}\right), \tag{64}$$

with C a finite positive constant. Let $(\alpha_T, T \geq 1)$ be a $(0,1)$-valued sequence which converges to zero when T grows to infinity. We deduce that:

$$\lim_{s,T\to+\infty} R^\star_{\rho(\alpha_T)} = 0.$$

By letting the sequence $(\alpha_T, T \geq 1)$ converge towards 0 as slow as we want, we deduce that for a sequence of separations $(\rho_{s,T},\, T \geq 1, s \geq 1)$ such that:

$$\lim_{s,T\to+\infty} \frac{\rho_{s,T}}{\frac{1}{T^{\frac{1}{4}}} \wedge \sqrt{\frac{s}{T}\log\left(\frac{c}{\sigma_T}\right)}} = +\infty, \tag{65}$$

we have:

$$\lim_{s,T\to+\infty} R^\star_{\rho_{s,T}} = 0.$$

Hence, we have obtained an asymptotic upper bound of the minimax separation associated to the detection of a mixture issued from a continuous dictionary.

We now compare this upper bound to the asymptotic lower bound obtained in the case where the dictionary contains a finite number of features instead of a continuum. Assume that the dictionary is fixed, known and contains p features parametrized by the parameters in the known and fixed set $\mathcal{Q}^0 = \{\theta_1^0, \cdots, \theta_p^0\} \subset \Theta_T$. We consider the high dimensional linear regression model:

$$y = \beta^\star\Phi_T(\vartheta^0) + w_T \quad \text{in } L^2(\lambda_T),$$

with $\vartheta^0 = (\theta_1^0, \cdots, \theta_p^0) \in \Theta_T^p$ and where $\beta^\star \in \mathbb{R}^p$ is a s-sparse vector. Notice that in this model the entries of $\beta^\star$ can take the value 0. The high dimension comes from the fact that p can be much larger than T. Under coherence assumptions on the finite dictionary and for a sequence of separations $(\rho_{s,T},\, T \geq 1, s \geq 1)$ such that:

$$\lim_{s,T\to+\infty} \frac{\rho_{s,T}}{\frac{1}{T^{\frac{1}{4}}} \wedge \sqrt{\frac{s}{T}\log(p)} \wedge \frac{p^{\frac{1}{4}}}{\sqrt{T}}} = 0, \tag{66}$$

the authors of [14] showed for different hypotheses on the design matrix $\Phi_T(\vartheta^0)$ that:

$$\lim_{s,T\to+\infty} R^\star_{\rho_{s,T}} = 1.$$

It means that the hypotheses (62) cannot be distinguished asymptotically when the separation converges to zero faster than the rate given by (66). We remark that in the high dimensional framework (*i.e.* $T < p$), we get only the first two regimes in (66) since $1/T^{1/4} < p^{1/4}/\sqrt{T}$.

4. Goodness-of-fit of the dictionary

In spectroscopy, a prescribed material has known chemical components and a list of s_0 corresponding location parameters of the features is provided. From a sampled material we want to decide whether its chemical components are included in the prescribed list. The linear coefficients are non-negative in this case and they are not given, which makes the null hypothesis composite, that is, fixed location parameters and varying positive linear coefficients. We generalize this setup to real valued linear coefficients. Under the null hypothesis the location parameters are still fixed, but the linear coefficients vary with fixed sign.

More precisely, let $s^0 \in \mathbb{N}$ and let $\mathcal{Q}^0 = \{\theta_1^0, \cdots, \theta_{s_0}^0\} \subset \Theta_T$ be a set of known location parameters pairwise separated by a distance $\delta^0 \geq 0$ so that the model is identifiable, see Lemma 2.4. We set the vector $\vartheta^0 = (\theta_1^0, \cdots, \theta_{s^0}^0)$. Let $v^0 = (v_1^0, \ldots, v_{s^0}^0)$ be a vector in $\{-1, 1\}^{s^0}$ that contains the common signs of all linear coefficients under the null hypothesis. Consider two disjoint subsets of the set $\mathcal{Q}^0$ associated to linear coefficients with sign $\epsilon = \pm 1$: $\mathcal{Q}^{0,\epsilon} = \{\theta_k^0 \in \mathcal{Q}^0 : \epsilon v_k^0 > 0\}$. Let $s \in \mathbb{N}^*$. Assume that we observe a random element y issued from the model (3) with linear coefficients $\beta^\star \in (\mathbb{R}^*)^s$ and non-linear parameters $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star) \in \Theta_T^s$. We test if the unknown set $\mathcal{Q}^{\star,\epsilon} = \{\theta_k^\star \in \mathcal{Q}^\star : \epsilon\beta_k^\star > 0\}$ is a subset of $\mathcal{Q}^{0,\epsilon}$ for each $\epsilon = \pm 1$. If $s^0 = 0$, this amounts to testing that $\mathcal{Q}^\star$ is empty, which corresponds to the signal detection framework presented in Section 3 in the case $s^0 = 0$. Hence, we shall assume in this section that $s_0 \geq 1$. For example, if $\mathcal{Q}^{0,-}$ is empty, this amounts to testing that $\mathcal{Q}^\star$ is a subset of $\mathcal{Q}^0$ and $\beta^\star$ has positive entries.

4.1. A measure of discrepancy between dictionaries.

We define the closed balls centered at $\theta \in \Theta_T$ with radius r by:

$$\mathcal{B}_T(\theta, r) = \{\theta' \in \Theta_T : \mathfrak{d}_T(\theta, \theta') \leq r\} \subseteq \Theta_T.$$

Let us define for $\epsilon = \pm 1$ the set of indices $\mathcal{I}^\epsilon = \{k \in \{1, \ldots, s^0\}, \epsilon v_k^0 > 0\}$. We introduce for $r > 0$, $k \in \mathcal{I}^\epsilon$ and $\epsilon \in \{-1, +1\}$ the set $S_k^\epsilon(r)$ gathering the indices of the elements of $\mathcal{Q}^{\star,\epsilon}$ that are close to the element θ_k^0 of $\mathcal{Q}^{0,\epsilon}$:

$$S_k^\epsilon(r) = \left\{\ell \in \{1, \cdots, s\} : \theta_\ell^\star \in \mathcal{B}_T(\theta_k^0, r) \text{ and } \epsilon\beta_\ell^\star > 0\right\}. \tag{67}$$

Notice that the sets $S_k^\epsilon(r)$ can be empty. We assume that $r < \min_{\ell \neq k} \mathfrak{d}_T(\theta_\ell^0, \theta_k^0)/2$ so that the sets $S_k^\epsilon(r)$ with $\epsilon = \pm 1$ and $k \in \mathcal{I}^\epsilon$ are pairwise disjoint. We also set:

$$S(r) = \bigcup_{\epsilon \in \{-1,+1\}} S^\epsilon(r) \quad \text{with} \quad S^\epsilon(r) = \bigcup_{k \ \in \mathcal{I}^\epsilon} S_k^\epsilon(r).$$

We now define a discrepancy measure between the model and any approximation by a linear combination of features having their parameters in $\mathcal{Q}^0$:

$$\mathcal{D}_{T,r}(\beta^\star, \vartheta^\star, v^0, \vartheta^0) = \sum_{\epsilon \in \{-1,+1\}} \sum_{k \in \mathcal{I}^\epsilon} \sum_{\ell \in S_k^\epsilon(r)} |\beta_\ell^\star| \, \mathfrak{d}_T(\theta_\ell^\star, \theta_k^0)^2 + \sum_{k \in S(r)^c} |\beta_k^\star| \quad \text{for } r > 0,$$

where $S(r)^c$ denotes the complementary set of $S(r)$ in $\{1, \ldots, s\}$. Notice that $\mathcal{D}_{T,r}(\beta^\star, \vartheta^\star, v^0, \vartheta^0) = 0$ if and only if $\mathcal{Q}^{\star,+} \subseteq \mathcal{Q}^{0,+}$ and $\mathcal{Q}^{\star,-} \subseteq \mathcal{Q}^{0,-}$.

4.2. The testing hypotheses.

We shall test the following hypotheses:

$$\begin{cases} H_0 : & (\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star), \qquad \mathcal{Q}^{\star,+} \subseteq \mathcal{Q}^{0,+} \text{ and } \mathcal{Q}^{\star,-} \subseteq \mathcal{Q}^{0,-}, \\ H_1(\rho) : & (\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star) \quad \text{and} \quad \mathcal{D}_{T,r}(\beta^\star, \vartheta^\star, v^0, \vartheta^0) \geq \rho, \end{cases} \tag{68}$$

where ρ and $\delta^\star$ are separation parameters depending *a priori* on T, s and s^0 that need to be evaluated. Notice that the null hypothesis is also composite. We recall the definitions (35) and (36) of the minimax testing risk $R_\rho^\star$ and the minimax separation $\rho^\star$. In the following, we give upper bounds on the testing risk and on the minimax separation $\rho^\star(\alpha)$ for any $\alpha \in (0, 1)$.

4.3. **Main result.** In this section, we build a test for (68). Under Assumptions 2.1 and 2.2, we define the element of $L^2(\lambda_T)$:

$$p_0 = \sum_{k=1}^{s^0} \alpha_k \phi_T(\theta_k^0) + \sum_{k=1}^{s^0} \xi_k \, \tilde{D}_{1,T}[\phi_T](\theta_k^0), \tag{69}$$

where $\alpha, \xi \in \mathbb{R}^{s^0}$ solve the system:

$$\left\langle \phi_T(\theta_k^0), p_0 \right\rangle_{L^2(\lambda_T)} = v_k^0 \quad \text{and} \quad \left\langle \partial_\theta \phi_T(\theta_k^0), p_0 \right\rangle_{L^2(\lambda_T)} = 0, \text{ for all } k \in \{1, \cdots, s^0\}. \tag{70}$$

Remark 4.1. The element p_0 of $L^2(\lambda_T)$ coincides with the vanishing derivative pre-certificate which appears in [10, Section 4] and is the solution of (70) with minimal norm $\|p_0\|_{L^2(\lambda_T)}$.

Following [5], we give the existence and properties of the interpolating certificate function.

Lemma 4.2 (Interpolating certificate)**.** *Let $T \in \mathbb{N}$, let $s \in \mathbb{N}^*$, $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$, $\eta \in (0,1)$ and $\mathcal{Q} = \{\theta_1, \cdots, \theta_s\} \subset \Theta_T$. Suppose that Assumption 2.4 holds.*

Then, there exist finite positive constants C_N, C_F, C_B with $C_F < 1$, depending on r and the function F, such that for any application $v : \mathcal{Q} \mapsto \{-1, 1\}$, there exist unique $\alpha, \xi \in \mathbb{R}^s$ such that $p \in L^2(\lambda_T)$ uniquely defined by:

$$\begin{cases} p = \sum_{k=1}^{s} \alpha_k \phi_T(\theta_k) + \sum_{k=1}^{s} \xi_k \, \tilde{D}_{1,T}[\phi_T](\theta_k), \\ \langle \phi_T(\theta), p \rangle_{L^2(\lambda_T)} = v(\theta) \quad \text{and} \quad \langle \partial_\theta \phi_T(\theta), p \rangle_{L^2(\lambda_T)} = 0, \quad \text{for all} \quad \theta \in \mathcal{Q}, \end{cases} \tag{71}$$

satisfies:

(i) For all $\theta \in \mathcal{Q}$ and $\theta' \in \mathcal{B}_T(\theta, r)$, we have $|\langle \phi_T(\theta'), p \rangle_{L^2(\lambda_T)}| \leq 1 - C_N\, \mathfrak{d}_T(\theta, \theta')^2$.
(ii) For all θ in Θ_T, $\theta \notin \bigcup_{\theta' \in \mathcal{Q}} \mathcal{B}_T(\theta', r)$ (far region), we have $|\langle \phi_T(\theta), p \rangle_{L^2(\lambda_T)}| \leq 1 - C_F$.
(iii) We have $\|p\|_{L^2(\lambda_T)} \leq \sqrt{s}\, C_B$.

Proof. Using similar arguments as those developed in the proof of Theorem 2.3, we get that all the hypotheses of [5, Proposition, 7.4] are satisfied. The existence and uniqueness of p is then guaranteed by [5, Lemma, 10.1]. The properties satisfied by p are direct consequences of [5, Proposition, 7.4]. □

Using the estimator $\hat{\beta}$ from (28) for a given value of the tuning parameter κ, we define the test statistic:

$$\mathcal{T}_3 = \left\| \hat{\beta} \right\|_{\ell_1} - \langle y, p_0 \rangle_{L^2(\lambda_T)} \, . \tag{72}$$

and the corresponding test $\Psi_{\mathcal{T}_3}(t) = \mathbf{1}_{\{|\mathcal{T}_3| > t\}}$.

Theorem 4.3. *Let $T \in \mathbb{N}$, $s^0 \in \mathbb{N}^*$ and choose $K \in \mathbb{N}$ such that $s_0 \leq K$. Let also $\eta \in (0,1)$ and $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$. Let $\delta^\star \geq \sigma_T\, \Sigma(\eta, r, s)$ and $\delta^0 \geq \sigma_T\, \Sigma(\eta, r, s^0)$. Assume we observe the random element y of $L^2(\lambda_T)$ under the regression model (3) with unknown parameters $s \in \mathbb{N}^*$ such that $s \leq K$, $\beta^\star \in (\mathbb{R}^*)^s$ and $\vartheta^\star = (\theta_1^\star, \cdots, \theta_s^\star) \in \Theta_T^s(\delta^\star)$. Let $v^0 \in \{-1, 1\}^{s^0}$ be a sign vector and let $\vartheta^0 = (\theta_1^0, \cdots, \theta_{s^0}^0) \in \Theta_T^{s^0}(\delta^0)$. Suppose that Assumption 2.4 holds for the sets $\mathcal{Q}^\star = \{\theta_1^\star, \cdots, \theta_s^\star\} \subset \Theta_T$ of cardinal s and $\mathcal{Q}^0 = \{\theta_1^0, \cdots, \theta_{s^0}^0\} \subset \Theta_T$ of cardinal s^0. Suppose also that the noise process w_T satisfies Assumption 1.1 for a noise level $\overline{\sigma} > 0$ and a decay rate for the noise variance $\Delta_T > 0$.*

Then, the test statistic $\mathcal{T}_3$ is uniquely defined and there exist finite positive constants, a and $\mathcal{C}_i$ with $i = 1, \cdots, 5$, (depending on r and on the function F) such that for any $\tau > 1$ and any tuning parameter κ:

$$\kappa \geq \mathcal{C}_1 \overline{\sigma} \sqrt{\Delta_T \log(\tau)}, \tag{73}$$

the test $\Psi_{\mathcal{T}_3}$ satisfies:

$$R_\rho\left(\Psi_{\mathcal{T}_3}(t)\right) \le 2\,\mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\,\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right) + \frac{2}{\tau^{a\,s_0}}, \tag{74}$$

for any threshold $t > 0$ and any separation $\rho > 0$ satisfying:

$$t \ge 2\,\mathcal{C}_3\,s^0\,\kappa \quad \text{ and } \quad \rho \ge \mathcal{C}_4\,s\,\kappa + \mathcal{C}_5\,t. \tag{75}$$

Proof. Recall the test problem given by (68). Assumption 2.4 holds for the set $\mathcal{Q}^0$. Thanks to Lemma 4.2, the element p_0 of $L^2(\lambda_T)$ is uniquely defined by v^0, (69) and (70). Hence, the test statistic $\mathcal{T}_3$ from (72) is well-defined.

We first bound the type I error probability. Let us fix $(\beta^\star, \vartheta^\star) \in (\mathbb{R}^*)^s \times \Theta_T^s(\delta^\star)$ such that H_0 holds. Using that $y = \beta^\star\Phi_T(\vartheta^\star) + w_T$ and the triangle inequality, we obtain:

$$\begin{aligned} |\mathcal{T}_3| &= \left| \left\|\hat\beta\right\|_{\ell_1} - \|\beta^\star\|_{\ell_1} + \|\beta^\star\|_{\ell_1} - \langle \beta^\star\Phi_T(\vartheta^\star), p_0\rangle_{L^2(\lambda_T)} - \langle w_T, p_0\rangle_{L^2(\lambda_T)}\right| \\ &\le \left| \left\|\hat\beta\right\|_{\ell_1} - \|\beta^\star\|_{\ell_1}\right| + |B| + \left|\langle w_T, p_0\rangle_{L^2(\lambda_T)}\right|, \end{aligned} \tag{76}$$

where:

$$B = \|\beta^\star\|_{\ell_1} - \langle \beta^\star\Phi_T(\vartheta^\star), p_0\rangle_{L^2(\lambda_T)}. \tag{77}$$

Since $\mathcal{Q}^{\star,+} \subseteq \mathcal{Q}^{0,+}$, $\mathcal{Q}^{\star,-} \subseteq \mathcal{Q}^{0,-}$, we have for all $k \in \{1, \cdots, s\}$:

$$|\beta_k^\star| - \langle \beta_k^\star \phi_T(\theta_k^\star), p_0\rangle_{L^2(\lambda_T)} = 0,$$

we deduce that $B = 0$ under H_0. Hence, we have that:

$$\mathbb{E}_{(\beta^\star,\,\vartheta^\star)}[\Psi_{\mathcal{T}_3}(t)] \le \mathbb{P}\left(\left| \left\|\hat\beta\right\|_{\ell_1} - \|\beta^\star\|_{\ell_1}\right| > t/2\right) + \mathbb{P}\left(\left|\langle w_T, p_0\rangle_{L^2(\lambda_T)}\right| > t/2\right). \tag{78}$$

Recall that under H_0, we have $s \le s^0$. Therefore, since $\mathcal{C}_3\,\kappa\,s^0 \le t/2$, we have $\mathcal{C}_3\,\kappa\,s \le t/2$. We get from Theorem 2.3 that:

$$\mathbb{P}\left(\left| \left\|\hat\beta\right\|_{\ell_1} - \|\beta^\star\|_{\ell_1}\right| > t/2\right) \le \mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right). \tag{79}$$

Then, thanks to Assumptions 1.1 and Lemma 4.2, the quantity $\langle w_T, p_0\rangle_{L^2(\lambda_T)}$ is a centered Gaussian random variable of variance bounded by $\overline{\sigma}^2 C_B^2 \Delta_T s_0$ where C_B is the finite positive constant from Lemma 4.2. Hence we have, provided that $t \ge 2\mathcal{C}_3\,\kappa\,s^0$ with $\kappa \ge \mathcal{C}_1\overline{\sigma}\sqrt{\Delta_T\log(\tau)}$, that is, $t^2 \ge (2\mathcal{C}_1\mathcal{C}_3\overline{\sigma}s_0)^2\Delta_T\log(\tau)$:

$$\mathbb{P}\left(\langle w_T, p_0\rangle_{L^2(\lambda_T)} > t/2\right) \le \int_{t/2}^{+\infty} \frac{\mathrm{e}^{-x^2/(2\overline{\sigma}^2\Delta_T C_B^2 s_0)}}{\sqrt{2\pi\overline{\sigma}^2\Delta_T C_B^2 s_0}}\,\mathrm{d}x \le \frac{1}{2}\,\mathrm{e}^{-\frac{t^2}{8(\overline{\sigma}^2\Delta_T C_B^2 s_0)}} \le \frac{1}{2\tau^{a s_0}},$$

with $a = (\mathcal{C}_1\mathcal{C}_3/C_B)^2/2$ and where we used the tail bound (45). It gives by symmetry that:

$$\mathbb{P}\left(|\,\langle w_T, p_0\rangle_{L^2(\lambda_T)}\,| > t/2\right) \le \frac{1}{\tau^{a\,s_0}}. \tag{80}$$

Plugging (79) and (80) in (78), we get:

$$\sup_{(\beta^\star,\,\vartheta^\star)\in H_0} \mathbb{E}_{(\beta^\star,\,\vartheta^\star)}\left[\Psi_{\mathcal{T}_3}(t)\right] \le \mathcal{C}_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\,\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right) + \frac{1}{\tau^{a\,s_0}}. \tag{81}$$

We now bound the type II error probability. Assume that H_1 holds, that is $\mathcal{D}_{T,r}(\beta^\star, \vartheta^\star, v^0, \vartheta^0) \geq \rho$. We have, using the first equality of (76) and the triangle inequality, that:

$$|\mathcal{T}_3| \geq |B| - \left|\langle w_T, p_0\rangle_{L^2(\lambda_T)}\right| - \left| \left\|\hat{\beta}\right\|_{\ell_1} - \|\beta^\star\|_{\ell_1} \right|,$$

with B defined in (77). Using the definitions (67) of $S(r)$ and $S_k^\epsilon(r)$ with $\epsilon \in \{-1,+1\}$ and $k \in \mathcal{I}^\epsilon$, we get:

$$B = \sum_{\substack{\epsilon\in\{-1,+1\}\\ k\in\mathcal{I}^\epsilon,\, \ell\in S_k^\epsilon(r)}} |\beta_\ell^\star| \left(1 - \operatorname{sign}(\beta_\ell^\star)\ \langle \phi_T(\theta_\ell^\star), p_0\rangle_{L^2(\lambda_T)}\right) + \sum_{k\in S(r)^c} |\beta_k^\star| \left(1 - \operatorname{sign}(\beta_k^\star)\ \langle \phi_T(\theta_k^\star), p_0\rangle_{L^2(\lambda_T)}\right).$$

Thanks to Lemma 4.2 (i)-(ii) of , we obtain:

$$\begin{aligned} B &\geq \sum_{\substack{\epsilon\in\{-1,+1\}\\ k\in\mathcal{I}^\epsilon,\, \ell\in S_k^\epsilon(r)}} C_N |\beta_\ell^\star| \mathfrak{d}_T(\theta_\ell^\star, \theta_k^0)^2 + \sum_{k\in S(r)^c} C_F |\beta_k^\star| \\ &\geq (C_N \wedge C_F) \mathcal{D}_{T,r}(\beta^\star, \vartheta^\star, v^0, \vartheta^0) \geq (C_N \wedge C_F)\rho, \end{aligned}$$

where the constants C_N and C_F are defined in Lemma 4.2 and depend on r and on the function F. Therefore, we have with $a_t = (C_N \wedge C_F)\rho - t$:

$$\begin{aligned} \mathbb{E}_{(\beta^\star,\vartheta^\star)}\left[1 - \Psi_{\mathcal{T}_3}(t)\right] &\leq \mathbb{P}\left(\left|\langle w_T, p_0\rangle_{L^2(\lambda_T)}\right| + \left|\|\beta^\star\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right| \geq a_t\right) \\ &\leq \mathbb{P}\left(\left|\langle w_T, p_0\rangle_{L^2(\lambda_T)}\right| \geq a_t/2\right) + \mathbb{P}\left(\left|\|\beta^\star\|_{\ell_1} - \left\|\hat{\beta}\right\|_{\ell_1}\right| \geq a_t/2\right). \end{aligned}$$

Provided that $\rho \geq \mathcal{C}_4\, s\, \kappa + \mathcal{C}_5\, t$ with $\mathcal{C}_4 = 2\,\mathcal{C}_3/(C_N \wedge C_F)$ and $\mathcal{C}_5 = 2/(C_N \wedge C_F)$ we have $a_t/2 \geq (\mathcal{C}_3 \kappa s) \vee (t/2)$. By using (79) and (80), we obtain:

$$\sup_{(\beta^\star,\,\vartheta^\star)\in H_1(\rho)} \mathbb{E}_{(\beta^\star,\,\vartheta^\star)}\left[1 - \Psi_{\mathcal{T}_3}(t)\right] \leq \mathcal{C}_2 \left(\frac{|\Theta_T|}{\sigma_T\, \tau\, \sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right) + \frac{1}{\tau^{a s_0}}. \tag{82}$$

Finally, by adding both sides of (81) and (82), we get (74). This concludes the proof. □

4.4. **Separation rates.** We give in this section an upper bound on the minimax separation $\rho^\star$ to test the goodness-of-fit of the dictionary, that is to distinguish the assumptions H_0 and $H_1(\rho)$ presented in Section 4.

Corollary 4.4. *Under the framework and the assumptions of Theorem 4.3, there exist finite positive constants c and C (depending on r and the function F) such that provided that $|\Theta_T|/\sigma_T \geq 1$, we have for any $\alpha \in (0,1)$:*

$$\rho^\star(\alpha) \leq C\, \overline{\sigma}\, (s \vee s^0) \sqrt{\Delta_T \log\left(\frac{c\,|\Theta_T|}{\alpha\, \sigma_T}\right)}. \tag{83}$$

Proof. The result is a direct consequence of Theorem 4.3. We fix the tuning parameter $\kappa = \mathcal{C}_1 \overline{\sigma} \sqrt{\Delta_T \log(\tau)}$ by taking the equality in (73). Then, for $\rho \geq \mathcal{C}_4\, s\, \kappa + \mathcal{C}_5\, t$ and $t = 2\,\mathcal{C}_3\, s^0 \kappa$ we have by Theorem 4.3 for $\tau > 1$ and since $s_0 \geq 1$:

$$R_\rho^\star \leq R_\rho\left(\Psi_{\mathcal{T}_3}(t)\right) \leq 2\mathcal{C}_2 \left(\frac{|\Theta_T|}{\sigma_T\, \tau\, \sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right) + \frac{2}{\tau^a},$$

where the finite positive constants a, $\mathcal{C}_i$ with $i \in \{1, \cdots, 5\}$, from Theorem 4.3 depend on r and the function F.

Hence, by taking $\tau = c'/(\sigma_T\alpha/(2|\Theta_T|))^{c''}$ with $c'' = 1 \vee (1/a)$ and $c' = (2\mathcal{C}_2) \vee \mathrm{e} \vee 2^{1/a}$, we get for $\rho \geq 2\mathcal{C}_1((2\,\mathcal{C}_3\,\mathcal{C}_5) \vee \mathcal{C}_4)\overline{\sigma}(s \vee s^0)\sqrt{\Delta_T \log(c'/(\sigma_T\alpha/(2|\Theta_T|))^{c''})}$ and $\alpha \in (0,1)$ that $R^\star_\rho \leq \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$. We readily deduce (83) with $c = 2c'^{(1/c'')}$. □

5. Gaussian scaled spikes deconvolution

In this section, we consider the discrete time process observed on a regular grid given in Section 1.2.1. We recall that Assumption 1.1 holds with:

$$\lambda_T = \Delta_T \sum_{j=1}^{T} \delta_{t_j} \quad \text{with} \quad t_j = -b_T + j\Delta_T \quad \text{and} \quad \Delta_T = \frac{2b_T}{T},$$

and w_T given by (5), where $T \in \mathbb{N}^*$. We consider the scaled Gaussian features associated to the function:

$$h(t,\sigma) \mapsto \frac{\exp(-t^2/2\sigma^2)}{\pi^{1/4}\sigma^{1/2}} \quad \text{defined on} \quad \Theta \times \mathfrak{S} = \mathbb{R} \times \mathbb{R}^*_+.$$

We shall see below that the natural choice for the function F appearing in (10) is given by:

$$F = h^0 * h^0 = h^0(\cdot/\sqrt{2}) \quad \text{with} \quad h^0(\cdot) = h(\cdot, 1).$$

In the following, we check that Assumption 2.4 holds. Then, using Theorem 2.3 on a particular example, we provide a prediction bound for the estimator of $(\beta^\star, \vartheta^\star)$ solution of the optimization problem (28).

5.1. Choice of the approximating kernel.

We denote the unscaled feature φ^0 on $\theta \in \Theta$ by:

$$\varphi^0(\theta) = h(\theta - \cdot, 1) = h^0(\theta - \cdot).$$

We define the mapping $f_T : \Theta \to \Theta$ by $f_T(\theta) = \theta/\sigma_T$ for any $\theta \in \Theta$ and the (pushforward) measure $\lambda_T^0 = \lambda_T \circ f_T^{-1}$ so that for any $g \in L^1(\lambda_T^0)$:

$$\int g(\theta/\sigma_T)\lambda_T(\mathrm{d}\theta) = \int g(\theta)\, \lambda_T^0(\mathrm{d}\theta).$$

The Hilbert space $L^2(\lambda_T^0)$ is endowed with its natural scalar product $\langle\cdot,\cdot\rangle_{L^2(\lambda_T^0)}$ and norm $\|\cdot\|_{L^2(\lambda_T^0)}$. We define on Θ^2 the kernel:

$$\mathcal{K}_T^0(\theta,\theta') = \langle \phi_T^0(\theta), \phi_T^0(\theta')\rangle_{L^2(\lambda_T^0)} \quad \text{with} \quad \phi_T^0(\theta) = \varphi^0(\theta)/\left\|\varphi^0(\theta)\right\|_{L^2(\lambda_T^0)}.$$

The kernel $\mathcal{K}_T$ can be seen as a scaled kernel derived from $\mathcal{K}_T^0$ as for any $\theta, \theta' \in \Theta$:

$$\mathcal{K}_T(\theta,\theta') = \mathcal{K}_T^0(\theta/\sigma_T, \theta'/\sigma_T).$$

When the measure λ_T^0 converges in some sense, as T goes to infinity, towards the Lebesgue measure Leb on $\mathbb{R}$, it is natural to consider the approximation $\mathcal{K}_\infty^0$ of $\mathcal{K}_T^0$ on Θ^2 by:

$$\mathcal{K}_\infty^0(\theta,\theta') = \left\langle \phi_\infty^0(\theta), \phi_\infty^0(\theta')\right\rangle_{L^2(\mathrm{Leb})} \quad \text{with} \quad \phi_\infty^0(\theta) = \varphi^0(\theta)/\left\|\varphi^0(\theta)\right\|_{L^2(\mathrm{Leb})}.$$

Thanks to the definition of F, we also have on Θ^2 that:

$$F(\theta - \theta') = \mathcal{K}_\infty^0(\theta, \theta').$$

The approximating kernel $\mathcal{K}_T^{\mathrm{prox}}$ is then given by (10) on Θ^2.

5.2. Checking Assumption 2.4.

5.2.1. *Regularity of the dictionary.* We refer to [5, Section 8] to check that Assumption 2.4 *(i)* holds for the feature φ_T defined by (1) and any scale parameter $\sigma_T \in \mathfrak{S} = \mathbb{R}^*_+$.

5.2.2. *Boundedness and local concavity on the diagonal.* Elementary calculations show that $g_\infty = -F''(0) = 1/2$. By definition of F, we directly deduce that Assumption 2.3 holds. We also get that for $r \in (0, \sqrt{2})$:

$$\varepsilon(r) = 1 - \mathrm{e}^{-r^2/4} > 0 \quad \text{and} \quad \nu(r) = \left(1 - \frac{r^2}{2}\right) \mathrm{e}^{-r^2/4}.$$

We fix $r \in (0, 1/2)$. We readily check that Assumption 2.4 *(ii)* is verified.

5.2.3. *Proximity to the approximating kernel.* In order for the kernel $\mathcal{K}_T^{\mathrm{prox}}$ to be a good approximation of $\mathcal{K}_T$ in the sense of Assumption 2.4 *(iii)*, we shall consider the set Θ_T over which the optimization is performed:

$$\Theta_T = [(1-\xi)a_T, (1-\xi)b_T] \subset [a_T, b_T] \quad \text{with a given shrinkage parameter } \xi \in (0,1).$$

Intuitively, one does not expect the estimation of the location parameter to perform well near the lower and upper bounds of the observation grid (given by the support of λ_T). Following [5, Section 8], we set:

$$\gamma_T = 2\Delta_T\, \sigma_T^{-1} + \sqrt{\pi}\, \mathrm{e}^{-\xi^2 b_T^2/2\sigma_T^2}. \tag{84}$$

Recall $\mathcal{V}_T$ and C_T defined by (20) and (22). Using Lemma [5, Lemma 8.1], there exist finite positive universal constants c_0, c_1 and c_2, such that $\gamma_T < c_0$ implies:

$$\mathcal{V}_T \leq c_1 \gamma_T \quad \text{and} \quad |1 - C_T| \leq c_2 \gamma_T. \tag{85}$$

Assume that $(b_T, T \geq 2)$ and $(\sigma_T, T \geq 2)$ are sequences of positive numbers, such that:

$$\lim_{T\to\infty} b_T = +\infty, \quad \lim_{T\to\infty} \sigma_T = 0 \quad \text{and} \quad \lim_{T\to\infty} \Delta_T\, \sigma_T^{-1} = 0. \tag{86}$$

Therefore, we have $\lim_{T\to+\infty} \mathcal{V}_T = 0$ and $\lim_{T\to+\infty} C_T = 1$.

Let $\eta \in (0,1)$ be fixed. We deduce that under (86), Assumption 2.4 *(iii)* is satified provided that T is larger than some constant depending on η, r, the sparsity s and the sequences $(b_T, T \geq 2)$ and $(\sigma_T, T \geq 2)$.

5.2.4. *Separation of the non-linear parameters.* We remark that $\lim_{r''\to\infty} \sup_{|r'|\geq r''} |F^{(i)}(r')| = 0$ for all $i \in \{0, \ldots, 3\}$. Thus, we deduce from the definition (25) of δ that $\delta(u, s)$ is finite for all $s \in \mathbb{N}^*$ and $u > 0$. Let us stress that $\sup_{s\in\mathbb{N}^*} \delta(u, s) \leq M/u$ for some universal finite constant M, see [5, Remark 8.2]. Therefore, the quantity $\Sigma(\eta, r, s)$ is bounded by a constant depending only on η and r.

So Assumption 2.4 *(iv)* is verified as soon as $|\theta - \theta'| > \sigma_T\, \Sigma(\eta, r, s)$ for all for all $\theta \neq \theta' \in \mathcal{Q}^\star$. (Notice this happens for the scaling parameter σ_T small enough depending on $\mathcal{Q}^\star$.)

5.3. **Prediction error bound in a particular case.** Recall the shrinkage parameter $\xi \in (0,1)$ in (84). Let us assume that:

$$b_T = \log(T) \quad \text{and} \quad \sigma_T = 1/\sqrt{\xi \log(T)}.$$

In particular, condition (86) holds. In this case, there exists a finite positive constant c depending on r, η and ξ such that for $T \geq c\sqrt{\log(T)}\, s$, Assumption 2.4 holds (notice that the separation condition (27) of the location parameters in $\mathcal{Q}^\star$ is also verified for T large enough, depending on $\mathcal{Q}^\star$, as $\lim_{T\to+\infty} \sigma_T = 0$). By Theorem 2.3 with $\tau = T$ and κ given by the equality in (29), we get that:

$$\frac{1}{\sqrt{T}} \left\| \hat{\beta} \Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_{\ell_2} \leq \mathcal{C}_0\, \mathcal{C}_1\, \overline{\sigma} \sqrt{\frac{s \log(T)}{T}},$$

with probability larger than $1 - \mathcal{C}_2 \left(\frac{2\sqrt{\xi}\log(T)}{T} \vee \frac{1}{T} \right)$, where the constants $\mathcal{C}_0$, $\mathcal{C}_1$ and $\mathcal{C}_2$ do not depend on T.

6. Low-pass filter

In this section, we consider the continuous-time process described in Section 1.2.2 on the torus $\Theta = \mathbb{R}/\mathbb{Z}$ with $\lambda_T = \text{Leb}$ the Lebesgue measure on Θ and the noise:

$$w_T = \sum_{k\in\mathbb{N}} \sqrt{\xi_k}\, G_k\, \psi_k,$$

where $(G_k, k \in \mathbb{N})$ are independent centered Gaussian random variables with variance $\overline{\sigma}^2$, $(\psi_k, k \in \mathbb{N})$ is an o.n.b. of $L^2(\text{Leb})$ on Θ and $\xi = (\xi_k, k \in \mathbb{N})$ is a square summable sequence of non-negative real numbers depending on $T \in 2\mathbb{N}^* + 1$. Recall that the noise satisfies Assumption 1.1 for a positive noise level $\overline{\sigma}$ and a decay on the noise variance $\Delta_T = \sup_{k\in\mathbb{N}} \xi_k$.

We consider the normalized Dirichlet kernel, see (8), on Θ:

$$h(t, \sigma) = \frac{\sin(T\pi t)}{\sqrt{T}\, \sin(\pi t)} \quad \text{defined for} \quad t \in \Theta = \mathbb{R}/\mathbb{Z} \quad \text{and} \quad \sigma = \frac{1}{T}, \quad T \in 2\mathbb{N}^* + 1. \tag{87}$$

The parameter T is related to the so-called cut-off frequency $f_c \in \mathbb{N}^*$ by $T = 2f_c + 1$. We shall see below that the natural choice for the function F appearing in (10) is given by:

$$F(t) = \frac{\sin(\pi t)}{\pi t} \quad \text{for} \quad t \in \mathbb{R}.$$

We get from the definition (15) that $g_\infty = -F''(0) = \pi^2/3$.

In the following, we check that Assumption 2.4 hold. Then, using Theorem 2.3, we provide a prediction bound for the estimator of $(\beta^\star, \vartheta^\star)$ solution of the optimization problem (28).

6.1. **The approximating kernel.** We define the features φ_T using (1) with $\sigma_T = 1/T$. Elementary calculations give that for $\theta, \theta' \in \Theta$:

$$\mathcal{K}_T(\theta, \theta') = \frac{\sin(T\pi(\theta - \theta'))}{T\, \sin(\pi(\theta - \theta'))}.$$

Recall that by convention $|\theta - \theta'|$ is the Euclidean distance between θ and θ' in Θ, and in particular it belongs to $[0, 1/2]$. We define the approximating kernel $\mathcal{K}_T^{\text{prox}}$ on Θ by:

$$\mathcal{K}_T^{\text{prox}}(\theta, \theta') = F(T|\theta - \theta'|) \quad \text{with} \quad |\theta - \theta'| \in [0, 1/2].$$

Since F is even, we get also that $F(T|\theta - \theta'|) = F(T(\theta - \theta'))$ where, for $\theta, \theta' \in \Theta$, their representers in $\mathbb{R}$ a chosen so that $\theta - \theta'$ belongs to $[-1/2, 1/2]$.

6.2. **Checking Assumption 2.4.**

6.2.1. *Regularity of the dictionary.* It is elementary to check that g_T is a constant function on Θ equal to $g_\infty\,(T^2 - 1)$ and that Assumption 2.4 *(i)* on the regularity of the dictionary holds.

6.2.2. *Boundedness and local concavity on the diagonal.* There exists $R > 0$ such that for any $r \in (0, R)$:

$$\varepsilon(r) = 1 - \frac{\sin(\pi r)}{\pi r} > 0 \quad \text{and} \quad \nu(r) = -\left(\frac{6}{\pi^3 r^3} - \frac{3}{\pi\, r}\right) \sin(\pi r) + \frac{6\cos(\pi r)}{\pi^2 r^2} > 0.$$

We fix $r \in (0, (1/\sqrt{2g_\infty L_2}) \wedge (R/2))$. This and the fact that F is $\mathcal{C}^\infty$ with bounded derivatives implies that Assumption 2.4 *(ii)* on the boundedness and the local concavity of the approximating kernel holds.

6.2.3. *Proximity to the approximating kernel.* We set $\Theta_T = \Theta$. The proof of the next lemma on the uniform approximation of $\mathcal{K}_T$ by $\mathcal{K}_T^{\text{prox}}$ on the torus is postponed to Section 6.3.1.

Lemma 6.1. *There exists a universal positive finite constant c_3 such that for any $T \in 2\mathbb{N}^* + 1$:*

$$\mathcal{V}_T \leq \frac{c_3}{T} \quad \textit{and} \quad |1 - C_T| \leq \frac{1}{2(T^2-1)}. \tag{88}$$

Let $\eta \in (0, 1)$ be fixed. We deduce from (88) that Assumption 2.4 *(iii)* is satified provided that T is larger than some constant depending on η, r, the sparsity s.

6.2.4. *Separation of the non-linear parameters.* Notice that $\lim_{r''\to\infty} \sup_{|r'|\geq r''} |F^{(i)}(r')| = 0$ for all $i \in \{0, \cdots, 3\}$. Thus, we deduce from the definition (25) of δ that $\delta(u, s)$ is finite for all $s \in \mathbb{N}^*$ and $u > 0$.

So Assumption 2.4 *(iv)* is verified as soon as $|\theta - \theta'| > \sigma_T\, \Sigma(\eta, r, s)$ for all $\theta \neq \theta' \in \mathcal{Q}^\star$. (Notice this happens for T large enough depending on $\mathcal{Q}^\star$ as $\sigma_T = 1/T$.)

6.3. **Prediction error bound.** There exists a constant c depending on η and r such that for any $T \in 2\mathbb{N}^*+1$ such that $T \geq c\, s$, and provided that (27) is satisfied, Assumption 2.4 holds. Using Theorem 2.3 with κ given by an equality in (29) with $\tau > 1$, we obtain the prediction bound:

$$\left\| \hat{\beta}\Phi_T(\hat{\vartheta}) - \beta^\star \Phi_T(\vartheta^\star) \right\|_{L^2(\text{Leb})} \leq \mathcal{C}_0\, \mathcal{C}_1\, \overline{\sigma}\, \sqrt{s\, \Delta_T\, \log(\tau)},$$

with probability larger than $1 - \mathcal{C}_2 \left(\frac{T}{\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau} \right)$, where the constants $\mathcal{C}_0$, $\mathcal{C}_1$ and $\mathcal{C}_2$ do not depend on T.

Remark 6.2. Exact support recovery results were obtained in [10]. The authors considered a small noise regime , that is $\|w_T\|_{L^2(\text{Leb})}/\kappa$ less than a constant). They assumed that the location parameters satisfy for any $k, \ell \in \{1, \cdots, s\}$ such that $k \neq \ell$, the separation condition $|\theta_k^\star - \theta_\ell^\star| \geq C/f_c$ for $T = 2f_c + 1$, for some positive constant C and with $f_c \geq s$ (s being the number of active features in the mixture). They showed that there exist finite constants C' and C'' such that for all $k \in \{1, \cdots, s\}$:

$$|\tilde{\theta}_k - \theta_k^\star| \leq C' \|w_T\|_{L^2(\text{Leb})} \quad \text{and} \quad |\tilde{\beta}_k - \beta_k^\star| \leq C'' \|w_T\|_{L^2(\text{Leb})},$$

for some estimators $(\tilde{\beta}, \tilde{\vartheta} = (\tilde{\theta}_1, \cdots, \tilde{\theta}_s))$ obtained by solving the BLasso problem.

However the small noise regime assumption is restrictive as it does not encompass the example of Section 1.2.2 where for all $k \in \mathbb{N}$, $\xi_k = T^{-1}\mathbf{1}_{\{1\leq k\leq T\}}$ and thus $\Delta_T = 1/T$ and $\mathbb{E}[\|w_T\|_{L^2(\text{Leb})}]$ is of order 1. Recall that in (31) we obtain that our estimators satisfy:

$$\Big| \|\, \hat{\beta}\|_{\ell_1} - \|\, \beta^\star\|_{\ell_1} \Big| \leq C \frac{s\,\sqrt{\log(T)}}{\sqrt{T}}$$

for some constant $C > 0$ with high probability. Thus our prediction and estimation rates are smaller by a factor $\sqrt{\log(T)}/\sqrt{T}$ due to the probabilistic bounds on linear functionals of the noise process that we used in the proof, and this holds under analogous separation condition on any $\theta_k^\star$ and $\theta_\ell^\star$, for $k \neq \ell$ in $\{1, ..., s\}$.

6.3.1. *Proof of Lemma 6.1.* It is easy to check that the functions g_T and $g_{\mathcal{K}_T^{\text{prox}}}$ are constant functions with:

$$g_T = g_\infty\, (T^2 - 1) \quad \text{and} \quad g_{\mathcal{K}_T^{\text{prox}}} = g_\infty\, T^2. \tag{89}$$

Thus, we easily deduce the second inequality of (88) from the definition (20) of C_T.

We now consider the bound on $\mathcal{V}_T$. For $i, j \in \{0, \cdots, 3\}$ and $\ell = i + j$, we have with $\alpha_T = 1 - 1/T^2$:

$$\sup_{\Theta^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}_T^{\text{prox}[i,j]}| = g_\infty^{-\ell/2} (T^2 \alpha_T)^{-\ell/2} \sup_{t\in[-\frac{1}{2},\frac{1}{2}]} \left| \partial_t^\ell \left[D_T(t) + \left(1 - \alpha_T^{\ell/2}\right) \frac{\sin(T\pi t)}{T\pi t} \right] \right|, \tag{90}$$

where, for $t \in [-1/2, 1/2]$ and the convention $J(0) = 0$:

$$D_T(t) = \frac{\sin(T\pi t)}{T} J(t) \quad \text{and} \quad J(t) = \frac{1}{\sin(\pi t)} - \frac{1}{\pi t}.$$

It is easy to check that the function J can be expanded as a power series at 0 with positive convergence radius, and thus is of class $\mathcal{C}^\infty$ on $[-1/2, 1/2]$. Thus the following constant is finite:

$$M = \sup_{0 \le \ell \le 6} \sup_{[-1/2,1/2]} |J^{(\ell)}| < +\infty.$$

Using the Leibniz rule, we have that for $\ell \in \{1, \cdots, 6\}$ and $t \in [-1/2, 1/2]$:

$$|\partial_t^\ell D_T(t)| = \frac{1}{T} \left| \sum_{j=0}^{\ell} \binom{\ell}{j} (T\pi)^j \sin^{(j)}(T\pi t)\, J^{(\ell-j)}(t) \right| \le M \frac{(T\pi + 1)^\ell}{T}.$$

We deduce from (90) that for $i, j \in \{0, \cdots, 3\}$ and $\ell = i + j$:

$$\sup_{\Theta^2} |\mathcal{K}_T^{[i,j]} - \mathcal{K}_T^{\mathrm{prox}[i,j]}| \le g_\infty^{-\ell/2} (T^2 \alpha_T)^{-\ell/2} \left(M \frac{(T\pi + 1)^\ell}{T} + (1 - \alpha_T^{\ell/2}) \right) \le M 3^\ell T^{-1},$$

where we used that $T \ge 3$ and $g_\infty \alpha_T \ge 1$, and that $1 - \alpha_T^{\ell/2} = 0$ for $\ell = 0$. Recall the definition (22) of $\mathcal{V}_T$ to get $\mathcal{V}_T \le M 3^\ell T^{-1}$. This finishes the proof.

References

[1] Milton Abramowitz and Irene A. Stegun, editors. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, Inc., New York, 1992. Reprint of the 1972 edition. 13

[2] Ery Arias-Castro, Emmanuel J. Candès, and Yaniv Plan. Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.*, 39(5):2533–2556, 2011. 5

[3] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002. 4

[4] Claire Boyer, Yohann De Castro, and Joseph Salmon. Adapting to unknown noise level in sparse deconvolution. *Inf. Inference*, 6(3):310–348, 2017. 4

[5] C. Butucea, J-F. Delmas, A. Dutfoy, and C. Hardy. Off-the-grid learning of sparse mixtures from a continuous dictionary. *arXiv preprint arXiv:2207.00171*, 2022. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 16, 20, 23, 24, 28, 29

[6] Cristina Butucea, Jean-François Delmas, Anne Dutfoy, and Clément Hardy. Modeling infra-red spectra: an algorithm for an automatic and simultaneous analysis. In *In Proceedings of the 31st European Safety and Reliability Conference*, pages 3359–3366, 2021. 1, 4

[7] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *J. Fourier Anal. Appl.*, 19(6):1229–1254, 2013. 4, 6

[8] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Comm. Pure Appl. Math.*, 67(6):906–956, 2014. 4

[9] Yohann de Castro and Fabrice Gamboa. Exact reconstruction using Beurling minimal extrapolation. *J. Math. Anal. Appl.*, 395(1):336–354, 2012. 4, 10

[10] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Found. Comput. Math.*, 15(5):1315–1355, 2015. 1, 4, 6, 20, 26

[11] M. S. Ermakov. Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(4):704–715, 1990. 4

[12] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016. 4

[13] Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003. 4

[14] Yuri I. Ingster, Alexandre B. Tsybakov, and Nicolas Verzelen. Detection boundary in sparse regression. *Electron. J. Stat.*, 4:1476–1526, 2010. 5, 17, 18

[15] Béatrice Laurent, Jean-Michel Loubes, and Clément Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.*, 6:91–122, 2012. 4

[16] Stéphane Mallat. *A wavelet tour of signal processing : the sparse way*. Elsevier/Academic Press, Amsterdam, third edition, 2009. With contributions from Gabriel Peyré. 6

[17] Clarice Poon, Nicolas Keriven, and Gabriel Peyré. The geometry of off-the-grid compressed sensing. *Foundations of Computational Mathematics*, 2021. 4, 7, 8
[18] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Trans. Inform. Theory*, 61(1):499–512, 2015. 4

Appendix A. Proof of Theorem 2.3

This section is devoted to the proof of Theorem 2.3. Let $T \in \mathbb{N}$ and consider a positive scaling σ_T. In order to prove the theorem we shall apply [5, Theorem 2.1] replacing the limit kernel, noted $\mathcal{K}_\infty$ therein, by the approximating kernel $\mathcal{K}_T^{\text{prox}}$ defined on Θ^2 by:

$$\mathcal{K}_T^{\text{prox}}(\theta, \theta') = F(|\theta - \theta'|/\sigma_T). \tag{91}$$

We check that all the hypotheses of [5, Theorem 2.1] hold in our framework. Since Assumption 1.1 holds, the noise w_T is admissible and satisfies Point (i) of [5, Theorem 2.1]. Then, recall that Assumptions 2.1 and 2.2 are in force thanks to Assumption 2.4 (i). Therefore Point (ii) of [5, Theorem 2.1] on the regularity of the dictionary φ_T is verified. We shall check Point (iii) of [5, Theorem 2.1] on the regularity of the kernel with $\mathcal{K}_\infty$ replaced by $\mathcal{K}_T^{\text{prox}}$. Since Assumption 2.3 holds, we readily check that [5, Assumption 5.1] as $\mathcal{K}_T^{\text{prox}}(\theta, \theta) = F(0) = 1$ and $\mathcal{K}_T^{\text{prox}[2,0]}(\theta, \theta) = F''(0)/g_\infty = -1$. Thus, Points (iii) therein holds on Θ (with $\Theta_\infty = \Theta$). Point (iv) on the proximity between the kernels $\mathcal{K}_T$ and $\mathcal{K}_T^{\text{prox}}$ is verified since Assumption 2.4 (iii) holds and implies [5, Assumption 5.2].

It remains to show that Point (v) on the existence of certificate functions also holds. To do so, we shall apply [5, Propositions 7.4 and 7.5] that give sufficient conditions for Point (v) to hold. Let us first focus on the hypotheses of [5, Proposition 7.4]. We fix $r \in \left(0, 1/\sqrt{2\, g_\infty\, L_2}\right)$ (we stress that the quantities "r" and "ρ"from [5, Propositions 7.4 and 7.5] are respectively taken equal to $r\sqrt{g_\infty}$ and 2). It is straightforward to see that Point (i) of [5, Proposition 7.4] on the regularity of the dictionary is satisfied thanks to Assumption 2.2 and 2.1.

The Riemannian metric noted $\mathfrak{d}_\infty$ in [5] is given by, for any $\theta, \theta' \in \Theta$:

$$\mathfrak{d}_\infty(\theta, \theta') = |G_{\mathcal{K}_T^{\text{prox}}}(\theta) - G_{\mathcal{K}_T^{\text{prox}}}(\theta')| = \sqrt{g_{\mathcal{K}_T^{\text{prox}}}}|\theta - \theta'| = \sqrt{g_\infty}\sigma_T^{-1}|\theta - \theta'|, \tag{92}$$

where $G_{\mathcal{K}_T^{\text{prox}}}$ is a primitive of the function $\sqrt{g_{\mathcal{K}_T^{\text{prox}}}}$ defined by (11) and we used (18) for the second inequality. Following [5, (Eq.39-40)], we define the quantities for $r' > 0$,

$$\varepsilon_\infty(r') = 1 - \sup\left\{|\mathcal{K}_T^{\text{prox}}(\theta, \theta')|; \quad \theta, \theta' \in \Theta \text{ such that } \mathfrak{d}_\infty(\theta', \theta) \geq r'\right\},$$
$$\nu_\infty(r') = -\sup\left\{\mathcal{K}_T^{\text{prox}[0,2]}(\theta, \theta'); \quad \theta, \theta' \in \Theta \text{ such that } \mathfrak{d}_\infty(\theta', \theta) \leq r'\right\}.$$

We readily check that for any $r' > 0$, $\varepsilon(r'/\sqrt{g_\infty}) = \varepsilon_\infty(r')$ and $\nu(r'/\sqrt{g_\infty}) = \nu_\infty(r')$. Thus, $\varepsilon_\infty(r\sqrt{g_\infty}/2) > 0$ and $\nu_\infty(2r\sqrt{g_\infty}) > 0$. Furthermore, Assumption 2.3 on the properties of the function F is in force which corresponds to [5, Assumption 5.1]. Hence, Point (ii) of [5, Proposition 7.4] on the regularity of the "limit" kernel $\mathcal{K}_T^{\text{prox}}$ holds.

Following [5, (Eq.42)], we define for $u > 0$:

$$\delta_\infty(u, s) = \inf\left\{\delta > 0 : \max_{1 \leq \ell \leq s} \sum_{k=1, k \neq \ell}^{s} |\mathcal{K}_T^{\text{prox}[i,j]}(\theta_\ell, \theta_k)| \leq u \text{ for all } (i,j) \in \{0,1\} \times \{0,1,2\}\right.$$
$$\left.\text{and for all } \ell \neq k,\ \mathfrak{d}_\infty(\theta_k, \theta_\ell) > \delta\right\}.$$

Elementary calculations using (17) and (92) show that for any $u > 0$, $\delta_\infty(u, s) = \sqrt{g_\infty}\,\delta(u, s)$ where δ is defined in (25). We fix $u_\infty = \eta H_\infty^{(2)}(r)$. By assumption, we have that $\delta(u_\infty, s) < +\infty$. Therefore, $\delta_\infty(u_\infty, s)$ is finite and Point (iii) of [5, Proposition 7.4] holds. Recall that we have from Assumption 2.4 (iii) that $C_T \leq 2$ which gives that Point (iv) of [5, Proposition 7.4] holds with $\rho = 2$ therein.

We verify Point (v) of [5, Proposition 7.4] on the proximity between the kernels $\mathcal{K}_T$ and $\mathcal{K}_T^{\text{prox}}$ thanks to Point (iii) of Assumption 2.4. We have verified all the assumptions of [5, Proposition 7.4]. Similarly Points $(i)-(iv)$ of [5, Proposition 7.5] hold with $u'_\infty = u_\infty$.

Finally, acording to [5, Propositions 7.4 and 7.5] Point (v) of Theorem [5, Theorem 2.1] on the existence of certificate functions holds for any subset $\mathcal{Q}^\star$ such that for all $\theta \neq \theta$, we have

$$\mathfrak{d}_T(\theta,\theta) > 2\max(r, 2\delta_\infty(\eta H_\infty^{(2)}(r), s)) = 2\max(r, 2\, g_\infty^{1/2}\delta(\eta H_\infty^{(2)}(r), s)), \tag{93}$$

where $\mathfrak{d}_T$ is defined in (19).

Recall that by assumption $C_T \leq 2$. Since for any $\theta \neq \theta' \in \mathcal{Q}^\star$ we get from the bound (21) on $\mathfrak{d}_T$ and Assumption 2.4 (iv) that:

$$|\theta - \theta'|/C_T > 4\,\sigma_T\, g_\infty^{-1/2}\, \max(r, 2\, g_\infty^{1/2}\delta(\eta H_\infty^{(2)}(r), s)).$$

Thus, inequality (93) holds. We deduce that Point (v) of Theorem [5, Theorem 2.1] is verified. Finally, by [5, Theorem 2.1], there exist finite positive constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}'_2$, $\mathcal{C}_3$, depending on $\mathcal{K}_T^{\text{prox}}$ and on r such that for any $\tau > 0$ and a tuning parameter: $\kappa \geq \mathcal{C}_1\overline{\sigma}\sqrt{\Delta_T \log(\tau)}$, we have the prediction error bound of the estimators $\hat{\beta}$ and $\hat{\vartheta}$ defined in (28) given by (30) with probability larger than $1 - 2\sqrt{g_\infty}\,\mathcal{C}'_2\left(\frac{|\Theta_T|}{\sigma_T\,\tau\sqrt{\log(\tau)}} \vee \frac{1}{\tau}\right)$, where the diameter $|\Theta_T|_{\mathfrak{d}_T}$ of the set Θ_T with respect to the metric $\mathfrak{d}_T$ is bounded by $2\sqrt{g_\infty}|\Theta_T|/\sigma_T$ using (21) and the fact that $C_T \leq 2$. We set $\mathcal{C}_2 = 2\sqrt{g_\infty}\,\mathcal{C}'_2$. In addition, we have (31) with the same probability. A careful reading of the proof of Theorem [5, Theorem 2.1] shows that the constants $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}'_2$, $\mathcal{C}_3$ appearing in its statement depend only on the quantities $M_{i,j} = \sup_{\Theta_\infty^2} |\mathcal{K}_T^{\text{prox}[i,j]}|$ with $i,j \in \{0,1,2,3\}$ and on some contants appearing in the properties of the certificates (denoted $C_N, C'_N, C_F, C_B, c_N, c_F, c_B$ in [5]). By [5, Propositions 7.4 and 7.5], we have that the latter depend only on $\varepsilon_\infty(r\sqrt{g_\infty})$, $\nu_\infty(r\sqrt{g_\infty})$ and $M_{i,j}$ with $i,j \in \{0,1,2,3\}$. We readily show, using (17) to see that $M_{i,j}$ depend only on F, that they do not depend on T but on only r and F. This finishes the proof.

Cristina Butucea, Centre de Recherche en Économie et de Statistiques (CREST), CNRS, École polytechnique, GENES, ENSAE Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
Email address: cristina.butucea@ensae.fr

Jean-François Delmas, CERMICS, École des Ponts, France
Email address: jean-francois.delmas@enpc.fr

Anne Dutfoy, EDF R&D, Palaiseau, France
Email address: anne.dutfoy@edf.fr

Clément Hardy, CERMICS École des Ponts and EDF R&D Palaiseau, France
Email address: clement.hardy@enpc.fr