Model selection for optimal estimation of the mean function of random fields

Fabienne Comte, Céline Duval, Omar Kassi, Valentin Patilea

▶ **To cite this version:**

Fabienne Comte, Céline Duval, Omar Kassi, Valentin Patilea. Model selection for optimal estimation of the mean function of random fields. 2026. ⟨hal-05705767⟩

HAL Id: hal-05705767

https://hal.science/view/index/docid/5705767

Preprint submitted on 28 Jul 2026

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

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

Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Model selection for optimal estimation of the mean function of random fields

Comte, F.* Duval, C.† Kassi, O.‡ Patilea, V.§

July 28, 2026

Abstract

We consider the problem of estimating the mean function of a random field over a compact multidimensional domain, framed as a nonparametric regression problem with highly correlated errors. We adapt the model selection paradigm to this setting, relying on linear-algebra tools for hypermatrices (tensors) to formalize a new nonparametric estimation method, and obtain risk bounds for projection estimators defined as minimizers of a least-squares contrast. The centered random field realizations play the role of nuisance parameters and induce specific theoretical challenges. Optimizing the $\mathbb{L}^2$-risk bounds yields minimax-optimal estimation rates on anisotropic regularity spaces in the common design context, which is new in the functional data literature. For random design, we propose model selection methods in an isotropic setting, providing a new data-driven choice of projection spaces. The new estimator is shown to be minimax-optimal, either exactly or up to logarithmic factors, according to the tail behaviour of the uniform norm of the centered random fields. Numerical examples illustrate that the resulting estimators are parsimonious, outperform competitors, and are automatically selected by a fast algorithm.

MSC2020 subject classifications: Primary 62G05, 62R10; secondary 62M40
Keywords: Adaptive least-squares. Functional data. Minimax optimality. Nonparametric regression.

1 Introduction

In this paper, we consider the problem of estimating, in a nonparametric context, the mean function of a random field over a compact domain $\mathcal{T} \subset \mathbb{R}^d$. The data are generated by n independent and identically distributed realizations of a real-valued random field over $\mathcal{T}$, each one discretely observed at finitely many design points within $\mathcal{T}$ and corrupted by additive, possibly heteroscedastic noise. Two frameworks are considered: first, the case where the design points in $\mathcal{T}$ are deterministic common observation points for all the random fields; this is generally called "*common* (or *synchronous*) design". Second, the case where the design points are random, and their number may differ across the random field realizations; this is generally called "*random* (or *independent*) design".

These two settings are standard in the literature of Functional Data Analysis (FDA), since [9] established the first optimal rates in both cases in dimension $d = 1$, relying on

*Université Paris Cité, CNRS, MAP5, F-75006 Paris, France; fabienne.comte@u-pariscite.fr.
†Sorbonne Université and Université Paris Cité, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France; celine.duval@sorbonne-universite.fr.
‡Bocconi University, Department of Decision Sciences, Milan, Italy; omar.kassi@unibocconi.it.
§Univ. Rennes, Ensai, CREST–UMR 9194, F-35000 France; valentin.patilea@ensai.fr. V. Patilea acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-24-CE40-2439 (FUNMathStat project).

smoothing splines for the upper bounds. See also [41] for extensions to penalized splines. Most results are related to the case $d = 1$, see [43, 41, 14, 8]. Very few works in FDA consider a multidimensional setting; we may mention [24], where general theory of random functions is developed, but estimation questions are studied for $d = 1$ only. Assuming the regularity of the mean to be known and the random fields Hölder continuous, [7] recently derived supremum-norm optimal rates for the Hölder-smooth mean function under a multivariate, synchronous design points, while [26] proposed a plug-in adaptive estimation method under a multivariate, independent design, relying on tensorized Fourier bases together with an estimator of the regularity. We focus here on the estimation of the mean function, under either the common or the independent design, with several new proposals and results: this is the consequence of the meeting of two worlds, FDA questions and model selection methods.

There is a variety of applications in medicine, biophysics, neuroscience or environmental science, that consider recording multiple observations obtained as discretized measurements of curves or surfaces, and all these contexts are consistent with our framework. For example, the Argo project has become a major source of data for oceanography and climate research and has recently attracted attention in the statistics literature, see e.g. [42]. The usual modeling paradigm considers that the curves or surfaces are realizations of an underlying random function defined on a possibly multidimensional domain, see also [40, 29] and [18] for recent surveys, and [13] for a recent textbook.

In this paper, we consider the problem of estimating the mean of the random field as a nonparametric multidimensional regression problem in the presence of highly correlated errors, to which we adapt the model selection paradigm. General model selection strategies have been introduced in [5], later synthesized in the book [32]. The refined application to fixed design regression, related to the common design setting in FDA, was developed in [2], and stability conditions in the setup of random design were later formalized in [11]. The multidimensional regression setting in the presence of random design is studied in [3, 16]. This last paper develops linear-algebra tools for hypermatrices (tensors), on which we also rely. This allows us to present general estimation results and risk bounds for projection estimators relying on the minimization of a least-squares contrast. The centered random field realizations play the role of nuisance parameters and induce dependent error terms, leading to specific theoretical challenges. Optimizing the $\mathbb{L}^2$-risk bounds obtained in a first step leads to general rates on anisotropic regularity spaces in the common design context, which are completely new in an FDA setting and can be proved to be optimal in the minimax sense. In addition, we propose model selection methods in the presence of random design and in an isotropic setting, leading to a new data-driven choice of the projection spaces. Lastly, we illustrate through numerical examples that the estimation strategy leads to parsimonious estimators automatically selected by a fast algorithm. It is worth mentioning a fundamental difference between the fixed and the random design cases: for the risk to be small, the fixed design case requires both the number n of independent random field realizations and the number of design points per realization to be large, while in the random design case, only n must be large and the numbers of design points per realization can be bounded. We also underline that we set our assumptions on the mean function while the centered random field is left unconstrained in terms of regularity (it may be discontinuous or of low functional regularity, e.g. a jump process or fractional Brownian motion).

Our method is different from [26], our main competitor for mean estimation in a multidimensional setting with random design. In that paper, the authors propose a weighted average strategy for the reconstruction of Fourier coefficients, assuming a known density for the design. Their results include evaluation of confidence sets, and adaptivity is obtained by plug-in of an estimator of the isotropic Hölder-regularity index. However, the adaptivity in such plug-in-based approaches relies on rather asymptotic considerations, while our model

selection device appears to be the first model selection approach in FDA genuinely relying on finite-sample theoretical guarantees. [25] proposed strategies for estimating the regularity of the random field in the multivariate anisotropic FDA context, extending the local intrinsic stationarity concept introduced by [23] for spatial statistics. The approach in [25] implicitly assumes that the mean function is smoother than the random field realizations, and is not yet applied to construct plug-in adaptive estimation for the mean.

The paper is organized as follows. The remainder of this introduction provides general notation and definitions, specifically those related to hypermatrices. In Section 2, we study the common (fixed) design setting. We derive a risk bound for the least-squares estimator in Section 2.3. Although unsurprising, it appears to be new in the multidimensional context and yields new rates on anisotropic regularity spaces, together with corresponding sparse and dense regimes, see Section 2.4. We discuss their minimax optimality, since the best estimator depends on the unknown regularity of the multivariate mean function. In Section 2.5, a data-driven model selection procedure is defined and proved to achieve an automatic trade-off over isotropic regularity spaces. In that procedure, the uniform norm of the error variance function appears in the penalty; a natural estimator of it is proposed and shown to suffice. In Section 3, the procedure is extended to the independent random design framework. The risk bound is expressed in terms of a squared empirical norm or a $\mathbb{L}^2$-norm weighted by the design density, see Section 3.2. General anisotropic rates can be derived and the sparse and dense regimes can be defined accordingly. The rates are proved to be minimax optimal (Section 3.3) and a model selection procedure is implemented in an isotropic context (Section 3.4). The adaptive estimator is proved to be either optimal or nearly optimal (*i.e.* up to logarithmic terms), depending on the moment conditions imposed on the random field. The proof relies on a nonstandard analysis of empirical processes, conditional on the random field realizations. Section 4 proposes a fully data-driven implementation of the model selection approach in different contexts and provides numerical evidence on the performance of the new adaptive estimators. Our model selection estimators perform well compared with data-driven kernel-based and spline methods in random design settings, and with interpolation in fixed design. Section 5 contains the main proofs. Additional technical results are gathered in an Appendix.

Notation and definitions

In what follows, unless stated otherwise, we take $\mathcal{T} = [0,1]^d$ as the reference domain. Moreover, we will use the following notations: $\lesssim$ means that the left side is bounded by a positive constant times the right side; for $a, b \in \mathbb{R}$, $a \asymp b$ means $a \lesssim b$ and $b \lesssim a$. The vectors are column matrices and, for any vector $\mathbf{a}$, $\mathbf{a}^\top$ and $\|\mathbf{a}\|$ denote its transpose and the Euclidean norm, respectively. For any continuous real-valued function f defined on $\mathcal{T}$, let $\|f\|_\infty = \sup_{t\in\mathcal{T}} |f(t)|$ denote its uniform norm. For $d \geq 1$, let $\mathbb{N}_{>0}^d$ be the set of d-tuples of positive integers. Finally, $C, C', C_1, C_2, \mathfrak{C}, \mathfrak{c}, \ldots$, denote generic constants that may differ from line to line.

The model selection procedures that we propose below for functional data on the multidimensional domain $\mathcal{T}$ are based on tensorized bases. The associated projections in such bases are formalized using hypermatrix calculus, which naturally extends matrix calculus by allowing entries indexed by three or more indices instead of only two. We briefly recall the definitions and main properties that will be used in the sequel. For detailed presentations of hypermatrices and their properties, see [21, Ch. 15] and [16].

In the following, we denote $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{N}_{>0}^d$ a $d-$multi-index, and we denote by $D_{\boldsymbol{m}} = \prod_{l=1}^d m_l$ its dimension. For $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{N}_{>0}^d$, $\boldsymbol{k} \leq \boldsymbol{m}$ means $k_l \leq m_l$, $\forall 1 \leq l \leq d$. For $d, d' \geq 1$, let $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{N}_{>0}^d$ and $\boldsymbol{m}' = (m'_1, \ldots, m'_{d'}) \in \mathbb{N}_{>0}^{d'}$ be $d-$ and d'-multi-indices, respectively. Let $\mathcal{M}_{\boldsymbol{m},\boldsymbol{m}'}(\mathbb{R})$ be the set of $\boldsymbol{m} \times \boldsymbol{m}'-$hypermatrices. The entries of a hypermatrix $A \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}'}(\mathbb{R})$ are real numbers denoted $[A]_{\boldsymbol{k},\boldsymbol{k}'}$, $\boldsymbol{k} \leq \boldsymbol{m}, \boldsymbol{k}' \leq \boldsymbol{m}'$. The set

$\mathcal{M}_{\boldsymbol{m},\boldsymbol{m}'}(\mathbb{R})$ is a vector space with operations $[A+B]_{\boldsymbol{k},\boldsymbol{k}'} = [A]_{\boldsymbol{k},\boldsymbol{k}'} + [B]_{\boldsymbol{k},\boldsymbol{k}'}$, $[\alpha A]_{\boldsymbol{k},\boldsymbol{k}'} = \alpha [A]_{\boldsymbol{k},\boldsymbol{k}'}$. For $A \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}'}(\mathbb{R})$ we define its transpose $A^\top$ as the hypermatrix in $\mathcal{M}_{\boldsymbol{m}',\boldsymbol{m}}(\mathbb{R})$ with entries

$$\forall \boldsymbol{k} \leq \boldsymbol{m}, \boldsymbol{k}' \leq \boldsymbol{m}' \qquad \left[A^\top\right]_{\boldsymbol{k}',\boldsymbol{k}} = \left[A\right]_{\boldsymbol{k},\boldsymbol{k}'}.$$

The *contraction product* of $A \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}'}(\mathbb{R})$ and $B \in \mathcal{M}_{\boldsymbol{m}',\boldsymbol{m}''}(\mathbb{R})$, with $\boldsymbol{m} \in \mathbb{N}_{>0}^d$, $\boldsymbol{m}' \in \mathbb{N}_{>0}^{d'}$ and $\boldsymbol{m}'' \in \mathbb{N}_{>0}^{d''}$, is the hypermatrix $AB \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}''}(\mathbb{R})$ with entries

$$\forall \boldsymbol{k} \leq \boldsymbol{m},\ \boldsymbol{k}'' \leq \boldsymbol{m}'', \qquad [AB]_{\boldsymbol{k},\boldsymbol{k}''} = \sum_{\boldsymbol{k}' \leq \boldsymbol{m}'} [A]_{\boldsymbol{k},\boldsymbol{k}'} [B]_{\boldsymbol{k}',\boldsymbol{k}''}.$$

The identity hypermatrix in $\mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$, denoted as $\mathrm{Id}_{\boldsymbol{m}}$, is defined as :

$$\forall \boldsymbol{k}, \boldsymbol{k}' \leq \boldsymbol{m}, \qquad [\mathrm{Id}_{\boldsymbol{m}}]_{\boldsymbol{k},\boldsymbol{k}'} = \delta_{\boldsymbol{k},\boldsymbol{k}'},$$

where $\delta_{\boldsymbol{k},\boldsymbol{k}'}$ is the Kronecker symbol defined for the multi-indices $\boldsymbol{k}, \boldsymbol{k}'$. A hypermatrix $A \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$ is invertible if a hypermatrix $B \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$ exists such that $AB = BA = \mathrm{Id}_{\boldsymbol{m}}$. The inverse of A is denoted by A^{-1}.

Since $\mathcal{M}_{\boldsymbol{m},1}(\mathbb{R})$ is an extension of the space of column matrices we also denote it by $\mathbb{R}^{\boldsymbol{m}}$. On this space we consider the scalar product and the norm defined as follows : for $a, b \in \mathbb{R}^{\boldsymbol{m}}$,

$$\langle a, b \rangle_{\mathbb{R}^{\boldsymbol{m}}} = a^\top b = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} [a]_{\boldsymbol{k},1} [b]_{\boldsymbol{k},1}, \qquad \|a\|_{\mathbb{R}^{\boldsymbol{m}}}^2 = a^\top a = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \left([a]_{\boldsymbol{k},1}\right)^2.$$

Moreover, the space $\mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$ can be identified with the endomorphism ring of $\mathbb{R}^{\boldsymbol{m}}$. Therefore, the notions of trace, eigenvalues and eigenvectors, positive semi-definiteness, and operator norms can be extended to hypermatrices. See [31]. In particular, for $A \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$

$$\|A\|_{\mathrm{op}} = \sup_{a \in \mathbb{R}^{\boldsymbol{m}} \setminus \{0\}} \frac{\|Aa\|_{\mathbb{R}^{\boldsymbol{m}}}}{\|a\|_{\mathbb{R}^{\boldsymbol{m}}}} = \sqrt{\lambda_{\max}(A^\top A)}.$$

2 Common fixed design

First, we develop a new model selection approach for estimating the mean function from functional data observed on a common fixed grid of points, where the observation noise may have non-constant variance.

2.1 Model and assumptions

In this section, we will consider the case of a common fixed design, also known as a synchronous design. The model is defined by: for $j = 1, \ldots, J$ and $i = 1, \ldots, n$,

$$Y_{i,j} = g(T_j) + X_i(T_j) + \sigma_\varepsilon(T_j)\varepsilon_{i,j}, \tag{1}$$

where the non-random function $g : \mathcal{T} \to \mathbb{R}$ is the quantity of interest to be estimated from observations $(Y_{i,j}, T_j)$, and $\sigma_\varepsilon^2 : \mathcal{T} \to [0, \infty)$ is an unknown variance function allowing for heteroscedastic errors. The X_i are random fields.

Assumption 1. *(i) the random fields $\{(X_i(t))_{t\in\mathcal{T}}\}$ are i.i.d. such that*

$$\mathbb{E}[X_1(t)] = 0 \quad \textit{and} \quad \mathfrak{C}_X := \sup_{t\in\mathcal{T}} \mathbb{E}[X_1^2(t)] < +\infty;$$

(ii) the error variance function is bounded, i.e., $\|\sigma_\varepsilon^2\|_\infty := \sup_{t\in\mathcal{T}} \sigma_\varepsilon^2(t) < +\infty$*;*

(iii) the variables $\{\varepsilon_{i,j}\}$ are i.i.d. centered with unit variance;

(iv) the $\{T_j\} \subset \mathcal{T}$ are deterministic common design points for all X_i;

(v) the random fields $\{(X_i(t))_{t\in\mathcal{T}}\}$ and the variables $\{\varepsilon_{i,j}\}$ are independent.

2.2 Least-squares estimator

We first define the estimator as the nonparametric least-squares contrast by setting, for a function $h : \mathcal{T} \to \mathbb{R}$,

$$\gamma_J(h) = \frac{1}{J} \sum_{j=1}^{J} \left\{ \bar{Y}_{\bullet,j} - h(T_j) \right\}^2,$$

where

$$\bar{Y}_{\bullet,j} = \frac{1}{n} \sum_{i=1}^{n} Y_{i,j} = g(T_j) + \frac{1}{n} \sum_{i=1}^{n} X_i(T_j) + \sigma_\varepsilon(T_j)\, \bar{\varepsilon}_{\bullet,j}, \quad \bar{\varepsilon}_{\bullet,j} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{i,j}. \tag{2}$$

For $\boldsymbol{m} = (m_1, \ldots, m_d) \in \mathbb{N}_{>0}^d$, we consider h belonging to a approximation space $S_{\boldsymbol{m}}$ defined by

$$S_{\boldsymbol{m}} = \operatorname{span}\{\varphi_{\boldsymbol{k}} : \boldsymbol{k} \leq \boldsymbol{m}\},$$

with $\{\varphi_{\boldsymbol{k}}\}$ a tensorized orthonormal family of functions in $\mathbb{L}^2(\mathcal{T}, dt)$ defined by $\varphi_{\boldsymbol{k}} = \varphi_{k_1} \otimes \ldots \otimes \varphi_{k_d}$. By definition, $S_{\boldsymbol{m}} \subset S_{\boldsymbol{m}'}$ if $\boldsymbol{m} \leq \boldsymbol{m}'$. Below, $S_{\boldsymbol{m} \vee \boldsymbol{m}'}$ is the space corresponding to the multi-index $\boldsymbol{m} \vee \boldsymbol{m}' = (m_1 \vee m_1', \ldots, m_d \vee m_d')$.

We define, whenever it exists, the following estimator:

$$\widehat{g}_{\boldsymbol{m}} = \arg \min_{h \in S_{\boldsymbol{m}}} \gamma_J(h). \tag{3}$$

An element of $S_{\boldsymbol{m}}$ is entirely determined by the coefficients of its expansion in the tensorized orthonormal family. Let

$$\mathbf{Y}_{\bullet} = (Y_{\bullet,1}, \ldots, Y_{\bullet,J})^\top \in \mathbb{R}^J = \mathcal{M}_{J,1}(\mathbb{R}), \qquad \Phi_{\boldsymbol{m}} = [\varphi_{\boldsymbol{k}}(T_j)]_{1 \leq j \leq J,\, \boldsymbol{k} \leq \boldsymbol{m}} \in \mathcal{M}_{J,\boldsymbol{m}}(\mathbb{R}),$$

and consider the hypermatrix $\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}} \in \mathcal{M}_{\boldsymbol{m},\boldsymbol{m}}(\mathbb{R})$. Note that if $\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}}$ is invertible, then necessarily $D_{\boldsymbol{m}} \leq J$. By [16, Lemma 8], if $\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}}$ is invertible, then the problem (3) has the unique solution

$$\widehat{g}_{\boldsymbol{m}} = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \widehat{a}_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}, \quad \vec{a}^{(\boldsymbol{m})} = [\widehat{a}_{\boldsymbol{k}}]_{\boldsymbol{k} \leq \boldsymbol{m}} = (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \mathbf{Y}_{\bullet} \in \mathbb{R}^{\boldsymbol{m}}.$$

Note that $\widehat{g}_{\boldsymbol{m}}$ equivalently minimizes $(nJ)^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n} \{Y_{i,j} - h(T_j)\}^2$.

2.3 Risk bound on a fixed projection space

We define the empirical norm as $\|h\|_J^2 = J^{-1} \sum_{j=1}^{J} h^2(T_j)$. The following risk bound holds.

Proposition 1. *Consider the estimator (3) in model (1) under Assumption 1. Let $\boldsymbol{m} \in \mathbb{N}_{>0}^d$ and $S_{\boldsymbol{m}}$ be an approximation space. If $\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}}$ is invertible (thus $D_{\boldsymbol{m}} \leq J$), then*

$$\mathbb{E}(\|\widehat{g}_{\boldsymbol{m}} - g\|_J^2) \leq \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_J^2 + \frac{\|\sigma_\varepsilon^2\|_\infty D_{\boldsymbol{m}} + \mathfrak{C}_X\, J}{nJ}. \tag{4}$$

In the context of model selection, Proposition 1 provides a non-asymptotic bound for the empirical norm, thereby complementing the asymptotic bounds derived in [9, Th.2.1] and [41, Th.3.1]. Note that for the upper bound in (4) to be small, both J and n must be large. Indeed, at a given location T_j, large n ensures that the contribution of X is cancelled. On the contrary, large J ensures that $\mathcal{T}$ is densely covered and allows to estimate g.

Remark 1 (Comparison with fixed design regression model in dimension $d = 1$)**.** *The above results can be connected to those in usual nonparametric regression estimation in presence of dependent errors. See [4, Sec. 3.3], [20, 17]. Indeed, by (2) it holds that*

$$\bar{Y}_{\bullet,j} = g(T_j) + \bar{e}_{\bullet,j} \quad \textit{with} \quad \bar{e}_{\bullet,j} = \frac{1}{n}\sum_{i=1}^{n} X_i(T_j) + \sigma_\varepsilon(T_j)\,\bar{\varepsilon}_{\bullet,j}.$$

Then $\mathbb{E}[\bar{e}_{\bullet,j}] = 0$ *and* $\mathbb{E}[\bar{e}_{\bullet,j}\bar{e}_{\bullet,k}] = n^{-1}\sigma_X(T_j, T_k)$ *if* $j \neq k$*, with* $\sigma_X(s,t) = \mathbb{E}[X_1(s)X_1(t)]$*. If* n *is fixed and* J *increases, the rate will depend on the properties of* σ_X *(long memory or not). However, if* n *increases the role of the properties of* σ_X *disappears.*

2.4 Discussion of the rate of convergence

Let $\mathbb{R}_+ = (0, +\infty)$. Assume that g belongs to a regularity class of smoothness $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and radius R associated with the basis $\{\varphi_{\boldsymbol{k}}\}$, such that

$$\inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_J^2 \leq c(\boldsymbol{\alpha}, R) \sum_{j=1}^{d} m_j^{-2\alpha_j}, \tag{5}$$

for some $c(\boldsymbol{\alpha}, R)$ independent of $\boldsymbol{m}$. Let $\overline{\boldsymbol{\alpha}} \in \mathbb{R}_+$ defined by

$$\frac{1}{\overline{\boldsymbol{\alpha}}} = \frac{1}{d}\sum_{j=1}^{d} \frac{1}{\alpha_j}. \tag{6}$$

Then, the solution $\boldsymbol{m}_{\text{opt}}$ of the constraint optimization problem

$$\inf_{\{\boldsymbol{m}: D_{\boldsymbol{m}} \leq J\}} \left(\sum_{j=1}^{d} m_j^{-2\alpha_j} + \frac{D_{\boldsymbol{m}}}{nJ} \right)$$

is given by the choice

$$m_{j,\text{opt}} = \begin{cases} (nJ)^{\overline{\boldsymbol{\alpha}}/\{\alpha_j(2\overline{\boldsymbol{\alpha}}+d)\}} & \text{while } \prod_{j=1}^{d} m_j = (nJ)^{d/(2\overline{\boldsymbol{\alpha}}+d)} \leq J \quad \textit{i.e.,} \quad J^{2\overline{\boldsymbol{\alpha}}/d} \geq n, \\ J^{\overline{\boldsymbol{\alpha}}/(d\alpha_j)} & \text{otherwise.} \end{cases} \tag{7}$$

Note that in the case $J^{2\overline{\boldsymbol{\alpha}}/d} \geq n$ the rate of the bound in (4) is

$$(nJ)^{-2\overline{\boldsymbol{\alpha}}/(2\overline{\boldsymbol{\alpha}}+d)} + n^{-1} \asymp n^{-1},$$

and when $m_j = J^{\overline{\boldsymbol{\alpha}}/(d\alpha_j)}$ then $\prod_{j=1}^{d} m_j = J$ and the rate is $J^{-2\overline{\boldsymbol{\alpha}}/d} + n^{-1} \asymp J^{-2\overline{\boldsymbol{\alpha}}/d}$. We refer to [27] for computations. To sum up, this leads to the rate

$$\begin{aligned} \max\left[(nJ)^{-2\overline{\boldsymbol{\alpha}}/\{2\overline{\boldsymbol{\alpha}}+d\}}, n^{-1}\right] \mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} \geq n\}} + \max\left[J^{-2\overline{\boldsymbol{\alpha}}/d}, n^{-1}\right] \mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} < n\}} \\ = n^{-1}\mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} \geq n\}} + J^{-2\overline{\boldsymbol{\alpha}}/d}\mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} < n\}}, \end{aligned}$$

for the estimator defined with $\boldsymbol{m} = \boldsymbol{m}_{\text{opt}}$, as stated formally in the following result.

Proposition 2. *Let* $\widehat{g}_{\boldsymbol{m}}$ *be the least-squares estimator of* g *obtained if* $\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}}$ *is invertible and Assumption 1 holds. Assume that* g *belongs to a regularity class of smoothness* $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ *and radius* R *such that Condition (5) holds. Then,* $\boldsymbol{m} = \boldsymbol{m}_{\text{opt}}$ *with the components as in (7), yields*

$$\mathbb{E}(\|\widehat{g}_{\boldsymbol{m}_{\text{opt}}} - g\|_J^2) \leq C(\boldsymbol{\alpha}, R)\left(n^{-1}\mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} \geq n\}} + J^{-2\overline{\boldsymbol{\alpha}}/d}\mathbf{1}_{\{J^{2\overline{\boldsymbol{\alpha}}/d} < n\}}\right),$$

where $C(\boldsymbol{\alpha}, R)$ *is a positive constant depending on* $c(\boldsymbol{\alpha}, R)$*,* $\mathfrak{C}_X$ *and* $\|\sigma_\varepsilon^2\|_\infty$*.*

The term of order n^{-1} in the upper bound in Proposition 2 corresponds to the classical parametric rate and is therefore optimal (see also the proof of Theorem 2 in the random design case). The term of order $J^{-2\overline{\boldsymbol{\alpha}}/d}$ corresponds to the approximation rate in squared $\mathbb{L}^2$-norm for a function of regularity $\boldsymbol{\alpha}$ observed at J points, and is proved to be optimal in dimension $d = 1$ (see [9, Theorem 2.1]). For $d \geq 1$ and the uniform norm in an isotropic setting, a related optimality result is established in [7, Theorem 2] (with $\boldsymbol{p}_{\min}$ corresponding to $J^{-1/d}$).

In dimension $d = 1$, with $\boldsymbol{\alpha}$ and $\boldsymbol{m}$ reducing to scalars α and m, the case $J^{2\alpha} \geq n$ is usually referred to as the *dense regime* ([43] proposes to refine the definition of the dense regime by calling the case $n^{-1}J^{2\alpha} \to \infty$ the ultra-dense regime; we do not follow this refinement for the sake of simplicity). This can generally be defined as the case where the estimator achieves the parametric rate, that of the infeasible empirical estimator with perfectly observed X_i. By contrast the case $J^{2\overline{\boldsymbol{\alpha}}/d} < n$ is the so-called *sparse regime*. See [9]. Additionally, [2, Section 4] shows that, when $d = 1$, Condition (5) is satisfied when g belongs to a Besov space, namely $g \in \mathcal{B}_{\alpha,2,\infty}$ with $\alpha > 1/2$. In that case,

$$\inf_{h \in S_m} \|h - g\|_J^2 \leq c(\alpha)\, |g|_{\alpha,2}^2 \left\{\dim(S_m)\right\}^{-2\alpha},$$

where S_m is either the space spanned by piecewise polynomials of degree at most r (*e.g.*, $r = 3$) on dyadic partitions with step size 2^{-m}, or the space of trigonometric polynomials of degree 2^m, and $|\cdot|_{\alpha,2}$ denotes the seminorm of the Besov space $\mathcal{B}_{\alpha,2,\infty}$. See also [15, p.354, 205].

In the following, let $L \in \mathbb{R}_+$ and $\boldsymbol{\alpha} \in \mathbb{R}_+^d$. In the multidimensional setting, consider the anisotropic Sobolev spaces defined by

$$S(\boldsymbol{\alpha}, L) = \Bigg\{ f \in \mathbb{L}^2(\mathcal{T}, dt) : f = \sum_{\boldsymbol{k} \in \mathbb{N}_{>0}^d} a_{\boldsymbol{k}} \varphi_{\boldsymbol{k}} \quad \text{such that} \quad \sum_{\boldsymbol{k} \in \mathbb{N}_{>0}^d} \left\{2\lfloor k_l/2 \rfloor\right\}^{2\alpha_l} a_{\boldsymbol{k}}^2 \leq L \text{ for any } l = 1, \ldots, d \Bigg\}, \tag{8}$$

where $a_{\boldsymbol{k}} = \langle f, \varphi_{\boldsymbol{k}} \rangle$ and $\{\varphi_{\boldsymbol{k}}\}$ is a tensorized basis: $\varphi_{\boldsymbol{k}} = \varphi_{k_1} \otimes \cdots \otimes \varphi_{k_d}$. For $\boldsymbol{m} \in \mathbb{N}_{>0}^d$, and $f = \sum_{\boldsymbol{k} \in \mathbb{N}_{>0}^d} a_{\boldsymbol{k}} \varphi_{\boldsymbol{k}} \in S(\boldsymbol{\alpha}, L)$, let $f_{\boldsymbol{m}} = \sum_{\mathbf{k} \leq \mathbf{m}} a_{\mathbf{k}} \varphi_{\mathbf{k}}$. If g belongs to $S(\boldsymbol{\alpha}, L)$, it holds

$$\|g - g_{\mathbf{m}}\|^2 \leq \sum_{l=1}^d \sum_{\{\boldsymbol{k}: k_l \geq m_l\}} a_{\boldsymbol{k}}^2 = \sum_{l=1}^d \sum_{\{\boldsymbol{k}: k_l \geq m_l\}} \left\{2\lfloor k_l/2 \rfloor\right\}^{-2\alpha_l} \left[\left\{2\lfloor k_l/2 \rfloor\right\}^{2\alpha_l} a_{\boldsymbol{k}}^2\right] \leq L \sum_{l=1}^d m_l^{-2\alpha_l}. \tag{9}$$

Consider $\{\varphi_{\boldsymbol{k}}\}$ as the tensorized trigonometric basis with univariate basis functions $\varphi_1(x) = \mathbf{1}_{[0,1]}(x)$, $\varphi_{2j}(x) = \sqrt{2}\cos(2\pi j x)\mathbf{1}_{[0,1]}(x)$, $\varphi_{2j+1}(x) = \sqrt{2}\sin(2\pi j x)\mathbf{1}_{[0,1]}(x)$, $j \geq 1$. Then, in dimension $d = 1$, for equally spaced design points $T_j = j/J$, $j \leq J$, the trigonometric identity $\sum_{j=1}^J \exp(i2\pi\ell j/J) = 0$ for any $1 \leq \ell \leq J - 1$ yields $\langle \varphi_k, \varphi_{k'} \rangle_J = \delta_{k,k'} = \langle \varphi_k, \varphi_{k'} \rangle$, $1 \leq k, k' \leq J$, provided J is odd. Consequently, it holds $\|f\| = \|f\|_J$. This can be extended to the multidimensional setting for $J = \mathfrak{J}^d$ for $\mathfrak{J} \in \mathbb{N}_{>0}$ and $T_{\boldsymbol{j}} = (j_1/\mathfrak{J}, \ldots, j_d/\mathfrak{J})$, with $j_1, \ldots, j_d \leq \mathfrak{J}$. It follows that

$$\inf_{h \in S_{\boldsymbol{m}}} \|h - f\|_J^2 = \inf_{h \in S_{\boldsymbol{m}}} \|h - f\|^2 = \|f_{\boldsymbol{m}} - f\|^2.$$

Thus, if $g \in \mathcal{S}(\boldsymbol{\alpha}, L)$ and $\{\varphi_{\boldsymbol{k}}\}$ is the tensorized trigonometric basis, Condition (5) holds true and the rate in Proposition 2 is achieved.

For $f \in \mathcal{S}(\boldsymbol{\alpha}, L)$, a condition is imposed on the decay rate of the basis expansion coefficients of f. For the trigonometric basis, a connection with the regularity of f, namely the existence

of derivatives up to order $\boldsymbol{\alpha}$ is established in [36] for $d = 1$ and in [34, Section 4.2.2 and Proposition A.1] for $d \geq 2$ (see also Section 3.3 below).

However, as $\boldsymbol{\alpha}$ is unknown, the proposed choice of $\mathbf{m}_{\text{opt}}$ cannot be made in practice. The aim of Section 2.5 hereafter is precisely to propose a model selection procedure that provides an automatic finite-sample trade-off.

2.5 Model selection for common design on isotropic spaces

We restrict the study to isotropic projection spaces $S_{\boldsymbol{m}}$; in this case, multi-indices $\boldsymbol{m} \in \mathbb{N}_{>0}^d$ reduce to $\boldsymbol{m} = m\mathbf{1}_d$, where $\mathbf{1}_d = (1, \ldots, 1)$ and $m \in \mathbb{N}_{>0}$, and $D_{\boldsymbol{m}} = m^d$.

2.5.1 Procedure and adaptive bound

Let

$$\mathcal{M}_J = \{\boldsymbol{m} = m\mathbf{1}_d : 1 \leq m \leq J \text{ and } \Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}} \text{ is invertible}\}.$$

The model selection procedure is defined as follows : for some $\kappa \geq 0$, define

$$\widehat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m} \in \mathcal{M}_J} \left\{\gamma_J(\widehat{g}_{\boldsymbol{m}}) + \text{pen}(\boldsymbol{m})\right\}, \quad \text{pen}(\boldsymbol{m}) = \kappa \|\sigma_\varepsilon^2\|_\infty \frac{m^d}{nJ}. \tag{10}$$

Theorem 1. *Assumption 1 holds. There exists κ_0 such that for $\kappa \geq \kappa_0$,*

$$\mathbb{E}(\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2) \leq \inf_{\boldsymbol{m} \in \mathcal{M}_J} \left\{3 \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_J^2 + 4\kappa \|\sigma_\varepsilon^2\|_\infty \frac{m^d}{nJ}\right\} + \frac{C'}{n}\left(\|\sigma_\varepsilon^2\|_\infty + \mathfrak{C}_X\right),$$

where C' is a constant that does not depend on n, J or $\|\sigma_\varepsilon^2\|_\infty$.

Remark 2. *Note that in the case where $\sigma_\varepsilon \equiv 0$, there is no penalty involved in the choice of the model in (10). Then, the choice will correspond to the largest model,* i.e., *$\widehat{\boldsymbol{m}} = J\mathbf{1}_d$. This choice recovers the ideal empirical estimator $\widehat{g}(T_j) = \bar{Y}_{\bullet,j}$ (no smoothing).*

Remark 3. *From a practical point of view, the model selection procedure requires the choice of a value for κ_0 and the knowledge of $\|\sigma_\varepsilon^2\|_\infty$ or an estimate of it. Following the lines of the proof of Theorem 1, it can be seen that a possible theoretical value is $\kappa_0 = 64$. However, smaller values of this parameter can be calibrated in practice by means of preliminary numerical experiments. See the discussion in Section 4. Finally, a simple data-driven choice concerning $\|\sigma_\varepsilon^2\|_\infty$ is proposed in Section 2.5.2 and implemented in Section 4.*

2.5.2 Error variance estimator

We distinguish two cases, depending on whether σ_ε is constant (homoscedasticity) or not (heteroscedasticity). If σ_ε is constant (possibly $\sigma_\varepsilon = 0$), for $J \geq 2$ let

$$V_{\varepsilon,n} = \frac{1}{2n(J-1)} \sum_{i=1}^{n} \sum_{j=1}^{J-1} (Y_{i,j+1} - Y_{i,j})^2. \tag{11}$$

Under mild assumptions on the fields X_i and the mean g, it can be shown that there exists a constant $a > 0$ and a map $\sigma_\varepsilon \to C(\sigma_\varepsilon)$, bounded on bounded intervals, such that

$$\mathbb{E}\left|V_{\varepsilon,n} - \sigma_\varepsilon^2\right| \leq C(\sigma_\varepsilon) J^{-a}. \tag{12}$$

The details are provided in the Appendix (see Proposition 4).

Regarding the impact on the model selection and the risk bound derived in Theorem 1 by replacing σ_ε^2 with the estimate $V_{\varepsilon,n}$, note that the selection step (10) becomes

$$\widehat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}\in\mathcal{M}_J} \left\{\gamma_J(\widehat{g}_{\boldsymbol{m}}) + \widehat{\mathrm{pen}}(\boldsymbol{m})\right\}, \quad \widehat{\mathrm{pen}}(\boldsymbol{m}) = \kappa V_{\varepsilon,n}\frac{m^d}{nJ}. \tag{13}$$

A careful inspection of the proof of Theorem 1 reveals that the risk bound remains valid at the cost of adding a term that is at most of the same order, and possibly inflating the value of κ_0, if there exists a function $\sigma_\varepsilon \to c(\sigma_\varepsilon)$, bounded on bounded intervals, such that

$$\mathbb{E}\max_{\boldsymbol{m}\in\mathcal{M}_J}\left|\widehat{\mathrm{pen}}(\boldsymbol{m}) - \mathrm{pen}(\boldsymbol{m})\right| \le \frac{c(\sigma_\varepsilon)J^{-a}}{n}. \tag{14}$$

Note that, since $m^d \le J$ for every $\boldsymbol{m} = m\mathbf{1}_d \in \mathcal{M}_J$, it holds

$$\forall \boldsymbol{m}\in\mathcal{M}_J, \quad \left|\widehat{\mathrm{pen}}(\boldsymbol{m}) - \mathrm{pen}(\boldsymbol{m})\right| = \kappa\frac{m^d}{nJ}\left|V_{\varepsilon,n} - \sigma_\varepsilon^2\right| \le \frac{\kappa}{n}\left|V_{\varepsilon,n} - \sigma_\varepsilon^2\right|. \tag{15}$$

Then, (14) follows from (15) and (12), *regardless of whether* $\sigma_\varepsilon = 0$ *or* $\sigma_\varepsilon > 0$. In conclusion, under mild conditions on the mean function and the increments of the fields X_i, the bound in Theorem 1 remains valid up to terms of the same order, and possibly inflating the value of κ_0, if the selection step (10) is replaced by that in (13). Moreover, since (15) holds for every $\sigma_\varepsilon \ge 0$, the model selection based on (13) adapts to the presence or absence of noise.

In the heteroscedastic case (σ_ε^2 is no longer constant), let $W_{\varepsilon,n}(j) = (2n)^{-1}\sum_{i=1}^n (Y_{i,j+1} - Y_{i,j})^2$. Then, the estimator of $\|\sigma_\varepsilon^2\|_\infty$ can be defined as $V_{\varepsilon,n} = \max_{1\le j\le J-1} W_{\varepsilon,n}(j)$. The properties of this estimator are discussed in the Appendix.

3 Independent random design

In this section, we consider the case where the design points are drawn randomly and their number may vary from curve to curve. This setting differs substantially from the synchronous design case. In the synchronous case, a common observation grid restricts the available information to a fixed finite set of locations, identical across curves. By contrast, random design points spread across the entire domain as their number grows, so the data carry richer information about the global behavior of the underlying processes.

3.1 Model and projection estimator

Let us reconsider the model

$$Y_{i,j} = g(T_{i,j}) + X_i(T_{i,j}) + \sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}, \qquad j = 1,\dots,N_i, \quad i = 1,\dots,n. \tag{16}$$

The data points are $(Y_{i,j}, T_{i,j})$, $j = 1,\dots,N_i$, $i = 1,\dots,n$. Let $N = \sum_{i=1}^n N_i$. With a random design, for simplicity let q be an index replacing the pair of indices (i,j). Thus, when no confusion arises, instead of $T_{1,1},...,T_{1,N_1},T_{2,1},...,T_{n,N_n}$ we write $T_1,...,T_N$ or T_q, $1 \le q \le N$. We use the same convention for the $Y_{i,j}$'s. With this notation, $\sum_{q=1}^N$ replaces $\sum_{i=1}^n\sum_{j=1}^{N_i}$.

Assumption 2. *The conditions Assumption 1(i) to Assumption 1(iii) and Assumption 1(v) hold. In addition:*

(iv') *the* $\{T_{i,j}\} \subset \mathcal{T}$ *are i.i.d., independent of the random fields* $\{(X_i(t))_{t\in\mathcal{T}}\}$ *and the errors* $\{\varepsilon_{i,j}\}$. *Moreover, the* $T_{i,j}$*'s admit a density* ρ *and there exist* ρ_0, ρ_1 *such that*

$$\forall t\in\mathcal{T}, \quad 0 < \rho_0 \le \rho(t) \le \rho_1 < +\infty. \tag{17}$$

Below, for squared integrable functions ϕ, ψ defined on $\mathcal{T}$, let $\langle \phi, \psi \rangle_\rho = \int_{\mathcal{T}} \phi(t)\psi(t)\rho(t)dt$ and $\|\phi\|_\rho^2 = \int_{\mathcal{T}} \phi^2(t)\rho(t)dt$. We define the estimator as the nonparametric least-squares estimation of g on $S_{\boldsymbol{m}}$ based on all the observations $(Y_{i,j}, T_{i,j})$:

$$\widehat{g}_{\boldsymbol{m}} = \arg\min_{h \in S_{\boldsymbol{m}}} \gamma_N(h), \quad \gamma_N(h) := \frac{1}{N}\sum_{i=1}^{N} \{Y_q - h(T_q)\}^2. \tag{18}$$

Let

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)^\top \in \mathbb{R}^N, \qquad \mathbf{T} = (T_1, T_2, \ldots, T_N)^\top \in \mathbb{R}^N,$$

and analogously, for $u_{i,j} = X_i(T_{i,j}) + \sigma(T_{i,j})\varepsilon_{i,j}$,

$$\mathbf{u} = (u_1, \ldots, u_N)^\top \in \mathbb{R}^N, \qquad \bar{\boldsymbol{\varepsilon}} = (\sigma_\varepsilon(T_1)\varepsilon_1, \ldots, \sigma_\varepsilon(T_N)\varepsilon_N)^\top \in \mathbb{R}^N. \tag{19}$$

For any function h defined on $\mathcal{T}$, let $h(\mathbf{T}) \in \mathbb{R}^N$ be the vector with components $h(T_{i,j})$.

For $\boldsymbol{m} \in \mathbb{N}_+^d$ with $D_{\boldsymbol{m}} \leq N$, and a family of orthonormal functions $\varphi_{\boldsymbol{k}}$, $\boldsymbol{k} \leq \boldsymbol{m}$, in $\mathbb{L}^2(\mathcal{T}, dt)$, define

$$\widehat{\Phi}_{\boldsymbol{m}} = [\varphi_{\boldsymbol{k}}(T_q)]_{1 \leq q \leq N, \boldsymbol{k} \leq \boldsymbol{m}} \in \mathcal{M}_{N,\boldsymbol{m}}(\mathbb{R}).$$

We also set

$$\widehat{\Psi}_{\boldsymbol{m}} = \frac{1}{N}\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}}, \quad \Psi_{\boldsymbol{m}} = \mathbb{E}\big(\widehat{\Psi}_{\boldsymbol{m}}\big) = (\langle \varphi_{\boldsymbol{k}}, \varphi_{\boldsymbol{k}'} \rangle_\rho)_{\boldsymbol{k}, \boldsymbol{k}' \leq \boldsymbol{m}}.$$

The empirical norm is

$$\|h\|_N^2 := \frac{1}{N}\sum_{q=1}^{N} h^2(T_q) = \frac{1}{N} h(\mathbf{T})^\top h(\mathbf{T}), \quad \text{and it holds} \quad \mathbb{E}(\|h\|_N^2) = \|h\|_\rho^2.$$

Under condition (17), it holds : $\forall x \in \mathbb{R}^{\boldsymbol{m}}$, for $h = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} x_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}$,

$$\rho_0 \|x\|_{\mathbb{R}^{\boldsymbol{m}}}^2 \leq x^\top \Psi_{\boldsymbol{m}} x = \|h\|_\rho^2 \leq \rho_1 \|x\|_{\mathbb{R}^{\boldsymbol{m}}}^2.$$

With the notation from Section 1, this implies

$$\|\Psi_{\boldsymbol{m}}\|_{\text{op}} \leq \rho_1, \quad \text{and} \quad \|\Psi_{\boldsymbol{m}}^{-1}\|_{\text{op}} \leq \frac{1}{\rho_0}. \tag{20}$$

In particular, under Condition (17), the matrix $\Psi_{\boldsymbol{m}}$ is invertible, and therefore so is the matrix $\widehat{\Psi}_{\boldsymbol{m}}$ with probability close to 1, for sufficiently large N.

The estimate $\widehat{g}_{\boldsymbol{m}}$ is defined as $\widehat{g}_{\boldsymbol{m}} = \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \widehat{a}_{\boldsymbol{k}} \varphi_{\boldsymbol{k}}$, with

$$\vec{a}^{(\boldsymbol{m})} := [\widehat{a}_{\boldsymbol{k}}]_{\boldsymbol{k} \leq \boldsymbol{m}} = \big(\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}}\big)^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{Y} = \widehat{\Psi}_{\boldsymbol{m}}^{-1}\Big(\frac{1}{N}\widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{Y}\Big) \in \mathbb{R}^{\boldsymbol{m}}.$$

Invertibility is not sufficient to ensure well-conditioned $\Psi_{\boldsymbol{m}}^{-1}$ and stable behavior for $\widehat{\Psi}_{\boldsymbol{m}}^{-1}$. [11] introduce a "security zone" through the following theoretical stability condition :

$$L(\boldsymbol{m})(\|\Psi_{\boldsymbol{m}}^{-1}\|_{\text{op}} \vee 1) \leq c_p \frac{N}{\log(N)}, \quad c_p = \frac{3\log(3/2) - 1}{2(p+1)}, \tag{21}$$

where p will be fixed later, and where

$$L(\boldsymbol{m}) := \sup_{t \in \mathcal{T}} \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \varphi_{\boldsymbol{k}}^2(t) \geq D_{\boldsymbol{m}} = \int_{\mathcal{T}} \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \varphi_{\boldsymbol{k}}^2(t) dt.$$

The last inequality holds with $\mathcal{T} = [0,1]^d$. The quantity $L(\boldsymbol{m})$ depends on the basis, and is finite if the basis functions are bounded. Under condition (17), and in view of (20), the stability condition (21) is fulfilled if

$$L(\boldsymbol{m}) \leq c_p \rho_0 \frac{N}{\log(N)}. \tag{22}$$

This is a weak condition on the maximal model dimension $\boldsymbol{m}$ that can be considered. For instance, if $L(\boldsymbol{m}) \asymp D_{\boldsymbol{m}}$, then it requires $D_{\boldsymbol{m}} \lesssim N/\log(N)$. The piecewise constant functions and the trigonometric functions are two examples with $L(\boldsymbol{m}) \asymp D_{\boldsymbol{m}}$; see [11].

Nevertheless, it is more convenient to define a data-driven version of this constraint, which does not require knowledge of ρ_0. Consider the event

$$\Lambda_{\boldsymbol{m}} := \left\{ L(\boldsymbol{m})(\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1) \leq 2c_p \frac{N}{\log(N)} \right\}, \tag{23}$$

with c_p as given in (21), and the factor 2 in front of it is added for technical reasons. We prove that, under (21), the empirical condition associated with $\Lambda_{\boldsymbol{m}}$ is satisfied with high probability. More precisely, using Chernoff's inequality as given in [35], we establish the following lemma.

Lemma 1. *Consider the event*

$$\Omega_{\boldsymbol{m}} := \left\{ \left| \frac{\|h\|_N^2}{\|h\|_\rho^2} - 1 \right| \leq \frac{1}{2}, \forall h \in S_{\boldsymbol{m}}, h \neq 0 \right\}. \tag{24}$$

If Assumption 2(iv′) and Condition (22) hold, then, for any $p \geq 0$,

$$\mathbb{P}(\Lambda_{\boldsymbol{m}}^c) \leq \mathbb{P}(\Omega_{\boldsymbol{m}}^c) \leq c\, N^{-p},$$

where $c > 0$ is a constant depending only on p.

On the set $\Omega_{\boldsymbol{m}}$, the empirical norm and the $\mathbb{L}^2(\mathcal{T}, \rho\, dt)$-norm $\|\cdot\|_\rho$ are equivalent for elements of $S_{\boldsymbol{m}}$.

3.2 Risk bound on a fixed model

Using the sample-based set $\Lambda_{\boldsymbol{m}}$, we define the data-driven estimator $\widetilde{g}_{\boldsymbol{m}} = \widehat{g}_{\boldsymbol{m}} \mathbf{1}_{\Lambda_{\boldsymbol{m}}}$ as a counterpart to the estimator defined in (18).

Proposition 3. *Assumption 2 holds. Under the Condition (22) (which implies (21)) and if $\mathbb{E}[g^4(T_{1,1})] < +\infty$, then*

$$\mathbb{E}(\|\widetilde{g}_{\boldsymbol{m}} - g\|_N^2) \leq \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_\rho^2 + \frac{2M_1\, D_{\boldsymbol{m}}}{N} + \frac{2M_2}{\rho_0} \frac{\sum_{i=1}^n N_i(N_i-1)}{N^2} + \|\sigma_\varepsilon^2\|_\infty \frac{D_{\boldsymbol{m}}}{N} + \frac{c}{N^p},$$

with $p \geq 1$ and c the constant from Lemma 1, $\sigma_X(u,v) = \mathbb{E}[X_1(u)X_1(v)]$, $M_1 = \sup_{\mathcal{T}} \sigma_X(u,u)$ and $M_2 = \int_{\mathcal{T}} \sigma_X(u,u)\rho^2(u)du$. A similar bound holds for $\mathbb{E}(\|\widetilde{g}_{\boldsymbol{m}} - g\|_\rho^2)$.

Remark 4. *By an alternative argument, the term $2M_1\, D_{\boldsymbol{m}}/N$ in the bound of Proposition 3 can be replaced by $2M_3 L(\boldsymbol{m})/(\rho_0 N)$, where $M_3 = \int_{\mathcal{T}} \sigma_X(u,u)\rho(u)\, du$. In the case of an orthonormal family $\{\varphi_{\boldsymbol{k}}\}$ supported on $[0,1]^d$, one has $L(\boldsymbol{m}) \geq D_{\boldsymbol{m}}$, so this alternative bound is no better in order. However, allowing families with larger support can lead to $L(\boldsymbol{m})$ of a much lower order than $D_{\boldsymbol{m}}$; the family of Hermite polynomials is one such example. See the proof of Proposition 3 for the details.*

Remark 5. *In the case where*

$$\frac{\sum_{i=1}^n N_i(N_i-1)}{N^2} \asymp \frac{1}{n}, \tag{25}$$

and if $L(\boldsymbol{m}) = D_{\boldsymbol{m}}$ *(e.g.,* $\{\varphi_{\boldsymbol{k}}\}$ *is the tensorized trigonometric basis), the bound of Proposition 3 reduces to*

$$\mathbb{E}(\|\widetilde{g}_{\boldsymbol{m}} - g\|_N^2) \lesssim \inf_{h \in S_{\boldsymbol{m}}} \|h-g\|_\rho^2 + (2M_1 + \|\sigma_\varepsilon^2\|_\infty)\frac{D_{\boldsymbol{m}}}{N} + \frac{2M_2}{\rho_0}\frac{1}{n} + \frac{c}{N^p},$$

and a similar bound holds for $\mathbb{E}(\|\widetilde{g}_{\boldsymbol{m}} - g\|_\rho^2)$*. Condition (25) is commonly imposed in the literature (see [43, Corollary 3.2]). Since* $\sum_{i=1}^n N_i/N^2 = 1/N \le 1/n$*, Condition (25) is equivalent to* $\sum_{i=1}^n N_i^2/N^2 \lesssim n^{-1}$*.*

3.3 Discussion about rates and optimality

As in Section 2.4, assume that g belongs to a regularity class of smoothness $\boldsymbol{\alpha} \in \mathbb{R}_+^d$ and radius R associated with the basis $\{\varphi_{\boldsymbol{k}}\}$, such that Condition (5) holds, and that $L(\boldsymbol{m}) \asymp D_{\boldsymbol{m}} = \prod_{l=1}^d m_l$. Moreover, we assume that Condition (25) holds. Then the optimization

$$\inf_{\{\boldsymbol{m}:\, D_{\boldsymbol{m}} \le N\}} \left(\sum_{j=1}^d m_j^{-2\alpha_j} + \frac{D_{\boldsymbol{m}}}{N} \right) + \frac{1}{n}$$

gives $\boldsymbol{m}_{\text{opt}}$ with components $m_{j,\text{opt}} = N^{\overline{\boldsymbol{\alpha}}/\{\alpha_j(2\overline{\boldsymbol{\alpha}}+d)\}}$, where $\overline{\boldsymbol{\alpha}}$ is as in (6) This leads to the rate-optimal upper bound

$$\mathbb{E}\big(\|\widehat{g}_{\boldsymbol{m}_{\text{opt}}} - g\|_\rho^2\big) \le C(\boldsymbol{\alpha}, R)\Big(N^{-\frac{2\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}} + n^{-1}\Big). \tag{26}$$

The optimal rate in the upper bound is attained, in particular, on the Sobolev spaces (8) introduced in Section 2.4. As pointed out there, for $g \in S(\boldsymbol{\alpha}, L)$, Inequality (9), which controls the bias term in the upper bound, holds true. Moreover, for the specific choice of the tensorized trigonometric basis, assuming that g belongs to a certain regularity class is equivalent to assuming that $g \in S(\boldsymbol{\alpha}, L)$ for some $\boldsymbol{\alpha}$ and L. Indeed, from the computations of [34, Proposition A.1] (considering special cases of type $\beta_l = \alpha_l$ and $\beta_{l'} = 0$ if $l' \neq l$, for all $1 \le l \le d$), we obtain that $S(\boldsymbol{\alpha}, L)$ coincides with the space

$$W^{\text{per}}(\boldsymbol{\alpha}, \mathcal{Q}) = \left\{ f = \sum_{\boldsymbol{k} \in \mathbb{N}_{>0}^d} a_{\boldsymbol{k}} \varphi_{\boldsymbol{k}} \in \mathbb{L}^2(\mathcal{T}, dt) : \left\| \frac{\partial^{\alpha_l} f}{\partial x_l^{\alpha_l}} \right\|^2 \le \mathcal{Q} \right\},$$

for some $\mathcal{Q}$ depending on $\boldsymbol{\alpha}$ and L (see also [36, Definition 1.11 and Proposition 1.14]). Using this alternative definition, the following result shows that the rate of the risk bound in (26) is optimal in a multidimensional and anisotropic setting; see also [41] for the case $d = 1$.

Theorem 2. *Let* $\{(X_i(t))_{t \in \mathcal{T}}\}$*,* $1 \le i \le n$*, be Gaussian random fields satisfying Assumption 1(i). Let* $\boldsymbol{\alpha} \in \mathbb{N}_{>0}^d$ *and* $L > 0$*. Then it holds*

$$\inf_{\widehat{g}} \sup_{g \in \mathcal{S}(\boldsymbol{\alpha}, L)} \mathbb{E}(\|\widehat{g} - g\|_\rho^2) \ge c\left(N^{-\frac{2\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}} + n^{-1}\right),$$

where $c > 0$ *is a positive constant and the infimum is taken over all estimators* $\widehat{g}$ *of* g *build from data generated as in (16).*

Remark 6. *The empirical norm we use is based on the simple average $N^{-1}\sum_{q=1}^{N}$, which assigns the same weight to every observation $(Y_{i,j}, T_{i,j})$. As a result, the bound in (26) depends on N, that is, on the product of the average of the N_i and the number n of curves. Alternatively, the empirical norm could be defined through the curve-weighted average $n^{-1}\sum_{i=1}^{n} N_i^{-1}\sum_{j=1}^{N_i}$, in which case the bound in (26) would depend on n times the harmonic mean of the N_i. Depending on the configuration of the N_i, these two bounds can be of very different orders of magnitude. This is precisely the comparison between the OBS (simple average) and SUBJ (curve-weighted average) schemes of [43], therein in dimension one and under known regularity. When the regularity is known, the comparison simplifies: it reduces to comparing the arithmetic or the harmonic mean of the N_i to a known power of n. Our model selection approach extends naturally to a weighted least-squares loss whose weights follow the OBS scheme, the SUBJ scheme, or a convex combination of the two. However, the OBS scheme is never dominated in rate, the SUBJ scheme improves upon it only by a constant factor in the variance, and only when both attain the parametric rate. Therefore, we focus in this paper on the empirical norm defined through the simple average $N^{-1}\sum_{q=1}^{N}$.*

Remark 7. *Our results require no regularity assumption on the fields X_i or on their covariance, in contrast to [9, 7]. There, the X_i are required to be smooth, as smooth as the mean in [9] (with $d = 1$), and Hölder continuous with Hölder-smooth covariance in [7]. Here, the smoothness is imposed solely on the target g and the X_i's may be rough, or even discontinuous.*

3.4 Isotropic model selection for independent random design

The data-driven choice of $\boldsymbol{m}$ lies at the core of model selection. In the common design case with J distinct points, the dimension $D_{\boldsymbol{m}}$ is bounded above by J, and letting $\boldsymbol{m}$ attain this bound corresponds to no regularization. With random design, there are N observed design points making the upper bound for $D_{\boldsymbol{m}}$ and the model much larger. Generally, it happens that the optimal $\boldsymbol{m}$ corresponds to a much smaller model, which makes the selection of $\boldsymbol{m}$ very useful. Here, like in Section 2.5, we restrict the study to isotropic spaces $S_{\boldsymbol{m}}$ with $\boldsymbol{m} = m\mathbf{1}_d \in \mathbb{N}_{>0}^d$, and $D_{\boldsymbol{m}} = m^d$. We define the collection of models by the set of multi-indices

$$\mathcal{M}_N = \left\{ \boldsymbol{m} = m\mathbf{1}_d : m \in \{1, \dots, N\}, \text{ such that } L(\boldsymbol{m})(\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1) \le c_p \frac{N}{\log(N)} \right\}, \quad (27)$$

and its empirical counterpart by the following set of multi-indices

$$\widehat{\mathcal{M}}_N = \left\{ \boldsymbol{m} = m\mathbf{1}_d : m \in \{1, \dots, N\}, \text{ such that } L(\boldsymbol{m})(\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1) \le 2c_p \frac{N}{\log(N)} \right\}, \quad (28)$$

and take $p = 5$. Next, we define

$$\widehat{\boldsymbol{m}} = \arg \min_{\boldsymbol{m} \in \widehat{\mathcal{M}}_N} \left\{ \gamma_N(\widehat{g}_{\boldsymbol{m}}) + \mathrm{pen}(\boldsymbol{m}) \right\}, \qquad \text{with } \gamma_N(h) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} \left\{ Y_{i,j} - h(T_{i,j}) \right\}^2 = \|h\|_N^2.$$

We study the risk of $\widehat{g}_{\widehat{\boldsymbol{m}}}$.

Theorem 3. *Assume that Assumption 2 holds, $\mathbb{E}[g^4(T_{1,1})] < \infty$, $\mathbb{E}(\varepsilon_{1,1}^4) < \infty$, and that*

1. *the collection of projection spaces $S_{\boldsymbol{m}}$ is nested and such that $L(\boldsymbol{m}) \le m^d$,*
2. *the fields X_i are separable and $\mathbb{E}\big(B_{n,X}^2\big) < \infty$, where $B_{n,X} := \max_{1 \le i \le n} \|X_i\|_\infty$.*

Then there exists $\kappa_0 > 0$ such that for all $\kappa \geq \kappa_0$ and with

$$\mathrm{pen}(\boldsymbol{m}) = \kappa \left(B_{n,X}^2 + \|\sigma_\varepsilon\|_\infty^2\right) \frac{m^d}{N}, \tag{29}$$

we have

$$\mathbb{E}\left(\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2\right) \leq C \inf_{\boldsymbol{m} \in \mathcal{M}_N} \left(\|g - g_{\boldsymbol{m}}\|_\rho^2 + \kappa \left\{ \mathbb{E}\left(B_{n,X}^2\right) + \|\sigma_\varepsilon\|_\infty^2 \right\} \frac{m^d}{N} \right) + C' \left(\frac{\sum_{i=1}^n N_i^2}{N^2} + \frac{\mathbb{E}\left(B_{n,X}^2\right) + \log^2(N)}{N} \right),$$

where C is a numerical positive constant and C' is a positive constant depending on ρ_0, $\mathbb{E}[g^4(T_{1,1})]$, $\mathbb{E}(\varepsilon_{1,1}^4)$, $\|\sigma_\varepsilon\|_\infty$ and M_3.

The separability of the X_i's is a "technical blanket assumption" commonly imposed in stochastic process theory to avoid measurability issues, see [1, p. 8] for this remark and the separability definition. Note that the penalty in (29) is random and depends on unknown quantities. See Section 2.5.2 and Remark 8 for possible estimators that are considered in Section 4. To understand the adaptive upper bound, we must investigate the order of the term $\mathbb{E}(B_{n,X}^2)$. For instance, if we suppose that $\|X_1\|_\infty$ (resp. $\|X_1\|_\infty^2$) is sub-Gaussian for some variance proxy σ_1^2 (resp. σ_2^2), Lemma 6 in the Appendix (see also [38, Lemma 2.2.2]) allows us to conclude that $\mathbb{E}(B_{n,X}^2)$ can be bounded by a constant multiplied by $\sigma_1^2 \log(n)$ (resp. $\sqrt{\sigma_2^2 \log(n)}$), provided $n \geq 2$. In that case, the adaptive estimator reaches the optimal rate up to a logarithmic loss. The sub-Gaussianity condition on the uniform norm is standard in nonparametric statistics [19, 10], Bayesian nonparametrics [37], and machine learning [30]. Fields with sub-Gaussian $\|X_1\|_\infty^2$ have a uniform norm with lighter *sub-Weibull* tail. See [39] for a study of such distributions and their application to deep neural networks. When $\|X_1\|_\infty \leq B$ *a.s.*, we have $\mathbb{E}(B_{n,X}^2) \leq B^2$ and the adaptive estimator attains the optimal rate.

Remark 8. *The random quantity $B_{n,X}$ appearing in the model selection criterion is unknown in practice, since the fields X_i are latent. We propose a crude estimate of $B_{n,X}$ via a two-stage, leave-one-field-out residual least-squares procedure. For each curve i, we first estimate the mean function g from the pooled observations of all other fields $\{(T_{l,j}, Y_{l,j}) : l \neq i\}$, by projecting onto the half-cosine basis $S_{\mathbf{m}}$ truncated at a fixed level $m = \lfloor N^{1/4} \rfloor$, yielding $\widehat{g}_{-i}$. We then form the residuals of curve i with respect to this leave-one-out mean estimate, $r_{i,j} = Y_{i,j} - \widehat{g}_{-i}(T_{i,j})$, $j = 1, \ldots, N_i$. In a second step, we regress these residuals on the same cosine system, now truncated at a field-specific level $m_i' = \lfloor N_i^{1/3} \rfloor$, giving the estimate $\widehat{X}_i$. We estimate the sup-norm X_i by*

$$\widehat{B}_i = \max_{j=1,\ldots,N_i} \left|\widehat{X}_i(T_{ij})\right|.$$

The construction ensures that the $r_{i,j}$ used to approximate X_i are not contaminated by X_i's own contribution to the mean estimate. Finally, $B_{n,X}$ is estimated by $\widehat{B}_{n,X} = \max_{1 \leq i \leq n} \widehat{B}_i$.

4 Numerical illustrations

We assess the finite-sample performance of the proposed model selection procedure in both the one-dimensional ($d = 1$) and two-dimensional ($d = 2$) settings, with both common and random design. The approximation spaces of the model collection are built from the half-cosine basis $\varphi_0(x) = 1$, $\varphi_k(x) = \sqrt{2}\cos(\pi k x)$, $x \in [0,1]$, $k = 1, 2, \ldots$, and its d-fold tensor product base functions $\varphi_{\boldsymbol{k}}(t) = \prod_{l=1}^d \varphi_{k_l}(t_l)$ with $\boldsymbol{k} = (k_1, \ldots, k_d)$, $t = (t_1, \ldots, t_d) \in \mathcal{T} = [0,1]^d$.

The true mean function g is defined, for $d = 2$, by

$$g(t_1, t_2) = 5\,\big|t_1 - 0.5\cos(2\pi t_2)\big|^{0.7} \exp\big(-0.7\,(t_1 t_2 - 0.5)^2\big) + t_1 t_2, \qquad t_1, t_2 \in [0, 1],$$

and, for $d = 1$, g is obtained by imposing $t_1 = t_2$. The two functions are plotted in Figure 1.

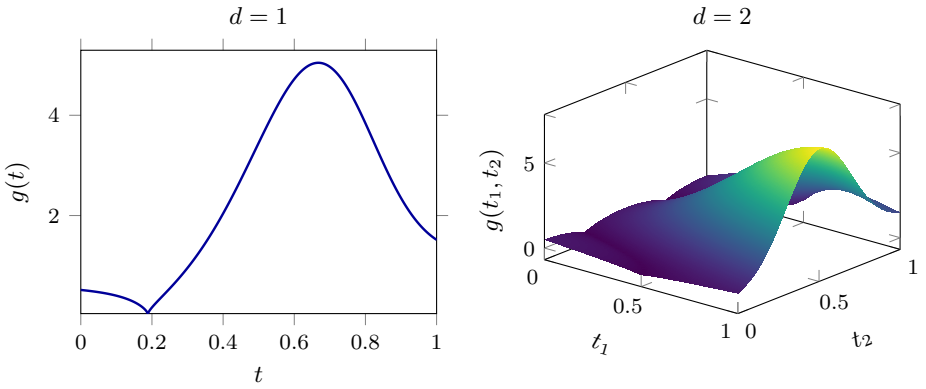

Figure 1: The true mean function g for $d = 1$ (left panel) and $d = 2$ (right panel).

We consider two models for the centered fields X_i. First, we generate the X_i's from a truncated Karhunen–Loève expansion. Specifically, for K a positive integer,

$$X_i(t) = \sum_{\boldsymbol{k} \le K\mathbf{1}_d} \zeta_{i,\boldsymbol{k}}\, \phi_{\boldsymbol{k}}(t), \quad \text{where} \quad \phi_{\boldsymbol{k}}(t) = \prod_{l=1}^{d} \cos\big(\pi(k_l - 1/2)\, t_l\big), \quad t \in [0, 1]^d,$$

and the coefficients are independent Gaussian, $\zeta_{i,\boldsymbol{k}} \sim \mathcal{N}\big(0,\, D_{\boldsymbol{k}}^{-1-4/d}\big)$, $D_{\boldsymbol{k}} = \prod_{l=1}^{d} k_l$. We also consider a setting where the X_i are piecewise-constant random fields, with jumps across randomly oriented hyperplanes. Given a positive integer M, for each i we draw i.i.d. directions $a_{i,1}, \ldots, a_{i,M}$ uniformly on $\mathbb{S}^{d-1}$ ($d \ge 1$) and offsets $b_{i,1}, \ldots, b_{i,M}$, where each $b_{i,p}$ is uniform over the projected domain $[\min_{t\in\mathcal{T}} a_{i,p}^\top t,\ \max_{t\in\mathcal{T}} a_{i,p}^\top t]$. For each i, these define a random partition of $\mathcal{T}$, and the field X_i is obtained as

$$X_i(t) = \sum_{p=1}^{M} \delta_{i,p} \mathbf{1}\{a_{i,p}^\top t \le b_{i,p}\}, \qquad t \in \mathcal{T}, \tag{30}$$

with independent jump sizes $\delta_{i,p} \sim \mathcal{N}(0, 1)$. In the random design case, the N_i are drawn from a Poisson distribution with parameter N_{mean} with any zero draw replaced by 1, and each component of $T_{i,j}$ is independently drawn from a Beta(q, q) distribution. With common design, all the N_i are equal to J. The noise variables $\varepsilon_{i,j}$ are drawn independently from a standard normal distribution. This means we consider the homoscedastic case with $\sigma_\varepsilon = 1$.

All estimators are evaluated on a regular grid $\{s_1, \ldots, s_{L^d}\} \subset \mathcal{T}$ obtained from L equally spaced points per coordinate. For each mean function estimator $\widehat{g}$, we consider the integrated squared error (ISE), approximated by the Riemann sum $\mathrm{ISE}(\widehat{g}) = L^{-d} \sum_{\ell=1}^{L^d} \big\{\widehat{g}(s_\ell) - g(s_\ell)\big\}^2$. We report the empirical mean integrated squared error (MISE), obtained by averaging $\mathrm{ISE}(\widehat{g})$ over the B Monte Carlo replications $\widehat{g}^{(1)}, \ldots, \widehat{g}^{(B)}$, that is,

$$\mathrm{MISE}(\widehat{g}) = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{L^d} \sum_{\ell=1}^{L^d} \big\{\widehat{g}^{(b)}(s_\ell) - g(s_\ell)\big\}^2.$$

In our experiments we consider $B = 1000$ for $d = 1$ and $B = 500$ for $d = 2$, $K = 20$, $L = 201$ and $M = 8$. Moreover, $N_{\text{mean}} \in \{10, 25, 50\}$ with random design, $J \in \{49, 100, 169\}$ with

common design. The Beta distribution parameters used to draw the random design $T_{i,j}$ are $q \in \{1, 2\}$. Finally, $n \in \{50, 100\}$ for random design and $n \in \{100, 200\}$ for common design.

The model selection penalty depends on several quantities that have to be estimated or fixed in practice. Throughout the simulations, the constant $B_{n,X}$ is estimated by the procedure described in Remark 8, and the noise level is estimated as proposed in Section 2.5.2. In the following we describe a simple simulation-based procedure which was used to determine κ.

4.1 Data-driven selection of κ

Here, to calibrate a value of κ, we use simulated data generated according to the model under study, with some arbitrary fixed mean functions g^* that are plotted in Figure 2. We select the value of κ providing the lowest MISE over the collection of model, with the MISE averaged over 300 simulated samples on a grid of 100 values of κ in a compact interval. The calibration procedure is fast and the selected value of κ is expected to be satisfactory regardless of the true mean function and the distribution of the random fields. We consider the four settings determined by the dimension $d \in \{1, 2\}$ and the type of design (common and random). We allow for anisotropic projection spaces $S_{\boldsymbol{m}}$ with $d = 2$. In the settings with random design, we generated data points $(Y_{i,j}, T_{i,j})$ using $n = 100$ and $N_i \sim \text{Poisson}(100)$ i.i.d. design points with uniformly distributed components, and random fields generated according to (30). In the settings with common design, we consider $J = 100$. The results of this preliminary calibration of the value of κ are presented in Table 1. We notice the much smaller selected values of κ compared to the theoretical bound discussed in Remark 3. The illustration of the calibration results with fixed design is presented in Figure 3. We note that the MISE is quite insensitive to the choice of κ. The selected dimension is more dispersed at small κ and concentrates as κ grows. We also refer to [6] for other data-driven approaches for calibrating κ from simulated data, namely the slope heuristics and the jump dimension strategy.

Table 1: Data-driven selection of the κ parameter

	Design	
Dimension	Common	Random
$d = 1$	$\kappa = 0.62$	$\kappa = 0.63$
$d = 2$	$\kappa = 2.57$	$\kappa = 0.29$

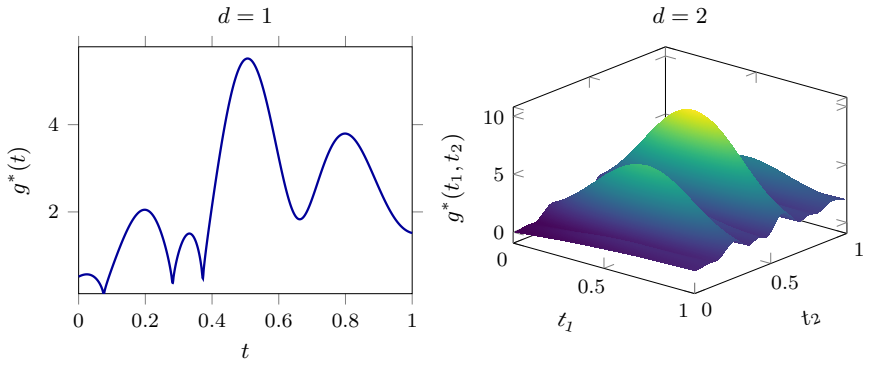

Figure 2: The mean function g^* used for calibrating κ for $d = 1$ (left) and $d = 2$ (right).

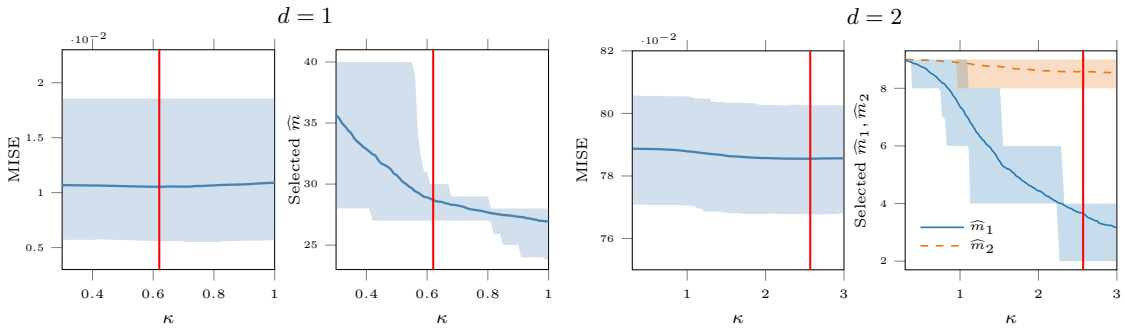

Figure 3: Calibration of κ (fixed design, $d \in \{1, 2\}$). Panels 1 and 3: MISE of the model-selection estimator against κ. Panels 2 and 4: mean selected dimension against κ. The coordinate-wise dimensions $\widehat{m}_1, \widehat{m}_2$ for $d = 2$ may differ (Panel 4). In each panel the line is the mean over the 300 repetitions and the shaded band gives the pointwise 10%–90% quantiles. The vertical line marks the value of κ minimising the mean MISE.

4.2 Simulation results

The purpose of this first simulation study, carried out using `R`, is twofold. First, we provide evidence that our model selection approach performs well relative to standard nonparametric benchmarks with publicly available implementations, such as data-driven kernel or spline-based estimators and the interpolation estimator. Second, we compare our data-driven implementation of model selection with the oracle choice, defined as the model minimizing the MISE over the collection of models we consider. The list of all comparisons is given in Table 2.

Table 2: Competitors by setting: OR (Oracle), SP (Splines), LL (Local linear), Int (Interpolation)

	Design	
Dimension	Common	Random
$d = 1$	OR, SP, Int	OR, SP, LL
$d = 2$	OR, SP, Int	OR, SP

The penalized-spline estimator, denoted $\widehat{g}_{\mathrm{SP}}$, is obtained using the `mgcv` package applied to the pooled sample of (T_{ij}, Y_{ij})'s, with tensor-product splines for $d = 2$ and a generalized cross-validation (GCV)-selected smoothing parameter. The local linear estimator $\widehat{g}_{\mathrm{LL}}$ of the mean function (see [43]), with cross-validated bandwidth selection, is computed using the `fdapace` package. In the common design case, all estimators are computed using the average responses $\bar{Y}_{\bullet,j} = n^{-1} \sum_{i=1}^{n} Y_{ij}$, $j = 1, \ldots, J$, at each design point. The interpolation estimator of the pairs $(T_j, \bar{Y}_{\bullet,j})$, denoted $\widehat{g}_{\mathrm{Int}}$, is piecewise linear for $d = 1$ and is obtained by Delaunay triangulation for $d = 2$, computed using the `akima` package.

The performance comparison is done using the log-ratios

$$R_{\mathrm{Comp}} = \log\Big(\frac{\mathrm{MISE}(\widehat{g}_{\mathrm{MS}})}{\mathrm{MISE}(\widehat{g}_{\mathrm{Comp}})}\Big), \qquad \text{with} \quad \mathrm{Comp} \in \{\mathrm{OR}, \mathrm{SP}, \mathrm{LL}, \mathrm{Int}\}. \tag{31}$$

With this convention, negative values indicate that $\widehat{g}_{\mathrm{MS}}$ achieves a smaller MISE and thus outperforms the benchmark, and values near zero indicate comparable performance.

Across the four settings (Figures 4 to 7), the data-driven estimator $\widehat{g}_{\mathrm{MS}}$ tracks the oracle closely: the log-ratio R_{OR} stays small and positive throughout, growing only mildly in the harder random design and $d = 2$ regimes. Since the oracle is, by construction, the best model in our collection, it appears that the simple error variance estimator and the simulation-based

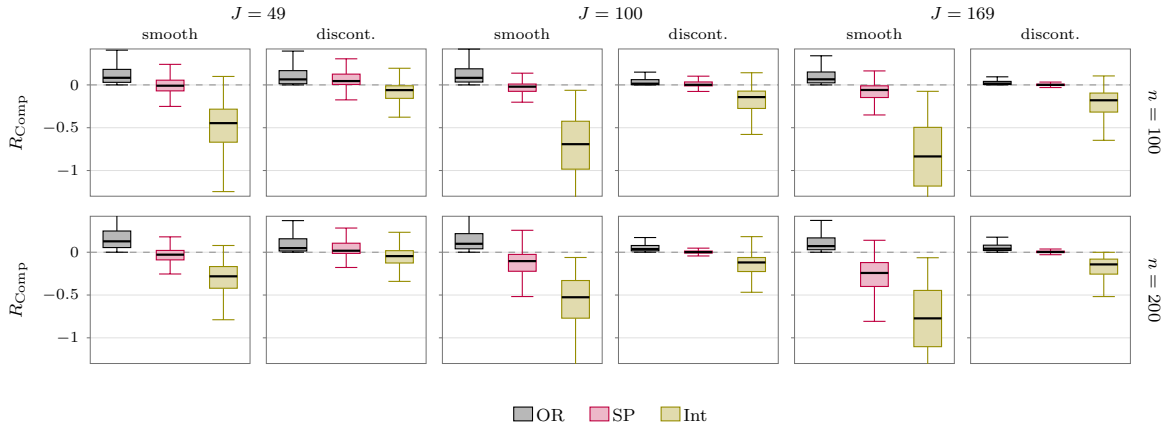

Figure 4: Common design, $d = 1$. Boxplots of R_{Comp} from (31) over $B = 1000$ Monte Carlo replications, for Comp $\in \{\text{OR}, \text{SP}, \text{Int}\}$. Columns vary the number of common design points $J \in \{49, 100, 169\}$ and the curve type (smooth/discontinuous); rows vary the number of curves $n \in \{100, 200\}$.

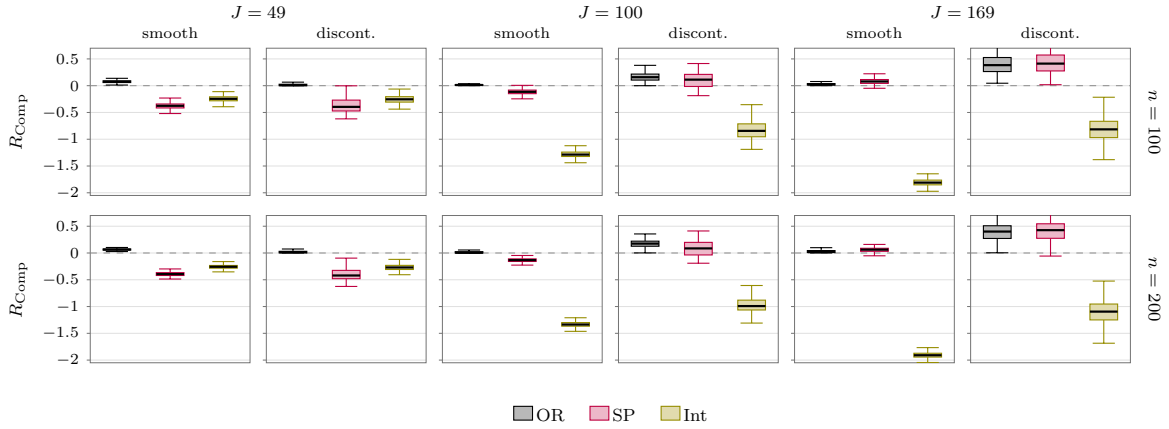

Figure 5: Common design, $d = 2$. As in Figure 4, over 500 replications, with tensor-product splines (SP) and Delaunay interpolation (Int).

rule of thumb for κ already allow us to recover near-oracle behaviour. Against the off-the-shelf competitors, $\widehat{g}_{\text{MS}}$ is at least as accurate and often markedly more so: interpolation is dominated by a wide margin (most visibly for $d = 2$ and discontinuous fields), local linear is consistently outperformed, and penalized splines are matched, with R_{SP} concentrated around zero. In a few cells, most notably random design and $d = 2$, $\widehat{g}_{\text{SP}}$ edges out even the oracle: as the oracle is optimal only within our collection, this suggests that the choice of approximating basis is itself worth optimizing, a direction we leave to future work. Overall, model selection is effective and computationally fast in both $d = 1$ and $d = 2$. The unknown quantities entering the penalty are estimated here through a deliberately simple device whose performance is already reasonably good. Sharpening these estimates is a natural avenue for improvement.

5 Proofs

Below, for any symmetric matrices $\mathbf{A}, \mathbf{B}$ of the same dimension, $\mathbf{A} \preceq \mathbf{B}$ (resp. $\mathbf{A} \succeq \mathbf{B}$) denotes the Loewner order and means that $\mathbf{B} - \mathbf{A}$ (resp. $\mathbf{A} - \mathbf{B}$) is positive semi-definite.

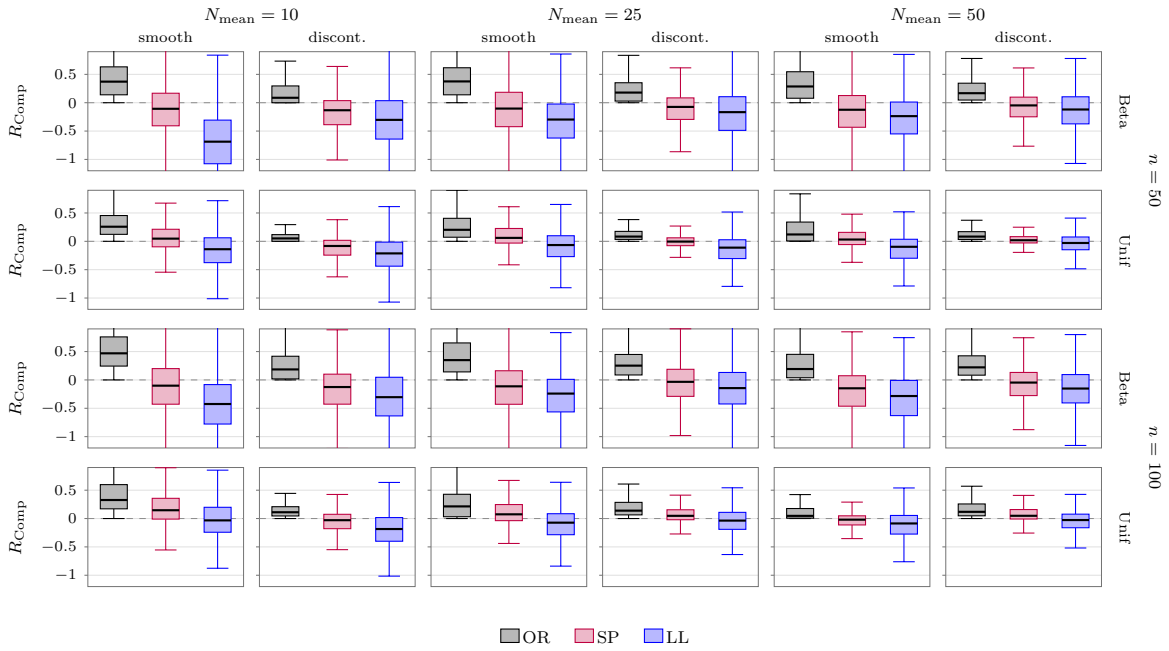

Figure 6: Random design, $d = 1$. Boxplots of R_{Comp} from (31) over $B = 1000$ Monte Carlo replications, for Comp $\in \{\text{OR}, \text{SP}, \text{LL}\}$. Columns vary the average number of points per curve $N_{\text{mean}} \in \{10, 25, 50\}$ and the curve type (smooth / discontinuous); rows vary the number of curves $n \in \{50, 100\}$ and the sampling design (Beta/Uniform).

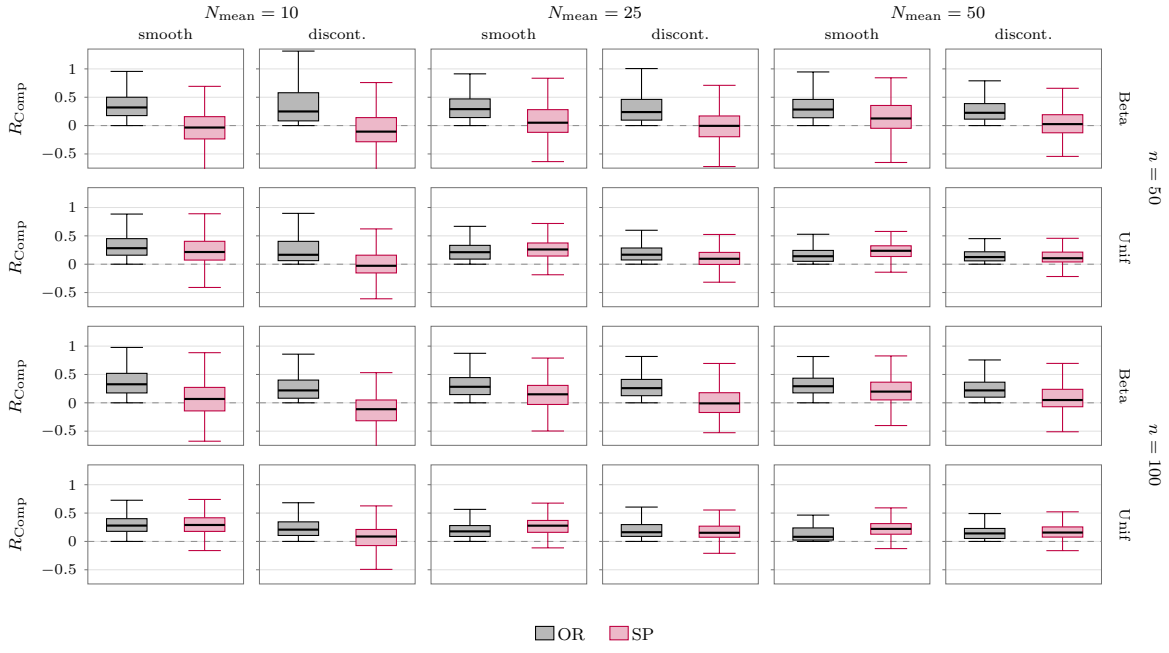

Figure 7: Random design, $d = 2$. As in Figure 6, over 500 replications, with Comp $\in \{\text{OR}, \text{SP}\}$ (no local-linear competitor).

5.1 Proofs for the common design case

Proof of Proposition 1. Let $\Pi_{\boldsymbol{m}} g \in S_{\boldsymbol{m}}$ be such that $\Pi_{\boldsymbol{m}} g(\mathbf{T})$ is the vector of $\mathbb{R}^J$ corresponding to the orthogonal projection of $g(\mathbf{T}) := (g(T_1), \ldots, g(T_J))^\top$ in the subspace of $\mathbb{R}^J$ spanned

by $\varphi_{\boldsymbol{k}}(\mathbf{T})$, for $\boldsymbol{k} \leq \boldsymbol{m}$, and $\mathbf{T} = (T_1, \dots, T_J)^\top \in \mathcal{M}_{J,d}([0,1])$. The coefficients of $\Pi_{\boldsymbol{m}} g$ are given by $(\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top g(\mathbf{T}) \in \mathcal{M}_{\boldsymbol{m},1}(\mathbb{R})$ and $\Pi_{\boldsymbol{m}} g(\mathbf{T}) = \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top g(\mathbf{T})$. Note that

$$\widehat{g}_{\boldsymbol{m}}(\mathbf{T}) - \Pi_{\boldsymbol{m}} g(\mathbf{T}) = \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \mathbf{u},$$

where $\mathbf{u} = (u_1, \dots, u_J)^\top$ with

$$u_j = \bar{\eta}_j + \sigma_\varepsilon(T_j)\, \bar{\varepsilon}_{\bullet,j}, \qquad \bar{\eta}_j = \frac{1}{n} \sum_{i=1}^n X_i(T_j), \qquad 1 \leq j \leq J.$$

and $\bar{\varepsilon}_{\bullet,j}$ defined in (2). By Pythagoras, we get

$$\|\widehat{g}_{\boldsymbol{m}} - g\|_J^2 = \|\Pi_{\boldsymbol{m}} g - g\|_J^2 + \|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_J^2.$$

The bias term is given by

$$\|\Pi_{\boldsymbol{m}} g - g\|_J^2 = \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_J^2,$$

and we need to upper bound the variance term:

$$\mathbb{E}\left(\|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_J^2\right) = \frac{1}{J} \mathbb{E}\left(\mathbf{u}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \mathbf{u}\right).$$

Let $\bar{\boldsymbol{\eta}} = (\bar{\eta}_1, \dots, \bar{\eta}_J)^\top \in \mathbb{R}^J$, $\bar{\boldsymbol{\varepsilon}} = (\bar{\varepsilon}_{\bullet,1}, \dots, \bar{\varepsilon}_{\bullet,J})^\top \in \mathbb{R}^J$ and

$$\bar{\boldsymbol{\epsilon}} = \left(\sigma_\varepsilon(T_1)\bar{\varepsilon}_{\bullet,1}, \dots, \sigma_\varepsilon(T_J)\bar{\varepsilon}_{\bullet,J}\right)^\top = \mathrm{diag}(\sigma_\varepsilon(T_1), \dots, \sigma_\varepsilon(T_J))\bar{\boldsymbol{\varepsilon}} =: \Delta_{\sigma_\varepsilon} \bar{\boldsymbol{\varepsilon}}.$$

As the $\varepsilon_{i,j}$ are independent of the X_i and centered, we get

$$\begin{aligned}\mathbb{E}\left(\mathbf{u}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \mathbf{u}\right) &= \mathbb{E}\left(\bar{\boldsymbol{\eta}}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \bar{\boldsymbol{\eta}}\right) \\ &\quad + \mathbb{E}\left(\bar{\boldsymbol{\epsilon}}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \bar{\boldsymbol{\epsilon}}\right) =: \mathbb{V}_1 + \mathbb{V}_2.\end{aligned}$$

It follows that

$$\mathbb{E}\left(\|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_J^2\right) = \frac{1}{J}\left(\mathbb{V}_1 + \mathbb{V}_2\right). \tag{32}$$

We start with computing $\mathbb{V}_2$. First note that $\Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \in \mathcal{M}_{J,J}(\mathbb{R})$ is a real projection matrix onto a space of dimension $D_{\boldsymbol{m}}$. Recall that for any symmetric real matrices A and B with respective eigenvalues $\lambda_1 \geq \ldots \geq \lambda_J \geq 0$ and $\mu_1 \geq \ldots \geq \mu_J \geq 0$ it holds that $\mathrm{Tr}(AB) \leq \sum_{j=1}^J \lambda_j \mu_j$. See [22, Theorem 4.3.53]. Therefore, we get for $A = \mathbb{E}[\bar{\boldsymbol{\epsilon}} \bar{\boldsymbol{\epsilon}}^\top]$ and $B = \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top$

$$\begin{aligned}\mathbb{V}_2 &= \mathbb{E}\left[\mathrm{Tr}\left(\bar{\boldsymbol{\epsilon}}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top \bar{\boldsymbol{\epsilon}}\right)\right] = \mathbb{E}\left[\mathrm{Tr}\left(\bar{\boldsymbol{\epsilon}} \bar{\boldsymbol{\epsilon}}^\top \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top\right)\right] \\ &= \mathrm{Tr}\left(\mathbb{E}[\bar{\boldsymbol{\epsilon}} \bar{\boldsymbol{\epsilon}}^\top] \Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top\right) \leq \frac{\|\sigma_\varepsilon^2\|_\infty}{n} \sum_{j=1}^J \mu_j = \|\sigma_\varepsilon^2\|_\infty \frac{D_{\boldsymbol{m}}}{n},\end{aligned} \tag{33}$$

where we used that $\Phi_{\boldsymbol{m}} (\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1} \Phi_{\boldsymbol{m}}^\top$ is a projection matrix of rank $D_{\boldsymbol{m}}$ with eigenvalues μ_j in $\{0, 1\}$ and that

$$\mathbb{E}[\bar{\boldsymbol{\epsilon}} \bar{\boldsymbol{\epsilon}}^\top] = \mathrm{diag}\left(\sigma_\varepsilon^2(T_1)\mathbb{E}[\bar{\varepsilon}_{\bullet,1}^2], \dots, \sigma_\varepsilon^2(T_J)\mathbb{E}[\bar{\varepsilon}_{\bullet,J}^2]\right) = \frac{1}{n}\mathrm{diag}\left(\sigma_\varepsilon^2(T_1), \dots, \sigma_\varepsilon^2(T_J)\right).$$

For $\mathbb{V}_1$, we have

$$\mathbb{E}(\bar{\eta}_j \bar{\eta}_\ell) = \frac{1}{n^2} \sum_{i=1}^n \sum_{i'=1}^n \mathrm{Cov}(X_i(T_j), X_{i'}(T_\ell)) = \frac{1}{n} \mathrm{Cov}(X_1(T_j), X_1(T_\ell)),$$

so that

$$\mathbb{V}_1 = \frac{1}{n}\mathbb{E}\Big[X_1(\mathbf{T})^\top \Phi_{\boldsymbol{m}}(\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1}\Phi_{\boldsymbol{m}}^\top X_1(\mathbf{T})\Big] \leq \frac{1}{n}\|\Phi_{\boldsymbol{m}}(\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1}\Phi_{\boldsymbol{m}}^\top\|_{\mathrm{op}}\mathbb{E}\left[\sum_{j=1}^J X_1^2(T_j)\right].$$

As $\Phi_{\boldsymbol{m}}(\Phi_{\boldsymbol{m}}^\top \Phi_{\boldsymbol{m}})^{-1}\Phi_{\boldsymbol{m}}^\top$ is a (symmetric) projection matrix, its operator norm is 1. We get

$$\mathbb{V}_1 \leq \frac{J}{n}\mathbb{E}\left[\frac{1}{J}\sum_{j=1}^J X_1^2(T_j)\right] \leq \frac{J}{n}\mathfrak{C}_X. \tag{34}$$

Plugging (33) and (34) in (32) gives the variance bound, and ends the proof. □

Proof of Theorem 1. Since

$$\gamma_J(h) - \gamma_J(g) = \|g-h\|_J^2 + \frac{2}{J}\sum_{j=1}^J \{g(T_j) - h(T_j)\}u_j,$$

the relationship

$$\forall \boldsymbol{m} \in \mathcal{M}_J, \quad \gamma_J(\widehat{g}_{\widehat{\boldsymbol{m}}}) + \mathrm{pen}(\widehat{\boldsymbol{m}}) \leq \gamma_J(g_{\boldsymbol{m}}) + \mathrm{pen}(\boldsymbol{m}),$$

together with the fact that the spaces $S_{\boldsymbol{m}}$ are nested in the isotropic case, implying $S_{\boldsymbol{m}} + S_{\boldsymbol{m}'} = S_{\boldsymbol{m}\vee\boldsymbol{m}'}$, leads to

$$\begin{aligned}
\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2 &\leq \|g_{\boldsymbol{m}} - g\|_J^2 + \mathrm{pen}(\boldsymbol{m}) + \frac{2}{J}\sum_{j=1}^J (\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}})(T_j)(\bar{\eta}_j + \sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}) - \mathrm{pen}(\widehat{\boldsymbol{m}}) \\
&\leq \|g_{\boldsymbol{m}} - g\|_J^2 + \mathrm{pen}(\boldsymbol{m}) + \frac{1}{8}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_J^2 + \frac{8}{J}\sum_{j=1}^J \bar{\eta}_j^2 \\
&\quad + \frac{1}{8}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_J^2 + 8\sup_{h\in S_{\widehat{\boldsymbol{m}}\vee\boldsymbol{m}}, \|h\|_J=1}\left(\frac{1}{J}\sum_{j=1}^J h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\right)^2 - \mathrm{pen}(\widehat{\boldsymbol{m}}).
\end{aligned}$$

Since $\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_J^2 \leq 2(\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2 + \|g_{\boldsymbol{m}} - g\|_J^2)$, we get

$$\begin{aligned}
\frac{1}{2}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2 &\leq \frac{3}{2}\|g_{\boldsymbol{m}} - g\|_J^2 + \mathrm{pen}(\boldsymbol{m}) \\
&\quad + 8\sup_{h\in S_{\widehat{\boldsymbol{m}}\vee\boldsymbol{m}}, \|h\|_J=1}\left(\frac{1}{J}\sum_{j=1}^J h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\right)^2 - \mathrm{pen}(\widehat{\boldsymbol{m}}) + \frac{8}{J}\sum_{j=1}^J \bar{\eta}_j^2.
\end{aligned}$$

Then,

$$\mathbb{E}(\bar{\eta}_j^2) = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i(T_j)\right)^2\right] = \frac{1}{n}\mathbb{E}[X_1^2(T_j)],$$

and thus

$$\mathbb{E}\left(\frac{8}{J}\sum_{j=1}^J \bar{\eta}_j^2\right) \leq \frac{8}{n}\sup_{t\in\mathcal{T}}\mathbb{E}[X_1^2(t)] = \frac{8}{n}\mathfrak{C}_X.$$

For any quantity $p(\boldsymbol{m}, \widehat{\boldsymbol{m}})$, we obtain

$$\mathbb{E}(\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2) \le 3\|g_{\boldsymbol{m}} - g\|_J^2 + 2\mathrm{pen}(\boldsymbol{m}) + 16\mathbb{E}\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_J = 1} \left(\frac{1}{J}\sum_{j=1}^{J} h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\right)^2 - p(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+ + \mathbb{E}\big[16p(\boldsymbol{m}, \widehat{\boldsymbol{m}}) - 2\mathrm{pen}(\widehat{\boldsymbol{m}})\big] + 16\frac{\mathfrak{C}_X}{n}. \quad (35)$$

Next, use Lemma 2 with $p(\boldsymbol{m}, \boldsymbol{m}')$ as in (36). Then, $16p(\boldsymbol{m}, \boldsymbol{m}') \le 2\mathrm{pen}(\boldsymbol{m}) + 2\mathrm{pen}(\boldsymbol{m}')$ if $\kappa \ge 8$. Plugging this in (35), for $\kappa \ge 64$ we get

$$\mathbb{E}(\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_J^2) \le 3\|g_{\boldsymbol{m}} - g\|_J^2 + 4\mathrm{pen}(\boldsymbol{m}) + \frac{C\|\sigma_\varepsilon^2\|_\infty + 16\mathfrak{C}_X}{n}.$$

Taking $C' = \max(C, 16)$, the proof of Theorem 1 is complete. □

Lemma 2. *Let $\boldsymbol{m} = m\mathbf{1}_d$ and $\boldsymbol{m}' = m'\mathbf{1}_d$, $m, m' \in \mathbb{N}^*$, so that $\boldsymbol{m} \vee \boldsymbol{m}' = (m \vee m')\mathbf{1}_d$. Let*

$$p(\boldsymbol{m}, \boldsymbol{m}') = 8\|\sigma_\varepsilon^2\|_\infty \frac{(m \vee m')^d}{nJ}. \quad (36)$$

There exists a constant $C > 0$, depending on $\|\sigma_\varepsilon^2\|_\infty$ but not n, J such that

$$\mathbb{E}\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_J = 1} \left(\frac{1}{J}\sum_{j=1}^{J} h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\right)^2 - p(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+ \le C\frac{\|\sigma_\varepsilon^2\|_\infty}{n}.$$

Proof of Lemma 2. Define the empirical process

$$\nu_J(h) = \frac{1}{J}\sum_{j=1}^{J} h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}.$$

and split it as $\nu_J(h) = \nu_J^{(1)}(h) + \nu_J^{(2)}(h)$ where $\bar{\varepsilon}_{\bullet,j} := \bar{\varepsilon}_{\bullet,j}^{(1)} + \bar{\varepsilon}_{\bullet,j}^{(2)}$ and

$$\nu_J^{(1)}(h) = \frac{1}{J}\sum_{j=1}^{J} h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}^{(1)}, \qquad \bar{\varepsilon}_{\bullet,j}^{(1)} = \bar{\varepsilon}_{\bullet,j}\mathbf{1}_{\{|\bar{\varepsilon}_{\bullet,j}| \le B_{n,J}\}} - \mathbb{E}\left(\bar{\varepsilon}_{\bullet,j}\mathbf{1}_{\{|\bar{\varepsilon}_{\bullet,j}| \le B_{n,J}\}}\right),$$

with $\nu_J^{(2)}$ accordingly. The truncation threshold $B_{n,J}$ will be determined below. Then

$$\begin{aligned}
&\mathbb{E}\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_J = 1} \left(\frac{1}{J}\sum_{j=1}^{J} h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\right)^2 - p(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+ \\
&\le 2\mathbb{E}\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_J = 1} [\nu_J^{(2)}(h)]^2\right) + 2\mathbb{E}\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_J = 1} [\nu_J^{(1)}(h)]^2 - \frac{1}{2}p(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+ \\
&=: 2\mathbf{T}_1 + 2\mathbf{T}_2.
\end{aligned}$$

For $\|h\|_J = 1$,

$$[\nu_J^{(2)}(h)]^2 \le \left(\frac{1}{J}\sum_{j=1}^{J} h^2(T_j)\right)\left(\frac{1}{J}\sum_{j=1}^{J} \sigma_\varepsilon^2(T_j)[\bar{\varepsilon}_{\bullet,j}^{(2)}]^2\right) = \frac{1}{J}\sum_{j=1}^{J} \sigma_\varepsilon^2(T_j)[\bar{\varepsilon}_{\bullet,j}^{(2)}]^2.$$

Thus

$$\mathbf{T}_1 \leq \|\sigma_\varepsilon^2\|_\infty \mathbb{E}\bigg(\frac{1}{J}\sum_{j=1}^{J}[\bar{\varepsilon}^{(2)}_{\bullet,j}]^2\bigg) = \|\sigma_\varepsilon^2\|_\infty \mathbb{E}\big([\bar{\varepsilon}^{(2)}_{\bullet,1}]^2\big).$$

As by Markov Inequality, $\mathbb{E}(\bar{\varepsilon}^2_{\bullet,1}\mathbf{1}_{\{|\bar{\varepsilon}_{\bullet,1}|>B_{n,J}\}}) \leq \mathbb{E}\big(\bar{\varepsilon}^4_{\bullet,1}B^{-2}_{n,J}\big)$, we get

$$\mathbf{T}_1 \leq \|\sigma_\varepsilon^2\|_\infty \mathbb{E}(\bar{\varepsilon}^2_{\bullet,1}\mathbf{1}_{\{|\bar{\varepsilon}_{\bullet,1}|>B_{n,J}\}}) \leq \|\sigma_\varepsilon^2\|_\infty \bigg(\frac{1}{n} \wedge C\frac{\mathbb{E}(\bar{\varepsilon}^4_{1,1})}{n^2 B^2_{n,J}}\bigg), \tag{37}$$

where the second bound follows from Rosenthal Inequality. This second term will be useful in the random design setting. Here, with the common design, we use

$$\mathbf{T}_1 \leq \frac{\|\sigma_\varepsilon^2\|_\infty}{n}.$$

Secondly, for the bounded $\bar{\varepsilon}^{(1)}_{\bullet,j}$ we write that

$$\mathbf{T}_2 \leq \sum_{\boldsymbol{m}'\in\mathcal{M}_J} \mathbb{E}\bigg(\sup_{h\in S_{\boldsymbol{m}\vee\boldsymbol{m}'},\|h\|_J=1} [\nu^{(1)}_J(h)]^2 - \frac{1}{2}p(\boldsymbol{m},\boldsymbol{m}')\bigg)_+.$$

Now, we apply the Talagrand Inequality, as recalled in Theorem 5 in Section 6, to the term inside the sum. With $\boldsymbol{m}^\star = \boldsymbol{m}\vee\boldsymbol{m}'$, we compute the terms denoted by $\mathbb{M}$, $\mathbb{H}$ and v in Theorem 5. We decompose $h \in S_{\boldsymbol{m}^\star}$ in a basis $(\theta_{\boldsymbol{k}})_{\boldsymbol{k}\leq\boldsymbol{m}^\star}$ orthonormal with respect to the empirical scalar product, $h = \sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star} b_{\boldsymbol{k}}\theta_{\boldsymbol{k}}$, $\|h\|_J^2 = \sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star} b^2_{\boldsymbol{k}}$, $\langle\theta_{\boldsymbol{k}},\theta_{\boldsymbol{k}'}\rangle_J = \delta_{\boldsymbol{k},\boldsymbol{k}'}$. (In these writings, $\boldsymbol{k}$ and $\boldsymbol{k}'$ can have different components.) Then,

$$\begin{aligned}
\mathbb{E}\bigg[\sup_{h\in S_{\boldsymbol{m}^\star},\|h\|_J=1}\bigg(\frac{1}{J}\sum_{j=1}^{J}h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}^{(1)}_{\bullet,j}\bigg)^2\bigg] &\leq \sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star}\mathbb{E}\bigg[\bigg(\frac{1}{J}\sum_{j=1}^{J}\theta_{\boldsymbol{k}}(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}^{(1)}_{\bullet,j}\bigg)^2\bigg] \\
&= \sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star}\mathbb{E}\bigg[\frac{1}{J^2}\sum_{j=1}^{J}\sum_{\ell=1}^{J}\theta_{\boldsymbol{k}}(T_j)\theta_{\boldsymbol{k}}(T_\ell)\sigma_\varepsilon(T_j)\bar{\varepsilon}^{(1)}_{\bullet,j}\sigma_\varepsilon(T_\ell)\bar{\varepsilon}^{(1)}_{\bullet,\ell}\bigg] \\
&\leq \sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star}\mathbb{E}\bigg[\frac{1}{J^2}\sum_{j=1}^{J}\sum_{\ell=1}^{J}\theta_{\boldsymbol{k}}(T_j)\theta_{\boldsymbol{k}}(T_\ell)\sigma_\varepsilon(T_j)\sigma_\varepsilon(T_\ell)\{n^{-1}\delta_{j,\ell}\}\bigg] \\
&\leq \|\sigma_\varepsilon^2\|_\infty\sum_{\boldsymbol{k}\leq\boldsymbol{m}^\star}\mathbb{E}\bigg[\frac{1}{nJ^2}\sum_{j=1}^{J}\theta^2_{\boldsymbol{k}}(T_j)\bigg] = \|\sigma_\varepsilon^2\|_\infty\frac{\{m^\star\}^d}{nJ} := \mathbb{H}^2,
\end{aligned}$$

as $J^{-1}\sum_{j=1}^{J}\theta^2_{\boldsymbol{k}}(T_j) = 1$. Next, since $\bar{\varepsilon}^{(1)}_{\bullet,j}$ is bounded by $2B_{n,J}$, we get

$$\begin{aligned}
\sup_{h\in S_{m^\star},\,\|h\|_J=1}\sup_{|x|\leq 2B_{n,J}}\sup_{1\leq j\leq J}|h(T_j)\sigma_\varepsilon(T_j)x| &\leq 2B_{n,J}\sup_{h\in S_{m^\star},\,\|h\|_J=1}\sqrt{\sum_{j=1}^{J}h^2(T_j)\sigma^2_\varepsilon(T_j)} \\
&\leq 2B_{n,J}\sqrt{J\,\|\sigma^2_\varepsilon\|_\infty} =: \mathbb{M}.
\end{aligned}$$

Lastly,

$$\begin{aligned}
\sup_{h\in S_{\boldsymbol{m}^\star},\,\|h\|_J=1}\frac{1}{J}\sum_{j=1}^{J}\mathrm{Var}\big(h(T_j)\sigma_\varepsilon(T_j)\bar{\varepsilon}_{\bullet,j}\big) &= \sup_{h\in S_{\boldsymbol{m}^\star},\,\|h\|_J=1}\frac{1}{J}\sum_{j=1}^{J}h^2(T_j)\sigma^2_\varepsilon(T_j)\mathrm{Var}(\bar{\varepsilon}_{\bullet,j}) \\
&\leq \frac{\|\sigma^2_\varepsilon\|_\infty}{n} := v.
\end{aligned}$$

The Talagrand Inequality writes

$$\mathbb{E}\Big(\sup_{h\in S_{\boldsymbol{m}'\vee\boldsymbol{m}},\|h\|_J=1}[\nu_J^{(1)}(h)]^2-\frac{1}{2}p(\boldsymbol{m},\boldsymbol{m}')\Big)_+ \\ \leq C_1\frac{\|\sigma_\varepsilon^2\|_\infty}{nJ}\left(\exp\big(-C_2\{m^\star\}^d\big)+nB_{n,J}^2\exp\left(-C_3\sqrt{\frac{\{m^\star\}^d}{nB_{n,J}^2}}\right)\right),$$

for C_1, C_2, C_3 positive constants.

Next, we note that

$$\sum_{\boldsymbol{m}\in\mathcal{M}_J}\exp\big(-C_2m^d\big)\leq\sum_{m\in\mathbb{N}^*}\exp\big(-C_2m^d\big)\leq C_4<\infty.$$

Moreover, taking $B_{n,J}=\sqrt{1/n}$ leads to

$$\sum_{\boldsymbol{m}\in\mathcal{M}_J}nB_{n,J}^2\exp(-C_3m^{d/2})\leq\sum_{m\in\mathbb{N}^*}\exp\big(-C_3m^{d/2}\big)\leq C_5<\infty.$$

Thus, we get, for $J\geq 1$,

$$\mathbb{E}\Big(\sup_{h\in S_{\boldsymbol{m}'\vee\boldsymbol{m}},\|h\|_J=1}[\nu_J^{(1)}(h)]^2-\frac{1}{2}p(\boldsymbol{m},\boldsymbol{m}')\Big)_+\leq C\frac{\|\sigma_\varepsilon^2\|_\infty}{nJ},$$

for some constant C, and we conclude

$$\mathbb{E}\left[\sup_{h\in S_{\widehat{\boldsymbol{m}}\vee\boldsymbol{m}},\|h\|_J=1}[\nu_J^{(1)}(h)]^2-4\|\sigma_\varepsilon^2\|_\infty\frac{(m\vee\widehat{m})^d}{nJ}\right]_+\leq C\frac{\|\sigma_\varepsilon^2\|_\infty}{nJ}. \tag{38}$$

Therefore, (38) implies that $\mathbf{T}_2\leq C\|\sigma_\varepsilon^2\|_\infty/(nJ)$, which together with (37) ends the proof of Lemma 2, with $(1/2)p(\boldsymbol{m},\boldsymbol{m}')=4\|\sigma_\varepsilon^2\|_\infty(m\vee m')^d/(nJ)$. □

5.2 Proofs for the random design case

Proof of Lemma 1. Let $\mathbf{H}_{\boldsymbol{m}}$ and $\widehat{\mathbf{H}}_{\boldsymbol{m}}$denote the $D_{\boldsymbol{m}}\times D_{\boldsymbol{m}}$ matrix corresponding to the developed version of the hypermatrices $\Psi_{\boldsymbol{m}}$ and $\widehat{\Psi}_{\boldsymbol{m}}$ respectively. We also denote by $(\phi_k)_{1\leq k\leq D_{\boldsymbol{m}}}$ the developed version of the basis $\varphi_{\boldsymbol{k}}$, so that for $h\in S_{\boldsymbol{m}}$, we have the two writings

$$h=\sum_{1\leq\boldsymbol{k}\leq\boldsymbol{m}}a_{\boldsymbol{k}}\varphi_{\boldsymbol{k}}=\sum_{k=1}^{D_{\boldsymbol{m}}}b_k\phi_k$$

and

$$[\mathbf{H}_{\boldsymbol{m}}]_{k_1,k_2}=\langle\phi_{k_1},\phi_{k_2}\rangle_\rho,\quad\left[\widehat{\mathbf{H}}_{\boldsymbol{m}}\right]_{k_1,k_2}=\langle\phi_{k_1},\phi_{k_2}\rangle_N=\frac{1}{N}\sum_{q=1}^N\phi_{k_1}(T_q)\phi_{k_2}(T_q).$$

Then we can write

$$\begin{aligned}\sup_{h\in S_{\boldsymbol{m}},h\neq0}\left|\frac{\|h\|_N^2}{\|h\|_\rho^2}-1\right| &= \sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}\big|\|t\|_N^2-1\big| \\ &= \sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}\left|\frac{1}{N}\sum_{q=1}^N\sum_{1\leq k_1,k_2\leq D_{\boldsymbol{m}}}b_{k_1}b_{k_2}\phi_{k_1}(T_q)\phi_{k_2}(T_q)-1\right| \\ &= \sup_{\mathbf{x}\in\mathbb{R}^{D_{\boldsymbol{m}}},\mathbf{x}^\top\mathbf{H}_{\boldsymbol{m}}\mathbf{x}=1}\left|\mathbf{x}^\top\left(\widehat{\mathbf{H}}_{\boldsymbol{m}}-\mathbf{H}_{\boldsymbol{m}}\right)\mathbf{x}\right|\end{aligned}$$

We denote by $\mathbf{H}_{\boldsymbol{m}}^{1/2}$ a symmetric square root of $\mathbf{H}_{\boldsymbol{m}}$, and we get

$$\begin{aligned}\sup_{h\in S_{\boldsymbol{m}},h\neq 0}\left|\frac{\|h\|_N^2}{\|h\|_\rho^2}-1\right| &= \sup_{\mathbf{x}\in\mathbb{R}^{D_{\boldsymbol{m}}},\|\mathbf{x}\|=1}\left|\mathbf{x}^\top\left(\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\right)\mathbf{x}\right| \\ &= \left\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\right\|_{\mathrm{op}}\end{aligned}$$

As a consequence

$$\Omega_{\boldsymbol{m}}:=\left\{\left|\frac{\|h\|_N^2}{\|h\|_\rho^2}-1\right|\leq\frac{1}{2},\forall h\in S_{\boldsymbol{m}},h\neq 0\right\}=\left\{\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\|_{\mathrm{op}}\leq\frac{1}{2}\right\}.$$

Next, we aim proving (with $\mathrm{Id}_{D_{\boldsymbol{m}}}$ the identity matrix of dimension $D_{\boldsymbol{m}}$),

$$\mathbb{P}\left(\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\|_{\mathrm{op}}>\frac{1}{2}\right)\leq cN^{-p}. \tag{39}$$

Moreover, under Condition (21), as $\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}=\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}$, it is easy to see that

$$\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\|_{\mathrm{op}}\leq\frac{1}{2}\ \Rightarrow\ \frac{1}{2}\mathbf{H}_{\boldsymbol{m}}\preceq\widehat{\mathbf{H}}_{\boldsymbol{m}}\ \Rightarrow\ \|\widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}\leq 2\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}.$$

Thus, since $\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}=\|\widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}$, we would have $\Omega_{\boldsymbol{m}}\subset\Lambda_{\boldsymbol{m}}$ and consequently $\mathbb{P}(\Lambda_{\boldsymbol{m}}^c)\leq\mathbb{P}(\Omega_{\boldsymbol{m}}^c)$.

We now prove (39) under condition (21), by applying the Matrix Chernoff inequality to the N independent symmetric matrices (see [35]; see also Theorem 4 in the Appendix for the notation)

$$\mathbf{M}_q=\frac{1}{N}\mathbf{H}_{\boldsymbol{m}}^{-1/2}\left[\phi_k(T_q)\phi_{k'}(T_q)\right]_{1\leq k,\,k'\leq D_{\boldsymbol{m}}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}.$$

By definition, $\mathbb{E}(\mathbf{M}_q)=\mathrm{Id}_{D_{\boldsymbol{m}}}$ and thus, $\mu_{\min}=\mu_{\max}=1$. Moreover, considering $\mathbf{y}=\mathbf{H}_{\boldsymbol{m}}^{-1/2}\mathbf{x}$ with $\|\mathbf{x}\|=1$, it holds that

$$\lambda_{\max}(\mathbf{M}_q)=\sup_{\|\mathbf{x}\|=1}\mathbf{x}^\top\mathbf{M}_q\mathbf{x}\leq\frac{L(\boldsymbol{m})}{N}\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}},$$

because

$$\left(\sum_{k\leq D_{\boldsymbol{m}}}y_k\phi_k(T_q)\right)^2\leq\sum_{k\leq D_{\boldsymbol{m}}}y_k^2\sum_{k'\leq D_{\boldsymbol{m}}}\phi_{k'}^2(T_q)\leq L(\boldsymbol{m})\|\mathbf{y}\|^2\leq L(\boldsymbol{m})\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}=L(\boldsymbol{m})\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}.$$

Proceeding as in [11], [12], we apply the Matrix Chernoff inequality (see Th. 4) to obtain

$$\mathbb{P}\left(\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\|_{\mathrm{op}}>\frac{1}{2}\right)\leq 2D_{\boldsymbol{m}}\exp\left(-c(\delta)\frac{N}{L(\boldsymbol{m})\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}}\right),$$

where, for $0\leq\delta<1$, we set $c(\delta)=(1+\delta)\log(1+\delta)-\delta\geq 0$. Then, it holds

$$c(1/2)=\frac{3}{2}\log\frac{3}{2}-\frac{1}{2}=\frac{3\log(3/2)-1}{2}=(p+1)c_p.$$

Under Condition (21), it holds

$$c(1/2)\frac{N}{L(\boldsymbol{m})\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}}\geq\frac{c(1/2)}{c_p}\log N=(p+1)\log N.$$

Since $D_{\boldsymbol{m}}\leq N$, we conclude

$$\mathbb{P}\left(\|\mathbf{H}_{\boldsymbol{m}}^{-1/2}\widehat{\mathbf{H}}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1/2}-\mathrm{Id}_{D_{\boldsymbol{m}}}\|_{\mathrm{op}}>\frac{1}{2}\right)\leq 2N\cdot N^{-(p+1)}=2\,N^{-p},$$

which establishes (39) with the constant $c=2$. □

Proof of Proposition 3. Let $\Pi_{\boldsymbol{m}} g \in S_{\boldsymbol{m}}$ be the orthogonal projection of g on $S_{\boldsymbol{m}}$ with respect to the empirical norm. With $\Lambda_{\boldsymbol{m}}$ the set defined in (23), we have

$$\begin{aligned}
\|\widetilde{g}_{\boldsymbol{m}} - g\|_N^2 &= \|\widehat{g}_{\boldsymbol{m}} - g\|_N^2 \mathbf{1}_{\Lambda_{\boldsymbol{m}}} + \|g\|_N^2 \mathbf{1}_{\Lambda_{\boldsymbol{m}}^c} \\
&= \left\{\|\Pi_{\boldsymbol{m}} g - g\|_N^2 + \|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_N^2\right\} \mathbf{1}_{\Lambda_{\boldsymbol{m}}} + \|g\|_N^2 \mathbf{1}_{\Lambda_{\boldsymbol{m}}^c} \\
&\leq \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_N^2 + \|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_N^2 \mathbf{1}_{\Lambda_{\boldsymbol{m}}} + \|g\|_N^2 \mathbf{1}_{\Lambda_{\boldsymbol{m}}^c} \\
&=: \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3.
\end{aligned}$$

Clearly,

$$\mathbb{E}(\mathbb{T}_1) \leq \inf_{h \in S_{\boldsymbol{m}}} \|h - g\|_\rho^2. \tag{40}$$

As $\mathbb{E}[g^4(T_{1,1})] < +\infty$, we obtain straightforwardly that

$$\mathbb{E}(\mathbb{T}_3) \leq \mathbb{E}^{1/2}[g^4(T_{1,1})] c^{1/2} N^{-p/2}, \tag{41}$$

with p and c from Lemma 1. Next, since

$$\widehat{g}_{\boldsymbol{m}}(\mathbf{T}) - \Pi_{\boldsymbol{m}} g(\mathbf{T}) = \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{u},$$

with $\mathbf{u} \in \mathbb{R}^N$ defined in (19), it holds that

$$\begin{aligned}
\|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_N^2 &= \frac{1}{N} \{\widehat{g}_{\boldsymbol{m}}(\mathbf{T}) - \Pi_{\boldsymbol{m}} g(\mathbf{T})\}^\top \{\widehat{g}_{\boldsymbol{m}}(\mathbf{T}) - \Pi_{\boldsymbol{m}} g(\mathbf{T})\} \\
&= \frac{1}{N} \mathbf{u}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{u}.
\end{aligned}$$

Let decompose $\mathbf{u} = \mathbf{X} + \bar{\boldsymbol{\varepsilon}}$ with $\bar{\boldsymbol{\varepsilon}} \in \mathbb{R}^N$ defined in (19) and

$$\mathbf{X} = (X_1(T_{1,1}), \ldots, X_1(T_{1,N_1}), X_2(T_{2,1}), \ldots X_2(T_{2,N_2}), \ldots X_n(T_{n,N_n}))^\top \in \mathbb{R}^N.$$

The cross-terms involving $\mathbf{X}$ and $\bar{\boldsymbol{\varepsilon}}$ are centered thanks to the fact that the $\varepsilon_{i,j}$'s are centered and independent of $\mathbf{X}$ and $\mathbf{T}$. Therefore

$$\mathbb{E}(\mathbb{T}_2) = \mathbb{E}(\|\widehat{g}_{\boldsymbol{m}} - \Pi_{\boldsymbol{m}} g\|_N^2 \, \mathbf{1}_{\Lambda_{\boldsymbol{m}}}) = \mathbb{V}_1 + \mathbb{V}_2,$$

with

$$\mathbb{V}_1 := \mathbb{E}\left(\frac{1}{N} \mathbf{X}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}}}\right),$$

and

$$\mathbb{V}_2 := \frac{1}{N} \mathbb{E}\left(\bar{\boldsymbol{\varepsilon}}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \bar{\boldsymbol{\varepsilon}} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}}}\right).$$

Since $\Lambda_{\boldsymbol{m}}$ depends only on $\mathbf{T}$, and the $\varepsilon_{i,j}$'s are independent of $\mathbf{T}$, we can write

$$\mathbb{E}\big(\bar{\boldsymbol{\varepsilon}}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \bar{\boldsymbol{\varepsilon}} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}}}\big) = \mathbb{E}\left(\mathrm{Tr}\left(\mathbb{E}(\bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}^\top | \mathbf{T}) \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top\right) \mathbf{1}_{\Lambda_{\boldsymbol{m}}}\right).$$

Then, using the same reasoning as in (33), we get

$$\mathbb{V}_2 \leq \|\sigma_\varepsilon^2\|_\infty \frac{D_{\boldsymbol{m}}}{N}. \tag{42}$$

Next, with $\Omega_{\boldsymbol{m}}$ the set defined in (24), we further decompose the first term

$$\begin{aligned}
\mathbb{V}_1 = \mathbb{E}\left(\frac{1}{N} \mathbf{X}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X} \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) \\
+ \mathbb{E}\left(\frac{1}{N} \mathbf{X}^\top \widehat{\Phi}_{\boldsymbol{m}} (\widehat{\Phi}_{\boldsymbol{m}}^\top \widehat{\Phi}_{\boldsymbol{m}})^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X} \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}^c}\right) := \mathbb{V}_{1,1} + \mathbb{V}_{1,2}.
\end{aligned}$$

The term $\mathbb{V}_{1,2}$ is a remainder that can be handled by repeatedly applying Cauchy-Schwarz inequality as follows: with the notation from Section 1 and the simplified indexing introduced in (18),

$$\begin{aligned}
\mathbb{V}_{1,2} &= \frac{1}{N^2}\mathbb{E}\left(\mathbf{X}^\top \widehat{\Phi}_{\boldsymbol{m}} \widehat{\Psi}_{\boldsymbol{m}}^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X} \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}^c}\right) \\
&\leq \frac{1}{N^2}\mathbb{E}\left(\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \mathbf{1}_{\Lambda_{\boldsymbol{m}}} \|\widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X}\|_{\mathbb{R}^{\boldsymbol{m}}}^2 \mathbf{1}_{\Omega_{\boldsymbol{m}}^c}\right) \\
&= \frac{1}{N^2}\mathbb{E}\left(\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \mathbf{1}_{\Lambda_{\boldsymbol{m}}} \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \left(\sum_{q=1}^N \varphi_{\boldsymbol{k}}(T_q) X_i(T_q)\right)^2 \mathbf{1}_{\Omega_{\boldsymbol{m}}^c}\right) \\
&\leq \frac{1}{N^2}\mathbb{E}\left(NL(\boldsymbol{m})\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \mathbf{1}_{\Lambda_{\boldsymbol{m}}} \sum_{q=1}^N [X_i(T_q)]^2 \mathbf{1}_{\Omega_{\boldsymbol{m}}^c}\right) \\
&\leq 2c_p \frac{N}{\log(N)} \frac{1}{N}\mathbb{E}\left(\sum_{q=1}^N [X_i(T_q)]^2 \mathbf{1}_{\Omega_{\boldsymbol{m}}^c}\right) \\
&\leq 2c_p \frac{N}{\log(N)} \mathbb{E}^{1/2}[X_1^4(T_{1,1})] \sqrt{\mathbb{P}(\Omega_{\boldsymbol{m}}^c)} \lesssim N^{-p/2+1}. \qquad (43)
\end{aligned}$$

For the last inequality, we appeal to Lemma 1. Moreover, the index i in $X_i(T_q)$ implicitly display depends on q, since X_i denotes the field associated to T_q, but this does not affect the bound as the X_i are independent of the design points.

For the term $\mathbb{V}_{1,1}$, we note that on $\Omega_{\boldsymbol{m}}$ the eigenvalues of $\mathbf{H}_{\boldsymbol{m}}^{-1/2} \widehat{\mathbf{H}}_{\boldsymbol{m}} \mathbf{H}_{\boldsymbol{m}}^{-1/2}$ belong to $[1/2, 3/2]$ and thus the eigenvalues of its inverse are in $[2/3, 2]$. We also denote by $\widehat{\mathbf{F}}_{\boldsymbol{m}}$ the $D_{\boldsymbol{m}} \times N$ extended version of $\widehat{\Phi}_{\boldsymbol{m}}$. From this and the inequality $\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} = \|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \leq 1/\rho_0$, see (20), we deduce

$$\begin{aligned}
\mathbb{V}_{1,1} &= \mathbb{E}\left(\frac{1}{N^2}\mathbf{X}^\top \widehat{\Phi}_{\boldsymbol{m}} \widehat{\Psi}_{\boldsymbol{m}}^{-1} \widehat{\Phi}_{\boldsymbol{m}}^\top \mathbf{X} \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) = \mathbb{E}\left(\frac{1}{N^2}\mathbf{X}^\top \widehat{\mathbf{F}}_{\boldsymbol{m}} \widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1} \widehat{\mathbf{F}}_{\boldsymbol{m}}^\top \mathbf{X} \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) \\
&= \frac{1}{N^2}\mathbb{E}\left(\mathbf{X}^\top \widehat{\mathbf{F}}_{\boldsymbol{m}} \mathbf{H}_{\boldsymbol{m}}^{-1/2} \left(\mathbf{H}_{\boldsymbol{m}}^{1/2} \widehat{\mathbf{H}}_m^{-1} \mathbf{H}_{\boldsymbol{m}}^{1/2}\right) \mathbf{H}_{\boldsymbol{m}}^{-1/2} \widehat{\mathbf{F}}_{\boldsymbol{m}}^\top \mathbf{X} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) \\
&\leq \frac{2}{N^2}\mathbb{E}\left(\mathbf{X}^\top \widehat{\mathbf{F}}_{\boldsymbol{m}} \mathbf{H}_{\boldsymbol{m}}^{-1} \widehat{\mathbf{F}}_{\boldsymbol{m}}^\top \mathbf{X} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) \qquad (44) \\
&\leq \frac{2}{\rho_0 N^2}\mathbb{E}\left(\mathbf{X}^\top \widehat{\mathbf{F}}_{\boldsymbol{m}} \widehat{\mathbf{F}}_{\boldsymbol{m}}^\top \mathbf{X} \, \mathbf{1}_{\Lambda_{\boldsymbol{m}} \cap \Omega_{\boldsymbol{m}}}\right) \\
&\leq \frac{2}{\rho_0 N^2}\mathbb{E}\left[\sum_{k=1}^{D_{\boldsymbol{m}}} \left(\sum_{q=1}^N \phi_k(T_q) X_i(T_q)\right)^2\right] \\
&= \frac{2}{\rho_0 N^2} \sum_{\boldsymbol{k} \leq \boldsymbol{m}} \sum_{i=1}^n \mathbb{E}\left[\left(\sum_{j=1}^{N_i} \varphi_{\boldsymbol{k}}(T_{i,j}) X_i(T_{i,j})\right)^2\right].
\end{aligned}$$

For the last equality we use the fact that the terms $A_i := \sum_{j=1}^{N_i} \varphi_{\boldsymbol{k}}(T_{i,j}) X_i(T_{i,j})$, $1 \leq i \leq n$,

are independent and centered. We deduce that

$$\begin{aligned}\mathbb{V}_{1,1} &\leq \frac{2}{\rho_0\, N^2}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{N_i}\varphi_{\boldsymbol{k}}^2(T_{i,j})X_i^2(T_{i,j})\right.\\ &\qquad\left.+\sum_{1\leq j\neq j'\leq N_i}\varphi_{\boldsymbol{k}}(T_{i,j})X_i(T_{i,j})\varphi_{\boldsymbol{k}}(T_{i,j'})X_i(T_{i,j'})\right]\\ &= \frac{2}{\rho_0\, N^2}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\sum_{i=1}^{n}\left(N_i\int\varphi_{\boldsymbol{k}}^2(u)\sigma_X(u,u)\rho(u)du\right.\\ &\qquad\left.+N_i(N_i-1)\iint\varphi_{\boldsymbol{k}}(u)\varphi_{\boldsymbol{k}}(v)\sigma_X(u,v)\rho(u)\rho(v)dudv\right),\end{aligned}$$

with $\sigma_X(u,v)=\mathbb{E}[X_1(u)X_1(v)]$. In the last bound, the term containing the simple integrals can be bounded as follows:

$$\frac{2}{\rho_0\, N}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\int\varphi_{\boldsymbol{k}}^2(u)\sigma_X(u,u)\rho(u)du\leq\frac{2L(\boldsymbol{m})}{\rho_0\, N}\int\sigma_X(u,u)\rho(u)du.$$

On the other hand, for the term containing the double integrals, by a projection argument which entails $\sum_{\boldsymbol{k}\leq\boldsymbol{m}}(\int\varphi_{\boldsymbol{k}}g)^2\leq\int g^2$, we deduce

$$\begin{aligned}&\frac{2}{\rho_0}\frac{\sum_{i=1}^n N_i(N_i-1)}{N^2}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\mathbb{E}\left[\left(\int\varphi_{\boldsymbol{k}}(u)\rho(u)X_1(u)du\right)^2\right]\\ &\qquad\leq\frac{2}{\rho_0}\frac{\sum_{i=1}^n N_i(N_i-1)}{N^2}\mathbb{E}\left(\int\{X_1(u)\rho(u)\}^2du\right)\\ &\qquad=\frac{2}{\rho_0}\frac{\sum_{i=1}^n N_i(N_i-1)}{N^2}\int\sigma_X(u,u)\rho^2(u)du.\end{aligned}$$

Combining the terms gives a first bound for $\mathbb{V}_{1,1}$, that is

$$\mathbb{V}_{1,1}\leq\frac{2}{\rho_0}\left(\frac{M_3L(\boldsymbol{m})}{N}+\frac{M_1}{N^2}\sum_{i=1}^{n}N_i(N_i-1)\right).\tag{45}$$

The term $\mathbb{V}_{1,1}$ can also be bounded in a different way. Let $Z_{i,k}=\sum_{j=1}^{N_i}X_i(T_{i,j})\phi_k(T_{i,j})$, which are centered and independent in $i=1,\ldots,n$, identically distributed for fixed k. Starting from the inequality (44), we have

$$\begin{aligned}\mathbb{V}_{1,1} &\leq \frac{2}{N^2}\sum_{1\leq k,k'\leq D_{\boldsymbol{m}}}\left[\mathbf{H}_{\boldsymbol{m}}^{-1}\right]_{k,k'}\mathbb{E}\left[\left(\sum_{i=1}^{n}Z_{i,k}\right)\left(\sum_{i=1}^{n}Z_{i,k'}\right)\right]\\ &= \frac{2}{N^2}\left\{\left(\sum_{i=1}^{n}N_i\right)\int\sigma_X(u,u)\left(\sum_{1\leq k,k'\leq D_{\boldsymbol{m}}}\phi_k(u)\left[\mathbf{H}_{\boldsymbol{m}}^{-1}\right]_{k,k'}\phi_{k'}(u)\right)\rho(u)du\right.\\ &\left.+\left(\sum_{i=1}^{n}N_i(N_i-1)\right)\iint\sigma_X(u,v)\left(\sum_{1\leq k,k'\leq D_{\boldsymbol{m}}}\phi_k(u)\left[\mathbf{H}_{\boldsymbol{m}}^{-1}\right]_{k,k'}\phi_{k'}(v)\right)\rho(u)\rho(v)\,dudv\right\}\\ &:= \mathbb{S}_1+\mathbb{S}_2.\end{aligned}$$

For the first term, we get, denoting by $\boldsymbol{\phi}(u) = (\phi_1(u), \ldots, \phi_{D_{\boldsymbol{m}}}(u))^\top \in \mathbb{R}^{D_{\boldsymbol{m}}}$,

$$\begin{aligned}
\mathbb{S}_1 &= \frac{2}{N}\int \sigma_X(u,u)\boldsymbol{\phi}(u)^\top \mathbf{H}_{\boldsymbol{m}}^{-1}\boldsymbol{\phi}(u)\rho(u)du \\
&\leq \frac{2M_1}{N}\int \boldsymbol{\phi}(u)^\top \mathbf{H}_{\boldsymbol{m}}^{-1}\boldsymbol{\phi}(u)\rho(u)du \\
&= \frac{2M_1}{N}\sum_{1\leq k,k'\leq D_{\boldsymbol{m}}}\int \phi_k(u)\phi_{k'}(u)\rho(u)du\left[\mathbf{H}_{\boldsymbol{m}}^{-1}\right]_{k,k'} \\
&= \frac{2M_1}{N}\sum_{1\leq k,k'\leq D_{\boldsymbol{m}}}[\mathbf{H}_{\boldsymbol{m}}]_{k,k'}\left[\mathbf{H}_{\boldsymbol{m}}^{-1}\right]_{k,k'} \\
&= \frac{2M_1}{N}\operatorname{Tr}(\mathbf{H}_{\boldsymbol{m}}\mathbf{H}_{\boldsymbol{m}}^{-1}) = \frac{2M_1\, D_{\boldsymbol{m}}}{N},
\end{aligned}$$

where for the inequality we use the fact that $\mathbf{H}_{\boldsymbol{m}}$ is positive definite and by definition $\sigma_X(u,u) \geq 0$. For $\mathbb{S}_2$, we write that

$$\begin{aligned}
\mathbb{S}_2 &= \frac{2}{N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right)\mathbb{E}\left[\left(\int X_1(u)\boldsymbol{\phi}(u)^\top\rho(u)du\right)\mathbf{H}_{\boldsymbol{m}}^{-1}\left(\int X_1(v)\boldsymbol{\phi}(v)\rho(v)dv\right)\right] \\
&\leq \frac{2}{N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right)\|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}\mathbb{E}\left[\left\|\left(\int X_1(u)\boldsymbol{\phi}(u)\rho(u)du\right)\right\|^2_{\mathbb{R}^{D_{\boldsymbol{m}}}}\right] \\
&\leq \frac{2}{\rho_0\, N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right)\mathbb{E}\left[\sum_{k=1}^{D_{\boldsymbol{m}}}\left(\int X_1(u)\phi_k(u)\rho(u)du\right)^2\right] \\
&\leq \frac{2}{\rho_0\, N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right)\mathbb{E}\left[\int X_1^2(u)\rho^2(u)du\right] = \frac{2M_2}{\rho_0\, N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right).
\end{aligned}$$

Combining the two terms yields a second bound for $\mathbb{V}_{1,1}$, that is

$$\mathbb{V}_{1,1} \leq \frac{2M_1\, D_{\boldsymbol{m}}}{N} + \frac{2M_2}{\rho_0\, N^2}\left(\sum_{i=1}^n N_i(N_i-1)\right). \tag{46}$$

Gathering (40), (41), (42), (43) and (46) (resp. (45)) gives the result stated in Proposition 3 (resp. Remark 4) provided $p = 4$ or larger and we redefine $p \leftarrow p/2 - 1 \geq 1$. □

Proof of Theorem 2. First, note that, due to Assumption 2 and (17), it holds

$$\inf_{\widehat{g}}\sup_{g\in\mathcal{S}(\boldsymbol{\alpha},L)}\mathbb{E}[\|\hat{g}-g\|_\rho^2] \geq \frac{1}{\rho_0^2}\inf_{\hat{g}}\sup_{g\in\mathcal{S}(\boldsymbol{\alpha},L)}\mathbb{E}[\|\hat{g}-g\|^2],$$

and therefore we may establish a lower bound for the standard $\mathbb{L}^2$-norm denoted by $\|\cdot\|$. The term n^{-1} in the lower bound do raises any difficulty as it corresponds to the parametric rate, which is attained considering g a constant function, $X \equiv 0$ and $\sigma_\varepsilon \equiv 1$ (see also [9, Th. 2.1]).

Following [36, Th. 2.5 and Th. 2.7], the remaining part of the lower bound is established by considering a decision problem between $K+1$ competing functions $g_0 = g$ and $g_1, \ldots, g_K$, all contained in $\mathcal{S}(\boldsymbol{\alpha}, L)$ with $\boldsymbol{\alpha} = (\alpha_1 \ldots, \alpha_d) \in \mathbb{N}_{>0}^d$ and $L > 0$, that satisfy the following two conditions, for some increasing $K \in \mathbb{N}_{>0}$ to be specified later:

(i) $\|g_{k'} - g_k\|^2 \geq \mathfrak{c}_g N^{-\frac{2\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}}$, $\forall\, 0 \leq k' < k \leq K$ for some positive constant $\mathfrak{c}_g$,

(ii) $K^{-1}\sum_{k=1}^K \mathrm{KL}(\mathbb{P}_k^{\otimes N}, \mathbb{P}_0^{\otimes N}) \leq \mathfrak{c}\log K$, for some $0 < \mathfrak{c} < 1/8$,

where $\mathbb{P}_k$ denotes the probability measure of the observations (16) corresponding to the function g_k and $\mathrm{KL}(\cdot,\cdot)$ is the Kullback-Leibler divergence. Fix $\sigma_\varepsilon \equiv 1$ and let $\rho(x) = \mathbf{1}_{\mathcal{T}}(x)$ be the uniform density function on $\mathcal{T}$. Consider $\mathcal{N}(0,1)$ as the common distribution of $\varepsilon_{i,j}$ and the X_i's are i.i.d. centered Gaussian fields with bounded covariance function Γ_X. Let $Q = (Q_1, \ldots, Q_d) \in \mathbb{N}_{>0}^d$ to be specified later. Let $D_Q = \prod_{l=1}^d Q_l$ and $\chi : \mathbb{R}^d \to \mathbb{R}$, which is zero outside $\mathcal{T}$ and admits bounded partial derivatives on $\mathcal{T}$ up to order α_l in the l-direction, for $1 \le l \le d$. Consider the functions with non overlapping support

$$f_{\boldsymbol{q}}(x) = \chi(xQ - \boldsymbol{q} + \mathbf{1}_d), \quad x \in \mathcal{T},\ \boldsymbol{q} \in \mathbb{N}_{>0}^d,\ \boldsymbol{q} \le Q,$$

where $xQ - \boldsymbol{q} + \mathbf{1}_d$ has the components $(x_l Q_l - q_l + 1)$, $1 \le l \le d$. (Recall that $\boldsymbol{q} = (q_1, \ldots, q_d) \in \mathbb{N}_{>0}^d$ with $\boldsymbol{q} \le Q$ means $1 \le q_l \le Q_l$, $\forall 1 \le l \le d$.) Note that the functions $f_{\boldsymbol{q}}$ depend also on Q. For $\boldsymbol{q} \in \mathbb{N}_{>0}^d$ with $\boldsymbol{q} \le Q$, the function $f_{\boldsymbol{q}}(x)$ is supported on $\mathcal{R}_{\boldsymbol{q}} = [(q_1-1)/Q_1, q_1/Q_1] \times \ldots \times [(q_d-1)/Q_d, q_d/Q_d] \subset \mathcal{T}$. Therefore, for any $\boldsymbol{q} \neq \boldsymbol{q}'$, the intersection of the supports of $f_{\boldsymbol{q}}$ and $f_{\boldsymbol{q}'}$ has measure zero, implying

$$\int_{\mathcal{T}} f_{\boldsymbol{q}}(x) f_{\boldsymbol{q}'}(x) dx = D_Q^{-1} \|\chi\|^2 \mathbf{1}_{\{\boldsymbol{q}=\boldsymbol{q}'\}}.$$

Let $\boldsymbol{\theta} \in \{0,1\}^{D_Q}$ with components $\theta_{\boldsymbol{q}} \in \{0,1\}$, $\boldsymbol{q} \le Q$, and let $\delta, A > 0$ to be specified later. Using these quantities and the D_Q functions $f_{\boldsymbol{q}}$ we define

$$g_{\boldsymbol{\theta}}(x) = \delta A \sum_{\boldsymbol{q} \le Q} \theta_{\boldsymbol{q}} f_{\boldsymbol{q}}(x).$$

First, we derive conditions guaranteeing that $g_{\boldsymbol{\theta}}$ belongs to $\mathcal{S}(\boldsymbol{\alpha}, L)$. By definition, $g_{\boldsymbol{\theta}}$ admits α_l bounded derivatives in the l-direction, $1 \le l \le d$, and by a change of variables, for any $\alpha_l' \in \mathbb{N}$ such that $\alpha_l' \le \alpha_l$, any $1 \le l \le d$, and any $\boldsymbol{\theta} \in \{0,1\}^{D_Q}$,

$$\|\partial_l^{\alpha_l'} g_{\boldsymbol{\theta}}\|^2 = \delta^2 A^2 Q_l^{2\alpha_l'} \sum_{\boldsymbol{q} \le Q} \theta_{\boldsymbol{q}}^2 \int_{\mathcal{T}} \{\partial_l^{\alpha_l'} f_{\boldsymbol{q}}(x))\}^2 dx \le \delta^2 A^2 Q_l^{2\alpha_l'} \|\partial_l^{\alpha_l'} \chi\|^2.$$

Here, $\partial_l^{\alpha_l}$ denotes the partial derivative operator of order α_l with respect to the l-th variable. Then $g_{\boldsymbol{\theta}} \in \mathcal{S}(\boldsymbol{\alpha}, L)$ if $(\delta, A, Q_1, \ldots, Q_d)$ and the function χ are such that

$$\delta^2 A^2 \max_{1 \le l \le d} \left\{ Q_l^{2\alpha_l} \|\partial_l^{\alpha_l} \chi\|^2 \right\} \le L.$$

Next, we establish (i). For any $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \{0,1\}^{D_Q}$, by definition and a change of variables, it holds

$$\|g_{\boldsymbol{\theta}'} - g_{\boldsymbol{\theta}}\|^2 = \delta^2 A^2 D_Q^{-1} \|\chi\|^2 H(\boldsymbol{\theta}, \boldsymbol{\theta}'),$$

where $H(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sum_{\boldsymbol{q} \le Q} |\theta_{\boldsymbol{q}} - \theta_{\boldsymbol{q}}'|$ is the Hamming distance between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. The Varshamov–Gilbert bound (see [36, Lemma 2.9]) ensures that for $D_Q \ge 8$ there exist $K \ge 2^{D_Q/8}$ elements $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$ of $\{0,1\}^{D_Q}$ such that $H(\boldsymbol{\theta}^{(k')}, \boldsymbol{\theta}^{(k)}) \ge D_Q/8$, for all $0 \le k' < k \le K$, where $\boldsymbol{\theta}^{(0)} = (0, \ldots, 0)$. Then, it holds

$$\|g_{\boldsymbol{\theta}^{(k')}} - g_{\boldsymbol{\theta}^{(k)}}\|^2 \ge \delta^2 \frac{A^2}{8} \|\chi\|^2, \quad \forall 0 \le k' < k \le K.$$

As a consequence, to establish (i) it suffices to select (δ, A) such that $\delta^2 A^2 = O\big(N^{-\frac{2\overline{\alpha}}{2\overline{\alpha}+d}}\big)$.

Second, to derive (ii) note that the following variables are independent for $1 \le i \le n$,

$$Y_{i,\cdot}^{\boldsymbol{\theta}} \mid T_{i,\cdot} = \left(Y_{i,1}^{\boldsymbol{\theta}} \ , \ldots, \ Y_{i,N_i}^{\boldsymbol{\theta}} \ \mid \ T_{i,1} \ , \ldots, \ T_{i,N_i}\right)^\top \sim \mathcal{N}\left(\mu_{\boldsymbol{\theta}}, \mathrm{Id}_{N_i} + \Sigma_X^i\right),$$

where

$$Y_{i,j}^{\boldsymbol{\theta}} = g_{\boldsymbol{\theta}}(T_{i,j}) + X_i(T_{i,j}) + \varepsilon_{i,j}, \quad \mu_{\boldsymbol{\theta}} = \begin{pmatrix} g_{\boldsymbol{\theta}}(T_{i,1}) & \dots & g_{\boldsymbol{\theta}}(T_{i,N_i}) \end{pmatrix}^\top,$$

and $\Sigma_X^i = \big(\Gamma_X(T_{i,j}, T_{i,j'})\big)_{1\le j,j'\le N_i}$. By standard computations, for all $1 \le k \le K$ and with $\mathbb{P}_k^{\otimes N}$ the distribution of $(Y_{i,j}^{\boldsymbol{\theta}_k})_{1\le i\le n, 1\le j\le N_i}$, conditionally on $\mathbf{T}$, it holds

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^{K} \mathrm{KL}(\mathbb{P}_k^{\otimes N}, \mathbb{P}_0^{\otimes N} \mid \mathbf{T}) &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{2}\sum_{i=1}^{n} \mu_{\boldsymbol{\theta}^{(k)}}^\top \left(\mathrm{Id}_{N_i} + \Sigma_X^i\right)^{-1} \mu_{\boldsymbol{\theta}^{(k)}} \\ &\le \frac{1}{2K}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{N_i} g_{\boldsymbol{\theta}^{(k)}}(T_{i,j})^2.\end{aligned}$$

Integrating with respect to $\mathbf{T}$ and using that $\|g_{\boldsymbol{\theta}}\|^2 \le \delta^2 A^2 \|\chi\|^2$, for $K \ge 2^{D_{Q_N}/8}$ it holds

$$\frac{1}{K}\sum_{k=1}^{K} \mathrm{KL}(\mathbb{P}_k^{\otimes N}, \mathbb{P}_0^{\otimes N}) \le \frac{N\delta^2 A^2 \|\chi\|^2}{2}.$$

Combining facts, $g \in \mathcal{S}(\boldsymbol{\alpha}, L)$, (i) and (ii) are satisfied as soon as $(\delta, A, Q_1, \dots, Q_d)$ depend on N such that:

$$\begin{cases} \delta^2 A^2 = O\Big(N^{-\frac{2\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}}\Big) & \text{(Condition (i))}, \\ \delta^2 A^2 \max_{1\le l\le d} Q_l^{2\alpha_l} \|\partial_l^{\alpha_l}\chi\|^2 \le L & \text{(Condition } \mathcal{S}(\boldsymbol{\alpha}, L)\text{)}, \\ NA^2 = D_Q \quad \text{and} \quad \frac{\delta^2\|\chi\|^2}{2} < \frac{\log 2}{8} & \text{(Condition (ii))}. \end{cases}$$

A possible choice is to select

$$A := A_N = N^{-\frac{\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}}, \quad Q_l = Q_{l,N} = N^{\frac{\overline{\boldsymbol{\alpha}}}{\alpha_l(2\overline{\boldsymbol{\alpha}}+d)}}, \quad \forall 1 \le l \le d,$$

implying that

$$D_{Q_N} = \prod_{l=1}^{d} Q_{l,N} = N^{\sum_{l=1}^{d}\frac{1}{\alpha_l}\frac{\overline{\boldsymbol{\alpha}}}{2\overline{\boldsymbol{\alpha}}+d}} = N^{\frac{d}{2\overline{\boldsymbol{\alpha}}+d}} \quad \text{and} \quad NA_N^2 = D_{Q_N},$$

and to take $\delta > 0$ such that

$$\delta < \frac{L}{\sqrt{\max_{1\le l\le d}\|\partial_l^{\alpha_l}\chi\|}} \wedge \frac{\sqrt{\log 2}}{2\|\chi\|}.$$

Then, Theorem 2.7 in [36] allows us to conclude. □

5.2.1 Risk bound for model selection with independent design

This section contains the proof of Theorem 3. Before proceeding with that proof, we introduce some definitions and technical results. Recall that the collection of models $\mathcal{M}_N$ and its empirical counterpart $\widehat{\mathcal{M}}_N$ are defined in (27) and (28), respectively. Let us define

$$\mathcal{M}_N^+ = \left\{ \boldsymbol{m} = m\mathbf{1}_d, m \in \{1, \dots, N\}, \quad L(\boldsymbol{m})(\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \vee 1) \le \frac{4c_p N}{\log(N)} \right\},$$

and with the sets $\Omega_{\boldsymbol{m}}$ as defined in (24), let

$$\Omega_N = \bigcap_{\boldsymbol{m}\in\mathcal{M}_N^+} \Omega_{\boldsymbol{m}}. \tag{47}$$

To bound the risk of the estimator from Theorem 3, we also define the set of events

$$\Xi_N = \{\mathcal{M}_N \subset \widehat{\mathcal{M}}_N \subset \mathcal{M}_N^+\}.$$

The following lemmas will be used in the proof of Theorem 3. Below, for any matrix $\mathbf{A}$, $\mathrm{Sp}(\mathbf{A})$ denotes the spectrum of $\mathbf{A}$.

Lemma 3. *Under the assumptions of Theorem 3, it holds that $\Omega_N \subset \Xi_N$.*

Lemma 4. *Assume that the $\varepsilon_{i,j}$'s have a finite fourth order moment, and set*

$$p_2(\boldsymbol{m}, \boldsymbol{m}') = 4\|\sigma_\varepsilon^2\|_\infty \frac{(m \vee m')^d}{N}.$$

Then

$$\mathbb{E}\left\{\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_N = 1} \left(\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}\right)^2 - p_2(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+ \mathbf{1}_{\Omega_N}\right\} \leq C\frac{\log^2(N)}{N},$$

where C is a constant depending on ρ_0, $\|\sigma_\varepsilon^2\|_\infty$ and $\mathbb{E}(\varepsilon_{1,1}^4)$.

The proofs of Lemmas 3 and 4 are given in Section 6.2.

Lemma 5. *The assumptions of Theorem 3 hold. Let*

$$p_1(\boldsymbol{m}, \boldsymbol{m}') = 4B_{n,X}^2 \frac{(m \vee m')^d}{N}.$$

Then, there exists a constant c depending on ρ_0 such that, for $M_3 = \int \sigma_X(u,u)\rho(u)du$, it holds

$$\mathbb{E}\left\{\left(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N}\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i} h(T_{i,j})X_i(T_{i,j})\right)\right)^2 - p_1(\boldsymbol{m}, \widehat{\boldsymbol{m}})\right)_+\right\} \leq \frac{c\,\mathbb{E}\left(B_{n,X}^2\right)}{N} + 2M_3\frac{\sum_{i=1}^n N_i^2}{N}.$$

Proof of Lemma 5. Define $\nu_N(h) := N^{-1}\sum_{i=1}^n \sum_{j=1}^{N_i} h(T_{i,j})X_i(T_{i,j})$. Use the decomposition

$$\begin{aligned}\nu_N(h) &= \frac{1}{N}\sum_{i=1}^{n}\left(\sum_{j=1}^{N_i} h(T_{i,j})X_i(T_{i,j}) - \int X_i(u)h(u)\rho(u)du\right) \\ &\quad + \frac{1}{N}\sum_{i=1}^{n} N_i \int X_i(u)h(u)\rho(u)du := \nu_N^{(X)}(h) + R_n(h).\end{aligned}$$

Let $\mathcal{S}_N$ be a nesting space ($\forall m, S_{\boldsymbol{m}} \subset \mathcal{S}_N$), with dimension $\mathcal{D}_N \leq N$. Then

$$\begin{aligned}
\mathbb{E}\left(\sup_{h\in S_{\widehat{\boldsymbol{m}}\vee\boldsymbol{m}},\|h\|_\rho=1} R_n^2(h)\right) &\leq \mathbb{E}\left(\sup_{h\in\mathcal{S}_N,\|h\|_\rho=1} R_n^2(h)\right) \leq \sum_{p=1}^{\mathcal{D}_N}\mathbb{E}(R_n^2(\varphi_p^{(\rho)})) \\
&\leq \sum_{p=1}^{\mathcal{D}_N}\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^n N_i\int X_i(u)\varphi_p^{(\rho)}(u)\rho(u)du\right) \\
&= \sum_{p=1}^{\mathcal{D}_N}\sum_{i=1}^n\left(\frac{N_i}{N}\right)^2\mathrm{Var}\left(\int X_1(u)\varphi_p^{(\rho)}(u)\rho(u)du\right) \\
&\leq \frac{\sum_{i=1}^n N_i^2}{N^2}\mathbb{E}\left[\sum_{p=1}^{\mathcal{D}_N}\left(\int X_1(u)\varphi_p^{(\rho)}(u)\rho(u)du\right)^2\right] \\
&\leq \frac{\sum_{i=1}^n N_i^2}{N^2}\mathbb{E}\left[\int X_1^2(u)\rho(u)du\right] = M_3\frac{\sum_{i=1}^n N_i^2}{N^2}.
\end{aligned}$$

Here, $\varphi_p^{(\rho)}$, $1 \leq p \leq \mathcal{D}_N$, are orthonormal with respect to the ρ-weighted inner product.

Next, we consider the empirical process $\nu_N^{(X)}$ conditionally on fixed realizations $X_i = x_i$, $1 \leq i \leq n$. The variables involved in $\nu_N^{(X)}$ are the $T_{i,j}$'s, which are independent. Thus, conditionally on the fixed realizations of the X_i's, $\nu_N^{(X)}$ is an average of independent, non-identically distributed variables, for which Talagrand's Inequality can be applied. We can then apply Talagrand's Inequality conditionally on the fixed realizations of the X_i's as follows. First,

$$\begin{aligned}
\mathbb{E}\left(\sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}[\nu_N^{(x)}(h)]^2\right) &\leq \sum_{\boldsymbol{k}\leq\boldsymbol{m}}\mathbb{E}\left\{[\nu_N^{(x)}(\varphi_{\boldsymbol{k}}^{(\rho)})]^2\right\} \\
&= \frac{1}{N^2}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\sum_{i=1}^n N_i\mathrm{Var}\left(\varphi_{\boldsymbol{k}}^{(\rho)}(T_{1,1})x_i(T_{1,1})\right) \\
&\leq \frac{1}{N^2}\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\sum_{i=1}^n N_i\int x_i^2(u)(\varphi_{\boldsymbol{k}}^{(\rho)}(u))^2\rho(u)du \\
&\leq \frac{1}{N}B_{n,x}^2\sum_{\boldsymbol{k}\leq\boldsymbol{m}}\int(\varphi_{\boldsymbol{k}}^{(\rho)}(u))^2\rho(u)du \\
&\leq B_{n,x}^2\,\frac{m^d}{N} := \mathbb{H}^2.
\end{aligned}$$

Here, given the realizations of the X_i's, $B_{n,x} = \sup_{1\leq i\leq n}\sup_{t\in\mathcal{T}}|x_i(t)|$. Next,

$$\begin{aligned}
&\sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}\mathrm{Var}\left(h(T_{i,j})\,x_i(T_{i,j})\right) \\
&\qquad\leq \sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}\int\left(\frac{1}{N}\sum_{i=1}^n N_i x_i^2(u)\right)h^2(u)\rho(u)du \leq B_{n,x}^2 := v,
\end{aligned}$$

using that $\int h^2(u)\rho(u)du = 1$. Lastly

$$\sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}\sup_{1\leq i\leq n}\sup_{t\in\mathcal{T}}(|x_i(t)h(t)|) \leq \frac{B_{n,x}\sqrt{L(\boldsymbol{m})}}{\sqrt{\rho_0}} := \mathbb{M}.$$

Let $\mathbb{E}_x$ denote the conditional expectation operator given $X_i = x_i$, $1 \le i \le n$.Then, we obtain that, for $L(\boldsymbol{m}) = m^d \le N$,

$$\mathbb{E}_x\left[\left(\sup_{h\in S_{\boldsymbol{m}},\|h\|_\rho=1}[\nu_N^{(x)}(h)]^2 - 4\frac{m^d}{N}B_{n,x}^2\right)_+\right] \le \frac{C_1}{N}B_{n,x}^2\left\{\exp\left(-C_2\, m^d\right) + \exp\left(-C_3\sqrt{N}\right)\right\},$$

where C_1, C_2, C_3 are numerical constants which do not depend on $B_{n,x}$, but possibly depending on ρ_0. We can decondition since nothing in the exponentials depends on the x_i's and get

$$\mathbb{E}\left[\left(\sup_{h\in S_{\widehat{\boldsymbol{m}}\vee\boldsymbol{m}},\|h\|_\rho=1}[\nu_N^{(X)}(h)]^2 - 4\frac{(m\vee\widehat{m})^d}{N}B_{n,X}^2\right)_+\right]$$
$$\le C_1\frac{\mathbb{E}\big(B_{n,X}^2\big)}{N}\left(\sum_{m^d\le N}\exp(-C_2m^d) + \exp(-C_3\sqrt{N})\right) \le C'\,\frac{\mathbb{E}\big(B_{n,X}^2\big)}{N}.$$

This ends the proof of Lemma 5. □

Proof of Theorem 3. We treat separately the bound on the events Ω_N and Ω_N^c.

Bound on the event Ω_N. Similarly to the decomposition used in the proof of Theorem 1, on the event Ω_N defined in (47), for all $\boldsymbol{m} \in \widehat{\mathcal{M}}_N$, we have:

$$\begin{aligned}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \;\le\; & \|g_{\boldsymbol{m}} - g\|_N^2 + \mathrm{pen}(\boldsymbol{m}) + \frac{2}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}(\widehat{g}_{\widehat{\boldsymbol{m}}} - g_m)(T_{i,j})X_i(T_{i,j})\\ & + \frac{2}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}(\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}})(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j} - \mathrm{pen}(\widehat{\boldsymbol{m}}).\end{aligned}$$

Next, we write

$$\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}(\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}})(T_{i,j})X_i(T_{i,j}) \le \|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_\rho \sup_{h\in S_{\boldsymbol{m}\vee\widehat{\boldsymbol{m}}},\|h\|_\rho=1}\left|\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}h(T_{i,j})X_i(T_{i,j})\right|,$$

and

$$\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}(\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}})(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}$$
$$\le \|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_N \sup_{h\in S_{\boldsymbol{m}\vee\widehat{\boldsymbol{m}}},\,\|h\|_N=1}\left|\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}h(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}\right|.$$

Then by the inequality $2|ab| \le a^2/\theta + \theta b^2$ for $\theta = 8$, and $\theta = 16$, we get

$$\begin{aligned}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \;\le\; & \|g_{\boldsymbol{m}} - g\|_N^2 + \mathrm{pen}(\boldsymbol{m}) + \frac{1}{16}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_\rho^2 + \frac{1}{8}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_N^2\\ & + 16\sup_{h\in S_{\boldsymbol{m}\vee\widehat{\boldsymbol{m}}},\|h\|_\rho=1}\left(\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}h(T_{i,j})X_i(T_{i,j})\right)^2\\ & + 8\sup_{h\in S_{\boldsymbol{m}\vee\widehat{\boldsymbol{m}}},\|h\|_N=1}\left(\frac{1}{N}\sum_{i=1}^n\sum_{j=1}^{N_i}h(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}\right)^2 - \mathrm{pen}(\widehat{\boldsymbol{m}}).\end{aligned}$$

Note that due to Lemma 3 and since in our isotropic context the spaces $S_{\boldsymbol{m}}$ are nested, it holds $\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_\rho^2 \leq 2\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_N^2$. Thus, we get that, on Ω_N, for all $\boldsymbol{m} \in \widehat{\mathcal{M}}_N$,

$$\begin{aligned}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \quad \leq \quad & \|g_{\boldsymbol{m}} - g\|_N^2 + \mathrm{pen}(\boldsymbol{m}) + \frac{1}{4}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_N^2 \\ & +16 \sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) X_i(T_{i,j}) \right)^2 \\ & +8 \sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) \sigma_\varepsilon(T_{i,j}) \varepsilon_{i,j} \right)^2 - \mathrm{pen}(\widehat{\boldsymbol{m}}).\end{aligned}$$

Now we use that

$$\frac{1}{4}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g_{\boldsymbol{m}}\|_N^2 \leq \frac{1}{2}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 + \frac{1}{2}\|g_{\boldsymbol{m}} - g\|_N^2$$

to gather similar terms and introduce the two functions $p_1(\boldsymbol{m}, \boldsymbol{m}')$ and $p_2(\boldsymbol{m}, \boldsymbol{m}')$ defined in Lemmas 5 and 4, and deduce

$$\begin{aligned}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \quad \leq \quad & 3\|g_{\boldsymbol{m}} - g\|_N^2 + 2\mathrm{pen}(\boldsymbol{m}) + \\ & +32\Bigg(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) X_i(T_{i,j}) \right)^2 - p_1(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \Bigg)_+ \\ & +16\Bigg(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) \sigma_\varepsilon(T_{i,j}) \varepsilon_{i,j} \right)^2 - p_2(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \Bigg)_+ \\ & +32p_1(\boldsymbol{m}, \widehat{\boldsymbol{m}}) + 16p_2(\boldsymbol{m}, \widehat{\boldsymbol{m}}) - 2\mathrm{pen}(\widehat{\boldsymbol{m}}).\end{aligned}$$

Using that $32p_1(\boldsymbol{m}, \widehat{\boldsymbol{m}}) + 16p_2(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \leq 2\mathrm{pen}(\widehat{\boldsymbol{m}}) + 2\mathrm{pen}(\boldsymbol{m})$ for $\kappa \geq 64$, we get

$$\begin{aligned}\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \quad \leq \quad & 3\|g_{\boldsymbol{m}} - g\|_N^2 + 4\mathrm{pen}(\boldsymbol{m}) + \\ & +32\Bigg(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) X_i(T_{i,j}) \right)^2 - p_1(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \Bigg)_+ \\ & +16\Bigg(\sup_{h \in S_{\widehat{\boldsymbol{m}} \vee \boldsymbol{m}}, \|h\|_\rho = 1} \left(\frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{N_i} h(T_{i,j}) \sigma_\varepsilon(T_{i,j}) \varepsilon_{i,j} \right)^2 - p_2(\boldsymbol{m}, \widehat{\boldsymbol{m}}) \Bigg)_+ .\end{aligned}$$

The last two terms in the previous inequality are bounded using Lemmas 5 and 4.

Bound on the event Ω_N^c. Note that, as a consequence of Lemma 1, and c_p as in (21),

$$\mathbb{P}(\Omega_N^c) \leq c\,\mathrm{Card}(\mathcal{M}_N^+) N^{-p}.$$

Since by definition, $\mathrm{Card}(\mathcal{M}_N^+) \leq N$, we obtain $\mathbb{P}(\Omega_N^c) \lesssim N^{-p+1}$. Next, we write

$$\mathbb{E}\big[\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \mathbf{1}_{\Omega_N^c}\big] \leq 2\sqrt{2}\Big\{ \sqrt{\mathbb{E}[\|\widehat{g}_{\widehat{\boldsymbol{m}}}\|_N^4]} + \sqrt{\mathbb{E}[\|g\|_N^4]} \Big\} \sqrt{\mathbb{P}(\Omega_N^c)}.$$

It holds

$$\mathbb{E}\big[\|g\|_N^4\big] = \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{N_i} g^2(T_{i,j}) \right)^2 \right] \leq \int g^4(u)\rho(u)du.$$

Next,

$$\|\widehat{g}_{\widehat{\boldsymbol{m}}}\|_N^2 = \frac{1}{N^2} \mathbf{Y}^\top \widehat{\Phi}_{\widehat{\boldsymbol{m}}} \widehat{\Psi}_{\widehat{\boldsymbol{m}}}^{-1} \widehat{\Psi}_{\widehat{\boldsymbol{m}}} \widehat{\Psi}_{\widehat{\boldsymbol{m}}}^{-1} \widehat{\Phi}_{\widehat{\boldsymbol{m}}}^\top \mathbf{Y} = \frac{1}{N^2} \mathbf{Y}^\top \widehat{\Phi}_{\widehat{\boldsymbol{m}}} \widehat{\Psi}_{\widehat{\boldsymbol{m}}}^{-1} \widehat{\Phi}_{\widehat{\boldsymbol{m}}}^\top \mathbf{Y}.$$

As $\widehat{\boldsymbol{m}} \in \widehat{\mathcal{M}}_N$, we have $L(\widehat{\boldsymbol{m}})\|\widehat{\Psi}_{\widehat{\boldsymbol{m}}}^{-1}\|_{\mathrm{op}} \leq c_p N/\log(N)$ and we note that

$$\|\widehat{\Phi}_{\widehat{\boldsymbol{m}}}\widehat{\Phi}_{\widehat{\boldsymbol{m}}}^{\top}\|_{\mathrm{op}} = \sup_{\|x\|_{\mathbb{R}^N}^2=1} x^{\top}\widehat{\Phi}_{\widehat{\boldsymbol{m}}}\widehat{\Phi}_{\widehat{\boldsymbol{m}}}^{\top} x \leq NL(\widehat{\boldsymbol{m}}).$$

Thus,

$$\|\widehat{g}_{\widehat{m}}\|_N^2 \leq \frac{c_p}{\log(N)}\mathbf{Y}^{\top}\mathbf{Y}.$$

It follows that

$$\mathbb{E}[\|\widehat{g}_{\widehat{\boldsymbol{m}}}\|_N^4] \leq c_p^2\mathbb{E}\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{N_i} Y_{i,j}^2\right)^2\right] \leq c_p^2 N^2\mathbb{E}(Y_{1,1}^4).$$

As a consequence, taking $p=5$ yields $\mathbb{E}[\|\widehat{g}_{\widehat{\boldsymbol{m}}} - g\|_N^2 \mathbf{1}_{\Omega_N^c}] \lesssim N^{-1}$. This ends the proof. □

6 Appendix

6.1 Complements on the error variance estimation

Here, we detail and complete the discussion from Section 2.5.2 on the error variance estimation. Below we assume that g is bounded, $\sup_{t\in\mathcal{T}}\mathbb{E}[X_1^4(t)] < \infty$ and $\mathbb{E}[\varepsilon_{i,j}^4] < \infty$.

For $1 \leq j \leq J-1$, let

$$A_j := g(T_{j+1}) - g(T_j), \qquad B_j := X_1(T_{j+1}) - X_1(T_j), \qquad D_j := \varepsilon_{1,j+1} - \varepsilon_{1,j}.$$

By construction, $Y_{1,j+1} - Y_{1,j} = A_j + B_j + \sigma_\varepsilon D_j$, with A_j deterministic, $\mathbb{E}[B_j] = 0$, $\mathbb{E}[D_j] = 0$, $\mathbb{E}[D_j^2] = 2$, and B_j independent of D_j (by Assumption 1(v)). Hence

$$E_j := \mathbb{E}\big[(Y_{1,j+1} - Y_{1,j})^2\big] = A_j^2 + \mathbb{E}[B_j^2] + 2\sigma_\varepsilon^2.$$

For $V_{\varepsilon,n}$ in (11), averaging $E_j/2$ over j gives

$$0 \leq \mathbb{E}[V_{\varepsilon,n}] = \sigma_\varepsilon^2 + \frac{1}{2(J-1)}\sum_{j=1}^{J-1}\mathbb{E}[B_j^2] + \frac{1}{2(J-1)}\sum_{j=1}^{J-1} A_j^2,$$

so that

$$0 \leq \mathbb{E}[V_{\varepsilon,n}] \leq \sigma_\varepsilon^2 + \frac{1}{2}\max_{1\leq j\leq J-1}\mathbb{E}[B_j^2] + \frac{1}{2}\max_{1\leq j\leq J-1} A_j^2. \tag{48}$$

We now turn to $\mathrm{Var}(V_{\varepsilon,n})$, and first we write the centered squared increment as

$$W_j := (Y_{1,j+1} - Y_{1,j})^2 - E_j = U_j + V_j + N_j,$$

where

$$U_j := B_j^2 - \mathbb{E}[B_j^2] + 2A_jB_j, \qquad V_j := 2\sigma_\varepsilon D_j(A_j + B_j), \qquad N_j := \sigma_\varepsilon^2(D_j^2 - 2),$$

and let $Z_{j,j'} := W_jW_{j'}$. Then

$$\mathrm{Var}(V_{\varepsilon,n}) = \frac{1}{4n}\frac{1}{(J-1)^2}\left[\sum_{|j-j'|\geq 2}\mathbb{E}[Z_{j,j'}] + \sum_{1\leq j\leq J-1}\{\mathbb{E}[Z_{j,j}] + \mathbb{E}[Z_{j,j+1}]\}\right]. \tag{49}$$

We first investigate the second sum, over j. Since D_j is independent of (A_j, B_j) and $\mathbb{E}[D_j] = \mathbb{E}[D_j^2 - 2] = 0$, the cross products $\mathbb{E}[U_jV_j]$, $\mathbb{E}[U_jN_j]$ and $\mathbb{E}[V_jN_j]$ all vanish (the last one using $\mathbb{E}[D_j^3] = 0$, which is a consequence of the definition of D_j). Thus

$$\mathbb{E}[W_j^2] = \mathbb{E}[U_j^2] + \mathbb{E}[V_j^2] + \mathbb{E}[N_j^2].$$

Using $(x+y)^2 \le 2(x^2+y^2)$ and Jensen's inequality, we get $\mathbb{E}[U_j^2] \lesssim A_j^4 + \mathbb{E}[B_j^4]$. Next, using $\mathbb{E}[D_j^2] = 2$, we get $\mathbb{E}[V_j^2] = 8\sigma_\varepsilon^2(A_j^2 + \mathbb{E}[B_j^2])$. Finally, using $\mathbb{E}[D_j^4] = 2\mathbb{E}[\varepsilon_{1,j}^4] + 6$ (a consequence of the independence and the standardization of $\varepsilon_{1,j}, \varepsilon_{1,j+1}$), it holds $\mathbb{E}[N_j^2] = \sigma_\varepsilon^4\big(\mathbb{E}[D_j^4] - 4\big) = 2\sigma_\varepsilon^4(\mathbb{E}[\varepsilon_{1,j}^4] + 1)$. Combining the three terms gives

$$\mathbb{E}[W_j^2] \lesssim A_j^4 + \mathbb{E}[B_j^4] + \mathbb{E}[\varepsilon_{1,j}^4] + 1, \tag{50}$$

with the constant in $\lesssim$ depending on σ_ε. As $Z_{j,j} = W_j^2$, this directly bounds $\mathbb{E}[Z_{j,j}]$. For the adjacent term, D_j and D_{j+1} share $\varepsilon_{1,j+1}$ and are not independent, so we instead use Cauchy-Schwarz with $\sqrt{ab} \le (a+b)/2$ together with (50):

$$\big|\mathbb{E}[Z_{j,j+1}]\big| \le \big\{\mathbb{E}(W_j^2)\mathbb{E}(W_{j+1}^2)\big\}^{1/2} \le \frac{1}{2}\big[\mathbb{E}(W_j^2) + \mathbb{E}(W_{j+1}^2)\big].$$

We deduce that

$$\frac{1}{J-1}\left|\sum_{1\le j\le J-1}\{\mathbb{E}[Z_{j,j}] + \mathbb{E}[Z_{j,j+1}]\}\right| \lesssim \mathbb{E}[\varepsilon_{1,1}^4] + \max_{1\le j\le J-1} A_j^4 + \max_{1\le j\le J-1}\mathbb{E}[B_j^4] + 1. \tag{51}$$

For $|j - j'| \ge 2$, D_j is independent of $(B_j, B_{j'}, D_{j'})$ jointly (since $\varepsilon_{1,j}, \varepsilon_{1,j+1}$ share no index with $\varepsilon_{1,j'}, \varepsilon_{1,j'+1}$). Expanding $Z_{j,j'} = W_jW_{j'} = (U_j + V_j + N_j)(U_{j'} + V_{j'} + N_{j'})$, every cross term other than $U_jU_{j'}$ contains a factor D_j or $D_{j'}$ (through $V_j, N_j, V_{j'}$ or $N_{j'}$) that is independent of the remaining factors and centered ($\mathbb{E}[D_j] = \mathbb{E}[D_j^2 - 2] = 0$); each such term therefore vanishes in expectation. Hence

$$\mathbb{E}[Z_{j,j'}] = \mathbb{E}[U_jU_{j'}], \qquad |j - j'| \ge 2,$$

with no contribution from the noise. By Cauchy–Schwarz and $\sqrt{ab} \le (a+b)/2$,

$$\big|\mathbb{E}[U_jU_{j'}]\big| \le \frac{1}{2}\big[\mathbb{E}(U_j^2) + \mathbb{E}(U_{j'}^2)\big] \lesssim A_j^4 + A_{j'}^4 + \mathbb{E}[B_j^4] + \mathbb{E}[B_{j'}^4].$$

We deduce the simplified bound

$$\frac{1}{(J-1)^2}\left|\sum_{|j-j'|\ge 2}\mathbb{E}[Z_{j,j'}]\right| \lesssim \max_{1\le j\le J-1} A_j^4 + \max_{1\le j\le J-1}\mathbb{E}[B_j^4]. \tag{52}$$

Combining (49), (51) and (52), and using (48) for the bias, we deduce, without any further assumption on the regularity of g or the field X_1, that

$$\mathbb{E}\big[\{V_{\varepsilon,n} - \sigma_\varepsilon^2\}^2\big] \le C\left[n^{-1} + 1\right],$$

for C a constant depending on $\|g\|_\infty$, $\sup_{t\in\mathcal{T}}\mathbb{E}[X_1^4(t)]$ and $\mathbb{E}[\varepsilon_{i,j}^4]$.

The consistency of $V_{\varepsilon,n}$ as J increases can be obtained by imposing some mild additional assumptions. Before introducing these assumptions, let us assume that the set of common design points $\{T_j\}$ is ordered in the following way: set T_1 to be the point closest to the origin in Euclidean norm. Then, for each $j \ge 2$, T_j is defined as the point closest to T_{j-1} among the remaining design points. Ties are broken by lexicographic order. With this ordering, we assume that the Euclidean distance between any two consecutive points is bounded by $C_d(J^{1/d} - 1)^{-1}$, where C_d is a constant depending only on d, and without loss of generality we consider $J \ge 2^d$. Finally, the $Y_{i,j}$ are re-indexed if necessary according to the ordering of the T_j.

Proposition 4. *Assume that for some $\beta > 0$, $\mathbb{E}[\Lambda_X^4(\beta)]$ is finite, where*

$$\Lambda_X(\beta) = \sup_{s,t\in[0,1]^d, s\neq t} \frac{|X_1(s) - X_1(t)|}{\|s-t\|^\beta}, \tag{53}$$

and that, for some $\xi > 0$,

$$L_g = \sup_{s,t\in\mathcal{T}, s\neq t} \frac{|g(s) - g(t)|}{\|s-t\|^\xi} < \infty. \tag{54}$$

Then, there exists a constant $C' > 0$ such that

$$\mathbb{E}\big[\{V_{\varepsilon,n} - \sigma_\varepsilon^2\}^2\big] \le C' \left[\frac{1}{nJ^\nu} + \frac{1}{J^{4(\beta\wedge\xi)/d}}\right], \tag{55}$$

where $\nu = \min(1, 4\beta/d, 4\xi/d)$. The constant C' depends on d, L_g, $\|g\|_\infty$, and the moments of order 4 of $\Lambda_X(\beta)$ and $\varepsilon_{i,j}$ (but not on J or n).

Proof of Proposition 4. From (53), for $1 \le j \le J-1$,

$$|B_j| = |X_1(T_{j+1}) - X_1(T_j)| \le \Lambda_X(\beta)\,\|T_{j+1} - T_j\|^\beta \le \frac{C_d^\beta}{(J^{1/d}-1)^\beta}\Lambda_X(\beta),$$

so that, taking second and fourth moments,

$$\mathbb{E}[B_j^2] \le \frac{C_d^{2\beta}}{(J^{1/d}-1)^{2\beta}}\mathbb{E}[\Lambda_X^2(\beta)], \qquad \mathbb{E}[B_j^4] \le \frac{C_d^{4\beta}}{(J^{1/d}-1)^{4\beta}}\mathbb{E}[\Lambda_X^4(\beta)]. \tag{56}$$

Similarly, from (54),

$$|A_j| = |g(T_{j+1}) - g(T_j)| \le \frac{C_d^\xi}{(J^{1/d}-1)^\xi} L_g,$$

$$\text{so that} \qquad A_j^2 \le \frac{C_d^{2\xi}}{(J^{1/d}-1)^{2\xi}} L_g^2, \quad A_j^4 \le \frac{C_d^{4\xi}}{(J^{1/d}-1)^{4\xi}} L_g^4. \tag{57}$$

Plugging the second-moment bounds in (56)-(57) into (48) gives the bias bound

$$\big|\mathbb{E}[V_{\varepsilon,n}] - \sigma_\varepsilon^2\big| \lesssim \frac{1}{(J^{1/d}-1)^{2\beta}}\mathbb{E}[\Lambda_X^2(\beta)] + \frac{1}{(J^{1/d}-1)^{2\xi}} L_g^2. \tag{58}$$

For the variance, plugging the fourth-moment bounds into (51) and (52) gives

$$\frac{1}{(J-1)^2}\left|\sum_{|j-j'|\ge 2} \mathbb{E}[Z_{j,j'}]\right| \lesssim (J^{1/d}-1)^{-4\beta} + (J^{1/d}-1)^{-4\xi},$$

$$\text{and} \qquad \frac{1}{J-1}\left|\sum_{1\le j\le J-1} \{\mathbb{E}[Z_{j,j}] + \mathbb{E}[Z_{j,j+1}]\}\right| \lesssim 1,$$

where the second bound follows because $\mathbb{E}[\varepsilon_{1,1}^4]$ and the additive constant in (51) do not decay with J, while the A_j^4 and $\mathbb{E}[B_j^4]$ contributions are dominated by this constant. Combining these two bounds in (49),

$$\mathrm{Var}(V_{\varepsilon,n}) \lesssim \frac{1}{n}\Big[J^{-4\beta/d} + J^{-4\xi/d} + J^{-1}\Big]. \tag{59}$$

Since $\mathbb{E}\big[\{V_{\varepsilon,n} - \sigma_\varepsilon^2\}^2\big] = \mathrm{Var}(V_{\varepsilon,n}) + \big(\mathbb{E}[V_{\varepsilon,n}] - \sigma_\varepsilon^2\big)^2$, combining (58) and (59), the bound (55) follows. □

Note that a moment of order 4 of $\Lambda_X(\beta)$ is finite whenever there exist $\gamma, \zeta > 0$ such that $\gamma \geq 4$, $\zeta > \gamma\beta$ and

$$\mathbb{E}\big[|X_1(t) - X_1(s)|^\gamma\big] \lesssim \|t - s\|^{d+\zeta}, \quad \forall s, t \in [0,1]^d.$$

See [33, Th.2.1]. The bound (55) proves the consistency of $V_{\varepsilon,n}$ as an estimator of σ_ε^2 as $J \to \infty$.

In the heteroscedastic case, that is, if σ_ε^2 is not constant, let

$$W_{\varepsilon,n}(j) = \frac{1}{2n} \sum_{i=1}^{n} (Y_{i,j+1} - Y_{i,j})^2.$$

The estimator of $\|\sigma_\varepsilon^2\|_\infty$ can be defined as

$$V_{\varepsilon,n} = \max_{1 \leq j \leq J-1} W_{\varepsilon,n}(j). \tag{60}$$

Note that, for all $1 \leq j \leq J-1$,

$$\begin{aligned}\mathbb{E}[W_{\varepsilon,n}(j)] &= \{\sigma_\varepsilon^2(T_{j+1}) + \sigma_\varepsilon^2(T_j)\}/2 + \mathbb{E}\big(|X_1(T_{j+1}) - X_1(T_j)|^2\big)/2 + |g(T_{j+1}) - g(T_j)|^2 \\ &= \sigma_\varepsilon^2(\{T_{j+1} + T_j\}/2) + R_\varepsilon(j)/2 + \mathbb{E}\big(|X_1(T_{j+1}) - X_1(T_j)|^2\big)/2 \\ &\qquad + |g(T_{j+1}) - g(T_j)|^2,\end{aligned}$$

where,

$$R_\varepsilon(j) = \max\big\{|\sigma_\varepsilon^2(\{T_{j+1} + T_j\}/2) - \sigma_\varepsilon^2(T_j)|,\, |\sigma_\varepsilon^2(\{T_{j+1} + T_j\}/2) - \sigma_\varepsilon^2(T_{j+1})|\big\}.$$

The remainder $R_\varepsilon(j)$ can be controlled by a Hölder continuity assumption on the function σ_ε^2. Using assumptions on X_i and g as in the homoscedastic case, $\mathbb{E}[W_{\varepsilon,n}(j)] - \sigma_\varepsilon^2(\{T_{j+1} + T_j\}/2)$ can be bounded by $C(J^{1/d} - 1)^{-\alpha}$ for some $C, \alpha > 0$ not depending on j. Moreover, the Hölder continuity of σ_ε^2 allows to similarly bound the absolute difference between $\max_{1 \leq j \leq J-1} \mathbb{E}[W_{\varepsilon,n}(j)]$ and $\|\sigma_\varepsilon^2\|_\infty$. Finally, under suitable moment assumptions on the increments of X_i and the error $\varepsilon_{i,j}$, the independence assumptions along the i index allow one to derive a concentration bound for $W_{\varepsilon,n}(j)$ around its mean. Combining these facts, and provided n and J are such that the union bound applied to the concentration bound is negligible, the risk of the estimator in (60) can again be controlled. Then, one can conclude once more that Theorem 1 remains valid modulo negligible terms if the selection step (10) is replaced by that in (13). We omit the details of this analysis in the heteroscedastic case for brevity.

6.2 Proofs of technical lemmas

In this section we provide the proofs for Lemmas 3 and 4 used in the proof of Theorem 3.

Proof of Lemma 3. Let $\omega \in \Omega_N$ and, with the notation from the proof of Lemma 1, let

$$\widehat{\mathbf{G}}_{\boldsymbol{m}}(\omega) := \mathbf{H}_{\boldsymbol{m}}^{-\frac{1}{2}} \widehat{\mathbf{H}}_{\boldsymbol{m}}(\omega) \mathbf{H}_{\boldsymbol{m}}^{-\frac{1}{2}}, \quad \forall \boldsymbol{m} \in \mathcal{M}_N^+.$$

For any $\boldsymbol{m} \in \mathcal{M}_N^+$, since $\omega \in \Omega_{\boldsymbol{m}}$,

$$\mathrm{Sp}\big(\widehat{\mathbf{G}}_{\boldsymbol{m}}(\omega)\big) \subset [1/2, 3/2], \quad \text{and then} \quad \mathrm{Sp}\big(\widehat{\mathbf{G}}_{\boldsymbol{m}}^{-1}(\omega)\big) \subset [2/3, 2].$$

Moreover,

$$\widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}(\omega) = \mathbf{H}_{\boldsymbol{m}}^{-\frac{1}{2}} \widehat{\mathbf{G}}_{\boldsymbol{m}}^{-1}(\omega) \mathbf{H}_{\boldsymbol{m}}^{-\frac{1}{2}}.$$

So, thanks to a well-known property of the Loewner order,

$$\frac{2}{3}\mathbf{x}^\top \mathbf{H}_{\boldsymbol{m}}^{-1}\mathbf{x} \leqslant \mathbf{x}^\top \widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}(\omega)\mathbf{x} \leqslant 2\mathbf{x}^\top \mathbf{H}_{\boldsymbol{m}}^{-1}\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^{D_{\boldsymbol{m}}},$$

leading to

$$\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} = \|\mathbf{H}_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}} \leqslant \frac{3}{2}\|\widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}(\omega)\|_{\mathrm{op}} = \frac{3}{2}\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}(\omega)\|_{\mathrm{op}}$$

and

$$\|\widehat{\Psi}_{\boldsymbol{m}}^{-1}(\omega)\|_{\mathrm{op}} = \|\widehat{\mathbf{H}}_{\boldsymbol{m}}^{-1}(\omega)\|_{\mathrm{op}} \leqslant 2\|\mathbf{H}_{m}^{-1}\|_{\mathrm{op}} = 2\|\Psi_{\boldsymbol{m}}^{-1}\|_{\mathrm{op}}.$$

Therefore, $\omega \in \Xi_N$. □

Proof of Lemma 4. We can proceed conditionally on the $T_{i,j}$ (independent of the $\varepsilon_{i,j}$), as in the proof of Lemma 2 for the fixed design case: we split the empirical process into two parts by truncating the $\varepsilon_{i,j}$'s, then apply Talagrand's Inequality (Theorem 5) to the main part. For $h \in S_{\boldsymbol{m}}$, let

$$\nu_N^{(\varepsilon)}(h) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} h(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j} = \nu_N^{(1,\varepsilon)}(h) + \nu_N^{(2,\varepsilon)}(h)$$

corresponding to the splitting $\varepsilon_{i,j} = \varepsilon_{i,j}^{(1)} + \varepsilon_{i,j}^{(2)}$ with

$$\varepsilon_{i,j}^{(1)} = \varepsilon_{i,j}\mathbf{1}_{\{|\varepsilon_{i,j}|\leq B_N\}} - \mathbb{E}(\varepsilon_{i,j}\mathbf{1}_{\{|\varepsilon_{i,j}|\leq B_N\}}) \text{ and } \varepsilon_{i,j}^{(2)} = \varepsilon_{i,j} - \varepsilon_{i,j}^{(1)}.$$

Similarly to what has been done in the proof of Lemma 2, it holds

$$\mathbb{E}\left(\sup_{h\in S_{\boldsymbol{m}}, \|h\|_N=1} [\nu_N^{(2,\varepsilon)}(h)]^2\right) \leq \|\sigma_\varepsilon^2\|_\infty \frac{\mathbb{E}\big(\varepsilon_{1,1}^{2(p+1)}\big)}{B_N^{2p}}. \tag{61}$$

For $\nu_N^{(1,\varepsilon)}(h)$, we intend to apply Talagrand's Inequality. We have, conditionally on the $T_{i,j}$,

$$\mathbb{E}\left(\sup_{h\in S_{\boldsymbol{m}}, \|h\|_N=1} [\nu_N^{(1,\varepsilon)}(h)]^2\right) \leq \|\sigma_\varepsilon^2\|_\infty \frac{m^d}{N} := \mathbb{H}^2.$$

Next, still conditionally on the $T_{i,j}$,

$$\frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{N_i} \mathrm{Var}(h(T_{i,j})\sigma_\varepsilon(T_{i,j})\varepsilon_{i,j}^{(1)}) \leq \|\sigma_\varepsilon^2\|_\infty := v.$$

Lastly, on Ω_N, we have $\|h\|_\rho^2 \leq 2\|h\|_N^2$ and under Condition (17) $\|h\|^2 \leq \|h\|_\rho^2/\rho_0$, so for $h \in S_{\boldsymbol{m}}$ such that $\|h\|_N^2 = 1$, it holds that $\|h\|_\infty \leq \sqrt{m^d}\|h\| \leq \sqrt{2m^d/\rho_0}$. Thus, conditionally on the $T_{i,j}$, we get

$$\sup_{h\in S_{\boldsymbol{m}},\, \|h\|_N=1} \sup_{|x|\leq 2B_N} \big|h(T_{i,j})\sigma_\varepsilon(T_{i,j})x\big| \leq 2B_N\sqrt{\frac{2m^d\|\sigma_\varepsilon^2\|_\infty}{\rho_0}} := \mathbb{M}.$$

Therefore, conditionally on the $T_{i,j}$ and on Ω_N, applying Talagrand's Inequality (Theorem 5) yields:

$$\begin{aligned}
&\mathbb{E}\left(\sup_{h\in S_{\boldsymbol{m}}, \|h\|_N=1} [\nu_N^{(1,\varepsilon)}(h)]^2 - 4\|\sigma_\varepsilon^2\|_\infty\frac{m^d}{N}\right)_+ \\
&\quad\leq \frac{C_1}{N}\left(\|\sigma_\varepsilon^2\|_\infty \exp\left(-C_2 m^d\right) + \frac{2m^d B_N^2\|\sigma_\varepsilon^2\|_\infty}{N\rho_0}\exp\left(-C_3\sqrt{\frac{\rho_0}{\|\sigma_\varepsilon^2\|_\infty}}\frac{\sqrt{N}}{B_N}\right)\right).
\end{aligned}$$

Now we choose

$$B_N = \frac{C_3}{2}\sqrt{\frac{\rho_0}{\|\sigma_\varepsilon^2\|_\infty}}\frac{\sqrt{N}}{\log(N)},$$

and get, using that $m^d \leq cN$,

$$\mathbb{E}\Big(\sup_{h\in S_{\boldsymbol{m}},\|h\|_N=1}[\nu_N^{(1,\varepsilon)}(h)]^2 - 4\|\sigma_\varepsilon^2\|_\infty\frac{m^d}{N}\Big)_+ \leq \frac{C_1}{N}\Big(\|\sigma_\varepsilon^2\|_\infty \exp\big(-C_2 m^d\big) + \frac{cC_3^2}{2}\frac{1}{N}\Big).$$

Combining the terms yields an upper bound of order $1/N$, and coming back to (61) gives for $p=1$

$$\mathbb{E}\Big(\sup_{h\in S_{\boldsymbol{m}},\|h\|_N=1}[\nu_N^{(2,\varepsilon)}(h)]^2\Big) \leq \frac{4}{C_3^2\rho_0}\|\sigma_\varepsilon^2\|_\infty^2\mathbb{E}(\varepsilon_{1,1}^4)\frac{\log^2(N)}{N}.$$

Gathering the terms and deconditioning ends the proof of Lemma 4. □

Lemma 6. *Assume that there exist $\sigma^2, c > 0$ such that*

$$\mathbb{P}\Big(\sup_{t\in\mathcal{T}}\big|X_1(t)\big| > x\Big) \leq c\exp\big(-x^2/(2\sigma^2)\big), \qquad \forall x \geq 0.$$

Then, for any $\alpha \in (0,2]$, there exists n_0 (depending on c,α) such that, for all $n \geq n_0$,

$$\mathbb{E}\left[\max_{1\leq i\leq n}\sup_{t\in\mathcal{T}}\big|X_i(t)\big|^\alpha\right] \leq (3\sigma)^\alpha\{\log n\}^{\alpha/2}. \tag{62}$$

Proof of Lemma 6. Define $U_i(\alpha) = \sup_{t\in\mathcal{T}}\big|X_i(t)\big|^\alpha$, which are i.i.d. because the fields X_i are i.i.d. Then, for any $\mu > 0$,

$$\begin{aligned}
\mathbb{E}\Big[\max_{1\leq i\leq n}U_i(\alpha)\Big] &= \int_0^\infty \mathbb{P}\Big(\max_{1\leq i\leq n}U_i(\alpha) > x\Big)\,dx = \int_0^\infty \mathbb{P}\Big(\max_{1\leq i\leq n}U_i(\alpha)^{1/\alpha} > x^{1/\alpha}\Big)\,dx \\
&\leq \int_0^\mu \underbrace{\mathbb{P}\Big(\max_{1\leq i\leq n}U_i(\alpha)^{1/\alpha} > x^{1/\alpha}\Big)}_{\leq 1}\,dx + cn\int_\mu^\infty \exp\big(-x^{2/\alpha}/(2\sigma^2)\big)\,dx \\
&\leq \mu + cn\alpha\sigma^2\mu^{1-2/\alpha}\exp\big(-\mu^{2/\alpha}/(2\sigma^2)\big).
\end{aligned}$$

With $\mu = \{4\sigma^2\log n\}^{\alpha/2}$ we obtain

$$\begin{aligned}
\mathbb{E}\left[\max_{1\leq i\leq n}U_i(\alpha)\right] &\leq \{4\sigma^2\log n\}^{\alpha/2} + cn\alpha\sigma^2\{4\sigma^2\log n\}^{\alpha/2-1}\exp\big(-\{4\sigma^2\log n\}/(2\sigma^2)\big) \\
&= (2\sigma)^\alpha\log^{\alpha/2}n + c\frac{\alpha\sigma^2}{n}\{4\sigma^2\log n\}^{\alpha/2-1}.
\end{aligned}$$

Since the second term equals $(2\sigma)^\alpha\log^{\alpha/2}n$ times $c\alpha/(4n\log n) \to 0$, there exists n_0 (depending on c,α) such that, for all $n \geq n_0$, the inequality (62) holds. □

6.3 Concentration Inequalities

For the sake of completeness, we recall the following deviation inequality for the extreme eigenvalues of a sum of positive semi-definite matrices; see [35, Theorem 1.1].

Theorem 4 (Matrix Chernoff Inequality)**.** *Consider a finite sequence* $\{\mathbf{X}_k\}$ *of independent, random, self-adjoint matrices with dimension* D*. Assume that each random matrix satisfies*

$$\mathbf{X}_k \succeq 0 \quad \textit{and} \quad \lambda_{\max}(\mathbf{X}_k) \leq R, \quad \textit{almost surely.}$$

Define

$$\mu_{\min} := \lambda_{\min}\left(\sum_k \mathbb{E}(\mathbf{X}_k)\right) \quad \textit{and} \quad \mu_{\max} := \lambda_{\max}\left(\sum_k \mathbb{E}(\mathbf{X}_k)\right).$$

Then

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_k \mathbf{X}_k\right) \leq (1-\delta)\mu_{\min}\right) \leq D\left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu_{\min}/R} \qquad \textit{for } \delta \in [0,1],$$

$$\mathbb{P}\left(\lambda_{\max}\left(\sum_k \mathbf{X}_k\right) \geq (1+\delta)\mu_{\max}\right) \leq D\left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu_{\max}/R} \qquad \textit{for } \delta \geq 0.$$

Next, we recall Talagrand's Concentration Inequality given in [28].

Theorem 5. *Consider* $n \in \mathbb{N}_{>0}$*,* $\mathcal{F}$ *a class of at most countable many measurable functions, and* $(X_i)_{i\in\{1,\dots,n\}}$ *a family of real independent random variables. Define, for* $f \in \mathcal{F}$*,* $\nu_n(f) = (1/n)\sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)])$*, and assume that there are three positive constants* $\mathbb{M}$*,* $\mathbb{H}$ *and* v *such that*

$$\sup_{f\in\mathcal{F}} \|f\|_\infty \leq \mathbb{M}, \quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}} |\nu_n(f)|\Big] \leq \mathbb{H} \quad \textit{and} \quad \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(f(X_i)) \leq v.$$

Then for all $\alpha > 0$*,*

$$\mathbb{E}\left[\left(\sup_{f\in\mathcal{F}} |\nu_n(f)|^2 - 2(1+2\alpha)\mathbb{H}^2\right)_+\right] \leq \frac{4}{b}\left(\frac{v}{n}\exp\left(-b\alpha\frac{n\mathbb{H}^2}{v}\right) + \frac{49\mathbb{M}^2}{bC^2(\alpha)n^2}\exp\left(-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7}\frac{n\mathbb{H}}{\mathbb{M}}\right)\right),$$

with $C(\alpha) = (\sqrt{1+\alpha} - 1) \wedge 1$*, and* $b = 1/6$*.*

By density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f \to \nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

[1] R. J. Adler and J. E. Taylor. *Random Fields and Geometry.* Springer Monographs in Mathematics. Springer, New York, 2007.

[2] Y. Baraud. Model selection for regression on a fixed design. *Probab. Theory Related Fields*, 117(4):467–493, 2000.

[3] Y. Baraud. Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6:127–146, 2002.

[4] Y. Baraud, F. Comte, and G. Viennet. Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.*, 29(3):839–875, 2001.

[5] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[6] J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470, 2012.

[7] M. Berger, P. Hermann, and H. Holzmann. From dense to sparse design: optimal rates under the supremum norm for estimating the mean function in functional data analysis. *Bernoulli*, 31(3):1858–1888, 2025.

[8] T. T. Cai, D. Kim, and H. Pu. Transfer learning for functional mean estimation: Phase transition and adaptive algorithms. *Ann. Statist.*, 52(2):654–678, 2024.

[9] T. T. Cai and M. Yuan. Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Ann. Statist.*, 39(5):2330–2355, 2011.

[10] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42(4):1564–1597, 2014.

[11] A. Cohen, M. A. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13(5):819–834, 2013.

[12] A. Cohen, M. A. Davenport, and D. Leviatan. Correction to: On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 19(1):239, 2019.

[13] C. M. Crainiceanu, J. Goldsmith, A. Leroux, and E. Cui. *Functional data analysis with R*, volume 174 of *Monogr. Stat. Appl. Probab.* Boca Raton, FL: CRC Press, 2024.

[14] H. Dette, K. Kokot, and A. Aue. Functional data analysis in the Banach space of continuous functions. *Ann. Statist.*, 48(2):1168–1192, 2020.

[15] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303. Springer Science & Business Media, 1993.

[16] F. Dussap. Nonparametric multiple regression by projection on non-compactly supported bases. *Ann. Inst. Statist. Math.*, 75(5):731–771, 2023.

[17] S. Efromovich. How to overcome the curse of long-memory errors. *IEEE Trans. Inform. Theory*, 45(5):1735–1741, 1999.

[18] J. Gertheiss, D. Rügamer, B. X. W. Liew, and S. Greven. Functional data analysis: An introduction and recent developments. *Biom. J.*, 66(7):e202300363, 2024.

[19] E. Giné and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016.

[20] P. Hall and J. D. Hart. Nonparametric regression with long-range dependence. *Stochastic Process. Appl.*, 36(2):339–351, 1990.

[21] L. Hogben, editor. *Handbook of linear algebra*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, second edition, 2014.

[22] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[23] T. Hsing, T. Brown, and B. Thelen. Local intrinsic stationarity and its inference. *Ann. Statist.*, 44(5):2058 – 2088, 2016.

[24] T. Hsing and R. Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2015.

[25] O. Kassi, N. Klutchnikoff, and V. Patilea. Learning the regularity of multivariate functional data. *Electron. J. Stat.*, 19(2):4174–4229, 2025.

[26] O. Kassi and V. Patilea. Optimal inference for the mean of random functions. *arXiv:2504.11025*, 2025.

[27] G. Kerkyacharian, O. Lepski, and D. Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2):137–170, 2001.

[28] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.

[29] S. Koner and A.-M. Staicu. Second-generation functional data. *Annu. Rev. Stat. Appl.*, 10:547–572, 2023.

[30] A. Lederer, J. Umlauft, and S. Hirche. Uniform error bounds for Gaussian process regression with application to safe control. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

[31] L. Lim. Singular values and eigenvalues of tensors: A variational approach. *arXiv:math/0607648*, 2006. SCCM Technical Report 05-10.

[32] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[33] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren Math. Wiss.* Berlin: Springer, 3rd ed. edition, 2005.

[34] F. Schürzinger. Nonparametric estimation of the transition density in dependence of time. Preprint HAL-05441227, Jan. 2026.

[35] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, 12(4):389–434, 2012.

[36] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.

[37] A. W. van der Vaart and J. H. van Zanten. Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, 36(3):1435–1463, 2008.

[38] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes—with applications to statistics*. Springer Series in Statistics. Springer, Cham, second edition, 2023.

[39] M. Vladimirova, S. Girard, H. Nguyen, and J. Arbel. Sub-Weibull distributions: Generalizing sub-Gaussian and sub-exponential to heavier tails. *Stat*, 9(1):e318, 2020.

[40] J.-L. Wang, J.-M. Chiou, and H.-G. Müller. Functional data analysis. *Annu. Rev. Stat. Appl.*, 3:257–295, 2016.

[41] L. Xiao. Asymptotic properties of penalized splines for functional data. *Bernoulli*, 26(4):2847–2875, 2020.

[42] D. Yarger, S. Stoev, and T. Hsing. A functional-data approach to the Argo data. *Ann. Appl. Stat.*, 16(1):216 – 246, 2022.

[43] X. Zhang and J.-L. Wang. From sparse to dense functional data and beyond. *Ann. Statist.*, 44(5):2281–2321, 2016.