HAL open science

Local asymptotics of cross-validation around the optimal model

Guillaume Maillard

▶ To cite this version:

Guillaume Maillard. Local asymptotics of cross-validation around the optimal model. 2022. hal-03263396v2

HAL Id: hal-03263396

https://hal.science/hal-03263396v2

Preprint submitted on 30 Nov 2022

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.

Local asymptotics of cross-validation around the optimal model

Guillaume Maillard*

November 30, 2022

Abstract

When used to estimate the risk of a single predictor, the behaviour of cross-validation can often be understood through a central limit theorem. In model selection however, cross-validation is applied simultaneously to many different estimators in order to compare them. Thus, analyzing CV in this context requires a multi-dimensional or functional CLT. Since the mean and variance may vary widely over the model collection, careful attention must be paid to how the process is centered and scaled. In this article, we conduct the first such analysis of cross-validation in the context of least-squares density estimation by Fourier polynomials. Our results characterize the fluctuations of some CV criteria in the vicinity of the optimal model, at the critical scale at which they become significant. Asymptotically, CV is approximately the sum of a convex function and a symmetrized, time-changed Wiener process. For a slowly increasing number of folds V, the variance decreases proportionally to $\frac{1}{V}$: the folds are asymptotically independent. This analysis presents some unusual challenges which we overcome through a combination of tools including strong approximation, concentration inequalities and coupling of Gaussian vectors.

1 Introduction

Cross-validation is a widely-used class of methods for risk estimation and model selection. Despite its popularity, its behaviour is still not completely understood and much practical advice is based on heuristic arguments or numerical simulations. This sometimes leads to disagreement between experts [3], such as on the question of whether V-fold [13] or leave-1-out [8] has better performance.

The fluctuations of CV along a given sequence of predictors (indexed by the sample size) can be assessed by computing the variance and establishing a central limit theorem. Precise variance computations were carried out in various settings by Burman [8], Célisse and Robin [9], Arlot and Lerasle [4] among others - see the review paper by [3] for a more complete overview. On the other hand, central limit theorems were proved in a general setting by [11, 5] and most recently by [6]. Together, these results can provide accurate information about the deviations of cross-validation when estimating the risk of a single predictor, or sequence of predictors.

*INRIA Céleste, Université Paris-Saclay, Université du Luxembourg

However, in many cases, cross-validation is applied simultaneously to a whole class of estimators rather than just one: for example, empirical risk minimizers on a nested sequence of models, or Lasso predictors with various values of the penalty parameter. This class of predictors may change with the sample size: for example, models will typically be allowed to grow bigger and penalty parameters to become smaller as n grows. Practically relevant quantities, such as the model or parameter selected by cross-validation and its risk, depend on the whole joint distribution of cross-validation applied to the given class of predictors. The behaviour of cross-validation in the neighbourhood of the optimal model is particularly significant as we typically want and expect model selection estimators to concentrate around the optimal model, at least in the sense of having nearly optimal risk. For cross-validation based model selection, such results, known as oracle inequalities, were established by [23] and [19] in regression and by [4] in least-squares density estimation. However, oracle inequalities do not tell the whole story as they "only" provide an upper bound on the rate of concentration of the selected estimator. Ideally, we would like to characterize the fluctuations of the CV selected model and estimator around their optimal ("oracle") counterparts, as central limit theorems do for the CV risk estimator at a single model. This raises the question of establishing a multi-dimensional or functional CLT for cross-validation, at least locally around the optimal model.

In this article, we focus on the special case of least-squares density estimation, with a collection of linear models generated by cosine functions with increasing frequencies, a model collection which is both practically useful and allows for explicit computations. The study of cross-validation in that context has several original characteristics which distinguish it from textbook applications of empirical process theory, even in the simplest case of simple validation. The hold-out risk estimator, centered at the optimal model, can be written as

$$\mathrm{HO}_T(k) - \mathrm{HO}_T(k*) = \left\|\hat{s}_k^T - s\right\|^2 - \left\|\hat{s}_{k*}^T - s\right\|^2 - 2\left(P_n^{T^c} - P\right)(\hat{s}_k^T - \hat{s}_{k*}^T),$$

where $\hat{s}_k^T$ is the empirical risk minimizer on the $k-$th model, computed using the training data, s is the true density with corresponding probability measure P and $P_n^{T^c}$ is the empirical measure on the test data. Hence, it can be remarked that, though $\mathrm{HO}_T(k) - \mathrm{HO}_T(k*)$ is an empirical process *conditionally on the training data*,

- The class of functions, of the form $(\hat{s}_k)_{k\in K_n}$, is random, depends on n and the estimators $\hat{s}_k$ range over models of unbounded dimension.
- The variance, $\mathrm{Var}_P\left(\hat{s}_k^T - \hat{s}_{k*}^T\right)$, converges to 0 since, for k close enough to the optimum, $\hat{s}_k$ and $\hat{s}_{k_*}$ converge to s.
- The process, suitably renormalized, may fail to converge in distribution. If so, we want to approximate it by a sequence of "simple" Gaussian process.

In order to find a Gaussian approximation in this non-standard setting, we use the celebrated Komlós-Major-Tusnády (KMT) Theorem [15], which provides a coupling between the empirical process $(\sqrt{n}(F_n - F)(t))_{t\in\mathbb{R}}$, where F is the cdf and F_n the empirical cdf, and a Brownian bridge process $B_n(t)$, such that

$$\mathbb{E}\left[\left\|\sqrt{n}(F_n - F) - B_n \circ F\right\|_\infty\right] \leq C\frac{\log n}{\sqrt{n}},$$

for some numerical constant C. The KMT theorem can be used to approximate a general empirical process of the form $\sqrt{n}(P_n - P)(f)_{f\in\mathcal{F}}$ by the Gaussian process

$$\left(\int B_n(F(t))df(t)\right)_{f\in\mathcal{F}}, \tag{1}$$

whenever the functions in $\mathcal{F}$ are of bounded variation. This construction was first used by Hall [12] to prove a law of the iterated logarithm for kernel density estimators, then by Rio [22] in a more general empirical process setting. Other methods were later used to obtain strong approximations to general empirical processes under different assumptions [14, 7, 10].

However, the simple construction (1) is sufficient in the present setting.Its main technical advantage, in addition to its simplicity (modulo the KMT construction, used as a black box), is that the underlying Brownian bridge process can clearly be constructed independently of the class of estimators $\hat{s}_k^T$.

The resulting Gaussian process has a complicated, random covariance function. By approximating it through explicit computations and concentration inequalities, we show that the Gaussian process itself can be approximated by a symmetrized Brownian motion changed in time, W_{g_n}, where g_n is an increasing function. In combination with concentration inequalities for the risk, $\left\|\hat{s}_k^T - s\right\|^2$, this implies that the hold-out process, centered at the optimal model and appropriately rescaled, is approximately the sum of a convex function f_n and the zero-mean Gaussian process W_{g_n}. The same result applies to "incomplete" V-fold cross-validation (Definition 3), with g_n replaced by g_n/V. Both functions are deterministic: as a consequence, the distribution of the hold-out is approximately independent from the training data, and the folds of cross-validation are approximately independent. We establish upper and lower bounds on the functions f_n, g_n which guarantee that f_n, W_{g_n} do not vanish as $n \to +\infty$ and remain bounded /tight, at least when restricted to intervals of non-vanishing length containing the optimal parameter. At larger scales and outside these intervals, the limiting behaviour is trivial in the sense that the (centered and rescaled) hold-out concentrates around a deterministic function. Thus, our results characterize the critical scale at which randomness appears in the asymptotic.

This analysis was conducted under the assumption that there is a unique optimal model for each value of the sample size. Indeed, our main hypothesis is that the Fourier coefficients of the true density, s, are non-decreasing in absolute value, which is equivalent to the convexity of the squared bias, and implies in particular the approximate convexity of the risk with respect to the model dimension and hence the uniqueness of its minimizer (at least at the critical scale). While this non-standard hypothesis is quite strong, we believe that it need only hold approximately, and discuss several ways in which it could be relaxed.

The results of this article allow to make use of the abundant theory available on Brownian motion in order to study the model selection step.

2 L^2 density estimation

Let $s \in L^2([0;1])$ be a probability density function. Given a sample $X_1, \ldots, X_n$ drawn according to the density s, the L^2 density estimation problem consists in

constructing an estimator $\hat{s}_n$ that approaches s in terms of the L^2 norm.

Although it is not obvious at first glance (this is not true for the other L^p norms), this non-parametric density estimation problem can be reformulated as a risk minimization problem, with a contrast function: $\gamma(t,x) = \|t\|^2 - 2t(x)$, which yields the *risk* $\mathbb{E}[\gamma(t,X)] = \|t\|^2 - 2\int s(x)t(x)dx = \|t-s\|^2 - \|s\|^2$. It follows that s is indeed the minimizer of the risk corresponding to the γ contrast function, and furthermore the *excess risk* $\ell(s,t) := \mathbb{E}[\gamma(t,X)] - \mathbb{E}[\gamma(s,X)]$ coïncides with the L^2 norm:

$$\ell(s,t) = \|t-s\|^2 .$$

As a result, it is possible to construct an *empirical risk estimator*,

$$P_n\gamma(t) = \frac{1}{n}\sum_{i=1}^{n}\gamma(t,X_i) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} t(X_i),$$

which can in particular be used to perform cross-validation.

Here we will consider as a family of non-parametric estimators the *empirical orthogonal series* estimators [1, Section 3.1] on a trigonometric basis. To ease the presentation, we consider only cosine functions, which is equivalent to assuming that s is symmetrical with respect to $\frac{1}{2}$. This restriction is of no fundamental importance — it is reasonable to conjecture that the results remain valid with the complete trigonometric basis.

For every $j \in \mathbb{N}^*$, let $\psi_j : x \mapsto \sqrt{2}\cos(2\pi jx)$ and let $\psi_0 : x \mapsto 1$. The collection $(\psi_j)_{j\in\mathbb{N}}$ is an orthonormal basis of the subset of $L^2([0;1])$ of functions symmetrical with respect to $\frac{1}{2}$.

Let $D_n = (X_1, ..., X_n)$ be a sample. For any $n \in \mathbb{N}$ and any $T \subset \{1,\dots,n\}$, we will denote, for any real-valued measurable function t,

$$P_n^T(t) = \frac{1}{|T|}\sum_{i\in T} t(X_i).$$

Consider the estimators defined as follows.

Definition 1 *For all $k \in \mathbb{N}$ and all $T \subset \{1,\dots,n\}$,*

$$\hat{s}_k^T = \sum_{j=0}^{k} P_n^T(\psi_j)\psi_j,$$

where $\psi_0 = 1$ and for all $j \geq 1$, $\psi_j(x) = \sqrt{2}\cos(2\pi jx)$.

The estimators $\hat{s}_k^T$ are empirical risk minimizers on the *models*

$$E_k = \left\{ \sum_{j=0}^{k} v_j\psi_j : v \in \mathbb{R}^{k+1} \right\}.$$

The problem of parameter choice k is therefore a problem of *model selection* within the model collection $(E_k)_{k\geq 0}$. Here, the models are nested, meaning $E_k \subset E_{k'}$ for every $k \leq k'$.

3 Risk estimation for the hold-out

The larger k is, the better the approximation of s by the functions of E_k, but the more difficult it is to estimate the best approximation to s within E_k. The choice of k is therefore subject to a bias-variance trade-off which, if properly carried out, allows adaptation to the smoothness of s, simultaneously reaching the minimax risk on Lipschitz spaces of periodic functions [1].

3.1 Cross-validation

Since the risk, except for a constant, is expressed as the expectation of a contrast function

$$P\gamma(\hat{s}_k^T) := E_X\left[\gamma(\hat{s}_k^T, X)\right] = \left\|\hat{s}_k^T - s\right\|^2 - \|s\|^2,$$

it can be estimated by hold-out and cross-validation as in regression and classification.

This is the subject of the following definition.

Definition 2 *Let D_n be an i.i.d sample drawn from the distribution $s(x)dx$. Let $n_t \in \{1, \ldots, n-1\}$. Let $T \subset \{1..n\}$ be a subset with cardinality $|T| = n_t$. Then, for all $k \in \mathbb{N}$, we define the hold-out estimator of the risk of $\hat{s}_k$ with training sample indices T by*

$$HO_T(k) = \left\|\hat{s}_k^T\right\|^2 - 2P_n^{T^c}(\hat{s}_k^T).$$

$\mathrm{HO}_T(\cdot)$ is indeed an estimator since the norm $\|\cdot\|$ is computed with respect to a known dominating measure (in this case the Lebesgue measure) and so does not depend on the distribution of X. Moreover,

$$\mathrm{HO}_T(k) = \left\|\hat{s}_k^T - s\right\|^2 - \|s\|^2 - 2(P_n^{T^c} - P)(\hat{s}_k^T): \tag{2}$$

the hold-out risk estimator can be expressed as the sum of the risk and a centered empirical process. The risk which the hold-out naturally estimates is that of the estimator $\hat{s}_k^T$, which is trained on a sub-sample. If n_t is close enough to n, we expect its risk to be close to that of the same estimator trained on the full sample and this can indeed be proved in our setting (see Arlot and Lerasle [4, Lemma 14]).

The hold-out risk estimator depends on the choice of a subset T of $\{1, \ldots, n\}$, but its distribution depends only on the cardinality of that subset. The precise choice of a subset T of cardinality n_t will thus play no role in the sequel. We will therefore denote by T any subset of $\{1, \ldots, n\}$ of cardinality n_t.

Since the distribution of $\mathrm{HO}_T(\cdot)$ only depends on T through its cardinality n_t, it is possible to construct an estimator with smaller variance by averaging several $\mathrm{HO}_{T_i}(\cdot)$. This is the idea behind cross-validation. In the V-fold scheme presented below, the T_i are chosen such that the test sets T_i^c are disjoint.

Definition 3 *Let n_t, V be integers such that $\frac{V-1}{V}n \leq n_t \leq n-1$. Let $(I_i)_{1\leq i\leq V}$ be a collection of disjoint subsets of $\{1, \ldots, n\}$ of equal cardinality $|I_i| = n - n_t$, chosen independently from the data. For all $i \in \{1, \ldots, V\}$, let $T_i = \{1, \ldots, n\} \backslash I_i$. Let $\mathcal{T} = (T_1, \ldots, T_V)$. The "incomplete" V-fold CV risk estimator is*

$$\mathrm{CV}_{\mathcal{T}}(k) = \frac{1}{V}\sum_{i=1}^{V} HO_{T_i}(k).$$

Similarly to the hold-out, the distribution of $\mathrm{CV}_{\mathcal{T}}(\cdot)$ only depends on n_t, V and in the rest of this article, we will denote by $\mathcal{T}$ any collection $T_1, \ldots, T_V$ which satisfies the assumptions of Definition 3. Compared to standard $V-$fold, the cross-validation scheme defined above retains the constraint that the I_i be disjoint and of equal size, but decouples the size of the test sets $|I_i| = n - n_t$ from the number of splits V. In particular, the hold-out (Definition 2) is a special case of Definition 3 (for $V = 1$). As the collection $(I_i)_{1 \leq i \leq V}$ may be "completed" into a partition by adding sets I_j, we shall call $\mathrm{CV}_{\mathcal{T}}(k)$ "incomplete $V-$fold cross-validation".

To analyze the asymptotics of these risk estimators, fix a sequence of integers $(n_t(n))_{n \in \mathbb{N}}$ such that, for all $n \in \mathbb{N}$, $\frac{n}{2} \leq n_t(n) \leq n$, and define $n_v(n) = n - n_t(n)$. In the following, we shall denote $n_t = n_t(n)$ and $n_v = n_v(n)$ for a generic value of n. Whenever n, n_v, n_t appear in the same expression, it will be understood that $n_t = n_t(n)$ and $n_v = n_v(n) = n - n_t(n)$.

Since $\mathrm{CV}_{\mathcal{T}}(\cdot)$ can be expressed as an average of $\mathrm{HO}_{T_i}(\cdot)$, we first focus on analyzing the hold-out risk estimator $\mathrm{HO}_T(\cdot)$. Consequences for cross-validation will be derived in section 5.3 .

If a sequence k_n of model indices is not too badly chosen (such that $k_n \to +\infty$ and $\frac{k_n}{n} \to 0$), then the sequence of estimators $\hat{s}_{k_n}$ is consistent in L^2, which implies (under sufficient moment assumptions on s), that

$$\mathrm{HO}_T(k_n) \approx \left\| \hat{s}_{k_n}^T - s \right\|^2 - \|s\|^2 - 2(P_n^{T^c} - P)(s),$$

where the random error $(P_n^{T^c} - P)(s)$ is now independent of the model parameter k_n. While this error term is unavoidable when estimating the risk of a single estimator, when *comparing* several different models $k_{1,n}, \ldots, k_{j,n}$, the quantities that matter are the pairwise differences, $\mathrm{HO}_T(k_{i,n}) - \mathrm{HO}_T(k_{j,n})$, wherein the leading error term, $(P_n^{T^c} - P)(s)$, cancels out. To account for this effect, we consider a centered and scaled process of the following generic form:

$$\begin{aligned} \frac{1}{\mathfrak{e}}\left(\mathrm{HO}_T(k_n + \alpha\Delta) - \mathrm{HO}_T(k_n)\right) = \frac{1}{\mathfrak{e}}\left(\left\|\hat{s}_{k_n+\alpha\Delta}^T - s\right\|^2 - \left\|\hat{s}_{k_n}^T - s\right\|^2\right) \\ - \frac{2}{\mathfrak{e}}(P_n^{T^c} - P)(\hat{s}_{k_n+\alpha\Delta}^T - \hat{s}_{k_n}^T), \end{aligned} \tag{3}$$

where $k_n, \Delta = \Delta(n)$ and $\mathfrak{e} = \mathfrak{e}(n)$ are sequences depending on $n_t(n)$ and the density s.

3.2 Centering sequence

While any centering sequence $(k_n)_{n \in \mathbb{N}}$ may a priori be chosen, they are not equally interesting. Sufficiently good model selection estimators are expected to concentrate, in a certain sense, around the optimal model $m_*(n)$ corresponding to a given sample size (assuming it is well defined). In the case of cross-validation, the relevant sample size is rather that of the training set, since cross-validation estimates the risk of estimators trained on a sub-sample. Concentration results of this kind were proved for cross-validation by van der Vaart et al. [23] and Navarro and Saumard [19] in regression. These works establish in various frameworks that the risk of the cross-validation based estimator approaches the optimal risk to within an arbitrarily small constant factor, as $n \to +\infty$. In contrast, it is known that, relative to the oracle computed on the

full sample, cross-validation can be sub-optimal by a constant factor [2] unless the fraction of data in the training set, $\frac{n_t}{n}$, approaches 1. Though in that case, concentration around $m_*(n)$ can be established (as was done in least-squares density estimation by Arlot and Lerasle [4]), it may well be that concentration around $m_*(n_t)$ remains asymptotically tighter than around $m_*(n)$. The question is in any case left open, as non-asymptotic oracle inequalities provide only an upper bound on the concentration of cross-validation, with no indication whether it is optimal beyond the first order. By characterizing the scale of the random fluctuations of the CV criterion in the vicinity of the optimal parameter, a local asymptotic analysis of CV around the optimal model could provide precise estimates of the deviation between the model selected by CV and the optimal one. For these reasons, we will seek to center the hold-out risk estimator close to the model $m_*(n_t)$.

Thus far, we have talked of the optimal model parameter as if it were a well-defined quantity. Under a few conditions, it is possible to guarantee its uniqueness and to give moreover a simple, deterministic approximant to it.

For any $j \in \mathbb{N}$, let $\theta_j = \langle s, \psi_j \rangle$ denote the Fourier coefficients of s on the cosine basis. The L^2 risk can be approximated as follows with large probability (a precise statement can be found in claim 3):

$$\left\| \hat{s}_k^T - s \right\|^2 \approx \frac{k}{n_t} + \sum_{j=k+1}^{+\infty} \theta_j^2.$$

If the squared Fourier coefficients θ_j^2 form a non-increasing sequence, then the approximating function $k \mapsto \frac{k}{n_t} + \sum_{j=k+1}^{+\infty} \theta_j^2$ is convex and, in particular, has a unique minimizer $k_*(n_t)$. As a further consequence, the level sets of "near-optimal" values of k form a nested collection of intervals. These properties greatly simplify the analysis of the hold-out procedure by avoiding situations where the hold-out "jumps" between two widely separated regions. For this reason, in the remainder of the article, we shall always assume that the sequence θ_j^2 is non-increasing. We will discuss later how this assumption might be relaxed.

Assuming now that θ_j^2 is non-decreasing, it is possible to give simple approximate formulas for the argmin and minimum of the true risk, $\left\| \hat{s}_k^T - s \right\|^2$ and its expectation.

Definition 4 *For all $n \in \mathbb{N}$, let*

$$k_*(n) = \max \left\{ k \in \mathbb{N} : \theta_k^2 \geq \frac{1}{n} \right\}$$

$$\textit{and} \qquad \mathrm{or}(n) = \inf_{k \in \mathbb{N}} \left\{ \sum_{j=k+1}^{+\infty} \theta_j^2 + \frac{k}{n} \right\}.$$

Equivalently,

$$k_*(n) = \max \operatorname*{argmin}_{k \in \mathbb{N}} \left\{ \sum_{j=k+1}^{+\infty} \theta_j^2 + \frac{k}{n} \right\}$$

$$\textit{and} \qquad \mathrm{or}(n) = \sum_{j=k_*(n)+1}^{+\infty} \theta_j^2 + \frac{k_*(n)}{n}.$$

$k_*(n_t)$ and $\mathrm{or}(n_t)$ are thus, approximately, the minimizer and the minimum in k of the L^2 risk of the estimators $\hat{s}_k^T$, which explains the name $\mathrm{or}(n_t)$ (oracle). Thus, we will analyze the hold-out centered at the sequence $k_n = k_*(n_t(n))$ (cf equation (3)).

3.3 Scaling

It remains now to choose the scaling sequences $\Delta(n), \mathfrak{e}(n)$. Since $\mathfrak{e}(n)$ will be chosen so as to obtain a bounded, non-vanishing process, its value is essentially determined by that of $\Delta(n)$. Let us now discuss how this sequence is defined. It has already been remarked that, provided that the sequence $n_t(n)$ is properly chosen, the hold-out risk estimator, $\mathrm{HO}_T(k)$, concentrates around the risk of $\hat{s}_k^T$, which itself concentrates around its expectation. Thus, if the scale Δ is too large, the centered and rescaled hold-out process (equation (3)) will converge to a deterministic function. The scale chosen is then too large to distinguish the hold-out estimator from the quantity that it estimates. As the scale decreases, the effect of the random fluctuations is magnified and at a certain critical scale, randomness should appear in the asymptotic. Our goal is now to characterize this critical scale, as well as the corresponding limit process.

The appropriate choice of $\Delta, \mathfrak{e}$ is given in the following definition.

Definition 5 *For all $n \in \mathbb{N}$, let*

$$\begin{aligned}
\Delta_d(s, n_t, n) &= \max\left\{ l \in \mathbb{N} : \theta^2_{k_*(n_t)+l} \geq \left[1 - \sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{l}}\right]\frac{1}{n_t} \right\} \\
\Delta_g(s, n_t, n) &= \min\left\{ l \in \{0, \dots, k_*(n_t)\} : \theta^2_{k_*(n_t)-l} \geq \left[1 + \sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{l}}\right]\frac{1}{n_t} \right\} \\
\Delta(s, n_t, n) &= \max\left(\Delta_d(s, n_t, n), \Delta_g(s, n_t, n)\right) \\
\mathcal{E}(s, n_t, n) &= \frac{\Delta(s, n_t, n)}{n_t}. \\
\mathfrak{e}(s, n_t, n) &= \sqrt{\frac{\mathcal{E}(s, n_t, n)}{n - n_t}}.
\end{aligned}$$

Definition 5 also introduces the quantity $\mathcal{E}(s, n_t, n)$. This quantity appears often in the proofs, so it is helpful to have notation for it; it also has an interpretation as the order of magnitude of the fluctuations in the variance term,

$$\mathbb{E}\left[\left\|\hat{s}_k^T - \mathbb{E}[\hat{s}_k^T]\right\|^2\right] - \mathbb{E}\left[\left\|\hat{s}_{k_*}^T - \mathbb{E}[\hat{s}_{k_*}^T]\right\|^2\right],$$

and the bias term,

$$\left\|\mathbb{E}[\hat{s}_k^T] - s\right\|^2 - \left\|\mathbb{E}[\hat{s}_{k_*}^T] - s\right\|^2 = -\mathrm{sign}(k - k_*) \sum_{j=k_* \wedge k+1}^{k \vee k_*} \theta_j^2,$$

of the estimators $\hat{s}_k^T$, for $k - k_*$ "of order" Δ (in a sense to be made precise later).

As the sequence $n_t(n)$ and the density s are considered to be fixed once and for all, the notation $\Delta(s, n_t, n), \mathcal{E}(s, n_t, n), \mathfrak{e}(s, n_t, n)$ will frequently be replaced by the abbreviations $\Delta, \mathcal{E}, \mathfrak{e}$.

Definition 5 does not make clear how large $\Delta, \mathcal{E}$ and $\mathfrak{e}$ are. Their order of magnitude may depend on the sequence $(\theta_j)_{j\in\mathbb{N}}$ of Fourier coefficients of s as well as on $n_t(n)$. However, the following inequalities always hold.

Lemme 3.1 *For any density s such that the sequence $\theta_j^2 = \langle s, \psi_j\rangle^2$ is non-increasing,*

$$\Delta \geq \frac{n_t}{n-n_t} \tag{4}$$

$$\mathcal{E} \geq \frac{1}{n-n_t} \tag{5}$$

$$\mathfrak{e} \geq \frac{1}{n-n_t} \tag{6}$$

$$\mathfrak{e} \leq \mathcal{E} \tag{7}$$

$$\mathcal{E} \leq 2\mathrm{or}(n_t) + \frac{1}{n-n_t}. \tag{8}$$

This lemma is proved in section 7.1.1. The following two examples show that in extreme cases, lemma 3.1 may be optimal, at least up to constants.

Two examples

- Let $n_t(n)$ and u_n be two integer sequences, such that $\frac{n_t(n)}{n} \to 1$, $u_n \to +\infty$ and $u_n \leq \frac{\sqrt{n}}{2}$ for all n. Assume also that $\frac{n-n_t}{n_t} = o(\frac{u_{n_t}}{n_t})$. Let for all $j \in \mathbb{N}$

$$\theta_{j,n}^2 = \begin{cases} 1 \text{ if } j = 0 \\ \frac{1}{n_t} \text{ if } 1 \leq j \leq u_{n_t} \\ 0 \text{ if } j \geq u_{n_t}+1, \end{cases} \tag{9}$$

  corresponding for example to the pdf $s_n = 1 + \sum_{j=1}^{u_{n_t}} \sqrt{\frac{1}{n_t}}\psi_j$. Remark that equation (9) implies that $k_*(n_t) = u_{n_t}$. Then as $n \to +\infty$, $\mathcal{E}(s_n, n_t, n) \sim \frac{u_{n_t}}{n_t} \sim \mathrm{or}(n_t)$ and $\frac{n-n_t}{n_t} = o(\mathrm{or}(n_t))$, so $\mathfrak{e}(s_n, n_t, n) = o(\mathcal{E}(s_n, n_t, n))$.

- Let s be the pdf associated with the Fourier coefficients

$$\forall j \in \mathbb{N}, \langle s, \psi_j\rangle = \theta_j = \frac{1}{3^j}. \tag{10}$$

  Let $n_t(n)$ be a sequence of integers such that $\frac{n_t(n)}{n} \to 1$. Then by Lemma 3.1, $\Delta \geq \frac{n_t}{n-n_t}$, but as

$$9^{-\frac{n_t}{n-n_t}} = o\left(1 - \sqrt{\frac{1}{1+\frac{n-n_t}{n_t}}}\right),$$

  it follows that $\Delta(s, n_t, n) \sim \frac{n_t}{n-n_t}$, hence

$$\mathcal{E}(s, n_t, n) \sim \frac{n_t}{(n-n_t)n_t} \sim \frac{1}{n-n_t}.$$

  As a result, $\mathcal{E}(s, n_t, n) \sim \mathfrak{e}(s, n_t, n) \sim \frac{1}{n-n_t}$, and this for *any* sequence $n_t(n)$ such that $n_t(n) \sim n$.

Now that $\Delta, \mathfrak{e}$ are defined, the hold-out process can be rescaled as in equation (6). More precisely, the rescaled hold-out process is given by Definition 6 below.

Definition 6 *For all $j \in [-k_*; +\infty[\cap\mathbb{Z}$, let*

$$\hat{R}_T^{ho}\left(\frac{j}{\Delta}\right) = \frac{1}{\mathfrak{e}}\left(HO_T\left(k_* + j\right) - HO_T\left(k_*\right)\right),$$

in other words (by definition 2)

$$\hat{R}_T^{ho}\left(\frac{j}{\Delta}\right) = \frac{1}{\mathfrak{e}}\left(\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2\right) - \frac{2}{\mathfrak{e}}\left(P_n^{T^c} - P\right)\left(\hat{s}_{k_*+j}^T - \hat{s}_{k_*}^T\right).$$

The $\hat{R}_T^{ho}$ function is extended by linear interpolation to all $\alpha \in \left[\frac{-k_*(n_t)}{\Delta}; +\infty\right[$. Let $\hat{R}_{\mathcal{T}}^{cv}$ be defined in a similar manner, i.e

$$\hat{R}_{\mathcal{T}}^{cv}(\tfrac{j}{\Delta}) = \frac{1}{\mathfrak{e}}\left(\mathrm{CV}_{\mathcal{T}}(k_* + j) - \mathrm{CV}_{\mathcal{T}}(k_*)\right)$$

for all $j \in [-k_*; +\infty[\cap\mathbb{Z}$, extended by linear interpolation to $\left[\frac{-k_*(n_t)}{\Delta}; +\infty\right[$. Note that by linearity of the interpolation operation, $\hat{R}_{\mathcal{T}}^{cv} = \frac{1}{V}\sum_{i=1}^{V}\hat{R}_{T_i}^{ho}$.

The extension of $\hat{R}_T^{ho}, \mathrm{CV}_{\mathcal{T}}()$ by linear interpolation simplifies their approximation by a continuous process. Notice that any minimizer of $\hat{R}_T^{ho}$ (resp. $\mathrm{CV}_{\mathcal{T}}()$) on the grid $\frac{1}{\Delta}\left([-k_*(n_t); +\infty[\cap\mathbb{Z}\right)$ remains a minimizer of $\hat{R}_T^{ho}$ (resp. $\mathrm{CV}_{\mathcal{T}}()$) on the interval $\left[\frac{-k_*(n_t)}{\Delta}; +\infty\right[$. In particular, this applies to the hold-out parameter obtained by minimisation of the hold-out risk estimator.

The quantity $\hat{R}_T^{ho}\left(\frac{j}{\Delta}\right)$ can be expressed as the sum of the (rescaled) excess risk,

$$\frac{1}{\mathfrak{e}}\left(\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2\right)$$

and a centered empirical process: $\frac{2}{\mathfrak{e}}\left(P_n^{T^c} - P\right)\left(\hat{s}_{k_*+j}^T - \hat{s}_{k_*}^T\right)$. Though the excess risk is a priori random (it depends on D_n^T), the proof will show that it concentrates around a deterministic function f_n, depending on n, which is given by definition 7 below.

Definition 7 *For all $k \in \mathbb{N}$, let $R(k) = \sum_{j=k+1}^{+\infty}\theta_j^2$. Extend R to $\mathbb{R}_+$ by linear interpolation:*

$$\forall x \in \mathbb{R}_+, R(x) = (1 + \lfloor x\rfloor - x)R(\lfloor x\rfloor) + (x - \lfloor x\rfloor)R(\lfloor x\rfloor + 1).$$

$f_n :]-\frac{k_*(n_t)}{\Delta}; +\infty[\to \mathbb{R}_+$ is now defined by:

$$f_n(\alpha) = \frac{1}{\mathfrak{e}}\left(R(k_*(n_t) + \alpha\Delta) - R(k_*(n_t)) + \frac{\alpha\Delta}{n_t}\right). \tag{11}$$

Thus, for all $k \in \mathbb{N}$, $k \neq k_*(n_t)$,

$$\mathfrak{e} f_n\left(\frac{k - k_*(n_t)}{\Delta}\right) = \sum_{j=k\wedge k_*(n_t)+1}^{k\vee k_*(n_t)}\left|\theta_j^2 - \frac{1}{n_t}\right|. \tag{12}$$

It is clear by equation (12) that f_n reaches its minimum at 0, moreover the assumption that the sequence $(\theta_j^2)_{j\in\mathbb{N}}$ is non-increasing implies that f_n is convex. In particular, f_n is non-increasing on $]-\frac{k_*(n_t)}{\Delta};0]$ and non-decreasing on $[0;+\infty[$. Moreover, the definition of Δ and $\mathfrak{e}$ (Definition 5) implies the following bounds on the increments of f_n:

Lemme 3.2 *For any $\alpha_1,\alpha_2\in\mathbb{R}$ such that $\alpha_1\alpha_2\geq 0$ and $|\alpha_2|\geq|\alpha_1|\geq 1$,*

$$f_n(\alpha_2)-f_n(\alpha_1)\geq|\alpha_2|-|\alpha_1|.$$

In particular, since $f_n(0)=0$, for all $\alpha\in\mathbb{R}$,

$$f_n(\alpha)\geq(|\alpha|-1)_+.$$

Moreover, using the notation from Definition 5,

- *If $\Delta=\Delta_d$, then for any $\alpha_1,\alpha_2\in[0;1]$ such that $\alpha_1\leq\alpha_2$, $f_n(\alpha_2)-f_n(\alpha_1)\leq\alpha_2-\alpha_1$.*
- *If $\Delta=\Delta_g$, then for any $\alpha_1,\alpha_2\in[-1;0]$ such that $\alpha_1\leq\alpha_2$, $f_n(\alpha_1)-f_n(\alpha_2)\leq\alpha_2-\alpha_1$.*

This lemma is proved in section 7.1.2. It guarantees that f_n remains in a sense of "finite order" and "non zero" as $n\to+\infty$, which means that f_n remains uniformly bounded on $[-1;0]$ or on $[0;1]$, and is lower-bounded on $\mathbb{R}$ by the non-zero function $(|x|-1)_+$.

4 Hypotheses

The main hypothesis of this article is that the squared Fourier coefficients θ_j^2 are non-increasing, which is admittedly strong, though a natural assumption in the context of this article, as discussed above. It seems likely that the desirable effects of this hypothesis can be retained under weaker conditions, as we will discuss later (section 6.2).

In addition to this "shape constraint" on the θ_j^2 , the main theorem of this article requires a number of more technical assumptions. First, approximating the process $HO_T(k)$, which is a sum over Fourier coefficients, inevitably involves bounding "tail sums" of the Fourier series, such as $\sum_{j=k+1}\theta_j^2$. To control these, we assume that the Fourier coefficients decay fast enough.

Hypothesis 1 *There exists constants $c_1\geq 0$ and $\delta_1\geq 0$ such that for all $k\in\mathbb{N}$, $\sum_{j=k+1}^{+\infty}\theta_j^2\leq\frac{c_1}{k^{2+\delta_1}}$.*

This upper bound is satisfied for some c_1,δ_1 if and only if the smoothness assumption $s\in H^\beta$ holds for some $\beta>1$, where H^β denotes the Sobolev Hilbert space. This implies in particular that $(k_*(n_t)\leq n_t)$ is satisfied for all sufficiently large n_t (thus also for all large enough n).

Secondly, since we seek to approximate the discrete process $HO_T(k)$ by a continuous one, it is necessary to make sure that the number of "points" $k\in[k_*-\alpha\Delta,k_*+\alpha\Delta]$ tends to infinity as $n\to+\infty$, for all α. Similarly, one must assume that $f_n(\frac{j}{\Delta})\sim f_n(\frac{j-1}{\Delta})$, otherwise the continuous function f_n is not sufficiently close to its discrete version $j\mapsto f_n(\frac{j}{\Delta})$. For technical reasons, we will assume a polynomial growth rate, using the following three hypotheses.

Hypothesis 2 *There exists constants $c_2 \geq 0$, $\delta_2 \geq 0$ such that for all $k \in \mathbb{N}$, $\sum_{j=k+1}^{+\infty} \theta_j^2 \geq \frac{c_2}{k^{\delta_2}}$*

This hypothesis states that the Fourier coefficients θ_j^2 cannot decay faster than polynomially, and excludes in particular analytic functions. This guarantees that $k_*(n_t) \to +\infty$ at a polynomial rate. Hypothesis 1 holds for example if s or one of its derivatives has a point of discontinuity.

Hypothesis 3 *There exists constants $c_3 > 0$, $\delta_3 > 0$ such that for all $k \geq 1$,*

$$\theta_{k+k^{\delta_3}}^2 \geq c_3 \theta_{k-k^{\delta_3}}^2 .$$

Hypothesis 3 means that the sequence θ_j^2 cannot decrease too abruptly, excluding in particular a locally exponential decrease such as $\theta_{k_n+j}^2 = 2^{-j} w_n$, for $j \in \{1, \ldots, \varepsilon \log k_n\}$ and $k_n \to +\infty$. Without this hypothesis, f_n may have "asymptotically sharp" discontinuities at ± 1, violating the condition $f_n(\frac{j}{\Delta}) \sim f_n(\frac{j-1}{\Delta})$, see claim 3.2 and example (10) . Hypothesis 3 is satisfied by polynomially decreasing sequences, $\theta_j^2 = \kappa j^{-\beta}$, with $\delta_3 = 1$, but also by sequences $\theta_j^2 = \kappa \exp(-j^\alpha)$, as long as $\alpha < 1$. Locally, θ_j^2 can thus decrease much faster than the polynomial lower bound given by hypothesis 2.

Together, hypotheses 1, 2, 3 basically mean that the Fourier coefficients of s on the cosine basis decrease polynomially. For example, they are satisfied if there are two strictly positive constants μ, L and a constant $\beta > 1$, such that

$$\forall k \in \mathbb{N}, \mu k^{-2\beta} \leq \sum_{j=k+1}^{+\infty} \theta_j^2 \leq L k^{-2\beta}.$$

The two remaining hypotheses 4 and 5 do not bear on s, but on the parameter n_t which is chosen by the statistician. They serve to establish upper and lower bounds on Δ using lemma 3.1.

Hypothesis 4 *There exists a constant $\delta_4 > 0$ such that $n - n_t \leq n^{1-\delta_4}$.*

By lemma 3.1, hypothesis 4 guarantees that Δ grows at least at a polynomial rate n^{δ_4}, as announced above. For technical reasons, it is also necessary to upper bound Δ, which is accomplished using hypothesis 5 below.

Hypothesis 5 *There exists a constant $\delta_5 > 0$ such that $n_v = n - n_t \geq n^{\frac{2}{3}+\delta_5}$.*

The statistician can always choose n_t such that hypotheses 4 and 5 hold. One should however check that this is compatible with good performance of the hold-out. The oracle inequalities of Arlot and Lerasle [4] show that the risk of the hold-out in model selection for L^2 density estimation is (at most) of order $\frac{n}{n_t}\mathrm{or}(n) + \frac{\log(n-n_t)}{n-n_t}$. If $\mathrm{or}(n)$ decreases in n with rate $\frac{1}{n^\alpha}$ ($\alpha \in (0,1)$), which is the case under assumptions 1 and 2, $n - n_t$ can be chosen within the interval $[\frac{1}{2} n^{\frac{2+\alpha}{3}}; n^{\frac{4+\alpha}{5}}]$ —so that assumptions 4 and 5 are satisfied— without changing the order of magnitude of the risk.

5 Main theorems

The purpose of this article is to find simple approximants $Y_{n,V}(u)$ to the rescaled CV estimators $\hat{R}_{\mathcal{T}}^{cv}(u)$ (including the hold-out when the number of splits $V = 1$).

5.1 Domain of approximation

Given that the process $\hat{R}^{cv}_{\mathcal{T}}$ is unbounded, it is unrealistic to expect a uniform approximation to hold on the whole real line. Rather, we consider uniform approximations on intervals containing the optimum $\alpha = 0$ defined as sub-level sets of f_n. More precisely, these intervals are defined as follows.

Definition 8 *Let $T \subset \{1 \dots n\}$ be a subset of cardinality n_t and let $k_* = k_*(n_t)$. For any $x > 0$, let*

$$a_x = \min\{\tfrac{j}{\Delta} : j \in \{-k_*(n_t), \dots, 0\}, f_n(\tfrac{j}{\Delta}) \leq x\}$$
$$b_x = \max\{\tfrac{j}{\Delta} : j \in \mathbb{N}, f_n(\tfrac{j}{\Delta}) \leq x\}.$$

Up to the constraint that $a_x, b_x \in \frac{1}{\Delta}\mathbb{Z}$, the sets $[a_x, b_x]$ are just the sublevel sets of f_n. Moreover, claim 7 implies that

$$x - o(1) \leq \min(f_n(a_x), f_n(b_x)) \leq \max(f_n(a_x), f_n(b_x)) \leq x.$$

By lemma 3.2, for $x \geq 1$, $[a_x, b_x]$ either contains $[0, 1]$ (if $\Delta = \Delta_d$) or $[-1, 0]$ (if $\Delta = \Delta_g$). On the other hand, by lemma 7.1, $b_x - a_x \leq 2(1 + x)$. This shows that the intervals $[a_x, b_x]$ are bounded and of non-vanishing size as $n \to +\infty$.

Not much is lost by restricting to such intervals, since we will show later that the process $\hat{R}^{cv}_{\mathcal{T}}(u)$ concentrates around its expectation for $u \notin [a_{x_n}, b_{x_n}]$ and $x_n \to +\infty$.

5.2 Simple validation

The following theorem shows that the process $\hat{R}^{ho}_T(\cdot)$ can be approximated on $[a_x, b_x]$ by the sum of f_n and a time-changed Brownian motion.

Theorem 1 *Assume that the hypotheses of section 4 hold. There exists an increasing continuous function $g_n : [-\frac{k_*(n_t)}{\Delta}; +\infty[\to \mathbb{R}$ and for any $x > 0$, there exists a two-sided Brownian motion $(W_t)_{t \in [a_x; b_x]}$ independent from D^T_n such that, for any $y > 0$, with probability greater than $1 - e^{-y}$,*

$$\mathbb{E}\left[\sup_{u \in [a_x; b_x]} \left|\hat{R}^{ho}_T(u) - (f_n(u) - W_{g_n(u)})\right| \,\middle|\, D^T_n\right] \leq \kappa_0 (1+y)^2 (1+x)^{\frac{3}{2}} n^{-u_1}, \quad (13)$$

where $u_1 > 0$ and $\kappa_0 \geq 0$ are two constants which depend only on $\delta_1, \delta_3, \delta_5, \delta_2$ and c_1, c_2, δ_1, c_3, respectively. Moreover, g_n and W can be chosen so as to satisfy the following conditions.

1. *$g_n(0) = 0$, $W_0 = 0$,*
2. *$\forall (\alpha_1, \alpha_2) \in \left[\frac{-k_*}{\Delta}; +\infty\right[^2$, $\alpha_2 < \alpha_1 \implies g_n(\alpha_1) - g_n(\alpha_2) \geq 4 \|s\|^2 [\alpha_1 - \alpha_2]$.*
3. *For all $(\alpha_1, \alpha_2) \in \left[\frac{-k_*}{\Delta}; +\infty\right[^2$ such that $\alpha_1 < \alpha_2 < 0$ or $0 < \alpha_1 < \alpha_2$,*

$$g_n(\alpha_2) - g_n(\alpha_1) \leq -\frac{8 \|s\|_\infty}{(n - n_t)\mathfrak{e}} [f_n(\alpha_2) - f_n(\alpha_1)] + \left(8 \|s\|_\infty + 4 \|s\|^2\right) [\alpha_2 - \alpha_1]. \quad (14)$$

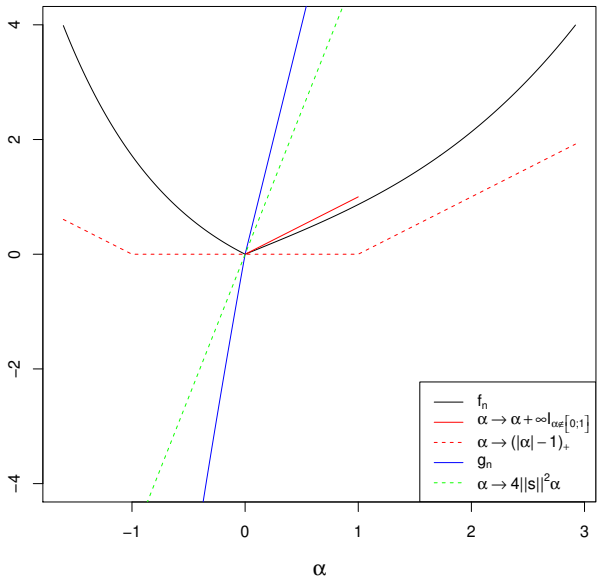

Figure 1: A plot of f_n, g_n on $[a_6; b_6]$ with upper and lower bounds, for $\|s\|^2 = 1.2$.

This theorem is proved in section 7. It states that the rescaled hold-out process, $\hat{R}_T^{ho}$, can be approximated uniformly in expectation on $[a_x, b_x]$ by a continuous process $Y_{n,1}$,which is the sum of a convex non-negative function f_n and a time-changed Brownian motion W_{g_n}. f_n and g_n depend on n_t and n, but not on the data (they are deterministic functions), while W depends on the data only through the test sample $D_n^{T^c}$. In particular, in this asymptotic setting, $\hat{R}_T^{ho}$ doesn't depend on D_n^T, the training data.

The function g_n increases on its domain and has a Lipschitz-continuous inverse. By lemma 3.2 and equation (3), f_n, g_n are Lipschitz continuous either on $[-1, 0]$ or on $[0, 1]$ (depending on whether $\Delta = \Delta_d$ or $\Delta_g = \Delta$), with Lipschitz constants that depends only on $\|s\|^2, \|s\|_\infty$. In particular, f_n, g_n are both of constant order on $[a_x, b_x]$ for $x > 1$. Figure 1 illustrates the bounds that hold on f_n, g_n in a situation where $\Delta = \Delta_d$.

On the one hand, by definition of a_x, b_x, $0 \leq f_n \leq x$ on $[a_x, b_x]$ and by equation (3),

$$\sup_{\alpha \in [a_x, b_x]} |g_n(\alpha)| \leq 8 \|s\|_\infty x + \left(8 \|s\|_\infty + 4 \|s\|^2\right) \max(|a_x|, |b_x|) \leq 20 \|s\|_\infty (1+x) \tag{15}$$

On the other hand, $|g_n(\alpha)| \geq 4 \|s\|^2 |\alpha|$ by equation (2) and $f_n(b_x) \geq x - o(1)$ by claim 7 . Thus, the mean and standard deviation of $Y_{n,1}$ are both of constant order on $[a_x, b_x]$, as $n \to +\infty$.

This leaves the question of what happens outside the window $[a_x, b_x]$ and/or at larger scales. The following proposition answers this question: as $x \to +\infty$,

the hold-out process concentrates around f_n.

Proposition 5.1 *Let $(M_n)_{n\in\mathbb{N}}$ be an integer sequence and*

$$I_n = \left[-\frac{k_*(n_t)}{\Delta}, \frac{M_n - k_*(n_t)}{\Delta}\right].$$

There exists constants κ and $\delta > 0$ (depending on $(c_i, \delta_i)_{i\in\{1,\dots,5\}}$) such that, for all $x \geq e^2$,

$$\mathbb{E}\left[\sup_{u\in I_n\setminus[a_x,b_x]}\left|\frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1\right|\right] \leq \sqrt{\frac{2\kappa\log\log(\kappa x)}{x}} + \frac{\kappa\log^2(nM_n)}{n^\delta}.$$

In particular, for any sequence $\lambda_n \to +\infty$ and any $\delta > 0$,

$$\sup_{u\in I_n:|u|>\delta}\left|\frac{\hat{R}_T^{ho}(\lambda_n u)}{f_n(\lambda_n u)} - 1\right| \to 0$$

in probability, as long as M_n isn't too large (for example, $M_n = o(n^r)$ for some $r > 0$).

Because a union bound is used, proposition 5.1 applies only on a bounded interval I_n corresponding to values of the dimension parameter $k \leq M_n$. M_n can grow polynomially or even super-polynomially without invalidating the result. In practice, this assumption is mild since by definition 4, $k_*(n_t) \leq \|s\|^2 n_t \leq n^2$ for any $n \geq \|s\|^2$.

Proof Since $\hat{R}_T^{ho}$ and f_n are piecewise linear, the supremum is reached on $\frac{1}{\Delta}\mathbb{Z}$. Let $x_n = \frac{n^{\frac{u_1}{2}\wedge u_2}}{4\kappa_3}$. Consider first $x \in [e^2, x_n]$ and $u \in [a_{x_n}, b_{x_n}]\setminus[a_x, b_x]$. By the above Theorem, the sequence of random variables

$$V_n^1 = \sup_{u\in[a_{x_n},b_{x_n}]} n^{u_1/4}\left|\hat{R}_T^{ho}(u) - f_n(u) + W_{g_n(u)}\right|$$

satisfies $\mathbb{E}[V_n^1] \leq \kappa$ for some constant κ. By definition,

$$\left|\frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1\right| \leq \frac{|W_{g_n(u)}|}{f_n(u)} + \frac{1}{e^2}V_n^1 n^{-u_1/4}.$$

Define the function

$$w : t \mapsto \sqrt{2(t\vee e^2)\log\log(t\vee e^2)}.$$

w is clearly non-decreasing, and one can easily show that $t \mapsto \frac{w(t)}{t}$ is non-increasing.

By the law of the iterated logarithm, the random variable

$$V_n^2 = \sup_{t\in\mathbb{R}}\frac{|W_t^n|}{w(t)}$$

is almost surely bounded. Since the Gaussian process $\frac{W_t^n}{w(t)}$ is almost surely continuous and bounded, with variance bounded by $\frac{1}{2\log 2}$, the Gaussian isoperimetric theorem can be used together with the monotone convergence theorem to establish that

$$\mathbb{P}\left(V_n^2 \geq M(V_n^2) + 2\sqrt{y\log 2}\right) \leq e^{-y}$$

where $M(V_n^2)$ is the median of V_n^2. In particular, $\mathbb{E}[V_n^2] = \kappa$ is a finite constant. Moreover, by definition,

$$\left|\frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1\right| \leq V_n^2 \frac{w(g_n(u))}{f_n(u)} + \frac{1}{e^2} V_n^1 n^{-u_1/4}.$$

From equation (3) and lemma 3.2, it follows that $g_n(u) \leq \kappa f_n(u)$, for some constant κ depending only on $\|s\|_\infty, \|s\|$. Moreover, since $u \notin [a_x, b_x]$, by claim 7,

$$f_n(u) \geq \min(f_n(a_x), f_n(b_x)) \geq x - \kappa_3(1+x)^2 n^{-u_2} \geq x - 2\kappa_3 x^2 n^{-u_2} = \left[1 - 2\kappa_3 x n^{-u_2}\right] x \geq \frac{x}{2}$$

Hence,

$$\begin{aligned}\left|\frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1\right| &\leq V_n^2 \frac{w(\kappa f_n(u))}{f_n(u)} + \frac{1}{e^2} V_n^1 n^{-u_1/4} \\ &\leq V_n^2 \frac{2w(\kappa x/2)}{x} + \frac{1}{e^2} V_n^1 n^{-u_1/4} \\ &= V_n^2 \sqrt{\frac{4\kappa \log\log(\kappa x)}{x}} + \frac{1}{e^2} V_n^1 n^{-u_1/4},\end{aligned}$$

which yields

$$\mathbb{E}\left[\sup_{u \in I_n : x < f_n(u) \leq x_n} \left|\frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1\right|\right] \leq 2\sqrt{\frac{\kappa \log\log(\kappa x)}{x}} + \kappa n^{-\frac{u_1}{4}}. \tag{16}$$

Assume now that $u \in I_n \backslash [a_{x_n}, b_{x_n}]$. By Claim 7,

$$f_n(u) \geq \min\left(f_n(a_{x_n}), f_n(b_{x_n})\right) \geq x_n - \kappa_3(1+x_n)^2 n^{-u_2} \geq 4\kappa_3 n^{\frac{u_1}{2} \wedge u_2}.$$

Let $j = u\Delta$. Then

$$\begin{aligned}\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2 &= \operatorname{sign}(j) \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left[\left(\hat{\theta}_i^T - \theta_i\right)^2 - \theta_i^2\right] \\ &= \operatorname{sign}(j) \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left[\frac{1}{n_t} - \theta_i^2\right] + \operatorname{sign}(j) \left(\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 - \frac{|j|}{n_t}\right).\end{aligned}$$

By definition of k_* and of f_n,

$$\operatorname{sign}(j) \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left[\frac{1}{n_t} - \theta_i^2\right] = \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left|\frac{1}{n_t} - \theta_i^2\right| = \mathfrak{e} f_n(u).$$

Let $w_n(y) = \kappa_1[\log(nM_n) + y]^2 n^{-\frac{1}{12} \wedge \frac{\delta_1}{4}}$. By proposition 8.3, on an event $\Omega_{n,y}$ with probability greater than $1 - e^{-y}$,

$$\left|\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2 - \mathfrak{e} f_n(u)\right| \leq u w_n(y) \mathfrak{e} \tag{17}$$

for all $u \in \left[\frac{-k_*}{\Delta}, \frac{M_n - k_*}{\Delta}\right] \cap \frac{1}{\Delta}\mathbb{Z}$. Consider now the term

$$\left(P_n^{T^c} - P\right)\left(\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right).$$

On $\Omega_{n,y}$,

$$\begin{aligned}
\mathrm{Var}\left(\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right) &\leq \|s\|_\infty \left\|\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right\|^2 \\
&= \|s\|_\infty \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T\right)^2 \\
&\leq 2\|s\|_\infty \left(\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \theta_i^2 + \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2\right) \\
&\leq 2\|s\|_\infty \left(\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left|\theta_i^2 - \frac{1}{n_t}\right| + 2\frac{|j|}{n_t} + \left|\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 - \frac{|j|}{n_t}\right|\right) \\
&\leq 2\|s\|_\infty \left(\mathfrak{e} f_n(u) + 2|u|\mathcal{E} + |u| w_n(y)\mathfrak{e}\right).
\end{aligned}$$

Since $|u| \leq 2f_n(u)$ by lemma 3.2 and $\mathfrak{e} \leq \mathcal{E}$ by lemma 3.1,

$$\mathrm{Var}\left(\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right) \leq \|s\|_\infty \left\|\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right\|^2 \leq 2\|s\|_\infty \left(3 + 2w_n(y)\right)\mathcal{E} f_n(u).$$

Moreover, by the Cauchy-Schwarz inequality, on $\Omega_{n,y}$,

$$\begin{aligned}
\left\|\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right\|_\infty &= \left\|\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \hat{\theta}_i^T \psi_i\right\|_\infty \\
&\leq \sqrt{2|j|}\sqrt{\sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T\right)^2} \\
&\leq \sqrt{2|u|\Delta}\sqrt{2|u|\mathcal{E} + u w_n(y)\mathfrak{e}} \\
&\leq 4\sqrt{1 + w_n(y)} f_n(u)\sqrt{\Delta\mathcal{E}}.
\end{aligned}$$

It follows by Bernstein's inequality that on $\Omega_{n,y}$, with probability greater than $1 - e^{-y}$ conditionally on D_n^T, for all $u \in I_n \backslash \left[a_{x_n}, b_{x_n}\right]$,

$$\left(P_n^{T^c} - P\right)\left(\hat{s}^T_{k_*+u\Delta} - \hat{s}^T_{k_*}\right) \leq \kappa[\log^{3/2}(nM_n) + y^{3/2}]\sqrt{\|s\|_\infty}\mathfrak{e}\sqrt{f_n(u)} + \kappa[\log^2(nM_n) + y^2] f_n(u)\sqrt{\frac{\Delta}{n_v}}\mathfrak{e} \tag{18}$$

for some constant κ. By lemma 3.1,

$$\begin{aligned}
\sqrt{\frac{\Delta}{n_v}} &\leq \sqrt{\frac{2n_t or(n_t) + \frac{n_t}{n_v}}{n_v}} \\
&\leq \sqrt{\frac{2n_t or(n_t)}{n_v}} + \frac{\sqrt{n_t}}{n_v} \\
&\leq \sqrt{\frac{2n_t or(n_t)}{n^{\frac{2}{3}}}} + \frac{1}{n^{\frac{1}{6}}} \text{ by hypothesis5.}
\end{aligned}$$

Moreover, choosing $k = \left\lceil (c_1 n_t)^{\frac{1}{3+\delta_1}} \right\rceil$ yields, by hypothesis 1,

$$or(n_t) \leq 2\frac{c_1^{\frac{1}{3+\delta_1}}}{n_t^{\frac{2+\delta_1}{3+\delta_1}}} + \frac{1}{n_t}.$$

It follows that

$$\sqrt{\frac{\Delta}{n_v}} \leq \sqrt{4\frac{c_1^{\frac{1}{3+\delta_1}}}{n^{\frac{1}{3}}} + \frac{1}{n^{\frac{2}{3}}}} + \frac{1}{n^{\frac{1}{6}}} \leq \frac{\kappa}{n^{\frac{1}{6}}} \tag{19}$$

for some constant κ. Hence, by equations (17), (18), with probability greater than $(1-e^{-y})^2$, for all $u \in I_n \backslash \left[a_{x_n}, b_{x_n}\right]$,

$$\left| \frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1 \right| \leq \kappa[\log^2(nM_n) + y^2]n^{-\frac{1}{12}\wedge\frac{\delta_1}{4}} + \kappa[\log^{3/2}(nM_n) + y^{3/2}]\sqrt{\frac{\|s\|_\infty}{f_n(u)}}.$$

It follows that

$$\mathbb{E}\left[\sup_{u\in I_n\backslash\left[a_{x_n},b_{x_n}\right]} \left| \frac{\hat{R}_T^{ho}(u)}{f_n(u)} - 1 \right|\right] \leq \kappa\sqrt{\|s\|_\infty}\log^2(nM_n)n^{-\frac{1}{12}\wedge\frac{\delta_1}{4}\wedge\frac{u_1}{4}\wedge\frac{u_2}{2}}.$$

Together with equation (16), this yields the proposition. ■

Figure 2 gives an illustration of the situation for $x = 25$ and

$$f_n : \alpha \mapsto \begin{cases} e^{-\alpha} - 1 \text{ if } \alpha \leq 0 \\ \frac{8}{10}\alpha + \frac{8}{30}\alpha^3 \text{ if } \alpha \geq 0 \end{cases}$$

$$g_n : \alpha \mapsto \begin{cases} 7.8\alpha \text{ if } \alpha \geq 0 \\ 7.8\alpha - 3f_n(\alpha) \text{ if } \alpha \leq 0 \end{cases}$$

(which satisfy the properties of lemma 3.2 and Theorem 1 when $\|s\|^2 \leq 1.2$ and $\|s\|_\infty \leq 1.5$).

Figure 2 shows how for large x, $\sqrt{g_n}$ and hence W_{g_n} become negligible compared to f_n outside the interval $[a_x, b_x]$, as stated by proposition 5.1 . Hence, for sufficiently large x, the random term W_{g_n} becomes negligible relative to the deterministic f_n outside the interval $[a_x; b_x]$.

5.3 Incomplete $V-$fold cross-validation

Since the cross-validation risk estimator $\mathrm{CV}_{\mathcal{T}}(k)$ can be written as an average of hold-out risk estimators $\mathrm{HO}_{T_i}(k)$, Theorem 1 has direct implications for CV.

Corollary 2 *Assume that the hypotheses of section 4 hold. Let f_n, g_n be as in definition 7 and theorem 1. For any $x > 0$,*

$$\mathbb{E}\left[\sup_{u\in[a_x;b_x]} \left|\hat{R}_{\mathcal{T}}^{cv}(u) - (f_n(u) - W_{\frac{g_n(u)}{V}})\right|\right] \leq 5\kappa_0(1+x)^{\frac{3}{2}}n^{-u_1},$$

with the same constants κ_0, u_1 as in Theorem 1.

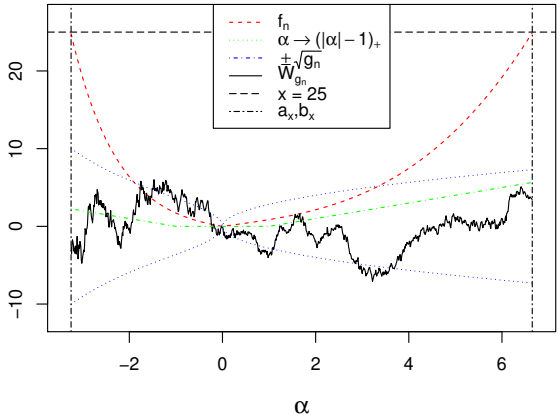

Figure 2: A plot of f_n, W_{g_n} on $[a_x; b_x]$, for $x = 25$, $g_n : \alpha \mapsto 7.8\alpha - 3f_n(\alpha)\mathbb{I}_{\alpha<0}$.

Proof By integrating the bound of Theorem 1,

$$\mathbb{E}\left[\sup_{u\in[a_x;b_x]}\left|\hat{R}^{ho}_{T_i}(u) - (f_n(u) - W^i_{\frac{g_n(u)}{V}})\right|\right] \leq 5\kappa_0(1+x)^{\frac{3}{2}}n^{-u_1}$$

for each $i \in \{1, \ldots, V\}$, where the W^i are symmetrical BMs that are independent of $D_n^{T_i}$. We can construct W^i such that $W^i = H(D_n^{T_i^c}, U_i)$, where H is a measurable function and U_i is an auxiliary uniform random variable. Taking independent U_i yields i.i.d W^i, since the sets $T_i^c = I_i$ are disjoint. Let $\bar{W} = \frac{1}{V}\sum_{i=1}^{V} W^i$. By Jensen's inequality,

$$\mathbb{E}\left[\sup_{u\in[a_x;b_x]}\left|\hat{R}^{cv}_{\mathcal{T}}(u) - (f_n(u) - \bar{W}_{g_n(u)})\right|\right] \leq 5\kappa_0(1+x)^{\frac{3}{2}}n^{-u_1}.$$

Conclude by noting that $(\bar{W}_t)_{t\in\mathbb{R}}$ is equal in distribution to $(W_{t/V})_{t\in\mathbb{R}}$, as a continuous random process. ■

Corollary 2 proves that cross-validation is effective at reducing the variance of risk estimation, compared to simple validation (the hold-out). The process approximating $\hat{R}^{cv}_{\mathcal{T}}$ is of the same form as that approximating $\hat{R}^{ho}_{T}$, but with its variance reduced by a factor V, as would be the case if the hold-out estimators $\mathrm{HO}_{T_i}(\cdot)$ were independant. Importantly, this reduction in variance occurs for the rescaled process $\hat{R}^{cv}_{\mathcal{T}}$, and so can be expected to reflect the model selection performance. Corollary 2 is sharp when V is fixed as $n \to +\infty$, since in that case, the approximating process $f_n - W_{g_n/V}$ remains nontrivial (random and of bounded size) as $n \to +\infty$. When $V = V_n \to +\infty$, corollary 2 implies that $\hat{R}^{cv}_{\mathcal{T}}$ concentrates around the deterministic function f_n. If the seauence V_n grows

slowly enough, such that $V_n = o(n^{u_1})$, corollary 2 also provides the rate of this convergence (which is $\frac{1}{\sqrt{V_n}}$).

An analog to Proposition 5.1 for cross-validation follows likewise from Jensen's inequality. Hence, at scales asymptotically larger than Δ or outside intervals $[a_{x_n}, b_{x_n}]$, $x_n \to +\infty$, cross-validation also concentrates around the excess risk function f_n.

6 Discussion

In this section, we interpret the results of the article and discuss how they could be extended.

6.1 Implications of our results

Theorem 1 provides an approximation to the rescaled hold-out process, with scale factor Δ given by Definition 5. The approximating process is asymptotically random, tight along a sequence of intervals of length greater than 1 and trends away from zero at $\pm\infty$ (as discussed in section 5.2). In the case of "incomplete" cross-validation for fixed V, the process is of identical type, but with the variance reduced by a factor V. This means that CV performs significantly better than the hold-out when comparing $\hat{s}_k^T$ and $\hat{s}_{k_*}^T$ for k close to $k_*(n_t)$. If $V \to +\infty$, then rescaled cross-validation concentrates around the deterministic function f_n, which by the argument of section 3.3 means that CV has asymptotically better performance than the hold-out. This supports the belief that cross-validation improves when V is increased. Though $k_*(n_t)$ is not the oracle based on the full sample $(k_*(n))$, the greater concentration of CV around $k_*(n_t)$ makes it possible to choose n_t closer to n than would be reasonable for the hold-out, resulting in improved overall performance.

6.2 Hypotheses

Theorem 1 relies on the assumptions of section 4, most importantly on the assumption that the sequence of squared Fourier coefficients θ_j^2 is non-increasing.

This could be weakened in various ways. First, given the local nature of our analysis, what is really required is that the coefficients θ_j^2 be non-increasing in a neighbourhood $[k_* - r_n, k_* + r_n]$ of k_* of radius r_n which dominates Δ ($\Delta = o(r_n)$). The only difference in that case is that there may be other local minima outside $[k_* - r_n, k_* + r_n]$, which diminishes the usefulness of Theorem 1 for studying the whole process.

Moreover, since the process $\hat{R}_{\mathcal{T}}^{cv}(\alpha)$ consists of sums $\sum_{j=1}^{k_*+\alpha\Delta}$, it is probably sufficient to replace the hypotheses on the individual coefficients θ_j^2 with hypotheses bearing on local averages $\bar{\theta}_j^2 = \frac{1}{2m_n}\sum_{r=j-m_n}^{j+m_n}\theta_r^2$, at some scale $m_n = o(\Delta)$. Depending on the scale m_n, the hypothesis that a smoothed sequence $\bar{\theta}_j^2$ is non-decreasing may be quite plausible, considering the fact that the Fourier coefficients tend to 0 at a prescribed rate for sufficiently smooth functions s. For example, it suffices to assume that $\sum_{j=k+1}^{+\infty}\theta_j^2 \sim L(k)k^{-2\alpha}$ (for some $\alpha > 0$ and a slowly varying function L) in order for the local averages to be non-increasing at some scale $m_n = o(k_*(n_t))$. Note that Δ may be of order $k_*(n_t)$ for small enough n_v.

6.3 Perspectives

Model selection The CV risk estimator is usually used to select a model, $\hat{k}^{cv}_{\mathcal{T}}$, which minimizes it. The final result of CV is then the estimator $\hat{s}_{\hat{k}^{cv}_{\mathcal{T}}}$, or $\hat{s}^T_{\hat{k}_T}$ in the case of simple validation. Thus, what we are most interested in practice is the risk $\left\|\hat{s}_{\hat{k}^{cv}_{\mathcal{T}}} - s\right\|^2$ of this final estimator, and how it depends on the CV method used (at least when CV is used with a goal of *estimation*, as opposed to *identification* of the best model). There are several ways our results can contribute to answering these questions. First, it follows from proposition 5.1 and Markov's inequality that $\frac{\hat{k}^{cv}_{\mathcal{T}}}{\Delta} \in [a_x, b_x]$ with high probability for x large enough. Since $\hat{R}^{cv}_{\mathcal{T}}(u)$ can be uniformly approximated by $f_n(u) - W_{g_n(u)/V}$ on $[a_x, b_x]$, it is natural to approximate $\frac{\hat{k}^{cv}_{\mathcal{T}} - k_*}{\Delta}$ (the minimizer of $\hat{R}^{cv}_{\mathcal{T}}(u)$) by the minimizer $\hat{\alpha}_{n,V}$ of $f_n - W_{g_n/V}$). The minima of $f_n - W_{g_n/V}$ can be studied using the theory of Wiener processes. Together with our results about f_n and g_n (in lemma 3.2 and Theorem 1), this makes the analysis of $\hat{\alpha}_{n,V}$ much easier than that of $\hat{k}^{cv}_{\mathcal{T}}$. Moreover, claim 3 of this article proves that the (excess) risk $\left\|\hat{s}^T_k - s\right\|^2 - \left\|\hat{s}^T_{k_*} - s\right\|^2$ concentrates around $\mathfrak{e} f_n\left(\frac{k-k_*}{\Delta}\right)$ for k "close enough" to k_*. This removes the dependency on the training sample D^T_n and reduces the analysis of $\left\|\hat{s}^T_k - s\right\|^2 - \left\|\hat{s}^T_{k_*} - s\right\|^2$ to that of the deterministic function $f_n\left(\frac{k-k_*}{\Delta}\right)$.

Other cross-validation methods In this article, we only considered a particular type of cross-validation. Corollary 2 relies on decomposing "incomplete" $V-$fold CV as a finite average of asymptotically independent hold-out estimators, which can be analysed more easily than general cross-validation because conditionally on their training data, they are empirical processes.

In general, for projection estimators in L^2 density estimation, the cross-validation risk estimator is not a (conditional) empirical process but a (weighted) U-statistic of order 2 (more precisely, a weighted sum of the terms $\psi_k(X_i)\psi_k(X_j)$). Thus, new methods are required to approximate general CV estimators by gaussian processes. We conjecture that more general cross-validation methods also behave locally like the sum of the (rescaled) excess risk and a time-changed Wiener process, though we expect the scaling and the time-change g_n to be different for different versions of CV.

Theorem 1 can also shed light on the behaviour of other methods which use simple validation as a key ingredient, such as Aggregated hold-out [17].

Acknowledgements

While finishing the writing of this article, the author (Guillaume Maillard) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 811017.

7 Proofs

In this section, the term constant means a function of $\|s\|_\infty, \|s\|^2$ and the constants $c_1, c_2, c_3, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5$. which appear in the hypotheses of Theorem 1. Note that by hypothesis (1), $\|\theta\|_{\ell^1}, \|s\|_\infty, \|s\|^2$ are finite and can be bounded by functions of c_1, δ_1. The letter u will denote strictly positive constants that only depend on $(\delta_i)_{1\leq i\leq 5}$ (they will generally appear as exponents of $\frac{1}{n}$). The letter κ denotes a non-negative constant. The notation $n_v = n - n_t$ will also be used frequently.

7.1 Preliminary results

The results of this section are independent from the rest. They will be used in the rest of the proof of Theorem 1, as well as in the Appendix. Let's start by proving some basic properties of a_x, b_x and f_n that will be used repeatedly in the main proofs.

7.1.1 Proof of lemma 3.1

- By definition and non-negativity of θ_j^2, $\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta_d}} \leq 1$, therefore $\Delta \geq \Delta_d \geq \frac{n_t}{n-n_t}$.
- $\mathcal{E} = \frac{\Delta}{n_t} \geq \frac{1}{n-n_t}$.
- $\mathfrak{e} = \sqrt{\frac{\mathcal{E}}{n-n_t}} \geq \sqrt{\frac{1}{(n-n_t)^2}} = \frac{1}{n-n_t}$.
- $\frac{\mathcal{E}}{\mathfrak{e}} = \mathcal{E}\sqrt{\frac{n-n_t}{\mathcal{E}}} = \sqrt{(n-n_t)\mathcal{E}} \geq 1$.
- By definition, $\Delta_g \leq k_*$. Thus $\frac{\Delta_g}{n_t} \leq \frac{k_*}{n_t} \leq \mathrm{or}(n_t)$. Moreover,

$$\Delta_d\left[1 - \sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta_d}}\right]\frac{1}{n_t} \leq \sum_{j=k_*+1}^{k_*+\Delta_d}\theta_j^2 \leq \sum_{j=k_*+1}^{+\infty}\theta_j^2 \leq \mathrm{or}(n_t).$$

  Thus

$$\begin{aligned} n_t\mathrm{or}(n_t) &\geq \Delta_d - \sqrt{\frac{n_t}{n-n_t}}\sqrt{\Delta_d} \\ &\geq \Delta_d - \frac{1}{2}\frac{n_t}{n-n_t} - \frac{1}{2}\Delta_d \\ &\geq \frac{1}{2}\Delta_d - \frac{1}{2}\frac{n_t}{n-n_t}. \end{aligned}$$

  It follows that

$$\Delta_d \leq 2n_t\mathrm{or}(n_t) + \frac{n_t}{n-n_t},$$

  so since $\frac{n_t}{n-n_t}\frac{1}{n_t} = \frac{1}{n-n_t}$,

$$\frac{\Delta_d}{n_t} \leq 2\mathrm{or}(n_t) + \frac{1}{n-n_t},$$

  which proves the result.

7.1.2 Proof of lemma 3.2

f_n is continuous and piecewise linear by definition 7. f_n is convex because the sequence θ_j^2 is non-increasing by assumption. Let $j \in \mathbb{Z}$ and $\alpha \in \left]\frac{j}{\Delta}; \frac{j+1}{\Delta}\right[$ be two numbers. By definition, f_n is linear on the interval $\left]\frac{j}{\Delta}; \frac{j+1}{\Delta}\right[$, in particular f_n is differentiable on this interval and

$$\begin{aligned} f_n'(\alpha) &= \Delta\left[f_n\left(\tfrac{j+1}{\Delta}\right) - f_n\left(\tfrac{j}{\Delta}\right)\right] \\ &= \frac{\Delta}{\mathfrak{e}}\left[\frac{1}{n_t} - \theta_{k_*+j+1}^2\right]. \end{aligned} \tag{20}$$

Because the sequence θ_j^2 is non-increasing, it follows from the definition of $k_*(n_t)$ that f_n is increasing on $\left]\frac{j}{\Delta}; \frac{j+1}{\Delta}\right[$ if $j \geq 0$ and non-increasing if $j < 0$. This implies that f_n reaches its minimum at $k_*(n_t)$. If $\alpha \geq 1$, then $j = \lfloor \alpha\Delta \rfloor \geq \Delta$, therefore by definition of $\Delta_d \leq \Delta$,

$$\begin{aligned} f_n'(\alpha) &\geq \frac{\Delta}{\mathfrak{e}}\left[\frac{1}{n_t} - \theta_{k_*+\Delta+1}^2\right] \\ &\geq \frac{\Delta}{\mathfrak{e}}\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta}}\frac{1}{n_t} \\ &= \Delta\sqrt{\frac{(n-n_t)n_t}{\Delta}}\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta}}\frac{1}{n_t} \\ &= 1. \end{aligned}$$

In the same way, if $\alpha < -1$, then $j+1 = \lceil \alpha\Delta \rceil \leq -\Delta \leq -\Delta_g$, so

$$\begin{aligned} f_n'(\alpha) &\leq \frac{\Delta}{\mathfrak{e}}\left[\frac{1}{n_t} - \theta_{k_*-\Delta}^2\right] \\ &\leq -\frac{\Delta}{\mathfrak{e}}\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta}}\frac{1}{n_t} \\ &\leq -1. \end{aligned}$$

Furthermore,

- If $\Delta = \Delta_d$, then for all $\alpha \in [0;1]$, $j+1 = \lceil \alpha\Delta \rceil \leq \Delta = \Delta_d$, therefore by definition of Δ_d,

$$\begin{aligned} f_n'(\alpha) &\leq \frac{\Delta}{\mathfrak{e}}\left[\frac{1}{n_t} - \theta_{k_*+\Delta}^2\right] \\ &\leq \frac{\Delta}{\mathfrak{e}}\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta}}\frac{1}{n_t} \\ &\leq 1. \end{aligned}$$

- If $\Delta = \Delta_g$, then for all $\alpha \in [-1;0]$, $j = \lfloor \alpha\Delta \rfloor \geq -\Delta = -\Delta_g$, therefore by definition of Δ_g and since the sequence $(\theta_j^2)_{j\in\mathbb{N}}$ is non-increasing,

$$\begin{aligned} f_n'(\alpha) &\geq \frac{\Delta}{\mathfrak{e}}\left[\frac{1}{n_t} - \theta_{k_*-\Delta+1}^2\right] \\ &\geq -\frac{\Delta}{\mathfrak{e}}\sqrt{\frac{n_t}{n-n_t}}\frac{1}{\sqrt{\Delta}}\frac{1}{n_t} \\ &\geq -1. \end{aligned}$$

By continuity of f_n, this proves the lemma.

7.1.3 Properties of the interval $[a_x; b_x]$

Lemme 7.1 *Let a_x, b_x be as defined in Theorem1. Then for all $x > 0$,*

$$[b_x - a_x] \leq 2(1+x)$$
$$\sum_{k_*+a_x\Delta}^{k_*+b_x\Delta} \theta_j^2 \leq 4(1+x)\mathcal{E}.$$

Proof Either $b_x \leq 1$, or $b_x > 1$ and by lemma 3.2, $f_n(b_x) \geq b_x - 1$ which implies that $b_x \leq f_n(b_x) + 1 \leq x + 1$. In all cases, $b_x \leq x + 1$. In the same way, $a_x > -1 - x$. Thus $b_x - a_x \leq 2(1+x)$. Moreover,

$$\begin{aligned}\sum_{k_*+a_x\Delta}^{k_*+b_x\Delta} \theta_j^2 &\leq \sum_{k_*+a_x\Delta}^{k_*+b_x\Delta} \left|\theta_j^2 - \frac{1}{n_t}\right| + \frac{(b_x - a_x)\Delta}{n_t} \\ &\leq \mathfrak{e}[f_n(a_x) + f_n(b_x)] + [b_x - a_x]\mathcal{E} \\ &\leq 2x\mathfrak{e} + 2\,[1+x]\,\mathcal{E}.\end{aligned}$$

Since $\mathfrak{e} \leq \mathcal{E}$ by lemma 3.1,

$$\sum_{k_*+a_x\Delta}^{k_*+b_x\Delta} \theta_j^2 \leq (4x+2)\mathcal{E}.$$

This proves lemma 7.1. ■

We now introduce some notation which will be used in the remainder of this chapter.

Definition 9 *Let an i.i.d sample D_n be given, with distribution P and pdf s on $[0;1]$. For all $j \in \mathbb{N}$ and any $T \subset \{1, \ldots, n\}$, let*

$$\theta_j = P\psi_j = \langle s, \psi_j \rangle$$
$$\hat{\theta}_j^T = P_n^T(\psi_j).$$

This notation will be used very often in the remainder of the chapter.

The hold-out risk estimator can be expressed as the sum of two terms. Definition 10 below gives a name to each of these terms.

Definition 10 *For all $j \in [-k_*(n_t); +\infty[\cap\mathbb{Z}$, let*

$$L(\tfrac{j}{\Delta}) = \frac{1}{\mathfrak{e}}\left(\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2\right).$$

The function L is extended to the interval $[-\frac{k_*(n_t)}{\Delta}; +\infty[$ by linear interpolation. Let Z be the random function defined for all $j \in [-\frac{k_*(n_t)}{\Delta}; +\infty[\cap\mathbb{Z}$ by

$$Z\left(\frac{j}{\Delta}\right) = \frac{2}{\mathfrak{e}}\left(P_n^{T^c} - P\right)\left(\hat{s}_{k_*+j}^T - \hat{s}_{k_*}^T\right)$$

and extended by linear interpolation to the interval $[-\frac{k_*(n_t)}{\Delta}; +\infty[$, so that for all α, $\hat{R}_T^{ho}(\alpha) = L(\alpha) - Z_\alpha$.

Thus, L is the rescaled excess risk, and Z is a centered empirical process. These two terms will be approximated separately.

7.2 Approximation of the excess risk

Let $x > 0$ be fixed for the entirety of this section. We now prove the following claim.

Claim 3 *Let L be the function introduced in definition 10, and f_n be given by definition 7. There exists a constant κ_1 such that, for any $y > 0$, with probability greater than $1 - e^{-y}$,*

$$\sup_{\alpha \in [a_x; b_x]} |L(\alpha) - f_n(\alpha)| \leq \kappa_1 (1 + x)[\log(2 + x) + y + \log n]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})}.$$

Proof Let $j \in \{a_x\Delta, \ldots, b_x\Delta\}$. Since $\hat{s}_k^T = \sum_{j=1}^k P_n^T(\psi_j)\psi_j = \sum_{j=1}^k \hat{\theta}_j^T \psi_j$,

$$\left\|\hat{s}_{k_*+j}^T - s\right\|^2 - \left\|\hat{s}_{k_*}^T - s\right\|^2 = \operatorname{sgn}(j) \sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 - \theta_i^2.$$

It is known [4, Lemma 14] [16, Proposition 6.3] that the process

$$\sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 = \sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} (P^T - P)(\psi_j)^2$$

concentrates around its expectation, so that

$$\sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 \sim \sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \frac{\operatorname{Var}(\psi_j)}{n_t}.$$

Furthermore, by lemma 8.1 in the appendix, $\operatorname{Var}(\psi_j) \sim 1$, therefore

$$\sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 \sim \frac{|j|}{n_t}.$$

More precisely, proposition 8.3 in the appendix and a union bound show that, with probability greater than $1 - e^{-y}$, for any $j \in \mathbb{Z} \cap [a_x\Delta; b_x\Delta[$,

$$\left| \sum_{i=k_*+(j)_-+1}^{k_*+(j)_+} \left(\hat{\theta}_i^T - \theta_i\right)^2 - \frac{|j|}{n_t} \right| \leq \kappa_1 (y + \log n + \log((b_x - a_x)\Delta))^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})} \frac{j}{\Delta} \mathfrak{e}.$$

Let $r_n = \kappa_1(y + \log n + \log((b_x - a_x)\Delta))^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})}$. Then for any $j \geq 1$,

$$\begin{aligned}
\left\|\hat{s}^T_{k_*+j} - s\right\|^2 - \left\|\hat{s}^T_{k_*} - s\right\|^2 &= -\sum_{i=k_*+1}^{k_*+j} \theta_i^2 + \sum_{i=k_*+1}^{k_*+j} \left(\hat{\theta}_i^T - \theta_i\right)^2 \\
&= -\sum_{i=k_*+1}^{k_*+j} \theta_i^2 + \frac{j}{n_t} \pm \frac{j}{\Delta} r_n \mathfrak{e} \\
&= \sum_{i=k_*+1}^{k_*+j} \left[\frac{1}{n_t} - \theta_i^2\right] \pm \frac{j}{\Delta} r_n \mathfrak{e} \\
&= \mathfrak{e} f_n\left(\frac{j}{\Delta}\right) \pm \frac{j}{\Delta} r_n \mathfrak{e}.
\end{aligned}$$

On this same event, for any $j \in \{-k_*(n_t), \ldots, -1\}$,

$$\begin{aligned}
\left\|\hat{s}^T_{k_*+j} - s\right\|^2 - \left\|\hat{s}^T_{k_*} - s\right\|^2 &= \sum_{i=k_*+j+1}^{k_*} \theta_i^2 - \sum_{i=k_*+j+1}^{k_*} \left(\hat{\theta}_i^T - \theta_i\right)^2 \\
&= \sum_{i=k_*+j+1}^{k_*} \theta_i^2 - \frac{|j|}{n_t} \pm \frac{|j|}{\Delta} r_n \mathfrak{e} \\
&= \sum_{i=k_*+j+1}^{k_*} \left[\theta_j^2 - \frac{1}{n_t}\right] \pm \frac{j}{\Delta} r_n \mathfrak{e} \\
&= \mathfrak{e} f_n\left(\frac{j}{\Delta}\right) \pm \frac{j}{\Delta} r_n \mathfrak{e}.
\end{aligned}$$

Thus, since f_n and $\hat{R}_T^{ho}$ are linear between the points of $\frac{1}{\Delta}\mathbb{Z}$,

$$\begin{aligned}
\sup_{\alpha \in [a_x; b_x]} |L(\alpha) - f_n(\alpha)| &= \frac{1}{\mathfrak{e}} \max_{a_x\Delta \leq j \leq b_x\Delta} \left| \left\|\hat{s}^T_{k_*+j} - s\right\|^2 - \left\|\hat{s}^T_{k_*} - s\right\|^2 - \mathfrak{e} f_n\left(\tfrac{j}{\Delta}\right) \right| \\
&\leq \max(|a_x|, |b_x|) r_n.
\end{aligned}$$

By lemma 7.1, $\max(|a_x|, |b_x|) \leq b_x - a_x \leq 2(1 + x)$ so

$$\begin{aligned}
\max(|a_x|, |b_x|) r_n &\leq 2(1+x)\kappa_1(y + \log n + \log(2(1+x)) + \log(\Delta))^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})} \\
&\leq \kappa(1+x)[\log(2+x) + \log n + y]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})}
\end{aligned}$$

for some constant κ, since by lemma 3.1 and hypothesis (5) of Theorem 1,

$$\begin{aligned}
\Delta &= n_t \mathcal{E} \\
&\leq 2n_t or(n_t) + \frac{n_t}{n - n_t} \\
&\leq 2(\|s\|^2 - 1)n_t + n_t^{\frac{1}{3}}.
\end{aligned}$$

This proves claim 3. ■

We will now seek to approximate the process Z given by definition 10.

7.3 Strong approximation of the hold-out process

Let us start by showing that the empirical process Z (definition 10) can be approximated by a gaussian process, uniformly on $[a_x; b_x]$. This is the purpose of the following result, which will be proven in this section.

Claim 4 *Let Z be the process given by definition 10. There exists a gaussian process $(Z^1_\alpha)_{\alpha\in[a_x;b_x]}$ with the same variance-covariance function as Z: for any $(\alpha_1,\alpha_2)\in[a_x;b_x]^2$, $\mathrm{Cov}(Z^1_{\alpha_1}, Z^1_{\alpha_2}) = \mathrm{Cov}(Z_{\alpha_1}, Z_{\alpha_2})$ and such that for all $n\geq 1$, for all $x>0$, with probability greater than $1-e^{-y}$,*

$$\mathbb{E}\left[\sup_{\alpha\in[a_x;b_x]} |Z_\alpha - Z^1_\alpha| \middle| D_n^T\right] \leq \kappa_5(c_1,\delta_5)(1+y)(1+x)^{\frac{3}{2}} n^{-\frac{\delta_5}{3}}.$$

Furthermore, Z^1 can be expressed as $Z^1 = H(D_n^{T^c}, \nu)$, with ν a uniform random variable independent from D_n and H a measurable function on $C([0;1],\mathbb{R})$.

Let $n_v = |T^c| = n - |T| = n - n_t$. Let $F : x \to \int_0^x s(t)dt$ be the cumulative distribution function of the given X_i. Let $F_{T^c} : x \to \frac{1}{n_v}\sum_{i\notin T} \mathbb{I}_{X_i\leq x}$ be the empirical cumulative distribution function of the sample $D_n^{T^c}$. By the Komlós-Major-Tusnády approximation theorem [15, Theorem 3], there exist a universal constant C and a standard Brownian bridge process B_{T^c} such that for all $y>0$, with probability greater than $1-e^{-y}$, $\left\|B_{T^c}\circ F - \sqrt{n_v}(F_{T^c}-F)\right\|_\infty \leq \frac{C(\log n_v + y)}{\sqrt{n_v}}$ (remark that since F is continuous, $F(X_i)\sim\mathcal{U}([0;1])$, which means that the result for general F follows from the result for the uniform distribution). Furthermore, B_{T^c} can always be realized as a measurable function of $D_n^{T^c}$ and an auxiliary, uniformly distributed random variable ν: $B_{T^c} = H(D_n^{T^c},\nu)$, with ν independant from D_n. Let B^{T^c} be obtained in this way. From $B_{T^c}\circ F$, one can define an operator on the Sobolev space $W^1(\mathbb{R})$:

Definition 11 *For any function f such that $f'\in L^1([0;1])$, let*

$$G_{T^c}(f) = -\int_0^1 f'(x) B_{T^c}(F(x))dx.$$

G_{T^c} "approximates" the empirical process $\sqrt{n_v}(P_n^{T^c} - P)$ on the space W^1. Lemma 7.2 below gives a bound on the error made with this approximation.

Lemme 7.2 *For any function f such that $f'\in L^1([0;1])$,*

$$\left|G_{T^c}(f) - \sqrt{n_v}(P_n^{T^c} - P)(f)\right| \leq \|B_{T^c} - \sqrt{n_v}(F_{T^c}-F)\|_\infty \|f'\|_{L^1}.$$

Furthermore, for all functions f,g such that $f',g'\in L^1([0;1])$,

$$\mathrm{Cov}(G_{T^c}(f), G_{T^c}(g)) = P[fg] - P[f]P[g] = \mathrm{Cov}\left(\sqrt{n_v}(P_n^{T^c}-P)(f), \sqrt{n_v}(P_n^{T^c}-P)(g)\right).$$

Proof Let f be a function such that $f' \in L^1([0;1])$. Then

$$\begin{aligned}
(P_n^{T^c} - P)(f) &= \int f d(P_n^{T^c} - P) \\
&= \int [f - f(0)] d(P_n^{T^c} - P) \\
&= \int_0^1 \int_0^1 \mathbb{I}_{t<x} f'(t) dt \; d(F_{T^c} - F)(x) \\
&= \int_0^1 f'(t)(P_n^{T^c} - P)((t, +\infty)) \\
&= -\int_0^1 f'(t)(F_{T^c} - F)(t) dt. \qquad (21)
\end{aligned}$$

Il follows that for all functions f such that $f' \in L^1([0;1])$,

$$\begin{aligned}
\left| G_{T^c}(f) - \sqrt{n_v}(P_n^{T^c} - P)(f) \right| &= \left| \int_0^1 f'(t) \left[\sqrt{n_v}(F_{T^c} - F) - B_{T^c} \circ F \right](t) dt \right| \\
&\leq \|f'\|_{L^1([0;1])} \left\| B_{T^c} \circ F - \sqrt{n_v}(F_{T^c} - F) \right\|_\infty .
\end{aligned}$$

By definition, it is clear that $\mathbb{E}[G_{T^c}(f)] = 0$. Thus,

$$\begin{aligned}
\mathrm{Cov}\left(G_{T^c}(f), G_{T^c}(g)\right) &= \mathbb{E}\left[G_{T^c}(f) G_{T^c}(g)\right] \\
&= \mathbb{E}\left[\int_0^1 \int_0^1 f'(u) g'(v) B_{T^c}(F(u)) B_{T^c}(F(v)) \right] \\
&= \int_0^1 \int_0^1 f'(u) g'(v) [F(u) \wedge F(v)][1 - F(u) \vee F(v)] du dv \\
&= \int_0^1 \int_0^1 f'(u) g'(v) \left(\mathbb{E}[\mathbb{I}_{X \leq u} \mathbb{I}_{X \leq v}] - \mathbb{E}[\mathbb{I}_{X \leq u}] \mathbb{E}[\mathbb{I}_{X \leq v}] \right) \\
&= n_v \int_0^1 \int_0^1 f'(u) g'(v) \mathbb{E}\left[(F_{T^c} - F)(u)(F_{T^c} - F)(v) \right] \\
&= \mathrm{Cov}\left(\sqrt{n_v}(P_n^{T^c} - P)(f), \sqrt{n_v}(P_n^{T^c} - P)(g) \right) \text{ by equation (21)}
\end{aligned}$$

■

Let the process Z^1 be defined for all $j \in \{a_x \Delta, \ldots, b_x \Delta\}$ by

$$Z^1\left(\frac{j}{\Delta}\right) = \frac{2}{\sqrt{n_v}\mathfrak{e}} G_{T^c}\left(\hat{s}^T_{k_*+j} - \hat{s}^T_{k_*}\right).$$

Z^1 is extended to the interval $[a_x; b_x]$ by linear interpolation, as for Z. By lemma 7.2, the variance-covariance function of Z^1 conïncides with that of Z at the points $\frac{j}{\Delta}, j \in \mathbb{Z} \cap [a_x \Delta; b_x \Delta]$, and this property extends by bilinearity to the

whole interval $[a_x; b_x]$. Furthermore,

$$\begin{aligned}
\sup_{a_x \le \alpha \le b_x} |Z^1_\alpha - Z_\alpha| &\le \max_{j \in \mathbb{Z} \cap [a_x\Delta; b_x\Delta]} \left| Z^1\left(\frac{j}{\Delta}\right) - Z\left(\frac{j}{\Delta}\right)\right| \\
&\le \frac{4\sqrt{2}\pi}{\mathfrak{e}\sqrt{n_v}} \left\| B_{T^c} \circ F - \sqrt{n_v}(F_{n_v} - F)\right\|_\infty \times \max_{a_x\Delta \le j \le b_x\Delta} \left\| \sum_{i=k_*+1}^{k_*+j} i\hat{\theta}^T_i \sin(2i\pi\cdot)\right\|_1 \\
&\le \frac{4\pi}{\mathfrak{e}\sqrt{n_v}} \left\| B_{T^c} \circ F - \sqrt{n_v}(F_{n_v} - F)\right\|_\infty \times \sqrt{\sum_{i=k_*+a_x\Delta+1}^{k_*+b_x\Delta} i^2(\hat{\theta}^T_i)^2}
\end{aligned}$$

By construction, the process $B_{T^c} \circ F - \sqrt{n_v}(F_{n_v} - F)$ is independent from D^T_n. As a result,

$$\begin{aligned}
\mathbb{E}\left[\sup_{a_x \le \alpha \le b_x} |Z^1_\alpha - Z_\alpha| | D^T_n\right] &\le \frac{4\pi}{\mathfrak{e}\sqrt{n_v}} \mathbb{E}\left[\left\| B_{T^c} \circ F - \sqrt{n_v}(F_{n_v} - F)\right\|_\infty\right] \\
&\quad \times (k_* + b_x\Delta)\sqrt{\sum_{i=k_*+a_x\Delta+1}^{k_*+b_x\Delta} \left(\hat{\theta}^T_j\right)^2} \\
&\le \frac{4\pi C \log n_v}{\mathfrak{e} n_v} \times (k_* + b_x\Delta)\sqrt{\sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta} 2\theta^2_j + 2\left(\hat{\theta}^T_j - \theta_j\right)^2}.
\end{aligned} \tag{22}$$

By proposition 8.3, there exists an event $E_1(y)$ of probability greater than $1-e^{-y}$ such that, for all $D^T_n \in E_1(y)$,

$$\sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta} \left(\hat{\theta}^T_j - \theta_j\right)^2 \le [b_x - a_x]\frac{\Delta}{n_t} + \kappa_1(b_x - a_x)[\log n + y]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})}\mathfrak{e}(n),$$

therefore by lemma 7.1 and equation (22), for all $D^T_n \in E_1(y)$,

$$\mathbb{E}\left[\sup_{a_x \le \alpha \le b_x} |Z^1_\alpha - Z_\alpha| \Big| D^T_n\right] \le \frac{4\pi C \log n_v}{\mathfrak{e} n_v} \times (k_* + b_x\Delta) 2\sqrt{1+x}\left[2\sqrt{\mathcal{E}} + \sqrt{2\kappa_1}(\log n + y) n^{-\min(\frac{1}{24}, \frac{\delta_4}{4})}\sqrt{\mathfrak{e}}\right].$$

Since $\mathfrak{e} \le \mathcal{E}$ and $n^{-\min(\frac{1}{24}, \frac{\delta_4}{4})} \log n \to 0$, there exists therefore a constant κ such that for all $D^T_n \in E_1(y)$:

$$\begin{aligned}
\mathbb{E}\left[\sup_{a_x \le \alpha \le b_n} |Z^1_\alpha - Z_\alpha| \Big| D^T_n\right] &\le \kappa \frac{\log n_v}{\sqrt{n_v}} \frac{\sqrt{\mathcal{E}}}{\mathfrak{e}\sqrt{n_v}} \times (k_* + b_x\Delta)(1+y)\sqrt{1+x} \\
&\le \kappa \frac{\log n_v}{\sqrt{n_v}} \times (k_* + b_x\Delta)(1+y)\sqrt{1+x}.
\end{aligned} \tag{23}$$

By lemma 3.1, $\mathcal{E} \le 2\mathrm{or}(n_t) + \frac{1}{n_v}$ therefore $\Delta \le 2n_t\mathrm{or}(n_t) + \frac{n_t}{n_v}$ and by definition of *or*, $k_*(n_t) = n_t\frac{k_*}{n_t} \le n_t\mathrm{or}(n_t)$ therefore $\frac{k_*+b_x\Delta}{\sqrt{n_v}} \le (2b_x+1)\frac{n_t\mathrm{or}(n_t)}{\sqrt{n_v}} + \frac{b_x n_t}{n_v\sqrt{n_v}}$. By hypothesis 5 of section 4, $n_v \ge n^{\frac{2}{3}+\delta_5}$, so

$$\frac{k_* + b_x\Delta}{\sqrt{n_v}} \le (2b_x+1)\frac{n_t\mathrm{or}(n_t)}{n^{\frac{1}{3}+\frac{\delta_5}{2}}} + \frac{b_x}{n^{\frac{3\delta_5}{2}}}.$$

Moreover, by hypothesis 1 of Theorem 1, $\sum_{j=k+1}^{+\infty}\theta_j^2 \leq \frac{c_1}{k^{2+\delta_1}}$, therefore

$$\mathrm{or}(n_t) \leq \min_{k\in\mathbb{N}^*} \frac{c_1}{k^{2+\delta_1}} + \frac{k}{n_t} \leq 2\inf_{x\geq 1}\frac{c_1}{x^{2+\delta_1}} + \frac{x}{n_t} \leq 3\frac{c_1^{\frac{1}{3+\delta_1}}}{n_t^{\frac{2+\delta_1}{3+\delta_1}}},$$

whence (since $c_1 \geq 1$) $n_t\mathrm{or}(n_t) \leq 3(c_1 n_t)^{\frac{1}{3+\delta_1}}$. It follows that:

$$\log n\frac{k_* + b_x\Delta}{\sqrt{n_v}} \leq 3(2b_x+1)c_1^{\frac{1}{3+\delta_1}}\log n n^{-\frac{\delta_5}{2}} + \frac{b_x\log n}{n_t^{\frac{3\delta_5}{2}}}. \tag{24}$$

Since $\frac{\log n}{n^{\frac{\delta_5}{2}}} = o\left(n^{-\frac{\delta_5}{3}}\right)$, by equations (23), (24) and lemma 7.1, there exists a constant $\kappa(c_1,\delta_5)$ such that for any n, with probability greater than $1-e^{-y}$,

$$\mathbb{E}\left[\sup_{a_x\leq\alpha\leq b_n}\left|Z_\alpha^1 - Z_\alpha\right|\middle| D_n^T\right] \leq \kappa(1+y)(1+x)^{\frac{3}{2}}n^{-\frac{\delta_5}{3}}.$$

7.4 Approximation of the covariance function

We will now seek to approximate the process Z^1 given by claim 4 by a time-changed Wiener process. To this end, we first approximate the variance-covariance function of Z^1 (which is the same as that of Z).

Claim 5 *There exists a function g_n satisfying the hypotheses of Theorem 1 and a constant $u_5 > 0$ such that, for all $x, y > 0$, with probability greater than $1-e^{-y}$,*

$$\max_{(j_1,j_2)\in\{0,\ldots,b_x\Delta\}^2}\left|\mathrm{Cov}\left(Z\left(\tfrac{j_1}{\Delta}\right), Z\left(\tfrac{j_2}{\Delta}\right)|D_n^T\right) - \min\left(g_n\left(\tfrac{j_1}{\Delta}\right), g_n\left(\tfrac{j_2}{\Delta}\right)\right)\right| \leq \kappa_6(1+x)^2[y+\log n]^2 n^{-u_5}$$

$$\max_{(j_1,j_2)\in\{a_x\Delta,\ldots,0\}^2}\left|\mathrm{Cov}\left(Z\left(\tfrac{j_1}{\Delta}\right), Z\left(\tfrac{j_2}{\Delta}\right)|D_n^T\right) - \max\left(g_n\left(\tfrac{j_1}{\Delta}\right), g_n\left(\tfrac{j_2}{\Delta}\right)\right)\right| \leq \kappa_6(1+x)^2[y+\log n]^2 n^{-u_5}$$

$$\max_{(j_1,j_2)\in\{a_x\Delta,\ldots,0\}\times\{0,\ldots,b_x\Delta\}}\left|\mathrm{Cov}\left(Z\left(\tfrac{j_1}{\Delta}\right), Z\left(\tfrac{j_2}{\Delta}\right)|D_n^T\right)\right| \leq \kappa_6(1+x)^2[y+\log n]^2 n^{-u_5}.$$

We introduce the following definition.

Definition 12 *Let $(W_t)_{t\in\mathbb{R}}$ be a two-sided Wiener process such that $W_0 = 0$. For any function $g : I \to \mathbb{R}$, where I is an interval containing 0, let $K(g) : I^2 \to \mathbb{R}$ be defined for any $(s,t) \in I^2$ by*

$$K(g)(s,t) = \begin{cases} g(s\wedge t) \text{ if } (s,t)\in(I\cap\mathbb{R}_+)^2 \\ -g(s\vee t) \text{ if } (s,t)\in(I\cap\mathbb{R}_-)^2 \\ 0 \text{ else .} \end{cases} \tag{25}$$

For all $j \in \mathbb{Z}\cap[a_x\Delta; b_x\Delta]$, by definition 10 of Z:

$$Z\left(\frac{j}{\Delta}\right) = \frac{2}{\mathfrak{e}}\left(P_n^{T^c} - P\right)\left(\hat{s}_{k_*+j}^T - \hat{s}_{k_*}^T\right) = \begin{cases} 0 \text{ if } j=0, \\ \frac{2}{\mathfrak{e}}\left(P_n^{T^c}-P\right)\sum_{i=k_*+1}^{k_*+j}\hat{\theta}_i^T\psi_i \text{ if } j>0, \\ \frac{2}{\mathfrak{e}}\left(P_n^{T^c}-P\right)\sum_{i=k_*+j+1}^{k_*}\hat{\theta}_i^T\psi_i \text{ if } j<0 \end{cases} \tag{26}$$

In other words, for all $j \in \mathbb{Z} \cap [a_x\Delta; b_x\Delta]$,

$$Z\left(\frac{j}{\Delta}\right) = \mathrm{sgn}(j)\frac{2}{\mathfrak{e}} \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \hat{\theta}_i^T \left(P_n^{T^c} - P\right)(\psi_i).$$

Let $n_v = |T^c| = n - n_t$. Thus, for any $(j_1, j_2) \in \{a_x\Delta, \ldots, b_x\Delta\}^2$ and any variable X with distribution $s(x)dx$,

$$\begin{aligned}
&\mathrm{Cov}\left(Z\left(\tfrac{j_1}{\Delta}\right), Z\left(\tfrac{j_2}{\Delta}\right)|D_n^T\right) \\
&\quad = \mathrm{sgn}(j_1)\mathrm{sgn}(j_2)\frac{4}{n_v\mathfrak{e}^2} \sum_{i_1=k_*-(j_1)_-+1}^{k_*+(j_1)_+} \sum_{i_2=k_*-(j_2)_-+1}^{k_*+(j_2)_+} \hat{\theta}_{i_1}^T \hat{\theta}_{i_2}^T \,\mathrm{Cov}(\psi_{i_1}(X), \psi_{i_2}(X)).
\end{aligned}$$

Let us now introduce the following definition.

Definition 13 *Let $(m_1, m_2, m_3) \in \mathbb{N}^3$ be three integers. Let $m_{(1)} \leq m_{(2)} \leq m_{(3)}$ be their non-decreasing rearrangement. Define $E_{m_1,m_2,m_3} = 0$ if $m_{(1)} = m_{(2)}$ or $m_{(2)} = m_{(3)}$ and*

$$E_{m_1,m_2,m_3} = \sum_{j_1=m_{(1)}+1}^{m_{(2)}} \sum_{j_2=m_{(2)}+1}^{m_{(3)}} \hat{\theta}_{j_1}^T \hat{\theta}_{j_2}^T \,\mathrm{Cov}(\psi_{j_1}(X), \mathrm{Cov}(\psi_{j_2}(X)) \tag{27}$$

if $m_{(1)} < m_{(2)} < m_{(3)}$.

Let $n_v = n - n_t$. The covariance can be broken down as follows: If $0 < j_1 \leq j_2$, conditionally on D_n^T.

$$\begin{aligned}
\mathrm{Cov}\left(Z\left(\frac{j_1}{\Delta}\right), Z\left(\frac{j_2}{\Delta}\right)|D_n^T\right) &= \mathrm{Var}\left(Z\left(\frac{j_1}{\Delta}\right)\right) \\
&\quad + \frac{4}{n_v\mathfrak{e}^2} \sum_{i_1=k_*+1}^{k_*+j_1} \sum_{i_2=k_*+j_1+1}^{k_*+j_2} \hat{\theta}_{i_1}^T \hat{\theta}_{i_2}^T \,\mathrm{Cov}(\psi_{i_1}(X), \psi_{i_2}(X)).
\end{aligned}$$

If $j_1 \leq j_2 < 0$, symetrically,

$$\begin{aligned}
\mathrm{Cov}\left(Z\left(\frac{j_1}{\Delta}\right), Z\left(\frac{j_2}{\Delta}\right)|D_n^T\right) &= \mathrm{Var}\left(Z\left(\frac{j_2}{\Delta}\right)\right) \\
&\quad + \frac{4}{n_v\mathfrak{e}^2} \sum_{i_2=k_*+j_2+1}^{k_*} \sum_{i_1=k_*+j_1+1}^{k_*+j_2} \hat{\theta}_{i_1}^T \hat{\theta}_{i_2}^T \,\mathrm{Cov}(\psi_{i_1}(X), \psi_{i_2}(X)).
\end{aligned}$$

Finally, if $j_1 < 0 < j_2$,

$$\mathrm{Cov}\left(Z\left(\frac{j_1}{\Delta}\right), Z\left(\frac{j_2}{\Delta}\right)|D_n^T\right) = \frac{-4}{n_v\mathfrak{e}^2} \sum_{i_1=k_*+j_1+1}^{k_*} \sum_{i_2=k_*+1}^{k_*+j_2} \hat{\theta}_{i_1}^T \hat{\theta}_{i_2}^T \,\mathrm{Cov}(\psi_{i_1}(X), \psi_{i_2}(X)).$$

It follows from the previous equations that for any $(k_1, k_2) \in \mathbb{N}^2$,

$$\mathrm{Cov}\left(Z\left(\frac{k_1-k_*}{\Delta}\right), Z\left(\frac{k_2-k_*}{\Delta}\right)\right) = \begin{cases} \mathrm{Var}(Z\left(\frac{k_1-k_*}{\Delta}\right)) + 4\frac{E_{k_*,k_1,k_2}}{n_v\mathfrak{e}^2} & \text{if } k_* < k_1 \leq k_2 \\ 4\frac{E_{k_*,k_1,k_2}}{n_v\mathfrak{e}^2} & \text{if } k_1 < k_* < k_2 \\ \mathrm{Var}(Z\left(\frac{k_2-k_*}{\Delta}\right)) + 4\frac{E_{k_*,k_1,k_2}}{n_v\mathfrak{e}^2} & \text{if } k_1 \leq k_2 < k_* \end{cases} . \tag{28}$$

Let $I, J \subset \{a_x\Delta, \dots, b_x\Delta\}$. Assuming concentration around the expectation yields

$$\sum_{i\in I}\sum_{j\in J}\hat{\theta}_i^T\hat{\theta}_j^T \operatorname{Cov}(\psi_i(X),\psi_j(X)) \sim \sum_{i\in I}\sum_{j\in J}\theta_i\theta_j \operatorname{Cov}(\psi_i(X),\psi_j(X)) + \frac{1}{n_t}\sum_{i\in I}\sum_{j\in J}\operatorname{Cov}(\psi_i(X),\psi_j(X))^2.$$

Moreover, for any $(i_1, i_2) \in \mathbb{N}^2$,

$$\psi_{i_1}(X)\psi_{i_2}(X) = 2\cos(2i_1\pi X)\cos(2i_2\pi X) = \cos\left(2(i_1+i_2)\pi X\right)+\cos(2(i_1-i_2)\pi X)$$

and by definition, for all $i \in \mathbb{N}^*$, $\psi_i = \sqrt{2}\cos(2i\pi X)$, while $\psi_0 = 1 = \cos(0\pi x)$. As a result, $\psi_{i_1}(X)\psi_{i_2}(X) = \frac{\psi_{i_1+i_2}(X)+\psi_{|i_1-i_2|}(X)}{\sqrt{2}}$ if $i_1 \neq i_2$ and

$$\operatorname{Cov}(\psi_{i_1}(X),\psi_{i_2}(X)) = \frac{\theta_{i_1+i_2}}{\sqrt{2}} + \left(\frac{1-\delta_{i_1,i_2}}{\sqrt{2}} + \delta_{i_1,i_2}\right)\theta_{|i_2-i_1|} - \theta_{i_1}\theta_{i_2}.$$

By assumption, the sequence $|\theta_k|$ tends to 0 with a polynomial rate of convergence, hence for sequences $i_1 \sim i_2$ tending to $+\infty$, $\theta_{|i_1-i_2|}$ dominates $\theta_{i_1}\theta_{i_2}$ and $\theta_{i_1+i_2}$. Heuristically, it can thus be expected that

$$\sum_{i\in I}\sum_{j\in J}\hat{\theta}_i^T\hat{\theta}_j^T \operatorname{Cov}(\psi_i(X),\psi_j(X)) \sim \sum_{i\in I}\sum_{j\in J}\theta_i\theta_j\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|j-i|} + \sum_{i\in I}\sum_{j\in J}\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)^2\theta_{|j-i|}^2.$$

This leads to the following proposition, the rigourous proof of which can be found in the appendix (proposition 8.6).

Proposition 7.3 *Let P be the probability measure with pdf s on $[0;1]$, let $\theta_j = \langle s, \psi_j\rangle = P(\psi_j)$ and assume that the coefficients θ_j satisfy the hypotheses of section 4. Let $\hat{\theta}_j^T = P^T(\psi_j)$. Let $I_k^1, I_k^2 \subset \{k_* + a_x\Delta, \dots, k_* + b_x\Delta\}$ be two intervals. Then the statistics*

$$U_{I_k^1,I_k^2} = \sum_{i\in I_k^1}\sum_{j\in I_k^2}\hat{\theta}_i^T\hat{\theta}_j^T[P(\psi_i\psi_j) - P\psi_i P\psi_j]$$

can be approximated in the following way: there exists two constants κ_4 and $u_3 > 0$ such that, with probability greater than $1 - e^{-y}$,

$$U_{I_k^1,I_k^2} = \frac{1}{2}\frac{|I_k^1 \cap I_k^2|}{n_t} + \left(1 - \frac{1}{\sqrt{2}}\right)\sum_{i\in I_k^1\cap I_k^2}\theta_i^2 + \frac{1}{\sqrt{2}}\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\theta_{|i-j|} + \frac{1}{2n_t}\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_{|i-j|}^2 \pm \kappa_4(y+\log n)^2(1+x)n^{-u_3}\mathcal{E}.$$

It is now possible to show that the terms E_{k_*,k_1,k_2} which appear in equation (28) are negligible compared to $\mathcal{E}$. That is the point of the following claim.

Claim 6 *Under the assumptions of Theorem 1, there exists constants $\kappa_7 \geq 0$ and $u_4 > 0$ such that for all $n \in \mathbb{N}$, $x > 0$ and $(m_1, m_2, m_3) \in \{a_x\Delta, \ldots, b_x\Delta\}^3$ such that $m_1 < m_2 < m_3$,*

$$\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} \leq \kappa_7(1+x)^2 n^{-u_4}\mathcal{E} \tag{29}$$

$$\frac{1}{n_t} \sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta^2_{|j_1-j_2|} \leq \kappa_7(1+x)^2 n^{-u_4}\mathcal{E}. \tag{30}$$

and moreover, for all $x > 0$, with probability greater than $1 - e^{-y}$, for any integers $(m_1, m_2, m_3) \in \{a_x\Delta, \ldots, b_x\Delta\}^3$,

$$|E_{m_1,m_2,m_3}| \leq \kappa_7(1+x)^2(y + \log n)^2 n^{-u_4}\mathcal{E}. \tag{31}$$

Proof Assume without loss of generality that $m_1 < m_2 < m_3$. We start by proving equation (29). First, changing variables from j_1, j_2 to $i = j_1, r = j_2 - j_1$ yields

$$\begin{aligned}
\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} &= \sum_{r\in\mathbb{N}} \theta_r \sum_{i=m_2+1-r}^{m_2} \mathbb{I}_{i\geq m_1+1}\mathbb{I}_{i+r\leq m_3}\theta_i\theta_{i+r} \\
&\leq \frac{1}{2}\sum_{r\leq r_0} |\theta_r| \sum_{i=(m_2+1-r)\vee(m_1+1)}^{m_2\wedge(m_3-r)} \theta_i^2 + \theta_{i+r}^2 \\
&\quad + \frac{1}{2}\sum_{r>r_0} |\theta_r| \left[\sum_{j_1=m_1+1}^{m_2} \theta_{j_1}^2 + \sum_{j_2=m_2+1}^{m_3} \theta_{j_2}^2\right] \\
&\leq \frac{\|\theta\|_{\ell^1}}{2} \max_{1\leq r\leq r_0} \sum_{i=(m_2+1-r)\vee(m_1+1)}^{m_2\wedge(m_3-r)} \theta_i^2 + \theta_{i+r}^2 \\
&\quad + \frac{1}{2}\sum_{r>r_0} |\theta_r| \sum_{i=a_x\Delta+1}^{b_x\Delta} \theta_{k_*+i}^2.
\end{aligned}$$

By claim 7 in appendix, for any $k \in \{m_1+1, \ldots, m_3\} \subset \{a_x\Delta+1, \ldots, b_x\Delta\}$, $\theta_k^2 \leq \kappa_3(1+x)^2 n^{-u_2}\mathfrak{e}$, therefore

$$\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} \leq r_0\frac{\|\theta\|_{\ell^1}}{2}\kappa_3(1+x)^2 n^{-u_2}\mathfrak{e} + \frac{1}{2}\sum_{r>r_0} |\theta_r| \sum_{i=a_x\Delta+1}^{b_x\Delta} \theta_{k_*+i}^2.$$

By hypothesis 1 of section 4,

$$\sum_{j\geq r_0+1} |\theta_j| \leq \sum_{j=r_0+1}^{+\infty} \sqrt{\sum_{i=j}^{+\infty} \theta_i^2} \leq \sum_{j=r_0+1}^{+\infty} \frac{c_1}{(j-1)^{1+\frac{\delta_1}{2}}} \leq \frac{2c_1}{\delta_1} r_0^{\frac{-\delta_1}{2}}.$$

Thus, by lemma 7.1,

$$\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} \leq \frac{r_0\,\|\theta\|_{\ell^1}}{2}\kappa_3(1+x)^2 n^{-u_2}\mathcal{E} + \frac{2c_1}{\delta_1} r_0^{-\frac{\delta_1}{2}} 4(1+x)\mathcal{E}. \tag{32}$$

Let $r_0 = \lceil n^{\frac{2u_2}{2+\delta_1}} \rceil \leq 2n^{\frac{2u_2}{2+\delta_1}}$ and $u = \frac{\delta_1(u_2)}{2+\delta_1} > 0$. For all $n \geq 2$,

$$\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} \leq \left[\|\theta\|_{\ell^1} \kappa_3 + 8\frac{c_1}{\delta_1} \right] (1+x)^2 n^{-u} \mathcal{E}, \tag{33}$$

which proves equation (29).

Moreover,

$$\begin{aligned}
\sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta^2_{|j_1-j_2|} &= \sum_{r\in\mathbb{N}} \theta_r^2 \left|\{j_1 : (m_1+1 \leq j_1 \leq m_2) \wedge (m_2+1 \leq j_1+r \leq m_3)\}\right| \\
&\leq \sum_{r\in\mathbb{N}} \theta_r^2 \left[(m_3-m_1)\wedge r\right] \\
&\leq r_0 \sum_{r=0}^{r_0} \theta_r^2 + (m_3-m_1) \sum_{r>r_0} \theta_r^2 \\
&\leq r_0 \|s\|^2 + (b_x-a_x)\Delta \sum_{r>r_0} \theta_r^2 \\
&\leq r_0 \|s\|^2 + 2(1+x)\Delta \frac{c_1}{r_0^{2+\delta_1}},
\end{aligned}$$

by hypothesis 1 of Theorem 1 and lemma 7.1. Let now $r_0 = \lceil \Delta^{\frac{1}{3}} \rceil$. Since $\Delta \geq 1$, it follows that:

$$\begin{aligned}
\frac{1}{n_t} \sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta^2_{|j_1-j_2|} &\leq \frac{\Delta^{\frac{1}{3}}+1}{n_t} \|s\|^2 + 2c_1(1+x)\frac{\Delta}{n_t} (\Delta)^{\frac{-2}{3}} \\
&\leq \left[2\frac{\|s\|^2}{(\Delta)^{\frac{2}{3}}} \mathcal{E} + 2c_1(1+x)\mathcal{E}(\Delta)^{-\frac{2}{3}} \right] \\
&\leq \left[2\|s\|^2 + 2c_1(1+x) \right] \frac{\mathcal{E}}{(\Delta)^{\frac{2}{3}}}.
\end{aligned}$$

On the other hand, $\Delta \geq \frac{n_t}{n-n_t} \geq n^{\delta_4}$ by hypothesis 4 of Theorem 1. There exists therefore $\kappa(c_1, \|s\|^2)$ such that, for any n,

$$\frac{1}{n_t} \sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta^2_{|j_1-j_2|} \leq \kappa(1+x) n^{-\frac{2\delta_4}{3}} \mathcal{E}, \tag{34}$$

which proves equation (30). Since for all $x > 0$,

$$(b_x - a_x)\Delta \leq 2(1+x)\Delta \leq \kappa(1+x) \left[2n_t or(n_t) + \frac{n_t}{n-n_t} \right] \leq \kappa(1+x) \left[2\|s\|^2 n_t + n_t^{\frac{1}{3}} \right],$$

by proposition 7.3 and a union bound, there exists an event A of probability greater than $1-e^{-y}$ and a constant κ such that, if $a_x \leq m_1 < m_2 < m_3 \leq b_x$, then

$$\begin{aligned}
|E_{m_1,m_2,m_3}| \leq \frac{1}{\sqrt{2}} \sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta_{j_1}\theta_{j_2}\theta_{|j_1-j_2|} + \frac{1}{2n_t} \sum_{j_1=m_1+1}^{m_2} \sum_{j_2=m_2+1}^{m_3} \theta^2_{|j_1-j_2|} \\
+ \kappa(y + \log(2+x) + \log n)^2 (1+x) n^{-u_3} \mathcal{E}.
\end{aligned} \tag{35}$$

From equations (35), (33) and (34), equation (31) follows with $u_4 = \min\left(u_3, \frac{2\delta_4}{3}, \frac{\delta_1 u_2}{2+\delta_1}\right)$.
■

Let then $g_n^0 : \left[\frac{-k_*}{\Delta}; +\infty\right[\to \mathbb{R}$ be defined first for all $\alpha \in \left\{\frac{j}{\Delta} : j \in \mathbb{N} - k_*\right\}$ by

$$\forall \alpha \in \left\{\frac{j}{\Delta} : (j+k_*) \in \mathbb{N}\right\}, g_n^0(\alpha) = \operatorname{sgn}(\alpha)\operatorname{Var}(Z_\alpha), \tag{36}$$

then for all $\alpha \in \left[\frac{-k_*}{\Delta}; +\infty\right[$ by linear interpolation (hence in general, $g_n^0(\alpha) \neq \operatorname{Var}(Z_\alpha)$). Let $K(g_n^0)$ be given by definition 12, then by equation (28) and claim 6, with probability greater than $1 - e^{-y}$, for any $x > 0$,

$$\begin{aligned}
&\max_{(j_1,j_2)\in\{a_x\Delta,\dots,b_x\Delta\}^2} \left|\operatorname{Cov}\left(Z\left(\frac{j_1}{\Delta}\right), Z\left(\frac{j_2}{\Delta}\right)\right) - K\left(g_n^0, \frac{j_1}{\Delta}, \frac{j_2}{\Delta}\right)\right| \\
&\quad \leq 4\kappa_7(1+x)^2(y+\log n)^2 n^{-u_4}\frac{\mathcal{E}}{n_v\mathfrak{e}^2} \\
&\quad \leq 4\kappa_7(1+x)^2(y+\log n)^2 n^{-u_4}.
\end{aligned} \tag{37}$$

Moreover, for any $j \in \mathbb{Z} \cap [a_x\Delta; b_x\Delta]$, by definition of Z,

$$\begin{aligned}
\operatorname{sgn}(j) g_n^0\left(\frac{j}{\Delta}\right) = \operatorname{Var}\left(Z\left(\frac{j}{\Delta}\right)\right) = \frac{4}{n_v\mathfrak{e}^2} \sum_{i_1=k_*-(j)_-+1}^{k_*+(j)_+} \sum_{i_2=k_*-(j)_-+1}^{k_*+(j)_+} \hat{\theta}_{i_1}^T \hat{\theta}_{i_2}^T \Big[\frac{\theta_{i_1+i_2}}{\sqrt{2}} \\
+ \left(\frac{1-\delta_{i_1,i_2}}{\sqrt{2}} + \delta_{i_1,i_2}\right)\theta_{|i_1-i_2|} - \theta_{i_1}\theta_{i_2}\Big].
\end{aligned} \tag{38}$$

Moreover, since $\mathfrak{e}^2 = \frac{\mathcal{E}}{n_v}$,

$$\begin{aligned}
\frac{4}{n_v\mathfrak{e}^2}\frac{1}{2}\frac{j}{n_t} &= \frac{4}{n_v\mathfrak{e}^2}\frac{1}{2}\frac{j}{\Delta}\frac{\Delta}{n_t} \\
&= \frac{\mathcal{E}}{n_v\mathfrak{e}^2} 2\frac{j}{\Delta} \\
&= 2\frac{j}{\Delta} \ .
\end{aligned}$$

Thus, by proposition 7.3, with probability greater than $1 - e^{-y}$,

$$\begin{aligned}
\operatorname{sgn}(j) g_n^0\left(\frac{j}{\Delta}\right) &= 2\frac{|j|}{\Delta} + \frac{4}{n_v\mathfrak{e}^2}\left(1 - \frac{1}{\sqrt{2}}\right) \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \theta_i^2 \\
&+ \frac{4}{n_v\mathfrak{e}^2} \sum_{i_1=k_*-(j)_-+1}^{k_*+(j)_+} \sum_{i_2=k_*-(j)_-+1}^{k_*+(j)_+} \theta_{i_1}\theta_{i_2}\frac{\theta_{|i_1-i_2|}}{\sqrt{2}} \\
&+ \frac{4}{n_v\mathfrak{e}^2}\frac{1}{2n_t} \sum_{i_1=k_*-(j)_-+1}^{k_*+(j)_+} \sum_{i_2=k_*-(j)_-+1}^{k_*+(j)_+} \theta_{|i_1-i_2|}^2 \\
&\pm 4\kappa_4(y+\log n)^2(1+x)n^{-u_3}.
\end{aligned} \tag{39}$$

Let g_n^1 be defined for all $\alpha = \frac{j}{\Delta}, j \in \mathbb{Z} \cap [-k_*(n_t); +\infty)$ by

$$\begin{aligned}
\operatorname{sgn}(j) g_n^1\left(\frac{j}{\Delta}\right) &= \frac{4}{n_v \mathfrak{e}^2} \sum_{i_1=k_*-(j)_-+1}^{k_*+(j)_+} \sum_{i_2=k_*-(j)_-+1}^{k_*+(j)_+} \theta_{i_1}\theta_{i_2} \frac{\theta_{|i_1-i_2|}}{\sqrt{2}} \\
&\quad + \frac{4}{n_v \mathfrak{e}^2}\left(1-\frac{1}{\sqrt{2}}\right) \sum_{i=k_*-(j)_-+1}^{k_*+(j)_+} \theta_i^2 \\
&= \frac{4}{n_v \mathfrak{e}^2} \sum_{i_1=k_*-(j)_-+1}^{k_*+(j)_+} \sum_{i_2=k_*-(j)_-+1}^{k_*+(j)_+} \theta_{i_1}\theta_{i_2} \left(\frac{1-\delta_{i_1,i_2}}{\sqrt{2}} + \delta_{i_1,i_2}\right) \theta_{|i_1-i_2|},
\end{aligned} \tag{40}$$

and for all $\alpha \in \left[-\frac{k_*}{\Delta}; +\infty\right[$ by linear interpolation.

We will now apply lemma 8.8 to g_n^1. Let $x > 0$ and $(k_1, k_2) \in \{k_* + a_x\Delta, \ldots, k_* + b_x\Delta\}^2$ be such that $k_1 < k_2$. Thus:

- If $k_* \leq k_1$,

$$\begin{aligned}
g_n^1\left(\frac{k_2-k_*}{\Delta}\right) - g_n^1\left(\frac{k_1-k_*}{\Delta}\right) &= \frac{4}{n_v \mathfrak{e}^2}\Bigg(\sum_{i=k_*+1}^{k_2} \sum_{j=k_*+1}^{k_2} \theta_i\theta_j \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} \\
&\quad - \sum_{i=k_*+1}^{k_1} \sum_{j=k_*+1}^{k_1} \theta_i\theta_j \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|}\Bigg) \\
&= \frac{4}{n_v \mathfrak{e}^2}\Bigg(\sum_{i=k_1+1}^{k_2} \sum_{j=k_1+1}^{k_2} \theta_i\theta_j \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} \\
&\quad + 2\sum_{i=k_1+1}^{k_2} \sum_{j=k_*+1}^{k_1} \theta_i\theta_j \frac{\theta_{|i-j|}}{\sqrt{2}}\Bigg).
\end{aligned}$$

By lemma 8.7 in appendix,

$$0 \leq \sum_{i=k_1+1}^{k_2} \sum_{j=k_1+1}^{k_2} \theta_i\theta_j \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} \leq \|s\|_\infty \sum_{i=k_1+1}^{k_2} \theta_i^2.$$

Thus by equation (29) from claim 6,

$$-\kappa_7(1+x)^2 n^{-u_3}\frac{4\sqrt{2}}{n_v \mathfrak{e}^2}\mathcal{E} \leq g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right) - g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right) \leq \frac{4\|s\|_\infty}{n_v \mathfrak{e}^2} \sum_{i=k_1+1}^{k_2} \theta_i^2 + \kappa_7(1+x)^2 n^{-u_4}\frac{4\sqrt{2}}{n_v \mathfrak{e}^2}\mathcal{E}.$$

- If $k_2 \leq k_*$,

$$g_n^1\left(\frac{k_2-k_*}{\Delta}\right)-g_n^1\left(\frac{k_1-k_*}{\Delta}\right)=\frac{4}{n_v\mathfrak{e}^2}\Bigg(\sum_{i=k_1+1}^{k_*}\sum_{j=k_1+1}^{k_*}\theta_i\theta_j\left(\frac{1-\delta_{i,j}}{\sqrt{2}}+\delta_{i,j}\right)\theta_{|i-j|}$$
$$-\sum_{i=k_2+1}^{k_*}\sum_{j=k_2+1}^{k_*}\theta_i\theta_j\left(\frac{1-\delta_{i,j}}{\sqrt{2}}+\delta_{i,j}\right)\theta_{|i-j|}\Bigg)$$
$$=\frac{4}{n_v\mathfrak{e}^2}\Bigg(\sum_{i=k_1+1}^{k_2}\sum_{j=k_1+1}^{k_2}\theta_i\theta_j\left(\frac{1-\delta_{i,j}}{\sqrt{2}}+\delta_{i,j}\right)\theta_{|i-j|}$$
$$+2\sum_{i=k_1+1}^{k_2}\sum_{j=k_2+1}^{k_*}\theta_i\theta_j\frac{\theta_{|i-j|}}{\sqrt{2}}\Bigg).$$

In the same way, by lemma 8.7 and equation (29) from claim 6,

$$g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right)-g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right)\geq-\kappa_7(1+x)^2n^{-u_3}\frac{4\sqrt{2}}{n_v\mathfrak{e}^2}\mathcal{E},$$
$$g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right)-g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right)\leq\frac{4\left\|s\right\|_\infty}{n_v\mathfrak{e}^2}\sum_{i=k_1+1}^{k_2}\theta_i^2+\kappa_7(1+x)^2n^{-u_4}\frac{4\sqrt{2}}{n_v\mathfrak{e}^2}\mathcal{E}.$$

- If $k_1 \leq k_2 \leq k_2$,

$$g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right)-g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right)=g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right)-g_n(0)+g_n(0)-g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right),$$

therefore by the two previous cases,

$$0\leq g_n^1\left(\tfrac{k_2-k_*}{\Delta}\right)-g_n^1\left(\tfrac{k_1-k_*}{\Delta}\right)\leq\frac{4\left\|s\right\|_\infty}{n_v\mathfrak{e}^2}\sum_{i=k_1+1}^{k_2}\theta_i^2+\kappa_7(1+x)^2n^{-u_4}\frac{8\sqrt{2}}{n_v\mathfrak{e}^2}\mathcal{E}.$$

By definition $\mathfrak{e}^2=\frac{\mathcal{E}}{n_v}$ therefore for any $x>0$ and $(j_1,j_2)\in[a_x\Delta;b_x\Delta]^2$ such that $j_1\leq j_2$,

$$-4\sqrt{2}\kappa_7(1+x)^2n^{-u_4}\leq g_n^1\left(\tfrac{j_2}{\Delta}\right)-g_n^1\left(\tfrac{j_1}{\Delta}\right)\leq\frac{4\left\|s\right\|_\infty}{n_v\mathfrak{e}^2}\sum_{i=j_1+1}^{j_2}\theta_{k_*+i}^2+8\sqrt{2}\kappa_7(1+x)^2n^{-u_4}. \tag{41}$$

Moreover, for any j_1,j_2 such that $0<j_1<j_2$, $\theta_{k_*+j_i}^2\leq\frac{1}{n_t}$ hence

$$\left\|s\right\|_\infty\sum_{i=j_1+1}^{j_2}\theta_{k_*+i}^2\leq\left\|s\right\|_\infty\sum_{j=k_*+j_1+1}^{k_*+j_2}[\theta_j^2-\frac{1}{n_t}]+\frac{j_2-j_1}{\Delta}\left\|s\right\|_\infty\frac{\Delta}{n_t}$$
$$=-\left\|s\right\|_\infty\mathfrak{e}[f_n\left(\frac{j_2}{\Delta}\right)-f_n\left(\frac{j_1}{\Delta}\right)]+\frac{j_2-j_1}{\Delta}\left\|s\right\|_\infty\mathcal{E}. \tag{42}$$

For j_1,j_2 such that $j_1<j_2\leq0$, $\theta_{k_*+j_i}^2\geq\frac{1}{n_t}$ hence

$$\left\|s\right\|_\infty\sum_{i=j_1+1}^{j_2}\theta_{k_*+i}^2\leq\left\|s\right\|_\infty\sum_{j=k_*+j_1+1}^{k_*+j_2}[\theta_j^2-\frac{1}{n_t}]+\frac{j_2-j_1}{\Delta}\left\|s\right\|_\infty\frac{\Delta}{n_t}$$
$$=\mathfrak{e}\left\|s\right\|_\infty[f_n\left(\frac{j_1}{\Delta}\right)-f_n\left(\frac{j_2}{\Delta}\right)]+\frac{j_2-j_1}{\Delta}\left\|s\right\|_\infty\mathcal{E}. \tag{43}$$

By equations (41), (42) and (42), it follows that, for any $(j_1, j_2) \in \big([a_x\Delta; b_x\Delta] \cap \mathbb{Z}\big)^2$,

$$\begin{aligned} g_n^1\left(\frac{j_2}{\Delta}\right) - g_n^1\left(\frac{j_1}{\Delta}\right) &\leq -4\frac{\|s\|_\infty}{n_v\mathfrak{e}}\left[f_n\left(\frac{j_2}{\Delta}\right) - f_n\left(\frac{j_1}{\Delta}\right)\right] + 4\|s\|_\infty \frac{j_2 - j_1}{\Delta} + 8\sqrt{2}\kappa_7(1+x)^2 n^{-u_4} \\ &\leq -4\frac{\|s\|_\infty}{n_v\mathfrak{e}}\left[f_n\left(\frac{j_2}{\Delta}\right) - f_n\left(\frac{j_1}{\Delta}\right)\right] + 4\|s\|_\infty \frac{j_2 - j_1}{\Delta} + 8\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}. \end{aligned} \tag{44}$$

To extend the lower bound given by equation (41), notice that for any $(\alpha_1, \alpha_2) \in [a_x; b_x]^2$ such that $\alpha_1 < \alpha_2$,

- if $\lfloor \alpha_1\Delta \rfloor = \lfloor \alpha_2\Delta \rfloor \leq \alpha_1\Delta < \alpha_2\Delta \leq \lfloor \alpha_1\Delta \rfloor + 1$, by linearity of g_n^1 on $[\lfloor \alpha_1\Delta \rfloor; \lfloor \alpha_1\Delta \rfloor + 1]$,

$$g_n^1(\alpha_2) - g_n^1(\alpha_1) \geq -\left[g_n^1\left(\frac{\lfloor \alpha_1\Delta \rfloor + 1}{\Delta}\right) - g_n^1\left(\frac{\lfloor \alpha_1\Delta \rfloor}{\Delta}\right)\right]_-,$$

- otherwise, $\lfloor \alpha_1\Delta \rfloor + 1 \leq \lfloor \alpha_2\Delta \rfloor$, therefore by linearity of $(u, v) \mapsto g_n^1(u) - g_n^1(v)$ on $\frac{1}{\Delta}[\lfloor \alpha_1\Delta \rfloor; \lfloor \alpha_1\Delta \rfloor + 1] \times \frac{1}{\Delta}[\lfloor \alpha_2\Delta \rfloor; \lfloor \alpha_2\Delta \rfloor + 1]$,

$$\begin{aligned} g_n^1(\alpha_2) - g_n^1(\alpha_1) \geq \min\Big\{ & g_n^1(u) - g_n^1(v) : \\ & (u, v) \in \Big\{\frac{\lfloor \alpha_2\Delta \rfloor}{\Delta}; \frac{\lfloor \alpha_2\Delta \rfloor + 1}{\Delta}\Big\} \times \Big\{\frac{\lfloor \alpha_1\Delta \rfloor}{\Delta}; \frac{\lfloor \alpha_1\Delta \rfloor + 1}{\Delta}\Big\}\Big\}. \end{aligned}$$

In all cases,

$$g_n^1(\alpha_2) - g_n^1(\alpha_1) \geq -\max\left\{\left[g_n^1\left(\frac{j_2}{\Delta}\right) - g_n^1\left(\frac{j_1}{\Delta}\right)\right]_- : j_1 \leq j_2, (j_1, j_2) \in \{a_x\Delta, \ldots, b_x\Delta\}^2\right\}. \tag{45}$$

Thus by equation (41), for any $x > 0$:

$$\forall (\alpha_1, \alpha_2) \in [a_x; b_x]^2, \alpha_1 < \alpha_2 \implies g_n^1(\alpha_2) - g_n^1(\alpha_1) \geq -4\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}. \tag{46}$$

By the same argument applied to the function

$$\alpha \mapsto g_n^1(\alpha) + 4\frac{\|s\|_\infty}{n_v\mathfrak{e}} f_n(\alpha) - 4\|s\|_\infty \alpha,$$

which is piecewise linear on the partition $\big\{\big[\frac{j}{\Delta}; \frac{j+1}{\Delta}\big[: j \in \{a_x\Delta, \ldots, b_x\Delta\}\big\}$, equation (44) extends to $[a_x; b_x]$ for any $x > 0$:

$$\begin{aligned} \forall (\alpha_1, \alpha_2) \in [a_x; b_x]^2, g_n^1(\alpha_2) - g_n^1(\alpha_1) \leq 4\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(\alpha_1) - f_n(\alpha_2)] + 4\|s\|_\infty [\alpha_2 - \alpha_1] \\ + 8\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}. \end{aligned} \tag{47}$$

Let $\varepsilon_d : \alpha \mapsto \inf_{x \in \mathbb{R}_+ : b_x \geq \alpha} 8\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}$. The function ε_d is non-decreasing by definition, and $\varepsilon_d(0) \geq 8\kappa_7 n^{-u_4} > 0$. Furthermore, by equations (46) and

(47),

$$\forall(\alpha_1,\alpha_2)\in\mathbb{R}_+^2, \alpha_1\le\alpha_2 \implies$$
$$-\varepsilon_d(\alpha_2)\le g_n^1(\alpha_2)-g_n^1(\alpha_1)\le 4\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(\alpha_1)-f_n(\alpha_2)]+4\|s\|_\infty[\alpha_2-\alpha_1]+\varepsilon_d(\alpha_2).$$

In this situation, lemma 8.8 applies with $g=g_n^1$, $h_+=-\frac{8\|s\|_\infty}{n_v\mathfrak{e}}f_n+8\|s\|_\infty\,\mathrm{Id}$ and $\varepsilon=2\varepsilon_d$. Note that h_+ is indeed non-decreasing since by equation (20), $f_n'\le\frac{\Delta}{n_t\mathfrak{e}}$ on $\mathbb{R}_+$ which leads to $\frac{f_n'}{n_v\mathfrak{e}}\le\frac{\Delta}{n_vn_t\mathfrak{e}^2}=1$. This guarantees existence of a non-decreasing function $g_{n,+}^2:\mathbb{R}_+\to\mathbb{R}_+$ such that $g_{n,+}^2(0)=0$,

$$\sup_{\alpha\in\mathbb{R}_+}\frac{|g_n^1(\alpha)-g_{n,+}^2(\alpha)|}{2\varepsilon_d(\alpha)}\le 6$$

and for all α_1,α_2 such that $\alpha_1\le\alpha_2$,

$$g_{n,+}^2(\alpha_2)-g_{n,+}^2(\alpha_1)\le 8\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(\alpha_1)-f_n(\alpha_2)]+8\|s\|_\infty[\alpha_2-\alpha_1]$$

Symetrically, let $\varepsilon_g:\alpha\mapsto\inf_{x\in\mathbb{R}_+:-a_x\ge\alpha}8\sqrt{2}\kappa_7(1+x)^2n^{-u_4}$, defined on $\left[0;\frac{k_*(n_t)}{\Delta}\right]$. ε_g is non-decreasing by definition. Furthermore, $\varepsilon_g(0)\ge 8\kappa_7n^{-u_4}>0$. By equations (46) and (47),

$$\begin{aligned}\forall(\alpha_1,\alpha_2)\in\mathbb{R}_+^2,\alpha_1\le\alpha_2 \implies -\varepsilon_g(\alpha_2)&\le g_n^1(-\alpha_1)-g_n^1(-\alpha_2)\\&\le 4\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(-\alpha_2)-f_n(-\alpha_1)]\\&\quad+4\|s\|_\infty[\alpha_2-\alpha_1]+\varepsilon_g(\alpha_2).\end{aligned}$$

In this situation, lemma 8.8 applies with $g=-g_n^1(-\cdot)$, $h_+=\frac{8\|s\|_\infty}{n_v\mathfrak{e}}f_n(-\cdot)+8\|s\|_\infty\,\mathrm{Id}$, $\varepsilon=2\varepsilon_g$. It guarantees existence of a function $g_{n,-}^2:\left[0;\frac{k_*(n_t)}{\Delta}\right]\to\mathbb{R}_+$ such that $g_{n,-}^2(0)=0$,

$$\sup_{\alpha\in\left[0;\frac{k_*(n_t)}{\Delta}\right]}\frac{|-g_n^1(-\alpha)-g_{n,-}^2(\alpha)|}{2\varepsilon_g(\alpha)}\le 6$$

and for any α_1,α_2 such that $\alpha_1\le\alpha_2$,

$$g_{n,-}^2(\alpha_2)-g_{n,-}^2(\alpha_1)\le 8\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(-\alpha_2)-f_n(-\alpha_1)]+8\|s\|_\infty[\alpha_2-\alpha_1].$$

Let then $g_n^2:\alpha\mapsto g_{n,+}^2(\alpha)\mathbb{I}_{\alpha\ge0}-g_{n,-}^2(-\alpha)\mathbb{I}_{\alpha<0}$ and $\varepsilon(\alpha)=\varepsilon_d(\alpha)\mathbb{I}_{\alpha\ge0}+\varepsilon_g(-\alpha)\mathbb{I}_{x<0}$, which yields

$$\left\|\frac{g_n^2-g_n^1}{2\varepsilon}\right\|_\infty\le 6 \tag{48}$$

and

$$\forall(\alpha_1,\alpha_2)\in\mathbb{R}^2, g_n^2(\alpha_2)-g_n^2(\alpha_1)\le 8\frac{\|s\|_\infty}{n_v\mathfrak{e}}[f_n(\alpha_1)-f_n(\alpha_2)]+8\|s\|_\infty[\alpha_2-\alpha_1]. \tag{49}$$

By definition of ε, for any $x > 0$ and any $\alpha \in [a_x, b_x]$, $\varepsilon(\alpha) \leq 8\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}$, hence

$$\forall x > 0, \forall \alpha \in [a_x; b_x], |g_n^2(\alpha) - g_n^1(\alpha)| \leq 96\sqrt{2}\kappa_7(1+x)^2 n^{-u_4}. \tag{50}$$

Let then:

$$g_n : \alpha \mapsto g_n^2(\alpha) + 4\|s\|^2 \alpha. \tag{51}$$

Since g_n^2 is non-decreasing, $g_n(\alpha_2) - g_n(\alpha_1) \geq 4\|s\|^2 [\alpha_2 - \alpha_1]$, which proves equation 2 of Theorem 1. Moreover, equation (49) yields equation (14) of Theorem 1.

Let now $x > 0$ be fixed until the end of this section. By definition of g_n^1 (equation (40)), equations (39), (50) and since the functions g_n^0 and $\alpha \mapsto 4\|s\|^2 \alpha$ are piecewise linear on the partition $\left(\left[\frac{j}{\Delta}, \frac{j+1}{\Delta}\right)\right)_{a_x\Delta \leq j \leq b_x\Delta - 1}$, with probability greater than $1 - e^{-y}$,

$$\begin{aligned}
\left\|g_n^0 - g_n\right\|_\infty &\leq \left\|g_n^1 - g_n^2\right\|_\infty + \kappa_4(y + \log n)^2(1+x)n^{-u_3} \\
&\quad + \max_{a_x\Delta \leq j \leq b_x\Delta} \left| \frac{\operatorname{sgn}(j)}{n_v \mathfrak{e}^2} \frac{4}{2n_t} \sum_{i_1 = k_* + (j)_- + 1}^{k_* + (j)_+} \sum_{i_2 = k_* + (j)_- + 1}^{k_* + (j)_+} \theta^2_{|i_1 - i_2|} - \frac{(4\|s\|^2 - 2)j}{\Delta} \right| \\
&\leq \max_{a_x\Delta \leq j \leq b_x\Delta} \left| \frac{4}{n_v \mathfrak{e}^2} \frac{1}{2n_t} \sum_{i_1 = k_* + (j)_- + 1}^{k_* + (j)_+} \sum_{i_2 = k_* + (j)_- + 1}^{k_* + (j)_+} \theta^2_{|i_1 - i_2|} - 4\left(\|s\|^2 - \tfrac{1}{2}\right) \frac{|j|}{\Delta} \right| \\
&\quad + 96\sqrt{2}\kappa_7(1+x)^2 n^{-u_4} + \kappa_4(y + \log n)^2(1+x)n^{-u_3}.
\end{aligned} \tag{52}$$

It remains to bound the max. By parity in j of the sum, one can assume $0 \leq j \leq \max(|a_x|, |b_x|)\Delta$ instead of $a_x\Delta \leq j \leq b_x\Delta|]$. Let therefore $j \in \{0, \ldots, \max(|a_x|, |b_x|)\Delta\}$, then

$$\begin{aligned}
\frac{1}{2n_t} \sum_{i_1 = k_* + 1}^{k_* + j} \sum_{i_2 = k_* + 1}^{k_* + j} \theta^2_{|i_1 - i_2|} &= \frac{|j|}{2n_t} + \frac{1}{2n_t} \sum_{r \in \mathbb{N}^*} 2\left|\{i : k_* \leq i \leq i + r \leq k_* + j\}\right| \theta_r^2 \\
&= \frac{|j|}{2n_t} + \frac{1}{n_t} \sum_{r=1}^{+\infty} (j - r)_+ \theta_r^2.
\end{aligned} \tag{53}$$

Furthermore, for all $r_0 \in \mathbb{N}^*$,

$$\begin{aligned}
\left| \frac{1}{n_t} \sum_{r=1}^{+\infty} (j - r)_+ \theta_r^2 - \frac{j}{n_t}(\|s\|^2 - 1) \right| &\leq \frac{1}{n_t} \sum_{r=1}^{+\infty} \theta_r^2 \left|(j - r)_+ - j\right| \\
&\leq \frac{r_0}{n_t} \sum_{r=1}^{r_0} \theta_r^2 + \frac{j}{n_t} \sum_{r=r_0+1}^{+\infty} \theta_r^2 \\
&\leq \|s\|^2 \frac{r_0}{n_t} + \max(|a_x|, |b_x|) \frac{\Delta}{n_t} \times \frac{c_1}{r_0^2},
\end{aligned}$$

by hypothesis 1 of section 4. By setting $r_0 = \lceil (\Delta)^{\frac{1}{3}} \rceil \leq 2(\Delta)^{\frac{1}{3}}$ (since $\Delta \geq 1$), it follows that

$$\begin{aligned}
\left| \frac{1}{n_t} \sum_{r=1}^{+\infty} (j - r)_+ \theta_r^2 - \frac{j}{n_t}(\|s\|^2 - 1) \right| &\leq \left[2\|s\|^2 + c_1 \max(|a_x|, |b_x|)\right] \frac{(\Delta)^{\frac{1}{3}}}{n_t} \\
&\leq \left[2\|s\|^2 + 2c_1(1+x)\right] (\Delta)^{-\frac{2}{3}} \mathcal{E} \text{ by lemma 7.1.}
\end{aligned}$$

Let $\kappa = \kappa(c_1, \|s\|)$. Since by hypothesis 4 of Theorem 1, $\Delta \geq \frac{n_t}{n-n_t} \geq n^{\delta_4}$,

$$\left| \frac{1}{n_t} \sum_{r=1}^{+\infty} (j-r)_+ \theta_r^2 - \frac{j}{n_t}(\|s\|^2 - 1) \right| \leq \kappa(1+x) n^{-\frac{2\delta_4}{3}} \mathcal{E}.$$

By equation (53) and since $\frac{j}{n_t} = \frac{j}{\Delta}\mathcal{E}$, for any $j \in [0; \max(|a_x|, |b_x|)\Delta]$.

$$\left| \frac{1}{2n_t} \sum_{i_1=k_*+1}^{k_*+j} \sum_{i_2=k_*+1}^{k_*+j} \theta_{|i_1-i_2|}^2 - \frac{j}{\Delta}\left(\|s\|^2 - \frac{1}{2}\right)\mathcal{E} \right| \leq \kappa(1+x) n^{-\delta_4} \mathcal{E}.$$

By equation (52) and since $\frac{\mathcal{E}}{n_v \mathfrak{e}^2} = 1$, it follows that, with probability greater than $1 - e^{-y}$,

$$\left\| g_n^0 - g_n \right\|_\infty \leq 96\sqrt{2}\kappa_7 (1+x)^2 n^{-u_4} + \kappa_4 (y + \log n)^2 (1+x) n^{-u_3} + 4\kappa(1+x) n^{-\delta_4}. \tag{54}$$

Let $\kappa = 96\sqrt{2}\kappa_7 + \kappa_4 + 4\kappa$ and $u_5 = \min(u_4, u_3, \delta_4)$, it then follows from definition 12 of K that with probability greater than $1 - e^{-y}$,

$$\left\| K(g_n^0) - K(g_n) \right\|_\infty \leq \left\| g_n^0 - g_n \right\|_\infty \leq \kappa(1+x)^2 (y + \log n)^2 n^{-u_5}.$$

By equation (37), it follows that that, with probability greater than $1 - e^{-y}$, for any $(j_1, j_2) \in \{a_x\Delta, \ldots, b_x\Delta\}^2$,

$$\begin{aligned} \left| \mathrm{Cov}\left(Z\left(\frac{j_1}{\Delta}\right), Z\left(\frac{j_2}{\Delta}\right)\right) - K(g_n)\left(\frac{j_1}{\Delta}, \frac{j_2}{\Delta}\right) \right| &\leq 4\kappa_7 (1+x)^2 (y + \log n)^2 n^{-u_3} \\ &\quad + \kappa(1+x)^2 (y + \log n)^2 n^{-u_5} \\ &\leq \kappa(1+x)^2 (y + \log n)^2 n^{-u_5} \end{aligned} \tag{55}$$

by setting $\kappa = \kappa + 4\kappa_7$ and since $u_5 \leq u_3$. Claim 5 follows. It remains to upper bound g_n on $[a_x; b_x]$ in order to check equation 15 of Theorem 1. This is the subject of the following lemma.

Lemme 7.4 *For any $\alpha \in \mathbb{R}$,*

$$|g_n(\alpha)| \leq 20 \|s\|_\infty f_n(\alpha) + 12 \|s\|_\infty \leq \max\left(40 \|s\|_\infty f_n(\alpha), 24 \|s\|_\infty\right).$$

In particular, for all $x > 0$, $\max(|g_n(a_x)|, |g_n(b_x)|) \leq 20 \|s\|_\infty (1+x)$.

Proof Since $\|s\|_\infty \geq \|s\|^2 \geq 1$ and $n_v \mathfrak{e} \leq 1$ by lemma 3.1, point 3 of Theorem 1 which we already proved implies that for any $\alpha \in \mathbb{R}$,

$$|g_n(\alpha))| \leq 8 \|s\|_\infty f_n(\alpha) + 12 \|s\|_\infty |\alpha|.$$

If $|\alpha| < 1$, then

$$|g_n(\alpha))| \leq 8 \|s\|_\infty |f_n(\alpha)| + 12 \|s\|_\infty \leq \max\left(16 \|s\|_\infty f_n(\alpha), 24 \|s\|_\infty\right),$$

else $|f_n(\alpha) - f_n(1)| \geq |\alpha| - 1$, therefore $|\alpha| \leq f_n(\alpha) + 1$, which yields

$$|g_n(\alpha))| \leq 20 \|s\|_\infty f_n(\alpha) + 12 \|s\|_\infty \leq \max\left(40 \|s\|_\infty f_n(\alpha), 24 \|s\|_\infty\right).$$

■

7.5 Construction of a Wiener process W such that $W \circ g_n$ approximates Z

Let E_y be the event of probability greater than $1 - e^{-y}$ on which the equations of claim 5 are satisfied. Let $x > 0$. Given $D_n^T \in E_y$, Z^1 is a piecewise linear gaussian process on the partition $([\frac{j}{\Delta}; \frac{j+1}{\Delta}))_{a_x\Delta \leq j \leq b_x\Delta}$, such that for any $j \in \{a_x\Delta, \ldots, b_x\Delta\}$,

$$\max_{(j_1,j_2)\in\{0,\ldots,b_x\Delta\}^2} \left|\mathrm{Cov}\left(Z\left(\tfrac{j_1}{\Delta}\right), Z\left(\tfrac{j_2}{\Delta}\right)\right) - K(g_n)\left(\tfrac{j_1}{\Delta}, \tfrac{j_2}{\Delta}\right)\right| \leq \kappa_6(1+x)^2[y+\log n]^2 n^{-u_5}, \tag{56}$$

where $K(g_n)$ is given by definition 12. Since g_n is non-decreasing, $K(g_n)(s,t) = \mathrm{Cov}(W_{g_n(s)}, W_{g_n(t)})$ for any two-sided Wiener process W on $\mathbb{R}$ such that $W_0 = 0$. In particular, $K(g_n)$ is a positive-definite function. Furthermore, by definition, $\forall(\alpha_1, \alpha_2) \in [a_x; b_x]^2$,

$$K(g_n)(\alpha_1,\alpha_1) + K(g_n)(\alpha_2,\alpha_2) - 2K(g_n)(\alpha_1,\alpha_2) = |g_n(\alpha_2) - g_n(\alpha_1)|.$$

Moreover, for all $j \in \{a_x\Delta, \ldots, b_x\Delta - 1\}$, since $n_v\mathfrak{e} \leq 1$,

$$\begin{aligned}
|g_n(\tfrac{j+1}{\Delta}) - g_n(\tfrac{j}{\Delta})| &\leq 8\,\|s\|_\infty\,|f_n(\tfrac{j+1}{\Delta}) - f_n(\tfrac{j}{\Delta})| + \frac{12\,\|s\|_\infty}{\Delta} \text{ by equations (51) and (49)}\\
&\leq 8\kappa_3\,\|s\|_\infty\,(1+x)^2 n^{-u_2} + 12\,\|s\|_\infty\,\frac{n-n_t}{n_t} \text{ by claim 7}\\
&\leq 8\kappa_3\,\|s\|_\infty\,(1+x)^2 n^{-u_2} + 12\,\|s\|_\infty\,n^{-\delta_4} \text{ by hypothesis 4.}
\end{aligned}$$

Finally, by lemma 7.4 and since g_n is non-decreasing,

$$\sup_{\alpha\in[a_x;b_x]} K(g_n)(\alpha,\alpha) \leq \max(|g_n(a_x)|, |g_n(b_x)|) \leq 20\,\|s\|_\infty\,(1+x).$$

In this situation, proposition 8.9 in the appendix (applied to $Y = Z^1$, $K_X = K(g_n)$ with $h = g_n$) guarantees the existence of a continuous gaussian process $Z^2(D_n^T)$, with variance-covariance function $K(g_n)$ and such that for some constant κ and for $u = \min(u_5, u_2, \delta_4)$,

$$\forall y > 0, \forall D_n^T \in E_y, \mathbb{E}\left[\sup_{a_x\leq t\leq b_x} |Z^1(t) - Z^2(t)| | D_n^T\right] \leq \kappa(1+x)^{\frac{7}{6}}[y+\log n]^{\frac{2}{3}} \times n^{-\frac{u}{12}}. \tag{57}$$

Since the conditional distribution of $Z^2(D_n^T)$ given D_n^T is entirely determined by the function g_n which does not depend on D_n^T, Z^2 is independent from D_n^T. Moreover, since g_n increases, $W = Z^2 \circ g_n^{-1}$ is a continuous, centered gaussian process with covariance function

$$\mathrm{Cov}(Z_s, Z_t) = K(g_n)(g_n^{-1}(s), g_n^{-1}(t)) = \begin{cases} s \wedge t \text{ if } 0 \leq s,t \\ -(s \vee t) \text{ if } s,t \leq 0 \\ 0 \text{ else ,} \end{cases} \tag{58}$$

it is therefore a two-sided Wiener process on $[g_n(a_x); g_n(b_x)]$ taking value 0 at 0. W can be extended to $\mathbb{R}$ by placing independent Wiener processes W_g, W_d on its left and on its right, by the equations $W(u) = W(g_n(a_x)) + W_g(u) - W_g(g_n(a_x))$

for $u < a_x$, $W(u) = W(g_n(b_x)) + W_d(u) - W_d(g_n(b_x))$ for $u > b_x$. Thus, by claim 4 and equation (57), with probability greater than $1 - 2e^{-y}$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{a_x \le t \le b_x} |Z(t) - W_{g_n(t)}| | D_n^T\right] &= \mathbb{E}\left[\sup_{a_x \le t \le b_x} |Z(t) - Z^2(t)| | D_n^T\right] \\
&\le \mathbb{E}\left[\sup_{a_x \le t \le b_x} |Z^1(t) - Z(t)| | D_n^T\right] + \mathbb{E}\left[\sup_{a_x \le t \le b_x} |Z^1(t) - W_{g_n(t)}| | D_n^T\right] \\
&\le \kappa(1+x)^{\frac{7}{6}}[y + \log n]^{\frac{2}{3}} n^{-\frac{u_2}{12}} + \kappa_5(c_1)(1+y)(1+x)^{\frac{3}{2}} n^{-\frac{\delta_5}{3}} \\
&\le \kappa(1+y)(1+x)^{\frac{3}{2}} n^{-u},
\end{aligned}$$

for all $u < \min\left(\frac{u_5}{12}, \frac{u_2}{12}, \frac{\delta_4}{12}, \frac{\delta_5}{3}\right)$ and a constant $\kappa(u)$. Finally, by claim 3, with probability greater than $1 - 3e^{-y}$,

$$\begin{aligned}
&\sup_{\alpha \in [a_x; b_x]} \left| \hat{R}_T^{ho}(\alpha) - [f_n(\alpha) - W_{g_n(\alpha)}] \right| \\
&\le \sup_{\alpha \in [a_x; b_x]} |L(\alpha) - f_n(\alpha)| + \sup_{\alpha \in [a_x; b_x]} |Z(\alpha) - W_{g_n(\alpha)}| \\
&\le \kappa_1(1+x)[\log(n) + \log(2+x) + y]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})} + \kappa(1+y)(1+x)^{\frac{3}{2}} n^{-u} \\
&\le \kappa(1+y)^2(1+x)^{\frac{3}{2}} n^{-u_1},
\end{aligned}$$

for all $u_1 < \min\left(\frac{u_5}{12}, \frac{u_2}{12}, \frac{\delta_4}{12}, \frac{\delta_5}{3}\right)$ and a constant κ. This proves Theorem 1.

8 Appendix

Lemme 8.1 *Let X be a random variable belonging to $[-1; 1]$, with pdf s. For all $j \in \mathbb{N}$, let $\theta_j = \langle s, \psi_j \rangle$. Then*

$$\mathrm{Var}\left(\psi_j(X)\right) \underset{j \to +\infty}{\longrightarrow} 1$$

$$\forall k_0 \le k, \sum_{j=k_0}^{k} |\mathrm{Var}(\psi_j) - 1| \le \|\theta\|_{\ell^1} = \sum_{j=0}^{+\infty} |\langle s, \psi_j \rangle|.$$

Proof $\mathbb{E}[\psi_j(X)] = \int_0^1 \psi_j(x) s(x) dx = \theta_j$. Moreover, $\psi_j(X)^2 = 2\cos^2(2\pi j X) = 1 + \cos(2\pi j X)$, therefore $\mathrm{Var}(\cos(\pi j X)) = 1 + \frac{\theta_j}{\sqrt{2}} - \theta_j^2$, therefore since $|\theta_j| \le \sqrt{2}$, $|\mathrm{Var}(\cos(jX)) - 1| \le \left|\sqrt{2} - \frac{1}{\sqrt{2}}\right| |\theta_j| \le |\theta_j|$. ■

Lemme 8.2 *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a function, $g, h : \mathbb{R}_+ \to \mathbb{R}$ be two non-increasing functions. Then*

$$\inf_{x \in \mathbb{R}_+} \{f(x) + g(x) + h(x)\} \le \inf_{x \in \mathbb{R}_+} \{f(x) + g(x)\} + \inf_{x \in \mathbb{R}_+} \{f(x) + h(x)\}.$$

Proof Let $\delta > 0$. Let x_g be such that $f(x_g) + g(x_g) \le \delta + \inf_{x \in \mathbb{R}_+} \{f(x) + g(x)\}$. Let x_h be such that $f(x_h) + h(x_h) \le \inf_{x \in \mathbb{R}_+} \{f(x) + h(x)\}$. Let $x_* = \max(x_g, x_h)$.

If $x_* = x_g$, then

$$\begin{aligned} f(x_*) + g(x_*) + h(x_*) &\leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} + \delta + h(x_*) \\ &\leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} + \delta + h(x_h) \text{ since } h \text{ is non-increasing} \\ &\leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} + \delta + f(x_h) + h(x_h) \\ &\leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} + \inf_{x\in\mathbb{R}_+} \{f(x) + h(x)\} + 2\delta \end{aligned}$$

Symetrically, if $x_* = x_h$, then $f(x_*) + g(x_*) + h(x_*) \leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} + \inf_{x\in\mathbb{R}_+} \{f(x) + h(x)\} + 2\delta$. As a result,

$$\begin{aligned} \inf_{x\in\mathbb{R}_+} \{f(x) + g(x) + h(x)\} \leq f(x_*) + g(x_*) + h(x_*) \leq \inf_{x\in\mathbb{R}_+} \{f(x) + g(x)\} \\ + \inf_{x\in\mathbb{R}_+} \{f(x) + h(x)\} + 2\delta. \end{aligned}$$

Since no assumptions were made about $\delta > 0$, lemma 8.2 is proved. ■

Proposition 8.3 *For any integers $k_0 \leq k$, with probability greater than $1-e^{-y}$:*

$$\begin{aligned} \left| \sum_{j=k_0+1}^{k} (\hat{\theta}_j^T - \theta_j)^2 - \frac{|k-k_0|}{n_t} \right| &\leq \frac{3\,\|\theta\|_{\ell^1}}{n_t} + (1+\kappa)\,\|s\|_\infty\,(y + \log n) \\ &\quad \times \left[\frac{\sqrt{|k-k_0|}}{n_t} + (y+\log n)\frac{|k-k_0|}{n_t^{\frac{5}{4}}} \right]. \end{aligned} \tag{59}$$

In particular, there exists a constant $\kappa_1 = \kappa_1(\|s\|_\infty, c_1, \|\theta\|_{\ell^1})$ such that for any α_1, α_2 such that $(\alpha_1\Delta, \alpha_2\Delta) \in \mathbb{N}^2$ and $\alpha_1 < \alpha_2$, with probability greater than $1 - e^{-y}$,

$$\left| \sum_{j=k_*+\alpha_1\Delta}^{k_*+\alpha_2\Delta} (\hat{\theta}_j^T - \theta_j)^2 - [\alpha_2 - \alpha_1]\mathcal{E} \right| \leq \kappa_1(\alpha_2 - \alpha_1)[\log n + y]^2 \times n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})}\mathfrak{e}(n). \tag{60}$$

Proof Let $(k_0, k) \in \mathbb{N}^2$ be such that $k_0 < k$. The proof rests on lemma 14 of Arlot and Lerasle [4] applied to $S_m = \langle \psi_{k_0+1}, \ldots, \psi_k \rangle$. Let us compute $b_m = \sup_{u\in\mathbb{R}^{|k-k_0|}:\|u\|\leq 1} \sum_{j=k_0}^{k} u_j\psi_j(x) \leq \sup_x \sqrt{\sum_{j=k_0}^{k} \psi_j^2(x)} \leq \sqrt{|k-k_0|}$ and

$$\mathcal{D}_k = \sum_{j=k_0+1}^{k} \mathrm{Var}\left(\psi_j(X)\right) = |k-k_0| \pm \frac{\|\theta\|_{\ell^1}}{n_t}$$

(by lemma 8.1). Furthermore, $\mathcal{D}_k \leq \sqrt{2}|k-k_0|$ since $\psi_j = \sqrt{2}\cos(2\pi j\cdot) : [0;1] \to [-\sqrt{2}; \sqrt{2}]$. By [4, lemma 14], with probability greater than $1-e^{-y}$, for any $\varepsilon > 0$,

$$\Big| \sum_{j=k_0+1}^{k} (\hat{\theta}_j^T - \theta_j)^2 - \frac{\mathcal{D}_k}{n_t} \Big| \leq \varepsilon\frac{\mathcal{D}_k}{n_t} + \kappa\left(\frac{\|s\|_\infty\,[\log n + y]}{(\varepsilon\wedge 1)n_t} + \frac{|k-k_0|[\log n + y]^2}{(\varepsilon\wedge 1)^3 n_t^2} \right).$$

Let $\varepsilon_1 = \sqrt{\frac{\|s\|_\infty (\log n + y)}{|k-k_0|}} \wedge 1$. If $\varepsilon_1 = 1$, then $|k-k_0| \leq \|s\|_\infty (y + \log n)$ therefore $\varepsilon_1 \frac{|k-k_0|}{n_t} + \kappa \frac{\|s\|_\infty [\log n + y]}{(\varepsilon_1 \wedge 1) n_t} \leq (1+\kappa) \frac{\|s\|_\infty (y + \log n)}{n_t}$. If $\varepsilon_1 < 1$, then $\varepsilon_1 \frac{|k-k_0|}{n_t} + \kappa \frac{\|s\|_\infty [\log n + y]}{(\varepsilon_1 \wedge 1) n_t} = (1+\kappa) \sqrt{\|s\|_\infty (y + \log n)} \frac{\sqrt{|k-k_0|}}{n_t}$. In all cases, if $k > k_0$,

$$\varepsilon_1 \frac{|k-k_0|}{n_t} + \kappa \frac{\|s\|_\infty [\log n + y]}{(\varepsilon_1 \wedge 1) n_t} \leq (1+\kappa) \|s\|_\infty (y + \log n) \frac{\sqrt{|k-k_0|}}{n_t}. \tag{61}$$

Let $\varepsilon_2 = \frac{\sqrt{\log n + y}}{n_t^{\frac{1}{4}}} \wedge 1$. If $\frac{\sqrt{y + \log n}}{n_t^{\frac{1}{4}}} \geq 1 = \varepsilon_2$, then $\varepsilon_2 \frac{|k-k_0|}{n_t} + \kappa \frac{|k-k_0| [\log n + y]^2}{(\varepsilon_2 \wedge 1)^3 n_t^2} \leq \sqrt{y + \log n} \frac{|k-k_0|}{n_t^{\frac{5}{4}}} + \kappa \frac{|k-k_0| (y + \log n)^2}{n_t^2} \leq (1+\kappa)(y + \log n)^2 \frac{|k-k_0|}{n_t^{\frac{5}{4}}}$. If $\varepsilon_2 = \frac{\sqrt{y + \log n}}{n_t^{\frac{1}{4}}} < 1$, then

$$\begin{aligned}
\varepsilon_2 \frac{|k-k_0|}{n_t} + \kappa \frac{|k-k_0| [\log n + y]^2}{(\varepsilon_2 \wedge 1)^3 n_t^2} &= \sqrt{y + \log n} \frac{|k-k_0|}{n_t^{\frac{5}{4}}} + \kappa (y + \log n)^2 \frac{|k-k_0|}{n_t^2} \frac{n_t^{\frac{3}{4}}}{(y + \log n)^{\frac{3}{2}}} \\
&\leq (1+\kappa) \sqrt{y + \log n} \frac{|k-k_0|}{n_t^{\frac{5}{4}}}.
\end{aligned}$$

In all cases,

$$\varepsilon_2 \frac{|k-k_0|}{n_t} + \kappa \frac{|k-k_0| [\log n + y]^2}{(\varepsilon_2 \wedge 1)^3 n_t^2} \leq (1+\kappa)(y + \log n)^2 \frac{|k-k_0|}{n_t^{\frac{5}{4}}}. \tag{62}$$

By lemma 8.2,

$$\begin{aligned}
\Big| \sum_{j=k_0+1}^{k} (\hat{\theta}_j^T - \theta_j)^2 - \frac{\mathcal{D}_k}{n_t} \Big| &\leq \inf_{\varepsilon \geq 0} \left\{ \varepsilon \frac{\mathcal{D}_k}{n_t} + \kappa \frac{\|s\|_\infty [\log n + y]}{(\varepsilon \wedge 1) n_t} \right\} \\
&\quad + \inf_{\varepsilon \geq 0} \left\{ \varepsilon \frac{\mathcal{D}_k}{n_t} + \kappa \frac{|k-k_0| [\log n + y]^2}{(\varepsilon \wedge 1)^3 n_t^2} \right\} \\
&\leq \varepsilon_1 \frac{|k-k_0|}{n_t} + \kappa \frac{\|s\|_\infty [\log n + y]}{(\varepsilon_1 \wedge 1) n_t} + \varepsilon_2 \frac{|k-k_0|}{n_t} \\
&\quad + \kappa \frac{|k-k_0| [\log n + y]^2}{(\varepsilon_2 \wedge 1)^3 n_t^2} + (\varepsilon_1 + \varepsilon_2) \frac{\|\theta\|_{\ell^1}}{n_t} \\
&\leq (1+\kappa) \|s\|_\infty (y + \log n) \frac{\sqrt{|k-k_0|}}{n_t} \\
&\quad + (1+\kappa)(y + \log n)^2 \frac{|k-k_0|}{n_t^{\frac{5}{4}}} + \frac{2 \|\theta\|_{\ell^1}}{n_t},
\end{aligned}$$

by equations (61), (62). In conclusion, on an event E_y of probability greater than $1 - e^{-y}$,

$$\begin{aligned}
\Big| \sum_{j=k_0+1}^{k} (\hat{\theta}_j^T - \theta_j)^2 - \frac{|k-k_0|}{n_t} \Big| &\leq \Big| \sum_{j=k_0+1}^{k} (\hat{\theta}_j^T - \theta_j)^2 - \frac{\mathcal{D}_k}{n_t} \Big| + \frac{\|\theta\|_{\ell^1}}{n_t} \\
&\leq \frac{3 \|\theta\|_{\ell^1}}{n_t} + (1+\kappa) \|s\|_\infty (y + \log n) \\
&\quad \times \left[\frac{\sqrt{|k-k_0|}}{n_t} + (y + \log n) \frac{|k-k_0|}{n_t^{\frac{5}{4}}} \right].
\end{aligned}$$

This proves equation (59).

If $k_0 = k_* + \alpha_1 \Delta$ and $k = k_* + \alpha_2 \Delta$, then by hypothesis 4 of Theorem 1,

$$\frac{\sqrt{|k-k_0|}}{n_t} = \sqrt{\alpha_2 - \alpha_1}\sqrt{\frac{\Delta}{n_v n_t}}\sqrt{\frac{n_v}{n_t}} = \sqrt{\alpha_2 - \alpha_1}\sqrt{\frac{n-n_t}{n_t}}\mathfrak{e} \leq \sqrt{\alpha_2 - \alpha_1} n^{-\frac{\delta_4}{2}}\mathfrak{e}. \tag{63}$$

Furthermore,

$$\begin{aligned}\frac{|k-k_0|}{n_t^{\frac{5}{4}}} &= (\alpha_2 - \alpha_1)\frac{\mathcal{E}}{n_t^{\frac{1}{4}}} \\ &= (\alpha_2 - \alpha_1)\sqrt{\frac{\mathcal{E}}{n_v}}\frac{\sqrt{n_v \mathcal{E}}}{n_t^{\frac{1}{4}}} \\ &\leq (\alpha_2 - \alpha_1)\mathfrak{e}\frac{\sqrt{2n_v \mathrm{or}(n_t) + 1}}{n_t^{\frac{1}{4}}}.\end{aligned}$$

Let $k_1 = \lceil n_t^{\frac{1}{3+\delta_1}} \rceil$, so that $n_t^{\frac{1}{3+\delta_1}} \leq k_1 \leq 2n_t^{\frac{1}{3+\delta_1}}$. By hypothesis 1 of Theorem 1, $\sum_{j=k+1}^{+\infty} \theta_j^2 \leq \frac{c_1}{k^{2+\delta_1}}$ therefore

$$\mathrm{or}(n_t) \leq \inf_{k \in \mathbb{N}^*} \frac{c_1}{k^{2+\delta_1}} + \frac{k}{n_t} \leq \frac{c_1}{k_1^{2+\delta_1}} + \frac{k_1}{n_t} \leq \frac{c_1}{n_t^{\frac{2}{3+\delta_1}}} + \frac{2n_t^{\frac{1}{3+\delta_1}}}{n_t} \leq \frac{2+c_1}{n_t^{\frac{2}{3+\delta_1}}}.$$

Thus $1 + 2n_v \mathrm{or}(n_t) \leq 1 + 2n_t \mathrm{or}(n_t) \leq (5 + 2c_1) n_t^{\frac{1+\delta_1}{3+\delta_1}}$, hence

$$\begin{aligned}\frac{|k-k_0|}{n_t^{\frac{5}{4}}} &\leq (\alpha_2 - \alpha_1)\mathfrak{e}\sqrt{5+2c_1}\frac{n_t^{\frac{1+\delta_1}{6+2\delta_1}}}{n_t^{\frac{1}{4}}} \\ &\leq (\alpha_2 - \alpha_1)\sqrt{5+2c_1} n_t^{-\frac{1}{12}}\mathfrak{e} \\ &\leq (\alpha_2 - \alpha_1)\sqrt{5+2c_1}\frac{2^{\frac{1}{12}}}{n^{\frac{1}{12}}}\mathfrak{e}.\end{aligned} \tag{64}$$

Finally, $\frac{\|\theta\|_{\ell^1}}{n_t} = \frac{n_v}{n_t}\frac{\|\theta\|_{\ell^1}}{n_v} \leq \|\theta\|_{\ell^1} \frac{n-n_t}{n_t}\mathfrak{e} \leq \|\theta\|_{\ell^1} n^{-\delta_4}\mathfrak{e}$. Equation (60) follows from equations (59), (63) and (64).

■

Lemme 8.4 *Let $(c_{i,j})_{(i,j)\in\mathbb{N}^2}$ be real coefficients. Let $I_1, I_2 \subset \mathbb{N}$ be two finite sets. Let $(\theta_j)_{j\in\mathbb{N}}$ be a sequence. Let $C = \max\left\{\sup_{i\in I_1}\sum_{j\in I_2}|c_{i,j}|, \sup_{i\in I_2}\sum_{j\in I_1}|c_{i,j}|\right\}$. Then*

$$\sum_{i\in I_1}\left(\sum_{j\in I_2} c_{i,j}\theta_j\right)^2 \leq C^2 \sum_{j\in I_2}\theta_j^2$$

and

$$\left|\sum_{i\in I_1}\sum_{j\in I_2}\theta_i\theta_j c_{i,j}\right| \leq C \max\left\{\sum_{i\in I_1}\theta_i^2, \sum_{j\in I_2}\theta_j^2\right\}.$$

Proof Let $C_i = \sum_{j\in I_2} |c_{i,j}|$. Then

$$\begin{aligned}
\sum_{i\in I_1}\left(\sum_{j\in I_2} c_{i,j}\theta_j\right)^2 &= \sum_{i\in I_1} C_i^2\left(\frac{1}{C_i}\sum_{j\in I_2}\mathrm{sgn}(c_{i,j})|c_{i,j}|\theta_j\right)^2 \\
&\le \sum_{i\in I_1}\frac{C_i^2}{C_i}\sum_{j\in I_2}|c_{i,j}|\theta_j^2 \text{ by the Jensen inequality} \\
&\le \left(\max_{i\in I_1} C_i\right)\sum_{j\in I_2}\theta_j^2\sum_{i\in I_1}|c_{i,j}| \\
&\le C^2\sum_{j\in I_2}\theta_j^2.
\end{aligned}$$

This proves the first equation. Furthermore,

$$\begin{aligned}
\left|\sum_{i\in I_1}\sum_{j\in I_2}\theta_i\theta_j c_{i,j}\right| &\le \sum_{i\in I_1}\sum_{j\in I_2}\frac{\theta_i^2+\theta_j^2}{2}|c_{i,j}| \\
&= \frac{1}{2}\sum_{i\in I_1}\theta_i^2\sum_{j\in I_2}|c_{i,j}| + \frac{1}{2}\sum_{j\in I_2}\theta_j^2\sum_{i\in I_1}|c_{i,j}| \\
&\le C\max\left\{\sum_{i\in I_1}\theta_i^2, \sum_{j\in I_2}\theta_j^2\right\},
\end{aligned}$$

which proves the second equation. ■

Lemme 8.5 *Under the assumptions of Theorem 1, there exists a constant* $\kappa(c_1, c_2) > 0$ *such that for any* $x \ge 0$,

$$k_* + a_x\Delta \ge \frac{\kappa}{(1+x)^{\frac{1}{\delta_2}}} n_t^{\frac{2}{3\delta_2}}.$$

Proof By hypothesis 2 of Theorem 1,

$$\begin{aligned}
c_2(k_* + a_x\Delta)^{-\delta_2} &\le \sum_{j=k_*+a_x\Delta+1}^{+\infty}\theta_j^2 \\
&\le \sum_{j=k_*+a_x\Delta+1}^{k_*}\left[\theta_j^2 - \frac{1}{n_t}\right] + |a_x|\mathcal{E} + \sum_{j=k_*+1}^{+\infty}\theta_j^2 \\
&\le \mathfrak{e} f_n(a_x) + |a_x|\mathcal{E} + \mathrm{or}(n_t). \qquad (65)
\end{aligned}$$

By definition, $f_n(a_x) \le x$ and by lemma 7.1, $|a_x| \le 2(1+x)$. Furthermore, by lemma 3.1, $\mathfrak{e} \le \mathcal{E} \le 2\mathrm{or}(n_t) + \frac{1}{n_v}$. Since by hypothesis 5 of Theorem 1, $n_v \ge n^{\frac{2}{3}+\delta_5}$, it follows that: $\mathcal{E} \le 2\mathrm{or}(n_t) + \frac{1}{n^{\frac{2}{3}+\delta_5}}$. Equation (65) thus yields

$$c_2(k_* + a_x\Delta)^{-\delta_2} \le 6(1+x)\left[\mathrm{or}(n_t) + \frac{1}{n^{\frac{2}{3}}}\right].$$

On the other hand, by hypothesis 1 of Theorem 1,

$$\mathrm{or}(n_t) \leq \min_{k\in\mathbb{N}} \frac{c_1}{k^2} + \frac{k}{2n_t} \leq \frac{3c_1^{\frac{1}{3}}}{n_t^{\frac{2}{3}}}.$$

It follows finally that, for some constant $\kappa(c_1, c_2)$,

$$k_* + a_x\Delta \geq \frac{\kappa}{(1+x)^{\frac{1}{\delta_2}}} n_t^{\frac{2}{3\delta_2}}.$$

■

Claim 7 *Let $u_2 = \min\left(\frac{2\delta_3}{3\delta_2}, \delta_4\right)$. Let x be a non-negative real number. Let a_x, b_x be such that $a_x \leq 0 \leq b_x$ and $\max(f_n(a_x), f_n(b_x)) \leq x$. Assume also that $a_x\Delta - 1 \geq \frac{-k_*}{\Delta}$. There exists a constant $\kappa_3 \geq 0$ such that for all $j \in [a_x\Delta; b_x\Delta + 1]$,*

$$\left|f_n\left(\tfrac{j}{\Delta}\right) - f_n\left(\tfrac{j-1}{\Delta}\right)\right| \leq \kappa_3(1+x)^2 n^{-u_2} \tag{66}$$

$$\theta^2_{k_*+j} \leq \kappa_3(1+x)^2 n^{-u_2}\mathfrak{e}. \tag{67}$$

Proof By hypothesis 3 of Theorem 1, for all $k \geq 1$, $\theta^2_{k+k^{\delta_3}} \geq c_3\theta^2_{k-k^{\delta_3}}$. Thus, for all $k \geq 1$ and any $j \in [k - k^{\delta_3}; k + k^{\delta_3}]$,

$$\begin{aligned}\max\left(\theta_k^2, \frac{1}{n_t}\right) &\leq \max\left(c_3\theta_j^2, \frac{1}{n_t}\right)\\ &\leq \frac{1+c_3}{n_t} + c_3\left|\theta_j^2 - \frac{1}{n_t}\right|.\end{aligned} \tag{68}$$

Let $k \in [k_* + a_x\Delta; k_* + b_x\Delta + 1]$. Assume without loss of generality (up to a change in the constant κ_9) that $x \geq 1$. Thus by lemma 3.2, $\max(-a_x, b_x) \geq 1$.

- If $|b_x| \geq 1$, then two cases can be distinguished.
  - If $k \leq k_* + \frac{\Delta}{2}$, then $k + k^{\delta_3} \wedge \frac{\Delta}{2} \leq k_* + \Delta \leq k_* + b_x\Delta$, therefore by definition of a_x, b_x,

$$2x\mathfrak{e} \geq \mathfrak{e}[f_n(a_x)+f_n(b_x)] = \sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\left|\theta_j^2 - \frac{1}{n_t}\right| \geq \sum_{j=k+1}^{k+k^{\delta_3}\wedge\frac{\Delta}{2}}\left|\theta_j^2 - \frac{1}{n_t}\right|.$$

  - If $k_* + \frac{\Delta}{2} < k \leq k_* + b_x\Delta + 1$, then $k - k^{\delta_3} \wedge \frac{\Delta}{2} \geq k_*$, therefore

$$2x\mathfrak{e} \geq \sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\left|\theta_j^2 - \frac{1}{n_t}\right| \geq \sum_{j=k-k^{\delta_3}\wedge\frac{\Delta}{2}}^{k-1}\left|\theta_j^2 - \frac{1}{n_t}\right|.$$

- If $|a_x| \geq 1$, then we likewise consider two possibilities.
  - If $k > k_* - \frac{\Delta}{2}$, then $k - k^{\delta_3} \wedge \frac{\Delta}{2} > k_* - \Delta \geq k_* + a_x\Delta$, therefore by definition of a_x, b_x,

$$2x\mathfrak{e} \geq \sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\left|\theta_j^2 - \frac{1}{n_t}\right| \geq \sum_{j=k-k^{\delta_3}\wedge\frac{\Delta}{2}}^{k-1}\left|\theta_j^2 - \frac{1}{n_t}\right|.$$

– If $k \leq k_* - \frac{\Delta}{2}$, then $k + k^{\delta_3} \wedge \frac{\Delta}{2} \leq k_*$, therefore

$$2x\mathfrak{e} \geq \sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta} \left|\theta_j^2 - \frac{1}{n_t}\right| \geq \sum_{j=k+1}^{k+k^{\delta_3}\wedge\frac{\Delta}{2}} \left|\theta_j^2 - \frac{1}{n_t}\right|.$$

In all cases, by equation (68),

$$\left(k^{\delta_3} \wedge \frac{\Delta}{2}\right) \max\left(\theta_k^2, \frac{1}{n_t}\right) \leq k^{\delta_3} \wedge \frac{\Delta}{2} \frac{1+c_3}{n_t} + 2x\mathfrak{e},$$

in other words

$$\max\left(\theta_k^2, \frac{1}{n_t}\right) \leq \frac{1+c_3}{n_t} + \frac{2x\mathfrak{e}}{k^{\delta_3} \wedge \frac{\Delta}{2}}.$$

Furthermore, by hypothesis 4 of Theorem 1, $\Delta \geq \frac{n_t}{n-n_t} \geq n^{\delta_4}$, and by lemma 8.5,

$$k^{\delta_3} \geq (k_* + a_x\Delta)^{\delta_3} \geq \frac{\kappa}{(1+x)^{\frac{\delta_3}{\delta_2}}} n_t^{\frac{2\delta_3}{3\delta_2}}.$$

Let $u_2 = \min\left(\frac{2\delta_3}{3\delta_2}, \delta_4\right)$ Since $\delta_3 \leq \delta_2$, there exists therefore a constant κ such that

$$\max\left(\theta_k^2, \frac{1}{n_t}\right) \leq \kappa(1+x)^2 n^{-u_2}\mathfrak{e}.$$

In conclusion, for all $j \in \{a_x\Delta, \ldots, b_x\Delta + 1\}$,

$$\theta_{k_*+j}^2 \leq \max\left(\theta_{k_*+j}^2, \frac{1}{n_t}\right) \leq \kappa(1+x)^2 n^{-u_2}\mathfrak{e}$$

$$\begin{aligned}\left|f_n\left(\tfrac{j}{\Delta}\right) - f_n\left(\tfrac{j-1}{\Delta}\right)\right| &= \frac{1}{\mathfrak{e}}|\theta_{j+k_*}^2 - \frac{1}{n_t}| \\ &\leq \frac{1}{\mathfrak{e}} \max\left(\theta_{k_*+j}^2, \frac{1}{n_t}\right) \\ &\leq \kappa(1+x)^2 n^{-u_2}.\end{aligned}$$

This proves claim 7. ■

Proposition 8.6 *Let P be the probability measure with pdf s on $[0;1]$. Let $\theta_j = \langle s, \psi_j\rangle = P(\psi_j)$ and $\theta_j^2 = \theta_j^2$, and assume that they satisfy the hypotheses of Theorem 1. Let $\hat{\theta}_j^T = P^T\psi_j$. Let $I_k^1, I_k^2 \subset \{k_* + a_x\Delta + 1, \ldots, k_* + b_x\Delta\}$ be two intervals. Then the statistics*

$$U_{I_k^1, I_k^2} = \sum_{i \in I_k^1} \sum_{j \in I_k^2} \hat{\theta}_i^T \hat{\theta}_j^T [P(\psi_i\psi_j) - P\psi_i P\psi_j]$$

can be approximated in the following way. There exists two constants κ_4 and $u_3 > 0$ such that, with probability greater than $1 - e^{-y}$,

$$\begin{aligned}U_{I_k^1, I_k^2} = \frac{1}{2}\frac{|I_k^1 \cap I_k^2|}{n_t} + \left(1 - \frac{1}{\sqrt{2}}\right) \sum_{i \in I_k^1 \cap I_k^2} \theta_i^2 + \frac{1}{\sqrt{2}} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \theta_i\theta_j\theta_{|i-j|} + \frac{1}{2n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \theta_{|i-j|}^2 \\ \pm \kappa_4(y + \log n)^2(1+x)n^{-u_3}\mathcal{E}.\end{aligned}$$

Proof First, by lemma 7.1,

$$\max\left(\sum_{i\in I_k^1}\theta_i^2, \sum_{j\in I_k^2}\theta_j^2\right) \leq \sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\theta_j^2 \leq 4(1+x)\mathcal{E}. \tag{69}$$

Let $c_{i,j} = \frac{\theta_{i+j}}{\sqrt{2}} + \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} - \theta_i\theta_j$. $U_{I_k^1,I_k^2}$ can be expressed as the sum of 6 terms: $U_{I_k^1,I_k^2} = V_1 + V_2 + V_3 + V_4 + V_5 + V_6$, where

$$\begin{aligned}
V_1 &= \sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\left[\frac{\theta_{i+j}}{\sqrt{2}} + \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} - \theta_i\theta_j\right]\\
V_2 &= (P^T - P)\sum_{i\in I_k^1}\psi_i\sum_{j\in I_k^2}\theta_j c_{i,j}\\
V_3 &= (P^T - P)\sum_{j\in I_k^2}\psi_j\sum_{i\in I_k^1}\theta_i c_{i,j}\\
V_4 &= \frac{1}{\sqrt{2}}\sum_{i\in I_k^1}\sum_{j\in I_k^2}(P^T - P)\psi_i(P^T - P)\psi_j\theta_{|i-j|}\\
V_5 &= \left(1 - \frac{1}{\sqrt{2}}\right)\sum_{j\in I_k^1\cap I_k^2}\left(\hat{\theta}_j^T - \theta_j\right)^2\\
V_6 &= \sum_{i\in I_k^1}\sum_{j\in I_k^2}(P^T - P)\psi_i(P^T - P)\psi_j\left[\frac{\theta_{i+j}}{\sqrt{2}} - \theta_i\theta_j\right]
\end{aligned}$$

The first term is

$$V_1 = \left(1 - \frac{1}{\sqrt{2}}\right)\sum_{i\in I_k^1\cap I_k^2}\theta_i^2 + \frac{1}{\sqrt{2}}\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\theta_{|i-j|} + \sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\left[\frac{\theta_{i+j}}{\sqrt{2}} - \theta_i\theta_j\right].$$

For all $i \in I_k^1$,

$$\sum_{j\in I_k^2}\frac{|\theta_{i+j}|}{\sqrt{2}} + |\theta_i||\theta_j| \leq 2\sum_{j\geq k_*+a_x\Delta+1}|\theta_j|.$$

Furthermore, for all $k \geq 2$, by hypothesis 1 of Theorem 1,

$$\sum_{j\geq k}|\theta_j| \leq \sum_{j=k}^{+\infty}\sqrt{\sum_{i=j}^{+\infty}\theta_i^2} \leq \sum_{j=k}^{+\infty}\frac{c_1}{(j-1)^{1+\frac{\delta_1}{2}}} \leq \frac{2c_1}{\delta_1}(k-1)^{\frac{-\delta_1}{2}}. \tag{70}$$

Since $k_* + a_x\Delta \geq \frac{\kappa}{(1+x)^{\frac{1}{\delta_2}}}n_t^{\frac{2}{3\delta_2}}$ by lemma 8.5, there is a constant $\kappa(c_1, c_2)$ such that

$$\sum_{j\in I_k^2}\frac{|\theta_{i+j}|}{\sqrt{2}} + |\theta_i||\theta_j| \leq \kappa\frac{(1+x)^{\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}. \tag{71}$$

The same argument applies to $\sum_{i\in I_k^1}\frac{|\theta_{i+j}|}{\sqrt{2}} + |\theta_i||\theta_j|$. Thus, by lemma 8.4,

$$\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\left[\theta_{i+j} - \theta_i\theta_j\right] \leq 2\kappa\frac{(1+x)^{\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\left[\sum_{i\in I_k^1}\theta_i^2 + \sum_{j\in I_k^2}\theta_j^2\right].$$

By equation (69), it follows that for a certain constant $\kappa(c_1, c_2)$,

$$\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\left[\frac{\theta_{i+j}}{\sqrt{2}}-\theta_i\theta_j\right]\leq\kappa\frac{(1+x)^{1+\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\mathcal{E}.$$

Thus

$$V_1=\left(1-\frac{1}{\sqrt{2}}\right)\sum_{i\in I_k^1\cap I_k^2}\theta_i^2+\frac{1}{\sqrt{2}}\sum_{i\in I_k^1}\sum_{j\in I_k^2}\theta_i\theta_j\theta_{|i-j|}\pm\kappa\frac{(1+x)^{1+\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\mathcal{E} \tag{72}$$

Bernstein's inequality applies to V_2 and V_3. By symmetry, let us only consider V_2. Its variance satisfies the following inequality.

$$\begin{aligned}\mathrm{Var}\left(\sum_{i\in I_k^1}\psi_i\sum_{j\in I_k^2}\theta_jc_{i,j}\right)&\leq\|s\|_\infty\left\|\sum_{i\in I_k^1}\psi_i\sum_{j\in I_k^2}\theta_jc_{i,j}\right\|^2\\&\leq\|s\|_\infty\sum_{i\in I_k^1}\left(\sum_{j\in I_k^2}\theta_jc_{i,j}\right)^2.\end{aligned}$$

Let us now apply lemma 8.4. For all $i\in I_k^1$,

$$\begin{aligned}\sum_{j\in I_k^2}|c_{i,j}|&\leq\frac{1}{\sqrt{2}}\sum_{j\in I_k^2}|\theta_{i+j}|+\frac{1}{\sqrt{2}}\sum_{j\in I_k^2}|\theta_{|i-j|}|+|\theta_i|\sum_{j\in I_k^2}|\theta_j|\\&\leq\left(\sqrt{2}+\sup_{i\in\mathbb{N}}|\theta_i|\right)\sum_{r\in\mathbb{N}}|\theta_r|\\&\leq3\|\theta\|_{\ell^1}\end{aligned}$$

In the same way, for all $j\in I_k^2$, $\sum_{i\in I_k^1}|c_{i,j}|\leq3\|\theta\|_{\ell^1}$, hence by lemma 8.4,

$$\begin{aligned}\mathrm{Var}\left(\sum_{i\in I_k^1}\psi_i\sum_{j\in I_k^2}\theta_jc_{i,j}\right)&\leq3\|\theta\|_{\ell^1}\|s\|_\infty\sum_{j\in I_k^2}\theta_j^2\\&\leq12\|\theta\|_{\ell^1}\|s\|_\infty(1+x)\mathcal{E}\text{ by equation (69).}\end{aligned} \tag{73}$$

As for the upper bound on the uniform norm, it follows from lemma 8.4 and the elementary upper bound $\|\psi_i\|_\infty\leq\sqrt{2}$ that

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\sum_{i\in I_k^1}\psi_i(x)\sum_{j\in I_k^2}\theta_jc_{i,j}\right|&\leq\sqrt{\sum_{i\in I_k^1}\left(\sum_{j\in I_k^2}\theta_jc_{i,j}\right)^2}\sup_{x\in\mathbb{R}}\sqrt{\sum_{i\in I_k^1}\psi_i(x)^2}\\&\leq3\|\theta\|_{\ell^1}\sqrt{2|I_k^1|}\sqrt{\sum_{j\in I_k^2}\theta_j^2}\\&\leq3\|\theta\|_{\ell^1}\sqrt{2(b_x-a_x)\Delta}\sqrt{4(1+x)\mathcal{E}}\\&\leq\kappa(1+x)\sqrt{\Delta\mathcal{E}}\text{ by lemma 7.1,}\end{aligned} \tag{74}$$

for some constant $\kappa = \kappa(\|\theta\|_{\ell^1})$. By Bernstein's inequality, there exists an event $E_2(y) \subset \mathbb{R}^{n_t}$ with probability $\mathbb{P}(D_n^T \in E_2(y)) \geq 1 - e^{-y}$ such that, for any $D_n^T \in E_2(y)$,

$$|V_2| \leq \sqrt{\frac{2y}{n_t}} \sqrt{\mathrm{Var}\left(\sum_{i \in I_k^1} \psi_i \sum_{j \in I_k^2} \theta_j c_{i,j}\right)} + \frac{y}{3n_t} \sup_{x \in \mathbb{R}} \left| \sum_{i \in I_k^1} \psi_i(x) \sum_{j \in I_k^2} \theta_j c_{i,j} \right|$$
$$\leq \sqrt{24 \|\theta\|_{\ell^1} \|s\|_\infty} \sqrt{\frac{y(1+x)\mathcal{E}}{n_t}} + \frac{\kappa y}{3n_t}(1+x)\sqrt{\Delta \mathcal{E}} \text{ by (73), (74).}$$

Setting $\kappa = \max\left(\sqrt{24 \|\theta\|_{\ell^1} \|s\|_\infty}, \frac{\kappa}{3}\right)$, it follows that on $E_2(y)$,

$$|V_2| \leq \kappa \sqrt{y(1+x)} \sqrt{\frac{n_v}{n_t}} \mathfrak{e} + \kappa y (1+x) \sqrt{\frac{n_v}{n_t}} \mathfrak{e} \sqrt{\mathcal{E}}.$$

By lemma 3.1, $\mathcal{E}$ is uniformly bounded: $\mathcal{E} \leq 2\sum_{j=1}^{n_t} \theta_j^2 + \frac{1}{n-n_t} \leq 1 + 2\|s\|^2 \leq 1 + 2\|s\|_\infty$. Furthermore, by hypothesis 4 of Theorem 1, $\sqrt{\frac{n_v}{n_t}} = \sqrt{\frac{n-n_t}{n_t}} \leq n^{-\frac{\delta_4}{2}}$. Thus, there exists a constant $\kappa(\|\theta\|_{\ell^1}, \|s\|_\infty)$ such that, on $E_2(y)$,

$$|V_2| \leq \kappa y (1+x) n^{-\frac{\delta_4}{2}} \mathfrak{e}. \tag{75}$$

Symmetrically, there exists an event $E_3(y)$ of probability greater than $1 - e^{-y}$, such that for any $D_n^T \in E_3(y)$,

$$|V_3| \leq \kappa y (1+x) n^{-\frac{\delta_4}{2}} \mathfrak{e}. \tag{76}$$

Now consider V_4. This term can be expressed as a finite sum of sums of squares:

$$V_4 = \frac{1}{\sqrt{2}} \sum_{r \in \mathbb{Z}} \sum_{i \in I_k^1 \cap (I_k^2 - r)} (P^T - P)\psi_i (P^T - P)\psi_{i+r} \theta_{|r|}$$
$$= \frac{1}{4\sqrt{2}} \sum_{r \in \mathbb{Z}} \theta_{|r|} \sum_{i \in I_k^1 \cap (I_k^2 - r)} \left[(P^T - P)(\psi_i + \psi_{i+r})\right]^2 - \left[(P^T - P)(\psi_i - \psi_{i+r})\right]^2.$$

Let $J_0 = \{j \in \mathbb{N} : \lfloor \frac{j}{r} \rfloor \text{ is even}\}$ and $J_1 = \{j \in \mathbb{N} : \lfloor \frac{j}{r} \rfloor \text{ is odd}\}$. Thus

$$V_4 = \frac{1}{4\sqrt{2}} \sum_{r \in \mathbb{Z}} \theta_{|r|} \sum_{(z,\varepsilon) \in \{0;1\} \times \{-1;1\}} \sum_{j \in J_z} \varepsilon (P^T - P)(\psi_i + \varepsilon \psi_{i+r})^2 \mathbb{I}_{I_k^1}(i) \mathbb{I}_{I_k^2}(i+r).$$

For any fixed $r \neq 0$, $(z, \varepsilon) \in \{0;1\} \times \{-1;1\}$, $\frac{1}{\sqrt{2}}\left(\psi_i + \varepsilon \psi_{i+r}\right)_{i \in J_z}$ is an orthonormal collection of functions, since for any $(i,j) \in J_z^2$,

$$\begin{aligned} < \psi_i + \varepsilon\psi_{i+r}, \psi_j + \varepsilon\psi_{j+r} > &= < \psi_i, \psi_j > + \varepsilon < \psi_{i+r}, \psi_j > + \varepsilon < \psi_i, \psi_{j+r} > + < \psi_{i+r}, \psi_{j+r} > \\ &= 2\delta_{i,j} + \varepsilon < \psi_{i+r}, \psi_j > + \varepsilon < \psi_i, \psi_{j+r} > \\ &= 2\delta_{i,j} \text{ since } i, j \in J_z \text{ and } i+r, j+r \in J_{1-z}. \end{aligned}$$

[4, Lemma 14] applied to $S_m = \langle(\psi_i + \varepsilon\psi_{i+r})_{i\in J_z\cap I_k^1\cap I_k^2}\rangle$ for all $(z,\varepsilon) \in \{0;1\} \times \{-1;1\}$, $r \in \{-n_t,\dots,n_t\}$ and a union bound yield an event $E_4(y)$ of probability $\mathbb{P}(D_n^T \in E_4(y)) \geq 1 - e^{-y}$ such that, for some absolute constant κ and for all $D_n^T \in E_4(y)$, $(z,\varepsilon) \in \{0;1\} \times \{-1;1\}$ and $r \in \mathbb{Z}$,

$$\begin{aligned}\sum_{i\in J_z\cap I_k^1\cap(I_k^2-r)} \varepsilon(P^T - P)(\psi_i + \varepsilon\psi_{i+r})^2 = (1\pm\delta)\frac{\varepsilon}{n_t}\sum_{i\in J_z\cap I_k^1\cap(I_k^2-r)} \big[\mathrm{Var}(\psi_i) + \mathrm{Var}(\psi_{i+r}) \\ + 2\varepsilon\,\mathrm{Cov}(\psi_i,\psi_{i+r})\big] + \kappa\frac{\|s\|_\infty\,[\log(1+r) + \log n_t + y]}{(\delta\wedge 1)n_t} \\ + \kappa\frac{|I_k^1|[\log(1+r) + \log n_t + y]^2}{(\delta\wedge 1)^3 n_t^2}.\end{aligned}$$

By summing on $(r,z,\varepsilon) \in \mathbb{Z}\times\{0;1\}\times\{-1;1\}$ and since $\|\psi_i\|_\infty \leq \sqrt{2}$, it follows that for all $D_n^T \in E_4(y)$,

$$\begin{aligned}\left|V_4 - \frac{1}{n_t}\sum_{r\in\mathbb{Z}}\frac{\theta_{|r|}}{\sqrt{2}}\sum_{i\in I_k^1\cap(I_k^2-r)} c_{i,i+r}\right| &= \left|V_4 - \frac{1}{n_t}\sum_{r=-n_t}^{n_t}\frac{\theta_{|r|}}{\sqrt{2}}\sum_{i\in I_k^1\cap(I_k^2-r)} c_{i,i+r}\right| \\ &\leq \frac{\delta}{n_t\sqrt{2}}\sum_{r\in\mathbb{Z}}|\theta_{|r|}|\sum_{i\in I_k^1\cap(I_k^2-r)}[\mathrm{Var}(\psi_i) + \mathrm{Var}(\psi_{i+r})] \\ &\quad + \kappa\sum_{r\in\mathbb{Z}}\frac{|\theta_{|r|}|}{\sqrt{2}} \times \frac{\|s\|_\infty\,[\log n_t\log(1+r) + y]}{(\delta\wedge 1)n_t} \\ &\quad + \kappa\sum_{r\in\mathbb{Z}}\frac{|\theta_{|r|}|}{\sqrt{2}} \times \frac{|I_k^1|[\log n_t + \log(1+r) + y]^2}{(\delta\wedge 1)^3 n_t^2}\end{aligned}$$

By hypothesis 1, $|\theta_j| \leq \frac{\sqrt{c_1}}{j^{1+\frac{\delta_1}{2}}}$, hence the sum $\sum_{r\in\mathbb{Z}}|\theta_{|r|}|\log(1+r)^2$ converges to a finite value $\|\theta\|_{1,\log^2}$. Moreover, by lemma 7.1, $|I_k^1| \leq (b_x - a_x)\Delta \leq 2(1+x)\Delta$, hence

$$\begin{aligned}\left|V_4 - \frac{1}{n_t}\sum_{r\in\mathbb{Z}}\frac{\theta_{|r|}}{\sqrt{2}}\sum_{i\in I_k^1\cap(I_k^2-r)} c_{i,i+r}\right| &\leq 6\,\|\theta\|_{\ell^1}\,(1+x)\frac{\delta}{n_t}\Delta + 2\kappa\,\|\theta\|_{1,\log^2}\,\frac{\|s\|_\infty\,[1+y]}{(\delta\wedge 1)n_t} \\ &\quad + 8\kappa(1+x)\,\|\theta\|_{1,\log^2}\,\frac{\Delta[1+y]^2}{(\delta\wedge 1)^3 n_t^2}.\end{aligned}$$

There exists therefore a constant $\kappa(\|\theta\|_{1,\log^2})$ such that, for all $D_n^T \in E_4(y)$,

$$\left|V_4 - \frac{1}{n_t}\sum_{r\in\mathbb{Z}}\frac{\theta_{|r|}}{\sqrt{2}}\sum_{i\in I_k^1\cap(I_k^2-r)} c_{i,i+r}\right| \leq \kappa\delta(1+x)\mathcal{E} + \frac{[\log n_t + y]}{(\delta\wedge 1)n_t} + \kappa(1+x)\frac{[\log n_t + y]^2}{(\delta\wedge 1)^3 n_t}\mathcal{E}.$$

Let now $\delta = \max\left\{\frac{n-n_t}{n_t}, n^{-\frac{1}{3}}\right\}^{\frac{3}{4}}$. By hypothesis 4 of section 4, $\frac{n-n_t}{n_t} \leq n^{-\delta_4}$, therefore $\delta\mathcal{E} \leq n^{-\min\left(\frac{1}{4},\frac{3\delta_4}{4}\right)}\mathcal{E}$. Moreover, $\mathcal{E} \geq \frac{1}{n_v}$ therefore $\frac{1}{\delta n_t} \leq \left(\frac{n-n_t}{n_t}\right)^{\frac{1}{4}}\frac{1}{n_v} \leq$

$n^{-\frac{\delta_4}{4}}\mathcal{E}$. Finally, since $\delta \geq n^{-\frac{1}{4}}$ and $n_t \geq \frac{n}{2}$, $\frac{\mathcal{E}}{\delta^3 n_t} \leq 2n^{-\frac{1}{4}}\mathcal{E}$. Since $\delta_4 \leq 1$, there exists therefore a constant κ such that for all $D_n^T \in E_4(y)$,

$$\left| V_4 - \frac{1}{n_t\sqrt{2}} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \theta_{|i-j|} c_{i,j} \right| \leq \kappa(1+x)[\log n_t + y]^2 n^{-\frac{\delta_4}{4}} \mathcal{E}. \tag{77}$$

Moreover, since $c_{i,j} = \frac{\theta_{i+j}}{\sqrt{2}} + \left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} - \theta_i\theta_j$ and $\theta_0 = 1$,

$$\begin{aligned}\frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}}{\sqrt{2}} c_{i,j} = & \sum_{i \in I_k^1 \cap I_k^2} \frac{1}{\sqrt{2}} \left(1 - \frac{1}{\sqrt{2}}\right) \frac{1}{n_t} + \frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}^2}{2} \\ & + \frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}\theta_{i+j}}{2} - \frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}}{\sqrt{2}} \theta_i \theta_j.\end{aligned}$$

Since for all $j \in \mathbb{N}$, $|\theta_j| \leq 1$,

$$\begin{aligned}\left| \frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}}{\sqrt{2}} c_{i,j} - \left(1 - \frac{1}{\sqrt{2}}\right) \frac{|I_k^1 \cap I_k^2|}{n_t\sqrt{2}} - \frac{1}{n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \frac{\theta_{|i-j|}^2}{2} \right| & \leq \frac{2}{n_t} \left(\sum_{r \in \mathbb{N}} |\theta_r| \right)^2 \\ & \leq 2 \frac{n - n_t}{n_t} \frac{1}{n_v} \|\theta\|_{\ell^1}^2 \\ & \leq 2 \|\theta\|_{\ell^1}^2 n^{-\delta_4} \mathcal{E}\end{aligned} \tag{78}$$

since $\mathcal{E} \geq \frac{1}{n_v}$ and $\frac{n-n_t}{n_t} \geq n^{-\delta_4}$, by hypothesis 4 of Theorem 1. From equations (77) and (78), it follows that, for some constant $\kappa(\|\theta\|_{1,\log^2})$,

$$\left| V_4 - \left(1 - \tfrac{1}{\sqrt{2}}\right) \frac{|I_k^1 \cap I_k^2|}{n_t\sqrt{2}} - \frac{1}{2n_t} \sum_{i \in I_k^1} \sum_{j \in I_k^2} \theta_{|i-j|}^2 \right| \leq \kappa(1+x)[\log n_t + y]^2 n^{-\frac{\delta_4}{4}} \mathcal{E}. \tag{79}$$

V_5 can be expressed as

$$V_5 = \left(1 - \frac{1}{\sqrt{2}}\right) \sum_{j \in I_k^1 \cap I_k^2} \left(\hat{\theta}_j^T - \theta_j\right)^2,$$

therefore by proposition 8.3, there exists an event $E_5(y)$ of probability greater than $1 - e^{-y}$ such that for all $D_n^T \in E_5(y)$,

$$V_5 = \left(1 - \frac{1}{\sqrt{2}}\right) \sum_{i \in I_k^1 \cap I_k^2} \frac{1}{n_t} \pm \left(1 - \frac{1}{\sqrt{2}}\right) \kappa_1 (b_x - a_x)[\log n + y]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})} \mathfrak{e}(n)$$

It follows by lemma 7.1 that on $E_5(y)$,

$$\left| V_5 - \left(1 - \frac{1}{\sqrt{2}}\right) \frac{|I_k^1 \cap I_k^2|}{n_t} \right| \leq 4\kappa_1(1+x)[\log n + y]^2 n^{-\min(\frac{1}{12}, \frac{\delta_4}{2})} \mathfrak{e}(n). \tag{80}$$

Finally, V_6 can be bounde in the following manner.

$$V_6 \leq \frac{1}{2}\sum_{i\in I_k^1}\sum_{j\in I_k^2}\left[(P^T-P)^2\psi_i+(P^T-P)^2\psi_j\right]\left[\frac{|\theta_{i+j}|}{\sqrt{2}}+|\theta_i||\theta_j|\right]$$
$$= \frac{1}{2}\sum_{i\in I_k^1}(P^T-P)^2\psi_i\sum_{j\in I_k^2}\left[\frac{|\theta_{i+j}|}{\sqrt{2}}+|\theta_i||\theta_j|\right]+\frac{1}{2}\sum_{j\in I_k^2}(P^T-P)^2\psi_j\sum_{i\in I_k^1}\left[\frac{|\theta_{i+j}|}{\sqrt{2}}+|\theta_i||\theta_j|\right].$$

Thus, by equation (71),

$$V_6 \leq \kappa\frac{(1+x)^{\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\times\left[\frac{1}{2}\sum_{i\in I_k^1}(P^T-P)^2\psi_i+\frac{1}{2}\sum_{j\in I_k^2}(P^T-P)^2\psi_j\right]$$
$$\leq \kappa\frac{(1+x)^{\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\times\sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\left(\hat{\theta}_j^T-\theta_j\right)^2. \tag{81}$$

By proposition 8.3, there exists an event $E_6(y)$ of probability greater than $1-e^{-y}$, such that for any $D_n^T\in E_6(y)$,

$$\sum_{j=k_*+a_x\Delta+1}^{k_*+b_x\Delta}\left(\hat{\theta}_j^T-\theta_j\right)^2 \leq (b_x-a_x)\mathcal{E}+\kappa_1(b_x-a_x)(y+\log n)^2 n^{-\min(\frac{1}{12},\frac{\delta_4}{2})}\mathfrak{e}$$
$$\leq 2\max(\kappa_1,1)(b_x-a_x)(y+\log n)^2\mathcal{E}$$
$$\leq 8\max(\kappa_1,1)(1+x)(y+\log n)^2\mathcal{E} \text{ by lemma 7.1.}$$

It follows by equation (81) that on $E_6(y)$, for a certain constant $\kappa(\kappa_1,\delta_1,c_1,\kappa_6)$,

$$V_6 \leq \kappa[y+\log n]^2\frac{(1+x)^{\frac{\delta_1}{2\delta_2}}}{n_t^{\frac{\delta_1}{3\delta_2}}}\mathcal{E}. \tag{82}$$

Combining equations (72), (75), (76), (79), (80), (82) on the event $\cap_{i=2}^6 E_i(\log 6+y)$ yields the result. ■

Lemme 8.7 *Let $s\in L^\infty([0;1])$ be a probability density function. For all $j\in\mathbb{N}$, let $\theta_j=\langle s,\psi_i\rangle$, where $\psi_0(x)=1$ and $\psi_j(x)=\sqrt{2}\cos(2j\pi x)$ for all $j\in\mathbb{N}^*$. Thus for any finite set $I\subset\mathbb{N}$ and for all functions $u\in\mathbb{R}^I$,*

$$0\leq\sum_{i\in I}\sum_{j\in I}u(i)u(j)\left(\frac{1-\delta_{i,j}}{\sqrt{2}}+\delta_{i,j}\right)\theta_{|i-j|}\leq\|s\|_\infty\sum_{i\in I}u(i)^2.$$

Proof Let $X\sim s$ be a random variable with distribution $s(x)dx$ on $[0;1]$. For any $x\in\mathbb{R}$, and any $i\neq j$,

$$\psi_i(x)\psi_j(x)=2\cos(2i\pi x)\cos(2j\pi x)=\cos(2(i+j)\pi x)+\cos(2(i-j)\pi x)=\frac{\psi_{i+j}+\psi_{|i-j|}}{\sqrt{2}}.$$

If $i \neq j$, then $\mathrm{Cov}(\psi_i(X), \psi_j(X)) = \frac{\theta_{i+j}+\theta_{|i-j|}}{\sqrt{2}} - \theta_i\theta_j$. If $i = j$, $\mathrm{Var}(\psi_i(X)) = 1 + \frac{\theta_{2i}}{\sqrt{2}} - \theta_i^2$. Let $u \in \mathbb{R}^I$, $k \in \mathbb{N}$ and $t_k = \sum_{i\in I} u(i)\psi_{i+k}$, then

$$\mathrm{Var}(t_k(X)) = \sum_{i\in I}\sum_{j\in I} u(i)u(j)\left[\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} + \frac{\theta_{i+j+2k}}{\sqrt{2}} - \theta_{i+k}\theta_{j+k}\right]$$

Furthermore, $\lim_{n\to+\infty}\theta_n = 0$, hence

$$\lim_{k\to+\infty}\mathrm{Var}(t_k(X)) = \sum_{i\in I}\sum_{j\in I} u(i)u(j)\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|}.$$

It immediately follows that $\sum_{i\in I}\sum_{j\in I} u(i)u(j)\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} \geq 0$. Moreover, for all $k \in \mathbb{N}$,

$$\begin{aligned}\mathrm{Var}(t_k(X)) &\leq \mathbb{E}\left[t_k(X)^2\right]\\ &= \int_0^1 t_k(x)^2 s(x)dx\\ &\leq \|s\|_\infty \|t_k\|^2\\ &\leq \|s\|_\infty \sum_{i\in I} u(i)^2.\end{aligned}$$

Thus

$$\sum_{i\in I}\sum_{j\in I} u(i)u(j)\left(\frac{1-\delta_{i,j}}{\sqrt{2}} + \delta_{i,j}\right)\theta_{|i-j|} \leq \|s\|_\infty \sum_{i\in I} u(i)^2.$$

■

Lemme 8.8 *Let $\varepsilon : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function such that $\varepsilon(0) > 0$ and $h_+ : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous, non-decreasing function. Let $g_0 : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function such that, for any $s < t$,*

$$-\varepsilon(\max(s,t)) \leq g_0(t) - g_0(s) \leq \max\{h_+(t) - h_+(s), \varepsilon(\max(s,t))\}.$$

Assume that $\varepsilon(0) > 0$. Then there exists a continuous, non-decreasing function $g : \mathbb{R}_+ \to \mathbb{R}_+$ such that $g_0(0) = g(0)$,

$$\left\|\frac{g_0 - g}{\varepsilon}\right\|_\infty \leq 6,$$

and moreover

$$\forall x, y, |g(y) - g(x)| \leq |h_+(y) - h_+(x)|.$$

Proof Assume to begin with that ε is right-continuous. Let $r > 0, \delta > 0$. We define by induction a sequence $(x_i)_{i\in\mathbb{N}}$ and a function g on $[x_i; x_{i+1}]$. Let $x_0 = 0$

and $g(x_0) = g_0(x_0)$. For any $i \in \mathbb{N}$, assuming x_i and $g(x_i)$ have been defined, let

$$x_{i+1} = \inf\left\{x \geq x_i : g_0(x) \geq g_0(x_i) + 2\varepsilon(x_i) \text{ or } \varepsilon(x) \geq \frac{3}{2}\varepsilon(x_i)\right\}$$

$$\forall x \in (x_i, x_{i+1}], g(x) = \begin{cases} g(x_i) \text{ if } \varepsilon(x_{i+1}) \geq \frac{3}{2}\varepsilon(x_i) \\ g(x_i) + \frac{g_0(x_{i+1}) - g_0(x_i)}{h_+(x_{i+1}) - h_+(x_i)}[h_+(x) - h_+(x_i)] \text{ else.} \end{cases} \tag{83}$$

If $x_{i+1} = +\infty$, the above definitions still make sense and the induction stops. Notice first that for any $x \in [x_i; x_{i+1})$, $g_0(x) - g_0(x_i) \leq [h_+(x) - h_+(x_i)] \vee \varepsilon(x) \leq [h_+(x) - h_+(x_i)] \vee \frac{3}{2}\varepsilon(x_i)$. Thus,by continuity of g_0,

$$g_0(x_{i+1}) - g_0(x_i) \leq [h_+(x_{i+1}) - h_+(x_i)] \vee \frac{3}{2}\varepsilon(x_i).$$

By assumption, ε is right-continuous, therefore if $\varepsilon(x_{i+1}) < \frac{3}{2}\varepsilon(x_i)$, it must be that $\inf\{x \geq x_i : \varepsilon(x) \geq \frac{3}{2}\varepsilon(x_i)\} > x_{i+1}$. Then by definition of x_{i+1} and continuity of g_0, $g_0(x_{i+1}) = g_0(x_i) + 2\varepsilon(x_i)$, therefore

$$2\varepsilon(x_i) = g_0(x_{i+1}) - g_0(x_i) \leq [h_+(x_{i+1}) - h_+(x_i)] \vee \frac{3}{2}\varepsilon(x_i),$$

which implies that

$$0 < 2\varepsilon(x_i) = g_0(x_{i+1}) - g_0(x_i) \leq [h_+(x_{i+1}) - h_+(x_i)]. \tag{84}$$

This proves that g is well defined. g is non-decreasing and continuous since h_+ has these properties. If $\varepsilon(x_{i+1}) < \frac{3}{2}\varepsilon(x_i)$, then the previous equation implies that

$$\forall i \in \mathbb{N}, \forall (x, y) \in (x_i, x_{i+1}], x \leq y \implies g(y) - g(x) \leq h_+(y) - h_+(x),$$

else g is constant on $]x_i; x_{i+1}]$ and the above equation is trivially true. Hence, since g, h_+ are non-decreasing and continuous,

$$\forall (x, y) \in \mathbb{R}, x \leq y \implies g(y) - g(x) \leq h_+(y) - h_+(x).$$

We will now prove by induction that for all $i \in \mathbb{N}^*$,

$$0 \leq g_0(x_i) - g(x_i) \leq 4\varepsilon(x_i). \tag{85}$$

Base case: This equation is true for $i = 1$ since $x_0 = 0$ and $g(0) = g_0(0) = 0$, therefore by definition of g, x_1, $0 \leq g(x_1) \leq g_0(x_1) \leq 2\varepsilon(x_0)$.

Inductive step: Assume that equation (85) is true for some $i \in \mathbb{N}$. Then by definition of x_{i+1} and g,

- If $\varepsilon(x_{i+1}) \geq \frac{3}{2}\varepsilon(x_i)$, then $g(x_{i+1}) = g(x_i)$ therefore $g_0(x_{i+1}) - g(x_{i+1}) = g_0(x_{i+1}) - g_0(x_i) + g_0(x_i) - g(x_i)$. By the induction hypothesis and the definition of x_{i+1},

  $$0 \leq g_0(x_{i+1}) - g(x_{i+1}) \leq 2\varepsilon(x_i) + 4\varepsilon(x_i) \leq 6\varepsilon(x_i) \leq 4\varepsilon(x_{i+1}),$$

  which proves equation (85) for $i + 1$.

- Otherwise, by definition of g, $g(x_{i+1}) = g(x_i) + [g_0(x_{i+1}) - g_0(x_i)]$ therefore by the induction hypothesis and since ε is non-decreasing,

$$0 \leq g_0(x_{i+1}) - g(x_{i+1}) = g_0(x_i) - g(x_i) \leq 4\varepsilon(x_i) \leq 4\varepsilon(x_{i+1}).$$

  This proves equation (85) for $i+1$.

By induction, equation (85) is therefore true for all $i \in \mathbb{N}$ (such that $x_i < +\infty$). Let now $i \in \mathbb{N}$ and $x \in (x_i, x_{i+1}]$. By definition of g,

$$g(x_i) \leq g(x) \leq g(x_i) + (g_0(x_{i+1}) - g_0(x_i))_+.$$

By equation (85) and definition of x_{i+1},

$$\begin{aligned} g(x) - g_0(x) &\leq g(x) - g_0(x_i) \\ &\leq g(x_i) - g_0(x_i) + (g_0(x_{i+1}) - g_0(x_i))_+ \\ &\leq 2\varepsilon(x_i) \\ &\leq 2\varepsilon(x). \end{aligned}$$

Moreover, by equation (85) and definition of the x_i,

$$\begin{aligned} g(x) - g_0(x) &\geq g(x_i) - g_0(x_{i+1}) \\ &\geq g(x_i) - g_0(x_i) - [g_0(x_{i+1}) - g_0(x_i)] \\ &\geq -4\varepsilon(x_i) - 2\varepsilon(x_i) \\ &\geq -6\varepsilon(x_i) \\ &\geq -6\varepsilon(x). \end{aligned}$$

It has been proved that for all $i \in \mathbb{N}$ such that x_i is finite,

$$\forall x \in (x_i; x_{i+1}], |g(x) - g_0(x)| \leq 6\varepsilon(x).$$

It must now be proved that $\lim_{n\to+\infty} x_n = +\infty$. Since ε is non-decreasing and right-continuous, by definition of x_n, $g_0(x_{n+1}) \geq g_0(x_n) + 2\varepsilon(x_n) \geq g_0(x_n) + \varepsilon(0)$ or $\varepsilon(x_{n+1}) \geq \frac{3}{2}\varepsilon(x_n)$. Since $\varepsilon(0) > 0$ by assumption, this implies that $\max(g_0, \varepsilon)(x_n) \to +\infty$. The function $\max(g_0, \varepsilon)$ is non-decreasing, thus it is bounded on every interval of the form $[0; x]$, which implies that $x_n \to +\infty$. This proves the proposition under the assumption that ε is right-continuous.

In the general case, let $\varepsilon_+ : x \mapsto \inf_{y>x} \varepsilon(y)$, which is non-decreasing and right-continuous. Since ε is non-decreasing, $\varepsilon_+ \geq \varepsilon$, therefore the assumptions of the proposition hold with ε_+ instead of ε. By the right-continuous case of the proposition, which we already proved, there exists a non-decreasing function g such that $\left\|\frac{g-g_0}{\varepsilon_+}\right\|_\infty \leq 6$ and

$$\forall x, y, x \leq y \implies g(y) - g(x) \leq h_+(y) - h_+(x).$$

Let $x \in \mathbb{R}_+$. For all $y < x$, $|g(y) - g_0(y)| \leq 6\varepsilon_+(y)$, therefore by continuity of g, g_0,

$$|g(x) - g_0(x)| \leq 6 \sup_{y<x} \varepsilon_+(y) = 6 \sup_{y<x} \inf_{y'>y} \varepsilon(y') \leq 6\varepsilon(x).$$

This proves the proposition in the general case.

■

Proposition 8.9 *Let $([x_i, x_{i+1}))_{1\leq i\leq M-1}$ be a partition of the interval $[a, b)$. Let $Y : \{x_1, \ldots, x_M\} \to \mathbb{R}$ be such that $(Y(x_j))_{1\leq j\leq M}$ is a zero-mean gaussian vector. Abusing notation, we also denote by Y the extension of Y to $[a; b]$ by linear interpolation. Let $K_Y : [a; b]^2 \to \mathbb{R}$ be the variance-covariance function of Y. Let $h : [a; b] \to \mathbb{R}$ be a continuous, increasing function and let $K_X : [a; b]^2 \to \mathbb{R}$ be a positive semi-definite function such that:*

$$\forall (s, t) \in [a; b]^2,\ |K_X(s, s) + K_X(t, t) - 2K_X(s, t)| \leq |h(s) - h(t)|.$$

Assume that there exists constants $L > 0$ and $\varepsilon \in [0; 1]$ such that:

- $\sup_{t\in[a;b]} \sqrt{K_X(t, t)} \leq L$
- *For any $i \in \{1, \ldots, M - 1\}, h(x_{i+1}) - h(x_i) \leq \varepsilon$*
- $\max_{(i,j)\in\{1,\ldots,M\}^2} |K_X(x_i, x_j) - K_Y(x_i, x_j)| \leq \varepsilon$.

There exists a universal constant κ and a measurable function $f : C([a; b], \mathbb{R}) \to C([a; b], \mathbb{R})$ such that for all random variables $\nu \sim \mathcal{U}([0; 1])$ independent from Y, $X = f(Y, \nu)$ is a zero-mean gaussian process with variance-covariance function K_X and moreover,

$$\mathbb{E}\left[\sup_{a\leq t\leq b} |X_t - Y_t|\right] \leq \kappa\sqrt{(1 + L) \log M\, [(h(b) - h(a)) \vee 1]}\varepsilon^{\frac{1}{12}}.$$

Proof We assume without loss of generality that $h(b) - h(a) \geq 1$. We shall moreover use the following notation. For A, B two symmetric matrices, $A \prec B$ means that $B - A$ is positive definite. $\|A\|_{op}$ denotes the matrix operator norm corresponding to the euclidean norm, i.e $\|A\|_{op} = \sup_{x:\|x\|_2\leq 1} \|Ax\|_2$. We will need the following lemmas:

Lemme 8.10 *For all $A \in \mathbb{R}^{m\times m}$, $\|A\|_{op} \leq m \max_{1\leq i,j\leq m} |A_{i,j}|$.*

Proof Let $v \in \mathbb{R}^m$ be such that $\sum_{i=1}^m v_i^2 = 1$. By the Cauchy-Schwartz inequality,

$$\|Av\|^2 = \sum_{i=1}^m \left(\sum_{j=1}^m A_{i,j}v_j\right)^2 \leq \sum_{i=1}^m \sum_{j=1}^m A_{i,j}^2 \leq m^2 \max_{i=1,\ldots,m} A_{i,j}^2.$$

This is true for any v, which proves lemma 8.10. ■

Lemma 8.11 below is a special case of Mc-Carthy's trace inequality [18, Lemma 2.6].

Lemme 8.11 *Let A, B be two symmetric, positive semi-definite matrices, then*

$$Tr(\sqrt{A + B}) \leq Tr(\sqrt{A}) + Tr(\sqrt{B}).$$

The hypotheses imply that h is bijective from $[a; b]$ to $[h(a); h(b)]$. Let $m \in \mathbb{N}$. For all $j \in \{1, \ldots, m\}$, let

$$t_j = \max\left\{x_i | i \in \{1, \ldots, M\}, h(x_i) \leq h(a) + \frac{j - 1}{m - 1}[h(b) - h(a)]\right\}. \qquad (86)$$

Let $K_{X,m} = \big(K_X(t_i,t_j)\big)_{1\le i,j\le m}$ and $K_{Y,m} = \big(K_Y(t_i,t_j)\big)_{1\le i,j\le m}$. The Wasserstein distance between two gaussian vectors is known [20]: there exists a coupling $\tilde{X}^m, \tilde{Y}^m$ of the distributions $\mathcal{N}(0, K_{X,m})$ and $\mathcal{N}(0, K_{Y,m})$ such that:

$$\mathbb{E}\left[\sum_{i=1}^m (\tilde{X}_i^m - \tilde{Y}_i^m)^2\right] = \mathrm{Tr}\left(K_{X,m} + K_{Y,m} - 2\big(K_{X,m}^{\frac{1}{2}} K_{Y,m} K_{X,m}^{\frac{1}{2}}\big)^{\frac{1}{2}}\right).$$

Thus

$$\begin{aligned} K_{X,m}^2 &= K_{X,m}^{\frac{1}{2}} K_{Y,m} K_{X,m}^{\frac{1}{2}} + K_{X,m}^{\frac{1}{2}}\big(K_{Y,m} - K_{X,m}\big) K_{X,m}^{\frac{1}{2}} \\ &\prec K_{X,m}^{\frac{1}{2}} K_{Y,m} K_{X,m}^{\frac{1}{2}} + \|K_{Y,m} - K_{X,m}\|_{op} K_{X,m}. \end{aligned}$$

By lemma 8.11,

$$\mathrm{Tr}(K_{X,m}) \le \mathrm{Tr}\left(\big(K_{X,m}^{\frac{1}{2}} K_{Y,m} K_{X,m}^{\frac{1}{2}}\big)^{\frac{1}{2}}\right) + \|K_{Y,m} - K_{X,m}\|_{op}^{\frac{1}{2}} \mathrm{Tr}\left(K_{X,m}^{\frac{1}{2}}\right).$$

By the same argument (exchangeing X and Y),

$$\mathrm{Tr}(K_{Y,m}) \le \mathrm{Tr}\left(\big(K_{Y,m}^{\frac{1}{2}} K_{X,m} K_{Y,m}^{\frac{1}{2}}\big)^{\frac{1}{2}}\right) + \|K_{Y,m} - K_{X,m}\|_{op}^{\frac{1}{2}} \mathrm{Tr}\left(K_{Y,m}^{\frac{1}{2}}\right).$$

It follows that

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^m (\tilde{X}_i^m - \tilde{Y}_i^m)^2\right] &\le \|K_{Y,m} - K_{X,m}\|_{op}^{\frac{1}{2}} \mathrm{Tr}\left(K_{X,m}^{\frac{1}{2}} + K_{Y,m}^{\frac{1}{2}}\right) \\ &\le m^{\frac{1}{2}} \|K_{Y,m} - K_{X,m}\|_{\infty}^{\frac{1}{2}} \left(\sqrt{\mathrm{Tr}(K_{X,m})} + \sqrt{\mathrm{Tr}(K_{Y,m})}\right) \\ &\le m^{\frac{1}{2}} \|K_{Y,m} - K_{X,m}\|_{\infty}^{\frac{1}{2}} m^{\frac{1}{2}} \max_{1\le i,j\le m} \left\{\sqrt{K_X(t_i,t_i)} + \sqrt{K_Y(t_j,t_j)}\right\} \\ &\le 2m\sqrt{\varepsilon}\sqrt{L^2+\varepsilon}. \end{aligned} \tag{87}$$

By the transfer principle (Kallenberg, Theorem 5.10), there exists f_1 such that for all uniform random variables variables ν_1 independent from Y, $(\tilde{X}^m, \tilde{Y}^m) \sim (f_1(Y^m, \nu_1), Y^m)$.

Let X_0 be a gaussian process with variance-covariance function K_X. Let $W_0 = X_0 \circ h^{-1}$. For any $(s,t) \in [h(a); h(b)]^2$, $W_0(t) - W_0(s)$ is a centred gaussian random variable, hence for all $r > 0$, there exists a universal constant $C(r)$ such that

$$\begin{aligned} \mathbb{E}\left[(W_0(t) - W_0(s))^r\right] &\le C(r)\mathbb{E}\left[(W_0(t) - W_0(s))^2\right]^{\frac{r}{2}} \\ &\le C(r)\Big(K_X(h^{-1}(t), h^{-1}(t)) + K_X(h^{-1}(s), h^{-1}(s)) \\ &\qquad - 2K_X(h^{-1}(s), h^{-1}(t))\Big)^{\frac{r}{2}} \\ &\le C(r)|t-s|^{\frac{r}{2}}. \end{aligned}$$

By the Kolmogorov continuity theorem [21, Chapter 1, Theorem 2.1], applied to $x \mapsto \frac{W_0(h(a)+(h(b)-h(a))x)}{\sqrt{h(b)-h(a)}}$, there exists a continuous version W_1 of W_0 such that for any $\theta \in [0;1)$ and all $(s,t) \in [h(a); h(b)]^2$,

$$\mathbb{E}\left[\left(\sup_{s\neq t} \frac{|W_1(t) - W_1(s)|}{|t-s|^{\theta\left(\frac{1}{2}-\frac{1}{r}\right)}}\right)^r\right] \le \frac{(h(b)-h(a))^{\frac{r}{2}}}{(h(b)-h(a))^{\theta r\left(\frac{1}{2}-\frac{1}{r}\right)}} B(\theta, r) < +\infty,$$

where $B(\theta, r)$ is a universal constant. Let $X_1 = W_1 \circ h$, which is still a gaussian process, with variance-covariance function K_X. Then, since by assumption $h(b) - h(a) \geq 1$, for all $r \geq 2$,

$$\mathbb{E}\left[\left(\sup_{s\neq t} \frac{|X_1(t) - X_1(s)|}{|h(t) - h(s)|^{\theta\left(\frac{1}{2}-\frac{1}{r}\right)}}\right)^r\right] \leq [h(b) - h(a)]^{\frac{r}{2}} B(\theta, r) < +\infty. \tag{88}$$

The $C([a;b], \mathbb{R})$-valued process X_1 induces a probability distribution Q on the Borel space $C([a;b], \mathbb{R})$. Furthermore, $(X_1(t_j))_{1\leq j\leq m} \sim \tilde{X}^m \sim f_1(Y^m, \nu_1)$. By (Kallenberg, Theorem 5.10), there exists a measurable function f_2 such that for all uniform random variables ν_3 independent from Y^m, ν_1, $(X_1, (X_1(t_j))_{1\leq j\leq m}) \sim (f_2(f_1(Y^m, \nu_1), \nu_3), f_1(Y^m, \nu_1))$. Let $X = f_2(f_1(Y^m, \nu_1), \nu_3)$ and $X^m = (X(t_j))_{1\leq j\leq m}$. Almost surely,

Claim 8 *1. $X^m = (X(t_j))_{1\leq j\leq m} = f_1(Y^m, \nu_1)$ p.s, so*

2. $(X^m, Y^m) \sim (\widetilde{X}^m, \widetilde{Y}^m)$, in particular by equation (87),

$$\mathbb{E}\left[\|X^m - Y^m\|^2\right] \leq 2m\sqrt{\varepsilon}\sqrt{L^2 + \varepsilon}.$$

3. $X \sim X^1$ as a random continuous function, in particular by equation (88) with $\theta = \frac{3}{4}$ and $r = 6$,

$$\forall \delta > 0, \mathbb{E}\left[\sup_{(s,t)\in[a;b]:|h(t)-h(s)|\leq\delta} |X(t) - X(s)|^6\right]^{\frac{1}{6}} \leq \sqrt{h(b) - h(a)} B(\tfrac{3}{4}, 6)^{\frac{1}{6}} \delta^{\frac{1}{4}}. \tag{89}$$

By abuse of notation, denote X^m, Y^m the random processes obtained by linear interpolation between the points (t_j, X_j^m) and (t_j, Y_j^m), respectively. For all $t \in [a;b]$, there exists $j \in \{1, \ldots, m\}$ such that $t_j \leq t \leq t_{j+1}$, therefore $|h(t) - h(t_j)| \leq h(t_{j+1}) - h(t_j)$, since h is non-decreasing. By definition of t_{j+1} (equation (86)), $h(t_{j+1}) \leq h(a) + \frac{j}{m-1}(h(b) - h(a))$ and furthermore, there exists $i \in \{1, \ldots, M\}$ such that $x_i = t_j$. By equation (86) which defines t_j, $h(x_{i+1}) > h(a) + \frac{j-1}{m-1}(h(b) - h(a))$, which yields

$$\begin{aligned} |h(t) - h(t_j)| &\leq h(a) + \frac{j}{m-1}(h(b) - h(a)) - h(x_{i+1}) + h(x_{i+1}) - h(x_i) \\ &\leq \frac{h(b) - h(a)}{m-1} + h(x_{i+1}) - h(x_i). \end{aligned}$$

By assumption, $h(x_{i+1}) - h(x_i) \leq \varepsilon$, which yields

$$\forall t \in [a;b], \exists j(t) \in \{1, \ldots, m\}, t_j \leq t \leq t_{j+1} \text{ and } |h(t) - h(t_{j(t)})| \leq \frac{h(b) - h(a)}{m} + \varepsilon. \tag{90}$$

Since Y is piecewise linear on the partition $([x_i, x_{i+1}))_{1\le i\le M-1}$,

$$\begin{aligned}\sup_{a\le t\le b}|X(t)-Y(t)| &\le \sup_{t\in[a;b]}|X(t)-X(t_{j(t)})| + \max_{j\in\{1,\dots,m\}}|X(t_j)-Y(t_j)| \\ &\quad + \sup_{t\in[a;b]}|Y(t)-Y(t_{j(t)})| \\ &\le \sup_{(s,t):|h(s)-h(t)|\le\varepsilon+\frac{h(b)-h(a)}{m}}|X(s)-X(t)| + \sqrt{\sum_{j=1}^m |X(t_j)-Y(t_j)|^2} \\ &\quad + \max_{j\in\{1,\dots,m\}}\max\left\{|Y(x_i)-Y(t_j)| : i\in\{1,\dots,M\}\cap[t_j;t_{j+1}]\right\}.\end{aligned}$$

Thus, by claim 8,

$$\begin{aligned}&\mathbb{E}\left[\sup_{a\le t\le b}|X(t)-Y(t)|\right] \\ &\le \mathbb{E}\Big[\sup_{(s,t):|h(s)-h(t)|\le\varepsilon+\frac{h(b)-h(a)}{m}}|X(s)-X(t)|\Big] + \mathbb{E}\Big[\sum_{i=1}^m (X_i^m - Y_i^m)^2\Big]^{\frac12} \\ &\quad + \mathbb{E}\left[\max_{j\in\{1,\dots,m\}}\max\left\{|Y(x_i)-Y(t_j)| : i\in\{1,\dots,M\}\cap[t_j;t_{j+1}]\right\}\right] \\ &\le \sqrt{h(b)-h(a)}B(\tfrac34,6)^{\frac16}\left(\varepsilon+\frac{h(b)-h(a)}{m}\right)^{\frac14} + \left(2m\sqrt{\varepsilon}\sqrt{L^2+\varepsilon}\right)^{\frac12} \\ &\quad + \sqrt{2\log M}\max_{j\in\{1,\dots,m\}}\max\left\{\sqrt{\mathbb{E}[|Y(x_i)-Y(t_j)|^2]} : i\in\{1,\dots,M\}\cap[t_j;t_{j+1}]\right\}.\end{aligned} \tag{91}$$

Furthermore, for any $(i,j)\in[1;M]^2$,

$$\begin{aligned}\mathbb{E}[(Y(x_i)-Y(x_j))^2] &= K_Y(x_i,x_i)+K_Y(x_j,x_j)-2K_Y(x_i,x_j) \\ &\le K_X(x_i,x_i)+K_X(x_j,x_j)-2K_X(x_i,x_j) \\ &\quad + 4\max_{(r,s)\in[1;M]^2}|K_X(x_r,x_s)-K_Y(x_r,x_s)| \\ &\le |h(x_i)-h(x_j)| + 4\varepsilon.\end{aligned}$$

Setting $\kappa = B(\frac34,6)^{\frac16}$, it follows by equation (91) and the non-decreasing nature of h that

$$\begin{aligned}\mathbb{E}\left[\sup_{a\le t\le b}|X_t-Y_t|\right] &\le \sqrt{h(b)-h(a)}B(\tfrac34,6)^{\frac16}\left(\varepsilon+\frac{h(b)-h(a)}{m}\right)^{\frac14} + \sqrt{2m(L+1)}\varepsilon^{\frac14} \\ &\quad + \sqrt{2\log M}\sqrt{h(t_{j+1})-h(t_j)+4\varepsilon} \\ &\le \kappa\sqrt{h(b)-h(a)}\varepsilon^{\frac14} + \kappa\frac{[h(b)-h(a)]^{\frac34}}{m^{\frac14}} + \sqrt{2m(L+1)}\varepsilon^{\frac14} \\ &\quad + \sqrt{2\log M}\sqrt{\frac{h(b)-h(a)}{m}+5\varepsilon} \text{ by equation (90).}\end{aligned}$$

Let now $m = \left\lceil \frac{h(b)-h(a)}{\varepsilon^{\frac13}}\right\rceil$. Since by assumption $\varepsilon\le 1$, $h(b)-h(a)\ge 1$ it follows finally, by keeping only the largest powers of $[h(b)-h(a)], \varepsilon, L$ and $\log M$, that

$$\mathbb{E}\left[\sup_{a\le t\le b}|X_t-Y_t|\right] \le \kappa\sqrt{h(b)-h(a)}\sqrt{2(L+1)\log M}\varepsilon^{\frac{1}{12}}.$$

for an absolute constant κ. ■

References

[1] *Filtering and Asymptotics*, pages 259–322. Springer New York, New York, NY, 1999. ISBN 978-0-387-22638-5. doi: 10.1007/978-0-387-22638-5_7. URL https://doi.org/10.1007/978-0-387-22638-5_7.

[2] Sylvain Arlot. V-fold cross-validation improved: V-fold penalization. 40 pages, plus a separate technical appendix., February 2008. URL https://hal.archives-ouvertes.fr/hal-00239182.

[3] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4:40–79, 2010. doi: 10.1214/09-SS054. URL https://doi.org/10.1214/09-SS054.

[4] Sylvain Arlot and Matthieu Lerasle. Choice of V for V-fold cross-validation in least-squares density estimation. *Journal of Machine Learning Research (JMLR)*, 17(208):1–50, 2016.

[5] Morgane Austern and Wenda Zhou. Asymptotics of cross-validation. *arXiv e-prints*, art. arXiv:2001.11111, January 2020.

[6] Pierre Bayle, Alexandre Bayle, Lucas Janson, and Lester Mackey. Cross-validation confidence intervals for test error. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 16339–16350. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/bce9abf229ffd7e570818476ee5d7dde-Paper.

[7] Philippe Berthet and David Mason. Revisiting two strong approximation results of dudley and philipp. 51, 01 2007. doi: 10.1214/074921706000000824.

[8] Prabir Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514, 1989. ISSN 00063444. URL http://www.jstor.org/stable/2336116.

[9] Alain Célisse and Stephane Robin. Nonparametric density estimation by exact leave-p-out cross-validation. *Computational Statistics and Data Analysis*, 52(5):2250–2368, 2008. doi: 10.1016/j.csda.2007.10.002. URL https://hal.archives-ouvertes.fr/hal-01197590.

[10] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximation of suprema of empirical processes. *The Annals of Statistics*, 42(4):1564 – 1597, 2014. doi: 10.1214/14-AOS1230. URL https://doi.org/10.1214/14-AOS1230.

[11] Sandrine Dudoit and Mark J. van der Laan. Asymptotics of cross-validated risk estimation in estimator selection and performance

assessment. *Statistical Methodology*, 2(2):131–154, 2005. ISSN 1572-3127. doi: https://doi.org/10.1016/j.stamet.2005.02.003. URL https://www.sciencedirect.com/science/article/pii/S1572312705000158.

[12] Peter Hall. Laws of the iterated logarithm for nonparametric density estimators. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 56(1):47–61, Mar 1981. ISSN 1432-2064. doi: 10.1007/BF00531973. URL https://doi.org/10.1007/BF00531973.

[13] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *Basis Expansions and Regularization*, pages 139–189. Springer New York, New York, NY, 2009. ISBN 978-0-387-84858-7. doi: 10.1007/978-0-387-84858-7_5. URL https://doi.org/10.1007/978-0-387-84858-7_5.

[14] Vladimir I. Koltchinskii. Komlos-major-tusnady approximation for the general empirical process and haar expansions of classes of functions. *Journal of Theoretical Probability*, 7(1):73–118, Jan 1994. ISSN 1572-9230. doi: 10.1007/BF02213361. URL https://doi.org/10.1007/BF02213361.

[15] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv'-s, and the sample df. i. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(1):111–131, Mar 1975. ISSN 1432-2064. doi: 10.1007/BF00533093. URL https://doi.org/10.1007/BF00533093.

[16] Matthieu Lerasle. Optimal model selection for stationary data under various mixing conditions. *Annals of Statistics - ANN STATIST*, 39, 11 2009. doi: 10.1214/11-AOS888.

[17] Guillaume Maillard, Sylvain Arlot, and Matthieu Lerasle. Aggregated Hold-Out. working paper or preprint, September 2019. URL https://hal.archives-ouvertes.fr/hal-02273193.

[18] Charles A. McCarthy. Cp. *Israel Journal of Mathematics*, 5(4): 249–271, 1967. ISSN 1565-8511. doi: 10.1007/BF02771613. URL https://doi.org/10.1007/BF02771613.

[19] Fabien Navarro and Adrien Saumard. Slope heuristics and v-fold model selection in heteroscedastic regression using strongly localized bases. *ESAIM: Probability and Statistics*, 21:412–451, 2017. ISSN 1262-3318. doi: 10.1051/ps/2017005. URL http://dx.doi.org/10.1051/ps/2017005.

[20] I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257 – 263, 1982. ISSN 0024-3795. doi: https://doi.org/10.1016/0024-3795(82)90112-4. URL http://www.sciencedirect.com/science/article/pii/0024379582901124.

[21] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*. Springer Berlin Heidelberg, 1999. doi: 10.1007/978-3-662-06400-9. URL https://doi.org/10.1007/978-3-662-06400-9.

[22] Emmanuel Rio. Local invariance principles and their application to density estimation. *Probability Theory and Related Fields*, 98(1):21–45, Mar 1994. ISSN 1432-2064. doi: 10.1007/BF01311347. URL https://doi.org/10.1007/BF01311347.

[23] Aad W. van der Vaart, Sandrine Dudoit, and Mark J. van der Laan. Oracle inequalities for multi-fold cross validation. *Statist. Decisions*, 24(3):351–371, 2006. ISSN 0721-2631.