Minimax Estimation of Linear Functionals Under Squared Error Loss

Meng Zhao
and
K. B. Kulasekera
Department of Mathematical Sciences
Clemson University
Clemson, SC 29634-0975

March 8, 2005

Abstract

Under a very general setting, we consider the problem of estimating a linear functional of an unknown vector in a Hilbert space from indirect data contaminated by noise. We give formulas for the minimax affine risk of the estimation showing that for parameter spaces satisfying certain conditions, the minimax affine risk is the supremum of the minimax affine risks for one-dimensional subproblems. We then apply these results to two situations: estimating $f^{(k)}(t_0)$ for a given t_0 in the fractional Brownian motion model and the regression model with correlated errors. The fractional Brownian motion model has the form $y(t) = \int_0^t f(s)\,ds + \sigma Z_t$ where f belongs a convex set of functions $\mathcal{F}$, and Z_t is a fractional Brownian motion with index $H \in (1/2, 1)$. The regression model with correlated errors is defined as $y(t_i) = f(t_i) + z_i$, $i = 1, \ldots, n$ where f is as above and z_i are mean zero random errors satisfying $Cov(z_i, z_j) \sim C_1 |j-i|^{-\alpha}$, $|j-i| \to \infty$. For both estimation problems, we obtain the asymptotic rate for the minimax affine risks over certain types of parameter spaces. In each case, we also show that the minimax affine risk is bounded by 1.25 times the minimax risk.

1 Introduction

Suppose that we observe data of the form

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{Z} \tag{1}$$

where $\mathbf{x}$ belongs to a convex subset $\mathcal{F}$ of a separable real Hilbert space $\mathbf{X}$ and $\mathbf{K}$ is a linear operator from $\mathcal{F}$ to $\mathbf{U}$, another separable real Hilbert space. $\mathbf{Z}$ is a bounded linear operator from $\mathbf{U}$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$, the space of all random variables define on the probability space $(\Omega, \mathscr{F}, P)$ that have finite variance. It is assumed that $\mathbf{Z}$ is invertible (the inverse might be unbounded) and $\mathbf{Zw}$ has mean zero for $\mathbf{w} \in \mathbf{U}$. By defining $\mathbf{y} = \mathbf{Kx} + \mathbf{Z}$, we are treating $\mathbf{Kx}$ as an operator from $\mathbf{U}$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. This is justified since $\mathbf{Kx}$ defines a functional on $\mathbf{U}$, and real numbers can be treated as constant random variables in $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. Suppose that L is a real affine functional on $\mathbf{X}$, that is $L = L_1 + l$ where L_1 is a linear functional on $\mathbf{X}$ and l is a constant. We consider the problem of estimating the value of L at some $\mathbf{x} \in \mathcal{F}$ by affine estimators of the form

$$\hat{L}(\mathbf{w}, d) = \mathbf{yw} + d = \langle \mathbf{w}, \mathbf{Kx} \rangle + \mathbf{Zw} + d, \tag{2}$$

where $\mathbf{w}$ is in $\mathbf{U}$. We will evaluate the performance of an affine estimator by the mean squared error. We want to obtain the minimax affine risk for estimating L

$$\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}; L, K\right),$$

where

$$R_{\mathcal{A}}\left(\hat{L}; L, K\right) = \sup_{\mathbf{x} \in \mathcal{A}} E_{\mathbf{x}}\left(\hat{L} - L(\mathbf{x})\right)^2.$$

Two special cases of model (1) are the white noise model

$$Y(t) = \int_{-1/2}^{t} f(s)\, ds + \sigma W(t), \qquad -1/2 \le t \le 1/2, \tag{3}$$

where $W(t)$ is a standard Brownian motion, and the regression model

$$y_i = f(t_i) + \sigma z_i, \qquad i = 1, \ldots, n; t_i \in [-1/2, 1/2] \tag{4}$$

where z_i's are i.i.d. noises and $f \in \mathcal{F}$, where $\mathcal{F}$ is a convex class of functions in each model. In Ibragimov and Khasminskii (1984) the minimax linear risk for the white noise model (3) was given together with the minimax linear estimator for a hypothesis set that is symmetric. A relationship between the minimax linear risk and the minimax risk was also established. Donoho and Liu (1991) removed the symmetry constraint on the set $\mathcal{F}$ and established the minimax affine risk and rate of convergence for the asymptotic minimax risk in the white noise model. In their work, the minimax affine risk was expressed in term of modulus of continuity. An interesting result of Donoho and Liu (1991) is that the minimax affine risk for the full problem is just the minimax affine risk of the hardest one-dimensional subproblem. Using this result, it was readily shown that the ratio of the minimax affine risk to the minimax risk is bounded by 1.25. Donoho and Liu (1991) then applied their result on the white noise model to regression data with independent errors. Donoho (1994) showed the same results for a generalization of model (3) in which the data $\mathbf{y}$ have the form

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{z}$$

where $\mathbf{x}$ is from a convex subset $\mathbf{X}$ of l_2, the space of all square summable sequences, $\mathbf{K}$ is a linear operator and $\mathbf{z}$ is a noise vector. That generalization was in two senses: first, the data are not observed directly, but through the operator $\mathbf{K}$. Second, the Gaussian process $\mathbf{z}$ can be non-white noise. However, it was still required that the covariance has a bounded inverse. Note that this is not the case for fractional Brownian motion, which is a special cases of the model (1) above. Fan (1993) discussed the estimation of a regression function in a slightly different framework, and it was shown that under some restrictive conditions the local linear smoother, with a proper choice of kernel and bandwidth, is near minimax. By applying Ibragimov and Khasminskii (1984)'s and Donoho and Liu (1991)'s results, Zhao (1997) gave the exact linear minimax estimator of $f(0)$ under the white noise model with f known to be in the closure of the class $\{f : |f''(t)| \leq B\}$, and by comparing the kernel of this linear minimax estimator with the Epanechinikov kernel, it was shown that the Epanechinikov kernel is 99% efficient. This confirmed Ylvisaker's conjecture (see Sacks and Ylvisaker (1981)), which states that the Epanechinikov kernel is nearly minimax.

Several authors have considered the estimation of the function f itself, a different problem than estimating a linear functional of f, in the regression model. Among them are Donoho and Johnstone (1998), Wang (1996,1997), and Johnstone (1999). In these papers, the performance of an estimator was evaluated by the expectation of the L_2 distance between the estimator and f. In Donoho and Johnstone (1998), it was shown that in the case of estimating f, the linear estimator no longer has near minimax risk. The authors proposed the use of a wavelet transformation to convert the function space $\mathcal{F}$ into a sequence space, and an estimator for f was then obtained in the wavelet domain by simple nonlinear shrinkage of the empirical wavelet coefficients. Besides showing the near minimaxity of the wavelet shrinkage estimator over a wide range of Triebel and Besov-type (Donoho and Johnstone, 1998) smoothness constraints and asymptotic minimaxity over certain Besov bodies, the authors also showed that more practical simple threshold nonlinear estimators are nearly minimax. This work of estimating f was extended by Wang (1996) and Wang (1997) for the case when long-range dependency appears in the data. In these works, Wang worked with the long-memory counterparts of models (3) and (4) – the fractional Gaussian noise model and the regression model with dependent errors. In the fractional Gaussian noise model, the white noise in model (3) was replaced by fractional Gaussian noise. In Wang (1996), it was shown that with proper scaling, the fractional Gaussian noise model is an approximation to the nonparametric regression model with correlated errors. Wavelet estimates with proper choices of thresholds were shown to achieve minimax rates over a wide range of function spaces. Wang (1997) extended the results of Wang (1996) to the case in which the data are indirect. Johnstone (1999) discussed the choice of the threshold in the wavelet estimators in Wang (1996) that adapts to a broad range of Besov classes. He also proposed an extension to the case of indirect data similar to (and independently from) Wang (1997).

Deo (1997) discussed the estimation of a linear functional for data with long-memory errors. A kernel estimator was studied and, under some rather restrictive conditions, the asymptotic normality of the estimator was shown. We will show that their result coincides with our lower bound on the asymptotic rate for the minimax risk.

Other generalizations of the white noise model include Cai and Low

(2003) and Cai and Low (2004). Based on the results of Donoho (1994), Cai and Low (2003) gave precise asymptotic descriptions of the minimax affine risks and bias variance trade-offs for estimating linear functionals for what the authors called regular modulus. Cai and Low (2004) extended the minimax theory for estimating linear functionals to the case of a finite union of convex parameter spaces. In this extension, an interesting contrast to the case of convex parameter spaces is that linear estimators no longer have optimal rates of convergence.

In this article, we extend the technique of Donoho and Liu (1991) and prove that, in our generalized setting (1),

$$\inf_{\hat{L}\text{ affine}} R_{\mathcal{F}}\left(\hat{L}; L, \mathbf{K}\right) = \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}} \inf_{\hat{L}\text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}; L, \mathbf{K}\right) \tag{5}$$

where $[\mathbf{x}_1, \mathbf{x}_2] = \{\mathbf{x} : \mathbf{x} = \alpha\mathbf{x}_1 + \beta\mathbf{x}_2, 0 \leq \alpha \leq 1, \beta = 1 - \alpha\}$ is the convex hull of $\{\mathbf{x}_1, \mathbf{x}_2\}$. This is the same as saying that the minimax affine risk is the supremum of the minimax affine risks for one-dimenssional subproblems. To our knowledge, this is the most generalized version of the model for convex parameter spaces that has been considered and our results reduce to all existing results including those of Donoho and Liu (1991) and Donoho (1994).

By applying this result to the fractional Gaussian noise model and the nonparametric regression model with correlated data, we give the rate of convergence for the asymptotic minimax affine risk of estimating the value of a function $f^{(k)}$ at a fixed point t_0. Under both these settings, we also show that the ratio of the minimax affine risk to the minimax risk is bounded above by 1.25. Notice that although Wang (1996), Wang (1997) and Johnstone (1999) discussed estimation with long-range dependent data, their discussion was focused on estimating f itself, and the performance was evaluated in L_2 distance, which does not tell us a lot about the error of estimation at a fixed point. Also, the relationship between the minimax affine risk and the minimax risk was not obtained for such settings.

We prove (5) in few steps. First, we notice that for an affine estimator $\hat{L}(\mathbf{w}, d)$, we can find $c \in \mathbb{R}$ and $\mathbf{w}_0 \in S(\mathbf{U})$, the unit sphere of $\mathbf{U}$, such that $\mathbf{w} = c\mathbf{w}_0$. Thus the affine estimator becomes $\hat{L}(c\mathbf{w}_0, d)$. For calculating the minimax affine risk, we can take the infimum over $\mathbf{w}_0 \in S(\mathbf{U})$ and

$c, d \in \mathbb{R}$. Next, for technical reasons to be explained in the sequel, every vector $\mathbf{w} \in S(\mathbf{U})$ is approximated by some $c\mathbf{w}'$ with $\mathbf{w}' \in W_a$, where

$$W_a = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = a, 0 < a < 1, \|\mathbf{w}\| \leq 1\} \tag{6}$$

for a fixed unit vector $\mathbf{v} \in S(\mathbf{U})$, thus solving the reduced problem of finding the minimax risk with respect to the subset $\left\{\hat{L}(c\mathbf{w}, d) : c, d \in R, \mathbf{w} \in W_a\right\}$ of all affine estimators. Finally, we extend the result to the full problem – finding the minimax affine risk (with the infimum taken over all $\mathbf{w} \in S(\mathbf{U})$).

The article is organized as follows. Section 2 gives some general results on the model described by (1). Section 3 applies the results to the fractional Brownian model and the nonparametric regression model with correlated errors respectively. Finally most of the technical proofs for the results presented in these two sections are given in Section 4.

2 Minimax Risk for the Hardest one dimensional sub-problem and the Full Problem

In this section we first consider the minimax risk for the hardest one dimensional problem. These results are then extended to the general $\mathcal{F}$. We start with a few definitions. Suppose that T is an operator from $\mathcal{F}$ to another Banach space H. The modulus of continuity of T is defined as

$$\omega(\epsilon; T, \mathcal{F}) = \sup\{\|T(\mathbf{x}_2) - T(\mathbf{x}_1)\| : \|\mathbf{x}_2 - \mathbf{x}_1\| \leq \epsilon \text{ and } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}\}$$

where the $\|\ \|$'s are the norms of the respective normed spaces. As in Donoho (1994), we define the modulus of continuity of L with respect to the seminorm $\|v\|_{\mathbf{K}} \equiv \|\mathbf{K}v\|$ as

$$\omega(\epsilon; L, \mathbf{K}, \mathcal{F}) = \sup\{|L(\mathbf{x}_2) - L(\mathbf{x}_1)| : \|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbf{K}} \leq \epsilon \text{ and } \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}\}.$$

We assume that $\omega(\epsilon; L, \mathbf{K}, \mathcal{F})$ is finite for all $\epsilon > 0$. We also make the following assumptions

Assumption 1. (a) $\lim_{\epsilon \to 0} \omega(\epsilon; L, \mathcal{F}) \to 0$ and (b) $\lim_{\epsilon \to 0} \omega(\epsilon; \mathbf{K}, \mathcal{F}) \to 0$.

2.1 The hardest one dimensional sub-problem

To address the hardest one dimensional subproblem, we first look at a one dimensional sub-problem. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$. Let $[\mathbf{x_1}, \mathbf{x_2}]$ denote the convex span of $\{\mathbf{x}_1, \mathbf{x}_2\}$. Since $\mathcal{F}$ is convex, this is a subfamily of $\mathcal{F}$. Now we consider the problem of estimating $L(\mathbf{x})$ with affine estimators when we know that $\mathbf{x}$ is in $[\mathbf{x}_1, \mathbf{x}_2]$. Let $\hat{L}$ be an affine estimator defined as in (2). The risk of $\hat{L}$ is

$$\begin{aligned}
E_{\mathbf{x}}\left(\hat{L} - L\left(\mathbf{x}\right)\right)^2 &= E\left(L(\mathbf{x}) - \langle \mathbf{w}, \mathbf{K}\mathbf{x}\rangle - \mathbf{Z}\mathbf{w} - d\right)^2 \\
&= \left(L\left(\mathbf{x}\right) - \langle \mathbf{w}, \mathbf{K}\mathbf{x}\rangle - d\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2 \\
&= bias\left(\hat{L}, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2 .
\end{aligned}$$

Then the maximum risk for $\hat{L}$ over $[\mathbf{x}_1, \mathbf{x}_2]$ is

$$R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}; L, \mathbf{K}) = \sup_{\mathbf{x}\in[\mathbf{x}_1,\mathbf{x}_2]} bias\left(\hat{L}, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2. \tag{7}$$

Later in the text, we may omit part or all of the secondary arguments of $R_{[\mathbf{x}_1,\mathbf{x}_2]}$ in (7) and simply write $R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)$ when they are clear from the context. We will also do the same in other notations with secondary arguments.

For fixed $\mathbf{w} \in \mathbf{U}$, $R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L})$ is minimized for estimators with

$$bias\left(\hat{L}, \left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\right) = 0$$

and this gives

$$d = L(\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2) - \langle \mathbf{w}, \mathbf{K}\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\rangle.$$

For the estimator $\hat{L}\left(\mathbf{w}, d\right)$, the maximum risk is attained at either $\mathbf{x} = \mathbf{x}_1$ or $\mathbf{x}_2$. Thus

$$\inf_{d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}(\mathbf{w}, d)) = \left(L_1\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2 - \langle \mathbf{w}, \mathbf{K}\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2\rangle\right)^2 + \|\mathbf{Z}\mathbf{w}\|^2.$$

Notice that in the above equation, we extend the definition of L_1 and $\mathbf{K}$ to the algebra generated by $\mathcal{F}$, and we will do the same later on whenever necessary. Now, let $\mathbf{w} = c\mathbf{w}_0$, where $\mathbf{w}_0 \in S(\mathbf{U}) = \{\mathbf{w} \in \mathbf{U} : \|\mathbf{w}\| = 1\}$. Then

$$\inf_{c,d} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) = \inf_{c}\Big\{c^2\|\mathbf{Z}\mathbf{w}_0\|^2 + \left[L_1\left((\mathbf{x}_2 - \mathbf{x}_1)/2\right) - \langle c\mathbf{w}_\mathbf{0}, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2\rangle\right]^2\Big\}.$$

A straightforward calculation shows that

$$\inf_{c,d} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) = \frac{\left[L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)\right]^2}{1 + \left[g_{\mathbf{w}_0}(\mathbf{x}_2 - \mathbf{x}_1)\right]^2}$$

with $g_{\mathbf{w}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{K}\mathbf{x}/2\rangle / \|\mathbf{Z}\mathbf{w}\|$, where the above minimum is achieved at

$$c = c_0 = \frac{L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)\langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2\rangle}{\langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2\rangle^2 + \|\mathbf{Z}\mathbf{w}_0\|^2}.$$

Hence, the minimax affine risk for the one-dimensional subfamily is

$$\begin{aligned}\inf_{\mathbf{w}\in\mathbf{U}, d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(\mathbf{w}, d)\right) &= \inf_{\mathbf{w}_0\in S(\mathbf{U}); c,d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}_0, d)\right) \\ &= \inf_{\mathbf{w}_0\in S(\mathbf{U})} \frac{\left[L_1((\mathbf{x}_2 - \mathbf{x}_1)/2)\right]^2}{1 + \left[g_{\mathbf{w}_0}(\mathbf{x}_2 - \mathbf{x}_1)\right]^2}. \qquad (8)\end{aligned}$$

Thus finding the minimax risk for the one-dimensional subfamily is the same as finding the maximum of $|g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)|$ over all $\mathbf{w} \in S(\mathbf{U})$. Also notice that in (8), $g_{\mathbf{w}_0}$ is determined by the "direction" of $\mathbf{w}_0$. This means that instead of taking the infimum in (8) over $S(\mathbf{U})$, we can take it over a set W of vectors that "covers all the directions" in the sense that for any $\mathbf{w} \notin W$, we can find $\mathbf{w}' \in W$ such that $\mathbf{w} = c\mathbf{w}'$ for some $c \in \mathbb{R}$. Donoho (1994), under his setting, showed that the minimax affine risk is achieved by the estimator of the form $\hat{L}(c_0\mathbf{w}_0, d_0)$ with $\mathbf{w}_0 = (\mathbf{x}_2 - \mathbf{x}_1)/\|\mathbf{x}_2 - \mathbf{x}_1\|$ using a sufficiency argument. This approach is not generally possible in our setting. However, we can still find a sequence $\mathbf{w}_n \in S(\mathbf{U})$ such that

$$\lim_n g_{\mathbf{w}_n}(\mathbf{x}_2 - \mathbf{x}_1) = \sup_{\mathbf{w}\in S(\mathbf{U})} g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1).$$

Now, by the weak sequential compactness of the unit ball of a separable Hilbert space, we can find a subsequence $\mathbf{w}_{n_k}$ that converge weakly to a $\mathbf{w}' \in \mathbf{U}$. A problem that can arise here is that this weak limit may be $\mathbf{0}$. As a way of getting around the zero limit issue, instead of taking the supremum over $S(\mathbf{U})$, we take the supremum over an indexed collection of bounded closed convex subsets of $\mathbf{U}$ that do not contain $\mathbf{0}$. Of course such sets will not "cover all the directions" in the above sense. However, by taking a suitable limit on the indexing parameter, such condition can be "almost" satisfied. Our choice of these subsets are of the form $W_a = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = a, \|\mathbf{w}\| \leq 1\}$, where $0 < a < 1$ and $\mathbf{v}$ is a fixed unit vector in $\mathbf{U}$. This collection of subsets has the property that every element in $S(\mathbf{U})$ can be approximated (in norm of $\mathbf{U}$) by some $c\mathbf{w}$ with $\mathbf{w} \in W_a$ and $c \in [-1, 1]$, and the error of this approximation goes to zero uniformly over $S(\mathbf{U})$ as $a \to 0$. Now we have the following result.

Lemma 1. *For every pair* $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$, *and* $0 < a < 1$ *there exists a* $\mathbf{w}_a(\mathbf{x}_2 - \mathbf{x}_1) \in W_a$ *such that* $\left|g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| = \sup_{\mathbf{w}\in W_a} \left|g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right|.$

Next, for $0 < a < 1$ and $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$, define

$$\begin{aligned} G_a\left(\mathbf{x}_2 - \mathbf{x}_1\right) &= G_a\left(\mathbf{x}_2 - \mathbf{x}_1; \mathbf{Z}, \mathbf{K}\right) \\ &= \sup_{\mathbf{w}\in W_a} \left|g_{\mathbf{w}}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| = \left|g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right| \end{aligned}$$

and

$$\begin{aligned} \rho_a\left(\mathbf{x}_2 - \mathbf{x}_1\right) &= \rho_a(\mathbf{x}_2 - \mathbf{x}_1; \mathbf{Z}, L, \mathbf{K}) \\ &= \inf_{\mathbf{w}\in W_a} \frac{\left(L_1(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2)\right)^2}{1 + g_{\mathbf{w}}^2\left(\mathbf{x}_2 - \mathbf{x}_1\right)}. \end{aligned}$$

Extending this notation we also define

$$G_0\left(\mathbf{x}\right) = \sup_{\mathbf{w}\in W_0} \left|g_{\mathbf{w}}\left(\mathbf{x}\right)\right|$$

and

$$\rho_0\left(\mathbf{x}\right) = \inf_{\mathbf{w}\in W_0} \frac{\left(L_1\left(\mathbf{x}/2\right)\right)^2}{1 + g_{\mathbf{w}}^2\left(\mathbf{x}\right)},$$

where $W_0 \equiv S\left(\mathbf{U}\right)$. Note that $\rho_a\left(\mathbf{x}\right)$ is non-increasing when $a \downarrow 0$. By Lemma 1, $\rho_a(\mathbf{x}) = \left(L_1(\mathbf{x}/2)\right)^2 \Big/ \left(1 + g_{\mathbf{w}_a(\mathbf{x})}^2\left(\mathbf{x}\right)\right) = \left(L_1\left(\mathbf{x}/2\right)\right)^2 / \left(1 + G_a^2\left(\mathbf{x}\right)\right),$ $0 < a < 1$. Thus

$$\begin{aligned} \inf_{\mathbf{w}\in W_a; c,d\in\mathbb{R}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}(c\mathbf{w}, d)\right) &= \rho_a(\mathbf{x}_2 - \mathbf{x}_1) \\ &= \frac{\left[L_1(\left(\mathbf{x}_2 - \mathbf{x}_1\right)/2)\right]^2}{1 + \left[g_{\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)}\left(\mathbf{x}_2 - \mathbf{x}_1\right)\right]^2} \end{aligned}$$

for $a \in (0,1)$, where $R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)$ is defined in (7).

Having examined the minimax risk for a one dimensional sub-problem, we next find the minimax risk for the hardest 1-dimensional sub-problem. The lemma given below will be used in the sequel. It is a slightly modified version of Lemma 5 in Donoho (1994), and can be proven along the same lines.

Lemma 2. *Let* $\mathbf{V}$ *be a closed bounded convex set in a separable Hilbert space* $\mathbf{H}$*, and* $J(\mathbf{v})$ *a continuous convex functional on* $\mathbf{V}$*. Suppose that* $(\mathbf{v}_n)$ *is a sequence in* $\mathbf{V}$ *converging weakly to* $\mathbf{v}$*. Then*

$$J(\mathbf{v}) \leq \liminf J(\mathbf{v}_n).$$

We define

$$\begin{aligned}\rho_a(\mathcal{F}) &= \rho_a(\mathcal{F};\mathbf{Z},L,\mathbf{K}) \\ &= \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}} \rho_a(\mathbf{x}_2-\mathbf{x}_1;\mathbf{Z},L,\mathbf{K})\end{aligned}$$

for $0 < a < 1$ and

$$\begin{aligned}\rho_0(\mathcal{F}) &= \rho_0(\mathcal{F};\mathbf{Z},L,\mathbf{K}) \\ &= \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}} \rho_0(\mathbf{x}_2-\mathbf{x}_1;\mathbf{Z},L,\mathbf{K}).\end{aligned}$$

It is easy to see that $\rho_a(\mathcal{F})$ is non-increasing as $a \downarrow 0$. Now, we can prove the following lemma that finds the minimax risk for the hardest one dimensional sub-problem for the restricted case $\mathbf{w} \in W_a$.

Lemma 3. *If* $\mathcal{F}$ *is convex, closed, and bounded, then for each* $0 < a < 1$ *there exists a pair* $\mathbf{x}_1(a;\mathbf{Z},L,\mathbf{K},\mathcal{F}), \mathbf{x}_2(a;\mathbf{Z},L,\mathbf{K},\mathcal{F})$ *(which we simply write as* $\mathbf{x}_1(a), \mathbf{x}_2(a)$ *when there is no confusion) such that* $\rho_a(\mathbf{x}_2(a)-\mathbf{x}_1(a)) = \rho_a(\mathcal{F})$.

2.2 The Full Problem

We obtain the minimax risk for the full problem in this section. This is examined in several steps. First, we obtain the minimax risk for estimating

L with affine estimators of the form $L\left(c\mathbf{w},d\right)$, with $\mathbf{w}\in W_a$ over a bounded closed convex parameter space $\mathcal{F}$. Once the minimax risk over such sets are derived, our goal is to remove the boundedness and closedness for $\mathcal{F}$ and then extend the results to estimators with $\mathbf{w}\in S\left(\mathbf{U}\right)$. In particular, we now give the following result, which states that for a bounded closed convex parameter space, when we restrict our attention to affine estimators of the form $\hat{L}(c\mathbf{w}+d)$ with $\mathbf{w}\in W_a$, the minimax risk is the supremum of the minimax risks of one-dimensional subproblems.

Theorem 1. *If $\mathcal{F}$ is a bounded closed convex subset of $\mathbf{X}$, then*

$$\inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right)=\rho_a\left(\mathcal{F}\right)=\rho_a\left(\mathbf{x}_2\left(a\right)-\mathbf{x}_1\left(a\right)\right)$$

where $0<a<1$ and $\hat{L}=\hat{L}\left(c\mathbf{w},d\right)$. The above infimum is achieved at $\mathbf{w}_0=\mathbf{w}_a\left(\mathbf{x}_2\left(a\right)-\mathbf{x}_1\left(a\right)\right)$ with $c=c_0$ where

$$c_0=\frac{L_1((\mathbf{x}_2\left(a\right)-\mathbf{x}_1\left(a\right))/2)\langle\mathbf{w}_0,\mathbf{K}\left(\mathbf{x}_2\left(a\right)-\mathbf{x}_1\left(a\right)\right)/2\rangle}{\langle\mathbf{w}_0,\mathbf{K}\left(\mathbf{x}_2\left(a\right)-\mathbf{x}_1\left(a\right)\right)/2\rangle^2+\left\|\mathbf{Z}\mathbf{w}_0\right\|^2},$$

and

$$d=d_0=L((\mathbf{x}_1\left(a\right)+\mathbf{x}_2\left(a\right))/2)-\langle\mathbf{w}_0,\mathbf{K}\left(\mathbf{x}_1\left(a\right)+\mathbf{x}_2\left(a\right)\right)/2\rangle,$$

and $\mathbf{x}_1\left(a\right)$ and $\mathbf{x}_2\left(a\right)$ are defined in Lemma 3.

Now we argue that the boundedness and closedness constraints for $\mathcal{F}$ in Theorem 1 can be removed. To this end, we have the following theorem which is proven in a fashion similar to the proof of Theorem 2 in Donoho (1994).

Theorem 2. *Let $\mathcal{F}$ be a convex subset of $\mathbf{X}$, and $0<a<1$. Then*

$$\inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right)=\rho_a\left(\mathcal{F}\right)$$

and there exists $\tilde{\mathbf{w}}\in W_a$ and $c,d\in\mathbb{R}$ such that the affine estimator $\hat{L}_0=\hat{L}\left(c\tilde{\mathbf{w}},d\right)$ achieves the above minimax risk.

Theorem 2 shows that when we restrict the affine estimator to be determined by some $\mathbf{w}\in W_a$, the minimax affine risk of estimating $L\left(\mathbf{x}\right)$ is just the supremum over the minimax affine risks (the infimum is taken over estimators of the form $\hat{L}(c\mathbf{w}, d)$ with $\mathbf{w} \in W_a$, and $c, d \in \mathbb{R}$) of all one-dimensional subproblems. We also notice that every vector in $S(\mathbf{U})$ can be approximated (in norm) by $c\mathbf{w}$ with a $c \in \mathbb{R}$ and $\mathbf{w} \in W_a$, and the error of this approximation goes to zero as a goes to zero. Thus it is natural to think that the minimax affine risk $\inf_{\mathbf{w}\in S(\mathbf{U});c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right)$ can be obtained by taking the limit in a of $\inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right)$. The following theorem, which is the main result of this article, will show that this is indeed the case.

Theorem 3. *Suppose that $\mathcal{F}$ is convex. Then the $\mathbf{v}$ in (6) can be chosen so that*

$$\inf_{\hat{L}\ \textit{affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \lim_{a\to 0} \inf_{\mathbf{w}\in W_a;c,d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}(c\mathbf{w}, d)\right) = \lim_{a\to 0} \rho_a(\mathcal{F}).$$

In Theorem 3, the minimax affine risk is expressed as the limit, $\lim_{a\to 0} \rho_a(\mathcal{F})$. It will be desirable if we can express the minimax affine risk in term of $\rho_0(\mathcal{F})$. This can be done for bounded $\mathcal{F}$ as given in the following lemma.

Lemma 4. *If $\mathcal{F}$ is convex, closed and bounded, then*

$$\lim_{a\to 0} \rho_a(\mathcal{F}) = \rho_0(\mathcal{F}). \tag{9}$$

The result proven in the next lemma shows that we can actually relax the boundedness restriction on $\mathcal{F}$. To proceed with this avenue, we make the following assumption.

Assumption 2. For any positive M, we have

$$\chi(M) = \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F},\|\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)\|\le M} \|\mathbf{x}_2 - \mathbf{x}_1\| < \infty$$

Lemma 5. *If $\mathcal{F}$ is closed and symmetric and Assumption 2 is satisfied, then* (9) *holds.*

With the above Theorem and the previous lemma, and also the fact that $\rho_a(\mathcal{F}) = \rho_a(cl(\mathcal{F}))$, and $\rho_0(\mathcal{F}) = \rho_0(cl(\mathcal{F}))$, which are not hard to prove, the following corollary is immediate.

Corollary 4. *If Assumption 2 is satisfied and $\mathcal{F}$ is convex and symmetric, then* $\inf_{\hat{L}\ \textit{affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_0(\mathcal{F})$.

3 Applications

Now we apply the above results to two examples. The first is the fractional Brownian motion model, and the second is the nonparametric regression model with correlated errors.

3.1 Fractional Brownian Motion Model

Let $\Omega = \mathscr{C}_0\left([0,T],\mathbb{R}\right)$ be the space of real-valued continuous functions on the interval $[0,T]$ with the topology of local uniform convergence and initial value zero. Let $\mathscr{F}$ be the Borel σ-algebra. There is a probability measure P on $(\Omega,\mathscr{F})$ under which the coordinate process $(Z_t, t\in[0,T])$ is a Gaussian process that has stationary increments and satisfies the following.

i. $Z_0 = 0$.

ii. $EZ_t = 0$ for all $t \geq 0$.

iii. $R\left(s,t\right) = EZ_tZ_s = \left(t^{2H} + s^{2H} - |s-t|^{2H}\right)/2$ for every $s,t \geq 0$.

The process $(Z_t, t\in[0,T])$ is called the fractional Brownian motion on the interval $[0,T]$ with an index H, $1/2 < H < 1$. Now consider the following model

$$y\left(t\right) = \int_0^t f\left(u\right)du + \sigma Z_t, t\in[0,T] \tag{10}$$

where $\sigma > 0$ and f belongs to $\mathbf{W}_{[0,T]}\left(m,p,C\right)$ (Donoho and Liu, 1991) with $m \geq 1$ and $1 \leq p \leq \infty$. That is, f is defined on $[0,T]$ and satisfies

a. $f,\ldots,f^{(m-1)}$ are absolutely continuous, and

b. $f \in \mathscr{L}_2\left[0,T\right]$ and $\left\|f^{(m)}\right\|_p \leq C$.

We are interested in the problem of estimating $f^{(k)}\left(t_0\right)$ for $0 \leq k < m$ and $t_0 \in (0,T)$ from observing $y\left(t\right), t\in[0,T]$. It is assumed that either $k < m-1$ or $p > 1$. First we verify that this is a special case of the model and the estimation problem discussed in section 2. For a function

$g \in \mathscr{L}_2[0,T]$, the integral $\int_0^T g(u)\,dZ_u$ is well defined, and it can be shown that

$$E\left(\left|\int_0^T g(u)\,dZ_u\right|^2\right) = \int_0^T \int_0^T g(u)\,g(v)\,\phi(u,v)\,dudv < \infty$$

where $\phi(u,v) = H(2H-1)|u-v|^{2H-2}$. Following Duncan et al. (2000), we define $|g|_\phi := \left(\int_0^T \int_0^T g(s)\,g(t)\,\phi(s,t)\,dsdt\right)^{1/2}$ and let

$$\mathscr{L}_\phi^2 = \mathscr{L}_\phi^2([0,T]) = \left\{ f | f : [0,T] \to \mathbb{R}, |f|_\phi^2 < \infty \right\}.$$

An inner product $\langle \cdot, \cdot \rangle_\phi$ can also be defined on $\mathscr{L}_\phi^2$:

$$\langle f, g \rangle_\phi = \int_0^T \int_0^T f(u)\,g(v)\,\phi(u,v)\,dudv.$$

Memin et al. (2001) showed that for $g \in \mathscr{L}_{1/H}[0,T]$,

$$E\left(\left|\int_0^T g(u)\,dZ_u\right|^2\right) \le c(H,2)\,\|g\|_{1/H}^2.$$

This enables us to treat the integration with respect to Z as a bounded operator from $\mathscr{L}_{1/H}[0,T]$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. For $f \in \mathscr{L}_2[0,T]$, $|f|^{1/H} \in \mathscr{L}_{2H}[0,T]$. Thus it can be shown that

$$\begin{aligned}
\|f\|_{1/H} &= \left(\int_0^T |f|^{1/H}\,dt\right)^H \\
&\le \left(\left(\int_0^T \left(|f|^{1/H}\right)^{2H}\right)^{1/2H} \left(\int_0^T 1dt\right)^{1-1/2H}\right)^H \\
&= T^{H-1/2}\,\|f\|_2\,.
\end{aligned}$$

Thus the identity map I from $\mathscr{L}_2[0,T]$ to $\mathscr{L}_{1/H}[0,T]$ is bounded. Combining these two operators we get a bounded linear operator from $\mathscr{L}_2[0,T]$ to $\mathscr{L}_2(\Omega, \mathscr{F}, P)$. We write this operator as $\mathbf{Z}_1$ and let $\mathbf{Z}_\sigma = \sigma \mathbf{Z}_1$. To show that with $\mathcal{F} = \mathbf{W}_{[0,T]}(m,p,C)$, $\mathbf{Z} = \mathbf{Z}_\sigma$, and $\mathbf{K} = \mathbf{I}$, the model described by (10) is a special case of the model described by (1), now we only need to show that Assumption 1 is satisfied, which is immediate from the following lemma.

Lemma 6. $\sup_{f\in \mathbf{W}_{[0,T]}(m,p,C),\|f\|_r\le\epsilon} \left\|f^{(k)}\right\|_\infty = O\left(\epsilon^{\alpha_k}\right)$ *where*

$$\alpha_k = \left(m-k-p^{-1}\right)/\left(m-p^{-1}+r^{-1}\right) \; and 0\le k<m.$$

Now, we examine the rate of convergence of the minimax affine risk for estimating $f^{(k)}(t_0)$ for model (10). For $g \in \mathscr{L}_2[0,T]$ and $d\in\mathbb{R}$, we can define an affine estimator

$$\hat{L}(g,d) = \int_0^T g(u)\,y(u)\,du + d = \int_0^T g(u)\,f(u)\,du + \sigma\int_0^T g(u)\,dZ_t + d.$$

The main result in this section is to obtain the rate of convergence of the minimax affine risk when $\sigma\to 0$, and to show that the minimax affine risk and the minimax risk have the same rate of convergence. Define

$$\upsilon(\epsilon;\mathbf{Z},L,\mathbf{K},\mathcal{F}) = \inf\left\{\left|G_0(\mathbf{x}_2-\mathbf{x}_1)\right| : \mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}, \left|L_1((\mathbf{x}_2-\mathbf{x}_1)/2)\right| = \epsilon\right\}.$$

If $\{\mathbf{x}_2-\mathbf{x}_1 : \mathbf{x}_1,\mathbf{x}_2\in\mathcal{F},\ L_1((\mathbf{x}_2-\mathbf{x}_1)/2)=\epsilon\} = \emptyset$, then we let $\upsilon(\epsilon)=\infty$. It is easy to check that $\upsilon(\epsilon)$ is convex. Clearly,

$$\begin{aligned}
\rho_0(\mathcal{F}) &= \sup_{\mathbf{x}_1,\mathbf{x}_2\in\mathcal{F}} \frac{\left[L_1((\mathbf{x}_2-\mathbf{x}_1)/2)\right]^2}{1+\left[G_0(\mathbf{x}_2-\mathbf{x}_1)\right]^2} \\
&= \sup_{\epsilon>0}\sup\left\{\frac{\left[L_1(\mathbf{x}_2-\mathbf{x}_1)/2\right]^2}{1+\left[G_0(\mathbf{x}_2-\mathbf{x}_1)\right]^2} : \left|L_1((\mathbf{x}_2-\mathbf{x}_1)/2)\right| = \epsilon\right\} \\
&= \sup_{\epsilon>0}\frac{\epsilon^2}{1+\upsilon^2(\epsilon)}.
\end{aligned}$$

If $\mathcal{F}$ is symmetric and $\mathbf{K}$ satisfies Assumption 2, then by Corollary 4, we have

$$\inf_{\hat{L}\text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \sup_{\epsilon>0}\frac{\epsilon^2}{1+\upsilon^2(\epsilon)}. \tag{11}$$

In the fractional Brownian model described above, $\mathcal{F}=\mathbf{W}_{[0,T]}(m,p,C)$ and $\mathbf{K}=\mathbf{I}$. It is easy to check that $\mathbf{W}_{[0,T]}(m,p,C)$ is symmetric. For a function $h(\epsilon)$ of ϵ, we use the notation $h(\epsilon)\asymp\epsilon^\alpha$ to denote $A_1\epsilon^\alpha\le h(\epsilon)\le A_2\epsilon^\alpha$ for ϵ sufficiently small, and A_1, A_2 two constants free of ϵ. Now we have the following lemma that characterizes υ for this specific model.

Lemma 7. $\upsilon\left(\epsilon;\mathbf{Z}_\sigma,f^{(k)}(t_0),\mathbf{I},\mathbf{W}_{[0,T]}(m,p,C)\right)\asymp\epsilon^{\gamma_k}/\sigma$, *with*

$$\gamma_k = \frac{m-p^{-1}+1-H}{m-k-p^{-1}}.$$

With Lemma 7, we can now obtain the rate of the minimax affine risk given in (11) above. In particular, we have the following result.

Theorem 5. *The minimax affine risk for the model described in* (10) *satisfies*

$$\inf_{\hat{L}\ affine} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}\right) \asymp \sigma^{2/\gamma_k}. \tag{12}$$

Next, we will get an upper bound for the ratio of the minimax affine risk and the minimax risk, and show that the rate of convergence for minimax risk for estimating $f^{(k)}(t_0)$ is also given by (12).

Theorem 6. *The minimax affine risk for the model described in* (10) *satisfies*

$$\frac{\inf_{\hat{L}\ affine} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}\right)}{\inf_{\hat{T}\ measurable} R_{\mathbf{W}^r(m,p,C)}\left(\hat{T}\right)} \leq 1.25.$$

With Theorem 5 and Theorem 6, we have the following corollary.

Corollary 7. *The minimax risk for the model described in* (10) *satisfies*

$$\inf_{\hat{T}\ measurable} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}\right) \asymp \sigma^{2/\gamma_k}.$$

A close look at the results of Theorem 5 and Corollary 7 reveals that if the index of the fractional Brownian motion H is taken to be 1/2, in which case it reduces to Brownian motion, then our results agree with those presented in Donoho and Liu (1991).

3.2 Regression Model

Nonparametric regression with long-range dependent errors was studied in Wang (1996), who also established asymptotics for minimax risk with respect to the L_2 norm. In this article, we consider the pointwise minimax risk. The regression model is described as

$$y_i = f(t_i) + z_i, \quad i = 1, \ldots, n, \tag{13}$$

where the t_i's are equispaced on $[0,T]$ and $z_1,\ldots,z_n$ are observational errors with mean 0 and finite variance. We assume that $(z_i)_{1\le i\le n}$ have long-range dependence (Wang, 1996),

$$R(j-i)=\operatorname{Cov}(z_i,z_j)\sim C_1\,|j-i|^{-\alpha}, \qquad j-i\to\infty$$

with $0<\alpha<1$. The regression function f is known to belong to $\mathbf{W}_{[0,T]}(m,p,C)$. The problem of interest is the estimation of $f^{(k)}(t_0)$ for some $t_0\in(0,T)$, $0\le k<m$, by affine estimators of the form (2) with $\mathbf{x}=f$, $\mathbf{y}=(y_1,\ldots,y_n)'\in\mathbb{R}^n$, and $\mathbf{K}_n:\mathbf{W}_{[0,T]}(m,p,C)\to\mathbb{R}^n$ and $\tilde{\mathbf{Z}}_n:\mathbb{R}^n\to\mathscr{L}_2(\Omega,\mathscr{F},P)$ being operators defined by $\mathbf{K}_nf=(f(t_1),\ldots,f(t_n))'$ and $\tilde{\mathbf{Z}}_n((c_1,\ldots,c_n)')=\frac{1}{n^{1/2}}\sum_{i=1}^n c_iz_i$ respectively. For two vectors $\mathbf{u}=(u_1,\ldots,u_n)'$ and $\mathbf{v}=(v_1,\ldots,v_n)'$ in $\mathbb{R}^n$, we define the inner product

$$\langle\mathbf{u},\mathbf{v}\rangle=\frac{1}{n}\sum u_iv_i$$

and the norm

$$\|\mathbf{u}\|_2=\sqrt{\langle\mathbf{u},\mathbf{u}\rangle}.$$

By the smoothness of the functions in $\mathbf{W}_{[0,T]}(m,p,C)$, it is easy to see that $\mathbf{K}_n$ is well defined. We will also show that $\mathbf{K}_n$ also satisfies Assumption 1 and 2. Throughout this section we assume,

$$L(f)=f^{(k)}(t_0)$$

and

$$\mathcal{F}=\mathbf{W}_{[0,T]}(m,p,C).$$

By Lemma 6, the $\mathbf{K}_n$ and $\mathcal{F}$ defined above satisfies part (b) of Assumption 1. The fact that Assumption 2 is satisfied follows readily from the following lemma.

Lemma 8. *If $n\ge m$, then there exists a positive K independent of f such that*

$$\left\|f^{(k)}\right\|_\infty\le K\,\|\mathbf{K}_nf\|_2+CT^{m-k-p^{-1}} \tag{14}$$

for $k=0,\ldots,m-1$.

Theorem 8. *Let $\tilde{H} = 1 - \alpha/2$. Then the minimax affine risk for the model described in* (13) *satisfies*

$$\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) \asymp n^{2\left(\tilde{H}-1\right)/\gamma_k}$$

where $\gamma_k = \left(m - p^{-1} + 1 - \tilde{H}\right) / \left(m - k - p^{-1}\right)$.

By a sufficiency discussion similar to that in the proof of Theorem 6, the following theorem can be proven.

Theorem 9. *For the model described in* (13)*, we have*

$$\frac{\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right)}{\inf_{\hat{T} \text{ measurable}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right)} \leq 1.25.$$

By Theorem 8 and Theorem 9, we have

Corollary 10. *The minimax risk for the model described in* (13) *satisfies*

$$\inf_{\hat{T} \text{ measurable}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{T}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) \asymp n^{2\left(\tilde{H}-1\right)/\gamma_k}.$$

Deo (1997) discussed two kernel estimators for estimating the value of a function $f : [0,1] \to \mathbb{R}$ at $x \in (0,1)$ from data with long-memory errors. According to Theorem 3 of Deo (1997), if f is assumed to be twice differentiable on $[0,1]$ with both derivatives bounded, then the two kernel estimators studied have asymptotic risk of order $(nh_n)^{-\alpha}$ where h_n satisfies $nh_n^{1+4/\alpha} \to 0$, $nh_n^{1+\eta} \to \infty$ for some $\eta > 0$. It can be seen that in this case the parameter space is a subset of $W_{[0,1]}(2, \infty, C)$ for some $C > 0$, and according to our result, the rate of convergence for the minimax risk is of order $n^{-4\alpha/(4+\alpha)}$, which is a lower bound for the rate for the kernel estimators given by Deo (1997).

4 Proofs

In this section we provide proofs of the results stated in the previous sections.

4.1 Proofs for results in Section 2.1

Proof of Lemma 1

Suppose that $\mathbf{w}_n \in W_a$, $n = 1, 2, \ldots$, and

$$|g_{\mathbf{w}_n}(\mathbf{x}_2 - \mathbf{x}_1)| \to \sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)|.$$

By the weak sequential compactness of the unit ball $B(\mathbf{U})$, we can find a subsequence $\mathbf{w}_{n_k}$ which converge weakly to $\mathbf{w}_0 \in W_a$. Then

$$\langle \mathbf{w}_{n_k}, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2 \rangle \to \langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2 \rangle$$

and $\mathbf{Z}\mathbf{w}_{n_k} \xrightarrow{w} \mathbf{Z}\mathbf{w}_0$, which gives

$$\liminf \|\mathbf{Z}\mathbf{w}_{n_k}\| \geq \|\mathbf{Z}\mathbf{w}_0\|.$$

Thus

$$\begin{aligned}
\sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)| &= \lim_{k \to \infty} |\langle \mathbf{w}_{n_k}, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2 \rangle / \|\mathbf{Z}\mathbf{w}_{n_k}\||\\
&= |\langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2 \rangle| / \liminf \|\mathbf{Z}\mathbf{w}_{n_k}\|\\
&\leq |\langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)/2 \rangle| / \|\mathbf{Z}\mathbf{w}_0\|\\
&= |g_{\mathbf{w}_0}(\mathbf{x}_2 - \mathbf{x}_1)|\\
&\leq \sup_{\mathbf{w} \in W_a} |g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)|.
\end{aligned}$$

Letting $\mathbf{w}_a(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{w}_0$ finishes the proof.

Proof of Lemma 3

Suppose that $\mathbf{x}_1^n, \mathbf{x}_2^n, n = 1, 2, \ldots$ satisfy

$$\rho_a(\mathbf{x}_2^n - \mathbf{x}_1^n) \to \rho_a(\mathcal{F}).$$

Then we find subsequences $\mathbf{x}_1^{n_k}, \mathbf{x}_2^{n_k}$ such that $\mathbf{x}_i^{n_k} \xrightarrow{w} \mathbf{x}_i(a), i = 1, 2$. With Assumption 1, and by Lemma 2, we have $L_1(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}) \to L_1(\mathbf{x}_2(a) - \mathbf{x}_1(a))$

and $\langle \mathbf{w}, \mathbf{K}(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k})\rangle \to \langle \mathbf{w}, \mathbf{K}(\mathbf{x}_2(a) - \mathbf{x}_1(a))\rangle$ for any $\mathbf{w} \in \mathbf{U}$. Hence,

$$\begin{aligned}
\rho_a(\mathcal{F}) &\geq \rho_a(\mathbf{x}_2(a) - \mathbf{x}_1(a)) \\
&= \frac{\left[L_1\left((\mathbf{x}_2(a) - \mathbf{x}_1(a))/2\right)\right]^2}{1 + \left[g_{\mathbf{w}_a(\mathbf{x}_2(a)-\mathbf{x}_1(a))}(\mathbf{x}_2(a) - \mathbf{x}_1(a))\right]^2} \\
&= \lim \frac{\left[L_1\left((\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k})/2\right)\right]^2}{1 + \left[g_{\mathbf{w}_a(\mathbf{x}_2(a)-\mathbf{x}_1(a))}(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k})\right]^2} \\
&\geq \limsup \rho_a(\mathbf{x}_2^{n_k} - \mathbf{x}_1^{n_k}) = \rho_a(\mathcal{F})
\end{aligned}$$

proving the lemma.

4.2 Proofs for Results in Section 2.2

To prove Theorem 1, we need the following lemmas, Lemma 9 – Lemma 14. We state each lemma, give a proof and then prove Theorem 1.

Lemma 9. *Suppose that $\mathcal{F}$ is a bounded closed convex subset of $\mathbf{X}$. For $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ and $h \in (0,1)$, Let $\mathbf{x}_h = h\mathbf{x} + (1-h)\mathbf{x}_2 - \mathbf{x}_1$ and $\mathbf{x}_0 = \mathbf{x}_2 - \mathbf{x}_1$. Then $\left|\left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_h)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_0)}(\mathbf{x}_0)\right|\right| = O(h)$.*

Proof of Lemma 9

Since $\left|g_{\mathbf{w}_a(\mathbf{x}_0)}(\mathbf{x}_0)\right| \geq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_0)\right|$, we have

$$\begin{aligned}
\left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_h)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_0)}(\mathbf{x}_0)\right| &\leq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_h)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_0)\right| \\
&\leq \left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_h) - g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_0)\right| \\
&= \left|\frac{\langle \mathbf{w}_a(\mathbf{x}_h), h\mathbf{K}\tilde{\mathbf{x}}/2\rangle}{\|\mathbf{Z}\mathbf{w}_a(\mathbf{x}_h)\|}\right| \\
&\leq \frac{h\|\mathbf{K}\tilde{\mathbf{x}}\|/2}{M} \\
&= O(h)
\end{aligned}$$

where $M = \inf_{\mathbf{w} \in W_a} \|\mathbf{Z}\mathbf{w}\| > 0$, and $\tilde{\mathbf{x}} = \mathbf{x} - \mathbf{x}_2$. Similarly, we can show that

$$\left|g_{\mathbf{w}_a(\mathbf{x}_0)}(\mathbf{x}_0)\right| - \left|g_{\mathbf{w}_a(\mathbf{x}_h)}(\mathbf{x}_h)\right| = O(h).$$

This proves the lemma.

Now we introduce a few new terms. For $\mathbf{u} \in \mathbf{U}$, define

$$I_t^{\mathbf{u}} = \left\{ \mathbf{w} \in W_a : \langle \mathbf{w}, \mathbf{u} \rangle = t \sup_{\mathbf{w}' \in W_a} \left\langle \mathbf{w}', \mathbf{u} \right\rangle + (1-t) \inf_{\mathbf{w}' \in W_a} \left\langle \mathbf{w}', \mathbf{u} \right\rangle \right\}$$

where $t \in [0,1]$. Now let $\mathbf{u} = \mathbf{u}_{\mathbf{v}} + \mathbf{u}_{\perp}$, where $\mathbf{v}$ is as in (6), $\mathbf{u}_{\mathbf{v}} = \alpha \mathbf{v}$ and $\mathbf{u}_{\perp} \in \ker(\mathbf{v}) = \{\mathbf{w} \in \mathbf{U} : \langle \mathbf{w}, \mathbf{v} \rangle = 0\}$. For $\mathbf{w} \in W_a$, we have

$$\begin{aligned} \langle \mathbf{w}, \mathbf{u} \rangle &= \langle \mathbf{w}, \mathbf{u}_{\mathbf{v}} \rangle + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = a\alpha + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle \\ &\in [a\alpha - \|\mathbf{u}_{\perp}\| \theta, a\alpha + \|\mathbf{u}_{\perp}\| \theta] \end{aligned}$$

where $\theta = \sqrt{1-a^2}$. Hence, $\sup_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle = a\alpha + \|\mathbf{u}_{\perp}\| \theta$ and $\inf_{\mathbf{w}' \in W_a} \langle \mathbf{w}', \mathbf{u} \rangle = a\alpha - \|\mathbf{u}_{\perp}\| \theta$. For $\mathbf{w} \in I_t^{\mathbf{u}}$,

$$\langle \mathbf{w}, \mathbf{u} \rangle = a\alpha + \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = a\alpha + (2t-1) \|\mathbf{u}_{\perp}\| \theta.$$

Thus $\langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = (2t-1) \|\mathbf{u}_{\perp}\| \theta$. For $U, V \subset \mathbf{X}$, define

$$D(U,V) = \max \left\{ \sup_{\mathbf{x} \in U} \inf_{\mathbf{y} \in V} \|\mathbf{x} - \mathbf{y}\|, \sup_{\mathbf{x} \in V} \inf_{\mathbf{y} \in U} \|\mathbf{x} - \mathbf{y}\| \right\}.$$

Now we have the following lemma.

Lemma 10. *For $t \in [0,1]$ and $\mathbf{u} \in \mathbf{U}$, if $\mathbf{u}_{\perp} \neq \mathbf{0}$, then $D\left(I_t^{\mathbf{u}}, I_t^{\mathbf{u}'}\right) \to 0$ as $\mathbf{u}' \to \mathbf{u}$, and the convergence is uniform in t.*

Proof of Lemma 10

Let $J_t^{\mathbf{u}} = I_t^{\mathbf{u}} - a\mathbf{v}, J_t^{\mathbf{u}'} = I_t^{\mathbf{u}'} - a\mathbf{v}$. It is not hard to see that $D\left(I_t^{\mathbf{u}}, I_t^{\mathbf{u}'}\right) = D\left(J_t^{\mathbf{u}}, J_t^{\mathbf{u}'}\right)$. It can be shown that

$$J_t^{\mathbf{u}} = \{\mathbf{w} \in B_\theta(\ker(\mathbf{v})) : \langle \mathbf{w}, \mathbf{u}_{\perp} \rangle = (2t-1) \|\mathbf{u}_{\perp}\| \theta\}$$

where $B_\theta(\ker(\mathbf{v})) = \{\mathbf{w} \in \ker(\mathbf{v}) : \|\mathbf{w}\| \leq \theta\}$ and remember that $\theta = \sqrt{1-a^2}$. Now we will show that $diam(J_t^{\mathbf{u}}) \to 0$ uniformly in $\mathbf{u}$ if $t \to 1$ or $t \to 0$. Suppose that $\mathbf{w} \in J_t^{\mathbf{u}}$. Let $\mathbf{w}' = (2t-1) \theta \mathbf{u}_{\perp} / \|\mathbf{u}_{\perp}\|$ Then we have $\mathbf{w}' \in J_t^{\mathbf{u}}$. Therefore,

$$\left\langle \mathbf{w} - \mathbf{w}', \mathbf{w}' \right\rangle = \frac{(2t-1) \theta (\langle \mathbf{w}, \mathbf{u}_{\perp} \rangle - \langle \mathbf{w}', \mathbf{u}_{\perp} \rangle)}{\|\mathbf{u}_{\perp}\|} = 0$$

giving $\|\mathbf{w}\|^2 = \|\mathbf{w}-\mathbf{w}'+\mathbf{w}'\|^2 = \|\mathbf{w}-\mathbf{w}'\|^2+\|\mathbf{w}'\|^2$. We know that $\|\mathbf{w}'\| = |(2t-1)|\,\theta$ and $\|\mathbf{w}\| \le \theta$. Thus $\|\mathbf{w}-\mathbf{w}'\|^2 \le \theta^2 - (2t-1)^2\theta^2 = 4t(1-t)\theta^2$ so that

$$\begin{aligned} \mathit{diam}\left(J_t^{\mathbf{u}}\right) &= \sup_{\mathbf{w}_1,\mathbf{w}_2\in J_t^{\mathbf{u}}} \|\mathbf{w}_1-\mathbf{w}_2\| \\ &\le \sup_{\mathbf{w}_1,\mathbf{w}_2\in J_t^{\mathbf{u}}} \left(\|\mathbf{w}_1-\mathbf{w}'\| + \|\mathbf{w}_2-\mathbf{w}'\|\right) \\ &\le 4\sqrt{t(1-t)}\theta. \end{aligned}$$

Since $4\sqrt{t(1-t)}\theta \to 0$ if $t\to 0$ or $t \to 1$, and it does not depend on $\mathbf{u}$, we have proven the claim that $\mathit{diam}\left(J_t^{\mathbf{u}}\right) \to 0$ uniformly in $\mathbf{u}$.

Now for any $1 > \epsilon > 0$, we can find δ such that when $t \ge 1-\delta$ or $t \le \delta$, $\mathit{diam}\left(J_t^{\mathbf{u}}\right) < \epsilon/3$. When $\|\mathbf{u}'-\mathbf{u}\| < \epsilon\|\mathbf{u}_\perp\|/6$, we have

$$\begin{aligned} \left\|\frac{(2t-1)\,\theta\mathbf{u}_\perp}{\|\mathbf{u}_\perp\|} - \frac{(2t-1)\,\theta\mathbf{u}'_\perp}{\|\mathbf{u}'_\perp\|}\right\| &= \left\|\frac{(2t-1)\,\theta\left(\left(\|\mathbf{u}'_\perp\|-\|\mathbf{u}_\perp\|\right)\mathbf{u}_\perp + \|\mathbf{u}_\perp\|\left(\mathbf{u}_\perp-\mathbf{u}'_\perp\right)\right)}{\|\mathbf{u}_\perp\|\,\|\mathbf{u}'_\perp\|}\right\| \\ &\le \frac{|2t-1|\,\theta}{\|\mathbf{u}'_\perp\|}\left(\left|\|\mathbf{u}'_\perp\|-\|\mathbf{u}_\perp\|\right| + \|\mathbf{u}_\perp-\mathbf{u}'_\perp\|\right) \\ &\le \frac{2\|\mathbf{u}_\perp-\mathbf{u}'_\perp\|}{\|\mathbf{u}_\perp\|} \le \frac{2\|\mathbf{u}-\mathbf{u}'\|}{\|\mathbf{u}_\perp\|} \\ &< \frac{2\|\mathbf{u}_\perp\|\,\epsilon}{6\|\mathbf{u}_\perp\|} = \epsilon/3. \end{aligned}$$

Since $(2t-1)\,\theta\mathbf{u}_\perp/\|\mathbf{u}_\perp\| \in J_t^{\mathbf{u}}$, and likewise, $(2t-1)\,\theta\mathbf{u}'_\perp/\|\mathbf{u}'_\perp\| \in J_t^{\mathbf{u}'}$, we have that for $t \ge 1-\delta$ or $t \le \delta$, $D\left(J_t^{\mathbf{u}}, J_t^{\mathbf{u}'}\right) < \epsilon$. Now suppose that $t \in (\delta, 1-\delta)$. For $\mathbf{w} \in J_t^{\mathbf{u}}$, we have $\langle \mathbf{w}, \mathbf{u}_\perp\rangle = (2t-1)\|\mathbf{u}_\perp\|\,\theta$. First, we consider the case in which $\langle \mathbf{w}, \mathbf{u}'_\perp\rangle \le (2t-1)\|\mathbf{u}'_\perp\|\,\theta$. Suppose that $\|\mathbf{u}-\mathbf{u}'\| < \delta_1 = \delta\theta\|\mathbf{u}_\perp\|\,\epsilon/4$. Then $\|\mathbf{u}_\perp - \mathbf{u}'_\perp\| < \delta_1, |\langle\mathbf{w},\mathbf{u}_\perp\rangle - \langle\mathbf{w},\mathbf{u}'_\perp\rangle| < \delta_1$, and $|(2t-1)\|\mathbf{u}_\perp\|\,\theta - (2t-1)\|\mathbf{u}'_\perp\|\,\theta| < \delta_1$. Therefore,

$$\begin{aligned} \left\langle \mathbf{w}, \mathbf{u}'_\perp\right\rangle &\ge \langle \mathbf{w}, \mathbf{u}_\perp\rangle - \delta_1 = (2t-1)\|\mathbf{u}_\perp\|\,\theta - \delta_1 \\ &\ge (2t-1)\|\mathbf{u}'_\perp\|\,\theta - 2\delta_1. \end{aligned}$$

Next, let $\mathbf{w}' = p\mathbf{w} + q\theta\mathbf{u}'_\perp/\|\mathbf{u}'_\perp\|$, where $p = \left((2-2t)\,\theta\|\mathbf{u}'_\perp\|\right)/\left(\theta\|\mathbf{u}'_\perp\| - \langle\mathbf{w},\mathbf{u}'_\perp\rangle\right) \le 1$ and $q = 1-p$. We then have

$$
\begin{aligned}
\left\langle \mathbf{w}', \mathbf{u}'_{\perp} \right\rangle &= \frac{(2-2t)\,\theta\,\|\mathbf{u}'_{\perp}\|}{\theta\left\|\mathbf{u}'_{\perp}\right\| - \left\langle \mathbf{w}, \mathbf{u}'_{\perp}\right\rangle}\left\langle \mathbf{w}, \mathbf{u}'_{\perp}\right\rangle + \left(1 - \frac{(2-2t)\,\theta\,\|\mathbf{u}'_{\perp}\|}{\theta\left\|\mathbf{u}'_{\perp}\right\| - \left\langle \mathbf{w}, \mathbf{u}'_{\perp}\right\rangle}\right)\theta\left\|\mathbf{u}'_{\perp}\right\| \\
&= (2t-2)\,\theta\left\|\mathbf{u}'_{\perp}\right\| + \theta\left\|\mathbf{u}'_{\perp}\right\| = (2t-1)\,\theta\left\|\mathbf{u}'_{\perp}\right\|
\end{aligned}
$$

and

$$
\begin{aligned}
\left\|\mathbf{w} - \mathbf{w}'\right\| &= \left\|\mathbf{w} - \left(p\mathbf{w} + q\theta\mathbf{u}'_{\perp}/\left\|\mathbf{u}'_{\perp}\right\|\right)\right\| \leq \|q\mathbf{w}\| + q\theta \\
&\leq 2q \\
&= 2\left(1 - \frac{(2-2t)\,\theta\,\|\mathbf{u}'_{\perp}\|}{\theta\left\|\mathbf{u}'_{\perp}\right\| - \left\langle \mathbf{w}, \mathbf{u}'_{\perp}\right\rangle}\right) \\
&= 2\frac{(2t-1)\,\theta\,\|\mathbf{u}'_{\perp}\| - \langle \mathbf{w}, \mathbf{u}'_{\perp}\rangle}{\theta\left\|\mathbf{u}'_{\perp}\right\| - \left\langle \mathbf{w}, \mathbf{u}'_{\perp}\right\rangle} \\
&\leq 2\frac{2\delta_1}{\theta\left\|\mathbf{u}'_{\perp}\right\| - (2t-1)\,\theta\left\|\mathbf{u}'_{\perp}\right\|} \\
&\leq \frac{4\delta_1}{2\delta\theta\left\|\mathbf{u}'_{\perp}\right\|} \leq \frac{2\delta_1}{\delta\theta\left\|\mathbf{u}_{\perp}\right\| - \delta_1} < \epsilon.
\end{aligned}
$$

For the case of $\langle \mathbf{w}, \mathbf{u}'_{\perp}\rangle \geq (2t-1)\,\|\mathbf{u}'_{\perp}\|\,\theta$, we just need to replace the p above by $2t\,\|\mathbf{u}'_{\perp}\|\,\theta/(\langle \mathbf{w}, \mathbf{u}'_{\perp}\rangle + \theta\,\|\mathbf{u}'_{\perp}\|)$ and $\mathbf{w}'$ by $p\mathbf{w} - q\theta\mathbf{u}'_{\perp}/\|\mathbf{u}'_{\perp}\|$ and repeat the discussion. This shows that $\sup_{\mathbf{w}\in J_t^{\mathbf{u}}} \inf_{\mathbf{w}'\in J_t^{\mathbf{u}'}} \|\mathbf{w} - \mathbf{w}'\| \to 0$. Exchanging the role of $\mathbf{u}$ and $\mathbf{u}'$, we can show that $\sup_{\mathbf{w}'\in J_t^{\mathbf{u}'}} \inf_{\mathbf{w}\in J_t^{\mathbf{u}}} \|\mathbf{w} - \mathbf{w}'\| \to 0$. This finishes the proof.

Lemma 11. *Define the function $\varphi_{\mathbf{u}}(t) = dist\left(\mathbf{0}, \mathbf{Z}\left(I_t^{\mathbf{u}}\right)\right)$. Then $\varphi_{\mathbf{u}}$ is strictly convex.*

Proof of Lemma 11

For any $t_1, t_2 \in [0,1]$, suppose that $\mathbf{z}_i \in \mathbf{Z}\left(I_{t_i}^{\mathbf{u}}\right)$, $\|\mathbf{z}_i\| = dist\left(\mathbf{0}, \mathbf{Z}\left(I_{t_i}^{\mathbf{u}}\right)\right)$, $i = 1, 2$. Then for $p \in (0,1)$, $q = 1-p$, we have $p\mathbf{z}_1 + q\mathbf{z}_2 \in \mathbf{Z}\left(I_{pt_1+qt_2}^{\mathbf{u}}\right)$. Thus $\varphi_{\mathbf{u}}(pt_1 + qt_2) = dist\left(\mathbf{0}, \mathbf{Z}\left(I_{pt_1+qt_2}^{\mathbf{u}}\right)\right) \leq \|p\mathbf{z}_1 + q\mathbf{z}_2\| \leq p\,\|\mathbf{z}_1\| + q\,\|\mathbf{z}_2\|$. Since $\mathbf{0}, \mathbf{z}_1$ and $\mathbf{z}_2$ cannot be on the same line, the last inequality is strict. Hence, $\varphi_{\mathbf{u}}$ is strictly convex.

Lemma 12. *Let U, V be two closed, bounded and convex subsets of a Hilbert space $\mathbf{H}$, $\mathbf{h} \in \mathbf{H}$ and $dist(\mathbf{h}, U) = l$. Suppose that $D(U, V) < \epsilon$ for some $0 <$*

$\epsilon < l/2$. *Let* $\mathbf{h}_1 \in U, \mathbf{h}_2 \in V$ *satisfy* $\|\mathbf{h}_1 - \mathbf{h}\| = dist(\mathbf{h}, U)$ *and* $\|\mathbf{h}_2 - \mathbf{h}\| = dist(\mathbf{h}, V)$. *Then* $\|\mathbf{h}_1 - \mathbf{h}_2\| < 4\sqrt{l}\sqrt{\epsilon}$.

Proof of Lemma 12

Clearly, $dist(\mathbf{h}, V) > dist(\mathbf{h}, U) - \epsilon$. Since $D(U, V) < \epsilon$, we find $\mathbf{h}' \in V$ such that $\|\mathbf{h}_1 - \mathbf{h}'\| < \epsilon$. Thus $\left\|\mathbf{h} - \mathbf{h}'\right\| < \|\mathbf{h} - \mathbf{h}_1\| + \epsilon < \|\mathbf{h} - \mathbf{h}_2\| + 2\epsilon$. Suppose that $\mathbf{h}' - \mathbf{h}_2 = \mathbf{g}_1 + \mathbf{g}_2$, where $\mathbf{g}_1 = p(\mathbf{h} - \mathbf{h}_2)$, and $\mathbf{g}_2 \perp \mathbf{g}_1$. Now assume that $p > 0$. Let $\mathbf{h}_\alpha = (1-\alpha)\mathbf{h}_2 + \alpha\mathbf{h}'$, and

$$\begin{aligned} h(\alpha) &= \|\mathbf{h} - \mathbf{h}_\alpha\|^2 \\ &= (1 - \alpha p)^2 \|\mathbf{h} - \mathbf{h}_2\|^2 + \alpha^2 \|\mathbf{g}_2\|^2. \end{aligned}$$

We have

$$h'(\alpha) = 2\|\mathbf{g}_2\|^2 \alpha - 2p(1 - \alpha p),$$

which gives $h'(0) = -2p < 0$, which contradicts the fact that $\mathbf{h}_2$ achieves the distance between $\mathbf{h}$ and V. This shows that $p \le 0$. Thus

$$\begin{aligned} \left\|\mathbf{h} - \mathbf{h}'\right\| &= \sqrt{(\|\mathbf{h} - \mathbf{h}_2\| + \|\mathbf{g}_1\|)^2 + \|\mathbf{g}_2\|^2} \\ &\ge \max\left\{\sqrt{\|\mathbf{h} - \mathbf{h}_2\|^2 + \|\mathbf{g}_2\|^2}, \|\mathbf{h} - \mathbf{h}_2\| + \|\mathbf{g}_1\|\right\}, \end{aligned}$$

which gives $\|\mathbf{h} - \mathbf{h}_2\| + \|\mathbf{g}_1\| < \|\mathbf{h} - \mathbf{h}_2\| + 2\epsilon$. Thus $\|\mathbf{g}_1\| < 2\epsilon$ and

$$\sqrt{\|\mathbf{h} - \mathbf{h}_2\|^2 + \|\mathbf{g}_2\|^2} < \|\mathbf{h} - \mathbf{h}_2\| + 2\epsilon,$$

or

$$\|\mathbf{g}_2\|^2 < 4\epsilon^2 + 4\epsilon \|\mathbf{h} - \mathbf{h}_2\|.$$

Thus

$$\begin{aligned} \left\|\mathbf{h}' - \mathbf{h}_2\right\| &= \sqrt{\|\mathbf{g}_1\|^2 + \|\mathbf{g}_2\|^2} \\ &= 2\sqrt{2\epsilon^2 + \epsilon \|\mathbf{h} - \mathbf{h}_2\|} \\ &\le 3\sqrt{l}\sqrt{\epsilon}. \end{aligned}$$

Now,

$$\begin{aligned} \|\mathbf{h}_1 - \mathbf{h}_2\| &\le \left\|\mathbf{h}_1 - \mathbf{h}'\right\| + \left\|\mathbf{h}' - \mathbf{h}_2\right\| \\ &< \epsilon + 3\sqrt{l}\sqrt{\epsilon} \\ &\le 4\sqrt{l}\sqrt{\epsilon}. \end{aligned}$$

Lemma 13. *For every pair* $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathcal{F}$, *there are at most two choices for* $\mathbf{w}_a(\mathbf{x}_2 - \mathbf{x}_1)$; *and if there are two choices, say,* $\mathbf{w}_1$ *and* $\mathbf{w}_2$, *then* $\mathbf{w}_1$ *and* $\mathbf{w}_2$ *have different signs.*

Proof of Lemma 13

Let $\phi_{\mathbf{u}}(t) = a\alpha + (2t-1)\theta\|\mathbf{u}_\perp\| = \langle \mathbf{w}, \mathbf{u}\rangle$ for some $\mathbf{w} \in I_t^{\mathbf{u}}$. First, suppose that $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_0) = 0$ for some $t_0 \in (0,1)$. We have $G_a(\mathbf{x}_2 - \mathbf{x}_1) = \sup_{t\in[0,1]} \sup_{\mathbf{w}\in I_t^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}} |g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)|$. It is easy to see that $\sup_{\mathbf{w}\in I_t^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}} |g_{\mathbf{w}}(\mathbf{x}_2 - \mathbf{x}_1)| = \left|\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)\right| / \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$. Since $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t) < 0$ for $t \in [0, t_0)$ and $\phi_{\mathbf{k}(\mathbf{x}_2-\mathbf{x}_1)}(t) > 0$ for $t \in (t_0, 1]$, we have

$$
\begin{aligned}
G_a(\mathbf{x}_2 - \mathbf{x}_1) &= \sup_{t\in[0,1]} \frac{\left|\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)\right|}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)} \\
&= \max\left\{\sup_{t\in[0,t_0]} -\frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}, \sup_{t\in[t_0,1]} \frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}\right\}.
\end{aligned}
$$

Let k_1 be the smallest positive number k such that the line segment $y = k(t - t_0), t \in [t_0, 1]$ intersects $y = \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$. Since, by Lemma 11, $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$ is strictly convex, we know that the line segment intersects $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$ at only one point, say, $\left(t_1, \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)\right)$. Thus elementary calculations give us

$$
\sup_{t\in[t_0,1]} \frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)} = \frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)}.
$$

Similarly, we can find $t_2 \in [0, t_0]$ such that

$$
\sup_{t\in[0,t_0]} -\frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)} = -\frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)}.
$$

By convexity, we know that for each t, $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t) = \|\mathbf{Z}\mathbf{w}\|$ for only one $\mathbf{w} \in I_t^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$, and that there can be at most two choices for $\mathbf{w}_a(\mathbf{x}_2 - \mathbf{x}_1)$, they are the $\mathbf{w}_1$ and $\mathbf{w}_2$ that satisfy $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_i) = \|\mathbf{Z}\mathbf{w}_i\|, i = 1, 2$. Now suppose that $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t) \geq 0$ or $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t) \leq 0$ for all $t \in [0,1]$. Since these two cases can be discussed the same way, we will only do the former, $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t) \geq 0$.

If $\|(\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1))_\perp\| = 0$, then $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$ is constant. There is a unique $t' \in [0,1]$ that minimizes $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$, and we find $\mathbf{w}' \in I_{t'}^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$

such that $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t') = \|\mathbf{Z}\mathbf{w}'\|$. Then $\mathbf{w}'$ is the unique choice for $\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)$. If $\|(\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1))_\perp\| \neq 0$, then suppose that $t'' \notin (0,1)$ satisfies

$$a\alpha + \left(2t''-1\right)\theta\,\|(\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)_\perp)\| = 0.$$

Now, let k_2 be the smallest positive number k such that the line segment $k(t-t'') = y$, $t \in [0,1]$ intersects $y = \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$. Again, by the convexity of $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$, these two intersect at exactly one point, which is denoted by t_3. Suppose that $\mathbf{w}_3$ is the unique element in $I_{t_3}^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$ such that $\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_3) = \|\mathbf{Z}\mathbf{w}_3\|$. Then $\mathbf{w}_3$ is the unique choice for $\mathbf{w}_a((\mathbf{x}_2-\mathbf{x}_1))$. This proves the lemma.

Lemma 14. *Let $\mathbf{x}_1$, $\mathbf{x}_2$; $\mathbf{x}_1^{(n)}$, $\mathbf{x}_2^{(n)}$, $i = 1,\ 2$; $n = 1,2,\ldots$ be elements of $\mathcal{F}$ and $\mathbf{x}_i^{(n)} \to \mathbf{x}_i$, $\mathbf{K}\mathbf{x}_i^{(n)} \to \mathbf{K}\mathbf{x}_i$ for $i = 1,2$. Suppose that $\mathbf{w}_1, \mathbf{w}_2 \in W_a$ are the two choices of $\mathbf{w}_a(\mathbf{x}_2-\mathbf{x}_1)$ defined in Lemma 13 (they could be the same). Then for at least one of them, which, without loss of generality, we assume to be $\mathbf{w}_1$, we can find subsequences $\mathbf{x}_1^{(n_k)}$, $\mathbf{x}_2^{(n_k)}$ such that $\mathbf{Z}\mathbf{w}_a\left(\mathbf{x}_2^{(n_k)}-\mathbf{x}_1^{(n_k)}\right) \to \mathbf{Z}\mathbf{w}_1$.*

Proof of Lemma 14

We only give the proof for the case that a $t_0 \in (0,1)$ can be found satisfying $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_0) = 0$, where $\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$ is defined in the proof of the previous lemma. The proofs for other cases are similar, in fact simpler. Let $t_0^{(n)}$ be such that $\phi_{\mathbf{K}\left(\mathbf{x}_2^{(n)}-\mathbf{x}_1^{(n)}\right)}\left(t_0^{(n)}\right) = 0$. It can be easily shown that $\phi_{\mathbf{K}\left(\mathbf{x}_2^{(n)}-\mathbf{x}_1^{(n)}\right)}(t) \to \phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)$ if $\mathbf{K}\left(\mathbf{x}_2^{(n)}-\mathbf{x}_1^{(n)}\right) \to \mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)$, and this convergence is uniform in t for $t \in [0,1]$. This gives $t_0^{(n)} \to t_0$. Let t_1, t_2 be such that $0 \le t_1 < t_0 < t_2 \le 1$ and

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)} = \sup_{t\in[0,t_0]} \frac{-\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)},$$

$$\frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)} = \sup_{t\in[t_0,1]} \frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t)},$$

and $\mathbf{w}_1, \mathbf{w}_2$ are the elements in $I_{t_1}^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$, $I_{t_2}^{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}$ respectively satisfying $\|\mathbf{Z}\mathbf{w}_1\| = \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)$ and $\|\mathbf{Z}\mathbf{w}_2\| = \varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)$. By discarding the initial terms, we can assume that $t_0^{(n)} \in (0,1)$ for all n. Similar to t_1, t_2 and $\mathbf{w}_1, \mathbf{w}_2$, we define $0 \leq t_1^{(n)} < t_0^{(n)} < t_2^{(n)} \leq 1$ and $\mathbf{w}_1^{(n)} \in I_{t_1^{(n)}}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}, \mathbf{w}_2^{(n)} \in I_{t_2^{(n)}}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}$ satisfying

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}\left(t_1^{(n)}\right)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}\left(t_1^{(n)}\right)} = \sup_{t \in \left[0, t_0^{(n)}\right]} \frac{-\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)},$$

$$\frac{\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}\left(t_2^{(n)}\right)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}\left(t_2^{(n)}\right)} = \sup_{t \in \left[t_0^{(n)}, 1\right]} \frac{\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)},$$

and $\left\|\mathbf{Z}\mathbf{w}_i^{(n)}\right\| = \varphi_{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}\left(t_i^{(n)}\right)$, $i = 1, 2$. By Lemma 10 it can be shown that $\varphi_{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}(t) \to \varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)$ uniformly in $t \in [0,1]$. With the uniform (over t) convergence of $\varphi_{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}(t)$ and $\phi_{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}(t)$ and the strict convexity of $\varphi_{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}(t)$ and $\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t)$, we can show that $t_1^{(n)} \to t_1$ and $t_2^{(n)} \to t_2$. If

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)} \neq \frac{\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)},$$

then there will be only one choice for $\mathbf{w}_a(\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1))$. If

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)} > \frac{\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)},$$

then $\mathbf{w}_1$ is chosen, and eventually, we will also choose $\mathbf{w}_1^{(n)}$ for $\mathbf{w}_a\left(\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)\right)$. By Lemma 10 we have

$$D\left(I_{t_1}^{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}, I_{t_1}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}\right) \to 0.$$

Since $t_1^{(n)} \to t_1$, it is easy to show that

$$D\left(I_{t_1}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}, I_{t_1^{(n)}}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}\right) \to 0.$$

Thus

$$D\left(I_{t_1}^{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}, I_{t_1^{(n)}}^{\mathbf{K}\left(\mathbf{x}_2^{(n)} - \mathbf{x}_1^{(n)}\right)}\right) \to 0.$$

By Lemma 12, we know that $\mathbf{Z}\mathbf{w}_1^{(n)} \to \mathbf{Z}\mathbf{w}_1$. For the case of

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_1)} < \frac{\phi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)}(t_2)},$$

the proof is similar. Now suppose that

$$\frac{-\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_1)} = \frac{\phi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)}{\varphi_{\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1)}(t_2)}.$$

Since we have to pick either $\mathbf{w}_1^{(n)}$ or $\mathbf{w}_2^{(n)}$ for $\mathbf{w}_a\left(\mathbf{K}\left(\mathbf{x}_2^{(n)}-\mathbf{x}_1^{(n)}\right)\right)$, we must pick infinitely many from either of the two sequences $\left(\mathbf{w}_1^{(n)}\right)$ or $\left(\mathbf{w}_2^{(n)}\right)$. Without loss of generality, we assume that we pick an infinite number elements from $\left(\mathbf{w}_1^{(n)}\right)$ and form a subsequence $\left(\mathbf{w}_1^{(n_k)}\right)$. Then we pick $\mathbf{w}_1$ for $\mathbf{w}_a(\mathbf{K}(\mathbf{x}_2-\mathbf{x}_1))$. Again, we can show that $\mathbf{Z}\mathbf{w}_1^{(n_k)} \to \mathbf{Z}\mathbf{w}_1$.

With these results, we can now prove Theorem 1 .

Proof of Theorem 1

For simplicity, let $\mathbf{x}_i = \mathbf{x}_i(a), i = 1,2$. Suppose that $\mathbf{x} \in \mathcal{F}$. We want to show that $R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}_0) \geq \left[E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)\right]^2$, where $\hat{L}_0 = \hat{L}(c_0\mathbf{w}_0, d_0)$. We have

$$E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 = bias\left(\hat{L}_0, \mathbf{x}\right)^2 + \|c_0\mathbf{Z}\mathbf{w}_0\|^2$$

and

$$\begin{aligned} R_{[\mathbf{x}_1,\mathbf{x}_2]}(\hat{L}_0) &= E_{\mathbf{x}_i}\left(\hat{L}_0 - L(\mathbf{x}_i)\right)^2 \\ &= bias\left(\hat{L}_0, \mathbf{x}_i\right)^2 + \|c_0\mathbf{Z}\mathbf{w}_0\|^2, i = 1,2. \end{aligned}$$

Therefore, we only need to show that $\left|bias\left(\hat{L}_0, \mathbf{x}\right)\right| \leq \left|bias\left(\hat{L}_0, \mathbf{x}_i\right)\right|, i = 1,2$. Since $bias\left(\hat{L}_0, \mathbf{x}_1\right)$ and $bias\left(\hat{L}_0, \mathbf{x}_2\right)$ have opposite signs, without loss of generality, we may assume that $bias\left(\hat{L}_0, \mathbf{x}_2\right)$ and $bias\left(\hat{L}_0, \mathbf{x}\right)$ have the same sign. We will only deal with the case in which $bias\left(\hat{L}_0, \mathbf{x}_2\right) > 0$. The other case can be proven similarly. Let $\mathbf{x}_h = h\mathbf{x} + (1-h)\mathbf{x}_2$. If $bias\left(\hat{L}_0, \mathbf{x}\right) > bias\left(\hat{L}_0, \mathbf{x}_2\right)$, then let $bias\left(\hat{L}_0, \mathbf{x}\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right) = \Delta$. We have

$$\begin{aligned} bias\left(\hat{L}_0, \mathbf{x}_h\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right) &= h\left(bias\left(\hat{L}_0, \mathbf{x}\right) - bias\left(\hat{L}_0, \mathbf{x}_2\right)\right) \\ &= h\Delta. \end{aligned} \tag{15}$$

Our technique will be to show that this cannot be true. Let $\psi(s,t) = s^2/(1+t^2)$, $\mathbf{s}_h = \mathbf{x}_h - \mathbf{x}_1$, $\mathbf{s}_0 = \mathbf{x}_2 - \mathbf{x}_1$, $\mathbf{s} = \mathbf{x} - \mathbf{x}_2$, and $\mathbf{w}_h = \mathbf{w}_a(\mathbf{s}_h)$,

$\mathbf{w}_0 = \mathbf{w}_a(\mathbf{s}_0)$. Since $\mathbf{x}_h \to \mathbf{x}_2$, by the continuity of $\mathbf{K}$ on $\mathcal{F}$, we have $\mathbf{K}\mathbf{x}_h \to \mathbf{K}\mathbf{x}_2$, and thus $\mathbf{K}\mathbf{s}_h \to \mathbf{K}\mathbf{s}_0$. Hence, by Lemma 14 we can find $\mathbf{w}_{h_k}$ such that $\mathbf{Z}\mathbf{w}_{h_k} \to \mathbf{Z}\mathbf{w}_0$. Now we can find a subsequence of $\mathbf{w}_{h_k}$, which we write as $\mathbf{w}_{h_k}$ also for simplicity, such that $\mathbf{w}_{h_k} \overset{w}{\to} \mathbf{w}'$ for some $\mathbf{w}'$ in W_a. It is easy to see that $\mathbf{w}' = \mathbf{w}_0$. Thus

$$g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k}) = \frac{\langle \mathbf{w}_{h_k}, \mathbf{K}\mathbf{s}_{h_k}\rangle}{\|\mathbf{Z}\mathbf{w}_{h_k}\|} \to \frac{\langle \mathbf{w}_0, \mathbf{K}\mathbf{s}_0\rangle}{\|\mathbf{Z}\mathbf{w}_0\|} = g_{\mathbf{w}_0}(\mathbf{s}_0),$$

which means that they eventually will have the same sign. With this and Lemma 9, we have $g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k}) - g_{\mathbf{w}_0}(\mathbf{s}_0) = O(h_k)$. Also,

$$\begin{aligned}\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0}\right)(\mathbf{s}_0) &= g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k}) - g_{\mathbf{w}_0}(\mathbf{s}_0) - g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k}) + g_{\mathbf{w}_{h_k}}(\mathbf{s}_0) \\ &= O(h_k).\end{aligned}$$

Let $\psi^0 = \psi(L_1(\mathbf{s}_0/2), g_{\mathbf{w}_0}(\mathbf{s}_0))$, $\psi_1(s,t) = \frac{\partial\psi}{\partial s}(s,t)$, $\psi_2(s,t) = \frac{\partial\psi}{\partial t}(s,t)$, $\psi_1^0 = \psi_1(L_1(\mathbf{s}_0/2), g_{\mathbf{w}_0}(\mathbf{s}_0))$, and $\psi_2^0 = \psi_2(L_1(\mathbf{s}_0/2), g_{\mathbf{w}_0}(\mathbf{s}_0))$. We then have

$$\begin{aligned}&\psi\left(L_1(\mathbf{s}_{h_k}/2), g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k})\right) \\ &= \psi^0 + \psi_1^0 L_1(h_k\mathbf{s}/2) + \psi_2^0\left(g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k}) - g_{\mathbf{w}_0}(\mathbf{s}_0)\right) + o(h_k) \\ &= \psi^0 + \psi_2^0\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0}\right)(\mathbf{s}_0) + \psi_1^0 L_1(h_k\mathbf{s}/2) \\ &+ h_k\psi_2^0\left(g_{\mathbf{w}_{h_k}} - g_{\mathbf{w}_0}\right)(\mathbf{s}) + \psi_2^0 g_{\mathbf{w}_0}(h_k\mathbf{s}) + o(h_k). \\ &= \psi\left(L_1(\mathbf{s}_0/2), g_{\mathbf{w}_{h_k}}(\mathbf{s}_0)\right) + \psi_1^0 L_1(h_k\mathbf{s}/2) + \psi_2^0 g_{\mathbf{w}_0}(h_k\mathbf{s}) + o(h_k).\end{aligned}$$

By the definition of $\mathbf{x}_2(a), \mathbf{x}_1(a)$, and $\mathbf{w}_0$, we know that

$$\psi\left(L_1(\mathbf{s}_{h_k}/2), g_{\mathbf{w}_{h_k}}(\mathbf{s}_{h_k})\right) \le \psi^0 \le \psi\left(L_1(\mathbf{s}_0/2), g_{\mathbf{w}_{h_k}}(\mathbf{s}_0)\right).$$

Thus $\psi_1^0 L_1(h_k\mathbf{s}/2) + \psi_2^0 g_{\mathbf{w}_0}(h_k\mathbf{s}) + o(h_k) \le 0$. However,

$$\begin{aligned}&\psi_1^0 L_1(h_k\mathbf{s}/2) + \psi_2^0 g_{\mathbf{w}_0}(h_k\mathbf{s}) \\ &= \frac{2L_1(\mathbf{s}_0/2)}{1+g_{\mathbf{w}0}^2(\mathbf{s}_0)} L_1(h_k\mathbf{s}/2) - \frac{2\left[L_1(\mathbf{s}_0/2)\right]^2 g_{\mathbf{w}_0}(\mathbf{s}_0)}{\left(1+g_{\mathbf{w}_0}^2(\mathbf{s}_0)\right)^2} g_{\mathbf{w}_0}(h_k\mathbf{s}) \\ &= \frac{2L_1(\mathbf{s}_0/2)}{1+g_{\mathbf{w}0}^2(\mathbf{s}_0)}\left(L_1(h_k\mathbf{s}/2) - c_0\langle \mathbf{w}_0, h_k\mathbf{K}\mathbf{s}/2\rangle\right).\end{aligned}$$

We have $bias\left(\hat{L}_0, \mathbf{x}_2\right) > 0$. Also

$$\left|E_{\mathbf{x}_2}\left(\hat{L}_0\right) - L\left(\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\right)\right| < \left|L\left(\mathbf{x}_2\right) - L\left(\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\right)\right|$$

and they have the same sign. Therefore, we must have

$$L\left(\mathbf{x}_2\right) - L\left(\left(\mathbf{x}_1 + \mathbf{x}_2\right)/2\right) = L_1\left(\mathbf{s}_0/2\right) < 0.$$

Hence,

$$\frac{2L_1\left(\mathbf{s}_0/2\right)}{1 + g_{\mathbf{w}0}^2\left(\mathbf{s}_0\right)}\left(L_1\left(h_k\mathbf{s}/2\right) - c_0\left\langle\mathbf{w}_0, h_k\mathbf{K}\mathbf{s}/2\right\rangle\right) + o\left(h_k\right) \leq 0,$$

which gives

$$L_1\left(h_k\mathbf{s}/2\right) - c_0\left\langle\mathbf{w}_0, h_k\mathbf{K}\mathbf{s}/2\right\rangle + \frac{1 + g_{\mathbf{w}0}^2\left(\mathbf{s}_0\right)}{2L_1\left(\mathbf{s}_0/2\right)}o\left(h_k\right) \geq 0,$$

or

$$L_1\left(h_k\mathbf{s}\right) - c_0\left\langle\mathbf{w}_0, h_k\mathbf{K}\mathbf{s}\right\rangle + o\left(h_k\right) \geq 0.$$

But by (15), we know that $L_1\left(h_k\mathbf{s}\right) - c_0\left\langle\mathbf{w}_0, h_k\mathbf{K}\mathbf{s}\right\rangle = -h_k\Delta$, giving a contradiction to the last inequality above, hence proving the desired result.

Proof of Theorem 2

Showing that the closedness condition can be removed is done exactly the same way as the corresponding part in the proof of Theorem 2 in Donoho (1994). We will now show that the boundedness constraint can also be dropped. Let $\mathcal{F}_k = \{\mathbf{x} \in \mathcal{F} : \|\mathbf{x}\| \leq k\}$. It is easy to see that $\rho_a\left(\mathcal{F}_k\right) \uparrow \rho_a\left(\mathcal{F}\right)$. Clearly, $\inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) \geq \rho_a\left(\mathcal{F}\right)$. We can assume that $\rho_a\left(\mathcal{F}\right) < \infty$, for, if not, the result is trivial. Also, for a non-trivial setting, $\rho_a\left(\mathcal{F}_k\right) > 0$ for sufficiently large k, so that by ignoring the first few terms, we can assume that $\rho_a\left(\mathcal{F}_k\right) > 0$ for all k. Since $\mathcal{F}_k \neq \phi$ for sufficiently large k, we will assume this to be true for all k. By Theorem 1, we can find $\mathbf{w}_k \in W_a$ and $c_k, d_k \in \mathbb{R}$ such that the estimator $\hat{L}_k = \hat{L}\left(c_k\mathbf{w}_k, d_k\right)$ satisfies $R_{\mathcal{F}_k}\left(\hat{L}_k\right) = \inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}_k}\left(\hat{L}\left(c\mathbf{w}, d\right)\right) = \rho_a\left(\mathcal{F}_k\right)$. Now, we will show that $\sup_k |c_k| < \infty$. In fact,

$$\begin{aligned}\rho_a\left(\mathcal{F}_k\right) &= \inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}_k}\left(\hat{L}\right) = R_{\mathcal{F}_k}\left(\hat{L}_k\right) \\ &= \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2 + \left\|c_k\mathbf{Z}\mathbf{w}_k\right\|^2 \geq \left\|c_k\mathbf{Z}\mathbf{w}_k\right\|^2,\end{aligned}$$

so that

$$\sup_k c_k^2 \|\mathbf{Z}\mathbf{w}_k\|^2 \leq \sup_k \rho_a (\mathcal{F}_k) < \infty.$$

Since $\inf_k \|\mathbf{Z}\mathbf{w}_k\| \geq \inf_{\mathbf{w}\in W_a} \|\mathbf{Z}\mathbf{w}\| > 0$, we have $\sup_k c_k^2 < \infty$. Next, we will show that $\sup_k |d_k| < \infty$. From the above discussion, we observe that

$$\begin{aligned}\infty > \sup_k \rho_a (\mathcal{F}_k) &\geq \sup_k \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2 \\ &\geq \sup_k bias\left(\hat{L}_k, \mathbf{x}_0\right)^2 = \sup_k \left(\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}_0\rangle + d_k - L(\mathbf{x}_0)\right)^2,\end{aligned}$$

where $\mathbf{x}_0$ is any vector in $\mathcal{F}_1$. Since $\sup_k |\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}_0\rangle - L(\mathbf{x}_0)| < \infty$, we must have $\sup_k |d_k| < \infty$. Thus we can find a subsequence which for simplicity, we continue to write as c_k, d_k, and $\mathbf{w}_k$, such that $c_k \to c_0, d_k \to d_0$, and $\mathbf{w}_k \to \mathbf{w}_0 \in W_a$ weakly. We claim that $\hat{L}_0 = \hat{L}(c_k\mathbf{w}_k, d_k)$ is the affine estimator that we are looking for. If fact, we have $\|c_0\mathbf{Z}\mathbf{w}_0\| \leq \liminf_k \|c_k\mathbf{Z}\mathbf{w}_k\| \leq \limsup_k \|c_k\mathbf{Z}\mathbf{w}_k\|$, and

$$\begin{aligned}bias\left(\hat{L}_0, \mathbf{x}\right) &= |\langle c_0\mathbf{w}_0, \mathbf{K}\mathbf{x}\rangle + d_0 - L(\mathbf{x})| = \lim_k |\langle c_k\mathbf{w}_k, \mathbf{K}\mathbf{x}\rangle + d_k - L(\mathbf{x})| \\ &= \lim_k bias\left(\hat{L}_k, \mathbf{x}\right) \leq \liminf \sup_{\mathbf{x}'\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}'\right), \forall \mathbf{x} \in \mathcal{F}.\end{aligned}$$

Here we used the fact that $\mathbf{x} \in \mathcal{F}_k$ for k large enough. Hence,

$$\sup_{\mathbf{x}\in\mathcal{F}} bias\left(\hat{L}_0, \mathbf{x}\right)^2 \leq \liminf \sup_{\mathbf{x}\in\mathcal{F}_k} bias\left(\hat{L}_k, \mathbf{x}\right)^2.$$

Thus we have

$$\begin{aligned}\inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right) &\leq \sup_{\mathbf{x}\in\mathcal{F}} E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 \\ &= \sup_{\mathbf{x}\in\mathcal{F}} bias\left(\hat{L}_0, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}_0\|^2 \\ &\leq \liminf \sup_{\mathbf{x}\in\mathcal{F}_k}\left(bias\left(\hat{L}_k, \mathbf{x}\right)^2 + \|\mathbf{Z}\mathbf{w}_k\|^2\right) \\ &= \lim_k \rho_a(\mathcal{F}_k) \leq \inf_{\mathbf{w}\in W_a, c, d\in\mathbb{R}} R_{\mathcal{F}}\left(\hat{L}\right)\end{aligned}$$

proving the desired result.

Proof of Theorem 3

Clearly $M = \inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) \leq \lim_{a\to 0} \rho_a(\mathcal{F})$, since $\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) \leq \rho_a(\mathcal{F})$ for all $a \in (0,1)$. Suppose that $\hat{L}_0 = \hat{L}(\mathbf{w}_0, d_0)$ is an affine estimator such that $R_{\mathcal{F}}\left(\hat{L}_0\right) < \infty$. We will use $\mathbf{w}_0/\|\mathbf{w}_0\|$ as the $\mathbf{v}$ in (6). For any $\epsilon \in (0,1)$, we find an affine estimator $\hat{L}_\epsilon = \hat{L}(\mathbf{w}_\epsilon, d_\epsilon)$ such that $R_{\mathcal{F}}\left(\hat{L}_\epsilon\right) < M + \epsilon$. If $\mathbf{w}_\epsilon \notin \ker(\mathbf{v})$, then there exists $a_\epsilon > 0$ and c_ϵ such that $\mathbf{w}_\epsilon \in c_\epsilon W_{a_\epsilon}$. Then $\rho_{a_\epsilon}(\mathcal{F}) \leq R_{\mathcal{F}}\left(\hat{L}_\epsilon\right) < M + \epsilon$. If $\mathbf{w}_\epsilon \in \ker(\mathbf{v})$, then let $\hat{L}' = p\hat{L}_0 + q\hat{L}_\epsilon$ where $0 < p < \epsilon \Big/ \left(M + \epsilon + R_{\mathcal{F}}\left(\hat{L}_0\right)\right)$ and $q = 1 - p$. For any $\mathbf{x} \in \mathcal{F}$, we have

$$
\begin{aligned}
E_{\mathbf{x}}\left(\hat{L}' - L(\mathbf{x})\right)^2 &= E_{\mathbf{x}}\left(p\hat{L}_0 - pL(\mathbf{x}) + q\hat{L}_\epsilon - qL(\mathbf{x})\right)^2 \\
&\leq p^2 E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 + q^2 E_{\mathbf{x}}\left(\hat{L}_\epsilon - L(\mathbf{x})\right)^2 \\
&+ pq\left(E_{\mathbf{x}}\left(\hat{L}_0 - L(\mathbf{x})\right)^2 + E_{\mathbf{x}}\left(\hat{L}_\epsilon - L(\mathbf{x})\right)^2\right) \\
&\leq \epsilon + q^2 E_{\mathbf{x}}\left(\hat{L}_\epsilon - L(\mathbf{x})\right)^2 + \epsilon \\
&\leq 2\epsilon + E_{\mathbf{x}}\left(\hat{L}_\epsilon - L(\mathbf{x})\right)^2 < M + 3\epsilon.
\end{aligned}
$$

Since $R_{\mathcal{F}}\left(\hat{L}'\right) \geq \rho_a(\mathcal{F})$ for some $a > 0$, we have

$$\lim_{a\to 0} \rho_a(\mathcal{F}) \leq M + 3\epsilon.$$

This finishes the proof.

Proof of Lemma 4

Since $\rho_0(\mathbf{x}_2 - \mathbf{x}_1) \leq \rho_a(\mathbf{x}_2 - \mathbf{x}_1)$ for any $0 < a < 1$, it is clear that

$$\lim_{a\to 0} \rho_a(\mathcal{F}) \geq \rho_0(\mathcal{F}).$$

Suppose that there is a positive ϵ such that $\lim_{a\to 0} \rho_a(\mathcal{F}) > \rho_0(\mathcal{F}) + \epsilon$. Then there exist sequences $\mathbf{x}_1^n$, $\mathbf{x}_2^n \in \mathcal{F}$, and $a_n \downarrow 0$ such that $\rho_{a_n}(\mathbf{x}_2^n - \mathbf{x}_1^n) > \rho_0(\mathcal{F}) + \epsilon, n = 1, 2, \ldots$. By passing to a subsequence, we can assume that $\mathbf{x}_1^n \xrightarrow{w} \tilde{\mathbf{x}}_1$ and $\mathbf{x}_1^n \xrightarrow{w} \tilde{\mathbf{x}}_2$, where $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2 \in \mathcal{F}$. Clearly, $\rho_0(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) \leq \rho_0(\mathcal{F})$. We can find $\mathbf{w}_0 \in S(\mathbf{U})$ such that

$$\frac{\left[L_1\left((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2\right)\right]^2}{1 + g^2_{\mathbf{w}_0}(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)} < \rho_0(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) + \epsilon/3.$$

With Assumption 1 and Lemma 2, we have $[L_1(\mathbf{x}_2^n - \mathbf{x}_1^n)]^2 \to [L_1(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)]^2$. Also, $\langle \mathbf{w}_0, \mathbf{K}(\mathbf{x}_2^n - \mathbf{x}_1^n)\rangle^2 \to \langle \mathbf{w}_0, \mathbf{K}(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)\rangle^2$ so that we can find N such that for all $n \geq N$,

$$\frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} \leq \frac{[L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)} + \epsilon/3.$$

Now, suppose that $\mathbf{w}_0$ is not in $\ker(\mathbf{v})$. We can find N' such that when $n \geq N'$, $c\mathbf{w}_0 \in W_{a_n}$ for some $c \in [-1, 1]$. Thus for $n \geq \max\{N, N'\}$ we have that

$$\begin{aligned}
\rho_{a_n}(\mathbf{x}_2^n - \mathbf{x}_1^n) &\leq \frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} \\
&\leq \frac{[L_1((\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)} + \epsilon/3 \\
&\leq \rho_0(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) + 2\epsilon/3,
\end{aligned}$$

which is impossible. This shows that $\mathbf{w}_0 \in \ker(\mathbf{v})$. It is not hard to prove that we can find $\mathbf{w}_n \in W_{a_n}$ such that the sequence $\{\mathbf{w}_n\}$ converges to $\mathbf{w}_0$. Since $\mathbf{K}(\mathcal{F} - \mathcal{F})$ is bounded and $\mathbf{Z}$ is a bounded operator, we have $g_{\mathbf{w}_n}(\mathbf{x}_2 - \mathbf{x}_1) \to g_{\mathbf{w}_0}(\mathbf{x}_2 - \mathbf{x}_1)$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ and this convergence is uniform. Also, $\{L_1(\mathbf{x}_2^n - \mathbf{x}_1^n) : n = 1, 2, \ldots\}$ is bounded. Thus there exists N'' such that when $n \geq N''$, we can find $\mathbf{w}_n \in W_{a_n}$ close enough to $\mathbf{w}_0$ such that

$$\frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_n}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} \leq \frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} + \epsilon/3.$$

Hence, for $n \geq \max\{N, N', N''\}$, we have

$$\begin{aligned}
\rho_{a_n}(\mathbf{x}_2^n - \mathbf{x}_1^n) &\leq \frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_n}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} \\
&\leq \frac{[L_1((\mathbf{x}_2^n - \mathbf{x}_1^n)/2)]^2}{1 + g_{\mathbf{w}_0}^2(\mathbf{x}_2^n - \mathbf{x}_1^n)} + \epsilon/3 \\
&\leq \frac{L_1[(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)/2]^2}{1 + g_{\mathbf{w}_0}^2(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1)} + 2\epsilon/3 \\
&< \rho_a(\tilde{\mathbf{x}}_2 - \tilde{\mathbf{x}}_1) + \epsilon \\
&\leq \rho_0(\mathcal{F}) + \epsilon,
\end{aligned}$$

which contradicts the assumption. Thus

$$\lim_{a \to 0} \rho_a(\mathcal{F}) \leq \rho_0(\mathcal{F})$$

completing the proof.

Proof of Lemma 5

Let $\mathcal{F}_s = \mathcal{F} \cap B_s(\mathbf{X})$ where $B_s(\mathbf{X}) = \{\mathbf{x} : \|\mathbf{x}\| \leq s\}$. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$. Then since $\mathcal{F}$ is symmetric, $-\mathbf{x}_1, -\mathbf{x}_2$ are also in $\mathcal{F}$. Let $\mathbf{x}_1' = (\mathbf{x}_1 - \mathbf{x}_2)/2$, and $\mathbf{x}_2' = (\mathbf{x}_2 - \mathbf{x}_1)/2$. We see that $\mathbf{x}_1', \mathbf{x}_2' \in \mathcal{F}$, and $\mathbf{x}_2' - \mathbf{x}_1' = \mathbf{x}_2 - \mathbf{x}_1$.

Therefore, $\rho_a(\mathbf{x}_2 - \mathbf{x}_1) = \rho_a(\mathbf{x}_2' - \mathbf{x}_1') \leq \rho_a\left(\mathcal{F}_{\|(\mathbf{x}_2 - \mathbf{x}_1)/2\|}\right)$. Again, we only need to prove that $\lim_{a\to 0} \rho_a(\mathcal{F}) \leq \rho_0(\mathcal{F})$. If not, we find $\delta > 0$ and a sequence a_n such that $a_n \downarrow 0$ and

$$\lim_{n\to\infty} \rho_{a_n}(\mathcal{F}) = \tau(\mathcal{F}) \geq \rho_0(\mathcal{F}) + \delta. \tag{16}$$

Let$M > \sqrt{64\|\mathbf{Z}\|^2 \rho_0(\mathcal{F})/\delta + 32\|\mathbf{Z}\|^2}$, and $M' > \chi(M)$. By Lemma 4,

$$\lim_{n\to\infty} \rho_{a_n}(\mathcal{F}_{M'}) = \rho_0(\mathcal{F}_{M'}).$$

Hence, for $\epsilon \in (0, \delta/4)$we find m' such that

$$\begin{aligned}\rho_{a_{m'}}(\mathcal{F}_{M'}) &< \rho_0(\mathcal{F}_{M'}) + \epsilon \\ &\leq \rho_0(\mathcal{F}) + \epsilon\end{aligned} \tag{17}$$

and $a_{m'} < 1/4$. Equation (16) gives us $\rho_{a_{m'}}(\mathcal{F}) \geq \rho_0(\mathcal{F}) + \delta$. Thus there exists a pair $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{F}$ such that

$$\begin{aligned}\rho_{a_{m'}}(\mathbf{x}_2 - \mathbf{x}_1) &> \rho_{a_{m'}}(\mathcal{F}) - \epsilon \\ &\geq \rho_0(\mathcal{F}) + \delta - \epsilon.\end{aligned} \tag{18}$$

Let $\mathbf{x}_1' = (\mathbf{x}_1 - \mathbf{x}_2)/2$, and $\mathbf{x}_2' = (\mathbf{x}_2 - \mathbf{x}_1)/2$. If $\|(\mathbf{x}_2 - \mathbf{x}_1)/2\| \leq M'$, then $\mathbf{x}_1', \mathbf{x}_2' \in \mathcal{F}_{M'}$, and this gives

$$\begin{aligned}\rho_0(\mathcal{F}) + \delta - \epsilon &< \rho_{a_{m'}}(\mathbf{x}_2 - \mathbf{x}_1) \\ &= \rho_{a_{m'}}\left(\mathbf{x}_2' - \mathbf{x}_1'\right) \\ &\leq \rho_{a_{m'}}(\mathcal{F}_{M'}) \\ &< \rho_0(\mathcal{F}) + \epsilon,\end{aligned}$$

which is a contradiction. Hence, $\|(\mathbf{x}_2 - \mathbf{x}_1)/2\| > M'$.

Now, let $\hat{\mathbf{x}}_i = M'\mathbf{x}'_i / \|(\mathbf{x}_2 - \mathbf{x}_1)/2\|$, $i = 1, 2$. We have $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \in \mathcal{F}_{M'}$. Let $\mathbf{w}_0 = \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1) / \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|$. Then either $c\mathbf{w}_0 \in W_{a_{m'}}$ for some $c \in [-1, 1]$ or there exists $\mathbf{w}' \in W_{a_{m'}}$ such that $\|\mathbf{w}' - c\mathbf{w}_0\| < 2a_{m'}$ for $c = 1$ or -1. In either case, we can find a $\mathbf{w}'$ such that $\|\mathbf{w}' - c\mathbf{w}_0\| / \|c\mathbf{w}_0\| < 2a_{m'}$, with $0 \neq c \in [-1, 1]$. Thus

$$\begin{aligned}
|g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| &= \left| \frac{\langle \mathbf{w}', \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle}{\|\mathbf{Z}\mathbf{w}'\|} \right| \\
&= \left| \frac{\langle \mathbf{w}' - c\mathbf{w}_0 + c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0 + c\mathbf{w}_0)\|} \right| \\
&\geq \frac{|\langle c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle| - |\langle \mathbf{w}' - c\mathbf{w}_0, \mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)/2 \rangle|}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0)\| + \|c\mathbf{Z}\mathbf{w}_0\|} \\
&\geq \frac{|c| \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\| / 2 - |c| a_{m'} \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\|}{\|\mathbf{Z}(\mathbf{w}' - c\mathbf{w}_0)\| + \|c\mathbf{Z}\mathbf{w}_0\|} \\
&> \frac{c \|\mathbf{K}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)\| / 4}{2c \|\mathbf{Z}\|} \\
&\geq \frac{M}{4 \|\mathbf{Z}\|},
\end{aligned}$$

and

$$\left| g_{\mathbf{w}'}(\mathbf{x}'_2 - \mathbf{x}'_1) \right| = \frac{\|(\mathbf{x}_2 - \mathbf{x}_1)/2\|}{M'} |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| .$$

It can be shown that $\mathbf{w}_{a_{m'}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)$ and $\mathbf{w}_{a_{m'}}(\mathbf{x}'_2 - \mathbf{x}'_1)$ can be chosen to be the same, which is denoted $\tilde{\mathbf{w}}$. Therefore,

$$|g_{\tilde{\mathbf{w}}}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq |g_{\mathbf{w}'}(\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1)| \geq \frac{M}{4 \|\mathbf{Z}\|}$$

and

$$\left| g_{\tilde{\mathbf{w}}}(\mathbf{x}'_2 - \mathbf{x}'_1) \right| \geq \left| g_{\mathbf{w}'}(\mathbf{x}'_2 - \mathbf{x}'_1) \right| \geq \frac{M}{4 \|\mathbf{Z}\|},$$

giving

$$\begin{aligned}
&\left| \rho_{a_{m'}}(\mathbf{x}'_2 - \mathbf{x}'_1) - \frac{[L_1((\mathbf{x}'_2 - \mathbf{x}'_1)/2)]^2}{g^2_{\tilde{\mathbf{w}}}(\mathbf{x}'_2 - \mathbf{x}'_1)} \right| \\
&= \left| \frac{[L_1((\mathbf{x}'_2 - \mathbf{x}'_1)/2)]^2}{1 + g^2_{\tilde{\mathbf{w}}}(\mathbf{x}'_2 - \mathbf{x}'_1)} - \frac{[L_1((\mathbf{x}'_2 - \mathbf{x}'_1)/2)]^2}{g^2_{\tilde{\mathbf{w}}}(\mathbf{x}'_2 - \mathbf{x}'_1)} \right| \\
&= \rho_{a_{m'}}(\mathbf{x}'_2 - \mathbf{x}'_1) \frac{1}{g^2_{\tilde{\mathbf{w}}}(\mathbf{x}'_2 - \mathbf{x}'_1)} \\
&\leq \frac{16 \|\mathbf{Z}\|^2}{M^2} \rho_{a_{m'}}(\mathbf{x}'_2 - \mathbf{x}'_1) .
\end{aligned}$$

Similarly,

$$\left|\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)-\frac{\left[L_1((\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1)/2)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)}\right|\leq\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right).$$

However, noticing that

$$\begin{aligned}\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\mathbf{x}_2'-\mathbf{x}_1'\right)}&=\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{\left\langle\tilde{\mathbf{w}},\mathbf{x}_2'-\mathbf{x}_1'/2\right\rangle^2/\left\|\mathbf{Z}\tilde{\mathbf{w}}\right\|^2}\\&=\frac{\left[L_1\left(\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)/2\right)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)},\end{aligned}$$

we have

$$\begin{aligned}&\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right|\\&=\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\mathbf{x}_2'-\mathbf{x}_1'\right)}+\frac{\left[L_1\left(\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)/2\right)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)}-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right|\\&\leq\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{\left[L_1((\mathbf{x}_2'-\mathbf{x}_1')/2)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\mathbf{x}_2'-\mathbf{x}_1'\right)}\right|+\left|\frac{\left[L_1\left(\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)/2\right)\right]^2}{g_{\tilde{\mathbf{w}}}^2\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)}-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right|\\&\leq\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right)\\&\leq\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\right).\end{aligned}$$

Thus

$$\begin{aligned}\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)&\geq\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\\&\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\left|\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\rho_{a_{m'}}\left(\hat{\mathbf{x}}_2-\hat{\mathbf{x}}_1\right)\right|\\&\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\left(\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)+\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\right),\end{aligned}$$

or

$$\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\geq\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right)-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right),$$

giving

$$\left(1+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\rho_{a_{m'}}\left(\mathcal{F}_{M'}\right)\geq\left(1-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\rho_{a_{m'}}\left(\mathbf{x}_2'-\mathbf{x}_1'\right).$$

By (17) and (18), we have

$$\left(1+\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\left(\rho_0\left(\mathcal{F}\right)+\epsilon\right)\geq\left(1-\frac{16\left\|\mathbf{Z}\right\|^2}{M^2}\right)\left(\rho_0\left(\mathcal{F}\right)+\delta-\epsilon\right).$$

And this gives

$$\frac{64\,\|\mathbf{Z}\|^2\,\rho_0\left(\mathcal{F}\right)}{\delta}+32\,\|\mathbf{Z}\|^2\geq M^2$$

proving the lemma

4.3 Proofs for Section 3

For convenience, in the proofs of this and the next section, we shall use $K, K_1, \ldots$ and $M, M_1, \ldots$ as generic constants which may wary from line to line. They may depend on the fixed numbers such as m, p, or C, but not on the function f or the number of sampling points n.

For the proof of Lemma 6, we need the following results.

Lemma 15. *Let $f\in\mathbf{W}_{[0,T]}\left(m,p,C\right)$, and $0\leq k<m$. If $\left|f^{(k)}\left(t\right)\right|\geq A$, for $t\in\left[t_1,t_2\right]$ with $t_2-t_1=B$, then we can find $\tau_1,\tau_2\in\left[t_1,t_2\right]$, with $\tau_2-\tau_1\geq\lambda_k B$, such that $\left|f\left(t\right)\right|\geq\mu_k A B^k$. Here, λ_k and μ_k are constants depend only on k.*

Proof

The result for $k=0$ is clear. Now let $k=1$. By the continuity of $f'\left(t\right)$, $\left|f'\left(t\right)\right|\geq A$ for $t\in\left[t_1,t_2\right]$ implies that either $f'\left(t\right)\geq A$ or $f'\left(t\right)\leq -A$. We will only provide the proof for the former case, since the proof for the latter case is similar. Suppose that $t'=\left(t_2+t_1\right)/2$. If $f\left(t'\right)\geq 0$, then let $t_1^{(1)}=3t_2/4+t_1/4$ and $t_2^{(1)}=t_2$. We have

$$\frac{f\left(t_1^{(1)}\right)-f\left(t'\right)}{t_1^{(1)}-t'}\geq A,$$

so that,

$$\begin{aligned}f\left(t_1^{(1)}\right)&\geq A\left(t_1^{(1)}-t'\right)+f\left(t'\right)\geq A\left(t_1^{(1)}-t'\right)\\&=\frac{1}{4}A\left(t_2-t_1\right)\geq\frac{1}{4}AB.\end{aligned}$$

Thus for $t\in\left[t_1^{(1)},t_2^{(1)}\right]$, $f\left(t\right)\geq AB/4$. If $f\left(t'\right)\leq 0$, then let $t_1^{(1)}=t_1, t_2=3t_1/4+t_2/4$. Then we have $f\left(t\right)\leq -AB/4$ for $t\in\left[t_1^{(1)},t_2^{(1)}\right]$. The proof of the lemma can be completed by an induction on k.

Proof of Lemma 6

First, we prove the lemma for $1 < p < \infty$ and $k = m-1$. Suppose that $\left\|f^{(m-1)}\right\|_\infty = \left|f^{(m-1)}(t')\right| = 2A$. Since the discussion for $f^{(m-1)}(t') = 2A$ and that for $f^{(m-1)}(t') = -2A$ are almost the same, we only give a proof for the case $f^{(m-1)}(t') = 2A$. Let $\delta = (A/C)^{p/(p-1)}$ and $\delta' = \min\{\delta, T/2\}$. Then we have either $[t', t'+\delta'] \subseteq [0,T]$ or $[t'-\delta', t'] \subseteq [0,T]$. Without loss of generality, we may assume that $[t', t'+\delta'] \subseteq [0,T]$. We claim that $f^{(m-1)}(t) \geq A$ for $t \in [t', t'+\delta']$. If not, we can find a $t'' \in [t', t'+\delta']$ such that $f^{(m-1)}(t'') < A$. Then we have

$$\begin{aligned}\int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right| dt &\geq \left|\int_{t'}^{t''} f^{(m)}(t)\, dt\right| \\ &= \left|f^{(m-1)}(t'') - f^{(m-1)}(t')\right| > A,\end{aligned}$$

giving

$$\begin{aligned}\left\|f^{(m)}\right\|_p^p &\geq \int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right|^p dt \\ &\geq \delta' \left(\frac{1}{\delta'}\int_{t'}^{t'+\delta'} \left|f^{(m)}(t)\right| dt\right)^p \\ &> \delta'^{1-p} A^p \geq \delta^{1-p} A^p = \left(\frac{A}{C}\right)^{-p} A^p = C^p,\end{aligned}$$

which is impossible. Hence, by Lemma 15 we find $\tau_1, \tau_2 \in [t', t'+\delta]$ with $\tau_2 - \tau_1 = \delta'' \geq \lambda_{m-1}\delta'$ such that for $t \in [\tau_1, \tau_2]$, we have $|f(t)| \geq \mu_{m-1} A \delta'^{m-1}$. Now,

$$\begin{aligned}\|f\|_r &\geq \left(\lambda_{m-1}\delta' \left(\mu_{m-1} A \delta'^{m-1}\right)^r\right)^{1/r} \\ &= K A \delta'^{(1/r+m-1)}.\end{aligned}$$

If $\left\|f^{(m-1)}\right\|_\infty / 2 \geq C\,(T/2)^{(p-1)/p}$, then $\delta' = T/2$ and

$$\|f\|_r \geq KA\left(\frac{T}{2}\right)^{(1/r+m-1)}$$

or

$$\|f\|_r \geq K\left\|f^{(m-1)}\right\|_\infty,$$

which means that $\left\| f^{(m-1)} \right\|_\infty /2 \le C\,(T/2)^{(p-1)/p}$ for $\|f\|_r$ small enough. If this is the case, we have $\delta' = \delta$ and

$$\begin{aligned}\|f\|_r &\ge KA\delta^{(1/r+m-1)} \\ &= K\,(2A)^{1+\frac{p}{p-1}(1/r+m-1)}.\end{aligned}$$

Therefore

$$\begin{aligned}\left\| f^{(m-1)} \right\|_\infty &\le K\,\|f\|_r^{\frac{1}{1+\frac{p}{p-1}\left(\frac{1}{r}+m-1\right)}} \\ &= K\,\|f\|_r^{\alpha_{m-1}}.\end{aligned}$$

Now, suppose that we have proven the lemma for k, $1 < k \le m-1$. Let $\left\| f^{(k-1)} \right\|_\infty = \left| f^{(k-1)}(t') \right| = 2A$. As before, we assume that $f^{(k-1)}(t') = 2A$. Let $\delta = A / \left\| f^{(k)} \right\|_\infty$ and $\delta' = \min\{\delta, T/2\}$. Then either $[t', t'+\delta'] \subset [0,T]$ or $[t'-\delta', t'] \subset [0,T]$. Without loss of generality, we may assume the former. For every $t \in [t', t'+\delta']$, we have $f^{(k-1)}(t) \ge A$. Now, by applying Lemma 15, we can find $[\tau_1, \tau_2] \subset [t', t'+\delta']$ with $\tau_2 - \tau_1 = \delta'' \ge \lambda_{k-1}\delta'$ such that for $t \in [\tau_1, \tau_2]$, $|f(t)| \ge \mu_{k-1} A \delta'^{k-1}$. Next,

$$\begin{aligned}\|f\|_r &\ge \left(\int_{\tau_1}^{\tau_2} |f(t)|^r \, dt \right)^{1/r} \\ &\ge \left(\delta'' \left(\mu_{k-1} A \delta'^{k-1} \right)^r \right)^{1/r} \\ &= KA\delta'^{r^{-1}+k-1}.\end{aligned}$$

If $\left\| f^{(k-1)} \right\|_\infty > T \left\| f^{(k)} \right\|_\infty$, we have $\delta' = T/2$. Then

$$\|f\|_r \ge KA \left(\frac{T}{2} \right)^{r^{-1}+k-1}$$

or

$$\left\| f^{(k-1)} \right\|_\infty = 2A \le K\,\|f\|_r \left(\frac{T}{2} \right)^{-\left(r^{-1}+k-1\right)}. \tag{19}$$

If $\|f\|_r \le 1$ then (19) gives

$$\left\| f^{(k-1)} \right\|_\infty \le K\,\|f\|_r^{\alpha_{k-1}}$$

since $\alpha_{k-1} \le 1$. If (19) is violated, then $\left\| f^{(k-1)} \right\|_\infty \le t \left\| f^{(k)} \right\|_\infty$, which means that $\delta' = \delta$. Therefore

$$\begin{aligned}\|f\|_r &\ge KA\delta^{r^{-1}+k-1} \\ &= KA^{r^{-1}+k} \left\| f^{(k)} \right\|_\infty^{1-r^{-1}-k}.\end{aligned} \tag{20}$$

By the assumption, we can find ϵ and K_1 such that when $\|f\|_r \leq \epsilon$, we have $\left\|f^{(k)}\right\|_\infty \leq K_1 \|f\|_r^{\alpha_k}$. Then by (20), when $\|f\|_r \leq \epsilon$ we have

$$\begin{aligned}\left\|f^{(k-1)}\right\|_\infty &= 2A \leq K \|f\|_r^{\frac{1}{r^{-1}+k}} \left\|f^{(k)}\right\|_\infty^{\frac{r^{-1}+k-1}{r^{-1}+k}} \\ &\leq K \|f\|_r^{\frac{1}{r^{-1}+k}} \left(K_1 \|f\|_r^{\alpha_k}\right)^{\frac{r^{-1}+k-1}{r^{-1}+k}} \\ &= K \|f\|_r^{\alpha_{k-1}}.\end{aligned}$$

This proves the lemma.

The following lemmas are needed for proving Lemma 7. Assume that $f \in \mathscr{L}_2\,[0,T]$ and $F(x) = \int_0^x f(t)\,dt$ for $x \in [0,T]$.

Lemma 16.

$$\begin{aligned}\int_0^y \int_0^u \int_0^{T-t} f(x+t) f(x)\,dxdtdu = \frac{1}{2}\int_{1-y}^T (F(T) - F(x))^2\,dx+ \\ +\frac{1}{2}\int_0^{T-y} (F(x+y) - F(x))^2\,dx + \frac{1}{2}\int_0^y F^2(x)\,dx.\end{aligned} \tag{21}$$

Proof

Let $F(x) = \int_0^x f(t)\,dt$. We have

$$\begin{aligned}&\int_0^u \int_0^{T-t} f(x+t) f(x)\,dxdt \\ &= \int_0^{T-u} F(x+u) f(x)\,dx + F(T)\left(F(T) - F(T-u)\right) - \int_0^T F(x) f(x)\,dx.\end{aligned}$$

Thus

$$\begin{aligned}&\int_0^y \int_0^u \int_0^{T-t} f(x+t) f(x)\,dxdtdu \\ &= \int_0^y \left(\int_0^{T-u} F(x+u) f(x)\,dx + F(T)\left(F(T) - F(T-u)\right) - \int_0^T F(x) f(x)\,dx.\right) du \\ &= \frac{1}{2}\int_{T-y}^T (F(T) - F(x))^2\,dx + \frac{1}{2}\int_0^{T-y} (F(x+y) - F(x))^2\,dx + \frac{1}{2}\int_0^y F^2(x)\,dx.\end{aligned}$$

Lemma 17. *There exists a constant A_c such that for any $0 < \delta < T$, and $s \in [0,T]$ such that $[s, s+\delta] \subset [0,T]$, we have*

$$\int_0^T \int_0^T f(u) f(v) |u-v|^{2H-2}\,dudv \geq A_c \delta^{2H-2} \left|F(s+\delta) - F(s)\right|^2.$$

Proof

Let $I = \int_0^T \int_0^T f(u) f(v) |u-v|^{2H-2} \, dudv$. It is enough to prove the lemma for bounded f, since such functions are dense in $\mathscr{L}_2[0,T]$, and hence in $\mathscr{L}_\phi^2[0,T]$. Let $g(t) = \int_0^{T-t} f(x+t) f(x) \, dx$, $G(t) = \int_0^t g(s) \, ds$, and $Q(t) = \int_0^t G(s) \, ds$ for $t \in [0,T]$. First, notice that

$$\begin{aligned} I &= \int_0^T \int_0^v f(u) f(v) (v-u)^{2H-2} \, dudv + \int_0^T \int_v^T f(u) f(v) (u-v)^{2H-2} \, dudv \\ &= 2 \int_0^T x^{2H-2} g(x) \, dx \\ &= 2T^{2H-2} G(T) - 2(2H-2) T^{2H-3} Q(T) + 2(2H-2)(2H-3) \int_0^T x^{2H-4} Q(x) \, dx. \end{aligned}$$

It can be shown that $G(T) > 0$, and by Lemma 16, $Q(T)$ is also positive. So we only need to show that there exists K such that $\int_0^T x^{2H-4} Q(x) \, dx \geq K\delta^{2H-2} |F(s+\delta) - F(s)|^2$. By Lemma 16 we have

$$\int_0^T x^{2H-4} Q(x) \, dx \geq \frac{1}{2} I_1 = \int_0^T x^{2H-4} \int_0^{T-x} (F(y+x) - F(y))^2 \, dydx.$$

Without loss of generality, we assume that $F(s+\delta) > F(s)$. Let $h_0 = F(s+\delta) - F(s)$, $\delta_0 = \delta$, $s_0 = s$ and $t_0 = s+\delta$. Beginning at $k=0$, we repeat the process below recursively until certain condition (inequality (22)) is met. Define $\eta_1 = \sup\left\{F(t) - F(s_k) : s_k \leq t \leq \frac{\lambda-1}{\lambda} s_k + \frac{1}{\lambda} t_k\right\}$ and $\eta_2 = \inf\left\{F(t) - F(s_k) : \frac{1}{\lambda} s_k + \frac{\lambda-1}{\lambda} t_k \leq t \leq t_k\right\}$ where $\lambda = 3^{1/(1-H)}$. If

$$\eta_2 - \eta_1 \geq \frac{1}{3} h_k \tag{22}$$

is not satisfied, we have either $\eta_1 > \frac{1}{3} h_k$ or $\eta_2 < \frac{2}{3} h_k$. Without loss of generality, we assume that the former is true. Then we find $\tau \in \left(s_k, \frac{\lambda-1}{\lambda} s_k + \frac{1}{\lambda} t_k\right]$ such that $F(\tau) - F(s_k) > \frac{1}{3} h_k$. Now let $s_{k+1} = s_k$, $t_{k+1} = \tau$, $\delta_{k+1} = t_{k+1} - s_{k+1}$, and $h_{k+1} = F(t_{k+1}) - F(s_{k+1})$.

We claim that the condition (22) will be met after a finite number of repetitions, because if not, then for any positive integer m, we have $h_m \geq (1/3)^m h_0$ and $\delta_m \leq \frac{1}{\lambda^m} \delta$. Thus $\frac{F(t_m) - F(s_m)}{t_m - t_s} = \frac{1}{3^m} \Big/ \frac{1}{\lambda^m}$ which goes to infinity, contradicting the assumption that f is bounded. Thus there is a k

such that condition (22) is met. We have

$$\begin{aligned}
I_1 &\geq \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4} \int_{s_k}^{s_k+\frac{\lambda-1}{\lambda}\delta_k} (F(y+x)-F(y))^2\,dydx \\
&\geq \frac{\lambda}{\lambda-1}\delta_k^{-1} \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4} \left(\int_{s_k}^{s_k+\frac{\lambda-1}{\lambda}\delta_k} (F(y+x)-F(y))\,dy\right)^2 dx \\
&= \frac{\lambda}{\lambda-1}\delta_k^{-1} \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4} \left(\int_0^x \left(F\left(u+s_k+\frac{\lambda-1}{\lambda}\delta_k\right)-F(u+s_k)\right)du\right)^2 dx \\
&\geq \frac{\lambda}{\lambda-1}\delta_k^{-1} \int_0^{\frac{1}{\lambda}\delta_k} x^{2H-4} \left(\int_0^x \frac{1}{3}h_k du\right)^2 dx \\
&\geq \frac{1}{2H-1}\frac{\lambda}{\lambda-1}\left(\frac{1}{\lambda}\right)^{2H-1}\frac{1}{9}\delta^{2H-2}h_0^2.
\end{aligned}$$

This finishes the proof.

Proof of Lemma 7

Suppose that $f^{(k)}(t_0) = 2\epsilon$ with $\epsilon < 1$, and $\left\|f^{(k)}\right\|_\infty = \left|f^{(k)}(t')\right| = 2A \geq 2\epsilon$. Again, we assume that $f^{(k)}(t') = 2A$, since the discussion for the case of $f^{(k)}(t') = -2A$ is almost the same. Like in the proof of Lemma 6, we can find $[\tau_1, \tau_2] \subset [0, T]$ with $\tau_2 - \tau_1 = l \geq \lambda_{k-1}\delta'$, $\delta' = \min\{\delta, T/2\}$ where

$$\delta = \begin{cases} \left(\frac{A}{C}\right)^{\frac{p}{p-1}} & \text{if } \; 1 < p < \infty \text{ and } k = m-1, \\ A/\left\|f^{(k+1)}\right\|_\infty & \text{if } \; k < m-1, \end{cases}$$

and $|f(t)| \geq \mu_k A\delta'^k$ for $t \in [\tau_1, \tau_2]$. If we take

$$g(t) = \begin{cases} l^{-1/2} & \text{if } t \in [\tau_1, \tau_2], \\ 0 & \text{o/w.} \end{cases}$$

We have $\|g\|_2 = 1$ and

$$\begin{aligned}
|\langle g, f\rangle| &\geq l\left(l^{-1/2}\right)\mu_k A\delta'^k \\
&= l^{1/2}\mu_k A\delta'^k.
\end{aligned}$$

We also have

$$\begin{aligned}
\|\mathbf{Z}_\sigma g\|^2 &= \sigma^2 H(2H-1)\int_0^T\int_0^T g(u)g(v)|u-v|^{2H-2}dudv \\
&= \sigma^2 l^{2H-1}.
\end{aligned}$$

Then

$$\begin{aligned}\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} &\geq \frac{l^{1/2}\mu_k A\delta'^k}{2\sigma l^{H-1/2}} \\ &\geq \frac{1}{2\sigma}\left(\lambda_{k-1}\delta'\right)^{1-H}\mu_k A\delta'^k \\ &= KA\delta'^{1-H+k}/\sigma. \end{aligned} \tag{23}$$

If $\delta' = T/2$, then (23) gives

$$\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} \geq KA/\sigma \geq K\epsilon/\sigma \geq K\epsilon^{\gamma_k}/\sigma.$$

If $k = m-1$ and $\delta' = (A/C)^{p/(p-1)}$, then (23) gives

$$\begin{aligned}\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} &\geq KA\,(A/C)^{\frac{p}{p-1}(1-H+(m-1))}/\sigma \\ &= KA^{\gamma_{m-1}}/\sigma \geq K\epsilon^{\gamma_{m-1}}/\sigma.\end{aligned}$$

If $k < m-1$ and $\delta' = A\left/\left\|f^{(k+1)}\right\|_\infty\right.$, because $f^{(k)} \in \mathbf{W}_{[0,T]}(m-k,p,C)$, we apply Lemma 6 and get positive ϵ_0 and K' such that

$$\left\|f^{(k+1)}\right\|_\infty \leq K'\left\|f^{(k)}\right\|_\infty^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}} \leq K'A^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}}$$

when $\epsilon \leq \epsilon_0$. Thus

$$\begin{aligned}\frac{|\langle g,f\rangle|/2}{\|\mathbf{Z}_\sigma g\|} &\geq KA\left(A\left/\left\|f^{(k+1)}\right\|_\infty\right.\right)^{1-H+k}/\sigma \\ &\geq KA^{2-H+k}\left(K'\left\|f^{(k)}\right\|_\infty^{\frac{m-k-1-p^{-1}}{m-k-p^{-1}}}\right)^{-(1-H+k)}/\sigma \\ &= KA^{\gamma_k}/\sigma \geq K\epsilon^{\gamma_k}/\sigma.\end{aligned}$$

Up to now, we have proven that there is a constant K such that when ϵ is small enough, $\upsilon\left(\epsilon;\mathbf{W}_{[0,T]}\left(m,p,\frac{1}{2}C\right)\right) \geq K\epsilon^{\gamma_k}/\sigma$. Next we will complete the other part of the proof. Let $f \in \mathbf{W}_{[0,T]}(m,p,C)$ satisfying $f^{(i)}(0) = f^{(i)}(T) = 0$ for $0 \leq i \leq m$ and $f^{(k)}(t_0)/2 \neq 0$. Let

$$f_\delta(t) = \begin{cases} \delta^{m-1/p} f\left(\frac{t-t_0}{\delta} + t_0\right), & \text{if } t \in I_\delta = (t_0 - t_0\delta, t_0 + (T-t_0)\delta), \\ 0, & \text{otherwise}\end{cases}$$

where $1 \geq \delta > 0$. We have $\left\| f_\delta^{(m)} \right\|_p = \left\| f^{(m)} \right\|_p$. so that $f_\delta \in \mathbf{W}_{[0,T]}(m, p, C)$. For any bounded function g in $\mathscr{L}_2[0,T]$, we have

$$\begin{aligned} |\langle g, f_\delta \rangle| &= \left| \int_{I_\delta} f_\delta(t)\, g(t)\, dt \right| \\ &\leq \left\| f_\delta' \right\|_1 \left\| \mathbf{1}_{I_\delta} G \right\|_\infty , \end{aligned}$$

where $G(t) = \int_{t_0 - t_0\delta}^{t} g(s)\, ds$ for $t \in [t_0 - t_0\delta, t_0 + (T - t_0)\delta]$. Suppose that t_1 and $t_2 \in I_\delta$, $t_1 < t_2$ satisfy that $\left\| \mathbf{1}_{I_\delta} G \right\|_\infty = |G(t_2) - G(t_1)|$. By Lemma 17, we have

$$\begin{aligned} \int_0^T \int_0^T g(u)\, g(v)\, |u - v|^{2H-2}\, du dv &\geq A_c (t_2 - t_1)^{2H-2} |G(t_2) - G(t_1)|^2 \\ &\geq A_c \delta^{2H-2} |G(t_2) - G(t_1)|^2 . \end{aligned}$$

Thus

$$\begin{aligned} \frac{|\langle g, f_\delta \rangle / 2|}{\|\mathbf{Z}_1 g\|} &\leq \frac{\|f_\delta'\|_1 |G(t_2) - G(t_1)|}{A_c^{1/2} \delta^{H-1} |G(t_2) - G(t_1)|} \\ &\leq A_c^{-1/2} \left\| f' \right\|_1 \delta^{m - p^{-1} + 1 - H}. \end{aligned}$$

Notice that the right hand side of the above inequality is free of g, so we have

$$G_0(f_\delta) \leq A_c^{-1/2} \|f\|_1 \delta^{m - p^{-1} + 1 - H} / \sigma.$$

Let

$$\epsilon = \frac{d^k}{dt^k} f_\delta(t_0)/2 = \delta^{m - k - p^{-1}} f^{(k)}(t_0)/2.$$

Since $f_\delta \in \mathbf{W}_{[0,T]}(m, p, C) - \mathbf{W}_{[0,T]}(m, p, C)$, we have

$$\begin{aligned} \upsilon\left(\epsilon; \mathbf{Z}_\sigma, \mathbf{W}_{[0,T]}(m, p, C)\right) &\leq G_0(f_\delta) \\ &\leq A_c^{-1/2} \|f\|_1 \delta^{m - p^{-1} + 1 - H} / \sigma \\ &= K \epsilon^{\frac{m - p^{-1} + 1 - H}{m - k - p^{-1}}} / \sigma \end{aligned}$$

proving the lemma.

Proof of Theorem 5

By Lemma 7 we know that there exists $\epsilon_0 > 0$ and constants A and B such that $A\epsilon^{\gamma_k}/\sigma \leq \upsilon(\epsilon; \mathbf{Z}_\sigma) \leq B\epsilon^{\gamma_k}/\sigma$ when $0 < \epsilon \leq \epsilon_0$. By (11) and Corollary 4, we have

$$\begin{aligned}
\inf_{\hat{L}\text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}\right) &= \sup_{\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)} \\
&= \max\left\{\sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)}, \sup_{\epsilon\geq\epsilon_0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)}\right\} \\
&\leq \max\left\{\sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)}, \sup_{\epsilon\geq\epsilon_0} \frac{\epsilon^2}{\upsilon^2(\epsilon;\mathbf{Z}_\alpha)}\right\} \\
&= \max\left\{\sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)}, \frac{(\epsilon_0)^2\sigma^2}{\upsilon^2((\epsilon_0);\mathbf{Z}_1)}\right\}.
\end{aligned}$$

The last step follows due to the convexity of $\upsilon(\epsilon)$. We now have

$$\sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+B\epsilon^{2\gamma_k}/\sigma^2} \leq \sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)} \leq \sup_{\epsilon_0>\epsilon>0} \frac{\epsilon^2}{1+A\epsilon^{2\gamma_k}/\sigma^2}.$$

A straight forward calculation gives that for σ sufficiently small, we have

$$\begin{aligned}
B^{1/\gamma_k}\gamma_k^{-1}(\gamma_k-1)^{1-1/\gamma_k}\sigma^{2/\gamma_k} &\leq \sup_{1/2>\epsilon>0} \frac{\epsilon^2}{1+\upsilon(^2\epsilon;\mathbf{Z}_\alpha)} \\
&\leq A^{1/\gamma_k}\gamma_k^{-1}(\gamma_k-1)^{1-1/\gamma_k}\sigma^{2/\gamma_k}.
\end{aligned}$$

This yields,

$$\sup_{1/2>\epsilon>0} \frac{\epsilon^2}{1+\upsilon^2(\epsilon;\mathbf{Z}_\alpha)} \asymp \sigma^{2/\gamma_k}.$$

Since

$$\frac{\epsilon_0^2\sigma^2}{\upsilon^2((\epsilon_0);\mathbf{Z}_1)} = o\left(\sigma^{2/\gamma_k}\right),$$

we have (12).

Some preparation is needed for the proof of Theorem 6. Like in Duncan et al. (2000), we define $\varepsilon : \mathscr{L}_\phi^2 \to \mathscr{L}_1(\Omega, \mathscr{F}, P)$ as

$$\varepsilon(f) := \exp\left\{\int_0^T f(t)\,dZ_t - \frac{1}{2}|f|_\phi^2\right\},$$

and

$$\mathcal{E} = \left\{\sum_{k=1}^n a_k\varepsilon(f_k), n\in\mathbb{N}, a_k\in\mathbb{R}, f_k\in\mathscr{L}_\phi^2 \text{ for } k\in\{1,\dots,n\}\right\}.$$

As in Duncan et al. (2000), for $f \in \mathscr{L}_\phi^2$, we have $\varepsilon(f) \in \mathscr{L}_p(\Omega, \mathscr{F}, P)$ for each $p \geq 1$. Following the same lines as in the proof of Theorem 3.1 in Duncan et al. (2000), we can prove the following lemma.

Lemma 18. *$\mathcal{E}$ is a dense set of $\mathscr{L}_p(\Omega, \mathscr{F}, P)$ for each $p \geq 1$.*

Proof of Theorem 6

For simplicity, we assume that $\sigma = 1$. By Corollary 4 we have

$$\inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) = \rho_0(\mathcal{F}).$$

Suppose that $[\mathbf{x}_1, \mathbf{x}_2]$ is a one-dimensional subproblem such that

$$\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right) > \inf_{\hat{L} \text{ affine}} R_{\mathcal{F}}\left(\hat{L}\right) + \epsilon.$$

We will show that

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\hat{T} \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{T}\right)} \leq 1.25.$$

Let $\hat{T} \in \mathscr{L}_1(\Omega, \mathscr{F}, P)$. Without loss of generality, we can assume that $\hat{T}$ is bounded. By Lemma 18, $\mathcal{E}$ is dense in $\mathscr{L}_4(\Omega, \mathscr{F}, P)$. Thus we can find $X = \sum_{i=1}^r a_i \varepsilon(f_i)$ such that $E\left[\Delta^4\right] \leq \epsilon^2$ and $E\left[\Delta^2\right] \leq \epsilon$ where $\Delta = \hat{T} - X$. By Theorem 30.7 in Samko et al. (1993), we can find functions $g_1(t), g_2(t)$ on $[0, T]$ such that $\mathbf{x}_i(t) = \int_0^T \phi(t,u)\, g_i(u)\, du, i = 1, 2$. For $f \in [\mathbf{x}_1, \mathbf{x}_2]$, suppose that $f = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2$. We have $f(t) = \int_0^T \phi(t,u)\, g(u)\, du$ where $g = \theta g_1 + (1-\theta) g_2$. By Theorem 3.3 in Duncan et al. (2000)

$$\begin{aligned}
&E\left\{\left|\Delta^2\left(Z_\cdot + \int_0^\cdot f(s)\, ds\right)\right|\right\} = E\left\{\Delta^2(Z_\cdot)\, e^{\int_0^T g(s) dZ_s - \frac{1}{2}|g|_\phi^2}\right\} \\
&\leq \left(E\left(\Delta^4\right)\right)^{1/2} \left(E\left(\exp 2\left(\int_0^T g(s)\, dZ_s - \frac{1}{2}|g|_\phi^2\right)\right)\right)^{1/2} \\
&\leq \epsilon e^{\frac{1}{2}|g|_\phi^2} \\
&\leq \epsilon \exp\left(\frac{1}{2}\max\left\{|g_1|_\phi^2, |g_2|_\phi^2\right\}\right). \qquad (24)
\end{aligned}$$

Let $X_i = \int_0^T f_i(t)\, dY(t), i = 1, \ldots, r$. Now, consider the minimax risk of estimating $L(f)$ by functions of $X_1, \ldots, X_r$, knowing that $f \in [\mathbf{x}_1, \mathbf{x}_2]$. Without loss of generality, we can assume that the f_i's are linearly independent. Under such assumptions, it is easy to show that $\mathbf{X} = (X_1, \ldots, X_r)'$ is a Gaussian vector with positive definite covariance matrix

$$\Sigma = \begin{pmatrix} |f_1|_\phi^2 & \langle f_1, f_2 \rangle_\phi & \ldots & \langle f_1, f_r \rangle_\phi \\ \langle f_2, f_1 \rangle_\phi & |f_2|_\phi^2 & \ldots & \langle f_2, f_r \rangle_\phi \\ \ldots & \ldots & \ldots & \ldots \\ \langle f_r, f_1 \rangle_\phi & \langle f_r, f_2 \rangle_\phi & \ldots & |f_r|_\phi^2 \end{pmatrix}$$

and mean vector $\mu_f = E_f(\mathbf{X}) = \left(\int_0^T f_1(t) f(t)\, dt, \ldots, \int_0^T f_r(t) f(t)\, dt \right)' = (\theta\mu_1 + (1-\theta)\mu_2)$, where $\mu_i = E_{\mathbf{x}_i}(\mathbf{X}), i = 1, 2$. It is easy to see that $L(f)$ is a linear function of θ, and the estimation of $L(f)$ is equivalent to the estimation of θ. Since a sufficient statistic for θ is $S = \left(\mu_1^T - \mu_2^T\right) \Sigma^{-1} \mathbf{X}$, which is distributed $N\left(\left(\mu_1^T - \mu_2^T\right) \Sigma^{-1} \mu_f, \left(\mu_1^T - \mu_2^T\right) \Sigma^{-1} (\mu_1 - \mu_2)\right)$, we have

$$\inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1, \mathbf{x}_2]}(\eta(\mathbf{X})) = \inf_{\kappa \text{ measurable}} \sup_{\theta \in [0,1]} E_\theta \left\{ (\kappa(S) - L(f))^2 \right\}.$$

But, we know that

$$\frac{\inf_{\kappa \text{ affine}} \sup_{\theta \in [0,1]} E_\theta \left\{ (\kappa(S) - L(f))^2 \right\}}{\inf_{\kappa \text{ measurable}} \sup_{\theta \in [0,1]} E_\theta \left\{ (\kappa(S) - L(f))^2 \right\}} \leq 1.25,$$

and since $\kappa(S)$ is an affine estimator if κ is affine, we have

$$\frac{\inf_{\hat{L} \text{ affine}} R_{[\mathbf{x}_1, \mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1, \mathbf{x}_2]}(\eta(\mathbf{X}))} \leq 1.25. \tag{25}$$

Clearly, for $f \in [\mathbf{x}_1, \mathbf{x}_2]$,

$$E_f \left(\left(X \left(Z. + \int_0^{\cdot} f(s)\, ds \right) - L(f) \right)^2 \right) \geq \inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1, \mathbf{x}_2]}(\eta(\mathbf{X})).$$

By (24) we have

$$\left(E\left(\left(X\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)\right)^{1/2}$$
$$\leq \left(E\left(\left(\hat{T}\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)\right)^{1/2}+\left(E\left\{\left|\Delta^2\left(Z.+\int_0^{\cdot} f(s)\,ds\right)\right|\right\}\right)^{1/2}$$
$$\leq \left(E\left(\left(\hat{T}\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)\right)^{1/2}+\left(\epsilon\exp\left(\frac{1}{2}\max\left\{|g_1|_\phi^2,|g_2|_\phi^2\right\}\right)\right)^{1/2}$$
$$= \left(E\left(\left(\hat{T}\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)\right)^{1/2}+O\left(\epsilon^{1/2}\right).$$

This gives us

$$\inf_{\eta \text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\eta\left(\mathbf{X}\right)\right)\leq E\left(\left(\hat{T}\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)+O\left(\epsilon^{1/2}\right). \tag{26}$$

By (25) and (26) we have

$$\frac{\inf_{\hat{L}\text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{E\left(\left(\hat{T}\left(Z.+\int_0^{\cdot} f(s)\,ds\right)-L(f)\right)^2\right)+O\left(\epsilon^{1/2}\right)}\leq 1.25.$$

Taking infimum over $f\in[\mathbf{x}_1,\mathbf{x}_2]$ we have

$$\frac{\inf_{\hat{L}\text{ affine}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{L}\right)}{\inf_{\hat{T}\text{ measurable}} R_{[\mathbf{x}_1,\mathbf{x}_2]}\left(\hat{T}\right)+O\left(\epsilon^{1/2}\right)}\leq 1.25.$$

Since the ϵ in the above inequality is arbitrary, the proof is completed.

Proof of Lemma 8

For simplicity, we work with the cases in which $n=2lm$ for some $l\geq 1$. The result can be easily generalized to other cases. Let $\|\mathbf{K}_n f\|_2=M$. We have

$$\min_{1\leq i\leq l}\left\{f^2\left(t_{2(j-1)l+i}\right)\right\}\leq \frac{1}{l}\Sigma_{i=1}^{l} f^2\left(t_{2(j-1)l+i}\right)\leq \frac{nM^2}{l}=2mM^2, \tag{27}$$

where $j = 1, \ldots, m$. Let ξ_j, $j = 1, \ldots, m$ be such that

$$\min_{1 \le i \le l} \left\{ f^2 \left(t_{2(j-1)l+i} \right) \right\} = f^2 \left(\xi_j \right).$$

We have $|\xi_i - \xi_j| \ge T/2m$ for any $1 \le i, j \le m$, $i \ne j$. Let $P_k(f)$ be the Lagrangian interpolation polynomial for $(\xi_i, f(\xi_i)), \ldots, (\xi_k, f(\xi_k))$, $k = 2, \ldots, m$. By (27) it can be shown that there exists a K independent of f such that

$$\begin{aligned} (P_k(f))^{(k-1)}(t) &\le K \max \{ |f(\xi_i)| : i = 1, \ldots, k \} \\ &\le KM. \end{aligned} \tag{28}$$

Notice that we can find $\zeta_k \in (0, T), k = 1, \ldots m-1$ such that $f^{(k)}(\zeta_k) = (P_{k+1}(f))^{(k)}(\zeta_k)$. With this observation and also by combining (28) and the fact that $\left\| f^{(m)} \right\|_p \le C$, we have

$$\begin{aligned} \left| \left\| f^{(m-1)} \right\|_\infty - f^{(m-1)}(\zeta_k) \right| &\le \int_0^T \left| f^{(m)} \right| dt \\ &\le T^{1-p^{-1}} \left(\int_0^T \left| f^{(m)} \right|^p dt \right)^{1/p} \\ &= CT^{1-p^{-1}}. \end{aligned}$$

Thus

$$\begin{aligned} \left\| f^{(m-1)} \right\|_\infty &\le (P_m(f))^{(m-1)}(\zeta_{m-1}) + CT^{1-p^{-1}} \\ &= KM + CT^{1-p^{-1}}. \end{aligned}$$

Now, (14) can be readily shown by induction.

For the proof of Theorem 8, we need the following lemmas.

Lemma 19. *Let $h = T/n$. Then* $\left| \sqrt{T} \left\| \mathbf{K}_n f \right\|_2 - \|f\|_2 \right| = \left\| \mathbf{K}_n f \right\|_2 O_1(h) + O_2(h)$.

Proof

For $t \in [t_i - h/2, t_i + h/2]$, $i = 1, \ldots, n$,

$$f(t) = f(t_i) + (t - t_i) f'(\xi).$$

Thus

$$\begin{aligned}|f(t)-f(t_i)| &\le \left|(t-t_i)f'(\xi)\right| \\ &\le \frac{h}{2}\left(KM\|\mathbf{K}_n f\|_2 + CT^{m-1-p^{-1}}\right)\end{aligned}$$

where K is as in Lemma 8. Now,

$$\int_{t_i-h/2}^{t_i+h/2} |f(t)-f(t_i)|^2\, dt \le \left(\frac{h}{2}\left(KM\|\mathbf{K}_n f\|_2 + CT^{m-1-p^{-1}}\right)\right)^2 h.$$

Since

$$\|\mathbf{K}_n f\|_2 = \left(\frac{1}{n}\sum_{i=1}^{n} f^2(t_i)\right)^{1/2} = \sqrt{\frac{1}{T}}\,\|f_n\|_2\,,$$

where function f_n is defined on $[0,T)$ as $f_n = f(t_i)$ for $t \in [t_i - h/2, t_i + h/2)$, $i = 1,\ldots,n$. Thus

$$\begin{aligned}\left|\sqrt{T}\,\|\mathbf{K}_n f\|_2 - \|f\|_2\right| &= |\|f_n\|_2 - \|f\|_2| \le \|f_n - f\|_2 \\ &= \left(\Sigma_{i=1}^{n}\int_{t_i-h/2}^{t_i+h/2} |f(t)-f(t_i)|^2\, dt\right)^{1/2} \\ &\le \left(n\left(\frac{h}{2}\left(K\|\mathbf{K}_n f\|_2 + CT^{m-2} + T^{m-1}\right)\right)^2 h\right)^{1/2} \\ &= \|\mathbf{K}_n f\|_2\, O_1(h) + O_2(h)\,.\end{aligned}$$

Now we have the following result.

Lemma 20. *Suppose that $f \in \mathcal{F}$ with $\|f\|_2 \le M$ and $L(f)/2 = \epsilon \ge Bn^{(H-1)/\gamma_k}$ for some positive number B. Then there exists B' independent of f and n such that for n sufficiently large and ϵ sufficiently small, we have $G_0\left(f;\tilde{\mathbf{Z}}_n,\mathbf{K}_n,\mathcal{F}\right) \ge B'\epsilon^{\gamma_k}\Big/ n^{\tilde{H}-1}$.*

Proof

For any $f \in \mathcal{F}_M$, and $|L(f/2)| = \epsilon$, we have

$$G_0\left(f;\tilde{\mathbf{Z}}_n,\mathbf{K}_n,\mathcal{F}\right) \ge \frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|}$$

where g is as in the proof of Lemma 7. Let $1 \le i_1 < i_2 \le n$ be the smallest and largest integer i such that $t_i \in [\tau_1, \tau_2]$ respectively. We know that

$$\lambda_k \delta' \le \tau_2 - \tau_1 \le \delta' = \min\{\delta, T/2\}$$

where

$$\delta = \begin{cases} (A/C)^{p/(p-1)} & \text{if } k = m-1, \\ A \big/ \left\| f^{(k+1)} \right\|_\infty & \text{if } k < m-1, \end{cases}$$

with $A = \left\| f^{(k)} \right\|_\infty /2 \geq L(f)/2 = \epsilon$. Then we have

$$\langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle = \frac{1}{2n} \sum_{i=i_1}^{i_2} l^{-1/2} f(t_i)$$

where $l = \tau_2 - \tau_1$. Now we will show that $[\tau_1, \tau_2]$ becomes dense with t_i's as n goes to infinity. If $\delta' = T/2$, then clearly the number of t_i's in the interval $[\tau_1, \tau_2]$ will go to infinity. Now assume that $\delta' = \delta$. First, assume that $k = m-1$. We then have $\delta' = (A/C)^{p/(p-1)} \geq (\epsilon/C)^{p/(p-1)} \geq Kn^{\frac{\tilde{H}-1}{\gamma_k}\frac{p}{p-1}} = Kn^{\frac{(\tilde{H}-1)}{m-p^{-1}+1-\tilde{H}}}$. Since $\left(\tilde{H}-1\right) \Big/ \left(m - p^{-1} + 1 - \tilde{H}\right)$ is larger than -1, the t_i's will also become dense in $[\tau_1, \tau_2]$ as n goes to infinity. Next, if $k < m-1$, we have $\delta' = A \big/ \left\| f^{(k+1)} \right\|_\infty = \left\| f^{(k)} \right\|_\infty /2 \left\| f^{(k+1)} \right\|_\infty$. By Lemma 6, again we have that δ' is bigger than $Kn^{\left(\tilde{H}-1\right)/\left(m-p^{-1}+1-\tilde{H}\right)}$ for some constant K and sufficiently large n. Thus, again, the t_i's will become dense in $[\tau_1, \tau_2]$ as n goes to infinity. We have

$$\begin{aligned}
&T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle - \langle g, f/2 \rangle \\
&= \frac{T}{2n} \sum_{i=i_1}^{i_2} l^{-1/2} f(t_i) - \frac{1}{2} \int_{\tau_1}^{\tau_2} l^{-1/2} f(t)\, dt \\
&= \frac{T}{2n} \sum_{i=i_1}^{i_2} l^{-1/2} f(t_i) - \frac{1}{2} l^{-1/2} \left(\int_{\tau_1}^{t_{i_1}} f(t)\, dt + \sum_{i=i_1}^{i_2-1} \int_{t_i}^{t_{i+1}} f(t)\, dt + \int_{t_{i_2}}^{\tau_2} f(t)\, dt \right) \\
&= \frac{1}{2l^{1/2}} \sum_{i=i_1}^{i_2-1} \int_{t_i}^{t_{i+1}} \left(f(t_i) - f(t) \right) dt + \frac{T}{2nl^{1/2}} f(t_{i_2}) - \frac{1}{2l^{1/2}} \left(\int_{\tau_1}^{t_{i_1}} f(t)\, dt + \int_{t_{i_2}}^{\tau_2} f(t)\, dt \right),
\end{aligned}$$

so that

$$\begin{aligned}
&\left| T \langle \mathbf{K}_n g, \mathbf{K}_n f/2 \rangle - \langle g, f/2 \rangle \right| \\
&\leq \frac{1}{2l^{1/2}} \left(\sum_{i=i_1}^{i_2-1} \int_{t_i}^{t_{i+1}} \left| f(t_i) - f(t) \right| dt + \frac{T}{n} \left| f(t_{i_2}) \right| + \int_{\tau_1}^{t_{i_1}} |f(t)|\, dt + \int_{t_{i_2}}^{\tau_2} |f(t)|\, dt \right) \\
&\leq \frac{1}{2l^{1/2}} \left(l \frac{T}{n} \left\| f' \right\|_\infty + \frac{3T}{n} \|f\|_\infty \right) = l^{1/2} \frac{T}{2n} \left\| f' \right\|_\infty + \frac{3T}{2l^{1/2} n} \|f\|_\infty. \qquad (29)
\end{aligned}$$

From Lemma 6 we know that $\lim_{\epsilon \to 0} \sup \left\{ \|f\|_\infty : \|f\|_2 = \epsilon \right\} = 0$. From this and the fact that $\sup \left\{ \|f\|_\infty : \|f\|_2 = \epsilon \right\}$ is a concave function of ϵ, we know that $\sup\{\|f\|_\infty : \|f\|_2 = \epsilon\}$ is finite for every ϵ. And since $\|f\|_2 \le M$, we have $\|f\|_\infty \le M_1 = \sup\{\|f\|_\infty : \|f\|_2 = M\} < \infty$. Similarly, we have $\|f'\|_\infty \le M_2 = \sup\{\|f'\|_\infty : \|f\|_2 = M\} < \infty$. It is shown in the proof of Lemma 7 that $|\langle g, f\rangle| \ge l^{1/2}\mu_k A\delta'^k$. Hence, (29) gives

$$\begin{aligned}\left|\frac{T\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right| &\le \frac{l^{1/2}\frac{T}{n}M_2 + \frac{3T}{l^{1/2}n}M_1}{l^{1/2}\mu_k A\delta'^k} \\ &= \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{ln\mu_k A\delta'^k} \\ &\le \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{\lambda_k n\mu_k A\delta'^{k+1}}. \qquad (30)\end{aligned}$$

If $k = m-1$, we have

$$\delta = (A/C)^{p/(p-1)}$$

and (30) becomes

$$\begin{aligned}&\left|\frac{T\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right| \\ &\le \max\left\{\frac{TM_2}{n\mu_{m-1}A\,(T/2)^{m-1}}, \frac{TM_2}{n\mu_{m-1}A\left((A/C)^{p/(p-1)}\right)^{m-1}}\right\} \\ &+ \max\left\{\frac{3TM_1}{\lambda_{m-1}n\mu_{m-1}A\,(T/2)^{m}}, \frac{3TM_1}{\lambda_{m-1}n\mu_{m-1}A\left((A/C)^{p/(p-1)}\right)^{m}}\right\} \\ &\le \max\left\{K_2 n^{\frac{1-\tilde{H}}{\gamma_{m-1}}-1}, K_3 n^{\frac{1-\tilde{H}}{\gamma_{m-1}}\left(\frac{(m-1)p}{p-1}+1\right)-1}\right\} + \max\left\{K_4 n^{\frac{1-\tilde{H}}{\gamma_{m-1}}-1}, K_5 n^{\frac{1-\tilde{H}}{\gamma_{m-1}}\left(\frac{mp}{p-1}+1\right)-1}\right\} \\ &= O\left(n^{\frac{1-\tilde{H}}{\gamma_{m-1}}\left(\frac{mp}{p-1}+1\right)-1}\right) \\ &= O\left(n^{\beta}\right),\end{aligned}$$

where $\beta = \left(\tilde{H}p^{-1} - \tilde{H}m\right)\Big/\left(m - p^{-1} + 1 - \tilde{H}\right)$. If $k < m-1$, we have

$$\delta = A\Big/\left\|f^{(k+1)}\right\|_\infty .$$

For ϵ sufficiently small, by (30) and Lemma 6 we have

$$
\begin{aligned}
&\left|\frac{T\left\langle \mathbf{K}_n g, \mathbf{K}_n f/2\right\rangle - \left\langle g, f/2\right\rangle}{\left\langle g, f/2\right\rangle}\right| \\
&\leq \frac{TM_2}{n\mu_k A\delta'^k} + \frac{3TM_1}{\lambda_k n\mu_k A\delta'^{k+1}} \\
&\leq \max\left\{K_2 n^{\frac{1-\tilde{H}}{\gamma_k}-1}, \frac{TM_2}{n\mu_k A\left(A/\left\|f^{(k+1)}\right\|_\infty\right)^k}\right\} \\
&+ \max\left\{K_4 n^{\frac{1-\tilde{H}}{\gamma_k}-1}, \frac{3TM_1}{\lambda_k n\mu_k A\left(A/\left\|f^{(k+1)}\right\|_\infty\right)^{k+1}}\right\} \\
&\leq \max\left\{K_2 n^{\frac{1-\tilde{H}}{\gamma_k}-1}, K_3' n^{\left(\frac{1-\tilde{H}}{\gamma_k}\right)\left(\frac{k}{m-k-p^{-1}}+1\right)-1}\right\} \\
&+ \max\left\{K_4 n^{\frac{1-\tilde{H}}{\gamma_k}-1}, K_5' n^{\left(\frac{1-\tilde{H}}{\gamma_k}\right)\left(\frac{k+1}{m-k-p^{-1}}+1\right)-1}\right\} \\
&= O\left(n^{\left(\frac{1-\tilde{H}}{\gamma_k}\right)\left(\frac{k+1}{m-k-p^{-1}}+1\right)-1}\right) = O\left(n^\beta\right).
\end{aligned}
$$

Since we assume that either $m > 1$ or $p > 1$, β is always negative. Now notice that

$$
\begin{aligned}
\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|^2 &= \frac{1}{n^2} l^{-1} \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} R\left(i-j\right) \\
&= \frac{1}{n^2} n'^2 \left(n'\right)^{2\tilde{H}-2} l^{-1} \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} \frac{1}{n'^2} R\left(i-j\right)\left(\frac{1}{n'}\right)^{2\tilde{H}-2}
\end{aligned}
$$

where $n' = i_2 - i_1 + 1$. Let

$$
S = \sum_{j=i_1}^{i_2} \sum_{i=i_1}^{i_2} \frac{1}{n'^2} R\left(i-j\right)\left(\frac{1}{n'}\right)^{2\tilde{H}-2},
$$

and we can see that

$$
S \to C_1 \int_0^1 \int_0^1 \left|u-v\right|^{2\tilde{H}-2} du dv \tag{31}
$$

as n' goes to infinity. In fact, we define the function

$$
f_{n'}\left(u,v\right) = \begin{cases} R\left(i-j\right)\left(\frac{1}{n'}\right)^{2\tilde{H}-2} \text{ if } \left(u,v\right) \in \left(\frac{i-1}{n'}, \frac{i}{n'}\right] \times \left(\frac{j-1}{n'}, \frac{j}{n'}\right] & \text{for } i = 1, \ldots, n'; \\ 0 & \text{otherwise.} \end{cases}
$$

Then we can see that $f_{n'}\left(u,v\right) \to C_1 \left|u-v\right|^{2\tilde{H}-2}$ a.e. on $[0,1] \times [0,1]$, and $S = \int_0^1 \int_0^1 f_{n'}\left(u,v\right) du dv$. Further more, Since $R\left(k\right) \sim C_1 \left|k\right|^{2\tilde{H}-2}$, we can find K such that $R(k) \leq K|k|^{2\tilde{H}-2}$ for $k > 0$. Now, letting $K' = \max\left\{2^{2-2\tilde{H}}K, R(0)\right\}$, we have $R(k) \leq K'(|k|+1)^{2\tilde{H}-2}$ for $k \geq 0$. Thus

$$R(i-j)\left(\frac{1}{n'}\right)^{2\tilde{H}-2} \leq K'\left(\frac{|i-j|+1}{n'}\right)^{2\tilde{H}-2} \leq K'|u-v|^{2\tilde{H}-2}$$

for $(u,v) \in ((i-1)/n', i/n'] \times ((j-1)/n', j/n']$, which means that the function $f_n(u,v)$ is dominated by $K'|u-v|^{2\tilde{H}-2}$. By dominated convergence theorem (31) follows. Now, letting $\sigma_n = \left(C_1^{1/2}T^{1-\tilde{H}}n^{\tilde{H}-1}\right) \Big/ \left(\tilde{H}\left(2\tilde{H}-1\right)\right)^{1/2}$, we have

$$\begin{aligned}
&\frac{\left|T^2\left\|\tilde{\mathbf{Z}}_n\mathbf{K}_n g\right\|^2 - \|\mathbf{Z}_{\sigma_n}g\|^2\right|}{\|\mathbf{Z}_{\sigma_n}g\|^2} \\
&= \frac{\left|T^2n^{-2}n'^2n'^{2\tilde{H}-2}l^{-1}S - \sigma_n^2\tilde{H}\left(2\tilde{H}-1\right)l^{-1}\int_0^l\int_0^l|u-v|^{2\tilde{H}-2}\,dudv\right|}{\sigma_n^2 l^{2\tilde{H}-1}} \\
&= \frac{n^{2-2\tilde{H}}T^{2-2\tilde{H}}\left(\left(\frac{n'T}{nl}\right)^{2\tilde{H}}S - C_1\int_0^1\int_0^1|s-t|^{2\tilde{H}-2}\,dsdt\right)}{\sigma_n^2} \\
&= C_1^{-1}\left(\tilde{H}\left(2\tilde{H}-1\right)\right)^{-1}\left(\left(\frac{n'T}{nl}\right)^{2\tilde{H}}S - C_1\int_0^1\int_0^1|s-t|^{2\tilde{H}-2}\,dsdt\right).
\end{aligned}$$

Hence, using (31) and $n'T/(nl) \to 1$ we have

$$\left|\frac{T^2\left\|\tilde{\mathbf{Z}}_n g_n\right\|^2 - \|\mathbf{Z}_{\sigma_n}g\|^2}{\|\mathbf{Z}_{\sigma_n}g\|^2}\right| \to 0.$$

Since

$$\begin{aligned}
\left|\frac{T^2\left\|\tilde{\mathbf{Z}}_n g_n\right\|^2 - \|\mathbf{Z}_{\sigma_n}g\|^2}{\|\mathbf{Z}_{\sigma_n}g\|^2}\right| &= \left|\frac{\left(T\left\|\tilde{\mathbf{Z}}_n g_n\right\| - \|\mathbf{Z}_{\sigma_n}g\|\right)\left(T\left\|\tilde{\mathbf{Z}}_n g_n\right\| + \|\mathbf{Z}_{\sigma_n}g\|\right)}{\|\mathbf{Z}_{\sigma_n}g\|^2}\right| \\
&\geq \left|\frac{T\left\|\tilde{\mathbf{Z}}_n g_n\right\| - \|\mathbf{Z}_{\sigma_n}g\|}{\|\mathbf{Z}_{\sigma_n}g\|}\right|,
\end{aligned}$$

we have

$$\left|\frac{T\left\|\tilde{\mathbf{Z}}_n g_n\right\| - \|\mathbf{Z}_{\sigma_n}g\|}{\|\mathbf{Z}_{\sigma_n}g\|}\right| \to 0,$$

which also gives

$$\frac{T\left\|\tilde{\mathbf{Z}}_n g_n\right\|}{\left\|\mathbf{Z}_{\sigma_n} g\right\|} \to 1.$$

Hence,

$$\frac{\left|\frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} - \frac{\langle g, f/2\rangle}{\left\|\mathbf{Z}_{\sigma_n} g\right\|}\right|}{\left|\frac{\langle g, f/2\rangle}{\left\|\mathbf{Z}_{\sigma_n} g\right\|}\right|} = \frac{\left|\frac{T\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle \left\|\mathbf{Z}_{\sigma_n} g\right\| - T\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\| \langle g, f/2\rangle}{T\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\| \left\|\mathbf{Z}_{\sigma_n} g\right\|}\right|}{\left|\frac{\langle g, f/2\rangle}{\left\|\mathbf{Z}_{\sigma_n} g\right\|}\right|}$$

$$= \left|\frac{\left\|\mathbf{Z}_{\sigma_n} g\right\|}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} \frac{T\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle - \langle g, f/2\rangle}{\langle g, f/2\rangle}\right.$$

$$\left. + \frac{\left\|\mathbf{Z}_{\sigma_n} g\right\| - T\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|}\right| \to 0.$$

In the proof of Lemma 7 we have shown that

$$\frac{\langle g, f/2\rangle}{\left\|\mathbf{Z}_{\sigma_n} g\right\|} \geq K_1 \epsilon^{\gamma_k}/\sigma_n$$

for some positive K_1 independent of f, and for sufficiently small ϵ. Thus for n sufficiently large and ϵ sufficiently small we have

$$G_0\left(f; \tilde{\mathbf{Z}}_n, \mathbf{K}_n, \mathcal{F}\right) \geq \frac{\langle \mathbf{K}_n g, \mathbf{K}_n f/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{K}_n g\right\|} \geq K'' \epsilon^{\gamma_k}/\sigma_n = B' \epsilon^{\gamma_k} \Big/ n^{\tilde{H}-1}.$$

The following lemma is the discrete version of Lemma 17. Since its proof is similar to that of Lemma 17 except that integrations are replaced by summations, we omit it.

Lemma 21. *For $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$, and $1 \leq i_1 \leq i_2 \leq n$, we can find a positive A_d independent of n and $\mathbf{x}$ such that*

$$\sum_{j=1}^{n} \sum_{i=1}^{n} x_i x_j R(i-j) \geq A_d (i_2 - i_1 + 1)^{2\tilde{H}-2} \left|\sum_{i=i_1}^{i_2} x_i\right|^2.$$

Proof of Theorem 8

By Corollary 4 we have

$$\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C)}\left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right) = \rho_0\left(\mathcal{F}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n\right).$$

By Lemma 8, there exists M such that if $\|f\|_2 > M$, then $\|\mathbf{K}_n f\|_2 > M' > 0$. Hence,

$$\rho_0\left(\mathcal{F};\mathbf{K}_n,\tilde{\mathbf{Z}}_n\right) = \sup_{f\in\mathcal{F}-\mathcal{F}} \rho_0\left(f;\mathcal{F},\tilde{\mathbf{Z}}_n,L,\mathbf{K}_n\right)$$
$$= \max\{\rho_1,\rho_2,\rho_3\}$$

where $\rho_i = \sup\{\rho_0(f) : f\in\mathcal{F}_i\}$, $i=1,2,3$ with

$$\mathcal{F}_1 = \{f\in\mathcal{F}-\mathcal{F} : \|f\|_2 > M\},$$
$$\mathcal{F}_2 = \left\{f\in\mathcal{F}-\mathcal{F} : |L(f)/2| < Kn^{(\tilde{H}-1)/\gamma_k}\right\},$$

and

$$\mathcal{F}_3 = \left\{f\in\mathcal{F}-\mathcal{F} : \|f\|_2 \le M, |L(f)/2| \ge Kn^{(\tilde{H}-1)/\gamma_k}\right\}.$$

By Lemma 20 we have

$$\rho_3 \le \sup_{\epsilon\ge Kn^{(\tilde{H}-1)/\gamma_k}, f\in\mathcal{F}_3} \frac{\epsilon^2}{1+G_0^2(f)}$$
$$\le \sup_{\epsilon\ge Kn^{(\tilde{H}-1)/\gamma_k}} \frac{\epsilon^2}{1+\left(B_1\frac{\epsilon^{\gamma_k}}{n^{\tilde{H}-1}}\right)^2}$$
$$\le \sup_{\epsilon} \frac{\epsilon^2}{1+\left(B_1\frac{\epsilon^{\gamma_k}}{n^{\tilde{H}-1}}\right)^2} = O\left(n^{(2\tilde{H}-2)/\gamma_k}\right)$$

$$\rho_1 \le \sup_{f\in\mathcal{F}_1} \frac{(L(f)/2)^2}{1+\|\mathbf{K}_n f/2\|_2^2 \Big/ \left\|\tilde{\mathbf{Z}}_n\right\|^2}$$
$$= \sup_{f\in\mathcal{F}_1} \frac{(L(f))^2 \Big/ \|\mathbf{K}_n f\|_2^2}{4\Big/\|\mathbf{K}_n f\|_2^2 + 1\Big/\left\|\tilde{\mathbf{Z}}_n\right\|^2}.$$

By Lemma 8, we can find K such that $\left\|f^{(k)}\right\|_\infty \le K\|\mathbf{K}_n f\|_2$. Hence

$$\frac{|L(f)|}{\|\mathbf{K}_n f\|_2} = \frac{\left|f^{(k)}(t_0)\right|}{\|\mathbf{K}_n f\|_2} \le \frac{\left\|f^{(k)}\right\|_\infty}{\|\mathbf{K}_n f\|_2} \le K.$$

Therefore,

$$\rho_1 \le \sup_{f\in\mathcal{F}_2} \frac{K^2}{1\Big/\left\|\tilde{\mathbf{Z}}_n\right\|^2} = K^2\left\|\tilde{\mathbf{Z}}_n\right\|^2 = O\left(n^{2\tilde{H}-2}\right).$$

Also,

$$\rho_2 \le \left(Kn^{\left(\tilde{H}-1\right)/\gamma_k}\right)^2 = O\left(n^{\left(2\tilde{H}-2\right)/\gamma_k}\right).$$

Thus $\max\{\rho_1, \rho_2, \rho_3\} = O\left(n^{\left(2\tilde{H}-2\right)/\gamma_k}\right)$. This proves one side of the inequality.

Now, let f_δ be defined as in the proof of Lemma 7, and let

$$i_1 = \min\{1 \le i \le n : t_i \ge t_0 - t_0\delta\}$$

and

$$i_2 = \max\{1 \le i \le n : t_i \le t_0 + (1-t_0)\,\delta\}.$$

For any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{K}_n f_\delta \rangle &= \frac{1}{n}\sum_{i_1}^{i_2} f_\delta(t_i)\, x_i \\
&= \frac{1}{n}\left(f_\delta(t_{i_2})\left(\sum_{i=i_1}^{i_2} x_i\right) - \sum_{i=i_1}^{i_2-1}\left(f_\delta(t_{i+1}) - f_\delta(t_i)\right)\left(\sum_{j=i_1}^{i} x_i\right)\right) \\
&\le \frac{1}{n}\left(\sum_{i=i_1}^{i_2-1}\left|f_\delta(t_{i+1}) - f_\delta(t_i)\right|\right)\sup_{i=i_1}^{i_2-1}\left|\sum_{j=i_1}^{i} x_i\right|.
\end{aligned}$$

Suppose that $i_1 \le j_1 \le j_2 \le i_2 - 1$, and $\left|\sum_{i=j_1}^{j_i} x_i\right| = \sup_{i=i_1}^{i_2-1}\left|\sum_{j=i_1}^{i} x_i\right|$. By Lemma 21 we have

$$\begin{aligned}
\left\|\tilde{\mathbf{Z}}_n \mathbf{x}\right\|^2 &= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} x_i x_j R(i-j) \\
&\ge \frac{1}{n^2} A_d\,(j_2 - j_1 + 1)^{2\tilde{H}-2}\left|\sum_{i=j_1}^{j_2} x_i\right|^2 \\
&\ge \frac{1}{n^2} A_d\,(i_2 - i_1)^{2\tilde{H}-2}\sup_{i=i_1}^{i_2-1}\left|\sum_{j=i_1}^{i} x_i\right|^2.
\end{aligned}$$

Thus

$$\begin{aligned}
G_0(f_\delta) &= \sup_{\mathbf{x}\in\mathbb{R}^n}\frac{\langle \mathbf{x}, \mathbf{K}_n f_\delta/2\rangle}{\left\|\tilde{\mathbf{Z}}_n \mathbf{x}\right\|} \\
&\le A_d^{1/2}\sum_{i=i_1}^{i_2-1}\left|f_\delta(t_{i+1}) - f_\delta(t_i)\right|(i_2 - i_1)^{1-\tilde{H}}. \qquad (32)
\end{aligned}$$

Since

$$\begin{aligned}
&\sum_{i=i_1}^{i_2-1} \left| f_\delta\left(t_{i+1}\right) - f_\delta\left(t_i\right) \right| - \int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta} \left| f_\delta'\left(t\right) \right| dt \\
&= \frac{T}{n} \sum_{i=i_1}^{i_2-1} f_\delta'\left(\xi_i\right) - \int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta} f_\delta'\left(t\right) dt \\
&\leq \frac{T}{n} \int_{t_0-t_0\delta}^{t_0+(1-t_0)\delta} \left\| f_\delta'' \right\|_\infty dt \\
&= \frac{T}{n} \delta \left\| f_\delta'' \right\|_\infty \\
&= \frac{T}{n} \delta^{m-1-p^{-1}} \left\| f'' \right\|_\infty ,
\end{aligned}$$

by (32)

$$\begin{aligned}
G_0\left(f_\delta\right) &\leq A_d^{-1/2} \left(\left\| f_\delta' \right\|_1 + \frac{T}{n} \delta^{m-1-p^{-1}} \left\| f'' \right\|_\infty \right) \left(i_2 - i_1\right)^{1-\tilde{H}} \\
&= A_d^{-1/2} \left(\delta^{m-p^{-1}} \left\| f' \right\|_1 + \frac{T}{n} \delta^{m-1-p^{-1}} \left\| f'' \right\|_\infty \right) \left(i_2 - i_1\right)^{1-\tilde{H}} .
\end{aligned}$$

Now let $\delta = n^{\left(\tilde{H}-1\right)/\left(m-p^{-1}+1-\tilde{H}\right)}$. Since $i_2 - i_1 \sim n\delta$, we have

$$G_0\left(f_\delta\right) = O\left(1\right).$$

Thus

$$\begin{aligned}
\inf_{\hat{L} \text{ affine}} R_{\mathbf{W}_{[0,T]}(m,p,C/2)} \left(\hat{L}; \mathbf{K}_n, \tilde{\mathbf{Z}}_n \right) &= \rho_0 \left(\mathbf{W}_{[0,T]}\left(m, p, C/2\right) \right) \\
&\geq \frac{\left(L\left(f_\delta^{(k)} /2 \right) \right)^2}{1 + G_0^2\left(f\right)} \\
&= \frac{\left(\delta^{m-k-p^{-1}} \right)^2}{1 + G_0^2\left(f\right)} \\
&= O\left(n^{\left(2\tilde{H}-2\right)/\gamma_k} \right),
\end{aligned}$$

proving the theorem.

REFERENCES

CAI, T. T. and LOW, M. G. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.* **31** 1140-1153.

CAI, T. T. and LOW, M. G. (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* **32** 552-576.

DEO, R. S. (1997). Nonparametric regression with long-memory errors. *Statist. Probab. Lett.* **33** 89-94.

DONOHO, D. L. (1994). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238-270.

DONOHO, D. L. and JOHNSTONE, L. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26** 879-921.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence III. *Ann. Statist.* **19** 668-701.

DUNCAN, T. E., HU, Y. and PASIK-DUNCAN, B. (2000). Stochastic calculus for fractional Brownian motion. I. Theory. *SIAM J. Control Optim.* **38** 582-612.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21** 196-216.

IBRAGIMOV, I. A. and KHASMINSKII, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18-32.

JOHNSTONE, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sinica.* **9** 51-83.

MEMIN, J., MISHURA, Y. and VALKEILA, E. (2001). Inequalities for the moments of Wiener integrals with respect to a fractional Brownian motion. *Statist. Probab. Lett.* **51** 197-206.

SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimum kernels for density estimation at a point. *Ann. Statist.* **9** 334-346.

SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional integrals and derivatives: theory and applications.* Gordon and Breach Science Publishers.

WANG, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24** 466-484.

WANG, Y. (1997). Minimax estimation via wavelets for indirect long-memory data. *J. Statist. Plann. Inference.* **64** 45-55.

ZHAO, L. H. (1997). Minimax linear estimation in a white noise problem. *Ann. Statist.* **25** 745-755.