**Electronic Journal of Statistics**
Vol. 9 (2015) 2976–3045
ISSN: 1935-7524
DOI: 10.1214/15-EJS1101

# Drift estimation with non-gaussian noise using Malliavin Calculus

**Christian Krein**

*Mathematics Research Unit*
*University of Luxembourg*
*Luxembourg*
*e-mail:* christian.krein@education.lu

**Abstract:** The aim of this paper is to show the existence of drift estimators dominating the standard one in continuous-time models of the form $X_t = u_t + Z_t$, where $u_t$ is the drift and $Z_t$ is either a Brownian martingale or a non-martingale noise living in the second Wiener chaos. Our results are based on the use of Malliavin calculus techniques, and extend previous findings of Privault and Réveillac (2008).

**MSC 2010 subject classifications:** 60H07, 62M09, 62M20.
**Keywords and phrases:** Malliavin calculus, Cramer-Rao bound, drift estimator, Stein estimator, Rosenblatt process.

Received May 2015.

## Contents

## 1. Introduction

### *1.1. Overview*

In this paper we consider the problem of estimating a (possibly random) drift $(u_t)_{t\in[0,T]}$ in continuous-time statistical models of the type

$$X_t = u_t + Z_t, \quad t \in [0,T], \tag{1.1}$$

where $(X_t)_{t\in[0,T]}$ is the observation process, and $(Z_t)_{t\in[0,T]}$ is a noise having the form of a process living in a fixed Wiener chaos of a Brownian motion. Our results allow in particular to deal in detail with the following cases:

(i) $(Z_t)_{t\in[0,T]}$ is a chaotic Brownian martingale,
(ii) $(Z_t)_{t\in[0,T]}$ is a Rosenblatt process (living in the second Wiener chaos).

Our main finding (see Theorem 4.1) is that, under fairly general circumstances, it is possible to find an estimator of the drift, whose risk is smaller than the one of the standard estimator $X_t$.

Our results generalize the recent findings by Privault and Réveillac [20] which only dealt with the case of a Gaussian noise. As such, our work can be regarded as an infinite-dimensional extension of the seminal works by Stein [23], [24] and James-Stein [12], that first described analogous phenomena in a finite dimensional setting.

### 1.2. History and motivation

In his famous paper [23], Stein has shown that for a normally distributed $d$-dimensional random vector $X$ with unknown mean $\mu$ and covariance matrix $I$, the standard (unbiased) estimator $X$ for the mean $\mu$ is *inadmissible* with respect to the quadratic loss function if and only if $d \geqslant 3$. This means that there exists an estimator $\xi$ for the mean, verifying:

$$\mathbb{E}\left[\|\xi - \mu\|^2\right] \leqslant \mathbb{E}\left[\|X - \mu\|^2\right], \quad \mu \in \mathbb{R}^d,$$

with a strict inequality for at least one $\mu$ if and only if $d \geqslant 3$. In this case one says that $\xi$ *dominates* $X$. For $d = 1$ and $d = 2$, the standard unbiased estimator $X$ is admissible, see [23]. We recall that an estimator $\delta_1$ is said to *dominate an estimator* $\delta_2$, if for every $\mu \in \mathbb{R}^d$ we have $\mathbb{E}\left[\|\delta_1 - \mu\|^2\right] \leqslant \mathbb{E}\left[\|\delta_2 - \mu\|^2\right]$, and one has a strict inequality for at least one $\mu$. In other words, there are estimators that dominate the standard estimator $X$ for the mean if and only if $d \geqslant 3$. More precisely, Stein has shown that for $d \geqslant 3$, for sufficiently large $a > 0$ and sufficiently small $b > 0$, the estimator

$$\delta(X) = \left(1 - \frac{b}{a + \|X\|^2}\right) X = X - \frac{bX}{a + \|X\|^2} \tag{1.2}$$

is possibly biased, but has smaller risk than the standard estimator, that is:

$$\mathbb{E}\left[\|X - \mu\|^2\right] = d > \mathbb{E}\left[\|\delta(X) - \mu\|^2\right]. \tag{1.3}$$

In 1962, James and Stein [12] have proved that estimators of the form

$$\delta(X) = \left(1 - \frac{b}{\|X\|^2}\right) X = X - \frac{bX}{\|X\|^2} \tag{1.4}$$

dominate $X$ for every $0 < b < 2(d - 2)$, where $d \geqslant 3$ is again the dimension of the random vector $X$. In 1981, Stein published an important article (see [24]): he was able to give a much easier proof of the earlier result using the technique of integration by parts. He created a link to superharmonic functions and was able to give a criterion for an estimator to have smaller risk than the standard estimator.

In parallel with these developments, which concern normally distributed random variables, research has concentrated mostly on two aspects:

(i) considering spherically symmetric distributions more general than the Gaussian distribution,
(ii) finding more general estimators that dominate the standard unbiased estimator (in the normal case and in the more general symmetric distributed case).

An overview of the theory can be found in [3], [4] and [5]. We give only a few examples of these extensions. For instance, Brandwein has proved (see [2]) that for the quadratic loss function and a spherically symmetric distributed $d$-dimensional random vector $X$ such that $\mathbb{E}[\|X\|^2] < \infty$ and $\mathbb{E}[\|X\|^{-2}] < \infty$,

$$\delta(X) = \left(1 - \frac{a}{\|X\|^2}\right) X$$

dominates $X$ if $d \geqslant 4$ and $0 < a < [2(d-2)/d]/\mathbb{E}[\|X\|^{-2}]$. Brandwein and Strawderman have generalized the Stein-type estimator in 1991 by considering again the quadratic loss and more general estimators of the form

$$\delta_a(X) = X + ag(X),$$

see [4]. They show that $\delta_a(X)$ has smaller risk than $X$ if $d \geqslant 4$, $0 < a < 1/\left(d\,\mathbb{E}[\|X\|^{-2}]\right)$ and if some conditions on $g$ hold (see [4] for these conditions). It is worth noticing that superharmonic functions and the divergence theorem play a crucial role. We notice that the divergence theorem is closely connected to integration by parts. The technique of integration by parts was also used by Shinozaki [22] to prove the existence of estimators that dominate the standard estimator for the location parameters. Moreover the existence of estimators that dominate the standard estimator for the location parameters of $Z$ is proved if $\mathbb{E}[Z_i] = \mathbb{E}[Z_i^3] = 0$, $\mathbb{E}[Z_i^2] = 1$ and $\mathbb{E}[Z_i^4] = k$ and $Z_i$ are independent and identically distributed.

Apart from these results that concern all classical probability theory, integration by parts has found applications in a paper by Evans and Stark (see [10]). In connection with stochastic processes, Girsanov's theorem is used to prove the following general result concerning the existence of an estimator of the form given in Eq. (1.2) satisfying the relation (1.3): if $X = Z + \theta$ is a $d$-dimensional random vector with $d \geqslant 3$, and if $Z$ is not almost surely 0, $\mathbb{E}[Z] = 0$, $\mathbb{E}[\|Z\|^2] < \infty$ and

$$\mathbb{E}[\|Z + \theta\|^{2-d}] \leqslant \|\theta\|^{2-d}, \quad \theta \in \mathbb{R}^d,$$

then $\delta(X) = \left(1 - a/(1 + \|X\|^2)\right) X$ dominates $X$ for every sufficiently small $a > 0$. The techniques used in this last paper are non standard and are very different from those used in previous works.

In recent years, research has turned to stochastic processes. Stein's approach has shown to be effective in this field as well. Privault and Réveillac have considered the problem of estimating the drift of a Gaussian process. We sketch below the main aspects of the setting considered by the authors (for details, see [20]). For $T > 0$, the authors consider a real-valued Gaussian process $(X_t)_{t \in [0,T]}$ with covariance function

$$\gamma(s,t) = \mathbb{E}[X_s X_t], \quad s, t \in [0, T]$$

on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the $\sigma$-algebra generated by $X$. The process $(X_t)_{t \in [0,T]}$ is represented as an isonormal Gaussian process on the real separable Hilbert space $H$ generated by the functions $\chi_t : s \mapsto \min\{s, t\}$ for $s, t \in [0, T]$, with the scalar product $\langle \cdot, \cdot \rangle_H$ and the norm $\| \cdot \|_H$ defined by

$$\langle \chi_t, \chi_s \rangle_H := \gamma(s, t).$$

Then

$$X(\chi_t) := X_t, \quad t \in [0,T]$$

and $\{X(h) : h \in H\}$ is a family of centered Gaussian random variables satisfying

$$\mathbb{E}[X(h)X(g)] = \langle h, g\rangle_H, \quad h, g \in H.$$

The authors consider a one-dimensional Gaussian process $(X_t)_{t\in[0,T]}$ with

$$dX_t = \dot{u}_t dt + dX_t^u$$

where $(u_t)_{t\in[0,T]}$ is an adapted process of the form

$$u_t = \int_0^t \dot{u}_s ds, \quad t \in [0,T] \text{ with } \dot{u} \in L^2\left(\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}|[0,T]\right),$$

and $(X_t^u)_{t\in[0,T]}$ is a centered Gaussian process under a probability $\mathbb{P}_u$ which is the translation of $\mathbb{P}$ on $\Omega$ by $u$. Malliavin calculus and the integration by parts formula (see formula (3.2)) are used to construct an estimator with lower risk than the standard estimator $X$ for the drift. The key to this approach is an infinite-dimensional version of Stein's lemma. Their estimator $\delta(X)$ for a deterministic drift is *biased and anticipating*, but has smaller risk than the standard estimator $X_t$ for the drift:

$$\mathbb{E}\left[\|\delta(X) - u\|^2_{L^2([0,T],\lambda_T)}\right] < \mathbb{E}\left[\|X - u\|^2_{L^2([0,T],\lambda_T)}\right]. \tag{1.5}$$

For the ease of notation, we use the shortform $\lambda_T := \lambda|[0,T]$. The estimator $\delta(X)$ is given by

$$\begin{aligned}\delta(X) &= X_t - (d-2)\frac{[\Pi_d X]_t}{\|\Pi_d X\|_{L^2([0,T],\lambda_T)}} \\ &= X_t - (d-2)\frac{\tilde{h}_1(t)X(h_1) + \ldots + \tilde{h}_d(t)X(h_d)}{\|\tilde{h}_1(t)X(h_1) + \ldots + \tilde{h}_d(t)X(h_d)\|^2_{L^2([0,T],\lambda_T)}},\end{aligned}$$

where $\Pi_d$ denotes the orthogonal projection on the space generated by $X(h_1), \ldots, X(h_d)$, and the functions $\tilde{h}_i$ are defined through $h_i$ and the covariance structure of the underlying Gaussian process (see [20] for details). This estimator has the same form than the James-Stein estimator in Eq. (1.4). In fact, letting $X = (X_1, \ldots, X_d)^\top$, the classical James-Stein estimator in Eq. (1.4) writes as:

$$X - b\,\frac{e_1X_1 + \ldots + e_dX_d}{\|e_1X_1 + \ldots + e_dX_d\|^2}, \quad 0 < b < 2(d-2),$$

where $e_i$ denotes the $i$-th vector of the canonical basis of $\mathbb{R}^d$. Privault and Réveillac prove also an interesting Cramer-Rao type bound that allows one to compare any *unbiased and adapted* drift estimator $\xi$ with the standard drift estimator $X$ in their setting:

$$\mathbb{E}\left[\|X - u\|^2_{L^2([0,T],\lambda_T)}\right] \leqslant \mathbb{E}\left[\|\xi - u\|^2_{L^2([0,T],\lambda_T)}\right]. \tag{1.6}$$

Inequality (1.6) shows for the Gaussian process $(X_t)_{t\in[0,T]}$ that $X_t = u_t + X_t^u$ is the best unbiased estimator for the drift $u_t$ and realizes the minimal risk among all adapted drift estimators. On the other hand inequality (1.5) shows that there are *biased and anticipating* estimators that improve upon this estimator. The fact that the estimators verifying inequality (1.5) are biased has its analogue in the classical situation considered by Stein. The methods used in the setting for Gaussian processes have been applied to Poisson processes and to the Fractional Brownian motion as well (see [21] and [9]).

### *1.3. Main results and plan*

In the framework of the continuous-time model (see Eq. (1.1)), our main findings will be the following:

(1) a Cramer-Rao type bound under the assumption that $(Z_t)_{t\in[0,T]}$ is a Brownian martingale, a situation in which $X_t$ is the best unbiased adapted drift estimator, see Theorem 2.1 and Theorem 2.2;
(2) the existence of biased estimators with smaller risk than the one of the standard estimator in the case of a process living in the second Wiener chaos and under fairly general circumstances, see Theorem 4.1;
(3) applications and examples that illustrate the previous result, see Section 6.

The paper is organized as follows:

- In Section 2, we introduce the necessary notations. In this section we consider processes of the form

$$X_t = \int_0^t \dot{u}_s ds + \int_0^t b_s dW_s^u, \quad t \in [0,T].$$

  The noise $(Z_t)_{t\in[0,T]}$ is a martingale with respect to the filtration generated by the Brownian motion $(W_t^u)_{t\in[0,T]}$. For the case of a deterministic drift $(u_t)_{t\in[0,T]}$ with $\dot{u} \in L^2([0,T],\lambda_T)$, we show that the risk of an unbiased adapted drift estimator $(\xi_t)_{t\in[0,T]}$ cannot be lower than the risk of the standard estimator:

$$\int_0^T \mathbb{E}_u\left[(X_t-u_t)^2\right]dt \leqslant \int_0^T \mathbb{E}_u\left[(\xi_t-u_t)^2\right]dt.$$

- In Section 3, we give a brief introduction to Malliavin calculus. We need the basic elements of this theory for the forthcoming proofs.
- In Section 4, we consider (for a deterministic drift) processes of the form

$$X_t = u_t + Z_t = \int_0^t \dot{u}_s ds + \int_0^T\int_0^T f(x_1,x_2;t)dW_{x_1}^u dW_{x_2}^u, \quad t \in [0,T].$$

  The noise $(Z_t)_{t\in[0,T]}$ is not necessarily a martingale. We define an estimator whose risk is smaller than the risk of the standard estimator. Combining this result with the Cramer-Rao type bound of Section 2, we conclude that in the martingale case our estimator is *superefficient* and that the standard estimator is inadmissible.

- In Section 6, we apply our main result to the case where the noise $(Z_t)_{t\in[0,T]}$ is a Gaussian process, a Rosenblatt process or a chaotic Brownian martingale.
- In Section 7, we estimate a constant $a$ that is needed to define our James-Stein estimator

$$X_t - \xi_t = X_t - \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \tag{1.7}$$

for the case of a Rosenblatt process and for the case of a chaotic Brownian martingale. In Eq. (1.7), $a$ is a positive constant, $g(t)$ is a vector $(g_i(t), \dots, g_d(t))^\top$ and every $g_i : [0,T] \to \mathbb{R}$ is a continuous function, $B$ is a positive definite matrix and $X(h)$ is a vector of $d$ observations.
- In Section 8, we give without proof a discrete version of the main theorem for the case of a particular non Gaussian noise.

## 2. Cramer-Rao type bound under martingale assumptions

### 2.1. Notations

In this section we study Cramer-Rao bounds for the model in Eq. (1.1), in the case where $(Z_t)_{t\in[0,T]}$ is a square-integrable Brownian martingale. This result covers, in particular, the case of chaotic martingales — these are special cases of the noise processes studied in Section 6.1. More specifically, we consider $T > 0$, a measurable space $(\Omega, \mathcal{F})$ and a measurable mapping

$$u : (\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}|[0,T]) \to (\mathbb{R}, \mathcal{B}|[0,T]) .$$

The stochastic process $(u_t)_{t\in[0,T]}$ is called the *drift*. We suppose that we have a representation of the form

$$u_t = \int_0^t \dot{u}_s ds, \quad t \in [0,T]. \tag{2.1}$$

Suppose that there is a probability measure $\mathbb{P}_u$ on $(\Omega, \mathcal{F})$, a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ and stochastic processes $(b_t)_{t\in[0,T]}$, $(W_t^u)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ such that:

- $(\Omega, \mathcal{F}, \mathbb{P}_u)$ is a complete probability space,
- $u_t$ is $\mathcal{F}_t$-adapted,
- $b : (\Omega \times [0,T], \mathcal{F} \otimes \mathcal{B}|[0,T]) \to (\mathbb{R}, \mathcal{B}|[0,T])$ is measurable and $b_t$ is $\mathcal{F}_t$-adapted,
- $(W_t^u)_{t\in[0,T]}$ is a standard Brownian motion with respect to $\mathbb{P}_u$ and $(\mathcal{F}_t)_{t\in[0,T]}$,
- $X_t = u_t + \int_0^t b_s dW_s^u$ and $\left(\int_0^t b_s dW_s^u\right)_{t\in[0,T]}$ is a $(\mathcal{F}_t)_{t\in[0,T]}$-martingale with respect to $\mathbb{P}_u$.

We suppose moreover that:

$$\mathbb{E}_u\left[\int_0^T b_s^2 ds\right] < \infty; \quad \mathbb{E}_u\left[\int_0^T \dot{u}_s^2 ds\right] < \infty; \quad \mathbb{E}_u[b_s^2] \neq 0, \quad \forall s \in (0,T].$$

From now on, we suppose that the process $(u_t)_{t\in[0,T]}$ has a representation as in Eq. (2.1) and is square-integrable with respect to $\mathbb{P}_u \otimes \lambda_T$. The process $\left(\int_0^t b_s dW_s^u\right)_{t\in[0,T]}$ is a $(\mathcal{F}_t)_{t\in[0,T]}$-martingale with respect to $\mathbb{P}_u$. Conversely every $(\mathcal{F}_t)_{t\in[0,T]}$-martingale $(M_t)_{t\in[0,T]}$ (with respect to $\mathbb{P}_u$) has such a representation if $\mathbb{E}_u[M_0] = 0$.

A process $(\xi_t)_{t\in[0,T]}$ is called *unbiased drift estimator* if

$$\mathbb{E}_u\left[\xi_t\right] = \mathbb{E}_u[u_t], \quad t \in [0,T]$$

for all square-integrable adapted processes $(u_t)_{t\in[0,T]}$ as defined above. It is called adapted if the process $(\xi_t)_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted (see [20]).

### *2.2. Cramer-Rao type bound*

We are now going to state the two main findings of the present section, namely:

(i) Theorem 2.1, containing a Cramer-Rao bound for possibly random drifts, under some restrictive assumptions on $(b_t)_{t\in[0,T]}$;

(ii) Theorem 2.2, dealing with Cramer-Rao bounds for deterministic drifts, but with no technical assumptions on $(b_t)_{t\in[0,T]}$.

**Theorem 2.1.** *Consider a drift $(u_t)_{t\in[0,T]}$ with $u \in L^2(\Omega\times[0,T], \mathbb{P}_u \otimes \lambda_T)$ and the situation of Section 2.1 with*

$$X_t = u_t + \int_0^t b_s dW_s^u, \quad t \in [0,T].$$

*If the following conditions hold:*

(1) $\mathbb{E}_u\left[\exp\left(\frac{1}{2}\int_0^T \left(\frac{\dot{u}_s + \epsilon b_s^2}{b_s}\right)^2 ds\right)\right] < \infty$ *for every $\epsilon \in U$, where $U$ is an arbitrary small neighbourhood of 0,* *(Novikov condition)*

(2) $\mathbb{E}_u\left[\exp\left(\left|\int_0^T b_s dW_s^u\right|\right)^4\right] < \infty,$

(3) $\mathbb{E}_u\left[\left(\int_0^T b_s^2 ds\right)^4\right] < \infty,$

(4) $\int_0^\cdot b_s^2 ds \in L^2(\Omega\times[0,T], \mathbb{P}_u \otimes \lambda_T),$

*we have for every square-integrable, unbiased adapted drift estimator $(\xi_t)_{t\in[0,T]}$:*

$$\mathbb{E}_u\left[(\xi_t - u_t)^2\right] \geqslant \mathbb{E}_u\left[\int_0^t b_s^2 ds\right] = \mathbb{E}_u\left[(X_t - u_t)^2\right].$$

*In particular, $X$ realizes the minimal risk for this class of estimators.*

*Proof.* See Appendix 9.1. The proof is based on the proof given in [20]. The conditions are used to verify the conditions of Girsanov's theorem and Lebesgue's dominated convergence theorem (to interchange derivatives and integrals). □

**Theorem 2.2.** *Consider the situation of Section 2.1 with*

$$X_t = u_t + \int_0^t b_s dW_s^u, \quad t \in [0,T],$$

*and a square-integrable, unbiased adapted drift-estimator $(\xi_t)_{t\in[0,T]}$, i.e. $\xi_t$ is $\mathcal{F}_t$-adapted and:*

$$\mathbb{E}_u[\xi_t] = \mathbb{E}_u[u_t], \quad t \in [0,T],$$

*for all square-integrable $\mathcal{F}_t$-adapted processes $(u_t)_{t\in[0,T]}$. If $(u_t)_{t\in[0,T]}$ is deterministic with $\int_0^T \dot{u}_t^2 dt < \infty$, we have:*

$$\int_0^T \mathbb{E}_u\left[(\xi_t - u_t)^2\right] dt \geqslant \int_0^T \mathbb{E}_u\left[(X_t - u_t)^2\right] dt.$$

*Proof.* See Appendix 9.2. The proof is based on an approximation argument and elementary adapted processes. □

*Remark* 2.3. The techniques used in the proofs of Theorems 2.1 and 2.2 make use of the martingale property of the noise $\left(\int_0^t b_s dW_s^u\right)_{t\in[0,T]}$. For the general case, it is unknown to the author whether a Cramer-Rao type bound continues to hold if the martingale assumption is dropped. However, there are special cases outside the martingale setting for which a bound of this type holds. In [9], the authors prove the existence of a Cramer-Rao type bound if the noise $(X_t - u_t)_{t\in[0,T]}$ is a Fractional Brownian motion with Hurst parameter $0 < H < 1/2$. The proof is based on a version of Girsanov's Theorem for the Fractional Brownian motion (see [18] or [6]). Under the martingale assumptions, the theorems above stress the optimality of the standard estimator in the class of all unbiased adapted drift estimators. We will now provide some comparisons (essentially concerning the techniques in the proof) with other related bounds in the literature.

(a) In the framework of stochastic calculus and stochastic processes, Girsanov's theorem is used to prove a Cramer-Rao type inequality (see [20], [21] and [9]). The underlying idea is to use Girsanov's theorem to interchange expectation and differentiation. An inequality is found that allows the direct comparison of adapted and unbiased drift estimators with the standard estimator.
(b) Evans and Stark also use Girsanov's theorem but make no statement about the optimality of their estimator. They only affirm that for the considered class of processes, their estimator dominates the standard estimator (see [10]).
(c) James and Stein considered the special case of a random variable $X$ that is normally distributed. For this particular case, it is known that the standard estimator is the best unbiased estimator for the expectation of $X$.

(d) In the framework of classical statistics, Brandwein, Strawderman and Shinozaki do not affirm that their estimators are uniformly minimum-variance unbiased estimator. Instead they consider the class of *invariant estimators*. An overview of the theory of (location) invariant estimators can be found in [13]. The authors suppose that the random variable $X$ has a density of the form $f(x-\mu)$ for $x,\mu \in \mathbb{R}^d$. The *location parameter* $\mu$ is the mean of $X$. Then for every $a \in \mathbb{R}^d$, with $x' = x + a$ and $\mu' = \mu + a$, we have:

$$f(x' - \mu') = f(x - \mu).$$

The squared loss function shares this invariance property and the problem of estimating $\mu$ is thus called *location invariant*. Estimators $\delta$ that verify

$$\delta(X + a) = \delta(X) + a$$

are called *location invariant estimators*. In [3], [2] and [22], Brandwein, Strawderman and Shinozaki affirm that the standard estimator $X$ is the best invariant estimator for the location parameter without affirming that these estimators are uniformly minimum-variance unbiased estimators. The main reason for the consideration of location invariant estimators seems to be the fact that it is easier to make statements about their existence.

Both approaches have disadvantages for our (continuous) setting: a version of Girsanov's theorem is not always available (for instance in the case of the Rosenblatt process) and it is not clear how the (finite dimensional) discrete concept of (location) invariant estimators can be adapted to the (infinite dimensional) continuous-time setting that we consider.

## 3. Malliavin Calculus

Consider the Cameron-Martin space, that is the real separable Hilbert space

$$H = \left\{ v : [0,T] \to \mathbb{R} : v(t) = \int_0^t \dot{v}(s)ds \; ; \; \dot{v} \in L^2([0,T], \lambda_T) \right\},$$

endowed with the scalar product

$$\langle h, g \rangle_H := \int_0^T \dot{h}(t)\dot{g}(t)dt = \langle \dot{h}, \dot{g} \rangle_{L^2([0,T],\lambda_T)}$$

for all $h, g \in H$. The Hilbert space $H$ is generated by the functions

$$\chi_t : s \mapsto \min\{s,t\} = s \wedge t, \quad s,t \in [0,T].$$

Consider a standard Wiener process $(W_t^u)_{t\in[0,T]}$ (see Section 2) and a random variable $F = f_n\left(W^u(h_1), \dots, W^u(h_n)\right)$ where $n \geqslant 1$ and:

- $f_n$ is an infinitely differentiable rapidly decreasing function on $\mathbb{R}^n$ for $n \geqslant 1$,
- $h_1, \dots, h_n \in H$,
- $W^u(h_i) = \int_0^T \dot{h}(t)dW_t^u$.

Then $F$ is called a *smooth random variable*. Some authors, including Privault and Réveillac (see [20, Definition 3.4]), call

$$\sum_{i=1}^{n} h_i(t)\,\partial_i f_n(W^u(h_1),\dots,W^u(h_n)) = \sum_{i=1}^{n} h_i(t)\,\partial_i f_n\left(\int_0^T \dot{h}_1(s)dW^u_s,\dots,\int_0^T \dot{h}_n(s)dW^u_s\right)$$

the Malliavin derivative of $F$, whereas other authors such as Nualart and Øksendal (see for instance [17] or [7]) call

$$\sum_{i=1}^{n} \dot{h}_i(t)\,\partial_i f_n(W^u(h_1),\dots,W^u(h_n)) = \sum_{i=1}^{n} \dot{h}_i(t)\,\partial_i f_n\left(\int_0^T \dot{h}_i(s)dW^u_s,\dots,\int_0^T \dot{h}_n(s)dW^u_s\right) \tag{3.1}$$

the Malliavin derivative of $F$. Both definitions are equivalent. We use the definition given in Eq. (3.1) and write $D_tF$ for the Malliavin derivative of $F$. We have that $D_t$ is closable from $L^2(\Omega,\mathbb{P}_u)$ to $L^2(\Omega\times[0,T],\mathbb{P}_u\otimes\lambda_T)$, that is (see [7] or [17, Proposition 1.2.1]): *If a sequence $(H_n)_{n\in\mathbb{N}}\subset L^2(\Omega,\mathbb{P}_u)$ converges to $0$, that is $\mathbb{E}_u\left[H_n^2\right]\to 0$ if $n\to\infty$, and $D_tH_n$ converges in $L^2(\Omega\times[0,T],\mathbb{P}_u\otimes\lambda_T)$ as $n\to\infty$, then* $\lim_{n\to\infty} D_tH_n = 0$. We write Dom $D$ for the closed domain of $D$.

Moreover the Malliavin derivative has a closable adjoint $\delta$ (under $\mathbb{P}_u$). The operator $\delta$ is called the *divergence operator* or, in the white noise case, the Skorohod integral. The domain of $\delta$ is denoted by Dom $\delta$, it is the set of square-integrable random variables $v\in L^2(\Omega\times[0,T],\mathbb{P}_u\otimes\lambda_T)$ with:

$$\mathbb{E}_u\left[\langle DF,v\rangle_{L^2([0,T],\lambda_T)}\right]\leqslant c\sqrt{\mathbb{E}_u[F^2]}$$

for a constant $c$ depending on $v$ and all $F\in D^{1,2}$ where $D^{1,2}$ is the closure of the class of smooth random variables with respect of the norm

$$\|F\|_{1,2} := \left(\mathbb{E}_u\left[F^2\right]+\mathbb{E}_u\left[\|DF\|^2_{L^2([0,T],\lambda_T)}\right]\right)^{1/2}.$$

With the scalar product

$$\langle F;G\rangle_{1,2} = \mathbb{E}_u[FG]+\mathbb{E}_u[\langle DF,DG\rangle_{L^2([0,T],\lambda_T)}],$$

$D^{1,2}$ is a Hilbert space. If $v\in$ Dom $\delta$, then $\delta(v)$ is the element of $L^2(\Omega,\mathbb{P}_u)$ characterized by

$$\mathbb{E}_u[F\delta(v)] = \mathbb{E}_u[\langle v,DF\rangle_{L^2([0,T],\lambda_T)}]. \tag{3.2}$$

This relation is often called the *integration by parts formula*. We have the more general rule (see [17, Proposition 1.3.3]) for $F\in D^{1,2}$, $v\in$ Dom $\delta$ such that $Fv\in$ Dom $\delta$:

$$\delta(Fv) = F\delta(v) - \langle DF,v\rangle_{L^2([0,T],\lambda_T)}. \tag{3.3}$$

For multiple Wiener integrals, we have for symmetric and square-integrable functions $f_n$ (see for instance [17, p.35]):

$$D_{x_1} D_{x_2} \dots D_{x_k} \int_0^T \dots \int_0^T f_n(y_1, \dots, y_n) dW^u_{y_1} \dots dW^u_{y_n} \tag{3.4}$$
$$= \frac{n!}{(n-k)!} \int_0^T \dots \int_0^T f_n(y_1, \dots, y_{n-k}, x_1, \dots, x_k) dW^u_{y_1} \dots dW^u_{y_{n-k}}.$$

Formula (3.2) can be generalized by considering the multiple divergence (see [16, p.33]): If $v \in \mathrm{Dom}\ \delta^n$ and $F \in D^{n,2}$:

$$\mathbb{E}_u \left[ F \delta^n(v) \right] = \mathbb{E}_u [\langle D^n F, v \rangle_{L^2([0,T]^n,\, \lambda^n_T)}], \tag{3.5}$$

see [16] or [17] for details. For the ease of notations, we have used the shortform $\lambda^n_T := \lambda^n | [0,T]^n$.

## 4. Second chaos noise

### 4.1. Preliminary remarks

In this section we consider primarily noises $(Z_t)_{t \in [0,T]}$ that live in the second Wiener-Itô chaos and processes $(X_t)_{t \in [0,T]}$ with $X_0 = 0$ and

$$X_t = u_t + Z_t = u_t + \int_0^T \int_0^T f(x_1, x_2; t) dW^u_{x_1} dW^u_{x_2}, \quad t \in (0,T]. \tag{4.1}$$

We consider stochastic integrals that do not necessarily define martingales. The kernels of the stochastic integrals depend on the time parameter $t \in [0,T]$. This setting includes the well-known Rosenblatt process as well as a class of Brownian martingales living in the second Wiener-Itô chaos. As we already stressed in Remark 2.3, Theorem 2.1 and Theorem 2.2 cannot be applied in this section where we are outside the martingale setting. Therefore we shall compare our estimators only to the standard estimator $X$. We construct a stochastic integral with respect to our noise. Since our construction aims to be generally applicable, we consider a special class of observations. This restrictions can be relaxed when considering particular cases of our setting (see Section 6).

In Paragraph 4.1.1, we recall the basic facts of the construction of the Lebesgue-Stieltjes integral. In Paragraph 4.1.2, we construct a stochastic integral with respect to the noise $(Z_t)_{t \in [0,T]}$ following the ideas of Tudor (see [26]). In Paragraph 4.1.3, we calculate the first two Malliavin derivatives for observations $X(h)$ and we define functions $g$ which play a crucial role in the proof of the main result. In Paragraph 4.1.4, we give a brief overview of our setting and definitions.

*4.1.1. Lebesgue-Stieltjes integration*

Consider a function $\tilde{f} : [0,T] \to \mathbb{R}$ that is right-continuous and of bounded variation. We define the Lebesgue-Stieltjes integral with respect to $\tilde{f}$ and give a brief summary of the well known construction of this integral. Since $\tilde{f}$ is of bounded variation and right-continuous, we can find right-continuous non decreasing functions $\tilde{f}_1$ and $\tilde{f}_2$ with $\tilde{f} = \tilde{f}_1 - \tilde{f}_2$. For every $\tilde{f}_i$ there is a unique measure $\mu_i$ on the Borel sets on $[0,T]$ defined by:

$$\mu_i((\alpha,\beta]) = \tilde{f}_i(\beta) - \tilde{f}_i(\alpha), \qquad \text{for every } 0 \leqslant \alpha < \beta \leqslant T,$$

and $\mu_i(\varnothing) = 0$. We consider the signed measure $\mu := \mu_1 - \mu_2$. For a measurable function $g : [0,T] \to \mathbb{R}$, the Lebesgue-Stieltjes integral of $g$ with respect to $\tilde{f}$ is defined as the Lebesgue integral of $g$ with respect to $\mu$:

$$\int_0^T g d\tilde{f} = \int_0^T g d\mu.$$

The Lebesgue-Stieltjes integral above exists if we have:

$$g \in L^1([0,T],\mu) := L^1([0,T],\mu_1) \cap L^1([0,T],\mu_2).$$

*4.1.2. Stochastic integrals with respect to $(Z_t)_{t\in[0,T]}$*

We consider a process of the form

$$X_t = u_t + Z_t = u_t + \int_0^T \int_0^T f(x_1,x_2;t) dW^u_{x_1} dW^u_{x_2}, \quad t \in (0,T]$$

and $X_0 = 0$. We make the following assumptions about the drift and the kernel of the stochastic integral:

(i) the drift $(u_t)_{t\in[0,T]}$ is supposed to be deterministic and in the Cameron-Martin space

$$H = \left\{ v : [0,T] \to \mathbb{R} : v(t) = \int_0^t \dot{v}(s) ds \ ; \ \dot{v} \in L^2([0,T],\lambda_T) \right\},$$

(ii) the kernel $f(\cdot,\cdot;t)$ is supposed to be symmetric in the first two variables for every $t \in (0,T]$:

$$f(x_1,x_2;t) = f(x_2,x_1;t) \text{ for almost every } (x_1,x_2) \in [0,T]^2,$$

(iii) the kernel $f(\cdot,\cdot;t)$ is supposed to be square-integrable for every $t \in [0,T]$:

$$\int_0^T f(x_1,x_2;t)^2 dx_1 dx_2 < \infty,$$

(iv) for almost every $(x_1, x_2) \in [0,T]^2$, the function $t \mapsto f(x_1, x_2; t)$ is right-continuous and of bounded variation,

(v) the total variation $V_0^T(f(x_1, x_2; \cdot))$ is square-integrable:

$$\int_0^T \int_0^T V_0^T(f(x_1, x_2; \cdot))^2 dx_1 dx_2 < \infty,$$

(vi) for $s, t \in [0,T]$ with $|t-s| \to 0$, we have $\mathbb{E}_u[(Z_t - Z_s)^2] \to 0$,

(vii) we have $\int_0^T \mathbb{E}_u[Z_t^2] dt < \infty$.

We construct a stochastic integral with respect to $(Z_t)_{t\in[0,T]}$ by following the ideas of Tudor (see [26]). We first define the stochastic integral with respect to the noise $(Z_t)_{t\in[0,T]}$ for step functions. We define $\int_0^T 1_{(0,t]}(s) dZ_s = Z_t - Z_0 = Z_t$, and more generally for $0 \leqslant \alpha < \beta \leqslant T$:

$$\int_0^T 1_{(\alpha,\beta]}(s) dZ_s = \int_0^T 1_{(0,\beta]}(s) dZ_s - \int_0^T 1_{(0,\alpha]}(s) dZ_s = Z_\beta - Z_\alpha.$$

We have on the other hand:

$$\begin{aligned} Z_\beta - Z_\alpha &= \int_0^T \int_0^T \left(f(x_1, x_2; \beta) - f(x_1, x_2; \alpha)\right) dW_{x_1}^u dW_{x_2}^u \\ &= \int_0^T \int_0^T \left( \int_0^T 1_{(\alpha,\beta]}(s) df(x_1, x_2; s) \right) dW_{x_1}^u dW_{x_2}^u, \end{aligned}$$

thus:

$$\int_0^T 1_{(\alpha,\beta]}(s) dZ_s = \int_0^T \int_0^T \left( \int_0^T 1_{(\alpha,\beta]}(s) df(x_1, x_2; s) \right) dW_{x_1}^u dW_{x_2}^u. \tag{4.2}$$

By linearity, we can extend Eq. (4.2) to step functions $\varphi : t \mapsto \sum_i \gamma_i 1_{(\alpha_i,\beta_i]}(t)$:

$$\int_0^T \varphi(s) dZ_s = \int_0^T \int_0^T \left( \int_0^T \varphi(s) df(x_1, x_2; s) \right) dW_{x_1}^u dW_{x_2}^u \tag{4.3}$$

We extend Eq. (4.3) to a larger class of functions. In our general setting, we limit ourselves to regulated function $\varphi : [0,T] \to \mathbb{R}$. This means that the left and right limits $\varphi(x-)$ and $\varphi(x+)$, as well as $\varphi(0+)$ and $\varphi(T-)$, exist for every $x \in (0,T)$. Dieudonné [8] proved that $\varphi$ is a regulated function if and only if $\varphi$ is the limit in $L^\infty([0,T], d\lambda_T)$ of a series of step functions $(\varphi_n)_{n\in\mathbb{N}}$. We use the following inequality (see [1, p.177]):

$$\left| \int_0^T g d\tilde{f} \right| \leqslant \|g\|_\infty V_0^T(\tilde{f}), \tag{4.4}$$

where $\tilde{f}$ is a function of bounded variation and $g$ is a regulated function.

We notice that the measurability of $f(\cdot,\cdot;t)$ implies the measurability of $\int_0^T \varphi_n(t)df(x_1,x_2;t)$. Moreover, Lebesgue's theorem of dominated convergence proves that for almost every $(x_1,x_2)$:

$$\lim_{n\to\infty}\int_0^T \varphi_n(s)df(x_1,x_2;s)=\int_0^T \varphi(s)df(x_1,x_2;s). \tag{4.5}$$

We prove now that convergence in Eq. (4.5) holds in $L^2([0,T]^2,\lambda_T^2)$:

$$\begin{aligned}
&\int_0^T\int_0^T\left(\int_0^T \varphi_n(s)df(x_1,x_2;s)-\int_0^T \varphi(s)df(x_1,x_2;s)\right)^2 dx_1dx_2\\
&=\int_0^T\int_0^T\left(\int_0^T (\varphi_n(s)-\varphi(s))df(x_1,x_2;s)\right)^2 dx_1dx_2\\
&\leqslant \int_0^T\int_0^T \left(\|\varphi_n(s)-\varphi(s)\|_\infty V_0^T(f(x_1,x_2;))\right)^2 dx_1dx_2\\
&=\|\varphi_n(s)-\varphi(s)\|_\infty^2\int_0^T\int_0^T \left(V_0^T(f(x_1,x_2;))\right)^2 dx_1dx_2\\
&\to 0,\qquad \text{for } n\to\infty.
\end{aligned}$$

We have used inequality (4.4) and assumption (v) above. We conclude that $\int_0^T \varphi(s)df(x_1,x_2;s)$ is square-integrable and the convergence in Eq. (4.5) holds in $L^2([0,T]^2,\lambda_T^2)$. The stochastic integral

$$\int_0^T\int_0^T\int_0^T \varphi(s)df(x_1,x_2;s)dW_{x_1}^u dW_{x_2}^u$$

is thus well defined and by the Itô isometry we have:

$$\lim_{n\to\infty}\mathbb{E}_u\left[\left(\int_0^T \varphi_n(s)dZ_s-\int_0^T\int_0^T\left(\int_0^T \varphi(s)df(x_1,x_2;s)\right)dW_{x_1}^u dW_{x_2}^u\right)^2\right]=0.$$

We can thus define for any regulated function $\varphi$:

$$\int_0^T \varphi(s)dZ_s:=\lim_{n\to\infty}\int_0^T \varphi_n(s)dZ_s=\int_0^T\int_0^T\left(\int_0^T \varphi(s)df(x_1,x_2;s)\right)dW_{x_1}^u dW_{x_2}^u.$$

As a direct consequence we find that $\int_0^T \varphi_n(s)\dot{u}_s ds+\int_0^T \varphi_n(s)dZ_s$ converges in $L^2(\Omega,\mathbb{P}_u)$ against $\int_0^T \varphi(s)\dot{u}_s ds+\int_0^T \varphi(s)dZ_s$. We define thus:

$$\int_0^T \varphi(s)dX_s:=\lim_{n\to\infty}\int_0^T \varphi_n(s)dX_s=\lim_{n\to\infty}\left(\int_0^T \varphi_n(s)\dot{u}_s ds+\int_0^T \varphi_n(s)dZ_s\right).$$

### 4.1.3. Malliavin derivatives

We consider an absolutely continuous function $h$ with:

$$h_i(t) = \int_0^t \dot{h}_i(s)ds, \quad t \in [0,T],$$

such that $\dot{h}$ can be chosen as a regulated function. We define $u(h) := \int_0^T \dot{h}(s)\dot{u}_s ds$ and $Z(h) := \int_0^T \int_0^T \left(\int_0^T \dot{h}(s)df(x_1,x_2;s)\right) dW^u_{x_1} dW^u_{x_2}$ and similarly for the observation $X(h)$:

$$\begin{aligned} X(h) &= \int_0^T \dot{h}_i(s)dX_s = u(h) + Z(h) \\ &:= \int_0^T \dot{h}_i(s)\dot{u}_s ds + \int_0^T \int_0^T \left(\int_0^T \dot{h}_i(s)df(x_1,x_2;s)\right) dW^u_{x_1} dW^u_{x_2}. \end{aligned}$$

The first two Malliavin derivatives of $X(h)$ exist and we have:

$$D_{x_1} \int_0^T \dot{h}(s)dX_s = 2\int_0^T \left(\int_0^T \dot{h}(s)df(x_1,x_2;s)\right) dW^u_{x_2}, \tag{4.6}$$

$$D_{x_1}D_{x_2} \int_0^T \dot{h}(s)dX_s = 2\left(\int_0^T \dot{h}(s)df(x_1,x_2;s)\right). \tag{4.7}$$

We define a function $g$ by

$$g(t) := \mathrm{Cov}_u(X(h), X_t) = \int_0^T \int_0^T f(x_1,x_2;t) D_{x_1}D_{x_2}X(h)dx_1dx_2.$$

We drop the dependence on $h$ which is clear in the context. The second equality above is a consequence of Eq. (4.7) and the Itô isometry. We prove that $g$ is continuous on $(0,T)$, left-continuous in $T$ and right-continuous in 0 using assumption (vi). We have for $s,t \in [0,T]$ with $|t-s| \to 0$:

$$|g(t)-g(s)|^2 = \mathbb{E}_u[Z(h)\,(Z_t - Z_s)]^2 \leqslant \mathbb{E}_u[Z(h)^2]\ \mathbb{E}_u[(Z_t-Z_s)^2] \to 0.$$

### 4.1.4. Setting and notations

We summarize the setting and introduce some notations. We consider a stochastic process $(X_t)_{t\in[0,T]}$ with $X_0 := 0$ and for $t \in (0,T]$:

$$X_t = u_t + Z_t = u_t + \int_0^T \int_0^T f(x_1,x_2;t)dW^u_{x_1}dW^u_{x_2}.$$

We suppose moreover that:

(i) the drift $(u_t)_{t\in[0,T]}$ is supposed be deterministic and in the Cameron-Martin space

$$H = \left\{ v : [0,T] \to \mathbb{R} : v(t) = \int_0^t \dot{v}(s)ds \ ; \ \dot{v} \in L^2([0,T], \lambda_T) \right\},$$

(ii) the kernel $f(\cdot,\cdot;t)$ is supposed to be symmetric in the first two variables for every $t \in [0,T]$:

$$f(x_1,x_2;t) = f(x_2,x_1;t) \text{ for almost every } (x_1,x_2) \in [0,T]^2,$$

(iii) the kernel $f(\cdot,\cdot;t)$ is supposed to be square-integrable for every $t \in (0,T]$:

$$\int_0^T f(x_1,x_2;t)^2 dx_1 dx_2 < \infty,$$

(iv) for almost every $(x_1,x_2) \in [0,T]^2$, the function $t \mapsto f(x_1,x_2;t)$ is right-continuous and of bounded variation,

(v) the total variation $V_0^T(f(x_1,x_2;\cdot))$ is square-integrable:

$$\int_0^T \int_0^T V_0^T(f(x_1,x_2;\cdot))^2 dx_1 dx_2 < \infty,$$

(vi) for $s,t \in [0,T]$ with $|t-s| \to 0$, we have $\mathbb{E}_u[(Z_t - Z_s)^2] \to 0$,

(vii) we have $\int_0^T \mathbb{E}_u[Z_t^2]dt < \infty$.

We consider absolutely continuous functions $h_i$ for $i = 1,\dots,d$ and $d \geqslant 3$ with:

$$h_i(t) = \int_0^t \dot{h}_i(s)ds, \quad t \in [0,T],$$

such that every $\dot{h}_i$ is a regulated function. We define for $i = 1,\dots,d$:

$$\begin{aligned} X(h_i) := u(h_i) + Z(h_i) = &\int_0^T \dot{h}_i(s)\dot{u}_s ds \\ &+ \int_0^T \int_0^T \left( \int_0^T \dot{h}_i(s) df(x_1,x_2;s) \right) dW_{x_1}^u dW_{x_2}^u. \end{aligned}$$

We define for every $i = 1,\dots,d$ continuous functions $g_i$ with

$$g_i(t) := \mathrm{Cov}_u(X(h_i), X_t) = \int_0^T \int_0^T f(x_1,x_2;t)\, D_{x_1} D_{x_2} X(h_i) dx_1 dx_2, \quad t \in [0,T].$$

We finally introduce some vector notations:

$$\begin{aligned} &g(t) := (g_1(t),\dots,g_d(t))^\top, \\ &u(h) = (u(h_1),\dots,u(h_d))^\top, \quad u(h_i) := \int_0^T \dot{u}_s \dot{h}_i(s)ds, \end{aligned}$$

$$\begin{aligned} X(h) &:= (X(h_1), \ldots, X(h_d))^\top, \quad \mathbb{E}_u[X(h_i)^2] \neq 0, \\ D_{x_1}X(h) &:= (D_{x_1}X(h_1), \ldots, D_{x_1}X(h_d))^\top, \\ D_{x_1}D_{x_2}X(h) &:= (D_{x_1}D_{x_2}X(h_1), \ldots, D_{x_1}D_{x_2}X(h_d))^\top. \end{aligned}$$

We suppose that the matrix

$$\int_0^T g(t)g(t)^\top dt = \left(\int_0^T g_i(x)g_j(x)dx\right)_{i,j=1,\ldots,d}$$

is invertible (see Remark 4.3). Since this matrix is clearly symmetric and positive semi-definite, it is positive definite. We write $B$ for its inverse, $B$ is symmetric and positive definite as the inverse of a symmetric and positive definite matrix. We define for $a > 0$:

$$\xi_t := \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \quad t \in [0,T].$$

In Theorem 4.1, we consider an estimator of the form

$$X_t - \xi_t = X_t - \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \quad t \in [0,T],$$

where $a$ is a positive constant. Since $a > 0$ and $X(h)^\top BX(h) \geqslant 0$, we have $0 < 1/(a + X(h)^\top BX(h)) \leqslant 1/a$ and $0 < \mathbb{E}_u\left[1/(a + X(h)^\top BX(h))\right] \leqslant 1/a$.

### *4.2. Construction of an estimator for a second chaos noise*

We formulate now the main result of this section.

**Theorem 4.1.** *For the model discussed above with a deterministic drift and $d \geqslant 3$, we consider the following drift estimator:*

$$\begin{aligned} X_t - \xi_t &= X_t - \frac{g(t)^\top BX(h)}{a + \|g^\top BX(h)\|^2_{L^2([0,T],\lambda_T)}} \\ &= X_t - \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \quad t \in [0,T]. \end{aligned} \tag{4.8}$$

*The drift estimator $X_t - \xi_t$ has smaller risk than the standard estimator $X_t$:*

$$\int_0^T \mathbb{E}_u\left[(X_t - \xi_t - u_t)^2\right] dt < \int_0^T \mathbb{E}_u\left[(X_t - u_t)^2\right] dt \tag{4.9}$$

*for every value of $a$ that is greater than a positive constant $A$.*

*Remark* 4.2. (1) In Theorem 4.1, it is essential to find positive constants $a$ such that inequality (4.9) holds. The proof of Theorem 4.1 shows that every $a$ that is greater than some positive constant $A$, depending on $f$, $h_1, \ldots, h_d$ and $T$, satisfies inequality (4.9). The problem of finding $A$ is non trivial and is discussed in Section 7 for two special cases.

(2) Another essential assumption in Theorem 4.1 is that the number of observations $X(h_i)$ used to construct the estimator is at least 3, that is $d \geqslant 3$. In the context of stochastic processes and drift estimation, this assumption can also be found in [20] and in [9]. In the context of classical statistics and for $d$-dimensional spherical symmetric distributions with a Lebesgue density, the assumption $d \geqslant 3$ is also very common, see for instance [24], [10] or [22]. For some results, it is even necessary that the dimension satisfies the condition $d \geqslant 4$, see for instance [4]. It is worth noticing that the usual condition $d \geqslant 3$ may not be needed for some discrete settings. In [21], the authors estimate the intensities of a Poisson process and the estimators constructed dominate the standard estimator for every dimension $d \geqslant 1$.

*Proof of Theorem 4.1.* We notice that $\xi_t$ is square-integrable with respect to $\mathbb{P}_u \otimes \lambda_T$. We show below that $\int_0^T \mathbb{E}_u[\xi_t^2]dt \leqslant \mathbb{E}_u\left[(a + X(h)^\top BX(h))^{-1}\right] < \infty$. We have:

$$
\begin{aligned}
&\int_0^T \mathbb{E}_u\left[(X_t - u_t)^2\right] dt - \int_0^T \mathbb{E}_u\left[(X_t - \xi_t - u_t)^2\right] dt \\
&= 2\int_0^T \mathbb{E}_u\left[(X_t - u_t)\,\xi_t\right] dt - \int_0^T \mathbb{E}_u\left[\xi_t^2\right] dt.
\end{aligned}
$$

Before we prove the theorem, we transform both terms in this expression. The proof is complete if we find that the expression above is positive for some $a > A > 0$. We use that $B$ is symmetric and have:

$$
\begin{aligned}
\|g^\top BX(h)\|^2_{L^2([0,T],\lambda_T)} &= \int_0^T \left(g(t)^\top BX(h)\right)^2 dt \\
&= \int_0^T X(h)^\top B^\top g(t)\, g(t)^\top BX(h)dt \\
&= X(h)^\top B^\top \int_0^T g(t)g(t)^\top dt BX(h) \\
&= X(h)^\top B^\top B^{-1} BX(h) \\
&= X(h)^\top BX(h) \qquad (B = B^\top \text{ since } B \text{ is symmetric}).
\end{aligned}
$$

We have for the second term:

$$
\begin{aligned}
\int_0^T \mathbb{E}_u\left[\xi_t^2\right] dt &= \int_0^T \mathbb{E}_u\left[\frac{\left(g(t)^\top BX(h)\right)^2}{\left(a + X(h)^\top BX(h)\right)^2}\right] dt \\
&= \mathbb{E}_u\left[\frac{X(h)^\top BX(h)}{\left(a + X(h)^\top BX(h)\right)^2}\right].
\end{aligned}
$$

We transform the first term using Malliavin calculus. We use in particular integration by parts, see Eq. (3.5). Notice also that, for deterministic functions, the iterated divergence operator $\delta^2$ coincides with the double Wiener-Itô integral, hence $\delta^2(f(\cdot,\cdot;t)) = \int_0^T\int_0^T f(x_1,x_2;t)dW^u_{x_1}dW^u_{x_2}$. Thus, with the classical Fubini theorem:

$$\begin{aligned}\int_0^T \mathbb{E}_u\left[(X_t-u_t)\,\xi_t\right]dt &= \int_0^T \mathbb{E}_u\left[\int_0^T\int_0^T f(x_1,x_2;t)dW^u_{x_1}dW^u_{x_2}\,\xi_t\right]dt\\ &= \int_0^T \mathbb{E}_u\left[\delta^2(f(,;t))\,\xi_t\right]dt\\ &= \int_0^T \mathbb{E}_u\left[\int_0^T\int_0^T f(x_1,x_2;t)D_{x_1}D_{x_2}\xi_t dx_1dx_2\right]dt\\ &= \int_0^T\int_0^T\int_0^T f(x_1,x_2;t)\,\mathbb{E}_u\left[D_{x_1}D_{x_2}\xi_t\right]dx_1dx_2dt\\ &= 2\int_0^T\int_0^T\int_0^{x_2} f(x_1,x_2;t)\,\mathbb{E}_u\left[D_{x_1}D_{x_2}\xi_t\right]dx_1dx_2dt.\end{aligned}$$

This result can also be found using the Wiener-Itô chaos decomposition of $\xi_t$ and the fact that $X_t-u_t$ lives in the second Wiener-Itô chaos. We notice that $B$ is defined to be symmetric positive semi-definite. Since $B$ is moreover supposed to be invertible, we have that $B$ is symmetric positive definite. Thus $B$ has a matrix square root $C$ that is again symmetric positive definite and $B=C^2=C^\top C$. We use $X(h)=u(h)+Z(h)$ and the following inequalities with $a>0$ and $k\in(0,1)$ (see Theorem 9.1 for the proofs):

$$\frac{a+u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} \geqslant \left(1+\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^{-2},$$

$$\frac{a+u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} \leqslant \frac{1}{k^2}+\frac{Z(h)^\top BZ(h)}{(k-1)^2a},$$

and for $Q=\frac{1}{a+X(h)^\top BX(h)}$ respectively $Q=\frac{1}{k^2}+\frac{Z(h)^\top BZ(h)}{a(1-k)^2}$:

$$\int_0^T\int_0^T\int_0^{x_2}\mathbb{E}_u\left[|f(x_1,x_2;t)|\,\|Cg(t)\|\,\|CD_{x_2}X(h)\|\,\|CD_{x_1}X(h)\|\,Q\right]dx_1dx_2dt<\infty,$$

where $\|\cdot\|$ is the standard euclidean norm. For the first and second Malliavin derivative of $\xi_t := [g(t)^\top BX(h)]/[a+X(h)^\top BX(h)]$, we obtain by the means of the chain rule:

$$\begin{aligned}D_{x_2}\xi_t &= \frac{g(t)^\top BD_{x_2}X(h)}{a+X(h)^\top BX(h)} - 2\,\frac{\left[g(t)^\top BX(h)\right]\left[X(h)^\top BD_{x_2}X(h)\right]}{\left(a+X(h)^\top BX(h)\right)^2},\\ D_{x_1}D_{x_2}\xi_t &= \frac{g(t)^\top BD_{x_1}D_{x_2}X(h)}{a+X(h)^\top BX(h)}\\ &\quad -2\,\frac{\left[g(t)^\top BD_{x_2}X(h)\right]\left[X(h)^\top BD_{x_1}X(h)\right]}{\left(a+X(h)^\top BX(h)\right)^2}\end{aligned}$$

$$
\begin{aligned}
&- 2\, \frac{\left[g(t)^\top B D_{x_1} X(h)\right]\left[X(h)^\top B D_{x_2} X(h)\right]}{\left(a + X(h)^\top B X(h)\right)^2} \\
&- 2\, \frac{\left[g(t)^\top B X(h)\right]\left[D_{x_1} X(h)^\top B D_{x_2} X(h)\right]}{\left(a + X(h)^\top B X(h)\right)^2} \\
&- 2\, \frac{\left[g(t)^\top B X(h)\right]\left[X(h)^\top B D_{x_1} D_{x_2} X(h)\right]}{\left(a + X(h)^\top B X(h)\right)^2} \\
&+ 8\, \frac{\left[g(t)^\top B X(h)\right]\left[X(h)^\top B D_{x_1} X(h)\right]\left[X(h)^\top B D_{x_2} X(h)\right]}{\left(a + X(h)^\top B X(h)\right)^3}.
\end{aligned}
$$

We estimate $\int_0^T \mathbb{E}_u\left[\xi_t^2\right] dt$ and the six terms of

$$
2\int_0^T \mathbb{E}_u\left[(X_t - u_t)\xi_t\right] dt = 4\int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t)\,\mathbb{E}_u\left[D_{x_1} D_{x_2} \xi_t\right] dx_1 dx_2 dt.
$$

(i) The sum of entries of $A * A^{-1}$ equals $d$ for any invertible symmetric matrix $A \in \mathbb{R}^{d\times d}$ and the Hadamard product $*$ of matrices (see Theorem 9.2). Thus:

$$
\int_0^T g(t)^\top B g(t) dt = \sum_{i,j} B_{i,j} \int_0^T g_i(t) g_j(t) dt = \sum_{i,j} B_{i,j} \left(B^{-1}\right)_{i,j} = d, \quad (4.10)
$$

and for $a > 0$:

$$
\begin{aligned}
& 4\int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t)\,\mathbb{E}_u\left[\frac{g(t)^\top B D_{x_1} D_{x_2} X(h)}{a + X(h)^\top B X(h)}\right] dx_1 dx_2 dt \\
&= 2\int_0^T \int_0^T \int_0^T f(x_1, x_2; t)\,\mathbb{E}_u\left[\frac{g(t)^\top B D_{x_1} D_{x_2} X(h)}{a + X(h)^\top B X(h)}\right] dx_1 dx_2 dt \\
&= 2\int_0^T g(t)^\top B \int_0^T \int_0^T f(x_1, x_2; t) D_{x_1} D_{x_2} X(h) dx_1 dx_2 dt \\
&\quad \times \mathbb{E}_u\left[\frac{1}{a + X(h)^\top B X(h)}\right] \\
&= 2\int_0^T g(t)^\top B g(t) dt\; \mathbb{E}_u\left[\frac{1}{a + X(h)^\top B X(h)}\right] \\
&= \mathbb{E}_u\left[\frac{2d}{a + X(h)^\top B X(h)}\right].
\end{aligned}
$$

The first equality follows from the symmetry in $x_1, x_2$, the second equality follows since $X(h_i)$ is a random variable in the second Wiener-Itô chaos and has therefore a deterministic second Malliavin derivative. The third equality follows from the definition of

$$
g_i : t \mapsto \int_0^T \int_0^T f(x_1, x_2; t) D_{x_1} D_{x_2} X(h_i) dx_1 dx_2.
$$

The last equality follows with Eq. (4.10).

(ii) For the ease of notation, we write $\|v_1\|$ for the standard euclidean norm of a vector $v \in \mathbb{R}^d$. We have $B = C^2 = C^\top C$ and the Cauchy-Schwarz inequality yields for arbitrary vectors $v_1, v_2 \in \mathbb{R}^d$:

$$\left|v_1^\top B v_2\right| = \left|(Cv_1)^\top (Cv_2)\right| \leqslant \|Cv_1\| \, \|Cv_2\|.$$

Thus for $a > 0$:

$$\begin{aligned}
&8 \left| \int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t) \right. \\
&\quad \left. \times \mathbb{E}_u \left[ \frac{[g(t)^\top B D_{x_2} X(h)] \, [X(h)^\top B D_{x_1} X(h)]}{\left(a + X(h)^\top B X(h)\right)^2} \right] dx_1 dx_2 dt \right| \\
&\leqslant 8 \int_0^T \int_0^T \int_0^{x_2} |f(x_1, x_2; t)| \\
&\quad \times \mathbb{E}_u \left[ \frac{[\|Cg(t)\| \, \|CD_{x_2} X(h)\|] \, [\|CX(h)\| \, \|CD_{x_1} X(h)\|]}{\left(a + X(h)^\top B X(h)\right)^2} \right] dx_1 dx_2 dt \\
&\leqslant 8 \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \frac{\|Cg(t)\| \, \|CD_{x_2} X(h)\| \, \|CD_{x_1} X(h)\|}{\sqrt{a}} \right. \\
&\quad \left. \times \sqrt{\frac{\|CX(h)\|^2}{a + X(h)^\top B X(h)}} \, \frac{|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \right] dx_1 dx_2 dt \\
&= 8 \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \frac{\|Cg(t)\| \, \|CD_{x_2} X(h)\| \, \|CD_{x_1} X(h)\|}{\sqrt{a}} \right. \\
&\quad \left. \times \sqrt{\frac{X(h)^\top B X(h)}{a + X(h)^\top B X(h)}} \, \frac{|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \right] dx_1 dx_2 dt \\
&\leqslant \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \frac{\|Cg(t)\| \, \|CD_{x_2} X(h)\| \, \|CD_{x_1} X(h)\|}{\sqrt{a}} \right. \\
&\quad \left. \times \frac{8|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \right] dx_1 dx_2 dt.
\end{aligned}$$

(iii) The next two terms can be estimated similarly. We find for $a > 0$:

$$\begin{aligned}
&8 \left| \int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t) \right. \\
&\quad \left. \times \mathbb{E}_u \left[ \frac{[g(t)^\top B D_{x_1} X(h)] \, [X(h)^\top B D_{x_2} X(h)]}{\left(a + X(h)^\top B X(h)\right)^2} \right] dx_1 dx_2 dt \right| \\
&\leqslant \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \frac{8|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \right. \\
&\quad \left. \times \frac{\|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CX(h)\| \, \|CD_{x_2} X(h)\|}{\sqrt{a} \sqrt{a + \|CX(h)\|^2}} \right] dx_1 dx_2 dt
\end{aligned}$$

$$\leqslant \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \Bigg[ \frac{\|Cg(t)\| \, \|CD_{x_2} X(h)\| \, \|CD_{x_1} X(h)\|}{\sqrt{a}} \times \frac{8|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \Bigg] dx_1 dx_2 dt,$$

and:

$$8 \left| \int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t) \times \mathbb{E}_u \left[ \frac{[g(t)^\top B X(h)] \, [D_{x_1} X(h)^\top B D_{x_2} X(h)]}{(a + X(h)^\top B X(h))^2} \right] dx_1 dx_2 dt \right|$$
$$\leqslant \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \Bigg[ \frac{8|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \times \frac{\|Cg(t)\| \, \|CX(h)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_2} X(h)\|}{\sqrt{a} \sqrt{a + \|CX(h)\|^2}} \Bigg] dx_1 dx_2 dt$$
$$\leqslant \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \Bigg[ \frac{\|Cg(t)\| \, \|CD_{x_2} X(h)\| \, \|CD_{x_1} X(h)\|}{\sqrt{a}} \times \frac{8|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \Bigg] dx_1 dx_2 dt.$$

(iv) We have for $a > 0$:

$$8 \left| \int_0^T \int_0^T \int_0^{x_2} f(x_1, x_2; t) \times \mathbb{E}_u \left[ \frac{[g(t)^\top B X(h)] \, [X(h)^\top B D_{x_1} D_{x_2} X(h)]}{(a + X(h)^\top B X(h))^2} \right] dx_1 dx_2 dt \right|$$
$$= 4 \left| \int_0^T \mathbb{E}_u \Bigg[ [g(t)^\top B X(h)] \times \frac{[X(h)^\top B \int_0^T \int_0^T f(x_1, x_2; t) D_{x_1} D_{x_2} X(h) dx_1 dx_2]}{(a + X(h)^\top B X(h))^2} \Bigg] dt \right|$$
$$= 4 \left| \int_0^T \mathbb{E}_u \left[ \frac{[g(t)^\top B X(h)] \, [X(h)^\top B g(t)]}{(a + X(h)^\top B X(h))^2} \right] dt \right|.$$

We transform this last expression using the definition of $B$:

$$4 \left| \int_0^T \mathbb{E}_u \left[ \frac{[g(t)^\top B X(h)] \, [X(h)^\top B g(t)]}{(a + X(h)^\top B X(h))^2} \right] dt \right|$$
$$= 4 \left| \mathbb{E}_u \left[ \frac{X(h)^\top B \int_0^T g(t) g(t)^\top dt B X(h)}{(a + X(h)^\top B X(h))^2} \right] \right|$$

$$
\begin{aligned}
&= 4\left|\mathbb{E}_u\left[\frac{X(h)^\top BX(h)}{\left(a+X(h)^\top BX(h)\right)^2}\right]\right| \\
&= 4\left|\mathbb{E}_u\left[\frac{X(h)^\top BX(h)}{a+X(h)^\top BX(h)}\,\frac{1}{a+X(h)^\top BX(h)}\right]\right| \\
&\leqslant 4\,\mathbb{E}_u\left[\frac{1}{a+X(h)^\top BX(h)}\right].
\end{aligned}
$$

(v) Analogue estimations to the ones above show for $a > 0$:

$$
\begin{aligned}
&32\left|\int_0^T\int_0^T\int_0^{x_2} f(x_1,x_2;t)\,\mathbb{E}_u\left[\left[g(t)^\top BX(h)\right]\right.\right. \\
&\quad\left.\left.\times\,\frac{\left[X(h)^\top BD_{x_1}X(h)\right]\left[X(h)^\top BD_{x_2}X(h)\right]}{\left(a+X(h)^\top BX(h)\right)^3}dx_1dx_2dt\right|\right. \\
&\leqslant \int_0^T\int_0^T\int_0^{x_2}\mathbb{E}_u\left[\frac{32|f(x_1,x_2;t)|\,\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|}{\sqrt{a}\,[a+X(h)^\top BX(h)]}\right. \\
&\quad\left.\times\,\frac{\|CX(h)\|\,\|CX(h)\|\,\|CX(h)\|}{\left(a+\|CX(h)\|^2\right)^{3/2}}\right]dx_1dx_2dt \\
&\leqslant \int_0^T\int_0^T\int_0^{x_2}\mathbb{E}_u\left[\frac{\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|}{\sqrt{a}}\right. \\
&\quad\left.\times\frac{32|f(x_1,x_2;t)|}{a+X(h)^\top BX(h)}\right]dx_1dx_2dt.
\end{aligned}
$$

(vi) We have for $a > 0$:

$$
\begin{aligned}
\int_0^T\mathbb{E}_u\left[\xi_t^2\right]dt &= \mathbb{E}_u\left[\frac{X(h)^\top BX(h)}{a+X(h)^\top BX(h)}\,\frac{1}{a+X(h)^\top BX(h)}\right] \\
&\leqslant \mathbb{E}_u\left[\frac{1}{a+X(h)^\top BX(h)}\right].
\end{aligned}
$$

We combine now the estimations found above for $a > 0$, $d \geqslant 3$ and $k \in (0,1)$:

$$
\begin{aligned}
&2\int_0^T\mathbb{E}_u\left[(X_t-u_t)\,\xi_t\right]dt - \int_0^T\mathbb{E}_u\left[\xi_t^2\right]dt \\
&\geqslant \mathbb{E}_u\left[\frac{2d}{a+X(h)^\top BX(h)}\right] \\
&\quad -(8+8+8+32)\int_0^T\int_0^T\int_0^{x_2}\mathbb{E}_u\left[\frac{|f(x_1,x_2;t)|}{a+X(h)^\top BX(h)}\right. \\
&\quad\left.\times\,\frac{\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|}{\sqrt{a}}\right]dx_1dx_2dt \\
&\quad -\mathbb{E}_u\left[\frac{4}{a+X(h)^\top BX(h)}\right]-\mathbb{E}_u\left[\frac{1}{a+X(h)^\top BX(h)}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_u \left[ \frac{2d-5}{a + X(h)^\top B X(h)} \right] - 56 \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \frac{|f(x_1, x_2; t)|}{a + X(h)^\top B X(h)} \right.\\
&\quad \left. \times \frac{\|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_2} X(h)\|}{\sqrt{a}} \right] dx_1 dx_2 dt \\
&= \frac{1}{a + u(h)^\top B u(h)} \left\{ (2d-5) \ \mathbb{E}_u \left[ \frac{a + u(h)^\top B u(h)}{a + X(h)^\top B X(h)} \right] \right.\\
&\quad - 56 \int_0^T \int_0^T \int_0^{x_2} |f(x_1, x_2; t)| \ \mathbb{E}_u \left[ \frac{a + u(h)^\top B u(h)}{a + X(h)^\top B X(h)} \right.\\
&\quad \left.\left. \times \frac{\|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_2} X(h)\|}{\sqrt{a}} \right] dx_1 dx_2 dt \right\}.
\end{aligned}
$$

Together with the inequalities of Theorem 9.1, we find:

$$
\begin{aligned}
&2 \int_0^T \mathbb{E}_u \left[ (X_t - u_t) \, \xi_t \right] dt - \int_0^T \mathbb{E}_u \left[ \xi_t^2 \right] dt \\
&\geqslant \frac{1}{a + u(h)^\top B u(h)} \left\{ (2d-5) \, \mathbb{E}_u \left[ \left( 1 + \sqrt{\frac{Z(h)^\top B Z(h)}{a}} \right)^{-2} \right] \right.\\
&\quad - \frac{56}{\sqrt{a}} \int_0^T \int_0^T \int_0^{x_2} |f(x_1, x_2; t)| \ \mathbb{E}_u \left[ \|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_2} X(h)\| \right.\\
&\quad \left.\left. \times \left( \frac{1}{k^2} + \frac{Z(h)^\top B Z(h)}{a(1-k)^2} \right) \right] dx_1 dx_2 dt \right\} \\
&= \frac{1}{a + u(h)^\top B u(h)} \left\{ (2d-5) \, \mathbb{E}_0 \left[ \left( 1 + \sqrt{\frac{X(h)^\top B X(h)}{a}} \right)^{-2} \right] \right.\\
&\quad - \frac{56}{\sqrt{a}} \int_0^T \int_0^T \int_0^{x_2} |f(x_1, x_2; t)| \ \mathbb{E}_0 \left[ \|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_2} X(h)\| \right.\\
&\quad \left.\left. \times \left( \frac{1}{k^2} + \frac{X(h)^\top B X(h)}{a(1-k)^2} \right) \right] dx_1 dx_2 dt \right\}.
\end{aligned}
$$

In the last step we have used that $X_t = u_t + Z_t$ and thus $X_t = Z_t$ if $u = 0$. The expression in brackets is positive if $a$ is chosen large enough, more precisely for $a > A > 0$ (see Section 7). The expectations in the last inequality above can be calculated without knowing the drift $u$. We notice that:

$$
\lim_{a \to +\infty} \mathbb{E}_0 \left[ \left( 1 + \sqrt{\frac{X(h)^\top B X(h)}{a}} \right)^{-2} \right] = 1.
$$

The results of Theorem 9.1 imply that:

$$
\int_0^T \int_0^T \int_0^{x_2} |f(x_1, x_2; t)| \, \mathbb{E}_0 \left[ \|Cg(t)\| \, \|CD_{x_1} X(h)\| \, \|CD_{x_1} X(h)\| \right.
$$

$$\times \left(\frac{1}{k^2} + \frac{X(h)^\top BX(h)}{a(1-k)^2}\right)\Bigg] dx_1 dx_2 dt$$

is (for a fixed $k \in (0,1)$) finite and bounded as a function of $a \geqslant \epsilon > 0$ (for any $\epsilon > 0$). This completes the proof. □

*Remark* 4.3. The problem of finding an optimal set of functions $g_i$ that define an invertible matrix

$$\left(\int_0^T g_i(t)g_j(t)dt\right)_{i,j=1,\dots,d}$$

is non trivial. From a practical point of view, $g_i$ is best accessible if $\dot{h}_i = 1_{(0,t_i]}$ for $0 < t_1 < t_2 < \dots < t_d \leqslant T$. If $\mathrm{Cov}_u(Z_{t_i}, Z_{t_j})_{i,j}$ is an invertible matrix, then the $g_i$ are continuous, linearly independent functions. The Gram-Schmidt algorithm with the standard scalar product on $L^2([0,T]^2, \lambda_T^2)$ can be used to find an invertible lower triangular matrix $L$ such that $(Lg)_1, \dots, (Lg)_d$ are orthonormalized. We have:

$$I = \int_0^T (Lg(t))(Lg(t))^\top dt = L\int_0^T g(t)g(t)^\top dt L^\top.$$

This shows that $\int_0^T g(t)g(t)^\top dt$ is invertible and equal to $L^{-1}L^{-\top} = \left(L^\top L\right)^{-1}$. We conclude that $\int_0^T g(t)g(t)^\top dt$ and the inverse $B$ are symmetric positive definite.

## 5. Extensions

In this section, we point out extensions of the previous results. Since the methods of Section 4 are useful in more general settings but the calculations are lengthy, we do not provide complete proofs for the results of this section.

### 5.1. A more general setting for the noise

An analogue version of Theorem 4.1 holds for a process $(X_t)_{t\in[0,T]}$ with $X_0 = 0$ and for $t \in (0,T]$:

$$\begin{aligned} X_t &= u_t + Z_t \\ &= u_t + \int_0^T \cdots \int_0^T f(x_1,\dots,x_n;t)dW^u_{x_1}\dots dW^u_{x_n}, \end{aligned} \tag{5.1}$$

if the analogue, $n$-dimensional version of conditions (i)-(vii) hold. Moreover a version of Theorem 4.1 holds if $(X_t)_{t\in[0,T]}$ is a process with $X_0 = 0$ and

$$\begin{aligned} X_t &= u_t + Z_t \\ &= \int_0^T f_1(x_1;t)dW^u_{x_1} + \int_0^T\int_0^T f_2(x_1,x_2;t)dW^u_{x_1}dW^u_{x_2}, \quad t \in (0,T], \end{aligned} \tag{5.2}$$

if conditions (i)-(vii) hold for $f_1$ and $f_2$. We have thus:

**Proposition 5.1.** *For the settings of Eq.* (5.1) *respectively Eq.* (5.2) *and a deterministic drift, we consider the following drift estimator:*

$$\begin{aligned} X_t - \xi_t &= X_t - \frac{g(t)^\top BX(h)}{a + \|g^\top BX(h)\|^2_{L^2([0,T],\lambda_T)}} \\ &= X_t - \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \quad t \in [0,T], \end{aligned}$$

*where* $g(t) := (g_1(t), \ldots, g_d(t))^\top$ *with* $g_i : t \mapsto Cov_u(X_t, X(h_i))$ *and* $d \geqslant 3$. *The drift estimator* $X_t - \xi_t$ *has smaller risk than the standard estimator* $X_t$*:*

$$\int_0^T \mathbb{E}_u \left[(X_t - \xi_t - u_t)^2\right] dt < \int_0^T \mathbb{E}_u \left[(X_t - u_t)^2\right] dt$$

*for every value of* $a$ *greater than a positive constant* $A$.

*Remark* 5.2. (1) For the setting of Eq. (5.1) and a noise living in the Wiener-Itô chaos of order $n$, the Malliavin derivative $D_{x_1} \ldots D_{x_n} \xi_t$ for

$$\xi_t = \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}$$

is needed. Calculating these derivatives becomes increasingly complicated as $n$ grows. We show below that only two of the terms appearing in $D_{x_1} \ldots D_{x_n} \xi_t$ are relevant for the proof and that all the terms can be estimated as in the proof of Theorem 4.1. We follow an idea that goes back to Meyer [15]. The Malliavin derivative satisfies the product rule, thus for smooth random variables $F$ and $G$ and $\underline{n} := \{1, \ldots, n\}$:

$$D_{x_1} \ldots D_{x_n}(F\,G) = \sum_{S \subseteq \underline{n}} (D_{x_S} F) \left(D_{x_{\overline{S}}} G\right).$$

We use the notation $D_{x_S} F = D_{x_1} \ldots D_{x_l} F$ for any subset $S = \{1, \ldots, l\}$ of $\underline{n}$. Thus:

$$\begin{aligned} D_{x_1} \ldots D_{x_n}(g(t)^\top BX(h)) &= D_{x_1} \ldots D_{x_n} \left[\xi_t \left(a + X(h)^\top BX(h)\right)\right], \\ g(t)^\top BD_{x_1} \ldots D_{x_n} X(h) &= \sum_{S \subsetneqq \underline{n}} (D_{x_S} \xi_t) \left(D_{x_{\overline{S}}}(a + X(h)^\top BX(h))\right) \\ &\quad + (D_{x_1} \ldots D_{x_n} \xi_t)\,(a + X(h)^\top BX(h)). \end{aligned}$$

We have thus:

$$\begin{aligned} &D_{x_1} \ldots D_{x_n} \xi_t \\ &= \frac{g(t)^\top BD_{x_1} \ldots D_{x_n} X(h)}{a + X(h)^\top BX(h)} - \sum_{S \subsetneqq \underline{n}} (D_{x_S} \xi_t) \frac{D_{x_{\overline{S}}}(a + X(h)^\top BX(h))}{a + X(h)^\top BX(h)} \\ &= \frac{g(t)^\top BD_{x_1} \ldots D_{x_n} X(h)}{a + X(h)^\top BX(h)} \qquad (5.3) \end{aligned}$$

$$
\begin{aligned}
&- 2\,\frac{\left[g(t)^\top BX(h)\right]\left[X(h)^\top BD_{x_1}\dots D_{x_n}X(h)\right]}{\left[a+X(h)^\top BX(h)\right]^2} \qquad (5.4)\\
&- \sum_{\varnothing\neq T\subsetneq \underline{n}} \frac{\left[g(t)^\top BX(h)\right]\left[D_{x_T}X(h)^\top BD_{x_{\underline{n}-T}}X(h)\right]}{\left[a+X(h)^\top BX(h)\right]^2}\\
&- \sum_{\varnothing\neq S\subsetneq \underline{n}} (D_{x_S}\xi_t)\,\frac{D_{x_{\overline{S}}}\left(a+X(h)^\top BX(h)\right)}{a+X(h)^\top BX(h)}.
\end{aligned}
$$

It can be seen by induction over $n$ that

$$
\begin{aligned}
&\left|\sum_{\varnothing\neq T\subsetneq \underline{n}} \frac{\left[g(t)^\top BX(h)\right]\left[D_{x_T}X(h)^\top BD_{x_{\underline{n}-T}}X(h)\right]}{\left[a+X(h)^\top BX(h)\right]^2}\right.\\
&\qquad\left.+\sum_{\varnothing\neq S\subsetneq \underline{n}} (D_{x_S}\xi_t)\,\frac{D_{x_{\overline{S}}}\left(a+X(h)^\top BX(h)\right)}{a+X(h)^\top BX(h)}\right|\\
&\leqslant \frac{R(x_1,\dots,x_n)}{\left[a+X(h)^\top BX(h)\right]^{3/2}},
\end{aligned}
$$

where $R$ is square-integrable with respect to $\mathbb{P}_u\otimes\lambda_T^n$ and does not depend on $a$ or $u$. If we choose again $g_i(t) := \operatorname{Cov}_u(X_t, X(h_i))$ and $B = \left(\int_0^T g(t)g(t)^\top dt\right)^{-1}$, we can proceed with the terms in lines (5.3) and (5.4) as in Theorem 4.1 and prove:

$$
2\int_0^T \mathbb{E}_u[(X_t-u_t)\,\xi_t]dt - \int_0^T \mathbb{E}_u\left[\xi_t^2\right]dt > 0,
$$

for $a>0$ large enough and $d\geqslant 3$. We conclude that the estimator given by

$$
X_t - \xi_t := X_t - \frac{g(t)^\top BX(h)}{a+X(h)^\top BX(h)}, \quad t\in[0,T]
$$

has smaller risk than $X_t$ if $a$ is large enough, $d\geqslant 3$ and $X_t = u_t + Z_t$ and $(Z_t)_{t\in[0,T]}$ has the form given by Eq. (4.1) but lives in a Wiener chaos of higher order.

(2) For the setting of Eq. (5.2), the method used in Theorem 4.1 is applicable as well. Notice however that in this situation extra terms appear that can be estimated applying once again the integration by parts formula.

### *5.2. Absolutely continuous kernels*

We consider a stochastic process $(X_t)_{t\in[0,T]}$ with $X_0 := 0$ and:

$$
X_t = u_t + Z_t = u_t + \int_0^T\int_0^T f(x_1,x_2;t)dW_{x_1}^u dW_{x_2}^u, \quad t\in(0,T], \qquad (5.5)
$$

where $t \mapsto f(x_1, x_2; t)$ is absolutely continuous with:

$$f(x_1, x_2; t) = \int_0^t k(x_1, x_2; a) da.$$

We can replace the assumptions of Section 4 by the following assumptions that are more appropriate for the case of absolutely continuous kernels. Notice that assumptions (iv)-(vii) are used to define stochastic integrals, guarantee the existence of expectations and show the continuity of the functions $g_i$. These properties can be proved more efficiently with conditions (i)-(iii) and the following conditions (iv') and (v'):

(iv') the function $f$ is absolutely continuous with respect to $t$:

$$f(x_1, x_2; t) = \int_0^t k(x_1, x_2; v) dv, \quad t \in [0, T],$$

(v') the function

$$\gamma : (a, b) \mapsto 2 \int_0^T \int_0^T k(x_1, x_2; a) k(x_1, x_2; b) dx_1 dx_1$$

is in $L^{q_1}([0, T]^2, \lambda_T^2)$ for a $q_1 > 1$ (then $(a, b) \mapsto \mathbb{E}_u[Z_a Z_b]$ has mixed second-order derivatives a.e. and they are equal to $\gamma$ a.e.), we write $q_2$ to indicate the real such that $1/q_1 + 1/q_2 = 1$.

Theorem 4.1 holds and proves the existence of estimators with smaller risk than the standard estimator.

**Proposition 5.3.** *For the settings of Eq.* (5.5) *and a deterministic drift, we consider the following drift estimator:*

$$\begin{aligned} X_t - \xi_t = X_t - \frac{g(t)^\top B X(h)}{a + \|g^\top B X(h)\|^2_{L^2([0,T],\lambda_T)}} \\ = X_t - \frac{g(t)^\top B X(h)}{a + X(h)^\top B X(h)}, \quad t \in [0, T], \end{aligned}$$

*where* $g(t) := (g_1(t), \ldots, g_d(t))^\top$ *with* $g_i : t \mapsto Cov_u(X_t, X(h_i))$ *and* $d \geqslant 3$*. The drift estimator* $X_t - \xi_t$ *has smaller risk than the standard estimator* $X_t$*:*

$$\int_0^T \mathbb{E}_u \left[ (X_t - \xi_t - u_t)^2 \right] dt < \int_0^T \mathbb{E}_u \left[ (X_t - u_t)^2 \right] dt$$

*for every value of* $a$ *than a positive constant* $A$*.*

In the following paragraphs, we review the construction of Section 4 and adapt it to the setting of absolutely continuous kernels.

#### 5.2.1. The covariance function

We calculate $\mathbb{E}_u[Z_aZ_b]$ for $a, b \in [0, T]$:

$$\begin{aligned}
\mathbb{E}_u[Z_aZ_b] &= \mathbb{E}_u\left[\int_0^T\int_0^T f(x_1,x_2;a)dW^u_{x_1}dW^u_{x_2}\int_0^T\int_0^T f(x_1,x_2;b)dW^u_{x_1}dW^u_{x_2}\right]\\
&= 2\int_0^T\int_0^T f(x_1,x_2;a)f(x_1,x_2;b)dx_1dx_2\\
&= 2\int_0^T\int_0^T\left(\int_0^a k(x_1,x_2;\alpha)d\alpha\right)\left(\int_0^b k(x_1,x_2;\beta)d\beta\right)dx_1dx_2\\
&= 2\int_0^b\int_0^a\left(\int_0^T\int_0^T k(x_1,x_2;\alpha)k(x_1,x_2;\beta)dx_1dx_2\right)d\alpha d\beta. \qquad (5.6)
\end{aligned}$$

This is the two-dimensional version of an absolutely continuous function. Since

$$\gamma : (\alpha,\beta) \mapsto 2\int_0^T\int_0^T k(x_1,x_2;\alpha)k(x_1,x_2;\beta)dx_1dx_2$$

is in $L^1([0,T]^2, \lambda_T^2)$, the covariance function

$$(a,b) \mapsto \mathbb{E}_u[Z_aZ_b]$$

has mixed second-order derivatives almost everywhere (see [27, Theorem 3.1, Remark 3.3]) and we have for almost every $(\alpha, \beta)$:

$$\begin{aligned}
\gamma(\alpha,\beta) &= 2\int_0^T\int_0^T k(x_1,x_2;\alpha)k(x_1,x_2;\beta)dx_1dx_2\\
&= \frac{\partial^2}{\partial\alpha\partial\beta}\,\mathbb{E}_u[Z_\alpha Z_\beta] = \frac{\partial^2}{\partial\beta\partial\alpha}\,\mathbb{E}_u[Z_\alpha Z_\beta]. \qquad (5.7)
\end{aligned}$$

We have moreover since $\gamma \in L^1([0,T]^2, \lambda_T^2)$:

$$\int_0^T \mathbb{E}_u[Z_t^2]dt = \left|\int_0^T\int_0^t\int_0^t \gamma(a,b)\,da\,db\,dt\right| \leqslant \int_0^T\int_0^T\int_0^T |\gamma(a,b)|\,da\,db\,dt < \infty.$$

#### 5.2.2. Hölder continuity

We notice that $(Z_t)_{t\in[0,T]}$ has a version that is $k$-Hölder continuous for every $k \in (0, 1/q_2)$. This result is proved in Theorem 9.3 using that $\gamma \in L^{q_1}\left([0,T]^2, \lambda_T^2\right)$.

#### 5.2.3. Stochastic integrals with respect to $(Z_t)_{t\in[0,T]}$

We follow a similar approach to the one of Section 4.1 and extend the definition of $\int_0^T \varphi(s)dZ_s$ from regulated functions to "sufficiently integrable functions" $\varphi$. For real numbers $q_1 > 1$ and $q_2 > 0$ with $1/q_1 + 1/q_2 = 1$, we suppose that:

(a) the functions $\varphi$ that are in $L^{q_2}([0,T],\lambda_T)$,
(b) the function $\gamma$ is in $L^{q_1}([0,T]^2,\lambda_T^2)$ (then the mixed second order derivatives of $(a,b)\mapsto \mathbb{E}_u[Z_aZ_b]$ exist and are equal to $\gamma$ almost everywhere).

It is known that the set of step functions is dense in $L^{q_2}([0,T],\lambda_T)$ (see for instance [25, Theorem 3.4 and Theorem 4.3]). For $\varphi\in L^{q_2}([0,T],\lambda_T)$, we choose a sequence $\varphi_n$ of step functions such that:

$$\|\varphi-\varphi_n\|_{L^{q_2}([0,T],\lambda_T)}\to 0,\qquad \text{for } n\to\infty.$$

We prove below, as $n\to\infty$:

$$\left\|\int_0^T \varphi_n dZ_y-\int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)dW^u_{x_1}dW^u_{x_2}\right\|_{L^2(\Omega,\mathbb{P}_u)}\to 0.$$

Then $\int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)^2dx_1dx_2<\infty$ and the double Wiener integral $\int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)dW^u_{x_1}dW^u_{x_2}$ exists and is square-integrable. Thus we can extend the definition of $\int_0^T\varphi(s)dZ_s$ to functions $\varphi\in L^{q_2}([0,T],\lambda_T)$:

$$\begin{aligned}\int_0^T\varphi(y)dZ_y &:= \lim_{n\to\infty}\int_0^T\varphi_n(y)dZ_y\\ &= \int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)dW^u_{x_1}dW^u_{x_2}.\end{aligned}\tag{5.8}$$

We have:

$$\begin{aligned}&\left\|\int_0^T \varphi_n dZ_y-\int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)dW^u_{x_1}dW^u_{x_2}\right\|^2_{L^2(\Omega,\mathbb{P}_u)}\\ &=2\int_0^T\int_0^T\left(\int_0^T(\varphi_n(a)-\varphi(a))k(x_1,x_2;a)da\right)^2dx_1dx_2\\ &=2\int_0^T\int_0^T\left(\int_0^T(\varphi_n(a)-\varphi(a))k(x_1,x_2;a)da\right)\\ &\quad\times\left(\int_0^T(\varphi_n(b)-\varphi(b))k(x_1,x_2;b)db\right)dx_1dx_2.\end{aligned}$$

Thus:

$$\begin{aligned}&\left\|\int_0^T \varphi_n dZ_y-\int_0^T\int_0^T\left(\int_0^T\varphi(s)k(x_1,x_2;s)ds\right)dW^u_{x_1}dW^u_{x_2}\right\|^2_{L^2([0,T],\lambda_T)}\\ &=\int_0^T\int_0^T(\varphi_n(a)-\varphi(a))(\varphi_n(b)-\varphi(b))\end{aligned}$$

$$
\begin{aligned}
&\times \left(\int_0^T \int_0^T 2k(x_1,x_2;a)k(x_1,x_2;b)dx_1dx_2\right) dadb \\
&\leqslant \left[\int_0^T \int_0^T |\varphi_n(a)-\varphi(a)|^{q_2}|\varphi_n(b)-\varphi(b)|^{q_2} dadb\right]^{1/q_2} \\
&\quad \times \left(\int_0^T \int_0^T |\gamma(a,b)|^{q_1} dadb\right)^{1/q_1} \\
&= \|\varphi_n - \varphi\|^2_{L^{q_2}([0,T],\lambda_T)} \underbrace{\left(\int_0^T \int_0^T |\gamma(a,b)|^{q_1} dadb\right)^{1/q_1}}_{<\infty} \to 0, \qquad \text{for } n \to \infty.
\end{aligned}
$$

The calculations above show that:

$$
\begin{aligned}
&2\int_0^T \int_0^T \left(\int_0^T \varphi(s)k(x_1,x_2;s)ds\right)^2 dx_1dx_2 \\
&= \int_0^T \int_0^T \varphi(a)\varphi(b)\gamma(a,b)dadb < \infty.
\end{aligned} \tag{5.9}
$$

*5.2.4. Observations and functions $g_i$.*

We define:

$$
H_q := \left\{ f : [0,T] \to \mathbb{R} : f(t) = \int_0^t \dot{f}(x)dx \; ; \; \dot{f} \in L^q\left([0,T],\lambda_T\right) \right\}.
$$

As above, we consider a deterministic drift $u \in H_2$. We define $X(h)$. We have $\mathbb{E}_u[Z(h)^2] < \infty$ for $h \in H_{q_2}$ as exposed in the previous section. We have $u(h) := \int_0^T \dot{u}_s \dot{h}(s)ds < \infty$ if $h \in H_2$. This shows the necessity to choose $h \in H_{q_2} \cap H_2 = H_{\max\{2,q_2\}}$. For $h \in H_{\max\{2,q_2\}}$, we use the following notations in agreement with (5.8):

$$
\begin{aligned}
X(h) &= u(h) + Z(h) \\
&= \int_0^T \dot{u}_s \dot{h}(s)ds + \int_0^T \int_0^T \left(\int_0^T \dot{h}(s)k(x_1,x_2;s)ds\right) dW^u_{x_1} dW^u_{x_2}.
\end{aligned}
$$

We notice that the right side of the last equation can be approximated in $L^2(\Omega,\mathbb{P}_u)$ by $X(h_n)$ where $h_n \in H_{\max\{2,q_2\}}$ such that $\dot{h}_n$ can be chosen as a sequence of step functions that converge to $\dot{h}$ in $L^{\max\{2,q_2\}}([0,T],\lambda_T)$. For the functions $g_i$ we have:

$$
\begin{aligned}
g_i(t) &= \mathbb{E}_u[Z_t\, Z(h_i)] = \operatorname{Cov}_u(X_t, X(h_i)) \\
&= 2\int_0^T \int_0^T \left(\int_0^t k(x_1,x_2;v)dv\right)\left(\int_0^T \dot{h}_i(v)k(x_1,x_2;v)dv\right) dx_1dx_2.
\end{aligned}
$$

We have for $0 \leqslant s \leqslant t \leqslant T$, $\gamma \in L^{q_1}([0,T]^2, \lambda_T^2)$, $h \in H_{\max\{2,q_2\}}$, $q_1 > 1$ and $1/q_1 + 1/q_2 = 1$:

$$\begin{aligned}
|g_i(t) - g_i(s)|^2 &= \mathrm{Cov}_u(X(h_i), X_t - X_s)^2 \\
&= 4\left[\int_0^T\int_0^T \left(\int_0^T 1_{(s,t]}(a) k(x_1,x_2;a)da\right)\right. \\
&\quad \left.\times \left(\int_0^T \dot{h}_i(a) k(x_1,x_2;a)da\right) dx_1 dx_2\right]^2 \\
&\leqslant 4\int_0^T\int_0^T \left(\int_0^T 1_{(s,t]}(a) k(x_1,x_2;a)da\right)^2 dx_1 dx_2 \\
&\quad \times \int_0^T\int_0^T \left(\int_0^T \dot{h}_i(a) k(x_1,x_2;a)da\right)^2 dx_1 dx_2 \\
&\leqslant \int_s^t\int_s^t |\gamma(x_1,x_2)| dx_1 dx_2 \ \mathbb{E}_u[Z(h_i)^2].
\end{aligned}$$

We have used Eq. (5.9) and find:

$$\begin{aligned}
|g_i(t) - g_i(s)| &\leqslant \left(\int_s^t\int_s^t |\gamma(x_1,x_2)| dx_1 dx_2\right)^{1/2} \sqrt{\mathbb{E}_u[Z(h_i)^2]} \\
&= \left(\int_0^T\int_0^T |1_{(s,t]}(x_1) 1_{(s,t]}(x_2) \gamma(x_1,x_2)| dx_1 dx_2\right)^{1/2} \sqrt{\mathbb{E}_u[Z(h_i)^2]} \\
&\leqslant \left(\int_0^T\int_0^T |\gamma(x_1,x_2)|^{q_1} dx_1 dx_2\right)^{1/(2q_1)} \\
&\quad \times \left(\int_0^T\int_0^T 1_{(s,t]}(x_1) 1_{(s,t]}(x_2) dx_1 dx_2\right)^{1/(2q_2)} \sqrt{\mathbb{E}_u[Z(h_i)^2]} \\
&= \left(\int_0^T\int_0^T |\gamma(x_1,x_2)|^{q_1} dx_1 dx_2\right)^{1/(2q_1)} |t-s|^{1/q_2} \sqrt{\mathbb{E}_u[Z(h_i)^2]}.
\end{aligned}$$

The functions $g_i$ are thus bounded on $[0,T]$, continuous on $(0,T)$, right-continuous in 0 and left-continuous in $T$.

### 5.2.5. *Conclusion*

Proposition 5.3 holds under the assumptions (i), (ii), (iii), (iv'), (v') and Eq. (4.8) gives an estimator that dominates the standard estimator if $d \geqslant 3$.

## 6. Applications

In this section we apply the results of Section 4 and Section 5 to concrete situations. We suppose that conditions (i)-(vii) hold, unless other conditions are specified.

In Paragraph 6.1 we consider a noise that has a finite Wiener-Itô chaos decomposition and that defines a martingale. In Paragraph 6.2 we consider the Gaussian case and show moreover that for this particular situation, we can choose $a = 0$ in Eq. (4.8). In Paragraph 5.2 we consider the case of an absolutely continuous kernel and adapt the construction of Section 4 to this situation. In Paragraph 6.3 we consider in particular the case of the Rosenblatt process. In all these cases, Eq. (4.8) gives an estimator whose risk is smaller risk than the risk of the standard drift estimator.

### *6.1. The martingale case*

In the martingale case, the Cramer-Rao bounds of Section 2 hold and our estimator is *superefficient*.

(1) Theorem 4.1 and Proposition 5.1 can be used to handle the case where the noise $(Z_t)_{t\in[0,T]}$ defines a martingale with respect to the filtration of the Brownian motion with $Z_0 = 0$ and

$$Z_t = X_t - u_t = \int_0^t f_1(x_1)dW^u_{x_1} + \int_0^t \int_0^{x_2} f_2(x_1, x_2)dW^u_{x_1}dW^u_{x_2}, \quad t \in (0, T],$$

for square-integrable function functions $f_1$ and $f_2$. The function $f_2$ is symmetric. We define:

$$\begin{aligned} f_1(x_1; t) &:= f_1(x_1)\, 1_{[0\leqslant x_1\leqslant t]}, \\ f_2(x_1, x_2; t) &:= \frac{1}{2} f_2(x_1, x_2)\, 1_{[0\leqslant x_1\leqslant t]}\, 1_{[0\leqslant x_2\leqslant t]} \\ &= \frac{f_2(x_1, x_2)}{2}\, 1_{[0\leqslant \min\{x_1,x_2\}\leqslant \max\{x_1,x_2\}\leqslant t]}. \end{aligned}$$

The functions $t \mapsto f_1(x_1; t)$ and $t \mapsto f_2(x_1, x_2; t)$ are monotone and right-continuous. Furthermore $f_2(x_1, x_2; t)$ is symmetric in the first two variables. Consider $h : t \mapsto \int_0^t \dot{h}_i(s)ds$ for a regulated function $\dot{h}_i$. With a result from [11], we have:

$$\begin{aligned} \int_0^T \dot{h}_i(t)df_1(x_1; t) &= \dot{h}_i(x_1)f_1(x_1), \\ \int_0^T \dot{h}_i(t)df_2(x_1, x_2; t) &= \dot{h}_i(\max\{x_1, x_2\})\frac{f_2(x_1, x_2)}{2}. \end{aligned}$$

Since $\dot{h}_i$ is bounded as a regulated function, the Lebesgue-Stieltjes integrals above are square-integrable. We have:

$$\begin{aligned}
X(h_i) &= \int_0^T \dot{u}_s\, \dot{h}_i(s)ds + \int_0^T \left( \int_0^T \dot{h}_i(t) df_1(x_1;t) \right) dW^u_{x_1} \\
&\quad + \int_0^T \int_0^T \left( \int_0^T \dot{h}_i(t) df_2(x_1,x_2;t) \right) dW^u_{x_1} dW^u_{x_2} \\
&= \int_0^T \dot{u}_s\, \dot{h}_i(s)ds + \int_0^T \dot{h}_i(x_1) f_1(x_1) dW^u_{x_1} \\
&\quad + \int_0^T \int_0^T \dot{h}_i(\max\{x_1,x_2\}) \frac{f_2(x_1,x_2)}{2} dW^u_{x_1} dW^u_{x_2} \\
&= \int_0^T \dot{u}_s\, \dot{h}_i(s)ds + \int_0^T \dot{h}_i(x_1)\, f_1(x_1) dW^u_{x_1} \\
&\quad + 2\,\frac{1}{2} \int_0^T \int_0^{x_2} \dot{h}_i(x_2) f_2(x_1,x_2) dW^u_{x_1} dW^u_{x_2} \\
&= \int_0^T \dot{u}_s\, \dot{h}_i(s)ds + \int_0^T \dot{h}_i(x_1)\, f_1(x_1) dW^u_{x_1} \\
&\quad + \int_0^T \dot{h}_i(x_2) \int_0^{x_2} f_2(x_1,x_2) dW^u_{x_1} dW^u_{x_2},
\end{aligned}$$

and in particular for $h_i : s \mapsto \min\{s, t_i\}$, $\dot{h}_i = 1_{(0,t_i]}$:

$$X_{t_i} = \int_0^{t_i} \dot{u}_s ds + \int_0^{t_i} f_1(x_1) dW^u_{x_1} + \int_0^{t_i} \int_0^{x_2} f_2(x_1,x_2) dW^u_{x_1} dW^u_{x_2}.$$

The analogue of assumption (v) holds, since:

$$\begin{aligned}
\int_0^T V_0^T (f_1(x_1;\cdot))^2 dx_1 &= \int_0^T f_1(x_1)^2 dx_1 < \infty, \\
\int_0^T \int_0^T V_0^T (f_2(x_1,x_2;\cdot))^2 dx_1 dx_2 &= \frac{1}{4} \int_0^T \int_0^T f_2(x_1,x_2)^2 dx_1 dx_2 < \infty.
\end{aligned}$$

It is easy to check that $\mathbb{E}_u[(Z_t - Z_s)^2] \to 0$ for $|t-s| \to 0$. We have for $s \leqslant t$:

$$\begin{aligned}
\mathbb{E}_u[(Z_t - Z_s)^2] &= \int_s^t f_1(x_1)^2 dx_1 + 2 \int_s^t \int_0^{x_2} f_2(x_1,x_2)^2 dx_1 dx_2 \\
&\to 0, \text{ for } |t-s| \to 0,
\end{aligned}$$

since $f_1$ and $f_2$ are square-integrable. We find similarly that $\int_0^T \mathbb{E}_u[Z_t]^2 dt < \infty$. We conclude that Theorem 4.1 can be used in this situation to find estimators with smaller risk than the standard drift estimator $X_t$.

We notice that $(Z_t)_{t\in[0,T]}$ has a version that is locally $k$-Hölder continuous for every $k \in (0, 1/(2q_2))$ if the function $x_1 \mapsto f_1(x_1)^2 + 2\int_0^{x_1} f_2(x_1, x_2)^2 dx_2$ is in $L^{q_1}([0,T], d\lambda_T)$ for positive reels $q_1, q_2$ with $1/q_1 + 1/q_2 = 1$.

(2) As a direct consequence, we can choose $f_2 = 0$ and recover the case of Gaussian processes of the form

$$X_0 := 0 \ ; \ X_t = u_t + \int_0^t f_1(x) dW_x^u, \quad t \in (0,T],$$

for a square-integrable function $f_1$.

### *6.2. The Gaussian case*

Choosing $f_2 = 0$ in Proposition 5.1, we have

$$X_0 := 0 \ ; \ X_t = u_t + \int_0^T f_1(x_1;t) dW_{x_1}^u, \quad t \in (0,T]. \tag{6.1}$$

We prove for this class of Gaussian processes that we can choose $a = 0$. As in Eq. (4.8), we define $\xi_t := (g(t)^\top BX(h))/(a + X(h)^\top BX(h))$ and $g_i(t) = \mathrm{Cov}_u(X(h_i), X_t)$. We have for the model given in Eq. (6.1) and $d \geqslant 3$:

$$\begin{aligned}
2\int_0^T \mathbb{E}_u[(X_t - u_t)\xi_t]dt &= 2\int_0^T \int_0^T \mathbb{E}_u\left[f_1(x_1;t)\,\frac{g(t)^\top BD_{x_1}X(h)}{a + X(h)^\top BX(h)}\right]dx_1 dt \\
&\quad - 4\int_0^T\int_0^T \mathbb{E}_u\left[f_1(x_1;t)\,\frac{g(t)^\top BX(h)\,X(h)^\top BD_{x_1}X(h)}{(a + X(h)^\top BX(h))^2}\right]dx_1 dt \\
&= 2\,\mathbb{E}_u\left[\frac{\int_0^T g(t)^\top Bg(t)dt}{a + X(h)^\top BX(h)}\right] - 4\int_0^T \mathbb{E}_u\left[\frac{g(t)^\top BX(h)\,X(h)^\top Bg(t)}{(a + X(h)^\top BX(h))^2}\right]dt \\
&= 2d\,\mathbb{E}_u\left[\frac{1}{a + X(h)^\top BX(h)}\right] - 4\,\mathbb{E}_u\left[\frac{X(h)^\top B\int_0^T g(t)g(t)^\top dtBX(h)}{(a + X(h)^\top BX(h))^2}\right] \\
&= 2d\,\mathbb{E}_u\left[\frac{1}{a + X(h)^\top BX(h)}\right] - 4\,\mathbb{E}_u\left[\frac{X(h)^\top BX(h)}{(a + X(h)^\top BX(h))^2}\right] \\
&\geqslant 2d\,\mathbb{E}_u\left[\frac{1}{a + X(h)^\top BX(h)}\right] - 4\,\mathbb{E}_u\left[\frac{1}{a + X(h)^\top BX(h)}\right] \\
&= \mathbb{E}_u\left[\frac{2d - 4}{a + X(h)^\top BX(h)}\right],
\end{aligned}$$

thus

$$2\int_0^T \mathbb{E}_u[(X_t - u_t)\xi_t]dt - \int_0^T \mathbb{E}_u[\xi_t^2]dt \geqslant \mathbb{E}_u\left[\frac{2d - 5}{a + X(h)^\top BX(h)}\right].$$

The right-hand expectation is positive for every $d \geqslant 3$ and $a > 0$. We can choose $a = 0$ for this special case if $d \geqslant 3$ and $\mathrm{Cov}_u(X(h))$ is invertible. This follows from the well known fact that

$$\int_{\mathbb{R}^d} \frac{1}{n_1^2 + \ldots + n_d^2} \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{n_1^2 + \ldots + n_d^2}{2}\right) dn_1 \ldots dn_d$$

is finite if and only if $d \geqslant 3$.

### 6.3. The Rosenblatt process

An important process $(Z_t)_{t\in[0,T]}$ verifying the conditions of Section 4 respectively of Section 5.2 is the Rosenblatt process on a compact interval $[0, T]$.

(1) Tudor [26] gives the following representation for the Rosenblatt process on $[0, T]$:

$$Z_t = d(H) \int_0^T \int_0^T \left( \int_0^t 1_{[v>y_1 \vee y_2]} \frac{\partial K^{H'}}{\partial v}(v, y_1) \frac{\partial K^{H'}}{\partial v}(v, y_2) dv \right) dW^u_{y_1} dW^u_{y_2},$$

$$K^H(t, s) := \left( \frac{H(2H-1)}{\beta(2-2H, H-\frac{1}{2})} \right)^{\frac{1}{2}} s^{\frac{1}{2}-H} \int_s^t (v-s)^{H-\frac{3}{2}} v^{H-\frac{1}{2}} dv \text{ for } t > s,$$

$$\frac{1}{2} < H < 1, \quad H' = \frac{H+1}{2}, \quad d(H) = \frac{1}{H+1} \left( \frac{H}{2(2H-1)} \right)^{-1/2}.$$

The covariance function is given by

$$(t, s) \mapsto \mathrm{Cov}_u(Z_t, Z_s) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right),$$

and we have almost everywhere the mixed second order derivative:

$$(s, t) \mapsto H(2H-1)|t-s|^{2H-2}.$$

This function is in $L^{q_1}([0,T]^2, \lambda_T^2)$ for every $1 < q_1 < 1/(2-2H)$. The function $t \mapsto \mathbb{E}_u[Z_t^2] = t^{2H}$ is an integrable function over $[0, T]$. The conditions of Section 5.2 are easy to check (see also [26]). It is thus possible to apply Theorem 4.1 to construct an estimator with smaller risk than the standard estimator in the case of a Rosenblatt process with Hurst parameter $1/2 < H < 1$.

(2) If $(Z_t)_{t\in[0,T]}$ is a Rosenblatt process with $1/2 < H < 1$, the conditions about $\gamma$ can be specified. For this particular process, we have with Eq. (5.7) that $\gamma(s, t) = H(2H-1)|t-s|^{2H-2}$ for almost every $(s, t)$. It is easy to see that $\gamma \in L^1([0,T]^2, \lambda_T^2)$ and this is enough for the results above to hold in the case of the Rosenblatt process. Consider $\varphi \in L^2([0,T], \lambda_T) \subset L^{1/H}([0,T], \lambda_T)$. We have with formula (5.9) and [19, p.19, (4.8)]:

$$\int_0^T \int_0^T \left( \int_0^T \varphi(s) k(x_1, x_2; s) ds \right)^2 dx_1 dx_2 = \frac{1}{2} \int_0^T \int_0^T \varphi(a)\varphi(b)\gamma(a, b) da db$$

$$\leqslant \frac{1}{2}\int_0^T\int_0^T |\varphi(a)\varphi(b)\gamma(a,b)|dadb$$
$$\leqslant \|\varphi\|^2_{L^{1/H}([0,T],\lambda_T)}\, c(H).$$

The constant $c(H)$ depends only on $H$. This relation is sufficient to prove for instance that $g_i$ are continuous if $\dot{h}_i \in L^2([0,T],\lambda_T)$. Consider $0 \leqslant s \leqslant t \leqslant T$:

$$\begin{aligned}
&|g_i(t) - g_i(s)| = |\,\mathbb{E}_u[Z(h_i)(Z_t - Z_s)]| \\
&= 2\left|\int_0^T\int_0^T \left(\int_0^T \dot{h}_i(s)k(x_1,x_2;s)da\right)\left(\int_0^T 1_{(s,t]}(a)k(x_1,x_2;a)da\right)dx_1dx_2\right| \\
&\leqslant 2\left[\int_0^T\int_0^T \left(\int_0^T \dot{h}_i(a)k(x_1,x_2;a)da\right)^2 dx_1dx_2\right. \\
&\quad \left.\times \int_0^T\int_0^T \left(\int_0^T 1_{(s,t]}(a)k(x_1,x_2;s)da\right)^2 dx_1dx_2\right]^{1/2} \\
&= 2\left(\int_0^T\int_0^T \left(\int_0^T \dot{h}_i(a)k(x_1,x_2;a)da\right)^2 dx_1dx_2\right)^{1/2} \\
&\quad \times \|1_{(s,t]}\|_{L^{1/H}([0,T],\lambda_T)}\sqrt{c(H)} \\
&= 2\left(\int_0^T\int_0^T \left(\int_0^T \dot{h}_i(a)k(x_1,x_2;a)da\right)^2 dx_1dx_2\right)^{1/2} |t-s|^H\sqrt{c(H)}.
\end{aligned}$$

We can also make a more specific statement about Hölder continuity. We can adapt the proof of Theorem 9.3 for $0 \leqslant s \leqslant t \leqslant T$:

$$\begin{aligned}
\mathbb{E}_u[|Z_t - Z_s|^{2n}] &\leqslant \tau(n)\left[\int_s^t\int_s^t \gamma(a,b)dadb\right]^n \\
&\leqslant \tau(n)2^n\left[\frac{1}{2}\int_0^T\int_0^T 1_{(s,t]}(a)1_{(s,t]}(b)\gamma(a,b)dadb\right]^n \\
&\leqslant 2^n\tau(n)c(H)^n\left(\|1_{(s,t]}\|^2_{L^{1/H}([0,T],\lambda_T)}\right)^n \\
&\leqslant 2^n\tau(n)c(H)^n|s-t|^{2Hn}.
\end{aligned}$$

We find as in Theorem 9.3, that the Rosenblatt process has a $k$-Hölder continuous version for every $k \in (0,H)$. This result is also stated in [14].

## 7. Estimation of the constant $A$

In this section we discuss the problem of finding an optimal value for the constant $A$ used to define the estimator defined in Eq. (4.8):

$$X_t - \xi_t = X_t - \frac{g(t)^\top BX(h)}{a + X(h)^\top BX(h)}, \quad t \in [0,T].$$

For the case $d \geqslant 3$, Theorem 4.1 states that the risk of $X_t - \xi_t$ is smaller than the risk of the standard estimator for every $a$ that is large enough, more precisely $a > A > 0$, where $A$ is a constant depending on $h_1, \ldots, h_d$, $f$ and $T$. In this section we discuss the problem of estimating $A$. In Proposition 7.3 and Proposition 7.4, we give estimates for two special settings, one being related to the Rosenblatt process, the other related to the martingale case.

### *7.1. Estimating A in the Rosenblatt case*

We estimate $A$ in the case of a Rosenblatt process $(Z_t)_{t\in[0,T]}$.

*Remark* 7.1. For the proof of the main result, it is essential that $d \geqslant 3$ and

$$\begin{aligned}(2d-5)\,\mathbb{E}_0 &\left[\left(1+\sqrt{\frac{X(h)^\top BX(h)}{a}}\right)^{-2}\right] - \frac{56}{\sqrt{a}}\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)| \\ &\times \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|\right. \\ &\times \left.\left(\frac{1}{k^2} + \frac{X(h)^\top BX(h)}{a(1-k)^2}\right)\right] dx_1dx_2dt > 0. \end{aligned} \tag{7.1}$$

We have already proved that this inequality holds whenever $a$ is large enough, more precisely $a > A > 0$. Finding the *smallest* possible value of $A$ is a non trivial problem. We show how a value for $A$ can be found using numerical calculations. *Notice that this approach does not provide the smallest possible value for A.*

We consider $T > 0$, $d \geqslant 3$ and $k \in (0,1)$. We use the notations of Theorem 4.1, for instance $h(t) = (h_1(t), \ldots, h_d(t))^\top$. We notice that $X_t = u_t + Z_t$, thus $(X_t)_{\in[0,T]} = (Z_t)_{t\in[0,T]}$ if $u = 0$. We estimate the terms of inequality (7.1).

(a) We give a lower bound for

$$\mathbb{E}_0\left[\left(1+\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^{-2}\right].$$

We notice that $Z(h)^\top BZ(h) \geqslant 0$ and that $\varphi : [0,+\infty) \to (0,1]$, $x \mapsto (1+x)^{-2}$ is a convex function. Jensen's inequality yields for every $a > 0$:

$$(2d-5)\;\mathbb{E}_0\left[\left(1+\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^{-2}\right]$$

$$
\begin{aligned}
&= (2d-5)\ \mathbb{E}_0\left[\varphi\left(\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)\right] \\
&\geqslant (2d-5)\,\varphi\left(\mathbb{E}_0\left[\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right]\right) \\
&= (2d-5)\ \left(1+\mathbb{E}_0\left[\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right]\right)^{-2} \\
&\geqslant (2d-5)\ \left(1+\frac{\sqrt{\mathbb{E}_0[Z(h)^\top BZ(h)]}}{\sqrt{a}}\right)^{-2}.
\end{aligned}
$$

In the last step we have used that $\mathbb{E}_0[\sqrt{Z(h)^\top BZ(h)}] \leqslant \sqrt{\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]}$.

(b) We now estimate:

$$
\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|\right]dx_1dx_2dt.
$$

We have with $C^\top C = C^2 = B$ and $\int_0^T g(t)^\top Bg(t)dt = d$ (see proof of Theorem 4.1):

$$
\begin{aligned}
&\left|\int_0^T\int_0^T\int_0^{x_2}|f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}Z(h)\|\,\|CD_{x_2}Z(h)\|\right]dx_1dx_2dt\right|^2 \\
&\leqslant \frac{1}{4}\int_0^T\int_0^T\int_0^T f(x_1,x_2;t)^2dx_1dx_2dt\int_0^T\|Cg(t)\|^2dt \\
&\quad\times\int_0^T\int_0^T\mathbb{E}_0\left[\|CD_{x_1}Z(h)\|\,\|CD_{x_2}Z(h)\|\right]^2dx_1dx_2 \\
&\leqslant \frac{1}{4}\int_0^T\frac{1}{2}\ \mathbb{E}_0[Z_t^2]dt\int_0^T g(t)^\top Bg(t)dt\left(\int_0^T\mathbb{E}_0\left[\|CD_{x_1}Z(h)\|^2\right]dx_1\right)^2 \\
&= \frac{d}{8}\int_0^T\mathbb{E}_0[Z_t^2]dt\left(\int_0^T\mathbb{E}_0\left[(D_{x_1}Z(h))^\top B(D_{x_1}Z(h))\right]dx_1\right)^2.
\end{aligned}
$$

We calculate $\int_0^T\mathbb{E}_0\left[(D_{x_1}Z(h))^\top B(D_{x_1}Z(h))\right]dx_1$ with the integration by parts formula:

$$
\begin{aligned}
&\int_0^T\mathbb{E}_0\left[(D_{x_1}Z(h))^\top B(D_{x_1}Z(h))\right]dx_1 \\
&= \sum_{i,j=1}^d B_{i,j}\int_0^T\mathbb{E}_0\left[D_{x_1}Z(h_i)D_{x_1}Z(h_j)\right]dx_1 \\
&= \sum_{i,j=1}^d B_{i,j}\,\mathbb{E}_0\left[\langle DZ(h_i),DZ(h_j)\rangle_{L^2([0,T],\lambda_T)}\right]
\end{aligned}
$$

$$= \sum_{i,j=1}^{d} B_{i,j}\,\mathbb{E}_0\left[Z(h_i)\,\delta D Z(h_j)\right]$$

$$= \sum_{i,j=1}^{d} B_{i,j}\,\mathbb{E}_0\left[Z(h_i)\,2Z(h_j)\right] = 2\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right].$$

We get:

$$\left|\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}X(h)\|\,\|CD_{x_2}X(h)\|\right]dx_1dx_2dt\right|$$

$$\leqslant \left(\frac{d}{8}\int_0^T \mathbb{E}_u[Z_t^2]dt\,4\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]^2\right)^{1/2}$$

$$\leqslant \left(\frac{d}{2}\int_0^T \mathbb{E}_u[Z_t^2]dt\right)^{1/2}\mathbb{E}_0\left[Z(h)^\top BZ(h)\right].$$

(c) We estimate now:

$$\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}X(h)\|\right.$$
$$\left.\times\ \|CD_{x_2}X(h)\|\,X(h)^\top BX(h)\right]dx_1dx_2dt.$$

We have:

$$\left(\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}Z(h)\|\right.\right.$$
$$\left.\left.\times\ \|CD_{x_2}Z(h)\|\,Z(h)^\top BZ(h)\right]dx_1dx_2dt\right)^2$$
$$\leqslant \frac{1}{4}\int_0^T\int_0^T\int_0^T f(x_1,x_2,t)^2\,\mathbb{E}_0\left[(Z(h)^\top BZ(h))^2\right]dx_1dx_2dt$$
$$\times\int_0^T\int_0^T\int_0^T \|Cg(t)\|^2\,\mathbb{E}_0\left[\|CD_{x_1}Z(h)\|^2\|CD_{x_2}Z(h)\|^2\right]dx_1dx_2dt$$
$$\leqslant \frac{1}{4}\int_0^T \frac{1}{2}\mathbb{E}_0[Z_t^2]dt\ \mathbb{E}_0\left[(Z(h)^\top BZ(h))^2\right]\int_0^T g(t)^\top Bg(t)dt$$
$$\times\int_0^T\int_0^T \mathbb{E}_0\left[\left((D_{x_1}Z(h))^\top BD_{x_1}Z(h)\right)\ \left((D_{x_2}Z(h))^\top BD_{x_2}Z(h)\right)\right]dx_1dx_2$$
$$\leqslant \frac{d}{8}\int_0^T \mathbb{E}_0[Z_t^2]dt\,\mathbb{E}_0\left[(Z(h)^\top BZ(h))^2\right]$$
$$\times\int_0^T\int_0^T \mathbb{E}_0\left[\left((D_{x_1}Z(h))^\top BD_{x_1}Z(h)\right)\left((D_{x_2}Z(h))^\top BD_{x_2}Z(h)\right)\right]dx_1dx_2.$$

We define $\tilde{h}(t) := Ch(t)$, then $CZ(h) = Z(\tilde{h})$. To simplify notations we write $Z(\tilde{h}_i) = I_2(\tilde{q}_i)$. We have:

$$\mathbb{E}_0\left[(Z(h)^\top BZ(h))^2\right] = \mathbb{E}_0\left[\|CZ(h)\|^4\right] = \mathbb{E}_0\left[\|Z(\tilde{h})\|^4\right]$$

$$
\begin{aligned}
&= \mathbb{E}_0\left[\left(Z(\tilde{h}_1)^2 + \ldots + Z(\tilde{h}_d)^2\right)^2\right] \\
&= \mathbb{E}_0\left[\left(I_2(\tilde{q}_1)^2 + \ldots + I_2(\tilde{q}_d)^2\right)^2\right].
\end{aligned}
$$

We prove:

$$
\mathbb{E}_0\left[\left(Z(h)^\top B Z(h)\right)^2\right] \leqslant 15\, \mathbb{E}_0\left[Z(h)^\top B Z(h)\right]^2.
$$

We write $\tilde{\otimes}.$ for the symmetrization of the contraction of two functions and find:

$$
\begin{aligned}
&\mathbb{E}_0[Z(\tilde{h}_i)^2 Z(\tilde{h}_j)^2] \\
&= \mathbb{E}_0\left[\left(\sum_{r=0}^{2} r! \binom{2}{r}^2 I_{4-2r}\left(\tilde{q}_i \tilde{\otimes}_r \tilde{q}_i\right)\right)\left(\sum_{r'=0}^{2} r'! \binom{2}{r'}^2 I_{4-2r'}\left(\tilde{q}_j \tilde{\otimes}_{r'} \tilde{q}_j\right)\right)\right] \\
&= \sum_{r=0}^{2} r!^2 \binom{2}{r}^4 \mathbb{E}_0\left[I_{4-2r}\left(\tilde{q}_i \tilde{\otimes}_r \tilde{q}_i\right) I_{4-2r}\left(\tilde{q}_j \tilde{\otimes}_r \tilde{q}_j\right)\right] \\
&\leqslant \sum_{r=0}^{2} r!^2 \binom{2}{r}^4 \left(\mathbb{E}_0\left[I_{4-2r}\left(\tilde{q}_i \tilde{\otimes}_r \tilde{q}_i\right)^2\right] \mathbb{E}_0\left[I_{4-2r}\left(\tilde{q}_j \tilde{\otimes}_r \tilde{q}_j\right)^2\right]\right)^{1/2}.
\end{aligned}
$$

We use the inequality

$$
\|\tilde{q}_i \tilde{\otimes}_r \tilde{q}_i\|^2_{L^2([0,T]^{4-2r}, \lambda_T^{4-2r})} \leqslant \|\tilde{q}_i \otimes_r \tilde{q}_i\|^2_{L^2([0,T]^{4-2r}, \lambda_T^{4-2r})} \leqslant \|\tilde{q}_i\|^4_{L^2([0,T]^2, \lambda_T^2)},
$$

and find:

$$
\begin{aligned}
&\mathbb{E}_0[Z(\tilde{h}_i)^2 Z(\tilde{h}_i)^2] \\
&\leqslant \sum_{r=0}^{2} r!^2 \binom{2}{r}^4 \left[(4-2r)!^2 \|\tilde{q}_i\|^4_{L^2([0,T], \lambda_T)} \|\tilde{q}_j\|^4_{L^2([0,T], \lambda_T)}\right]^{1/2} \\
&= \sum_{r=0}^{2} r!^2 \binom{2}{r}^4 (4-2r)! \frac{1}{4}\, \mathbb{E}_0[I_2(\tilde{q}_i)^2]\, \mathbb{E}_0[I_2(\tilde{q}_j)^2] \\
&= \sum_{r=0}^{2} r!^2 \binom{2}{r}^4 (4-2r)! \frac{1}{4}\, \mathbb{E}_0[Z(\tilde{h}_i)^2]\, \mathbb{E}_0[Z(\tilde{h}_j)^2] \\
&= 15\, \mathbb{E}_0[Z(\tilde{h}_i)^2]\, \mathbb{E}_0[Z(\tilde{h}_j)^2].
\end{aligned}
$$

We conclude:

$$
\begin{aligned}
\mathbb{E}_0\left[\left(Z(h)^\top B Z(h)\right)^2\right] &= \sum_{i,j=1}^{d} \mathbb{E}_0\left[Z(\tilde{h}_i)^2 Z(\tilde{h}_j)^2\right] \\
&\leqslant 15 \sum_{i,j=1}^{d} \mathbb{E}_0[Z(\tilde{h}_i)^2]\, \mathbb{E}_0[Z(\tilde{h}_j)^2]
\end{aligned}
$$

$$= 15\left(\mathbb{E}_0[Z(\tilde{h}_1)^2] + \ldots + \mathbb{E}_0[Z(\tilde{h}_j)^2]\right)^2$$
$$= 15\left(\mathbb{E}_0\left[\|Z(\tilde{h})\|^2\right]\right)^2 = 15\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]^2.$$

We can prove similarly:

$$0 \leqslant \int_0^T\int_0^T \mathbb{E}_0\left[\left((D_{x_1}Z(h))^\top BD_{x_1}Z(h)\right)\left((D_{x_2}Z(h))^\top BD_{x_2}Z(h)\right)\right]dx_1dx_2$$
$$\leqslant 12\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]^2,$$

using the inequality

$$\mathbb{E}_0[I_1(l_1)^2\, I_1(l_2)^2] \leqslant 3\,\mathbb{E}_0[I_1(l_1)^2]\ \mathbb{E}_0[I_1(l_2)^2],$$

for any square-integrable functions $l_1, l_2$ and random variables $I_1(l_1), I_1(l_2)$ in the first Wiener-Itô chaos. This inequality yields:

$$\mathbb{E}_0\left[D_{x_1}Z(\tilde{h}_i)^2 D_{x_2}Z(\tilde{h}_j)^2\right] \leqslant 3\,\mathbb{E}_0\left[D_{x_1}Z(\tilde{h}_i)^2\right]\,\mathbb{E}_0\left[D_{x_2}Z(\tilde{h}_j)^2\right]. \tag{7.2}$$

We have:

$$\int_0^T\int_0^T \mathbb{E}_0\left[\left((D_{x_1}Z(h))^\top BD_{x_1}Z(h)\right)\ \left((D_{x_2}Z(h))^\top BD_{x_2}Z(h)\right)\right]dx_1dx_2$$
$$= \int_0^T\int_0^T \mathbb{E}_0\left[\|C(D_{x_1}Z(h))\|^2\,\|C(D_{x_2}Z(h))\|^2\right]dx_1dx_2$$
$$= \int_0^T\int_0^T \mathbb{E}_0\left[\|D_{x_1}Z(\tilde{h})\|^2\|D_{x_2}Z(\tilde{h})\|^2\right]dx_1dx_2$$
$$= \sum_{i,j=1}^d \int_0^T\int_0^T \mathbb{E}_0\left[\left(D_{x_1}Z(\tilde{h}_i)\right)^2\left(D_{x_2}Z(\tilde{h}_j)\right)^2\right]dx_1dx_2.$$

With inequality (7.2), we find:

$$\int_0^T\int_0^T \mathbb{E}_0\left[\left((D_{x_1}Z(h))^\top BD_{x_1}Z(h)\right)\ \left((D_{x_2}Z(h))^\top BD_{x_2}Z(h)\right)\right]dx_1dx_2$$
$$\leqslant 3\sum_{i,j=1}^d \int_0^T\int_0^T \mathbb{E}_0\left[\left(D_{x_1}Z(\tilde{h}_i)\right)^2\right]\mathbb{E}_0\left[\left(D_{x_2}Z(\tilde{h}_j)\right)^2\right]dx_1dx_2$$
$$= 3\int_0^T \mathbb{E}_0\left[\|D_{x_1}Z(\tilde{h})\|^2\right]dx_1\int_0^T \mathbb{E}_0\left[\|D_{x_2}Z(\tilde{h})\|^2\right]dx_2$$
$$= 3\left(\int_0^T \mathbb{E}_0\left[\|CD_{x_1}Z(h)\|^2\right]dx_1\right)^2$$
$$= 3\left(\int_0^T \mathbb{E}_0\left[(D_{x_1}Z(h)^\top B(D_{x_1}Z(h)\right]dx_1\right)^2$$

$$= 3\left(2\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]\right)^2 = 12\,\mathbb{E}_0\left[Z(h)^\top BZ(h)\right]^2.$$

Hence:

$$\Bigg|\int_0^T\int_0^T\int_0^{x_2} |f(x_1,x_2;t)|\ \mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}Z(h)\|\right.$$
$$\left.\times\ \|CD_{x_2}Z(h)\|\,Z(h)^\top BZ(h)\right]dx_1dx_2dt\Bigg|$$
$$\leqslant \left(\frac{d}{8}\int_0^T \mathbb{E}_0[Z_t^2]dt\,15\,\mathbb{E}_0[Z(h)^\top BZ(h)]^2\,12\,\mathbb{E}_0[Z(h)^\top BZ(h)]^2\right)^{1/2}$$
$$= \left(\frac{45d}{2}\int_0^T \mathbb{E}_0[Z_t^2]dt\ \mathbb{E}_0[Z(h)^\top BZ(h)]^4\right)^{1/2}$$
$$= \left(\frac{45d}{2}\int_0^T \mathbb{E}_0[Z_t^2]dt\right)^{1/2}\mathbb{E}_0[Z(h)^\top BZ(h)]^2.$$

(d) We find for the left-hand side of inequality (7.1), $a > 0$ and $k \in (0,1)$:

$$(2d-5)\,\mathbb{E}_0\left[\left(1+\sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^{-2}\right] - \frac{56}{\sqrt{a}}\int_0^T\int_0^T\int_0^{x_2}|f(x_1,x_2;t)|$$
$$\times\,\mathbb{E}_0\left[\|Cg(t)\|\,\|CD_{x_1}Z(h)\|\,\|CD_{x_2}Z(h)\|\ \left(\frac{1}{k^2}+\frac{Z(h)^\top BZ(h)}{a(1-k)^2}\right)\right]dx_1dx_2dt$$
$$\geqslant (2d-5)\left(1+\sqrt{\frac{\mathbb{E}_0[Z(h)^\top BZ(h)]}{a}}\right)^{-2}$$
$$-\ \frac{56\sqrt{d}}{k^2\sqrt{2a}}\sqrt{\int_0^T \mathbb{E}_0[Z_t^2]dt}\ \mathbb{E}_0[Z(h)^\top BZ(h)]$$
$$-\ \frac{56\sqrt{45d}}{a(1-k)^2\sqrt{2a}}\sqrt{\int_0^T \mathbb{E}_0[Z_t^2]dt}\ \mathbb{E}_0[Z(h)^\top BZ(h)]^2. \tag{7.3}$$

For concrete situations, this last expression can be useful to find possible values for $a$. It is however obvious that the calculations above do not provide the optimal value for $a$.

*Example* 7.2. (a) In practice it may be useful to consider a noise that has a constant variance. We consider thus the model given by the equation

$$X_0 = 0;\quad X_t = u_t + \epsilon\, t^{-H}Z_t,\quad t \in (0,T], \tag{7.4}$$

where $\epsilon > 0$ and $(Z_t)_{t\in[0,T]}$ *is a standard Rosenblatt process with Hurst parameter* $H \in (1/2,1)$. We have then $\mathbb{E}_u[(X_t-u_t)^2] = \epsilon^2 t^{-2H}t^{2H} = \epsilon^2$ for every $t \in (0,T]$. To simplify notations and to avoid confusion with the initial model, we write:

$$\tilde{X}(t,\epsilon,T) = u_t + \epsilon\, t^{-H}Z_t = u_t + \tilde{Z}(t,\epsilon,T),\quad t \in (0,T], \tag{7.5}$$

and we define $\tilde{X}(0,\epsilon,T)=0$. We use only simple random variables $\tilde{X}(t_i,\epsilon,T)$ to construct an estimator of the form

$$\tilde{X}(t,\epsilon,T)-\tilde{\xi}_a(t,\epsilon,T)=\tilde{X}(t,\epsilon,T)-\frac{\tilde{g}(t,\epsilon,T)^\top\tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}{a+\tilde{X}(t_i,\epsilon,T)_i^\top\tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}, \tag{7.6}$$

where $t_i=iT/d$ for $i=1,\dots,d$ and $d\geqslant 3$. Clearly $t_i$ depends on $T$ and we do not insist on this obvious dependence in the calculations below. We have used the following notations:

$$\tilde{X}(t_i,\epsilon,T)_i=\left(\tilde{X}(t_1,\epsilon,T),\dots,\tilde{X}(T,\epsilon,T)\right)^\top,$$
$$\tilde{B}(\epsilon,T)=\left(\int_0^T\tilde{g}(t,\epsilon,T)\tilde{g}(t,\epsilon,T)^\top dt\right)^{-1},$$

and for $t\in(0,T]$:

$$\tilde{g}(t,\epsilon,T)=\left(\tilde{g}_1(t,\epsilon,T),\dots,\tilde{g}_d(t,\epsilon,T)\right)^\top,$$
$$\tilde{g}_i(t,\epsilon,T)=\mathbb{E}_u\left[\tilde{Z}(t,\epsilon,T)\tilde{Z}(t_i,\epsilon,T)\right].$$

We define $\tilde{g}_i(0,\epsilon,T)=0$. The functions $\tilde{g}_i(\cdot,\epsilon,T)$ are right-continuous in $t=0$. We have seen in Theorem 4.1 that the estimator defined in Eq. (7.6) has smaller risk than the standard estimator if $a>A(\epsilon,T)$ where $A(\epsilon,T)$ does not depend on the drift and is supposed to be chosen as the infimum of all possible (positive) values. We prove the relation

$$A(\epsilon,T)\,\epsilon^2T=A(1,1) \tag{7.7}$$

in Section 9.5. We have obviously:

- $A(\epsilon,T)$ decreases if $\epsilon>0$ is fixed and $T$ increases,
- $A(\epsilon,T)$ decreases if $\epsilon>0$ increases and $T$ is fixed.

This observation reflects our intuition: if $\epsilon>0$ is small, $\tilde{X}(t,\epsilon,T)\approx u_t$. Since it is not possible to improve upon the "estimator" $(u_t)_{t\in[0,T]}$, the term $\tilde{\xi}_a(t,\epsilon,T)$ in our estimator should be small. This is realized in particular if $A(\epsilon,T)\to\infty$ for $\epsilon\downarrow 0$. On the other hand, a large value for $\epsilon$ reflects an important noise. The standard estimator for the drift is thus not very good and the term $\tilde{\xi}_a(t,\epsilon,T)$ in $\tilde{X}(t,\epsilon,T)-\tilde{\xi}_a(t,\epsilon,T)$ should allow an improvement upon $\tilde{X}(t,\epsilon,T)$. This is realized in particular if $A(\epsilon,T)\downarrow 0$ for $\epsilon\to\infty$. If $A(\epsilon,T)$ or $A(1,1)$ cannot be chosen as the infimum of all possible values, Eq. (7.7) becomes an inequality. For instance, if $A(1,1)$ is estimated by $A'(1,1)$ and is not optimal, we do not have an optimal value for $A(\epsilon,T)$ either, but we can state:

$$A(\epsilon,T)\leqslant\frac{A'(1,1)}{T\epsilon^2}.$$

For $d = 3$ and $H = 0.55$, numerical calculations and the estimations of Remark 7.1 give for instance $A'(1,1) \approx 4.727 \times 10^7$, with $T = 1\,000$ and $\epsilon = 100$, we find $A(\epsilon, T) \leqslant 4.727$. We notice that $\epsilon^2 T$ is the variance of the noise integrated over the interval $[0,T]$. We have $\int_0^T \mathbb{E}_u[\tilde{Z}(t,\epsilon,T)^2]dt = \epsilon^2 T$.

(b) It should be noticed that similar calculations can be made for the process given by

$$X_t = u_t + \epsilon Z_t, \quad t \in [0,T], \tag{7.8}$$

where $(Z_t)_{t\in[0,T]}$ is a standard Rosenblatt process as above. We find then:

$$A(\epsilon, T)\,\epsilon^2 T^{2H+1} = A(1,1).$$

We have for the process defined by Eq. (7.8):

$$\int_0^T \mathbb{E}_u[(X_t - u_t)^2]dt = \int_0^T \epsilon^2\, t^{2H} dt = \frac{\epsilon^2\, T^{2H+1}}{2H+1}.$$

The factor $\epsilon^2 T^{2H+1}$ is again a multiple of the variance of the process integrated over $[0,T]$.

We conclude:

**Proposition 7.3.** *Consider a Rosenblatt process $(Z_t)_{t\in[0,T]}$ with Hurst parameter $1/2 < H < 1$ and the situation defined in Eq. (7.4) respectively in Eq. (7.5) by:*

$$\begin{aligned} X_0 = 0; &\quad X_t = u_t + \epsilon\, t^{-H} Z_t, \quad t \in (0,T], \\ \tilde{X}(0,\epsilon,T) = 0; &\quad \tilde{X}(t,\epsilon,T) = u_t + \epsilon\, t^{-H} Z_t = u_t + \tilde{Z}(t,\epsilon,T), \quad t \in (0,T], \end{aligned}$$

*with $\epsilon = 100$ and $T = 1\,000$. Consider the functions*

$$\tilde{g}_i(\cdot,\epsilon,T) : t \mapsto \mathbb{E}_u[\tilde{Z}(t,\epsilon,T)\tilde{Z}(t_i,\epsilon,T)]$$

*and $d = 3$. The estimator defined in Eq. (7.6) by*

$$\tilde{X}(t,\epsilon,T) - \tilde{\xi}_a(t,\epsilon,T) = \tilde{X}(t,\epsilon,T) - \frac{\tilde{g}(t,\epsilon,T)^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}{a + \tilde{X}(t_i,\epsilon,T)_i^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i},$$

*has smaller risk than the standard estimator for $a > 4.727$.*

### *7.2. Estimating A in the martingale case*

As we did in Section 7.1, we can estimate the constant $A$ in the martingale case. We notice that inequality (7.3) holds generally for a noise living in the second chaos and is not specific for the case of the Rosenblatt process.

We choose in Eq. (4.1) $f(x_1,x_2;t) := 1_{[0\leqslant \min\{x_1,x_2\}\leqslant \max\{x_1,x_2\}\leqslant t]}/2$. This leads to the process $(X_t)_{t\in[0,T]}$ defined by:

$$X_t = u_t + \int_0^T \int_0^T \frac{1}{2}\, 1_{[0\leqslant \min\{x_1,x_2\}\leqslant \max\{x_1,x_2\}\leqslant t]} dW^u_{x_1} dW^u_{x_2} \tag{7.9}$$

$$
\begin{aligned}
&= u_t + \int_0^T \int_0^T 1_{[0\leqslant x_1 \leqslant x_2 \leqslant t]} dW^u_{x_1} dW^u_{x_2}, \\
X_t &= u_t + \frac{1}{2}\left[(W^u_t)^2 - t\right], \quad t \in [0,T].
\end{aligned}
$$

The noise $Z_t = 1/2\left[\left(W^u_t\right)^2 - t\right]$ is a self-similar process. This follows from the self-similarity of the Brownian motion. We prove the equivalence of all finite dimensional distributions. Consider $s > 0$ and $t_1, \ldots, t_n \in \mathbb{R}$, then:

$$
\begin{aligned}
\mathbb{P}_u\left(Z_{s\,t_1} \leqslant a_1, \ldots, Z_{s\,t_n} \leqslant a_n\right) &= \mathbb{P}_u\left((W^u_{s\,t_i})^2 \leqslant 2a_i + t_i s\ ,\ i = 1, \ldots, n\right) \\
&= \mathbb{P}_u\left(s(W^u_{t_i})^2 \leqslant 2a_i + t_i s\ ,\ i = 1, \ldots, n\right) \\
&= \mathbb{P}_u\left(\frac{s}{2}[(W^u_{t_i})^2 - t_i] \leqslant a_i\ ,\ i = 1, \ldots, n\right) \\
&= \mathbb{P}_u\left(sZ_{t_1} \leqslant a_1, \ldots, sZ_{t_n} \leqslant a_n\right).
\end{aligned}
$$

Thus $(Z_{ct}) \overset{(d)}{=} (c\,Z_t)$ where $\overset{(d)}{=}$ means equivalence of all finite dimensional distributions.

We consider the analogue of Eq. (7.4) in the present context:

$$
\tilde{X}(0,\epsilon,T) := 0; \quad \tilde{X}(t,\epsilon,T) = u_t + \frac{\sqrt{2}\epsilon}{t}\frac{(W^u_t)^2 - t}{2}, \quad t \in (0,T]. \tag{7.10}
$$

The estimation of the constant $A$ for the model Eq. (7.10) is analogue to the estimation in Section 7.1. We consider again a drift estimator of the form

$$
\tilde{X}(t,\epsilon,T) - \tilde{\xi}_a(t,\epsilon,T) = \tilde{X}(t,\epsilon,T) - \frac{\tilde{g}(t,\epsilon,T)^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}{a + \tilde{X}(t_i,\epsilon,T)_i^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i} \tag{7.11}
$$

and $t_i = iT/d$ for $d \geqslant 3$. Using the relation

$$
\mathbb{E}_u\left[\left((W^u_a)^2 - a\right)\ \left((W^u_b)^2 - b\right)\right] = 2\min\{a,b\}^2,
$$

we find that:

$$
\mathrm{Cov}_u(\tilde{X}(t,\epsilon,T), \tilde{X}(t_i,\epsilon,T)) = \frac{\epsilon^2}{t_i\,t}\min\{t,t_i\}^2
$$

With the self-similarity of the noise $(Z_t)_{t\in[0,T]}$ we can prove an analogue of Eq. (7.7) and find numerically as in Section 7.1 for $d = 3$, $T = 1\,000$ and $\epsilon = 100$ that $A(\epsilon,T) \leqslant 1.057$. We conclude:

**Proposition 7.4.** *Consider the situation defined in Eq. (7.10) by:*

$$
\tilde{X}(0,\epsilon,T) := 0; \quad \tilde{X}(t,\epsilon,T) = u_t + \frac{\sqrt{2}\epsilon}{t}\frac{(W^u_t)^2 - t}{2} = u_t + \tilde{Z}(t,\epsilon,T), \quad t \in (0,T],
$$

*with $\epsilon = 100$ and $T = 1\,000$. Consider the functions*

$$
\tilde{g}_i(\cdot,\epsilon,T) : t \mapsto \mathbb{E}_u[\tilde{Z}(t,\epsilon,T)\tilde{Z}(t_i,\epsilon,T)]
$$

*and* $d = 3$. *The estimator defined in Eq.* (7.11) *by*

$$\tilde{X}(t,\epsilon,T) - \tilde{\xi}_a(t,\epsilon,T) = \tilde{X}(t,\epsilon,T) - \frac{\tilde{g}(t,\epsilon,T)^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}{a + \tilde{X}(t_i,\epsilon,T)_i^\top \tilde{B}(\epsilon,T)\tilde{X}(t_i,\epsilon,T)_i}$$

*has smaller risk than the standard estimator for* $a > 1.057$.

## 8. Discrete-time results

In this section, we give, without proof, a discrete version of the continuous-time models considered so far.

### 8.1. Discrete-time version of the estimator

Consider the martingale case of Section 6.1. In Proposition 5.1, we have noticed that the results of Theorem 4.1 hold in a Wiener-Itô chaos of higher order. We consider a stochastic process $(X_t)_{t\in[0,T]}$ with:

$$\begin{aligned} X_t &= u_t + \int_0^t \int_0^{x_n} \cdots \int_0^{x_2} dW^u_{x_2} \cdots dW^u_{x_n} \\ &= u_t + \frac{1}{n!} H_n\left(\frac{W^u_t}{\sqrt{t}}\right)\sqrt{t}^n, \quad t \in [0,T], \end{aligned} \tag{8.1}$$

where $H_n$ is the Hermite polynomial of order $n$. Using the properties of the Hermite polynomials and the classical version of the integration by parts formula, we can show the following discrete version of Theorem 4.1:

**Theorem 8.1.** *Consider the* $d$*-dimensional random vector*

$$X^\top = (\ldots, X_i, \ldots) = \left(\ldots, \frac{1}{n!} H_n\left(\frac{Y_i}{\sqrt{\sigma_{i,i}}}\right)\sqrt{\sigma_{i,i}}^n + \mu_i, \ldots\right),$$

*where*

$$Y \sim \mathcal{N}(0,\Sigma), \quad \Sigma = \begin{pmatrix} \sigma_{1,1} & \cdots & \sigma_{1,d} \\ \cdots & \cdots & \cdots \\ \sigma_{d,1} & \cdots & \sigma_{d,d} \end{pmatrix},$$

*and the matrix* $\Sigma$ *is positive definite and symmetric. We write* $A^{*n}$ *for the Hadamard product of* $n$ *matrices* $A$. *The estimator*

$$\delta(X) = X - \frac{n!\,(\Sigma^{*n})^{-1} X}{a + \|n!\,(\Sigma^{*n})^{-1} X\|^2} = X - \frac{\mathrm{Cov}_\mu(X)^{-1} X}{a + \|\mathrm{Cov}_\mu(X)^{-1} X\|^2} \tag{8.2}$$

*has smaller risk than* $X$ *as an estimator for* $\mu$ *with respect to the quadratic loss function, provided that* $a$ *is large enough and* $d \geqslant 3$.

### 8.2. Approximation of the discrete-time estimator

Consider the discrete situation with $Y \sim \mathcal{N}(0, \Sigma)$ and

$$X_i = \frac{1}{n!} H_n\left(\frac{Y_i}{\sqrt{\sigma_{i,i}}}\right) \sqrt{\sigma_{i,i}}^n + \mu_i, \quad i = 1, \ldots, d. \tag{8.3}$$

Suppose that $\Sigma = (\min\{t_i, t_j\})_{i,j=1,\ldots,d}$ for $0 < t_1 < t_2 < \ldots < t_d \leqslant T$. This corresponds to $Y_i = W_{t_i}$ for a standard Brownian motion. Eq. (8.3) can be regarded as the discrete version of Eq. (8.1). The estimator given in Eq. (8.2) is connected to the estimator for the continuous-time model. We can recover the estimator for the discrete-time model by considering the continuous-time version and choosing $h_i : t \mapsto \min\{t_i, t\}$ and approximating the integrals $\int_0^T g_i(s) g_j(s) ds$ appearing in $B$ by Riemann sums. We find that the estimator for the continuous version becomes for $t = t_j$:

$$\delta'(X) \approx X_{t_j} - \frac{dm!\,(\Sigma^{*m})^{-1}_{j\text{-th row}} X(h)}{a + d\|m!\,(\Sigma^{*m})^{-1} X(h)\|^2}.$$

This corresponds to the $j$-th component of the estimator for the discrete version up to a factor $d$ (which can be eliminated by replacing $a$ by $ad$ in the estimator for the continuous version).

## 9. Appendix

### 9.1. Proof of Theorem 2.1

Let $(\xi_t)_{t\in[0,T]}$ be an *unbiased adapted* drift estimator:

$$\mathbb{E}_u[\xi_t] = \mathbb{E}_u[u_t], \quad t \in [0, T],$$

where $u_t = \int_0^t \dot{u}_s ds$. Consider $v_t := \int_0^t b_s^2 ds$. Condition 4 implies that $u + \epsilon v \in L^2(\Omega \times [0,T], \mathbb{P}_{u+\epsilon v} \otimes \lambda_T)$ and $u_t + \epsilon v_t = \int_0^t \left(\dot{u}_s + \epsilon b_s^2\right) ds$. We can suppose without loss of generality that $U \subset (-1, 1)$. We have for the unbiased adapted drift estimator for every $|\epsilon| < 1$ and $t \in [0, T]$:

$$\mathbb{E}_{u+\epsilon v}[\xi_t] = \mathbb{E}_{u+\epsilon v}[u_t + \epsilon v_t]$$

and

$$\mathbb{E}_{u+\epsilon v}[\xi_t - u_t] = \epsilon\,\mathbb{E}_{u+\epsilon v}[v_t] = \epsilon\,\mathbb{E}_{u+\epsilon v}\left[\int_0^t b_s^2 ds\right] < \infty. \tag{9.1}$$

(a) We have:

$$dX_t = \dot{u}_t dt + b_t dW_t^u, \tag{9.2}$$

where $(W_t^u)_{t\in[0,T]}$ is a standard Wiener process with respect to $\mathbb{P}_u$. Since condition 1 holds, we can use Girsanov's theorem and define $Q$ by:

$$\frac{dQ}{d\,\mathbb{P}_u} = M_T \text{ and } \frac{d\,\mathbb{P}_u}{dQ} = M_T^{-1}$$

for

$$M_t := \exp\left(-\int_0^t \frac{\dot{u}_s}{b_s} dW_s^u - \frac{1}{2}\int_0^t \left(\frac{\dot{u}_s}{b_s}\right)^2 ds\right), \quad t \in [0,T].$$

We then have with Eq. (9.2):

$$\begin{aligned}
\frac{d\,\mathbb{P}_u}{dQ} &= \exp\left(\int_0^T \frac{\dot{u}_s}{b_s} dW_s^u + \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s}{b_s}\right)^2 ds\right) \\
&= \exp\left(\int_0^T \frac{\dot{u}_s}{b_s^2} b_s dW_s^u + \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s}{b_s}\right)^2 ds\right) \\
&= \exp\left(\int_0^T \frac{\dot{u}_s}{b_s^2} dX_s - \int_0^T \frac{\dot{u}_s}{b_s^2}\dot{u}_s ds + \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s}{b_s}\right)^2 ds\right) \\
&= \underbrace{\exp\left(\int_0^T \frac{\dot{u}_s}{b_s^2} dX_s - \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s}{b_s}\right)^2 ds\right)}_{=:\Lambda(u)}.
\end{aligned}$$

(b) We have:

$$\begin{aligned}
&\frac{d}{d\epsilon}\,\mathbb{E}_{u+\epsilon v}[\xi_t - u_t]_{\epsilon=0} \\
&= \frac{d}{d\epsilon}\,\mathbb{E}_Q\left[(\xi_t - u_t)\exp\left(\int_0^T \frac{\dot{u}_s + \epsilon b_s^2}{b_s^2} dX_s - \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s + \epsilon b_s^2}{b_s}\right)^2 ds\right)\right]_{\epsilon=0} \\
&= \mathbb{E}_Q\left[(\xi_t - u_t)\frac{d}{d\epsilon}\left(\Lambda(u+\epsilon v)\right)_{\epsilon=0}\right] \\
&= \mathbb{E}_Q\left[(\xi_t - u_t)\frac{d}{d\epsilon}\left(\log\Lambda(u+\epsilon v)\right)_{\epsilon=0}\Lambda(u)\right] \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\frac{d}{d\epsilon}\left(\log\Lambda(u+\epsilon v)\right)_{\epsilon=0}\right].
\end{aligned}$$

We have then for $\frac{d}{d\epsilon}E_{u+\epsilon v}[\xi_t - u_t]_{\epsilon=0}$:

$$\begin{aligned}
&\frac{d}{d\epsilon}\,\mathbb{E}_{u+\epsilon v}[\xi_t - u_t]_{\epsilon=0} \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\frac{d}{d\epsilon}\left(\int_0^T \frac{\dot{u}_s + \epsilon b_s^2}{b_s^2} dX_s - \frac{1}{2}\int_0^T \frac{\left(\dot{u}_s + \epsilon b_s^2\right)^2}{b_s^2} ds\right)_{\epsilon=0}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_u\left[(\xi_t - u_t)\left(\int_0^T dX_s - \frac{1}{2}\int_0^T \frac{2\dot{u}_s b_s^2}{b_s^2}ds\right)\right] \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\left(\int_0^T dX_s - \int_0^T \dot{u}_s ds\right)\right] \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\int_0^T b_s dW_s^u\right].
\end{aligned}
$$

We give a justification for the interchange of differentiation and integration in (d) below. We condition on $\mathcal{F}_t$ and use the fact that $\left(\int_0^t b_s dW_s^u\right)_{t\in[0,T]}$ is a $(\mathcal{F}_t)_{t\in[0,T]}$-martingale to get:

$$
\begin{aligned}
\frac{d}{d\epsilon}\,\mathbb{E}_{u+\epsilon v}[\xi_t - u_t]_{\epsilon=0} &= \mathbb{E}_u\left[\mathbb{E}_u\left[(\xi_t - u_t)\int_0^T b_s dW_s^u | \mathcal{F}_t\right]\right] \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\,\mathbb{E}_u\left[\int_0^T b_s dW_s^u | \mathcal{F}_t\right]\right] \\
&= \mathbb{E}_u\left[(\xi_t - u_t)\int_0^t b_s dW_s^u\right]. \qquad (9.3)
\end{aligned}
$$

(c) We apply the same procedure to calculate $\frac{d}{d\epsilon}\epsilon\,\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0}$. We give a justification for the existence of $\frac{d}{d\epsilon}\,\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0}$ in (d).

$$
\begin{aligned}
\frac{d}{d\epsilon}\epsilon\,\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0} &= 1\;\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0} + 0\,\frac{d}{d\epsilon}\,\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0} = \mathbb{E}_u[v_t] \\
&= \mathbb{E}_u\left[\int_0^t b_s^2 ds\right]. \qquad (9.4)
\end{aligned}
$$

Comparing Eq. (9.1), Eq. (9.3) and Eq. (9.4), we get:

$$
\mathbb{E}_u\left[\int_0^t b_s^2 ds\right] = \mathbb{E}_u\left[(\xi_t - u_t)\int_0^t b_s dW_s^u\right].
$$

We apply the Cauchy-Schwarz inequality and the Itô isometry to get:

$$
\begin{aligned}
\mathbb{E}_u\left[\int_0^t b_s^2 ds\right]^2 &= \mathbb{E}_u\left[(\xi_t - u_t)\int_0^t b_s dW_s^u\right]^2 \\
&\leqslant \mathbb{E}_u\left[(\xi_t - u_t)^2\right]\,\mathbb{E}_u\left[\left(\int_0^t b_s dW_s^u\right)^2\right] \\
&\leqslant \mathrm{Var}_u(\xi_t)\,\underbrace{\mathbb{E}_u\left[\int_0^t b_s^2 ds\right]}_{<\infty}.
\end{aligned}
$$

Thus:

$$\mathbb{E}_u\left[\int_0^t b_s^2 ds\right] \leqslant \mathrm{Var}_u(\xi_t). \tag{9.5}$$

We calculate $\mathrm{Var}_u(X_t) = \mathbb{E}_u\left[(X_t - u_t)^2\right]$ by applying the Itô isometry:

$$\mathrm{Var}_u(X_t) = \mathbb{E}_u\left[(X_t - u_t)^2\right] = \mathbb{E}_u\left[\left(\int_0^t b_s dW_s^u\right)^2\right] = \mathbb{E}_u\left[\int_0^t b_s^2 ds\right]. \tag{9.6}$$

Comparing inequality (9.5) and Eq. (9.6), we get:

$$\mathrm{Var}_u(X_t) \leqslant \mathrm{Var}_u(\xi_t).$$

(d) Justification for the interchange of the expectation and the differentiation.

(i) We calculate $\Lambda(u+\epsilon v)$ and $\frac{d}{d\epsilon}\Lambda(u+\epsilon v)$.

$$\begin{aligned}
\Lambda(u+\epsilon v) &= \exp\left[\int_0^T \frac{\dot{u}_s + \epsilon b_s^2}{b_s^2} dX_s - \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s + \epsilon b_s^2}{b_s}\right)^2 ds\right] \\
&= \exp\Bigg[\int_0^T \frac{\dot{u}_s}{b_s^2} dX_s + \epsilon \int_0^T dX_s - \frac{1}{2}\int_0^T \left(\frac{\dot{u}_s}{b_s}\right)^2 ds \\
&\qquad - \frac{1}{2}2\epsilon \int_0^T \frac{\dot{u}_s b_s^2}{b_s^2} ds - \frac{1}{2}\epsilon^2 \int_0^T \frac{b_s^4}{b_s^2} ds\Bigg] \\
&= \Lambda(u)\, \exp(\epsilon X_T - \epsilon u_T) \exp\left(-\frac{1}{2}\epsilon^2 \int_0^T b_s^2 ds\right) \\
&= \Lambda(u)\, \exp\left[\epsilon(X_T - u_T)\right] \exp\left(-\frac{1}{2}\epsilon^2 \int_0^T b_s^2 ds\right).
\end{aligned}$$

The derivative $\frac{d}{d\epsilon}\Lambda(u+\epsilon v)$ equals then:

$$\begin{aligned}
&\Lambda(u)\Bigg[(X_T - u_T)\exp\left[\epsilon(X_T - u_T)\right] \exp\left(-\frac{1}{2}\epsilon^2 \int_0^T b_s^2 ds\right) \\
&\qquad + \exp\left[\epsilon(X_T - u_T)\right] \left(-\frac{1}{2}2\epsilon \int_0^T b_s^2 ds\right) \exp\left(-\frac{1}{2}\epsilon^2 \int_0^T b_s^2 ds\right)\Bigg] \\
&= \Lambda(u)\, \exp\left[\epsilon\left(X_T - u_T\right)\right] \exp\left(-\frac{1}{2}\epsilon^2 \int_0^T b_s^2 ds\right) \\
&\qquad \times \left[(X_T - u_T) - \epsilon \int_0^T b_s^2 ds\right] \\
&= \Lambda(u+\epsilon v)\, \left(\int_0^T b_s dW_s^u - \epsilon \int_0^T b_s^2 ds\right).
\end{aligned}$$

(ii) We have to show:

$$\left|(\xi_t - u_t)\frac{d}{d\epsilon}\Lambda(u+\epsilon v)\right| \leqslant M \in L^1(Q).$$

Then it is possible to interchange the expectation and the derivative. We have $\exp(|\epsilon x|) \leqslant \exp(|x|)$ for every $\epsilon$ with $|\epsilon| < 1$ and thus:

$$\begin{aligned}
&\left|(\xi_t - u_t)\frac{d}{d\epsilon}\Lambda(u+\epsilon v)\right| \\
&\leqslant |\xi_t - u_t|\Lambda(u)\exp\left[\epsilon\left(X_T - u_T\right)\right]\exp\left(-\frac{1}{2}\epsilon^2\int_0^T b_s^2 ds\right) \\
&\quad \times \left|\int_0^T b_s dW_s^u - \epsilon\int_0^T b_s^2 ds\right| \\
&\leqslant |\xi_t - u_t|\Lambda(u)\exp\left[\left|\epsilon\left(X_T - u_T\right)\right|\right]\left(\left|\int_0^T b_s dW_s^u\right| + |\epsilon|\left|\int_0^T b_s^2 ds\right|\right) \\
&\leqslant |\xi_t - u_t|\Lambda(u)\exp\left(|X_T - u_T|\right)\left(\left|\int_0^T b_s dW_s^u\right| + \left|\int_0^T b_s^2 ds\right|\right) \\
&\leqslant |\xi_t - u_t|\Lambda(u)\exp\left(\left|\int_0^T b_s dW_s^u\right|\right)\left(\left|\int_0^T b_s dW_s^u\right| + \left|\int_0^T b_s^2 ds\right|\right) \\
&\leqslant |\xi_t - u_t|\Lambda(u)\exp\left(\left|\int_0^T b_s dW_s^u\right|\right) \\
&\quad \times \left[\exp\left(\left|\int_0^T b_s dW_s^u\right|\right) + \left|\int_0^T b_s^2 ds\right|\right] \\
&=: M.
\end{aligned}$$

We have with conditions 2 and 3:

$$\begin{aligned}
\mathbb{E}_Q[M] &= \mathbb{E}_u\left\{\underbrace{|\xi_t - u_t|}_{\in L^2(\Omega,\mathbb{P}_u)}\underbrace{\left[\exp\left(\left|\int_0^T b_s dW_s^u\right|\right)\right]^2}_{\in L^2(\Omega,\mathbb{P}_u)}\right\} \\
&\quad + \mathbb{E}_u\left\{\underbrace{|\xi_t - u_t|}_{\in L^2(\Omega,\mathbb{P}_u)}\underbrace{\exp\left(\left|\int_0^T b_s dW_s^u\right|\right)}_{\in L^4(\Omega,\mathbb{P}_u)}\underbrace{\left|\int_0^T b_s^2 ds\right|}_{\in L^4(\Omega,\mathbb{P}_u)}\right\} \\
&< \infty.
\end{aligned}$$

(iii) A similar argument proves the existence of $\frac{d}{d\epsilon}\mathbb{E}_{u+\epsilon v}[v_t]_{\epsilon=0}$. □

### 9.2. Proof of Theorem 2.2

(1) We know that (see [17, p.15-16]) the class of elementary adapted processes is dense in the space of adapted and square-integrable (with respect to $\mathbb{P}_u \otimes \lambda_T$) processes $L_a^2(\Omega \times [0,T])$. Every elementary adapted process $(v_t)_{t\in[0,T]}$ can be written as

$$\sum_{i=1}^{n} F_i 1_{(t_i,t_{i+1}]}(t), \quad t \in [0,T],$$

where $0 \leqslant t_1 < \ldots < t_{n+1} \leqslant T$ are elements of $[0,T]$ and every $F_i$ is $\mathcal{F}_{t_i}$-measurable and square-integrable (with respect to $\mathbb{P}_u$). We can find a sequence of elementary adapted processes $(b(s,n))_{s\in[0,T]}$ with

$$\lim_{n\to+\infty} \mathbb{E}_u \left[ \int_0^T (b_s - b(s,n))^2 \, ds \right] = 0. \tag{9.7}$$

We define for $1 > \epsilon > 0$, $M > \epsilon$ and $n \in \mathbb{N}$:

$$X_{t,\epsilon,M,n} = u_t + \int_0^t b(s,n)_\epsilon^M dW_s^u, \quad t \in [0,T]$$

where $b(\cdot,n)_\epsilon^M$ is the truncated version of the process $b(\cdot,n)$. We define:

$$b(s,n)_\epsilon^M = \begin{cases} b(s,n) & \text{if } \epsilon <\mid b(s,n) \mid \leqslant M \\ \epsilon & \text{if } 0 \leqslant b(s,n) \leqslant \epsilon \\ -\epsilon & \text{if } 0 > b(s,n) \geqslant -\epsilon \\ M & \text{if } b(s,n) > M \\ -M & \text{if } b(s,n) < -M \end{cases}$$

and

$$b(s,n)^M = \begin{cases} b(s,n) & \text{if } \mid b(s,n) \mid \leqslant M \\ M & \text{if } b(s,n) > M \\ -M & \text{if } b(s,n) < -M. \end{cases}$$

(a) We show:

$$\lim_{n\to\infty} \lim_{M\to\infty} \lim_{\epsilon\to 0^+} \mathbb{E}_u \left[ \int_0^T \left( b_s - b(s,n)_\epsilon^M \right)^2 ds \right] = 0.$$

We calculate

$$\lim_{\epsilon\to 0^+} \mathbb{E}_u \left[ \int_0^T \left( b_s - b(s,n)_\epsilon^M \right)^2 ds \right].$$

We have for every $1 > \epsilon > 0$ and every fixed $n \in \mathbb{N}$, $M > \epsilon$:

$$\left( b_s - b(s,n)_\epsilon^M \right)^2 \leqslant 2b_s^2 + 2\left( b(s,n)_\epsilon^M \right)^2 \leqslant 2b_s^2 + 2|b(s,n)|^2 + 2.$$

Since $\mathbb{E}_u\left[\int_0^T b_s^2 ds\right] < \infty$ and $(b(s,n))_{s\in[0,T]}$ is an elementary adapted process, we can apply Lebesgue's theorem and we get:

$$\lim_{\epsilon\to 0^+} \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)_\epsilon^M\right)^2 ds\right] = \mathbb{E}_u\left[\int_0^T \lim_{\epsilon\to 0^+} \left(b_s - b(s,n)_\epsilon^M\right)^2 ds\right]$$
$$= \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)^M\right)^2 ds\right].$$

We notice that $\lim_{\epsilon\to 0^+} b(s,n)_\epsilon^M = b(s,n)^M$ holds almost surely. Indeed:

- If $b(s,n)^M(\omega) \neq 0$, we have $b(s,n)_\epsilon^M(\omega) = b(s,n)^M(\omega)$ for all sufficiently small $\epsilon > 0$.
- If $b(s,n)^M(\omega) = 0$, we have $b(s,n)(\omega) = 0$, $b(s,n)_\epsilon^M(\omega) = \epsilon$ and thus:

$$\lim_{\epsilon\to 0^+} b(s,n)_\epsilon^M(\omega) = \lim_{\epsilon\to 0^+} \epsilon = 0.$$

(b) Now we calculate:

$$\lim_{M\to\infty}\lim_{\epsilon\to 0^+} \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)_\epsilon^M\right)^2 ds\right].$$

We have for every $M > 1$ and every fixed $n \in \mathbb{N}$:

$$\left(b_s - b(s,n)^M\right)^2 \leqslant 2b_s^2 + 2\left(b(s,n)^M\right)^2 \leqslant 2b_s^2 + 2|b(s,n)|^2.$$

Since $\mathbb{E}_u\left[\int_0^T b_s^2 ds\right] < \infty$ and $(b(s,n))_{s\in[0,T]}$ is an elementary adatped process, we can apply Lebesgue's theorem and we get together with (a):

$$\lim_{M\to\infty}\lim_{\epsilon\to 0^+} \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)_\epsilon^M\right)^2 ds\right]$$
$$= \lim_{M\to\infty} \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)^M\right)^2 ds\right]$$
$$= \mathbb{E}_u\left[\int_0^T \lim_{M\to\infty} \left(b_s - b(s,n)^M\right)^2 ds\right]$$
$$= \mathbb{E}_u\left[\int_0^T (b_s - b(s,n))^2 ds\right].$$

(c) We finally calculate

$$\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+} \mathbb{E}_u\left[\int_0^T \left(b_s - b(s,n)_\epsilon^M\right)^2 ds\right].$$

We have together with (a) and (b) and Eq. (9.7):

$$\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+}\mathbb{E}_u\left[\int_0^T\left(b_s-b(s,n)_\epsilon^M\right)^2ds\right]$$
$$=\lim_{n\to\infty}\mathbb{E}_u\left[\int_0^T\left(b_s-b(s,n)\right)^2ds\right]=0.$$

(2) We have for every $t\in[0,T]$:

$$\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+}\mathbb{E}_u\left[\left((X_{t,\epsilon,M,n}-u_t)-(X_t-u_t)\right)^2\right]$$
$$=\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+}\mathbb{E}_u\left[\int_0^t\left(b_s-b(s,n)_\epsilon^M\right)^2ds\right]$$
$$\leqslant\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+}\mathbb{E}_u\left[\int_0^T\left(b_s-b(s,n)_\epsilon^M\right)^2ds\right]=0.$$

Thus for every $t\in[0,T]$:

$$\lim_{n\to\infty}\lim_{M\to\infty}\lim_{\epsilon\to 0^+}\mathbb{E}_u\left[(X_{t,\epsilon,M,n}-u_t)^2\right]=\mathbb{E}_u\left[(X_t-u_t)^2\right]. \tag{9.8}$$

(3) We check the conditions of Theorem 2.1 for the process $b(\cdot,n)_\epsilon^M$. We suppose that $(u_t)_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted and $\mathbb{P}_u\left(\int_0^T\dot{u}_t^2dt\leqslant C\right)=1$ for some positive constant $C$. We suppose that the elementary adapted process $b(\cdot,n)$ can be written as:

$$b(s,n)=\sum_{j=1}^{m_n-1}F_j1_{(t_j,t_{j+1}]}(s).$$

where without loss of generality $t_1=0$ and $t_{m_n}=T$.

(a) With $\mathbb{P}_u\left(\int_0^T\dot{u}_t^2dt\leqslant C\right)=1$ and $z\in U$, where $U\subset(-1,1)$ is a neighbourhood of 0, we obtain:

$$\mathbb{E}_u\left[\exp\left(\int_0^T\frac{1}{2}\left(\frac{\dot{u}_s+z\,\left(b(s,n)_\epsilon^M\right)^2}{b(s,n)_\epsilon^M}\right)^2ds\right)\right]$$
$$\leqslant\mathbb{E}_u\left[\exp\left(\frac{1}{2}\int_0^T\epsilon^{-2}\,\dot{u}_s^2ds+z\int_0^T\dot{u}_sds+\frac{z^2}{2}\int_0^T\left(b(s,n)_\epsilon^M\right)^2ds\right)\right]$$
$$\leqslant\mathbb{E}_u\left[\exp\left(\frac{1}{2\epsilon^2}\int_0^T\dot{u}_s^2ds+|z|\sqrt{T}\left(\int_0^T\dot{u}_s^2ds\right)^{1/2}+\frac{z^2}{2}T\,M^2\right)\right]$$
$$<\infty.$$

The first inequality follows with the definition of $b(s,n)_\epsilon^M$, the second inequality follows with the Cauchy-Schwarz inequality and since $x\mapsto\exp(x)$ is an increasing function.

(b) We have:

$$\left|\int_0^T b(s,n)_\epsilon^M dW_s^u\right| = \left|\sum_{i=1}^m (F_i)_\epsilon^M \left(W_{t_{i+1}}^u - W_{t_i}^u\right)\right|$$
$$\leqslant \sum_{i=1}^m \underbrace{|(F_i)_\epsilon^M|}_{\leqslant M} |W_{t_{i+1}}^u - W_{t_i}^u|$$
$$\leqslant \sum_{i=1}^m M\,|W_{t_{i+1}}^u - W_{t_i}^u|.$$

$$\mathbb{E}_u\left[\exp\left(\mid \int_0^T b(s,n)_\epsilon^M dW_s^u \mid\right)^4\right] \leqslant \mathbb{E}_u[\exp(4M\sum_{i=1}^m \mid W_{t_{i+1}}^u - W_{t_i}^u \mid)]$$
$$= \prod_{i=1}^m \mathbb{E}_u\left[\exp\left(4M\,|Z_i|\right)\right] < \infty.$$

We have used that the standard Wiener process has independent increments and considered $Z_i \sim \mathcal{N}\left(0, t_{i+1} - t_i\right)$ with respect to $\mathbb{P}_u$.

(c) We have:

$$\mathbb{E}_u\left[\left(\int_0^T \left(b(s,n)_\epsilon^M\right)^2 ds\right)^4\right] \leqslant \left(M^2T\right)^4 < \infty.$$

(d) Finally:

$$\int_0^T \mathbb{E}_u\left[\left(\int_0^t \left(b(s,n)_\epsilon^M\right)^2 ds\right)^2\right] dt \leqslant \int_0^T M^4t^2dt = \frac{1}{3}M^4T^3 < \infty.$$

(4) We suppose that $(u_t)_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted and $\mathbb{P}_u\left(\int_0^T \dot{u}_s^2 ds \leqslant C\right) = 1$ for a positive constant $C$. We apply the results of Theorem 2.1 and get for every unbiased adapted drift estimator $\xi_{t,\epsilon,M,n}$ in the situation $X_{t,\epsilon,M,n} = u_t + \int_0^t b(s,n)_\epsilon^M dW_s^u$:

$$\mathbb{E}_u\left[(\xi_{t,\epsilon,M,n} - u_t)^2\right] \geqslant \mathbb{E}_u\left[(X_{t,\epsilon,M,n} - u_t)^2\right], \quad t \in [0,T]. \tag{9.9}$$

(5) We suppose as above that $(u_t)_{t\in[0,T]}$ is $\mathcal{F}_t$-adapted and $\mathbb{P}_u\left(\int_0^T \dot{u}_s^2 ds \leqslant C\right) = 1$ for a positive constant $C$. We now consider an unbiased adapted drift estimator $(\xi_t)_{t\in[0,T]}$ for the situation

$$X_t = u_t + \int_0^t b_s dW_s^u, \quad t \in [0,T].$$

We also consider the "truncated model" defined above:

$$X_{t,\epsilon,M,n} = u_t + \int_0^t b(s,n)_\epsilon^M dW_s^u, \quad t \in [0,T].$$

From (4) above, for a deterministic drift $(u_t)_{t\in[0,T]}$ which satisfies the conditions of Section 2.1, we have in particular that inequality (9.9) holds.

(a) We define $\xi_{t,\epsilon,M,n} = \mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]$. The estimator $\xi_{t,\epsilon,M,n}$ is a function of $X_{t,\epsilon,M,n}$ only, $\xi_{t,\epsilon,M,n}$ is an unbiased adapted drift estimator for the "truncated model":

$$\mathbb{E}_u[\xi_{t,\epsilon,M,n}] = \mathbb{E}_u\left[\mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]\right] = \mathbb{E}_u\left[\xi_t\right] = \mathbb{E}_u\left[u_t\right] = u_t,\ \ t \in [0,T].$$

(b) Since the conditional variance is a.s. non-negative we have for every $t \in [0,T]$:

$$0 \leqslant \mathrm{Var}_u\left[\xi_t | X_{t,\epsilon,M,n}\right] = \mathbb{E}_u\left[\xi_t^2 | X_{t,\epsilon,M,n}\right] - \mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]^2.$$

Thus for every $t \in [0,T]$:

$$\mathbb{E}_u\left[\xi_t^2 | X_{t,\epsilon,M,n}\right] \geqslant \mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]^2.$$

This is also immediate with Jensen's inequality.

(c) We get for every $t \in [0,T]$:

$$\begin{aligned}
&\mathbb{E}_u\left[\xi_t^2 | X_{t,\epsilon,M,n}\right] \geqslant \mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]^2 \\
\Rightarrow\ &\mathbb{E}_u\left[\mathbb{E}_u\left[\xi_t^2 | X_{t,\epsilon,M,n}\right]\right] \geqslant \mathbb{E}_u\left[\mathbb{E}_u\left[\xi_t | X_{t,\epsilon,M,n}\right]^2\right] \\
\Rightarrow\ &\mathbb{E}_u\left[\xi_t^2\right] \geqslant \mathbb{E}_u\left[\xi_{t,\epsilon,M,n}^2\right] \\
\Rightarrow\ &\mathbb{E}_u\left[\xi_t^2\right] - u_t^2 \geqslant \mathbb{E}_u\left[\xi_{t,\epsilon,M,n}^2\right] - u_t^2.
\end{aligned}$$

For a deterministic drift $(u_t)_{t\in[0,T]}$, we deduce that:

$$\mathbb{E}_u[(\xi_t - u_t)^2] \geqslant \mathbb{E}_u[(\xi_{t,\epsilon,M,n} - u_t)^2],$$

and with inequality (9.9), we find for every $t \in [0,T]$:

$$\begin{aligned}
&\mathbb{E}_u\left[(\xi_t - u_t)^2\right] \geqslant \mathbb{E}_u\left[(\xi_{t,\epsilon,M,n} - u_t)^2\right] \geqslant \mathbb{E}_u\left[(X_{t,\epsilon,M,n} - u_t)^2\right] \\
\Rightarrow\ &\mathbb{E}_u\left[(\xi_t - u_t)^2\right] \geqslant \mathbb{E}_u\left[(X_{t,\epsilon,M,n} - u_t)^2\right].
\end{aligned} \tag{9.10}$$

(d) Inequality (9.10) holds for every $\epsilon, M, n$ and taking into account Eq. (9.8), we finally find:

$$\mathbb{E}_u\left[(\xi_t - u_t)^2\right] \geqslant \mathbb{E}_u\left[(X_t - u_t)^2\right].$$

This completes the proof. ∎

### *9.3. Inequalities and a property of the Hadamard product*

We prove the inequalities used to demonstrate the main result and that the sum of elements of the Hadamard product $M * M^{-1}$ equals the dimension of the matrix.

**Theorem 9.1.** *We have for $a > 0$, $k \in (0,1)$ and $X(h) = u(h) + Z(h)$:*

$$\frac{a + u(h)^\top Bu(h)}{a + X(h)^\top BX(h)} \geqslant \left(1 + \sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^{-2},$$

$$\frac{a + u(h)^\top Bu(h)}{a + X(h)^\top BX(h)} \leqslant \frac{1}{k^2} + \frac{Z(h)^\top BZ(h)}{(k-1)^2 a},$$

*and for $Q = [a + X(h)^\top BX(h)]^{-1}$ respectively $Q = k^{-2} + [Z(h)^\top BZ(h)]/[a(1-k)^2]$:*

$$\int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[|f(x_1, x_2; t)| \, \|Cg(t)\| \, \|CD_{x_2}X(h)\| \, \|CD_{x_1}X(h)\| \, Q\right] dx_1 dx_2 dt < \infty.$$

*Proof.* (a) In the inequalities below, all numerators and denominators are positive:

$$\begin{aligned}
\frac{a + u(h)^\top Bu(h)}{a + X(h)^\top BX(h)} &= \frac{a + u(h)^\top Bu(h)}{a + u(h)^\top Bu(h) + 2u(h)^\top BZ(h) + Z(h)^\top BZ(h)} \\
&\geqslant \frac{a + u(h)^\top Bu(h)}{a + u(h)^\top Bu(h) + 2\|Cu(h)\| \, \|CZ(h)\| + \|CZ(h)\|^2} \\
&= \frac{1}{1 + 2\frac{\|Cu(h)\|}{\sqrt{a+\|Cu(h)\|^2}} \frac{\|CZ(h)\|}{\sqrt{a+\|Cu(h)\|^2}} + \frac{\|CZ(h)\|^2}{a+\|Cu(h)\|^2}} \\
&\geqslant \frac{1}{1 + 2\,\frac{\|CZ(h)\|}{\sqrt{a}} + \frac{\|CZ(h)\|^2}{a}} \\
&= \frac{1}{\left(1 + \sqrt{\frac{Z(h)^\top BZ(h)}{a}}\right)^2}.
\end{aligned}$$

(b) We show that $X(h)^\top BX(h) \leqslant k^2 u(h)^\top Bu(h)$ implies

$$u(h)^\top Bu(h) \leqslant \frac{Z(h)^\top BZ(h)}{(1-k)^2}$$

for $k \in (0,1)$. We have:

$$\begin{aligned}
&X(h)^\top BX(h) \leqslant k^2 u(h)^\top Bu(h) \\
&\Rightarrow (CX(h))^\top (CX(h)) \leqslant k^2 (Cu(h))^\top (C(u(h)) \\
&\Rightarrow \|CX(h)\|^2 \leqslant k^2 \|Cu(h)\|^2
\end{aligned}$$

$$
\begin{aligned}
&\Rightarrow \|CX(h)\| \leqslant k\|Cu(h)\| \\
&\Rightarrow \|Cu(h)\| - \|CZ(h)\| \leqslant \|C(u(h)+Z(h))\| \leqslant k\|Cu(h)\| \\
&\Rightarrow -\|CZ(h)\| \leqslant (k-1)\|Cu(h)\|.
\end{aligned}
$$

We have used that $B = C^2 = C^\top C$. We can now easily prove the desired inequality:

$$
\begin{aligned}
X(h)^\top BX(h) \leqslant k^2 u(h)^\top Bu(h) &\Rightarrow \|CZ(h)\| \geqslant (1-k)\|Cu(h)\| \\
&\Rightarrow \|CZ(h)\|^2 \geqslant (1-k)^2\|Cu(h)\|^2 \\
&\Rightarrow \frac{Z(h)^\top BZ(h)}{(1-k)^2} \geqslant u(h)^\top Bu(h).
\end{aligned}
$$

We have for $X(h)^\top BX(h) \leqslant k^2 u(h)^\top Bu(h)$:

$$
\begin{aligned}
\frac{a+u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} &= \frac{a}{a+X(h)^\top BX(h)} + \frac{u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} \\
&\leqslant 1 + \frac{u(h)^\top Bu(h)}{a} \leqslant 1 + \frac{Z(h)^\top BZ(h)}{(1-k)^2 a}.
\end{aligned}
$$

We have for $X(h)^\top BX(h) > k^2 u(h)^\top Bu(h)$:

$$
\frac{a+u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} < \frac{a+u(h)^\top Bu(h)}{a+k^2 u(h)^\top Bu(h)} < \frac{1}{k^2}.
$$

We combine both inequalities using that $k \in (0,1)$ and get:

$$
\frac{a+u(h)^\top Bu(h)}{a+X(h)^\top BX(h)} \leqslant \frac{1}{k^2} + \frac{Z(h)^\top BZ(h)}{(k-1)^2 a}.
$$

(c) We use below that $B = C^2 = C^\top C$ and that there is a positive constant $k'$ with:

$$
\|Cx\| \leqslant k'\|x\|, \quad x \in \mathbb{R}^d.
$$

Choose for instance $k'$ as the largest eigenvalue of $C$. We use also that for any matrix $M \in \mathbb{R}^{d\times d}$ there is a positive constant $k_M$ such that:

$$
\|Mg(t)\| \leqslant k_M, \quad t \in [0,T].
$$

This holds since every $g_i$ is bounded as a continuous function on $[0,T]$. We have for $a > 0$ and $k \in (0,1)$:

$$
\begin{aligned}
&\Bigg(\int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \Big[\|Cg(t)\| \, \|CD_{x_2}X(h)\| \, \|CD_{x_1}X(h)\| \\
&\quad \times \frac{|f(x_1,x_2;t)|}{a+X(h)^\top BX(h)}\Big] dx_1 dx_2 dt\Bigg)^2
\end{aligned}
$$

$$= \frac{1}{\left(a+u(h)^\top B u(h)\right)^2} \left( \int_0^T \int_0^T \int_0^{x_2} |f(x_1,x_2;t)| \, \|Cg(t)\| \right.$$
$$\left. \times\, \mathbb{E}_u \left[ \|CD_{x_2}X(h)\| \, \|CD_{x_1}X(h)\| \, \frac{a+u(h)^\top B u(h)}{a+X(h)^\top B X(h)} \right] dx_1 dx_2 dt \right)^2$$
$$\leqslant \frac{1}{\left(a+u(h)^\top B u(h)\right)^2} \left( \int_0^T \int_0^T \int_0^{x_2} |f(x_1,x_2;t)| \, \|Cg(t)\| \right.$$
$$\left. \times\, \mathbb{E}_u \left[ \|CD_{x_2}X(h)\| \, \|CD_{x_1}X(h)\| \, \left( \frac{1}{k^2} + \frac{Z(h)^\top B Z(h)}{a(1-k)^2} \right) \right] dx_1 dx_2 dt \right)^2 .$$

With the Cauchy-Schwarz inequality, the inequalities $\|Cg(t)\| \leqslant k_C$ and $\|Cx\| \leqslant k'\|x\|$ for positive real numbers $k_C$ and $k'$, we prove that the expression in parenthesis is finite. We notice that for $u = 0$, we have $X(h) = u(h) + Z(h) = Z(h)$, thus:

$$\left( \int_0^T \int_0^T \int_0^{x_2} |f(x_1,x_2;t)| \, \|Cg(t)\| \right.$$
$$\left. \times\, \mathbb{E}_u \left[ \|CD_{x_2}X(h)\| \, \|CD_{x_1}X(h)\| \, \left( \frac{1}{k^2} + \frac{Z(h)^\top B Z(h)}{a(1-k)^2} \right) \right] dx_1 dx_2 dt \right)^2$$
$$\leqslant \int_0^T\!\!\int_0^T\!\!\int_0^{x_2} \mathbb{E}_u \left[ f(x_1,x_2;t)^2 \, \|Cg(t)\|^2 \, \left( \frac{1}{k^2} + \frac{Z(h)^\top B Z(h)}{a(1-k)^2} \right)^2 \right] dx_1 dx_2 dt$$
$$\times \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \|CD_{x_2}X(h)\|^2 \, \|CD_{x_1}X(h)\|^2 \right] dx_1 dx_2 dt$$
$$\leqslant k_C^2 \, k'^4 \int_0^T \int_0^T \int_0^{x_2} f(x_1,x_2;t)^2 dx_1 dx_2 dt \; \mathbb{E}_u \left[ \left( \frac{1}{k^2} + \frac{Z(h)^\top B Z(h)}{a(1-k)^2} \right)^2 \right]$$
$$\times \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_u \left[ \|D_{x_2}X(h)\|^2 \, \|D_{x_1}X(h)\|^2 \right] dx_1 dx_2 dt$$
$$= k_C^2 \, k'^4 \int_0^T \int_0^T \int_0^{x_2} f(x_1,x_2;t)^2 dx_1 dx_2 dt \; \mathbb{E}_0 \left[ \left( \frac{1}{k^2} + \frac{X(h)^\top B X(h)}{a(k-1)^2} \right)^2 \right]$$
$$\times \int_0^T \int_0^T \int_0^{x_2} \mathbb{E}_0 \left[ \|D_{x_2}X(h)\|^2 \, \|D_{x_1}X(h)\|^2 \right] dx_1 dx_2 dt < \infty.$$

For the last inequality, we need that $\int_0^T \mathbb{E}_u[Z_t^2]dt < \infty$ (this was one of the assumptions) and

$$\int_0^T \int_0^T \int_0^T \mathbb{E}_0 \left[ \left(D_{x_1}X(h_i)\right)^2 \, \left(D_{x_2}X(h_j)\right)^2 \right] dx_1 dx_2 dt < \infty$$

for every $i, j = 1, \dots, d$. We can prove this last inequality with:

$$\mathbb{E}_0 \left[ \left(I_1(q)\, I_1(l)\right)^2 \right] \leqslant 3\, \mathbb{E}_0 \left[ I_1(q)^2 \right] \mathbb{E}_0 \left[ I_1(l)^2 \right].$$

We have:

$$\begin{aligned}
&\int_0^T \int_0^T \int_0^T \mathbb{E}_0 \left[ (D_{x_1} X(h_i))^2 \ (D_{x_2} X(h_j))^2 \right] dx_1 dx_2 dt \\
&\leqslant \int_0^T \int_0^T \int_0^T 3\, \mathbb{E}_0 \left[ (D_{x_1} X(h_i))^2 \right] \mathbb{E}_0 \left[ (D_{x_2} X(h_j))^2 \right] dx_1 dx_2 dt \\
&= 3 \int_0^T \int_0^T \mathbb{E}_0 \left[ (D_{x_1} X(h_i))^2 \right] dx_1 \int_0^T \mathbb{E}_0 \left[ (D_{x_2} X(h_j))^2 \right] dx_2 dt \\
&= 3 \int_0^T 2\, \mathbb{E}_0[Z(h_i)^2]\, 2\, \mathbb{E}_0[Z(h_j)^2] dt < \infty.
\end{aligned}$$

This completes the proof. □

**Theorem 9.2.** *Consider an arbitrary invertible symmetric matrix $M \in \mathbb{R}^{d\times d}$ with inverse $M^{-1}$. We write $*$ for the Hadamard product of two matrices. We prove that the sum of all elements of $M * M^{-1}$ is always $d$:*

$$\sum_{i,j=1}^{d} M_{i,j}\, (M^{-1})_{i,j} = d.$$

*Proof.* We use the notation $M = M^\top = (u_1, \ldots, u_d)$ and $M^{-1} = (v_1, \ldots, v_d)$ where $u_1, \ldots, u_d, v_1, \ldots, v_d$ are $d$-dimensional vectors. Since $M\, M^{-1} = I$, we have $u_i^\top v_j = \delta_{i,j}$. Since $M$ is symmetric, $M^{-1}$ is symmetric as well and we have $M^{-1} = \left(M^{-1}\right)^\top = (v_1, \ldots, v_d)$. We have then:

$$\sum_{i,j=1}^{d} M_{i,j} \left(M^{-1}\right)_{i,j} = u_1^\top v_1 + \ldots + u_d^\top v_d = 1 + \ldots + 1 = d.$$

This concludes the proof. □

### *9.4. Hölder continuity*

We prove that the processes considered in Section 5.2 have Hölder continuous paths.

**Theorem 9.3.** *The process $Z$ defined in Section 5.2 has a version with $k$-Hölder continuous paths for every $k \in (0, 1/q_2)$.*

*Proof.* The proof is divided in two steps:

(a) We consider a symmetric function $f \in L^2([0,T]^2, \lambda_T^2)$ and prove for every $n \geqslant 1$ :

$$\mathbb{E}_u \left[ (I_2(f))^{2n} \right] \leqslant \tau(n)\ \mathbb{E}_u \left[ (I_2(f))^2 \right]^n, \tag{9.11}$$

where $\tau(n)$ is a constant that is independent of $f$.

(b) We use this result to prove that $(Z_t)_{t\in[0,T]}$ has a version with $k$-Hölder continuous paths for every $k \in (0, 1/q_2)$.

We now prove these statements.

(a) We use the well known multiplication formula for multiple Wiener integrals:

$$I_2(f)^2 = \sum_{r_1=0}^{2} r_1! \binom{2}{r_1}^2 I_{4-2r_1}(f\tilde{\otimes}_{r_1} f) = \sum_{r_1=0}^{2} \alpha_2(r_1) I_{4-2r_1}(f\tilde{\otimes}_{r_1} f).$$

We write $f\tilde{\otimes}_{r_1} f$ for the symmetrization of the contraction $f\otimes_{r_1} f$ and

$$\alpha_2(r_1) := r_1! \binom{2}{r_1}^2.$$

Using the multiplication formula once again, we find for $I_2(f)^3$:

$$\sum_{r_1=0}^{2} \sum_{r_2=0}^{2\wedge(4-2r_1)} r_1! r_2! \binom{2}{r_1}^2 \binom{2}{r_2} \binom{4-2r_1}{r_2} I_{6-2(r_1+r_2)}((f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2} f)$$
$$= \sum_{r_1=0}^{2} \sum_{r_2=0}^{2\wedge(4-2r_1)} \alpha_3(r_1, r_2) I_{6-2(r_1+r_2)}((f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2} f).$$

We have used the notation

$$\alpha_3(r_1, r_2) := r_1! r_2! \binom{2}{r_1}^2 \binom{2}{r_2} \binom{4-2r_1}{r_2}.$$

Using the multiplication formula $n-1$ times, we find similarly for $I_2(f)^n$:

$$\sum_{r_1=0}^{2} \sum_{r_2=0}^{2\wedge(4-2r_1)} \cdots \sum_{r_{n-1}=0}^{2\wedge(2(n-1)-2(r_1+\ldots+r_{n-2}))} \alpha_n(r_1, \ldots, r_{n-1})$$
$$\times I_{2n-2(r_1+\ldots+r_{n-1})}((f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2} \ldots)$$
$$= \sum_{(r_1,\ldots,r_{n-1})^\top \in A_n} \alpha_n(r_1, \ldots, r_{n-1}) I_{2n-2(r_1+\ldots+r_{n-1})}((f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2} \ldots).$$

For the ease of notation we write $A_n$ for the set of $r^\top := (r_1, \ldots, r_{n-1})^\top \in \mathbb{N}^{n-1}$ verifying:

$$0 \leqslant r_1 \leqslant 2, 0 \leqslant r_2 \leqslant 2\wedge(4-2r_1), \ldots, 0 \leqslant r_{n-1} \leqslant 2\wedge\left(2(n-1) - 2\sum_{i=1}^{n-2} r_i\right).$$

We have with $(x_1 + \ldots + x_m)^2 \leqslant m\big(x_1^2 + \ldots + x_m^2\big)$:

$$\mathbb{E}_u\left[\left(I_2(f)^n\right)^2\right]$$

$$
\begin{aligned}
&\leqslant |A_n| \sum_{r^\top \in A_n} \alpha_n(r_1,\ldots,r_{n-1})^2 \; \mathbb{E}_u\left[I_{2n-2(r_1+\ldots+r_{n-1})}((f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2}\ldots)^2\right] \\
&= |A_n| \sum_{r^\top \in A_n} \alpha_n(r_1,\ldots,r_{n-1})^2 \,[2n-2(r_1+\ldots+r_{n-1})]! \\
&\quad\times \|(f\tilde{\otimes}_{r_1} f)\tilde{\otimes}_{r_2}\ldots\|^2_{L^2([0,T]^{2n-2(r_1+\ldots+r_{n-1})},\lambda_T^{2n-2(r_1+\ldots+r_{n-1})})} \\
&\leqslant |A_n| \sum_{r^\top \in A_n} \alpha_n(r_1,\ldots,r_{n-1})^2 \left(2n - 2\sum_{i=1}^{n-1} r_i\right)! \left(\|f\|^2_{L^2([0,T]^2,\lambda_T^2)}\right)^n .
\end{aligned}
$$

For symmetric functions $f$ and $g$ with

$$
f \in L^2([0,T]^{k_1}, \lambda_T^{k_1}) \; ; \; g \in L^2([0,T]^{k_2}, \lambda_T^{k_2})
$$

and $r \leqslant \min\{k_1,k_2\}$, we have used the following inequality:

$$
\|f\tilde{\otimes}_r g\|^2_{L^2([0,T]^{k_1+k_2-2r},\lambda_T^{k_1+k_2-2r})} \leqslant \|f\|^2_{L^2([0,T]^{k_1},\lambda_T^{k_1})} \|g\|^2_{L^2([0,T]^{k_2},\lambda_T^{k_2})}.
$$

With constants $B_n := \sum_{(r_1,\ldots,r_{n-1})^\top \in A_n} \alpha_n(r_1,\ldots,r_{n-1})^2 \,[2n-2(r_1+\ldots+r_{n-1})]!$ and $\tau(n) := 2^{-n}\,|A_n|\,B_n$, we get:

$$
\begin{aligned}
\mathbb{E}_u\left[(I_2(f)^n)^2\right] &\leqslant |A_n|\,B_n \left(\|f\|^2_{L^2([0,T]^2,\lambda_T^2)}\right)^n = |A_n|\,B_n \left(\frac{\mathbb{E}_u\left[I_2(f)^2\right]}{2}\right)^n \\
&= \tau(n)\left(\mathbb{E}_u\left[I_2(f)^2\right]\right)^n .
\end{aligned}
$$

(b) We use Kolmogorov's continuity theorem to prove that $(Z_t)_{t\in[0,T]}$ has a version that is locally $k$-Hölder continuous for every $k \in (0, 1/q_2)$. We consider $n \in \mathbb{N}_{\geqslant 2}$. We have for every $0 \leqslant s \leqslant t \leqslant T$ with inequality (9.11):

$$
\begin{aligned}
&\mathbb{E}_u\left[|Z_t - Z_s|^{2n}\right] \\
&= \mathbb{E}_u\left[\left(\int_0^T\int_0^T (f(x_1,x_2;t) - f(x_1,x_2;s))\,dW^u_{x_1}dW^u_{x_2}\right)^{2n}\right] \\
&\leqslant \tau(n)\,\mathbb{E}_u\left[\left(\int_0^T\int_0^T (f(x_1,x_2;t) - f(x_1,x_2;s))\,dW^u_{x_1}dW^u_{x_2}\right)^2\right]^n \\
&= \tau(n)\left(2\int_0^T\int_0^T (f(x_1,x_2;t) - f(x_1,x_2;s))^2\,dx_1dx_2\right)^n \\
&= \tau(n)\left(2\int_0^T\int_0^T \left(\int_s^t k(x_1,x_2;a)da\right)^2 dx_1dx_2\right)^n .
\end{aligned}
$$

This yields for $\mathbb{E}_u\left[|Z_t - Z_s|^{2n}\right]$:

$$
\mathbb{E}_u\left[|Z_t - Z_s|^{2n}\right]
$$

$$
\begin{aligned}
&\leqslant \tau(n)\left(2\int_0^T\int_0^T\left(\int_s^t k(x_1,x_2;a)da\right)\left(\int_s^t k(x_1,x_2;b)db\right)dx_1dx_2\right)^n\\
&= \tau(n)\left[\int_s^t\int_s^t\left(2\int_0^T\int_0^T k(x_1,x_2;a)k(x_1,x_2;b)dx_1dx_2\right)dadb\right]^n\\
&= \tau(n)\left[\int_s^t\int_s^t \gamma(a,b)dadb\right]^n.
\end{aligned}
$$

We find with the Cauchy-Schwarz inequality:

$$
\begin{aligned}
\mathbb{E}_u\left[|Z_t-Z_s|^{2n}\right] &\leqslant \tau(n)\left[\left(\int_s^t\int_s^t 1dadb\right)^{1/q_2}\left(\int_s^t\int_s^t|\gamma(a,b)|^{q_1}dadb\right)^{1/q_1}\right]^n\\
&\leqslant \tau(n)\left[(t-s)^{2/q_2}\left(\int_0^T\int_0^T|\gamma(a,b)|^{q_1}dadb\right)^{1/q_1}\right]^n\\
&= \tau(n)\left(\int_0^T\int_0^T|\gamma(a,b)|^{q_1}dadb\right)^{n/q_1}(t-s)^{(2n)/q_2}.
\end{aligned}
$$

Kolmogorov's continuity theorem implies that $(Z_t)_{t\in[0,T]}$ has a version that is locally $k$-Hölder continuous for every $k\in(0,(2n/q_2-1)/2n)$. This holds for every $n>q_2/2$ and thus $(Z_t)_{t\in[0,T]}$ has a version that is locally $k$-Hölder continuous for every $k\in(0,1/q_2)$. Since $[0,T]$ is compact, it follows that $(Z_t)_{t\in[0,T]}$ has a version that is $k$-Hölder continuous for every $k\in(0,1/q_2)$.

□

### *9.5. Proof of Eq.* (7.7)

$A(\epsilon,T)\,\epsilon^2T = A(1,1)$.

(a) We prove that $\tilde{g}_i(\cdot,\epsilon,T)$ is right-continuous in $t=0$. We use the series expansion of $|t-t_i|^{2H}$ in a neighbourhood of $t=0$. We have for $t>0$ sufficiently small:

$$
\begin{aligned}
\tilde{g}_i(t,\epsilon,T) &= \frac{\epsilon^2}{2t_i^H\,t^H}\left(t^{2H}+t_i^{2H}-|t-t_i|^{2H}\right)\\
&= \frac{\epsilon^2}{2t_i^H\,t^H}\left(t^{2H}+t_i^{2H}-t_i^{2H}+2H\,t_i^{2H-1}\,t+\mathcal{O}(t^2)\right)\to 0 \text{ for } t\downarrow 0.
\end{aligned}
$$

Hence $\tilde{g}(\cdot,\epsilon,T)$ is right-continuous in $t=0$.

(b) We transform the condition

$$
2\int_0^T\mathbb{E}_u\left[(\tilde{X}(t,\epsilon,T)-u_t)\,\tilde{\xi}_a(t,\epsilon,T)\right]dt-\int_0^T\mathbb{E}_u[\tilde{\xi}_a(t,\epsilon,T)^2]dt>0
$$

using the properties of the Rosenblatt process. The Rosenblatt process $(Z_t)_{t\in[0,T]}$ with Hurst parameter $H$ is a self-similar process and for every $c>0$, we have

$$(Z_{ct}) \overset{(d)}{=} (c^H Z_t),$$

where $\overset{(d)}{=}$ means equivalence of all finite dimensional distributions (see [26]). We consider $s=t/T$ and we have as a direct consequence that: $(\tilde{Z}(t,\epsilon,T)) \overset{(d)}{=} (\epsilon\tilde{Z}(s,1,1))$:

$$\begin{aligned}(\tilde{Z}(t,\epsilon,T)) &\overset{(d)}{=} (\epsilon\, t^{-H} Z_t) \overset{(d)}{=} (\epsilon\, t^{-H} Z_{sT}) \overset{(d)}{=} (\epsilon\, t^{-H} T^H Z_s) \overset{(d)}{=} (\epsilon\, s^{-H} Z_s)\\ &\overset{(d)}{=} (\epsilon\, \tilde{Z}(s,1,1)).\end{aligned}$$

We define $\tilde{u}:[0,1]\to\mathbb{R}$ by $\tilde{u}_x=\epsilon^{-1}\, u_{xT}$. For $s=t/T$, we have the relations:

$$\tilde{g}_i(t,\epsilon,T)=\epsilon^2\tilde{g}_i(s,1,1), \quad \tilde{B}(\epsilon,T)=\frac{1}{T\epsilon^4}\tilde{B}(1,1). \tag{9.12}$$

We have with $s=t/T$ and $s_i=t_i/T=i/d$:

$$\begin{aligned}\tilde{g}_i(t,\epsilon,T) &= \mathrm{Cov}_u(\tilde{X}(t_i,\epsilon,T),\tilde{X}(t,\epsilon,T)) = \mathbb{E}_u\left[\tilde{Z}\,(t_i,\epsilon,T)\ \tilde{Z}(sT,\epsilon,T)\right]\\ &= \epsilon^2\,\mathbb{E}_u\left[\tilde{Z}\,(s_i,1,1)\ \tilde{Z}(s,1,1\right] = \epsilon^2\tilde{g}_i(s,1,1);\\ \tilde{B}(\epsilon,T) &= \left(\int_0^T \tilde{g}(t,\epsilon,T)\tilde{g}(t,\epsilon,T)^\top dt\right)^{-1}\\ &= \left(\epsilon^4\int_0^1 \tilde{g}(s,1,1)\tilde{g}(s,1,1)^\top\, T ds\right)^{-1} = T^{-1}\epsilon^{-4}\tilde{B}(1,1).\end{aligned}$$

(c) We know that, for $a>A(\epsilon,T)$:

$$2\int_0^T \mathbb{E}_u\left[(\tilde{X}(t,\epsilon,T)-u_t)\,\tilde{\xi}_a(t,\epsilon,T)\right]dt - \int_0^T \mathbb{E}_u[\tilde{\xi}_a(t,\epsilon,T)^2]dt > 0.$$

On the other hand we use Eq. (9.12) and transform the left-hand side of the inequality above. We have:

$$\begin{aligned}&2\int_0^T \mathbb{E}_u\left[(\tilde{X}(t,\epsilon,T)-u_t)\,\tilde{\xi}_a(t,\epsilon,T)\right]dt\\ &= 2\int_0^T \mathbb{E}_u\Bigg[\tilde{Z}(t,\epsilon,T)\\ &\quad\times \frac{\tilde{g}(t,\epsilon,T)^\top\tilde{B}(\epsilon,T)\left(u_{t_i}+\tilde{Z}(t_i,\epsilon,T)\right)_i}{a+\left(u_{t_i}+\tilde{Z}(t_i,\epsilon,T)\right)_i^\top\tilde{B}(\epsilon,T)\left(u_{t_i}+\tilde{Z}(t_i,\epsilon,T)\right)_i}\Bigg]dt\\ &= 2T\int_0^1 \mathbb{E}_u\Big[\tilde{Z}(sT,\epsilon,T)\end{aligned}$$

$$\times \left. \frac{\epsilon^2 \, \tilde{g}(s,1,1)^\top \tilde{B}(\epsilon,T) \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i}{a + \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i^\top \tilde{B}(\epsilon,T) \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i} \right] ds$$

$$= 2T \int_0^1 \mathbb{E}_u \Big[ \epsilon \, \tilde{Z}(s,1,1)$$

$$\times \left. \frac{\epsilon^2 \tilde{g}(s,1,1)^\top T^{-1} \epsilon^{-4} \tilde{B}(1,1) \, \epsilon \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i}{a + \epsilon \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top T^{-1} \epsilon^{-4} \tilde{B}(1,1) \epsilon \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i} \right] ds$$

$$= 2T\epsilon^2 \int_0^1 \mathbb{E}_u \Big[ \tilde{Z}(s,1,1)$$

$$\times \left. \frac{\tilde{g}(s,1,1)^\top \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i}{aT\epsilon^2 + \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i} \right] ds$$

$$= 2T\epsilon^2 \int_0^1 \mathbb{E}_{\tilde{u}} \left[ \left( \tilde{X}(s,1,1) - \tilde{u}_s \right) \tilde{\xi}_{aT\epsilon^2}(s,1,1) \right] ds,$$

and similarly:

$$\int_0^T \mathbb{E}_u \left[ \tilde{\xi}_a(t,\epsilon,T)^2 \right] dt$$

$$= \mathbb{E}_u \left[ \frac{\left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i^\top \tilde{B}(\epsilon,T) \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i}{\left( a + \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i^\top \tilde{B}(\epsilon,T) \left( u_{t_i} + \tilde{Z}(t_i,\epsilon,T) \right)_i \right)^2} \right]$$

$$= \mathbb{E}_u \left[ \frac{\epsilon^2 \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top T^{-1} \epsilon^{-4} \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i}{\left( a + \epsilon^2 \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top T^{-1} \epsilon^{-4} \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i \right)^2} \right]$$

$$= T\epsilon^2 \, \mathbb{E}_u \left[ \frac{\left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i}{\left( aT\epsilon^2 + \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i^\top \tilde{B}(1,1) \left( \tilde{u}_{\frac{i}{d}} + \tilde{Z}(\frac{i}{d},1,1) \right)_i \right)^2} \right]$$

$$= T\epsilon^2 \int_0^1 \mathbb{E}_{\tilde{u}} \left[ \tilde{\xi}_{aT\epsilon^2}(t,1,1)^2 \right] dt.$$

Combining these results, we find:

$$2 \int_0^T \mathbb{E}_u \left[ (\tilde{X}(t,\epsilon,T) - u_t) \, \tilde{\xi}_a(t,\epsilon,T) \right] dt - \int_0^T \mathbb{E}_u [\tilde{\xi}_a(t,\epsilon,T)^2] dt$$

$$= T\epsilon^2 \Big\{ 2 \int_0^1 \mathbb{E}_{\tilde{u}} \left[ (\tilde{X}(t,1,1) - \tilde{u}_t) \, \tilde{\xi}_{aT\epsilon^2}(t,1,1) \right] dt$$

$$- \int_0^1 \mathbb{E}_{\tilde{u}}[\tilde{\xi}_{aT\epsilon^2}(t,1,1)^2]dt \Bigg\} . \tag{9.13}$$

(d) We use Eq. (9.13) to prove Eq. (7.7) The expression on the left-hand side of Eq. (9.13) is positive for $a > A(\epsilon, T)$ whereas the expression on the right-hand side of the equation is positive if $aT\epsilon^2 > A(1,1)$. Assuming that $A(1,1)$ and $A(\epsilon, T)$ are both chosen as the infimum of all possible values, we have:

$$A(\epsilon, T) = \frac{A(1,1)}{T\epsilon^2}.$$

This completes the proof. □

**Acknowledgements**

I would like to thank two anonymous referees for their comments and remarks, as well as Prof. Dr. G. Peccati and Prof. Dr. A. Thalmaier for their time and many interesting discussions.

**References**

[1] Apostol, T. M. (1974). *Mathematical analysis*, second ed. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont. MR0344384 (49 ##9123)

[2] Brandwein, A. C. (1979). Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multivariate Anal.* **9** 579–588. MR556913 (81f:62012)

[3] Brandwein, A. C. and Strawderman, W. E. (1990). Stein estimation: the spherically symmetric case. *Statist. Sci.* **5** 356–369. MR1080957 (92a:62114)

[4] Brandwein, A. C. and Strawderman, W. E. (1991). Generalizations of James-Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650. MR1126343 (92i:62137)

[5] Brandwein, A. C. and Strawderman, W. E. (2012). Stein estimation for spherically symmetric distributions: recent developments. *Statist. Sci.* **27** 11–23. MR2953492

[6] Decreusefond, L. and Üstünel, A. S. (1999). Stochastic analysis of the fractional Brownian motion. *Potential Anal.* **10** 177–214. MR1677455 (2000b:60133)

[7] Di Nunno, G., Øksendal, B. and Proske, F. (2009). *Malliavin calculus for Lévy processes with applications to finance*. *Universitext*. Springer-Verlag, Berlin. MR2460554 (2010f:60001)

[8] Dieudonné, J. (1969). *Foundations of modern analysis*. Academic Press, New York-London Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I. MR0349288 (50 ##1782)

[9] Es-Sebaiy, K., Ouassou, I. and Ouknine, Y. (2009). Estimation of the drift of fractional Brownian motion. *Statist. Probab. Lett.* **79** 1647–1653. MR2547425 (2010j:60094)

[10] Evans, S. N. and Stark, P. B. (1996). Shrinkage estimators, Skorokhod's problem and stochastic integration by parts. *Ann. Statist.* **24** 809–815. MR1394989 (98f:62016)

[11] Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. John Wiley & Sons, Inc., New York. MR1100924 (92i:60091)

[12] James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, Calif. MR0133191 (24 ##A3025)

[13] Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR1639875 (99g:62025)

[14] Maejima, M. and Tudor, C. A. (2007). Wiener integrals with respect to the Hermite process and a non-central limit theorem. *Stoch. Anal. Appl.* **25** 1043–1056. MR2352951 (2008h:60209)

[15] Meyer, F. (1890). Über die höheren Ableitungen eines Quotienten zweier Functionen. *Monatsh. Math. Phys.* **1** 33–38. MR1546141

[16] Nourdin, I. and Peccati, G. (2012). *Normal approximations with Malliavin calculus*. *Cambridge Tracts in Mathematics* **192**. Cambridge University Press, Cambridge From Stein's method to universality. MR2962301

[17] Nualart, D. (2006). *The Malliavin calculus and related topics*, second ed. *Probability and its Applications (New York)*. Springer-Verlag, Berlin. MR2200233 (2006j:60004)

[18] Nualart, D. and Ouknine, Y. (2002). Regularization of differential equations by fractional noise. *Stochastic Process. Appl.* **102** 103–116. MR1934157 (2004b:60151)

[19] Pipiras, V. and Taqqu, M. S. (2000). Integration questions related to fractional Brownian motion. *Probab. Theory Related Fields* **118** 251–291. MR1790083 (2002c:60091)

[20] Privault, N. and Réveillac, A. (2008). Stein estimation for the drift of Gaussian processes using the Malliavin calculus. *Ann. Statist.* **36** 2531–2550. MR2458197 (2009k:62092)

[21] Privault, N. and Réveillac, A. (2009). Stein estimation of Poisson process intensities. *Stat. Inference Stoch. Process.* **12** 37–53. MR2486115 (2010c:62117)

[22] Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss. *Ann. Statist.* **12** 322–335. MR733517 (85k:62013)

[23] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I* 197–206. University of California Press, Berkeley and Los Angeles. MR0084922 (18,948c)

[24] STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR630098 (83a:62080)

[25] STEIN, E. M. and SHAKARCHI, R. (2005). *Real analysis. Princeton Lectures in Analysis, III.* Princeton University Press, Princeton, NJ Measure theory, integration, and Hilbert spaces. MR2129625 (2005k:28024)

[26] TUDOR, C. A. (2008). Analysis of the Rosenblatt process. *ESAIM Probab. Stat.* **12** 230–257. MR2374640 (2008m:60067)

[27] ŠREMR, J. (2010). Absolutely continuous functions of two variables in the sense of Carathéodory. *Electron. J. Differential Equations* No. 154, 11. MR2740595 (2011m:26016)