

Reiß, Markus

**Working Paper**

## Minimax rates for nonparametric estimation of the drift functional in affine stochastic delay equations

SFB 373 Discussion Paper, No. 2000,77

**Provided in Cooperation with:**

Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

*Suggested Citation:* Reiß, Markus (2000) : Minimax rates for nonparametric estimation of the drift functional in affine stochastic delay equations, SFB 373 Discussion Paper, No. 2000,77, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10048034>

This Version is available at:

<https://hdl.handle.net/10419/62267>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Minimax rates for nonparametric estimation of the drift functional in affine stochastic delay equations

Markus Reiß\*  
Institut für Mathematik  
Humboldt-Universität Berlin  
reiss@mathematik.hu-berlin.de

September 7, 2000

## Abstract

Let  $X$  be a stationary process satisfying the stochastic differential equation with time delay

$$dX(t) = \left( \int_{-r}^0 X(t+s)g(s) ds \right) dt + dW(t), \quad t \geq 0.$$

The function  $g \in L^2([-r, 0])$  is estimated nonparametrically from the continuous observation of a trajectory of  $X$  up to time  $T > 0$ . The estimation problem is transformed into an illposed inverse problem with stochastically perturbed operator and data. The estimator is then constructed by the Ritz-Galerkin projection method. The  $L^2$ -risk of the estimator is asymptotically for  $T \rightarrow \infty$  of order  $T^{-\frac{s}{2s+3}}$  in a minimax sense, where  $g$  is assumed to lie in some Sobolev ball of order  $s$ . This rate is shown to be optimal for the estimation problem.

**Keywords:** stochastic delay equations; drift estimation; stationary Gaussian process; illposed problem; Ritz-Galerkin method; minimax rates

**AMS(2000) subject classification:** 62G20, 62M09, 65R32, 34K50

## 1 Introduction

Delay equations, also known as functional differential equations, arise when the dynamics of a system not only depend on its present, but also on its past state. Many “real life phenomena” exhibit such a memory effect, e.g. population growth with an age-dependent fertility. From a mathematical point of view

---

\*This paper was printed using funds made available by the Deutsche Forschungsgemeinschaft (SFB 373). Helpful comments by Matthias Fengler on the final draft are gratefully acknowledged.

the deterministic, one-dimensional, linear and autonomous delay equations with finite memory form the fundamental and best understood subclass among these equations. They are of the following form:

$$\begin{aligned} \dot{x}(t) &= \int_{-r}^0 x(t+s) da(s), \quad t \geq 0, \\ x(t) &= f(t), \quad t \in [-r, 0], \end{aligned} \quad (1.1)$$

with a positive constant  $r$ , an element  $a$  of  $M([-r, 0])$ , the space of signed, finite measures on  $[-r, 0]$ , and an initial function  $f \in C([-r, 0])$ , the space of continuous functions on  $[-r, 0]$ .

A basic model for such a dynamic corrupted by stochastic noise is obtained by adding white noise. With a standard Brownian motion  $(W(t), t \geq 0)$  and a – possibly random – initial function  $F$  the following stochastic differential equation will be considered

$$\begin{aligned} dX(t) &= \left( \int_{-r}^0 X(t+s) da(s) \right) dt + dW(t), \quad t \geq 0, \\ X(t) &= F(t), \quad t \in [-r, 0]. \end{aligned} \quad (1.2)$$

In view of applications this may be interpreted as a time continuous analogue of autoregressive models.

Restricting to stationary processes and assuming absolute continuity of the measure  $a$  with respect to Lebesgue measure (i.e.  $da(s) = g(s)ds$ ), we aim at an estimator of  $g$  given the continuous observation of one trajectory up to time  $T$ . The asymptotic behaviour of this estimator for  $T \rightarrow \infty$  is examined. It is shown that for densities  $g$  in a Sobolev ball of smoothness  $s$  the  $L^2$ -risk of the estimator has order  $T^{-\frac{2s}{2s+3}}$ , which is worse than the classical  $T^{-\frac{2s}{2s+1}}$ -rate for the corresponding white noise model  $dY(t) = g(t)dt + T^{-\frac{1}{2}}dW(t)$  (cf. [IbrHas81]). However, this rate is optimal in a minimax sense.

Wherever it is possible, the results are stated for the general case of signed measures  $a$ , which should simplify the generalisation from measures with densities to other parameter sets, e.g. sums of measures with densities and point measures. Our estimator will be derived from the quantities  $Q_T$  and  $b_T$ , which depend on  $(X(t), -r \leq t \leq T)$  and lead to approximations of the covariance operator  $Q$  of the stationary solution and the image  $Qa$  of the true, but unknown parameter  $a$  in (1.2).

**1.1. Definition.** For the process  $(X(t), -r \leq t \leq T)$  in (1.2) define

$$q_T(u, v) := \int_0^T X(t+u)X(t+v) dt, \quad u, v \in [-r, 0], \quad (1.3)$$

$$(Q_T \mu)(s) := \int_{-r}^0 q_T(s, u) d\mu(u), \quad \mu \in M([-r, 0]), s \in [-r, 0], \quad (1.4)$$

$$b_T(s) := \int_0^T X(t+s) dX(t), \quad s \in [-r, 0]. \quad (1.5)$$

The covariance operator  $Q = Q_a : M([-r, 0]) \rightarrow C([-r, 0])$  of a stationary solution  $(X(t), -r \leq t \leq 0)$  is implicitly defined by the bilinear form

$$\langle Q_a \mu, \nu \rangle = \mathbb{E}_a \left[ \int_{-r}^0 X(u+r) d\mu(u) \int_{-r}^0 X(v+r) d\nu(v) \right], \quad \mu, \nu \in M([-r, 0])$$

or explicitly given in terms of the covariance function  $q_a(t) = \mathbb{E}_a[X(0)X(|t|)]$

$$Q_a \mu(s) := \int_{-r}^0 q_a(s-u) d\mu(u), \quad \mu \in M([-r, 0]), s \in [-r, 0].$$

Basic results on stochastic delay equations, its stationary solutions and the continuous dependence on the parameters are presented in the next section. The rate of convergence of  $\frac{1}{T}q_T$  to  $q_a$  and of  $\frac{1}{T}b_T$  to  $Q_a a$  is examined in section 3. By a nice variant of the Kolmogorov continuity test via Sobolev embeddings a  $\frac{1}{\sqrt{T}}$ -rate of convergence in function space norms ( $C^\alpha$  denotes the Hölder space of order  $\alpha$ ) is obtained in Corollaries 3.2 and 3.4:

*For  $\alpha < \frac{1}{2}$  and  $p \geq 1$  there is a constant  $C_{\alpha p}$  such that for all  $T > 0$*

$$\mathbb{E}_a[\|T^{-1}q_T(u, v) - q_a(u-v)\|_{C^\alpha([-r, 0]^2)}^p] \leq C_{\alpha p}^p T^{-p/2}.$$

*There is a constant  $K$  such that for all  $T > 0$*

$$\mathbb{E}_a[\|T^{-1}b_T - Q_a a\|_{L^2([-r, 0])}^2] \leq KT^{-1}.$$

The main results about the covariance operator  $Q$  are gathered in section 4, where it is shown that  $Q$  is injective (Theorem 4.1) and  $Q$  maps densities in  $L^2([-r, 0])$  to the Sobolev space  $H^2([-r, 0])$  of order 2 (Theorem 4.2). Assume for a moment  $Q$  to be known. Then we end up with the illposed inverse problem to determine the true density  $g \in L^2([-r, 0])$  from  $\frac{1}{T}b_T$  which is close to  $Qg$ , but  $Q^{-1}$  is unbounded and may inflate even small errors in the data without control. The usual deterministic methods for such problems, as for example Tychonov regularization or projection methods, will nevertheless give rates of convergence of order  $T^{-\frac{\alpha}{2(\alpha+1)}}$ , where  $\alpha$  is the degree of illposedness of the operator (in our case  $\alpha = 2$ ), given the true value is known to lie in a Sobolev ball of order  $s > 0$  (cf. [Baumeister87]).

In our case a method independent of the involved operator is advisable, and since the covariance operator is positive definite, the Ritz-Galerkin method is appropriate. Given approximation spaces  $V_n$  (e.g. based on splines, cf. Definition 5.5 and Example 5.6), in section 5 an estimator  $G_{T,n} \in V_n$  of  $g$  is obtained by solving the linear system

$$\langle Q_T G_{T,n}, v_n \rangle = \langle b_T, v_n \rangle, \quad \forall v_n \in V_n.$$

This projection method even leads to an estimator  $G_{T,n}$  which has an interpretation as maximum likelihood estimator (cf. Remark 5.9).

In the deterministic case some results for only approximately known operators exist (cf. [Hämarik83]). For known operators, but stochastic noise in the data, which however has to be strongly linked to the operator (the Gaussian covariance operator must have the same eigenfunctions), convergence results are proved in [NusPer99] and [MatPer00] using a spectral cut-off method. Interesting enough, they obtain the rate  $T^{-\frac{s}{2(s+\alpha-\beta)}}$  where  $\beta$  is the regularity of the noise. For the white noise model the identity operator is considered ( $\alpha = 0$ ), but with white noise ( $\beta = -\frac{1}{2}$ ), which exactly gives the  $T^{-\frac{s}{2s+1}}$ -rate. As it turns out, the stochastic error in our case behaves like Brownian motion (i.e.  $\beta = \frac{1}{2}$ ) and we indeed end up with the minimax rate  $T^{-\frac{s}{2(s+2-1/2)}}$ . Our proof of the upper bound is based on a generalization of the deterministic results for approximately known operators. The set of parameters is the intersection of a Sobolev ball with the set of densities which give rise to stationary solutions (condition  $v_0(g) \leq -\delta$  below, cf. Definition 2.1). The relation  $A \lesssim B$  means  $A$  is less than a uniform constant times  $B$ ,  $A \gtrsim B$  means the converse and  $A \sim B$  stands for  $A \lesssim B$  as well as  $A \gtrsim B$ . Theorem 5.8 then reads as follows:

*For  $S > 0$ ,  $s > \frac{1}{2}$  assume the true parameter  $g$  to lie in  $H^s([-r, 0])$  with  $\|g\|_s \leq S$ . Rescale  $G_{T,n}$  to  $\frac{S}{\|G_{T,n}\|} G_{T,n}$  in the case  $\|G_{T,n}\| > S$ . Then choosing  $n(T) \sim T^{\frac{1}{2s+3}}$  yields the uniform rate  $T^{-\frac{2s}{2s+3}}$  for the mean square error. More precisely, for  $\delta > 0$  the following holds:*

$$\sup_{\substack{\|g\|_s \leq S \\ v_0(g) \leq -\delta}} \mathbb{E}_g [\|g - G_{T,n(T)}\|_{L^2([-r,0])}^2] \lesssim T^{-\frac{2s}{2s+3}}.$$

Finally, in section 6 the corresponding lower bound is obtained in Theorem 6.2 using Assouad's cube:

*For  $s > 0$ ,  $S > 0$  and  $\delta > 0$  the statement*

$$\inf_{G_T} \sup_{\substack{\|g\|_s \leq S \\ v_0(g) \leq -\delta}} \mathbb{E}_g [\|g - G_T\|_{L^2([-r,0])}^2] \gtrsim T^{-\frac{2s}{2s+3}}$$

*holds, where the infimum is taken over all  $\sigma(X(t), 0 \leq t \leq T)$ -measurable mappings with values in  $L^2([-r, 0])$ .*

The connection of the nonparametric estimation problem with illposed problems was first observed in [Rothkirch93]. The deterministic counterpart to the nonparametric estimation problem, the question whether the trajectory of a deterministic delay equation uniquely determines the measure  $a$ , has been addressed by [Lunel99]. For small noise a nonparametric estimator for the deterministic trajectory has been proposed in [KutoMou94]. Parametric inference for affine stochastic delay equations in the case of signed point measures  $a$  has been considered in [KüchMen92], [KuMoBo94], [GushKüch99] and [KüchKuto00].

The notation follows the usual conventions.  $\mathbb{P}_a$ ,  $\mathbb{E}_a$  and  $\text{Var}_a$  denote probability, expected value and variance with respect to the underlying parameter  $a$ .

Dimension, codimension and range are abbreviated by  $\dim$ ,  $\text{codim}$  and  $\text{ran}$ , respectively. For  $f \in C([-r, 0])$  and  $\mu \in M([-r, 0])$  we shall write  $\langle f, \mu \rangle := \int f d\mu$ .  $H^s([-r, 0])$  denotes the Sobolev space of order  $s$  on  $[-r, 0]$ , which is obtained by interpolation from the classical spaces  $W^{m,p} = \{f \in L^p([-r, 0]) \mid f^{(i)} \in L^p, 1 \leq i \leq m\}$ ,  $m \in \mathbb{N}$ , for  $p = 2$ . The Sobolev norm of order  $s$  is written as  $\|\cdot\|_s$ , no subscript stands for the  $L^2$ -case, hence order 0. For the sake of brevity the space  $H^{s-} = \cup_{\alpha < s} H^\alpha$  is introduced. The norm of a measure is always its total variation norm. Weak convergence of  $(a_n) \subset M([-r, 0])$  to  $a \in M([-r, 0])$  means  $\langle f, a_n \rangle \rightarrow \langle f, a \rangle$  for all  $f \in C([-r, 0])$ . We say that a constant depends weakly continuously on the parameter  $a \in M([-r, 0])$  if weak convergence of a sequence in  $M([-r, 0])$  implies convergence of the associated constants.

It is a pleasure for me to acknowledge the support of many people. In particular, Prof. K uchler introduced me to stochastic delay equations and advised me constantly, Dr. Plato and Prof. H amarik gave me many hints for the inverse problem and Dr. Marc Hoffmann pushed the work considerably by showing me how to prove lower bounds and insisting on an optimal rate, which is exactly the one he expected. Many thanks to them and to all the other people I have not mentioned.

## 2 Stochastic delay equations

For the deterministic equation (1.1) one can define a fundamental solution and a characteristic function. These two concepts play the key role for representing and analysing the solutions, comparable to their counterparts in the theory of ordinary differential equations.

**2.1. Definition.** A (weak) solution  $x_0$  of (1.1) with initial values  $x_0(0) = 1$  and  $x_0(t) = 0$  for  $-r \leq t < 0$  is called *fundamental solution* of (1.1).

The function  $\chi : \mathbb{C} \rightarrow \mathbb{C}$  defined by

$$\chi(\lambda) := \chi_a(\lambda) := \lambda - \int_{-r}^0 e^{\lambda s} da(s), \quad (2.1)$$

is called the *characteristic function* associated to equation (1.1) or just to the measure  $a$ . We set

$$v_0 := v_0(a) := \sup\{\text{Re}(\lambda) \mid \chi_a(\lambda) = 0\}. \quad (2.2)$$

The set of all measures  $a$  with  $v_0(a) < 0$  will be denoted by  $M^-([-r, 0])$  or  $\mathcal{M}^-$  for short. Its elements will be called  $\mathcal{M}^-$ -measures.

The following facts concerning the deterministic equation (1.1) are well known (cf. [HaleLun93] or [DGLW95]). There exists a unique solution of (1.1) for continuous initial functions  $f$ . This solution  $x$  can be represented by the – always existing and uniquely determined – fundamental solution  $x_0$  via

$$x(t) = f(0)x_0(t) + \int_{-r}^0 \int_{-s}^0 x_0(t+s-u)f(u) du da(s), \quad t \geq 0. \quad (2.3)$$

The zero set of  $\chi_a$  is always discrete and the value  $v_0(a)$  always finite. For all initial functions one can estimate the asymptotic growth of  $x$  by

$$|x(t)| \lesssim e^{\delta t} \quad (2.4)$$

with  $\delta > v_0(a)$  arbitrary. Thus, if  $a$  lies in  $\mathcal{M}^-$ , the solutions of (1.1) converge exponentially fast to zero.

Stochastic equations like (1.2) are treated for example in [Mao97], [Mohammed84] and [KüchMen92]. Assuming a standard Brownian motion  $(W(t), t \geq 0)$  adapted to the filtration  $(\mathcal{F}_t)$  on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , this equation has a unique strong solution  $X$ , if the initial function  $F = F(t, \omega)$  is a.s. continuous in  $t$  as well as  $\mathcal{F}_0$ -measurable in  $\omega$  with  $\mathbb{E}[\|F\|_{C([-r,0])}^2] < \infty$ . By a variation of constants formula a version of  $X$  for  $t \geq 0$  is obtained via

$$X(t) = F(0)x_0(t) + \int_{-r}^0 \int_{-s}^0 x_0(t+s-u)F(u) du da(s) + \int_0^t x_0(t-s) dW(s). \quad (2.5)$$

Crucial for statistical purposes is the likelihood ratio of the measures induced by the processes on the interval  $[0, T]$ .

**2.2. Theorem.** *The measure  $\mu_X$  on  $C([0, T])$ , induced by  $X$  defined by (2.5) and thus satisfying (1.2), is equivalent to the translated Wiener measure  $\mu_W$  induced by  $X(0) + W$ . The likelihood ratio or Radon-Nikodym derivative is given by*

$$\begin{aligned} \Lambda_T(X, X(0) + W) &:= \frac{d\mu_X}{d\mu_W}(a, T, X) \\ &= \mathbb{E}_a[\exp(\langle b_T, a \rangle - \frac{1}{2}\langle Q_T a, a \rangle) | \sigma(X(s), 0 \leq s \leq T)], \end{aligned} \quad (2.6)$$

where  $a$  is the signed measure in (1.2).

*Proof.* The proof is an extension of the Girsanov theorem. Note that partial integration of the representation (2.5) shows that  $|X(t)|$  may be bounded by a multiple of  $W^*(t) := \sup_{0 \leq s \leq t} |W(s)|$  which implies  $\mathbb{E}[\exp(X(t) - \frac{1}{2}\langle X \rangle_t)] = 1$  (cf. the proof of Cor. 5.16 in [KarShr99]). Then in the general case of Itô processes the statement in [LipShir77, Thm. 7.1] – adapted to nonzero initial values – asserts that the Radon-Nikodym derivative is given by the respective conditional expectation of the expression

$$\exp\left(\int_0^T \int_{-r}^0 X(t+s) da(s) dX(t) - \frac{1}{2} \int_0^T \left(\int_{-r}^0 X(t+s) da(s)\right)^2 dt\right).$$

An application of the deterministic and the stochastic Fubini theorem (cf. [LipShir77, Thm. 5.15]) gives (2.6).  $\square$

By [GushKüch00] or general semigroup theory [DaPrZab92, Chap. 11] the stochastic equation (1.2) admits a stationary solution if and only if the measure

$a$  is in  $M^-([-r, 0])$ . A stationary process is obtained by extending the Brownian motion  $W$  to the whole real axis and setting

$$X(t) = \int_{-\infty}^t x_0(t-s) dW(s). \quad (2.7)$$

The integral may be understood in the Wiener sense, that is by partial integration, which immediately shows that the paths of  $X$  are a.s.  $\alpha$ -Hölder continuous for  $\alpha < \frac{1}{2}$ . It is a stationary Gaussian process with covariance function

$$q_a(t) = q(t) = \mathbb{E}[X(0)X(|t|)] = \int_0^\infty x_0(s)x_0(s+t) ds = \frac{1}{2\pi} \int_{-\infty}^\infty |\chi(i\xi)|^{-2} e^{it\xi} d\xi \quad (2.8)$$

for  $t \in \mathbb{R}$ . Note that the spectral density of  $X$  is uniquely determined by the characteristic function  $\chi$ . One can further show that  $X$  is  $\beta$ -mixing. If  $a$  is a  $\mathcal{M}^-$ -measure, then trajectories corresponding to arbitrary initial functions converge exponentially fast in  $L^2(\Omega)$  to the trajectories of the stationary solution (cf. (2.5) or [MSW86]). From now on, we shall assume  $X$  to be stationary.

For the establishment of a minimax rate it is necessary to establish the continuous dependence of the covariance function with respect to the underlying parameter  $a$  in (1.2). First a lemma is needed.

**2.3. Lemma.** *If a sequence of measures  $(a_n) \subset M([-r, 0])$  converges weakly to  $a \in M([-r, 0])$  then the associated characteristic functions converge uniformly on compact subsets of  $\mathbb{C}$  to the characteristic function of  $a$ .*

*Proof.* Let  $K \subset \mathbb{C}$  be compact. Since  $a_n \rightarrow a$  weakly the associated characteristic functions  $\chi_n$  converge pointwise to the characteristic function  $\chi$  of  $a$  by definition. For  $\mathbb{C}$ -valued continuous functions  $f \in C_{\mathbb{C}}([-r, 0])$  set

$$l_n(f) := \int_{-r}^0 f d(a_n - a),$$

such that  $l_n(e^{i\lambda \cdot}) = \chi_n(\lambda) - \chi(\lambda)$ . Now observe that the set of functions  $\{f_\lambda = e^{i\lambda \cdot} \mid \lambda \in K\} \subset C_{\mathbb{C}}([-r, 0])$  is equicontinuous, hence precompact by the Arzelà-Ascoli theorem, since for all  $s \in [-r, 0]$

$$\sup_{\lambda \in K} |f'_\lambda(s)| \leq \max_{(\lambda, s) \in K \times [-r, 0]} |\lambda e^{i\lambda s}| < \infty.$$

Moreover, the weakly convergent sequence  $(a_n - a)$  is bounded in norm, hence  $S := \sup_n \|a_n - a\| < \infty$ . If the convergence  $l_n(f_\lambda) \rightarrow 0$  were not uniformly in  $\lambda \in K$ , there would exist  $\varepsilon > 0$  and  $(\lambda_n) \subset K$  such that  $|l_n(f_{\lambda_n})| > \varepsilon$  for all  $n \in \mathbb{N}$ . Due to compactness one can assume that  $f_{\lambda_n}$  converges to an  $f \in C_{\mathbb{C}}([-r, 0])$  by passing to a convergent subsequence. But then weak convergence gives

$$|l_n(f_{\lambda_n})| \leq |l_n(f_{\lambda_n} - f)| + |l_n(f)| \leq S \|f_{\lambda_n} - f\| + |l_n(f)| \rightarrow 0,$$

which contradicts the assumption and hence shows uniform convergence.  $\square$



**2.4. Proposition.** *Let  $(a_n)$  be a sequence in  $M^-([-r, 0])$  with covariance functions  $q_n$  according to (2.8). If  $a_n \rightarrow a$  weakly and  $a \in M^-([-r, 0])$  with covariance function  $q$ , then  $\|q_n - q\|_\alpha \rightarrow 0$  for  $\alpha < \frac{5}{2}$ . The covariance functions itself only lie in  $H^{\frac{3}{2}-}(\mathbb{R})$ .*

*Proof.* In view of equation (2.8) the associated characteristic functions  $\chi_n$  and  $\chi$  respectively are considered first. The aim is to find a function dominating  $|\chi_n|^{-1}$  on the imaginary axis. Setting  $S := \sup_n \|a_n\|$  (which is finite) the estimate  $|\chi_n(iy)| \geq |y| - \|a_n\| \geq |y| - S$  is obtained. An application of Lemma 2.3 for the compactum  $[-2iS, 2iS]$  shows the convergence

$$\min_{-2S \leq y \leq 2S} |\chi_n(iy)| \rightarrow \min_{-2S \leq y \leq 2S} |\chi(iy)|$$

using a classical argument from analysis for uniform convergence. By definition of  $\mathcal{M}^-$  the functions  $\chi_n$  and  $\chi$  do not vanish on the imaginary axis, so that

$$m := \inf_n \min_{-2S \leq y \leq 2S} |\chi_n(iy)| > 0.$$

Hence  $|\chi_n(iy)|^{-1}$  is uniformly dominated by

$$g(y) := \begin{cases} m^{-1}, & -2S \leq y \leq 2S \\ \left|\frac{y}{2}\right|^{-1}, & \text{otherwise} \end{cases}.$$

Formula (2.8) and the inverse triangle inequality yield for  $y \in \mathbb{R}$

$$\begin{aligned} |\hat{q}(y) - \hat{q}_n(y)| &= \frac{1}{\sqrt{2\pi}} \left| \frac{1}{|\chi(iy)|^2} - \frac{1}{|\chi_n(iy)|^2} \right| \\ &\leq \frac{|\chi(iy) - \chi_n(iy)| |\chi(iy) + \chi_n(iy)|}{|\chi(iy)|^2 |\chi_n(iy)|^2} \\ &\leq |\chi(iy) - \chi_n(iy)| (2|y| + 2S) g^4(y) \\ &\lesssim (1 + y^2)^{-3/2} |\chi(iy) - \chi_n(iy)|. \end{aligned}$$

The constant involved is independent of  $n$  and  $y$ . Taking further into account  $\chi_n \rightarrow \chi$  pointwise by the weak convergence of the measures the dominated convergence theorem shows that

$$\lim_{n \rightarrow \infty} \|(1 + y^2)^\alpha (\hat{q}(y) - \hat{q}_n(y))^2\|_{L^1(\mathbb{R})} = 0$$

holds for  $\alpha < \frac{5}{2}$ . By the characterisation of Sobolev spaces on the real axis via the Fourier transformation this convergence result is equivalent to  $q - q_n \rightarrow 0$  in  $H^\alpha(\mathbb{R})$  (cf. [Adams75]).

Finally, note that the point measure  $a = -\delta_0$  leads to an Ornstein-Uhlenbeck process with covariance function  $q_{OU}(t) = \frac{1}{2}e^{-|t|}$  and spectral density  $\hat{q}_{OU}(y) = (2\pi)^{-1/2}(1 + y^2)^{-1}$ , so that this covariance function is an element of  $H^s(\mathbb{R})$  if and only if  $s < \frac{3}{2}$ . The result  $q_a - q_{OU} \in H^{\frac{5}{2}-}(\mathbb{R})$  shows that  $q_a \in H^{\frac{3}{2}-}(\mathbb{R})$  is the exact order of regularity for any covariance function  $q_a$ .  $\square$

**2.5. Corollary.** *The covariance functions are Lipschitz continuous with a Lipschitz constant depending weakly continuously on the underlying parameter  $a \in M^-([-r, 0])$ .*

*Proof.* As in the preceding proof, consider the covariance function  $q_{OU}(t) = \frac{1}{2}e^{-|t|}$  of the Ornstein-Uhlenbeck process. This function is obviously Lipschitz continuous. By the Sobolev embedding theorem ([Adams75]) and the preceding proposition, the difference  $q_a - q_{OU}$  is continuously differentiable for every covariance function  $q_a$ , hence  $q_a$  is Lipschitz continuous as well. The same argument shows that the Lipschitz-norm of a difference  $q_{a_n} - q_a$  of covariance functions associated to measures  $a_n$  and  $a$  tends to zero if  $(a_n)$  converges weakly to  $a$ .  $\square$

With the estimation problem in mind it is desirable to have a good description of the set of  $\mathcal{M}^-$ -measures in order to ensure that the estimator lies in  $\mathcal{M}^-$ . However, the characteristic function is nonlinear and two-dimensional sections of  $M([-r, 0])$  show that  $\mathcal{M}^-$  has a non-differentiable boundary and is not even convex (e.g., see [DGLW95]). Fortunately, two topological properties can be proved, the second one implies that a consistent estimator eventually lies in  $\mathcal{M}^-$ .

**2.6. Theorem.** *The set  $M^-([-r, 0])$  is pathwise connected in  $M([-r, 0])$ -norm.*

*Weak convergence of  $(a_n) \subset M([-r, 0])$  to  $a \in M([-r, 0])$  implies convergence of  $(v_0(a_n))$  to  $v_0(a)$ . In particular,  $a_n \in \mathcal{M}^-$  holds for sufficiently large  $n$  if  $a$  is a  $\mathcal{M}^-$ -measure.*

*Proof.* The key for proving connectedness is the following translation relation:

$$\chi_a(\lambda + \tau) = \lambda + \tau - \int_{-r}^0 e^{\lambda s} e^{\tau s} da(s) = \chi_{a_\tau}(\lambda)$$

for  $da_\tau(s) = -\tau d\delta_0(s) + e^{\tau s} da(s)$  and  $\tau \in \mathbb{R}$ . The map  $\tau \mapsto a_\tau$  is norm continuous and the associated characteristic functions are translated by  $\tau$  to the left.

Given a  $\mathcal{M}^-$ -measure  $a$  we construct a path to the  $\mathcal{M}^-$ -measure  $-T\delta_0$  for  $T > \|a\|$ . Two given  $\mathcal{M}^-$ -measures can then be connected by a path through  $-T\delta_0$  where  $T$  is larger than the maximum of both norms. Now choose the path from  $a$  to  $a_T$  along the described translation map, which obviously remains in  $\mathcal{M}^-$ . For  $h \in [0, 1]$  construct another path from  $a_T$  to  $-T\delta_0$  by  $h \mapsto c_h$  with  $dc_h(s) = -T d\delta_0(s) + (1-h)e^{T s} da(s)$ . This path is again continuous and  $c_h$  is a  $\mathcal{M}^-$ -measure since for  $\text{Re}(\lambda) \geq 0$

$$|\chi_{c_h}(\lambda)| \geq |\lambda + T| - \|c_h + T\delta_0\| \geq T - \|a\| > 0.$$

Therefore the asserted connectedness has been shown by construction.

Choose  $S$  such that  $S > \|a_n\|$  for all  $n$ . Let  $\varepsilon > 0$  be arbitrary under the condition that there are no zeroes of the characteristic function  $\chi$  of  $a$  on the lines  $\{\lambda \mid \text{Re}(\lambda) = v_0(a) \pm \varepsilon\} \subset \mathbb{C}$ . It is then shown that the number of zeroes of  $\chi_n$ , the characteristic function associated to  $a_n$ , is eventually at least one in the strip  $\Sigma := \{\lambda \mid |\text{Re}(\lambda) - v_0(a)| < \varepsilon\} \subset \mathbb{C}$  and eventually zero in the half plane  $H := \{\lambda \mid \text{Re}(\lambda) > v_0(a) + \varepsilon\}$ , which yields the statement of the theorem.

For arbitrary  $n \in \mathbb{N}$  and  $\lambda \in \mathbb{C}$  with real part  $\operatorname{Re}(\lambda) \geq v$  the estimate  $|\chi_n(\lambda)| > |\lambda| - e^{-vr} S$  holds. Therefore the region

$$R := \{\lambda \mid \operatorname{Re}(\lambda) > v_0(a) - \varepsilon, |\lambda| > S e^{-(v_0(a) - \varepsilon)r}\}$$

does neither contain a zero of  $\chi$  nor a zero of any  $\chi_n$ . The statements about  $\Sigma$  and  $H$  can consequently be reduced to bounded subsets of these sets by intersecting with the complement of  $R$ .  $\chi_n$  converges on these subsets uniformly to  $\chi$ , as was shown in Lemma 2.3. By the classical argument principle in complex analysis [Rudin87, Thm. 10.43] one can calculate the number of zeroes of a holomorphic function in a bounded region by integrating the quotient of the function and its derivative over the boundary. By uniform convergence this integral for  $\chi_n$  converges to the one for  $\chi$ , but the value is a natural number, counting the zeroes, so that it becomes constant for sufficiently large  $n$  and equals the number for  $\chi$ . Since  $\chi$  has no zeroes in  $H$  and at least one zero in  $\Sigma$  the same is eventually true for  $\chi_n$ , which proves the second statement of the theorem.  $\square$

**2.7. Remark.** Using the idea of the proof one can more generally show that the number of zeroes of  $\chi_a$  on any bounded open set in  $\mathbb{C}$  depends weakly continuously on the measure  $a$ . Properties of the zero set of  $\chi_a$  then imply the result for any right half plane in  $\mathbb{C}$ .

### 3 Convergence results

Since  $X$  is ergodic in the stationary case, the convergence result

$$T^{-1}q_T(u, v) = T^{-1} \int_0^T X(t+u)X(t+v) dt \rightarrow q(u-v) = \mathbb{E}[q_T(u, v)]$$

for  $T \rightarrow \infty$  is easily established; similarly,

$$T^{-1}b_T(s) = T^{-1}Q_T a + T^{-1} \int_0^T X(t+s) dW(t) \rightarrow Q a(s).$$

However, uniformity with respect to  $u$  and  $v$  as well as the speed of convergence are important for the asymptotic behaviour of the estimator. First, a classical moment estimate is needed.

**3.1. Theorem.** *Set  $Y_T(u, v) := T^{-1}q_T(u, v) - q(u-v)$ . Then for  $u, v, u', v' \in [-r, 0]$ ,  $m \in \mathbb{N}$  and  $T > 0$  the estimate*

$$\mathbb{E}_a[(Y_T(u, v) - Y_T(u', v'))^{2m}] \lesssim T^{-m} \|(u, v) - (u', v')\|^m$$

*holds with a constant independent of  $T$ , which may be chosen to depend weakly continuously on the underlying parameter  $a \in M^-([-r, 0])$ . The norm symbol on the right hand side denotes any norm on  $\mathbb{R}^2$ .*

*Proof.* The following short hand notations will be used:

$$\begin{aligned} A(t) &:= X(t+u), & B(t) &:= X(t+v) - X(t+v'), & \alpha &:= q(u-v) - q(u-v'), \\ C(t) &:= X(t+v'), & D(t) &:= X(t+u) - X(t+u'), & \gamma &:= q(u-v') - q(u'-v'). \end{aligned}$$

By Definition 1.1 one obtains

$$\begin{aligned} &Y_T(u, v) - Y_T(u', v') \\ &= \frac{1}{T} \int_0^T (X(t+u)X(t+v) - q(u-v) - X(t+u')X(t+v') + q(u'-v')) dt \\ &= \frac{1}{T} \int_0^T (A(t)B(t) - \alpha) dt + \frac{1}{T} \int_0^T (C(t)D(t) - \gamma) dt. \end{aligned}$$

Using the Minkowski inequality for  $L^{2m}(\Omega)$  yields

$$\begin{aligned} \mathbb{E}_a[(Y_T(u, v) - Y_T(u', v'))^{2m}]^{1/2m} &\leq \mathbb{E}_a[(\frac{1}{T} \int_0^T (A(t)B(t) - \alpha) dt)^{2m}]^{1/2m} \\ &\quad + \mathbb{E}_a[(\frac{1}{T} \int_0^T (C(t)D(t) - \gamma) dt)^{2m}]^{1/2m}. \end{aligned} \tag{3.1}$$

Both summands can be estimated analogously, so that only the estimate for the first one will be presented. By symmetry one gets

$$\begin{aligned} &\mathbb{E}_a[(\frac{1}{T} \int_0^T (A(t)B(t) - \alpha) dt)^{2m}] \tag{3.2} \\ &= \frac{1}{T^{2m}} \mathbb{E}_a[\int_0^T \cdots \int_0^T (A(t_1)B(t_1) - \alpha) \cdots (A(t_{2m})B(t_{2m}) - \alpha) dt_{2m} \cdots dt_1] \\ &= \frac{(2m)!}{T^{2m}} \int_0^T \int_0^{t_1} \cdots \int_0^{t_{2m-1}} \mathbb{E}_a[(A_1B_1 - \alpha) \cdots (A_{2m}B_{2m} - \alpha)] dt_{2m} \cdots dt_1, \end{aligned}$$

where  $A_i := A(t_i)$ ,  $B_i := B(t_i)$  are centered, jointly Gaussian random variables. In general, for a centered Gaussian random vector  $(N_1, \dots, N_{2n})$  the following formula holds true

$$\mathbb{E}[N_1 \cdots N_{2n}] = \sum_{\Gamma \in P_2(1, \dots, 2n)} \prod_{\{k, l\} \in \Gamma} \mathbb{E}[N_k N_l], \tag{3.3}$$

where  $P_2(1, \dots, 2n)$  denotes the set of all partitions of  $\{1, \dots, 2n\}$  into subsets of cardinality 2. In our case choose  $n := 2m$ ,  $N_{2i-1} := A_i$ ,  $N_{2i} := B_i$ . In addition to the terms in (3.3), there are also summands involving  $\alpha$ , but the expectations of these summands cancel with the expectations of the summands, where neighbouring random variables  $N_{2i-1}$ ,  $N_{2i}$  turn up, since  $\alpha = \mathbb{E}[N_{2i-1}N_{2i}]$ . Thus, one obtains

$$\mathbb{E}[(N_1N_2 - \alpha) \cdots (N_{4m-1}N_{4m} - \alpha)] = \sum_{\substack{\Gamma \in P_2(1, \dots, 4m) \\ \forall i: \{2i-1, 2i\} \notin \Gamma}} \prod_{\{k, l\} \in \Gamma} \mathbb{E}[N_k N_l]. \tag{3.4}$$

According to (2.4) there are  $\delta > 0$  and  $C > 0$  such that  $|q(t)| \leq Ce^{-\delta t}$  for all  $t \in \mathbb{R}$ . Proposition 2.4 and Theorem 2.6 show that  $\delta$  and  $C$  can be chosen to be weakly continuous. Moreover,  $q$  is Lipschitz continuous with a Lipschitz constant  $L$  by Corollary 2.5. Assuming  $k < l$ , i.e.  $t_k \geq t_l$ , the appearing covariances can be bounded as follows:

$$\begin{aligned}
\mathbb{E}[A_k A_l] &= q(t_k - t_l) \leq Ce^{-\delta(t_k - t_l)}, \\
\mathbb{E}[A_k B_l] &= q(t_k - t_l + u - v) - q(t_k - t_l + u - v') \\
&\leq \min(2Ce^{-\delta(t_k - t_l - r)}, L|v - v'|), \\
\mathbb{E}[B_k B_l] &= 2q(t_k - t_l) - q(t_k - t_l + v - v') - q(t_k - t_l + v' - v) \\
&\leq \min(4Ce^{-\delta(t_k - t_l - r)}, 2L|v - v'|).
\end{aligned}$$

Every product in (3.4) consists of  $2m$  factors. The bounds for the covariances decrease with increasing difference  $t_k - t_l$ , which shows that the bound for the product is the larger the closer the indices  $\{k, l\}$  in the partition  $\Gamma$  are. Moreover, there are at least  $m$  factors involving a covariance formed with some  $B_k$ , so that for all considered partitions  $\Gamma$

$$\prod_{\{k, l\} \in \Gamma} \mathbb{E}[N_k N_l] \lesssim \prod_{i=1}^m e^{-\delta(t_{2i-1} - t_{2i})} \min(|v - v'|, e^{-\delta(t_{2i-1} - t_{2i})}).$$

The constant in this estimate depends continuously on  $\delta$ ,  $C$  and  $L$ . Applying this bound to (3.2), the multiple integrals can be evaluated pairwise and one obtains with  $|P_2|$  as number of 2-partitions:

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{T} \int_0^T A(t)B(t) - \alpha dt\right)^{2m}\right] \\
&\lesssim \frac{(2m)!|P_2|}{T^{2m}} \prod_{i=1}^m \int_0^T \int_0^{t_{2i-1}} e^{-\delta(t_{2i-1} - t_{2i})} |v - v'| dt_{2i} dt_{2i-1} \\
&= \frac{(2m)!|P_2|}{T^{2m}} \left( \int_0^T \int_0^t e^{-\delta s} |v - v'| ds dt \right)^m \\
&\lesssim \frac{|v - v'|^m}{T^m} \left( \int_0^\infty e^{-\delta s} ds \right)^m \\
&\lesssim \frac{|v - v'|^m}{T^m}. \tag{3.5}
\end{aligned}$$

Note that the constants in this estimate again only depend upon  $\delta$ ,  $C$ ,  $L$  and  $m$ . Treating the other summand of (3.1) in exactly the same way, the assertion has been proved.  $\square$

**3.2. Corollary.** *For  $\alpha < \frac{1}{2}$  and  $p \geq 1$  there is a constant  $C_{\alpha p}$ , depending continuously on the parameter  $a$ , such that for all  $T > 0$*

$$\mathbb{E}_a[\|T^{-1}q_T(u, v) - q(u - v)\|_{C^\alpha([-r, 0]^2)}^p] \leq C_{\alpha p}^p T^{-p/2}.$$

*Proof.* Following the lines of the proof of the Kolmogorov continuity theorem in [DaPrZab92, Thm. 3.4], the estimate of the preceding theorem and an estimate similar to (3.5) show that for all  $\beta < \frac{1}{2}$  and  $m \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}_a[\|Y_T\|_{W^{\beta,2m}}^{2m}] &\lesssim \mathbb{E}_a[|Y_T(0,0)|^{2m}] + \int_{[-r,0]^2} \int_{[-r,0]^2} \frac{\mathbb{E}_a[|Y_T(x) - Y_T(y)|^{2m}]}{\|x - y\|^{2m\beta+2}} dx dy \\
&\lesssim \mathbb{E}_a\left[\frac{1}{T} \int_0^T (X(t)^2 - q(0)) dt\right]^{2m} \\
&\quad + T^{-m} \int_{[-r,0]^2} \int_{[-r,0]^2} \|x - y\|^{m(1-2\beta)-2} dx dy \\
&\lesssim T^{-m}.
\end{aligned}$$

An application of the Sobolev embedding theorem with the norm estimate  $\|\cdot\|_{C^\alpha} \leq C(\alpha, \beta, m) \|\cdot\|_{W^{\beta,2m}}$  for  $\alpha < \beta - \frac{1}{2m}$  then proves the corollary.  $\square$

**3.3. Lemma.** *For  $X$  satisfying the stochastic equation (1.2) introduce the random function*

$$r_T(s) := \int_0^T X(t+s) dW(t), \quad s \in [-r, 0].$$

*For continuous versions of  $b_T$  and  $r_T$  the following holds true in the stationary case with parameter  $a \in M^-([-r, 0])$ :*

$$b_T(s) = Q_T a(s) + r_T(s), \quad \forall s \in [-r, 0], \quad \mathbb{P}_a - a.s., \quad (3.6)$$

$$\mathbb{E}_a[\langle r_T, \mu \rangle] = 0, \quad \mu \in M([-r, 0]), \quad (3.7)$$

$$\mathbb{E}_a[\langle r_T, \mu \rangle^2] = T \langle Q_a \mu, \mu \rangle, \quad \mu \in M([-r, 0]). \quad (3.8)$$

*Proof.* By Kolmogorov's continuity theorem versions of  $b_T$  and  $r_T$  may be chosen to be continuous functions on  $[-r, 0]$ . Since  $dX(t) = (\int_{-r}^0 X(t+u) da(u)) dt + dW(t)$  holds under  $\mathbb{P}_a$ , the first equality then follows from the definition of  $b_T$ . The second and the third statement are obtained using the stochastic Fubini theorem

$$\langle r_T, \mu \rangle = \int_0^T \int_{-r}^0 X(t+s) d\mu(s) dW(t)$$

and then applying basic facts about stochastic integrals.  $\square$

**3.4. Corollary.** *There is a constant  $K$ , depending weakly continuously on the parameter  $a$ , such that for all  $T > 0$  the following bound is obtained*

$$\mathbb{E}_a[\|T^{-1}b_T - Q_a a\|_{L^2([-r,0])}^2] \leq KT^{-1}.$$

*Proof.* The following calculation, using (3.6), triangle inequality, the Fubini theorem and Corollary 3.2, yields the result:

$$\begin{aligned}
\mathbb{E}_a[\|T^{-1}b_T - Q_a a\|^2] &= \mathbb{E}_a[\|T^{-1}r_T - \int_{-r}^0 (q_a(\cdot - s) - T^{-1}q_T(\cdot, s)) da(s)\|^2] \\
&\leq 2T^{-2} \mathbb{E}_a[\|r_T\|^2] + \\
&\quad 2 \mathbb{E}_a[\|q_a(u - v) - T^{-1}q_T(u, v)\|_{C([-r, 0]^2)}^2] \|a\|^2 r \\
&\leq 2T^{-2} \int_{-r}^0 \int_0^T \mathbb{E}_a[X(u)^2] du ds + 2r \|a\|^2 C_{02}^2 T^{-1} \\
&= 2r(q(0) + \|a\|^2 C_{02}^2) T^{-1}.
\end{aligned}$$

□

## 4 Properties of the covariance operator

Motivated by the convergence results of the preceding section, the covariance operator will be analysed in detail. Note that a covariance operator is always positive semi-definite and compact.

**4.1. Proposition.** *The covariance operator  $Q$  is strictly positive definite, i.e. for all  $\mu \in M([-r, 0])$ ,  $\mu \neq 0$ , the inequality  $\langle Q\mu, \mu \rangle > 0$  holds.*

*Proof.* Since  $Q$  is positive semi-definite, it suffices to show injectivity. By Theorem 2.3 the process  $X$  induces a measure  $\mu_X$  on  $C([0, r])$  equivalent to the Wiener measure  $\mathbb{P}$  translated by  $X(0)$ .  $X(0)$  is distributed with normal law  $N(0, \sigma^2)$ ,  $\sigma > 0$ . Hence for  $\mu \in M([-r, 0])$  independence of  $X(0)$  from  $W$  yields

$$\begin{aligned}
\langle Q_a \mu, \mu \rangle = 0 &\Leftrightarrow \langle X(\cdot + r), \mu \rangle = 0 \quad \mu_X - a.s. \\
&\Leftrightarrow \langle W(\cdot + r) + X(0), \mu \rangle = 0 \quad \mathbb{P} \otimes N(0, \sigma^2) - a.s. \\
&\Leftrightarrow \mathbb{E}[\langle W(\cdot + r), \mu \rangle^2] = 0 \text{ and } \int_{-r}^0 \mathbf{1} d\mu(s) = 0.
\end{aligned}$$

Girsanov's theorem implies ([RevYor99, Cor. VII.2.3]) that the support of the Wiener measure on  $C([0, r])$  is the space  $C_0([0, r])$  of all continuous functions starting in zero. Therefore  $\mathbb{E}[\langle W(\cdot + r), \mu \rangle^2] = 0$  implies  $\mu = \alpha \delta_{-r}$  for some  $\alpha \in \mathbb{R}$ . The argument relies upon the fact that for all other measures  $\mu$  the set  $\{f \in C_0([0, r]) \mid \langle f, \mu \rangle \neq 0\}$  is a nonempty open subset of  $C_0([0, r])$ . The second condition in the above equivalence implies then  $\alpha = 0$  proving that only  $\mu \equiv 0$  satisfies  $\langle Q\mu, \mu \rangle = 0$ . □

From this point on, we shall mainly consider the restriction of the parameter space  $M^-([-r, 0])$  to the set of  $\mathcal{M}^-$ -measures with an  $L^2([-r, 0])$ -integrable Lebesgue density. Abusing the notation, everything defined so far will be applied to the  $L^2$ -density directly and not to its generated measure. For instance, we

shall consider the covariance operator  $Q$  to be defined on  $L^2([-r, 0])$  and write for  $f, g \in L^2([-r, 0])$

$$\langle Qf, g \rangle = \int_{-r}^0 \int_{-r}^0 q(t-s) f(t) g(s) ds dt.$$

From general theory it is known that the covariance operator  $Q$  is a trace class operator. In our case the range can be determined explicitly with respect to the scale of Sobolev spaces.

**4.2. Theorem.** *The covariance operator is a continuous linear operator  $Q_a : L^2([-r, 0]) \rightarrow H^2([-r, 0])$ . Its range is closed in  $H^2([-r, 0])$ , so that there are constants  $C_Q \geq c_Q > 0$  with*

$$c_Q \|f\| \leq \|Q_a f\|_2 \leq C_Q \|f\|, \quad \forall f \in L^2([-r, 0]). \quad (4.1)$$

*The mapping  $a \mapsto Q_a$  is weakly-operator norm-continuous.  $C_Q$  and  $c_Q$  may thus be chosen to depend weakly continuously on the parameter  $a \in M^-([-r, 0])$ .*

*The range of  $Q_a$  is the whole space  $H^2([-r, 0])$  if  $Q_a$  is viewed as an operator on the space  $L^2([-r, 0]) \oplus \text{span}(\delta_{-r}, \delta_0)$ .*

*Proof.* In the case of the Ornstein-Uhlenbeck process one obtains  $q_{OU}(t) = \frac{1}{2}e^{-|t|}$ , which yields  $q'_{OU}(t) = q_{OU}(t) - \delta_0$  in the distributional sense. Using Proposition 2.4, this shows that  $k(t) := q''_a(t) + \delta_0$  can be interpreted as an element of  $H^{\frac{1}{2}-}([-r, 0])$  for all parameters  $a$ . Denoting the derivative operator by  $D$  one obtains

$$\begin{aligned} D^2 Q_a f(t) &= D^2 \int_{-r}^0 q_a(t-s) f(s) ds = \int_{-r}^0 q''_a(t-s) f(s) ds \\ &= \int_{-r}^0 k(t-s) f(s) ds - f(t), \quad f \in L^2([-r, 0]), t \in [-r, 0]. \end{aligned}$$

Therefore we may write  $D^2 Q = D^2 Q_a = -\text{Id} + K : L^2([-r, 0]) \rightarrow L^2([-r, 0])$  with an operator  $K$  on  $L^2([-r, 0])$  which is compact since  $H^\alpha([-r, 0])$  embeds compactly into  $L^2([-r, 0])$  for  $\alpha > 0$ .  $D^2 Q$  is thus, as a compactly disturbed identity operator, a Fredholm operator of index 0 on  $L^2([-r, 0])$  (see e.g. [Werner95] for the functional analytic techniques used in this proof).

Let  $V$  denote the kernel of  $D^2 Q$  and  $V^\perp$  a complement of  $V$  in  $L^2([-r, 0])$ . Then the restriction  $D^2 Q|_{V^\perp}$  is an injective operator with the same range as  $D^2 Q$  on  $L^2([-r, 0])$ , which is closed by Fredholm theory. The closed graph theorem then shows the existence of a constant  $c > 0$  such that

$$\|D^2 Q w\| \geq c \|w\| \text{ for all } w \in V^\perp.$$

In particular, one obtains  $\|Q w\|_2 \geq c \|w\|$  for all  $w \in V^\perp$ , which proves that  $Q|_{V^\perp}$  has closed range in  $H^2([-r, 0])$ .



Now,  $V$  is finite dimensional by Fredholm theory, so that  $Q_a|_V$  has finite dimensional, hence closed range. The representation of the range as a direct sum of closed subspaces

$$\text{ran } Q = \text{ran } Q|_V \oplus \text{ran } Q|_{V^\perp}$$

shows immediately that  $\text{ran } Q$  is closed itself. The equivalence of the norms in (4.1) then follows again from the closed graph theorem by the injectivity of  $Q$ .

The estimate for the operator norm  $\|Q_a - Q_{a_n}\| \leq \|q_a - q_{a_n}\|_2$  in connection with Proposition 2.4 proves that  $Q_{a_n} \rightarrow Q_a$  in operator norm if  $a_n \rightarrow a$  weakly. Hence, the operator norms do converge which may be taken for  $C_Q$ . Putting  $c_n := \inf_{\|f\|=1} \|Q_{a_n} f\|$  the estimate

$$\|Q_a f\| \geq \|Q_{a_n} f\| - \|(Q_a - Q_{a_n})f\| \geq (c_n - \|Q_a - Q_{a_n}\|)\|f\|$$

implies  $\|Q_a f\| \geq \limsup_n c_n \|f\|$ . The converse estimate

$$\inf_{\|f\|=1} \|Q_a f\| \leq \inf_{\|f\|=1} \|Q_{a_n} f\| + \|Q_a - Q_{a_n}\|$$

shows that the choice  $c_Q := \inf_{\|f\|=1} \|Q_a f\|$  leads to a weakly continuous constant  $c_Q$ .

Set  $W^\perp := \text{ran } Q|_{V^\perp} \subset H^2$ . Since  $D^2 Q|_{V^\perp}$  is injective, we can choose a subspace  $W \subset H^2$  such that  $\ker D^2 \subset W$  and  $H^2 = W \oplus W^\perp$ . Then, by surjectivity of  $D^2 : H^2 \rightarrow L^2$ ,  $D^2(W)$  is a complement to  $D^2(W^\perp) = D^2 Q(V^\perp)$  in  $L^2$ . The kernel of  $D^2 : H^2 \rightarrow L^2$  is two-dimensional and the codimension of  $\text{ran } D^2 Q|_{L^2}$  equals the dimension of  $V$  by Fredholm theory, so that

$$\dim V = \text{codim } \text{ran } D^2 Q|_{L^2} = \text{codim } D^2 Q(V^\perp) = \dim D^2(W) = \dim W - 2.$$

The injectivity of  $Q$  now implies that  $\dim W = \dim V + \text{codim } \text{ran } Q$ . Altogether  $\text{codim } \text{ran } Q = 2$  follows. A glance at Proposition 2.4 in connection with  $q_{OU}(t) = \frac{1}{2}e^{-|t|}$  shows that the covariance function is an element of  $H^{\frac{5}{2}-}([0, r])$  and  $H^{\frac{5}{2}-}([-r, 0])$ , if suitably restricted. Hence  $Q\delta_{-r}$  and  $Q\delta_0$  lie in  $H^2(\mathbb{R})$  and using the injectivity of  $Q$  once again yields the surjectivity of  $Q$  defined on the space spanned by  $L^2([-r, 0])$  and the measures  $\delta_{-r}, \delta_0$ .  $\square$

**4.3. Remark.** That  $Q$  is smoothing of order 2, i.e. has degree 2 of ill-posedness, becomes plausible if the covariance operator of Brownian motion is considered, whose kernel is  $k(s, t) = \min(s, t)$  and which essentially integrates twice.

Similar results for operators acting on the space of continuous functions have been obtained by [Sakhnovich96].

Later, for the right choice of the approximation spaces the following lemma is important, because it shows a very strong uniformity in the behaviour of  $Q_a$  with respect to  $a$ .

**4.4. Lemma.** *If the element  $a_0 \in M^-([-r, 0])$  and the weakly compact set  $A \subset M^-([-r, 0])$  are given, then the associated covariance operators satisfy*

$$\inf_{a \in A} \inf_{\substack{f \in L^2([-r, 0]) \\ f \neq 0}} \frac{\langle Q_a f, f \rangle}{\langle Q_{a_0} f, f \rangle} > 0.$$

*Proof.* By the positive definiteness of the covariance operator  $Q$  one can define its square root  $Q^{1/2}$ , which is a bounded operator, and the inverse  $Q^{-1/2}$  of the square root, which is an unbounded operator. Since in our case the Gaussian measures are equivalent, the Feldman-Hajek theorem [DaPrZab, Thm. 2.23] implies that  $\text{ran } Q_a^{1/2}$  and  $\text{ran } Q_{a_0}^{1/2}$  agree, so that  $T_a := Q_{a_0}^{-1/2} Q_a^{1/2}$  is an isomorphism on  $L^2([-r, 0])$  by the closed graph theorem. Moreover,  $Q_a^{1/2}$  and thus  $T_a$  and its adjoint  $T_a^*$  depend continuously on  $Q_a$  and therefore weakly continuously on  $a$ . By compactness there is a constant  $C_T > 0$  independent of  $a \in A$  and  $h \in L^2([-r, 0])$  such that

$$\|T_a^* h\| \geq C_T \|h\|.$$

Thus, setting  $h := Q_{a_0}^{1/2} f$  yields

$$\frac{\langle Q_a f, f \rangle}{\langle Q_{a_0} f, f \rangle} = \frac{\langle Q_a Q_{a_0}^{-1/2} h, Q_{a_0}^{-1/2} h \rangle}{\langle h, h \rangle} = \frac{\langle T_a^* h, T_a^* h \rangle}{\|h\|^2} \geq C_T^2,$$

which is the desired bound.  $\square$

**4.5. Remark.** As is obvious from the proof, we only need that  $Q_{a_0}$  is a covariance operator of a Gaussian measure on  $L^2([-r, 0])$  which is equivalent to  $Q_a$ . Hence, the operator  $Q_W$  with

$$Q_W f(t) = \int_0^r \min(t+r, s) f(s-r) ds + \int_{-r}^0 f(s) ds,$$

derived from  $\mu_W$  in Theorem 2.2 with  $\mathbb{E}[X(0)^2] = 1$ , will do.

## 5 Construction of the estimator

The situation discussed so far is the following: We know

$$\lim_{T \rightarrow \infty} \frac{1}{T} Q_T = Q_a, \quad \lim_{T \rightarrow \infty} \frac{1}{T} b_T = Q_a a.$$

If we knew the limits we would end up with the inverse problem to determine  $a$  from  $Q_a a$ , which in particular shows that  $a$  is identifiable after an infinitely long observation time. However, we only know the operator  $Q_a$ , which is compact, and the image  $Q_a a$  approximately. Hence we face a delicate illposed inverse problem.

Since  $Q_a$  is positive definite, but unknown, the obvious choice for treating the illposed problem is to use the Ritz-Galerkin projection method. The estimator  $G_{T,n}$  of  $g$  from the observation  $(X(t), -r \leq t \leq T)$  is hence constructed as the unique element in a finite-dimensional approximation space  $V_n \subset L^2([-r, 0])$  which satisfies

$$\langle Q_T G_{T,n}, v_n \rangle = \langle b_T, v_n \rangle, \quad \forall v_n \in V_n. \quad (5.1)$$

Choosing a basis of  $V_n$  only a system of linear equations has to be solved. The better  $Q_T$  and  $b_T$  approximate  $Q_a$  and  $Q_a a$  the better  $V_n$  has to approximate functions from the space  $H^s([-r, 0])$ . This means that a whole range of approximation spaces comes into play and one has to choose  $n = n(T)$  dependent on  $T$ .

The error analysis will be based upon the next general theorem, which has inherited many ideas from an analogous statement in [Hämarik83]. The occurring operator  $R_{n\eta}$  corresponds to the solution operator for (5.1),  $u_n$  to the estimator and  $u$  to the true parameter in our context.

**5.1. Theorem.** *Let  $A$  be a positive definite operator on the Hilbert space  $H$  and  $A_\eta$  be another operator on  $H$  with  $\|A - A_\eta\| \leq \eta$ . Let  $u, f, f_\delta \in H$  be given with  $Au = f$ . For a finite-dimensional subspace  $V_n \subset H$  denote the orthogonal projection on this subspace by  $P_n$  and assume that  $P_n A_\eta|_{V_n}$  is invertible. Set  $R_n := (P_n A|_{V_n})^{-1} P_n$ ,  $R_{n\eta} := (P_n A_\eta|_{V_n})^{-1} P_n$  and  $u_n := R_{n\eta} f_\delta$ . Then*

$$\begin{aligned} \|u - u_n\| &\leq [1 + \|R_{n\eta}\|(\|(\text{Id} - P_n)A\| + \eta)] \|(\text{Id} - P_n)u\| \\ &\quad + (1 + \|R_{n\eta}\|\eta) \|R_n(A_\eta u - f_\delta)\| \end{aligned} \quad (5.2)$$

holds. If  $\eta < \|R_n\|^{-1}$  is satisfied, then  $P_n A_\eta|_{V_n}$  is invertible with

$$\|R_{n\eta}\| \leq \frac{\|R_n\|}{1 - \eta\|R_n\|}. \quad (5.3)$$

*Proof.* Suitable applications of the triangle inequality are the main tool for the first estimate. Note that  $(P_n A|_{V_n})^{-1}$  is well-defined since  $\langle Av_n, v_n \rangle > 0$  for all  $v_n \in V_n$ . Further observe that  $R_{n\eta} A_\eta$  is a projection on  $V_n$  by definition, that  $(\text{Id} - P_n)^2 = \text{Id} - P_n$  holds by the projection property of  $\text{Id} - P_n$  and that  $\|A(\text{Id} - P_n)\| = \|(\text{Id} - P_n)^* A^*\| = \|(\text{Id} - P_n)A\|$  is satisfied by selfadjointness. Thus the estimate (5.2) is obtained:

$$\begin{aligned} \|u - u_n\| &\leq \|u - P_n u\| + \|P_n u - R_{n\eta} A_\eta u\| + \|R_{n\eta}(A_\eta u - f_\delta)\| \\ &\leq \|(\text{Id} - P_n)u\| + \|R_{n\eta} A_\eta (P_n - \text{Id})u\| \\ &\quad + \|(R_{n\eta} - R_n)(A_\eta u - f_\delta)\| + \|R_n(A_\eta u - f_\delta)\| \\ &\leq [1 + \|R_{n\eta}\|(\|A(\text{Id} - P_n)\| + \|A_\eta - A\|)] \|(\text{Id} - P_n)u\| \\ &\quad + [\|R_{n\eta}(A - A_\eta)\| + 1] \|R_n(A_\eta u - f_\delta)\| \\ &\leq [1 + \|R_{n\eta}\|(\|(\text{Id} - P_n)A\| + \eta)] \|(\text{Id} - P_n)u\| \\ &\quad + (1 + \|R_{n\eta}\|\eta) \|R_n(A_\eta u - f_\delta)\|. \end{aligned}$$

Subsequently, the observation

$$\|R_n\| = \sup_{v_n \in V_n \setminus \{0\}} \frac{\|R_n v_n\|}{\|v_n\|} = \sup_{v_n} \frac{\|v_n\|}{\|P_n A v_n\|} = \sup_{\|v_n\|=1} \langle Av_n, v_n \rangle^{-1} \quad (5.4)$$

and the analogue for  $R_{n\eta}$  will be useful. The inequality for  $v_n \in V_n$

$$\|P_n A_\eta v_n\| \geq \|P_n A v_n\| - \|P_n(A_\eta - A)v_n\| \geq (\|R_n\|^{-1} - \eta) \|v_n\|$$

shows that  $P_n A_\eta|_{V_n}$  is invertible for  $\eta < \|R_n\|^{-1}$ . Further calculations give

$$\begin{aligned}
\|R_{n\eta}\| &= \sup_{v_n \in V_n \setminus \{0\}} \frac{\|v_n\|}{\|P_n A_\eta v_n\|} \\
&\leq \sup_{v_n} \frac{\|v_n\|}{\|P_n A v_n\| - \eta \|v_n\|} \\
&\leq \frac{\sup_{v_n} \frac{\|v_n\|}{\|P_n A v_n\|}}{1 - \eta \sup_{v_n} \frac{\|v_n\|}{\|P_n A v_n\|}} \\
&= \frac{\|R_n\|}{1 - \eta \|R_n\|},
\end{aligned}$$

which is the second estimate (5.3).  $\square$

In our case,  $Q_T$  will play the role of  $A_\eta$  in the preceding theorem and the restricted operator  $P_n Q_T|_{V_n}$  will always be invertible almost surely. Before proving it in great generality we need an almost obvious lemma, the proof of which is based on a nice idea by Prof. Behrends.

**5.2. Lemma.** *Let  $n$  linear independent functions  $f_j : [-r, T] \rightarrow \mathbb{R}$ ,  $j = 1, \dots, n$ , be given, which vanish in a neighbourhood of  $T$ . Define for  $t \geq 0$  the translation operator  $T_t$  by*

$$T_t f_j(s) := \begin{cases} f_j(s-t), & \text{if } s \in [-r+t, T] \\ 0, & \text{if } s \in [-r, -r+t) \end{cases}, \quad s \in [-r, T].$$

*Then for any  $T > 0$  there are points  $0 = t_1 < t_2 < \dots < t_m < T$  such that the  $m \cdot n$  functions  $(T_{t_i} f_j)_{1 \leq i \leq m, 1 \leq j \leq n}$  are linearly independent on  $[-r, T]$ .*

*Proof.* To prove the lemma it suffices to show the linear independence of the family  $(T_{t_i} f_j)$  for  $m = 2$  with a point  $t_2$  which lies arbitrarily close to 0, since then the family  $\mathcal{F} := \{f_1, \dots, f_n, T_{t_2} f_1, \dots, T_{t_2} f_n\}$  satisfies the conditions of the theorem and a simple induction over  $m$  will yield the result.

For  $t \in (-r, T]$  consider the subspaces

$$N_t := \left\{ \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \mid \sum_{j=1}^n \lambda_j f_j|_{[-r, t]} = 0 \right\} \subset \mathbb{R}^n.$$

Note  $N_t \subset N_s$  for  $t > s$  and  $N_{T-\varepsilon} = \{0\}$  for some  $\varepsilon > 0$  by linear independence and the condition on the functions  $f_j$  near  $T$ . Introduce the points  $\tau_j$  where the dimension of  $N_t$  drops, i.e.  $\tau_j := \sup\{t \in (-r, T] \mid \dim N_t \geq j\}$ ,  $j = 0, \dots, n$ . A choice of  $t_2 \in (0, \varepsilon)$  with  $t_2$  not equal to any distance  $|\tau_j - \tau_{j'}|$  will then produce a linear independent family  $\mathcal{F}$ . To prove this, consider functions

$$F_1 = \sum_{j=1}^n \alpha_j f_j \quad \text{and} \quad F_2 = \sum_{j=1}^n \beta_j T_{t_2} f_j \quad \text{with } F_1 + F_2 = 0.$$

Assume  $F_1 \neq 0$  and set  $t^* := \inf\{t \in [-r, T] \mid F_1(t) \neq 0\}$ . It is easily established that  $t^* \geq -r + t_2$  since  $F_2|_{[-r, -r+t_2]} = 0$ . Thus,  $\alpha \in N_{t^*}$  and by the monotonicity of the  $N_t$  even  $\alpha \in N_{\tau_j}$  is obtained for the smallest  $\tau_j$  with  $\tau_j \geq t^*$ . This implies  $F_1|_{[-r, \tau_j]} = 0$ . Now,  $F_2$  is the image of some  $\Phi_2$  under  $T_{t_2}$  and  $F_2|_{[-r, \tau_j]} = 0$  then gives  $\Phi_2|_{[-r, \tau_j - t_2]} = 0$ . However, the same reasoning applies to  $\Phi_2$  and therefore  $\Phi_2$  vanishes on some interval  $[-r, \tau_{j'}]$ , where  $\tau_{j'}$  is strictly larger than  $\tau_j - t_2$  by the choice of  $t_2$ . Then  $F_2$  and hence  $F_1$  vanish on  $[-r, \tau_{j'} + t_2]$  with  $\tau_{j'} + t_2 > \tau_j \geq t^*$ . This is a contradiction to the definition of  $t^*$  and thus  $F_1 = 0$  everywhere. Therefore also  $F_2 = 0$ , but the families  $(f_j)$  and  $(T_{t_2} f_j)$  are each linearly independent by assumption and by  $t_2 < \varepsilon$  respectively, so that  $\alpha = \beta = 0$  and the entire family  $\mathcal{F}$  is linearly independent.  $\square$

**5.3. Proposition.** *Let  $A_n \subset M([-r, 0])$  be a subspace with  $\dim A_n = n < \infty$  and consider the operator  $K_T : A_n \rightarrow A_n$  which is implicitly defined by*

$$\langle K_T b, c \rangle := \langle Q_T b, c \rangle = \int_{-r}^0 \int_{-r}^0 q_T(u, v) db(u) dc(v), \quad b, c \in A_n.$$

*Then for  $T > 0$  the operator  $K_T$  is bijective with probability one.*

*Proof.* Let the points  $0 < t_1 < \dots < t_n < T$  be fixed and chosen later. Then by positive semidefiniteness of  $K_T$ , which is obvious from the second line of calculations below, the following probability  $P$  has to be shown to vanish:

$$\begin{aligned} P &:= \mathbb{P}_g(\exists a_n \in A_n \setminus \{0\} : \langle K_T a_n, a_n \rangle = 0) \\ &= \mathbb{P}_g(\exists a_n \in A_n \setminus \{0\} : \int_0^T \left( \int_{-r}^0 X(t+u) da_n(u) \right)^2 dt = 0) \\ &= \mathbb{P}_g(\exists a_n \in A_n \setminus \{0\} \forall t \in [0, T] : \int_{-r}^0 X(t+u) da_n(u) = 0) \\ &\leq \mathbb{P}_g(\exists a_n \in A_n \setminus \{0\} \forall i = 1, \dots, n : \int_{-r}^0 X(t_i + u) da_n(u) = 0) \\ &= \mathbb{P}_g(\text{the matrix } M := \left( \int_{-r}^0 X(t_i + u) db_j(u) \right)_{i,j=1, \dots, n} \text{ is singular}), \end{aligned}$$

where  $(b_1, \dots, b_n)$  is a basis of  $A_n$ . On the interval  $[-r + t_1, T]$   $\mathbb{P}_g$  is equivalent to  $\mathbb{P}_B$ , the measure of Brownian motion  $(B(t), -r \leq t \leq T)$  starting at  $t = -r$  in zero (cf. Theorem 2.3). Hence it suffices to show that the matrix  $M$  is almost surely nonsingular under  $\mathbb{P}_B$ , which will be accomplished by showing that it has – as a random vector in  $\mathbb{R}^{n \times n}$  – a density with respect to  $n \times n$ -dimensional Lebesgue measure, since the set of singular matrices has nonempty interior in  $\mathbb{R}^{n \times n}$ . The covariance matrix  $C$  of this normally distributed random vector has

shown to be singular. Its entries are

$$\begin{aligned}
C_{ijkl} &= \mathbb{E}_B[M_{ij}M_{kl}] \\
&= \int_{-r}^0 \int_{-r}^0 (r + \min(t_i + u, t_k + v)) db_j(u) db_l(v) \\
&= \int_{-r}^T \left( \int_{-r}^0 \mathbf{1}_{[-r, t_i+u]}(s) db_j(u) \right) \left( \int_{-r}^0 \mathbf{1}_{[-r, t_k+u]}(s) db_l(u) \right) ds.
\end{aligned}$$

If  $C$  is singular, then – again due to positive semidefiniteness – there are coefficients  $(\alpha_{ij})_{i,j=1,\dots,n}$ , not all equal to zero, such that

$$\begin{aligned}
0 &= \sum_{i,j,k,l} \alpha_{ij} \alpha_{kl} C_{ijkl} \\
&= \int_{-r}^T \left( \sum_{i,j} \alpha_{ij} \int_{-r}^0 \mathbf{1}_{[-r, t_i+u]}(s) db_j(u) \right)^2 ds \\
&= \int_{-r}^T \left( \sum_{i,j} \alpha_{ij} b_j([s - t_i, 0]) \right)^2 ds,
\end{aligned}$$

where  $[s - t_i, 0]$  should be understood as the empty set for  $s > t_i$ . Putting  $B_j(u) := b_j([u, 0])$  with this convention, the calculation shows that the functions  $B_j(\cdot - t_i)$  are linearly dependent on  $[-r, T]$ . However, the measures  $b_j$  are linearly independent on  $[-r, 0]$  and so are their primitives  $B_j$  on  $[-r, 0]$ , hence on  $[-r, T]$ . The functions  $B_j$  vanish on the interval  $(0, T]$  and an application of the preceding lemma shows the existence of  $t_1 \dots, t_n \in (0, T]$  such that  $B_j(\cdot - t_i)$  is linearly independent, so that this choice proves  $P = 0$  as asserted.  $\square$

**5.4. Remark.** Whether the operator  $Q_T$  is injective on  $L^2([-r, 0])$  or even on  $M([-r, 0])$ , remains an interesting open problem. It is equivalent to the question whether the linear span of the segments  $(W(t+s), 0 \leq s \leq r)$ ,  $0 \leq t \leq T-r$ , of a Brownian trajectory is a.s. dense in  $L^2([0, r])$  or  $C([0, r])$ , respectively. However, we shall only need the result proved above.

For the convergence of the estimator the approximation spaces must satisfy the Jackson- and Bernstein-inequalities, also called direct and inverse estimates.

**5.5. Definition.** A sequence  $(V_n)_{n \geq 1}$  of subspaces of  $L^2([-r, 0])$  with  $\dim V_n = n$  will be called *s-approximating*, if it satisfies the following inequalities with fixed constants  $C_J$  and  $C_B$ , where  $P_n$  denotes the orthogonal projection onto  $V_n$ :

$$\|(\text{Id} - P_n)u\| \leq C_J \|u\|_\alpha n^{-\alpha}, \quad \forall u \in H^\alpha, \alpha \in \{s, 2\}, \quad (5.5)$$

$$\langle Qv_n, v_n \rangle \geq C_B n^{-2} \|v_n\|^2, \quad \forall v_n \in V_n, \quad (5.6)$$

where  $Q$  denotes any covariance operator  $Q_a$  for  $a \in M^-([-r, 0])$  or the operator  $Q_W$  from Remark 4.5.

**5.6. Example.** A whole class of such  $s$ -approximating sequences is provided by the spaces  $V_n$  of splines of order  $m \geq \max(s, 2)$  with  $n + 1$  uniformly spaced knots. It is well known that these spaces satisfy the Jackson-inequality (5.5) for all  $\alpha \leq m$  (compare e.g. [Schumaker81, Thm. 6.27]). To prove the Bernstein-inequality (5.6) twofold partial integration shows that for  $f \in L^2([-r, 0])$

$$\begin{aligned}
\langle Q_W f, f \rangle &= \int_{-r}^0 \int_{-r}^0 (r + \min(t, s)) f(s) f(t) ds dt + \left( \int_{-r}^0 f(s) ds \right)^2 \\
&\geq 2 \int_{-r}^0 \int_{-r}^t (r + s) f(s) ds f(t) dt \\
&= -2 \int_{-r}^0 \int_{-r}^t F(s) ds f(t) dt \\
&= 2 \int_{-r}^0 F(t)^2 dt
\end{aligned}$$

holds with  $F(t) := -\int_t^0 f(s) ds$ .

For a polynomial  $p(x) = \sum_{k=0}^{m-1} a_k x^k$  with primitive  $P(x) = -\int_x^0 p(y) dy$ ,  $x < 0$ , one obtains

$$\begin{aligned}
\int_{-h}^0 p(x)^2 dx &= - \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k a_l (k + l + 1)^{-1} (-h)^{k+l+1}, \\
\int_{-h}^0 P(x)^2 dx &= -h^2 \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \frac{a_k}{k+1} \frac{a_l}{l+1} (k + l + 3)^{-1} (-h)^{k+l+1}.
\end{aligned}$$

Substituting  $\alpha_k := (-h)^k a_k$  in these expressions shows that the quotient of the  $L^2([-h, 0])$ -norms of  $P$  and  $h \cdot p$  is bounded from below by a constant, since it can be minimized independently of  $h$ . Let us now consider a piecewise polynomial  $f$  on  $[-r, 0]$  of degree at most  $m - 1$  with (possible) jumps at  $n - 1$  uniformly spaced interior knots. Then the analogous quotient between the primitive  $F$  of  $f$  on  $[-r, 0]$  and  $f$  itself is bounded from below by a constant times the width of the subdivision  $rn^{-1}$ , which becomes clear as soon as one realizes that the minimal quotient is attained by the function  $f$  that vanishes outside  $[-r, -r + \frac{r}{n}]$  and minimizes the quotient within this interval. Since splines of order  $m$  are special cases of these piecewise polynomials the Bernstein-inequality holds true for them.

Further subspaces which may be used could be those arising from a wavelet multiresolution analysis. However, the proof of the Bernstein-inequality will not always be immediate and one must pay attention to boundary corrections. Unfortunately, the covariance operator  $Q : L^2([-r, 0]) \rightarrow H^2([-r, 0])$  is not surjective so that the usual duality arguments cannot be used.

**5.7. Proposition.** *Assume that  $g \in H^s([-r, 0])$  with  $v_0(g) < 0$ ,  $\|g\|_s \leq S$  is the true parameter,  $s, S > 0$ . Let  $s$ -approximating subspaces  $(V_n)$  be given and*

determine the function  $G_{T,n} \in V_n$  according to

$$\langle Q_T G_{T,n}, v_n \rangle = \langle b_T, v_n \rangle, \quad \forall v_n \in V_n. \quad (5.7)$$

Introduce the random set

$$\mathcal{H} := \{ \|q_g(u-v) - T^{-1}q_T(u,v)\| \leq \frac{1}{2} \|(P_n Q_g|_{V_n})^{-1}\|^{-1} \}.$$

Then  $G_{T,n}$  satisfies

$$\mathbb{E}_g [\|G_{T,n} - g\|^2 \mathbf{1}_{\mathcal{H}}] \lesssim n^{-2s} + n^3 T^{-1}. \quad (5.8)$$

The constant involved is independent of  $n$  and  $T$  and depends  $L^2$ -continuously on the parameter  $g$ .

*Proof.* With regard to Theorem 5.1 set  $H := L^2([-r, 0])$ ,  $A := Q_a$ ,  $A_\eta := T^{-1}Q_T$ ,  $\eta := \|q_g - T^{-1}q_T\|$  and  $f_\delta := T^{-1}b_T$ , which implies  $u_n = G_{T,n}$ . By Proposition 5.3 the system (5.7) is almost surely solvable. The condition in  $\mathcal{H}$  ensures  $\eta \|R_n\| \leq \frac{1}{2}$ ,  $\eta \|R_{n\eta}\| \leq 1$  holds by (5.3) and the following estimate is obtained on  $\mathcal{H}$ :

$$\|g - G_{T,n}\| \leq (2 + 2\|R_n\| \|(\text{Id} - P_n)Q_g\|) \|(\text{Id} - P_n)g\| + \frac{2}{T} \|R_n(Q_T g - b_T)\|. \quad (5.9)$$

By the  $s$ -approximating property of  $(V_n)$  the Jackson-inequality implies  $\|(\text{Id} - P_n)g\| \leq C_J S n^{-s}$ ,  $\|(\text{Id} - P_n)Q_g\| \leq C_J C_Q n^{-2}$  and the Bernstein-inequality implies (cf. (5.4))

$$\|R_n\| = \sup_{\|v_n\|=1} \langle Q_g v_n, v_n \rangle^{-1} \lesssim n^2 \quad (5.10)$$

uniformly in  $g \in L^2([-r, 0])$  for a bounded parameter set in  $L^2([-r, 0])$  (cf. Lemma 4.4 and note the weak compactness of closed bounded subsets of  $\mathcal{M}^-$ ). Therefore estimate (5.9) reduces to

$$\begin{aligned} \|g - G_{T,n}\| &\lesssim 2(1 + n^2 C_Q C_J n^{-2}) S n^{-s} + \frac{2}{T} \|R_n(Q_T g - b_T)\| \\ &\lesssim S n^{-s} + T^{-1} \|R_n(Q_T g - b_T)\|, \end{aligned} \quad (5.11)$$

where the constant does not depend on  $n$  and  $T$  and may be chosen continuously with respect to  $g$ , since  $L^2$ -convergence implies weak convergence of the induced measures.

The last summand is pointwise of order  $n^2$ , but in a stochastic sense the rate  $n^{\frac{3}{2}}$  is obtained. With  $(e_1, \dots, e_n)$  as orthonormal basis of  $V_n$  we obtain by the



selfadjointness of  $Q_g$ , Lemma 3.3 and (5.10)

$$\begin{aligned}
\mathbb{E}_g[\|R_n(Q_T g - b_T)\|^2] &= \mathbb{E}_g\left[\sum_{i=1}^n \langle (P_n Q_g|_{V_n})^{-1} P_n r_T, e_i \rangle^2\right] \\
&= \sum_{i=1}^n \mathbb{E}_g[\langle r_T, (P_n Q_g|_{V_n})^{-1} e_i \rangle^2] \\
&= T \sum_{i=1}^n \langle Q_g (P_n Q_g|_{V_n})^{-1} e_i, (P_n Q_g|_{V_n})^{-1} e_i \rangle \\
&= T \sum_{i=1}^n \langle P_n Q_g (P_n Q_g|_{V_n})^{-1} e_i, (P_n Q_g|_{V_n})^{-1} e_i \rangle \\
&= T \sum_{i=1}^n \langle e_i, R_n e_i \rangle \\
&\lesssim T n^3.
\end{aligned}$$

An application of the general inequality  $(A + B)^2 \leq 2(A^2 + B^2)$  yields

$$E_g[\|G_{T,n} - g\|^2 \mathbf{1}_{\mathcal{H}}] \lesssim n^{-2s} + T^{-1} n^3$$

with the properties of the constant as asserted.  $\square$

**5.8. Theorem.** *Assume the hypotheses of the preceding proposition with  $s > \frac{1}{2}$  and with  $G_{T,n}$  rescaled to  $\frac{S}{\|G_{T,n}\|} G_{T,n}$  in the case  $\|G_{T,n}\| > S$ . Then choosing  $n(T) \sim T^{\frac{1}{2s+3}}$  yields the uniform rate  $T^{-\frac{2s}{2s+3}}$  for the mean squared error. More precisely, for  $\delta > 0$  the following holds:*

$$\sup_{\substack{\|g\|_s \leq S \\ v_0(g) \leq -\delta}} \mathbb{E}_g[\|g - G_{T,n(T)}\|^2] \lesssim T^{-\frac{2s}{2s+3}}. \quad (5.12)$$

*Proof.* The rescaling always produces a better approximation for  $g$  since we know a priori that  $\|g\| \leq \|g\|_s \leq S$  holds and since  $L^2([-r, 0])$  provides a scalar product, so that this may be considered as an obvious three-point problem in the Euclidean space  $\mathbb{R}^2$ .

The choice of  $n(T)$  gives for the estimate (5.8) asymptotically

$$\mathbb{E}_g[\|g - G_{T,n(T)}\|^2 \mathbf{1}_{\mathcal{H}}] \lesssim T^{-\frac{2s}{2s+3}},$$

the constant depending  $L^2$ -continuously on the parameter  $g$ . Since the Sobolev ball  $\{f \in H^s([-r, 0]) \mid \|f\|_s \leq S\}$  is compact in  $L^2([-r, 0])$  and the set  $\{f \in L^2([-r, 0]) \mid v_0(f) \leq -\delta\}$  is closed in  $L^2([-r, 0])$  by Theorem 2.6, the supremum of the constants over the compact intersection of these sets remains bounded and the statement

$$\sup_{\substack{\|g\|_s \leq S \\ v_0(g) \leq -\delta}} \mathbb{E}_g[\|g - G_{T,n(T)}\|^2 \mathbf{1}_{\mathcal{H}}] \lesssim T^{-\frac{2s}{2s+3}}$$

is obtained.

By Theorem 3.1 the estimate  $\mathbb{E}_g[\|q_g - T^{-1}q_T\|^{2m}] \lesssim T^{-m}$  holds and by (5.10) the asymptotic estimate  $\|R_{n(T)}\|^{2m} \lesssim n(T)^{4m} \sim T^{\frac{4m}{2s+3}}$  holds, so that

$$\limsup_{T \rightarrow \infty} T^{\frac{(2s-1)m}{2s+3}} \mathbb{E}_g[\|q_g - T^{-1}q_T\|^{2m} \|R_{n(T)}\|^{2m}] < \infty.$$

An application of the generalized Chebyshev inequality yields

$$\limsup_{T \rightarrow \infty} T^{\frac{(2s-1)m}{2s+3}} \mathbb{P}_g(\|q_g - T^{-1}q_T\| \|R_{n(T)}\| > \frac{1}{2}) < \infty.$$

Applying the condition  $s > \frac{1}{2}$ , the considered probability tends to zero faster than any polynomial in  $T$ . By construction, the norm of  $G_{T,n}$  is bounded by  $S$  so that in particular for the complement  $\mathcal{H}^C$  of  $\mathcal{H}$

$$\mathbb{E}_g[\|g - G_{T,n}\|^2 \mathbf{1}_{\mathcal{H}^C}] \lesssim \mathbb{P}_g(\mathcal{H}^C) \lesssim T^{-\frac{2s}{2s+3}}$$

holds with a constant that can again be chosen to depend  $L^2$ -continuously on the parameter  $g$ . The same compactness argument as above applies and the theorem is obtained by putting the separate estimates on  $\mathcal{H}$  and  $\mathcal{H}^C$  together.  $\square$

### 5.9. Remark.

- The restriction  $s > \frac{1}{2}$  is used for controlling the probability of the set  $\mathcal{H}^C$ . It is not clear whether the upper bound holds already for  $s > 0$ . The proof of the lower bound works for all  $s > 0$ .
- The estimator  $G_{T,n}$ , constructed in Theorem 5.4, is approximately the maximum-likelihood estimator of  $g$  if  $g$  is an element of  $V_n$ . This follows from the fact that  $\frac{1}{T} \log \Lambda_T$  for large  $T$  is well approximated by  $\langle T^{-1}b_T, g \rangle - \frac{1}{2} \langle T^{-1}Q_T g, g \rangle$  (cf. Theorem 2.2). This function does attain its maximum on the finite-dimensional space  $V_n$  due to the compactness of balls in  $V_n$  and the decay for  $\|g\| \rightarrow \infty$ . Setting the derivative with respect to the functions  $g \in V_n$  to be zero leads exactly to the equation  $b_T - Q_T G_{T,n} = 0$  for the estimator  $G_{T,n}$ . The proposed nonparametric estimator may thus be interpreted as a (quasi-)maximum-likelihood estimator for a misspecified model, where the misspecification, due to the approximation properties of  $V_n$ , becomes smaller with increasing  $n$ .
- An implementation of  $G_{T,n}$  on a computer can deal with discrete observations only. Moreover, round-off errors in the calculations will occur. However, as long as these errors have the order  $T^{-1/2}$ , the asymptotic rate of convergence remains the same. This is due to the fact that the Ritz-Galerkin projection method genuinely deals with errors in the data. Hence a robust implementation seems feasible.

## 6 Asymptotic lower bound

It will now be shown that the rate of convergence found in Theorem 5.8 is asymptotically optimal. The proof of this lower bound is based upon the so called Assouad cube using wavelet techniques. First, a lemma simplifying the likelihood ratio in the stationary case will be proved.

**6.1. Lemma.** *If  $a_1$  and  $a_2$  are  $\mathcal{M}^-$ -measures then the likelihood ratio  $\Lambda_T(a_1, a_2)$  under  $\mathbb{P}_{a_2}$  of the distributions of the stationary solution  $X$  in  $C([0, T])$  according to  $\mathbb{P}_{a_1}$  and  $\mathbb{P}_{a_2}$  respectively is given by*

$$\begin{aligned} \log \Lambda_T(a_1, a_2, X) &= \log \Lambda(a_1, a_2, X(0)) \\ &\quad - \frac{1}{2} \langle \tilde{Q}_T(a_1 - a_2), a_1 - a_2 \rangle + \langle \tilde{r}_T, a_1 - a_2 \rangle, \end{aligned} \quad (6.1)$$

where  $\Lambda(a_1, a_2, X(0))$  denotes the likelihood ratio of  $X(0)$  under  $\mathbb{P}_{a_1}$  and  $\mathbb{P}_{a_2}$  and  $\tilde{Q}_T$  and  $\tilde{r}_T$  are formed using  $\Xi(t) := \mathbb{E}[X(t) | \sigma(X(s), 0 \leq s \leq T)]$  instead of  $X(t)$  in the definition of  $Q_T$  and  $r_T$ .

*Proof.* Since  $X$  is a Gaussian process the drift term in the stochastic delay equation (1.2) is Gaussian and the likelihood ratio (2.6) is simplified:

$$\Lambda_T(X, X(0) + W) = \exp(\langle \tilde{b}_T, a \rangle - \frac{1}{2} \langle \tilde{Q}_T a, a \rangle).$$

For this result use [Thm 7.15, LipShir77], the linearity of conditional expectations

$$\mathbb{E}[\int_{-r}^0 X(t+s) da(s) | \sigma(X(u), 0 \leq u \leq T)] = \int_{-r}^0 \Xi(t+s) da(s)$$

and substitute  $\Xi$  for  $X$  in the calculations of the proof for Theorem 2.2 ( $\tilde{b}_T$  formed analogously). With the notation  $X_i$  for  $X$  with parameter  $a_i$ , one obtains under the law  $\mathbb{P}_{a_2}$ :

$$\begin{aligned} \log \Lambda_T(a_1, a_2, X) &= \log \Lambda_T(X_1, X_1(0) + W) + \log \Lambda_T(X_1(0) + W, X_2(0) + W) \\ &\quad - \log \Lambda_T(X_2, X_2(0) + W) \\ &= \langle \tilde{b}_T, a_1 - a_2 \rangle - \frac{1}{2} \langle \tilde{Q}_T a_1, a_1 \rangle + \frac{1}{2} \langle \tilde{Q}_T a_2, a_2 \rangle \\ &\quad + \log \Lambda(a_1, a_2, X(0)) \\ &= \langle \tilde{Q}_T a_2, a_1 - a_2 \rangle + \langle \tilde{r}_T, a_1 - a_2 \rangle - \frac{1}{2} \langle \tilde{Q}_T a_1, a_1 \rangle \\ &\quad + \frac{1}{2} \langle \tilde{Q}_T a_2, a_2 \rangle + \log \Lambda(a_1, a_2, X(0)) \\ &= \log \Lambda(a_1, a_2, X(0)) \\ &\quad - \frac{1}{2} \langle \tilde{Q}_T(a_1 - a_2), a_1 - a_2 \rangle + \langle \tilde{r}_T, a_1 - a_2 \rangle, \end{aligned}$$

where the decomposition (3.6) and the selfadjointness of  $\tilde{Q}_T$  were used.  $\square$

**6.2. Theorem.** For  $s > 0$ ,  $S > 0$  and  $\delta > 0$  the statement

$$\inf_{G_T} \sup_{\substack{\|g\|_s \leq S \\ v_0(g) \leq -\delta}} \mathbb{E}_g[\|g - G_T\|^2] \gtrsim T^{-\frac{2s}{2s+3}} \quad (6.2)$$

holds, where the infimum is taken over all  $\sigma(X(t), 0 \leq t \leq T)$ -measurable mappings with values in  $L^2([-r, 0])$ .

*Proof.* Take  $g_0 \in H^s([-r, 0])$  with  $\|g_0\| < S$  and  $v_0(g_0) < -\delta$ . Let  $(\psi_{j,k})$  be a compactly supported wavelet basis of  $L^2(\mathbb{R})$  which is  $s$ -regular. Denote by  $R_j$  a maximal subset of  $\mathbb{Z}$  such that  $\text{supp } \psi_{jk} \subset [-r, 0]$  and  $\text{supp } \psi_{jk} \cap \text{supp } \psi_{jk'} = \emptyset$  hold for  $k, k' \in R_j$  and  $k \neq k'$ . The cardinality of  $R_j$ , denoted by  $S_j$ , grows like  $2^j$  with increasing  $j$ , hence  $S_j \sim 2^j$ . For  $\varepsilon \in \{-1, 1\}^{R_j}$  set

$$g_\varepsilon := g_0 + \gamma \sum_{k \in R_j} \varepsilon_k \psi_{jk}, \quad (6.3)$$

where  $\gamma = \gamma(j, T)$  is – for the moment – arbitrary under the condition  $\gamma \lesssim 2^{-j(s+\frac{1}{2})}$  such that  $\|g_\varepsilon\|_s \leq S$  is satisfied. Moreover, this implies  $\gamma 2^{j/2} \rightarrow 0$  for  $j \rightarrow \infty$  so that  $\|g_\varepsilon - g_0\| \rightarrow 0$  and hence  $v_0(g_\varepsilon) < -\delta$  holds for sufficiently large  $j$ . Only these values of  $j$  shall be considered in the sequel.

Let us now quote [HKPT98, Lemma 10.2] with an adapted notation:

**6.3. Lemma (Assouad).** Let  $\delta := \frac{1}{2} \inf_{\varepsilon \neq \varepsilon'} \|g_\varepsilon - g_{\varepsilon'}\|$ . For  $\varepsilon = (\varepsilon_i) \in \{-1, 1\}^{R_j}$  put  $\varepsilon_{k^*} = (\varepsilon'_i)$  such that

$$\varepsilon'_i := \begin{cases} \varepsilon_i, & \text{if } i \neq k \\ -\varepsilon_i, & \text{if } i = k \end{cases}.$$

If there exist  $\lambda, p > 0$  such that for the likelihood ratio  $\Lambda_T$

$$\mathbb{P}_{g_\varepsilon}(\Lambda_T(g_{\varepsilon_{k^*}}, g_\varepsilon) > e^{-\lambda}) \geq p, \quad \forall \varepsilon, T,$$

then for any  $\sigma(X(t), 0 \leq t \leq T)$ -measurable estimator  $G_T$  the following lower bound is obtained:

$$\max_{g_\varepsilon} \mathbb{E}_{g_\varepsilon}[\|G_T - g_\varepsilon\|^2] \geq \frac{1}{2} S_j \delta^p e^{-\lambda} p. \quad (6.4)$$

In our case, obviously  $\delta$  equals  $\gamma$  and Lemma 6.1 shows that under  $\mathbb{P}_{g_\varepsilon}$  the likelihood ratio is given by (use  $g_\varepsilon - g_{\varepsilon_{k^*}} = \pm 2\gamma \psi_{jk}$ )

$$\begin{aligned} \log \Lambda_T(g_{\varepsilon_{k^*}}, g_\varepsilon) &= \log \Lambda(g_{\varepsilon_{k^*}}, g_\varepsilon, X(0)) \\ &\quad - \left[ \frac{4\gamma^2}{2} \langle \tilde{Q}_T \psi_{jk}, \psi_{jk} \rangle \pm 2\gamma \langle \tilde{r}_T, \psi_{jk} \rangle \right]. \end{aligned}$$

Denote the expression inside the squared brackets by  $Z = Z(T, j)$  and note that the remainder is the log-likelihood ratio of  $X(0)$ , whose moments may

be uniformly bounded for arbitrary weakly compact sets  $K$  in  $\mathcal{M}^-$ , since its distribution only depends upon  $q(0)$  and  $\{q_a(0) \mid a \in K\}$  is compact in  $(0, \infty)$ . By compactness of the considered set of parameters  $g$  the asymptotic behaviour of the log-likelihood ratio will consequently be governed by  $Z$ .

The general inequalities  $\text{Var}[X+Y] \leq 2(\text{Var}[X]+\text{Var}[Y])$  and  $\text{Var}[\mathbb{E}[X \mid \mathcal{F}]] \leq \text{Var}[X]$  show in combination with Corollary 3.2 and Lemma 3.3

$$\begin{aligned} \mathbb{E}_{g_\varepsilon}[Z] &= 2T\gamma^2 \langle Q\psi_{jk}, \psi_{jk} \rangle, \\ \text{Var}_{g_\varepsilon}[Z] &\leq 8\gamma^4 \mathbb{E}_{g_\varepsilon}[\langle (Q_T - TQ)\psi_{jk}, \psi_{jk} \rangle^2] + 8T\gamma^2 \langle Q\psi_{jk}, \psi_{jk} \rangle \\ &\leq 8T^2\gamma^4 \mathbb{E}[\|q - T^{-1}q_T\|_\infty^2] \|\psi_{jk}\|_{L^1}^4 + 8T\gamma^2 \langle Q\psi_{jk}, \psi_{jk} \rangle \\ &\lesssim T\gamma^4 2^{-2j} + T\gamma^2 \langle Q\psi_{jk}, \psi_{jk} \rangle. \end{aligned}$$

At this point the convergence of  $T^{-1}q_T$  in  $C^\alpha$  is crucial. If  $\|\psi_{jk}\|_{L^2}$  instead of  $\|\psi_{jk}\|_{L^1}$  were used, the proof would only work for  $s > \frac{1}{2}$  (cf. Remark 5.9). In Corollary 2.7 it was shown that the covariance function  $q$  is Lipschitz continuous, so that one can estimate further using  $|q(t) - q(0)| \leq L|t|$  and  $\int_{-r}^0 \psi_{jk} = 0$ :

$$\begin{aligned} \langle Q\psi_{jk}, \psi_{jk} \rangle &= \int_{-r}^0 \int_{-r}^0 (q(t-s) - q(0))\psi_{jk}(t)\psi_{jk}(s) ds dt \\ &\leq L \int_{-r}^0 \int_{-r}^0 |t-s| |\psi_{jk}(t)| |\psi_{jk}(s)| ds dt \\ &= L \int_{\mathbb{R}} \int_{\mathbb{R}} |2^{-j}(t-s)| |2^{j/2}\psi(t-k)| |2^{j/2}\psi(s-k)| 2^{-2j} ds dt \\ &= 2^{-2j} L \int_{\mathbb{R}} \int_{\mathbb{R}} |t-s| |\psi(t)| |\psi(s)| ds dt \\ &\lesssim 2^{-2j}. \end{aligned}$$

The expectation of  $\log \Lambda_T$  remains therefore bounded if  $\gamma \lesssim T^{-\frac{1}{2}} 2^j$  holds asymptotically for  $T, j \rightarrow \infty$ . Moreover, the Chebyshev inequality then shows that  $\log \Lambda_T$  stays with uniform positive probability in the neighbourhood of its expectation if the variance remains bounded, i.e. in addition  $\gamma \lesssim T^{-\frac{1}{4}} 2^{\frac{j}{2}}$  must be satisfied. Now choose  $j = j(T)$  such that  $2^j \sim T^{\frac{1}{2s+3}}$  and  $\gamma \sim 2^{-j(s+\frac{1}{2})}$  as demanded at the beginning. This choice implies  $\gamma \sim T^{-\frac{1}{2}} 2^j$  and  $\gamma \sim 2^{\frac{j}{2}} T^{-\frac{s+1}{2s+3}} \lesssim 2^{\frac{j}{2}} T^{-\frac{1}{4}}$ . All conditions on  $\gamma$  are consequently satisfied and Assouad's lemma gives the asymptotic estimate

$$\max_{g_\varepsilon} \mathbb{E}_{g_\varepsilon}[\|G_T - g_\varepsilon\|^2] \gtrsim 2^j \gamma^2 \sim T^{-\frac{2s}{2s+3}}.$$

Since the constant depends continuously on  $g_\varepsilon$  and the parameter space is compact, the theorem has been proved.  $\square$

Strictly speaking, the proposed estimator  $G_{T,n}$  depends on  $(X(t), -r \leq t \leq T)$ , but due to stationarity it may equivalently be defined to depend upon  $(X(t), 0 \leq t \leq T+r)$ . This additional segment of length  $r$  is asymptotically negligible so that  $G_{T,n}$  is really asymptotically optimal in the above minimax sense.

## References

- Adams75:** Robert A. Adams, *Sobolev Spaces*, Academic Press, 1975
- Baumeister87:** Johann Baumeister, *Stable Solution of Inverse Problems*, Vieweg Advanced Lectures in Math., 1987
- DaPrZab92:** Giuseppe Da Prato, Jerzy Zabczyk, *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, 1992
- DGLW95:** Odo Diekmann, Stephan van Gils, Sjoerd Verduyn Lunel, Hans-Otto Walther, *Delay Equations*, Applied math. sciences 110, Springer, 1995
- GushKüch99:** Alexander A. Gushchin, Uwe Küchler, *Asymptotic Inference for a Linear Stochastic Differential Equation with Time Delay*, Bernoulli 5, No. 6, 1059-1098, 1999
- GushKüch00:** Alexander A. Gushchin, Uwe Küchler, *On stationary solutions of delay differential equations driven by a Lévy process*, Stochastic Proc. Appl. 88, 195-211, 2000
- HaleLun93:** Jack K. Hale, Sjoerd M. Verduyn Lunel, *Introduction to Functional differential equations*, Springer, 1993
- Hämarik83:** Uno A. Hämarik, *Projection methods for regularizing ill-posed problems*, in Russian, Tr. Vychisl. Tsentra Tartusk. Un-ta 50, 69-90, 1983
- HKPT98:** Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, Alexander Tsybakov, *Wavelets, Approximation, and Statistical Applications*, Lecture Notes in Stat. 129, Springer, 1998
- IbrHas81:** Il'dar A. Ibragimov, Rafail Z. Has'minskiĭ, *Statistical Estimation: Asymptotic Theory*, Springer, 1981
- KarShr99:** Ioannis Karatzas, Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Math., Springer, 2nd Edition, 1999
- KüchKuto00:** Uwe Küchler, Yuri A. Kutoyants, *Delay estimation for some stationary diffusion-type processes*, Scand. J. Statist. 27, 1-10, to appear 2000
- KüchMen92:** Uwe Küchler, Beatrice Mensch, *Langevin's stochastic differential equation extended by a time-delayed term*, Stochastics Stochastics Rep. 40, No. 1/2, 23-42, 1992
- KuMoBo92:** Yuri A. Kutoyants, T. Mourid, D. Bosq, *Estimation paramétrique d'un processus de diffusion avec retards*, Ann. Inst. H. Poincaré 28, 95-106, 1992

- KutoMou94:** Yury A. Kutoyants, Tahar Mourid, *Estimation par la Distance Minimale pour un Processus de Type Diffusion avec Retards*, Pub. Inst. Stat. Univ. Paris XXXVIII, fasc. 2, 3-18, 1994
- LipShir77:** Robert S. Liptser, Albert N. Shiryaev, *Statistics of Random Processes I*, Appl. of Mathematics 5, Springer, 1st Edition, 1977
- Lunel99:** Sjoerd Verduyn Lunel, *Identification of parameters of functional differential equations*, unpublished manuscript, 1999
- Mao97:** Xuerong Mao, *Stochastic Differential Equations and their Applications*, Harwood Publishing Chichester, 1997
- MatPer00:** Peter Mathé, Sergei V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales*, Preprint No. 548, Weierstraß-Institut Berlin (WIAS), 2000
- Mohammed84:** Salah Mohammed, *Stochastic functional differential equations*, Pitman, 1984
- MSW86:** Salah Mohammed, Michael Scheutzow, Heinrich von Weizsäcker, *Hyperbolic state space decomposition for a linear stochastic delay equation*, SIAM J. Control and Optimization 24, No. 3, 543-551, May 1986
- NusPer99:** Michael Nussbaum, Sergei V. Pereverzev, *The degrees of ill-posedness in stochastic and deterministic noise models*, Preprint No. 509, Weierstraß-Institut Berlin (WIAS), 1999
- RevYor99:** Daniel Revuz, Marc Yor, *Continuous martingales and Brownian motion*, Grundlehren 293, Springer, 3rd Edition, 1999
- Rothkirch93:** Ulf Rothkirch, *Schätzung von Parametern in linearen stochastischen Funktionaldifferentialgleichungen*, Dissertation, Universität Jena, Germany, 1993
- Rudin87:** Walter Rudin, *Real and complex analysis*, Mc-Graw-Hill, 3rd Edition, 1987
- Sakhnovich96:** Lev A. Sakhnovich, *Integral Equations with Difference Kernels on Finite Intervals*, Operator Theory Adv. & Appl. 84, Birkhäuser, 1996
- Schumaker81:** Larry L. Schumaker, *Spline Functions*, John Wiley & Sons, 1981
- Werner95:** Dirk Werner, *Funktionalanalysis*, Springer, 1st Edition, 1995