Bernoulli **11**(1), 2005, 67–102

Adaptive estimation for affine stochastic delay differential equations

MARKUS REISS

Weierstrass Institute of Applied Analysis and Stochastics (WIAS), Mohrenstrasse 39, D-10117 Berlin, Germany. E-mail: mreiss@wias-berlin.de

For stationary solutions of the affine stochastic delay differential equation

$$\mathrm{d}X(t) = \left(\gamma_0 X(t) + \gamma_r X(t-r) + \int_{-r}^{0} X(t+u)g(u)\mathrm{d}u\right)\mathrm{d}t + \sigma\,\mathrm{d}W(t),$$

we consider the problem of nonparametric inference for the weight function g and for γ_0, γ_r from the continuous observation of one trajectory up to time $T > 0$. For weight functions in the scale of Besov spaces $B^s_{p,1}$ and L^ρ-type loss functions, convergence rates are established for long-time asymptotics. The estimation problem is equivalent to an ill-posed inverse problem with error in the data and unknown operator. We propose a wavelet thresholding estimator that achieves the rate $(T/\log T)^{-s/(2s+3)}$ under certain restrictions on p and ρ. This rate is shown to be optimal in a minimax sense.

Keywords: Besov space; ill-posed inverse problem; minimax rates; spatial adaptivity; wavelet thresholding

1. Introduction

1.1. The model

Stochastic delay differential equations (SDDEs) appear naturally in the description of many processes: for example, in population dynamics with a time lag due to an age-dependent birth rate (Scheutzow 1984), in economics where a certain 'time to build' is needed (Kydland and Prescott 1982), in finance (Hobson and Rogers 1998) and in many engineering applications; see Kolmanovskii and Myshkis (1992) for a broad range of models appearing in applications and their mathematical analysis. They are also obtained as continuous-time limits of time series models. Among the huge variety of types of equations, the so-called affine stochastic delay differential equations form the fundamental class. They generalize the Langevin equation leading to the Ornstein–Uhlenbeck process and appear as continuous-time limits of linear autoregressive schemes. A general scalar affine SDDE is of the form

$$\begin{aligned} \mathrm{d}X(t) &= \left(\int_{-r}^{0} X(t+u)\mathrm{d}a(u)\right)\mathrm{d}t + \sigma\,\mathrm{d}W(t), \qquad t \geqslant 0, \\ X(u) &= F(u), \qquad u \in [-r, 0]. \end{aligned} \tag{1.1}$$

The drift coefficient depends linearly on the past trajectory $(X(u), u \in [t-r, t])$ by means of an integration with respect to the finite signed Borel measure a on $[-r, 0]$. The values r and σ are supposed to be positive and $(W(t), t \geqslant 0)$ denotes a standard Wiener process adapted to a filtration $(\mathcal{F}_t)_{t\geqslant 0}$, satisfying the usual conditions. In order to ensure well-posedness of the differential equation, a whole initial function F independent of the Wiener process is prescribed. The Langevin equation without memory effect is obtained if a is taken to be a point measure at zero.

The asymptotic properties and the existence of stationary solutions for affine SDDEs, even with more general driving processes, have been studied in detail by Mohammed *et al.* (1986) and Gushchin and Küchler (2000). Our goal here is to estimate the weight measure a nonparametrically from the observation $(X(t), t \in [-r, T])$ of one realization of a stationary solution to (1.1). In many applications the drift term consists of both instantaneous feedback and distributed delayed feedback. Having nonparametric estimation and these kinds of models in mind, our parameter class should thus contain weight measures a with a Lebesgue density except for a possible point mass at zero. Our estimation procedure will more generally cope with weight measures having a Lebesgue density in the interior and possible point masses at both interval end-points $-r$ and 0.

1.2. The estimation procedure

The main idea for the construction of the estimator is that the weight measure a solves the integral equation

$$Q_a a(t) := \int_{-r}^{0} q_a(t-u)\mathrm{d}a(u) = q_a'(-t), \qquad t \geqslant 0, \tag{1.2}$$

where the function q_a denotes the covariance function of the process X. From observations up to time T we can construct estimators of the kernel of the operator Q_a and of the data q_a' and solve the integral equation for a. We shall show that Q_a is regularizing in the manner of twofold integration independently of the weight a such that the solution of equation (1.2) does not depend continuously on the data. We face an ill-posed inverse problem of degree 2 with errors occurring in the operator and on the right-hand side.

A linear estimation technique for L^2-risk and weight functions in L^2-Sobolev balls of regularity $s > \frac{1}{2}$ was been presented in Reiss (2002), which is close in spirit to the work by Efromovich and Koltchinskii (2001). The ill-posed inverse problem was solved by the Galerkin projection method, and a minimax risk of order $T^{-s/(2s+3)}$ for observation times $T \to \infty$ was established. Here, we strive for adaptive estimation, that is, we do not suppose the regularity of the unknown weight function to be known and we automatically adapt to spatial inhomogeneity of the function. Moreover, we allow for more general L^ρ-loss, $\rho \in (1, \infty)$. As usual in adaptive estimation theory, we are led to consider density functions in Besov spaces $B^s_{p,\alpha}([-r, 0])$ and to use nonlinear approximation techniques. Under suitable conditions on p and ρ we shall find for our adaptive estimator an asymptotic risk of order $(T/\log T)^{-s/(2s+3)}$, which will be shown to be minimax with respect to the Besov classes considered.

For the construction of the estimator, wavelet thresholding techniques in a suitable image domain are used. Our approach is related to the wavelet–vaguelette decomposition and vaguelette–wavelet decomposition methods proposed by Donoho (1995) and Abramovich and Silverman (1998). In fact, the latter paper presents the main idea: we first threshold the wavelet coefficients and then invert the approximation of equation (1.2). Since our operator is not exactly known (and for each observation different), the inversion should not be performed by calculating the corresponding vaguelettes exactly, but rather by applying a numerical inversion algorithm. For this we can allow for numerical errors up to the order of the statistical error in the first step and even rely on adaptive procedures; see Cohen *et al.* (2004) for a suitable algorithm.

Mathematically speaking, we denoise the data in a certain Sobolev space along the lines of the abstract results obtained by Kerkyacharian and Picard (2000) for heteroscedastic noise. For the right choice of the thresholding level and for theoretical purposes the mapping properties of the covariance operator have to be studied in detail.

1.3. Related approaches

Except for Reiss (2002), statistical inference for the weight measure in affine SDDEs has so far focused on parametric models (see, for example Kutoyants *et al.* 1992; Gushchin and Küchler 1999; 2003), where for sufficiently smooth parametrizations of the weight measure $a = a_\theta$ the classical local asymptotic normality property with rate $T^{-1/2}$ holds under stationarity assumptions. On the other hand, nonparametric and even adaptive estimation of the drift coefficient b in ergodic diffusions

$$dX(t) = b(X(t))dt + \sigma dW(t), \qquad t \in [0, T],$$

is well established (Hoffmann 1999; Dalalyan 2001). Although the estimation problem has a similar structure, under recurrency conditions the minimax rate for estimating drift functions b of regularity s is $T^{-s/(2s+1)}$, indicating a close relationship with classical regression estimation. In our SDDE case the worse, because smaller, exponent $s/(2s+3)$ can be explained intuitively by the presence of an integration in the drift term, which leads to additional smoothing of the observation and thus makes the inference more difficult.

More correctly, the interpretation via the ill-posed problem (1.2) and the framework developed by Nussbaum and Pereverzev (1999) allows a deeper understanding of this rate. To give a rough summary, the model consists of a linear operator K on $L^2([0, 1])$ that is regularizing of degree τ and a centred Gaussian noise Γ on $L^2([0, 1])$ with regularity γ ($\gamma = 0$ corresponds to white noise, $\gamma = 1$ to Brownian motion, and so on). Under suitable conditions their main result states that functions f from a Sobolev ball of regularity s can be estimated from the observation $Y_\varepsilon = Kf + \varepsilon\Gamma$ with an L^2-risk of order $\varepsilon^{2s/(2s+2(\tau-\gamma)+1)}$. The values $\tau = \gamma = 0$ correspond to the classical white noise model for direct observations, which should still be called statistically ill-posed because convergence rates can only be ensured under compactness assumptions on the function class. Translated to our situation, we formally set $\varepsilon = T^{-1/2}$, $K = Q_a$ and $\tau = 2$ and the noise process has regularity $\gamma = 1$ (Theorem 3.1) yielding the exponent $s/(2s + 2(2 - 1) + 1)$, though it should be emphasized that neither the conditions such as exact knowledge of the operator and the Gaussian noise structure nor the methods of proof using singular value decomposition can be suitably adapted.

Ill-posed problems with stochastic error structure have recently attracted increasing attention and the case of weight estimation in SDDEs provides a good example from statistical inference with the special feature of a merely approximately known operator and dependent non-Gaussian noise structure. The archetype of inverse problems in statistics is density deconvolution, where ill-posedness arises due to the smoothness of the contaminating noise density. Our problem might be compared to the case of an error in this density; see Butucea and Matias (2003) for a study in the case of an unknown scaling parameter. For idealized situations, like the Gaussian model sketched above, adaptive estimation procedures based on wavelet thresholding or general oracle inequalities have been obtained; see, for instance, Donoho (1995), Kalifa and Mallat (2003), Cavalier and Tsybakov (2002) or Cavalier *et al.* (2002).

Inverse problems with approximately known operator have already been considered in a deterministic setting by Hämarik (1983) and recently in a stochastic framework by Efromovich and Koltchinskii (2001) and Cavalier and Hengartner (2002). The latter, however, suppose all unknown operators to have the same eigenfunctions and can thus transfer their problem to a sequence space model. Although the operators Q_a are of convolution type, their kernels are not periodic and their eigenfunctions are not given by the Fourier basis, but depend on the weight a such that this approach is not feasible in our case. The general procedure of Efromovich and Koltchinskii (2001) most closely resembles the way we proceed. They also use suitable basis functions and establish the same risk bounds as in the case of known operators whenever the level of the noise in the operator is not larger than that of the noise in the data. We also recover the same minimax rates, although we do not dispose of a training sample and the two noise sources are correlated. The main difference in the estimation procedure is that we use a nonlinear projection method and can thus adapt to spatial inhomogeneity, whereas Efromovich and Koltchinskii (2001) employ a linear Lepski method.

1.4. Structure and notation

In Section 2 we introduce the theory of affine SDDEs and their stationarity behaviour and present results on the mapping properties of the covariance operator. Section 3 is devoted to the construction of the estimator and a discussion of its properties. In Section 4 we assess the optimality of our estimator by the minimax approach. The proofs of the statements are deferred to Sections 5 to 7, Section n providing the proofs for Section $n-3$. The Appendix presents some essentials on function spaces and wavelet bases.

Let us fix some notation. $\mathbb{P}_a(\cdot)$ and $\mathbb{E}_a[\cdot]$ denote the probability measure and the expectation operator depending on the parameter a. By $\mathcal{F}_T^X$ we denote the σ-field generated by the process X up to time T. The space of continuous (or p-integrable) functions on the interval I is denoted by $C(I)$ ($L^p(I)$). The space of finite signed Borel measures on I is written as $M(I)$ and equipped with the total variation norm norm $\|\cdot\|_{\mathrm{TV}}$. δ_x is the Dirac measure in x and $g \circ \lambda$ denotes the measure with Lebesgue density $g \in L^1$. Usually, the density g is identified with the measure $g \circ \lambda$ and thus operators acting on measures are considered to act on the densities themselves. For $f \in C(I)$ and $\mu \in M(I)$ we introduce the dual pairing $\langle f, \mu \rangle := \int_I f \, d\mu$. The cardinality of a set M is denoted by $|M|$. Finally, the symbol $A(T) \lesssim B(T)$ means that $A(T)$ is bounded by a multiple of $B(T)$ independently of T, that is, $A(T) = O(B(T))$ in the O-notation. Equally, $A(T) \gtrsim B(T)$ stands for $B(T) \lesssim A(T)$, and $A(T) \sim B(T)$ for $A(T) \lesssim B(T)$ as well as $A(T) \gtrsim B(T)$.

2. Affine stochastic delay differential equations

For the theory of deterministic delay equations we refer to the monograph by Hale and Verduyn Lunel (1993), while fundamental results on stochastic delay equations can be found in Mohammed (1984) and Mao (1997). If we put $\sigma = 0$ in (1.1), we obtain the deterministic linear delay equation

$$\dot{x}(t) = \int_{-r}^{0} x(t+u) \mathrm{d}a(u), \qquad t \geqslant 0. \tag{2.1}$$

As for linear ordinary differential equations, the ansatz $x(t) = \mathrm{e}^{\lambda t}$ gives rise to a characteristic function the zeros of which determine the long-time behaviour of general solutions x.

Definition 2.1. *The characteristic function associated with* (1.1) *is given by*

$$\chi_a(\lambda) := \lambda - \int_{-r}^{0} \mathrm{e}^{\lambda u} \, \mathrm{d}a(u), \qquad \lambda \in \mathbb{C}.$$

The maximal real part of its zeros is denoted by

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

Without loss of generality we shall henceforth assume $\sigma = 1$; otherwise we rescale X and consider $\tilde{X}(t) = \sigma^{-1} X(t)$ instead. Gushchin and Küchler (2003) then prove the following result:

Theorem 2.1. *A stationary solution of the affine SDDE* (1.1) *exists if and only if* $v_0(a) < 0$ *holds. In this case the stationary solution X is unique. It is a centred Gaussian process with (auto)covariance function* $q_a(t) := \mathbb{E}_a[X(0)X(|t|)]$, $t \in \mathbb{R}$, *satisfying*

$$q_a'(t) = \int_{-r}^{0} q_a(t+u) \mathrm{d}a(u) \qquad \textit{for all } t \geqslant 0. \tag{2.2}$$

Its spectral density is given by

$$\hat{q}_a(\xi) := \int_{-\infty}^{\infty} q_a(t) \mathrm{e}^{\mathrm{i}\xi t} \, \mathrm{d}t = \frac{1}{|\chi_a(\mathrm{i}\xi)|^2}, \qquad \xi \in \mathbb{R}. \tag{2.3}$$

Example 2.1. For the point easure $a = \alpha\delta_0$, equation (1.1) reduces to a stochastic ordinary differential equation with the Ornstein–Uhlenbeck process as solution. We obtain $\chi_a(\lambda) = \lambda - \alpha$ and $v_0(a) = \alpha$. For $\alpha < 0$ a stationary solution exists with covariance function $q_a(t) = (2|\alpha|)^{-1}\mathrm{e}^{-|\alpha t|}$ and spectral density $\hat{q}_a(\xi) = (\xi^2 + \alpha^2)^{-1}$.

The law μ_X of the solution process X on the interval $[0, T]$ and the law μ_W of Brownian motion starting at $X(0)$ are mutually absolutely continuous in the canonical space $C([0, T])$. We express the likelihood ratio in terms of sufficient statistics b_T and Q_T that will be of major importance later.

Definition 2.2. *For the solution process X of* (1.1), *define*

$$b_T(u) := \int_0^T X(t+u)\mathrm{d}X(t) \qquad u \in [-r, 0],$$

$$q_T(u, v) := \int_0^T X(t+u)X(t+v)\mathrm{d}t \qquad u, v \in [-r, 0],$$

$$Q_T\mu(u) := \int_{-r}^0 q_T(u, v)\mathrm{d}\mu(v) \qquad u \in [-r, 0], \mu \in M([-r, 0]),$$

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

The operator Q_a is the covariance operator of the stationary process X on $[-r, 0]$, *regarded as an element of* $C([-r, 0])$.

It is understood that for b_T a continuous version in $u \in [-r, 0]$ is chosen, which is possible since the Kolmogorov continuity theorem applies due to the moment bound:

$$\begin{aligned} &\mathbb{E}\left[\left(\int_0^T X(t+u_1)\mathrm{d}W(t) - \int_0^T X(t+u_2)\mathrm{d}W(t)\right)^4\right] \\ &\quad \lesssim \mathbb{E}\left[\left(\int_0^T (X(t+u_1) - X(t+u_2))^2\,\mathrm{d}t\right)^2\right] \\ &\quad \lesssim T^2(u_1 - u_2)^2, \end{aligned} \tag{2.4}$$

which follows from the Burkholder–Davis–Gundy inequality and the uniform Lipschitz continuity of the covariance function q_a; see Proposition 5.1.

Theorem 2.2. *For a deterministic initial function F in* (1.1), *the Radon–Nikodym derivative $\Lambda_T(X, X(0) + W)$ of μ_X with respect to μ_W is given by*

$$\begin{aligned}\Lambda_T(X, X(0)+W) &:= \frac{\mathrm{d}\mu_X}{\mathrm{d}\mu_W}\\ &= \exp\left(\int_0^T\int_{-r}^0 X(t+u)\mathrm{d}a(u)\mathrm{d}X(t) - \frac{1}{2}\int_0^T\left(\int_{-r}^0 X(t+u)\mathrm{d}a(u)\right)^2\mathrm{d}t\right)\\ &= \exp\big(\langle b_T, a\rangle - \tfrac{1}{2}\langle Q_T a, a\rangle\big).\end{aligned}$$

This result is the basis for the maximum-likelihood theory developed by Gushchin and Küchler (1999). Its proof is derived from the Girsanov theorem for diffusion-type processes and the stochastic Fubini theorem; see Lipster and Shiryaev (2001) or Küchler and Sørensen (1997).

The first impulse in defining a nonparametric estimator $\hat{a}_T$ of a would thus be to maximize the likelihood function which amounts to solving the infinite-dimensional equation $Q_T\hat{a}_T = b_T$. However, the empirical covariance operator Q_T need not be invertible, and although the covariance operator Q_a, obtained in the limit, is invertible, its inverse Q_a^{-1} is an unbounded operator, as will be shown later. Hence, we are in a classical nonparametric situation and smoothing methods need to be employed. Our basic idea is to smooth first and then to solve the maximum likelihood equation in terms of the smoothed quantities. The convergences $T^{-1}Q_T \to Q_a$ and $T^{-1}b_T \to Q_a a$ for $T \to \infty$, a consequence of Theorems 6.1 and 3.1, show that in the limit of an infinitely long observation period the weight measure a is always identifiable. Having adapted an asymptotic viewpoint, we proceed by analysing the covariance operator Q_a in detail. From this analysis and the exact convergence properties all subsequent results will be derived.

For the notion of Besov spaces $B^s_{p,\alpha}$ of functions with L^p-regularity s and fine-tuning parameter α we refer to the Appendix. Just recall the identity $B^s_{2,2} = W^{s,2}$ so that the subsequent results are valid, in particular, for the scale of L^2-Sobolev spaces $W^{s,2}$. Küchler and Mensch (1992) show that the covariance function q_a is twice differentiable on $\mathbb{R}\backslash\{0\}$, but its derivative has a jump at zero which implies that, roughly speaking, the covariance operator Q_a is smoothing of order 2, that is, measures with density $B^s_{p,\alpha}$ are mapped to $B^{s+2}_{p,\alpha}$; cf. Theorem 2.3. Now let us consider the images of point masses δ_ρ, $\rho \in [-r, 0]$:

$$Q_a\delta_\rho(u) = q_a(u-\rho) = q_a(|u-\rho|), \qquad u \in [-r, 0].$$

This shows that for values of ρ in the interior $(-r, 0)$ the image $Q_a\delta_\rho \in C([-r, 0])$ has a jump in its first derivative at $u=\rho$, whereas at the boundary $\rho \in \{-r, 0\}$ the image $Q_a\delta_\rho$ is as regular as the covariance function q_a on $[-r, 0]$. As a consequence, the inclusion of the point measures at the boundary will not complicate the estimator and even imply that the covariance operator is onto. In anticipation of the precise mapping properties we introduce suitable spaces of weight measures; just recall that $g \circ \lambda$ denotes the measure with Lebesgue density g on $[-r, 0]$.

Definition 2.3. (i) *Besov scale. For* $s>0$*,* $p \in (1, \infty)$*,* $\alpha \in [1, \infty]$*,* $v<0$*, set*

$$\mathcal{B}^s_{p,\alpha} := \left\{\gamma_0\delta_0 + \gamma_r\delta_{-r} + g \circ \lambda | \gamma_0, \gamma_r \in \mathbb{R},\ g \in B^s_{p,\alpha}([-r, 0])\right\},$$

$$\mathcal{B}^s_{p,\alpha}(v) := \left\{a \in \mathcal{B}^s_{p,\alpha} | v_0(a) \geqslant v\right\}.$$

On $\mathcal{B}^s_{p,\alpha}$ we introduce the norm

$$\|\{\gamma_0\delta_0 + \gamma_r\delta_{-r} + g \circ \lambda\|_{s,p,\alpha} := |\gamma_0| + |\gamma_r| + \|g\|_{B^s_{p,\alpha}([-r,0])}.$$

(ii) *L^p-scale. For $p \in (1, \infty)$, $v < 0$, set*

$$\mathcal{L}^p := \{\gamma_0\delta_0 + \gamma_r\delta_{-r} + g \circ \lambda | \gamma_0, \gamma_r \in \mathbb{R},\ g \in L^p([-r, 0])\},$$

$$\mathcal{L}^p(v) := \{a \in \mathcal{L}^p | v_0(a) \leqslant v\}.$$

On $\mathcal{L}^p$ we introduce the norm

$$\|\gamma_0\delta_0 + \gamma_r\delta_{-r} + g \circ \lambda\|_p := |\gamma_0| + |\gamma_r| + \|g\|_{L^p([-r,0])}.$$

The space $\mathcal{B}^s_{p,\alpha}$ is isomorphic to the tensor product $\mathbb{R}^2 \otimes B^s_{p,\alpha}$ and thus $(\mathcal{B}^s_{p,\alpha}, \|\cdot\|_{s,p,\alpha})$ is a Banach space. The SDDE (1.1) with weight $a \in \mathcal{B}^s_{p,\alpha}$ typically takes the form

$$\mathrm{d}X(t) = \left(\gamma_0 X(t) + \gamma_r X(t-r) + \int_{-r}^0 X(t+u)g(u)\mathrm{d}u\right)\mathrm{d}t + \mathrm{d}W(t), \qquad t \geqslant 0.$$

The set $\mathcal{B}^s_{p,\alpha}(v)$ is a closed subset of $\mathcal{B}^s_{p,\alpha}$, due to $v < 0$ consisting of weights with a uniform mixing behaviour. This follows from a result in Reiss (2002) adapted to more general weight measures. By the same arguments these properties also hold for $\mathcal{L}^p$. The Besov-type weights form the nonparametric class $M(s, p, S, \delta)$ for which our estimator will be shown to be rate-optimal, whereas the space $\mathcal{L}^p$ will merely occur in the context of mapping properties of the covariance operator.

Definition 2.4. *For $s < 0$, $S > 0$, $p \in [1, \infty]$ and $> \delta > 0$, set*

$$M(s, p, S, \delta) := \{a \in \mathcal{B}^s_{p,1}(-\delta) \mid \|a\|_{s,p,1} \leqslant S\}.$$

The choice $\alpha = 1$ in the definition will be discussed in Section 3.3. In Reiss (2001) and Gushchin and Küchler (2003) it was shown that the covariance operator is always one-to-one on $M([-r, 0])$ and maps densities in $L^2([-r, 0])$ to $W^{2,2}([-r, 0])$. We show that for a certain range of Besov spaces the covariance operator is also smoothing of order 2.

Theorem 2.3. *For weight measures a in $\mathcal{B}^s_{p,\alpha}(v)$ and the parameters as before, the covariance operator is a bijective bounded linear operator on the appropriate spaces:*

$$Q_a : \mathcal{B}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}([-r, 0]) \quad \text{and} \quad Q_a : \mathcal{L}^p \to W^{2,p}.$$

In order to obtain upper bounds in a minimax sense for our estimator, we shall need uniform norm bounds in the preceding statement.

Theorem 2.4. *For* $s, S, \delta > 0$, $p \in (1, \infty)$, *the following operator norms are uniformly bounded:*

$$\sup_{a \in M(s,p,S,\delta)} \|Q_a\|_{\mathcal{B}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}} < \infty, \qquad \sup_{a \in M(s,p,S,\delta)} \|Q_a^{-1}\|_{\mathcal{B}^{s+2}_{p,\alpha} \to B^s_{p,\alpha}} < \infty,$$

$$\sup_{a \in M(s,p,S,\delta)} \|Q_a\|_{\mathcal{L}^p \to W^{2,p}} < \infty, \qquad \sup_{a \in M(s,p,S,\delta)} \|Q_a^{-1}\|_{W^{2,p} \to \mathcal{L}^p} < \infty.$$

3. Construction of the estimator

3.1. The general idea

We start by smoothing the statistic b_T adaptively. To this end kernel and wavelet methods are equally applicable, but since an integral equation with this estimator as data has to be solved later, wavelet techniques avoid a second numerical discretization step. For the notation (ψ_λ) of an s-regular wavelet basis on $[-r, 0]$ we refer to the Appendix.

First, we need to clarify the functional nature of the noise in the estimate $T^{-1}b_T$ of $Q_a a$. A good intuition is provided by the decomposition (recall $\sigma = 1$)

$$\begin{aligned} b_T(u) &= \int_0^T \left(X(t+u) \int_{-r}^0 X(t+v) \mathrm{d}a(v) \right) \mathrm{d}t + \int_0^T X(t+u) \mathrm{d}W(t) \\ &= (Q_T \mathrm{a})(u) + \int_0^T X(t+u) \mathrm{d}W(t). \end{aligned} \tag{3.1}$$

Suppose for a moment that $T^{-1}Q_T$ equals Q_a exactly (its kernel is in fact easier to estimate). The error term is then due to the stochastic integral term which is as regular with respect to u as Brownian motion due to the Kolmogorov continuity theorem. Thus, we do not face the classical 'signal plus white noise' model, but rather an integrated form involving the signal and a perturbation by Brownian motion; one may think of recovering the function f from the noisy observation

$$Y(u) = f(u) + \varepsilon W(u), \qquad u \in [-r, 0].$$

However, in our setting the noise is not Gaussian and we are not interested in the signal $f = Q_a a$ itself, but rather in $Q_a^{-1} f = a$. Nevertheless the analoguous ill-posed problem in white noise would be

$$\mathrm{d}Y(u) = \mathrm{D}Q_a a(u)\mathrm{d}t + \varepsilon \mathrm{d}W(u), \qquad u \in [-r, 0],$$

where D denotes differentiation. The operator DQ_a has only one degree of ill-posedness, and this explains the exponent $s/(2s+3)$ in the minimax rate. This formal argument could be substantiated by considering the stochastic differential $\mathrm{d}b_T(u)$ instead of $b_T(u)$. The main drawback of looking at differentials in our situation is that DQ_a is no longer one-to-one.

Taking account of Theorem 2.3, we will minimize the expected L^ρ-loss in estimating the signal $Q_a a$ with minimal $W^{2,\rho}$-loss. Although our noise process is more regular than white noise, the noise is still less regular than the norm we want the signal to be estimated in.

To solve our problem in practice we can adapt the abstract wavelet thresholding results of Kerkyacharian and Picard (2000). Our estimator $\hat{b}_T$ is obtained by expanding b_T in a wavelet basis up to a certain level and only keeping the significant coefficients (hard thresholding).

Definition 3.1. *Let $s_{\max} > 2$ be fixed. With b_T from Definition 2.2, introduce for any multi-index λ the wavelet coefficient*

$$\beta_{\lambda,T} := \left\langle \frac{1}{T} b_T, \psi_\lambda \right\rangle,$$

where $(\psi_\lambda)_\lambda$ is a compactly supported $s_{\max}$-regular wavelet basis in $L^2([-r, 0])$ (see Section A.2). Define the thresholding estimator

$$\hat{b}_T := \hat{b}_{T,J(T),\kappa(T)} := \sum_{|\lambda| \leqslant J(T)} \left(\beta_{\lambda,T} \mathbf{1}_{|\beta_{\lambda,T}| > \kappa_\lambda(T)}\right) \psi_\lambda$$

for a certain resolution level $J(T)$ and thresholds $(\kappa_\lambda(T))_\lambda$.

3.2. Results

How should we choose the threshold values $\kappa_\lambda(T)$? The second term in the decomposition (3.1) gives for $\beta_{\lambda,T}$ the variance estimate $T^{-1}\langle Q_a \psi_\lambda, \psi_\lambda \rangle \sim T^{-1} 2^{-2|\lambda|}$ by Lemma A.5. In Section 3.3 we comment on the choice for a specific sample; here, however, we strive for asymptotically optimal threshold values $\kappa_\lambda(T)$, $T \to \infty$. We note that the overall noise level is $T^{-1/2}$ as expected, and that level-wise the noise intensity decays as $2^{-|\lambda|}$, which indicates that the noise is one degree smoother than white noise. A detailed analysis confirms this picture and even yields uniform Gaussian tail estimates, which cannot be obtained by standard large-deviation techniques.

Theorem 3.1. *Let $(\psi_\lambda)_\lambda$ be a compactly supported 2-regular wavelet basis of $L^2([-r, 0])$ and let $R, \delta, \rho > 0$. Then there is a universal bound $\kappa^* > 0$ such that we obtain uniformly for all weight measures a with $v_0(a) \leqslant -\delta$, $\|a\|_{\mathrm{TV}} \leqslant R$, all multi-indices λ and all sufficiently large T,*

$$\mathbb{P}_a\left(2^{|\lambda|} T^{1/2} |\langle T^{-1} b_T - Q_a a, \psi_\lambda \rangle| \geqslant \frac{\kappa}{2}\sqrt{\log T}\right) \leqslant T^{-3\rho} \qquad \text{for all } \kappa \geqslant \kappa^*. \tag{3.2}$$

We shall therefore set $\kappa_\lambda(T) = \kappa 2^{-|\lambda|} T^{-1/2} \sqrt{\log T}$. As is classical in wavelet methods, we choose the maximal frequency level $J(T)$ such that $J(T) 2^{J(T)}$ is inversely proportional to the variance level T^{-1}.

Proposition 3.2. *Let $s \in (0, s_{\max} - 2]$, $S > 0$, $\rho, p \in (1, \infty)$ and $\delta > 0$ be given satisfying*

$$\frac{1}{p} - \frac{1}{\rho} \leqslant \frac{2}{\rho}\frac{s}{3}. \tag{3.3}$$

Set $2^{J(T)} \sim T/\log T$ and $\kappa_\lambda(T) = \kappa 2^{-|\lambda|} T^{-1/2} \log(T)$ *with* κ *chosen as in Theorem 3.1. Then we obtain the following asymptotic estimate for the estimator* $\hat{b}_T$ *from Definition 3.1:*

$$\sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{b}_T - Q_a a\|_{W^{2,\rho}}] \lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}.$$

In the next step we construct an operator $\hat{Q}_T$ from the observations up to time T, which is close to the true covariance operator. We could, of course, use the results for Q_T from Theorem 6.1, but it is even simpler to use (1.2), that is, $q_a'(t) = Q_a a(-t)$ obtained from (2.2). Writing $q_a(t) = q_a(0) + \int_0^t q_a'(u)\mathrm{d}u$, we can thus determine q_a from $q_a(0)$ and $Q_a a$ and derive an estimator from estimators for these two quantities. The convolution kernel of $\hat{Q}_T$ will have a jump in the first derivative at zero, but will be regular elsewhere. This explains why the mapping properties of $\hat{Q}_T$ resemble those of Q_a. In particular, $\hat{Q}_T$ is with high probability an isomorphism between $\mathcal{L}^\rho$ and $W^{2,\rho}$. This gives our main result.

Theorem 3.3. *Let the parameters* s, S, p *and* δ *be as before. Introduce the integral operator* $\hat{Q}_T : \mathcal{L}^\rho \to W^{2,\rho}$ *with convolution kernel*

$$\hat{q}_T(u) := \frac{1}{T}\int_0^T X(t)^2\,\mathrm{d}t + \int_{-|u|}^0 \hat{b}_T(v),\,\mathrm{d}v, \qquad u \in [-r, r],$$

that is,

$$\hat{Q}_T\mu(t) := \int_{-r}^0 \hat{q}_T(t-u)\mathrm{d}\mu(u) \qquad \textit{for } t \in [-r, 0],\ \mu \in \mathcal{L}^p.$$

Define the estimator $\hat{a}_T$ *by*

$$\hat{a}_T := \begin{cases} \min\Big(S\|\hat{Q}_T^{-1}\hat{b}_T\|_{\mathcal{L}^\rho}^{-1}, 1\Big)\hat{Q}_T^{-1}\hat{b}_T, & \textit{if } \hat{Q}_T : \mathcal{L}^\rho \to W^{2,\rho} \textit{ is invertible}, \\ 0, & \textit{otherwise}. \end{cases}$$

Then the following asymptotic upper bound holds for $T \to \infty$*:*

$$\sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{\mathcal{L}^\rho}] \lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}.$$

3.3. Discussion

Our method differs from the classical wavelet thresholding algorithm due to the ill-posedness involved. Our threshold κ_λ depends on the resolution level $|\lambda|$, because the intensity of the noise coefficients is of order $2^{-|\lambda|}T^{-1/2}$. Furthermore, it is not necessary to impose additionally the linear constraint that the weight lies in $W^{s/(2s+3),\rho}$ because the restriction (3.3) is much stronger than in the classical setting. If we call $\beta = s/3$ the effective smoothness (as in multidimensional smoothing) in our model, then (3.3) is well known in terms of β (Härdle *et al.* 1998, Section 10.4). We have chosen the Besov scale ($B^s_{p,\alpha}$) with $\alpha = 1$ for simpler embedding relations. In fact, $\alpha = p/\rho$ would do, as can be seen from (6.3). It is not known whether this value is the maximal possible.

Example 3.1. What rate of convergence do we obtain for the L^ρ-risk of the weight function $g = -\mathbf{1}_{[-\frac{1}{2},0]}$ with delay $r = 1$? This might be seen as a toy example for estimating a change point or the maximal delay time. The function g lies in $B^1_{1,\infty}([-1, 0])$ and thus, by embedding in $W^{s,\rho}([-1, 0])$ for $s < 1/p$. This shows that linear methods cannot converge faster than at rate $T^{-1/(2+3\rho)}$ whereas, due to $g \in B^s_{p,1}$ for any $s < 1/p$, relation (3.3) is satisfied for all $s < 3/(3\rho - 2)$. Owing to $s \leqslant 1$, our wavelet thresholding estimator achieves (almost) the rate $T^{-1/5}$ for $\rho < 5/3$ and the rate $T^{-1/(3\rho)}$ for $\rho \geqslant 5/3$, which is a significant gain. If our results could be generalized to cover the case of the quasi-Banach spaces L^p for $p < 1$ (which is to be expected), then $g \in B^s_{1/s,\infty}$ would yield (almost) the L^1-rate $T^{-1/3}$. The latter is in fact the minimax rate for nonparametric change point detection; cf. Reiss (2004).

In the mathematical results we have focused on the spatial adaptivity of our estimator, but the construction is clearly independent of major a priori knowledge of the unknown parameter. However, we had to assume some maximal domain of regularity ($s_{\max}$), some bound on the size (S) and some uniform mixing behaviour (δ). The resulting minimal asymptotically optimal threshold κ^* depends in a complicated way on these quantities. So clearly the question arises how to choose κ_λ for a specific observation.

First of all, note that $T\beta_{\lambda,T} - \langle Q_T a\psi_\lambda\rangle$ is a martingale with respect to T with quadratic variation $\langle Q_T\psi_\lambda, \psi_\lambda\rangle$. Asymptotically for $T \to \infty$, the random variable

$$\eta_{\lambda,T} := T^{1/2}\frac{\beta_{\lambda,T} - \langle T^{-1}Q_T a, \psi_\lambda\rangle}{\sigma_{\lambda,T}} \qquad (\text{with } \sigma^2_{\lambda,T} := \langle T^{-1}Q_T\psi_\lambda, \psi_\lambda\rangle \sim 2^{-2|\lambda|})$$

is therefore $N(0, 1)$-distributed by the martingale central limit theorem. In other words, we observe the coefficient $\langle T^{-1}Q_T a, \psi_\lambda\rangle$ under the noise $T^{-1/2}\sigma_{\lambda,T}\eta_{\lambda,T}$. Because $\sigma^2_{\lambda,T}$ converges stochastically, we conclude that the noise is approximately normally distributed with variance $T^{-1}\sigma^2_{\lambda,T}$, which is observable. Thus, we are led to apply the usual threshold rules in the Gaussian shift setting; see Donoho and Johnstone (1994) for a detailed discussion and Neumann and von Sachs (1995) for the case of only asymptotically Gaussian noise. It only remains to take account of the $W^{2,\rho}$-norm used so that we have to be a little bit more conservative and should choose in the Hilbertian case $\rho = 2$ the thresholds $\kappa_j = T^{-1/2}\sigma_{\lambda,T}\sqrt{6\log(T\sigma^{-2}_{\lambda,T}}$, provided the isomorphism constants in Theorem 2.3 are close to one when measured in the corresponding wavelet coefficient norms; cf. equation (17) of Abramovich and Silverman (1998). The maximal frequency J should be chosen such that $J^{-1}2^{-J} \approx T^{-1}\max_\lambda \sigma^2_{\lambda,T}$, which is an estimate of the squared noise level.

The cut-off in the definition of $\hat{a}_T$ avoids uncontrolled errors in the inversion and is quite natural; cf. Efromovich and Koltchinskii (2001). If only discrete observations (X_{t_i}) are available with $0 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_N = T$, then it can be shown that the error in approximating the stochastic integral b_T does not increase the asymptotics as long as $\Delta := \max_i(t_{i+1} - t_i)$ satisfies $\Delta \lesssim T^{-1/2}$. For low-frequency observations, that is, $\Delta > 0$ fixed and $N \to \infty$, it is an open question whether a consistent estimator exists at all.

One might wish to consider the submodel in which the weights do not include any point masses, that is, the weight space $B^s_{p,1}([-r, 0])$ instead of $\mathcal{B}^s_{p,1}$. For this purpose one can project the estimator $\hat{a}_T$ onto $L^\rho([-r, 0])$ by neglecting the point measure part. The asymptotic risk bound remains the same. Conversely, adding finitely many point masses with known location to our model does not increase the asymptotic rates for the risk. It must only be taken into account that the image functions of the weight measures under the covariance operator are less regular at these locations.

Finally, note that the approach can be extended naturally to multidimensional affine SDDEs where a matrix A of weight measures is to be estimated. In this case, we use the matrix-valued statistics b_T and q_T formed by applying the one-dimensional definition to all cross terms and we are led to the analogous inverse problem $Q_T\hat{A}_T \approx b_T$ to determine an estimator $\hat{A}_T$. A mathematical analysis of an adaptive version of $\hat{A}_T$ will be accomplished by similar methods, putting a wide range of applications within reach, the model being the counterpart of vector autoregressive processes in time series analysis.

4. Optimality of the estimator

We show that the adaptive wavelet thresholding estimator is rate-optimal with respect to L^ρ-risk functions, in the sense that one cannot improve on the restriction (3.3) in order to obtain the speed of convergence $(T/\log T)^{-s/(2s+3)}$ for weights in $\mathcal{B}^s_{p,1}$. For smaller values of p the rate of convergence is indeed worse and is obtained by embedding $\mathcal{B}^s_{p,1}$ in $\mathcal{B}^u_{\pi,1}$ with some properly chosen $u < s$ and $\pi > p$. In what follows, we merely assume $s + 1/\rho - 1/p \geqslant 0$ in order to have the embedding $\mathcal{B}^s_{p,1} \subset \mathcal{L}^\rho$ and thus a well-defined risk. For the sake of simplicity we do not present the proofs for the stationary case, but for fixed deterministic initial functions in (1.1). Due to ergodicity the initial segment is not significant for asymptotic statements, but the proofs for stochastic initial conditions are lengthy and tedious; see Reiss (2001).

Theorem 4.1. *Let $s > 0$, $p > 0$, $S > 0$ and $\delta > 0$ be given with $s + 1/\rho - 1/p \geqslant 0$ and such that $M(s, p, S, \delta)$ has non-empty interior in $\mathcal{B}^s_{p,1}$. Then the following asymptotic minimax lower bound holds for $T \to \infty$:*

$$\inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{\mathcal{L}^\rho}] \gtrsim \left(\frac{T}{\log T}\right)^{-(s+1/\rho-1/p)/(2s+3-2/p)},$$

where the infimum is taken over all $\mathcal{F}^X_T$-measurable estimators $\hat{a}_T$.

We obtain a fairly complete picture of the minimax rates for the L^ρ-risk and certain Besov regularity classes $M(s, p, S, \delta)$. Again replacing $s/3$ by the efficient regularity β, we obtain the usual picture (Härdle *et al.* 1998, Section 10.4).

Corollary 4.2. *Assume that $s > 0$, $p \in (1, \infty)$, $S > 0$ and $\delta > 0$ are given such that $M(s, p, S, \delta)$ has non-empty interior in $\mathcal{B}^s_{p,1}$. In what follows the infima are taken over all $\mathcal{F}^X_T$-measurable estimators $\hat{a}_T$.*

(i) *(Sparse case.) For $1/p - 1/\rho \geqslant 2s/(3\rho)$ the risk lower and upper bound match, that is, our estimator is rate-optimal in a minimax sense. We find*

$$\inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{\mathcal{L}^\rho} \sim \left(\frac{T}{\log T}\right)^{-u/(2u+3)},$$

with

$$u = \frac{3\rho}{3\rho - 2}\left(s + \frac{1}{\rho} - \frac{1}{p}\right) \leqslant s.$$

(ii) *(Regular case.) For $1/p - 1/\rho < 2s/(3\rho)$ we have*

$$T^{-s(2s+3)} \lesssim \inf_{\hat{a}_T} \sup_{a \in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{\mathcal{L}^\rho}] \lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}.$$

It should be noted that the actual minimax rates in the regular case are of course $T^{-s/(2s+3)}$, which are attained by linear estimators.

5. Proofs for the covariance operator

5.1. Proof of Theorem 2.3

We first need precise regularity and tail behaviour results for the covariance function q_a. Roughly speaking, the covariance function is three times more regular than the weight itself and decreases exponentially fast.

Proposition 5.1. *For $a \in \mathcal{B}^s_{p,\alpha}(v)$ with $v < 0$, $s > 0$, $p \in (1, \infty)$ and $\alpha \in [1, \infty]$, the covariance function q_a lies in $B^{s+3}_{p,\alpha}([0, r])$.*

Proof. Let us write $a = g + \alpha\delta_0 + \beta\delta_{-r}$ as in Definition 2.3. The covariance function satisfies (2.2) such that, for $t \in (0, r)$,

$$q_a'(t) = \int_{-r}^{0} q_a(|t+u|)\mathrm{d}a(u) = \int_{-r}^{0} q_a(|t+u|)g(u)\mathrm{d}u + \alpha q_a(t) + \beta q_a(r-t).$$

The properties $q_a(-t) = q_a(t)$, $q_a'(0+) = -\frac{1}{2}$ and $q_a \in C^2([0, \infty))$ from Küchler and Mensch (1992) imply for $t \in (0, r)$ that

$$q_a'''(t) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\int_{-r}^{0} q_a'(|t+u|)\mathrm{sgn}(t+u)g(u)\mathrm{d}u + \alpha q_a'(t) - \beta q_a'(r-t)\right)$$
$$= \int_{-r}^{0} q_a''(|t+u|)g(u)\mathrm{d}u + 2q_a'(0+)g(-t) + \alpha q_a''(t) + \beta q_a''(r-t)$$
$$= \int_{t-r}^{0} q_a''(v)g(v-t)\mathrm{d}v + \int_{0}^{t} q_a''(v)g(v-t)\mathrm{d}v - g(-t) + \alpha q_a''(t) + \beta q_a''(r-t).$$

This shows that the third derivative q_a''' is at least as regular in a Besov space sense as the most irregular term among the terms on the right-hand side. For the two convolution terms, note that the regularity result from Lemma A.2 with $f = g$, $k = q_a''$ and with obvious modifications of the interval boundaries applies: $g \in B^s_{p,\alpha}$ and $q_a \in B^w_{p,\alpha}$ for some $w \geqslant 1$, imply, by embedding, $q_a'' \in B^{\tilde{w}-2}_{p',\alpha'}$ for all $\tilde{w} < w - 1 + 2(p-1)/p$ and thus, by Lemma A.2, that the convolutions lie in $B^{\tilde{w}-2\}}_{p,\alpha}$. Formally, we now proceed by putting $w := \sup\{s \geqslant 2 | q_a \in B^s_{p,\alpha}\}$ and by deducing that the convolution terms are in $B^{\tilde{w}-2}_{p,\alpha}$ for some $\tilde{w} > w - 1$ while g is in $B^s_{p,\alpha}$ and q_a'' in $B^{w-2-\varepsilon}_{p,\alpha}$ for all $\varepsilon > 0$. Consequently, the right-hand side is an element of $B^{(\tilde{w}-2)\wedge s}_{p,\alpha}$, hence q_a is in $B^{(\tilde{w}+1)\wedge(s+3)}_{p,\alpha}$. By definition of w we know that q_a is not an element of $B^{\tilde{w}+1}_{p,\alpha}$ and therefore $q_a \in B^s_{p,\alpha}$ follows. For the $\mathcal{L}^p$-scale the proof is the same except that Lemma A.3 can be applied immediately to the expression for q_a'.

□

The following lemma extends well-known non-uniform results; see, for example, Hale and Verduyn Lunel (1993), Küchler and Sørensen (1997).

Lemma 5.2. *For all $S > 0$ and $v > \delta > 0$, we have the uniform bound*

$$\sup_{\|a\|_{\mathrm{TV}} \leqslant S, v_0(a) \leqslant -v} \|E_\delta q_a\|_{C^1([0,\infty))} < \infty.$$

Proof. We consider the formula $\hat{q}_a(\xi) = |\chi_a(\mathrm{i}\xi)|^{-2}$ from (2.3). Due to $|\chi_a(\mathrm{i}\xi)|^2 = \chi_a(\mathrm{i}\xi)\chi_a(-\mathrm{i}\xi)$, the Fourier transform $\hat{q}$ can be extended to a holomorphic function on the strip $\{z \in \mathbb{C} | |\mathrm{Im}(z)| < v_0(a)\}$ (Katznelson 1976, Section VI.7.1) and satisfies

$$\widehat{E_\delta(q_a)}(\xi) = \hat{q}_a(\xi + \mathrm{i}\delta) = \chi_a(\mathrm{i}\xi - \delta)^{-1}\chi_a(-\mathrm{i}\xi + \delta)^{-1}.$$

The assumptions guarantee $|\chi_a(\pm\mathrm{i}\xi \mp p\delta)| \geqslant |\mathrm{i}\xi + \delta| - S\mathrm{e}^{\delta r}$. Since subsets $U \subset M([-r, 0])$ that are bounded and closed in total variation norm are compact in the weak*-topology of $M([-r, 0])$ by the Banach–Alaoglu theorem, the set of characteristic functions $\{\chi_a | a \in U\}$ is compact in the space of entire functions equipped with the convergence on compact sets. Consequently, the classical result from calculus about the convergence of maxima on compact sets yields, for the respective choice of signs,

$$K_\pm := \sup_a \max_{|\xi| \leqslant 2S\mathrm{e}^{\delta r}} |\chi_a(\pm\mathrm{i}\xi \mp p\delta)|^{-1} < \infty,$$

where the supremum is taken over all measures a as in the statement of the lemma. We compare with the covariance function $q_{\mathrm{OU}}(t) = (2v)^{-1}\mathrm{e}^{-v|t|}$ of the Ornstein–Uhlenbeck process with parameter $-v$ and conclude that

$$\begin{aligned}
&\sup_a \|E_\delta(q_a - q_{\mathrm{OU}})\|_{C^1(\mathbb{R})} \\
&\quad\leqslant \sup_a \int_{-\infty}^{\infty} (1+|\xi|)\left|\frac{\chi_a(\mathrm{i}\xi-\delta)\chi_a(-\mathrm{i}\xi+\delta) - \chi_{\mathrm{OU}}(\mathrm{i}\xi-\delta)\chi_{\mathrm{OU}}(-\mathrm{i}\xi+\delta)}{\chi_a(\mathrm{i}\xi-\delta)\chi_a(-\mathrm{i}\xi+\delta)\chi_{\mathrm{OU}}(\mathrm{i}\xi-\delta)\chi_{\mathrm{OU}}(-\mathrm{i}\xi+\delta)}\right|, \mathrm{d}\xi \\
&\quad\leqslant \sup_a \int_{-\infty}^{\infty} (1+|\xi|)\frac{|\chi_a(-\mathrm{i}\xi+\delta) - \chi_{\mathrm{OU}}(-\mathrm{i}\xi+\delta)|}{|\chi_a(-\mathrm{i}\xi+\delta)\chi_{\mathrm{OU}}(\mathrm{i}\xi-\delta)\chi_{\mathrm{OU}}(-\mathrm{i}\xi+\delta)|}\mathrm{d}\xi \\
&\quad + \sup_a \int_{-\infty}^{\infty} (1+|\xi|)\frac{|\chi_a(\mathrm{i}\xi-\delta) - \chi_{\mathrm{OU}}(\mathrm{i}\xi-\delta)|}{|\chi_a(\mathrm{i}\xi-\delta)\chi_a(-\mathrm{i}\xi+\delta)\chi_{\mathrm{OU}}(\mathrm{i}\xi-\delta)|}\mathrm{d}\xi \\
&\quad\leqslant (\mathrm{e}^{\delta r}S + v)\left((S + \mathrm{S}^2\mathrm{e}^{\delta r})(K_+^3 + K_-^3) + \int_{|\xi|>2S\mathrm{e}^{\delta r}} (1+|\xi|)(|\mathrm{i}\xi+\delta| - S\mathrm{e}^{\delta r})^{-3}\,\mathrm{d}\xi\right) \\
&\quad< \infty.
\end{aligned}$$

Because $\|E_\delta(\mathrm{q}_{\mathrm{OU}})\|_{C^1([0,\infty))} < \infty$ the result follows. □

Proposition 5.3. *Let* E_δ *be the multiplication operator with the exponential:* $E_\delta(f)(t) := f(t)\mathrm{e}^{\delta t}$. *Then for* $S, v > 0$ *and* $p \geqslant 1$, *the covariance function* q_a *has for any* $\delta \in (0, v)$ *the property*

$$\sup_{a \in \mathcal{L}^p(-v): \|a\|_p \leqslant S} \|E_\delta(q_a)\|_{C^{2,1}([0,\infty))} < \infty.$$

In particular, the second derivative of $E_\delta(q_a)$ *is uniformly bounded on* $[0, \infty)$ *and Lipschitz continuous.*

Proof. The solution property (2.2) of the covariance function q_a implies, for $t \geqslant r$,

$$|q_a''(t)| \leqslant \|q_a'\|_{C([t-r,t])}\|a\|_{\mathrm{TV}} \leqslant \|q_a\|_{C([t-2r,t])}\|a\|_{\mathrm{TV}}^2.$$

By Lemma 5.2 and the product rule of differentiation we conclude that $\sup_{t \geqslant r}|E_\delta(q_a)''(t)|$ is uniformly bounded over the weights a. Similarly, the bound $|q_a''(t+v) - q_a''(t)| \leqslant \|q_a(\cdot+v) - q_a\|_{C([t-2r,t])}\|a\|_{\mathrm{TV}}^2$ for $t \geqslant r$ and $v \geqslant 0$ yields the existence of a uniform Lipschitz constant for q_a'' on $[r, \infty)$ by Lemma 5.2. It remains to apply Proposition 5.1, the uniformity result of Proposition 5.4 and Sobolev embeddings to obtain the same results on the interval $[0, r]$. □

We are now in a position to prove Theorem 2.3. First, let us see how Q_a acts on functions $f \in B^s_{p,\alpha}([-r, 0])$. Since Q_a maps $M([-r, 0])$ continuously to $C([-r, 0])$ by general properties of covariance operators (Vakhaniya *et al.* 1987, Theorem III.2.2), we only need to estimate $\|(Q_a f)''\|_{s,p,\alpha}$. By symmetry of q_a and by the regularity result

$q_a \in B^{s+3}_{p,\alpha}([0, r])$ (Proposition 5.1) we obtain for $t \in [-r, 0]$, as in the proof of Lemma 5.1, $(Q_a f)''(t) = \int_{-r}^{0} q_a''(t-s) f(s) \mathrm{d}s - f(t)$. From Lemma A.2 we infer further that

$$\|(Q_a f)''\|_{s,p,\alpha} \lesssim \|f\|_{s,p,\alpha} + \|f\|_{s-1,p,\alpha} \|q_a''\|_{s,p',\alpha'} \lesssim (1 + \|q_a\|_{s+3,p,\alpha}) \|f\|_{s,p,\alpha},$$

which shows that Q_a maps $B^s_{p,\alpha}$ continuously to $B^{s+2}_{p,\alpha}$. Writing the derivative operator as D, we find, for any $\varepsilon \in (0, 2 - (\frac{2}{p} \vee 1))$, by Lemma A.2 with $k = q_a''$ and by the embedding $B^{s+1}_{p,\alpha} \subset B^{s+\varepsilon}_{p',\alpha'}$, that

$$\|(D^2 Q_a + \mathrm{Id}) f\|_{s+\varepsilon} \lesssim \|f\|_{s+\varepsilon-1,p,\alpha} \|q_a'\|_{s+\varepsilon,p',\alpha'} \lesssim \|q_a\|_{s+3,p,\alpha} \|f\|_{s,p,\alpha}.$$

Hence, $D^2 Q_a + \mathrm{Id}$ is a compact operator on $B^s_{p,\alpha}$ ($B^{s+\varepsilon}_{p,\alpha}([-r,0]) \subset B^s_{p,\alpha}([-r,0])$ compactly).

Let $V \subset B^s_{p,\alpha}$ denote the kernel of $D^2 Q_a$ and let V^c be a complementary subspace of V. By Fredholm theory (Rudin 1991) the range of $D^2 Q_a$ is closed and its codimension equals the finite dimension of V. Therefore there exists a complementary subspace U of $D^2 Q_a(B^s_{p,\alpha})$ with $\dim U = \dim V$. The situation is illustrated by the following diagram:

$$\begin{array}{ccccc}
B^s_{p,\alpha} & = & V^c & \oplus & V \\
 & & & \Big\downarrow Q_a & \\
B^{s+2}_{p,\alpha} & = & Q_a(B^s_{p,\alpha}) & + & (D^2)^{-1}(U) \\
 & & & \Big\downarrow D^2 & \\
B^s_{p,\alpha} & = & D^2 Q_a(B^s_{p,\alpha}) & \oplus & U
\end{array}$$

While the decomposition in the first and in the third line hold by definition, the representation of $B^{s+2}_{p,\alpha}$ in the second line follows from the third line due to $(D^2)^{-1}(B^s_{p,\alpha}) = (D^2)^{-1}(D^2 Q_a(B^s_{p,\alpha}) \oplus U) \subset Q_a(B^s_{p,\alpha}) + (D^2)^{-1}(U)$. The fact that $Q_a(\mathrm{V})$ is contained in the kernel of D^2 implies that the operators Q_a and D^2 each map the vertically corresponding subspaces into each other.

This argument shows that $Q_a(B^s_{p,\alpha})$ is a closed subspace of $B^{s+2}_{p,\alpha}$ of codimension not larger than 2. Due to $Q_a \delta_{-r} = q_a(\cdot + r)$ and $Q_a \delta_0 = q_a$ we have $Q_a(\mathrm{span}(\delta_{-r}, \delta_0)) \subset B^{s+3}_{p,\alpha} \subset B^{s+2}_{p,\alpha}$ by Proposition 5.1. The injectivity of Q_a on $M([-r,0])$ implies that $Q_a(\mathrm{span}(\delta_{-r}, \delta_0))$ is a two-dimensional subspace of $B^{s+2}_{p,\alpha}$ in the complement of $Q_a(B^s_{p,\alpha})$. Owing to $\operatorname{codim} Q_a(B^s_{p,\alpha}) \leqslant 2$ this codimension must equal two and $Q_a : \mathcal{B}^s_{p,\alpha} \to B^{s+2}_{p,\alpha}$ is onto, hence bijective. Because Q_a is separately continuous on these two subspaces, it is continuous on its span $\mathcal{B}^s_{p,\alpha}$ and by the open mapping theorem it is an isomorphism.

Exactly the same reasoning applies for L^p and $\mathcal{L}^p$ instead of $B^s_{p,\alpha}$ and $\mathcal{B}^s_{p,\alpha}$; just apply Lemma A.3 to $(Q_a f)'$.

5.2. Proof of Theorem 2.4

This will be a consequence of the following continuity properties.

Proposition 5.4. *Suppose $s > 0$, $1 < p < \infty$, $\alpha \in [1, \infty]$ and $v < 0$ are given. If (a_n) is a sequence in $\mathcal{B}^s_{p,\alpha}(v)$ that converges in $\mathcal{B}^s_{p,\alpha}$-norm to the $\mathcal{B}^s_{p,\alpha}(v)$-weight a, then $\|q_{a_n} - q_a\|_{B^{s+3}_{p,\alpha}([0,r])} \to 0$ follows.*

Proof. Put $f_n := q_{a_n} - q_a$ and $a_n = g_n + \gamma_{r,n}\delta_{-r} + \gamma_{0,n}\delta_0$. As before, the following identities hold for $t \in (0, r)$:

$$
\begin{aligned}
f_n''(t) &= \left(\int_{-r}^0 q_{a_n}(\cdot + u)\mathrm{d}a_n(u) - \int_{-r}^0 q_a(\cdot + u)\mathrm{d}a(u)\right)'(t)\\
&= \left(\int_{-r}^0 f_n(\cdot + u)\mathrm{d}a_n(u)\right)'(t) + (Q_a(a_n - a)(-\cdot))'(t)\\
&= -\int_{-r}^{-t} f_n'(-t-u)g_n(u)\mathrm{d}u + \int_{-t}^0 f_n'(t+u)g_n(u)\mathrm{d}u\\
&\quad - \gamma_{r,n}f_n'(r-t) + \gamma_{0,n}f_n'(t) - (Q_a(a_n - a))'(-t)\\
&= -\int_0^{r-t} f_n'(u)g_n(-u-t)\mathrm{d}u + \int_{-t}^0 f_n'(u)g_n(u-t)\mathrm{d}u\\
&\quad - \gamma_{r,n}f_n'(r-t) + \gamma_{0,n}f_n'(t) - (Q_a(a_n - a))'(-t).
\end{aligned}
$$

By Lemma A.2 we obtain for all $w > 0$ (allowing the value ∞) the estimate

$$
\begin{aligned}
\|f_n''\|_{w,p,\alpha} \lesssim{}& \|f_n'\|_{w,p',\alpha'}\|g_n\|_{w-1,p,\alpha} + (|\gamma_{1,n}| + |\gamma_{2,n}|)\|f_n'\|_{w,p,\alpha}\\
&+ \|Q_a\|_{B^{(w-1)\vee 0}_{p,\alpha} \to B^{w+1}_{p,\alpha}}\|a_n - a\|_{(w-1)\vee 0,p,\alpha}.
\end{aligned}
\tag{5.1}
$$

For $a_n \to a$ weakly the covariance functions converge in $W^{\rho,2}([0, r])$ for all $\rho < \frac{5}{2}$, which was established in Reiss (2002) by spectral methods. Hence, $\|f_n\|_{w,p,\alpha} \to 0$ holds for all $w < 2 + 1/p$. In particular, the convergence $\|f_n\|_{L^p} \to 0$ follows. The right-hand side of estimate (5.1) is thus finite for all $w \in (0, 1/p)$. Once again using $B^w_{p,\alpha} \subset B^{w-1+\varepsilon}_{p',\alpha'}$ for any $\varepsilon \in (0, 2 - (2/p \vee 1))$, we obtain, for all $w \leqslant s + 1$,

$$
\|f_n\|_{w+2,p,\alpha} \lesssim \|f_n\|_{L^p} + \|a_n\|_{s,p,\alpha}\|f_n\|_{w+2-\varepsilon,p,\alpha} + \|Q_a\|\,\|a_n - a\|_{s,p,\alpha}.
$$

Starting with $w_0 = \varepsilon$, we can iterate this estimate ($w_{n+1} := (w_n + \varepsilon) \wedge (s + 1)$). Hence $\|f_n\|_{s+3,p,\alpha}$ is bounded by a multiple of $\|f_n\|_{L^p} + \|f_n\|_{2,p,\alpha} + \|a_n - a\|_{s,p,\alpha}$, which tends to zero for $n \to \infty$. This proves that $\|f_n\|_{s+3,p,\alpha} \to 0$. □

Proposition 5.5. *If (a_n) is a sequence of $\mathcal{B}^s_{p,\alpha}(v)$-weights that converges in $\mathcal{B}^w_{p,\alpha}$-norm to the*

$\mathcal{B}^s_{p,\alpha}$*-weight* a *for some* $w > s-2+(1\vee 2/p)$ *and* $s>0$, $p\in(1,\infty)$, $\alpha\in[1,\infty]$, $v<0$, *then the covariance operators converge in operator norm:*

$$\lim_{n\to\infty}\|Q_{a_n}-Q_a\|_{\mathcal{B}^s_{p,\alpha}\to B^{s+2}_{p,\alpha}}=0.$$

This statement remains true if $\mathcal{B}^s_{p,\alpha}$ *is replaced by* $\mathcal{L}^p$ *and* $B^{s+2}_{p,\alpha}$ *by* $W^{2,p}$ *and under the condition that* (a_n) *converges in* $\mathcal{L}^p$*-norm to* a.

Proof. By linearity we have, for $f\in B^s_{p,\alpha}$ and $t\in[-r,0]$,

$$((Q_{a_n}-Q_a)f)''(t)=\int_0^{r+t}f(t-u)(q_{a_n}-q_a)''(u)\mathrm{d}u+\int_0^{-t}f(t+u)(q_{a_n}-q_a)''(u)\mathrm{d}u.$$

By Lemma A.2 and by the norm estimates $\|\cdot\|_{s+2,p',\alpha'}\lesssim\|\cdot\|_{w+3,p,\alpha}$ and $\|\cdot\|_{L^{p'}}\lesssim\|\cdot\|_{w+3,p,\alpha}$, we infer the bound

$$\begin{aligned}\|(Q_{a_n}-Q_a)f\|_{s+2,p,\alpha}&\lesssim\|(Q_{a_n}-Q_a)f\|_{L^\infty}+\|f\|_{s-1,p,\alpha}\|(q_{a_n}-q_a)''\|_{s,p',\alpha'}\\&\lesssim\big(\|q_{a_n}-q_a\|_{L^{p'}}+\|q_{a_n}-q_a\|_{w+3,p,\alpha}\big)\|f\|_{s,p,\alpha}\\&\lesssim\|q_{a_n}-q_a\|_{w+3,p,\alpha}\|f\|_{s,p,\alpha}.\end{aligned}$$

The proof in the L^p-case is proceeds in a completely analogous way. □

The following proof of Theorem 2.4 works for the Besov scale as well as for the L^p-scale. Since the set $M(s,p,S,\delta)$ is bounded in $\mathcal{B}^s_{p,\alpha}$, it is relatively compact in any $\mathcal{B}^w_{p,\alpha}$ for $w<s$. By the preceding result the operator norm of Q_a depends continuously on a in $\mathcal{B}^w_{p,\alpha}$-norm for some $w<s$ such that the supremum of $\|Q_a\|$ over $M(s,p,S,\delta)$ is attained and finite. The norm continuity of the mapping $Q_a\mapsto Q_a^{-1}$ (Rudin 1991, Thm. 10.11) yields the bound for the inverse, and the proof of Theorem 2.4 is complete.

6. Proof of the upper bound

6.1. Proof of Theorem 3.1.

This result follows from an exponential moment bound on $T^{-1}Q_T$, which for small exponents α yields a Gaussian moment behaviour.

Theorem 6.1. *Let* δ, R >0 *be given. Then there are constants* K, $T_0>0$ *such that for all weight measures* a *with* $v_0(a)\leqslant-\delta$, $\|a\|_{\mathrm{TV}}\leqslant R$, *all multi-indices* λ, *all measures* $\mu\in M([-r,0])$, *all* $T\geqslant T_0$ *and all* $\alpha\in[0,T^{1/2}(K\|\mu\|_{\mathrm{TV}})^{-1})$, *the following moment bound holds:*

$$\mathbb{E}_a\Big[\cosh\Big(\alpha T^{1/2}2^{3|\lambda|/2}\langle(T^{-1}Q_T-Q_a)\mu,\psi_\lambda\rangle\Big)\Big]\leqslant\exp\Big(K\|\mu\|^2_{\mathrm{TV}}\alpha^2\Big).$$

In particular, using $x^{2m}\lesssim\cosh(x)$, we obtain

$$\mathbb{E}_a[\langle (T^{-1}Q_T - Q_a)\mu,\, \psi_\lambda\rangle^{2m}]^{1/2m} \lesssim T^{-1/2}2^{-3|\lambda|/2}\|\mu\|_{\mathrm{TV}}. \tag{6.1}$$

Proof. Due to $\cosh(x) = \sum_m x^{2m/(2m)!}$, we shall estimate polynomial moments. Using the finiteness of $\mathbb{E}[\|X\|^{2m}_{C([-r,T])}]$ by the Fernique theorem on $C([-r, T])$ as a requirement for the Fubini theorem we obtain:

$$\begin{aligned}
&\mathbb{E}_a[\langle (T^{-1}Q_T - Q_a)\mu,\, \psi_\lambda\}^{2m}] \\
&\quad= \mathbb{E}_a[\langle \mu,\, (T^{-1}Q_T - Q_a)\psi_\lambda\}^{2m}] \\
&\quad= \int_{[-r,0]^{2m}} \mathbb{E}_a\left[\prod_{i=1}^{2m}(T^{-1}Q_T - Q_a)\psi(u_i)\right] \mathrm{d}\mu(u_{2m})\ldots \mathrm{d}\mu(u_1) \\
&\quad\lesssim \|\mu\|^{2m}_{\mathrm{TV}} \sup_{u_1,\ldots,u_{2m}}\left|\int_{[-r,0]^{2m}} \mathrm{d}v_{2m}\ldots \mathrm{d}v_1 \mathbb{E}_a\left[\prod_{i=1}^{2m}\big(T^{-1}q_T(u_i,\, v_i) - q_a(u_i,\, v_i)\big)\right]\prod_{i=1}^{2m}\psi_\lambda(v_i)\right| \\
&\quad= \|\mu\|^{2m}_{\mathrm{TV}} T^{-2m} \sup_{u_1,\ldots,u_{2m}}\left|\int_{[-r,0]^{2m}} \mathrm{d}v_{2m}\ldots \mathrm{d}v_1 \int_{[0,T]^{2m}} \mathrm{d}t_{2m}\ldots \mathrm{d}t_1\right. \\
&\qquad\left.\mathbb{E}_a\left[\prod_{i=1}^{2m}(X(t_i+u_i)X(t_i+v_i) - q_a(u_i - v_i))\right]\prod_{i=1}^{2m}\psi_\lambda(v_i)\right|.
\end{aligned}$$

In order to evaluate the expected value of the product, let us introduce the set $P_2(2n)$ of all partitions of the set $\{1, \ldots, 2n\}$ into subsets with two elements. An easy argument based on the characteristic function shows that for a centred Gaussian random vector $(N_1, \ldots, N_{2n})$ the formula

$$\mathbb{E}\left[\prod_{i=1}^{2n} N_i\right] = \sum_{\Gamma\in P_2(2n)} \prod_{(k,l)\in\Gamma} \mathbb{E}[N_k N_l]$$

is valid. Let us set $n = 2m$, $A_i = N_{2i-1}$, $B_i = N_{2i} =$ and $\alpha = \mathbb{E}[A_i B_i]$. Then we obtain the following formula because terms involving neighbouring random variables N_{2i-1}, N_{2i} cancel (proof by induction over n):

$$\mathbb{E}_a\left[\prod_{i=1}^{2m}(A_i B_i - \alpha)\right] = \sum_{\substack{\Gamma\in P_2(4m)\\ \forall i:\{2i-1,2i\}\notin\Gamma}} \prod_{(k,l)\in\Gamma} \mathbb{E}_a[N_k N_l]. \tag{6.2}$$

In our case the expected value of the product equals

$$\sum_{\substack{\Gamma\in P_2(4m)\\ \forall i:\{2i-1,2i\}\notin\Gamma}} \prod_{(k,l)\in\Gamma} q_a(z_k - z_l),$$

with $z_{2i-1} = t_i + u_i$ and $z_{2i} = t_i + v_i$. Changing the order of integration, we start with the integration over v_i, $i = 1, \ldots, 2m$. Since any v_i appears only once in the product, we have to deal with products over terms which have one of the following three forms:

(i) $q_a(t_i + u_i - t_j - u_j)$,
(ii) $\int_{-r}^{0} q_a(t_i + u_i - t_j - v_j)\psi_\lambda(v_j)\mathrm{d}v_j$,
(iii) $\int_{-r}^{0}\int_{-r}^{0} q_a(t_i + v_i - t_j - v_j)\psi_\lambda(v_i)\psi_\lambda(v_j)\mathrm{d}v_i\,\mathrm{d}v_j$.

For factor (i) we shall use $|q_a(t_i + u_i - t_j - u_j) \leqslant C_1 \mathrm{e}^{-\delta|t_i - t_j|}$ derived from Proposition 5.1 for $\delta < -v_0(a)$.

The Lipschitz constant of $q_a(t_i + u_i - t_j - \cdot)$ on $[-r, 0]$ is of order $\mathrm{e}^{-\delta(|t_i - t_j| - r)}$ by Proposition 5.1, which implies the existence of a constant C_2 such that the modulus of integral (ii) is smaller than $C_2 2^{-3|\lambda|/2}\mathrm{e}^{-\delta|t_i - t_j|}$ (Lemma A.6).

For the estimation of integral (iii) we let S denote the length of the minimal interval supporting ψ and distinguish the cases (1) $|t_i - t_j| > 2^{-|\lambda|}S$ and (2) $|t_i - t_j| \geqslant 2^{-|\lambda|}S$. A substitution gives

$$\int_{-r}^{0}\int_{-r}^{0} q_a(t_i + v_i - t_j - v_j)\psi_\lambda(v_i)\psi_\lambda(v_j)\mathrm{d}v_i\,\mathrm{d}v_j$$

$$= \iint_{|v_i - v_j| \leqslant S} q_a(t_i - t_j + 2^{-|\lambda|}(v_i - v_j))2^{-|\lambda|}\psi(v_i)\psi(v_j)\,\mathrm{d}v_i\,\mathrm{d}v_j,$$

which shows that in case (1) q_a needs only to be evaluated at either positive arguments or at negative ones. Due to the Lipschitz continuity of q_a' with exponentially decaying norm (Proposition 5.1) the estimate in Lemma A.6 shows that the modulus of (iii) is in case (1) smaller than $C_3 2^{-3|\lambda|}\mathrm{e}^{-\delta|t_i - t_j|}$, $C_3 > 0$ a constant. In case (2) q_a is at least Lipschitz continuous and the modulus of (iii) is by the same arguments smaller than $C_4 2^{-2|\lambda|}\mathrm{e}^{-\delta|t_i - t_j|}$, $C_4 > 0$ a constant.

Finally, note that each u_i and each v_i appears exactly once in the product and that each t_i appears twice so that, with $C := \max_j C_j$,

$$\left| \int_{[-r,0]^{2m}} \mathrm{d}v_{2m} \ldots \mathrm{d}v_1 \mathbb{E}_a\left[\prod_{i=1}^{2m}(X(t_i + u_i)X(t_i + v_i) - q_a(u_i - v_i))\right]\prod_{i=1}^{2m}\psi_\lambda(v_i)\right|$$

$$\leqslant \sum_{\Gamma} 2^{-3|\lambda|m}C^{2m}\prod_{(k,l)\in\Gamma}\left(1 + 2^{|\lambda|}\mathbf{1}_{k,l\ \mathrm{even}|t_{k/2} - t_{l/2}| \leqslant S2^{-|\lambda|}\}}\right)\mathrm{e}^{-\delta|t_{\lceil k/2\rceil} - t_{\lceil l/2\rceil}|}.$$

The partitions Γ can also be described by fixed point-free permutations. Let us denote $2k - 1$ and $2k$ by the same symbol $s(k)$. The idea is to start with one pair $\{k_0, k_1\} \in \Gamma$, to look for $\{k_1', k_2\} \in \Gamma$ with $s(k_1') = s(k_1)$, then for $\{k_2', k_3\}$ with $s(k_2') = s(k_2)$ and so on until $s(k_l)$ equals $s(k_0)$. This describes a cyclic permutation of $\{s(k_0), \ldots, s(k_{l-1})\}$. Proceeding in the same manner for the remaining elements of Γ and identifying $s(k)$ with $\lceil k/2\rceil$, a fixed point-free permutation $\pi = \pi(\Gamma)$ of $\{1, \ldots, 2m\}$ is defined. To clarify the construction look at the following example ($m = 6$):

$$\Gamma = \{\{1, 3\}, \{2, 11\}, \{4, 7\}, \{5, 10\}, \{6, 9\}, \{8, 12\}\}$$
$$\Rightarrow s(1) \mapsto s(2) \mapsto s(4) \mapsto s(6);\ s(3) \mapsto s(5) \Rightarrow \pi(G) = (1\,2\,4\,6)\ (3\,5).$$

Let us denote by $C(\pi)$ the set of cycles in π and by $|\tau|$ the length of a cycle τ. Then we can easily evaluate the integral over the product:

$$\begin{aligned}
&\int_{[0,T]^{2m}} \prod_{(k,l)\in\Gamma} (1 + 2^{|\lambda|}\mathbf{1}_{\{k,l \text{ even},|t_{k/2}-t_{l/2}|\leqslant S2^{-|\lambda|}\}})\mathrm{e}^{-\delta|t_{\lceil k/2\rceil}-t_{\lceil l/2\rceil}|}\,\mathrm{d}t_1\ldots\mathrm{d}t_{2m}\\
&= \prod_{\tau\in C(\pi(\Gamma))} \int_{[0,T]^{|\tau|}} \prod_{k=1}^{|\tau|}(1 + 2^{|\lambda|}\mathbf{1}_{|s_{k+1}-s_k|\leqslant S2^{-|\lambda|}\}})\mathrm{e}^{-\delta|s_{k+1}-s_k|}\,\mathrm{d}s_1\ldots\mathrm{d}s_{|\tau|}\\
&\leqslant \prod_{\tau\in C(\pi(\Gamma))} \int_0^T \mathrm{d}s_1 \int_{[T,T]^{|\tau|-1}} \prod_{k=1}^{|\tau|-1}(1 + 2^{|\lambda|}\mathbf{1}_{|u_k|\leqslant S2^{-|\lambda|}\}})\mathrm{e}^{-\delta|u_k|}\,\mathrm{d}u_1\ldots,\mathrm{d}u_{|\tau|-1}\\
&\leqslant \prod_{\tau\in C(\pi(\Gamma))} (T(2\delta^{-1}+2S)^{|\tau|-1})\\
&\leqslant T^{|C(\pi(\Gamma))|}(2\delta^{-1}+2S)^{2m}.
\end{aligned}$$

So far we have shown that

$$\mathbb{E}_a[\langle(2(T^{-1}Q_T - Q_a)\mu, \psi_\lambda\rangle^{2m}] \leqslant \|\mu\|_{\mathrm{TV}}^{2m} T^{-2m}(2C(\delta^{-1}+S))^{2m}2^{-3|\lambda|m}\sum_\Gamma T^{|C(\pi(\Gamma))|}.$$

It remains to solve the combinatorial problem to determine the number $a_{n,k}$ of fixed point-free permutations of $\{1,\ldots,n\}$ with exactly k cycles. We claim that the following recursive relation is true for all $n \geqslant 3$, $k \geqslant 1$:

$$a_{n,k} = (n-1)a_{n-1,k} + (n-1)a_{n-2,k-1}, \qquad a_{n,1} = a_{n,0} = 1, \quad a_{1,k} = 0.$$

We classify with respect to the element n. If in a permutation n is in a cycle of length at least 3, then by leaving n out, we obtain a fixed-point free permutation of $\{1,\ldots,n-1\}$ with k cycles. Since n can stand in front of every other element, there are exactly $n-1$ ways to generate such an n-permutation from a valid $(n-1)$-permutation. This explains the first term. The second stems from the permutations where n lies in a cycle of length 2. By removing this 2-cycle we obtain a fixed point-free permutation of $n-2$ elements with $k-1$ cycles. Since the other element of the cycle involving n can be chosen from all other elements, we find the second summand.

From this recursive relationship we infer by an easy induction argument that the generating function satisfies

$$\sum_{k=1}^{2m} a_{2m,k}x^k \leqslant \frac{(2m)!}{m!}(x+1)\ldots(x+m), \qquad x \geqslant 0.$$

We are now in a position to prove the assertion of the theorem:

$$\begin{aligned}
&\mathbb{E}_a[\cosh(\alpha T^{1/2}2^{3|\lambda|/2}\langle (T^{-1}Q_T - Q_a)\mu, \psi_\lambda\rangle)] \\
&\quad = \sum_{m=0}^{\infty} \frac{\mathbb{E}_a[(\alpha T^{1/2}2^{3|\lambda|/2}\langle (T^{-1}Q_T - Q_a)Q_a, \psi_\lambda\rangle)^{2m}]}{(2m)!} \\
&\quad \leqslant \sum_{m=0}^{\infty} \frac{\|\mu\|_{\mathrm{TV}}^{2m} T^{-m}(2C(\delta^{-1}+S))^{2m}\alpha^{2m}}{(2m)!} \sum_{k=1}^{2m} a_{2m,k}T^k \\
&\quad \leqslant \sum_{m=0}^{\infty} \frac{(\|\mu\|_{\mathrm{TV}}^2 T^{-1}(2C(\delta^{-1}+S))^2\alpha^2)^m}{m!}(T+1)\cdots(T+m) \\
&\quad \leqslant \sum_{m=0}^{\infty} (\|\mu\|_{\mathrm{TV}} T^{-1/2} 2C(\delta^{-1}+S)\alpha)^{2m}\binom{T+m}{m} \\
&\quad = (1 - (\|\mu\|_{\mathrm{TV}} 2C(\delta^{-1}+S)\alpha)^2 T^{-1})^{-(T+1)} \\
&\quad \leqslant \exp(K\|\mu\|_{\mathrm{TV}}^2\alpha^2),
\end{aligned}$$

where $K := 2(2C(\delta^{-1}+S))^2$ and $T \geqslant T_0$ large enough. The estimates of the covariance function rely only on Proposition 5.1, whence, by Proposition 5.4, the uniformity of the constant follows. □

We can now prove Theorem 3.1. The moment inequality in Theorem 6.1 yields

$$\mathbb{P}_a\Big(T^{1/2}2^{3|\lambda|/2}|\langle (T^{-1}Q_T - Q_a)\mu, \psi_\lambda\rangle| \geqslant \frac{\kappa}{2}\sqrt{\log T}\Big) \leqslant \frac{\exp(K\|\mu\|_{\mathrm{TV}}^2\alpha^2}{\cosh(\frac{1}{2}\alpha\kappa\sqrt{\log T})}$$

for any suitable α. The choice $\alpha = \frac{1}{2}K\sqrt{\log T}/(2K\|\mu\|_{\mathrm{TV}}^2)$ yields the bound $2T^{-\kappa^2/(16K\|\mu\|_{\mathrm{TV}}^2)}$. From the decomposition (3.1) it follows that

$$|\beta_{\lambda,T} - \langle Q_a a, \psi_\lambda\rangle| \leqslant |\langle (T^{-1}Q_T - Q_a)a, \psi_\lambda\rangle| + \left|T^{-1}\int_0^T \langle X(t+\cdot), \psi_\lambda\rangle \mathrm{d}W(t)\right|.$$

The stochastic integral has quadratic variation $\langle Q_T\psi_\lambda, \psi_\lambda\rangle$ and by the exact deviation probability bound found by Lipster and Spokoiny (2000) we infer, for any $\kappa > 0$ and large T, that

$$\mathbb{P}_a\left(\frac{2^{|\lambda|}}{T}\left|\int_0^T \langle X(t+\cdot), \psi_\lambda\rangle \mathrm{d}W(t)\right| > \frac{\kappa}{2}\sqrt{2^{2|\lambda|}\langle Q_a\psi_\lambda, \psi_\lambda\rangle T\log T}\right)$$

$$\leqslant 4\sqrt{\mathrm{e}}\kappa(\log T)T^{-\kappa^2/(8+2\kappa(\log T)^{1/2}T^{-1/2})}$$

$$+\mathbb{P}_a\left(2^{2|\lambda|}|\langle(T^{-1}Q_T - Q_a)\psi_\lambda, \psi_\lambda\rangle| > \frac{\kappa}{2}\sqrt{T^{-1}\log T}\right)$$

$$\lesssim T^{-\kappa^2/9} + T^{-\kappa^2/(16K\|\psi\|^2_{L^1})}.$$

By Lemma A.5 the expression $2^{2|\lambda|}\langle Q_a\psi_\lambda, \psi_\lambda\rangle$ is uniformly bounded from below by some $m > 0$ and we obtain the uniform estimate

$$\mathbb{P}_a\left(2^{|\lambda|}T^{1/2}|\beta_{\lambda,T} - \langle Q_a a, \psi_\lambda\rangle| > \frac{\kappa}{2}\sqrt{\log T}\right)$$

$$\lesssim T^{-\kappa^2 2|\lambda|/(16K\|a\|^2_{\mathrm{TV}})} + T^{-\kappa^2/(9m^2)} + T^{-\kappa^2/(16K\|\psi\|^2_{L^1}m^2)}.$$

If we choose $\kappa^2 \geqslant \max(48KR^2, 48m^2\|\psi\|^2_{L^1}, 27m^2)\rho =: \kappa^*$, then the right-hand side is of maximal order $T^{-3\rho}$. □

6.2. Proof of Proposition 3.2

Without loss of generality we assume $p < \rho$ and we omit the T-dependence of the quantities. Let us introduce the true coefficients $(b_{j,k})$ and error coefficients $(e_{j,k})$

$$b_{j,k} := \langle Q_a a, \psi_{j,k}\rangle, \qquad e_{j,k} := \beta_{j,k} - b_{j,k} = \langle T^{-1}b_T - Q_a a, \psi_{j,k}\rangle.$$

We split the risk according to the usual bias–variance decomposition:

$$\mathbb{E}_a[\|\hat{b}_T - Q_a a\|_{W^{2,\rho}}] \leqslant \mathbb{E}_a[\|\hat{b}_T - P_J Q_a a\|_{W^{2,\rho}}] + \|(\mathrm{Id} - P_J)Q_a a\|_{W^{2,\rho}}.$$

The second (bias) term can be dealt with by linear approximation theory. The Besov space embeddings (A.1) yield under the restriction (3.3) that $B^{s+2}_{p,1} \subset W^{\frac{s}{2s+3}+2,\rho}$ together with Jackson's inequality (A.2) in $W^{2,\rho}([-r, 0])$,

$$\|(\mathrm{Id} - P_J)Q_a a\|_{W^{2,\rho}} \lesssim 2^{-Js/(2s+3)}\|Q_a a\|_{W^{s/(2s+3)+2,\rho}} \lesssim (T/\log T)^{-s/(2s+3)}\|Q_a a\|_{B^{s+2}_{p,1}}.$$

Due to $Q_a : \mathcal{B}^s_{p,1} \to B^{s+2}_{p,1}$ isomorphically (Theorem 2.3) with uniform constants (Theorem 2.4), this second term is of order $(T/\log T)^{-s/(2s+3)}$ uniformly over $M(s, p, S, \delta)$.

The first term can be estimated using the imbedding $B^2_{\rho,1}([-r, 0]) \subset W^{2,\rho}([-r, 0])$, the characterization of $B^2_{\rho,1}$ by 2-regular wavelets (Section A.2) and Jensen's inequality:

$$\begin{aligned}\mathbb{E}_a[\|\hat{b}_T - P_J Q_a a\|_{W^{2,\rho}}] &\lesssim \mathbb{E}_a\left[\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \left|\beta_{j,k}\mathbf{1}_{|\beta_{j,k}|>\kappa_j} - b_{j,k}\right|^\rho\right)^{1/\rho}\right] \\ &\leqslant \sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{E}_a\left[\left|\beta_{j,k}\mathbf{1}_{|\beta_{j,k}|>\kappa_j} - b_{j,k}\right|^\rho\right]\right)^{1/\rho}.\end{aligned}$$

The term $\beta_{j,k}\mathbf{1}_{|\beta_{j,k}|>\kappa_j} - b_{j,k}|^\rho$ can be split according to whether thresholding takes place or not and whether the true coefficient is large or not. It equals

$$\begin{aligned}&|\beta_{j,k} - b_{j,k}|^\rho \mathbf{1}_{|\beta_{j,k}|>\kappa_j} + |b_{j,k}|^\rho \mathbf{1}_{|\beta_{j,k}|\leqslant\kappa_j} \\ &\quad = |e_{j,k}|^\rho \mathbf{1}_{\substack{|\beta_{j,k}|>\kappa_j \\ |b_{j,k}|\leqslant\kappa_j/2}} + |e_{j,k}|^\rho \mathbf{1}_{\substack{|\beta_{j,k}|>\kappa_j \\ |b_{j,k}|>\kappa_j/2}} + |b_{j,k}|^\rho \mathbf{1}_{\substack{|\beta_{j,k}|\leqslant\kappa_j \\ |b_{j,k}|>2\kappa_j}} + |b_{j,k}|^\rho \mathbf{1}_{\substack{|\beta_{j,k}|\leqslant\kappa_j \\ |b_{j,k}|\leqslant 2\kappa_j}} \\ &\quad \leqslant |e_{j,k}|^\rho \mathbf{1}_{|e_{j,k}|>\kappa_n/2} + |e_{j,k}|^\rho \mathbf{1}_{|b_{j,k}|>\kappa_j/2} + |b_{j,k}|^\rho \mathbf{1}_{|e_{j,k}|>\kappa_j} + |b_{j,k}|^\rho \mathbf{1}_{|b_{j,k}|\leqslant 2\kappa_j} \\ &\quad =: S_1(j, k) + S_2(j, k) + S_3(j, k) + S_4(j, k).\end{aligned}$$

By the Cauchy–Schwarz inequality, the large deviation bound (3.2) on $e_{j,k}$ and (6.1), we obtain a fast decay for the sum involving $S_1(j, k)$:

$$\begin{aligned}&\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{E}_a[S_1(j, k)]\right)^{1/\rho} \\ &\quad \leqslant \sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{P}_a\left(|e_{j,k}| > \frac{\kappa_j}{2}\right)^{1/2} \mathbb{E}_a[e_{j,k}^{2\rho}]^{1/2}\right)^{1/\rho} \\ &\quad \lesssim \sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k T^{-3\rho/2} 2^{-j\rho} T^{-\rho/2}\right)^{1/\rho} \sim T^{-2} 2^{3J/2} \lesssim T^{-1/2}.\end{aligned}$$

The large-deviation estimate also bounds the sum over $S_3(j, k)$:

$$\begin{aligned}\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{E}_a[S_3(j, k)]\right)^{1/\rho} &= \sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{P}_a\left(|e_{j,k}| > \frac{\kappa_j}{2}\right)|b_{j,k}|^\rho\right)^{1/\rho} \\ &\lesssim \|Q_a a\|_{2,\rho,1} T^{-3} \lesssim T^{-3}\|a\|_{B^2_{p,1}}.\end{aligned}$$

The remaining estimates rely on nonlinear approximation theory. Using the characterization by $(s+2)$-regular wavelets (Section A.2),

$$\|Q_a a\|_{B^{s+2}_{p,p/\rho}} \sim \left(\sum_{j\geqslant 0} 2^{jp(s+5/2-1/p)/\rho}\left(\sum_k |b_{jk}|^p\right)^{1/\rho}\right)^{\rho/p},$$

we infer for all $j\in\mathbb{N}_0$ and $\tau_j>0$ by a Chebyshev inequality-type argument the following bound on the cardinality of large wavelet coefficients:

$$\sum_{j\geqslant 0} 2^{jp(s+5/2-1/p)/\rho}|\{k\,|\,|b_{jk}|\geqslant\tau_j\}|^{1/\rho}\tau_j^{p/\rho} \lesssim \|Q_a a\|^{p/\rho}_{B^{s+2}_{p,p/\rho}} \leqslant \|Q_a a\|^{p/\rho}_{B^{s+2}_{p,1}}. \tag{6.3}$$

The sum involving $S_2(j,k)$ can be bounded by separate estimates, where j_0 is such that $2^{j_0}\sim T^{1/(2s+3)}$:

$$\begin{aligned}
&\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{E}_a[S_2(j,k)]\right)^{1/\rho}\\
&\quad=\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\left(\sum_k \mathbb{E}_a[|e_{j,k}|^\rho]\mathbf{1}_{|b_{j,k}|>\kappa_j/2}\right)^{1/\rho}\\
&\quad\lesssim \sum_{j\leqslant j_0} 2^{j(5/2-1/\rho)}2^{j/\rho}T^{-1/2}2^{-j}+\sum_{j>j_0}2^{j(5/2-1/\rho)}\left(\sum_k T^{-\rho/2}2^{-j\rho}\mathbf{1}_{|b_{j,k}|>\kappa_j/2}\right)^{1/\rho}\\
&\quad\lesssim T^{-1/2}2^{3j_0/2}+T^{-1/2+p/2\rho}\sum_{j>j_0}2^{j(3/2-1/\rho+p/\rho)}\left(\frac{\kappa_j}{2}\right)^{p/\rho}|\{k\,|\,|b_{j,k}|>\kappa_j/2\}|^{1/\rho}\\
&\quad\lesssim T^{-1/2}2^{3j_0/2}+T^{-1/2+p/2\rho}2^{-j_0(ps/\rho+3p/2\rho-3/2)}\\
&\quad\sim T^{-1/2}2^{3j_0/2}(1+T^{p/2\rho}2^{-j_0p(s+3/2)/\rho})\sim T^{-s/(2s+3)}.
\end{aligned}$$

In the fifth line we have used the sparsity estimate (6.3) and the fact that $ps/\rho+3p/2\rho-3/2$ is non-negative due to assumption (3.3).

The slightly extended technique also applies to the estimate of the sum over $S_4(j,k)$. Here, one must choose $2^{j_0}\sim(T/\log T)^{1/(2s+3)}$ for balancing the two sums appearing in the following calculations:

$$\sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\Big(\sum_k \mathbb{E}_a[S_4(j,\,k)]\Big)^{1/\rho}$$

$$= \sum_{j\leqslant J} 2^{j(5/2-1/\rho)}\Big(\sum_k |b_{j,k}|^\rho \mathbf{1}_{|b_{j,k}|\leqslant 2\kappa_j}\Big)^{1/\rho}$$

$$\leqslant \sum_{j\leqslant j_0} 2^{j(5/2-1/\rho)} 2^{j/\rho} 2\kappa_j + \sum_{j>j_0} 2^{j(5/2-1/\rho)}\Big(\sum_k |b_{j,k}|^\rho \sum_{m\geqslant 0} \mathbf{1}_{2^{-m}\kappa_j<|b_{j,k}|\leqslant 2^{-m+1}\kappa_j}\Big)^{1/\rho}$$

$$\lesssim 2^{3j_0/2} T^{-1/2} (\log T)^{1/2} + \sum_{j>j_0} 2^{j(5/2-1/\rho)}\Big(\sum_k \sum_{m\geqslant 0} 2^{(-m+1)\rho}\kappa_j^\rho \mathbf{1}_{|b_{j,k}|>2^{-m}\kappa_j}\Big)^{1/\rho}$$

$$\lesssim \Big(\frac{T}{\log T}\Big)^{-s/(2s+3)} + \sum_{m\geqslant 0}\sum_{j>j_0} 2^{j(5/2-1/\rho)} 2^{-m}\kappa_j |\{k \,|\, |b_{j,k}| > 2^{-m}\kappa_j\}|^{1/\rho}$$

$$\lesssim \Big(\frac{T}{\log T}\Big)^{-s/(2s+3)} + \sum_{m\geqslant 0} 2^{-j_0(ps/\rho+3p/2\rho-3/2)}\Big(\frac{2^{-m}(\log T)^{1/2}}{T^{1/2}}\Big)^{(\rho-p)/\rho}$$

$$\sim \Big(\frac{T}{\log T}\Big)^{-s/(2s+3)}.$$

All estimates together yield $\mathbb{E}_a[\|\hat{b}_T - Q_a a\|_{W^{2,\rho}}] \lesssim (T/(\log T))^{-s/(2s+3)}$ with uniform constants.

6.3. Proof of Theorem 3.3

Due to $\hat{b}_T \in W^{2,\rho}([-r,0])$, the kernel $\hat{q}_T|_{[0,r]}$ is an element of $W^{3,\rho}([0,r])$ and the continuity of $\hat{Q}_T : \mathcal{L}^\rho \to W^{2,\rho}([-r,0])$ follows from Lemma A.3. Formally, the von Neumann series expansion yields for $\hat{Q}_T^{-1}$,

$$\hat{Q}_T^{-1} = (\mathrm{Id} - Q_a^{-1}(Q_a - \hat{Q}_T))^{-1} Q_a^{-1} = \sum_{m=0}^{\infty} (Q_a^{-1}(Q_T - \hat{Q}_T))^m Q_a^{-1}.$$

Introducing the random set

$$\mathcal{C}_T := \{\|Q_a^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho} \|Q_a - \hat{Q}_T\|_{\mathcal{L}^\rho\to W^{2,\rho}} \leqslant \tfrac{1}{2}\},$$

the operator $\hat{Q}_T$ is therefore invertible on $\mathcal{C}_T$ with

$$\|\hat{Q}_T^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho} \leqslant 2\|Q_a^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho},$$
$$\|\hat{Q}_T^{-1} - Q_a^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho} \leqslant 2\|Q_a^{-1}\|^2_{W^{2,\rho}\to\mathcal{L}^\rho} \|\hat{Q}_T - Q_a\|_{\mathcal{L}^\rho\to W^{2,\rho}}.$$

In order to bound the probability of $\mathcal{C}_T$ from below, we use the estimate $\|Q_a - \hat{Q}_T\|_{\mathcal{L}^\rho \to W^{2,\rho}} \lesssim \|q_a - \hat{q}_T\|_{W^{3,\rho}([-r,0])}$, derived from Lemma A.3 with $k(t) = (q_a - \hat{q}_T)(-t)$ and $W^{3,\rho} \subset W^{2,\rho'}$. From Proposition 3.2 we know that

$$\mathbb{E}_a[\|q_a' - \hat{q}_T'\|_{W^{2,\rho}}] = \mathbb{E}_a[\|Q_a a(-\cdot) - \hat{b}_T(-\cdot)\|_{W^{2,\rho}}] \lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}.$$

Furthermore, we infer from Propositions 5.1 and 5.4 that

$$\mathbb{E}_a\left[\left|q_a(0) - \frac{1}{T}\int_0^T X(u)^2\,\mathrm{d}u\right|^2\right] = \frac{1}{T^2}\int_0^T\int_0^T 2q_a^2(u-v)\mathrm{d}u\,\mathrm{d}v \lesssim \frac{1}{T}$$

with uniform constants. We conclude that

$$\mathbb{E}_a[\|q_a - \hat{q}_T\|_{W^{3,\rho}}] \lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}. \tag{6.4}$$

Finally, Markov's inequality yields, for suitable $c > 0$,

$$\begin{aligned}
\sup_{a\in M(s,p,S,\delta)} \mathbb{P}_a(\Omega\backslash\mathcal{C}_T) &\leqslant \sup_{a\in M(s,p,S,\delta)} \mathbb{P}_a(\|q_a - \hat{q}_T\|_{W^{3,\rho}} > c)\\
&\leqslant \sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|q_a - \hat{q}_T\|_{W^{3,\rho}}]c^{-1}\\
&\lesssim \left(\frac{T}{\log T}\right)^{-s/(2s+3)}.
\end{aligned}$$

It therefore suffices to work on the set $\mathcal{C}_T$, because on its complement the loss is bounded by $2S$. Since our renormalization uses the a priori knowledge $\|a\|_{\mathcal{L}^\rho} \leqslant S$, our estimator is on $\mathcal{C}_T$ only up to a constant factor worse than the estimator obtained by pure inversion. We obtain on $\mathcal{C}_T$,

$$\begin{aligned}
\|\hat{a}_T - a\|_{\mathcal{L}^\rho} &\lesssim \hat{Q}_T^{-1}\hat{b}_T - Q_a^{-1}Q_a a\|_{\mathcal{L}^\rho}\\
&\leqslant \|\hat{Q}_T^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho}\|\hat{b}_T - Q_a a\|_{W^{2,\rho}}\\
&\quad + \|\hat{Q}_T^{-1} - Q_a^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho}\|Q_a a\|_{W^{2,\rho}}\\
&\leqslant 2\|Q_a^{-1}\|_{W^{2,\rho}\to\mathcal{L}^\rho}\|\hat{b}_T - Q_a a\|_{W^{2,\rho}}\\
&\quad + 2\|Q_a^{-1}\|^2_{W^{2,\rho}\to\mathcal{L}^\rho}\|Q_a - \hat{Q}_T\|_{\mathcal{L}^\rho\to W^{2,\rho}}\|Q_a a\|_{W^{2,\rho}}\\
&\lesssim \|\hat{b}_T - Q_a a\|_{W^{2,\rho}} + \|q_a - \hat{q}_T\|_{W^{3,\rho}}.
\end{aligned}$$

By Theorem 2.4 the last estimate holds uniformly for all $a \in M(s, p, S, \delta)$.

From Proposition 3.2 and the estimate (6.4) we conclude that

$$\sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{\mathcal{L}^\rho}\mathbf{1}_{\mathcal{C}_T}] \leqslant T^{-s/(2s+3)},$$

which accomplishes the proof of the asymptotic risk upper bound.

7. Proof of the lower bound

Proof of Theorem 4.1. We build from a weight a_0 in the interior of $M(s, p, S, \delta)$ a family of local alternatives (a_{jk}). Choose a compactly supported s-regular wavelet basis in $L^2(\mathbb{R})$ and denote by R_j a maximal set of integers with $\operatorname{supp}(\psi_{jk}) \subset [-r, 0]$ and $\operatorname{supp}(\psi_{jk}) \cap \operatorname{supp}(\psi_{jk'}) = \varnothing$ for all $k, k' \in R_j$, $k \neq k'$. For any $k \in R_j$ we set $a_{jk} := a_0 + \gamma\psi_{jk}$ with $\gamma = \gamma(T) \sim 2^{-j(T)(s+1/2-1/p)}$ such that $\|a_{jk}\|_{s,p,1} \leqslant S$ and $v_0(a_{jk}) \leqslant -\delta$ are satisfied, hence $a_{jk} \in B(s, S, p, \delta)$. We are using a classical lemma on lower bounds in the sparse case:

Lemma 7.1. *Suppose the likelihood ratio satisfies*

$$\mathbb{P}_{a_{jk}}(\log(\Lambda_T(X^{(a_0)}, X^{(a_{jk})})) \geqslant -j) \geqslant p_* > 0$$

uniformly for all a_{jk}. *Then for* $\mathcal{F}_T^X$*-measurable estimators* $\hat{a}_T$ *we have the lower bound*

$$\inf_{\hat{a}_T} \sup_{a\in M(s,p,S,\delta)} \mathbb{E}_a[\|\hat{a}_T - a\|_{L^\rho}] \geqslant \gamma(T)2^{j(T)(1/2-1/\rho)} \sim 2^{-j(T)(s+1/\rho-1/p)}.$$

This is an adapted version of Korostelev and Tsybakov (1993, Theorem 2.4.3). Note the relations $M \sim 2^j$ and $s_T \sim \gamma 2^{j(1/2-1/\rho)}$ in their statement, having substituted n by T.

We use the likelihood ratio from Theorem 2.5 with some fixed initial condition and apply Lemma A.5 and estimate (6.1):

$$\begin{aligned}
&\mathbb{E}_{a_{jk}}[\log(\Lambda_T(X^{(a_0)}, X^{(a_{jk})}))^2] \\
&\quad = \mathbb{E}_{a_{jk}}\left[\left(\int_0^T \langle X(t+\cdot), a_{jk} - a_0\rangle \mathrm{d}W(t) - \tfrac{1}{2}\langle Q_T(a_{jk} - a_0), a_{jk} - a_0\rangle\right)^2\right] \\
&\quad \leqslant 2\gamma^2 T\langle Q_{a_{jk}}\psi_{jk}, \psi_{jk}\rangle + \tfrac{1}{2}\mathbb{E}_{a_{jk}}[\langle Q_T(a_{jk} - a_0), a_{jk} - a_0\rangle^2] \\
&\quad \leqslant 2\gamma^2 T\langle Q_{a_{jk}}\psi_{jk}, \psi_{jk}\rangle + \gamma^4 T^2\langle Q_{a_{jk}}\psi_{jk}, \psi_{jk}\rangle^2 \\
&\qquad + \gamma^4\mathbb{E}_{a_{jk}}[\langle (Q_T - TQ_{a_{jk}})\psi_{jk}, \psi_{jk}\rangle^2] \\
&\quad \lesssim \gamma^2 T 2^{-2j} + \gamma^4 T^2 2^{-4j} + \gamma^4 T 2^{-4j}
\end{aligned}$$

with a uniform constant for all a_{jk}. Thus, by Chebyshev's inequality the requirements of Lemma 7.1 are satisfied, when we balance the restrictions on γ by choosing $2^{(2s+3-2/p)j(T)} \sim T/\log T$ such that

$$\gamma(T)^4 T^2 2^{-4j(T)} \sim T^2 2^{-j(T)(4s+6-4/p)} \sim (\log T)^2 \sim j(T)^2.$$

The lower bound follows. □

Proof of Corollary 4.2. (i) The lower bound is just Theorem 4.1 properly rewritten. For the upper bound use the embedding $\mathcal{B}^s_{p,1} \subset \mathcal{B}^w_{\pi,1}$ with $1/\pi := w - s + 1/p < 1/p$. Due to $1/\pi - 1/\rho = 2/\rho \times w/3$, we can apply Theorem 3.3 to the class $M(w, \pi, S', \delta)$, S' chosen appropriately.

(ii) The upper bound is the content of Theorem 3.3, whereas the lower bound follows along the lines of the L^2-lower bound proof using Assouad's cube in Reiss (2002). The details are omitted. □

Appendix

A.1. Function spaces

For a more detailed account, see Triebel (1983). Let us introduce the scale of Sobolev spaces $W^{m,p}(I)$, $m \in \mathbb{N}$, $p \in [1, \infty]$, $I \subset \mathbb{R}$ an interval: $W^{m,p}(I) := \{f \in L^p(I) | f^{(i)} \in L^p(I) \text{ for all } i = 0, \ldots, m\}$, where $f^{(i)}$ denotes the ith derivative of f in a weak sense. These are Banach spaces with respect to the norm $\|f\|_{m,p} := (\sum_{i=0}^m \|f^{(i)}\|^p_{L^p})^{1/p}$.

Besov spaces $B^s_{p,\alpha}$, measure the regularity s in an L^p-sense with an additional fine-tuning parameter $\alpha \in [1, \infty]$.

Definition A.1. *Let $I \subset \mathbb{R}$ be an interval, $\Delta_h f(x) := f(x+h) - f(x)$ and $I_h := \{x \in I | x \pm h \in I\}$. Then the nth-order L^p-modulus of smoothness is defined by*

$$\omega_n(f, \varepsilon)_p := \sup_{|h| \leqslant \varepsilon} \|\Delta_h^n f\|_{L^p(I_{nh})},$$

with Δ_h^n denoting the n-fold application of Δ_h. For $p, \alpha \in [1, \infty]$ and $s > 0$, set

$$\|f\|_{s,p,\alpha} := \|f\|_{L^p(I)} + \left(\int_0^1 \left(\frac{\omega_n(f, t)_p}{t^s}\right)^\alpha \frac{\mathrm{d}t}{t}\right)^{1/\alpha}$$

with the usual modification $\sup_t \omega_n(f, t)_p t^{-s}$ for $\alpha = \infty$ and with $n = \lfloor s \rfloor + 1$. The Besov space $B^s_{p,\alpha}(I) := \{f \in L^p(I) | \, \|f\|_{s,p,\alpha} < \infty\}$ is a Banach space when equipped with the norm $\|\cdot\|_{s,p,\alpha}$ On a bounded interval an equivalent norm is given by $\|f\|_{B^s_{p,\alpha}} \sim \|f\|_{L^p} + \|f^{(n-1)}\|_{s-(n-1),p,\alpha}$, with n as above.

Proposition A.1. *The following embedding relations hold true: $B^s_{p,\alpha} \subset B^{s'}_{p,\alpha'}$, $s > s'$, any α, α'; $B^s_{p,\alpha} \subset B^s_{p',\alpha}$, $p > p'$; $B^s_{p,\alpha} \subset B^s_{p,\alpha'}$, $\alpha < \alpha'$. The Sobolev embedding theorem generalizes to*

$$B^s_{p,\alpha} \subset B^{s'}_{p',\alpha} \qquad \text{for } s \geqslant s' \text{ and } s - \frac{1}{p} \geqslant s' - \frac{1}{p'}. \tag{A.1}$$

As a special case $B^s_{p,\alpha} \subset C^{s'}$ for $s - \frac{1}{p} > s'$ follows. The first embedding is compact for Besov spaces on bounded intervals.

The regularity property of convolutions with variable integral bound seems obvious, but does not appear to have been treated in the literature.

Lemma A.2. *For functions* $f \in B^s_{p,\alpha}([-r, 0])$ *and* $k \in B^{s+1}_{p',\alpha'}([0, r])$*, where* $s > 0$*,* $p, p' \in (1, \infty)$ *and* $\alpha, \alpha' \in [1, \infty]$ *with* $1/p + 1/p = 1/\alpha + 1/\alpha' = 1$*, set*

$$L(f, k)(t) := \int_0^t f(u - t)k(u)\mathrm{d}u, \qquad t \in [0, r].$$

Then L *is a bilinear mapping from* $B^s_{p,\alpha}([-r, 0]) \times B^{s+1}_{p',\alpha'}([0, r])$ *to* $B^{s+1}_{p,\alpha}([0, r])$ *with* $L(f, k)\|_{s+1,p,\alpha} \lesssim \|f\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}$.

Proof. First, we show for a fixed function f in $L^p([-r, 0])$ that $Tk := L(f, k)$ maps $B^s_{p',\alpha'}([0, r])$ to $B^s_{p,\alpha}([0, r])$ for $s \in (0, 1)$ and all p and α.

In order to apply abstract interpolation theory, we consider the case $s = 1$ in a Sobolev scale first:

$$\begin{aligned}
\|Tk\|_{W^{1,p}} &\sim \|Tk\|_{L^p} + \|(Tk)'\|_{L^p} \\
&\lesssim \|Tk\|_{L^\infty} + \left\| \left(\int_{-\cdot}^0 f(v)k(v + \cdot)\mathrm{d}v \right)' \right\|_{L^p} \\
&\lesssim \|f\|_{L^p}\|k\|_{L^{p'}} + \left\| f(-\cdot)k(0) + \int_0^{\cdot} f(u - \cdot)k'(u)\mathrm{d}u \right\|_{L^p} \\
&\lesssim \|f\|_{L^p}\|k\|_{L^{p'}} + \|f\|_{L^p}\|k\|_\infty + \|T(k')\|_{L^p} \\
&\lesssim \|f\|_{L^p}\|k\|_{W^{1,p'}}
\end{aligned}$$

Due to $\|Tk\|_\infty \leqslant \|f\|_{L^p}\|k\|_{L^{p'}}$ the real interpolation theory (Triebel 1983, Theorem 3.3.6) yields, for all $s \in (0, 1)$,

$$\|Tk\|_{s,p,\alpha} \lesssim \|f\|_{L^p}\|k\|_{s,p',\alpha'}.$$

In a second step, we use an induction argument from s to $s + 1$ for non-integer $s > 0$. Suppose $f \in B^s_{p,\alpha}$ and $k \in B^{s+1}_{p',\alpha}$ The weak derivative of $L(f, k)$ is given by

$$L(f, k)'(t) = f(-t)k(0) + L(f, k')(t), \qquad t \in [0, r]$$

(see above), which yields, for $s \in (0, 1)$,

$$\|L(f, k)'\|_{s,p,\alpha} \leqslant \|f\|_{s,p,\alpha}\|k\|_\infty + \|T(k')\|_{s,p,\alpha} \lesssim \|f\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}$$

and a fortiori for $s > 1$, $s \notin \mathbb{N}$, by induction

$$\|L(f, k)'\|_{s,p,\alpha} \leqslant \|f\|_{s,p,\alpha}\|k\|_\infty + \|f\|_{s-1,p,\alpha}\|k'\|_{s,p',\alpha'}\| \lesssim \|f\|_{s,p,\alpha}\|k\|_{s+1,p',\alpha'}.$$

Since the very first argument provided an estimate for $\|L(f, k)\|_{L^p}$ of the same type, the norm $\|L(f, k)\|_{s+1,p,\alpha}$ is bounded.

Finally, the same induction argument for $s \in \mathbb{N}$ requires an extra estimate for $\|T(k^{(s)})\|_{0,p,\alpha}$. Since f is in $B^s_{p,\alpha} \subset L^\infty$ and k in $B^{s+1}_{p',\alpha'} \subset C^s$, we infer, from

$$\left(\int_0^t f(u-t)k^{(s)}(u)\mathrm{d}u\right)'(t) = f(0)k^{(s)}(t) - \int_0^t f'(u-t)k^{(s)}(u)\mathrm{d}u$$

and the convolution estimate (use Triebel 1983, Theorem 2.11.2, Proposition 3.3.2)

$$\sup_{t\in\mathbb{R}} |(f'\mathbf{1}_{\mathbb{R}^-}) * (k^{(s)}\mathbf{1}_{\mathbb{R}^+})(t)| \lesssim \|f'\|_{0,p,\alpha}\|k^{(s)}\|_{0,p',\alpha'},$$

that $\|T(k^{(s)})\|_{C^1} \lesssim \|f\|_{1,p,\alpha}\|k\|_{s+1,p',\alpha'}$. □

Lemma A.3. *Suppose that k is a function in* $W^{2,\rho'}([0,r])$, $\rho' \in (1,\infty)$, *and that* ρ *satisfies* $1/\rho + 1/\rho' = 1$. *Then the integral operator*

$$Kf(t) := \int_{-r}^0 k(|t-s|)f(s)\mathrm{d}s, \qquad t \in [-r, 0],$$

is continuous from $L^\rho([-r,0])$ *to* $W^{2,\rho}([-r,0])$ *with* $\|K\|_{L^\rho \to W^{2,\rho}} \lesssim \|k\|_{W^{2,\rho'}}$.

Proof. First consider the following identities in an almost everywhere sense for $t \in [-r, 0]$ and $f \in L^\rho([-r,0])$:

$$(Kf)'(t) = \int_{-r}^0 k'(|t-s|)\mathrm{sgn}(t-s)f(s)\mathrm{d}s,$$

$$(Kf)''(t) = \left(\int_{-r}^{\cdot} k'(\cdot - s)f(s)\mathrm{d}s - \int_{\cdot}^0 k'(s-\cdot)f(s)\mathrm{d}s\right)'(t)$$

$$= \int_{-r}^0 k''(|t-s|)f(s)\mathrm{d}s + 2k'(0)f(t).$$

By the Hölder inequality, we obtain

$$\|Kf\|_{L^\rho} \leqslant 2\|k\|_{L^{\rho'}}\|f\|_{L^\rho}, \qquad \|(Kf)''\|_{L^\rho} \leqslant \rho\|k''\|_{L^{\rho'}}\|f\|_{L^\rho} + \rho\|k'\|_\infty\|f\|_{L^\rho}.$$

The Sobolev embedding $W^{2,\rho'} \subset C^1$ proves $\|Kf\|_{W^{2,\rho}} \lesssim \|k\|_{W^{2,\rho'}}\|f\|_{L^\rho}$. □

A.2. Wavelets

In the following definition we largely follow Cohen (2000).

Definition A.2. *For* $j, k \in \mathbb{Z}$, *introduce the multi-index* $\lambda = (j,k)$ *and put* $|\lambda| := |(j,k)| := j$. *A wavelet basis* $(\psi_\lambda)_\lambda$ *is an orthonormal basis of functions in* $L^2(\mathbb{R})$, *derived from one function* $\psi \in L^2(\mathbb{R})$ *by translations and dilations*

$$\psi_\lambda(x) := \psi_{jk}(x) := 2^{j/2}\psi(2^j x - k).$$

Furthermore, set V_j *as the closure of* $\mathrm{span}(\psi_\lambda, |\lambda| \leqslant j)$. *By* $P_j : L^2([-r, 0]) \to V_j$ *we denote the orthogonal projection onto* V_j.

Cohen *et al.* (1993) constructed orthonormal wavelet bases on a bounded interval I. The basis functions are obtained by restriction. Wavelet functions ψ_λ whose support crosses the boundary of I are suitably corrected in order to keep the orthogonality and approximation properties. These corrected functions are still denoted by ψ_λ even if they are not directly derived from ψ. A consequence of this construction is that only multi-indices $\lambda = (j, k)$ with $|k| \lesssim 2^j$ are used and that the spaces V_j are finite-dimensional, whence we can start off with a space V_{-1} and an orthonormal basis $(\psi_{-1,k})_k$ of V_{-1}. Then any function $f \in L^2(I)$ has the wavelet decomposition

$$f = \sum_\lambda \langle f, \psi_\lambda \rangle \psi_\lambda = \sum_{j \geqslant -1} \sum_k \langle f, \psi_{jk} \rangle \psi_{jk}.$$

Note that summation over $|\lambda| \leqslant j_0$ will always mean summation over (j, k) for all $j \leqslant j_0$ and all possible values of k.

Definition A.3. *A wavelet basis* (ψ_λ) *will be called s-regular on the interval I if the following two conditions are satisfied:*

(i) *For all* $w \in (0, s]$, $p, \alpha \in [1, \infty]$, *the function f is an element of* $B^w_{p,\alpha}(I)$ *if and only if*

$$\|P_{-1}f\|_{L^p} + \left(\sum_{j=0}^{\infty} 2^{\alpha j (w+1/2-1/p)} \left(\sum_k |\langle f, \psi_{jk} \rangle|^p \right)^{\alpha/p} \right)^{1/\alpha} < \infty.$$

The above expression constitutes a norm equivalent to $\|\cdot\|_{w,p,\alpha}$.

(ii) *For all* $k = 0, \ldots, \lfloor s \rfloor$, *the vanishing moment property*

$$\int_{-\infty}^{\infty} x^k \psi(x) \mathrm{d}x = 0.$$

is fulfilled.

Sufficiently regular wavelets guarantee that, for $m, s > 0$ and $\rho \in [1, \infty]$, the general Jackson inequality

$$\|f - P_J d\|_{W^{m,\rho}} \lesssim 2^{-Js} \|f\|_{W^{m+s,\rho}} \qquad \text{for all } f \in W^{m+s,\rho} \tag{A.2}$$

holds (Cohen 2000). Despite their different notion of s-regularity, from Cohen *et al.* (1993) we immediately obtain:

Theorem A.4. *s-regular wavelet bases exist for any* $s > 0$. *Moreover, they may be chosen to have compact support.*

Lemma A.5. *Let* (ψ_λ) *be a 1-regular wavelet basis. Then for any weight measure a with*

$\|a\|_{\mathrm{TV}} \leqslant R < \infty$ *and* $v_0(a) \leqslant -\delta < 0$ *and for any multi-index* λ, *we have* $\langle Q_a\psi_\lambda, \psi_\lambda\rangle \sim 2^{-2|\lambda|}$ *uniformly.*

Proof. Using the formula for the spectral density (2.3), estimates as in the proof of Lemma 5.2 and the spectral characterization of the space $W^{1,2}$, we obtain

$$\langle Q_a\psi_\lambda, \psi_\lambda\rangle = \int_{-\infty}^{\infty}\left|\frac{\hat{\psi}_\lambda(\xi)}{\mathrm{i}\xi - \int_{-r}^{0} \mathrm{e}^{\mathrm{i}\xi u} > \mathrm{d}a(u)}\right|^2 \mathrm{d}\xi \sim \int_{-\infty}^{\infty}(1+\xi^2)^{-1}|\hat{\psi}_\lambda(\xi)|^2\,\mathrm{d}\xi$$

$$= \sup_{\|f\|_{L^2}=1}\left|\int_{-\infty}^{\infty}(1+\xi^2)^{-1/2}f(\xi)\hat{\psi}_\lambda(\xi)\mathrm{d}\xi\right|^2 = \sup_{\|h\|_{W^{1,2}}=1}\langle h, \psi_\lambda\rangle^2.$$

The last expression is clearly of order $2^{-2|\lambda|}$. □

Lemma A.6. *Given that* $f \in C^{m,1}([-r, r])$, $m \in \mathbb{N}_0$, *i.e.* $f^{(m)}$ *is Lipschitz continuous, suppose that* (ψ_λ) *is a compactly supported* $(m+1)$*-regular wavelet basis of* $L^2([-r, 0])$. *Then*

$$\left|\int_{-r}^{0}\int_{-r}^{0} f(x-y)\psi_\lambda(x)\psi_\lambda(y)\mathrm{d}y\,\mathrm{d}x\right| \lesssim \|f\|_{C^{m,1}}2^{-|\lambda|(m+2)}$$

with a constant independent of f and of the multi-index λ.

Proof. Note that for $y \in [-r, 0]$ the function $f(\cdot - y)|_{[-r,0]}$ lies in $C^{m,1}([-r, 0])$. By the $(m+1)$-regularity of (ψ_λ) we find that

$$\left|\int_{-r}^{0}\int_{-r}^{0} f(x-y)\psi_\lambda(x)\psi_\lambda(y)\mathrm{d}y\,\mathrm{d}x\right|$$

$$\leqslant \sup_{y\in[-r,0]}|\langle f(\cdot - y), \psi_\lambda\rangle|\,\|\psi_\lambda\|_{L^1}$$

$$\lesssim \sup_{y\in[-r,0]}\|f(\cdot - y)\|_{C^{m,1}}2^{-|\lambda|(m+1+1/2)}2^{-|\lambda|/2}\|\psi\|_{L^1}.$$

Note that we have used the embedding $C^{m,1} \subset B^{m+1}_{\infty,\infty}$. □

References

Abramovich, F. and Silverman, B. (1998) Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

Butucea, C. and Matias, C. (2003) Minimax estimation of the noise level and of the signal density in a semiparametric convolution model. Preprint no. 795, Laboratoire de Probabilités et Modèles Aléatoires, Paris.

Cavalier, L. and Hengartner, N.W. (2002) Adaptive estimation for inverse problems with noisy operators. Manuscript.

Cavalier, L. and Tsybakov, A. (2002) Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, **123**(3), 323–354.

Cavalier, L., Golubev, G., Picard, D. and Tsybakov, A. (2002) Oracle inequalities for inverse problems. *Ann. Statist.*, **30**, 843–874.

Cohen, A. (2000). Wavelet methods in numerical analysis. In P.G. Ciarlet (ed.), *Handbook of Numerical Analysis, Vol. 7*. Amsterdam: North-Holland/Elsevier.

Cohen, A., Daubechies, I. and Vial, P. (1993) Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, **1**(1), 54–81.

Cohen, A., Hoffmann, M. and Reiss, M. (2004) Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Num. Anal.* To appear.

Dalalyan, A. (2001) Asymptotically efficient estimation for ergodic diffusion processes. Doctoral thesis, University of Le Mans, France.

Donoho, D.L. (1995) Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, **2**(2), 101–126.

Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Efromovich, S. and Koltchinskii, V. (2001) On inverse problems with unknown operators. *IEEE Trans. Inform. Theory*, **47**, 2876–2894.

Gushchin, A.A. and Küchler, U. (1999) Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli,* **5**, 1059–1098.

Gushchin, A.A. and Küchler, U. (2000) On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Process. Appl.*, **88**, 195–211.

Gushchin, A.A. and Küchler, U. (2003) On parametric statistical models for stationary solutions of affine stochastic delay differential equations. *Math. Methods Statist.*, **12**(1), 31–61.

Hale, J.K. and Verduyn Lunel, S.M. (1993) *Introduction to Functional Differential Equations*. New York: Springer-Verlag.

Hämarik, U. (1983) Projection methods for regularizing ill-posed problems (in Russian). *Tr. Vychisl. Tsentra Tartu*, **50**, 69–90.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*. Berlin: Springer-Verlag.

Hobson, D.G. and Rogers, L. (1998) Complete models with stochastic volatility. *Math. Finance*, **8**(1), 27–48.

Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, **79**(1), 135–163.

Kalifa, J. and Mallat, J. (2003) Thresholding estimators for linear inverse problems. *Ann. Statist.*, **31**(1), 58–109.

Katznelson, Y. (1976) *An Introduction to Harmonic Analysis.* 2nd corr. ed. New York: Dover.

Kerkyacharian, G. and Picard, D. (2000) Thresholding algorithms, maxisets and well-concentrated bases. *Test*, **9**(2), 283–344.

Kolmanovskii, V. and Myshkis, A. (1992) *Applied Theory of Functional Differential Equations*, Math. Appl. (Soviet Ser.) 85. Dordrecht: Kluwer.

Korostelev, A. and Tsybakov, A. (1993) *Minimax Theory of Image Reconstruction*, Lecture Notes in Statist. 82. New York: Springer-Verlag.

Küchler, U. and Mensch, B. (1992) Langevin's stochastic differential equation extended by a time-delayed term. *Stochastic Stochastics Rep.*, **40**(1/2), 23–42.

Küchler, U. and Sørensen, M. (1997) *Exponential Families of Stochastic Processes.* New York: Springer-Verlag.

Kutoyants, Y., Mourid, T. and Bosq, D. (1992) Estimation paramétrique d'un processus de diffusion avec retards. *Ann. Inst. H. Poincaré Probab. Statist.*, **28**(1), 95–106.

Kydland, F.E. and Prescott, E.C. (1982) Time to build and aggregate fluctuations. *Econometrica*, **50**, 1345–1370.

Lipster, R. and Spokoiny, V. (2000) Deviation probability bound for martingales with applications to statistical estimation. *Statist. Probab. Lett.*, **46**(4), 347–357.

Lipster, R.S. and Shiryaev, A.N. (2001) *Statistics of Random Processes. 1: General Theory* (2nd edition). Berlin: Springer-Verlag.

Mao, X. (1997) *Stochastic Differential Equations and Their Applications.* Chichester: Horwood.

Mohammed, S.-E. (1984) *Stochastic Functional Differential Equations*. Boston: Pitman.

Mohammed, S.-E.A., Scheutzow, M. and von Weizsäcker, H. (1986) Hyperbolic state space decomposition for a linear stochastic delay equation. *SIAM J. Control Optim.*, **24**, 543–551.

Neumann, M. and von Sachs, R. (1995) Wavelet thresholding: Beyond the Gaussian i.i.d. situation. In A. Antoniadis (ed.), *Wavelets and Statistics. Proceedings of the 15th French-Belgian Meeting of Statisticians, 1994*. New York: Springer-Verlag.

Nussbaum, M. and Pereverzev, S.V. (1999) The degrees of ill-posedness in stochastic and deterministic noise models. Preprint 509, Weierstrass Institute Berlin.

Reiss, M. (2001) Nonparametric estimation for stochastic delay differential equations. Doctoral thesis, Humboldt Universität zu Berlin, http://dochost.rz.hu-berlin.de.

Reiss, M. (2002) Minimax rates for nonparametric drift estimation in affine stochastic delay equations. *Statist. Inference Stochastic Process.*, **5**(2), 131–152.

Reiss, M. (2004) Nonparametric estimation of estimation of the delay time in affine stochastic differential equations. *Internat. J. Wavelets Multiresolution Inform. Processing.* To appear.

Rudin, W. (1991) *Functional Analysis* (2nd edition). New York: McGraw-Hill.

Scheutzow, M. (1984) Qualitative behaviour of stochastic delay equations with a bounded memory., *Stochastics*, **12**, 41–80.

Triebel, H. (1983) *Theory of Function Spaces*, Mathematik und ihre Anwendungen in Physik und Technik 38. Leipzig: Akademische Verlagsgesellschaft Geest & Portig.

Vakhaniya, N., Tarieladze, V. and Chobanyan, S. (1987) *Probability Distributions on Banach Spaces.* Dordrecht: Reidel.

Received May 2003 and revised January 2004