Gushchin, Alexander A.; Küchler, Uwe

Working Paper

On parametric statistical models for stationary solutions of affine stochastic delay differential equations

SFB 373 Discussion Paper, No. 2001,91

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

On parametric statistical models for stationary solutions of affine stochastic delay differential equations*

Alexander A. Gushchin
Steklov Mathematical Institute
Gubkina 8
117966 Moscow GSP-1, Russia

Uwe Küchler
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
D-10099 Berlin, Germany

November 8, 2001

Abstract

Assume that we observe a stationary Gaussian process $X(t)$, $t \in [-r, T]$, which satisfies the affine stochastic delay differential equation

$$dX(t) = \int_{[-r,0]} X(t+u)\, a_\vartheta(du)\, dt + dW(t), \qquad t \geq 0,$$

where a_ϑ is a finite signed measure, the parameter ϑ belongs to an open set $\Theta \subseteq \mathbb{R}^k$, is unknown and has to be estimated based on the observation. Conditions are derived under which this model satisfies the local asymptotic normality property for every $\vartheta \in \Theta$ as $T \to \infty$ and the maximum likelihood and Bayesian estimators of ϑ are asymptotically normal and efficient.

Keywords: affine stochastic delay differential equation, stationary Gaussian process, local asymptotic normality, maximum likelihood estimator, Bayesian estimator, Hellinger distance

*Support of Deutsche Forschungsgemeinschaft (SFB 373, “Quantification and Simulation of Economic Processes”, Humboldt-Universität zu Berlin) and of DFG–RFBR Project 98-01-04108 is gratefully acknowledged. The first author would like to thank Vladimir Bogachev for useful discussions.

1 Introduction

Let J be a fixed finite interval $[-r, 0]$, $r \geq 0$. Denote by M the set of all finite signed measures on J. For $a \in M$ consider the equation

$$\begin{aligned} X(t) &= X(0) + \int_0^t \int_J X(s+u)\, a(du)\, ds + W(t), && t \geq 0, \\ X(t) &= X_0(t), && t \in J, \end{aligned} \tag{1.1}$$

where $W = (W(t),\, t \geq 0)$ is a standard Wiener process and $X_0 = (X_0(t),\, t \in J)$ is a continuous initial process *independent* of W. If W, X_0 and a are given, then there is a unique continuous process $X = (X(t),\, t \geq -r)$ satisfying (1.1) for all ω, and the law of X is uniquely determined by a and the law of X_0, see the details in Section 3.
Given $a \in M$, one can ask whether a stationary solution to (1.1) exists under an appropriate choice of the law of X_0. A necessary and sufficient condition in terms of a for the existence of a stationary solution is given in Gushchin and Küchler (2000), see also Section 3. The stationary solution is a Gaussian process with zero mean and a strictly positive spectral density uniquely determined by a. We denote by M_s the class of all $a \in M$ for which a stationary solution to (1.1) exists.
Now assume that we observe continuously a single realization $(X(t),\, -r \leq t \leq T)$ satisfying (1.1) with an unknown a which is assumed to belong to a parametric family $\mathcal{A} = \{a_\vartheta \in M \colon \vartheta \in \Theta\}$, where Θ is an open subset of $\mathbb{R}^k$. The problem under consideration is to estimate ϑ based on the observation and to study asymptotic properties of the estimators as $T \to \infty$. It is implicitly assumed in the subsequent discussion that the statistical information about ϑ in the initial data is negligible for the considered asymptotics.
This statistical problem has been considered only for concrete models and in few papers up to now. We mention the papers by Dietz (1992) (though the model considered there slightly differs from (1.1)), Gushchin and Küchler (1999, 2001), Küchler and Kutoyants (2000), Putschke (2000).
In Gushchin and Küchler (1999, 2001) we considered the model

$$\vartheta = (\vartheta_1, \vartheta_2) \in \Theta = \mathbb{R}^2, \qquad a_\vartheta = \vartheta_1 \delta_{\{0\}} + \vartheta_2 \delta_{\{-1\}}$$

($\delta_{\{x\}}$ stands for the Dirac measure at x). An important feature of this model is that the mapping $\vartheta \rightsquigarrow a_\vartheta$ from Θ to M is affine. Then the log-likelihood function is quadratic in ϑ, which allowed us to perform a direct analysis of the asymptotic behaviour of the likelihood ratios and the maximum likelihood estimator. It seems to us that a similar analysis can be performed for any model such that the mapping $\vartheta \rightsquigarrow a_\vartheta$ is affine (and one-to-one), see e.g. Putschke (2000) for the case $\vartheta = (\vartheta_0, \ldots, \vartheta_N) \in \Theta = \mathbb{R}^{N+1}$, $a_\vartheta = \sum_{i=0}^N \vartheta_i \delta_{\{-r_i\}}$.
In this paper we study the general problem under two basic assumptions: (1) the model is ergodic (i.e. $a_\vartheta \in M_s$) for every $\vartheta \in \Theta$; (2) at every $\vartheta \in \Theta$ the family $\mathcal{A}$ is differentiable in the following (weak) sense: there is a linear one-to-one operator A from $\mathbb{R}^k$ to M such that

$$a_\eta - a_\vartheta = A(\eta - \vartheta) + \|\eta - \vartheta\| r_\eta,$$

where the sequence of the measures r_{η_n} (considered as functionals on the space of continuous functions) $*$-weakly converges to 0 every time as η_n tends to ϑ. Under these two assumptions our parametric model will satisfy the local asymptotic normality property with scaling $T^{-1/2}$ at every $\vartheta \in \Theta$. Thus, the second assumption is similar to the differentiability in quadratic mean condition in the case of independent and identically distributed observations.

In fact, the assumptions under which our main results are proved are slightly more restrictive than stated above. The reason is that, under mild additional conditions, it turns out to be possible to prove asymptotic normality and efficiency of maximum likelihood and Bayesian estimators using the method suggested by Ibragimov and Has'minskii (1981).

In Küchler and Kutoyants (2000) the following model was considered:

$$\Theta = (\pi/2b, 0), \qquad a_\vartheta = b\delta_{\{\vartheta\}},$$

where $b < 0$ is a (known) constant. Here we have $a_\vartheta \in M_s$ for every $\vartheta \in \Theta$ but the second basic assumption is not satisfied in the above form (note that the mapping $\vartheta \rightsquigarrow a_\vartheta$ is differentiable in the sense of generalized functions but the derivatives cannot be represented as finite signed measures). In this model the local asymptotic normality property does not hold, the rate of convergence of estimators is T (not $T^{1/2}$), the maximum likelihood and Bayesian estimators have different limit distributions.

We do not discuss here what may happen if $a_\vartheta \in M \setminus M_s$. Let us only mention that, in general, one cannot expect to prove the local asymptotic normality property or to check Ibragimov–Has'minskii conditions.

Nonparametric estimation of the delay measure $a \in M_s$ based on observations of a stationary solution to (1.1) was considered in Reiss (2000).

The paper is organized as follows. In Section 2 the necessary notation and the main results are given. In Section 3 we recall some basic facts concerning affine delay differential equations and prove preliminary results for stationary solutions. In Section 4 we deal with pairs of arbitrary Gaussian measures (with zero means); the aim is to estimate from above the Hellinger-type distances between them and to specify this estimate in terms of spectral densities in the case of stationary Gaussian processes. In Section 5 we study an upper bound for the Hellinger integral for distributions of solutions to the equation (1.1). Finally, in Section 6 the main results are proved.

2 Notation and the main results

To formulate our assumptions and the main results let us introduce some notation.

Unless otherwise specified, $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ stand for the Euclidean distance and scalar product respectively. We also use the symbol $\|\cdot\|$ for the operator norm of an operator, while $\|\cdot\|_{(2)}$ denotes the Hilbert–Schmidt norm of a Hilbert–Schmidt operator. $\|\cdot\|_v$ stands for the total variation norm of a finite signed measure.

Let P and $\widetilde{P}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. If Q is a measure dominating P and $\widetilde{P}$, put $z = dP/dQ$ and $\widetilde{z} = d\widetilde{P}/dQ$. Recall that the Hellinger distance $\rho_m(P,\widetilde{P})$ of order m, $m \geq 1$, is defined by

$$\rho_m^m(P,\widetilde{P}) := \int\limits_\Omega |\widetilde{z}^{1/m} - z^{1/m}|^m \, dQ,$$

and the quantity

$$H(P,\widetilde{P}) := \int\limits_\Omega \sqrt{z\widetilde{z}}\, dQ = 1 - \frac{1}{2}\rho_2^2(P,\widetilde{P})$$

is the Hellinger integral. Define also

$$J(P,\widetilde{P}) := \begin{cases} \int\limits_\Omega \{z\log(z/z') + z'\log(z'/z)\}\, dQ, & \text{if } P \sim \widetilde{P}, \\ +\infty, & \text{otherwise.} \end{cases}$$

All these definitions do not depend on the choice of Q.
$\mathbf{C}(I)$ is the space of all continuous real-valued functions on an interval $I \subseteq \mathbb{R}$, and we write $\mathbf{C}_T$ instead of $\mathbf{C}([-r,T])$ and $\mathbf{C}$ instead of $\mathbf{C}(J)$.
Let $M^k = \underbrace{M \times \cdots \times M}_{k \text{ times}}$ and denote by $M^k_\#$ the set of all k-tuples $(a_1,\ldots,a_k) \in M^k$ such that $a_1, \ldots, a_k$ are linearly independent in M. The elements of M^k are considered as column-vectors, and integrals with respect to them are understood coordinatewise. We shall state the conditions on the family $\mathcal{A} = \{a_\vartheta \in M \colon \vartheta \in \Theta \subseteq \mathbb{R}^k\}$ under which the properties of estimators will be investigated.
Condition 1. Θ is a nonempty open bounded subset of $\mathbb{R}^k$.
Condition 2. There is a constant $C < \infty$ such that $\|a_\theta - a_\eta\|_v \leq C\|\vartheta - \eta\|$ for all $\vartheta, \eta \in \Theta$.
According to Condition 2, for any $\vartheta \in \overline{\Theta} \setminus \Theta$, where $\overline{\Theta}$ is the closure of Θ, the limit $\lim_{\eta\to\vartheta,\eta\in\Theta} a_\eta$ in $(M, \|\cdot\|_v)$ exists. We denote it by a_ϑ, $\vartheta \in \overline{\Theta} \setminus \Theta$, again.
Condition 3. $a_\vartheta \in M_s$ for any $\vartheta \in \overline{\Theta}$.
Condition 4. For any $g \in \mathbf{C}$, the function

$$\vartheta \rightsquigarrow \int\limits_J g(u)\, a_\vartheta(du)$$

is continuously differentiable in Θ with the gradient

$$\int\limits_J g(u)\, \dot{a}_\vartheta(du),$$

where $\dot{a}_\vartheta \in M^k_\#$, $\vartheta \in \Theta$.
Condition 5. $a_\vartheta \neq a_\eta$ for all $\vartheta \in \Theta$ and $\eta \in \overline{\Theta}$, $\vartheta \neq \eta$.

Remark 2.1: Comparing Conditions 2 and 4, one can see that they are not independent. Indeed, it follows from Condition 4 and the uniform boundedness principle that for any compact $\mathbf{K} \subset \Theta$ there is a constant $C = C(\mathbf{K})$ such that $\|a_\theta - a_\eta\|_v \leq C\|\vartheta - \eta\|$

for all $\vartheta, \eta \in \mathbf{K}$. On the other hand, given Condition 2, it is enough to assume in Condition 4 that g belongs to a total subset of $\mathbf{C}$. However, we prefer to formulate Condition 4 as above because in this form it allows us to prove the uniform local asymptotic normality property. Condition 2 is more technical.

Example: Let $\Theta = (\varepsilon, r)$, $0 < \varepsilon < r$, and $a_\vartheta(du) = b\mathbf{1}_{[-\vartheta,0]}(u)\,du$, where $b \in (-\pi^2/(2r^2), 0)$ is a (known) constant. Then Conditions 1–5 are satisfied. Note that the family (a_ϑ) is not differentiable in ϑ with respect to the total variation norm.

Let now P_T^ϑ be the distribution on $(\mathbf{C}_T, \mathcal{B}(\mathbf{C}_T))$ of a stationary solution $X = (X(t),\, t \in [-r, T])$ to (1.1) with $a = a_\vartheta$. Let also $K_\vartheta(t) = E_T^\vartheta X(s)X(s+t)$ be the covariance function of this solution. Define a $k \times k$-matrix $\Sigma(\vartheta) = (\Sigma_{ij}(\vartheta))_{i,j=1,\ldots,k}$ by

$$\Sigma_{ij}(\vartheta) = \iint\limits_{J\times J} K_\vartheta(u-v)\dot{a}_{\vartheta,i}(du)\dot{a}_{\vartheta,j}(dv). \tag{2.1}$$

Remark 2.2: It will be shown in Lemma 3.4 below that $\Sigma(\vartheta)$ is nondegenerate for any $\vartheta \in \Theta$ and is continuous in ϑ on Θ.

Put

$$\varphi_T(\vartheta) := T^{-1/2}\Sigma^{-1/2}(\vartheta).$$

Define $w_\vartheta(t) := x(t) - x(0) - \int\limits_0^t \int\limits_J x(s+v)\, a_\vartheta(dv)\, ds$, $\vartheta \in \Theta$. The process $(w_\vartheta(t),\, t \in [0,T])$ is a Wiener process with respect to P_T^ϑ. Therefore, one may define

$$\Delta_{T,\vartheta} := \varphi_T(\vartheta) \int\limits_0^T \int\limits_J x(t+s)\, \dot{a}_\vartheta(ds)\, dw_\vartheta(t).$$

Put $U_{T,\vartheta} := \varphi_T^{-1}(\vartheta)(\Theta - \vartheta)$, $\overline{U}_{T,\vartheta} := \varphi_T^{-1}(\vartheta)(\overline{\Theta} - \vartheta)$.

Theorem 2.1 *Let Conditions 1–5 be satisfied. Then*

(1) $P_T^\vartheta \sim P_T^\eta$ *for any* $T \in \mathbb{R}_+$ *and* $\vartheta, \eta \in \overline{\Theta}$;

(2) *the normalized likelihood functions* $Z_{T,\vartheta}(u) := \frac{dP_T^{\vartheta+\varphi_T(\vartheta)u}}{dP_T^\vartheta}$, $u \in \overline{U}_{T,\vartheta}$, *have continuous modifications in* u *for any* $T > 0$ *and* $\vartheta \in \Theta$;

(3) *for any compact set* $\mathbf{K} \subset \Theta$ *and arbitrary sequences* $T_n \to \infty$, $\vartheta_n \in \mathbf{K}$, $u_n \in U_{T_n,\vartheta_n}$, $u_n \to u$,

$$\log Z_{T_n,\vartheta_n}(u_n) - \left(\langle \Delta_{T_n,\vartheta_n}, u\rangle - \frac{\|u\|^2}{2}\right)$$

converges in $P_{T_n}^{\vartheta_n}$*-probability to zero and*

$$\mathcal{L}(\Delta_{T_n,\vartheta_n} \mid P_{T_n}^{\vartheta_n}) \Longrightarrow N(0, I_k)$$

as $n \to \infty$;

(4) *for any* $T_0 > 0$, *any compact set* $\mathbf{K} \subset \Theta$ *and any* $m \geq 2$ *there is a constant* $B = B(T_0, \mathbf{K}, m)$ *such that*

$$\sup_{\vartheta \in \mathbf{K}} \sup_{u,v \in U_{T,\vartheta}} \rho_m^m(P_T^{\vartheta+\varphi_T(\vartheta)u}, P_T^{\vartheta+\varphi_T(\vartheta)v}) \leq B\|u-v\|^m \qquad \text{for any} \quad T \geq T_0; \quad (2.2)$$

(5) *for any compact set* $\mathbf{K} \subset \Theta$ *there is a constant* $C = C(\mathbf{K})$ *such that for any* $T > 0$, $\vartheta \in \mathbf{K}$, $u \in U_{T,\vartheta}$,

$$H(P_T^\vartheta, P_T^{\vartheta+\varphi_T(\vartheta)u}) \leq \exp(-C\|u\|^2). \tag{2.3}$$

Theorem 2.1 (together with Remark 2.2) states that all the conditions of Theorems 5.1 and 5.2 in Chapter I, Theorems 1.1, 2.1 and Corollary 1.1 in Chapter III of Ibragimov and Has'minskii (1981) are satisfied. As a corollary, we obtain the next result. Define the maximum likelihood estimator $\widehat{\vartheta}_T(x)$, $x \in \mathbf{C}_T$, as any measurable solution of the equation

$$\frac{dP_T^{\widehat{\vartheta}_T(x)}}{dP_T^\vartheta}(x) = \sup_{\eta \in \overline{\Theta}} \frac{dP_T^\eta}{dP_T^\vartheta}(x),$$

where $\vartheta \in \Theta$ is some fixed value, and a continuous modification of the likelihood ratio random field is considered.

Theorem 2.2 *Let Conditions* 1–5 *be satisfied and let* $\mathbf{K}$ *be an arbitrary compact in* Θ. *Then uniformly in* $\vartheta \in \mathbf{K}$

(1)

$$\mathcal{L}(T^{1/2}(\widehat{\vartheta}_T - \vartheta) \mid P_T^\vartheta) \Longrightarrow N(0, \Sigma^{-1}(\vartheta)), \quad T \to \infty;$$

(2) *all the moments of* $T^{1/2}(\widehat{\vartheta}_T - \vartheta)$ *under* P_T^ϑ *converge as* $T \to \infty$ *to the corresponding moments of the normal distribution with parameters* $(0, \Sigma^{-1}(\vartheta))$;

(3) *there are positive constants* b_0 *and* B_0 *such that for* T *large enough*

$$\sup_{\vartheta \in \mathbf{K}} P_T^\vartheta\{T^{1/2}\|\widehat{\vartheta}_T - \vartheta\| > R\} \leq B_0 \exp(-b_0 R), \quad R > 0;$$

(4) *the maximum likelihood estimator* $\widehat{\vartheta}_T$ *is asymptotically efficient in* $\mathbf{K}$.

The efficiency in (4) is to be understood in the sense of Corollary 1.1, p. 177, in Chapter III of Ibragimov and Has'minskii (1981).
The same statement is true for the Bayesian estimators $\widetilde{\vartheta}_T$ corresponding to a continuous and positive prior density and to loss functions of the form $l_T(u) = l(T^{1/2}x)$, where l satisfies the assumptions of Theorem 2.1, p. 179, in Chapter III of Ibragimov and Has'minskii (1981) (in particular, one can take $l(u) = \|u\|^p$, $p > 0$).

3 Affine delay differential equations

3.1 Deterministic equations

Since the equation (1.1) involves no stochastic integrals and can be treated pathwise, we will formulate a number of results for solutions of the deterministic equation

$$\begin{aligned} X(t) &= X(0) + \int_0^t \int_J X(s+u)\, a(du)\, ds + F(t), && t \geq 0, \\ X(t) &= X_0(t), && t \in J, \end{aligned} \tag{3.1}$$

$F \in \mathbf{C}(\mathbb{R}_+)$, $F(0) = 0$, $X_0 \in \mathbf{C}$, for which we refer to Myschkis (1972), Hale and Verduyn Lunel (1993), Diekmann *et al.* (1995), and also to Mohammed and Scheutzow (1990) and Gushchin and Küchler (2000).
The equation (3.1) has a unique solution X in the class $\mathbf{C}([-r, \infty))$.
If $F \equiv 0$ and

$$X_0(t) = \begin{cases} 1, & t = 0, \\ 0, & -r \leq t < 0, \end{cases}$$

then (3.1) can also be solved uniquely, and the solution of (3.1) is called the *fundamental solution* (corresponding to a) and denoted by x_0. In other words, a function $x_0(t)$, $t \geq -r$, is the fundamental solution if it is absolutely continuous on $\mathbb{R}_+$, $x_0(t) = 0$ for $t \in [-r, 0)$, $x_0(0) = 1$, and

$$\dot{x}_0(t) = \int_J x_0(t+u)\, a(du) \tag{3.2}$$

for Lebesgue almost all $t > 0$. To facilitate some notation in the sequel it is convenient to put $x_0(t) = 0$ for $t < -r$.
The solution of (3.1) can be represented via the fundamental solution x_0 by the formula

$$X(t) = \begin{cases} x_0(t)X_0(0) + \int_J \int_u^0 X_0(s)x_0(t+u-s)\, ds\, a(du) + \int_{[0,t]} F(t-s)\, dx_0(s), & t \geq 0, \\ X_0(t), & t \in J, \end{cases} \tag{3.3}$$

where the domain of integration in the last integral in (3.3) includes zero:

$$\int_{[0,t]} F(t-s)\, dx_0(s) = F(t) + \int_{]0,t]} F(t-s)\, dx_0(s).$$

The asymptotic behaviour of solutions of the equation (3.1) for $t \to \infty$ is connected with the set of complex solutions of the so-called *characteristic equation*

$$h(\lambda) := \lambda - \int_J e^{\lambda u} a(du) = 0. \tag{3.4}$$

Note that a complex number λ solves (3.4) if and only if $(e^{\lambda t},\ t \geq -r)$ solves (3.1) for $F \equiv 0$ and $X_0(t) = e^{\lambda t}$, $t \in J$.
The set $\Lambda := \{\lambda \in \mathbb{C}: h(\lambda) = 0\}$ is not empty; moreover, it is infinite except the case where a is concentrated at 0. The set $\{\lambda \in \Lambda: \operatorname{Re}\lambda \geq c\}$ is finite for every $c \in \mathbb{R}$. In particular, it holds

$$v_0 := \max\{\operatorname{Re}\lambda: \lambda \in \Lambda\} < \infty. \tag{3.5}$$

For any $\gamma > v_0$ there is a constant L such that

$$|x_0(t)| \leq Le^{\gamma t}, \quad t \in \mathbb{R}_+, \tag{3.6}$$

and the Laplace transform of x_0 is given by

$$\int_{\mathbb{R}} e^{-\lambda t} x_0(t)\, dt = \frac{1}{h(\lambda)}, \quad \operatorname{Re}\lambda > v_0. \tag{3.7}$$

3.2 Stationary solutions

In this subsection we assume that X is a stationary solution of the equation (1.1), in particular, $a \in M_s$. According to Theorem 3.1 in Gushchin and Küchler (2000), $a \in M_s$ if and only if

$$v_0 < 0, \tag{3.8}$$

and then X is a Gaussian process with zero mean and covariance

$$K(t) = EX(u)X(t+u) = \int_{\mathbb{R}} x_0(s)x_0(s+t)\, ds, \quad t \in \mathbb{R}, \tag{3.9}$$

(recall that $x_0(s) = 0$ if $s < 0$ and $x_0(s)$ is exponentially small as $s \to +\infty$ due to (3.6) and (3.8)). Obviously, $K(-t) = K(t)$, $t \in \mathbb{R}$.
Since the Fourier transforms of $x_0(\cdot)$ and $x_0(t + \cdot)$ are known, see (3.7), we can apply Parseval's identity to (3.9) to obtain

$$K(t) = \int_{\mathbb{R}} e^{i\lambda t} f(\lambda)\, d\lambda, \quad t \in \mathbb{R}, \tag{3.10}$$

where

$$f(\lambda) = \frac{1}{2\pi |h(i\lambda)|^2}, \quad \lambda \in \mathbb{R}, \tag{3.11}$$

is the spectral density of X. The function f is strictly positive and continuous. By the definition of h

$$|\lambda| - \|a\|_v \leq |h(i\lambda)| \leq |\lambda| + \|a\|_v, \quad \lambda \in \mathbb{R}. \tag{3.12}$$

Hence there are positive constants B_* and B^* (depending on a) such that

$$\frac{B_*}{1+\lambda^2} \leq f(\lambda) \leq \frac{B^*}{1+\lambda^2}, \quad \lambda \in \mathbb{R}. \tag{3.13}$$

Lemma 3.1 *The covariance function K of the stationary solution of* (1.1) *has the following properties*:

(i) *K is differentiable on $(-\infty,0)\cup(0,\infty)$ and has one-sided derivatives at* 0 :

$$\dot{K}(t)=\int_J K(t+u)\,a(du),\quad t>0, \tag{3.14}$$

$$\dot{K}(\pm 0)=\mp 1/2,$$
$$\dot{K}(t)=-\dot{K}(-t),\quad t<0;$$

(ii) *Put $\dot{K}(0)=-1/2$. Then the function $\dot{K}$ is absolutely continuous on $\mathbb{R}_+$ and*

$$\ddot{K}(t)=\int_J \dot{K}(t+u)\,a(du)\quad \textit{a.e. for Lebesgue measure on } (0,\infty). \tag{3.15}$$

Proof: Since $K(t)=K(-t)$, it is sufficient to prove the first statement of the lemma for $t\in\mathbb{R}_+$. Let $\gamma\in(v_0,0)$. It follows from (3.6) and (3.2) that there is a constant c such that $|x_0(s)|\le ce^{\gamma s}$ and $|x_0(t+s+h)-x_0(t+s)|\le ce^{\gamma s}|h|$ for $s\ge 0$, $t\ge 0$, $t+h\ge 0$. Thus by the Lebesgue dominated convergence theorem we can differentiate under the integral sign in (3.9). This yields that K is differentiable in the required sense and

$$\dot{K}(t)=\int_{\mathbb{R}} x_0(s)\dot{x}_0(s+t)\,ds,\quad t\in\mathbb{R}_+$$

(where the right-hand side derivative is taken if $t=0$). Inserting (3.2) into the last equality, changing the order of integration and using (3.9) we get (3.14). In particular,

$$\dot{K}(0+)=\int_0^\infty x_0(s)\dot{x}_0(s)\,ds=-\frac{1}{2}x_0^2(0)=-\frac{1}{2}.$$

It follows from the first part of the lemma and from (3.14) that K is continuously differentiable on $(-\infty,0)\cup(0,\infty)$. Therefore, it is absolutely continuous on $\mathbb{R}$ and we obtain, for $t>0$,

$$\begin{aligned}\dot{K}(t) &= \int_J K(t+u)\,a(du)=\dot{K}(0+)+\int_J\{K(t+u)-K(u)\}\,a(du)\\ &= \dot{K}(0+)+\int_J\int_0^t \dot{K}(s+u)\,ds\,a(du)=\dot{K}(0+)+\int_0^t\int_J \dot{K}(s+u)\,a(du)\,ds.\end{aligned}$$

The claim follows.

Now let $\widetilde{a}$ be another measure from M_s. Denote by $\widetilde{X}$ the corresponding stationary solution to (1.1). Let $\widetilde{K}$ and $\widetilde{f}$ be the covariance function and the spectral density of $\widetilde{X}$ respectively. Put $R(t)=K(t)-\widetilde{K}(t)$.

By Lemma 3.1 the function R is differentiable on $\mathbb{R}$ and its derivative $\dot{R}$ is absolutely continuous on $\mathbb{R}$. Moreover, it follows from (3.15) and (3.14) that $\ddot{R}$ is bounded a.e. for Lebesgue measure. It is known, see e.g. Theorem 13 in (Ibragimov and Rozanov, 1978, Chapter III, p. 99), that these properties together with the inequality (3.13) imply the following statement. Denote by P_T^a and $P_T^{\widetilde{a}}$ the distributions of the processes $X(t)$ and $\widetilde{X}(t)$, $t \in [-r, T]$, respectively in the space $\mathbf{C}_T$.

Corollary 3.2 *The measures P_T^a and $P_T^{\widetilde{a}}$ are equivalent for every $T \in \mathbb{R}_+$.*

Following the idea of the proof of the above mentioned theorem one can obtain even an estimate for closedness of P_T^a and $P_T^{\widetilde{a}}$. The proof is based on the results in Section 4.

Theorem 3.3 *For every $m \geq 2$ there is a constant C_m (depending only on m) such that*

$$\rho_m^m(P_T^a, P_T^{\widetilde{a}}) \leq C_m B_*^{-m} B^{*2m} (1 + r + T)^{m/2} (1 + \|a\|_v^m + \|\widetilde{a}\|_v^m) \|\widetilde{a} - a\|_v^m,$$

where the constants B_* and B^* are such that

$$\frac{B_*}{1+\lambda^2} \leq f(\lambda) \leq \frac{B^*}{1+\lambda^2}, \quad \widetilde{f}(\lambda) \leq \frac{B^*}{1+\lambda^2}, \qquad \lambda \in \mathbb{R}.$$

Proof: It follows from (3.11) and (3.4) that

$$\begin{aligned} |f(\lambda) - \widetilde{f}(\lambda)| &= 2\pi \Big| |h(i\lambda)|^2 - |\widetilde{h}(i\lambda)|^2 \Big| f(\lambda) \widetilde{f}(\lambda) \\ &\leq 4\pi B^{*2} \|\widetilde{a} - a\|_v (|\lambda| + \|a\|_v + \|\widetilde{a}\|_v)(1 + \lambda^2)^{-2}. \end{aligned}$$

Now the claim follows from Corollaries 4.4 and 4.6.

3.3 First properties of the parametric model

Assume that $\mathcal{A} = \{a_\vartheta\}_{\vartheta \in \Theta}$ is a parametric family satisfying Conditions 1–5. Let $x_0^\vartheta(t)$ be the fundamental solution corresponding to a_ϑ, $h_\vartheta(\lambda) = \lambda - \int_J e^{\lambda u} a_\vartheta(du)$, $v_0(\vartheta) = \max\{\operatorname{Re}\lambda\colon h_\vartheta(\lambda) = 0\}$, $K_\vartheta(t) = \int_{\mathbb{R}} x_0^\vartheta(s) x_0^\vartheta(s+t)\, ds$, $f_\vartheta(\lambda) = \frac{1}{2\pi |h_\vartheta(i\lambda)|^2}$. The matrix $\Sigma(\vartheta)$ is defined in (2.1).

We shall need also the dual Lipschitz norm $\|\cdot\|_D$ on the space M. Let Lip_1 be the class of all real-valued functions g on J satisfying $|g(x)| \leq 1$ and $|g(x) - g(y)| \leq |x - y|$ for all $x, y \in J$. Now, for $a \in M$, define

$$\|a\|_D = \sup_{g \in \mathrm{Lip}_1} \int_J g(u)\, a(du). \tag{3.16}$$

In the case $a(J) = 0$ this norm coincides with the Kantorovich–Rubinstein norm of a, see Kantorovich and Akilov (1982). Note that the normed linear space $(M, \|\cdot\|_D)$ is not complete if $r > 0$.

Lemma 3.4 *Let Conditions* 1–5 *be satisfied. There are constants* $0 < B_* \le B^* < \infty$, $\delta > 0$, $L < \infty$, *such that*

$$\frac{B_*}{1+\lambda^2} \le f_\vartheta(\lambda) \le \frac{B^*}{1+\lambda^2} \quad \textit{for any} \quad \vartheta \in \overline{\Theta},\ \lambda \in \mathbb{R}, \tag{3.17}$$

$$v_0(\vartheta) < -\delta \quad \textit{for any} \quad \vartheta \in \overline{\Theta}, \tag{3.18}$$

$$|x_0^\vartheta(t)| \le Le^{-\delta t} \quad \textit{for any} \quad \vartheta \in \overline{\Theta},\ t \in \mathbb{R}_+. \tag{3.19}$$

For every compact $\mathbf{K} \subset \Theta$ *there is a constant* $d = d(\mathbf{K}) > 0$ *such that*

$$\|a_\vartheta - a_\eta\|_D \ge d\|\vartheta - \eta\| \quad \textit{for any} \quad \vartheta \in \mathbf{K},\ \eta \in \Theta. \tag{3.20}$$

Moreover, $\Sigma(\vartheta)$ *is nondegenerate for every* $\vartheta \in \Theta$ *and is continuous in* ϑ *on* Θ.

Proof: The existence of a constant $B_* > 0$ satisfying (3.17) follows easily from the right-hand inequality in (3.12) and from the inequality

$$\sup_{\vartheta \in \overline{\Theta}} \|a_\vartheta\|_v < \infty, \tag{3.21}$$

see Condition 2. To prove the existence of a finite constant B^* satisfying (3.17), we use the left-hand inequality in (3.12), but in that case we have to show additionally that

$$\inf_{\vartheta \in \overline{\Theta}} \inf_{\lambda \in \mathbb{R}} |h_\vartheta(i\lambda)| > 0.$$

Assume the converse. Then there are sequences $\{\vartheta_n\}$, $\vartheta_n \in \overline{\Theta}$, and $\{\lambda_n\}$, $\lambda_n \in \mathbb{R}$, such that $\vartheta_n \to \vartheta \in \overline{\Theta}$ and $h_\vartheta(i\lambda_n) \to 0$. In view of (3.12) and (3.21), the sequence $\{\lambda_n\}$ is bounded. Extracting a convergent subsequence, we may assume that $\lambda_n \to \lambda$. But it is easy to see that $0 = \lim_n h_{\vartheta_n}(i\lambda_n) = h_\vartheta(i\lambda)$, which contradicts Condition 3.
The existence of a $\delta > 0$ satisfying (3.18) is proved similarly.
Now let us show (3.19). The uniform boundedness of $x_0^\vartheta(t)$ on any finite interval, say $t \in [0,1]$, follows from the Gronwall lemma applied to $\bar{x}^\vartheta(t) = \sup_{0 \le s \le t} |x_0^\vartheta(s)|$ and from (3.21). Next, we use the formula (3.7) for the Laplace transform of $x_0^\vartheta(t)$. Using the inversion formula and integrating by parts we get

$$x_0^\vartheta(t) = \frac{1}{2\pi i} \lim_{w \to \infty} \int_{-\delta - iw}^{-\delta + iw} \frac{e^{\mu t}}{h_\vartheta(\mu)} d\mu = \frac{1}{2\pi i t} \lim_{w \to \infty} \int_{-\delta - iw}^{-\delta + iw} \frac{h'_\vartheta(\mu) e^{\mu t}}{[h_\vartheta(\mu)]^2} d\mu.$$

Hence, if $t \ge 1$,

$$|x_0^\vartheta(t)| \le \frac{e^{-\delta t}(1 + r\|a_\vartheta\|_v)}{2\pi} \int_{-\infty}^{\infty} \frac{d\lambda}{|h_\vartheta(-\delta + i\lambda)|^2},$$

where the last integral is bounded uniformly in $\vartheta \in \overline{\Theta}$ due to the estimate

$$\sup_{\vartheta \in \overline{\Theta}} \sup_{\lambda \in \mathbb{R}} \frac{1 + \lambda^2}{|h_\vartheta(-\delta + i\lambda)|^2} < \infty,$$

which is proved similarly to the right-hand inequality in (3.17).
If (3.20) does not hold, then there exist two sequences $\{\vartheta_n\}$, $\vartheta_n \in \mathbf{K}$, and $\{\eta_n\}$, $\eta_n \in \Theta$, $\vartheta_n \neq \eta_n$, such that $\vartheta_n \to \vartheta \in \mathbf{K}$, $\eta_n \to \eta \in \overline{\Theta}$, and

$$\frac{\|a_{\vartheta_n} - a_{\eta_n}\|_D}{\|\vartheta_n - \eta_n\|} \to 0, \quad n \to \infty. \tag{3.22}$$

Since $\|\cdot\|_D \le \|\cdot\|_v$, the distance $\|a_\vartheta - a_\eta\|_D$ is a continuous function of ϑ and η in view of Condition 2. Since $\|a_\vartheta - a_\eta\|_D > 0$ if $\vartheta \neq \eta$ due to Condition 5, we obtain $\vartheta = \eta$. Extracting a subsequence if necessary, we may assume that $\frac{\vartheta_n - \eta_n}{\|\vartheta_n - \eta_n\|} \to \zeta$, say. Since the segment containing ϑ_n and η_n lies in Θ for n large enough, it follows from Condition 4 that

$$\int_J g(u) \frac{(a_{\vartheta_n} - a_{\eta_n})(du)}{\|\vartheta_n - \eta_n\|} \longrightarrow \int_J g(u) \langle \dot a_\vartheta, \zeta \rangle (du)$$

for any $g \in \mathbf{C}$. Since $\dot a_\vartheta \in M^k_\#$, the measure $\langle \dot a_\vartheta, \zeta\rangle$ does not equal to zero identically. Hence there is a function $g_0 \in \mathrm{Lip}_1$ such that $\int_J g_0(u) \langle \dot a_\vartheta, \zeta\rangle(du) > 0$, which implies

$$\liminf_{n\to\infty} \frac{\|a_{\vartheta_n} - a_{\eta_n}\|_D}{\|\vartheta_n - \eta_n\|} = \liminf_{n\to\infty} \sup_{g \in \mathrm{Lip}_1} \int_J g(u) \frac{(a_{\vartheta_n} - a_{\eta_n})(du)}{\|\vartheta_n - \eta_n\|} \ge \int_J g_0(u) \langle \dot a_\vartheta, \zeta \rangle (du) > 0,$$

which contradicts (3.22).
If the matrix $\Sigma(\vartheta)$ is degenerate then there is a non-zero vector $\zeta \in \mathbb{R}^k$ such that $\langle \zeta, \Sigma(\vartheta)\zeta\rangle = 0$. Then, denoting $b_\vartheta = \langle \zeta, \dot a_\vartheta \rangle$, we have

$$0 = \iint_{J\times J} K_\vartheta(u-v) b_\vartheta(du) b_\vartheta(dv) = \int_{\mathbb{R}} \Big| \int_J e^{i\lambda u}\, b_\vartheta(du) \Big|^2 f_\vartheta(\lambda)\, d\lambda$$

in view of (2.1) and (3.10). Since $f_\vartheta(\lambda)$ is strictly positive, we obtain that the Fourier transform of b_ϑ is equal to zero identically, hence $b_\vartheta = 0$, which contradicts the condition $\dot a_\vartheta \in M^k_\#$.
Similarly, we obtain from (2.1) and (3.10) that

$$\Sigma_{ij}(\vartheta) = \int_{\mathbb{R}} \Big(\int_J e^{i\lambda u}\, \dot a_{\vartheta,i}(du) \Big) \Big(\int_J e^{-i\lambda v}\, \dot a_{\vartheta,j}(dv) \Big) f_\vartheta(\lambda)\, d\lambda.$$

The integrals over J in the last expression and the function $f_\vartheta(\lambda)$ are continuous in ϑ for a fixed λ due to Condition 4, (3.11) and the definition of $h_\vartheta(\lambda)$. Moreover, we can use the dominated convergence theorem taking into account (3.17) and the fact that the norms $\|\dot a_{\vartheta,i}\|_v$, $i = 1, \ldots, k$, $\vartheta \in \Theta$, are uniformly bounded by Conditions 2 and 4. This proves the continuity of $\Sigma(\vartheta)$.

4 Some inequalities for Gaussian measures

Let P and $\widetilde{P}$ be probability measures on a measurable space $(\Omega, \mathcal{F})$. Assume $X = (X(t),\, t \in I)$, where I is an arbitrary index set, is a Gaussian process on $(\Omega, \mathcal{F})$ with

respect to both P and $\widetilde{P}$ and $EX(t) = \widetilde{E}X(t) = 0$, $t \in I$ (E and $\widetilde{E}$ are expectations relative to P and $\widetilde{P}$ respectively). Suppose also that $\mathcal{F} = \sigma\{X(t),\, t \in I\}$.
These assumptions will be satisfied throughout this section.
It is well known that P and $\widetilde{P}$ are either equivalent or singular. The aim of this section is to provide upper bounds for the Hellinger distances $\rho_m(P, \widetilde{P})$ in terms of covariances of X with respect to P and $\widetilde{P}$. These bounds fit well if P and $\widetilde{P}$ are close to each other. The main result is Theorem 4.3 in subsection 4.4. In subsection 4.5 we specify these bounds in a special case of stationary Gaussian processes. To estimate $\rho_2(P, \widetilde{P})$ we use arguments that are very similar to those used to prove conditions for equivalence of Gaussian measures, see e.g. Shepp (1966), Rozanov (1971), Ibragimov and Rozanov (1978), Gihman and Skorohod (1980), Kuo (1975).
In the next two subsections $N(0, 1+\beta)$, $\beta > -1$, stands for the normal distribution on $\mathbb{R}$ with mean zero and variance $1+\beta$. Its density is denoted by $\varphi_\beta(x)$; $\varphi(x) := \varphi_0(x)$ is the standard normal density.

4.1 One-dimensional Gaussian distributions

We start with a preliminary case of one-dimensional distributions.

Lemma 4.1 *Let $P = N(0,1)$ and $\widetilde{P} = N(0, 1+\beta)$, $\beta > -1$. Then*

$$H(P, \widetilde{P}) = \left(1 + \frac{\beta^2}{4(1+\beta)}\right)^{-1/4}, \qquad J(P, \widetilde{P}) = \frac{\beta^2}{2(1+\beta)}. \tag{4.1}$$

For every $m \geq 1$ there is a constant C_m (depending only on m) such that

$$\rho_m^m(P, \widetilde{P}) \leq C_m |\beta|^m. \tag{4.2}$$

Proof: It is easy to check (4.1) by computation.
Let $f(\beta; x)$ be the density of $\widetilde{P}(dx)$ with respect to $P(dx)$, i.e.

$$f(\beta; x) = \varphi_\beta(x)/\varphi(x) = (1+\beta)^{-1/2} e^{\frac{\beta x^2}{2(1+\beta)}}.$$

Differentiating $f(\beta; x)$ with respect to β, one obtains

$$\frac{\partial f(\beta; x)}{\partial \beta} = f(\beta; x) g(\beta; x), \qquad \text{where} \quad g(\beta; x) = \frac{1}{2}\left(\frac{x^2}{(1+\beta)^2} - \frac{1}{1+\beta}\right).$$

By Newton–Leibniz formula,

$$f(\beta; x)^{1/m} = 1 + \frac{\beta}{m} \int_0^1 f(\beta t; x)^{1/m} g(\beta t; x)\, dt.$$

Hence

$$\begin{aligned}\rho_m^m(P,\widetilde{P}) &= \int_{\mathbb{R}} |f(\beta;x)^{1/m}-1|^m\varphi(x)\,dx \le \frac{|\beta|^m}{m^m}\int_{\mathbb{R}}\int_0^1 f(\beta t;x)|g(\beta t;x)|^m\varphi(x)\,dt\,dx\\ &= \frac{|\beta|^m}{(2m)^m}\int_0^1\int_{\mathbb{R}}\Big|\frac{x^2}{(1+\beta t)^2}-\frac{1}{1+\beta t}\Big|^m\varphi_{\beta t}(x)\,dx\,dt\\ &= \frac{|\beta|^m}{(2m)^m}\int_0^1\int_{\mathbb{R}}\Big|\frac{x^2}{1+\beta t}-\frac{1}{1+\beta t}\Big|^m\varphi(x)\,dx\,dt\\ &= \frac{|\beta|^m}{(2m)^m}\int_{\mathbb{R}}|x^2-1|^m\varphi(x)\,dx\int_0^1\frac{dt}{(1+\beta t)^m}.\end{aligned}$$

Fix a number $\beta_0\in(-1,0)$. Since the integral $\int_0^1\frac{dt}{(1+\beta t)^m}$ decreases in β and $\rho_m^m(P,\widetilde{P})\le 2$, we obtain (4.2) with

$$C_m=\max\Big\{(2m)^{-m}\int_{\mathbb{R}}|x^2-1|^m\varphi(x)\,dx\int_0^1\frac{dt}{(1+\beta_0 t)^m},2|\beta_0|^{-m}\Big\}.$$

4.2 Independent Gaussian variables

In this subsection we consider the case where P and $\widetilde{P}$ are the distributions of sequences of independent Gaussian variables. Namely, we assume that $\Omega=\mathbb{R}^\infty$ with the product σ-algebra, $P=\mu_1\times\mu_2\times\cdots$, $\widetilde{P}=\widetilde{\mu}_1\times\widetilde{\mu}_2\times\cdots$, where

$$\mu_n=N(0,1)\qquad\text{and}\qquad\widetilde{\mu}_n=N(0,1+\beta_n),\quad \beta_n>-1.$$

Lemma 4.2 *Let P and $\widetilde{P}$ be as described above.*

(a) *Assume that $\sum_{n=1}^\infty\beta_n^2<\infty$. Then $P\sim\widetilde{P}$, $J(P,\widetilde{P})<\infty$, and*

$$\frac{\sum_{n=1}^\infty\beta_n^2}{1+\sup\limits_{n:\,\beta_n\neq 0}\beta_n}\le 2J(P,\widetilde{P})\le\frac{\sum_{n=1}^\infty\beta_n^2}{1+\inf\limits_{n:\,\beta_n\neq 0}\beta_n}. \tag{4.3}$$

(b)

$$\exp\Big(-\frac{1}{8}J(P,\widetilde{P})\Big)\le H(P,\widetilde{P})\le\Big(1+\frac{1}{2}J(P,\widetilde{P})\Big)^{-\frac{1}{4}}. \tag{4.4}$$

(c) *For every $m\ge 2$ there is a constant C_m (depending only on m) such that*

$$\rho_m^m(P,\widetilde{P})\le C_m\rho_2^m(P,\widetilde{P}). \tag{4.5}$$

Proof: Let $\mathcal{F}_n$ be the sub-σ-algebra of $\mathcal{F}$ generated by the first n coordinates and P_n and $\widetilde{P}_n$ the restrictions of P and $\widetilde{P}$ respectively to $\mathcal{F}_n$. Then

$$H(P_n,\widetilde{P}_n)=\prod_{k=1}^{n}H(\mu_k,\widetilde{\mu}_k),\qquad J(P_n,\widetilde{P}_n)=\sum_{k=1}^{n}J(\mu_k,\widetilde{\mu}_k) \tag{4.6}$$

and

$$H(P,\widetilde{P})=\lim_{n\to\infty}H(P_n,\widetilde{P}_n),\qquad J(P,\widetilde{P})=\lim_{n\to\infty}J(P_n,\widetilde{P}_n) \tag{4.7}$$

see e.g. Liese and Vajda (1987). By (4.1) and (4.6),

$$H(P_n,\widetilde{P}_n)=\prod_{k=1}^{n}\left(1+\frac{\beta_k^2}{4(1+\beta_k)}\right)^{-1/4},\qquad J(P_n,\widetilde{P}_n)=\frac{1}{2}\sum_{k=1}^{n}\frac{\beta_k^2}{1+\beta_k}. \tag{4.8}$$

Now it is clear that the convergence of $\sum_{n=1}^{\infty}\beta_n^2$ implies $J(P,\widetilde{P})<\infty$, hence $P\sim\widetilde{P}$, and (4.3) follows.
To show (4.4), we use (4.8) and the inequality

$$1+\sum_{k=1}^{n}x_k\le\prod_{k=1}^{n}(1+x_k)\le e^{\sum_{k=1}^{n}x_k},\quad x_k\ge 0,\quad k=1,\dots,n,$$

which yields

$$1+\frac{1}{2}J(P_n,\widetilde{P}_n)\le H^{-4}(P_n,\widetilde{P}_n)\le\exp\Bigl(\frac{1}{2}J(P_n,\widetilde{P}_n)\Bigr).$$

Passing to the limit as $n\to\infty$ with the help of (4.7), we obtain (4.4).
Let us prove (c). In the rest of the proof the letter C_m will denote positive constants depending only on m; they need not be the same in different expressions.
Since $\rho_m^m(P,\widetilde{P})\le 2$, inequality (4.5) holds with $C_m=2$ if $\rho_2^2(P,\widetilde{P})\ge 1$. Thus, we will assume that $\rho_2^2(P,\widetilde{P})\le 1$ or, equivalently, $H(P,\widetilde{P})\ge 1/2$. Under this assumption, (4.4) implies

$$J(P,\widetilde{P})\le 2\left(H^{-4}(P,\widetilde{P})-1\right)\le 60(1-H(P,\widetilde{P}))=30\rho_2^2(P,\widetilde{P}). \tag{4.9}$$

The key point of the proof of this part is the inequality proved by Dzhaparidze and Valkeila (1990, Corollary 3.1). In our context it states that

$$\rho_m^m(P,\widetilde{P})\le C_m\left\{\Bigl(\frac{1}{2}\sum_{n=1}^{\infty}\rho_2^2(\mu_n,\widetilde{\mu}_n)\Bigr)^{m/2}+\sum_{n=1}^{\infty}\rho_m^m(\mu_n,\widetilde{\mu}_n)\right\}. \tag{4.10}$$

Let us estimate the sums in (4.10). Since $\frac{1}{2}\rho_2^2(\mu_n,\widetilde{\mu}_n)=1-H(\mu_n,\widetilde{\mu}_n)$, we get from (4.1), (4.9) and the inequality $1-(1+x)^{-1/4}\le x/4$ that

$$\begin{aligned}\sum_{n=1}^{\infty}\rho_2^2(\mu_n,\widetilde{\mu}_n) &= 2\sum_{n=1}^{\infty}\left\{1-\left(1+\frac{\beta_n^2}{4(1+\beta_n)}\right)^{-1/4}\right\}\le\frac{1}{8}\sum_{n=1}^{\infty}\frac{\beta_n^2}{1+\beta_n}\\ &= \frac{1}{4}J(P,\widetilde{P})\le\frac{15}{2}\rho_2^2(P,\widetilde{P}).\end{aligned} \tag{4.11}$$

Since $H(\mu_n, \widetilde{\mu}_n) \geq H(P, \widetilde{P}) \geq 1/2$, we get from (4.1) that $\beta_n \leq \bar{\beta} := 30 + 8\sqrt{15}$ for every $n \geq 1$. Using (4.2) and (4.9), we obtain

$$\sum_{n=1}^{\infty} \rho_m^m(\mu_n, \widetilde{\mu}_n) \leq C_m \sum_{n=1}^{\infty} |\beta_n|^m \leq C_m \Big(\sum_{n=1}^{\infty} \beta_n^2\Big)^{m/2} \leq C_m (1+\bar{\beta})^{m/2} \Big(\sum_{n=1}^{\infty} \frac{\beta_n^2}{1+\beta_n}\Big)^{m/2}$$
$$= C_m 2^{m/2} (1+\bar{\beta})^{m/2} J^{m/2}(P, \widetilde{P}) \leq C_m 60^{m/2} (1+\bar{\beta})^{m/2} \rho_2^m(P, \widetilde{P}). \quad (4.12)$$

Combining (4.10)– (4.12), we arrive at (4.5).

4.3 The Hilbert–Schmidt operator

In this subsection we consider the general setting described at the beginning of the section. We assume here that

$$P \sim \widetilde{P}. \quad (4.13)$$

Let $\mathcal{K}$ be the linear space of random variables ξ on $(\Omega, \mathcal{F})$ that can be represented as finite linear combinations $\xi = \sum_k c_k X(t_k)$, $c_k \in \mathbb{R}$, $t_k \in I$. As usual, we deal with equivalence classes of random variables that are equal P-a.s. and/or $\widetilde{P}$-a.s. but continue to speak of these classes as of random variables. It is known that under (4.13) there are constants $0 < d_1 \leq d_2 < \infty$ such that

$$d_1 E\xi^2 \leq \widetilde{E}\xi^2 \leq d_2 E\xi^2 \quad (4.14)$$

for any $\xi \in \mathcal{K}$, see e.g. Rozanov (1971, Chapter II) or Ibragimov and Rozanov (1978, Chapter III). Let $\mathcal{H}$ (resp. $\widetilde{\mathcal{H}}$) be the closure of $\mathcal{K}$ in $L^2(P)$ (resp. in $L^2(\widetilde{P})$). Then, due to (4.14) and (4.13), $\mathcal{H}$ and $\widetilde{\mathcal{H}}$ consist of the same elements and are equipped with different scalar products $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle^{\sim}$:

$$\langle \xi, \eta \rangle = E\xi\eta, \qquad \langle \xi, \eta \rangle^{\sim} = \widetilde{E}\xi\eta.$$

Of course, if $\xi \in \mathcal{H}$ (or $\widetilde{\mathcal{H}}$) then ξ is a Gaussian random variable with respect to P and $\widetilde{P}$ with $E\xi = \widetilde{E}\xi = 0$, and (4.14) is still valid.
Define a linear bounded operator B from $\mathcal{H}$ to $\widetilde{\mathcal{H}}$ by $B\xi = \xi$. Denote by B^* the adjoint operator of B and put $A = E - B^*B$, where E is the identity operator in $\mathcal{H}$. The operator A is uniquely determined by the relation

$$\begin{aligned} \langle A\xi, \eta \rangle &= \langle \xi, \eta \rangle - \langle B^*B\xi, \eta \rangle = \langle \xi, \eta \rangle - \langle B\xi, B\eta \rangle^{\sim} \\ &= \langle \xi, \eta \rangle - \langle \xi, \eta \rangle^{\sim} = E\xi\eta - \widetilde{E}\xi\eta, \quad \xi, \eta \in \mathcal{H}. \end{aligned} \quad (4.15)$$

B^*B is a positive self-adjoint operator, hence A is a self-adjoint operator, whose eigenvalues lie in the interval $[1-d_2, 1-d_1]$ because of (4.15) and (4.14).
A fundamental fact is that A is a Hilbert–Schmidt operator under (4.13), see Rozanov (1971, Chapter II, Theorem 5); cf. also Ibragimov and Rozanov (1978, Chapter III, Theorem 4, p. 80). In particular, the subspace $(\ker A)^{\perp}$ orthogonal to the kernel of A in $\mathcal{H}$ is separable, and there is a complete (in this subspace) orthonormal sequence consisting of eigenvectors of A.

4.4 General case

In this subsection we extend Lemma 4.2 to the general case of Gaussian measures with zero means. We are in the setting introduced at the beginning of this section. The operator A has been introduced in the previous subsection (under assumption $P \sim \widetilde{P}$). Let s_A be the set of all non-zero eigenvalues of A. According to previous subsection, $-\infty < \inf s_A \leq \sup s_A < 1$ (if $s_A = \emptyset$, we put $\inf s_A = \sup s_A = 0$). Recall that $\|A\|_{(2)}$ is the Hilbert–Schmidt norm of A.

Theorem 4.3 (a) *Assume that $P \sim \widetilde{P}$. Then*

$$\frac{\|A\|_{(2)}^2}{1-\inf s_A} \leq 2J(P,\widetilde{P}) \leq \frac{\|A\|_{(2)}^2}{1-\sup s_A}.$$

(b)

$$\exp\Big(-\frac{1}{8}J(P,\widetilde{P})\Big) \leq H(P,\widetilde{P}) \leq \Big(1+\frac{1}{2}J(P,\widetilde{P})\Big)^{-\frac{1}{4}};$$

(c) *for every $m \geq 2$ there is a constant C_m (depending only on m) such that*

$$\rho_m^m(P,\widetilde{P}) \leq C_m \rho_2^m(P,\widetilde{P}).$$

Corollary 4.4 *Let $P \sim \widetilde{P}$. For every $m \geq 2$ there is a constant C_m (depending only on m) such that*

$$\rho_m^m(P,\widetilde{P}) \leq C_m \|A\|_{(2)}^m.$$

Proof of Corollary 4.4: Since $\sup \sigma_A \leq \|A\| \leq \|A\|_{(2)}$, it follows from parts (a) and (b) of Theorem 4.3 that if $\|A\|_{(2)} \leq 4\sqrt{5}-8 < 1$ then we have (with $C_2 = [8(9-4\sqrt{5})]^{-1}$)

$$\begin{aligned}\rho_2^2(P,\widetilde{P}) &= 2(1-H(P,\widetilde{P})) \leq 2\Big(1-\exp\Big(-\frac{1}{8}J(P,\widetilde{P})\Big)\Big) \\ &\leq \frac{1}{4}J(P,\widetilde{P}) \leq \frac{\|A\|_{(2)}^2}{8(1-\|A\|_{(2)})} \leq C_2\|A\|_{(2)}^2.\end{aligned}$$

But if $\|A\|_{(2)} > 4\sqrt{5}-8$ the inequality $\rho_2^2(P,\widetilde{P}) \leq C_2\|A\|_{(2)}^2$ is still true since its right-hand side is greater than 2. The statement for $m > 2$ follows from part (c) of Theorem 4.3.

Proof of Theorem 4.3: If $P \perp \widetilde{P}$, then $J(P,\widetilde{P}) = \infty$, $H(P,\widetilde{P}) = 0$, $\rho_m^m(P,\widetilde{P}) = 2$, so the statements (b) and (c) are trivial in that case. Thus, let us assume that $P \sim \widetilde{P}$.
Let I_0 be a countable subset of I such that a version Z of the density $d\widetilde{P}/dP$ is measurable relative to the σ-algebra $\mathcal{G} := \sigma\{X(t),\, t \in I_0\}$.
Let $\mathcal{H}_{\mathcal{G}} := \{\xi \in \mathcal{H} \colon \xi = E(\xi \mid \mathcal{G})\ P\text{-a.s.}\}$. Then $\mathcal{H}_{\mathcal{G}}$ is a closed separable subspace of $\mathcal{H}$. By the theorem on normal correlation, the projection of any $\xi \in \mathcal{H}$ on $\mathcal{H}_{\mathcal{G}}$ is $E(\xi \mid \mathcal{G})$;

moreover, if ξ is orthogonal to $\mathcal{H}_{\mathcal{G}}$ then ξ is independent of $\mathcal{G}$. Therefore, $A\xi = 0$ if ξ is orthogonal to $\mathcal{H}_{\mathcal{G}}$. Indeed, using (4.15), for any $\eta \in \mathcal{H}$ we get

$$\begin{aligned}
\langle A\xi, \eta\rangle &= E\xi\eta - \widetilde{E}\xi\eta = E\{(1-Z)\xi\eta\} \\
&= EE\{(1-Z)\xi E(\eta \mid \mathcal{G}) + (1-Z)\xi(\eta - E(\eta \mid \mathcal{G}))| \,\mathcal{G}\} \\
&= E\{(1-Z)E(\xi \mid \mathcal{G})E(\eta \mid \mathcal{G})\} + E(1-Z)E\{\xi(\eta - E(\eta \mid \mathcal{G}))\} = 0.
\end{aligned}$$

Due to the spectral decomposition theorem, there is an orthonormal system $\{\xi_n\}_{n=1,2,\ldots}$, $\xi_n \in \mathcal{H}_{\mathcal{G}}$, which consists of eigenvectors of the self-adjoint Hilbert–Schmidt operator A and is complete in $\mathcal{H}_{\mathcal{G}}$. If we denote the corresponding eigenvalues by $-\beta_n$, then

$$A\xi = -\sum_{n=1}^{\infty} \beta_n \langle \xi, \xi_n\rangle \xi_n, \quad \xi \in \mathcal{H}, \tag{4.16}$$

$$\|A\|_{(2)}^2 = \sum_{n=1}^{\infty} \beta_n^2. \tag{4.17}$$

As it was explained in the previous subsection, $\beta_n > -1$. It follows from (4.16) and (4.15) that $\{\xi_n\}$ is an orthogonal system in $\widetilde{\mathcal{H}}$ with $\langle \xi_n, \xi_n\rangle^{\sim} = 1 + \beta_n$.
In other words, $\{\xi_n\}$ is a sequence of independent Gaussian variables with respect to both P and $\widetilde{P}$, $E\xi_n = \widetilde{E}\xi_n = 0$, $E\xi_n^2 = 1$, $\widetilde{E}\xi_n^2 = 1 + \beta_n$. Define a mapping $T\colon \Omega \to \mathbb{R}^\infty$ by the formula

$$T(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega), \ldots).$$

It is clear that the images P^T and $\widetilde{P}^T$ of P and $\widetilde{P}$ under T satisfy the assumptions of Lemma 4.2. On the other hand, if $\sigma\{T\}$ is the sub-σ-algebra of $\mathcal{F}$ generated by T, $\sigma\{T\}$ and $\mathcal{G}$ coincide up to P-null or $\widetilde{P}$-null sets since $\{\xi_n\}$ is complete in $\mathcal{H}_{\mathcal{G}}$, i.e.

$$Z = \frac{d\widetilde{P}|_{\sigma\{T\}}}{dP|_{\sigma\{T\}}} \quad P\text{- and } \widetilde{P}\text{-a.s.}.$$

In view of the general formula

$$\frac{d\widetilde{P}|_{\sigma\{T\}}}{dP|_{\sigma\{T\}}}(\omega) = \frac{d\widetilde{P}^T}{dP^T}(T(\omega)) \quad P\text{- and } \widetilde{P}\text{-a.s.},$$

we obtain

$$\rho_m(P,\widetilde{P}) = \rho_m(P^T,\widetilde{P}^T), \qquad H(P,\widetilde{P}) = H(P^T,\widetilde{P}^T), \qquad J(P,\widetilde{P}) = J(P^T,\widetilde{P}^T).$$

The claim follows from Lemma 4.2 and (4.17).

4.5 Stationary case

Corollary 4.4 gives an estimate from above for $\rho_m(P,\widetilde{P})$ in terms of the Hilbert–Schmidt norm $\|A\|_{(2)}$ of the operator A. In this subsection we shall find an estimate from above for $\|A\|_{(2)}$ in a special case of stationary Gaussian processes.

In addition to the assumptions of Theorem 4.3, we shall assume that I is a closed interval, say, $[0,\tau]$, $\tau > 0$, of $\mathbb{R}$, $X(t)$ is a *stationary* Gaussian process with respect to both P and $\widetilde{P}$ with the covariances $K(t-s) = EX(t)X(s)$ and $\widetilde{K}(t-s) = \widetilde{E}X(t)X(s)$, $s,t \in [0,\tau]$, respectively. Put $R(t) = K(t) - \widetilde{K}(t)$, $t \in [-\tau,\tau]$.

Theorem 4.5 *Assume that* $X = (X(t),\, t \in [0,\tau])$, *is a stationary Gaussian process with respect to both* P *and* $\widetilde{P}$, X *has a spectral density* $f(\lambda)$ *with respect to* P *satisfying the inequality*

$$f(\lambda) \geq \frac{B}{1+\lambda^2}, \quad \lambda \in \mathbb{R}, \tag{4.18}$$

with some $B > 0$. *Suppose also that* $R(t)$ *is differentiable on the interval* $(-\tau,\tau)$, *the derivative* $\dot{R}(t)$ *is absolutely continuous on* $(-\tau,\tau)$, *and the second derivative* $\ddot{R}(t)$ *satisfies*

$$\iint_{[0,\tau]\times[0,\tau]} \ddot{R}^2(s-t)\,ds\,dt < \infty. \tag{4.19}$$

Then $P \sim \widetilde{P}$ *and*

$$\begin{aligned} \|A\|^2_{(2)} \;\leq\; & \frac{C}{B^2}\Bigg\{R^2(0) + R^2(\tau) + \int_0^\tau [R^2(s) + \dot{R}^2(s)]\,ds \\ & + \iint_{[0,\tau]\times[0,\tau]} [R^2(s-t) + 2\dot{R}^2(s-t) + \ddot{R}^2(s-t)]\,ds\,dt\Bigg\}, \end{aligned} \tag{4.20}$$

where C *is an absolute constant.*

Proof: It is known that $P \sim \widetilde{P}$ under the assumptions of Theorem 4.5, see Ibragimov and Rozanov (1978, Chapter III, Theorem 13, p. 99). The subsequent proof is inspired by the proof of the just mentioned theorem.
Let $b(s,t) := R(s-t)$, $(s,t) \in V := [0,\tau] \times [0,\tau]$. Our first step is to extend the function $b(s,t)$ to the whole plane so that it has a finite support and satisfies some additional properties, see inequality (4.21) below.
First, define $b(s,t) = 0$ if $(s,t) \notin U$, where, say, $U = (-1,\tau+1) \times (-1,\tau+1)$. Then define $b(s,t)$ for $(s,t) \in U \setminus V$, $t \in [0,\tau]$, in such a way that, for any fixed $t \in [0,\tau]$, the function $b(s,t)$ is linear in s on the intervals $[-1,0]$ and $[\tau,\tau+1]$. Finally, define $b(s,t)$ for $(s,t) \in U \setminus V$, $t \notin [0,\tau]$, in such a way that, for any fixed $s \in (-1,\tau+1)$, the function $b(s,t)$ is linear in t on the intervals $[-1,0]$ and $[\tau,\tau+1]$.
With this definition of $b(s,t)$, one can directly check that

- the partial derivatives $b_{10}(s,t) := \frac{\partial b(s,t)}{\partial s}$ and $b_{01}(s,t) := \frac{\partial b(s,t)}{\partial t}$ are defined everywhere except the lines $s=-1$, $s=0$, $s=\tau$, $s=\tau+1$ and $t=-1$, $t=0$, $t=\tau$, $t=\tau+1$ respectively;

- the mixed derivative $b_{11}(s,t) := \frac{\partial b(s,t)}{\partial s \partial t}$ is defined almost everywhere for Lebesgue measure on $\mathbb{R}^2$ and belongs to $L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ due to (4.19);

- $b_{10}(s,t) = \int\limits_{-\infty}^{t} b_{11}(s,u)\,du$ if $s \notin \{-1,0,\tau,\tau+1\}$;

- $b_{01}(s,t) = \int\limits_{-\infty}^{s} b_{11}(u,t)\,du$ if $t \notin \{-1,0,\tau,\tau+1\}$;

- $b_{00}(s,t) := b(s,t) = \int\limits_{-\infty}^{s} b_{10}(u,t)\,du = \int\limits_{-\infty}^{t} b_{01}(s,u)\,du = \iint\limits_{(-\infty,s]\times(-\infty,t]} b_{11}(u,v)\,du\,dv$, $(s,t)\in\mathbb{R}^2$;

- $$\begin{aligned}\sum_{k,l=0}^{1}\iint\limits_{\mathbb{R}\times\mathbb{R}} b_{kl}^2(s,t)\,ds\,dt &\le \frac{32}{9}(R^2(0)+R^2(\tau)) + \frac{16}{3}\int\limits_0^\tau [R^2(s)+\dot{R}^2(s)]\,ds \\ &+ \iint\limits_{[0,\tau]\times[0,\tau]} [R^2(s-t)+2\dot{R}^2(s-t)+\ddot{R}^2(s-t)]\,ds\,dt. \quad (4.21)\end{aligned}$$

Let now $\psi_{kl}(\lambda,\mu)$, $k,l=0,1$, be the Fourier transform of $b_{kl}(s,t)$:

$$\psi_{kl}(\lambda,\mu) = \frac{1}{4\pi^2}\iint\limits_{\mathbb{R}\times\mathbb{R}} e^{i(\lambda s-\mu t)} b_{kl}(s,t)\,ds\,dt. \quad (4.22)$$

Integrating by parts, we obtain

$$\begin{aligned}\psi_{10}(\lambda,\mu) &= \frac{1}{4\pi^2}\int\limits_{\mathbb{R}}\Big(\int\limits_{\mathbb{R}} e^{i(\lambda s-\mu t)} b_{10}(s,t)\,ds\Big)\,dt \\ &= -\frac{1}{4\pi^2}\int\limits_{\mathbb{R}}\Big(\int\limits_{\mathbb{R}} i\lambda e^{i(\lambda s-\mu t)} b_{00}(s,t)\,ds\Big)\,dt = -i\lambda\psi_{00}(\lambda,\mu) \quad (4.23)\end{aligned}$$

and, similarly,

$$\psi_{01}(\lambda,\mu) = i\mu\psi_{00}(\lambda,\mu), \qquad \psi_{11}(\lambda,\mu) = \lambda\mu\psi_{00}(\lambda,\mu). \quad (4.24)$$

Using (4.23), (4.24), and Parseval's equality, we obtain

$$\begin{aligned}\iint\limits_{\mathbb{R}\times\mathbb{R}} |\psi_{00}(\lambda,\mu)|^2(1+\lambda^2)(1+\mu^2)\,d\lambda\,d\mu &= \iint\limits_{\mathbb{R}\times\mathbb{R}}\sum_{k,l=0}^{1}|\psi_{kl}(\lambda,\mu)|^2\,d\lambda\,d\mu \\ &= \frac{1}{4\pi^2}\iint\limits_{\mathbb{R}\times\mathbb{R}}\sum_{k,l=0}^{1} b_{kl}^2(s,t)\,ds\,dt < \infty. \quad (4.25)\end{aligned}$$

Let $L^2_{F\times F}$ be the complex Hilbert space of complex-valued functions $\varphi(\lambda,\mu)$, $(\lambda,\mu)\in\mathbb{R}^2$, such that $\iint_{\mathbb{R}\times\mathbb{R}}|\varphi(\lambda,\mu)|^2 f(\lambda)f(\mu)\,d\lambda\,d\mu < \infty$ with the scalar product

$$\langle\varphi_1,\varphi_2\rangle_{F\times F} = \iint\limits_{\mathbb{R}\times\mathbb{R}} \varphi_1(\lambda,\mu)\overline{\varphi_2(\lambda,\mu)}f(\lambda)f(\mu)\,d\lambda\,d\mu,$$

where a bar means the complex conjugate. Define $\psi(\lambda,\mu)=\frac{\psi_{00}(\lambda,\mu)}{f(\lambda)f(\mu)}$. Then $\psi\in L^2_{F\times F}$ and $\|\psi\|^2_{F\times F}$ does not exceed the right-hand side of (4.20). Indeed, use (4.18) to obtain

$$\begin{aligned}\iint\limits_{\mathbb{R}\times\mathbb{R}} |\psi(\lambda,\mu)|^2 f(\lambda)f(\mu)\,d\lambda\,d\mu &= \iint\limits_{\mathbb{R}\times\mathbb{R}} \frac{|\psi_{00}(\lambda,\mu)|^2}{f(\lambda)f(\mu)}\,d\lambda\,d\mu \\ &\le \frac{1}{B^2}\iint\limits_{\mathbb{R}\times\mathbb{R}} |\psi_{00}(\lambda,\mu)|^2(1+\lambda^2)(1+\mu^2)\,d\lambda\,d\mu\end{aligned}$$

and then apply (4.25) and (4.21).

The concluding part of the proof is to show that $\|A\|_{(2)}\le\|\psi\|_{F\times F}$. We use some notation from subsections 4.3 and 4.4. In particular, let $\mathcal{K}$ be the space of linear combinations $\sum_k c_k X(t_k)$, $c_k\in\mathbb{R}$, $t_k\in[0,\tau]$. As it follows from the proof of Theorem 4.3, there is an orthonormal system $\{\xi_n\}_{n=1,2,\ldots}$ such that $\xi_n\in\mathcal{K}$ for all n and the closure of the linear span generated by $\{\xi_n\}$ contains all eigenvectors of A with non-zero eigenvalues. By the well-known property of Hilbert-Schmidt operators we have

$$\|A\|^2_{(2)}=\sum_{n,m=1}^{\infty}|\langle A\xi_n,\xi_m\rangle|^2.$$

Let L^2_F be the complex Hilbert space of complex-valued functions $\varphi(\lambda)$, $\lambda\in\mathbb{R}$, such that $\int_{\mathbb{R}}|\varphi(\lambda)|^2 f(\lambda)\,d\lambda<\infty$ with the scalar product

$$\langle\varphi_1,\varphi_2\rangle_F=\int\limits_{\mathbb{R}}\varphi_1(\lambda)\overline{\varphi_2(\lambda)}f(\lambda)\,d\lambda.$$

It is easy to check that the mapping $\Phi\colon\mathcal{K}\to L^2_F$ defined by $\Phi(\xi)(\lambda)=\sum_k c_k e^{i\lambda t_k}$, where $\xi=\sum_k c_k X(t_k)$, is well defined and preserves the scalar product. Thus, $\overline{\{\varphi_n\}_{n=1,2,\ldots}}$ with $\varphi_n=\Phi(\xi_n)$ is an orthonormal system in L^2_F. Put $\varphi_{nm}(\lambda,\mu)=\varphi_n(\lambda)\overline{\varphi_m(\mu)}$. Then, evidently, $\{\varphi_{nm}\}_{n,m=1,2,\ldots}$ is an orthonormal system in $L^2_{F\times F}$.

If $\xi=X(s)$ and $\eta=X(t)$, $s,t\in[0,\tau]$, then by (4.15)

$$\langle A\xi,\eta\rangle=EX(s)X(t)-\widetilde{E}X(s)X(t)=R(s-t)=b_{00}(s,t).$$

Inverting the Fourier transform in (4.22) and using the definition of ψ, we get

$$\begin{aligned}\langle A\xi,\eta\rangle &= \iint\limits_{\mathbb{R}\times\mathbb{R}} e^{-i(\lambda s-\mu t)}\psi_{00}(\lambda,\mu)\,d\lambda\,d\mu=\iint\limits_{\mathbb{R}\times\mathbb{R}} e^{-i(\lambda s-\mu t)}\psi(\lambda,\mu)f(\lambda)f(\mu)\,d\lambda\,d\mu \\ &= \iint\limits_{\mathbb{R}\times\mathbb{R}} \overline{\Phi(\xi)(\lambda)}\Phi(\eta)(\mu)\psi(\lambda,\mu)f(\lambda)f(\mu)\,d\lambda\,d\mu.\end{aligned}$$

The expressions on the left and on the right in the last equality are bilinear in ξ and η, hence the equality holds for all $\xi,\eta\in\mathcal{K}$. Therefore,

$$\langle A\xi_n,\xi_m\rangle=\iint\limits_{\mathbb{R}\times\mathbb{R}}\overline{\varphi_{nm}(\lambda,\mu)}\psi(\lambda,\mu)f(\lambda)f(\mu)\,d\lambda\,d\mu=\langle\psi,\varphi_{nm}\rangle_{F\times F}$$

and

$$\|A\|_{(2)}^2 = \sum_{n.m=1}^{\infty} |\langle A\xi_n, \xi_m\rangle|^2 = \sum_{n.m=1}^{\infty} |\langle \psi, \varphi_{nm}\rangle_{F\times F}|^2 \leq \|\psi\|_{F\times F}^2.$$

Corollary 4.6 *If in Theorem* 4.5 X *has a spectral density* $\widetilde{f}(\lambda)$ *with respect to* $\widetilde{P}$ *then*

$$\|A\|_{(2)}^2 \leq \frac{C}{B^2}\Big\{\Big(\int_{-\infty}^{+\infty} |f(\lambda) - \widetilde{f}(\lambda)|\, d\lambda\Big)^2 + (1+\tau)\int_{-\infty}^{+\infty} (1+\lambda^4)(f(\lambda) - \widetilde{f}(\lambda))^2 d\lambda\Big\}, \quad (4.26)$$

where C *is an absolute constant.*

Proof: If the second integral on the right in (4.26) equals ∞, then the statement is trivial. Thus, let us assume that

$$\int_{-\infty}^{+\infty} (1+\lambda^4)(f(\lambda) - \widetilde{f}(\lambda))^2 d\lambda < \infty. \quad (4.27)$$

By the definition of $R(t)$,

$$R(t) = \int_{-\infty}^{+\infty} e^{i\lambda t}(f(\lambda) - \widetilde{f}(\lambda))\, d\lambda, \quad t \in [-\tau, \tau]. \quad (4.28)$$

Extend the definition of $R(t)$ to $\mathbb{R}$ according to (4.28). It follows from (4.27) that $\int_{-\infty}^{+\infty} |\lambda||f(\lambda) - \widetilde{f}(\lambda)|\, d\lambda < \infty$, hence we may differentiate under the integral sign in (4.28) to obtain that $R(t)$ is differentiable on $\mathbb{R}$ and

$$\dot{R}(t) = i\int_{-\infty}^{+\infty} \lambda e^{i\lambda t}(f(\lambda) - \widetilde{f}(\lambda))\, d\lambda, \quad t \in \mathbb{R}. \quad (4.29)$$

Then

$$n[\dot{R}(t+1/n) - \dot{R}(t)] = i\int_{-\infty}^{+\infty} \lambda e^{i\lambda t}[n(e^{i\lambda/n} - 1)](f(\lambda) - \widetilde{f}(\lambda))\, d\lambda.$$

The sequence of functions $i\lambda[n(e^{i\lambda/n} - 1)](f(\lambda) - \widetilde{f}(\lambda))$ converges in $L^2(-\infty, +\infty)$ as $n \to \infty$ to $-\lambda^2(f(\lambda) - \widetilde{f}(\lambda))$ due to (4.27) by the dominated convergence theorem. By Plancherel's theorem $n[\dot{R}(t+1/n) - \dot{R}(t)]$ converges in $L^2(-\infty, +\infty)$ to, say, $Q(t)$ and

$$Q(t) = -\int_{-\infty}^{+\infty} \lambda^2 e^{i\lambda t}(f(\lambda) - \widetilde{f}(\lambda))\, d\lambda \quad \text{a.e.}, \quad (4.30)$$

where the integral on the right can be understood in the sense of the principal value. But it is assumed in Theorem 4.5 that $n[\dot{R}(t+1/n) - \dot{R}(t)]$ converges a.e. in $(-\tau, \tau)$ to $\ddot{R}(t)$, hence $\ddot{R}(t) = Q(t)$ a.e. in $(-\tau, \tau)$.

Using Parseval's equality, we get from (4.28)–(4.30) that

$$R^2(0) + R^2(\tau) \le 2\Big(\int_{-\infty}^{+\infty} |f(\lambda) - \widetilde{f}(\lambda)|\, d\lambda\Big)^2, \tag{4.31}$$

$$\int_0^\tau [R^2(s) + \dot{R}^2(s)]\, ds \le \int_{-\infty}^{+\infty} [R^2(s) + \dot{R}^2(s)]\, ds = 2\pi \int_{-\infty}^{+\infty} (1+\lambda^2)(f(\lambda) - \widetilde{f}(\lambda))^2 d\lambda, \tag{4.32}$$

$$\begin{aligned} &\iint_{[0,\tau]\times[0,\tau]} [R^2(s-t) + 2\dot{R}^2(s-t) + \ddot{R}^2(s-t)]\, ds\, dt \\ &\quad = 2\int_0^\tau (\tau - s)[R^2(s) + 2\dot{R}^2(s) + \ddot{R}^2(s)]\, ds \le 2\tau \int_{-\infty}^{+\infty} [R^2(s) + 2\dot{R}^2(s) + Q(s)]\, ds \\ &\quad = 4\pi\tau \int_{-\infty}^{+\infty} (1+\lambda^2)^2 (f(\lambda) - \widetilde{f}(\lambda))^2 d\lambda. \end{aligned} \tag{4.33}$$

Now (4.26) follows from (4.20) and (4.31)–(4.33).

5 An upper estimate for the Hellinger integral

In this section, if $a \in M$ and μ is a Borel measure on $\mathbf{C}$, we denote by $P_T^{a,\mu}$ the distribution of the solution $X = (X(t),\, -r \le t \le T)$ of (1.1) in $\mathbf{C}_T$, where the distribution of the initial process $X_0 = (X_0(t),\, t \in J)$ in $\mathbf{C}$ is μ. If $a \in M_s$ and μ is such that X is a stationary process, we shall write P_T^a instead of $P_T^{a,\mu}$. The aim of this section is to provide a formula for the likelihood ratio $d\widetilde{P}/dP$ for $P = P_T^{a,\mu}$ and $\widetilde{P} = P_T^{\widetilde{a},\widetilde{\mu}}$ and to estimate from above the Hellinger integral $H(P, \widetilde{P})$ in the case where $a \in M_s$ and $P = P_T^a$, i.e. the initial condition μ corresponds to the stationary solution of (1.1).

5.1 Formula for the likelihood ratio

In this subsection we assume that a and $\widetilde{a}$ belong to M and μ and $\widetilde{\mu}$ are probability measures on $\mathbf{C}$. Given $T \in \mathbb{R}_+$, define non-anticipating functionals $(A_t(x),\, t \in [0,T])$ and $(\widetilde{A}_t(x),\, t \in [0,T])$ on the space $\mathbf{C}_T$ by

$$A_t(x) = \int_J x(t+u)\, a(du), \quad \widetilde{A}_t(x) = \int_J x(t+u)\, \widetilde{a}(du), \quad t \in [0,T], \quad x \in \mathbf{C}_T.$$

Since $\int_0^T [A_t^2(x) + \widetilde{A}_t^2(x)]\, dt < \infty$ for any $x \in \mathbf{C}_T$, the next result is, of course, not surprising. But we supply it with a short proof since we have not found a direct reference.

If $x \in \mathbf{C}_T$, we denote by $\hat{x}$ the restriction of x to the interval J.

Proposition 5.1 *If $\widetilde{\mu} \ll \mu$ (resp. $\widetilde{\mu} \sim \mu$) then $P_T^{\widetilde{a},\widetilde{\mu}} \ll P_T^{a,\mu}$ (resp. $P_T^{\widetilde{a},\widetilde{\mu}} \sim P_T^{a,\mu}$) and $P_T^{a,\mu}$-a.s.*

$$\frac{dP_T^{\widetilde{a},\widetilde{\mu}}}{dP_T^{a,\mu}}(x) = \frac{d\widetilde{\mu}}{d\mu}(\hat{x}) \exp\left\{ \int_0^T [\widetilde{A}_t(x) - A_t(x)]\, dx(t) - \frac{1}{2}\int_0^T [\widetilde{A}_t^2(x) - A_t^2(x)]\, dt \right\}$$
$$= \frac{d\widetilde{\mu}}{d\mu}(\hat{x}) \exp\left\{ \int_0^T [\widetilde{A}_t(x) - A_t(x)]\, dw(t) - \frac{1}{2}\int_0^T [\widetilde{A}_t(x) - A_t(x)]^2\, dt \right\}, \quad (5.1)$$

where $w = (w(t),\, t \in [0,T])$ is a Wiener process on $(\mathbf{C}_T, \mathcal{B}(\mathbf{C}_t), P_T^{a,\mu})$ defined by $w(t) = x(t) - x(0) - \int_0^t A_s(x)\, ds$.

Proof: Assume first that $\widetilde{a} = 0$. If $\widetilde{\mu} \ll \mu$, our statement follows from Theorem III.5.34 in Jacod and Shiryaev (1987) (note that the measure $P_T^{0,\widetilde{\mu}}$ satisfies the local unicity property as it is required in that theorem). Since the exponent in (5.1) is strictly positive, we obtain that $P_T^{0,\widetilde{\mu}} \sim P_T^{a,\mu}$ if $\widetilde{\mu} \sim \mu$.
In the general case we have now that $P_T^{\widetilde{a},\widetilde{\mu}} \sim P_T^{0,\widetilde{\mu}} \ll P_T^{a,\mu}$. We also know expressions for the densities $dP_T^{\widetilde{a},\widetilde{\mu}}/dP_T^{0,\widetilde{\mu}} = \left(dP_T^{0,\widetilde{\mu}}/dP_T^{\widetilde{a},\widetilde{\mu}}\right)^{-1}$ and $dP_T^{0,\widetilde{\mu}}/dP_T^{a,\mu}$. Since $dP_T^{\widetilde{a},\widetilde{\mu}}/dP_T^{a,\mu} = \left(dP_T^{\widetilde{a},\widetilde{\mu}}/dP_T^{0,\widetilde{\mu}}\right)\left(dP_T^{0,\widetilde{\mu}}/dP_T^{a,\mu}\right)$ $P_T^{a,\mu}$-a.s., the claim easily follows.

5.2 An upper estimate

Recall that the dual Lipschitz norm $\|\cdot\|_D$ is defined in subsection 3.3.

Theorem 5.2 *Assume that $P = P_T^a$, $\widetilde{P} = P_T^{\widetilde{a},\widetilde{\mu}}$, where $a \in M_s$, $\widetilde{a} \in M$, $\widetilde{\mu}$ is a probability measure on $\mathbf{C}$, and $T \in \mathbb{R}_+$. Let $L > 0$ and $\gamma < 0$ be numbers such that the fundamental solution $x_0(t)$ corresponding to a satisfies (3.6), and B_* a number such that the spectral density $f(\lambda)$ of the stationary solution to (1.1) corresponding to a satisfies the left-hand inequality in (3.13). Then*

$$H(P, \widetilde{P}) \le \exp\left\{ -\frac{C_r B_* \|\widetilde{a} - a\|_D^2 T}{1 + C_{\gamma,L}\|\widetilde{a} - a\|_v^2} \right\}, \quad (5.2)$$

where $C_r > 0$ depends only on r and $C_{\gamma,L} = \gamma^{-2}L^2 e^{-\gamma r}$.

Proof: Put $b := \widetilde{a} - a$ and define a stochastic process $Y(t)$, $0 \le t \le T$, on $(\mathbf{C}_T, \mathcal{B}(\mathbf{C}_T), P)$ by

$$Y(t) = \int_J x(t+u)\, b(du).$$

Let μ be the initial distribution on $\mathbf{C}$ of a stationary solution to (1.1) corresponding to a. If $\widetilde{\mu} \ll \mu$, one can use (5.1) and a standard trick with Hölder's inequality to obtain

$$H(P, \widetilde{P}) \le \left\{ E \exp\left(-\frac{1}{3}\int_0^T Y^2(t)\, dt \right) \right\}^{1/4}, \quad (5.3)$$

where E is expectation with respect to P. In fact, this inequality is true even if $\widetilde{\mu}$ is not absolutely continuous with respect to μ. For instance, one can apply Corollary V.4.19 and Theorem IV.3.39 in Jacod and Shiryaev (1987).
Since $(x(t),\, t \in [-r,T])$ is a stationary Gaussian process with zero mean (relative to P), $(Y(t),\, t \in [0,T])$ is also a stationary Gaussian process with zero mean and covariance

$$\overline{R}(t-s) := R(s,t) := E(Y(s)Y(t)) = \iint\limits_{J\times J} K(t-s+u-v)\, b(du)\, b(dv), \qquad s,t \in [0,T], \tag{5.4}$$

where $K(t) := Ex(s)x(t+s)$. Let R_T be the covariance operator in $L^2[0,T]$ corresponding to R, i.e. the integral operator with the kernel $R(s,t)$. By Lemma 3.1 in Kallianpur and Selukar (1991),

$$E\exp\Big(-\frac{1}{3}\int_0^T Y^2(t)\,dt\Big) \le \exp\Big(-\frac{\mathrm{tr}\,(R_T)}{3+2\|R_T\|}\Big), \tag{5.5}$$

where $\|R_T\|$ is the operator norm of R_T and

$$\mathrm{tr}\,(R_T) = \int_0^T R(t,t)\,dt = T\overline{R}(0) \tag{5.6}$$

is the trace of R_T.
We shall show that

$$\overline{R}(0) \ge C_r B_* \|b\|_D^2 \qquad \text{and} \qquad \|R_T\| \le C_{\gamma,L}\|b\|_v^2. \tag{5.7}$$

Then the statement of the theorem (with a different C_r) follows from (5.3) and (5.5)–(5.7).
Let us first check the second inequality in (5.7). It follows from (5.4), (3.9) and (3.6) that

$$|\overline{R}(t)| \le (2|\gamma|)^{-1}L^2 e^{-\gamma r} e^{\gamma|t|}\|b\|_v^2, \qquad t \in [-T,T].$$

Hence

$$\int_0^T |\overline{R}(t-s)|\,ds \le C_{\gamma,L}\|b\|_v^2 \qquad \text{for any} \quad t \in [0,T].$$

Now, for any $g \in L^2[0,T]$, cf. the proof of Lemma 3.2 in Kallianpur and Selukar (1991),

$$[(R_T g)(t)]^2 = \Big(\int_0^T \overline{R}(t-s)g(s)\,ds\Big)^2 \le C_{\gamma,L}\|b\|_v^2 \int_0^T |\overline{R}(t-s)|g^2(s)\,ds, \qquad t \in [0,T],$$

and

$$\int_0^T [(R_T g)(t)]^2 dt \le C_{\gamma,L}\|b\|_v^2 \int_0^T\!\!\int_0^T |\overline{R}(t-s)|g^2(s)\,ds\,dt \le C_{\gamma,L}^2\|b\|_v^4 \int_0^T g^2(s)\,ds.$$

The second inequality in (5.7) follows.
The proof of the first inequality in (5.7) is direct if $r = 0$, so we shall assume that $r > 0$. The key idea is to replace the covariance K in (5.4) by the function $\widehat{K}$ of a special form. To realize it, put $\widehat{K}(t) := (r - |t|)^+$. Then $\widehat{K}(t) = \int_{\mathbb{R}} e^{i\lambda t} \widehat{f}(\lambda)\, d\lambda$, $t \in \mathbb{R}$, where $\widehat{f}(\lambda) = \frac{1-\cos(r\lambda)}{\pi\lambda^2}$. There is a positive constant C_r, depending on r, such that $(1+\lambda^2)^{-1} \geq C_r \widehat{f}(\lambda)$, $\lambda \in \mathbb{R}$. Then, using (5.4), (3.10) and (3.13), we get

$$\begin{aligned}\overline{R}(0) &= \int_{\mathbb{R}} \left| \int_J e^{i\lambda u}\, b(du) \right|^2 f(\lambda)\, d\lambda \geq C_r B_* \int_{\mathbb{R}} \left| \int_J e^{i\lambda u}\, b(du) \right|^2 \widehat{f}(\lambda)\, d\lambda \\ &= C_r B_* \iint_{J\times J} \widehat{K}(u-v)\, b(du)\, b(dv). \end{aligned} \tag{5.8}$$

Note that

$$\begin{aligned}\iint_{J\times J} \widehat{K}(u-v)\, b(du)\, b(dv) &= \iint_{J\times J} (r - |u-v|)\, b(du)\, b(dv) \\ &= \iint_{J\times J} \int_{-r}^{0} \left(\mathbf{1}_{\{u\vee v \leq t\}} + \mathbf{1}_{\{u\wedge v > t\}}\right) dt\, b(du)\, b(dv) \\ &= \int_{-r}^{0} \left\{ b^2([-r,t]) + b^2((t,0]) \right\} dt \\ &\geq \frac{1}{2} \int_{-r}^{0} \left\{ |b([-r,t])| + |b((t,0])| \right\}^2 dt. \end{aligned} \tag{5.9}$$

To complete the proof, we shall estimate the right-hand side of (5.9) in terms of $\|b\|_D$. We extend the arguments used by Vallander (1973). Let $b = b^+ - b^-$ be the Hahn–Jordan decomposition of the measure b. Put $F(t) := b^+([-r,t])$, $G(t) := b^-([-r,t])$, $t \in J$, and $F^{-1}(s) := \inf\{t\colon F(t) > s\}$, $s \in [0, F(0))$, $G^{-1}(s) := \inf\{t\colon G(t) > s\}$, $s \in [0, G(0))$. Then, for any $g \in \mathrm{Lip}_1$,

$$\begin{aligned}\int_J g(t)\, b(dt) &= \int_J g(t)\, F(dt) - \int_J g(t)\, G(dt) = \int_0^{F(0)} g(F^{-1}(s))\, ds - \int_0^{G(0)} g(G^{-1}(s))\, ds \\ &= \int_0^{F(0)\wedge G(0)} \{g(F^{-1}(s)) - g(G^{-1}(s))\}\, ds + \int_{F(0)\wedge G(0)}^{F(0)} g(F^{-1}(s))\, ds \\ &\quad - \int_{F(0)\wedge G(0)}^{G(0)} g(G^{-1}(s))\, ds \leq \int_0^{F(0)\wedge G(0)} |F^{-1}(s) - G^{-1}(s)|\, ds \end{aligned}$$

$$+ |F(0) - G(0)| \le \int_J |F(t) - G(t)|\, dt + |F(0) - G(0)|$$

$$= \int_J |b([-r,t])|\, dt + |b(J)|.$$

Therefore,

$$\|b\|_D^2 \le \Big\{ \int_J |b([-r,t])|\, dt + |b(J)| \Big\}^2 \le \Big\{ (1+r^{-1}) \int_J |b([-r,t])|\, dt + r^{-1} \int_J |b((t,0])|\, dt \Big\}^2$$

$$\le 2r(1+r^{-1})^2 \int_J \{|b([-r,t])| + |b((t,0])|\}^2\, dt. \tag{5.10}$$

Combining (5.8)–(5.10), we obtain the first inequality in (5.7).

6 Proof of Theorem 2.1

(1) The statement follows from Corollary 3.2.
(4) The statement follows from Theorem 3.3 applied to $a = a_{\vartheta+\varphi_T(\vartheta)u}$ and $\widetilde{a} = a_{\vartheta+\varphi_T(\vartheta)v}$, if we take into account (3.17), Condition 2, and the fact that

$$\|\varphi_T(\vartheta)(u-v)\| = T^{-1/2}\|\Sigma^{-1/2}(\vartheta)(u-v)\| \le C'T^{-1/2}\|u-v\|,$$

where C' does not depend on ϑ, u, v, or T. Indeed, $\Sigma(\vartheta)$ is continuous and non-degenerate on Θ by Lemma 3.4, hence its minimal eigenvalue is separated from 0 on $\mathbf{K}$.
(2) The existence of a continuous modification of the random field $Z_{T,\vartheta}(u)$ follows from (2.2) with $m > k$ and from e.g. Theorem 19, Appendix I, p. 372 in Ibragimov and Has'minskii (1981).
(5) In this part of the proof the letter B with subscripts will be used for positive constants depending only on Θ, r, the family $\mathcal{A}$, and the compact $\mathbf{K}$.
Let $T > 0$, $\vartheta \in \mathbf{K}$, and $u \in U_{T,\vartheta}$. By Condition 2 and the same argument as in the proof of part (4),

$$\|a_{\vartheta+\varphi_T(\vartheta)u} - a_\vartheta\|_v \le B_1 T^{-1/2}\|u\|. \tag{6.1}$$

Next,

$$\|a_{\vartheta+\varphi_T(\vartheta)u} - a_\vartheta\|_D \ge B_2\|\varphi_T(\vartheta)u\| = B_2 T^{-1/2}\|\Sigma^{-1/2}(\vartheta)u\| \ge B_3 T^{-1/2}\|u\|, \tag{6.2}$$

where we used the inequality (3.20) and the fact that the maximal eigenvalue of $\Sigma(\vartheta)$ is bounded on $\mathbf{K}$ by Lemma 3.4.
Applying Theorem 5.2 and using (3.17)–(3.19), (6.1), and (6.2), we get

$$H(P_T^\vartheta, P_T^{\vartheta+\varphi_T(\vartheta)u}) \le \exp\Big\{ -\frac{B_4\|u\|^2}{1 + B_5\|u\|^2 T^{-1}} \Big\}.$$

Since Θ is a bounded set, we have $\|u\| \le B_6 T^{1/2}$, therefore

$$H(P_T^{\vartheta}, P_T^{\vartheta+\varphi_T(\vartheta)u}) \le \exp\left\{-\frac{B_4\|u\|^2}{1+B_5B_6}\right\}.$$

The claim follows.
(3) To simplify the notation put $P^n = P_{T_n}^{\vartheta_n}$. Let us fix for a while a vector $\gamma \in \mathbb{R}^k$ and introduce a stochastic process $Y_n(t)$, $0 \le t \le T_n$, on $(\mathbf{C}_{T_n}, \mathcal{B}(\mathbf{C}_{T_n}), P^n)$ by

$$Y_n(t) = \int_J x(t+s)\, b_n(ds), \qquad \text{where} \quad b_n = \langle \gamma, \dot a_{\vartheta_n} \rangle. \tag{6.3}$$

Since the coordinate process $x(t)$ is a stationary Gaussian process with respect to P^n, it is easy to check that $Y_n(t)$ is a stationary Gaussian process with the covariance function (E^n is the expectation with respect to P^n)

$$K_n(t) = E^n Y_n(s) Y_n(t+s) = \iint_{J\times J} K_{\vartheta_n}(t+u-v) b_n(du) b_n(dv). \tag{6.4}$$

Note that

$$\sup_n \|b_n\|_v < \infty. \tag{6.5}$$

Indeed, otherwise we can find a subsequence $\{\vartheta_{n_k}\}$ such that $\|b_{n_k}\|_v \to \infty$ and $\vartheta_{n_k} \to \vartheta \in \mathbf{K}$ as $k \to \infty$. But it follows from Condition 4 that $\{b_{n_k}\}$ $*$-weakly converges to $\langle \gamma, \dot a_{\vartheta} \rangle$ as $k \to \infty$, hence the norms $\|b_{n_k}\|_v$ are bounded.
Evidently, we have

$$E^n \frac{1}{T_n} \int_0^{T_n} Y_n^2(t)\, dt = K_n(0) = \langle \gamma, \Sigma(\vartheta_n)\gamma \rangle.$$

An easy calculation shows that

$$E^n \left(\frac{1}{T_n} \int_0^{T_n} Y_n^2(t)\, dt - K_n(0) \right)^2 = \frac{2}{T_n^2} \int_0^{T_n}\int_0^{T_n} [K_n(t-s)]^2 ds dt.$$

It is easy to check that the expression on the right tends to 0 as $n \to \infty$: use (6.4), (6.5), (3.9) and the estimate (3.19). Therefore,

$$\frac{1}{T_n} \int_0^{T_n} Y_n^2(t)\, dt - \langle \gamma, \Sigma(\vartheta_n)\gamma \rangle \xrightarrow{P^n} 0.$$

Since $\gamma \in \mathbb{R}^k$ is arbitrary, we obtain from (6.3) by polarization that

$$\frac{1}{T_n} \int_0^{T_n} \Big(\int_J x(t+s)\, \dot a_{\vartheta_n,i}(ds) \Big) \Big(\int_J x(t+s)\, \dot a_{\vartheta_n,j}(ds) \Big) dt - \Sigma_{ij}(\vartheta_n) \xrightarrow{P^n} 0.$$

In other words, since the elements of the matrices $\Sigma^{-1/2}(\vartheta_n)$ are uniformly bounded by Lemma 3.4,

$$\Big(\varphi_{T_n}(\vartheta_n)\int_J x(t+s)\,\dot a_{\vartheta_n}(ds)\Big)\Big(\varphi_{T_n}(\vartheta_n)\int_J x(t+s)\,\dot a_{\vartheta_n}(ds)\Big)^\top \xrightarrow{P^n} I_k. \tag{6.6}$$

By the central limit theorem for stochastic integrals, see e.g. Basawa and Prakasa Rao (1980), Appendix 2, Theorem 2.1, p. 405,

$$\mathcal{L}(\Delta_{T_n,\vartheta_n} \mid P^n) \Longrightarrow N(0, I_k), \qquad n \to \infty. \tag{6.7}$$

Now put

$$\widehat{Y}_n(t) = \int_J x(t+s)\,\widehat{b}_n(ds), \qquad \text{where} \quad \widehat{b}_n = a_{\vartheta_n+\varphi_{T_n}(\vartheta_n)u_n} - a_{\vartheta_n} - \langle \varphi_{T_n}(\vartheta_n)u_n, \dot a_{\vartheta_n}\rangle.$$

Our next step is to show that

$$\int_0^{T_n} \widehat{Y}_n^2(t)\,dt \xrightarrow{P^n} 0. \tag{6.8}$$

Indeed, the process $\widehat{Y}_n(t)$, $0 \le t \le T_n$, is a stationary Gaussian process with respect to P^n with the covariance function

$$\widehat{K}_n(t) = E^n \widehat{Y}_n(s)\widehat{Y}_n(t+s) = \iint_{J\times J} K_{\vartheta_n}(t+u-v)\widehat{b}_n(du)\widehat{b}_n(dv)$$

and the spectral density

$$\widehat{f}_n(\lambda) := \Big|\int_J e^{i\lambda u}\,\widehat{b}_n(du)\Big|^2 f_{\vartheta_n}(\lambda).$$

Hence,

$$E^n \int_0^{T_n} \widehat{Y}_n^2(t)\,dt = T_n\widehat{K}_n(0) = \int_{-\infty}^{\infty} \Big|T_n^{1/2}\int_J e^{i\lambda u}\widehat{b}_n(du)\Big|^2 f_{\vartheta_n}(\lambda)d\lambda. \tag{6.9}$$

Since $\Sigma^{-1/2}(\vartheta_n)u_n$ is a bounded sequence due to Lemma 3.4, we obtain from Condition 4 that

$$T_n^{1/2}\int_J g(s)\widehat{b}_n(ds) \longrightarrow 0, \quad n\to\infty, \quad \text{for any} \quad g \in \mathbf{C}. \tag{6.10}$$

Using the upper estimate in (3.17), we obtain that the integrand in the right-hand side of (6.9) converges to zero as $n \to \infty$. On the other hand, (6.10) also implies that $\sup_n T_n^{1/2}\|b_n\|_v < \infty$. Using (3.17) again we obtain that the integral on the right in (6.10) tends to zero by the dominated convergence theorem, and (6.8) follows.

Let $\mu_n := P_0^{\vartheta_n}$ (resp. $\mu_n' := P_0^{\vartheta_n+\varphi_{T_n}(\vartheta_n)u_n}$) be the initial distribution on $(\mathbf{C}, \mathcal{B}(\mathbf{C}))$ of the stationary solution to (1.1) with $a = a_{\vartheta_n}$ (resp. $a = a_{\vartheta_n+\varphi_{T_n}(\vartheta_n)u_n}$). Introduce new probability measures Q^n and $\widehat{Q}^n$ on $(\mathbf{C}_{T_n}, \mathcal{B}(\mathbf{C}_{T_n}))$, which are the distributions of solutions $X = (X(t),\, t \in [-r, T_n])$ of the equation (1.1): Q^n corresponds to the initial distribution μ_n and the delay measure $a_{\vartheta_n} + \langle \varphi_{T_n}(\vartheta_n)u_n, \dot{a}_{\vartheta_n}\rangle$, and $\widehat{Q}^n$ corresponds to the initial distribution μ_n and the delay measure $a_{\vartheta_n+\varphi_{T_n}(\vartheta_n)u_n}$.
It has been already proved that $P^n \sim P'^n := P_{T_n}^{\vartheta_n+\varphi_{T_n}(\vartheta_n)u_n}$. By Proposition 5.1 $P^n \sim Q^n \sim \widehat{Q}^n$. Moreover, due to (5.1) we have

$$\log \frac{dQ^n}{dP^n} = \langle \Delta_{T_n,\vartheta_n}, u_n\rangle - \frac{1}{2}\int_0^{T_n} \left(\left\langle \varphi_{T_n}(\vartheta_n) \int_J x(t+s)\, \dot{a}_{\vartheta_n}(ds), u_n\right\rangle\right)^2 dt,$$

and it follows from (6.6) and (6.7) that

$$\log \frac{dQ^n}{dP^n} - \left(\langle \Delta_{T_n,\vartheta_n}, u\rangle - \frac{\|u\|^2}{2}\right) \xrightarrow{P^n} 0, \quad n \to \infty. \tag{6.11}$$

Let $h_n(t)$, $0 \le t \le T_n$, be the Hellinger process of order 1/2 for Q^n and $\widehat{Q}^n$ (with respect to the natural filtration). Using formula (5.1) for the density process of $\widehat{Q}^n$ with respect to Q^n, we obtain

$$h_n(t) = \frac{1}{8}\int_0^t \widehat{Y}_n^2(s)\, ds, \quad t \in [0, T_n].$$

By (6.8), $h_n(T_n) \xrightarrow{P^n} 0$. But, in view of (6.11), the sequences $\{P^n\}$ and $\{Q^n\}$ are mutually contiguous, hence $h_n(T_n) \xrightarrow{Q^n} 0$. By Theorem V.4.32 in Jacod and Shiryaev (1987)

$$\|\widehat{Q}^n - Q^n\|_v \to 0, \quad n \to \infty.$$

By Kraft's inequality, see e.g. Jacod and Shiryaev (1987), Proposition V.4.4,

$$\rho_2(\widehat{Q}^n, Q^n) \to 0, \quad n \to \infty. \tag{6.12}$$

Furthermore, since

$$\frac{dP'^n}{d\widehat{Q}^n}(x) = \frac{d\mu_n'}{d\mu^n}(\hat{x})$$

by Proposition 5.1, we have

$$\rho_2(\widehat{Q}^n, P'^n) = \rho_2(\mu_n, \mu_n') \to 0, \quad n \to \infty, \tag{6.13}$$

where the convergence statement follows from Theorem 3.3. Combining (6.12) and (6.13), we get $\rho_2(Q^n, P'^n) \to 0$, which implies

$$\frac{dP'^n}{dQ^n} \xrightarrow{Q^n} 1, \quad n \to \infty.$$

Hence,

$$\frac{dP'^n}{dQ^n} \xrightarrow{P^n} 1, \quad n \to \infty.$$

due to the mutual contiguity of $\{P^n\}$ and $\{Q^n\}$. Now we get

$$\log \frac{dP'^n}{dP^n} - \log \frac{dQ^n}{dP^n} = \log \frac{dP'^n}{dQ^n} \xrightarrow{P^n} 0, \quad n \to \infty,$$

and the first statement in (3) follows from (6.11).

References

Basawa, I. V. and Prakasa Rao, B. L. S. (1980) *Statistical inference for stochastic processes.* London: Academic Press.

Diekmann, O., van Gils, S. A., Verduyn Lunel, S. M., and Walther, H.-O. (1995) *Delay equations: functional-, complex-, and nonlinear analysis.* New York: Springer-Verlag.

Dietz, H. M. (1992) A non-Markovian relative of the Ornstein–Uhlenbeck process and some of its local statistical properties. *Scand. J. Stat.*, **19**, 363–379.

Dzhaparidze, K. and Valkeila, E. (1990) On the Hellinger type distances for filtered experiments. *Probab. Theory Relat. Fields*, **85**, 105–117.

Gihman, I. I. and Skorohod, A. V. (1980) *The theory of stochastic processes I.* Grundlehren der mathematischen Wissenschaften, 210. Berlin, Heidelberg, New York: Springer-Verlag.

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 Processes Appl.*, **88**, 195–211.

Gushchin, A. A. and Küchler, U. (2001) Addendum to 'Asymptotic inference for a linear stochastic differential equation with time delay'. *Bernoulli*, **7**, 629–632.

Hale, J. K. and Verduyn Lunel, S. M. (1993) *Introduction to functional-differential equations.* New York: Springer-Verlag.

Ibragimov, I. A. and Has'minskii, R. Z. (1981) *Statistical estimation. Asymptotic theory.* New York etc.: Springer-Verlag.

Ibragimov, I. A. and Rozanov, Y. A. (1978) *Gaussian random processes.* New York etc.: Springer-Verlag.

Jacod, J. and Shiryaev, A.N. (1987). *Limit theorems for stochastic processes.* Berlin New York: Springer-Verlag.

Kallianpur, G. and Selukar, R. S. (1991) Parameter estimation in linear filtering. *J. Multivariate Anal.*, **39**, 284–304.

Kantorovich, L. V. and Akilov, G. P. (1982) *Functional analysis*, 2nd ed. Oxford etc.: Pergamon Press.

Küchler, U. and Kutoyants, Yu. A. (2000) Delay estimation for some stationary diffusion-type processes. *Scand. J. Stat.*, **27**, 405–414.

Kuo, H.-H. (1975). *Gaussian measures in Banach spaces.* Lecture Notes in Mathematics. 463. Berlin - Heidelberg - New York: Springer-Verlag.

Liese, F. and Vajda, I. (1987) *Convex statistical distances.* Leipzig: Teubner.

Mohammed, S.-E. A. and Scheutzow, M. K. R. (1990) Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics and Stochastics Reports*, **29**, 259–283.

Myschkis, A. D. (1972) *Linear differential equations with delayed argument.* Moscow: Nauka (in Russian).

Putschke, U. (2000) Affine stochastische Funktionaldifferentialgleichungen und lokal asymptotische Eigenschaften ihrer Parameterschätzungen. Doctoral dissertation, Humboldt-Universität zu Berlin, Institute of Mathematics.

Reiss, M. (2000) Minimax rates for nonparametric estimation of the drift functional in affine stochastic delay equations. Discussion Paper 77/2000, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Rozanov, Ju. A. (1971) *Infinite-dimensional Gaussian distributions.* Proceedings of the Steklov Institute of Mathematics, No. 108 (1968). Providence, RI: AMS.

Shepp, L. A. (1966) Gaussian measures in function space. *Pac. J. Math.*, **17**, 167–173.

Vallander, S. S. (1973) Calculation of the Wasserstein distance between probability distributions on the line. *Theory Probab. Appl.*, **18**, 784–786.