JOURNAL OF MULTIVARIATE ANALYSIS 39, 284–304 (1991)

Parameter Estimation in Linear Filtering*

G. KALLIANPUR AND R. S. SELUKAR

University of North Carolina at Chapel Hill

Communicated by C. R. Rao

Suppose on a probability space $(\Omega, \mathbf{F}, \mathbf{P})$, a partially observable random process (x_t, y_t), $t \geqslant 0$; is given where only the second component (y_t) is observed. Furthermore assume that (x_t, y_t) satisfy the following system of stochastic differential equations driven by independent Wiener processes $(W_1(t))$ and $(W_2(t))$:

$$dx_t = -\beta x_t\, dt + dW_1(t), \qquad x_0 = 0$$

$$dy_t = \alpha x_t\, dt + dW_2(t), \qquad y_0 = 0; \quad \alpha, \beta \in (a, b), a > 0.$$

We prove the local asymptotic normality of the model and obtain a large deviation inequality for the maximum likelihood estimator (m.l.e.) of the parameter $\theta = (\alpha, \beta)$. This also implies the strong consistency, efficiency, asymptotic normality and the convergence of moments for the m.l.e. The method of proof can be easily extended to obtain similar results when vector valued instead of one-dimensional processes are considered and θ is a k-dimensional vector.

1. INTRODUCTION

Suppose on a probability space $(\Omega, \mathbf{F}, \mathbf{P})$, a partially observable random process (x_t, y_t), $t \geqslant 0$; is given where only the second component (y_t) is observed (both the components could be vector valued). Furthermore assume that (x_t, y_t) satisfy the following system of stochastic differential equations (SDE):

$$\begin{aligned} dx_t &= Fx_t\, dt + G\, dW_1(t), \qquad x_0 = X_0, \\ dy_t &= Hx_t\, dt + dW_2(t), \qquad y_0 = 0, \end{aligned} \tag{1}$$

where $(W_1(t))$ and $(W_2(t))$ are *independent* standard Wiener processes and F, G, H are nonrandom matrices of appropriate order. The initial value X_0

Received April 20, 1990; revised February 8, 1991.

AMS 1980 subject classifications: primary 62G05; secondary 62M99.

Key words and phrases: Kalman filter, maximum likelihood estimation, large deviation inequality, local asymptotic normality.

* Research partially supported by the Air Force Office of the Scientific Research Contract F49620 85 C 0144.

0047-259X/91 $3.00

is assumed to be a Gaussian random variable independent of both $(W_1(t))$ and $(W_2(t))$.

Let θ (a point in R^k) be the vector of unknown parameters in H, G, and F. Estimation of θ using observations $(y_t, 0 \leqslant t \leqslant T)$ is known as system identification. It appears that this problem of system identification was first considered by Balakrishnan [2], who proved the weak local consistency of the maximum likelihood estimator (m.l.e.) under suitable regularity and identifiability assumptions. Later Bagchi and Borkar [1] showed the strong global consistency of the m.l.e. for a slightly more general model. In their case the signal process could be an infinite dimensional process of the following kind:

$$x_t = \int_0^t S_{t-s} D \, dW_1(s),$$

where S_t, $t \geqslant 0$, is a strongly continuous semigroup with generator A on a separable Hilbert space $\mathbf{H}$, W_1 is a Brownian motion on a separable Hilbert space $\mathbf{K}$, and D is a bounded linear operator from $\mathbf{K}$ to $\mathbf{H}$. The observation process y_t, however, is finite dimensional and satisfies the following SDE:

$$y_t = \int_0^t C x_s \, ds + W_2(t)$$

where C is a bounded linear operator from $\mathbf{H}$ to R^q and W_2 is R^q valued Brownian motion independent of W_1. The vector of unknown system parameters θ is assumed to be a point from a compact set in R^k.

It is important to note that even if the signal process is assumed to be finite dimensional their proof cannot be easily extended to obtain either the rate of convergence or the asymptotic normality of the m.l.e. Also, in addition to standard smoothness (with respect to θ), stability and controllability assumptions on A, D, and C, the smoothness of the error covariance operator P and uniform growth requirements on the semigroup Y_t, a perturbation of S_t introduced by them, are assumed rather than proved (see Bagchi and Borkar [1, (3.1a), (3.1b), p. 210). These assumptions are automatically satisfied in our case. No example of an infinite dimensional system satisfying the above mentioned assumptions is given.

It was Kutoyants [8] who first considered the question of asymptotic normality of the m.l.e. in this setting. However, he only considered the following special case of the model in (1):

$$\begin{aligned} dx_t &= -\beta x_t \, dt + dW_1(t), & x_0 &= 0 \\ dy_t &= \alpha x_t \, dt + dW_2(t), & y_0 &= 0; \quad \alpha, \beta \in (a, b), a > 0. \end{aligned} \tag{2}$$

(All the processes involved in (2) are assumed to be one-dimensional). In the above model, *when β is a known constant*, Kutoyants obtained a large deviation inequality for the m.l.e. of α which in turn implies the strong consistency, asymptotic normality, and the convergence of moments. In fact, without mentioning so explicitly, the "local asymptotic normality" (LAN) property of the model is also established, thus proving the efficiency of the m.l.e. in a large class of estimators (see, for instance, Basawa and Scott [5, Chap. 2]).

Here we extend this result to the m.l.e. of the bivariate parameter $\theta = (\alpha, \beta)$. It should be emphasized that Kutoyants's technique cannot be applied to this bivariate estimation problem (not even for the univariate estimation of β when α is a known constant). On the other hand, it will be clear that the method we have used can be applied without any major modification to the general model considered in (1) if, besides identifiability, the following standard conditions are satisfied: (i) The parameter space Θ is an open, *bounded* subset of R^k. (ii) F, G, H are continuously differentiable with respect to θ. (iii) For every θ the pair (F, H) is completely observable and the pair (F, G) is completely controllable. (iv) The eigenvalues of the matrix F lie in the *open* left-half of the complex plane (i.e., F is stable).

The main results along with the necessary notation are given in Section 2 and the proofs are in Section 3.

2. Notation and Statements of Results

From now on, unless mentioned otherwise, the signal and observation processes x_t, y_t, $t \geq 0$; will refer to the solution of the SDE in (2). Also assume that the bivariate parameter $\theta = (\alpha, \beta)$ is an element of $\Theta = (a, b) \times (a, b)$, $a > 0$, $b < \infty$. The letter C (with or without a subscript) will denote a positive constant *independent* of T (the time parameter); it need not be the same in two different expressions.

For $0 \leq t$ let $\hat{x}_t$ be the conditional expectation of x_t given the observations up to time t, i.e.,

$$\hat{x}_t = E(x_t \mid \mathbf{F}_t^y), \tag{3}$$

where $\mathbf{F}_t^y$ is the σ-field generated by $\{y_s, 0 \leq s \leq t\}$ and all the $\mathbf{P}$-null sets; furthermore, let

$$dv_t = dy_t - \alpha \hat{x}_t \, dt. \tag{4}$$

Then it is well known that (v_t) is a Wiener process and, moreover, the process $(\hat{x}_t)$ satisfies the SDE

$$d\hat{x}_t = -\beta \hat{x}_t \, dt + \alpha a_t \, dv_t, \qquad \hat{x}_0 = 0;$$

where a_t is the (unique) solution of a (deterministic) differential equation known as the Riccati equation. More precisely, a_t is the solution of the following nonlinear differential equation:

$$\frac{d}{dt} a_t = 1 - \alpha^2 a_t^2 - 2\beta a_t, \qquad a_0 = 0.$$

It is also known that as $t \to \infty$, $a_t \to a_\theta$, $a_\theta = (-\beta + \sqrt{\alpha^2 + \beta^2})/\alpha^2$. All these facts can be found in Liptser and Shiryayev [9, Vol. II, 16.2]. Now we make a simplifying assumption commonly made in the literature (see, e.g., Kutoyants [8, p. 103]). We assume that the system has reached the steady state, i.e., we assume that $\hat{x}_t$ satisfies the SDE:

$$d\hat{x}_t = -\beta \hat{x}_t\, dt + \alpha a_\theta\, dv_t, \qquad \hat{x}_0 = 0. \tag{5}$$

Then from (4) and (5), it is easy to verify that

$$\hat{x}_t = \alpha a_\theta \int_0^t e^{-b_\theta (t-s)}\, dy_s, \tag{6}$$

where $b_\theta = \sqrt{\alpha^2 + \beta^2}$.

Let C_T denote the space of real valued continuous functions defined on $[0, T]$ endowed with the sup-norm topology and let $\mathbf{C}_\mathrm{T}$ be the σ-field of Borel sets in C_T. Furthermore, let P_T^θ denote the measure induced by the paths $(y_s, 0 \leqslant s \leqslant T)$ on $(C_T, \mathbf{C}_\mathrm{T})$.

Then in view of the relation

$$dy_t = \alpha \hat{x}_t\, dt + dv_t^\theta \tag{7}$$

and the fact that v_t^θ is a Wiener process, it follows that P_T^θ is equivalent to the standard Wiener measure μ_W defined on $(C_T, \mathrm{C}_\mathrm{T})$. Furthermore the density or the likelihood function of the data $(y_s, 0 \leqslant s \leqslant T)$ at θ is given by

$$\frac{dP_\theta^T}{d\mu_W}(y) = \exp\left(\int_0^T \alpha \hat{x}_t\, dy_t - 1/2 \int_0^T \alpha^2 \hat{x}_t^2\, dt\right). \tag{8}$$

The verification of this fact is quite straightforward; for example, it follows from the combination of two results (Theorems 7.3.1 and 7.3.2, p. 176) from Kallianpur [7].

Let $l(\theta) = l(\theta, y)$ be the *loglikelihood function* of the data. For each $T > 0$ define a random function $Z_T(u)$ as

$$U_T = \sqrt{T}(\Theta - \theta) = \sqrt{T}(a - \alpha, b - \alpha) \times (a - \beta, b - \beta). \tag{9}$$

For $u \in U_T$,

$$Z_T(u) = l(\theta + u/\sqrt{T}) - l(\theta). \tag{10}$$

(Clearly Z_T and U_T depend on θ but this dependence is suppressed for notational convenience.) Now we can state the main result of this section.

THEOREM 2.1. *The random functions $Z_T(u)$ have continuous sample paths and satisfy the following three properties*:

(I) $$\sup_{\substack{|u|,|w| \leqslant M \\ u, w \in U_T}} E_\theta |Z_T(u) - Z_T(w)|^4 \leqslant CM^4 |u-w|^4$$

(II) *For $u \in U_T$ and T large, $T > T_0$,*

$$E_\theta \exp(1/4 Z_T(u)) \leqslant \exp(-C' |u|^2).$$

(III) *As $T \to \infty$ the finite dimensional distributions of $Z_T(u)$ converge to the finite dimensional distributions of $Z(u)$, where for $u \in R^2$,*

$$Z(u) = u' Y_\theta - 1/2 u' \Sigma_\theta u;$$

here Y_θ is a zero mean bivariate normal variable with invertible covariance matrix Σ_θ and u' denotes the transpose of the vector u. (Note that $Z(u)$ is a real valued, continuous random function defined on R^2 which attains its maximum at a unique (random) point $\Sigma_\theta^{-1} Y_\theta$.)

The elements of the matrix Σ_θ are described as follows: Suppose that H_θ, G_θ, and R_θ are trace class operators (R_θ is self-adjoint) defined on $L^2[0, T]$ with kernels:

$$\begin{aligned} H_\theta(t,s) &= \frac{\alpha}{\sqrt{\alpha^2+\beta^2}} [1 + \beta(t-s) - \sqrt{\alpha^2+\beta^2}(t-s)]\, e^{-\sqrt{\alpha^2+\beta^2}(t-s)}, \\ G_\theta(t,s) &= \left(-1 + \frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right) [1 + \beta(t-s)]\, e^{-\sqrt{\alpha^2+\beta^2}(t-s)}, \end{aligned} \tag{11}$$

if $0 \leqslant s \leqslant t \leqslant T$ and equal to zero otherwise;

$$R_\theta(t,s) = \frac{e^{-\beta|t-s|} - e^{-\beta(t+s)}}{2\beta}, \qquad 0 \leqslant s, t \leqslant T. \tag{12}$$

Let H^* and G^* denote the corresponding adjoint operators. Then it is easy to verify the following (see Lemma 3.2 in the next section):

(i) $\lim_{T\to\infty} 1/T \operatorname{trace}[HRH^* + HH^*] = \sigma_1^2 < \infty$,

(ii) $\lim_{T\to\infty} 1/T \operatorname{trace}[GRG^* + GG^*] = \sigma_2^2 < \infty$, and

(iii) $\lim_{T\to\infty} 1/T \operatorname{trace}[HRG^* + HG^*] = \sigma_{12} < \infty$.

Then Σ_θ denotes the 2×2 symmetric matrix with $\Sigma_{11} = \sigma_1^2$, $\Sigma_{22} = \sigma_2^2$, and $\Sigma_{12} = \sigma_{12}$. It is easy to check that Σ is strictly positive definite.

The above result verifies all the conditions of two results in Ibragimov and Hasminski [6, Theorem 10.1, Chap. 1, and Corollary 1.1 in Chap. 3]. The conditions are slightly modified in the sense that loglikelihood ratios are used rather than the likelihood ratios. The equivalence of these two versions is not very difficult to see (see, e.g., Selukar [11, Theorem 2.2, p. 22]). The property (III) is the uniform LAN property. The next theorem which is the result of practical interest follows as a corollary.

Let $\hat{\theta}_T(y)$ be the m.l.e. of θ based on observations (y_t), $0 \leqslant t \leqslant T$; i.e., the maximum of the likelihood ratio is attained at $\hat{\theta}_T$.

THEOREM 2.2. *The m.l.e. of θ has the following properties*:

(i) *$\hat{\theta}_T$ is a strongly consistent and asymptotically efficient estimator of θ.*

(ii) *As T tends to infinity the distribution of $\sqrt{T}(\hat{\theta}_T - \theta)$ converges to the normal distribution with zero mean and covariance matrix Σ^{-1}. Furthermore, for every $p > 0$, the pth moment of the norm of $\sqrt{T}(\hat{\theta}_T - \theta)$ converges to the pth moment of the norm of this normal variable.*

(iii) *For $h > 0$ and large T, $T > T_0$,*

$$P_\theta^T\{\sqrt{T}\,|(\hat{\theta}_T - \theta)| > h\} \leqslant B_0 \exp(-b_0 h^2),$$

where B_0, $b_0 > 0$ are constants.

The efficiency in (i) above is to be understood in the sense of the second result from Ibragimov and Hasminski mentioned above. In particular, the m.l.e. is efficient under the quadratic loss.

3. PROOF

It suffices to verify the conditions of Theorem 2.1 which will be done using several lemmas. The first two lemmas are technical. Let (Y_t), $0 \leqslant t \leqslant T$, be a mean square integrable *zero* mean Gaussian process with covariance $R(t, s)$. Let R be the corresponding covariance operator. It is well known that R is a self-adjoint, non-negative definite trace class operator and

$$\operatorname{trace}(R) = \int_0^T R(t, t)\, dt = E \int_0^T Y_t^2\, dt. \tag{13}$$

LEMMA 3.1. (i) *For all* $k \geqslant 1$,

$$E\left(\int_0^T Y_t^2\,dt\right)^k \leqslant k^k\left[E\left(\int_0^T Y_t^2\,dt\right)\right]^k.$$

(ii)

$$E\exp\left(-\int_0^T Y_t^2\,dt\right) \leqslant \exp\left(-\frac{\operatorname{trace}(R)}{1+2\,\|R\|}\right),$$

where $\|R\|$ *is the operator norm of* R.

Proof. Let $\{f_j\}$ be the system of eigenfunctions of R and λ_j's be the corresponding eigenvalues then, using the Karhunen–Loeve expansion of (Y_t), we obtain

$$\int_0^T Y_t^2\,dt = \sum_{j=1}^{\infty} \lambda_j X_j^2 \qquad \text{and} \qquad E\int_0^T Y_t^2\,dt = \sum_{j=1}^{\infty} \lambda_j = \lambda^* \qquad \text{say.} \tag{14}$$

The coefficients (X_j) are i.i.d. $N(0, 1)$ random variables. From (14)

$$E\left(\int_0^T Y_t^2\,dt\right)^k = E\left(\sum_{j=-}^{\infty} \lambda_j X_j^2\right)^k = (\lambda^*)^k\, E\left[(\lambda^*)^{-1} \sum_{j=1}^{\infty} \lambda_j X_j^2\right]^k.$$

Then the application of Jensen's inequality in the square bracket and the i.i.d. nature of (X_j)s imply

$$E\left(\int_0^T Y_t^2\,dt\right)^k \leqslant (\lambda^*)^{k-1}\, E\left(\sum_{j=1}^{\infty} \lambda_j X_j^{2k}\right) = (\lambda^*)^k\, EX_1^{2k}.$$

Moreover, because X_1 is $N(0, 1)$, using *Stirling's* approximation (see Rao [10, p. 59]),

$$E\left(\int_0^T Y_t^2\,dt\right)^k \leqslant (\lambda^*)^k \frac{(2k)!}{2^k k!} \leqslant k^k (\lambda^*)^k.$$

This, together with (14) proves (i). The last inequality is easy to deduce from (14) and the fact that if X is a $N(0, 1)$ variable then, for $\beta < \frac{1}{2}$, $E\exp(\beta X^2) = (1-2\beta)^{-1/2}$. ▮

For $\lambda > 0$ and m a non-negative integer, let L be an integral operator defined on $L^2[0, T]$ with kernel $L(t, s)$ given by

$$\begin{aligned} L(t, s) &= (t-s)^m\, e^{-\lambda(t-s)}, && 0 \leqslant s \leqslant t \leqslant T \\ &= 0 && \text{otherwise.} \end{aligned}$$

That is, for $f \in L^2[0, T]$,

$$(Lf)(t) = \int_0^T L(t, s)\, f(s)\, ds. \tag{15}$$

L is a special case of a *Volterra* operator. Let $\|L\|$ be the operator norm of L and L^* denote the adjoint of L. Then LL^* is a self-adjoint trace class operator and it is easy to check that

$$\begin{aligned}\operatorname{trace}(LL^*) &= \int_0^T \int_0^s L^2(t, s)\, ds\, dt \\ &= T \int_0^T u^{2m} e^{-2\lambda u}\, du - \int_0^T u^{2m+1} e^{-2\lambda u}\, du.\end{aligned}$$

The conclusions of the following lemma are simple consequences of the above expression for the trace of LL^* and the definition of L.

LEMMA 3.2. (i) $\|L\| \leqslant \sqrt{2}(\Gamma(m+1)/\lambda^{m+1})$

(ii) $\operatorname{trace}(LL^*) \leqslant (\Gamma(2m+1)/(2\lambda)^{2m+1})\, T$

(iii) *For large* T, $\operatorname{trace}(LL^*) \geqslant (\Gamma(2m+1)/(2\lambda)^{2m+1})\, T/2$

(iv) $\lim_{T\to\infty} 1/T \operatorname{trace}(LL^*) = \Gamma(2m+1)/(2\lambda)^{2m+1}$

(v) *If L_1 and L_2 are two Volterra operators of the above type (i.e., with Gamma kernels) then the operator $L_1 L_2$ is also a Volterra operator which is a finite linear combination of the operators of the above type.*

Proof. First observe that $L(t, s) \geqslant 0$ and for each t, $\int_0^T L(t, s)\, ds \leqslant \Gamma(\alpha+1)/\lambda^{\alpha+1}$. Hence

$$(Lf)^2(t) = \left(\int_0^t L(t, s) f(s)\, ds\right)^2 \leqslant \Gamma(\alpha+1)/\lambda^{\alpha+1} \left(\int_0^t L(t, s) f^2(s)\, ds\right).$$

Therefore,

$$\begin{aligned}\|Lf\|^2 &= \int_0^T (Lf)^2(t)\, dt \\ &\leqslant \Gamma(\alpha+1)/\lambda^{\alpha+1} \int_0^T \int_0^t L(t, s) f^2(s)\, ds\, dt \\ &\leqslant 2(\Gamma(\alpha+1)/\lambda^{\alpha+1})^2\, \|f\|^2.\end{aligned}$$

This proves the first assertion. The others are even simpler to prove so the proofs are omitted. ∎

Remark 3.1. Note that the operator norm of L has a bound *independent* of T and the trace of LL^* is of the same order as that of T. It is obvious that statements of the above lemma can also be obtained for an operator M which is a linear combination of L_is.

Let $\hat{X}_\theta(t, y) = \alpha \hat{x}_t$. Then, using (6),

$$\hat{X}_\theta(t, y) = (-\beta + \sqrt{\alpha^2 + \beta^2}) \int_0^t e^{-\sqrt{\alpha^2+\beta^2}(t-s)}\, dy_s. \tag{16}$$

From (8) and (10)

$$l(\theta, y) = \int_0^T \hat{X}_\theta(t, y)\, dy_t - 1/2 \int_0^T \hat{X}_t^2(\theta, y)\, dt.$$

Therefore,

$$\begin{aligned} Z_T(u) &= l\left(\theta + \frac{u}{\sqrt{T}}\right) - l(\theta) \\ &= \int_0^T \left[\hat{X}_t\left(\theta + \frac{u}{\sqrt{T}}, y\right) - \hat{X}_t(\theta, y)\right] dy_t \\ &\quad - 1/2 \int_0^T \left[\hat{X}_t^2\left(\theta + \frac{u}{\sqrt{T}}, y\right) - \hat{X}_t^2(\theta, y)\right] dt. \end{aligned}$$

If we complete the square in the second term of the RHS and rearrange the terms we obtain

$$\begin{aligned} Z_T(u) &= \int_0^T \left[\hat{X}_t\left(\theta + \frac{u}{\sqrt{T}}, y\right) - \hat{X}_t(\theta, y)\right] dv_t^\theta \\ &\quad - 1/2 \int_0^T \left[\hat{X}_t\left(\theta + \frac{u}{\sqrt{T}}, y\right) - \hat{X}_t(\theta, y)\right]^2 dt, \end{aligned} \tag{17}$$

where $dv_t^\theta = dy_t - \hat{X}_t(\theta, y)\, dt$. Recall that under P_θ^T, v_t^θ is a standard Wiener process.

From now on, unless stated otherwise, all the expectations are taken w.r.t. the true probability measure P_θ^T. Also, in order to simplify the notation we may sometimes write, $Z(u) = Z_T(u)$, $\hat{X}_t(u) = \hat{X}_t(\theta + u/\sqrt{T}, y)$ and $dv_t = dv_t^\theta$. The next lemma verifies the first assertion of Theorem 3.1.

LEMMA 3.3. *For $u, w \in U_T$, $|u|, |w| \leqslant M$,*

$$E(Z_T(u) - Z_T(w))^4 \leqslant CM^4\, |u - w|^4$$

where C depends only on a and b. Recall that $\Theta = (a, b) \times (a, b)$.

Proof. From (17),

$$\begin{aligned} Z_T(u) - Z_T(w) &= \int_0^T [\hat{X}_t(u) - \hat{X}_t(w)]\, dv_t \\ &\quad - 1/2 \int_0^T \{[\hat{X}_t(u) - \hat{X}_t(0)]^2 - [\hat{X}_t(w) - \hat{X}_t(0)]^2\}\, dt \\ &= \text{TERM1} - 1/2\, \text{TERM2} \qquad \text{say.} \end{aligned}$$

Then,

$$E(Z(u)-Z(w))^4 \leqslant 16\{E(\text{TERM1})^4 + 1/16\, E(\text{TERM2})^4\}. \tag{18}$$

Consider $E(\text{TERM1})^4$:

$$\begin{aligned} E(\text{TERM1})^4 &= E\left\{\int_0^T [\hat{X}_t(u)-\hat{X}_t(w)]\, dv_t\right\}^4 \\ &\leqslant 16E\left\{\int_0^T [\hat{X}_t(u)-\hat{X}_t(w)]^2\, dt\right\}^2 \\ &\leqslant 64\left\{\int_0^T E[\hat{X}_t(u)-\hat{X}_t(w)]^2\, dt\right\}^2. \end{aligned} \tag{19}$$

(The first step follows from Burkholder's martingale inequality and the fact that (v_t) is a Wiener process under P_θ^T. The last step is a consequence of Lemma 3.1(i).)

Note that from (16),

$$\hat{X}_t(u)-\hat{X}_t(w) = \int_0^t (L_u - L_w)(t, s)\, dy_s \tag{20}$$

where $L_u = L_{\theta + u/\sqrt{T}}$ is given as

$$\begin{aligned} L_u(t, s) &= A_u e^{-B_u(t-s)}, && 0 \leqslant s \leqslant t \leqslant T, \\ &= 0, && \text{otherwise}, \end{aligned} \tag{21}$$

$$A_u = A_{\theta + u/\sqrt{T}} = -\left(\beta + \frac{u_2}{\sqrt{T}}\right) + \left\{\left(\beta + \frac{u_2}{\sqrt{T}}\right)^2 + \left(\alpha + \frac{u_1}{\sqrt{T}}\right)^2\right\}^{1/2},$$

and

$$B_u = B_{\theta + u/\sqrt{T}} = \left\{\left(\beta + \frac{u_2}{\sqrt{T}}\right)^2 + \left(\alpha + \frac{u_1}{\sqrt{T}}\right)^2\right\}^{1/2}.$$

(Note that $u = (u_1, u_2)$ is a point in $U_T \subset R^2$).

Now recall that (see (2))

$$dy_t = \alpha x_t\, dt + dW_2(t), \tag{22}$$

where the "signal" (x_t) and the observation "noise" $(W_2(t))$ are independent. Let $R(t, s)$ denote the covariance function of x_t, i.e., $R(t, s) = E(x_t x_s)$. Then, since x_t is the familiar Ornstein–Uhlenbeck process, it follows that

$$R(t, s) = \frac{e^{-\beta|t-s|} - e^{-\beta(t+s)}}{2\beta}. \tag{23}$$

If R is the integral operator with the kernel $R(t, s)$. Then it is easy to check that $R = R_1 R_1^*$, where R_1 is the Volterra operator with kernel

$$R_1(t, s) = e^{-\beta(t-s)} \qquad \text{for} \quad 0 \leqslant s \leqslant t \leqslant T.$$

Therefore, by Lemma 3.2, R is a trace class operator and

$$\|R\| \leqslant 2/\beta^2. \tag{24}$$

From (20), (22), and the independence of (x_t) and $(W_2(t))$ it follows that

$$\begin{aligned}
&\int_0^T E[\hat{X}_t(u) - \hat{X}_t(w)]^2 \, dt \\
&\quad = \int_0^T \int_0^t \int_0^t (L_u - L_w)(t, v) \, R(v, s)(L_u - L_w)(t, s) \, ds \, dv \, dt \\
&\quad\quad + \int_0^T \int_0^t (L_u - L_w)^2 \, (t, s) \, ds \, dt \\
&\quad = \text{Trace}\{(L_u - L_w) \, R(L_u - L_w)^*\} + \text{Trace}\{(L_u - L_w)(L_u - L_w)^*\},
\end{aligned}$$

where L_u denotes the integral operator with kernel $L_u(t, s)$ and $(L_u)^*$, its adjoint. Next, using (24) and the fact that for any two non-negative definite trace class operators J_1 and J_2,

$$\text{Trace}(J_1 J_2) \leqslant \text{Trace}(J_1) \cdot \|J_2\|, \tag{25}$$

where $\|J_2\| =$ the operator norm of J_2, it follows that

$$\begin{aligned}
\int_0^T E[\hat{X}_t(u) - \hat{X}_t(w)]^2 \, dt &\leqslant (1 + 2/\beta^2) \, \text{Trace}(L - u - L_w)(L_u - L_w)^* \\
&= (1 + 2/\beta^2) \int_0^T \int_0^t (L_u - L_w)^2 \, (t, s) \, ds \, dt \\
&\leqslant C\{(u_1 - w_1)^2 + (u_2 - w_2)^2\}. \qquad (26)
\end{aligned}$$

The last step is obtained using the Taylor expansion and Lemma 3.2 as

$$\begin{aligned}
L_u(t, s) - L_w(t, s) = (u_1 - w_1)/\sqrt{T} \, \frac{\delta}{\delta\alpha} \, L_{\mathbf{u}}(t, s) \\
+ (u_2 - w_2)/\sqrt{T} \, \frac{\delta}{\delta\beta} \, L_{\mathbf{u}}(t, s),
\end{aligned}$$

where $\mathbf{u} \in (\alpha + u_1/\sqrt{T}, \alpha + w_1/\sqrt{T}) \times (\beta + u_2/\sqrt{T}, \beta + w_2/\sqrt{T}) \subset \Theta$. ($\mathbf{u}$ may depend on t and s.) Therefore,

$$(L_u - L_w)^2\,(t, s) \leqslant 4/T[(u_1 - w_1)^2 + (u_2 - w_2)^2]$$
$$\times \sup_{\theta \in \Theta} \left\{ \left(\frac{\delta}{\delta\alpha} L_\theta\right)^2 + \left(\frac{\delta}{\delta\beta} L_\theta\right)^2 (t, s) \right\}$$
$$\leqslant 4/T[(u_1 - w_1)^2 + (u_2 - w_2)^2]$$
$$\times \{(C_1 + C_2(t-s) + C_3(t-s)^2) \exp(-C_4(t-s))\};$$

C_i's are positive constants which depend on $a > 0$ and b only. (Recall that $\Theta = (a, b) \times (a, b)$). Therefore from Lemma 3.2(ii) and Remark 3.1,

$$(1 + 2/\beta^2) \int_0^T \int_0^t (L_u - L_w)^2\,(t, s)\, ds\, dt$$
$$\leqslant 4/T[(u_1 - w_1)^2 + (u_2 - w_2)^2]$$
$$\times \int_0^T \int_0^t \{(C_1 + C_2(t-s) + C_3(t-s)^2) \exp(-C_4(t-s))\}\, ds\, dt$$
$$\leqslant C\, \|u - w\|^2/T \times T.$$

Thus, from (19) and (26) it follows that

$$E(\text{TERM1})^4 \leqslant C\, \|u - w\|^4. \tag{27}$$

Now consider TERM2:

$$(\text{TERM2})^4 = \left(\int_0^T \{[\hat{X}_t(u) - \hat{X}_t(0)]^2 - [\hat{X}_t(w) - \hat{X}_t(0)]^2\}\, dt \right)^2$$
$$\leqslant \left(\int_0^T ([\hat{X}_t(u) - \hat{X}_t(0)] + [\hat{X}_t(w) - \hat{X}_t(0)])^2\, dt \right)^2$$
$$\times \left(\int_0^T (\hat{X}_t(u) - \hat{X}_t(w))^2\, dt \right)^2.$$

We have used the identity $(x^2 - y^2) = (x + y)(x - y)$ and then applied the Cauchy–Schwartz inequality. Again applying the Cauchy–Schwartz inequality, Lemma 3.1(i) and from calculations similar to TERM1 we have

$$E(\text{TERM2})^4 \leqslant \left\{ E\left(\int_0^T ([\hat{X}_t(u) - \hat{X}_t(0)] + [\hat{X}_t(w) - \hat{X}_t(0)])^2\, dt \right)^4 \right.$$
$$\left. \times E\left(\int_0^T (\hat{X}_t(u) - \hat{X}_t(w))^2\, dt \right)^4 \right\}^{1/2}$$
$$\leqslant \{16(C4^4[\|u - 0\|^8 + \|w - 0\|^8]) \times C\, \|u - w\|^8\}^{1/2}.$$

Thus, since $\|u\|$, $\|w\| < M$,

$$E(\text{TERM2})^4 < CM^4 \|u-w\|^4. \tag{28}$$

The statement of the lemma follows from (18), (27), and (28). ∎

LEMMA 3.4. *For $u \in U_T$ and T large,*

$$E \exp(1/4 Z_T(u)) \leqslant \exp(-C' \|u\|^2),$$

where $C' > 0$, depends on a and b only.

Proof. From (17) and an application of the Cauchy–Schwartz inequality,

$$\begin{aligned} E \exp(1/4 Z_T(u)) &= E \exp\left(1/4 \int_0^T [\hat{X}_t(u) - \hat{X}_t(0)]\, dv_t^\theta \right.\\ &\qquad \left. - 1/8 \int_0^T [\text{over } T\hat{X}_t(u) - \hat{X}_t(0)]^2\, dt\right) \\ &\leqslant E \left\{ \exp\left(1/2 \int_0^T [\hat{X}_t(u) - \hat{X}_t(0)]\, dv_t^\theta \right.\right.\\ &\qquad \left. - 1/8 \int_0^T [\hat{X}_t(u) - \hat{X}_t(0)]^2\, dt\right) \\ &\qquad \left. \times E \exp\left(-1/8 \int_0^T [\hat{X}_t(u) - \hat{X}_t(0)]^2\, dt\right)\right\}^{1/2}. \end{aligned}$$

Note that the first term of the product in the bracket is a density (w.r.t. P_θ^T). Therefore,

$$E \exp(1/4 Z_T(u)) \leqslant \left\{(1) \cdot E \exp\left(-1/8 \int_0^T [\hat{X}_t(u) - \hat{X}_t(0)]^2\, dt\right)\right\}^{1/2}.$$

The RHS above can be easily bounded by applying Lemma 3.1(ii), since $\xi_t = [\hat{X}_t(u) - \hat{X}_t(0)]$ is a Gaussian process. Let $F(t, s)$ be the covariance function of this process and F be the corresponding covariance operator. Then,

$$E \exp(1/4 Z_T(u)) \leqslant \exp\left(-1/16 \frac{\text{trace } F}{1 + 2\|F\|}\right).$$

From the calculations made in order to bound $E(\text{TERM1})^2$ in Lemma 3.3 (with $u = u$ and $w = 0$) we can make the following observations:

(i) $F=(L_u-L_0)\,R(L_u-L_0)^*+(L_u-L_0)(L_u-L_0)^*$

(ii) $\|F\| \leqslant C'$

(iii) trace F = trace$(L_u - L_0)\,R(L_u - L_0)^*$ + trace$(L_u - L_0)(L_u - L_0)^*$ and, since the first term is always positive, trace $F \geqslant$ trace$(L_u-L_0)(L_u-L_0)^* \geqslant C\,\|u\|^2$.

In order to see (iii) first recall that

$$\operatorname{trace}(L_u-L_0)(L_u-L_0)^*=\int_0^T\int_0^t (L_u-L_0)^2\,(t,s)\,ds\,dt.$$

Now

$$L_u(t,s)-L_0(t,s)=(u_1)/\sqrt{T}\,\frac{\delta}{\delta\alpha}\,L_{\mathbf{u}}(t,s)$$
$$+(u_2)/\sqrt{T}\,\frac{\delta}{\delta\beta}\,L_{\mathbf{u}}(t,s),$$

where $\mathbf{u}\in(\alpha+u_1/\sqrt{T},\alpha)\times(\beta+u_2/\sqrt{T},\beta)\subset\Theta$. ($\mathbf{u}$ may depend on t and s.) Therefore,

$$\begin{aligned}(L_u-L_0)^2\,(t,s)&\\ &\geqslant 4/T[(u_1)^2+(u_2)^2]\left\{\inf_{\theta\in\Theta}\left(\frac{\delta}{\delta\alpha}L_\theta\right)^2+\inf_{\theta\in\Theta}\left(\frac{\delta}{\delta\beta}L_\theta\right)^2(t,s)\right\}\\ &\geqslant 4/T[(u_1)^2+(u_2)^2]\{(C_1+C_2(t-s)+C_3(t-s)^2)\exp(-C_4(t-s))\};\end{aligned}$$

C_i's are positive constants which depend on $a>0$ and b only. (Recall that $\Theta=(a,b)\times(a,b)$.) Therefore, from Lemma 3.2(iii) and Remark 3.1,

$$\begin{aligned}&\int_0^T\int_0^t (L_u-L_0)^2\,(t,s)\,ds\,dt\\ &\qquad\geqslant 4/T\,\|u\|^2\int_0^T\int_0^t (C_1+C_2(t-s)+C_3(t-s)^2)\\ &\qquad\quad\times\exp(-C_4(t-s))\,ds\,dt\\ &\qquad\geqslant 4/T\,\|u\|^2\,C\cdot T=C\cdot\|u\|^2.\end{aligned}$$

for large T. Thus (iii) follows and we finally obtain that, for large T,

$$E\exp(1/4Z_T(u))\leqslant\exp(-C'\,\|u\|^2),$$

where $C' > 0$, depends on a and b only. This is the statement of the lemma. ∎

For $u = (u_1, u_2) \in R^2$ let

$$Z_\theta(u) = u' Y_\theta - 1/2 u' \Sigma u, \tag{29}$$

where Y_θ is a zero mean bivariate normal variable with covariance matrix Σ.

Now we will show that finite dimensional distributions of $Z_T(u)$ converge to the finite dimensional distributions of $Z_\theta(u)$ as $T \to \infty$. This is done in two steps. First we define random functions $Z_T^*(u)$ such that for any fixed $u \in U_T$,

$$E(Z_T(u) - Z_T^*(u))^2 \to 0 \qquad \text{as} \quad T \to \infty. \tag{30}$$

Next we show that finite dimensional distributions of $Z_T^*(u)$ converge to the finite dimensional distributions of $Z_\theta(u)$. These two steps are clearly sufficient for our purpose.

Let us begin the first step: For $u \in U_T$ let

$$Z_T^*(u) = \int_0^T [u_1/\sqrt{T}\, h_t(\theta) + u_2/\sqrt{T}\, g_t(\theta)]\, dv_t^\theta$$

$$- 1/2 \int_0^T [u_1/\sqrt{T}\, h_t(\theta) + u_2/\sqrt{T}\, g_t(\theta)]^2\, dt,$$

where

$$h_t(\theta) = \int_0^t \frac{\delta}{\delta\alpha} L_\theta(t, s)\, dy_s$$

$$= \int_0^t \alpha/\sqrt{\alpha^2 + \beta^2}[1 + \beta(t - s) - \sqrt{\alpha^2 + \beta^2}(t - s)]\, e^{-\sqrt{\alpha^2 + \beta^2}(t - s)}\, dy_s$$

$$= \int_0^t H_\theta(t, s)\, dy_s$$

and

$$g_t(\theta) = \int_0^t \frac{\delta}{\delta\beta} L_\theta(t, s)\, dy_s$$

$$= \int_0^t (-1 + \beta/\sqrt{\alpha^2 + \beta^2})[1 + \beta(t - s)]\, e^{-\sqrt{\alpha^2 + \beta^2}(t - s)}\, dy_s$$

$$= \int_0^t G_\theta(t, s)\, dy_s. \tag{31}$$

(Recall the two integral operators H_θ, G_θ with kernels $H_\theta(t, s)$ and $G_\theta(t, s)$ defined in Section 2 (see (11).) It turns out that $H_\theta(t, s) = (\delta/\delta\alpha)\, L_\theta(t, s)$ and $G_\theta(t, s) = (\delta/\delta\beta)\, L_\theta(t, s)$.)

LEMMA 3.5. *For* $u \in U_T$,

$$E(Z_T(u) - Z_T^*(u))^2 \to 0 \qquad as \quad T \to \infty.$$

Proof. This proof is almost identical to that of Lemma 3.3. From (17) and (31)

$$Z_T(u) - Z_T^*(u) = \text{TERM } A - 1/2 \text{ TERM } B, \tag{32}$$

where

$$\text{TERM } A = \int_0^T [\hat{X}_t(u) - \hat{X}_t(0) - (u_1/\sqrt{T}\, h_t(\theta) + u_2/\sqrt{T}\, g_t(\theta))]\, dv_t^\theta$$

and

$$\text{TERM } B = \int_0^T ([\hat{X}_t(u) - \hat{X}_t(0)]^2 - [u_1/\sqrt{T}\, h_t(\theta) + u_2/\sqrt{T}\, g_t(\theta)]^2)\, dt.$$

Let us consider $E(\text{TERM } A)^2$. Since (v_t) is a Wiener process,

$$\begin{aligned} &E(\text{TERM } A)^2 \\ &\quad = E\int_0^T [\hat{X}_t(u) - \hat{X}_t(0) - (u_1/\sqrt{T}\, h_t(\theta) + u_2/\sqrt{T}\, g_t(\theta))]^2\, dt \\ &\quad = E\int_0^T \left[\int_0^t (L_u - L_0 - (u_1/\sqrt{T}\, H + u_2/\sqrt{T}\, G))(t, s)\, dy_s\right]^2 dt \\ &\quad \leqslant (1 + 2/\beta^2) \int_0^T \int_0^t (L_u - L_0 - (u_1/\sqrt{T}\, H + u_2/\sqrt{T}\, G))^2\, (t, s)\, ds\, dt. \end{aligned}$$

The last step is obtained using the same arguments as in (17)–(26). Now note that, using the Taylor expansion,

$$\begin{aligned} &(L_u - L_0 - (u_1/\sqrt{T}\, H + u_2/\sqrt{T}\, G))(t, s) \\ &\quad = \left(u_1^2/2T \frac{\delta^2}{\delta\alpha^2} L_{\mathbf{u}} + u_2^2/2T \frac{\delta^2}{\delta\beta^2} L_{\mathbf{u}} + u_1 u_2/2T \frac{\delta^2}{\delta\alpha\, \delta\beta} L_{\mathbf{u}}\right)(t, s) \end{aligned}$$

for some $\mathbf{u} \in (\theta, \theta + u/\sqrt{T}) \subset \Theta$. Therefore,

$$\begin{aligned} &(L_u - L_0 - (u_1/\sqrt{T}\, H + u_2/\sqrt{T}\, G))^2\, (t, s) \\ &\quad \leqslant C/T^2 (u_1^2 + u_2^2)^2 \\ &\qquad \times \sup_{\theta \in \Theta} \left[\left(\frac{\delta^2}{\delta\alpha^2} L_\theta\right)^2 + \left(\frac{\delta^2}{\delta\beta^2} L_\theta\right)^2 + \left(\frac{\delta^2}{\delta\alpha\, \delta\beta} L_\theta\right)^2\right](t, s) \\ &\quad \leqslant C\, \|u\|^4/T^2 \left[\sum_{i=0}^4 (t - s)^i\right] \exp(-C_1(t - s)). \end{aligned}$$

(The positive constants C and C_1 depend on a and b only.) Therefore

$$(1+2/\beta^2)\int_0^T\int_0^t (L_u - L_0 - (u_1/\sqrt{T}\,H + u_2/\sqrt{T}\,G))^2\,(t,s)\,ds\,dt$$

$$\leqslant C\,\|u\|^4/T^2 \int_0^T\int_0^t \left[\sum_{i=0}^{4}(t-s)^i\right]\exp(-C_1(t-s))\,ds\,dt$$

$$\leqslant C\,\|u\|^4/T^2\cdot T \to 0 \qquad \text{as} \quad T\to\infty$$

(since $u\in U_T$ is fixed). Thus $E(\text{TERM } A)^2\to 0$ as $T\to\infty$.

The steps to show that $E(\text{TERM } B)^2\to 0$ as $T\to\infty$ are also very similar. This concludes the proof of Lemma 3.5. ∎

Now let us begin the second step: First we state a version of the central limit theorem which is useful for our purpose (see Basawa and Prakasa Rao [4, Theorem 2.1, Appendix 2, p. 405]).

Let $\{W(t),\ t\geqslant 0\}$ denote the standard m-dimensional Brownian motion. Suppose that $F(s)=((f_{kj}(s)))_{n\times m}$ is a random matrix valued function such that its elements $f_{kj}\in H[0,T]$ for all $T>0$. (A random function $f\in H[0,T]$ iff it is adapted to the Wiener filtration and

$$E\int_0^T f^2(t)\,dt<\infty.)$$

Set $\mathbf{f_k}(s)=(f_{k1}(s),\ldots,f_{km}(s))$, $1\leqslant k\leqslant n$.

THEOREM 3.1. *Suppose that the random matrix valued function $F(s)$ satisfies*

$$1/T\int_0^T (\mathbf{f_k}(s))'\,\mathbf{f_j}(s)\,ds\to c_{kj}$$

in probability as $T\to\infty$, where c_{kj}, $1\leqslant k, j\leqslant n$, are finite. Then the distribution of

$$T^{-1/2}\int_0^T F(s)\,dW(s)$$

converges to the normal distribution with mean zero and covariance matrix $C=(c_{kj})$ as $T\to\infty$.

Note that just as $Z_\theta(u)$ (see (29)) we can write

$$Z_T^*(u)=u'Q_T(\theta)-1/2u'\Delta_T(\theta)u,$$

where $Q_T(\theta)$ is a bivariate normal random variable and $\Delta_T(\theta)$ is a 2×2 random symmetric matrix described below:

$$Q_T(\theta)=\left(1/\sqrt{T}\int_0^T h_t(\theta)\,dv_t,\ 1/\sqrt{T}\int_0^T g_t(\theta)\,dv_t\right)',$$

$$\Delta_T(1,1)=1/T\int_0^T h_t^2(\theta)\,dt,\qquad \Delta_T(2,2)=1/T\int_0^T g_t^2(\theta)\,dt,$$

$$\Delta_T(1,2)=1/T\int_0^T g_t(\theta)\,h_t(\theta)\,dt.$$

From the above result, the fact that (v_t) is a Wiener process, and the special forms of $Z_T^*(u)$ and $Z_\theta(u)$, it is clear that we only have to show that $\Delta_T(\theta)$ converges to Σ in probability and $Q_T(\theta)$ converges to Y_θ in distribution. Since the former implies the latter we only have to show that $\Delta_T(\theta)$ converges to Σ in probability.

Let us first show $1/T\int_0^T h_t^2(\theta)\,dt\to\sigma_1^2$. We will show that

$$E\left(1/T\int_0^T h_t^2(\theta)\,dt-\sigma_1^2\right)^2\to 0\qquad\text{as}\quad T\to\infty. \tag{33}$$

Let $M_T=E1/T\int_0^T h_t^2(\theta)\,dt$. Then

$$E\left(1/T\int_0^T h_t^2(\theta)\,dt-\sigma_1^2\right)^2$$
$$\leqslant 2E\left(1/T\int_0^T h_t^2(\theta)\,dt-M_T\right)^2+2(M_T-\sigma_1^2)^2.$$

Consider the first term on the RHS:

$$E\left(1/T\int_0^T h_t^2(\theta)\,dt-M_T\right)^2=E\left(1/T\int_0^T h_t^2(\theta)\,dt\right)^2-(M_T)^2$$
$$=1/T^2\int_0^T\int_0^T (Eh_t^2h_s^2)\,ds\,dt-(M_T)^2.$$

Let $J(s,t)=Eh_th_s$; then, since (h_t) is a Gaussian process,

$$Eh_t^2h_s^2=2J^2(s,t)+J(t,t)\,J(s,s).$$

This, together with the fact that $M_T=1/T\int_0^T J(t,t)\,dt$, implies that

$$E\left(1/T\int_0^T h_t^2(\theta)\,dt-M_T\right)^2=2/T^2\int_0^T\int_0^T J^2(s,t)\,ds\,dt$$
$$=2/T^2\,\mathrm{trace}(JJ^*)$$
$$\leqslant 2/T^2\,\mathrm{trace}(J)\,\|J\|,$$

where J is the integral operator corresponding to the symmetric kernel $J(s, t)$.

Since $h_t = \int_0^t H_\theta(t, s)\, dy_s$, it is easy to check that $J = HRH^* + HH^*$. Therefore, using Lemma 3.3(i) and (ii) we obtain that $\|J\| < C$ and trace $J \leqslant CT$. Thus

$$E\left(1/T \int_0^T h_t^2(\theta)\, dt - M_T\right)^2 \leqslant 2/T^2 CT \to 0$$

as $T \to \infty$. Therefore (33) is proved if we show that

$$M_T \to \sigma_1^2 \qquad \text{as} \quad T \to \infty.$$

However, note that

$$M_T = E 1/T \int_0^T h_t^2(\theta)\, dt = 1/T \int_0^T J(t, t)\, dt$$

$$= 1/T \operatorname{trace} J = 1/T[\operatorname{trace}(HRH^*) + \operatorname{trace}(HH^*)].$$

Therefore

$$\lim_{T \to \infty} M_T = \sigma_1^2.$$

The verification of $1/T \int_0^T g_t^2(\theta)\, dt \to \sigma_2^2$ is exactly identical. Then only $1/T \int_0^T g_t(\theta)\, h_t(\theta)\, dt \to \sigma_{1,2}$ remains to be verified. Let

$$V_T = E 1/T \int_0^T g_t(\theta)\, h_t(\theta)\, dt.$$

Consider

$$E\left(1/T \int_0^T g_t(\theta)\, h_t(\theta)\, dt - V_T\right)^2$$

$$= E\left(1/T \int_0^T g_t(\theta)\, h_t(\theta)\, dt\right)^2 - (V_T)^2$$

$$= 1/T^2 \int_0^T \int_0^T (E g_t g_s h_t h_s)\, ds\, dt - (V_T)^2.$$

Note that if $\xi_1, \xi_2, \xi_3, \xi_4$ are jointly normal then

$$E(\xi_1 \xi_2 \xi_3 \xi_4) = E(\xi_1 \xi_2)\, E(\xi_3 \xi_4) + E(\xi_1 \xi_3)\, E(\xi_2 \xi_4) + E(\xi_1 \xi_4)\, E(\xi_2 \xi_3).$$

Therefore, if $I(s, t) = E(g_t g_s)$ and $K(s, t) = E(h_s g_t)$,

$$E g_t g_s h_t h_s = I(t, s)\, J(t, s) + K(t, t)\, K(s, s) + K(t, s)\, K(s, t).$$

Therefore,

$$
\begin{aligned}
1/T^2 & \int_0^T \int_0^T (Eg_t g_s h_t h_s)\, ds\, dt - (V_T)^2 \\
&= 1/T^2 \int_0^T \int_0^T (I(t, s)\, J(t, s) + K(t, t)\, K(s, s) + K(t, s)\, K(s, t))\, ds\, dt \\
&\quad - 1/T^2 \int_0^T \int_0^T (K(t, t)\, K(s, s))\, ds\, dt \\
&= 1/T^2 \int_0^T \int_0^T (I(t, s)\, J(t, s) + K(t, s)\, K(s, t))\, ds\, dt \\
&\leqslant 1/T^2 \left[\left(\int_0^T \int_0^T I^2(t, s)\, ds\, dt \right)^{1/2} \left(\int_0^T \int_0^T J^2(t, s)\, ds\, dt \right)^{1/2} \right] \\
&\quad + 1/T^2 \left[\left(\int_0^T \int_0^T K^2(t, s)\, ds\, dt \right) \right].
\end{aligned}
$$

Using the facts from Lemma 3.2 and Remark 3.1 it is easy to see that the above terms tend to zero as $T \to \infty$. Let K be the integral operator with kernel $K(s, t)$; then

$$K = HRG^* + HG^*.$$

Thus the verification is complete, since

$$
\begin{aligned}
\lim_{T \to \infty} V_T &= 1/T \lim_{T \to \infty} \int_0^T K(t, t)\, dt \\
&= 1/T \lim_{T \to \infty} \operatorname{trace}[HRG^* + HG^*] = \sigma_{1,2}.
\end{aligned}
$$

This concludes the proof of Theorem 2.1. ∎

References

[1] Bagchi, A., and Borkar, V. (1984). Parameter identification in infinite dimensional linear systems. *Stochastics* **12** 201–213.

[2] Balakrishnan, A. V. (1973). *Stochastic Differential Systems.* Lecture notes in Economics and Mathematical Systems. Springer-Verlag, Berlin.

[3] Balakrishnan, A. V. (1976). *Applied Functional Analysis.* Springer-Verlag, New York.

[4] Basawa, I. V., and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

[5] Basawa, I. V., and Scott, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models.* Lecture notes in statistics, Springer-Verlag, Berlin.

[6] IBRAGIMOV, I. A., AND HASMINSKI, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.
[7] KALLIANPUR, G. (1980). *Stochastic Filtering Theory*. Springer-Verlag, New York.
[8] KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes*. R & E, Research and Exposition in Mathematics, No. 6. Heldermann Verlag, Berlin.
[9] LIPTSER, R. S., AND SHIRYAYEV, A. N. (1977). *Statistics of Random Processes I, II*. Springer-Verlag, Berlin.
[10] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.
[11] SELUKAR, R. S. (1989). *On Estimation of Hilbert Space Valued Parameters*. Ph.D. thesis. University of North Carolina, Chapel Hill.