

# Cross-validation estimation of covariance parameters under fixed-domain asymptotics

F. Bachoc<sup>\*1</sup>, A. Lagnoux<sup>1</sup>, and T.M.N. Nguyen<sup>1</sup>

<sup>1</sup>Institut de Mathématiques de Toulouse, Université Paul Sabatier, France  
118 route de Narbonne, 31062 TOULOUSE Cedex 9  
{francois.bachoc,agnes.lagnoux,thi\_mong\_ngoc.nguyen}@math.univ-  
toulouse.fr

March 22, 2017

## Abstract

We consider a one-dimensional Gaussian process having exponential covariance function. Under fixed-domain asymptotics, we prove the strong consistency and asymptotic normality of a cross validation estimator of the microergodic covariance parameter. In this setting, Ying [40] proved the same asymptotic properties for the maximum likelihood estimator. Our proof includes several original or more involved components, compared to that of Ying. Also, while the asymptotic variance of maximum likelihood does not depend on the triangular array of observation points under consideration, that of cross validation does, and is shown to be lower and upper bounded. The lower bound coincides with the asymptotic variance of maximum likelihood. We provide examples of triangular arrays of observation points achieving the lower and upper bounds. We illustrate our asymptotic results with simulations, and provide extensions to the case of an unknown mean function. To our knowledge, this work constitutes the first fixed-domain asymptotic analysis of cross validation.

*Keywords:* Kriging, cross validation, strong consistency, asymptotic normality, spatial sampling, fixed-domain asymptotics

## 1 Introduction

Kriging [35, 28] consists in inferring the values of a Gaussian random field given observations at a finite set of observation points. It has become a popular method for a large range

---

<sup>\*</sup>Corresponding author. francois.bachoc@math.univ-toulouse.fr, Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 TOULOUSE Cedex 9

of applications, such as geostatistics [25], numerical code approximation [29, 30, 8] and calibration [27, 9] or global optimization [20].

Before Kriging can be applied, a covariance function must be chosen. The most common practice is to statistically estimate the covariance function, from a set of observations of the Gaussian process, and to plug [35, Ch.6.8] the estimate in the Kriging equations. Usually, it is assumed that the covariance function belongs to a given parametric family (see [1] for a review of classical families). In this case, the estimation boils down to estimating the corresponding covariance parameters. For covariance parameter estimation, maximum likelihood (ML) is the most studied and used method, while cross validation (CV) [36, 43, 5] is an alternative technique. CV has been shown to have attractive properties, compared to ML, when the parametric family of covariance functions is misspecified [5, 7].

There is a fair amount of literature on the asymptotic properties of ML. In this regard, the two main frameworks are increasing-domain and fixed-domain asymptotics [35, p.62]. Under increasing-domain asymptotics, the average density of observation points is bounded, so that the infinite sequence of observation points is unbounded. Under fixed-domain asymptotics, this sequence is dense in a bounded domain.

Consider first increasing-domain asymptotics. Generally speaking, for all (identifiable) covariance parameters, the ML estimator is consistent and asymptotically normal under some mild regularity conditions. The asymptotic covariance matrix is equal to the inverse of the (asymptotic) Fisher information matrix. This result was first shown in [24], and then extended in different directions in [12, 13, 31, 6, 16].

The situation is significantly different under fixed-domain asymptotics. Indeed, two types of covariance parameters can be distinguished: microergodic and non-microergodic parameters [18, 35]. A covariance parameter is microergodic if, for two different values of it, the two corresponding Gaussian measures are orthogonal, see [18, 35]. It is non-microergodic if, even for two different values of it, the two corresponding Gaussian measures are equivalent. Non-microergodic parameters cannot be estimated consistently, but have an asymptotically negligible impact on prediction [32, 33, 34, 42]. On the other hand, it is at least possible to consistently estimate microergodic covariance parameters, and misspecifying them can have a strong negative impact on prediction.

Under fixed-domain asymptotics, there exist results indicating which covariance parameters are microergodic, and providing the asymptotic properties of the corresponding ML estimator. Most of these available results are specific to particular covariance models. In dimension  $d = 1$  when the covariance model is exponential, only a reparameterized quantity obtained from the variance and scale parameters is microergodic. It is shown in [40] that the ML estimator of this microergodic parameter is strongly consistent and asymptotically normal. These results are extended in [11], by taking into account measurement errors, and in [10], by taking into account both measurement errors and an unknown mean function. When  $d > 1$  and for a separable exponential covariance function, all the covariance parameters are microergodic, and the asymptotic normality of the ML estimator is proved in [41]. Other results in this case are also given in [37, 2]. Consistency of ML is shown as well in [23] for the scale parameters of the Gaussian covariance function and in [22] for all the covariance parameters of the separable Matérn 3/2 covariance function. Finally, for

the entire isotropic Matérn class of covariance functions, all parameters are microergodic for  $d > 4$  [3], and only reparameterized parameters obtained from the scale and variance are microergodic for  $d \leq 3$  [42]. In [21], the asymptotic normality of the ML estimators for these microergodic parameters is proved, from previous results in [14] and [39]. Finally we remark that, beyond ML, quadratic variation-based estimators have also been extensively studied, under fixed-domain asymptotics (see for instance [19]).

In contrast to ML, CV has received less theoretical attention. Under increasing-domain asymptotics, the consistency and asymptotic normality of a CV estimator is proved in [6]. Also, under increasing-domain asymptotics, it is shown in [7] that this CV estimator asymptotically minimizes the integrated square prediction error. To the best of our knowledge, no fixed-domain asymptotic analysis of CV exists in the literature.

In this paper, we provide a first fixed-domain asymptotic analysis of the CV estimator minimizing the CV logarithmic score, see [28] Equation (5.11) and [43]. We focus on the case of the one-dimensional exponential covariance function, which was historically the first covariance function for which the asymptotic properties of ML were derived [40]. This covariance function is particularly amenable to theoretical analysis, as its Markovian property yields an explicit (matrix-free) expression of the likelihood function. It turns out that the CV logarithmic score can also be expressed in a matrix-free form, which enables us to prove the strong consistency and asymptotic normality of the corresponding CV estimator. We follow the same general proof architecture as in [40] for ML, but our proof, and the nature of our results, contain several new elements.

In terms of proofs, the random CV logarithmic score, and its derivatives, have more complicated expressions than for ML. [This is because the CV logarithm score is based on the conditional distributions of the observations, from both their nearest left and right neighbors, while the likelihood function is solely based on the nearest left neighbors. See Lemma 3.1 and Lemma 1 in [40] for details.] As a consequence, the computations are more involved, and some other tools than in [40] are needed. In particular, many of our asymptotic approximations rely on Taylor expansions of functions of several variables (where each variable is an interpoint distance going to zero, see the proofs for details). In contrast, only Taylor approximations with one variable are needed in [40]. In addition, we use central limit theorems for dependent random variables, while only independent variables need to be considered in [40].

The nature of our asymptotic normality result also differs from that in [40]. In this reference, the asymptotic variance does not depend on the triangular array of observation points. On the contrary, in our case, different triangular arrays of observation points can yield different asymptotic variances. We exhibit a lower and an upper bound for these asymptotic variances, and provide examples of triangular arrays reaching them. The lower bound is in fact equal to the asymptotic variance of ML in [40]. Interestingly, the triangular array given by equispaced observation points attains neither the lower nor the upper bound. It is also pointed out in [6] that equispaced observation points need not provide the smallest asymptotic variance for covariance parameter estimation.

Finally, the fact that the asymptotic variance is larger for CV than for ML is a standard finding in the well-specified case considered here, where the covariance function of

the Gaussian process does belong to the parametric family of covariance functions under consideration. In contrast, as mentioned above, CV has attractive properties compared to ML when this well-specified case does not hold [5, 7].

The rest of the paper is organized as follows. In Section 2, we present in more details the setting and the CV estimator under consideration. In Section 3, we give our strong consistency result for this estimator. In Section 4, we provide the asymptotic normality result, together with the analysis of the asymptotic variance. In Section 5, we present numerical experiments, illustrating our theoretical findings. In Section 6, we extend the results of Sections 3 and 4 to the case of an unknown mean function. In Section 7, we give a few concluding remarks. All the proofs are postponed to Section 8.

## 2 The context and the cross-validation estimators

We consider a centered Gaussian process  $Y$  on  $[0, 1]$  with covariance function

$$K_0(t_1, t_2) = \sigma_0^2 \exp\{-\theta_0|t_1 - t_2|\}$$

for some fixed and unknown parameters  $\theta_0 > 0$  and  $\sigma_0^2 > 0$ . This process is commonly known as the Ornstein-Uhlenbeck process. It satisfies the following stochastic differential equation, called the Langevin's equation,

$$dY(t) = -\theta_0 Y(t)dt + \sqrt{2\theta_0}\sigma_0 dB(t),$$

where  $(B(t))_t$  denotes a standard Brownian motion process. The Ornstein-Uhlenbeck process has been widely used to model physical, biological, social, and many other phenomena. It also possesses many useful mathematical properties that simplify the analysis.

We introduce the parametric set of covariance functions  $\{K_{\theta, \sigma^2}, a \leq \theta \leq A, b \leq \sigma^2 \leq B\}$  for some fixed  $0 < a \leq A < +\infty$  and  $0 < b \leq B < +\infty$  where

$$K_{\theta, \sigma^2}(t_1, t_2) = \sigma^2 \exp\{-\theta|t_1 - t_2|\}.$$

For any  $n \in \mathbb{N}$ , we consider a design of observation points  $\{s_1, \dots, s_n\}$ . Without loss of generality, we may assume that  $0 = s_1 < \dots < s_n = 1$ . Similarly as in [40], there is no need to assume that the sequences of observation points are nested. We consider the vector of observations at locations  $s_1, \dots, s_n$ ,  $(Y(s_1), \dots, Y(s_n))'$ . Now let  $\Delta_i := s_i - s_{i-1}$ , for  $i = 2, \dots, n$ , and  $y_i := Y(s_i)$ , for  $i = 1, \dots, n$ . For ease of redaction, we do not mention in  $s_i$  and  $\Delta_i$  the dependency in  $n$ . We define  $\mathbf{R}_\theta$  as the variance-covariance matrix of  $(y_1, \dots, y_n)'$  under covariance function  $K_{\theta, 1}$ ,

$$\mathbf{R}_\theta := \begin{pmatrix} 1 & e^{-\theta\Delta_2} & \dots & e^{-\theta\sum_{i=2}^n \Delta_i} \\ e^{-\theta\Delta_2} & 1 & \dots & e^{-\theta\sum_{i=3}^n \Delta_i} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\theta\sum_{i=2}^n \Delta_i} & e^{-\theta\sum_{i=3}^n \Delta_i} & \dots & 1 \end{pmatrix}.$$

From [4], we have

$$\mathbf{R}_\theta^{-1} = \begin{pmatrix} \frac{1}{1-e^{-2\theta\Delta_2}} & \frac{-e^{-\theta\Delta_2}}{1-e^{-2\theta\Delta_2}} & 0 & \cdots & 0 \\ \frac{-e^{-\theta\Delta_2}}{1-e^{-2\theta\Delta_2}} & \frac{1}{1-e^{-2\theta\Delta_2}} + \frac{e^{-2\theta\Delta_3}}{1-e^{-2\theta\Delta_3}} & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & & \frac{1}{1-e^{-2\theta\Delta_{n-1}}} + \frac{e^{-2\theta\Delta_n}}{1-e^{-2\theta\Delta_n}} & \frac{-e^{-\theta\Delta_n}}{1-e^{-2\theta\Delta_n}} \\ 0 & \cdots & 0 & \frac{-e^{-\theta\Delta_n}}{1-e^{-2\theta\Delta_n}} & \frac{1}{1-e^{-2\theta\Delta_n}} \end{pmatrix}. \quad (1)$$

We now address the CV estimators of  $\theta_0$  and  $\sigma_0^2$  considered in [28, 43]. Let

$$\hat{Y}_{\theta,-i}(s_i) = \mathbb{E}_{\theta,\sigma^2}(Y(s_i)|Y(s_1), \dots, Y(s_{i-1}), Y(s_{i+1}), \dots, Y(s_n)),$$

where the conditional expectation  $\mathbb{E}_{\theta,\sigma^2}$  is calculated assuming that  $Y$  is centered and has covariance function  $K_{\theta,\sigma^2}$ . We remark that  $\hat{Y}_{\theta,-i}(s_i)$  does not depend on  $\sigma^2$ . We define similarly

$$\hat{\sigma}_{\theta,\sigma^2,-i}^2(s_i) = \text{Var}_{\theta,\sigma^2}(Y(s_i)|Y(s_1), \dots, Y(s_{i-1}), Y(s_{i+1}), \dots, Y(s_n)).$$

Then, the CV estimators are given by

$$(\hat{\theta}, \hat{\sigma}^2) \in \underset{a \leq \theta \leq A, b \leq \sigma^2 \leq B}{\text{argmin}} S_n(\theta, \sigma^2),$$

where

$$S_n(\theta, \sigma^2) = \sum_{i=1}^n \left[ \log(\hat{\sigma}_{\theta,\sigma^2,-i}^2(s_i)) + \frac{(y_i - \hat{Y}_{\theta,-i}(s_i))^2}{\hat{\sigma}_{\theta,\sigma^2,-i}^2(s_i)} \right] \quad (2)$$

is the logarithmic score. The rationale for minimizing the logarithmic score is that  $\log(2\pi) + \log(\hat{\sigma}_{\theta,\sigma^2,-i}^2(s_i)) + \frac{(y_i - \hat{Y}_{\theta,-i}(s_i))^2}{\hat{\sigma}_{\theta,\sigma^2,-i}^2(s_i)}$  is equal to  $-2$  times the conditional log-likelihood of  $y_i$ , given  $(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)'$ , with covariance parameters  $\theta, \sigma^2$ . The term cross-validation underlines the fact that we consider leave-one-out quantities.

As already known [18, 40, 42], it is not possible to consistently estimate simultaneously  $\theta_0$  and  $\sigma_0^2$  (the ML estimator of  $\theta_0$  is a non-degenerate random variable, even if  $(Y(t))_{t \in [0,1]}$  is observed continuously [44]), but it is possible to consistently estimate  $\theta_0 \sigma_0^2$ . As a consequence, we have considered three different cases, as in [40]. (i) Set  $\sigma^2 = \sigma_1^2$  in (2) with  $\sigma_1^2 > 0$  being a predetermined constant and consider the CV estimator  $\hat{\theta}_1$  of  $\theta_1 = \theta_0 \sigma_0^2 / \sigma_1^2$  that minimizes (2) with  $\sigma^2 = \sigma_1^2$ . (ii) Set  $\theta = \theta_2$  in (2) with  $\theta_2 > 0$  being a predetermined constant and consider the CV estimator  $\hat{\sigma}_2^2$  of  $\sigma_2^2 = \theta_0 \sigma_0^2 / \theta_2$  that minimizes (2) with  $\theta = \theta_2$ . (iii) Consider the estimator  $\hat{\theta} \hat{\sigma}^2$  of  $\theta_0 \sigma_0^2$ , where  $\hat{\theta}$  and  $\hat{\sigma}^2$  are the CV estimators of  $\theta_0$  and  $\sigma_0^2$ .

Ying [40] considers the ML estimators of  $\theta$  and  $\sigma^2$  and establishes their consistency and asymptotic normality. We carry out a similar asymptotic analysis for the above CV estimators. More precisely, we prove that  $\hat{\theta} \hat{\sigma}^2$  (resp.  $\hat{\theta}_1$  and  $\hat{\sigma}_2^2$ ) converges almost surely

to  $\theta_0\sigma_0^2$  (resp.  $\theta_1$  and  $\sigma_2^2$ ) in the next section. In section 4, we establish that, for a sequence  $\tau_n$  which is lower and upper-bounded,  $(\sqrt{n}/[\theta_1\tau_n])(\hat{\theta}_1 - \theta_1)$ ,  $(\sqrt{n}/[\sigma_2^2\tau_n])(\hat{\sigma}_2^2 - \sigma_2^2)$  and  $(\sqrt{n}/[\theta_0\sigma_0^2\tau_n])(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2)$  all converge in distribution to a standard Gaussian random variable. We remark that the asymptotic variance  $\tau_n^2$  depends on how the underlying design points  $\{s_1, \dots, s_n\}$  are chosen. On the contrary, considering the ML estimators [40], the asymptotic variance is the same for any triangular array of design points.

### 3 Consistency

In this section, we establish the strong consistency of the CV estimator  $\hat{\theta}\hat{\sigma}^2$  of  $\theta_0\sigma_0^2$  described in the previous section. In that view, we consider  $S_n(\theta, \sigma^2)$  defined by (2). As done in [40], we base our analysis on the Markovian property of the Ornstein-Uhlenbeck process in order to handle the fact that, as  $n$  increases, the observed sample  $(y_1, \dots, y_n)'$  becomes more and more correlated. We have

$$\hat{Y}_{\theta, -i}(s_i) = - \sum_{\substack{j=1, \dots, n; \\ j \neq i}} \frac{(\mathbf{R}_\theta^{-1})_{ij}}{(\mathbf{R}_\theta^{-1})_{ii}} y_j \quad (3)$$

and

$$\hat{\sigma}_{\theta, \sigma^2, -i}^2(s_i) = \frac{\sigma^2}{(\mathbf{R}_\theta^{-1})_{ii}},$$

from [43, 5, 15]. Then, using Equation (1), we get the following lemma after some tedious computations.

**Lemma 3.1** (Logarithmic score). *With  $S_n(\theta, \sigma^2)$  as in (2), we have*

$$\begin{aligned} S_n(\theta, \sigma^2) &= n \log(\sigma^2) + \log(1 - e^{-2\theta\Delta_2}) + \log(1 - e^{-2\theta\Delta_n}) \\ &+ \frac{(y_1 - e^{-\theta\Delta_2}y_2)^2}{\sigma^2(1 - e^{-2\theta\Delta_2})} + \frac{(y_n - e^{-\theta\Delta_n}y_{n-1})^2}{\sigma^2(1 - e^{-2\theta\Delta_n})} - \sum_{i=2}^{n-1} \log \left( \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right) \\ &+ \frac{1}{\sigma^2} \sum_{i=2}^{n-1} \left[ \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right] \left[ y_i - \frac{\frac{e^{-\theta\Delta_i}}{1 - e^{-2\theta\Delta_i}}y_{i-1} + \frac{e^{-\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}}y_{i+1}}{\frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}}} \right]^2. \end{aligned}$$

Based on Lemma 3.1, we prove the following theorem in Section 8.2.

**Theorem 3.2** (Consistency). *Assume that*

$$\limsup_{n \rightarrow +\infty} \max_{i=2, \dots, n} \Delta_i = 0. \quad (4)$$

*Let  $J = [a, A] \times [b, B]$ , where  $a, A, b$  and  $B$  are fixed and have been defined in the previous section. Assume that there exists  $(\tilde{\theta}, \tilde{\sigma}^2)$  in  $J$  so that  $\tilde{\theta}\tilde{\sigma}^2 = \theta_0\sigma_0^2$ . Define  $(\hat{\theta}, \hat{\sigma}^2) \in J$  as a solution of*

$$S_n(\hat{\theta}, \hat{\sigma}^2) = \min_{(\theta, \sigma^2) \in J} S_n(\theta, \sigma^2). \quad (5)$$

Then  $(\hat{\theta}, \hat{\sigma}^2)$  exists and

$$\hat{\theta}\hat{\sigma}^2 \xrightarrow{a.s.} \theta_0\sigma_0^2. \quad (6)$$

In particular, let  $\sigma_1^2 > 0$  and  $\theta_2 > 0$  be predetermined constants satisfying  $\sigma_0^2\theta_0/\sigma_1^2 \in [a, A]$  and  $\sigma_0^2\theta_0/\theta_2 \in [b, B]$ . Define  $\hat{\theta}_1 \in [a, A]$  and  $\hat{\sigma}_2^2 \in [b, B]$  as solutions of

$$S_n(\hat{\theta}_1, \sigma_1^2) = \min_{\theta \in [a, A]} S_n(\theta, \sigma_1^2) \quad (7)$$

and

$$S_n(\theta_2, \hat{\sigma}_2^2) = \min_{\sigma^2 \in [b, B]} S_n(\theta_2, \sigma^2). \quad (8)$$

Then  $\hat{\theta}_1 \xrightarrow{a.s.} \theta_1 := \sigma_0^2\theta_0/\sigma_1^2$  and  $\hat{\sigma}_2^2 \xrightarrow{a.s.} \sigma_2^2 := \sigma_0^2\theta_0/\theta_2$ .

**Remark 3.3.** It is worth remarking that the asymptotically preponderant terms in Lemma 3.1 are the same as those obtained in the context of ML estimation (see [40] and Section 8.2 for more details).

## 4 Asymptotic normality

Once the consistency has been established, the natural question of the convergence speed arises. We address this point in this section. We first provide a central limit result in the following theorem.

**Theorem 4.1** (Central Limit Theorem). *Consider the same notation and assumptions as in Theorem 3.2. Assume further that either  $aB < \theta_0\sigma_0^2$ ;  $Ab > \theta_0\sigma_0^2$  or  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$  hold. Then the estimators are asymptotically normal. More precisely, we have*

$$\frac{\sqrt{n}}{\theta_0\sigma_0^2\tau_n}(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1). \quad (9)$$

Also, when  $(\sigma_0^2\theta_0)/\sigma_1^2 \in (a, A)$  we have

$$\frac{\sqrt{n}}{\theta_1\tau_n}(\hat{\theta}_1 - \theta_1) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1). \quad (10)$$

Finally, when  $(\sigma_0^2\theta_0)/\theta_2 \in (b, B)$  we have

$$\frac{\sqrt{n}}{\sigma_2^2\tau_n}(\hat{\sigma}_2^2 - \sigma_2^2) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1). \quad (11)$$

The quantity  $\tau_n^2$  depends on how the underlying design points  $\{s_1, \dots, s_n\}$  have been chosen. More precisely,

$$\tau_n^2 = \frac{2}{n} \sum_{i=3}^{n-1} \left[ \left( \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + \frac{\Delta_{i-1}}{\Delta_i + \Delta_{i-1}} \right)^2 + 2 \frac{\Delta_i \Delta_{i+1}}{(\Delta_i + \Delta_{i+1})^2} \right]. \quad (12)$$

**Remark 4.2.** The condition  $aB < \theta_0\sigma_0^2$ ;  $Ab > \theta_0\sigma_0^2$  or  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$  ensures that  $(\partial/\partial\theta)S_n(\hat{\theta}, \hat{\sigma}^2)$  or  $(\partial/\partial\sigma^2)S_n(\hat{\theta}, \hat{\sigma}^2)$  will be equal to zero for  $n$  large enough almost surely, by applying Theorem 3.2. This is used in the proof of Theorem 4.1. A similar assumption is made in [40], where the parameter domain for  $(\theta, \sigma^2)$  is  $(0, \infty) \times [b, B]$  or  $[a, A] \times (0, \infty)$ .

In the following proposition, we show that the quantity  $\tau_n^2$  in Theorem 4.1 is lower and upper bounded, so that the rate of convergence is always  $\sqrt{n}$  in this theorem.

**Proposition 4.3.** *We have, for any choice of the triangular array of design points  $\{s_1, \dots, s_n\}$  satisfying (4),*

$$2 \leq \liminf_{n \rightarrow \infty} \tau_n^2 \leq \limsup_{n \rightarrow \infty} \tau_n^2 \leq 4. \quad (13)$$

**Remark 4.4.** 1. The asymptotic variance of the limiting distribution of  $\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2$  can be easily estimated. By the previous proposition, this asymptotic variance is always larger than the one of the ML estimator. Indeed, with  $\hat{\theta}_{ML}$  and  $\hat{\sigma}_{ML}^2$  the ML estimators of  $\theta$  and  $\sigma^2$  we have  $(\sqrt{n}/[\theta_0\sigma_0^2])(\hat{\theta}_{ML}\hat{\sigma}_{ML}^2 - \theta_0\sigma_0^2) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 2)$ , see [40]. This fact is quite expected as ML estimates usually perform best when the covariance model is well-specified, as is the case here.

2. As one can check easily, the regular design  $\Delta_i \equiv \frac{1}{n-1}$  for all  $i = 2, \dots, n$ , does not yield the limiting variance of the ML estimator. Instead, we have  $\tau_n^2 \rightarrow_{n \rightarrow \infty} 3$  for this design. However, in Proposition 4.5, we exhibit a particular design realizing the limiting variance of the ML estimator:  $\lim_{n \rightarrow \infty} \tau_n^2 = 2$ .

In fact, the bounds in (13) are sharp as shown in the following proposition.

**Proposition 4.5.** (i) *Let  $\{s_1, \dots, s_n\}$  be such that  $s_1 = 0$ , for  $i = 2, \dots, n-1$ ,*

$$\Delta_i = \begin{cases} (1 - \gamma_n) \frac{2}{n} & \text{if } i \text{ is even,} \\ \frac{2\gamma_n}{n} & \text{if } i \text{ is odd,} \end{cases}$$

where  $\gamma_n \in (0, 1)$ , and  $\Delta_n = 1 - \sum_{i=2}^{n-1} \Delta_i$ . Then, taking  $\gamma_n = 1/n$ , we get  $\tau_n^2 \xrightarrow[n \rightarrow \infty]{} 4$ .

(ii) *Let  $\{s_1, \dots, s_n\}$  and  $0 < \alpha < 1$  be such that  $s_1 = 0$ ,  $\Delta_i = 1/(i!)$  for  $i = \lfloor n^\alpha \rfloor + 1, \dots, n$  and  $\Delta_2 = \dots = \Delta_{\lfloor n^\alpha \rfloor} \equiv (1 - r_n)/(\lfloor n^\alpha \rfloor - 1)$  with  $r_n := \sum_{i=\lfloor n^\alpha \rfloor + 1}^n \Delta_i$ . Then  $\sum_{i=2}^n \Delta_i = 1$  and*

$$\tau_n^2 \xrightarrow[n \rightarrow \infty]{} 2.$$

**Remark 4.6.** Intuitively, in Proposition 4.5 (ii),  $\Delta_{i+1}$  will be much smaller than  $\Delta_i$  for most of the indices  $i$ , so that the quantities  $\frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}}$  and  $\frac{\Delta_i \Delta_{i+1}}{(\Delta_i + \Delta_{i+1})^2}$  in (12) will be negligible. We refer to the proof of Proposition 4.5 for further details.



## 5 Numerical experiments

We illustrate Theorem 4.1 by a Monte Carlo simulation. We set  $\theta_0 = 3$  and  $\sigma_0^2 = 1$  and we consider three sample size values,  $n = 12, 50, 200$ . For the sample size  $n = 12$ , we address three designs  $\{s_1, \dots, s_n\}$ . The first one is the ‘minimal’ design given by Proposition 4.5 (ii) with  $\alpha = 0.5$ , which asymptotically achieves the minimal estimation variance. The second one is the ‘regular’ design given by  $\{s_1, \dots, s_n\} = \{0, 1/(n-1), \dots, 1\}$ . The third one is the ‘maximal’ design given by Proposition 4.5 (i) with  $\gamma_n = 1/n$ , which asymptotically achieves the maximal estimation variance. These three designs are show in Figure 1. For the sample sizes  $n = 50$  and  $n = 200$ , the ‘minimal’ design is not amenable to numerical computation anymore, as the values of  $\Delta_i$  become too small; so that we only address the ‘regular’ and ‘maximal’ designs.

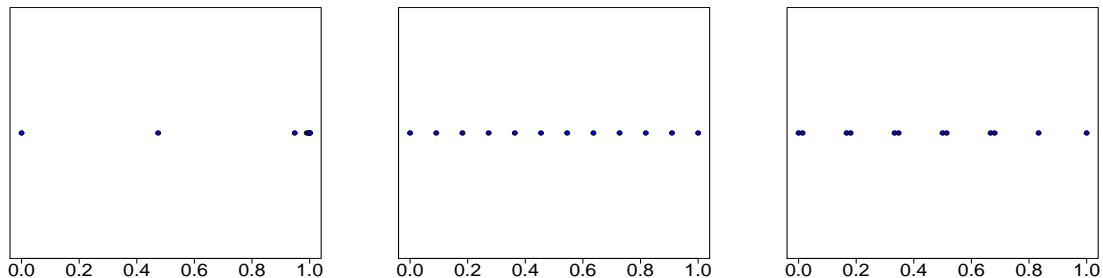


Figure 1: Plot of the points  $\{s_1, \dots, s_n\}$  for the ‘minimal’, ‘regular’ and ‘maximal’ designs of the numerical experiments for  $n = 12$  (from left to right). For the ‘minimal’ design, nine points form a dense cluster around one and the asymptotic variance of the CV estimator is 2 (Proposition 4.5 (i)), for the ‘regular’ design, the asymptotic variance is 3, and for the ‘maximal’ design, the asymptotic variance is 4 (Proposition 4.5 (ii)).

For a given configuration of  $n$  and a given design  $\{s_1, \dots, s_n\}$ , we repeat  $N = 2.000$  data generations and estimations. That is, we independently sample  $N$  Gaussian vectors of size  $n$  with zero mean vector and covariance matrix  $[\sigma_0^2 e^{-\theta_0 |s_i - s_j|}]_{1 \leq i, j \leq n}$ . For each of these Gaussian vectors, we compute the CV estimators  $\hat{\theta}$  and  $\hat{\sigma}^2$ , with parameter space  $[0.1, 10] \times [0.3, 30]$ , so that we consider case (9) of Theorem 4.1. The computation of  $\hat{\theta}$  and  $\hat{\sigma}^2$  is not challenging since, from Lemma 3.1, the logarithmic score  $S_n(\theta, \sigma^2)$  can be computed quickly, with a  $O(n)$  complexity. [For more general covariance functions, the computation of CV or ML criteria is more costly, with a  $O(n^3)$  complexity.] The criterion  $S_n$  is minimized over  $(\theta, \sigma^2)$  by repeating the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, with several starting points for  $(\theta, \sigma^2)$ , and by keeping the value of  $(\theta, \sigma^2)$  with smallest logarithmic score, over all the repetitions. The **R** software was used, with the **optim** function. [We remark that, for fixed  $\theta$ , one could find an explicit expression of  $\hat{\sigma}^2(\theta) \in \operatorname{argmin}_{\sigma^2 > 0} S_n(\theta, \sigma^2)$  (see also [5] for a different CV criterion). Hence, it would be possible to minimize the profile logarithmic score  $\min_{\sigma^2 > 0} S_n(\theta, \sigma^2)$  over  $\theta$  only. As

mentioned earlier, this improvement is not needed here, since the criterion  $S_n(\theta, \sigma^2)$  can be computed quickly.]

For the  $N$  values of  $(\hat{\theta}, \hat{\sigma}^2)$ , we compute the  $N$  values of  $(\sqrt{n}/[\theta_0\sigma_0^2])(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2)$ . In Figure 2, we report the histograms of these latter  $N$  values, for the seven configurations under consideration. In addition, we report the probability density functions of the seven corresponding Gaussian distributions with mean 0 and variance  $\tau_n^2$ , to which the histograms converge when  $n \rightarrow \infty$ , in view of Theorem 4.1.

In Figure 2, we observe that, for  $n = 12$ , the asymptotic Gaussian distributions are already reasonable approximations of the empirical histograms. For  $n = 50$ , the asymptotic distributions become very close to the histograms, and for  $n = 200$  the asymptotic distributions are almost identical to the histograms. Hence, the convergence in distribution of Theorem 4.1 provides a good approximation of the finite sample situation already for small to moderate  $n$ . The case  $n = 12$  illustrates the benefit of the ‘minimal’ design for estimation, as the histogram is most concentrated around zero for this design. Similarly, the value of  $\tau_{12}^2$  is the smallest for this design, compared to the ‘regular’ and ‘maximal’ designs. For  $n = 50$  and 200, we also observe that the estimation is more accurate for the ‘regular’ design than for the ‘maximal’ design, which also confirms Remark 4.4 and Proposition 4.5.

Finally, we have obtained similar conclusions for the case where either  $\theta_0$  or  $\sigma_0^2$  is known in the computation of  $\hat{\theta}, \hat{\sigma}^2$  (cases of (10) and (11)). We do not report the corresponding results for concision.

## 6 Extension to regression models

In this section, we extend Theorems 3.2 and 4.1 to the case of regression models. We assume that, instead of  $Y$ , we observe the Gaussian process  $Z$  defined by  $Z(t) = \beta_{01}f_1(t) + \dots + \beta_{0p}f_p(t) + Y(t)$ . In the definition of  $Z$ ,  $\beta_0 = (\beta_{01}, \dots, \beta_{0p})' \in \mathbb{R}^p$  is fixed and unknown and, for  $k = 1, \dots, p$ ,  $f_k : [0, 1] \rightarrow \mathbb{R}$  is a known function. Hence, we estimate jointly  $(\theta, \sigma^2, \beta_0)$  from the observation vector  $\mathbf{z} = (z_1, \dots, z_n)'$ , with  $z_i = Z(s_i)$ .

Let  $\mathbf{F}$  be the  $n \times p$  matrix  $[f_j(s_i)]_{i=1, \dots, n, j=1, \dots, p}$ . Then  $\hat{\beta} = (\mathbf{F}'\mathbf{R}_\theta^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{R}_\theta^{-1}\mathbf{z}$  is the best linear unbiased predictor of  $\beta_0$  given  $\mathbf{z}$ , under covariance function  $K_{\theta, \sigma^2}$  for all  $\sigma^2$ , see e.g. [30].

We now address CV estimation. Let  $\mathbf{f}_i = (f_1(s_i), \dots, f_p(s_i))'$ , let  $\mathbf{z}_{-i} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)'$ , let  $\mathbf{F}_{-i}$  be the matrix obtained by removing line  $i$  of  $\mathbf{F}$ , and let  $\mathbf{R}_{\theta, -i}$  be the  $(n-1) \times (n-1)$  matrix obtained by removing line and column  $i$  of  $\mathbf{R}_\theta$ . Then, for all  $\sigma^2$ ,  $\hat{\beta}_{-i} = (\mathbf{F}_{-i}'\mathbf{R}_{\theta, -i}^{-1}\mathbf{F}_{-i})^{-1}\mathbf{F}_{-i}'\mathbf{R}_{\theta, -i}^{-1}\mathbf{z}_{-i}$  is the best linear unbiased predictor of  $\beta_0$  given  $\mathbf{z}_{-i}$ , under covariance function  $K_{\theta, \sigma^2}$ .

We also let  $\mathbf{r}_{\theta, -i} = (K_{\theta, 1}(s_i, s_1), \dots, K_{\theta, 1}(s_i, s_{i-1}), K_{\theta, 1}(s_i, s_{i+1}), \dots, K_{\theta, 1}(s_i, s_n))'$ . Then, from e.g. [30],

$$\hat{Z}_{\theta, -i}(s_i) = \mathbf{f}_i' \hat{\beta}_{-i} + \mathbf{r}_{\theta, -i}' \mathbf{R}_{\theta, -i}^{-1} (\mathbf{z}_{-i} - \mathbf{F}_{-i} \hat{\beta}_{-i}) \quad (14)$$

is the best linear unbiased predictor of  $z_i$  given  $\mathbf{z}_{-i}$ . We let

$$\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i) = \mathbb{E}_{\theta, \sigma^2} \left( \left[ \hat{Z}_{\theta, -i}(s_i) - Z(s_i) \right]^2 \right).$$

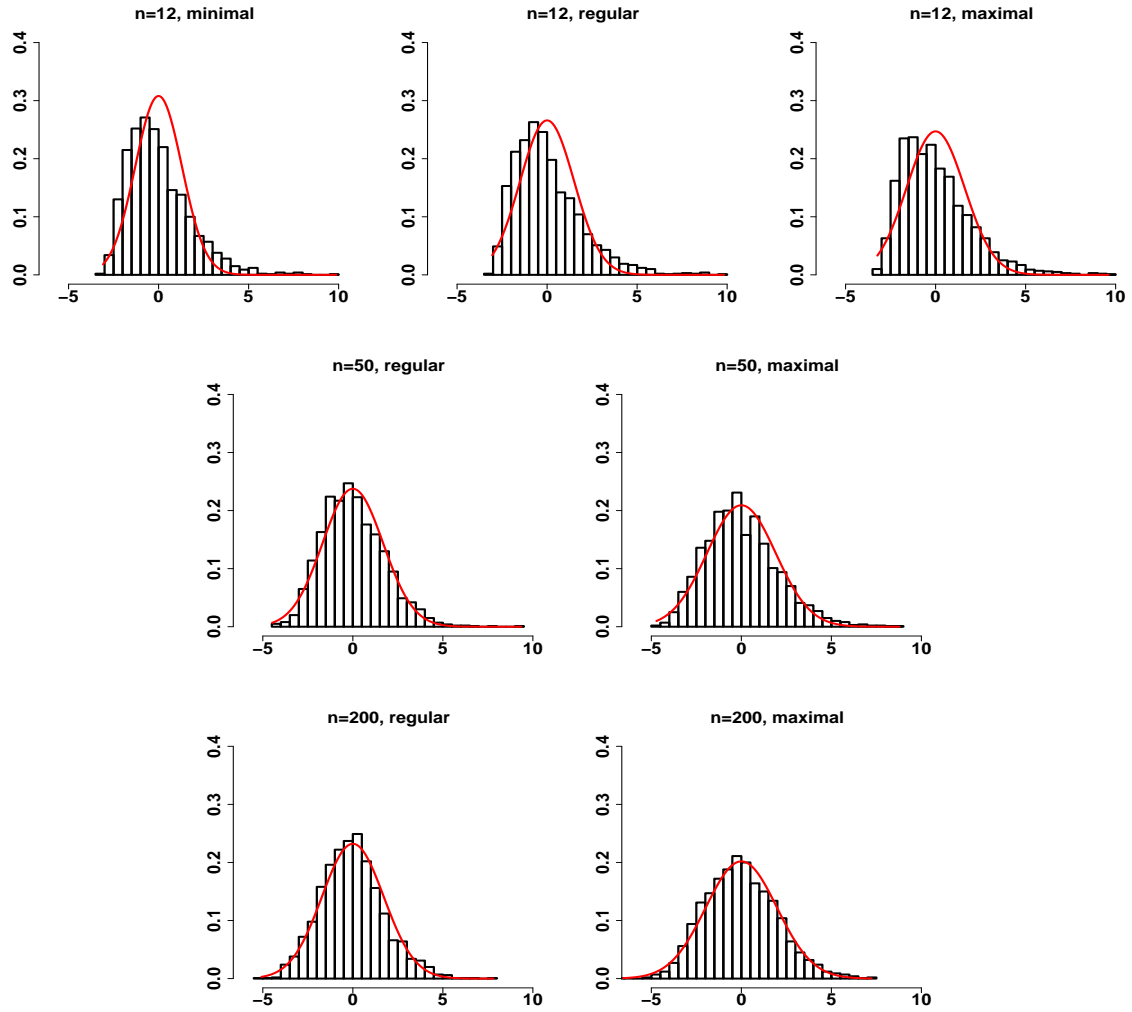


Figure 2: Illustration of Theorem 4.1. Histograms of  $N = 2,000$  independent realizations of  $(\sqrt{n}/[\theta_0\sigma_0^2])(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2)$ , together with the corresponding asymptotic Gaussian probability density functions with mean 0 and variances  $\tau_n^2$  (red lines). The sample size is  $n = 12$  (top row),  $n = 50$  (middle row) and  $n = 200$  (bottom row). For the top row, the designs are the ‘minimal’ design (left), achieving the smallest asymptotic variance; the ‘regular’ design (middle), with equispaced observation points; and the ‘maximal’ design (right), achieving the largest asymptotic variance. For the middle and bottom rows, the designs are the ‘regular’ design (left) and the ‘maximal’ design (right).

Then, the CV estimator of  $(\theta, \sigma^2)$  we shall study in this section is

$$(\check{\theta}, \check{\sigma}^2) \in \operatorname{argmin}_{a \leq \theta \leq A, b \leq \sigma^2 \leq B} \bar{S}_n(\theta, \sigma^2)$$

with

$$\bar{S}_n(\theta, \sigma^2) = \sum_{i=1}^n \left[ \log(\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i)) + \frac{(z_i - \hat{Z}_{\theta, -i}(s_i))^2}{\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i)} \right].$$

We remark that [43] suggests to use a similar CV criterion, with the notable difference that  $\hat{\beta}_{-i}$  is replaced by  $\hat{\beta}$  in (14). The benefit of the CV predictor (14), compared to that considered in [43], is that, in (14), no use of  $z_i$  is made at all for predicting  $z_i$ . In [15], the following relations are shown, extending those of Section 3. We have

$$\hat{Z}_{\theta, -i}(s_i) = - \sum_{\substack{j=1, \dots, n; \\ j \neq i}} \frac{(\mathbf{Q}_{\theta}^-)_{ij}}{(\mathbf{Q}_{\theta}^-)_{ii}} z_j$$

and

$$\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i) = \frac{\sigma^2}{(\mathbf{Q}_{\theta}^-)_{ii}}, \quad (15)$$

with  $\mathbf{Q}_{\theta}^- = \mathbf{R}_{\theta}^{-1} - \mathbf{R}_{\theta}^{-1} \mathbf{F} (\mathbf{F}' \mathbf{R}_{\theta}^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{R}_{\theta}^{-1}$ .

Based on the two displays above, and again using the explicit matrix inverse in (1), we are able to prove the consistency and asymptotic normality of  $\check{\theta} \check{\sigma}^2$  where the asymptotic distribution is identical to that of Section 4.

**Theorem 6.1** (Consistency). *Assume that  $\limsup_{n \rightarrow +\infty} \max_{i=2, \dots, n} \Delta_i = 0$  and that there exists  $(\tilde{\theta}, \tilde{\sigma}^2)$  in  $J$  so that  $\tilde{\theta} \tilde{\sigma}^2 = \theta_0 \sigma_0^2$ . Assume also that  $f_1, \dots, f_p$  are twice continuously differentiable and are linearly independent on  $[0, 1]$ . Then  $(\check{\theta}, \check{\sigma}^2)$  exists and*

$$\check{\theta} \check{\sigma}^2 \xrightarrow{a.s.} \theta_0 \sigma_0^2. \quad (16)$$

**Theorem 6.2** (Central Limit Theorem). *Assume that the conditions of Theorem 6.1 hold and that  $aB > \theta_0 \sigma_0^2$  and  $Ab < \theta_0 \sigma_0^2$ . Then, with  $\tau_n$  as in (12), we have*

$$\frac{\sqrt{n}}{\theta_0 \sigma_0^2 \tau_n} (\check{\theta} \check{\sigma}^2 - \theta_0 \sigma_0^2) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1). \quad (17)$$

In Theorems 6.1 and 6.2, the twice differentiability condition for  $f_1, \dots, f_p$  is mostly technical, and could be replaced by a continuous differentiability condition, at the price of more technical proofs. [We remark that Theorems 6.1 and 6.2 apply in particular to polynomial functions  $f_1, \dots, f_p$  which are widely considered, for instance in the framework of computer experiments [30].] As remarked in [40], when  $f_1, \dots, f_p$  are continuously differentiable, the parameter  $\beta_0$  is non-microergodic and can not be consistently estimated.

Finally, assume now that  $f_1, \dots, f_p$  satisfy the conditions given in Theorem 3 (ii) in [40]. Then, one can show from the proof of this theorem that, for any sequence or random variables  $(\bar{\theta}, \bar{\sigma}^2) \in J$  (and in particular for  $(\check{\theta}, \check{\sigma}^2)$ ), the estimator  $\hat{\beta} = \hat{\beta}(\bar{\theta})$  given above is consistent and asymptotically normal, with asymptotic distribution given in (4.5) in [40]. In this setting, it would be interesting to study the joint asymptotic distribution of  $(\check{\theta}, \check{\sigma}^2, \hat{\beta})$ .

## 7 Concluding remarks

We have proved the consistency and asymptotic normality of the CV estimator of the microergodic parameter  $\theta_0\sigma_0^2$ , based on the logarithmic score. While the ML asymptotic variance of  $(\sqrt{n}/[\theta_0\sigma_0^2])(\hat{\theta}_{ML}\hat{\sigma}_{ML}^2 - \theta_0\sigma_0^2)$  is 2 for any triangular array of observation points, the corresponding CV asymptotic variance is simply bounded between 2 and 4, those bounds being tight. The triangular array we exhibit, achieving the asymptotic variance 2 for CV, is based on some ratios between interpoint distances (of the form  $(s_{i+1} - s_i)/(s_{j+1} - s_j)$ ) going to zero as  $n \rightarrow \infty$ , which makes it too challenging to simulate numerically for large  $n$ . It would be interesting to find the smallest possible asymptotic variance for CV, when the ratios between interpoint distances are bounded away from zero.

One interesting agenda for future research would be to extend this asymptotic analysis of CV in the other settings where such an analysis was possible for ML. These settings include the case of measurement errors for the exponential covariance function in dimension one [11, 10], the case of the separable exponential covariance function in higher dimension [41] (consistency and asymptotic normality), of the separable Matérn 3/2 covariance function [22] (consistency) and of the Gaussian covariance function [23] (consistency). In these references, tractable approximations of the inverse covariance matrices are provided, which could also be exploited in the case of CV. Finally, using techniques which are more spectral in nature, [14, 21, 39] prove central limit theorems for the ML estimation of the microergodic parameter in the isotropic Matérn model. An extension to CV would also be valuable.

## 8 Proofs

### 8.1 Notation and auxiliary results

Remind that  $\Delta_i = s_i - s_{i-1}$  and introduce  $W_i := [y_i - e^{-\theta\Delta_i}y_{i-1}]$  and its normalized version  $\bar{W}_i := [y_i - e^{-\theta\Delta_i}y_{i-1}]/[\sigma^2(1 - e^{-2\theta\Delta_i})]^{1/2}$  for  $i = 2, \dots, n$  (the dependency in  $n$  is not mentioned in  $W_i$  and  $\bar{W}_i$  to lighten notation). When  $(\theta, \sigma^2) = (\theta_0, \sigma_0^2)$ , the random variables will be denoted  $W_{i,0}$  and  $\bar{W}_{i,0}$ . By the Markovian and Gaussian properties of  $Y$ , it follows that for each  $i \geq 2$ ,  $\bar{W}_{i,0}$  is independent of  $\{y_j, j \leq i-1\}$ . Therefore,  $\{\bar{W}_{i,0}, 2 \leq i \leq n\}$  is an i.i.d. sequence of random variables having the standard normal distribution  $\mathcal{N}(0, 1)$ .

It is convenient to have short expressions for terms that converge in probability to zero. We follow [38]. The notation  $o_{\mathbb{P}}(1)$  (respectively  $O_{\mathbb{P}}(1)$ ) stands for a sequence of random variables that converges to zero in probability (resp. is bounded in probability) as  $n \rightarrow \infty$ . More generally, for a sequence of random variables  $R_n$ ,

$$\begin{aligned} X_n &= o_{\mathbb{P}}(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{with} \quad Y_n \xrightarrow{\mathbb{P}} 0 \\ X_n &= O_{\mathbb{P}}(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{with} \quad Y_n = O_{\mathbb{P}}(1). \end{aligned}$$

For deterministic sequences  $X_n$  and  $R_n$ , the stochastic notation reduce to the usual  $o$  and  $O$ . Throughout the paper, by  $K$ , we denote a generic constant (i.e.  $K$  may or may not be the same at different occurrences) that does not depend on  $(Y, \theta, \sigma^2, n)$ .

We also denote by  $\delta_n$  and  $\bar{\delta}_n$  two sequences of random variables satisfying

$$\sup_{(\theta, \sigma^2) \in J} |\delta_n| = O_{\mathbb{P}}(1)$$

and

$$\sup_{(\theta, \sigma^2) \in J} |\bar{\delta}_n| = o(n), \text{ a.s.}$$

The definition of  $\delta_n$  and  $\bar{\delta}_n$  may change from one line to the other. Similarly, we denote by  $\delta_{i,n}$  a triangular array of deterministic scalars satisfying

$$\sup_{\substack{n \in \mathbb{N}, i=1, \dots, n \\ (\theta, \sigma^2) \in J}} |\delta_{i,n}| \leq K.$$

The definition of  $\delta_{i,n}$  may change from one line to the other, and possibly also in different occurrences within the same line. We also use several times that,

$$\sum_{i=2}^{n-1} \Delta_i = 1, \tag{18}$$

$$\sup_{t \in [0,1]} |Y(t)| < +\infty \text{ a.s.} \tag{19}$$

Before turning to the proof of Theorems 3.2 and 4.1, we state five auxiliary lemmas that will be required in the sequel.

**Lemma 8.1.** (i) *Let  $\lambda_i > 0$ ,  $i = 1, 2$  be fixed. Then as  $x \rightarrow 0$ ,*

$$\frac{1 - e^{-\lambda_1 x}}{1 - e^{-\lambda_2 x}} = \frac{\lambda_1}{\lambda_2} + O(x).$$

(ii) *Let  $\lambda > 0$  be fixed. Then as  $x, y \rightarrow 0$ ,*

$$1 + e^{-\lambda y} \frac{1 - e^{-\lambda x}}{1 - e^{-\lambda y}} = \frac{x + y}{y} (1 + O(x + y)).$$

**Lemma 8.2.** *For any constant  $\delta > 0$ , there exists a constant  $\eta > 0$  such that*

$$\inf_{\substack{|x-1| \geq \delta \\ x > 0}} (x - 1 - \log x) \geq \eta. \tag{20}$$

**Lemma 8.3.** *Suppose that for each  $n$ , the random variables  $Z, Z_{1,n}, \dots, Z_{n,n}$  are independent and identically distributed and centered. Suppose also that for some  $t > 0$  and  $p > 0$ ,  $\mathbb{E}[\exp\{t|Z|^p\}] < \infty$ . Then for all  $\alpha > 0$ , a.s*

$$\sup_{1 \leq k \leq n} |Z_{k,n}| = o(n^\alpha) \tag{21}$$

$$\sum_{k=2}^n Z_{k,n} = o(n^{(1/2)+\alpha}). \tag{22}$$

The proof of Lemma 8.1 is direct and these of Lemmas 8.2 and 8.3 can be found in [40].

**Lemma 8.4.** *Let for any  $i \in \{2, \dots, n-1\}$ ,*

$$\eta_i = \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n}(\Delta_i + \Delta_{i+1}))$$

and

$$\tau_i = \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n}(\Delta_i + \Delta_{i+1})).$$

Then

$$(i) \quad \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \eta_i W_{i,0} y_i \right| = O_{\mathbb{P}}(1);$$

$$(ii) \quad \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \eta_i W_i y_i \right| = O_{\mathbb{P}}(1);$$

$$(iii) \quad \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \tau_i W_{i+1,0} y_{i-1} \right| = O_{\mathbb{P}}(1);$$

$$(iv) \quad \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \tau_i W_{i+1} y_{i-1} \right| = O_{\mathbb{P}}(1).$$

*Proof of Lemma 8.4.* We only prove (iii) and (iv), the proof of (i) and (ii) being similar.

(iii) We have

$$\left| \sum_{i=2}^{n-1} \tau_i W_{i+1,0} y_{i-1} \right| \leq \left| \sum_{i=2}^{n-1} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} W_{i+1,0} y_{i-1} \right| + K \sum_{i=2}^{n-1} \Delta_i |W_{i+1,0} y_{i-1}| =: L_1 + L_2.$$

Using the fact that  $\frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \leq 1$  and that for  $i \neq j$ ,

$$\mathbb{E}[W_{i+1,0} y_{i-1} W_{j+1,0} y_{j-1}] = 0,$$

we get

$$\begin{aligned} \mathbb{E}[L_1^2] &\leq \sum_{i=2}^{n-1} \mathbb{E}[W_{i+1,0}^2 y_{i-1}^2] = \sum_{i=2}^{n-1} \mathbb{E}[y_{i-1}^2] \mathbb{E}[W_{i+1,0}^2] \\ &\leq K \sum_{i=2}^{n-1} (1 - e^{-2\theta_0 \Delta_{i+1}}) \leq K \sum_{i=2}^{n-1} \Delta_{i+1} = O(1), \end{aligned}$$

using the fact that  $W_{i+1,0}$  and  $y_{i-1}$  are independent and a Taylor expansion of  $\Delta_{i+1} \mapsto 1 - e^{-2\theta_0 \Delta_{i+1}}$ . Now, by Cauchy-Schwarz,

$$\mathbb{E}(|L_2|) \leq K \sum_{i=2}^{n-1} \Delta_i \mathbb{E}(|W_{i+1,0}| |y_{i-1}|) \leq K \sum_{i=2}^{n-1} \Delta_i \sqrt{\mathbb{E}(W_{i+1,0}^2)} \sqrt{\mathbb{E}(y_{i-1}^2)} = O(1).$$

Hence, since  $L_1$  and  $L_2$  do not depend on  $\theta$  and  $\sigma^2$ , we have  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \tau_i W_{i+1,0} y_{i-1} \right| = O_{\mathbb{P}}(1)$ .

(iv) Now, we use the decomposition

$$W_{i+1} = W_{i+1,0} + (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) y_i \quad (23)$$

to get

$$\begin{aligned} \left| \sum_{i=2}^{n-1} \tau_i W_{i+1} y_{i-1} \right| &= \left| \sum_{i=2}^{n-1} \tau_i W_{i+1,0} y_{i-1} + \sum_{i=2}^{n-1} \tau_i (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) y_i y_{i-1} \right| \\ &\leq \left| \sum_{i=2}^{n-1} \tau_i W_{i+1,0} y_{i-1} \right| + \sum_{i=2}^{n-1} \tau_i |e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}| |y_i y_{i-1}|. \end{aligned}$$

The first sum is  $O_{\mathbb{P}}(1)$  by (iii). The second sum is  $O_{\mathbb{P}}(1)$  by using the fact that  $\sup_{t \in [0,1]} Y^2(t) < \infty$  a.s. and

$$\sup_{n \in \mathbb{N}, i=2, \dots, n-1, (\theta, \sigma^2) \in J} \frac{|\tau_i| |e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}|}{\Delta_{i+1}} \leq K.$$

Hence  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \tau_i W_{i+1} y_{i-1} \right| = O_{\mathbb{P}}(1)$ .  $\square$

We can show, after some tedious but straightforward calculations, the following Taylor expansions.

**Lemma 8.5.** *Let*

$$\alpha_i := (\sigma^2 (1 - e^{-2\theta \Delta_i}))^{-1}, \quad \text{for } i = 2, \dots, n, \quad (24)$$

$$\alpha_{i,0} := (\sigma_0^2 (1 - e^{-2\theta_0 \Delta_i}))^{-1}, \quad \text{for } i = 2, \dots, n, \quad (25)$$

$$q_i := \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + \frac{\Delta_{i-1}}{\Delta_i + \Delta_{i-1}}, \quad \text{for } i = 3, \dots, n-1, \quad (26)$$

$$A_i := \alpha_i + \alpha_{i+1} e^{-2\theta \Delta_{i+1}}, \quad \text{for } i = 2, \dots, n-1, \quad (27)$$

$$B_i := \frac{1}{A_i} + \frac{e^{-2\theta \Delta_i}}{A_{i-1}}, \quad \text{for } i = 3, \dots, n-1, \quad (28)$$

$$C_i := (\alpha_i \alpha_{i+1} e^{-\theta \Delta_{i+1}}) / A_i, \quad \text{for } i = 2, \dots, n-1. \quad (29)$$

Let  $\alpha'_i = \partial \alpha_i / \partial \theta$ ,  $A'_i = \partial A_i / \partial \theta$ ,  $B'_i = \partial B_i / \partial \theta$  and  $C'_i = \partial C_i / \partial \theta$ . Note that we have

$$C'_i = \frac{e^{-\theta \Delta_{i+1}}}{A_i} \left\{ \alpha'_i \alpha_{i+1} + \alpha_i \alpha'_{i+1} - \alpha_i \alpha_{i+1} \Delta_{i+1} - \frac{\alpha_i \alpha_{i+1} A'_i}{A_i} \right\}. \quad (30)$$



We have,

$$\sup_{\substack{n \in \mathbb{N}, i=2, \dots, n, \\ (\theta, \sigma^2) \in J}} \left| \alpha_i - \frac{1}{2\sigma^2\theta\Delta_i} \right| \leq K \quad (31)$$

$$\sup_{\substack{n \in \mathbb{N}, i=2, \dots, n, \\ (\theta, \sigma^2) \in J}} \left| \alpha'_i - \frac{1}{2\sigma^2\theta^2\Delta_i} \right| \leq K \quad (32)$$

$$\alpha_i^2 = \frac{1}{4\sigma^4\theta^2\Delta_i^2} + \delta_{i,n} \frac{1}{\Delta_i} \quad (33)$$

$$\sup_{\substack{n \in \mathbb{N}, i=2, \dots, n-1, \\ (\theta, \sigma^2) \in J}} \left| A_i - \frac{1}{2\sigma^2\theta} \left( \frac{1}{\Delta_i} + \frac{1}{\Delta_{i+1}} \right) \right| \leq K \quad (34)$$

$$\sup_{\substack{n \in \mathbb{N}, i=2, \dots, n-1, \\ (\theta, \sigma^2) \in J}} \left| A'_i - \frac{1}{2\sigma^2\theta^2} \left( \frac{1}{\Delta_i} + \frac{1}{\Delta_{i+1}} \right) \right| \leq K \quad (35)$$

$$\frac{A'_i}{A_i} = \frac{1}{\theta} + \delta_{i,n} \frac{\Delta_i\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} \quad (36)$$

$$\alpha_i B'_i = -\frac{1}{\theta} q_i + \delta_{i,n} (\Delta_{i-1} + \Delta_i + \Delta_{i+1}) \quad (37)$$

$$\alpha'_i B_i = \frac{1}{\theta} q_i + \delta_{i,n} (\Delta_{i-1} + \Delta_i + \Delta_{i+1}) \quad (38)$$

$$\sup_{\substack{n \in \mathbb{N}, i=2, \dots, n-1, \\ (\theta, \sigma^2) \in J}} \left| C'_i - \frac{1}{2\sigma^2\theta^2} \frac{1}{\Delta_i + \Delta_{i+1}} \right| \leq K. \quad (39)$$

## 8.2 Proof of Theorem 3.2

The existence of  $\hat{\theta}$  and  $\hat{\sigma}^2$  is a consequence of the fact that  $S_n$  is a continuous function defined on a compact set. Now, it suffices to show (16) since  $\hat{\theta}_1 \rightarrow \theta_1$  (resp.  $\hat{\sigma}_2^2 \rightarrow \sigma_2^2$ ) is a particular case of (16) with  $b = B = \sigma_1^2$  (resp.  $a = A = \theta_2$ ). Moreover, in view of (5), the result (16) holds if we can show that for every  $\varepsilon > 0$ , a.s.

$$\inf_{\substack{(\theta, \sigma^2) \in J, \\ |\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \varepsilon}} \{S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2)\} \rightarrow \infty \quad (40)$$

where  $(\tilde{\theta}, \tilde{\sigma}^2) \in J$  can be any non-random vector such that  $\tilde{\theta}\tilde{\sigma}^2 = \theta_0\sigma_0^2$ .

Let us compute the difference  $S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2)$  and determine the preponderant terms. After some computations, we naturally have

$$y_i - \frac{\alpha_i e^{-\theta\Delta_i} y_{i-1} + \alpha_{i+1} e^{-\theta\Delta_{i+1}} y_{i+1}}{A_i} = \frac{1}{A_i} (\alpha_i W_i - \alpha_{i+1} e^{-\theta\Delta_{i+1}} W_{i+1}), \quad (41)$$

where  $\alpha_i$  and  $A_i$  have already been defined in Equations (24) and (27). Hence, from Lemma 3.1,

$$S_n(\theta, \sigma^2) = n \log(\sigma^2) + \log(1 - e^{-2\theta\Delta_2}) + \log(1 - e^{-2\theta\Delta_n}) - \sum_{i=2}^{n-1} \log(\sigma^2 A_i) \quad (42)$$

$$+ \frac{(y_1 - e^{-\theta\Delta_2} y_2)^2}{\sigma^2(1 - e^{-2\theta\Delta_2})} + \frac{W_n^2}{\sigma^2(1 - e^{-2\theta\Delta_n})} \quad (43)$$

$$+ \sum_{i=2}^{n-1} \frac{1}{A_i} (\alpha_i W_i - \alpha_{i+1} e^{-\theta\Delta_{i+1}} W_{i+1})^2. \quad (44)$$

In the following, we prove that the terms in (43) and those obtained by developing (44), except one, are  $o(n)$  uniformly in  $(\theta, \sigma^2) \in J$ , a.s. More precisely, we establish the following lemma (see the proof in Section 8.3).

**Lemma 8.6.** *One has*

$$S_n(\theta, \sigma^2) = n \log(\sigma^2) + \log(1 - e^{-2\theta\Delta_2}) + \log(1 - e^{-2\theta\Delta_n}) - \sum_{i=2}^{n-1} \log(\sigma^2 A_i) \\ + \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 + \bar{\delta}_n.$$

As a consequence, we find that,

$$S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2) = n \log \frac{\sigma^2}{\tilde{\sigma}^2} + \log \frac{1 - e^{-2\theta\Delta_2}}{1 - e^{-2\tilde{\theta}\Delta_2}} + \log \frac{1 - e^{-2\theta\Delta_n}}{1 - e^{-2\tilde{\theta}\Delta_n}} - \sum_{i=2}^{n-1} \log \left( \frac{A_i}{\tilde{A}_i} \right) \\ + \sum_{i=3}^{n-1} \left( \alpha_i^2 B_i - \tilde{\alpha}_i^2 \tilde{B}_i \right) W_{i,0}^2 + \bar{\delta}_n \quad (45)$$

where  $\tilde{\alpha}_i$ ,  $\tilde{B}_i$  and  $\tilde{A}_i$  are the analogs of  $\alpha_i$ ,  $B_i$  and  $A_i$  defined in Equations (24), (27) and (28) with  $\theta = \tilde{\theta}$  and  $\sigma^2 = \tilde{\sigma}^2$ . More precisely, they are naturally defined by  $\tilde{\alpha}_i = (\tilde{\sigma}^2(1 - e^{-2\tilde{\theta}\Delta_i}))^{-1}$ ,  $\tilde{A}_i = \tilde{\alpha}_i + \tilde{\alpha}_{i+1} e^{-\tilde{\theta}\Delta_{i+1}}$  and  $\tilde{B}_i = \left[ \frac{1}{\tilde{A}_i} + \frac{e^{-2\tilde{\theta}\Delta_i}}{\tilde{A}_{i-1}} \right]$ .

Using the fact that  $(\tilde{\theta}, \tilde{\sigma}^2)$  has been chosen such as  $\tilde{\theta}\tilde{\sigma}^2 = \theta_0\sigma_0^2$  and making some more computations, we get the following lemma (see the proof in Section 8.3).

**Lemma 8.7.** *Uniformly in  $(\theta, \sigma^2) \in J$ , a.s.*

$$S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2) = n \left[ \frac{\theta_0\sigma_0^2}{\theta\sigma^2} - 1 - \log \frac{\theta_0\sigma_0^2}{\theta\sigma^2} \right] + o(n).$$

Hence by Lemma 8.7, a.s.

$$\inf_{\substack{(\theta, \sigma^2) \in J, \\ |\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \varepsilon}} \{S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2)\} \geq \inf_{\substack{(\theta, \sigma^2) \in J, \\ |\theta\sigma^2 - \tilde{\theta}\tilde{\sigma}^2| \geq \varepsilon}} n \left[ \frac{\theta_0\sigma_0^2}{\theta\sigma^2} - 1 - \log \frac{\theta_0\sigma_0^2}{\theta\sigma^2} \right] + o(n) \quad (46)$$

which by Lemma 8.2, for every  $\varepsilon > 0$ , is strictly positive, for  $n$  large enough, a.s. Then the proof of Theorem 3.2 is now complete.

### 8.3 Proofs of the lemmas of Section 8.2

*Proof of Lemma 8.6.* (i) First, we study the terms in (43). We have, from (31)

$$\begin{aligned} \frac{(y_1 - e^{-\theta\Delta_2}y_2)^2}{\sigma^2(1 - e^{-2\theta\Delta_2})} &\leq K \frac{(y_1 - e^{-\theta\Delta_2}y_2)^2}{\sigma_0^2(1 - e^{-2\theta_0\Delta_2})} \\ &\leq K \frac{(y_1 - e^{-\theta_0\Delta_2}y_2)^2}{\sigma_0^2(1 - e^{-2\theta_0\Delta_2})} + K \frac{(e^{-\theta_0\Delta_2} - e^{-\theta\Delta_2})^2}{\sigma_0^2(1 - e^{-2\theta_0\Delta_2})} y_2^2 \\ &\leq K \sup_{2 \leq i \leq n} \bar{W}_{i,0}^2 + K\Delta_2 \sup_{t \in [0,1]} Y^2(t) \\ &= o(n) \quad \text{a.s.} \end{aligned} \quad (47)$$

from Lemma 8.3. The random variable  $W_n^2/(\sigma^2(1 - e^{-2\theta\Delta_n}))$  can be treated in the same manner leading to the same result.

(ii) Second, we turn to the term in (44) that we aim at approximating by a sum of independent random variables. In this goal, we first show the relation

$$W_i^2 = W_{i,0}^2 + [e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i}]^2 y_{i-1}^2 + 2[e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i}] y_{i-1} W_{i,0}. \quad (49)$$

Hence, by (49), one has

$$\begin{aligned} &\sum_{i=2}^{n-1} \frac{1}{A_i} (\alpha_i W_i - \alpha_{i+1} e^{-\theta\Delta_{i+1}} W_{i+1})^2 \\ &= \sum_{i=2}^{n-1} \frac{\alpha_i^2}{A_i} W_i^2 + \sum_{i=2}^{n-1} \frac{\alpha_{i+1}^2 e^{-2\theta\Delta_{i+1}}}{A_i} W_{i+1}^2 - 2 \sum_{i=2}^{n-1} \frac{\alpha_i \alpha_{i+1} e^{-\theta\Delta_{i+1}}}{A_i} W_i W_{i+1} \\ &= \sum_{i=3}^{n-1} \alpha_i^2 \left[ \frac{1}{A_i} + \frac{e^{-2\theta\Delta_i}}{A_{i-1}} \right] W_i^2 + \frac{\alpha_2^2}{A_2} W_2^2 + \frac{\alpha_n^2 e^{-2\theta\Delta_n}}{A_{n-1}} W_n^2 - 2 \sum_{i=2}^{n-1} C_i W_i W_{i+1} \\ &= \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 + \sum_{i=3}^{n-1} \alpha_i^2 B_i (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2 y_{i-1}^2 + 2 \sum_{i=3}^{n-1} \alpha_i^2 B_i (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i}) W_{i,0} y_{i-1} \\ &\quad + \frac{\alpha_2^2}{A_2} W_2^2 + \frac{\alpha_n^2 e^{-2\theta\Delta_n}}{A_{n-1}} W_n^2 - 2 \sum_{i=2}^{n-1} C_i W_i W_{i+1} \\ &=: \Sigma_1 + \Sigma_2 + 2\Sigma_3 + T_2 + T_n - 2\Sigma_4, \end{aligned} \quad (50)$$

where  $\alpha_i$ ,  $A_i$ ,  $B_i$  and  $C_i$  have been defined in (24), (27), (28) and (29). We prove that all the previous terms are  $o(n)$ , uniformly in  $\theta$  and  $\sigma^2$  a.s, except  $\Sigma_1$  that still appears in the expression of  $S_n(\theta, \sigma^2)$  in Lemma 8.6.

- **Term  $T_2$ :** For  $n$  large enough, since  $\frac{\alpha_2}{A_2} \leq 1$ , we get

$$|T_2| = \left| \frac{\alpha_2^2}{A_2} W_2^2 \right| \leq \alpha_2 W_2^2.$$

Hence, we can show  $\sup_{(\theta, \sigma^2) \in J} |T_2| = o(n)$  a.s. in the same way as for (47).

- **Term  $T_n$ :** For  $n$  large enough, since  $\frac{\alpha_n e^{-2\theta\Delta_n}}{A_{n-1}} \leq 1$ , we get

$$|T_n| = \left| \frac{\alpha_n^2 e^{-2\theta\Delta_n}}{A_{n-1}} W_n^2 \right| \leq \alpha_n W_n^2.$$

Hence, we can show  $\sup_{(\theta, \sigma^2) \in J} |T_n| = o(n)$  a.s. in the same way as for (47).

- **Term  $\Sigma_2$ :** The deterministic quantity  $\frac{\alpha_i}{\Delta_i} (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2$  is bounded for  $n$  large enough, uniformly in  $(\theta, \sigma^2) \in J$  (trivial inequalities and (31)) while  $\frac{\alpha_i}{A_i} \leq 1$  and  $\frac{\alpha_i e^{-2\theta\Delta_i}}{A_{i-1}} \leq 1$ . Then, we are led to

$$\begin{aligned} \sum_{i=3}^{n-1} \alpha_i^2 B_i (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2 y_{i-1}^2 &= \sum_{i=3}^{n-1} \Delta_i \alpha_i B_i \frac{\alpha_i}{\Delta_i} (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2 y_{i-1}^2 \\ &\leq K \sup_{t \in [0,1]} Y(t)^2 \sum_{i=3}^{n-1} \Delta_i = K \sup_{t \in [0,1]} Y(t)^2 \end{aligned}$$

from which, by (19), we deduce  $\sup_{(\theta, \sigma^2) \in J} \Sigma_2 = o(n)$  a.s.

- **Term  $\Sigma_3$ :** By the Cauchy-Schwarz inequality,

$$|\Sigma_3| \leq \left( \sum_{i=3}^{n-1} \frac{\alpha_i^4}{\alpha_{i,0}} B_i^2 (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2 \right)^{1/2} \left( \sum_{i=3}^{n-1} y_{i-1}^2 \overline{W}_{i,0}^2 \right)^{1/2}. \quad (51)$$

As already mentioned, the deterministic term

$$\frac{\alpha_i}{\Delta_i} (e^{-\theta_0\Delta_i} - e^{-\theta\Delta_i})^2$$

is bounded uniformly in  $(\theta, \sigma^2) \in J$ . Furthermore,  $\alpha_i/\alpha_{i,0}$  is bounded uniformly in  $(\theta, \sigma^2) \in J$  from (31). Finally,  $\alpha_i^2 B_i^2 \leq K \alpha_i^2 (1/\alpha_i)^2 = K$ . Hence the first term on the right-hand side of (51) is bounded uniformly in  $(\theta, \sigma^2) \in J$ . Now Lemma 8.3 yields that  $\sum_{i=3}^{n-1} (\overline{W}_{i,0}^2 - 1) = o(n^{(1/2)+\alpha})$  a.s. for any  $\alpha > 0$  leading to  $\sum_{i=3}^{n-1} \overline{W}_{i,0}^2 = O(n)$  a.s. and

$$\left( \sum_{i=3}^{n-1} y_{i-1}^2 \overline{W}_{i,0}^2 \right)^{1/2} \leq \left( \sup_{t \in [0,1]} Y^2(t) \sum_{i=3}^{n-1} \overline{W}_{i,0}^2 \right)^{1/2} = O(n^{1/2}) \text{ a.s.}$$

As a consequence,  $\sup_{(\theta, \sigma^2) \in J} |\Sigma_3| = O(n^{1/2})$  a.s. and naturally  $\sup_{(\theta, \sigma^2) \in J} |\Sigma_3| = o(n)$  a.s.

• **Term  $\Sigma_4$ :** Using the trivial equality  $ab = ab - a_0b_0 + a_0b_0 = a_0b_0 + a_0(b - b_0) + b(a - a_0)$ , one gets

$$\begin{aligned} W_i W_{i+1} &= [y_i - e^{-\theta \Delta_i} y_{i-1}][y_{i+1} - e^{-\theta \Delta_{i+1}} y_i] = [y_i - e^{-\theta_0 \Delta_i} y_{i-1}][y_{i+1} - e^{-\theta_0 \Delta_{i+1}} y_i] \\ &+ [y_i - e^{-\theta_0 \Delta_i} y_{i-1}](e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) y_i + [y_{i+1} - e^{-\theta \Delta_{i+1}} y_i](e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) y_{i-1} \\ &= W_{i,0} W_{i+1,0} + (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) W_{i,0} y_i + (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_{i+1} y_{i-1}. \end{aligned} \quad (52)$$

Thus  $\Sigma_4$  rewrites

$$\sum_{i=2}^{n-1} C_i W_{i,0} W_{i+1,0} + \sum_{i=2}^{n-1} C_i (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) W_{i,0} y_i + \sum_{i=2}^{n-1} C_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_{i+1} y_{i-1}. \quad (53)$$

We can show that

$$\frac{C_i}{\alpha_{i,0}^{1/2} \alpha_{i+1,0}^{1/2}} = \frac{\sigma_0^2 \theta_0}{\sigma^2 \theta^2} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} + \delta_{i,n}(\Delta_i + \Delta_{i+1}).$$

Hence the first random variable of (53) rewrites

$$\begin{aligned} &\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i W_{i,0} W_{i+1,0} \right| = \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \frac{C_i}{\alpha_{i,0}^{1/2} \alpha_{i+1,0}^{1/2}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| \\ &\leq \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} \frac{\sigma_0^2 \theta_0}{\sigma^2 \theta^2} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| + \sum_{i=2}^{n-1} \delta_{i,n}(\Delta_i + \Delta_{i+1}) |\bar{W}_{i,0} \bar{W}_{i+1,0}| \\ &\leq K \left| \sum_{i=2}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| + \sum_{i=2}^{n-1} \delta_{i,n}(\Delta_i + \Delta_{i+1}) |\bar{W}_{i,0} \bar{W}_{i+1,0}| \\ &= K \left| \sum_{i=2}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| + \bar{\delta}_n, \end{aligned}$$

from (21). Now, we have

$$\begin{aligned} &\left| \sum_{i=2}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| \\ &\leq \left| \sum_{\substack{i=2, \dots, n-1 \\ i \text{ odd}}} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right| + \left| \sum_{\substack{i=2, \dots, n-1 \\ i \text{ even}}} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} \right|. \end{aligned}$$

In each of the two sums above, the summands constitute two triangular arrays of independent random variables. Thus, applying Theorem 2.1 in [17] with  $a_n = n$ , each of the two sums is a  $o(n)$  a.s. Hence finally  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i W_{i,0} W_{i+1,0} \right| = o(n)$  a.s.

Let us now address the second term in (53) that by the Cauchy-Schwarz inequality is bounded from above by

$$\left( \sum_{i=2}^{n-1} \frac{C_i^2}{\alpha_{i,0}} (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}})^2 \right)^{1/2} \left( \sum_{i=2}^{n-1} y_i^2 \bar{W}_{i,0}^2 \right)^{1/2} = o(n) \quad \text{a.s.}$$

where the last equality comes from similar computations as from the term  $\Sigma_3$  above, and from the fact that

$$\sup_{n \in \mathbb{N}} \sup_{2 \leq i \leq n-1} \sup_{(\theta, \sigma^2) \in J} \left| \frac{C_i^2}{\alpha_{i,0} \Delta_{i+1}} (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}})^2 \right| \leq K.$$

The third term in the right-hand side of (53) yields

$$\begin{aligned} & \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_{i+1} y_{i-1} \right| \leq \\ & \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_{i+1,0} y_{i-1} \right| + \sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i})^2 y_i y_{i-1} \right|. \end{aligned} \quad (54)$$

Since trivially  $C_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) = \delta_{i,n}$ , the second term in (54) is bounded by  $K \sum_{i=2}^{n-1} \Delta_i |y_i y_{i-1}| = O(1)$  a.s.

The first term in (54) is bounded by

$$\begin{aligned} K \sum_{i=2}^{n-1} |W_{i+1,0} y_{i-1}| & \leq K \sup_{t \in [0,1]} |Y(t)| \sum_{i=2}^{n-1} |\bar{W}_{i+1,0}| \alpha_{i,0}^{-1/2} \\ & \leq K \sup_{t \in [0,1]} |Y(t)| \sqrt{\sum_{i=2}^{n-1} \bar{W}_{i+1,0}^2} \sqrt{\sum_{i=2}^{n-1} \alpha_{i,0}^{-1}} \\ & = O(\sqrt{n}) \quad \text{a.s.} \end{aligned}$$

since  $\sum_{i=2}^{n-1} \bar{W}_{i+1,0}^2 = O(n)$  a.s. has been shown when handling  $\Sigma_3$  after (51).

One may conclude that  $\sup_{(\theta, \sigma^2) \in J} \Sigma_4 = o(n)$  a.s. The proof of Lemma 8.6 is then complete.  $\square$

*Proof of Lemma 8.7.* We address each of the terms in (45).

• Using Lemma 8.1(i), we get that

$$\sup_{\theta \in [a, A]} \left| \log \frac{1 - e^{-2\theta \Delta_2}}{1 - e^{-2\tilde{\theta} \Delta_2}} \right| = O(1).$$

In the same way,  $\sup_{\theta \in [a, A]} \left| \log \frac{1 - e^{-2\theta\Delta_n}}{1 - e^{-2\tilde{\theta}\Delta_n}} \right| = O(1)$ .

- We have, using Lemmas 8.1 (i) and (ii),

$$\begin{aligned} \log(A_i/\tilde{A}_i) &= \log \left[ \left( \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right) / \left( \frac{1}{1 - e^{-2\tilde{\theta}\Delta_i}} + \frac{e^{-2\tilde{\theta}\Delta_{i+1}}}{1 - e^{-2\tilde{\theta}\Delta_{i+1}}} \right) \right] \\ &= \log \frac{1 - e^{-2\tilde{\theta}\Delta_i}}{1 - e^{-2\theta\Delta_i}} + \log \left[ \left( 1 + e^{-2\theta\Delta_{i+1}} \frac{1 - e^{-2\theta\Delta_i}}{1 - e^{-2\theta\Delta_{i+1}}} \right) / \left( 1 + e^{-2\tilde{\theta}\Delta_{i+1}} \frac{1 - e^{-2\tilde{\theta}\Delta_i}}{1 - e^{-2\tilde{\theta}\Delta_{i+1}}} \right) \right] \\ &= \log \frac{\tilde{\theta}}{\theta} + \delta_{i,n}(\Delta_i + \Delta_{i+1}). \end{aligned}$$

Thus, by summation we have,

$$\sup_{\theta \in [a, A]} \left| \sum_{i=2}^{n-1} \log \frac{A_i}{\tilde{A}_i} - n \log \frac{\tilde{\theta}}{\theta} \right| = O(1) = o(n).$$

- We want to show that

$$\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 - n \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \right| = o(n) \quad \text{a.s.}$$

By (28), one has

$$\begin{aligned} \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 &= \sum_{i=3}^{n-1} \frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} \bar{W}_{i,0}^2 + \sum_{i=3}^{n-1} \frac{\alpha_i^2}{\alpha_{i,0}} \frac{e^{-2\theta\Delta_i}}{A_{i-1}} \bar{W}_{i,0}^2 \\ &= \sum_{i=3}^{n-2} \left( \frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} \bar{W}_{i,0}^2 + \frac{\alpha_{i+1}^2}{\alpha_{i+1,0}} \frac{e^{-2\theta\Delta_{i+1}}}{A_i} \bar{W}_{i+1,0}^2 \right) + \frac{\alpha_3^2 e^{-2\theta\Delta_3}}{A_2} W_{3,0}^2 + \frac{\alpha_{n-1}^2}{A_{n-1}} W_{n-1,0}^2. \end{aligned}$$

Then we use (31) to develop  $\alpha_i/\alpha_{i,0}$  (respectively  $\alpha_{i+1}/\alpha_{i+1,0}$ ) and Lemma 8.1 (ii) to develop  $\alpha_i/A_i$  (respectively  $\alpha_{i+1}e^{-2\theta\Delta_{i+1}}/A_i$ ). We get

$$\frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} = \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + \delta_{i,n}(\Delta_i + \Delta_{i+1}) \quad (55)$$

$$\frac{\alpha_{i+1}^2}{\alpha_{i+1,0}} \frac{e^{-2\theta\Delta_{i+1}}}{A_i} = \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} + \delta_{i,n}(\Delta_i + \Delta_{i+1}). \quad (56)$$

In addition, we easily show, as in (47), that  $\sup_{(\theta, \sigma^2) \in J} \left| \frac{\alpha_3^2 e^{-2\theta\Delta_3}}{A_2} W_{3,0}^2 \right| = o(n)$  a.s. and that

$\sup_{(\theta, \sigma^2) \in J} \left| \frac{\alpha_{n-1}^2}{A_{n-1}} W_{n-1,0}^2 \right| = o(n)$  a.s. Then,

$$\begin{aligned}
& \left| \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 - n \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \right| = \left| \sum_{i=3}^{n-2} \left( \frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} \overline{W}_{i,0}^2 + \frac{\alpha_{i+1}^2}{\alpha_{i+1,0}} \frac{e^{-2\theta \Delta_{i+1}}}{A_i} \overline{W}_{i+1,0}^2 \right) - \sum_{i=3}^{n-2} \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} + \bar{\delta}_n \right| \\
&= \left| \sum_{i=3}^{n-2} \left( \frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} \overline{W}_{i,0}^2 + \frac{\alpha_{i+1}^2}{\alpha_{i+1,0}} \frac{e^{-2\theta \Delta_{i+1}}}{A_i} \overline{W}_{i+1,0}^2 \right) - \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \sum_{i=3}^{n-2} \left( \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \right) + \bar{\delta}_n \right| \\
&\leq \sum_{i=3}^{n-2} \left| \frac{\alpha_i^2}{\alpha_{i,0}} \frac{1}{A_i} - \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} \right| \overline{W}_{i,0}^2 + \sum_{i=3}^{n-2} \left| \frac{\alpha_{i+1}^2}{\alpha_{i+1,0}} \frac{e^{-2\theta \Delta_{i+1}}}{A_i} - \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \right| \overline{W}_{i+1,0}^2 \\
&\quad + \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \left| \sum_{i=3}^{n-2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i,0}^2 - 1) \right| + \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \left| \sum_{i=3}^{n-2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i+1,0}^2 - 1) \right| + |\bar{\delta}_n| \\
&\leq K \sum_{i=3}^{n-2} (\Delta_i + \Delta_{i+1}) \overline{W}_{i,0}^2 + K \sum_{i=3}^{n-2} (\Delta_i + \Delta_{i+1}) \overline{W}_{i+1,0}^2 + |\bar{\delta}_n| \tag{57} \\
&\quad + \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \left| \sum_{i=3}^{n-2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i,0}^2 - 1) \right| + \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \left| \sum_{i=3}^{n-2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i+1,0}^2 - 1) \right|.
\end{aligned}$$

Let us show that the terms in the right-hand side of (57) are a.s.  $o(n)$ . We have

$$\sum_{i=3}^{n-2} (\Delta_i + \Delta_{i+1}) \overline{W}_{i,0}^2 \leq K \sup_{i=3, \dots, n-2} \overline{W}_{i,0}^2 \leq 1 + \sup_{i=3, \dots, n-2} |\overline{W}_{i,0}^2 - 1| = o(n),$$

a.s. from Lemma 8.3. Similarly

$$\sum_{i=3}^{n-2} (\Delta_i + \Delta_{i+1}) \overline{W}_{i+1,0}^2 = o(n)$$

a.s. Also, using theorem 2.1 in [17] with  $a_n = n$ , we have

$$\sum_{i=3}^{n-2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i,0}^2 - 1) = o(n)$$

a.s. and

$$\sum_{i=3}^{n-2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (\overline{W}_{i+1,0}^2 - 1) = o(n)$$

a.s. Hence finally

$$\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} \alpha_i^2 B_i W_{i,0}^2 - n \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} \right| = o(n) \quad \text{a.s.}$$



• We can now conclude the proof. We have

$$S_n(\theta, \sigma^2) - S_n(\tilde{\theta}, \tilde{\sigma}^2) = n \log \frac{\sigma^2}{\tilde{\sigma}^2} - n \log \frac{\tilde{\theta}}{\theta} + n \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - n \frac{\theta_0 \sigma_0^2}{\tilde{\theta} \tilde{\sigma}^2} + \delta_n = n \left( \log \frac{\sigma^2 \theta}{\sigma_0^2 \theta_0} + \frac{\theta_0 \sigma_0^2}{\theta \sigma^2} - 1 \right) + \delta_n,$$

by reminding that  $\tilde{\theta} \tilde{\sigma}^2 = \theta_0 \sigma_0^2$ . The proof of Lemma 8.7 is thus complete.  $\square$

## 8.4 Proof of Theorem 4.1

Let us first prove (9) in the case  $aB < \theta_0 \sigma_0^2$ ;  $Ab > \theta_0 \sigma_0^2$ . We shall then discuss the other cases at the end. In that view, let

$$\psi(\theta, \sigma^2) = \frac{\partial}{\partial \theta} S_n(\theta, \sigma^2). \quad (58)$$

Then from Theorem 3.2, a.s. for  $n$  large enough,  $\hat{\theta} \in (a, A)$ . Thus a.s. for  $n$  large enough  $(\hat{\theta}, \hat{\sigma}^2)$  satisfies  $\psi(\hat{\theta}, \hat{\sigma}^2) = 0$ . We shall approximate  $\psi(\theta, \sigma^2)$  uniformly in  $(\theta, \sigma^2) \in J$ . Starting from (42), (43), (44) and (50) we can write

$$\begin{aligned} \psi(\theta, \sigma^2) &= \frac{2\Delta_2 e^{-2\theta\Delta_2}}{1 - e^{-2\theta\Delta_2}} + \frac{2\Delta_n e^{-2\theta\Delta_n}}{1 - e^{-2\theta\Delta_n}} + \frac{2\Delta_2 e^{-\theta\Delta_2} y_2 (y_1 - e^{-\theta\Delta_2} y_2)}{\sigma^2 (1 - e^{-2\theta\Delta_2})} \\ &\quad - \frac{2\Delta_2 e^{-2\theta\Delta_2} (y_1 - e^{-\theta\Delta_2} y_2)^2}{\sigma^2 (1 - e^{-2\theta\Delta_2})^2} + \frac{2\Delta_n e^{-\theta\Delta_n} y_{n-1} W_n}{\sigma^2 (1 - e^{-2\theta\Delta_n})} - \frac{2\Delta_n e^{-2\theta\Delta_n} W_n^2}{\sigma^2 (1 - e^{-2\theta\Delta_n})^2} \\ &\quad - \sum_{i=2}^{n-1} \frac{A'_i}{A_i} + \Sigma'_1 + \Sigma'_2 + 2\Sigma'_3 + T'_2 + T'_n - 2\Sigma'_4 \end{aligned}$$

where

$$\begin{aligned} \circ \Sigma'_1 &= \frac{\partial}{\partial \theta} \Sigma_1 = \sum_{i=3}^{n-1} \alpha_i D_i W_{i,0}^2; \\ \circ \Sigma'_2 &= \frac{\partial}{\partial \theta} \Sigma_2 = \sum_{i=3}^{n-1} \alpha_i [D_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) + 2\alpha_i B_i \Delta_i e^{-\theta \Delta_i}] y_{i-1}^2 (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}); \\ \circ \Sigma'_3 &= \frac{\partial}{\partial \theta} \Sigma_3 = \sum_{i=3}^{n-1} \alpha_i [D_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) + \alpha_i B_i \Delta_i e^{-\theta \Delta_i}] W_{i,0} y_{i-1}; \\ \circ T'_2 &= \frac{\partial}{\partial \theta} T_2 = 2\alpha'_2 \alpha_2 \frac{W_2^2}{A_2} + \alpha_2^2 \left( \frac{2\Delta_2 e^{-\theta \Delta_2} y_1 W_2}{A_2} - W_2^2 \frac{A'_2}{A_2^2} \right); \\ \circ T'_n &= \frac{\partial}{\partial \theta} T_n = \frac{\alpha_n e^{-2\theta \Delta_n}}{A_{n-1}} \left[ (2\alpha'_n - 2\alpha_n \Delta_n - \frac{A'_{n-1}}{A_{n-1}} \alpha_n) W_n + 2\alpha_n \Delta_n e^{-\theta \Delta_n} y_{n-1} \right] W_n; \\ \circ \Sigma'_4 &= \frac{\partial}{\partial \theta} \Sigma_4 = \sum_{i=2}^{n-1} C'_i W_i W_{i+1} + \sum_{i=2}^{n-1} C_i \Delta_i e^{-\theta \Delta_i} W_{i+1} y_{i-1} + \sum_{i=2}^{n-1} C_i \Delta_{i+1} e^{-\theta \Delta_{i+1}} W_i y_i \end{aligned}$$

where  $C'_i$  is the derivative of  $C_i$  w.r.t.  $\theta$  defined in (30) and

$$D_i := 2\alpha'_i B_i + \alpha_i B'_i, \quad \text{for } i = 3, \dots, n-1. \quad (59)$$

First, we consider the terms  $\Sigma'_1$  and  $\Sigma'_4$  in the following lemma (proved in Section 8.6).

**Lemma 8.8.** *We have*

$$\Sigma'_1 = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=3}^{n-1} q_i \overline{W}_{i,0}^2 + \delta_n$$

and

$$\Sigma'_4 = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=3}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0} + \delta_n,$$

where  $q_i$  and  $C'_i$  have been defined in (26) and (30).

Now we prove that the remaining terms in  $\psi(\sigma^2, \theta)$  are  $O_{\mathbb{P}}(1)$ , uniformly in  $(\theta, \sigma^2) \in J$ , at the exception of  $\sum_{i=2}^{n-1} A'_i/A_i$ , leading to the following lemma (proved in Section 8.6).

**Lemma 8.9.** *We obtain*

$$\psi(\theta, \sigma^2) = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=3}^{n-1} \left[ q_i \overline{W}_{i,0}^2 - 2 \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0} \right] - \frac{n}{\theta} + \delta_n.$$

Since  $\hat{\theta}^2 \hat{\sigma}^2 \psi(\hat{\theta}, \hat{\sigma}^2) = 0$  with probability going to 1, and since we can show that  $\sum_{i=3}^{n-1} q_i = n + O(1)$ , we have

$$n(\hat{\theta}^2 \hat{\sigma}^2 - \theta_0 \sigma_0^2) = \theta_0 \sigma_0^2 \sum_{i=3}^{n-1} \left[ q_i (\overline{W}_{i,0}^2 - 1) - 2 \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0} \right] + O_{\mathbb{P}}(1). \quad (60)$$

We want to establish a Central Limit Theorem for  $\sqrt{n}(\hat{\theta}^2 \hat{\sigma}^2 - \theta_0 \sigma_0^2)$ . In that view, we define  $X_i := q_i (\overline{W}_{i,0}^2 - 1) - 2 \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0}$  and we apply Theorem 2.1 in [26] for weakly dependent variables (since  $X_i$  is not necessarily independent with  $X_{i-1}$  and  $X_{i+1}$  but is independent with  $X_k$  for  $|i - k| \geq 2$ ).

Note that we can show easily that  $\tau_n^2 = \frac{1}{n} \text{Var}(\sum_{i=3}^{n-1} X_i)$ , and assume

$$\sqrt{n} \frac{(\hat{\theta}^2 \hat{\sigma}^2 - \theta_0 \sigma_0^2)}{\theta_0 \sigma_0^2 \tau_n} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1).$$

By Proposition 4.3, we can extract a subsequence  $\epsilon_n$  so that  $\tau_{\epsilon_n}^2 \xrightarrow[n \rightarrow \infty]{} \tau^2$  with  $\tau^2 \in [2, 4]$  and so that

$$\sqrt{\epsilon_n} \frac{(\hat{\theta}^2 \hat{\sigma}^2 - \theta_0 \sigma_0^2)}{\theta_0 \sigma_0^2 \tau_{\epsilon_n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1).$$

The triangular array  $(X_i/\sqrt{\varepsilon_n})_{i=3,\dots,\varepsilon_n-1}$  satisfies the conditions of [26, Theorem 2.1], thus we obtain

$$\frac{1}{\sqrt{\varepsilon_n}} \sum_{i=3}^{\varepsilon_n-1} X_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \tau^2).$$

Now, from (60),

$$\sqrt{\varepsilon_n} \frac{(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2)}{\theta_0\sigma_0^2\tau_{\varepsilon_n}} = \frac{1}{\sqrt{\varepsilon_n}} \sum_{i=3}^{\varepsilon_n-1} \frac{X_i}{\tau_{\varepsilon_n}} + o_{\mathbb{P}}(1) = \frac{1}{\sqrt{\varepsilon_n}} \sum_{i=3}^{\varepsilon_n-1} \frac{X_i}{\tau} + \left( \frac{1}{\tau_{\varepsilon_n}} - \frac{1}{\tau} \right) \frac{1}{\sqrt{\varepsilon_n}} \sum_{i=3}^{\varepsilon_n-1} X_i + o_{\mathbb{P}}(1).$$

Since  $\frac{1}{\sqrt{\varepsilon_n}} \sum_{i=3}^{\varepsilon_n-1} X_i = O_{\mathbb{P}}(1)$  and  $\left( \frac{1}{\tau_{\varepsilon_n}} - \frac{1}{\tau} \right) = o(1)$ , we get by Slutsky's lemma

$$\sqrt{\varepsilon_n} \frac{(\hat{\theta}\hat{\sigma}^2 - \theta_0\sigma_0^2)}{\theta_0\sigma_0^2\tau_{\varepsilon_n}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1),$$

which is contradictory and ends the proof of (9).

Now (10) is under consideration only when  $b = B = \sigma_1^2$  and so when  $aB < \theta_0\sigma_0^2$ ;  $Ab > \theta_0\sigma_0^2$ . Thus (10) is a special case of (9). Now, when  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$ , we have almost surely for  $n$  large enough  $(\partial/\partial\sigma^2)S_n(\hat{\theta}, \hat{\sigma}^2) = 0$ , so that the estimator  $\hat{\sigma}_2^2$  can be expressed explicitly, by differentiating the terms in (42), (43) and (44) w.r.t.  $\sigma^2$ . Hence, (9) can be proved in the case  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$  by using identical techniques as in the case  $aB < \theta_0\sigma_0^2$ ;  $Ab > \theta_0\sigma_0^2$ . We omit the details to save space. Finally, (11) is under consideration only when  $a = A = \theta_2$  and so when  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$ . Thus (11) is a special case of (9).

## 8.5 Proof of Propositions 4.3 and 4.5

*Proof of Proposition 4.3.* We have

$$\tau_n^2 = \frac{2}{n} \sum_{i=3}^{n-1} \left[ q_i^2 + 2 \frac{\Delta_i \Delta_{i+1}}{(\Delta_i + \Delta_{i+1})^2} \right].$$

(i) Upper bound for  $\tau_n^2$ . Let  $a_i = \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}}$ , note that  $\frac{\Delta_{i-1}}{\Delta_i + \Delta_{i-1}} = 1 - a_{i-1}$  and  $q_i = a_i + 1 - a_{i-1}$ . First, we have after some trivial computations,

$$\begin{aligned} \tau_n^2 &= \frac{2}{n} \sum_{i=3}^{n-1} [(a_i + 1 - a_{i-1})^2 + 2a_i(1 - a_i)] \\ &= \frac{2}{n} \sum_{i=3}^{n-1} (1 + 2a_i - 2a_i a_{i-1}) + o(1) \\ &\leq 2 + 4m + o(1), \end{aligned} \tag{61}$$

where  $m := \frac{1}{n-3} \sum_{i=3}^{n-1} a_i$ .

Also, since for  $k = 2, \dots, n-1$ ,  $0 \leq a_k \leq 1$ , we have  $1 + 2(1 - a_{i-1})a_i \leq 3 - 2a_{i-1}$ . Thus, from (61),

$$\tau_n^2 \leq \frac{2}{n} \sum_{i=3}^{n-1} (3 - 2a_{i-1}) + o(1) = 6 - 4m + o(1).$$

Finally,  $\tau_n^2 \leq \min(2 + 4m, 6 - 4m) + o(1)$ . Since  $\sup_{m \in [0,1]} \min(2 + 4m, 6 - 4m) = 4$ ,  $\tau_n^2 \leq 4 + o(1)$ .

(ii) Lower bound for  $\tau_n^2$ . Note that  $\frac{1}{n} \sum_{i=3}^{n-1} q_i = 1 + o(1)$ . Since  $\frac{\Delta_i \Delta_{i+1}}{(\Delta_i + \Delta_{i+1})^2} \geq 0$ , we get

$$\tau_n^2 \geq \frac{2}{n} \sum_{i=3}^{n-1} q_i^2 \geq 2 \left( \frac{1}{n} \sum_{i=3}^{n-1} q_i \right)^2 + o(1) = 2 + o(1). \quad (62)$$

□

*Proof of Proposition 4.5.* (i) After some computation, we have

$$\tau_n^2 = 4\gamma_n^2 - 4\gamma_n + 4 + o(1).$$

Since  $\gamma_n = 1/n$ , then  $\tau_n^2 \xrightarrow{n \rightarrow \infty} 4$ .

(ii) We have

$$\tau_n^2 \leq \frac{2}{n} \sum_{i=[n^\alpha]+2}^{n-1} \left( \left( \frac{1}{i+2} + \frac{i}{i+1} \right)^2 + \frac{2(i+1)}{(i+2)^2} \right) + o(1) \leq 2 + o(1).$$

As a consequence, this particular design realizes  $\tau_n^2 = 2 + o(1)$  by (62). □

## 8.6 Proofs of the lemmas of Section 8.4

*Proof of Lemma 8.8.* (i) From (37) and (38),

$$\Sigma'_1 = \frac{1}{\theta} \sum_{i=3}^{n-1} \frac{\alpha_i}{\alpha_{i,0}} q_i \overline{W}_{i,0}^2 + \sum_{i=3}^{n-1} \frac{\alpha_i}{\alpha_{i,0}} \delta_{i,n} (\Delta_{i-1} + \Delta_i + \Delta_{i+1}) \overline{W}_{i,0}^2$$

where  $q_i$  has been defined in (26). By (31), we have

$$\frac{\alpha_i}{\alpha_{i,0}} = \frac{\theta_0 \sigma_0^2}{\sigma^2 \theta^2} + \delta_{i,n} \Delta_i.$$

Moreover, since  $\mathbb{E}[\overline{W}_{i,0}^2] = 1$ , one clearly has

$$\sum_{i=3}^{n-1} \frac{\alpha_i}{\alpha_{i,0}} \delta_{i,n} (\Delta_{i-1} + \Delta_i + \Delta_{i+1}) \overline{W}_{i,0}^2 = O_{\mathbb{P}}(1)$$

that leads to the desired result.

(ii) Now we study the first sum of  $\Sigma'_4$  that rewrites  $\sum_{i=2}^{n-1} C'_i (M_{i,1} + M_{i,2} + M_{i,3})$  using (52) and where  $C'_i$  has been defined in (30) and

$$\begin{cases} M_{i,1} := \frac{\overline{W}_{i,0} \overline{W}_{i+1,0}}{\alpha_{i,0}^{1/2} \alpha_{i+1,0}^{1/2}} \\ M_{i,2} := (e^{-\theta_0 \Delta_{i+1}} - e^{-\theta \Delta_{i+1}}) W_{i,0} y_i \\ M_{i,3} := (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_{i+1} y_{i-1}. \end{cases}$$

• First, we consider  $\sum_{i=2}^{n-1} C'_i M_{i,1}$ . By (39) and (31) we can show

$$\frac{C'_i}{\alpha_{i,0}^{1/2} \alpha_{i+1,0}^{1/2}} = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} + \delta_{i,n} (\Delta_i + \Delta_{i+1}).$$

Furthermore we have

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{i=2}^{n-1} \delta_{i,n} (\Delta_i + \Delta_{i+1}) \overline{W}_{i,0} \overline{W}_{i+1,0} \right| \right] &\leq K \sum_{i=2}^{n-1} (\Delta_i + \Delta_{i+1}) \sqrt{\mathbb{E}[\overline{W}_{i,0}^2] \mathbb{E}[\overline{W}_{i+1,0}^2]} \\ &= K \sum_{i=2}^{n-1} (\Delta_i + \Delta_{i+1}) \leq K. \end{aligned}$$

Thus

$$\sum_{i=2}^{n-1} \frac{C'_i}{\alpha_{i,0}^{1/2} \alpha_{i+1,0}^{1/2}} \overline{W}_{i,0} \overline{W}_{i+1,0} = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=2}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0} + \delta_n.$$

Hence

$$\sum_{i=2}^{n-1} C'_i M_{i,1} = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=2}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \overline{W}_{i,0} \overline{W}_{i+1,0} + \delta_n.$$

• Second, one clearly has

$$\sum_{i=2}^{n-1} C'_i M_{i,2} = \frac{\theta - \theta_0}{2\theta^2 \sigma^2} \sum_{i=2}^{n-1} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n} (\Delta_i + \Delta_{i+1})) W_{i,0} y_i.$$

Hence  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C'_i M_{i,2} \right| = O_{\mathbb{P}}(1)$  by Lemma 8.4 (i).

- Third, we get

$$\sum_{i=2}^{n-1} C'_i M_{i,3} = \frac{\theta - \theta_0}{2\theta^2 \sigma^2} \sum_{i=2}^{n-1} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n}(\Delta_i + \Delta_{i+1})) W_{i+1} y_{i-1}$$

and  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C'_i M_{i,3} \right| = O_{\mathbb{P}}(1)$  by Lemma 8.4 (iv).

(iii) We now consider the second and third sums in  $\Sigma'_4$ . Using (31) and (34), we can show

$$C_i \Delta_i e^{-\theta \Delta_i} = \frac{1}{2\theta \sigma^2} \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n}(\Delta_i + \Delta_{i+1}))$$

and

$$C_i \Delta_{i+1} e^{-\theta \Delta_{i+1}} = \frac{1}{2\theta \sigma^2} \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} (1 + \delta_{i,n}(\Delta_i + \Delta_{i+1})).$$

Hence by Lemma 8.4 (iv) and (ii),  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=2}^{n-1} C_i (\Delta_i e^{-\theta \Delta_i} W_{i+1} y_{i-1} + \Delta_{i+1} e^{-\theta \Delta_{i+1}} W_i y_i) \right| = O_{\mathbb{P}}(1)$ .  $\square$

*Proof of Lemma 8.9.* • We have

$$\sup_{a \leq \theta \leq A} \left| \frac{2\Delta_2 e^{-2\theta \Delta_2}}{1 - e^{-2\theta \Delta_2}} + \frac{2\Delta_n e^{-2\theta \Delta_n}}{1 - e^{-2\theta \Delta_n}} \right| = O(1).$$

- For  $n$  large enough,

$$\sup_{(\theta, \sigma^2) \in J} \left| \frac{2\Delta_2 e^{-\theta \Delta_2} y_2 [y_1 - e^{-\theta \Delta_2} y_2]}{\sigma^2 (1 - e^{-2\theta \Delta_2})} \right| \leq K \sup_{a \leq \theta \leq A} |y_2| |y_1 - e^{-\theta \Delta_2} y_2| \leq K \sup_{t \in [0,1]} Y(t)^2 = O_{\mathbb{P}}(1).$$

- Using  $W_i^2 = W_{i,0}^2 + (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) y_{i-1} W_i + W_{i,0} (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) y_{i-1}$  we can easily show  $\sup_{(\theta, \sigma^2) \in J} \left| \frac{2\Delta_2 e^{-2\theta \Delta_2} (y_1 - e^{-\theta \Delta_2} y_2)^2}{\sigma^2 (1 - e^{-2\theta \Delta_2})^2} \right| = O_{\mathbb{P}}(1)$ ,  $\sup_{(\theta, \sigma^2) \in J} \left| \frac{2\Delta_n e^{-\theta \Delta_n} y_{n-1} W_n}{\sigma^2 (1 - e^{-2\theta \Delta_n})} \right| = O_{\mathbb{P}}(1)$  and  $\sup_{(\theta, \sigma^2) \in J} \left| \frac{2\Delta_n e^{-2\theta \Delta_n} W_n^2}{\sigma^2 (1 - e^{-2\theta \Delta_n})^2} \right| = O_{\mathbb{P}}(1)$ .

• **Term  $\Sigma'_2$ :** First, using (37), (38) and the definition (26) of  $q_i$ , the deterministic quantity  $D_i$  is bounded uniformly in  $(\theta, \sigma^2) \in J$ , and so is  $(e^{-2\theta_0 \Delta_i} - e^{-2\theta \Delta_i})^2 \alpha_i / \Delta_i$  from (31). By (19), we are led to

$$\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} \Delta_i \frac{\alpha_i}{\Delta_i} (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i})^2 y_{i-1}^2 D_i \right| = O_{\mathbb{P}}(1).$$

Similarly,  $\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} 2\alpha_i^2 B_i \Delta_i e^{-\theta \Delta_i} (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) y_{i-1}^2 \right| = O_{\mathbb{P}}(1)$  and thus  $\sup_{(\theta, \sigma^2) \in J} |\Sigma'_2| = O_{\mathbb{P}}(1)$ .

• **Term  $\Sigma'_3$ :** First, from (37), (38) and (31), we have

$$\sup_{n \in \mathbb{N}, i=2, \dots, n-1, (\theta, \sigma^2) \in J} \left| \alpha_i D_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) - \frac{\theta - \theta_0}{2\sigma^2 \theta^2} q_i \right| \leq K.$$

Hence, proceeding as in the proof of Lemma 8.4, we can show

$$\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} \alpha_i D_i (e^{-\theta_0 \Delta_i} - e^{-\theta \Delta_i}) W_i y_{i-1} \right| = O_{\mathbb{P}}(1).$$

Second, we can show

$$\sup_{n \in \mathbb{N}, i=3, \dots, n-1, (\theta, \sigma^2) \in J} \left| \alpha_i^2 B_i \Delta_i e^{-\theta \Delta_i} - \frac{1}{2\sigma^2 \theta} \left( \frac{\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + \frac{\Delta_{i-1}}{\Delta_i + \Delta_{i-1}} \right) \right| \leq K.$$

Hence we can show

$$\sup_{(\theta, \sigma^2) \in J} \left| \sum_{i=3}^{n-1} \alpha_i^2 B_i \Delta_i e^{-\theta \Delta_i} W_{i,0} y_{i-1} \right| = O_{\mathbb{P}}(1),$$

as in the proof of Lemma 8.4. Hence finally,  $\sup_{(\theta, \sigma^2) \in J} \Sigma'_3 = O_{\mathbb{P}}(1)$ .

• **Term  $T'_2$ :** From (27), we have

$$|T'_2| \leq 2|\alpha'_2|W_2^2 + |\alpha_2|\Delta_2 e^{-\theta \Delta_2} |y_1| + W_2^2 |A'_2|.$$

We can show

$$\sup_{(\theta, \sigma^2) \in J} (|\alpha'_2|W_2^2 + W_2^2 |A'_2|) = O_{\mathbb{P}}(1)$$

by using  $W_2^2 = W_{2,0}^2 + (e^{-\theta_0 \Delta_2} - e^{-\theta \Delta_2}) y_1 W_2 + W_{2,0} (e^{-\theta_0 \Delta_2} - e^{-\theta \Delta_2}) y_1$ .

Finally,  $\sup_{(\theta, \sigma^2) \in J} |\alpha_2| \Delta_2 e^{-\theta \Delta_2} = O_{\mathbb{P}}(1)$ , which finally shows  $\sup_{(\theta, \sigma^2) \in J} |T'_2| = O_{\mathbb{P}}(1)$ .

• **Term  $T'_n$ :** Using (31), (34) and (35), we get

$$\sup_{(\theta, \sigma^2) \in J} |2\alpha_n \Delta_n| \leq K \quad \text{and} \quad \sup_{(\theta, \sigma^2) \in J} |2\alpha_n \Delta_n e^{-2\theta \Delta_n}| \leq K.$$

Moreover, one has  $|\frac{\alpha_n e^{-2\theta \Delta_n}}{A_{n-1}}| \leq 1$ . Finally, we have  $\sup_{(\theta, \sigma^2) \in J} W_n^2 = O_{\mathbb{P}}(1)$  and  $\sup_{(\theta, \sigma^2) \in J} |y_{n-1} W_n| = O_{\mathbb{P}}(1)$ . Hence, in order to show  $\sup_{(\theta, \sigma^2) \in J} |T'_n| = O_{\mathbb{P}}(1)$  it remains to show  $\sup_{(\theta, \sigma^2) \in J} |\alpha'_n W_n^2| = O_{\mathbb{P}}(1)$ .

$O_{\mathbb{P}}(1)$  and  $\sup_{(\theta, \sigma^2) \in J} |(A'_{n-1}/A_{n-1})\alpha_n W_n^2| = O_{\mathbb{P}}(1)$ . This is shown by using  $W_n^2 = W_{n,0}^2 + (e^{-\theta_0 \Delta_n} - e^{-\theta \Delta_n})y_{n-1}W_n + W_{n,0}(e^{-\theta_0 \Delta_n} - e^{-\theta \Delta_n})y_{n-1}$ .

• **Term**  $\sum_{i=2}^{n-1} \frac{A'_i}{A_i}$ : By (36),

$$\sum_{i=2}^{n-1} \frac{A'_i}{A_i} = \frac{n}{\theta} + \delta_n.$$

Finally,

$$\psi(\sigma^2, \theta) = \frac{\theta_0 \sigma_0^2}{\theta^2 \sigma^2} \sum_{i=3}^{n-1} q_i \bar{W}_{i,0}^2 - 2 \sum_{i=3}^{n-1} \frac{\sqrt{\Delta_i \Delta_{i+1}}}{\Delta_i + \Delta_{i+1}} \bar{W}_{i,0} \bar{W}_{i+1,0} - \frac{n}{\theta} + \delta_n$$

using Lemma 8.8. □

## 8.7 Proof of Theorems 6.1 and 6.2

In this section and the next one, we let  $\|\mathbf{A}\|$  denote the largest singular value of a matrix  $\mathbf{A}$  and  $\|\mathbf{v}\|$  denote the Euclidean norm of a vector  $\mathbf{v}$ . Finally, for  $k \in \{1, \dots, p\}$ , we let  $\mathbf{f}^{(k)} = (f_k(s_1), \dots, f_k(s_n))'$ . We first provide a decomposition of  $\bar{S}_n$  in the following lemma.

**Lemma 8.10.** *We have*

$$\bar{S}_n(\theta, \sigma^2) = S_n(\theta, \sigma^2) - r_1(\theta) + \frac{r_2(\theta) + 2r_3(\theta) - r_4(\theta)}{\sigma^2},$$

with

$$\begin{aligned} r_1(\theta) &= \sum_{i=1}^n [\log((\mathbf{R}_\theta^{-1})_{ii} - \bar{\epsilon}_{-i}) - \log((\mathbf{R}_\theta^{-1})_{ii})], \\ r_2(\theta) &= \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \epsilon_{-i}^2, \\ r_3(\theta) &= \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \epsilon_{-i} (y_i - \hat{Y}_{\theta, -i}(s_i)), \\ r_4(\theta) &= \sum_{i=1}^n \bar{\epsilon}_{-i} (y_i - \hat{Y}_{\theta, -i}(s_i) + \epsilon_{-i})^2, \\ \epsilon_{-i} &= [\mathbf{f}'_i - \mathbf{r}'_{\theta, -i} \mathbf{R}_{\theta, -i}^{-1} \mathbf{F}_{-i}] (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{-i}), \\ \bar{\epsilon}_{-i} &= \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{F} (\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{e}_{i,n}, \end{aligned}$$

where  $\mathbf{e}_{i,n}$  is the  $i$ -th base column vector of  $\mathbb{R}^n$ . We remark that  $r_k(\theta)$  does not depend on  $\sigma^2$  for  $k = 1, \dots, 4$ .



We now show that, in Lemma 8.10, the term  $S_n$  only, corresponding to the zero-mean case, is preponderant.

**Lemma 8.11.** *With the notation of Lemma 8.10, we have for  $k = 1, \dots, 4$*

$$\sup_{\theta \in [a, A]} |r_k(\theta)| = O(1) \text{ a.s.}$$

Because of Lemmas 8.10 and 8.11, we have a.s.

$$\sup_{(\theta, \sigma^2) \in J} |\bar{S}_n(\theta, \sigma^2) - S_n(\theta, \sigma^2)| = O(1).$$

Hence Theorem 6.1 follows from (46).

Also, from Lemmas 3.1, 8.10 and 8.11, we have

$$\begin{aligned} \bar{S}_n(\theta, \sigma^2) &= n \log(\sigma^2) + \log(1 - e^{-2\theta\Delta_2}) + \log(1 - e^{-2\theta\Delta_n}) \\ &\quad - \sum_{i=2}^{n-1} \log \left( \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right) - r_1(\theta) \\ &\quad + \frac{1}{\sigma^2} \left\{ \frac{(y_1 - e^{-\theta\Delta_2}y_2)^2}{(1 - e^{-2\theta\Delta_2})} + \frac{(y_n - e^{-\theta\Delta_n}y_{n-1})^2}{(1 - e^{-2\theta\Delta_n})} \right. \\ &\quad \left. + \sum_{i=2}^{n-1} \left[ \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right] \left[ y_i - \frac{\frac{e^{-\theta\Delta_i}y_{i-1} + \frac{e^{-\theta\Delta_{i+1}}y_{i+1}}{1 - e^{-2\theta\Delta_{i+1}}}}{\frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}}} \right]^2 + \delta_n(\theta) \right\}, \end{aligned}$$

where  $\delta_n(\theta)$  does not depend on  $\sigma^2$  and satisfies  $\sup_{a \leq \theta \leq A} |\delta_n(\theta)| = O_{\mathbb{P}}(1)$ . Since  $\check{\theta}\check{\sigma}^2 \rightarrow \theta_0\sigma_0^2$  a.s. from Theorem 6.1, and since  $aB > \theta_0\sigma_0^2$ ;  $Ab < \theta_0\sigma_0^2$ , we have  $[\partial/\partial\sigma^2]\bar{S}_n(\check{\theta}, \check{\sigma}^2) = 0$  for  $n$  large enough almost surely. Hence we obtain for  $n$  large enough almost surely,

$$\begin{aligned} 0 &= \frac{n}{\check{\sigma}^2} - \frac{1}{\check{\sigma}^4} \left\{ \frac{(y_1 - e^{-\check{\theta}\Delta_2}y_2)^2}{(1 - e^{-2\check{\theta}\Delta_2})} + \frac{(y_n - e^{-\check{\theta}\Delta_n}y_{n-1})^2}{(1 - e^{-2\check{\theta}\Delta_n})} \right. \\ &\quad \left. + \sum_{i=2}^{n-1} \left[ \frac{1}{1 - e^{-2\check{\theta}\Delta_i}} + \frac{e^{-2\check{\theta}\Delta_{i+1}}}{1 - e^{-2\check{\theta}\Delta_{i+1}}} \right] \left[ y_i - \frac{\frac{e^{-\check{\theta}\Delta_i}y_{i-1} + \frac{e^{-\check{\theta}\Delta_{i+1}}y_{i+1}}{1 - e^{-2\check{\theta}\Delta_{i+1}}}}{\frac{1}{1 - e^{-2\check{\theta}\Delta_i}} + \frac{e^{-2\check{\theta}\Delta_{i+1}}}{1 - e^{-2\check{\theta}\Delta_{i+1}}}} \right]^2 + \delta_n(\check{\theta}) \right\}. \end{aligned}$$

As noted at the end of the proof of Theorem 4.1, using identical techniques as for proving Theorem 4.1, one can finish the proof of Theorem 6.2.

## 8.8 Proofs of the lemmas in Section 8.7

*Proof of Lemma 8.10.* We have, with  $\mathbf{y}_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)'$ ,

$$\begin{aligned} Z(s_i) - \hat{Z}_{\theta, -i}(s_i) &= y_i + \mathbf{f}'_i \boldsymbol{\beta}_0 - \mathbf{f}'_i \hat{\boldsymbol{\beta}}_{-i} - \mathbf{r}'_{\theta, -i} \mathbf{R}_{\theta, -i}^{-1} (\mathbf{y}_{-i} + \mathbf{F}_{-i} \boldsymbol{\beta}_0 - \mathbf{F}_{-i} \hat{\boldsymbol{\beta}}_{-i}) \\ &= y_i - \mathbf{r}'_{\theta, -i} \mathbf{R}_{\theta, -i}^{-1} \mathbf{y}_{-i} + (\mathbf{f}'_i - \mathbf{r}'_{\theta, -i} \mathbf{R}_{\theta, -i}^{-1} \mathbf{F}_{-i}) (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{-i}) \\ &= y_i - \hat{Y}_{\theta, -i}(s_i) + \epsilon_{-i}. \end{aligned}$$

Also

$$(\mathbf{Q}_\theta^-)_{ii} = \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{e}_{i,n} - \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{F} (\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{e}_{i,n} = (\mathbf{R}_\theta^{-1})_{ii} - \bar{\epsilon}_{-i}.$$

Then, from (15),

$$\begin{aligned} \bar{S}_n(\theta, \sigma^2) &= \sum_{i=1}^n \left[ \log(\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i)) + \frac{(z_i - \hat{Z}_{\theta, -i}(s_i))^2}{\check{\sigma}_{\theta, \sigma^2, -i}^2(s_i)} \right] \\ &= n \log(\sigma^2) + \sum_{i=1}^n \left[ \log \left( \frac{1}{(\mathbf{Q}_\theta^-)_{ii}} \right) \right] + \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{Q}_\theta^-)_{ii} [y_i - \hat{Y}_{\theta, -i}(s_i) + \epsilon_{-i}]^2 \\ &= S_n(\theta, \sigma^2) - \sum_{i=1}^n [\log((\mathbf{R}_\theta^{-1})_{ii} - \bar{\epsilon}_{-i}) - \log((\mathbf{R}_\theta^{-1})_{ii})] + \frac{1}{\sigma^2} \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \epsilon_{-i}^2 \\ &\quad + \frac{2}{\sigma^2} \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \epsilon_{-i} (y_i - \hat{Y}_{\theta, -i}(s_i)) - \frac{1}{\sigma^2} \sum_{i=1}^n \bar{\epsilon}_{-i} (y_i - \hat{Y}_{\theta, -i}(s_i) + \epsilon_{-i})^2 \\ &= S_n(\theta, \sigma^2) - r_1(\theta) + \frac{r_2(\theta) + 2r_3(\theta) - r_4(\theta)}{\sigma^2}. \end{aligned}$$

□

Before proving Lemma 8.11, we state and prove some intermediary results.

**Lemma 8.12.** *We have*

$$\|\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{y}\| \leq K \sup_{t \in [0,1]} |Y(t)|.$$

*Proof of Lemma 8.12.* From (1) we have, with  $k = 1, \dots, p$  and with  $\mathbf{f} = (f_1, \dots, f_n)' = \mathbf{f}^{(k)}$ ,

$$\begin{aligned} [\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{y}]_k &= \mathbf{f}' \mathbf{R}_\theta^{-1} \mathbf{y} \\ &= \frac{1}{1 - e^{-2\theta\Delta_2}} f_1 y_1 + \frac{1}{1 - e^{-2\theta\Delta_n}} f_n y_n - \sum_{i=2}^n \frac{e^{-\theta\Delta_i}}{1 - e^{-2\theta\Delta_i}} f_i y_{i-1} - \sum_{i=2}^n \frac{e^{-\theta\Delta_i}}{1 - e^{-2\theta\Delta_i}} f_{i-1} y_i \\ &\quad + \sum_{i=2}^{n-1} \left( \frac{1}{1 - e^{-2\theta\Delta_i}} + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \right) f_i y_i \\ &= \frac{f_1 y_1 - f_2 y_1 e^{-\theta\Delta_2} - f_1 y_2 e^{-\theta\Delta_2} + f_2 y_2}{1 - e^{-2\theta\Delta_2}} \\ &\quad + \sum_{i=3}^n \frac{f_i y_i - f_i y_{i-1} e^{-\theta\Delta_i} - f_{i-1} y_i e^{-\theta\Delta_i} + e^{-2\theta\Delta_i} f_{i-1} y_{i-1}}{1 - e^{-2\theta\Delta_i}} \\ &= a_1 + \sum_{i=3}^n m_i, \end{aligned}$$

say. We have, using (31), and that  $f_k$  is continuously differentiable,

$$\begin{aligned} a_1 &= \left( \frac{1}{2\theta\Delta_2} + \delta_{1,n} \right) (f_1 y_1 (1 - e^{-\theta\Delta_2}) + (f_1 - f_2) y_1 e^{-\theta\Delta_2} + y_2 f_2 (1 - e^{-\theta\Delta_2}) + e^{-\theta\Delta_2} y_2 (f_2 - f_1)) \\ &= \frac{1}{2} (f_1 y_1 + y_2 f_2) + \frac{f_1 - f_2}{2\theta\Delta_2} (y_1 - y_2) + \delta_{1,n} \Delta_2 \sup_{t \in [0,1]} |Y(t)| \end{aligned}$$

and

$$\begin{aligned} m_i &= y_i \left( \frac{f_i - e^{-\theta\Delta_i} f_{i-1}}{1 - e^{-2\theta\Delta_i}} \right) + y_{i-1} \left( \frac{e^{-2\theta\Delta_i} f_{i-1} - e^{-\theta\Delta_i} f_i}{1 - e^{-2\theta\Delta_i}} \right) \\ &= y_i \left( \frac{f_i - f_{i-1} + f_{i-1} (1 - e^{-\theta\Delta_i})}{1 - e^{-2\theta\Delta_i}} \right) + y_{i-1} \left( \frac{f_{i-1} (e^{-2\theta\Delta_i} - e^{-\theta\Delta_i}) + (f_{i-1} - f_i) e^{-\theta\Delta_i}}{1 - e^{-2\theta\Delta_i}} \right) \\ &= (y_i - y_{i-1}) \frac{f_i - f_{i-1}}{2\theta\Delta_i} + \frac{1}{2} (y_i - y_{i-1}) f_{i-1} + \delta_{i,n} \Delta_i \sup_{t \in [0,1]} |Y(t)|. \end{aligned}$$

Hence, using that  $f_k$  is continuously differentiable,

$$\begin{aligned} \mathbf{f}' \mathbf{R}_\theta^{-1} \mathbf{y} &= \frac{1}{2} (f_1 y_1 + y_2 f_2) + \frac{f_1 - f_2}{2\theta\Delta_2} (y_1 - y_2) + \delta_{1,n} \Delta_2 \sup_{t \in [0,1]} |Y(t)| \\ &\quad + \sum_{i=3}^n \left( (y_i - y_{i-1}) \frac{f_i - f_{i-1}}{2\theta\Delta_i} + \frac{1}{2} (y_i - y_{i-1}) f_{i-1} + \delta_{i,n} \Delta_i \sup_{t \in [0,1]} |Y(t)| \right) \\ &= \delta_{1,n} \sup_{t \in [0,1]} |Y(t)| + \sum_{i=3}^{n-1} \left( \frac{f_i - f_{i-1}}{2\theta\Delta_i} (y_i - y_{i-1}) \right) + \frac{1}{2} \sum_{i=3}^{n-1} (y_i - y_{i-1}) f_{i-1}. \end{aligned} \quad (63)$$

The absolute value of first sum in (63) is equal to, after a summation by part, and using that  $f_k$  is twice continuously differentiable,

$$\begin{aligned} &\left| - \sum_{i=3}^{n-1} y_i \left( \frac{f_{i+1} - f_i}{2\theta\Delta_{i+1}} - \frac{f_i - f_{i-1}}{2\theta\Delta_i} \right) + \delta_{n,n} \sup_{t \in [0,1]} |Y(t)| \right| \\ &\leq \frac{1}{2\theta} \left| \sum_{i=3}^{n-1} y_i (f'(s_i) + \delta_{i,n} \Delta_{i+1} - f'(s_i) + \delta_{i,n} \Delta_i) \right| + \delta_{n,n} \sup_{t \in [0,1]} |Y(t)| \\ &= \delta_{n,n} \sup_{t \in [0,1]} |Y(t)|. \end{aligned}$$

The second sum in (63) is equal to, using that  $f_k$  is continuously differentiable,

$$\sum_{i=3}^{n-1} (y_i f_{i-1} - y_{i-1} f_{i-2} + y_{i-1} (f_{i-2} - f_{i-1})) = \delta_{n,n} \sup_{t \in [0,1]} |Y(t)|.$$

Hence

$$|\mathbf{f}' \mathbf{R}_\theta^{-1} \mathbf{y}| \leq K \sup_{t \in [0,1]} |Y(t)|.$$

□

**Lemma 8.13.** *We have, for  $k = 1, \dots, p$ ,*

$$|e'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}^{(k)}| \leq \begin{cases} K(\Delta_i + \Delta_{i+1}) & \text{when } i \notin \{1, n\} \\ K & \text{when } i \in \{1, n\} \end{cases}.$$

*Proof of Lemma 8.13.* Let  $\mathbf{f} = (f_1, \dots, f_n)' = \mathbf{f}^{(k)}$ .

i) When  $i \notin \{1, n\}$ , from (1),

$$\begin{aligned} e'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f} &= -\frac{e^{-\theta\Delta_i}}{1 - e^{-2\theta\Delta_i}} f_{i-1} + \frac{1}{1 - e^{-2\theta\Delta_i}} f_i + \frac{e^{-2\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} f_i - \frac{e^{-\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} f_{i+1} \\ &= \frac{f_i(1 - e^{-\theta\Delta_i}) + e^{-\theta\Delta_i}(f_i - f_{i-1})}{1 - e^{-2\theta\Delta_i}} + \frac{f_i(e^{-2\theta\Delta_{i+1}} - e^{-\theta\Delta_{i+1}}) + (f_i - f_{i+1})e^{-\theta\Delta_{i+1}}}{1 - e^{-2\theta\Delta_{i+1}}} \\ &= \frac{f_i}{2} + \frac{f'(s_i)}{2\theta} + \delta_{i,n}\Delta_i - \frac{f_i}{2} - \frac{f'(s_i)}{2\theta} + \delta_{i,n}\Delta_{i+1} \\ &= \delta_{i,n}(\Delta_i + \Delta_{i+1}), \end{aligned}$$

where we have used (31) and that  $f$  is twice continuously differentiable.

ii) Similarly,

$$e'_{1,n} \mathbf{R}_\theta^{-1} \mathbf{f} = \frac{f(0)}{2} - \frac{f'(0)}{2\theta} + \delta_{1,n}\Delta_2.$$

and

$$e'_{n,n} \mathbf{R}_\theta^{-1} \mathbf{f} = \frac{f(1)}{2} + \frac{f'(1)}{2\theta} + \delta_{n,n}\Delta_n.$$

□

**Lemma 8.14.** *We have*

$$\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{F} = \mathbf{I}_f + \mathbf{W}(\theta)$$

with  $\sup_{\theta \in [a, A]} \|\mathbf{W}(\theta)\| \rightarrow_{n \rightarrow \infty} 0$  and with  $\mathbf{I}_f$  the  $p \times p$  matrix defined by

$$[\mathbf{I}_f]_{kl} = f_k(0)f_l(0) + \frac{1}{2\theta} \int_0^1 f'_k(t)f'_l(t)dt + \frac{\theta}{2} \int_0^1 f_k(t)f_l(t)dt + \frac{1}{2} \int_0^1 (f'_k(t)f_l(t) + f_k(t)f'_l(t))dt.$$

Furthermore,  $\mathbf{I}_f$  is invertible.

*Proof of Lemma 8.14.* Let  $k, l \in \{1, \dots, p\}$  and let  $g = f_k$ ,  $h = f_l$ ,  $\mathbf{g} = (g_1, \dots, g_n)' = \mathbf{f}^{(k)}$

and  $\mathbf{h} = (h_1, \dots, h_n)' = \mathbf{f}^{(l)}$ . From (1), we have,

$$\begin{aligned}
[\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{F}]_{kl} &= \mathbf{g}' \mathbf{R}_\theta^{-1} \mathbf{h} \\
&= \frac{g_1 h_1}{1 - e^{-2\theta \Delta_2}} + \frac{g_n h_n}{1 - e^{-2\theta \Delta_n}} - \sum_{i=2}^n \frac{e^{-\theta \Delta_i}}{1 - e^{-2\theta \Delta_i}} g_i h_{i-1} - \sum_{i=2}^n \frac{e^{-\theta \Delta_i}}{1 - e^{-2\theta \Delta_i}} g_{i-1} h_i \\
&\quad + \sum_{i=2}^{n-1} \left( \frac{1}{1 - e^{-2\theta \Delta_i}} + \frac{e^{-2\theta \Delta_{i+1}}}{1 - e^{-2\theta \Delta_{i+1}}} \right) g_i h_i \\
&= \frac{g_1 h_1 - e^{-\theta \Delta_2} (g_2 h_1 + g_1 h_2) + g_2 h_2}{1 - e^{-2\theta \Delta_2}} + \sum_{i=3}^n \left( \frac{(g_i - e^{-\theta \Delta_i} g_{i-1}) (h_i - e^{-\theta \Delta_i} h_{i-1})}{1 - e^{-2\theta \Delta_i}} \right) \\
&= \frac{(g_1 - g_2)(h_1 - h_2) + \theta \Delta_2 (g_2 h_1 + g_1 h_2) + \delta_{1,n} \Delta_2^2}{2\theta \Delta_2 + \delta_{1,n} \Delta_2^2} \\
&\quad + \sum_{i=3}^n \frac{1}{1 - e^{-2\theta \Delta_i}} \left( (g_i - g_{i-1})(h_i - h_{i-1}) + g_{i-1} (1 - e^{-\theta \Delta_i}) h_{i-1} (1 - e^{-\theta \Delta_i}) \right. \\
&\quad \left. + g_{i-1} (1 - e^{-\theta \Delta_i}) (h_i - h_{i-1}) + (g_i - g_{i-1}) h_{i-1} (1 - e^{-\theta \Delta_i}) \right), \tag{64}
\end{aligned}$$

where we have used (31). Since  $f_k$  and  $f_l$  are continuously differentiable we have

$$\frac{(g_1 - g_2)(h_1 - h_2) + \theta \Delta_2 (g_2 h_1 + g_1 h_2)}{2\theta \Delta_2 + \delta_{1,n} \Delta_2^2} = g(0)h(0) + \delta_{1,n} \Delta_2.$$

Also, the element  $i$  of the sum in (64) is equal to

$$\begin{aligned}
&\frac{1}{2\theta} \frac{(g_i - g_{i-1})}{\Delta_i} (h_i - h_{i-1}) + g_{i-1} h_{i-1} \frac{\theta}{2} \Delta_i + \frac{1}{2} g_{i-1} (h_i - h_{i-1}) + \frac{1}{2} h_{i-1} (g_i - g_{i-1}) + \delta_{i,n} \Delta_i^2 \\
&= \frac{1}{2\theta} g'(s_i) h'(s_i) \Delta_i + \frac{\theta}{2} g(s_{i-1}) h(s_{i-1}) \Delta_i + \frac{1}{2} g(s_i) h'(s_i) \Delta_i + \frac{1}{2} h(s_i) g'(s_i) \Delta_i + \delta_{i,n} \Delta_i^2,
\end{aligned}$$

since  $f$  is twice continuously differentiable. Hence we have

$$\begin{aligned}
\mathbf{g}' \mathbf{R}_\theta^{-1} \mathbf{h} &= g(0)h(0) + \delta_{1,n} \Delta_2 \\
&\quad + \sum_{i=3}^n \left[ \Delta_i \left( \frac{1}{2\theta} g'(s_i) h'(s_i) + \frac{\theta}{2} g(s_{i-1}) h(s_{i-1}) + \frac{1}{2} g'(s_i) h(s_i) + \frac{1}{2} g(s_i) h'(s_i) \right) + \delta_{i,n} \Delta_i^2 \right] \\
&= [\mathbf{I}_f]_{kl} + w(\theta),
\end{aligned}$$

where  $\sup_{\theta \in [a, A]} |w(\theta)| \rightarrow_{n \rightarrow \infty} 0$ , by dominated convergence. Finally, we have, for  $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ ,

$$\sum_{k,l=1}^p \lambda_k \lambda_l [I_f]_{kl} = \left( \sum_{k=1}^p \lambda_k f_k(0) \right)^2 + \frac{1}{2} \int_0^1 \left( \frac{1}{\sqrt{\theta}} \sum_{k=1}^p \lambda_k f'_k(t) + \sqrt{\theta} \sum_{k=1}^p \lambda_k f_k(t) \right)^2 dt.$$

Hence, if  $\sum_{k,l=1}^p \lambda_k \lambda_l [I_f]_{kl} = 0$ , then  $\sum_{k=1}^p \lambda_k f_k(0) = 0$  and for all  $t \in [0, 1]$ ,  $\sum_{k=1}^p \lambda_k f'_k(t) = -\theta \sum_{k=1}^p \lambda_k f_k(t)$  so that, by linear independence,  $\lambda_1 = \dots = \lambda_p = 0$ . Hence  $\mathbf{I}_f$  is invertible.  $\square$

**Lemma 8.15.** *We have*

$$\max_{i=1,\dots,n} \sup_{a \leq \theta \leq A} |y_i - \hat{Y}_{\theta,-i}(s_i)| \leq K \sup_{t \in [0,1]} |Y(t)|.$$

*Proof of Lemma 8.15.* For  $i \notin \{1, \dots, n\}$ , using (41),

$$\begin{aligned} y_i - \hat{Y}_{\theta,-i}(s_i) &= \frac{1}{A_i} (\alpha_i W_i - \alpha_{i+1} e^{-\theta \Delta_{i+1}} W_{i+1}) \\ &= \delta_{i,n} W_i + \delta_{i,n} W_{i+1} \\ &= \delta_{i,n} \sup_{t \in [0,1]} |Y(t)|. \end{aligned}$$

Also

$$y_1 - \hat{Y}_{\theta,-1}(s_1) = y_1 - e^{-\theta \Delta_2} y_2 = \delta_{1,n} \sup_{t \in [0,1]} |Y(t)|$$

and similarly  $y_n - \hat{Y}_{\theta,-n}(s_n) = \delta_{n,n} \sup_{t \in [0,1]} |Y(t)|$ . □

We now prove Lemma 8.11.

*Proof of Lemma 8.11.* • **Term**  $r_1(\theta)$ :

$$\begin{aligned} r_1(\theta) &= \sum_{i=1}^n [\log((\mathbf{R}_\theta^{-1})_{ii} - \bar{\epsilon}_{-i}) - \log((\mathbf{R}_\theta^{-1})_{ii})] = \sum_{i=1}^n \log \left( 1 - \frac{\bar{\epsilon}_{-i}}{(\mathbf{R}_\theta^{-1})_{ii}} \right) \\ &= \sum_{i=1}^n \log \left( 1 - \frac{(\mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{F})(\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{F})^{-1} (\mathbf{F}' \mathbf{R}_\theta^{-1} \mathbf{e}_{i,n})}{(\mathbf{R}_\theta^{-1})_{ii}} \right) \\ &= \sum_{i=1}^n \log \left( 1 - \frac{\delta_{i,n}}{(\mathbf{R}_\theta^{-1})_{ii}} \right), \end{aligned}$$

from Lemmas 8.13 and 8.14.

Now, from (1), for  $i = 2, \dots, n$ , we have  $\sup_\theta \frac{1}{(\mathbf{R}_\theta^{-1})_{ii}} = \delta_{i,n} \Delta_i$  and  $\sup_\theta \frac{1}{(\mathbf{R}_\theta^{-1})_{11}} = \delta_{1,n} \Delta_2$ . Hence  $r_1(\theta) = \log(1 - \delta_{1,n} \Delta_2) + \sum_{i=2}^n \log(1 - \delta_{i,n} \Delta_i)$  so that  $\sup_{\theta \in [a,A]} |r_1(\theta)|$  is bounded as  $n \rightarrow \infty$ .

• **Term**  $r_2(\theta)$ :

For  $k \in \mathbb{N}$ , let  $\mathbf{f}_{-i}^{(k)}$  be obtained by removing component  $i$  of  $\mathbf{f}^{(k)}$ . We observe that  $f_k(s_i) - \mathbf{r}'_{\theta,-i} \mathbf{R}_{\theta,-i}^{-1} \mathbf{f}_{-i}^{(k)}$  can be interpreted as a leave-one-out prediction error for a  $n$ -dimensional observation vector equal to  $\mathbf{f}^{(k)}$ . Hence from (3),

$$\begin{aligned} r_2(\theta) &= \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \bar{\epsilon}_{-i}^2 \\ &= \sum_{k,l=1}^p \left[ \sum_{i=1}^n \frac{1}{[(\mathbf{R}_\theta^{-1})_{ii}]} (\beta_0 - \hat{\beta}_{-i})_k (\beta_0 - \hat{\beta}_{-i})_l \left( \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}^{(k)} \right) \left( \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}^{(l)} \right) \right]. \end{aligned}$$

For  $i \in \{1, \dots, n\}$ , one can show that Lemmas 8.12 and 8.14 remain true with  $\mathbf{F}$ ,  $\mathbf{R}_\theta$ ,  $\mathbf{y}$  replaced by  $\mathbf{F}_{-i}$ ,  $\mathbf{R}_{\theta,-i}$ ,  $\mathbf{y}_{-i}$ . In addition, these (modified) lemmas can be shown to be uniform over  $i = 1, \dots, n$ . As a consequence, we have, for  $i = 1, \dots, n$ ,

$$\|\hat{\boldsymbol{\beta}}_{-i} - \boldsymbol{\beta}_0\| = \|(\mathbf{F}'_{-i}\mathbf{R}_{\theta,-i}^{-1}\mathbf{F}_{-i})^{-1}\mathbf{F}'_{-i}\mathbf{R}_{\theta,-i}^{-1}\mathbf{y}_{-i}\| \leq K \sup_{t \in [0,1]} |Y(t)|. \quad (65)$$

Also, since from (1)  $\max_{i=1, \dots, n} \frac{1}{(\mathbf{R}_\theta^{-1})_{ii}} \leq K$ , we have

$$\sum_{i=1}^n \frac{1}{(\mathbf{R}_\theta^{-1})_{ii}} \left| \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}^{(k)} \right| \left| \mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}^{(l)} \right| \leq K + \sum_{i=2}^{n-1} \delta_{i,n} (\Delta_i + \Delta_{i+1})^2,$$

from Lemma 8.13. Hence, we have

$$r_2(\theta) \leq K \sup_{t \in [0,1]} Y(t)^2 \left( K + \sum_{i=2}^{n-1} (\Delta_i + \Delta_{i+1})^2 \right).$$

Hence finally,  $\sup_{\theta \in [a, A]} |r_2(\theta)|$  is a.s. bounded as  $n \rightarrow \infty$ .

• **Term**  $r_3(\theta)$ :

$$\begin{aligned} r_3(\theta) &= \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} \epsilon_{-i} (y_i - \hat{Y}_{\theta,-i}(s_i)) \\ &= \sum_{k=1}^p \sum_{i=1}^n (\mathbf{R}_\theta^{-1})_{ii} (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{-i})_k [f_k(s_i) - \mathbf{r}'_{\theta,-i} \mathbf{R}_{\theta,-i}^{-1} \mathbf{f}_{-i}^{(k)}] (y_i - \hat{Y}_{\theta,-i}(s_i)). \end{aligned}$$

We make the same observation on  $f_k(s_i) - \mathbf{r}'_{\theta,-i} \mathbf{R}_{\theta,-i}^{-1} \mathbf{f}_{-i}^{(k)}$  as for  $r_2(\theta)$ . Hence we have

$$r_3(\theta) = \sum_{k=1}^p \sum_{i=1}^n (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{-i})_k (\mathbf{e}'_{i,n} \mathbf{R}_\theta^{-1} \mathbf{f}_{-i}^{(k)}) (y_i - \hat{Y}_{\theta,-i}(s_i)).$$

Hence, from (65) and Lemmas 8.13 and 8.15, we have

$$|r_3(\theta)| \leq \sum_{k=1}^p \left( \sum_{i=2}^{n-1} \left( \delta_{i,n} (\Delta_i + \Delta_{i+1}) \sup_{t \in [0,1]} Y(t)^2 \right) + \delta_{1,n} \sup_{t \in [0,1]} Y(t)^2 \right).$$

Hence,  $r_3(\theta)$  is almost surely bounded as  $n \rightarrow \infty$ .

• **Term**  $r_4(\theta)$ :

From Lemmas 8.13 and 8.14, we have for  $i \in \{2, \dots, n-1\}$ ,  $\bar{\epsilon}_{-i} \leq K(\Delta_i + \Delta_{i+1})^2$  and for  $i \in \{1, n\}$ ,  $\bar{\epsilon}_{-i} \leq K$ . Hence

$$\begin{aligned} |r_4(\theta)| &\leq K(y_1 - \hat{Y}_{\theta,-1}(s_1) + \epsilon_{-1})^2 + K(y_n - \hat{Y}_{\theta,-n}(s_n) + \epsilon_{-n})^2 \\ &\quad + \sum_{i=2}^{n-1} (\Delta_i + \Delta_{i+1})^2 (y_i - \hat{Y}_{\theta,-i}(s_i) + \epsilon_{-i})^2 \\ &\leq K \sup_{t \in [0,1]} Y(t)^2 + K \sum_{i=2}^{n-1} (\Delta_i + \Delta_{i+1})^2 \sup_{t \in [0,1]} Y(t)^2, \end{aligned}$$

using also Lemma 8.15, and remarking that the previous treatment of  $r_2(\theta)$  implies that  $\epsilon_{-i}^2 \leq K \sup_{t \in [0,1]} Y(t)^2$ . Hence  $r_4(\theta)$  is a.s. bounded as  $n \rightarrow \infty$ . □

## Acknowledgements

We are grateful to Jean-Marc Azais, for constructive discussions on the topic of this manuscript. This work was partly funded by the ANR project PEPITO. We thank the associate editor and two referees, whose suggestions lead to a broader scope and an improved presentation of the manuscript.

## References

- [1] P. Abrahamsen. A review of Gaussian random fields and correlation functions. Technical report, Norwegian computing center, 1997.
- [2] M. Abt and W. J. Welch. Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes. *The Canadian Journal of Statistics*, 26:127–137, 1998.
- [3] E. Anderes. On the consistent separation of scale and variance for Gaussian random fields. *The Annals of Statistics*, 38:870–893, 2010.
- [4] A. B. Antognini and M. Zagoraïou. Exact optimal designs for computer experiments via Kriging metamodeling. *Journal of Statistical Planning and Inference*, 140:2607–2617, 2010.
- [5] F. Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics and Data Analysis*, 66:55–69, 2013.
- [6] F. Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35, 2014.
- [7] F. Bachoc. Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *To appear in Bernoulli*, 2017.
- [8] F. Bachoc, K. Ammar, and J.M. Martinez. Improvement of code behavior in a design of experiments by metamodeling. *Nuclear science and engineering*, 183(3):387–406, 1016.
- [9] F. Bachoc, G. Bois, J. Garnier, and J.M. Martinez. Calibration and improved prediction of computer models by universal Kriging. *Nuclear Science and Engineering*, 176(1):81–97, 2014.



- [10] C-H. Chang, H-C. Huang, and C-K. Ing. Mixed domain asymptotics for a stochastic process model with time trend and measurement error. *Bernoulli*, 23(1):159–190, 2017.
- [11] H-S. Chen, D. G. Simpson, and Z. Ying. Infill asymptotics for a stochastic process model with measurement error. *Statistica Sinica*, 10:141–156, 2000.
- [12] N. Cressie and S.N. Lahiri. The asymptotic distribution of REML estimators. *Journal of Multivariate Analysis*, 45:217–233, 1993.
- [13] N. Cressie and S.N. Lahiri. Asymptotics for REML estimation of spatial covariance parameters. *Journal of Statistical Planning and Inference*, 50:327–341, 1996.
- [14] J. Du, H. Zhang, and V. S. Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics*, 37:3330–3361, 2009.
- [15] O. Dubrule. Cross validation of Kriging in a unique neighborhood. *Mathematical Geology*, 15:687–699, 1983.
- [16] R. Furrer, F. Bachoc, and J. Du. Asymptotic properties of multivariate tapering for estimation and prediction. *Journal of Multivariate Analysis*, 149:177–191, 2016.
- [17] T-C. Hu and R.L. Taylor. On the strong law for arrays and for the bootstrap mean and variance. *International Journal of Mathematics and Mathematical Sciences*, 20(2):375–382, 1997.
- [18] I.A. Ibragimov and Y.A. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, 1978.
- [19] J. Istas and G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l’Institut Henri Poincaré*, 33:407–436, 1997.
- [20] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black box functions. *Journal of Global Optimization*, 13:455–492, 1998.
- [21] C. G. Kaufman and B. A. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484, 2013.
- [22] W-L. Loh. Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *The Annals of Statistics*, 33:2344–2394, 2005.
- [23] W-L. Loh and T-K. Lam. Estimating structured correlation matrices in smooth Gaussian random field models. *The Annals of Statistics*, 28:880–904, 2000.
- [24] K. V. Mardia and R. J. Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71:135–146, 1984.

- [25] G. Matheron. *La Théorie des Variables Régionalisées et ses Applications*. Fascicule 5 in Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau. Ecole Nationale Supérieure des Mines de Paris, 1970.
- [26] M. H. Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134, 2013.
- [27] R. Paulo, G. Garcia-Donato, and J. Palomo. Calibration of computer models with multivariate output. *Computational Statistics and Data Analysis*, 56:3959–3974, 2012.
- [28] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, 2006.
- [29] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409–423, 1989.
- [30] T.J. Santner, B.J. Williams, and W.I. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.
- [31] B. A. Shaby and D. Ruppert. Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, 21(2):433–452, 2012.
- [32] M.L. Stein. Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics*, 16:55–63, 1988.
- [33] M.L. Stein. Bounds on the efficiency of linear predictions using an incorrect covariance function. *The Annals of Statistics*, 18:1116–1138, 1990.
- [34] M.L. Stein. Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *The Annals of Statistics*, 18:850–872, 1990.
- [35] M.L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.
- [36] S. Sundararajan and S.S. Keerthi. Predictive approaches for choosing hyperparameters in Gaussian processes. *Neural Computation*, 13:1103–1118, 2001.
- [37] A. W. van der Vaart. Maximum likelihood estimation under a spatial sampling scheme. *The Annals of Statistics*, 24(5):2049–2057, 1996.
- [38] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [39] D. Wang and W-L. Loh. On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electronic Journal of Statistics*, 5:238–269, 2011.

- [40] Z. Ying. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, 36:280–296, 1991.
- [41] Z. Ying. Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics*, 21:1567–1590, 1993.
- [42] H. Zhang. Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261, 2004.
- [43] H. Zhang and Y. Wang. Kriging and cross validation for massive spatial data. *Environmetrics*, 21:290–304, 2010.
- [44] H. Zhang and D. L. Zimmerman. Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92:921–936, 2005.