The Annals of Statistics
1993, Vol. 21, No. 3, 1567–1590

MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS UNDER A SPATIAL SAMPLING SCHEME[1]

BY ZHILIANG YING

University of Illinois, Urbana–Champaign

We study in detail asymptotic properties of maximum likelihood estimators of parameters when observations are taken from a two-dimensional Gaussian random field with a multiplicative Ornstein–Uhlenbeck covariance function. Under the complete lattice sampling plan, it is shown that the maximum likelihood estimators are strongly consistent and asymptotically normal. The asymptotic normality here is normalized by the fourth root of the sample size and is obtained through higher order expansions of the likelihood score equations. Extensions of these results to higher-dimensional processes are also obtained, showing that the convergence rate becomes better as the dimension gets higher.

1. Introduction. In their modeling of computer experiments, Sacks, Schiller and Welch (1989) and Sacks, Welch, Mitchell and Wynn (1989) have proposed a use of spatial Gaussian processes. Suppose that U is the set of all possible computer inputs, usually called design points, upon which experiments may be conducted. Let $Y(u)$ be the computer response after an experiment is run at $u \in U$. Typically, each component in u represents a factor that is related to the output function Y. In these experiments, no measurement error exists, that is, repeating an experiment at the same design value gives the same response. Suppose N such experiments have been conducted at N different points $u_1, \dots, u_N$ with their corresponding responses $Y(u_1), \dots, Y(u_N)$ being observed. It is often desirable to characterize behavior of Y and to predict responses $Y(s)$ at unexperimented input points $s \in U$, based on observations $Y(u_1), \dots, Y(u_N)$. To do so, Sacks, Schiller and Welch (1989) and Sacks, Welch, Mitchell and Wynn (1989) modeled Y as a realization of a multidimensional Gaussian spatial process (random field) whose covariance function belongs to a parametric family. Specifically, with certain rescaling, they considered without loss of generality the situation with $U = [0, 1]^d \subset R^d$ and found a class of homogeneous Gaussian random fields with multiplicative covariance functions of the form (1.1) below to be widely applicable to the analysis of computer experiments as well as to be computationally amenable.

Received January 1990; revised June 1992.
[1]Research supported by NSF Grants DMS-86-09819 and DMS-90-01726 and AT & T Affiliated Program.
AMS 1991 *subject classifications*. Primary 62F12; secondary 60G60, 60G15.
Key words and phrases. Gaussian random fields, multiplicative covariance functions, lattice sampling, maximum likelihood estimation, computer experiments, consistency, asymptotic normality.

The parametric family of covariance functions is defined by

$$(1.1)\quad V_q(\sigma^2,\theta,t,s) = \sigma^2 \prod_{i=1}^{d} \exp\{-\theta_i|t_i - s_i|^q\} = \sigma^2 \exp\left\{-\sum_{i=1}^{d} \theta_i|t_i - s_i|^q\right\},$$

where $t = (t_1,\dots,t_d)^T$ and $s = (s_1,\dots,s_d)^T \in U$, $\sigma^2 > 0$, $\theta = (\theta_1,\dots,\theta_d) \in (0,\infty)^d$ and $q \in (0,2]$. Here q usually indicates smoothness of the process Y. For example, $q = 1$ implies that the process is nowhere differentiable, whereas $q = 2$ implies that the process is infinitely many times differentiable. Unfortunately, likelihood functions with $q > 1$ are rather difficult to handle analytically, due mainly to the fact that as the sample size increases, the observations become very highly correlated. However, when $q = 1$, which will be assumed throughout this paper, certain Markovian properties can be exploited to conduct asymptotic analysis. Once q is fixed, the probabilistic structure of Y relies on the parameters θ and σ^2. Since θ and σ^2 are unknown, Sacks, Schiller and Welch (1989) proposed a natural approach of estimating them by their maximum likelihood estimators using observations $Y(u_1),\dots,Y(u_N)$. Modeling deterministic functions via spatial processes is also examined by Currin, Mitchell, Morris and Ylvisaker (1991). A wide range of statistical applications of spatial processes can be found in Cressie (1991) and Ripley (1981).

Maximum likelihood estimation under spatial sampling of Gaussian processes was studied by Mardia and Marshall (1984). Their results on asymptotic behavior of maximum likelihood estimators rely on the assumptions that observations of neighboring points are not highly correlated and that the Fisher information matrix satisfies certain regularity conditions. Since the set of all design points here is compact, Mardia and Marshall's (1984) results are not applicable to our model. Earlier, Sweeting (1980) established a general result on asymptotic behavior of maximum likelihood estimates for possibly non-Gaussian processes under certain regularity conditions. Again these conditions make his result unsuitable for our model. In fact, we are not aware of any results on asymptotic properties of maximum likelihood estimates when observations are taken from a Gaussian random field on a compact region.

When $d = 1$, the parameters θ and σ^2 in (1.1), recalling $q = 1$, cannot be identified simultaneously because if $\theta\sigma^2 = \tilde{\theta}\tilde{\sigma}^2$, then the induced measures with (θ,σ^2) and $(\tilde{\theta},\tilde{\sigma}^2)$ are mutually absolutely continuous [cf. Ibragimov and Rozanov (1978)]. However, Ying (1991) showed that the product $\hat{\theta}\hat{\sigma}^2$ of the maximum likelihood estimators $\hat{\theta}$ and $\hat{\sigma}^2$, as an estimator of $\theta\sigma^2$, is strongly consistent and asymptotically normal as the number of sample points tends to ∞. For $d \geq 2$, (1.1) gives a rather different structure because all the parameters θ_i, $i = 1,\dots,d$ and σ^2 are identifiable. Indeed, the main theme of this investigation is to show the consistency and asymptotic normality of the maximum likelihood estimators of these parameters.

The paper is organized as follows. In the next section some notation and assumptions are introduced. In Section 3 we investigate the large sample behavior of the maximum likelihood estimators and establish their consistency and asymptotic normality, whose proofs are provided in Section 4. These results are extended, in Section 5, to higher-dimensional processes under somewhat more stringent conditions on the sampling scheme.

2. Notation and assumptions. For notational convenience as well as technical reasons we first deal with the two-dimensional situation. Let $X(s,t)$, $s,t \in [0,1]$ denote a zero-mean Gaussian process with a multiplicative covariance function

$$\begin{aligned}\Gamma_{\lambda,\mu,\sigma^2}(s_1,t_1;s_2,t_2) &\triangleq E[X(s_1,t_1)X(s_2,t_2)]\\ &= \sigma^2\exp\{-\lambda|s_1-s_2|-\mu|t_1-t_2|\},\end{aligned} \tag{2.1}$$

where λ, μ and σ^2 are parameters. Some technical difficulties force us to consider the situations in which observations are taken from a domain which forms a complete lattice with m and n partitions in the first and the second coordinates, respectively. In other words, the set of all observations may be written as $\{X(u_i,v_k): i=1,\dots,m;\ k=1,\dots,n\}$. Without loss of generality, both $\{u_i\}$ and $\{v_k\}$ will be arranged in ascending order. The design need not be nested in the sense that u_i, v_k may depend on m and n and we do not assume that $\{u_i,\ i=1,\dots,m\}=\{u_i(m),\ i=1,\dots,m\}$ is subset of $\{u_i(m+1),\ i=1,\dots,m+1\}$ or $\{v_k(n),\ k=1,\dots,n\}$ is a subset of $\{v_k(n+1),\ i=1,\dots,n+1\}$.

The requirement that observations be taken from a complete lattice-type subset is crucial to our analysis. In particular, it results in a convenient dimension reduction as well as an inheritance of a useful Markovian-type property from the corresponding one-dimensional process. Such a dimension reduction is easily seen from the likelihood function below and Lemma 2 in Section 4. Let

$$A(\lambda)=(e^{-\lambda|u_i-u_j|})_{1\le i,j\le m},\qquad B(\mu)=(e^{-\mu|v_k-v_l|})_{1\le k,l\le n}, \tag{2.2}$$

$$x=(x_1^T,\dots,x_n^T)^T\quad\text{with } x_i=(x_{i1},\dots,x_{in})^T,\quad x_{ik}=X(u_i,v_k). \tag{2.3}$$

It is straightforward that the covariance matrix

$$E(xx^T)=\sigma^2A(\lambda)\otimes B(\mu). \tag{2.4}$$

Here and in the sequel, $\otimes$ is used to denote the Kronecker product. Therefore, the likelihood function of x can be written as

$$\begin{aligned}L_{m,n}(\lambda,\mu,\sigma^2) &= (2\pi\sigma^2)^{-mn/2}[\det(A(\lambda)\otimes B(\mu))]^{-1/2}\\ &\quad\times\exp\left\{-\frac{1}{2\sigma^2}x^T(A(\lambda)\otimes B(\mu))^{-1}x\right\}\end{aligned} \tag{2.5}$$

and the loglikelihood function becomes

$$\begin{aligned}l_{m,n}(\lambda,\mu,\sigma^2) &\triangleq -2\log L_{m,n}(\lambda,\mu,\sigma^2)\\ &= mn\log(2\pi\sigma^2)+\log[\det(A(\lambda)\otimes B(\mu))]\\ &\quad+\frac{1}{\sigma^2}x^T(A(\lambda)\otimes B(\mu))^{-1}x.\end{aligned} \tag{2.6}$$

Consistency and asymptotic normality of maximum likelihood estimators of λ, μ, and σ^2 will be proved through various approximations to the loglikelihood function $l_{m,n}$ and its derivatives. To this end, the following assumptions will be made:

(2.7) Both $m \to \infty$ and $n \to \infty$ in such a way that m/n are bounded away from 0 and ∞.

(2.8) Both $\{u_i, i = 1, \dots, m\}$ and $\{v_k, k = 1, \dots, n\}$ become dense in $[0, 1]$ as $m \to \infty$ and $n \to \infty$.

3. Consistency and asymptotic normality. In this section, main results on asymptotic behavior of the maximum likelihood estimators $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}^2$ will be presented in two theorems. The first theorem gives a strong consistency result while the second one shows that estimators are asymptotically normal. Proofs of these two theorems are given in Section 4.

THEOREM 1. *Let λ_0, μ_0 and σ_0^2 denote the true parameters and C be a compact region in R_+^2 (positive orthant) that contains (λ_0, μ_0) as an interior point. Then $(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2)$, the maximum likelihood estimator that maximizes $L_{m,n}$ over $C \times R_+$, is strongly consistent*

$$(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2) \to (\lambda_0, \mu_0, \sigma_0^2) \quad a.s. \tag{3.1}$$

In particular, there always exists a strongly consistent local maximizer of $L_{m,n}$.

REMARK 1. The above consistency result shows that parameter estimation for this particular two-dimensional process is very different from its one-dimensional counterpart. All unknown parameters in the former model are clearly identifiable whereas parameters in the latter are not, as mentioned in Section 1. In fact, even without conditions (2.7) and (2.8) on the sampling procedure, we can easily provide a naive consistent estimator of $(\lambda_0, \mu_0, \sigma_0^2)$ as follows. First it is not hard to see, by taking a rather crude first order expansion of $l_{m,n}$ as in Ying (1991), that $\hat{\lambda}\hat{\mu}\hat{\sigma}^2 \to \lambda_0\mu_0\sigma_0^2$ a.s. Moreover, with u_1 fixed, $X(u_1, v)$ is a one-dimensional Gaussian process having $\sigma^2 \exp\{-u|t - s|\}$ as the covariance function. Thus Theorem 1 of Ying (1991) shows that the maximum likelihood estimator of $\mu_0\sigma_0^2$ using $X(u_1, v_k)$, $k = 1, \dots, n$ is consistent. Likewise, a consistent estimator of $\lambda_0\sigma_0^2$ can also be constructed from $X(u_i, v_1)$, $i = 1, \dots, m$. By combining these three estimators we can easily get consistent estimators of λ_0, μ_0 and σ_0^2.

REMARK 2. The identifiability of parameters in the Ornstein–Uhlenbeck covariance functions is somewhat intriguing. The preceding paragraph shows that it depends on the dimension. We now show that the particular product form is also important. Consider the corresponding isotropic (d-dimensional) covariance function

$$V_d(\sigma^2, \alpha; t, s) = \sigma^2 \exp\left\{-\alpha\left[\sum_{i=1}^{n} (t_i - s_i)^2\right]^{1/2}\right\}.$$

Its spectral density function can be written as

$$g_d(\sigma^2, \alpha; x) = \frac{C\sigma^2\alpha}{\left(\alpha^2 + \sum_{i=1}^d x_i^2\right)^{(d+1)/2}},$$

where C is a constant depending only on d; compare Yaglom [(1987), page 362]. Thus for $d \le 3$, $V_d(\sigma^2, \alpha; \cdot\,)$ and $V_d(\tilde{\sigma}^2, \tilde{\alpha}; \cdot\,)$ induce two Gaussian measures that are mutually absolutely continuous if and only if $\sigma^2\alpha = \tilde{\sigma}^2\tilde{\alpha}$; compare Skorokhod and Yadrenko (1973). In other words, σ^2 and α are not identifiable when $d \le 3$. The proof of Theorem 1 in Section 3 indicates that identification of λ and μ relies heavily on higher order expansion of score equations to separate effects of the two parameters. The product form effectively produces such separation while the Euclidean distance does not.

REMARK 3. The Gaussian assumption is also crucial to the identifiability of σ^2, λ and μ. Consider two independent stationary Gaussian processes Z_1 and Z_2 on [0, 1] with covariance functions $\sigma^2 e^{-\lambda|\cdot|}$ and $e^{-\mu|\cdot|}$, respectively. Define a two-dimensional random field $Z(t, s) = Z_1(t)Z_2(s)$. Then Z has the same covariance function as X but is non-Gaussian. Moreover, σ^2, λ and μ cannot be identified simultaneously, since even if we could observe entire sample paths of Z_1 and Z_2 (therefore Z too), only $\lambda\sigma^2$ and μ can be identified. An intuitive explanation for this is that higher order expansions need moment conditions higher than two, making the Gaussian assumption necessary.

Under some additional regularity conditions on the sampling scheme, the estimators $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}^2$ are asymptotically normal. However, their convergence rates are rather different from that of the usual maximum likelihood estimation. This is the content of the following theorem.

THEOREM 2. *With the same notation and assumptions as in Theorem* 1, *we have*

$$\sqrt{mn}\left(\hat{\lambda}\hat{\mu}\hat{\sigma}^2 - \lambda_0\mu_0\sigma_0^2\right) \to_{\mathscr{D}} N\left(0, 2\left(\lambda_0\mu_0\sigma_0^2\right)^2\right). \tag{3.2}$$

Furthermore, suppose that $\{u_i\}$ *and* $\{v_k\}$ *are so chosen that*

$$\max_{2\le i\le m} \xi_i = o(m^{-1/2}) \quad \text{and} \quad \max_{2\le k\le n} \zeta_k = o(n^{-1/2}), \tag{3.3}$$

where $\xi_i = u_i - u_{i-1}$ *and* $\zeta_k = v_k - v_{k-1}$. *Then*

$$\begin{pmatrix} \sqrt{n}\,(\hat{\lambda} - \lambda_0) \\ \sqrt{m}\,(\hat{\mu} - \mu_0) \end{pmatrix} \to_{\mathscr{D}} N\left(0, \begin{pmatrix} 2\lambda_0^2/(1+\lambda_0) & 0 \\ 0 & 2\mu_0^2/(1+\mu_0) \end{pmatrix}\right). \tag{3.4}$$

If, in addition, $n/m \to \rho$, then

$$\sqrt{n}\left[\begin{pmatrix}\hat{\lambda}\\ \hat{\mu}\\ \hat{\sigma}^2\end{pmatrix} - \begin{pmatrix}\lambda_0\\ \mu_0\\ \sigma_0^2\end{pmatrix}\right] \to_{\mathscr{D}} N(0, \Sigma), \tag{3.5}$$

where

$$\Sigma = \begin{pmatrix} 2\lambda_0^2/(1+\lambda_0) & 0 & -2\sigma_0^2\lambda_0/(1+\lambda_0) \\ 0 & 2\rho\mu_0^2/(1+\mu_0) & -2\rho\sigma_0^2\mu_0/(1+\mu_0) \\ -2\sigma_0^2\lambda_0/(1+\lambda_0) & -2\rho\sigma_0^2\mu_0/(1+\mu_0) & 2\sigma_0^4\left[(1+\lambda_0)^{-1} + \rho(1+\mu_0)^{-1}\right] \end{pmatrix}.$$

Note that Σ is singular ($\text{rank}(\Sigma) = 2$).

REMARK 4. The weak convergence results (3.4) and (3.5) are rather interesting since they are normalized not by the usual square root of the sample size. For fixed $\mu = \mu_0$ and $\sigma^2 = \sigma_0^2$, the Fisher information number for the parameter λ can be shown to be proportional to the sample size $N = mn$. The same is true for the other two parameters. Therefore, it is the collinearity among the three unknown parameters that brings down the rate of convergence.

REMARK 5. The weak convergence result (3.4) also shows that $\hat{\lambda}$ and $\hat{\mu}$ are asymptotically independent and that the value of one parameter does not affect the accuracy of the estimator of the other. Moreover, none of the asymptotic variances in (3.2), (3.4) and (3.5) involves the underlying design. Therefore, as far as the estimation is concerned, any choice of $\{u_i,\ i = 1, \dots, m\}$ and $\{v_k,\ k = 1, \dots, n\}$ satisfying (2.7), (2.8) and (3.3) results in the same asymptotic accuracy for the maximum likelihood estimators.

REMARK 6. Both Theorems 1 and 2 require that observations be made on a complete lattice subset. This is certainly a restrictive assumption, but it seems to be necessary for our proofs to go through. From these two theorems it appears very plausible that similar results should hold for more general, nonlattice-type sampling schemes of Y. Further developments in this direction will certainly be of great interest.

REMARK 7. In view of its long history, it is natural to ask whether available results on maximum likelihood estimates can be used to derive our results. However, we have not been able to find any of those directly applicable to our setting. Most results deal with independent observations. For those covering dependent data, some kind of rapidly decreasing dependency (as in time series) is usually assumed. Other more general results, such as Theorem 1.1 of Ibragimov and Has'minskii [(1981), page 174], require conditions that are difficult to verify and we feel a direct analysis as given here is more appropriate.

REMARK 8. It is also natural to ask whether or not the maximum likelihood estimators here are asymptotically efficient. In a recent unpublished manuscript, A. van der Vaart showed, via certain reparametrization and asymptotic expansions similar to the proof of Theorem 2 given in Section 4, that it is indeed the case in the sense that the usual convolution and asymptotic minimax properties hold.

4. Proofs of Theorems 1 and 2. The proofs of Theorems 1 and 2 are based on approximations of the loglikelihood function $l_{m,n}$ and its derivatives. To do so, we first introduce some lemmas.

LEMMA 1. *Let* $\eta > 0$ *and* $-\infty < s_1 < \cdots < s_r < \infty$. *Define* $r \times r$ *matrix* $G \triangleq (\exp\{-\eta|s_i - s_j|\})_{1 \le i,j \le r}$ *and* $r \times 1$ *vector* $g(s) = (\exp\{-\eta|s_1 - s|\}, \ldots, \exp\{-\eta|s_r - s|\})^T$, *where* $s \ge s_r$. *From* f_i, h_i *form two* $rl \times 1$ *vectors* $f = (f_1^T, \ldots, f_r^T)^T$ *and* $h = (h_1^T, \ldots, h_r^T)^T$. *Then for any* $l \times l$ *matrix* H,

$$G^{-1}g(s) = (0, \ldots, 0, e^{-\eta(s-s_r)})^T, \tag{4.1}$$

$$\begin{aligned} f^T&(G \otimes H)^{-1}h \\ &= f_1^T H^{-1} h_1 \\ &\quad + \sum_{i=2}^{r} \frac{\left(f_i - e^{-\eta(s_i - s_{i-1})} f_{i-1}\right)^T H^{-1}\left(h_i - e^{-\eta(s_i - s_{i-1})} h_{i-1}\right)}{1 - e^{-2\eta(s_i - s_{i-1})}}, \end{aligned} \tag{4.2}$$

$$\det G = \prod_{i=2}^{r} \left(1 - e^{-2\eta(s_i - s_{i-1})}\right). \tag{4.3}$$

PROOF. Since $s \ge s_i$, $i = 1, \ldots, r$, we can write

$$g(s) = \exp\{-\eta(s - s_r)\} g(s_r).$$

But $g(s_r)$ is the last column of G. Thus $G^{-1}g(s_r) = (0, \ldots, 0, 1)^T$ and (4.1) follows.

To show (4.2), we first partition the matrix G into

$$G = \begin{pmatrix} G_1 & a \\ a^T & 1 \end{pmatrix}$$

and then use a matrix inversion formula [cf. Anderson (1984), Theorem A.3.1] to get

$$G^{-1} = \begin{pmatrix} I_{r-1} & -G_1^{-1}a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} G_1^{-1} & 0 \\ 0 & \left(1 - a^T G_1^{-1} a\right)^{-1} \end{pmatrix} \begin{pmatrix} I_{r-1} & 0 \\ -a^T G_1^{-1} & 1 \end{pmatrix}, \tag{4.4}$$

where I_{r-1} denotes the $(r-1)\times(r-1)$ identity matrix. From (4.1), we see that $G_1^{-1}a=(0,\dots,0,\exp\{-\eta|s_r-s_{r-1}|\})^T$. Meanwhile, from (4.4) and applying multiplication and inversion formulae (22) and (23) of Anderson [(1984), page 600] for the Kronecker product, we can get

$$(G\otimes H)^{-1}=\begin{pmatrix} I_{r-1}\otimes I_l & -(G_1^{-1}a)\otimes I_l \\ 0 & I_l \end{pmatrix}$$

$$\times\begin{pmatrix} (G_1\otimes H)^{-1} & 0 \\ 0 & (1-a^TG_1^{-1}a)^{-1}H^{-1} \end{pmatrix} \tag{4.5}$$

$$\times\begin{pmatrix} I_{r-1}\otimes I_l & 0 \\ -a^TG_1^{-1}\otimes I_l & I_l \end{pmatrix}.$$

With $f(r-1)=(f_1,\dots,f_{r-1})^T$ and $h(r-1)=(h_1,\dots,h_{r-1})^T$, (4.5) implies

$$f^T(G\otimes H)^{-1}h=f^T(r-1)(G_1\otimes H)^{-1}h(r-1)$$
$$+\frac{\left(f_r-e^{-\eta(s_r-s_{r-1})}f_{r-1}\right)^TH^{-1}\left(h_r-e^{-\eta(s_r-s_{r-1})}h_{r-1}\right)}{1-e^{-2\eta(s_r-s_{r-1})}},$$

which becomes (4.2) by induction. Finally (4.4) implies (4.3) by induction. □

LEMMA 2. *Let A, B, x, x_i and x_{ik} be defined as (2.2) and (2.3) and ξ_i and ζ_k as in Theorem 2. Then the loglikelihood function $l_{m,n}$ defined by (2.6) has the following representation*:

$$l_{m,n}(\lambda,\mu,\sigma^2)=mn\log(2\pi)+\log\sigma^2+\sum_{i=2}^{m}\log\left[\sigma^2(1-e^{-2\lambda\xi_i})\right]$$
$$+\sum_{k=2}^{n}\log\left[\sigma^2(1-e^{-2\mu\zeta_k})\right]$$
$$+\sum_{i=2}^{m}\sum_{k=2}^{n}\log\left[\sigma^2(1-e^{-2\lambda\xi_i})(1-e^{-2\mu\zeta_k})\right]$$
$$+\frac{1}{\sigma^2}\Bigg[x_1^TB^{-1}(\mu)x_1$$
$$+\sum_{i=2}^{m}\frac{\left(x_i-e^{-\lambda\xi_i}x_{i-1}\right)^TB^{-1}(\mu)\left(x_i-e^{-\lambda\xi_i}x_{i-1}\right)}{1-e^{-2\lambda\xi_i}}\Bigg].$$

PROOF. From formula (25) of Anderson [(1984), page 600] we get

$$\log\det[A(\lambda)\otimes B(\mu)] = \log\{[\det A(\lambda)]^n[\det B(\mu)]^m\}$$

$$= n\sum_{i=2}^{m}\log(1-e^{-2\lambda\xi_i}) + m\sum_{k=2}^{n}\log(1-e^{-2\mu\zeta_k}), \tag{4.6}$$

where the last equality follows from (4.3). Moreover, (4.2) implies that

$$x^T[A(\lambda)\otimes B(\mu)]^{-1}x = x_1^TB^{-1}(\mu)x_1$$

$$+\sum_{i=2}^{m}\frac{\left(x_i-e^{\lambda\xi_i}x_{i-1}\right)^TB^{-1}(\mu)\left(x_i-e^{-\lambda\xi_i}x_{i-1}\right)}{1-e^{-2\lambda\xi_i}}. \tag{4.7}$$

Hence Lemma 2 follows from (4.6), (4.7) and (2.6). □

LEMMA 3. *Let $b>0$. Then for all λ_1, λ_2 and δ in $[-b,b]$,*

$$M^{-1}|\delta| \le |1-e^{\delta}| \le M|\delta|, \tag{4.8}$$

$$|(1-e^{-\delta})^{-1}-\delta^{-1}-1/2| \le M|\delta|, \tag{4.9}$$

$$\left|\frac{1-e^{-\lambda_2\delta}}{1-e^{-\lambda_1\delta}}-\frac{\lambda_2}{\lambda_1}-\frac{\lambda_2(\lambda_1-\lambda_2)}{2\lambda_1}\delta\right| \le M\delta^2. \tag{4.10}$$

LEMMA 4. *For any constant $\delta>0$, there exists an $\eta>0$ such that*

$$\inf_{|\gamma-1|\ge\delta,\gamma>0}(\gamma-1-\log\gamma)\ge\eta.$$

Both Lemmas 3 and 4 follow easily via the Taylor expansion method. Moreover, through evaluations of the relevant covariances, the following lemma can be obtained.

LEMMA 5. *Let x_i, ξ_i and ζ_k be the same as those in Lemma 2.*

(i) *Let $\eta_i=(x_i-e^{-\lambda_0\xi_i}x_{i-1})/[(1-e^{-2\lambda_0\xi_i})^{1/2}\sigma_0]$, $i=2,\dots,m$. Then for each i, η_i is independent of $\{x_j,\ j=1,\dots,i-1\}$. Moreover, $\{\eta_i,\ i=2,\dots,m\}$ is a sequence of i.i.d. random vectors with $N(0,B(\mu_0))$ as their common distribution.*

(ii) *Let $w_{ik}=(\eta_{ik}-e^{-\mu_0\zeta_k}\eta_{i,k-1})/(1-e^{-2\mu_0\zeta_k})^{1/2}$, where η_{ik} is the kth component of η_i defined in* (i). *Then w_{ik} is independent of $\{x_{jl},\ j\le i-1$ or $l\le k-1\}$ for each i, k, and $\{w_{ik},\ i=2,\dots,m;\ k=2,\dots,n\}$ is a sequence of i.i.d. random variables with the standard normal distribution.*

Proofs of Theorems 1 and 2 involve approximations of the loglikelihood function by double arrays of normal random variables. The next lemma will be used to control magnitudes of these double arrays.

LEMMA 6. *Let $\{z_{ik}\}$ be a sequence of standard normal random variables.*

(i) *If z_{ik} are also independent and $m = O(n^q)$ for some $q > 0$, then*

$$\sup_{1 \le i \le m,\, 1 \le k \le n} |z_{ik}| = o(n^{\varepsilon}) \quad a.s., \tag{4.11a}$$

$$\sup_{1 \le i \le m} \left| \sum_{k=1}^{n} z_{ik} \right| = o(n^{1/2+\varepsilon}) \quad a.s., \tag{4.11b}$$

for every $\varepsilon > 0$.

(ii) *Suppose that $E(z_{i_1k_1} z_{i_2k_2} z_{i_3k_3} z_{i_4k_4}) \ne 0$ if and only if a permutation of $(i_1, k_1), \ldots, (i_4, k_4)$ can be made so that $i_1 = i_2$, $k_1 = k_2$, $i_3 = i_4$ and $k_3 = k_4$ and that $m = O(n)$. Then*

$$\sum_{i=1}^{m} \sum_{k=1}^{n} z_{ik} = o(n^{2-\varepsilon_0}) \quad a.s., \tag{4.12a}$$

$$\sum_{i=1}^{m} \delta_i^{1/2} \sum_{k=1}^{n} z_{ik} = o(n^{1-\varepsilon_0}) \quad a.s., \tag{4.12b}$$

for some $\varepsilon_0 > 0$, where $\delta_i = \delta_i(m, n) \ge 0$ satisfying $\sum_{i=1}^{m} \delta_i \le 1$.

PROOF. By Chebyschev's inequality,

$$P\left\{ \sup_{1 \le i \le m,\, 1 \le k \le n} |z_{i,k}| \ge n_0^{\varepsilon} \right\} \le mne^{-tn^{\varepsilon_0}} Ee^{t|z|}.$$

Since $m = O(n^q)$, the right-hand side of the above equation is summable and (4.11a) follows from the Borel–Cantelli lemma. Now let $\delta > 0$, $y_{ik} = \max(-n^{\delta}, \min(n^{\delta}, z_{ik}))$ and $S_i = \sum_{k=1}^{i}(y_{ik} - Ey_{ik})$. By Corollary 4.3.2 of Chow and Teicher (1978), we have $\mathrm{Var}(y_{ik}) \le \mathrm{Var}(z_{ik}) = \mathrm{Var}(z)$. Since $|y_{ik} - Ey_{ik}| \le 2n^{\delta}$, Lemma 10.2.1 of Chow and Teicher (1978) implies that, for some constant $c > 0$,

$$P\left\{ \sup_{1 \le i \le m} |S_i| \ge cn^{1-\delta} \right\} \le me^{-n^{1-2\delta}}.$$

Letting $1/2 > \delta > \max\{1/2 - \varepsilon_0, 0\}$, the Borel–Cantelli lemma again implies that $\sup_{1 \le i \le m} |S_i| = o(n^{1/2+\varepsilon_0})$ a.s. Thus (4.11b) holds, since $\sup_{1 \le i \le m} |S_i - \sum_{k=1}^{n} z_{ik}| = o(1)$ a.s. by (4.11a).

To show (4.12a), we again use Chebyshev's inequality and the assumption in (ii) to get

$$P\left\{ \left| \sum_{i=1}^{m} \sum_{k=1}^{n} z_{ik} \right| \ge n^{2-\varepsilon_0} \right\} \le n^{-4(2-\varepsilon_0)} E \left| \sum_{i=1}^{m} \sum_{k=1}^{n} z_{ik} \right|^4 = O(n^{-4(1-\varepsilon_0)}),$$

which is summable for $\varepsilon_0 < 3/4$. Hence (4.12a) follows from the Borel–Cantelli lemma. Equation (4.12b) can be shown similarly. □

PROOF OF THEOREM 1. To prove the theorem, we need the following two approximations that will be verified later on:

$$\begin{aligned}
&\sum_{i=2}^{m} \frac{\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)}{1 - e^{-2\lambda\xi_i}} \\
&\quad = \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu}\sum_{i=2}^{m}\sum_{k=2}^{n} w_{ik}^2 \\
&\qquad + \left[\frac{\lambda_0\sigma_0^2}{\lambda} + \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu}(\mu - \mu_0) + \frac{\lambda_0(\mu - \mu_0)^2\sigma_0^2}{2\lambda\mu}\right]m \qquad (4.13)\\
&\qquad + \left[\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu}(\lambda - \lambda_0) + \frac{\mu_0(\lambda - \lambda_0)^2\sigma_0^2}{2\lambda\mu}\right]n + o(n) \quad \text{a.s.},
\end{aligned}$$

$$x_1^T B^{-1}(\mu) x_1 = \frac{\mu_0\sigma_0^2}{\mu} n + o(n) \quad \text{a.s.}, \tag{4.14}$$

where w_{ik} are defined in Lemma 5(ii). Applying (4.10) we get

$$\begin{aligned}
&\sum_{i=2}^{m} \log\frac{\sigma^2(1 - e^{-2\lambda\xi_i})}{\sigma_0^2(1 - e^{-2\lambda_0\xi_i})} + \sum_{k=2}^{n} \log\frac{\sigma^2(1 - e^{-2\mu\zeta_k})}{\sigma_0^2(1 - e^{-2\mu_0\zeta_k})} \\
&\qquad + \sum_{i=2}^{m}\sum_{k=2}^{n} \log\frac{\sigma^2(1 - e^{-2\lambda\xi_i})(1 - e^{-2\mu\zeta_k})}{\sigma_0^2(1 - e^{-2\lambda_0\xi_i})(1 - e^{-2\mu_0\zeta_k})} \qquad (4.15)\\
&\quad = (m - 1)(n - 1)\log\frac{\lambda\mu\sigma^2}{\lambda_0\mu_0\sigma_0^2} + n(\lambda_0 - \lambda) + m(\mu_0 - \mu) \\
&\qquad + m\log\frac{\lambda\sigma^2}{\lambda_0\sigma_0^2} + n\log\frac{\mu\sigma^2}{\mu_0\sigma_0^2} + o(n).
\end{aligned}$$

Note that in applying (4.10), the requirement $(\lambda, \mu) \in C$, which is compact and does not contain the origin, is needed to ensure that the above approximation holds uniformly. Although it will not be pointed out explicitly, many subsequent approximations, derived from Lemma 3, also implicitly require this for the same purpose. From (4.13)–(4.15) follows, with probability 1,

$$\begin{aligned} l_{m,n}(\lambda, \mu, \sigma^2) &- l_{m,n}(\lambda_0, \mu_0, \sigma_0^2) \\ &= \left(\frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} - 1 \right) \sum_{i=2}^{m} \sum_{k=2}^{n} w_{ik}^2 - (m-1)(n-1) \log \frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} \\ &\quad + m \left[\frac{\lambda_0 \sigma_0^2}{\lambda \sigma^2} - 1 - \log \frac{\lambda_0 \sigma_0^2}{\lambda \sigma^2} + \left(\frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} - 1 \right)(\mu - \mu_0) \right. \\ &\qquad \left. + \frac{\lambda_0 (\mu - \mu_0)^2 \sigma_0^2}{2 \sigma^2 \lambda \mu} \right] \\ &\quad + n \left[\frac{\mu_0 \sigma_0^2}{\mu \sigma^2} - 1 - \log \frac{\mu_0 \sigma_0^2}{\mu \sigma^2} + \left(\frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} - 1 \right)(\lambda - \lambda_0) \right. \\ &\qquad \left. + \frac{\mu_0 \sigma_0^2 (\lambda - \lambda_0)^2}{2 \sigma^2 \lambda \mu} \right] + o(n). \end{aligned} \tag{4.16}$$

Since $\sum_{i=2}^{m} \sum_{k=2}^{n} (w_{ik}^2 - 1) = o(mn)$ a.s., (4.16) equals

$$(m-1)(n-1) \left(\frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} - 1 - \log \frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu \sigma^2} \right) + o(mn).$$

This and Lemma 4 entail

$$\hat{\lambda} \hat{\mu} \hat{\sigma}^2 \to \lambda_0 \mu_0 \sigma_0^2 \quad \text{a.s.} \tag{4.17}$$

Applying (4.17) to (4.16) we get

$$\begin{aligned} l_{m,n}(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2) &- l_{m,n}\left(\lambda_0, \mu_0, \frac{\hat{\lambda} \hat{\mu} \hat{\sigma}^2}{\lambda_0 \mu_0} \right) \\ &= m \left[\frac{\lambda_0 \sigma_0^2}{\hat{\lambda} \hat{\sigma}^2} - 1 - \log \frac{\lambda_0 \sigma_0^2}{\hat{\lambda} \hat{\sigma}^2} + \frac{\lambda_0 (\hat{\mu} - \mu_0)^2 \sigma_0^2}{2 \hat{\sigma}^2 \hat{\lambda} \hat{\mu}} \right] \\ &\quad + n \left[\frac{\mu_0 \sigma_0^2}{\hat{\mu} \hat{\sigma}^2} - 1 - \log \frac{\mu_0 \sigma_0^2}{\hat{\mu} \hat{\sigma}^2} + \frac{\mu_0 \sigma_0^2 (\hat{\lambda} - \lambda_0)^2}{2 \hat{\sigma}^2 \hat{\lambda} \hat{\mu}} \right] + o(n) \quad \text{a.s.,} \end{aligned} \tag{4.18}$$

which, by Lemma 4, converges to ∞ unless $\hat{\lambda} \to \lambda_0$ and $\hat{\mu} \to \mu_0$ a.s. Hence Theorem 1 holds in view of the definitions of $\hat{\lambda}$ and $\hat{\mu}$.

It remains to show (4.13) and (4.14). By writing $x_i - e^{-\lambda \xi_i} x_{i-1} = x_i - e^{-\lambda_0 \xi_i} x_{i-1} + (e^{-\lambda_0 \xi_i} - e^{-\lambda \xi_i}) x_{i-1}$,

$$\begin{aligned} \text{l.h.s. of } (4.13) &= \sigma_0^2 \sum_{i=2}^m \frac{1 - e^{-2\lambda_0 \xi_i}}{1 - e^{-2\lambda \xi_i}} \eta_i^T B^{-1}(\mu) \eta_i \\ &\quad + \sum_{i=2}^m \frac{\left(e^{-\lambda_0 \xi_i} - e^{-\lambda \xi_i}\right)^2}{1 - e^{-2\lambda \xi_i}} x_{i-1}^T B^{-1}(\mu) x_{i-1} \\ &\quad + 2\sigma_0^2 \sum_{i=2}^m \frac{\left(e^{-\lambda_0 \xi_i} - e^{-\lambda \xi_i}\right)\left(1 - e^{-2\lambda_0 \zeta_i}\right)^{1/2}}{1 - e^{-2\lambda \xi_i}} x_{i-1}^T B^{-1}(\mu) \eta_i \\ &= \text{I} + \text{II} + \text{III}, \quad \text{say}, \end{aligned} \tag{4.19}$$

where η_i are as in Lemma 5(i). From Lemma 1 with $l = 1$ and $H = 1$,

$$\begin{aligned} \text{II} &= \sum_{i=2}^m \frac{e^{-2\lambda_0 \xi_i}\left(1 - e^{-(\lambda - \lambda_0)\xi_i}\right)^2}{1 - e^{-2\lambda \xi_i}} \\ &\quad \times \left[\sum_{k=2}^n \frac{\left(x_{i-1,k} - e^{-\mu \zeta_k} x_{i-1,k-1}\right)^2}{1 - e^{-2\mu \zeta_k}} + x_{i-1,1}^2 \right] \\ &= (1 + o(1)) \frac{(\lambda - \lambda_0)^2}{2\lambda} \\ &\quad \times \sum_{i=2}^m \xi_i \sum_{k=2}^n \frac{\left(x_{i-1,k} - e^{-\mu \zeta_k} x_{i-1,k-1}\right)^2}{1 - e^{-2\mu \zeta_k}} + O(1) \quad \text{a.s.}, \end{aligned} \tag{4.20}$$

where the second equality follows from Lemma 3. Applying Lemma 3 first and then Lemma 6(i), we get

$$\sum_{i=2}^m \xi_i \sum_{k=2}^n \frac{\left(e^{-\mu_0 \zeta_k} - e^{-\mu \zeta_k}\right) x_{i-1,k-1}}{1 - e^{-2\mu \zeta_k}} \left(x_{i-1,k} - e^{-\mu_0 \zeta_k} x_{i-1,k-1}\right) = O(\sqrt{n}) \quad \text{a.s.},$$

which combined with (4.20) gives

$$\begin{aligned} \mathrm{II} &= (1+o(1))\frac{(\lambda-\lambda_0)^2}{2\lambda} \\ &\quad\times \sum_{i=2}^{m}\xi_i\sum_{k=2}^{n}\frac{\left(x_{i-1,k}-e^{-\mu_0\zeta_k}x_{i-1,k-1}\right)^2}{1-e^{-2\mu\zeta_k}} + O(\sqrt{n}) \\ &= (1+o(1))\frac{(\lambda-\lambda_0)^2\mu_0\sigma_0^2}{2\lambda\mu} \\ &\quad\times \sum_{i=2}^{m}\xi_i\sum_{k=2}^{n}\frac{\left(x_{i-1,k}-e^{-\mu_0\zeta_k}x_{i-1,k-1}\right)^2}{(1-e^{-2\mu_0\zeta_k})\sigma_0^2} + O(\sqrt{n}) \\ &= (1+o(1))\frac{(\lambda-\lambda_0)^2\mu_0\sigma_0^2}{2\lambda\mu}n + O(n^{1/2+\varepsilon}) \quad \text{a.s. for every } \varepsilon>0, \end{aligned} \tag{4.21}$$

where the last two equalities follow from Lemmas 3 and 6. For III we use Lemma 1 again to write

$$\begin{aligned} \mathrm{III} &= 2\sigma_0\sum_{i=2}^{m}\frac{(e^{-\lambda_0\xi_i}-e^{-\lambda\xi_i})(1-e^{-2\lambda_0\xi_i})^{1/2}}{1-e^{-2\lambda\xi_i}}x_{i-1,1}\eta_{i,1} \\ &\quad + 2\sigma_0\sum_{i=2}^{m}\frac{(e^{-\lambda_0\xi_i}-e^{-\lambda\xi_i})(1-e^{-2\lambda_0\xi_i})^{1/2}}{1-e^{-2\lambda\xi_i}} \\ &\quad\times\sum_{k=2}^{n}\frac{\left(x_{i-1,k}-e^{-\mu\zeta_k}x_{i-1,k-1}\right)\left(\eta_{ik}-e^{\mu\zeta_k}\eta_{i,k-1}\right)}{1-e^{-2\mu\zeta_k}}. \end{aligned} \tag{4.22}$$

Similar to the previous derivation of (4.21), Lemmas 3 and 6 can be applied repeatedly to (4.22) to get

$$\mathrm{III} = o(n) \quad \text{a.s.} \tag{4.23}$$

In view of (4.19) and Lemma 1,

$$\begin{aligned} \mathrm{I} &= \sigma_0^2\sum_{i=1}^{m}\frac{1-e^{-2\lambda_0\xi_i}}{1-e^{-2\lambda\xi_i}}\left[\eta_{i,1}^2+\sum_{k=2}^{n}\frac{\left(\eta_{ik}-e^{-\mu\zeta_k}\right)^2\eta_{i,k-1}}{1-e^{-2\mu\zeta_k}}\right] \\ &= (1+o(1))\frac{\lambda_0\sigma_0^2}{\lambda}\sum_{i=2}^{m}\eta_{i,1}^2 + \sigma_0^2\sum_{i=2}^{m}\frac{1-e^{-2\lambda_0\xi_i}}{1-e^{-2\lambda\xi_i}}\sum_{k=2}^{n}\frac{1-e^{-2\mu_0\zeta_k}}{1-e^{-2\mu\zeta_k}}w_{ik}^2 \\ &\quad + \sigma_0^2\sum_{i=2}^{m}\frac{1-e^{-2\lambda_0\xi_i}}{1-e^{-2\lambda\xi_i}}\sum_{k=2}^{n}\frac{\left(e^{-\mu_0\zeta_k}-e^{-\mu\zeta_k}\right)^2}{1-e^{-2\mu\zeta_k}}\eta_{i,k-1}^2 \\ &\quad + 2\sigma_0^2\sum_{i=2}^{m}\frac{1-e^{-2\lambda_0\xi_i}}{1-e^{-2\lambda\xi_i}}\sum_{k=2}^{n}\frac{e^{-\mu_0\zeta_k}-e^{-\mu\zeta_k}}{1-e^{-2\mu\zeta_k}}\eta_{i,k-1}\left(\eta_{ik}-e^{-\mu_0\zeta_k}\eta_{i,k-1}\right), \end{aligned} \tag{4.24}$$

recalling that $w_{ik} = (\eta_{ik} - e^{-\mu_0\zeta_k}\eta_{i,k-1})/(1 - e^{-2\mu_0\zeta_k})^{1/2}$. Applying Lemmas 3 and 6 we can get

$$\sum_{i=2}^{m} \eta_{i,1}^2 = \sum_{i=2}^{m} \left(\eta_{i,1}^2 - 1\right) + m - 1 = m + o(m^{1/2+\alpha}) \quad \text{a.s.,} \tag{4.25}$$

$$\begin{aligned} &\sum_{i=2}^{m} \sum_{k=2}^{n} \frac{1 - e^{-2\lambda_0\xi_i}}{1 - e^{-2\lambda\xi_i}} \frac{1 - e^{-2\mu_0\zeta_k}}{1 - e^{-2\mu\zeta_k}} w_{ik}^2 \\ &\quad = \frac{\lambda_0\mu_0}{\lambda\mu} \sum_{i=2}^{m} \sum_{k=2}^{n} w_{ik}^2 + \frac{\lambda_0\mu_0}{\lambda\mu}(\lambda - \lambda_0)n \\ &\qquad + \frac{\lambda_0\mu_0}{\lambda\mu}(\mu - \mu_0)m + o(n) \quad \text{a.s.,} \end{aligned} \tag{4.26}$$

$$\begin{aligned} &\sum_{i=2}^{m} \frac{1 - e^{-2\lambda_0\xi_i}}{1 - e^{-2\lambda\xi_i}} \sum_{k=2}^{n} \frac{\left(e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}\right)^2}{1 - e^{-2\mu\zeta_k}} \eta_{i,k-1}^2 \\ &\quad = \frac{\lambda_0(\mu - \mu_0)^2}{2\lambda\mu} m + o(n) \quad \text{a.s.,} \end{aligned} \tag{4.27}$$

$$\begin{aligned} &\sum_{i=2}^{m} \sum_{k=2}^{n} \frac{1 - e^{-2\lambda_0\xi_i}}{1 - e^{-2\lambda\xi_i}} \frac{e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}}{1 - e^{-2\mu\zeta_k}} \eta_{i,k-1}\left(\eta_{ik} - e^{-\mu_0\zeta_k}\eta_{i,k-1}\right) \\ &\quad = o(n) \quad \text{a.s.} \end{aligned} \tag{4.28}$$

Combining (4.24)–(4.28) with (4.21) and (4.23) we get (4.13). From Lemma 1 we get

$$\begin{aligned} x_1^T B^{-1}(\mu) x_1 &= \sum_{k=2}^{n} \frac{\left(x_{1,k} - e^{-\mu\zeta_k} x_{1,k-1}\right)^2}{1 - e^{-2\mu\zeta_k}} + x_{1,1}^2 \\ &= \sigma_0^2 \sum_{k=2}^{n} \frac{1 - e^{-2\mu_0\zeta_k}}{1 - e^{-2\mu\zeta_k}} \frac{\left(x_{1,k} - e^{-\mu_0\zeta_k} x_{1,k-1}\right)^2}{\sigma_0^2(1 - e^{-2\mu_0\zeta_k})} \\ &\quad + \sum_{k=2}^{n} \frac{\left(e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}\right)^2}{1 - e^{-2\mu\zeta_k}} x_{1,k-1}^2 \\ &\quad + 2\sum_{k=2}^{n} \frac{e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}}{1 - e^{-2\mu\zeta_k}} x_{1,k-1}\left(x_{1,k} - e^{-\mu_0\zeta_k} x_{1,k-1}\right) \end{aligned} \tag{4.29}$$

which, along with a similar argument using Lemmas 3 and 6(i) as before, gives (4.14). □

PROOF OF THEOREM 2. By setting the derivative of $l_{m,n}(\lambda, \mu, \sigma^2)$ with respect to σ^2 to be 0 at $(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2)$, we get

$$\hat{\sigma}^2 = \frac{1}{mn}\left\{x_1^T B^{-1}(\hat{\mu})x_1 + \sum_{i=2}^{m} \frac{\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)}{1 - e^{-2\hat{\lambda}\xi_i}}\right\} \tag{4.30}$$

$$= \frac{1}{mn}\left[\frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}} \sum_{i=2}^{m}\sum_{k=2}^{n} w_{ik}^2 + m\frac{\lambda_0\sigma_0^2}{\hat{\lambda}} + n\frac{\mu_0\sigma_0^2}{\hat{\mu}} + o_p(m)\right],$$

where the last equality follows from expansions (4.13) and (4.14). Therefore

$$\sqrt{(m-1)(n-1)}\left(\hat{\lambda}\hat{\mu}\hat{\sigma}^2 - \lambda_0\mu_0\sigma_0^2\right)$$

$$= \lambda_0\mu_0\sigma_0^2[(m-1)(n-1)]^{-1/2} \sum_{i=2}^{m}\sum_{k=2}^{n}\left(w_{ik}^2 - 1\right) + o_p(1).$$

Hence (3.2) holds by the central limit theorem.

To show (3.4), take differentiation

$$\frac{\partial}{\partial\lambda} l_{m,n}(\lambda, \mu, \sigma^2)$$

$$= n\sum_{i=2}^{m} \frac{2\xi_i e^{-2\lambda\xi_i}}{1 - e^{-2\lambda\xi_i}} + \frac{2}{\sigma^2}\sum_{i=2}^{m} \frac{\xi_i e^{-\lambda\xi_i} x_{i-1}^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)}{1 - e^{-2\lambda\xi_i}} \tag{4.31}$$

$$- \frac{2}{\sigma^2}\sum_{i=2}^{m} \frac{\xi_i e^{-2\lambda\xi_i}}{\left(1 - e^{-2\lambda\xi_i}\right)^2}\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right).$$

Applying expansion $\xi_i e^{-2\lambda\xi_i}/(1 - e^{-2\lambda\xi_i}) = (2\lambda)^{-1} - 2^{-1}\xi_i + O(\xi_i^2)$, we see that

$$\sum_{i=2}^{m} \frac{\xi_i e^{-2\hat{\lambda}\xi_i}}{\left(1-e^{-2\hat{\lambda}\xi_i}\right)^2}\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)$$

$$= \frac{1}{2\hat{\lambda}}\sum_{i=2}^{m} \frac{\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)}{1-e^{-2\hat{\lambda}\xi_i}}$$

$$-\frac{1}{2}\sum_{i=2}^{m} \frac{\xi_i}{1-e^{-2\hat{\lambda}\xi_i}}\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)$$

$$+ O(1)\sum_{i=2}^{m} \frac{\xi_i^2}{1-e^{-2\hat{\lambda}\xi_i}}\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right) \tag{4.32}$$

$$= \frac{1}{2\hat{\lambda}}\left(mn\hat{\sigma}^2 - x_1^T B^{-1}(\hat{\mu})x_1\right)$$

$$-\frac{1}{4\hat{\lambda}}\sum_{i=2}^{m}\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)$$

$$+ O(1)\sum_{i=2}^{m}\xi_i\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right) + o_p(n^{1/2}),$$

where we have used the first equation in (4.30) to get the second equality. Through tedious approximations involving Lemmas 1 and 3 and certain variance calculations, it can be shown that

$$\sum_{i=2}^{m}\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)$$

$$= \frac{2\lambda_0\mu_0\sigma_0^2}{\mu}\sum_{i=2}^{m}\xi_i\sum_{k=2}^{n}w_{ik}^2 + o_p(n^{1/2}), \tag{4.33}$$

$$\sum_{i=2}^{m}\xi_i\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right) = o_p(n^{1/2}), \tag{4.34}$$

$$x_1^T B^{-1}(\mu)x_1 = \frac{\mu_0\sigma_0^2}{\mu}\sum_{k=2}^{n} z_{2,k}^2 + O_p(1), \tag{4.35}$$

where $z_{ik} = (x_{i-1,k} - e^{\mu_0\zeta_k}x_{i-1,k-1})/[\sigma_0^2(1 - e^{-2\mu_0\zeta_k})]^{1/2}$. From (4.32)–(4.35) follows

$$\sum_{i=2}^{m} \frac{\xi_i e^{-2\hat{\lambda}\xi_i}}{(1-e^{-2\hat{\lambda}\xi_i})^2}\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)^T B^{-1}(\hat{\mu})\left(x_i - e^{-\hat{\lambda}\xi_i}x_{i-1}\right)$$

$$(4.36)\quad = \frac{1}{2\hat{\lambda}}\left(mn\hat{\sigma}^2 - \frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{k=2}^{n} z_{2,k}^2\right) - \frac{\lambda_0\mu_0\sigma_0^2}{2\hat{\lambda}\hat{\mu}}\sum_{i=2}^{m}\xi_i\sum_{k=2}^{n} w_{ik}^2 + o_p(n^{1/2}).$$

Likewise,

$$\sum_{i=2}^{m} \frac{\xi_i e^{-\lambda\xi_i}}{1-e^{-2\lambda\xi_i}} x_{i-1}^T B^{-1}(\mu)\left(x_i - e^{-\lambda\xi_i}x_{i-1}\right)$$

$$(4.37)\quad = \frac{\sqrt{2\lambda_0}\,\mu_0\sigma_0^2}{2\lambda\mu}\sum_{i=2}^{m}\xi_i^{1/2}\sum_{k=2}^{n} z_{ik}w_{ik} + \frac{(\lambda-\lambda_0)\mu_0\sigma_0^2}{2\lambda\mu}\sum_{i=2}^{m}\xi_i\sum_{k=2}^{n} z_{ik}^2 + o_p(n^{1/2}),$$

which, combined with (4.31) and (4.36), implies

$$(4.38)\quad \begin{aligned}\frac{\partial}{\partial\lambda} l_{m,n}\left(\hat{\lambda},\hat{\mu},\hat{\sigma}^2\right) = {} & -n\sum_{i=2}^{m}\xi_i + \frac{n(m-1)}{\hat{\lambda}} \\ & + \frac{\sqrt{2\lambda_0}\,\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{i=2}^{m}\sqrt{\xi_i}\sum_{k=2}^{n} z_{ik}w_{ik} \\ & + \frac{\left(\hat{\lambda}-\lambda_0\right)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{i=2}^{m}\xi_i\sum_{k=2}^{n} z_{ik}^2 \\ & - \frac{1}{\hat{\lambda}\hat{\sigma}^2}\left(mn\hat{\sigma}^2 - \frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{k=2}^{n} z_{2,k}^2\right) \\ & + \frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{i=2}^{m}\xi_i\sum_{k=2}^{n} w_{ik}^2 + o_p(n^{1/2}).\end{aligned}$$

Since l.h.s. of (4.38) = 0 by definition, it follows that

$$(4.39)\quad \begin{aligned}\sqrt{n}\left(\hat{\lambda}-\lambda_0\right) = {} & -\frac{\lambda_0}{(1+\lambda_0)\sqrt{n}} \\ & \times\sum_{k=2}^{n}\left[\sqrt{2\lambda_0}\sum_{i=2}^{m}\sqrt{\xi_i}\,z_{ik}w_{ik} + z_{2k}^2 - 1\right] + o_p(n^{1/2}).\end{aligned}$$

Interchanging λ with μ, we have

$$\sqrt{m}(\hat{\mu}-\mu_0) = -\frac{\mu_0}{(1+\mu_0)\sqrt{m}} \times \sum_{i=2}^{m}\left[\sqrt{2\mu_0}\sum_{k=2}^{n}\sqrt{\zeta_k}\, y_{ik}w_{ik} + y_{i,2}^2 - 1\right] + o_p(n^{1/2}), \tag{4.40}$$

where $y_{ik} = (x_{i,k-1} - e^{-\lambda_0 \xi_i} x_{i-1,k-1})/[\sigma_0^2(1-e^{-2\lambda_0\xi_i})]^{1/2}$. To prove (3.4), it suffices to show that for every t,

$$\sqrt{n}(\hat{\lambda}-\lambda_0) + t\sqrt{m}(\hat{\mu}-\mu_0) \to_{\mathscr{D}} N\big(0, 2\lambda_0^2/(1+\lambda_0) + 2t^2\mu_0^2/(1+\mu_0)\big). \tag{4.41}$$

In view of (4.39) and (4.40), (4.41) is equivalent to

$$\frac{1}{\sqrt{n}}\sum_{k=2}^{n}\left[\sqrt{2\lambda_0}\sum_{i=2}^{m}\sqrt{\xi_i}\, z_{ik}w_{ik} + z_{2,k}^2 - 1\right] + \frac{t}{\sqrt{m}}\sum_{i=2}^{m}\left[\sqrt{2\mu_0}\sum_{k=2}^{n}\zeta_k^{1/2} y_{ik}w_{ik} + y_{i,2}^2 - 1\right] \to_{\mathscr{D}} N\big(0, 2[\lambda_0 + 1 + t^2(\mu_0+1)]\big). \tag{4.42}$$

The left-hand side of (4.42) can be written as

$$\begin{aligned} &\sum_{k=2}^{n}\left[\frac{\sqrt{2\lambda_0}}{\sqrt{n}}\sum_{i=2}^{m}\sqrt{\xi_i}\, z_{ik}w_{ik} + z_{2,k}^2 - 1 + \frac{t\sqrt{2\mu_0}}{\sqrt{m}}\sum_{i=2}^{m}\zeta_k^{1/2}y_{ik}w_{ik}\right] \\ &\quad + \frac{t}{\sqrt{m}}\sum_{i=2}^{m}(y_{i,2}^2-1) \\ &= \sum_{k=2}^{n}\varepsilon_k(m,n) + \frac{t}{\sqrt{m}}\sum_{i=2}^{m}(y_{i,2}^2-1), \quad \text{say.} \end{aligned} \tag{4.43}$$

From Lemma 5 it is not difficult to see that for each fixed pair m and n, conditioning on $\{y_{i,2},\ i=1,\dots,m\}$, $\{\varepsilon_k(m,n)\}$, $k=2,\dots,n\}$ is a martingale difference sequence with respect to the σ-filtration $\mathscr{F}_k = \sigma\{x_{il},\ l \le k,\ i = 1,\dots,n\}$. Moreover, it can also be shown easily that

$$\sum_{k=2}^{n} E\{\varepsilon_k^2(m,n)|\mathscr{F}_{k-1}\} \to_P 2(\lambda_0 + 1 + t^2\mu_0), \tag{4.44}$$

$$\sum_{k=2}^{n} E\{\varepsilon_k^4(m,n)|\mathscr{F}_{k-1}\} \to_P 0. \tag{4.45}$$

From (4.44), (4.45) and a martingale central limit theorem (cf. Pollard, 1984, page 171) we obtain

$$\sum_{k=2}^{n}\varepsilon_k(m,n) \to_{\mathscr{D}} N\big(0, 2(\lambda_0 + 1 + t^2\mu_0)\big),$$

which implies (4.42) since $y_{i,2}^2 - 1$, $i = 2, \dots, m$ are i.i.d. with mean 0 and variance 2. Finally, (3.5) follows from (3.4) and (3.2). □

5. Extensions to higher-dimensional spaces. In this section we generalize the results of Section 3 to higher-dimensional spatial processes. Let $X(t)$, $t \in [0,1]^d$, denote a spatial Gaussian process with a multiplicative covariance function

$$\Gamma(t,s) = \sigma^2 \exp\{-\theta_1|t_1 - s_1| - \cdots - \theta_d|t_d - s_d|\}. \tag{5.1}$$

Suppose a sample of size $N = \prod_{i=1}^d n_i$ is taken at $\{(u_{k_1}^{(1)}, \dots, u_{k_d}^{(d)})\colon 1 \le k_i \le n_i, 1 \le i \le d\}$, where, without loss of generality, $u_1^{(i)} < u_2^{(i)} < \cdots < u_{n_i}^{(i)}$, $i = 1, \dots, d$. Therefore the set of observations consists of

$$x_{k_1 \cdots k_d} \triangleq X\left(u_{k_1}^{(1)}, \dots, u_{k_d}^{(d)}\right), \qquad 1 \le k_i \le n_i,\ i = 1, \dots, d. \tag{5.2}$$

Again we do not assume sets of design points at different stages to be nested. By large sample, we shall mean that the set of sample observations from X becomes denser in the following sense:

$$\begin{aligned}&n_i \text{ are of the same order in the sense that}\\ &0 < \liminf_{N\to\infty} n_i/N^{1/d} \le \limsup_{N\to\infty} n_i/N^{1/d} < \infty \text{ for } i = 1, \dots, d.\end{aligned} \tag{5.3}$$

$$\max_{1 \le i \le p;\, 1 \le k_i \le n_i + 1} \left(u_{k_i}^{(i)} - u_{k_i - 1}^{(i)}\right) \to 0, \quad \text{where } u_0^{(i)} = 0,\ u_{n_i+1}^{(i)} = 1. \tag{5.4}$$

Let $\Delta_k^{(i)} = u_k^{(i)} - u_{k-1}^{(i)}$ and

$$A_i(\theta_i) = \left(\exp\{-\theta_i|u_k^{(i)} - u_l^{(i)}|\}\right)_{1 \le k, l \le n_i}. \tag{5.5}$$

Then the likelihood function can be written as

$$\begin{aligned} L(\theta, \sigma^2) &= (2\pi\sigma^2)^{-N/2}\left[\det\left(\bigotimes_{i=1}^d A_i(\theta_i)\right)\right]^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2} x^T \left[\bigotimes_{i=1}^d A_i(\theta_i)\right]^{-1} x\right\} \end{aligned} \tag{5.6}$$

with x being the $N \times 1$ column vector with entries $x_{k_1 \cdots k_d}$. Similar to Theorems 1 and 2 for two-dimensional processes, the following consistency and asymptotic normality results hold for the maximum likelihood estimate.

THEOREM 3. *Let C be a compact set containing $\theta^{(0)}$ as an interior point and let $(\hat\theta, \hat\sigma^2)$ be the maximum likelihood estimator that maximizes the likelihood function L over $C \times (0, \infty)$. Then under the assumptions* (5.3) *and* (5.4),

$$\left(\hat\theta_1, \dots, \hat\theta_d, \hat\sigma^2\right) \to \left(\theta_1^{(0)}, \dots, \theta_d^{(0)}, \sigma_0^2\right) \quad a.s., \tag{5.7}$$

where $(\theta_1^{(0)}, \dots, \theta_d^{(0)}, \sigma_0^2)$ denotes the true parameter vector.

Because of space limitation and its similarity to the proof of Theorem 1, the proof of Theorem 3 will be omitted.

THEOREM 4. *With the same notation and assumptions as in Theorem* 3, *suppose that* $\Delta_2^{(i)} = \cdots = \Delta_{n_i}^{(i)} = \Delta^{(i)}$ *for every* $i = 1, \dots, d$. *Then*

$$N^{1/2}\left(\hat{\sigma}^2 \prod_{i=1}^{d} \hat{\theta}_i - \sigma_0^2 \prod_{i=1}^{d} \theta_i^{(0)}\right) \to_{\mathscr{D}} N\left(0, 2\left(\sigma_0^2 \prod_{i=1}^{d} \theta_i^{(0)}\right)^2\right), \tag{5.8}$$

$$\begin{pmatrix} (N/n_1)^{1/2}(\hat{\theta}_1 - \theta_1^{(0)}) \\ \vdots \\ (N/n_d)^{1/2}(\hat{\theta}_d - \theta_d^{(0)}) \end{pmatrix} \to_{\mathscr{D}} N(0, \Sigma_d), \tag{5.9}$$

where Σ_d *is a diagonal matrix* $\operatorname{diag}\{2\theta_1^{(0)^2}/(1 + \theta_1^{(0)}), \dots, 2\theta_d^{(0)^2}/(1 + \theta_d^{(0)})\}$. *Moreover, if* $N^{1/d}/n_i \to \rho_i > 0$, $i = 1, \dots, d$, *then*

$$N^{(d-1)/(2d)}\begin{pmatrix} \hat{\theta} - \theta^{(0)} \\ \hat{\sigma}^2 - \sigma_0^2 \end{pmatrix} \to_{\mathscr{D}} N\left(0, \begin{pmatrix} \tilde{\Sigma}_d & b \\ b^T & 2\sigma_0^4 \sum_{i=1}^{d} \left(\prod_{j \neq i} \rho_j\right) / (1 + \theta_i^{(0)}) \end{pmatrix}\right), \tag{5.10}$$

where $\tilde{\Sigma}_d = \operatorname{diag}\{2(\prod_{i=2}^{d}\rho_i)\theta_1^{(0)^2}/(1 + \theta_1^{(0)}), \dots, 2(\prod_{i=1}^{d-1}\rho_i)\theta_d^{(0)^2}/(1 + \theta_d^{(0)})\}$, $b^T = (b_1, \dots, b_d)$ *with* $b_i = -2(\prod_{j \neq i}^{d}\rho_j)\sigma_0^2\theta_i^{(0)}/(1 + \theta_i^{(0)})$.

The strong consistency and weak convergence (5.7)–(5.10) are multivariate versions of (3.1), (3.2), (3.4) and (3.5). Although Theorem 3 is a complete extension of Theorem 1 to d-dimensional spatial processes, the assumption $\Delta_k^{(i)} = \Delta^{(i)}$ in Theorem 4 is a much stronger requirement than (3.3). It is likely that our conditions may be relaxed. The convergence rate in (5.9) and (5.10) is $N^{-(d-1)/(2d)}$, which becomes faster as d gets larger. One intuitive explanation for this is that as d increases, the level of correlation decreases.

PROOF OF THEOREM 4. The proof of Theorem 4 is more delicate since we need to handle all the terms in the approximations of $(\partial/\partial\theta_i)l(\theta, \sigma^2)$ except $o(N^{(d-1)/(2d)})$. As we shall see, it not only uses techniques in the proof of Theorem 2, but also depends on the assumption that the partition in each coordinate is equally spaced. The full proof is tedious and notationally prohibitive. So only a brief sketch will be presented.

First from equation $(\partial/\partial\sigma^2)l(\hat{\theta},\hat{\sigma}^2) = 0$ we get

$$\hat{\sigma}^2 = \frac{1}{N}\left\{x_1^T B_1^{-1}(\hat{\theta})x_1 + \sum_{k=2}^{n_1} \frac{\left(x_k - e^{-\hat{\theta}^{(1)}\Delta^{(1)}}x_{k-1}\right)^T B_1^{-1}(\hat{\theta})\left(x_k - e^{-\hat{\theta}^{(1)}\Delta^{(1)}}x_{k-1}\right)}{1 - e^{-2\hat{\theta}_1\Delta^{(1)}}}\right\}, \tag{5.11}$$

where x_k are $(N/n_1) \times 1$ vectors derived from the vector x by deleting all entries $x_{k_1,\dots,k_p}$ with $k_1 \neq k$ and $B_1(\theta) = \otimes_{j=2}^{p} A_j(\theta_j)$. Since $(1 - e^{-2\theta_i^{(0)}\Delta^{(i)}})/(1 - e^{-2\theta_i\Delta^{(i)}}) = 1 + e^{-2\theta_i\Delta^{(i)}}(1 - e^{2(\theta_i - \theta_i^{(0)})\Delta^{(i)}})/(1 - e^{-2\theta_i\Delta^{(i)}})$, we can use a similar argument as that leads to (4.10) to get

$$\frac{1 - e^{-2\theta_i^{(0)}\Delta^{(i)}}}{1 - e^{-2\theta_i\Delta^{(i)}}} = \frac{\theta_i^{(0)}}{\theta_i} + (1 + o(1))\frac{\theta_i^{(0)}}{\theta_i}(\theta_i - \theta_i^{(0)})\Delta^{(i)}. \tag{5.12}$$

From (5.11), (5.12) and Theorem 3, it can be shown that

$$\begin{aligned}\hat{\sigma}^2 = \frac{1}{N}\Bigg\{&\frac{\sigma_0^2\prod_{i=1}^d\theta_i^{(0)}}{\prod_{i=1}^d\hat{\theta}_i}\sum_{k_1=2}^{n_1}\cdots\sum_{k_d=2}^{n_d} w_{k_1\cdots k_d}^2 \\ &+ \sum_{i=1}^{d}\frac{N}{n_i}\frac{\sigma_0^2\prod_{j=1}^d\theta_j^{(0)}}{\prod_{j=1}^d\hat{\theta}_j}(\hat{\theta}_i - \theta_i^{(0)})(1 + o(1)) \\ &+ \sum_{i=1}^{d}\sum_{1\le l_1<\cdots<l_i\le d}\left(\prod_{j\notin\{l_1\cdots l_i\}}(n_j - 1)\right)\frac{\sigma_0^2\prod_{j\notin\{l_1\cdots l_i\}}\theta_j^{(0)}}{\prod_{j\notin\{l_1\cdots l_i\}}\hat{\theta}_j} + o_p(N^{1/2})\Bigg\},\end{aligned} \tag{5.13}$$

where $w_{k_1\cdots k_d}$ and i.i.d. $N(0,1)$ random variables. From (5.11), it follows that

$$\begin{aligned}\left(\hat{\sigma}^2\prod_{i=1}^d\hat{\theta}_i - \sigma_0^2\prod_{i=1}^d\theta_i^{(0)}\right) &= \frac{\sigma_0^2\prod_{i=1}^d\theta_i^{(0)}}{N}\sum_{k_1=2}^{n_1}\cdots\sum_{k_d=2}^{n_d}\left(w_{k_1\cdots k_d}^2 - 1\right) \\ &\quad + O_p\left(\frac{\max_{1\le i\le d}|\hat{\theta}_i - \theta_i^{(0)}|}{N^{1/d}}\right) + o_p(N^{-1/2}) \\ &= O_p(N^{-1/2}) + O_p\left(\frac{\max_{1\le i\le d}|\hat{\theta}_i - \theta_i^{(0)}|}{N^{1/d}}\right).\end{aligned} \tag{5.14}$$

Thus (5.8) follows if (5.9) holds.

We now prove (5.9). From (5.11),

$$\begin{aligned}\frac{\partial}{\partial\theta_1}l(\hat{\theta},\hat{\sigma}^2) = &\frac{N(n_1 - 1)}{n_1}\frac{2\Delta^{(1)}e^{-2\hat{\theta}_i\Delta^{(1)}}}{1 - e^{-2\hat{\theta}_1\Delta^{(1)}}} \\ &+ \frac{2\Delta^{(1)}e^{-\hat{\theta}_1\Delta^{(1)}}}{\hat{\sigma}^2(1 - e^{-2\hat{\theta}_1\Delta^{(1)}})}\sum_{k=2}^{n_1} x_{k-1}^T B_1^{-1}(\hat{\theta})\left(x_k - e^{-\hat{\theta}^{(1)}\Delta^{(1)}}x_{k-1}\right) \\ &- \frac{2\Delta^{(1)}e^{-2\hat{\theta}_1\Delta^{(1)}}}{\hat{\sigma}^2(1 - e^{-2\hat{\theta}_1\Delta^{(1)}})}\left[\left(\prod_{j=1}^d n_j\right)\hat{\sigma}^2 - x_1^T B_1^{-1}(\hat{\theta})x_1\right] = 0.\end{aligned} \tag{5.15}$$

Applying (5.15) and (5.12) we can show that

$$
\begin{aligned}
\frac{\partial}{\partial\theta_1} l(\hat{\theta}, \hat{\sigma}^2) = {} & \frac{\sigma_0^2 \prod_{j=2}^d \theta_j^{(0)}}{\hat{\sigma}^2 \prod_{j=1}^d \hat{\theta}_j} \sum_{k_2=2}^{n_d} \cdots \sum_{k_d=2}^{n_d} \left(z_{k_2 \cdots k_d}^2(1) - 1 \right) \\
& + \prod_{j=2}^d (n_j - 1) \left[\frac{\sigma_0^2 \prod_{j=2}^d \theta_j^{(0)}}{\hat{\sigma}^2 \prod_{j=1}^d \hat{\theta}_j} - \frac{1}{\hat{\theta}_1} \right] \\
& + \left(\hat{\theta}_1 - \theta_1^{(0)} \right) \frac{\sigma_0^2 \prod_{j=2}^d \theta^{(0)}}{\hat{\sigma}^2 \prod_{j=1}^d \hat{\theta}_j} \sum_{k_1=2}^{n_1} \Delta^{(1)} \sum_{k_1=2}^{n_2} \cdots \sum_{k_k=2}^{n_d} z_{k_2 \cdots k_d}^2(k_1 - 1) \\
& + \frac{\sigma_0^2 \prod_{j=2}^d \theta_j^{(0)}}{\hat{\sigma}^2 \prod_{j=1}^d \hat{\theta}_j} \sum_{k_1=2}^{n_1} \sqrt{1 - e^{-2\theta_1^{(0)} \Delta^{(1)}}} \\
& \times \sum_{k_2=2}^{n_2} \cdots \sum_{k_d=2}^{n_d} z_{k_2 \cdots k_d}(k_1 - 1) w_{k_1 \cdots k_d} \\
& + \sum_{i=1}^d o_p\left(\left(\hat{\theta}_i - \theta_i^{(0)} \right) N^{(d-1)/d} \right) + o_p\left(N^{(d-1)/(2d)} \right),
\end{aligned} \tag{5.16}
$$

where for each $1 \le k \le n_1 - 1$, $z_{k_2 \cdots k_d}(k)$ is defined as follows. Let

$$
z_{k_2 \cdots k_d}^{(1)}(k) = x_{k k_2 \cdots k_d}, \qquad z_{k_2 \cdots k_d}^{(2)}(k) = z_{k_2 \cdots k_d}^{(1)}(k) - e^{\Delta^{(2)}} z_{k_2 - 1, k_3 \cdots k_d}^{(1)}(k), \ldots,
$$

$$
z_{k_2 \cdots k_d}^{(d)}(k) = z_{k_2 \cdots k_d}^{(d-1)}(k) - e^{\Delta^{(d)}} z_{k_2 \cdots k_{d-1}, k_{d-1}}^{(d)}(k).
$$

Then $z_{k_2 \cdots k_d}(k) = z_{k_2 \cdots k_d}^{(d)}(k)$. In view of (5.14) and Theorem 3, (5.16) implies that

$$
\begin{aligned}
& \left(\frac{1}{\theta_1^{(0)}} + 1 \right) \left(\prod_{i=2}^d (n_i - 1) \right) \left(\theta_1^{(0)} - \hat{\theta}_1 \right) \\
& \quad = \sum_{k_2=2}^{n_2} \cdots \sum_{k_d=2}^{n_d} \left(z_{k_2 \cdots k_d}^2(1) - 1 \right) \\
& \qquad + \sum_{k_1=2}^{n_1} \sqrt{2\theta_1^{(0)} \Delta^{(1)}} \sum_{k_2=2}^{n_2} \cdots \sum_{k_d=2}^{n_d} z_{k_2 \cdots k_d}(k_1 - 1) w_{k_1 \cdots k_d} \\
& \qquad + \sum_{i=1}^d o_p\left(N^{(d-1)/d} \left(\hat{\theta}_i - \theta_i^{(0)} \right) \right) + o_p\left(N^{(d-1)/(2d)} \right).
\end{aligned} \tag{5.17}
$$

By symmetry, approximations similar to (5.17) hold for θ_i, $i = 2, \ldots, d$. From these d equations and a similar argument as in the proof of Theorem 2 we get (5.9). Finally, (5.10) is an immediate consequence of (5.8) and (5.9). □

Acknowledgments. I wish to thank Professor Jerome Sacks for introducing me to this problem and for many subsequent discussions. I also thank Professor Adam Martinsek and an anonymous referee for their constructive comments. The example in Remark 2 of an isotropic random field with nonidentifiable parameters was suggested by the referee.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability and Martingales*. Springer, New York.

CRESSIE, N. (1991). *Statistics for Spatial Data*. Wiley, New York.

CURRIN, C., MITCHELL, T., MORRIS, M. and YLVISAKER, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.* **86** 953–963.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes*. Springer, New York.

MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

RIPLEY, B. D. (1981). *Spatial Statistics*. Wiley, New York.

SACKS, J., SCHILLER, S. B. and WELCH, W. J. (1989). Designs for computer experiments. *Technometrics* **31** 41–47.

SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments (with discussion). *Statist. Sci.* **4** 409–435.

SACKS, J. and YLVISAKER, D. (1970). Statistical designs and integral approximation. In *Proceedings of the 12th Biennial Seminar of the Canadian Mathematical Congress* (R. Pyke, ed.) 115–136. Canadian Math. Congress, Montreal.

SKOROKHOD, A. V. and YADRENKO, M. I. (1973). On absolute continuity of measures corresponding to homogeneous Gaussian fields. *Theory Probab. Appl.* **18** 27–40.

STEIN, M. L. (1987). Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.* **82** 765–772.

SWEETING, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8** 1357–1381.

YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions I*. Springer, New York.

YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.* **36** 280–296.

DEPARTMENT OF STATISTICS
101 ILLINI HALL
725 SOUTH WRIGHT STREET
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820