The Annals of Probability
1991, Vol. 19, No. 4, 1664–1678

A NONHOMOGENEOUS MARKOV PROCESS FOR THE ESTIMATION OF GAUSSIAN RANDOM FIELDS WITH NONLINEAR OBSERVATIONS

BY YALI AMIT[1] AND MAURO PICCIONI[2]

Brown University and Universitá di Roma II

We consider an estimation problem in which the signal is modelled by a continuous Gaussian random field and is observed through smooth and bounded nonlinear sensors. A nonhomogeneous Markov process is defined in order to sample the conditional distribution of the signal given the observations. At any finite time the process takes values in a finite-dimensional space, although the dimension goes to infinity in time. We prove that the empirical averages of any bounded functional continuous w.p.1 converge in the mean square to the conditional expectation of the functional.

1. Introduction. In this paper the following general estimation problem is considered. The signal is modelled by an n-dimensional continuous Gaussian random field on a bounded and smooth domain in R^d, whose covariance is the Green's tensor of some uniformly elliptic differential system. In order to obtain a measure on the space of continuous functions, the order of the system must be greater than some constant which depends on the dimension of the space. In particular, such a field can be the solution of a system of stochastic partial differential equations [see Dembo and Zeitouni (1990)]. This signal is observed through a finite set of bounded and smooth nonlinear sensors corrupted by noise. The continuity of the signal also allows for pointwise measurements. In order to recover information on the original signal it is then required to compute conditional expectations of functionals of the signal given the observations or to sample from the conditional distribution of the signal given the observations.

Due to the nonlinearity in the sensors, the conditional distribution of the field is no longer Gaussian. Therefore estimation cannot be reduced to computing the linear regression [see Piccioni and Roma (1990)], nor is there any direct way to sample from such a distribution. In the finite-dimensional case a common approach has been to simulate a homogeneous Markov process whose unique invariant distribution is the desired one [see Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) and Geman and Geman (1984)]. The same idea has been extended to infinite dimensions in the framework of

Received April 1989; revised June 1990.

[1]Mainly supported by the Weizmann Fellowship and partially supported by ARO DAAL-03-86-K-0110 and ARO DAAL-03-86-K-0171.

[2]Supported by a NATO–CNR Advanced Fellowship at the Division of Applied Mathematics, Brown University.

AMS 1980 *subject classifications*. Primary 60G60; secondary 60H15.

Key words and phrases. Gaussian random fields, nonhomogeneous Markov processes, estimation, Galerkin approximations.

stochastic quantization of field theory [Jona-Lasinio and Mitter (1983) and Borkar, Chari and Mitter (1988)]. The random fields we deal with are similar to those treated in the preceding references, with the log-likelihood interpreted as an interaction term, but we only consider bounded interactions.

An alternative to the construction of an infinite-dimensional process is to work with an increasing sequence of finite-dimensional subspaces of the space of continuous functions, on each of which a Galerkin approximation to the conditional field is constructed. Under appropriate conditions these approximations converge in distribution. In each of these subspaces one can then simulate a finite-dimensional homogeneous Markov process in order to obtain an approximation to the infinite-dimensional distribution.

In this paper a more refined simulation scheme is suggested. A nonhomogeneous Markov process is constructed which passes through this sequence of subspaces in time. In each subspace the process evolves for a fixed interval of time like the finite-dimensional homogeneous Markov process corresponding to that subspace. At the end of each time interval the process moves up to the next subspace according to a prescribed rule. We are able to show that this process converges in distribution to the conditional field and that the empirical averages of any bounded measurable function, continuous w.p.1, converge in mean square to the conditional expectation of the function.

The main ingredients in obtaining such a result are the fact that the second eigenvalue of the generator of all the finite-dimensional Markov processes can be uniformly bounded from below together with the boundedness assumptions on the interaction. This allows us to obtain stronger results than in the theory of nonhomogeneous Markov processes arising in simulated annealing [see Gidas (1985) and Holley and Stroock (1988)].

The motivation for this study comes from certain problems in image analysis in which Gaussian random fields model prior information on the variability of shapes [see Grenander (1970) and Amit, Grenander and Piccioni (1990)].

The model described in the later paper considers an image as a function from the unit square $I^2 \in R^2$ to the unit interval I, and assumes that the family of images under consideration are deformations of a fixed smooth and bounded function r_0: $R^2 \to I$ with support in I^2 called the template. More precisely any image r: $I^2 \to I$ is given as $r(x) = r_0(x + U(x))$, where $U(x)$ is a continuous mapping from I^2 into R^2. A Gaussian prior distribution is assumed on the space of continuous mappings $C(I^2, R^2)$, namely, a mean zero Gaussian random field on I^2 with continuous sample paths. This prior reflects the assumptions on the variability to be expected within the given family of images. The zero mean implies that all images in the family will be centered in some sense around the template image r_0.

It is assumed that we observe a degraded version of the real image r at a finite number of pixels, that is, $y_i = r(x_i) + n_i$, $i = 1, \dots, q$, where n_i are i.i.d. $N(0, \sigma)$. Using Bayes' formula one obtains a posterior distribution of the field U given an observation y. This posterior has the form of a perturbation of the Gaussian prior with a nonlinear term given by $1/(\sigma^2)\Sigma_{i=1}^{q}(y_i - r_0(x_i +$

$U(x_i)))^2$. This term can be considered as the interaction term. Clearly the boundedness assumption on the interaction is natural in this setting due to the assumptions on r_0.

The restoration consists of sampling from the posterior distribution via a process similar to the one described above. Any sample mapping U is applied to the template to obtain a restored image $r(x) = r_0(x + U(x))$.

This model for image restoration allows us to incorporate knowledge of what we expect to find in the image, that is, variations of the template r_0. In that sense it is more powerful than restoration techniques that only employ smoothness assumptions on the image itself. Moreover the restoration provides a transformation from the template to the true image, thus allowing for a structural understanding of the picture. For more details see Amit, Grenander and Piccioni (1990).

2. Continuous Gaussian random fields. We first address the issue of which differential operators generate Gaussian random fields with continuous sample configurations [see Dudley (1973) and Itô (1984)]. The material in this section is known but is included to make the paper accessible to nonexperts. For this purpose we recall some basic notions on linear random functionals.

Let $(H, (\cdot, \cdot))$ be a real separable Hilbert space. Let $\mathscr{L}_0^2$ denote the Hilbert space of all random variables with zero mean and finite variance on some probability space $(\Omega, \mathscr{F}, P)$. A linear random functional (LRF) ϕ on H is a bounded linear mapping from H into $\mathscr{L}_0^2$. Two LRF's ϕ and ψ are considered the same if, for any $f \in H$, $\phi(f)$ and $\psi(f)$ have the same distribution. Given a LRF ϕ and a bounded linear transformation A from H into another Hilbert space K we define a LRF $A\phi$ on K by $A\phi(h) = \phi(A^*h)$.

If $\phi(f)$ is Gaussian for any $f \in H$, the LRF is called Gaussian. It is nondegenerate if the variance of $\phi(f)$ is nonzero for all $f \in H$. The mapping $(f, g) \to E(\phi(f)\phi(g))$ is a bounded positive symmetric bilinear form on H. Consequently there exists a bounded strictly positive self-adjoint linear operator C, for short a covariance, which satisfies $E(\phi(f)\phi(g)) = (Cf, g)$.

A Gaussian LRF is completely specified by its covariance operator. Moreover it is easy to produce a Gaussian LRF with any prescribed covariance C. First observe that if ϕ has covariance S, then $A\phi$ has covariance ASA^*. Thus if n has covariance I, then $C^{1/2}n$ has covariance C. The LRF n is easily obtained by taking a sequence $\{n_i\}_{i=1}^{\infty}$ of i.i.d. standard Gaussian random variables and defining

$$n(f) = \sum_{i=1}^{\infty} (f, e_i) n_i, \tag{1}$$

where e_i, $i = 1, \dots,$ is any complete orthonormal basis of H.

Now observe that $(Cf, f)^{1/2} = |f|_C$ is a Hilbertian norm in H, generally weaker than $|f|_H = (f, f)^{1/2}$. Let $\tilde{V}$ be the completion of H with respect to $|\cdot|_C$. The LRF ϕ can be continuously extended to $\tilde{V}$. By taking $V = C^{1/2}H$ with the Hilbertian norm $B(f, f) = |C^{-1/2}f|^2$, then $\tilde{V}$ is isomorphic to the dual V^* of V, and we have the dense continuous embeddings of Hilbert spaces

$$V \hookrightarrow_i H \hookrightarrow_{i^*} V^*. \tag{2}$$

Conversely, any closed Hilbertian norm $B(\cdot, \cdot)$ on a dense subspace V of H such that $B(f, f) \geq c(f, f)$ for some $c > 0$ corresponds to a unique covariance C. In fact (2) is immediately obtained by identifying H with its dual. By the Riesz representation theorem there is an isomorphism J of V^* onto V such that $B(J\eta, v) = \eta(v)$ for all $v \in V$, $\eta \in V^*$ and the operator $C = iJi^*$ is a covariance on H. Moreover $V = C^{1/2}H$ and $B(f, f) = |C^{-1/2}f|^2$. We call B the energy form of the LRF.

In what follows we assume that $[H_0^m(D)]^n \subset V \subset [H^m(D)]^n$ for some positive integers m and n, D being a bounded and smooth domain in R^d. We take

$$B(u,v) = \sum_{|p|,|q| \leq m} \sum_{i,j=1}^{n} \int_D a_{p,q}^{i,j} D^q u_j D^p v_i \, dx,$$

where $a_{p,q}^{i,j}$ are smooth in $\overline{D}$, and assume it to be equivalent in V to the Sobolev norm $\|f\|_m^2 = \sum_{j=1}^n |f_j|_m^2$ where $|g|_m^2 = \sum_{|\beta| \leq m} \int |D^\beta g|^2$. The preceding conditions are equivalent to the well posedness of the differential system $B(u,v) = \eta(v)$, $\forall\ v \in V$, where $\eta \in V^*$ [see Agmon, Douglis and Nirenberg (1964)]. By choosing $H = [L_2(D)]^n$ the embeddings in (2) are obtained.

Observe that the preceding setting includes the following two relevant examples:

1. $B(u,v) = (Pu, v)$, where P is a symmetric uniformly elliptic system of differential operators of order $2m$ with Dirichlet boundary conditions. In this case $V = [H_0^m(D)]^n$.
2. $B(u,v) = (Qu, Qv)$, where Q is a uniformly elliptic system of differential operators of order m (even) with Dirichlet boundary conditions (not necessarily symmetric). By elliptic regularity [see Agmon, Douglis and Nirenberg (1964)], V in this case is $[H_0^{m/2}(D) \cap H^m(D)]^n$. In this case the corresponding field can be considered as a weak solution of the stochastic partial differential equation $Q\phi = n$, with n as in (1) [see Dembo and Zeitouni (1990) and Piccioni (1987)].

The following kind of Sobolev embedding theorem will allow us to realize LRF's defined on $L_2(D)$ as $\overline{D}$ indexed Gaussian random fields [see Adams (1975)].

PROPOSITION 1. *For any $m \geq m_d^* = [d/2 + 1]$ the space $H^m(D)$ is continuously embedded into a space $C^\alpha(\overline{D})$ for some $0 < \alpha < 1$, that is, there exists a constant $\kappa > 0$ depending only on the domain D such that*

$$\left(\sup_{x \in \overline{D}} |g(x)| \right) + \left(\sup_{x, y \in \overline{D}} \frac{|g(x) - g(y)|}{|x - y|^\alpha} \right) \leq \kappa |g|_m, \tag{3}$$

for all $g \in H^m(D)$.

From (3) it is immediately seen that the linear functional δ_x^i corresponding to the evaluation of the jth component at a point $x \in \overline{D}$ is bounded on V, so it lies in the dual V^*, for any $m \geq m_d^*$. Given a LRF ϕ with energy form B we can therefore define an n-dimensional random vector field X indexed by $x \in \overline{D}$ as

$$X(x) = \left(\phi(\delta_x^1), \dots, \phi(\delta_x^n)\right).$$

The field X has continuous sample configurations. To establish this let us recall the Kolmogorov criterion for random fields [see, e.g., Karatzas and Shreve (1988), Chapters 2, 2.9 and 4.11].

PROPOSITION 2. *Let $X^{(k)}(x)$, $k = 1, 2, \dots,$ be n-dimensional random fields indexed by $\overline{D} \in R^d$. Assume there exist positive constants γ, c and β such that*

$$E(|X^{(k)}(x) - X^{(k)}(y)|^\gamma) \leq c|x - y|^{d+\beta}, \tag{4}$$

for all $x, y \in \overline{D}$ and for all k. Then $X^{(k)}$ have continuous versions and they are tight in $[C(\overline{D})]^n$.

REMARK. In the Gaussian case there are much weaker conditions for continuity. In fact d is not needed in the exponent [see Dudley (1973), Theorem 2.10]. However, tightness requires control of the random constant multiplying the modulus of continuity.

For Gaussian fields we have an easy way of estimating the right-hand side of (4), since for any positive integer l,

$$\begin{aligned} E|X(x) - X(y)|^{2l} &\leq n^l \sum_{j=1}^n E\left(X_j(x) - X_j(y)\right)^{2l} \\ &= \frac{n^l(2l)!}{2^l l!} \sum_{j=1}^n \left[\operatorname{var}\left(X_j(x) - X_j(y)\right)\right]^l. \end{aligned}$$

Now set

$$G_{ij}(x, y) = E\left(X_i(x) X_j(y)\right) = \delta_y^j(J\delta_x^i).$$

Then we get

$$\begin{aligned} &E|X(x) - X(y)|^{2l} \\ &\quad\leq \frac{n^l(2l)!}{2^l l!} \sum_{j=1}^n \left[G_{jj}(x,x) - 2G_{jj}(x,y) + G_{jj}(y,y)\right]^l \\ &\quad\leq \frac{n^l(2l)!}{2l!} \sum_{j=1}^n \Big[\left(G_{jj}(x,x) - G_{jj}(x,y)\right)^l \\ &\qquad\qquad + \left(G_{jj}(y,y) - G_{jj}(y,x)\right)^l\Big]. \end{aligned} \tag{5}$$

Now since $G_j(x,\cdot) = J(\delta_x^j) \in V$, it is Hölder continuous and its Hölder norm is bounded by $\kappa\|J(\delta_x^j)\|_m$. Moreover, by (3) $\{\delta_x, x \in \overline{D}\}$ is contained in a ball of radius κ in V^* so that the Hölder norm of $G_j(x,\cdot)$ is uniformly bounded in x by $\kappa^2\|J\|$. Therefore, by taking l large enough we get (4). We have therefore proved the following.

THEOREM 1. *Let $B(\cdot,\cdot)$ be a Hilbertian norm equivalent to the Sobolev norm in a subspace $[H_0^m(D)]^n \subset V \subset [H^m(D)]^n$ with $m \geq m^* = [d/2+1]$. Let J be defined by*

$$B(J\eta, v) = \eta(v),$$

for $\eta \in V^*$ and $v \in V$. Let $G_{ij}(x,y) = \delta_x^j(J\delta_x^i)$ for $1 \leq i, j \leq n$. Then there exists a Gaussian random field $\{X(x): x \in \overline{D}\}$ with $E(X_i(x), X_j(y)) = G_{ij}(x,y)$ and with $C(\overline{D})$ sample configurations.

Because of the preceding theorem we may call X the random field with energy B on V. Its covariance matrix $G(x,y)$ is nothing but the Green's tensor of the boundary value problem $B(u,v) = \eta(v)$, $v \in V$.

3. The finite-dimensional approximations. We now introduce the finite-dimensional approximations we are interested in. For a general reference on this type of approximation in numerical analysis of partial differential equations, see Raviart and Thomas (1983). For any positive integer h, let V_h be a finite-dimensional subspace of V and let $B_h(\cdot,\cdot)$ be the restriction of $B(\cdot,\cdot)$ to $V_h \times V_h$. Since any $\eta \in V^*$ restricted to V_h is bounded, again by the Riesz representation theorem there exists a unique $J_h\eta \in V_h$ such that $B(J_h\eta, v_h) = \eta(v_h)$ for all $v_h \in V_h$.

Next let us define the random field $\{X^{(h)}(x), x \in \overline{D}\}$ with covariance matrix

$$E\left(X_i^{(h)}(x)X_j^{(h)}(y)\right) = G_{ij}^{(h)}(x,y) = \delta_y^j\left(J_h\delta_x^i\right),$$

for $1 \leq i, j \leq n$. Such a field exists by the general theorem on the existence of Gaussian random fields with given covariance. In fact let $\{\varphi_k^{(h)}\}$, $k = 1,\dots,d_h$, be an orthonormal basis of V_h with respect to $B(\cdot,\cdot)$, and set I_h to be the mapping from R^{d_h} into V_h defined by $I_h(u) = \sum_{k=1}^{d_h} u_k\varphi_k^{(h)}$. Then we can write

$$X^{(h)}(x) = \left[I_h(\xi^{(h)})\right](x), \tag{6}$$

where $\xi^{(h)} = (\xi_1^{(h)},\dots,\xi_{d_h}^{(h)})$ is a mean zero Gaussian random vector with identity covariance matrix.

THEOREM 2. *Let the random field X be defined as in Theorem 1 and the fields $X^{(h)}$ as in (6). ($X^{(h)}$ and X are not necessarily defined on the same probability space.) Suppose $\bigcup_{h=1}^{\infty} V_h$ is dense in V. Then $X, X^{(h)}$ all have continuous sample paths and the distribution of $X^{(h)}$ converges weakly to the distribution of X in $[C(\overline{D})]^n$, as $h \to \infty$.*

PROOF. From the equality $B(J_h\eta - J\eta, v_h) = 0$, for all $v_h \in V_h$ and $\eta \in V^*$, we see that $J_h\eta$ is nothing but the projection of $J\eta$ onto V_h. Thus J_h converges strongly to J which implies that the finite-dimensional distributions of $X^{(h)}$ converge to those of X. Moreover the norms $\|J_h\|$ are uniformly bounded by some positive constant C. Consequently

$$\begin{aligned}
&\sup_{1\le j\le n} \sum_{x,y\in\overline{D}} \frac{|G_{jj}^{(h)}(x,x) - G_{jj}^{(h)}(x,y)|}{|x-y|^\alpha} \\
&\quad \le \sup_{1\le j\le n} \sup_{x,y,z\in\overline{D}} \frac{|G_{jj}^{(h)}(x,z) - G_{jj}^{(h)}(x,y)|}{|z-y|^\alpha} \\
&\quad \le \kappa \sup_{x\in\overline{D}} \|J_h(\delta_x^j)\|_m \le \kappa C \sup_{x\in\overline{D}} \|\delta_x^j\|_{V^*} \le \kappa^2 C.
\end{aligned}$$

From Proposition 2, the inequality (5) and the preceding estimate it follows that the $X^{(h)}$ are tight in $[C(\overline{D})]^n$. □

4. The observation model. Let us now specify the statistical properties of the observed variables $Y_1, \dots, Y_q$. The weak convergence we have established will enable us to approximate the conditional distribution of the signal field given the observation.

Assume that there exists a jointly continuous function $Q(y; f)$ such that

$$P(Y \in A | X = f) = \int_A e^{-Q(y;f)} \rho(dy),$$

where ρ is a σ-finite Borel measure on R^q. In addition, we assume that for any value of y, $Q(y; \cdot)$ is bounded from above and below and has two Frechet derivatives uniformly bounded in norm on $[C(\overline{D})]^n$. The simplest example is when $Y_i = g_i(X) + n_i$, $i = 1, \dots, q$, where g_i $i = 1, \dots, q$, are continuous functionals on $[C(\overline{D})]^n$ with bounded continuous first and second derivatives, and n_i, $i = 1, \dots, q$, are independent standard Gaussian random variables.

Let X and $X^{(h)}$ be as in Theorem 2. Let P_X and $P_{X^{(h)}}$ be their respective probability distributions and set $\Lambda(y; f) = \exp(-Q(y; f))$. Let

$$P_X^y(df) = \frac{\Lambda(y; f) P_X(df)}{\int \Lambda(y; f) P_X(df)} \quad \text{and} \quad P_{X^{(h)}}^y(df) = \frac{\Lambda(y; f) P_{X^{(h)}}(df)}{\int \Lambda(y; f) P_{X^{(h)}}(df)}.$$

Then P_X^y is the regular conditional probability distribution of X given y. Observe that P_X^y remains unchanged if ρ is replaced by any measure $\tilde{\rho}$ such that $\int e^{-Q(y;f)}\tilde{\rho}(dy) = 1$ for any f. Let μ_h denote the Gaussian measure on R^{d_h} with mean zero and identity covariance and define

$$\mu_h^y(d\xi^{(h)}) = \frac{\exp(-Q(y; I_h(\xi^{(h)})\varphi^{(h)}))}{Z_y^{(h)}} \mu_h(d\xi^{(h)}),$$

where $Z_y^{(h)}$ is the appropriate normalizing constant. Then $P_{X^{(h)}} = \mu_h \circ I_h^{-1}$ and $P_{X^{(h)}}^y = \mu_h^y \circ I_h^{-1}$. Since $\Lambda(y; f)$ is bounded continuous and bounded away from zero, the following corollary to Theorem 2 is easily established.

COROLLARY 1. *For any* $y \in R^q$, *the sequence of measures* $P^y_{X^{(h)}}$ *converges weakly to* P^y_X *in* $[C(\overline{D})]^n$.

In the sequel y is considered fixed and we omit the dependence of Q and $Z^{(h)}$ on y. Moreover, since μ^y_h is not changed by adding to Q a constant which depends only on y, we can assume without loss of generality that Q is nonnegative.

5. The finite-dimensional Markov processes. In this section for each fixed h a Markov process of Langevin type whose invariant distribution is $P^y_{X^{(h)}}$ is introduced.

Let $W_{t,h}$, $h = 1, \dots,$ be a sequence of independent scalar standard Brownian motions. Denote by $W^{(h)}_t$ the vector $(W_{t,1}, \dots, W_{t,d_h})$ and set $Q^{(h)}(\xi^{(h)}) = Q(\xi^{(h)} \cdot \varphi^{(h)})$. Consider the Langevin equation

$$dx^{(h)}_t(z) = \left(-M^{(h)}x^{(h)}_t(z) - \nabla Q^{(h)}\left(x^{(h)}_t(z)\right)\right)dt + \sqrt{2}\,dW^{(h)}_t. \tag{7}$$

Under our hypotheses this equation has a unique solution for any initial point z and defines a Feller process whose unique invariant measure is μ^y_h. Its generator is the closure of

$$L^{(h)}\psi(x) = \Delta\psi(x) - \left(M^{(h)}x + \nabla Q^{(h)}(x)\right) \cdot \nabla\psi(x), \qquad \psi \in C^\infty_0(R^{d_h}) \tag{8}$$

[see Ethier and Kurtz (1986)].

In order to estimate the rate of convergence of the transition probabilities to the invariant distribution we consider the semigroup generated by $L^{(h)}$ on the Hilbert space $L_2(\mu^y_h)$. Through integration by parts we get

$$\left(L^{(h)}\psi, \phi\right)_{L_2(\mu^y_h)} = \left(L^{(h)}\psi, \phi\right)_{L_2(\mu^y_h)} = -\int \nabla\psi \cdot \nabla\phi \, d\mu^y_h = -\mathscr{E}^{(h)}(\psi, \phi), \tag{9}$$

for $\psi, \phi \in C^\infty(R^{d_h})$. $L^{(h)}$ is a symmetric dissipative operator and has a self-adjoint closure on $L_2(\mu^y_h)$ which generates a C_0-semigroup $\{T^{(h)}_t\}$ of self-adjoint contractions, such that $T^{(h)}_t f(z) = Ef(x^{(h)}_t(z))$. It is clear that 0 is an eigenvalue of $L^{(h)}$ and that $\mathbf{1}$ is its eigenvector. Let $P_h\psi = \int \psi(x)\mu^y_h(dx)$. Then we have

$$\left|\left(T^{(h)}_t - P_h\right)\psi\right|_{L_2(\mu^y_h)} \le e^{-\lambda(L^{(h)})t}|\psi - P_h\psi|_{L_2(\mu^y_h)} \le e^{-\lambda(L^{(h)})t}|\psi|_{L_2(\mu^y_h)},$$

where

$$\lambda(L^{(h)}) = \inf\left\{\mathscr{E}^{(h)}(\psi, \psi); P_h\psi = 0, |\psi|_{L_2(\mu^y_h)} = 1, \psi \in C^\infty_b(R^{d_h})\right\}. \tag{10}$$

We will obtain a lower bound for $\lambda(L^{(h)})$ which is independent of h.

Observe that if $Q(f) = 0$, then $\lambda(L^{(h)})$ becomes the Ornstein–Uhlenbeck (OU) operator in R^{d_h} for which the spectral gap is 1 [see Reed and Simon (1972), page 142]. Now we are ready to prove the following lemma.

LEMMA 1. *Let* $\mathscr{Q} = \max\{Q(f), f \in [C(\overline{D})]^n\}$. *Then* $\lambda(L^{(h)}) \ge \lambda \equiv e^{-\mathscr{Q}}$.

PROOF. For any function $\phi \in C_b^\infty(R^{d_h})$ such that $\int \phi(x)\mu_h^y(dx) = 0$, let $\bar{\phi} = \int \phi(x)\mu_h(dx)$. Observe that

$$\begin{aligned}\int (\phi(x))^2 \mu_h^y(dx) &= \frac{1}{Z^{(h)}} \int (\phi(x))^2 e^{-Q^{(h)}(x)} \mu_h(dx)\\ &\le \frac{1}{Z^{(h)}} \int \left(\phi(x) - \bar{\phi}\right)^2 e^{-Q^{(h)}(x)} \mu_h(dx)\\ &\le \frac{1}{Z^{(h)}} \int \left(\phi(x) - \phi\right)^2 \mu_h(dx)\\ &\le \frac{1}{Z^{(h)}} \int |\nabla \phi(x)|^2 \mu_h(dx)\\ &\le e^{\mathscr{D}} \int |\nabla \phi(x)|^2 \mu_h^y(dx).\end{aligned}$$

The second inequality follows from the fact that $Q^{(h)}$ is positive. The third follows from the fact that the spectral gap of the OU operator is 1. □

In the preceding discussion it is not essential that the Markov process be a diffusion. For example, we could replace the generator (8) by

$$L^{(h)}\psi(x) = Z^{(h)} \int [\psi(z) - \psi(x)] k_h(x, z) \mu_h^y(dz), \tag{11}$$

where $0 < k_{\min} \le k_h(x, z) \le e^{Q^{(h)}(z)}$ is a symmetric continuous function on $R^{d_h} \times R^{d_h}$. In this case

$$\begin{aligned}\mathscr{E}^{(h)}(\psi, \psi) &= \frac{1}{2} Z^{(h)} \int \int [\psi(z) - \psi(x)]^2 k_h(x, z) \mu_h^y(dz) \mu_h^y(dx)\\ &\ge \frac{k_{\min}}{e^{\mathscr{D}}} \left(|\psi|^2_{L_2(\mu_h^y)} - (P_h \psi)^2\right),\end{aligned}$$

from which we get that $\lambda(L^{(h)}) \ge k_{\min}/e^{\mathscr{D}}$. The operator (11) is the generator of a jump process with transition probability $q_h(x, dz) = k_h(x, z)\mu_h^y(dz)/K_h(x)$ and jump intensities $\lambda(x) = Z^{(h)}K_h(x)$. Such a process can be simulated by picking a sample z from μ_h at i.i.d. mean 1 exponential times and accepting it as a new state with probability $k_h(x, z)e^{Q^{(h)}(z)}$, x being the current state. In particular the choice

$$k_h(x, z) = \exp\left(Q^{(h)}(x) \wedge Q^{(h)}(z)\right)$$

resembles the Metropolis algorithm [see Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953)] since the probability of acceptance is then $\exp(-[Q^{(h)}(z) - Q^{(h)}(x)]^+)$.

6. The multilevel Markov process. In this section we are going to present our main result, which is Theorem 4. In order to do so, more stringent assumptions on the finite-dimensional approximations are needed. We will consider only Galerkin approximations, assuming that the subspaces V_h are increasing. To keep notation simple, let us also assume $d_h = h$. Then it is possible to choose a sequence $\varphi_h \in V$, orthonormal with respect to the scalar product B such that $V_h = \operatorname{span}\{\varphi_1, \dots, \varphi_h\}$.

Now we define a nonhomogeneous Markov process x_t with state space $\bigcup_{h=1}^{\infty} R^h$. Let c be a positive constant to be determined below and define a sequence of times $t_h = c(h-1)$, $h = 1, \dots$. For any time t let $h_t = [t/c] + 1$. Now define

$$x_t = x_t^{(1)}(\omega_0), \qquad 0 = t_1 \le t < t_2,$$

$$x_t = x_{t-t_h}^{(h)}(x_{t_h -}, \omega_h), \qquad t_h \le t < t_{h+1},\ h = 1, 2, \dots,$$

where ω_h, $h = 0, 1, \dots$, are i.i.d. standard Gaussian variables. The process $\{x_t\}$ adds a component at each of the times t_h. The process $I(x_t) = I_{h_t}(x_t)$ is a process in $[C(\overline{D})]^n$. Let $\pi_{h-1,h}$ be the operator from $L_2(R^h, \mu_h^y)$ to $L_2(R^{h-1}, \mu_{h-1}^y)$ defined as

$$\pi_{h-1,h} f(\xi_1, \dots, \xi_{h-1}) = \int f(\xi_1, \dots, \xi_{h-1}, \xi_h) \mu_1(d\xi_h).$$

For $s < t$ define the transition operator

$$U(t,s) f(z) = E(f(x_t) | x_s = z),$$

f being a bounded and measurable function on R^{h_t}. Setting $k = h_s$ and $h = h_t$ it is easily seen that

$$U(t,s) = T_{t_{k+1}-s}^{(k)} \pi_{k,k+1} T_{t_{k+2}-t_{k+1}}^{(k+1)} \cdots \pi_{h-1,h} T_{t-t_h}^{(h)}. \tag{12}$$

Set $\varepsilon_j = |\psi_j|_{[C(\overline{D})]^n}$, for any $j = 1, \dots$. Note that since V is compactly embedded in $[C(\overline{D})]^n$, the sequence ε_j converges to zero as $j \to \infty$.

THEOREM 3. *If $c > \mathscr{Q}/\lambda$, then*:

(a) *The distribution of $I(x_t)$ converges weakly to P_X^y as $t \to \infty$.*

(b) *For any bounded measurable functional F whose set of discontinuities is of P_X^y probability zero on $[C(\overline{D})]^n$, the process*

$$L_t(F) = \frac{1}{t} \int_0^t \big(F(I(x_u)) - P_X^y(F) \big)\, du$$

converges in the mean square to zero.

PROOF. (a) Let s, t, k, h be as in (12). Let $T_j = T_{t_{j+1}-t_j}^{(j)}$ for $j = k+1, \dots, h-1$, $T_k = T_{t_{k+1}-s}^{(k)}$ and $T_h = T_{t-t_h}^{(h)}$. Since $P_j T_j = P_j$ and since $T_k \pi_{k,k+1} \cdots$

$\pi_{j-1,j}P_j = P_j$, we write

$$
\begin{aligned}
U(t,s) - P_{h_t} &= T_k \pi_{k,k+1} T_{k+1} \pi_{k+1,k+2} \cdots \pi_{h-1,h} T_h - P_h \\
&= (T_k - P_k)\pi_{k,k+1}(T_{k+1} - P_{k+1}) \cdots \pi_{h-1,h}(T_h - P_h) \\
&\quad + \sum_{j=k}^{h-1} P_j \pi_{j,j+1}(T_{j+1} - P_{j+1}) \\
&\quad\quad \times \pi_{j+1,j+2}(T_{j+2} - P_{j+2}) \cdots \pi_{h-1,h}(T_h - P_h) \\
&= (T_k - P_k)\pi_{k,k+1}(T_{k+1} - P_{k+1}) \cdots \pi_{h-1,h}(T_h - P_h) \\
&\quad + \sum_{j=k}^{h-1} (P_j \pi_{j,j+1} - P_{j+1})(T_{j+1} - P_{j+1}) \\
&\quad\quad \times \pi_{j+1,j+2} \cdots \pi_{h-1,h}(T_h - P_h),
\end{aligned}
$$

from which we obtain

$$
\begin{aligned}
\|U(t,s) - P_{h_t}\| &\le \|T_k - P_k\| \, \|\pi_{k,k+1}\| \, \|T_{k+1} - P_{k+1}\| \cdots \|\pi_{h-1,h}\| \, \|T_h - P_h\| \\
&\quad + \sum_{j=k}^{h-1} \|P_j \pi_{j,j+1} - P_{j+1}\| \, \|T_{j+1} - P_{j+1}\| \\
&\quad\quad \times \|\pi_{j+1,j+2}\| \cdots \|\pi_{h-1,h}\| \, \|T_h - P_h\|.
\end{aligned}
\tag{13}
$$

Let us estimate the various norms. Let $f \in L_2(R^j, \mu_j^y)$. Then

$$
\begin{aligned}
|\pi_{j-1,j} f|^2_{L_2(R^{j-1}, \mu_{j-1}^y)} &\le \frac{1}{Z^{(j-1)}} \int\int \left| f(\xi^{(j-1)}, \xi_j) \right|^2 \mu_1(d\xi_j) \\
&\quad \times \exp(-Q^{(j-1)}(\xi^{(j-1)})) \mu_{j-1}(d\xi^{(j-1)}) \\
&\le e^{\mathscr{Q}} \int \left| f(\xi^{(j)}) \right|^2 \mu_j(d\xi^{(j)}) \\
&\le \frac{e^{2\mathscr{Q}}}{Z^{(j)}} \int \left| f(\xi^{(j)}) \right|^2 \exp(-Q^{(j)}(\xi^{(j)})) \mu_j(d\xi^{(j)}) \\
&= e^{2\mathscr{Q}} |f|^2_{L_2(R^j, \mu_j^y)}.
\end{aligned}
$$

Thus $\|\pi_{j-1,j}\| \le e^{\mathscr{Q}}$ for any j. Moreover by Lemma 1, $\|T_j - P_j\| \le e^{-\lambda(t_{j+1} - t_j)}$, for $j = k+1, \dots, h-1$, $\|T_h - P_h\| \le e^{-\lambda(t - t_h)}$ and $\|T_k - P_k\| \le e^{-\lambda(t_{k+1} - s)}$. Finally we want to estimate $\|P_j\pi_{j,j+1} - P_{j+1}\|$. This expression can be bounded as follows. Let $f \in L_2(\mu_{j+1}^y)$ such that $|f|_{L_2(\mu_{j+1}^y)} = 1$. Then

$$\begin{aligned}
&\left|(P_j\pi_{j,j+1} - P_{j+1}) f\right| \\
&\quad\le \frac{1}{Z^{(j)}}\left|\int f(\xi^{(j+1)})\big(\exp(-Q^{(j)}(\xi^{(j)}))\right. \\
&\qquad\qquad\qquad \left. -\exp(-Q^{(j+1)}(\xi^{(j+1)}))\big)\mu_{j+1}(d\xi^{(j+1)})\right| \\
&\qquad + \left|\frac{1}{Z^{(j)}} - \frac{1}{Z^{(j+1)}}\right| |P_{j+1}f| \\
&\quad\le 2e^{2\mathscr{D}}\bigg(\int\Big|\big(\exp(-Q^{(j+1)}(\xi^{(j)},0)) \\
&\qquad\qquad -\exp(-Q^{(j+1)}(\xi^{(j+1)}))\big)\Big|^2\mu_{j+1}(d\xi^{(j+1)})\bigg)^{1/2} \\
&\quad\le 2e^{2\mathscr{D}}\left(\int\left|\big(Q^{(j+1)}(\xi^{(j)},0) - Q^{(j+1)}(\xi^{(j+1)})\big)\right|^2\mu_{j+1}(d\xi^{(j+1)})\right)^{1/2} \\
&\quad\le 2e^{2\mathscr{D}}\tilde{\mathscr{D}}\left(\int \xi_{j+1}^2|\varphi_{j+1}|^2_{[C(\overline{D})]^n}\mu_{j+1}(d\xi^{(j+1)})\right)^{1/2} \\
&\quad\le 2e^{2\mathscr{D}}\tilde{\mathscr{D}}\varepsilon_{j+1},
\end{aligned}$$

where $\tilde{\mathscr{D}}$ denotes the uniform bound on the norm of the derivative of Q.

Tying these estimates together we get

$$\begin{aligned}
\|U(t,s) - P_{h_t}\| &\le (e^{\mathscr{D}})^{h_t-h_s}e^{-\lambda(t-s)} \\
&\qquad + 2e^{2\mathscr{D}}\tilde{\mathscr{D}}\sum_{j=h_s+1}^{h_t}\varepsilon_j(e^{\mathscr{D}})^{h_t-j}\exp(-\lambda(t-t_j)) \\
&\le e^{\mathscr{D}}\exp((-\delta(h_t-h_s)+\lambda c)) \\
&\qquad + 2\exp((2\mathscr{D}+\lambda c))\tilde{\mathscr{D}}\sum_{j=h_s+1}^{h_t}\varepsilon_j\exp(-\delta(h_t-j)) \qquad (14)\\
&= e^{\mathscr{D}}\exp((-\delta(h_t-h_s)+\lambda c)) \\
&\qquad + 2\exp(2\mathscr{D}+\lambda c)\tilde{\mathscr{D}}\sum_{j=1}^{h_t-h_s-1}\varepsilon_{h_t-h_s-j}e^{-\delta j},
\end{aligned}$$

where $\delta = \lambda c - \mathscr{D} > 0$. Since ε_j converges to zero and the series $e^{-\delta j}$ is summable, the above sum converges to zero as $t - s \to \infty$. Therefore

$$\|U(t,s) - P_{h_t}\| \le \Delta(t-s), \qquad 0 \le s \le t < \infty, \tag{15}$$

where $\Delta(\tau) \to 0$ as $\tau \to \infty$.

Finally let $F^{(h)}(\xi^{(h)}) = (F \circ I_h)(\xi^{(h)})$ where F is a bounded measurable functional whose set of discontinuities is of P_X^y probability zero on $[C(\overline{D})]^n$. Setting $s = 0$ in (15) we get

$$\int \left|E\left(F^{(h_t)}(x_t)|x_0 = z\right) - P_{h_t}F^{(h_t)}\right|^2 \mu_1(dz) \to 0,$$

as $t \to \infty$. Since $P_{X^{(h)}}^y F = P_h F^{(h)}$ and $P_{X^{(h)}}^y$ converges weakly to P_X^y we have proved that

$$\lim_{t\to\infty} EF^{(h_t)}(x_t) = P_X^y(F).$$

(b) To prove the last statement let $|F|_\infty = K$ and write

$$(16)\qquad \begin{aligned} E\left(L_t^2(F)\right) \le 2E&\left(\frac{1}{t}\int_0^t \left((F\circ I)(x_u) - P_{h_u}F^{(h_u)}\right) du\right)^2 \\ &+ 2\left(\frac{1}{t}\int_0^t \left(P_{h_u}F^{(h_u)} - P_X^y(F)\right) du\right)^2. \end{aligned}$$

The second term in (16) goes to zero as $t \to \infty$ since the integrand goes to zero as $u \to \infty$. As for the first integral we write

$$\begin{aligned} E&\left(\frac{1}{t}\int_0^t \left((F\circ I)(x_u) - P_{h_u}F^{(h_u)}\right) du\right)^2 \\ &= \frac{2}{t^2}\int_0^t\int_u^t E\Big\{\left((F\circ I)(x_u) - P_{h_u}F^{(h_u)}\right) \\ &\qquad\qquad \times\left((F\circ I)(x_s) - P_{h_s}F^{(h_s)}\right)\Big\}\, ds\, du \\ (17)\qquad &= \frac{2}{t^2}\int_0^t\int_u^t E\Big\{\left((F\circ I)(x_u) - P_{h_u}F^{(h_u)}\right) \\ &\qquad\qquad \times E\left((F\circ I)(x_s) - P_{h_s}F^{(h_s)}\middle|x_u\right)\Big\}\, ds\, du \\ &= \frac{2}{t^2}\int_0^t\int_u^t \int U(u,0)\left(\left(F^{(h_u)} - P_{h_u}F^{(h_u)}\right)G_s^{(h_u)}\right)(x_0)\mu_1(dx_0)\, ds\, du \\ &\le \frac{2e^Q}{t^2}\int_0^t\int_u^t P_1 U(u,0)\left(\left(F^{(h_u)} - P_{h_u}F^{(h_u)}\right)G_s^{(h_u)}\right) ds\, du, \end{aligned}$$

where

$$\begin{aligned} G_s^{(h_u)}(\xi) &= E\left((F\circ I)(x_s) - P_{h_s}F^{(h_s)}\middle|x_u = \xi\right) \\ &= U(s,u)\left(F^{(h_s)} - P_{h_s}F^{(h_s)}\right)(\xi). \end{aligned}$$

Now write the last term in (17) as

$$\begin{aligned}(18)\qquad &\frac{1}{t^2}\int_0^t\int_u^t P_1\big(U(u,0)-P_{h_u}\big)\Big(\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)G_s^{(h_u)}\Big)\,ds\,du\\ &+\frac{1}{t^2}\int_0^t\int_u^t P_{h_u}\Big(\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)G_s^{(h_u)}\Big)\,ds\,du,\end{aligned}$$

and estimate the integrands:

$$\begin{aligned}&P_1\big(U(u,0)-P_{h_u}\big)\Big(\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)G_s^{(h_u)}\Big)\\ &\quad\le \|U(u,0)-P_{h_u}\|\,\Big|\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)G_s^{(h_u)}\Big|_\infty\\ &\quad\le 4K^2\|U(u,0)-P_{h_u}\|.\end{aligned}$$

By (15) this expression goes to zero as $u\to\infty$; hence the first integral in (18) converges to zero as $t\to\infty$. In addition

$$\begin{aligned}&\Big|P_{h_u}\Big(\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)G_s^{(h_u)}\Big)\Big|\\ &\quad=\Big|P_{h_u}\Big(\big(F^{(h_u)}-P_{h_u}F^{(h_u)}\big)\big(U(s,u)-P_{h_s}\big)F^{(h_s)}\Big)\Big|\\ &\quad\le 2K^2\|U(s,u)-P_{h_s}\|.\end{aligned}$$

Again by (15), the second integral in (18) will be bounded by

$$\begin{aligned}\frac{1}{t^2}\int_0^t\int_u^t\Delta(s-u)\,ds\,du&=\frac{1}{t^2}\int_0^t\int_0^{t-u}\Delta(v)\,dv\,du\\ &\le\frac{1}{t^2}\int_0^t\int_0^t\Delta(v)\,dv\,du=\frac{1}{t}\int_0^t\Delta(v)\,dv,\end{aligned}$$

so that it goes to zero as $t\to\infty$, since $\Delta(v)\to 0$ as $v\to\infty$. Consequently (16) goes to zero as $t\to\infty$. □

REMARK. To illustrate a simple application of the preceding result let us consider a concrete example arising in Grenander (1970). Fix a point $z\in D\subset R^d$ and let S be a measurable set in R^n with boundary of Lebesgue measure zero. We approximate $P_X^y(X(z)\in S)$ with $(1/t)\int_0^t 1_S(x_u\cdot\varphi(z))\,du$. The function $f\to 1_S(f(z))$ is bounded and measurable on $[C(\overline{D})]^n$ and its set of discontinuities is $A=\{f\colon f(z)\in\partial S\}$. Since the conditional distribution of $X(z)$ given y is absolutely continuous with respect to a Gaussian measure it does not change the boundary of S so that $P_X^y(A)=0$. Therefore the preceding approximation is convergent.

Acknowledgments. We would like to thank Ulf Grenander for having suggested the problem. We would also like to thank him and Basilis Gidas for many useful conversations.

REFERENCES

ADAMS, R. A. (1975). *Sobolev Spaces*. Academic, Orlando.

AGMON, S., DOUGLIS, A. and NIRENBERG, L. (1964). Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions. II. *Comm. Pure Appl. Math.* **17** 35–92.

AMIT, Y., GRENANDER, U. and PICCIONI, M. (1990). Structural image restoration through deformable templates. Center for Intelligent Control Systems, M.I.T.

BORKAR, V., CHARI, R. T. and MITTER, S. K. (1988). Stochastic quantization of field theory in finite and infinite volume. *J. Funct. Anal.* **81** 184–206.

DEMBO, A. and ZEITOUNI, O. (1990). Maximum a posteriori estimation of elliptic Gaussian fields observed via a non-linear channel. *J. Multivariate Anal.* To appear.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* **1** 66–103.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

GEMAN, S and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intelligence* **6** 721–741.

GIDAS, B. (1985). Non-stationary Markov chains and convergence of the annealing algorithm. *J. Statist. Phys.* **39** 73–131.

GRENANDER, U. (1970). A unified approach to pattern analysis. *Adv. Comput.* **10** 175–216.

HOLLEY, R. and STROOCK, D. (1988). Simulated annealing via Sobolev inequalities. *Commun. Math. Phys.* **115** 553–569.

ITÔ, K. (1984). *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces.* SIAM, Philadelphia.

JONA-LASINIO, G. and MITTER, P. K. (1983). On the stochastic quantization of field theory. *Commun. Math. Phys.* **101** 409–436.

KARATZAS, I. and SHREVE, S. E. (1988). *Brownian Motion and Stochastic Calculus.* Springer, New York.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1091.

PICCIONI, M. (1987). A generalization of Wahba's theorem on the equivalence between spline smoothing and Bayesian estimation. *Stochastic Partial Differential Equations. Applications. Lecture Notes in Math.* **1236**. Springer, New York.

PICCIONI, M. and ROMA, M. (1990). Stochastic regularization of linear equations and the realization of Gaussian fields. *J. Multivariate Anal.* **33** 143–150.

RAVIART, P. A. and THOMAS, J. M. (1983). *Introduction a l'Analyse Numerique des Equations aux Derives Partielles.* Masson, Paris.

REED, M. and SIMON. B. (1972). *Methods of Modern Mathematical Physics* **1** 142. Academic, New York.

DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02904

DIPARTIMENTO DI MATEMATICA
E CENTRO VITO VOLTERRA
UNIVERSITÀ DI ROMA II
00173 ROMA
ITALIA