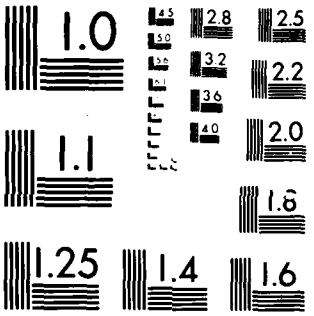
AD-A164 054 ESTIMATION THEORY AND STATISTICAL PHYSICS(U) 1/1
MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR
INFORMATION AND DECISION SYSTEMS S K MITTER DEC 85
UNCLASSIFIED LIDS-P-1518 ARO-20980.38-MA F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

2

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: ARO 20980.38-MA	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): ESTIMATION THEORY AND STATISTICAL PHYSICS	5. TYPE OF REPORT & PERIOD COVERED: Technical Paper
	6. PERFORMING ORG. REPORT NUMBER: LIDS-P-1518
7. AUTHOR(s): S.K. Mitter	8. CONTRACT OR GRANT NUMBER(s): U.S. Army Research Office Grant DAAG29-84-K-0005
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709	12. REPORT DATE: December 1985 13. NUMBER OF PAGES: 20
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/ DOWNGRADING SCHEDULE: NA

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ARO 20980.38-MA

December, 1985 LIDS-P-1518

ESTIMATION THEORY AND STATISTICAL PHYSICS

by

S.K. Mitter

Research supported by the U.S. Army Research Office under Grant DAAG29-84-K-0005 and by the Air Force Office of Scientific Research under grant AFOSR-85-0227.

This paper is a detailed version of an invited lecture given at the conference on Stochastic Processes and Applications held at Nagoya, Japan in July 1985.

ESTIMATION THEORY AND STATISTICAL PHYSICS

S.K. Mitter
Department of Electrical Engineering and Computer Science and
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

1. INTRODUCTION

In my previous work on non-linear filtering for diffusion processes [Mitter 1980,1983] I have discussed the analogies that exist between these problems and problems of quantum physics from the Feynman point of view [cf. Glimm-Jaffe 1981]. The basic idea here is that construction of a non-linear filter involves an integration over function space which is exactly analogous to the construction of a measure on path-space via the Feynman-Kac-Nelson Formula. ~~According to this~~ viewpoint, the Kalman-Bucy filtering problem, ~~namely~~ the filtering of Gauss-Markov processes in the presence of additive white Gaussian noise occupies the same role as the Ornstein-Uhlenbeck process (finite or infinite-dimensional) in Quantum Mechanics or Quantum Field Theory. That this analogy ~~may be deeper~~ is borne out by the fact that a solvable Lie algebra, the oscillator algebra which contains the Heisenberg algebra as a derived algebra is intrinsically attached to the Kalman-Bucy Filtering problem. I have also shown that the problem of non-linear filtering of diffusion processes admits a stochastic variational interpretation ~~[Fleming-Mitter 1982]~~. The objective of this paper is to strengthen these analogies further with a view to showing the close relationship of estimation theory to statistical mechanics. The motivation for this comes from problems of estimation and inverse problems related to image processing.

In order to carry out this program it is necessary to generalize these ideas to filtering problems for infinite-dimensional processes where we are forced to work in the context of generalized random fields. There are two types of processes involved: continnous processes such as infinite-dimensional Ornstein-Uhlenbeck processes and their L2-functionals which represent intensities of images and processes of a "discrete"-nature which will represent "boundaries" of images. The most interesting models are obtained when these processes are coupled according to a probabilistic mechanisms. The "discrete" processes should be thought of as gauge fields and will be a process on connection forms.

Although the estimation problems of interest are naturally viewed in the context of random fields which are independent of time, the best way to obtain sample functions for these fields is to simulate it via finite or infinite-dimensional stochastic differential equations whose invariant distribution coincides with the

distribution of the time-independent random field. This Monte-Carlo simulation procedure is the same idea as stochastic quantization, an idea advanced by Parisi (cf. Parisi-Wu 1981) and recently studied in a rigorous manner for $(P\phi)_2$ fields in a finite volume by Jona-Lasinio and Mitter, 1984. The problem which are of interest here are filtering problems associated with these stochastic fields obtained by introducing observations which are "local" and studying the behaviour of these filters as $t \to \infty$. To make any progress however, one would have to work with lattice approximations of these stochastic fields and reduce the filtering problems to a finite dimensional situation and even here there are severe technical difficulties. When the observations are however local and considered to be on the stationary Gibbs field the problems amounts to looking at the invariant distribution of a stochastic differential equation with a coupling coming from the observations.

The main objective of this paper is to explore these relationships between problems of estimation and stochastic quantization (stochastic mechanics) at a conceptual level. The detailed technical discussion will appear elsewhere.

2. SIGNAL MODELS FOR IMAGE PROCESSING

In order to treat problems of Image Processing in a probabilistic framework we need probabilistic models for the signals in question. These signals are various attributes of images such as intensity of images and boundaries between smooth parts of images. The probabilistic models we choose are Gibbsian random fields and are borrowed from statistical mechanics. These models for image processing have been used by several authors recently (c.f. Geman and Geman 1984, Grenander 1984, Marroquin 1985). The exposition of signals models given below follows Sinai 1982. The signal models we wish to consider correspond to statistical mechanical models on a finite lattice and we shall take this lattice to be Z^2 with the Euclidean norm. The sample space Ω consists of functions $\phi: Z^2 \to \bar{\Phi}: x=(x_1,x_2) \to \phi(x)$, where $\bar{\Phi}$ is a finite set, a homogeneous space of a compact Lie Group with the natural σ-algebra of Borel Sets or R^1. The sample space Ω is termed a configuration space in statistical physics. For $V \subset Z^2$, a finite subset, we denote by $\phi(V) = \{\phi(x) \,|\, x \in V\}$ and $V \subset Z^2 = \{\phi(V) \,|\, V \subset Z^2\}$, finite.

For a non-empty finite subset $V \subset Z^2$ we are given a function $I: \Omega(V) \to R: \phi(V) \to I(\phi(V))$ which is called the potential. $I(\phi(V)$ is the joint interation energy of $\phi(x)$ inside the domain V.

For an arbitrary finite $W \subset Z^2$, we define the energy $H(\phi(w))$ of the configuration ϕ in the domain W as

$$H(\phi(w)) = \sum_{v \subset w} I(\phi(v)). \tag{2.1}$$

The sum

$$H(\phi(w)\,|\,\phi(Z^2\backslash W)) = \sum_{\substack{V \cap W = \phi \\ V \cap Z^2\backslash W = \phi}} I(\phi(V)) \tag{2.2}$$

is the interaction energy between the configuration $\phi(W)$ and $\phi(Z^2\backslash W)$, where $\phi(Z^2\backslash W)$ is the boundary condition. The total energy of the configuration $\phi(W)$ is the sum $H(\phi(W)) + H(\phi(Z^2\backslash W))$.

The Hamiltonian is defined as the formal series.

$$H(\phi) = \sum_{V} I(\phi(V)), \tag{2.3}$$

where V ranges over all finite subsets of Z^2. The model of most interest to us is the 2-dimensional Ising model when $\Phi = \{1, -1\}$ (spins), and $I(\phi(V)) = 0$ unless $V = \{x,y\}$ with $||x-y|| = 1$.

We take $I(\phi(V)) = J\phi(x)\phi(y)$ where $J = \pm 1$.

$$H(\phi) = -J \sum_{\substack{\{x,y\} \\ ||x-y||=1}} \phi(x)\phi(y) \tag{2.4}$$

This Hamiltonian is translation invariant and reflection invariant. If J=+1, the model is ferromagnetic and J=-1 corresponds to an antiferromagnetic model.

The other model which will be of interest to us is the Ising model with an external magnetic field with Hamiltonian

$$H(\phi) = -J \sum_{\substack{\{x,y\} \\ ||x-y||=1}} \phi(x)\phi(y) - h \sum_{x \varepsilon Z^2} \phi(x). \tag{2.5}$$

where h may be random.

Finally a case of interest to us is the Hamiltonian

$$H(\phi) = \sum_{\substack{\{x,y\} \\ ||x-y||=1}} J(x,y)\phi(x)\phi(y) - h \sum_{x \in Z^2} \phi(x) \tag{2.6}$$

where $J(x,y)$ is random. This corresponds to a spin glass.

Let a measure be given on $\bar{\Omega}$ and for every $V \subset Z^2$ we consider the product measure

$$\prod_{s \in V} d\chi(\phi(s)) = d\mu.$$

We are interested in Gibbs distributions in the domain V which is a probability distribution on $\Omega(V)$ whose density with respect to $d\mu$ is given by

$$\frac{\exp(-H(\phi(V)))}{\int \exp(-H(\phi(V))d\mu} \tag{2.7}$$

$Z = \int \exp(-H(\phi(v))d\mu$ is called the partition function.

This corresponds to the so-called Gibbs distributions in V with free boundary conditions (Dirichlet Boundary conditions in the literature of quantum field theory). The signals ϕ we shall be interested in will have Gibbs distributions given by (2.7) and is defined by prescribing the potential I.

For modelling intensities of images we shall typically use a Gibbs distribution corresponding to a Hamiltonian of the 2-dimensional Ising model. Boundaries between smooth patches of images will be modelled as Gibbs distributions on the dual lattice. We shall discuss this in a later section.

2.1 Simulation of the Gibbs Distribution

Sample function's of the Gibbs distribution are obtained by constructing a Markov chain whose states correspond to the configurations of the Lattice field at time points 1,2,... in such a way that it has a unique invariant measure as the Gibbs measure $\exp(-H(\phi(V)))d\mu$. This chain clearly has to be reversible. Various algorithms for creating such a chain are known of which the Metropolis algorithm is the best known. In practice one may have to deal with a very large subset of a lattice and the random variables at the lattice points may take values in R^1. In this case it may be useful to make a diffusion approximation and simulate a stochastic differential equation with the same properties as above. This is the analogy to stochastic quantization we referred to before.

One therefore has to study the stochastic differential equation of the basic form:

$$d\phi(t,x) = -D_\phi H(\phi(t,x)) + dw(t,x)$$
$$\phi(0,x) = \phi_0(x) \tag{2.8}$$

where in general $\phi(t,\cdot)$ is a generalized random field and D_ϕ denotes "functional" derivative with respect to ϕ. The questions of interest is to construct the measure in the path space of ϕ and prove that the stochastic differential equation (2.8) has a unique invariant measure with density (with respect to an appropriate measure) $\exp(-H(\phi))$. The interest in this model for generating Gibbs distributions is that only Gaussian random numbers need to be generated and the computation is amenable to parallel processing.

3. STOCHASTIC MECHANICS, STOCHASTIC QUANTIZATION AND SIMULATION OF IMAGE INTENSITIES

We start with some well-known facts relating the Feynman-Kac Formula and the Girsanov Formula. (Carmona 1979, Simon 1980, Mitter 1980).

Let us suppose that $V:R^n \to R$, be measurable, bounded below and tends to $+\infty$ as $|x| \to \infty$ and consider the Schrodinger operator $H = -\Delta + V$ where Δ is the n-dimensional Laplacian. Then H defines a self-adjoint operator on $L^2(R^n; dx)$ which is bounded below and the lower bound λ (assumed to be 0) of the spectrum of H is an eigenvalue of H. Let $\varphi(x)$ be the corresponding eigenfunction of H, the so-called ground state and assume $\varphi(x) > 0$. We normalize $\varphi(x)$ i.e. $\int |\varphi(x)|^2 dx = 1$. Define the probability measure $d\mu = |\varphi(x)|^2 dx$, and the unitary operator

$$U : L^2(R^n; dx) \to L^2(R^n; d\mu(x))$$
$$: f \to \varphi^{-1} f.$$

If we define the Dirichlet form for $f, g \in C_c^\infty(R^n)$

$$\delta(f,g) = \frac{1}{2} \int_{R^n} \nabla f(x) \cdot \nabla g(x) dx$$

then a calculation shows

$$\delta(f,g) = (\mathcal{L} f, g)_\mu$$

where $(\ ,\ )_\mu$ denotes the scalar product in $L^2(R^n;d\mu)$ and $\mathcal{L}$ is the diffusion operator (self-adjoint, positive)

$$\mathcal{L}\phi = -\frac{1}{2}\Delta\phi + \nabla b \cdot \nabla\phi \tag{3.1}$$
$$b = -\log \varphi$$

Since ψ satisfies the equation

$$\frac{1}{2}\Delta\psi + V(x)\psi = 0, \tag{3.2}$$

in the sense of distributions (note that we have taken $\lambda=0$), a direct calculation shows

$$V(x) = \frac{1}{2}\left(|\nabla b(x)|^2 - \Delta b(x)\right). \tag{3.3}$$

Now using the Feynman-Kac formula for (3.2)

$$\psi(x) = E^{\omega}[\psi(x(t))\exp(-\int_0^t V(x(s))ds)\,|\,x(t) = x], \tag{3.4}$$

where E^{ω} denotes expectation w.r. to Wiener measure, the properties of V, equation (3.3) and the generalized Ito-differential rule (Meyer 1978), we see

$$\begin{aligned} L(t) &= \psi^{-1}(x(0)\psi(x(t))\exp(-\int_0^t V(x(x))ds) \\ &= \exp[-\int_0^t \nabla b(x(s)).dx(s) - \frac{1}{2}\int_0^t |\nabla b(x(s))|^2 ds] \end{aligned} \tag{3.5}$$

is a $(\Omega, \mathcal{F}_t, \mu^{\omega})$ martingale, where $\mathcal{F}_t$ is the σ-field generated by $(x(s)\,|\,0 \leq s \leq t)$ and μ^{ω} denotes Wiener measure. Therefore the process $(w(t)\,|\,t \geq 0)$ defined by

$$w(t) = x(t) - x(0) + \int_0^t \nabla b(x(s))ds \tag{3.6}$$

is standard Brownian motion with respect to the measure $\mu^{\mathcal{X}}$ defined by

$$\frac{d\mu^{\mathcal{X}}}{d\mu^{\omega}} = L(t) \tag{3.7}$$

Hence x(t) considered as a stochastic process on $(\Omega, \mathcal{F}_t, \mu^{\mathcal{X}})$ is a weak solution of the equation

$$\begin{aligned} dx(t) &= -\nabla b(x(t))dt + dw(t) \\ x(0) &= x. \end{aligned} \tag{3.8}$$

Indeed, the stochastic process defined by (3.8) is a Feller process, which is recurrent and has μ as its unique invariant measure. Therefore with the ground

state of a Schrodinger operator we have attached a diffusion process which is ergodic.

The converse procedure is of interest to us. Suppose we start with the stochastic differential equation on R^n given by (3.8) where $-\nabla b(.)$ is a singular drift. The case of interest to us is where $b(.)$ is a polynomial which is bounded below. Now, typically the Novikov condition

$$E \exp(\frac{1}{2} \int_0^t |\nabla b(x(s))|^2 ds) < \infty \tag{3.9}$$

will fail for these drifts and hence the Girsanov functional

$$L(t) = \exp(- \int_0^t \nabla b(x(s)).dx(s) - \frac{1}{2} \int_0^t |\nabla b(x(s))|^2 ds) \tag{3.10}$$

although a super-martingale, need not satisfy

$$E^{\mu} [L(t)] = 1$$

and hence fail to be a martingale. Therefore, the method of proof to show that equation (3.8) has a weak solution using the Girsanov formula will not work. To show, however that (3.8) has a weak solution and the process defined by (3.8) is recurrent and possesses a unique invariant measure, we consider the operator

$$\mathcal{L} = - \frac{1}{2} \Delta + \nabla b.\nabla, \tag{3.11}$$

and transform it to

$$H = - \frac{1}{2} \Delta + V(x), \tag{3.12}$$

where

$$V(x) = \frac{1}{2} (|\nabla b(x)|^2 - \Delta b(x)),$$

using the Gauge transformation,

$$\phi(x) \rightarrow \exp(b(x))\phi(x). \tag{3.13}$$

Then using functional-analytic arguments (cf. Segal 1970), one shows that the semigroup e^{-tH} is indeed, a strongly continuous, self-adjoint contractive semigroup on $L^2(d\mu)$ where μ is an appropriate probability measure, $e^{-tH}1=1$, is positivity preserving and improving and 1 is the unique ground state. It then follows that

e^{-tH} is ergodic with μ as its unique invariant measure. To carry out this program rigorously, one would use hyper-contractive estimates of Nelson and Segal and hence it is natural to work with the stochastic differential equation:

$$dx(t) = -\nabla b(x(t))dt + d\xi(t), \tag{3.14}$$

where

$$d\xi(t) = -A\xi(t)dt + dw(t), \tag{3.15}$$

where A is a symmetric nxn matrix. Therefore (3.15) defines a generalized Ornstein-Uhlenbeck process. Let us define formally the semigroup,

$$(e^{-tL}f)(x) = E^{\omega}[f(x(t))L(t)R(t)\,|\,x(t) = x], \tag{3.16}$$

where

$$L(t) = \exp(-\int_0^t \nabla b(x(x)).dx(s) - \frac{1}{2}\int_0^t |\nabla b(x(s))|^2 ds)$$

$$R(t) = \exp(-\int_0^t Ax(s).ds(s) - \frac{1}{2}\int_0^t |Ax(s)|^2 ds)$$

and E^{ω} denotes expectation with respect to Wiener measure. One wants to write (3.16) in Feynman-Kac form, i.e. in the form

$$(e^{-tL}f)(x) = E^{\omega}[f(x(t))\exp(-\int_0^t V(x(s)ds)\,|\,x(t) = x]\ . \tag{3.17}$$

To do this we need an Ito differential rule for b. For this we need that b is continuous,

$$\nabla b \in L^2_{loc.}(R^n;\, dx) \text{ and } \Delta b \in L^1_{loc.}(R^n;\, dx).$$

The measure $d\mu$ referred to previously then is

$$d\mu = \exp(-\int_0^t V(x(s))ds)d\mu^{\xi},$$

where

$$\frac{d\mu^{\xi}}{d\mu^{\omega}} = R(t).$$

The rest then follows from Nelson 1973, Segal 1970, Glimm-Jaffe 1981, Gross 1972.

These ideas can be generalized to random fields (cf. Jona Lasinio-Mitter loc.cit). Let $\Lambda \subset R^2$ be a square and consider the Laplacian Δ on R^2 with Dirichlet Boundary conditions, which is a self-adjoint operator on $V = L^2(\Lambda)$. Let $H^1(\Lambda)$ be the Sobolev space of functions f on R^2 with norm

$$||f||^2_{H^1} = \int_\Lambda \left|(-\Delta+1)^{1/2} f(x)\right|^2 dx$$

and let $V' = H^{-1}(\Lambda)$ be the dual-space of distributions with norm

$$||\phi||^2_{H^{-1}} = \int_\Lambda |(-\Delta+1)^{-1/2} \phi(x)|^2 dx.$$

Let μ_c denote the Gaussian measure on V' with mean zero and covariance operator C given by

$$C = (-\Delta+1)^{-1}. \tag{3.19}$$

Consider the stochastic differential equation on V':

$$\begin{aligned} d\phi(t) &= -\frac{1}{2} C^{-\varepsilon} \phi(t) dt + dw(t), \\ \phi(0) &= \phi, \ 0 < \varepsilon < \frac{1}{2} \end{aligned} \tag{3.20}$$

interpreted in the weak sense, where w(t) is a V'-valued Brownian motion with covariance $C^{1-\varepsilon}$ min(t,s) (defined using test functions).

One can show that this stochastic process defines a measure on the path space $C(0,\infty;V')$ and we denote this measure by P. Define the *semigroup*

$$(e^{-tL_0} f)(\phi) = E[f(\phi(t)) \,|\, \phi(t) = \phi], \tag{3.21}$$

where f is a "suitable" function and E denotes expectation w.r. to P-measure. One shows using the work Segal and Nelson referred to above that:

(i) e^{-tL_0} is a strongly continuous semigroup on $L^2(d\mu_c)$, which is contractive, self-adjoint.

(ii) is positivity preserving and improving on $L^2(d\mu_c)$; $e^{-tL_0}1 = 1$.

(iii) is a contraction on $L^p(d\mu_c)$; $1 \leq p < \infty$

(iv) is hypercontractive: $||e^{-tL_0}\phi||_{L^4(d\mu_c)} \leq C||\phi||_{L^2(d\mu_c)}$

for $\phi \in L^2(d\mu_c)$ and t sufficiently large. Moreover the stochastic process defined by (3.20) is ergodic and has μ_c as its unique invariant measure. Therefore the Ornstein-Uhlenbeck process defined by (3.20) is the stochastic quantization of the free euclidean field. In view of the well known relationship between the free field and the Ising model via the lattice approximation (cf. Guerra-Rosen-Simon 1975), we see that simulating the stochastic differential equation (3.20) is a powerful way of obtaining sample functions of the free field (Ising field) when the lattice is large. This can be accomplished in a parallel machine using multi-grid methods.

As remarked by Guerra-Simon-Rosen, the Ising (nearest-neighbour) nature of the lattice fields are not destroyed if we use non-gaussian random variables at the lattice points. This can be accomplished by studying the stochastic differential equation on V'

$$\begin{aligned} d\overset{\bullet}{\phi}(t) &= -\frac{1}{2}C^{-\varepsilon}\overset{\bullet}{\phi}(t)dt + \lambda C^{1-\varepsilon}:\overset{\bullet}{\phi}(t)^3:dt + dw(t) \\ \overset{\bullet}{\phi}(0) &= \phi,\ 0 < \varepsilon < \frac{1}{10} \end{aligned} \tag{3.22}$$

where : : denotes Wick ordering with respect to the covariance C (cf. Glimm-Jaffe 1981). Using an appropriate Ito rule (proved by approximation using tame functions), considering an aprpoximation of (3.22) using a spectral basis related to $(-\Delta+1)$ on $L^2(\Lambda)$, and using limiting arguments coupled with hyper-contractive estimates (see the discussion in the first part of this section), Mitter-Jona Lasinio show that the semi-group defined by

$$(e^{-tL}f)(\phi) = E(f(\phi(t))\exp(\xi(t))\,|\,\phi(t) = \phi \tag{3.23}$$

where

$$\xi(t) = -\frac{\lambda}{2}\int_0^t (:\phi(s)^3:,\, dw(s)) - \frac{\lambda^2}{8}\int_0^t (:\phi(s)^3:,\, C^{1-\varepsilon}:\phi(s)^3:)ds \tag{3.24}$$

where $\phi(t)$ is the Ornstein-Uhlenbeck process defined by (3.20) satisfy

(i) $e^{-tL}1=1$ and e^{-tL} is a contraction on $L^\infty(d\mu_c)$.

(ii) e^{-tL} is a strongly continuous, self-adjoint semigroup on $L^2(d\mu)$ with $d\mu = \exp(-\frac{\lambda}{4}\int_\Lambda :\phi^4(x):dx)d\mu_c$, positivity preserving and improving and 1 is the unique ground state.

(iii) e^{-tL} is ergodic and mixing.

This then allows them to show that the stochastic differential equation (3.22) has a weak solution and has as its unique invariant measure μ.

Since the stochastic differential equations under consideration defines a Markov process which is ergodic (and mixing), we have

$$\lim_{t\uparrow\infty} E_\phi((\hat{\phi}(t), f_1)(\hat{\phi}(t), f_2) \ldots (\hat{\phi}(t), f_n))$$

$$= \frac{\int(\phi, f_1)\ldots(\hat{\phi}, f_n)d\mu}{\int d\mu}$$

where the f_i are test functions. This enables us to compute spatial statistics of the time-in independent random field from the simulation of the stochastic differential equation.

4. Some Estimation Problems for Random Fields on a Lattice (cf. Marroquin 1985)

Let $V \subset Z^2$ be a finite subset. Consider a random field

$$f: V \to \bar{\Phi} \tag{4.1}$$

with Gibbs distribution having a density with respect to $d\mu$ (see (2.7)) given by

$$P_f = Z^{-1}\exp(-\frac{1}{T_0}H_0(f)), \tag{4.2}$$

and $T_0 > 0$ is a parameter (temperature).

We observe a corrupted version of f given by

$$g(j) = \Phi[G(j;f), \nu(j)], \quad j \in S \subset V, \tag{4.3}$$

where H(j,.) is a function with "local" support and Φ is invertible in the sense that $\nu(j) = \Phi^{-1}(g(j), G(j;f))$. We shall assume that $\nu(i)$ and $\nu(j)$ are indepenent and also $\nu(.)$ and f(.) are independent. Let us suppose that the distribution of $\nu(.)$ with respect to $d\mu$ has density

$$p_\nu > 0 \tag{4.4}$$

Define the functions

$$K(i;f,g(i)) = -\ln p_\nu(\Phi^{-1}(g(i), G(i;f)). \tag{4.5}$$

Then the conditional density $p_{f/g}$ can be written as

$$p_{f/g} = Z_p^{-1} \exp(-H_p(f,g)), \tag{4.6}$$

$$H_p(f,g) = \frac{1}{T_0} H_0(f) + \sum_{i \in S} K(i;f,g(i)) \tag{4.7}$$

and Z_p is a normalizing constant.

We can now provide a physical interpretation of the posterior distribution, by considering that, while the prior distribution (4.2) describes the behavior of a free field in thermal equilibrium, the distribution (4.7) describes the behavior of the same field coupled with a fixed (but spatially varying) external field whose value is given by g. The functions K whose magnitude depends on the noise variance, can then be interpreted as the coupling strengths between the two fields. This coupled system is also Gibbsian and if

$$G(i;f) = G(i;f(i))$$

the Markovian structure of this field will be identical to that of the original field.

The importance of this interpretation lies in the fact that the optimal estimate of f can be obtained simply by observing the equilibrium behavior of this coupled field.

In particular if $H_0(f)$ is the Hamiltonian of a ferromagnetic Gaussian Ising field and the observation is,

$$g(i) = f(i) + n(i), \tag{4.8}$$

with n(i) Gaussian, then the coupled field corresponds to a Ising field coupled to a random external field.

The estimates that are of interest to us depends on the choice of the problem. The two most important estimates for image reconstruction purposes are:

(i) $\hat{f} = E(f|g)$ (Conditional mean)

(ii) $\hat{f} = \text{Arg Max } p_{f|g}(f;g)$ (MAP)

From (4.6) and (4.7) the MAP estimate corresponds to minimizing the Hamiltonian $H_p(f,g)$ with respect to f.

4.1 Block Spin Transformation for MAP Estimation (Marroquin 1985)

In order to illustrate the analogy with Statistical Physics further we consider the MAP estimation of a binary Ising field withe the observations taken as the output of a binary symmetric channel with error rate ε. Therefore $\Phi = \{1,-1\}$ and the observation model is given by

$$P(g(i)|f(i)) = \begin{cases} 1-\varepsilon & \text{if } g_i = f_i \\ \varepsilon & \text{if } g_i \neq f_i. \end{cases}$$

Then it is easy to see that

$$H_p(f,g) = \frac{1}{T_0} \sum_{\substack{\{i,j\} \\ ||i-j||=1}} f(i)f(j) + \alpha \sum_i f(i)g(i) \tag{4.8}$$

and

$$\alpha = \ln\left(\frac{1-\varepsilon}{\varepsilon}\right).$$

Minimizing H_p is now a combinatorial optimization problem.

4.1.1 Simulated Annealing and Global Minimization

Simulated annealing is a new technique, developed by Kirkpatrick et al (1983) for the solution of combinatorial optimization problems. It is based on the idea that any cost functional of N variables, each of which can take values on some finite

set, can be considered as the Hamiltonian (Energy) of a physical system whose state corresponds to a particular value of these variables. Therefore, we can use, say, the Metropolis algorithm to generate, at any given "temperature" T (which now becomes a parameter of the optimization process) samples from the corresponding Gibbs measure. As $T \downarrow 0$ this measure should converge to a measure which concentrates on the states of minimum energy, the state of the system in thermal equilibrium at zero temperature will correspond to the value of f that minimizes the energy H(f) globally.

One serious difficulty, however, is that attaining thermal equilibrium might take a very long time at low temperatures. Kirkpatrick's idea was to start at a relatively high temperature (where thermal equilibrium is reached very fast), and then, to slowly cool the system, until "freezing" occurs and the state stops changing.

The analysis of this algorithm is presented in the Appendix.

4.1.2 Block Spin Transformations

In the case of the MAP estimator, the efficiency of the Simulated Annealing algorithm for the minimization of H_p can be improved by defining large "blocks" of sites (in a manner that is reminiscent of the "block-spin" strategy used by Wilson (1975) in connection with the renormalization group approach to the study of critical phenomena); the optimal estimate for the average value of the field in each of these blocks is found, and then progressively refined by subdividing the blocks in successive annealing stages. We will now show that, if we use a maximum entropy assumption, the structure of the MAP estimation process for Ising models is invariant under the "blocking" transformation; this means that the ground state (i.e., the MAP estimator) of the aggregated process (with blocks of size L) also corresponds to that of an Ising model with a coupled external field, in which the natural temperature is scaled by a factor of 1/L, and the noise (coupling) parameter by a factor of L^2. As a consequence of this scaling, the final temperature for the simulated annealing of this smaller network will be approximately L times larger than for the original problem.

If we denote by V(f(i), f(j)) = f(i)f(j) and q(f(i), g(i)) = f(i)g(i), let $V_c(f(i), f(j))$ and $q_c(f(i), g(i))$ be the extension to RxR of V and q respectively. We then write

$$H_p(f,g) = \frac{1}{T_0} \sum_{\substack{i,j=1 \\ ||i-j||=1}} V_c(f(i),f(j)) + \alpha \sum_i q_c(f(i),g(i)) \qquad (4.9)$$

We will now derive an expression for the energy in the "block spin" case. Let us

partition the original lattice into square blocks of side L. The "block observations" g_L will now be the density of 1's on each block, i.e.,

$$g_L(i) = \frac{1}{L^2} = \sum_{j \in B_i} g(j) \ .$$

where B_i is the i^{th} block. The "block field" f_L is defined in a similar way.

For a given f_L, we compute the energy by assuming a maximum entropy configuration, which occurs when the 1's that correspond to the given density $f_L(i)$ are randomly distributed within the block. The energy will have three terms:

1. Interactions between adjacent blocks:
The interaction between two adjacent blocks i and j will be:

$$I_{ij} = [-1 \ . \ (P_{11} + P_{00}) + 1 \ . \ (P_{10} + P_{01})] \ . \ L$$

where P_{kl} is the probability of having an element with state k on block i adjacent to an element with state l on block j:

$$P_{11} = f_1(i)f_L(j)$$

$$P_{01} = f_L(j)(1 - f_L(i))$$

$$P_{10} = f_L(i)(1 - f_L(j))$$

$$P_{00} = (1 - f_L(i))(1 - f_L(j))$$

Substituting these values we get:

$$I_{ij} = L[2(f_L(i) + f_1(j)) - 4f_L(i)f_L(j) - 1]$$

2. Interactions within each block: the internal interaction I_i is:

$$I_i = 2L(L-1)(-4f_L(i)^2 + 4f_L(i) - 1)$$

3. Interaction with the observations:
Assuming that the 1's in the observations and in the field are independently distributed we get:

$$I_{obs}(i) = \alpha L^2[f_L(i)(1 - g_L(i)) + (f - f_L(i))g_L(i)] =$$

$$= \alpha L^2[f_L(i) + g_L(i) - 2f_L(i)g_L(i)]$$

Finally, the Hamiltonian takes the form

$$H_L(f_L) = \frac{1}{T_0} \sum_{i,j} I_{ij} + \sum_i (\frac{1}{T_0} I_i + I_{obs}(i)) =$$

$$= L\{\frac{1}{T_0} \sum_{i,j} [2(f_L(i) + f_L(j)) - 4f_1(i)f_L(j) - 1] +$$

$$+ \frac{2}{T_0} (L - 1) \sum_i (-4f_L(i)^2 + 4f_L(i) - 1) +$$

$$+ \alpha L \sum_i (f_L(i) + g_L(i) - 2f_L(i)g_L(i)\}$$

note that the sums are taken over pairs of adjacent blocks, and over all the blocks, respectively. For L = 1, this expression reduces to (4.9) with

$$V_c(f(i),f(j)) = 2(f(i) + f(j)) - 4f(i)f(j) - 1$$

$$q_c(f(i), g(i)) = f(i) + g(i) - 2f(i)g(i).$$

For L > 1, the quadratic terms of H_L are:

$$\frac{L}{T_0} [-4 \sum_{i,j} f_L(i)f_L(j) - 8(L - 1) \sum_i f_L(i)^2]$$

and since

$$- 2 \sum_{i,j} f_L(i)f_L(j) + 2 \sum_i f_L(i)^2 = \sum_{ij} (f_L(i) - f_L(j))^2 \geq 0$$

it follows that

$$\sum_i f_L(i)^2 \geq \sum_{i,j} f_L(i)f_L(j)$$

and

$$-4 \sum_{i,j} f_L(i) f_L(j) - 8(L-1) \sum_i f_L(i)^2 <$$

$$< -(4 + 8(L-1)) \sum_i f_L(i)^2 \leq 0$$

which implies that H_L is negative definite for $L > 1$, and therefore, its minima, constrained to the hypercube $[0,1]^N$ (N_L is the total number of blocks) will always lie in a corner of such hypercube which means that we can use simulated annealing to find the global minimum of H_L, constraining the search to $\{0,1\}^N$. In this case, the energy to be minimized takes the simpler equivalent form (up to an additive constant):

$$U_L = \frac{1}{T_0/L} \sum_{i,j} V(f_L(i), f_L(j)) + \alpha L^2 \sum_i q(f_L(i), g_L(i))$$

The minimum energy solutions for each L can be interpreted as "coarse scale" representations of the original pattern f. Once a solution is obtained, the next refinement (for blocks of size L/2) can be efficiently obtained using the previous solution as a starting point, and initiating the annealing process at a lower temperature.

5. FINAL REMARKS

Due to lack of space we are unable to discuss:

(i) Estimation of boundaries using coupled models on the lattice Z^2 and the dual lattice of bonds on Z^2,

(ii) Estimation of the field and temperature parameter T using the innovations field

(iii) other problems in computation vision such as depth from stero-images, shape from shading etc.

A preliminary account can be found in Marroquin (1985).

Appendix on Simulated Annealing.

Let Ω be a finite set and let $|\Omega|$ denote the cardinality of Ω. Consider the problem of minimizing the energy function:

$$U: \Omega \to R: \; i \to U_i.$$

Let $N_0 = N \cup \{0\}$ where N are the natural numbers, $T_k > 0$, $k \in N_0$ be a sequence of real numbers. Consider a Markov chain $\{x_k\}_{k \in N}$ with 1-step transition matrices

$\{P^{(k,k+1)}\}_{k\in N}$ and some initial distribution constructed on a probability space and let $P_i^{(k)} = P\{x_k=i\}$, $i\in\Omega$, $k\in N_0$. The "annealing chain" is simulated as follows:

Suppose $x_k=i$. Then generate a random variable y with $P\{y=j\} = q_{ij}$ where $Q = \{q_{ij}\}_{i,j\in\Omega}$ a stochastic matrix. Suppose y=j, and then define

$$x_{k+1} = \begin{cases} j \text{ if } U_j \leq U_i \\ j \text{ if } U_j > U_i \text{ with probability } e^{(U_i-U_j)/T_k} \\ i \text{ otherwise} \end{cases}$$

We may think of the annealing algorithm as a probabilistic descent algorithm where the Q-matrix represents some prior distribution of "directions", transitions to some or lower energies are always allowed and transitions to higher energies are allowed with positive probability which tends to 0 as $k \to \infty$.

Hajek 1985 has given necessary and sufficient conditions on the rate at which T_k should go to zero such that $P\{x_k \in S^*\} \to 1$ as $k \to \infty$ where S^* is the set of global minimizing energy states. In this analysis the stochastic matrix Q has to be irreducible and satisfy a weak reversibility condition.

In Gelfand-Mitter 1985 a finite-time analysis of the annealing chain has been performed as well as a result on the rate of convergence of $P\{x_k \in S^*\} \to 1$ as $k \to \infty$ has been given, under somewhat weaker hypotheses on Q.

Finally, in Tsitsiklis 1985, necessary and sufficient conditions for $P\{x_k \in S^*\} \to 1$ as $k \to \infty$ are given by considering the annealing chain as a singularly perturbed Markov chain operating under different time-scales (under hypotheses weaker than that of Hajek).

Space does not permit us to give a detailed account of these results.

REFERENCES

1. Carmona, R. (1980): Processes de Diffusion Gouverne par la Forme de Dinichlet de 1' Operateur de Schrodinger in Seminaire de Probabilite's XIII, Springer-Verlag, Berlin, New York.

2. Fleming, W. and S.K. Mitter (1982): Optimal Control and Nonlinear Filtering for Nondegenerate Diffusion Processes, Stochastics, 8, pp. 63-78.

3. Geman, S. and D. Geman (1984): "Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images". IEEE Trans. Pattern Analysis and Machine Intelligence 6, 721-741 (1984).

4. Gelfand, S. and S.K. Mitter (1985): "Analysis of Simulated Annealing for Optimization", LIDS-P-1494.

5. Glimm, J. and A. Jaffe (1981): Quantum Physics, Springer-Verlag.

6. Grenander, U. (1984), "Tutorial in Pattern Theory", Div. of Applied Math. Brown University.

7. Gross, L. (1972): Existence and Uniqueness of Physical Ground States, J. Functional Analysis, 10, pp. 52-109.

8. Guerra, F., L. Rosen and B. Simon (1975): The $P(\phi)_2$ Euclidean Quantum Field Theory as Classical Statistical Mechanics, Annals of Mathematics, 101, pp. 111-259.

9. Hajek, B. (1985): "Cooling Schedules for Optimal Annealing," Preprint, Dept. Elec. Eng. and Coor. Science Lab., U. Illinois at Champaign-Urbana.

10. Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi (1983): "Optimization by Simulated Annealing". Science 220 671-680.

11. Lasinio, G. Jona, P.K. Mitter (1984): On the Stochastic Quantization of Field Theory, Preprint, *LPTHE 84/36, University of Paris VI.*

12. Marroquin, J.L. (1985): Probabilistic Solution of Inverse Problems, Doctoral Dissertation, Department of Electrical Engineering and Computer Science, M.I.T.

13. Meyer, P.A. (1978): La Formula de Ito pour le Mouvement Brownian d'apres G. Brosmaler, Lecture Notes in Mathematics, 649, 763-769, Springer-Verlag.

14. Mitter, S.K. (1980): Filtering Theory and Quantum Fields, in Analyse des Systemes, Astreisque, Vol. 75-76, pp. 199-205.

15. Mitter, S.K. (1983): Lectures on Nonlinear Filtering and Stochastic Control in Nonlinear Filtering and Stochastic Control, eds. Mitter, S.K., A. Moro, Springer Lecture Notes in Mathematics, 972, Springer-Verlag, Berlin, New York.

16. Mitter, S.K. (1979): "On the Analogy Between Mathematical Problems of Non-Linear Filtering and Quantum Physics", Ricerche Di Automatica, Vol. 10, No. 2, December.

17. Nelson, E. (1973): Probability Theory and Euclidean Field Theory in : Constructive Quantum Field Theory, G. Velo and A.S. Wightman, eds., Springer-Verlag.

18. Parisi, G. and Wu Yongshi (1981): Perturbation Theory without Gauge Fixing, Scientia Sinica, XXIV, pp. 483-496.

19. Segal, I. (1970): Construction of Nonlinear Local Quantum Processes I, Annals of Mathematics 92, pp. 462-481.

20. Simon, B. (1979): Functional Integration and Quantum Physics, Academic Press, New York.

21. Sinai, Ya G. (1982): Theory of Phase Transitions: Rigorous Results, Pargamon Press, Oxford, New York.

22. Tsitsiklis, J. (1985): "Markov Chains with Rare Transitions and Simulated Annealing", LIDS-P-1497, Laboratory for Information and Decision Systems, M.I.T.

23. Wilson, K.G. (1973): "The Renormalization Group: Critical Phenomena and the Kondo Problem", Rev. Mod. Phys. 47, 4 .

END

FILMED

3-86

DTIC