Journal of Applied Mathematics and Stochastic Analysis
8, Number 3, 1995, 249-260

IDENTIFICATION OF LINEAR STOCHASTIC SYSTEMS BASED ON PARTIAL INFORMATION

N.U. AHMED
University of Ottawa
Department of Electrical Engineering and Department of Mathematics
Ottawa, Ontario
Canada

S.M. RADAIDEH
University of Ottawa
Department of Electrical Engineering
Ottawa, Ontario
Canada

(Received October, 1994; Revised May, 1995)

ABSTRACT

In this paper, we consider an identification problem for a system of partially observed linear stochastic differential equations. We present a result whereby one can determine all the system parameters including the covariance matrices of the noise processes. We formulate the original identification problem as a deterministic control problem and prove the equivalence of the two problems. The method of simulated annealing is used to develop a computational algorithm for identifying the unknown parameters from the available observation. The procedure is then illustrated by some examples.

Key words: Identification, Stochastic Systems, Partial Information, Simulated Annealing.

AMS (MOS) subject classifications: 93P30, 93E12.

1. Introduction

Over the last several years, considerable attention has been focused on an identification problem of stochastic systems governed by linear or nonlinear Itô equations [2, 3, 7, 8, 10]. In [10], the identification problem for partially observed linear time-invariant systems was considered. Using linear filter theory, maximum likelihood approach, and the smoothness of solutions of an algebraic Riccati equation, sufficient conditions were obtained for the consistency of the likelihood estimate.

In [8], Liptser and Shiryayev considered the identification problem for a class of completely observed systems governed by a stochastic differential equation of the form

$$dX(t) = h(t, X(t))\alpha dt + dW(t), \qquad t \geq 0, \tag{1}$$

where X is a real-valued stochastic process and α is some unknown parameter. Using the maximum likelihood approach, the authors [8] obtained an explicit expression for the maximum likeli-

Printed in the U.S.A. ©1995 by North Atlantic Science Publishing Company

hood estimate $\hat{\alpha}$. An extension of this result to a multi-parameter problem $\alpha \in R^m$ for stochastic systems in R^n was considered by Ahmed [1]. In [7], Legland considered an identification problem for a more general class of systems governed by stochastic differential equations of the form

$$dy(t) = h(\alpha, X(t))dt + dW(t), \qquad t \geq 0, \tag{2}$$

where α is an unknown parameter and $X(t)$ is a diffusion process. Utilizing the maximum likelihood approach along with forward and backward Zakai equations, a numerical scheme was developed for computing the parameter α given the output history $y(s)$, $s \leq t$.

In [3], Dabbous and Ahmed considered the problem of identification of drift and dispersion parameters for a general class of partially observed systems governed by the following system of Itô equations

$$\begin{aligned} dX(t) &= a(t, X(t), \alpha)dt + b(t, X(t), \alpha)dW(t), \quad t \in [0, T], \ X(0) = X_0, \\ dy(t) &= h(X(t), \alpha)dt + \sigma_0(t, y(t))dW_0(t), \quad t \in [0, T], \ y(0) = 0. \end{aligned} \tag{3}$$

Using the pathwise description of Zakai equation, they formulated the original identification problem as a deterministic control problem in which the unnormalized conditional density (solution of Zakai equation) is treated as a state, the unknown parameters as control and the likelihood ratio as an objective functional.

In [2], Bagchi considered a situation with an unknown observation covariance noise in which case the likelihood functional cannot be apparently defined. Bagchi proposed a functional analogous to the likelihood functional by giving an apriori guess of the observation covariance noise. However, from the numerical point of view, an a priori guess should be close to the true value.

Newton's method is the usual procedure for computing the maximum likelihood estimates (MLE) [4] which involved recursive calculation of the gradient vector and Hessian matrix of the (MLE) at a fixed valued of the parameter vector. The drawback of this method is that the convergence to the desired optimum fails whenever Hessian matrix has some negative eigenvalues or nearly singular.

In [7, 8, 9], identification of drift parameters for completely observed systems were considered. In [3], which considers partially observed identification problem, the authors used the Zakai equation as the basic state equation which, of course, is a partial differential equation. For n-dimensional problems, $n \geq 2$, the associated computational problem becomes nontrivial. It appears that for partially observed nonlinear problems there is no escape from PDE. In this paper we consider partially observed linear problems and develop techniques for identification of all the parameters including the covariance matrices of the Wiener processes without resorting to PDE.

2. Identification Problem (IP)

To introduce the identification problem, we shall need some basic notations. For each pair of integers $n, m \in N$, let $M(n \times m)$ denote the space of $n \times m$ matrices with all real entries and let $M^+(m \times m)$, a subset of $M(m \times m)$, denote the class of all positive definite matrices, and $I(d \times d)$ denote the space of $d \times d$ identity matrices. Let $*$ denote the transpose of a matrix or a vector. Define

$$M_0(p \times q) \equiv \{\sigma \in M(p \times q) : \sigma\sigma^* \in M^+(p \times p)\},$$

and

$$\Sigma \equiv M(d \times d) \times M(d \times m) \times M(p \times d) \times M_0(p \times q).$$

We shall denote our identification problem by (IP) which is described as follows: We are given a class of linear stochastic systems governed by

$$\begin{cases} dX(t) = AX(t)dt + \sigma dW(t), & X(0) = X_0 \\ dy(t) = HX(t)dt + \sigma_0 dW_0(t), & y(0) = 0, \end{cases} \tag{4}$$

where X is an R^d-valued signal process, and y is an R^p-valued observation process. The processes $\{W, W_0\}$ are $\{R^m, R^q\}$-valued independent standard Wiener processes. In general, each $\pi = \{A, \sigma, H, \sigma_0\} \in \Sigma$, determines a distinct linear stochastic system of form (4).

The (IP) is to estimate the unknown parameters $\pi = \{A, \sigma, H, \sigma_0\}$, based on the observation $\{y(t), 0 \le t \le T\}$, and the knowledge of the mean $\bar{X}_0$ and covariance $P_0 \equiv E\{(X_0 - \bar{X}_0)(X_0 - \bar{X}_0)^*\}$. Let $\pi^0 \in \Sigma$ denote the true system parameters. Our objective is to develop a method including an algorithm for identification of the true parameter. We formulate this problem as a deterministic control problem and use a simulated annealing algorithm to estimate the unknown parameters.

3. Formulation of the Identification Problem as a Deterministic Control Problem

In this section, we shall show the (IP) is equivalent to an optimal control problem. This is given in the following theorem.

Theorem 1: *Consider the (IP) as stated above. This problem is equivalent to the following optimal control problem:*

estimate $\pi = \{A, \sigma, H, \sigma_0\} \in \Sigma$ that minimizes the objective functional

$$J(\pi, T) = \int_0^T Tr\{(\tilde{K}_\pi(t) + K_\pi(t) - P_\pi(t))(\tilde{K}_\pi(t) + K_\pi(t) - P_\pi(t))^*\}dt,$$

subject to the dynamic constraints

$$\begin{cases} de(t,\pi) = (A - K_\pi H^* R^{-1} H)e(t,\pi)dt + K_\pi H^* R^{-1}[dy - H\bar{X}(t,\pi)dt], \quad e(0,\pi) = 0, \\ \dot{K}_\pi(t) = AK_\pi + K_\pi A^* + \sigma\sigma^* - K_\pi H^* R^{-1} H K_\pi, \quad K_\pi(0) = K_0 = P_0 \\ \dot{\bar{X}}(t,\pi) = A\bar{X}(t,\pi), \bar{X}(0,\pi) = EX_0 \\ \dot{P}_\pi(t) = AP_\pi + P_\pi A^* + \sigma\sigma^*, P(0,\pi) = P_0, \end{cases} \tag{5}$$

where $R \equiv \sigma_0\sigma_0^*$ and $\tilde{K}_\pi(t) = E(e(t,\pi)e(t,\pi)^*)$.

Proof: Let $\pi \in \Sigma$ constitute the system given by (4). Then by Kalman-Bucy filter theory, the estimate is given by $\hat{X}(t,\pi) = E(X(t,\pi)/\mathcal{F}_t^y)$ which satisfies the following stochastic differential equation (SDE):

$$\begin{cases} d\hat{X}(t,\pi) = A\hat{X}(t,\pi)dt + K_\pi(t)H^* R^{-1} d\nu(t,\pi), \hat{X}(0,\pi) = \bar{X}_0, \\ \nu(t,\pi) = y(t) - \int_0^t H\hat{X}(s,\pi)ds, \end{cases} \tag{6}$$

where K_π is the state estimation error covariance and it satisfies the following matrix Riccati differential equation

$$\dot{K}_\pi(t) = AK_\pi + K_\pi A^* + \sigma\sigma^* - K_\pi H^* R^{-1} H K_\pi, K_\pi(0) = K_0 = P_0. \tag{7}$$

Here, $\mathfrak{F}_t^y$ is the filtering adapted to $\{y(s); t \in [0,T]\}$, and $\nu(t,\pi)$ is a Wiener process. The latter is a so-called innovations process, with

$$E\{\nu(t,\pi)\nu^*(s,\pi)\} = R\,min(t,s). \tag{8}$$

The mean of $X = \{X(t,\pi), t \geq 0\}$, given by $\bar{X}(t,\pi) = E(X(t,\pi))$, satisfies the following deterministic differential equation

$$\dot{\bar{X}}(t,\pi) = A\bar{X}(t,\pi), \bar{X}(0,\pi) = EX_0. \tag{9}$$

Defining $e(t,\pi) = \hat{X}(t,\pi) - \bar{X}(t,\pi)$, we have from equations (6) and (9) that e satisfies the following (SDE):

$$de(t,\pi) = (A - K_\pi H^* R^{-1} H)e(t,\pi)dt + K_\pi H^* R^{-1}[dy - H\bar{X}(t,\pi)dt], e(0,\pi) = 0. \tag{10}$$

In terms of the innovations process, one can write system (10) as:

$$de(t,\pi) = Ae(t,\pi)dt + K_\pi H^* R^{-1} d\nu(t,\pi), e(0,\pi) = 0. \tag{11}$$

Further, the process $e = \{e(t,\pi), t \geq 0\}$ and the error covariance matrix K_π are related through the equation

$$(K_\pi(t)\eta,\eta) = (P_\pi(t)\eta,\eta) - E(e(t,\pi),\eta)^2, \text{ for all } \eta \in R^d, \tag{12}$$

where P_π is the covariance of the process $X = \{X(t,\pi), t \geq 0\}$ and it satisfies the following differential equation:

$$\dot{P}_\pi(t) = AP_\pi(t) = P_\pi(t)A^* + \sigma\sigma^*, \;\; P_\pi(0) = P_0. \tag{13}$$

This is justified as follows: by definition, for each $\eta \in R^d$, we have

$$\begin{aligned}(K_\pi(t)\eta,\eta) &\equiv E(X(t,\pi) - \hat{X}(t,\pi),\eta)^2 \\ &= E(X(t,\pi) - \bar{X}(t,\pi) + \bar{X}(t,\pi) - \hat{X}(t,\pi),\eta)^2 \\ &= E(X(t,\pi) - \bar{X}(t,\pi),\eta)^2 + E(\bar{X}(t,\pi) - \hat{X}(t,\pi),\eta)^2 \\ &+ 2E((X(t,\pi) - \bar{X}(t,\pi),\eta)(\bar{X}(t,\pi) - \hat{X}(t,\pi),\eta)).\end{aligned} \tag{14}$$

Since $((\bar{X}(t,\pi) - \hat{X}(t,\pi))$ is $\mathfrak{F}_t^y$-measurable, we have

$$E\{(X(t,\pi) - \bar{X}(t,\pi),\eta)(\bar{X}(t,\pi) - \hat{X}(t,\pi),\eta)\} = -E(\bar{X}(t,\pi) - \hat{X}(t,\pi),\eta)^2, t \in [0,T]. \tag{15}$$

Using this in the third term of the preceding equation, we obtain that

$$(K_\pi(t)\eta,\eta) = (P_\pi(t)\eta,\eta) - E(e(t,\pi),\eta)^2$$

$$= (P_\pi(t\eta,\eta) - (\widetilde{K}_\pi(t)\eta,\eta), \tag{16}$$

for each $t \in [0,T]$. This validates equation (12). For the identification of the system parameters, equations (11) and (16) are most crucial. Suppose the process $\{y^0(t), t \in [0,T]\}$, as observed from laboratory (field) measurements, corresponds to the true system parameters, say π^0. If one uses the same observed process as an input to the model system (11) with the arbitrary choice of the parameter π, it is clear that one can not expect equality (16) to hold. On the other hand, (16) must hold if the trial parameter π coincides with the true parameter π^0. Hence it is logical to adjust the parameter π to have this equality satisfied. This can be achieved by choosing for the cost function, the functional given by

$$J(\pi,T) = \int_0^T Tr\{(\widetilde{K}^0_\pi(t) + K_\pi(t) - P_\pi(t))(\widetilde{K}^0_\pi(t) + K_\pi(t) - P_\pi(t))^*\}dt, \tag{17}$$

where K_π and P_π are solutions of equations (7) and (13), and $\widetilde{K}^0_\pi$ is the covariance of the process $e_0(t,\pi) = e(t,\pi,y^0)$ given by the solution of equation (10) driven by the observed process y^0. This functional is to be minimized on Σ subject to dynamic equations (5) as proposed in the theorem. This proves that the (IP) is equivalent to the optimal control problem as stated. This completes the proof.

Remark 1: (Uniqueness) Let Σ_0 be a subset of Σ. Define $m_0 \equiv inf\{J(\pi,T),\ \pi \in \Sigma_0\}$. Given that the actual physical system is governed by a linear Itô equation, in general we may expect that $m_0 = 0$. In any case, let $M \equiv \{\pi \in \Sigma_0 : J(\pi,T) = m_0\}$ denote the set of points in Σ_0 at which the infimum is attained. It is easy to verify that this set is closed. If the set M is singleton, the system is uniquely defined. In general, for (IP's), which are basically inverse problems, we may not expect uniqueness since the same natural behavior may be realized by many different parameters.

Remark 2: (Weighted cost functional) The cost functional $J(\pi,T)$, given by equation (17), can be generalized by introducing a positive semidefinite weighting matrix (valued function) $\Gamma(t)$ in the cost integrand giving

$$J(\pi,T) \equiv \int_0^T Tr\{L_\pi \Gamma(t) L^*_\pi\}, \quad L_\pi \equiv \widetilde{K}^0_\pi + K_\pi - P_\pi.$$

By choosing Γ suitably, one can assign weights as required for any specific problem.

4. Measurement of Autocovariance Function of $\{e\}$

In the real world, we can never measure the actual covariance $\widetilde{K}^0_\pi = E(e_0(t,\pi)e^*_0(t,\pi))$ because we can never have all sample functions of the process $\{e_0\}$. One obtains only a sample path $\{y^0(t), t \in [0,T]\}$ corresponding to the true system parameters, π^0. Thus, our only recourse is to determine time average based on observation of one sample path of finite length. The time interval is taken large enough so that the ensemble average equals the time average. This is possible if the process is ergodic. In the following discussion we will establish sufficient conditions for ergodicity of $\{e\}$ which will be presented in Proposition 1.

We extend the Brownian motions W and W_0 over the entire real line by standard techniques, that is, we introduce two independent Brownian motions $\widetilde{W}$ and $\widetilde{W}_0$ which are also independent of W and W_0 and define

$$W(t) = \begin{cases} W(t), & t \geq 0 \\ \widetilde{W}(-t), & t < 0 \end{cases} \tag{18}$$

$$W_0(t) = \begin{cases} W_0(t), & t \geq 0 \\ \widetilde{W}_0(-t), & t < 0. \end{cases} \tag{19}$$

Therefore, for $t_0 > 0$, systems (4) and (11) can be rewritten as

$$\begin{aligned} dX(t) &= AX(t)dt + \sigma dW(t), \quad X(-t_0) = X_0 \\ dy(t) &= HX(t)dt + \sigma_0 dW_0(t), \quad y(-t_0) = 0, \end{aligned} \tag{20}$$

and

$$de(t,\pi) = Ae(t,\pi)dt + K_\pi H^* R^{-1} d\nu(t,\pi), \quad e(-t_0,\pi) = 0. \tag{21}$$

Suppose the following conditions hold:

Condition I: For every $\pi = \{A, \sigma, H, \sigma_0\} \in \Sigma_0$, A is a stable matrix, i.e., all eigenvalues of A have negative real parts.

Condition II: For every $\pi = \{A, \sigma, H, \sigma_0\} \in \Sigma_0$, the pair (A, H) is completely observable, that is, the rank $[H^*, A^*H^*, \ldots, (A^*)^{d-1}H^*] = d$.

Condition I implies that the initial condition of the state has no effect on the asymptotic behavior of the system. Conditions I and II imply that $\lim_{t\to\infty} K_\pi(t)$ exists and is unique. We denote this limit by K_π^0, which satisfies the algebraic Riccati equation

$$AK_\pi^0 + K_\pi^0 A^* + \sigma\sigma^* - K_\pi^0 H^* R^{-1} H K_\pi^0 = 0. \tag{22}$$

Furthermore, the matrix $A - K_\pi^0 H^* R^{-1} H$ is stable [5, Theorem 4.11, p. 367].

Using the steady state version of Kalman-Bucy filter (6), that is, using K_π^0 instead of $K_\pi(t)$ in equation (6), one can write $\widetilde{K}_\pi(t)$ as

$$\begin{aligned} \widetilde{K}_\pi(t) &\equiv E(\widehat{X}(t,\pi)\widehat{X}^*(t,\pi)) - \bar{X}(t,\pi)\bar{X}^*(t,\pi) \\ &= V_1(t) - GV(t)G^* - \bar{X}(t,\pi)\bar{X}^*(t,\pi), \end{aligned} \tag{23}$$

where the matrices G, V_1 and V are given as follows:

The matrix G is a $d \times 2d$ with elements $g_{i,i} = 1$, $g_{i,i+d} = -1$, for $1 \leq i \leq d$, and 0 everywhere else. The matrix $V_1(t) = E(X(t,\pi)X^*(t,\pi))$ and it satisfies the matrix differential equation

$$\frac{dV_1(t)}{dt} = AV_1(t) + V_1(t)A^* + \sigma\sigma^*. \tag{24}$$

The matrix $V(t) = \begin{bmatrix} E(X(t,\pi)X^*(t,\pi)) & E(X(t,\pi)\widehat{X}^*(t,\pi)) \\ E(\widehat{X}(t,\pi)X^*(t,\pi)) & E(\widehat{X}(t,\pi)\widehat{X}^*(t,\pi)) \end{bmatrix}$ and it satisfies the matrix differential equation

$$\frac{dV(t)}{dt} = \mathcal{A}_\pi V(t) + V(t)\mathcal{A}_\pi^* + C_\pi C_\pi^*, \tag{25}$$

where

$$\mathcal{A}_\pi = \begin{bmatrix} A & 0 \\ K_\pi^0 H^* R^{-1} H & A - K_\pi^0 H^* R^{-1} H \end{bmatrix}, \quad C_\pi = \begin{bmatrix} \sigma & 0 \\ 0 & K_\pi^0 H^* R^{-1} \sigma_0 \end{bmatrix}.$$

Under the conditions I and II, the matrices $\{A, \mathcal{A}_\pi\}$ are both stable for every $\pi \in \Sigma_0$, and therefore, equations (24) and (25) have steady state solutions V_1^0 and V^0, respectively. They are given by the solution of the following algebraic Lyapunov equations

$$AV_1^0 + V_1^0 A^* + \sigma\sigma^* = 0 \tag{26}$$

and

$$\mathcal{A}_\pi V^0 + V^0 \mathcal{A}_\pi^* + C_\pi C_\pi^* = 0, \tag{27}$$

respectively. We shall show that the process $e(t, \pi)$, given by equation (11), is ergodic. This is presented in the following proposition.

Proposition 1: *Suppose that Conditions I and II are satisfied, and the processes W and W_0 are the extended Brownian motions in* (18) *and* (19). *Then for each $\pi \in \Sigma_0$, the process $\{e(t, \pi), t \in R\}$ is stationary and ergodic.*

Proof: It is clear from equation (11) that the random process $e(\cdot, \pi)$ is a zero mean Gaussian process. It is stationary if we can show that the corresponding autocovariance matrix $R(s, t)$ is dependent only on the time difference. For this purpose define

$$\begin{aligned} R(s,t) &\equiv E(e(s,\pi)e^*(t,\pi)) \\ &\equiv E(\hat{X}(s,\pi)\hat{X}^*(t,\pi)) - \bar{X}(s,\pi)\bar{X}^*(t,\pi) \\ &= I_1(s,t) - G\, I_2(s,t) G^* - \bar{X}(s,\pi)\bar{X}^*(t,\pi), \end{aligned} \tag{28}$$

where the matrices I_1 and I_2 are given by

$$\begin{aligned} I_1(s,t) = E \int_{-t_0}^{s} \int_{-t_0}^{t} & e^{A(s-\theta)} \sigma dW(\theta) dW^*(\gamma) \sigma^* e^{A^*(t-\gamma)} \\ & + e^{A(s+t_0)} V_1(-t_0) e^{A^*(t+t_0)} \end{aligned} \tag{29}$$

and

$$\begin{aligned} I_2(s,t) = E \int_{-t_0}^{s} \int_{-t_0}^{t} & e^{\mathcal{A}_\pi(s-\theta)} C_\pi d\beta(\theta) d\beta^*(\gamma) C_\pi^* e^{\mathcal{A}_\pi^*(t-\gamma)} \\ & + e^{\mathcal{A}_\pi(s+t_0)} V(-t_0) e^{\mathcal{A}_\pi^*(t+t_0)} \end{aligned} \tag{30}$$

for $\beta = [W, W_0]^*$. Setting $s - t \equiv \tau$, after some elementary calculations, we have

$$I_1(s,t) = \begin{cases} e^{A\tau} V_1(t), & \tau \geq 0 \\ V_1(s) e^{-A^*\tau}, & \tau \leq 0 \end{cases} \tag{31}$$

$$I_2(s,t) = \begin{cases} e^{\mathcal{A}_\pi \tau} V(t), & \tau \geq 0 \\ V(s) e^{-\mathcal{A}_\pi^* \tau}, & \tau \leq 0. \end{cases} \tag{32}$$

Then, from (9), (28), (31), and (32), we have

$$R(s,t)=\begin{cases} e^{A\tau}V_1(t)-Ge^{\mathcal{A}_\pi\tau}V(t)G^*-e^{A\tau}\bar{X}(t,\pi)\bar{X}^*(t,\pi), & \tau\geq 0\\ V_1(s)e^{-A^*\tau}-GV(s)e^{-\mathcal{A}_\pi^*\tau}G^*-e^{A\tau}\bar{X}(t,\pi)\bar{X}^*(t,\pi), & \tau\leq 0. \end{cases} \tag{33}$$

Since A and $\mathcal{A}_\pi$ are stable matrices, letting $t_0\to+\infty$, we obtain

$$R(\tau)=\begin{cases} e^{A\tau}V_1^0-Ge^{\mathcal{A}_\pi\tau}V^0G^*, & \tau\geq 0\\ V_1^0e^{-A^*\tau}-GV^0e^{-\mathcal{A}_\pi^*\tau}G^*, & \tau\leq 0. \end{cases} \tag{34}$$

The latter proves that the process $\{e(t,\pi), t\in R\}$ is stationary. It is well known that the zero mean stationary Gaussian process is ergodic if the corresponding autocovariance matrix $R(\tau)$ satisfies [6, Theorem 7.6.1, p. 484]

$$\int_{-\infty}^{\infty} \| R(\tau) \| \, d\tau < \infty. \tag{35}$$

It is clear from (34) that $R(\tau)=R^*(-\tau)$ and hence

$$\int_{-\infty}^{\infty} \| R(\tau) \| \, d\tau = 2\int_{0}^{\infty} \| R(\tau) \| \, d\tau. \tag{36}$$

For any $\tau>0$, we have

$$\| R(\tau) \| \leq \| e^{A\tau} \| \, \| V_1^0 \| + \| G \|^2 \| e^{\mathcal{A}_\pi\tau} \| \, \| V^0 \|. \tag{37}$$

Since, for every $\pi\in\Sigma_0$, A and $\mathcal{A}_\pi$ are stable matrices, it holds true that

$$\| e^{A\tau} \| \leq e^{\lambda_1\tau}, \; \| e^{\mathcal{A}_\pi\tau} \| \leq e^{\lambda_2\tau},$$

where $\lambda_1<0$, $\lambda_2<0$ are the real parts of the largest eigenvalues of the matrices A and $\mathcal{A}_\pi$, respectively. Hence, it follows that

$$\int_{0}^{\infty} \| R(\tau) \| \, d\tau < \infty, \tag{38}$$

and therefore (35) holds, proving the ergodicity of the process $\{e\}$.

Therefore, under conditions I and II and by taking the observation time T large enough, one can approximate the ensemble average of $e(t,\pi)e^*(t,\pi)$ by its time average. This is what has been done in estimating the unknown parameters in our simulation experiments as given in section 6.

If the stability and observability conditions are not satisfied, one must use Monte-Carlo techniques to produce an ensemble average.

5. Numerical Algorithm

In this paper we applied the method of simulated annealing to determine the optimal parameters that minimize the cost function. The method of simulated annealing is an iterative improvement technique that is suitable for large scale minimization problems. The method avoids being trapped in local minima by using stochastic approach for making moves, based on Metropolis optimization algorithm to minimize the cost function [9]. It works by analogy to the physical annealing of molten material. In the physical situation, the material is cooled slowly, allowing it to coalesce into the lowest possible energy state giving the strongest physical structure. If a liquid metal is cooled quickly, it may end up in a polycrystalline state having a higher energy.

The main idea behind this algorithm is while being at a high temperature, τ_a called the annealing temperature, where most moves are accepted, then slowly reduce the temperature, while reducing the cost function until only "good" moves are accepted. The pseudo-code of the algorithm is presented as follows:

Step 1: Generate an initial scheduling order randomly and set the temperature at high level.

Step 2: Randomly pick one of the elements of $\pi = \{c_1, c_2, \ldots, c_n\} \in \Sigma_0$. A picked parameter moves as

$$c_i = c_i + \alpha U_{c_i} \tag{39}$$

where α is the maximal allowed displacement, which for the sake of this argument is arbitrary; U_{c_i} is a random number uniformly distributed in the interval $[-1, +1]$, and U_{c_i} is independent of U_{c_j}, for $i \neq j$.

Step 3: Calculate the change in the cost function, ΔJ, which is caused by the move of c_i into $c_i + \alpha U_{c_i}$.

Step 4: If $\Delta J < 0$ (i.e., the move would bring the system to a state of lower energy) we allow the move.

Step 5: If $\Delta J > 0$ we allow the move with probability $exp(-\Delta J/\tau_a)$; i.e., we take a random number U uniformly distributed between 0 and 1, and if $U < exp(-\Delta J/\tau_a)$, we allow the move. If $U > exp(-\Delta J/\tau_a)$, we return it to its old value.

Step 6: Go to step 2 until the cost function stabilizes.

Step 7: If $\tau_a = 0$, then stop; otherwise reduce the temperature, and repeat steps 2-6.

6. Examples and Illustrations

In this section we will present a two-dimensional example illustrating our results. We assume that the observation data $\{y^0(t), t \in [0, T]\}$ for the real system is generated by the true parameters $\pi^o = \{A^o, \sigma^o, H^o, \sigma_0^o\}$ where

$$A^o = \begin{bmatrix} -2.0 & 2.0 \\ 0.5 & -2.0 \end{bmatrix}, \ \sigma^9 = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}, \ H^0 = \begin{bmatrix} 0.0 & 1.0 \end{bmatrix}, \ \sigma_0^o = [1.0].$$

The basic procedure used to obtain the best estimate of the unknown parameters using the algorithm as proposed in section 4 is as follows: Let τ_c be the initial choice for the true parameter π^0. Using the algorithm with this choice of π_c and starting the annealing temperature at τ_a we arrive

at $\pi_{\tau a}$ by decreasing τ_a step by step (slowly) to zero. The distance between the computed parameter $\pi_{\tau a}$ (using the algorithm) and the true parameter π^0 is denoted by $d(\pi^0, \pi_{\tau a})$. The simulation was carried out with sampling interval $\delta = 0.01 sec.$, and the observation time $T \in [0, 120] sec.$, and the weighting matrix $\Gamma(t) = 1000I$.

Example 1: In general, the (system) dynamic noise and measurement noise are modelled as Wiener processes but the noise power and hence the system and measurement noise covariance matrices may be unknown. In the Itô equation, the martingale terms may then be modelled as σdW and $\sigma_0 dW_0$ where W and W_0 are standard Wiener processes and σ and σ_0 are constant but unknown matrices. We assume also that the matrices A and H are unknown. The problem is to determine A, σ, H and σ_0 based on the observation data $\{y^0(t), t \in [0, T]\}$. Then end results are given in table 1 which are quite close to the true values. Figure 1 shows the estimation error as a function, $\tau_a \rightarrow d(\pi^0, \pi_{\tau a})$, of the starting annealing temperature τ_a. For fixed observation time T, three curves are plotted for three different initial choices π_c for the true parameter π^0. It is clear from this graph that the larger the discrepancy is between the true value and the initial choice, the larger is the starting annealing temperature required to reach the true values.

Table 1

	Starting value	Estimated value	Actual value
a_{11}	-1.0	-1.994406	-2.0
a_{12}	1.0	1.995625	2.0
a_{21}	2.0	0.504037	0.5
a_{22}	-3.0	-2.064103	-2.0
s_{11}	0.1	1.010031	1.01
s_{12}	0.5	0.2000456	0.2
s_{22}	0.1	1.010027	1.01
h_{11}	-0.7	-0.000276	0.0
h_{12}	2.0	1.009440	1.0
σ_0^2	0.1	0.992090	1.0

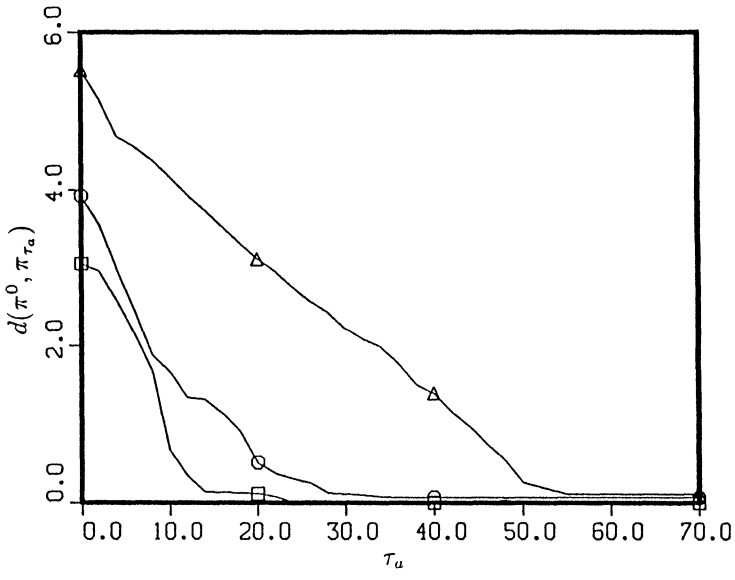


Figure 6.1: The distance between the computed parameter $\pi_{\tau a}$ and the true parameter π^0, $d(\pi^0, \pi_{\tau a})$, as a function of the starting annealing temperature τ_a, for three different initial choices π_c.

Figure 2 shows the estimated error as a function of the observation time T, $T \rightarrow d(\pi^0, \pi_T)$, where π_T is the estimated value of π^0 based on the observation $\{y^0(t), 0 \leq t \leq T\}$ until time T. As expected, it is a nonincreasing function of T and tends to a limit (saturation) as T becomes larger and π_T comes closer to π^0. The best starting annealing temperature required to obtain the estimate π_T, in this example, was found to be 25. In other words, the choice of a starting annealing temperature beyond 25 doesn't improve the estimate; it only consumes more CPU time.

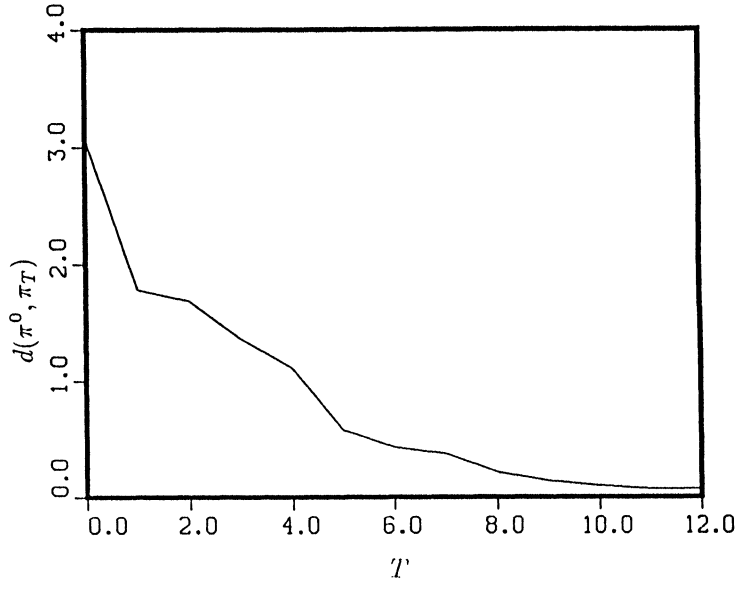


Figure 6.2: The distance between the computed parameter π_T and the true parameter π^0, $d(\pi^0, \pi_T)$, as a function of observation time T.

Remark 3: Since it is well known that the probability law of any Itô process is determine by $\sigma\sigma^*$ rather than σ itself, it is not possible to uniquely identify σ or σ_0. Therefore the results shown in the table are those for $\sigma\sigma^*$ and $\sigma_0\sigma_0^*$. In the table, s_{ij} are the entries of the matrix $S = \sigma\sigma^*$.

7. Summary and Conclusion

We have presented a formulation of the identification problem for partially observed linear stochastic systems as a deterministic control problem. For this purpose, an appropriate and also natural objective functional has been introduced for the first time in the literature. Using this method, we successfully identified the system parameter π simultaneously, as shown in section 6.

References

[1] Ahmed, N.U., *Elements of Finite Dimensional Systems and Control Theory*, Longman Scientific and Technical, U.K.; Co-publisher: John Wiley, New York 1988.

[2] Bagchi, A., Continuous time systems identification with unknown noise covariance, *Automatica* **11** (1975), 533-536.

[3] Dabbous, T.E. and Ahmed, N.U., Parameter identification for partially observed diffusion, *J. of Opt. Theory and Appl.* **75**:1 (1992), 33-50.

[4] Gupta, N.K. and Mehra, R.K., Computational aspects of maximum likelihood estimation and and reduction in sensitivity function calculations, *IEEE Trans. on Autom. Control* **AC-19**:6 (1974), 774-783.

[5] Kwakernaak, H. and Sivan, R., *Linear Optimal Control Systems*, John Wiley, New York 1972.

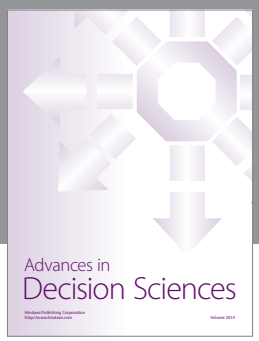
[6] Larson, H.J. and Shubert, B.O., *Probabilistic Models in Engineering Sciences*, Vol. II, John Wiley, New York 1979.

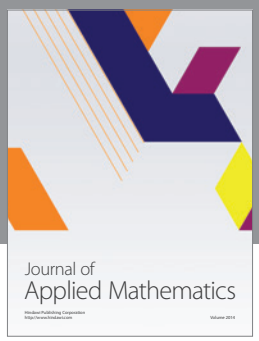
[7] Legland, F., Nonlinear filtering and problem of parametric estimation, In: *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, (ed. by M. Hazewinkel and J. Willems), D. Reidel Publishing Col., Boston, MA (1980), 613-620.

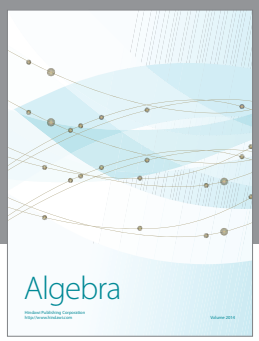
[8] Liptser, R.S. and Shiryayev, A.N., *Statistics of Random Processes*, Vols. I and II, Springer-Verlag, Berlin 1978.

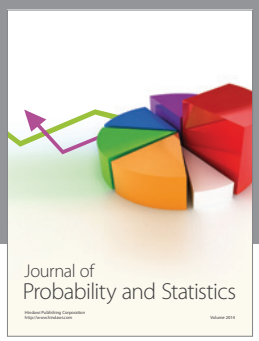
[9] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**:6 (1953), 1087-1092.

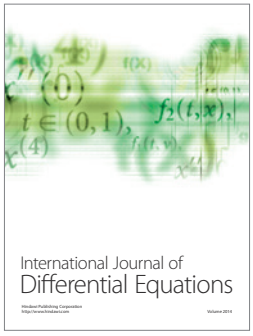
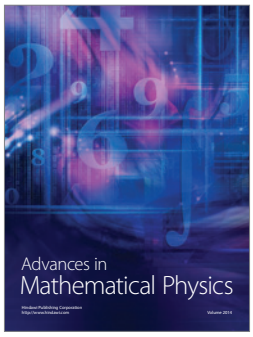
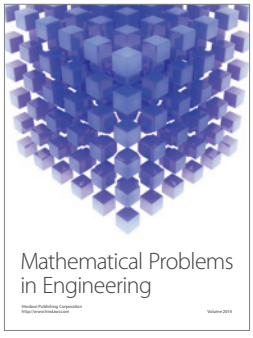
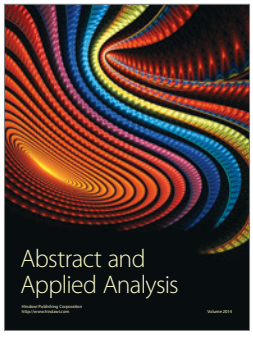
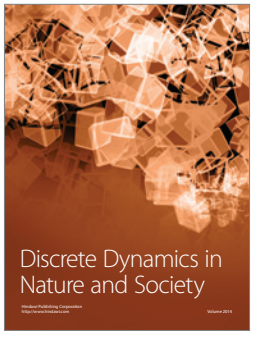
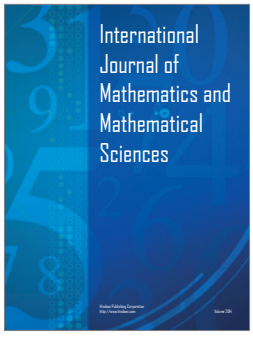
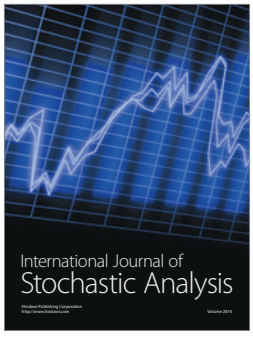
[10] Tungait, J.K., Continuous-time system identification on compact parameter sets, *IEEE Trans. on Info. Theory* **IT-31** (1985), 652-659.









Hindawi

Submit your manuscripts at
http://www.hindawi.com