AN ABSTRACT OF THE THESIS OF

Wojciech Jan Kolodziej for the degree of Doctor of Philosophy

in Electrical and Computer Engineering presented on June 2, 1980

Title: CONDITIONALLY GAUSSIAN PROCESSES IN STOCHASTIC CONTROL THEORY

Abstract approved: Redacted for Privacy

R. R. Mohler

An application of the theory of conditionally Gaussian random processes to filtering and stochastic control problems is presented here. The results due to Liptser and Shiryayev are proved to hold in the multidimensional case under somewhat relaxed conditions, when compared to the original ones. Such a generalization is required from the point of view of modelling real engineering systems. The concept of a weak solution to the stochastic differential equations involved in the problem formulation is used. A detailed filter derivation for conditionally Gaussian multidimensional processes is presented. In this derivation both conditionally Gaussian processes and nonlinear filtering theories are used. A finite dimensional, recursive formula (filter) for calculating the optimal mean-square estimate of the unobservable part of the process is obtained.

An application of the derived filter to an optimal stochastic control problem is presented. The class of systems under consideration includes linear, partially observable control systems with quadratic criteria, that have random coefficients which are certain functionals of a Wiener process. All stochastic processes involved in the problem formulation are assumed to be strong solutions to the corresponding stochastic differential equations. Separation of filtering and control is shown to hold, and the optimal regulator is a function of both the observable part and the estimate of the unobservable part of the process. Sufficient conditions for an optimal control to exist are expressed through the existence of a solution to a certain Cauchy problem of the parabolic type partial differential equation. The existence and uniqueness of a solution to the above mentioned partial differential equation is studied. The references to the results used in the text are given. A simple simulation example, which gives an illustration of the obtained results, is also presented.

CONDITIONALLY GAUSSIAN PROCESSES IN STOCHASTIC CONTROL THEORY

by

Wojciech Jan Kolodziej

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1980

APPROVED:

Redacted for Privacy

Professor of Electrical and Computer Engineering
in charge of major

Redacted for Privacy

Head of Department of Electrical and Computer Engineering

Redacted for Privacy

Dean of Graduate School

Date thesis is presented June 2, 1980

Typed by Lyndalu Sikes for Wojciech Jan Kolodziej

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my major professor, Dr. R. R. Mohler. His friendship, guidance and patience over the past several years have contributed to the successful completion of this study.

This research was supported by NSF Grant No. ENG 77-07027.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CONDITIONALLY GAUSSIAN PROCESSES IN STOCHASTIC CONTROL THEORY

1. INTRODUCTION

Systems arising in applications are often modelled by differential equations:

$$\frac{dx(t)}{dt} = f(t, x(t)),$$

where the vector function x(t) describes the system state at time t. However, many systems are subject to imperfectly known disturbances, which may be taken as random. A stochastic model is now appropriate in which the system states evolve according to some vector-valued stochastic process x_t. Such a stochastic process may be interpreted as a "solution" to a system of ordinary differential equations containing a term $g(t, x_t)v_t$ representing the effect of disturbances

$$\frac{dx_t}{dt} = f(t, x_t) + g(t, x_t)\, v_t. \tag{1.1}$$

v_t is very often meant to be a stationary Gaussian process with a spectral density that is flat over a very wide range of frequencies. If v_t has well-behaved sample functions, there is no difficulty in interpreting (1.1) as an ordinary differential equation for each sample function. However, if v_t is taken to be a process with well-behaved sample functions some of the simple statistical properties of x_t, the primary one being the Markov property, are lost.

In practice the following interpretation of (1.1) can be given. Take a sequence of Gaussian processes (v_t^n) which "converges" in some suitable sense to a white Gaussian noise, yet for each n, v_t^n has well-behaved sample functions. Now for each n the equation

$$\frac{dx_t^n}{dt} = f(t, x_t^n) + g(t, x_t^n)\, v_t^n, \quad t \in [0, T],$$

can be solved. Thus a sequence of processes (x_t^n), $t \in [0, T]$, is obtained. Suppose now that as $n \to \infty$, (v_t^n) converges in a suitable sense to white noise, and the sequence (x_t^n), $t \in [0, T]$, converges, say in quadratic mean, to a process x_t, $t \in [0, T]$. Then x_t is said to be a solution to (1.1), where v_t is a Gaussian white noise.

In order to make precise the notion of v_t^n converging to white noise define

$$w_t^n = \int_0^t v_s^n \, ds$$

Since Gaussian white noise is the formal derivative of Brownian motion, the convergence of v_t^n to Gaussian white noise will be understood as convergence of w_t^n to the Wiener process w_t. w_t was suggested by Wiener as a mathematical model for the motion of particles suspended in a fluid. The nonexistence of $\frac{dw_t}{dt}$ implies, in Wiener's model of physical Brownian motion, that particles do not have well defined velocities. This corresponds roughly to physical observation. In engineering literature, the formal time derivative of Brownian motion is called white noise. Proceeding in a purely formal way $v_t = \frac{dw_t}{dt}$, $t \in [0, T]$, can be regarded as a stationary process in which the random variables v_t are independent for different time instants t, with $E(v_t) = 0$. The covariance function $E(v_t v_s)$ turns out to be Dirac's delta function and its Fourier transform, the spectral density, is constant. Thus the average power with which various frequencies appear in the spectral resolution is constant (compare with the spectrum of a "white" light).

Now (1.1) can be rewritten as follows:

$$x_t = x_0 + \int_0^t f(s, x_s)ds + \int_0^t g(s, x_s)dw_s, \quad t \in [0, T], \tag{1.2}$$

where

x_0 is an initial condition for x_t.

The last integral in (1.2) is interpreted as a stochastic integral and needs to be defined. Since w_t has realizations of unbounded variation, the stochastic integral cannot be defined in the usual Lebesgue-Stieltjes sense. The one generally accepted definition is due to Ito and is often referred to as the Ito integral. It turns out that a calculus based on Ito's definition is not compatible with results of ordinary differential calculus. Other definitions for stochastic integrals have been proposed (Fisk, 1963; Stratonovich, 1966), for which rules of ordinary calculus apply. However, this approach has the disadvantage that conditions which guarantee the convergence of Fisk-Stratonovich's integral are less natural and more difficult to verify than those of the stochastic Ito integral. But first of all the martingale property (see Def. A3) of Ito's integral could be lost. As the most important results in this text are based on this property, it plays a crucial role. From now on Ito's definition of the stochastic integral will be used.

Very often (1.2) is written in a symbolic, "differential", form

$$dx_t = f(t, x_t)\, dt + g(t, x_t)\, dw_t, \tag{1.3}$$

and is called a stochastic differential equation.

A process x_t, $t \in [0, T]$, is said to satisfy (1.3) with initial condition x_0 if

i. for each $t \in [0, T]$, $\int_0^t g(s, x_s)\, dw_s$ can be interpreted as the stochastic integral,

ii. for each $t \in [0, T]$, x_t is almost surely equal to

$$x_0 + \int_0^t f(s, x_s)\, ds + \int_0^t g(s, x_s)\, dw_s .$$

For a more formal definition of a solution to a stochastic differential equation see Def. A1. The character of this text as well as the limitation of space, makes it very difficult to give even a flavor of the subject of stochastic differential equations. [W4] may serve as an excellent introduction to this topic, while [G1] presents a rather formal and advanced approach.

Most of the results in stochastic control and filtering theory were obtained with the assumption that the processes under consideration satisfy linear stochastic differential equations. With certain assumptions about initial conditions and the structure of the linear stochastic equation the resulting process is Gaussian. This is crucial for the property of finite-dimensionality of filter equations.

Motivated by application, nonlinear stochastic differential equations play a very important role. This is seen particularly in control problems, where even a linear system becomes very often nonlinear after the feedback regulator is applied.

Conditionally Gaussian processes defined below, and satisfying equations of the type (1.4) seem to be natural generalizations of linear stochastic systems. They offer considerably more flexibility in control applications than linear models and yet enjoy the property of a finite dimensional filter (see discussion below).

This dissertation presents an application of the theory of conditionally Gaussian random processes to filtering and stochastic control problems. The original study of this subject is due to Liptser and Shiryayev [L1, L2, L3, L4].

To summarize their result consider a partially observable random process (x_t, y_t), $t \geq 0$, with only the second component y_t, $t \geq 0$, being observed. At any time t it is required to estimate x_t based on $(y_s : 0 \leq s \leq t)$.

It is a well known fact that if $E(x_t^2) < \infty$, $t \geq 0$, then the problem of finding the optimal mean square estimate m_t of x_t from $(y_s : 0 \leq s \leq t)$ is reduced to finding the conditional expectation $m_t = E(x_t \mid \mathcal{Y}_t)$, where $\mathcal{Y}_t$ is the σ-algebra generated by the observations $(y_s : 0 \leq s \leq t)$.

Assume that (x_t, y_t), $t \geq 0$, have an Ito differential:

$$
\begin{aligned}
dx_t &= (a(t, y)x_t + b(t, y))\, dt + g_1(t, y)dw_t^1 + g_2(t, y)dw_t^2, \\
dy_t &= (c(t, y)x_t + d(t, y))dt + r_1(t, y)dw_t^1 + r_2(t, y)dw_t^2,
\end{aligned}
\tag{1.4}
$$

where each of the functionals $a(t, y), \ldots, r_2(t, y)$ is $\mathcal{Y}_t$ measurable at any $t \geq 0$. The Wiener processes w_t^1 and w_t^2, and the random variables (x_0, y_0) which are an initial condition for (1.4) are assumed to be independent. It should be noticed that the unobservable process x_t enters into (1.4) in a linear way. It can be proved under certain assumptions that if the conditional distribution $(P(x_0 \leq \alpha \mid y_0)$, $\alpha \in R$, is Gaussian then the process (x_t, y_t), $t \geq 0$, satisfying (1.4) is conditionally Gaussian in the sense that for any $t \geq 0$ the conditional distributions

$$P(x_{t_0} \leq \alpha_0, \ldots, x_{t_k} \leq \alpha_k \mid \mathcal{Y}_t), \quad 0 \leq t_0 < t_1 < \ldots < t_k \leq t ,$$

are Gaussian [L2].

This result allows a closed system of equations for $m_t = E(x_t \mid \mathcal{Y}_t)$ and $\Gamma_t = E((x_t - m_t)^2 \mid \mathcal{Y}_t)$ which completely characterize the distribution $P(x_t \leq \alpha \mid \mathcal{Y}_t)$, $t \geq 0$. These equations are a recursive formula (filter) for updating the estimator m_t every time a new observation is made. The result of [L2] can be considered as a generalization of Kalman-Bucy filtering theory for Gaussian stochastic processes satisfying linear stochastic equations.

As the modelling of engineering systems very often involves vectorial processes, Chapter 2 of this thesis presents a detailed derivation of an optimal filter for the multidimensional case of (1.4). Assumptions made here are somewhat less restrictive than those of [L2]. For example, conditions $|a(t,\xi)| \leq \text{const} < \infty$, $|c(t,\xi)| \leq \text{const} < \infty$ for all ξ which are continuous functions on $[0, T]$ are replaced by the conditions

$$\int_0^T a^2(t,\xi)dt \leq \text{const} < \infty , \text{ and } \int_0^T c^4(t,\xi)dt < \infty .$$

The concept of a weak solution to (1.4) (see Def. A1 and A2) is used leaving open the question of a physical interpretation of (1.4). However, if necessary, with additional assumption about the structure of (1.4) (for example Lipschitz conditions in ξ satisfied by $a(t, \xi), \ldots, r_2(t, \xi)$) a solution to (1.4) is a strong one and all results obtained here hold. Chapter 3 presents an application of the above filtering result to an optimal stochastic control problem. The class of systems under consideration includes linear, partially observable control systems with quadratic criteria, having random coefficients which are certain functionals of a Wiener process.

The class of linear stochastic systems with random coefficients has attracted the attention of several authors [A1, B3, B4]. One of the most formal approaches to the stochastic control of such systems was presented by Bismut [B2]. The linear stochastic differential equation considered here is

$$dx_t = (Ax_t + Bu_t + D)dt + (Hx_t + Gu_t + F)dw_t ,$$

where w_t is a Brownian motion and all the random coefficients are supposed to be observable by the controller u_t. These random coefficients are certain bounded random variables adapted to the same σ-algebra as w_t, $t \in [0, T]$.

For such completely observable systems a problem of minimizing the criteria

$$I(u) = \left(\int_0^T |M_t x_t|^2 dt + \int_0^T < N_t u_t, u_t > dt + |M_t x_t|^2\right),$$

where N_t, $t \in [0, T]$, is a family of self-adjoint positive operators and where again all the coefficients are observable by the controler, is discussed. Functional analysis techniques were used to prove that I(u) is differentiable. The necessary condition for u_t to be optimal was expressed through dual variables. The optimal control was found in a random feedback form linear in x_t. The formal Riccati equation determining the gain coefficient for the optimal regulator was not proved to have a solution in a general case and only special cases, which cover the results presented in [W2] are given here.

Methods used in Chapter 3 are of the dynamic stochastic programming type. All stochastic processes involved in the problem formulation are assumed to be the strong solutions to the corresponding stochastic differential equations. The main theorem shows a separation of the filtering and control problem as the optimal regulator is a (linear) function of the estimate of the unobservable part of the process and a (nonlinear) function of the observable part. This result can be compared with results presented in [W3, D3, D4]. Sufficient conditions for an optimal control are expressed through the existence of a bounded solution to a certain Cauchy problem for parabolic type of partial differential equations.

In Chapter 4 the existence and uniqueness of a solution to the above mentioned partial differential equation is studied.

Certain similarities between this equation and the Riccati equation resulting from a solution to the stochastic linear-quadratic control problem [W1] are used in the proof construction. Results developed for linear and semi-linear partial differential equations, [L5], are used here.

As an interesting by-product [Lemma 4.3], positive definiteness of a solution to a certain Cauchy problem for linear partial differential equations is proved.

Appendix A is thought to be an easy reference to the main results used in this text.

Chapter 5 presents a simple digital simulation of an optimal stochastic control system, for which the regulator was obtained by using the result of Chapters 3 and 4. Concluding remarks point out still open questions and possible research directions for applications of the obtained results.

Notation

The following notation will be used throughout:

R^n denotes Euclidean n-dimensional space;

C_T^m denotes the space of all continuous m-dimensional functions on $[0, T]$;

$*$ denotes transposition of a vector or a matrix;

tr denotes trace of a matrix;

$\|A\|$ the Euclidean norm is defined for A being a vector or a matrix as follows

$$\|A\|^2 = tr(A\, A^*);$$

$R > Q$ says that R-Q is a positive definite matrix, assuming that R, Q are square, symmetric matrices; similarly $\geq$; for a square, symmetric matrix A the following notation is used

$$A^2 = AA, \; A^{-2} = (A^{-1})(A^{-1}), \; A^{-1} = (A^{-1/2})(A^{-1/2}) ,$$

where in the second and the third of the above equalities A^{-1} is assumed to exist;

$A^\dagger$ denotes the pseudoinverse of a matrix A;

$[A]_{ij}$ denotes i, j th element of a matrix A;

$[a]_i$ denotes ith element of a vector a;

$a \geq b$ for $a, b \in R^n$ is understood to hold for the corresponding components of a and b;

c, c_i stand for positive constants;

$t_1 \wedge t_2$ denotes $\min(t_1, t_2)$;

$\frac{\partial}{\partial \xi}$ denotes the gradient vector;

$\frac{\partial}{\partial \xi \, \partial \xi^*}$ denotes the Jacobian matrix;

Random variables or stochastic processes are tacitly referred to an underlying probability space (Ω, F, P) and the generic argument $\omega \in \Omega$ will not be written;

$E(\cdot)$ stands for the expectation;

$E(\cdot \mid \cdot)$ stands for the conditional expectation;

σ - alg $(y_s: 0 \leq s \leq t)$ denotes the σ - algebra generated by the process y up to time t;

$\mu_1 \sim \mu_2$ denotes the equivalence of the (probability) measures μ_1 and μ_2;

$\mu_1 << \mu_2$ denotes the absolute continuity of the measure μ_1 with respect to the measure μ_2;

$\frac{d\mu_1}{d\mu_2}$ denotes, for $\mu_1 << \mu_2$, the Radon-Nikodym derivative [W4] of μ_1 with respect to μ_2;

2. FILTERING FOR MULTIDIMENSIONAL CONDITIONALLY GAUSSIAN PROCESSES

In this chapter the derivation of the optimal, in the mean-square sense, filter for multidimensional conditionally-Gaussian processes is presented. The filtering equations are obtained under somewhat less-constraining assumptions than those used in [L2]. A detailed proof shows the places in which successive assumptions are used. This gives an opportunity for the study of further relaxation of the assumptions made.

Let $(x_t, y_t), t \in [0, T]$, be a continuous (P-a. s.) stochastic process of the following differential representation,

$$dx_t = A(t, y)x_t dt + B(t, y)dt + G_1(t, y)dw_t^1 + G_2(t, y)dw_t^2, \qquad (2.1)$$

$$dy_t = C(t, y)x_t dt + D(t, y)dt + R_1(t, y)dw_t^1 + R_2(t, y)dw_t^2. \qquad (2.2)$$

Consider as given some complete probability space $(\Omega, \mathcal{F}, P)$ with a nondecreasing, right-continuous family of sub-σ-algebras $\mathcal{F}_t \subseteq \mathcal{F}$, $t \in [0,T]$. Let $(w_t^i, \mathcal{F}_t)$, $i = 1, 2$, be mutually independent Wiener processes of the dimensions ℓ_i, $i = 1, 2$ respectively. The stochastic process x_t, $t \in [0, T]$, of the dimension n is the unobservable component of (x_t, y_t), while y_t, $t \in [0, T]$ is the m-dimensional stochastic process that is observed. The random variables x_0 and y_0 which form the initial conditions for (2.1) and (2.2) are assumed to be independent of the Wiener processes w_t^i, $i = 1, 2$. The matrices $A(t,\xi)$, $B(t,\xi)$, $G_1(t,\xi)$, $G_2(t,\xi)$, $C(t,\xi)$, $D(t,\xi)$, $R_1(t,\xi)$, $R_2(t,\xi)$ are of the dimensions $n \times n$, $n \times 1$, $n \times \ell_1$, $n \times \ell_2$, $m \times n$, $m \times 1$, $m \times \ell_1$, $m \times \ell_2$ respectively and their elements are assumed to be measurable nonanticipative functionals on $[0,T] \times C_T^m$.

Listed below are sufficient conditions for the derivation of a recursive formula (filter) for an optimal, in the mean-square sense, estimator of x_t given $(y_s: 0 \leq s \leq t)$.

For all $\xi \in C_T^m$

$$\int_0^T \|A(t,\xi)\|^2 \, dt \leq c < \infty \text{ , and}$$

$$\int_0^T (\|B(t,\xi)\|^4 + \|C(t,\xi)\|^4 + \|D(t,\xi)\|^2 + \|G_1(t,\xi)\|^4 + \|G_2(t,\xi)\|^4) dt < \infty. \tag{2.3}$$

For all $\xi, \eta \in C_T^m$, $t \in [0,T]$ define

$$R^2(t,\xi) = R_1(t,\xi)\, R_1^*(t,\xi) + R_2(t,\xi)\, R_2^*(t,\xi)$$

then

$$\| R^{-2}(t,\xi) \| \leq c < \infty \text{ ,} \tag{2.4}$$

$$\|R(t,\xi)-R(t,\eta)\|^2 \leq c\left(\int_0^t \|\xi(s)-\eta(s)\|^2 dK(s) + \|\xi(t)-\eta(t)\|^2\right), \text{ and}$$

$$\|R(t,\xi)\|^2 \leq c\left(\int_0^t (1 + \|\xi(s)\|^2) dK(s) + 1 + \|\xi(t)\|^2\right), \tag{2.5}$$

where $0 \leq K(s) \leq 1$ is a right-continuous, nondecreasing function on $[0, T]$, and c is a positive constant.

The condition (2.3) is somewhat relaxed when compared to the conditions given in [L2], where it was assumed that

$$\|A(t,\xi)\| \leq c < \infty, \quad \|C(t,\xi)\| \leq c < \infty, \text{ for all } t \in [0,T], \xi \in C_T^m .$$

The condition (2.3) also implies that

$$\int_0^T \| B(t,\xi) \|^4 dt < \infty .$$

For a stochastic control problem based on the equation (2.1) the above will restrict the additive controls to the class which satisfies

$$E(\int_0^T \| u_t \|^4 dt) < \infty . \text{ (see Chapter 3).}$$

This is a stronger condition that of the square integrability, which is generally assumed in control theory.

Condition (2.4) is made, roughly speaking, to avoid degenerated stochastic measures associated with y. One may think here about the situation where some of the equations of (2.2) do not have the noise terms. In [D3] an approximation of a degenerate system of stochastic equations is discussed. This approximation satisfies condition (2.4). It is worth noticing that one of the most important results, Thm A3 and Thm A4, used to prove that x_t is conditionally

Gaussian, do not require (2.4) to hold.

Condition (2.5) restricts the noise coefficients R_1, R_2 in (2.2) to the class of "smooth" functionals of ξ. Also the rate of growth of these coefficients is limited to, at most, linear growth in ξ. Any relaxation of the condition (2.5) will lead to the problem of finding another sufficient condition, under which equation (2.16) has a strong solution. This is necessary to show an equivalence of certain probability measures and to give a well defined Bayes formula (see (2.38)).

Now the following theorem can be stated

Theorem 2.1

Let (2.1) and (2.2) have a weak solution (x_t, y_t), $t \in [0,T]$, with the initial conditions (x_0, y_0) which satisfy

$$\mathcal{E}(\| x_0 \|^4) < \infty, \quad P(\| y_0 \| < \infty) = 1,$$

where $\mathcal{E}(\cdot)$ denotes the expectation operator.

Also let x_0, y_0 be independent of w_t^i, $i = 1, 2$ and let the conditional distribution $P(x_0 \leq \alpha_0 \mid y_0)$ be P-a.s. Gaussian with the parameters $m_0 = \mathcal{E}(x_0 | y_0)$, $\Gamma_0 = \mathcal{E}((x_0 - m_0)(x_0 - m_0)^* | y_0)$ and $\mathrm{tr}(\Gamma_0) < \infty$ P-a.s. In the above the inequality $x_0 \leq \alpha_0$, $\alpha_0 \in R^n$ is understood to hold for the corresponding components of x_0 and α_0.

If the conditions (2.3), (2.4) and (2.5) are satisfied then the processes (x_t, y_t), $t\in[0,T]$ are conditionally Gaussian, i.e., for any $t\in[0,T]$ and any finite partition t_j, $j = 0, 1, \ldots, k$, of $[0,T]$ such that $0 \le t_0 < t_1 < \ldots < t_k \le t$, the conditional distribution $P(x_{t_0} \le \alpha_0, \ldots, x_{t_k} \le \alpha_k \mid \mathcal{Y}_t)$, $\alpha_j \in \mathcal{R}^n$ is P-a.s. Gaussian. $\mathcal{Y}_t$, $t\in[0,T]$ denotes the σ - algebra generated by $(y_s : 0 \le s \le t)$.

Further m_t, Γ_t, which denote conditional mean and covariance respectively, i.e., $m_t = \mathcal{E}(x_t \mid \mathcal{Y}_t)$ and $\Gamma_t = \mathcal{E}((x_t - m_t)(x_t - m_t)^* \mid \mathcal{Y}_t)$, are unique, continuous solutions with initial condition m_0, Γ_0 to

$$dm_t = (A\, m_t + B)\, dt + K d\nu_t\ , \tag{2.6}$$

$$K = (G_1 R_1 + G_2 R_2 + \Gamma_t C^*)\, R^{-1}\ ,$$

$$d\nu_t = R^{-1}(dy_t - (Cm_t + D)\, dt)\ ,$$

$$d\Gamma_t = (A\,\Gamma_t + \Gamma_t\, A^* + G_1 G_1{}^* + G_2 G_2{}^* - KK^*)\, dt\ , \tag{2.7}$$

where arguments (t, y) are omitted for brevity.

Proof of Thm 2.1

The above filtering result was proved in [L2] under somewhat more restrictive conditions than (2.3), (2.4) and (2.5). The scalar case was discussed and only a few steps of the proof for the multi-dimensional case were given. That x_t given $\mathcal{Y}_t$ is conditionally

Gaussian can be shown to hold with less restrictive conditions than those of Thm 2.1. However, as the filter itself is the main result of this chapter the whole proof uses the same assumptions. The proof follows the main steps of the scalar case proof in [L2] and appropriate references to pages in [L1] and [L2] are provided. Also all results which are used here in the form of theorems and lemmas are given in the Appendix A together with detailed references to the original sources.

The outline of the proof is as follows:

Step 1. It is shown, Lemma 2.1a, that (2.1), (2.2) can be written in the equivalent form (2.8), (2.9).

Step 2. A generalized Bayes formula is used to show that the conditional characteristic function of x_t, given $\mathcal{Y}_t$, is a characteristic function of a Gaussian process. To this end Theorems A3, A4, and Lemma A3 are used. To show that all assumptions necessary for the application of these results are satisfied, an auxiliary result, Lemma 2.1b, is proved. After rather cumbersome notational changes, Lemma A7 is used to obtain the desired result.

Step 3. In order to apply the results of nonlinear filtering theory, a certain representation (see (2.45)) of the equation (2.9) is proved to hold. This is made by using Thm A5.

Next Lemmas 2.1c and 2.1d allow for a certain representations of $E(x_t | \mathcal{Y}_t)$ and $E(x_t x_t^* | \mathcal{Y}_t)$ (see (2.60) and (2.61)).

Step 4. The property that x_t is conditionally Gaussian is used to derive the final form of the filter equations (2.6) and (2.7).

Step 5. The uniqueness of a continuous solution to (2.6) and (2.7) is proved, as well as the property that if $\Gamma_0 > 0$, P-a.s., then Γ_t is, P-a.s., uniformly positive definite on $[0, T]$.

As the first step the equations (2.1) and (2.2) are transformed to the form

$$dx_t = A(t, y)x_t dt + B(t, y)dt + H_1(t, y)d\tilde{w}_t^1 + H_2(t, y)d\tilde{w}_t^2, \tag{2.8}$$

$$dy_t = C(t, y)x_t dt + D(t, y)dt + R(t,y)d\tilde{w}_t^2 , \tag{2.9}$$

where $H_2 = (G_1 R_1^* + G_2 R_2^*) R^{-1}$,

$H_1 = (G_1 G_1^* + G_2 G_2^* - H_2 H_2^*)^{1/2}$,

and $(\tilde{w}_t^i, \mathcal{F}_t)$, $i = 1, 2$, are mutually independent Wiener processes of the dimensionality n and m respectively.

Lemma 2.1a

The equations (2.8) and (2.9) are an equivalent form of the equations (2.1) and (2.2).

Proof of Lemma 2.1a

It is enough to show that P-a.s. for all $t\in[0, T]$

$$\int_0^t G_1(s, y)dw_s^1 + \int_0^t G_2(s, y)dw_s^2 = \int_0^t H_1(s, y)d\tilde{w}_s^1 + \int_0^t H_2(s, y)d\tilde{w}_s^2, \tag{2.10}$$

and

$$\int_0^t R_1(s, y)dw_s^1 + \int_0^t R_2(s, y)dw_s^2 = \int_0^t R(s, y)d\tilde{w}_s^2 . \tag{2.11}$$

The notation used here is as follows (arguments are omitted for brevity)

$$P = \begin{bmatrix} G_1 & G_2 \\ R_1 & R_2 \end{bmatrix}, \qquad S = \begin{bmatrix} H_1 & H_2 \\ \emptyset & R \end{bmatrix}.$$

Matrices P and S are of the dimension $(n + m) \times (\ell_1 + \ell_2)$ and $(n + m) \times (n + m)$ respectively. $\emptyset$ denotes an $m \times n$ matrix of zero elements. By definition of R, H_1, H_2, P and S satisfy

$$PP^* = SS^*. \tag{2.12}$$

Denote
$$w_t = \begin{bmatrix} w_t^1 \\ w_t^2 \end{bmatrix} \quad \text{and} \quad \tilde{w}_t = \begin{bmatrix} \tilde{w}_t^1 \\ \tilde{w}_t^2 \end{bmatrix}$$

Now the equations (2.10) and (2.11) can be rewritten in the more compact form

$$\int_0^t P dw_s = \int_0^t S d\tilde{w}_s . \tag{2.13}$$

As the properties of the pseudoinverse $S^\dagger$ of the matrix S will be used it is recalled that by definition $S^\dagger$ satisfies

$$SS^\dagger S = S,$$

and $S^\dagger = US^* = S^*V$ for some matrices U and V.

$S^\dagger$ always exists and is unique [L2, p. 51]. If the inverse S^{-1} of S exists then $S^{-1} = S^\dagger$.

Let (v_t, F_t) be an n + m dimensional Wiener process independent of w_t^i, i = 1, 2. Then define

$$\tilde{w}_t = \int_0^t S^\dagger P \, dw_t + \int_0^t (I - S^\dagger S) \, dv_s , \tag{2.14}$$

where I is an (n + m) x (n + m) unit matrix. Using Levy's theorem (Theorem A1) it will be shown that (2.14) defines an n + m dimensional Wiener process with independent components. By (2.14), (2.3) and (2.4) $\tilde{w}_t$ is a square integrable martingale with continuous trajectories and

$$P(\tilde{w}_0 = 0) = 1 .$$

It remains to show that for $t \geq s$ P-a. s.

$$E((\tilde{w}_t - \tilde{w}_s)(\tilde{w}_t - \tilde{w}_s)^* | F_s) = (t - s) \cdot I \tag{2.15}$$

The left hand side of (2.15) is equal to

$$\mathcal{E}\Big(\int_s^t S^\dagger PP^*(S^\dagger)^* d\tau \,\Big|\, F_s\Big) + \mathcal{E}\Big(\int_s^t (I - S^\dagger S)(I - S^\dagger S)^* d\tau \,\Big|\, F_s\Big).$$

From (2.12) and the properties of $S^\dagger$ it follows that

$$S^\dagger PP^* (S^\dagger)^* = S^\dagger SS^*(S^\dagger)^* = S^\dagger S (S^\dagger S)^* = (S^\dagger S)^2 = S^\dagger S ,$$

and

$$(I - S^\dagger S)(I - S^\dagger S)^* = I - S^\dagger S - (S^\dagger S)^* + S^\dagger S (S^\dagger S)^* =$$

$$I - S^\dagger S - S^\dagger S + S^\dagger S = I - S^\dagger S .$$

The above equalities prove (2.15).

To show that (2.13) holds note that

$$\int_0^t S d\tilde{w}_s = \int_0^t SS^\dagger P \, dw_s + \int_0^t S(I - S^\dagger S) \, dv_s =$$

$$\int_0^t SS^\dagger P \, dw_s = \int_0^t P \, dw_s + \lambda_t ,$$

where
$$\lambda_t = \int_0^t (SS^\dagger - I) P \, dw_s .$$

But $\mathcal{E}(\lambda_t \lambda_t^*) = \mathcal{E}\Big(\int_0^t (SS^\dagger - I) PP^* (SS^\dagger - I)^* ds\Big) =$

$$\mathcal{E}\Big(\int_0^t (SS^\dagger - I)SS^*(SS^\dagger - I)^* ds\Big) = \mathcal{E}\Big(\int_0^t (SS^\dagger S - S)(SS^\dagger S - S)^* ds\Big) = 0.$$

This ends the proof of Lemma 2.1a.

In the sequel the Wiener processes in (2.8) and (2.9) will be written without tildes.

Next it will be shown that the process $\tilde{y}_t$, $t \in [0, T]$, defined by (2.16) below, generates the measure $\mu_{\tilde{y}}$ which is equivalent to the measure μ_y generated by y_t, $t \in [0, T]$. Moreover the measures μ and $\tilde{\mu}$ generated respectively by (x_0, w^1, y) and $(x_0, w^1, \tilde{y})$ are also shown to be equivalent. To this end Thm A2 is used to show that the assumption (2.5) and the condition $P(\|y_0\| < \infty) = 1$ give a unique continuous solution to

$$d\tilde{y}_t = R(t, \tilde{y})\, dw_t^2 \quad , \quad \tilde{y}_0 = y_0 \, . \tag{2.16}$$

Let z_t, $t \in [0, T]$, denote any of the processes y_t or $\tilde{y}_t$. Then equation (2.8), with y replaced by z can be rewritten in the following form:

$$\tilde{x}_t = x_0 + \int_0^t (\tilde{A}(s, z)\tilde{x}_s + \tilde{B}(s, z))ds + \int_0^t H_1(s, z)dw_s^1 + \int_0^t H_2(s, z)R^{-1}(s, z)dz_s \, , \tag{2.17}$$

where $\tilde{A} = A - H_2 R^{-1} C$, and

$\tilde{B} = B - H_2 R^{-1} C$.

Also note that

$$\tilde{x}_t = \Phi_t(z)(x_0 + \int_0^t \Phi_s^{-1}(z)\tilde{B}(s,z)ds + \int_0^t \Phi_s^{-1}(z)H_1(s,z)dw_s^1 +$$

$$\int_0^t \Phi_s^{-1}(z)H_2(s,z)R^{-1}(s,z)dz_s), \tag{2.18}$$

where $\Phi_t(z) = \exp\left(\int_0^t \tilde{A}(s,z)ds\right)$.

All the integrals in (2.17) and (2.18) are well defined as a consequence of (2.3) and (2.4). Using the Ito formula, it is easy to check that $\tilde{x}_t$ given by (2.18) satisfies (2.17).

To show the uniqueness of a solution to (2.17), Lemma A1 will be applied. Denote by e_t the difference between two continuous solutions to (2.17). From (2.17) it follows that

$$e_t = \int_0^t \tilde{A}(s,z)\, e_s\, ds, \quad t \in [0,T] .$$

To prove uniqueness, it is enough to show that

$$P\left(\sup_{0 \le t \le T} \| e_t \| > 0\right) = 0 \tag{2.19}$$

The equation for e_t results in the following inequality:

$$\| e_t \| \le \sqrt{n} \int_0^t \| \tilde{A}(s,z) \| \; \| e_s \| ds, \quad t \in [0,T].$$

To apply Lemma A1 it must be shown that P-a. s.

$$\int_0^T \|\tilde{A}\| \, dt < \infty \ .$$

The above results from the following inequalities and the conditions (2.3) and (2.4):

$$\int_0^T \|\tilde{A}\| dt \le \int_0^T \|A\| dt + \int_0^T \|R^{-1}\| \ \|H_2\| \ \|C\| dt \le$$

$$\int_0^T \|A\| dt + 0.5 \int_0^T \|R^{-1}\| (\|H_2\|^2 + \|C\|^2) dt \ ,$$

$$\|H_2\|^2 = tr((G_1R_1^* + G_2R_2^*)R^{-2}(R_1G_1^* + R_2G_2^*)) \le \|G_1\|^2 + \|G_2\|^2 \ .$$

Now application of Lemma A1 results in

$$P(\|e_t\| > 0) = 0, \quad t \in [0, T],$$

which together with continuity of e_t gives (2.19).

By Lemma A2 there exists a representation of a solution x_t to (2.17) with $z_t = y_t$, such that for almost all $t \in [0, T]$

$$x_t = f_t\,(x_0, \ w^1, \ y),$$

where f_t denotes a measurable functional on $[0, T] \times R^n \times C_T^n \times C_T^m$.

To show that μ is equivalent to $\tilde{\mu}$ ($\mu \sim \tilde{\mu}$), Theorems A3, A4 and Lemma A3 will be used.

To check that all the assumptions of the theorems A3 and A4 are satisfied the following Lemma will be proved.

Lemma 2.1.b. Under the assumptions of Theorem 2.1.

$$E(\sup_{0\leq t\leq T} \| x_t \|^4) < \infty . \tag{2.20}$$

Proof of Lemma 2.1.b:

Let $\tau_p = \inf (t : \sup_{0\leq s\leq t} \| x_s \|^4 \geq p)$, taking $\tau_p = T$ if $\sup_{0\leq t\leq T} \|x_t\|^4 < p$.

From (2.1) it follows that

$$\|x_{t\wedge\tau_p}\|^4 = \|x_0 + \int_0^{t\wedge\tau_p} (Ax_s + B)ds + \int_0^{t\wedge\tau_p} G_1 dw^1{}_s + \int_0^{t\wedge\tau_p} G_2 dw^2{}_s \|^4 \leq$$

$$5^3 (\|x_0\|^4 + \| \int_0^{t\wedge\tau_p} Ax_s ds \|^4 + \| \int_0^{t\wedge\tau_p} B ds \|^4 + \| \int_0^{t\wedge\tau_p} G_1 dw^1{}_s \|^4 +$$

$$\| \int_0^{t\wedge\tau_p} G_2 dw^2{}_s \|^4) , \tag{2.21}$$

where $t\wedge\tau_p$ = minimum (t, τ_p) and arguments (t, y) are omitted.

Denote

$$g^i{}_{jk} = \int_0^{t\wedge\tau_p} [G_i]_{jk} [dw^i{}_s]_k , \quad i = 1, 2,$$

where $[G_i]_{jk}$ $([dw^i{}_s]_k)$ is j, k-th, (k-th) element of $G_i(dw^i{}_s)$.

By Lemma A4, applied with $\alpha = 2$, it follows that

$$E((g^i_{jk})^4) \leq 6^2 T \int_0^T E(([G_i]_{jk})^4)\, ds . \tag{2.22}$$

The inequality

$$\| \int_0^{t\wedge\tau} {}^p G_i dw^i_s \|^4 \leq n\, \ell_i^3 \sum_{j=1}^{n} \sum_{k=1}^{\ell_i} (g^i_{jk})^4$$

and the inequality (2.22) now can be combined to give

$$E(\| \int_0^{t\wedge\tau} {}^p G^i dw^i_s \|^4) \leq 36 \cdot T \cdot n\, \ell_i^3 \sum_{j=1}^{n} \sum_{k=1}^{\ell_i} \int_0^T E(([G_i]_{jk})^4) ds \leq$$

$$36 T\, n\, \ell_i^3 \int_0^T E(\|G_i\|^4) ds = c_i, \quad i = 1,\ 2. \tag{2.23}$$

The upper boundary of the second and the third term in (2.21) are given by

$$\| \int_0^{t\wedge\tau} {}^p A x_s ds \|^4 + \| \int_0^{t\wedge\tau} {}^p B ds \|^4 \leq n^2 ((\int_0^T \|A\|^2 ds)^2 (\int_0^{t\wedge\tau} {}^p \|x_s\|^2 ds)^2 +$$

$$(\int_0^{t\wedge\tau} {}^p \|B\| ds)^4) \leq n^2 ((\int_0^T \|A\|^2 ds)^2 \cdot T \int_0^{t\wedge\tau} {}^p \|x_s\|^4 ds + (\int_0^T \|B\| ds)^4) \tag{2.24}$$

Denote $\quad c_3 = 125(E(\|x_0\|^4) + n^2 E(\int_0^T \|B_s\| ds)^4 + c_1 + c_2)$,

$$c_4 = 125\, n^2\, T (\int_0^T \|A\|^2 ds)^2 .$$

By (2.3) and the assumption about x_0, both c_3 and c_4 are finite.

Taking expectation of the both sides of (2.21) the following inequality is obtained

$$E(\|x_{t\wedge\tau_p}\|^4) \le c_3 + c_4 \int_0^t E(\|x_{s\wedge\tau_p}\|^4)\, ds\,. \tag{2.25}$$

By Lemma A1 it follows from (2.25) that

$$E(\|x_{t\wedge\tau_p}\|^4) \le c_3 \exp(c_4 T)\,,$$

and by the Fatou Lemma (Lemma A5)

$$E(\|x_t\|^4) \le \liminf_{p\to\infty} E(\|x_{t\wedge\tau_p}\|^4) \le c_3 \exp(c_4 T)\,.$$

From the above

$$\sup_{0\le t\le T} E(\|x_t\|^4) < \infty\,. \tag{2.26}$$

To prove (2.20), inequality (2.21) is used again with $t\wedge\tau_p$ replaced by t or by T.

$$\sup_{0\le t\le T} \|x_t\|^4 \le 125(\|x_0\|^4 + n^2\Big(\int_0^T \|B\|\,ds\Big)^4 + \sup_{0\le t\le T}\Big\|\int_0^t G_1 dw^1_s\Big\|^4 +$$

$$\sup_{0\le t\le T}\Big\|\int_0^t G_2 dw^2_s\Big\|^4 + c_4 \int_0^T \|x_s\|^4 ds)\,. \tag{2.27}$$

Lemma A6 applied with $\alpha = 4$ and (2.23) give

$$\mathcal{E}\left(\sup_{0\le t\le T} \left\| \int_0^t G_i dw^i_s \right\|^4\right) \le \left(\frac{4}{3}\right)^4 36\cdot T\cdot n\, \ell_i^3 \int_0^T \mathcal{E}(\|G_i\|^4)\, ds.$$

Hence from (2.26), (2.27) and the above it follows that

$$\mathcal{E}\left(\sup_{0\le t\le T} \|x_t\|^4\right) \le 125\ \Big(\mathcal{E}(\|x_0\|^4) + n^2\, \mathcal{E}\Big(\int_0^T \|B\| ds\Big)^4 +$$

$$\left(\frac{4}{3}\right)^4 36Tn\Big(\ell_1^3 \int_0^T \mathcal{E}(\|G_1\|^4)ds + \ell_2^3 \int_0^T \mathcal{E}(\|G_2\|^4)\Big) + c_4 T \sup_{0\le t\le T} \mathcal{E}(\|x_t\|^4)\Big) < \infty.$$

<u>The above inequality ends proof of Lemma 2.1.b.</u>

The conditions to be checked before Theorems A3 and A4 can be used are

$$\alpha_t = \|R^{-1}(t,y)(C(t,y)x_t + D(t,y))\| < \infty, \text{ P-a.s., for almost all } t\in[0,T], \tag{2.28}$$

and

$$P\Big(\int_0^T \|R^{-1}(t,y)(C(t,y)x_t + D(t,y))\|^2 dt < \infty\Big) = 1. \tag{2.29}$$

From (2.28)

$$\alpha_t \le \|R^{-1}(t,y)\| (0.5(\|x_t\|^2 + \|C(t,y)\|^2) + \|D(t,y)\|)$$

By (2.4), $R^{-1}(t,y)$ is P-a.s. uniformly bounded which implies

$$E(\alpha_t) \leq c(E\|D(t,y)\| + 0.5E(\|x_t\|^2) + 0.5E(\|C(t,y)\|^2)). \qquad (2.30)$$

By (2.3) $\|D(t,\xi)\|$ and $\|C(t,\xi)\|^2$ are bounded for almost all $t \in [0,T]$ and all $\xi \in C_T^m$. Now from (2.20) and (2.30) condition (2.28) follows.

Similarly as the above steps show that (2.29) also holds. From Theorem A4, it follows now that $\mu << \tilde{\mu}$. According to Lemma A2 there exists a measurable functional $\tilde{f}_t(a,\eta,\xi)$ defined on $[0,T]$ x R^n x C_T^n x C_T^m such that P-a.s.

$$\tilde{x}_t = \tilde{f}_t(x_0, w^1, \tilde{y}), \text{ for almost all } t \in [0,T],$$

where $\tilde{x}_t$ results from (2.18) with $z_t = \tilde{y}_t$, $t \in [0,T]$.

In the identical way to that in Lemma 2.1.b, it can be shown that $E(\sup_{0 \leq t \leq T} \|\tilde{x}_t\|^4) < \infty$. This inequality and the conditions (2.3) and (2.4) give

$$P\left(\int_0^T \|R^{-1}(t,\tilde{y})((C(t,\tilde{y})\tilde{f}_t(x_o, w^1, \tilde{y}) + D(t,\tilde{y}))\|^2 dt < \infty\right) = 1.$$

Now, Theorem A3 can be used to show that $\mu \sim \tilde{\mu}$ and

$$\phi_t(x_0, w^1, \tilde{y}) = \frac{d\mu}{d\tilde{\mu}} = \exp\left(\int_0^t Q(s,\tilde{y})R^{-1}(s,\tilde{y})d\tilde{y}_s - 0.5\int_0^t Q(s,\tilde{y})Q^*(s,\tilde{y})ds\right), \quad (2.31)$$

$$\psi_t(x_0, w^1, y) = \frac{d\tilde{\mu}}{d\mu} = \exp\left(-\int_0^t Q(s,y)R^{-1}(s,y)dy_s + 0.5\int_0^t Q(s,y)Q^*(s,y)ds\right), \quad (2.32)$$

where $\quad Q(t,\xi)=(C(t,\xi)f_t(a,\eta,\xi)+D(t,\xi))^*R^{-1}(t,\xi),$

$$(t,a,\eta,\xi)\in[0,T]\times R^n\times C_T^{\,n}\times C_T^{\,m}\ .$$

Denote by $m_t(y)$ a measurable functional (see Lemma A2) such that for almost all $t\in[0,T]$

$$m_t(y) \;=\; E\,(x_t\,|\,\mathcal{Y}_t)\quad \text{P-a. s.}$$

Define

$$\overline{w}_t=\int_0^t R^{-1}(s,y)dy_s-\int_0^t R^{-1}(s,y)(C(s,y)m_s(y)+D(s,y))ds. \qquad (2.33)$$

It can be shown (L2, Lemma 11.3, p. 6) that $(\overline{w}_t,\mathcal{Y}_t)$, $t\in[0,T]$, is a Wiener process, and further (L2, Lemma 11.4, p. 7) that $\mu_y\sim\mu_{\tilde{y}}$ with the following densities

$$\lambda_t(\tilde{y})=\frac{d\mu_y}{d\mu_{\tilde{y}}}=\exp\left(\int_0^t V(s,\tilde{y})R^{-1}(s,\tilde{y})d\tilde{y}_s-0.5\int_0^t V(s,\tilde{y})V^*(s,\tilde{y})ds\right), \qquad (2.34)$$

$${}_t(y)=\frac{d\mu_{\tilde{y}}}{d\mu_y}=\exp\left(\int_0^t V(s,y)R^{-1}(s,y)dy_s-0.5\int_0^t V(s,y)V^*(s,y)ds\right), \quad (2.35)$$

where $V(t,\xi)=(C(t,\xi)m_t(\xi)+D(t,\xi))^*R^{-1}(t,\xi),\ (t,\xi)\in[0,T]\times C_T^{\,m}$.

Now let $(\overline{y}_t,F_t)$ be a random process with the differential

$$d\overline{y}_t=R(t,\overline{y})d\overline{w}_t\ ,\quad \overline{y}_0=y_0\ .$$

By (2.5) the above equation has a unique strong solution. Hence (see (2.16)) the measures $\mu_{\bar{y}}$ and $\mu_{\tilde{y}}$ coincide.

Write (2.9) in the equivalent form

$$dy_t = (C(t,y)m_t(y) + D(t,y))dt + R(t,y)d\bar{w}_t \; . \tag{2.36}$$

Now define

$$\bar{\rho}_t\,(x_0, w^1, \bar{y}) = \frac{\phi_t\,(x_0, w_1, \bar{y})}{\lambda_t(\bar{y})}$$

From (2.31), (2.33) and (2.34), it follows that

$$\rho_t(x_0, w^1, y) \equiv \bar{\rho}_t(x_0, w^1, \bar{y}) \mid_{\bar{y} = y} =$$

$$\exp\left(\int_0^t P_s(x_0, w^1, y)d\bar{w}_s - 0.5\int_0^t P_s(x_0, w^1, y)P_s^{\,*}(x_0, w^1, y)ds\right), \tag{2.37}$$

where

$$P_t(a,\eta,\xi) = (C(t,\xi)(f_t(a,\eta,\xi) - m_t(\xi)))^* R^{-1}(t,\xi),$$

$$(t,a,\eta,\xi) \in [0,T] \times R^n \times C_T^{\,n} \times C_T^{\,m} \; .$$

Now using Bayes Formula (L1, Theorem 7.23, p. 289) the following representation of a conditional expectation can be derived. Let $F_t(a,\eta,\xi)$ denote a measurable functional on $[0,T]\times R^n \times C_T^{\,n} \times C_T^{\,m}$, such that $E(\| F_t(x_0, w^1, y)\|) < \infty$, for all $t \in [0,T]$. Then

$$E(F_t(x_0, w^1, y) \mid \mathcal{Y}_t) = \int_{R^n} \int_{C_T^n} F_t(a,\eta, y)\, \rho_t(a, \eta, y) d\mu_{w^1}(\eta) dP_{y_0}(a), \tag{2.38}$$

where $\mu_{w^1}(\cdot)$ is a Wiener measure on C_T^m and $P_{y_0}(a) = P(x_0 \leq a \mid y_0)$.

Let $0 \leq t_0 < t_1 \ldots < t_k \leq t \leq T$ be some decomposition of the interval $[0, T]$. Then for $z_i \in R^n$, $i = 0, 1, \ldots, k$, denote $Z = (z_0, \ldots, z_k)$ and

$$F_t(x_0, w^1, y, Z) = \exp\left(j \sum_{\ell=0}^{k} (z_\ell)^* f_{t_\ell}(x_0, w^1, y)\right)$$

where j is the imaginary unit.

By (2.38) and the above

$$\phi_t(Z, y) = E(F_t \mid \mathcal{Y}_t) = \int_{R^n} \Delta_t(a, y, Z) dP_{y_0}(a), \tag{2.39}$$

where

$$\Delta_t(a, y, Z) = \int_{C_T^n} F_t(a,\eta, y, Z)\, \rho_t(a, \eta, y) d\mu_{w^1}(\eta) .$$

It will be shown that $\Delta_t(a, y, Z)$ has the following form

$$\Delta_t(a, y, Z) = \exp\left(Q_t(a, y, Z)\right)$$

where Q_t is quadratic in the variables a, Z, and is nonnegative definite in Z. Then, because $P_{y_0}(a)$ is Gaussian and $\phi_t(Z, y)$ is the conditional characteristic function of x_t given $\mathcal{Y}_t$, equation (2.39) shows x_t is conditionally Gaussian.

To this end Lemma A7 will be used. To show that $\Delta_t(a, y, Z)$ has the form which allows for the application of the mentioned Lemma, a number of notational changes will be made. All of them involve simple algebraic equations and the equations (2.8), (2.9), (2.18), (2.33), (2.37) and (2.39).

Denote (arguments (t, y) on the right hand sides are omitted)

$$h_1(t, y) = (C(\Phi_t(\int_0^t \Phi_s^{-1}\tilde{B}_s ds + \int_0^t \Phi_s^{-1} H_2 R^{-1} dy_s) - m_t))^* R^{-1},$$

$$h_2(t, y) = \Phi_t^* C^* R^{-1}, \qquad h_3(t, y) = \Phi_t^{-1} H_1,$$

$$q_1(t, y) = \int_0^t h_1 d\bar{w}_s - 0.5 \int_0^t h_1 h_1^* ds,$$

$$q_2(t, y) = \int_0^t h_2 d\bar{w}_s - \int_0^t h_2 h_1^* ds,$$

$$q_3(t, y) = \int_0^t h_2 h_2^* ds,$$

$$q_4(t, w^1, y) = \int_0^t (\int_0^s h_3 dw^1_\tau)^* h_2 d\bar{w}_s - \int_0^t h_1 h_2^* (\int_0^s h_3 dw^1_\tau) ds,$$

$$q_5(t,w^1,y) = -\int_0^t h_2h_2^*\left(\int_0^s h_3dw^1_\tau\right)ds \quad .$$

The last two equalities can be rewritten as

$$q_4(t,w^1,y)=\int_0^t (h_4(t,y)-h_5(t,y)-h_4(s,y)+h_5(s,y))h_3(s,y)dw^1_s, \tag{2.40}$$

where
$$h_4(t,y) = \left(\int_0^t h_2d\bar{w}_s\right)^* \quad ,$$

$$h_5(t,y) = \int_0^t h_1h_2^*ds \quad ,$$

and as

$$q_5(t,w^1,y) = -\int_0^t (q_3(t,y)-q_3(s,y))h_3(s,y)dw^1_s \quad . \tag{2.41}$$

Now from (2.37), (2.18) and the above equations note that

$$\ln\rho_t(x_0,w^1,y)=q_1(t,y) + x_0^*(q_2(t,y) + q_5(t,w^1,y)) + q_4(t,w^1,y) - 0.5\,x_0^*q_3(t,y)x_0 - 0.5\int_0^t\left(\int_0^s h_3dw^1_\tau\right)^*h_2h_2^*\left(\int_0^s h_3dw^1_\tau\right)ds \;. \tag{2.42}$$

Equivalence of the measures μ and $\tilde{\mu}$, Lemma A3 and (2.42) give

$$\ln\rho_t(x_0,w^1,\tilde{y})=q_1(t,\tilde{y})+x_0^*(q_2(t,\tilde{y})+q_5(t,w^1,\tilde{y}))+q_4(t,w^1,\tilde{y}) - 0.5\,x_0^*q_3(t,\tilde{y})x_0 - 0.5\int_0^t\left(\int_0^s h_3dw^1_\tau\right)^*h_2h_2^*\left(\int_0^s h_3dw^1_\tau\right)ds \;, \tag{2.43}$$

where in h_2, h_3 arguments $(t, \tilde{y})$ are omitted.

Using the fact that the processes w^1 and $\tilde{y}$ are independent (see (2.16)) it is seen from (2.40) and (2.41) that the conditional distribution of q_4 and q_5 given $\tilde{y}$ is Gaussian. By the equivalence of measures μ_y and $\mu_{\tilde{y}}$ it is enough to show that $Q_t(a, \tilde{y}, Z)$ has the desired quadratic form. But equations (2.39) and (2.43), together with the discussion above allow for the application of Lemma A7 from which it is concluded that

$$\phi_t(Z, \tilde{y}) = \exp\left(j \sum_{\ell=0}^{k} (z_\ell)^* m_\ell(t, y) - 0.5 \sum_{\ell=0}^{k} \sum_{i=0}^{k} (z_\ell)^* \Gamma_{\ell i}(t, y)(z_i)\right),$$

where $\Gamma_{\ell i}(t, y)$ are some special nonnegative definite symmetric, n x n, matrices and $m_\ell(t, y)$ are some special n dimensional vectors. It will be shown in the sequel that they are a solution to the certain stochastic differential equations (filter).

Because of the arbitrariness of z_i, $i = 0, 1, \ldots, k$ this ends the proof that the conditional distribution of $x_{t_0}, \ldots, x_{t_k}$, given $\mathcal{Y}_t$ is P-a.s. Gaussian. This result, together with some results from the nonlinear filtering theory will be now used to derive the filter equation (2.6) and (2.7).

Let (z_t, y_t), $t \in [0, T]$, be a partially observable random process with only the second component $(y_t, \mathcal{F}_t)$ being observed. The stochastic differential representation of the a posteriori mean

$E(z_t | \mathcal{Y}_t)$ is to be found. Assume that z_t has the following representation

$$z_t = z_0 + \int_0^t \Lambda_s d_s + \lambda_t, \quad t \in [0, T] \tag{2.44}$$

where

$(\lambda_t, \mathcal{F}_t)$ is a martingale and $(\Lambda_t, \mathcal{F}_t)$ is a random process

such that

$$\int_0^T \|\Lambda_t\| dt < \infty \quad \text{P-a. s.}$$

It will be shown that y_t given by (2.9) can be represented as an Ito process

$$y_t = y_0 + \int_0^t \overline{C}_s ds + \int_0^t R_s \, d\overline{w}_s, \tag{2.45}$$

where

$(\overline{w}_t, \mathcal{F}_t)$ is a Wiener process, and the processes $(\overline{C}_t, \mathcal{F}_t)$ and $(R_t, \mathcal{F}_t)$ are assumed to be nonanticipative and such that

$$P\left(\int_0^T \|\overline{C}_t\| dt < \infty\right) = 1, \quad P\left(\int_0^T \|R_t\|^2 dt < \infty\right) = 1 .$$

To prove that y_t permits the representation (2.45) Theorem A5 is used. To satisfy its assumptions it must be shown that

$$\int_0^T E(\|C(t, y)x_t + D(t, y)\|)dt < \infty .$$

But the above is the straightforward consequence of the assumption (2.3) and Lemma 2.1b.

Then the condition that $R^2(t, y) > 0$ follows directly from (2.4). From (2.5) and the fact that y_t satisfies (2.9), it follows that

$$\int_0^T \|R(t, y)\|^2 dt < \infty \quad P\text{-a. s.}$$

Summarizing, conditions (2.3), (2.4), (2.5) and Lemma 2.1b allow for an application of Theorem A5 and the following result is obtained:

For y_t, $t \in [0, T]$, satisfying (2.9) there exists a Wiener process $(\bar{w}_t, F_t)$, $t \in [0,T]$, such that y_t has the representation

$$y_t = y_0 + \int_0^t (C(s, y)m_s + D(s, y))ds + \int_0^t R(s, y)d\bar{w}_s \,, \tag{2.46}$$

where $\qquad m_t = E(x_t \mid \mathcal{Y}_t)$.

Now represent $z_t^1 = x_t$ and $z_t^2 = x_t x_t^*$ in the form given by (2.44). From (2.8), it follows that

$$x_t = x_0 + \int_0^t (Ax_s + B)\, ds + \lambda_t^1 \,, \tag{2.47}$$

and

$$x_t x_t^* = x_0 x_0^* + \int_0^t (x_s(Ax_s + B)^* + (Ax_s + B)x_s^* + H_1H_1^* + H_2H_2^*)ds + \lambda_t^2 , \tag{2.48}$$

where
$$\lambda_t^1 = \int_0^t (H_1H_1^* dw_s^1 + H_2H_2^* dw_s^2) \ ,$$

$$\lambda_t^2 = \int_0^t (x_s(d\lambda_s^1)^* + d\lambda_s^1 x_s^*) \ ,$$

and the arguments (t, y) are omitted for brevity.

It will be shown now that λ_t^i, i = 1, 2, are square integrable martingales. This is the first step towards application of Theorem A6.

The martingale properties follow straightforward from the above equations which define λ_t^i, i = 1, 2. To show square integrability of λ_t^1 proceed as follows. From (2.47)

$$\|\lambda_t^1\|^2 = \| x_t - x_0 - \int_0^t (Ax_s + B)ds \|^2 \leq 3(\|x_t\|^2 + \|x_0\|^2 + \|\int_0^t (Ax_s + B)ds\|^2) \leq$$

$$\leq 3(\|x_t\|^2 + \|x_0\|^2 + 2n((\int_0^T \|A\|^2 ds)(\int_0^t \|x_s\|^2 ds) + (\int_0^T \|B\| ds)^2) \ .$$

Hence
$$\sup_{0 \leq t \leq T} E(\|\lambda_t^1\|^2) \leq c_1 + c_2 \sup_{0 \leq t \leq T} E(\|x_t\|^2) < \infty \ , \tag{2.49}$$

where
$$c_1 = 3(E(\|x_0\|^2) + 2n(\int_0^T \|B\| ds)^2) \ ,$$

$$c_2 = 6n\,T \int_0^T \|A\|^2 ds + 3 \ ,$$

and the last inequality results from (2.3)and (2.20).

Similarly, using (2. 48) it can be shown that

$$\|\lambda_t^2\|^2 \leq 4(\|x_t\|^4+\|x_0\|^4+8(\int_0^T \|A\|^2 ds+4T)\int_0^t \|x_s\|^4 ds+4(\int_0^T \|B\|^2 ds)^2 +$$

$$2n^2(\int_0^T \|G_1G_1^*\| ds)^2 + 2n^2(\int_0^T \|G_2G_2^*\| ds)^2) \;.$$

From the above, (2. 3) and (2. 20) it follows that

$$\sup_{0\leq t\leq T} E(\|\lambda_t^2\|^2) < \infty \;. \tag{2.50}$$

Now take the conditional expectation $E(\cdot\,|Y_t)$ of the both sides of (2. 47) and (2. 48), and apply Fubini's theorem to obtain

$$\pi_1(t,t)= \pi_1(0,t)+\int_0^t (A\pi_1(s,t)+B)ds+ (\lambda_t^1|Y_t) \;, \tag{2.51}$$

and

$$\pi_2(t,t)= \pi_2(0,t)+\int_0^t (\pi_1(s,t)B^*+B\pi_1^*(s,t)+\pi_2(s,t)A^* +$$

$$A\pi_2(s,t) + H_1H_1^* + H_2H_2^*)ds + (\lambda_t^2|Y_t) \;, \tag{2.52}$$

where

$$\pi_i(s,t) = E(z_s^i\,|\,Y_t),\; i = 1,\; 2,\; s,t\in[0,T] \;.$$

By Theorem A6 together with (2.3) $\pi_i(0,t)$, i = 1, 2, are square integrable martingales with the representation

$$\pi_1(0,t) = \pi_1(0,0) + \int_0^t F_s^1 \, d\bar{w}_s \quad , \tag{2.53}$$

where

$$\int_0^T E(\|F_s^1\|^2)ds < \infty \quad ,$$

and

$$\pi_2(0,t) = \pi_2(0,0) + \int_0^T \sum_{k=1}^m F_s^{2k} \, [d\bar{w}_s]_k \quad , \tag{2.54}$$

where

$$\int_0^T \sum_{k=1}^m E(\|F_s^{2k}\|^2)ds < \infty \; .$$

The purpose of the following Lemma 2.1c is to show that $E(\lambda_t^i | \mathcal{Y}_t)$ i = 1, 2, admit similar representations as the one given above.

Lemma 2.1c

The processes $E(\lambda_t^i | \mathcal{Y}_t)$, i = 1, 2, are the square integrable martingales with the representation

$$E(\lambda_t^1 | \mathcal{Y}_t) = \int_0^t \Lambda_s^1 \, d\bar{w}_s \quad , \tag{2.55}$$

and

$$E(\lambda_t^2 | \mathcal{Y}_t) = \int_0^t \sum_{k=1}^m \Lambda_s^{2k} [d\overline{w}_s]_k \; , \tag{2.56}$$

where
$$\int_0^T E(\|\Lambda_s^1\|^2)\, ds < \infty \; ,$$

and
$$\int_0^T \sum_{k=1}^m E(\|\Lambda_s^{2k}\|^2) ds < \infty \; .$$

Proof of Lemma 2.1c

First it is shown that $E(\lambda_t^1 | \mathcal{Y}_t)$ is a square integrable process

$$\sup_{0 \le t \le T} E(\| E(\lambda_t^1 | \mathcal{Y}_t)\|^2) \le \sup_{0 \le t \le T} E(E(\|\lambda_t^1\|^2 | \mathcal{Y}_t) =$$

$$\sup_{0 \le t \le T} E(\|\lambda_t^1\|^2) < \infty \; .$$

The inequality above results from (2.49).

To check a martingale property of $E(\lambda_t^1 | \mathcal{Y}_t)$ it must be shown that

$$E(E(\lambda_t^1 | \mathcal{Y}_t) | \mathcal{Y}_s) = E(\lambda_s^1 | \mathcal{Y}_t) \; , \quad t \ge s \; . \tag{2.57}$$

Notice that by definition $\mathcal{Y}_s \subseteq \mathcal{Y}_t$ which allows to rewrite the left hand side of (2.57) as follows

$$E(\lambda_t^1 | \mathcal{Y}_s) = E(E(\lambda_t^1 | F_s) | \mathcal{Y}_s) = E(\lambda_s^1 | \mathcal{Y}_s) \; .$$

The above equality which proves (2.57) follows from the fact that $\mathcal{Y}_s \subseteq \mathcal{F}_s$ and that $(\lambda_t^1, \mathcal{F}_t)$ is a martingale (see (2.47)).

Apply now Theorem A6, which together with the obvious equality $E(\lambda_0^1 | \mathcal{Y}_0) = 0$ gives (2.55). Proof of (2.56) is identical.

<u>This ends the proof of Lemma 2.1c</u>

Lemma 2.1d below gives the desired representation of the second terms in (2.51) and (2.52).

<u>Lemma 2.1d</u>

Let
$$\Delta_s^i = \pi_i(s,t) - \pi_i(s,s)\ , \quad i = 1,\ 2$$

and
$$\eta_t^1 = \int_0^t A\,\Delta_s^1\, ds\ ,$$

$$\eta_t^2 = \int_0^t (\Delta_s^1 B^* + B(\Delta_s^1)^* + \Delta_s^2 A^* + A\Delta_s^2)\, ds\ .$$

Then η_t^1, η_t^2 are the square-integrable martingales $(\eta_t^i, \mathcal{Y}_t)$, having the representation

$$\eta_t^1 = \int_0^T M_s^1\, d\bar{w}_s\ , \tag{2.58}$$

and

$$\eta_t^2 = \int_0^T \sum_{k=1}^m M_s^{2k}\, [d\bar{w}_s]_k\ , \tag{2.59}$$

where

$$\int_0^T E(\|M_s^1\|^2)ds < \infty ,$$

and

$$\int_0^T \sum_{k=1}^m E(\|M_s^{2k}\|^2)\, ds < \infty .$$

Proof of Lemma 2.1d

Similarly, as in the proof of Lemma 2.1c, it is shown that η_t^i are square integrable processes and further that they possess the martingale property. The square integrability of η_t^1 is proved to hold as follows

$$\|\eta_t^1\|^2 = \|\int_0^t A\Delta_s^1 ds\|^2 \leq n \int_0^T \|A\|^2 ds \int_0^t \|\Delta_s^1\|^2\, ds \leq$$

$$c_1 \int_0^t (\|\pi_1(s,t)\|^2 + \|\pi_1(s,s)\|^2)ds ,$$

where

$$c_1 = 2n \int_0^T \|A\|^2\, ds .$$

Hence

$$\sup_{0\leq t\leq T} E(\|\eta_t^1\|^2) \leq 2\, T\, c_1 \sup_{0\leq t\leq T} E(\|x_t\|^2) < \infty .$$

Proving the above inequality (2.3) and (2.26) are used. To show martingale property of η_t^1 note that

$$E(\eta_t^1 \mid Y_s) = E(\int_0^t A\pi_1(\tau, t)\, d\tau \mid Y_s) - E(\int_0^t A\pi_1(\tau,\tau)\, d\tau \mid Y_s) =$$

$$E(\int_0^t A x_\tau d\tau \mid Y_s) - \int_0^t E(A\pi_1(\tau,\tau) \mid Y_s) d\tau = E(\int_0^s A x_\tau d\tau \mid Y_s) +$$

$$E(\int_0^t A x_\tau d\tau \mid Y_s) - \int_0^s E(A\,\pi_1(\tau,\tau) d\tau \mid Y_s) d\tau - \int_s^t E(A\,\pi_1(\tau,\tau) \mid Y_s) d\tau =$$

$$\int_0^s A\,\pi_1(\tau, s) d\tau + \int_s^t E(A x_\tau \mid Y_s) d\tau - \int_0^s A\pi_1(\tau,\tau) d\tau - \int_s^t E(A x_\tau \mid Y_s) d\tau =$$

$$= \int_0^s A(\pi_1(\tau, s) - \pi_1(\tau,\tau))\, d\tau = \eta_s^1$$

Similar calculations apply to η_t^2 where the following inequality is used

$$\|\eta_t^2\|^2 \le 8n^2\left(\int_0^T \|A\|^2 ds \int_0^t \|\Delta_s^2\|^2 ds + 0.5\left(\left(\int_0^T \|B\|^2 ds\right)^2 + T\int_0^t \|\Delta_s^1\|^4 ds\right)\right)$$

Application of Theorem A6 to both η_t^i, $i = 1, 2$, results in (2.58) and (2.59).

This ends the proof of Lemma 2.1d

From (2.53), (2.54), (2.55), (2.56), (2.58), and (2.59) equations (2.47) and (2.48), take the following representation

$$m_t = m_0 + \int_0^t (A\, m_s + B)\, ds + \int_0^t V_s^{\,1}\, d\overline{w}_s\ , \tag{2.60}$$

where

$$m_t = \pi_1(t,t) \text{ (see (2.33) and (2.47))}\ ,$$

$$V_s^{\,1} = F_s^{\,1} + \Lambda_s^{\,1} + M_s^{\,1}\ ,$$

and

$$\pi_2(t,t) = \pi_2(0,0) + \int_0^t (m_s B^* + B m_s^{\,*} + \pi_2(s,s) A^* + A\pi_2(s,s) +$$

$$H_1 H_1^{\,*} + H_2 H_2^{\,*})\, ds + \int_0^t \sum_{k=1}^m V_s^{\,2k}\, [d\overline{w}_s]_k\ , \tag{2.61}$$

where $\qquad V_s^{\,2k} = F_s^{\,2k} + \Lambda_s^{\,2k} + M_s^{\,2k}.$

By Lemma 2.1c and Lemma 2.1d

$$\int_0^T E(\| V_s^{\,1} \|^2 ds < \infty\ , \text{ and } \int_0^T \sum_{k=1}^m E(\| V_s^{\,2k} \|^2) ds < \infty\ .$$

To derive the filter equations from (2.60) and (2.61) the following notation will be used. For $t \in [0,T]$,

$$\alpha_t^{\,1} = \int_0^t V_s^{\,1} d\overline{w}_s\ , \quad \alpha_t^{\,2} = \int_0^t \sum_{k=1}^m V_s^{\,2k}\, [d\overline{w}_s]_k\ ,$$

and

$$\beta_t^{\,1} = \int_0^t (d\overline{w}_s)^* \gamma_s^{\,1}, \quad \beta_t^{\,2} = \int_0^t \sum_{k=1}^m \gamma_s^{\,2k}\, [d\overline{w}_s]_k\ ,$$

where $(\gamma_t^1, \mathcal{Y}_t)$ is some bounded, m x n dimensional, process such that $\|\gamma_t^1\| \leq c < \infty$ P-a. s. almost everywhere on $[0, T]$, and $(\gamma_t^{2k}, \mathcal{Y}_t)$, $k = 1, \ldots, m$ are some bounded random, n x n dimensional processes such that $\sum_{k=1}^{m} \|\gamma_t^{2k}\| \leq c < \infty$ P-a. s. almost everywhere on $[0, T]$.

By the properties of the stochastic integrals it follows that

$$E(\alpha_t^1 \beta_t^1) = E\left(\int_0^t V_s^1 \gamma_s^1 ds\right) . \tag{2.62}$$

Note that

$$E(m_0 \beta_t^1) = E(m_0 E(\beta_t^1 | \mathcal{Y}_0)) = 0 . \tag{2.63}$$

The above is implied by the fact that β_t^1 is a martingale and $\beta_0^1 = 0$ P-a. s.

Further

$$E\left(\int_0^t (Am_s + B) ds \beta_t^1\right) = \int_0^t E((Am_s + B) E(\beta_t^1 | \mathcal{Y}_s)) ds =$$

$$= \int_0^t E((Am_s + B) \beta_s^1) ds . \tag{2.64}$$

In (2.64) again the martingale property of β_t^1 were used. From (2.60) it follows that

$$\alpha_t^1 = m_t - m_0 - \int_0^t (Am_s + B) ds .$$

The above, together with (2.63) and (2.64) give

$$\mathcal{E}(\alpha_t^1\beta_t^1) = \mathcal{E}(m_t\beta_t^1) - \int_0^t \mathcal{E}((Am_s + B)\beta_s^1)ds =$$

$$\mathcal{E}(\mathcal{E}(x_t\beta_t^1 | \mathcal{Y}_t)) - \int_0^t \mathcal{E}(\mathcal{E}((Ax_s + B)\beta_s^1 | \mathcal{Y}_s))ds =$$

$$\mathcal{E}(x_t\beta_t^1 - \int_0^t (Ax_s + B)\beta_s^1 ds) . \tag{2.65}$$

From (2.46) it follows that

$$\bar{w}_t = \int_0^t R^{-1}(dy_s - (Cm_s + D)\,ds) , \tag{2.66}$$

and next from (2.9)

$$\bar{w}_t = w_t^2 + \int_0^t R^{-1} C(x_s - m_s)\,ds . \tag{2.67}$$

Using (2.67) expression for β_t^1 can be written as follows

$$\beta_t^1 = \bar{\beta}_t^1 + \int_0^t (R^{-1}C(x_s - m_s))^* \gamma_s^1\,ds , \tag{2.68}$$

where

$$\bar{\beta}_t^1 = \int_0^t (dw_s^2)^* \gamma_s^1 .$$

Now (2.65) takes the form

$$\mathcal{E}(\alpha_t^1\beta_t^1) = \mathcal{E}(x_t\bar{\beta}_t^1 - \int_0^t (A x_s + B)\bar{\beta}_s^1 ds) +$$

$$E\left(\int_0^t x_t(R^{-1}C(x_s-m_s))^* \gamma_s^1 ds\right) - E\left(\int_0^t\int_0^s (Ax_s+B)(R^{-1}C(x_\tau-m_\tau))^* \gamma_\tau^1 d\tau ds\right) \tag{2.69}$$

$(\bar{\beta}_t^1, F_t)$ is by definition and by (2.68) a square integrable martingale and similarly to (2.63).

$$E(x_0\bar{\beta}_t^1) = E(x_0 E(\bar{\beta}_t^1 | F_0) = E(x_0\bar{\beta}_0^1) = 0 \quad \text{P-a.s.}$$

Analogously to (2.64) the following equalities hold

$$E\left(\int_0^t (Ax_s+B)\bar{\beta}_s^1 ds\right) = E\left(\int_0^t (Ax_s + B) E(\bar{\beta}_t^1 | F_s) ds\right) =$$

$$E\left(\int_0^t (A x_s + B) ds \, \bar{\beta}_t^1\right).$$

Hence the first term of the right hand side of (2.69) can be written as

$$E\left(x_t\bar{\beta}_t^1 - \int_0^t (Ax_s + B)\bar{\beta}_s^1 ds\right) = E\left(\left(x_t - x_0 - \int_0^t (A x_s + B) ds\right) \bar{\beta}_t^1\right) =$$

$$E\left(\int_0^t (H_1 dw_s^1 + H_2 dw_s^2) \int_0^t (dw_s^2)^* \gamma_s^1\right) = E\left(\int_0^t H_2 \gamma_s^1 ds\right). \tag{2.70}$$

The second term of (2.69) takes the form

$$E\left(\int_0^t x_t\,(R^{-1}C(x_s-m_s))^*\gamma_s^1\,ds\right) =$$

$$E\left(\int_0^t x_s(R^{-1}C(x_s-m_s))^*\gamma_s^1 ds\right) + E\left(\int_0^t (x_t-x_s)(R^{-1}C(x_s-m_s))^*\gamma_s^1 ds\right) =$$

$$E\left(\int_0^t \Gamma_s C^* R^{-1}\gamma_s^1 ds\right) + e_t^1\,, \tag{2.71}$$

where

$$\Gamma_t = \pi_2(t,t) - m_t m_t^*\,, \tag{2.72}$$

and

$$e_t^1 = E\left(E\left(\int_0^t (x_t - x_s)\,(R^{-1}C(x_s-m_s))^*\gamma_s^1 ds \mid F_s\right)\right).$$

From (2.47) it follows by the martingale property of λ_t^1 that

$$E((x_t - x_s) \mid F_s) = E\left(\int_s^t (A\,x_\tau + B)\,d\tau \mid F_s\right) + E\left((\lambda_t^1 - \lambda_s^1) \mid F_s\right) =$$

$$E\left(\int_0^t (A\,x_\tau + B)\,d\tau \mid F_s\right).$$

Hence

$$e_t^1 = E\left(\int_0^t\int_s^t (A\,x_\tau + B)\,(R^{-1}C(x_s-m_s))^*\gamma_s^1\,d\tau ds\right) =$$

$$E\left(\int_0^t\int_0^s (Ax_s + B)\,(R^{-1}C(x_\tau-m_\tau))^*\gamma_\tau^1\,d\tau ds\right). \tag{2.73}$$

Now from (2.62), (2.69), (2.70) and (2.73) the following equations can be obtained:

$$E\left(\int_0^t (V_s^1 - H_2 - \Gamma_s C^* R^{-1})\gamma_s^1 ds\right) = 0 . \tag{2.74}$$

Since γ_s^1 was arbitrary it is concluded that

$$\Gamma_t C^* R^{-1} + H_2 = V_t^1 \quad \text{P-a.s. for almost all } t \in [0,T].$$

Now as the value of the integral

$$\int_0^t V_s^1 \, dw_s$$

does not change with the change of V_s^1 on the set of Lebesgue measure zero, then for each $t \in [0,T]$, (2.60) takes the following form

$$m_t = m_0 + \int_0^t (Am_s + B)ds + \int_0^t (H_2 + \Gamma_s C^* R^{-1}) d\overline{w}_s . \tag{2.75}$$

To derive a similar equation for Γ_t (see (2.72)) calculations of the type (2.62) to (2.75) are repeated for α_t^2 and β_t^2. Anologously to (2.62),

$$E(\alpha_t^2 \beta_t^2) = E\left(\sum_{k=1}^{m} \int_0^t V_s^{2k} \gamma_s^{2k} \, ds\right) . \tag{2.76}$$

As in (2.63), it is shown that

$$E(\pi_2(0,0)\beta_t^2) = E(\pi_2(0,0)E(\beta_t^2 \mid \mathcal{Y}_0)) = 0. \tag{2.77}$$

To simplify rather complicated equations, the following notation is introduced:

$$K_s = x_s B_s^* + B x_s^* + x_s x_s^* A^* + A x_s x_s^* + H_1 H_1^* + H_2 H_2^*, \qquad (2.78)$$

and

$$L_s = (K_s \mid \mathcal{Y}_s) = m_s B^* + B m_s^* + \pi_2(s,s) A^* + A \pi_2(s,s) + H_1 H_1^* + H_2 H_2^* . \qquad (2.79)$$

Now similarly to (2.64)

$$E\left(\int_0^t L_s ds \beta_t^2\right) = E\left(\int_0^t L_s \beta_s^2 ds\right)$$

and further from (2.61), (2.77) and the above

$$E(\alpha_t^2 \beta_t^2) = E(\pi_2(t,t)\beta_t^2) - E\left(\int_0^t L_s \beta_s^2 ds\right) =$$

$$E\left(x_t x_t^* \beta_t^2 - \int_0^t K_s \beta_s^2 ds\right). \qquad (2.80)$$

Let r_s^k denotes the k-th element of the m x 1 vector $R^{-1}C(x_s - m_s)$. Define $\bar{\beta}_t^2$ as follows

$$\bar{\beta}_t^2 = \int_0^t \sum_{k=1}^m \gamma_s^{2k} [dw_s^2]_k .$$

Now β_t^2 can be decomposed into two parts

$$\beta_t^2 = \bar{\beta}_t^2 + \int_0^t \sum_{k=1}^m \gamma_s^{2k} r_s^k ds .$$

The above follows from the definition of β_t^2 and (2.67). Now (2.80) can be rewritten as follows:

$$E(\alpha_t^2\beta_t^2) = E\left(x_t x_t^* \bar{\beta}_t^2 - \int_0^t K_s \bar{\beta}_s^2 ds\right) + E\left(x_t x_t^* \int_0^t \sum_{k=1}^m \gamma_s^{2k} r_s^k ds - \right.$$

$$\left. E\left(\left(\int_0^t \int_0^s K_s \sum_{k=1}^m \gamma_\tau^{2k} r_\tau^k d\tau\, ds\right)\right)\right. . \tag{2.81}$$

Using the equalities

$$E(x_0 x_0^* \bar{\beta}_t^2) = E(x_0 x_0^* E(\bar{\beta}_t^2 (F_0)) = E(x_0 x_0^* \bar{\beta}_0^2) = 0 ,$$

and

$$E\left(\int_0^t K_s \bar{\beta}_s^2\right) = E\left(\int_0^t K_s E(\bar{\beta}_t^2 \mid F_s) ds\right) = E\left(\int_0^t K_s ds \bar{\beta}_t^2\right) ,$$

the first term of the right hand side of (2.81) can be written in the following form

$$E\left(\left(x_t x_t^* - x_0 x_0^* - \int_0^t K_s ds\right) \bar{\beta}_t^2\right) = E(\Lambda_t^2 \bar{\beta}_t^2) =$$

$$E\left(\int_0^t (x_s(H_1 dw_s^1 + H_2 dw_s^2)^* + (H_1 dw_s^1 + H_2 dw_s^2) x_s^*) \int_0^t \sum_{k=1}^m \gamma_s^{2k} [dw_s^2]_k\right) =$$

$$E\left(\int_0^t \sum_{k=1}^m (x_s (h_2^k)^* + h_2^k x_s^*) \gamma_s^{2k} ds\right) , \tag{2.82}$$

where

h_2^k denotes the k-th column of H_2.

The second term of the right hand side of (2.81) is equal to

$$E\left(\int_0^t x_s x_s^* \sum_{k=1}^m \gamma_s^{2k} r_s^k\, ds\right) + e_t^2 , \tag{2.83}$$

where

$$e_t^2 = E\left(\int_0^t (x_t x_t^* - x_s x_s^*) \sum_{k=1}^m \gamma_s^{2k} r_s^k\, ds\right) .$$

From (2.48) and (2.78) it follows that

$$E((x_t x_t^* - x_s x_s^*) \mid F_s) = E\left(\int_0^t K_\tau d\tau \mid F_s\right) + E((\lambda_t^2 - \lambda_s^2) \mid F_s) =$$

$$E\left(\int_s^t K_\tau d\tau \mid F_s\right).$$

Hence

$$e_t^2 = E\left(\int_0^t \int_s^t K_\tau \sum_{k=1}^m \gamma_s^{2k} r_s^k d\tau ds\right) = E\left(\int_0^t \int_0^s K_s \sum_{k=1}^m \gamma_\tau^{2k} r_\tau^k d\tau ds\right) . \tag{2.84}$$

(2.82) has the following equivalent form

$$E\left(\int_0^t \sum_{k=1}^m (m_s (h_2^k)^* + h_2^k m_s^*) \gamma_s^{2k}\, ds\right) , \tag{2.85}$$

and the first term in (2.83) is equal to

$$E\left(\int_0^t E(x_s x_s^* r_s^k \mid Y_s) \sum_{k=1}^m \gamma_s^{2k}\, ds\right) . \tag{2.86}$$

Finally (2.76), (2.81), (2.83), (2.84), (2.85) and (2.86) give

$$\mathcal{E}\left(\sum_{k=1}^{m}\int_0^t V_s^{2k}\gamma_s^{2k}ds\right) = \mathcal{E}\left(\sum_{k=1}^{m}\int_0^t (m_s(h_2^k)^* + h_2^k m_s^*)\gamma_s^{2k}\right) +$$

$$\mathcal{E}\left(\sum_{k=1}^{m}\int_0^t \mathcal{E}(x_s x_s^* r_s^k | \mathcal{Y}_s)\gamma_s^{2k}ds\right). \tag{2.87}$$

By reasoning similar to that used for derivation of (2.74), it can be shown that

$$V_t^{2k} = m_t(h_2^k)^* + h_2^k m_t^* + \mathcal{E}(x_t x_t^* r_t^k | \mathcal{Y}_t), \quad k = 1, \ldots, m,$$

P-a. s. for almost all $t \in [0, T]$. This implies

$$\sum_{k=1}^{m} V_t^{2k}[d\bar{w}_s]_k = m_s(d\bar{w}_s)^* H_2^* + H_2 d\bar{w}_s m_s^* +$$

$$\mathcal{E}(x_s x_s^* (x_s - m_s)^* C^* R^{-1} d\bar{w}_s | \mathcal{Y}_s). \tag{2.88}$$

From (2.75), using Ito's formula, it follows that

$$m_t m_t^* = m_0 m_0^* + \int_0^t (m_s B^* + B m_s^* + m_s m_s^* A^* + A m_s m_s^*)ds +$$

$$+\int_0^t \Lambda_s \Lambda_s^* ds + \int_0^t (m_s (d\bar{w})^* \Lambda_s^* + \Lambda_s d\bar{w}\, m^*{}_s), \tag{2.89}$$

where

$$\Lambda_s = H_2 + \Gamma_s C^* R^{-1}.$$

Subtracting (2.89) from (2.61) and using (2.72), (2.88) equation for Γ_t is obtained

$$\Gamma_t = \Gamma_0 + \int_0^t (\Gamma_s A^* + A \Gamma_s + H_1 H_1^* + H_2 H_2^* - \Lambda_s \Lambda_s^*) ds + \int_0^t \Pi_s ds, \qquad (2.90)$$

where $\Pi_s = E(x_s x_s^* (x_s - m_s)^* C^* R^{-1} d\bar{w}_s \mid \mathcal{Y}_s) -$

$$m_s (d\bar{w}_s)^* R^{-1} C \Gamma_s - \Gamma_s C^* R^{-1} d\bar{w}_s m_s^* .$$

Now the property that x_t is conditionally Gaussian will be used. In particular it means that

$$E([x_s]_i [x_s]_j ([x_s]_\ell - [m_s]_\ell \mid \mathcal{Y}_t) = [m_s]_i [\Gamma_s]_{j\ell} + [m_s]_j [\Gamma_s]_{\ell i},$$

where $[(\cdot)]_i$ denotes the i-th component of a vector $(\cdot)$. Hence for any $(n \times 1)$ dimensional process g_t, which is $\mathcal{Y}_t$ measurable, $t \in [0, T]$, the above implies

$$E(x_s x_s^* (x_s - m_s)^* g_s \mid \mathcal{Y}_s) = m_s (\Gamma_s g_s)^* + \Gamma_s g_s m_s^* .$$

Let $g_s = C^* R^{-1} d\bar{w}_s$. Then $\Pi_s = 0$ P-a. s.

Now from (2.90) and the definitions of H_1, H_2 it follows that (2.75) and (2.90) are identical with the equations (2.6) and (2.7).

It will be shown now that (2.6) and (2.7) have a unique, continuous, $\mathcal{Y}_t$ measurable solution for any $t \in [0, T]$. $\mathcal{Y}_t$ measurability of m_t and Γ_t follows immediately from the structure of respective equations. Assume now that Γ_t^1 and Γ_t^2, $t \in [0, T]$, are two nonnegative definite, continuous solutions to (2.7). Then if $\Delta_t = \Gamma_t^1 - \Gamma_t^2$, $\tilde{A} = A - H_2R^{-1}C$, $\Pi = C^*R^{-2}C$ it follows that

$$\Delta_t = \int_0^t (\tilde{A}\,\Delta_s + \Delta_s\tilde{A}^* - \Gamma_s^2\Pi\,\Delta_s + \Delta_s\Pi\Gamma_s^1)\,ds\,. \tag{2.91}$$

From (2.91) the following inequality results

$$\|\Delta_t\| \le \int_0^t r_s\,\|\Delta_s\|\,ds\,, \tag{2.92}$$

where

$$r_s = n(2\|\tilde{A}\| + \|\Gamma_s^2\,\Pi\| + \|\Pi\quad\Gamma_s^1\|)\,.$$

In order to apply Lemma A1 to (2.92) it must be shown that P-a.s.

$$\int_0^T r_s\,ds < \infty\,. \tag{2.93}$$

As the first step in this direction it will be shown that for any solution Γ_t to (2.7) P-a.s.

$$\|\Gamma_t\| \le c < \infty\,,\ t \in [0,T]\,. \tag{2.94}$$

From (2.7) it follows that

$$\mathrm{tr}(\Gamma_t) \le \mathrm{tr}(\Gamma_0) + 2\int_0^t \mathrm{tr}(\Gamma_sA)ds + \int_0^t(\|G_1\|^2 + \|G_2\|^2)ds \le$$

$$c_1 + 2\int_0^t \mathrm{tr}(\Gamma_sA)\,ds\,,$$

where

$$c_1 = \mathrm{tr}\,(\Gamma_0) + \int_0^T (\|G_1\|^2 + \|G_2\|^2)ds < \infty .$$

Next, using symmetricity and nonnegative definiteness of Γ_t the following inequalities are obtained:

$$\|\Gamma_t\|^2 \leq (\mathrm{tr}(\Gamma_t))^2 \leq 2c_1^2 + 4\left(\int_0^t \|\Gamma_s\|\ \|A\|\ ds\right)^2 \leq$$

$$2c_1^2 + 4\int_0^T \|A\|^2 ds \int_0^t \|\Gamma_s\|^2 ds .$$

By Lemma A1 $\qquad \|\Gamma_t\|^2 \leq 2c_1^2 \exp(c_2 T)$,

where

$$c_2 = 4\int_0^T \|A\|^2 ds .$$

By (2.3) $c_2 < \infty$ which implies (2.94).

Now note that

$$\int_0^T r(s)ds = n\left(2\int_0^T \|\tilde{A}\|ds + \int_0^T \|\Gamma_s^2\,\Pi\|ds + \int_0^T \|\Pi\ \Gamma_s^1\|ds\right) \leq$$

$$2n\left(\int_0^T \|A\|ds + \int_0^T \|H_2 R^{-1} C\|ds + c\int_0^T \|C^* R^{-1} C\|ds\right) \leq$$

$$n\left(2\int_0^T \|A\|ds + \int_0^T (\|G_1\|^2 + \|G_2\|^2)ds + (c + 4c^3)\int_0^T \|C\|^2 ds\right) < \infty .$$

The above inequalities, which give (2.93), follow from (2.94), (2.3), (2.4) and the inequality

$$\|H_2\|^2 = \mathrm{tr}((G_1R_1^* + G_2R_2^*)R^{-2}(R_1G_1^* + R_2G_2^*)) \leq \mathrm{tr}(G_1G_1^* + G_2G_2^*).$$

Now Lemma A1 can be applied to (2.92) resulting in

$$P(\|\Delta_t\|^2 = 0) = 1, \quad t \in [0, T] .$$

Because of the continuity of Γ_t^1, Γ_t^2 the above implies

$$P(\sup_{0 \leq t \leq T} \|\Delta_t\|^2 = 0) = 1.$$

The above proves the uniqueness of a continuous solution to (2.7).

Assume now that m_t^1 and m_t^2 are continuous solutions to (2.6).

Then

$$m_t^1 - m_t^2 = \int_0^t (\tilde{A} + \Gamma_s \Pi)(m_s^1 - m_s^2)\, ds.$$

Using the results obtained above and Lemma A1 the following is obtained:

$$P(\sup_{0 \leq t \leq T} \|m_t^1 - m_t^2\| = 0) = 1,$$

which proves the uniqueness of a continuous solution to (2.6).

As the final step it will be shown that if $\Gamma_0 > 0$ P-a.s. then Γ_t, $t \in [0, T]$ is uniformly positive definite. Γ_t^{-1} exists for sufficiently small t due to the nonsingularity of the matrix Γ_0 and the continuity, P-a.s., of the matrix Γ_t in $t \in [0, T]$.

Let $\hat{t} = \inf(0 < t \leq T : \det \Gamma_t = 0)$, with $\hat{t} = T$ if $\inf_{0 \leq t \leq T}(\det \Gamma_t) > 0$.

Then for $t < \hat{t} \wedge T$, Γ_t^{-1} exists and is nonnegative definite. To show that all the elements of Γ_t^{-1} are bounded note that for $t < \hat{t} \wedge T$

$$0 = \frac{d}{dt} I = \frac{d}{dt} (\Gamma_t \Gamma_t^{-1}) = \dot{\Gamma}_t \Gamma_t^{-1} + \Gamma_t \dot{\Gamma}_t^{-1} .$$

Hence

$$\dot{\Gamma}_t^{-1} = - \Gamma_t^{-1} \dot{\Gamma}_t \Gamma_t^{-1} ,$$

and from (2.7)

$$\dot{\Gamma}_t^{-1} = - \Gamma_t^{-1} (\Gamma_t A^* + A \Gamma_t + G_1 G_1^* + G_2 G_2^* - \Lambda_t \Lambda_t^*) \Gamma_t^{-1} =$$

$$-\tilde{A}^* \Gamma_t^{-1} - \Gamma_t^{-1} \tilde{A} - \Gamma_t^{-1} H_1 H_1^* \Gamma_t^{-1} + C^* R^{-2} C. \tag{2.95}$$

From (2.95) it follows that

$$\frac{d}{dt} \operatorname{tr} (\Gamma_t^{-1}) \leq -2 \operatorname{tr} (\tilde{A} \Gamma_t^{-1}) + \| R^{-1} C \|^2,$$

or after integration

$$\operatorname{tr}(\Gamma_t^{-1}) \leq \operatorname{tr}(\Gamma_0^{-1}) + \int_0^T \| R^{-1} C \|^2 ds - 2 \int_0^t \operatorname{tr}(\tilde{A} \Gamma_s^{-1}) ds .$$

Nonnegative definiteness of Γ_t^{-1} gives

$$\| \Gamma_t^{-1} \|^2 \leq (\operatorname{tr}(\Gamma_t^{-1}))^2 \leq 2 c_1^2 + 4 c_2 \int_0^t \| \Gamma_s^{-1} \|^2 ds, \tag{2.96}$$

where

$$c_1 = \operatorname{tr}(\Gamma_0^{-1}) + \int_0^T \| R^{-1} C \|^2 ds < \infty ,$$

and $c_2 = \int_0^T \| \tilde{A} \|^2 ds \leq 2 \int_0^T \| A \|^2 ds + \int_0^T \| R^{-1} \|^2 (\| G_1 \|^4 + \| G_2 \|^4 + \| C \|^4) ds < \infty.$

Both of the above inequalities are implied by (2.3) and (2.4). Using Lemma A1 the following boundary for $\|\Gamma_t^{-1}\|$ can be derived from (2.96):

$$\|\Gamma_t^{-1}\|^2 \leq 2\, c_1^2 \exp(4\, c_2 T) < \infty .$$

Boundedness of $\|\Gamma_t^{-1}\|$ for $t < \hat{t} \wedge T$ contradicts definition of $\hat{t}$ unless $\hat{t} = T$. Therefore $P(\hat{t} < T) = 0$ and

$$\|\Gamma_t^{-1}\| \leq c < \infty , \quad t \in [0, T]. \tag{2.97}$$

<u>This ends the proof of Theorem 2.1.</u>

3. AN OPTIMAL CONTROL PROBLEM USING INCOMPLETE DATA

An application of the filtering results derived in Chapter 2 to the stochastic control of a certain class of systems is discussed here. The separation of estimation and control for linear, partially observable systems and quadratic criteria, both having random coefficients is shown to hold. The existence of an optimal control is shown to depend on the existence of a solution to a certain "Riccati-like" partial differential equation.

A control problem is stated as follows. Let the state of some controlled system be described by the stochastic processes $x_t, y_t, z_t, t \in [0, T]$, of the dimensions n, m and k respectively, where only y_t and z_t are being observed. It is assumed that $x_t, y_t, z_t, t \in [0, T]$ admit the following differential representation

$$dx_t = (A(t, z_t)\, x_t + B(t, z_t)u_t)\, dt + G(t, z_t)\, dw_t^1, \tag{3.1}$$

$$dy_t = (C(t, z_t)\, x_t + D(t, z_t))\, dt + R(t, z_t)\, dw_t^2, \tag{3.2}$$

$$dz_t = F(t, z_t)\, dt + H(t, z_t)\, dw_t^3, \tag{3.3}$$

where w_t^i, i = 1, 2, 3, are independent Wiener processes of dimensions ℓ_i respectively. The matrices A, B, G, C, D, R, F, H have the dimensions $n \times n$, $n \times p$, $n \times \ell_i$, $m \times n$, $m \times 1$, $m \times \ell_2$, $k \times 1$, $k \times \ell_3$

respectively. If e denotes an element of any of these matrices then $e(t,\xi)$ as a function of $(t,\xi) \in [0,T] \times R^k$, is assumed to be Borel measurable.

Let $\mathcal{Y}_t$ denotes the σ - algebra generated by $(y_s, z_s : 0 \leq s \leq t)$. The controls u_t of the dimension p, are assumed to be $\mathcal{Y}_t$ measurable for every $t \in [0,T]$. It is desired to minimize the criterion

$$I(u) = E\left(\int_0^T x_t^* Q(t,z_t)x_t + u_t^* P(t,z_t)u_t)dt + x_t^* S(z_t)x_t\right), \qquad (3.4)$$

where the symmetric matrices Q, P, S of the dimensions n x n, p x p, n x n, have again Borel measurable elements and minimization is performed over certain class $\mathcal{U}$ of admissible controls. The control $u \in \mathcal{U}$ is called optimal if

$$I(u) = \inf_{u \in \mathcal{U}} I(u) .$$

One of the interpretations of the above control problem is that included are linear, partially-observable control systems with quadratic criteria which have random coefficients being certain functionals of the Wiener process w_t^3. The controller is getting information about realizations of y_t and z_t. The optimal control in a random feedback form is to be derived.

In order to apply Theorem 2.1 and to obtain the optimal regulator the following assumptions are made.

For $t \in [0,T]$, $\xi, \eta \in R^k$,

$$\|A(t,\xi)\| + \|B(t,\xi)\| + \|C(t,\xi)\| + \|G(t,\xi)\| \leq c < \infty , \quad (3.5)$$

here and below $c < \infty$ denotes a positive constant,

$$\int_0^T \|D(t,\xi)\|^2 dt < \infty , \quad (3.6)$$

$$\|(R(t,\xi)R^*(t,\xi))^{-1}\| + \|(H(t,\xi)H^*(t,\xi))^{-1}\| \leq c < \infty , \quad (3.7)$$

$$\|R(t,\xi) - R(t,\eta)\|^2 + \|H(t,\xi) - H(t,\eta)\|^2 \leq c \|\xi - \eta\|^2, \quad (3.8)$$

$$\|R(t,\xi)\|^2 + \|H(t,\xi)\|^2 \leq c\,(1 + \|\xi\|^2) , \quad (3.9)$$

F and z_0 are such that a unique, strong solution (3.10)

z_t, $t \in [0,T]$ to (3) exists and $P(\int_0^T \|z_t\|^2 dt < \infty) = 1$,

$E(\|x_0\|^4) < \infty$, $P(\|y_0\| < \infty) = 1$, x_0, y_0, z_0 are

independent of w_t^i, $i = 1, 2, 3$, and $P(x_0 \leq a | y_0, z_0)$ (3.11)

is P-a.s. Gaussian,

$Q(t,\xi)$, $S(\xi)$ are non-negative definite matrices, and

$P(t,\xi)$ is uniformly positive definite, i.e., elements (3.12)

of its inverse are uniformly bounded,

the controls $u \in \mathcal{U}$ satisfy

$$\int_0^T E(\|u_t\|^4)\, dt < \infty , \tag{3.13}$$

and are such that (3.1) has a unique, strong solution.

Comments:

Conditions (3.5) to (3.13) are chosen to compromise between generality of the class of systems given by (3.1) to (3.4) and simplicity of the proof of the main theorem. They can be somewhat relaxed with complication of proof. For example, condition $\|G(t,\xi)\| \leq c < \infty$, for $(t,\xi) \in [0,T] \times R^k$ can be replaced by

$$\int_0^T \|G(t,\xi)\|^4\, dt < \infty , \quad \xi \in R^k .$$

Sufficient conditions for (3.10) may take the following form: for $t \in [0,T]$, $\xi, \eta \in R^k$

$$\|F(t,\xi) - F(t,\eta)\|^2 \leq c\|\xi - \eta\|^2 ,$$

$$\|F(t,\xi)\|^2 \leq c(1 + \|\xi\|^2) ,$$

$$E(\|z_0\|^2) < \infty .$$

For $u \in \mathcal{U}$ the following notation is used

$$m_t^u = \mathcal{E}(x_t^u | \mathcal{Y}_t) ,$$

$$\Gamma_t = \mathcal{E}((x_t^u - m_t^u)(x_t^u - m_t^u)^* | \mathcal{Y}_t) ,$$

where x_t^u, y_t^u are corresponding to u solution to (3.1) and (3.2), and Γ_t will be shown to not depend on u.

Theorem 3.1

Let (3.5) to (3.13) be satisfied. If there exists a uniformly bounded solution $V(t,\xi)$, $(t,\xi) \in [0,T] \times R^k$, to the Cauchy problem

$$L(V) + A^* V + VA + Q - VBP^{-1} B^* V = 0 , \tag{3.14}$$

where

$$L(\cdot) = \frac{\partial}{\partial t}(\cdot) + (F^* \frac{\partial}{\partial \xi})(\cdot) + 0.5 \operatorname{tr} (HH^* \frac{\partial^2}{\partial \xi \partial \xi^*})(\cdot) ,$$

(the arguments (t,ξ) are omitted for brevity), and

$$V(T,\xi) = S(\xi) ,$$

then the optimal control exists and is given by

$$\hat{u}_t = -P^{-1}(t,z_t)B^*(t,z_t)V(t,z_t)\hat{m}_t, \quad t \in [0,T] . \tag{3.15}$$

Here $\hat{m}_t = m_t^{\hat{u}}$ and

$$dm_t^u = (Am_t^u + Bu_t)dt + \Lambda\, d\nu_t \,, \tag{3.16}$$

$$d\Gamma_t = (A\Gamma_t + \Gamma_t A^* + GG^* - \Lambda\Lambda^*)\, dt, \tag{3.17}$$

$$\Lambda = \Gamma_t C^*(RR^*)^{-1/2} \,,$$

$$d\nu_t = (RR^*)^{-1/2}\,(dy_t - (Cm_t + D)dt) \,.$$

In the above equations arguments (t, z_t) were omitted.

Proof of Theorem 3.1

Under assumptions made equations (3.16) and (3.17) result straight-forward from the filter equations of the Theorem 2.1. Equation (3.17) does not depend on the choice of u hence $\Gamma_t^u = \Gamma_t$ for all $u \in \mathcal{U}$. This proof involves the following steps:

Step 1: The "candidate" $\tilde{u}_t$ for an optimal control is found using the dynamic stochastic programming approach (see (3.21)).

Step 2: It is checked that $\tilde{u}_t$ given by (3.21) satisfies condition (3.13).

Step 3: It is shown that (3.1) has a unique strong solution for $\tilde{u}_t$ defined by (3.21).

The optimality of $\hat{u}$ given by (3.15) will be proved by stochastic dynamic programming methods. Introduce the value function

$J(t, \alpha, \xi)$ on $[0,T] \times R^n \times R^k$ by the following definition

$$J(t, \alpha, \xi) = \alpha^* V(t,\xi)\alpha + v(t,\xi),$$

where $V(t,\xi)$ is a solution to (3.14) and $v(t,\xi)$ satisfies the Cauchy equation:

$$L(v) + \mathrm{tr}(GG^* V) = 0 , \tag{3.18}$$

$$v(T,\xi) = 0 .$$

Both, the above equation and (3.14) have the same differential operator L . It is assumed tacitly at this point that $G(t,\xi)G^*(t,\xi)$ possesses the same regularity property as $Q(t,\xi)$ so the existence of a solution to (3.14) implies the existence of v.

Using Ito's formula it follows that

$$J(T, x_T^u, z_T) - J(0, x_0, z_0) = \int_0^T ((x_t^u)^* \frac{\partial V}{\partial t} x_t^u + \frac{\partial v}{\partial t} +$$

$$2(Ax_t^u + Bu_t)^* Vx_t^u + \mathrm{tr}(GG^* V) +$$

$$F^* (\frac{\partial}{\partial \xi} ((x_t^u)^* Vx_t^u + v)|_{\xi = z_t}) +$$

$$0.5\ \mathrm{tr}(HH^* (\frac{\partial}{\partial\xi\partial\xi^*} ((x_t^u)^* Vx_t^u + v)|_{\xi = z_t}))) \, dt +$$

$$\int_0^T (2(x_t^u)^* VG\, dw_t^1 + (\frac{\partial}{\partial \xi}((x_t^u)^* V x_t^u + v)|_{\xi = z_t})^* H dw_t^3 . \tag{3.19}$$

Arguments (t, z_t) in (3.19) are omitted. Taking the expectation of the both sides of (3.19) and noticing that the expectation of the second integral in (3.19) is equal zero the following equation is obtained.

$$\mathcal{E}(J(T, x_t^u, z_T) - J(0, x_0, z_0)) - \mathcal{E}\left(\int_0^T (\mathrm{tr}((L(V)|_{\xi=z_t} + A^* V + VA) x_t^u (x_t^u)^*) + \right.$$

$$\left. (L(v)|_{\xi=z_t} + \mathrm{tr}(GG^* V) + 2u_t^* B^* V x_t^u)\, dt. \right. \qquad (3.20)$$

Now as V and v satisfy (3.14) and (3.18), and the relation $\mathcal{Y}_s \subseteq \mathcal{Y}_t$ for $t \geq s$ is satisfied equation (3.20) takes the following form

$$\mathcal{E}(J(T, x_T^u, z_T) - J(0, x_0, z_0)) =$$

$$-\mathcal{E}\left(\int_0^T (\mathrm{tr}((Q - VBP^{-1}B^* V)(m_t^u (m_t^u)^* + \Gamma_t)) + 2u_t^* B^* V m_t^u) dt\right) =$$

$$I_1(u) + I_2 ,$$

where

$$I_1(u) = \mathcal{E}\left(\int_0^T ((m_t^u)^* (VBP^{-1}B^* V - Q) + 2u_t^* B^* V) m_t^u dt\right) ,$$

$$I_2 = \mathcal{E}\left(\int_0^T \mathrm{tr}((VBP^{-1}B^* V - Q)\, \Gamma_t) dt\right)$$

Note that I_2 does not depend on u.

For all $\alpha \in R^n$ and $\eta \in R^p$ the following inequality holds

$$\alpha^* VBP^{-1}B^* V\alpha + 2\eta^* BV\mu \geq -\eta^* P\eta$$

with equality iff

$$\tilde{\eta} = -P^{-1} B^* V \alpha .$$

Hence for

$$\tilde{u}_t = -P^{-1}(t, z_t) B^* (t, z_t) V (t, z_t) \tilde{m}_t , \tag{3.21}$$

where $\tilde{m}_t = m_t^{\tilde{u}}$,

it follows that

$$E (J(T, \tilde{x}_T, z_T) - J(0, x_0, z_0)) = -E(\int_0^T (\tilde{m}_t^* Q\tilde{m}_t + \tilde{u}_t^* P\tilde{u}_t)dt) + I_2, \tag{3.22}$$

where again $\tilde{x}_t = x_t^{\tilde{u}}$.

For any $u \in U$

$$E(J(T, z_T^u z_T) - J(0, x_0, z_0)) \geq -E(\int_0^T ((m_t^u)^* Qm_t^u + u_t^* Pu_t)dt) + I_2. \tag{3.23}$$

Let $I_3 = E(\int_0^T tr(VBP^{-1}B^* V \Gamma_t)dt)$. Notice that I_3 does not depend on u, and that

$$E(J(T, x_T^u, z_T) = E((x_T^u)^* S(z_T)x_T^u) .$$

Now from (3.22) and (3.23) it follows that for all $u \in U$

$$E(J(0,x_0,z_0)) = I(\tilde{u}) - I_3 \leq I(u) - I_3.$$

The above results in the following inequality

$$I(\tilde{u}) \leq I(u)\ , \quad u \in \mathcal{U} \quad . \tag{3.24}$$

To show that $\tilde{u} \in \mathcal{U}$, and hence $\tilde{u} = \hat{u}$ it is first shown that ν_t, which is a Wiener process with respect to ${}_t$ (see Theorem 2.1), does not depend on u. Let x_t^0, y_t^0, m_t^0, ν_t^0 correspond to $u_t \equiv 0$ on $[0,T]$. Denote by $e_t^u = x_t^u - m_t^u$ and $e_0^u = x_0 - m_0$. From (3.1) and (3.16) it follows that

$$de_t^u = Ae_t^u\,dt + G\,dw_t^1 - \Gamma_t\,\Lambda(RR^*)^{-1/2}((Ce_t^u + D)dt + R\,dw_t^2)\ .$$

Hence e_t^u and e_t^0 coincide. Write now

$$d\nu_t^u = (RR^*)^{-1/2}((C\,e_t^u + R\,dw_t^3)\ .$$

The above imply that $\nu_t^u = \nu_t^0 = \nu_t$ P-a.s. (the arguments (t, z_t) are omitted).

Let u be any admissible control. By construction of ν_t the following holds

$$\mathcal{Y}_t^u = \sigma\text{-alg}(y_s^u, z_s : 0 \leq s \leq t) \subseteq \sigma\text{-alg}\,(\nu_s, z_s : 0 \leq s \leq t) = \mathcal{F}_t$$

By (3.21) $\tilde{u}_t$ is $\mathcal{F}_t$ measurable, and by the above $\tilde{u}_t$ is $\mathcal{Y}_t^u$ measurable for any $u \in \mathcal{U}$.

Next it will be shown that $\tilde{u}_t$ satisfies (3.13). First for Γ_t which is a unique continuous, nonnegative definite, symmetric solution to (3.17) (see Theorem 2.1) it follows that

$$\frac{1}{\sqrt{n}} \operatorname{tr}(\Gamma_t) \le \|\Gamma_t\| \le \operatorname{tr}(\Gamma_t), \quad t \in [0,T] .$$

From (3.17) and the above write

$$\|\Gamma_t\| \le \operatorname{tr}(\Gamma_0) + \int_0^t (2\operatorname{tr}(A\,\Gamma_s) + \|G\|^2 - \|\Gamma_s \Lambda\|^2) ds \le c\,(1 + \int_0^t \|\Gamma_s\| ds), t \in [0,T],$$

where

$$0 \le c = \max\,(\operatorname{tr}(\Gamma_0);\ 2\|A\|;\ \int_0^T \|G\| ds) < \infty \text{ P-a. s. by (3.5) and (3.11).}$$

From Lemma A1 and the above

$$\|\Gamma_t\| \le c \exp(c\,T) , \tag{3.25}$$

which shows the uniform, P-a. s., boundedness of Γ_t.

For $\tilde{u}_t$ given by (3.21), (3.16) takes the following form

$$d\tilde{m}_t = (A - BP^{-1}B^*V)\,\tilde{m}_t\,dt + \Gamma_t C^*(RR^*)^{-1/2} d\nu_t, \quad \tilde{m}_0 = m_0 . \tag{3.26}$$

From (3.5), (3.7), (3.12) and (3.25) it follows that

$$\|A - BP^{-1}B^*V\| \le c_1 < \infty ,$$

and

$$\|\Gamma_t\,C^*(RR^*)^{-1/2}\| \le c_1 < \infty .$$

Now by Theorem A6 (3.26) has a unique, strong solution. To prove that $\int_0^T E(\|\tilde{u}_t\|^4)dt < \infty$, it is enough to show that

$$\int_0^T E(\|\tilde{m}_t\|^4)dt < \infty \text{ . (see (3.5) and (3.12))} . \tag{3.27}$$

Write (3.26) in the following form

$$d\tilde{m}_t = \tilde{A}\,\tilde{m}_t\,dt + \tilde{B}\,d\nu_t\ ,$$

where

$$\tilde{A} = A - BP^{-1}B^{*}V\ ,$$

$$\tilde{B} = \Gamma_t\,C^{*}(RR^{*})^{-1/2}\quad .$$

By Ito's formula $\|\tilde{m}_t\|^4 = \|m_0\|^4 + 4\int_0^t \|\tilde{m}_s\|^2 \tilde{m}_s^{\,*}\tilde{A}\tilde{m}_s\,ds +$

$$2\int_0^t (\|\tilde{m}_s\|^2 \mathrm{tr}(\tilde{B}\tilde{B}^{*}) + \mathrm{tr}(\tilde{B}\tilde{B}^{*}\tilde{m}_s\tilde{m}_s^{\,*}))ds + 4\int_0^t \|\tilde{m}_s\|^2 \tilde{m}_s^{\,*}\tilde{B}d\nu_s\ . \tag{3.28}$$

Define $\tau_N = \inf\,(t: \sup_{0\le s\le t}\|\tilde{m}_s\| \ge N)$, assuming $\tau_N = T$ if $\sup_{0\le s\le T}\|\tilde{m}_s\| < N$.

With the above definition of τ_N and the boundedness of $\tilde{A}$ and $\tilde{B}$ the following inequality can be derived after taking the expectation of the both sides of (3.28)

$$E(\|\tilde{m}_{t\wedge\tau_N}\|^4) \le E(\|m_0\|^4) + 4E(c_1\int_0^{t\wedge\tau_N} (\|\tilde{m}_s\|^4 + c_1\|\tilde{m}_s\|^2)ds) \le$$

$$E(\|m_0\|^4) + c_1^3 T + 8 c_1 \int_0^t E(\|\tilde{m}_{s\wedge\tau_N}\|^4) ds .$$

Lemma A1 applied to the above gives

$$E(\|\tilde{m}_{t\wedge\tau_N}\|^4) \le (E(\|m_0\|^4) + c_1^3 T) \exp(8 c_1 t) .$$

By Fatou's Lemma

$$E(\|\tilde{m}_t\|^4) \le \liminf_{N\to\infty} E(\|\tilde{m}_{t\wedge\tau_N}\|^4) \le (E(\|m_0\|^4 + c_1^3 T)\exp(8 c_1 t).$$

From the above and (3.11) it follows that

$$E(\|\tilde{m}_t\|^4) \le (E\|x_0\|^4 + c_1^3 T) \exp(8c_1 T) < \infty , \qquad (3.29)$$

which proves (3.27).

As the final step it will be shown that (3.1) has for $\tilde{u}_t$ a strong unique solution. To show the uniqueness assume that x_t^1 and x_t^2 are two continuous strong solutions to (3.1) with the same initial condition x_0. It follows that

$$x_t^1 - x_t^2 = \int_0^t A(x_s^1 - x_s^2) ds, \ t \in [0,T] .$$

Denote

$$\chi_t^p = \chi(\sup_{0\le s\le t}(\|x_s^1\|^2 + \|x_s^2\|^2) \le p) ,$$

where $\chi(\cdot)$ denotes the characteristic function.

Since $\chi_t^p = \chi_t^p \chi_s^p$ for $t \geq s$ then

$$\chi_t^p \| x_t^1 - x_t^2 \| = \chi_t^p \| \int_0^t \chi_s^p A(x_s^1 - x_s^2) ds \| . \tag{3.30}$$

From the definition of χ_t^p it follows that $\chi_t^p \| x_t^1 - x_t^2 \|$ is bounded, and consequently the mathematical expectation of the both sides of (3.30) exists. Hence

$$E(\chi_t^p \| x_t^1 - x_t^2 \|) \leq \sqrt{n} \int_0^t E(\chi_s^p \| A(x_s^1 - x_s^2) \| ds \leq$$

$$\sqrt{n} \cdot c \int_0^T E(\chi_s^p \| x_s^1 - x_s^2 \|) ds, \quad t \in [0, T].$$

The last inequality is implied by (3.5). Applying Lemma A1 to the above it follows that

$$E(\chi_t^p \| x_t^1 - x_t^2 \|) = 0, \text{ for all } t \, [0, T].$$

Therefore

$$P(\| x_t^1 - x_t^2 \| > 0) \leq P(\sup_{0 \leq s \leq T} (\| x_s^1 \|^2 + \| x_s^2 \|^2) > p) .$$

The last term of the above inequality goes to zero as p goes to infinity, because of the continuity of the processes x_t^1 and x_t^2. Hence for any $t \in [0, T]$,

$$P(\| x_t^1 - x_t^2 \| > 0) = 0,$$

and therefore for any countable, dense set $\mathcal{D} \subset [0, T]$

$$P(\sup_{t \in \mathcal{D}} \| x_t^1 - x_t^2 \| > 0) = 0.$$

Using again the continuity of the processes x_t^1 and x_t^2 the following equality

$$P(\sup_{0 \leq t \leq T} \| x_t^1 - x_t^2 \| > 0) = P(\sup_{t \in \mathcal{D}} \| x_t^1 - x_t^2 \| > 0) = 0,$$

proves the uniqueness of a strong, continuous solution to (3.1).

To show the existence of a solution to (3.1) write (3.1) in the following equivalent form

$$dx_t = df_t + Ax_t \, dt , \tag{3.31}$$

where $df_t = B\tilde{u}_t \, dt + Gdw_t^1$, $f_0 = x_0$.

First it will be shown that

$$E(\sup_{0 \leq t \leq T} \| f_t \|^2) < \infty . \tag{3.32}$$

From the definition of f_t the following inequalities result

$$\| f_t \|^2 \leq 3(\| f_0 \|^2 + \| \int_0^t B\tilde{u}_s \, ds \|^2 + \| \int_0^t G \, dw_s^1 \|^2) ,$$

and

$$E(\sup_{0\le t\le T}\|f_t\|^2) \le 3(\,E(\|x_0\|^2) + E(\sup_{0\le t\le T}\|\int_0^t B\tilde{u}_s ds\|^2) +$$

$$E(\sup_{0\le t\le T}\|\int_0^t G dw_s^1\|^2\,). \tag{3.33}$$

To show that the second term of the right hand side of (3.33) is finite it is enough to show that

$$E(\sup_{0\le t\le T}\int_0^t \|\tilde{m}_s\|^2\, ds) < \infty .$$

From (3.29) it follows that $\sup_{0\le t\le T} E(\|\tilde{m}_t\|^2) < \infty$. (3.34)

It will be shown that actually

$$E(\sup_{0\le t\le T}\|\tilde{m}_t\|^2) < \infty . \tag{3.35}$$

From (3.26) write

$$\sup_{0\le t\le T}\|m_t\|^2 \le 3(\|m_0\|^2 + c^2\, nT\int_0^T \|\tilde{m}_s\|^2 ds + \sup_{0\le t\le T}\|\int_0^t \tilde{B} d\nu_s\|^2). \tag{3.36}$$

Denote $V_t^{ij} = \int_0^t [\tilde{B}]_{ij}\ [d\nu_s]_j$. According to Lemma A4 (a = 2) and (3.5)

$$E((V_t^{ij})^4) \le 36\, T^2\, c^4 ,$$

which implies

$$E((V_t^{ij})^2) \leq 6\ T\ c^2 .$$

Using now Lemma A6 (a = 2) and the above the following inequality is obtained.

$$E(\sup_{0\leq t\leq T}(V_t^{ij})^2) \leq 24T\ c^2 ,$$

and since

$$\Big\|\int_0^t \tilde{B}\, d\nu_s\Big\| = \sum_{i=1}^{n} (\sum_{j=1}^{m} V_t^{ij})^2 \leq m \sum_{i=1}^{n}\sum_{j=1}^{m} (V_t^{ij})^2$$

then from (3.34) and (3.36) it follows that

$$E(\sup_{0\leq t\leq T}\|\tilde{m}_t\|^2) \leq 3(E(\|m_0\|^2) + c^2T^2n \sup_{0\leq t\leq T} E(\|\tilde{m}_t\|^2) + 24nm^2Tc^2) < \infty .$$

The above proves (3.35). To show that (3.32) holds it remains to prove that the third term of the right hand side of (3.33) is bounded. Similarly like in the proof of Theorem 2.1 (see Chapter 2, (2.23)).

$$E(\Big\|\int_0^t Gdw_s^1\Big\|^4) \leq 36Tn\ \ell_1^3 \int_0^T E(\|G\|^4)ds \leq 36T^2n\,\ell_1^3c^4 ,$$

and by Lemma A6

$$E(\sup_{0\leq t\leq T}\Big\|\int_0^t Gdw_s^1\Big\|^2) \leq 24\ T\ c^2\ell_1\ \sqrt{n}\,\ell_1 < \infty .$$

The last inequality which follows from (3.5) is the last step in proving that (3.32) holds.

Consider now the sequence of continuous process $x_t^{(i)}$, $i = 0, 1, \ldots, t \in [0, T]$, defined by

$$x_t^{(0)} = f_t$$

$$x_t^{(i)} = f_t + \int_0^t A\, x_s^{(i-1)}\, ds \quad , \quad i = 1, 2, \ldots \tag{3.37}$$

To show that $E(\| x_t^{(i)} \|^2) \leq c < \infty$, where c depends neither on i nor on t, write using (3.32)

$$g^{(i)}(t) = E(\| x_t^{(i)} \|^2) \leq 2(E(\| f_t \|^2) + nc^2 \int_0^t E(\| x_t^{(i-1)} \|^2) ds) .$$

From the above and (3.32) it follows that

$$g^{(i)}(t) \leq c_1 (1 + \int_0^t g^{(i-1)}(s) ds), \tag{3.38}$$

where

$$c_1 = 2 \max (nc^2 ; \sup_{0 \leq t \leq T} E(\| f_t \|^2)) < \infty .$$

By mathematical induction (3.38) results in

$$g^{(i)}(t) \leq c_1 \exp (c_1 t) \leq c_1 \exp (c_1 T) . \tag{3.39}$$

Consider now the following inequality implied by (3.37)

$$h^{(i+1)}(t) = E(\|x_t^{(i+1)} - x_t^{(i)}\|^2) = E(\|\int_0^t A(x_s^{(i)} - x_s^{(i-1)})ds\|^2) \leq nc^2 \int_0^t h^{(i)}(s)ds. \tag{3.40}$$

Since $E(\sup_{0\leq t\leq T}\|x_t^{(1)} - x_t^{(0)}\|^2) = E(\sup_{0\leq t\leq T}\|\int_0^T A\ f_s ds\|^2) \leq$

$$nc^2\ E(\sup_{0\leq t\leq T}\|f_t\|^2) \leq c_2 < \infty ,$$

it follows from (3.40) that

$$h^{(i+1)}(t) \leq \frac{c_2\ (nc^2 t)^i}{i!} . \tag{3.41}$$

Next $\sup_{0\leq t\leq T}\|x_t^{(i+1)} - x_t^{(i)}\|^2 \leq nc^2 \sup_{0\leq t\leq T}\int_0^t \|x_s^{(i)} - x_s^{(i-1)}\|^2 ds \leq$

$$nc^2 \int_0^T \|x_s^{(i)} - x_s^{(i-1)}\|^2 ds,$$

which together with the inequality (3.41) give

$$E(\sup\|x_t^{(i+1)} - x_t^{(i)}\|^2) \leq nc^2 \int_0^T \frac{c_2(nc^2 t)^{i-1}}{(i-1)!} = c_2 \frac{(nc^2 T)^i}{i!} .$$

Markov inequality gives

$$\sum_{i=1}^{\infty} P(\sup_{0\leq t< T}\|x_t^{(i+1)} - x_t^{(i)}\| \geq \frac{1}{i^2}) \leq c_2 \sum_{i=1}^{\infty} \frac{(nc^2 T)^i}{i!}\ i^4 < \infty .$$

Hence by Borel-Cantelli Lemma the series

$$\|x_t^{(0)}\| + \sum_{i=0}^{\infty} \| x_t^{(i+1)} - x_t^{(i)} \| \ ,$$

converges P-a.s. uniformly over $t \in [0, T]$. Therefore the sequence of the random processes $(x_t^{(i)})$, $t \in [0, T]$, $i = 0, 1, \ldots,$ converges P-a.s. uniformly to the continuous process x_t given by

$$x_t = x^{(0)} + \sum_{i=0}^{\infty} (x_t^{(i+1)} - x_t^{(i)}), \quad t \in [0, T] \ . \tag{3.42}$$

From (3.31) and Fatou's lemma it follows that

$$E(\|x_t\|^2) \leq c_1 \exp(c_1 T) \ .$$

To show that x_t defined by (3.42) is the solution to (3.1) for $t \in [0, T]$, i.e. for each $t \in [0, T]$, P-a.s.

$$x_t - f_t - \int_0^t A x_s \, ds = 0 \ , \tag{3.43}$$

define

$$e_t^{(i+1)} = x_t - x_t^{(i+1)} + \int_0^t A(x_s^{(i)} - x_s) ds \ .$$

By (3.37), (3.43) will be proved if it is shown that $e_t^{(i)}$ converges in probability to zero as i goes to infinity. But

$$\left\| \int_0^t A(x_s^{(i)} - x_s) ds \right\|^2 \leq nc^2 T \sup_{0 \leq t \leq T} \| x_t^{(i)} - x_t \|^2$$ and as was shown

above

$$\forall \ \epsilon > 0 \ \ P\left(\sup_{0 \le t \le T} \| x_t^{(i)} - x_t \|^2 > \epsilon\right) \to 0 \text{ as } i \to \infty .$$

Hence $\|e_t^{(i)}\|$ converges in probability to zero as $i \to \infty$. This ends the proof of the existence of a strong solution to (3.1). This ends also proof that $\tilde{u} \in \mathcal{U}$ and since (3.24) holds for arbitrary $u \in \mathcal{U}$ it is concluded that $\tilde{u} = \hat{u}$.

This ends the proof of Theorem 3.1.

4. ON A RICCATI-LIKE EQUATION OF STOCHASTIC CONTROL

The sufficient condition for the existence and uniqueness of a solution to the Cauchy problem stated in Theorem 3.1 (Chapter 3, (3.14)) is given. The result presented here is a generalization of a solution to the Riccati differential equation. To state the main theorem of this chapter it is necessary to define two Banach spaces with elements which are matrix functions $A(t,\xi)$, $(t,\xi) \in [0,T] \times R^k$ and $B(\xi)$, $\xi \in R^k$ respectively. With the norms defined here these spaces are called Holder spaces and are denoted by H_i^r, $i = 1, 2$, where r is a certain, positive noninteger number.

Definition 4.1

A matrix function $A(t,\xi)$ is said to belong to H_1^r if all elements of $A(t,\xi)$ are continuous on $[0,T] \times R^k$, together with continuity of all derivatives $\partial_t^i \partial_\xi^j$, $2i + j < r$, where ∂_t^i denotes the partial derivative with respect to t of order i, ∂_ξ^j denotes any derivative with respect to ξ of order j, and A has a finite norm defined as follows:

$$\|A\|_{H_1} = \|A\|_r + \sum_{q=0}^{[r]} \|A\|_q ,$$

where

$$\|A\|_0 = \max_{(t,\xi)\in[0,T]\times R^k} \|A(t,\xi)\|,$$

$$\|A\|_q = \sum_{(2i+j=q)} \|\partial_t^{\,i}\partial_\xi^{\,j} A\|_0 .$$

In the above, summation is taken over all possible combinations of i and j such that $2i + j = q$. $[r]$ denotes the largest integer value such that $[r] < r$.

Further $\|A\|_r$ is defined by

$$\|A\|_r = \|A\|_\xi + \|A\|_t ,$$

where

$$\|A\|_\xi = \sum_{(2i+j=[r])} \|\partial_t^{\,i}\partial_\xi^{\,j} A\|_\xi^{(r-[r])} ,$$

$$\|A\|_t = \sum_{(0< r-2i-j< 2)} \|\partial_t^{\,i}\partial_\xi^{\,j} A\|_t^{\left(\frac{r-2i-j}{2}\right)},$$

(again summations are taken over all possible combinations of i and j such that $2i + j = [r]$, and $0 < r - 2i - j < 2$), and for $0 < a < 1$

$$\|A\|_\xi^{(a)} = \sup_{(t,\xi),(t,\eta)\in[0,T]\times R^k} \frac{\|A(t,\xi)-A(t,\eta)\|}{\|\xi-\eta\|^a} ,$$

$$\|A\|_t^{(a)} = \sup_{(t,\xi),(s,\xi) \in [0,T] \times R^k} \frac{\|A(t,\xi)-A(s,\xi)\|}{|t-s|^a}$$

Definition 4.2

A matrix function $B(\xi)$ is said to belong to $H_2^{\,r}$ if all elements of $B(\xi)$ are continuous functions on R^k together with continuity of all derivatives $\partial_\xi^{\,j}$ up to order $[r]$, and B has a finite norm defined as follows:

$$\|B\|_{H_2} = |B|_r + \sum_{j=0}^{[r]} |B|_j \ ,$$

where

$$|B|_0 = \max_{\xi \in R^k} \|B(\xi)\|$$

$$|B|_j = \sum_{(j)} |\partial_\xi^{\,j} B|_0$$

$$|B|_r = \sum_{([r])} |\partial_\xi^{[r]} B|^{(r-[r])}$$

and for $0 < a < 1$,

$$|B|^{(a)} = \sup_{\xi,\eta \in R^k} \frac{\|B(\xi)-B(\eta)\|}{\|\xi-\eta\|^a} \ .$$

Remark: The last equations of definitions 4.1 and 4.2 may be given with an additional condition, $\|\xi-\eta\| \leq \rho$, where ρ is a certain positive number. However the norms defined on H_1^r and H_2^r for different values of ρ are equivalent and hence their dependence on ρ is not indicated.

Now the following result can be stated.

Theorem 4.1

Let $r > 0$ be a noninteger number. If $A(t,\xi)$, $B(t,\xi)$, $F(t,\xi)$, $H(t,\xi)$, $P(t,\xi)$, $Q(t,\xi)$ belong to H_1^r, $Q(t,\xi)$ is nonnegative definite and $P(t,\xi)$ is uniformly positive definite on $[0,T] \times R^k$, then for any $S(\xi) \in H_2^{r+2}$ which is nonnegative definite on R^k the following Cauchy problem

$$\frac{\partial V}{\partial t} + L(V) + A^* V + VA + Q - VBP^{-1} B^* V = 0, \tag{4.1}$$

where

$$L(V) = (F^* \frac{\partial}{\partial \xi}) V + 0.5 \operatorname{tr} (HH^* \frac{\partial^2}{\partial_\xi \partial_\xi^*}) V,$$

$$V = V(t,\xi), \ (t,\xi) \in [0,T] \times R^k \text{ and } V(T,\xi) = S(\xi), \ \xi \in R^k,$$

has a unique solution $V(t,\xi)$ which belongs to H_1^{r+2}. Also, this solution is nonnegative definite and uniformly bounded on $[0,T] \times R^k$.

Proof of Theorem 4.1

By the transformation $s = T-t$, (4.1) becomes the classical

Cauchy problem with the initial, instead of the terminal, condition $S(\xi)$

$$\frac{\partial V}{\partial s} = L(V) + A^* V + VA + Q - VBP^{-1} B^* V, \tag{4.2}$$

$$V|_{s=0} = S(\xi) .$$

All the matrix functions in the above have the arguments of the form $(T-s, \xi)$, $(s, \xi) \in [0, T] \times R^k$.

Write (4.2) in the more convenient form:

$$\overset{o}{V} = L(V) + N(V, K) , \tag{4.3}$$

where $\overset{o}{V} = \frac{\partial V}{\partial s}$, $K = P^{-1} B^* V$,

$$N(V, K) = (A-BK)^* V + V(A-BK) + K^* PK + Q .$$

Further, for any matrix $\tilde{K}$ of the same dimension as K, it follows that

$$(A-BK)^* V + V(A-BK) + K^* PK =$$

$$(A-B\tilde{K})^* V + V(A-B\tilde{K}) + \tilde{K}^* P\tilde{K} - (\tilde{K}-K)^* P (\tilde{K}-K), \tag{4.4}$$

which implies

$$N(V, K) \leq N(V, \tilde{K}) ,$$

with equality only if $\tilde{K} = K$.

Define now the following sequence of matrices (V^i), $i = 1, 2, \ldots,$ where V^1 is a solution to the Cauchy problem

$$\overset{o}{V}{}^1 = L(V^1) + A^*V^1 + V^1A + Q,$$

$$V^1|_{s=0} = S(\xi)$$

and for i = 2, 3, ..., V^i is a solution to

$$\overset{o}{V}{}^i = L(V^i) + N(V^i, K^i), \tag{4.5}$$

$$V^i|_{s=0} = S(\xi), \ K^i = P^{-1}B^*V^{i-1}.$$

Now the following results will be used:

Lemma 4.1

[Kl, p. 189] Let (V^i), i = 1, 2, ..., be a sequence of n x n symmetric matrices such that

$$V^1 \geq V^2 \geq \ldots, \text{ and } V^i \geq V, \ i = 1, 2, \ldots,$$

for some matrix V. Then $V^\infty = \lim_{i\to\infty} V^i$ exists and $V^\infty \geq V$.

△

Lemma 4.2 [L3, Theorem 10.2, p. 617]

Let the assumptions of Theorem 4.1 hold and Let $K^{i-1} \in H_1^{\,r}$, then the Cauchy problem (4.5) has for i = 2, 3, ..., a unique solution in the class of matrix functions from $H_1^{\,r+2}$, and this solution is subject to the inequality

$$\|V^i\|_{H_1^{r+2}} \leq c(\|(K^{i-1})^* P K^{i-1} + Q\|_{H_1^r} + \|S\|_{H_2^{r+2}}),$$

with the constant $c < \infty$ not depending on K^i, P, Q and S.

$\triangle$

Notice that for $i = 1$, $K^0 = 0$ can be taken and the result of Lemma 4.2 holds giving $V^1 \in H_1^{r+2}$ and the inequality

$$\|V^1\|_{H_1^{r+2}} \leq c\,(\|Q\|_{H_1^r} + \|S\|_{H_2^{r+2}}) = c_1 < \infty .$$

Now $K^2 = P^{-1}B^*V^1$ belongs to ${}_1^{r}$, which by Lemma 4.2 implies that V^2 belongs to H_1^{r+2}. Continuing this procedure for $i = 2, 3, \ldots,$ it follows that $V^i \in H_1^{r+2}$.

Next the following result is proven to hold.

<u>Lemma 4.3</u>

Suppose that the Cauchy problem

$$\overset{o}{W} = L(W) + \tilde{A}W + W\tilde{A}^* + \tilde{Q}, \tag{4.6}$$

where $\tilde{Q} = \tilde{Q}^* \geq 0$ and $W(0,\xi) = S(\xi) = S^*(\xi) \geq 0$,

has a solution $W(t,\xi)$, $(t,\xi) \in [0,T] \times R^k$, such that $W \in H_1^{r+2}$ for some noninteger number $r > 0$. Then $W(t,\xi) \geq 0$ on $[0,T] \times R^k$.

Proof of Lemma 4.3

The following auxiliary result will be used.

Let λ_i, $i = 1, \ldots, n$, denote the real eigenvalues of a symmetric matrix W, ordered as follows $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. It will be shown that if for some $\alpha \in R^n$

$$\alpha^* W \alpha = \lambda_1 \| \alpha \|^2, \tag{4.7}$$

then

$$\alpha^* \tilde{A} W \alpha = \lambda_1 \alpha^* \tilde{A} \alpha, \tag{4.8}$$

for any n x n matrix $\tilde{A}$.

It is seen that for any real, symmetric matrix W there exists a nonsingular matrix U with the property $U^{-1} = U^*$, and such that $\hat{W} = UWU^*$ is a diagonal matrix with the real eigenvalues of W on the diagonal (see [Cl, p. 412]). Consequently, (4.7) may be rewritten

$$\alpha^* U^* UWU^* U\alpha = \eta^* \hat{W}\eta = \sum_{i=1}^{n} \lambda_i |\eta_i|^2 = \sum_{i=1}^{n} \lambda_1 |\eta_i|^2, \tag{4.9}$$

where

$\eta = U \alpha \in R^n$, and the last equality follows from (4.7) and

$$\|\eta\|^2 = \eta^* \eta = \alpha^* U^* U \alpha = \|\alpha\|^2 .$$

From (4.9) it follows that if $\lambda_i > \lambda_1$ for some $i = 2, \ldots, n$, then $\eta_j \equiv 0$ for $j = i, \ldots, n$. This gives η of the form

$$\eta^* = (\eta_1, \ldots, \eta_{i-1}, 0, \ldots, 0),$$

and
$$\lambda_1 = \lambda_2 = \ldots = \lambda_{i-1}$$

Rewrite (4.8) in the equivalent form

$$\alpha^* \tilde{A} W \alpha = \alpha^* U^* U \tilde{A} U^* U W U^* U \alpha = \eta^* U \tilde{A} U^* \hat{W} \eta =$$

$$\sum_{q=1}^{i-1} \sum_{p=1}^{i-1} \lambda_q [U\tilde{A}U^*]_{qp} \eta_q \eta_p = \lambda_1 \sum_{q=1}^{i-1} \sum_{p=1}^{i-1} [UAU^*]_{qp} \eta_q \eta_p =$$

$$\lambda_1 \eta^* U\tilde{A}U^* \eta = \lambda_1 \alpha^* \tilde{A} \alpha .$$

The last equality proves that (4.7) implies (4.8).

Now let $W^\epsilon(0,\xi) = S(\xi) + \epsilon I$ for some $\epsilon > 0$. It is obvious that if a solution to (4.6) exists with the initial condition $S(\xi)$, then there also exists a solution W^ϵ with the above initial condition for any $\epsilon > 0$.

Because of the continuity of $W^\epsilon(t,\xi)$, there exists $\hat{t} > 0$ such that

$$\hat{t} = \sup_{t \in [0,T]} (t : W(s,\xi) \geq 0, \ (s,\xi) \in [0,t] \times R^k). \tag{4.10}$$

Denote by $g(t,\xi,\alpha) = \alpha^* W^\epsilon(t,\xi)\alpha$. $g(t,\xi,\alpha)$ is a continuous function on $[0,T] \times R^k \times R^n$, and has continuous first order t-derivatives and second order ξ-derivatives. $g(t,\xi,\alpha)$ is nonnegative on $[0,\hat{t}] \times R^k \times R^n$, and for some $(\hat{t}, \hat{\xi}, \hat{\alpha})$ is equal to zero by (4.10). Hence $(\hat{t},\hat{\xi},\hat{\alpha})$ is a minimum point of g in the domain $[\hat{t}] \times R^k \times [\hat{\alpha}]$. It implies that

$$\frac{\partial g(\hat{t},\xi,\hat{\alpha})}{\partial \xi} \Big|_{\xi=\hat{\xi}} = 0, \tag{4.11}$$

and
$$\frac{\partial^2 g(\hat{t},\xi,\hat{\alpha})}{\partial_\xi \partial_\xi^*} \Big|_{\xi=\hat{\xi}} \geq 0 . \tag{4.12}$$

It will be shown that

$$\operatorname{tr}(H(\hat{t},\hat{\xi})H^*(\hat{t},\hat{\xi}) \frac{\partial^2}{\partial_\xi \partial_\xi^*}) \, g(\hat{t},\xi,\hat{\alpha})\Big|_{\xi=\hat{\xi}} \geq 0 . \tag{4.13}$$

By a linear transformation $\eta = U^*\xi$, where $U^{-1} = U^*$, and $UHH^*U^*=\hat{H}$, where $\hat{H}$ is a diagonal matrix with the real, positive eigenvalues λ_i, $i = 1, \ldots, k$, of HH^*, on the diagonal (see [C1]), R^k is transformed onto (U^* is nonsingular) R^k. Now the left hand side of (4.13) can be rewritten as follows:

$$\operatorname{tr}((H^* \frac{\partial}{\partial \xi})^* (H^* \frac{\partial}{\partial \xi})) \, g(\hat{t},\xi,\hat{\alpha})\Big|_{\xi=\xi} =$$

$$\operatorname{tr}((H^*U^* \frac{\partial}{\partial \eta})^* (H^*U^* \frac{\partial}{\partial \eta}))g(\hat{t},U\eta,\hat{\alpha})\Big|_{\eta=\hat{\eta}=U^*\hat{\xi}} =$$

$$\operatorname{tr}(UHH^* \frac{\partial^2 g(\hat{t},U\eta,\hat{\alpha})}{\partial \eta \, \partial \eta^*}\Big|_{\eta=\hat{\eta}}) . \tag{4.14}$$

Obviously $(\hat{t}, \hat{\eta}, \hat{\alpha})$ is a minimum point of $g(\hat{t}, U\eta, \hat{\alpha})$ in $[\hat{t}] \times R^k \times [\hat{\alpha}]$ which implies that

$$\frac{\partial^2 g(\hat{t}, U\eta, \hat{\alpha})}{\partial \eta \partial \eta^*} \Big|_{\eta = \hat{\eta}} \geq 0,$$

and from (4.14) it follows that

$$\mathrm{tr}(\hat{H} \frac{\partial^2 g}{\partial \eta \partial \eta^*} \Big|_{(\eta=\hat{\eta})} \sum_{i=1}^{k} \lambda_i \frac{\partial^2 g}{\partial \eta_i^2} \Big|_{\eta = \hat{\eta}} \geq 0,$$

which proves (4.13).

Now $g(\hat{t}, \hat{\xi}, \hat{\alpha}) = \hat{\alpha}^* W^{\epsilon}(\hat{t}, \hat{\xi})\hat{\alpha} = 0 = \lambda_1 \|\hat{\alpha}\|^2$,

where λ_1 denotes the smallest eigenvalue of the nonnegative definite matrix $W^{\epsilon}(\hat{t}, \hat{\xi})$.

According to (4.7) and (4.8) the above gives

$$\hat{\alpha}^* (\tilde{A}(\hat{t}, \hat{\xi})\, W(\hat{t}, \hat{\xi}) + W(\hat{t}, \hat{\xi})\tilde{A}^*(\hat{t}, \hat{\xi}))\hat{\alpha} = 0 . \tag{4.15}$$

From (4.6), (4.11), (4.13) and (4.15) it follows that

$$\overset{o}{g}(\hat{t}, \hat{\xi}, \hat{\alpha}) \geq 0 .$$

The above contradicts definition of $\hat{t}$ unless $\hat{t} = T$. Finally, letting ϵ go to zero

ends the proof of Lemma 4.3.

Using the result of Lemma 4.3, it will be shown that

$$V^i \geq V^{i+1}, \text{ for } i = 1, 2, 3, \ldots \tag{4.16}$$

Denote

$$M^{i+1} = V^i - V^{i+1} .$$

From (4.5) the following equation is obtained.

$$\overset{o}{M}^{i+1} = L(M^{i+1}) + N(V^i, K^i) - N(V^{i+1}, K^{i+1}).$$

Then from (4.4) it follows that

$$\overset{o}{M}^{i+1} = L(M^{i+1}) + (A-BK^i)^* M^{i+1} + M^{i+1}(A-BK^i) + (K^{i+1}-K^i)^* P(K^{i+1}-K^i),$$

and

$$M^{i+1}(0, \xi) = 0 .$$

From Lemma 4.3 it follows that $M^{i+1} \geq 0$ proving (4.16). Now by Lemma 4.1 $V(t,\xi) = \lim_{i \to \infty} V^i(t,\xi)$ exists and since $V^i \geq 0$ for $i = 1, 2, 3, \ldots$ (see 4.5) and apply Lemma 4.3) it follows that

$$V(t,\xi) \geq 0, \ (t,\xi) \in [0,T] \times R^k .$$

Further for $i = 1, 2, 3, \ldots$ the nonnegative definiteness of V^i implies

$$\|V^i\| \leq tr(V^i) \leq tr(V^1) \leq \sqrt{n} \ \|V^1\| \leq c_1 \sqrt{n}$$

which by (4.5) gives

$$\|K^i\| \leq c < \infty .$$

Let $\Phi\,(t,\xi\,;s,\eta)$ be the fundamental solution to

$$\overset{o}{W} = L(W) .$$

Then the following equation can be written

$$V^i(t,\xi) = \int_{R^k} \Phi(t,\xi;0,\eta)S(\eta)d\eta + \int_0^t \int_{R^k} \Phi(t,\xi;s,\eta)N(V^i(s,\eta),K^i(s,\eta))dsd\eta .$$

Notice that by Lemma 4.2 $\Phi \in H_1^{r+2}$. Applying the dominated convergence theorem to the above integral it is concluded that it holds with V^i, K^i replaced by V and K respectively where $K = \lim_{i\to\infty} K^i = P^{-1}B^*V$. This shows that V is a solution to (4.1) and that $V \in H_1^{r+2}$.

To show the uniqueness of a continuous solution to (4.1), is equivalent to proving that if V^1 and V^2 are two such solutions then $M = V^1 - V^2 = 0$ on $[0,T] \times R^k$.

According to (4.3) and (4.4), M satisfies the following equation

$$\overset{o}{M} = L(M) + N(V^1, K^1) - N(V^2, K^2) =$$

$$L(M) + (A-BK^1)^* M + M(A-BK^1) + (K^1-K^2)^* P(K^1-K^2),$$

$$M(0,\xi) = 0, \ \xi \in R^k.$$

From Lemma 4.3 it follows that $M \geq 0$ on $[0, T] \times R^k$. Denote by $M^- = -M = V^2 - V^1$ and using again (4.3) and (4.4) write

$$\overset{o}{M}{}^- = L(M^-) + (A-BK^2)^* M^- + M^-(A-BK^2) + (K^2-K^1)^* P(K^2-K^1),$$

$$M^-(0,\xi) = 0, \ \xi \in R^k.$$

From the above and Lemma 4.3 it follows that $M^- \geq 0$ on $[0, T] \times R^k$. Taking into account symmetry of V^1 and V^2 this results in $M = 0$ on $[0, T] \times R^k$ proving the desired uniqueness of a solution to (4.1).

<u>This ends the proof of Theorem 4.1.</u>

5. CONCLUDING REMARKS

To illustrate the result of Theorem 3.1, digital simulation of a simple continuous stochastic control system is presented. An implementation of a continuous stochastic integral on a computer creates some difficulties. A "convergence" criteria of a series of pseudo-random variables to a stochastic process is an example of such a difficulty. Also the generation of a large population of "independent" samples from "random" number generators suffers from pseudo-randomness and cross-correlation effects. This inconsistency of a discrete representation of a continuous stochastic process should be kept in mind while interpreting the results obtained in the example below.

<u>Example</u>

Consider the following stochastic differential equation:

$$dx_t = a(w_t)x_t \, dt + u_t \, dt, \quad x_0 = \text{const}, \tag{5.1}$$

where w_t, $t \in [0,1]$, is a Wiener process, and

$$a(\xi) = a_0 + a_1 \operatorname{arctg}(\xi) , \quad \xi \in R .$$

The controller u_t uses the information about realizations of w_t, and is of a form of a nonanticipative, measurable functional on the space of all continuous functions on $[0,1]$. It is desired to find a control $\hat{u}_t$ that minimizes the criteria

$$Q = E\left(\int_0^1 (x_t^2 + u_t^2)\, dt\right) \tag{5.2}$$

Equation (5.1) can be rewritten in the form below which corresponds to the notation used in Theorem 3.1.

$$dx_t = a(z_t)\, x_t\, dt + u_t\, dt, \quad x_0 = \text{const},$$

$$dz_t = dw_t, \quad z_0 = 0 .$$

According to the results obtained in Chapter 3, stochastic control problem (5.1), (5.2) has solution of the form

$$\hat{u}_t = - V(t, w_t)\, \hat{m}_t , \tag{5.3}$$

where

$$d\hat{m}_t = (a(w_t) - V(t, w_t))\hat{m}_t dt, \quad \hat{m}_0 = x_0,$$

and $V(t,\xi)$, $\xi \in R$ satisfies

$$\frac{\partial V}{\partial t} + 0.5\, \frac{\partial^2 V}{\partial \xi^2} + 2a(\xi)\, V + 1 - V^2 = 0, \tag{5.4}$$

$$V(1, \xi) = 0$$

The partial differential equation (5.4) has, according to Theorem 4.1, a bounded, unique solution. The above equation was solved numerically using an algorithm presented in [S1]. Discretization step in the t-variable $\Delta_t = 0.001$ and in the ξ-variable $\Delta_\xi = 0.1$. These values were chosen as the ones for which changes in solution became relatively insensitive with respect to finer discretization. Generation of a Wiener process w_t involved 1,000 pseudo-random Gaussian variables, $v_i \sim N(0,1)$. Standard procedure from the IMSL library called GGNML was used. Increments of w_t were approximated by the formula $dw_{i\Delta_t} \simeq \sqrt{\Delta_t}\, v_i$. Performance Q of the regulator (5.3) was compared to the performance $\tilde{Q}$ of a "classical" linear-quadratic stochastic regulator obtained for (5.1) with $a(w_t)$ replaced by $E(a(w_t))$. The Table 5.1 below contains results of 5 simulation runs for different values of a_0 and a_1. The initial condition was $x_0 = 1$ for all runs.

Table 5.1. Comparison of optimal and suboptimal regulators.

Run #	1	2	3	4	5
a_0	-1.25	1.25	1.25	6.25	3.75
a_1	-3.	3.	1.	5.	1.5
Q	0.205	5.26	2.77	15.2	8.62
$\tilde{Q}$	0.210	6.43	2.80	97.7	10.7

As was expected for the cases of "strong" disturbances and unstable systems, the performance Q of the controller (5.3) is superior over the performance $\tilde{Q}$ resulting from the classic regulator (see runs number 2, 4, 5). The figures 1, 2, and 3 show the sample paths of w_t, u_t and x_t respectively (run number 5). The enlarged part of a pseudo-Wiener process sample path on Figure 1 shows its highly irregular nature, as would be expected from the properties of a "truly" random Wiener process. The figures 2 and 3 show that the "critical" time for the control system is $t \simeq .4$, when the optimal control path intersects the suboptimal one and converges rapidly to zero. Beginning from this point the trajectory of the suboptimally controlled system diverges towards the larger values, as compared to the value of initial condition.

A natural direction for future research is the simulation of examples more complex than those presented here. When done properly, this may not only illustrate the results of this dissertation, but modify them towards applications with implementation on small computers and minicomputers.

The class of systems given by equations (2.1) and (2.2) in Chapter 2 includes the following stochastic control systems

$$\begin{aligned} dx_t &= (A(t,y,u)x_t + B(t,y,u))dt + G_1(t,y,u)dw_t^1 + G_2(t,y,u)dw_t^2, \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.5) \\ dy_t &= (C(t,y,u)x_t + D(t,y,u))dt + R_1(t,y,u)dw_t^1 + R_2(t,y,u)dw_t^2, \end{aligned}$$

where the state (x_t, y_t), $t \in [0,T]$, is divided into the unobservable

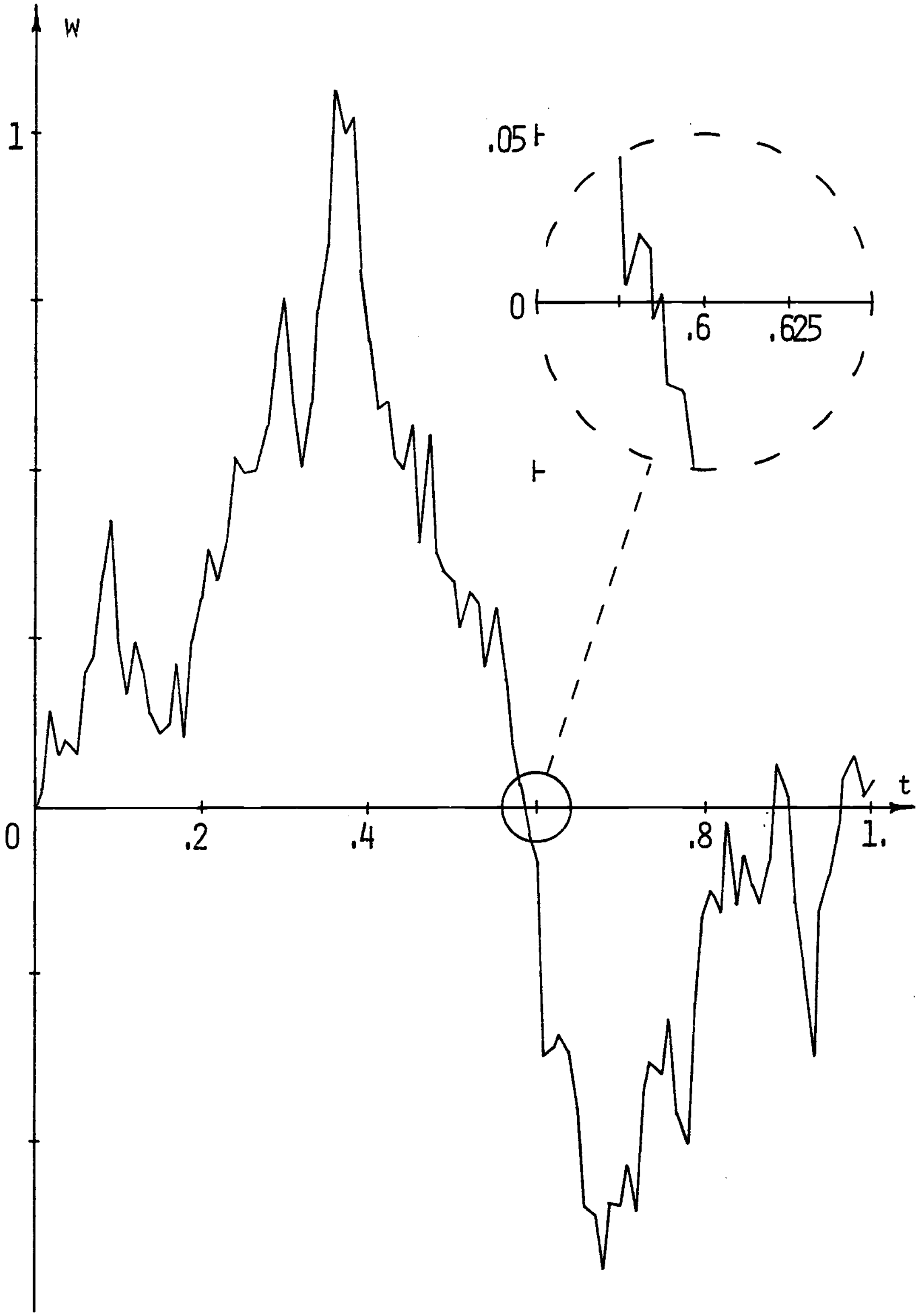

FIG 1. SAMPLE PATH OF A PSEUDO-WIENER PROCESS

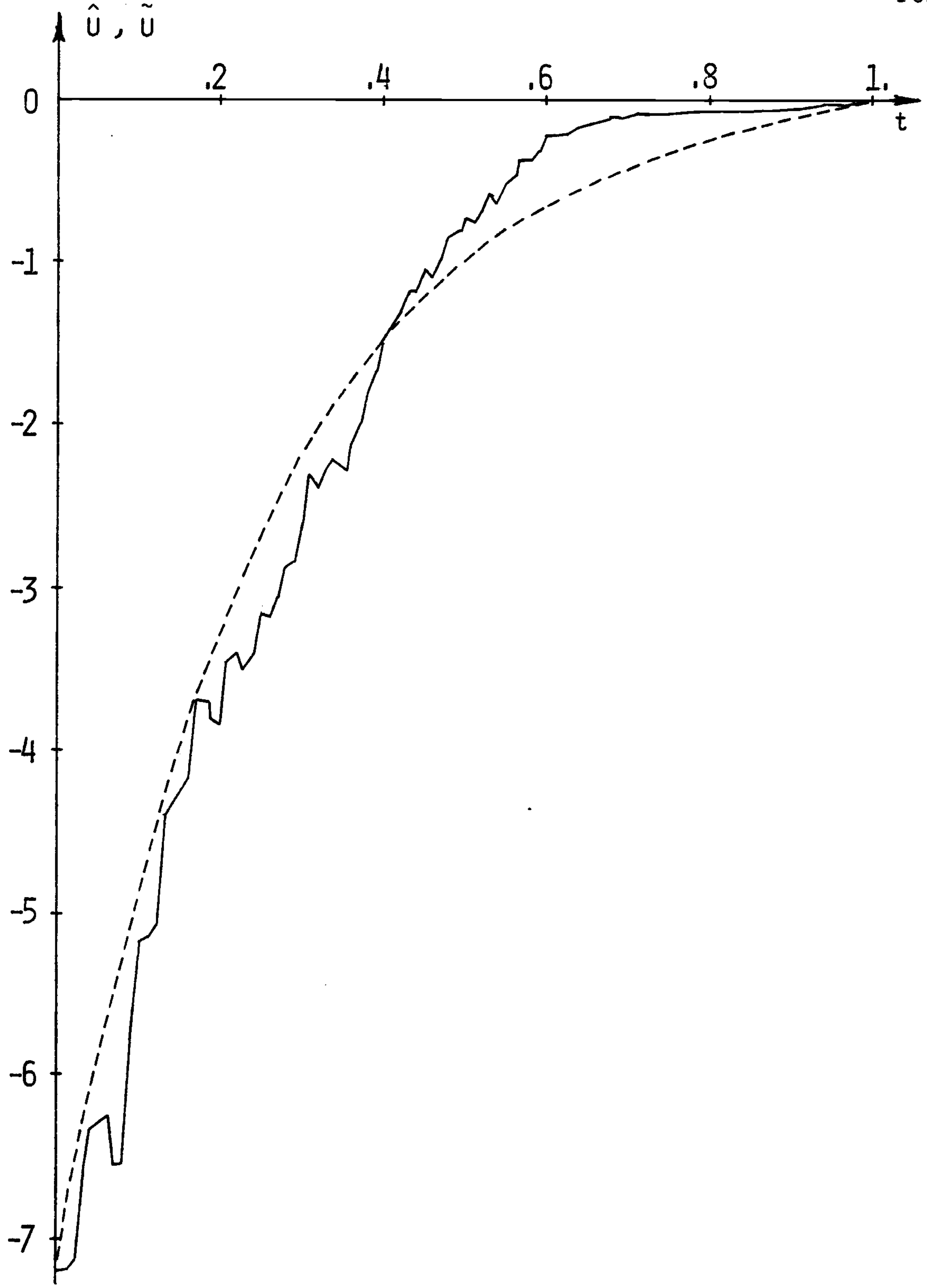

FIG 2. REALIZATIONS OF OPTIMAL û (——) AND SUBOPTIMAL ũ (- - -) REGULATORS

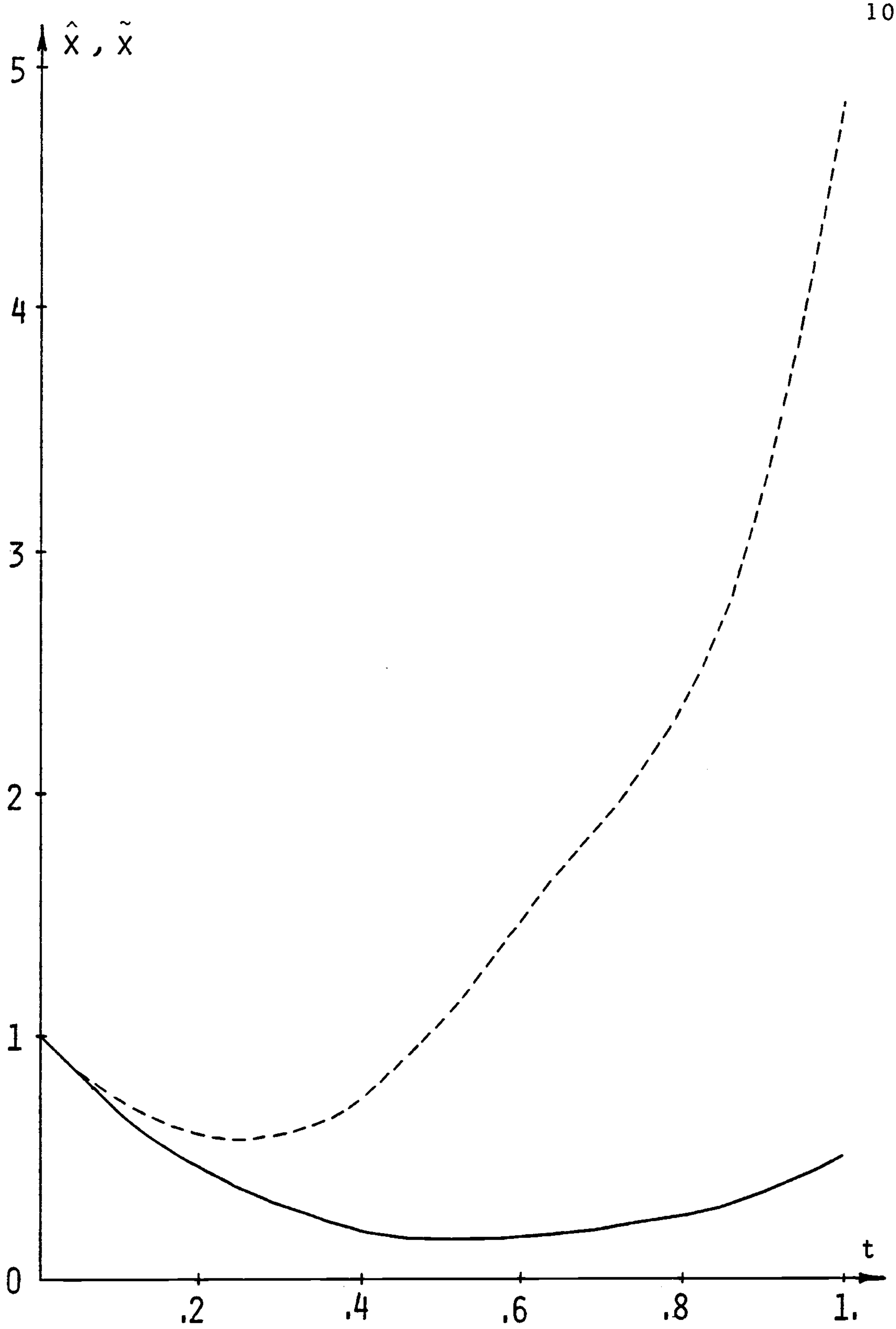

FIG 3. REALIZATIONS OF x UNDER OPTIMAL (——) AND SUBOPTIMAL (- - -) CONTROLS

part x_t of the dimension n and the observable part y_t of the dimension m. w_t^1, i = 1, 2, are mutually independent Wiener processes on a certain probability space $(\Omega, \mathcal{F}, P)$. The p-dimensional control vector u_t is assumed to have $\mathcal{Y}_t$ measurable components for all $t \in [0,T]$, where $\mathcal{Y}_t$ denotes the σ - algebra generated by $(y_s : 0 \leq s \leq t)$. Denote by $\mathcal{C}_T^m$ the space of m dimensional continuous functions on $[0,T]$. Let the components of a p-dimensional vector $h(t,\xi)$ be measurable, nonanticipative functionals on $[0,T] \times \mathcal{C}_T^m$. Then if $g(t,\xi,h(t,\xi))$ denotes any of the elements of A, B, G_1, G_2, C, D, R_1, R_2, it is assumed that $g(\cdot)$ is a measurable, nonanticipative functional on $[0,T] \times \mathcal{C}_T^m$.

The stochastic differential equations (5.5) describe control systems which are linear in the unobservable part of the state variables and nonlinear, in a very general, functional manner in the observable part of the state and control variables.

It is interesting to interpret a special form of (5.5) as a linear-in-state control system with unknown parameters which are to be estimated from the observations $(y_s : 0 \leq s \leq t)$, $t \in [0,T]$.

As (5.5) has well defined filter equation (see Theorem 2.1), the transformation of a partially observable control problem into a completely observable one becomes possible. To justify this statement it is shown that the control criteria of the form

$$I(u) = E\left(\int_0^T Q(t, y, x_t, u_t)dt + S(T, y, x_T)\right) , \tag{5.6}$$

where Q and S are integrable, nonanticipative functionals, and x_t, y_t satisfy (5.5), can be written in the equivalent form

$$I(u) = E\left(\int_0^T \tilde{Q}(t, y, m_t, \Gamma_t, u_t)dt + \tilde{S}(T, y, m_T, \Gamma_T)\right) . \tag{5.7}$$

In the above, m_t, Γ_t are solutions to the filter equations of the type given by Theorem 2.1. The functionals $\tilde{Q}$, $\tilde{S}$ in (5.7) can be obtained from Q and S in (5.6) by using the fact that x_t, given $(y_s: 0 \leq s \leq t)$ is conditionally Gaussian. Consequently

$$\tilde{Q}(t, y, m_t, \Gamma_t, u_t) = \int_{R^n} Q(t, y, a, u_t)f(m_t, \Gamma_t, a)\, d\, a,$$

$$\tilde{S}(T, y, m_T, \Gamma_T) = \int_{R^n} S(T, y, a)f(m_T, \Gamma_T, a)\, d\, a ,$$

where $f(m_t, \Gamma_t, a)$, $a \in R^n$, $t \in [0, T]$, is the probability density function of the Gaussian n-dimensional random variable with mean m_t and covariance Γ_t. This transformation resulted in Chapter 3 in the extension of the separation principle and the synthesis of the optimal regulator. The above transformation might provide the first step towards finding the existence criteria for control problem of the form (5.5), (5.6) [R1, D1, D2, F1, F2, D5].

One area of particularly significant applications of Theorem 1 includes a class of coupled bilinear stochastic systems which appear in such diverse areas as biology, socioeconomics, chemistry and physics [Ml]. As an example of such an application consider the following bilinear stochastic system:

$$\begin{aligned} dx_t &= a(t)\, x_t\, dt + g(t)\, x_t\, dw_t^1 , \\ dy_t &= c(t)\, x_t\, dt + dw_t^2 , \end{aligned} \tag{5.8}$$

where w_t^i, $i = 1, 2$, $t \in [0, T]$, are mutually independent Wiener processes. It is a simple exercise to show that the estimation of x_t from $(y_s: 0 \leq s \leq t)$ involves infinite dimensional filter equations. Consider now the following approximation of (x_t, y_t)

$$\begin{aligned} dx_t &= a(t)\, x_t\, dt + g(t)\, m_t(y)\, dw_t^1 , \\ dy_t &= c(t)\, x_t\, dt + dw_t^2 , \end{aligned} \tag{5.9}$$

where $m_t(y)$ is the mean-square optimal estimate of x_t given $(y_s: 0 \leq s \leq t)$. From (5.9) and Theorem 2.1 the filter equations for (5.9) result

$$dm_t = a(t) m_t\, dt + c(t)\, \Gamma_t\, d\nu_t ,$$

$$d\nu_t = dy_t - c(t)\, m_t\, dt,$$

$$d\Gamma_t = (2a(t)\, \Gamma_t + g^2(t)\, m_t^2 - c^2(t)\, \Gamma_t^2)\, dt.$$

The above may serve as an approximation to the original filter equations. As the results of Theorem 2.1 have a solid mathematical base an approximation like the one above involves less heuristical arguments than, for example, arbitrary truncation in the number of filter equations.

Bilinear systems are a natural generalization of linear systems, and similarly, conditionally Gaussian processes, which satisfy equations of the type given by (5.5) seem to be a natural generalization of Gaussian processes satisfying linear stochastic equations.

BIBLIOGRAPHY

A1 Ariaratnam, S., Graefe, P. "Linear Systems with Stochastic Coefficients", Int'l. J. Cont. 1 (1965), pp. 239-250.

B1 Benes, V. E. "Existence of Optimal Stochastic Control Laws", SIAM, J. Control 9 (1971), pp. 446-472.

B2 Bismut, J. M. "Linear Quadratic Optimal Stochastic Control with Random Coefficients", SIAM J. Control and Optimization 14 (1976), pp. 419-444.

B3 Blankenship, G. "Stability of Linear Differential Equations with Random Coefficients", IEEE Trans. Auto. Cont. 22 (1977), pp. 834-838.

B4 Brockett, R. W. "Parametrically Stochastic Linear Differential Equations", Math. Progr. Study 5 (1976) pp. 8-21.

C1 Chen, C. T. "Introduction to Linear System Theory", Holt, Rinehart and Winston, Inc., New York, 1970.

D1 Davis, M. H. A., Varaiya, P. "Dynamic Programming Conditions for Partially Observable Stochastic Systems", SIAM J. Control 11 (1973), pp. 226-261.

D2 Davis, M. H. A. "On the Existence of Optimal Policies in Stochastic Control", SIAM J. Control 11 (1973), pp. 587-594.

D3 Davis, M. H. A. "The Separation Principle in Stochastic Control via Girsanov Solutions", SIAM J. Control and Optimization 14 (1976), pp. 176-188.

D4 Davis, M. H. A. "The Separation Theorem for Linear Systems with Quadratic Cost Function", Publication 72/36 (Lecture Notes) Department of Computing and Control, Imperial College, London SW7 2BT.

D5 Davis, M. H. A. "Optimal Control of a Degenerate Markovian System", Recent Mathematical Developments in Control, D. J. Bell, Ed., Academic Press, New York, 1973.

F1 Fleming, W. H., Rishel, R. W. "Deterministic and Stochastic Optimal Control", Springer-Verlag, New York, 1975.

F2 Fleming, W. M. "Optimal Continuous Parameter Stochastic Control", SIAM Rev. 11 (1969), pp. 470-509.

G1 Gihman, I. I., Skorohod, A. V. "The Theory of Stochastic Processes III", Springer-Verlag, New York, 1979.

K1 Kantorovich, L. V., Akilov, G. P. "Functional Analysis in NormedSpaces", MacMillan, New York, 1964.

L1 Liptser, R. S., Shiryayev, A. N. "Statistics of Random Processes I. General Theory", Springer-Verlag, New York, 1977.

L2 Liptser, R. S., Shiryayev, A. N. "Statistics of Random Processes II. Applications", Springer-Verlag, New York, 1978.

L3 Liptser, R. S., Shiryayev, A. N. "On Filtering, Interpolation and Extrapolation of Diffusion Type Markov Processes with Incomplete Data", Teoria Verojatn i Primenen. XIII 3 (1968), pp. 569-570.

L4 Liptser, R. S., Shiryayev, A. N. "Nonlinear Filtering of Diffusion Type Markov Processes", Trudy Matem. In-ta in. V. A. Steklova, A. N., SSSR 104 (1968), pp. 135-180.

L5 Ladyzenskaja, O. A., Solonnikov, V. A., Uralceva, N. N. "Linear and Quasilinear Equations of Parabolic Type", Vol. 23, Transactions of Mathematical Monographs, AMS, Providence, Rhode Island, 1968.

M1 Mohler, R. R., Kolodziej, W. T. "Bilinear Systems in Theory and Practice", Proc. J. Auto. Cont. Conf., Denver, 1979.

N1 Neveu, T. "Mathematical Foundations of the Calculus of Probability", Holden-Day Inc., San Francisco, 1965.

R1 Rishel, R. "Necessary and Sufficient Dynamic Programming Conditions for Continuous Time Stochastic Optimal Control", SIAM J. Control 8 (1970), pp. 559-571.

S1 Sincovec, R. F., Madsen, N. K. "Software for Nonlinear Partial Differential Equations", ACM Transactions on Mathematical Software 1 (1975), pp. 232-260.

W1 Wonham, W. M. "On a Matrix Riccati Equation of Stochastic Control", SIAM J. Control 6 (1968), pp. 681-697.

W2 Wonham, W. M. "Random Differential Equations in Control Theory", Probabilistic Methods in Applied Mathematics, Vol. 2, ed. by Bharucha-Reid, Academic Press, 1970, pp. 131-242.

W3 Wonham, W. M. "On the Separation Theorem of Stochastic Control", SIAM J. Control 6 (1968), pp. 312-326.

W4 Wong, E. "Stochastic Processes in Information and Dynamical Systems", McGraw-Hill, New York, 1971.

W5 Wong, E. "Martingale Theory and Applications to Stochastic Problems in Dynamical Systems", Publication 72/19, Imperial College Dept. of Computing and Control, 1972.

APPENDIX A

Selected Results from Stochastic Processes Theory

Theorem A1 (Levy) (L1, Thm 4.1 on p. 82, G1 p. 73)

Define a Wiener process as follows

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{F}_t)$, $t \in [0,T]$ be a nondecreasing family of sub-σ-algebras of $\mathcal{F}$. The random process $(w_t, \mathcal{F}_t)$, $t \in [0,T]$, is called a Wiener process if

(1) the trajectories w_t, $t \in [0,T]$ are continuous (P-a.s.) on $[0,T]$,

(2) w_t, $t \in [0,T]$ is a square integrable martingale with $w_0 = 0$, P-a.s. and

$$\mathcal{E}((w_t - w_s)(w_t - w_s)^*) = (t-s)\, I, \; t \geq s.$$

Any Wiener process $(w_t, \mathcal{F}_t)$, $t \in [0,T]$ is a Brownian motion process.

□

Theorem A2 (L1, Thm 4.6 p. 128, vector case p. 137)

Let the nonanticipative functionals $A(t,\xi)$, $B(t,\xi)$, $t \in [0,T]$, $\xi \in \mathcal{C}_T^m$, satisfy the Lipschitz condition for $\eta \in \mathcal{C}_T^m$

$$\|A(t,\xi)-A(t,\eta)\|^2 + \|B(t,\xi)-B(t,\eta)\|^2 \leq \tag{A.1}$$

$$c\left(\int_0^t \|\xi(s)-\eta(s)\|^2 dK(s) + \|\xi(t) - \eta(t)\|^2\right), \text{ and}$$

$$\|A(t,\xi)\|^2 + \|B(t,\xi)\|^2 \leq c\left(\int_0^t (1 + \|\xi(s)\|^2) dK(s) + (1 + \|\xi_t\|^2)\right), \tag{A.2}$$

where $c < \infty$ is a constant, $K(s)$ is a nondecreasing right continuous function $0 \leq K(s) \leq 1$. Let x_0 be a random variable such that $P(\|x_0\| < \infty) = 1$. Then:

(1) the equation

$$dx_t = A(t,x)\, dt + B(t,x)\, dw_t ,$$

has a unique, strong solution $x_t, t \in [0,T]$, with the initial condition x_0;

(2) if $E(\|x_0\|^{2a}) < \infty$, $a \geq 1$, then there exists a constant c_a such that

$$E(\|x_t\|^{2a}) \leq (1 + E(\|x_0\|^{2a}))\, e^{c_a t} - 1 .$$

□

<u>Lemma A1</u> (Gronwall's inequality, see also L1, Lemma 4.13 p. 130)

Let c_0, c_1, be a nonnegative constants, $u(t)$ be a nonnegative bounded function, and $v(t)$ be a nonnegative integrable function on $[0,T]$ such that

$$u(t) \leq c_0 + c_1 \int_0^t v(s)\, u(s)\, ds .$$

Then

$$u(t) \leq c_0 \exp \left(c_1 \int_0^t v(s)\, ds\right) .$$

□

Lemma A2 (Yershov, M. P. "On representations of Ito processes". Teoria Verojatn. i Primenen, XVII, 1 (1972), 167-172) (see also L1, Lemma 4.9 p. 114).

Let y_t, $t \in [0,T]$ be a continuous random process defined on the complete probability space (Ω, F, P). Next let the measurable process z_t, $t \in [0,T]$ be adapted to the family of the σ - algebras $\mathcal{Y}_t$ generated by $(y_s : 0 \leq s \leq t)$. Then there exists a measurable functional $f_t(\xi)$ defined on $[0,T] \times C_T^m$, such that $\mu \times P((t,\omega)$: $z_t(\omega) \neq f_t(y(\omega))) = 0$, where μ is the Lebesque measure on $[0,T]$ and $\mu \times P$ is the direct product of the measures μ and P.

□

Theorem A3 (L1, Thm 7.19 p. 277, multidimensional formulation p. 279).

Let y_t and $\tilde{y}_t$, $t \in [0,T]$, be two processes of the diffusion type with

$$dy_t = C(t,y)dt + R(t,y)\, dw_t, \tag{A.3}$$

$$d\tilde{y}_t = \tilde{C}(t,\tilde{y})dt + R(t,\tilde{y})dw_t, \quad \tilde{y}_0 = y_0 \tag{A.4}$$

Let the following assumptions be satisfied. The nonananticipative functionals R and C satisfy (A.1), (A.2) (see Thm A2), providing the existence and uniqueness of a strong solution to (A. 4); (A. 5)

for any $t \in [0, T]$ the equation (A. 6)

$$R(t, y)g_t = C(t, y) - \tilde{C}(t, y)$$

has with respect to g_t (P-a. s.) a bounded solution;

$$P\left(\int_0^T (C^*(RR^*)\ C + \tilde{C}(RR^*)\ \tilde{C})dt < \infty \right) = 1, \tag{A.7}$$

the above is assumed to hold for both (t, y) and $(t, \tilde{y})$; denotes the pseudo-inverse. Then $\mu_{\tilde{y}} \sim \mu_y$ and the densities $\frac{d\mu_{\tilde{y}}}{d\mu_y}$, $\frac{d\mu_y}{d\mu_{\tilde{y}}}$ are given by the formulas analogous to (2. 34) and (2. 35) of Chapter 2, and formula (A. 8) Thm A4.

□

Theorem A4 (L1, Thm 7.20 p. 278, multidimensional formulation p. 279).

Let the assumptions of Theorem A3 be fulfilled with the exception that (A. 7) holds only for (t, y). Then $\mu_y << \mu_{\tilde{y}}$ and the density $f_t(\tilde{y}) = \frac{d\mu_y}{d\mu_{\tilde{y}}}$ is given by

$$f_t(\tilde{y}) = \exp\left(\int_0^t (C-\tilde{C})^*(RR^*)\ d\tilde{y}_s - 0.5\int_0^t (C-\tilde{C})^*(RR^*)\ (C-\tilde{C})ds\right), \tag{A.8}$$

where the arguments $(s, \tilde{y})$ are omitted for brevity.

□

Lemma A3 (L1, Lemma 4.10 p. 116)

Let y_t, $\tilde{y}_t$, $t \in [0,T]$ be two processes of a differential representation (A.3), (A.4). Let $G(y)$, $\tilde{G}(\tilde{y})$, $F(y)$, $\tilde{F}(\tilde{y})$ be well defined functionals given by the following formulas

$$G(y) = \int_0^T g(t,y)dt \;, \qquad \tilde{G}(\tilde{y}) = \int_0^T g(t,\tilde{y})dt \;,$$

$$F(y) = \int_0^T f(t,y)dy_t \;, \qquad \tilde{F}(\tilde{y}) = \int_0^T f(t,\tilde{y})d\tilde{y}_t \;,$$

where $g(t,\xi)$, $f(t,\xi)$, $\xi \in C_T^m$ are nonanticipative functionals.

If the measure μ_y is absolutely continuous with respect to the measure $\mu_{\tilde{y}}$ ($\mu_y << \mu_{\tilde{y}}$), then $G(y) = \tilde{G}(y)$, $F(y) = \tilde{F}(y)$ P-a.s. If $\mu_{\tilde{y}} << \mu_y$, then $G(\tilde{y}) = \tilde{G}(\tilde{y})$, $F(\tilde{y}) = \tilde{F}(\tilde{y})$ P-a.s.

□

Lemma A 4 (L1, Lemma 4.12, p. 125)

Let w_t, $t \in [0,T]$, be a Wiener process and let g_t be a nonanticipative function with

$$\int_0^T E(g_t^{2a})\,dt < \infty \,, \quad a > 1 \,.$$

Then

$$E((\int_0^t g_s dw_s)^{2a}) \leq (a(2a-1))^a t^{a-1} \int_0^t E(g_s^{2a})ds \,.$$

□

Lemma A5 (Fatou's Lemma)

If the sequence of P-a.s. positive random variables Z_n, $n = 1, 2, 3, \ldots$, i is such that $Z_n \leq Z$, where Z is an integrable random variable, then

$$E(\liminf_{n\to\infty} (Z_n)) \leq \liminf_{n\to\infty} (E(Z_n)) .$$

□

Lemma A6 (L1, continuous analog of Theorem 3.2 (3.8) p. 58)

If (z_t, F_t), $t \in [0, T]$, is a nonnegative submartingale with

$E(z_t^a) < \infty$, for some a, $1 < a < \infty$, then

$$E((\sup_{0\leq t\leq T} z_t)^a) \leq \left(\frac{a}{a-1}\right)^a E(z_T^a) .$$

□

Lemma A 7 (L2, Lemma 11.6 p. 12, see also L1, Thm 721 p. 280).

Consider a random vector $x = (x_1, \ldots, x_n)$ and an n-dimensional Wiener process w_t, $t \in [0, T]$, with independent components and suppose that the system (x, w_t) is Gaussian. Let $r = (r_1, \ldots, r_n)$ be a row vector and h_3 and $h_2h_2^*$ be (m x m) matrices and

$$\operatorname{tr}\left(\int_0^t (h_3h_3^* + h_2h_2^*)\, dt\right) < \infty .$$

Then

$$E\left(\exp\left(rx - \int_0^T \left(\int_0^t h_3 dw_s\right)^* h_2 h_2^* \left(\int_0^t h_3 dw_s\right) dt\right)\right) =$$

$$\exp\left(r E(x) + 0.5\ r P r^* + 0.5\ \mathrm{tr}\left(\int_0^T h_3 h_3^* S\ dt\right)\right),$$

where P is a nonnegative definite matrix.

□

Theorem A5 (L1, Thm 7.17, p. 270)

Let (y_t, F_t), $t \in [0,T]$, be an Ito process given by

$$y_t = y_0 + \int_0^t C_s\, ds + \int_0^t R_s\, dw_s\ .$$

Let (ν_t, F_t), $t \in [0,T]$, be some Wiener process independent of a Wiener process (w_t, F_t) and the processes (C_t, F_t), (R_t, F_t). Let the following condition be satisfied

$$\int_0^T E(\|C_t\|)dt < \infty\ .$$

Then there will be:

- the measurable functionals $\overline{C}$ and $\overline{R}$ satisfying P-a. s., $t \in [0,T]$ the equalities

$$\overline{C}_t = E(C_t | y_t), \quad \overline{R}_t = (R_t R_t^*)^{1/2}\ ,$$

where

$$\mathcal{Y}_t = \sigma\text{-alg}\,(y_s : 0 \le s \le t);$$

- a Wiener process $(\overline{w}_t, \overline{\mathcal{Y}}_t)$, $t \in [0,T]$, $\overline{\mathcal{Y}}_t = \sigma\text{-alg}\,(\nu_s, y_s :$ $0 \le s \le t)$, such that the process y permits the representation

$$y_t = y_0 + \int_0^t \overline{C}_s(y)ds + \int_0^t \overline{R}_s\, d\overline{w}_s .$$

If, in addition $R_t R_t^* > 0$, P-a.s., for almost all $t \in [0,T]$, then the Wiener process $\overline{w}_t$ is adapted to $\mathcal{Y}_t$, $t \in [0,T]$.

□

<u>Theorem A6</u> (L1, Thm 5.18 p. 197)

Let $(y_t, \mathcal{F}_t)$, $t \in [0,T]$, be a process of the diffusion type with the differential

$$dy_t = C_t(y)\,dt + R_t(y)\,dw_t ,$$

where $C_t(\xi)$, $R_t(\xi)$, $\xi \in C_T^m$, are nonanticipative functionals. Assume that $R_t(\xi)$ satisfies (A.1), (A.2) and that for almost all $t \in [0,T]$, $R_t R_t^*$ be uniformly positive definite. Suppose that

$$P\left(\int_0^T \|C_t(y)\|^2\, dt < \infty\right) = 1 .$$

Then any martingale $(g_t, \mathcal{Y}_t)$ has a continuous modification with the representation

$$g_t = g_0 + \int_0^t f_s \, dw_s ,$$

with $\mathcal{Y}_t$ adapted process f_t, such that

$$P \left(\int_0^T \|f_s\|^2 ds < \infty \right) = 1.$$

If $(g_t, \mathcal{Y}_t)$, $t \in [0, T]$, is a square integrable martingale then also

$$\int_0^T E \left(\|f_s\|^2 \right) ds < \infty .$$

□

Definition A1 [L1]

The P-a. s. continuous random process x_t, $t \in [0, T]$, is a strong solution of the stochastic differential equation

$$dx_t = A(t, x) \, dt + B(t, x) \, dw_t , \qquad (A.9)$$

where $A(t, x)$, $B(t, x)$ are measurable, nonanticipative functionals, $(w_t, \mathcal{F}_t)$, $t \in [0, T]$, is a Wiener process on a given probability space $(\Omega, \mathcal{F}, P)$, and $(\mathcal{F}_t)$, $t \in [0, T]$, is a nondecreasing family of the sub-σ-algebras of $\mathcal{F}$, with $\mathcal{F}_0$-measurable initial condition $x_0 = \eta$ P-a. s., if for each $t \in [0, T]$, the random variables x_t are $\mathcal{F}_t$-measurable,

$$P \left(\int_0^T \|A(t, x)\| dt < \infty \right) = 1, \qquad (A.10)$$

$$P\left(\int_0^T \|B(t,x)\|^2 dt < \infty\right) = 1, \tag{A.11}$$

and with probability 1, for each $t \in [0,T]$

$$x_t = \eta + \int_0^t A(s,x)ds + \int_0^t B(s,x)dw_s . \tag{A.12}$$

Definition A2 [L1]

The stochastic differential equation (A.9) has a weak solution, with the initial condition η of the prescribed distribution function $F(a)$, if there are: a probability space $(\Omega, \mathcal{F}, P)$, a nondecreasing family of the sub-σ-algebras $(\mathcal{F}_t)$, $t \in [0,T]$, a continuous random process $(x_t, \mathcal{F}_t)$, and a Wiener process $(w_t, \mathcal{F}_t)$ such that (A.10), (A.11), (A.12) are satisfied and $P(x_0 \leq a) = F(a)$.

Remark

The principal difference between the concepts of a strong and weak solutions is that if one speaks about the solution in a strong sense, then it is implied that there have been prescribed some probability space $(\Omega, \mathcal{F}, P)$, the family $(\mathcal{F}_t)$, $t \in [0,T]$, of the sub-σ-algebras, and the Wiener process $(w_t, \mathcal{F}_t)$. When one speaks about the weak solution of (A.9) with the prescribed nonanticipative functionals $A(t,\cdot)$, $B(t,\cdot)$, then it is assumed that we may construct a probability space $(\Omega, \mathcal{F}, P)$, a family $(\mathcal{F}_t)$, $t \in [0,T]$, of the sub-σ-algebras, a process (x_t, F_t), and a Wiener process (w_t, F_t) for which (A. 12) is satisfied P-a. s. The weak solution is, actually, an aggregate of the system $(\Omega, F, F_t, P, w_t, x_t)$, where for brevity the process x_t, $t \in [0, T]$ is called a weak solution.

The question of using or not using a concept of weak solutions in stochastic control and filtering problems is almost exclusively the question of a system modelling approach. While a strong solution deals with a "given" Wiener process as a model of wide-spectrum random noise, a weak solution says only that there exists a Wiener process which can be used as such a model. If the physical nature of the problem implies construction of an underlying probability space (for example it specifies the basic space as a set of possible outcomes of a random experiment and gives a family of finite distributions which define a probability measure), then a weak solution approach may not be appropriate. On the other hand, a Wiener process is nothing other than an abstract model of some random phenomena. From this point of view, equation (A. 9), which for a weak solution case takes the following form:

$$dx_t = A(t, x)\, dt + B(t, x)\, d\, (\text{a Wiener process})_t\ ,$$

seems to satisfy modelling purposes quite well. It should be noticed at this point that the concept of a weak solution was used in stochastic control theory by Benes, Davis, Varaiya [Bl, Dl, D2, D3, D4, D5], and others, (see references in [F1] and [L1]).

Definition A3 [L1, W4, G1, N1]

The random process (x_t, F_t), $t \in [0, T]$, is called a martingale (with respect to the family (F_t), $t \in [0, T]$, of σ-algebras) if $E(\|x_t\|) < \infty$, $t \in [0, T]$, and

$$E(x_t \mid F_s) = x_s \quad \text{P-a. s.} \quad t \geq s .$$