ON THE CONVERGENCE OF THE GAUSS-GALERKIN METHOD FOR THE DENSITY OF SOME MARKOV PROCESSES

By

Ali Haj Jafar

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1984

ABSTRACT

ON THE CONVERGENCE OF THE
GAUSS-GALERKIN METHOD FOR THE DENSITY
OF SOME MARKOV PROCESSES

By

Ali Haj Jafar

This dissertation concerns the convergence of the Gauss-Galerkin method for some Markov processes. This method provides approximations to $p(t,x)$, the solution of the Fokker-Planck equation $\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(ap) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b^2 p)$ corresponding to the stochastic differential equation $dX(t) = a(X(t),t)dt + b(X(t),t)dW(t)$ defined on (r_1, r_2) with $0 \leq r_1 < r_2 \leq \infty$. The approximation is in the sense of approximation of probability measures

We first show that when the coefficients of the stochastic differential equation are polynomials, the resulting Gauss-Galerkin system is equivalent to the Hankel system of moments which is closed in an appropriate manner. We then compare the Gauss-Christoffel approximations to $p(t,x)$ with the Gauss-Galerkin approximations and derive upper bounds for the inherent errors.

The Gauss-Galerkin measures, which are atomic in nature, form the basis for numerical integration quadrature formulas. We show that under suitable conditions on the coefficients of the stochastic differential equation, these integration formulas converge to the true value of the integral. The proofs rely on the use of differential inequalities, Helly's theorems on weak compactness of measures and the spectral theory of linear operators.

Numerical examples are presented that illustrate close agreements between the numerical and theoretical results.

ACKNOWLEDGMENTS

I wish to express my deep gratitude to Professor David Yen and Professor Habib Salehi, my thesis advisors, for all their patience, expert guidance and encouragement during the course of my work. I also wish to offer thanks to Professors L.M. Sonneborn, S. Dragosh and W.E. Kuan, other members of my thesis committee.

I am grateful to Professor D. Dawson of Carleton University for many valuable conversations and for making his work on Galerkin approximation accessible to me. I also extend my appreciation to Professor S. Ethier for useful conversations and suggesting several related references.

A special thanks goes to my wife, Fattaneh, for her unwavering support and infinite patience.

I wish to thank Mrs. Berle Reiter, the Librarian of the Mathematics-Statistics library, for all her help during my graduate work at Michigan State University. Finally, I would like to thank Cindy Lou Smith and Tammy Hatfield for typing this dissertation.

DEDICATED TO

MY PARENTS

AND

MY WIFE

TABLE OF CONTENTS

CHAPTER 0

INTRODUCTION

0.1 Statement of Problem and Its Motivation

We are concerned in this dissertation with numerical solutions of the probability law of a class of stochastic differential equations of the type

$$dX(t) = a(X(t),t)dt + b(X(t),t)dW(t) \tag{0.1.1}$$

where X is the spatial coordinate in one-dimension, t is time, $W(t)$ is the standard Brownian motion, and the functions $a(x,t)$ and $b(x,t)$ are known as the "drift" and "diffusion" coefficients respectively. The stated equation models many problems in the physical, engineering and biological sciences.

Since $W(t)$ is a random variable, the solution $X(t)$ of the above equation for $t > 0$ subject to given initial condition $X(0)$, whether random or deterministic, is a random process. Under suitable assumptions on $a(x,t)$ and $b(x,t)$, the existence and uniqueness of the above initial value problem for $X(t)$ is known.

Furthermore, under somewhat more stringent conditions, a probability density function $p(t,x)$ for the process $X(t)$ can be shown to exist and satisfies the deterministic "Fokker-Planck" equation

$$(0.1.2) \qquad \frac{\partial p(t,x)}{\partial t} = -\frac{\partial (a(x,t)p)}{\partial x} + \frac{1}{2}\frac{\partial^2 (b^2(x,t)p)}{\partial x^2}$$

$$(0.1.3) \qquad p(0,x) = \text{given}$$

where x may lie in the infinite interval $(-\infty,\infty)$, the semi-infinite interval $[0,\infty)$ or a finite interval $[r_1,r_2]$. It is clear that additional appropriate boundary conditions on a, b and/or p must be posed.

D. Dawson in [6] suggested a Galerkin type method for approximating $p(t,x)$ by atomic measures. This method transforms the problem to one for a system of nonlinear ordinary differential equations for the "nodes" and "weights" of the atomic measures. The atomic measures provide approximations to the Gauss-Christoffel measures of the exact $p(t,x)$. This method has been referred to as the Gauss-Galerkin method and the atomic measures the Gauss-Galerkin measures.

Our aim in this dissertation is to make a detailed analysis of the Gauss-Galerkin method above. We shall establish some convergence theorems, in both the finite interval and semi-infinite interval case, for the

convergence of the Gauss-Galerkin measures to $p(t,x)$, as $n \to \infty$ (n = number of atoms), when the coefficients $a(x,t)$ and $b(x,t)$ satisfy appropriate continuity and growth conditions.

The proofs of these theorems require techniques from both analysis and probability theory. Specifically, the proof involves the use of differential inequalities, Helley's theorems on the weak compactness of probability measures, the problem of moment, criteria for unique determinism of a measure by its moments, and the eigen-function expansion of a second order differential operator.

A number of examples are presented and, whenever possible, compared with the exact solutions. We have also included examples which do not satisfy the assumptions of the convergence theorems but for which the Gauss-Galerkin method seems to work well.

A Galerkin method based on fixed nodes is also developed and illustrated by applying it to an example. Unfortunately we do not have a convergence theorem for this method at this time.

0.2 Organization of the Dissertation

This dissertation is organized as follows. Chapter I contains background materials from the probability theory.

We begin with a review of the notions of random variables, distribution functions and conditional expectations and probability. We then proceed to a discussion of stochastic processes including the Markov processes and Brownian motions. Diffusion processes and their governing backward and forward equations are then discussed. These are followed by stochastic integrals and stochastic differential equations, and, finally, we state several theorems on the existence and uniqueness of solutions of stochastic differential equations of the type we wish to consider.

In Chapter II we develop the Gauss-Galerkin method by first presenting the weak form of the Fokker-Planck equation. We then review the Gauss-Christoffel approximations to the probability measures. This consideration leads to the development of the Gauss-Galerkin equations. An alternative formulation of the numerical problem in terms of the moments for the case where the Fokker-Planck equation involves polynomial coefficients leading to the so-called Hankel system is also presented. We conclude Chapter II with a discussion of the inherent errors in the Gauss-Galerkin approximations.

The main convergence theorems of the Gauss-Galerkin method are presented in Chapter III for the cases of a semi-infinite interval and of a finite interval. We begin Chapter III with a discussion of the convergence of the Gauss-Christoffel approximations. This is followed

by a detailed formulation of the assumptions involved in the Fokker-Planck equation and the boundary conditions. After some differential inequalities are presented, we state two Helly's theorems dealing with the weak compactness of probability measures. The proofs of the main convergence theorems are preceded immediately by a number of preparatory lemmas.

Several numerical examples are presented in Chapter IV. The numerical solutions, whenever possible, are compared with known exact solutions.

Appendix A contains the numerical solution of a nonlinear stochastic equation which is not of the type of equations studied in this dissertation. The numerical results for this problem, which was previously studied in [5] are encouraging enough and suggest that the Gauss-Galerkin method may indeed converge for much wider classes of stochastic equations.

Appendix B contains the Galerkin method based on fixed nodes as mentioned before and includes numerical results for the problem treated in Appendix A.

CHAPTER I

PRELIMINARIES

The following preliminaries are taken from the books on stochastic differential equations by Arnold [1] and Friedman [11].

1.1 Events and Random Variables: Probability theory deals with mathematical models of trials whose outcomes depend on chance. We group together the possible outcomes (the elementary events) in a set Ω with typical element $\omega \in \Omega$. An observable event A is a subset of Ω; however, not every subset of Ω is in general an observable or interesting event. Let Σ denote the set of observable events for a single trial. Of course, Σ must include Ω, $\emptyset$ and for every event A, its complement $\bar{A}$. Furthermore, given two events A and B in Σ, the union $A \cup B$ and the intersection $A \cap B$ also belong to Σ; thus, Σ is an algebra of events. An algebra Σ of events is called a sigma algebra if

$$\bigcup_{n=1}^{\infty} A_n \in \Sigma$$

when $A_n \in \Sigma$ for $n \geq 1$. The elements of Σ are called measurable sets and the pair (Ω,Σ) is called a measurable space.

Let $\mathcal{A}$ denote a family of subsets of Ω. There exists in Ω a smallest sigma algebra $\Sigma(\mathcal{A})$ that contains all sets belonging to $\mathcal{A}$. This $\Sigma(\mathcal{A})$ is called the sigma algebra (σ-algebra) generated by $\mathcal{A}$. Let (Ω,Σ) and (Ω',Σ') denote measurable spaces. A mapping $X : \Omega \to \Omega'$ is said to be $(\Sigma - \Sigma')$-measurable (and is called an Ω'-valued random variable on (Ω,Σ)) if the preimage of measurable sets in Ω' are measurable sets in Ω, that is, for $A' \in \Sigma'$

$$\{\omega : X(\omega) \in A'\} = [X \in A'] = X^{-1}(A') \in \Sigma .$$

If Ω' is the d-dimensional Euclidean space R^d with the usual distance function, we shall always choose as the sigma algebra Σ' of events the sigma algebra $\mathcal{B}^d$ of Borel sets in R^d generated by the d-dimensional intervals.

1.2 Probability and Distribution Functions: Let (Ω,Σ) denote a measurable space. A set function P defined on Σ is called a probability measure or simply a probability if

a) $0 \leq P(A) \leq 1$ for all $A \in \Sigma$,

b) $P(\emptyset) = 0$,

c) $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ if $A_n \in \Sigma$ for all $n \geq 1$, and $A_n \cap A_m = \emptyset$ $(n \neq m)$ (sigma-additivity)

d) $P(\Omega) = 1$.

The triple (Ω,Σ,P) is called a probability space.

Now, suppose that (Ω,Σ,P) is a probability space, that (Ω',Σ') is a measurable space, and that X is a random variable on (Ω,Σ) with values in Ω'. X induces the probability P_X on the measurable space of the images by

$$P_X(A') = P(X^{-1}(A')) = P\{\omega : X(\omega) \in A'\} = P[X \in A']$$

for all $A' \in \Sigma'$. The function P_X is called the distribution measure, or briefly the distribution of X. For an R^d-valued random variable, the distribution P_X is uniquely defined on $\mathcal{B}^d$ by its distribution function

$$F(x) = F(x_1,\ldots,x_d) = P\{\omega : X_1(\omega) \leq x_1,\ldots,X_d(\omega) \leq x_d\} = P[X \leq x] .$$

It is called the joint distribution function of the d scalar random variables $X_1,\ldots,X_d$ which are the components of X.

1.3 Conditional Expectation and Conditional Probabilities:

For a real-valued random variable X, defined on (Ω,Σ,P) we define its expectation $\mathcal{E}X$ to be

$$\mathcal{E}X = \int_{\Omega} X(\omega)\,dP(\omega) \ .$$

For R^d-valued random variables, we define expectation componentwise. For $p \geq 1$, we define the space

$$L^p(\Omega,\Sigma,P) = \{X : X \text{ is an } R^d\text{-valued random variable }, \ \mathcal{E}|X|^p < \infty \text{ where } |.| \text{ is the usual Euclidean norm }\}$$

Let $X \in L^1(\Omega,\Sigma,P)$ be an R^d-valued random variable and let $\mathcal{F} \subset \Sigma$ be a sub σ-algebra of Σ. The probability space $(\Omega,\mathcal{F},P)$ is a coarsening of the original one and X is, in general, no longer $(\mathcal{F}-\mathcal{B}^d)$ measurable ($\mathcal{F}$-measurable). We seek now an $\mathcal{F}$-measurable coarsening Y of X that assumes, on the average, the same values as X, that is, an integrable random variable Y such that Y is $\mathcal{F}$-measurable and

$$\int_C Y\,dP = \int_C X\,dP \quad \text{for all} \quad C \in \mathcal{F}.$$

According to the Radon-Nikodym theorem, there exists exactly one such Y, almost surely (a.s.) unique. It is called the conditional expectation of X given the σ-algebra $\mathcal{F}$. We write

$$Y = \mathcal{E}(X \mid \mathcal{F}) .$$

The conditional probability $P(A \mid \mathcal{F})$ of an event A given the σ-algebra $\mathcal{F} \subset \Sigma$ is defined by

$$P(A \mid \mathcal{F}) = \mathcal{E}(I_A \mid \mathcal{F}) .$$

1.4 Stochastic Processes: Let I denote an arbitrary nonempty index set and let (Ω,Σ,P) denote a probability space. A family $\{X(t,\cdot);t \in I\}$ of R^d-valued random variables on (Ω,Σ,P) is called a stochastic process with index set I and state space R^d. Sometimes we write $X_t(\cdot)$ instead of $X(t,\cdot)$. If $\{X(t,\cdot);t \in [t_0,T]\}$ is a stochastic process, then for every fixed $\omega \in \Omega$, $X_{\cdot}(\omega)$ is an R^d-valued function defined on $[t_0,T]$ (sample functions). We wish to include the possibilities $t_0 = -\infty$ or $T = \infty$ in which case we write $[t_0,\infty)$, $(-\infty,T]$ or $(-\infty,\infty)$.

The finite dimensional distribution functions of the stochastic process $\{X(t,\cdot);t \in [t_0T]\}$ are given by

$$P[X(t_1) \leq x_1] = P\{\omega : X_1(t_1,\omega) \leq x_{11}, X_2(t_1,\omega) \leq x_{12},\ldots,X_d(t_1,\omega) \leq x_{1d}\} = F_{t_1}(x_1)$$

$$P[X(t_1) \leq x_1, X(t_2) \leq x_2] = F_{t_1,t_2}(x_1,x_2)$$

$$\vdots$$

$$P[X(t_1) \leq x_1,\ldots,X(t_n) \leq x_n] = F_{t_1,\ldots,t_n}(x_1,\ldots,x_n)$$

where t_i belongs to $[t_0,T]$ and $x_i = (x_{i1},x_{i2},\ldots,x_{id})$ belong to R^d (the symbol $\leq$ applies to the components), and $n \geq 1$.

This system of distribution functions satisfies the following two conditions:

a) Condition of symmetry: If $\{i_1,\ldots,i_n\}$ is a permutation of the numbers $1,\ldots,n$, then for arbitrary instants $t_{i_1},\ldots,t_{i_n}$ and for arbitrary $n \geq 1$,

$$F_{t_{i_1},\ldots,t_{i_n}}(x_{i_1},\ldots,x_{i_n}) = F_{t_1,\ldots,t_n}(x_1,\ldots,x_n).$$

b) Condition of compatibility: For $m < n$ and arbitrary $t_{m+1},\ldots,t_n \in [t_0,T]$,

$$F_{t_1,\ldots,t_m,t_{m+1},\ldots,t_n}(x_1,\ldots,x_m,\infty,\ldots,\infty)$$

$$= F_{t_1,\ldots,t_m}(x_1,\ldots,x_m).$$

Conversely by Kolmogorov's fundamental theorem, for every family of distribution functions that satisfies the symmetry and compatibility conditions, there exists a probability space (Ω,Σ,P) and a stochastic process $\{X(t,\cdot);t \in [t_0,T]\}$ defined on it that posesses the given distribution as its finite-dimensional distributions. For $\{X(t,\cdot);t \in [t_0,T]\}$ we shall write briefly X_t or $X(t)$, usually omitting the variable ω. Two

stochastic process $X(t)$ and $\overline{X}(t)$ defined on the same probability space are said to be (stochastically) equivalent if, for every $t \in [t_0,T]$, we have $X(t) = \overline{X(t)}$ with probability 1. Then $X(t)$ is called a version of $\overline{X(t)}$ and vice versa. The finite dimensional distributions of $X(t)$ and $\overline{X(t)}$ coincide. However, since the set N_t for exceptional values of ω for which $X(t) \neq \overline{X(t)}$ depends in general on t, the sample functions of equivalent processes can have quite different analytical properties.

1.5 Markov Processes: A stochastic process $\{X(t);t \in [t_0,T]\}$ defined on the probability space (Ω,Σ,P) with index set $[t_0,T] \subset [0,\infty)$ and with state space R^d is called a Markov Process if the following so-called Markov Property is satisfied: For $t_0 \leq s \leq t \leq T$ and all $B \in \mathcal{B}^d$ (the Borel sets in R^d), the equation

$$(1.5.1) \quad P(X(t) \in B \mid \Sigma([t_0,s])) = P(X(t) \in B \mid X_s)$$

holds with probability 1.

Here $P(X(t) \in B \mid X_s)$ is $P(X(t) \in B \mid \mathcal{F})$ where $\mathcal{F}$ is the σ-algebra generated by X_s which is the smallest σ-algebra w.r.t. which X_s is measurable, and $\Sigma([t_0,s])$ is the smallest σ-algebra generated by $X(t)$, $t_0 \leq t \leq s$.

For given Markov Process $X(t)$, equation (1.5.1) is equivalent to saying that "the past and future are statistically independent when the present is known".

1.6 Transition Probability and Density: Let $X(t)$, $t_0 \leq t \leq T$, denote a Markov Process. The conditional probability $P(X(t) \in B \mid X_s)$ determines a function $P(s,x,t,B)$ of four arguments $s,t \in [t_0,T]$, $x \in R^d$ and $B \in \mathcal{B}^d$. It has the following properties: (Arnold [1])

(1.6.1) For fixed $s \leq t$ and $B \in \mathcal{B}^d$, we have

$$P(s,X_s,t,B) = P(X_t \in B \mid X_s) \quad \text{with probability 1.}$$

Here $P(s,X_s,t,B)$ is the conditional distribution (Arnold [1], p. 29)

(1.6.2) $P(s,x,t,\cdot)$ is a probability on $\mathcal{B}^d$ for fixed $s \leq t$ and $x \in \mathcal{B}^d$.

(1.6.3) $P(s,\cdot,t,B)$ is $\mathcal{B}^d$-measurable for fixed $s \leq t$ and $B \in \mathcal{B}^d$.

(1.6.4) For $t_0 \leq s \leq u \leq t \leq T$ and $B \in \mathcal{B}^d$ and for all $x \in R^d$ with the possible exception of a set $N \subset R^d$ such that $P[X_s \in N] = 0$, we have the so-called Chapman-Kolmogorov equation

$$(1.6.5) \quad P(s,x,t,B) = \int_{R^d} P(u,y,t,B)P(s,x,u,dy).$$

It is always possible to choose $P(s,x,t,B)$ in such a way that for all $s \in [t_0,T]$ and $B \in \mathcal{B}^d$, we have

$$(1.6.6) \qquad P(s,x,s,B) = I_B(x) = \begin{cases} 1 & \text{for } x \in B \\ 0 & \text{for } x \notin B \end{cases}$$

Definition 1.6.1: A function $P(s,x,t,B)$ with the properties (1.6.2-6) is called a transition probability. If $X(t)$ is a Markov Process and $P(s,x,t,B)$ is a transition probability so that (1.6.1) is also satisfied, then $P(s,x,t,B)$ is called a transition probability of the Markov Process $X(t)$. We use the notation $P(s,x,t,B) = P(X(t) \in B \mid X_s = x)$ which is the probability that the observed process will be in the set B at time t if at time s, where $s \leq t$, it was in the state x.

Definition 1.6.2: If $p(s,x,t,y)$ is a non-negative function that is measurable with respect to (w.r.t.) y and whose integral is 1 and for all $s,t \in [t_0,T]$, where $s < t$, all $x \in R^d$ and all $B \in \mathcal{B}^d$, we have

$$P(s,x,t,B) = \int_B p(s,x,t,y)\,dy ,$$

then we call $p(s,x,t,y)$ a density for $P(s,x,t,B)$

Remark 1.6.3: According to Definition 1.6.2 equation (1.6.5) reduces to

$$p(s,x,t,y) = \int_{R^d} p(s,x,u,z)p(u,z,t,y)\,dz.$$

Definition 1.6.4: A Markov Process $X(t)$, $t \in [t_0,T]$, is said to be homogeneous (w.r.t. time) if its transition probability $P(s,x,t,B)$ is stationary, that is, if the condition

$$P(s+u,x,t+u,B) = P(s,x,t,B)$$

is identically satisfied for $t_0 \leq s \leq t \leq T$ and $t_0 \leq s+u \leq t+u \leq T$. In this case the transition probability is then a function only of x, $t-s$ and B. Hence we can write it in the form

$$P(t-s,x,B) = P(s,x,t,B), \quad 0 \leq t-s \leq T-t_0.$$

Therefore, $P(t,x,B)$ is the probability of transition from x to B in time t, regardless of the actual position of the interval of length t on the time axis.

Remark 1.6.5: Every Markov Process $X(t)$ can, by assuming time to be a state component, be transferred into a homogeneous Markov Process $Y(t) = (t,X(t))$ with state space $[t_0,T] \times R^d$. The transition probability $Q(t,y,B)$ for $Y(t)$ for the special sets $B = C \times D$ is then given by

$$Q(t,y,C \times D) = Q(t,(s,x),C \times D) = P(s,x,s+t,D)I_C(s+t)$$

This uniquely determines the probability $Q(t,y,\cdot)$ on the entire set $\mathcal{B}^1([t_0,T]) \times \mathcal{B}^d$.

As an example of a Markov Process, we cite the Brownian motion.

1.7 The Wiener Process (Brownian motion): A Brownian motion is a stochastic process $W(t)$, $t \geq 0$, satisfying

(i) $W(0) = 0$

(ii) For any $0 \leq t_0 < t_1 < \cdots < t_n$ the random variables $W(t_k) - W(t_{k-1})$ $(1 \leq k \leq n)$ are independent.

(iii) If $0 \leq s < t$, $W(t) - W(s)$ is normally distributed with $E(W(t) - W(s)) = (t-s)\mu$, $Var(W(t) - W(s)) = (t-s)\sigma^2$ where μ, σ are real constants, $\sigma \neq 0$. μ is called the drift and σ^2 is called the variance.

As is well known, Brownian motion can be realized on the space of continuous functions with the property that its paths are nowhere differentiable with probability 1. For this process the transition density is given by

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t}$$

Definition 1.7.1: A d-dimensional process $W(t) = (W_1(t), \ldots, W_d(t))$ is called a d-dimensional Brownian motion if each process $W_i(t)$ is a Brownian motion and if the σ-algebras $\Sigma(W_i(t), t \geq 0)$ $1 \leq i \leq d$, are independent.

1.8 The Infinitesimal Generator: We can assign to a general Markov Process $X(t)$ a family of operators defined on a function space. Let $X(t), t \in [t_0, T]$ be a homogeneous Markov Process with transition probability $P(t,x,B)$. Define the operator T_t on the space $B(R^d)$ of bounded measurable scalar functions defined on R^d and equipped with the norm $\|g\| = \sup_{x \in R^d} |g(x)|$ as follows:

For $t \in [0, T - t_0]$, let $T_t g$ denote the function defined by

$$T_t g(x) = \delta_x g(X(t)) = \int_{R^d} g(y) P(t,x,dy).$$

Since $T_t I_B(x) = P(t,x,B)$, we can derive the transition probability from the operator T_t. These operators have the following properties:

For $t \in [0, T - t_0]$ the operator T_t maps the space $B(R^d)$ into itself, is linear, positive and continuous, and has norm $\|T_t\| = 1$. The operator T_0 is the identity, and $T_{s+t} = T_s T_t = T_t T_s$ whenever $t, s, t+s \in [0, T - t_0]$. In particular, in the case $[t_0, \infty)$ the T_t constitutes a commutative one-parameter semigroup of operators, the so-called semigroup of Markov transition operators.

Definition 1.8.1: The infinitesimal operator (generator) A of a homogeneous Markov Process $X(t)$ for $t_0 \leq t \leq T$ is defined by

$$Ag(x) = \lim_{t \downarrow 0} \frac{T_t g(x) - g(x)}{t}, \quad g \in B(R^d)$$

where the limit is uniform with respect to x. The domain of definition $D_A \subset B(R^d)$ consists of all functions for which the limit exists. The operator A is in general an unbounded closed linear operator. If the transition probabilities of $X(t)$ are stochastically continuous, that is, if for every $x \in R^d$ and every $\varepsilon > 0$

$$\lim_{t \downarrow 0} P(t,x,U_\varepsilon) = 1, \quad U_\varepsilon = \{y : |y - x| < \varepsilon,$$

then $P(t,x,B)$ is uniquely defined by A. (Arnold [1], p. 39) In the nonhomogeneous case, let $X(t)$ for $t \in [t_0,T]$ denote an arbitrary Markov Process with transition probability $P(s,x,t,B)$. We refer to Remark 1.6.5 according to which $Y(t) = (t,X(t))$ is a homogeneous Markov process with the state space $[t_0,T] \times R^d \subset R^{d+1}$. We now define the Markov transition operator T_t and the infinitesimal operator A of $X(t)$ as being equal to the same quantities as in the case of the corresponding homogeneous proces $Y(t) = (t,X(t))$ under the definition given earlier, namely

$$T_t g(s,x) = \mathcal{E}_{s,x} g(s + t, X(t + s)) = \int_{R^d} g(s + t, y) P(s,x,t+s,dy),$$

$0 \leq t \leq T - s$, where $g(s,x)$ is a bounded measurable

function in $[t_0,T] \times R^d$ and $Ag(s,x) = \lim_{t \downarrow 0} \frac{T_t g(s,x) - g(s,x)}{t}$ where the limit means the uniform limit in $(s,x) \in [t_0,T]^t \times R^d$.

1.9 Diffusion Process: A Markov Process $X(t)$, for $t_0 \leq t \leq T$, with values in R^d and almost certainly continuous sample functions is called a diffusion process if its transition probability $P(s,x,t,B)$ satisfies the following three conditions:

For every $s \in [t_0,T]$, $x \in R^d$, and $\varepsilon > 0$

(a) $\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} P(s,x,t,dy) = 0$;

(b) There exists an R^d-valued function $a(x,s)$ such that

$$\lim_{t \downarrow s} \frac{1}{t-s} \int (y-x) P(s,x,t,dy) = a(x,s);$$

(c) There exists a $d \times d$ matrix-valued function $b(s,x)$ such that

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|y-x| \leq \varepsilon} (y-x)(y-x)^T P(s,x,t,dy) = b(x,s).$$

The functions a and b are called the coefficients of the diffusion process. In particular, a is called the drift vector and b is called the diffusion matrix. $b(x,s)$ is symmetric and non-negative-definite. The

Brownian motion is a diffusion with drift term being zero and diffusion term being σ^2.

1.10 Backward and Forward Equations: To each diffusion process with coefficients a and $b = (b_{ij})$ is assigned the second order differential operator

$$L \equiv \sum_{i=1}^{d} a_i(x,s) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} b_{ij}(x,s) \frac{\partial^2}{\partial x_i \partial x_j}$$

Lg can be formally written for every twice partially differentiable function $g(x)$ and is determined by a and b. Every diffusion process is uniquely determined by its infinitesimal operator A. We calculate this operator from

$$(1.10.1) \quad Ag(s,x) = \lim_{t \downarrow 0} \frac{1}{t} \int_{R^d} (g(s+t,y) - g(s,x)) \, P(s,x,t+s,dy)$$

by means of a Taylor expansion of $g(s+t,y)$ about (s,x) under the assumption that g is defined and bounded on $[t_0,T] \times R^d$ and is, on the set, twice continuously differentiable w.r.t. x_i and once continuously differentiable w.r.t.s. When we use conditions (b) and (c) of the definition of diffusion process (1.9) we obtain for the right-hand members of (1.10.1) the operator $\frac{\partial}{\partial s} + L$. Under certain conditions on a and b we have, for all functions in DA

$$A = \frac{\partial}{\partial s} + L \ .$$

The following theorems and results are basic in this area and for our later results. (Gikhman and Skorokhod [12a]), and Arnold [1], p. 42).

Theorem 1.10.1: Let $X(t)$ for $t_0 \leq t \leq T$, denote a d-dimensional diffusion process with continuous coefficients $a(x,s)$ and $b(x,s)$ and suppose the limit relation in definition (1.9) holds uniformly in $s \in [t_0,T]$. Let $g(x)$ denote a continuous and bounded scalar function and define

$$u(s,x) = E_{s,x}g(X(t)) = \int_{R^d} g(y)P(s,x,t,dy),$$

where t is fixed, $s < t$ and $x \in R^d$.

Suppose u, $\frac{\partial u}{\partial x_i}$ and $\frac{\partial^2 u}{\partial x_i \partial x_j}$ for $1 \leq i, j \leq d$ are continuous and bounded. Then, $u(s,x)$ is differentiable w.r.t. s and satisfies Kolmogorov's backward equation

$$\frac{\partial u}{\partial s} + L\,u = 0$$

with the end condition $\lim_{s \uparrow t} u(s,x) = g(x)$.

Theorem 1.10.2: Suppose that the assumption of Theorem 1.10.1 regarding $X(t)$ holds. If $P(s,x,t,\cdot)$ has a density $p(s,x,t,y)$ which is continuous with

respect to s and if the derivatives $\frac{\partial p}{\partial x_i}$ and $\frac{\partial^2 p}{\partial x_i \partial x_j}$ exist and are continuous with respect to s, then p is a so-called FUNDAMENTAL SOLUTION of the backward equation

$$\frac{\partial p}{\partial s} + L\, p = 0 ;$$

with the end condition $\lim_{s \uparrow t} p(s,x,t,y) = \delta(x-y)$.

Theorem 1.10.3: For $t \in [t_0 T]$, let $X(t)$ denote a d-dimensional diffusion process for which the limit relation in (1.9) holds uniformly in s and x and which posesses a transition density $p(s,x,t,y)$. If the derivatives $\frac{\partial p}{\partial t}$, $\partial(a_i(y,t)p)/\partial y_i$ and $\partial^2(a_i(y,t)p)/\partial y_i \partial y_j$ exist and are continuous functions, then for fixed s and $x \in R^d$ such that $s \leq t$, this transition density $p(s,x,t,y)$ is a fundamental solution of Kolmogorov's forward equation (the Fokker-Planck equation)

$$(1.10.2) \quad \frac{\partial p}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial y_i}(a_i(y,t)p) - \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{\partial} \frac{\partial^2}{\partial y_i \partial y_j}(b_{ij}(y,t)p) = 0.$$

If we define the distribution $X(t_0)$ in terms of the initial probability P_{t_0} we obtain from $p(s,x,t,y)$ the probability density $p(t,y)$ of $X(t)$ itself:

(1.10.3) $$p(t,y) = \int_{R^d} p(t_0,x,t,y)\, P_{t_0}(dx).$$

If we integrate (1.10.2) w.r.t. $P_{t_0}(dx)$, we see that $p(t,y)$ also satisfies the Fokker-Planck equation (1.10.2) For a Brownian motion with drift term μ and diffusion term σ^2 the Kolmogorov backward equation is

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2} + \mu \frac{\partial p}{\partial x} \quad \text{for} \quad t > 0, \quad -\infty < x, \quad y < \infty$$

and the Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2} - \mu \frac{\partial p}{\partial x}.$$

1.11 Stochastic integrals and stochastic differential Equations: Let $W(t)$, $t \geq 0$, be a one-dimensional Brownian motion on a probability space $(\Omega,\mathcal{F},P)$. Let $\mathcal{F}_t (t \geq 0)$ be an increasing family of σ-algebras such that the σ-algebra $\mathcal{F}(W(s), 0 \leq s \leq t)$ is contained in $\mathcal{F}_t$, and $\mathcal{F}(W(\lambda + t) - W(t), \lambda \geq 0)$ is independent of $\mathcal{F}_t$ for all $t \geq 0$ (e.g., $\mathcal{F}_t = \mathcal{F}(W(s), 0 \leq s \leq t)$). Let $0 \leq \alpha < \beta < \infty$. A stochastic process $f(t)$ defined for $\alpha \leq t \leq \beta$ is called a non-anticipative function with respect to $\mathcal{F}_t$ if:

(i) $f(t)$ is a separable process, i.e. there exists a countable dense set $M = \{t_1,t_2,\ldots\} \subset [\alpha,\beta]$ and a set $N \in \mathcal{F}$ of P-measure zero such that for every

open subinterval (a,b) of (α,β) and every closed subset A of R^d, the two sets

$$\{\omega : f(t,\omega) \in A \text{ for all } t \in (a,b) \cap M\} \in \mathcal{F}$$

and

$$\{\omega : f(t,\omega) \in A \text{ for all } t \in (a,b)\} \quad \text{(not necessarily in } \mathcal{F})$$

differ, if at all, only on a subset of N.

(ii) $f(t)$ is a measurable process, i.e., the function $(t,\omega) \to f(t,\omega)$ from $[\alpha,\beta] \times \Omega$ into R^1 is measurable.

(iii) for each $t \in [\alpha,\beta]$, $f(t)$ is $\mathcal{F}_t$ measurable.

We denote by $L^p_\omega[\alpha,\beta]$ $(1 \leq p \leq \infty)$ the class of all non-anticipative stochastic processes $f(t)$ satisfying:

$$P\{\int_\alpha^\beta |f(t)|^p \, dt < \infty\} = 1$$

$$(P\{\operatorname*{ess\,sup}_{\alpha \leq t \leq \beta} |f(t)| < \infty\} = 1 \quad \text{if} \quad p = \infty)$$

We denote by $M^p_\omega[\alpha,\beta]$ the subset of $L^p_\omega[\alpha,\beta]$ consisting of all stochastic processes f with

$$\mathcal{E} \int_\alpha^\beta |f(t)|^p \, dt < \infty$$

$$(E[\operatorname*{ess\,sup}_{\alpha \leq t \leq \beta} |f(t)| < \infty \quad \text{if} \quad p = \infty)$$

(Friedman [8])

Since any Brownian motion $W(t)$ is nowhere differentiable with probability 1, the integral $\int_0^T f(t)dW(t)$ for the stochastic function $f(t)$ cannot be defined in the usual Lebesgue-Stieltjes sense. K. Ito has given the definition of the integral above which we recall here briefly.

Definition 1.11.1: (Friedman [11]) A stochastic process $f(t)$ defined on $[\alpha,\beta]$ is called a step function if there exists a partition $\alpha = t_0 < t_1 < \cdots < t_r = \beta$ of $[\alpha,\beta]$ such that

$$f(t) = f(t_i) \quad \text{if} \quad t_i \leq t < t_{i+1}, \quad 0 \leq i \leq r-1.$$

Definition 1.11.2: Let $f(t)$ be a step function in $L_w^2[\alpha,\beta]$. The random variable

$$\sum_{k=0}^{r-1} f(t_k)[W(t_{k+1}) - W(t_k)]$$

is denoted by

$$\int_\alpha^\beta f(t)dW(t)$$

and is called the stochastic integral of f w.r.t. the Brownian motion $W(t)$; it is also called the Ito integral.

In a series of lemmas and theorems Ito proves that for a given process $f(t) \in L_w^2[\alpha,\beta]$, there exists a

sequence of step functions $f_n(t)$ in $L^2_\omega[\alpha,\beta]$ such that

$$\lim_{n\to\infty} \int_\alpha^\beta |f(t) - f_n(t)|^2 dt = 0 \quad \text{a.s.}$$

and the sequence

$$\{\int_\alpha^\beta f_n(t)\,dW(t)\}$$

is convergent in probability. The limit is denoted by

$$\int_\alpha^\beta f(t)\,dW(t)$$

and is called the stochastic integral of $f(t)$ w.r.t. Brownian motion $W(t)$.

For an $m \times d$ matrix $b(t) = (b_{ij}(t))$ where any of its elements belong to $L^2_\omega[\alpha,\beta]$ and for $W(t) = (W_1(t),\ldots,W_d(t))$ a d-dimensional Brownian motion, the stochastic integral $\int_\alpha^\beta b(t)\,dW(t)$ is the m-vector defined by

$$\int_\alpha^\beta b(t)\,dW(t) = \{\sum_{j=1}^{d} \int_\alpha^\beta b_{ij}(t)\,dW_j(t)\}_{i=1,\ldots,m}$$

Definition 1.11.3: Assume $a(x,t) = (a_1(x,t),\ldots,a_m(x,t))$ and $b(x,t) = (b_{ij}(x,t))_{i,j=1}^{m}$. Suppose $a_i(x,t)$, $b_{ij}(x,t)$ are measurable in $(x,t) \in R^d \times [0,T]$. Let $X(t)$ be an m-dimensional process for $0 \le t \le T$, and suppose that for any

$0 \leq t_1 < t_2 \leq T$,

$$X(t_2) - X(t_1) = \int_{t_1}^{t_2} a(X(t),t)\,dt + \int_{t_1}^{t_2} b(X(t),t)\,dW(t)$$

where $a(X(t),t)$ and $b(X(t),t)$ belong to $L_w^1[0,T]$ and $L_w^2[0,T]$ respectively. Then we say that $X(t)$ has a stochastic differential

$$dX(t) = a(X(t),t)\,dt + b(X(t),t)\,dW(t).$$

1.12 Existence and uniqueness: For $a(x,t)$ and $b(x,t)$ as in definition 1.11.3, we write $|b|^2 = \sum_{i,j} |b_{ij}|^2$. If $X(t)$ $(0 \leq t \leq T)$ is a stochastic process such that

(1.12.1) $dX(t) = a(X(t),t)\,dt + b(X(t),t)\,dW(t)$,

(1.12.2) $X(0) = X_0$ a.s.,

then we say that $X(t)$ satisfies the system of stochastic differential equations (1.12.1) and the initial condition (1.12.2).

Theorem 1.12.1: Suppose $a(x,t)$ and $b(x,t)$ are measurable in $(x,t) \in R^d \times [0,T]$ and there exist constants K and K_* such that

$$|a(x,t)| \leq K(1 + |x|), \quad |b(x,t)| \leq K(1 + |x|),$$

and

$$|a(x,t) - a(\bar{x},t)| \leq K_* |x - \bar{x}|,$$

$$|b(x,t) - b(\bar{x},t)| \leq K_* |x - \bar{x}|$$

for $x \in R^n$ and $0 \leq t \leq T$. Let X_0 be any d-dimensional random variable independent of $\mathcal{F}(W(t), 0 \leq t \leq T)$ such that $\mathcal{E}|X_0|^2 < \infty$. Then there exists a unique solution $X(t)$ of (1.12.1), (1.12.2) in $M_w^2[0,T]$.

If the conditions of theorem 1.12.1 hold for arbitrary $T > 0$ with constants K and K_* depending only on T and in addition if $a(x,t)$, $b(x,t)$ are continuous in $(x,t) \in R^d \times [0,\infty)$, then the solution of (1.12.1), (1.12.2) is a diffusion process with drift $a(x,t)$ and diffusion matrix $\sigma(x,t) = b(x,t)b^*(x,t)$; therefore, theorems 1.10.2 and 1.10.3 are valid for the solution.

Remark 1.12.2: Theorems 1.10.2 and 1.10.3 are valid provided the transition density of the transition probability exists and is continuous. The following theorem guarantees the existence of a density. (Theorem 5.4, Friedman [11]).

Theorem 1.12.3: If

(i) There is a positive constant c such that

$$\sum \xi_i \xi_j b_{ij}(x,t) \geq c|\xi|^2 \quad \text{for all} \quad (x,t) \in R^d \times [0,T],$$
$$\xi \in R^d,$$

(ii) The functions a_i, b_{ij} are bounded on $R^d \times [0,T]$ and uniformly Lipschitz continuous in (x,t) in compact subsets of $R^d \times [0,T]$.

(iii) The functions b_{ij} are Hölder continuous in x (of order α), uniformly with respect to $(x,t) \in R^d \times [0,T]$.

Then the transition probability of the solution of (1.12.1) has a density.

We close this chapter by stating the definition of a linear stochastic differential equation, a theorem concerning existence of moments of a linear stochastic differential equation and Ito's formula. (Arnold [1] and Friedman [11]).

Definition 1.12.4: A stochastic differential equation

$$dX(t) = a(X(t),t)dt + b(X(t),t)dW(t)$$

for the d-dimensional process $X(t)$ on the interval $[t_0,T]$ is said to be linear if the functions $a(x,t)$ and $b(x,t)$ are linear functions of $x \in R^d$ on $R^d \times [t_0,T]$, i.e. if $a(x,t) = A(t)x + \alpha(t)$, where $A(t)$ is a $(d \times d)$ matrix-valued $\alpha(t)$ is an R^d-valued and if $b(x,t) = (B_1(t)x + b_1(t), \ldots, B_m(t)x + b_m(t))$ where $B_k(t)$ is $(d \times d)$ matrix-valued and $b_k(t)$ is R^d-valued. Thus a linear stochastic differential equation has the form

$$dX(t) = (A(t)X(t) + \alpha(t))dt + \sum_{i=1}^{m} (B_i(t)X(t) + b_i(t))dW_i(t)$$

where $W(t) = (W_1(t),\ldots,W_m(t))$. It is said to be homogeneous if $\alpha(t) = b_1(t) = \cdots = b_m(t) = 0$.

Theorem 1.12.5: The solution $X(t)$ of the linear stochastic differential equation has for all $t \in [t_0,T]$, a p-th order moment if and only if $E|X(0)|^p < \infty$ (Arnold [1], p. 138).

The following theorem is very essential in the calculus of stochastic differential equations (Friedman [11] p. 90).

Theorem 1.12.6: (Ito's formula). Let $u(x,t)$ be a function in $(x,t) \in R^d \times [0,\infty)$ and suppose $u(x,t)$ and all its derivatives u_t, u_{x_i}, $u_{x_i x_j}$ are continuous. Let $X(t)$ be an d-dimensional process having a stochastic differential

$$dX(t) = a(t)dt + b(t)dW(t)$$

where $a = (a_1,\ldots,a_d)$ and $b = (b_{ij})$ $(1 \leq i \leq d,\ 1 \leq j \leq n)$ belong to $L_w^1[0,T]$ and $L_w^2[0,T]$ respectively. Then $u(X(t),t)$ has a stochastic differential

$$du(X(t),t) = [u_t(X(t),t) + \sum_{i=1}^{d} u_{x_i}(X(t),t)a_i(t) +$$

$$+ \frac{1}{2} \sum_{p=1}^{n} \sum_{i,j=1}^{d} u_{x_i x_j}(X(t),t)b_{i\ell}(t)b_{j\ell}(t)]dt$$

$$+ \frac{1}{2} \sum_{p=1}^{n} \sum_{i=1}^{d} u_{x_i}(X(t),t)b_{i\ell}(t)dW_\ell(t) .$$

In the more interesting one dimensional case the formula above becomes

$$du(X(t),t) = [u_t(X(t),t) + u_x(X(t),t)a(X(t),t) +$$

$$+ \frac{1}{2} u_{xx}(X(t),t)b^2(X(t),t)]dt +$$

$$+ u_x(X(t),t)b(X(t),t)dW(t) .$$

CHAPTER II

NUMERICAL SOLUTIONS TO THE PROBABILITY DENSITY OF STOCHASTIC DIFFERENTIAL EQUATIONS IN ONE-DIMENSION

2.1 The Mathematical Problem for the Density and its Weak Form: Consider the stochastic differential equation

$$dX(t) = a(X(t),t)\,dt + b(X(t),t)\,dW(t) \tag{2.1.1}$$

$$X(0) = X_0 \tag{2.1.2}$$

on the interval $I = (r_1, r_2)$ where $0 \le r_1 \le r_2 \le \infty$ with continuous coefficients. X_0 is a random variable having a density and with finite moments of all orders.

According to Theorem 1.12.1 the solution to (2.1.1) and (2.1.2) is a Markov process and the density $p(x,t)$ of the law of the process satisfies the Fokker-Planck equation

$$\frac{\partial p(t,x)}{\partial t} = L^* p = -\frac{\partial (a(x,t)p)}{\partial x} + \frac{1}{2}\frac{\partial^2 (b^2(x,t)p)}{\partial x^2} \tag{2.1.3}$$

$$p(0,x) = \text{given} . \tag{2.1.4}$$

It is possible for the process to exhibit various type of behavior at the boundary r_1 or r_2 (Mandl [15]). We assume the coefficients are such that the probability is always preserved, i.e. $\int_{r_1}^{r_2} p(t,x)dx = 1$. In Chapter III we shall discuss more about the boundary conditions.

Equations (2.1.3) and (2.1.4) form a parabolic boundary value problem. We note that the formal adjoint L of the operator L^* is given by the operator

$$(2.1.5) \qquad L = a(x,t)\frac{\partial}{\partial x} + \frac{1}{2}b^2(x,t)\frac{\partial^2}{\partial x^2} .$$

We multiply (2.1.3) by some functions $v(x)$ and integrate over (r_1,r_2) to yield

$$(2.1.6) \qquad (p_t,v) = (L^*p,v) .$$

This is to hold for each function v in some appropriate space V. We assume V is such that integrations by parts can be performed in (L^*p,v) and that all the resulting boundary terms drop out. This leads to

$$(2.1.7) \qquad (p_t,v) = (p,Lv), \quad \text{for all} \quad v \in V .$$

Equation (2.1.7) above is the weak or Galerkin form of (2.1.3).

The density $p(t,x)$ can be used to compute the expected value

$$\mathcal{E}f(X(t)) = \int_{r_1}^{r_2} f(x)p(t,x)dx \ ,$$

of a function $f(x)$. Quite often the integration above is done numerically. In the next section we shall discuss one such method for the numerical approximation to a positive measure μ.

2.2 The Gauss-Christoffel Approximation of Measures:

An n-point Gauss-Christoffel approximation to a measure μ with density $p(x)$ is given by

$$\tilde{\mu}_n = \sum_{k=1}^{n} \tilde{a}_k \, \delta_{\tilde{x}_k}$$

where the $\{\delta_{\tilde{x}_k}\}$ denote atoms at the Gauss-Christoffel points $\{\tilde{x}_k\}$ and the $\{\tilde{a}_k\}$ denote the Gauss-Christoffel weights. The main results concerning this approximation are summarized in the following (Stroud [21]).

Theorem 2.2.1:

(a) For a given measure μ defined on a finite interval $[r_1, r_2]$, the n points $\{\tilde{x}_k\}_1^n$ and the n weights $\{\tilde{a}_k\}$ can be uniquely chosen so that

$$(2.2.1) \qquad \int_{r_1}^{r_2} f(x)d\mu = \sum_{k=1}^{n} \tilde{a}_k \, f(\tilde{x}_k)$$

holds for all polynomials of degree less than or equal to $2n-1$.

(b) Let $\{P_m(x)\}$ be the family of orthogonal polynomials with leading coefficients unity associated with the measure μ; that is, for each m, $P_m(x)$ is a polynomial of degree m and

$$(2.2.2) \qquad \int_{r_1}^{r_2} P_m(x) P_n(x)\, d\mu = 0 \quad \text{if} \quad m \neq n \ .$$

Then the Gauss-Christoffel points $\{\tilde{x}_k; k = 1,\ldots,n\}$ are the zeros of the polynomial $P_n(x)$.

(c) The positive Gauss-Christoffel weights are uniquely determined as the solution of the equations

$$(2.2.3) \qquad \int_{r_1}^{r_2} f(x)\, d\mu = \sum_{k=1}^{n} \tilde{a}_k\, f(\tilde{x}_k)$$

for all polynomials of degree less than or equal to $n-1$.

(d) If $f(x)$, $f^{(1)}(x),\ldots,f^{(2n)}(x)$ are continuous on $[r_1\ r_2]$, then there exists a function $K(s)$ such that the error

$$(2.2.4) \qquad \begin{aligned} E[f] &= \int_{r_1}^{r_2} f(x)\, d\mu - \sum_{k=1}^{n} \tilde{a}_k\, f(\tilde{x}_k) \\ &= \int_{r_1}^{r_2} K(s) f^{(2n)}(s)\, ds \\ &= \frac{1}{(2n)!}\, f^{(2n)}(\zeta) \int_{r_1}^{r_2} [P_n(x)]^2 d\mu \end{aligned}$$

for some $\zeta \in (r_1, r_2)$, and

$$K(s) = E[(x-s)_+^{2n-1}/(2n-1)!]$$

where $(x-s)_+^{2n-1} = \begin{cases} (x-s)^{2n-1} & \text{for } x \geq s \\ 0 & \text{for } x < s \end{cases}$.

From (2.2.4) we obtain the estimate

$$|E[f]| \leq eM \tag{2.2.5}$$

where

$$e = \int_{r_1}^{r_2} K(s)ds \quad \text{and} \quad M = \sup_{[r_1,r_2]} |f^{(2n)}(x)| .$$

The nonnegative function $K(s)$ is called a Peano's kernel function. The inequality (2.2.5) is called Peano's estimate for $E[f]$.

In the theorem above we assumed that the interval $[r_1,r_2]$ was finite. Stroud and Kwan-Wei Chen [22] give Peano's error estimates for Gauss-Laguerre formulas ($p(x) = e^{-x}$). Their idea could also be used to obtain similar estimates for other densities for the intervals $[0,\infty)$ and $(-\infty,\infty)$ (Stroud [21] p. 204). In the generalization of Theorem 2.4.1, parts (a), (b) and (c) remain unchanged (Krylov [12]). However, in part (d) for the $[0,\infty)$ case in addition to $f(x)$, $f^{(1)}(x),\ldots,f^{(2n)}(x)$ being continuous, we must assume $|f^{(2n)}(x)| \leq \beta x^{\alpha}$ (where α is a constant) for $x \geq x_n$. In this case the Peano's estimate for $E[f]$ becomes

$$E[f] \leq eM + \beta \int_{x_n}^{\infty} K(s)s^{\alpha}ds \tag{2.2.6}$$

where

$$e = \int_0^{x_n} K(s)\,ds \quad \text{and} \quad M = \sup_{[0,x_n]} |f^{(2n)}(x)| .$$

For example, if $d\mu = e^{-x}dx$ we see that for $s > x_n$ we have

$$\begin{aligned}(2n-1)!K(s) &= \int_0^{\infty} e^{-x}(x-s)_+^{2n-1}\,dx \\ &= \int_s^{\infty} e^{-x}(x-s)^{2n-1}\,dx \\ &= e^{-s}\int_0^{\infty} e^{-u}\,u^{2n-1}\,du = (2n-1)!e^{-s} .\end{aligned}$$

Therefore,

$$K(s) = e^{-s} \quad \text{for} \quad s > x_n . \tag{2.2.7}$$

This shows that the integral in (2.2.6) is

$$\int_{x_n}^{\infty} K(s)s^{\alpha}ds = \int_{x_n}^{\infty} e^{-s}\,s^{\alpha}ds = \Gamma(\alpha+1,x_n) .$$

So in this case we have shown that

$$|E[f]| \le eM + \beta\Gamma(\alpha+1,x_n) . \tag{2.2.8}$$

To estimate e in (2.2.8) we note that

$$E[P_n^2(x)] = \int_0^{\infty} e^{-x}\,P_n^2(x)\,dx = (n!)^2$$

(Krylov [14], p. 35).

On the other hand,

$$(n!)^2 = E[P_n^2] = E[x^{2n}] = (2n)! \int_0^\infty K(s)\,ds$$

$$= (2n)!\left(\int_0^{x_n} K(s)\,ds + \int_{x_n}^\infty K(s)\,ds\right)$$

and

$$\int_{x_n}^\infty K(s)\,ds = \int_{x_n}^\infty e^{-t}\,dt = e^{-x_n}\ .$$

Thus in (2.2.8) we have

$$e = (n!)^2/(2n)! - e^{-x_n}\ .$$

Now let $p(t,x)$ be the solution to the equations (2.1.3) and (2.1.4). By theorem 2.2.1 for any t the n-point Gauss Christoffel approximation to $p(t,x)\,dx$ is

$$\tilde{\mu}_n(t) = \sum_{k=1}^n \tilde{a}_k(t)\,\delta_{\tilde{x}_k(t)}$$

and by (2.2.4) (2.1.7) can be written as

$$(2.2.9)\qquad \frac{d}{dt}\left(\sum_{k=1}^n \tilde{a}_k(t)\,v(\tilde{x}_k) + E[v]\right)$$

$$= \sum_{k=1}^n \tilde{a}_k(t)\,(Lv)(\tilde{x}_k(t)) + E[Lv]$$

If $f_i(x)$ $(i = 1,2,\ldots,2n)$ is a basis for polynomials of degree less than or equal to $2n-1$, equation (2.2.9) for $f_i(x)$'s becomes

$$(2.2.10)\qquad \frac{d}{dt}\sum_{k=1}^{n}\tilde{a}_k(t)f_i(\tilde{x}_k(t)) = \sum_{k=1}^{n}\tilde{a}_k(t)(Lf_i)(\tilde{x}_k(t)) + E[Lf_i]$$

$$\text{for}\quad i = 1,2,\ldots,2n$$

The system (2.2.10) is a system of ordinary differential equations for the Gauss-Christoffel weights and nodes. The system (2.2.10) however, is not closed as $p(t,x)$ is involved in $E[Lf_i]$ and thus cannot be used for finding $\tilde{\mu}_n(t)$ without knowing $p(t,x)$.

2.3 The Gauss-Galerkin Method: The Gauss-Galerkin method for approximating $p(t,x)$ (as introduced by D.A. Dawson [6]) is the following system obtained from (2.2.10) with the terms $E[Lf_i]$ dropped.

$$(2.3.1)\qquad \frac{d}{dt}\sum_{k=1}^{n}a_k(t)f_i(x_k(t)) = \sum_{k=1}^{n}a_k(t)(Lf_i)(x_k(t))$$

$$\text{for}\quad i = 1,2,\ldots,2n$$

where $\{f_i(x)\}_1^{2n}$ is a basis for polynomials of degree less than or equal to $2n-1$.

We take the initial condition

$$\sum_{k=0}^{n}\tilde{a}_k(0)\delta_{\tilde{x}_k(0)} = \tilde{\mu}(0)$$

as the Gauss-Christoffel approximation to $p(0,x)$.

In matrix form the nonlinear system (2.3.1) with given initial condition $p(0,x)$ can be written as

$$AX' = BX \tag{2.3.2}$$

$$X(0) = \text{given} \tag{2.3.3}$$

where

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$B = \begin{pmatrix} B_1 & 0 \\ B_2 & 0 \end{pmatrix},$$

$$X^T = (a_1, a_2, \ldots, a_n, x_1, \ldots, x_n),$$

$$A_{11} = \begin{pmatrix} f_1(x_1) & f_1(x_2), \ldots & f_1(x_n) \\ f_2(x_1) & f_2(x_2), \ldots & f_2(x_n) \\ \vdots & & \\ f_n(x_1) & f_n(x_2), \ldots & f_n(x_n) \end{pmatrix},$$

$$A_{12} = \begin{pmatrix} a_1 f_1'(x_1) & a_2 f_1'(x_2) & \cdots & a_n f_1'(x_n) \\ a_1 f_2'(x_1) & a_2 f_2'(x_2) & \cdots & a_n f_2'(x_n) \\ \vdots & & & \\ a_1 f_n'(x_1) & a_2 f_n'(x_1) & \cdots & a_n f_n'(x_n) \end{pmatrix},$$

$$A_{21} = \begin{pmatrix} f_{n+1}(x_1) & f_{n+1}(x_2) & \cdots & f_{n+1}(x_n) \\ f_{n+2}(x_1) & f_{n+2}(x_2) & \cdots & f_{n+2}(x_n) \\ \vdots & & & \\ f_{2n}(x_1) & f_{2n}(x_2) & \cdots & f_{2n}(x_n) \end{pmatrix},$$

$$A_{22} = \begin{pmatrix} a_1 f_{n+1}'(x_1) & a_2 f_{n+1}'(x_2) & \cdots & a_n f_{n+1}'(x_n) \\ a_1 f_{n+2}'(x_1) & a_2 f_{n+2}'(x_2) & \cdots & a_n f_{n+2}'(x_n) \\ \vdots & & & \\ a_1 f_{2n}'(x_1) & a_2 f_{2n}'(x_2) & \cdots & a_n f_{2n}'(x_n) \end{pmatrix},$$

$$B_1 = \begin{pmatrix} Lf_1(x_1) & Lf_1(x_2) & \cdots & Lf_1(x_n) \\ Lf_2(x_1) & Lf_2(x_2) & \cdots & Lf_2(x_n) \\ \vdots & & & \\ Lf_n(x_1) & Lf_n(x_2) & \cdots & Lf_n(x_n) \end{pmatrix},$$

and

$$B_2 = \begin{pmatrix} Lf_{n+1}(x_1) & Lf_{n+1}(x_2) & \cdots & Lf_{n+1}(x_n) \\ Lf_{n+2}(x_1) & Lf_{n+2}(x_2) & \cdots & Lf_{n+2}(x_n) \\ \vdots & & & \\ Lf_{2n}(x_1) & Lf_{2n}(x_2) & \cdots & Lf_{2n}(x_n) \end{pmatrix} .$$

Throughout our work we shall take $f_i(x) = x^{i-1}$ $(i = 1,\ldots,2n)$. Note that the system (2.3.2) is nonlinear.

In the remainder of this section we are concerned with the question of solvability of the system (2.3.2). In particular we shall show that under appropriate assumptions the matrix A in (2.3.2) is nonsingular.

Lemma 2.3.1. If $x_1,x_2,\ldots,x_n$ are distinct, then the matrix

$$\hat{A} = \begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{pmatrix}$$

where $\hat{A}_{11} = A_{11}$, $\hat{A}_{21} = A_{22}$, $\hat{A}_{12} = A_{12}D^{-1}$, $\hat{A}_{22} = A_{22}D^{-1}$ and $D = \mathrm{dig}(a_1,a_2,\ldots,a_n)$, is nonsingular.

Proof: Let $f : R \to R^{2n}$ be defined by

$$f(x)^T = (1,x,\ldots,x^{2n-1})$$

Note that

$$\hat{A} = (f(x_1) \quad f(x_2) \quad \dots \quad f(x_n) \quad f'(x_1) \quad \dots \quad f'(x_n)) \ .$$

Let $F(x_1,x_2,\dots,x_n) = \det \hat{A}$. It is easy to see that

$$\frac{\partial F}{\partial x_1} = \det(f(x_1) \ f(x_2) \dots f(x_n) \ f''(x_1) \ f'(x_2) \dots f'(x_n)),$$

$$\frac{\partial^2 F}{\partial x_1^2} = \det(f'(x_1) \ f(x_2) \dots f(x_n) f''(x_1) \ f'(x_2) \dots f'(x_n))$$

$$+ \det(f(x_1) \ f(x_2) \dots f(x_n) \ f'''(x_1) \ f'(x_1) \dots f'(x_n))$$

and

$$\frac{\partial^3 F}{\partial x_1^3} = 2 \det(f'(x_1) \ f(x_2) \dots f(x_n) \ f'''(x_1) \ f'(x_2) \dots f'(x_n))$$

$$+ \det(f(x_1) \ f(x_2) \dots f(x_n) \ f^{(4)}(x_1) \ f'(x_1) \dots f'(x_n))$$

$$= (x_2 - x_1)h(x_1,\dots,x_n) \quad \text{because} \quad \frac{\partial^3 F}{\partial x_1^3} = 0 \quad \text{if we put} \quad x_1 = x_2.$$

Thus $F(x_1,\dots,x_n) = (x_2 - x_1)^4 \varphi(x_1,\dots,x_n)$. Similarly for $i \neq j$ we have $F(x_1,\dots,x_n) = (x_i - x_j)^4 \psi(x_1,\dots,x_n)$. Thus

$$\det \hat{A} = C(X) \prod_{i>j}^{n} (x_i - x_j)^4 \ .$$

Now since the degree of $\det \hat{A}$ is the same as the degree of $\prod_{i>j}^{n} (x_i - x_j)^4$, $C(X)$ must be a constant and the proof is complete.

Lemma 2.3.2: With the same assumptions as in Lemma 2.3.1 the matrix $\hat{A}_{22} - \hat{A}_{21}\hat{A}_{11}^{-1}\hat{A}_{12}$ is nonsingular.

Proof: Note that $\hat{A}_{11}$ is the well known Vandermond matrix which is nonsingular. Now consider

$$\hat{B} = \begin{pmatrix} I & 0 \\ -\hat{A}_{21}\hat{A}_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} \\ 0 & -\hat{A}_{21}\hat{A}_{11}^{-1}\hat{A}_{12} + \hat{A}_{22} \end{pmatrix} .$$

Since

$$\det \hat{B} = \det \hat{A} = (\det \hat{A}_{11})(\det(\hat{A}_{22} - \hat{A}_{21}\hat{A}_{11}^{-1}\hat{A}_{12}) \neq 0 ,$$

the proof is complete.

Now with the assumption as in Lemma 2.3.1, if we let $y_1^T = (a_1', a_2', \ldots, a_n')$, $y_2^T = (a_1x_1', \ldots, a_nx_n')$ and $a^T = (a_1, a_2, \ldots, a_n)$, then the system (2.3.2) becomes

$$\begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} B_1 a \\ B_2 a \end{pmatrix}$$

which is linear in y_1, y_2 and can be written as

$$y_1^T = \hat{A}_{11}^{-1}[B_1 - \hat{A}_{12}(\hat{A}_{22} - \hat{A}_{21}\hat{A}_{11}^{-1}\hat{A}_{12})^{-1}(B_2 - \hat{A}_{21}\hat{A}_{11}^{-1}B_1)]a$$

(2.3.4)

$$y_2^T = (\hat{A}_{22} - \hat{A}_{21}\hat{A}_{11}^{-1}\hat{A}_{12})^{-1}(B_2 - \hat{A}_{21}\hat{A}_{11}^{-1}B_1)a \ .$$

The above system is a system of nonlinear differential equations for the Gauss-Galerkin points $\{x_k(t)\}$ coupled with a linear system of differential equations for the Gauss-Galerkin weights $\{a_k(t)\}$.

Now assume that the system (2.3.2), (2.3.3) has a solution $\{x_k(t)\}$ and $\{a_k(t)\}$ where

$$X(0)^T = (a_1(0),\ldots,a_n(0),x_1(0),\ldots,x_n(0))$$

has been chosen as the Gauss-Christoffel approximation of $p(0,x)$. By the continuity of $x_k(t)$ and $a_k(t)$ it is obvious that there exists an interval $[0,T]$ so that for each $t \in [0,T]$, $a_k(t)$ is positive for each k, and $x_1(t), \ldots, x_n(t)$ are distinct. Also by Lemma 2.3.1 the systems (2.3.2) and (2.3.4) are equivalent on $[0,T]$. On such an interval the system (2.3.2) can be written as

$$X' = A^{-1}BX = F(X) \tag{2.3.5}$$

because $A = \hat{A} \cdot \mathrm{diag}(1,\ldots,1,a_1,\ldots,a_n)$ which is invertible. Also

$$\frac{\partial F(X)}{\partial x_i} = \frac{\partial (A^{-1}BX)}{\partial x_i} = \frac{\partial A^{-1}}{\partial x_i} BX + A^{-1} \frac{\partial BX}{\partial x_i}$$

$$= -A^{-1} \frac{\partial A}{\partial x_i} A^{-1}BX + A^{-1} \frac{\partial BX}{\partial x_i}$$

$$= -A^{-1} \frac{\partial A}{\partial x_i} X' + A^{-1} \frac{\partial BX}{\partial x_i}$$

Since $x_i(t)$ and $a_i(t)$ for $i = 1,\ldots,n$ are continuous functions on $[0,T]$ and A^{-1} exists; it follows that A^{-1}, $\frac{\partial A}{\partial x_i}$ and $\frac{\partial BX}{\partial x_i}$ are all bounded. Therefore, $F(X)$ is Lipschitz w.r.t x_i for each i. It is easy to prove that F is Lipschitz w.r.t. X. Thus on $[0,T]$ the system (2.3.4) has a unique solution

$$\mu_n(t) = \sum_{k=1}^{n} a_k(t) \delta_{x_k(t)}$$

which approximates the law of the process of the solution to the stochastic differential equation (2.1.1), (2.1.2).

The corresponding approximated moments of the process are

$$(2.3.6) \qquad m_k(t) = \int_{r_1}^{r_2} x^k(t) d\mu_n(t)$$

$$= \sum_{i=1}^{n} a_i(t) x_i^k(t) \qquad k = 0,1,\ldots \ .$$

In the next section we discuss an alternative way to compute the above approximated moments, in the case where the coefficients of the stochastic differential

equations are polynomials.

2.4 The Hankel System of Moments: In this section we assume that the coefficients of the stochastic differential equation (2.1.1) are polynomials in x. From equation (2.3.6) we have

$$(2.4.1)\qquad \frac{dm_k(t)}{dt} = \frac{d}{dt}\sum_{i=1}^{n} a_i(t)x_i^k(t)$$

$$= \sum_{i=1}^{n} a_i'(t)x_i^k(t) + \sum_{i=1}^{n} ka_i(t)x_i^{k-1}x_i'(t)$$

for $k = 0,1,\ldots,2n-1$. By (2.3.6) the sum $\sum_{i=1}^{n} a_i(t)(Lf_k)(x_i(t))$ becomes a polynomial $g_k(m_1,m_2,\ldots)$, involving finitely many moments. From the system (2.3.2) and (2.3.3) we have

$$(2.4.2)\qquad \frac{dm_k}{dt} = g_k(m_1(t),m_2(t),\ldots) \quad k = 0,1,2,\ldots,2n-1$$

$$(2.4.3)\qquad m_k(0) = \text{given} \quad k = 0,1,\ldots,2n-1$$

Let us suppose that $a(x,t)$ is a polynomial of degree q in x and $b(x,t)$ is a polynomial of degree ℓ in x. Recall that $Lf(x)$ is given by

$$(Lf)(x) = a(x,t)\frac{\partial f}{\partial x} + \frac{1}{2}b^2(x,t)\frac{\partial^2 f}{\partial x^2}.$$

For $f(x) = x^k$, the polynomial g_k in (2.4.2) involves $m_{k-2}, m_{k-1},\ldots,m_{p_k}$ where $p_k = \max\{k+q-1,\ 2\ell+k-2\}$.

For $q \leq 1$ and $\ell \leq 1$, we have $p_k \leq 2n-1$ so the system (2.4.2) is closed. On the other hand, if $q \geq 2$ or $\ell \geq 2$, $p_{2n-1} = \max\{2n+q-2,\ 2\ell+2n-3\} \geq 2n$, and the system (2.4.2) is not closed.

Theorem 2.4.1: Suppose that the system (2.3.2) and (2.3.3) has a solution in $[0,T]$ with nodes that remain distinct as we have assumed. Then the system (2.4.2) may be made closed.

Proof: We shall show that it is possible to express all the moments that appear in (2.4.2) with order higher than $2n-1$ in terms of the lower moments. To do this we define

$$P_n(x,t) = (x-x_1(t))(x-x_2(t)) \ldots (x-x_n(t)) .$$

Since $\mu_n(t)$ is a measure concentrated on $x_1(t),\ldots,x_n(t)$, we have

$$(2.4.4) \qquad \int_{r_1}^{r_2} x^k\, p_n(x,t)\, d\mu_n(t) = 0 \qquad k = 0,1,2,\ldots .$$

Let $\sigma_i(t)$ $i = 1,2,\ldots,n$ be the sum of all products of i of the numbers $x_1(t), x_2(t),\ldots,x_n(t)$, without permutation or repetition. Then (2.4.4) becomes

$$(2.4.5) \qquad m_{n+k} - \sigma_1 m_{n+k-1} + \cdots + (-1)^i \sigma_i m_{n+k-i} + \cdots + (-1)^n \sigma_n m_k = 0 \qquad \text{for } k = 0,1,2,\ldots$$

Let

$$M_{n,k} = \begin{bmatrix} m_k & m_{k+1} & \cdots & m_{n+k} \\ m_{k+1} & m_{k+2} & \cdots & m_{n+k+1} \\ \vdots & & & \\ m_{n+k} & m_{n+k+1} & \cdots & m_{2n+k} \end{bmatrix}$$

Any n consecutive equations of the system (2.4.5) form a linear system of equations with $\sigma_1,\sigma_2,\ldots,\sigma_n$ as unknowns. Therefore, we get

$$(2.4.6) \qquad \Delta_{n,k} = \det M_{n,k} = 0 \quad k = 0,1,2,\ldots$$

By the assumption that the nodes are distinct we see that for $\ell < n$ and for real numbers $y_0,y_1,\ldots,y_\ell$, we have $\int_{r_1}^{r_2} x^k (y_0 + y_1 x + \cdots + y_\ell x^\ell)^2 \, d\mu_n$ positive, i.e. for any $\ell < n$, $Q_{\ell,k}(y) = \sum_{i,j=0}^{\ell} m_{i+j+k}\, y_i y_j > 0$, where $y = (y_1,\ldots,y_\ell)$. This is equivalent to saying that in $M_{n,k}$ the determinants of all principal minor submatrices are positive. It is thus possible to determine m_{2n} in terms of $m_0,m_1,\ldots,m_{2n-1}$ from $\Delta_{n,0} = 0$. Similarly, from $\Delta_{n,1} = 0$ we get m_{2n+1} in terms of $m_1,m_2,\ldots,m_{2n}$ and in general from $\Delta_{n,k} = 0$ we get m_{2n+k} in terms of previous moments. Thus the system (2.4.2) may be made closed.

Remark 2.4.2. Given a sequence of real numbers $\{m_k\}_{k=1}^{\infty}$ which satisfy (2.4.6) with the additional assumption that the determinant of all minor submatrices of $M_{n,k}$ $(k = 0,1,\ldots)$ are positive, we can obtain a unique measure concentrated on n distinct points in $(0,\infty)$ whose moments are the given sequence. For a proof we refer to a paper written by Ernest Fischer [9]. Actually the proof is based on the fact that the quadratic forms $\sum_{i,j=0}^{n-1} m_{i+j}x_ix_j$ and $\sum_{i,j=0}^{n-1} m_{i+j+1}x_ix_j$ are positive definite. It is interesting to know that if the determinants of $M_{n,0}$ and $M_{n,1}$ are non-negative for all n, then there exist a measure μ with $\{m_k\}_{k=1}^{\infty}$ as moments. If the determinants of $M_{n,0}$ and $M_{n,1}$ are positive for all n, then μ is a measure whose spectrum cannot be reduced to finitely many point.

Remark 2.4.3: To close the system (2.4.2) in practice we need a simple computer algorithm to write higher moments in terms of previous ones. To do this we use the following theorem (George E. Forsythe and Cleve B. Moler [10]).

Theorem 2.4.4: For a given square matrix A of order n, let A_k denote the principal minor submatrix made from the first k rows and column. Assume that

$\det(A_k) \neq 0$ for $k = 1,2,\ldots,n-1$. Then there exist a unique lower triangular matrix $L = (\ell_{ij})$, with $\ell_{11} = \ell_{22} = \cdots = \ell_{nn} = 1$ and a unique upper triangular matrix $U = (u_{ij})$ so that $LU = A$. Moreover, $\det A = u_{11} \cdot u_{22} \cdots u_{nn}$. Here $A = M_{n,k}$ is a symmetric matrix of order $n+1$; therefore, by $A = LU = U^T L^T$ and by uniqueness of L and U we have $\ell_{ij} = u_{ji}/u_{jj}$ if $i \neq j$. Thus

$$u_{1i} = a_{1i} \quad \text{for} \quad 1 \leq i \leq n+1 .$$

Now multiplying all rows of L with index greater than or equal to 2 by the second column of U we get

$$u_{22} = a_{22} - u_{12}^2/u_{11}$$

and

$$u_{2i} = a_{i2} - u_{1i} \cdot u_{12}/u_{11}, \quad 3 \leq i \leq n+1 .$$

Continuing this procedure we get the following algorithm to compute $u_{11}, u_{22}, \ldots, u_{n+1}$

For $I = 1, n+1$ Do

$$u_{1i} = a_{1i}$$

For $j = 2, n+1$ Do

$$u_{jj} = a_{jj} - \sum_{k=1}^{j-1} u_{kj}^2/u_{kk} \qquad (2.4.7)$$

For $I = j+1, n+1$ Do

$$u_{ij} = a_{ij} - \sum_{k=1}^{j-1} u_{ki}u_{kj}/u_{kk} \cdot$$

In order to obtain m_{2n+k} in terms of $m_k, m_{k+1}, \ldots,$ m_{2n+k-1} which is a consequence of the equation $\Delta_{n,k} = 0$, we modify the last step which completes the algorithm (2.4.7) for the case $j = n+1$ and set $u_{n+1\ n+1} = 0$ which gives

$$m_{2n+k} = a_{n+1\ n+1} = \sum_{k=1}^{n} u_{n+1k}^2/u_{kk} \cdot$$

We now prove the following theorem.

Theorem 2.4.5: With the same assumptions as in Theorem 2.4.1 the systems (2.3.2) implies the system (2.4.2). Conversely the system 2.4.2 implies the system (2.3.2) provided $\Delta_{nk} = 0$ $k = 0,1,2,\ldots$ and the principal submatrices of M_{nk} have positive determinants.

Proof: It is understood that we use equivalent initial values. We obtain the system (2.4.2) from the system (2.3.4) and it is obvious that any solution to the system (2.3.4) is a solution for the system (2.4.2).

Conversely, by Remark 2.4.2 any solution of the closed system (2.4.2) induces a unique measure $\mu_n(t)$ concentrated on n points $x_1(t),\ldots,x_n(t)$ with corresponding weights $a_1(t),\ldots,a_n(t)$. Therefore by (2.3.6) we obtain the system (2.3.4) from (2.4.2) with $\mu_n(t)$ as a solution. Therefore any solution of the system (2.4.2) is also a solution to the system (2.3.4) and the proof is complete.

2.5 Comparison of the Gauss-Christoffel and the Gauss-Galerkin Measures: The Gauss-Christoffel approximation which was discussed in Section 2.2 can be considered as the "best possible" approximation of a measure because it determines the first $2n-1$ moments the same as exact moments. Now assuming that the system (2.3.4) has a solution on the interval $[0,T]$ we wish to compare the Gauss-Christoffel approximation

$$\tilde{\mu}_n(t) = \sum_{k=1}^{n} \tilde{a}_k(t)\,\delta_{\tilde{x}_k(t)} ,$$

with the Gauss-Galerkin approximation

$$\mu_n(t) = \sum_{k=1}^{n} a_k(t)\,\delta_{x_k(t)} .$$

To do this we need the following lemma and corollary (Hale [13]).

Lemma 2.5.1: Let $w(t,u)$ be continuous on an open connected set $\Omega \subset R^2$ and such that the initial value problem for the scalar equation

$$u' = w(t,u) \tag{2.5.1}$$

has a unique solution. If $u(t)$ is a solution of (2.5.1) on $a \leq t \leq b$ and $v(t)$ is a solution of (2.5.1) on $a \leq t < b$ with $v(a) \leq u(a)$, then $v(t) \leq u(t)$ for $t \in [a,b]$.

Corollary 2.5.2: Suppose that $w(t,u)$ satisfies the conditions of Lemma 2.5.1 for $a \leq t < b$, $u \geq 0$, and let $u(t) \geq 0$ be a solution of (2.5.1) on $a \leq t < b$. If $f : [a,b] \times R^n \to R^n$ is continuous and

$$|f(t,X)| \leq w(t,|X|), \quad a \leq t < b, \quad X \in R^n$$

then the solution of

$$X' = f(t,X), \quad |X(a)| \leq u(a)$$

exists on $[a,b)$ and $|X(t)| \leq u(t)$ for $t \in [a,b]$. Now the system (2.2.10) can be written in matrix form as

$$\tilde{A}\tilde{X}' = \tilde{B}\tilde{X} + e(t) \tag{2.5.2}$$

where $\tilde{A}$ and $\tilde{B}$ are of the same form as A and B in (2.3.2) with the Gauss-Galerkin nodes and weights

replaced by Gauss-Christoffel ones and

$$e(t)^T = (e_1(t), e_2(t), \ldots, e_n(t))$$

where

$$e_i(t) = E[Lf_i] \ .$$

Note that the errors $e_i(t)$ depend on the nodes $\{\tilde{x}_i(t)\}$ and the weights $\{\tilde{a}_i(t)\}$. By Lemma 2.3.1, $\tilde{A}$ is invertible and equation (2.5.2) can be written as

$$\tilde{X}' = \tilde{A}^{-1}\tilde{B}\tilde{X} + \tilde{A}^{-1}e(t) \tag{2.5.3}$$

or

$$\tilde{X}' = F(\tilde{X}) + E(t) \tag{2.5.3)'}$$

where $E(t) = E(\tilde{X}, \tilde{a}, t)$ is regarded as a known function. Again note that by neglecting $E(t)$ from the system (2.5.3) we obtain the Gauss-Galerkin system. In the same way as before we can show that $F(\tilde{X})$ is Lipschitz with respect to $\tilde{X}$ with some Lipschitz constant M_n. The two systems

$$\tilde{X}'(t) = F(X(t)) + E(t)$$

$$X'(t) = F(X(t))$$

lead to the system

$$(\tilde{X} - X)' = F(\tilde{X}) - F(X) + E(t) \ . \tag{2.5.4}$$

We now consider the scalar initial value problem

$$u' = Mu + |E(t)|$$

$u(0) = 0$, which has the solution

$$u(t) = e^{M_n t} \int_0^t |E(s)| e^{-M_n s}\, ds .$$

Also we have

$$|(\tilde{X} - X)'| \leq |F(\tilde{X}) - F(X)| + |E(t)| \leq M_n |\tilde{X} - X| + |E(t)| .$$

Thus by Corollary 2.5.2, with $\tilde{X}(0) = X(0)$ (as we always take Gauss-Christoffel nodes and weights as initial values), we have for $t \in [0,T]$

$$(2.5.5) \quad |\tilde{X}(t) - X(t)| \leq u(t) = e^{M_n t} \int_0^t |E(s)| e^{-M_n s}\, ds .$$

Thus we have proved

Theorem 2.5.3: Assuming that the system (2.3.2) has a solution $\mu_n(t) = \sum_{k=1}^{n} a_k(t) \delta_{x_k(t)}$ for $t \in [0,T]$ with distinct nodes and with $\mu_n(0)$ the Gauss-Christoffel approximation of $p(0,x)$ and $\tilde{\mu}_n(t) = \sum_{k=1}^{n} \tilde{a}_k(t) \delta_{\tilde{x}_k(t)}$ the Gauss-Christoffel approximation of $p(t,x)$, then we have

$$(2.5.6) \qquad |\tilde{X}(t) - X(t)| \leq e^{M_n t} \int_0^t |E(s)| e^{-M_n s} \, ds$$

where

$$\tilde{X}(t) = (\tilde{a}_1(t), \ldots, \tilde{a}_n(t), \tilde{x}_1(t), \ldots, \tilde{x}_n(t))$$

and

$$X(t) = (a_1(t), \ldots, a_n(t), x_1(t), \ldots, x_n(t))$$

Remark 2.5.4. In equation (2.5.6) $E(s) = \tilde{A}^{-1} e(s)$ where $e(s)$ is the Gauss-Christoffel error and $|E(s)| \leq c_n |e(s)|$ for some constant c_n. One may wish to use (2.5.6) to prove convergence of the Gauss-Galerkin weights and nodes to the Gauss-Christoffel ones as $n \to \infty$. However the dependence of c_n and the Lipschitz constant M_n on n does not make such convergence theorem possible.

Remark 2.5.5. If the coefficients of the stochastic differential equation are polynomials, then the vector $e(t)$ has its components equal to zero except for finitely many of some components. This number of non-zero components depends on the degree of the coefficients. If the degree of the coefficients is less than or equal to one, then $e(t) = 0$ and therefore, $\tilde{X}(t) = X(t)$ for all t. Thus for a linear stochastic differential equation the Gauss-Galerkin and Gauss-Christoffel approximations are the same.

2.6 Steady-State Solutions: For some stochastic differential equations the density $p(t,x)$ of the solution $X(t)$ as $t \to \infty$ becomes independent of t. In other words, the influence of the initial state fades away and the system tends to a steady-state governed by the stationary solution. In this case the density satisfies (2.1.3) with left hand side replaced by 0. Now assuming that a stochastic differential equation has a steady-state solution the corresponding system to (2.5.2) becomes

$$\tilde{B}\tilde{X} + e = 0 \tag{2.6.1}$$

or

$$F(\tilde{X}) + e = 0 \tag{2.6.1}'$$

which is a nonlinear system of equations and yields the Gauss-Christoffel steady-state approximation. It is to be noted that the system (2.6.1) is not closed in so far as $\tilde{X}$ is concerned as the steady-state density is involved in e and thus cannot be used for finding the steady-state approximated solution. The corresponding system of equations given by the Gauss-Galerkin method has the form

$$BX = 0 \tag{2.6.2}$$

or

$$F(X) = 0 \tag{2.6.2}'$$

which is a closed system of nonlinear equations for X. Numerical methods such as Newton's method or the bisection method can be used to obtain the Gauss-Galerkin steady-state solution. Also an upper bound for the term $|\tilde{X} - X|$ as in Theorem 2.5.5 may be derived as follows. We define the function

$$H(X,s) = F(X) + se = 0 \quad . \tag{2.6.3}$$

Obviously for $s = 0$ we have the Gauss-Galerkin steady-state equations and for $s = 1$ the Gauss-Christoffel equations. From (2.6.3) we have

$$(DF(X))X'(s) + e = 0$$

and this leads to (Chow and Hale [3], p. 21)

$$|\tilde{X} - X| \leq \|DF(X)\|\,|e| \quad .$$

CHAPTER III

CONVERGENCE OF THE GAUSS-GALERKIN METHOD

For a given interval (r_1, r_2) where $-\infty \leq r_1 < r_2 \leq \infty$ and given $p(x)$, assume that we have a sequence of integration formulas

$$\Sigma_n(f) \approx I(f)$$

where

$$\Sigma_n(f) = \sum_{k=1}^{n} a_k^{(n)} f(x_k^{(n)}) \quad \text{and} \quad I(f) = \int_{r_1}^{r_2} f(x)p(x)dx$$

$n = 1,2,\ldots$. It is of importance to know under what conditions the sequence $\Sigma_n(f)$ converges to the true value of the integral as $n \to \infty$. In this chapter we study the convergence of the integration formulas obtained by the Gauss-Galerkin method in Chapter II.

3.1 Convergence of the Gauss-Christoffel Integration Formulas: We shall begin by stating some results about the convergence of the Gauss-Christoffel integration formulas.

If (r_1, r_2) is a finite interval, it is known that for a given positive measure μ with finite moments, we

have

$$(3.1.1)\qquad \lim_{n\to\infty} \Sigma_n(f) = \lim_{n\to\infty} \int_{r_1}^{r_2} f(x)\, d\tilde{\mu}_n$$

$$= \int_{r_1}^{r_2} f(x)\, d\mu = I(f) \ .$$

for any continuous function $f(x)$ defined on $[r_1, r_2]$ (Stroud [21] p. 142).

In the infinite interval case we generally do not have the same result as in the above. We state some theorems about the convergence of the Gauss-Christoffel integration formulas in the $(0,\infty)$ and $(-\infty,\infty)$ cases.

Theorem 3.1.1: Let μ be a positive measure defined on $[0,\infty)$ whose moments of all order are finite. Let $\tilde{\mu}_n$ be the n-point Gauss-Christoffel approximation to μ. Then there exists a positive measure ν defined on $[0,\infty)$ such that

$$(3.1.2)\qquad \lim_{n\to\infty} \int_0^\infty f(x)\, d\tilde{\mu}_n = \int_0^\infty f(x)\, d\nu$$

for any continuous function $f(x)$ on $[0,\infty)$ such that as $x \to \infty$, $f(x)$ is dominated by a polynomial. (Shohat and Tamarkin [19], p. 121).

Theorem 3.1.2: Let μ be a positive measure defined on $(-\infty,\infty)$ whose moments of all orders are finite. Let $\tilde{\mu}_n$ be the n-point Gauss-Christoffel approximation to μ. Then there exists a positive measure ν defined on

$(-\infty,\infty)$ and a subsequence $\{\tilde{\mu}_{n_k}\}$ such that

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} f(x)\, d\tilde{\mu}_{n_k} = \int_{-\infty}^{\infty} f(x)\, d\nu \tag{3.1.3}$$

for any continuous function $f(x)$ on $(-\infty,\infty)$ such that as $|x| \to \infty$, $f(x)$ is dominated by a polynomial with non-negative coefficients.

Corollary 3.1.3: If the measure μ in Theorem 3.1.1 or 3.1.2 is uniquely determined by its moments, then the measure ν will be the measure μ "substantially".

Remark 3.1.4: By "substantially" here we mean

$$\int f(t)\, d\mu = \int f(t)\, d\nu \tag{3.1.4}$$

for any continuous function $f(t)$ which vanishes for all sufficiently large values of $|t|$.

J.V. Uspenski [25] has given sufficient conditions for a measure μ with density $p(x)$ that is determined uniquely by its moments (in both $(0,\infty)$ and $(-\infty,\infty)$ cases). For example in the $(0,\infty)$ case he proves that if there exist constants C and R such that

$$m_n = \int_0^{\infty} x^n d\mu < c(2n+1)!R^{2n} \quad \text{for all} \quad n \geq 1 , \tag{3.1.5}$$

then μ is uniquely determined by its moments. We refer to Shohat and Tamarkin ([19], p.22) for examples showing that these upper bounds are in fact sharp. In the next

sections we shall see how these bounds are essential in the theory of convergence of the Gauss-Galerkin integration formulas.

3.2 Assumptions and Boundary Conditions: The Gauss-Galerkin method proposed in Chapter II is for the approximation of the density function $p(t,x)$ associated with a process governed by the stochastic differential equation

$$dX(t) = a(X(t),t)\,dt + b(X(t),t)\,dW(t) \tag{3.2.1}$$

$$X(0) = \text{given} \tag{3.2.2}$$

on the interval $I = (r_1, r_2)$ where $0 \leq r_1 < r_2 \leq \infty$. In this section we shall state precisely the class of functions of $a(x,t)$ and $b(x,t)$ that are allowed and discuss the types of boundary conditions that are considered in this dissertation.

We suppose that $a(x,t)$ and $b(x,t)$ are continuous functions on $I \times [0,T]$ which satisfy assumptions of the existence and uniqueness theorems of Chapter I. More precisely, we assume that there are constants K and K_* such that

$$|a(x,t)| \leq K(1+|x|), \quad |b(x,t)| \leq K(1+|x|) \tag{3.2.3}$$

and

$$|a(x,t) - a(\bar{x},t)| \leq K_*|x-\bar{x}|, \quad |b(x,t) - b(\bar{x},t)| \leq K_*|x-\bar{x}|, \tag{3.2.4}$$

for $x \in I$ and $0 \leq t \leq T$. We assume that $X(0)$ is a random variable having a density with finite moments of all order and is independent of $\mathcal{F}(W(t), 0 \leq t \leq T)$.

It is clear that with these restrictions on the $a(x,t)$ and $b(x,t)$, there are non-negative constants a_1, a_2, b_1 and b_2 such that on $I \times [0,T]$ we have

$$(3.2.3) \quad |a(x,t)| \leq a_1 x + b_1 \quad \text{and} \quad |b(x,t)| \leq a_2 x + b_2 .$$

Also it is well known (Arnold [1], p. 116) that with these assumptions, for any $t \in [0,T]$, $X(t)$ has finite moments of all order.

Now consider the Gauss-Galerkin approximations

$$\mu_n(t) = \sum_{k=1}^{n} a_k^{(n)}(t) \, \delta_{x_k^{(n)}(t)} \qquad n = 1,2,\ldots$$

as a sequence of integration formulas $\Sigma_{n,t}(f) \approx I_t(f)$ where

$$I_t(f) = \int_{r_1}^{r_2} f(x) p(t,x) \, dx$$

and

$$\Sigma_{n,t}(f) = \sum_{k=1}^{n} a_k^{(n)}(t) f(x_k^{(n)}(t)) .$$

Before discussing the convergence of the integration formulas at any $t > 0$, it is essential to have convergence of $\Sigma_{n,0}(f)$ to $I_0(f)$ for $f(x)$ belonging to some appropriate class of functions. For example

it is seen in Section 3.1 that if $p(0,x)$ is determined uniquely by its moments, then we have convergence of $\Sigma_{n,0}(f)$ to $I_0(f)$ for all continuous functions $f(x)$ such that as $x \to \infty$, f is dominated by a polynomial.

Throughout this chapter by "the stochastic differential equation satisfies the condition A on $(r_1,r_2) \times [0,T]$" we mean:

Condition A

1. The coefficients $a(x,t)$ and $b(x,t)$ are continuous on $(r_1,r_2) \times [0,T]$. They are such that a unique solution exists and satisfy the inequality (3.2.3).

2. $X(0)$ is a random variable independent of $\mathfrak{F}(W(t), 0 \leq t \leq T)$ with density $p(0,x)$ having finite moments of all order.

3. The Gauss-Christoffel integration formulas $\Sigma_n(f)$ for $p(0,x)$ converge to the true value of the integral $I(f)$ for each function f that is continuous and such that as $x \to \infty$, f is bounded by a polynomial.

4. The boundary conditions on $p(t,x)$ for each t is such that

$$(L^* p, f_i) = (p, L f_i) \tag{3.2.5}$$

for $i = 1,2,\ldots,2n$ and $f_i(x) = x^{i-1}$.

5. The eigenvalue problem defined by

$$Lu = \lambda u$$

and the boundary conditions above generates an infinite set of eigenfunctions $\{u_n(x)\}$ such that the set $\{u_n e^{-\theta_o x}\}$ is complete in $L_2(r_1,r_2)$ for some $\theta_o > 0$ and such that $u_n' e^{-\theta_o x} \to 0$ as $n \to \infty$. Furthermore any sufficiently smooth function $\varphi(x)$ can be approximated uniformly by combinations of the set $\{u_n e^{-\theta_o x}\}$.

6. The density $p(t,x)$ governed by

$$\frac{\partial p}{\partial t} = L^* p \qquad x \in (r_1,r_2) \ ,$$

and the boundary conditions above exists and is unique.

We shall now make some comments on these conditions.

Remark 3.2.1. Equation (3.2.5) for $f_1(x) = 1$ implies that

$$\frac{d}{dt}(p,1) = 0.$$

This means that boundary conditions are such that the probability in (r_1,r_2) is preserved.

Remark 3.2.2. The assumptions made in Condition 4 may be analyzed by appealing to the spectral theory of the operator L involved. For a smooth function $\varphi(x)$, the uniform convergence of its formal eigenfunction

expansion depends on the behavior of the expansion of the Green's function in terms of the eigenfunctions. See [4] or [17]. Actually for the finite interval case condition A5 is automatically true. See [4] Theorem 4.1.

Remark 3.2.3. As Feller [6,b] points out in diffusion theory one usually starts with the assumption that the transition probability $P(x,t,B) = P(0,x,t,B)$ has a probability density $p(x,t,y) = p(0,x,t,y)$ which for fixed y satisfies the Fokker-Planck equation. As we have seen in Chapter I (equation (1.10.3)) the density $p(0,x)$ and $p(x,t,y)$ determines $p(t,y)$ the solution of the Fokker-Planck equation. For simplicity we assume condition 6 here and refer to Gikhman and Skorokhod ([9,b]) where the existence of the density $p(x,t,y)$ under more restrictive conditions on a and b is discussed in detail.

Remark 3.2.4. The classical boundary conditions that make L and L^* adjoint to each other are discussed by Feller [8,a]. Equation (3.2.5) is a special case of the above for polynomials.

Let us present some examples of stochastic differential equations defined on (r_1,r_2) where

$0 \leq r_1 < r_2 \leq \infty$ for which under suitable boundary conditions equation (3.2.3) is valid. We give examples defined on $(0,1)$ and $(0,\infty)$ only since by Ito's formula (using linear functions) any (r_1,r_2) can be treated similarly. The idea for the proof of the following examples are discussed in the work of S. Ethier [7].

Example 3.2.4. Consider the stochastic differential equation

$$dX(t) = a(X(t))dt + b(X(t))dW(t) \tag{3.2.6}$$

$$X(0) = \text{given} \tag{3.2.7}$$

define on $(0,1)$. Assume that $a(x)$ and $b(x)$ are continuous on $(0,1)$, $b(0) = b(1) = 0$, $a(0) \geq 0$ and $a(1) \leq 0$. Define $\sigma(x) = b^2(x)$ and

$$\rho(x) = \begin{cases} 1 & \text{if } x \geq 1 \\ x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x \leq 0 \end{cases}$$

since $a \circ \rho$ and $\sigma \circ \rho$ are bounded, the stochastic differential equation

$$dX(t) = a \circ \rho(X(t))dt + (\sigma \circ \rho)^{\frac{1}{2}}(X(t))dW(t)$$

$$X(0) = \text{given}$$

has a solution $(\Omega,\mathcal{F},X(t),P)$. Now the equation

$$(3.2.8)\qquad X_{t_1} - X_{t_0} = \int_{t_0}^{t_1} a \circ \rho(X_s)\,ds + \int_{t_0}^{t_1} (\sigma \circ \rho)^{\frac{1}{2}}(X_s)\,dW(s)$$

implies that

$$P\{X(s) \in (-\infty,0) \text{ for } t_0 \le s \le t_1 \text{ and } X(t_1) - X(t_0) < 0\} = 0$$

and

$$P\{X(s) \in (1,\infty) \text{ for } t_0 \le s \le t_1 \text{ and } X(t_1) - X(t_0) > 0\} = 0$$

whenever $0 < t_0 < t_1$. Summing over all rationals t_0 and t_1 we obtain

$$P\{X_s \in (-\infty,0) \cup (1,\infty) \text{ for some } s > 0\} = 0 .$$

Therefore the probability is preserved in the interval $(0,1)$. Thus if $p(t,x)$ is the density of the stochastic differential equation (3.2.6) and (3.2.7) we have

$$(3.2.9)\qquad \int_0^1 p(t,x)\,dx = 1$$

Now (3.29) implies that (3.25) is valid for $f_1(x) = 1$. For $f(x) = x^k$, $k \ge 1$, we have

$$(L^{*}p,x^{k}) = \int_{0}^{1} \left(-\frac{\partial ap}{\partial x} + \frac{I}{2} \frac{\partial^{2} b^{2} p}{\partial x^{2}} \right) x^{k} dx$$

$$= x^{k}\left(-ap + bb'p + \frac{1}{2} b^{2} \frac{\partial p}{\partial x}\right)\Big|_{0}^{1} - \frac{1}{2} kx^{k-1} b^{2} p\Big|_{0}^{1}$$

$$+ (p,Lx^{k}) = (p,Lx^{k}) - \lim_{x \to 1} ap \ .$$

Therefore if $a(1) \neq 0$ the only boundary condition that we need here is $\lim_{x \to 1} p(t,x) = 0$.

Example 3.2.5. Consider the stochastic differential equation

$$dX(t) = a(X(t))dt + b(X(t))dW(t) \tag{3.2.10}$$

$$X(0) = \text{given} \tag{3.2.11}$$

defined on $(0,\infty)$. Assume that $b(0) = 0$ and $a(0) \geq 0$. Also assume that $a(x)$ and $b(x)$ satisfy the existence and uniqueness conditions. Define $\sigma(x)$ and $\rho(x)$ the same as in Example 3.2.4. Therefore the stochastic differential equation

$$dX(t) = a \circ \rho(X(t))dt + (\sigma \circ \rho)^{\frac{1}{2}}(X(t))dW(t)$$

$$X(0) = \text{given}$$

has a solution $(\Omega,\mathcal{F},X(t),P)$. The same equation as (3.2.8) implies that

$$P\{X(s) \in (-\infty,0) \quad t_0 \le s \le t_1 \text{ and } X(t_1) - X(t_0) < 0\} = 0$$

and therefore with the same argument we have

$$P\{X(s) \in (-\infty,0) \quad \text{for some} \quad s > 0\} = 0$$

and therefore equation (3.2.9) is true (only ∞ is replaced by 1). Thus (3.2.5) is valid for $f_1(x) = 1$. For $f(x) = x^k$, $k \ge 1$ we have

$$(L^*p,x^k) = (p,Lx^k) + x^k(-ap + bb'p + \frac{1}{2} b^2 \frac{\partial p}{\partial x}\Big|_0^\infty - \frac{1}{2} kx^{k-1}b^2 p\Big|_0^1$$

$$= (p,Lx^k) + \lim_{x\to\infty} x^k(-ap + bb'p + \frac{1}{2} b^2 \frac{\partial p}{\partial x})$$

$$- \lim_{x\to\infty} kx^{k-1}b^2 p \ .$$

Therefore here if we impose the boundary condition

$$(3.2.12) \quad \lim_{x\to\infty} x^k p = 0 \quad \text{and} \quad \lim_{x\to\infty} x^k \frac{\partial p}{\partial x} = 0$$

the equation (3.2.5) is valid. In Section 3.6 we see that if $a(x) = bx + c$, $b(x) = a_2\sqrt{x}$, then the boundary conditions (3.2.12) are automatically valid.

3.3 Some Results from Differential Inequalities: To continue our preparations for the convergence theorems of the Gauss-Galerkin formulas we need some lemmas from differential inequalities (Szarski [24]).

Let $Y = (y_1,\ldots,y_n)^T$ and $\tilde{y} = (\tilde{y}_1,\ldots,\tilde{y}_n)$ be two points of the n-dimensional space. We write

$$Y \leq \tilde{Y} \quad \text{if} \quad y_i \leq \tilde{y}_i \quad (i = 1,2,\ldots,n)$$

and

$$Y < \tilde{Y} \quad \text{if} \quad y_i < \tilde{y}_i \quad (i = 1,2,\ldots,n) \ .$$

For a fixed index j we write

$$Y \leq^j \tilde{Y} \quad \text{if} \quad y_i \leq \tilde{y}_i \quad (i = 1,\ldots,n) \quad \text{and} \quad y_j = \tilde{y}_j$$

Let a system of functions $f_i(t,y) = f_i(t,y_1,\ldots,y_n)$ $(i = 1,\ldots,n)$ be defined in a region D. We have

Condition V+ (V-): System $f_i(t,y)$ $i = 1,2,\ldots,n$ is said to satisfy condition V+ (V-) with respect to y in D if for every fixed index j the function $f_j(t,Y)$ is increasing (decreasing) w.r.t. each variable $y_1,\ldots,y_{j-1}, y_{j+1},\ldots,y_n$ separately.

Condition W+ (W-): System $f_i(t,y)$ $(i = 1,2,\ldots,n)$ is said to satisfy condition W+ (W-) with respect to Y in D if for every fixed index j, the following implication holds true

$$Y \leq^j \tilde{Y}, (t,Y) \in D, (t,\tilde{Y}) \in D \Rightarrow f_j(t,Y) \leq f_j(t,\tilde{Y})$$

$$(Y \leq^j \tilde{Y}, \ (t,Y) \in D, \ (t,\tilde{Y}) \in D \Rightarrow f_j(t,Y) \geq f_j(t,\tilde{Y})) \ .$$

Definition 3.3.1 (Dini's derivatives) For a real-valued function $\varphi(t)$, defined in some neighborhood of the point t_0, we denote $D^+\varphi(t_0)$, $D_+\varphi(t_0)$, $D^-\varphi(t_0)$ and $D_-\varphi(t_0)$ by

$$D^+\varphi(t_0) = \limsup_{h\to 0^+} \frac{\varphi(t_0+h) - \varphi(t_0)}{h}$$

$$D_+\varphi(t_0) = \liminf_{h\to 0^+} \frac{\varphi(t_0+h) - \varphi(t_0)}{h}$$

$$D^-\varphi(t_0) = \limsup_{h\to 0^-} \frac{\varphi(t_0+h) - \varphi(t_0)}{h}$$

$$D_-\varphi(t_0) = \liminf_{h\to 0^-} \frac{\varphi(t_0+h) - \varphi(t_0)}{h}$$

(the values $+\infty$ and $-\infty$ being not excluded). Obviously if φ is differentiable at t_0 all four derivatives are equal.

Lemma 3.3.2. Let the right-hand side of the system

$$(3.3.1) \qquad \frac{dy_i}{dt} = \sigma_i(t,y_1,\ldots,y_n) \quad i = 1,2,\ldots,n$$

be defined in some open region D and satisfy in D condition W+ with respect to Y. Let $(t_0,Y_0) \in D$. Assume that $\varphi(t) = (\varphi_1(t),\ldots,\varphi_n(t))^T$ is continuous in $[t_0,\alpha)$ and that the curve $Y = \varphi(t)$ lies in D. Let $Y(t) = (y_1(t),\ldots,y_n(t))^T$ be an arbitrary solution of the system (3.3.1) passing through (t_0,Y_0) and defined in some interval $[t_0,\beta)$. Under these assumptions, if

(3.3.2) $$\varphi(t_0) < Y_0$$

and

$$D_-\varphi_i(t) \leq \sigma_i(t,\varphi_1(t),\ldots,\varphi_n(t)) \quad i = 1,\ldots,n$$

for $t \in (t_0,\alpha)$, then we have the inequality

$$\varphi(t) < Y(t) \quad \text{for} \quad t \in [t_0,\alpha) \cap [t_0,\beta) \ .$$

Corollary 3.3.3. If the system (3.3.1) in Lemma 3.3.2 has a unique solution, then Lemma 3.3.2 holds true when the strict inequalities $(<)$ are replaced by $(\leq)$.

For a proof of the lemma and corollary we refer to Szarski [24].

Corollary 3.3.4. Let the right hand sides of the systems

(3.3.3) $$\frac{dy_i}{dt} = \sigma_i(t,y_1,\ldots,y_n) \quad i = 1,2,\ldots,n$$

and

(3.3.4) $$\frac{dy_i}{dt} = f_i(t,y_1,\ldots,y_n) \quad i = 1,\ldots,n$$

be defined in some region D and let the right-hand side of the system (3.3.3) satisfy condition W+ with respect to Y. Let (3.3.3) have a unique solution and assume that $Y(t) = (y_1(t),\ldots,y_n(t))^T$ and $\varphi(t) = (\varphi_1(t),\ldots,\varphi_n(t))^T$ are solutions to (3.3.3) and (3.3.4)

respectively on $[t_0,\alpha)$. With these conditions and if

$$\varphi(t_0) \leq Y(t_0)$$

$$f_i(t,\varphi_1,\ldots,\varphi_n) \leq \sigma_i(t,\varphi_1,\ldots,\varphi_n) \quad (t,Y) \in D ,$$

then $\varphi(t) \leq Y(t)$ for $t \in [t_0,\alpha)$.

Proof. Notice that $\varphi(t)$ and $Y(t)$ satisfies all conditions of Corollary 3.3.4.

3.4 Helly's Theorems: The following two theorems are needed in the proof of the preparatory lemmas in Section 3.5. First we state

Definition 3.4.1. A sequence of measures $\{\mu_n\}$ is said to converge to a measure μ substantially if

$$\lim_{n\to\infty} \mu_n(I) = \mu(I)$$

for all (finite) intervals of continuity of μ.

Theorem 3.4.2. (The first theorem of Helly). Given a sequence $\{\mu_n\}$ of positive measures and uniformly bounded, then there exists a subsequence $\{\mu_{k_n}\}$ and a measure μ to which this subsequence converges substantially. Furthermore, if the sequence $\{\mu_n\}$ itself does not converge substantially to μ, then there exists another subsequence $\{\mu_{k_n'}\}$ converging substantially to another measure μ' which is not substantially equal to μ.

Theorem 3.4.3. (The second theorem of Helly).
Given a sequence $\{\mu_n\}$ of positive measures defined on $(0,\infty)$ and uniformly bounded, which converges substantially to a measure μ. Then

$$\lim_{n\to\infty} \int_0^\infty f(x)\,d\mu_n = \int_0^\infty f(x)\,d\mu$$

for any function $f(x)$ continuous in $(0,\infty)$ and such that, as intervals $I_N \uparrow (0,\infty)$, $\int_{I_N} f(x)\,d\mu_n \to \int_0^\infty f(x)\,d\mu_n$ uniformly in n.

3.5 Preparatory Lemmas: We shall develop some lemmas that are needed in the proofs of the convergence theorems in the next section. In these lemmas we shall consider measures $\mu_n(t)$ defined on the semi-infinite interval $(0,\infty)$. It is clear that these lemmas also hold for measures defined on any finite interval (r_1,r_2) contained in $(0,\infty)$.

Lemma 3.5.1. Assume that the stochastic differential equation

$$dX(t) = a(X(t),t)\,dt + b(X(t),t)\,dW(t) \tag{3.5.1}$$

$$X(0) = \text{given} \tag{3.5.2}$$

satisfies condition A on $(0,\infty) \times [0,T]$. For $t \in [0,T]$ let

$$\mu_n(t) = \sum_{k=1}^{n} a_k^{(n)}(t)\, \delta_{x_k^{(n)}(t)} \qquad n = 1,2,\ldots$$

denotes the n-point Gauss-Galerkin solution to the corresponding system (3.5.1) with non-negative weights and distinct nodes. Let ℓ be any fixed integer. Let $m_\ell^{(n)}(t) = \int_0^\infty x^\ell d\mu_n(t)$; then the set

$$\{m_\ell^{(n)}(t);\quad n > \tfrac{1}{2}(1+\ell),\quad t \in [0,T]\}$$

is bounded and equi-continuous.

Proof. We take $f_i(x) = x^i$, $i = 0,1,\ldots,2n-1$. The system (2.3.1) can be written as

$$dm_k^{(n)}(t)/dt = \int_0^\infty (Lf_k)(x)\, d\mu_n \qquad k = 0,1,\ldots,2n-1 .$$

Using condition A we have

$$\begin{aligned} dm_k^{(n)}(t)/dt &\le \int_0^\infty (a_1 x + b_1)\, \frac{\partial f_k}{\partial x}\, d\mu_n \\ &\quad + \tfrac{1}{2}\int_0^\infty (a_2 x + b_2)^2\, \frac{\partial^2 f_k}{\partial x^2}\, d\mu_n \\ &= g_k(m_{k-2}^{(n)}, m_{k-1}^{(n)}, m_k^{(n)}) . \end{aligned} \tag{3.5.3}$$

We let $m_0, m_1, \ldots, m_{2n-1}$ be the solutions to the system

$$\frac{dm_k}{dt} = g_k(m_{k-2}, m_{k-1}, m_k) \qquad k = 0,1,\ldots,2n-1 \tag{3.5.4}$$

with the same initial values as $m_k^{(n)}(0)$ $k = 0,1,\ldots,2n-1$. Then by Corollary 3.3.4 we have

$$(3.5.5) \qquad m_\ell^{(n)}(t) \leq m_\ell(t) \qquad n > \tfrac{1}{2}(\ell+1), \quad t \in [0,T] .$$

On the other hand the system (3.5.4) is the Hankel system for the stochastic differential equation

$$(3.5.6) \qquad dX(t) = (a_1 x + b_1)dt + (a_2 x + b_2)dW(t)$$

$$(3.5.7) \qquad X(0) = \text{given} .$$

If $\mu_t = \mathcal{L}(X(t))$ is the law of the solution to the equations (3.5.6) and (3.5.7) we know by Remark 2.5.5 that the linearity of (3.5.6) implies

$$m_k(t) = \int_0^\infty x^k d\mu_t \qquad k = 0,1,\ldots,2n-1 ,$$

that is, $m_k(t)$ is the exact k-th moment of the process. For each $0 \leq k \leq 2n-1$ $m_k(t)$ is a continuous function on $[0,T]$ and therefore bounded. Thus there exists a K_ℓ such that for all $n > \frac{1}{2}(\ell+1)$ and for all $0 \leq t \leq T$

$$(3.5.8) \qquad m_\ell^{(n)}(t) \leq m_\ell(t) \leq K_\ell .$$

The non-negativeness of $m_\ell^{(n)}(t)$ implies that $\{m_\ell^{(n)}(t),\ n > \frac{1}{2}(\ell+1),\ t \in [0,T]\}$ is bounded. To prove

the equicontinuity, we note that for given $n > \frac{1}{2}(1+\ell)$, by the mean-value theorem, we have

$$|m_\ell^{(n)}(t_2) - m_\ell^{(n)}(t_1)| = \left|\frac{dm_\ell^{(n)}(\xi)}{dt}\right| |t_2 - t_1|$$

$$\text{for } t_1 < \xi < t_2$$

but

$$\left|\frac{dm_\ell^{(n)}(\xi)}{dt}\right| = \left|\int_0^\infty a(x,\xi)\ \frac{\partial f_\ell}{\partial x} + \frac{1}{2}\, b^2(x,\xi)\ \frac{\partial^2 f_\ell}{\partial x^2}\, d\mu_n\right|$$

$$\leq \int_0^\infty |a(x,\xi)|\ \frac{\partial f_\ell}{\partial x}\, dx + \frac{1}{2}\int_0^\infty |b^2(x,\xi)|\ \frac{\partial^2 f_\ell}{\partial x^2}\, d\mu_n$$

$$\leq \int_0^\infty (a_1 x + b_1)\ \frac{\partial f_\ell}{\partial x} + \frac{1}{2}\int_0^\infty (a_2 x + b_2)^2\ \frac{\partial^2 f_\ell}{\partial x^2}\, d\mu_n$$

$$= g_\ell(m_{\ell-2}^{(n)}(\xi),\ m_{\ell-1}^{(n)}(\xi),\ m_\ell^{(n)}(\xi))\ .$$

Since g_ℓ is a polynomial with constant coefficients and $m_{\ell-2}^{(n)}(t)$, $m_{\ell-1}^{(n)}(t)$ and $m_\ell^{(n)}(t)$ are all uniformly bounded on t, there exists an M_ℓ such that

$$\left|\frac{dm_\ell^{(n)}(\xi)}{dt}\right| \leq |g_\ell(m_{\ell-2}^{(n)}(\xi),\ m_{\ell-1}^{(n)}(\xi),\ m_\ell^{(n)}(\xi)| \leq M_\ell\ .$$

Thus for fixed ℓ and all $n > \frac{1}{2}(\ell+1)$ there exists an M_ℓ such that

$$|m_\ell^{(n)}(t_1) - m_\ell^{(n)}(t_2)| \leq M_\ell |t_2 - t_1|$$

which implies the equicontinuity of $\{m_\ell^{(n)}(t),\ n > \frac{1}{2}(\ell+1),\ 0 \leq t \leq T\}$ and the proof is complete.

Lemma 3.5.2. Assume that the condition A holds. Then there exists a subsequence $\{\mu_{k_n}(t)\}$ of the Gauss-Galerkin measures $\{\mu_n(t)\}$ and a sequence of functions $\{m_\ell^*(t)\}$ such that for any positive integer ℓ, we have

$$\lim_{k_n \to +\infty} m_\ell^{(k_n)}(t) = m_\ell^*(t)$$

uniformly on $t \in [0,T]$, where $m_\ell^{(k_n)}(t) = \int_0^\infty x^\ell d\mu_{k_n}(t)$.

Proof. By Lemma 3.5.1 the set $\{m_1^{(n)}(t),\ n > 1,\ t \in [0,T]\}$ is bounded and equicontinuous; therefore, by the Ascoli theorem there exists a subsequence $\{k_{1,n}\}$ contained in $\{n\}$ such that $\{m_1^{(k_{1,n})}(t)\}$ converges uniformly to a limit function $m_1^*(t)$. We can assume $k_{1,n} > 3/2$. The set $\{m_2^{(k_{1,n})}(t),\ k_{1,n} > 3/2,\ t \in [0,T]\}$ is again bounded and equicontinuous; therefore, there exists a subsequence $\{k_{2,n}\}$ contained in $\{k_{1,n}\}$ such that $\{m_2^{(k_{2,n})}(t)\}$ converges uniformly to a function $m_2^*(t)$. Continuing this process we get a subsequence $\{k_{\ell,n}\}$ such that (for any positive integer ℓ) $\{m_\ell^{(k_{\ell,n})}\}$ converges uniformly to a function $m_\ell^*(t)$. Now consider the subsequence $\{k_{n,n}\}$ which we rename it $\{k_n\}$ that is contained in $\{n\}$. It is obvious that $\{m_\ell^{(k_n)}(t)\}$ converges uniformly to $m_\ell^*(t)$ and the proof is complete.

Remark 3.5.3. In Lemma 3.5.2 if we start with any subsequence $\{k_n\} \subset \{n\}$, in the same way we can show the existence of a further subsequence $\{s_{k_n}\} \subset \{k_n\}$ and a sequence of functions $\{m_\ell^*(t)\}$ such that

$$\lim_{s_{k_n} \to \infty} m_\ell^{(k_n)}(t) = m_\ell^*(t)$$

uniformly in $t \in [0,T]$.

Lemma 3.5.4. For any $t \in [0,T]$, the elements of the sequence $\{m_\ell^*(t)\}$ are indeed moments of a measure, i.e. there exists a measure $P^*(t)$ such that for any ℓ

$$\int_0^\infty x^\ell dP^*(t) = m_\ell^*(t) \ .$$

Proof. Assume $\{\mu_{k_n'}\}$ is a subsequence of $\{\mu_n\}$ such that for any positive integer ℓ, we have

$$\lim_{k_n \to \infty} m_\ell^{(k_n')}(t) = m_\ell^*(t) \ .$$

By the first theorem of Helly there exists a subsequence of $\{k_n'\}$ which we rename $\{k_n\}$ and a measure $P^*(t)$ such that $\{\mu_{k_n}(t)\}$ converges to $P^*(t)$ substantially. Now we claim that for any ℓ and any $\epsilon > 0$, there exists an interval $(0,b)$ such that $\int_0^b x^\ell d\mu_{k_n} > m_\ell^{(k_n)}(t) - \epsilon$ for all n, i.e.

$$\int_b^\infty x^\ell d\mu_{k_n} < \epsilon \quad \text{for all} \quad n \ .$$

To prove the above claim, let us assume the contrary. Therefore, there exists an $\epsilon > 0$ and an ℓ such that for any b there exists an $k_n^* \in \{k_n\}$ such that

$$\int_b^\infty x^\ell d\mu_{k_n}^* \geq \epsilon \ .$$

Now consider

$$m_{\ell+1}^{(k_n^*)}(t) = \int_0^\infty x^{\ell+1} d\mu_{k_n}^*(t) \geq \int_b^\infty x x^\ell d\mu_{k_n}^*(t)$$

$$\geq b \int_b^\infty d\mu_{k_n}^*(t) \geq \epsilon \cdot b \ .$$

Thus we have shown that for any M, there exists a k_n^* such that $m_{\ell+1}^{(k_n^*)}(t) \geq M$ or

$$\lim_{k_n \to \infty} M_{\ell+1}^{(k_n)} = \infty = M_{\ell+1}^*(t)$$

which is a contradiction. Thus by second theorem of Helly we have

$$\operatorname{Lim}_{k_n \to \infty} \int_0^\infty x^\ell d\mu_{k_n}(t) = \int_0^\infty x^\ell dP^*(t) = m_\ell^*(t)$$

and the proof is complete.

Note in Lemma 3.5.4 we have shown only the existence of a measure $P^*(t)$ which need not be unique. The following Lemma shows that if the measures $P^*(t)$ are uniquely determined by the moments $m_\ell^*(t)$, then the Gauss-Galerkin integration formulas converge to the true values of integrals.

Lemma 3.5.5. Under condition A if for each t the measure $P^*(t)$ is uniquely determined by $\{m_\ell^*(t)\}$, then we have

$$m_\ell^*(t) - m_\ell^*(0) = \int_0^t \int_0^\infty \ell x^{\ell-1} a(x,s)\, dP^*(s)\, ds + \frac{1}{2} \int_0^t \int_0^\infty \ell(\ell-1) x^{\ell-2} b^2(x,s)\, dP^*(s)\, ds .$$

Proof. Let $\{k_n\}$ be a subsequence of positive integers such that for any ℓ, $\lim_{k_n \to \infty} m_\ell^{(k_n)} = m_\ell^*(t)$ uniformly in t. By the assumption that for any t, $\{m_\ell^*(t)\}$ determines the measure $P^*(t)$ uniquely, the subsequence $\{\mu_{k_n}(t)\}$ converges to $P^*(t)$ substantially; otherwise, by the first theorem of Helly for a fixed t there exists a subsequence $\{\mu_{s_{k_n}}\}$ which converges to a measure $P_1^*(t)$ and $P_1^*(t)$ is not $P^*(t)$ substantially. In the same way as we did in Lemma 3.5.4 we can show that

$$\lim m_\ell^{(s_{n_k})}(t) = m_\ell^*(t) = \int_0^\infty x^\ell dP_1^*(t) = \int_0^\infty x^\ell dP^*(t) ,$$

and this is not possible. Also in the same way as we proved in Lemma 3.5.4, we can show that for any t and and any $\epsilon > 0$ there exists $b > 0$ such that $\int_b^\infty f(x)\, d\mu_{k_n}(t)$ is less than ϵ for all k_n, where $f(x)$ is a polynomial with non-negative coefficients;

$$\lim_{k_n \to \infty} \int_0^\infty xf(x)\, d\mu_{k_n} = \infty$$

which is a contradiction. Now let $\varphi(x)$ be any continuous function which is dominated by a polynomial $f(x)$ with non-negative coefficients. For any $\epsilon > 0$, there exists $b > 0$ such that

$$\left| \int_b^\infty \varphi(x)\, d\mu_{k_n}(t) \right| \leq \int_b^\infty |\varphi(x)|\, d\mu_{k_n}(t)$$

$$\leq \int_b^\infty f(x)\, d\mu_{k_n}(t) < \epsilon$$

for all k_n. Thus by the second theorem of Helly, we have

$$\lim_{k_n \to \infty} \int_0^\infty \varphi(x)\, d\mu_{k_n}(t) = \int_0^\infty \varphi(x)\, dP^*(t) \quad . \tag{3.5.9}$$

Recall that

$$\frac{dm_\ell^{(k_n)}(t)}{dt} = \int_0^\infty \ell x^\ell a(x,t)\, d\mu_{k_n}(t)$$

$$+ \int_0^\infty \frac{1}{2}\, \ell(\ell-1) x^{\ell-2} b^2(x,t)\, d\mu_{k_n}(t)$$

$$\leq \ell \int_0^\infty (a_1 x + b_1) x^{\ell-1} d\mu_{k_n}(t)$$

$$+ \frac{1}{2}\, \ell(\ell-1) \int_0^\infty (a_2 x + b_2)^2 x^{\ell-2} d\mu_{k_n}(t)$$

$$= g_\ell(m_{\ell-2}^{(k_n)}(t),\ m_{\ell-1}^{(k_n)}(t),\ m_\ell^{(k_n)}(t)) \quad .$$

We know that g_ℓ is uniformly bounded. By the continuity of $\frac{dm_\ell^{(k_n)}(t)}{dt}$ we have

$$m_\ell^{(k_n)}(t) - m_\ell^{(k_n)}(0) = \int_0^t \left(\int_0^\infty \ell x^{\ell-1} a(x,s)\, d\mu_{k_n}(s) + \int_0^\infty \frac{1}{2}\, \ell(\ell-1)x^{\ell-2} b^2(x,s)\, d\mu_{k_n}^{(s)}\right) ds$$

$$\leq \int_0^t g_\ell(m_{\ell-2}^{(k_n)}(s), \ldots, m_\ell^{(k_n)}(s))\, ds \ .$$

Thus by the Dominated Convergence Theorem and (3.5.9) we have

$$m_\ell^*(t) - m_\ell^*(0)$$

$$= \int_0^t \left(\int_0^\infty \ell x^{\ell-1} a(x,s)\, dP^*(s) + \int_0^\infty \frac{1}{2}\, \ell(\ell-1)x^{\ell-2} b^2(x,s)\, dP^*(s)\right) ds \ .$$

3.6 Main Convergence Theorems: As we have mentioned before all the Lemmas in section 3.5 are true for finite interval. Thus we can prove the following theorem.

Theorem 3.6.1. Under condition A if $(r_1, r_2) \subset (0, \infty)$ is a finite interval, then the Gauss-Galerkin integral formulas $\Sigma_{n,t}(f)$ converge to $I_t(f)$ where $f(x)$ is any continuous function defined on $[r_1, r_2]$.

Proof. To prove the theorem, first we show that for any positive integer ℓ and $0 \leq t \leq T$, the sequence

$\{m_\ell^{(n)}(t)\}$ converges to the exact moment $\mathcal{E}(X^\ell(t))$. If we consider the sequence $\{m_\ell^{(n)}(t)\}$ as a sequence in the complete metric space $C([0,T])$, it suffices to prove that any subsequence of $\{m_\ell^{(n)}(t)\}$ has a further subsequence which converges to the exact moment uniformly. Let $\{k_n^*\}$ be a subsequence of positive integers. By Lemma 3.5.2 there exists a subsequence $\{k_n\} \subset \{k_n^*\}$ and a sequence $\{m_\ell^*(t)\}$ such that for any ℓ, $\lim_{k_n \to \infty} m_\ell^{(k_n)}(t) = m_\ell^*(t)$ uniformly in t. By Lemma 3.5.4, $m_\ell^*(t)$, $\ell = 0,1,2,\ldots$ are the moments of a measure $P^*(t)$. Now since $[r_1,r_2]$ is finite, the coefficients a and b in (3.5.1) are bounded and therefore, the equation (3.5.6) reduces to

(3.6.1) $dX(t) = a_1 dt + b_1 dW(t)$, a_1 and b_1 constants

(3.6.2) $X(0) = \text{given}$.

Since the density of the solution to the above stochastic differential equation is normally distributed (Arnold [1], p. 133), the moments of the solution satisfy

$$m_\ell(t) \leq C(2\ell+1)!R^\ell$$

for some constants C and R independent of ℓ that may depend on t. Thus for any ℓ and n we have

$$m_\ell^{(n)}(t) \le m_\ell(t) \le C(2\ell+1)!R^\ell$$

and $\{m_\ell^*(t)\}$ determines the measure $P^*(t)$ uniquely. Thus by Lemma 3.5.5 for any ℓ we have

(3.6.3) $m_\ell^*(t) - m_\ell^*(0)$

$$= \int_0^t \int_0^\infty \ell x^{\ell-1} a(x,s)\, dP^*(s)\, ds + \frac{1}{2}\int_0^t \int_0^\infty \ell(\ell-1)x^{\ell-2} b^2(x)\, dP^*(t))\, ds$$

$$= \int_0^t (P^*(s), Lx^\ell)\, ds \ .$$

On the other hand $P(t) = p(t,x)\,dx$, the exact measure, satisfies the same equation as above. Thus we have

$$(P(t),f) - (P(0),f) = \int_0^t (P(s),Lf)\, ds$$

for all polynomials f.

Now let $\varphi(x)$ be any C_2-function on $[r_1,r_2]$. By the Weierstrass approximation theorem and the finiteness of the interval $[r_1,r_2]$, for any $\varepsilon > 0$ there exists a polynomial $P_m(x)$ such that

$$\|\varphi(x) - P_m(x)\| + \|\varphi'(x) - P_m'(x)\| + \|\varphi''(x) - P_m''(x)\| < \varepsilon$$

for all $x \in [r_1,r_2]$. Now we wish to prove that

(3.6.4) $$(P^*(t),\varphi) - (P^*(0),\varphi) = \int_0^t (P^*(s),L\varphi)\, ds \ .$$

To do this, note that by (3.6.3)

$$\begin{aligned}
&|(P^*(t),\varphi) - (P^*(0),\varphi) - \int_0^t (P^*(s),L\varphi)\,ds| \\
&= |(P^*(t),\varphi - P_m + P_m) - (P^*(0),\varphi + P_m - P_m) \\
&\qquad - \int_0^t (P^*(s),L(\varphi - P_m) + LP_m)\,ds| \\
&\leq |(P^*(t),\varphi - P_m) - (P^*(0),\varphi - P_m) - \int_0^t (P^*(s),L(\varphi - P_m)\,ds| \\
&\leq |(P^*(t),\varphi - P_m)| + |(P^*(0),\varphi - P_m)| \\
&\qquad + \int_0^t |(P^*(s),L(\varphi - P_m))ds| < C\varepsilon
\end{aligned}$$

where C is a constant depending on the bounds of the coefficient a and b. Thus equation (3.6.4) holds for any C_2-function $\varphi(x)$ defined on $[r_1,r_2]$.

Now let $u_\lambda(x)$ be an eigenfunction for the eigenvalue problem

$$\lambda u = Lu \ . \tag{3.6.5}$$

Equation (3.6.4) for u_λ becomes

$$\begin{aligned}
&(P^*(t),u_\lambda) - (P^*(0),u_\lambda) \\
&\quad = \int_0^t (P^*(s),Lu_\lambda)\,ds = \lambda \int_0^t (P^*(s),u_\lambda)\,ds \ .
\end{aligned}$$

Thus $(P^*(t),u_\lambda) = c_\lambda e^{\lambda t}$. Obviously the equation (3.6.4) is true for $P(t)$ and in the same way as we showed for p^* we

can prove $(P(t),u_\lambda) = C'_\lambda e^{\lambda t}$. Since $P^*(0) = P(0)$, we have proved

$$(3.6.6) \qquad (P^*(t),u_\lambda) = (P(t),u_\lambda) \ .$$

Thus by condition A5, (3.6.6) holds for any $\varphi(x)$ sufficiently smooth; therefore we have

$$P^*(t) = P(t) \quad \text{and} \quad m^*_\ell(t) = \mathcal{E}(X^\ell(t)) \ ,$$

for any positive integer ℓ. Now this and equation (3.5.9) complete the proof of the theorem.

In the remainder of this section we shall discuss the convergence of the Gauss-Galerkin method in the $(0,\infty)$ case. First let us consider the special stochastic differential equation

$$(3.6.7) \qquad dX(t) = X(t)\,dt + X(t)\,d\bar{W}(t)$$

$$(3.6.8) \qquad X(0) = \text{given} \ .$$

The Fokker-Planck equation for the above equation is

$$\frac{\partial p}{\partial t} = -\frac{\partial (xp)}{\partial x} + \frac{1}{2}\frac{\partial^2 (x^2 p)}{\partial x^2}$$

$$p(0,x) = \text{given} \ .$$

The Hankel system for the moments of $p(t,x)$ is

$$\frac{dm_n(t)}{dt} = \frac{1}{2}\,(n+n^2)\,m_n(t) \qquad n = 0,1,\ldots$$

with the solution

$$m_n(t) = m_n(0)\, e^{\frac{1}{2}(n+n^2)t} \qquad n = 0,1,\ldots$$

As these moments do not satisfy the sufficient conditions (3.1.5); $p(t,x)dx$ may not be uniquely determined by these moments. Also it is obvious that the Gauss-Galerkin solutions of (3.6.7) and (3.6.8) are the same as the Gauss-Christoffel ones. Thus the sequence of functions $\{m_\ell^*(t)\}$ is the sequence of the exact moments. Thus, in general even if a linear stochastic differential equation satisfies condition A, Lemma 3.5.5 may not be true and therefore we do not have a convergence theorem similar to the Theorem 3.6.1. However, we have convergence for a more restrictive class of stochastic differential equations defined on $(0,\infty)$. First consider the stochastic differential equations

$$(3.6.9) \qquad dX(t) = (bX(t) + a)dt + \sqrt{2aX(t)}\, dW(t)$$

$$(3.6.9)' \qquad X(0) = \text{given where } p(0,x) = \beta e^{-\beta x}, \quad \beta > 0$$

for $a \geq 0$ defined on $(0,\infty) \times [0,T]$. The Fokker-Planck equation of (3.6.9) is

$$\frac{\partial p}{\partial t} = -\frac{\partial (bx+a)p}{\partial x} + \frac{\partial^2 axp}{\partial x^2} .$$

$$= -bp + (a - bx)\frac{\partial p}{\partial x} + ax\frac{\partial^2 p}{\partial x^2} .$$

By the separation of variables, we seek solutions to above partial differential equation in the form $e^{-\lambda t}v(x)$. Obviously $v(x)$ satisfies the differential equation.

$$xv''(x) + (1 - \frac{b}{a}x)v'(x) - \frac{b}{a}v(x) + \frac{\lambda}{a}v(x) = 0 .$$

If $b > 0$ substituting $u(\xi) = v(x)$ where $x = \frac{a}{b}\xi$ we have

$$(3.6.13) \quad \xi u''(\xi) + (1 - \xi)u'(\xi) + (\frac{\lambda}{b} - 1)u(\xi) = 0 .$$

It is well known (see [16], p. 243) that a necessary and sufficient condition for the differential equation

$$xy'' + (1 - x)y' + \mu y = 0$$

to have a polynomial solution is that $\mu = n$. Furthermore, $L_n(x)$, the Laguerre polynomial of degree n is the only solution. Now if we assume $\frac{\lambda_n}{b} - 1 = n$, i.e. $\lambda_n = b(n+1)$ we may write

$$(3.6.11) \quad p(t,x) = \sum_{n=0}^{\infty} a_n e^{-\lambda_n t} L_n(\frac{b}{a}x)$$

as a solution for the Fokker-Planck equation. With $p(0,x) = \beta e^{-\beta x}$ and since ([18], p. 135 and 205))

$$(3.6.12) \quad \int_0^{\infty} e^{-\xi} L_n(\xi) L_m(\xi) d\xi = 0 \quad m \neq n ,$$

$$(3.6.13) \quad \int_0^{\infty} e^{-\xi} [L_n(\xi)]^2 d\xi = 1$$

(3.6.14) $\frac{1}{1-z}\exp(-xz/1-z) = \sum z^n L_n(z)$, $|z| < 1$, and

(3.6.15) $$\int_0^\infty e^{-z\xi} L_n(\xi)\,d\xi = z^{-n-1}(z-1)^n ,$$

we obtain from (3.6.11)

$$\int_0^\infty \beta e^{-\beta x} e^{-bx/a} L_m(\tfrac{b}{a}x)\,d(\tfrac{b}{a}x)$$

$$= \sum_{n=0}^{\infty} e^{-bx/a} L_n(\tfrac{b}{a}x) L_m(\tfrac{b}{a}x)\,d(\tfrac{b}{a}x)$$

$$= a_m \int_0^\infty e^{-bx/a} [L_m(\tfrac{b}{a}x)]^2\, d(\tfrac{b}{a}x) = a_m .$$

Thus by (3.6.13) we have

$$a_n = \beta(1+\tfrac{a\beta}{b})^{-n-1}(\tfrac{a\beta}{b})^n = \beta(1+\tfrac{a\beta}{b})^{-1}\ (\tfrac{a\beta/b}{1+a\beta/a})^n$$

and

(3.6.16) $$p(t,x) = \frac{\beta e^{-bt}}{1+\frac{a\beta}{b}} \sum_{n=0}^{\infty} (\frac{a\beta e^{-bt}}{b+a\beta})^n L_n(\tfrac{b}{a}x)$$

$$= \frac{\beta b e^{-bt}}{b+a\beta-a\beta e^{-bt}} \exp(-b\beta e^{-bt}x/(b+a\beta-a\beta e^{-bt}))$$

In the case when $b < 0$ equation (3.6.10) becomes

$$\xi u'' + (1+\xi)u' + (1+\frac{\lambda}{|b|})u = 0 .$$

In this case for $1+\frac{\lambda}{|b|} = 1+n$ the solution of the above equation is of the form $u(\xi) = e^{-\xi} L_n(\xi)$. Thus equation (3.6.11) becomes

$$p(t,x) = \sum_{n=0}^{\infty} a_n e^{-n|b|t} e^{-\frac{|b|}{a}x} L_n(\tfrac{|b|}{a}x) .$$

With $p(0,x) = \beta e^{-\beta x}$ using equations (3.6.12) - (3.6.15) we get $a_n = \beta(\frac{\beta a}{|b|})^{-n-1}(\frac{\beta a}{|b|}-1)^n$. Thus formally

$$p(t,x) = \sum_{n=0}^{\infty} \frac{|b|}{a}\left(\frac{\frac{\beta a}{|b|}-1}{\frac{\beta a}{|b|}}\right)^n e^{-|b|nt} e^{-\frac{|b|}{a}x} L_n(\tfrac{|b|}{a}x)$$

$$= \frac{|b|}{a} e^{-\frac{|b|}{a}x} \sum_{n=0}^{\infty} (\frac{\beta a - |b|}{\beta a} e^{-|b|t})^n L_n(\tfrac{|b|}{a}x) .$$

Note that the series above converges for $\beta > |b|/2a$ to the limit

$$p(t,x) = \frac{\beta b e^{-bt}}{b + a\beta - a\beta e^{-bt}} \exp \frac{-b\beta e^{-bt}x}{b + a\beta - a\beta e^{-bt}}$$

which is the same as equation (3.6.16). However the limit is defined for all β without any restriction. Thus for any b, $p(t,x)$ is exponential. In particular there exists γ such that

$$(3.6.17) \quad \int_0^{\infty} e^{\gamma x} p(t,x)\,dx + \int_0^{\infty} x e^{\gamma x} p(t,x)\,dx + \int_0^{\infty} x^2 e^{\gamma x} p(t,x)\,dx \le M < \infty$$

which is needed later. It is interesting to see that the stochastic differential equation (3.6.9), (3.6.10) satisfies Condition A. It is easy to verify Conditions

A1-A3. By integration by parts we can verify that A4 is valid. Also in this special case $p(t,0)$ need not be zero. The eigenvalue problem of condition A5 for equation (3.6.9) becomes

$$(bx + a)\frac{\partial u}{\partial x} + ax\frac{\partial^2 u}{\partial x^2} = \lambda u \ .$$

As we saw above, depending on whether b is negative or positive, the eigenfunctions are $L_n(x)$ or $e^{-x}L_n(x)$ where $L_n(x)$ is the Laguerre polynomial of degree n and they satisfy the requirements of the Condition A5. Finally we have derived the density explicitly (Equation (3.6.16)). We refer to (Feller [8,a], p. 516) for the uniqueness of the solution and the validity of Condition A6 holds.

Now consider any stochastic differential equation of the form

$$(3.6.18) \qquad dX(t) = a(X(t)) + b(X(t))dW(t)$$

$$(3.6.19) \qquad X(0) = \text{given}$$

which satisfies Condition A on $(0,\infty) \times [0,T]$. In addition we assume

$$(3.6.20) \qquad b(x) \leq \sqrt{2\sigma x} \quad (\sigma \text{ a positive constant})$$

and $X(0)$ has a moment generating function, i.e.

$$M = \sum_{n=0}^{\infty} \frac{\theta^n m_n(0)}{n!} < \infty \quad \text{for some} \quad \theta > 0 \ .$$

There exists a $\beta_1 > 0$ such that $\frac{1}{\beta_1} > (1+M)\theta^{-1}$ or

$$\int_0^\infty \beta_1 x^n e^{-\beta_1 x} dx = n!\left(\frac{1}{\beta_1}\right)^n \geq M\theta^{-n} n! \geq m_n(0) \ .$$

In the same way as in Lemma 3.5.1 we can show that for any n, the n-th moment of the solution to the equation (3.6.18) is less than or equal to the n-th moment of the solution of an stochastic differential equation of type $dX(t) = (bX+\sigma)dt + \sqrt{2\sigma X}\, dW(t)$ with the initial density $\beta e^{-\beta x}$ whose n-th moment is greater than or equal to $m_n(0)$ and therefore the functions $\{m_\ell^*(t)\}$ determine a measure uniquely. Thus all the Lemmas of Section 5 are valid. Let θ_1 be

$$\theta_1 = \min_{0\leq t\leq T} \frac{\beta b e^{-bt}}{b+\sigma\beta-\sigma\beta e^{-bt}}$$

We are now ready to prove the following theorem.

Theorem 3.6.2. Assume that the stochastic differential equation (3.6.18) and the initial value (3.6.19) satisfy Condition A together with (3.6.20). If θ_0, the constant in A5, is less than or equal to θ_1, then the Gauss-Galerkin integral formulas converge to the true value of integral, i.e.

$$\lim_{n\to\infty} \Sigma_{n,t}(f) = I_t(f)$$

where $f(x)$ is any continuous function dominated by a polynomial.

Proof: The proof is similar to that of the Theorem 3.6.1. Let $\{k_n^*\}$ be a subsequence of positive integers. By Lemma 3.5.2 there exists a subsequence $\{k_n\} \subset \{k_n^*\}$ and a sequence $\{m_\ell^*(t)\}$ such that for any ℓ, $\lim_{k_n \to \infty} m_\ell^{(k_n)}(t) = m_\ell^*(t)$ uniformly in t. Similar to what we did in Theorem 3.6.1 we can show that

$$m_\ell^*(t) - m_\ell^*(0) = \int_0^t \int_0^\infty x^{\ell-1} a(x)\, dP^*(s)\, ds$$

$$+ \frac{1}{2} \int_0^t \int_0^\infty \ell(\ell-1) x^{\ell-2} b^2(x)\, dP^*(s)\, ds$$

$$= \int_0^t (P^*(s), Lx^\ell)\, ds$$

On the other hand $P(t) = p(t,x)\,dx$, the exact measure, satisfies the same equation as above. Thus we have

$$(P^*(t), f) - (P^*(0), f) = \int_0^t (P^*(s), Lf)\, ds$$

and

$$(P(t), f) - (P(0), f) = \int_0^t (P(s), Lf)\, ds$$

for all polynomials f. Now let $\varphi(x)$ be a C_2-function defined on $[0,\infty)$ such that $\lim_{x \to \infty} \varphi''(x) e^{-\theta_0 x} = 0$. By the generalized Weierstrass approximation theorem (see Buck [2], p. 74), for any $\varepsilon > 0$ there exists a polynomial $P_m(x)$ such that

$$|\varphi(x) - P_m(x)| + |\varphi'(x) - P_m'(x)| + |\varphi''(x) - P_m''(x)| < \varepsilon e^{\theta_0 x}$$

for all $x \in [0,\infty)$. Now we wish to show that

$$(3.6.21)\quad (P^*(t),\varphi) - (P^*(0),\varphi) = \int_0^t (P^*(s),\varphi)ds \ .$$

To do this note that

$$\begin{aligned}
&\left|(P^*(t),\varphi) - (P^*(0),\varphi) - \int_0^t (P^*(s),L\varphi)ds\right| \\
&= \left|(P^*(t),\varphi - P_m + P_m) - (P^*(0),\varphi - P_m + P_m)\right. \\
&\qquad \left. - \int_0^t (P^*(s),L(\varphi - P_m) + LP_m)ds\right| \\
&\le \left|(P^*(t),\varphi - P_m\right| - (P^*(0),\varphi - P_m) \\
&\qquad \left. - \int_0^t (P^*(s),L(\varphi - P_m))\right| \\
&\le \left|(P^*(t),\varphi - P_m)\right| + \left|P^*(0),\varphi - P_m)\right| \\
&\qquad + \int_0^t \left|(P^*(s),L(\varphi - P_m))\right|ds \\
&< C\varepsilon
\end{aligned}$$

where C is a constant depending on the bounds of

$$\int_0^\infty e^{\theta_0 x} dP^*(t), \quad \int_0^\infty xe^{\theta_0 x} dP^*(t)$$

$$\text{and} \quad \int_0^\infty x^2 e^{\theta_0 x} dP^*(t)$$

which is independent of t because of (3.6.17). Thus equation (3.6.4) holds for any C_2-function $\varphi(x)$ such that $\varphi''(x)e^{-\theta_0 x} \to 0$ as $x \to \infty$. We now let u_n be an eigenfunction for the eigenvalue problem

(3.6.22) $$\lambda u = Lu \ .$$

Since u_n is a C_∞-function and by Condition A5 $\lim_{x\to\infty} u_n'' e^{-\theta_0 x} = 0$, equation (3.6.21) for u_n becomes

$$(P^*(t),u_n) - (P^*(0),u_n) = \int_0^t (P^*(s),Lu_n)ds$$

$$= \lambda \int_0^t (P^*(s),u_n)ds$$

Thus in the same as in Theorem (3.6.1) we have

(3.6.23) $$(P^*(t),u_n) = (P(t),u_n)$$

Now let $\psi(x)$ be any C_∞ function with compact support contained in $(0,\infty)$. By Condition A5 for any $\varepsilon > 0$ there exists a finite sum of the form $\sum_{n=1}^{N} c_n u_n e^{-\theta_0 x}$ such that

$$|\psi - \sum_{n=1}^{N} c_n u_n e^{-\theta_0 x}| < \varepsilon \ .$$

Then for $\varphi(x) = \psi(x)e^{\theta_0 x}$ we have

$$|\varphi - \sum_{n=1}^{N} c_n u_n| < \varepsilon e^{\theta_0 x} \ .$$

Thus

$$(\varphi,P^*(t) - P(t)) = (\varphi - \sum_{n=1}^{N} c_n u_n, P^* - P) + \sum_{n=1}^{N} (c_n u_n, P^* - P)$$

$$= (\varphi - \sum_{n=1}^{N} c_n u_n, P^* - P)$$

implies that

$$|(\varphi, P^*(t) - P(t))| < \varepsilon\left(\int_0^\infty e^{\theta_0 x} dP^*(t) + \int_0^\infty e^{\theta_0 x} dP(t)\right)$$

$$< C\varepsilon$$

Thus (3.6.23) holds for any $\varphi(x)$ with compact support; therefore,

$$P^*(t) = P(t), \quad \text{and} \quad m_\ell^*(t) = \beta(x^\ell(t)) ,$$

for any positive integer ℓ. Now this and equation (3.5.9) completes the proof of the theorem.

CHAPTER IV

NUMERICAL EXAMPLES

We present in this chapter several numerical examples that serve to illustrate the Gauss-Galerkin method developed in the preceding chapters. We note that the Kolmogorov equation corresponding to nonlinear stochastic equations has been solved explicitly only in a few simple cases. We have thus included examples for some simple stochastic equations where the exact solution are known so that the numerical results may be compared with the exact ones. We have also included examples for problems which do not satisfy the hypotheses in the convergence theorem but whose Gauss-Galerkin approximation seem to be accurate nonetheless.

The Gauss-Galerkin solutions are obtained by using the well-known programs (e.g the International Mathematical Statistical Library) to compute the initial Gauss-Christoffel weights and nodes and then using standard ordinary differential equation solvers (Dgear).

4.1 Example 1: Consider the stochastic differential equation

$$(4.1.1) \qquad dX(t) = (1 - X(t))\,dt + \sqrt{2X(t)}\,dW(t)$$

$$(4.1.2) \qquad p(0,x) = 2\exp(-2x)$$

studied in Chapter 3, by equation (3.6.16) $p(t,x)$, the exact density, is given by

$$(4.1.3) \qquad p(t,x) = \frac{2e^t}{2e^t - 1} \exp\left(\frac{-2e^t x}{2e^t - 1}\right)$$

and the exact n-th moment is given by

$$(4.1.4) \qquad M_n(t) = n!\left(1 - \frac{1}{2}e^{-t}\right)^n .$$

From (4.1.3), the exact 5-point nodes and weights are computed and shown in Tables 4.1.5 and 4.1.6. The numerical solutions for the Gauss-Galerkin 5-point nodes and weights are given in Tables 4.1.7 and 4.1.8. Using 5-point Gauss-Galerkin nodes and weights the first five moments are computed. Tables 4.1.9 and 4.1.10 show the exact and above computed moments.

Table 4.1.5

The exact 5-point nodes for Example 1

t	x_1	x_2	x_3	x_4	x_5
0	.13178	.70670	1.7982	3.5429	6.3204
.5	.18363	.98477	2.5058	4.9369	8.8073
1.0	.21508	1.1534	2.9349	5.7824	10.316
1.5	.23416	1.2557	3.1952	6.2953	11.231
2.0	.24573	1.3178	3.3531	6.6063	11.785
2.5	.25274	1.3554	3.4489	6.7950	12.122
3.0	.25699	1.3782	3.5069	6.9094	12.326
5.0	.26267	1.4086	3.5843	7.0619	12.598
7.0	.26344	1.4127	3.5948	7.0825	12.635
9.0	.26354	1.4133	3.5962	7.0853	12.640

Table 4.1.6

The exact 5-point weights for Example 1

t	a_1	a_2	a_3	a_4	a_5
0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
.5	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
1.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
1.5	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
2.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
2.5	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
3.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
5.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
7.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4
9.0	.52176	.39867	.75942E-1	.36118E-2	.2337OE-4

Table 4.1.7

The Gauss-Galerkin 5-point nodes for Example 1

t	x_1	x_2	x_3	x_4	x_5
0	.13178.	.70670	1.7982	3.5429	6.3204
.5	.18366	.98490	2.5062	4.9380	8.8105
1.0	.21507	1.1533	2.9347	5.7821	10.315
1.5	.23410	1.2554	3.1944	6.2937	11.228
2.0	.24560	1.3171	3.3514	6.6030	11.780
2.5	.25259	1.3546	3.4467	6.7908	12.115
3.0	.25686	1.3774	3.5049	6.9056	12.319
5.0	.26258	1.4081	3.5830	7.0593	12.594
7.0	.26338	1.4124	3.5939	7.0808	12.632
9.0	.26353	1.4132	3.5959	7.0848	12.639

Table 4.1.8

The Gauss-Galerkin 5-point weights for Example 1

t	a_1	a_2	a_3	a_4	a_5
0	.52176	.39867	.75942E-1	.36118E-2	.23370E-4
.5	.52180	.39865	.75917E-1	.36082E-2	.23312E-4
1.0	.52176	.39866	.75939E-1	.36112E-2	.23360E-4
1.5	.52176	.39867	.75942E-1	.36117E-2	.23368E-4
2.0	.52176	.39867	.75942E-1	.36117E-2	.23368E-4
2.5	.52176	.39867	.75942E-1	.36117E-2	.23370E-4
3.0	.52176	.39867	.75943E-1	.36118E-2	.23370E-4
5.0	.52176	.39867	.75943E-1	.36118E-2	.23370E-4
7.0	.52176	.39867	.75943E-1	.36118E-2	.23370E-4
9.0	.52176	.39867	.75942E-1	.36118E-2	.23370E-4

Table 4.1.9

The exact moments for Example 1

t	m_0	m_1	m_2	m_3	m_4
0	1.0000	.50000	.50000	.75000	1.5000
1.0	1.0000	.81606	1.3319	3.2608	10.644
2.0	1.0000	.93233	1.7385	4.8625	18.134
3.0	1.0000	.97511	1.9017	5.5630	21.698
4.0	1.0000	.99084	1.9635	5.8367	23.133
5.0	1.0000	.99663	1.9865	5.9396	23.678
6.0	1.0000	.99876	1.9950	5.9777	23.881
7.0	1.0000	.99954	1.9982	5.9918	23.956
8.0	1.0000	.99983	1.9993	5.9970	23.984
9.0	1.0000	.99994	1.9998	5.9989	23.994

Table 4.1.10

The Gauss-Galerkin moments for Example 1

t	m_0	m_1	m_2	m_3	m_4
0	1.0000	.50000	.50000	.75000	1.5000
1.0	1.0000	.81599	1.3317	3.2599	10.640
2.0	1.0000	.93186	1.7367	4.8551	18.097
3.0	1.0000	.97456	1.8995	5.5537	21.650
4.0	1.0000	.99029	1.9613	5.8269	23.081
5.0	1.0000	.99626	1.9851	5.9333	23.643
6.0	1.0000	.99848	1.9939	5.9727	23.854
7.0	1.0000	.99930	1.9972	5.9874	23.933
8.0	1.0000	.99967	1.9987	5.9940	23.968
9.0	1.0000	.99986	1.9994	5.9974	23.986

The agreement above appear to be excellent. We also note the rapid convergence of the above solutions to their steady-state values.

4.2 Example 2: Consider the stochastic differential equation

$$dX(t) = X(t)(1 - X(t)dW(t) \tag{4.2.1}$$

$$X(0) = \text{given} \tag{4.2.2}$$

defined on $(0,1) \times [0,T]$ where the initial process $X(0)$ is a random variable whose density is

$$(4.2.3) \quad p(0,x) = \begin{cases} cx \exp(-1/(1-x)^2) & \text{if } 0 < x < 1 \\ \text{where } C \text{ is normalizing constant} & \\ 0 & \text{if } x \leq 0 \text{ or } x \geq 1 . \end{cases}$$

We note that an explicit solution of above problem does not seem to be available (however the solution exists). We present the Gauss-Galerkin 5-point nodes and weights in Tables 4.2.4 and 4.2.5 below. We also present the values of the first five moments in Table 4.2.6.

Table 4.2.4

The Gauss-Galerkin 5-point nodes for Example 2

t	x_1	x_2	x_3	x_4	x_5
0	.46931E-1	.15656	.29809	.44651	.58476
1.0	.49810E-1	.20298	.42683	.67040	.87236
2.0	.38880E-1	.20435	.45225	.71633	.91882
3.0	.31840E-1	.20430	.46443	.73763	.93942
4.0	.26985E-1	.20398	.47172	.75017	.95158
5.0	.23428E-1	.20350	.47652	.75846	.95952

Table 4.2.5

The Gauss-Galerkin 5-point weights for Example 2

t	a_1	a_2	a_3	a_4	a_5
0	.90525E-1	.33506	.39231	.16635	.15759E-1
1.0	.33524	.34166	.20516	.95342E-1	.22596E-1
2.0	.43649	.25838	.15546	.10050	.49179E-1
3.0	.50310	.20095	.12300	.9616E-1	.76791E-1
4.0	.55148	.15953	.99061E-1	.87519E-1	.10241
5.0	.58793	.12871	.80720E-1	.77590E-1	.12505

Table 4.2.6

The Gauss-Galerkin 5-point moments for Example 2

t	m_0	m_1	m_2	m_3	m_4
1.0	1.0000	.25725	.11233	.62580E-1	.39736E-1
2.0	1.0000	.25726	.13633	.91698E-1	.68467E-1
3.0	1.0000	.25727	.15552	.11631	.94347E-1
4.0	1.0000	.25726	.17107	.13695	.11687
5.0	1.0000	.25727	.18375	.15415	.13606

Our convergence theorem in Chapter 3 (Theorem 3.6.1) guarantees the convergence of the numerical results here to the exact ones as n become large. Our experience suggests that with $n = 5$ here we expect the first ten moments to be good approximation to the first ten exact moments.

4.3 Example 3. Consider the stochastic differential equation

$$dX(t) = \frac{1}{2} X(t)\,dt + \frac{\sqrt{2}}{10} X(t)\,dW(t) \tag{4.3.1}$$

$$X(0) = \text{given} \tag{4.3.2}$$

defined on $(0,\infty) \times [0,T]$ where the initial process X_0 is a random variable whose corresponding density is

$$p(0,x) = \begin{cases} cx \exp(-1/(x-1)^2) & \text{if } 0 \leq x < 1 \\ \text{where } C \text{ is normalizing factor} & \\ 0 & \text{if } x \geq 1 \end{cases} \tag{4.3.3}$$

Equation (4.3.1) is a homogeneous linear stochastic differential equation. The boundaries are inaccessible (see Feller [6,a]) and the n-th moment of its solution is given by (see Arnold [1], p. 139)

$$m_n(t) = m_n(0) \exp\left(.01(n^2 + 49n)t\right) \tag{4.3.4}$$

The exact 5-point nodes and weights are given in Tables 4.3.5 and 4.3.6. Also the Gauss-Galerkin 5-point nodes and weights are given in Tables 4.3.7 and 4.3.8. Using 5-point and 6-point Gauss-Galerkin weights and nodes the first 10 moments are computed. Tables 4.3.9-4.3.11 show the exact moments and the Gauss-Galerkin moments.

Table 4.3.5

The exact 5-point nodes for Example 3

t	x_1	x_2	x_3	x_4	x_5
0	.46931E-1	.15656	.29809	.44651	.58476
1	.91699E-1	.30071	.57356	.88306	1.2263
2	.17088	.55329	1.0635	1.7039	2.6209
3	.31106	.99813	1.9483	3.2573	5.3333
4	.55735	1.7795	3.5488	6.1883	10.7165
5	.98691	3.1485	6.4420	11.703	21.343

Table 4.3.6

The exact 5-point weights for Example 3

t	a_1	a_2	a_3	a_4	a_5
0	.90525E-1	.33506	.39231	.16635	.15759E-1
1	.12685	.42126	.36904	.80708E-1	.21504E-2
2	.16357	.48110	.31398	.40818E-1	.52822E-3
3	.20200	.52022	.25577	.21849E-1	.16605E-3
4	.24190	.54137	.20441	.12249E-1	.59992E-4
5	.28273	.54801	.16212	.71172E-2	.23716E-4

Table 4.3.7

The Gauss-Galerkin 5-point nodes for Example 3

t	x_1	x_2	x_3	x_4	x_5
0	.46931E-1	.15656	.29809	.44651	.58476
1	.91622E-1	.30122	.57380	.88338	1.2628
2	.17070	.55464	1.0648	1.7064	2.6245
3	.31090	1.0013	1.9527	3.2669	5.3508
4	.55809	1.7879	3.5645	6.2252	10.791
5	.98905	3.1648	6.4760	11.787	21.525

Table 4.3.8

The Gauss-Galerkin 5-point weights for Example 3

t	a_1	a_2	a_3	a_4	a_5
0	.90525E-1	.33506	.39231	.16635	.15759E-1
1	.12657	.41745	.36874	.80571E-1	.21670E-2
2	.16292	.47811	.31332	.40617E-1	.53997E-3
3	.20102	.51785	.25479	.21671E-1	.17067E-2
4	.24099	.53953	.20287	.12046E-1	.61669E-4
5	.28177	.54625	.16052	.69323E-1	.22346E-4

Table 4.3.9

The first 10 exact moments for Example 3

m/t	0	1	2	3	4
m_0	1.0000	1.0000	1.0000	1.0000	1.0000
m_1	.25714	.42396	.69899	1.1524	1.90000
m_2	.81828E-1	.22692	.62931	1.7452	4.8397
m_3	.29647E-1	.14109	.67140	3.1951	15.205
m_4	.11755E-1	.97930E-1	.81587	6.7971	56.628
m_5	.49851E-2	.74176E-1	1.1037	16.423	244.37
m_6	.22287E-2	.60425E-1	1.6383	44.418	1204.3
m_7	.104400E-2	.52414E-1	2.6417	133.14	6710.5
m_8	.50289E-3	.48068E-1	4.5945	439.16	41976
m_9	.25067E-3	.46357E-1	8.5731	1585.5	293210

Table 4.3.11

The first 10 moments for Example 3 using
5-point Gauss-Galerkin weights and nodes

m/t	0	1	2	3	4
m_0	1.0000	1.0000	1.0000	1.0000	1.0000
m_1	.25714	.42400	.69923	1.1533	1.9029
m_2	.81828E-1	.22698	.62986	1.7484	4.8568
m_3	.29647E-1	.14116	.67253	3.2059	15.300
m_4	.11755E-1	.98017	.81821	6.8345	57.196
m_5	.49851E-2	.74281E-1	1.1087	16.562	248.09
m_6	.22287E-2	.60552E-1	1.6495	44.970	1231.0
m_7	.104400E-2	.52568E-1	2.6676	135.49	6920.7
m_8	.50289E-3	.48256E-1	4.6562	449.73	43776
m_9	.25067E-3	.46589E-1	8.7239	1635.6	309770

Table 4.3.12

The first 10 moments for Example 3 using
6-point Gauss-Galerkin weights and nodes

m/t	0	1	2	3	4
m_0	1.0000	1.0000	1.0000	1.0000	1.0000
m_1	.25714	.42397	.69918	1.1531	1.9013
m_2	.81828E-1	.22694	.62971	17.474	4.8469
m_3	.29647E-1	.14111	.67217	3.2026	15.246
m_4	.11755E-1	.97960E-1	.81735	6.8234	56.876
m_5	.49851E-2	.74215E-1	1.1068	16.522	246.04
m_6	.22287E-2	.60475E-1	1.6449	44.822	1216.6
m_7	.104400E-2	.52479E-1	2.6568	134.91	6810.2
m_8	.50289E-3	.48152E-1	4.6308	447.45	42863
m_9	.25067E-3	.46469E-1	8.6648	1627.0	301820

Although the coefficients of this example do not satisfy the conditions of our convergence theorem, the agreement above is very good. We note also that the moments computed from the Gauss-Galerkin 6-point nodes and weights are closer to the exact moments than the moments computed from the Gauss-Galerkin 5-point nodes and weights.

4.4 Example 4. Consider the stochastic differential equation

$$dX(t) = (1 - X(t))\,dt + \sqrt[3]{2X(t)}\,dW(t) \tag{4.4.1}$$

$$X(0) = \text{given} \tag{4.4.2}$$

(A)

defined on $(0,\infty) \times [0,T]$. We note that an explicit solution of above problem does not seem available and its coefficients do not satisfy conditions of our convergence theorem. However we expect that the solution above to be close to the solution of the stochastic differential equation

$$dX(t) = (1 - X(t))\,dt + b(X(t))\,dW(t) \tag{4.4.3}$$

(B)

where $b(x) = \begin{cases} \sqrt{2x} & \text{if } x < 1 \\ \sqrt[3]{2x} & \text{if } x \geq 1 \end{cases}$ with the same initial value as (4.4.2). Equation (4.4.3) satisfies our

convergence theorem; therefore the convergence of the numberical results are guaranteed. The Gauss-Galerkin 5-point nodes and weights for the equations (4.4.1) and (4.4.3) with the initial value $p(0,x) = 2\exp(-2x)$ are computed and used to compute the first eight moments of each problem. These moments are shown in Tables 4.4.4 and 4.4.5.

Table 4.4.4

The first eight moments of Example 4(A)

m/t	1	2	4	6	8	10
m_0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m_1	.81605	.93266	.99094	.99866	1.0000	1.0000
m_2	1.2034	1.5268	1.6997	1.7230	1.7271	1.7271
m_3	2.3729	3.3101	3.8432	3.9167	3.9296	3.9296
m_4	5.7459	8.7543	10.571	10.826	10.871	10.871
m_5	16.317	27.010	33.854	34.832	35.004	35.004
m_6	52.821	94.572	122.85	126.95	127.68	127.68
m_7	191.14	368.64	495.67	514.41	517.75	517.74

Table 4.4.5

The first eight moments of Example 4(B)

m/t	1	2	4	6	8	10
m_0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
m_1	.81607	.93218	.99048	.99847	.99975	1.0000
m_2	1.2183	1.5582	1.7433	1.7694	1.7736	1.7744
m_3	2.4470	3.4561	4.0467	4.1317	4.1455	4.1482
m_4	5.9740	9.2345	11.273	11.573	11.621	11.631
m_5	16.991	28.565	36.262	37.417	37.604	37.642
m_6	54.946	99.964	131.71	136.57	137.36	137.52
m_7	198.47	389.11	531.37	535.55	557.17	557.90

The moments are reasonably close and tend to their steady-state values at the same rate.

APPENDIX A

A GEOSTOCHASTIC MODEL

In This's Appendix we consider a geostochastic model which is not covered by the kind of stochastic differential equations discussed in previous chapters

$$dX(t) = (X(t) - X^2(t))dt + \rho(\delta(X(t)) - X(t))dt + (\gamma X(t))^{1/2}\, dW(t) \tag{A.1}$$

$$X(0) = \text{given} \tag{A.2}$$

where $\gamma > 0$ and $0 \leq \rho \leq 1$ defined on $(0,\infty)$. The main difference is that in (A.1), the drift term depends on the law of the process if $\rho \neq 0$. The convergence theorems developed in this dissertation do not apply to this model. However the Gauss-Galerkin method can still be applied to such problems, as we shall illustrate below.

The Fokker-Planck equation corresponding to the equation (A.1) is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left[(\rho \int yp(t,y)\,dy + (1-\rho)x - x^2)p(t,x)\right] + \frac{1}{2}\gamma \frac{\partial^2}{\partial x^2}(xp(t,x)) = L^* p(t,x) \ .$$

The steady-state density of equation (A.1) is given by (see Dawson [5])

$$(A.3)\quad P_e(x) = Cx^{(\frac{2e}{\gamma}-1)} \exp(-(x-1+\rho)^2/\gamma), \quad x \geq 0$$

where C is a constant which makes $C\int_0^\infty P_e(x)\,dx = 1$ and e is a positive constant which satisfies the equation

$$(A.4)\qquad e = m(e) = \rho \int_0^\infty x\, P_e(x)\, dx \ .$$

For fixed ρ, γ and any positive number e, $m(e)$ is an integral which can be computed using the Gauss-Laguerre integration formula. Then we plot the graph of $y = m(e)$ and $y = e$ where the intersection provides an initial value for finding e^*, the root of the equation $e = m(e)$. Having $P_{e^*}(x)$, we can compute the exact moments and therefore the corresponding Gauss- Christoffel nodes and weights for comparison with the Gauss-Galerkin nodes and weights.

The Hankel system of moments corresponding to equation (A.1) is given by

$$(A.5)\qquad \frac{dm_1(t)}{dt} = \rho m_1(t) + a m_1(t) - m_2(t)$$

$$(A.6)\qquad \frac{dm_n(t)}{dt} = (n\rho m_1(t) + \tfrac{1}{2}n(n-1)\gamma)\, m_{n-1}(t) + anm_n(t) - nm_{n+1}(t), \quad n \geq 2$$

$$m_n(0) = \mathcal{E}(X_0^n), \quad a = 1-\rho$$

As we can see, the system (A.5) is not closed. Using the algorithm given in (2.9.1) we may close the system, solve it numerically and compare the steady-state results with exact steady-state moments. We studied the case of $\rho = .5$ and $\nu = 1$, i.e. for the stochastic differential equation

$$dX(t) = (.5X(t) + .5\delta(X(t)) - X^2(t))dt + (X(t))^{\frac{1}{2}}dW(t) \quad . \tag{A.7}$$

The first six exact steady-state moments are

$$\begin{aligned} m_0 &= 1.0000 \\ m_1 &= .60284 \\ m_2 &= .60434 \\ m_3 &= .77234 \\ m_4 &= 1.1602 \\ m_5 &= 1.9805 \end{aligned} \tag{A.8}$$

The 5-point Gauss-Galerkin steady-state nodes and weights of the equation (A.7) are computed. Also using the Gauss-Galerkin nodes and weights, the first six steady-state moments are computed. The results are shown below.

$$
\begin{aligned}
m_0 &= 1.0000 \\
m_1 &= .56044 \\
m_2 &= .56044 \\
m_3 &= .71749 \\
m_4 &= 1.0762 \\
m_5 &= 1.8154
\end{aligned}
\tag{A.9}
$$

The system (4.4.5) corresponding to equation (A.7) for $n = 10$ is solved and the first six moments are shown below.

$$
\begin{aligned}
m_0 &= 1.0000 \\
m_1 &= .56044 \\
m_2 &= .56044 \\
m_3 &= .71748 \\
m_4 &= 1.0762 \\
m_5 &= 1.8153
\end{aligned}
\tag{A.10}
$$

By the equivalence of the system (2.2.4) and (2.5.12) we expect that the results in (A.9) and (A.10) to be very close and indeed they are. The numerical results for first three moments are accurate within 5%. We have done computations for this example with Laguerre polynomials as a basis for polynomials of degree less than or equal to $2n-1$ instead of $1, x, \cdots, x^{2n-1}$, the numerical results are identical.

APPENDIX B

GALERKIN METHOD WITH FIXED NODES

In Section 2.2 we discussed the n-point Gauss-Christoffel approximation

$$\tilde{\mu}_n = \sum_{k=1}^{n} \tilde{a}_k \, \delta_{\tilde{x}_k}$$

for a measure μ defined on (r_1, r_2) where $-\infty \leq r_1 < r_2 \leq \infty$. Since the nodes $\{\tilde{x}_k\}$ are the zeros of orthogonal polynomials, the weights $\{\tilde{a}_k\}$ can be chosen so that

$$\int_{r_1}^{r_2} f(x)\,d\mu = \sum_{k=1}^{n} \tilde{a}_k f(\tilde{x}_k) + E[f]$$

is exact for all polynomials of degree less than or equal $2n-1$. This was the basis for the Gauss-Galerkin method that we developed in Chapter III. We shall consider in this Appendix another Galerkin method based on nodes that are fixed in time. Let $x_1, x_2, \ldots, x_n$ be any n distinct points. We can find constants $\hat{a}_1, \ldots, \hat{a}_n$ uniquely so that

$$\int_{r_1}^{r_2} f(x)\,d\mu = \sum_{k=1}^{n} \hat{a}_k f(x_k) + E[f]$$

is exact for all polynomials of degree less than or equal to $n-1$. (Stroud [21], p. 107). The sum $\sum_{k=1}^{n} a_k f(x_k)$ is called a quadrature sum. Also, as before, if $f(x), f^{(1)}(x), \ldots, f^{(n)}(x)$ are continuous on $[r_1, r_2]$, then there exists a function $K(s)$ so that

$$E[f] = \int_{r_1}^{r_2} f(x)\,d\mu - \sum_{k=1}^{n} \hat{a}_k f(x_k) = \int_{r_1}^{r_2} K(s) f^{(n)}(s)\,ds \quad .$$

Let $p(t,x)$ be the solution to the equation (2.1.3) and (2.1.4). For given nodes $x_1, \ldots, x_n$, that do not change with t, the n-point quadrature sum $\sum_{k=1}^{n} \hat{a}_k(t)\,\delta_{x_k}$ is an approximation to $p(t,x)\,dx$ in the sense that equation (2.1.7) can be written as

$$\text{(B.1)} \qquad \frac{d}{dt}\Big(\sum_{k=1}^{n} \hat{a}_k(t) v(x_k) + E(v)\Big)$$

$$= \sum_{k=1}^{n} \hat{a}_k(t) (Lv)(x_k) + E[Lv]$$

If $\{f_i(x)\}$, $i = 1, \ldots, n$ is a basis for polynomials of degree less than or equal to $n-1$, equation (B.1) for $f_i(x)$'s becomes

$$\text{(B.2)} \qquad \frac{d}{dt} \sum_{k=1}^{n} \hat{a}_k(t) f_i(x_k) = \sum_{k=1}^{n} \hat{a}_k(t) (Lf_i)(x_k)$$

$$+ E[Lf_i], \quad i = 1, \ldots, n$$

The system (B.2) is a system of ordinary differential equations for the weights $\hat{a}_k(t)$. It is not a closed system as $p(t,x)$ is involved in $E[Lf_i]$ and thus can not be used for the solution of the weights $\{\hat{a}_k(t)\}$ without explicit knowledge of $p(t,x)$.

The Galerkin method with fixed nodes for approximation $p(t,x)$ is obtained from (B.2) with the terms $E[Lf_i]$ dropped

$$\text{(B.3)} \qquad \frac{d}{dt} \sum_{k=1}^{n} \hat{a}_k(t) f_i(x_k) = \sum_{k=1}^{n} a_k(t) (Lf_i(x_k)) \ ,$$

for $i = 1,2,\ldots,n$ where $\{f_i(x)\}_1^n$ is a basis for polynomials of degree less than or equal to $n-1$, which we shall take to be $1,x,\ldots x^{n-1}$. The system (B.3) with given initial $p(0,x)$ may be cast in matrix form as

$$\text{(B.4)} \qquad AX' = BX$$

$$\text{(B.5)} \qquad X(0) = \text{given}$$

where

$$X^T = (\hat{a}_1(t),\ldots,\hat{a}_n(t)) \ ,$$

$$A = \begin{pmatrix} f_1(x_1) & f_1(x_2) & \cdots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \cdots & f_2(x_n) \\ \vdots & & & \\ f_n(x_n) & f_n(x_2) & \cdots & f_n(x_n) \end{pmatrix}$$

and

$$B = \begin{pmatrix} Lf_1(x_1) & Lf_1(x_2) & \cdots & Lf_1(x_n) \\ Lf_2(x_1) & Lf_2(x_2) & \cdots & Lf_2(x_n) \\ \vdots & & & \\ Lf_n(x_1) & Lf_n(x_2) & \cdots & Lf_n(x_n) \end{pmatrix}$$

The matrix A in (B.4) is the well known Vandermonde matrix which is non-singular.

In the case when the coefficients of the stochastic differential equation are polynomials, we can find as in Gauss-Galerkin method the corresponding Hankel system of ordinary differential equations.

$$\text{(B.6)} \qquad \frac{dm_k(t)}{dt} = g_k(m_1, m_2, \cdots) \qquad k = 0,1,\cdots,n-1$$

$$\text{(B.7)} \qquad m_k(0) = \text{given} \qquad k = 0,1,\cdots,n-1$$

In this case also if the degree of at least one of the coefficients is greater than one, moments of order higher than $n-1$ appears in some of the equations of the system.

Theorem B.1. The system (B.6) can be made closed.

Proof. As before we shall show that it is possible to express all the moments that appear in (B.6) with order higher than $n-1$ in terms of the lower order moments.

Recall that

$$m_n(t) = \sum_{k=1}^{n} a_k(t) x_k^n \qquad n = 0,1,\cdots$$

In matrix form, we can write

$$\text{(B.8)} \qquad \begin{pmatrix} m_0 \\ m_1 \\ \vdots \\ m_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & & & \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$\text{(B.9)} \qquad \begin{pmatrix} m_n \\ m_{n+1} \\ \vdots \\ m_{2n-1} \end{pmatrix} = \begin{pmatrix} x_1^n & x_2^n & \cdots & x_n^n \\ x_1^{n+1} & x_2^{n+1} & \cdots & x_n^{n+1} \\ \vdots & \vdots & & \\ x_1^{2n-1} & x_2^{2n-1} & \cdots & x_n^{2n-1} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

In general if

$$M_k^T = (m_{kn}, m_{kn+1}, \cdots, m_{kn+n-1})$$

and

$$A_k = \begin{pmatrix} x_1^{kn} & x_2^{kn} & \cdots & x_n^{kn} \\ x_1^{kn+1} & x_2^{kn+1} & \cdots & x_n^{kn+1} \\ \vdots & & & \\ x_1^{kn+n-1} & x_2^{kn+n-1} & \cdots & x_n^{kn(n-1)} \end{pmatrix} ,$$

then we have

$$M_k = A_k X \qquad k = 0,1,2,\ldots \quad .$$

From (A.1.8) we have

$$X = A_0^{-1} M_0 \ .$$

Therefore,

$$M_k = A_k A_0^{-1} M_0 \qquad k = 1,2,\ldots$$

and the proof is complete.

As we see from equation (B.4) and (B.8) $A_0 = A$ is independent of time in the present case and therefore we invert A_0 only once while in the Gauss-Galerkin method. We must perform matrix inversion at each time step. The above method is easy to implement and often yields satisfactory results. Unfortunately we do not have convergence theorem as $n \to \infty$ as those we established

in Chapter III for the Gauss-Galerkin method there. This should not come as a surprise as it is well known that quadrature formulas with fixed nodes named as the Newton-Cotes formulas do not have convergence in general though at each n we do have error bounds.

As a numerical example we solve the system (B.6) for the geostochastic model (A.7) with $n = 10$ and an initial atomic measure chosen randomly with $x_i = \frac{i}{4}$, $a_i = .1$ $i = 1,\cdots,10$. The first six steady-state moments are shown below.

$$m_0 = 1.0000$$
$$m_1 = .56242$$
$$m_2 = .56242$$
$$m_3 = .72058$$
$$m_4 = 1.08086$$
$$m_5 = 1.82393$$

The results are close to those obtained by the Gauss-Galerkin method.

LIST OF REFERENCES

LIST OF REFERENCES

1. Arnold: Stochastic Differential Equations, Theory and Application, Wiley Intersciences, 1974.

2. Buck: Studies in Modern Analysis, A.M.S. 1962.

3. Chow and Hale: Methods of Bifurcation Theory, Springer, 1982.

4. Coddington and Levinson: Theory of Differential Equations, New York, McGraw Hill, 1955.

5. Dawson: Qualitative behavior of geostochastic systems, J. Stochastic Processes and their Application, 10 (1980) pp. 1-31.

6. _____: Galerkin Approximation of Nonlinear Markov Processes; Statistics and Related Topics. Proceeding of the international symposium on Statistics and Related Topics held in Ottawa, Canada, May 5-7. 1980. North-Holland Publishing Company.

7. Ethier: A Class of Degenerate Diffusion Processes Occuring in Population Genetics. Comm. Pure Appl. Math. 1976, pp. 483-493.

8. Feller: (a) The Parabolic Differential Equations and the Associated Semi-Group of Transformation, Ann. Vol. 55, 1952, pp. 468-519.

   _____: (b) Diffusion Processes in One Dimension, Amer. Math. Soc. Trans. Vol. 77, 1954, pp. 1-31.

9. Fischer: Ueber das Caratheodorysche Problem, Potenzreihen mit Positive reellen Teil betreffend, Rendiconti del Circolo Matematica di Palermo, 32 (1911), 240-256.

10. Forsythe and Moler: Computer Solution of Linear Algebraic Systems, Prentice-Hall, 1977.

11. Friedman: Stochastic Differential Equations and Application to P.D.E.'s, Vol. I, Academic Press, Inc. 1975.

12. Gikhman and Skorokhod: (a) Introduction to the Theory of Random Processes, Philadelphia W.B. Saunder, 1969.

_____: (b) Stochastic Differential Equations, Springer 1972.

13. Hale: Ordinary Differential Equations, Wiley-Intersciences, 1969.

14. Krylov: Approximation Calculations of Integrals, Macmillan, New York, 1962 (Trans. from 1-st Russian ed. 1959 by A. Stroud).

15. Mandl: Analytic treatment of One-Dimensional Markov Processes, Springer (1968).

16. Mangnus, Oberhettinger and Soni: Formulas and Theorems for Special Functions of Mathematical Physics, Springer, 1966.

17. Muller-Pfeiffer: Spectral Theory of Ordinary Differential Operators, John Wiley and Sons, 1981.

18. Rainville: Special Functions, the Macmillan Co., 1960.

19. Shohat and Tamarkin: The Problem of Moments, A.M.S. 1950.

20. Strang and Fix: An Analysis of the Finite Element Method, Prentice Hall, Inc., 1973.

21. Stroud: Numerical Quadrature and Solutions of O.D.E.'s, Springer, 1974.

22. Stroud and Kwan-wei Chen: Peano Error Estimates for Gauss-Lagrange Quadrature Formulas, SIAM. Numer. Vol. 9, No. 2 (1972).

23. Stroud and Stanch: Quadrature Formulas with Multiple Gaussian nodes, J. SIAM. Numer. Anal. Ser. B - Vol. 2, No. 1, pp. 129-143.

24. Szarski: Differential Inequalities, Wars Zawa, 1965.

25. Uspensky: On the Convergence of Quadrature Formulas Related to an Infinite Interval, Trans. Am. Math. Soc. 30 (1928), pp. 542-559.