◊

Affine Processes

and their Applications to Financial Mathematics

By

ELIE FAYEZ SALIBA

Thesis Advisor: Dr. Ramez Maalouf

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Mathematics
in the Department of Mathematics and Statistics
in the Faculty of Natural and Applied Sciences
of Notre Dame University-Louaizé
Lebanon

July 9, 2015

NDU LIBRARY
20 JUN 2016
RECEIVED

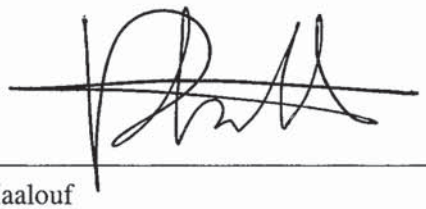
NDU FACULTY OF NATURAL & APPLIED SCIENCES

Affine Processes

and their
Applications to Financial Mathematics

ELIE FAYEZ SALIBA

Approved by:



Dr. Ramez Maalouf

Dr. George Eid

Dr. Bassem Ghalayini

Dr. John Haddad

Acknowledgements

I would first like to thank my parents for their support throughout the period of my study at NDU towards the Master of Science degree in Mathematics.

In addition, I owe a special thanks to Zeina Kadri for her great support during my journey.

I would also like to thank the jury members of my thesis defense: Dr. George Eid, Dr. Bassem Ghalayini and Dr. John Haddad.

Finally I want to express my gratitude to Dr. Ramez N Maalouf for suggesting this topic on affine processes, and for his follow up during the different stages of my work.

Abstract

In this work we discuss a special kind of stochastic processes $\mathbf{X} = \{X_t\}_{t\in\mathbf{R}_{\geq 0}}$, that are of exceptional interest from both the theoretical and the applied points of view. These processes are called **Affine Processes**, and are characterised by the fact that their characteristic function has the form of an exponential of an affine function, i.e. by

$$\phi^x_{\mathbf{X}}(u)|_t = \mathbb{E}^x[e^{<X_t,u>}] = e^{\Phi(t,u)+<x,\Psi(t,u)>}, \tag{1}$$

where the exponent

$$\Phi(t,u)+ < x, \Psi(t,u) > \tag{2}$$

is an affine function of its initial state x in the state-space $\mathbf{E} = \mathbf{R}^m_{\geq 0} \times \mathbf{R}^n$. The above expectation $\mathbb{E}^x$ is the expectation with respect to the law $\mathbb{P}^x$ of the process started at x.

In chapter 2 we introduce affine processes and discuss the main properties associated with these processes, and we give examples of such processes. In addition, we discuss in detail the *semi-flow property* and the *Feller property* for affine processes. We also discuss the important class of *regular affine processes.*

Chapter 3 discusses the application of affine processes to financial mathematics. In sectio 3.1 we introduce basic aspects of the math of finance, while in section 3.2 we discuss some applications of affine processes in financial mathematics.

Finally, given that the general subjects of stochastic processes, stochastic calculus, and stochastic differential equations are highly technical subjects, and so many definitions are needed for a smooth reading of such a work, we give a quite detailed first chapter on the basics of stochastic processes and stochastic calculus, to pave the way for a clear understanding of the rest of the thesis.

Contents

Chapter 1

Foundations on Stochastic Processes, Stochastic Calculus, and SDE's

For clear discussions related to Stochastic Processes and Stochastic Calculus, you first need

" ... a six months course (only) on definitions ..."

A.P.Meyer

In this chapter we introduce some basic definitions and elements for general stochastic processes, stochastic calculus, and stochastic differential equations, to make the rest of this thesis a useful and understandable piece of work.

1.1 Basics on Stochastic Processes and Martingales

In this section we closely follow the discussion and definitions given in "*Diffusions, Markov Processes and Martingales*" by Rodgers and Williams [RW].

1.1.1 Basic Elements for Stochastic Processes

Stochastic Processes, Sample Functions, and the $\mathscr{E}^{\mathrm{T}}$ Sigma-Algebra

♣ A DEFINITION (CHARACTERIZATION) OF STOCHASTIC PROCESSES. We first recall that a **stochastic process** is a family $\mathbf{X}$ of random variables, written as

$$\mathbf{X} = \{X_t, t \in \mathrm{T}\}, \tag{1.1}$$

where T is a parameter set, defined over a probability space $(\Omega, \sigma, \mathbf{P})$, with values in a **state space** (a measurable space) $(\mathbf{E}, \mathscr{E})$. Thus a stochastic process is a collection of $(\mathbf{E}, \mathscr{E})$-valued random variables over $(\Omega, \sigma, \mathbf{P})$, i.e.

$$\boxed{X_t : \Omega \longrightarrow \mathbf{E}, \quad \forall t \in \mathrm{T}} \quad \text{or} \quad \boxed{\mathbf{X} : \mathrm{T} \times \Omega \longrightarrow \mathbf{E}} \tag{1.2}$$

where X_t^{-1} defines a mapping from the sigma-algebra $\mathscr{E}$ to the sigma-algebra σ, i.e.

$$X_t^{-1} : \mathscr{E} \longrightarrow \sigma, \tag{1.3}$$

given that each X_t is measurable.

♣ SAMPLE FUNCTIONS & ANOTHER CHARACTERIZATION OF STOCHASTIC PROCESSES. Given a stochastic process $\mathbf{X}$ one defines the **sample function** (also called the **sample path** if $\mathrm{T} = \mathbf{R}_{\geq 0}$), of $\mathbf{X}$ corresponding to the point $\omega \in \Omega$, as being the function

$$\mathbf{X}(\omega) : \mathrm{T} \longrightarrow \mathbf{E} \tag{1.4}$$

such that

$$\boxed{\mathbf{X}(\omega)(t) = X_t(\omega), \quad \forall t \in \mathrm{T}} \tag{1.5}$$

Thus $\mathbf{X}(\omega) \in \mathbf{E}^{\mathrm{T}}$, where $\mathbf{E}^{\mathrm{T}}$ is *the space of all functions from* T *to* $\mathbf{E}$.

This idea of sample function establishes another characterization (or definition) of a stochastic process as a function

$$\mathbf{X} : \Omega \longrightarrow \mathbf{E}^{\mathrm{T}}, \tag{1.6}$$

instead of a function $\mathbf{X} : \mathrm{T} \times \Omega \longrightarrow \mathbf{E}$, *that maps every* $\omega \in \Omega$ *to an element (a function)* $\mathbf{X}(\omega)$ *in* $\mathbf{E}^{\mathrm{T}}$, to be called the "sample function" associated with ω, such that $\mathbf{X}$ is measurable with respect to the sigma-algebra σ on Ω and a sigma-algebra $\mathscr{E}^{\mathrm{T}}$ (to be defined next) on $\mathbf{E}^{\mathrm{T}}$. This would characterize a stochastic process as just being a measurable function

$$\boxed{\mathbf{X} : (\Omega, \sigma) \longrightarrow (\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})} \tag{1.7}$$

and this would make the stochastic process $\mathbf{X}$ as (simply) being a *random variable* from the probability space $(\Omega, \sigma, \mathbf{P})$ to the state-space $(\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})$. The (measurable) function $\mathbf{X}$ thus establishes a mapping $\mathbf{X}^{-1}$ from $\mathscr{E}^{\mathrm{T}}$ on $\mathbf{E}^{\mathrm{T}}$ to σ, i.e.

$$\mathbf{X}^{-1} : \mathscr{E}^{\mathrm{T}} \longrightarrow \sigma, \tag{1.8}$$

that maps any $\Sigma \in \mathscr{E}^{\mathrm{T}}$, i.e. maps any measurable collection Σ of functions in $\mathscr{E}^{\mathrm{T}}$, to the (measurable set in σ consisting of) points $\omega \in \Omega$ whose "sample functions" $\mathbf{X}(\omega)$ are in Σ.

♣ THE $\mathscr{E}^{\mathrm{T}}$ SIGMA-ALGEBRA. To properly establish the above characterization of a stochastic process, we only need to define $\mathscr{E}^{\mathrm{T}}$ as being the smallest sigma-algebra on the space $\mathbf{E}^{\mathrm{T}}$ of functions from T to $\mathbf{E}$ such that $\forall t \in \mathrm{T}$, the *evaluation map*

$$\pi_t : \mathbf{E}^{\mathrm{T}} \longrightarrow \mathbf{E}, \tag{1.9}$$

defined by

$$\pi_t(f) = f(t), \qquad \forall f \in \mathbf{E}^{\mathrm{T}}, \tag{1.10}$$

is measurable. Thus

$\mathscr{E}^{\mathrm{T}}$ is the smallest sigma-algebra on $\mathbf{E}^{\mathrm{T}}$ that makes every evaluation map π_t (as above), measurable.

This would establish any stochastic process $\mathbf{X}$, defined on the probability space $(\Omega, \sigma, \mathbf{P})$ with state space $(\mathbf{E}, \mathscr{E})$, as a measurable function from (Ω, σ) to $(\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})$, and thus making $\mathbf{X}$ an $(\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})-$valued random variable on the probability space $(\Omega, \sigma, \mathbf{P})$.

And indeed, one can show that the two definitions (given above) for a stochastic process are equivalent. Thus

$$\boxed{\begin{array}{l} X_t : (\Omega, \sigma) \longrightarrow (\mathbf{E}, \mathscr{E}) \text{ is measurable for every } t \in \mathrm{T}, \\ \text{if and only if } \ \mathbf{X} : (\Omega, \sigma) \longrightarrow (\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}}) \text{ is measurable.} \end{array}}$$

Laws of Stochastic Processes

♣ THE LAW OF A STOCHASTIC PROCESS. We recall that, for a stochastic process $\mathbf{X}$, the **law of the process** is the probability measure μ on the measurable space $(\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})$ (defined in the previous subsection), given by

$$\boxed{\mu(\Sigma) = (\mathbf{P} \circ \mathbf{X}^{-1})(\Sigma)} \tag{1.11}$$

for every measurable subset $\Sigma \in \mathscr{E}^{\mathrm{T}}$ of functions from T to E. Thus the law of the stochastic process $\mathbf{X}$ is the law (the "distribution") of the *random variable* $\mathbf{X} : (\Omega, \sigma) \longrightarrow (\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})$.

♣ THE LAW OF A STOCHASTIC PROCESS STARTED AT $x \in \mathbf{E}$. We define the **law of the stochastic process started at** $x \in \mathbf{E}$ as being the probability measure, denoted by $\mathbb{P}^x$, on (Ω, σ) defined as follows. First let $\Omega_x \subset \Omega$ be defined as the set of all $\omega \in \Omega$ such that the corresponding sample function $\mathbf{X}(\omega)$ starts at $x \in \mathbf{E}$, i.e.

$$\Omega_x = \{\omega \in \Omega : X_0(\omega) = x\}. \tag{1.12}$$

Roughly speaking, $\mathbb{P}^x$ measures, as described below, and in a way related to $\mathbf{P}$, the intersection with Ω_x of any measurable $B \in \sigma$, with respect to the $\mathbf{P}-$measure of Ω_x.

- *The case where* $\mathbf{P}(\Omega_x) > 0$. For every measurable $B \in \sigma$ we simply have

$$\boxed{\mathbb{P}^x(B) = \mathbf{P}(B \cap \Omega_x)/\mathbf{P}(\Omega_x)} = \frac{\mathbf{P}(B \cap \Omega_x)}{\mathbf{P}(\Omega_x)}. \tag{1.13}$$

  In other words $\mathbb{P}^x$ measures (by $\mathbf{P}$, and with respect to $\mathbf{P}(\Omega_x)$) the subset of B (i.e. $B \cap \Omega_x$) that consists of all ω whose sample functions start at x.

- *The case where* $\mathbf{P}(\Omega_x) = 0$. In this case the definition of $\mathbb{P}^x$ is more intricate, in which it can be imagined to be a probability measure on (Ω, σ) with $\mathbb{P}^x(\Omega_x) = 1$, such that for any two points $\omega_1, \omega_2 \in \Omega_x$, the "relative density" of $\mathbf{P}$ at ω_1 and ω_2, is the same as the "relative density" of $\mathbb{P}^x$ at these two points. The exact procedure to construct such a $\mathbb{P}^x$ is not straightforward.

Given the previous definition of $\mathbf{X}^{-1} : \mathscr{E}^{\mathrm{T}} \longrightarrow \sigma$ (as above), one can also define the law of the process started at x as being the probability measure on $(\mathbf{E}^{\mathrm{T}}, \mathscr{E}^{\mathrm{T}})$, such that for every $\Sigma \in \mathscr{E}^{\mathrm{T}}$, $\mathbb{P}^x$ measures the functions (from T to $\mathbf{E}$) in Σ that are sample functions, associated with the process $\mathbf{X}$, and that start at $x \in \mathbf{E}$. This is given as follows:

- *The case where* $\mathbf{P}(\Omega_x) > 0$.

$$\boxed{\mathbb{P}^x(\Sigma) = \underline{\mathbf{P}}(\mathbf{X}^{-1}(\Sigma) \cap \Omega_x)/\mathbf{P}(\Omega_x)} = \frac{\mathbf{P}(\mathbf{X}^{-1}(\Sigma) \cap \Omega_x)}{\mathbf{P}(\Omega_x)}, \qquad \forall \Sigma \in \mathscr{E}^{\mathrm{T}}. \tag{1.14}$$

- *The case where* $\mathbf{P}(\Omega_x) = 0$. This is also more intricate, and needs similar arguments as those for the previous part (above).

Stochastically Continuous Processes

A Stochastic Process $\mathbf{X} = \{X_t\}$, where $t \in \mathrm{T} = \mathbf{R}_{\geq 0}$, defined over a probability space $(\Omega, \sigma, \mathbf{P})$ and with values in a state space $\mathbf{E}$ (a topological space, with the Borel sigma-algebra) is said to be a **Stochastically Continuous Process** if for every sequence $\{t_n\}_{\geq 0}$ that converges to t, the sequence $\{X_{t_n}\}$ converges in probability (see next paragraph), with respect to $\mathbf{P}$ (or equivalently with respect

to every $\mathbb{P}^x, x \in \mathbf{E}$, where $\mathbb{P}^x$ is the law of the stochastic process started at x) to X_t.

This means that for every $t \in \mathrm{T}$, the set of $\omega \in \Omega$ where there exists a sequence $t_n \to t$ with $X_{t_n}(\omega) \not\to X_t(\omega)$ has $\mathbf{P}$-measure zero (or equivalently has $\mathbb{P}^x$-measure zero for every $x \in \mathbf{E}$). Thus, almost everywhere in Ω, one has that $X_{t_n}(\omega) \to X_t(\omega)$ for every $t_n \to t$ and any $t \in \mathrm{T}$.

This is equivalent to having, for any $t \in \mathrm{T}$ and every $t_n \to t$, that

$$\boxed{\forall \mathrm{A}_{open} \subset \mathbf{E}, \quad \lim_{n\to\infty} \mathbf{P}(X_{t_n}^{-1}(\mathrm{A})) = \mathbf{P}(X_t^{-1}(\mathrm{A}))} \tag{1.15}$$

This is also equivalent to having, for every $x \in \mathbf{E}$ and every $\mathrm{A}_{open} \subset \mathbf{E}$, that

$$\boxed{\lim_{n\to\infty} \mathbb{P}^x(X_{t_n}^{-1}(\mathrm{A})) = \mathbb{P}^x(X_t^{-1}(\mathrm{A}))} \tag{1.16}$$

1.1.2 Basic Elements for Martingales

Roughly, a (for example discrete-time) martingale is a sequence of random variables (i.e., a stochastic process) for which, at a particular time in the realized sequence, the expectation of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values, as is specified next.

♣ CONDITIONAL EXPECTATION. The idea of *conditional expectation* is basic for defining martingales. Thus, given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, the conditional expectation of a random variable $Y : \Omega \longrightarrow \mathbf{R}^n$ with respect to a family $\{X_\alpha : \Omega \longrightarrow \mathbf{R}^n\}$ of random variables, denoted by

$$\mathbb{E}[Y/\{X_\alpha\}], \tag{1.17}$$

is given by a random variable $Z : \Omega \longrightarrow \mathbf{R}^n$, such that the expectation of Y over any measurable subset S (i.e. element) in the

$$\boxed{\text{smallest sigma-algebra } \mathcal{F}(\{X_\alpha\}) \subset \mathcal{F} \text{ over } \Omega \text{ that makes all } X_\alpha \text{ measurable}}$$

is equal to the expectation of Z over S, i.e.

$$\mathbb{E}_S[Y] = \mathbb{E}_S[Z], \qquad \forall S \in \mathcal{F}(\{X_\alpha\}), \tag{1.18}$$

where (also) Z has $\mathcal{F}(\{X_\alpha\})$ as the smallest sigma-algebra to make it measurable. In this case we write that

$$\boxed{\mathbb{E}[Y/\{X_\alpha\}] = Z} \tag{1.19}$$

In fact this conditional expectation is also written as

$$\boxed{\mathbb{E}[Y/\mathcal{F}(\{X_\alpha\})] = Z} \tag{1.20}$$

Discrete-Time Martingales

♣ DISCRETE-TIME MARTINGALES. A **discrete-time martingale** is defined to be a discrete-time stochastic process (i.e., a sequence of random variables) $X_1, X_2, X_3, \cdots$ that satisfies the following:

1. $\mathbb{E}[|X_n|] < \infty, \quad \forall\, n = 1, 2, 3, \cdots$
2. $\mathbb{E}[X_{n+1}|X_1, X_2 \cdots X_n] = X_n. \quad \forall\, n = 1, 2, 3, \cdots$

Thus, in a discrete-time martingale, *the conditional expected value of the next observation X_{n+1}, given all the past observations $X_1, \cdots, X_n$, is equal to the last observation X_n*. Due to the linearity of expectation, this second requirement is equivalent to:

$$\mathbb{E}[X_{n+1} - X_n|X_1, X_2 \cdots X_n] = 0. \tag{1.21}$$

which states that the average "gains" from observation n to observation $n+1$ are 0.

♣ MARTINGALE $\{Y_n\}$ WITH RESPECT TO ANOTHER SEQUENCE $\{X_n\}$. Along the line of the above definition, one has that a sequence $Y_1, Y_2, Y_3 \cdots$ is said to be a (discrete) **martingale with respect to another sequence** $X_1, X_2, X_3 \cdots$ if:

- $\mathbb{E}[|Y_n|] < \infty, \quad \forall\, n = 1, 2, 3, \cdots$
- $\mathbb{E}[Y_{n+1}|X_1, X_2 \cdots X_n] = Y_n. \quad \forall\, n = 1, 2, 3, \cdots$

Continuous-Time Martingales

♣ CONTINUOUS-TIME MARTINGALES. A **continuous-time martingale** $\mathbf{Y} = \{Y_t\}$ with respect to the stochastic process $\mathbf{X} = \{X_t\}$, is a stochastic process such that for all $t \in \mathbf{R}_{\geq 0}$:

1. $\mathbb{E}[|Y_t|] < \infty$,
2. $\mathbb{E}[Y_t | \{X_\tau, \tau \leq s\}] = Y_s, \quad \forall s \leq t$

This expresses the property that *the conditional expectation of the observation X_t at time t, given all the observations X_τ up to time s , is equal to the observation X_s at time s.*

Note: Given the above definitions, one can in fact give a general definition for a stochastic process to be an appropriate martingale.

Martingales with Respect to Filtrations

♣ FILTRATIONS OF PROBABILITY SPACES. We recall that a **filtration** $\{\mathscr{F}_t : t \in \mathrm{T} = \mathbf{R}_{\geq 0}\}$ of a probability space$(\Omega, \sigma, \mathbf{P})$ is a family of sigma-algebras $\mathscr{F}_t$ on Ω such that

$$\boxed{\mathscr{F}_t \subseteq \mathscr{F}_{t'} \subseteq \sigma = \mathscr{F}_\infty, \qquad \forall t < t' \in \mathrm{T}} \tag{1.22}$$

This definition for filtration allows for the following definition.

♣ MARTINGALES WITH RESPECT TO FILTRATIONS. A process $\mathbf{X} = \{X_t\}_{t \geq 0}$ over a probability space $(\Omega, \sigma, \mathbf{P})$, i.e. where

$$\mathbf{X} : \mathbf{R}_{\geq 0} \times \Omega \longrightarrow \mathbf{E} \tag{1.23}$$

is called a **martingale with respect to a filtration** $\{\mathscr{F}_t\}_{t \geq 0}$ of the given probability space if :

1. $\{X_t\}$ is **adapted** to the the filtration $\{\mathscr{F}_t\}$, i.e. X_t is $\mathscr{F}_t-$measurable, for every $t \geq 0$,

2. $\mathbb{E}[|X_t|] < \infty$,

3. $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$.

♣ FILTRATIONS GENERATED BY PROCESSES. A filtration $\{\mathscr{F}_t\}$ of a probability space $(\Omega, \sigma, \mathbf{P})$ is said to be **generated by a process** $\mathbf{X} = \{X_t\}_{t\geq 0}$ over $(\Omega, \sigma, \mathbf{P})$, if for every t, $\mathscr{F}_t$ is the smallest sigma-algebra to make all X_s, $s \leq t$, measurable with respect to this $\mathscr{F}_t$. We shall write that

$$\boxed{\mathscr{F}_t = \mathscr{F}(\{X_s\}_{s\leq t})} \tag{1.24}$$

If a filtration on a probability space, on which a process $\mathbf{X} = \{X_t\}_{t\geq 0}$ is defined, is mentioned without further restrictions, then it will be understood to be the filtration generated by the process.

Note that the property of being a martingale involves both the filtration $\{\mathscr{F}_t\}$ and the probability measure σ (with respect to which the expectations $\mathbb{E}$ are taken). Note also that it is possible that $\mathbf{X}$ could be a martingale with respect to one measure but not another one.

Supermartingales and Submartingales

♣ $\mathbb{E}[Y/\{X_\alpha\}] \leq Z$ AND $\mathbb{E}[Y/\{X_\alpha\}] \geq Z$. Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a random variable $Y : \Omega \longrightarrow \mathbf{R}^n$ and a family $\{X_\alpha : \Omega \longrightarrow \mathbf{R}^n\}$ of random variables, we shall mean by

$$\mathbb{E}[Y/\{X_\alpha\}] \leq Z, \tag{1.25}$$

or by

$$\mathbb{E}[Y/\{X_\alpha\}] \geq Z, \tag{1.26}$$

where $Z : \Omega \longrightarrow \mathbf{R}^n$ is another random variable, that the expectation of Y over any measurable subset S (element) in the sigma-algebra $\mathcal{F}(\{X_\alpha\}) \subset \mathcal{F}$ (as defined above), is *smaller than or equal to* (respectively *larger than or equal to*) the expectation of Z over S, where Z is measurable with respect to $\mathcal{F}(\{X_\alpha\})$.

♣ DISCRETE SUPERMARTINGALES. A **discrete supermartingale** is defined as in a discrete martingale, except that the equality in the last condition is replaced by $\leq$, i.e. one would have

$$\mathbb{E}[X_{n+1}|X_1, \cdots, X_n] \leq X_n \qquad \forall\, n = 1, 2, 3, \cdots . \tag{1.27}$$

♣ CONTINUOUS SUPERMARTINGALES. While in a **continuous supermartingale** one would have that

$$\boxed{\mathbb{E}[X_t|\{X_\tau, \tau \leq s\}] \leq X_s} \qquad \forall s \leq t. \tag{1.28}$$

♣ SUBMARTINGALES. Similarly one defines a **discrete submartingale**, or a **continuous submartingale**, in a similar way, only replacing $\leq$ by $\geq$, to have

$$\mathbb{E}[X_{n+1}|X_1, \cdots, X_n] \geq X_n \qquad \forall\, n = 1, 2, 3, \cdots , \tag{1.29}$$

and

$$\boxed{\mathbb{E}[X_t|\{X_\tau, \tau \leq s\}] \geq X_s} \qquad \forall s \leq t. \tag{1.30}$$

respectively.

Immediate Properties

1. $\mathbf{X}$ is a supermartingale if and only if $-\mathbf{X}$ is a submartingale.
2. $\mathbf{X}$ is a martingale if and only if $\mathbf{X}$ is both a supermartingale and a submartingale.

3. The process $\mathbf{X}$ for which $X_0 \in \mathcal{L}^1(\Omega, \mathscr{F}_0, \mathbf{P})$ is a martingale (respectively, supermartingale, submartingale) if and only if the process $X - X_0 = (X_n - X_0 \ : \ n = 1, 2, 3, \cdots)$ has the same property. So one can concentrate on processes that are null at 0.

Martingales and Stopping Times

♣ STOPPING TIMES. A stopping time with respect to a sequence of random variables $X_1, X_2, X_3, \cdots$ is a random variable T such that for each t, the occurrence or non-occurrence of the event $T = t$ depends only on the values of $X_1, X_2, X_3, \cdots, X_t$. The idea is that at any particular time t, you can look at the sequence so far and tell if it is time to stop.

The concept of stopping time is defined by requiring only that the occurrence or non-occurrence of the event $T = t$ be probabilistically independent of $X_{t+1}, X_{t+2}, \cdots$ but not that it be completely determined by the history of the process up to time t.

♣ BASIC PROPERTIES AND THEOREMS FOR STOPPING TIMES. Moreover, One of the basic properties of martingales is that, if $(X_t)_{t>0}$ is a (sub-/super-) martingale and τ is a stopping time, then the corresponding stopped process $(X_t^\tau)_{t>0}$ defined by $X_t^\tau := X_{\min\{\tau,t\}}$ is also a (sub-/super-) martingale.

The concept of a stopped martingale leads to a series of important theorems, including, for example, the *Optional Stopping Theorem* which states that, under certain conditions, the expected value of a martingale at a stopping time is equal to its initial value.

1.1.3 Markov Operators and Markov Semigroups

Defining Markov Operators and Markov Semigroups

♣ MARKOV OPERATORS. Let (Ω, σ, μ) be a $\sigma-$finite measure space. Then a linear mapping

$$P : L^1(\Omega, \sigma, \mu) \longrightarrow L^1(\Omega, \sigma, \mu) \tag{1.31}$$

is said to be a **Markov operator**, if it maps the set

$$D = \{f \in L^1(\Omega, \sigma, \mu) : f \geq 0 \text{ and } \|f\| = 1\} \tag{1.32}$$

(called the set of *densities*) into itself, i.e. if

$$f(D) \subset D. \tag{1.33}$$

♣ MARKOV SEMIGROUPS. Now a family $\{P_t\}_{t\geq 0}$ of Markov operators over a $\sigma-$finite measure space (Ω, σ, μ) is said to be a **Markov semigroup**, if it satisfies the following:

1. $P(0) = I$, i.e. $P(0)$ is the identity operator on $L^1(\Omega, \sigma, \mu)$ (that maps every element in $L^1(\Omega, \sigma, \mu)$ to itself).
2. $P(t + s) = P(t) \circ P(s)$ for any $s, t \geq 0$.
3. For any fixed $f \in L^1(\Omega, \sigma, \mu)$, one has that the mapping

$$t \longrightarrow P_t f \tag{1.34}$$

is a continuous function of t (into $L^1(\Omega, \sigma, \mu)$).

Examples of Markov Operators and Markov Semigroups

Assume a $\sigma-$finite measure space (Ω, σ, μ) in what follows.

- **Integral Operators**: Integral operators P of the form

$$(Pf)(x) = \int_\Omega k(x, y) f(y) \mu(dy), \tag{1.35}$$

where $k : \Omega \times \Omega \longrightarrow [0, \infty)$ is a measurable function such that

$$\int_\Omega k(x, y) \mu(dx) = 1 \tag{1.36}$$

for almost all $y \in \Omega$, define Markov operators. For these cases, the function $k(x, y)$ is called the *kernel* of P.

- **Markov Operators and Semigroups Associated with an SDE**: (*this example requires reading other parts of this chapter*) Consider the Stratonovich stochastic differential equation (SDE) given by

$$dX_t = \sigma dB_t + \sigma_0(X_t)dt, \tag{1.37}$$

where $\mathbf{B} = \{B_t\}$ is $n-$dimensional Brownian motion, with $\sigma(x)$ an $m \times n$ matrix with components ${\sigma_j}^i(x)$ and $\sigma_0(x) \in \mathbf{R}^m$ with components ${\sigma_0}^i(x)$. Then if a solution X_t of the above SDE, with X_0 having a distribution that is absolutely continuous with density $v(x)$, one finds that the density $u(x,t)$ of X_t satisfies the Fokker-Planck PDE given by

$$\frac{\partial u}{\partial t} = \sum_{i=1}^{m} \frac{\partial}{\partial x_i} \left(\sum_{j=1}^{m} a_{ij}(x) \frac{\partial u}{\partial x_j} \right) - \sum_{i=1}^{m} \frac{\partial({\sigma_0}^i(x)u)}{\partial x_i}, \tag{1.38}$$

where $a_{ij}(x) = \frac{1}{2} \sum_{k=1}^{n} {\sigma_k}^i(x){\sigma_k}^j(x)$.

Now for any $v(x) \in C_b^2(\mathbf{R}^m)$, i.e. $v(x)$ has continuous and bounded derivatives of orders 1 and 2 on $\mathbf{R}^m$, one can show that the above Fokker-Planck equation has, for any t in any interval $[0,T]$, a unique solution $u(x,t)$ such that $u(x,0) = v(x)$, with $u : [0,T] \times \mathbf{R}^m \longrightarrow \mathbf{R}$ being, together with its spatial derivatives of order 1 and 2, uniformly bounded.

Now if one defines an operator $P(t)$ of the form

$$P(t)v(x) = u(x,t) \tag{1.39}$$

for every $v(x) \in C_b^2(\mathbf{R}^m)$, then one can show that $P(t)$ can be extended to an operator $P(t)$ defined on $L^1(\mathbf{R}^m)$ (where $\Omega = \mathbf{R}^m$ in this case). This operator $P(t)$ is a Markov operator with the family $\{P(t)\}_{t \geq 0}$ being a Markov semigroup.

1.1.4 Characteristic Functions

The *characteristic function* of any real-valued random variable completely defines its probability distribution. If a random variable admits a probability density function, then the characteristic function is the inverse Fourier transform of

the probability density function. Thus characteristic functions provide the basis for an alternative way to analytical results compared with working directly with probability density functions.

Characteristic Functions of Random Variables

♣ CHARACTERISTIC FUNCTIONS OF REAL-VALUED RANDOM VARIABLES. Let X be a real-valued random variable defined over a probability space $(\Omega, \sigma, \mathbf{P})$, and let i be the imaginary unit (where $i^2 = -1$). The function

$$\phi_X : \mathbf{R} \longrightarrow \mathbf{C}, \tag{1.40}$$

defined by (for $t \in \mathbf{R}$)

$$\boxed{\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX(\omega)} dP(\omega)} \tag{1.41}$$

is called the **characteristic function** of X.

♣ CHARACTERISTIC FUNCTIONS OF MULTIVARIATE RANDOM VARIABLES. One also has generalizations of the above definition for the case of *multivariate* random variables

$$X = (X^{(1)}, X^{(2)}, \cdots, X^{(n)}), \tag{1.42}$$

taking values in (say) $\mathbf{R}^n$. In this case, and for $t = (t_1, \cdots, t_n) \in \mathbf{R}^n$ (considered as a column vector), one defines the characteristic function of X (in terms of t) to be given by

$$\boxed{\phi_X(t) = \mathbb{E}[e^{i(X^{(1)}t_1 + \cdots + X^{(n)}t_n)}] = \mathbb{E}[e^{itX}] = \mathbb{E}[e^{i<X,t>}] = \int_{\Omega} e^{i<X(\omega),t>} dP(\omega)} \tag{1.43}$$

where T denotes transposition, and $< X, t >= \sum_{j=1}^{n} X^{(j)} t_j$. Similarly one can define characteristic functions of complex-valued random variables, or multivariate complex random variables (i,e, with values in $\mathbf{C}^n$), by replacing (above) the exponent itX by $i\mathrm{Re}(\bar{t}^T X)$, where the bar over t indicates complex conjugation.

Note, for example for real-valued random variables X, that the characteristic function $\phi_X(t)$ exists $\forall t \in \mathbf{R}$, since

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i\sin(tX)] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)], \quad (1.44)$$

where the last two expected values are well defined, given that the $\sin(y)$ and $\cos(y)$ functions are bounded in $[-1, 1]$.

Characteristic Functions of Stochastic Processes

For a stochastic process $\mathbf{X} = \{\mathbf{X_t}\}_{t\in \mathrm{T}}$, one can either define the "characteristic function" of $\mathbf{X}$ for a given t, in which case it would just be the characteristic function of the random variable X_t (as above), or one can define a characteristic function of $\mathbf{X}$ (as an entire process).

♣ CHARACTERISTIC FUNCTIONS OF PROCESSES: INDEX-SET DEPENDENT. Consider a (multivariate with values in $\mathbf{R}^n$) stochastic process $\mathbf{X} = \{X_s\}$, with (say) $s \in \mathbf{R}^m$, defined over a probability space (Ω, σ, P), i.e.

$$\mathbf{X} : \mathbf{R}^m \times \Omega \longrightarrow \mathbf{R}^n. \quad (1.45)$$

First, one can define characteristic functions of such processes, in terms of the characteristic functions of the random variables that make the process, i.e. *in terms of* $s \in \mathbf{R}^m$, and for any $t \in \mathbf{R}^n$, as was done above, i.e. by

$$\boxed{\phi_{\mathbf{X}}(t)|_s = \phi_{X_s}(t) = \mathbb{E}\left[e^{i<X_s,t>}\right] = \int_{\Omega} e^{i<X_s(\omega),t>} dP(\omega)} \quad (1.46)$$

♣ CHARACTERISTIC FUNCTIONS OF PROCESSES: INDEX-SET INDEPENDENT. On the other hand, the characteristic function for such a process $\mathbf{X}$ (not in terms of the characteristic functions of the individual random variables X_s that make $\mathbf{X}$) can be defined as follows: For every $\mathbf{R}^n-$valued function $t(s)$ i.e.

$$t : \mathbf{R}^m \longrightarrow \mathbf{R}^n, \quad (1.47)$$

such that

$$\int_{\mathbf{R}^m} < X_s(\omega), t(s) > ds \tag{1.48}$$

converges for almost all $\omega \in \Omega$, one defines the characteristic function of $\mathbf{X}$ (in terms of $t(s)$) to be given by

$$\boxed{\phi_{\mathbf{X}}(t) = \mathbb{E}\left[e^{i\int_{\mathbf{R}^m} < X_s, t(s) > ds}\right] = \int_{\Omega} e^{i\int_{\mathbf{R}^m} < X_s(\omega), t(s) > ds} dP(\omega)} \tag{1.49}$$

Simple Examples on Characteristic Functions of Random Variables

1. If X is a discrete random variable with support $\mathbf{R}_X$ and probability (mass) function $p_X(x)$,then its characteristic function is:

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \sum_{x \in \mathbf{R}_X} e^{itx} p_X(x). \tag{1.50}$$

   Thus, all we need to do, is to sum the complex numbers $e^{itx}p_X(x)$ over all values of x belonging to the support of X.

2. If X is an *absolutely continuous* random variable (on $\mathbf{R}$, where the sum of the absolute differences of the values associated with any finite division of any given interval is bounded by a finite constant dependent only on the given interval) with probability density function $f_X(x)$,then its characteristic function is:

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{+\infty} e^{itx} f_X(x) dx \tag{1.51}$$

   which is a contour integral of a complex function along the real axis.
   Thus

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{+\infty} \cos(tx) f_X(x) dx + i \int_{-\infty}^{+\infty} \sin(tx) f_X(x) dx. \tag{1.52}$$

   In this case, the characteristic function is also known as the *Fourier transform* of the function $f_X(x)$.

Basic Properties of the Characteristic Function (for real-valued X)

1. $\phi_X(t)$ exists for any ditribution X, even when the probability density function and the moment generating function do not exist.
2. $\phi_X(0) = 1$, with $\phi_X(t)$ non-vanishing (i.e. not 0) in a neighbourhood of 0.
3. $|\phi_X(t)| \leq 1 \quad \forall t \in \mathbf{R}$.
4. $\phi_X(t)$ is a uniformly continuous function (in the sense that its $(\epsilon\delta)-$continuity has ϵ dependent only on δ and not on t).
5. The characteristic function of $a + bX$ is $e^{iat}\phi_X(bt)$.

1.2 Important Types of Stochastic Processes

In this section we give some important and usefull examples of stochastic processes, that will be needed below, such as Brownian Motion, the Weiner Process, Markov Processes, Lévy processes, and diffusions.

1.2.1 Brownian Motion

"Brownian motion" was first physically observed in the 19^{th} century (by Brown) as the random motion of tiny paricles in fluids, as a result of random bombardement with fluid molecules. This kind of behaviour was then described as a stochastic process that has the following specifications.

Definition 1.2.1. *Let $(\Omega, \sigma, \mathbf{P})$ be a probability space. An $\mathbf{R}^d$-valued process*

$$\boxed{\mathbf{B} = \{B_t\}_{t\geq 0} : \mathbf{R}_{\geq 0} \times \Omega \longrightarrow \mathbf{R}^d} \tag{1.53}$$

is said to be a **Brownian motion** *if it satisfies the following properties:*

1. *$B_0 = 0$, i.e. $B_0(\omega) = 0$ for every $\omega \in \Omega$.*
2. *All sample functions of $\mathbf{B}$ are continuous, i.e. $B_t(\omega)$ is a continuous function of $t \in \mathbf{R}_{\geq 0}$ for every given $\omega \in \Omega$.*

3. *For every* $0 \leq u \leq t < s$, *one has that* $B_s - B_t$ *is independent of* B_u, *and has Gaussian distribution with mean* 0 *and variance* $s - t$, *i.e.*

$$B_s - B_t \approx N(0, s - t). \tag{1.54}$$

We note the following known facts for Brownian motion:

1. Every Brownian motion is a Martingale. In fact for a Brownian motion $\mathbf{B} = \{B_t\}_{t \geq 0}$ one has that $B_t^2 - t$ is a martingale, i.e.

$$\mathbb{E}[B_t^2 - t | \mathscr{B}_s] = B_s^2 - s, \tag{1.55}$$

   where $\mathscr{B}_s$ is the σ-algebra on Ω associated with B_s (i.e. the smallest σ-algebra on Ω to make B_s a measurable function).

2. Every Brownian motion is a Guassian process.

3. Every Brownian motion is a Markov process.

4. Every Brownian motion is a diffusion process.

5. Every Brownian motion is a Lévy process.

We also note that almost every interesting class of processes contains a Brownian motion and that Brownian motion can be used as a building block for very general processes (in a certain class of processes) by using certain sequences of transformation.

1.2.2 Markov Processes

In this section we give the general definition of Markov Process and Transition Function. This section is restrict to the case where $\mathrm{T} = \mathbf{R}_{\geq 0}$.

Markov Processes

A Markov process can be thought of as a "memoryless" process, in the sense that one can make predictions for the future of the process based slowely on its present state.

♣ DEFINITION OF A MARKOV PROCESS. The following definition gives the most general formulation of a Markov process.

Definition 1.2.2. *A Markov Process, with state space* $(\mathbf{E}, \mathscr{E})$*, is a collection* $\mathcal{X}$ *given by*

$$\boxed{\mathcal{X} = (\Omega, \sigma, \mathbf{P}, \{\mathscr{F}_t\}, \{X_t\}, \{P_t\}, \{\mathbb{P}^x : x \in \mathbf{E}\})_{t \in \mathbf{R}_{\geq 0}}} \tag{1.56}$$

where
1) (Ω, σ, P) *is a probability space,*

2) $\{\mathscr{F}_t\}$ *is a filtration of* $(\Omega, \sigma, \mathbf{P})$*,*

3) $\mathbf{X} = \{X_t\}_{t \in \mathrm{T}}$ *is an* $(\mathbf{E}, \mathscr{E})$*-valued stochastic process, with the family* $\{X_t\}$ *adapted to the filtration* $\{\mathscr{F}_t\}$*, i.e.*$\forall t \in \mathrm{T}$*,* X_t *is measurable with respect to* $\mathscr{F}_t$ *(and hence* X_t *is measurable with respect to* $\mathscr{F}_{t'}$ *for every* $t' > t$*),*

4) $\mathbb{P}^x$ *is the law of the stochastic process* $\mathbf{X}$ *started at* $x \in \mathbf{E}$*,*

5) $\{P_t\}$ *define, what are called as, the* **(probability) transition functions** *on* $(\mathbf{E}, \mathscr{E})$*, that satisfy the following:*

- *For every* $t \geq 0$ *one has that*

$$P_t : \mathbf{E} \times \mathscr{E} \longrightarrow [0, 1], \tag{1.57}$$

such that for any $x \in \mathbf{E}$*,*

$$P_t(x, -) : \mathscr{E} \longrightarrow [0, 1] \tag{1.58}$$

is a probability measure on $\mathscr{E}$*, i.e. with* $P_t(x, \mathbf{E}) = 1$*, and for any* $\Gamma \in \mathscr{E}$*,*

$$P_t(-, \Gamma) : \mathbf{E} \longrightarrow [0, 1] \tag{1.59}$$

is $\mathscr{E}$-measurable.

- *for $s, t \geq 0$, $x \in \mathbf{E}$ and $\Gamma \in \mathscr{E}$, P_s, P_t and P_{s+t} satisfy the* **Chapman-Kolmogorov equation** *given by*

$$\boxed{P_{t+s}(x,\Gamma) = \int_{\mathbf{E}} P_t(-,\Gamma) dP_s(x,-)} \quad \left(= \int_{\mathbf{E}} P_t(y,\Gamma) P_s(x,dy) \right) \tag{1.60}$$

i.e. that $P_{t+s}(x,\Gamma)$ is the integral, over $\mathbf{E}$, of the function $P_t(-,\Gamma)$ with respect to the measure $P_s(x,-)$.

It is the Chapman-Kolmogorov equation above that gives the process $\mathbf{X}$ its "memoryless" property, in the sense that the transition function P_{t+s} for time $t+s$ depends on a past (any past) transition function P_t at time t, and on the transition function P_s associated with the difference in times between $t+s$ and t.

Note: That sometimes we will refer to $\mathbf{X}$ (as above, which is part of a Markov process $\mathscr{X}$), as being *the Markov process*. Thus the "Markov process $\mathbf{X}$" will mean that $\mathbf{X}$ is the stochastic-process part of a Markov process $\mathscr{X}$ that includes all the elements given above (for a Markov process).

♣ THE TRANSITION FUNCTIONS AS A SEMI-GROUP. One can show that the Chapman-Kolmogorov equation becomes a **semi-group** equation

$$\boxed{P_{s+t} = P_s P_t} \tag{1.61}$$

if one regards every P_t as acting on the bounded $\mathscr{E}-$measurable functions $f : \mathbf{E} \longrightarrow \mathbf{R}$ where the action is given by

$$(P_t f)(x) = \int_{\mathbf{E}} f(-) dP_t(x,-), \qquad \forall x \in \mathbf{E}, \tag{1.62}$$

Note that the resulting $(P_t f)$ is indeed a function $(P_t f) : \mathbf{E} \longrightarrow \mathbf{R}$, on which another P_s can act, to establish the semi-group property. Thus one would have that

$$(P_{s+t} f)(x) = (P_s P_t f)(x), \tag{1.63}$$

and one would transform the family $\{P_t\}$ to a semi-group.

Discret-Time Markov Processes

For a discrete-time process $\mathbf{X}$ (where $\mathrm{T} = \mathbf{N}$), $\mathbf{X}$ would be a Markov process (i.e. the process part of a Markov process $\mathcal{X}$) if it satisfies (as one can show, given the above general definition for a Markov process) the property that

$$\begin{aligned} Pr(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2} \cdots, X_1 = x_1) \\ = Pr(X_n = x_n | X_{n-1} = x_{n-1}) \end{aligned} \tag{1.64}$$

indicating that the values at time n depend only on the values at time $n-1$.

Time-Homogeneous Markov Processes

A **time homogeneous Markov process** is a Markov process such that the transition probability functions depend only on the *difference* s between $t+s$ and t and not on the actual times t and $t+s$.

Thus for a time-homogeneous discrete-time Markov process one would have that

$$Pr(X_{n+1} = x | X_n = y) = Pr(X_n = x | X_{n-1} = y), \tag{1.65}$$

for every n. While for a time-homogeneous continuous-time Markov process one would have that

$$Pr(X_{t+s} = x | X_t = y) = Pr(X_s = x | X_0 = y). \tag{1.66}$$

1.2.3 Lévy Processes

In this section we define Lévy processes, which provide important examples on Markov processes.

Lévy Processes

Definition 1.2.3. *Let $(\Omega, \sigma, \mathbf{P})$ be a probability triple. An $\mathbf{R}^d$-valued process $\mathbf{X} = \{X_t\}_{t\geq 0}$ is said to be a Lévy processes (or a process with stationary independent increments) if it satisfies the following:*

1. *For almost every $\omega \in \Omega$, the function $X_t(\omega)$ (as a function of t) is right continuous on $\mathbf{R}_{\geq 0}$, with left limits on $\mathbf{R}_{>0}$.*
2. *For every $0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$, the random variables $Y_j = X_{t_j} - X_{t_{j-1}}$, $j = 1, \cdots, n$, are independent random variables*
3. *The law (or distribution) of $X_s - X_t$ depends only on $t - s$.*

♣ THE CENTRAL RESULT ON LÉVY PROCESSES. The following is the central result in the theory of Lévy processes, and it concerns the characteristic function of such a process.

Lévy-Khintchine representation: An $\mathbf{R}^d-$valued lévy process $\mathbf{X} = \{X_t\}_{t\geq 0}$ has a characteristic function defined by

$$\phi_X(\theta)|_{t=1} = \phi_{X_1}(\theta), \tag{1.67}$$

and this has the form (for any $\theta \in \mathbf{R}^d$) given by

$$\boxed{\phi_{X_1}(\theta) = \exp\left(i\alpha^T\theta - \tfrac{1}{2}\theta^T M\theta + \textstyle\int_{\mathbf{R}^d}(e^{i\theta^T x} - 1 - i\theta^T x I_{|x|<1})\nu(dx)\right)} \tag{1.68}$$

where $\alpha \in \mathbf{R}^d$, T denotes transposition, M is a non-negative definite symmetric $d \times d$ matrix, and $I_{|x|<1}$ is the indicator function for the domain $|x| < 1$ in $\mathbf{R}^d$. While ν is a certain measure called the **lévy measure** of the process X, and satisfies

$$\int (|x|^2 \wedge 1)\nu(dx) < \infty. \tag{1.69}$$

Infinitely Divisible Measures and Distributions

It is known that the theory of analytic semigroups associated with Lévy processes is equivalent to the theory of infinitely divisible distributions. We first define infinitely divisible measures.

♣ INFINITELY DIVISIBLE MEASURES. A probability measure μ on $\mathbf{R}^d$ is said to be an **infinitely divisible measure** if for every integer n there exists a probability measure μ_n on $\mathbf{R}^d$ such that if $W_1, \cdots, W_n$ are independent $\mathbf{R}^d$ random variables on a probability space$(\Omega, \sigma, \mathbf{P})$ with identical laws μ_n, then the law of $W_1 + \cdots + W_n$ is given by μ, i.e, $W_1 + \cdots + W_n$ is equal, *in distribution*, to μ. This is written as

$$W_1 + \cdots + W_n \stackrel{\mathcal{D}}{=} \mu. \tag{1.70}$$

♣ INFINITELY DIVISIBLE DISTRIBUTIONS. In a similar way, one says that a distribution F is an **infinitely divisible distribution** if and only if for every integer n there exist n independent and identically distributed random variable $X_1, \cdots, X_n$ such that $X_1 + \cdots + X_n$ has distribution F. In this case we write that

$$X_1 + \cdots + X_n \stackrel{\mathcal{D}}{=} F. \tag{1.71}$$

And if X is a random variable with distribution F, we write that

$$X_1 + \cdots + X_n \stackrel{\mathcal{D}}{=} X. \tag{1.72}$$

If F is an infinitely divisible distribution, with $X = X_1 + \cdots + X_n$ having distribution F, where each X_i has distribution $\bar{F}$, then the characteristic function of X is given by

$$\phi_X(t) = [\phi_{X_i}(t)]^n. \tag{1.73}$$

1.2.4 Diffusions

Diffusions in stochastic processes provide models for what one can *physically* understand by the term "diffusion". One readily available example of a stochastic

diffusion process is Brownian motion itself (as defined in section (1.2.1)). It is worth mentioning that at the heart of stochastic (and physical) diffusions is what is known as the **diffusion equation** of mathematical physics.

♣ THE DIFFUSION EQUATION. To briefly describe the diffusion equation, one can $\rho(t, \vec{x})$ be the concentration, as a function of time t and position $\vec{x}$, of a substance (whether of a physical or of any other nature) that can "diffuse" from one point to another. Then, with a "flux-vector" given by

$$\vec{F}(t, \vec{x}) = -\frac{1}{2} a \vec{\nabla} \rho(t, \vec{x}), \tag{1.74}$$

where a describes a diffusivity constant, which may even be a variable depending on position, or even a matrix depending on position and dictating different diffusivity properties along different directions at the same point in space, the diffusion equation for this substance is given by

$$\begin{aligned} \frac{\partial \rho}{\partial t}(t, \vec{x}) &= -\vec{\nabla} \cdot (F(t, \vec{x})) \\ &= \frac{1}{2} \vec{\nabla} \cdot (a \vec{\nabla}(t, \vec{x})). \end{aligned} \tag{1.75}$$

♣ PROBABILISTIC INTERPRETATIONS & SEMI-GROUPS FOR THE CASE $a = 1$. The special case where $a = 1$ establishes the **Kolmogorov forward equation** for the **Brownian transition density**. In such cases, a probabilistic interpretation of the diffusion equation involves a diffusion whose infinitesimal generator $\mathscr{G}$ has an adjoint given by

$$\mathscr{G}^* = \frac{1}{2} \vec{\nabla} \cdot (a \vec{\nabla}), \tag{1.76}$$

in the sense that there exists a *transition semigroup*

$$\{P_t\}_{t \geq 0}, \tag{1.77}$$

such that

$$\mathscr{G} f = \lim_{t \downarrow o} t^{-1} (P_t f - f), \tag{1.78}$$

for some class of functions f, where $\mathscr{G}$ and $\mathscr{G}^*$ are related by

$$\int f\mathscr{G}g = \int g\mathscr{G}^* f, \tag{1.79}$$

for every $f, g \in C_K^\infty$ (where the K indicates "with compact support").

♣ PHYSICAL ASPECTS OF DIFFUSION & THE ORNSTEIN-UHLENBECK SDE. If one is discussing physical aspects of diffusion associated with a "particle", then one can model the velocity v_t of this particle as a Brownian motion, or by modelling it through the **Ornstein-Uhlenbeck** stochastic differential equation

$$dv_t = dB_t - \lambda v_t dt, \tag{1.80}$$

involving Brownian motion B on $\mathbf{R}$, where $\lambda > 0$ is the viscous drag coefficient, where the SDE means (see a detailed discussion in section 1.3) that

$$\int_0^s dB_t \text{ (Itô integral)} - \int_0^s \lambda v_t dt \text{ (usual integral)} = v_s - v_0. \tag{1.81}$$

This SDE would have an explicit solution given by the stochastic integral

$$v_t = v_0 e^{-\lambda t} + e^{-\lambda t} \int_0^t e^{\lambda s} dB_s \text{ (Itô integral)}, \tag{1.82}$$

where

$$\int_0^t e^{\lambda s} dB_s \text{ (Itô integral)} := e^{\lambda t} B_t - \lambda \int_0^t e^{\lambda s} B_s ds. \tag{1.83}$$

1.3 Basic Elements of Itô's Calculus & Stochastic Differential Equations

In this section we just introduce some basic elements of stochastic calculus, and of stochastic differential eqution (SDE's), that will be needed later on.

1.3.1 Stochastic Calculus versus the Usual Calculus

It is essential that one first highlights the difference between *stochastic calculus* and the *usual calculus*, in terms of the differences in the fundamental operations of integration and differentiation that occur in these two forms of calculus.

On Stochastic Integrals and Stochastic Integration

In what is known as "stochastic calculus", where integrating (smooth) functions $f(X_t, t)$ of stochastic processes $\mathbf{X}$ and time t, where

$$\mathbf{X} : \mathbf{R} \times [0, \infty) \longrightarrow \mathbf{R}, \tag{1.84}$$

and in particular where the process $\mathbf{X}$ is in fact just a function of another underlying process such as Brownian motion $\mathbf{B} : \Omega \times [0, \infty) \longrightarrow \mathbf{R}$ as well as on time, is a main objective, the specification of *how* (as will be briefly discussed below) to do a definite integration of the form

$$\boxed{I = \int_a^b f(X_t(B_t), t)\, dB_t} \tag{1.85}$$

associated with every $\omega \in \Omega$, i.e. where for every $\omega \in \Omega$ one has

$$I(w) = \int_a^b f(X_t(B_t(\omega)), t)\, dB_t(\omega), \tag{1.86}$$

between times a and b ($0 \leq a \leq b$), of the function f over the stochastic paths $B_t(\omega)$, $t \geq 0$, associated with the underlying process $\mathbf{B}$ (which could be Brownian motion), makes a great difference over the end result of this integration.

If one thinks in terms of integrations, as inspired by the definition of a Riemann integral, where one considers partitions $[t_0, t_1, \cdots, t_n]$ of the interval $[a, b]$ and considers a value (at $\omega \in \Omega$), associated with this partition, of the form

$$\Psi_n = \sum_{i=0}^{n-1} \mathcal{H}_i(f)(B_{t_{i+1}} - B_{t_i}), \tag{1.87}$$

i.e.

$$\Psi_n(\omega) = \sum_{i=0}^{n-1} \mathcal{H}_i(f,\omega)(B_{t_{i+1}}(\omega) - B_{t_i}(\omega)), \tag{1.88}$$

where $\mathcal{H}_i(f,\omega)$ depends on the values of f in the subinterval $[t_i, t_{i+1}]$, then the limit

$$\lim_{n\to\infty} \Psi_n, \tag{1.89}$$

assuming it exists (at ω) for a given form for $\mathcal{H}_i(f)$, depends greatly on this form of $\mathcal{H}_i(f)$, when one deals with functions of stochastic processes, and stochastic paths.

In contrast, if one considers Riemann integrals of (e.g. smooth) functions f dependent on "reasonable" functions X, this limit of Ψ_n will not depend on the choices of $\mathcal{H}_i(f)$ if that choice varies (for example) from

$$\mathcal{H}_i(f)^{(1)} = f(X_{t_i}(B_{t_i}), t_i), \tag{1.90}$$

to

$$\mathcal{H}_i(f)^{(2)} = \frac{f(X_{t_{i+1}}(B_{t_{i+1}}), t_{i+1}) + f(X_{t_i}(B_{t_i}), t_i)}{2}, \tag{1.91}$$

to

$$\mathcal{H}_i(f)^{(3)} = f(X_{t_{i+1}}(B_{t_{i+1}}), t_{i+1}) \tag{1.92}$$

to

$$\mathcal{H}_i(f)^{(4)} = f(X_{(t_i+t_{i+1})/2}(B_{(t_i+t_{i+1})/2}), (t_i + t_{i+1})/2), \tag{1.93}$$

etc.... This exactly is what establishes a "Riemann integral".

However, the choice of the particular function $\mathcal{H}(f)$ (as in the cases above, for example) greatly affects the limit $\lim_{n\to\infty} \Psi_n$, for (even smooth) functions f dependent on stochastic processes $\mathbf{X}$ with underlying processes $\mathbf{B}$, as the stochastic paths of $\mathbf{B}$ (e.g. when $\mathbf{B}$ is Brownian motion) can have infinite length between any two finite times a and b, and they can be differentiable almost nowhere even though continuous. Thus one might find out that the two limits

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \mathcal{H}_i(f)^{(1)}(B_{t_{i+1}} - B_{t_i}) \qquad \text{and} \qquad \lim_{n\to\infty} \sum_{i=0}^{n-1} \mathcal{H}_i(f)^{(2)}(B_{t_{i+1}} - B_{t_i}) \tag{1.94}$$

might both exist but have different values, at a given $\omega \in \Omega$.

These facts dictate that the resulting values for a definite integral involving a stochastic process, as in (1.94) above, are **very sensitive** on how one defines an integral, i.e. **on the particular choice of** $\mathcal{H}_i(f)$, if one intends a definition that resembles the idea of Riemann integration.

On Stochastic Differentials and Stochastic Differentiation

What is more, is that the *differential* of a function $f(X(B,t),t)$, dependent on a stochastic process $\mathbf{X}$ and time t, where $\mathbf{X}$ itself depends on some underlying process $\mathbf{B}$, i.e. a (differential) process of the form

$$df(X_t(B_t),t) = \rho(X_t(B_t),t)dt + \sigma(X_t(B_t),t)dB_t, \tag{1.95}$$

at $\omega \in \Omega$, i.e.

$$df(X_t(B_t),t)|_\omega = \rho(X_t(B_t),t)|_\omega dt + \sigma(X_t(B_t),t)|_\omega dB_t(\omega), \tag{1.96}$$

or

$$df(X_t(B_t(\omega)),t) = \rho(X_t(B_t(\omega)),t)dt + \sigma(X_t(B_t(\omega)),t)dB_t(\omega), \tag{1.97}$$

depends on how one has defined integration (as discussed above) so that one, in an inverse process, obtains

$$\int_a^b [\rho(X_t(B_t,t))dt + \sigma(X_t(B_t),t)dB_t] = f(X_b(B_b),b) - f(X_a(B_a),a), \tag{1.98}$$

at almost every $\omega \in \Omega$. Thus, for example, if for two different functions $\sigma_1(X_t(B_t),t)$ and $\sigma_2(X_t(B_t),t)$, and for two different "definitions" for integration (depending on different $\mathcal{H}(\sigma)$ as discussed above) one has that

$$\left[\int_a^b \sigma_1(X_t(B_t),t)dB(t)\right]^{(1)} = R(X_b(B_b),b) - R(X_a(B_a),a), \tag{1.99}$$

for almost every $\omega \in \Omega$, while

$$\left[\int_a^b \sigma_2(X_t(B_t),t)dB(t)\right]^{(2)} = R(X_b(B_b),b) - R(X_a(B_a),a), \qquad (1.100)$$

for almost every $\omega \in \Omega$, i.e. if both integrations, by using different definitions for integration and the two different functions, give the same result (for example) almost everywhere in Ω, then one would have, *for the same function* $R(X_t(B_t),t)$, that

$$\left[\frac{\partial R(X_t(B_t),t)}{\partial B}\right]^{(1)} = \sigma_1(X_t(B_t),t), \qquad (1.101)$$

if one has defined integration by the first method, while

$$\left[\frac{\partial R(X_t(B_t),t)}{\partial B}\right]^{(2)} = \sigma_2(X_t(B_t),t), \qquad (1.102)$$

if one has defined integration by the second method.

Thus differentiation (or partial differentiation) involving stochastic processes depends on how one defines integration, and thus **has no absolute meaning**, as in the "usual calculus", and thus **differentiation is only with respect to the definition of integration**. This is all about

"Stochastic Calculus" versus the "Usual Calculus"

1.3.2 Basic Elements of Itô's Calculus

Defining Itô Integration

Itô's calculus, which is the main formulation of stochastic calculus, is essentially based on the form of integration given by the definition of this process (as above) using

$$\mathcal{H}_i(f) = f(X_{t_i}(B_{t_i}),t_i), \qquad (1.103)$$

at every $\omega \in \Omega$ i.e. where the integrals (in this calculus) are defined by

$$\boxed{\int_0^b f(X_t(B_t),t)dB_t = \lim_{n\to\infty}\left[\sum_{i=0}^{n-1} f(X_{t_i}(B_{t_i}),t_i)[B_{t_{i+1}} - B_{t_i}]\right]} \tag{1.104}$$

at every $\omega \in \Omega$. This is the starting point in what is known as **Itô's calculus**

But to have meaningful results within this formulation, one would have to assume the following.

Basic Assumptions to Have Itô Integrals

1. There is given a probability space $(\Omega, \mathscr{F}, \mathbf{P})$, with $\mathbf{B}$ standard Brownian motion on Ω where $\mathbf{B} = B_-(-) : \Omega \times [0,\infty) \longrightarrow \mathbf{R}$, and one has the filtration $\{\mathscr{F}_t\}$, $t \in [0,\infty)$, of $(\Omega, \mathscr{F}, \mathbf{P})$ where $\mathscr{F}_t$ is the sigma algebra generated by the collection of random variables $\{B_s : s \leq t\}$.

2. There is given a function

$$F : \Omega \times [0,\infty) \longrightarrow \mathbf{R} \tag{1.105}$$

that is *adapted* to the above Brownian motion $\mathbf{B}$, in the sense that for every $t \in [0,\infty)$, one has that

$$F(-,t) : \Omega \longrightarrow \mathbf{R} \tag{1.106}$$

is measurable with respect to $\mathscr{F}_t$ (in the filtration $\{\mathscr{F}_t\}$ above generated by the Brownian motion).

In practice, and for the relevant cases, F will have the form

$$F(\omega,t) = f(X_t(B_t(\omega)),t), \tag{1.107}$$

for some two functions X and f, where X depends on Brownian motion $\mathbf{B}$ and time t, while f depends on the values of X and time t.

3. For almost every $\omega \in \Omega$, the sample paths $F(\omega, t)$, $t \geq 0$, are continuous paths, with the integral, from 0 to any time t, of the expectations of the random variables $F^2(-, s) : \Omega \longrightarrow \mathbf{R}$ being finite, i.e

$$\int_0^t \mathbf{E}\left[F^2(-, s)\right] ds = \int_0^t \mathbf{E}\left[f^2(X_s(B_s(-)), t)\right] ds < \infty. \tag{1.108}$$

Within these assumptions, one defines the Itô integral of F with respect to Brownian motion $\mathbf{B}$, at the point $\omega \in \Omega$, to be given (as above) by

$$\begin{aligned} I^{(\text{itô})}(F, \mathbf{B}, \omega, b) &= \int_0^b F(\omega, t) dB_t(\omega) \\ &= \int_0^b f(X_t(B_t(\omega)), t) dB_t(\omega) \\ &= \lim_{n \to \infty} \left[\sum_{i=0}^{n-1} f(X_{t_i}(B_{t_i}(\omega)), t_i)[B_{t_{i+1}}(\omega) - B_{t_i}(\omega)] \right] \end{aligned} \tag{1.109}$$

Note that the above integral $I^{(\text{itô})}(F, \mathbf{B}, \omega)$, for a fixed $F, \mathbf{B}$ and $b \in [0, \infty)$, defines a random variable

$$I^{(\text{itô})}(F, \mathbf{B}, -, b) : \Omega \longrightarrow \mathbf{R}. \tag{1.110}$$

Consequences of Itô Integration and the Resulting Differential Calculus

Some consequences are the following (see [RW2] for proves).

1. Given a function F and Brownian motion $\mathbf{B}$ as above, one has that the expectation over Ω of the random variable $I^{(\text{itô})}(F, \mathbf{B}, -, b)$ is zero, i.e.

$$\mathbf{E}\left[I^{(\text{itô})}(F, \mathbf{B}, -, b)\right] = \mathbf{E}\left[\int_0^b F(-, t) dB_t(-)\right] = 0 \tag{1.111}$$

2. For a random variable $Z_0 : \Omega \longrightarrow \mathbf{R}$, one has that the process $\mathbf{Z} = \{Z_b\}$, $b \in [0, \infty)$, given by

$$Z_b = Z_0 + I^{(\text{itô})}(F, \mathbf{B}, -, b) = Z_0 + \int_0^b F(-, t) dB_t(-), \tag{1.112}$$

defines a martingale, i.e. where

$$\mathbf{E}[M_b | \mathscr{F}_s] = M_s, \tag{1.113}$$

where $0 \leq s < b$, and $\{\mathscr{F}_s\}$ is the filtration given above. In particular one has that

$$\mathbf{I}^{(\text{itô})}(F, \mathbf{B}) = \{I^{(\text{itô})}(F, \mathbf{B}, -, b)\} \tag{1.114}$$

where $b \in [0, \infty)$, defines a martingale.

3. **Itô's Formula**: Assume a function F as above, where

$$F(\omega, t) = f(X_t(B_t(\omega)), t), \tag{1.115}$$

and assume that $\mathbf{X}$ satisfies **the Itô Stochastic Differential Equation (SDE)**

$$\boxed{dX_t = \rho(X, t)dt + \sigma(X, t)dB_t} \tag{1.116}$$

in the sense that the usual (Riemann) integral from 0 to t of $\rho(X, t)dt$, and the Itô integral (as above) from 0 to t of $\sigma(X, t)dB_t$, give $X_t - X_0$ (after using s as a variable for integration instead of t). Then one obtains **Itô's Formula** given by

$$dF = \boxed{\left[\frac{\partial f}{\partial t} + \rho(X, t)\frac{\partial f}{\partial X}(X, t) + \frac{1}{2}\sigma^2(X, t)\frac{\partial^2 f}{\partial X^2}(X, t)\right] dt} + \\ + \boxed{\left[\sigma(X, t)\frac{\partial f}{\partial X}(X, t)\right] dB_t} \tag{1.117}$$

also in the sense that the usual (Riemann) integral from 0 to t of the first box, and the Itô integral from 0 to t of the second box, give $F(\omega, t) - F(\omega, 0)$

(after using s as a variable for integration instead of t).

For example, if $X = B$, i.e. $X_t = B_t$ for every $t \in [0, \infty)$, and thus indeed one has that $dX_t = dB_t$, with $\rho(X,t) = 0$ and $\sigma(X,t) = 1$, then one has, for $F = f(X_t(B_t), t) = X_t^2 = B_t^2$, that

$$dF = d(B_t^2) = dt + 2B_t dB_t, \tag{1.118}$$

where $\partial^2 f/\partial X^2 = 2$ and $\partial f/\partial X = X = B$. In the "usual calculus" one would have obtained $dF = 2B_t dB_t$, instead of the above.

1.3.3 Introducing Systems of Stochastic Differential Equations

In this section we introduce a system of differential equations, where a stochastic differential equation would be a special case where only one equation exists in the system. A **system of stochastic differential equations (SDE's)** for a vector-valued stochastic process

$$\vec{\mathbf{X}} = \{\vec{X}(t)\}_{t\geq 0} = \{(X_1(t), X_2(t), \cdots, X_n(t))\}_{t\geq 0} = \{(\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)\}, \tag{1.119}$$

with respect to n-dimensional Brownian motion

$$\vec{\mathbf{B}} = \{\vec{B}(t)\}_{t\geq 0} = \{(B_1(t), B_2(t), \cdots, B_n(t))\}_{t\geq 0} = \{(\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_n)\}, \tag{1.120}$$

where the $\mathbf{B}_i$'s are independent 1-dimensional Brownian motions such that

$$\mathbf{E}[B_i(t)B_j(s)] = \begin{cases} \min(t,s) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1.121}$$

is a differential expression of the form

$$\boxed{d\vec{\mathbf{X}} = \vec{b}(\vec{\mathbf{X}}, t)\, dt + \sigma(\vec{\mathbf{X}}, t)\, d\vec{\mathbf{B}}}. \tag{1.122}$$

The coefficient functions $\vec{b}$ and σ in the above differential expression are, respectively, a **drift vector**

$$\vec{b} = (b_1, b_2, \cdots, b_n) : \mathbf{R}^n \times [0, \infty) \longrightarrow \mathbf{R}^n, \tag{1.123}$$

i.e. where for every i

$$b_i(\vec{\mathbf{X}}, t) : \mathbf{R}^n \times [0, \infty) \longrightarrow \mathbf{R}, \tag{1.124}$$

and a **diffusion matrix**

$$\sigma = (\sigma_{ij}) : \mathbf{R}^n \times [0, \infty) \longrightarrow \mathbf{R}^{n \times n}, \tag{1.125}$$

i.e. where for every (i, j)

$$\sigma_{ij}(\vec{\mathbf{X}}, t) : \mathbf{R}^n \times [0, \infty) \longrightarrow \mathbf{R}, \tag{1.126}$$

which satisfy some smoothness conditions.

The differential expression above is to be understood in terms of the integral expression

$$\vec{X}(t) = \vec{X}_0 + \int_0^t \vec{b}(\vec{X}(s), s)ds + \int_0^t \sigma(\vec{X}(s), s)d\vec{B}(s), \tag{1.127}$$

where the second integral is understood to be an Itô integral. In component form the above integral can be given as

$$\begin{aligned} X_i(t) = X_{i0} \;&+\; \int_0^t b_i(X_1(s), \cdots, X_n(s), s)ds \\ &+\; \sum_{j=1}^{n} \int_0^t \sigma_{ij}(X_1(s), \cdots, X_n(s), s)dBJ(s), \end{aligned} \tag{1.128}$$

for $1 \leq i \leq n$.

At this point we note that if

$$f : \mathbf{R}^n \times [0, \infty) \longrightarrow \mathbf{R} \tag{1.129}$$

is a smooth function $f(X_1, \cdots, X_n, t)$ and

$$Y(t) = f(X_1(t), \cdots, X_n(t), t), \tag{1.130}$$

where $\vec{X}(t) = (X_1(t), \cdots, X_n(t))$ is a solution of the above system of stochastic differential equations then Itô formula becomes

$$\boxed{dY = \left(\frac{\partial f}{\partial X_i} + \sum_{i=1}^{n} b_i \frac{\partial f}{\partial X_i} + \frac{1}{2} \sum_{i,j,k=1}^{n} \sigma_{ik}\sigma_{jk} \frac{\partial^2 f}{\partial X_i \partial X_j} \right) dt + \sum_{i,j=1}^{n} \sigma_{ij} \frac{\partial f}{\partial X_i} dB_j}, \tag{1.131}$$

Which follows from the "Itô rule" that

$$dB_i dB_j \begin{cases} dt & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1.132}$$

1.4 Some Basic Results from Measure Theory and General Analysis

1.4.1 Some Basic Results from Measure Theory

Lebesgue Dominated Convergence Theorem

Theorem 1.4.1. *Let (Ω, σ, μ) be a meaure space and assume that $f_n : \Omega \longrightarrow \mathbf{R}$, $n = 1, 2, \cdots$, and $f : \Omega \longrightarrow \mathbf{R}$, are measurable functions with $\{f_n\} \longrightarrow f$ as $n \longrightarrow \infty$, pointwise almost everywhere in ω. If the sequence $\{|f_n|\}$ is dominated above by a positive integrable $g \in \mathcal{L}^1(\Omega, \sigma, \mu)$ then*

$$f_n \longrightarrow f \quad \text{in } \mathcal{L}^1(\Omega, \sigma, \mu) \tag{1.133}$$

i.e.

$$\int_\Omega |f_n - f| d\mu \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty, \tag{1.134}$$

and hence that

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu. \tag{1.135}$$

1.4.2 Some Basic Results from Gerenal Analysis

Reimann-Lebesgue Lemma

Let f be a function in $\mathcal{L}^1(\mathbf{R}^n)$. The **Fourier Transform** of f, denoted by $\mathcal{F}(f)$, is defined as the function on $\mathbf{R}^n$ given by

$$\begin{aligned} \mathcal{F}(f)(k) &= \int_{\mathbf{R}^n} e^{-2\pi i <k,x>} f(x) dx \\ &= \int_{\mathbf{R}^n} e^{-2\pi i \sum_{j=1}^n (k_j x_j)} f(x) dx. \end{aligned} \tag{1.136}$$

We note that the Fourier transform is a linear operator on f that satisfies the following simple facts:

1. $\mathcal{F}(f) \in \mathcal{L}^\infty(\mathbf{R}^n)$
2. $||f||_\infty \leq ||f||_1$
3. $\mathcal{F}(f)$ is a continuous, and hence measurable, function.

The following is the Riemann-Lebesgue Lemma.

Lemma 1.4.1. *For $f \in \mathcal{L}^1(\mathbf{R}^n)$ one has that*

$$\mathcal{F}(f)(k) \longrightarrow 0 \qquad \text{as } |k| \longrightarrow \infty. \tag{1.137}$$

Chapter 2

Affine Processes

It is thought that the following quote will apply at all times, not only at a specific time !

> **"One wonders if the present theory of stochastic processes is not still too difficult for applications"**

K.L.Chung

In this chapter we discuss the basic elements of the theory of affine processes. While the next chapter will consider some applications of affine processes to financial mathematics.

2.1 Affine Process and Basic Properties

Roughly speacking, an Affine Process can be described as a Markov Process

$$\mathbf{X} = \{\Omega, \sigma, \mathbf{P}, \{\mathscr{F}_t\}, \{X_t\}, \{P_t\}, \{\mathbb{P}^x : x \in \mathbf{E}\}\}_{t \in \mathbf{R}_{\geq 0}} \tag{2.1}$$

whose Log-charachteristic function is an affine function of its initial state vector x in its state space $\mathbf{E}$. It is also a stochastically continuous time homogeneous markov process where the state space $\mathbf{E}$ is given by the Borel-measurable space

$$\mathbf{E} = \mathbf{R}^{\mathrm{m}}_{\geq 0} \times \mathbf{R}^{\mathrm{n}}. \tag{2.2}$$

We first discuss the characteristic function of an affine process.

2.1.1 Characteristic Function of an Affine Process

Briefly the characteristic function of an affine process is an *exponentially affine-function of the starting state vector.*

Consider a starting state vector

$$x \in \mathbf{E} = \mathbf{R}^{\mathrm{m}}_{\geq 0} \times \mathbf{R}^{\mathrm{n}} \subset \mathbf{R}^d \quad (\text{where } d = m + n). \tag{2.3}$$

For this given x and any $t \in \mathbf{R}_{\geq 0}$, the values of a random variable X_t in $\mathbf{E}$ have a certain distribution (depending on x) as determined by the law $\mathbb{P}^x$ (on the measurable space (Ω, σ)) of the process started at x. The characteristic function for the process $\mathbf{X} = \{X_t\}$, that is *started at* x, and as a function of t, will be defined by

$$\boxed{\phi^x_{\mathbf{X}}(u)|_t = \phi^x_{X_t}(u)} \tag{2.4}$$

which is also a function of

$$x \in \mathbf{E} \quad \text{and} \quad u \in i\mathbf{R}^d, \tag{2.5}$$

and is given by

$$\boxed{\phi^x_{\mathbf{X}}(u)|_t = \mathbb{E}^x[e^{<X_t, u>}] = \mathbb{E}^x[e^{X_{t,1}u_1 + \cdots + X_{t,d}u_d}] = \mathbb{E}^x[e^{i(X_{t,1}\hat{u}_1 + \cdots + X_{t,d}\hat{u}_d)}]} \tag{2.6}$$

where

$$u = (u_1, \cdots, u_d) = i(\hat{u}_1, \cdots, \hat{u}_d), \tag{2.7}$$

with $(\hat{u}_1, \cdots, \hat{u}_d) \in \mathbf{R}^d$, and the expectation $\mathbb{E}^x$ is with respect to $\mathbb{P}^x$, and where for vectors x, y in $\mathbf{R}^d$ or $\mathbf{C}^d$ we have used that

$$\boxed{< x, y >= \sum_{i=1}^{d} x_i y_i} \tag{2.8}$$

(with no complex conjugation used in $\sum_{i=1}^{d} x_i y_i$ for the case of $\mathbf{C}^d$).

Note: As will be discussed below, one could have used u in a larger set in $\mathbf{C}^d$ other than $i\mathbf{R}^d$. In particular u could have been in the following set $\mathcal{U}$, which will be used below, and which has a special role:

$$\boxed{\mathcal{U} = \{u \in \mathbf{C}^d : \mathrm{Re}(u_i) \leq 0, \mathrm{Re}(u_j) = 0\}} \tag{2.9}$$

where $i \in I = \{1, 2, \cdots, m\}$ and $j \in J = \{m+1, m+2, \cdots, m+n = d\}$, with

$$d = m + n. \tag{2.10}$$

Affine Processes and their Characteristic Functions

This would be the definition of an affine process.

Definition 2.1.1. *An* **affine process** *is defined to be a stochastically-continuous time-homogeneous Markov process that has an affine log-characteristic function with respect to a starting state vector x in the state space $\mathbf{E}$, given by*

$$\boxed{\boxed{\phi^x_{\mathbf{X}}(u)|_t = \mathbb{E}^x[e^{<X_t, u>}] = e^{(\Phi(t,u) + <x, \Psi(t,u)>)}}} \tag{2.11}$$

$\forall x \in \mathbf{E}$, and $\forall (t, u) \in \mathbf{R}_{\geq 0} \times i\mathbf{R}^d$, and where the functions Φ and Ψ satisfy

$$\boxed{\Phi : \mathbf{R}_{\geq 0} \times i\mathbf{R}^d \longrightarrow \mathbf{C}} \quad \text{and} \quad \boxed{\Psi : \mathbf{R}_{\geq 0} \times i\mathbf{R}^d \longrightarrow \mathbf{C}^d} \tag{2.12}$$

We immediately note that the exponent

$$\Phi(t,u)+ < x, \Psi(t,u) >, \tag{2.13}$$

that appears in the above characteristic function, is an **affine function** of x, given that it is a sum of a linear function of x (the part $< x, \Psi(t,u) >$) plus a constant (independent of x) translation (the part $\Phi(t,u)$). This would establish the "affine log-characteristic" nature of affine processes.

(Markov) Semigroups Associated with Markov and Affine Processes

It is known that one can associate to each time-homogenous Markov process $\mathbf{X} = \{X_t\}_{t\in\mathbf{R}_{\geq 0}}$ a (Markov) semigroup (see section 1.1.3 for all details) $\{P_t\}_{t\in\mathbf{R}_{\geq 0}}$ acting on the bounded Borel functions $b\mathcal{B}(\mathbf{E})$ (defined on the Borel-measurable state space $\mathbf{E}$), where for every $f \in b\mathcal{B}(\mathbf{E})$ and for every $t \in \mathbf{R}_{\geq 0}$, $P_t f$ is given by

$$\boxed{(P_t f)(x) = \mathbb{E}^x[f(X_t)]} \qquad \forall x \in \mathbf{E} \tag{2.14}$$

where $\mathbb{E}^x$ is the expectation with respect to the Law $\mathbb{P}^x$.

In particular, with every affine process $\mathbf{X}$ one can associate such semigroups. We will refer to these semigroups at some points in what follows.

2.1.2 Basic Elementary Properties for Affine Processes

The study of affine processes is essentially the study of the two functions Ψ and Φ that form the (affine exponent of the) characteristic function of these processes. In what follows, we basically discuss some properties associated with these two functions.

- We first start by stating a lemma which gives a continuous extension of $\Phi(t,u)$ and $\Psi(t,u)$ to a set $\mathcal{O} \subset \mathbf{R}_{\geq 0} \times \mathbf{C}^d$ that contains $\mathbf{R}_{\geq 0} \times i\mathbf{R}^d$ (where Φ and Ψ initially existed by definition).

- It is these extended functions Φ and Ψ that lead to some important collection of properties, including, for example, the **semi-flow property** for affine processes, and which also help (later on) in establishing that every affine process is (what is known as) a **Feller process**.

- If, furthermore, one assumes some **regularity** (i.e. differentiability) assumptions on Φ and Ψ, one establishes what are known as **regular affine processes**, and these

regular affine processes are a very favourable kind of processes for doing calculations in the context of mathematical finance, as will be discussed in chapter 3.

We first begin with basic properties for the function Φ and Ψ associated (as above) with any affine process $\mathbf{X}$. In particular we start with the following lemma which gives a continuous extension of $\Phi(t,u)$ and $\Psi(t,u)$ to a set $\mathbf{R}_{\geq 0} \times \mathcal{U}$, where we use $f_u(x)$ to mean

$$f_u(x) = e^{<u,x>}. \tag{2.15}$$

Lemma 2.1.1. *Let $\mathbf{X} = \{X_t\}_{t\geq 0}$ be an affine process, and the set $\mathcal{U}$ as defined above. Then the set*

$$\boxed{\mathcal{O} = \{(t,u) \in \mathbf{R}_{\geq 0} \times \mathcal{U} : \forall s \in [0,t], P_s f_u(0) = \mathbb{E}^x[e^{<u,X_s>}] \neq 0\}}, \tag{2.16}$$

where

$$\mathbf{R}_{\geq 0} \times i\mathbf{R}^d \subset \mathcal{O} \subset \mathbf{R}_{\geq 0} \times \mathcal{U}, \tag{2.17}$$

is open in $\mathbf{R}_{\geq 0} \times \mathcal{U}$, and there exists a unique continuous extension of $\Phi(t,u)$ and $\Psi(t,u)$ to $\mathcal{O}$, and that (2.11) holds $\forall (t,u) \in \mathcal{O}$

(see Duffie [2003, lemma 3.1 for a proof])

Given the above basic extension lemma to the set $\mathcal{O}$, we have the following properties for the Φ and Ψ functions for an affine process.

Lemma 2.1.2. *The functions Φ and Ψ associated with an affine prosses $\mathbf{X}$ satisfy the following:*

1. *Φ maps $\mathcal{O}$ to $\mathbf{C}_- := \{u \in \mathbf{C} : \mathrm{Re}(u) \leq 0\}$.*

2. *Ψ maps $\mathcal{O}$ to $\mathcal{U}$.*

3. *$\Phi(0,u) = 0$ and $\Psi(0,u) = u$ for all $u \in \mathcal{U}$*

4. *For every $t, s \geq 0$ such that $(t+s, u) \in \mathcal{O}$, Φ and Ψ satisfy the* **semi-flow property** *consisting of the functional equation*

$$\boxed{\Psi(t+s,u) = \Psi(s, \Psi(t,u))}, \tag{2.18}$$

which is called the semi-flow property for Ψ, and of the equation

$$\boxed{\Phi(t+s,u) = \Phi(s, \Psi(t,u)) + \Phi(t,u)}, \tag{2.19}$$

which is the functional equation for Φ called the **additive cocycle** *of the semi-flow Ψ.*

5. *For a fixed $u_J = (u_{m+1}, u_{m+2}, \cdots, u_d)$ and for a fixed t, the functions*

$$\Phi(t, (-, u_J)) \quad \text{and} \quad \Psi(t, (-, u_J)), \tag{2.20}$$

are analytic functions of $u_I = (u_1, u_2, \cdots, u_m)$ (i.e. can be written in the form of a power-series in terms of these variables) in the domain where $\mathrm{Re}(u_I) < 0$ and $(t,u) \in \mathcal{O}$.

6. *Φ and Ψ are jointly continuous on $\mathcal{O}$.*

7. *Let $(t,u), (t,w) \in \mathcal{O}$ with $\mathrm{Re}(u) \leq \mathrm{Re}(w)$. Then*

$$\mathrm{Re}(\Phi(t,u)) \leq \Phi(t, \mathrm{Re}(w)), \tag{2.21}$$

$$\mathrm{Re}(\Psi(t,u)) \leq \Psi(t, \mathrm{Re}(w)) \tag{2.22}$$

Proof: 1) and 2):first we note that the semigroup $\{P_t\}_{\geq 0}$ (defined above) is a contractive semi-group of operators, since $(P_t f)(x) = \mathbb{E}^x[f(X_t)]$, giving that

$$|(P_t f)(x)| = |\mathbb{E}^x[f(X_t)]| \leq |f|, \tag{2.23}$$

and hence that $||P_t|| \leq 1$. Thus we have

$$||P_t f_u||_\infty \leq ||f_u||_\infty = 1, \tag{2.24}$$

given that $\forall x \in \mathbf{E}$ and $\forall u \in \mathcal{U}$ (refer to the definition of $\mathbf{E}$ and $\mathcal{U}$ above) one has that $f_u(x)$ has the form

$$f_u(x) = e^{<u,x>} = e^{a+ib} \tag{2.25}$$

where $a \leq 0$. Also, since

$$P_t f_u(x) = \mathbb{E}^x[f_u(X_t)] = \mathbb{E}^x[e^{<u,X_t>}], \tag{2.26}$$

and by the affine property we have that

$$\mathbb{E}^x[e^{<u,X_t>}] = \exp(\Phi(t,u) + < x, \Psi(t,u) >), \tag{2.27}$$

then

$$\begin{aligned} P_t f_u(x) &= \exp(\Phi(t,u) + < x, \Psi(t,u) >) \\ &= e^{\Phi(t,u)} f_{\Psi(t,u)}(x). \end{aligned} \tag{2.28}$$

Now since $||f_u||_\infty \leq 1$ if and only if $u \in \mathcal{U}$ then, first, $\Phi(t,u)$ must be in $\mathbf{C}_-$ as otherwise, if it is in $\mathbf{C}_{Re>0}$, then $|e^{\Phi(t,u)}| > 1$ which gives that

$$|e^{\Phi(t,u)} f_{\Psi(t,u)}(x)| > |f_{\Psi(t,u)}(x)|, \tag{2.29}$$

and thus for some subset of $\mathbf{E}$ of positive measure $|e^{\Phi(t,u)} f_{\Psi(t,u)}(x)| > 1$ giving that $||P_t f_u||_\infty = ||e^{\Phi(t,u)} f_{\Psi(t,u)}|| > 1$ contradicting that $||P_t f_u||_\infty \leq 1$. this proves 1).

For 2) note that if $\Psi(t,u)$ is not in $\mathcal{U}$ then $|f_{\Psi(t,u)}(x)| > 1$ giving that $|e^{\Phi(t,u)} f_{\Psi(t,u)}(x)| \geq 1$ for a subset of $\mathbf{E}$ of positive measure and hence that $||e^{\Phi(t,u)} f_{\Psi(t,u)}||_\infty > 1$ contradicting, as above, that it should be ≤ 1.

3) By letting $t = 0$ in $P_t f_u(x) = e^{\Phi(t,u)} f_{\Psi(t,u)}(x)$ one obtains that

$$P_0 f_u(x) = e^{\Phi(0,u)} f_{\Psi(0,u)}(x), \tag{2.30}$$

$\forall x \in \mathbf{E}$ and $\forall u \in \mathcal{U}$. But $P_0 f_u(x) = f_u(x)$, hence

$$f_u(x) = e^{\Phi(0,u)} f_{\Psi(0,u)}(x), \tag{2.31}$$

$\forall x \in \mathbf{E}$ and $\forall u \in \mathcal{U}$, and thus $\Phi(0,u) = 0$ and $\Psi(0,u) = u$

4) By the semi-group property

$$\begin{aligned} P_{t+s} f_u(x) &= P_t P_s f_u(x) \\ &= P_t(e^{\Phi(s,u)} f_{\Psi(s,u)}(x)) \\ &= e^{\Phi(s,u)} P_t f_{\Psi(s,u)}(x) \\ &= e^{\Phi(s,u)} e^{\Phi(t,\Psi(s,u))} f_{\Psi(t,\Psi(s,u))}(x) \qquad (2.32) \\ &= e^{\Phi(s,u)+\Phi(t,\Psi(s,u))} f_{\Psi(t,\Psi(s,u))}(x), \qquad (2.33) \end{aligned}$$

$\forall x \in \mathbf{E}$ and $\forall (t+s,u) \in \mathcal{O}$. But

$$P_{t+s} f_u = e^{\Phi(t+s,u)} f_{\Psi(t+s,u)}(x), \tag{2.34}$$

$\forall x \in \mathbf{E}$ and $\forall (t+s,u) \in \mathcal{O}$. Hence

$$\Phi(t+s,u) = \Phi(s,u) + \Phi(t,\Psi(s,u)), \tag{2.35}$$

and

$$\Psi(t+s,u) = \Psi(t,\Psi(s,u)). \tag{2.36}$$

This establishes the semi-flow property.

5) The analyticity of $\Phi(t,(-,u_J))$ and $\Psi(t,(-,u_J))$ in $\mathrm{Re}(u_I) < 0$ where $(t,u) \in \mathcal{O}$, follows from the analyticity properties of the extended moment generating function $\mathbb{E}^x[f_u(X_t)]$ mentioned in the remark after the proof.

6) Consider any sequence $\{(t_n,u_n)\} \subset \mathcal{O}$ that converges to $(t,u) \in \mathcal{O}$. Since $\mathbf{X}$ is a stochastically continuous process, then $\{X_{t_n}\} \longrightarrow X_t$, as $n \longrightarrow \infty$, in probability, and hence $\{X_{t_n}\} \longrightarrow X_t$ in distribution. Thus

$$e^{<X_{t_n},u_n>} \longrightarrow e^{<X_t,u>} \tag{2.37}$$

in distribution as $n \longrightarrow \infty$. Thus by using the dominated convergence theorem (see section 1.4.1) we find that

$$\begin{aligned} P_{t_n} f_{u_n} &= \mathbb{E}^x[e^{(X_{t_n},u_n)}] \\ &= e^{\Phi(t_n,u_n)+<x,\Psi(t_n,u_n)>} \\ &\longrightarrow \mathbb{E}^x[e^{(X_t,u)}] \\ &= e^{\Phi(t,u)+<x,\Psi(t,u)>} = P_t f_u(x). \end{aligned} \tag{2.38}$$

Hence

$$\Phi(t_n,u_n) \longrightarrow \Phi(t,u), \qquad \text{and} \qquad \Psi(t_n,u_n) \longrightarrow \Psi(t,u), \tag{2.39}$$

and thus Φ and Ψ are jointly continuous functions.

7) For $(t,u),(t,w) \in \mathcal{O}$ with $\mathrm{Re}(u) \leq \mathrm{Re}(w)$, one has,$\forall x \in \mathbf{E}$, that

$$\begin{aligned} \left|\mathbb{E}^x[e^{<X_t,u>}]\right| &\leq \mathbb{E}^x\left[\left|e^{<X_t,u>}\right|\right] \\ &= \mathbb{E}^x[e^{<X_t,Re(u)>}] \\ &\leq \mathbb{E}^x[e^{<X_t,Re(w)>}]. \end{aligned} \tag{2.40}$$

Since $(t,u),(t,w) \in \mathcal{O}$ then one finds, from the affine property and the above inequality that

$$Re(\Phi(t,u))+<x,Re(\Psi(t,u))> \leq \Phi(t,Re(w))+<x,\Psi(t,Re(w))>. \tag{2.41}$$

Now by taking $x = 0$ in the previous inequality one obtains

$$Re(\Phi(t,u)) \leq \Phi(t, Re(w)), \tag{2.42}$$

and by taking $x = ke_i$ one finds that

$$Re(\Phi(t,u)) + kRe(\Psi_i(t,u)) \leq \Phi(t, Re(w)) + k\Psi_i(t, Re(w)), \tag{2.43}$$

and now by letting $k \longrightarrow \infty$ we find that $kRe(\Psi_i(t,u)) \leq k\Psi_i(t, Re(w))$, and hence that

$$Re(\Psi_i(t,u)) \leq \Psi_i(t, Re(w)). \tag{2.44}$$

This completes the proof. •

Extended Moment Generating Functions & Extended Cumulant Generating Functions

For part 5) above one can first define the **extended moment generating function** given by $\Phi : \mathcal{D}_{\mathbf{C}} \longrightarrow \mathbf{C}$, where

$$\mathcal{D}_{\mathbf{C}} = \{u \in \mathbf{C}^d : \Phi(Re(u)) < \infty\} \tag{2.45}$$

by

$$\mathbf{\Phi}(u) = \int_{\mathbf{R}^d} e^{<u,\xi>} \mu(d\xi), \tag{2.46}$$

if μ in an infinitely divisible (see section 1.2.3) sub-stochastic measure on $\mathbf{R^d}$, (recall that a Borel measure μ on $\mathbf{R^d}$ is called sub-stochastic measure, if $0 < \mu(\mathbf{R^d}) \leq 1$). It is known (see for example Lemma C.3 page 93 in [KR]) that $\mathbf{\Phi}(u)$ can not have zeroes in $\mathbf{C^n}$ and thus one can define the **extended cumulant generating function**

$$\log(\mathbf{\Phi})(\mathbf{u}) = \log\left(\int_{\mathbf{R^d}} \mathbf{e}^{<\mathbf{u},\xi>} \mu(\mathbf{d}\xi)\right), \tag{2.47}$$

and that (see proposition B.4 page 92 in [KR]) the extended moment generating function Φ and the extended cumulant generating function $\log(\Phi)$ are analytic in the interior $Int(\mathcal{D}_{\mathbf{C}})$ of $\mathcal{D}_{\mathbf{C}}$.

2.1.3 The Semi-Flow Property for Affine Processes

A very important property of Φ and Ψ that appear in the charcteristic function of an affine process is the semi-flow property mentioned in lemma(2.1.2). In general semi-flows, that appear in many contents in mathematics such as in dynamical system and in topological transformation groups on manifolds, are functions f that satisfy the following functional equation

Definition 2.1.2. *A function*

$$f : G \times X \longrightarrow X, \tag{2.48}$$

where G is a semi-group (with binary operation $*$), and X is some space, is said to have the **semi-flow property**, *if it satisfies the functional equation*

$$\boxed{f(s * t, x) = f(s, f(t, x))} \tag{2.49}$$

for every $s, t \in G$ and $x \in X$.

As mentioned in lemma (2.1.2) above Ψ satisfies a semi-flow property while Φ satisfies an additive cocycle property associated with the semi-flow Ψ. One can combine Φ and Ψ in a 2-component vector Γ that satisfies a **"global" semi-flow property for the affine process X** on

$$\hat{\mathcal{O}} = \mathcal{O} \times \mathbf{C}, \tag{2.50}$$

with values in

$$\hat{\mathcal{U}} = \mathcal{U} \times \mathbf{C}, \tag{2.51}$$

i.e. a *semi-flow*

$$\Gamma : \hat{\mathcal{O}} \longrightarrow \hat{\mathcal{U}}. \tag{2.52}$$

We first define the vectors $\xi \in \hat{\mathcal{O}}$ by

$$\xi = \begin{pmatrix} t \\ u_1 \\ \vdots \\ u_d \\ u_{d+1} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} t \\ u \end{pmatrix} \\ u_{d+1} \end{pmatrix} \in \mathcal{O} \times \mathbf{C} = \hat{\mathcal{O}}, \tag{2.53}$$

and similarly we have $\rho \in \hat{\mathcal{U}}$ given by

$$\rho = \begin{pmatrix} u_1 \\ \vdots \\ u_d \\ u_{d+1} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} u \end{pmatrix} \\ u_{d+1} \end{pmatrix} \in \mathcal{U} \times \mathbf{C} = \hat{\mathcal{U}}. \tag{2.54}$$

Now $\Gamma : \hat{\mathcal{O}} \longrightarrow \hat{\mathcal{U}}$ is defined by

$$\boxed{\Gamma(\xi) = \Gamma \begin{pmatrix} t \\ u_1 \\ \vdots \\ u_d \\ u_{d+1} \end{pmatrix} = \begin{pmatrix} \Psi(t, (u_1, \cdots, u_d)) \\ \Phi(t, (u_1, \cdots, u_d)) + u_{d+1} \end{pmatrix} = \rho \in \hat{\mathcal{U}}.} \tag{2.55}$$

We note the following: the function Γ satisfies

1. for $u = (u_1, \cdots, u_d, u_{d+1}) \in \hat{\mathcal{U}}$, one has that

$$\begin{aligned} \Gamma \begin{pmatrix} 0 \\ u \end{pmatrix} &= \begin{pmatrix} \Psi(0, (u_1, \cdots, u_d)) \\ \Phi(0, (u_1, \cdots, u_d)) + u_{d+1} \end{pmatrix} = \begin{pmatrix} (u_1, \cdots, u_d) \\ u_{d+1} \end{pmatrix} \\ &= \begin{pmatrix} u_1 \\ \vdots \\ u_d \\ u_{d+1} \end{pmatrix} = u. \end{aligned} \tag{2.56}$$

2. also for $u = (u_1, \cdots, u_d, u_{d+1}) \in \hat{\mathcal{U}}$, one has that

$$\Gamma\begin{pmatrix} t+s \\ u \end{pmatrix} = \begin{pmatrix} \Psi(t+s,(u_1,\cdots,u_d)) \\ \Phi(t+s,(u_1,\cdots,u_d)) + u_{d+1} \end{pmatrix}$$
$$= \begin{pmatrix} \Psi(s,\Psi(t,(u_1,\cdots,u_d))) \\ \Phi(s,\Psi(t,(u_1,\cdots,u_d))) + \Phi(t,(u_1,\cdots,u_d)) + u_{d+1} \end{pmatrix},$$

while

$$\Gamma\begin{pmatrix} s \\ \Gamma(t,u) \end{pmatrix} = \Gamma\left(\begin{pmatrix} s \\ \begin{pmatrix} \Psi(t,(u_1,\cdots,u_d)) \\ \Phi(t,(u_1,\cdots,u_d)) + u_{d+1} \end{pmatrix} \end{pmatrix}\right.$$
$$= \begin{pmatrix} \Psi(s,\Psi(t,(u_1,\cdots,u_d))) \\ \Phi(s,\Psi(t,(u_1,\cdots,u_d))) + \Phi(t,(u_1,\cdots,u_d)) + u_{d+1} \end{pmatrix}.$$

Thus

$$\boxed{\Gamma\begin{pmatrix} t+s \\ u \end{pmatrix} = \Gamma\begin{pmatrix} s \\ \Gamma(t,u) \end{pmatrix}} \tag{2.57}$$

and Γ is indeed a semi-flow function. Thus we have established that

Lemma 2.1.3. *The above defined function Γ establishes a (global) semi-flow for any given affine process* **X**.

2.1.4 The Feller Property for Affine Processes

The Feller property for affine processes is simply the fact that

Every affine process is a Feller process

As such, we note that the family $\{P_t\}_{t\geq 0}$ of (probability) transition functions associated with any affine process (being itself a Markov process), and which forms a semi-group (as discussed in section 1.1.3), in fact forms a *Feller semi-group*, as per the definition (below) for a Feller process.

We start by defining Feller semigroups and Feller processes.

Feller Semigroups and Feller Processes

Let $bC(X)$ be the space of all real valued continuous functions on a locally compact topological space X with a countable base. And let $bC(X)$ be given the supremum norm $||f||$. A **Feller Semigroup** on $bC(X)$ is a collection $\{P_t\}_{t\geq 0}$ of positive linear maps from $bC(X)$ to $bC(X)$, i.e. $\forall t \geq 0$

$$P_t : bC(X) \longrightarrow bC(X), \tag{2.58}$$

such that the following are satisfied:

1. For every $t \geq 0$, and for every $f \in bC(X)$, one has that

$$\boxed{||P_t f|| \leq ||f||} \tag{2.59}$$

   i.e. the family $\{P_t\}_{t\geq 0}$ is a family of contractions with respect to the supremum norm on $bC(X)$.

2. The family $\{P_t\}_{t\geq 0}$ has the semigroup property

$$\boxed{P_{t+s} = P_t \circ P_s} \tag{2.60}$$

   for every $t, s \geq 0$.

3. For every $f \in bC(X)$, one has that

$$\boxed{\lim_{t\to 0} ||P_t f - f|| = 0} \tag{2.61}$$

Definition 2.1.3. *A* **Feller Process** *is defined as being a Markov Process* $\mathbf{X}$ *with a Feller semigroup of transition function* $\{P_t\}_{t\geq 0}$ *i.e. with a (probability) transition functions* $\{P_t\}_{t\geq 0}$ *forming a Feller semigroup.*

One can define infinitesimal generators associated with Feller processes as follows. A function $f \in C_0(X)$ is said to be in the domain of the **generator** A associated with a Feller Process $\mathbf{X}$ with probability transition function derived from a Feller semigroup $\{P_t\}_{t\geq 0}$ if

$$Af = \lim_{t\to 0} \frac{P_t f - f}{t} \tag{2.62}$$

exists uniformly. A is said to be a generator of the Feller semigroup$\{P_t\}_{t\geq 0}$ with domain D_A being the subset of $C_0(X)$ consisting of all $f \in C_0(X)$ where the above limits Af exist uniformly.

It is known that Brownian motions, Poisson processes, Lévy processes, and solutions to SDE's with Lipschitz continuous coefficients are Feller processes.

We also have the following proposition.

Proposition 2.1.1. *Every Feller process* $\mathbf{X}$ *satisfies the strong Markov property.*

Every Affine Process is a Feller process

We now prove the following lemma

Lemma 2.1.4. *Every affine process is a Feller process*

we first start by stating the following proposition (see proof page 14 in (KR)) that we shall need.

Proposition 2.1.2. *Let* $\mathbf{X} = \{X_t\}_{t\geq 0}$, *be an affine process on* $\mathbf{E} = \mathbf{R}^{\mathrm{m}}_{\geq 0} \times \mathbf{R}^{\mathrm{n}}$, *and let* J *denote the components* $(m+1, \cdots, m+n)$ *in* $\mathbf{E}$ *then there exists an* $n \times n$ *matrix* M *such that*

$$\Psi_J(t,u) = (\Psi_{m+1}, \cdots, \Psi_{m+n}) = e^{tM} u_J, \tag{2.63}$$

where

$$u_J = \begin{pmatrix} u_{m+1} \\ \vdots \\ u_{m+n} \end{pmatrix}, \tag{2.64}$$

for every $(t,u) \in \mathcal{O}$.

We also need the following proposition (for proof see proposition 1.10 page 15 in [KR])

Proposition 2.1.3. *Let $\Psi(t,u)$ be the semi-flow associated with the affine process* $\mathbf{X}$. *Then $\forall(t,u) \in \mathcal{O}$, such that $u \in Int(\mathcal{U})$, one has that $\Psi(t,u) \in Int(\mathcal{U})$.*

Proof of Lemma(2.1.4):

Let $\mathbf{X} = \{X_t\}_{t\geq 0}$ be an affine process. Using the stochastic continuity of $\mathbf{X}$ and using the dominated convergence theorem (see section 1.4.1) one can immediatly show that

$$P_t f(x) = \mathbb{E}^x[f(X_t)] \longrightarrow f(x) \qquad \text{as } t \longrightarrow 0, \tag{2.65}$$

for every $f \in C_0(\mathbf{E})$ and for every $x \in \mathbf{E}$ (since $\mathbb{E}^x$ is the expectation with respect to the probability measure $\mathbb{P}^x$ which is the law of the process started at x).

Now since $\{P_t\}_{t\geq 0}$ is already known to be a family of contractions on $C_0(x)$ and satisfies the semigroup property then these (together with what was mentioned in the previous paragraph) show that the family $\{P_t\}_{t\geq 0}$ satisfies the three axioms for the Feller property. Only one thing remains to show that $\forall t \geq 0, P_t$ maps $C_0(\mathbf{E})$ to $C_0(\mathbf{E})$.

Consider the set Θ of functions given by

$$\Theta = \left\{ h_{(u_I,g)}(x) = e^{<u_I,x_I>} \int_{\mathbf{R}^n} f_{iz}(x_J) g(z) dz : u_i \in \mathrm{Int}(\mathcal{U}_I), g \in C_c^\infty(\mathbf{R}^n) \right\}, \tag{2.66}$$

and let $\mathcal{L}(\Theta)$ be the set of (complex and finite) linear combinations of functions in Θ. Note that $\mathcal{L}(\Theta)$ is an algebra of functions and is also closed under complex conjugation. It is also easy to check that $\mathcal{L}(\Theta)$ separates points in $\mathbf{E}$ and vanishes nowhere in $\mathbf{E}$. It also follows from the Reimann-Lebesgue Lemma (see section 1.4.2) that $\int_{\mathbf{R}^n} f_{iz}(x_J) g(z) dz$ vanishes at ∞, and thus that every $h_{(u_I,g)}(x) \in \Theta$ belongs to $C_0(\mathbf{E})$, and hence that every element in $\mathcal{L}(\Theta)$ is in $C_0(\mathbf{E})$. Thus $\mathcal{L}(\Theta)$ is a subalgebra of $C_0(\mathbf{E})$ and therefore, by using the Stone-Weierstrass Theorem, $\mathcal{L}(\Theta)$ is dense in $C_0(\mathbf{E})$.

For any given $t \in \mathbf{R}_{\geq 0}$ we have, by proposition(2.1.2), that $\Psi_{u_J}(t,u) = e^{\beta(t)} u_J$ where β is a real $n \times n$ matrix, whenever $(t,u) \in \mathcal{O}$, and that by lemma

(2.1.1) we have $\mathbb{E}^x[f_u(X_t)] = \Phi(t,u) + < x, \Psi(t,u) >$ if $(t,u) \in \mathcal{O}$, with $\mathbb{E}[f_{(u_I,iz)}(X_t)] = 0$ if $(t,u) \notin \mathcal{O}$. Hence, for a given $h(x) \in \Theta$, we find that

$$\begin{aligned} P_t h(x) &= \mathbb{E}^x \left[e^{<u_I,x_I>} \int_{\mathbf{R}^n} f_{iz}(x_J) g(z) dz \right] \\ &= \mathbb{E}^x \left[\int_{\mathbf{R}^n} e^{<u_I,x_I>} e^{<iz,x_J>} g(z) dz \right] \\ &= \mathbb{E}^x \left[\int_{\mathbf{R}^n} e^{<u_I,x_I>+<iz,x_J>} g(z) dz \right] \\ &= \mathbb{E}^x \left[\int_{\mathbf{R}^n} e^{<(u_I,iz),(x_I,x_J)>} g(z) dz \right] \\ &= \mathbb{E}^x \left[\int_{\mathbf{R}^n} f_{(u_I,iz)}(X_t) g(z) dz \right] \end{aligned} \tag{2.67}$$

and thus we find (using Fubini's theorem) that

$$\begin{aligned} P_t h(x) &= \int_{\mathbf{R}^n} \mathbb{E}^x[f_{(u_I,iz)}(X_t)] g(z) dz \\ &= \int_{\{u \in \mathcal{U}:(t,u) \in \mathcal{O}\}} P_t f_{(u_I,iz)}(x) g(z) dz \\ &= \int_{\{u \in \mathcal{U}:(t,u) \in \mathcal{O}\}} e^{(\Phi(t,u_I,iz)+<x_I,\Psi_I(t,u_I,iz)>+<x_J,e^{t\beta}iz>)} g(z) dz. \end{aligned} \tag{2.68}$$

Now since $(u_I, iz) \in Int(\mathcal{U})$, it follows by the previous proposition(2.1.3) that $\Psi(t,(u_I,iz)) \in Int(\mathcal{U})$, and thus that

$$Re(\Psi_I(t,(u_I,iz))) < 0 \tag{2.69}$$

for any $z \in \mathbf{R}^n$. Therefore, as $|x_I| \to \infty$, one finds that $P_t h(x) \to 0$. In addition, since $P_t h(x)$ is the Fourier Transform (over the J components) of a compactly supported density then by the Reimann-Lebesgue Lemma one finds that $P_t h(x) \to 0$ as $|x_J| \to \infty$. Hence $P_t h(x) \to 0$ as$|x| \to \infty$, which means that $P_t h(x) \in C_0(\mathbf{E})$, and thus that $P_t r(x) \to 0$ as $|x| \to \infty$ for any $r(x) \in \mathcal{L}(\Theta)$.

Finally this shows that $P_t r(x) \to 0$ as $|x| \to \infty$ for every $r(x) \in C_0(\mathbf{E})$ given that $\mathcal{L}(\Theta)$ is dense in $C_0(\mathbf{E})$. This shows that $\forall t \geq 0$, P_t maps $C_0(\mathbf{E})$ to $C_0(\mathbf{E})$ and the result follows. •

2.2 The Semi-Flows Versus the Nature of the Affine Processes

This part illustrates the *interplay* between the semi-flow properties and the nature of the given affine process $\mathbf{X}$.

We begin with Lévy processes.

2.2.1 Every Conservative Affine Process is a Lévy Process

Definition 2.2.1. *A* **Lévy process** *on a triple* (Ω, σ, P)*, is an* $\mathbf{R}^d$*−valued process* $\mathbf{Y} = \{Y_t\}_{t \geq 0}$ *such that*

- *The sample paths* $Y_t(\omega)$ *are continuous in terms of* t*.*
- *The law of* $Y_{t+h} - Y_t$ *is dependent on* h*, and not on* t*.*
- *If* $s < t \leq w < z$*, then* $Y_t - Y_s$ *and* $Y_z - Y_w$ *are independent.*

Now suppose that $\mathbf{X} = \{X_t\}_{t \geq 0}$ is a **conservative affine process**, i.e. that $\mathbf{X}$ is affine with $\Phi(t, 0) = \Psi(t, 0) = 0$ for every $t \geq 0$. Assume further that $\mathbf{X}$ has a stationary semi-flow Ψ given by

$$\Psi(t, u) = u, \qquad \forall (t, u) \in \mathcal{O}. \tag{2.70}$$

Then the functional equation for the cocycle Φ of Ψ becomes

$$\begin{aligned} \Phi(t + s, u) &= \Phi(s, \Psi(t, u)) + \Phi(t, u) \\ &= \Phi(s, u) + \Phi(t, u). \end{aligned} \tag{2.71}$$

Since Φ is continuous and satisfies $\Phi(0,u) = 0$ (a property), then $\Phi(-,u)$ is a linear function for any fixed u, and hence $\Phi(t,u)$ has the form

$$\Phi(t,u) = kt = m(u)t \tag{2.72}$$

where $k = m(u)$ is some constant dependent on u. Hence $P_t f_u(0)$ has the form

$$\begin{aligned} P_t f_u(0) &= \mathbb{E}^0[e^{<X_t,u>}] \\ &= e^{\Phi(t,u)+<0,\Psi(t,u)>} \\ &= e^{m(u)t}, \end{aligned} \tag{2.73}$$

and thus $e^{m(u)t}$ is a characteristic function $\forall t \geq 0$. Hence $\mathbb{E}^0[e^{<X_t,u>}]$ is an *infinitely divisible characteristic function* and thus, by the **central theorem on Lévy processes** (see [RW]) p.74), $m(u)$, as a function of u, has to be of Lévy-Khintchine form, i.e. of the form

$$m(u) = b^T \cdot u + \frac{1}{2} u^T M u + \int (e^{u \cdot x} - 1 - u \cdot x I_{|x| \leq 1}) \nu(dx) \tag{2.74}$$

where $b \in \mathbf{R}^d$, M is some symmetric $d \times d$ matrix, and ν is called a *Lévy measure* which satisfies

$$\int_{\mathbf{R}^d - \{0\}} (1 \wedge x^2) \nu(dx) < \infty. \tag{2.75}$$

Hence (by this theorem) the process $\mathbf{X}$ must be a Lévy Process.

2.2.2 Every Ornstein-Uhlenbeck-Type Process is an Affine Process

Let $\mathbf{X} = \{X_t\}_{t \geq 0}$ be a *conservative affine process* on the state space $\mathbf{E} = \mathbf{R}$, defined as a *semi-martingale*, i.e. where each X_t can be decomposed as the sum of a local martingale M_t and an adapted finite-variation process A_t, giving for every $t \geq 0$ that

$$X_t = M_t + A_t. \tag{2.76}$$

Then by (proposition 1.9 in the paper [KR]), $\Psi(t,u)$ has the form

$$\Psi(t,u) = e^{t\beta}u, \tag{2.77}$$

for some constant (i.e. 1×1 matrix) $\beta \in \mathbf{R}$.

Now consider the stochastic differential equation

$$\boxed{dX_t = \beta X_t dt + dL_t} \tag{2.78}$$

where $X_0 = x \in \mathbf{R}$, and where L_t is a Lévy Process with characteristic exponent κu, and $\beta \in \mathbf{R}$. this SDE has a unique solution $\mathbf{X} = \{X_t\}_{t\geq 0}$ (see Sato [1999]) called the **Ornstein-Uhlenbeck-Type Process**. It can be shown that the characteristic function of this process is given by

$$\mathbb{E}^x[e^{<X_t,u>}] = exp\left(\int_0^t \kappa(e^{s\beta}u)ds + xe^{t\beta}u\right) \tag{2.79}$$

(where $xe^{t\beta}u =< x, e^{t\beta}u >$). Given this characteristic function, one can conclude that $\mathbf{X}$ is a conservative affine process where

$$\Phi(t,u) = \int_0^t \kappa(e^{s\beta}u)ds, \tag{2.80}$$

and

$$\Psi(t,u) = e^{t\beta}u, \tag{2.81}$$

as should be the case for the semi-flow Ψ for a conservative affine process (above).

Note: That if $\mathbf{X} = \{X_t\}_{t\geq 0}$ is a conservative affine process on the state space $\mathbf{E} = \mathbf{R}$, then by (proposition 1.9 in the paper [KR]), $\Psi(t,u)$ must have the form

$$\Psi(t,u) = e^{t\beta}u, \tag{2.82}$$

for some constant (i.e. 1×1 matrix) $\beta \in \mathbf{R}$, which is the same as the one for an OU-process (above). However the $\Phi(t,u)$ part of this affine process may not have the same form as that for an OU-process, and thus it would not be the case that a (conservative) affine process satisfies the above SDE.

2.2.3 Every Squared Bessel Process is an Affine Process

Consider the SDE

$$\boxed{dZ_t = 2\sqrt{Z_t}\, dW_t + \delta dt} \tag{2.83}$$

where $Z_0 = z \geq 0$. There exists (see Revuz and Yor [1999]) a unique solution that is non-negative and has the extended moment generating function

$$\mathbb{E}^z[e^{uZ_t}] = exp\left(\frac{\delta}{2}\log(1-2ut) + z\frac{u}{1-2ut}\right) \tag{2.84}$$

defined for all $u \in \mathbf{C}$ with $\mathrm{Re}(u) < \frac{1}{2t}$. This process $\mathbf{Z} = \{Z_t\}_{t\geq 0}$, called the **Squared Bessel Process** of dimension δ, is clearly an affine Process on $\mathbf{E} = \mathbf{R}_{\geq 0}$ where

$$\Phi(t,u) = \frac{\delta}{2}\log(1-2ut), \tag{2.85}$$

and

$$\Psi(t,u) = \frac{u}{1-2ut}. \tag{2.86}$$

Note that for every $t \geq 0$, $\Psi(t,u)$ is a Möbius transformation and thus is a bijective conformal map of the extended complex plane to itself.

2.3 Regular Affine Processes

In this section we discuss regularity properties for the semi-flows Ψ and the cocycles Φ of the semi-flows, for affine processes. These regularity properties will be very important in the study of affine processes, roughly speaking

The regularity properties are about the differentiability of the functions $\Phi(t,u)$ and $\Psi(t,u)$ with respect to the time parameter t

For affine processes $\mathbf{X}$, where Φ and Ψ are parts of the characteristic function, differentiability in the u variable is equivalent to the existance of momements, and

the existance of such bounded moments should also exhibit in these semi-flows some regularity in the time variable t.

The existance of regularity assumptions for affine processes would enable one to fully characterize these processes in terms of their infinitesimal generators, defined as the operators

$$Af(x) = \lim_{t \to 0} \frac{\mathbb{E}^x[f(X_t)] - f(x)}{t}, \tag{2.87}$$

for functions $f : \mathbf{R}^n \to \mathbf{R}$, and would result in other important results.

2.3.1 Regularity Assumption for Affine Processes

Definition 2.3.1. *An affine process is said to be a* **regular affine process** *if the (right-sided) derivatives*

$$F(u) = \frac{\partial \Phi}{\partial t}(t,u)\bigg|_{t=0+}, \qquad R(u) = \frac{\partial \Psi}{\partial t}(t,u)\bigg|_{t=0+}, \tag{2.88}$$

exist for all $u \in \mathcal{U}$ and are continuous at $u = 0$

for a regular affine process $\mathbf{X}$ the functions $F(u)$ and $R(u)$ are called the **functional characteristics** of x, since these functions (as can be demonstrated) completely characterize the process $\mathbf{X}$.

For a regular affine process $\mathbf{X}$ one can differentiate the semi-flow equations

$$\begin{aligned} \Phi(t+s,u) &= \Phi(t,u) + \Phi(s, \Psi(t,u)) \\ \Psi(t+s,u) &= \Psi(s, \Psi(t,u)) \end{aligned}$$

with respect to s and evaluate at $s = 0$ to obtain

$$\frac{\partial \Phi}{\partial s}(t+s,u)\bigg|_{s=0} = \frac{\partial \Phi}{\partial s}(t,u)\bigg|_{s=0} + \frac{\partial \Phi}{\partial s}(s, \Psi(t,u))\bigg|_{s=0} = F(\Psi(t,u)) \tag{2.89}$$

$$\frac{\partial \Psi}{\partial s}(t+s,u)\bigg|_{s=0} = \frac{\partial \Psi}{\partial s}(s, \Psi(t,u))\bigg|_{s=0} = R(\Psi(t,u)). \tag{2.90}$$

Now since $\frac{\partial\Phi}{\partial t}(t+s,u) = \frac{\partial\Phi}{\partial s}(t+s,u)$ and similarly $\frac{\partial\Psi}{\partial t}(t+s,u) = \frac{\partial\Psi}{\partial s}(t+s,u)$, then

$$\frac{\partial\Phi}{\partial t}(t,u) = \frac{\partial\Phi}{\partial t}(t+s,u)\Big|_{s=0} = \frac{\partial\Phi}{\partial s}(t+s,u)\Big|_{s=0} = F(\Psi(t,u)) \qquad (2.91)$$

and

$$\frac{\partial\Psi}{\partial t}(t,u) = \frac{\partial\Psi}{\partial t}(t+s,u)\Big|_{s=0} = \frac{\partial\Psi}{\partial s}(t+s,u)\Big|_{s=0} = R(\Psi(t,u)) \qquad (2.92)$$

Then one obtains the following two ODE's

$$\frac{\partial\Phi}{\partial t}(t,u) = F(\Psi(t,u)) \qquad \Phi(0,u) = 0 \qquad (2.93)$$

$$\frac{\partial\Psi}{\partial t}(t,u) = R(\Psi(t,u)) \qquad \Psi(0,u) = u \qquad (2.94)$$

that are called the **generalized Riccati differential equations** for Φ and Ψ.

2.3.2 The Main Result for Regular Affine Processes

The main result for regularity, concerning Φ and Ψ that are associated with an affine process $\mathbf{X}$, indicates that $F(u)$ and $R(u)$ as defined above have a specific form given by Log-Characteristic functions of sub-stochastic infinitely divisible measures (for an explanation of the terminology, see Appendix page 90 in [KR]) satisfying some additional admissibility conditions.

As was mentioned in section (1.2.3) the characteristic function of an infinitely divisible probability measure is characterized by the Lévy-triplet (a,b,m) of parameters where a is a positive definite (diffusion) matrix, b is a (drift) vector and

m is a Lévy measure, i.e. a (σ-finite Borel) measure satisfying the integrability condition

$$\int_{\mathbf{E}} (1 \wedge |\xi|^2) m(d\xi) < \infty. \tag{2.95}$$

While for the case of a sub-stochastic infinitely divisible measures a fourth parameter c in $\mathbf{R}_{\geq 0}$ given by $c = -log\mu(\mathbf{E})$, where (μ is the measure on $\mathbf{E}$).

Given this, we find that the function $F(u)$, which is one-dimensional is characterized by the Lévy-quadruplet (a, b, c, m), while $R(u)$, which is d-dimensional is characterized by the d Lévy-quadruplets given by

$$\left((\alpha_1, \beta_1, \gamma_1, \mu_1), \cdots, (\alpha_d, \beta_d, \gamma_d, \mu_d)\right), \tag{2.96}$$

with the quadruplet $(\alpha_i, \beta_i, \gamma_i, \mu_i)$ corresponding to the i-th component $R_i(u)$ of $R(u), i = 1, 2, \cdots, d$.

We define the following admissibility conditions:

Definition 2.3.2. *A* **parameter set** *for an affine process* $\mathbf{X}$ *positive semi-definite real* $d \times d$*-matrices* $a, \alpha^1, \cdots, \alpha^d$*; by* $\mathbf{R}^d$ *valued vectors* $b, \beta^1, \cdots, \beta^d$*; by non-negative numbers* $c, \gamma^1 \cdots, \gamma^d$ *and by Lévy measures* $m, \mu^1, \cdots, \mu^d$ *on* $\mathbf{R}^d$*, forming a one-dimensional quadruplet* (a, b, c, m) *and a* d*-dimensional quadruplet* $\left((\alpha^1, \beta^1, \gamma^1, \mu^1), \cdots, (\alpha^d, \beta^d, \gamma^d, \mu^d)\right)$.

A parameter set as above is said to be **an admissible parameter set** *for an affine process* $\mathbf{X}$ *with state space* $\mathbf{E}$*, if the following are satisfied*

$$a_{kl} = 0 \qquad \text{if } k \in I \text{ or } l \in I \tag{2.97}$$

$$\alpha^j = 0 \qquad \forall j \in J \tag{2.98}$$

$$\alpha^i_{kl} = 0 \qquad \text{if } k \in I \backslash \{i\} \text{ or } l \in I \backslash \{i\} \tag{2.99}$$

$$b \in \mathbf{E} \tag{2.100}$$

$$\beta^i_k \geq 0 \qquad \forall\ i \in I \text{ and } k \in I \backslash \{i\} \tag{2.101}$$

$$\beta^j_k = 0 \qquad \forall\ j \in J \text{ and } k \in I \tag{2.102}$$

$$\gamma^j = 0 \qquad \forall\ j \in J \tag{2.103}$$

$$\text{supp}\{m\} \subseteq \mathbf{E} \qquad \text{and} \qquad \int_{\mathbf{E}\setminus\{0\}} \{(|x_I| + |x_J|^2) \wedge 1\} m(dx) < \infty \tag{2.104}$$

$$\mu^j = 0 \qquad j \in J \tag{2.105}$$

$$\text{supp}\{\mu^i\} \subseteq \mathbf{E} \qquad \forall i \in I \tag{2.106}$$

$$\int_{\mathbf{E}\setminus\{0\}} \{(|x_{I\setminus\{i\}}| + |x_{J\cup\{i\}}|^2) \wedge 1\} \mu_i(dx) < \infty \quad i \in I \tag{2.107}$$

Thus, in simpler matrix-form, one has that

$$a = \begin{pmatrix} O & \vdots & O \\ \cdots\cdots & \vdots & \cdots\cdots \\ O & \vdots & \geq\geq \end{pmatrix}, \qquad b = \begin{pmatrix} \geq \\ \vdots \\ \geq \\ \cdots\cdots \\ * \\ \vdots \\ * \end{pmatrix}$$

$$\alpha^i = \begin{pmatrix} & 0 & & \vdots & \\ & \vdots & & \vdots & \\ & 0 & & \vdots & \\ 0\cdots 0 & \alpha^i_{ii} & 0\cdots 0 & \vdots & *\cdots * \\ & 0 & & \vdots & \\ & \vdots & & \vdots & \\ & 0 & & \vdots & \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \vdots & \cdots\cdots \\ & * & & \vdots & \\ & \vdots & & \vdots & \geq\geq \\ & * & & \vdots & \end{pmatrix} \text{ where } \alpha^i_{ii} \geq 0, \qquad \alpha^j = O \text{ where } j \in J$$

$$\beta^i = \begin{pmatrix} \geq \\ \vdots \\ \geq \\ \beta_i^i \\ \geq \\ \vdots \\ \geq \\ \cdots\cdots \\ * \\ \vdots \\ * \end{pmatrix} \text{ where } i \in I \text{ and } \beta_i^i \in \mathbf{R}, \qquad \beta^j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cdots\cdots \\ * \\ \vdots \\ * \end{pmatrix} \text{ where } j \in J$$

for the structure of a, b, α^i and β^i. Note that stars denote arbitrary real numbers; the $\geq$-signs denote non-negative real numbers and the $\geq\geq$-signs positive semi-definite matrices. A big O stands for a zero-matrix, and also empty regions in a matrix denote all-zero elements. The dotted lines indicate the boundary between the first m and the last n coordinates.

Before giving the main result of regularity we define the following functions. We let $h : \mathbf{R}^d \longrightarrow [-1,1]^d$ be given, component-wise, by the following

$$h_k(\xi) = \begin{cases} 0 & k \in I \\ \frac{\xi_k}{1+\xi_k^2} & k \in J \end{cases} \qquad \forall \xi \in \mathbf{R}^d. \tag{2.108}$$

We also define functions $\chi^1, \cdots, \chi^m : \mathbf{R}^d \longrightarrow [-1,1]^d$, component-wise, by

$$\chi_k^i(\xi) = \begin{cases} 0 & k \in I \backslash \{i\} \\ \frac{\xi_k}{1+\xi_k^2} & k \in J \cup \{i\} \end{cases} \qquad \forall \xi \in \mathbf{R}^d, i \in I. \tag{2.109}$$

Now we state the following main theorem concerning the characterization of an affine process in terms of admissible functions

Theorem 2.3.1. *Let $\{\mathbf{X}\}_{t\geq 0}$ be a regular affine process with a state space* $\mathbf{E}$. *Then there exists sets of admissible parameter quadruplets*

$$(a, b, c, m) \quad \text{and} \quad (\alpha^i, \beta^i, \gamma^i, \mu^i)_{i\in\{1,\cdots,d\}}, \tag{2.110}$$

as defined bove such that

- *The functions $F(u)$ and $R(u)$ are of the Lévy-Khintchine form*

$$F(u) = \frac{1}{2} < u, au > \; + \; < b, u > -c + \int_{\mathbf{R}^d\setminus\{0\}} \left(e^{<\xi,u>} - 1 - < h(\xi), u >\right) m(d\xi) \tag{2.111}$$

$$R(u) = \frac{1}{2} < u, \alpha^i u > \; + \; < \beta^i, u > -\gamma^i + \int_{\mathbf{R}^d\setminus\{0\}} \left(e^{<\xi,u>} - 1 - < \chi^i(\xi), u >\right) \mu^i(d\xi) \tag{2.112}$$

- *The generator $\mathcal{A}$ of the affine process $\{\mathbf{X}\}_{t\geq 0}$*

$$\begin{aligned} \mathcal{A}f(x) \;&=\; \frac{1}{2}\sum_{k,l=1}^{d}\left(a_{kl} + \sum_{i=1}^{m}\alpha_{kl}^i x_i\right)\frac{\partial^2 f(x)}{\partial x_k \partial x_l} + \\ &+\; < b, \sum_{i=1}^{d}\beta^i x_i \nabla f(x) > - \left(c + \sum_{i=1}^{m}\gamma^i x_i\right) f(x) + \\ &+\; \int_{\mathbf{E}\setminus\{0\}} [f(x+\xi) - f(x) - < h(\xi), \nabla f(x) >] m(d\xi) + \\ &+\; \sum_{i=1}^{m}\int_{\mathbf{E}\setminus\{0\}} [f(x+\xi) - f(x) - < \chi^i(\xi), \nabla f(x) >] x_i \mu^i(d\xi) \end{aligned} \tag{2.113}$$

$\forall f \in C_0^2(\mathbf{E})$ *and* $x \in \mathbf{E}$.

Chapter 3

Affine Processes in Financial Mathematics

The first systematic introduction of affine process in financial mathematics came with the paper by D. Duffie, D. Filipovic & W. Schachermayer, *Affine Processes and Applications in Finance* [DFS].

We start by introducing some basic elements in financial mathematics.

3.1 Basic Elements of Financial Mathematics

3.1.1 Basic Financial Terms and Models

In this section we will define some basic financial terms and the notion of financial modeling:

Definition 3.1.1.

1. **Derivative**: *is a financial instrument whose value depends on the value of other basic assets, such as common stock.*

2. **Option**: *is a contract which gives the buyer (the owner) the right, but not the obligation, to buy or sell an underlying asset at a specified strike price on or before a specified date.*

3. **Long position**: *is when a trader buys an option contract.*

4. **Short position**: *is when a trader sells an option contract.*

5. **European call option**: *is a financial contract between two parties, the buyer and the seller of this type of option. The buyer of the call option has the right, but not the obligation to buy the underlying asset from the seller of the option at the expiration date for a certain price (strike price). The seller (or "writer") is obligated to sell the commodity or financial instrument to the buyer if the buyer decides so. The buyer pays a fee (called premium) for this right.*

Characteristics of Call Option

The characteristic of call option are:

1. **Strike price**: this is the price at which you can buy the stock (if you have bought a call option) or the price at which you must sell your stock (if you have sold a call option)

2. **Expiry date**: this is the date on which the option expires, or becomes worthless, if the buyer doesn't excersice it.

3. **Premium**: this is the price you pay when you buy an option and the price you receive when you sell an option.

Now, we will define the financial modeling.

Definition 3.1.2. Financial modeling: *is the task of building an abstract representation (a model) of a real world financial situation. Which will be explained as follows:*

- *This is a mathematical model designed to represent (a simplified version of) the performance of a financial asset or portfolio of a business, project, or any other investment. Financial modeling is a general term that means different things to different users; the reference usually relates either to accounting and corporate finance applications, or to quantitative finance applications. While there has been some debate in the industry as to the nature of financial modeling whether it is a tradecraft, such as welding, or a science the task of financial modeling has been gaining acceptance and rigor over the years.*

- *Typically, financial modeling is understood to mean an exercise in either asset pricing or corporate finance, of a quantitative nature. In other words, financial modelling is about translating a set of hypothesis about the behavior of markets or agents into numerical predictions.*

Stock Price Model

In order to determine the stock price model we assume the following:

1. The history of the stock is fully reflected in the present price.
2. Markets respond immediately to new information about the stock.

With these two assumptions, one can notice that changes in a stock price follow a Markov process (where only the present value of the variable is relevant for predicting the future).

3.1.2 Black-Scholes-Merton Equation

Black Scholes Merton model is the most famous formula in all finance, it is a mathematical model of a financial market containing certain derivative investment instruments. This formula gives a theoretical estimate of the price of European style options. The theory is derived from a partial differential equation called the Black Scholes equation, which estimates the price of the option over time. The

key idea behind the model is to hedge the option by buying and selling the underlying asset in just the right way, and as a consequence to eliminate risk.

Notation

- t is a time in years, we generally use: now$= 0$ and expiry$= T$.
- S is the price of the stock, which will sometimes be a random variable and other times a constant.(where S_0 is the price of the stock at $t = 0$)
- $V(S, t)$ is the price of a derivative as a function of time and stock price.
- $C(S, t)$ is the European call option.
- X is the strike price (exercise price).
- r is the free-risk interest rate.
- σ is the standard deviation of Log returns (volatility).

Volatility

The volatility σ of a stock is a measure of our uncertainty about returns provided by the stock. It can be defined as the standard deviation of the return provided by a stock in 1 year when the return is expressed using continuous compounding. The Black-Scholes model explains how volatility can be either estimated from historical data or implies from option prices.

The Black Scholes Equation

The BlackScholes equation is a partial differential equation, which describes the price of the option over time. Broadly speaking, the term may refer to a similar PDE that can be derived for a variety of options, or more generally, derivatives. The equation is:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \tag{3.1}$$

This equation (that gained Scholes and Merton the Nobel Prize in economics in 1997) results from the analysis that the option does have a *real* price $V(S,t)$ dependent on the *current price* S of the stock, *irrespective of the (random) behaviour* of the stock's price until the time T of the option's expiry, that can be achieved by following a certain **hedging** (i.e. protection) strategy of buying and selling the stock and option, to completely eliminate risk (on the option) that would otherwise exist.

Black Scholes Formula

The BlackScholes formula calculates the price of European call options, by solving this partial differential equation for the corresponding terminal and the following boundary conditions:

- $C(0,t) = 0$ for all t.
- $C(S,t) \to S$ as $S \to \infty$.
- $C(S,T) = \max\{S - X, 0\}$, which gives the value of the option at the time that the option matures.

Or else the partial differential equation does not have a unique solution. (Knowing that the European call option can be exercised one time on the exercise date, while the American call option can be exercised at any point). Inorder to get the following formula:

$$C(S,t) = N(d_1)S - N(d_2)Xe^{-r(T-t)} \tag{3.2}$$

(which is the difference between what you get and what you pay), where

- $N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz$ is the cumulative probability distribution function for a standard distribution function. In other words it is the probability that a variable with a standard normal distribution will be less than x,$(0 < N(x) < 1)$.

- d_1 is the first input given by:

$$d_1 = \frac{1}{\sigma\sqrt{T-t}} \left[\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t) \right] \tag{3.3}$$

- d_2 is the second input given by:

$$d_2 = \frac{1}{\sigma\sqrt{T-t}} \left[\ln\left(\frac{S}{X}\right) + \left(r - \frac{\sigma^2}{2}\right)(T-t) \right] \tag{3.4}$$

Now we note the following results, based on the conditions where some of the parameters take extreme values:

1. If d_1 and d_2 are very big numbers, due to the reason where S is a very big number or when $\sigma \longrightarrow 0$, then $N(d_1) = N(d_2) = 1$ and thus the Black-Scholes formula becomes $C(S,t) = S - Xe^{-r(T-t)}$.

2. If the ratio of the stock price to the excercise price $\frac{S}{X}$ is increasing, then d_1 and d_2 are increasing so we will get high probabilities which will higher the exercises price. (So that makes sense the relation between the stock price and the exercise price).

3. People who work with options focus on the volatility (the higher the volatility the higher the option price), let us verify these factors in the equation:

   - in d_1: if σ increases then d_1 will be increased.
   - in d_2: if σ increases then d_2 will be decreased.

   which means that what we should get $(N(d_1)S)$ increases and what we should pay $(N(d_2)Xe^{-r(T-t)})$ decreases.

$$\text{as a result:} \begin{cases} \text{if } \sigma \uparrow & \Rightarrow \quad C \uparrow \\ \text{if } \sigma \downarrow & \Rightarrow \quad C \downarrow \end{cases} \qquad \text{(conversely)}$$

Conclusion

- **Advantage:** The main advantage of the Black Scholes model is speed, it lets you calculate a very large number of option prices in a very short time.
- **Limitation:** The Black Scholes model has one major limitaion, it can not be used to accurately price options with American-style exercise as it only calculates the option price at one point in time at expiration. It does not consider the steps along the way where there could be the possibility of early exercise of an American option.

3.2 Affine Processes and Financial Mathematics

Affine processes are very useful in financial modeling, due to the flexibility of their properties when used to *calculate* terms that are relevant in the stochastic description of the financial markets.

3.2.1 Discounting

Discounting is a financial mechanism in which a debtor obtains the right to delay payments to a creditor, in return for charges that are usually calculated in terms of interest rates.

Such (interest) rates are formulated according to an (affine) function involving an $\mathbf{R}^d-$valued process $\mathbf{X} = \{X_t\}_{t\geq 0}$ by

$$r_t = r(X_t) = r + < \lambda, X_t >, \qquad r \in \mathbf{R}, \quad \lambda \in \mathbf{R}^d. \tag{3.5}$$

Given this, the price of a claim, given in terms of the interest rate process $\mathbf{X}$, in the form $f(X_t)$, where $f \in b\mathbf{E}$ (the Banach space of bounded complex-valued Borel-measurable functions f on $\mathbf{E}$), would be given by an expectation (where $t = 0$ is the present, and t is in the future)

$$\boxed{Q_t f(x) = \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds} f(X_t) \right]} \tag{3.6}$$

where the term $e^{-\int_0^t r(X_s)ds}$ is a continuous compounding term, and x is the starting state (i.e. the starting price now). It is clear that $Q_t f(x) - Q_0 f(x) = Q_t f(x) - x$ is about *discounting*, and that the family $\{Q_t\}_{t \geq 0}$ forms a semi-group.

In this model for discounting, it is essential to be able to calculate the integral for $Q_t f(x)$ above, and this seems to be very tractable if the process $\mathbf{X}$ is taken to be a **regular affine process** (with the resulting properties) (see [DFS]).

Several financial modeling use the special properties of affine processes, because of their flexibility and tractability in capturing certain stochastic properties that are obvious in many financial markets such as jumps, stochastic volatility, optimal dynamic portfolio and consumption choices.

3.2.2 The Term Structure of Interest Rates

One of the central objects in finance is the term structure $t \longrightarrow Q_t 1$ of prices of "bonds", assets that pay 1 unit of account at a given maturity t. Then we can build up the prices of bonds that make payments at multiple dates, and other "fixed-income" securities.

A typical model of the price processes of bonds of various maturities begins with a discount rate process $\{r(X_t) : t \geq 0\}$ defined by the above affine function $x \longrightarrow r(x)$, and where the discount factor $\mathbb{E}\left[e^{-\int_0^t r(X_s)ds} | X_s \right]$ is well defined and is of the anticipated exponential-affine form in X_s.

In general, since $e^{<0,x>} = 1$, the bond price

$$Q_t 1(x) = e^{A(t) + <B(t),x>} \tag{3.7}$$

is calculated from the generalized Recatti equation for a broad range of affine processes. However for $(d > 1)$ we have variant results, beside this there is "infinite-dimensional affine term structure model".

The Vasiček (Gaussian Ornstein-Uhlenbeck) process or the Cox-Ingersoll-Ross process are important simple models of interest short rate $r(X)$, which is the continuous branching diffusion of Feller. (These short rate processes are affine $(r(x) = x)$).

Vasiček Short Rate Model

The state space is $\mathbf{R}$, i.e.$(d = 1)$, and we set $r = X$ for the Vasiček short rate model:

$$dr_t = a(b - r_t)dt + \sigma dB_t \tag{3.8}$$

where

- $a(b - r_t)$ is a drift factor that represents the expected instantaneous change in the interest rate at time t.
- a and b are some constants.
- σ is the instantaneous volatility that measures the amplitude of randomness entering the system.
- B is a brownian motion.

Vasiček gives an explicit characterization of the term structure of interest rates in an efficient market. The model is widely use for pricing the bond. Additionally it uses Ornstein-Uhlenbeck process to compute the spot interest rate. This model is a one-factor model which means that rates depend on the spot interest rate. Thus the spot rate defined the whole term structure.

Beside the general characteristics, the main advantage of this model is to value all interest rates contingent claims in a consistent way. While the disadvantage is that it involves unobservable parameters and do not provide a perfect fit to the initial term structure of interest rate.

3.2.3 Default Risk

This is about the risk that a financial contract *defaults*, i.e. *fails to pay what it has to pay*. The idea is to model the *time-of-default* τ of financial contracts, and (for example) the probability distribution of the total default losses on a portfolio of financial contracts.

Doubly Stochastic Models

In *doubly stochastic models*, the observed random variable can be modelled in two stages.

- The distribution of the observed outcome is represented in a fairly standard way using one or more parameters.
- Some of these parameters are treated as being themselves random variables.

This is essentially the same as the well-known concept of compounded distributions. As an example: the observed values in a point process might be modelled as a Poisson process in which the rate (the relevant underlying parameter) is treated as being the exponential of a Gaussian process.

Consider, for example, a defaultable bond with maturity time t, that pays $1_{\tau>t}$, i.e. it pays one-unit if t is smaller than the (expected) time τ of default. Then, it can be shown that this defaultable bond has a price given by

$$\mathbb{E}^x\left[e^{-\int_0^t r(X_s)ds}1_{\{\tau>t\}}\right] = \mathbb{E}^x\left[e^{-\int_0^t (r(X_s)+\Lambda(X_s))ds}\right] \tag{3.9}$$

where $r(X_t)$ is as in the previous case, with $\mathbf{X}$ being an interest rate process, with $\Lambda(x)$ being an affine function that is the *intensity* i.e.

$$\lim_{\Delta x\to 0}\frac{Pr(\text{one event in } [\mathrm{x}, \mathrm{x}+\Delta\mathrm{x}]}{\Delta x}, \tag{3.10}$$

of a process $\mathbf{N}$ that is Poisson, conditional on $\mathbf{X}$. ($\mathbf{N}$ would also be called as a *doubly stochastic process*.)

Also, the probability of survival, i.e. $\mathbb{P}^x(\tau > t)$, can be shown to be given by

$$\mathbb{P}^x(\tau > t) = \mathbb{E}^x\left[e^{-\int_0^t \Lambda(X_s)ds}\right]. \tag{3.11}$$

It is also found that having a regular affine process $\mathbf{X}$ to model interest rates is very effective in calculating the above expectation and probability in default risk.

Moreover, we can generalize for a model of the default times $\tau_1, \cdots, \tau_k$ of $k > 1$ different financial contracts, we suppose that τ_i is the first jump time of a non-explosive counting process N_i with respective intensity $\{\Lambda_i(X_{t-}) : t \geq 0\}$,for affine $x \longrightarrow \Lambda_i(x) \geq 0$, where $N_1, \cdots, N_k$ are doubly stochastic driven from X, one can view $(X, N_1, \cdots, N_k)$ as an affine process. Then for any sequence $t_1, \cdots, t_k \in \mathbf{R}_{\geq 0}$ we get

$$\mathbb{P}^x(\tau \geq t_1, \cdots, \tau \geq t_k) = \mathbb{E}^x\left[e^{-\int_0^{t_k} \Lambda(X_s,s)ds}\right], \tag{3.12}$$

where

$$\Lambda(x,s) = \sum_{\{i:s\leq t_i\}} \Lambda_i(x). \tag{3.13}$$

Using the law of iterated expectations, the joint distribution of the default time is given by:

$$\mathbb{P}^x(\tau \geq t_1, \cdots, \tau \geq t_k) = e^{\Phi_0 + <\Psi_0, x>}, \tag{3.14}$$

where Φ_i and Ψ_i are defined inductively by $\Phi_k = 0$ and $\Psi_k = 0$, and

$$e^{\Phi_i + <\Psi_i, x>} = \mathbb{E}^x\left[e^{-\int_0^{t_{i+1}-t_i} \Lambda(X_t, t_i+t)dt}\, e^{\Phi_{i+1} + <\Psi_{i+1}, X_{t_{i+1}-t_i}>}\right] \tag{3.15}$$

for $t_0 = 0$ and since $x \longrightarrow \Lambda(x,s) \geq 0$ is affine with constant coefficients for some $s \in [t_i; t_{i+1}]$, one can calculate the probability distribution of the total default losses on a portfolio of financial contracts.

However "the first default time"τ^* =inf$\{\tau_1, \cdots, \tau_k\}$,satisfies

$$\mathbb{P}^x(\tau^* \geq t) = \mathbb{E}^x \left[e^{-\int_0^t \Lambda^*(X_s)ds} \right], \tag{3.16}$$

where

$$\Lambda^*(x) = \sum_{i=1}^{k} \Lambda(x) \tag{3.17}$$

also has the same form as the bond price calculation (3.7).

3.2.4 Option Pricing

A put option gives its owner the right, but not the obligation, to sell an underlying asset at some future time t at a fixed pre-arranged exercise price $K \in \mathbf{R}_{\geq 0}$, which has the form $f(X_t)$ for some non-negative $f \in C(\mathbf{E})$ (the space of complex-valued continuous functions f on $\mathbf{E}$). Like wise we can treat the corresponding call option to buy the asset. We note that "the option is rationally exercised if and only if $f(X_t) \leq K$, with payoff

$$g(X_t) = max(K - f(X_t), 0), \tag{3.18}$$

and the initial price

$$\begin{aligned} Q_t g(x) &= \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds} g(X_t) \right] \\ &= K\, \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds} 1_{\{f(X_t) \leq K\}} \right] \\ &- \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds} f(X_t) 1_{\{f(X_t) \leq K\}} \right] \end{aligned} \tag{3.19}$$

By using the affine modeling approach to computational advantage provided $f(x) = ke^{<b,x>}$ for ($k \in \mathbf{R}_{\geq 0}$ and $b \in \mathbf{R}^d$), an example of which is the bond price $f(x) = e^{A(T-t)+<B(T-t),x>}$ of (3.7) of time t and maturity date $T > t$. Then $Q_t g(x)$ becomes:

$$G_{ab}(q) = \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds}\, e^{<a,X_t>} 1_{\{<b,X_t> \leq q\}} \right] \tag{3.20}$$

where $(a, b, q) \in \mathbf{R}^d \times \mathbf{R}^d \times \mathbf{R}$ such that ($q = \log K - \log k$).

We consider $G_{a,b}(.)$ the distribution function of $< b, X_t >$ with measure $e^{-\int_0^t r(X_s)ds}\, e^{<a,X_t>}\mathbb{P}^x$, so we can calculate the transform

$$\mathcal{G}_{a,b}(z) = \int_{-\infty}^{+\infty} e^{izq} G_{a,b}(dq) \tag{3.21}$$

then Fourier inversion method can be used to compute $G_{a,b}(q)$, one can notice that

$$\begin{aligned} \mathcal{G}_{a,b}(z) &= \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds}\, e^{<a,X_t>}\, e^{iz<b,X_t>} \right] \\ &= \mathbb{E}^x \left[e^{-\int_0^t r(X_s)ds} f_u(X_t) \right], \end{aligned} \tag{3.22}$$

where $u = a + izb$ and the generalized recatti equations give the solution under the non-negativity of $r(X)$.

Heston Approach to Option Pricing

Heston tried to extend the Black-Scholes model, where the underlying price process is a geometric brownian motion, the underlying asset price is e^{Z_t} and (Y, Z) is the affine $(m = n = 1)$ defined by the following 2 corrolated brownian motion model. The first talks about the evolution of the volatility (assuming that the volatility of the asset is not constant, nor even deterministic, but follows a random process). and the second talks about the evolution of the price.

$$\begin{aligned} dY_t &= (b_1 - \beta Y_t)dt + \sigma\sqrt{Y_t}dW_t^{(1)} \\ dZ_t &= b_2 dt + \sqrt{Y_t}(\rho dW_t^{(1)} + \sqrt{1-\rho^2}dW_t^{(2)}), \end{aligned} \tag{3.23}$$

for real constants $-1 < \rho = corr(dW_t^{(1)}, dW_t^{(2)}) < 1$, $b_1, \sigma \geq 0$ and b_2,β, and where $(W^{(1)}, W^{(2)})$ is a standard brownian motion in $\mathbf{R}^2$. In Black-Scholes case of a geometric Brownian price process e^Z, the stochastic volatility process Y is constant. For Heston's model, the Fourier transform $\mathcal{G}_{a,b}(.)$ is computed explicity.

Similarly, a defaultable option may be priced by replacing $r(X_t)$ with $r(X_t) + \Lambda(X_t)$, where $\{\Lambda(X_{t-}) : t \geq 0\}$ determines the intensity, as for defaultable bond pricing.

Affine Generalizations of the Heston-Model

Numerous affine generalizations of the Heston model have been directed toward more realistic stochastic volatility and jump behavior.

The work of Heston (1993) led to the development of stochastic volatility models. The Heston model is one of the most widely used stochastic volatility models today. There are many empirical, economic and mathematical reasons for using a model with such a form for investigation the volatility on the market. Empirical studies have shown that an assets log-return distribution is non-Gaussian. It is characterized by heavy tails and high peaks. It is also observed that equity returns and empirical volatility are negatively correlated.

Then we investigate classical and nonclassical stochastic volatility models with respect to their extreme behavior. We show that classical stochastic volatility models driven by Brownian motion can model heavy tails, but obviously they are not able to model volatility jumps. Such phenomena can be modelled by Lévy driven volatility processes as, for instance, by Lévy driven Ornstein-Uhlenbeck models. They can capture heavy tails and volatility jumps. Also volatility clusters can be found in such models, provided the driving Lévy process has regularly varying tails. This results then in a volatility model with similarly heavy tails.

Affine Jump-Diffusion and Numerical Methods Based on Fourier Inversion

We consider a Heston model with jumps in returns and variance and obtain closed-form solutions for a wide range of options on the realized variance. Analytical solution through Fourier transform for the density of the realized variance in the Heston model was obtained by Lipton (2001). However, by employing his method we need to discretisize the state space of future realized variance and for each state evaluate its probability by inverting the Fourier integral, and finally compute the expected value of the option by convoluting the pay-out in the given state with the probability of this state. We apply the generalized Fourier transform and reduce the computation of a single option price to numerical inversion

of a single Fourier integral. The generalized Fourier transform was extended for pricing options on the realized variance.

The purpose of introducing jumps in returns and variance dynamics is to make the Heston model consistent with short-term variance swaps with cap protection for which market prices are typically lower than theoretical prices implied by the Heston model with no jumps. Among others, the empirical study of the VIX time series shows that jumps in the dynamics of the S and P 500 index variance are statistically significant. Variance jumps are also necessary to produce positive volatility skews implied from market prices.

References

[RW1] L.C.G. Rogers & D. Williams, *Diffusions, Markov Processes and Martingales*. Cambridge University Press, 2000 (volume 1: Foundations)

[RW2] L.C.G. Rogers & D. Williams, *Diffusions, Markov Processes and Martingales*. Cambridge University Press, 2000 (volume 2: Itô Calculus)

[KR] M. Keller-Ressel, *Affine Processes - Theory and Applications in Finance*. Ph.D. Dissertation, Vienna University of Technology, 2008

[DFS] D. Duffie, D. Filipovic & W. Schachermayer, *Affine Processes and Applications in Finance*. The Annals of Applied Probability, 13(3), pp. 984-1053, 2003

[CT] C. Cuchiero & J. Teichmann, *Path Properties and Regularity of Affine Processes on General State Spaces*. arXiv:1107.1607v2 [math.PR], 16 Jan 2013.

[H] J.C. Hull, *Options, Futures, and other Derivatives*. Prentice-Hall, 7^{th}–edition, 2009

[HPS] P.G. Hoel, S.C. Port & C.J. Stone, *Introduction to Stochastic Proceses*. Waveland Press, 1987

[FKL] Vicky Fasen, Claudia Kluppelberg, Alexander Lindner, *Extremal behavior of stochastic volatility models*. Stochastic Finance, 2006.

[P] Ser-Huang Poon, *The Heston Option Pricing Model*. June 27, 2011

[L] Merrill Lynch, *Pricing Options on Realized Variance in the Heston Model with Jumps in Returns and Volatility*. Journal of Computational Finance, 2008.