Ca' Foscari
University
of Venice

Master's Degree programme
in **Economics and Finance**

Final thesis

Stochastic filtering: a functional and probabilistic approach

Supervisor
Ch. Prof. Antonella Basso

Graduand
Gianmarco Vidotto
Matriculation Number 842271

Academic Year
2016 / 2017

To my family

Contents

Introduction

In this thesis we will try to study the so-called "Stochastic filtering" problem. We will work in the continuous time setting and we will try to be mathematically rigorous.

In our context, filtering means using a set of "dirty" observations (dirty means "affected by noise") and try to gather some information about an unobservable process. Basically we have to filter this observations and try to distinguish the "real" compononent (which contains useful informations on the generator process) from the noise.

Nowadays, the filtering theory is extensively used in a broad range of applications: from the aerospace industry to GPS-tracking.

The structure is as follows.

In the first chapter we will try to give a sound and rigorous background on stochastic calculus and on the theory of Stochastic Differential Equations (SDEs). We will state the classical theorems and we will provide the proofs of the most interesting (and maybe less known) results. We will try to collect here all the probability one needs to know to read, in a profitable way, the remaining chapters.

In the second chapter we will study the filtering problem in a linear setting using a functional approach. We will prove and comment all the steps towards the derivation of the famous Kalman-Bucy filter. In the end, there will be two simple applications (with MATLAB) in order to show the main drawbacks of the continuous time setting and we will gave a little taste of the discrete time version of Kalman filter.

Finally, in the last chapter, we will study the non-linear case of the filtering problem using a probabilistic approach.

Chapter 1

Probability and SDE theory

In this chapter I introduce some useful results on probability and measure theory I will use later on and I summerize the most important results about the theory of Stochastic Differential Equation (SDE). This chapter is deliberately dense since I want to prepare the background and introduce all the notation.

1.1 Stochastic Process

A probability space, denoted as $(\Omega, \mathcal{F}, P)$ is a triple where Ω is a non empty sey, $\mathcal{F}$ is a σ-algebra on Ω and P is a probability measure. A set $A \subset \Omega$ is said to be measaurable (w.r.t $\mathcal{F}$) if $A \in \mathcal{F}$. A measaurable set B is said to be *negligible* if $P(B) = 0$. It is common, in probability theory, to denote the elements of $\mathcal{F}$ as *event*. Using this terminology, we say that an event A happens *almost surely* (a.s) if $P(A) = 1$. Finally, P is said to be complete if every subset of a negligible event is measurable. The assumption of completeness is not restrictive as we can always complete a measure space.

Let I be a real interval of the form $[0, T]$ or $[0, +\infty[$. A stochastic process on $\mathbb{R}^N$ is a collection of random variables with values in $\mathbb{R}^N$ such that the map

$$X : I \times \Omega \to \mathbb{R}^N, X(t, \omega) = X_t(\omega) \tag{1.1}$$

is measurable w.r.t the product σ-algebra $\mathcal{B}(I) \otimes \mathcal{F}$. X is integrable if $X_t \in$

$L^1(\Omega, P)$ for every $t \in I$.

It is useful to highlight that, for every fixed $\omega \in \Omega$, we have a so called *trajectory* (a function of time t).

A stochastic process is said to be *continuos* (a.s) if its trajectories are continuos function for all $\omega \in \Omega$ (for almost all $\omega \in \Omega$).

A stochastic process is said to be *right-continuos* (a.s) if its trajectories are right-continuos function for all $\omega \in \Omega$ (for almost all $\omega \in \Omega$).

A filtration $(\mathcal{F}_t)_{t\geq 0}$ is an increasing family of sub σ-algebras of $\mathcal{F}$. The natural filtration for a stochastic process X is defined as

$$\tilde{\mathcal{F}_t^X} = \sigma(X_s^{-1}(H)|0 \leq s \leq t, H \in \mathcal{B}) \tag{1.2}$$

We say that a process is *adapted* to $\{F\}$ if X_t is $\mathcal{F}_t$-measurable for every t. Every process is adapted to its own natural filtration.

Let X, Y be two stochastic process defined on the same probability space $(\Omega, \mathcal{F}, P)$. We say that X is a *modification* of Y if $X_t = Y_t$ almost surely for every t.

This means that, if we set

$$N_t = \{\omega \in \Omega| \quad X_t(\omega) \neq Y_t(\omega)\} \tag{1.3}$$

we have that $N_t \in \mathcal{N}$ for any $t \geq 0$. $\mathcal{N}$ represents the set of negligible event (those events whose probability is 0).

Two stochastic processes are indistinguishable if almost all the trajectories coincide. This means that, if we set

$$N = \{\omega \in \Omega| \quad X_t(\omega) \neq Y_t(\omega) \quad \forall t\} \tag{1.4}$$

we have that $N \in \mathcal{N}$.

Clearly, if X and Y are two indistinguishable processes, they are also modifications. The reverse is not true. Indeed, we can write the set N below as

$$N = \bigcup_{t\geq 0} N_t \tag{1.5}$$

but this is an uncountable union (not necessarility measurable). Anyway, it's easy to show that, in case of a.s. right-continuous processes the two notions coincide (see, for example, [10].

We state now another useful result we will use when we define the class of processes for which the Itô integral makes sense.
We say that a stochastic process X is *progressively measurable* w.r.t the filtration $\mathcal{F}_t$ if, for every t, $X : [0,t] \times \Omega$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$ measurable. This means that

$$\{(s,\omega) \in [0,t] \times \Omega | X_s(\omega) \in H\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t, \quad H \in \mathcal{B} \tag{1.6}$$

This definition will be used later. Here we only note that a progressively measurable process is measurable and adapted (Fubini-Tonelli theorem). The converse is not trivial.

However, we have that every right-continuous and adapted process is progressively measurable. The idea of the proof is to construct a sequence of simple processes who converges pointwise to X on $[0,t]$ and to show that they're measurable w.r.t. the product σ-algebra. Since X is defined as the pointwise limit of measurable functions, it's measurable as well.

1.2 Martingale spaces: $\mathcal{M}^2$ and $\mathcal{M}^2_c$

Before presenting the main results concerning the martingale spaces, we introduce the so-called *Doob's inequality.*

Let M be a right continuous martingale and $p > 1$. Then, for every T we have that

$$\mathbb{E}[\sup_{t\in[0,T]} |M_t|^p] \leq q^p \mathbb{E}[|M_T|^p] \tag{1.7}$$

where $q = \frac{p}{p-1}$ is the conjugate exponent to p.

The inequality can be proved easily in discrete time. Then, using the Beppo-Levi theorem (see, for instance, [2]), we can shift to the continuous case.

Now, we introduce the martingale spaces. These spaces are foundamental

because we'll se that the Itô Integral is defined as a limit in $\mathcal{M}_c^2$.

Thus, for a fixed $T > 0$ we denote by

1. $\mathcal{M}^2$ the linear space of right-continuous $\mathcal{F}_t$-martingales $(M_t)_{t\in[0,T]}$ such that $M_0 = 0$ a.s. and such that

$$[\![M]\!]_T = \sqrt{\mathbb{E}\big[\sup_{t\in[0,T]} |M_t|^2\big]} \tag{1.8}$$

   is finite

2. $\mathcal{M}_c^2$ the subspace of continuous martingales in M^2

Here the analogy with the L^p spaces is evident. In particular we can see that $[\![M]\!]_T$ is a seminorm in $\mathcal{M}^2$ because it's equal to zero for all those stochastic process which are indistinguishable from the null process.

For our purposes, the most relevant fact is that the space $(\mathcal{M}_c^2, [\![\cdot]\!]_T)$ is complete (every cauchy sequence converge). Moroever, $\mathcal{M}_c^2$ is a closed subspace of $\mathcal{M}^2$.

The proof is instructive so I reported it here.

Given a cauchy sequence (M^n) in $\mathcal{M}^2$, we want to prove that there exists a convergent subsequence to conclude that also the orginal sequence converge.

We construct a subsequence (M^{k_n}) such that

$$[\![M^{k_n} - M^{k_{n+1}}]\!]_T \leq \frac{1}{2^n}, \quad n \geq 1 \tag{1.9}$$

and we set $\nu_n = M^{k_n}$. Then we define

$$\kappa_N(\omega) = \sum_{n=1}^{N} \sup_{t\in[0,T]} |\nu_{n+1}(t,\omega) - \nu_n t, \omega|, \quad N \geq 1 \tag{1.10}$$

that is an increasing sequence of non-negative function.

Moreover, we remember that, if $f, g : A \subset \mathbb{R} \to \mathbb{R}$ are positive function in their domain, then $\sup(f \cdot g) = \sup(f) \cdot \sup(g)$. Thus, using this result, if we square the expression for $\kappa_N(\omega)$ and we take the expected value we have

that

$$\mathbb{E}[\kappa_N^2] \leq 2\sum_{n=1}^{N}[\![\nu_{n+1}-\nu_n]\!]_T^2 \tag{1.11}$$

which, using the way we constructed the sequence, is less than 2. Thus, we have that $\mathbb{E}[\kappa_N^2] \leq 2$. Now, using the Beppo-Levi theorem we have that

$$\kappa(\omega) = \lim_{n\to\infty} \kappa_N(\omega), \quad \omega \in \Omega \tag{1.12}$$

and $\mathbb{E}[\kappa^2] \leq 2$. Thus, even if $\kappa(\omega)$ can take infinite values, we know that the set in which can be equal to $+\infty$ is negligible.

Thus, formally, we can say that $\kappa(\omega) < +\infty$ for every $\omega \in \Omega \setminus F$ where $P(F) = 0$.

Now, using the triangle inequality of the absolute value and the fact that $sup(f+g) \leq sup(f) + sup(g)$ we can write

$$\sup_{t\in[0,T]} |\nu_{n+1}(t,\omega) - \nu_n t, \omega| \leq \kappa(\omega) - \kappa(\omega)_{m-1} \tag{1.13}$$

for $n \geq m \geq 2$. This means that $(\nu_n(t,\omega))$ is a Cauchy sequence in $\mathbb{R}$ for every $t \in [0,T]$ and $\omega \in \Omega \setminus F$. Since, for every $\omega \in \Omega \setminus F$, $\nu_n(t,\omega)$ is a right-continuous function of t and the convergence is uniform in t, there exist a process $M(t,\omega)$ which is indistinguishable from a right-continuous process (continuous in case $M^n \in \mathcal{M}_c^2$.

1.3 Usual hypotheses

We say that $(\mathcal{F}_t)$ satisfies the so-called "usual hypotheses" w.r.t P if:

1. $\mathcal{N} \subset \mathcal{F}_0$
2. the filtration is right continuous, i.e. for every $t \geq 0$

$$\mathcal{F}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon} \tag{1.14}$$

The two statements below are really strong and can lead, as we will see, to

counter-intuitive result. Intuitively, the first one means that we always know (even at time $t = 0$) what is possible and what is not.

Let X and Y be two random variables such that X is equal to Y almost surely ($P(X = Y) = 1$). If X is $\mathcal{F}_t$-measurable, is Y $\mathcal{F}_t$-measurable ? Not necessarily, unless we introduce the first assumption (and the completeness of the measure). Indeed, if we set $A = \{\omega \in \Omega | X \neq Y\}$, we have that

$$Y^{-1}(H) = \{\omega \in \Omega \setminus A | X^{-1}(H)\} \cup \{\omega \in A | Y^{-1}(H)\} \tag{1.15}$$

where the first set is measurable by assumption and the second has null measure (and so, by assumption, is contained in the filtration for every t)

The second assumption has even bigger implications for stochastic calculus. We'll see when we'll talk about stopping times. Here, we just note that if a random variable X is $\mathcal{F}_s$-measurable for every $s > t$, then X is also $\mathcal{F}_t$-measurable.

Given a stochastic process X on the space $(\Omega, \mathcal{F}, P)$, we set

$$\mathcal{F}_t^X = \bigcap_{\epsilon > 0} \hat{\mathcal{F}}_{t+\epsilon}^X \tag{1.16}$$

where $\hat{\mathcal{F}}_t^X = \sigma(\tilde{\mathcal{F}}_t^X \cup \mathcal{N})$. It can be verified that the filtration above verifies the usual hypotheses.

1.4 Stopping times

We introduce now a concept we will encounter a number of times later on. A random variable

$$\tau : \Omega \to \mathbb{R}_{\geq 0} \cup \{\infty\} \tag{1.17}$$

is called stopping time w.r.t the filtration $(\mathcal{F}_t)$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for every t.

It's easy to see that if τ is a stopping time then $\{\tau < t\} \in \mathcal{F}_t$.

To prove the reverse, we note that for every $\epsilon > 0$ we have that

$$\{\tau < t\} = \bigcap_{\{n \in \mathbb{N} | \frac{1}{n} < \epsilon\}} \{\tau < t + \frac{1}{n}\} \tag{1.18}$$

so that $\{\tau < t\} \in \mathcal{F}_{t+\epsilon}$. Consequently, in view of the usual hypotheses

$$\{\tau < t\} \in \mathcal{F}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon} \tag{1.19}$$

A particularly useful stopping time is the so-called hitting time of an open set in $\mathbb{R}^N$.

Let X a right-continuous, $\mathcal{F}_t$ adapted process in $\mathbb{R}^N$ and H an open set in $\mathbb{R}^N$. Let us set

$$I(\omega) = \{t \geq 0 | X_t(\omega) \in H\} \tag{1.20}$$

and

$$\tau(\omega) = \begin{cases} \inf I(\omega) & \text{if} I(\omega) \neq \emptyset \\ +\infty & \text{if} I(\omega) = \emptyset \end{cases} \tag{1.21}$$

So, if $\tau = t$, we know that the process touched the *boundary* of the open set at time t and then entered inside but we highlight that $X_t \notin H$.

In order to show that τ is stopping time, in light of the observations made above, we have to show that $\{\tau < t\} \in \mathcal{F}_t$.

We state that

$$\{\tau < t\} = \bigcup_{s \in \mathbb{Q} \cap [0,t[} \{X_s \in H\} \tag{1.22}$$

Indeed, let's suppose that there exist $\bar{\omega}$ and $t_0 < t \in \mathbb{R} \setminus \mathbb{Q}$ such that $X_{t_0}(\bar{\omega}) \in H$. Since H is open, there will exist $\bar{r}$ such that the ball $B(X_{t_0}(\bar{\omega}), \bar{r}) \subset H$. Moreover, since X_t is right continuous we know that there exist $\delta > 0$ such that $\forall t$ such that $t - t_0 < \delta$ we have $X_t \in B(X_{t_0}(\bar{\omega}), \bar{r})$. Now, pick $t_1 \in (t_0, t_0 + \delta)$ such that $t_1 < t$ and $t_1 \in \mathbb{Q}$ (We cal always do this because of the density of the irrational numbers). Then we have that $X_{t_1} \in H$. This concludes the proof.

We give now some usuful notion we will use later on when we define the stochastic integral with a stopping-time as upper integration limit.

Let τ be a stopping time which is finite on $\Omega \setminus N$ where N is a negligible event and let X be a stochastic process. We set

$$X_\tau(\omega) = X_{\tau(\omega}(\omega), \quad \omega \in \Omega \tag{1.23}$$

Thus, $\forall \omega \in \Omega \backslash N$ we have a random variable. When $\tau = \infty$ the expression above is not well defined (what is X_∞ ?) but since N is negligible, we don't care too much.

Then, we define the σ-algebra associated to the stopping time τ as

$$\mathcal{F}_\tau = \{F \in \mathcal{F} | F \cup \{\tau \leq t\} \in \mathcal{F}_t \forall t\} \tag{1.24}$$

For every stopping time τ and $n \in \mathbb{N}$, the equation

$$\tau_n(\omega) = \begin{cases} \frac{k+1}{2^n} & \text{if} \quad \frac{k}{2^n} \leq \tau(\omega) < \frac{k+1}{2^n} \\ +\infty & \text{if} \quad \tau(\omega) = +\infty \end{cases} \tag{1.25}$$

1.5 Brownian motion

We introduce now a foundamental process for stochastic calculus. Since Brownian motion and its properties are well-known, I just stated here without any proof (for a more detailde analysis see, for instance, [7]).

Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ be a filtered probability space. A real brownian motion is a stochatic process $W = (W_t)_{t \geq 0}$ in $\mathbb{R}$ such that

- $W_0 = 0$ a.s
- W is $\mathcal{F}_t$-adapted and continuos
- for $t > s \geq 0$, the random variable $W_t - W_s$ has normal distribution $\mathcal{N}_{0,t-s}$ and is independent of $\mathcal{F}_s$

It's worth noting that there exist more brownian motion defined on different probability spaces. The existence is not trivial and some construction can be found in the literature.

In order to understand the difference between ordinary calculus and stochastic calculus we need to remember the concept of bounded variation and quadratic variation.

Let us consider a function $g : [a, b] \to \mathbb{R}^n$ and a partition $\zeta = \{t_0, ..., t_N\}$ of $[a, b]$. The variation of g relative to ζ is defined by

$$V_{[a,b]}(g, \zeta) = \sum_{k=1}^{N} |g(t_k) - g(t_{k-1})| \tag{1.26}$$

The *first variation* of g over $[a, b]$ is the supremum taken over all the partitions of $[a, b]$, that is

$$V_{[a,b]}(g) = \sup_{\zeta} V_{[a,b]}(g, \zeta) \tag{1.27}$$

If this quantity is finite, we say that g is a bounded variation (BV) function ($g \in BV([a, b])$).

On the other hand, the quadratic variation of g relative to ζ is defined by

$$V_t^(2)(g, \zeta) = \sum_{k=1}^{N} |g(t_k) - g(t_{k-1})|^2 \tag{1.28}$$

Now, it's possible to show that if $g \in BV([0, t]) \cap C([0, t])$, then

$$\lim_{|\zeta| \to 0} V_t^{(2)}(g, \zeta) = 0 \tag{1.29}$$

Let's come back to our brownian motion.
It can be shown that for any sequence of partitions (ζ_n) with mesh converging to zero (that is $\lim_{n\to\infty} |\zeta_n| = 0$) we have that

$$\lim_{n\to\infty} V_t^{(2)}(W, \zeta_n) = t \tag{1.30}$$

in $L^2(\Omega, P)$.

Now, using a standard result of probability theory, we know that for any such sequence of partitions (ζ_n) there exists a subsequence (ζ_{k_n}) such that

$$\lim_{n\to\infty} V_t^{(2)}(W, \zeta_{k_n}) = t, \quad a.s. \tag{1.31}$$

Thus, in light of the previous result, almost all the trajectories of the brownian motion cannot have bounded variations.

Finally, it can be proved that with probability one, the trajectories of a Brownian Motion are nowhere differentiable.

1.6 Itô Integral

In this section we introduce the construction of the Itô Integral and the main problems in its definition.

Basically, we want to study what happens to this object

$$I_1 = \sum_{k=1}^{N} u_{t_{k-1}}(W_{t_k} - W_{t_{k-1}}) \tag{1.32}$$

when we shift from the discrete to the continuous case. Does there exist the limit for every $\omega \in \Omega$?

It turns out that this is a difficult object to deal with. Indeed, even if assume nice conditions on the integrand process u_t (for example, continuity of the trajectories) we cannot define those object as a Riemann Integral (we can't use the Lagrange theorem because W is nowhere differentiable) and we cannot even define it in the Riemann-Stieltjes sense (W has not bounded variation).

Here, the more general way to construct the Itô Integral. In what follows W is a real Brownian motion on the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ where the usual hypotheses hold and T is a fixed positive number.

We say that a stochastic process u belongs to the class $\mathbb{L}^2$ if:

- u is progressively measurable w.r.t the filtration $(\mathcal{F}_t)$
- $u \in L^2([0,T] \times \Omega)$ that is (using Fubini theorem)

$$\int_0^T \mathbb{E}[u_t^2]dt < +\infty \tag{1.33}$$

A process $u \in \mathbb{L}^2$ is called *simple* if it can be written as follows

$$u_t = \sum_{k=1}^{N} e_k \mathbb{1}_{]t_{k-1},t_k]}(t), \quad t \in [0,T] \tag{1.34}$$

where $0 = t_0 < t_1 < \cdots < t_N = T$ and e_k are random variables on $(\Omega, \mathcal{F}, P)$.

As a consequence of the assumption that $u \in \mathbb{L}^2$, we have that e_k is $\mathcal{F}_{t_{k-1}}$-measurable for every $k = 1, \cdots, N$. Further, $e_k \in L^2(\Omega, P)$.

For this kind of process, we define the Itô Integral as

$$\int u_t dW_t = \sum_{k=1}^{N} e_k (W_{t_k} - W_{t_{k-1}}) \tag{1.35}$$

and, for every $0 \leq a < b \leq T$,

$$\int_a^b u_t dW_t = \int u_t \mathbb{1}_{]a,b]}(t) dW_t \tag{1.36}$$

We briefly explain the main important properties concerning the Itô integral of simple processes. Recalling the tower property of conditional expectation, we have that all the properties listed below are still true in the "non-conditional" version.

Let $u, v \in \mathbb{L}^2, \alpha, \beta \in \mathbb{R}$ and $0 \leq a < b < c \leq T$. The following properties hold:

- Linearity:

$$\int (\alpha u_t + \beta v_t) dW_t = \alpha \int u_t dW_t + \beta \int v_t dW_t \tag{1.37}$$

- Additivity:

$$\int_a^b u_t dW_t + \int_b^c u_t dW_t = \int_a^c u_t dW_t \tag{1.38}$$

- Null expectation:

$$\mathbb{E}\left[\int_a^b u_t dW_t | \mathcal{F}_a\right] = 0 \tag{1.39}$$

- The stochastic process

$$X_t = \int_0^t u_s dWs, \quad t \in [0,T] \tag{1.40}$$

is a continuous $\mathcal{F}_t$-martingale.

- The Itô isometry

$$\mathbb{E}\left[\int_a^b u_t dW_t \int_a^b u_t dW_t | \mathcal{F}_a\right] = \mathbb{E}\left[\int_a^b u_t^2 dt | \mathcal{F}_a\right] \tag{1.41}$$

The last property, the *Itô isometry*, can be written as an equality between norms

$$\|\int_a^b u_t dW_t\|_{L^2(\Omega)} = \|u\|_{L^2([a,b]\times\Omega)} \tag{1.42}$$

This property is foundamental for the construction of the Itô Integral for a more general class of $\mathbb{L}^2$ processes.

Intuitively, if (u^n) is a cauchy sequence in $L^2([0,T]\times\Omega)$, then $(I_T(u^n))$ is a Cauchy sequence in $L^2(\Omega)$ too.

Luckily, it can be proved that, for every $u \in \mathbb{L}^2$ there exist a sequence of simple processes in $\mathbb{L}^2$ such that

$$\lim_{n\to\infty} \|u - u^n\|^2_{L^2([0,T]\times\Omega)} = 0 \tag{1.43}$$

Thus, we select a convergent sequence of simple processes (u^n) approximating $u \in \mathbb{L}^2$. Since it converges, (u^n) is a Cauchy sequence. Thanks to Itô Isometry, even $(I_T(u^n))$ is a Cauchy sequence in $L^2(\Omega)$. Thus, we define the Itô integral as the limit

$$\int_0^T u_t dW_t = \lim_{n\to\infty} I_T(u^n) \quad \text{in } L^2(\Omega) \tag{1.44}$$

With the procedure above, we have defined the stochastic integral only except for a negligible event $N_T \in \mathcal{N}$. Indeed, recall that every element of $L^2(\Omega)$ is unique up to a negligible event. This is a problem in defining the integral as a stochastic process since T belongs to un uncountable set.

Thus, we use the Doob's inequality. We consider the usual sequence of simple processes $(u^n) \in \mathbb{L}^2$ as before. Then, we define

$$I_t(u^n) = \int_0^t u_s^n dW_s, \quad t \in [0, T] \tag{1.45}$$

Using Doob's Inequality we have that

$$[\![I(u^n) - I(u^m)]\!]_T \leq 2\|u^n - v^n\|_{L^2([0,T]\times\Omega)} \tag{1.46}$$

Thus $I_t(u^n)$ is a Cauchy sequence in $(\mathcal{M}_c^2, [\![\cdot]\!])$ that is a complete space. Thus, finally, we can write

$$\int_0^t u_s dW_s = \lim_{n\to\infty} \int_0^t u_s^n dW_s \quad \text{in } \mathcal{M}_c^2 \tag{1.47}$$

The limit is defined up to indistinguishability.

All the properties outlined before in the case of simple processes are still true; in particular, using the Doob-Inequality, we have that if $u, v \in \mathbb{L}^2$ are modifications then their stochastic integrals coincide.

We give here also the definition in the case in which the upper integration limit is a stopping time.

If $u \in \mathbb{L}^2(\mathcal{F}_t)$ and τ is an $\mathcal{F}_t$-stopping time such that $0 \leq \tau \leq T$ a.s then $(u_t \mathbb{1}_{t\leq\tau}) \in \mathbb{L}^2$ and

$$X_\tau = \int_0^\tau u_s dW_s = \int_0^T u_s \mathbb{1}_{s\leq\tau} dW_s \quad \text{a.s} \tag{1.48}$$

The proof is constructive so I report it here.

The stochastic process $\mathbb{1}_{\{\tau\geq t\}}$ is bounded and progressively measurable. Thus, the product $u_t \mathbb{1}_{\{\tau\geq t\}} \in \mathbb{L}^2$. We define

$$Y = \int_0^T u_s \mathbb{1}_{\{\tau\geq s\}} dW_s \tag{1.49}$$

and we want to prove that $X_\tau = Y$ a.s.

First of all, we consider the case of a simple stopping time

$$\tau = \sum_{k=1}^{n} t_k \mathbb{1}_{F_k} \tag{1.50}$$

with $0 < t_1 < t_2 < \cdots < t_n = T$ and $F_k \in \mathcal{F}_{t_k}$ are disjoint events such that

$$F = \bigcup_{k=1}^{n} F_k \in \mathcal{F}_0 \tag{1.51}$$

.

Given the definition of X_τ, we have that $X_\tau = 0$ on $\Omega \setminus F$ and we can write

$$X_\tau = \mathbb{1}_F \int_0^T u_s dW_s - \sum_{k=1}^{n} \mathbb{1}_{F_k} \int_{t_k}^T u_s dW_s \tag{1.52}$$

Now, to go on, we need a useful lemma that we state without proof. Let $u \in \mathbb{L}^2$ and X a $\mathcal{F}_{t_0}$-measaurable random variable for some $t_0 > 0$. Then

$$X \int_{t_0}^T u_t dW_t = \int_{t_0}^T X u_t dW_t \tag{1.53}$$

Now, it's easy to see that $\mathbb{1}_{\{\tau \geq t\}} = (1 - \mathbb{1}_{\{\tau < t\}})$ so that we can write

$$\begin{aligned}
Y &= \int_0^T u_s 1 - \mathbb{1}_{\{\tau < s\}}) dW_s \\
&= \int_0^T u_s dW_s - \int_0^T u_s (\mathbb{1}_{\Omega \setminus F} + \sum_{k=1}^{n} \mathbb{1}_{F_k} \mathbb{1}_{\{s > t_k\}}) dW_s \\
&= \mathbb{1}_F \int_0^T u_s dW_s - \sum_{k=1}^{n} \int_{t_k}^T u_s \mathbb{1}_{F_k} dW_s
\end{aligned}$$

Now, $\mathbb{1}_{F_k}$ is a $\mathcal{F}_{t_k}$-measurable random variable for $k = 1, \cdots, n$ so that we have $X_\tau = Y$.

We can generalize to the case of a general stopping time τ such that

$0 \leq \tau \leq T$ a.s. We define the following decresing sequence of stopping time

$$\tau_n = \sum_{k=0}^{2^n-1} \frac{T(k+1)}{2^n} \mathbb{1}_{F_k} \tag{1.54}$$

where $F_k = \{\frac{Tk}{2^n} < \tau \leq \frac{T(k+1)}{2^n}\}$. Since τ is stopping time, we have that $F_k \in \mathcal{F}_{\frac{T(k+1)}{2^n}}$ and they are disjoint.

Finally, we can easily see that $\bigcup_{k=0}^{2^n-1} F_k$ is $\mathcal{F}_0$-measurable. This because we know that $\{\tau = 0\}$ and $\{\tau > T\}$ are $\mathcal{F}_0$-measurable (the first because of the definition of stopping times and the second because it's a negligible event). We can write

$$\Omega \setminus \{\{\tau = 0\} \cup \{\tau > T\}\} = \bigcup_{k=0}^{2^n-1} F_k \tag{1.55}$$

and this is measurable. So, we're in the previous situation.

Recalling that the stochastic integral is a continuous martingale, we have that X_{τ_n} converges to X_τ a.s. If we define

$$Y^n = \int_0^T u_s \mathbb{1}_{\{\tau_n > s\}} dW_s \tag{1.56}$$

we have that $X_{\tau_n} = Y^n, \forall n \in \mathbb{N}$.

We introduce here the concept of *quadratic variation* of a stochastic integral as

$$X_t = \int_0^t u_s dW_s \tag{1.57}$$

where $u \in \mathbb{L}^2$.

Let X as above. Then for any $t > 0$, there exists the limit

$$\lim_{|\zeta| \to \infty} \sum_{k=1}^{N} |X_{t_k} - X_{t_{k-1}}|^2 = \int_0^t u_s^2 ds \quad \text{in } L^2(\Omega, P) \tag{1.58}$$

and we call it the *quadratic variation process* of X.

Until now, we have defined the Itô Integral of $\mathbb{L}^2$-processes. We now

proceed to generalize this notion to a larger class of stochastic processes and we'll see the notion of *local martingale*.

We denote by $\mathbb{L}^2_{loc}$ the family of processes $(u_t)_{t\in[0,T]}$ that are progressively measurable w.r.t $(\mathcal{F}_t)_{t\in[0,T]}$ and such that

$$\int_0^T u_t^2 dt < \infty \quad \text{a.s} \tag{1.59}$$

The integral above is a random variable: $\forall\omega \in \Omega$ we have a trajectory and we compute the value of the integral. Morover, it is well defined: the assumption of progressively measurability togheter with the Tonelli Theorem ensure that $\forall\omega \in \Omega$ and $\forall t \in [0,T]$ the trajectory is Lebesgue integrable.

It's worth noting that:

- Every progressively measurable and (a.s) right continuous process belongs to $\mathbb{L}^2_{loc}$.
- The space is invariant w.r.t. equivalent probability measures.

This are the explained steps:

- Given $u \in \mathbb{L}^2_{loc}$, we define the process

$$A_t = \int_0^t u_s^2 ds \quad t \in [0,T] \tag{1.60}$$

  A_t is an increasing and continuous process (property of Lebesgue integral)

- For every $n \in \mathbb{N}$ we put

$$\tau_n = \inf\{t \in [0,T] | A_t \geq n\} \wedge T \tag{1.61}$$

  τ_n is a stopping time and (since $u \in \mathbb{L}^2_{loc}$)

$$\tau_n \nearrow T \quad \text{a.s as } n \to \infty \tag{1.62}$$

Moreover, if we set

$$F_n = \{\tau_n = T\} = \{A_T \leq n\} \tag{1.63}$$

we have that

$$\bigcup_{n\in\mathbb{N}} F_n = \Omega \setminus N, \quad N \in \mathcal{N} \tag{1.64}$$

- Now, we set

$$u_t^n = u_t \mathbb{1}_{\{t\leq\tau_n\}}, \quad t \in [0,T] \tag{1.65}$$

and we note that $u_t^n \in \mathbb{L}^2$ since

$$\mathbb{E}\left[\int_0^T (u_t^n)^2 dt\right] = \mathbb{E}\left[\int_0^{\tau_n} (u_t)^2 dt\right] \leq n \tag{1.66}$$

Thus, the process

$$X_t^n = \int_0^t u_t^n dW_t, \quad t \in [0,T] \tag{1.67}$$

is well-defined.

Now, for every $n, h \in \mathbb{N}$ we have that $u^n = u^{n+h} = u$ on F_n.

We give here the definition of local martingale.

A process $M = (M_t)_{t\in[0,T]}$ is a $\mathcal{F}_t$-local martingale if there exists an increasing sequence (τ_n) of $\mathcal{F}_t$-stopping times, called localizing sequence for M, such that

$$\lim_{n\to\infty} \tau_n = T \quad a.s. \tag{1.68}$$

and, for every $n \in \mathbb{N}$, the stochastic process $M_{t\wedge\tau_n}$ is a $\mathcal{F}_t$-martingale.

1.7 Itô process

An Itô process is a stochastic process X of the form

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \quad t \in [0,T] \tag{1.69}$$

where X_0 is a $\mathcal{F}_0$-measurable random variable, $\mu \in \mathbb{L}^1_{loc}$ and $\sigma \in \mathbb{L}^2_{loc}$.

The first integral is an ordinary Lebesgue integral for every $\omega \in \Omega$.

In what follows, we usually write the above expression in "differential" form

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1.70}$$

It's worth noting that every expression in the "integral form" has been justified while the differential form is merely a shortcut.

If X is the Itô process above, its quadratic variation process is given by

$$\langle X \rangle_t = \int_0^t \sigma_s^2 ds \tag{1.71}$$

or, in differential form

$$d \langle X \rangle_t = \sigma_t^2 dt \tag{1.72}$$

We will now introduce two "versions" of the so-called Itô-formula. The proof is not difficult. It is based on a Taylor series expansion. We'll not report it here (see, for instance, [10]).

Let $f \in C^2(\mathbb{R})$ and let W be a real Brownian motion. Then $f(W)$ is an Itô process and we have

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt \tag{1.73}$$

We observe the presence of the extra term $(\frac{1}{2}f''(W_t)dt)$ which makes this "differentiation" rule different from the ordinary one.

The general formulation is as follows.

Let X be the Itô process above and $f = f(t,x) \in C^{1,2}(\mathbb{R}^2)$. Then the stochastic process $Y_t = f(t, X_t)$ is an Itô process and we have

$$df(t, X_t) = \partial_t f(t, X_t)dt + \partial_x f(t, X_t)dX_t + \frac{1}{2}\partial_{xx} f(t, X_t)d \langle X \rangle_t \tag{1.74}$$

A little remark about the notation. Let X be an Itô process

$$dX_t = \mu_t dt + \sigma_t dW_t \tag{1.75}$$

If h is a stochastic process such that $h\mu \in \mathbb{L}^1_l oc$ and $h\sigma \in \mathbb{L}^2_l oc$, we usually write

$$dY_t = h_t dX_t \tag{1.76}$$

instead of

$$dY_t = h_t \mu_t dt + h_t \sigma_t dW_t \tag{1.77}$$

Thus, the stochastic integral w.r.t. an Itô process can be written as

$$\int_0^t h_d X_s = \int_0^t h_s \mu_s ds + \int_0^t h_s \sigma_s dW_s \tag{1.78}$$

We remark that, if $\mu \in \mathbb{L}^1_l oc, \sigma \in \mathbb{L}^2_l oc$ and h is a continuous *adapted* process, then $h\mu \in \mathbb{L}^1_l oc$ and $h\sigma \in \mathbb{L}^2_l oc$. More generally, it's enough that h is progresssively measurable and a.s. bounded.

The two formulas above are extremely useful in stochastic analysis. Indeed, using them, we can compute the differential for a large amount of functions depending on time and an Itô process. We will see later on that we can also solve some Stochastic Differential Equation (SDE) choosing wisely the function $f(x,t)$.

Particularly relevant is the case in which μ, σ are non stochastic processes but deterministic functions of time. Indeed, if $\mu \in L^1(0,+\infty)$ and $\sigma \in L^2(0,\infty)$, the process defined by

$$dS_t = \mu(t)dt + \sigma(t)dW_t \tag{1.79}$$

has a normal distibution with

$$\mathbb{E}[S_t] = S_0 + \int_0^t \mu(s)ds \quad VAR(S_t) = \int_0^t \sigma^2(s)ds \tag{1.80}$$

Computing the two moments above is not difficult: you write the integral form of the Itô process, and employ the property of the Itô integral (null expectation). To compute the variance you use the Itô isometry. On the other hand, to show that S_t has a normal distribution $\forall t$ you have to check that the characteristic function of S_t has the well-known expression of the

characteristic function of a normal variable.

We can also develop a Multi-dimensional Itô calculus. Conceptually, there is nothing new, so I will not focus too much on it. I just want to highlight the fact that, starting from an independent Brownian Motion, we can construct a *correlated* one.

Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ be a filtered probability space. A d-dimensional Brownian motion is a stochastic process in $\mathbb{R}^d$ such that

- $W_0 = 0$ P-a.s
- W is $\mathcal{F}_t$ adapted and continuous
- for $t > s \geq 0$ the random variable $W_t - W_s$ has multi-normal distribution $\mathcal{N}_{0,(t-s)I_d}$ and is independent of $\mathcal{F}_s$.

Intuitively, a d-dimensional Brownian Motion is a vector of 1-dimensional *independent* Brownian motions.

Sometimes, it can be useful to have a d-Brownian motion in which every component is correlated (to some degree) to each other. This is the so-called *correlated* Brownian motion.

Let A an $(N \times d)$ matrix with constant values and let us denote by A^T its transpose. Then let us define

$$\varrho = AA^T \tag{1.81}$$

Given $\mu \in \mathbb{R}^N$ and a d-dimensional Brownian motion W, we put

$$B_t = \mu + AW_t \tag{1.82}$$

Then, B_t is a N-dimensional Brownian motion starting from μ at $t = 0$ with correlation matrix ϱ.

In finance, usually, N is the number of assets on the market and d is the number of "risk" factors.

All the Itô formulas can be easily generalized to the multidimensional case.

1.8 SDE: Existence and Uniqueness

We consider $Z \in \mathbb{R}^N$ and two measurable functions

$$b = b(t,x) : [0,T] \times \mathbb{R}^N \to \mathbb{R}^N, \quad \sigma = \sigma(t,x) : [0,T] \times \mathbb{R}^N \to \mathbb{R}^{N \times d} \tag{1.83}$$

We refer to b as the drift and to σ as the diffusion coefficient.

Let W a d-dimensional Brownian Motion on the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ on which the usual hypotheses hold. A solution relative to W of the SDE with coefficent (Z, b, σ) is a $\mathcal{F}_t$-adapted continuous process $(X_t)_{t \in [0,T]}$ such that

- $b(t, X_t) \in \mathbb{L}^1_{loc}$ and $\sigma(t, X_t) \in \mathbb{L}^2_{loc}$
- we have

$$X_t = Z + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s, \quad t \in [0,T] \tag{1.84}$$

  or, in differential form,

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = Z \tag{1.85}$$

There are various notions of solutions to the previous SDE. Here, we only study *strong solutions*.

In particular, we say that the SDE with coefficents (Z, b, σ) is solvable in the strong sense if, for every fixed standard Brownian Motion W, there exist a solution.

The difference w.r.t. the solution in the weak sense is that, in that case, the Brownian Motion is part of the solution.

The most important result is the following:

The SDE

$$dX_t = b(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = Z \tag{1.86}$$

verifies the standard hypotheses if

- $Z \in L^2(\Omega, P)$ and it is $\mathcal{F}_0$-measurable;
- b, σ are locally Lipschitz continuous in x uniformly w.r.t t, i.e. for every $n \in \mathbb{N}$ there exist a constant K_n such that

$$|b(t,x) - b(t,y)|^2 + |\sigma(t,x) - \sigma(t,y)|^2 \leq K_n |x-y|^2 \tag{1.87}$$

  for $|x|, |y| \leq n, t \in [0,T]$
- b, σ have at most linear growth in x, i.e

$$|b(t,x)|^2 + |\sigma(t,x)|^2 \leq K(1+|x|^2) \quad x \in \mathbb{R}^N, t \in [0,T] \tag{1.88}$$

  for a postive constant K.

We have the following result.
If the standard conditions 1) and 2) hold, then the solution of the SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = Z \tag{1.89}$$

is pathwise unique (two strong solutions are indistinguishable).

The proof is constructive so I reported it here. Before reporting the proof, I state here (without proof) the Gronwall's lemma.

Let $\varphi \in C([0,T])$ be such that

$$\varphi(t) \leq a + \int_0^t f(s)\varphi(s)ds, \quad t \in [0,T] \tag{1.90}$$

where $a \in \mathbb{R}$ and f is a continuous, non-negative function. Then we have

$$\varphi(t) \leq a e^{\int_0^t f(s)ds}, \quad t \in [0,T] \tag{1.91}$$

Let X and $\tilde{X}$ two strong solutions with initial datum Z and $\tilde{Z}$.For $n \in \mathbb{N}$ and $\omega \in \Omega$ we define

$$s_n(\omega) = \inf\{t \in [0,T] | |X_t(\omega)| \geq n\} \tag{1.92}$$

and $\tilde{s}_n(\omega)$ analogously. Remember that $\inf\{\emptyset\} = \infty$ and that the two solutions are, by definition, *continuous*.

We have that $s_n(\omega)$ and $\tilde{s}_n(\omega)$ are an increasing sequence of stopping times converging to T a.s.

We define $\tau_n = s_n \wedge \tilde{s}_n$, and we have that τ_n is an increasing sequence converging to T a.s.

Now, we can write

$$\begin{aligned} X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n} = Z - \tilde{Z} + \int_0^{t\wedge\tau_n} (b(s, X_s) - b(s, \tilde{X}_s))ds + \\ + \int_0^{t\wedge\tau_n} (\sigma(s, X_s) - \sigma(s, \tilde{X}_s))dW_s \end{aligned}$$

Using the elementary inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$ and taking the expected value, we can write

$$\begin{aligned} \mathbb{E}\left[|X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n}|^2\right] &\leq 3\mathbb{E}[|Z - \tilde{Z}|^2] \\ &+ 3\mathbb{E}\left[|\int_0^{t\wedge\tau_n} (b(s, X_s) - b(s, \tilde{X}_s))ds|^2\right] \\ &+ 3\mathbb{E}\left[|\int_0^{t\wedge\tau_n} (\sigma(s, X_s) - \sigma(s, \tilde{X}_s))dW_s|^2\right] \end{aligned}$$

Now, to deal with the second integral we use the fact that $|\int| \leq \int|\cdot|$ togheter with the Holder Inequality (the t in the formula below appers because the other function in the Holder formula is 1). Instead, for what concern the third integral, we use the Itô isometry, toghether with the fact that $(\sigma(s, X_s) - \sigma(s, \tilde{X}_s))\mathbb{1}_{\{s\leq\tau_n\wedge t\}} \in \mathbb{L}^2$ (Remember the characterization of

th stochastic integral with a stopping time as upper integration limit).

$$\begin{aligned}\mathbb{E}\left[|X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n}|^2\right] &\leq 3\mathbb{E}[|Z - \tilde{Z}|^2] \\ &+ 3t\mathbb{E}\left[\int_0^{t\wedge\tau_n} |b(s, X_s) - b(s, \tilde{X}_s|^2 ds\right] \\ &+ 3\mathbb{E}\left[\int_0^{t\wedge\tau_n} |\sigma(s, X_s) - \sigma(s, \tilde{X}_s)|^2 ds\right]\end{aligned}$$

and using the assumption on the coefficents

$$\leq 3\left(\mathbb{E}[|Z - \tilde{Z}|^2] + K_n(T+1)\int_0^t \mathbb{E}\left[|X_{s\wedge\tau_n} - \tilde{X}_{s\wedge\tau_n}|^2\right] ds\right)$$

Now, using Gronwall's lemma, we know that

$$\mathbb{E}\left[|X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n}|^2\right] \leq 3\mathbb{E}[|Z - \tilde{Z}|^2]e^{3K_n(T+1)t} \tag{1.93}$$

If $Z = \tilde{Z}$ a.s., using Fatou's lemma, we have that

$$\begin{aligned}\mathbb{E}[|X_t - \tilde{X}_t|^2] &= \mathbb{E}\left[\lim_{n\to\infty} |X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n}|^2\right] \\ &\leq \lim_{n\to\infty} \inf\mathbb{E}\left[|X_{t\wedge\tau_n} - \tilde{X}_{t\wedge\tau_n}|^2\right] = 0\end{aligned}$$

This imply that $X_t = \tilde{X}_t$ a.s $\forall t \in [0, T]$. Thus, they're modifications and, since they are continuous stochastic processes, also indistinguishable.

Let's see now an example of SDE widely used in finance, the so-called *Geometric Brownian Motion.*

Let's consider the following SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{1.94}$$

where $\mu, \sigma \in \mathbb{R}$ and $S \in \mathbb{L}^2$.

We know that we have to find a process S such that, $\forall t \in [0, T]$ we have

$$S_t = S_0 + \mu \int_0^t S_s ds + \sigma \int_0^t S_s dW_s \tag{1.95}$$

To solve this SDE, we can use directly Itô formula with $f(t, X_t) = ln(X_t)$. We have that $f'(x) = \frac{1}{x}$ and $f''(x) = -\frac{1}{x^2}$.

Thus, applying Itô formula, we have

$$\begin{aligned} d\ln(S_t) &= \frac{1}{S_t} dS_t - \frac{1}{2} \frac{1}{S_t^2} d \langle S \rangle_t \\ &= \mu dt + \sigma dW_t - \frac{1}{2} \sigma^2 dt \\ &= (\mu - \frac{1}{2} \sigma^2) dt + \sigma dW_t \end{aligned}$$

Integrating, from 0 to t, we get

$$ln(S_t) = ln(S_0) + (\mu - \frac{1}{2} \sigma^2) t + \sigma W_t \tag{1.96}$$

and, finally,

$$S_t = S_0 e^{\sigma W_t + (\mu - \frac{\sigma^2}{2}) t} \tag{1.97}$$

Chapter 2

Stochastic filtering

2.1 The problem

In this section we explain what we mean by the term "stochastic filtering" and our main steps toward its "solution". We will recall some basic facts about Hilbert spaces when we deal with projections.
Suppose we have a stochastic process X_t with values in $\mathbb{R}^n$ which is a solution of the following SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \tag{2.1}$$

where $b : [0, +\infty[\times\mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0, +\infty[\times\mathbb{R}^n \to \mathbb{R}^{n\times p}$ satisfy the usual conditions on the coefficents for a SDE. Here W_t is a $p-$dimensional Brownian Motion on a given probability space $(\Omega, \mathcal{F}, P)$.

Now, assume that we cannot observe directly this process but only a noise version of it. In particular, $\forall t \geq 0$ we observe

$$H_t = c(t, X_t) + f(t, X_t)\tilde{W}_t \tag{2.2}$$

where $c : [0, +\infty[\times\mathbb{R}^n \to \mathbb{R}^m$ and $f : [0, +\infty[\times\mathbb{R}^n \to \mathbb{R}^{m\times r}$ satisfy the usual conditions on the coefficents for a SDE. Here $\tilde{W}_t$ is a $r-$dimensional Brownian motion independent from W_t and X_0.

In order to obtain a tractable mathematical model, we consider the fol-

lowing process

$$Z_t = \int_0^t H_s ds \tag{2.3}$$

so that Z_t obey to the following SDE

$$dZ_t = c(t, X_t)dt + f(t, X_t)d\tilde{W}_t \tag{2.4}$$

Our objective is to obtain the "best" possible estimate of the unobservable process X_t using the information obtained by observing Z_t. In particular, $\forall t \geq 0$ we would like to have an estimate $\hat{X}_t$ of X_t using the information given by the observation of Z_s in the interval $[0, t]$.

First of all, when we say that $\hat{X}_t$ should be based on the observations of Z we mean that $\hat{X}_t$ should be $\mathcal{G}_t$-measurable where

$$\mathcal{G}_t = \sigma(Z_s | s \leq t) \tag{2.5}$$

is the σ-algebra generated by the process Z_s until time t.

Then, when we say "best possible" estimate, we should specify in which sense. Since we are dealing with random variables we say that $\hat{X}_t$ is a good estimate if

$$\mathbb{E}[|X_t - \hat{X}_t|^2] = \min\{\mathbb{E}[|X_t - Y|^2], \quad Y \in \Lambda_t\} \tag{2.6}$$

where

$$\Lambda_t = \{Y : \Omega \to \mathbb{R}^n, Y \in L^2(\Omega, P), Y \text{ is } \mathcal{G}_t - \text{measurable}\} \tag{2.7}$$

Now, we'll use the theory of Hilbert space to characterize the solution $\hat{X}_t$.

Before going on, we recall this general result.

Let K be a convex, closed and non-empty subset of a Hilbert space H. Then, $\forall y \in H$ there exists a unique $x_0 \in K$ such that

$$\|y - x_0\| = \min_{x \in K} \|y - x\| \tag{2.8}$$

x_0 is called the projection of y on the convex K and is denoted by $P_K(y)$.

Moreover, if K is a closed "subspace" then we know that there

$$(x - P_K(x), v) = 0, \quad \forall v \in K, \quad \forall x \in H \tag{2.9}$$

where $(\cdot, \cdot)$ is the scalar product.

We are precisely in the second situation: $L^2(\Omega, P)$ is a Hilbert space, Λ_t is a linear subspace ($0 \in \Lambda_t$) of $L^2(\Omega, P)$ and thus it's an Hilbert space as well.

Thus, we know that

$$\hat{X}_t = P_{\Lambda_t}(X) \tag{2.10}$$

Now, from the general theory on Hilbert spaces, we know that

$$\int_\Omega Y(X_t - P_{\Lambda_t}(X)) = 0 \quad \forall Y \in \Lambda_t \tag{2.11}$$

Choosing $Y = 1$ we can say that

$$\int_A (X_t - P_{\Lambda_t}(X)) = 0 \quad \forall A \in \mathcal{G}_t \tag{2.12}$$

or equivalently

$$\int_A X_t = \int_A P_{\Lambda_t}(X) \quad \forall A \in \mathcal{G}_t \tag{2.13}$$

The expression above is the exact definition of conditional expectation.

Given that the conditional expectation is unique, we can write

$$\hat{X}_t = P_{\Lambda_t}(X) = \mathbb{E}[X_t|\mathcal{G}_t] \tag{2.14}$$

2.2 The 1-d linear case

Let us assume for the moment that the two process X_t and Z_t satisfy the two following linear SDEs

$$dX_t = a(t)X_t dt + b(t)dW_t$$
$$dZ_t = c(t)X_t dt + d(t)d\tilde{W}_t$$

We assume that $Z_0 = 0$, X_0 is normally distributed and independent from W_t and $\tilde{W}_t$ and the coefficents $(a(t), b(t), c(t), d(t))$ are bounded in bounded intervals. In this way the usual conditions for the existence and uniqueness are satisfied. We remember that, in the linear case above, we have an explicit solution for X_t.

Our ultimate objective is to write down an explicit expression for the stochastic differential of $\hat{X}_t$ in terms of Z_t. We proceed in five steps.

2.2.1 Step 1

Let's define $\mathcal{L} = \mathcal{L}(Z, t)$ as the closure in $L^2(\Omega, P)$ of the space of functions that can be written as

$$c_0 + \sum_{j=1}^{n} c_j Z_{s_j}(\omega), \quad c_0, c_j \in \mathbb{R} \; s_j \leq t \tag{2.15}$$

Thus, $\mathcal{L}$ is a closed (non-empty) subspace of L^2 and we can use the result about projections outlined before. $\mathcal{P}_{\mathcal{L}}$ will denote the projection operator from L^2 to $\mathcal{L}$.

Now, we want to show that

$$\hat{X}_t = \mathcal{P}_{\Lambda_t}(X_t) = \mathcal{P}_{\mathcal{L}}(X_t) \tag{2.16}$$

The fact above has an important implication: it shows that the best possible estimate of X_t is linear in Z_t. Remember that $\mathcal{P}_{\mathcal{L}}(X_t) \in \mathcal{L}$ so that $\hat{X}_t$ can be written as a linear combination (or as a limit in $L^2(\Omega, P)$ of a linear combination) of Z_{s_j} with $s_j \leq t$.

In order to prove this, we prove the following lemma:

Let X, Z_s with $s \leq t$ be $L^2(\Omega)$ random variables and *assume* that

$$(X, Z_{s_1}, Z_{s_2}, \ldots, Z_{s_n}) \in \mathbb{R}^{n+1} \tag{2.17}$$

is normal $\forall s_j \leq t$. Then,

$$\mathbb{E}[X|\mathcal{G}] = \mathcal{P}_{\Lambda}(X) = \mathcal{P}_{\mathcal{L}}(X) \tag{2.18}$$

where $\mathcal{G}$ is the σ-algebra generated by $Z_s, s \leq t$. There is no subscript in the expression above because we're considering t as fixed.

We define $\hat{X} = \mathcal{P}_{\mathcal{L}}(X)$ and we put $\tilde{X} = X - \hat{X}$. We want to show that $\tilde{X}$ is independent from $\mathcal{G}$.

We remember that a k-dimensional vector is (jointly) normal iff *every* linear combination of its elements is normal. Given that X is normal (by assumption) and $\hat{X}$ is normal because it's (the limit in $L^2(\Omega, P)$ of) a linear combination of normal random variables, we have that

$$(\tilde{X}, Z_{s_1}, Z_{s_2}, \ldots, Z_{s_n}) \tag{2.19}$$

is normal.

Now, using Hilbert theory, we know that $\mathbb{E}[\tilde{X} Z_{s_j}] = 0, \forall j = 1, \ldots, n$. This imply non correlation (remember that $\mathbb{E}[\tilde{X}] = 0$) and, since we're dealing with normal variables, independence.

Thus, we have that

$$\tilde{X} \text{ and } (Z_{s_1}, Z_{s_2}, \ldots, Z_{s_n}) \tag{2.20}$$

are independent $\forall s_j \leq t$ and, consequently, X is independent of $\mathcal{G}$.

Using the independence result we just found, we can write

$$\mathbb{E}[\mathbb{1}_A \tilde{X}] = \mathbb{E}[\tilde{X}]\mathbb{E}[\mathbb{1}_A] = 0 \quad \forall A \in \mathcal{G} \tag{2.21}$$

and, recalling the definition of conditional expectation,

$$\mathbb{E}[X|\mathcal{G}] = \mathcal{P}_{\mathcal{L}}(X) \tag{2.22}$$

Now, let's come back to our case. We need to show the jointly normality of X_t and Z_t. In order to do this, we note that the vector

$$M_t = \begin{bmatrix} X_t \\ Z_t \end{bmatrix} \tag{2.23}$$

can be seen as the solution of the 2-dimensional *linear* SDE

$$dM_t = H(t)M(t)dt + J(t)dB_t \tag{2.24}$$

where

$$H(t) = \begin{bmatrix} a(t) & 0 \\ c(t) & 0 \end{bmatrix}, \quad J(t) = \begin{bmatrix} b(t) & 0 \\ 0 & d(t) \end{bmatrix} \tag{2.25}$$

We know that the solution of a linear SDE is always a normal process. Knowing this, we can now apply the lemma we proved at the beginning of the section to conclude.

2.2.2 Step 2

Now, we will show that the space $\mathcal{L}(Z,T)$ (defined as above), can be written as

$$\mathcal{L}(Z,T) = \{c_0 + \int_0^T f(t)dZ_t; f \in L^2[0,T], c_0 \in \mathbb{R}\} \tag{2.26}$$

We call the right-hand side $\mathcal{N}(Z,T)$.

Now, from general topology, we know that the closure of a set A (which is usually denoted by $\bar{A}$) with respect to a norm $\|\cdot\|_1$ is the set itself togheter with all the limit points of the converging sequences (in the sense on the considered norm) contained in A.

Morover, $\bar{A}$ is the smallest closed set that contains the set A. Thus, if we want to prove the equality above, we have to show that

1. $\mathcal{N}(Z,T) \subset \mathcal{L}(Z,T)$
2. $\mathcal{N}(Z,T)$ is closed w.r.t the $L^2(\Omega,P)$-norm
3. $\mathcal{N}(Z,T)$ contain all the linear combinations of the form

$$c_0 + \sum_{j=1}^{n} c_j Z_{t_j} \quad c_0, c_j \in \mathbb{R} \quad t_j \leq T \quad \forall j \tag{2.27}$$

The proof of (3) is easy. Indeed, suppose $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$; we can

write

$$\sum_{j=1}^{n} c_j Z_{t_j} = \sum_{j=0}^{n-1} c_j^* \Delta Z_j = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} c_j^* dZ_t = \int_0^T (\sum_{j=0}^{n-1} c_j^* \mathbb{1}_{[t_j, t_{j+1})}) dZ_t$$

In order to show that $\mathcal{N}(Z,T)$ is closed, we have to show that every convergent sequence contained in $\mathcal{N}(Z,T)$ has a limit which is still in $\mathcal{N}(Z,T)$. More formally, let's $u_n \subset \mathcal{N}(Z,T)$ be a convergent sequence in the $L^2(\Omega, P)$-sense and let us denote its limit point as u. We have to show that $u \in \mathcal{N}(Z,T)$.

Before proving this, we state an important result without proof. Let $f \in L^2(\Omega, P)$ and let Z_t the Itô process presented in the introduction. Then,

$$A_0 \int_0^T |f(t)|^2 dt \leq \mathbb{E}[(\int_0^T f(t) dZ_t)^2] \leq A_1 \int_0^T |f(t)|^2 dt \tag{2.28}$$

Now, $\forall n \in \mathbb{N}$ we can write u_n as

$$u_n = c_n + \int_0^T f_n(t) dZ_t \tag{2.29}$$

We know that $\{u_n\}_{\{n \in \mathbb{N}\}}$ is a convergent sequence and, consequently, a Cauchy sequence. This means that $\forall \epsilon > 0$ there exist $\nu \in \mathbb{N}$ such that $\forall n, m > \nu$ we have

$$c_n - c_m + \mathbb{E}[(\int_0^T (f_n(t) - f_m(t)) dZ_t)] < \epsilon \tag{2.30}$$

Now, using the result we stated before (2.28), we see that $\{f_n(t)\}$ is a Cauchy sequence in $L^2(\Omega, P)$ which is a complete metric space. This means that the limit point is in $L^2(\Omega, P)$.

To show that $\mathcal{N}(Z,T) \subset \mathcal{L}(Z,T)$ we need to introduce an important concept of continuity wich has a relevant impact on the way Itô Integral is constructed.

We say that a stochastic process u_t is L^2-continuous at t_0 if

$$\lim_{t \to t_0} \mathbb{E}[(X_t - X_{t_0})^2] = 0 \tag{2.31}$$

In this case, if we put

$$u^{(\zeta)} = \sum_{k=1}^{N} u_{t_{k-1}} \mathbb{1}_{]t_{k-1},t_k]} \tag{2.32}$$

where $\zeta = \{t_0, t_1, \cdots, t_N\}$ is a partition of $[0,T]$, then u^ζ is a simple process in $\mathbb{L}^2$ and we have

$$\lim_{|\zeta|\to 0} u^\zeta = u \quad \text{in } L^2([0,T]\times\Omega) \tag{2.33}$$

Now, recalling the way in which we have defined the stochastic integral, we have that we can see the Itô integral (with fixed upper integration limit) simply as the limit (in L^2) of a Riemann-Stieltjes sum.

This is particularly useful in our case. Indeed,it's clear from the definition above that a continuous function is L^2-continuous. Remember that a continuous function $f(t), t \in [0,T]$ can be seen as a stochastic process with non dependence on Ω.

Thus, we have that if $f \in C[0,T]$

$$\int_0^T f(t)dZ_t = \lim_{n\to\infty} \sum_j f(j\cdot 2^{-n})(Z_{(j+1)2^{-n}} - Z_{j2^{-n}}) \tag{2.34}$$

which is in $\mathcal{L}(Z,T)$ because it's the $L^2(\Omega,P)$-limit of a sequence of element in $\mathcal{L}(Z,T)$.

Finally, we have proved that

$$\mathcal{L}(Z,T) = \mathcal{N}(Z,T) \tag{2.35}$$

This characterization of the space $\mathcal{L}(Z,T)$ will be usufel later on.

Now, we define the *innovation process* N_t as follows:

$$N_t = Z_t - \int_0^t c(s)\hat{X}_s ds \tag{2.36}$$

with dynamic given by the following SDE

$$dN_t = c(t)(X_t - \hat{X}_t)dt + d(t)d\tilde{W}_t \tag{2.37}$$

From now on we will often use the second specification.

Now, we will show the following facts about N_t

- N_t has orthogonal increments

  This means that

$$\mathbb{E}[(N_{t_2} - N_{t_1})(N_{s_2} - N_{s_1})] = 0 \tag{2.38}$$

  for every non overlapping interval $[t_1, t_2]$ and $[s_1, s_2]$.

  Now, if $s < t$ and $Y \in \mathcal{L}(Z, T)$ we have that

$$\begin{aligned}\mathbb{E}[(N_t - N_s)Y] &= \mathbb{E}[\left(\int_s^t c(r)(X_r - \hat{X}_r)dr + \int_s^t d(r)dV_r\right)Y] \\ &= \int_s^t c(r)\mathbb{E}[(X_r - \hat{X}_r)Y]dr + \mathbb{E}[(\int_s^t d(r)d\tilde{W}_r)Y] = 0\end{aligned}$$

  where we have used Fubini-Tonelli theorem to exchange the order of integration in the second line, the fact that $X_t - \hat{X}_t$ is ortoghonal in $\mathcal{L}(Z, r)$ which contain $\mathcal{L}(Z, s), \forall s \leq r$ and the property of the increment of the Brownian Motion (remember that $d(t)$ is a deterministic function of time and, consequently, the stochastic integral can be seen as the limit in $L^2(\Omega, P)$ of a Riemann sum).

- $\mathbb{E}[N_t^2] = \int_0^t d^2(s)ds$

  Applaying Itô's formula with $g(t, x) = x^2$ we have

$$d(N_t^2) = 2N_t dN_t + \frac{1}{2}2(dN_t)^2 \tag{2.39}$$

  Integrating from 0 to t and taking the expected value we get

$$\mathbb{E}[N_t^2] = \mathbb{E}[\int_0^t 2N_s dN_s] + \int_0^t d^2(s)ds \tag{2.40}$$

  Now, using the definition of integral w.r.t an Itô process and the prop-

erty of orthogonality of the increments of N_t, we have that

$$\int_0^t N_s dN_s = \lim_{\Delta t_j \to 0} \sum N_{t_j}[N_{t_{j+1}} - N_{t_j}] \tag{2.41}$$

which is equal to 0.

- $\mathcal{L}(N,t) = \mathcal{L}(Z,t) \quad \forall t \geq 0$

  The first inclusion ($\mathcal{L}(N,t) \subset \mathcal{L}(Z,t)$) is clear from the definition of N_t itself. To establish the reverse inclusion we use the characterization of the function in $\mathcal{L}(Z,t)$) we found earlier.

  In order to do this, let $f(t) \in L^2(\Omega, P)$ and let's try to see if we can write $\int_0^t f(s)dN_s$ as the sum of an integral w.r.t the stochastic process Z_t plus a constant. We have

$$\int_0^t f(s)dN_s = \int_0^t f(s)dZ_s - \int_0^t f(r)c(r)\hat{X}_r dr$$

  Now, we know that $\forall r$ we have that $c(r)\hat{X}_r \in \mathcal{L}(Z,r)$ and can be written as

$$c(r)\hat{X}_r = m(r) + \int_0^t g(r,s)dZ_s \tag{2.42}$$

  where $g(r,\cdot) \in L^2(\Omega, P)$ and $c(r) \in \mathbb{R}$

  Thus, we can write (remembering the previous expression)

$$\begin{aligned}\int_0^t f(s)dN_s &= \int_0^t f(s)dZ_s - \int_0^t f(r)[\int_0^r g(r,s)dZ_s]dr - \int_0^t f(r)c(r)dr \\ &= \int_0^t [f(s) - \int_s^t f(r)g(r,s)dr]dZ_s - \int_0^t f(r)c(r)dr\end{aligned}$$

  It's a well-know fact about Volterra integral equations that there exists for all $h \in L^2[0,t]$ an $f \in L^2[0,t]$ such that

$$f(s) - \int_s^t f(r)g(r,s)dr = h(s) \tag{2.43}$$

By choosing $h = \mathbb{1}_{[0,t_1]}, \quad t_1 \leq t$ we obtain

$$\int_0^t f(r)c(r)dr + \int_0^t f(s)dN_s = Z_{t_1} \tag{2.44}$$

and this conludes the proof.

2.2.3 Step 3

Let N_t the innovation process defined in the previous step. We have assumed that $d(t)$ is "bounded away from zero" on $[0, T]$. This essentialy means that there is a constant $c > 0$ such that $|d(t)| > c \quad \forall t \in [0, T]$.

We define the process R_t as

$$dR_t = \frac{1}{d(t)} dN_t \tag{2.45}$$

with $t \geq 0$ and $R_0 = 0$.

We claim that R_t is a 1-dimensional Brownian Motion.

We observe that R_t inherith from N_t the continuity of paths, the orthogonality of the increments and the Gaussian property. We have to show that $\mathbb{E}[R_t] = 0$ and $\mathbb{E}[R_r R_s] = \min(s, t)$

In order to do this, we apply Itô's formula to get

$$d(R_t^2) == 2R_t dR_t + (dR_t)^2 = 2R_t dR_t + dt \tag{2.46}$$

and (using again the orthogonality of the increments) we have

$$\mathbb{E}[R_t^2] = t \tag{2.47}$$

Thus, if $s < t$

$$\mathbb{E}[R_t R_s] = \mathbb{E}[(R_t - R_s)R_s] + \mathbb{E}[R_s^2] = s \tag{2.48}$$

Since

$$\mathcal{L}(N, t) = \mathcal{L}(R, t) \tag{2.49}$$

we conclude that

$$\hat{X}_t = \mathcal{P}_{\mathcal{L}(R,t)}(X_t) \tag{2.50}$$

The projection in the space $\mathcal{L}(R,t)$ can be written as

$$\hat{X}_t = \mathbb{E}[X_t] + \int_0^t \frac{\partial}{\partial s}\mathbb{E}[X_t R_s] dR_s \tag{2.51}$$

Indeed, recall the characterization of the functions in the space $\mathcal{L}(R,t)$. We know that

$$\hat{X}_t = c_0(t) + \int_0^t g(s) dR_s \quad g \in L^2[0,t], \quad c_0(t) \in \mathbb{R} \tag{2.52}$$

Taking expectation we see that

$$c_0(t) = \mathbb{E}[\hat{X}_t] = \mathbb{E}[X_t] \tag{2.53}$$

Recall that we have that $(X_t - \hat{X}_t)$ is orthogonal to every element of $\mathcal{L}(r,t)$. So,

$$(X_t - \hat{X}_t) \perp \int_0^t f(s) dR_s \quad \forall f \in L^2[0,t] \tag{2.54}$$

Thus, we can write (remember that X_t is independent from R_t and $\mathbb{E}[\hat{X}_t] = \mathbb{E}[X_t]$)

$$\mathbb{E}[X_t \int_0^t f(s) dR_s] = \mathbb{E}[\hat{X}_t \int_0^t f(s) dR_s] = \mathbb{E}[\int_0^t g(s) dR_s \int_0^t f(s) dR_s]$$

$$= \mathbb{E}[\int_0^t g(s) f(s) ds] = \int_0^t g(s) f(s) ds \quad \forall f \in L^2[0,t]$$

If we choose $f = \mathbb{1}_{[0,r]}, r \leq t$ we have

$$\mathbb{E}[X_t R_r] = \int_0^r g(s) ds \tag{2.55}$$

or equivalently

$$g(r) = \frac{\partial}{\partial r}\mathbb{E}[X_t R_r] \tag{2.56}$$

2.2.4 Step 4

Now, we want to find an explicit expression for X_t. We know that X_t solves the linear SDE

$$dX_t = a(t)X_t dt + b(t)dW_t \tag{2.57}$$

It's easy to see (using the Itô formula) that the solution is

$$X_t = \exp(\int_0^t a(s)ds)X_0 + \int_0^t \exp^{(} \int_s^t a(s)ds)b(s)dW_s \tag{2.58}$$

Moreover, we have that

$$\mathbb{E}[X_t] = E[X_0]\exp(\int_0^t a(s)ds) \tag{2.59}$$

2.2.5 Step 5

We now combine the previous steps to obtain the solution of the filtering problem: a stochastic differential equation for $\hat{X}_t$. Starting with the formula

$$\hat{X}_t = \mathbb{E}[X_t] + \int_0^t f(s,t)dR_s \tag{2.60}$$

where

$$f(s,t) = \frac{\partial}{\partial s}\mathbb{E}[X_t R_s] \tag{2.61}$$

we use that

$$R_s = \int_0^s \frac{c(r)}{d(r)}(X_r - \hat{X}_r)dr + V_s \tag{2.62}$$

and obtain

$$\mathbb{E}[X_t R_s] = \int_0^s \frac{c(r)}{d(r)}\mathbb{E}[X_t \tilde{X}_r] \tag{2.63}$$

where

$$\tilde{X}_r = X_r - \hat{X}_r \tag{2.64}$$

Using the explicit solution of X_t we obtain

$$\mathbb{E}[X_T \tilde{X}_r] = \exp(\int_r^t a(v)dv)\mathbb{E}[X_r \tilde{X}_r] = \exp(\int_r^t a(v)dv)S(r)$$

where

$$S(r) = \mathbb{E}[(\tilde{X}_r)^2] \tag{2.65}$$

Thus

$$\mathbb{E}[X_t R_s] = \int_0^t \frac{c(r)}{d(r)} \exp(\int_r^t a(v)dv) S(r)dr \tag{2.66}$$

so that

$$f(s,t) = \frac{c(s)}{d(s)} \exp(\int_s^t a(v)dv) S(s) \tag{2.67}$$

We claim that $S(t)$ satisfies the Riccati Equation

$$\frac{dS}{dt} = 2a(t)S(t) - \frac{c^2(t)}{d^2(t)} S^2(t) + b^2(t) \tag{2.68}$$

In order to prove this, note that (Pythagorean theorem) we have

$$S(t) = \mathbb{E}[(X_t - \hat{X}_t)^2] = \mathbb{E}[X_t^2] - 2\mathbb{E}[X_t \hat{X} t] + \hat{\mathbb{X}}_{\approx}^{\nvDash} = \mathbb{E}[X_t^2] - \mathbb{E}[\hat{X}_t]$$

which can be written as

$$= T(t) - \int_0^t f(s,t)^2 ds - \mathbb{E}[X_t]^2$$

where $T(t) = \mathbb{E}[X_t^2]$.

Now, using the explicit solution of X_t we have

$$T(t) = \exp(2\int_0^t a(s)ds)\mathbb{E}[X_0^2] + \int_0^t \exp(2\int_s^t a(u)du) b^2(s)ds \tag{2.69}$$

Thus,

$$\frac{dT}{dt} = 2a(t)\exp(2\int_0^t a(s)ds)\mathbb{E}[X_0^2] + \int_0^t 2a(t)\exp(2\int_s^t a(u)du) b^2(s)ds + b^2(t) \tag{2.70}$$

that is

$$\frac{dT}{dt} = 2a(t)T(t) + b^2(t) \tag{2.71}$$

Substituting in the expression for $S(t)$ derived above and taking the deriva-

tive (using step 4) we have

$$\begin{aligned}\frac{dS}{dt} &= \frac{dT}{dt} - f^2(t,t) - \int_0^t 2f(s,t)\cdot\frac{\partial}{\partial t}f(s,t)ds - 2a(t)\mathbb{E}[X_t]^2 \\ &= 2a(t)T(t) + b^2(t) - \frac{c^2(t)S^2(t)}{d^2(t)} - 2\int_0^t f^2(s,t)a(t)ds - 2a(t)\mathbb{E}[X_t]^2 \\ &= 2a(t)S(t) + b^2(t) - \frac{c^2(t)S^2(t)}{d^2(t)}\end{aligned}$$

From the formula

$$\hat{X}_t = m_0(t) + \int_0^t f(s,t)dR_s \quad m_0(t) = \mathbb{E}[X_t] \tag{2.72}$$

it follows that

$$d\hat{X}_t = m_0'(t)dt + f(t,t)dR_t + (\int_0^t \frac{\partial}{\partial t}f(s,t)dR_s)dt \tag{2.73}$$

since

$$\begin{aligned}\int_0^u (\int_0^t \frac{\partial}{\partial t}f(s,t)dR_s)dt &= \int_0^u (\int_s^u \frac{\partial}{\partial t}f(s,t)dt)dR_s \\ &= \int_0^u (f(s,u) - f(s,s))dR_s = \hat{X}_u - m_0(u) - \int_0^u f(s,s)dR_s\end{aligned}$$

Thus,

$$\begin{aligned}d\hat{X}_t &= c_0'(t)dt + a(t)\cdot(\hat{X}_t - c_0(t))dt + \frac{c(t)S(t)}{d(t)}dR_t \\ &= a(t)\hat{X}_t dt + \frac{c(t)S(t)}{d(t)}dR_t\end{aligned}$$

Now, we substitute

$$dR_t = \frac{1}{d(t)}[dZ_t - c(t)\hat{X}_t dt] \tag{2.74}$$

we obtain

$$d\hat{X}_t = (a(t) - \frac{c^2(t)S(t)}{d^2(t)})\hat{X}_t dt + \frac{c(t)S(t)}{d^2(t)}dZ_t \tag{2.75}$$

2.3 The Kalman Bucy Filter

Finally, collecting all the steps above, we can say that the solution $\hat{X}_t = \mathbb{E}[X_t|\mathcal{G}_t]$ of the 1-dimensional filtering problem

$$dX_t = a(t)X_t dt + b(t)dW_t$$
$$dZ_t = c(t)X_t dt + d(t)d\tilde{W}_t$$

with all the conditions on the coefficentis outlined above, satisfies the following SDE

$$d\hat{X}_t = (F(t) - \frac{c^2(t)S(t)}{d^2(t)})\hat{X}_t dt + \frac{c(t)S(t)}{d^2(t)}dZ_t \quad \hat{X}_0 = \mathbb{E}[X_0] \tag{2.76}$$

where $S(t) = \mathbb{E}[(X_t - \hat{X}_t)^2]$ satisfies the Riccati equation

$$\frac{dS}{dt} = 2a(t)S(t) - \frac{c^2(t)}{d^2(t)}S^2(t) + b^2(t) \tag{2.77}$$

with $S(0) = \mathbb{E}[(X_0 - E[X_0])^2]$

2.4 Some Applications

We present here two examples to show how to implement the formulas we find above.

2.4.1 Noisy observation of a constant

The problem we study here is deliberately easy but we can learn a lot from it. This is the setting.

We have a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. We cannot see its value. Instead, we can see a noisy version of it. For every time $t \in [0, T]$ we observe

$$Y_t = X_0 * t + \kappa * W_t \tag{2.78}$$

.

Even from this simple example, we can see the main difficulties in applying the formulas we found above:

- We have to discretize the two SDEs (observable and unobservable processes).
- Usually, we don't have an explicit formula for the Conditional Expectation but a SDE. So, we have to discretize it. Usually, this involves the computation of stochastic integrals and ordinary integrals with the consequent error.

In the script below I generate one value of X_0. Then, I generate m trajectories of Y_t so that I can see an approximation of the comditonal expectation (Remember that it's a random variable itself).

What we obtain is that the goodness of our filter depends critically on the valus of κ. I called this parameter the *amplification* factor. Indeed, an higher value of κ means that we have an higher noise (positive or negative) and consequently it's difficult to filter it. Here, we report the Matlab code used in the simulation.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%% NOISY OBSERVATION OF A COSTANT %%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

close all;
clear;

% Input:
mu=2;          %Expected value of X_0
sigma=1;       %sqrt(Variance) of X_0
kappa = 0.3;   %Amplification

T=1;           %Time lenght
n=10000;       %Number of time interval
m = 1000;      %Number of replication

dt = T/n;      %Lenght time interval

% generazione variabile casuale X_0

X_0 = mu + sigma*randn(1);

time=0:dt:T;

% generazione Brownian motion

W = zeros(m,n+1);
W(:,1)=0;
```

```
for k=2:n+1
    W(:,k)=W(:,k-1)+sqrt(dt)*randn(m,1);
end

% generazione observable process

Y = X_0*repmat(time,m,1)+kappa*W;

% compute the best estimate

Y_hat = ((kappa^2)*mu +(sigma^2)*Y)./repmat(kappa^2+(sigma^2)*time,m,1);

% plot a sample trajectories

plot(time,Y(1,:)')
title('Sample trajectory observable process');
xlabel('time');

figure
plot(time,Y_hat(1,:))
title('Conditional expectation');
xlabel('time');

figure
hist(Y_hat(:,100));
title('Empirical distribution at time t=0.1');

figure
hist(Y_hat(:,10000));
title('Empirical distribution at time t=0.1');
```

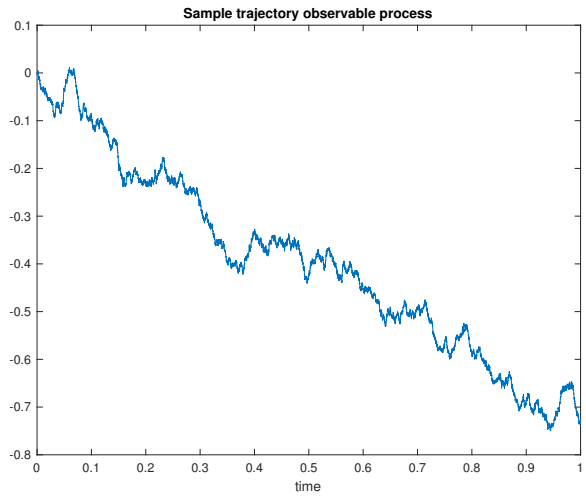

Figure 2.1: Simulated trajectory of observable process Y_t

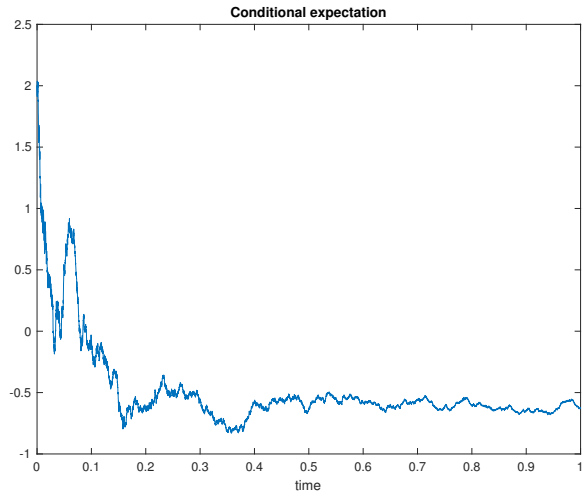

Figure 2.2: Conditional expectation process $\mathbb{E}[X_0|\mathcal{F}_t]$

2.4.2 Discretization problem: a solution

As we have seen in the previous example (and we will see again in the non-linear setting), to implement the Kalman-Bucy filter we *necessarily* have to

discretize the problem.

This raises a question. Are we sure there is no a better fiter in the discrete case? In this section we briefly present the so called Kalman filter (no Bucy) and an application with Matlab to show its potential.

From now on we essentially follows the book of Hamilton (1994). The key starting point is the so-called *state-space* representation of a dynamic model for $\mathbf{y}_t$

$$\boldsymbol{\epsilon}_{t+1} = \mathbf{F}\boldsymbol{\epsilon}_t + \mathbf{v}_{t+1} \tag{2.79}$$

$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\boldsymbol{\epsilon}_t + \mathbf{w}_t \tag{2.80}$$

In the expression above $\mathbf{y}_t$ represents a $(n \times 1)$ vector of *observed* variables at time t. $\boldsymbol{\epsilon}_t$, instead, is an *unobservable* process. We know its law but not its values. $\mathbf{A}', \mathbf{F}, \mathbf{H}'$ are matrices of *known* parameters of dimension $(n \times k), (r \times r), (n \times r)$ respectively. $\mathbf{x}_t$ is a $(k \times 1)$ vector of exogenous variables. $\mathbf{v}_t$ and $\mathbf{w}$ are vector white noise with corresponding variance-covariance matrix given by $\mathbf{Q}$ and $\mathbf{R}$. They are independent each other and intertemporraly (See the similarity with the Brownian motion).

As can be seen this a really general setting and an high number of continuous filtering problem can be approximated by this discrete version.

We will not study explicitly the derivation of the Filter (in any case it's not difficult at all and the mathematics is well beyond that developed above and below) but we highlight here three problems that can be solved using the Filter.

The first problem we want to tackle is the discrete analogous of the filtering problem in the continuous case. We want to compute

$$\boldsymbol{\epsilon}_{t+1|t} = \mathbb{E}[\boldsymbol{\epsilon}_{t+1}|\Omega_t] \tag{2.81}$$

where $\Omega_t = \{\mathbf{y}_t, \cdots, \mathbf{y}_1, \mathbf{x}_t, \cdots, \mathbf{x}_1\}$.

Once we compute the expectation above we can also forecast the next

value of $\mathbf{y}_t$ given Ω_{t-1}. More formally, we can compute

$$\mathbf{y}_{t|t-1} = \mathbb{E}[\mathbf{y}_t|\mathbf{x}_t, \Omega_{t-1}] \tag{2.82}$$

Forecasting $\mathbf{y}_t$ is really natural: we can observe it and, most part of the time, is the process we are interested in.

Sometimes, however, we may also be interested in the "latent" process $\boldsymbol{\epsilon}_t$. We remember here that we cannot know the values of the state process (even at time T when we collected all the observation of $\mathbf{y}_t$. Thus, it's natural to estimate the process $\boldsymbol{\epsilon}_t$ using all the information collected from time $t = 0$ to $t = T$. Formally, we want to compute

$$\boldsymbol{\epsilon}_{t|T} = \mathbb{E}[\boldsymbol{\epsilon}_t|\Omega_T] \tag{2.83}$$

This "procedure" is called *smoothing.*

Obtaining a solution of the smoothing problem in the continuous case (even when we are in the linear setting) is really difficult.

We show the potential of the Filter above in a simple simuleted example.

Here is the parameter of my example and its state-space representation. Anyway, I reported the code below.

The state (bivariate) process is given by

$$\begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 0.1 & 0.4 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} z_{t-1}^1 \\ z_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix} \tag{2.84}$$

while the observation process is

$$y_t = 0.5 * sin(t) + \begin{bmatrix} 0.6 & 0.9 \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} \tag{2.85}$$

Moreover, I assumed that the shock are independent and incorrelated with unit variance. We report here the code used. All the formulas and the algorithm are based on Hamilton, 1994 ([6]).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%% TESTING KALMAN FILTER %%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

n = 1000;                           % number of observations

% generation exogenous variables x

X = sin(linspace(0,10,1000))';      % simulated exogenous variable

% generation unobservable processes

Z = zeros(2,n);
Z(:,1) = 0;

F = [0.1, 0.4; 0.3, 0.7];

for t=1:n-1
    Z(:,t+1) = F*Z(:,t)+randn(2,1);
end

% generation observable process

A = 0.5;
H=[0.6 0.9];

y=zeros(1,n);

for t=1:n
    y(t) = 0.5 * X(t) + H*Z(:,t)+randn(1);
end

Y =  y';
R=1;            % Because we're considering normal gaussian
Q=eye(2);       % Two gaussian incorrelated random variable

% Kalman filter algorithm

[E,E_TT,Y_hat] = KalmanFilter_d(Y,F,X,A,H,Q,R);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%% KALMAN FILTER ALGORITHM %%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [E,E_TT,Y_hat] = KalmanFilter_d(Y,F,X,A,H,Q,R)

% Model:
% e_{t+1} = F*e_{t} + v_{t+1}     ---> State equation
% y_t = A*x_{t} + H*e_{t} + w_t   ---> Observation equation
%
% Y = vector of observations
% X = vector of exogenous variables
% A,F,H parameters matrix
% Q,R = variance-covariance matrix of the white noise (v_t and w_t)

% Dimensions control:

T=size(Y,1);            % Number of periods;
n=size(Y,1);            % Number of observations for every t;
r=size(F,1);            % Dimensions of the unobservable vector;

% Matrix initialization:

% LEGEND:
```

```
% E(t) = conditional exp. of the state vector given observation of y_t
% (t=1,...,t-1)
% E_tt = conditional exp. of the state vector given observation of y_t
% (t=1,...,t)
% E_TT = conditional exp. of the state vector given the entire observation
% of y_t (t=1,..., T)
% Y_hat = the forecast of y_t

E= zeros(r,T);
P= zeros(r^2,T);
P_tt = zeros(r^2,T);
E_TT = zeros(r,T);
E_tt = zeros(r,T);
J = zeros(r^2,T-1);
Y_hat=zeros(n,1);

% Starting values of the algorithm

E(:,1) = 0;
P(:,1) = inv(eye(r^2)-kron(F,F))*reshape(Q,r^2,1);

% Forecasting Algorithm for e_{t}

for k=1:T
    Z=reshape(P(:,k),r,r);
    E_tt(:,k)=E(:,k)+Z*H'*inv(H*Z*H'+R)*(Y(k)-A*X(k)-H*E(:,k));
    E(:,k+1)= F*E_tt(:,k);
    D=Z-Z*H'*inv(H * Z * H' + R)*H*Z;
    C=F*D*F' + Q;
    J(:,k)=reshape(D*F'*C,r^2,1);
    P_tt(:,k) = reshape(D,r^2,1);
    P(:,k+1) = reshape(C,r^2,1);
end

% Forecasting of y_{t}:

for k=1:T
    Y_hat(k) = A*X(k) + H*E(:,k);
end

% Smoothing Algorithm:

E_TT(:,end) = E_tt(:,end);

for k=1:T-1
    K=reshape(J(:,end-k+1),r,r);
    E_TT(:,end-k) = E_tt(:,end-k) + K*(E_TT(:,T-k+1)-E(:,end-k+1));
end

end
```

Here the results:

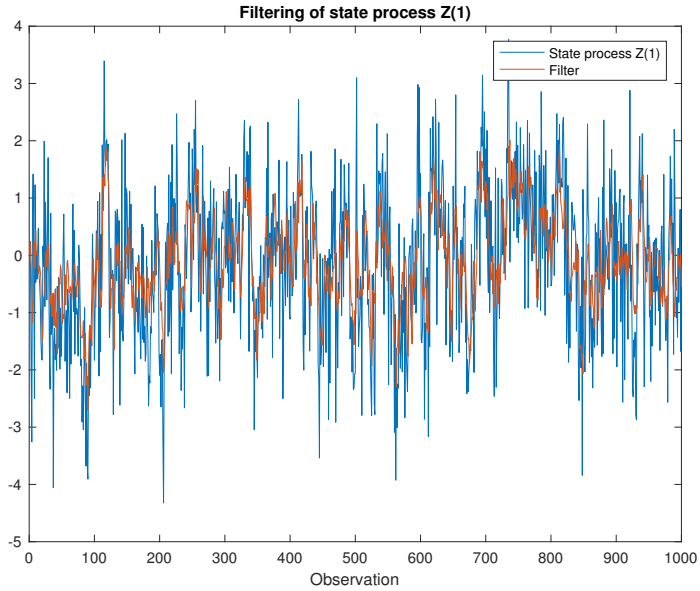

Figure 2.3: Simulated trajectory and filter of z_t^1

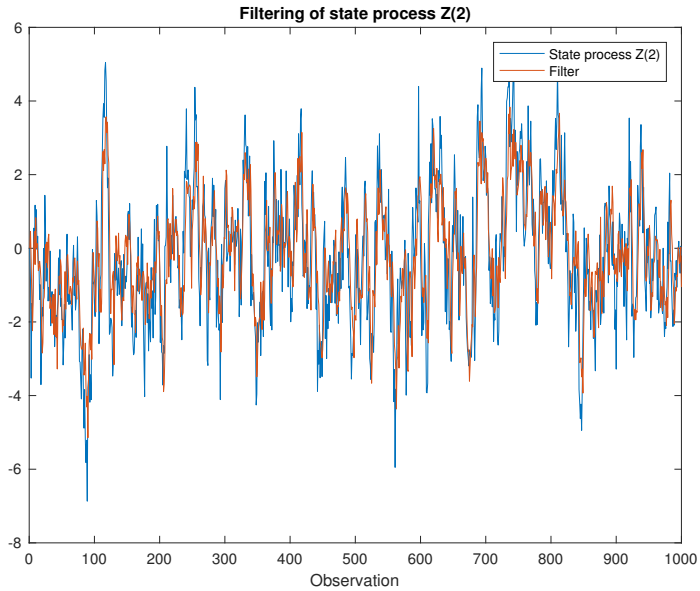

Figure 2.4: Simulated trajectory and filter of z_t^2

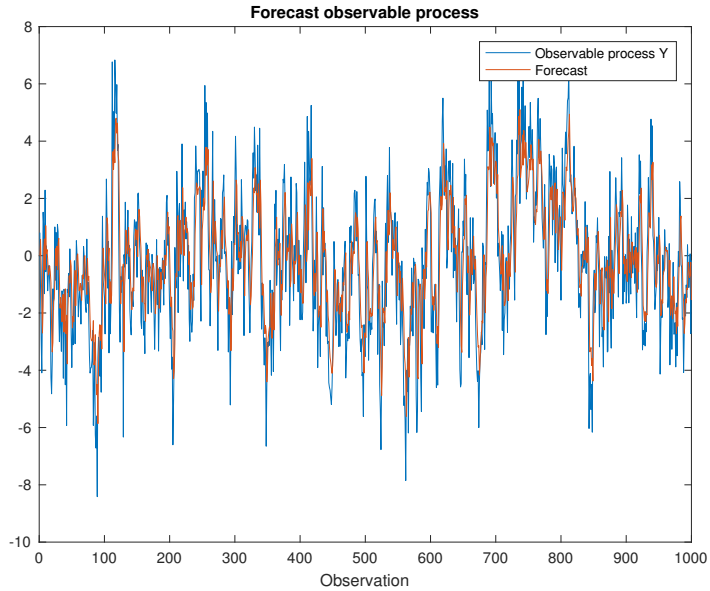

Figure 2.5: Simulated trajectory and forecast of y_t

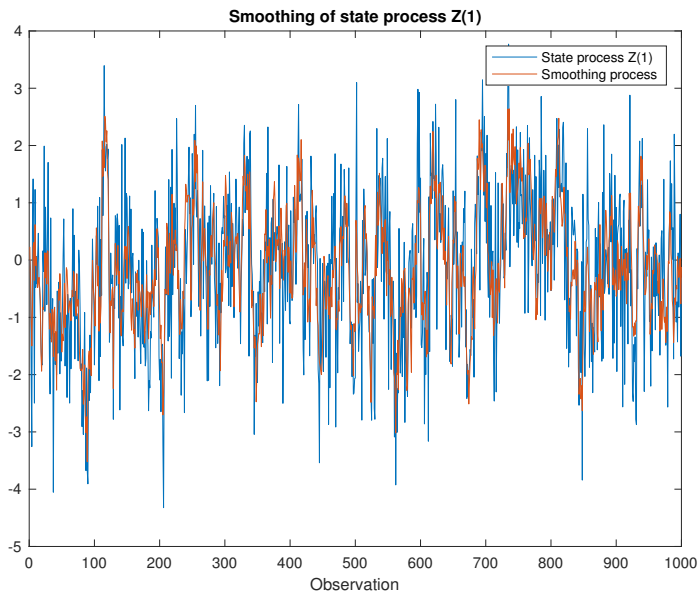

Figure 2.6: Smoothing process of z_t^1

Chapter 3

Non-linear case

In this section we will study the non-linear case of the stochastic filtering problem presented above.

Since we will use some results about change of measure, I recall some results in the first section. In this way there won't be any interruption or regression when we'll deal with the filtering problem.

3.1 Some results about change of measure

Let us consider a non-empty set X equipped with a σ-algebra $\mathcal{F}$. Let P and Q be two measures defined on $\mathcal{F}$.

We say that Q is *absolutely continuous* w.r.t. P on $\mathcal{F}$ if

$$\forall E \in \mathcal{F}, \quad P(E) = 0 \quad \Rightarrow \quad Q(E) = 0 \tag{3.1}$$

and we write $Q \ll P$.

Clearly, the definition depends on $\mathcal{F}$. For this reason it's more convenient to write $Q \ll_{\mathcal{F}} P$.

Indeed, if $\mathcal{G}$ is a sub σ-algebra of $\mathcal{F}$ and we have that $Q \ll_{\mathcal{G}} P$, this does not imply that $Q \ll_{\mathcal{F}} P$.

If $Q \ll P$ and $P \ll Q$ on $\mathcal{F}$, we say that Q and P are equivalent and we write $Q \sim P$.

We know that, for instance, $\mathcal{N}_{\mu,\sigma^2}$ is absolutely continuous w.r.t. the Lebesgue measure (m) (It follows from the property of Lebesgue integral).

Suppose now we have a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X : \Omega \to [0, \infty[$ such that $\mathbb{E}[X] = 1$. If we define another measure Q as

$$Q(A) = \int_A X dP, \quad A \in \mathcal{F} \tag{3.2}$$

we have that Q is a probability measure ($Q(\Omega) = \mathbb{E}[X] = 1$) and $Q \ll P$ on $\mathcal{F}$.

The requirement that $\mathbb{E}[X] = 1$ has been made only to obtain another *probability* measure.

We may wonder if all the measures absolutely continuous w.r.t. P have the above form (integral of a proper function). The answer is yes and it's contained on the foundamental theorem of Radon-Nikodym.

Let $(\Omega, \mathcal{F}, P)$ be a finite measure space (not necessarily a probability space). If Q is a finite measure ($Q(\Omega) < \infty$) on $(\Omega, \mathcal{F})$, then there exists $L : \Omega \to \mathbb{R}, L \geq 0$ such that

- L is $\mathcal{F}$-measurable;
- L is P-sommable;
- $Q(A) = \int_A L dP \quad \forall A \in \mathcal{F}$

We say that L is the density of Q with respect to P or, alternatively, the Radon-Nikodym derivative of Q w.r.t. P on $\mathcal{F}$. We write, indifferently,

$$L = \frac{dQ}{dP}, \quad dQ = L dP \tag{3.3}$$

or, to emphasize the dependence on $\mathcal{F}$,

$$L = \frac{dQ}{dP}|_{\mathcal{F}} \tag{3.4}$$

Of course, if we change the value of L in a negligible set, the value of the integral does not change. This means that L is P-almost surely unique: if L' verifies the same property, we necessarily have that $P(L = L') = 1$.

Moreover, let P and Q be probability measures on the space (Ω, P) with $Q \ll P$ and $L = \frac{dQ}{dP}$. It can be shown that $X \in L^1(\Omega, Q)$ if and only if $XL \in L^1(\Omega, P)$ and we have

$$\mathbb{E}^Q[X] = \mathbb{E}^P[XL] \tag{3.5}$$

We recall here the notion of conditional expectation. Let X be a real P-sommable random variable on a probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Let Y be a random variable such that

- Y is P-integrable and $\mathcal{G}$-measurable
- $\int_A X dP = \int_A Y dP, \quad \forall A \in \mathcal{G}$

We say that Y is the conditional expectation of X given $\mathcal{G}$ and we write $Y = \mathbb{E}[X|\mathcal{G}]$.

Of course, Y is defined P-almost surely. The proof of the existence of a random variable with these properties make use of the Radon-Nikodym theorem.

Suppose, for the moment, that X is a positive and P-integrable random variable on $(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ is a sub-σ-algebra. If we define a measure Q as

$$Q(A) = \int_A X dP \quad A \in \mathcal{G} \tag{3.6}$$

then $Q \ll P$ in $\mathcal{G}$. From the Radon Nikodym theorem, we know that there exist a $\mathcal{G}$-measurable and positive random variable L such that

$$\int_A X dP = Q(A) = \int_A L dP, \quad \forall A \in \mathcal{G}$$

so that $\mathbb{E}[X|\mathcal{G}] = L$ Pa.s. In the case of a general random variable we use the previous result on X^+ and X^-.

Now, we have all the ingredients to state the main results about the change of measure.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subset \mathcal{F}$ a σ-algebra, Q a probability measure such that $Q \ll_{\mathcal{F}} P$. We denote by $L^{\mathcal{F}}$ and $L^{\mathcal{G}}$ the Radon-Nikodym

derivatives of Q w.r.t. P on $\mathcal{F}$ and $\mathcal{G}$, respectively.

We have that

$$L^{\mathcal{G}} = \mathbb{E}^P[L^{\mathcal{F}}|\mathcal{G}] \tag{3.7}$$

Indeed, by the Radon-Nikodym theorem, $L^{\mathcal{G}}$ is P-integrable and $\mathcal{G}$-measurable and we have

$$\int_G L^{\mathcal{G}} dP = Q(G) = \int_G L^{\mathcal{F}} dP, \quad G \in \mathcal{G} \tag{3.8}$$

Let P, Q be probability measures on $(\Omega, \mathcal{F})$ with $Q \ll_{\mathcal{F}} P$. If $X \in L^1(\Omega, Q)$, $\mathcal{G} \subset \mathcal{F}$ a σ-algebra and $L = \frac{dQ}{dP}|_{\mathcal{F}}$. We have the so-called *Bayes' formula*

$$\mathbb{E}^Q[X|\mathcal{G}] = \frac{\mathbb{E}^P[XL|\mathcal{G}]}{\mathbb{E}^P[L|\mathcal{G}]} \tag{3.9}$$

It's worth making a couple of remark.

We say that the conditional expectation of X given $\mathcal{G}$ *under the measure* Q is given by formula (3.8). Thus, $\mathbb{E}^Q[X|\mathcal{G}]$ is a conditional expectation under the measure Q and not necessarily under the measure P. So, it's Q-unique (up to a negligible event under Q) and not necessarily P-unique. Obviously, if $P \sim Q$, $\mathbb{E}^Q[X|\mathcal{G}]$ is a conditional expectation under both measure.

I report here the proof because it's really nice and constructive. We define for simplicity $V = \mathbb{E}^Q[X|\mathcal{G}]$ and $\mathbb{E}^P[L|\mathcal{G}]$. First of all, we have to prove that the expression above makes senses (under Q) showing that the denominator is greater than zero Q a.s. Thus, we must have $Q(W > 0) = 1$. Since $L \geq 0$, it follows that we have to check that $Q(W = 0) = 0$. $\{W = 0\}$ is $\mathcal{G}$-measurable since L is $\mathcal{G}$-measurable. Thus, we can write

$$Q(W = 0) = \int_{W=0} L dP = \int_{W=0} W dP = 0$$

The first equality follows from the definition of L (Radon-Nikodym derivative of Q w.r.t. P), the second one from the definition of conditional expectation and the third equality is trivial.

Then, we have to show that the formula is right. Using the definition of

V and W we must have $VW = \mathbb{E}^P[XL|\mathcal{G}]$.

$$\begin{aligned}\int_G VW dP &= \int_G \mathbb{E}^P[VL|\mathcal{G}]dP = \int_G VLdP \\ &= \int_G \mathbb{E}^Q[X|\mathcal{G}]dQ = \int_G XdQ = \int_G XLdP\end{aligned}$$

In the first equality we used the definition of W and the $\mathcal{G}$-measurability of V (to bring V inside the conditional expectation). Then, we used the definition of conditional expectation together with the fact that $\mathbb{E}^P[VL\mathbb{1}_G] = \mathbb{E}^Q[V\mathbb{1}_G]$ and the definiton of V. This concludes the proof.

Finally, we present here the last result about the change of measure in the case we deal with stochastic process. It's known as *Girsanov Theorem*.

Let $(W_t)_{t\in[0,T]}$ be a d-dimensional Brownian motion on a given filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$. Let $\lambda \in \mathbb{L}^2_{loc}$ be a d-dimensional stochastic process.

We define the *exponential martingale* associated to λ as follows:

$$Z_t^\lambda = \exp(-\int_0^t \lambda_s \cdot dW_s - \frac{1}{2}\int_0^t |\lambda_s|^2 ds), \quad t \in [0,T] \tag{3.10}$$

Z_t^λ is a local martingale and it's a strict martingale if and only if $\mathbb{E}[Z_T^\lambda] = 1$.

The so-called *Novikov condition* is usefuel to verify if Z_t^λ is a martingale. Indeed, we have that if there exist a constant C such that

$$\int_0^T |\lambda_t|^2 dt \leq C \quad a.s. \tag{3.11}$$

then Z^λ is a martingale.

Now, assume that Z_t^λ is a P-martingale. We define a new measure Q such that

$$\frac{dQ}{dP} = Z_T^\lambda \tag{3.12}$$

Then, the process

$$W_t^\lambda = W_t + \int_0^t \lambda_s ds, \quad t \in [0,T] \tag{3.13}$$

is a Brownian motion on $(\Omega, \mathcal{F}, Q, (\mathcal{F}_t))$.

3.2 The main idea using a simple example

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X, Y : \Omega \to [0,1]$ be two random variables. Suppose that X and Y have a joint density $\Gamma(x,y)$. We recall that a density is a Borel measurable function from $\mathbb{R}^2$ to $\mathbb{R}$.

It's well know that, if $f : [0,1] \times [0,1] \to \mathbb{R}$ is a Borel measurable function such that $f\Gamma \in L^1(\mathbb{R}^2)$ we can write

$$\mathbb{E}[f(X,Y)] = \int_0^1 \int_0^1 f(x,y)\Gamma(x,y)dxdy \tag{3.14}$$

and, using Fubini theorem, we can change the order of integration.

We can also compute the conditional expectation given $Y = y$, that is

$$\mathbb{E}[f(X,Y)|Y = y] = \frac{\int_0^1 f(x,y)\Gamma(x,y)dx}{\int_0^1 \Gamma(x,y)dx} \tag{3.15}$$

where the quantity in the denominator is simply the density of Y.

Now, since we are working in a continuous setting, we may wonder the meaning of the expression above: indeed, we are conditioning on an event of measure 0. Please, note that the quantity above is not random.

Thus, we can try to extend the definition above and define a conditional expectation as follows:

$$\mathbb{E}[f(X,Y)|Y] = \frac{\int_0^1 f(x,Y)\Gamma(x,Y)dx}{\int_0^1 \Gamma(x,Y)dx} \tag{3.16}$$

The quantity above is a random variable: for every $\omega \in \Omega$ we have a value of Y and a consequent value of $\mathbb{E}[f(X,Y)|Y]$.

We recall that the conditional expectation with respect to a random variable has to be intended as the conditional expectation w.r.t. the σ-algebra generated by that random variable. So, above, we are conditioning on $\sigma(Y)$.

Is this a real conditional expectation?

To prove this we should verify that

$$\mathbb{E}[\mathbb{1}_A f(X,Y)] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[f(X,Y)|Y]] \tag{3.17}$$

for any $A \in \sigma(Y)$. We prove, more generally, that $\mathbb{E}[u(Y)f(X,Y)] = \mathbb{E}[u(Y)\mathbb{E}[f(X,Y)|Y]]$ for any bounded and Borel-measurable function u. We have that

$$\mathbb{E}[u(Y)\mathbb{E}[f(X,Y)|Y]] = \int_0^1 \int_0^1 \frac{\int_0^1 f(x,y)\Gamma(x,y)dx}{\int_0^1 \Gamma(x,y)dx} u(y)\Gamma(x,y)dxdy \tag{3.18}$$

$$= \int_0^1 \frac{\int_0^1 f(x,y)\Gamma(x,y)dx}{\int_0^1 \Gamma(x,y)dx} u(y) \left[\int_0^1 \Gamma(x,y)dx \right] dy \tag{3.19}$$

$$= \mathbb{E}[u(Y)f(X,Y)] \tag{3.20}$$

where we used Fubini theorem to exchange the order of integration.

An interesting case is when X and Y are independent. Indeed, we can write the joint density as the product of the two marginal densities and the expression above simplify to

$$\mathbb{E}[f(X,Y)|Y] = \int_0^1 f(x,Y)p(x)dx \tag{3.21}$$

where $p(x)$ is the density of X.

Now, let's try to formalize the concepts above in a measure-theory setting.

We define the following measurable space

- $\Omega = [0,1] \times [0,1]$
- $\mathcal{F} = \mathcal{B}([0,1]) \times \mathcal{B}([0,1])$
- P is a probability measure

Besides P, we introduce the product measure $Q = \mu_0 \times \mu_0$ where μ_0 is the uniform measure on the interval $[0,1]$.

We have two random variables Y and Z where

$$Y : \Omega \to [0,1], \quad Y(x,y) = y \tag{3.22}$$

and Z is a P-integrable random variable on Ω.

Suppose that $P \ll Q$. We know that (recall that the density of the uniform measure on $[0,1]$ is simply 1).

$$\mathbb{E}^P[Z] = \mathbb{E}^Q[Z\frac{dP}{dQ}] = \int_0^1 \int_0^1 Z(x,y)\frac{dP}{dQ}(x,y)dxdy \tag{3.23}$$

In expression (3.21) we can recognize that the $\frac{dP}{dQ}$ is the joint density of the example above. The latter equality is just a double ordinary integral and so it easier to compute.

Let's make a step further.

We claim that

$$\mathbb{E}^Q[Z|Y] = \int_{[0,1]} Z(x,y)\mu_0(dx) = \int_0^1 Z(x,y)dx \tag{3.24}$$

Let's check if it respects the definition of conditional expectation. First of all, we note that $\sigma(Y) = \{Y^{-1}(A), A \in \mathcal{B}([0,1])\}$ and, from the definition of Y, we have that

$$\sigma(Y) = \{[0,1] \times A, A \in \mathcal{B}([0,1])\} \tag{3.25}$$

and the indicator function a set $B \in \sigma(Y)$ is a function of only y.

Thus, we can write $\forall A \in \sigma Y$

$$\begin{aligned}\mathbb{E}^Q[\mathbb{1}_A \mathbb{E}^Q[Z|Y]] &= \int_\Omega \{\mathbb{1}_A(y) \int_{[0,1]} Z(x,y)\mu_0(dx)\}dQ \\ &= \int_0^1 \{\mathbb{1}_A(y) \int_{[0,1]} Z(x,y)\mu_0(dx)\}d\mu_0(dy) = \mathbb{E}^Q[\mathbb{1}_A Z]\end{aligned}$$

where we used Fubini theorem to see the last term as an expectation under Q.

What we obtained is that we can compute the conditional expectation under Q in a relatively straightforward manner (we have to compute an integral).

But what we are really interested in, is the conditional expectation under P. that is $\mathbb{E}^P[Z|Y]$.

In order to solve this problem, we use the Bayes formula we developed earlier. We have the following result

$$\mathbb{E}^P[Z|Y] = \frac{\mathbb{E}^Q[Z\frac{dP}{dQ}|Y]}{\mathbb{E}^Q[\frac{dP}{dQ}|Y]} = \frac{\int_0^1 Z(x,y)\frac{dP}{dQ}(x,y)dx}{\int_0^1 \frac{dP}{dQ}(x,y)dx} \tag{3.26}$$

Please, note that this is exactly (except for a slightly different notation) the epression we obtained before (without using measure theory).

This more abstract setting will be useful in the solution of the filtering problem.

Let us consider now this example. We are working on a given filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t))$ where it's defined a Brownian motion $(W_t)_{t\in[0,T]}$ and a $\mathcal{F}_0$-measurable random variable X_0.

The problem is that we cannot observe directly X_0 but, instead, a "noisy version" of it $y_t = X_0 + \kappa\epsilon_t$ where ϵ_t is a white noise.

As said in the previous chapter, we will work with the integrated form of y_t, that is

$$Y_t = X_0 t + \kappa W_t, \quad t \in [0,T] \tag{3.27}$$

We want to compute the best possible estimate of X_0 using the information obtained from the observation of the process Y_t. In mathematical terms, we want to compute

$$\pi_t(f) = \mathbb{E}^P[f(X_0)|\mathcal{F}_t^Y], \quad t \in [0,T] \tag{3.28}$$

where $\mathcal{F}_t^Y$ is the σ-algebra generated by the process Y.

We are not in the situation of the simple example studied before; here, the σ-algebra is generated by an entire process and not by a single random variable. However, we saw that what really simplified the example above, is

that X and Y were independent.

Let's proceed in this way. We note that

$$\kappa^{-1}Y_t = X_0\frac{t}{\kappa} + W_t = \int_0^t \frac{X_0}{\kappa}ds + W_t$$

is an ordinary Lebesgue integral plus a White noise under P and, consequently, we can apply the theorem of Girsanov (with $\lambda_t = \frac{X_0}{\kappa}\mathbb{1}_{[0,t]}$, using the notation of the previous section).

We suppose that the exponential martingale Z_t^λ associated to λ has $\mathbb{E}[Z_t^\lambda] = 1$ (This a requirement of the Girsanov theorem and it is necessary to have a Radon Nikodym density). As a consequence of the theorem, we know that there exists a measure $Q_T \ll P$ under which the process $\kappa^{-1}Y_t$ is a Brownian motion.

The Radon-Nikodym derivative of Q_T w.r.t P is

$$\frac{dQ}{dP} = \exp\Big(-\frac{X_0}{\kappa}W_T - \frac{1}{2}\frac{X_0^2}{\kappa^2}T\Big) = \Lambda_T^{-1}$$

After this change of measure, $\kappa^{-1}Y_t$ and X_0 are independent. Indeed, $\kappa^{-1}Y_t$ is a Brownian motion independent from $\mathcal{F}_0$ while X_0 is a $\mathcal{F}_0$-measurable random variable.

Moreover, we can show that the random variable X_0 has the same law under the two measures. In order to prove this, we should verify that

$$P(X_0 \in H) = Q_T(X_0 \in H) \quad \forall H \in \mathcal{B}(\mathbb{R}) \tag{3.29}$$

which is equivalenti to check that $\mathbb{E}^P[\mathbb{1}_H(X_0)] = \mathbb{E}^{Q_T}[\mathbb{1}_H(X_0)]$.

Using our toolkit in change of measure we have the following implications

$$\begin{aligned}\mathbb{E}^{Q_T}[\mathbb{1}_H(X_0)] &= \mathbb{E}^P[\mathbb{1}_H(X_0)\Lambda_T^{-1}] = \mathbb{E}^P\big[\mathbb{E}^P[\mathbb{1}_H(X_0)\Lambda_T^{-1}|\mathcal{F}_0]\big] \\ &= \mathbb{E}^P\big[\mathbb{1}_H(X_0)\mathbb{E}^P[\Lambda_T^{-1}|\mathcal{F}_0]\big] = \mathbb{E}^P[\mathbb{1}_H(X_0)]\end{aligned}$$

Finally, we can derive another important result. We know that

$$Q_T(A) = \int_A \Lambda_T^{-1} dP, \quad \forall A \in \mathcal{F}_T \tag{3.30}$$

We want to prove that $P \ll Q_T$. Suppose that there exist a set $G \in \mathcal{F}_T$ such that $Q(G) = 0$ but $P(G) > 0$. This means that $\Lambda_T^{-1} = 0$ almost everywhere in G. But this is not possible since Λ_T^{-1} is an exponential random variable with strictly positive values. Consequently, if $Q(G) = 0$, the only possibility is that $P(G) = 0$. This concludes the proof.

Using the Radon Nikodym theorem, we know there exists a density which we denote by Λ_T. It easy to see that since

$$1 = \mathbb{E}^P[\Lambda_T^{-1}] = \mathbb{E}^{Q_T}[\Lambda_T^{-1}\Lambda_T]$$

the density of P w.r.t. Q_T is

$$\frac{dP}{dQ_T} = \exp\left(\frac{X_0}{\kappa^2}Y_T - \frac{1}{2}\frac{X_0^2}{\kappa^2}T\right) = \Lambda_T \tag{3.31}$$

Now, we can use the Bayes formula (under the condition that $\mathbb{E}^P[|f(X_0)|] < \infty$) and we have that the filtered estimate is

$$\pi_t(f) = \mathbb{E}^P[f(X_0)|\mathcal{F}_t^Y] = \frac{\mathbb{E}^{Q_t}[f(X_0)\Lambda_t]|\mathcal{F}_t^Y}{\mathbb{E}^{Q_t}[\Lambda_t|\mathcal{F}_t^Y]}$$

Now, we claim that

$$\sigma_t(f) = \mathbb{E}^{Q_t}[f(X_0)\Lambda_t|\mathcal{F}_t^Y] = \int_{\mathbb{R}} f(x)\exp\left(\kappa^{-2}xY_t - \frac{1}{2}\kappa^{-2}x^2t\right)\mu_{X_0}(dx)$$

where μ_{X_0} is the law of the random variable X_0 and $\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}$.

Please note that the integral above is an abstract integral w.r.t. a probability measure.

Choosing a law for the random variable X_0 we end up with different results. For instance, if X_0 is a discrete random variable who takes the

values $x_1, \cdots, x_n$ with probabilties $p_1, \cdots, p_n$, we have

$$\pi_t(f) = \frac{\sum_{i=1}^n p_i f(x_i) \exp(\kappa^{-2} x_i Y_t - \frac{1}{2}\kappa^{-2} x_i^2 t)}{\sum_{i=1}^n p_i \exp(\kappa^{-2} x_i Y_t - \frac{1}{2}\kappa^{-2} x_i^2 t)} \tag{3.32}$$

3.3 Non linear filtering

In this section we will address the non-linear stochastic filtering problem. We will stick to the 1-dimensional case even if the generalization to multiple dimensions is not problematic.

We consider a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_t, P)$ where it's defined an *unobservable* process X_t which is the solution of the following SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \quad t \in [0, T] \tag{3.33}$$

where, as always, W_t is a Brownian motion and $X_0 \in \mathbb{R}$.

We can observe, continuously, a noisy version of X_t (usually refered to as "signal plus white noise") such as

$$y_t = h(t, X_t) + K(t)\epsilon_t, \quad t \in [0, T] \tag{3.34}$$

or, in its integrated form,

$$Y_t = \int_0^t h(s, X_s)ds + \int_0^t K(s)d\tilde{W}_t, \quad t \in [0, T] \tag{3.35}$$

where $\tilde{W}_t$ is a Brownian motion independent of W_t.

Of course, in the two expressions above, $b : [0, T] \times \mathbb{R} \to \mathbb{R}, \sigma : [0, T] \times \mathbb{R} \to \mathbb{R}, h : [0, T] \times \mathbb{R} \to \mathbb{R}, K : [0, T] \to \mathbb{R}$ are Borel measurable function.

In order to tackle the problem, we make these hypothesis:

- The equation for (X_t, Y_t) has a unique solution (so, the coefficents respect the usual hypotheses)
- $K(t)$ is invertible and $K(t)^{-1}$ is bounded
- b, σ, h are bounded functions

We define the filtration generated by the observable process Y_t as

$$\mathcal{F}_t^Y = \sigma\{Y_s : s \leq t\} \tag{3.36}$$

This filtration represents our information set. Please note that, at every time t, $\mathcal{F}_t^Y \subset \mathcal{F}_t$.

It's worth noting that we can relax some of the hypothesis above and still be able to solve the problem but the complexity of the problem increase exponentially.

The requirement that all the coefficients of the two SDEs are bounded is a strong requirement. For instance, the linear case we studied in the previous chapter does not obey to this condition.

Now, we will follow the main lines of the example above.

We define the following process

$$\bar{Y}_t = \int_0^t K(s)^{-1} h(s, X_s) ds + \tilde{W}_t \tag{3.37}$$

that is equal (recall here the definition of Integral w.r.t. an Ito process) to

$$\bar{Y}_t = \int_0^t K(s)^{-1} dY_s \tag{3.38}$$

Moreover, we define

$$Z_t = \begin{pmatrix} W_t \\ \bar{Y}_t \end{pmatrix} \tag{3.39}$$

and we have that

$$dZ_t = H(t)dt + dB_t \tag{3.40}$$

where $H(t) = (0, K(t)^{-1}(t, X_t))$ and $B_t = (W_t, \tilde{W}_t)$ is a 2-dimensional $(\mathcal{F}_t)$ Brownian motion.

Now, from the assumption on the boundedness on the coefficents, we know that

$$|K(t)^{-1} h(t, X_t)| < M, \quad M \in \mathbb{R} \tag{3.41}$$

and, consequently the Novikov condition below is satisfied

$$\mathbb{E}^P\left[\exp\left(\frac{1}{2}\int_0^T |K(s)^{-1}h(s,X_s)|ds\right)\right] < \mathbb{E}^P\left[\exp\left(\frac{1}{2}MT\right)\right] < \infty$$

We can thus apply the Girsanov theorem. We know that there exist a measure Q_T whose density w.r.t. P is given by

$$\frac{dQ_T}{dP} = \exp\left[-\int_0^T (K(t)^{-1}h(t,X_t)d\tilde{W}_t - \frac{1}{2}\int_0^T |K(t)^{-1}h(t,X_t)|^2 dt\right] = \Lambda_T^{-1}$$

such that the process Z_t is a 2-dimensional Brownian motion.

Consequently, we have that W_t and $\bar{Y}_t$ are two *independent* Brownian motions under Q_T and they are independent from X_0 (the argument is the same as the argument in the previous section).

Now, as in the previos section, the fact that $\frac{dQ_T}{dP}$ is strictly positive means that $P \ll Q_T$. The density of P w.r.t. Q_T is given by

$$\frac{dP}{dQ_T} = \exp\left[-\int_0^T (K(t)^{-1}h(t,X_t)d\bar{Y}_t - \frac{1}{2}\int_0^T |K(t)^{-1}h(t,X_t)|^2 dt\right] = \Lambda_T$$

Please, note that the first integral is w.r.t. $\bar{Y}_t$ which is a Brownian motion under Q_T. Λ_T is a Q_T-martingale. Λ_T^{-1} is a P-martingale.

Now, we proceed as before: we use the Bayes formula. We know that, if $\mathbb{E}^P[|f(X_t)|] < \infty$ we can write

$$\pi_t(f) = \mathbb{E}^P[f(X_t)|\mathcal{F}_t^Y] = \frac{\mathbb{E}^{Q_t}[f(X_t)\Lambda_t]|\mathcal{F}_t^Y]}{\mathbb{E}^{Q_t}[\Lambda_t|\mathcal{F}_t^Y]} = \frac{\sigma_t(f)}{\sigma_t(1)}$$

This is called the *Kallianpur-Striebel formula.*

A little remark about the notation. The measure Q_T is well defined above. When we write Q_t we mean the restriction of Q_T on $\mathcal{F}_t$. So that

$$Q_t = Q_T|_{\mathcal{F}_t} \tag{3.42}$$

Obviously, $Q_t(A) = Q_T(A), \forall A \in \mathcal{F}_t$ and $Q_t \ll P$. Please not that we can take the expectation under Q_t only of random variable which are $\mathcal{F}_t$-

measurable. We know from the previous section that the density of Q_t w.r.t. P is given by Λ_t. Moreover, we have the following chain of implication

$$\begin{aligned}\mathbb{E}^{Q_T}[f(X_t)\Lambda_t|\mathcal{F}_t^Y] = \mathbb{E}^{Q_T}[f(X_t)\mathbb{E}^{Q_T}[\Lambda_T|\mathcal{F}_t]|\mathcal{F}_t^Y] &= \mathbb{E}^{Q_T}[\mathbb{E}^{Q_T}[f(X_t)\Lambda_T|\mathcal{F}_t]|\mathcal{F}_t^Y] \\ &= \mathbb{E}^{Q_T}[f(X_t)\Lambda_T|\mathcal{F}_t^Y]\end{aligned}$$

Moreover, we have that

$$\mathbb{E}^{Q_t}[f(X_t)\Lambda_t|\mathcal{F}_t^Y] = \frac{\mathbb{E}^P[f(X_t)|\mathcal{F}_t^Y]}{\mathbb{E}^P[\Lambda_t^{-1}|\mathcal{F}_t^Y]} = \frac{\mathbb{E}^P[f(X_t)|\mathcal{F}_t^Y]}{\mathbb{E}^P[\Lambda_T^{-1}|\mathcal{F}_t^Y]} = \mathbb{E}^{Q_T}[f(X_t)\Lambda_T|\mathcal{F}_t^Y]$$

where in the first equality we used the Bayes formula, in the second one we simply use the fact that Λ_t^{-1} is a P-martingale w.r.t. $\mathcal{F}_t$. In what follows we will use frequently the expectation w.r.t. Q_t but recall the equality we proved above.

Now, we will focus on $\sigma_t(f)$ and we introduce the so-called *Zakai equation.*

Before going on, we state here two important results we will use later to prove the derivation of the Zakai equation.

Let $(W_t)_{t\in[0,T]}$ be a d-dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$ endowed with the Brownian filtration $(\mathcal{F}^W)$. For *every* real random variable $X \in L^2(\Omega, \mathcal{F}^W)$, there exists a unique $u \in \mathbb{L}^2(\mathcal{F}^W)$ such that

$$X = \mathbb{E}[X] + \int_0^T u_t dW_t \tag{3.43}$$

Let $(\Omega, \mathcal{F}, (\mathcal{F})_t, P)$ be a given probability space, $W_t, \tilde{W}_t$ be independent Brownian motion and let $u \in \mathbb{L}^2$. Now, define the sub-filtration generated by the Brownian motion $\mathcal{F}_t^W = \sigma\{W_s : s \leq t\} \subset \mathcal{F}_t$. Then, we have the following result:

$$\mathbb{E}\left[\int_0^t u_s dW_s|\mathcal{F}_t^W\right] = \int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W]dW_s, \quad \mathbb{E}\left[\int_0^t u_s d\tilde{W}_s|\mathcal{F}_t^W\right] = 0$$

and we have a similar result for the time integral

$$\mathbb{E}\left[\int_0^t u_s ds|\mathcal{F}_t^W\right] = \int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W] \tag{3.44}$$

The proof is constructive.

Let $A \in \mathcal{F}_t^W$ and denote by I_A its indicator function. Clearly, $I_A \in L^2(\Omega, \mathcal{F}_T^W)$ so that we can use the Itô representation and write

$$I_A = P(A) + \int_0^t h_s dW_s \tag{3.45}$$

where $h_s \in \mathbb{L}^2$.

Now, we can write

$$\begin{aligned}\mathbb{E}\left[I_A \int_0^t u_s dW_s\right] &= \mathbb{E}\left[P(A)\int_0^t h_s dW_s\right] + \mathbb{E}\left[\int_0^t h_s u_s ds\right] \\ &= \mathbb{E}\left[\int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W] h_s ds\right]\end{aligned}$$

where in the first equality we used the characterization of I_A and the null-expectation property of the stochastic integral while in the second we used the tower property and the Fubini tonelly theorem $(\mathbb{E}[\mathbb{E}[u_s|\mathcal{F}_s^W]]) = \mathbb{E}[u_s]$. Please note that u_t is *not* (necessarily) adapted to the filtration generated by the Brownian motion. It's adapted to the richer filtration $\mathcal{F}_t$.

Now, applying the same step to $\mathbb{E}[u_s|\mathcal{F}_s^W]$, we obtain

$$\mathbb{E}\left[I_A \int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W] dW_s\right] = \mathbb{E}\left[\int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W] h_s ds\right] \tag{3.46}$$

so that we can say that

$$\mathbb{E}\left[I_A \int_0^t u_s dW_s\right] = \mathbb{E}\left[\int_0^t \mathbb{E}[u_s|\mathcal{F}_s^W] h_s ds\right], \quad \forall A \in \mathcal{F}_t^W \tag{3.47}$$

Since this is just the definition of conditional expectation, the first statement is proved.

The second stament is trivial. Indeed, it's easy to show that $\forall A \in \mathcal{F}_t^W$

$$\mathbb{E}\left[I_A \int_0^t u_s d\tilde{W}_s\right] = \mathbb{E}\left[P(A) \int_0^t h_s d\tilde{W}_s\right] + \mathbb{E}\left[\int_0^t u_s d\tilde{W}_s \int_0^t h_s dW_s\right] = 0$$

where we used the the null-expectation property of the Stochastic integral in the second equality and the independence assumption in the third one.

To prove the expression 3.44, we note that (def. of conditional expectation)

$$\int_A \mathbb{E}[u_s|\mathcal{F}_t^W]dP = \int_A u_s dP, \quad \forall A \in \mathcal{F}_t^W \tag{3.48}$$

Thus, we can write

$$\int_0^t \int_A \mathbb{E}[u_s|\mathcal{F}_t^W]dPds = \int_0^t \int_A u_s dPds, \quad \forall A \in \mathcal{F}_t^W \tag{3.49}$$

and, using Fubini theorem and recalling the definition of conditional expectation, we have that

$$\mathbb{E}\left[\int_0^t u_s ds|\mathcal{F}_t^W\right] = \int_0^t \mathbb{E}[u_s|\mathcal{F}_t^W]ds \tag{3.50}$$

Now, we note that since u_s is $\mathcal{F}_s$-measurable is independent of $\mathcal{F}_{t,s}^W = \sigma\{W_r - W_s : s \leq r \leq t\}$ and $\mathcal{F}_t^W = \sigma\{\mathcal{F}_s^W, \mathcal{F}_{t,s}^W\}$.

Now, let us back to the derivation of the Zakai-equation.

We basically apply the Itô formula and take the conditional expectation. Recall that if we have a process $Z_t = X_t^1 X_t^2$, then its differential is given by

$$dZ_t = X_t^1 dX_t^2 + X_t^2 dX_t^1 \tag{3.51}$$

Let us apply the simple rule above to the process $f(X_t)\Lambda_t$. We have (recalling the expression for the differential of X_t and for the differential of the

exponential martingale Λ_t)

$$f(X_t)\Lambda_t = f(X_0) + \int_0^t \Lambda_s(f_x(X_s)b(s,X_s) + f_{xx}(X_s)\sigma(s,X_s))ds$$
$$+ \int_0^t \Lambda_s\sigma(s,X_s)dW_s + \int_0^t f(X_s)\Lambda_s(K(s)^{-1}h(s,X_s))d\bar{Y}_s$$

We recall here that, *under* Q_T, the process above is an Itô process because $\bar{Y}_t$ and W_t are independent Brownian motions. Moreover, thanks to our boundedness assumption, all the integrand are in $\mathbb{L}^2$ and we can take the expectation inside the integral.

Now, we can use the preliminary result we proved before. Indeed, we have that $\bar{Y}_t$ and W_t are two independent Brownian motions under Q_T. Moreover, we have that

$$\mathcal{F}_t^{\bar{Y}} = \sigma\{\bar{Y}_s : s \leq t\} = \sigma\{Y_s : s \leq t\} = \mathcal{F}_t^Y$$

Thus, after taking the conditional expectation, we employ the result we found before and we get

$$\mathbb{E}^{Q_t}[f(X_t)\Lambda_t|\mathcal{F}_t^Y] = \mathbb{E}^{Q_t}[f(X_0)|\mathcal{F}_t^Y]$$
$$+ \int_0^t \mathbb{E}^{Q_t}[\Lambda_s(f_x(X_s)b(s,X_s) + f_{xx}(X_s)\sigma(s,X_s))|\mathcal{F}_s^Y]ds$$
$$+ \int_0^t \mathbb{E}^{Q_t}[f(X_s)\Lambda_s(K(s)^{-1}h(s,X_s))|\mathcal{F}_s^Y]d\bar{Y}_s$$

We know that X_0 is $\mathcal{F}_0$-measurable and so independent from the Brownian Motions. Morevoer, recall that X_0 has the same law both under Q_T and P. This means that $\mathbb{E}^{Q_t}[f(X_0)|\mathcal{F}_t^Y] = \mathbb{E}^P[f(X_0)]$.

Now, recalling our definition of $\sigma_t(f)$, we can write

$$\sigma_t(f) = \sigma_0(f) + \int_0^t \sigma_s\{f_x(X_s)b(s,X_s) + f_{xx}(X_s)\sigma(s,X_s)\}ds$$
$$+ \int_0^t \sigma_s\{f(X_s)(K(s)^{-1}h(s,X_s))\}d\bar{Y}_s$$

Thus, we obtained an expression for the $\sigma_t(f)$. Now, we proceed to obtain an expression for $\pi_t(f)$.

We recall that if X_t, Y_t are two stochastic processes, then applying Itô formula we get

$$d\left(\frac{X_t}{Y_t}\right) = \frac{1}{Y_t}dX_t - \frac{X_t}{Y_t}dY_t - \frac{1}{Y_t^2}d(Y_t)d(X_t) + \frac{X_t}{Y_t^3}(dY_t)^2 \tag{3.52}$$

In our case, we have to compute the stochastic differential of $\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)}$. The differential of $\sigma_t(f)$ can be recovered easily from the expression above.

Instead, for $\sigma_t(1)$, we have

$$d\sigma_t(1) = \sigma_t(K(t)^{-1}h(t, X_t))d\bar{Y}_t \tag{3.53}$$

Applying the Itô formula as in 3.52 and integrating from 0 to t we get

$$\begin{aligned}\pi_t(f) = \pi_0(f) &+ \int_0^t \pi_s\{f_x(X_s)b(s, X_s) + f_{xx}(X_s)\sigma(s, X_s)\}ds \\ &+ \int_0^t (\pi_s\{f(X_s)K(s)^{-1}h(s, X_s)\} - \pi_s(f)\pi_s\{K(s)^{-1}h(s, X_s)\})dB_s\end{aligned}$$

where

$$dB_t = d\bar{Y}_t - \pi_t\{K(t)^{-1}h(t, X_t)\} \tag{3.54}$$

We highlighted B_t because it's called *innovation process* (you can see the similarity between the linear case we studied in the previous chapter).Moreover, it can be shown that, under P, B_t is an $\mathcal{F}_t$ Brownian motion.

The equation above, is called the *Kushner-Stratonovich* equation. Unfortunately, as one can easily see, this is not a stochastic differential equation for $\pi_t(f)$. Indeed we cannot, in general, express the integrand of the two integral as a function of $\pi_s(f)$. Thus, if we are not in the Kalman-Bucy filter case, we have to look for a numerical method. There are sevelar ways one can employ.

3.4 Numerical method

We present here a widely used numerical method for non-linear stochastic filtering. I advice the reader that in this section We deliberately skip some technical details; we present the idea and the implementation of the algorithm and not its convergence property. Moreover, from now on, we assume to work in a proper setting where all the conditions on the coefficents are satisfied (every function is Borel measurable, ecc...). This section and the algorithm are based on the paper of Crisan, 2016 ([3]).

3.4.1 Particle filtering

We present briefly the work setting.

As always, we have our two processes x_t, y_t

$$\begin{cases} dx_t = f(x_t)dt + \sigma(x_t)dW_t, & \text{unobservable signal process} \\ dy_t = h(x_t)dt + d\tilde{W}_t, & \text{observable process} \end{cases} \tag{3.55}$$

where $\tilde{W}, W$ are independent Brownian motion and our information set $\mathcal{F}_t^Y = \sigma\{y_s, s \in [0,t]\}$.

We know form previous section that

$$\pi_t(f) = \frac{\sigma_t(f)}{\sigma_t(1)} \tag{3.56}$$

where

$$\sigma_t(f) = \mathbb{E}^Q\left[f(x_t)\exp\left(\sum_{i=1}^{l}\int_0^t h^i(x_s)dy_s^i - \frac{1}{2}\sum_{i=1}^{l}\int_0^t (h^i(x_s))^2 ds\right)|\mathcal{F}_t^Y\right] \tag{3.57}$$

The idea of *particle filtering* is to approximate the conditional distribution $(\pi_t(f))$ using the empirical distribution of n weighted particles.

Basically, we generate n *indepedent* realizations of $x_t, t \geq 0$. For every t we have n *particles*, $x_t^j, j = 1, \cdots, n$.

Since these trajectories are generated using the dynamic of x_t under the

pysichal measure P, we have to weigh them in such a way that the weight are proportional to the probability of the occurances of the trajectory (to which the particle belongs) under Q_T.

The weights are given by

$$w_t^j = \exp\left(\sum_{i=1}^{l} \int_0^t h^i(x_s^j) dy_s^i - \frac{1}{2} \sum_{i=1}^{l} \int_0^t (h^i(x_s^j))^2 ds \right) \tag{3.58}$$

with $j = 1, \cdots, n$.

Finally, the approximation for π_t is defined as

$$\pi_t^n = \sum_{j=1}^{n} \bar{w}_t^j \delta_{x_t^j} \tag{3.59}$$

where $\delta_{x_t^j}$ is the Dirac measure concentrated at x_t^j and $\bar{w}_t^j$ are the *normalized* weights (normalized to 1).

We have that

$$\lim_{n \to \infty} \pi_t^n = \pi_t \tag{3.60}$$

Of course, in continuous time, we encounter several problems:

- The SDE for x_t may not be exactly computable (and so we have to resort to numerical schemes with the consequent discretization error);
- To comupute the weights we have to evaluate an Itô integral. This in not an easy object to compute because, as we learned in the first chapter, is not defined pathwise.
- More importantly, usually, we don't have the entire observation path but only a discrete subset of the observation (again, a discretization error).

For the first and the third problem, we can use a numerical scheme as the Euler method to approximate the SDE.

Instead, for the second problem, there is another remedy. Indeed, it's well known (see,for instance, Kushner,2009) that $\pi_t(f)$ and $\sigma_t(f)$ depend

continuously on the path $\{y_s, s \in [0,t]\}$. Moreover, it can be shown that we can approximate $\sigma_t(f)$ by

$$\tilde{\sigma}_t(f) = \mathbb{E}\left[f(x_t)\exp\left(\sum_{i=1}^{l} \int_0^t h^i(s, x_s) ds \right) \right] \tag{3.61}$$

where, for $s \in [t_i, t_{i+1}[$,

$$h^i(s, x_s) = h^i(x_s)\frac{y_{t_{i+1}} - y_{t_i}}{t_{i+1} - t_i} - \frac{1}{2}(h^i(x_s))^2 \tag{3.62}$$

Here, we try to give an idea on how to implement the model above and how to compute the relevant quantity.

The model is as follows:

$$dX_t = X_t dt + \sigma dW_t \tag{3.63}$$

$$dY_t = sin(X_t)dt + dW_t \tag{3.64}$$

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%% PARTICLE FILTER %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Model:
% dx_t = x_t dt + sigma dW_t
% dy_t = sin(x_t) dt + dW_t


T = 2;          % Time period
n=10000;        % Number of time interval
m = 1000;       % Number of Monte Carlo simulation
dt = T/n;       % Lenght time step
sigma = 0.3;    % Amplification factor

% Simulation of 'unobservable' process

X = zeros(m,n+1);       % Initilization matrix
X(:,1) = 1;

for k = 1:n
    X(:,k+1) = X(:,k) + X(:,k)*dt + sigma*sqrt(dt)*randn(m,1);
end

% Simulation of 'observable' process

Y = zeros(m,n+1);
Y(:,1) = 0;

for k=1:n
```

```
    Y(:,k+1) = Y(:,k) + sin(X(:,k))*dt + sqrt(dt)*randn(m,1);
end

% Weight computation

S = sin(X);
H = zeros(m,n+1);

for k=1:n
    H(:,k) = S(:,k).*(Y(:,k+1)-Y(:,k))*(1/dt)-0.5 * (S(:,k)).^2;
end

S = sin(X);
H = zeros(m,n);

for k=1:n
    H(:,k) = S(:,k).*(Y(:,k+1)-Y(:,k))*(1/dt)-0.5 * (S(:,k)).^2;
end

Int = [zeros(m,1), cumsum(dt*H,2)];

eInt = exp(Int);

% Normalization to 1

Tot = sum(eInt, 1);
weight = repmat((1./Tot),m,1) .* eInt;

Expected=weight.*X;

plot(sum(Expected,1))
```

Bibliography

[1] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, Springer, 2008.

[2] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011.

[3] D. Crisan, *Particle Filters in a Continuous Time Framework*, NSSPW - Nonlinear Statistical Signal Processing Workshop 2006, 2006.

[4] P. Glasserman, *Monte Carlo methods in financial engineering*, vol. 53 of Applications of Mathematics (New York), Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

[5] P. R. Halmos, *Measure Theory*, D. Van Nostrand Company, Inc., New York, N. Y., 1950.

[6] J.D. Hamilton, *Time series analysis*, Princeton University Press, Princeton, New Jersey, 1994.

[7] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, second ed., Springer, 1991.

[8] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer, 1992.

[9] B. Oksendal, *Stochastic Differential Equations, An Introduction with Applications*, Springer, Berlin, 2003.

[10] A. Pascucci, *PDE and martingale methods in option pricing*, vol. 2 of Bocconi Springer Series, Springer, Milan; Bocconi University Press, Milan, 2011.

[11] C. Pratik, S. Karaman, and E. Frazzoli, *Sampling-based algorithm for filtering using Markov chain approximations*,IEEE 51st IEEE Conference on Decision and Control (CDC), 5972-5978. Institute of Electrical and Electronics Engineers, 2012.

[12] R. van Handel, *Lecture notes on Stochastic Calculus, Filtering, and Stochastic Control*, Princeton, 2007.

[13] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.