On the Calculation of Mutual Information for Channels with Gauss-Markov Noise

by

Kasra Fallah

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this thesis, we study channels with additive Gauss-Markov noise. Such noise models are natural in many contexts and the characterization of the mutual information is unknown. We specifically study the relationship between mutual information and the Minimum Mean Square Error (MMSE) associated with the signal estimation in such channels. The early work of Duncan on channels with additive Brownian motion noise showed that the mutual information over an interval [0,T] of a channel with a general (not necessarily Gaussian) signal in the presence of additive Brownian noise is equal to the integral of the causal MMSE error over the period, where by causal we mean the filtered estimate. Our objective in this thesis is to expand upon this result in the presence of Ornstein-Uhlenbeck noise in the channel. It is shown that the same relation between mutual information and MMSE error holds. Additionally, the result is extended to a more general case in which the stochastic signal is a general function of the past. We derive the results using the machinery of stochastic calculus, specifically the Girsanov theorem and this is of independent interest.

Acknowledgments

I express my profound gratitude to Professor Ravi R. Mazumdar for his invaluable leadership, unwavering support, and profound inspiration throughout my program at the University of Waterloo. The guidance and support provided by my mentor have been beneficial, significantly influencing not just my present thesis but also my whole academic trajectory.

I would like to express my gratitude to Professor Andrew Heunis for his empathetic guidance and assistance throughout my pursuit of a master's degree, as well as to Professor Patrick Mitran and Professor Liang-Liang Xie for their valuable input as members of my thesis committee.

I express my gratitude to my fellow students Amirhossein Boreiri, Parsa Rangriz, and Sina Hoveida for their valuable cooperation, insightful recommendations, and unwavering support throughout my study.

Lastly, and most importantly, I would like to express my heartfelt gratitude to my parents for their unwavering support and encouragement. Their belief in me has been the cornerstone of my achievements.

Dedication

This is dedicated to my grandmother, Nahid Haghshenas whom I lost her final moments.

Table of Contents

List of Figures

List of Abbreviations

ACGN Additive Coloured Gaussian Noise 9

AWGN Additive White Gaussian Noise 8

CMMSE causal minimum mean square error (CMMSE) 20

KS formula Kallianpur–Striebel formula 43

MMSE minimum mean square error 20

NCMMSE non causal minimum mean square error 20

O-U Ornstein–Uhlenbeck 26

RKHS reproducing kernel Hilbert space 36

RND Radon-Nikodym derivative 10

SDE stochastic differential equation 26

Whether at Naishapur or Babylon,
Whether the Cup with sweet or bitter run,
The Wine of Life keeps oozing drop by drop,
The Leaves of Life keep falling one by one!

— Omar Khayyam, Iranian poet and mathematician, 11th century
Translated to English by Edward FitzGerald

Chapter 1

Introduction

One of the motivations for the development of Information Theory was to quantify how much information a signal contains and then to determine the maximum rate at which information could be transmitted across a noisy channel so that the signal could be recovered without error. In his classic 1948 paper, Shannon laid out the basic mathematical framework. In [1] Shannon introduced the concept of mutual information, which quantifies the amount of information shared between two random variables. The capacity of the channel is defined as the maximum of the mutual information over all distributions of the input signal. In a second publication in 1949 Shannon [2] discussed the capacity of noisy channels for additive Gaussian channels where the signal and noise are independent Gaussian random variables leading to the celebrated Shannon formula that the capacity is given by: $C \log(1+\text{SNR})$ where the SNR is the ratio between the signal and noise variance and C is a constant that depends on the power and signal bandwidth.

Later, Gelfand and Yaglom [3], along with Chiang [4], separately introduced a formula for mutual information in continuous stochastic input-output channels. This result was generalized by Duncan[5], in which he established the connection between the minimum mean square error and mutual information for additive channels with white Gaussian noise, commonly referred to as Duncan's theorem. Duncan's theorem states that the mutual information between the input and output can be represented as the integral over time of the causal minimum mean square error in a Gaussian channel. An interesting link between information theory and estimation theory is established by Duncan's theorem within the framework of a Gaussian channel. Several papers in information theory and statistical applications have demonstrated the value of Duncan's theorem. [6, 7, 8, 9, 10] In particular, the study by Guo, Shamai, and Verdu [11] used Duncan's theorem to establish the connection between causal and non-causal minimal mean square error in a Gaussian channel. Also in [12], Guo, Shamai, and Verdu showed the same relation holds for Poisson Channels. Recently, Piera and Parada [13] extended Duncan's theorem to include more general stochastic input-output systems with Gaussian noise. Their channel formulation allows for feedback in both the signal and the noise.

All the above articles assume that Gaussian distributed noise, or more specifically, the Wiener process, is the noise in the channel. While the Wiener process has many advantages as a noise process such as powerful mathematical methods to calculate the optimal filter, there are also some significant drawbacks. Balakrishnan in [14] argues that Wiener's process sample paths have unbounded variation with a probability of one is an inherent problem with this process as noise. This implies that the actual data samples have a probability of zero, which means the results obtained cannot be realizable in real-world applications such as communication channels without correction as for example in the Ito-Stratonovich approach. In addition, Nelson [15] argued that the Ornstein-Uhlenbeck (O-U) process can be seen as an approximation to the Wiener process, and it is possible to implement O-U processes in practice. Based on the Gaussian and Markovian properties of this process, the O-U process is also known as the Gauss-Markov process.

In this work we use the O-U processes as a noise and study the mutual information for channels with additive O-U noise. This approach is also compatible with the existing mathematical framework for Wiener processes. Recently, in the field of Quantum Communication, Schäfer et al.[16] and Cerf [17] separately argued that in the quantum communication channel modeling, Gauss-Markov process is more realistic to use as a noise source in the channel. Hence, it is important to study the extension of Duncan's results to channels with O-U noise.

This thesis focuses on analyzing an input-output stochastic observation channel that incorporates the O-U process as a source of noise. Furthermore, we also consider the Integrated O-U process as a source of noise. We first obtain the mathematical formulae for the mutual information of both observation channels. Next, we show how causal minimum mean square error is related to mutual information. In other words, extend Duncan's theorem for a stochastic input-output observation channel with the O-U process as a noise using the machinery of stochastic calculus, specifically the Girsanov theorem and reproducing kernel Hilbert space.

1.1 Contribution

The observation channel in we consider is similar to the one considered in [5], with the difference that we consider the O-U process as noise. Additionally, we also consider a model with the integrated O-U Process noise instead of the Wiener process employed in Duncan's theorem. Mandel et al.'s work in [18] is the most similar framework to the main problem in this thesis. They extended the Kallianpur-Striebel formula [19] for a stochastic input-output channel with an integrated Ornstein-Uhlenbeck process as noise. They derived a Bayes formula for the observation channel with an integrated Ornstein-Uhlenbeck process, which plays an important role in our problem.

Our approach involves finding the respective reproducing kernel Hilbert space (RKHS), which is a method developed by Aronszajn in [20]. Then, using the defined RKHS we can apply the extended Kallianpur-Striebel formula. We then obtain the Radon-Nikodym derivative of product measure space with respect to signal and observation measure. Then, using the definition of mutual information, we obtain a closed-form expression for mutual information over our stochastic input-output channel. In the next step, we define the minimum mean square error (MMSE) for our observation channel. This relationship enables us to understand the relationship between the mutual information and estimation error for our stochastic input-output channel. In other words, we prove a new version of Duncan's theorem for stochastic input-output channels with Ornstein-Uhlenbeck processes as noise.

1.2 Thesis outline

Chapter 2 begins with an overview to the mathematical framework from probability as well as the definition of mutual information in Section 2.1. This is followed by the establishment of the stochastic model for revisiting Duncan's problem in Section 2.2. In Section 2.3, we use the Girsanov theorem to prove Duncan's theorem[1], and in Section 2.4, we use the Girsanov theorem to show the connection between the casual and non-casual minimum mean square error shown by Guo, Shamai, and Verdu in [11][2].

Chapter 3 begins by providing detailed definitions of the Ornstein-Uhlenbeck (OU) process and the integrated OU process and the corresponding stochastic differential equations (SDE) in Section 3.1. The stochastic model of our problem is introduced in Section 3.2. In section 3.3, we compute the Radon-Nikodym derivative (RND) of the observation stochastic channel. Then, in section 3.4, we present the main result of this thesis, which is the mutual information across the observation channel. Lastly, in section 3.5, we discuss how the closed-form expression for mutual information is connected to the estimation error of the channel.

[1] In Duncan's original paper, a different approach was employed to obtain the result[5].

[2] In the original paper by Guo, Shamai, and Verdu[11], they used Duncan's results and employed an alternative approach to establishing their own result, different from our own method.

Finally, in Chapter 4, we present a series of observations. In Section 4.1, we summarize the results obtained and discuss potential future research directions in Section 4.2.

Chapter 2

Mutual information in the presence of Brownian Motion noise

This chapter discusses the calculation of mutual information of input-output stochastic systems with Gaussian noise. This is a common problem in several real-world domains, including economics and communication systems. We provide a comprehensive study of the previously described problem, which is derived from the influential paper by Duncan [5] in the 1970s. Duncan's theorem establishes a fundamental link between information theory and estimation theory within the framework of the stochastic input-output channel with Gaussian noise. Specifically, Duncan's theorem states that mutual information is equal to the integration of squared of causal minimum mean square error over the time domain. Also, we revisit the work of Guo, Shamai, and Verdu [11] which they used Duncan's theorem to find the relation between the causal and non-causal minimum mean square error

The sections of the chapter are structured in the following manner. In Section 2.1, we provide the fundamental concepts required for our investigation in this field and definitions that will be used throughout the remainder of the chapter. In Section 2.2, we study the stochastic model of the fundamental problem in its general structure and also present the necessary conditions for verifying Duncan's theorem. In Section 2.3, we review Duncan's theorem again by applying the Grisanov theorem and stochastic calculus. In Section 2.4, we investigate the correlation between the causal and non-causal minimum mean square error in the presence of Gaussian noise on the channel.

2.1 Preliminaries

2.1.1 Probability

The stochastic processes in this thesis, for example denoted as X_t, are defined on a σ-algebra $\mathcal{F}$ of the sample space Ω. The stochastic process $\{X_t\}_{t\in T}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a state space S. Each X_t is a measurable random variable that maps the sample space Ω to the state space S with respect to the sigma-algebra $\mathcal{F}$. More precisely, the process $\{X_t\}_{t\in T}$ belongs to the space L^2 if, for each $t \in T$, the random variable X_t has a finite second moment, denoted by $\mathbb{E}[X_t^2]$. The σ-algebra $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$ represents the natural filtration of the process, encompassing all available information up to time t. This arrangement guarantees that each X_t satisfies the condition of integrability, thereby establishing $\{X_t\}_{t\in T}$ as a well-defined stochastic process in $X_t \in L^2$.

Now we must establish the precise definitions of two key concepts: σ-finite measure which all probability measures are σ-finite and absolute continuity. These requirements are crucial for the Radon-Nikodym theorem, which is an essential concept enabling us to calculate mutual information.

Definition 2.1.1 (Absolute continuity)**.** A measure μ on Borel subsets of the real line is absolutely continuous with respect to the Lebesgue measure λ if for every λ-measurable set A, $\lambda(A) = 0$ implies $\mu(A) = 0$. Equivalently, $\mu(A) > 0$ implies $\lambda(A) > 0$. This condition is written as $\mu \ll \lambda$. We say μ is absolute continuous with respect to λ.

Theorem 2.1.2 (Radon–Nikodym theorem(Theorem 10.1.2 [21])**.** Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a function $f \in \mathcal{F}$ so that for all $A \in \mathcal{F}$

$$\int_A f d\mu = \nu(A)$$

f is usually denoted $d\nu/d\mu$ and called the Radon-Nikodym derivative.

2.1.2 Mutual information

Definition 2.1.3 (Finite expectation real-valued random variables space)**.** Let $\mathbb{R}^* = \mathbb{R} \cup \{\pm\infty\}$, and denote by $\Theta(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables θ taking values in $\mathbb{R}^*$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Furthermore, let $L^1(\Omega, \mathcal{F}, \mathbb{P})$ represent the subset of Θ consisting of random variables with finite expectation, formally defined as:

$$L^1(\Omega, \mathcal{F}, \mathbb{P}) = \{\theta \in \Theta : \mathbb{E}[|\theta|] < \infty\}.$$

Where $\mathbb{E}$ is the expectation with respect to measure $\mathbb{P}$.

Definition 2.1.4 (Mutual information)**.** For each X, Y which are stochastic processes defined on a measurable space which satifies:

$$\log\left[\frac{d\mu_{X,Y}}{d\left[\mu_X \times \mu_Y\right]}(X,Y)\right] \in L^1(\Omega, \mathcal{F}, \mathbb{P}),$$

where $\mu_{X,Y}$ is the (joint) measure induced by the pair of processes X, Y in the measurable space, and μ_X, μ_Y are respectively the measure defined by the stochastic processes X, Y on their measurable space. Now we define the mutual information between X, Y as $I : \mathbb{R}_+ \to \mathbb{R}$, by

$$I(X;Y) = \mathbb{E}\left[\log\left[\frac{d\mu_{X,Y}}{d\left[\mu_X \times \mu_Y\right]}(X,Y)\right]\right]$$

Observation 2.1.5. The mutual information between two discrete variables X and Y, represented as $I(X;Y)$, can be calculated using the joint probability distribution $P_{XY}(x,y)$ according to classical probability theory as below:

$$I(X;Y) = \sum_{x,y} P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} = E_{P_{XY}} \log \frac{P_{XY}}{P_X P_Y}.$$

Which was part of Shannon's paper[1, 2] formulation for finding channel capacity. We can see that the Radon-Nikodym derivative (RND) between the measures of X, Y, which are both random processes, is equal to $\frac{P_{XY}}{P_X P_Y}$. This means that we are essentially talking about the expected value of likelihood ratio (LR), which is a common way to decide between hypotheses.

Observation 2.1.6. Similarly, we can also achieve the same outcome in the context of continuous random variables. The mutual information between two continuous random variables, X and Y, is defined as follows [22]:

$$\mathrm{I}(X;Y) = \int_{\mathcal{Y}}\int_{\mathcal{X}} P_{XY}(x,y) \log\left(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}\right) dxdy$$

where $P_{X,Y}$ is now the joint probability density function of X and Y, and P_X and P_Y are the marginal probability density functions of X and Y respectively. So as we can see the result is similar to the discrete form, and stochastic processes.

In this thesis, we use the definition 2.1.4 to compute mutual information in both chapters 2 and 3. The concept applies to all channel formulations, regardless of their specific formulations or noise characteristics.

2.2 Stochastic Model

This section presents the stochastic model of channels that incorporate Gaussian noise. We begin by revisiting the general model provided by Piera et al in [13], followed by an analysis of specific cases, namely Duncan's theorem as presented in [5].

2.2.1 General stochastic Model

Consider the most general form of input-output stochastic systems with Brownian motion process as noise. Our system is represented as follows:

$$Y_t^r = \sqrt{r} \int_0^t A\left(s, X_s, Y_s^r\right) ds + \int_0^t B\left(s, Y_s^r\right) dB_s, \quad t \in [0, T] \tag{2.1}$$

Alternatively, expressed in the form of a Stochastic differential equation (SDE):

$$dY_t^r = \sqrt{r} A\left(s, X_s, Y_s^r\right) ds + B\left(s, Y_s^r\right) dB_s, \quad t \in [0, T] \tag{2.2}$$

Where X_t denotes the input stochastic process to the system, and let r be a non-negative real parameter, called signal-to-noise ratio which is the ratio between the signal and noise variance. Let Y_t^r represent the stochastic process that corresponds to the system output. Furthermore, let us regard A and B as predetermined time-varying adopted to filtering functions that serve as the most general depiction of input-output stochastic systems. Put simply, $A\left(t, X_t, Y_t^r\right)$ is exclusively influenced by the random trajectories of X and Y^r up until time t, and similarly, $B\left(t, Y_t^r\right)$. Additionally, dB_s represents the Brownian Motion measure, which is the source of noise in our problem description. The second integral in equation 2.1 represents Ito's integral, which is a widely used type of stochastic integral in information theory.

Piera and Parada have computed the mutual information of the observation channel 2.1 in their work [13]. They have demonstrated the equivalence of their formula with Duncan's theorem, which relates the mutual information to the causal minimum mean square error.

2.2.2 Special cases

Although the general stochastic model 2.1 and 2.2 offer clarity, it is essential to investigate specific scenarios in real-world applications to enhance our comprehension. Two well-known models in communication systems are the Additive White Gaussian Noise (AWGN) model and the Additive Coloured Gaussian Noise (ACGN) channel. The AWGN model is commonly used in classical communication systems, while the ACGN channel is found in both communication systems and traffic control. These models provide valuable insights, as discussed in the references [2][5].

2.2.2.1 Additive white Gaussian noise channel

In this section, we aim to demonstrate the application of a stochastic model to the well-known additive Gaussian noise channel (AWGN). By substituting $A(t, f, g) = f(t)$ and $B(t, g) = 1$ into the generic stochastic model described in equation 2.1, we obtain the following result:

$$dY_t^r = \sqrt{r} X_t dt + dB_t \tag{2.3}$$

$$Y_t^r = \int_0^t \sqrt{r} X_t dt + B_t \tag{2.4}$$

Our model now conforms to the typical communication noise model, which is fundamental to Shannon's renowned formula [1, 2]. The formulation of Duncan's theorem [5] also applies this model, which will be explored in more detail in section 2.3.

2.2.2.2 Additive colored Gaussian noise channel

Now, let us have a look at a more advanced model. We may construct the additive colored Gaussian noise channel (ACGN), which means that if we let $A(t, f, g) = f(t)$ and $B(t, g) = G(t)$ we will obtain the following:

$$dY_t^r = \sqrt{r} X_t dt + G(t) dW_t \tag{2.5}$$

$$Y_t^r = \int_0^t \sqrt{r} X_t dt + \int_0^t G(s)\, dW_s \tag{2.6}$$

As we can see, it is not difficult to determine the covariance of the noise, which is a well-known result that can be achieved by utilizing Ito's integral martingale properties that are cited in [23].

$$N_t = \int_0^t G(s) dW_s, \quad t \in [0, T]$$

$$\begin{aligned}
\text{cov}(N_{t_1,t_2}) &= \mathbb{E}\left[\int_0^{t_1} G(s)dW_s \int_0^{t_2} G(s)dW_s\right] \\
&= \mathbb{E}\left[\int_0^{t_1}\int_0^{t_2} G(r)G(s)dW_r dW_s\right] \\
&= \int_0^{\min\{t_1,t_2\}} G^2(s)ds
\end{aligned}$$

The covariance of the colored noise is determined by the square integrability of the function $G(s)$. This condition ensures that the covariance is meaningful. In order to obtain a reliable outcome, it is imperative to establish certain essential prerequisites for the situation at hand. It is important to acknowledge that as the complexity of a system increases, the conditions of the system also need to get more intricate.

2.3 Revisiting Duncan's Theorem

This section focuses on the problem of additive white Gaussian noise, as described in 2.2.2.1, and computes the mutual information between the signal and observation in this channel formulation, which serves as a revisiting of Duncan's theorem. In contrast to Duncan's original study [5], this section utilizes the Girsanov theorem to calculate the Radon-Nikodym derivative (RND) of the relevant measures, instead of using discrete samples. The Girsanov theorem is first proved in section 2.3.1 by the use of stochastic calculus tools such as Itô's lemma, martingale representations, and changing of measure. Subsequently, the theorem is applied to compute the RND of the AWGN channel. Finally, the mutual information is computed based on the definition provided in 2.1.4.

2.3.1 Calculation of Radon–Nikodym derivative

Theorem 2.3.1 (Girsanov theorem)**.** Let P and Q the law of X_t and Y_t which are the following SDE solutions

$$dX_t = \sigma(t, X_t)dB_t + b(t, X_t)dt \tag{2.7}$$

$$dY_t = \sigma(t, Y_t)dB_t + c(t, X_t)dt \tag{2.8}$$

Where $Y_0 = X_0 = x$ for our processes. Let us suppose that σ, b and c satisfy the Lipschitz and growth conditions. In addition let $a(t, x) > 0$ for all t and x, and

$$c(t, x) - b(t, x) = \sigma(t, x)e(t, x)$$

for some $e(t,x)$ satisfying $|e(t,x)| \leq C$. Then for P and Q are mutually absolutely continuous on the same σ-field $\mathcal{F}_t$ and the Radon-Nikodym derivative is given by

$$\frac{dQ}{dP} = \exp\left[\int_0^t e(s,X_s)d\widetilde{B_s} - \frac{1}{2}\int_0^t e^2(s,X_s)ds\right].$$

Where

$$d\widetilde{B_t} = \frac{1}{\sigma(t,X_t)}[dX_t - b(t,X_t)dt]$$

and the Radon-Nikodym derivative is measurable with respect to the σ-field $\mathcal{F}_t = \sigma\{x(u) : 0 \leq u \leq t\}$.

One lemma must be proven in order to prove the Girsanov theorem:

Lemma 2.3.2. Let $c(s,\omega)$ be adapted to filtering and bounded. Then

$$Z(t) = \exp\left[\int_0^t c(s,\omega)dX_s - \frac{1}{2}\int_0^t c^2(s,\omega)ds\right]$$

is a martingale with respect to $(\Omega, \mathcal{F}_s, P_0)$.

Proof. Restricting ourselves to an arbitrary yet finite interval $[0,T]$ does not compromise the generality of our analysis. The theorem holds unequivocally for constants c. Indeed, we can consider $c(t)$ as piece-wise constant, taking values c_i over finite intervals $[t_i, t_{i+1}]$, where $0 < t_1 < t_2 < \cdots < t_n = T$. Furthermore, each c_i may vary with ω, as long as it remains $\mathcal{F}_i$ measurable.

Consequently, if $c(s,\omega)$ is simple and bounded, the process $Z(t)$ defined a martingale. It's worth highlighting that:

$$Z_c(t)^2 \leq Z_{2c}e^{C^2t}$$

where C acts as an upper bound on $c(s,\omega)$. Given a bounded, adapted to filtering c, we can approximate it with simple functions c_n, each uniformly bounded. Hence, as c_n converges to c, the approximating martingales $Z_{c_n(\cdot)}(t)$ remain uniformly bounded in $L_2[P_0]$, ensuring that the limit $Z(t) = Z_{c(\cdot)}(t)$ remains a martingale. □

Girsanov theorem. Step 1. Let $e(s,\omega) = e(s,x(s,\omega))$. We know from the respective measure of Brownian motion that:

$$R(t,\omega) = \exp\left[\int_s^t e(s,x(s))d\beta(s) - \frac{1}{2}\int_s^t e^2(s,x(s))ds\right]$$

is a martingale with respect to $(\Omega, \mathcal{F}_t^s, P_{s,x})$. It defines a measure $Q = Q_{s,x}$ on $(\Omega, \mathcal{F}_\infty^s)$ by the relation $dQ = R(t,\omega)dP_{s,x}$ on $\mathcal{F}_t^s$. The function $\gamma(\cdot)$ defined by

$$\gamma(t) = \beta(t) - \int_0^t e(s,\omega)ds$$

is a Brownian motion relative to $(\Omega, \mathcal{F}_t, Q)$. More precisely under $Q, \gamma(t) - \gamma(s)$ is distributed independently of $\mathcal{F}_s$, and has the same distribution as Brownian motion. The proof is based on verifying that

$$\xi_\lambda(t) = \exp\left[\lambda\gamma(t) - \frac{\lambda^2}{2}t\right]$$

is a martingale relative to $(\Omega, \mathcal{F}_t, Q)$. By the definition of Q, this is the same as checking that $\xi_\lambda(t)R(t,\omega)$ is a martingale with respect to $(\Omega, \mathcal{F}_t, P)$. But

$$\begin{aligned}\xi_\lambda(t)R(t,\omega) &= \exp\left[\lambda\gamma(t) - \frac{\lambda^2}{2}t + \int_0^t e(s,\omega)d\beta(s) - \frac{1}{2}\int_0^t e^2(s,\omega)ds\right]\\ &= \exp\left[\int_0^t [\lambda + e(s,\omega)]d\beta(s) - \frac{1}{2}\int_0^t [\lambda + e(s,\omega)]^2 ds\right]\end{aligned}$$

is clearly a $(\Omega, \mathcal{F}_t, P)$ martingale. Now Let us rewrite the solution of

$$dx(t) = \sigma(t,x(t))d\beta(t) + b(t,x(t))dt$$

as

$$dx(t) = \sigma(t,x(t))d\gamma(t) + c(t,x(t))dt$$

with

$$\gamma(t) = \beta(t) - \int_0^t \sigma(t,x(t))e(t,x(t))dt$$

and compare it with the solution

$$dy(t) = \sigma(t,y(t))d\beta(t) + c(t,y(t))dt$$

Since $\beta(\cdot)$ under P and $\gamma(\cdot)$ under Q are both Brownian motions, the distribution of $y(\cdot)$ under P is the same as the distribution of $x(\cdot)$ under Q. So the Radon-Nikodym derivative of $Q_{0,x}$ with respect to $P_{0,x}$ can be computed by computing the Radon-Nikodym derivative (RND) of the distribution of $x(\cdot)$ under Q with respect the distribution of the same solution $x(\cdot)$ under P. If we can show that $R(t, \omega)$ is measurable with respect to $\widehat{\mathcal{F}}_t = \sigma\{x(s) : 0 \leq s \leq t\}$ we are done. To establish this all we need to note is that because $a(t, x) > 0, \sigma(t, x)$ is nonvanishing and we can invert the stochastic integral to represent

$$d\beta(t) = \frac{1}{\sigma(t, x(t))}[dx(t) - b(t, x(t))dt].$$

Therefore $R(t, \omega)$ can be expressed as a filtering adapted function of $x(\cdot)$. □

Our primary objective is to ascertain the RND of the measure on the product space of the signal (X) and observation (Y) with respect to the product of the individual measures on the signal and the observation, expressed as $\frac{d\mu_{YX}}{d(\mu_X \times \mu_Y)}$. Our methodology involves first computing $\frac{d\mu_{YX}}{d(\mu_B \times \mu_X)}$ by utilizing the Girsanov theorem. Following this, we determine $\frac{d\mu_Y}{d\mu_B}$. Finally, through the application of well-established lemmas in measure theory, we deduce $\frac{d\mu_{YX}}{d(\mu_X \times \mu_Y)}$. Note that μ_B represents the respective measure of Brownian motion, which serves as the noise in this particular stochastic input-output channel.

Proposition 2.3.3. For our stochastic input-output channel described in 2.2.2.1, the RND of the measure on the product space of the signal (X) and observation (Y) with respect to the product of the individual measures on the signal and the noise is given by :

$$\frac{d\mu_{YX}}{d\mu_B d\mu_X} = \exp\left[\sqrt{r}\int X_s dB_s - \frac{r}{2}\int X_s^2 ds\right]$$

Proof. By applying Girsanov's theorem, we can analyze the system described by equations (2.7) and (2.8) under a different measure. Specifically, we can transform the drift term b to zero and set the volatility term σ to 1. This transformation allows us to utilize Girsanov's theorem directly. Consequently, we obtain the following relationship:

$$e(t, x)\sigma(t, x) = c(t, x) - b(t, x) = c(t, x)$$

Simplifying further, we have:

$$e(t, x) = c(t, x) = \sqrt{r}X_s$$

Thus, by Girsanov's theorem, we derive the Radon-Nikodym derivative of the measure μ_{YX} with respect to the product measure $\mu_B \mu_X$:

$$\frac{d\mu_{YX}}{d\mu_B d\mu_X} = \exp\left[\sqrt{r}\int X_s\, dB_s - \frac{r}{2}\int X_s^2\, ds\right]$$

This result leverages Girsanov's theorem to facilitate the measure transformation necessary for the given stochastic process. □

Proposition 2.3.4. For our stochastic input-output channel described in 2.2.2.1, the RND of the measure on the product space of the observation (Y) with respect to the measure of the noise (B) is given by :

$$\frac{d\mu_Y}{d\mu_B} = \exp\left[\sqrt{r}\int \hat{X}_s dB_s - \frac{r}{2}\int \hat{X}_s^2 ds\right]$$

where $\hat{X}_s = \mathbb{E}\left[X_s \mid \mathcal{F}_s\right]$ represents the estimation of the signal process X_s based on the information available up to time s from the causal observation process $\mathcal{F}_s$.

Proof. As demonstrated in the case of $\mu_{YX} \ll \mu_{BX}$, we have the Radon-Nikodym derivative ϕ given by:

$$\phi = \exp\left[\sqrt{r}\int X_s dB_s - \frac{r}{2}\int X_s^2 ds\right]$$

Now, our objective is to establish that $\mu_Y \ll \mu_B$. Using the independence of the processes B and X, the measure μ_{BX} takes the form of the product measure $\mu_B \mu_X$. Thus, our focus shifts to integration over the measure μ_Z. Let's define

$$\psi = E_{\mu_X}[\phi],$$

where E_{μ_Z} denotes integration with respect to the measure μ_Z. Consequently, we have

$$\frac{d\mu_Y}{d\mu_B} = \psi.$$

So we will have:

$$\frac{d\mu_Y}{d\mu_B} = E_{\mu_X}[\phi] = E_{\mu_X}[\exp\left[\sqrt{r}\int X_s dB_s - \frac{r}{2}\int X_s^2 ds\right]]$$

By continuity of measures and Fubini's theorem we know that expected value will go inside the integral:

$$\frac{d\mu_Y}{d\mu_B} = \exp\left[\sqrt{r}\int E_{\mu_X}[X_s]dB_s - \frac{r}{2}\int E_{\mu_X}[X_s^2]ds\right]$$

So we will have:

$$\frac{d\mu_Y}{d\mu_B} = \exp\left[\sqrt{r}\int \hat{X}_s dB_s - \frac{r}{2}\int \hat{X}_s^2 ds\right]$$

where $\hat{X}_s = \mathbb{E}[X_s \mid \mathcal{F}_s]$ represents the estimation of the signal process X_s based on the information available up to time s from the causal observation process $\mathcal{F}_s$. □

Proposition 2.3.5. For our stochastic input-output channel described in 2.2.2.1, the RND of the measure on the product space of the signal (X) and observation (Y) with respect to the product of the individual measures on the signal and the observation is give by:

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \exp\left[\sqrt{r}\int (X_s - \hat{X}_s)dY_s - \frac{r}{2}\int (X_s - \hat{X}_s)^2 ds\right]$$

Before providing the proof of the proposition above, we must state the three lemmas used in the proof. This thesis does not provide proof of these lemmas, but it is available in the references mentioned.

Lemma 2.3.6 (Theorem 2.3.5 [23])**.** Let (X, M, μ) be a measure space, where X is the underlying set, M is the sigma-algebra of measurable sets, and μ is a measure defined on M. Let f be a nonnegative measurable function on (X, M). Define a measure ν on (X, M) by $\nu(E) = \int_E f\, d\mu$ for all $E \in M$. Then for any nonnegative, measurable function F on (X, M), we have

$$\int_X F\, d\nu = \int_X F f\, d\mu.$$

Lemma 2.3.7 (Theorem A, Chapter 6, [24])**.** Suppose μ, ν, and λ are finite measures on the measurable space $(X, \mathcal{M})$. If $\mu \ll \nu$ and $\nu \ll \lambda$, then $\mu \ll \lambda$, and we have the following Radon-Nikodym derivative relationship:

$$\frac{d\mu}{d\lambda} = \frac{d\mu}{d\nu} \cdot \frac{d\nu}{d\lambda}$$

for λ-almost every point.

Lemma 2.3.8. Let μ and ν be two finite measures on a measurable space. If $\mu \ll \nu$ and $\nu \ll \mu$, then the following statement holds almost everywhere with respect to μ (and therefore also with respect to ν):

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$$

Proof. By applying the Radon-Nikodym theorem to the conditions $\mu \ll \nu$ and $\nu \ll \mu$, and referencing Lemma 2.3.7, we obtain the following expression:

$$\nu = \frac{d\nu}{d\mu}\mu = \frac{d\nu}{d\mu}\frac{d\mu}{d\nu}\nu$$

Additionally, it is $\nu \ll \nu$ and $\nu = f\nu$ with $f \equiv 1$. From the $\mu-$ a.s.-uniqueness of the density of ν relating to ν it follows

$$\frac{d\nu}{d\mu}\frac{d\mu}{d\nu} = 1\nu - \text{ a.s.}$$

Finally, we will have:

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1} \nu - \text{ a.s..}$$

□

So with all of these lemmas let us go back to the proof of proposition 2.3.5

Proof. By lemma 2.3.7 we can write chain rule so we have:

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \frac{d\mu_{YX}}{d\mu_X d\mu_B} * \frac{d\mu_B}{d\mu_Y}$$

Then by lemma 2.3.8 we can write:

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \frac{d\mu_{YX}}{d\mu_X d\mu_B} * (\frac{d\mu_Y}{d\mu_B})^{-1}$$

So we can write from proposition 2.3.3, and proposition 2.3.4:

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \exp\left[\sqrt{r}\int X_s dB_s - \frac{r}{2}\int X_s^2 ds\right] * (\frac{d\mu_Y}{d\mu_B})^{-1}$$

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \exp\left[\sqrt{r}\int X_s dB_s - \frac{r}{2}\int X_s^2 ds\right] * (\exp\left[\sqrt{r}\int \hat{X}_s dB_s - \frac{r}{2}\int \hat{X}_s^2 ds\right])^{-1}$$

Then we can write:

$$\begin{aligned}\frac{d\mu_{YX}}{d\mu_Y d\mu_X} &= \exp\left[\sqrt{r}\int X_s\, dB_s - \frac{r}{2}\int X_s^2\, ds \quad -\sqrt{r}\int \hat{X}_s\, dB_s + \frac{r}{2}\int \hat{X}_s^2\, ds\right] \\ &= \exp\left[\sqrt{r}\int (X_s - \hat{X}_s) dB_s - \frac{r}{2}\int (X_s^2 - \hat{X}_s^2) ds\right]\end{aligned}$$

Then by $dY_t^r = \sqrt{r}\hat{X}_t dt + dB_t$ then we have $dB_t = dY_t^r - \sqrt{r}\hat{X}_t dt$

$$\begin{aligned}\frac{d\mu_{YX}}{d\mu_Y d\mu_X} &= \exp\left[\sqrt{r}\int (X_s - \hat{X}_s)(dY_t^r - \sqrt{r}X_t dt) - \frac{r}{2}\int (X_s^2 - \hat{X}_s^2) ds\right] \\ &= \exp\left[\sqrt{r}\int (X_s - \hat{X}_s)(dY_s^r) - r\int (X_s - \hat{X}_s)\hat{X}_s dt - \frac{r}{2}\int (X_s^2 - \hat{X}_s^2) ds\right]\end{aligned}$$

$$\frac{d\mu_{YX}}{d\mu_Y d\mu_X} = \exp\left[\sqrt{r}\int (X_s - \hat{X}_s)(dY_s^r) - \frac{r}{2}\int (X_s - \hat{X}_s)^2 ds\right]$$

□

2.3.2 Calculation of mutual information

Theorem 2.3.9. The mutual information between the input signal stochastic process and observation stochastic process of the stochastic input-output system described in 2.2.2.1 is given by:

$$I^r(X; Y^r) = \frac{r}{2} E\left[\int (X_s - \hat{X}_s)^2 ds\right]$$

where $\hat{X}_s = \mathbb{E}[X_s \mid \mathcal{F}_s]$ represents the estimation of the signal process X_s based on the information available up to time s from the causal observation process $\mathcal{F}_s$.

Before proving this proposition, we state a lemma whose proof is not provided here but can be found in the cited reference.

Lemma 2.3.10 (Theorem 5.4, [23])**.** Let $(X_t)_{t\geq 0}$ be a martingale with respect to a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. Let μ be a finite signed measure on $\mathcal{F}$ and $g : [0, \infty) \times \Omega \to \mathbb{R}$ be a measurable function such that $\int_0^t |g(s, \omega)| \, d|X_s(\omega)| < \infty$ for almost every $\omega \in \Omega$ and every $t \geq 0$. Additionally, for each fixed $s \geq 0$, the function $\omega \mapsto g(s, \omega)$ is $\mathcal{F}_s$-measurable.Suppose $Y_t(\omega) = \int_0^t g(s, \omega) \, dX_s(\omega)$ for $t \geq 0$. Then $(Y_t)_{t\geq 0}$ is a martingale with respect to $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$.

Now we can write the proof of theorem 2.3.9

Proof. By definition 2.1.4 now we have:

$$I^r(X; Y^r) \doteq \mathbb{E}\left[\log\left[\frac{d\mu_{X,Y^r}}{d\left[\mu_X \times \mu_{Y^r}\right]}(X, Y^r)\right]\right]$$

So by proposition 2.3.5 we will have:

$$I^r(X; Y^r) \doteq \mathbb{E}\left[\log\left[\exp\left[\sqrt{r}\int (X_s - \hat{X}_s)(dY_s^r) + \frac{r}{2}\int (X_s - \hat{X}_s)^2 ds\right]\right]\right]$$

By simple calculation, we can write the above equation as:

$$I^r(X; Y^r) \doteq \mathbb{E}\left[\sqrt{r}\int (X_s - \hat{X}_s)(dY_s^r) + \frac{r}{2}\int (X_s - \hat{X}_s)^2 ds\right]$$

By lemma 2.3.10, as we know Y_s^r is a martingale so, By Optimal stopping theorem, and also the fact that $dY_t^r = \sqrt{r}\hat{X}_t dt + dB_t$ we will have:

$$I^r(X; Y^r) = \frac{r}{2}E\left[\int (X_s - \hat{X}_s)^2 ds\right]$$

□

2.3.3 Analysis of the result

As demonstrated in Theorem 2.3.9, which is also called Duncan's theorem[1] [5], the mutual information of an input-output stochastic system is intrinsically linked to the minimum mean square error (MMSE) of the signal with respect to the observations. This relationship

[1]Duncan's result is a special case of our result where r (signal-to-noise ratio) is equal to 1

is fundamental in the field of communication systems, bridging between respective filtering problems and information theory. Duncan's theorem has inspired numerous extensions. For instance,Guo et al [6] showed the same relation for stochastic input-output systems with feedback, which is a key concept in the field of multiple-input multiple-output (MIMO) [25]. Moreover, Chayat et al., expanding directly on Duncan's work, derived a bound on the information rate in an AWGN channel, providing a valuable tool for diverse communication applications [8]. Also, Piera et al [13] find the same relation between the mutual information and MMSE for the stochastic input-output system described in 2.2.1.

2.4 Relationship of causal and noncausal MMSEs

In this section, we will explore the causal and noncausal minimum mean square error (MMSE) as discussed by Verdu et al. in their seminal work [11]. This integration is a cornerstone in communication theory, as it bridges the gap between two distinct types of MMSE estimators. Significantly, this result is applicable to any input signal distribution and additive white Gaussian noise channels. Moreover, recent advancements, such as Orthogonal Approximate Message-Passing (OAMP) [26], demonstrate the continued relevance and applicability of these foundational findings.

First, we begin by defining the causal and non-causal minimum mean square error (MMSE) in Section 2.4.1. Following this, Section 2.4.2 shows the relationship between causal MMSE (CMMSE) and non-causal MMSE (NCMMSE), revisiting the work of Guo, Shamai, and Verdu [11]. Finally, Section 2.4.3 discusses the significance of this finding for advancements in information theory.

2.4.1 Preliminaries

Definition 2.4.1 (causal minimum mean square error (CMMSE))**.** The causal minimum mean square error (CMMSE) (CMMSE) for any time $\forall t \in [0, T]$ is defined by

$$\text{CMMSE}\ (t, r) = \mathrm{E}\left\{\left(X_t - \hat{X}_t\right)^2\right\}$$

Where $\hat{X}_s = E\left[X_s \mid \mathcal{F}_s\right]$, and $\mathcal{F}_s$ is the sigma-algebra generated by Y_t^r

Definition 2.4.2 (noncausal minimum mean square error (NCMMSE))**.** The non causal minimum mean square error (NCMMSE) for any time $\forall t \in [0, T]$ is defined by

$$\text{NCMMSE}\ (t, r) = \mathrm{E}\left\{\left(X_t - \tilde{X}_t\right)^2\right\}$$

Where $\tilde{X}_s = E\left[X_s \mid \mathcal{F}_T\right]$, and $\mathcal{F}_T$ is the sigma-algebra generated by Y_t^r from 0 to T all over the time.

Definition 2.4.3 (Average CMMSE)**.** The average of CMMSE is defined by:

$$\text{CMMSE(r)} \ = \frac{1}{T} \int_0^T \text{CMMSE}(t, r) \mathrm{d}t$$

and similarly for the average of NCMMSE is defined by:

$$\text{NCMMSE(r)} \;= \frac{1}{T}\int_0^T \text{NCMMSE}(t,r)\text{d}t$$

Observation 2.4.4. We can write theorem 2.3.9 using definition 2.4.3 we can write:

$$I^r(X;Y^r) = \frac{r}{2}\text{CMMSE}(r)$$

As we can see it is only a function of r, so alternatively we can write:

$$I(r) = \frac{r}{2}\text{CMMSE}(r)$$

2.4.2 Calculation of derivative of Mutual information

To establish the relationship between CMMSE and NCMMSE, we begin by deriving the mutual information's derivative with respect to the signal-to-noise ratio. The stochastic input-output model employed in this analysis is detailed in Section 2.2.1. Notably, this model does not impose any specific distribution assumptions on the signal and, the noise is assumed as Brownian motion.

Theorem 2.4.5. If we are in AWGN channel as it is described in 2.2.2.1, and regardless of the distribution of the signal X_t when we have an expected value of the square of the signal is finite or $\int_0^T \text{E}X_t^2 \text{ d}t < \infty$, then we have:

$$\frac{\text{d}}{\text{d}r}I(r) = \frac{1}{2}\text{NCMMSE(r)}$$

Before proving theorem 2.4.5, we need to first prove a lemma

Lemma 2.4.6. If $r \to 0$, and the input-output the AWGN channel is given by 2.2.2.1:

$$\text{dY}_t^r = \sqrt{\delta}X_t \text{ d}t + \text{d}W_t, \quad t \in [0,T]$$

If we have $\int_0^T \text{E}X_t^2 \text{ d}t < \infty$ then the mutual information is given by:

$$\lim_{\text{r}\to 0}\frac{I(r)}{r} = \frac{1}{2}\int_0^T \text{E}\left(X_t - \text{E}X_t\right)^2 \text{ d}t$$

Proof. By theorem 2.3.9 we have:

$$I(r) = \frac{r}{2} E\left[\int (X_s - \hat{X}_s)^2 ds\right]$$

Then we can write:

$$\lim_{\mathrm{r}\to 0} \frac{I(r)}{r} = \lim_{\mathrm{r}\to 0} \frac{1}{2} E\left[\int (X_s - \hat{X}_s)^2 ds\right]$$

As we know $\lim_{\mathrm{r}\to 0} \hat{X}_s = \lim_{\mathrm{r}\to 0} E\left[X_s \mid \mathcal{F}_s\right] = \mathbb{E}X_t$, so the proof is done and we have:

$$\lim_{\mathrm{r}\to 0} \frac{I(r)}{r} = \frac{1}{2}\int_0^T \mathrm{E}\left(X_t - \mathrm{E}X_t\right)^2 \,\mathrm{d}t$$

□

Lemma 2.4.7. We have dW_1 and dW_2 which are independent Wiener processes with variances σ_1^2 and σ_2^2 respectively. For any two constants a, b, then:

$$\mathrm{Var}(a\, dW_1 + b\, dW_2) = a^2\sigma_1^2 dt + b^2\sigma_2^2 dt$$

Proof. Since both dW_1 and dW_2 are Wiener processes we know $\mathbb{E}[dW_1] = 0, \quad \mathbb{E}[dW_2] = 0$, and $\mathrm{Var}(dW_1) = \sigma_i^2 dt, \quad i = 1, 2,$, then also by their independence $Cov(dX_t, dY_t)] = 0$. Now let us Consider the random variable $Z = a\, dW_1 + b\, dW_2$. The variance of Z is:

$$\mathrm{Var}(Z) = \mathrm{Var}(a\, dW_1 + b\, dW_2).$$

Using the properties of variance for independent random variables:

$$\mathrm{Var}(a\, dW_1 + b\, dW_2) = \mathrm{Var}(a\, dW_1) + \mathrm{Var}(b\, dW_2).$$

Since $\mathrm{Var}(a\, dW_1) = a^2\mathrm{Var}(dW_1)$ and $\mathrm{Var}(b\, dW_2) = b^2\mathrm{Var}(dW_2)$, we have:

$$\mathrm{Var}(a\, dW_1 + b\, dW_2) = a^2\sigma_1^2\, dt + b^2\sigma_2^2\, dt.$$

□

Given that these lemmas are necessary for the proof, we now proceed to prove Theorem 2.4.5.

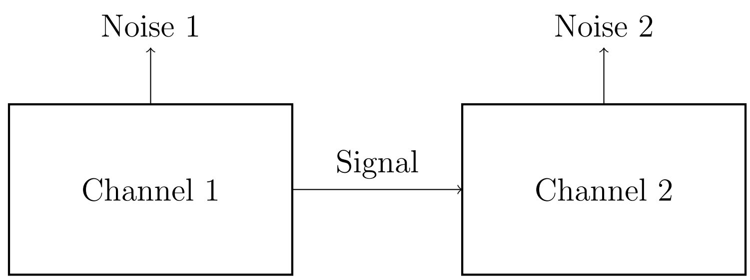

Figure 2.1: The model of the channel

Proof. Let's take the system in figure 2.1, where each of these channels has different variances so we can write this model in the form of below:

$$dY_1 = X_t dt + \sigma_1 dW_t \tag{2.9}$$

$$dY_2 = dY_1 + \sigma_2 d\tilde{W}_t \tag{2.10}$$

Where $Y_2(t)$ is the output of the system shown in figure 2.1, and X_t is the input of channel 1 in the system, also we know $\sigma_1^2 = \frac{1}{r+\delta}$, and $\sigma_1^2 + \sigma_2^2 = \frac{1}{r}$ By multiplying equation 2.9 by δ and equation 2.10 by r, then subtracting the latter from the former, we will have:

$$\delta dY_1 + r dY_1 + r\sigma_2 d\tilde{W}_t = r dY_2 + \delta X_t dt + \delta\sigma_1 dW_t$$

Now we can write it as:

$$(\delta + r) dY_1 = r dY_2 + \delta X_t dt + \delta\sigma_1 dW_t - r\sigma_2 d\tilde{W}_t$$

then it can be written as:

$$(\delta + r) dY_1 = r dY_2 + \delta X_t dt + \sqrt{\delta} d\bar{W}_t$$

where $d\bar{W}_t = \frac{\delta\sigma_1 dW_t - r\sigma_2 d\tilde{W}_t}{\sqrt{\delta}}$, which is also a wiener process itself based on the common properties of wiener processes.[27]. Now we should show that $d\bar{W}_t$ variance is also equal to 1 so we have by lemma 2.4.7 :

$$Var(d\bar{W}_t) = \frac{\delta^2\sigma_1^2 - r^2\sigma_2^2}{\delta} = \frac{\delta^2 \frac{1}{r+\delta} - r^2(\frac{1}{r} - \frac{1}{r+\delta})}{\delta}$$

Then we have:

$$Var(d\bar{W}_t) = \frac{\frac{\delta^2 - r^2}{r+\delta} + (\frac{1}{r+\delta})}{\delta} = \frac{\delta - r + r}{\delta} = 1$$

From the system structure in figure 2.1, we can write the information relation between the two sections of the system so we have:

$$I(X;Y_1^r) - I(X;Y_2^r) = I(r+\delta) - I(r) = \Delta I(r)$$

Then by Markov property of the X, Y_1, Y_2 are Markov chains so we can write it as [23]:

$$I(X;Y_1^r) - I(X;Y_2^r) = I\left(X;Y_1^r \mid Y_2^r\right)$$

Applying Lemma 2.4.6, we can then express the following for the entire interval from zero to a given T:

$$I\left(X;Y_1^r \mid Y_2^r\right) = \frac{\delta}{2}\int_0^T \mathrm{E}\big[\big(X_t - \mathrm{E}[X_t \mid Y_2^T]\big)^2\big]dt$$

So now by definition of NCMMSE in def 2.4.2 we can write it as we were supposed too:

$$\Delta I(r) = \frac{\delta}{2T}\int_0^T \mathrm{NCMMSE}(t, \mathrm{snr})dt$$

Then by taking limits, and definition of derivative we will have:

$$\frac{\mathrm{d}}{\mathrm{d}r} I(r) = \frac{1}{2}\mathrm{NCMMSE(r)}$$

This completes the proof.

□

2.4.3 Relation Between the NCMMSE and CMMSE

If we take a look at theorem 2.4.5, and theorem 2.3.9 we will get an interesting result which is the first mentioned in the work of Verdu et al. in their paper [11].

Proposition 2.4.8. In the classic form of a channel under a wiener process as noise we have:

$$\mathrm{CMMSE}(r) = \frac{1}{\mathrm{r}}\int_0^{\mathrm{r}} \mathrm{NCMMSE}(s)ds$$

Proof. The proof is straightforward. It involves using the results from both Theorem 2.4.5 and Theorem 2.3.9, and then formulating an equation for the mutual information. □

The most intriguing aspect of this result is that the relationship holds regardless of the input signal's distribution. In other words, this is a general property of a channel when the noise is modeled as Brownian motion. This result has been extended to various areas. For instance, the work of Montanari et al [28] established a fundamental relation between belief propagation (BP) and maximum a posteriori (MAP) decoding in general memory-less channels . Additionally, Verdú et al. extended this result to Poisson channels, discovering the relationship between Mutual Information and MMSE, which is significant for both communication and network problems [12]. In the Poisson channel context, the conditional expectation serves a role analogous to the MMSE in Gaussian channels.

Chapter 3

Mutual information in the presence of Ornstein–Uhlenbeck process as noise

In this chapter, we discuss the mutual information of input-output stochastic systems with Ornstein–Uhlenbeck (O-U) process. Diffusion processes, including special cases such as the O-U process and geometric Brownian motion process, have significant applications in fields like signal processing and finance [29]. This chapter will explore the relationship between mutual information and the estimation error of the input signal based on observations in a stochastic input-output channel influenced by O-U noise

The first section 3.1 defines the general diffusion process, focusing particularly on the O-U process as a special case, and introduces the general stochastic differential equation (SDE). Section 3.2 defines our channel equation and outlines the primary problem we aim to solve. In section 3.3, we will provide a closed-form RND for the stochastic channel with O-U noise. Subsequently, section 3.4 will introduce the relationship between mutual information over the channel and the estimation of the input signal. Finally, in section 3.5, we will analyze our results and demonstrate their compatibility with the findings discussed in chapter 2.

3.1 Preliminaries

In this chapter, the Brownian motions considered satisfy $B(0) = 0$, implying that all such Brownian motions are Wiener processes. Additionally, all stochastic integrals are expressed as Itô integrals, a standard tool in stochastic calculus.

3.1.1 General stochastic differential equation

Definition 3.1.1. A generalized SDE for a diffusion process is given by:

$$\mathrm{d}X_t = \mu(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t$$

where $\mu(X_t, t)$ is the drift term, $\sigma(X_t, t)$ is the diffusion term, and B_t denotes a Wiener process.

Observation 3.1.2. The definition 3.1.1 also can we written in terms of stochastic integral:

$$X_t = X_0 + \int_0^t \mu(X_t, t)\,\mathrm{d}t + \int_0^t \sigma(X_t, t)\,\mathrm{d}B_t$$

3.1.2 Ornstein Uhlenbeck stochastic differential equation

Definition 3.1.3 (Ornstein Uhlenbeck Process)**.** In definition 3.1.1 suppose the drift term is $\mu(X_t, t) = -\theta X_t$, and the diffusion term is $\sigma(X_t, t) = \sigma$, where $\theta \geq 0$, and $\sigma \geq 0$ are two constants then we will have:

$$\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

Then X_t is called the O-U process.

The O-U process, also known as the Gauss-Markov process, possesses Markovian properties and is a Gaussian process. These properties make it useful for modeling various real-world problems, including quantum communication.[16] Now we need to write the O-U process in the stochastic integral form which is common to use in channel modelling

Theorem 3.1.4. Let X_t be a stochastic process which satisfies the following SDE:

$$\mathrm{d}X_t = -\theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$$

Then we can write X_t as:

$$X_t = X_0 e^{-\theta t} + \int_0^t e^{\theta(s-t)}\sigma\,\mathrm{d}B_s$$

Proof. First we multiply $e^{\theta t}$ to both sides so we will have:

$$\mathrm{d}X_t e^{\theta t} = -e^{\theta t}\theta X_t \,\mathrm{d}t + e^{\theta t}\sigma \,\mathrm{d}B_t$$

So we can reform the expression as below:

$$\mathrm{d}X_t e^{\theta t} + e^{\theta t}\theta X_t \,\mathrm{d}t = e^{\theta t}\sigma \,\mathrm{d}B_t$$

So it can be written as:

$$d(X_t e^{\theta t}) = e^{\theta t}\sigma \,\mathrm{d}B_t$$

Now let us take integral from both sides:

$$X_t e^{\theta t} - X_0 e^{\theta 0} = \int_0^t e^{\theta s}\sigma \,\mathrm{d}B_s$$

So finally we get:

$$X_t = X_0 e^{-\theta t} + \int_0^t e^{\theta(s-t)}\sigma \,\mathrm{d}B_s$$

□

Lemma 3.1.5. (Theorem 11.1, [23]) Let (B_t) be a Brownian motion for all $0 \leq t$. Suppose X_t is an adapted stochastic process such that X_t is measurable with respect to the same σ-algebra, and X_t be square integrable for all $t \geq 0$. Then we have:

$$\mathbb{E}\left[\int_0^T X_s \, dB_s\right] = 0$$

Proposition 3.1.6. Let X_t be an O-U stochastic process then we have:

$$\mathbf{E}[X_t] = \mathbf{E}[X_0]e^{-\theta t}$$

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} + \frac{\sigma^2}{2\theta}\left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right)$$

Proof. For the first statement, we can easily take the expected value from the process, and by lemma 3.1.5 the second stochastic integral is zero so we can write:

$$\mathbf{E}[X_t] = \mathbf{E}[X_0 e^{-\theta t} + \int_0^t e^{\theta(s-t)}\sigma \,\mathrm{d}B_s]$$

$$\mathbf{E}[X_t] = \mathbf{E}[X_0]e^{-\theta t} + \mathbf{E}[\int_0^t e^{\theta(s-t)}\sigma \, \mathrm{d}B_s]] = \mathbf{E}[X_0]e^{-\theta t}$$

For the second statement, we can write based on the covariance definition:

$$\mathrm{Cov}(X_s, X_t) = \mathbb{E}[(X_t - \mathbb{E}[X_t])(X_s - \mathbb{E}[X_s])]$$

So based on proposition3.1.6 (a) ,and theorem 3.1.4 we can write:

$$\mathrm{Cov}(X_s, X_t) = \mathbb{E}[(X_0 e^{-\theta t} + \int_0^t e^{\theta(u-t)}\sigma \, \mathrm{d}B_u - \mathbf{E}[X_0]e^{-\theta t})(X_0 e^{-\theta s} + \int_0^s e^{\theta(r-t)}\sigma \, \mathrm{d}B_r - \mathbf{E}[X_0]e^{-\theta s})]$$

So by simplification, we can write:

$$\mathrm{Cov}(X_s, X_t) = \mathbb{E}[(X_0 - \mathbb{E}[X_0])^2]e^{-\theta(s+t)} + \mathbb{E}[(\int_0^t e^{\theta(u-t)}\sigma \, \mathrm{d}B_u)(\int_0^s e^{\theta(r-t)}\sigma \, \mathrm{d}B_r)]$$

By Fubini's theorem, we can write:

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} + \mathbb{E}[(\int_0^t \int_0^s \sigma^2 e^{\theta(u-t)} e^{\theta(r-t)} \, \mathrm{d}B_u \mathrm{d}B_r)]$$

By Ito's isometry lemma, we can change the stochastic integral to a Riemann integral. Then we can write:

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} + \mathbb{E}[(\int_0^{\min(t,s)} \sigma^2 e^{2\theta(u-t)} \, \mathrm{d}u)]$$

So easily we can calculate the integral:

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} - \frac{\sigma^2}{2\theta} e^{-\theta(s+t)} \left(1 - e^{2\theta \min(s,t)}\right)$$

By easy equality, we can write it as a proper function:

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} + \frac{\sigma^2}{2\theta} \left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right)$$

□

3.1.3 Integrated Ornstein Uhlenbeck stochastic differential equation

Definition 3.1.7. (Integrated Ornstein Uhlenbeck) Let Y_t be a stochastic process, and X_t be an O-U process, which satisfies the stochastic differential equation given below:

$$\mathrm{d}X_t = -\theta X_t \,\mathrm{d}t + \sigma \,\mathrm{d}B_t$$

If Y_t satisfies:

$$\mathrm{d}Y_t = X_t dt,$$

$$Y_t = Y_0 + \int_0^t X_s ds,$$

Then we call Y_t integrated Ornstein-Uhlenbeck process.

Proposition 3.1.8. Let Y_t be an integrated O-U process then we have:

$$Y_t = Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) + \int_0^t \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-s)}\right) dB_s$$

Proof. From definition 3.1.7 we have:

$$Y_t = Y_0 + \int_0^t X_u du$$

Then from theorem 3.1.4 we can write X_s as below:

$$Y_t = Y_0 + \int_0^t (X_0 e^{-\theta u} + \int_0^u e^{\theta(s-u)} \sigma \,\mathrm{d}B_s) du$$

Then we can expand it:

$$Y_t = Y_0 + \int_0^t X_0 e^{-\theta u} du + \int_0^t \int_0^u e^{\theta(s-u)} \sigma \,\mathrm{d}B_s du$$

So now we can integrate over u in our expression:

$$Y_t = Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) + \int_0^t \int_0^u e^{\theta(s-u)} \sigma \,\mathrm{d}B_s du$$

Let us take $\theta(s-u) = A$ then $dA = -\theta du$, then we have:

$$Y_t = Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) - \int_0^t \int_0^u e^{A} \frac{\sigma}{\theta} \,\mathrm{d}B_s dA$$

By Fubini's theorem, we can change the integrals but should change the integral domain. The first integral is from zero to t, and the second one is from zero to u, so $0 \leq r \leq u \leq t$, so when we change the integrals the outer one will be from 0 to t and the internal will be from 0 to $t - r$ where r is a dummy variable :

$$Y_t = Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) - \frac{\sigma}{\theta} \int_0^t \int_0^{t-r} e^A \, dA \mathrm{d}B_r$$

Then we can easily take the internal integral:

$$Y_t = Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) + \int_0^t \frac{\sigma}{\theta}(1 - e^{-\theta(t-r)})\mathrm{d}B_r$$

□

Proposition 3.1.9. Let Y_t be an integrated O-U. Then we have:

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_0] + \frac{\mathbb{E}[X_0]}{\theta}(1 - e^{-\theta t})$$

$$\mathrm{Cov}(Y_t, Y_s) = \int_0^{\min(s,t)} \phi(t, u)\phi(s, u)du$$

where $\phi(t, u) = \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-u)}\right)$

Proof. For the first statement, we can easily take the expected value by using proposition 3.1.8 so we have:

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_0 + \frac{X_0}{\theta}(1 - e^{-\theta t}) + \int_0^t \frac{\sigma}{\theta}(1 - e^{-\theta(t-r)})\mathrm{d}B_r]$$

Then by lemma 3.1.5 we have:

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_0] + \mathbb{E}[\frac{X_0}{\theta}(1 - e^{-\theta t})] + \mathbb{E}[\int_0^t \frac{\sigma}{\theta}(1 - e^{-\theta(t-r)})\mathrm{d}B_r]]$$

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_0] + \frac{\mathbb{E}[X_0]}{\theta}(1 - e^{-\theta t})$$

For the second statement, we can write From the definition of covariance:

$$\mathrm{Cov}(Y_t, Y_s) = \mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_s - \mathbb{E}[Y_s])]$$

Then by proposition 3.1.8, and proposition 3.1.9 (a) we have:

$$\mathrm{Cov}(Y_t, Y_s) = \mathbb{E}[(\int_0^t \frac{\sigma}{\theta}(1-e^{-\theta(t-r)})\mathrm{d}B_r)(\int_0^s \frac{\sigma}{\theta}(1-e^{-\theta(s-r)})\mathrm{d}B_s)]$$

Then by Ito's isometry lemma, we have:

$$\mathrm{Cov}(Y_t, Y_s) = \mathbb{E}[\int_0^t \int_0^s \left(\frac{\sigma}{\theta}\right)^2 \left(1-e^{-\theta(t-r)}\right)\left(1-e^{-\theta(s-u)}\right) dB_r dB_u]$$

$$\mathrm{Cov}(Y_t, Y_s) = \mathbb{E}[\int_0^t \int_0^s \left(\frac{\sigma}{\theta}\right)^2 \left(1-e^{-\theta(t-z)}\right)\left(1-e^{-\theta(s-u)}\right) \delta(z-u) dz du]$$

Then by taking the Riemann integral, and delta properties, we have:

$$\mathrm{Cov}(Y_t, Y_s) = \int_0^{\min(s,t)} \phi(t,u)\phi(s,u)du$$

where $\phi(t,u) = \frac{\sigma}{\theta}\left(1-e^{-\theta(t-u)}\right)$ □

3.1.4 Stationary Ornstein-Uhlenbeck stochastic differential equation

In general stationary noises are more common in modeling communication channels because we have more tools to deal with the model in the stationary form, and also the analysis of the noise in the stationary form is more common in literature and there are lots of algorithms to cancel that.[30] In this section, we want to introduce the stationary form of the O-U process.

Proposition 3.1.10. Let X_t be an O-U stochastic process, and assume that the $X_0 \overset{(d)}{=} N(0,1)$ where $N(0,1)$ is a Gaussian distribution, then the distribution of the process is given below:

$$X_t = X_0 e^{-\theta t} + \int_0^t e^{\theta(s-t)}\sigma\,\mathrm{d}B_s \overset{(d)}{=} N(0,1)e^{-\theta t} + \sigma\sqrt{\frac{1-e^{-2\theta t}}{2\theta}}N'(0,1), \quad \forall t \in T$$

Where $N'(0,1)$ is a normal distribution independent of $N(0,1)$

Before proving Proposition 3.1.10, we need to state and prove a Theorem that will be used in our proof.

Theorem 3.1.11. (Dambis-Dubins-Schwarz theorem) Let $(M_t)_{t\geq 0}$ be a continuous local martingale with $M_0 = 0$ and let $\langle M\rangle_t$ denote its quadratic variation process. Then there exists a standard Brownian motion $(B_t)_{t\geq 0}$ such that:

$$M_t = B_{\langle M\rangle_t} \quad \text{for all } t \geq 0.$$

Let us go back to the proof of proposition 3.1.10.

Proof. By theorem 3.1.4 we have:

$$X_t = X_0 e^{-\theta t} + \sigma \int_0^t e^{\theta(s-t)}\, \mathrm{d}B_s$$

Now we can write this in terms of Brownian motion by lemma 2.3.10 as we know stochastic Ito's integral is a martingale so we can write it as a Brownian motion with the quadratic variation $\int_0^t e^{2\theta(s-t)}ds$ then we can write as a result of Dambis-Dubins-Schwarz theorem we have:

$$X_t = X_0 e^{-\theta t} + \sigma B_{\int_0^t e^{-2\theta(t-s)}ds}$$

Then we can write the distribution based on these results easily because the Brownian motion is normally distributed which is independent of the X_0 distribution. So we have:

$$X_t = X_0 e^{-\theta t} + \sigma B_{\int_0^t e^{-2\theta(t-s)}ds} \overset{(\mathrm{d})}{=} N(0,1)e^{-\theta t} + \sigma\sqrt{\frac{1-e^{-2\theta t}}{2\theta}} N'(0,1), \quad \forall t \in T$$

□

Definition 3.1.12. (Stationary Ornstein-Uhlenbeck stochastic process) Let X_t be an O-U stochastic process, and let $X_0 \overset{(\mathrm{d})}{=} N(0,1)$, and $\sigma^2 = 2\theta$ then X_t is a stationary Ornstein-Uhlenbeck stochastic process

We can show the stationary property of the process easily. By proposition 3.1.10 we know that the Stationary Ornstein-Uhlenbeck stochastic process is distributed as below:

$$X_t \overset{(\mathrm{d})}{=} \mathbf{N}(0,1)e^{-\frac{\theta}{2}t} + \sqrt{\frac{1-e^{-2\theta t}}{2\theta}} N'(0,1), \quad \forall t \in T$$

Then based on independency of the $N(0,1)$, and $N'(0,1)$, and $\sigma^2 = 2\theta$ then by lemma 2.4.7 we have:

$$X_t \overset{(\mathrm{d})}{=} N''(0, e^{-\theta t} + \sigma^2 \frac{1-e^{-\theta t}}{2\theta}) \overset{(\mathrm{d})}{=} N''(0, \frac{2\theta e^{-\theta t} + \sigma^2 - \sigma^2 e^{-\theta t}}{2\theta}) \overset{(\mathrm{d})}{=} N''(0,1), \quad \forall t \in T$$

So as we can see the distribution of the stochastic process is normally distributed with a constant variance but to show it is stationary we need to show that the covariance of the process is only related to the difference of time it is pretty straightforward but because it is crucial for our further result we can show it in an observation

Observation 3.1.13. For the process which is defined in definition 3.1.12 we have:

$$\mathbf{E}[X_t] = 0$$

$$\mathrm{Cov}(X_s, X_t) = e^{-\theta|t-s|}$$

Proof. By proposition 3.1.6, and by the definition of our process we have where $\mathrm{Var}(X_0) = 1$, and $\sigma^2 = \theta$, we have:

$$\mathbf{E}[X_t] = \mathbf{E}[X_0]e^{-\theta t} = 0$$

$$\mathrm{Cov}(X_s, X_t) = \mathrm{Var}(X_0)e^{-\theta(s+t)} + \frac{\sigma^2}{2\theta}\left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right) = e^{-\theta(s+t)} + \left(e^{-\theta|t-s|} - e^{-\theta(t+s)}\right)$$

$$\mathrm{Cov}(X_s, X_t) = e^{-\theta|t-s|}$$

So as we can see the process is stationary.

□

3.2 Stochastic model

In this section, we will define our problem model which is technically similar to the previous model which was defined in the chapter 2. We are going to give 3 models which have different noises which are discussed in section 3.1.

3.2.1 Channel with integrated O-U process as noise

Let the observation channel represented below:

$$dY_t^{\theta,\sigma} = h\left(X_u\right)du + d\tilde{N}_t \tag{3.1}$$

$$Y_t^{\theta,\sigma} = \int_0^t h\left(X_u\right)du + \tilde{N}_t \tag{3.2}$$

Let $Y_t^{\theta,\sigma}$ denote the observation stochastic process, X_t the input stochastic process, and $\tilde{N}_t$ the integrated O-U process defined in Definition 3.1.7. Here, σ and θ are constants in the stochastic differential equation generating the O-U process which is inside generating the corresponding process. Additionally, let $h(u)$ be a differentiable function over its domain, with $h'(u) \in L^2[0,T]$

3.2.2 Channel with O-U process as noise

Consider the observation channel represented below[18]:

$$dY_t^{\theta,\sigma} = h\left(X_u\right) du + dN_t \tag{3.3}$$

$$Y_t^{\theta,\sigma} = \int_0^t h\left(X_u\right) du + N_t \tag{3.4}$$

Where $Y_t^{\theta,\sigma}$ is observation stochastic process, X_t is the input stochastic process, and N_t is an Ornstein Uhlenbeck process which follows the definition 3.1.3, and σ, and θ are the given constants in the stochastic differential equation which generates the Ornstein Uhlenbeck process. Also, $h(u)$ is differentiable function over the domain, and $h'(u) \in L^2[0,T]$.

3.2.3 Channel with Stationary O-U process as noise

Let the observation channel represented below:

$$dY_t^{\theta,\sigma} = h\left(X_u\right) du + d\tilde{\tilde{N}}_t \tag{3.5}$$

$$Y_t^{\theta,\sigma} = \int_0^t h\left(X_u\right) du + \tilde{\tilde{N}}_t \tag{3.6}$$

Let $Y_t^{\theta,\sigma}$ denote the observation stochastic process, X_t the input stochastic process, and $\tilde{\tilde{N}}_t$ the stationary Ornstein-Uhlenbeck process defined in Definition 3.1.12. Here, σ and θ are constants such that $\sigma^2 = 2\theta$ in the SDE. Additionally, let $h(u)$ be a differentiable function over its domain, with $h'(u) \in L^2[0,T]$ Technically this model as we discussed have a stationary noise inside it which is more common in the noise modeling in communication systems

3.3 Calculation of Radon–Nikodym derivative

In this section, we calculate the RND of the product measure of signal and observation signal with respect to the respective measures of the observation and signal.

3.3.1 RND over the channel with integrated O-U process as noise

In this section, we assume the initial value for the noise is $N_0 = 0$. Therefore, using the theorem 3.1.4, we can rewrite the equation 3.1, and 3.2 as:

$$dY_t^{\theta,\sigma} = h(X_t)\,dt + O_t^{\sigma,\theta}dt$$
$$Y_t^{\theta,\sigma} = \int_0^t h(X_u)\,du + \int_0^t O_s^{\sigma,\theta}ds,$$

where the $O_t^{\sigma,\theta}$ is an O-U process with respected σ, and θ. Therefore, using proposition 3.1.8, we can rewrite the SDE in terms of stochastic integral, which is a common representation for channel modeling:

$$Y_t^{\theta,\sigma} = \int_0^t h(X_u)\,du + \int_0^t \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-s)}\right)dB_s$$

Now we can rewrite the equation in terms of the function ϕ which we defined before:

$$Y_t^{\theta,\sigma} = \int_0^t h(X_u)\,du + \int_0^t \phi(t,u)dB_s, \tag{3.7}$$

where $\phi(t,u) = \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-u)}\right)$

In order to calculate RND of the stochastic input-output channel with integrated O-U process as noise, we need to define the respective reproducing kernel Hilbert space (RKHS)[18][20] [1]

[1]This theory is firstly developed by Aronszajn in his paper in 1950 but the version we are using is more connected to our case which is in Mandels paper [18]

3.3.1.1 Reproducing Kernel Hilbert space method

Definition 3.3.1 (Definition 1, [20]). A Hilbert space H consisting of real-valued functions defined on a set $\mathbf{T}$ is called a Reproducing Kernel Hilbert Space if there exists a function $K : \mathbf{T} \times \mathbf{T} \to \mathbb{R}$, called the reproducing kernel of H, such that the following properties hold for every $t \in \mathbf{T}$ and every $g \in H$:

1. For each $t \in \mathbf{T}$, the function $K(\cdot, t)$ belongs to H.
2. For each $g \in H$ and $t \in \mathbf{T}$,

$$(g(\cdot), K(\cdot, t)) = g(t),$$

where $(\cdot, \cdot)$ denotes the inner product in H.

The reproducing kernel K possesses the following fundamental properties:

1. If a reproducing kernel exists for H, then it is unique.
2. The set of kernel functions $\{K(\cdot, t) : t \in \mathbf{T}\}$ spans H.
3. The reproducing kernel K is nonnegative definite, meaning that for any finite set of points $t_1, \dots, t_n \in \mathbf{T}$ and any real numbers $a_1, \dots, a_n$,

$$\sum_{i,j=1}^{n} K(t_i, t_j) a_i a_j \geq 0.$$

While numerous intriguing properties, such as the uniqueness of these specific Reproducing Kernel Hilbert Spaces (RKHS), exist, they lie beyond the scope of this thesis. Our focus here is primarily on employing these tools to establish our model.[2] We will introduce our model with essential theorems, presented without proof.

Theorem 3.3.2 (Theorem 2.1, [18]). A symmetric nonnegative definite function K on $\mathbf{T} \times \mathbf{T}$ generates a unique Hilbert space, denoted by $H(K)$ or $H(K, \mathbf{T})$, of which K is the reproducing kernel.

Theorem 3.3.3 (Aronszajn, [20]). Suppose $K_{\mathbf{T}}$, defined on $\mathbf{T} \times \mathbf{T}$, is the reproducing kernel of the Hilbert space $H(K_{\mathbf{T}})$ with norm $\|\cdot\|$. Let $\mathbf{T}' \subset \mathbf{T}$, and let $K_{\mathbf{T}'}$ be the restriction of $K_{\mathbf{T}}$ on $\mathbf{T}' \times \mathbf{T}'$. Then $H(K_{\mathbf{T}'})$ consists of all functions f in $H(K_{\mathbf{T}})$ restricted to $\mathbf{T}'$. Furthermore, for such a restriction $f' \in H(K_{\mathbf{T}'})$, the norm $\|f'\|_{H(K_{\mathbf{T}'})}$ is the minimum of $\|f\|_{H(K_{\mathbf{T}})}$ for all $f \in H(K_{\mathbf{T}})$ whose restriction to $\mathbf{T}'$ is f'.

[2] A comprehensive definition of a Hilbert space is omitted here, as it is readily available in standard texts on functional analysis [31].

Let $K(s,t)$ denote a nonnegative definite function defined for $s,t \in \mathbf{T}$. According to Theorem 3.3.2, there exists a Reproducing Kernel Hilbert Space (RKHS) $H(K,\mathbf{T})$. If we restrict s and t to a subset $\mathbf{T}' \subseteq \mathbf{T}$, the function K remains nonnegative definite on $\mathbf{T}' \times \mathbf{T}'$, corresponding to an RKHS $H(K,\mathbf{T}')$ of functions defined on $\mathbf{T}'$.

Using Theorem 3.3.3, it follows that $H(K_{\mathbf{T}'})$ exactly contains all functions in $H(K_{\mathbf{T}})$ when restricted to $\mathbf{T}'$. Moreover, $\|f'\|_{H(K_{\mathbf{T}'})}$ is minimized among $\|f\|_{H(K_{\mathbf{T}})}$ norms for $f \in H(K_{\mathbf{T}})$ agreeing with f' on $\mathbf{T}'$.

If $K(s,t)$ serves as the covariance function for a mean zero process Z_t over $t \in \mathbf{T}$, Theorem 3.3.2 establishes a unique RKHS $H(K,\mathbf{T})$ with K as its reproducing kernel. Furthermore, there exists a linear, one-to-one, inner product-preserving map (congruence) between $H(K)$ and $\overline{\mathrm{sp}}^{L^2}\{Z_t, t \in \mathbf{T}\}$, mapping $K(\cdot,t)$ to Z_t. Denote the image of any $h \in H(K,\mathbf{T})$ under such a congruence as $\langle Z, h\rangle \in \overline{\mathrm{sp}}^{L^2}\{Z_t, t \in \mathbf{T}\}$. [18]

So now with this knowledge, we can develop our kernel for this problem and make it compatible with this theorem. So let's go back to the equation 3.7, and take the noise (O-U process) separately:

$$N_t = \int_0^t \phi(t,s) dB_s$$

Now we can define the K function which was mentioned before from the covariance in the proposition 3.1.9:

$$K(s,t) = \mathrm{Cov}(Y_t, Y_s) = \int_0^{\min(s,t)} \phi(t,u)\phi(s,u) du$$

Then we can define the reproducing kernel Hilbert space as below:

Definition 3.3.4. the RKHS for the stochastic input-output channel described in 3.1 is:

$$\mathrm{RKHS}(K) = \left\{ f \mid f(t) = \int_0^t \phi(t,s) f^*(s) ds \quad t \in [0,T] \right\}$$

for some (necessarily unique) $f^* \in \overline{\mathrm{span}}^{L^2}\left\{\phi(t,\cdot)1_{[0,t]}(\cdot),\ t \in [0,T]\right\}$ we can define inner product as below:

$$(f_1, f_2)_{\mathrm{RKHS}(K)} = \int_0^T f_1^*(s) f_2^*(s) ds$$

where

$$f_1(t) = \int_0^t \phi(t,u) f_1^*(u) du \quad \text{and} \quad f_2(t) = \int_0^t \phi(t,u) f_2^*(u) du$$

Now we can introduce the key concept that will allow us to redefine our system dynamics and leverage this to determine the corresponding measure over that space.

Theorem 3.3.5. For $0 \leq t \leq T$, let $K(\cdot, t)^*$ be defined as $\phi(t, \cdot)1_{[0,t]}(\cdot)$. Then, from equations (2.1) and (2.2), $K(\cdot, t) \in \text{RKHS}(K)$. Furthermore, the reproducing property holds as follows: Suppose $h(t) = \int_0^t \phi(t, u)h^*(u)\,du \in \text{RKHS}(K)$. Then,

$$(h, K(\cdot, t))_{H(K)} = \int_0^T h^*(u)K(\cdot, t)^*\,du = \int_0^t h^*(u)\phi(t, u)\,du = h(t).$$

Proof. By definition, for $0 \leq t \leq T$, let

$$K(\cdot, t)^* = \phi(t, \cdot)1_{[0,t]}(\cdot).$$

From the form of the kernel defined previously, we have that $K(\cdot, t) \in H(K)$. Then, we check the reproducing property. Let $h(t) = \int_0^t \phi(t, u)h^*(u)\,du \in H(K)$. Now based on the definition of the space in definition 3.3.4 we can write the inner product as below:

$$(h, K(\cdot, t))_{\text{RKHS}(K)} = \int_0^T h^*(u)K(\cdot, t)^*\,du.$$

Using $K(\cdot, t)^* = \phi(t, \cdot)1_{[0,t]}(\cdot)$, we obtain:

$$(h, K(\cdot, t))_{H(K)} = \int_0^T h^*(u)\phi(t, u)1_{[0,t]}(u)\,du.$$

Since $1_{[0,t]}(u)$ is 1 for $0 \leq u \leq t$ and 0 otherwise, the $1_{[0,t]}$ can be write as:

$$(h, K(\cdot, t))_{H(K)} = \int_0^t h^*(u)\phi(t, u)\,du.$$

By the definition of $h(t)$, this is exactly:

$$(h, K(\cdot, t))_{H(K)} = \int_0^t h^*(u)\phi(t, u)\,du = h(t).$$

□

Theorem 3.3.6. (thorem 4D, [32]) Over the same space $\text{RKHS}(K)$ and $\overline{\text{sp}}^{L^2}\{Z_t, t \in \mathbf{T}\}$ we w ill have:

$$\langle Z, g\rangle = \int_0^T g^*(u)\,dW_u.$$

Given the RKHS method discussed in this section, we can define an RKHS for the kernel mentioned in our problems, such as $\phi(t,u) = \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-u)}\right)$. This was the kernel function in equation 3.7 for a stochastic input-output channel with integrated Ornstein-Uhlenbeck (O-U) process noise. Alternatively, we can use the corresponding kernel for the O-U process as noise. This RKHS allows us to rewrite the channel model in a form that enables the calculation of the RND of the channel.

3.3.1.2 RKHS over int O-U process filtration

This section first establishes the reproducing kernel Hilbert space for the stochastic input-output channel with the integrated O-U process as noise. Then by utilizing this framework, we determine the RND of the product measure of the input and observation signals with respect to the measure of the signal and observation.

Theorem 3.3.7. Let $\phi(t,u) = \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-u)}\right)$, where $0 \leq u \leq t \leq T$. Consequently, the closure of the space spanned by $\phi(t,\cdot)1_{[0,t]}(\cdot)$ in $L^2[0,T]$ is $L^2[0,T]$.

Before proving the theorem, it is necessary to state the Leibniz integral rule. Although the proof of this lemma is a well-known result in calculus and will not be included in this thesis, it can be found in [33]. This rule is essential for our subsequent proof.

Lemma 3.3.8 (Chapter 8, theorem 3, [33])**.** The *Leibniz integral rule* for differentiation under the integral sign states that for an integral of the form

$$\int_{a(x)}^{b(x)} f(x,t)\,dt$$

where $a(x), b(x)$, are not infinite, then the derivative of that will be:

$$\frac{d}{dx}\left(\int_{a(x)}^{b(x)} f(x,t)\,dt\right) = f(x,b(x)) \cdot \frac{d}{dx}b(x) - f(x,a(x)) \cdot \frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)\,dt,$$

Proof. (3.3.7) Let $f \in L^2[0,T]$ such that $f \perp \phi(t,\cdot)1_{[0,t]}(\cdot)$ for all $\forall t \in [0,T])$. Then we can write the definition of the perpendicular function as an integral equation:

$$\int_0^t f(u)\phi(t,u)du = 0 \quad \text{for all } t.$$

Now let's substitute the function $\phi(t,u) = \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-u)}\right)$ inside the equation:

$$\frac{\sigma}{\theta}\int_0^t f(u)\left(1 - e^{-\theta(t-u)}\right) du = 0.$$

Both θ and σ are positive constants, allowing us to cancel them out. Thus, we can expand the integral as follows:

$$\int_0^t f(u)du - \int_0^t e^{-\theta(t-u)} f(u)du = 0$$

Now by lemma 3.3.8 we can take the derivative from the previous expression with respect to t so we got:

$$\frac{d}{dt}\left(\int_0^t f(u)\, du - \int_0^t e^{-\theta(t-u)} f(u)\, du\right) = f(t) - \left(\frac{d}{dt}\left(\int_0^t e^{-\theta(t-u)} f(u)\, du\right)\right)$$

By apply again lemma 3.3.8 we have:

$$\frac{d}{dt}\left(\int_0^t f(u)\, du - \int_0^t e^{-\theta(t-u)} f(u)\, du\right) = f(t) - \left(e^{-\theta(t-t)} f(t) + \int_0^t \frac{\partial}{\partial t}\left(e^{-\theta(t-u)}\right) f(u)\, du\right)$$

Then by taking the partial differentiation;

$$f(t) - \left(f(t) + \int_0^t -\theta e^{\theta u} f(u)\, du\right) = 0$$

By easy simplification, we got finally:

$$\int_0^t \theta e^{\theta u} f(u)\, du = 0$$

Since θ positive so:

$$\int_0^t e^{\theta u} f(u)\, du = 0$$

Because $e^{\theta u}$ is always positive over the domain of integral this implies that :

$$f(t) = 0 \quad \text{almost every where at } t$$

So we proved that the closure of the space spanned by $\phi(t,\cdot)1_{[0,t]}(\cdot)$ in $L^2[0,T]$ is $L^2[0,T]$ $\square$

So by theorem 3.3.7 we can write the reproducing kernel Hilbert space easily:

$$\text{RKHS}(K) = \left\{ f : f(s) = \int_0^s \phi(s,u) f^*(u) du, \quad \exists f^* \in L^2[0,T] \right\} \tag{3.8}$$

Proposition 3.3.9. Over the observation channel described in 3.7, and RKHS described in 3.8, we have:

$$\int_0^t h(X_u)\, du = \int_0^t \phi(t,u) \left[\frac{\theta}{\sigma} h(X_u) + \frac{1}{\sigma} h'(u) \right] du$$

Proof. We will show that the left hand side integral is equivalent to the right hand side one. First, let's substitute the function ϕ inside the integral equation:

$$\int_0^t \phi(t,u) \left[\frac{\theta}{\sigma} h(X_u) + \frac{1}{\sigma} h'(u) \right] du = \int_0^t \frac{\sigma}{\theta} \left(1 - e^{-\theta(t-u)}\right) \left[\frac{\beta}{\sigma} h(X_u) + \frac{1}{\sigma} h'(u) \right] du$$

Then we can separate the integrals:

$$= \int_0^t h(X_u) du + \frac{1}{\theta} \int_0^t h'(u) du - \frac{1}{\theta} \int_0^t e^{-\theta(t-u)} h'(u) du - \int_0^t e^{-\theta(t-u)} h(X_u) du$$

We leave the first integral unchanged, and the second one straightforward, thus focusing solely on the third and fourth expressions:

$$= \int_0^t h(X_u) du + \frac{1}{\theta}(h(X_t) - h(X_0)) - \frac{1}{\theta} \int_0^t e^{-\theta(t-u)} h'(u) du - \int_0^t e^{-\theta(t-u)} h(X_u) du$$

Now be apply the integration by parts method of integration to the third integral and take $d\nu = h'(u)du$, and $u = e^{-\theta(t-u)}$ which leads to $\nu = h(X_u)$, and now we can use $u\nu - \int \nu du$. SO we have:

$$\begin{aligned} &= \int_0^t h(X_u) du + \frac{1}{\theta}(h(X_t) - h(X_0)) \\ &\quad - \frac{1}{\theta} \left(e^{-\theta t} h(X_0) - e^{-\theta t} h(X_t) + \theta \int_0^t e^{-\theta(t-u)} h(X_u) du \right) \\ &\quad - \int_0^t e^{-\theta(t-u)} h(X_u) du \end{aligned}$$

so we finally have:

$$\int_0^t \phi(t,u) \left[\frac{\theta}{\sigma} h(X_u) + \frac{1}{\sigma} h'(u) \right] du = \int_0^t h(X_u)\, du$$

□

Now by proposition 3.3.9 we can rewrite our observation channel model using the kernel that we found:

$$Y_t^{\theta,\sigma} = \int_0^t \phi(t,u)\left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}h'(u)\right]du + \int_0^t \phi(t,u)dW_u \tag{3.9}$$

Then by theorem 3.3.6 we can write $Y_t^{\theta,\sigma}$ also using the kernel and rewrite the 3.3.1.2:

$$\int_0^t \phi(u)d\hat{Y}_u^{\sigma,\theta} = \int_0^t \phi(t,u)\left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}h'(u)\right]du + \int_0^t \phi(t,u)dB_u \tag{3.10}$$

Where

$$\hat{Y}_t^{\sigma,\theta} = \int_0^t (\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}h'(u))du + B_t$$

Now we need to state the extension of the Kallianpur–Striebel formula (KS formula)[3], the extended version is the result of the Mandal and Mandrekar in their paper in 2000 which they extended the Kallianpur–Striebel formula over a more general form of filtration which are generated by kernel which generates a RKHS.[18]

Theorem 3.3.10 (Theorem 3.7 [18][45])**.** Let the observation channel be as described in 3.10:

$$\int_0^t \phi(u)d\hat{Y}_u^{\sigma,\theta} = \int_0^t \phi(t,u)\left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}h'(u)\right]du + \int_0^t \phi(t,u)dB_u$$

where $\hat{Y}_t^{\sigma,\theta} = \int_0^t (\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}h'(u))du + B_t$ then we have:

$$E\left(g(T,X) \mid \mathcal{F}_t^Y\right) = \frac{\int g(T,x)\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}}d\mu_{X(x)}}{\int \frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}}d\mu_{X(x)}}$$

where :

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left(\int_0^t \tilde{h}\left(u,x_u\right)d\hat{Y}_u - \frac{1}{2}\int_0^t \left|\tilde{h}\left(u,x_u\right)\right|^2 du\right)$$

[3]Their work [19] is a classic result in stochastic filtering, focusing on estimation over a channel under the filtration of Brownian motion

[4]The notation is revisited to be compatible with this work

[5]The proof is available in the article so I am not going to provide the proof here

Using this theorem we are able to calculate the $\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}}$ for our observation channel and then we can find the $\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}}$ with is the central result for calculation of the mutual information over the channel

Proposition 3.3.11. for the observation stochastic input-output channel described in 3.7 we have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \psi(\theta,\sigma,X_u)dB_t + \frac{1}{2}\int_0^t [\psi(\theta,\sigma,X_u)]^2\, du\right\}$$

Where $\psi(\theta,\sigma,X_u) = [\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)]$

Proof. By the theorem 3.3.10 we know that:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right] d\tilde{Y}_u^{\theta,\sigma} - \frac{1}{2}\int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right]^2 du\right\}$$

where $\tilde{Y}_t^{\theta,\sigma}$ is a process with given integral equation:

$$\tilde{Y}_t^{\theta,\sigma} = \int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right] du + B_t$$

Let's name the $\int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right] du$ as $V_t^{\theta,\sigma}$ so then we will have:

$$d\tilde{Y}_t^{\theta,\sigma} = dV_t^{\theta,\sigma} + dB_t$$

So now we can substitute $d\tilde{Y}_t^{\theta,\sigma}$ with the integrand which we can take:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right] (dV_t^{\theta,\sigma} + dB_t) \right.$$
$$\left. -\frac{1}{2}\int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right]^2 du\right\}$$

$$
\begin{aligned}
\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\Bigg\{ & \int_0^t \left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u)\right] dB_u \\
& + \int_0^t \left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u)\right] dV_u^{\theta,\sigma} \\
& - \frac{1}{2} \int_0^t \left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u)\right]^2 du \Bigg\}
\end{aligned}
$$

$$
\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{ \int_0^t \left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u)\right] dB_t + \frac{1}{2} \int_0^t \left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u)\right]^2 du \right\}
$$

□

Now we can apply proposition 2.3.4 to our result to find the RND $\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B}$ by taking the expected value with respect to the past observations so let's write this result in the next proposition.

Proposition 3.3.12. For the observation stochastic input-output channel described in 3.7 we have:

$$
\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B} = \exp\left\{ \int_0^t \hat{\psi}(\theta,\sigma,X_u) dB_t + \frac{1}{2} \int_0^t \left[\hat{\psi}(\theta,\sigma,X_u)\right]^2 du \right\}
$$

Where $\hat{\psi}(\theta,\sigma,X_u) = \mathbb{E}\left[\frac{\theta}{\sigma} h\left(x_u\right) + \frac{1}{\sigma} h'(u) | \mathcal{F}_{Y_t^{\theta,\sigma}}\right]$

Proof. Proof as we mentioned is similar to proposition 2.3.4 □

So now we find the wanted RND which is $\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}}$ which we same apply the same method which we used in the section 2.3.1, and we can find this RND which can help us in the next section to calculate the mutual information over the channel:

Proposition 3.3.13. For the observation stochastic input-output channel described in 3.7 we have:

$$
\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \Gamma(\theta,\sigma,X_u) dB_t + \frac{1}{2} \int_0^t \Gamma(\theta,\sigma,X_u)^2 du \right]
$$

Where $\Gamma(\theta, \sigma, X_u) = \psi(\theta, \sigma, X_u) - \hat{\psi}(\theta, \sigma, X_u)$,and then ψ id defined as before in the previous propositions $\hat{\psi}(\theta, \sigma, X_u) = \mathbb{E}\left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)|\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\psi(\theta, \sigma, X_u) = [\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)]$

Proof. Similarly as section 2.3, we can write by lemma 2.3.7 we can write chain rule so we have:

$$\frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_{Y_t^{\theta,\sigma}}d\mu_X} = \frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_X d\mu_B} * \frac{d\mu_B}{d\mu_{Y_t^{\theta,\sigma}}}$$

Then by lemma 2.3.8 we can write:

$$\frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_{Y_t^{\theta,\sigma}}d\mu_X} = \frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_X d\mu_B} * (\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B})^{-1}$$

Then using proposition 3.3.16, and proposition 3.3.17, and $dY_t^r = \sqrt{r}\hat{X}_t dt + dB_t$ then we have $\tilde{Y}_t^{\theta,\sigma} = \int_0^t \left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)\right]du + B_t$,and easy as was mentioned in section 2.3 we will have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \Gamma(\theta, \sigma, X_u)dB_t + \frac{1}{2}\int_0^t \Gamma(\theta, \sigma, X_u)^2 du\right]$$

□

Technically the function Γ is the causal minimum mean square error of estimation of the $\bar{h}$ function which was the equivalent function to the h in the normal representation.

3.3.2 RND of the channel with O-U process as noise

Let's revisit our problem first from equation 3.3, and equation 3.4, so our observation channel is :

$$dY_t^{\theta,\sigma} = h\left(X_u\right)du + dN_t$$

$$Y_t^{\theta,\sigma} = \int_0^t h\left(X_u\right)du + N_t$$

Now from theorem 3.1.4, and by assumption the $N_0 = 0$, we can write our channel model as it is represented below:

$$Y_t^{\theta,\sigma} = \int_0^t h\left(X_u\right)du + \int_0^t e^{\theta(s-t)}\sigma\,\mathrm{d}B_s \tag{3.11}$$

Where θ, and σ, are the respective constants in the stochastic differential equation that generates this Ornstein Uhlenbeck noise process. So as we discussed previously in section 3.3.1, we need to define the reproducing kernel Hilbert space at another time and apply the same method to find the RND for this observation channel.

3.3.2.1 RKHS over O-U process filtration

As in the previous section let's write the covariance function for this process which we have done in proposition 3.1.6 but in the integral form

$$K'(s,t) = \mathrm{Cov}(N_s, N_t) = \mathbb{E}[(\int_0^{\min(t,s)} \phi'(t,u)\phi'(t,s)(\,\mathrm{d}u)]$$

Where $\phi'(t,u) = \sigma e^{-\theta(t-u)}$, so we should consider our new kernel the function $K'(s,t)$, and now we need to define the new RKHS for our new kernel based on the definition 3.3.4. So we need to prove that this kernel has the properties we need.

Theorem 3.3.14. Let $\phi'(t,u) = \sigma e^{-\theta(t-u)}$, where $0 \le u \le t \le T$. Consequently, the closure of the space spanned by $\phi'(t,\cdot)1_{[0,t]}(\cdot)$ in $L^2[0,T]$ is $L^2[0,T]$.

Proof. Similar to proof of theorem 3.3.7, Let $f \in L^2[0,T]$ such that $f \perp \phi'(t,\cdot)1_{[0,t]}(\cdot)$ for all $\forall t \in [0,T]$). Then we can write the definition of the perpendicular function as an integral equation:

$$\int_0^t f(u)\phi'(t,u)du = 0 \quad \text{for all } t$$

Now let's substitute the function $\phi'(t,u) = \sigma e^{-\theta(t-u)}$ inside the equation:

$$\int_0^t f(u)e^{-\theta(t-u)}du = 0$$

Where σ is positive, and then we have:

$$e^{-\theta t}\int_0^t f(u)\sigma e^{\theta u}du = 0$$

An we know $e^{\theta u}$ is positive over the domain so we got:

$$f(u) = 0$$

And by this, we showed that he closure of the space spanned by $\phi'(t,\cdot)1_{[0,t]}(\cdot)$ in $L^2[0,T]$ is $L^2[0,T]$

□

So by theorem 3.3.14 we can write the reproducing kernel Hilbert space easily:

$$\text{RKHS}(K') = \left\{ f : f(s) = \int_0^s \phi'(s,u) f^*(u) du, \quad \exists f^* \in L^2[0,T] \right\} \tag{3.12}$$

So by this RKHS we only need to show that there exists a function that over the kernel gives us the signal part of the observation channel by which we can use our method to find the RND from the theorem 3.3.10 which was the extended version of the Kallianpur–Striebel formula.

Proposition 3.3.15. Over the observation channel described in 3.7, and RKHS described in 3.8, we have:

$$\int_0^t h(X_u)\, du = \int_0^t \phi'(t,u) \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du$$

Proof. We will show that the left hand side integral is equivalent to the right hand side one. First, let's substitute the function ϕ' inside the integral equation:

$$\int_0^t \phi'(t,u) \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du = \int_0^t \sigma e^{-\theta(t-u)} \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du$$

Then we can separate the integrals:

$$= \int_0^t e^{-\theta(t-u)} h(X_u)\, du + \theta \int_0^t e^{-\theta(t-u)} \int_0^t h(X_s) ds du$$

We leave the first integral unchanged, and for the second integral we will use Fubini's theorem, and we change the two integrals' order because both of them are full regions the region of the integral will not change over changing orders:

$$= \int_0^t e^{-\theta(t-u)} h(X_u)\, du + \theta \int_0^t h(X_s) \left[\int_0^t e^{-\theta(t-u)} du \right] ds$$

The internal integral will be easily calculated $\int_0^s e^{-\theta(t-u)} du = \frac{1-e^{-\theta(t-s)}}{\theta}$ so we can substitute:

$$= \int_0^t e^{-\theta(t-u)} h(X_u)\, du + \theta \int_0^t h(X_s) \frac{1-e^{-\theta(t-s)}}{\theta} ds$$

We can separate the second integral:

$$= \int_0^t e^{-\theta(t-u)} h(X_u)\, du + \int_0^t h(X_s) ds - \int_0^t e^{-\theta(t-s)} h(X_s) ds = \int_0^t h(X_s) ds$$

□

Then by proposition 3.3.15 we can rewrite the observation channel equation 3.11 and write the kernel representation:

$$Y_t^{\theta,\sigma} = \int_0^t \phi'(t,u) \left[\frac{h\left(X_u\right)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du + \int_0^t \phi'(t,u) \, \mathrm{d}B_u$$

Then from the theorem 3.3.6 we can translate the observation process to make our observation channel ready to use the Kallianpur–Striebel formula extended formula to calculate the RND.

$$\int_0^t \phi'^*(u) d\tilde{Y}_u^{\sigma,\theta} = \int_0^t \phi'(t,u) \left[\frac{h\left(X_u\right)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du + \int_0^t \phi'(t,u) \, \mathrm{d}B_u \tag{3.13}$$

Where

$$\tilde{Y}_t^{\sigma,\theta} = \int_0^t \left[\frac{h\left(X_u\right)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right] du + B_t$$

Proposition 3.3.16. For the observation stochastic input-output channel described in 3.13 we have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{ \int_0^t \tilde{\psi}(\theta,\sigma,X_u) dB_t + \frac{1}{2} \int_0^t \tilde{\psi}(\theta,\sigma,X_u)^2 du \right\}$$

Where $\tilde{\psi}(\theta,\sigma,X_u) = \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma} \int_0^t h(X_s) ds \right]$

Proof. By the theorem 3.3.10 we know that:

$$\begin{aligned} \frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\Bigg\{ & \int_0^t \left[\frac{\theta}{\sigma} h(x_u) + \frac{1}{\sigma} \int_0^u h(X_s) \, ds \right] d\tilde{Y}_u^{\theta,\sigma} \\ & - \frac{1}{2} \int_0^t \left[\frac{\theta}{\sigma} h(x_u) + \frac{1}{\sigma} \int_0^u h(X_s) \, ds \right]^2 du \Bigg\} \end{aligned}$$

where $\tilde{Y}_t^{\theta,\sigma}$ is a process with the given integral equation:

$$\tilde{Y}_t^{\sigma,\theta} = \int_0^t \left[\frac{h\left(X_u\right)}{\sigma} + \frac{\theta}{\sigma} \int_0^u h(X_s) ds \right] du + B_t$$

Let's denote $\int_0^t \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma} \int_0^u h(X_s) ds \right] du$ as $V_t^{\theta,\sigma}$, so we have:

$$d\tilde{Y}_t^{\theta,\sigma} = dV_t^{\theta,\sigma} + dB_t$$

Now we can substitute $d\tilde{Y}_t^{\theta,\sigma}$ in the expression:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right]\left(dV_u^{\theta,\sigma} + dB_u\right) - \frac{1}{2}\int_0^t \left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right]^2 du\right\}$$

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \left[\frac{\theta}{\sigma}h(X_u) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right]^2 du + \int_0^t \left[\frac{\theta}{\sigma}h(X_u) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right] dB_u - \frac{1}{2}\int_0^t \left[\frac{\theta}{\sigma}h(X_u) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right]^2 du\right\}$$

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right] dB_t + \frac{1}{2}\int_0^t \left[\frac{\theta}{\sigma}h\left(X_u\right) + \frac{1}{\sigma}\int_0^u h(X_s)\,ds\right]^2 du\right\}$$

Define $\tilde{\psi}(\theta, \sigma, X_u) = \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^u h(X_s)ds\right]$, then:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_B \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left\{\int_0^t \tilde{\psi}(\theta, \sigma, X_u)dB_t + \frac{1}{2}\int_0^t \tilde{\psi}(\theta, \sigma, X_u)^2 du\right\}$$

□

Next, we can use Proposition 2.3.4 to determine the Radon-Nikodym derivative $\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B}$ by calculating the expected value based on the past observations. We will state this result formally in the following proposition.

Proposition 3.3.17. For the observation stochastic input-output channel described in 3.13, we have:

$$\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B} = \exp\left\{\int_0^t \hat{\tilde{\psi}}(\theta, \sigma, X_u)dB_t + \frac{1}{2}\int_0^t \left[\hat{\tilde{\psi}}(\theta, \sigma, X_u)\right]^2 du\right\}$$

Where $\hat{\tilde{\psi}}(\theta, \sigma, X_u) = \mathbb{E}\left[\left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right] |\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$

Proof. Proof as we mentioned is similar to proposition 2.3.4 □

So now we find the wanted RND which is $\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}}$ which we same apply the same method which we used in the section 2.3.1, and we can find this RND which can help us in the next section to calculate the mutual information over the channel:

Proposition 3.3.18. For the observation stochastic input-output channel described in 3.7 we have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \tilde{\Gamma}(\theta, \sigma, X_u)dB_t + \frac{1}{2}\int_0^t \tilde{\Gamma}(\theta, \sigma, X_u)^2 du\right]$$

Where $\tilde{\Gamma}(\theta, \sigma, X_u) = \tilde{\psi}(\theta, \sigma, X_u) - \hat{\tilde{\psi}}(\theta, \sigma, X_u)$,and then ψ id defined as before in the previous propositions $\hat{\tilde{\psi}}(\theta, \sigma, X_u) = \mathbb{E}\left[\left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right] |\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\tilde{\psi}(\theta, \sigma, X_u) = \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right]$

Proof. Similarly as section 2.3, we can write by lemma 2.3.7 we can write chain rule so we have:

$$\frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_{Y_t^{\theta,\sigma}}d\mu_X} = \frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_X d\mu_B} * \frac{d\mu_B}{d\mu_{Y_t^{\theta,\sigma}}}$$

Then by lemma 2.3.8 we can write:

$$\frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_{Y_t^{\theta,\sigma}}d\mu_X} = \frac{d\mu_{Y_t^{\theta,\sigma}X}}{d\mu_X d\mu_B} * \left(\frac{d\mu_{Y_t^{\theta,\sigma}}}{d\mu_B}\right)^{-1}$$

Then using proposition 3.3.16, and proposition 3.3.17, and $\tilde{Y}_t^{\sigma,\theta} = \int_0^t \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^u h(X_s)ds\right] du + B_t$,and easy as was mentioned in section 2.3 we will have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \tilde{\Gamma}(\theta, \sigma, X_u)dB_t + \frac{1}{2}\int_0^t \tilde{\Gamma}(\theta, \sigma, X_u)^2 du\right]$$

□

As we can see, it is the same as section 3.3.1.2, and we are left with a similar function, which is like casual minimum mean square error. The only difference between these two cases is changing the kernel which results in changing the $\bar{h}$ function which is pretty interesting. We could write it for a general form of noise such as described in 3.3.1 but these special cases are more important because of the natural behavior of noise in any kind of observation channel

3.4 Calculation of mutual information over the channel

In this section, we are going to use the definition of the mutual information over the observation channel from the definition 2.1.4 we can find the mutual information over the channel.

3.4.1 Mutual information over observation with int O-U process noise

Theorem 3.4.1. The mutual information over the observation channel described in equation 3.10 is:

$$I(X_t, Y_t^{\theta,\sigma}) = \frac{1}{2}\int_0^t \mathbb{E}\left[\Gamma(\theta,\sigma,X_u)^2\right] du$$

Where $\Gamma(\theta,\sigma,X_u) = \psi(\theta,\sigma,X_u) - \hat{\psi}(\theta,\sigma,X_u)$,and then ψ id defined as before in the previous propositions $\hat{\psi}(\theta,\sigma,X_u) = \mathbb{E}\left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)|\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\psi(\theta,\sigma,X_u) = [\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)]$

Proof. From the Proposition 3.3.13 we have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \Gamma(\theta,\sigma,X_u)dB_t + \frac{1}{2}\int_0^t \Gamma(\theta,\sigma,X_u)^2 du\right]$$

Where $\Gamma(\theta,\sigma,X_u) = \psi(\theta,\sigma,X_u) - \hat{\psi}(\theta,\sigma,X_u)$,and then ψ id defined as before in the previous propositions $\hat{\psi}(\theta,\sigma,X_u) = \mathbb{E}\left[\frac{\theta}{\sigma}h\left(x_u\right) + \frac{1}{\sigma}h'(u)|\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\psi(\theta,\sigma,X_u) = [\frac{\theta}{\sigma}h\left(x_u\right) +$

$\frac{1}{\sigma}h'(u)]$ From the definition 2.1.4 we can write:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\log\left(\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}}\right)\right]$$

Then by calculation of the logarithm we reach:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\int_0^t \Gamma(\theta,\sigma,X_u)dB_t + \frac{1}{2}\int_0^t \Gamma(\theta,\sigma,X_u)^2 du\right]$$

Then we can separate the integrals and by lemma 2.3.10 we have:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\frac{1}{2}\int_0^t \Gamma(\theta,\sigma,X_u)^2 du\right]$$

By the monotone convergence theorem, we have:

$$I(X_t, Y_t^{\theta,\sigma}) = \frac{1}{2}\int_0^t \mathbb{E}\left[\Gamma(\theta,\sigma,X_u)^2\right] du$$

□

3.4.2 Mutual information over the channel with O-U process noise

Theorem 3.4.2. The mutual information over the observation channel described in equation 3.13 is:

$$I(X_t, Y_t^{\theta,\sigma}) = \frac{1}{2}\int_0^t \mathbb{E}\left[\tilde{\Gamma}(\theta,\sigma,X_u)^2\right] du$$

Where $\tilde{\Gamma}(\theta,\sigma,X_u) = \tilde{\psi}(\theta,\sigma,X_u) - \hat{\tilde{\psi}}(\theta,\sigma,X_u)$,and then ψ id defined as before in the previous propositions $\hat{\tilde{\psi}}(\theta,\sigma,X_u) = \mathbb{E}\left[\left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right] |\mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\tilde{\psi}(\theta,\sigma,X_u) = \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right]$

Proof. From the Proposition 3.3.18 we have:

$$\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}} = \exp\left[\int_0^t \tilde{\Gamma}(\theta,\sigma,X_u)dB_t + \frac{1}{2}\int_0^t \tilde{\Gamma}(\theta,\sigma,X_u)^2 du\right]$$

Where $\tilde{\Gamma}(\theta,\sigma,X_u) = \tilde{\psi}(\theta,\sigma,X_u) - \hat{\tilde{\psi}}(\theta,\sigma,X_u)$,and then ψ id defined as before in the previous propositions $\hat{\tilde{\psi}}(\theta,\sigma,X_u) = \mathbb{E}\left[\left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right] | \mathcal{F}_{Y_t^{\theta,\sigma}}\right]$,and $\tilde{\psi}(\theta,\sigma,X_u) = \left[\frac{h(X_u)}{\sigma} + \frac{\theta}{\sigma}\int_0^t h(X_s)ds\right]$ From the definition 2.1.4 we can write:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\log\left(\frac{d\mu_{X,Y_t^{\theta,\sigma}}}{d\mu_X \times d\mu_{Y_t^{\theta,\sigma}}}\right)\right]$$

Then by calculation of the logarithm we reach:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\int_0^t \tilde{\Gamma}(\theta,\sigma,X_u)dB_t + \frac{1}{2}\int_0^t \tilde{\Gamma}(\theta,\sigma,X_u)^2 du\right]$$

Then we can separate the integrals and by lemma 2.3.10 we have:

$$I(X_t, Y_t^{\theta,\sigma}) = \mathbb{E}\left[\frac{1}{2}\int_0^t \tilde{\Gamma}(\theta,\sigma,X_u)^2 du\right]$$

By the monotone convergence theorem, we have:

$$I(X_t, Y_t^{\theta,\sigma}) = \frac{1}{2}\int_0^t \mathbb{E}\left[\tilde{\Gamma}(\theta,\sigma,X_u)^2\right] du$$

□

3.5 Analysis of the results

The primary findings of this thesis are the outcomes of theorems 3.4.1 and 3.4.2. This result holds regardless of the input signal's distribution; the only conditions needed are that the function $h(X_u)$ be differentiable and that it be integrable for the integrated O-U process and the O-U process respectively. Both theories demonstrate a relationship between the mutual information and the signal's estimate error, which is equivalent to Duncan's theorem in an environment of Gauss-Markov noise in a channel.

The $\psi(\theta,\sigma,X_u)$ function's estimation error with respect to the causal observation (shown with Γ) differs from the signal's estimation error, still in certain circumstances it can resemble the causal MMSE. Let's take a look at one example.

Observation 3.5.1. Let $h(X_u) = X_u$, and the integrated O-U process be the noise in the channel. which means our observation stochastic input-output channel will be :

$$dY_t^{\theta,\sigma} = X_t dt + dN_t$$

Where dN_t is an integrated O-U process then from theorem 3.4.1 we will have:

$$I(X, Y^{\theta,\sigma}) = \left(\frac{\theta}{\sqrt{2}\sigma}\right)^2 \int_0^T \mathbb{E}\left[X_u - \mathbb{E}[X_u|\mathcal{F}_{Y_u^{\sigma,\theta}}]\right]^2 du$$

This result means that in the case in which the signal is added to the integrated O-U noise, the result is similar to Duncan's theorem if we choose $\theta = \sigma$, and $h(X_u) = \sqrt{r}X_u$ where r is signal to noise ration we have:

$$dY_t^{\theta,\sigma} = \sqrt{r}X_t dt + dN_t$$

$$I(X, Y^{\theta,\sigma}) = \frac{r}{2}\int_0^T \mathbb{E}\left[X_u - \mathbb{E}[X_u|\mathcal{F}_{Y_u^{\sigma,\theta}}]\right]^2 du$$

As we know when $(\frac{\theta}{\sigma})^2 \to 1$ where both $\sigma, \theta \to \infty$ goes to infinity then from proposition 3.1.8 we have:

$$N_t = \int_0^t \frac{\sigma}{\theta}\left(1 - e^{-\theta(t-s)}\right) dB_s$$

Then by $(\frac{\theta}{\sigma})^2 \to 1$ where both $\sigma, \theta \to \infty$ we have[6]:

$$N_t = \int_0^t dB_s = B_t$$

As demonstrated above, when the two variables θ and σ approach infinity, our observation channel converges to Duncan's formulation. So According to Duncan's theorem, there is the same relationship between causal MMSE and mutual information. However, this new theory shows that even when the two variables θ and σ are equal and finite, the same theorem holds for the stochastic input-output channel.

In general, $\tilde{\Gamma}(\theta, \sigma, X_u)$ and $\Gamma(\theta, \sigma, X_u)$ are analogous to the non-causal MMSE because they represent the estimation error of a function of the main signal with respect to the σ-algebra generated by the observation in the respective channels. Consequently, we demonstrated that the relationship between mutual information and non-causal MMSE persists

[6]The convergence of O-U process to Brownian motion has been proved by Nelson in [15]

in stochastic input-output channels with either an Ornstein-Uhlenbeck process as noise or an integrated Ornstein-Uhlenbeck process as noise. The only conditions for this result are the independence between the input signal and noise, as well as the integrability and differentiability of the signal function in the channel model. In other words, the connection between the estimation problem and mutual information, as demonstrated by Duncan, is extended within the framework of Gauss-Markov noise.

Chapter 4

Conclusion

We conclude by summarizing the contributions of this thesis and discussing potential extensions and probelms that could be addressed in future work.

4.1 Summary

In this thesis, we commence with Chapter 2, which revisits Duncan's theorem and the relationship between Non-Causal Minimum Mean Square Error and Causal Minimum Mean Square Error in the AWGN channel, as established by Verdú and Guo. In Section 2.1, we define the measure space encompassing the stochastic processes and emphasize the significance of absolute continuity as a prerequisite for the existence of the RND of the respective measure. Additionally, we delineate mutual information for a general stochastic input-output channel.

Section 2.2 addresses the general form of the observation channel and introduces notable special cases frequently encountered in communication channel modeling. In Section 2.3, we present a proof of Duncan's theorem using the Girsanov theorem rather than Duncan's original method. This modification is motivated by the widespread contemporary use of the Girsanov theorem in stochastic calculus, rendering the proof more congruent with current literature. Section 2.4 demonstrates Verdú et al.'s result using our method, elucidating the relationship between causal and noncausal MMSE over a channel with Gaussian noise.

Chapter 3 begins with the definition of the O-U process and the integrated O-U process through their respective stochastic differential equations. Using these definitions, we derive the integrated representation of the process, facilitating the representation of a closed-form channel model. In Section 3.2, we formulate our problem for two cases: a stochastic input-output channel with O-U process noise and with integrated O-U process noise, both in SDE representation and stochastic integrated representation. Section 3.3 introduces the concept of the Reproducing Kernel Hilbert Space. By establishing RKHS for our channel stochastic model in both cases and applying an extended version of the KS formula, we derive the RND for the product space of observation and signal with respect to their individual measures.

Section 2.4 presents the primary theorem of our contribution, establishing the relationship between mutual information and the causal MMSE error of the observation channel. Finally, Section 3.5 discusses the parallels between this result and Duncan's theorem.

4.2 Future work and related problems

This result can be extended to encompass a general form of kernel noise generation and can also be adapted to observation channels with observation feedback. Furthermore, future research could explore the relationship between NCMMSE and CMMSE within the stochastic input-output channel with the O-U process as noise. Given the growing interest in employing Gaussian Markov processes as noise in quantum communication channels, these findings could also be extended to the realm of quantum communication.

References

[1] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[2] C.E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[3] I. M. Gel\cprime fand and A. M. Yaglom. Calculation of the amount of information about a random function contained in another such function. *Amer. Math. Soc. Transl. (2)*, 12:199–246, 1959.

[4] Tse-Pei Chiang. Remark on the Definition of the Quantity of Information. 1959. Google-Books-ID: V9m1YgEACAAJ.

[5] Tyrone E. Duncan. On the Calculation of Mutual Information. *SIAM Journal on Applied Mathematics*, 19(1):215–220, 1970. Publisher: Society for Industrial and Applied Mathematics.

[6] T. Kadota, M. Zakai, and J. Ziv. Mutual information of the white Gaussian channel with and without feedback. *IEEE Transactions on Information Theory*, 17(4):368–371, July 1971.

[7] T. Kadota, M. Zakai, and J. Ziv. Capacity of a continuous memoryless channel with feedback. *IEEE Transactions on Information Theory*, 17(4):372–378, July 1971.

[8] N. Chayat and S. Shamai. Bounds on the information rate of intertransition-time-restricted binary signaling over an AWGN channel. *IEEE Transactions on Information Theory*, 45(6):1992–2006, September 1999.

[9] I. Bar-David and S. Shamai. On information transfer by envelope-constrained signals over the AWGN channel. *IEEE Transactions on Information Theory*, 34(3):371–379, May 1988.

[10] Young-Han Kim, Haim H. Permuter, and Tsachy Weissman. Directed information, causal estimation, and communication in continuous time. In *2009 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 1–7, June 2009.

[11] Dongning Guo, S. Shamai, and S. Verdu. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, April 2005. Conference Name: IEEE Transactions on Information Theory.

[12] Dongning Guo, Shlomo Shamai, and Sergio Verdu. Mutual Information and Conditional Mean Estimation in Poisson Channels. *IEEE Transactions on Information Theory*, 54(5):1837–1849, May 2008.

[13] Francisco J. Piera and Patricio Parada. On the Relationship between Mutual Information and Minimum Mean-Square Errors in Stochastic Dynamical Systems, October 2007. arXiv:0710.1280 [cs, math].

[14] A.V. Balakrishnan. Nonlinear white noise theory. *Journal of Multivariate Analysis V,P. R. Krishnaiah*, Vol.5:97–106, 1980.

[15] Edward Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, 1967.

[16] Joachim Schäfer, David Daems, Evgueni Karpov, and Nicolas J. Cerf. Capacity of a bosonic memory channel with Gauss-Markov noise. *Physical Review A*, 80(6):062313, December 2009.

[17] Nicolas J. Cerf, Julien Clavareau, Chiara Macchiavello, and Jérémie Roland. Quantum entanglement enhances the capacity of bosonic channels with memory. *Physical Review A*, 72(4):042330, October 2005.

[18] Pranab K. Mandal and V. Mandrekar. A Bayes Formula for Gaussian Noise Processes and its Applications. *SIAM Journal on Control and Optimization*, 39(3):852–871, January 2000.

[19] G. Kallianpur and C. Striebel. Estimation of Stochastic Systems: Arbitrary System Process with Additive White Noise Observation Errors. *The Annals of Mathematical Statistics*, 39(3):785–801, June 1968.

[20] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.

[21] Sidney Resnick. *A Probability Path*. Birkhäuser Boston, Boston, MA, 1st ed. 2005. edition, 2005.

[22] A. Kolmogorov. On the Shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory*, 2(4):102–108, December 1956.

[23] J. F. Le Gall. *Brownian motion, martingales, and stochastic calculus*. Graduate texts in mathematics, 274. Springer, Switzerland, 2016.

[24] Paul R. Halmos. *Measure Theory*, volume 18 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1950.

[25] Yongpeng Wu, Chengshan Xiao, Zhi Ding, Xiqi Gao, and Shi Jin. A Survey on MIMO Transmission With Finite Input Signals: Technical Challenges, Advances, and Future Trends. *Proceedings of the IEEE*, 106(10):1779–1833, October 2018. Conference Name: Proceedings of the IEEE.

[26] Keigo Takeuchi. Orthogonal Approximate Message-Passing for Spatially Coupled Linear Models, May 2023. arXiv:2210.06019 [cs, math].

[27] Ioannis Karatzas. *Brownian motion and stochastic calculus*. Number 113 in Graduate Texts in Mathematics. Springer, New York, second edition. edition, 1998.

[28] Cyril Measson, Andrea Montanari, Thomas J. Richardson, and RÜdiger Urbanke. The Generalized Area Theorem and Some of its Consequences. *IEEE Transactions on Information Theory*, 55(11):4793–4821, November 2009.

[29] M. T. Barlow. A Diffusion Model for Electricity Prices. *Mathematical Finance*, 12(4):287–298, 2002.

[30] John G. Proakis. *Digital Communications, 5th edition*. McGraw-Hill Higher Education, 2008. Google-Books-ID: tOpKAQAACAAJ.

[31] Eberhard Zeidler. *Applied Functional Analysis*, volume 109 of *Applied Mathematical Sciences*. Springer, New York, NY, 1995.

[32] Emanuel Parzen. Statistical inference on time series by Hilbert space methods, I. 1959.

[33] Murray H. Protter and Charles B. Morrey. Differentiation under the Integral Sign. Improper Integrals. The Gamma Function. In Murray H. Protter and Charles B. Morrey, editors, *Intermediate Calculus*, pages 421–453. Springer, New York, NY, 1985.

[34] Tyrone E. Duncan. Mutual Information for Stochastic Signals and LÉvy Processes. *IEEE Transactions on Information Theory*, 56(1):18–24, January 2010.

[35] D.P. Palomar and S. Verdu. Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory*, 52(1):141–154, January 2006.

[36] T. M. Cover. *Elements of information theory.* Wiley-Interscience, Hoboken, N.J, 2nd ed. edition, 2006.

[37] N. Chayat and S. Shamai. Bounds on the capacity of a binary input AWGN channel with intertransition duration restrictions. In *17th Convention of Electrical and Electronics Engineers in Israel*, pages 227–229, March 1991.

[38] Y. Ephraim and N. Merhav. Lower and Upper Bounds on the Minimum Mean Square Error in Composite Source Estimation. In *Proceedings. 1991 IEEE International Symposium on Information Theory*, pages 163–163, June 1991.

[39] G. Huang, H. M. Jansen, M. Mandjes, P. Spreij, and K. De Turck. Markov-Modulated Ornstein-Uhlenbeck Processes. *Advances in Applied Probability*, 48(1):235–254, 2016.

[40] Y. Sung, L. Tong, and H.V. Poor. Neyman-pearson detection of gauss-Markov signals in noise: closed-form error exponentand properties. *IEEE Transactions on Information Theory*, 52(4):1354–1365, April 2006.

[41] Carlos Armando Mejía Vega. Calibration of the exponential Ornstein–Uhlenbeck process when spot prices are visible through the maximum log-likelihood method. Example with gold prices. *Advances in Difference Equations*, 2018(1):269, August 2018.

[42] N. Chayat and S. Shamai. Bounds on the information rate of intertransition-time-restricted binary signaling over an AWGN channel. *IEEE Transactions on Information Theory*, 45(6):1992–2006, September 1999. Conference Name: IEEE Transactions on Information Theory.

[43] Hailong Wang and Duni Hu. Heterogeneous beliefs with information processing capacity constraints and asset pricing in a monetary economy. *The North American Journal of Economics and Finance*, 72:102143, May 2024.

[44] S. Varadhan. Stochastic Processes, October 2007. ISBN: 9780821840856 9780821883556 9781470431167 ISSN: 1529-9031, 2472-4467 Publisher: American Mathematical Society Series: Courant Lecture Notes Volume: 16.