FEDERAL UNIVERSITY OF CEARÁ
CENTER OF TECHNOLOGY
DEPARTMENT OF TELEINFORMATICS ENGINEERING
POST-GRADUATION PROGRAM IN TELEINFORMATICS ENGINEERING
MASTER DEGREE IN TELEINFORMATICS ENGINEERING

FILIPE PEREIRA DE FARIAS

ON APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS

FORTALEZA
2025

FILIPE PEREIRA DE FARIAS

ON APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS

Dissertation submitted to the Post-Graduation Program in Teleinformatics Engineering of the Center of Technology of the Federal University of Ceará, as a partial requirement for obtaining the title of Master in Teleinformatics Engineering. Concentration Area: Signals and Systems with focus on Pattern Recognition and Dynamical Systems.

Advisor: Prof. Dr. Michela Mulas

FORTALEZA
2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

P491o Pereira de Farias, Filipe.
On Applications of Stochastic Differential Equations / Filipe Pereira de Farias. – 2025.
114 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2025.
Orientação: Prof. Dr. Michela Mulas.

1. equações diferenciais estocásticas. 2. termodinâmica estocástica. 3. computação numérica probabilística. 4. processos Gaussianos. 5. sistemas dinâmicos. I. Título.
CDD 621.38

NOTE TO THE READER

This thesis has animated content which is better visualized online. Thus, it is made available at the author's website https://filipepfarias.github.io/UFC_MSc_Thesis_presentation the presentation of the Thesis defense containing the animated version of the Figures 2, 3 and 6, as other visualizations.

FILIPE PEREIRA DE FARIAS

ON APPLICATIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS

Dissertation submitted to the Post-Graduation Program in Teleinformatics Engineering of the Center of Technology of the Federal University of Ceará, as a partial requirement for obtaining the title of Master in Teleinformatics Engineering. Concentration Area: Signals and Systems with focus on Pattern Recognition and Dynamical Systems.

Approved on 23rd July 2025.

EXAMINATION BOARD

Prof. Dr. Michela Mulas (Advisor)
Federal University of Ceará (UFC)

Prof. Dr. Charles Casimiro Cavalcante
Federal University of Ceará (UFC)

Prof. Dr. Leandro Chaves Rêgo
Federal University of Ceará (UFC)

For those who work hard to spread the knowledge.

ACKNOWLEDGEMENTS

To my parents, Fátima and Edson, who have always fostered me the vision of education as an essential element of human dignity, and who have always fought not only for mine, but also for society's. Especially to my mother, whose passion for teaching moves me and serves as an inspiration.

To Professors Michela and Francesco, who guided me and had a patience I can not begin to measure. I also thank them for years of friendship and for showing me that teaching goes beyond the walls of academia.

To the professors of the Department of Teleinformatics Engineering, especially Professor Charles whose classes were a cornerstone of my education and sparked my interest in these topics. I am also grateful for your attention and availability.

To my partner Daniel, for all the support, for standing witness to both the struggles and the victories, and for being present.

To the friends I made at the University and in the Department. Especially to our research group, for always making work light and enjoyable. Our discussions will be something I carry as a life lesson. Also to Mira, whose laughter echoes through the halls and will always be a reminder of home, even at work.

To Aalto University, for hosting me during my academic exchange and for the resources provided. Also, to the colleagues who welcomed me there and supported me throughout.

This work was carried out with support from the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Código de Financiamento 001, as well as from the *Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP).*

“Perhaps it will be wise to approach the subject cautiously.”

(Goodstein)

ABSTRACT

The study of dynamic systems represents a large field of mathematics that models the temporal evolution of phenomena in real life. The theory of differential equations largely deals with the modeling of these systems through the laws that govern the temporal rates at which the states describing dynamic systems change. In this context, we study those models in which a random factor is considered in the law that governs the rate of change of these systems over time. These are called stochastic differential equations. This work briefly presents the mathematical theory underlying this type of differential equation by constructing an analogy between the theory of ordinary differential equations and that of stochastic ones. Finally, and as the main objective, two applications of the theory are presented in problems at the state of the art: stochastic thermodynamics and probabilistic numerical methods. The first deals with an attempt to extend the classical theory of thermodynamics of equilibrium systems to systems out of equilibrium, such as chemical reactions presented in this work, for example. In the second problem, stochastic differential equations are used to probabilistically model numerical algorithms for solving ordinary differential equations, which can even account for uncertainties that the modeler might have regarding the system to be studied. These theories are reviewed and applied to systems already known in the literature. In the case of stochastic thermodynamics, we apply the theory to the study of enzymatic reactions governed by the Michaelis-Menten mechanism. For probabilistic numerical methods, we apply the theory to try to identify the parameters of a bioreactor model, which is part of a model of pilot wastewater treatment plant used as a test bench for water treatment systems.

Keywords: stochastic differential equations. stochastic thermodynamics. probabilistic numerics. Gaussian processes. dynamical systems.

RESUMO

O estudo de sistemas dinâmicos representa um grande campo da matemática que modela a evolução temporal de fenômenos na vida real. A teoria de equações diferenciais lida em grande parte com a modelagem desses sistemas pelas leis que regem as taxas temporais nas quais os estados que descrevem os sistemas dinâmicos variam. Nesse âmbito, estudamos aquelas modelagens na qual um fator aleatório é considerado na lei que que rege a taxa de variação temporal desses sistemas. A estas damos os nomes de equações diferenciais estocásticas. Nesse trabalho é apresentado superficialmente a teoria matemática que fundamenta esse tipo de equação diferencial, através da construção de uma analogia entre a teoria de equações diferenciais ordinárias e daquelas estocásticas. Por fim, e como maior objetivo, são apresentadas duas aplicações da teoria em problemas que estão no estado da arte: termodinâmica estocástica e cálculo numérico probabilístico. O primeiro trata de uma tentativa de extensão da teoria clássica de termodinâmica de sistemas em equilíbrio para sistemas fora do equilíbrio, no caso deste trabalho por exemplo, em reações químicas. No segundo problema, as equações diferenciais estocásticas são utilizadas para modelar probabilisticamente algoritmos numéricos de solução de equações diferenciais ordinárias, podendo até compreender incertezas que o modelador tenha a respeito do sistema a ser estudado. Estas teorias são revisadas e aplicadas em sistemas já conhecidos na literatura. No caso da termodinâmica estocástica, aplicamos a teoria ao estudo de reações enzimáticas regidas pelo mecanismo de Michaelis-Menten. Já para o cálculo numérico probabilistico, aplicamos a teoria para tentar identificar os parâmetros do modelo de um bioreator, este integrante de um modelo planta piloto de tratamento de água residual utilizado como de referência para sistemas de tratamento de água.

Palavras-chave: equações diferenciais estocásticas. termodinâmica estocástica. computação numérica probabilística. processos Gaussianos. sistemas dinâmicos.

LIST OF FIGURES

Figure 1 – Damped oscillator system and the graph of its displacement 23

Figure 2 – Brownian motion observed under the microscope 33

Figure 3 – Simulated collision of Brownian particle with smaller "gas" particles 34

Figure 4 – Sample path and the temporal probability density of the simulated Wiener process 43

Figure 5 – Representation of the volume collision by Gillespie (1977) 51

Figure 6 – Simulation of the Michaelis-Menten reaction mechanism 55

Figure 7 – Examples of covariance functions 66

Figure 8 – Pictorial representation of the enzymatic reaction of the Michaelis-Menten . 72

Figure 9 – Single trajectory of the Markov jump process for the Michaelis-Menten . . . 73

Figure 10 – Time evolution of the probability of the states of the Michaelis-Menten system 73

Figure 11 – Average of the stochastic energetic quatities 73

Figure 12 – Bioreactor schematic. 75

Figure 13 – Matérn covariance function $k(t,t')$ vs. the time difference $|t-t'|$ 105

LIST OF TABLES

Table 1 – Parameter absolute error for the LS objective function for the bioreactor parameters inference 78
Table 2 – Parameter absolute error for the marginal likelihood method 78
Table 3 – ASM1 - State variables 109
Table 4 – ASM1 - Parameters 110

LIST OF SYMBOLS

$\mathbf{x}$	Vector represented in boldface
$\mathbf{M}$	Matrix represented in boldface and uppercase
$[\mathbf{x}]_i$	i-th component of $\mathbf{x}$
$[\mathbf{M}]_{ij}$	i-th component of of the j-th column of $\mathbf{M}$
$\mathbf{A}^{-1}$	Inverse of the matrix $\mathbf{A}$
$\mathbf{A}^{\top}$	Transpose of the matrix $\mathbf{A}$
$\dot{x}$	First time derivative of x
$\ddot{x}$	Second time derivative of x
$\mathbb{R}$	Real numbers
$\mathbb{N}$	Natural numbers
$\mathcal{N}(\cdot \mid \mu, \Sigma)$	Gaussian distribution with mean μ and covariance Σ
P or P_X	Probability mass/distribution function
$X \sim \cdot$	X has probability distribution $\cdot$
p or p_X	Probability density function
$\mathbb{E}[X]$	Mean operator
$\mathrm{Cov}[X,Y]$	Covariance between X and Y
$\mathfrak{F}$	Fourier transform
$\mathsf{X}(i\omega)$	Fourier transform of X, with $i\omega \in \mathbb{C}$
R_X	Autocorrelarion function of X
$S_{X_t}(f)$	Spectral density of the process X_t
$X, Y, Z, \ldots$	Random variable
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$	Set
$\mathscr{A}, \mathscr{B}, \mathscr{C}, \ldots$	σ-algebra
$\mathscr{B}(\mathbb{R})$	Borel σ-algebra
$X_t, Y_t, W_t, \ldots$	Stochastic process
supp	Support of a function

$\mathcal{R}(f,\cdot)$	Riemann sum of f in the partitioning $\cdot$
$\sum_X$	Summation over the possible values of $X \in \Omega$
$a := b$	a is defined as b
$a =: b$	b is defined as a
$\lim_{x\searrow a}$	limit taken with $x > a$
$\operatorname{st-lim}$	limit taken in a stochastic sense
$\operatorname{as-lim}$	limit taken in an almost surely sense
$\operatorname{ms-lim}$	limit taken in a mean square sense

CONTENTS

1 INTRODUCTION . . . 18
1.1 Objectives . . . 21
1.2 Contributions . . . 21
1.3 Work organization . . . 22
2 OVERVIEW ON DIFFERENTIAL EQUATIONS . . . 23
2.1 Introduction to ODEs . . . 23
2.2 Integration of ODEs . . . 24
2.3 Existence and uniqueness of ODEs solutions . . . 26
2.4 State-space representation of ODEs . . . 27
2.4.0.1 Linear dynamical systems . . . 28
2.4.0.2 Linearization . . . 29
2.4.0.3 Solution of linear systems . . . 29
2.5 Numerical methods to solve ODEs . . . 30
3 STOCHASTIC DIFFERENTIAL EQUATIONS . . . 32
3.1 Motivation: The Brownian motion . . . 32
3.2 Stochastic integrals . . . 35
3.2.0.1 Itō's lemma . . . 38
3.2.0.2 Stratonovich integral . . . 39
3.2.1 Existence and uniqueness of solutions of SDEs . . . 40
3.2.2 Statistics of SDEs . . . 40
3.3 Numerical stochastic integration . . . 42
4 STOCHASTIC THERMODYNAMICS . . . 44
4.1 Review on classical thermodynamics . . . 44
4.2 Stochastic description of chemical reactions . . . 47
4.2.1 Macroscopic description . . . 48
4.2.2 Stochastic description . . . 50
4.2.2.1 Chemical master equation . . . 52
4.2.2.2 Stochastic simulation algorithm . . . 53
4.2.2.3 Chemical Langevin equation . . . 54
4.3 Energetics for Markov processes . . . 56
4.3.0.1 Stochastic entropy . . . 59

4.3.0.2 Thermodynamic consistency . . . 60

5 GAUSSIAN PROCESS STATE-SPACE MODEL . . . 61

5.1 Historical overview in signal processing . . . 61

5.2 Review on Bayesian inference . . . 61

5.3 Gaussian process regression . . . 62

5.3.0.1 Filtering and smoothing . . . 64

5.3.1 Gaussian process kernels . . . 65

5.3.1.1 Converting a covariance function to a dynamic matrix . . . 66

5.4 Gaussian ODE filters . . . 68

5.4.0.1 Estimating ODE solutions using Gaussian ODE filters . . . 69

6 THERMODYNAMICS OF THE MICHAELIS-MENTEN . . . 71

6.1 Michaelis-Menten theory . . . 71

6.2 Contributions and discussion . . . 72

7 INFERENCE OF THE PARAMETERS OF A BIOREACTOR . . . 75

7.1 Model description . . . 75

7.2 Inference via marginal likelihood . . . 77

7.3 Contributions and discussion . . . 77

8 CONCLUSION . . . 80

REFERENCES . . . 81

APPENDICES . . . 89

APPENDIX A – Measure-theoretic Probability Theory . . . 89

A.1 A primer on Measure theory . . . 89

A.2 Measure theory and Measure-theoretic probability . . . 94

A.3 Random variables and other definitions . . . 96

A.3.0.1 Distribution functions . . . 97

A.3.0.2 Limit theorems and convergence concepts . . . 100

A.3.0.3 Conditional probabilities and conditional expectations . . . 101

A.4 Stochastic processes . . . 103

A.4.0.1 Markov processes . . . 105

A.4.0.2 Master equation . . . 107

APPENDIX B – Parameters of the ASM No. 1 . . . 109

ANNEXES . . . 111

Index 111

1 INTRODUCTION

The field of dynamics is a huge and interdisciplinary area of study. It is the branch of physics that deals with the study of the temporal evolution of systems. When the dynamics of a system is deterministic, the equations of motion are described by ordinary differential equations (ODEs). However, when the dynamics are stochastic to comprise the uncertainty one has about it, the equations of motion are described by stochastic differential equations (SDEs), which is the subject reviewed in this work. The study of dynamics is fundamental to understand the behavior of systems in nature, from the motion of the planets to the behavior of the smallest particles. The equations of motion studied thoroughly in sciences are based temporal rates of displacement known as differential equations invented by Newton in mid-1600s while studying the motion of the planets (STROGATZ, 2018). Of course, those temporal rates are not restricted to those of displacement, but comprise any temporal evolution depending on the system we want to describe.

A system is defined as an object or a collection of objects whose properties one wants to study (LJUNG; GLAD, 1994). Questions about the system are answered through experimentation, but sometimes it is not possible to perform experiments due to the complexity of the system or the high cost of the experiment or even because the system is a conceptualization and thus nonexistent. In such cases, a model is used to represent the system, a tool which aims to answer the questions about the system without experimentation. However, as the model is not the system itself but a representation of it, the answers the model respond are directly related to the assumptions made to construct it.

There are, in principle, two sources from which the modeler can fetch knowledge to construct a model: the data and the theory. The data are the information collected from the system through experiments and fit a generic model such that its properties fit the ones of the system, a process called identification. On the other hand, the theory is the knowledge about the system that is already known, as the laws of nature representing subsystems of the system, and is used to construct the model from scratch. Both sources result in models that have a limited domain of validity, as the data may be noisy, or the theory may be incomplete, even wrong. The modeler must be aware of the limitations of the model and the assumptions made to construct it.

Models can be of different types depending on the nature of the system being studied. There are mathematical models which represent the relationships between quantities as distances, flows, concentrations, etc, that are observed in the system. For instance, the Newton's second law of motion is a mathematical model that relates the force acting on a body to its acceleration. The mathematical models can be classified inside the paradigms (LJUNG; GLAD, 1994):

- Deterministic if the model is able to predict the future of the system with an exact relationship between the quantities involved, or stochastic if there is an uncertainty in the prediction which is dealt with probability theory;
- if the model is time-invariant, or dynamic if the model is time-variant as the trajectories that are solutions of differential equations;
- Continuous-time if the model is described by functions continuous in time, or discrete-time if the model is described by sequences of values in time, in general due to the process of sampling a continuous-time process.

This list is not exhaustive, but it is sufficient for the discussion to come. More classifications as lumped versus distributed (i.e. if the system is described by a finite number of changing variables or not) and others are discussed by (LJUNG; GLAD, 1994). Our objective is to transform deterministic models into stochastic and analyze the implications of it. Regarding the classification above, most of the systems in nature are described by continuous-time models, as the motion of a body, the flow of a river, the concentration of a chemical in a solution, etc. as well as the systems we are going to deal in the applications.

It might be interesting to get the perspective that both the theory of dynamical systems and the theory of probability try to address the behavior of systems using different paradigms. Dynamics stem from physics when Isaac Newton in mid-1600s invented the differential equations to describe the motion of bodies and the universal gravitation to understand the Kepler's law(STROGATZ, 2018). Newton studied a system consisting of the earth translating around the sun comprising then just two bodies. But, it turned out that when including another celestial body, e.g. as the moon, the now called three-body problem was far more difficult, even being proved latter that no collection of algebraic formulas could be able to express its solution (WOLFRAM, 2002), making necessary a different approach. By the same period Pierre de Fermat and Blaise Pascal were studying the problem of determining how to divide the stakes of an unfinished game of dice when the game is interrupted (HACKING, 2006). Without using Newton's recently published ideas, Pascal discussed the chance of the expected outcome when throwing a die with no aid of their dynamics. Pierre-Simon de Laplace publishes then its

Théorie analytique des probabilités(LAPLACE, 1814) where he advocates for the principle of what would come to be the principle of causality, which is the idea that the future is determined by the past. Such principle will rule the models of dynamical systems and would come back posteriorly with Markov (2006) (1906) years later.

Still on the field of chances and long before the advent of statistical mechanics, Bernoulli (1738) suspected that the macroscopic pressure could be explained by “extremely small bodies agitated” causing “continually repeated impacts”, an idea by the time not commonly accepted. Clausius (1857) then introduces one of the first probabilistic modelings of a gas molecule by evaluating the mean free path that such a molecule travels before interacting with another one. Building on Clausius’ work, Maxwell (1860) assumed that, different from Clausius and other contemporary scientists, the collision between the molecules would imply that their speeds would not be equal at a given temperature, but a distribution of values. Maxwell successfully inaugurated the link between the macroscopic thermodynamic observables, pressure and temperature, to microscopic dynamics. Afterwards, it is Boltzmann (1872) the one responsible to ground the mathematical theory of gases along with the probabilistic perspective.

This description of molecular motion gains another chapter in the beginning 20th century with the developments made by Einstein (1905) on the Brownian movement. He evaluated the mean displacement of a particle of pollen in water, a problem unsolved since the observations made by Brown in 1827 who discarded the cause of the motion to be the microscopic living creatures. But it was Langevin (1908) to elaborate an equation of motion compatible with the Newton’s description of dynamics. His idea consisted of including the random behavior in the two century old Newton’s equations establishing a hallmark on the studies of stochastic processes, specifically in the new branch of the today called stochastic differential equations (SDEs). The works of Wiener (1923) started to set the mathematical grounds of the theory, leading to branches as the today’s signal processing field. Although, the solutions for such equations were not formalized until the work of Itō (1944) for the case when random part is a stochastic process with zero mean and constant spectral density, a useful mathematical idealization to describe the action of fluctuating forces acting randomly on a suspended particle as in the case of the Brownian motion.

This work is a summary of the research developed by this author during its master studies. The subject was twofold: the study of stochastic differential equations aiming a generalization of the deterministic models of dynamical systems, and the application of the theory to the stochastic thermodynamics to chemical enzymatic reactions. At first glance both subjects might seem unrelated when studying each area separately, but actually they are just examples of emergent behaviors of complex dynamical systems. Both are dynamical, i.e. they represent the evolution of a system in time, and both are stochastic, i.e. they represent temporal processes that are subject to some kind of “randomness”. These first concepts, dynamics and randomness, common to both the subjects will be dealt with in the first part of this work, specifically we review important aspects of ODEs, which used to model dynamical systems in Chapter 2. Afterwards, we include the randomness on those ODEs and review the theory of SDEs in Chapter 3. Concepts that may be needed from the modern theory of probability including the measure theoretic approach is reviewed in Appendix A.

1.1 Objectives

This work aims to:

1. Overview the theory of SDEs;
2. Showcase different fields in which the theory is applied, more specifically:
    1. Study the stochastic thermodynamics extending the classical thermodynamics, in the context of chemical reactions;
    2. Study the inference of parameters for a bioreactor model using stochastic differential equations to address the uncertainty over the dynamics.

1.2 Contributions

The main contributions of this thesis are:

1. The review the theory of SDEs;
2. Showcase the two works developed during the Masters program:
    1. The study of the thermodynamics of a single molecule Michaelis-Menten mechanism (FARIAS *et al.*, 2023a; FARIAS *et al.*, 2023b);
    2. The inference of parameters of a bioreactor model using SDEs and another on the stochastic thermodynamics of the Michaelis-Menten mechanism (FARIAS *et al.*, 2024).

The first contribution broadens the discussion of SDE theory beyond its traditional role in engineering, linking it with recent applications already recognized in the signal processing community. Both studies presented in Chapters 6 and 7 demonstrate how probability theory can be used to extract insights into the modeled phenomena.

The first study (Chapter 6) examines the stochastic thermodynamics of the Michaelis-Menten mechanism in the limiting case of a single enzyme within an equilibrium solution. This perspective contrasts with the more common models in the literature, which typically consider large enzyme populations (SEIFERT, 2012; QIAN, 2007; BARATO; SEIFERT, 2015).

The second study (Chapter 7) is, to the best of our knowledge, among the first applications of Probabilistic Numerics to an engineering task: parameter inference for a wastewater treatment bioreactor. This approach addresses data scarcity by incorporating expert knowledge directly into the inference process (COCKAYNE *et al.*, 2019; JR *et al.*, 2022; ZONTA *et al.*, 2014a; SHARIFI *et al.*, 2014a).

1.3 Work organization

This work overviews the theory of SDEs which in this sense it is not an easy task given the background literature needed to comprehend them. Thus, given the subjects as measure theory and real analysis are needed to understand both the SDE and modern probability theories, a quick review was included in the Appendix A

This thesis is organized as the following: In Chapter 2 we review the theory of ODEs in order to make a parallel with SDEs in the following chapter. Next, in Chapter 3 the SDEs are motivated using the Langevin dynamics. Then some schemes of solution via stochastic integration are reviewed. In Chapter 4 the application to the Stochastic Thermodynamics is presented, as also the theory for the thermodynamics for Markov Jump processes. In Chapter 5 the Gaussian process in the state-space model is presented as a stochastic process of the Gaussian type (also known as Gaussian processes) which solves certain stochastic differential equations. The Chapters 6 and 7 give the results of the two applications introduced are presented. The first, the inference of the parameters of a bioreactor has been published recently, and the stochastic thermodynamics of the Michaelis-Menten mechanism was presented in workshops.

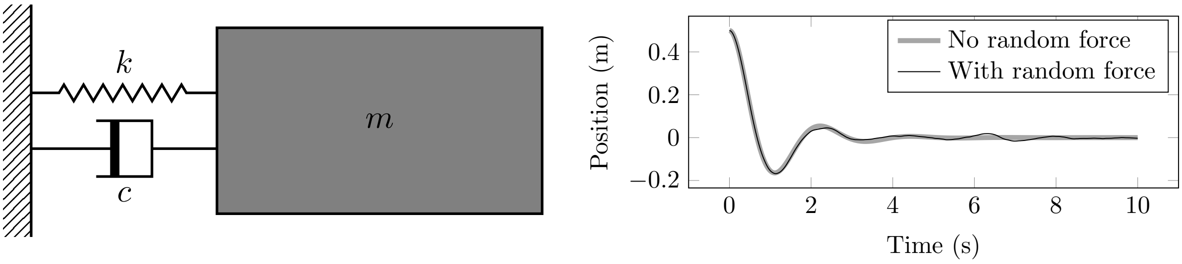
2 OVERVIEW ON DIFFERENTIAL EQUATIONS

This chapter provides a review of the mathematical domain of differential equations to support the study of those that are stochastic. We begin by revisiting core concepts from the theory of ODEs, building intuition and terminology that will later be extended to stochastic settings.

2.1 Introduction to ODEs

In the study of dynamical systems (whether mechanical, biological, or financial) differential equations serve as a foundational mathematical tool to describe how quantities evolve over time. Ordinary differential Equations (ODEs), in particular, provide a way to encode deterministic rules for systems where future states are entirely determined by the current state and rate of change.

Figure 1 – On the left a graphical representation of a damped oscillator, on the right is the graph of the function $x(t)$ which represents the position of the block of mass m versus the time t. On (—) the trajectory is shown when the mass m is not subject to some random force, while on (—) the mass m is subject to a random force F.



Source: Prepared by the author.

For example, in the Figure 1 we have a damped oscillator composed of a block of mass m, coupled to a spring and a dumper with constants k and c, respectively. The motion of this system is described by a differential equation, which is the Equation (2.1) involving derivative terms $\dot{x} \equiv \mathrm{d}x/\mathrm{d}t$ and $\ddot{x} \equiv \mathrm{d}^2x/\mathrm{d}t^2$. For such equations we give them the name of ordinary differential equations (ODEs), simply to distinguish them frompartial differential equations (PDE), which involve partial derivatives. We write the ODE for the motion of the the damped oscillator here omitting the dependence in time t

$$m\ddot{x} + c\dot{x} + kx = 0, \tag{2.1}$$

The terms $\dot{x}$ and $\ddot{x}$ are the first and second time derivatives of the position $x(t)$ corresponding to the velocity and the acceleration, respectively. We say that an ODE is of order n if the highest order derivative that appears in the equation is of order n, which for damped oscillator gives $n = 2$ characterizing a second order ODE. We may also say that the equation is linear if the dependent variable, in this case x, and its derivatives appear linearly in the equation. This equation is said because the right-hand side is zero. If the right-hand side is not zero, we say that the equation is inhomogeneous. A solution of an ODE is a function variable x that satisfies the equation, as shown in the right-plot on Figure 1. The solution might be a particular solution if it satisfies the equation for a given set of constants, or a general solution if it satisfies the equation for any set of constants. In practical terms, this means that to find a solution we need an initial condition, i.e. the value of the variable function x at a given time $t = t_0$. In general, t_0 is chosen to be time at which the motion of the studied systems starts, for instance the initial state (i.e. position and velocity) of the damped oscillator. If the mass m at t_0 is at rest and at a position to be defined as the origin of its trajectory, we have the initial conditions $x(t_0) = 0$ and $\dot{x}(t_0) = 0$. If we modify the initial conditions of the damped oscillator we obtain a different solution to the ODE modeling its motion. Dependence on an external force can be added on the right-hand side of the Equation (2.1). For instance

$$m\ddot{x} + c\dot{x} + kx = F,$$

where F is a force acting on the block of the damped oscillator. Such F must be a time-dependent function or even a random function, which is the main case we are interested in.

2.2 Integration of ODEs

There are practically useful ODEs, the solution to which can be found in a closed form are often referred to as analytical solutions. The most common methods are separation of variables, direct integration, and series solutions (here including the Laplace transform method). Of course, guessing a solution is also possible. The downside of these methods is that they solve only a group (although very large) of ODEs. We present the Examples 1 and 2 just to depict some of them.

Definition 2.2.1 (Separable ODEs (Kreyszig, 2018)) *An ODE is said to be separable if it can be written in the form*

$$g(y)\frac{dy}{dx} = f(x), \tag{2.2}$$

i.e. each side of the equality with only a single variable after algebraic manipulations, where $f : \mathbb{R} \to \mathbb{R}$ *and* $g : \mathbb{R} \to \mathbb{R}$ *are continuous functions. We integrate both sides of the equation with respect to* x *to obtain*

$$\int g(y)\frac{dy}{dx}dx = \int g(y)dy = \int f(x)dx + C,$$

where C *is a constant of integration that can be found given the initial conditions.*

Example 1 ((Kreyszig, 2018)) *Be the ODE*

$$\frac{dy}{dx} = 1 + y^2$$

which can be separable and then integrated as follows

$$\frac{dy}{1+y^2} = dx \Rightarrow \int \frac{dy}{1+y^2} = \int dx \Rightarrow \arctan(y) = x + C.$$

Example 2 ((Kreyszig, 2018)) *Be the homogeneous linear ODE*

$$\frac{dy}{dx} = -p(x)y$$

where p *is any continuous function. Applying the method of separation of variables we obtain*

$$\frac{dy}{dx} = -p(x)y \Rightarrow \frac{dy}{y} = -p(x)dx \Rightarrow \ln|y| = -\int p(x)dx + C^\star.$$

We rewrite the solution in terms of x *by applying the exponential function in both sides and obtaining that*

$$y = e^{-\int p(x)dx + C^\star} = e^{C^\star}e^{-\int p(x)dx} = Ce^{-\int p(x)dx},$$

where $C = \pm e^{C^\star}$ *when* $y \lesseqgtr 0$ *is a constant.*

Apart from the series solutions, most of the methods involve integrating the ODE at some point, as also proving if the solution exists and if it is unique. We are going to present in the next section the Picard-Lindelöf theorem to verify existence and uniqueness. In Chapter 3 we are going to see that there is an analog result for the stochastic case.

2.3 Existence and uniqueness of ODEs solutions

When trying to find functions that solve the ODE, we may wonder if that solution is unique. Furthermore, one may ask if the solution exists at all. The answer to these questions is given by the Picard-Lindelöf theorem. Knowing that the solution exists and is unique is important because it guarantees that the solution is well-defined and that the system is well-posed even if we do not have an algebraic treatment for the solution. We begin with the following definition.

Definition 2.3.1 (Picard iteration) *Let* $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ *be a continuous function. The Picard iteration is the sequence of functions* $\{x_n\}_{n=0}^{\infty}$ *defined by*

$$x_{n+1}(t) = x(0) + \int_{t_0}^{t} f(s, x_n(s))ds, \tag{2.3}$$

where x_0 *is a given function and* t_0 *is a given time.*

One can use the Picard iteration to find solutions. See for instance the following example.

Example 3 (Solving an ODE using Picard iteration) *Be the ODE*

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

with an initial condition $x(0) = 1$. *By the Definition 2.3.1, setting* $x_0(t) = x(0)$

$$x_1(t) = 1 + \int_0^t 1ds = 1 + t, \tag{2.4}$$

$$x_2(t) = 1 + \int_0^t (1+s)ds = 1 + t + 1/2t^2, \tag{2.5}$$

$$\cdots \tag{2.6}$$

$$x_n(t) = \sum_{i=0}^{n} \frac{t^i}{i!}. \tag{2.7}$$

If we take $n \to \infty$ *we obtain the exponential function* e^t *that solves the ODE.*

The Definition 2.3.1 is a recursive definition of a sequence of functions that converges to the solution of the ODE. The Picard iteration is a general method to solve ODEs, but it is not always convergent. The Picard-Lindelöf theorem guarantees that the Picard iteration converges to the solution of the ODE if the function f is Lipschitz continuous. We present the theorem here.

Definition 2.3.2 (The Picard-Lindelöf theorem) *Let* $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ *be a continuous function. If* f *is Lipschitz continuous in the second argument, i.e. there exists a constant* $K > 0$ *such that*

$$|f(t,x) - f(t,y)| \leq K|x-y|, \tag{2.8}$$

for all $t \in \mathbb{R}$ *and* $x, y \in \mathbb{R}^d$, *then for the ODE*

$$\dot{x} = f(t,x), \tag{2.9}$$

there exists an $\varepsilon > 0$ *such that the ODE has a unique solution in the interval* $[t_0 - \varepsilon, t_0 + \varepsilon]$ *for every initial condition* $x(t_0) = x_0$.

The proof is long and outside the scope of the current text, but easily found in references (CODDINGTON; LEVINSON, 1984).

2.4 State-space representation of ODEs

A general framework for the studying the dynamics of systems which the model is given by an ODE is the state-space representation. Generically in this representation, the state of the system is described by system of differential equations as

$$\begin{aligned}
\dot{x}_1 &= f_1(x_1, x_2, \ldots, x_n), \\
\dot{x}_2 &= f_2(x_1, x_2, \ldots, x_n), \\
&\vdots \\
\dot{x}_n &= f_n(x_1, x_2, \ldots, x_n),
\end{aligned}$$

The variables x_i are the state variables of the system and the functions f_i are the dynamics of the system. For instance, to describe the oscillator in the state space representation we make

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -\frac{c}{m}x_2 - \frac{k}{m}x_1,$$

where now we can describe the system in two new coordinates x_1 and x_2 representing the position and the velocity of the block, respectively. Observe that now the system had its order reduced at the cost of increasing the number of variables. This has the advantage that the solution is easily obtained similarly to the Example 2. We refer to Strogatz (2018) for further discussion on the insights we gain using this approach to analyze dynamical systems qualitatively.

We may not be able to observe all the variables of the system, or even the system may be nonautonomous system, i.e. it may depend on external input $u(t)$ and a random function W_t. To comprise these possibilities, we use the following representation of a system of differential equations

$$\dot{\mathbf{x}} = f\left(\mathbf{x}(t), \mathbf{u}(t), W_t, \theta_{\mathbf{x}}\right), \tag{2.10a}$$

$$\mathbf{y} = g\left(\mathbf{x}(t), \mathbf{u}(t), W_t, \theta_{\mathbf{y}}\right), \tag{2.10b}$$

being f and g functions whose images are compatible with the dimensions of $\dot{\mathbf{x}}$ and $\mathbf{y}$, respectively. The state at time t at which the trajectory of the system is located is denoted by $\mathbf{x}(t) := [x_1, ..., x_{N_x}] \in \mathbb{R}^{N_x}$, and the system is observed through the variables $\mathbf{y}(t) := [y_1, ..., y_{N_y}] \in \mathbb{R}^{N_y}$. The system is actuated by an external agent through an input denoted by $\mathbf{u}(t) \in \mathbb{R}^{N_u}$. The system may be disturbed by random interactions $W_t : \Omega \times [t_0, T] \to \mathbb{R}^{N_w}$. The function f describes the rate at which the states vary according to the relation between $\mathbf{x}$, $\mathbf{u}$, W_t and the parameters $\theta_{\mathbf{x}}$. The function g describes how the observations $\mathbf{y}(t) \in \mathbb{R}^{N_y}$ are related to the states, inputs, disturbances, and parameters θ_y of the system. It is commonly assumed in literature that the observation model g is autonomous and independent of $\mathbf{u}$.

2.4.0.1 Linear dynamical systems

A simplification that will be needed in the applications is the special case when the dynamics are described by a function f that is linear, as also the observation function g. The called linear dynamical system is defined as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{G}W_t, \tag{2.11a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}(t) + \mathbf{H}W_t, \tag{2.11b}$$

where $\mathbf{A} \in \mathbb{R}^{N_x \times N_x}$, $\mathbf{B} \in \mathbb{R}^{N_x \times N_u}$, $\mathbf{C} \in \mathbb{R}^{N_y \times N_x}$, and $\mathbf{G} \in \mathbb{R}^{N_x \times N_w}$, $\mathbf{H} \in \mathbb{R}^{N_y \times N_w}$ are the matrices that define the system and multiply the vectors $\mathbf{x}$, $\mathbf{u}$, W_t, and $\mathbf{y}$ with dimensions $1 \times N_x$, $1 \times N_u$, $1 \times N_w$, and $1 \times N_y$, respectively. The disturbances W_t in both dynamics and observation of the Equation (2.11) are not restricted to be the same W_t, we used the same process here just for simplicity. Notice the matrices defining the system are not restricted to be constant in time (leading to time-invariant model of the system). Even though the linear system seems to be very restrictive assuming the linearity of the system, it is a tool to describe the dynamics of many systems in nature and gives rise to the rich field linear systems (CHEN, 1984).

Although the theory of nonlinear systems is more general, much of the results are handled case by case (STROGATZ, 2018; VIDYASAGAR, 1993). Instead, we assume the systems we are dealing in this work to be linearizable if are nonlinear, i.e. the system can be approximated by a linear system around a point of operation.

2.4.0.2 Linearization

The systems in the real world are in general nonlinear, but they can be approximated until some extent by linear systems. The linearization of a system is the process of approximating the nonlinear system by a linear one around a point of operation. The linearization is done by the Taylor series expansion of the nonlinear function around the point of operation. Let $\mathbf{x}_o(t)$ and $\mathbf{u}_o(t)$ be solution of the system Equation (2.10) considering the system to be free of disturbances, i.e. in Equation (2.10a) $\mathbf{W}_0 = 0$. The system is then perturbed away from $\mathbf{x}_o$ and $\mathbf{u}_o$ by the vectors $\Delta\mathbf{x}(t) := \mathbf{x}(t) - \mathbf{x}_o(t)$ and $\Delta\mathbf{u}(t) := \mathbf{u}(t) - \mathbf{u}_o(t)$. Under this assumption, the perturbed solution is Taylor expanded

$$\begin{aligned} \Delta\dot{\mathbf{x}} + \dot{\mathbf{x}}_o &= f(\mathbf{x}_o + \Delta\mathbf{x}, \mathbf{u}_o + \Delta\mathbf{u}, \theta_x), \\ &= f(\mathbf{x}_o, \mathbf{u}_o, \theta_x) + \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{u}_o} \Delta\mathbf{x} + \left.\frac{\partial f}{\partial \mathbf{u}}\right|_{\mathbf{x}_o,\mathbf{u}_o} \Delta\mathbf{u} + O(\Delta\mathbf{x}^2, \Delta\mathbf{u}^2), \end{aligned} \tag{2.12}$$

where the terms of the order of $\Delta\mathbf{x}^2$ and $\Delta\mathbf{u}^2$ or higher represented by $O(\Delta\mathbf{x}^2, \Delta\mathbf{u}^2)$ are neglected. Being $\dot{\mathbf{x}}_o := f(\mathbf{x}_o, \mathbf{u}_o, \theta_x)$, the linearized dynamics are given by

$$\Delta\dot{\mathbf{x}} = \mathbf{A}\Delta\mathbf{x} + \mathbf{B}\Delta\mathbf{u},$$

where $\mathbf{A} = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}_o,\mathbf{u}_o}$ and $\mathbf{B} = \left.\frac{\partial f}{\partial \mathbf{u}}\right|_{\mathbf{x}_o,\mathbf{u}_o}$ are the Jacobian matrices of the system.

2.4.0.3 Solution of linear systems

The time trajectory of the linear system in Equation (2.11a) is given by the solution of the linear system of equations. We use the following property of the exponential matrix (CHEN, 1984)

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{\mathbf{A}t} = \mathbf{A} e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A},$$

and we premultiply $e^{-\mathbf{A}t}$ in both sides of the Equation (2.11a) to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-\mathbf{A}t}\mathbf{x}(t)\right) = e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t) + e^{-\mathbf{A}t}\mathbf{G}W_t.$$

We integrate both sides of the equation from 0 to t, and using that $(e^{-\mathbf{At}})^{-1}$ has as inverse $e^{\mathbf{At}}$, and that $e^{\mathbf{0}} = \mathbf{I}$ where $\mathbf{0}$ is the zero matrix and $\mathbf{I}$ is the identity matrix, we obtain

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{Bu}(\tau)\mathrm{d}\tau + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{G}W_\tau \mathrm{d}\tau. \tag{2.13}$$

If the system does not depend on the random function W_t, the last term would be zero, and we would have to deal with conventional integrals. The last term is a stochastic integral and is defined in the Section 3.3.

2.5 Numerical methods to solve ODEs

The integration of ODEs are limited to a small group of equations, for instance the ODEs of first order presented in the latter section. For the general case, we need to resort to numerical methods. The most common numerical methods are the Euler method and the Runge-Kutta method which we present in the Examples 4 and 5 below, respectively.

Example 4 (Euler's method) *Let us consider the ODE*

$$\dot{x} = f(t,x),$$

where $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ *is a continuous function. The Euler's method is a numerical method to solve the ODE by approximating the derivative by a finite difference. The method is defined by the recursive relation*

$$x_{n+1} = x_n + hf(t_n, x_n), \tag{2.14}$$

where h *is the step size and* $t_n = t_0 + nh$.

Example 5 (Runge-Kutta's method) *The Runge-Kutta's method is a generalization of the Euler method. The method is defined by the recursive relation*

$$x_{n+1} = x_n + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right), \tag{2.15}$$

where

$$k_1 = f(t_n, x_n),$$
$$k_2 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1),$$
$$k_3 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_2),$$
$$k_4 = f(t_n + h, x_n + hk_3),$$

where $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ *is a continuous function.*

Both Euler's and Runge-Kutta's methods are suited for first-order ODEs. While the first give the (global) error of the estimated solution due to the truncation of the Taylor series proportional to h, the second has it proportional to h^2. Consenquently, we need to sample in fewer points for the Runge-Kutta than for the Euler to achieve the same error for the solution. Notice that numerical methods may also be useful for computing analytical solutions given for instance by the linear system case in Equation (2.13). This solution involves computing the matrix exponential, which is given by the Taylor series of the exponential function and thus, being computationally costly (MOLER; LOAN, 2003). In Chapter 3 we will see that both methods have similar analogs for the stochastic case.

3 STOCHASTIC DIFFERENTIAL EQUATIONS

In this chapter the SDEs are motivated using the Langevin approach to Brownian motion. The stochastic integration to solve such differential equations are discussed as well as the Euler-Maruyama scheme for the numerical integration.

3.1 Motivation: The Brownian motion

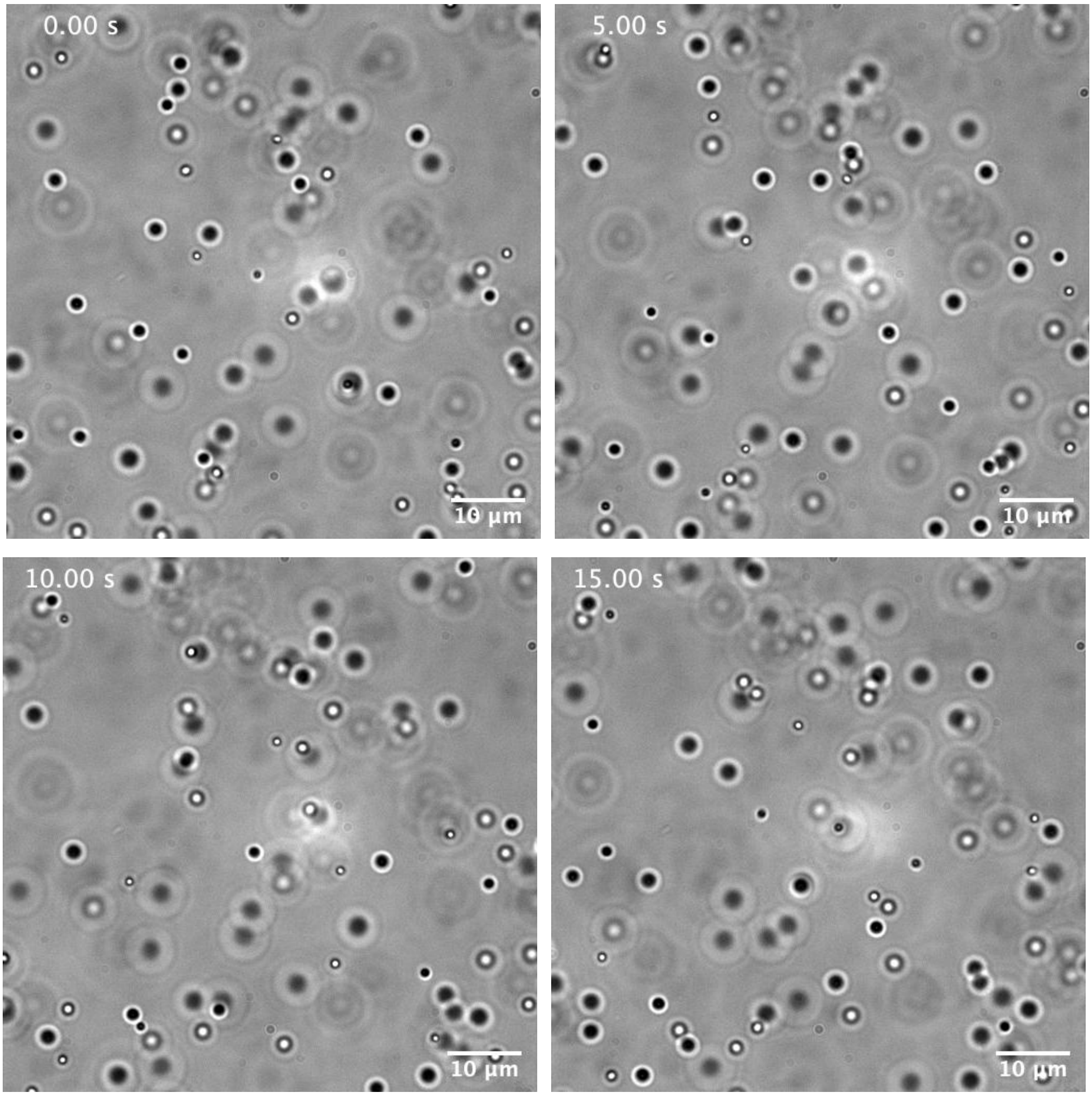
The phenomenon now known as Brownian motion was first rigorously documented by Brown (1828). While examining pollen grains suspended in water under a microscope, Brown observed an incessant, jittery movement of the particles. Crucially, he verified that this motion was not attributable to water currents or biological activity. To eliminate the hypothesis of life-driven movement, he repeated the experiment using inorganic particles extracted from the interior of quartz crystals, and still observed the same erratic behavior. This suggested a universal physical cause behind the motion, independent of the nature of the suspended material.

More than seventy years later, Einstein (1905) followed by Smoluchowski (1906) provided a theoretical explanation of this phenomenon. He modeled the suspended particles as being bombarded from all directions by water molecules undergoing random thermal motion. These incessant, uncoordinated impacts caused the particles to follow unpredictable paths. Einstein derived a statistical expression for the mean squared displacement of such particles over time, thereby linking microscopic molecular activity to macroscopic observable diffusion. His theory not only confirmed the molecular theory of matter, but also laid the foundation for modern stochastic modeling.

Soon after, Langevin (1908) proposed a more detailed physical model. He formulated an equation of motion for a single suspended particle, incorporating a deterministic drag force proportional to velocity and a stochastic force term accounting for the random impacts from surrounding molecules. This Langevin equation introduced the description of physical systems exposed to random influences mathematically by differential equations with added noise. Langevin's formulation was one of the first to explicitly recognize and incorporate stochastic forces into dynamical systems, paving the way for the theory of stochastic differential equations.

In Figures 2 and 3 we show the motion of a Brownian particle in a liquid. The first figure shows the motion of plastic particles observed by Miyazaki *et al.* (2022) under the

Figure 2 – Brownian motion observed under the microscope. The images show the motion of plastic particles with diameter of 1μm (micrometer) suspended in water. The time interval between frames is 0.5 seconds.

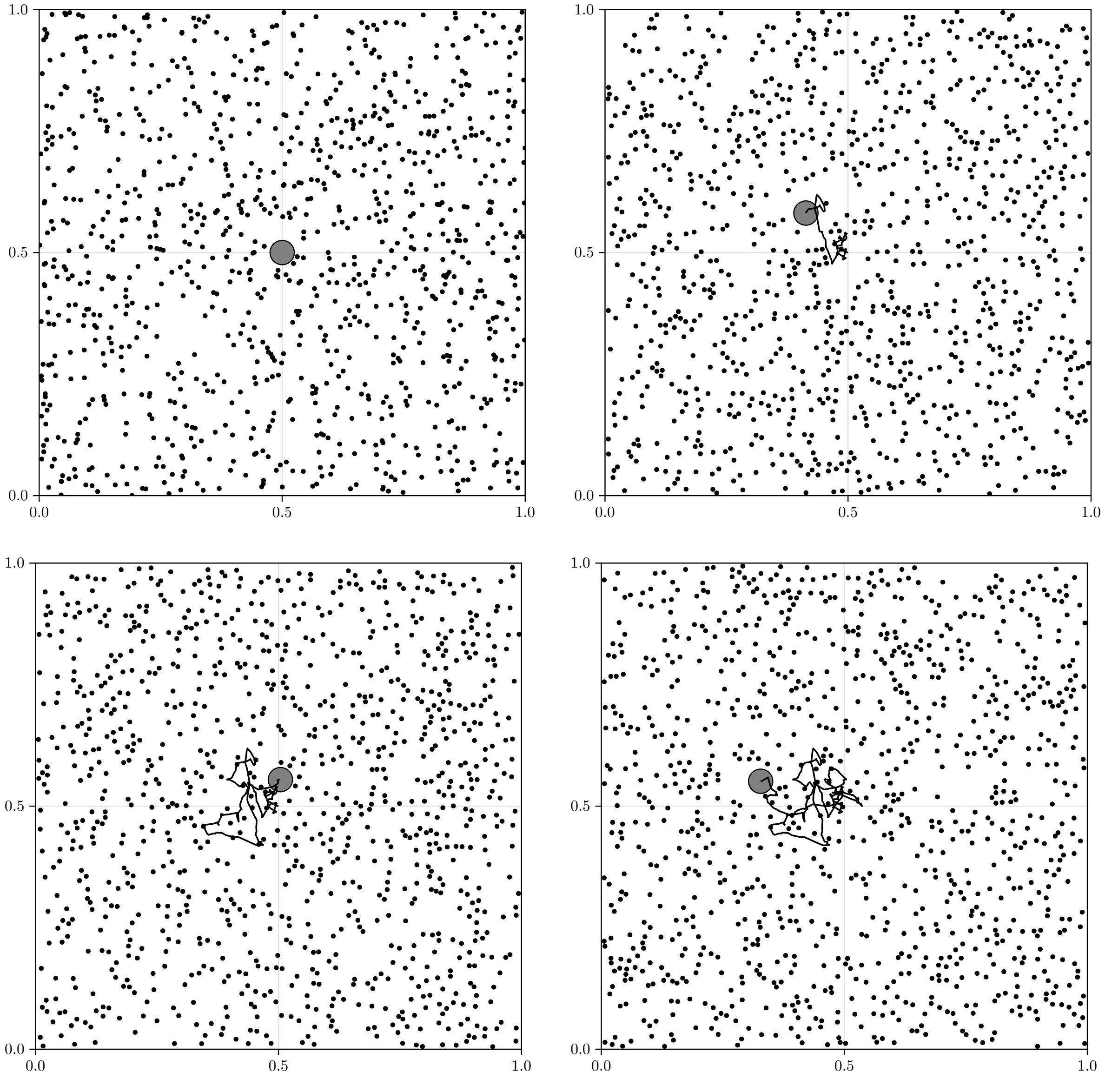


Source: Extracted from Miyazaki *et al.* (2022)

microscope. The second figure shows the simulation of a Brownian particle (gray circle) colliding with smaller particles (black dots) in a liquid. The simulation was performed considering hard collisions between the particles, as an ideal gas just like the Boltzmann assumption in the modelling proposed by Einstein.

We illustrate the Langevin's approach trying to solve the problem of a particle in a damping medium (BUTKOV, 1968). We represent the motion of a particle of mass m in a

Figure 3 – Simulated collision of Brownian particle (gray circle) with smaller "gas" particles (black dots). From left-right, top-down are shown the frames of the simulation at times $t = 0.0, 3.2, 6.4$ and 9.6 milliseconds. The black line is the trajectory of the Brownian particle. The particles are out of scale.



Source: Preparated by the author.

damping medium with resistance γ under influence of a force Ξ_t, and with velocity $v(t)$ with the following differential equation (or Langevin equation)

$$m\dot{v}(t) = -\gamma v(t) + \Xi_t.$$

If $\Xi_t = 0$ the solution is $v(t) = v_0 e^{-\gamma t/m}$ and after some time the particle with initial velocity v_0 rests.

Differently from the observation of Brown, the particle never rests as long as its surroundings does not alter its settings (i.e., it is in equilibrium meaning the absence of mechanical, thermal and chemical gradients). We conveniently denote the surroundings of the particle by heat bath with a constant temperature T. Some assumptions are made with respect to the interactions between the particle and the surroundings (ZWANZIG, 2001): their interactions are promoted via a force Ξ_t whose amplitude fluctuates, and it happens occasionally; the force varies very rapidly over the time of observation. We summarize these characteristics of Ξ_t by representing it as a random process whose moments are

$$\mathbb{E}[\Xi_t] = 0 \text{ and } \mathbb{E}[\Xi_t \Xi_{t'}] = 2\gamma k_B T \delta(t - t').$$

The velocity of the Brownian particle is then given by solving the Langevin equation

$$v(t) = v(0)e^{-\frac{\gamma}{m}} + \int_0^t \Xi_\tau e^{-\frac{\gamma}{m}(t-\tau)} \mathrm{d}\tau.$$

It is Paley and Wiener (1934) to bring attention to the need of setting a mathematical background to the theory. The step was accomplished by Itō (1944) who defined the stochastic integral.

In the next section, we discuss how to approach the modeling of systems that have random influences. The first step is to define the modeling itself. The modeling is the process of representing the system in a mathematical form. The second step is to define the stochastic differential equations that will be used to describe the dynamics of the system influenced by randomness. The third step is to define the numerical methods that will be used to solve the stochastic differential equations.

3.2 Stochastic integrals

Let a general kind of stochastic differential equation (SDE) be

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t) + g(x,t)W_t, \tag{3.1}$$

where f and g are given functions. For this kind of equation, we assume that W_t is a stochastic process as defined in Section A.4. The solution of the Equation (3.1) is given by the integral equation

$$x(t) = x(0) + \int_0^t f(x(s),s)\mathrm{d}s + \int_0^t g(x(s),s)W_s \mathrm{d}s.$$

We can deal with the first integral in a “classical” manner by integrating using the Riemann formulation. However, the second integral is problematic because the integrand is a stochastic process. Depending the assumptions made on W_t, the integral may not exist. In fact, assuming that W_t is a white noise process, the function is not measurable and the integral does not exist in a Lebesgue sense (see Section A.4).

We proceed following Oksendal (2013) and mount the SDE in a form that suggests the replacement of the white noise process for another stochastic process. We rewrite the Equation (3.1) in a discrete form. Let the partitioning of the interval $[0,T]$ be given by $0 = t_0 < t_1 < ... < t_N = T$ such that $\Delta t = t_{n+1} - t_n$, and the difference equation

$$X_{t_{n+1}} - X_{t_n} = f(X_{t_n}, t_n)\Delta t + g(X_{t_n}, t_n)\Xi_{t_n}\Delta t, \tag{3.2}$$

where Ξ_{t_n} is a random function that represents the value of W_t at time t_n. We assume the existence of a stochastic process B_t such that $\Xi_{t_n}\Delta t := \Delta B_{t_n}$, where $\Delta B_{t_n} = B_{t_{n+1}} - B_{t_n}$. Also, B_t is a stochastic process with independent increments and zero mean. The next step is to make sense of the limit of this discretisation when $\Delta t \to 0$ and substitute the summation in Equation (3.2) by an integral.

We begin by the same fashion as in Appendix A.1 when we defined the Lebesgue integral. We define a simple function and extend it by some approximation. Let us assume we want to approximate a function $f(t,\omega)$ by

$$\sum_{j\leq 0} f(t^\star, \omega)\mathbf{1}_{[j2^{-n},(j+1)2^{-n})}(t),$$

where $t^\star \in [j2^{-n}, (j+1)2^{-n})$. We want to define an integral

$$\int_{\mathcal{I}} f(t,\omega)\mathrm{d}B_t(\omega), \tag{3.3}$$

where $\mathcal{I} = [t_0, t]$ as a limit of

$$\sum_{j\leq 0} f(t^\star, \omega)\left(B_{(j+1)2^{-n}}(\omega) - B_{j2^{-n}}(\omega)\right), \tag{3.4}$$

as $n \to \infty$. To this end, we need the following definitions.

Definition 3.2.1 (Adapted process (OKSENDAL, 2013)) *A stochastic process X_t in the probability space $(\Omega, \mathscr{F}, P)$ is said to be adapted to the filtration $\mathscr{F}_t$ if X_t is $\mathscr{F}_t$-measurable, where $\{\mathscr{F}_t \subset \mathscr{F}\}_{t\geq 0}$ is an increasing family (i.e. $\mathscr{F}_s \subset \mathscr{F}_t$ for $s < t$) of σ-algebras.*

In other words, considering a random variable $X_t : \Omega \to \mathbb{R}$ in the probability space $(\Omega, \mathscr{F}, P)$ together with the Definition A.6, if for every Borel set $\mathcal{B} \in \mathscr{B}(\mathbb{R})$ we have that $X_t^{-1}(\mathcal{B}) \in \mathscr{F}_t$. For instance, the process $B_{t/2}$ is adapted to the filtration $\mathscr{F}_t$, while B_{2t} is not.

Be $\mathcal{M}$ the set of functions $G(t,\omega) : [t_0,t] \times \Omega \to \mathbb{R}$ such that it is $(\mathscr{B}([t_0,t)) \times \mathscr{F})$-measurable, $\mathscr{F}_t$-adapted and with $\mathbb{E}\left[\int_0^t G^2(s,\omega)\mathrm{d}s\right] < \infty$. Also, we say that $G(t,\omega) \in \mathcal{M}$ is a step function if there exists a decomposition $t_0 < t_1 < ... < t_n = t$ such that $G(s,\omega) = G(t_{i-1},\omega)$ for all $s \in [t_{i-1},t_i)$ where $i = 1,...,n$. It follows the definition

Definition 3.2.2 (Stochastic integral (ARNOLD, 1974)) *For every function $G \in \mathcal{M}$, the stochastic integral of G with respect to B_t over the interval $[t_0,t]$ is defined as the random variable*

$$\int_{t_0}^{t} G(s,\omega)dB_s(\omega) := \sum_{i=1}^{n} G(t_{i-1}\omega)(B_{t_i}(\omega) - B_{t_{i-1}}(\omega)).$$

Next, we obtain that for an arbitrary function $G(t,\omega)$ in $\mathcal{M}$, there exists a series of step functions $G_n(t,\omega) \in \mathcal{M}$ such that

$$\text{st-}\lim_{n\to\infty} \int_{t_0}^{t} |G(s,\omega) - G_n(s,\omega)|^2 \mathrm{d}s = 0. \tag{3.5}$$

The proof can be found in Arnold (1974). It can also be found the proof of Equation (3.5) implying that the convergence of the sequence of integrals

$$\int_{t_0}^{t} G_n(s,\omega)\mathrm{d}B_s$$

to a specific random variable. Then, follows our final definition

Definition 3.2.3 ((Itō's) stochastic integral (ARNOLD, 1974)) *For every function $G \in \mathcal{M}$, the (Itō's) stochastic integral of G with respect to the process B_t over the interval $[t_0,t]$ is defined as*

$$I(G) = \text{ms-}\lim_{n\to\infty} \int_{t_0}^{t} G_n(s,\omega)\mathrm{d}B_s,$$

where $\{G_n\}$ is a sequence of step functions in $\mathcal{M}$ that approximates G in the sense that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_{t_0}^{t} |G(s,\omega) - G_n(s,\omega)|^2 \mathrm{d}s\right] = 0.$$

Example 6 (From Arnold (1974)) *Let $G(t,\omega) = B_t(\omega)$ in the Definition 3.2.3, i.e.*

$$\int_{t_0}^{t} B_s dB_s.$$

If B_t is a.s. continuous then we may write it as a sequence of step functions, then we have by the Definition 3.2.3

$$\int_{t_0}^{t} B_s(\omega)dB_s(\omega) = \text{as-}\lim_{\delta_n\to\infty} \sum_{k=1}^{n} B_{t_{k-1}}(\omega)\left(B_{t_k}(\omega) - B_{t_{k-1}}(\omega)\right),$$

for $t_0 < t_1 < ... < t_n = t$ *and* $\delta_n = \max_k |t_k - t_{k-1}|$. *Let us write the sum (omitting the* ω *for simplicity)*

$$S_n = \sum_{k=1}^{n} B_{t_{k-1}} \left(B_{t_k} - B_{t_{k-1}}\right).$$

We use the identity

$$\left(B_{t_k} - (B_{t_k} - B_{t_{k-1}})\right)^2 = B_{t_k}^2 + (B_{t_k} - B_{t_{k-1}})^2 - 2B_{t_k}(B_{t_k} - B_{t_{k-1}}),$$

to obtain

$$S_n = \frac{1}{2}B_t^2 - \frac{1}{2}B_{t_0}^2 - \frac{1}{2}\sum_{k=1}^{n}\left(B_{t_k} - B_{t_{k-1}}\right)^2.$$

If S_n *converges in mean square, this implies that it converges stochastically. We look then for the mean square limit of* S_n *as* $n \to \infty$

$$\text{ms-}\lim_{\delta_n \to \infty} S_n = \frac{1}{2}B_t^2 - \frac{1}{2}B_{t_0}^2 - \frac{1}{2}(t - t_0).$$

3.2.0.1 Itō's lemma

Other consequence of the stochastic integral is the Itō's lemma, which is a generalization of the Taylor's theorem for stochastic processes. It is used to obtain the stochastic differential equation of a function $f(X_t,t)$, where X_t is a stochastic process defined by the stochastic differential

$$\mathrm{d}X_t = f(X_t,t)\mathrm{d}t + g(X_t,t)\mathrm{d}B_t.$$

The Itō's lemma states that if f is a twice continuously differentiable function, then

$$\mathrm{d}f(X_t,t) = \left(\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}f(X_t,t) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}g^2(X_t,t)\right)\mathrm{d}t + \frac{\partial f}{\partial x}g(X_t,t)\mathrm{d}B_t.$$

The d-dimensional version of the Itō's lemma is given by

$$\begin{aligned}\mathrm{d}f(X_t,t) = &\left(\frac{\partial f}{\partial t} + \sum_{i=1}^{d}\frac{\partial f}{\partial x_i}f(X_t,t) + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 f}{\partial x_i \partial x_j}g_i(X_t,t)g_j(X_t,t)\right)\mathrm{d}t \\ &+ \sum_{i=1}^{d}\frac{\partial f}{\partial x_i}g_i(X_t,t)\mathrm{d}B_t^i.\end{aligned}$$

where $g_i(X_t,t)$ is the i-th component of the vector $g(X_t,t)$ and B_t^i is the i-th component of the vector B_t. The proof of the Itō's lemma can be found in (ARNOLD, 1974).

3.2.0.2 Stratonovich integral

In the Definition 3.2.2 we assumed the decomposition of a function $G(t,\omega)$ in a series of step functions, such that $G(s,\omega)$. It turns out that this is not the only way to decompose the function. In fact, we can also decompose it as $G(s,\omega) = G(t_i,\omega)$ for all $s \in [t_{i-1}, t_i)$. This leads to a different definition of the stochastic integral, called Stratonovich integral which we are going to discuss below.

The same way we defined the Itō's integral in the Definition 3.2.3, we may define the Stratonovich integral, but with the decomposition $G(s,\omega) = 1/2\,(G(t_i,\omega) + G(t_{i-1},\omega))$. Also, we use $\circ$ to denote the Stratonovich integral as in the following example.

Example 7 (From Arnold (1974)) *Revisiting the Example 6, let $G(t,\omega) = B_t(\omega)$ in the Definition 3.2.2, i.e.*

$$\int_{t_0}^{t} B_s \circ dB_s,$$

but now using the Stratonovich integral. If B_t is a.s. continuous then we may write it as a sequence of step functions, then we have by the Definition 3.2.2

$$\int_{t_0}^{t} B_s(\omega) \circ dB_s(\omega) = \text{as-}\lim_{\delta_n \to \infty} \sum_{k=1}^{n} \frac{(B_{t_k} + B_{t_{k-1}})}{2} \left(B_{t_k}(\omega) - B_{t_{k-1}}(\omega)\right),$$

for $t_0 < t_1 < \ldots < t_n = t$ and $\delta_n = \max_k |t_k - t_{k-1}|$. Let us write the sum (omitting the ω for simplicity)

$$S_n = \sum_{k=1}^{n} \frac{(B_{t_k} + B_{t_{k-1}})}{2} \left(B_{t_k} - B_{t_{k-1}}\right).$$

If S_n converges in mean square, this implies that it converges stochastically. We look then for the mean square limit of S_n as $n \to \infty$

$$\text{ms-}\lim_{\delta_n \to \infty} S_n = \frac{1}{2} B_t^2 - \frac{1}{2} B_{t_0}^2.$$

The use of the Stratonovich integral is more appropriate when we want to apply the chain rule once it behaves like the classical calculus. Although, there are several properties of the Itō's integral that are not satisfied by the Stratonovich integral. For instance, the Itō's integral is a martingale (i.e. a stochastic process with independent increments, see Definition A.29 in Appendix A.4) while the Stratonovich integral is not.

3.2.1 Existence and uniqueness of solutions of SDEs

We saw in the Section 2.3 that the existence and uniqueness of solutions of ordinary differential equations is guaranteed by the Picard-Lindelöf theorem (see Section 2.3). For stochastic differential equations, the existence and uniqueness of solutions is guaranteed by the following definitions.

Definition 3.2.4 (Existence and uniqueness of solutions of SDEs (ARNOLD, 1974)) *Let the following stochastic differential equation be given by*

$$dX_t = f(X_t,t)dt + G(X_t,t)dB_t,$$

where $X_{t_0} = C$ *(where C is a random variable),* $t_0 \le t \le T < \infty$*. The function* f *is assumed to be* $\mathbb{R}$ *valued and G to be measurable on* $[t_0,T] \times \mathbb{R}$*. Both functions are assumed to follow the conditions:*

1. *f and G are Lipschitz continuous, i.e. there exists a constant* $k > 0$ *such that*

   $$|f(x,t) - f(y,t)| + |G(x,t) - G(y,t)| \le k|x-y|, \quad \forall x,y \in \mathbb{R}, t \in [t_0,T].$$

2. *f and G are of functions of constant growth, i.e. there exists a constant* $K > 0$ *such that*

   $$|f(x,t)|^2 + |G(x,t)|^2 \le K(1 + ||x||), \quad \forall x \in \mathbb{R}, t \in [t_0,T].$$

Then, the SDE has a unique solution X_t *in the interval* $[t_0,T]$ *which is continuous with probability one, i.e. if* X_t *and* Y_t *are two solutions of the SDE with initial condition* $X_{t_0} = C$*, then*

$$P\left(\sup_{t\in[t_0,T]} |X_t - Y_t| > 0\right) = 0.$$

The proof of this theorem is long and can be found in (ARNOLD, 1974). The conditions of the Definition 3.2.4 guarantee that f and G do not change faster with the change of x.

3.2.2 Statistics of SDEs

In the next discussion, we follow Särkkä and Solin (2019) and we define the probability density function for an SDE, whose partial differential equation describing its time evolution is known as Fokker-Planck-Kolmogorov equation. We begin by defining the infinitesimal generators which tells us how the expected values of a function of the stochastic process is changing in time.

Definition 3.2.5 (Infinitesimal generator (SÄRKKÄ; SOLIN, 2019)) *The infinitesimal generator of a stochastic process X_t for a function $\phi(X_t)$ may be defined as*

$$\mathfrak{A}(\phi(X_t)) = \lim_{s \searrow 0} \frac{\mathbb{E}[\phi(X_{t+s})] - \phi(X_t)}{s}.$$

For a process X_t that solves the SDE

$$dX_t = f(X_t,t)dt + G(X_t,t)dB_t,$$

which is also called Itō process, the generator operating on $\cdot$ is given by

$$\mathfrak{A}(\cdot) = \sum_i \frac{\partial \cdot}{\partial [X_t]_i}[f(X_t,t)]_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 \cdot}{\partial [X_t]_i \partial [X_t]_j}[G(X_t,t)G(X_t,t)^\top]_{ij},$$

where $[X_t]_i$ is the i-th component of X_t.

It can be shown that the Definiton 3.2.5 leads to the Fokker-Planck-Kolmogorov equation.

Definition 3.2.6 (Fokker-Planck-Kolmogorov (FPK) equation (SÄRKKÄ; SOLIN, 2019)) *From the Definition 3.2.5, $\phi(X_t) = p_{X_t}(x)$ is the probability density function of X_t, solution of the SDE*

$$dX_t = f(X_t,t)dt + G(X_t,t)dB_t,$$

solves the partial differential equation

$$\frac{\partial p_{X_t}(x)}{\partial t} = -\sum_i \frac{\partial [f(X_t,t)]_i p_{X_t}(x)}{\partial [X_t]_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2 [G(X_t,t)G(X_t,t)^\top]_{ij} p_{X_t}(x)}{\partial [X_t]_i \partial [X_t]_j},$$

known as the Fokker-Planck-Kolmogorov equation.

In some cases, instead of solving the complete FPK equation from Definition 3.2.6 we may obtain some moments as the mean and covariance(SÄRKKÄ; SOLIN, 2019). For instance, if we Taylor expand $\phi(X_t)$ and take the expectation we obtain

$$\frac{\mathrm{d}\mathbb{E}[\phi]}{\mathrm{d}t} = \mathbb{E}\left[\frac{\partial \phi}{\partial t}\right] + \mathbb{E}\left[\frac{\partial \phi}{\partial [X_t]_i}[f(X_t,t)]_i\right] + \frac{1}{2}\sum_{i,j}\mathbb{E}\left[\frac{\partial^2 \phi}{\partial [X_t]_i \partial [X_t]_j}[G(X_t,t)G(X_t,t)^\top]_{ij}\right].$$

If we choose $\phi(X_t,t) = [X_t]_i$ then $\frac{\mathrm{d}\mathbb{E}[[X_t]_i]}{\mathrm{d}t} = \mathbb{E}[[f(X_t,t)_i]]$, thus we denote the complete vector of the dynamics of the mean of the SDE as

$$\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t} = \mathbb{E}[f(X_t,t)], \text{ with } \mathbf{m}(t) = \mathbb{E}[X_t].$$

By a similar argument, choosing $\phi(X_t,t) = [X_t]_i[X_t]_j - [\mathbf{m}(t)]_i[\mathbf{m}(t)]_j$ gives the complete matrix of the dynamics of the covariance of the SDE as

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \mathbb{E}\left[f(X_t,t)(X_t - \mathbf{m}(t))^\top\right]\mathbb{E}\left[(X_t - \mathbf{m}(t))f(X_t,t)^\top\right] + \mathbb{E}\left[G(X_t,t)G(X_t,t)^\top\right].$$

It is important to note that this is relevant for the Brownian motion once mean and covariance are complete statistics for the Gaussian distribution, then we can describe the distribution of the Brownian motion completely, if the SDE is linear. In the nonlinear case, it may happen that the statistics to be dependent on higher order terms, turning the problem difficult to solve even for such lower moments. Also, there are exceptions for the Gaussian distribution, as the Black-Scholes model (BLACK; SCHOLES, 1973), which possesses multiplicative noise, but the first two moments are obtainable in a closed form (see the Examples 5.15 and 5.16 of Särkkä and Solin (2019)).

3.3 Numerical stochastic integration

In order to obtain numerical solutions for the SDEs the common method is a modification of the Euler's method for solving differential equations was proposed by (MARUYAMA, 1954). This method was named after him as Euler-Maruyama method and is defined as follows: for a stochastic differential equation defined as

$$\mathrm{d}X_t = f(X_t,t)\mathrm{d}t + g(X_t,t)\mathrm{d}B_t.$$

It is possible to approximate the stochastic process X_t which solves the stochastic differential equation defined above by using the following iteration scheme

$$X_{t_n} - X_{t_{n-1}} = f(X_{t_{n-1}},t_{n-1})(t_n - t_{n-1}) + g(X_{t_{n-1}},t_{t_{n-1}})(B_{t_n} - B_{t_{n-1}}),$$

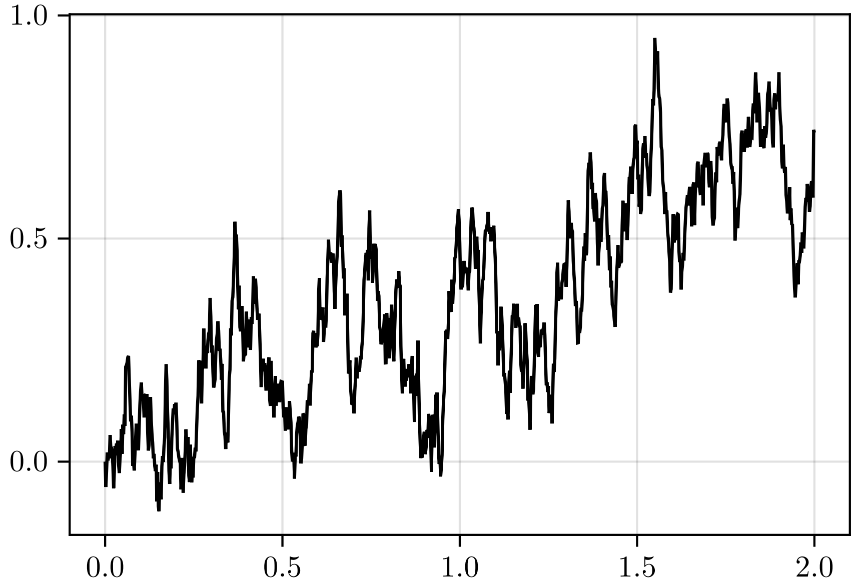
for $0 < t_1 < ... < t_n < ... < T$ such that $t_n = T/N$. From the definition of the Wiener process the term $B_{t_n} - B_{t_{n-1}}$ are Gaussian random variables with zero mean and variance $\sigma^2(t_n - t_{n-1})$, indicating that the solver samples from a distribution with those parameters at each iteration.

Example 8 (Simulated Wiener process) *In the Figure 4 the following stochastic differential equation is solved using the Euler-Maruyama scheme.*

$$dX_t = dB_t,$$

which represents the Wiener process.

Figure 4 – In continuous line a sample path of the simulated Wiener process. The heatmap shows the probability density function at each time step. The dashed line is the zero mean plus and minus the standard deviation.



Source: Preparated by the author.

There are other methods worth of mention but not used in this work as the stochastic versions of the Runge-Kutta method and of the Verlet method. We refer to Särkkä and Solin (2019) for a further discussion.

4 STOCHASTIC THERMODYNAMICS

In this chapter, the energetic quantities heat, work and entropy will be defined for the stochastic dynamics. We begin by defining the classical understanding about thermodynamics and them move to its stochastic description, giving place to the stochastic thermodynamics.

The theory tries to define the thermodynamics of small bodies in which the thermal fluctuations are non-negligible. One of the first to note such need was (EINSTEIN, 1905) while studying the Brownian motion. After, the emergence of the studies in biology and chemistry, in most cases involving mesoscopic systems, brought the attention to the applications of such theory (PELITI; PIGOLOTTI, 2021). Furthermore, it bridges the gap between the classical macroscopic thermodynamics and the microscopic quantum thermodynamics for systems between those scales (SEKIMOTO, 2010).

In this chapter we attempt to give a brief overview of the classical thermodynamics. Next, we try to apply the concepts of the classical thermodynamics to the stochastic dynamics. We will not go into details of the mathematical formalism, but we will try to give a general overview of the theory and its applications. The reader is referred to Seifert (2008), Broeck *et al.* (2013) and Peliti and Pigolotti (2021) for an introduction to the theory and its applications. Also, to the lectures of Lee (2022) for recent advances in the field. We apply the theory to the study of the thermodynamics of chemical reactions in order to make a parallel between the macroscopic and the mesoscopic scales (MOU *et al.*, 1986).

4.1 Review on classical thermodynamics

Thermodynamics is a branch of physics that deals with heat, work and their relation to energy, entropy and the physical properties of matter and radiation. The field started with the studies of Carnot (1824) of heat engines and their efficiency.

Primarily, the heat engines functioned using the steam generated by the boiling of water to expand inside some cylinder a moving piston, which in turn would move a wheel. Instead of wasting the steam after the expansion, the steam is cooled down and condensed to be then heated again. In this cycle, only heat is supplied to the system while mechanical work is produced by attaching any other machine to the wheel. Which substance to use in the engine, how to design the engine, how to make it more efficient, etc. were the questions that the field of thermodynamics aimed to answer(hence the name *therme*-heat and *dynamis*-power). As the result there were obtained statements about the functioning of such machines today known as the thermodynamic laws. In this work we attain to the first and second laws stated below.

Definition 4.1 (Thermodynamic laws)

- *1st law of thermodynamics: The energy exchanged through the boundaries of a system comprise heat and work;*

  $$\Delta E = Q + W$$

- *2nd law of thermodynamics: The variation of entropy in a system is non-negative.*

  $$\Delta S = \frac{Q_c}{T_c} - \frac{Q_h}{T_h} \geq 0$$

The 1$^{\text{st}}$ law deals with the conservation of energy in a closed system: the total energy in the system (denoted by E) amounts the heat put into it and the work done. It is sufficient to define the boundaries as any closed region of the space which we may study and that are not crossed by any flux of energy unless the work and heat. For instance, in the steam engine which transforms heat into work, the boundaries considered comprises the walls of the cylinder, the piston, the tubes connecting the cylinder to the boiler, the boiler itself as also the condenser, etc. It is conventional to consider the heat as a flow from/to a reservoir, which we call heat reservoir or heat bath. The work W is any energy that is applied to the system from an external agent. The reason for such distinction between energy in form of work and heat will be when we discuss the second law. What is left out of the boundary defined is conventionally denoted as universe where such an external agent realizing work may be in.

The 2$^{\text{nd}}$ law distinguishes work from heat as being a quantity of energy that can be totally used to perform transformations. For instance, the work done against the friction when pushing an object on a rugged surface. If the object comes to rest, all the work it was applied is converted to heat, which we could attest by touching the object and feeling it warm. On the other hand, we do not observe the object cooling down and putting itself into motion spontaneously. Although we could store in a reservoir the heat generated during the push to perform work later and then use this heat to set the object into motion again, this would require us to have some machinery allowing the heat to flow from a reservoir that is cold to the hot one. We can see this in the Example 9.

Example 9 *Let a system with temperature T_S^0. By applying some work to the system, e.g. pushing a mass in the rugged surface, the system goes to a temperature $T_S^1 > T_S^0$. Let a heat reservoir* *with temperature* $T_R^1 < T_S^1$ *be in contact with the system such that they exchange heat until both reach the equilibrium with the same temperature, say* $T_S^2 = T_R^2$ *the temperature of the system and the heat reservoir. As the physical properties of the system do not change, we require* $T_S^2 = T_S^0$ *such that all the heat gained by the system flows to the reservoir. Finally, we use the heat stored in the reservoir to perform work by assuming the existence of some engine which converts all the heat to work. Assume the engine extracts heat from the reservoir changing the temperature of the latter until it reaches* $T_R^1 < T_R^2$*. If the engine exists, the system should equilibrate with the reservoir at temperature* T_S^1*, which violates our assumption that* $T_R^1 \neq T_S^1$*.*

Note in the Example 9 that the contradiction occurs because we assume implicitly that the heat flows only from the hot reservoir to the cold one. This causes the system and the heat reservoir to equilibrate at a temperature which is not when $T_R^1 < T_S^1$ as in the initial state. The flow of heat from a cold body to the hot one is not observed in general, and due to that we can not treat both forms of energy, work and heat, in the same manner once we may use work to extract heat from the cold body. These machines are called refrigerators.

Another point to address with respect to the Example 9 is that the heat reservoirs have a heat capacity, i.e. a change of temperature with respect to the heat change, compared to the system. Otherwise, the temperature of the reservoir would not change when the system equilibrates with it (say, the system achieves thermal equilibrium). For instance, if we put the warmed object in the sea (considering the object very small compared to the sea), the temperature of the sea would not change as the heat capacity of the sea is much larger than the object. This is important to note since the heat reservoirs to be dealt from now on are idealized to have infinite heat capacity compare to the system, which resembles the difference in scale between the heat reservoir and the small systems we are going to deal with.

We understand from this that the first law of thermodynamics simply deals with energy conservation once any variation in the energy of the system is due to the energy variation in somewhere else, e.g. heat reservoir or universe. The second law of thermodynamics is not as simple. Clausius (1879) defines entropy in terms of the ability of exchanging heat between heat reservoirs with different temperatures, such that it is observed the heat to flow spontaneously from the warmer reservoir (with heat Q_h and temperature T_h) to the colder (Q_c and T_c). In other words, entropy deals with the reversibility of a process. As it is observed in the case of the cold and warm reservoirs, it is not expected the heat flow to "reverse" towards to the cold reservoir without some external action, we call such processes as irreversible.

Almost two centuries after Clausius, in the 1990s it was hypothesized that violations of the second law could happen for small-scaled systems (EVANS *et al.*, 1993) and a decade latter the first observations were made in laboratory experiments (WANG *et al.*, 2002). What happens is that the "classical" thermodynamics was first studied for systems (e.g. the heat engine). When the scale of the systems shrinks (i.e., mesoscopic and microscopic systems) the magnitude of the energy involved in the dynamics of such systems gets comparable to the energy from the thermal fluctuations (PELITI; PIGOLOTTI, 2021), for instance as in the case of the Brownian motion. A new thermodynamic theory is needed to comprise this "non-deterministic" behavior of the systems due to fluctuations.

The discipline of stochastic thermodynamics deals with systems below the macroscopic scale in which the thermal fluctuations are non-negligible to the behavior of the system. Another feature is that differently from classical thermodynamics which deals with systems in equilibrium, the stochastic thermodynamics studies some considered nonequilibrated systems and thus, it addresses the energetics of small systems via their dynamics. In the next section, we use the framework of chemical reactions to analyze the difference between macroscopic and mesoscopic systems.

4.2 Stochastic description of chemical reactions

In this work, we study the thermodynamics emerging from the fluctuations in chemical reactions. There are multiple ways to represent chemical reactions, each one focusing on a specific aspect of the process studied. In general, those representations differ themselves according to the level of detail of the reaction mechanism. For instance, one could model the solution of salt in water by using the Schrödinger equation. This would give a complete description of the system as the position of atoms and their electrons, but would not be practical once the dynamics of the electrons move at a timescale much faster than the atoms. Some models discard the electrons and focus on the atoms, which is known as Born-Oppenheimer approximation (BORN; OPPENHEIMER, 1927). At slower scales, we have the interaction between atoms forming molecules. As the level of agglomeration increases, the objects of study (e.g. the molecules, proteins, etc.) become more complex and heavy such that the faster dynamics of their smaller constituents can be neglected (KAMPEN, 1985). Thus, in order to be able yet to model the process occurring, we use abstract models which are able to capture the essential features of the system studied. In a chemical reaction where a huge number of molecules are involved (i.e. a number of the order of Avogadro's number), we do not describe the reacting components as the position of individual molecules but as the concentration of the species involved. In this work, we consider that our description is able to capture the mesoscopic dynamics of the system, i.e. the dynamics of the system at a scale where the thermal fluctuations are non-negligible, but fast enough to ignore the quantum effects of the atomic interactions.

This concept of timescale is important in the discussion to follow in order to define what the equilibrium is in the case of chemical reactions. In the Section 4.1 we saw the definition of equilibrium for a system in contact with a heat reservoir. But in the case of molecules of water, the equilibrium is relative to the timescale at which you observe them. For instance, the water molecules in a glass of water is at equilibrium with the room at a fixed temperature, but the water molecules are not at dynamic equilibrium once they are jiggling around due to the thermal fluctuation. Thus, we refer to equilibrium the variables (more specifically observables) which do not change with time at the timescale we are observing the system. This separation in timescales may occur also inside molecules whenever the conformal change in the atoms composing the molecule occurs at the same timescale as the system is observed. In this case, we may abstract our space of possible states of the system observed (i.e. the state-space) and consider states representing also the conformal changes of the molecule.

In the following sections, we review the macroscopic perspective of the theory of chemical reactions. Next, we introduce the stochastic approach to the theory leading to the description of the mesoscopic scale.

4.2.1 Macroscopic description

For the discussion that is going to follow, we consider a chemical reaction below.

Example 10 (Synthesis reaction) *Consider the reaction*

$$A + B \xrightarrow{k_{AB}} AB$$

where A and B are the reactants and AB is the product. The reaction is characterised by the rate constant k_{AB}.

The representation in the Example 10 denotes a simple example of a chemical reaction where two reactants, A and B, combine to form a product, AB. This may occur in several ways, e.g. a molecule of A may collide with a molecule of B and continue united forming the product AB.

The rate at which these collisions occur is given by the rate constant k_{AB}. It might be the case the molecules jiggles so much due to thermal fluctuations that both molecules dissociate and the molecules of A and B are formed again in the separated state. In this case, we should also consider this reverse reaction in the description of the reaction.

Example 11 ((Reversible) synthesis reaction) *Consider the reaction of the Example 10 but now with the reversible path*

$$A+B \underset{k_{AB}}{\overset{k_{AB}^{-}}{\rightleftharpoons}} AB,$$

where k_{AB}^{-} is the rate constant of the reverse reaction.

We now consider a (closed) system with a volume V containing the reactants A and B. The concentration of the species A, B and AB are denoted by $[A]$, $[B]$ and $[AB]$, and represent the number of molecules of each species per volume V. As the system is closed, the total number of species of the system is conserved, i.e.

$$[A]+[B]+[AB]=\text{constant}.$$

This system is considered to be at a temperature T and the spatial distribution of the molecules is homogeneous. The rate at which the concentration of the species of the Example 10 changes is given by the law of mass action.

Definition 4.2.1 (Law of mass action) *The rate at which the concentration of the species of a chemical reaction changes is proportional to the product of the concentrations of the reactants.*

By the Definition 4.2.1, the rate at which the concentration of the product AB changes equals

$$\frac{\mathrm{d}[AB]}{\mathrm{d}t}=k_{AB}[A][B]-k_{AB}^{-}[AB].$$

But notice that due to the conservation of the number of molecules, the rate at which the concentration of the reactants changes is the opposite of the rate at which the concentration of the product changes. Thus, we have

$$\frac{\mathrm{d}[A]}{\mathrm{d}t}+\frac{\mathrm{d}[B]}{\mathrm{d}t}=\frac{\mathrm{d}[AB]}{\mathrm{d}t}.$$

More specifically, as A only reacts in the presence of B, we can also conclude that $\mathrm{d}[A]/\mathrm{d}t=\mathrm{d}[B]/\mathrm{d}t$. When the change in the concentration of the species is ceased, we say the system is at chemical equilibrium and thus the derivatives go to zero. Such equilibrium is achieved whenever the concentration of the species get the ratio

$$\frac{[AB]}{[A][B]} = \frac{k_{AB}}{k_{AB}^{-}}.$$

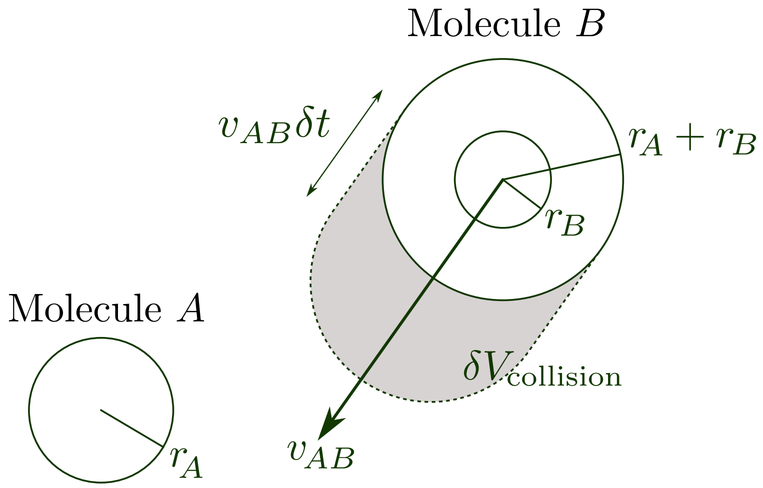
It must be pointed out that such equilibrium depends on the scale at which the system is observed. First, due to our assumption that the system is at a temperature $T \neq 0$, thus there is movement of the particles constituting it. This allows the species to react and dissociate, but at a rate which does not allow us to observe (macroscopically) any change in the concentration.

4.2.2 Stochastic description

One may notice that the concentration of the species is a continuous variable. This is a good approximation when the number of molecules is large. Also, by the law of large numbers (see Definition A.25), we may be led to believe that the concentration of the species is a deterministic variable. However, when one focuses on the processes occurring at the molecular level, the continuous assumption is not true simply by the fact that, if two molecules, say A and B from the Example 10, are in the same place at the same time, they will react and form the product AB. By the end of the collision, the number of molecules of A, B and AB will change by a discrete value. Furthermore, the fact that the molecules collide with each other is given by chance, and thus the number of molecules of each species is a stochastic variable. These two facts must be accounted when modelling in the mesoscopic scales where the number of molecules is not large enough to ignore the stochasticity of the system by considering its "average" behavior. In the discussion to follow, we review the approach by Gillespie (1976) to model systems in which the stochasticity in chemical reactions is accounted.

Considering that a chemical reaction occurs whenever the molecules of the species collide, we proceed by modeling from a probabilistic perspective how these collisions now considered to be random affect the number of molecules of each species. We assume that the system is composed by gas-like molecules, i.e. the molecules are not in contact with each other and the only interaction between them is due to the collisions. The system is also assumed to be in a heat reservoir at a fixed temperature T and the spatial distribution of the molecules is homogeneous in the volume V. In Figure 5 we consider the molecules to be spherical. The volume $\delta V_{\text{collision}}$ represents the volume in which the collision between the molecules s_A of species A and s_B of species B occurs. The chance of finding a molecule inside $\delta V_{\text{collision}}$ is

Figure 5 – The collision volume $\delta V_{\text{collision}}$ (gray area) is the volume in which the molecules s_A and s_B collide. The collision occurs because the molecules need to be far from at least the sum of their $r_{AB} = r_{s_A} + r_{s_B}$. The velocity of s_A relative to s_B is v_{AB}. If the center of the molecule s_A is found inside the volume $\delta V_{\text{collision}} = \pi r_{AB}^2 v_{AB} \delta t$ during the interval δt, the molecule s_B will have collided with molecule s_A.



Source: Adapted from Gillespie (1977).

given by the ratio $\delta V_{\text{collision}}/V$. If the molecules s_A and s_B move in random directions, the average of this ration is

$$\mathbb{E}\left[\frac{\delta V_{\text{collision}}}{V}\right] = \frac{\pi r_{AB}^2 \mathbb{E}\left[v_{AB}\right] \delta t}{V},$$

which is nothing more than the average probability that a single molecule of s_B collides with a molecule of s_A in the time interval δt. Obviously, if there are in the volume V N_A molecules of s_A and N_B molecules of s_B, the average number of collisions in the time interval δt is given by $N_A N_B V^{-1} \pi r_{AB}^2 \mathbb{E}\left[v_{AB}\right] \delta t$. This characterises the system by a probability of collision per unit time δt. If such collisions result in chemical alterations called reactive collisions, one finds the reaction probability rate. To connect the aforementioned discussion with the chemical reaction rate, assume that the species s_A and s_B react via a reaction, say r_m. Let $c_m \mathrm{d}t$ be the probability of a single reactive collision between the molecules s_A and s_B in the time interval $\mathrm{d}t$. Gillespie (1976) shows that if the reaction r_m occurs with a rate k_m, the reaction probability rate for a reaction of the type of the one in Example 10 is related to the reaction rate k_m by $V c_m = k_m$. Such connection is specific to the type of the reaction r_m. Another example is given below.

Example 12 (Dimerisation reaction (GILLESPIE, 1976)) *Consider the reaction*

$$2A \xrightarrow{k_{AA}} A_2$$

where A_2 *is the dimer of* A*. The reaction is characterised by the rate constant* k_{AA}*. The reaction probability rate is given by* $c_{AA} = 2V^{-1}k_{AA}$*.*

By having the probability rate, we may describe the temporal evolution of the system with a Markov process. In the following, we will describe the Markov process in terms of the master equation which describes the time evolution of the probability distribution of the system. Also, we will describe the system in terms of the stochastic trajectory which is a sequence of events that occur in the system.

4.2.2.1 Chemical master equation

In this section we are going to deal with a general case of the types of chemical reactions, called chemical reaction networks (CRNs), which are a general framework to describe and analyze chemical reactions involving multiple species and reactions. A CRN is defined by a set of species $\{s_1, s_2, \ldots, s_N\}$ and a set of reactions $\{r_1, r_2, \ldots, r_M\}$ (GILLESPIE, 1976; SCHMIEDL; SEIFERT, 2007). Each reaction r_m is represented as

$$\sum_{i=1}^{N} \nu_{im}^{-} s_i \xrightarrow{k_m} \sum_{i=1}^{N} \nu_{im}^{+} s_i,$$

where ν_{im}^{-} and ν_{im}^{+} are the stoichiometric coefficients of species S_i as reactants and products, respectively, in reaction r_m, and k_m is the rate constant of the reaction. The net change in the number of molecules of species S_i due to reaction r_m is given by the stoichiometric difference

$$\Delta\nu_{im} = \nu_{im}^{+} - \nu_{im}^{-}.$$

The state of the system at time t is described by the vector $X_t \in \mathbb{Z}_{+}^{N}$, where $X_t = [(X_1)_t, (X_2)_t, \ldots, (X_N)_t]$ with $(X_i)_t$ being the number of molecules of species S_i at time t. The time evolution of the probability distribution $P(X_t)$ is governed by the solution of chemical master equation (CME)

$$\frac{\partial P(X_t)}{\partial t} = \sum_{m=1}^{M} \left[a_m(X_t - \Delta\nu_m)P(X_t - \Delta\nu_m) - a_m(X_t)P(X_t)\right], \tag{4.1}$$

where $a_m(X)$ is called the propensity function of reaction r_m, defined as the probability per unit time that reaction r_m occurs given the state X, and the stoichiometric vector is $\Delta\nu_m =$

$[\Delta\nu_{1m},\ldots,\Delta\nu_{Nm}]$. For mass-action kinetics, the propensity function is

$$a_m(X) = c_m \prod_{i=1}^{N} \binom{X_i}{\nu_{im}^-} = c_m \prod_{i=1}^{N} \frac{X_i!}{\nu_{im}^-!(X_i-\nu_{im}^-)!},$$

To the complete derivation of the Equation (4.1), we refer to Gillespie (1976), although for the discussion to follow it is important to make sense of the terms in it. The propensity $a_m(X-\Delta\nu_m)$ quantifies the rate at which the probability of the system changes from the state $X-\Delta\nu_m$ to the state X given that the probability of the system to be in the state $X-\Delta\nu_m$ at time t is $P(X_t-\Delta\nu_m)$. In the opposite direction, the system may leave the state X to the state $X-\Delta\nu_m$ at probability rate $a_m(X)$. The difference between these two terms gives the net change in the probability of the system to be in the state X at time t. To put in terms of the common notation in the literature for master equations of Markov processes (see Equation (A.1) and Kampen (1992)),

$$\frac{\partial P(X_t)}{\partial t} = \sum_{X'_t} \left[W(X_t,X'_t)P(X'_t) - W(X'_t,X_t)P(X_t)\right],$$

where $W(X,X')$ is the transition rate from state X' to state X. The chemical formulation simplifies the states X' to $X-\Delta\nu_m$, i.e. the states are only the ones that are reachable by the reactions.

Solving Equation (4.1) is a difficult task in most of the cases. Jahnke and Huisinga (2007) present a method to solve the CME analytically for the case of monomolecular reactions, i.e. reactions that involve only one species. For other cases, one may resort to numerical methods, although these methods also suffer from the high dimensionality of representation of the solution $P(X_t)$ when the total number of molecules is large. For instance, a CRN containing 3 species with a 10^3 of molecules each results in a space of 10^9 possible states. Numerical solvers often resort to approximations to deal with this high dimensional state-space. To number a few: Munsky and Khammash (2006a) propose the Finite State Projection (FSP) method which truncates the state-space of possible states X in order to keep the solution memory-efficient. Dolgov and Khoromskij (2015) propose a specific construction of the states of the system using the reduction of dimensionality from tensor techniques.

4.2.2.2 Stochastic simulation algorithm

One may use the stochastic simulation algorithm (SSA) which is a Monte Carlo method to simulate the time evolution of the system proposed by Gillespie (1976). The SSA is a discrete event simulation algorithm that generates a stochastic trajectory of the system by simulating the occurrence of reactions in a CRN. The called Gillespie algorithm runs by generating a sequence of random numbers to determine the time intervals between reactions and the specific reactions that occur.

Definition 4.2.2 (Gillespie algorithm (GILLESPIE, 1976)) *The Gillespie algorithm is a stochastic simulation algorithm that generates a stochastic trajectory of the system by simulating the occurrence of reactions in a CRN. Let X_t be the state of the number of molecules at the time t. The algorithm proceeds as follows:*

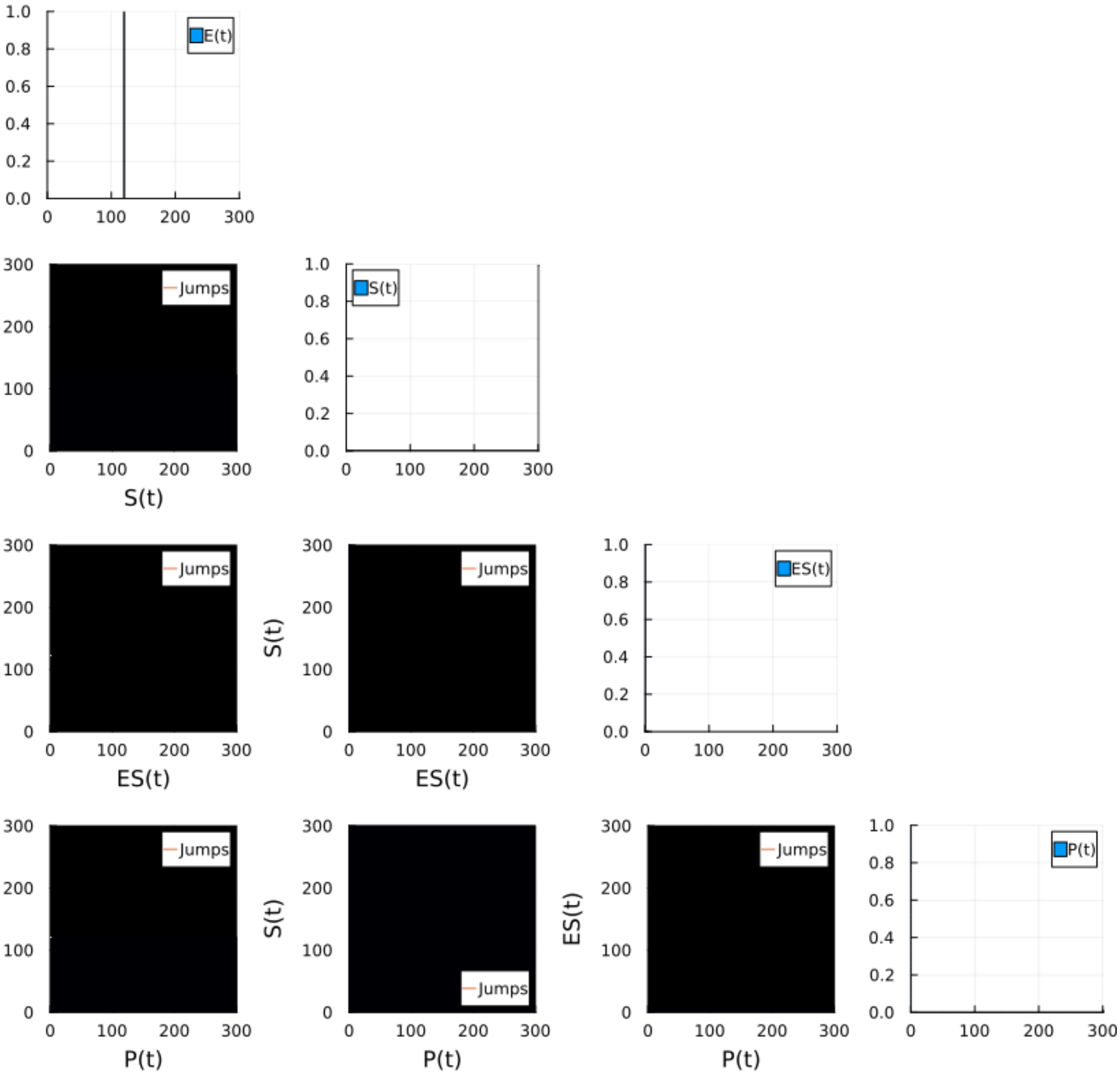
1. *Initialize the system with an initial state X_0 and time $t_0 = 0$.*
2. *Calculate the propensity functions $a_m(X_t)$ for all reactions r_m.*
3. *Generate two random numbers U_1 and U_2 uniformly distributed in the interval $(0,1)$.*
4. *Calculate the time increment $\Delta t = -\frac{1}{\sum_{m=1}^{M} a_m(X_t)} \ln(U_1)$.*
5. *Determine the next reaction to occur by finding the smallest integer j such that*
$$\sum_{m=1}^{j-1} a_m(X_t) < U_2 \sum_{m=1}^{M} a_m(X_t) \leq \sum_{m=1}^{j} a_m(X_t).$$
6. *Update the state of the system: $X_{t+\Delta t} = X_t + \Delta \nu_j$.*
7. *Update the time: $t = t + \Delta t$.*
8. *Repeat steps 2-7 until a stopping criterion is met (e.g., reaching a certain time or number of reactions).*

4.2.2.3 Chemical Langevin equation

Despite the handful of methods to solve the CME, describing the reacting system in terms of discrete states is yet computationally expensive when the number of molecules is large. It can be shown that under certain conditions, a system whose probability distribution is described by the CME can be approximated by a set of stochastic differential equations known as the chemical Langevin equation (CLE), named after the similarity with the equation of the Langevin dynamics (GILLESPIE, 2000).

Following Horowitz (2015), we consider a time-interval Δt such that many reactions occur, i.e. $a_m(X_t)\Delta t \gg 1$. At the same time, we consider the time-interval Δt to be small enough such that the propensity functions do not change significantly, i.e. $a_m(X_{t+\Delta t}) \approx a_m(X_t)$. Such assumptions are satisfied in general for systems with a large population of molecules. The

Figure 6 – Simulation of the Michaelis-Menten reaction mechanism (see Section 6.1) at some time instant t. As there are four reacting species involved (S, E, ES, P), the probability mass function describing the probability of a given number of molecules of each species is four-dimensional, which is difficult to represent graphically. Therefore, we opted for plotting just the two-by-two marginals of the species represented in the axes. Over the heatmap, a draw of a single realization of the stochastic simulation algorithm. The initial state is set at $(120, 0, 0, 301)$, and the rate constants are $k_{E+S \to ES} = 1.66 \times 10^{-3}$, $k_{E+S \to ES} = 1.66 \times 10^{-3}$, $k_{ES \to E+S} = 1.0 \times 10^{-4}$, $k_{ES \to E+P} = 0.1$ and $k_{E+P \to ES} = 1.0 \times 10^{-4}$. The CME is solved using Euler's method, concentrating all the probability in the single initial state previously mentioned. Notice that in the $E \times ES$ marginal (at the 3rd-row and 1st-column) the marginal is a flat distribution completely concentrated on planes parallel to the $E = -ES$ line. This is because the total number of enzymes E (i.e. $E + ES$) is fixed.



Source: Prepared by the author.

change in the number of molecules of all species ΔX_t in the time interval Δt is given by

$$\Delta X_t = \sum_{m=1}^{M} \Delta \nu_m \left(\Delta (N_m)_t\right),$$

where $\Delta(N_m)_t$ is a Poisson-distributed random variable representing the number of times the reaction r_m occurs in the time interval Δt. This formulation allows us to write the stochastic process resulted from the Gillespie algorithm in the differential form. Taking the limit of Equation (4.2) as $\Delta t \to \mathrm{d}t$, we have

$$\mathrm{d}X_t = \sum_{m=1}^{M} \Delta\nu_m \left(\mathrm{d}(N_m)_t\right), \tag{4.2}$$

where the increments $\mathrm{d}(N_m)_t$ are a Poisson-distributed random variables with mean and variance equal to $a_m(X_t)\mathrm{d}t$. As it was assumed the number of reactions to be large, we can approximate the Poisson distribution by a Gaussian distribution with mean and variance equal to $a_m(X_t)$. Thus, we have in the limit $\Delta t \to \mathrm{d}t$ that

$$\mathrm{d}X_t = \sum_{m=1}^{M} \Delta\nu_m \left(a_m(X_t)\mathrm{d}t + \sqrt{a_m(X_t)}\mathrm{d}(B_m)_t\right),$$

where $\mathrm{d}(B_m)_t$ is a Wiener process for each m-th reaction and independent of each other.

4.3 Energetics for Markov processes

In the introduction of this chapter we discussed the classical approach to thermodynamics. However, the classical approach is not able to describe the energetics of small systems where the thermal fluctuations are non-negligible. Nevertheless, we are still able to define the same thermodynamic quantities for such systems. To illustrate this, in this section we define the energetics of a Markov process applied in the context of chemical reactions.

Whenever a reaction occurs, some energetic quantity is exchanged with the surroundings. Such quantity we identify as heat, following our classical approach to thermodynamics. In order to quantify the heat exchanged, we modify our CRN description such that if we have defined a set of reactions $\{r_1, r_2, \dots, r_M\}$, there will be a set of reversed reactions $\{r_{1^-}, r_{2^-}, \dots, r_{M^-}\}$ such that the reaction r_{m^-} is the reverse of the reaction r_m. The reversed reaction r_{m^-} varies the state of the system in the opposite direction of the reaction r_m, thus we have that $\Delta\nu_{m^-} = -\Delta\nu_m$. If allow our reacting system to evolve, it will reach an equilibrium. In macroscopic terms, the concentration of reacting species will not change. But this only means that the average of the number of molecules of each species does not change. Let us assume that we have a reacting system that reached the equilibrium. We are going to identify the probability of the system to be in the equilibrium state as $P_{\mathrm{eq}}(X)$. By the Maxwell-Boltzmann statistics

(TOLMAN, 1979), this probability equals

$$P_{\text{eq}}(X) = \frac{1}{Z} e^{-\beta\varepsilon(X)},$$

where $Z = \sum_X e^{-\beta\varepsilon(X)}$ is the partition function, $\beta = 1/(k_B T)$ is the inverse temperature and $\varepsilon(X)$ is the energy of the system in the state X. If we have another state X' also in equilibrium, both relate to each other by the

$$\frac{P_{\text{eq}}(X)}{P_{\text{eq}}(X')} = e^{-\beta(\varepsilon(X)-\varepsilon(X'))},$$

which means that the ration between the probability of the system being in each state (while in equilibrium) depends only on the energy difference between the two states. Furthermore, the equilibrium property of the system implies that we can use the detailed balance condition defined in Equation (A.2) to write the relation

$$\frac{P_{\text{eq}}(X)}{P_{\text{eq}}(X')} = \frac{W(X,X')}{W(X',X)} = e^{-\beta(\varepsilon(X)-\varepsilon(X'))}.$$

From a tabletop perspective, the detailed balance is not able to identify uniquely the transition rates $W(X,X')$ and $W(X',X)$. However, it is able to identify the ratio between them if the energy difference between the states and the temperature are known.

Back to the system in equilibrium, we identify that when the system jumps between the two states, there is an energy unbalance we need to account for. We assume that, as our system is in contact with a heat reservoir, the energy unbalance is compensated by the heat exchanged with the reservoir.

The main result of the theory is to identify this quantity also when the system is out of equilibrium by its interactions with the heat reservoir, the latter remaining in equilibrium. We follow Esposito and Broeck (2010) and analyze the entropy of the master equation of the system using the definition of entropy given by Shannon (1948). Consider the entropy

$$s_t = -k_B \sum_X P(X_t) \ln P(X_t),$$

where $P(X_t)$ is the probability of the system to be in the state X at time t. The master equation leads us to write the time derivative of s_t

$$\begin{aligned}\dot{s}_t &= -k_B \sum_X \left[\frac{\partial P(X_t)}{\partial t} \ln P(X_t) + P(X_t) \frac{\partial}{\partial t} \ln P(X_t) \right], \\ &= -k_B \sum_X \left[\frac{\partial P(X_t)}{\partial t} \ln P(X_t) + P(X_t) \frac{1}{P(X_t)} \frac{\partial P(X_t)}{\partial t} \right].\end{aligned}$$

The second term in the right-hand side vanishes due the fact that $\sum_X \partial P(X_t)/\partial t = 0$, i.e. if there is conservation of probability the rate of change of probability must be such that when we sum over all the states, the result is zero leading to no "leak" or "gain" of probability. Next, we can use the master equation to find

$$\begin{aligned}
\dot{s}(t) &= -k_B \sum_{X_t} \left[\left(\sum_{X_t'} \left[W(X_t, X_t')P(X_t') - W(X_t', X_t)P(X_t) \right] \right) \ln P(X_t) \right] \\
&= \frac{k_B}{2} \sum_{X_t \neq X_t'} \left[W(X_t, X_t')P(X_t') - W(X_t', X_t)P(X_t) \right] \ln \frac{P(X_t')}{P(X_t)}, \\
&= \frac{k_B}{2} \sum_{X_t \neq X_t'} \left[W(X_t, X_t')P(X_t') - W(X_t', X_t)P(X_t) \right] \ln \frac{W(X_t, X_t')}{W(X_t', X_t)} \\
&+ \frac{k_B}{2} \sum_{X_t \neq X_t'} \left[W(X_t, X_t')P(X_t') - W(X_t', X_t)P(X_t) \right] \ln \frac{W(X_t, X_t')P(X_t')}{W(X_t', X_t)P(X_t)}, \\
&:= \dot{s}_e(t) + \dot{s}_i(t),
\end{aligned}$$

where $\dot{s}_e(t)$ is the average entropy flux rate and $\dot{s}_i(t)$ is the average entropy production rate[1]. The reason behind the naming will become clear in the following. We assume that our system is in contact with a heat reservoir at a fixed temperature T, thus the system is in (thermal) equilibrium with the reservoir. It is also assumed that if any reaction occurs, it is due to the energy exchanged with the reservoir as heat implying that the rate at which the system "switches" between the states X and X' relates to the heat exchanged with the reservoir by the Boltzmann factor. This allows us to identify

$$\begin{aligned}
\dot{s}_e(t) &= -\frac{1}{2T} \sum_{X_t \neq X_t'} \left[W(X_t, X_t')P(X_t') - W(X_t', X_t)P(X_t) \right] \left(\varepsilon(X_t) - \varepsilon(X_t') \right), \\
&= -\frac{1}{T} \sum_{X_t \neq X_t'} \left[W(X_t, X_t')P(X_t') \right] \left(\varepsilon(X_t) - \varepsilon(X_t') \right), \\
&= -\frac{1}{T} \sum_{X_t} \frac{\partial P(X_t)}{\partial t} \varepsilon(X_t) \\
&:= -\frac{\dot{q}}{T}
\end{aligned}$$

where $\dot{q}$ is the rate of heat exchanged with the reservoir. This resembles the second law of thermodynamics. If the system is driven slowly enough, is will keep the equilibrium at the same time it changes the energy of their states. By this count we have the energy delivered to

---

[1] The subscripts are chosen in accordance with bibliographic reference, the letters e and i representing "external" entropy change and "internal" entropy change, respectively.

the system as work, which average rate values

$$\dot{w} := P(X_t)\sum_{X_t}\frac{\partial\varepsilon(X_t)}{\partial t}.$$

We recover then the first law of thermodynamics as

$$\dot{\varepsilon} = \dot{w} + \dot{q}.$$

Notice that the first law is recovered as an average value once we have that

$$\varepsilon = \sum_{X_t} P(X_t)\varepsilon(X_t).$$

It remains for us to identify the average entropy production rate. From non-equilibrium thermodynamics (GROOT; MAZUR, 2013), whenever the entropy of the system changes, this change is the sum of the entropy supplied to the system by its surroundings and the entropy produced by the system. When the system is in equilibrium, the entropy produced by the system is zero. We observe this by the definition of $\dot{s}_i(t)$, that is zero when detailed balance is satisfied.

4.3.0.1 Stochastic entropy

The energetic quantities aforementioned represent average rates of the evolution of the system, but a pathwise description can also be obtained if, for instance we consider the average entropy rate as an average in $\mathbf{X}_t$ of Peliti and Pigolotti (2021)

$$S_t = -k_B \ln P(\mathbf{X}_t),$$

evaluated for the state of the system $\mathbf{X}_t$ at the instant t, considering we know the probability $P(\mathbf{X}_t)$ of the system to be at that state. Thus, S_t is a stochastic process representing the nonequilibrium entropy of a realization of the system. From non-equilibrium thermodynamics, the total variation of entropy S_t^{tot} is given by (PELITI; PIGOLOTTI, 2021; GROOT; MAZUR, 2013)

$$S_t^{\text{tot}} = \Delta S_t^i + S_t^e$$

where $\Delta S^i = S_t - S_0$ is the entropy production in a single trajectory given by the internal (with respect to the system) variation of entropy, and S^i is the entropy flux given externally by a heat reservoir, to the system. Thus, the entropy flux evaluates

$$S^e = \frac{1}{T}Q(\mathbf{X}_t),$$

where T is the classical thermodynamic description of the temperature of the heat reservoir, and $Q(\mathbf{X}_t)$ is the heat given by it to the system when the latter is found in the state $\mathbf{X}_t$.

To substantiate the reader unfamiliar with deep understanding with classical thermodynamics, the quantity $Q(\mathbf{X}_t)$ is taken from a well known system, as the heat reservoir, which is in equilibrium and thus all the classical thermodynamic description is available. We conclude then that such stochastic quantities mentioned above can not be fully described if there is no equilibrium system as a heat reservoir interacting with the system. Then the existence of such reservoir is assumed from the beginning.x

4.3.0.2 Thermodynamic consistency

In Section 4.2.2.3 we presented the CLE approximation of the CME. If the differential equations for both the CLE and the CME describe the same system, we should be able to identify the same thermodynamic quantities for both cases and expect them to be equal. This is not the case for the entropy production rate of the CLE and the CME, as shown by Horowitz (2015), unless the system is in equilibrium. This occurs because the CLE does not provide a precise description about which reactions took place, and then we loose information about the entropy production rate.

5 GAUSSIAN PROCESS STATE-SPACE MODEL

In this chapter the Gaussian process state-space model is revisited following the approach of Sarkka and Hartikainen (2012) using the Bayesian inference framework. The model is justified when the data is a time-series which allows us to make use of assumptions as the Markovian property to reduce the computational burden for processing the estimation. Also, the use of the state-space approach to model probabilistically solutions to ordinary differential equations is reviewed.

5.1 Historical overview in signal processing

Another of the contributions from Wiener (1949) is the Wiener filter named after him. He proposed to estimate a signal which is corrupted by (additive) Gaussian noise requiring the minimal error (in the mean squared sense) between the signal and the corrupted measurements. Inspired by this idea, Kalman (1960) proposes the filter named after him. Differently from the Wiener filter, the Kalman filter is able to approach also the dynamics (if given) of the process to be estimated due to its state-space representation which allows ones to include the information about the dynamics of the signal.

More recently, and with the growth of the field of machine learning, the Gaussian process model emerged as flexible tool for signal estimation, by the time more advantageous than neural networks due to the lack of expressivity of the models (RASMUSSEN; WILLIAMS, 2005). With the increase of the number of data, the shortcomings of the Gaussian process started to limit its progress and a variety of alternatives to approach the high-dimensional data were proposed (for a survey, see Liu *et al.* (2020)).

5.2 Review on Bayesian inference

The Bayesian inference framework lies on the Bayes' rule to solve the called inverse problems (TARANTOLA, 2005). The framework is summarized by the following: a prior probability density $p_X(x)$ is assumed to model the uncertainty one has about the value of the variable X. To decrease the "uncertainty" about X, a set of measurements are done in a variable Y that has its probability dependent on X via the conditional density $p_{Y|X}(x,y)$, also known as likelihood. This gives us the posterior density using the law of total probability and the conditional probability

$$p_{X|Y}(x,y) = \frac{p_{Y|X}(x,y)p_X(x)}{p_Y(y)} = \frac{p_{Y|X}(x,y)p_X(x)}{\int p_{Y|X}(x,y)p_X(x)\mathrm{d}x}.$$

The integral in the denominator in general can not be obtained analytically unless specific cases (MACKAY, 2003). We showcase in the next section how to obtain the posterior density of a Gaussian process.

5.3 Gaussian process regression

Given a dataset $\{(\mathbf{x}_n, Y_n)\}_{n=1}^{N}$, where $\mathbf{x}_n \in \mathbb{R}^{D_x}$ are input vectors and $Y_n \in \mathbb{R}^{D_y}$ are corresponding observations, Gaussian process regression aims to "learn" a function $f(\mathbf{x})$ that maps inputs to outputs. By learn, we mean that, given the set of inputs, we apply the Bayes' rule to obtain the posterior distribution of the function $f(\mathbf{x})$ given the observations. The Gaussian process regression is a non-parametric Bayesian approach to regression, where the function f is assumed to be drawn from a Gaussian process as a prior, i.e. $f(\mathbf{x}_n) = X_n$ where X_n are Gaussian distributed random variables. The multivariate marginal density of this process is assumed to be

$$p_{X_1,\ldots,X_N}(x_1,\ldots,x_N) = \mathcal{N}(\mathbf{0},\mathbf{K}),$$

where $\mathbf{0} \in \mathbb{R}^N$ is a vector whose entries equal 0 and $\mathbf{K}$ is the covariance matrix of the process measuring the covariance between $X_n = f(\mathbf{x}_n)$ and $X_m = f(\mathbf{x}_m)$.

The covariance function $k(\mathbf{x}_n, \mathbf{x}_m)$ is a function of the inputs $\mathbf{x}_n$ and $\mathbf{x}_m$ and is assumed to be positive-definite. The covariance function encodes the prior assumptions about the smoothness, periodicity, and other properties of the function f.

In a general setting, the observations y_n are assumed to be corrupted by Gaussian noise, i.e. $Y_n = X_n + \sigma_y \xi_n$, where ξ_n is a Gaussian random variable with zero mean and variance equal to 1. The likelihood of the observations given the function values is given by

$$p_{Y_n|X_n}(y_n, x_n) = \mathcal{N}(y_n \mid X_n, \sigma_y^2).$$

We assume that the observations are independent given the function values, so the joint likelihood of the observations is given by

$$p_{Y_1,\ldots,Y_N|X_1,\ldots,X_N}(y_1,\ldots y_N, x_1,\ldots,x_N) = \prod_{n=1}^{N} p_{Y_n|X_n}(y_n, x_n).$$

Our aim is to obtain the posterior density of the function values given the observations, i.e. $p_{X_1,\ldots,X_N|Y_1,\ldots,Y_N}(x_1,\ldots,x_N,y_1,\ldots,y_N)$. By applying the Bayes' rule we obtain the posterior density as

$$p_{X_1,\ldots,X_N|Y_1,\ldots,Y_N}(x_1,\ldots,x_N,y_1,\ldots,y_N) = \\ \frac{\prod_{n=1}^{N} p_{Y_n|X_n}(y_n,x_n)p_{X_1,\ldots,X_N}(x_1,\ldots,x_N)}{\int_{\mathbb{R}^{D_x}}\cdots\int_{\mathbb{R}^{D_x}}\prod_{n=1}^{N} p_{Y_n|X_n}(y_n,x_n)p_{X_1,\ldots,X_N}(x_1,\ldots,x_N)\mathrm{d}x_1\cdots\mathrm{d}x_N}.$$

One may also be interested in the posterior density of the function values at inputs $\mathbf{x}_*$ that were not observed, i.e. $p_{X_*|Y_1,\ldots,Y_N}(\cdot)$ for short. Notice that the random variables X_* and X_n represent the values of the same Gaussian process but at different inputs. Thus, we can write the wanted posterior density as the following marginalization

$$p_{X_*|Y_1,\ldots,Y_N}(x_*,y_1,\ldots,y_N) = \\ \int_{\mathbb{R}^{D_x}}\cdots\int_{\mathbb{R}^{D_x}}\big[p_{X_*|X_1,\ldots,X_N}(x_*,x_1,\ldots,x_N) \\ p_{X_1,\ldots,X_N|Y_1,\ldots,Y_N}(x_1,\ldots,x_N,y_1,\ldots,y_N)\big]\,\mathrm{d}x_1\cdots\mathrm{d}x_N.$$

Those integrals are tractable if we keep the assumption of a Gaussian process and Gaussian observations due to the Gaussian noise. Otherwise, one might run into trouble when trying to evaluate those posteriors analytically.

On the Gaussian case, we have the statistics of the posterior. The solution analytically obtained is also a Gaussian distribution whose sufficient parameters to describe it are given by (RASMUSSEN; WILLIAMS, 2005)

$$\begin{aligned} p_{X|Y}(x_1,\ldots x_N,y_1\ldots,y_N) &= \mathcal{N}(X \mid \mathbb{E}[X],\mathrm{Var}[X]) \\ \mathbb{E}[X] &= \mathbf{K}\left(\mathbf{K}+\sigma_y^2\mathbf{I}\right)^{-1}Y, \\ \mathrm{Var}[X] &= \mathbf{K}+\mathbf{K}^\top\left(\mathbf{K}+\sigma_y^2\mathbf{I}\right)^{-1}\mathbf{K} \end{aligned}$$

where $X=[X_1,\ldots,X_N]^\top$, $Y=[Y_1,\ldots,Y_N]^\top$ and the matrix $\mathbf{K}$ has entries $\mathbf{K}_{i,j}=\mathrm{Cov}[X_i,X_j]$.

The main drawback of this method is that the matrix inversion required to obtain the parameters of the posterior distribution has requires $O(N^3)$ for N observations, which is not scalable with the increase of number observations, considering we are just evaluating the correlations in time. If the correlations are also taken in other dimensions as the dimensions of X_t, this complexity scales up with $N\cdot\max(D_x,D_y)$, in the worst case. In the next section we define the problem in state-space form to overcome the cubic cost of the posterior computation.

5.3.0.1 Filtering and smoothing

The posterior obtained in the previous section brings more information than what might be necessary in real-time applications. This is the case for the called filtering problems in which one needs just a single estimate of X_N given all the measurements $Y_1, ..., Y_N$. In such cases, the covariance of X_N with another X_n for $n = 1, ..., N-1$ may not be important, but only its variance. With such requirements, we may assume the process X_n to be Markovian, leading to the simplification

$$p_{X_N|X_1,...,X_{N-1}}(\cdot) = p_{X_N|X_{N-1}}(x_N, x_{N-1}).$$

Using this fact together with the Bayes' rule and the Chapman-Kolmogorov gives the called Bayesian filtering equations (SÄRKKÄ; SVENSSON, 2023a).

Definition 5.3.1 (Bayesian filtering equations (SÄRKKÄ; SVENSSON, 2023a)) *The marginal posterior density* $p_{X_n|Y_1,...,Y_n}(\cdot)$ *is obtained recursively for* $n = 1, ..., N$ *through the following two equations*[1]

1. *The predictive density of* X_n *is computed through the Chapman-Kolmogorov equation*

$$p_{X_n|Y_1,...,Y_{n-1}} = \int_{\mathbb{R}^d} p_{X_n|X_{n-1}} p_{X_{n-1}|Y_1,...,Y_{n-1}} dx_{n-1}; \qquad \text{(Prediction equation)}$$

2. *The Bayes'rule updates the predictive density of* X_n *on the light of the "new" measurement* Y_n

$$p_{X_n|Y_1,...,Y_n} = \frac{p_{Y_n|X_{n-1}} p_{X_n|Y_1,...,Y_{n-1}}}{\displaystyle\int_{\mathbb{R}^d} p_{Y_n|X_{n-1}} p_{X_n|Y_1,...,Y_{n-1}} dx_n}; \qquad \text{(Update equation)}$$

assuming that for $n = 1$ *we have* $p_{X_1|Y_0} := p_{X_1}$.

A celebrated example of implementation of those equations is the Kalman filter, which is illustrated in the following example.

Example 13 (Kalman filter (SÄRKKÄ; SVENSSON, 2023a)) *The transition and the likelihood probability densities observations for a linear system as the one in Equation* (2.11) *with* $\mathbf{B} = \mathbf{0}$ *are, respectively*

$$p_{X_n|X_{n-1}}(x_n, x_{n-1}) = \mathcal{N}(X_n \mid \mathbf{A}X_{n-1}, \mathbf{Q}_{n-1}),$$
$$p_{Y_n|X_n}(y_n, x_n) = \mathcal{N}(Y_n \mid \mathbf{H}X_n, \mathbf{R}_n),$$

---

[1] The respective functions depend on the variables $x_1, ..., y_1, ...$, but were omitted for readability.

where $\mathbf{A}$ *is the dynamics matrix as in Equation* (2.11a)*,* $\mathbf{H}$ *is the observation matrix as in Equation* (2.11b)*,* $\mathbf{R}_n$ *is the covariance of the observations* Y_n *and* $\mathbf{Q}_{n-1}$ *is the covariance of the states* X_{n-1}*. The Kalman filter algorithm obtains iteratively the parameters of the posterior*

$$p_{X_n|Y_1,\dots,Y_n}(x_n) = \mathcal{N}(X_n \mid \mathbf{m}_n, \mathbf{P}_n)$$

via

$$\begin{aligned}
\mathbf{m}_n^p &= \mathbf{A}_{n-1}\mathbf{m}_{n-1},\\
\mathbf{P}_n^p &= \mathbf{A}_{n-1}\mathbf{P}_{n-1}\mathbf{A}_{n-1}^\top + \mathbf{Q}_{n-1},\\
\mathbf{S}_n &= \mathbf{H}_n\mathbf{P}_n^p\mathbf{H}_n^\top + \mathbf{R}_n,\\
\mathbf{m}_n &= \mathbf{m}_n^p + \mathbf{K}_n(Y_n - \mathbf{H}\mathbf{m}_n^p),\\
\mathbf{P}_n &= \mathbf{P}_n^p - \mathbf{K}_n\mathbf{S}_n\mathbf{K}_n^\top.
\end{aligned}$$

The function $p_{X_n|Y_1,\dots,Y_n}(x_n)$ obtained in the Example 13 is only on x_n given that the parameters $\mathbf{m}_n$ and $\mathbf{P}_n$ already comprise the dependence on $Y_1,\dots,Y_n$ via the statistics obtained recursively.

The computational burden to evaluate the matrices in the Kalman filter is still a single matrix inversion, but now only with $O\left(N \cdot \max(N_x,N_y)^3\right)$. Thus, if the process X_n as well as the observations Y_n are low dimensional, the problem scales linearly with the number of observations (SARKKA *et al.*, 2013).

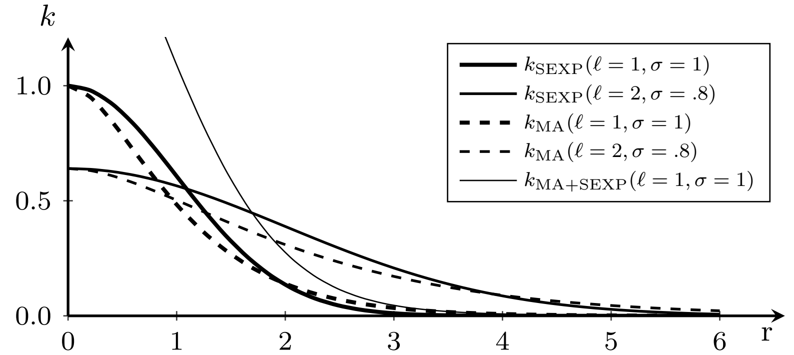
5.3.1 Gaussian process kernels

One of the main characters in the Gaussian process regression is the covariance function, or Gaussian process kernels, that models the correlations between different points of the process X_n. This because the Gaussian process is said to be a non-parametric model, once there are no parameters over which the inference is done to then, infer the latent variables, say to obtain the posterior in X_n. Besides that, the modeler is prone to choose a covariance function to obtain $\mathbf{K}$ for the prior density $\mathcal{N}(\mathbf{0},\mathbf{K})$. Some examples are given in the Figure 7 whose expressions of the functions are given in the Examples 14 and 15.

Example 14 *Squared exponential kernel (RASMUSSEN; WILLIAMS, 2005) The squared exponential covariance function has the form*

$$k_{SEXP}(r) = \sigma^2 \exp\left(\frac{r^2}{2\ell^2}\right),$$

Figure 7 – Examples of covariance functions. The Matérn function has a faster decay with the difference $\tau = |n-m|$ between the variables X_n and X_m, than the Squared Exponential function for the same set of parameters values.



Source: Prepared by the author.

with the parameter ℓ defining the characteristic length-scale and σ the intensity.

A property of the squared exponential in Example 14 that will be useful latter is that this kernel is infinitely mean squared differentiable.

Example 15 *Matérn kernel (RASMUSSEN; WILLIAMS, 2005) The class of Matérn covariance functions has the form*

$$\sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{r}{\rho} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{r}{\rho} \right),$$

with the parameter ℓ defining the characteristic length-scale, σ the intensity, $\Gamma(\nu)$ the Gamma function and K_ν the modified Bessel function. For $\nu = 3/2$, we have

$$k_{MA}(r) = \sigma^2 \left(1 + \frac{\sqrt{3}r}{\ell} \right) \exp\left(-\frac{\sqrt{3}r}{\ell} \right)$$

5.3.1.1 Converting a covariance function to a dynamic matrix

It is possible to construct a Gaussian process as result of the Kalman filter given a certain dynamic matrix $\mathbf{A}$. Following Sarkka and Hartikainen (2012), the strategy aims to construct $\mathbf{A}$ such that the posterior density given by Kalman filter has the same spectral density as the desired Gaussian process. This is possible due to the Wiener-Khinchin theorem.

The approach is to consider that a hypothetical dynamics in function of the random process B_t

$$a_q \frac{\mathrm{d}^q}{\mathrm{d}t^q} X_t + ... + a_1 \frac{\mathrm{d}}{\mathrm{d}t} X_t + a_0 X_t = \frac{\mathrm{d}}{\mathrm{d}t} B_t,$$

whose state-space format is

$$\mathrm{d}X_t = \begin{bmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & ... & ... & ... & 1 \\ a_0 & a_1 & ... & ... & a_q \end{bmatrix} X_t \mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \mathrm{d}B_t \tag{5.1}$$

with $X_t := \left[X_t^{(0)}, \;\; X_t^{(1)}, \;\; ..., \;\; X_t^{(q)} \right]^\top$ such that $X_t^{(q)}$ is a stochastic process of the same kind of X_t. Also, that $\mathrm{d}X_t := \left[\mathrm{d}X_t^{(0)}, \;\; \mathrm{d}X_t^{(1)}, \;\; ..., \;\; \mathrm{d}X_t^{(q)} \right]^\top$. Taking the Fourier transform of this equation gives

$$\mathsf{X}(i\omega)\left(a_n(i\omega)^n + ... + a_1(i\omega) + a_0\right) = \mathsf{W}(i\omega),$$

$$\mathsf{X}(i\omega)\mathsf{G}(i\omega) = \mathsf{B}(i\omega),$$

where the respective Fourier transforms are $\mathfrak{F}\{X_t\} = \mathsf{X}(i\omega)$, $\mathfrak{F}\{\mathrm{d}B_t\} = \mathsf{W}(i\omega)$ and $\mathsf{G}(i\omega) := (a_n(i\omega)^n + ... + a_1(i\omega) + a_0)$. This gives the spectral density of the process X_t

$$\mathsf{S}(\omega) := \mathsf{X}(i\omega)\mathsf{X}(-i\omega) = q_c |\mathsf{G}(i\omega)|^2$$

where we assumed $|\mathsf{W}(i\omega)|^2 = q_c{}^2$. The next step is to use the Wiener-Khinchin theorem to relate the spectral density $\mathsf{S}(\omega)$ with the correlation of the process X_t. Such relation is given by

$$R_X(\tau) = \mathfrak{F}^{-1}\{\mathsf{S}(\omega)\},$$

where $\mathfrak{F}^{-1}$ is the inverse Fourier transform of the spectral density $\mathsf{S}(\omega)$ and $R_X(\tau)$ the correlation between the X_t and $X_{t'}$, with $\tau = t - t'$. Notice that, this relation works in both ways, either we may construct stochastic differential equations from correlation function by applying the Fourier transform, or we may obtain correlation functions from stochastic differential equations via the inverse Fourier transform. This is not restrictive given we may construct correlation functions in a direct sense using the properties of the expectation (for correlation functions obtained from state-space models, see Grigorievskiy and Karhunen (2016)).

---

[2] We saw in Example 23 that such process is not physically possible once requires infinity energy, but it is mathematically convenient then we proceed with it.

Example 16 (Matérn covariance function (SARKKA; HARTIKAINEN, 2012)) *The family of Matérn covariance functions are given as*

$$C(\tau) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{\tau}{\ell}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\frac{\tau}{\ell}\right)$$

where $\nu, \ell > 0$ *are the smoothness and the length-scale parameters. The function* K_{ν} *is the modified Bessel function. The spectral density of* $C(\tau)$ *is*

$$S(\omega) \propto (\lambda^2 + \omega^2)^{-(\nu+1/2)}$$

where $\lambda = \sqrt{2\nu}/\ell$*. A possible factorization of the spectral density is* $S(\omega) \propto (\lambda^2+\omega^2)^{-(p+1)}(\lambda+\omega)^{-(p+1)}$*, where* $\nu = p+1/2$*. This results in the following transfer function for the stable part*

$$\mathsf{G}(i\omega) = (\lambda+\omega)^{-(p+1)}.$$

For instance for $p = 1$ *we obtain the corresponding stochastic differential equation*

$$dX_t = \begin{bmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{bmatrix} X_t dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dB_t,$$

where $X_t = [X_t^{(0)}, X_t^{(1)}]^{\top}$.

5.4 Gaussian ODE filters

Continuing in the state-space approach, we follow Kersting *et al.* (2020) to use the property that we are able to recast initial value problems such as the solution in time of an ODE as a SDE. This sets a probabilistic interpretation to the numerical methods used to solve the "classical" ODE problems, allowing us to employ the language of probability theory to treat the limited knowledge one may have about the numerical computations. We refer to the literature under the name of Probabilistic Numerics to a more comprehensive set of probabilistic approaches to classical numerical methods beyond ODE solvers, e.g. linear algebra, optimization, integration, etc (HENNIG *et al.*, 2022).

In Section 3.2.2 we saw how to obtain the statistics for SDEs that can be represented as Gaussian processes. Recall the state-space format in Equation (5.1) such that

$$\mathrm{d}X_t = \mathbf{F}X_t\mathrm{d}t + \mathbf{L}\mathrm{d}B_t.$$

Substituting by identification the terms $f(X_t,t) = \mathbf{F}$ and $G(X_t,t) = \mathbf{L}$, we obtain as solution of the dynamic equations of the statistics of the linear SDE above, the mean and the covariance time-functions

$$\begin{aligned}
&\mathbf{m}(t) = \mathbf{A}(t)\mathbf{m}(0),\\
&\mathbf{P}(t) = \mathbf{A}(t)\mathbf{P}(0)\mathbf{A}(t)^\top + \mathbf{Q}(t),\\
&\mathbf{A}(t) = e^{t\mathbf{F}},\\
&\mathbf{Q}(t) = \int_0^t e^{\mathbf{F}(t-\tau)}\mathbf{L}\mathbf{L}^\top \left(e^{\mathbf{F}(t-\tau)}\right)^\top \mathrm{d}\tau,
\end{aligned}$$

where $\mathbf{m}(0)$ and $\mathbf{P}(0)$ are assumed to be given. Notice that the algorithmic implementation of such Gaussian ODE filter breaks down simply to the application of the Kalman filter equations, as the ones shown in Example 13 considering just the prediction steps (the expressions for the matrices $\mathbf{m}_n^p$ and $\mathbf{P}_n^p$). The new idea here is to base the choice of the $\mathbf{A}$.

5.4.0.1 Estimating ODE solutions using Gaussian ODE filters

It is possible to obtain estimated solutions to ordinary differential equations (ODEs) using the technique introduced previously. One reason for such is that if one is interested in finding the solutions of ODEs as probability densities with a full Bayesian approach may be computationally heavy. The proposal of Tronarp *et al.* (2019) is to substitute the observation equation of the Kalman filter with measurements of the vector field of a given ODE. This means that if a system is modeled by the following ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,\theta,t) = \dot{x}$$

where the states $x \in \mathbb{R}^{D_x}$ and θ the parameters of the model of the system, the probability density of the observations Z_n of the Kalman filter is defined to be

$$p_{Z_n|X_n}(z_n,x_n) = \mathcal{N}(Z_n|\dot{x}_n - f(X_n,\theta,t_n),R).$$

If we choose $R \to 0$ this density approaches the Dirac delta and there is no uncertainty over the vector field $f(x,\theta,t)$. The state evolution proposed by Tronarp *et al.* (2019) assumes the Gaussian process to be a Wiener process which is proven to be comparable to the classic numerical methods for integration (SCHOBER *et al.*, 2019).

We use again the state-space approach and define the vector X_t as

$$X_t = \left[\left(X_t^{(0)}\right)^\top, \left(X_t^{(1)}\right)^\top, \ldots, \left(X_t^{(q)}\right)^\top\right]^\top. \tag{5.2}$$

being $X_t^{(q)}$ the q-th time derivative of X_t. Then, we define the following stochastic differential equation as

$$\mathrm{d}X_t = \mathbf{F}X_t\mathrm{d}t + \mathbf{L}\mathrm{d}B_t, \tag{5.3}$$

being $\mathbf{W}_t$ a q-dimensional Brownian motion. To use the matrix form we define the m-th canonical basis vector $\mathrm{e}_m{}^3$ in $\mathbb{R}^{q+1}$, I_{D_x} an identity matrix of size $D_x \times D_x$ and $\otimes$ the Kronecker product. $\mathbf{F} \in \mathbb{R}^{D_x(q+1)\times D_x(q+1)}$ is the state transition matrix of the stochastic differential equation, $\mathbf{L} \in \mathbb{R}^{D_x(q+1)}$ is a diffusion matrix and $\mathrm{d}B_t$ is a $D_x(q+1)$-th dimensional version of the Brownian motion. We rewrite the assumed observations Z_n as a stochastic process such that

$$Z_t := \mathrm{E}_1 X_t - f(\mathrm{E}_0 X_t, \theta, t) \equiv 0. \tag{5.4}$$

where $\mathrm{E}_m X_t$ with $\mathrm{E}_m = \mathrm{e}_m \otimes I_{D_x}$ selects the m-th row of equation (5.2) containing the vector with m-th time-derivative of all the D_x states. The densities for the implementation of the Kalman filter are then

$$p_{X_{t_{n+1}}|X_{t_n}} = \mathcal{N}(X_{t_{n+1}}|\mathbf{A}(h)\mu_n, \mathbf{A}(h)\Sigma_n\mathbf{A}^\top(h) + \mathbf{Q}(h)), \tag{5.5a}$$

$$p_{Z_{t_{n+1}}|X_{t_n}} = \mathcal{N}(\mathrm{E}_1 X_{t_n} - f(\mathrm{E}_0 X_{t_{n+1}}, \theta, t), \mathbf{R}), \tag{5.5b}$$

whose parameters are obtained by evaluating the mean and covariance of the solution of the Equation (5.3) and $h = t_{n+1} - t_n$. We obtain then

$$\Sigma(t) = \exp(\mathbf{F}t)\Sigma_0\left[\exp(\mathbf{F}t)\right]^\top + \mathbf{Q}, \tag{5.6a}$$

$$\mathbf{A}(t) = \exp(\mathbf{F}h), \tag{5.6b}$$

$$\mathbf{Q}(t) = \int_0^t \exp(\mathbf{F}(t-\tau))\mathbf{L}\mathbf{L}^\top \exp(\mathbf{F}^\top(t-\tau))\mathrm{d}\tau. \tag{5.6c}$$

The drawback of this method is that if $f(x,\theta,t)$ is not a linear function, the equations of the Kalman filter may not be evaluated analytically and approximations are required as the linearization introduced in Section 2.4, resulting in the Extended Kalman filter, for instance (SÄRKKÄ; SVENSSON, 2023b).

---

[3] Also called standard basis, defined as $\mathrm{e}_m = [\delta_{1,m}, ..., \delta_{D_x,m}]$.

6 THERMODYNAMICS OF THE MICHAELIS-MENTEN

The Michaelis-Menten model is a fundamental concept in enzymology, providing a mathematical framework to describe the kinetics of enzyme-catalyzed reactions. Named after the biochemists Leonor Michaelis and Maud Menten, who first proposed it in 1913, this model offers valuable insights into the mechanisms underlying enzyme-substrate interactions. Enzymes are biological catalysts that accelerate chemical reactions by facilitating the formation of enzyme-substrate complexes. In the Michaelis-Menten model, an enzyme (E) binds reversibly with a substrate (S) to form an enzyme-substrate complex (ES). The model describes the rate of formation of the ES complex and the subsequent conversion of ES into product (P). It assumes that the reaction proceeds through an initial rapid formation of ES followed by a slower conversion of ES into product.

In this chapter the theory of the Michaelis-Menten in the context of master equations is revised in Section 6.1 and the stochastic thermodynamic analysis is done in Section 6.2.

6.1 Michaelis-Menten theory

The Michaelis-Menten reacting system is depicted as

$$E+S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E+P.$$

In this analysis, it is assumed that there is a single molecule enzyme in either E or ES form and the number of molecules of S and P is large enough to not be affected by the change of the state of the enzyme. The reactions are modeled as random interactions in an environment at temperature T in contact with a thermal reservoir which is assumed to be in equilibrium (SEIFERT, 2010).

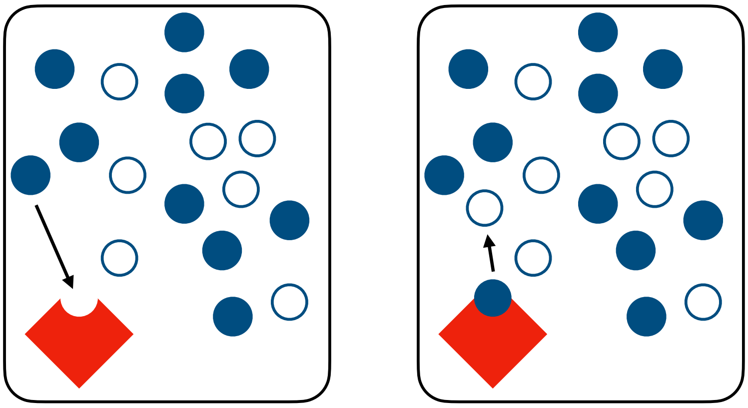
The system consists of 1 molecule of E or 1 molecule of ES, being the system states $\{X \in \mathcal{X} | X = (X_i, X_{i'})\}$ with $\mathcal{X} = \{0,1\} \times \{0,1\}$.

The evolution of the system is modeled as a Markov jump process between states x whose probability $P_X(t)$ solves the following the master equation (KAMPEN, 2007; GILLESPIE, 1976)

$$\frac{\mathrm{d}P_t(X)}{\mathrm{d}t} = \sum_{X,X'} W(X,X')P_t(X') - W(X',X)P_t(X) \tag{6.1}$$

$g_{xx'}(k_v)$ is the probability transition rate from a state x' to state x, it depends on the kinetics, and it is an element of the rate transition matrix G of size $|\mathcal{X}| \times |\mathcal{X}|$ is equal to

Figure 8 – Pictorial representation of the enzymatic reaction of the Michaelis-Menten, where the solution is composed by substrate (filled circle), product (empty circle) and single enzyme (red square). The single enzyme interacts with the solution and changes its state.



Source: Adapted from Seifert (2010).

(MUNSKY; KHAMMASH, 2006b)

$$-\sum_{\nu} a_\nu(x,k_\nu)\delta(x-x') + a_\nu(x-s_\nu,k_\nu)\delta(x-x'+s_\nu),$$

where the term

$$a_\nu(x,k_\nu) = k_\nu \frac{x_i!}{(x_i-s_{i,\nu})!}\frac{x_{i'}!}{(x_{i'}-s_{i',\nu})!}$$

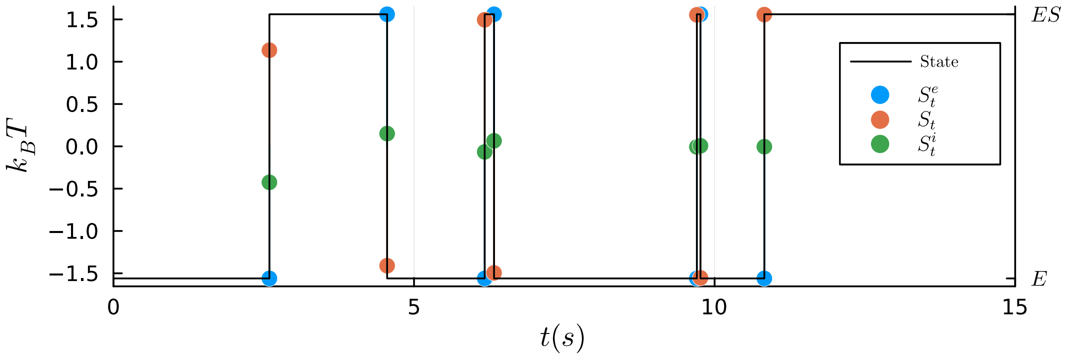
is the propensity function for the ν-th reaction with rate k_ν. $S_\nu = (s_{i,\nu}, s_{i',\nu})$ is the change in the number of molecules of $x_i \in x$ participating in the ν-th reaction (i.e., when ES or E changes from 1 to 0 and vice versa). $\delta(x) = 1$ only if $x \equiv (0,0)$, otherwise is 0.

Considering the system with $n_E + n_{ES} = 1$, it can be verified that the states $(0,0)$ and $(1,1)$ have no transitions coming to or leaving them, making the matrix G to have the size 2×2 and $P_t(X) = 0$ for those states.

6.2 Contributions and discussion

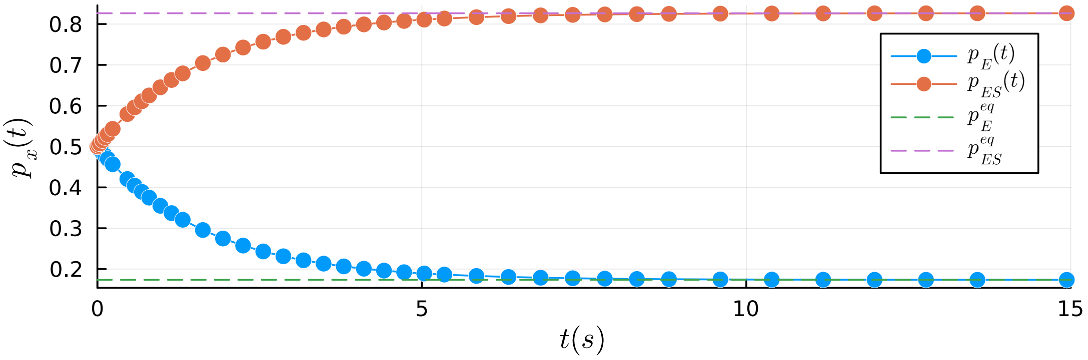
Some conclusions can be drawn from the results of the calculations shown in Figures 9, 10 and 11(FARIAS *et al.*, 2023a; FARIAS *et al.*, 2023b). Firstly, in Figure 9 we have a single realization of the Markov jump process parametrized by the Michaelis-Menten with parameters $k_1 = 0.5$, $k_{-1} = 0.005$, $k_2 = 0.1$ and $k_{-2} \approx 0.0$. For each jump, the entropy flux, balance and production defined in were evaluated as established in Section 4.3.

Figure 9 – On the right axis, a single trajectory of the Markov jump process for the Michaelis-Menten with $k_1 = 0.5$, $k_{-1} = 0.005$, $k_2 = 0.1$ and $k_{-2} \approx 0.0$, and entropy, entropy exchange and entropy production for that trajectory.



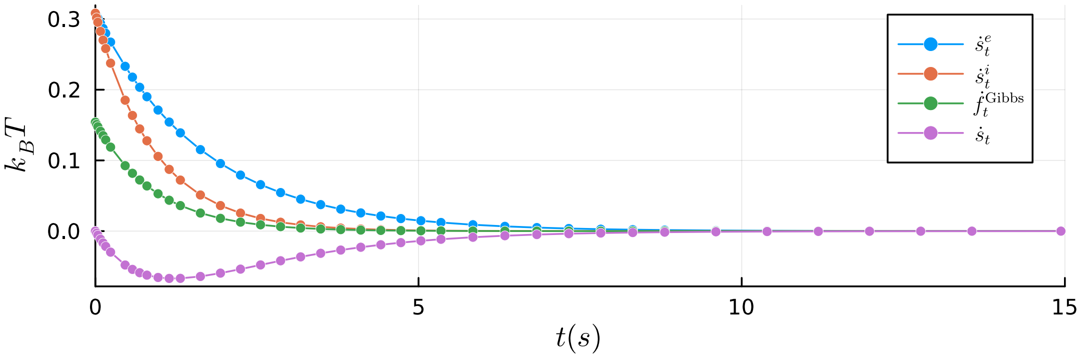
Source: Prepared by the author.

Figure 10 – Time evolution of the probability of the states of the system, p_E and p_{ES}, and the probability of the states for the system in equilibrium, p_E^{eq} and p_{ES}^{eq}.



Source: Prepared by the author.

Figure 11 – Average entropy flux rate s_t^e, average entropy production rate s_t^i, average free energy/work difference rate $\dot{f}_t^{\text{Gibbs}}$ and average entropy rate s_t.



Source: Prepared by the author.

The solution of the master equation is shown in Figure 10. Notice that this solution also provides the information about the probability of the states like $P(X_t = E) = p_E(t)$ and

$P(X_t = ES) = p_{ES}(t)$. The time-evolution of the solution of the master equation shows that the system relaxes to the equilibrium probability, denoted by $p_{\cdot}^{eq}$ once the rate of probability change reaches zero.

The corresponding average rates of the energetic quantities (free energy, entropy flux, production and balance) are shown in Figure 11. Integrating each of the graphs would give the total change of each quantity, e.g. the area above the $\dot{s}$ curve gives the average change in entropy of the system. We conclude the system reduces its entropy by giving entropy to the bath. Such entropy given is measured by the area under the graph of $\dot{s}_{\mathrm{e}}$. As all the rates tends to zero, this system is said to achieve the equilibrium (QIAN; GE, 2021). We denote $\dot{f}_t^{\mathrm{Gibbs}}$ the Kullback-Leibler divergence between the probability of the equilibrium reached and the probability of the system at time t. Esposito and Broeck (2011) suggest that this may be the available work to be extracted from the system, as the difference between work and Gibbs free energy, but it still is object of future investigations. The evaluation of the energetic quantities aforementioned is only possible if one obtains the solution of Equation (6.1) by analytical or numerical integration or by Monte Carlo using the sampling of the trajectories of the Markov jump process.

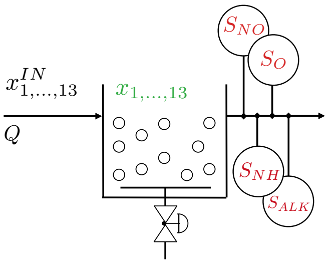
7 INFERENCE OF THE PARAMETERS OF A BIOREACTOR

In general, the approach used to estimate the parameters of bioreactors consists of simulating the system and evaluating the difference between the states and observed data available from the wastewater treatment plant (WWTP) (HOLMBERG, 1982; REICHERT, 1998; GADKAR *et al.*, 2003; ANTONOPOULOU *et al.*, 2015). This mostly involves four steps: i) characterisation of the influent wastewater; ii) definition of the dynamic influent loading data; iii) manual estimation of the model parameters; and iv) model validation. Few works adopt a probabilistic approach (SHARIFI *et al.*, 2014b; ZONTA *et al.*, 2014b; STENTOFT *et al.*, 2019), whereas other procedures use information about the gradient of the dynamics of the system to help with the computational burden (MUNACK, 1989). Systematic analysis approaches mainly consist of parameter identification (DOCHAIN; VANROLLEGHEM, 2005), sensitivity analysis (COSENZA *et al.*, 2013), and error propagation (GUJER, 2008), which have been developed and tested. A numerical optimal approaching procedure has been used for the systematic calibration of ASMs (ZHU *et al.*, 2015).

7.1 Model description

We explore the widely used Activated Sludge Model No. 1 (ASM1, (HENZE *et al.*, 1987)) which models a bioreactor consisting of the aerated continuous stirred tank reactor with perfect mixing shown in Figure 12, where the influent load, x^{IN}, is given in terms of the thirteen ASM1 state variables and the flow-rate, Q, that equals the effluent one, the set of sensors in the effluent stream gives the the measured variables: dissolved oxygen, nitrate and nitrite nitrogen, ammonia nitrogen and alkalinity, that is S_O, S_{NO}, S_{NH} and S_{ALK}, respectively, and K_La gives an adjustable variable considered representative of the diffused aeration in the bioreactor.

Figure 12 – Bioreactor schematic.



Source: Elaborated by the author.

ASM1 is a state-of-the-art model commonly used for representing the carbon and nitrogen removal, oxygen consumption and sludge production. The growth of the microbiological population responsible for the processes occurring inside the reactor is described by Monod kinetics (MONOD, 1949). The mass balances for the state variables lead to a system of ordinary differential equations of the thirteen state variables in Table 3 that comprises stoichiometric and kinetic parameters in Table 4. As an example of the nonlinearity in the ASM1, the differential equation for the readily biodegradable organic matter, S_S in equation (7.1), is provided.

$$\frac{\mathrm{d}S_S}{\mathrm{d}t} = \frac{Q}{V}\left[S_S^{IN} - S_S\right] + R_{S_S} \tag{7.1a}$$

$$\begin{aligned} R_{S_S} = & -\frac{\mu_H}{Y_H}\frac{S_S}{K_S + S_S}\left[\frac{S_O}{K_{OH} + S_O} + \eta_g \frac{K_{OH}}{K_{OH} + S_O}\frac{S_{NO}}{K_{NO} + S_{NO}}\right] X_{BH} \\ & + k_h \frac{X_S}{K_X X_{BH} + X_S}\left[\frac{S_O}{K_{OH} + S_O} + \eta_h \frac{K_{OH}}{K_{OH} + S_O}\frac{S_{NO}}{K_{NO} + S_{NO}}\right] X_{BH}. \end{aligned} \tag{7.1b}$$

The right-hand side of Equation (7.1a) represents the mass balance and the R_{S_S} term, fully given in Equation (7.1b), is the kinetic of the reaction producing (or consuming) S_S. S_S^{IN} is the inflow concentration of biodegradable substrate. We refer to the work in Neto *et al.* (2022) for the full description of the system of equations. The biological reactor is simulated considering the settings established by the Benchmark Simulation Model No. 1 (BSM1, (ALEX *et al.*, 2008)), a simulation environment defining a plant layout, the ASM1 as simulation model, the influent load concentrations and flow-rate, as well as the test procedures and evaluation criteria. We consider the steady-states of the BSM1 computed in Gernaey *et al.* (2014) as initial conditions to solve the differential equations of the ASM1.

We consider that, beyond the measurements of nitrate and nitrite nitrogen, S_{NO}, and dissolved oxygen, S_O, given by the sensors setting in the simulation platform (GERNAEY *et al.*, 2014) and commonly used in modern WWTPs, it is also possible to measure ammonia nitrogen, S_{NH}, (WANG *et al.*, 2020) and alkalinity, S_{ALK}, (VANROLLEGHEM; LEE, 2003) in real-time. The number of sensors is realistic and low avoiding to increase operational costs in a real plat setting. The measurements occur ten times a day and are evenly spaced.

7.2 Inference via marginal likelihood

Following Tronarp *et al.* (2023), the proposition is to calibrate the ASM model using the marginal likelihood defined as

$$\Omega(\{Y_{t_n}\}_{n=1}^{N}, \theta) = \int p(\{Y_{t_n}\}_{n=1}^{N} | \{X_{t_n}\}_{n=1}^{N}, \theta) p(\{X_{t_n}\}_{n=1}^{N} | \theta) \mathrm{d}X_{t_1} ... \mathrm{d}X_{t_N}. \tag{7.2}$$

for noisy measurements $\{Y_{t_n}\}_{n=1}^{N}$ of the real system. By $p(\{X_{t_n}\}_{n=1}^{N}|\theta)$ we represent the posterior $p(X_n|\{Z_{t_n}\}_{n=1}^{N})$ obtained by the solution of the ASM model using the method described in Section 2.5. Here, the dependence on Z_{t_n} is omitted to short the notation.

Once the posteriors $p(X_n|\{Z_{t_n}\}_{n=1}^{N})$ were obtained, we run another Kalman filter backwards in time to compute the marginal likelihood using the observations Y_n[1] (SCHWEPPE, 1965). We assume the following probability model

$$p(X_{t_n}|X_{t_{n+1}}) = \mathcal{N}(X_{t_n}|\xi_{t_n}, \mathbf{P}_{t_n}), \tag{7.3a}$$

$$p(Y_{t_n}|X_{t_n}) = \mathcal{N}(Y_{t_n}|\mathbf{B}\,\mathrm{E}_0 X_{t_n}, \sigma_y^2 \mathbf{I}). \tag{7.3b}$$

The parameters are then obtained via optimization of $\Omega(\{Y_{t_n}\}_{n=1}^{N}, \theta)$ algorithm (e.g, LBFGS).

7.3 Contributions and discussion

The computation of the solution of the system of differential equations for both the sensitivity analysis and the least squares method is done using the package DifferentialEquations.jl implemented in Julia language (BEZANSON *et al.*, 2017; RACKAUCKAS; NIE, 2017).

We decided to comprehensively identify the set of parameters, $\theta = Y_H$, Y_A, μ_A, b_H, μ_A, K_{NH} and K_{OA}, simultaneously and evaluate the performance of parameter estimation across them based in previous sensitivity analysis (FARIAS *et al.*, 2024).

To compare parameter estimation results, we employ the least squares (LS) method objective function, which is defined basically as the square distance between the solution obtained via a classical numerical method and the set of observations $\{Y_n\}$. The solution of the classical numerical method also varies with respect to the ASM1 parameters θ, then using the LBFGS algorithm and setting boundaries of the optimisation problem based on the range of

---

[1] We also refer to Wu and Lysy (2023) for a time-forward implementation.

values established by the literature review by Jr *et al.* (2022), we obtain the parameters for the LS method. The measurement of the states S_O, S_{NO}, S_{NH} and S_{ALK} are subject to additive Gaussian noise with standard deviation σ_y. The assumed range of σ_y in the simulated data encompasses values wider than the precision of the available sensors (GERNAEY *et al.*, 2014). Four scenarios were tested by varying the standard deviation of the noise in the observations, as detailed in the first column of Table 1. The absolute errors $|\theta^\star - \theta|$ are presented in Table 1, where $\theta^\star$ represents the result of the optimisation and θ denotes their true values shown in Table 4. Noticeably, LS was unable to reduce the error for σ_y values below 10 leading to the same parameter estimates on those scenarios.

Table 1 – Parameter absolute error for the LS objective function. The parameters for which the method performed better than the marginal likelihood method were highlighted.

σ_y	Y_A	Y_H	μ_H	b_H	μ_A	K_{NH}	b_A	K_{OA}
10	0.0400	0.0500	2.0000	1.3000	0.3000	**0.0000**	0.1000	1.6000
0.1	0.1679	0.0495	**0.0200**	0.2345	0.2920	0.2475	**0.0010**	0.0160
0.01	0.1679	0.0495	**0.0200**	**0.2345**	0.2920	0.2475	**0.0010**	0.0160
0.001	0.1679	0.0495	**0.0200**	0.2345	**0.2920**	0.2475	**0.0010**	0.0160

Source: Farias *et al.* (2024).

Table 2 – Parameter absolute error for the marginal likelihood method. The parameters for which the method performed better than the LS method were highlighted.

σ_y	Y_A	Y_H	μ_H	b_H	μ_A	K_{NH}	b_A	K_{OA}
10	**0.0039**	**0.0000**	**0.1659**	**1.3000**	**0.0975**	0.1674	**0.0999**	**0.0784**
0.1	**0.1404**	**0.0000**	0.1307	**0.1736**	**0.1930**	**0.2400**	0.0954	**0.0002**
0.01	**0.0636**	**0.0446**	0.0535	0.2500	**0.1590**	**0.2234**	0.0527	**0.0066**
0.001	**0.1505**	**0.0486**	0.0248	**0.2266**	0.3000	**0.2462**	0.0280	**0.0015**

Source: Farias *et al.* (2024).

The marginal likelihood evaluation in this study is conducted using the Fenrir package developed in the Julia programming language (`Fenrir.jl`, (TRONARP *et al.*, 2023)). It generates the negative log marginal likelihood as the objective function utilised in the optimisation problem. The optimisation is carried out using the same LBFGS settings as employed in the least squares method. Similarly to Table 1, Table 2 presents the absolute error $|\theta^\star - \theta|$ where $\theta^\star$ represents the result of the optimisation using the negative log marginal likelihood as the objective function and θ denotes their true values shown in Table 4.

Overall, the marginal likelihood objective function leads to parameter estimations that are closer to their true values in scenarios with higher noise. In contrast, the least squares method was unable to achieve parameter estimates closer to their true values for the values of σ_y below 10.

8 CONCLUSION

This work aimed to present some insights about the applications of SDEs, and it is far from being exhaustive. Some conclusions can be drawn from the Chapter 5:

1. SDEs are able to comprise the flexibility present in the already known and widely used Gaussian process;
2. The Kalman filter is able to deal with Gaussian processes for temporal series with a lower computational cost;
3. The application to parameter inference leads to higher precision than classical methods, but at the cost of increase the processing resources.

It is worth to mention that on the opposite approach of transforming Gaussian/Wiener processes into state-space models to leverage the computational burden, if SDEs may be represented in the state-space, one may use the Gaussian process to learn its solutions (GRIGORIEVSKIY; KARHUNEN, 2016). This may be the object of future works.

Yet with respect to the state-space representation, the approach is not limited to temporal but also the spatial-temporal case (SARKKA *et al.*, 2013) and may be the object of future works.

With respect to the Chapter 4:

1. The framework of stochastic thermodynamics is able to deal with systems out-of-equilibrium (differently from classical thermodynamics) in contact with a heat reservoir at a mesoscopic scale;
2. Energetic measures as heat, work and entropy can be calculated from the observation of the trajectory of such systems.

In special for the stochastic thermodynamics, the work is not complete. A non-exhaustive list of applications of the theory covers from the energetics of cellular structures (BROECK *et al.*, 2013), the inference of thermodynamic quantities from trajectories (SEIFERT, 2019), thermodynamics of computation (WOLPERT, 2019), thermodynamic efficiency of machine learning models (PARSI, 2024) and even economics (QIAN, 2009).

REFERENCES

ALEX, J.; BENEDETTI, L.; COPP, J.; GERNAEY, K.; JEPPSSON, U.; NOPENS, I.; PONS, M.; RIEGER, L.; ROSEN, C.; STEYER, J. *et al.* Benchmark simulation model no. 1 (bsm1). **Report by the IWA Taskgroup on benchmarking of control strategies for WWTPs**, Department of Industrial Electrical Engineering and Automation, Lund University, Sweden, v. 1, 2008.

ANTONOPOULOU, G.; ALEXANDROPOULOU, M.; LYTRAS, C.; LYBERATOS, G. Modeling of anaerobic digestion of food industry wastes in different bioreactor types. **Waste and Biomass Valorization**, Springer Science and Business Media LLC, v. 6, n. 3, p. 335–341, mar 2015. Available at: <https://doi.org/10.1007%2Fs12649-015-9362-7>.

ARNOLD, L. **Stocastic Differential Equations: theory and applications**. [S.l.]: John Wiley & Sons, 1974.

BARATO, A. C.; SEIFERT, U. Thermodynamic uncertainty relation for biomolecular processes. **Physical review letters**, APS, v. 114, n. 15, p. 158101, 2015.

BERNOULLI, D. **Hydrodynamica, sive, De viribus et motibus fluidorum commentarii**. [S.l.]: Sumptibus Johannis Reinholdi Dulseckeri, typis Joh. Henr. Deckeri, 1738.

BEZANSON, J.; EDELMAN, A.; KARPINSKI, S.; SHAH, V. B. Julia: A fresh approach to numerical computing. **SIAM review**, SIAM, v. 59, n. 1, p. 65–98, 2017.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. **Journal of Political Economy**, v. 81, n. 3, p. 637–654, 1973.

BOLTZMANN, L. **Weitere studien über das wärmegleichgewicht unter gasmolekülen**. [S.l.]: Aus der kk Hot-und Staatsdruckerei, 1872.

BORN, M.; OPPENHEIMER, J. R. On the quantum theory of molecules. **Annalen der Physik**, January 1927. Translated by S. M. Blinder, with emendations by Brian Sutcliffe and Wolf Geppert.

BROECK, C. Van den *et al.* Stochastic thermodynamics: A brief introduction. **Phys. Complex Colloids**, v. 184, p. 155–193, 2013.

BROWN, R. Xxvii. a brief account of microscopical observations made in the months of june, july and august 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. **The philosophical magazine**, Taylor & Francis, v. 4, n. 21, p. 161–173, 1828.

BUTKOV, E. **Mathematical Physics**. [S.l.]: Addison-Wesley Publishing Company, 1968. (A-W series in advanced physics). ISBN 9780201007275.

CARNOT, S. **Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance**. [S.l.]: Gauthier-Villars, 1824.

CHEN, C.-T. **Linear system theory and design**. [S.l.]: Saunders college publishing, 1984.

CLAUSIUS, R. Ueber die art der bewegung, welche wir wärme nennen. **Annalen der Physik**, WILEY-VCH Verlag Leipzig, v. 176, n. 3, p. 353–380, 1857.

CLAUSIUS, R. **The mechanical theory of heat**. [S.l.]: Macmillan, 1879.

COCKAYNE, J.; OATES, C. J.; SULLIVAN, T. J.; GIROLAMI, M. Bayesian probabilistic numerical methods. **SIAM review**, SIAM, v. 61, n. 4, p. 756–789, 2019.

CODDINGTON, A.; LEVINSON, N. **Theory of Ordinary Differential Equations**. [S.l.]: R.E. Krieger, 1984. (International series in pure and applied mathematics). ISBN 9780898747553.

COSENZA, A.; MANNINA, G.; VANROLLEGHEM, P. A.; NEUMANN, M. B. Global sensitivity analysis in wastewater applications: A comprehensive comparison of different methods. **Environmental Modelling & Software**, v. 49, p. 40–52, 2013.

DOCHAIN, D.; VANROLLEGHEM, P. **Dynamical Modelling and Estimation in Wastewater Treatment Processes**. [S.l.]: IWA Publishing, 2005.

DOLGOV, S.; KHOROMSKIJ, B. Simultaneous state-time approximation of the chemical master equation using tensor product formats. **Numerical Linear Algebra with Applications**, Wiley Online Library, v. 22, n. 2, p. 197–219, 2015.

EINSTEIN, A. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. **Annalen der physik**, v. 17, n. 549-560, p. 208, 1905. Translated by A. Beck.

ESPOSITO, M.; BROECK, C. V. den. Second law and landauer principle far from equilibrium. **Europhysics Letters**, v. 95, n. 4, p. 40004, aug 2011. Available at: <https://dx.doi.org/10.1209/0295-5075/95/40004>.

ESPOSITO, M.; BROECK, C. Van den. Three faces of the second law. i. master equation formulation. **Physical Review E—Statistical, Nonlinear, and Soft Matter Physics**, APS, v. 82, n. 1, p. 011143, 2010.

EVANS, D. J.; COHEN, E. G. D.; MORRISS, G. P. Probability of second law violations in shearing steady states. **Physical review letters**, APS, v. 71, n. 15, p. 2401, 1993.

FARIAS, F. P. de; CRUZ, D. L.; MULAS, M. Inference of the parameters of a bioreactor via approximate Bayesian evidence. **Learning & Nonlinear Models**, SBIC, v. 22, n. 1, p. 4–16, 2024.

FARIAS, F. P. de; MULAS, M.; CORONA, F. Mesoscopic thermodynamics of single-particle enzymatic reactions. In: **Book of Abstracts, Challenges in the Physics of Active and Biological Matter**. [S.l.: s.n.], 2023.

FARIAS, F. P. de; MULAS, M.; CORONA, F. Mesoscopic thermodynamics of single-particle enzymatic reactions. In: **Book of Abstracts, Nordic Process Control Workshop**. [S.l.: s.n.], 2023.

GADKAR, K.; DOYLE, F.; CROWLEY, T.; VARNER, J. Cybernetic model predictive control of a continuous bioreactor with cell recycle. **Biotechnology Progress**, Wiley, v. 19, n. 5, p. 1487–1497, oct 2003. Available at: <https://doi.org/10.1021%2Fbp025776d>.

GERNAEY, K. V.; JEPPSSON, U.; VANROLLEGHEM, P. A.; COPP, J. B. **Benchmarking of Control Strategies for Wastewater Treatment Plants**. [S.l.]: IWA Publishing, 2014.

GILLESPIE, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. **Journal of Computational Physics**, v. 22, n. 4, p. 403–434, 1976. ISSN 0021-9991. Available at: <https://www.sciencedirect.com/science/article/pii/0021999176900413>.

GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. **The journal of physical chemistry**, ACS Publications, v. 81, n. 25, p. 2340–2361, 1977.

GILLESPIE, D. T. The chemical langevin equation. **The Journal of Chemical Physics**, American Institute of Physics, v. 113, n. 1, p. 297–306, 2000.

GOODSTEIN, D. L. **States of matter**. [S.l.]: Courier Corporation, 1985.

GRIGORIEVSKIY, A.; KARHUNEN, J. **Gaussian Process Kernels for Popular State-Space Time Series Models**. 2016.

GROOT, S. R. D.; MAZUR, P. **Non-equilibrium thermodynamics**. [S.l.]: Courier Corporation, 2013.

GUJER, W. **Systems Analysis for Water Technology**. [S.l.]: Springer Berlin, Heidelberg, 2008.

HACKING, I. **The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference**. 2. ed. [S.l.]: Cambridge University Press, 2006.

HENNIG, P.; OSBORNE, M. A.; KERSTING, H. P. **Probabilistic Numerics: Computation as Machine Learning**. [S.l.]: Cambridge University Press, 2022.

HENZE, M.; GRADY, C. P. L. J.; GUJER, V.; MARAIS, G. v. R.; MATSUO, T. **Activated Sludge Model No. 1. IAWPRC Scientific and Technical Reports No. 1.** [S.l.], 1987.

HOLMBERG, A. On the practical identifiability of microbial growth models incorporating michaelis-menten type nonlinearities. **Mathematical Biosciences**, Elsevier BV, v. 62, n. 1, p. 23–43, nov 1982. Available at: <https://doi.org/10.1016%2F0025-5564%2882%2990061-x>.

HOROWITZ, J. M. Diffusion approximations to the chemical master equation only have a consistent stochastic thermodynamics at chemical equilibrium. **The Journal of chemical physics**, AIP Publishing, v. 143, n. 4, 2015.

ITŌ, K. 109. stochastic integral. **Proceedings of the Imperial Academy**, The Japan Academy, v. 20, n. 8, p. 519–524, 1944.

JAHNKE, T.; HUISINGA, W. Solving the chemical master equation for monomolecular reaction systems analytically. **Journal of mathematical biology**, Springer, v. 54, n. 1, p. 1–26, 2007.

JR, A. M.; SILVA, M.; BENETTI, A. Evaluation and optimization of asm1 parameters using large-scale wwtp monitoring data from a subtropical climate region in brazil. **Water Practice & Technology**, IWA Publishing, v. 17, n. 1, p. 268–284, 2022.

KALLIANPUR, G. **Stochastic Filtering Theory**. [S.l.]: Springer New York, 2013. (Stochastic Modelling and Applied Probability). ISBN 9781475765922.

KALMAN, R. E. A new approach to linear filtering and prediction problems. **Transactions of the ASME–Journal of Basic Engineering**, v. 82, n. Series D, p. 35–45, 1960.

KAMPEN, N. G. V. Elimination of fast variables. **Physics Reports**, Elsevier, v. 124, n. 2, p. 69–160, 1985.

KAMPEN, N. G. V. **Stochastic processes in physics and chemistry**. [S.l.]: Elsevier, 1992.

KAMPEN, N. V. **Stochastic processes in physics and chemistry**. [S.l.]: Elsevier, 2007. (North-Holland personal library).

KERSTING, H.; SULLIVAN, T. J.; HENNIG, P. Convergence rates of gaussian ode filters. **Statistics and computing**, Springer, v. 30, n. 6, p. 1791–1816, 2020.

LANGEVIN, P. Sur la théorie du mouvement brownien. **Compt. Rendus**, v. 146, p. 530–533, 1908.

LAPLACE, P. de. **Théorie analytique des probabilités**. [S.l.]: Mme Ve Courcier, 1814. (Théorie analytique des probabilités, v. 1).

LEE, J. S. Introduction to stochastic thermodynamics. Lecture presented at the ICTP-KIAS School on Statistical Physics for Life Sciences. 2022. Available at: <https://www.youtube.com/watch?v=MGk02Z6S_0M>.

LIU, H.; ONG, Y.-S.; SHEN, X.; CAI, J. When gaussian process meets big data: A review of scalable gps. **IEEE transactions on neural networks and learning systems**, IEEE, v. 31, n. 11, p. 4405–4423, 2020.

LJUNG, L.; GLAD, T. **Modeling of Dynamic Systems**. [S.l.]: PTR Prentice Hall, 1994. (Prentice-Hall information and system sciences series). ISBN 9780135970973.

MACKAY, D. J. **Information theory, inference and learning algorithms**. [S.l.]: Cambridge university press, 2003.

MARKOV, A. Extension of the law of large numbers to quantities, depending on each other (1906). reprint. **Journal Électronique d'Histoire des Probabilités et de la Statistique [electronic only]**, Séminaire d'Histoire du Calcul des Probabilités et de la Statistique, EHESS, Paris; Laboratoire de Probabilités et Modéles Aléatoires, Univerité Paris VI et VII, v. 2, n. 1b, p. Article 10, 12 p., electronic only–Article 10, 12 p., electronic only, 2006. Available at: <http://eudml.org/doc/128778>.

MARUYAMA, G. On the transition probability functions of the markov process. **Nat. Sci. Rep. Ochanomizu Univ**, v. 5, n. 1, p. 10–20, 1954.

MAXWELL, J. C. V. illustrations of the dynamical theory of gases.—part i. on the motions and collisions of perfectly elastic spheres. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, Taylor & Francis, v. 19, n. 124, p. 19–32, 1860.

MIYAZAKI, M.; YAMAZAKI, Y.; HASEGAWA, Y. Analysis of brownian motion by elementary school students. **The Physics Teacher**, v. 60, n. 6, p. 478–483, 09 2022. ISSN 0031-921X. Available at: <https://doi.org/10.1119/5.0059810>.

MOLER, C.; LOAN, C. V. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. **SIAM review**, SIAM, v. 45, n. 1, p. 3–49, 2003.

MONOD, J. The growth of bacterial cultures. **Annual review of microbiology**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 3, n. 1, p. 371–394, 1949.

MOU, C. Y.; LUO, J.-l.; NICOLIS, G. Stochastic thermodynamics of nonequilibrium steady states in chemical reaction systems. **The Journal of chemical physics**, AIP Publishing, v. 84, n. 12, p. 7011–7017, 1986.

MUNACK, A. Optimal feeding strategy for identification of monod-type models by fed-batch experiments. In: **Computer Applications in Fermentation Technology: Modelling and Control of Biotechnological Processes**. Springer Netherlands, 1989. p. 195–204. Available at: <https://doi.org/10.1007%2F978-94-009-1141-3_21>.

MUNSKY, B.; KHAMMASH, M. The finite state projection algorithm for the solution of the chemical master equation. **The Journal of chemical physics**, AIP Publishing, v. 124, n. 4, 2006.

MUNSKY, B.; KHAMMASH, M. The finite state projection algorithm for the solution of the chemical master equation. **The Journal of Chemical Physics**, AIP Publishing, v. 124, n. 4, jan 2006. Available at: <https://doi.org/10.1063%2F1.2145882>.

NETO, O. B.; MULAS, M.; CORONA, F. About the classical and structural controllability and observability of a common class of activated sludge plants. **Journal of Process Control**, v. 111, p. 8–26, 2022.

OKSENDAL, B. **Stochastic differential equations: an introduction with applications**. [S.l.]: Springer Science & Business Media, 2013.

PALEY, R.; WIENER, N. **Fourier Transforms in the Complex Domain**. [S.l.]: American Mathematical Society, 1934. (Amer. Math. Soc. Colloquium pub, v. 19). ISBN 9780821810194.

PARSI, S. S. Stochastic thermodynamics of learning parametric probabilistic models. **Entropy**, MDPI, v. 26, n. 2, p. 112, 2024.

PELITI, L.; PIGOLOTTI, S. **Stochastic Thermodynamics: An Introduction**. [S.l.]: Princeton University Press, 2021.

QIAN, H. Phosphorylation energy hypothesis: open chemical systems and their biological functions. **Annu. Rev. Phys. Chem.**, Annual Reviews, v. 58, n. 1, p. 113–142, 2007.

QIAN, H. Entropy demystified: The “thermo”-dynamics of stochastically fluctuating systems. **Methods in enzymology**, Elsevier, v. 467, p. 111–134, 2009.

QIAN, H.; GE, H. **Stochastic Chemical Reaction Systems in Biology**. Springer International Publishing, 2021. Available at: <https://doi.org/10.1007%2F978-3-030-86252-7>.

RACKAUCKAS, C.; NIE, Q. Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia. **Journal of Open Research Software**, Ubiquity Press, v. 5, n. 1, p. 15, 2017.

RASMUSSEN, C.; WILLIAMS, C. **Gaussian Processes for Machine Learning**. [S.l.]: MIT Press, 2005. (Adaptive Computation and Machine Learning series). ISBN 9780262182539.

REICHERT, P. User manual of AQUASIM 2.0 for the identification and simulation of aquatic systems. **Swiss Federal Institute for Environmental Science and Technology, Dubendorf, Switzerland**, 1998.

SARKKA, S.; HARTIKAINEN, J. Infinite-dimensional kalman filtering approach to spatio-temporal gaussian process regression. In: PMLR. **Artificial Intelligence and Statistics**. [S.l.], 2012. p. 993–1001.

SÄRKKÄ, S.; SOLIN, A. **Applied stochastic differential equations**. [S.l.]: Cambridge University Press, 2019.

SARKKA, S.; SOLIN, A.; HARTIKAINEN, J. Spatiotemporal learning via infinite-dimensional bayesian filtering and smoothing: A look at gaussian process regression through kalman filtering. **IEEE Signal Processing Magazine**, IEEE, v. 30, n. 4, p. 51–61, 2013.

SÄRKKÄ, S.; SVENSSON, L. **Bayesian Filtering and Smoothing**. [S.l.]: Cambridge University Press, 2023. (Institute of Mathematical Statistics Textbooks). ISBN 9781108926645.

SÄRKKÄ, S.; SVENSSON, L. **Bayesian filtering and smoothing**. [S.l.]: Cambridge university press, 2023.

SCHMIEDL, T.; SEIFERT, U. Stochastic thermodynamics of chemical reaction networks. **The Journal of chemical physics**, AIP Publishing, v. 126, n. 4, 2007.

SCHOBER, M.; SäRKKä, S.; HENNIG, P. A probabilistic model for the numerical solution of initial value problems. **Statistics and Computing**, Springer, v. 29, n. 1, p. 99–122, 2019.

SCHWEPPE, F. Evaluation of likelihood functions for gaussian signals. **IEEE transactions on Information Theory**, IEEE, v. 11, n. 1, p. 61–70, 1965.

SEIFERT, U. Stochastic thermodynamics: principles and perspectives. **The European Physical Journal B**, Springer, v. 64, p. 423–431, 2008.

SEIFERT, U. Stochastic thermodynamics of single enzymes and molecular motors. **Eur. Phys. J. E 34, 26, 2011**, 12 2010. Available at: <https://arxiv.org/pdf/1012.3119.pdf>.

SEIFERT, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. **Reports on progress in physics**, IOP Publishing, v. 75, n. 12, p. 126001, 2012.

SEIFERT, U. From stochastic thermodynamics to thermodynamic inference. **Annual Review of Condensed Matter Physics**, Annual Reviews, v. 10, p. 171–192, 2019.

SEKIMOTO, K. **Stochastic Energetics**. [S.l.]: Springer Berlin Heidelberg, 2010. (Lecture Notes in Physics). ISBN 9783642054112.

SHANNON, C. E. A mathematical theory of communication. **The Bell system technical journal**, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948.

SHARIFI, S.; MURTHY, S.; TAKÁCS, I.; MASSOUDIEH, A. Probabilistic parameter estimation of activated sludge processes using markov chain monte carlo. **Water Research**, Elsevier, v. 50, p. 254–266, 2014.

SHARIFI, S.; MURTHY, S.; TAKáCS, I.; MASSOUDIEH, A. Probabilistic parameter estimation of activated sludge processes using Markov Chain Monte Carlo. **Water Research**, v. 50, n. 1 March 2014, p. 254–266, 2014.

SHREVE, S. **Stochastic Calculus for Finance II: Continuous-Time Models**. [S.l.]: Springer New York, 2004. (Springer Finance). ISBN 9781441923110.

SMOLUCHOWSKI, M. V. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. **Annalen der physik**, Wiley Online Library, v. 326, n. 14, p. 756–780, 1906.

STENTOFT, P. A.; MUNK-NIELSEN, T.; VEZZARO, L.; MADSEN, H.; MØLLER, J. K.; MIKKELSEN, P. S. Towards model predictive control: Online predictions of ammonium and nitrate removal by using a stochastic asm. **Water Science and Technology**, v. 79, n. 1, p. 51 – 62, 2019.

STROGATZ, S. **Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering**. [S.l.]: CRC Press, 2018. ISBN 9780429961113.

TAO, T. **An Introduction to Measure Theory**. [S.l.]: American Mathematical Society, 2011. (Graduate studies in mathematics). ISBN 9780821869192.

TARANTOLA, A. **Inverse problem theory and methods for model parameter estimation**. [S.l.]: SIAM, 2005.

TOLMAN, R. C. **The principles of statistical mechanics**. [S.l.]: Courier Corporation, 1979.

TRONARP, F.; BOSCH, N.; HENNIG, P. **Fenrir: Physics-Enhanced Regression for Initial Value Problems**. 2023.

TRONARP, F.; KERSTING, H.; SäRKKä, S.; HENNIG, P. **Probabilistic Solutions To Ordinary Differential Equations As Non-Linear Bayesian Filtering: A New Perspective**. 2019.

VANROLLEGHEM, P. A.; LEE, D. S. On-line monitoring equipment for wastewater treatment processes: state of the art. **Water Science and Technology**, IWA Publishing, v. 47, n. 2, p. 1–34, 2003.

VIDYASAGAR, M. **Nonlinear Systems Analysis**. [S.l.]: Prentice-Hall International, 1993. (Prentice-Hall International editions). ISBN 9780136235132.

WANG, G.; SEVICK, E. M.; MITTAG, E.; SEARLES, D. J.; EVANS, D. J. Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales. **Physical review letters**, APS, v. 89, n. 5, p. 050601, 2002.

WANG, X.; FAN, Y.; HUANG, Y.; LING, J.; KLIMOWICZ, A.; PAGANO, G.; LI, B. Solving sensor reading drifting using denoising data processing algorithm (ddpa) for long-term continuous and accurate monitoring of ammonium in wastewater. **ACS ES&T Water**, ACS Publications, v. 1, n. 3, p. 530–541, 2020.

WIENER, N. Differential-space. **Journal of Mathematics and Physics**, v. 2, n. 1-4, p. 131–174, 1923.

WIENER, N. **Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications**. [S.l.]: The MIT press, 1949.

WOLFRAM, S. **A New Kind of Science**. Wolfram Media, 2002. ISBN 1579550088. Available at: <https://www.wolframscience.com>.

WOLPERT, D. H. The stochastic thermodynamics of computation. **Journal of Physics A: Mathematical and Theoretical**, IOP Publishing, v. 52, n. 19, p. 193001, 2019.

WU, M.; LYSY, M. **Data-Adaptive Probabilistic Likelihood Approximation for Ordinary Differential Equations**. 2023.

ZHU, A.; GUO, J.; NI, B.-J.; WANG, S.; YANG, Q.; PENG, Y. A novel protocol for model calibration in biological wastewater treatment. **Scientific Reports**, v. 5, p. 8493, 2015.

ZONTA, Ž. J.; FLOTATS, X.; MAGRÍ, A. Estimation of parameter uncertainty for an activated sludge model using bayesian inference: a comparison with the frequentist method. **Environmental technology**, Taylor & Francis, v. 35, n. 13, p. 1618–1629, 2014.

ZONTA, Z. J.; FLOTATS, X.; MAGRí, A. Estimation of parameter uncertainty for an activated sludge model using bayesian inference: a comparison with the frequentist method. **Environmental Technology**, v. 35, n. 1, p. 1618–1629, 2014.

ZWANZIG, R. **Nonequilibrium statistical mechanics**. [S.l.]: Oxford university press, 2001.

APPENDIX A – MEASURE-THEORETIC PROBABILITY THEORY

In this chapter the theory of probability is revisited to provide theoretical background to the subsequent discussions. In Section A.2 the notions of probability space are defined in light of the definitions presented on Appendix A.1. Section A.4 gives the definition of stochastic processes as it is used in this thesis, as well as examples.

A.1 A primer on Measure theory

As seen on the Chapter 2, to compute the trajectories of systems whose motion is ruled by differential equations there is always an integration involved. In Section 3 we are going to deal with functions that do not have the requirements to hold the existence and uniqueness presented. More specifically, with functions that are random and thus requiring a different kind of integration. Also, we are going to deal with probability spaces and operations on them. To do so, we introduce the basic concepts of measure theory.

As direct as may sound, measure theory deals with the measure of mathematical objects, for instance as the volume of physical world bodies (TAO, 2011). The approach in the physical world is to measure the volume of a body by partitioning it into finitely many components and then reassembling them to form a body whose volume is known. When no such partitioning is available, one may also give lower and upper bounds to its measure by using an inscribed and circumscribed body, respectively.

As we are going to deal with the measure of abstract objects as probability spaces, we need to define a measure in a rigorous way.

Definition A.1 (Measure (TAO, 2011)) *Let $\mathcal{X}$ be a set and $\mathscr{F}$ be a σ-algebra on $\mathcal{X}$. A measure is a function $\mu : \mathscr{F} \to [0, +\infty]$ satisfying:*

1. $\mu(\emptyset) = 0$,
2. *(Countable additivity) For any countable collection $\{\mathcal{A}_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathscr{F}$,*

$$\mu\left(\bigcup_{i=1}^{\infty} \mathcal{A}_i\right) = \sum_{i=1}^{\infty} \mu(\mathcal{A}_i).$$

To illustrate Definition A.1, consider these examples:

- **Counting measure**: Let $\mathcal{X}$ be any set, $\mathscr{F} = \wp(\mathcal{X})$, and $\mu(A) = |A|$ (number of elements in A)

- **Lebesgue measure**: Let $\mathcal{X} = \mathbb{R}$, $\mathscr{F}$ be the Lebesgue σ-algebra, and $\mu([a,b]) = b - a$

Notice that the original example with intervals $[a,b]$ alone does not satisfy Definition A.1 because:

- The collection of intervals is not a σ-algebra (not closed under complements or countable unions)
- The length function cannot be properly defined for arbitrary countable unions of intervals

Separately, we define outer measures:

Definition A.2 (Outer measure (TAO, 2011)) *An outer measure on a set* $\mathcal{X}$ *is a function* $\mu^* : \wp(\mathcal{X}) \to [0, +\infty]$ *satisfying:*

1. $\mu^*(\emptyset) = 0$
2. *(Monotonicity)* $A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$
3. *(Countable subadditivity)* $\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$

Until here, the theory posed is sufficient to integrate "well-behaved" functions, which is the case for the Riemann-Darboux integration that we commonly use. At the same time, we are going to deal with not so behaved functions (e.g. random functions), or even the integration in abstract sets as the ones that model the events when dealing with probabilities To define the kind of sets that we are going to measure, we start by defining an algebra. In layman's terms, an algebra is a collection of sets that is closed under the operations of union, intersection, and complement. More specifically, we have σ-algebras when we restrict those operations as we are going to see in the Definition A.3.

Definition A.3 (σ-algebra) *Given a set* $\mathcal{X}$*, a collection* $\mathscr{F} \subseteq \wp(\mathcal{X})$ *is called a* σ*-algebra if it satisfies the following properties*

1. $\mathcal{X} \in \mathscr{F}$,
2. *If* $\mathcal{A} \in \mathscr{F}$*, then its complement is also, i.e.* $\mathcal{A}^c := \mathcal{X} \setminus \mathcal{A} \in \mathscr{F}$,
3. *If* $\mathcal{A}_1, \mathcal{A}_2, \ldots \in \Sigma$*, then their union is also, i.e.* $\displaystyle\bigcup_{i=1}^{\infty} \mathcal{A}_i \in \Sigma$.

One can show using the properties of the Definition A.3 that a σ-algebra has the empty set in it and that is closed under the operation of countable[1] intersections (which is not required for generic (Boolean) algebras). The intersection of all σ-algebras that contain a given

---

[1] We use the definition of countability in the sense that a set is countable if it is finite or has the same cardinality as the set of natural numbers $\mathbb{N}$, i.e. we can assign a bijection between the set and $\mathbb{N}$.

σ-algebra $\mathscr{F}$ is called the generated σ-algebra by $\mathscr{F}$. We call the Borel σ-algebra $\mathscr{B}(\mathcal{X})$ the σ-algebra generated by the open subsets of a set $\mathcal{X}$. We give only more two definitions to be used in the next chapters. A product space is a Cartesian product of two or more measurable spaces. The product σ-algebra is the σ-algebra generated by the Cartesian product of the σ-algebras of the measurable spaces. Be the set $(\mathbb{R}^d)^{\mathcal{I}}$ the Cartesian product of d-dimensional Euclidean spaces $\mathbb{R}^d$ indexed by a set $\mathcal{I}$, the Borel σ-algebra on $(\mathbb{R}^d)^{\mathcal{I}}$ denoted as $\mathscr{B}(\mathbb{R}^d)^{\mathcal{I}}$ is the σ-algebra generated by the open sets of $(\mathbb{R}^d)^{\mathcal{I}}$.

As it is not our objective to develop this abstraction much further, it suffices to justify the use of the Borel σ-algebra by the fact that it avoids pathological cases leading to unmeasurable sets, which the converse case (i.e. the measurable ones) is discussed in the following.

Even we are not obliged to use the Lebesgue measure to perform the Lebesgue integral, we are going to use it because it is the most common measure used in the literature.

Definition A.4 (Lebesgue measure) *Let a set $\mathcal{X} = \mathbb{R}^d$ for some $d \in \mathbb{N}$ and a σ-algebra $\mathscr{F} = \mathscr{B}(\mathcal{X})$. Then, the Lebesgue measure is defined to be the (unique) map $\ell : \mathscr{F} \to \mathbb{R} \cup \{+\infty\}$ such that*

$$\ell([a_1,b_1] \times [a_2,b_2] \times ... \times [a_d,b_d]) = (b_1 - a_1)(b_2 - a_2)...(b_d - a_d), \textit{ for } a_i < b_i \in \mathbb{R}.$$

Some implications of the uniqueness on the Definition A.4 are that the Lebesgue measure is invariant under translation (i.e. $\ell(\mathcal{A}+x) = \ell(\mathcal{A})$ for all $\mathcal{A} \in \mathscr{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$), and that $\ell([0,1]^d) = 1$. Now, we possess a measure and a collection of sets to measure, we define the measurable sets.

Definition A.5 (μ-measurable sets) *Given a set $\mathcal{X}$ and a measure μ, a set $\mathcal{E}$ is said to be μ-measurable if for every set $\mathcal{A} \subset \mathcal{X}$, we have*

$$\mu(\mathcal{A}) = \mu(\mathcal{A} \cap \mathcal{E}) + \mu(\mathcal{A} \cap \mathcal{E}^c),$$

where $\mathcal{E}^c$ is the complement of $\mathcal{E}$.

It can be proven that the set $\mathcal{M}_\mu := \{\mathcal{E} \subset \mathcal{X}; \mathcal{E} \text{ is } \mu\text{-measurable } \}$ is a σ-algebra, and thus a set $\mathcal{X}$ equipped with a σ-algebra $\mathscr{F}$ constitutes a measurable space denoted $(\mathcal{X}, \mathscr{F})$. The elements of $\mathcal{M}_\mu$ are called measurable sets. We denote the triple $(\mathcal{X}, \mathscr{F}, \mu)$ as a measure space, i.e. the space of measurable sets equipped with a measure μ. Notice that one particular example the sets with measure zero. For instance, be a measure space $(\mathcal{X}, \mathscr{F}, \mu)$, a set $\mathcal{A} \in \mathscr{F}$ is called a null set if $\mu(\mathcal{A}) = 0$. Obviously, the empty set is a null set, but it is not the only one. This is important for what follows because we are going to consider that a condition (e.g. equality, convergence, etc.) holds "almost everywhere" (also denoted a.e.) if it holds for all $x \in \mathcal{X} \setminus \mathcal{Y}$, where $\mathcal{Y}$ is a null set.

We are now in position to define the Lebesgue integral. To do so, we need to define the concept of measurable functions.

Definition A.6 (Measurable functions) *Let the measurable spaces* $(\mathcal{X}, \mathscr{F})$ *and* $(\mathcal{Y}, \mathscr{G})$. *We have a (*$\mathscr{F}$*-)measurable function* $f : \mathcal{X} \to \mathcal{Y}$ *if the preimage of every measurable set in* $\mathcal{Y}$ *is a measurable set in* $\mathcal{X}$*, i.e. for every* $\mathcal{B} \in \mathscr{G}$*, we have* $f^{-1}(\mathcal{B}) \in \mathscr{F}$.

If it is not clear, we need measurable functions in order to assign a measure to the image of a set under a function which we are going to integrate over. Note that the construction so far is general enough to deal with a wide variety of sets, but from now on we are going to restrict ourselves to those which we are going to use later in the text, which are the real numbers $\mathbb{R}$ and the Borel σ-algebra $\mathscr{B}(\mathbb{R})$. The Riemann integral (given below in Definition A.9) assumes we are able to partition the domain of the function we are willing to integrate into finitely many intervals. This is not so different from the notion we have defined so far of an algebra (i.e. the interval of integration as being the union of the partitioned intervals), but it is a particular case. The Borel σ-algebra generalizes it to the case where we have countably many open intervals in $\mathbb{R}$ at the same time it avoids the pathological cases that could come up in such generic construction. The next step is to define a tool to integrate functions inside each element of the Borel σ-algebra.

Definition A.7 (Simple and indicator functions) *A measurable function* $s : \mathcal{X} \to \mathbb{R}$ *is called a simple function if it takes only a finite number of values* $s(\mathcal{X}) = \{s_n \in [0, \infty)\}_{n=1}^{N \in \mathbb{N}}$*, and thus it can be represented as*

$$s(x) = \sum_{z \in s(\mathcal{X})} z \mathbf{1}_{s^{-1}(\{z\})}(x), \text{ with } x \in \mathcal{X},$$

where $\mathbf{1}_{\mathcal{A}}$ *is the indicator function of the set* $\mathcal{A} \subseteq \mathcal{X}$*, such that*

$$\mathbf{1}_{\mathcal{A}}(a) = \begin{cases} 1, & \text{if } a \in \mathcal{A}, \\ 0, & \text{otherwise}. \end{cases}$$

Definition A.8 (Integral of a measurable non-negative function) *Let a measurable non-negative function* $f : \mathcal{X} \to \mathbb{R} \cap \{-\infty, \infty\}$*, i.e.* $f(x) \geq 0$ *for all* $x \in \mathcal{X}$*. The integral* $\int f d\mu$ *of* f *is the number*

$$\int f d\mu = \sup \left\{ \sum_{z \in s(\mathcal{X})} z \cdot \mu \left(s^{-1}(\{z\}) \right); 0 \leq s \leq f, s \textit{ is simple} \right\}.$$

The Definition A.8 is the Lebesgue integral of a non-negative function. The Lebesgue integral of a function that is not non-negative is defined as the difference between the Lebesgue integrals of its positive and negative parts. The Lebesgue integral is a generalization of the Riemann integral. For sake of comparison, we give the definition of the Riemann integral here.

Definition A.9 (Riemann integrability (TAO, 2011)) *Let* $[a,b]$ *be an interval of positive length, i.e.* $b > a$*. A tagged partition* $\mathcal{P} = \left((s_0, s_1, ..., s_n), (s_0^\star, s_1^\star, ..., s_n^\star)\right)$ *of* $[a,b]$ *is a finite sequence of real numbers such that* $a = s_0 < s_1 < ... < s_n = b$ *and* $s_{i-1} \leq s_i^\star \leq s_i$ *for* $i = 1, ..., n$*. We define* $s_i - s_{i-1} = \delta s_i$*. The quantity* $\Delta(\mathcal{P}) := \sup_{1 \leq i \leq \delta s_i} \delta s_i$ *is called the norm of the tagged partition* $\mathcal{P}$*. The Riemann sum* $\mathcal{R}(f, \mathcal{P})$ *of a function* f *with respect to* $\mathcal{P}$ *is defined as*

$$\mathcal{R}(f, \mathcal{P}) = \sum_{i=1}^{n} f(s_i^\star) \delta s_i.$$

The function f *is said to be Riemann integrable if there exists a real number, abbreviated as* $\int_a^b f(s) ds$ *and referred to as the Riemann integral of* f *on* $[a,b]$*, such that*

$$\int_a^b f(s) ds = \lim_{\Delta(\mathcal{P}) \to 0} \mathcal{R}(f, \mathcal{P}).$$

In other words, for every $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $\left| \mathcal{R}(f, \mathcal{P}) - \int_a^b f(s) ds \right| \leq \varepsilon$*, for every tagged partition* $\mathcal{P}$ *with* $\Delta(\mathcal{P}) < \delta$*.*

Using Definitions A.8 and A.9, we illustrate the difference between Lebesgue and Riemann integration with a classic example below.

Example 17 (Dirichlet function) *Consider* $f : [0,1] \to \mathbb{R}$ *defined as*

$$f(x) = \mathbf{1}_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \cap [0,1] \\ 0 & x \notin \mathbb{Q} \cap [0,1] \end{cases}$$

1. ***Riemann integral**: f is not Riemann integrable. For any partition, upper sums are 1 (since every interval contains rationals) and lower sums are 0 (since every interval contains irrationals), so they never converge to the same value.*
2. ***Lebesgue integral**: f is Lebesgue integrable with*

   $$\int_{[0,1]} f d\mu = 0$$

   since $\mathbb{Q} \cap [0,1]$ has Lebesgue measure zero ($\mu(\mathbb{Q} \cap [0,1]) = 0$).

The Example 17 shows that the Lebesgue integral is capable of handle uncountable many discontinuities if such discontinuities have measure zero.

A.2 Measure theory and Measure-theoretic probability

In general, the probability theory is introduced as the chance of discrete and finite events, as the toss of a coin or the roll of a die. However, in order to not extend too much, we are going to build the main results of the probability theory over the theory of measure presented in Section A.1. The main advantage is that we may collect the main results that we are going to use in the following sections without dealing with the difference of treatment between discrete and continuous versions.

The probability space consists of the triple $(\Omega, \mathscr{F}, P)$, being the sample space Ω, the σ-algebra $\mathscr{F}$ of subsets of Ω, and the probability measure P on $\mathscr{F}$. We already saw in the Definition A.3 that the σ-algebra is a family of subsets of Ω with the properties of closure under the set itself, complementing, and countable unions. Thus, with the aid of the measure P we can assign a probability value to each event in the σ-algebra $\mathscr{F}$, i.e. an element in the σ-algebra is a random event to which a probability value in $[0,1]$ is assigned. To the measure P we give the name of probability measure. Notice that, by Definition A.1, the probability measure inherits the properties of the measure, i.e. it is non-negative and countably additive. The only additional requirement of this measure is that the probability of the whole sample space is one, i.e. $P(\Omega) = 1$. It might occur that an event $\mathcal{A}^c \in \mathscr{F}$ satisfies $P(\mathcal{A}^c) = 0$. We saw in Section A.1 that this defines a null set, thus the event $\mathcal{A} = \Omega \setminus \mathcal{A}^c$ is said to hold almost everywhere, but in the case of the probability space we say that the event $\mathcal{A}$ holds almost surely (a.s.). We illustrate these concepts by applying the theory developed in Section A.1 on the following example.

Example 18 (Infinite, independent coin-toss space (SHREVE, 2004)) *Let a coin be tossed* *infinitely many times such we define the following σ-algebras and the respectivee probability measures.*

- *When $n = 0$ tosses are performed: Be the sample space*

  $$\Omega_\infty = \text{the set of infinite sequences of heads } (H) \text{ and tails } (T).$$

  Assume the probability of head in each toss is $p > 0$, equivalently the probability of tail is $q = 1 - p > 0$. Different tosses are independent, i.e. the outcome of one toss does not affect the outcome of another toss. We construct the probability measure corresponding to this random experiment step by step.

  Be the simpler case when no tosses have been performed, then the probability of no sequence to be drawn is $P(\emptyset) = 0$ and of any sequence is $P(\Omega_\infty) = 1$. These $2^{2^0} = 2$ sets form a σ-algebra $\mathscr{F}_0 = \{\emptyset, \Omega\}$, since (by Definition A.3) contains the empty set, is closed under complementing, and is closed under countable unions.

- *When $n = 1$ tosses are performed: Now, we define the sets*

  $$\mathcal{A}_H = \text{the set of all sequences that start with heads } H,$$
  $$\mathcal{A}_T = \text{the set of all sequences that start with tails } T.$$

  We denote $\omega = \omega_1\omega_2...$ an element in Ω_∞, where ω_n indicates the n-th toss of the coin. Thus, we write $\mathcal{A}_H = \{\omega; \omega_1 = H\}$ and $\mathcal{A}_T = \{\omega; \omega_1 = T\}$. We check if this choice forms a σ-algebra by checking the properties of Definition A.3. Let $\mathscr{F}_1 = \{\emptyset, \mathcal{A}_H, \mathcal{A}_T, \Omega_\infty\}$ be the σ-algebra generated by $\mathcal{A}_H$ and $\mathcal{A}_T$. See that $\mathcal{A}_H^c = \mathcal{A}_T$ thus $\mathcal{A}_H^c \in \mathscr{F}_1$ and $\mathcal{A}_H \cup \mathcal{A}_T = \Omega_\infty$ thus $\mathscr{F}_1$ is closed under complementing and countable unions. As the sets are then measurable define the probability measure P such that $P(\mathcal{A}_H) = p$ and $P(\mathcal{A}_T) = q$.

- *When $n = 2$ tosses are performed: Let us define the sets*

  $$\mathcal{A}_{HH} = \{\omega; \omega_1 = H, \omega_2 = H\}, \mathcal{A}_{HT} = \{\omega; \omega_1 = H, \omega_2 = T\},$$
  $$\mathcal{A}_{TH} = \{\omega; \omega_1 = T, \omega_2 = H\}, \mathcal{A}_{TT} = \{\omega; \omega_1 = T, \omega_2 = T\}.$$

  It is not hard to verify that they form the σ-algebra

  $$\mathscr{F}_2 = \left\{ \begin{array}{c} \emptyset, \mathcal{A}_H, \mathcal{A}_T, \mathcal{A}_{HH}, \mathcal{A}_{HT}, \mathcal{A}_{TH}, \mathcal{A}_{TT}, \mathcal{A}_{HH}^c, \mathcal{A}_{HT}^c, \mathcal{A}_{TH}^c, \mathcal{A}_{TT}^c, \\ \mathcal{A}_{HH} \cup \mathcal{A}_{TT}, \mathcal{A}_{HT} \cup \mathcal{A}_{TH}, \mathcal{A}_{HT} \cup \mathcal{A}_{TT}, \mathcal{A}_{HH} \cup \mathcal{A}_{TH}, \Omega_\infty \end{array} \right\}$$

  containing $2^{(2^2)} = 16$ sets. The probability measure P is defined such that $P(\mathcal{A}_{HH}) = p^2$, $P(\mathcal{A}_{HT}) = pq$, $P(\mathcal{A}_{TH}) = qp$, and $P(\mathcal{A}_{TT}) = q^2$.

Continuing this process, it is not difficult to see that the number of sets in the $\mathscr{F}_n$ *is* $2^{(2^n)}$ *and, for instance, a sequence of infinite tosses that only give heads is a subset of* $\mathcal{A}_H$, $\mathcal{A}_{HH}$, $\mathcal{A}_{HHH}$, *and so on, each having probability* $P(\mathcal{A}_H) = p$, $P(\mathcal{A}_{HH}) = p^2$, $P(\mathcal{A}_{HHH}) = p^3$, *etc. Thus, the probability measure of this sequence converges to zero. This is not particular to the heads. We can construct a* σ*-algebra* $\mathscr{F}_\infty$ *that contains all possible events of finite many tosses.*

A.3 Random variables and other definitions

Definition A.10 (Random variables and Distributions) *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space. A random variable is an* ($\mathscr{F}$*-)measurable function* $X : \Omega \to \mathbb{R}$, *i.e. for any Borel set* $\mathcal{B} \subset \mathscr{B}(\mathbb{R})$[2] *the set* $\{\omega \in \Omega : X(\omega) \in \mathcal{B}\} \in \mathscr{F}$. *The distribution measure of* X *is the probability measure* P_X *that assigns to each* $\mathcal{B} \in \mathscr{B}(\mathbb{R})$ *the probability* $P_X(\mathcal{B}) = P(\{X \in \mathcal{B}\})$.

By the Definition A.10 we see that distribution of a measure of a random variable makes sense for discrete random variables as well as for random variables that take continuous values. To illustrate this, consider the following example.

Example 19 (Stock prices (SHREVE, 2004)) *Let us define stock prices by*

$$S_0(\omega) = 4 \textit{ for all } \omega \in \Omega_\infty,$$

$$S_1(\omega) = \begin{cases} 8 \textit{ if } \omega_1 = H, \\ 2 \textit{ if } \omega_1 = T, \end{cases}$$

$$S_2(\omega) = \begin{cases} 16 \textit{ if } \omega_1 = \omega_2 = H, \\ 4 \textit{ if } \omega_1 \neq \omega_2, \\ 1 \textit{ if } \omega_1 = \omega_2 = T, \end{cases}$$

$$S_{n+1}(\omega) = \begin{cases} 2S_n(\omega) \textit{ if } \omega_{n+1} = H, \\ 0.5S_n(\omega)/2 \textit{ if } \omega_{n+1} = T, \end{cases}$$

where the probability space $(\Omega_\infty, \mathscr{F}_\infty, P)$ *is the same from the Example 18. The stock prices are random variables that depend on the outcomes of the coin tosses.*

---

[2] We may include the infinite when is needed, allowing the random variable to take values in the extended real line by doing $\mathcal{B} \subset \mathscr{B}(\mathbb{R} \cup \{-\infty, \infty\})$.

It is worth to mention that we may check the set $\mathcal{B}$ from which we evaluate the probability of the random variable X to take the values from could contain a single number. For instance, we may evaluate for $\mathcal{B} = \{4\}$ that $P(\{\omega \in \Omega : S_2 \in \mathcal{B}\})$ represents the events in which we obtain different outcomes in two consecutive tosses, i.e. $P(\{\mathcal{A}_{HT}, \mathcal{A}_{TH}\}) = pq + qp = 2pq$. We can also evaluate that for $\mathcal{B} = [2,5]$, $P(\{S_2 \in \mathcal{B}\}) = 2pq$. We stick to this notation omitting the fact that we are looking in the outcomes $\omega \in \Omega$. In the stock case of the Example 19 we say the "mass" of probability is lumped in a single value and not spread over an interval. Next, we give an example of a random variable that takes values in a continuous interval.

Example 20 (Uniform (Lebesgue) measure on the unit interval (SHREVE, 2004)) *We presented the Lebesgue measure in the Definition A.4. We define the probability of choosing a number in the unit interval* $[0,1]$ *as the Lebesgue measure of the interval, i.e.* $P([a,b]) = b - a$ *for* $0 \leq a < b \leq 1$.

The following definitions are given in the light of the notation we set so far, but they bring nothing new to the theory of probability presented in a standard course.

A.3.0.1 Distribution functions

Definition A.11 (Cumulative distribution function) *Given the probability space* $(\Omega, \mathscr{F}, P)$*, the cumulative distribution function (cdf) of a random variable X is the function* $P_X : \mathbb{R} \to [0,1]$ *defined by*

$$P_X(x) := P(\{X \leq x\}) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Also, $P_X(-\infty) = \lim_{x \to -\infty} P_X(x) = 0$ *and* $P_X(+\infty) = \lim_{x \to +\infty} P_X(x) = 1$.

If one chooses a probability space $(\Omega, \mathscr{F}, P) = (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), P)$ for some $d \in \mathbb{N}$, P_X is going to be the cumulative distribution function of an $\mathbb{R}^d$-valued random variable X. One thus may obtain the marginal distribution of each component X_i of $X = (X_1, X_2, ..., X_d)$ by evaluating $P_{X_i}(x_i) = P_X(\infty, ..., x_i, ..., \infty)$. It follows in general the definition below.

Definition A.12 (Marginal distribution function) *The marginal distribution function of a random variable* $X' = (X_i, ..., X_j)$ *from* $X = (X_1, ..., X_d)$ *in which* $\{i, ..., j\} \subset \{1, ..., d\}$ *is the function* $P_{X'} : \mathbb{R}^{|X'|} \to [0,1]$ *defined by*

$$P_{X'}(x_i, ..., x_j) := P_X(\infty, ..., x_i, ..., x_j, ..., \infty).$$

Definition A.13 (Probability density function) *The probability density function (pdf) of a random variable X is the function $p_X : \mathbb{R} \to [0,\infty)$ that satisfies*

$$P(\{X \in [a,b]\}) = \int_a^b p_X(x)dx, \text{ for } -\infty < a \leq b < \infty.$$

We may also refer to $P(\{X \in [a,b]\})$ as $P(a \leq X \leq b)$.

From this definition, we may use the same argument we used before to define the pdf for an $\mathbb{R}^d$-valued random variable X, obtaining

$$P_X(x_1,...,x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} p_X(\alpha_1,...,\alpha_d)\mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_d.$$

Notice also that the function p_X of the Definition A.13 is Lebesgue measurable (since it is Lebesgue integrable), thus by the fundamental theorem of calculus we collect the result

$$\frac{\partial^d}{\partial x_1 \cdots \partial x_d} P_X(x_1,...,x_d) = p_X(x_1,...,x_d).$$

Definition A.14 (Probability mass function) *The probability mass function (pmf) of a discrete random variable X is the function $p_X : \mathbb{R} \to [0,1]$ defined by*

$$P(\{X \in \mathcal{B}\}) = \sum_{x_i \in \mathcal{B}} p_X(x_i),$$

with $\mathcal{B} \in \mathbb{R}$. We may also refer to $P(\{X \in \mathcal{B}\}) = P(X \in \{x_i\})$ as $P(X = x_i) = p_X(x_i)$.

In the following definitions, we are going to operate with random variables.

Definition A.15 (Expectation of a random variable) *Let X be a random variable defined in a probability space $(\Omega, \mathscr{F}, P)$ with Ω finite. The expectation (or expected value) of X is defined as*

$$\mathbb{E}[X] = \int_\Omega X(\omega)dP(\omega),$$

if X is integrable, i.e. $\mathbb{E}[|X|] < \infty$.

Definition A.16 (Transformation of a random variable) *Let X be a random variable defined in a probability space $(\Omega, \mathscr{F}, P)$ with Ω finite. The transformation of X by a function $g : \mathbb{R}^d \to \mathbb{R}^m$ is the random variable $Y = g(X)$ whose expectation is*

$$\mathbb{E}[Y] = \int_\Omega g(X(\omega))dP(\omega).$$

We observe that Definition A.15 is a special case of the Definition A.16 when $d = m = 1$ and $g(x) = x$. Also, that the properties of linearity and monotonicity of the integral are valid for the expectation, i.e. for random variables X and Y and constants $a, b \in \mathbb{R}$ we have that

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y], \text{ and } X \leq Y \Rightarrow \mathbb{E}[X] \leq \mathbb{E}[Y].$$

Definition A.17 (Variance of a random variable) *Let X be a random variable defined in a probability space $(\Omega, \mathscr{F}, P)$. The variance of X is defined as*

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The number $\sigma = \sqrt{\mathrm{Var}[X]}$ is called the standard deviation of X[3].

Definition A.18 ((Central) moments of a random variable) *Let X be a random variable defined in a probability space $(\Omega, \mathscr{F}, P)$. The moment of order n of X is defined as*

$$\mathbb{E}[X^n] = \int_\Omega X^n(\omega) dP(\omega).$$

The central moment of order $n \in \mathbb{N}$ of X is defined as

$$\mathbb{E}[(X - \mathbb{E}[X])^n] = \int_\Omega (X(\omega) - \mathbb{E}[X])^n dP(\omega).$$

Definition A.19 (Covariance of random variables) *Let X and Y be random variables defined in a probability space $(\Omega, \mathscr{F}, P)$. The covariance of X and Y is defined as*

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Definition A.20 (Characteristic function of a random variable) *Let X be a random variable defined in a probability space $(\Omega, \mathscr{F}, P)$. The characteristic function of X is defined as*

$$\varphi_X(t) = \mathbb{E}[e^{itX}], \textit{ for } t \in \mathbb{R}.$$

---

[3] We may also refer to the standard deviation by the letter σ, depending on the context of the text do not confuse with σ of σ-algebras

A.3.0.2 Limit theorems and convergence concepts

We recover the discussion on convergence we introduced in the beginning of this chapter. We present the following theorems that are going to be useful in the following sections.

Consider X and X_n with $n \in \mathbb{N}$ random variables defined in a probability space $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), P)$. For this probability space, we define the following convergence concepts.

Definition A.21 (Almost sure (certainly) convergence (ARNOLD, 1974)) *If there exists a set $N \in \mathscr{B}(\mathbb{R}^d)$ with $P(N) = 0$ such that, for all $\omega \notin N$, the sequence $\{X_n(\omega)\}$ converges to $X(\omega)$ in the usual sense. Then we say that X_n converges to X almost surely (a.s.), or with probability one. We may also use*

$$\text{as-}\lim_{n\to\infty} X_n = X.$$

The Definition A.21 is nothing more than a consequence of the use of Measure theory for probability, since the null sets defined in the Definition A.1 are here the sets of probability zero. Sometimes it is hard to establish almost surely convergence, the following weaker concepts are given to deal with that.

Definition A.22 (Convergence in probability (ARNOLD, 1974)) *The sequence $\{X_n\}$ converges to X in probability if, for all $\varepsilon > 0$,*

$$\lim_{n\to\infty} P(\{\omega; |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

*We may also use the notation "$P-$" or "*st-*" as in*

$$\text{st-}\lim_{n\to\infty} X_n = X.$$

In order to define the convergence in mean square, we need to introduce the concept of L^p spaces. For short, consider the following set of measurable functions

$$\mathfrak{F}^p(\Omega, \mathscr{F}, P) := \{X : \Omega \to \mathbb{R}^d; \mathbb{E}[|X|^p] < \infty\},$$

that is the set of all random variables that have finite p-th moment. We say that any two functions in $\mathfrak{F}^p$ that differs between them only on sets $\mathcal{A} \in \mathscr{F}$ with $P(\mathcal{A}) = 0$, belongs to the same equivalence class. Finally, we define the L^p space as the set of all equivalence classes of $\mathfrak{F}^p$, i.e. the random variables that differ only on sets of probability zero and that have finite p-th moment.

Definition A.23 (Convergence in mean square (ARNOLD, 1974)) *Let a random variable* X *and a sequence of random variables* $\{X_n\}$ *with* $X, X_n \in L^p$. *Then the sequence* $\{X_n\}$ *converges to* X *in* p*-th mean if*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

More specifically for $p = 2$, *we use the notation*

$$\text{ms-}\lim_{n\to\infty} X_n = X,$$

and we say that the sequence $\{X_n\}$ *is convergent in mean square to* X.

Definition A.24 (Theorem on dominated convergence (ARNOLD, 1974)) *Let* $\{X_n\}$ *be a sequence of random variables and* Y *a random variable such that* $|X_n| \leq Y$ *for all* n *and* $\mathbb{E}[Y] < \infty$. *If* $X_n \to X$ *almost surely, then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n].$$

Definition A.25 ((Strong) law of large numbers (ARNOLD, 1974)) *For an arbitrary sequence* $\{\mathcal{A}_n\}$ *of events in a probability space* $(\Omega, \mathscr{F}, P)$, *the set* $\mathcal{A} = \{\omega : \omega \in \mathcal{A}_n \text{ for infinitely many } n\}$ *is also an event. The Borel-Cantelli lemma gives us that if* $\sum_n P(\mathcal{A}_n) < \infty$, *then* $P(\mathcal{A}) = 0$. *Moreover, if* $\{\mathcal{A}_n\}$ *is independent and* $\sum_n P(\mathcal{A}_n) = \infty$, *then* $P(\mathcal{A}) = 1$. *Consider*

$$S_n = X_1 + ... + X_n.$$

The strong law of large numbers gives us that

$$\text{as-}\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}[X_1]$$

if $\{X_n\}$ *is a sequence of independent and identically distributed of random variables.*

A.3.0.3 Conditional probabilities and conditional expectations

Definition A.26 (Conditional probability (ARNOLD, 1974)) *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space. The conditional probability of an event* $\mathcal{A} \in \mathscr{F}$ *given an event* $\mathcal{B} \in \mathscr{F}$ *with* $P(\mathcal{B}) > 0$ *is defined as*

$$P(\mathcal{A} \mid \mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})}.$$

The condition that $P(\mathcal{B}) > 0$ may be restrictive in some cases, for instance if we have a random variable X that takes values in a continuous interval, the probability of the event $\mathcal{B} = \{X = x\}$ is zero. In this case, we may use the concept of conditional expectation to define the conditional probability (ARNOLD, 1974).

Definition A.27 (Conditional expectation (OKSENDAL, 2013)) *Let X be a random variable defined in the probability space $(\Omega, \mathscr{F}, P)$. If $\mathscr{G} \subset \mathscr{F}$ is a σ-algebra then the conditional expectation of X given $\mathscr{G}$ denoted by $\mathbb{E}[X|\mathscr{G}]$ is a random variable that*

1. *is $\mathscr{G}$-measurable,*
2. *satisfies the property*

$$\int_{\mathcal{G}} \mathbb{E}[X|\mathscr{G}]dP = \int_{\mathcal{G}} XdP, \textit{ for all } \mathcal{G} \in \mathscr{G}.$$

Notice in Definition A.27 that $\mathbb{E}[X|\mathscr{G}]$ is unique almost surely due to the Radon-Nikodym theorem (see Appendix B in (OKSENDAL, 2013) for details). Thus, it is itself the random variable that best approximates X with the information contained in $\mathscr{G}$. We define then the conditional probability distribution of X given $\mathscr{G}$ the function $P(X \in \mathcal{B} \mid \mathscr{G})$ on ω (by fixing ω we fix $X(\omega)$) and $\mathcal{B}$ satisfying

$$P(\mathcal{C} \cap \{X \in \mathcal{B}\}) = \int_{\mathcal{C}} P(X \in \mathcal{B} \mid \mathscr{G}) \mathrm{d}P(\omega),$$

for all $\mathcal{C} \in \mathscr{G}$. To condition on a different random variable, let us define Y on the measurable space $(\Omega', \mathscr{F}')$ taking $\mathscr{G} = \mathscr{F}(Y)$ the σ-algebra generated by the Ω'-valued random variable Y. By the conditional expectation just defined, we have that

$$\mathbb{E}[X \mid \mathscr{G}] = \mathbb{E}[X \mid Y].$$

Next, we assume that there exists a function h for every $\mathscr{F}(Y)$-measurable random variable Z such that $Z = h(Y)$ (a.s.). This means that there exists a function h on Ω' such that $\mathbb{E}[X \mid Y] = h(Y)$ (a.s.). We can also write this as

$$\mathbb{E}[X \mid Y = y] = h(y).$$

We now assume that there exists a function $p_{X|Y}$ such that

$$P(X \in \mathcal{B} \mid Y = y) = \int_{\mathcal{B}} p_{X|Y}(x, y) \mathrm{d}x,$$

with $X = x$. This function is called conditional probability density of X given Y. It may be obtained also the called law of total probability as

$$P(X \in \mathcal{B}) = \int_{\Omega'} P(X \in \mathcal{B} \mid Y = y) \mathrm{d}P_Y(\omega') = \int_{\Omega'} p_{X|Y}(x,y) \mathrm{d}P_Y(\omega') =: \int_{\mathcal{B}} p_X(x) \mathrm{d}x,$$

where P_Y is the probability measure of the random variable Y.

Example 21 (From Arnold (1974)) *Let X and Y be two random variables defined in the sample spaces Ω and Ω', respectively, such that $\Omega = \mathbb{R}^d$ and $\Omega' = \mathbb{R}^m$. Their joint density function is $p_{X,Y}(x,y)$. If $P(X \in \mathcal{B} \mid Y = y)$ has a density function $p_{X|Y}(x,y)$ which, for all y such that the marginal density*

$$p_Y(y) = \int_{\mathbb{R}^d} p_{X,Y}(x,y) dx,$$

of Y is positive, has the form

$$p_{X|Y}(x,y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}.$$

A.4 Stochastic processes

To the collection of random variables indexed in time, we give the specific name of stochastic process. The definition follows.

Definition A.28 (Stochastic processes) *Let $(\Omega, \mathscr{F}, P) = (\mathbb{R}^d)$ be a probability space, and $t \in \mathcal{I} \subset \mathbb{R}$. A stochastic process is a collection of random variables $\{X_t\}_{t \in [0,T]}$ defined in the probability space $(\Omega, \mathscr{F}, P)$.*

Notice that X_t could be also called a random function, but since we are going to deal with processes (or functions) indexed in time, we stick with the name. For a fixed $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is a sample path of the stochastic process. If the set $\mathcal{I}$ is countable, we say that the stochastic process is discrete, otherwise, it is continuous. In the discrete case, the stochastic process $\{X_t\}_t; t \in \mathcal{I}\}$ in the probability space $(\Omega, \mathscr{F}, P)$ has finite-dimensional distributions

$$\begin{aligned} P(\{X_{t_1} \leq x_1\}) &:= P_{X_{t_1}}(x), \\ P(\{X_{t_1} \leq x_1, X_{t_2} \leq x_2\}) &:= P_{X_{t_1}, X_{t_2}}(x_1, x_2), \\ &\vdots \\ P(\{X_{t_1} \leq x_1, X_{t_2} \leq x_2, \ldots, X_{t_n} \leq x_n\}) &:= P_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, x_2, \ldots, x_n), \end{aligned}$$

for all $t_1, t_2, ..., t_n \in \mathcal{I}$. It is worth to mention that the Kolmogorov's fundamental theorem states that the finite-dimensional distributions of a stochastic process are sufficient to determine the stochastic process. This means that if we know the finite-dimensional distributions, we can reconstruct the stochastic process. A stochastic process is said stationary process (in a strict sense) if the finite-dimensional distributions are invariant under time shifts, i.e. for all $t_1, t_2, ..., t_n \in \mathcal{I}$ and $s \in \mathcal{I}$ we have

$$P_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_1, x_2, ..., x_n) = P_{X_{t_1+s}, X_{t_2+s}, ..., X_{t_n+s}}(x_1, x_2, ..., x_n).$$

It follows also that a strictly stationary process has $\mathbb{E}[X_t] = \mu$ and $\mathrm{Cov}[X_t, X_s] = C(t-s)$ for all $t, s \in \mathcal{I}$, where μ is a constant and C is a function of $t-s$. For a stochastic which all distributions do not match under time shifts, but only have the mean constant and the covariance dependent on the time difference $t-s$, we give the name of weak-sense stationary process. Note that every strict sense stationary process is also stationary in a weak sense, but the converse is not true.

Example 22 (Gaussian process (RASMUSSEN; WILLIAMS, 2005)) *Stochastic processes X_t in which any collection of its random variables are Gaussian distributed is called a Gaussian process and are completely specified my its mean $m(t) = \mathbb{E}[X_t]$ and covariance $k(t,t') =$* $\mathrm{Cov}[X_t, X_{t'}]$ *functions*

$$X_t \sim \mathcal{GP}(m(t), k(t,t')).$$

It is worth to note that this definition is not restricted to stochastic process but also to general random processes which may not be indexed in time. In Figure 13 some realizations of Gaussian processes are presented.

Recap the Example 18, where we have the σ-algebras $\mathscr{F}_n$ that contain all possible events of finite many tosses. Notice that between the tosses n and m with $n < m$, we have that $\mathscr{F}_n \subset \mathscr{F}_m$. For the discussion to follow, we interpret this as the information of the tosses n is contained in the tosses m, thus as the number of tosses increases the information increases. We define this concept formally in the following definition.

Definition A.29 (Filtration (SHREVE, 2004)) *Let Ω be a nonempty set and T a fixed positive number. Assume that for each $t \in [0,T]$ we have a σ-algebra $\mathscr{F}_t$ such that $\mathscr{F}_s \subset \mathscr{F}_t$ for $0 \leq s \leq t \leq T$. The collection $\{\mathscr{F}_t\}_{t \in [0,T]}$ is called a filtration.*

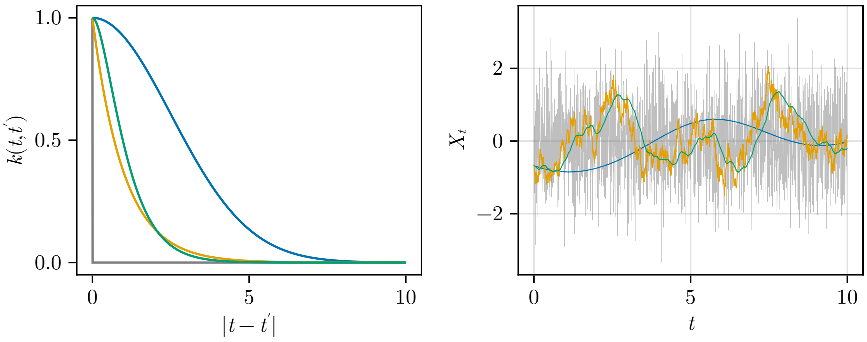


Figure 13 – On the left the values of the Matérn covariance function (RASMUSSEN; WILLIAMS, 2005) $k(t,t')$ vs. the time difference $|t-t'|$. Each color represents different parameters for the Matérn function. On the right, a realization of the process with zero mean and for different parameters of the Matérn covariance function.

An application of filtration is to the theory of martingales, i.e. a stochastic process $\{X_t\}_{t\in[0,T]}$ such that $\mathbb{E}[X_t|\mathscr{F}_s]=X_s$ for $0\leq s<t\leq T$. This means that the expected value of the process at time t given the information up to time s is equal to the value of the process at time s. This property is important in many applications, including finance and gambling once they represent fair games.

A.4.0.1 Markov processes

The next important concept is the Markov property, which states that the "future" of a general process is independent of the "past" given the "present". Such concept is similar to the causality principle in physics which we use to model the dynamics of a system[4] (ARNOLD, 1974). We define the Markov property in the following definition.

Definition A.30 (Markov process (ARNOLD, 1974)) *A stochastic process* $\{X_t\}_{t\in[t_0,T]}$ *with* $[t_0,T]\subset[0,\infty)$ *defined in the probability space* $(\Omega,\mathscr{F},P)$ *with* $\Omega=(\mathbb{R}^d)^{[t_0,T]}$ *and* $\mathscr{B}(\mathbb{R}^d)^{[t_0,T]}$ *is said to be a Markov process if, the Markov property holds, i.e. for all* $t_0\leq s\leq t\leq T$ *and all* $\mathcal{B}\in\mathscr{B}(\mathbb{R}^d)$ *it holds*

$$P(X_t\in\mathcal{B}|\mathscr{F}_s)=P(X_t\in\mathcal{B}|X_s),\ a.s.$$

An example of a Markov process is the Wiener process (or Brownian motion) which is a continuous-time stochastic process (OKSENDAL, 2013).

Example 23 (Wiener process) *A stochastic process* W_t *for* $t\leq 0$ *with the following properties*

---

[4] Of course the analogy is made ignoring systems with memory which depend on time delayed states. Markovian systems are said "memoryless".

(i) Initial distribution $p_{X_0}(x) = \delta(x)$*;*

(ii) $W_t - W_s \sim \mathcal{N}(0, \sigma^2(t-s))$*, for* $0 \geq s \geq t$*;*

(iii) $W_t - W_s$ *is independently distributed for any pair* t, s *with* $t \geq s$*.*

We may want to obtain the probability density function of the Markov process. For this, we need to define the transition probability of the Markov process.

Definition A.31 (Transition probability (of a Markov process) (ARNOLD, 1974)) *Let* $\{X_t\}_{t\in[t_0,T]}$ *be a Markov process defined in the probability space* $(\Omega, \mathscr{F}, P)$ *with* $\Omega = (\mathbb{R}^d)^{[t_0,T]}$ *and* $\mathscr{B}(\mathbb{R}^d)^{[t_0,T]}$*. The transition probability of the Markov process is defined as*

$$P(X_t \in \mathcal{B} | X_s = x),$$

which is the probability that X_t *will be in the set* $\mathcal{B}$ *at a time* t *given that the process is in the state* x *at a time* s*, with* $s \leq t$*.*

The function $P(X_t \in \mathcal{B} | X_s = x)$ has arguments $s, t \in [t_0, T]$ with $s \leq t$, $x \in \mathbb{R}^d$, and $\mathcal{B} \in \mathscr{B}(\mathbb{R}^d)$ and for fixed s, t and $\mathcal{B}$, it is a probability on $\mathscr{B}(\mathbb{R}^d)$.

The Markov process allows us to obtain its probability on different times given the probability at a time precedent to it. This is possible through the Chapman-Kolmogorov equation defined below.

Definition A.32 (Chapman-Kolmogorov equation (ARNOLD, 1974)) *Let* $\{X_t\}_{t\in[t_0,T]}$ *be a Markov process defined in the probability space* $(\Omega, \mathscr{F}, P)$ *with* $\Omega = (\mathbb{R}^d)^{[t_0,T]}$ *and* $\mathscr{B}(\mathbb{R}^d)^{[t_0,T]}$*. The Chapman-Kolmogorov equation is given by*

$$P(X_t \in \mathcal{B} \mid X_s = x) = \int_{\mathbb{R}^d} P(X_t \in \mathcal{B} \mid X_u = y) P(X_u \in \mathrm{d}y \mid X_s = x), \text{(a.s.)},$$

for all $t_0 \leq s \leq u \leq t \leq T$*.*

We may also define the respective density function for the transition probability. We require a function $p_{X_t|X_s}(y,x)$ that satisfies

$$P(X_t \in \mathcal{B} \mid X_s = x) = \int_{\mathcal{B}} p_{X_t|X_s}(y,x)\mathrm{d}y.$$

This allows us to write the Chapman-Kolmogorov equation in terms of the transition probability density function

$$p_{X_t|X_s}(y,x) = \int_{\mathbb{R}^d} p_{X_t|X_u}(y,z) p_{X_u|X_s}(z,x)\mathrm{d}z.$$

We can now obtain all finite-dimensional distributions of the Markov process given the transition probability density function. Let us assume that the probability of the process being initially in a set $\mathcal{A}$ is $P(X_{t_0} \in \mathcal{A})$, then the finite-dimensional distributions of the Markov process are given by

$$
\begin{aligned}
&P(X_{t_1} \in \mathcal{B}_1, X_{t_2} \in \mathcal{B}_2, ..., X_{t_n} \in \mathcal{B}_n) = \\
&\int_{\mathbb{R}^d} \int_{\mathcal{B}_1} \cdots \int_{\mathcal{B}_n} \big[P(X_{t_n} \in \mathcal{B}_n \mid X_{t_{n-1}} = x_{n-1}) P(X_{t_{n-1}} \in \mathrm{d}x_{n-1} \mid X_{t_{n-2}} = x_{n-2}) \cdots \\
&P(X_{t_1} \in \mathrm{d}x_1 \mid X_{t_0} = x_{t_0}) P(X_{t_0} \in \mathrm{d}x_{t_0}) \big],
\end{aligned}
$$

with $t_0 \leq t_1 \leq t_2 \leq ... \leq t_n \leq T$ and $\mathcal{B}_i \in \mathscr{B}(\mathbb{R}^d)$ for $i = 1, 2, ..., n$. We may also write the finite-dimensional distributions in terms of the transition probability density function as

$$
\begin{aligned}
&P(X_{t_1} \in \mathcal{B}_1, \ldots, X_{t_n} \in \mathcal{B}_n) = \\
&\quad \int_{\mathcal{B}_1} \cdots \int_{\mathcal{B}_n} \int_{\mathbb{R}^d} p_{X_{t_n} | X_{t_{n-1}}}(x_n, x_{n-1}) \cdots p_{X_{t_1} | X_{t_0}}(x_1, x_0) P(X_{t_0} \in \mathrm{d}x_{t_0}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n.
\end{aligned}
$$

Example 24 (White noise) *Let* $\{X_t\}_{t \in \mathcal{I}}$ *be a stochastic process with* $\mathbb{E}[X_t] = 0$, X_t *independent of* X_s, $\mathrm{Cov}[X_t, X_s] = 0$, *if* $t \neq s$, *and* $\mathrm{Cov}[X_t, X_t] = 1$. *Then we say that the process is a white noise process.*

We can show that there is no measurable process X_t equivalent to the white noise defined in the Definition 24 (KALLIANPUR, 2013), although it is a useful concept in the theory of stochastic processes. Differently, we can define a generalized stochastic process with such settings that has the measurability assured, but this subject is outside the scope of this text (see (ARNOLD, 1974)).

A.4.0.2 Master equation

We are going to discuss another way to represent Markov processes which is more conventional than the Chapman-Kolmogorov equation, the Master equation (KAMPEN, 1992). Let us consider a Markov process $\{X_t\}$ and the Chapman-Kolmogorov equation in Definition A.32 between the time instants t and $t + \tau$ with $\tau > 0$. We expand $P(X_{t+\tau} \in \mathcal{B} \mid X_t = y)$ in a Taylor series up to the first-order terms, giving

$$
P(X_{t+\tau} \in \mathcal{B} \mid X_t = y) = (1 - a_0 \tau) \mathbf{1}_{\{y \in \mathcal{B}\}} - \tau \int_{\mathcal{B}} W(x, y) \mathrm{d}x + o(\tau),
$$

where $o(\tau)$ stands for terms such that $o(\tau)/\tau \to 0$ as $\tau \to 0$, $W(x, y)$ is the transition probability rate between $X_{t+\tau}$ and X_t. The coefficient $(1 - a_0 \tau)$ represents the probability that the process does not change from a state X_t to any other state during τ. It is defined by satisfying

$$a_0(y) = \int_{\mathbb{R}^d} W(x,y)\mathrm{d}x.$$

By substituting $P(X_{t+\tau} \in \mathcal{B} \mid X_t = y)$ into the Chapman-Kolmogorov equation in Definition A.32 we obtain taking limit when $\tau \to 0$ that

$$\frac{\partial P(X_t \in \mathcal{B})}{\partial t} := \int_{\mathbb{R}^d} W(\mathcal{B},y)P(X_t \in \mathrm{d}y) - \int_{\mathcal{B}} a_0(x)P(X_t \in \mathrm{d}x).$$

Specially, when the stochastic process is on the discrete space (i.e. $\Omega = \mathbb{N}^d$ or $\Omega = \mathbb{Z}^d$), we have that the probability of $X_t = x$ chances with the rate

$$\frac{\partial P(X_t)}{\partial t} := \sum_{X_t} W(X_t,X_t')P(X_t') - W(X_t',X_t)P(X_t). \tag{A.1}$$

We can draw some specific probabilities from the Equation (A.1). For instance, if $P(X_t)$ is such that $\frac{\partial P(X_t)}{\partial t} = 0$ for all X_t, this probability distribution is said the stationary probability distribution. A more restrictive case that is worth to mention is that when $P(X_t) = \pi_X$ such that

$$\pi_X W(X_t,X_t') = \pi_{X'} W(X_t,X_t') \tag{A.2}$$

for all possible states X. Notice that this implies that π_X is also a stationary probability distribution, but in this case if we have a system modeled by a Markov process obeying such relation, this system is said reversible. This relation is known as detailed balance.

APPENDIX B – PARAMETERS OF THE ASM NO. 1

Symbol	Description	Units
S_I	Soluble inert organic matter	$g\,\mathrm{COD} \cdot m^{-3}$
S_S	Readily biodegradable substrate	$g\,\mathrm{COD} \cdot m^{-3}$
X_I	Particulate inert organic matter	$g\,\mathrm{COD} \cdot m^{-3}$
X_S	Slowly biodegradable substrate	$g\,\mathrm{COD} \cdot m^{-3}$
X_{BH}	Active heterotrophic biomass	$g\,\mathrm{COD} \cdot m^{-3}$
X_{BA}	Active autotrophic biomass	$g\,\mathrm{COD} \cdot m^{-3}$
X_P	Particulate products arising from biomass decay	$g\,\mathrm{COD} \cdot m^{-3}$
S_O	Dissolved oxygen	$g\,O_2 \cdot m^{-3}$
S_{NO}	Nitrate and nitrite nitrogen	$g\,N \cdot m^{-3}$
S_{NH}	$NH_4^+ + NH_3$ nitrogen	$g\,N \cdot m^{-3}$
S_{ND}	Soluble biodegradable organic nitrogen	$g\,N \cdot m^{-3}$
X_{ND}	Particulate biodegradable organic nitrogen	$g\,N \cdot m^{-3}$
S_{ALK}	Alkalinity	$\mathrm{mol}\,HCO_3^- \cdot m^{-3}$

Table 3 – ASM1 - State variables

Symbol	Description	Value	Units
Y_A	Autotrophic yield	0.24	g X_{BA} COD formed $\cdot$ (g N oxidised)$^{-1}$
Y_H	Heterotrophic yield	0.67	g X_{BH} COD formed $\cdot$ (g COD utilized)$^{-1}$
f_P	Fraction of biomass to particulate products	0.08	g X_P COD formed $\cdot$ (g N decayed)$^{-1}$
i_{XB}	Fraction nitrogen in biomass	0.08	g $N \cdot$ (g COD)$^{-1}$ in biomass
i_{XP}	Fraction nitrogen in particulate products	0.06	g $N \cdot$ (g COD)$^{-1}$ in X_P
μ_H	Maximum heterotrophic growth rate	4.00	d^{-1}
K_S	Half-saturation (heterotrophic growth)	10.0	g COD m^{-3}
K_{OH}	Half-saturation (heterotrophic oxygen)	0.20	g O_2 m^{-3}
K_{NO}	Half-saturation (nitrate)	0.50	g $NO_{3^-}N$ m^{-3}
b_H	Heterotrophic decay rate	0.30	d^{-1}
ν_g	Anoxic growth rate correction factor	0.80	-
ν_h	Anoxic hydrolysis rate correction factor	0.80	-
k_h	Maximum specific hydrolysis rate	3.00	g X_S (g X_{BH} COD d)$^{-1}$
K_X	Half-saturation (hydrolysis)	0.10	g X_S (g X_{BH} COD)$^{-1}$
μ_A	Maximum autotrophic growth rate	0.50	d^{-1}
K_{NH}	Half-saturation (autotrophic growth)	1.00	g $NH_{4^-}N$ m^{-3}
b_A	Autotrophic decay rate	0.05	d^{-1}
K_{OA}	Half-saturation (autotrophic oxygen)	0.40	g O_2 m^{-3}
k_a	Ammonification rate	0.05	g X_S (g COD d)$^{-1}$

Table 4 – ASM1 - Parameters

INDEX

μ-measurable, 91
σ-algebra, 90

adapted, 36
algebra, 90
almost everywhere, 92
almost surely, 94, 100
average entropy flux rate, 58
average entropy production rate, 58

Bayesian filtering equations, 64
Boltzmann factor, 57
Borel σ-algebra, 91
Borel-Cantelli lemma, 101
boundaries, 45
Brownian motion, 32

central moment, 99
Chapman-Kolmogorov equation, 106
characteristic function, 99
chemical equilibrium, 50
chemical Langevin equation, 54
chemical master equation, 52
chemical reaction networks, 52
conditional expectation, 102
conditional probability, 101
conditional probability density, 103
conditional probability distribution, 102
convergence in mean square, 101
convergence in probability, 100
countable, 90
covariance, 99
cumulative distribution function, 97

detailed balance, 57, 108
Deterministic, 19
distribution measure, 96
dynamic, 19

Euler method, 30
Euler-Maruyama method, 42
expectation, 98
expected value, 98

filtering, 64
filtration, 104
Fokker-Planck-Kolmogorov equation, 40

Gaussian process, 104
Gaussian process kernels, 65
generalized stochastic process, 107
generated σ-algebra, 91
Gillespie algorithm, 54

heat bath, 35, 45
heat capacity, 46
heat reservoir, 45

identification, 18
indicator function, 92
infinitesimal generators, 40
inverse problems, 61
irreversible, 46
Itō process, 41
Itō's lemma, 38

Jacobian, 29

Kalman filter, 64

Langevin equation, 32, 34
law of mass action, 49
law of total probability, 103
Lebesgue measure, 91
likelihood, 61
linear dynamical system, 28

macroscopic, 47
marginal distribution, 97
marginal distribution function, 97
marginal posterior density, 64
Markov process, 105
Markov property, 105
martingale, 39
Master equation, 107
master equation, 52
Matérn function, 66
Maxwell-Boltzmann statistics, 56
measurable
    function, 92
    sets, 91
    space, 91
measure, 89
measure space, 91
mesoscopic, 47
mesoscopic dynamics, 48
mesoscopic scales, 50
microscopic, 47
model, 18
moment, 99

nonautonomous system, 28
nonequilibrium entropy, 59
norm, 93
null set, 92

ODE
    analytical solution, 24
    general solution, 24
    homogeneous, 24
    inhomogeneous, 24
    initial condition, 24
    particular solution, 24
    separable, 25
ordinary differential equations, 23
outer measure, 90

partial differential equations, 23
Picard iteration, 26
Picard-Lindelöf theorem, 26
posterior probability, 61
Probabilistic Numerics, 68
probability density function, 98
probability mass function, 98
probability measure, 94
product σ-algebra, 91
propensity function, 52

random event, 94
random variable, 96
reaction probability rate, 51
reverse reaction, 49
reversed reactions, 56
Riemann integrable, 93
Riemann integral, 93
Riemann sum, 93

Runge-Kutta method, 30

sample space, 94
simple function, 92
Squared Exponential function, 66
standard deviation, 99
state-space representation, 27
Static, 19
stationary probability distribution, 108
stationary stochastic process, 104
step function, 37
stochastic, 19
  integral, 30
  process, 103
  simulation algorithm, 53
  trajectory, 52
stoichiometric coefficients, 52
Stratonovich integral, 39
system, 18

tagged partition, 93
thermal equilibrium, 46
thermodynamic laws, 45
transformation, 98
transition probability, 106
transition probability rate, 107

universe, 45

variance, 99

weak-sense stationary process, 104
white noise, 107
Wiener process, 105