Linear Control Theory and Structured Markov Chains

Yoni Nazarathy

Lecture Notes for a Course in the 2016 AMSI Summer School
(Separated into chapters).
Based on a book draft co-authored with
Sophie Hautphenne, Erjen Lefeber and Peter Taylor.

Last Updated: January 21, 2016.

Chapter 3:
Linear Dynamical Systems and Markov Chains
Preface

This booklet contains lecture notes and exercises for a 2016 AMSI Summer School Course: “Linear Control Theory and Structured Markov Chains” taught at RMIT in Melbourne by Yoni Nazarathy. The notes are based on a subset of a draft book about a similar subject by Sophie Hautphenne, Erjen Lefeber, Yoni Nazarathy and Peter Taylor. The course includes 28 lecture hours spread over 3.5 weeks. The course includes assignments, short in-class quizzes and a take-home exam. These assessment items are to appear in the notes as well.

The associated book is designed to teach readers, elements of linear control theory and structured Markov chains. These two fields rarely receive a unified treatment as is given here. It is assumed that the readers have a minimal knowledge of calculus, linear algebra and probability, yet most of the needed facts are summarized in the appendix, with the exception of basic calculus. Nevertheless, the level of mathematical maturity assumed is that of a person who has covered 2-4 years of applied mathematics, computer science and/or analytic engineering courses.

Linear control theory is all about mathematical models of systems that abstract dynamic behavior governed by actuators and sensed by sensors. By designing state feedback controllers, one is often able to modify the behavior of a system which otherwise would operate in an undesirable manner. The underlying mathematical models are inherently deterministic, as is suited for many real-life systems governed by elementary physical laws. The general constructs are system models, feedback control, observers and optimal control under quadratic costs. The basic theory covered in this book has reached relative maturity nearly half a century ago: the 1960’s, following some of the contributions by Kalman and others. The working mathematics needed to master basic linear control theory is centered around linear algebra and basic integral transforms. The theory relies heavily on eigenvalues, eigenvectors and other aspects related to the spectral decomposition of matrices.

Markov chains are naturally related to linear dynamical systems and hence linear control theory, since the state transition probabilities of Markov chains evolve as a linear dynamical system. In addition the use of spectral decompositions of matrices, the matrix exponential and other related features also resembles linear dynamical systems. The field of structured Markov chains, also referred to as Matrix Analytic Methods, goes back to the mid 1970’s, yet has gained popularity in the teletraffic, operations research...
and applied probability community only in the past two decades. It is unarguably a
more esoteric branch of applied mathematics in comparison to linear control theory and
it is currently not applied as abundantly as the former field.

A few books at a similar level to this one focus on dynamical systems and show that
the probabilistic evolution of Markov chains over finite state spaces behaves as linear
dynamical systems. This appears most notably in [?]. Yet, structured Markov chains
are more specialized and possess more miracles. In certain cases, one is able to ana-
lyze the behavior of Markov chains on infinite state spaces, by using their structure.
E.g. underlying matrices may be of block diagonal form. This field of research often
focuses on finding effective algorithms for solutions of the underlying performance anal-
ysis problems. In this book we simply illustrate the basic ideas and methods of the field.
It should be noted that structured Markov chains (as Markov chains in general) often
make heavy use of non-negative matrix theory (e.g. the celebrated Perron-Frobenius
Theorem). This aspect of linear algebra does not play a role in the classic linear control
theory that we present here, yet appears in the more specialized study of control of
non-negative systems.

Besides the mathematical relation between linear control theory and structured Markov
chains, there is also a much more practical relation which we stress in this book. Both
fields, together with their underlying methods, are geared for improving the way we
understand and operate dynamical systems. Such systems may be physical, chemical,
biological, electronic or human. With its styled models, the field of linear control theory
allows us to find good ways to actually control such systems, on-line. With its ability to
capture truly random behavior, the field of structured Markov chains allows us to both
describe some significant behaviors governed by randomness, as well as to efficiently
quantify (solve) their behaviors. But control does not really play a role.

With the exception of a few places around the world (e.g. the Mechanical Engineering
Department at Eindhoven University of Technology), these two fields are rarely taught
simultaneously. Our goal is to facilitate such action through this book. Such a unified
treatment will allow applied mathematicians and systems engineers to understand the
underlying concepts of both fields in parallel, building on the connections between the
two.

Below is a detailed outline of the structure of the book. Our choice of material to cover
was such as to demonstrate most of the basic features of both linear control theory and
structured Markov chains, in a treatment that is as unified as possible.

Outline of the contents:

The notes contains a few chapters and some appendices. The chapters are best read
sequentially. Notation is introduced sequentially. The chapters contain embedded short
exercises. These are meant to help the reader as she progresses through the book, yet at
the same time may serve as mini-theorems. That is, these exercises are both deductive
and informative. They often contain statements that are useful in their own right. The end of each chapter contains a few additional exercises. Some of these exercises are often more demanding, either requiring computer computation or deeper thought. We do not refer to computer commands related to the methods and algorithms in the book explicitly. Nevertheless, in several selected places, we have illustrated example MATLAB code that can be used.

For the 2016 AMSI summer school, we have indicated besides each chapter the in-class duration that this chapter will receive in hours.

**Chapter 1 (2h)** is an elementary introduction to systems modeling and processes. In this chapter we introduce the types of mathematical objects that are analyzed, give a feel for some applications, and describe the various use-cases in which such an analysis can be carried out. By a use-case we mean an activity carried out by a person analyzing such processes. Such use cases include “performance evaluation”, “controller design”, “optimization” as well as more refined tasks such as stability analysis, pole placement or evaluation of hitting time distributions.

**Chapter 2 (7h)** deals with two elementary concepts: Linear Time Invariant (LTI) Systems and Probability Distributions. LTI systems are presented from the viewpoint of an engineering-based “signals and systems” course. A signal is essentially a time function and system is an operator on functional space. Operators that have the linearity and time-invariance property are LTI and are described neatly by either their impulse response, step response, or integral transforms of one of these (the transfer function). It is here that the convolution of two signals plays a key role. Signals can also be used to describe probability distributions. A probability distribution is essentially an integrable non-negative signal. Basic relations between signals, systems and probability distributions are introduced. In passing we also describe an input–output form of stability: BIBO stability, standing for “bounded input results in bounded output”. We also present feedback configurations of LTI systems, showing the usefulness of the frequency domain (s-plane) representation of such systems.

**Chapter 3 (13h)** moves onto dynamical models. It is here that the notion of state is introduced. The chapter begins by introducing linear (deterministic) dynamical systems. These are basically solutions to systems of linear differential equations where the free variable represents time. Solutions are characterized by matrix powers in discrete time and matrix exponentials in continuous time. Evaluation of matrix powers and matrix exponentials is a subject of its right as it has to do with the spectral properties of matrices, this is surveyed as well. The chapter then moves onto systems with discrete countable (finite or infinite) state spaces evolving stochastically: Markov chains. The basics of discrete time and continuous time Markov chains are surveyed. In doing this a
few example systems are presented. We then move onto presenting input–state–output systems, which we refer to as \((A, B, C, D)\) systems. These again are deterministic objects. This notation is often used in control theory and we adopt it throughout the book. The matrices \(A\) and \(B\) describe the effect on input on state. The matrices \(C\) and \(D\) are used to describe the effect on state and input on the output. After describing \((A, B, C, D)\) systems we move onto distributions that are commonly called Matrix Exponential distributions. These can be shown to be directly related to \((A, B, C, D)\) systems. We then move onto the special case of phase type (PH) distributions that are matrix exponential distributions that have a probabilistic interpretation related to absorbing Markov chains. In presenting PH distributions we also show parameterized special cases.

**Chapter 4 (0h) is not taught as part of the course.** This chapter dives into the heart of Matrix Analytic Modeling and analysis, describing quasi birth and deaths processes, Markovian arrival processes and Markovian Binary trees, together with the algorithms for such models. The chapter begins by describing QBDs both in discrete and continuous time. Then moves onto Matrix Geometric Solutions for the stationary distribution showing the importance of the matrices \(G\) and \(R\). The chapter then shows elementary algorithms to solve for \(G\) and \(R\) focusing on the probabilistic interpretation of iterations of the algorithms. State of the art methods are summarized but are not described in detail. Markovian Arrival Point Processes and their various sub-classes are also survyed. As examples, the chapter considers the M/PH/1 queue, PH/M/1 queue as well as the PH/PH/1 generalization. The idea is to illustrate the power of algorithmic analysis of stochastic systems.

**Chapter 5 (4h)** focuses on \((A, B, C, D)\) systems as used in control theory. Two main concepts are introduced and analyzed: state feedback control and observers. These are cast in the theoretical framework of basic linear control theory, showing the notions of controllability and observability. The chapter begins by introducing two physical examples of \((A, B, C, D)\) systems. The chapter also introduces canonical forms of \((A, B, C, D)\) systems.

**Chapter 6 (2h)** deals with stability of both deterministic and stochastic systems. Notions and conditions for stability were alluded to in previous chapters, yet this chapter gives a comprehensive treatment. At first stability conditions for general deterministic dynamical systems are presented. The concept of a Lyapounov function is introduced. This is the applied to linear systems and after that stability of arbitrary systems by means of linearization is introduced. Following this, examples of setting stabilizing feedback control rules are given. We then move onto stability of stochastic systems (essentially positive recurrence). The concept of a Foster-Lyapounov function is given for showing positive recurrence of Markov chains. We then apply it to quasi-birth-death processes.
proving some of the stability conditions given in Chapter 4 hold. Further stability conditions of QBD’s are also given. The chapter also contains the Routh-Hourwitz and Jury criterions.

Chapter 7 (0h) is not taught as part of the course. is about optimal linear quadratic control. At first Bellman’s dynamic programming principle is introduced in generality, and then it is formulated for systems with linear dynamics and quadratic costs of state and control efforts. The linear quadratic regulator (LQR) is introduced together with its state feedback control mechanism, obtained by solving Ricatti equations. Relations to stability are overviewed. The chapter then moves onto Model-predictive control and constrained LQR.

Chapter 8 (0h) is not taught as part of the course. This chapter deals with fluid buffers. The chapter involves both results from applied probability (and MAM), as well as a few optimal control examples for deterministic fluid systems controlled by a switching server. The chapter begins with an account of the classic fluid model of Anick, Mitra and Sondhi. It then moves onto additional models including deterministic switching models.

Chapter 9 (0h) is not taught as part of the course. This chapter introduces methods for dealing with deterministic models with additive noise. As opposed to Markov chain models, such models behave according to deterministic laws, e.g. \((A, B, C, D)\) systems, but are subject to (relatively small) stochastic disturbances as well as to measurement errors that are stochastic. After introducing basic concepts of estimation, the chapter introduces the celebrated Kalman filter. There is also brief mention of linear quadratic Gaussian control (LQG).

The notes also contains an extensive appendix which the students are required to cover by themselves as demand arises. The appendix contains proofs of results in cases where we believe that understanding the proof is instructive to understanding the general development in the text. In other cases, proofs are omitted.

Appendix A touches on a variety of basics: Sets, Counting, Number Systems (including complex numbers), Polynomials and basic operations on vectors and matrices.

Appendix B covers the basic results of linear algebra, dealing with vector spaces, linear transformations and their associated spaces, linear independence, bases, determinants and basics of characteristic polynomials, eigenvalues and eigenvectors including the Jordan Canonical Form.
Appendix C covers additional needed results of linear algebra.

Appendix D contains probabilistic background.

Appendix E contains further Markov chain results, complementing the results presented in the book.

Appendix F deals with integral transforms, convolutions and generalized functions. At first convolutions are presented, motivated by the need to know the distribution of the sum of two independent random variables. Then generalized functions (e.g. the delta function) are introduced in an informal manner, related to convolutions. We then present the Laplace transform (one sided) and the Laplace-Stiltijes Transform. Also dealing with the region of convergence (ROC). In here we also present an elementary treatment of partial fraction expansions, a method often used for inverting rational Laplace transforms. The special case of the Fourier transform is briefly surveyed, together with a discussion of the characteristic function of a probability distribution and the moment generating function. We then briefly outline results of the z-transform and of probability generating functions.

Besides thanking Sophie, Erjen and Peter, my co-authors for the book on which these notes are based, I would also like to thank (on their behalf) to several colleagues and students for valuable input that helped improve the book. Mark Fackrell and Nigel Bean’s analysis of Matrix Exponential Distributions has motivated us to treat the subjects of this book in a unified treatment. Guy Latouche was helpful with comments dealing with MAM. Giang Nugyen taught jointly with Sophie Hautphenene a course in Vietnam covering some of the subjects. A Master’s student from Eindhoven, Kay Peeters, visiting Brisbane and Melbourne for 3 months and prepared a variety of numerical examples and illustrations, on which some of the current illustrations are based. Also thanks to Azam Asanjarani and to Darcy Bermingham. The backbone of the book originated while the authors were teaching an AMSI summer school course, in Melbourne during January 2013. Comments from a few students such as Jessica Yue Ze Chan were helpful.

I hope you find these notes useful,
Yoni.
# Contents

Preface ......................................................... 3

3 Linear Dynamical Systems and Markov Chains (13h) ............ 11
  3.1 Linear Dynamical Systems ........................................ 12
    3.1.1 Example Models ............................................ 14
    3.1.2 Finding the trajectory ...................................... 14
  3.2 Evaluation of the Matrix Exponential .......................... 17
    3.2.1 The Similarity Transformation ............................. 18
    3.2.2 Diagonalizable Matrices .................................... 19
    3.2.3 Jordan’s Canonical Form ................................... 20
  3.3 Markov Chains in Discrete Time .................................. 22
    3.3.1 Markov Chain Basics ....................................... 22
    3.3.2 First-Step Analysis ........................................ 25
    3.3.3 Class Structure, Periodicity, Transience and Recurrence .... 27
    3.3.4 Limiting Probabilities ..................................... 32
  3.4 Markov Chains in Continuous Time .............................. 34
    3.4.1 Continuous Time Basics .................................... 34
    3.4.2 Further Continuous Time Properties ........................ 38
  3.5 Elementary Structured Markov Models ......................... 38
    3.5.1 Birth-and-Death Processes .................................. 39
    3.5.2 The Poisson Process ....................................... 39
    3.5.3 The Birth-Death Stationary Distribution .................... 41
    3.5.4 Simple Queueing Models .................................... 42
  3.6 \((A, B, C, D)\) Linear Input-Output Systems ..................... 46
    3.6.1 Time-Domain Representation of the Output ................ 47
    3.6.2 The Resolvent .............................................. 49
    3.6.3 The Transfer Function Matrix ............................... 50
    3.6.4 Equivalent Representations of Systems ..................... 51
3.6.5 Rational Laplace-Stieltjes Transforms Revisited ............... 51
3.7 Phase-Type (PH) Distributions ........................................ 52
  3.7.1 The Absorption Time in an Absorbing CTMC ................... 52
  3.7.2 Examples .............................................. 53
  3.7.3 A Dense Family of Distributions ............................... 55
  3.7.4 Relationship to ME Distributions ............................... 55
  3.7.5 Operations on PH Random Variables ......................... 56
  3.7.6 Moment Matching ....................................... 60
3.8 Relations Between Discrete and Continuous Time .................. 62
  3.8.1 Different Discretizations of a CTMC .......................... 62
  3.8.2 Sampling a Continuous Time (A, B, C, D) System ............. 63
  3.8.3 Discrete/Continuous, PH/ME Distributions Relationships ..... 64
Bibliographic Remarks .................................................. 64
Exercises ............................................................... 64
Chapter 3

Linear Dynamical Systems and Markov Chains (13h)

In Chapter 1 we introduced four basic types of processes. These included the continuous time processes:

1. \( \{x(t)\} \) with \( t \in \mathbb{R} \) and \( x(t) \in \mathbb{R}^n \).

2. \( \{X(t)\} \) with \( t \in \mathbb{R} \) and \( X(t) \in \mathcal{S} \), where \( \mathcal{S} \) is some countable (finite or infinite set).

As well as their discrete time counter-parts:

3. \( \{x(\ell)\} \) with \( \ell \in \mathbb{Z} \) and \( x(\ell) \in \mathbb{R}^n \).

4. \( \{X(\ell)\} \) with \( \ell \in \mathbb{Z} \) and \( X(\ell) \in \mathcal{S} \), where \( \mathcal{S} \) is some countable (finite or infinite set).

We typically take the continuous valued processes (1) and (3) to be deterministic and the discrete valued processes (2) and (4) to be stochastic.

In this chapter we introduce the dynamics of the most fundamental classes of such processes. In the deterministic case we introduce the behaviors associated with linear dynamics. These types of processes are defined by means of linear differential or difference equations. In the stochastic case we introduce the behaviors associated with the Markovian property. In such cases we introduce the processes as Markov chains in continuous or discrete time.

In the previous chapter we treated systems as input-output relationships, generally ignoring the notion of their state. This chapter differs in the sense that it is almost all about state. We now treat the values of the processes as a state of a system.

After introducing the behaviors associated with linear dynamical systems and Markov chains, we move on to introduce some of the basic objects that will appear in the
continuation of the book. These include \((A, B, C, D)\) linear input-output systems. Such objects combine the notion of state with input and output. Further we show matrix exponential probability distributions which are closely related to \((A, B, C, D)\) systems. We close with phase type distributions which are a special case of matrix exponential probability distributions that is defined by means of a Markov chain.

As a note, the reader should observe that most of the processes introduced in this chapter (and in the remainder of the book) are time invariant. This concept was defined in terms of SISO LTI systems in Chapter 2. In the more general setting it implies that the behavior of the process is not influenced by the current time. Such processes are some times called time homogenous or stationary. Yet we caution the reader about the use of the term “stationary” since it has a different meaning in different contexts.

### 3.1 Linear Dynamical Systems

We now consider deterministic processes of the form:

1. \(\{x(t)\}\) with \(t \in \mathbb{R}_+\) and \(x(t) \in \mathbb{R}^n\).

3. \(\{x(\ell)\}\) with \(\ell \in \mathbb{Z}_+\) and \(x(\ell) \in \mathbb{R}^n\).

Observe the slight difference from our previous definition: We consider the time index as starting at 0, i.e. \(\mathbb{R}_+\) is the set of nonnegative reals and similarly for \(\mathbb{Z}_+\). This is useful since we will describe some initial value.

The standard way to describe the behavior of such processes is to suggest some Lipschitz continuous \(f : \mathbb{R}^n \to \mathbb{R}^n\) and set:

\[
\dot{x}(t) = f(x(t)) \quad \text{or} \quad x(\ell + 1) = f(x(\ell)),
\]  

(3.1)

together with a specified initial value \(x(0) = x_0\). The continuous time or discrete time equation (3.1) together with an initial value is sometimes referred to as an initial value problem.

Such processes are generally referred to as autonomous dynamical systems. In the dynamical system context, the use of the phrase “autonomous” is due to the fact that the evolution does not depend on time (as opposed to \(\dot{x}(t) = f(t, x(t))\) for example). Observe also that the use of the phrase “system” here is not in the input-output context used in Chapter 2. Rather the system is essentially the process \(x(\cdot)\) and its behaviors.

An alternative to looking at the differential/difference equation occurring in (3.1) is to look at the integral/summation version:

\[
x(t) = x_0 + \int_0^t f(x(s))ds \quad \text{or} \quad x(\ell) = x_0 + \sum_{k=0}^{\ell-1} \left( f(x(k)) - x(k) \right).
\]
Some of the theory of dynamical systems (and differential equations) deals with the existence and uniqueness of the continuous time system appearing in (3.1) (in the discrete time setting there are no such issues). To illustrate possible uniqueness problems, consider the following:

**Example 3.1.1.** Take \( \dot{x}(t) = x(t)^{1/3}, x(0) = 0 \), then there are at least two solutions:

\[
x(t) = 0 \quad \text{and} \quad x(t) = \left(\frac{4}{9} t\right)^{3/2}.
\]

We do not consider uniqueness and existence issues any further since our interest is in the special case:

\[
f(x) = Ax,
\]

for \( A \in \mathbb{R}^{n \times n} \). That is, we consider **linear dynamical systems** of the form:

\[
\dot{x}(t) = Ax(t) \quad \text{or} \quad x(\ell + 1) = Ax(\ell),
\]

(3.2)

together with,

\[
x(0) = x_0.
\]

(3.3)

For these systems, uniqueness and existence is not an issue:

**Theorem 3.1.2.** For any \( x_0 \in \mathbb{R}^n \) and \( A \in \mathbb{R}^{n \times n} \) there exists a unique \( x(\cdot) \) satisfying (3.2) and (3.3).

The proof for the discrete time case is immediate. We do not prove this result from first principles for the continuous time case, yet rather construct the unique solution in the sequel.

We now show two generic examples that bear significant importance in their own right. Further application examples appear in the exercises of this chapter.

**Example 3.1.3. Linearization around an equilibrium point:** Consider a general (non-linear) dynamical system,

\[
\dot{x}(t) = f(x(t)),
\]

where \( f(\cdot) \) is Lipschitz continuous. An equilibrium point of the system is a point \( \bar{x} \in \mathbb{R}^n \) such that,

\[
f(\bar{x}) = 0.
\]

Taking the Taylor series expansion of \( f(\cdot) \) at the point \( \bar{x} \), we have,

\[
f(x) = f(\bar{x}) + J(\bar{x})(x - \bar{x}) + o(||x - \bar{x}||).
\]

where \( J(\cdot) \) is the Jacobian matrix of \( f(\cdot) \). We can then analyze the linear dynamical system,

\[
\dot{y}(t) = Ay(t), \quad \text{with} \quad A = J(\bar{x}),
\]

with initial value, \( y(0) = x_0 - \bar{x} \). Then \( \{y(t)\} \) approximates \( \{x(t) - \bar{x}\} \) at the vicinity of \( 0 \).
Example 3.1.4. Higher order derivatives: Consider the linear, autonomous, homogeneous ordinary scalar differential equation of order $n$:

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \ldots + a_1 y^{(1)}(t) + a_0 y(t) = 0.$$  \hfill (3.4)

Denote now,

$$x_1(\cdot) := y(\cdot), \quad x_2(\cdot) := y^{(1)}(\cdot), \ldots \quad x_{n-1}(\cdot) := y^{(n-2)}(\cdot), \quad x_n(\cdot) := y^{(n-1)}(\cdot),$$

and consider the autonomous system,

$$\begin{bmatrix}
  \dot{x}_1(t) \\
  \dot{x}_2(t) \\
  \vdots \\
  \dot{x}_{n-1}(t) \\
  \dot{x}_n(t)
\end{bmatrix} =
\begin{bmatrix}
  0 & 1 & 0 & \ldots & 0 \\
  0 & 0 & 1 & 0 & \vdots \\
  \vdots \\
  0 & \ldots & \ldots & \ldots & 0 \\
  -a_0 & -a_1 & \ldots & -a_{n-1} & -1
\end{bmatrix}
\begin{bmatrix}
  x_1(t) \\
  x_2(t) \\
  \vdots \\
  x_{n-1}(t) \\
  x_n(t)
\end{bmatrix}.$$

Then it is clear that solutions of the $n$th dimensional system also satisfy (3.4). Note that the above matrix is called a companion matrix associated with $(a_0, \ldots, a_{n-1})$.

The idea of maintaining higher order derivatives as part of the state also comes up naturally in the modeling use case. For example, it may be very natural in certain cases to have the state record the location, speed and acceleration of a physical object.

### 3.1.1 Example Models

**Example 3.1.5.** Consider the following publication scenario for an academic researcher: Each year, each published paper of the researcher yields one new research direction which results in a new submission. Further each submitted paper becomes published. Let $x_1(\ell)$ and $x_2(\ell)$ denote the number of submitted and published papers of the academic in her $\ell$th year of research respectively. Then,

$$x_1(\ell) = x_2(\ell - 1), \quad \text{and} \quad x_2(\ell) = x_1(\ell - 1) + x_2(\ell - 1).$$

or,

$$\begin{bmatrix}
  x_1(\ell) \\
  x_2(\ell)
\end{bmatrix} =
\begin{bmatrix}
  0 & 1 \\
  1 & 1
\end{bmatrix}
\begin{bmatrix}
  x_1(\ell - 1) \\
  x_2(\ell - 1)
\end{bmatrix}.$$

### 3.1.2 Finding the trajectory

Given a system (3.2)-(3.3), one way of finding the behavior of $x(\cdot)$ is by successive iterations in the discrete time case or by approximation methods in the continuous time
case (we do not cover such methods here, see the bibliographic remarks at the end of the chapter). Alternatively (and often preferably) we would like to find a more insightful mathematical description of the solution. In the discrete time case this is elementary:

\[ x(\ell) = A x(\ell - 1) = A A x(\ell - 2) = A^3 x(\ell - 3) = \ldots = A^\ell x(\ell - \ell) = A^\ell x_0. \]

Hence,

\[ x(\ell) = A^\ell x_0 \quad (3.5) \]

In the continuous time case we need to introduce the continuous analog of the matrix power \( A^\ell \). We call this object the **matrix exponential** and denote it by \( e^{At} \). It is formally constructed below.

## Picard Iterations and Matrix Exponential

Given a continuous initial value problem (generally time varying):

\[ \dot{x}(t) = f(t, x(t)), \quad x(0) = x_0, \]

a **Picard iteration sequence** is a sequence of functions constructed as follows:

\[
\begin{align*}
\phi_0(t) &= x_0 \\
\phi_{m+1}(t) &= x_0 + \int_0^t f(s, \phi_m(s)) \, ds, \quad m = 0, 1, 2, 3, \ldots
\end{align*}
\]

It can be shown that if \( f(\cdot) \) satisfies some Lipschitz conditions then the successive approximations \( \phi_m(\cdot), m = 0, 1, 2, \ldots \) exist, are continuous and converge uniformly as \( m \to \infty \) to the unique solution which we denote \( \phi(\cdot) \). That is, for every \( \epsilon > 0 \) there exists an \( N \) such that for all \( t \) in the specified domain,

\[ ||\phi(t) - \phi_m(t)|| < \epsilon, \]

whenever \( m > N \).

To illustrate this, it is useful to briefly consider the time-dependent (non-autonomous) system (specified at initial time \( t_0 \), not necessarily 0):

\[ \dot{x}(t) = A(t)x(t), \quad x(t_0) = x_0. \quad (3.6) \]

By taking successive Picard iterations, it is observed that the solution of the above system is:

\[ \phi(t) = \Phi(t, t_0)x_0. \quad (3.7) \]
where the state transition matrix $\Phi(t, t_0)$ is defined as follows:

$$
\Phi(t, t_0) := I + \int_{t_0}^{t} A(s_1) ds_1 + \int_{t_0}^{t} A(s_1) \int_{t_0}^{s_1} A(s_2) ds_2 ds_1 + \int_{t_0}^{t} A(s_1) \int_{t_0}^{s_1} A(s_2) \int_{t_0}^{s_2} A(s_3) ds_3 ds_2 ds_1 + \ldots \\
\ldots + \int_{t_0}^{t} A(s_1) \int_{t_0}^{s_1} A(s_2) \ldots \int_{t_0}^{s_{m-1}} A(s_m) ds_m ds_{m-1} \ldots ds_1 + \ldots
$$

The above expression is sometimes called the Peano-Baker series. Note that, $\Phi(t, t) = I$.

Note that we can differentiate the Peano-Baker series with respect to $t$ to get:

$$
\dot{\Phi}(t, t_0) = A(t) \Phi(t, t_0).
$$

In the time-independent case of $A(t) = A$, the $m$’th term in the Peano-Baker series reduces to:

$$
A^m \int_{t_0}^{t} \int_{t_0}^{s_1} \int_{t_0}^{s_2} \ldots \int_{t_0}^{s_{m-1}} 1 ds_m \ldots ds_1 = \frac{(t - t_0)^m}{m!} A^m.
$$

Hence in this case, the state transition matrix reduces to the form,

$$
\Phi(t, t_0) = \sum_{k=0}^{\infty} \frac{(t - t_0)^k}{k!} A^k
$$

From the theory of differential equations and the result about the convergence of the Picard iteration sequence, the following can be deduced:

**Theorem 3.1.6.** Let $A \in \mathbb{R}^{n \times n}$. Denote,

$$
S_m(t) = \sum_{k=0}^{m} \frac{t^k}{k!} A^k.
$$

Then each element of the matrix $S_m(t)$ converges absolutely and uniformly on finite interval containing 0, as $m \to \infty$.

We can thus define the matrix exponential for any $t \in \mathbb{R}$ as,

$$
e^{At} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k.
$$

Thus for the linear autonomous system, from (3.8) we have,

$$
\Phi(t, t_0) = e^{A(t-t_0)}.
$$

Hence the behavior of $x(t) = A x$, with initial value $x(0) = x_0$ is,

$$
x(t) = e^{At} x_0.
$$
Exercise 3.1.7. Show the following elementary properties of the matrix exponential:

1. \( e^0 = I \).
2. For scalar \( t_1, t_2 \), \( e^{At_1}e^{At_2} = e^{A(t_1+t_2)} \).
3. \( e^{A'} = (e^A)' \).

Here are some further properties of the matrix-exponential

**Theorem 3.1.8.** The following holds:

1. For \( \lambda \) and eigenvalue of \( A \) it holds that \( e^\lambda \) is an eigenvalue of \( e^A \).
2. \( \det (e^A) = e^{tr(A)} \)
3. \( A^pe^{At} = e^{At}A^p \) for integer \( p \).
4. If \( AB = BA \) then, \( e^{A+B} = e^Ae^B \).
5. \( \frac{d}{dt}e^{At} = Ae^{At} \).
6. For non-singular \( A \), \( \int_0^t e^{A^\tau} d\tau = (e^{At} - I)A^{-1} \).

Exercise 3.1.9. Show by example that \( e^{A+B} = e^Ae^B \) does not necessarily imply that \( AB = BA \).

### 3.2 Evaluation of the Matrix Exponential

The straightforward method to compute \( A^\ell \) is by carrying out \( \ell - 1 \) successive multiplications of the matrix \( A \). This can be reduced to \( O(\log(\ell)) \) matrix multiplications by carrying out successive evaluations of,

\[
A_2 := AA, \quad A_4 := A_2A_2, \quad A_8 := A_4A_4, \quad \ldots \quad A_{2^k} := A_{2^{k-1}}A_{2^{k-1}},
\]

up to \( k = \lfloor \log_2 \ell \rfloor \), and then multiplying \( A_{2^k} \) by \( A \) (or other \( A_{2^j} \)) a few more times to “complete” the product from \( A_{2^k} \) to \( A^\ell \).

While this sort of algorithmic “divide and concur” approach yields a significant computation improvement, it still does not yield any insight about the structure of the sequence of matrices \( \{ A^\ell, \ \ell = 0, 1, 2, \ldots \} \).
The straightforward method to approximately compute $e^{At}$ is to choose a large $K$ and evaluate the finite sum,

$$
\sum_{k=0}^{K} \frac{(At)^k}{k!}.
$$

Since $e^{At}$ always exists, the finite sum converges to the correct value as $K \to \infty$. Here one can look for basic computational improvements. For example by using the relation,

$$
\frac{(At)^{k+1}}{(k+1)!} = \left( \frac{At}{k+1} \right) \frac{(At)^k}{k!},
$$

to compute the $k+1$’st term in the summation based on the $k$’th term. But here again, such computational improvements do not yield insight on the structure of $\{e^{At}, t \geq 0\}$.

We now consider more powerful and insightful linear-algebraic methods for effectively evaluating $A^\ell$ and $e^{At}$ as well as for gaining insight on the behavior of these matrix sequences.

The following exercise shows that in some cases, evaluation is straightforward and explicit.

**Exercise 3.2.1.** Take,

$$
A = \begin{bmatrix} 0 & 0 \\ \gamma & 0 \end{bmatrix},
$$

and find $\{A^\ell, \ell = 0, 1, 2, \ldots\}$ and $\{e^{At}, t \geq 0\}$ explicitly.

### 3.2.1 The Similarity Transformation

Given $P \in \mathbb{R}^{n \times n}$, with $\det(P) \neq 0$, the matrices $A$ and $\tilde{A}$ are said to be *similar* if,

$$
\tilde{A} = P^{-1}AP \quad \text{or alternatively} \quad A = P\tilde{A}P^{-1}. \tag{3.11}
$$

Here the action of replacing $A$ by $P\tilde{A}P^{-1}$ is called the *similarity transformation*.

Assume now that we can find $P$ such that evaluation of $\tilde{A}^\ell$ is in some way easier than $A^\ell$. In this case, carrying out a similarity transformation is beneficial since,

$$
A^\ell = (P\tilde{A}P^{-1})^\ell = (P\tilde{A}P^{-1}) \cdot (P\tilde{A}P^{-1}) \cdot \ldots \cdot (P\tilde{A}P^{-1}) = P\tilde{A}(P^{-1}P)\tilde{A}(P^{-1}P) \cdot \ldots \cdot (P^{-1}P)\tilde{A}P^{-1} = P\tilde{A}^\ell P^{-1}.
$$

Similarly, as a direct consequence, for the matrix exponential we have,

$$
e^{At} = Pe^{\tilde{A}t}P^{-1}.
$$
We now arrive at the question of what is a simple $\tilde{A}$? Well the simplest is a diagonal matrix.

Observe that the eigenvalues of $\tilde{A}$ and $A$ are the same because the characteristic polynomial is the same:

$$\det(P^{-1}AP - \lambda I) = \det(P^{-1}(A - \lambda I)P) = \det(P^{-1}) \det(A - \lambda I) \det(P) = \det(A - \lambda I).$$

### 3.2.2 Diagonalizable Matrices

If $\tilde{A}$ is a diagonal matrix,

$$\tilde{A} = \begin{bmatrix} \tilde{a}_1 & 0 & \cdots & 0 \\ 0 & \tilde{a}_2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \tilde{a}_n \end{bmatrix} := \text{diag}(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n),$$

then it is easy to see that, $\tilde{A}^\ell = \text{diag}(\tilde{a}_1^\ell, \tilde{a}_2^\ell, \ldots, \tilde{a}_n^\ell)$. Further finding the matrix exponential of a diagonal matrix is also simple:

**Exercise 3.2.2.** Show that if as above, $\tilde{A} = \text{diag}(\tilde{a}_1, \ldots, \tilde{a}_n)$, then,

$$e^{\tilde{A}t} = \text{diag}(e^{\tilde{a}_1t}, \ldots, e^{\tilde{a}_nt}).$$

We are thus motivated to find a similarity transformation matrix $P$ that will yield a diagonal $\tilde{A}$. When this can be done, the matrix $P$ that diagonalizes $A$, can be constructed by taking columns to be eigenvectors of $A$, each corresponding to a different eigenvalue. We illustrate the basic idea now.

The similarity transformation (3.11) may be read as,

$$AP = P\tilde{A}. \quad (3.12)$$

Now if the columns of $P$ are eigenvectors of $A$ and we impose on $\tilde{A}$ to be diagonal, then (3.12) is read as,

$$Ap_{i,i} = \tilde{a}_i p_{i,i}, \quad i = 1, \ldots, n,$$

where $p_{i,i}$ denotes the $i$'th column of $P$. In this case the diagonal elements of $\tilde{A}$ are nothing but the eigenvalues of $A$.

When is this possible? For start we have the following:

**Theorem 3.2.3.** If for $A \in \mathbb{R}^{n \times n}$ there are $n$ distinct eigenvalues then $A$ is diagonalizable.

Obviously having distinct eigenvalues is not a necessary condition for $A$ to be diagonalizable (consider for example certain diagonal matrices):
Exercise 3.2.4. Give an example of a matrix $A$ with non-distinct eigenvalues that is still diagonalizable.

The algebraic multiplicity of an eigenvalue, $\lambda_0$, denoted $m_a(\lambda)$ is the multiplicity of the root $\lambda_0$ in the characteristic equation $p_A(\cdot) = 0$. Namely, the polynomial $p_A(\lambda)$ is divisible by exactly $m_a(\lambda_0)$ powers of $(\lambda - \lambda_0)$.

Exercise 3.2.5. Argue why $n = \sum_i m_a(\lambda_i)$.

The geometric multiplicity, denoted $m_g(\lambda_0)$ is the dimension of the subspace $E_{\lambda_0}$ (spanning eigenvectors associated with $\lambda_0$).

Theorem 3.2.6. For a matrix $A$ with eigenvalue $\lambda$:

\[ 1 \leq m_g(\lambda) \leq m_a(\lambda). \]

Theorem 3.2.7. A matrix is diagonalizable if and only if for all its eigenvalues $m_a = m_g$.

3.2.3 Jordan’s Canonical Form

If $A$ is not diagonalizable, what can be done? We first define the power vector, $w$, (sometimes called generalized eigenvector) of a matrix $A$, if for some scalar $\lambda$ and positive integer $p$:

\[(A - \lambda I)^p w = 0.\]

The subspace spanned by power vectors corresponding to $\lambda$ is called the power space of $\lambda$. The order of a power vector is said to be $p$ if,

\[(A - \lambda I)^p w = 0, \quad \text{but} \quad (A - \lambda I)^{p-1} w \neq 0.\]

Exercise 3.2.8. Show that if $w$ is a power vector of order $p$ with $\lambda$, then $(L - \lambda I)w$ is a power vector of order $p - 1$, $(L - \lambda I)^2 w$ is a power vector of order $p - 2$ and so on through to $(L - \lambda I)^{p-1} w$ which is an eigenvector.

For eigenvalues $\lambda$ with eigenvector $v$ we have that $e^Av = e^{\lambda t}v$. For in the case of a power vector $w$ we have:

\[e^{\lambda\tau}w = e^{(A-\lambda I)\tau}w = e^\lambda \sum_{k=0}^{\infty} \frac{1}{k!} (A-\lambda I)^k w = e^\lambda \sum_{k=0}^{p-1} \frac{1}{k!} (A-\lambda I)^k w.\]

Example 3.2.9. Consider,

\[B = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}.\]
It holds that $p_A(\lambda) = (\lambda - 1)^2$. We have that $(1, -1)'$ is an eigenvector with eigenvalue 1 and $(1, 1)'$ is a power vector of order 2 corresponding to the eigenvalue 1. In this case,

$$e^{Bt} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = e^t \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$e^{Bt} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = e^t \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2te^t \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

**Theorem 3.2.10.** Let $A$ be a square matrix:

1. A collection of power vectors, each corresponding to distinct eigenvalues is linearly independent.
2. Every power vector has order less than or equal to $m_a(\lambda)$.
3. The power space corresponding to $\lambda$ is the same as the kernel of $(A - \lambda I)^{m_a(\lambda)}$.
4. The dimension of the power space corresponding to $\lambda$ is exactly $m_a(\lambda)$.
5. There exists a basis consisting of power vectors of $A$.

Now using power vectors, we can construct the Jordan Canonical Form of an arbitrary matrix $A$ (diagonalizable or not). Given an eignevalue $\lambda$, a Jordan Block is a block of size $m$ is the $m \times m$ matrix.

$$I_m(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & \cdots & \cdots & \lambda \end{bmatrix}$$

Note that a Jordan block of size 1 is the scalar $\lambda$.

Now the Jordan canonical form of a matrix $A$ with $r$ distinct eigenvalues, $\lambda_1, \ldots, \lambda_r$ each with geometric multiplicity denoted by $m_i$ is the matrix,

$$\tilde{A} = \begin{bmatrix} I_{m_1}(\lambda_1) & & & & \\ & I_{m_2}(\lambda_1) & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & I_{m_r}(\lambda_r) \end{bmatrix}$$

**Exercise 3.2.11.** Explain why the Jordan canonical form of a matrix reduces to a diagonal matrix if a matrix is diagonalizable.
Theorem 3.2.12. There always exists a non-singular \( P \) such that,
\[
\tilde{A} = P^{-1}AP,
\]
where \( \tilde{A} \) is the Jordan canonical form of \( A \).

Now in general we have \( e^A = P^{-1}e^{\tilde{A}}P \) and since \( \tilde{A} \) is either diagonal or block-diagonal (in the non-diagonalizable) case, it is “easier” to evaluate its matrix exponential (similarly for matrix powers).

3.3 Markov Chains in Discrete Time

Moving from the deterministic objects \( x(\cdot) \) to the stochastic ones \( X(\cdot) \), we now consider Markov chains. Specifically, *discrete time Markov chains* (DTMC) and *continuous time Markov chains* (CTMC).

3.3.1 Markov Chain Basics

A *stochastic process* is a random function \( X(\ell, \omega) \) where say \( t \in \mathbb{R} \) (or \( \ell \in \mathbb{Z} \)) represents time and \( \omega \in \Omega \) is a point in the probability sample space. An alternative view, is to think of a stochastic process as a family (sequence) of random variables \( \{X(t, \omega), \ t \in \mathbb{R}\} \) (or \( \ell \in \mathbb{Z} \)). Stochastic processes get interesting when the random variables are not independent. I.e. when there is some dependence structure between them. In the sequel we omit the fact that \( X(\cdot, \omega) \) depends on \( \omega \) from the notation, but keep in mind it is always there.

When analysing a stochastic process, we sometimes use the term *sample path* or alternatively *realisation* to refer to one instance of the time function \( X(\cdot, \omega) \) associated with a single \( \omega \).

An elementary, but highly useful stochastic process is the *time homogenous finite state space discrete time Markov chain* (finite DTMC for short). This is a sequence of random variables indexed by \( \ell \in \mathbb{Z}_+ \) with the following three properties:

1. Lack of memory (Markovian property):
   \[
P(X(\ell + 1) = j \mid X(\ell) = i_t, \ldots, X(0) = i_0) = P(X(\ell + 1) = j \mid X(\ell) = i_t).
   \]

2. Time Homogeneity (this makes the probability law of the the process time-homogenous):
   \[
P(X(\ell + 1) = j \mid X(\ell) = i) = P(X(1) = j \mid X(0) = i) := p_{i,j}.
   \]

3. Finite state space: There is some finite set (state space), \( S \), such that,
   \[
P(X(\ell) \notin S) = 0, \quad \forall \ell.
   \]
Since we are considering finite state space Markov chains, we may think of $S = \{1, \ldots, N\}$ for some fixed integer $N \geq 2$. At the end of section we briefly also discuss infinite (but still countable) state-spaces. As you read this, it may be a good idea that you occasionally ask yourself, where (and how) the finite state space assumption is used.

Based on properties (1) and (2) above, it can be seen that in order to specify the probability law of the evolution of $\{X(\ell)\}$ we need to specify, $p_{i,j}$ for $i,j \in S$ as well as the distribution of $X(0)$ (the initial distribution). The convenient way to specify the transition probabilities is by an $N \times N$ matrix $P = [p_{i,j}]$ with non-negative elements and with row sums $= 1$. I.e. each row $i$ can be treated as a PMF indicating the distribution of $X(\ell + 1)$ given that $X(\ell) = i$. A convenient way to is specify the initial distribution is by a row vector, $\mathbf{r}(0)$ of length $N$ having non-negative elements and summing to 1 with $i$'th entry, $\mathbf{r}_i(0)$ meaning: $\mathbb{P}(X(0) = i) = \mathbf{r}_i(0)$. This is can again be viewed as a PMF.

Note that a non-negative matrix with row sums equal to 1 is called a stochastic matrix. Don’t let the name confuse you; it isn’t a random variable or a random matrix, it is a deterministic object.

Exercise 3.3.1. Check that $\mathbb{P}(C | AB) = \mathbb{P}(C | A) \iff \mathbb{P}(CB | A) = \mathbb{P}(C | A)\mathbb{P}(B | A)$.

Now using basic conditional probability and the law of total probability we can get some very nice properties. First for $\ell = 0, 1, 2, \ldots$, denote,

$$p_{i,j}^{(\ell)} = \mathbb{P}(X(\ell) = j \mid X(0) = i),$$

and the matrix of these probabilities by $P^{(\ell)} = [p^{(\ell)}_{i,j}]$. Also denote,

$$r_{i}(\ell) = \mathbb{P}(X(\ell) = i),$$

with $\mathbf{r}(\ell)$ being the row vector of these probabilities.

Exercise 3.3.2. The basic dynamics of DTMCs is given by the following:

1. Show that $P^{(0)}$ is the identity matrix.
2. Show (arguing probabilistically) that $P^{(\ell)}$ is a stochastic matrix for any $\ell \in \mathbb{Z}_+.$
3. Show the Chapman-Kolmogorov equations hold:

$$p^{(m+n)}_{i,j} = \sum_{k=1}^{N} p^{(m)}_{i,k} p^{(n)}_{k,j}.$$ 
4. Show that $P^{(\ell)} = P^\ell$. I.e. $P^\ell = P \cdot P \cdot \ldots \cdot P$, where the product is of $\ell$ matrices.
5. Show that $\mathbf{r}(\ell) = \mathbf{r}(0)P^\ell$ (the right hand side here is a row vector multiplied by a matrix).
The next exercise, will ensure you got the point. I hope you are in the mood.

**Exercise 3.3.3.** Make a model of your feelings. Say $1 \equiv \text{“happy”}$, $2 \equiv \text{“indifferent”}$, $3 \equiv \text{“sad”}$. Assume that you are Markovian (i.e. the way you feel at day $\ell + 1$ is not affected by days prior to day $\ell$, if the feelings at day $\ell$ are known)

1. Specify the transition probabilities matrix $P$ which you think matches you best.

2. Assume that at day 0 you are sad with probability 1. What is the probability of being happy in day 3.

3. Assume that at day 0 you have a (discrete) uniform distribution of feelings, what is the probability of being happy in day 3.

4. Assuming again, that the initial distribution is uniform, what is the probability of “happy, happy, sad, sad, happy” (a sequence of 5 values on times $\ell = 0, 1, \ldots, 4$).

Markov chains generalised i.i.d. sequences:

**Exercise 3.3.4.** Assume you are given a PMF $p_X(\cdot)$ with support $\{1, \ldots, N\}$. How can you make a Markov chain such that $\{X(\ell)\}$ is an i.i.d. sequence of that PMF? I.e. what matrix $P$ will you use? Explain.

The fact that $r(\ell) = r(0)P^\ell$ is remarkable and beautiful. But in general it is quite hard to have an explicit analytic expression for $P^\ell$. With some effort, you can do this for a two-state Markov chain:

**Exercise 3.3.5.** Consider the Markov chain over $S = \{1, 2\}$.

1. How many free parameters are in this model (i.e. how many numbers specify $r(0)$ and $P$)?

2. Write an expression for $P^\ell$ in terms of the parameters (e.g. do this by diagonalising the matrix $P$ so that you can evaluate matrix powers easily).

3. Write an expression for $r(\ell)$.

4. What happens to $r(\ell)$ as $\ell \to \infty$?

5. Do you have any intuition on the previous result?
3.3.2 First-Step Analysis

Consider a gambler; one of those hard-core TAB types. She has \( X(\ell) \) dollars at day \( \ell \). Her goal is to reach \( L \) dollars, since this is the amount needed for the new tattoo she wants\(^1\). She attends the bookies daily and is determined to gamble her one dollar a day, until she reaches either \( L \) or goes broke, reaching \( 0 \). On each gamble (in each day) she has a chance of \( p \) of earning a dollar and a chance of \( 1 - p \) of loosing a dollar.

This problem is sometimes called the gambler’s ruin problem. We can view her fortune as the state of a Markov chain on state space, \( \mathcal{S} = \{0, 1, 2, \ldots, L - 1, L\} \).

Exercise 3.3.6. Specify the transition probabilities \( p_{i,j} \) associated with this model.

At day \( \ell = 0 \), our brave gambler begins with \( X(0) = x_0 \) dollars. As she drives to the bookies, Jimmy texts her: “Hey babe, I was wondering what is the chance you will eventually reach the desired \( L \) dollars?” She thinks while driving, but can’t concentrate, so she stops the car by the side of the road and sketches out the following in writing:

Define, 
\[
\tau_0 := \inf \{\ell \geq 0 : X(\ell) = 0\}, \quad \tau_L := \inf \{\ell \geq 0 : X(\ell) = L\}.
\]

These two objects are random variables which are called hitting times (the time it takes till hitting a state for the first time). They are random because different realisations of \( X(\cdot, \omega) \) imply different values for \( \tau_0 \) or \( \tau_L \). Note that the infimum of the empty set is defined to be \( \infty \). So if our gambler, for example reaches \( L \), then \( \tau_0 = \infty \) and similarly if the other case occurs.

In terms of hitting times, Jimmy’s question to our gambler, was to evaluate:
\[
q_i := \mathbb{P}(\tau_L < \tau_0 \mid X(0) = i), \quad \text{with} \quad i = x_0.
\]

We define \( q_i \) for all states \( i \), because to evaluate \( q_{x_0} \) we will need the other \( q_i \) also. It is obvious that \( q_0 = 0 \) and \( q_L = 1 \) but what if \( i \in \{1, \ldots, L - 1\} \)? Well here we can partition the event \( \{\tau_L > \tau_0\} \) based on the first step:

\[
q_i = \mathbb{P}(\tau_L < \tau_0 \mid X(0) = i) \\
= \mathbb{P}(\tau_L < \tau_0 \mid X(0) = i, X(1) = i + 1) p_{i,i+1} + \mathbb{P}(\tau_L < \tau_0, \mid X(0) = i, X(1) = i - 1) p_{i,i-1} \\
= \mathbb{P}(\tau_L < \tau_0 \mid X(1) = i + 1) p_{i, i+1} + \mathbb{P}(\tau_L < \tau_0, \mid X(1) = i - 1) p_{i, i-1} \\
= \mathbb{P}(\tau_L < \tau_0 \mid X(0) = i + 1) p_{i+1, i} + \mathbb{P}(\tau_L < \tau_0, \mid X(0) = i - 1) p_{i-1, i} \\
= q_{i+1} p + q_{i-1} (1-p).
\]

So using this first step analysis we end up with \( L + 1 \) equations for the \( L + 1 \) unknowns

---

\(^1\)The tattoo will feature the name of her boyfriend, “Jimmy” together with a picture of a Holden.
$q_0, q_1, \ldots, q_L$:

$$
\begin{bmatrix}
1 & 0 & \cdots & 0 \\
(1 - p) & -1 & p & \cdots & 0 \\
0 & (1 - p) & -1 & p & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
(1 - p) & -1 & p & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & (1 - p) & -1 & p & 0 & 1
\end{bmatrix}
\begin{bmatrix}
q_0 \\
q_1 \\
q_2 \\
\vdots \\
q_{L-2} \\
q_{L-1} \\
q_L
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
0 \\
\vdots \\
0 \\
0 \\
1
\end{bmatrix}
$$

The unique solution to these equations happens to be,

$$q_i = \begin{cases} 
i/L & \text{if } p = 1/2, \\ 
1 - \left(1 - \frac{1}{p}\right)^i & \text{if } p \neq 1/2. 
\end{cases} \quad (3.13)$$

**Exercise 3.3.7.** Verify that the solution given above is correct.

1. *(Analytically)* – Plug it in the equations and see it satisfies them.
2. *(Numerically)* – Make a $10 \times 10$ matrix in matlab (or anything else) and see that the vector $q_i$ solves the equations above.
3. *(Simulation)* – Simulate this gambler's ruin problem for some given parameters (say with $L = 9$) to verify that $q_i$ is indeed correct. Basically do this by generating sample paths $X(\cdot, \omega)$ for all times $\ell$, till $\min\{\tau_0, \tau_L\}$.

**Exercise 3.3.8.** Assume you didn’t know the formula in (3.13). Think of methods in which you can obtain it. Outline your methods. Try to start with $p = 1/2$ and then move onto $p \neq 1/2$.

The concept of *first step analysis* goes hand in hand with Markov chains and is useful for a variety of settings. When our gambler finished the calculations above, she texted Jimmy the result ($q_{x_0}$) and drove off. But then she got another text: “Honey love, for how many more days will you do this? Can’t wait babe!” She thinks, and then figures out that Jimmy wants to know,

$$m_i := \mathbb{E}[\min\{\tau_0, \tau_L\} \mid X(0) = i] \quad \text{with} \quad i = x_0.$$ 

By now our gambler knows how to do first step analysis, even while driving. She formulates the following: First,

$$m_0 = 0 \quad \text{and} \quad m_L = 0.$$
Even Jimmy can do this part. But further for \( i \in \{1, 2, \ldots, L - 1\} \):

\[
m_i = p_{i,i+1}(1 + m_{i+1}) + p_{i,i-1}(1 + m_{i-1})
= 1 + p_{i,i+1}m_{i+1} + p_{i,i-1}m_{i-1}
= 1 + pm_{i+1} + (1 - p)m_{i-1}
\]

So again we have \( L + 1 \) equations with \( L + 1 \) unknowns.

**Exercise 3.3.9.** Find the solution when \( p = 1/2 \).

**Exercise 3.3.10.** Find the solution when \( p \neq 1/2 \).

### 3.3.3 Class Structure, Periodicity, Transience and Recurrence

**Note:** Some of the derivations in this section are informal. Nevertheless, the reader should know that without much extra effort, all of the results can be proved in a precise manner.

One way to visualise the transition matrix of a finite DTMC is by drawing the weighted graph associated with \( P \). Edges associated with \((i, j)\) such that \( p_{i,j} = 0 \) are omitted. If you ignore the weights you simply get a directed graph. What does this graph tell you? Well, by studying it, you can see which paths the process may possibly take, and which paths are never possible. Of course, if \( p_{i,j} > 0 \) for all state pairs, then there is nothing to do because you have a complete graph. But in applications and theory, we often have \( p_{i,j} = 0 \) for a significant portion of the tuples \((i, j)\). This allows us to study the directed graph that has edge \((i, j)\) only when \( p_{i,j} > 0 \). This graph obviously doesn’t specify all of the information about the DTMC, but it does tell us the class structure. We describe this now.

![Figure 3.1: DTMC Transition Diagram](image)

We say that two states, \( i \) and \( j \) communicate if there are two non-negative integers \( t_1 \) and \( t_2 \) such that \( p_{i,j}^{(t_1)} > 0 \) and \( p_{j,i}^{(t_2)} > 0 \). This implies there is a path (in the directed graph) from \( i \) to \( j \) and a path from \( j \) to \( i \). We denote communication of \( i \) and \( j \) by \( i \leftrightarrow j \). The relation of communication is an equivalence relation\(^2\) over the set of states.

---

\(^2\)If for some reason you don’t know what an equivalence relation is, don’t stress. You’ll understand from the text.
Namely: $i \leftrightarrow i$ (reflexivity); if $i \leftrightarrow j$ then $j \leftrightarrow i$ (symmetry); and finally if $i \leftrightarrow j$ and $j \leftrightarrow k$ then $i \leftrightarrow k$ (transitivity).

**Exercise 3.3.11.** Use the Chapman-Kolmogorov equations to prove transitivity.

The implication of the fact that $\leftrightarrow$ is an equivalence relation is that it induces equivalence classes, $C_1, C_2, \ldots$ that are a partition of $S$. That is, $C_i$ and $C_j$ are disjoint for $i \neq j$ and $\cup C_i = S$. All states within class $C_i$ communicate with each other, but do not communicate with states that are not in $C_i$. Obviously for finite state spaces of size $N$, there can be at most $N$ classes and this upper bound is achieved only when $P = I$, the identity matrix. At the other extreme, we are often interested in Markov chains with only one class. Such Markov chains are said to be irreducible.

A state $i$ is said to have a period of $d$ if $p^{(\ell)}_{i,i} = 0$ for all integers $\ell$ that are not divisible by $d$, and further $d$ is the greatest integer with this property. E.g., assume, that $p^{(3)}_{i,i} > 0$, $p^{(6)}_{i,i} > 0$, $p^{(9)}_{i,i} > 0$ etc... and further $p^{(\ell)}_{i,i} = 0$ for $\ell \notin \{3, 6, 9, \ldots\}$. So if we start at time 0 in state $i$ we can only expect to be in state $i$ at the times 3, 6, 9, ..., etc. It isn’t guaranteed that at those times we visit state $i$, but we know that if we do visit state $i$, it is only at those times. It can be shown that all states in the same class have the same period. But we won’t ponder on that. In general, we aren’t so interested in periodic behaviour, but we need to be aware of it. In particular note that if $p_{i,i} > 0$ for all states $i$, then the Markov chain is guaranteed to be non-periodic.

Define now, the hitting time\(^3\) (starting at 1): $\tau_i = \inf\{\ell \geq 1 \mid X(\ell) = i\}$ and define,

$$f^{(\ell)}_{i,j} = \begin{cases} \Pr(\tau_j = \ell \mid X(0) = i) & \text{if } \ell \geq 1, \\ 0 & \text{if } \ell = 0. \end{cases}$$

Further define $f_{i,j} = \sum_{\ell=1}^{\infty} f^{(\ell)}_{i,j}$. This is the probability of ever making a transition into state $j$, when starting at state $i$:

$$f_{i,j} = \Pr\left(\sum_{\ell=1}^{\infty} 1\{X(\ell) = i\} \geq 1 \mid X(0) = i\right).$$

A state $i$ is said to be recurrent if $f_{i,i} = 1$. This means that if $X(0) = i$ we will continue visiting the state again and again. A state that is not recurrent is transient; i.e. i.e., $f_{i,i} < 1$ then there is a non-zero chance $(1 - f_{i,i})$ that we never return to the state.

**Exercise 3.3.12.** Assume that $X(0) = i$ and state $i$ is transient. Explain why the distribution of the number of visits to state $i$ after time 0, is geometric with success probability $1 - f_{i,i}$ and mean $1/(1 - f_{i,i})$. I.e.,

$$\Pr\left(\sum_{\ell=1}^{\infty} 1\{X(\ell) = i\} = n \mid X(0) = i\right) = (1 - f_{i,i})(f_{i,i})^n, \quad n = 0, 1, 2, \ldots$$

\(^3\)Some authors refer to the case starting at time 1 as a first passage time and to the case starting at time 0 as a hitting time. This distinction only matters if the initial state is $i$ itself.
Further, write an expression (in terms of \( f_{i,j} \) values) for,

\[
P\left( \sum_{\ell=1}^{\infty} 1\{X(\ell) = j\} \geq 1 \mid X(0) = i \right) = n
\]

In certain cases, it is obvious to see the values of \( f_{i,j} \):

**Exercise 3.3.13.** Consider the Markov chain with transition matrix,

\[
P = \begin{bmatrix}
0 & 0.3 & 0.7 & 0 \\
0 & 0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}.
\]

1. What are the classes of the Markov chain.
2. Which states are transient, and which are recurrent.
3. What are \( f_{i,j} \) for all \( i,j \)?

Consider now the following example,

\[
P = \begin{bmatrix}
0.1 & 0.7 & 0.2 & 0 \\
0.4 & 0.3 & 0.3 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}.
\]

The classes of this example are \( C_1 = \{1,2\} \), \( C_2 = \{3\} \) and \( C_3 = \{4\} \). Here without doing any calculations it is already obvious that \( f_{3,3} = 1 \) and \( f_{4,4} = 1 \), since states 3 and 4 are recurrent. They are even called absorbing, because once you get to state 3 or state 4, you never leave. So \( f_{3,i} = 0 \) for \( i \neq 3 \) and further \( f_{4,i} = 0 \) for \( i \neq 4 \). But the values \( f_{i,j} \) with \( i \in \{1,2\} \) are not as clear. Starting in state 1, for example, there is a 0.2 chance of absorbing in 3 and and with the complement there is a chance of staying within the class \( C_1 \). So how does this affect \( f_{1,1} \)?

The general mechanism we can use is first step analysis. This is the basic equation:

\[
f_{i,j} = P\left( \sum_{\ell=1}^{\infty} 1\{X(\ell) = j\} \geq 1 \mid X(0) = i \right)
= \sum_{k \neq j} \left[ \sum_{\ell=1}^{\infty} 1\{X(\ell) = j\} \geq 1 \mid X(0) = i, X(1) = k \right] p_{i,k}
+ \sum_{\ell=1}^{\infty} 1\{X(\ell) = j\} \geq 1 \mid X(0) = i, X(1) = j \right] p_{i,j}
= \sum_{k \neq j} f_{k,j} p_{i,k} + p_{i,j}
= \sum_{k \neq j} p_{i,k} f_{k,j} + p_{i,j}.
\]
Exercise 3.3.14. This exercise relates to the matrix $P$ in (3.14).

1. Find $f_{1,3}$ and $f_{1,4}$ (you’ll need to find out other $f_{i,j}$ values for this).

2. Explain why $f_{1,3} + f_{1,4} = 1$.

3. Run a simulation to verify your calculated value of $f_{1,3}$.

There are many characterisations of recurrent and transient states. One neat characterisation is the following:

State $i$ is recurrent if and only if
\[ \sum_{\ell=0}^{\infty} p_{i,i}^{(\ell)} = \infty. \] (3.15)

The idea of the derivation looks at the expected number of visits to the state:

\[ \mathbb{E} \left[ \sum_{\ell=0}^{\infty} \mathbb{1}_{\{X(\ell) = i\}} \mid X(0) = i \right] = \sum_{\ell=0}^{\infty} \mathbb{E} \left[ \mathbb{1}_{\{X(\ell) = i\}} \mid X(0) = i \right] = \sum_{\ell=0}^{\infty} p_{i,i}^{(\ell)} \]

Now for a recurrent state, we know that $\sum_{\ell=0}^{\infty} \mathbb{1}_{\{X(\ell) = i\}} = \infty$ and thus the expectation of this random variable should also be $\infty$. So this shows the direction $\Leftarrow$. For the other direction assume that state $i$ is transient (the contrapositive). In this case we saw that $\sum_{\ell=0}^{\infty} \mathbb{1}_{\{X(\ell) = i\}}$ is a geometric random variable with finite expectation, so $\sum_{\ell=0}^{\infty} p_{i,i}^{(\ell)} < \infty$.

In many cases, we can’t explicitly compute $p_{i,i}^{(\ell)}$ so there isn’t much computational use for (3.15). But one classic fascinating example is the simple random walk. For this we assume now a state is $S = \mathbb{Z}$ (not finite any more!). Take $p \in [0,1]$ and set,

\[ p_{i,j} = \begin{cases} 
  p & \text{if } j = i + 1, \\
  (1-p) & \text{if } j = i - 1, \\
  0 & \text{otherwise}. 
\end{cases} \]

The example is called a random walk because at every time step the walker takes either a step up with probability $p$ or a step down with probability $1-p$. It is called simple, because the change at each time point is a random variable with support $\{-1,1\}$. In the general random walk, steps would be of arbitrary magnitude.

A nice feature of this model is that we can actually calculate $p_{i,i}^{(\ell)}$.

Exercise 3.3.15. Verify the following:

1. If $p = 0$ or $p = 1$ there is an infinite number of classes, but if $p \in (0,1)$ the model is irreducible.

   For the rest of the items below, assume $p \in (0,1)$. 
2. The model is periodic with period 2. So now we will consider $P_{i,i}^{(2\ell)}$, since for $\ell \in \{1, 3, 5, 7, \ldots \}$, $P_{i,i}^{(\ell)} = 0$.

3. Explain why:

$$P_{i,i}^{(2\ell)} = \binom{2\ell}{\ell} p^\ell (1 - p)^\ell.$$  

4. Now use the Stirling approximation for $\ell!$ (see Appendix) to show,

$$P_{i,i}^{(2\ell)} \sim \frac{(4p(1-p))^\ell}{\sqrt{\pi \ell}},$$  

where the symbol $\sim$ implies that as $\ell \to \infty$ the ratio of the left hand side and the right hand side goes to 1.

5. Verify (using the definition of convergence of a series), that if $a_\ell \sim b_\ell$ then $\sum_\ell a_\ell < \infty$ if and only if $\sum_\ell b_\ell < \infty$.

6. Verify that

$$\sum_{\ell=0}^{\infty} \frac{(4p(1-p))^\ell}{\sqrt{\pi \ell}} = \infty,$$

if and only if $p = 1/2$ (otherwise the series converges).

With the results of the above exercise we know that state $i$ (for any $i$) is recurrent if and only if $p = 1/2$. That is if $p \neq 1/2$ then all states are transient. Loosely speaking, the chain will “drift off” towards $+\infty$ if $p > 1/2$ and towards $-\infty$ if $p < 1/2$. States may be revisited, but ultimately, each state $i$ will be revisited only a finite number of times.

In finite Markov chains, we can’t have all states transient:

**Exercise 3.3.16.** Argue why a finite DTMC, must have at least one recurrent state.

In the infinite state space case, we can sometimes have that,

$$E[\tau_i \mid X(0) = i] = \infty,$$

even when state $i$ is recurrent. Such is actually the case for the simple random walk in the symmetric case ($p = 1/2$). This cannot happen when the state space is finite. This phenomenon is called null-recurrence. The other case,

$$E[\tau_i \mid X(0) = i] < \infty,$$

is referred to as positive-recurrence. In finite state space DTMC all recurrent states are positive-recurrent. Further, in the finite state space case, if the DTMC is irreducible then all states are recurrent and thus all states are positive-recurrent. Further on this is in Chapter 6 dealing with stability.
3.3.4 Limiting Probabilities

We are often interested in the behaviour of \( \{X(\ell)\} \) over long time periods. In applied mathematics, infinity, is a good approximation for “long”. There is much to say here and we will only cover a small portion of the results and cases. Specifically, let us now assume that our DTMC has finite state-space, that it is irreducible, and that it is aperiodic (all states have a period of 1). Limiting probability results often hold when these assumptions are partially relaxed, but one needs to take more care in specifying the results.

To illustrate the main idea we return to exercise (3.3.3). If your example chain for that exercise had \( p_{i,i} \in (0, 1) \) then the above conditions are satisfied. Let us assume that this is the case. Now ask

“Over the long range, in what proportion of my days am I happy?”

Remembering that our code for “happy” was 1, the question can be posed as finding

\[
\pi_1 := \lim_{\ell \to \infty} \mathbb{E} \left[ \sum_{\ell=0}^{\ell} 1 \{ X(\ell) = 1 \} \right].
\]

The value \( \pi_1 \) is then referred to as the limiting probability of being in state 1. I should hope that for your Markov chain of exercise (3.3.3), \( \pi_1 \) is high (close to 1). How can we evaluate it? The key result is that we can solve the system of equations:

\[
\begin{align*}
\pi_1 &= \pi_1 p_{1,1} + \pi_2 p_{2,1} + \pi_3 p_{3,1}, \\
\pi_2 &= \pi_1 p_{1,2} + \pi_2 p_{2,2} + \pi_3 p_{3,2}, \\
\pi_3 &= \pi_1 p_{1,3} + \pi_2 p_{2,3} + \pi_3 p_{3,3}, \\
1 &= \pi_1 + \pi_2 + \pi_3.
\end{align*}
\]

Now the unique solution, \([\pi_1, \pi_2, \pi_3]\) gives the long range proportion during which state \( i \) is occupied. Note that we have 4 equations with only 3 unknowns, but we should in fact omit one (any one) of the first 3 equations (this is a consequence of the fact \( P \) is a singular matrix). These equations are called the balance equations. In matrix form they are compactly written with \( \pi \) taken as a row vector and \( 1 \) a column vector of 1’s.

\[
\begin{align*}
\pi &= \pi P, \\
1 &= \pi 1.
\end{align*}
\]  \( (3.16) \)

Remember that, \( r(\ell+1) = r(\ell) P \). What is a fixed point of this linear dynamical system? Fixed points \( r \) need to satisfy: \( r = r P \). One obvious such fixed point is \( 0' \). But this fixed point is not a probability distribution. Is it the only fixed point? What if \( P \) has an eigenvalue equal to 1? In this case any (left) eigenvector corresponding to the eigenvalue 1 is a fixed point. One such (left) eigenvector is \( \pi \). Indeed the Perron-Frobenious theorem implies that \( P \) has an eigenvalue of 1.

\( ^4 \) Beware of such questions if your current age is \( 10 + n \pm \epsilon \) where \( \epsilon \) is small. Such thoughts can throw you on soul adventures that you may end up regretting – or maybe not.
Exercise 3.3.17. Consider your matrix $P$ of exercise (3.3.3). Use a computer for the following:

1. Solve the balance equations for $\pi$.

2. Run a single simulation of the DTMC for $T = 10,000$ time points. Choose any initial distribution for $X(0)$. Evaluate for $i \in \{1, 2, 3\}$,
   \[ \hat{\pi}_i := \frac{\sum_{\ell=0}^{T} 1\{X(\ell) = i\}}{T}, \]
   compare these values to the answer of item 1.

3. Compute $P^5, P^{10}, P^{20}$ and $P^{100}$. Compare the rows of these matrices with the answer of item 1.

4. The numerical illustration of the previous item, indicates that the rows all converge to $\pi$. If this is indeed true (which it is), argue that for any initial distribution, $r(0)$,
   \[ \lim_{\ell \to \infty} r(\ell) = \pi. \]

The numerics of the above exercise, indicate the validity of the following (we omit the proof – note also that there are much more general formulations):

**Theorem 3.3.18.** Consider a finite DTMC that is irreducible and non-periodic. Then,

1. The balance equations (3.16) have a unique solution with $\pi_i \in (0, 1)$.

2. It holds that for any $i \in S$,
   \[ \lim_{\ell \to \infty} p_{i,j}^{(\ell)} = \pi_j. \]

3. It holds that,
   \[ \pi_i = \frac{1}{\mathbb{E}[\tau_i \mid X(0) = i]}. \]

4. For any function, $f : S \to \mathbb{R}$, we have with probability one,
   \[ \lim_{\ell \to \infty} \frac{\sum_{k=0}^{\ell} f(X(k))}{\ell} = \sum_{i \in S} \pi_i f(i). \]

So basically, knowing $\pi$ gives us much information about the long run or steady state behaviour of the system. When talking about long range behaviour it is $\pi$ that matters; the initial distribution, $r(0)$ becomes insignificant. Item 4 (also called the ergodic property) shows that long range behaviour can be summarised in terms of $\pi$.

One of the names of the distribution $\pi$ is the stationary distribution also known as the
**CHAPTER 3. LINEAR DYNAMICAL SYSTEMS AND MARKOV CHAINS (13H)**

**invariant distribution.** A process \( \{X(\cdot)\} \) (in discrete or continuous time) is stationary if for any integer \( k \geq 0 \) and any time values, \( t_1, \ldots, t_k \), and any integer \( \tau \),

\[
P(X(t_1) = i_1, \ldots, X(t_k) = i_k) = P(X(t_1 + \tau) = i_1, \ldots, X(t_k + \tau) = i_k).
\]

**Exercise 3.3.19.** Use the equations describing \( \pi \) to show:

1. If we start at time 0 with \( r(0) = \pi \), then \( r(1) = \pi \) and this holds for all \( r(\ell) \).
2. More generally, show that if we start at time 0 with \( r(0) = \pi \) then the process is stationary.

So when we look at a DTMC, we can consider the *stationary version* where we choose \( r(0) = \pi \). This means we are looking at the system which is already in “statistical equilibrium”. Such systems may not exactly occur in practice, but it is often a very sensible approximation for systems that have been running for some time.

If on the other hand \( r(0) \neq \pi \), then the DTMC is not stationary. But still, if we let it run for some time, it can be approximately considered to be stationary. This is due to item 2 of the theorem above.

### 3.4 Markov Chains in Continuous Time

Note that we use the phrase *Markov chain* for the case when the state space is countable, reserving the phrase *Markov process* for the case when the state space is continuous (this usage is not universal). We now discuss *Continuous Time Markov Chains (CTMC)*.

#### 3.4.1 Continuous Time Basics

Informally a finite state CTMC, is a process \( \{X(t)\} \) in continuous time that satisfies:

1. Lack of memory (Markovian property):

\[
P(X(t + s) = j \mid X(t) = i \text{ and further info about } X(u) \text{ for } u \in [0, t)) = P(X(t + s) = j \mid X(t) = i).
\]

2. Time Homogeneity:

\[
P(X(t + s) = j \mid X(t) = i) = P(X(s) = j \mid X(0) = i).
\]

3. Finite state space: There is some finite set (state space), \( \mathcal{S} \), such that,

\[
P(X(t) \not\in \mathcal{S}) = 0, \quad \forall t.
\]
In case $\mathcal{S}$ is not finite, but rather countably infinite, the process is still a CTMC and many of the results hold.

Suppose $X(0) = j$ and $T(j)$ is the first time the CTMC leaves $j$. Then

\[
\mathbb{P}(T(j) > t + s \mid T(j) > s) = \mathbb{P}(X(v) = j, 0 \leq v \leq t + s \mid X(u) = j, 0 \leq u \leq s) = \mathbb{P}(X(v) = j, s < v \leq t + s \mid X(s) = j) \quad \text{(Markov)}
\]

\[
= \mathbb{P}(X(v) = j, 0 < v \leq t \mid X(0) = j) \quad \text{(homogeneous)}
\]

\[
= \mathbb{P}(T(j) > t)
\]

So $T(j)$ has the memoryless property, and is thus exponentially distributed (see the extensive discussion about the exponential distribution in Chapter 2).

As in discrete time, we can specify an initial probability (row) vector, $r$ with

\[ r_i = \mathbb{P}(X(0) = i). \]

But how do we specify the transition rule? Instead of a probability transition matrix $P$, what governs the evolution of a continuous-time Markov chain is an infinitesimal generator $Q = [q_{ij}]_{i,j \in \mathcal{S}}$, with components $q_{ij}$, for $i \neq j$, being the instantaneous transition rate of going from $i$ to $j$, for $i,j \in \mathcal{S}$. In other words, for a sufficiently small interval of time $h > 0$

\[
\mathbb{P}(X(t+h) = j \mid X(t) = i) = q_{ij}h + o(h),
\]

where $o(h)$ goes to zero faster than $h$ does. The matrix $Q$ has non-positive elements on the diagonal ($q_{ii} \leq 0$ for $i \in \mathcal{S}$), and nonnegative elements off-diagonal ($q_{ij} \geq 0$ for $i \neq j$). Further, $Q1 = 0$. Since each row sums to 0 it implies that,

\[ q_{jj} = -\sum_{k \neq j} q_{jk}. \]

A consequence is that starting at $X(0) = i$, the Markov chain stays in this state for an exponentially-distributed amount of time, with rate $-q_{ii}$, then moves to $k$ with probability $q_{ik}/(-q_{ii})$ (see the discussion of a race between independent exponential random variables in Chapter 2). Then, it stays in state $k$ for an exponentially-distributed amount of time, with rate $-q_{kk}$, so on. Given that the chain is in state $i$, the exponential duration of time the chain will stay in a state and the next state to be jumped to are independent. Note that CTMCs are sometimes called Markov Jump Processes (MJP).

For small $h$,

\[
\mathbb{P}(X(t+h) = k \mid X(t) = j) = p^{(h)}_{j,k} \approx (I + hQ)_{jk}
\]

\[
= \begin{cases} 
    h q_{jk}, & \text{if } j \neq k, \\
    1 + hq_{jj}, & \text{if } j = k.
\end{cases}
\]
Figure 3.2: A CTMC Transition Diagram with probabilities of transitions on the arcs. The alternative is to mark transition rates on the arcs.

So we can think of $q_{j,k}$ as the rate of transition from $j$ to $k$, with

$$q_{j,k} \quad \text{being} \quad \begin{cases} \geq 0 \text{ if } k \neq j, \\ \leq 0 \text{ if } k = j. \end{cases}$$

The total rate of leaving state $j$ is $\sum_{k \neq j} q_{j,k} = -q_{j,j}$, so the exponential duration of time spent in state $j$ has parameter $\lambda_j = -q_{j,j}$.

To see where the CTMC moves upon leaving state $j$, observe that, for $k \neq j$,

$$\mathbb{P}(X(h) = k \mid X(h) \neq j, X(0) = j) = \frac{p^{(h)}_{j,k}}{\sum_{l \neq j} p^{(h)}_{j,l}} \to \frac{q_{j,k}}{-q_{j,j}} \quad \text{as} \quad h \to 0.$$  

That is, when the CTMC leaves state $j$, it has probability $-q_{j,k}/q_{j,j}$ of moving to state $k$.

An alternative way of thinking about a CTMC involves competing alarm clocks that ring at exponentially distributed times with different rates: for each state $i$, we set up a clock $C_{ij}$ for every state $j \neq i$ to which the system can move from $i$ in one transition. Each clock $C_{ij}$ rings after an interval of time exponentially distributed with rate $q_{ij}$, and we assume that these random intervals are mutually independent. Then, from $i$ the chain moves to whichever $j$ whose clock $C_{ij}$ rings first. As the minimum of independent exponential random variables is an exponential random variable, the time that system remains in $i$ is also exponentially distributed, with rate $-q_{ii} = \sum_{j \neq i} q_{ij}$. Thanks to the memoryless property of exponential distributions, we do not need to reset the clocks after the system moves to some state $k \neq i$—we just need to consider another set of clocks $C_{kj}$.

Note that the index of the winner of the clocks is independent of the value it got. This allows a basic mechanism to for simulating a CTMC.

**Exercise 3.4.1.** Describe how to simulate a CTMC based on generation of exponential random variables and generation of random variables $I_i, i \in S$ each with support $j \in S \setminus \{i\}$, distributed according to the probabilities $\{q_{i,j} / -q_{i,i}\}$.

Observe that
The Chapman-Kolmogorov equations for a CTMC are given by

\[ p_{i,j}^{(s+t)} = \sum_{k \in S} \mathbb{P}(X(s + t) = j | X(s) = k, X(0) = i) \mathbb{P}(X(s) = k | X(0) = i) = \sum_{k \in S} p_{i,k}^{(s)} p_{k,j}^{(t)}. \]

These equations express the transition probabilities in continuous time. In matrix form, we write

\[ P(t) = [p_{j,k}^{(t)}]. \]

Then, for \( s, t \geq 0 \), the Chapman-Kolmogorov equations can be expressed in the form,

\[ P(t+s) = P(t) P(s). \]

For non-explosive CTMCs, the matrix \( Q \) determines the transition probability completely by solving the backward or forward equations to get

\[ P(t) = e^{Qt}, \]

subject to \( P(0) = I \).

As a consequence,

\[ \mathbb{P}(X(t) = j | X(0) = i) = [e^{Qt}]_{ij} \quad \text{for } i, j \in S, \]

and the distribution vector \( r(t) \) with components \( r_i(t) = \mathbb{P}(X(t) = i) \) is given by

\[ r(t) = r(0)e^{Qt}. \]

Notice that the matrix \( e^{Qt} \) is stochastic, and plays a similar role as the probability transition matrix \( P \) in a discrete-time Markov chain.

There are, obviously, differences and similarities between discrete-time Markov chains (DTMC) and continuous-time Markov chains (CTMC). In both cases, given its current state, where the system will jump to next does not depend on its past trajectory; however, the time between two successive transitions is one unit of time for a DTMC, and is exponentially distributed with a state-dependent rate for a CTMC. Furthermore, the general concepts of limiting behaviours and state properties carry from discrete-time to continuous-time context, but the associated mathematical expressions differ.

The forward equations are:

\[ \frac{d}{dt} P(t) = Q P(t), \]

and the backward equations are:

\[ \frac{d}{dt} P(t) = P(t) Q. \]
This summary table is for discrete and continuous time:

<table>
<thead>
<tr>
<th></th>
<th>DTMC</th>
<th>CTMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit of time</td>
<td>One step</td>
<td>&quot;dt&quot;</td>
</tr>
<tr>
<td>Basic info</td>
<td>$P$</td>
<td>$Q$</td>
</tr>
<tr>
<td>Distribution propagation</td>
<td>$P(t) = P^t$</td>
<td>$P(t) = e^{Qt}$</td>
</tr>
<tr>
<td>Evolution</td>
<td>geometric times+jumps</td>
<td>exponential times+jumps</td>
</tr>
<tr>
<td>Stationarity</td>
<td>$\pi P = \pi$</td>
<td>$\pi Q = 0$</td>
</tr>
</tbody>
</table>

### 3.4.2 Further Continuous Time Properties

The concepts in discrete-time Markov chains of one state being accessible from another, of one state communicating with another, and of the chains being irreducible or reducible, are still applicable in continuous time.

A state $i$ is said to be absorbing if, once entering this state, the system remains in this state forever. That is, $q_{ij} = 0$ for all $j \in S$.

**Definition 3.4.2.** We say that a vector $\pi = (\pi_i)_{i \in S}$ is a stationary distribution of a continuous-time Markov chain $\{X(t)\}$ if $\pi$ satisfies the conditions

$$\pi Q = 0,$$

$$\pi 1 = 1.$$  \hspace{1cm} (3.19) \hspace{1cm} (3.20)

One cannot define periodicity for continuous-time chains in a similar fashion to discrete-time chains; somewhat unexpectedly, this means we have a stronger version of convergence to stationary for irreducible CTMCs.

**Theorem 3.4.3.** Every irreducible finite-state continuous-time Markov chain has a unique stationary distribution vector $\pi$, which is also the limiting distribution of the chain:

$$\lim_{t \to \infty} r(t) = \pi,$$  \hspace{1cm} (3.21)

for every initial distribution $r(0)$, where $r(t)$ denotes the probability distribution vector at time $t$.

Other results regarding the stationary distribution also hold. Namely time averages and the relation between mean return time to a state and the stationary distribution.

A phenomena that may occur in CTMCs with infinite (countable) state spaces is explosion. This means that the chain makes an infinite number of transitions in finite time.

### 3.5 Elementary Structured Markov Models

We now consider the most basic types of structured Markov Models.
3.5.1 Birth-and-Death Processes

A continuous-time Birth-and-Death process is a Markov chain \( \{ X(t) : t \geq 0 \} \) on the countably infinite state space \( S = \{ 0, 1, 2, 3, \ldots \} \) (i.e. \( S = \mathbb{Z}_+ \)) of which the generator has the following tridiagonal structure

\[
Q = \begin{bmatrix}
-\lambda_0 & \lambda_0 & 0 & 0 & \ldots \\
\mu_1 & -\mu_1 - \lambda_1 & \lambda_1 & 0 & \ldots \\
0 & \mu_2 & -\mu_2 - \lambda_2 & \lambda_2 & \ldots \\
0 & 0 & \mu_3 & -\mu_3 - \lambda_3 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix},
\tag{3.22}
\]

where \( \lambda_n \) and \( \mu_n \) are real nonnegative numbers for all \( n \).

![Figure 3.3: Birth Death Process on \( S = \{ 0, 1, \ldots \} \)](image)

In this process, the only possible transitions from a given state \( n \) are to the state \( n - 1 \) with rate \( \mu_n \), or to the state \( n + 1 \) with rate \( \lambda_n \). So, the process stays in state \( n \geq 1 \) for an exponentially distributed period of time with parameter \( (\mu_n + \lambda_n) \), at the end of which it moves to state \( n - 1 \) with probability \( \mu_n / (\mu_n + \lambda_n) \) (this corresponds to a “death” event), or to state \( n + 1 \) with probability \( \lambda_n / (\mu_n + \lambda_n) \) (this corresponds to a “birth” event). When in state 0, the process moves to state 1 with probability one after an exponentially distributed period of time with parameter \( \lambda_0 \).

3.5.2 The Poisson Process

Before discussing a few examples of Birth-and-Death processes, we start with an important example of a pure birth process called the Poisson process.

Consider a system with no death events and where the births correspond to the arrival of events occurring independently of each other and such that the interarrival times are identically distributed according to an exponential distribution with parameter \( \lambda \). Let \( X(t) \) represent the number of arrivals in the system in the time interval \([0, t]\). The process \( \{ X(t) : t \geq 0 \} \) is a counting process called a Poisson process with parameter (or rate) \( \lambda \).

It is a Markov chain with generator \( Q \) as in (3.22) where \( \mu_n = 0 \) and \( \lambda_n = \lambda > 0 \) for all \( n \). Indeed, recall from Section 3.3 that for any \( n \geq 0 \), the entry \( q_{n,n+1} \) of the generator \( Q \) is defined as

\[
P(X(t + h) = n + 1 | X(t) = n) = q_{n,n+1} h + o(h), \quad \text{for small } h.
\]
Then
\[
P[X(t + h) = n + 1 | X(t) = n] = \int_0^h (e^{-\lambda(h-u)}) \lambda e^{-\lambda u} du
\]
\[
= \lambda h e^{-\lambda h}
\]
\[
= \lambda h (1 - \lambda h + o(h))
\]
\[
= \lambda h + o(h),
\]
so that \(q_{n,n+1} = \lambda\) for all \(n\). The Markov chain \(\{X(t) : t \geq 0\}\) is transient, there is therefore no stationary distribution.

**Theorem 3.5.1.** The number of arrivals in the system in the time interval \([0, t]\), \(X(t)\), is Poisson distributed with parameter \(\lambda t\).

**Exercise 3.5.2.** Show that the above holds using the forward equations.

The Poisson process is a natural modelling tool for a variety of phenomena such as the arrival of phone calls at a switchboard, the particles emission by a radioactive substance, the arrival of cars at a roundabout, the arrival of items at a station of a manufacturing process (see the third application example we discussed in Section 1.3), or the arrival of customers at a counter.

There are many other basic properties of a Poisson process that one often studies in an introductory stochastic processes course. We do not go into further details here, but rather list these points (further details are to appear in the Markov Chains Appendix):

- A special (central) case within the class of Renewal-Processes where the time-stationary and event-stationary versions agree.
- Can be defined as the only simple counting process that is Levy.
- The uniform (order statistic) property.
- Poisson superposition.
- Poisson splitting.

Some generalizations of the Poisson Poisson are the compound Poisson process, time-varying Poisson process, doubly stochastic Poisson process (Cox Process) and general Poisson processes on Metric spaces. We do not discuss these further here.

**The Markovian branching process.** Consider a population model where all individuals behave independently of each other and according to the same rules: each individual lives for an exponentially distributed time with parameter \(\mu\) and generates new individuals during its lifetime according to a Poisson process with rate \(\lambda\) (that is,
there is one Poisson process per living individual. Let $X(t)$ represent the population size at time $t$. Then $\{X(t), t \geq 0\}$ is a Birth-and-Death process with $\mu_n = n\mu$ and $\lambda_n = n\lambda$ for $n \geq 0$.

We can show that the mean population size at time $t$, $m(t) = \mathbb{E}[X(t)]$, satisfies the ordinary differential equation

$$\dot{m}(t) = (\lambda - \mu) m(t),$$

so that $m(t) = e^{(\lambda - \mu)t} m(0)$. We thus see that the mean population size explodes if and only if $\lambda > \mu$.

**Exercise 3.5.3.** Draw a parallel between the previous result and Example 1.1.

Note that in a branching process in absence of immigration, the state 0 is absorbing and we can show that all other states $n \geq 1$ are transient. There is therefore no stationary distribution.

One quantity of interest in branching processes is the probability that, starting from a given state (initial population size) $n \geq 1$, the process eventually reaches the absorbing state 0 (that is, the population eventually becomes extinct). This probability is referred to the extinction probability of the branching process.

Branching processes form a fascinating branch of applied probability, and it is out of the scope of the present book to study them in more details here.

### 3.5.3 The Birth-Death Stationary Distribution.

Assume $\lambda_n, \mu_n > 0$ for all $n$. The infinite stationary distribution vector $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ satisfies the (infinite) system of equations

$$\pi Q = 0',$$

$$\pi 1 = 1,$$

of which the solution exists if and only if the process is positive recurrent.

**Theorem 3.5.4.** The stationary distribution vector $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ of the Birth-and-Death process with generator (3.22) satisfies the recurrence

$$\pi_n = \pi_{n-1} \frac{\lambda_{n-1}}{\mu_n}, \quad \text{for } n \geq 1.$$

As a consequence, $\pi_n$ can be expressed in terms of $\pi_0$ for any $n \geq 1$:

$$\pi_n = \pi_0 \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n}. \quad (3.23)$$

The Birth-and-Death process is positive recurrent if and only if

$$\sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} < \infty. \quad (3.24)$$
Proof: The system of equations

\[ \pi Q = 0', \]
\[ \pi 1 = 1, \]

becomes here

\[ \mu_1 \pi_1 = \lambda_0 \pi_0, \quad (3.25) \]
\[ (\mu_n + \lambda_n) \pi_n = \lambda_{n-1} \pi_{n-1} + \mu_{n+1} \pi_{n+1}, \quad n \geq 1, \quad (3.26) \]
\[ \pi_0 + \pi_1 + \pi_2 + \ldots = 1. \quad (3.27) \]

From (3.25) and (3.26) we obtain that the stationary probabilities satisfy the recurrence

\[ \pi_n = \pi_{n-1} \frac{\lambda_{n-1}}{\mu_n}, \quad \text{for } n \geq 1, \]

from which the expression (3.23) of \( \pi_n \) in terms of \( \pi_0 \) is straightforward. As a consequence, (3.27) can be rewritten as

\[ \pi_0 \left[ 1 + \sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \right] = 1, \]

or equivalently

\[ \pi_0 = \left[ 1 + \sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} \right]^{-1}, \]

and the process is positive recurrent if and only if

\[ \sum_{n \geq 1} \frac{\lambda_0 \lambda_1 \ldots \lambda_{n-1}}{\mu_1 \mu_2 \ldots \mu_n} < \infty. \]

\[ \square \]

3.5.4 Simple Queueing Models

Queueing models constitute some of the most basic (and useful) examples of birth-death processes. In general, a queueing system is composed of an arrival process, a service mechanism and of other rules that describe the operation of the system.
The M/M/1 queue. Consider a queueing system with a single server, in which customers arrive according to a Poisson process with rate \( \lambda \) and service times have an exponential distribution with parameter \( \mu \). Let \( X(t) \) denote the number of customers present in the system at time \( t \), including the one in service (if there is one). The process \( \{X(t) : t \geq 0\} \) is a Birth-and-Death process with \( \mu_n = \mu \) and \( \lambda_n = \lambda \) for all \( n \).

Let \( \rho = \lambda/\mu \) be the traffic intensity; this ratio represents the average number of new arrivals in the system during the service time of one customer. From (3.23), the stationary distribution \( \pi = (\pi_0, \pi_1, \pi_2, \ldots) \) of the queue length satisfies

\[
\pi_n = \pi_0 \rho^n.
\] (3.28)

The process is positive recurrent (or stable) if and only if

\[
\sum_{n \geq 1} \rho^n < \infty \iff \lambda < \mu,
\]

that is, if on average arrivals happen slower than service completions. In that case, from (3.27) and (3.28), \( \pi_0 \) satisfies

\[
\pi_0 \left[ 1 + \sum_{n \geq 1} \rho^n \right] = 1 \Rightarrow \pi_0 = 1 - \rho,
\]

and we finally obtain

\[
\pi_n = (1 - \rho) \rho^n, \quad \text{for } n \geq 0.
\] (3.29)

We thus see that the stationary queue length has a geometric distribution with parameter \( 1 - \rho \). The steady-state mean queue length is then given by

\[
\sum_{n=0}^{\infty} n \pi_n = \frac{\rho}{1 - \rho},
\]
which has a vertical asymptote at $\rho = 1$.

**The M/M/$\infty$ queue.** Consider a queueing system with infinitely many servers operating in parallel and independently of each other, so that every arriving customer is served immediately (there is no waiting time).

This model corresponds to a Birth-and-Death process with $\lambda_n = \lambda$ and $\mu_n = n\mu$ for all $n$.

**The M/M/$c$ queue.** Consider a multi-server queueing system with $c \geq 1$ servers operating in parallel and independently of each other, in which arrivals and service times follow the same rules as in the M/M/1 queue.

This model corresponds to a Birth-and-Death process with $\lambda_n = \lambda$ and $\mu_n = \min(n, c)\mu$ for all $n$. 

Figure 3.5: A finite service system with renegging.

Figure 3.6: The M/M/1 Transition Diagram.
Exercise 3.5.5. Determine the stationary distribution and the stability condition for the M/M/c queue.

The M/M/c/K queue. We can assume that a queueing system with c ≥ 1 servers has a finite capacity K (with K ≥ c). If the state of the system is K, every new arrival is considered as lost.

Exercise 3.5.6. Show that this queueing model corresponds to a Birth-and-Death process (by specifying $\lambda_n$ and $\mu_n$ for all n), and determine its stationary distribution and the stability condition.

Other birth-death Queueing Systems

The M/M/s/K+M feature customer abandonments at rate $\gamma$. The generator matrix is of this form:

\[
Q = \begin{bmatrix}
0 & 1 & 2 & 3 & \cdots & s & (s+1) & (s+2) & \cdots & (s+K-1) & (s+K) \\
-\lambda & \lambda & . & . & . & . & . & . & . & . & . \\
\mu - (\lambda + \mu) & \lambda & -2\lambda & . & . & . & . & . & . & . & . \\
2\mu - (\lambda + 2\mu) & -\lambda & \lambda & -\lambda & . & . & . & . & . & . & . \\
3\mu - (\lambda + 3\mu) & -\lambda & -\lambda & -\lambda & . & . & . & . & . & . & . \\
& & & & & & & & & & \ddots \\
&s\mu - (\lambda + s\mu) & -\lambda & -\lambda & . & . & . & . & . & . & . \\
(s+1)\mu - (\lambda + (s+1)\mu) & -\lambda & -\lambda & . & . & . & . & . & . & . \\
(s+2)\mu - (\lambda + (s+2)\mu) & -\lambda & -\lambda & . & . & . & . & . & . & . \\
\cdots & \cdots & \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
(s+K-1)\mu - (\lambda + (s+K-1)\mu) & -\lambda & -\lambda & . & . & . & . & . & . & . & . \\
(s+K)\mu & -\lambda & -\lambda & . & . & . & . & . & . & . & . \\
\end{bmatrix}
\]

Assume now that the system starts empty. Then the transient distributions follow:

Exercise 3.5.7. Consider a finite population queue: Inhabitants of an Island with a population of 15 occasionally go to use an internet stand on the Island, queueing when the stand is occupied. The rate of desire to use the stand is $\lambda$. The service rate is $\mu$. Assume exponential service times (and further typical assumptions). What is the stationary distribution?

Exercise 3.5.8. Compare the above to the Erlang Loss System.
3.6 \((A, B, C, D)\) Linear Input-Output Systems

In Chapter 2 we looked at input-output LTI systems (SISO versions). Then earlier in the current chapter (starting in Section 3.1) we looked at linear dynamical systems. In that case we did not consider input and output, instead we considered the state of the process. We now combine the two types of objects to get systems that we call \((A, B, C, D)\) Linear Input-Output Systems. Among other things, these systems will serve as the basis for linear control theory to be studied in Chapter 5.

These systems relate 3 processes: input \(u(\cdot)\), state \(x(\cdot)\) and output \(y(\cdot)\). As their name suggests, \((A, B, C, D)\) systems are parameterized by 4 matrices: \(A \in \mathbb{R}^{n \times n}\), \(B \in \mathbb{R}^{n \times m}\), \(C \in \mathbb{R}^{p \times n}\) and \(D \in \mathbb{R}^{p \times m}\). So the dimension of the state is \(n\), the dimension of the input is \(m\) and the dimension of the output is \(p\). The SISO case is when \(m = 1\) and \(p = 1\) (yet does not require that \(n = 1\)).

The input – state – output evolution of such systems is defined as follows:

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) \\
y(t) &= Cx(t) + Du(t)
\end{align*}
\]

\[
\begin{align*}
x(\ell + 1) &= Ax(\ell) + Bu(\ell) \\
y(\ell) &= Cx(\ell) + Du(\ell)
\end{align*}
\] (3.30)

for the continuous time and discrete time cases respectively. Our presentation here focuses primarily on the continuous time case. Observe that if \(B = 0\) and/or \(u(\cdot) \equiv 0\), then the state evolution is as that of an autonomous linear dynamical system (as the systems presented in Section 3.1). Yet if \(B \neq 0\) and \(u(\cdot)\) takes non-zero values then the state-evolution is modified/altered by the input.

In control theory applications (handled in Chapter 2), the matrix \(A\) indicates the “untouched behavior of the plant”, the matrix \(B\) indicates the “effect of actuators on the plant”, the matrix \(C\) indicates the “sensors in the system” and the matrix \(D\) indicates the “effect that the input has directly on the output”.

Observe that if \(C = I\) and \(D = 0\) then \(y(\cdot) = x(\cdot)\). I.e. the output of the system is exactly the state. Yet in applications, \(C\) is typically a “flat long” matrix \((p < n)\) while \(B\) is a “tall thin” matrix \((n > m)\). Such dimensions represent the fact that “there are not many sensors” and “there are not many actuators” respectively (“not many” is compared to the number of state variables). In the extreme SISO case, \(C\) is in fact a row vector (which we shall denote by \(c'\)) and \(B\) is a column vector (which we shall denote by \(b\)).

---

5This is a name we have given, it is not a standard name.
3.6. $(A, B, C, D)$ LINEAR INPUT-OUTPUT SYSTEMS

Further, $d$ is a scalar. In that case the $(A, b, c', d)$ system dynamics are,

$$
\dot{x}(t) = Ax(t) + bu(t), \\
y(t) = c'x(t) + du(t),
$$

or similarly for the discrete time case.

### 3.6.1 Time-Domain Representation of the Output

Specify now initial conditions,

$$
x(0) = x_0.
$$

If $x_0 = 0$ we say the system is starting at rest. We now have,
Theorem 3.6.1.

\[ y(t) = Ce^{At}x_0 + C \int_0^t e^{A(t-s)}Bu(s)ds + Du(t), \quad (3.32) \]

\[ y(\ell) = CA^\ell x_0 + C \sum_{k=0}^{\ell-1} A^{\ell-(k+1)}Bu(k) + Du(\ell). \quad (3.33) \]

The continuous time equation (3.32) can be obtained from Picard iterations. The discrete time equation is easily obtained by recursing the stem equations:

**Exercise 3.6.2.** Prove (3.33).

We can now verify using Theorem 3.6.1 that the mapping,

\[ y(\cdot) = O(u(\cdot)) \]

is LTI if the system is starting at rest. That is, \((A, B, C, D)\) systems yield linear time invariant input output systems (in the sense of Chapter 2).

In this MIMO-LTI setting the impulse response generalizes to the matrix impulse response. Focusing on the continuous time version, we assume it admits an integral representation,

\[ y(t) = O(u(\cdot))(t) = \int_{-\infty}^{\infty} h(t-\tau)u(\tau)d\tau = (h * u)(t), \]

with \(h(t) \in \mathbb{R}^{p \times m}\) being the impulse response matrix. Note that for inputs \(\{u(t)\}\) that have coordinates 0 except for the \(j\)'th coordinate, \(u_j(t)\), the \(i\)'th component of the output has the form,

\[ y_i(t) = \int_{-\infty}^{\infty} h_{ij}(t-\tau)u_j(\tau)d\tau, \]

as a SISO system with impulse response \(h_{ij}(t)\).

Any MIMO LTI system is said to be causal if and only if \(h(t) = 0_{p \times n}\) for \(t < 0\) and thus for inputs with positive support,

\[ y(t) = \int_0^t h(t-\tau)u(\tau)d\tau. \]

We can further get the following useful representations:

\[ h(t) = 1_{p \times p}(t)\left(Ce^{At}B + D\delta_{m \times m}(t)\right), \quad (3.34) \]

where we use a diagonal matrix of \(m\) delta-functions, \(\delta_{m \times m}(t)\).

**Exercise 3.6.3.** Argue the validity of (3.34) based on Theorem 3.6.1.
In the SISO case, (3.34) reads,

\[ h(t) = c'e^{At}b + d\delta(t). \]  

Further in this case, the step response is:

\[ H(t) = \int_0^t h(s)ds = d + c'\left(\int_0^t e^{As}ds\right)b. \]  

Hence when \( A \) is non-singular, the step response in the SISO case reads:

\[ H(t) = d - c'A^{-1}b + c'e^{At}A^{-1}b. \]  

### 3.6.2 The Resolvent

We now wish to get the Laplace transform of the impulse response matrix. As a first step, let us consider the autonomous system \( \dot{x} = Ax \) with \( x(0) = x_0 \) (studied in Section 3.1). Using the derivative property of Laplace transforms we have:

\[ s\hat{x}(s) - x_0 = A\hat{x}(s), \]  

and thus for \( s \) that are not eigenvalues of \( A \),

\[ \hat{x}(s) = (sI - A)^{-1}x_0. \]  

Hence the Laplace transform matrix of \( e^{At} \) is \( (sI - A)^{-1} \). This is called the resolvent of the system.

Note that the resolvent yields an additional method for computing \( e^{At} \). Here is an example:

**Example 3.6.4.** Consider,

\[ A = \begin{bmatrix} -1 & 3 \\ 0 & 1 \end{bmatrix}. \]

Then,

\[
(sI-A)^{-1} = \begin{bmatrix} s+1 & -3 \\ 0 & s-1 \end{bmatrix}^{-1} = \frac{1}{(s+1)(s-1)} \begin{bmatrix} s-1 & 3 \\ 0 & s+1 \end{bmatrix} = \begin{bmatrix} \frac{1}{s+1} & \frac{3/2}{s-1} - \frac{3/2}{s+1} \\ 0 & \frac{1}{s-1} \end{bmatrix}
\]

So,

\[ e^{At} = \begin{bmatrix} e^{-t} & \frac{3}{2}(e^t - e^{-t}) \\ 0 & e^t \end{bmatrix}. \]
3.6.3 The Transfer Function Matrix

The relation of convolutions and Laplace transforms carries over easily to the non-scalar version here. If the matrix Laplace transform, \( \hat{h}(s) \) of \( h(\cdot) \) exists then,

\[
\dot{y}(s) = \hat{h}(s) \hat{u}(s).
\]

A matrix Laplace transform such as this is simply a Laplace transform of each of the elements. In this case, \( \hat{h}(s) \) is the transfer function matrix.

Building on the idea of the resolvent, the transfer function takes on a very specific form for \((A, B, C, D)\) systems. We can extend the resolvent to \((A, B, C, D)\) systems by mimicking (3.37), this time for \( \dot{x}(t) = Ax(t) + Bu(t) \) (starting at rest):

\[
s \hat{x}(s) - 0 = A \hat{x} + B \hat{u}.
\]

This yields (for \( s \) values that are not eigenvalues of \( A \)):

\[
\hat{x}(s) = (sI - A)^{-1}B \hat{u}(s).
\]

Substitution in \( y(t) = Cx(t) + Du(t) \) we get,

\[
\hat{y}(s) = (C(sI - A)^{-1}B + D) \hat{u}(s).
\]

We have thus derived the following representation of the transfer function matrix for \((A, B, C, D)\) systems:

**Theorem 3.6.5.**

\[
\hat{h}(s) = C(sI - A)^{-1}B + D. \tag{3.38}
\]

In the SISO case, (3.38) is a scalar function and reads.

\[
\hat{h}(s) = c'(sI - A)^{-1}b + d. \tag{3.39}
\]

**Exercise 3.6.6.** Explain why the elements of (3.38) as well as (3.39) are rational functions.

We have just shown that all \((A, B, C, D)\) systems have rational Laplace transforms. We further have the following (without proof):

**Theorem 3.6.7.** Any matrix of rational functions (or a single scalar rational function) treated as a transfer function, is the transfer function of an \((A, B, C, D)\) system.

The action of finding an \((A, B, C, D)\) system that has a given rational transfer function is called realization (do not confuse this with the other meaning of the word “realization” that is synonymous with a sample path of a random process). In practice one often tries to realize a system by choosing “physical components” that have a given behavior (specified by the transfer function). A bit more on this is in Chapter 5.

Note that there is not a unique \((A, B, C, D)\) system corresponding to a transfer function. This is illustrated now through equivalent representations.
3.6.4 Equivalent Representations of Systems

Given \( P \in \mathbb{R}^{n \times n} \), with \( \det(P) \neq 0 \), we can change the coordinates of the state-space as follows:

\[ P\tilde{x} = x. \]

By substitution in the system equations, we see that resulting system is,

\[ \left( \tilde{A}, \tilde{B}, \tilde{C}, \tilde{D} \right) = \left( P^{-1}AP, P^{-1}B, CP, D \right). \]  (3.40)

Both systems have the same external representations (i.e. same impulse response/transfer function) and are thus called equivalent systems.

Note that the matrices \( A \) and \( \tilde{A} \) are similar, hence making such a change of coordinates is sometimes referred to as performing a similarity transform on the system.

**Exercise 3.6.8.** Prove (3.40) using either Theorem 3.6.1 or Theorem 3.6.5.

3.6.5 Rational Laplace-Stieltjes Transforms Revisited

In Chapter 2 we saw the straightforward connection between LTI input output systems and probability distributions. That chapter ended with a few example probability distributions whose Laplace transform is rational. In the previous section we have seen that all rational transfer functions may be realized as \((A, B, C, D)\) systems. The class of such systems whose step response, \( H(t) \) is a probability distribution is called a matrix exponential distribution (ME). We define this in detail now.

Given \( A \in \mathbb{R}^{n \times n} \) with \( \det(A) \neq 0 \), vectors \( b, c \in \mathbb{R}^n \) and a scalar \( c_0 \geq 0 \), consider the following function:

\[ F(t) = \begin{cases} 
0, & t < 0, \\
b, & t = 0, \\
d - c'A^{-1}b + c'e^{At}A^{-1}b, & t > 0.
\end{cases} \]

If \( F(t) \) satisfies the properties of a distribution then \( F(\cdot) \) is said to be a matrix exponential distribution of order \( n \).

The Laplace-Stieltjes transform (LST) is given by

\[ \hat{F}(s) = c'(sI - A)^{-1}b + d. \]  (3.41)

We have the following:

**Theorem 3.6.9.** If a distribution has a rational LST then it can be represented as a ME distribution.

In the next section we look at a special sub-set of ME distributions whose parameters bear probabilistic meaning.

Note: Given \( \hat{F}(\cdot) \) that is rational. It is not a trivial matter to check if it corresponds to a distribution (i.e. it is not easy to verify if the corresponding step-response is monotonic). If we know it corresponds to a distribution, then that distribution is ME.
3.7 Phase-Type (PH) Distributions

Having seen the family of matrix exponential distributions we now define a sub-class of these distributions whose construction is based on hitting times of Markov chains. We call such distributions phase-type distributions.

Phase-type distributions have had a profound effect on applied probability and stochastic modeling in the past few decades. As will be demonstrated in the next chapter, they essentially allow to incorporate behaviors of arbitrary distributions in stochastic models that are governed by CTMCs (remember that the “basic distribution” in CTMCs is the exponential distribution – this is quite restrictive from a modeling point of view).

3.7.1 The Absorption Time in an Absorbing CTMC

Consider a CTMC \( \{X(t) : t \geq 0\} \) on the finite state space \( S = \{0, 1, 2, \ldots, m\} \), where state 0 is absorbing and states \( \{1, 2, \ldots, m\} \) are transient. We denote the generator of the CTMC as the \((m+1) \times (m+1)\) matrix \(Q\) and define it shortly.

Let \( \tau \) denote the hitting time of state 0:

\[
\tau := \inf\{t : X(t) = 0\}.
\]

Regardless the initial (transient) state, the Markov chain will eventually hit the absorbing state in a finite time with probability one, therefore \( \tau < \infty \) almost surely (that is, the distribution of \( \tau \) is nondefective or proper).

To construct, \( Q \), take the vector \( c \in \mathbb{R}^m \) with \( c'1 \leq 1 \) and let \((c_0, c')'\) denote the probability distribution (row vector) of \( X(0) \). That is,

\[
c_0 = 1 - c'1,
\]

and,

\[
P(X(0) = i) = c_i.
\]

Now take \( A \in \mathbb{R}^{m \times m} \) with \( \det(A) \neq 0 \), negative entries on the diagonal positions, non-negative entries on the off-diagonal positions and \( A1 \leq 0 \). Such a matrix is called a sub-generator. Construct now \( Q \) as follows:

\[
Q = \begin{bmatrix}
0 & 0' \\
-b & A
\end{bmatrix},
\]

where the column vector \( b := -A1 \) can be interpreted as the absorption rate vector (in state 0).

**Exercise 3.7.1.** Argue why \( Q \) is a generator matrix where state 0 is absorbing (and thus recurrent) and states \( \{1, \ldots, m\} \) are transient.
We call the distribution of the random variable $\tau$, a phase type distribution. It is parameterized by $c$ and $A$. This is because for every $c$ and $A$ we have a CTMC and every CTMC implies the behavior of the hitting time random variable. We thus use the notation $PH(c', A)$.

Let $F(t) := \mathbb{P}(\tau \leq t)$ denote the distribution function of $PH(c', A)$. We have the following:

**Theorem 3.7.2.**

$$
F(t) = \begin{cases} 
0, & t < 0, \\
0, & t = 0, \\
1 - c' e^{At}, & t > 0.
\end{cases}
$$

**Exercise 3.7.3.** Show that the density (excluding the possible atom $c_0$ at 0) is

$$
f(t) = c' e^{At} b.
$$

**Exercise 3.7.4.** Show that the LST is

$$
\hat{f}(s) = c_0 + c'(sI - A)^{-1} b.
$$

Using the LST, it is a standard matter to obtain the moments:

**Theorem 3.7.5.** Let $\tau \sim PH(c', A)$, then

$$
\mathbb{E}[^{\tau^k}] = (-1)^k k! c' A^{-k} 1.
$$

### 3.7.2 Examples

- An exponential distribution with parameter $\lambda$ is a very special case of a PH distribution with $n = 1, c = 1, A = -\lambda$, and $b = \lambda$.

- Let $Z$ be distributed according to an Erlang $E(n, \lambda)$; in other words, $Z$ represents the sum of $n$ independent exponential random variables with parameter $\lambda$. Then the distribution of $Z$ can be seen as particular PH distribution of order $n$, that is, $Z \sim PH(c, A)$ with

$$
c' = [1, 0, \ldots, 0], \quad A = \begin{bmatrix} -\lambda & \lambda \\ -\lambda & -\lambda & \lambda \\ & \ddots & \ddots & \ddots \\ & & -\lambda & -\lambda & \lambda \\ & & & & -\lambda \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \lambda \end{bmatrix}.
$$

- Let $Z$ be distributed according to an hyperexponential distribution with density

$$
f_Z(z) = \sum_{1 \leq k \leq n} c_k \lambda_k e^{-\lambda_k z}, \quad \text{where } c_k > 0 \text{ for all } k, \text{ and } \sum_{1 \leq k \leq n} c_k = 1.
Then $Z$ is the convex mixture of $n$ exponential random variables, and $Z \sim PH(c', A)$ with

$$c' = [c_1, c_2, \ldots, c_n], \quad A = \begin{bmatrix} -\lambda_1 & -\lambda_2 & \cdots & -\lambda_n \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}, \quad b = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}.$$
3.7. PHASE–TYPE (PH) DISTRIBUTIONS

Figure 3.12: Erlang Distribution PDF

3.7.3 A Dense Family of Distributions

Any non-negative distribution may be approximated by a PH distribution. The approximation becomes better as the number of phases grows. In fact, if the space of all non-negative distributions is taken as a metric space (one needs to define a metric for this – we omit the details), it can be shown that the set of PH distributions is dense in that space (analogy: the rational numbers are dense in the reals).

Exercise 3.7.6. Think how to approximate an arbitrary non-negative distribution by using mixtures of Erlang distributions where the number of phases in the Erlang distributions is large.

3.7.4 Relationship to ME Distributions

Note that PH distributions form a special class of ME distributions since both distributions share the same structural properties. Yet for PH distributions, the vectors \( c \) and \( b \), and the matrix \( A \) characterizing the PH distribution are respectively probability mass vectors and a sub-generator matrix (i.e. they have probabilistic meaning). As opposed to that, the parameters of ME distributions bear no probabilistic meaning.

Exercise 3.7.7. In what way is a PH distribution a special case of an ME distribution? That is, given a PH\((c',A)\) distribution function, represent it as an ME\((\cdot)\) distribution function.

Theorem 3.7.8. There exists distributions that are ME, yet are not PH.
Figure 3.13: Mixture of Erlang Random Variables

Figure 3.14: A mixture of a Uniform random variable on (2,5), with p=0.7, and a Log Normal with $\mu = 0$ and $\sigma = 1$, with $1 - p = 0.3$ is approximated by $k$ Erlangs, each the sum of $n$ exponentials.

### 3.7.5 Operations on PH Random Variables

The class of PH distributions is closed with respect to:

1. Multiplication by a constant

2. Addition (of independent random variables)

3. Mixtures

4. Minimum
3.7. PHASE–TYPE (PH) DISTRIBUTIONS

Figure 3.15: Transitions of a State \((i, j) \in S_1 \times S_2\)

Figure 3.16: Two Dimensional Representation of Minimum of Two Phasetypes

Figure 3.17: PH Representation of Generalised Erlang Random Variable
Figure 3.18: Two Dimensional PH Representation of Minimum of Two Generalised Erlang Random Variables
3.7. PHASE-TYPE (PH) DISTRIBUTIONS

Figure 3.19: Two Dimensional PH Representation of Minimum of Two Hyperexponential Random Variables

Figure 3.20: The Minimum of Two Hyperexponential Random Variables is also Hyperexponential
3.7.6 Moment Matching

When modelling, we are often faced with the task of choosing a distribution based on some observed moments. In the simplest case we are given the mean and perhaps the variance. An alternative view is to use the squared coefficient of variation \( c_v^2 \). How can we match PH distributions for this?

\( c_v^2 < 1 \) Generalized Erlang Distribution Paramatrization

When processes with coefficients under 1 are considered, they can be modelled with the help of Generalized Erlang Distributions. This type of distribution is different from regular Erlang Distributions in the sense that it does not use \( n \) exponentially distributed steps with parameter \( \mu \), but one step with parameter \( \mu_1 \) and \( (n-1) \) steps with parameter \( \mu_2 \). These means are as follows:

\[
\mu_1 = \frac{n}{1 + \sqrt{(n-1)(nc^2 - 1)}} \\
\mu_2 = \frac{\mu_1 (n-1)}{\mu_1 - 1}
\]

In which the number of steps is based on \( c_v^2 \) as follows: \( n = \frac{1}{c_v^2} \).

\( c_v^2 > 1 \) Hyper Exponential Paramatrization

In order to generate distributions with a coefficient of variation \( c_v^2 > 1 \), a Hyper Exponential distribution can be used. In order to visualise the effects of an increasing \( c_v^2 \) on the queue length, the mean of the service process is fixed to be 1. By then varying the mean of the arrival process up to 1, it is possible to generate values for the queue length for various utilizations of the system.

In the Hyper Exponential case, two nodes are considered with rate parameters \( \mu_1 \) and \( \mu_2 \). The first node is entered with probability \( p \), while the second is entered with probability \( 1 - p \). As mentioned before, the mean of this process is fixed to be 1. This mean can be expressed as a function of the means of the nodes as follows:

\[
\frac{1}{\mu} = 1 = \frac{p}{\mu_1} + \frac{(1-p)}{\mu_2}
\]

(3.42)

Additionally, the coefficient of variation is a function of the variance and the mean: \( c_v^2 = \frac{\sigma^2}{\mu^2} \). However, with the mean set to be one, this equation simplifies to the variance alone. The formula can then be expressed as follows:
\[ c^2 = 2 \left( \frac{p}{\mu_1^2} + \frac{(1-p)}{\mu_2^2} \right) - 1 \]  
(3.43)

\[ \frac{c^2 + 1}{2} = \frac{p}{\mu_1^2} + \frac{(1-p)}{\mu_2^2} \]  
(3.44)

From 3.42 the following expression for \( \mu_2 \) is derived:

\[ \mu_2 = \frac{(1-p)}{(1 - \frac{p}{\mu_1^2})} \]  
(3.45)

Next, from 3.43 an expression for \( p \) can be derived, by filling in the previous equation:

\[ p = \frac{(c^2_v - 1)}{(c^2_v + 1 + \frac{2}{\mu_1^2} - \frac{4}{\mu_1})} \]  
(3.46)

This expression can then be used on 3.45 to derive an expression that solely depends on \( \mu_1 \) and \( c^2 \). This equation, together with 3.46, yields the following set of expressions:

\[ p = \frac{(c^2_v - 1)}{(c^2_v + 1 + \frac{2}{\mu_1^2} - \frac{4}{\mu_1})} \]  
(3.47)

\[ \mu_2 = \frac{2(1-\mu_1)}{2 - \mu_1(c^2_v + 1)} \]  
(3.48)

Here it should be noted that an extra restriction on \( \mu_1 \) should be imposed as \( 2 - \mu_1(c^2_v + 1) > 0 \). This restriction results in the following inequality:

\[ \mu_1 < \frac{2}{c^2_v + 1} \]  
(3.49)

This set of equations then allows us to choose values for any given \( c^2_v \). As an example one can consider a process with mean \( \mu = 1 \) and \( c^2_v = 7 \). The inequality then shows that \( \mu_1 < \frac{1}{4} \), so \( \mu_1 = \frac{1}{5} \) for example. Next \( p = \frac{3}{19} \) and \( \mu_2 = 4 \) are obtained. These values can be verified by filling them in for the definitions of the mean and coefficient of variation:
\[
\mu = \frac{p}{\mu_1} + \frac{(1-p)}{\mu_2} = \frac{15}{19} + \frac{4}{19} = 1
\]
\[
c_v^2 = 2 \left[ p(1-p) \begin{bmatrix} \frac{1}{\mu_1} & 0 \\ 0 & \frac{1}{\mu_2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right] - 1
\]
\[
c_v^2 = 2 \left( \frac{3}{19} \left( \frac{1}{5} \right)^2 + \frac{16}{19} \right) - 1 = 2 \left( \frac{76}{19} \right) - 1 = 7
\]

It should be noted that any value for \(c_v^2\) can be approximated with the aforementioned \(\mu_1\), as long as \(c_v^2 < \left( \frac{2-\mu_1}{\mu_1} \right)\) holds. Alternatively, a very small \(\mu_1\) allows for a large range of \(c_v^2\) to be approximated.

### 3.8 Relations Between Discrete and Continuous Time

We now explore relationships between CTMCs and related DTMCs as well as between \((A, B, C, D)\) systems and PH/ME distributions of both discrete and continuous time.

#### 3.8.1 Different Discretizations of a CTMC

When considering a CTMC, \(\{X(t)\}\) with generator matrix \(Q\) there are several ways in which we can associate a DTMC, \(\{\tilde{X}(\ell)\}\) with this DTMC. The first way is *discrete time sampling*. Here the DTMC is the CTMC sampled every \(T\) time units, that is,

\[
\tilde{X}(\ell) = X(\ell T).
\]

In this case, we have that the transition probability kernel of the DTMC \(\tilde{X}(\cdot)\) is,

\[
P_1 = e^{Q T}.
\]

A second way is to consider the so called *embedded Markov chain* (also known as the *jump chain*). If the CTMC is irreducible we have that \(q_{i,i} < 0\) (strictly) and thus we can define,

\[
P_2 = I - \text{diag}(Q)^{-1}Q.
\]

This is basically a stochastic matrix where for \(i \neq j\), \(P_{i,j} = q_{i,j} / -q_{i,i}\) and for \(i = j\) we have \(P_{i,j} = 0\). More generally if we have that \(q_{i,i} = 0\) for some \(i\) (this corresponds to an absorbing state \(i\) in the CTMC), we should set in the embedded Markov chain, \(P_{i,i} = 1\). In summary, the embedded Markov chain represents the CTMC sampled at jump points.
A third way corresponds to CTMCs where there is an upper bound to \( \{-q_{i,i}\} \):

\[
\max_{i \in S} -q_{i,i} \leq \gamma.
\]

This always holds when \( S \) is finite. In this case, a uniformized chain is a DTMC with,

\[
P_3 = I + \frac{1}{\gamma}Q.
\]

**Exercise 3.8.1.** Show that \( P_1, P_2 \) and \( P_3 \) are stochastic matrices.

Note that in the uniformized chain, transitions from a state to itself are possible (in a single discrete time step). This is not the case for the embedded Markov chain. The idea of uniformization is to have a single Poisson process at rate \( \gamma \) that marks the transitions of all types. Whenever the “clock” of this Poisson process “rings” a transition is made according to \( P_3 \) (sometimes allowing transitions from a state to itself).

Let \( \pi \) be the stationary distribution of \( P_3 \). Then,

\[
\pi(I + \frac{1}{\gamma}Q) = \pi,
\]

and hence,

\[
\pi Q = 0.
\]

So the stationary distribution of \( P_3 \) and \( Q \) is the same. This in general does not hold for \( P_2 \) nor for \( P_1 \):

**Exercise 3.8.2.** Show by example that \( P_2 \) and \( P_1 \) in general have a different stationary distribution than \( Q \).

### 3.8.2 Sampling a Continuous Time \((A, B, C, D)\) System

Consider a continuous time \((A, B, C, D)\) system that is sampled at time intervals of \( T \). Assume that a piecewise constant input is applied: \( u(t) = u(\ell) \) for \( t \in [\ell T, (\ell + 1)T) \). In this case, the discrete time system,

\[
(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}) := \left( e^{AT}, \int_0^T e^{A\tau}d\tau B, C, D \right)
\]

agrees with the continuous time system \((A, B, C, D)\) at the sampling points.

**Exercise 3.8.3.** Prove this.
3.8.3 Discrete/Continuous, PH/ME Distributions Relationships

In similar vein to Markov chains and \((A, B, C, D)\) systems, there are obvious relationships between discrete and continuous PH/ME distributions. At this point we only focus on the exponential and geometric distribution:

**Exercise 3.8.4.** Let \(X \sim \exp(\lambda)\). Denote \(N = \lfloor X \rfloor\). Show that \(N\) is geometrically distributed and find its parameter.

### Bibliographic Remarks

### Exercises

Suppose \(P(X(0) = 1) = 1/3, P(X(0) = 2) = 0, P(X(0) = 3) = 1/2, P(X(0) = 4) = 1/6\) and

\[
P = \begin{pmatrix}
\frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & \frac{2}{3} & \frac{1}{3} \\
\frac{1}{2} & 0 & \frac{1}{2} & 0
\end{pmatrix}.
\]

- Find the distribution of \(X(1)\),
- Calculate \(P(X(\ell + 2) = 2|X(\ell) = 4)\), and
- Calculate \(P(X(3) = 2, X(2) = 3, X(1) = 1)\).

Sometimes we want to model a physical system where the future does depend on part of the past. Consider following example. A sequence of random variables \(\{X_n\}\) describes the weather at a particular location, with \(X_n = 1\) if it is sunny and \(X_n = 2\) if it is rainy on day \(n\).

Suppose that the weather on day \(n + 1\) depends on the weather conditions on days \(n - 1\) and \(n\) as is shown below:

<table>
<thead>
<tr>
<th>(n - 1)</th>
<th>(n)</th>
<th>(n + 1)</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>rain</td>
<td>rain</td>
<td>0.6</td>
</tr>
<tr>
<td>sunny</td>
<td>sunny</td>
<td>sunny</td>
<td>0.8</td>
</tr>
<tr>
<td>sunny</td>
<td>rain</td>
<td>rain</td>
<td>0.5</td>
</tr>
<tr>
<td>rain</td>
<td>sunny</td>
<td>sunny</td>
<td>0.75</td>
</tr>
</tbody>
</table>

If we put \(Y(\ell) = (X_{\ell-1}, X_\ell)\), then \(Y(\cdot)\) is a DTMC. The possible states are \(1' = (1, 1)\), \(2' = (1, 2)\), \(3' = (2, 1)\) and \(4' = (2, 2)\).
We see that $\{Y(\ell) : \ell \geq 1\}$ is a DTMC with transition matrix

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.75 & 0.25 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \end{pmatrix}.$$ 

**Example 3.8.5.** Give a criterion for ergodicity of the DTMC with state space $\{0, 1, 2, \cdots\}$ and transition matrix

$$P = \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$ 

When the DTMC is ergodic, derive its stationary distribution.

We saw that this DTMC is irreducible, aperiodic and recurrent when $p \leq q$. Solve the linear equations

$$(\pi_0, \pi_1, \cdots) = (\pi_0, \pi_1, \cdots)P$$

to get $\pi_k = (p/q)^k \pi_0$.

We also need $\sum_{k \geq 0} \pi_k = 1$. The sum on the left hand side is finite only if $p < q$, in which case $\pi_0 = 1 - (p/q)$ and so $\pi_k = [1 - (p/q)](p/q)^k$. So there is a solution to $\pi = \pi P$ with $\sum_{k \geq 0} \pi_k = 1$, and hence the DTMC is ergodic, only if $p < q$, in which case

$$\mu_k = \frac{1}{(p/q)^k(1 - (p/q))}.$$ 

A manufacturing machine at a factory is required in the production process non-stop (24 hours a day and 7 days a week). Nevertheless, the machine experiences both “off periods” and “on periods”, where in the former it is not operating due to maintenance or malfunction and in the later it is operating as needed.

In analyzing the performance of the factory, an elementary model for the machine is that of an alternating sequence of independent random variables,

$$X_1, Y_1, X_2, Y_2, X_3, Y_3, \ldots,$$

where $X_i \sim F_X(\cdot)$ represents an “on period” and $Y_i \sim F_Y(0)$ represents “off periods”. It is known that at time $t = 0$ the machine has just changed from “off” to “on”. In such a case, the state of the machine at time $t$ is represented by $X(t)$ (where say 0 implies “off” and 1 implies “on”).

As a first step it is assumed that, $F_X(t) = 1 - e^{-\mu t}$ and $F_Y(t) = 1 - e^{-\lambda t}$. In this case:
2. Argue why $X(t)$ is a CTMC. What is the generator matrix?

3. Simulate a random sample path of $\{X(t), t \in [0, 20]\}$ with $\mu = 2$ and $\lambda = 1$. Plot the trajectory that you have simulated.

4. Calculate the long term proportion of time (i.e. over $t \in [0, \infty)$) during which the machine is “on” (respectively “off”). State your result in terms of the symbols $\mu$ and $\lambda$.

5. Simulate a long trajectory and use your simulation result to verify your answer to the question above.

6. Let $q(t) = \mathbb{P}(\text{“on” at time } t)$. Estimate $\{q(t), t \in [0, 10]\}$ by means of simulation. Plot your result.

7. Now calculate $\{q(t), t \in [0, 10]\}$ numerically (without simulation). You may compare to the result above.

8. Now try to find a precise analytic expression for $q(t)$ (in terms of $\lambda$ and $\mu$), compare your expression to the result above.

9. Is the information that a change occurred “exactly at time 0” important or are the results the same if it were simply stated that the machine is “on” at time 0? Explain your result.

Exercise 3.8.6. Determine whether each of the following matrices is the generator of a Markov chain and, if yes, describe how the CTMC evolves ($\lambda$ and $\mu$ are nonnegative):

$$Q_1 = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}; \quad Q_2 = \begin{pmatrix} -2 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 2 & -3 \end{pmatrix}; \quad Q_3 = \begin{pmatrix} -2 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 2 & -3 \end{pmatrix}.$$ 

Assume further that after the machine is in “off” state it needs to be in “warmup” state before moving to “on”. Thus the operation of the machine is determined by the sequence,

$$X_1, Y_1, Z_1, X_2, Y_2, Z_2, X_3, Y_3, Z_3, \ldots,$$

where $X_i$ and $Y_i$ are distributed as before and the “warmup” periods $Z_i$ are as follows: $Z_i \sim F_Z(\cdot)$ with $F_Z(t) = 1 - e^{-\gamma t}$.

13. Repeat now questions 5, 7, 10, 11 (assuming $\gamma = 3$ for questions 6 and 7).

It was found now that there is a chance of $p$ that at the end of the warmup period the machine will enter "off" instead of “on”.

14. Repeat now questions 5, 7. Leave your answer symbolic in terms of $p$. 

3.8. RELATIONS BETWEEN DISCRETE AND CONTINUOUS TIME

The above CTMC model appears restrictive as it assumes that the distribution of “on”, “off” and “warmup” durations is exponential. Comparison to data indicates that it is plausible to assume “on” and ”off” durations are exponentially distributed, yet this is not the case for “warmup”. In that case, it is suggested to use a PH distribution with $m$ phases, $PH(c', A)$.

15. Incorporate the assumption about the PH distribution of “warmup” in the CTMC model. You should now have a Markov chain where $|S| = m + 2$. Write down the generator matrix of this CTMC.

The last exercise illustrated one of the strengths of PH distributions: They allow to incorporate the distribution of arbitrary behavior in a CTMC. Often the construction of the PH distribution constitutes modeling in its own right:

16. Assume that the “warmup duration” is either “long” or “short”. In the “long” case it is exponentially distributed with $\gamma_1$. In the “short” case it is exponentially distributed with $\gamma_2$. We have $\gamma_1 < \gamma_2$. There is a chance of $r \in (0, 1)$ that it is long, and a chance of $1 - r$ that it is short. This is an hyper-exponential distribution. Show how it is a special case of the PH distribution and incorporate it in the CTMC model.

17. Assume now that it is measured that “warm up periods” have a mean of $m$ and a squared coefficient of variation of $c^2 > 1$ (the squared coefficient of variation of a random variable is the variance divided by the mean squared). Show how to incorporate this in the CTMC by means of a PH distribution of order 2 yielding arbitrary mean and arbitrary squared coefficient of variation $> 1$.

18. Why is the restriction of $c^2 > 1$ important? Can you answer the case of $c^2 \in (0, 1)$ with only 2 phases? If not argue why not. As a bonus you may try to find a PH distribution of higher order for this.
\[
\begin{bmatrix}
(1,1) & (1,2) & \cdots & (1,n) \\
(1,1) & -q_{1,1} & \mu_{1,2} & \cdots & \mu_{1,n} \\
(1,2) & \mu_{2,1} & -q_{1,2} & \cdots & \mu_{2,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
(1,n) & \mu_{n,1} & \mu_{n,2} & \cdots & -q_{1,n} \\
(2,1) & \lambda_{2,1} & -q_{2,1} & \mu_{1,2} & \cdots & \mu_{1,n} \\
(2,2) & \lambda_{2,1} & \mu_{2,1} & -q_{2,2} & \cdots & \mu_{2,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
(2,n) & \lambda_{2,1} & \mu_{n,1} & \mu_{n,2} & \cdots & -q_{2,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
(m,1) & \lambda_{m,1} & \lambda_{m,2} & -q_{m,1} & \mu_{1,2} & \cdots & \mu_{1,n} \\
(m,2) & \lambda_{m,1} & \lambda_{m,2} & \mu_{2,1} & -q_{m,2} & \cdots & \mu_{2,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
(m,n) & \lambda_{m,1} & \lambda_{m,2} & \mu_{n,1} & \mu_{n,2} & \cdots & -q_{m,n} \\
(0) & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Figure 3.21: Generator of Minimum of Two PH Random Variables

![Diagram of Generator of Minimum of Two PH Random Variables]

Figure 3.22: Generalized Erlang System

![Diagram of Generalized Erlang System]

Figure 3.23: 2-Hyperexponential Distribution

![Diagram of 2-Hyperexponential Distribution]
3.8. RELATIONS BETWEEN DISCRETE AND CONTINUOUS TIME

Figure 3.24: CTMC Uniformization