Control of Stochastic Systems

Serdar Yüksel
Queen’s University, Mathematics and Engineering and Mathematics and Statistics

July 29, 2018
This document is a set of supplemental lecture notes that has been used for MATH 472/872: Control of Stochastic Systems, at Queen’s University since 2009. It has also been used somewhat regularly at Bilkent University for EEE 446/546, and also at the University of Passau.

Serdar Yüksel
# Contents

## 1 Review of Probability

1.1 Introduction ................................................................................................. 1
1.2 Measurable Space ......................................................................................... 1
  1.2.1 Borel $\sigma$–field .................................................................................. 2
  1.2.2 Measurable Function ............................................................................. 2
  1.2.3 Measure .................................................................................................. 3
  1.2.4 The Extension Theorem ......................................................................... 3
  1.2.5 Integration ............................................................................................... 4
  1.2.6 Fatou’s Lemma, the Monotone Convergence Theorem and the Dominated Convergence Theorem .................................................. 5
1.3 Probability Space and Random Variables ..................................................... 5
  1.3.1 More on Random Variables and Probability Density Functions .......... 6
  1.3.2 Independence and Conditional Probability ........................................... 7
1.4 Stochastic Processes and Markov Chains ..................................................... 7
  1.4.1 Markov Chains ....................................................................................... 8
1.5 Exercises ........................................................................................................ 8

## 2 Controlled Markov Chains

2.1 Controlled Markov Models ........................................................................... 11
  2.1.1 Fully Observed Markov Control Problem Model ................................ 11
  2.1.2 Classes of Control Policies ................................................................... 12
2.2 Performance of Policies ................................................................................. 12
2.3 Markov Chain Induced by a Markov Policy .................................................. 14
2.4 Partially Observed Models and Reduction to a Fully Observed Model .......... 14
2.5 Decentralized Stochastic Systems .................................................................. 16
2.6 Controlled Continuous-Time Stochastic Systems ........................................ 16
2.7 Exercises ............................................................................................................. 16

3 Classification of Markov Chains ........................................................................... 19
3.1 Countable State Space Markov Chains ................................................................. 19
   3.1.1 Recurrence and transience ............................................................................. 21
   3.1.2 Stability and invariant measures ................................................................. 22
   3.1.3 Invariant measures via an occupational characterization ......................... 23
3.2 Uncountable (Complete, Separable, Metric) State Spaces ............................... 27
   3.2.1 Invariant probability measures ................................................................... 29
   3.2.2 Existence of an invariant probability measure ............................................ 31
   3.2.3 On small and petite sets: two sufficient conditions .................................... 32
   3.2.4 Dobrushin’s Ergodic coefficient for general state spaces ............................ 32
3.3 Further Conditions on the Existence and Uniqueness of Invariant Probability Measures ................................................................................................................. 33
   3.3.1 Markov chains with the Feller property ....................................................... 33
   3.3.2 Quasi-Feller chains ....................................................................................... 34
   3.3.3 Cases without the Feller condition ............................................................... 34
3.4 Ergodicity Properties ............................................................................................ 35
   3.4.1 Uniqueness of an invariant probability measure and ergodicity .................. 35
   3.4.2 Ergodic theorems for positive Harris recurrent chains ................................. 35
   3.4.3 Further ergodic theorems for Markov chains ............................................... 36
3.5 Exercises .............................................................................................................. 37

4 Martingales and Foster-Lyapunov Criteria for Stabilization of Markov Chains .......................... 39
4.1 Martingales .......................................................................................................... 39
   4.1.1 More on expectations and conditional probability ....................................... 39
   4.1.2 Some properties of conditional expectation ................................................ 40
   4.1.3 Discrete-time martingales ............................................................................. 41
   4.1.4 Doob’s optional sampling theorem ............................................................. 42
   4.1.5 An important martingale convergence theorem ......................................... 43
   4.1.6 Proof of Birkhoff’s individual ergodic theorem .......................................... 45
   4.1.7 This section is optional: Further martingale theorems ................................. 45
   4.1.8 Azuma-Hoeffding inequality for martingales with bounded increments .......... 46
4.2 Stability of Markov Chains: Foster-Lyapunov Techniques .................................. 46
   4.2.1 Criterion for positive Harris recurrence ...................................................... 46
   4.2.2 Criterion for finite expectations .................................................................. 48
## 4 Convergence to Equilibrium

4.2.3 Criterion for recurrence
4.2.4 Criterion for transience
4.2.5 State dependent drift criteria: Deterministic and random-time

## 4.3 Convergence Rates to Equilibrium

4.3.1 Rates of convergence: Geometric ergodicity
4.3.2 Subgeometric ergodicity

## 4.4 Conclusion

## 4.5 Exercises

## 5 Dynamic Programming

5.1 Dynamic Programming, Optimality of Markov Policies and Bellman’s Principle of Optimality

5.1.1 Optimality of Deterministic Markov Policies
5.1.2 Bellman’s principle of optimality
5.1.3 Examples

5.2 Existence of Minimizing Selectors and Measurability

5.3 The Linear Quadratic Gaussian (LOG) Problem

5.4 Optional: A Strategic Measures Approach

5.5 Infinite Horizon Optimal Discounted Cost Control Problems

5.5.1 Discounted cost problems with unbounded per-stage costs

5.6 Concluding Remarks

5.7 Exercises

## 6 Partially Observed Markov Decision Processes, Non-linear Filtering and the Kalman Filter

6.1 Enlargement of the State-Space and the Construction of a Controlled Markov Chain

6.2 The Linear Quadratic Gaussian (LOG) Problem and Kalman Filtering

6.2.1 A supporting result on estimation
6.2.2 The Linear Quadratic Gaussian Problem
6.2.3 Estimation and Kalman Filtering
6.2.4 Optimal Control of Partially Observed LOG Systems

6.3 On the Controlled Markov Construction in the Space of Probability Measures and Extension to General Spaces

6.3.1 Measurability Issues, Proof of Theorem 6.1.1 and its Extension to Polish Spaces
6.3.2 Continuity Properties of Belief-MDP
6.3.3 Concavity of the Value Function in the Priors

6.4 Bibliographic Notes

6.5 Exercises
# Contents

## 7 The Average Cost Problem

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Average Cost and the Average Cost Optimality Equation (ACOE) or Inequality (ACOI)</td>
<td>99</td>
</tr>
<tr>
<td>7.1.1</td>
<td>The Vanishing Discounted Cost Approach to the Average Cost Problem</td>
<td>101</td>
</tr>
<tr>
<td>7.1.2</td>
<td>Polish State and Action Spaces, ACOE and ACOI</td>
<td>102</td>
</tr>
<tr>
<td>7.2</td>
<td>The Convex Analytic Approach to Average Cost Markov Decision Problems</td>
<td>107</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Finite State/Action Setup</td>
<td>107</td>
</tr>
<tr>
<td>7.2.2</td>
<td>General State/Action Spaces</td>
<td>111</td>
</tr>
<tr>
<td>7.2.3</td>
<td>Extreme Points and the Optimality of Deterministic Policies</td>
<td>113</td>
</tr>
<tr>
<td>7.2.4</td>
<td>Sample-Path Optimality</td>
<td>113</td>
</tr>
<tr>
<td>7.3</td>
<td>Constrained Markov Decision Processes</td>
<td>114</td>
</tr>
<tr>
<td>7.4</td>
<td>The Linear Programming Approach to Discounted Cost Problems and Control until a Stopping Time</td>
<td>115</td>
</tr>
<tr>
<td>7.5</td>
<td>Bibliographic Notes</td>
<td>115</td>
</tr>
<tr>
<td>7.6</td>
<td>Exercises</td>
<td>115</td>
</tr>
</tbody>
</table>

## 8 Learning and Approximation Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Value and Policy Iteration Algorithms</td>
<td>119</td>
</tr>
<tr>
<td>8.1.1</td>
<td>Value Iteration</td>
<td>119</td>
</tr>
<tr>
<td>8.1.2</td>
<td>Policy Iteration</td>
<td>119</td>
</tr>
<tr>
<td>8.2</td>
<td>Stochastic Learning Algorithms</td>
<td>120</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Q-Learning</td>
<td>121</td>
</tr>
<tr>
<td>8.3</td>
<td>Approximation through Quantization of the State and the Action Spaces</td>
<td>122</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Finite Action Approximation to MDPs</td>
<td>122</td>
</tr>
<tr>
<td>8.3.2</td>
<td>Finite State Approximation to MDPs</td>
<td>126</td>
</tr>
<tr>
<td>8.4</td>
<td>Bibliographic Notes</td>
<td>130</td>
</tr>
<tr>
<td>8.5</td>
<td>Exercises</td>
<td>130</td>
</tr>
</tbody>
</table>

## 9 Decentralized Stochastic Control

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1</td>
<td>Witsenhausen’s Intrinsic Model</td>
<td>133</td>
</tr>
<tr>
<td>9.2</td>
<td>Classification of information structures</td>
<td>134</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Classical, quasiclassical and nonclassical information structures</td>
<td>134</td>
</tr>
<tr>
<td>9.3</td>
<td>Solutions to Static Teams</td>
<td>137</td>
</tr>
<tr>
<td>9.4</td>
<td>Dynamic Teams Quasi-Classical Information Patterns</td>
<td>141</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Non-classical information structures, signaling and its effect on lack of convexity</td>
<td>141</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Witsenhausen’s Counterexample and its variations</td>
<td>142</td>
</tr>
<tr>
<td>9.5</td>
<td>Static Reduction</td>
<td>142</td>
</tr>
</tbody>
</table>
1

Review of Probability

1.1 Introduction

Before discussing controlled Markov chains, we first discuss some preliminaries about probability theory.

Many events in the physical world are uncertain; that is, with a given knowledge up to a certain time (such as an initial condition) regarding a process, the future values of the process are not accurately predictable. Probability theory attempts to develop an understanding for such uncertainty in a consistent way given a number of properties to be satisfied.

Examples of stochastic processes include: a) The temperature in a city at noon in October 2014: This process takes values in \( \mathbb{R}^{31} \), b) The sequence of outputs of a communication channel modeled by an additive scalar Gaussian noise when the input is given by \( x = \{x_1, \ldots, x_n\} \in \mathbb{R}^n \) (the output process lives in \( \mathbb{R}^n \)), c) Infinite copies of a coin flip process (living in \( \{H, T\}^{\mathbb{Z}_+} \)), d) The trajectory of a plane flying from point A to point B (taking values in \( C([t_0, \infty)) \), the space of all continuous paths in \( \mathbb{R}^3 \) with \( x_{t_0} = A, x_{t_f} = B \) for some \( t_0 < t_f \in \mathbb{R} \), e) The exchange rate between the Canadian dollar and the American dollar in a given time index \( T \).

Some of these processes take values in countable spaces, some do not. If the state space \( X \) in which a random variable takes values is finite or countable, it suffices to associate with each point \( x \in X \) a number which determines the likelihood of the event that \( x \) is the value of the process. However, when \( X \) is uncountable, further technical intricacies arise; here the notion of an event needs to be carefully defined. First, if some event \( A \) takes place, it must be that the complement of \( A \) (that is, this event not happening) must also be defined. Furthermore, if \( A \) and \( B \) are two events, then the intersection must also be an event. This line of thought will motivate us for a more formal analysis below. In particular, one needs to construct probability values by first defining values for certain events and extending such probabilities to a larger class of events in a consistent fashion (in particular, one does not first associate probability values to single points as we do in countable state spaces). These issues are best addressed with a precise characterization of probability and random variables.

Hence, probability theory can be used to model uncertainty in the real world in a consistent way according some properties that we expect such measures should admit. In the following, we will develop a rigorous definition for probability. For a more complete exposition the reader could consult with the standard texts on probability theory, such as [49], [94], [8] and [77], and texts on stochastic processes, such as [130].

1.2 Measurable Space

Let \( X \) be a collection of points. Let \( \mathcal{F} \) be a collection of subsets of \( X \) with the following properties such that \( \mathcal{F} \) is a \( \sigma \)-field, that is:

- \( X \in \mathcal{F} \)
- If \( A \in \mathcal{F} \), then \( X \setminus A \in \mathcal{F} \)
If \( A_k \in \mathcal{F}, k = 1, 2, 3, \ldots, \) then \( \bigcup_{k=1}^{\infty} A_k \in \mathcal{F} \) (that is, the collection is closed under countably many unions).

By De Morgan’s laws, the set has to be closed under countable intersections as well.

For example, the full power-set of any set is a \( \sigma \)-field.

If the third item above holds for only finitely many unions or intersections, then, the collection of subsets is said to be a field or algebra.

With the above, \((X, \mathcal{F})\) is termed a measurable space (that is we can associate a measure to this space; which we will discuss shortly).

Remark 1.1. Subsets in \( \sigma \)-fields can be interpreted to represent information that a controlled has with regard to an underlying process. We will discuss this interpretation further and this will be a recurring theme in our discussions in the context of stochastic control problems.

A \( \sigma \)-field \( \mathcal{J} \) is generated by a collection of sets \( \mathcal{A} \), if \( \mathcal{J} \) is the smallest \( \sigma \)-field containing the sets in \( \mathcal{A} \), and in this case, we write \( \mathcal{J} = \sigma(\mathcal{A}) \). We consider an important special case in the following.

### 1.2.1 Borel \( \sigma \)-field

An important class of \( \sigma \)-fields is the Borel \( \sigma \)-field on a metric (or more generally topological) space. Such a \( \sigma \)-field is the one which is generated by open sets. The term open naturally depends on the space being considered. For this course, we will mainly consider spaces which are complete, separable and metric spaces (such as the space of real numbers \( \mathbb{R} \), or countable sets) see Appendix A. Recall that in a metric space with metric \( d \), a set \( U \) is open if for every \( x \in U \), there exists some \( \epsilon > 0 \) such that \( \{ y : d(x, y) < \epsilon \} \subset U \). We note also that the empty set is a special open set.

Thus, the Borel \( \sigma \)-field on \( \mathbb{R} \) is the one generated by sets of the form \((a, b)\), that is, open intervals (it is useful to note here that every open set in \( \mathbb{R} \) can be expressed a union of countably many open intervals). It is important to note that not all subsets of \( \mathbb{R} \) are Borel sets, that is, elements of the Borel \( \sigma \)-field.

We will denote the Borel \( \sigma \)-field on a space \( X \) as \( B(X) \).

We can also define a Borel \( \sigma \)-field on a product space. Let \( X \) be a complete, separable, metric space (with some metric \( d \)). Let \( X^{\mathbb{Z}^+} \) denote the infinite product of \( X \) so that with \( x = (x_0, x_1, x_2, \ldots) \), where \( x_k \in X \) for \( k \in \mathbb{Z}^+ \). If this space is endowed with the product metric (such a metric is defined as: \( \rho(x, y) = \sum_{i=0}^{\infty} 2^{-i} \frac{d(x_i, y_i)}{1+d(x_i, y_i)} \)), sets of the form \( \prod_{i \in \mathbb{Z}^+} A_i \), where only finitely many of these sets are not equal to \( X \), are open, and unions of such sets form open sets. We define cylinder sets in this product space as:

\[
B_{\{A_m, m \in I\}} = \{ x \in X^{\mathbb{Z}^+}, x_m \in A_m, A_m \in B(X) \},
\]

where \( I \subset \mathbb{Z} \) with \( |I| < \infty \), that is, the set has finitely many elements. Thus, in the above, if \( x \in B_{\{A_m, m \in I\}} \), then, \( x_m \in A_m \) and the remaining terms are arbitrary. The \( \sigma \)-field generated by such open cylinder sets is the Borel \( \sigma \)-field on the product space. Such a construction is important for stochastic processes (and is the reason why while studying certain properties of stochastic processes one often only considers finite dimensional distributions).

### 1.2.2 Measurable Function

If \((X, B(X))\) and \((Y, B(Y))\) are measurable spaces; we say a mapping from \( h : X \to Y \) is measurable if

\[
h^{-1}(B) = \{ x : h(x) \in B \} \in B(X), \quad \forall B \in B(Y)
\]

**Theorem 1.2.1** To show that a function is Borel measurable, it is sufficient to check the measurability of the inverses of sets that generate the \( \sigma \)-algebra on the image space.
for real valued functions, to check the measurability of the inverse images of open sets, it suffices to check the measurability of the inverse images of open sets. Furthermore, for Borel measurability, it suffices to check the measurability of the inverse images of open sets. Therefore, for Borel measurability, it suffices to check the measurability of the inverse images of open sets. Furthermore, for real valued functions, to check the measurability of the inverse images of open sets, it suffices to check the measurability of the inverse images of open sets.

Thus, the set of all subsets whose inverses are Borel:

\[ \mathcal{M} = \{ B \subset \mathcal{Y} : h^{-1}(B) \in \mathcal{B}(\mathcal{X}) \} \]

is a \( \sigma \)-algebra over \( \mathcal{Y} \) and this set contains the open sets. Thus, \( \mathcal{B}(\mathcal{Y}) \subset \mathcal{M} \).

1.2.3 Measure

A positive measure \( \mu \) on \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\) is a map from \( \mathcal{B}(\mathcal{X}) \) to \([0, \infty]\) which is countably additive such that for \( A_k \in \mathcal{B}(\mathcal{X}) \) and \( A_k \cap A_j = \emptyset \):

\[ \mu\left( \bigcup_{k=1}^{\infty} A_k \right) = \sum_{k=1}^{\infty} \mu(A_k). \]

Definition 1.2.1 \( \mu \) is a probability measure if it is positive and \( \mu(\mathcal{X}) = 1 \).

Definition 1.2.2 A measure \( \mu \) is finite if \( \mu(\mathcal{X}) < \infty \), and \( \sigma \)-finite if there exist a collection of subsets such that \( X = \bigcup_{k=1}^{\infty} A_k \) with \( \mu(A_k) < \infty \) for all \( k \).

On the real line \( \mathbb{R} \), the Lebesgue measure is defined on the Borel \( \sigma \)-field (in fact on a somewhat larger field obtained through adding all subsets of Borel sets of measure zero: this is known as completion of a \( \sigma \)-field) such that for \( A = (a, b) \), \( \mu(A) = b - a \). Borel field of subsets is a strict subset of Lebesgue measurable sets, that is there exist Lebesgue measurable sets which are not Borel sets. For a definition and the construction of Lebesgue measurable sets, see [27]. There exists Lebesgue measurable sets of measure zero which contain uncountably many elements, for a well-studied example see the Cantor set [45].

1.2.4 The Extension Theorem

Theorem 1.2.2 (The Extension Theorem) Let \( \mathcal{M} \) be an algebra, and suppose that there exists a measure \( P \) satisfying:

(i) There exists countably many sets \( A_n \in \mathcal{X} \) such that \( \mathcal{X} = \bigcup_n A_n \), with \( P(A_n) < \infty \), (ii) For pairwise disjoint sets \( A_n \), if \( \bigcup_n A_n \in \mathcal{M} \), then \( P(\bigcup_n A_n) = \sum_n P(A_n) \). Then, there exists a unique measure \( P' \) on the \( \sigma \)-field generated by \( \mathcal{M} \), \( \sigma(\mathcal{M}) \), which is consistent with \( P \) on \( \mathcal{M} \).

The above is useful since, when one states that two measures are equal it suffices to check if they are equal on the set of sets which generate the \( \sigma \)-algebra, and not necessarily on the entire \( \sigma \)-field.

The following is a refinement useful for stochastic processes. It in particular does not require a probability measure defined apriori before an extension. [4]:

Theorem 1.2.3 (Kolmogorov’s Extension Theorem) Let \( \mathcal{X} \) be a complete and separable metric space and for all \( n \in \mathbb{N} \) let \( \mu_n \) be a probability measure on \( \mathcal{X}^n \), the \( n \) product of \( \mathcal{X} \), such that
for every \( n \) and every sequence of Borel sets \( A_k \). Then, there exists a unique probability measure \( \mu \) on \( (\mathcal{X}^\mathbb{N}, \mathcal{B}(\mathcal{X}^\mathbb{N})) \) which is consistent with each of the \( \mu_n \)’s.

A further related result, which often in stochastic control is cited in the context of extensions, is the Ionescu Tulcea Extension Theorem [59, Appendix C]: where conditional probability measures (stochastic kernels) are defined (instead of probability measures on finite dimensional product spaces) as a starting assumption, before an extension to the infinite product space is established. Thus, if the \( \sigma \)–field on a product space is generated by the collection of finite dimensional cylinder sets, one can define a measure in the product space which is consistent with the finite dimensional distributions.

Likewise, we can construct the Lebesgue measure on \( \mathcal{B}(\mathbb{R}) \) by defining it on finitely many unions and intersections of intervals of the form \((a, b), [a, b), (a, b] \) and \([a, b]\), and the empty set, thus forming a field, and extending this to the Borel \( \sigma \)–field. Thus, the relation \( \mu(a, b) = b - a \) for \( b > a \) is sufficient to define the Lebesgue measure.

Remark 1.2. A more general result is as follows: Let \( \mathcal{S} \) be a \( \sigma \)–field. A class of subsets \( \mathcal{A} \subset \mathcal{S} \) is called a separating class if two probability measures that agree on \( \mathcal{A} \) agree on the entire \( \mathcal{S} \). A class of subsets is a \( \pi \)–system if it is closed under finite intersections. The class \( \mathcal{A} \) is a separating class if it is both a \( \pi \)–system and it generates the \( \sigma \)–field \( \mathcal{S} \); see [19] or [21].

1.2.5 Integration

Let \( h \) be a non-negative measurable function from \( \mathcal{X}, \mathcal{B}(\mathcal{X}) \) to \( \mathbb{R}, \mathcal{B}(\mathbb{R}) \). The Lebesgue integral of \( h \) with respect to a measure \( \mu \) can be defined in three steps:

First, for \( A \in \mathcal{B}(\mathcal{X}) \), define \( 1\{x \in A\} \) (or \( 1_{\{x \in A\}} \), or \( 1_A(x) \)) as an indicator function for event \( x \in A \), that is the value that the function takes is 1 if \( x \in A \), and 0 otherwise. In this case, define

\[
\int_{\mathcal{X}} 1_{\{x \in A\}}(x) \, d\mu(x) := \mu(A).
\]

Now, let us define simple functions \( h \) such that, there exist \( A_1, A_2, \ldots, A_n \) all in \( \mathcal{B}(\mathcal{X}) \) and positive numbers \( b_1, b_2, \ldots, b_n \) such that \( h_n(x) = \sum_{k=1}^n b_k 1_{\{x \in A_k\}} \). For such functions, define

\[
\int_{\mathcal{X}} h_n(x) \, d\mu(x) := \sum_{k=1}^n b_k \mu(A_k).
\]

Now, for any given measurable \( h \), there exists a sequence of simple functions \( h_n \) such that \( h_n(x) \to h(x) \) monotonically, that is \( h_{n+1}(x) \geq h_n(x) \) (for a construction, if \( h \) only takes non-negative values, consider partitioning the positive real line to two intervals \([0, n) \) and \([n, \infty)\), and partition \([0, n) \) to \( n2^n \) uniform intervals, define \( h_n(x) \) to be the lower floor of the interval that contains \( h(x) \): thus

\[
h_n(x) = k2^{-n}, \quad \text{if} \quad k2^{-n} \leq h(x) < (k+1)2^{-n}, \quad k = 0, 1, \ldots, n2^n - 1,
\]

and \( h_n(x) = n \) for \( h(x) \geq n \). By definition, and since \( h^{-1}(\{k2^{-n}, (k+1)2^{-n}\}) \) is Borel, \( h_n \) is a simple function. If the function takes also negative values, write \( h(x) = h_+(x) - h_-(x) \), where \( h_+ \) is the non-negative part and \(-h_- \) is the negative part, and construct the same for \( h_-(x) \). We define the limit as the Lebesgue integral:

\[
\lim_{n \to \infty} \int h_n(x) \, d\mu(x) = \lim_{n \to \infty} \sum_{k=1}^n b_k \mu(A_k)
\]
There are three important convergence theorems which we will not discuss in detail, the statements of which will be given in class.

### 1.2.6 Fatou’s Lemma, the Monotone Convergence Theorem and the Dominated Convergence Theorem

**Theorem 1.2.4 (Monotone Convergence Theorem)** If \( \mu \) is a \( \sigma \)-finite positive measure on \( (\mathcal{X}, \mathcal{B}(\mathcal{X})) \) and \( \{ f_n, n \in \mathbb{Z}_+ \} \) is a sequence of measurable functions from \( \mathcal{X} \) to \( \mathbb{R} \) which pointwise, monotonically, converge to \( f \): that is, \( 0 \leq f_n(x) \leq f_{n+1}(x) \) for all \( n \), and

\[
\lim_{n \to \infty} f_n(x) = f(x),
\]

for \( \mu \)-almost every \( x \), then

\[
\int_{\mathcal{X}} f(x) \mu(dx) = \lim_{n \to \infty} \int_{\mathcal{X}} f_n(x) \mu(dx)
\]

**Theorem 1.2.5 (Fatou’s Lemma)** If \( \mu \) is a \( \sigma \)-finite positive measure on \( (\mathcal{X}, \mathcal{B}(\mathcal{X})) \) and \( \{ f_n, n \in \mathbb{Z}_+ \} \) is a sequence of measurable functions (bounded from below) from \( \mathcal{X} \) to \( \mathbb{R} \), then

\[
\int_{\mathcal{X}} \liminf_{n \to \infty} f_n(x) \mu(dx) \leq \liminf_{n \to \infty} \int_{\mathcal{X}} f_n(x) \mu(dx)
\]

**Theorem 1.2.6 (Dominated Convergence Theorem)** If (i) \( \mu \) is a \( \sigma \)-finite positive measure on \( (\mathcal{X}, \mathcal{B}(\mathcal{X})) \), (ii) \( g \) is a Borel measurable function with

\[
\int_{\mathcal{X}} g(x) \mu(dx) < \infty,
\]

and (iii) \( \{ f_n, n \in \mathbb{Z}_+ \} \) is a sequence of measurable functions from \( \mathcal{X} \) to \( \mathbb{R} \) which satisfy \( |f_n(x)| \leq g(x) \) for \( \mu \)-almost every \( x \), and \( \lim_{n \to \infty} f_n(x) = f(x) \), then:

\[
\int_{\mathcal{X}} f(x) \mu(dx) = \lim_{n \to \infty} \int_{\mathcal{X}} f_n(x) \mu(dx)
\]

Note that for the monotone convergence theorem, there is no restriction on boundedness; whereas for the dominated convergence theorem, there is a boundedness condition. On the other hand, for the dominated convergence theorem, the pointwise convergence does not have to be monotone.

There also exist generalized versions of these theorems, where the measures themselves are time-varying, but converge to a limit measure in some appropriate sense; see e.g., [74, 111]. These will be discussed later in further detail.

### 1.3 Probability Space and Random Variables

Let \((\Omega, \mathcal{F})\) be a measurable space. If \( P \) is a probability measure, then the triple \((\Omega, \mathcal{F}, P)\) forms a probability space. Here \( \Omega \) is a set called the sample space, \( \mathcal{F} \) is called the event space, and this is a \( \sigma \)-field of subsets of \( \Omega \).

Let \((E, \mathcal{E})\) be another measurable space and \( X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{E}) \) be a measurable map. We call \( X \) an \( E \)-valued random variable. The image under \( X \) is a probability measure on \((E, \mathcal{E})\), called the law of \( X \).

The \( \sigma \)-field generated by the events \( \{ \{ w : X(w) \in A \}, A \in \mathcal{E} \} \) is called the \( \sigma \)-field generated by \( X \) and is denoted by \( \sigma(X) \).

Consider a coin flip process, with possible outcomes \( \{H, T\} \), heads or tails. We have a good intuitive understanding on the environment when someone tells us that, a coin flip takes the value \( H \) with probability \( 1/2 \). Based on the definition of a random variable, we view then a coin flip process as a deterministic function from some space \((\Omega, \mathcal{F}, P)\) to the binary space of a head and a tail event. Here, \( P \) denotes the uncertainty measure (you may think of the initial condition of the
coin when it is being flipped, the flow of air, the surface where the coin touches etc.; we encode all these aspects and all the uncertainty in the universe with the abstract space \((\Omega, \mathcal{F}, P)\). You can view then the \(\sigma\)-field generated by such a coin flip as a partition of \(\Omega\): if certain things take place the outcome is a \(H\) and otherwise it is a \(T\).

A useful fact about measurable functions (and thus random variables) is the following result.

**Theorem 1.3.1** Let \(f_n\) be a sequence of measurable functions from \((\Omega, \mathcal{F})\) to a complete separable metric space \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\). Then, \(\limsup_{n \to \infty} \{f_n(x)\}, \liminf_{n \to \infty} \{f_n(x)\}\) are measurable. Thus, if \(f(x) = \lim_{n \to \infty} f_n(x)\) exists, then \(f\) is measurable.

This theorem implies that to verify whether a real valued mapping \(f\) is a Borel measurable function, it suffices to check if \(f^{-1}(a, b) \in \mathcal{F}\) for \(a < b\) since one can construct a sequence of simple functions which will converge to any measurable \(f\), as discussed earlier. It suffices then to check if \(f^{-1}(-\infty, a) \in \mathcal{F}\) for \(a \in \mathbb{R}\).

### 1.3.1 More on Random Variables and Probability Density Functions

Consider a probability space \((\mathcal{X}, \mathcal{B}(\mathcal{X}), P)\) and consider an \(\mathbb{R}\)-valued random variable \(U\) measurable with respect to \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\).

This random variable induces a probability measure \(\mu\) on \(\mathcal{B}(\mathbb{R})\) such that for some \((a, b) \in \mathcal{B}(\mathbb{R})\):

\[
\mu((a, b)) = P(U \in (a, b]) = P(\{x : U(x) \in (a, b]\})
\]

When \(U\) is \(\mathbb{R}\)-valued, the *Expectation of \(U\)* is defined as:

\[
E[U] = \int_{\mathbb{R}} \mu(dx)x
\]

We define \(F(x) = \mu(-\infty, x]\) as the cumulative distribution function of \(U\). If the derivative of \(F(x)\) exists, we call this derivative the *probability density function* of \(U\), and denote it by \(p(x)\) (the distribution function is not always differentiable, for example when the random variable takes values only on integers). If a density function exists, we can also write:

\[
E[U] = \int_{\mathbb{R}} p(x)x\,dx
\]

If a probability density function \(p\) exists, the measure \(P\) is said to be *absolutely continuous with respect to the Lebesgue measure*. In particular, the density function \(p\) is the Radon-Nikodym derivative of \(P\) with respect to the Lebesgue measure in the sense that for all Borel \(A: \int_A p(x)\lambda(dx) = P(A)\).

A probability density does not always exist. In particular, whenever there is a probability *mass* on a given point, then a density does not exist; hence in \(\mathbb{R}\), if for some \(x\), \(P(\{x\}) > 0\), then we say there is a probability mass at \(x\), and a density function does not exist.

If \(\mathcal{X}\) is countable, we can write \(P(\{x = m\}) = p(m)\), where \(p\) is called the *probability mass function*.

Some examples of commonly encountered random variables are as follows:

- **Gaussian \(\mathcal{N}(\mu, \sigma^2)\):**

  \[
p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.
\]

- **Exponential:**

  \[
  F(x) = 1 - e^{-\lambda x}, \quad p(x) = \lambda e^{-\lambda x}.
  \]

- **Uniform on \([a, b]\) \((U([a, b]))\):**
1.3.2 Independence and Conditional Probability

Consider \( A, B \in \mathcal{B}(X) \) such that \( P(B) > 0 \). The quantity

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}
\]

is called the conditional probability of event \( A \) given \( B \). The measure \( P(\cdot | B) \) defined on \( \mathcal{B}(X) \) is itself a probability measure. If

\[
P(A|B) = P(A)
\]

Then, \( A, B \) are said to be independent events.

A countable collection of events \( \{A_n\} \) is independent if for any finitely many sub-collections \( A_{i_1}, A_{i_2}, \ldots, A_{i_m} \), we have that

\[
P(A_{i_1}, A_{i_2}, \ldots, A_{i_m}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_m}).
\]

Here, we use the notation \( P(A, B) = P(A \cap B) \). A sequence of events is said to be pairwise independent if for any two pairs \( (A_m, A_n) \): \( P(A_m, A_n) = P(A_m)P(A_n) \). Pairwise independence is weaker than independence.

Conditional probability and expectation will be discussed in more detail later in Chapter 4.

1.4 Stochastic Processes and Markov Chains

One can define a sequence of random variables as a single random variable living in the product set, that is we can consider \( \{x_1, x_2, \ldots, x_N, \ldots\} \) as an individual random variable \( X \) which is an \( \mathbb{X}^\mathbb{Z}_+ \)-valued random variable, where now the fields are to be defined on the product space.

Let \( \mathbb{X} \) be a complete, separable, metric space and let \( T = \mathbb{Z} \) or \( T = \mathbb{Z}_+ \). Let \( \mathcal{B}(\mathbb{X}) \) denote the Borel sigma-field over \( \mathbb{X} \). Let \( \Sigma = \mathbb{X}^T \) denote the sequence space of all one-sided (with \( T = \mathbb{Z}_+ \)) or two-sided (with \( T = \mathbb{Z} \)) infinitely many random variables drawn from \( \mathbb{X} \). Thus, if \( T = \mathbb{Z} \), \( x \in \Sigma \) then \( x = \{ \ldots, x_{-1}, x_0, x_1, \ldots \} \) with \( x_i \in \mathbb{X} \). Let \( X_n : \Sigma \to \mathbb{X} \) denote the coordinate function such that \( X_n(x) = x_n \). Let \( \mathcal{B}(\Sigma) \) denote the smallest sigma-field containing all cylinder sets of the form \( \{ x : x_i \in B_i, m \leq i \leq n \} \) where \( B_i \in \mathcal{B}(\mathbb{X}) \), for all integers \( m, n \). We can define a probability measure by a characterization of these finite dimensional cylinder sets, by (the extension) Theorem II.2.5.

A similar characterization also applies for continuous-time stochastic processes, where \( T \) is uncountable. The extension requires more delicate arguments, since finite-dimensional characterizations are too weak to uniquely define a sigma-field on a space of continuous-time paths which is consistent with such distributions. Such technicalities arise in the discussion for continuous-time Markov chains and controlled processes, typically requiring a construction from a separable product space. In this course, our focus will be on discrete-time processes; however, the analysis for continuous-time processes essentially follows from similar constructions with further structures that one needs to impose on continuous-time processes. More discussion on this topic is available in Chapter II.10.
1.4.1 Markov Chains

If the probability measure on an $\mathbb{X}^{\mathbb{Z}_+}$-valued sequence is such that

$$P(x_{k+1} \in A_{k+1} | x_k, x_{k-1}, \ldots, x_0) = P(x_{k+1} \in A_{k+1} | x_k) \quad P.a.s.,$$

then $\{x_k\}$ is said to be a Markov chain. Thus, for a Markov chain, the immediate state is sufficient to predict the future; past variables are not needed.

One way to construct a Markov chain is via the following: Let $\{x_t, t \geq 0\}$ be a random sequence with state space $(\mathbb{X}, B(\mathbb{X}))$, and defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $B(\mathbb{X})$ denotes the Borel $\sigma$-field on $\mathbb{X}$, $\Omega$ is the sample space, $\mathcal{F}$ a sigma field of subsets of $\Omega$, and $\mathbb{P}$ a probability measure. For $x \in \mathbb{X}$ and $D \in B(\mathbb{X})$, we let $P(x, D) := P(x_{t+1} \in D | x_t = x)$ denote the transition probability from $x$ to $D$, that is the probability of the event $\{x_{t+1} \in D\}$ given that $x_t = x$. Thus, the Markov chain is completely determined by the transition probability and the probability of the initial state, $P(x_0) = p_0$. Hence, the probability of the event $\{x_{t+1} \in D\}$ for any $t$ can be computed recursively by starting at $t = 0$, with $P(x_1 \in D) = \sum_x P(x_1 \in D | x_0 = x)P(x_0 = x)$, and iterating with a similar formula for $t = 1, 2, \ldots$.

Hence, if the probability of the same event given some history of the past (and the present) does not depend on the past, and hence is given by the same quantity, the chain is a Markov chain. As an example, consider the following linear system:

$$x_{t+1} = ax_t + w_t,$$

where $w_t$ is an independent random variable. This sequence is Markov.

We will continue our discussion on Markov chains after discussing controlled Markov chains.

1.5 Exercises

Exercise 1.5.1 a) Let $\mathcal{F}_\beta$ be a $\sigma$-field of subsets of some space $\mathbb{X}$, for all $\beta \in \Gamma$ where $\Gamma$ is a family of indices. Let

$$\mathcal{F} = \bigcap_{\beta \in \Gamma} \mathcal{F}_\beta$$

Show that $\mathcal{F}$ is also a $\sigma$-field.

For a space $\mathbb{X}$, on which a distance metric is defined, the Borel $\sigma$-field is generated by the collection of open sets. This means that, the Borel $\sigma$-field is the smallest $\sigma$-field containing open sets, and as such it is the intersection of all $\sigma$-fields containing open sets. On $\mathbb{R}$, the Borel $\sigma$-field is the smallest $\sigma$-field containing open intervals.

b) Is the set of rational numbers an element of the Borel $\sigma$-field on $\mathbb{R}$? Is the set of irrational numbers an element?

c) Let $\mathbb{X}$ be a countable set. On this set, let us define a metric as follows:

$$d(x, y) = \begin{cases} 
0, & \text{if } x = y \\
1, & \text{if } x \neq y 
\end{cases}$$

Show that, the Borel $\sigma$-field on $\mathbb{X}$ is generated by the individual sets $\{x\}, x \in \mathbb{X}$.

d) Consider the distance function $d(x, y)$ defined as above defined on $\mathbb{R}$. Is the $\sigma$-field generated by open sets according to this metric the same as the Borel $\sigma$-field on $\mathbb{R}$?

Exercise 1.5.2 Investigate the following limits in view of the convergence theorems.

a) Check if $\lim_{n \to \infty} \int_0^1 x^n dx = \int_0^1 \lim_{n \to \infty} x^n dx$. 
b) Check if \( \lim_{n \to \infty} \int_0^1 nx^n \, dx = \int_0^1 \lim_{n \to \infty} nx^n \, dx \).

c) Define \( f_n(x) = n1_{\{0 \leq x \leq \frac{1}{n}\}} \). Find \( \lim_{n \to \infty} \int f_n(x) \, dx \) and \( \lim_{n \to \infty} \int f_n(x) \, dx \). Are these equal?

**Exercise 1.5.3**

a) Let \( X \) and \( Y \) be real-valued random variables defined on a given probability space. Show that \( X^2 \) and \( X + Y \) are also random variables.

b) Let \( \mathcal{F} \) be a \( \sigma \)-field of subsets over a set \( \mathcal{X} \) and let \( A \in \mathcal{F} \). Prove that \( \{ A \cap B, B \in \mathcal{F} \} \) is a \( \sigma \)-field over \( A \) (that is a \( \sigma \)-field of subsets of \( A \)).

Hint for part a: The following equivalence holds: \( X + Y < x \equiv \bigcup \{ X < r, Y < x - r \} \). To check if \( X + Y \) is a random variable, it suffices to check if the event \( \{ X + Y < x \} = \{ \omega : X(\omega) + Y(\omega) < x \} \) is an element of \( \mathcal{F} \) for every \( x \in \mathbb{R} \).

**Exercise 1.5.4**

Let \( f_n \) be a sequence of measurable functions from \( (\Omega, \mathcal{F}) \) to \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). Show \( f(\omega) = \limsup_{n \to \infty} f_n(\omega) \) and \( g(\omega) = \liminf_{n \to \infty} f_n(\omega) \) define measurable functions.

**Exercise 1.5.5**

Let \( X \) and \( Y \) be real-valued random variables defined on a given probability space \( \Omega, \mathcal{F}, P \). Suppose that \( X \) is measurable on \( \sigma(Y) \). Show that there exists a function \( f \) such that \( X = f(Y) \).

This result also holds if \( \mathcal{X} \) is a Borel subset of a complete, separable and metric space. Such spaces are called standard Borel.

**Exercise 1.5.6**

Consider a random variable \( X \) which takes values in \( [0, 1] \) and is uniformly distributed. We know that for every set \( S = [a, b) \) \( 0 \leq a \leq b \leq 1 \), the probability of a random variable \( X \) taking values in \( S \) is \( P(X \in [a, b)) = U([a, b]) = b - a \). Consider now the following question: Does every subset \( S \subset [0, 1] \) admit a probability in the form above? In the following we will provide a counterexample, known as the Vitali set.

We can show that a set picked as follows does not admit a measure: Let us define an equivalence class such that \( x \equiv y \) if \( x - y \in \mathbb{Q} \). This equivalence definition divides \( [0, 1] \) into disjoint sets. There are countably many points in each equivalence class. Let \( A \) be a subset which picks exactly one element from each equivalent class (here, we adopt what is known as the Axiom of Choice \([20]\)). Since \( A \) contains an element from each equivalence class, each point of \( [0, 1] \) is contained in the union \( \bigcup_{q \in \mathbb{Q}} (A \oplus q) \). For otherwise there would be a point \( x \) which were not in any equivalence class. But \( x \) is equivalent with itself at least. Furthermore, since \( A \) contains only one point from each equivalence class, the sets \( A \oplus q \) are disjoint; for otherwise there would be two sets which could include a common point: \( A \oplus q \) and \( A \oplus q' \) would include a common point, leading to the result that the difference \( x - q = z \) and \( x - q' = z \) are both in \( A \), a contradiction, since there should be at most one point which is in the same equivalence class as \( x - q = z \). One can show that the uniform distribution is shift-invariant, therefore \( P(A) = P(A \oplus q) \). But \( [0, 1] = \bigcup_{q \in \mathbb{Q}} A \oplus q \). Since a countable sum of identical non-negative elements can either become \( \infty \), or 0, the contradiction follows: We can’t associate a number with this set and as a result, this set is not a Lebesgue measurable set.
In the following, we discuss controlled Markov models.

### 2.1 Controlled Markov Models

Consider the following model.

\[ x_{t+1} = f(x_t, u_t, w_t), \]  

(2.1)

where \( x_t \) is an \( \mathbb{X} \)-valued state variable, \( u_t \) is a \( \mathbb{U} \)-valued control action variable, \( w_t \) a \( \mathbb{W} \)-valued an i.i.d noise process, and \( f \) is a measurable function. We assume that \( \mathbb{X}, \mathbb{U}, \mathbb{W} \) are Borel subsets of complete, separable, metric (such spaces are called Polish) spaces; such spaces are also called standard Borel. We assume that all the random variables live in a probability space \((\Omega, \mathcal{F}, P)\).

Using stochastic realization results (see Lemma 1.2 in [54], or Lemma 3.1 of [26]), it can be shown that the model above in (2.1) contains the class of all \((\mathbb{X} \times \mathbb{U})^\mathbb{Z}_+\)-valued stochastic processes which satisfy the following for all Borel sets \( B \in \mathcal{B}(\mathbb{X}) \), \( t \geq 0 \), and \( (P \text{ almost all}) \) realizations \( x_{[0,t]}, u_{[0,t]} \):

\[
P(x_{t+1} \in B | x_{[0,t]} = a_{[0,t]}, u_{[0,t]} = b_{[0,t]}) = P(x_{t+1} \in B | x_t = a_t, u_t = b_t) =: T(x_{t+1} \in B | a_t, b_t) \]  

(2.2)

where \( T(\cdot|a, u) \) is a stochastic kernel from \( \mathbb{X} \times \mathbb{U} \) to \( \mathbb{X} \). A stochastic process which satisfies (2.2) is called a controlled Markov chain.

#### 2.1.1 Fully Observed Markov Control Problem Model

A Fully Observed Markov Control Problem is a five tuple

\( (\mathbb{X}, \mathbb{U}, \{\mathbb{U}(x), x \in \mathbb{X}\}, T, c) \),

where

- \( \mathbb{X} \) is the state space, a subset of a Polish space.
- \( \mathbb{U} \) is the action space, a subset of a Polish space.
- \( \mathbb{K} = \{(x, u) : u \in \mathbb{U}(x) \in \mathcal{B}(\mathbb{U}), x \in \mathbb{X}\} \) is the set of state, control pairs that are feasible. There might be different states where different control actions are possible.
- \( T \) is a state transition kernel, that is \( T(A|x, u) = P(x_{t+1} \in A | x_t = x, u_t = u) \).
- \( c : \mathbb{K} \to \mathbb{R} \) is the cost.
2.1.2 Classes of Control Policies

Admissible Control Policies

Let \( H_0 := \mathbb{X}, \quad H_t = H_{t-1} \times \mathbb{K} \) for \( t = 1, 2, \ldots \). We let \( h_t \) denote an element of \( H_t \), where \( h_t = \{x_{[0,t]}, u_{[0,t-1]}\} \). A deterministic admissible control policy \( \Pi \) is a sequence of functions \( \{\gamma_t\} \) from \( H_t \to U \); in this case \( u_t = \gamma_t(h_t) \).

A randomized admissible control policy is a sequence \( \Pi = \{\Pi_t, t \geq 0\} \) such that \( \Pi : H_t \to \mathcal{P}(U) \) (with \( \mathcal{P}(U) \) being the space of probability measures on \( U \)) such that
\[
\Pi_t(u_t \in C|h_t) = 1, \quad \forall h_t \in H_t.
\]

Markov Control Policies

A deterministic Markov control policy \( \Pi \) is a sequence of functions \( \{\gamma_t\} \) from \( \mathbb{X} \times \mathbb{Z}_+ \to U \) such that \( u_t = \gamma_t(x_t) \) for each \( t \in \mathbb{Z}_+ \).

A policy is randomized Markov if
\[
P^\Pi(u_t \in C|h_t) = \Pi_t(u_t \in C|x_t), \quad C \in \mathcal{B}(U),
\]
for all \( t \). Hence, the control action only depends on the state and the time, and not the past history.

Stationary Control Policies

A deterministic stationary control policy \( \Pi \) is a sequence of identical functions \( \{\gamma, \gamma, \gamma, \ldots\} \) from \( \mathbb{X} \to U \) such that \( u_t = \gamma(x_t) \) for each \( t \in \mathbb{Z}_+ \).

A policy is randomized stationary if
\[
P^\Pi(u_t \in C|h_t) = \Pi(u_t \in C|x_t), \quad C \in \mathcal{B}(U),
\]
for some stochastic kernel \( \Pi \).

Hence, the control selection is independent of the past history or time, given the current state \( x_t \).

Consider for now that the objective to be minimized is given by:
\[
J(\nu_0, \Pi) := E^\Pi_{\nu_0} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right],
\]
where \( \nu_0 \) is the initial probability measure, that is \( x_0 \sim \nu_0 \). The goal is to find a policy \( \Pi^* \) so that
\[
J(x_0, \Pi^*) \leq J(x_0, \Pi) \quad \forall \Pi \in \Pi_A.
\]
Such a \( \Pi^* \) is called an optimal policy. Here \( \Pi \) can also be called a strategy, or a law.

2.2 Performance of Policies

For stochastic control problems, according to the Ionescu - Tulcea theorem (or Kolmogorov’s extension theorem), an initial distribution \( \mu \) on \( \mathbb{X} \) and a policy \( \Pi \) define a unique probability measure \( P^\Pi_{\nu_0} \) on \( (\mathbb{X} \times U)^{\mathbb{Z}_+} \), which is called a strategic measure.

Consider a Markov control problem with an objective given as the minimization of
\[
J(\nu_0, \Pi) = E^\Pi_{\nu_0} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right].
\]
2.2 Performance of Policies

where $\nu_0$ denotes the distribution on $x_0$. For the case with $x_0 = x$ or $\nu_0 = \delta_x$, we then simply write

$$J(\delta_x, \Pi) = J(x, \Pi) = E^\Pi_\nu_0 \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right] = E^\Pi_\nu_0 \left[ \sum_{t=0}^{T-1} c(x_t, u_t) | x_0 = x \right]$$

Such a cost problem is known as a **Finite Horizon Optimal Control problem**.

We will also consider costs of the following form:

$$J_\beta(\nu_0, \Pi) = E^\Pi_\nu_0 \left[ \sum_{t=0}^{\infty} \beta^t c(x_t, u_t) \right],$$

for some $\beta \in (0, 1)$. This is called a **Discounted Optimal Control Problem**.

Finally, we will study costs of the following form:

$$J_\infty(\nu_0, \Pi) = \limsup_{T \to \infty} \frac{1}{T} E^\Pi_\nu_0 \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right]$$

Such a problem is known as the **Average Cost Optimal Control Problem**.

Let $\Pi_A$ denote the class of admissible policies, $\Pi_M$ denote the class of Markov policies, $\Pi_S$ denote the class of Stationary policies. These policies can be both randomized or deterministic. We will denote the randomizes ones with $\Pi_{RA}$, $\Pi_{RM}$ and $\Pi_{RS}$.

In a general setting, we have the following:

$$\inf_{\Pi \in \Pi_A} J(\nu_0, \Pi) \leq \inf_{\Pi \in \Pi_M} J(\nu_0, \Pi) \leq \inf_{\Pi \in \Pi_S} J(\nu_0, \Pi),$$

since the set of policies is progressively shrinking

$$\Pi_S \subset \Pi_M \subset \Pi_A.$$

We will show, however, that for the optimal control of a Markov chain, under mild conditions, Markov policies are always optimal (that is there is no loss in optimality in restricting the policies to be Markov); that is, it is sufficient to consider only Markov policies. This is an important result in stochastic control. That is,

$$\inf_{\Pi \in \Pi_A} J(\nu_0, \Pi) = \inf_{\Pi \in \Pi_M} J(\nu_0, \Pi)$$

We will also show that, under somewhat more restrictive conditions, stationary policies are optimal (that is, there is no loss in optimality in restricting the policies to be stationary). This will typically require $T \to \infty$ for discounted cost problems:

$$\inf_{\Pi \in \Pi_A} J_\beta(\nu_0, \Pi) = \inf_{\Pi \in \Pi_S} J_\beta(\nu_0, \Pi)$$

and

$$\inf_{\Pi \in \Pi_A} J_\infty(\nu_0, \Pi) = \inf_{\Pi \in \Pi_{RS}} J_\beta(\nu_0, \Pi)$$

Furthermore, we will show that, under some conditions,

$$\inf_{\Pi \in \Pi_S} J_\infty(\nu_0, \Pi)$$

is independent of the initial distribution (or initial condition) on $x_0$.

For further relations between such policies, see Chapter 5 and Chapter 7.
The last two results are computationally very important, as there are powerful computational algorithms that allow one to develop such stationary policies.

In the following set of notes, we will first consider further properties of Markov chains, since under a Markov control policy, the controlled state becomes a Markov chain. We will then get back to controlled Markov chains and the development of optimal control policies in Chapter 5.

The classification of Markov Chains in the next chapter will implicitly characterize the set of problems for which stationary policies contain optimal admissible policies.

### 2.3 Markov Chain Induced by a Markov Policy

The following is an important result:

**Theorem 2.3.1** Let the control policy be randomized Markov. Then, the controlled Markov chain induces an $\mathcal{X}$-valued Markov chain, that is, the state process itself becomes a Markov chain:

$$P_{x_0}^H(x_{t+1} \in B|x_t, x_{t-1}, \ldots, x_0) = Q_t^H(x_{t+1} \in B|x_t), \quad B \in \mathcal{B}(\mathcal{X}), \; t \geq 1, \; P.a.s.$$  

If the control policy is a stationary policy, the induced Markov chain is time-homogenous, that is the transition kernel for the Markov chain does not depend on time.

**Proof.** Let us consider the case where $\mathcal{U}$ is countable, the uncountable case follows similarly. Let $B \in \mathcal{B}(\mathcal{X})$. It follows that,

$$P_{x_0}^H(x_{t+1} \in B|x_t, x_{t-1}, \ldots, x_0) = \sum_{u_t} P_{x_0}^H(x_{t+1} \in B|u_t, x_t, x_{t-1}, \ldots, x_0) = \sum_{u_t} T(x_{t+1} \in B|u_t, x_t) \Pi_t(u_t|x_t) = \sum_{u_t} Q_t^H(x_{t+1} \in B, u_t|x_t) = Q_t^H(x_{t+1} \in B|x_t) \quad (2.3)$$

The essential issue here is that, the control only depends on $x_t$, and since $x_{t+1}$ depends stochastically only on $x_t$ and $u_t$, the desired result follows. In the above, we used properties from total probability and conditioning. If $\Pi_t(u_t|x_t) = H(u_t|x_t)$, that is, $\Pi_t = H$ for all $t$ values so that the policy is stationary, the resulting chain satisfies

$$P_{x_0}^H(x_{t+1} \in B|x_t, x_{t-1}, \ldots, x_0) = Q_t^H(x_{t+1} \in B|x_t),$$

for some $Q^H$. Thus, the transition kernel does not depend on time and the chain is time-homogenous. \hfill \Box

### 2.4 Partially Observed Models and Reduction to a Fully Observed Model

Consider a partially observable stochastic control problem with the following dynamics.

$$x_{t+1} = f(x_t, u_t, w_t), \quad y_t = g(x_t, v_t).$$
Here, $x_t$ is the $\mathbb{X}$-valued state, $u_t$ is the $\mathbb{U}$-valued the control, $y_t$ is the $\mathbb{Y}$-valued observation (measurement) process. Furthermore, $(w_t, v_t)$ are i.i.d. noise processes and $\{w_t\}$ is independent of $\{v_t\}$. The controller only has causal access to $\{y_t\}$.

As noted, $y_t$ denotes an observation variable taking values in $\mathbb{Y}$, a subset of $\mathbb{R}^n$ in the context of this review. The controller only has causal access to the second component $\{y_t\}$ of the process: A deterministic admissible control policy $\Pi$ is a sequence of functions $\{\gamma_t\}$ such that $u_t = \gamma(y_{t-1}; u_{t-1})$.

One could transform a partially observable Markov Decision Problem to a Fully Observed Markov Decision Problem by

$$\begin{align*}
\text{Theorem 2.4.1} & \quad \text{be discussed later):} \\
\text{set of probability measures on } & \mathbb{X} \\
\text{where} & \quad \text{for some} \\
\text{recursion (here, we assume that the state space is countable; the extension to more general spaces will be considered in} \\
\text{Chapter 6 in} \\
\text{Thus, the conditional measure process becomes a controlled Markov chain in } & \mathcal{P}(\mathbb{X}) \text{ (where } \mathcal{P}(\mathbb{X}) \text{ denotes the set of probability measures on } \mathbb{X}, \text{ we will endow this set with the metric giving rise to the weak convergence topology, to be discussed later):} \\
\text{Theorem 2.4.1} & \text{The process } \{\pi_t, u_t\} \text{ is a controlled Markov chain. That is, under any admissible control policy, given the action at time } t \geq 0 \text{ and } \pi_t, \pi_{t+1} \text{ is conditionally independent from } \{\pi_s, u_s, s \leq t-1\}. \\
\text{Proof. This follows from the observation that for any } B \in \mathcal{B}(\mathcal{P}(\mathbb{X})), \text{ by (2.4) } & \\
P(\pi_{t+1} \in B | \pi_s, u_s, s \leq t) \\
= \sum_{y_{t+1} = y} 1\{F(\pi_t, u_t, y) \in B\} P(y_{t+1} = y | \pi_s, u_s, s \leq t) \\
= \sum_{y_{t+1} = y} 1\{F(\pi_t, u_t, y) \in B\} \sum_x P(y_{t+1} = x_{t+1} = x | \pi_s, u_s, s \leq t) P(x_{t+1} = x | \pi_t, u_t) \\
= \sum_{y_{t+1} = y} 1\{F(\pi_t, u_t, y) \in B\} \sum_x P(y_{t+1} = x_{t+1} = x) \sum_{x'} P(x_{t+1} = x | x_t = x', u_t) \pi_t(x') \\
= P(\pi_{t+1} \in B | \pi_t, u_t) \\
\end{align*}$$

Let the cost function to be minimized be

$$E_{x_0}^{\Pi} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right],$$

where $E_{x_0}[\cdots]$ denotes the expectation over all sample paths with initial state given by $x_0$ under policy $\Pi$.

Now, using a property known as iterated expectations that we will be discussing in detail in Chapter 4, we can write:

$$E_{x_0}^{\Pi} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right] = E_{x_0}^{\Pi} \left[ \sum_{t=0}^{T-1} E[c(x_t, u_t) | y_t, y_s, u_s, s \leq t-1] \right] = E_{x_0}^{\Pi} \left[ \sum_{t=0}^{T-1} \tilde{c}(\pi_t, u_t) \right],$$

where

$$\begin{align*}
\tilde{c}(\pi_t, u_t) & = c(x_t, u_t) \\
E_{x_0}^{\Pi}[\cdots] & \text{ denotes the expectation over all sample paths with initial state given by } x_0 \text{ under policy } \Pi. \\
\end{align*}$$
\[ \hat{c}(\pi, u) = \sum_X c(x, u)\pi(x), \quad \pi \in \mathcal{P}(\mathcal{X}). \]

We can thus transform the system into a fully observed Markov model as follows. The stochastic transition kernel \( K \) is given by:

\[ K(B|\pi, u) = \sum_Y 1\{P(y|\pi, u, y) \in B\}P(y|x)\pi(x), \quad \forall B \in \mathcal{B}(\mathcal{P}(\mathcal{X})), \]

with \( 1\{\cdot\} \) denoting the indicator function.

It follows that \((\mathcal{P}(\mathcal{X}), U, K, \hat{c})\) defines a completely observable controlled Markov process.

Thus, the fully observed Markov Decision Model we will consider is sufficiently rich to be applicable to a large class of controlled stochastic systems.

### 2.5 Decentralized Stochastic Systems

We will consider situations in which there are multiple decision makers. We will leave this discussion to Chapter 9.

### 2.6 Controlled Continuous-Time Stochastic Systems

We will also study setups where the time index is a continuum. We leave this discussion to Chapter 10.

### 2.7 Exercises

**Exercise 2.7.1** A common example in mathematical finance applications is the portfolio selection problem where a controller (investor) would like to optimally allocate his wealth between a stochastic stock market and a market with a guaranteed income. Consider a stock with an i.i.d. random return \( \sigma_t \) and a bank account with fixed interest rate \( r > 0 \). These are modeled by:

\[ X_{t+1} = X_t u_t (1 + \sigma_t) + X_t (1 - u_t)(1 + r), \quad X_0 = 1 \]

and

\[ X_{t+1} = X_t (1 + r + u_t(\sigma_t - r)) \]

Here, \( u_t \in [0, 1] \) denotes the proportion of the money that the investor invests in the stock market. Suppose that the goal is to maximize \( E[\log(X_T)] \). Then, we can write:

\[ \log(X_T) = \log(\prod_{k=0}^{T-1} \frac{X_{k+1}}{X_k}) = \sum_{k=0}^{T-1} \log((1 + r + u_t(\sigma_t - r))) \]

Formula the problem as an optimal stochastic control problem by clearly identifying the state and the control action spaces, the information available at the controller, the transition kernel, and a cost functional mapping the actions and states to \( \mathbb{R} \).

**Exercise 2.7.2** Consider an inventory-production system given by

\[ x_{t+1} = x_t + u_t - w_t, \]

where \( x_t \) is \( \mathbb{R} \)-valued, with the one-stage cost
\[ c(x_t, u_t, w_t) = bu_t + h \max(0, x_t + u_t - w_t) + p \max(0, w_t - x_t - u_t) \]

Here, \( b \) is the unit production cost, \( h \) is the unit holding (storage) cost and \( p \) is the unit shortage cost; here we take \( p > b \).

At any given time, the decision maker can take \( u_t \in \mathbb{R}_+ \). The demand variable \( w_t \sim \mu \) is a \( \mathbb{R}_+ \)-valued i.i.d. process, independent of \( x_0 \), with a finite mean where \( \mu \) is assumed to admit a probability density function. The goal is to minimize

\[
J(x, \Pi) = E^{\Pi}_{x_t} \left[ \sum_{t=0}^{T-1} c(x_t, u_t, w_t) \right]
\]

The controller at time \( t \) has access to \( I_t = \{x_s, u_s, s \leq t - 1\} \cup \{x_t\} \).

Formulate the problem as an optimal stochastic control problem by clearly identifying the state, the control action spaces, the information available at the controller, the transition kernel and a cost functional mapping the actions and states to \( \mathbb{R} \).

**Exercise 2.7.3** A fishery manager annually has \( x_t \) units of fish and sells \( u_t x_t \) of these where \( u_t \in [0, 1] \). With the remaining ones, the next year’s production is given by the following model

\[
x_{t+1} = w_t x_t (1 - u_t) + w_t,
\]

with \( x_0 \) is given and \( w_t \) is an independent, identically distributed sequence of random variables and \( w_t \geq 0 \) for all \( t \) and therefore \( E[w_t] = \bar{w} \geq 0 \).

The goal is to maximize the profit over the time horizon \( 0 \leq t \leq T - 1 \). At time \( T \), he sells all of the fish.

Formulate the problem as an optimal stochastic control problem by clearly identifying the state, the control actions, the information available at the controller, the transition kernel and a cost functional mapping the actions and states to \( \mathbb{R} \).

**Exercise 2.7.4** An investor’s wealth dynamics is given by the following:

\[
x_{t+1} = u_t w_t,
\]

where \( \{w_t\} \) is an i.i.d. \( \mathbb{R}_+ \)-valued stochastic process with \( E[w_t] = 1 \). The investor has access to the past and current wealth information and his previous actions. The goal is to maximize:

\[
J(x_0, \Pi) = E^{\Pi}_{x_0} \left[ \sum_{t=0}^{T-1} \sqrt{x_t - u_t} \right].
\]

The investor’s action set for any given \( x \) is: \( U(x) = [0, x] \).

Formulate the problem as an optimal stochastic control problem by clearly identifying the state, the control action spaces, the information available at the controller, the transition kernel and a cost functional mapping the actions and states to \( \mathbb{R} \).

**Exercise 2.7.5** Consider an unemployed person who will have to work for years \( t = 1, 2, \ldots, 10 \) if she takes a job at any given \( t \).

Suppose that each year in which she remains unemployed; she may be offered a good job that pays 10 dollars per year (with probability \( 1/4 \)); she may be offered a bad job that pays 4 dollars per year (with probability \( 1/4 \)); or she may not be offered a job (with probability \( 1/2 \)). These events of job offers are independent from year to year (that is the job market is represented by an independent sequence of random variables for every year).

Once she accepts a job, she will remain in that job for the rest of the ten years. That is, for example, she cannot switch from the bad job to the good job.

Suppose the goal is maximize the expected total earnings in ten years, starting from year 1 up to year 10 (including year 10).

State the problem as a Markov Decision Problem, identify the state space, the action space and the transition kernel.
Exercise 2.7.6 (Zero-Delay Source Coding) Let \( \{x_t\}_{t \geq 0} \) be an \( \mathbb{X} \)-valued discrete-time Markov process where \( \mathbb{X} \) can be a finite set or \( \mathbb{R}^n \). Let there be an encoder which encodes (quantizes) the source samples and transmits the encoded versions to a receiver over a discrete noiseless channel with input and output alphabet \( \mathcal{M} := \{1, 2, \ldots, M\} \), where \( M \) is a positive integer. The encoder policy \( \Pi \) is a sequence of functions \( \{\eta_t\}_{t \geq 0} \) with \( \eta_t : \mathcal{M}^t \times (\mathbb{X})^{t+1} \rightarrow \mathcal{M} \). At time \( t \), the encoder transmits the \( \mathcal{M} \)-valued message

\[ q_t = \eta_t(I_t) \]

with \( I_0 = x_0, I_t = (q_{0:t-1}, x_{0:t}) \) for \( t \geq 1 \), where. The collection of all such zero-delay encoders is called the set of admissible quantization policies and is denoted by \( \Pi_A \). A zero-delay receiver policy is a sequence of functions \( \gamma = \{\gamma_t\}_{t \geq 0} \) of type \( \gamma_t : \mathcal{M}^{t+1} \rightarrow \mathbb{U} \), where \( \mathbb{U} \) denotes the finite reconstruction alphabet. Thus

\[ u_t = \gamma_t(q_{0:t}), \quad t \geq 0. \]

For the finite horizon setting the goal is to minimize the average cumulative cost (distortion)

\[ J_{\pi_0}(\Pi, \gamma, T) = E_{\pi_0}^{H,\gamma} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \right], \tag{2.7} \]

for some \( T \geq 1 \), where \( c_0 : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{R} \) is a nonnegative cost (distortion) function, and \( E_{\pi_0}^{H,\gamma} \) denotes expectation with initial distribution \( \pi_0 \) for \( x_0 \) and under the quantization policy \( \Pi \) and receiver policy \( \gamma \).

Express this problem as a controlled Markov chain problem. Later on, we will provide further refinements. There is a rich history behind this problem, see e.g., [126], [124], [114] and [131, 137].

Exercise 2.7.7 Suppose that there are two decision makers \( DM^1 \) and \( DM^2 \). Suppose that the information available to to \( DM^1 \) is a random variable \( Y^1 \) and the information available to \( DM^2 \) is \( Y^2 \), where these random variables are defined on a probability space \( (\Omega, \mathcal{F}, P) \). Suppose that \( Y^i \) is \( \mathbb{Y}^i \)-valued and these spaces admit a Borel \( \sigma \)-field \( \mathcal{B}(\mathbb{Y}^i) \).

Suppose that the sigma-field generated by \( Y^1 \) is a subset of the sigma-field generated by \( Y^2 \), that is \( \sigma(Y^1) \subset \sigma(Y^2) \).

Further, suppose that the decision makers wish to minimize the following cost function:

\[ E[c(\omega, u)], \]

where \( c : \Omega \times \mathbb{U} \rightarrow \mathbb{R}^+ \) is a measurable cost function (on \( \mathcal{F} \times \mathcal{B}(\mathbb{U}) \)), where \( \mathcal{B}(\mathbb{U}) \) is a \( \sigma \)-field over \( \mathbb{U} \). Here, for \( i = 1, 2 \), \( u^i = \gamma^i(Y^i) \) is generated by a measurable function \( \gamma^i \) on the sigma-field generated by the random variable \( Y^i \). Let \( \Gamma^i \) denote the space of all such policies.

Prove that

\[ \inf_{\gamma^1 \in \Gamma^1} E[c(\omega, u^1)] \geq \inf_{\gamma^2 \in \Gamma^2} E[c(\omega, u^2)]. \]

Hint: Make the argument that every policy \( u^1 = \gamma^1(Y^1) \) can be expressed as \( u^2 = \gamma^2(Y^2) \) for some \( \gamma^2 \in \Gamma^2 \). In particular, for any \( B \in \mathcal{B}(\mathbb{U}) \), \( \gamma^{-1}(B) \in \mathcal{B}(\mathbb{Y}^1) \). Furthermore, \( (Y^1)^{-1}(\gamma^{-1}(B)) \in \sigma(Y^1) \subset \sigma(Y^2) \). Following such a reasoning, the result can be obtained.
Classification of Markov Chains

3.1 Countable State Space Markov Chains

In this chapter, we first review Markov chains where the state takes values in a finite or a countably infinite set \( X \).

We assume that \( \nu_0 \) is an initial distribution for the Markov chain, so that \( x_0 \sim \nu_0 \). The process \( \Phi = \{ x_0, x_1, \ldots, x_n, \ldots \} \) is a (time-homogeneous) Markov chain with the probability measure on the sequence space satisfying:

\[
P_{\nu_0}(x_0 = a_0, x_1 = a_1, x_2 = a_2, \ldots, x_n = a_n) = \nu_0(x_0 = a_0)P(x_1 = a_1|x_0 = a_0)P(x_2 = a_2|x_1 = a_1)\ldots P(x_n = a_n|x_{n-1} = a_{n-1})
\]

(3.1)

If the initial condition is known to be \( x_0 \), we use \( P_{x_0} \) in place of \( P_{\delta x_0} \). The initial condition probability and the transition kernel uniquely identify the probability measure on the product space \( X^\mathbb{N} \), by the extension theorems presented in Chapter 1. We could represent the probabilistic evolution in terms of a matrix:

\[
P(i, j) = P(x_{t+1} = j|x_t = i) \geq 0, \forall i, j \in X
\]

Here \( P(\cdot, \cdot) \) is also a probability transition kernel, that is for every \( i \in X \), \( P(i, \cdot) \) is a probability measure on \( X \): \( \sum_j P(i, j) = 1 \) for every \( i \). Let \( P \) be a \( |X| \times |X| \) matrix with entries given with \( P(i, j) \). \( P \) is a stochastic matrix. By induction, we could verify that

\[
P^k(i, j) := P(x_{t+k} = j|x_t = i) = \sum_{m \in X} P(i, m)P^{k-1}(m, j)
\]

Let \( \pi_k(i) = P(x_k = i) \) for \( k \in \mathbb{Z}_+ \) and \( i \in X \). Let \( \pi_k = [\pi_k(i), i \in X] \). It follows that

\[
\pi_1(j) = \sum_i \pi_0(i)P(i, j)
\]

and with \( P \) denoting the transition matrix given with \( P(i, j) \) as defined above,

\[
\pi_1 = \pi_0 P.
\]

Note that we represent \( \pi_k \) as a row vector.

In the following, we characterize Markov Chains based on transience, recurrence and communication. We then consider the issue of the existence of an invariant distribution. Later, we will extend the analysis to uncountable space Markov chains.
Communication

If there exists an integer \( k \in \mathbb{Z}_+ \) such that 
\[
P(x_{t+k} = j | x_t = i) = P^k(i,j) > 0,
\]
and another integer \( l \) such that 
\[
P(x_{t+l} = i | x_t = j) = P^l(j,i) > 0
\]
then state \( i \) communicates with state \( j \).

A set \( C \subseteq \mathbb{X} \) is said to be communicating if every two elements (states) of \( C \) communicate with each other.

If every member of the set can communicate to every other member, such a chain is said to be irreducible.

The period of a state is defined to be the greatest common divisor of \( \{ k > 0 : P^k(i,i) > 0 \} \).

A Markov chain is called aperiodic if the period of all states is 1.

Absorbing Set

A set \( C \) is called absorbing if \( P(i,C) = 1 \) for all \( i \in C \). That is, if the state is in \( C \), then the state cannot get out of the set \( C \).

The Markov chain is irreducible if the smallest absorbing set is the entire \( \mathbb{X} \) itself.

The Markov chain is indecomposable if \( \mathbb{X} \) does not contain two disjoint absorbing sets.

Occupation, Hitting and Stopping Times

For any set \( A \in \mathbb{X} \), the occupation time \( \eta_A \) is the number of visits of \( \{x_t\} \) to set \( A \):
\[
\eta_A = \sum_{t=0}^{\infty} 1_{\{x_t \in A\}},
\]
where \( 1_E \) denotes the indicator function for an event \( E \), that is, it takes the value 1 when \( E \) takes place, and is otherwise 0.

Define
\[
\tau_A = \min\{ k > 0 : x_k \in A \},
\]
is the first time that the state visits state \( i \), known as the first return time.

We define also a hitting time:
\[
\sigma_A = \min\{ k \geq 0 : x_k \in A \}.
\]

A Brief Discussion on Stopping Times

The variable \( \tau_A \) defined above is an example for stopping times:

**Definition 3.1.1** A \( \mathbb{Z}_+ \)-valued random variable \( \tau \) is a stopping time (with respect to the process \( \{x_0, x_1, \ldots\} \)), if for all \( n \in \mathbb{Z}_+ \), the event \( \{ \tau = n \} \in \sigma(x_0, x_1, x_2, \ldots, x_n) \), that is the event is in the sigma-field generated by the random variables up to time \( n \).

Any realistic decision takes place at a time which is measurable. For example if an investor wants to stop investing when he stops profiting, he can stop at the time when the investment loses value, this is a stopping time. But, if an investor claims to stop investing when the investment is at the peak, this is not a stopping time because to find out whether the investment is at its peak, the next state value should be known, and this is not measurable in a causal fashion.

One important property of Markov chains is the so-called Strong Markov Property. This says the following: Consider a Markov chain which evolves on a countable set. If we sample this chain according to a stopping time rule, the sampled Markov chain starts from the sampled instant as a Markov chain:
Proof. We consider \( \tau \) and let us define \( \tau \) as a stopping time with \( P(\tau < \infty) = 1 \), then almost surely for any \( m \in \mathbb{N} \):

\[
P(x_{\tau+m} = a | x_{\tau} = b_0, x_{\tau-1} = b_1, \ldots) = P(x_{\tau+m} = a | x_{\tau} = b_0) = P^m(b_0, a).
\]

Note that the assumption \( \tau < \infty \) is critically used in the proof. In (3.8), we use the fact that \( \tau \) is a stopping time.

3.1.1 Recurrence and transience

Let us define

\[
U(x, A) := E_x \left[ \sum_{t=1}^{\infty} 1_{(x_t \in A)} \right] = \sum_{t=1}^{\infty} P^t(x, A)
\]

and let us define

\[
L(x, A) := P_x(\tau_A < \infty),
\]

which is the probability of the chain visiting set \( A \), once the process starts at state \( x \).

These two are important characterizations for Markov chains.

Definition 3.1.2 A set \( A \subset \mathbb{X} \) is recurrent if the Markov chain visits \( A \) infinitely often (in expectation), when the process starts in \( A \). This is equivalent to

\[
E_x[\eta_A] = \infty, \quad \forall x \in A
\]

That is, if the chain starts at a given state \( x \in A \), it comes back to the set \( A \), and does so infinitely often. If a state is not recurrent, it is transient.

Definition 3.1.3 A set \( A \subset \mathbb{X} \) is positive recurrent if the Markov chain visits \( A \) infinitely often, when the process starts in \( A \) and in addition:

\[
E_x[\tau_A] < \infty, \quad \forall x \in A
\]


**Definition 3.1.4** A state \( \alpha \) is transient if

\[
U(\alpha, \alpha) = E_\alpha[\eta_\alpha] < \infty.
\]  

(3.5)

The above is equivalent to the condition that

\[
\sum_{i=1}^\infty P_i(\alpha, \alpha) < \infty,
\]

which in turn is implied by

\[
P_i(\tau_i < \infty) < 1.
\]

While discussing recurrence, there is an equivalent, and often easier to check condition: If \( E_i[\tau_i] < \infty \), then the state \{i\} is positive recurrent. If \( L(i,i) = P_i(\tau_i < \infty) = 1 \), but \( E_i[\tau_i] = \infty \), then \( i \) is recurrent (also called, null recurrent).

The reader should connect the above with the strong Markov property: once the process hits a state, it starts from the state as if the past never happened; the process recurs itself.

We have the following result. The proof is presented later in the chapter, see Theorem 3.2.1.

**Theorem 3.1.1** If \( P_i(\tau_i < \infty) = 1 \), then \( P_i(\eta_i = \infty) = 1 \).

There is a more general notion of recurrence, named Harris recurrence:

**Definition 3.1.5** A set \( A \) is Harris recurrent if \( P_x(\eta_A = \infty) = 1 \) for all \( x \in A \). An irreducible Markov chain is Harris recurrent if \( P_x(\eta_A = \infty) = 1 \), \( \forall x \in X, A \subset X \).

We will investigate this property further below while studying uncountable state space Markov chains, however one needs to note that even for countable state space chains Harris recurrence is stronger than recurrence.

Let \( \tau_i(1) := \tau_i \) and for \( i \geq 1 \),

\[
\tau_i(k + 1) = \min\{n > \tau_i(k) : x_n = i\}
\]

For a finite state Markov chain, one can verify that (3.5) is equivalent to \( L(i,i) < 1 \). To show this, it suffices to first verify the relation

\[
P(\tau_i(k) < \infty) = P(\tau_i(k - 1) < \infty)P(\tau_i(1) < \infty),
\]

and then use the equality \( E[\eta] = \sum_{k=1}^\infty P(\eta \geq k) \).

**Remark 3.1.** Harris recurrence is stronger than recurrence. In one, an expectation is considered; in the other, a probability is considered. Consider the following example: Let \( X = \mathbb{N} \), \( P(1,1) = 1 \) and for \( x > 1 \): \( P(x, x+1) = 1 - 1/x^2 \) and \( P(x, 1) = 1/x^2 \). Then, for \( x \geq 2 \): \( P_x(\tau_1 = \infty) = \prod_{t \geq x} (1 - 1/t^2) > 0 \). Thus, the set \{1, 2\} is not Harris recurrent, but it is recurrent.

### 3.1.2 Stability and invariant measures

Stability is an important concept, but it has different meanings in different contexts. This notion will be made more precise in the following two chapters.

If a Markov chain starts at a given time, in the long-run, the chain may forget its initial condition, that is, the probability distribution at a later time will be less and less dependent on the initial distribution. Given an initial state distribution, the probability distribution on the state at time 1 is given by:

\[
\pi_1 = \pi_0 P
\]

And for \( t > 1 \):
\[ \pi_{t+1} = \pi_t P = \pi_0 P^{t+1} \]

One important property of Markov chains is whether the above iteration leads to a fixed point in the set of probability measures. Such a fixed point \( \pi \) is called an **invariant distribution**. A distribution in a countable state Markov chain is invariant if

\[ \pi = \pi P \]

This is equivalent to

\[ \pi(j) = \sum_{i \in \mathcal{X}} \pi(i) P(i, j), \quad \forall j \in \mathcal{X} \]

We note that, if such a \( \pi \) exists, it must be written in terms of \( \pi = \pi_0 \lim_{t \to \infty} P^t \), for some \( \pi_0 \). Clearly, \( \pi_0 \) can be \( \pi \) itself, but often \( \pi_0 \) can be any initial distribution under irreducibility conditions which will be discussed further. Invariant distributions are especially important in networking problems and stochastic control, due to the Ergodic Theorem (which shows that temporal averages converge to statistical averages), which we will discuss later in the semester.

### 3.1.3 Invariant measures via an occupational characterization

**Theorem 3.1.2** For a Markov chain, if there exists an element \( i \) such that \( E_i[\tau_i] < \infty \); the following is an invariant probability measure:

\[ \mu(j) = E \left[ \frac{\sum_{k=0}^{\tau_i-1} 1(x_k = j)}{E_i[\tau_i]} \middle| x_0 = i \right], \quad j \in \mathcal{X} \]

**Proof:**

We show that

\[ E \left[ \frac{\sum_{k=0}^{\tau_i-1} 1(x_k = j)}{E_i[\tau_i]} \middle| x_0 = i \right] = \sum_s P(s, j) E \left[ \frac{\sum_{k=0}^{\tau_i-1} 1(x_k = s)}{E_i[\tau_i]} \middle| x_0 = i \right] \]

Note that \( E[1_{x_{t+1} = j}] = P(x_{t+1} = j) \). Hence,

\[ \sum_s P(s, j) E \left[ \frac{\sum_{k=0}^{\tau_i-1} 1(x_k = s)}{E_i[\tau_i]} \middle| x_0 = i \right] = E \left[ \frac{\sum_{k=0}^{\tau_i-1} \sum_s P(s, j) 1(x_k = s)}{E_i[\tau_i]} \middle| x_0 = i \right] \]

\[ = \frac{E \left[ \sum_{k=0}^{\tau_i-1} \sum_s 1(x_k = s) E[1_{x_{k+1} = j}] \middle| x_k = s, x_{k-1}, \ldots, x_1, x_0 = i \right]}{E_i[\tau_i]} \]

\[ = \frac{E \left[ \sum_{k=0}^{\tau_i-1} \sum_s 1(x_k = s) E[1_{x_{k+1} = j}] \middle| x_k = s, x_{k-1}, \ldots, x_1, x_0 = i \right]}{E_i[\tau_i]} \]

\[ = \frac{E \left[ \sum_{k=0}^{\infty} E[1_{k<\tau_i}] \sum_s 1(x_k = s) E[1_{x_{k+1} = j}] \middle| x_k = s, x_{k-1}, \ldots, x_1, x_0 = i \right]}{E_i[\tau_i]} \]

\[ = \frac{E \left[ \sum_{k=0}^{\infty} 1_{k<\tau_i} \sum_s 1(x_k = s) E[1_{x_{k+1} = j}] \middle| x_k = s, x_{k-1}, \ldots, x_1, x_0 = i \right]}{E_i[\tau_i]} \]

\[ = \frac{E \left[ \sum_{k=0}^{\infty} 1_{k<\tau_i} \sum_s 1(x_k = s) \middle| x_0 = i \right]}{E_i[\tau_i]} \]

\[ = \frac{E \left[ \sum_{k=0}^{\infty} 1_{k<\tau_i} \right]}{E_i[\tau_i]} \]

\[ = \frac{E_i[\tau_i]}{E_i[\tau_i]} \]

\[ = 1 \]

\[ \Rightarrow \mu(j) = \pi(j) \]

\[ \mu \text{ is invariant.} \]
where we use the fact that the number of visits to a given set does not change whether we include or exclude time \( t = 0 \) or \( \tau_i \), since any state \( j \neq i \) is not visited at these times. Here, (3.6) follows from the fact that the process is a Markov chain, (3.7) and (3.8) follow from the properties of conditional expectation and that \( \tau_i \) is a stopping time (we will discuss such properties in Chapter 4), and (3.9) follows from the law of the iterated expectations, see Theorem 4.1.3. This concludes the proof.

\[ \sum \mu(j) = \sum_j E \left[ \sum_{k=0}^{\tau_i-1} 1\{x_k = j\} \bigg| x_0 = i \right] = 1. \]

For example, for a random walk \( E[\tau_i] = \infty \), hence there does not exist an invariant probability measure. But, it has an invariant measure defined with: \( \mu(\{i\}) = K \), for an arbitrary constant \( K \). The reader can verify this.

**Theorem 3.1.3** Every finite state space Markov chain admits an invariant probability measure. This invariant measure is unique if the chain is irreducible.

A common proof technique of this result builds on the Perron-Frobenius theorem. However, we will present a more general result in the context of Markov chains later in Theorem 3.3.1.

**Exercise 3.1.1** For an irreducible Markov chain if there exists an invariant measure, the invariant probability measure is unique.

**Proof:** Let \( \pi(i) \) and \( \pi'(i) \) be two different invariant probability measures. Define \( D := \{i : \pi(i) > \pi'(i)\} \). Then,

\[ \pi(D) = \sum_{i \in D} \pi(i) P(i, D) + \sum_{i \notin D} \pi(i) P(i, D) \]

\[ \pi'(D) = \sum_{i \in D} \pi'(i) P(i, D) + \sum_{i \notin D} \pi'(i) P(i, D) \]

implies that

\[ \pi(D) - \pi'(D) = \sum_{i \in D} (\pi(i) - \pi'(i)) P(i, D) + \sum_{i \notin D} (\pi(i) - \pi'(i)) P(i, D) \]

and thus

\[ \sum_{i \in D} (\pi(i) - \pi'(i))(1 - P(i, D)) = \sum_{i \notin D} (\pi(i) - \pi'(i)) P(i, D) \]

The first term is strictly positive (since \( P(i, D) = 1 \) cannot hold for all \( i \in D \) due to irreducibility, for otherwise \( D \) would be absorbing). The second term is not positive, hence a contradiction.

An implication of the above is the following important result, known as Kac’s lemma:
**Fact 3.1.1** Let \( \{x_t\} \) be irreducible and \( \pi \) be its invariant probability measure. Then,

\[
\pi(i) = \frac{1}{E_i}, \quad i \in X.
\]

**Invariant Distribution via Dobrushin’s Ergodic Coefficient**

Consider the iteration \( \pi_{t+1} = \pi_t P \). We would like to know when this iteration converges to a limit. We first review some notions from analysis.

**Review of Vector Spaces**

**Definition 3.1.6** A linear space is a space which is closed under (pointwise) addition and multiplication by a scalar.

**Definition 3.1.7** A normed linear space \( X \) is a linear vector space on which a functional (a mapping from \( X \) to \( \mathbb{R} \), that is a member of \( \mathbb{R}^X \)) called norm is defined such that:

- \( ||x|| \geq 0 \quad \forall x \in X \), \( ||x|| = 0 \) if and only if \( x \) is the null element (under addition and multiplication) of \( X \).
- \( ||x + y|| \leq ||x|| + ||y|| \) (this is the triangle inequality)
- \( ||\alpha x|| = ||\alpha|| ||x||, \quad \forall \alpha \in \mathbb{R}, \quad \forall x \in X \)

**Definition 3.1.8** In a normed linear space \( X \), an infinite sequence of elements \( \{x_n\} \) converges to an element \( x \) if the sequence \( \{|x_n - x||\} \) converges to zero.

**Definition 3.1.9** A sequence \( \{x_n\} \) in a normed space \( X \) is Cauchy if for every \( \epsilon \), there exists an \( N \) such that \( ||x_n - x_m|| \leq \epsilon \), for all \( n, m \geq N \).

The important observation on Cauchy sequences is that, every converging sequence is Cauchy, however, not all Cauchy sequences are convergent: This is because the limit might not live in the original space where the sequence lives in. This brings the issue of completeness:

**Definition 3.1.10** A normed linear space \( X \) is complete, if every Cauchy sequence in \( X \) has a limit in \( X \). A complete normed linear space is called Banach.

A map \( T \) from one complete normed linear space \( X \) to itself is called a contraction if for some \( 0 \leq \rho < 1 \)

\[
||T(x) - T(y)|| \leq \rho ||x - y||, \forall x, y \in X.
\]

**Theorem 3.1.4** A contraction map in a Banach space has a unique fixed point.

**Proof:** \( \{T^n(x)\} \) forms a Cauchy sequence, and by completeness, the Cauchy sequence has a limit. \( \diamond \)

**Contraction Mapping via Dobrushin’s Ergodic Coefficient**

Consider a countable state Markov Chain. Define

\[
\delta(P) = \min_{i,k} \sum_j \min(P(i,j), P(k,j))
\]

Observe that for two scalars \( a, b \)

\[
|a - b| = a + b - 2 \min(a, b).
\]
Let us define for a vector \( v \) the \( l_1 \) norm:

\[
\|v\|_1 = \sum_{i=1}^{N} |v_i|.
\]

It is known that the set of all countable index real-valued vectors (that is functions which map \( \mathbb{Z} \to \mathbb{R} \)) with a finite \( l_1 \) norm

\[
\{v : \|v\|_1 < \infty\}
\]

is a complete normed linear space, and as such, is a Banach space.

With these observations, we state the following:

**Theorem 3.1.5 (Dobrushin)** For any two probability measures \( \pi, \pi' \), it follows that

\[
\|\pi P - \pi' P\|_1 \leq (1 - \delta(P)) \|\pi - \pi'\|_1
\]

**Proof:** Let \( \psi(i) = \pi(i) - \min(\pi(i), \pi'(i)) \) for all \( i \). Further, let \( \psi'(i) = \pi'(i) - \min(\pi(i), \pi'(i)) \). Since

\[
0 = \sum_i \pi(i) - \pi'(i) = \sum_{i: \pi(i) > \pi'(i)} \pi(i) - \pi'(i) + \sum_{i: \pi(i) > \pi'(i)} \pi(i) - \pi'(i)
\]

we have that \( \|\psi\|_1 = \|\psi'\|_1 \), and since

\[
\sum_i |\pi(i) - \pi'(i)| = \sum_{i: \pi(i) > \pi'(i)} \psi(i) + \sum_{i: \pi(i) > \pi'(i)} \psi'(i)
\]

we have that

\[
\sum_i |\pi(i) - \pi'(i)| = \|\psi\|_1 + \|\psi'\|_1
\]

and thus

\[
\|\pi - \pi'\|_1 = \|\psi - \psi'\|_1 = 2 \|\psi\|_1 = 2 \|\psi'\|_1
\]

Now,

\[
\|\pi P - \pi' P\| = \|\psi P - \psi' P\|
\]

\[
= \sum_j \left| \sum_i \psi(i)P(i, j) - \sum_k \psi'(k)P(k, j) \right|
\]

\[
= \frac{1}{\|\psi'\|_1} \sum_j \left| \sum_k \sum_i \psi(i)\psi'(k)P(i, j) - \psi(i)\psi'(k)P(k, j) \right| - \delta(P)
\]

\[
\leq \frac{1}{\|\psi'\|_1} \sum_j \sum_k \sum_i \psi(i)\psi'(k)\left| P(i, j) - P(k, j) \right| - \delta(P)
\]

\[
= \frac{1}{\|\psi'\|_1} \sum_k \sum_i \psi(i)\psi'(k) \sum_j \left| P(i, j) - P(k, j) \right|
\]

\[
= \frac{1}{\|\psi'\|_1} \sum_i \sum_k |\psi(i)|\psi'(k) \left\{ \sum_j P(i, j) + P(k, j) - 2 \min(P(i, j), P(k, j)) \right\}
\]

\[
\leq \frac{1}{\|\psi'\|_1} \sum_i \sum_k |\psi(i)|\psi'(k)(2 - 2\delta(P))
\]

\[
= \|\psi'\|_1 (2 - 2\delta(P))
\]

\[
= \|\pi - \pi'\|_1 (1 - \delta(P))
\]
In the above, (3.10) follows from adding terms in the summation, (3.11) from taking the norm inside, (3.12) follows from the relation \( ||a - b|| = a + b - 2 \min(a, b) \), (3.13) from the definition of \( \delta(P) \) and finally (3.14) follows from the \( l_1 \) norms of \( \psi, \psi' \).

As such, the process \( P : \pi \in \mathcal{P}(\mathcal{X}) \rightarrow \pi \in \mathcal{P}(\mathcal{X}) \) is a contraction mapping if \( \delta(P) > 0 \). In essence, one proves that such a sequence is Cauchy, and as every Cauchy sequence in a Banach space has a limit, this process also has a limit. In our setting \( \{\pi_0P^n\} \) is a Cauchy sequence. We note that the set of probability measures is not a linear space, but viewed as a closed subset of \( l_1(\mathcal{X}; \mathbb{R}) \), the sequence will have a limit. Since \( \pi P \) is also a probability measure for every \( \pi \in \mathcal{P}(\mathcal{X}) \), the limit must also be a probability measure. The limit is the invariant distribution.

It should also be noted that Dobrushin’s theorem tells us how fast the sequence of probability distributions \( \{\pi_0P^n\} \) converges to the invariant distribution for any arbitrary \( \pi_0 \).

**Ergodic Theorem for Countable State Space Chains**

In Exercise 4.5.10, we will prove the ergodic theorem: Let \( \mathcal{X} \) be positive Harris recurrent with unique invariant distribution \( \mu \); we then have that almost surely
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} f(x_t) = \sum_i f(i)\mu(i)
\]
for bounded \( f \). This is a very important theorem, because, in essence, this property is what makes the connection with stochastic control, and Markov chains in a long time horizon. In particular, for a stationary control policy leading to a unique invariant distribution with bounded costs, it follows that, almost surely,
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} c(x_t, u_t) = \sum_{x,u} c(x, u)\mu(x, u),
\]
\( \forall \) bounded \( c \).

### 3.2 Uncountable (Complete, Separable, Metric) State Spaces

We now extend the discussion above to an uncountable state space setting. We will consider state spaces that are called standard Borel; these are Borel subsets of complete, separable and metric spaces. We note that the spaces that are complete, separable and metric are also called Polish space.

Let \( \{x_t, t \in \mathbb{Z}_+\} \) be a Markov chain with a Polish \( \mathcal{X}, B(\mathcal{X}) \), and defined on a probability space \((\Omega, \mathcal{F}, \mathcal{P})\), where \( B(\mathcal{X}) \) denotes the Borel \( \sigma \)-field on \( \mathcal{X} \), \( \Omega \) is the sample space, \( \mathcal{F} \) a sigma field of subsets of \( \Omega \), and \( \mathcal{P} \) a probability measure. Let \( P(x, D) := P(x_{t+1} \in D|x_t = x) \) denote the transition probability from \( x \) to \( D \), that is the probability of the event \( \{x_{t+1} \in D\} \) given that \( x_t = x \).

Consider a Markov chain with transition probability given by \( P(x, D) \), that is
\[
P(x_{t+1} \in D|x_t = x) = P(x, D)
\]
We could compute \( P(x_{t+k} \in D|x_t = x) \) inductively as follows:
\[
P(x_{t+k} \in D|x_t = x) = \int \cdots \int P(x_t, dx_{t+1}) \cdots P(x_{t+k-2}, dx_{t+k-1})P(x_{t+k-1}, D)
\]
As such, we have for all \( n \geq 1 \), \( P^n(x, A) = P(x_{t+n} \in A|x_t = x) = \int_X P^{n-1}(x, dy)P(y, A) \).
Definition 3.2.1 A Markov chain is $\mu$-irreducible, if for any set $B \in \mathcal{B}(\mathcal{X})$ such that $\mu(B) > 0$, and $\forall x \in \mathcal{X}$, there exists some integer $n > 0$, possibly depending on $B$ and $x$, such that $P^n(x, B) > 0$, where $P^n(x, B)$ is the transition probability in $n$ stages, that is $P(x_{t+n} \in B|x_t = x)$.

A maximal irreducibility measure $\psi$ is an irreducibility measure such that for all other irreducibility measures $\phi$, we have $\psi(B) = 0 \Rightarrow \phi(B) = 0$ for any $B \in \mathcal{B}(\mathcal{X})$ (that is, all other irreducibility measures are absolutely continuous with respect to $\psi$). In the text, whenever a chain is said to be irreducible, a maximal irreducibility measure is implied. We also define $\mathcal{B}^+(\mathcal{X}) = \{ A \in \mathcal{B}(\mathcal{X}) : \psi(A) > 0 \}$ where $\psi$ is a maximal irreducibility measure. A maximal irreducibility measure $\psi$ exists for a $\mu$-irreducible Markov chain, see [82, Proposition 4.2.2].

In the following, when irreducibility is invoked, irreducibility with respect to a maximal irreducibility measure is implied.

Example: Linear system with a drift. Consider the following linear system:

$$x_{t+1} = ax_t + w_t,$$

This chain is Lebesgue irreducible if $w_t$ is a Gaussian variable.

The definitions for recurrence and transience follow those in the countable state space setting.

Definition 3.2.2 A set $A \in \mathcal{B}(\mathcal{X})$ is called recurrent if

$$E_x\left[\sum_{t=1}^{\infty} 1_{x_t \in A}\right] = \sum_{t=1}^{\infty} P^t(x, A) = \infty, \quad \forall x \in \mathcal{X}.$$

A $\psi-$irreducible Markov chain is called recurrent if

$$E_x\left[\sum_{t=1}^{\infty} 1_{x_t \in A}\right] = \sum_{t=1}^{\infty} P^t(x, A) = \infty, \quad \forall x \in \mathcal{X},$$

whenever $\psi(A) > 0$.

A set $A \in \mathcal{B}(\mathcal{X})$ is Harris recurrent if the Markov chain visits $A$ infinitely often with probability 1, when the process starts in $A$:

Definition 3.2.3 A set $A \in \mathcal{B}(\mathcal{X})$ is Harris recurrent if

$$P_x(\eta_A = \infty) = 1, \quad A \in \mathcal{B}(\mathcal{X}), \quad \forall x \in A,$$

(3.16)

A $\psi$-irreducible Markov chain is Harris recurrent if

$$P_x(\eta_A = \infty) = 1, \quad A \in \mathcal{B}(\mathcal{X}), \quad \forall x \in \mathcal{X},$$

and $\psi(A) > 0$.

Theorem 3.2.1 Harris recurrence of a set $A$ is equivalent to

$$P_x(\tau_A < \infty) = 1, \quad \forall x \in A.$$

Proof: Let $\tau_A(1)$ be the first time the state hits $A$. By the Strong Markov Property, the Markov chain sampled at successive intervals $\tau_A(1)$, $\tau_A(2)$ and so on is also a Markov chain. Let $Q$ be the transition kernel for this sampled Markov Chain. Now, the probability of $\tau_A(2) < \infty$ can be computed recursively as

$$P(\tau_A(2) < \infty) = \int_A \int A Q_{x_{\tau_A(1)}}(x_{\tau_A(1)}, dy) P_y(\tau_A(1) < \infty).$$
3.2 Uncountable (Complete, Separable, Metric) State Spaces

\[ = \left( \int_A Q_{x_A(1)}(x_{\tau_A(1)}, dy) \right) P_y(\tau_A(1) < \infty) = 1 \quad (3.17) \]

By induction, for every \( n \in \mathbb{Z}_+ \)

\[ P(\tau_A(n + 1) < \infty) = \int_A Q_{x_A(1)}(x_{\tau_A(1)}, dy) P_y(\tau_A(n) < \infty) = 1 \quad (3.18) \]

Now,

\[ P_x(\eta_A \geq k) = P_x(\tau_A(k) < \infty), \]

since \( k \) times visiting a set requires \( k \) times returning to a set, when the initial state \( x \) is in the set. As such,

\[ P_x(\eta_A \geq k) = 1, \forall k \in \mathbb{Z}_+ \]

is identically equal to 1. Define \( B_k = \{ \omega \in \Omega : \eta(\omega) \geq k \} \), and it follows that \( B_{k+1} \subset B_k \). By the continuity of probability \( P(\bigcap B_k) = \lim_{k \to \infty} P(B_k) \), it follows that \( P_x(\eta_A = \infty) = 1 \).

The proof of the other direction for showing equivalence is left as an exercise to the reader. \( \diamond \)

**Definition 3.2.4** If a Harris recurrent Markov chain chain admits an invariant probability measure, then the chain is called positive Harris recurrent.

If a set is not recurrent, it is transient. A set \( A \) is transient if for some \( x \in A \)

\[ U(x, A) = E_x[\eta_A] < \infty. \quad (3.19) \]

Note that, this is equivalent to \( \sum_{i=1}^{\infty} P^i(x, A) < \infty. \)

3.2.1 Invariant probability measures

**Definition 3.2.5** For a Markov chain with transition probability defined as before, a probability measure \( \pi \) is invariant on the Borel space \((\mathbb{X}, \mathcal{B}(\mathbb{X}))\) if

\[ \pi(D) = \int_{\mathbb{X}} P(x, D) \pi(dx), \quad \forall D \in \mathcal{B}(\mathbb{X}). \]

Uncountable chains act like countable ones when there is a single atom \( \alpha \in \mathbb{X} \) which satisfies a finite mean return property to be discussed below.

**Definition 3.2.6** A set \( \alpha \) is called an atom if there exists a probability measure \( \nu \) such that

\[ P(x, A) = \nu(A), \quad \forall x \in \alpha, \forall A \in \mathcal{B}(\mathbb{X}). \]

If the chain is \( \mu \)-irreducible and \( \mu(\alpha) > 0 \), then \( \alpha \) is called an accessible atom.

In case there is an accessible atom \( \alpha \), we have the following:

**Theorem 3.2.2** For an \( \mu \)-irreducible Markov chain for which \( E_\alpha[\tau_\alpha] < \infty \); the following is the invariant probability measure:

\[ \pi(A) = E_\alpha \left[ \frac{\sum_{k=0}^{\tau_\alpha - 1} 1_{x_k \in A}}{E[\tau_\alpha]} | x_0 = \alpha \right], \quad \forall A \in \mathcal{B}(\mathbb{X}), \mu(A) > 0 \]

**Small Sets and Nummelin and Athreya-Ney’s Splitting Technique**

In case an atom is not present, one may construct an artificial atom:
A set \( A \in \mathcal{B}(\mathbb{X}) \) is \( n \)-small on \((\mathbb{X}, \mathcal{B}(\mathbb{X}))\) if for some positive measure \( \mu_n \)

\[
P^n(x, B) \geq \mu_n(B), \quad \forall x \in A, \text{ and } B \in \mathcal{B}(\mathbb{X}),
\]

where \( \mathcal{B}(\mathbb{X}) \) denotes the (Borel) sigma-field on \( \mathbb{X} \).

Small sets exist for irreducible Markov chains:

**Theorem 3.2.3** [85] Let \( \{x_t\} \) be \( \mu \)-irreducible. Then, for every Borel \( B \) with \( \mu(B) > 0 \), there exists \( m \geq 1 \) and a \( \nu_m \) small set \( C \) with \( \mu(C) > 0 \) and \( \nu_m(C) > 0 \).

**Definition 3.2.8** [85] A set \( A \in \mathcal{B}(\mathbb{X}) \) is \( \nu \)-petite on \((\mathbb{X}, \mathcal{B}(\mathbb{X}))\) if for some distribution \( T \) on \( \mathbb{N} \) (set of natural numbers), and some measure \( \mu \),

\[
\sum_{n=0}^{\infty} P^n(x, B)T(n) \geq \nu(B), \quad \forall x \in A, \text{ and } B \in \mathcal{B}(\mathbb{X}).
\]

A useful result is as follows.

**Theorem 3.2.4** [85] Let \( \{x_t\} \) be \( \mu \)-irreducible and let \( A \) be \( \nu \)-petite. Then, there exists a sampling distribution such that \( A \) is \( \psi \)-petite where \( \psi \) is a maximal irreducibility measure. Furthermore, \( A \) is \( \psi \)-petite for a sampling distribution with finite mean.

**Definition 3.2.9** A \( \psi \)-irreducible Markov chain is periodic with period \( d \) if there exists a partition of \( \mathbb{X} = \bigcup_{i=1}^{d} \mathbb{X}_i \cup D \) so that \( P(x, x_{t+1}) = 1 \) for all \( x \in \mathbb{X}_i \) and \( P(x, X_1) = 1 \) for all \( x \in \mathbb{X}_d \), for some \( d > 1 \) with \( \psi(D) = 0 \). If \( d = 1 \), the chain is aperiodic.

Another very useful result is the following:

**Theorem 3.2.5** [Theorem 5.5.3 of [85]] For an aperiodic and irreducible Markov chain \( \{x_t\} \) every petite set is \( \nu \)-small for some appropriate \( \nu \) (but now \( \nu \) may not be a maximal irreducibility measure; compare with Theorem 3.2.4).

The results on recurrence apply to uncountable chains with no atom provided there is a small set or a petite set. In the following, we construct an artificial atom through what is commonly known as the splitting technique, see [91] [92] (see also [10]).

Suppose a set \( A \) is 1-small. Define a process \( z_t = (x_t, \alpha_t) \), \( z_t \in \mathbb{X} \times \{0, 1\} \). That is we enlarge the state space. Suppose that when \( x_t \notin A \), \((\alpha_t, x_t)\) evolve independently from each other. However, when \( x_t \in A \), we pick a Bernoulli random variable, and with probability \( \delta \) the state visits \( A \times \{1\} \) and with probability \( 1 - \delta \) visits \( A \times \{0\} \). From \( A \times \{1\} \), the transition for the next time stage is given by \( \frac{\nu(dx_{t+1})}{\delta} \) and from \( A \times \{0\} \), it visits the future time stage with probability

\[
P(dx_{t+1}|x_t) = \nu(dx_{t+1}) \frac{1 - \delta}{\delta}.
\]

Now, pick \( \delta = \nu(\mathbb{X}) \). In this case, \( A \times \{1\} \) is an accessible atom, and one can verify that the marginal distribution of the original Markov process \( \{x_t\} \) has not been altered.

The following can be established using the above construction.

**Proposition 3.2.1** If

\[
\sup_{x \in A} E[\min(t > 0 : x_t \in A)|x_0 = x] \leq \infty
\]

then,

\[
\sup_{z \in (A \times \{1\})} E[\min(t > 0 : z_t \in (A \times \{1\}))|z_0 = z] \leq \infty.
\]
Now suppose that a set \( A \) is \( m \)-small. Then, we can construct a split chain for the sampled process \( x_{mn}, n \in \mathbb{N} \). Note that this sampled chain has a transition kernel as \( P^m \). We replace the discussion for the 1-small case with the sampled chain (also known as the \( m \)-skeleton of the original chain). If one can show that the sampled chain has an invariant measure \( \pi_m \), then (see Theorem 10.4.5 of [82]):

\[
\pi(B) := \frac{1}{m} \sum_{k=0}^{m-1} \int \pi_m(dx)P^k(x, B)
\]

(3.20)

is invariant for \( P \). Furthermore, \( \pi \) is also invariant for the sampled chain with kernel \( P^m \). Hence if \( P^m \) leads to a unique invariant probability measure, \( \pi = \pi_m \).

The above also applies for an arbitrary sampling distribution \( K \) on \( \mathbb{N} \). Suppose that we have

\[
\int \pi_K(dx) \left( \sum_n K(n)P^m(x, B) \right) = \pi_K(B), \quad B \in \mathcal{B}(\mathbb{X})
\]

Then,

\[
\pi(B) := \int \sum_m K(m) \sum_{k=0}^{m-1} \pi_K(dx)P^k(x, B)
\]

(3.21)

is an invariant measure for the original chain so that \( \pi = \pi_P \). By normalizing this measure, we obtain an invariant measure for the original chain, provided that \( \sum_n nK(n) < \infty \) (see Theorem 3.2.4).

### 3.2.2 Existence of an invariant probability measure

We state the following very useful result on the existence of invariant distributions for Markov chains.

**Theorem 3.2.6** Consider a Markov process \( \{x_t\} \) taking values in \( \mathbb{X} \). If there exists a set \( A \) which is also an \( m \)-small set for some \( m \in \mathbb{Z}_+ \), and if the set satisfies

\[
\sup_{x \in A} E[\min(t > 0 : x_t \in A)|x_0 = x] < \infty,
\]

then the Markov chain admits an invariant probability measure.

In the following, we relax the small set property, but impose irreducibility

**Theorem 3.2.7** (Meyn-Tweedie) Consider a \( \psi \)-irreducible Harris recurrent Markov process \( \{x_t\} \) taking values in \( \mathbb{X} \). If there exists a \( \mu \)-petite set \( A \) for some positive measure \( \mu \), \( P_x(\tau_A < \infty) = 1 \) for all \( x \), and if the set satisfies

\[
\sup_{x \in A} E[\min(t > 0 : x_t \in A)|x_0 = x] < \infty,
\]

then the Markov chain is positive Harris recurrent (and admits a unique invariant distribution).

**Remark 3.3.** The apriori assumption of the irreducibility of the chain can be eliminated in the theorem above (see [37]), since the other conditions specified already lead to an irreducible chain taking values in a subset of \( \mathbb{X} \): the small/petite set together with the return property \( P_x(\tau_A < \infty) = 1 \) for all \( x \), imply the presence of an irreducibility measure for a smaller state space in which the chain lives.

See Remark 4.3 on a positive Harris recurrence discussion for an \( m \)-skeleton and split chains: When a Markov chain has an invariant probability measure, the sampled chain (\( m \)-skeleton) also satisfies a drift condition, which then leads to the result
that an atom constructed through an $m$-skeleton has a finite return property, which can be used to establish the existence of an invariant probability measure.

In this case, the invariant measure satisfies the following, which is a generalization of Kac’s Lemma [44]:

**Theorem 3.2.8** For a $\mu$-irreducible Markov chain with a unique invariant probability measure $\pi$, the following holds:

$$\pi(A) = \int_C \pi(dx) E_x \left[ \sum_{k=0}^{\tau_C-1} 1_{\{x_k \in A\}} \right], \quad \forall A \in B(X), \mu(A) > 0, \pi(C) > 0$$

The above can also be extended to compute that expected values of a function of the Markov states. The above can be verified along the same lines used for the countable state space case (see Theorem 3.1.2).

### 3.2.3 On small and petite sets: two sufficient conditions

Establishing the smallness or petiteness of a set may be difficult to directly verify. In the following, we present two conditions that may be used to establish the petiteness properties.

By [82], p. 131: For a Markov chain with transition kernel $P$ and $K$ a probability measure on natural numbers, if there exists for every $E \in B(X)$, a lower semi-continuous function $N(\cdot, E)$ such that $\sum_{n=0}^{\infty} P^n(x, E) K(n) \geq N(x, E)$, for a sub-stochastic kernel $N(\cdot, \cdot)$, the chain is called a $T-$chain.

**Theorem 3.2.9** [82] For a $T-$chain which is irreducible, every compact set is petite.

For a countable state space, under irreducibility, every finite set $S$ is petite.

Tweedie [118] considers the following. If $S$ is such that, the following uniform countable additivity condition

$$\lim_{n \to \infty} \sup_{x \in S} P(x, B_n) = 0, \quad (3.22)$$

is satisfied for $B_n \downarrow \emptyset$, then, $S$ is petite (and for example, (4.6) in Chapter 4 implies the existence of an invariant probability measure). There exists at most finitely many invariant probability measures. By [82] Proposition 5.5.5 (iii), under irreducibility, the Harris recurrent component of the space can be expressed as a countable union of petite sets $C_n$ with $\bigcup_{n=1}^{\infty} C_n$, with $\bigcup_{n=1}^{\infty} C_n \to \emptyset$ as $n \to \infty$. By Lemma 4 of Tweedie (2001), under uniform countable additivity, any set $\bigcup_{n=1}^{\infty} C_n$ is uniformly accessible from $S$. Therefore, if the Markov chain is irreducible, the condition (3.22) implies that the set $S$ is petite. This may be easier to verify for a large class of applications. Under further conditions (such as if $S$ is compact and $V$ used in a drift criterion has compact level sets), then the analysis will lead sufficient conditions leading to (3.22). In particular, [118] Lemma 1] notes that if $S$ is bounded and $V$ is continuous (and thus uniformly bounded on $S$), it suffices to test (3.22) only for $B_n$ sets inside sets on which $V$ is bounded (that is with $B_1$ such that $\sup_{x \in B_1} V(x) < \infty$).

In applications, this is often much easier to apply, see e.g. [140].

### 3.2.4 Dobrushin’s Ergodic coefficient for general state spaces

We can extend Dobrushin’s contraction result for the uncountable state space case. By the property that $|a - b| = a + b - 2 \min(a, b)$, the Dobrushin’s coefficient is also related to the term (for the countable state space case):

$$\delta(P) = 1 - \frac{1}{2} \max_{i,k} \sum_j |(P(i, j) - P(k, j)|$$

In case $P(x, dy)$ is the stochastic transition kernel of a real-valued Markov process with transition kernels admitting densities (that is $P(x, A) = \int_A p(x, y)dy$ admits a density), the expression
### 3.3 Further Conditions on the Existence and Uniqueness of Invariant Probability Measures

**3.3.1 Markov chains with the Feller property**

This section uses certain properties of the space of probability measures, reviewed briefly in Section ??.

A Markov chain is weak Feller if $\int_X P(dz|x)v(z)$ is continuous in $x$ for every continuous $v$ on $X$.

**Theorem 3.3.1** Let $\{x_t\}$ be a weak Feller Markov process living in a compact subset of a complete, separable metric space. Then $\{x_t\}$ admits an invariant distribution.

**Proof.** Proof follows the observation that the space of probability measures on a compact set is tight (that is, it is weakly sequentially pre-compact), see Appendix D for a discussion on weak convergence. Consider a sequence $\mu_T = \frac{1}{T} \sum_{t=0}^{T-1} \mu_0 P_t$, $T \geq 1$. There exists a subsequence $\mu_{T_k}$ which converges weakly to some $\mu^*$. It follows that for every continuous and bounded function $f$

$$\langle \mu_{T_k}, f \rangle := \int \mu_{T_k}(dx) f(x) \to \langle \mu^*, f \rangle$$

Likewise, since $Pf(x) = \int f(x_1)P(dx_1|x_0 = x)$ is continuous in $x$ (by the weak Feller condition), it follows that

$$\langle \mu_{T_k}, Pf \rangle := \int \mu_{T_k}(dx) \left( \int P(dy|x)f(y) \right) \to \langle \mu^*, Pf \rangle.$$ 

Now,

$$\langle \mu_{T_k} - \mu_{T_k}P, f \rangle = \frac{1}{T_k} E_{\mu_0} \left( \sum_{k=0}^{T_k-1} P_x^k f - \sum_{k=0}^{T_k-1} P_x^{k+1} f \right)$$

$$= \frac{1}{T_k} E_{\mu_0} \left( f(x_0) - f(x_{T_k}) \right) \to 0.$$ \hspace{1cm} (3.23)

Thus,

$$\langle \mu_{T_k} - \mu_{T_k}P, f \rangle = \langle \mu_{T_k}, f \rangle - \langle \mu_{T_k}P, f \rangle = \langle \mu_{T_k}, f \rangle - \langle \mu_{T_k}, Pf \rangle \to \langle \mu^* - \mu^*P, f \rangle = 0.$$ 

Now, if the relation $\langle \mu^* - \mu^*P, f \rangle = 0$ holds for every continuous and bounded function, it also holds for any measurable function $f$: This is because continuous functions are dense in measurable functions under the supremum norm (in other words, continuous and bounded functions form a separating class for the space of probability measures, see e.g. p. 13 in [19] or Theorem 3.4.5 in [46]). Thus, $\mu^*$ is an invariant probability measure. \hfill \Box

**Remark 3.4.** The theorem applies identically if instead of a compact set assumption, one assumes that the sequence $\mu_k$ takes values in a weakly compact set; that is if the sequence admits a weakly converging subsequence.

**Remark 3.5.** Reference [75] gives the following example to emphasize the importance of the Feller property: Consider a Markov chain evolving in $[0, 1]$ given by: $P(x, x/2) = 1$ for all $x \neq 0$ and $P(0, 1) = 1$. This chain does not admit an invariant measure.
### 3.3.2 Quasi-Feller chains

Often, one does not have the Feller property, but the set of discontinuity is appropriately negligible.

**Assumption 3.3.1** \( P \) is continuous on \( \mathbb{X} \setminus D \) where \( D \) is a closed set with \( P(X_{t+1} \in D | x) = 0 \) for all \( x \). Furthermore, with \( D_m = \{ z : d(z, D) < \epsilon \} \) for \( \epsilon > 0 \) and \( d \) the metric on \( \mathbb{X} \), for some \( K < \infty \), we have that for all \( x \) and \( \epsilon > 0 \)

\[
P(X_{t+1} \in D_m | x_t = x) \leq Ke.
\]

**Theorem 3.3.2** Suppose that Assumption 3.3.1 holds. If the state space is compact, there exists an invariant probability measure for the Markov chain.

**Proof.** The sequence of expected empirical probability measures

\[
v_n(A) = E_x[\frac{1}{n} \sum_{k=0}^{n-1} 1\{X_k \in A\}]
\]

is tight, and thus there exists a weakly converging subsequence. Assumption 3.3.1 implies that every converging subsequence \( v_{n_k} \) of is such that for all \( \epsilon > 0 \)

\[
\lim_{n_k \to \infty} v_{n_k}(D_\epsilon) \leq K \epsilon.
\]

Note that with \( v = \lim_{n_k \to \infty} v_{n_k} \), it follows from the Portmanteau theorem (see e.g. [43, Thm. 11.1.1]) that

\[
v(D_\epsilon) \leq K \epsilon.
\]

Now, consider a weakly converging empirical occupation sequence \( v_{n_k} \) and let this sequence have an accumulation point \( v^* \). We will show that \( v^* \) is invariant.

Observe that the transitioned probability measure \( v_{n_k} P \) satisfies the following for every continuous and bounded \( f \): Consider \( \langle v_{n_k}, Pf \rangle = \langle v_{n_k}, g_f \rangle + \langle v_{n_k}, Pf - g_f \rangle \), where \( g_f \) is a continuous function which is equal to \( Pf \) outside an open neighborhood of \( D \) and is continuous with \( \|g_f\|_{\infty} = \|Pf\|_{\infty} \leq \|f\|_{\infty} \). The existence of such a function follows from the Tietze-Urysohn extension theorem [43], where the closed set is given by \( \mathbb{X} \setminus D \). It then follows from Assumption 3.3.1 that, for every \( \epsilon > 0 \) a corresponding \( g_f \) can be found so that \( \langle v_{n_k}, Pf - g_f \rangle \leq K \|f\|_{\infty} \epsilon \), and since \( \langle v_{n_k}, g_f \rangle \to \langle v^*, g_f \rangle \), it follows that

\[
\lim_{n_k \to \infty} \left| \langle v_{n_k}, Pf \rangle - \langle v^*, Pf \rangle \right| = \lim_{n_k \to \infty} \left| \langle v_{n_k}, Pf - g_f \rangle - \langle v^*, Pf - g_f \rangle \right| \\
\leq \lim_{n_k \to \infty} \left| \langle v_{n_k}, Pf - g_f \rangle \right| + \left| \langle v^*, Pf - g_f \rangle \right| \\
\leq 2K' \epsilon
\]

Here, \( K' = 2K \|f\|_{\infty} \) is fixed and \( \epsilon \) may be made arbitrarily small. We conclude that \( v^* \) is invariant.

**Remark 3.6.** In his definition for quasi-Feller chains, Lasserre assumes the state space to be locally compact. In the proof above [138] tightness is invoked directly with no use of convergence properties of the set of functions which decay to zero as is done in [61]; for a related result see Gersho [52].

### 3.3.3 Cases without the Feller condition

One can relax the weak Feller condition and instead consider spaces of probability measures which are setwise sequentially pre-compact. The proof of this result follows from a similar observation as (3.23) but with weak convergence replaced by
3.4 Ergodicity Properties

3.4.1 Uniqueness of an invariant probability measure and ergodicity

For a Markov chain, the uniqueness of an invariant probability measure implies the ergodicity of the measure.

Theorem 3.4.1 Let \( \{x_t\} \) be a \( \psi \)-irreducible Markov chain which admits an invariant probability measure. The invariant measure is unique.

Proof. Let there be two invariant probability measures \( \mu_1 \) and \( \mu_2 \). Then, there exists two mutually singular invariant probability measures \( \nu_1 \) and \( \nu_2 \), that is \( \nu_1(\mathcal{B}_1) = 1 \) and \( \nu_2(\mathcal{B}_2) = 1 \), \( \mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset \) and that \( P^n(x, \mathcal{B}_1) = 0 \) for all \( x \in \mathcal{B}_1 \) and \( n \in \mathbb{Z}_+ \) and likewise \( P^n(z, \mathcal{B}_2) = 0 \) for all \( z \in \mathcal{B}_1 \) and \( n \in \mathbb{Z}_+ \). This then implies that the irreducibility measure has zero support on \( \mathcal{B}_1 \) and zero support on \( \mathcal{B}_2 \) and thus on \( X \), leading to a contradiction.

A further result on uniqueness is given next.

Definition 3.7. For a Markov chain with transition kernel \( P \), a point \( x \) is accessible if for every \( y \) and every open neighborhood \( O \) of \( x \), there exists \( k > 0 \) such that \( P^k(y, O) > 0 \).

One can show that if a point is accessible, it belongs to the (topological) support of every invariant measure (see, e.g., Lemma 2.2 in [56]). The support (or spectrum) of a probability measure is defined to be the set of all points \( x \) for which every open neighbourhood of \( x \) has positive measure. A Markov chain \( V_t \) is said to have the strong Feller property if \( E[f|V_{t+1}|V_t = v] \) is continuous in \( v \) for every measurable and bounded \( f \).

Theorem 3.4.2 [56] If a Markov chain over a Polish space has the strong Feller property, and if there exists an accessible point, then the chain can have at most one invariant probability measure.

3.4.2 Ergodic theorems for positive Harris recurrent chains

Let \( c \in L_1(\mu) := \{ f : \int |f(x)|\mu(dx) < \infty \} \). Suppose that \( \mu \) is an invariant probability measure for a Markov chain. Then, by the individual ergodic theorem (e.g., Theorem 2.3.4 in [62]) it follows that for \( \mu \) almost everywhere \( x \in X \):

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} c(x_t) = \int c(x)\mu(dx),
\]

\( P_x \) almost surely (that is conditioned on \( x_0 = x \), with probability one, the above holds). Furthermore, again with \( c \in L_1(\mu) \), for \( \mu \) almost everywhere \( x \in X \).


\[
\lim_{T \to \infty} \frac{1}{T}E_x \left[ \sum_{t=1}^{T} c(x_t) \right] = \int c(x) \mu(dx),
\]

On the other hand, the positive Harris recurrence property allows the convergence to take place for every initial condition and is thus a refinement to the the result above: If \( \mu \) is the invariant probability measure for a positive Harris recurrent Markov chain, it follows that for all \( x \in X \) and every \( c \in L_1(\mu) \),

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} c(x_t) = \int c(x) \mu(dx),
\]

\( P_\epsilon \) almost surely (that is conditioned on \( x_0 = x \), with probability one, the above holds).

In fact, Theorem 4.2.13 of [62] establishes that if (3.26) holds for every \( c \in L_1(\mu) \) and every \( x \), then the chain must be positive Harris recurrent provided that an invariant probability measure \( \mu \) exists.

Furthermore, with \( c \) bounded, forall \( x \in X \)

\[
\lim_{T \to \infty} \frac{1}{T}E_x \left[ \sum_{t=1}^{T} c(x_t) \right] = \int c(x) \mu(dx),
\]

See Chapter 4 in [62] for further discussions.

### 3.4.3 Further ergodic theorems for Markov chains

Although beyond the scope of this course, for completeness, we state the following. When an invariant probability measure is known to exist for a Markov chain, we state the following ergodicity results.

**Theorem 3.4.3** [61] Theorems 2.3.4-2.3.5] Let \( \bar{P} \) be an invariant probability measure for a Markov process.

(i) [Individual ergodic theorem] Let \( X_0 = x \). For every \( f \in L_1(\bar{P}) \)

\[
\frac{1}{N} E_x \left[ \sum_{n=0}^{N-1} f(X_n) \right] \to f^*(x),
\]

for all \( x \in B_f \) where \( \bar{P}(B_f) = 1 \) (where \( B_f \) denotes that the set of convergence may depend on \( f \) for some \( f^* \)).

(ii) [Mean ergodic theorem] Furthermore, the convergence \( \frac{1}{N} E_x \left[ \sum_{n=0}^{N-1} f(X_n) \right] \to f^*(x) \) is in \( L_1(\bar{P}) \).

**Theorem 3.4.4** [61] Theorem 2.5.1] Let \( \bar{P} \) be an invariant probability measure for a Markov process. With \( X_0 = x \), for every \( f \in L_1(\bar{P}) \)

\[
\frac{1}{N} \sum_{n=0}^{N-1} f(X_n) \to f^*(x),
\]

for all \( x \in B_f \) where \( \bar{P}(B_f) = 1 \) for some \( f^*(x) \) with

\[
\int \bar{P}(dx) f^*(x) = \int \bar{P}(dx) f(x)
\]

One may state further refinements; see [61] for the locally compact case and [132] for the Polish state space case.

**Theorem 3.4.5** [61] [132] Let \( \bar{P} \) be an invariant probability measure for a Markov process.
3.5 Exercises

(i) [Ergodic decomposition and weak convergence] For $x$, $P$ a.s., $\frac{1}{N}E_x[\sum_{t=0}^{N-1} 1_{\{x_n \in \cdot\}}] \to P_x(\cdot)$ weakly and $P$ is invariant for $P_x(\cdot)$ in the sense that

$$\bar{P}(B) = \int P_x(B) \bar{P}(dx)$$

(ii) [Convergence in total variation] For all $\mu \in \mathcal{P}(\mathbb{N})$ which satisfies that $\mu \ll \bar{P}$ (that is, $\mu$ is absolutely continuous with respect to $\bar{P}$), there exists $v^*$ such that

$$\|E_{\mu}\frac{1}{N} \sum_{t=0}^{N-1} 1_{\{T_n X \in \cdot\}} - v^*(\cdot)\|_{TV} \to 0.$$ 

3.5 Exercises

**Exercise 3.5.1** For a countable state space Markov chain, prove that if $\{x_t\}$ is irreducible, then all states have the same period.

**Exercise 3.5.2** Prove that

$$P_x(\tau_A = 1) = P(x, A),$$

and for $n \geq 1$,

$$P_x(\tau_A = n) = \sum_{i \in A} P(x, i) P_i(\tau_A = n - 1)$$

**Exercise 3.5.3** Let $\{x_t\}$ be a Markov chain defined on state space $\{0, 1, 2\}$. Let the one-stage probability transition matrix be given by:

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 2/3 & 1/3 \end{bmatrix}$$

Compute $E[\min(t \geq 0 : x_t = 2)|x_0 = 0]$, that is the expected minimum number of stages for the state to move from 0 to 2.

**Exercise 3.5.4** Consider a line of customers at a service station (such as at an airport, a grocery store, or a communication network where customers are packets). Let $L_t$ be the length of the line, that is the total number of customers waiting in the line.

Let there be $M_t$ servers, serving the customers at time $t$. Let there be a manager (controller) who decides on the number of servers to be present. Let each of the servers be able to serve $N$ customers for every time-stage. The dynamics of the line can be expressed as follows:

$$L_{t+1} = L_t + A_t - M_t N I_{(L_t \geq N M_t)},$$

where $1_{(E)}$ is the indicator function for event $E$, i.e., it is equal to zero if $E$ does not occur and is equal to 1 otherwise. In the equation above, $A_t$ is the number of customers that have just arrived at time $t$. We assume $\{A_t\}$ to be an independent process, and to have an exponential distribution, with mean $\lambda$, that is for all $k \in \mathbb{Z}_+$

$$P(A(t) = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \ldots\}$$

The manager has only access to the information vector
Is there an absorbing set in this Markov chain? If so, what is it?

According to Policy A, the number of servers is given by:

\[ M_t = \frac{L_t + L_{t-1} + L_{t-3}}{2}. \]

According to Policy B, the number of servers is

\[ M_t = \frac{L_{t+1} + L_t + L_{t-1}}{2}. \]

According to Policy C,

\[ M_t = \lceil \lambda + 0.1 \rceil 1_{t \geq 10}. \]

Finally, according to Policy D, the update is

\[ M_t = [\lambda + 0.1]. \]

a) Which of these policies are admissible, that is measurable with respect to the \( \sigma \)-field generated by \( I_t \)? Which are Markov, or stationary?

b) Consider the model above. Further, suppose policy D is adopted by the manager. Consider the induced Markov chain by the policy, that is, consider a Markov chain with the following dynamics:

\[ L_{t+1} = L_t + A_t - \lceil \lambda + 0.1 \rceil N 1_{1 \leq 10}. \]

Is this Markov chain irreducible?

Is there an absorbing set in this Markov chain? If so, what is it?

**Exercise 3.5.5** Show that irreducibility of a Markov chain in a finite state space implies that every set \( A \) and every \( x \) satisfies \( U(x, A) = \infty \).

**Exercise 3.5.6** Show that for an irreducible Markov chain, either the entire chain is transient, or recurrent.

**Exercise 3.5.7** If \( P_a(\tau_a < \infty) < 1 \), show that \( E_a[\sum_{k=1}^{\infty} 1_{\{x_k = a\}}] < \infty \).

**Exercise 3.5.8** If \( P_a(\tau_a = \infty) < 1 \), show that \( P_a(\sum_{k=1}^{\infty} 1_{\{x_k = a\}} = \infty) = 1 \).

**Exercise 3.5.9 (Gambler’s Ruin)** Consider an asymmetric random walk defined as follows: \( P(x_{t+1} = x+1|x_t = x) = p \) and \( P(x_{t+1} = x-1|x_t = x) = 1-p \) for any integer \( x \). Suppose that \( x_0 = x \) is an integer between 0 and \( N \). Let \( \tau = \min(k > 0 : x_k \notin [1, N-1]) \). Compute \( P_x(x_\tau = N) \) (you may use Matlab for your solution).

Hint: Observe that one can obtain a recursion as \( P_x(x_\tau = N) = pP_{x+1}(x_\tau = N) + (1-p)P_{x-1}(x_\tau = N) \) for \( 1 \leq x \leq N-1 \) with boundary value conditions \( P_N(x_\tau = N) = 1 \) and \( P_0(x_\tau = N) = 0 \). One observes that

\[ P_{x+1}(x_\tau = N) - P_x(x_\tau = N) = \frac{1-p}{p} \left( P_x(x_\tau = N) - P_{x-1}(x_\tau = N) \right) \]

and in particular

\[ P_N(x_\tau = N) - P_{N-1}(x_\tau = N) = \left( \frac{1-p}{p} \right)^{N-1} \left( P_1(x_\tau = N) - P_0(x_\tau = N) \right) \]
Martingales and Foster-Lyapunov Criteria for Stabilization of Markov Chains

4.1 Martingales

In this chapter, we first discuss some martingale theorems. Only a few of these will be important within the scope of our coverage, some others are presented for the sake of completeness.

These are very important for us to understand stabilization of controlled stochastic systems. These also will pave the way to optimization of dynamical systems. The second half of this chapter is on the stabilization of Markov Chains.

4.1.1 More on expectations and conditional probability

Let \((\Omega, \mathcal{F}, P)\) be a probability space and let \(\mathcal{G}\) be a subset of \(\mathcal{F}\) which is itself a \(\sigma\)-field (such a collection is said to be a sub-\(\sigma\)-field of \(\mathcal{F}\)). Let \(X\) be a \(\mathbb{R}\)-valued random variable measurable with respect to \((\Omega, \mathcal{F})\) with a finite absolute expectation that is

\[
E[|X|] = \int_{\Omega} |X(\omega)| P(d\omega) < \infty,
\]

where \(\omega \in \Omega\). We call such random variables integrable.

We say that \(\Xi\) is the conditional expectation random variable (and is also called a version of the conditional expectation) \(E[X|\mathcal{G}]\), of \(X\) given \(\mathcal{G}\) if

1. \(\Xi\) is \(\mathcal{G}\)-measurable.
2. For every \(A \in \mathcal{G}\),

\[
E[1_A \Xi] = E[1_A X],
\]

where

\[
E[1_A \Xi] = \int_{\Omega} X(\omega) 1_{\omega \in A} P(d\omega)
\]

For example, if the information that we know about a process is whether an event \(A \in \mathcal{F}\) happened or not, then:

\[
X_A := E[X|A] = \frac{1}{P(A)} \int_A P(d\omega) X(\omega).
\]

If the information we have is that \(A\) did not take place:

\[
X_{A^c} := E[X|A^c] = \frac{1}{P(\Omega \setminus A)} \int_{\Omega \setminus A} P(d\omega) X(\omega).
\]

Thus, the conditional expectation given by the sigma-field generated by \(A\) (which is \(\mathcal{F}_A = \{\emptyset, \Omega, A, \Omega \setminus A\}\)) is given by:
It follows from the above that conditional probability can be expressed as

\[ P(A|G) = E[1_{\{x \in A\}}|G], \]

hence, conditional probability is a special case of conditional expectation.

The notion of conditional expectation is key for the development of stochastic processes which evolve according to a transition kernel. This is useful for optimal decision making when a partial information is available with regard to a random variable.

The following discussion is optional until the next subsection.

**Theorem 4.1.1 (Radon-Nikodym)** Let \( \mu \) and \( \nu \) be two \( \sigma \)-finite positive measures on \((\Omega, \mathcal{F})\) such that \( \nu(A) = 0 \) implies that \( \mu(A) = 0 \) (that is \( \mu \) is absolutely continuous with respect to \( \nu \)). Then, there exists a measurable function \( f : \Omega \to \mathbb{R}_+ \) such that for every \( A \):

\[ \mu(A) = \int_A f(\omega) \nu(d\omega) \]

The representation above is unique, up to points of measure zero. With the above discussion, the conditional expectation \( X = E[X|\mathcal{F}] \) exists for any sub-\( \sigma \)-field, subset of \( \mathcal{F} \), as the following discussion shows. Let \( X \) be an integrable non-negative random variable and observe that for any Borel \( A \in \mathcal{F}' \)

\[ \int_A \left( E[X|\mathcal{F}'](\omega) \right) P(d\omega) = \int_A X(\omega) P(d\omega). \]

We may view \( \zeta(A) := \int_A X(\omega) P(d\omega) \) as a measure which is absolutely continuous with respect to \( P \), and thus, \( \left( E[X|\mathcal{F}'](\omega) \right) \), is the Radon-Nikodym derivative of this measure with respect to \( P \) (This discussion extends to arbitrary integrable variables by considering the negative valued portion of the variable separately).

In case \( X \) is a random variable which is of second-order, another way to establish existence is through a Hilbert theoretic approach, by viewing the conditional expectation as the projection of \( X \) onto a subspace consisting of the set of all functions measurable on \( \mathcal{F}' \). We will revisit this later in the notes while deriving the Kalman Filter in Chapter 6. However, for this we would require that \( X \) to be square-integrable (that is, a second-order random variable).

It is a useful exercise to now consider the \( \sigma \)-field generated by an observation variable, and what a conditional expectation means in this case.

**Theorem 4.1.2** Let \( X \) be a \( \mathbb{K} \) valued random variable, where \( \mathbb{K} \) is a complete, separable, metric space and \( Y \) be another \( \mathbb{Y} \) valued random variable, Then, \( X \) is \( \mathcal{F}_Y \) (the \( \sigma \)-field generated by \( Y \)) measurable if and only if there exists a measurable function \( f : \mathbb{Y} \to \mathbb{K} \) such that \( X = f(Y(\omega)) \).

With the above, the expectation \( E[X|Y = y_0] \) can be defined, this expectation is a measurable function of \( Y \).

### 4.1.2 Some properties of conditional expectation:

One very important property is given by the following.

**Iterated expectations:**

**Theorem 4.1.3** If \( \mathcal{H} \subset \mathcal{G} \subset \mathcal{F} \), and \( X \) is \( \mathcal{F} \)-measurable and integrable, then it follows that:

\[ E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}] \]
Thus, the two conditional expectations are equal and for all $A$ the proof is complete by noting that any integrable expectations are the same.\hfill\diamondsuit

**Theorem 4.1.4** Let $G \subseteq F$, and $Y$ be $G$–measurable. Let $X$ be $F$–measurable and $XY$ be integrable. Then, $P$ almost surely

$$E[XY|G] = YE[X|G]$$

**Proof:** First assume that $Y = Y_n$ is a simple function (a simple random variable of the form: $Y_n(\omega) = \sum_{i=1}^{n} a_i 1_{\omega \in A_i}$). Let us call $E[X|G] = \eta$ and call $E[XY|G] = \zeta$.

Then, for all $A \in G$

$$\int_{A} Y_n(\omega)P(d\omega) = \int_{A} \sum_{i=1}^{n} a_i 1_{\omega \in A_i} \eta(\omega)P(d\omega) = \sum_{i=1}^{n} a_i \int_{A \cap A_i} \eta(\omega)P(d\omega) = \sum_{i=1}^{n} a_i \int_{A \cap A_i} X(\omega)P(d\omega)$$

(4.1)

Here, the last equality in (4.1) holds since $A \cap A_i \in G$. On the other hand,

$$\int_{A} \zeta(\omega)P(d\omega) = \int_{A} X(\omega)Y_n(\omega)P(d\omega) = \int_{A} \sum_{i=1}^{n} a_i 1_{\omega \in A_i} X(\omega)P(d\omega) = \sum_{i=1}^{n} a_i \int_{A \cap A_i} X(\omega)P(d\omega)$$

Thus, the two conditional expectations are equal and for all $A \in G$

$$\int_{A} E[XY_n|G](\omega)P(d\omega) = \int_{A} Y_nE[X|G](\omega)P(d\omega) = \int_{A} Y_nX(\omega)P(d\omega)$$

(4.2)

Now, the proof is complete by noting that any integrable $Y$ can be approached from below monotonically by a sequence of simple functions measurable on $G$. The monotone convergence theorem leads to the desired result.\hfill\diamondsuit

### 4.1.3 Discrete-time martingales

Let $(\Omega, F, P)$ be a probability space. An increasing family $\{F_n\}$ of sub-$\sigma$–fields of $F$ is called a filtration.

A sequence of random variables on $(\Omega, F, P)$ is said to be adapted to $F_n$ if $X_n$ is $F_n$-measurable, that is $X_n^{-1}(D) = \{w \in \Omega : X_n(w) \in D\} \in F_n$ for all Borel $D$. This holds for example if $F_n = \sigma(X_m, m \leq n), n \geq 0$.

Given a filtration $F_n$ and a sequence of real random variables adapted to it, $(X_n, F_n)$ is said to be a martingale if

$$E[|X_n|] < \infty$$

and

$$E[X_{n+1}|F_n] = X_n.$$ 

We will occasionally take the sigma-fields to be $F_n = \sigma(X_1, X_2, \ldots, X_n)$.

Let $n > m \in \mathbb{Z}_+$. Since $F_m \subseteq F_n$, it must be that $A \in F_m$ should also be in $F_n$. Thus, if $X_n$ is a martingale sequence,

$$E[1_A X_n] = E[1_A X_{n-1}] = \cdots = E[1_A X_m].$$
Thus, $E[X_n|F_m] = X_m$.

If we have that

$$E[X_n|F_m] \geq X_m$$

then $\{X_n\}$ is called a submartingale.

And, if

$$E[X_n|F_m] \leq X_m$$

then $\{X_n\}$ is called a supermartingale.

A useful concept related to filtrations is that of a stopping time, which we discussed while studying Markov chains. A stopping time is a random time, whose occurrence is measurable with respect to the filtration in the sense that for each $n \in \mathbb{N}$, $\{\tau \leq n\} \in F_n$.

Definition 4.1.1 (Filtration up to a stopping time) Let $F_t$ denote a filtration and $\tau$ be a stopping time with respect to this filtration so that for every $k$, $\{\tau \leq k\} \in F_k$. Then, the $\sigma$-field of events up to $\tau$, $F_\tau$, is the collection of all events $A \in F$ that satisfies:

$$A \cap \{\tau \leq t\} \in F_t \quad \forall t \in \mathbb{Z}^+.$$ 

4.1.4 Doob’s optional sampling theorem

Theorem 4.1.5 Suppose $(X_n, F_n)$ is a martingale sequence, and $\rho, \tau < n$ are (uniformly) bounded stopping times with $\rho \leq \tau$. Then,

$$E[X_\tau|F_\rho] = X_\rho$$

Proof: We observe that

$$E[X_\tau - X_\rho|F_\rho] = E\left[\sum_{k=\rho}^{\tau-1} X_{k+1} - X_k|F_\rho\right]$$

$$= E\left[\sum_{k=\rho}^{\infty} 1_{\{\tau > k\}} (X_{k+1} - X_k)|F_\rho\right]$$

$$= E\left[\sum_{k=\rho}^{n} 1_{\{\tau > k\}} (X_{k+1} - X_k)|F_\rho\right]$$

$$= E\left[\sum_{k=0}^{\tau-1} 1_{\{\tau > k\geq \rho\}} (X_{k+1} - X_k)|F_\rho\right]$$

$$= E\left[\sum_{k=0}^{n} E[1_{\{\tau > k\geq \rho\}} (X_{k+1} - X_k)|F_k]|F_\rho\right]$$

$$= E\left[\sum_{k=0}^{n} 1_{\{\tau > k\geq \rho\}} E[(X_{k+1} - X_k)|F_k]|F_\rho\right]$$

$$= E\left[\sum_{k=\rho}^{\tau-1} 0|F_\rho\right] = 0$$

Here, we invoke Theorem 4.1.3 and Theorem 4.1.4 since $1_{\{\tau > k\geq \rho\}}$ is $F_k$-measurable.

The statement of the theorem leads to inequalities for supermartingales or submartingales with the appropriate inequality signs.

In the above, the main properties we used were (i) the fact that the sub-fields are nested, (ii) $n$ is bounded so that the random variable $\sum_{k=\rho}^{\infty} 1_{\{\tau > k\}} (X_{k+1} - X_k)$ is integrable. Let us try to see why boundedness of the times are important: Consider
the following game. Suppose that one draws a fair coin; with equal probabilities of heads and tails. If we have a tail, we win a dollar, and a head will let us lose a dollar. Suppose we have 0 dollars at time 0 and we decide to stop when we have 5 dollars, that is at time $\tau = \min(n > 0 : X_n = 5)$. In this case, clearly $E[X_\tau] = 5$, as we will stop when we have 5 dollars. But $E[X_\tau] \neq X_0$!

For this example, $X_n = X_{n-1} + W_n$, where $W_n$ is either $-1$ or $1$ with equal probabilities and $X_n$ is the amount of money we have. Clearly $X_n$ is a martingale sequence for all finite $n \in \mathbb{Z}_+$. The problem is that one might have to wait for an arbitrarily long period of amount of time to be able to have the 5 dollars, the sequence $E[|X_n|]$ is not uniformly bounded, and $\sum_{k=\rho}^{\infty} 1_{\{\tau > k\}} (X_{k+1} - X_k)$ is not (uniformly) integrable, and the proof method adopted in Theorem 4.1.5 will not be applicable. Note that if we were able to claim that

$$\lim_{n \to \infty} E[\sum_{k=\rho}^{n} 1_{\{\tau > k\}} (X_{k+1} - X_k)|F_\rho] = E[\lim_{n \to \infty} \left( \sum_{k=\rho}^{n} 1_{\{\tau > k\}} (X_{k+1} - X_k) \right)|F_\rho] = E[X_\tau - X_\rho|F_\rho],$$

then the result would be applicable even if we didn’t have a finite upper bound on the stopping times. This requires in particular a convergence result on the martingale sequence and the finiteness of $\tau$. More on this will be discussed below.

### 4.1.5 An important martingale convergence theorem

We first discuss Doob’s upcrossing lemma. Let $(a, b)$ be a non-empty interval. Let $X_0 \in (a, b)$. Define a sequence of stopping times

$$T_1 = \min\{N; \min(0 \leq n \leq N, X_n \leq a)\} \quad T_2 = \min\{N; \min(T_1 \leq n \leq N, X_n \geq b)\}$$

$$T_3 = \min\{N; \min(T_2 \leq n \leq N, X_n \leq a)\} \quad T_4 = \min\{N; \min(T_3 \leq n \leq N, X_n \geq b)\}$$

and for $m \geq 1$:

$$T_{2m-1} = \min\{N; \min(T_{2m-2} \leq n \leq N, X_n \leq a)\} \quad T_{2m} = \min\{N; \min(T_{2m-1} \leq n \leq N, X_n \geq b)\}$$

The number of upcrossings of $(a, b)$ up to time $N$ is the random variable $\zeta_N(a, b) = \max I\{X_n \in (a, b)\}$ between 0 and $N$. The process $(X_n)$ crosses the strip $(a, b)$ from below $a$ to above $b$.

Note that $X_{T_2} - X_{T_1}$ has the expectation zero, if the sequence is a martingale!

**Theorem 4.1.6** Let $X_T$ be a supermartingale sequence. Then,

$$E[\zeta_N(a, b)] \leq E[\max(0, a - X_N)] \leq E[|X_N|] + |a|.$$

**Proof:**

There are three possibilities that might take place: The process can end below $a$, between $a$ and $b$ and above $b$. If it crosses above $b$, then we have completed an upcrossing. In view of this, we may proceed as follows: Let $\beta_N := \min(m : T_{2m} = N \text{ or } T_{2m-1} = N)$ (note that if $T_{2m-1} = N$, $T_{2m} = N$ as well). Here, $\beta_N$ is measurable on $\sigma(X_1, X_2, \ldots, X_N)$, so it is a stopping time. Then, by the supermartingale property

$$0 \geq E[\sum_{i=1}^{\beta_N} X_{T_{2i}} - X_{T_{2i-1}}] = E[\left( \sum_{i=1}^{\beta_N} X_{T_{2i}} - X_{T_{2i-1}} \right) 1_{\{T_{2\beta_N-1} \neq N\}} 1_{\{T_{2\beta_N} = N\}}] + E[\left( \sum_{i=1}^{\beta_N} X_{T_{2i}} - X_{T_{2i-1}} \right) 1_{\{T_{2\beta_N-1} = N\}} 1_{\{T_{2\beta_N} = N\}}]$$
Thus, for every fixed \(a, b\), by the monotone convergence theorem it follows that

\[
\begin{align*}
\mathbb{E} \left( \sum_{i=1}^{\beta_N-1} (X_{T_{2i}} - X_{T_{2i-1}}) + X_N - X_{T_{2\beta_N-1}} \right) &\leq \mathbb{E} \left( \sum_{i=1}^{\beta_N-1} (X_{T_{2i}} - X_{T_{2i-1}}) + (X_N - X_{T_{2\beta_N-1}}) \right) \\
&= \mathbb{E} \left( \sum_{i=1}^{\beta_N-1} X_{T_{2i}} - X_{T_{2i-1}} \right) + \mathbb{E} \left( (X_N - X_{T_{2\beta_N-1}}) \mathbb{1}_{\{T_{2\beta_N-1} \neq N\}} \mathbb{1}_{\{T_{2\beta_N} = N\}} \right)
\end{align*}
\]

Thus,

\[
\begin{align*}
\mathbb{E} \left( \sum_{i=1}^{\beta_N-1} X_{T_{2i}} - X_{T_{2i-1}} \right) &\leq -\mathbb{E} \left( (X_N - X_{T_{2\beta_N-1}}) \mathbb{1}_{\{T_{2\beta_N-1} \neq N\}} \mathbb{1}_{\{T_{2\beta_N} = N\}} \right) \\
&\leq \mathbb{E} \left( \max(0, a - X_N) \right) \mathbb{1}_{\{T_{2\beta_N} = N\}} \leq \mathbb{E} \left( \max(0, a - X_N) \right)
\end{align*}
\]

Since, \(\sum_{i=1}^{\beta_N-1} X_{T_{2i}} - X_{T_{2i-1}} \geq \mathbb{E} [\beta_N - 1] (b - a)\), it follows that \(\zeta_N(a, b) = (\beta_N - 1)\) satisfies:

\[
\mathbb{E} [\zeta_N](b - a) \leq \mathbb{E} \left[ \max(0, a - X_N) \right] \leq |a| + \mathbb{E}[|X_N|],
\]

and the result follows.

Recall that a sequence of random variables \(X_n\), defined on a probability space \((\Omega, \mathcal{F}, P)\) converges to \(X\) almost surely (a.s.) if

\[
P \left( w : \lim_{n \to \infty} X_n(w) = X(w) \right) = 1.
\]

**Theorem 4.1.7** Suppose \(X_n\) is a supermartingale and \(\sup_{n \geq 0} \mathbb{E}[\max(0, -X_n)] < \infty\). Then \(\lim_{n \to \infty} X_n = X\) exists (almost surely). The same result applies for submartingales, by considering \(-X_n\) regarded as a supermartingale and the condition \(\sup_{n \geq 0} \mathbb{E}[\max(0, X_n)] < \infty\). A sufficient condition for both cases is that

\[
\sup_{n \geq 0} \mathbb{E}[|X_n|] < \infty.
\]

**Proof:** The proof follows from Doob’s upcrossing lemma. Now, for any fixed \(a, b\) (independent of \(\omega\)); by the upcrossing lemma we have that

\[
\mathbb{E}[\zeta_N(a, b)] \leq \mathbb{E}[\max(0, a - X_N)] \leq \frac{\mathbb{E}[|X_N|] + |a|}{(b - a)},
\]

which is uniformly bounded. The above holds for every \(N\). Since \(\zeta_N(a, b)\) is a monotonically increasing sequence in \(N\), by the monotone convergence theorem it follows that

\[
\lim_{N \to \infty} \mathbb{E}[\zeta_N(a, b)] = \mathbb{E}[\lim_{N \to \infty} \zeta_N(a, b)] < \infty.
\]

Thus, for every fixed \(a, b\), the number of up-crossings has to be finite. Hence, the limsup cannot be above \(b\) and the liminf cannot be below \(a\), for otherwise the number of up-crossings would be infinite. It then follows that

\[
P(\omega : |\limsup X_n(\omega) - \liminf X_n(\omega)| > (b - a)) = 0,
\]

since this probability can be expressed also as

\[
P \left( \bigcup_{r \in \mathbb{Q}} \left\{ \omega : \limsup X_n(\omega) > (b - a + r), \liminf X_n(\omega) < r \right\} \right),
\]
By a union bound argument, the probability is upper bounded by a countable sum of zero probability events. Finally, a continuity of probability argument then leads to

\[ P(\omega : |\limsup_n X_n(\omega) - \liminf_n X_n(\omega)| > 0) = 0. \]

We can also show that the limit variable has finite absolute expectation.

**Theorem 4.1.8 (Submartingale Convergence Theorem)** Suppose \( X_n \) is a submartingale and \( \sup_{n \geq 0} E[|X_n|] < \infty \). Then \( X := \lim_{n \to \infty} X_n \) exists (almost surely) and \( E[|X|] < \infty \).

**Proof:** Note that, \( \sup_{n \geq 0} E[|X_n|] < \infty \), is a sufficient condition both for a submartingale and a supermartingale in Theorem 4.1.7. Hence \( X_n \to X \) almost surely. For finiteness, suppose \( E[|X|] = \infty \). By Fatou’s lemma,

\[ \limsup_n E[|X_n|] \geq E[\liminf_n |X_n|] = \infty. \]

But this is a contradiction as we had assumed that \( \sup_n E[|X_n|] < \infty \).

4.1.6 Proof of Birkhoff’s individual ergodic theorem

This will be discussed in class.

4.1.7 This section is optional: Further martingale theorems

This section is optional. If you wish not to read it, please proceed to the discussion on stabilization of Markov Chains.

**Theorem 4.1.9** Let \( X_n \) be a martingale such that \( X_n \) converges to \( X \) in \( L_1 \) that is \( E[|X_n - X|] \to 0 \). Then,

\[ X_n = E[X|\mathcal{F}_n], \quad n \in \mathbb{N} \]

We will use the following while studying the convex analytic method, as well as on the stabilization of Markov chains while extending the optional sampling theorem to situations where the sampling (stopping) time is not bounded from above by a finite number. Let us define uniform integrability:

**Definition 4.1.2** : A sequence of random variables \( \{X_n\} \) is uniformly integrable if

\[ \lim_{K \to \infty} \sup_n \int_{|X_n| \geq K} |X_n|P(dX_n) = 0 \]

This implies that

\[ \sup_{n} E[|X_n|] < \infty \]

Let for some \( \epsilon > 0 \),

\[ \sup_{n} E[|X_n|^{1+\epsilon}] < \infty. \]

This implies that the sequence is uniformly integrable as

\[ \sup_{n} \int_{|X_n| \geq K} |X_n|P(dX_n) \leq \sup_{n} \int_{|X_n| \geq K} \frac{|X_n|^\epsilon}{K^\epsilon}|X_n|P(dX_n) \leq \sup_{n} \frac{1}{K^\epsilon} E[|X_n|^{1+\epsilon}] \to_{K \to \infty} 0. \]

The following is a very important result:
Theorem 4.1.10 If $X_n$ is a uniformly integrable martingale, then $X = \lim_{n \to \infty} X_n$ exists almost surely (for all sequences with probability 1) and in $L_1$, and $X_n = E[X|\mathcal{F}_n]$.

Optional Sampling Theorem For Uniformly Integrable Martingales

Theorem 4.1.11 Let $(X_n, \mathcal{F}_n)$ be a uniformly integrable martingale sequence, and $\rho, \tau$ are finite stopping times with $\rho \leq \tau$. Then,

$$E[X_{\tau}|\mathcal{F}_{\rho}] = X_{\rho}.$$

Proof: By Uniform Integrability, it follows that $\{X_t\}$ has a limit. Let this limit be $X_{\infty}$. It follows that $E[X_{\infty}|\mathcal{F}_{\tau}] = X_{\tau}$ and

$$E[E[X_{\infty}|\mathcal{F}_{\tau}]|\mathcal{F}_{\rho}] = X_{\rho},$$

which is also equal to $E[X_{\tau}|\mathcal{F}_{\rho}] = X_{\rho}$. ⋄

In the preceding result, due to uniform integrability, notice that we do not require the stopping times to be bounded by a deterministic constant.

4.1.8 Azuma-Hoeffding inequality for martingales with bounded increments

The following is an important concentration result:

Theorem 4.1.12 Let $X_t$ be a martingale sequence such that $|X_t - X_{t-1}| \leq c$ for every $t$, almost surely. Then for any $x > 0$,

$$P\left(\frac{X_t - X_0}{t} \geq x\right) \leq 2e^{-\frac{tx^2}{2c}}$$

As a result, $\frac{X_t}{t} \to 0$ almost surely.

4.2 Stability of Markov Chains: Foster-Lyapunov Techniques

A Markov chain’s stability can be characterized by drift conditions, as we discuss below in detail.

4.2.1 Criterion for positive Harris recurrence

Theorem 4.2.1 (Foster-Lyapunov for Positive Recurrence) [82] Let $S$ be a petite set, $b \in \mathbb{R}$ and $V : X \to \mathbb{R}_+$. Let $\{x_n\}$ be a $\psi$-irreducible Markov chain on $X$. If the following is satisfied for all $x \in X$:

$$E[V(x_{t+1})|x_t = x] = \int_X P(x, dy)V(y) \leq V(x) - 1 + b1_{\{x \in S\}},$$

then the chain is positive Harris recurrent (and thus a unique invariant probability measure $\pi$ exists for the Markov chain).

Proof: We will first assume that $S$ is such that $\sup_{x \in S} V(x) < \infty$. Define $\tilde{M}_0 := V(x_0)$, and for $t \geq 1$

$$\tilde{M}_t := V(x_t) - \sum_{i=0}^{t-1} (-1 + b1_{\{x_i \in S\}}).$$

It follows that
It follows from (4.6) that $E[|M_t|] \leq \infty$ for all $t$ and thus, $\{M_t\}$ is a supermartingale. Now, define a stopping time: $\tau^N = \min(\tau, \min(k > 0 : V(x_k) + k \geq N))$, where $\tau = \min\{i > 0 : x_i \in S\}$. Note that the stopping time $\tau^N$ is bounded. Hence, we have, by the martingale optional sampling theorem:

$$E[M_{\tau^N} | x_0] \leq M_0.$$ 

Hence, we obtain

$$E_x[\sum_{i=0}^{\tau^N - 1} V(x_0) + bE_x[\sum_{i=0}^{\tau^N - 1} 1_{\{x_i \in S\}}]]$$

Thus, $E_x[\tau^N - 1 + 1] \leq V(x_0) + b$, and by the monotone convergence theorem,

$$\lim_{N \to \infty} E_x[\tau^N] = E_x[\tau] \leq V(x_0) + b$$

Now, if we had that

$$\sup_{x \in S} V(x) < \infty,$$ 

the proof would be complete in view of Theorem 3.2.7 (and the fact that $E_x[\tau] \leq V(x) + b < \infty$ for any $x \in \mathbb{X}$, leading to the Harris recurrence of $S$ since this implies that $P_x(\tau = \infty) = 1$ for every $x$) and thus the chain would be positive Harris recurrent.

Typically, condition (4.7) is satisfied. However, in case it is not easy to directly verify, we may need additional steps to construct a petite set. In the following, we will consider this: Following [82], Chapter 11, define for some $l \in \mathbb{Z}_+$

$$V_C(l) = \{x \in C : V(x) \leq l\}$$

We will show that $B := V_C(l)$ is itself a petite set which is recurrent and satisfies the uniform finite-mean-return property. Since $C$ is petite for some measure $\nu$, we have that

$$K_a(x, B) \geq 1_{\{x \in C\}} \nu(B), \quad x \in \mathbb{X},$$

where $K_a(x, B) = \sum_i a(i) P_i(x, B)$ and hence

$$1_{\{x \in C\}} \leq \frac{1}{\nu(B)} K_a(x, B)$$

Now, for $x \in B$,

$$E_{\tau_B} \leq V(x) + bE_{\tau_B} \left[ \sum_{k=0}^{\tau_B - 1} 1_{\{x_k \in C\}} \right] \leq V(x) + bE_{\tau_B} \left[ \sum_{k=0}^{\tau_B - 1} \frac{1}{\nu(B)} K_a(x_k, B) \right]$$

$$= V(x) + b\frac{1}{\nu(B)} E_{\tau_B} \left[ \sum_{k=0}^{\tau_B - 1} K_a(x_k, B) \right] = V(x) + b\frac{1}{\nu(B)} E_{\tau_B} \left[ \sum_{k=0}^{\tau_B - 1} \sum_i a(i) P_i(x_k, B) \right]$$

$$= V(x) + b\frac{1}{\nu(B)} \sum_i a(i) E_{\tau_B} \left[ \sum_{k=0}^{\tau_B - 1} 1_{\{x_{k+1} \in B\}} \right]$$

$$\leq V(x) + b\frac{1}{\nu(B)} \sum_i a(i)(1 + i),$$

where (4.10) follows since at most once the process can hit $B$ between 0 and $\tau_B - 1$. Now, the petiteness measure can be adjusted such that $\sum_i a_i < \infty$ (by Proposition 5.5.6 of [82]), leading to the result that
Finally, since $C$ is petite, so is $B$ and it can be shown that $P_x(\tau_B < \infty) = 1$ for all $x \in \mathbb{X}$. This concludes the proof.

**Remark 4.1.** We could relax the irreducibility of the Markov chain, as discussed in Remark 3.3 building on [37], the drift criterion and the small/petite nature of the set leads to an irreducible Markov chain taking values in a proper subset of $\mathbb{X}$.

**Remark 4.2.** Meyn and Tweedie [82, Theorem 13.0.1] show that under the hypotheses of Theorem 4.14, together with aperiodicity, it also follows that for any initial state $x \in \mathbb{X}$,

$$\lim_{n \to \infty} \sup_{B \in \mathfrak{B}(\mathbb{X})} |P^n(x, B) - \pi(B)| = 0,$$

that is $P^n(x, \ldots)$ converges to $\pi$ in total variation, for every $x \in \mathbb{X}$.

**Remark 4.3.** We note that if $x_t$ is aperiodic and irreducible and such that for some small set $\sup_{x \in A} E[\min(t > 0 : x_t \in A)|x_0 = x] < \infty$, then the sampled chain $\{x_{km}\}$ is such that $\sup_{x \in A} E[\min(km > 0 : x_{km} \in A)|x_0 = x] < \infty$, and the split chain discussion in Section 3.2.1 applies. The argument for this builds on the fact that, with $\sigma_C = \min(k \geq 0 : x_k \in C)$, $V(x) := 1 + E_x[\sigma_C]$, it follows that $E[V(x_{t+1})|x_t = x] \leq V(x) - 1 + b1_{x \in C}$ and iterating the expectation $m$ times we obtain that

$$E[V(x_{t+m})|x_t = x] \leq V(x) - mc + bE_x[\sum_{k=0}^{m-1} 1_{x_k \in C}].$$

By [52], it follows that $E_x[\sum_{k=0}^{m-1} 1_{x_k \in C}] \leq m1_{x \in C} + \epsilon$ for some petite set $C_\epsilon$ and $\epsilon > 0$. This set is petite also for the sampled chain (see Lemma 4.2.1). As a result, we have a drift condition for the $m$-skeleton, the return time for an artificial atom constructed through the split chain is finite and hence an invariant probability measure for the $m$-skeleton, and thus by (4.20), an invariant probability measure for the original chain exists.

There are other versions of Foster-Lyapunov criteria.

### 4.2.2 Criterion for finite expectations

**Theorem 4.2.2 [Comparison Theorem]** Let $V : \mathbb{X} \to \mathbb{R}_+$, $f, g : \mathbb{X} \to \mathbb{R}_+$ such that $E[f(x)] < \infty$, $E[g(x)] < \infty$. Let $\{x_n\}$ be a Markov chain on $\mathbb{X}$. If the following is satisfied:

$$\int_\mathbb{X} P(x, dy)V(y) \leq V(x) - f(x) + g(x), \quad \forall x \in \mathbb{X},$$

then, for any stopping time $\tau$ with $P(\tau < \infty) = 1$, it follows that

$$E[\sum_{t=0}^{\tau-1} f(x_t)] \leq V(x_0) + E[\sum_{t=0}^{\tau-1} g(x_t)].$$

The proof for the above follows from similar steps as that of Theorem 4.14 but this time with a careful construction of a supermartingale sequence (except for an integrability guarantee for the sequence), the optional sampling theorem, and a monotone convergence theorem argument. The above also allows for the computation of useful bounds. For example if $g(x) = b1_{x \in A}$, then one obtains that $E[\sum_{t=0}^{\tau-1} f(x_t)] \leq V(x_0) + b$. In view of the invariant measure properties, if $f(x) \geq 1$, this provides a bound on $\int \pi(dx) f(x)$.

**Theorem 4.2.3 [Criterion for finite expectations] [82] Let $S$ be a petite set, $b \in \mathbb{R}$ and $V(.) : \mathbb{X} \to \mathbb{R}_+$, $f(.) : \mathbb{X} \to [1, \infty)$. Let $\{x_n\}$ be a Markov chain on $\mathbb{X}$. If the following is satisfied:
Now, it follows that for any \( \alpha \neq 0 \), the optional sampling theorem (see Theorem 4.1.11) applies so that

\[
M \mapsto \frac{1}{T} \sum_{t=0}^{T-1} f(x_t) = \lim_{T \to \infty} E_x[f(x_t)] = \int \mu(dx)f(x) < \infty,
\]

almost surely, where \( \mu \) is the invariant probability measure on \( \mathbb{X} \).

**Theorem 4.2.4** Let (4.13) hold. Under every invariant probability measure \( \pi \), \( \int \pi(dx)f(x) \leq b \).

**Proof.** By Theorem 4.2.2, with taking \( T \) to be a deterministic stopping time,

\[
\limsup_{T} \frac{1}{T} E_x[T] \sum_{k=0}^{T} f(x_k) \leq \limsup_{T} \frac{1}{T} \left( V(x_0) + bT \right) = b.
\]

Now, suppose that \( \pi \) is any invariant probability measure. Fix \( N < \infty \), let \( f_N = \min(N, f) \), and apply Fatou’s Lemma as follows, where we use the notation \( \pi(f) = \int \pi(dx)f(x) \),

\[
\pi(f_N) = \limsup_{n \to \infty} \pi\left( \frac{1}{n} \sum_{t=0}^{n-1} P^t f_N \right) \\
\leq \pi\left( \limsup_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} P^t f_N \right) \leq b.
\]

Fatou’s Lemma is justified to obtain the first inequality, because \( f_N \) is bounded. The second inequality holds since \( f_N \leq f \). The monotone convergence theorem then gives \( \pi(f) \leq \pi(f) \).

We need to ensure that there exists an invariant probability measure, however. This is why we require that \( f : \mathbb{X} \to [1, \infty) \) and \( \pi \) be replaced with any positive number. If the existence of an invariant probability measure is known, then one may take the range of \( f \) to be \( \mathbb{R}_+ \).

### 4.2.3 Criterion for recurrence

**Theorem 4.2.5 (Foster-Lyapunov for Recurrence)** Let \( S \) be a compact set, \( b < \infty \), and \( V \) be an inf-compact functional on \( \mathbb{X} \) such that for all \( \alpha \in \mathbb{R}_+ \), \( \{ x : V(x) \leq \alpha \} \) is compact (note: this implies that \( \lim_{||x|| \to \infty} V(x) = \infty \)). Let \( \{ x_n \} \) be an irreducible Markov chain on \( \mathbb{X} \) with a positive irreducibility measure on \( S \). If the following is satisfied:

\[
\int_{\mathbb{X}} P(x, dy)V(y) \leq V(x) + b1_{x \in S}, \quad \forall x \in \mathbb{X},
\]

then, with \( \tau_S = \min(t > 0 : x_t \in S) \), \( P_x(\tau_S < \infty) = 1 \) for all \( x \in \mathbb{X} \).

**Proof:** Define two stopping times: Let \( \tau_S = \min(t > 0 : x_t \in S) \) and \( \tau_B = \min(t > 0 : x_t \in B_N) \) where \( B_N = \{ z : V(z) \geq N \} \) with \( N \geq V(x) \) where \( x_0 = x \). Note that \( V(x_t) \) is bounded until \( \tau_N := \min(\tau_S, \tau_B) \) and until this time \( E[V(x_{t+1})|F_t] \leq V(x_t) \). Note also that, due to irreducibility, \( \tau^N = \min(\tau_S, \tau_B) < \infty \) with probability 1. Define \( M_t = \min(t, \tau^N) \), which is a supermartingale sequence uniformly bounded. It follows then that a variation of the optional sampling theorem (see Theorem 4.1.11) applies so that

\[
E_x[M_{\tau^N}] = E_x[V(\min(\tau_S, \tau_B))] \leq V(x)
\]

Now, it follows that for \( x \notin (S \cup B_N) \), since when exiting into \( B_N \), the minimum value of the Lyapunov function is \( N \):
and show that the state does not come back to some set infinitely often. We state the following.

**Theorem 4.2.6** ([82], [56]) Let

\[ P_x(x_t = x) \leq \frac{V(x)}{N}. \]

for some finite positive \( M \). Hence,

\[ P_x(\tau_{BN} < \tau_S) \leq \frac{V(x)}{N}. \]

We also have that \( P(\min(\tau_S, \tau_{BN}) = \infty) = 0 \), since the chain is irreducible and it will escape any compact set in finite time. As a consequence, we have that

\[ P_x(\tau_S = \infty) \leq P(\tau_{BN} < \tau_S) \leq V(x)/N \]

and taking the limit as \( N \to \infty \), \( P_x(\tau_S = \infty) = 0 \).

\[ \diamond \]

**Remark 4.4.** If \( S \) is further petite, then once the petite set is visited, any other set with a positive measure (under the irreducibility measure, since the petiteness measure can be taken to be the maximal irreducibility measure) is visited with probability 1 infinitely often and hence the chain is Harris recurrent.

\[ \diamond \]

**Exercise 4.2.1** Show that the random walk on \( \mathbb{Z} \) is recurrent.

### 4.2.4 Criterion for transience

Criteria for transience is somewhat more difficult to establish. One convenient way is to construct a stopping time sequence and show that the state does not come back to some set infinitely often. We state the following.

**Theorem 4.2.6 ([82], [56])** Let \( V : \mathbb{X} \to \mathbb{R}_+ \). If there exists a set \( A \) such that \( E[V(x_{t+1})|x_t = x] \leq V(x) \) for all \( x \not\in A \) and \( \exists \bar{x} \not\in A \) such that \( V(\bar{x}) < \inf_{z \in A} V(z) \), then \( \{x_t\} \) is not recurrent, in the sense that \( P_x(\tau_A < \infty) \leq 1 \).

**Proof:** Let \( x = \bar{x} \). Proof now follows from observing that \( V(x) \geq \int_y V(y)P(x, dy) \geq (\inf_{z \in A} V(z))P(x, A) + \int_{y \not\in A} V(y)P(x, dy) \geq (\inf_{z \in A} V(z))P(x, A) \). It thus follows that

\[ P(\tau_A < 2) = P(x, A) \leq \frac{V(x)}{(\inf_{z \in A} V(z))}. \]

Likewise,

\[ V(\bar{x}) \geq \int_y V(y)P(\bar{x}, dy) \geq (\inf_{z \in A} V(z))P(\bar{x}, A) + \int_{y \not\in A} (\int_s V(s)P(y, ds))P(\bar{x}, dy) \geq (\inf_{z \in A} V(z))P(\bar{x}, A) + \int_{y \not\in A} P(\bar{x}, dy)((\inf_{s \in A} V(s))P(y, A) + \int_{s \not\in A} V(s)P(y, ds)) \geq (\inf_{z \in A} V(z))P(\bar{x}, A) + \int_{y \not\in A} P(\bar{x}, dy)((\inf_{s \in A} V(s))P(y, A)) = (\inf_{z \in A} V(z)) \left(P(\bar{x}, A) + \int_{y \not\in A} P(\bar{x}, dy)P(y, A)\right). \]  

(4.15)

Thus, observing that, \( P(\{\omega : \tau_{A}(\omega) < 3\}) = \int_A P(\bar{x}, dy) + \int_{y \not\in A} P(\bar{x}, dy)P(y, A) \), we observe that:

\[ P_x(\tau_{A} < 3) \leq \frac{V(\bar{x})}{(\inf_{z \in A} V(z))}. \]
4.2 Stability of Markov Chains: Foster-Lyapunov Techniques

Thus, this follows for any $n$: $P_{\bar{x}}(\tau_A < n) \leq \frac{V(\bar{x})}{\inf_{z \in A} V(z)} < 1$. Continuity of probability measures (by defining: $B_n = \{\omega : \tau_A < n\}$ and observing $B_n \subset B_{n+1}$ and that $\lim_n P(\tau_A < n) = P(\cup_n B_n) = P(\tau_A < \infty) < 1)$ now leads to $P_{\bar{x}}(\tau_A < \infty) < 1$.

Observe the difference with the inf-compactness condition leading to recurrence and the above condition, leading to non-recurrence.

We finally note that, a convenient way to verify instability or transience is to construct an appropriate martingale sequence.

4.2.5 State dependent drift criteria: Deterministic and random-time

It is also possible that, in many applications, the controllers act on a system intermittently. In this case, we have the following results [140]. These extend the deterministic state-dependent results presented in [82], [83]: Let $\tau_{z, z} \geq 0$ be a sequence of stopping times, measurable on a filtration, possibly generated by the state process.

**Theorem 4.2.7** [140] Suppose that $\{x_t\}$ is a $\varphi$-irreducible and aperiodic Markov chain. Suppose moreover that there are functions $V: \mathbb{X} \to (0, \infty)$, $\delta: \mathbb{X} \to [1, \infty)$, $f: \mathbb{X} \to [1, \infty)$, a small set $C$ on which $V$ is bounded, and a constant $b \in \mathbb{R}$, such that

$$E[V(x_{\tau_{i+1}}) \mid \mathcal{F}_{\tau_i}] \leq V(x_{\tau_i}) - \delta(x_{\tau_i}) + b1_C(x_{\tau_i})$$

$$E\left[ \sum_{k=\tau_i}^{\tau_{i+1}-1} f(x_k) \mid \mathcal{F}_{\tau_i} \right] \leq \delta(x_{\tau_i}), \quad i \geq 0. \quad (4.16)$$

Then the following hold:

(i) $\{x_t\}$ is positive Harris recurrent, with unique invariant distribution $\pi$

(ii) $\pi(f) := \int f(x) \pi(dx) < \infty$.

(iii) For any function $g$ that is bounded by $f$, in the sense that $\sup_x |g(x)| / f(x) < \infty$, we have convergence of moments in the mean, and the strong law of large numbers holds:

$$\lim_{t \to \infty} E_x [g(x_t)] = \pi(g)$$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} g(x_t) = \pi(g) \quad a.s., \quad x \in \mathbb{X}$$

By taking $f(x) = 1$ for all $x \in \mathbb{X}$, we obtain the following corollary to Theorem 4.2.7.

**Corollary 4.2.1** [140] Suppose that $\mathbb{X}$ is a $\varphi$-irreducible Markov chain. Suppose moreover that there is a function $V: \mathbb{X} \to (0, \infty)$, a petite set $C$ on which $V$ is bounded, and a constant $b \in \mathbb{R}$, such that the following hold:

$$E[V(x_{\tau_{z+1}}) \mid \mathcal{F}_{\tau_z}] \leq V(x_{\tau_z}) - 1 + b1_{\{x_{\tau_z} \in C\}}$$

$$\sup_{z \geq 0} E[\tau_{z+1} - \tau_z \mid \mathcal{F}_{\tau_z}] < \infty. \quad (4.17)$$

Then $\mathbb{X}$ is positive Harris recurrent. \quad \diamond

**More on invariant probability measures**

Without the irreducibility condition, if the chain is weak Feller, if $[4.6]$ holds with $S$ compact, then there exists at least one invariant probability measure as discussed in Section 3.3.1.
Suppose that $X$ is a Feller Markov chain, not necessarily $\varphi$-irreducible. If (4.16) holds with $C$ compact then there exists at least one invariant probability measure. Moreover, there exists $c < \infty$ such that, under any invariant probability measure $\pi$,

$$E_\pi[f(x)] = \int_X \pi(dx)f(x) \leq c. \quad (4.18)$$

### Petite sets and sampling

Unfortunately the techniques we reviewed earlier that rely on petite sets become unavailable in the random time drift setting considered in Section 4.2.5 as a petite set $C$ for $\{x_n\}$ is not necessarily petite for $\{x_{\tau_n}\}$.

**Lemma 4.2.1** [143] Suppose $\{x_t\}$ is an aperiodic and irreducible Markov chain. If there exists a sequence of stopping times $\{\tau_n\}$ independent of $\{x_t\}$, then any $C$ that is small for $\{x_t\}$ is petite for $\{x_{\tau_n}\}$.

**Proof.** Since $C$ is petite, it is small by Theorem 3.2.5 for some $m$. Let $C$ be $(m, \delta, \nu)$-small for $\{x_1\}$,

$$P^{\tau_1}(x, \cdot) = \sum_{k=1}^{\infty} P(\tau_1 = k) P^k(x, \cdot)$$

$$\geq \sum_{k=m}^{\infty} P(\tau_1 = k) \int P^m(x, dy) P^{k-m}(y, \cdot)$$

$$\geq \sum_{k=m}^{\infty} P(\tau_1 = k) \int 1_C(x) \delta \nu(dy) P^{k-m}(y, \cdot) \quad (4.19)$$

which is a well defined measure. Therefore defining $\kappa(\cdot) = \int \nu(dy) \sum_{k=m}^{\infty} P(\tau_1 = k) P^{k-m}(y, \cdot)$, we have that $C$ is $(1, \delta, \kappa)$-small for $\{x_{\tau_n}\}$. \hfill $\diamond$

Thus, one can relax the condition that $V$ is bounded on $C$ in Theorem 4.2.7 if the sampling times are deterministic. Another condition is when the sampling instances are hitting times to a set which contains $C$ [143].

### 4.3 Convergence Rates to Equilibrium

In addition to obtaining bounds on the rate of convergence through Dobrushin’s coefficient, one powerful approach is through the Foster-Lyapunov drift conditions.

Regularity and ergodicity are concepts closely related through the work of Meyn and Tweedie [85, 86] and Tuominen and Tweedie [117].

**Definition 4.3.1** A set $A \in \mathcal{B}(\mathcal{X})$ is called $(f, r)$-regular if

$$\sup_{x \in A} E_x[\sum_{k=0}^{\tau_B-1} r(k)f(x_k)] < \infty$$

for all $B \in \mathcal{B}^+(\mathcal{X})$. A finite measure $\nu$ on $\mathcal{B}(\mathcal{X})$ is called $(f, r)$-regular if

$$E_\nu[\sum_{k=0}^{\tau_B-1} r(k)f(x_k)] < \infty$$

for all $B \in \mathcal{B}^+(\mathcal{X})$, and a point $x$ is called $(f, r)$-regular if the measure $\delta_x$ is $(f, r)$-regular.
This leads to a lemma relating regular distributions to regular atoms.

**Lemma 4.5.** If a Markov chain \( \{x_t\} \) has an atom \( \alpha \in B^+(\mathcal{X}) \) and an \((f, r)\)-regular distribution \( \lambda \), then \( \alpha \) is an \((f, r)\)-regular set.

**Definition 4.3.2 \((f\text{-norm})\)** For a function \( f: \mathcal{X} \to [1, \infty) \) the \( f \)-norm of a measure \( \mu \) defined on \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\) is given by

\[
\|\mu\|_f = \sup_{g \leq f} \int \mu(dx)g(x).
\]

The total variation norm is the \( f \)-norm when \( f = 1 \), denoted by \( \|\cdot\|_{TV} \).

**Definition 4.3.3** A Markov chain \( \{x_t\} \) with invariant distribution \( \pi \) is \((f, r)\)-ergodic if

\[
r(n)\|P^n(x, \cdot) - \pi(\cdot)\|_f \to 0 \quad \text{as } n \to \infty \text{ for all } x \in \mathcal{X}. \tag{4.20}
\]

If (4.20) is satisfied for a geometric \( r \) (so that \( r(n) = M\zeta^n \) for some \( \zeta > 1, M < \infty \)) and \( f = 1 \) then the Markov chain \( \{x_t\} \) is called geometrically ergodic.

**Coupling inequality and moments of return times to a small set** The main idea behind the coupling inequality is to bound the total variation distance between the distributions of two random variables by the probability they are different. Let \( X \) and \( Y \) be two jointly distributed random variables on a space \( \mathcal{X} \) with distributions \( \mu_x, \mu_y \) respectively. Then we can bound the total variation between the distributions by the probability the two variables are not equal.

\[
\|\mu_x - \mu_y\|_{TV} = \sup_A |\mu_x(A) - \mu_y(A)|
\]

\[
= \sup_A |P(X \in A, X = Y) + P(X \in A, X \neq Y)| - |P(Y \in A, X = Y) - P(Y \in A, X \neq Y)|
\]

\[
\leq \sup_A |P(X \in A, X \neq Y) - P(Y \in A, X \neq Y)|
\]

\[
\leq P(X \neq Y)
\]

The coupling inequality is useful in discussions of ergodicity when used in conjunction with parallel Markov chains. Later, we will see that the coupling inequality is also useful to establish the existence of optimal solutions to average cost optimization problems.

One creates two Markov chains having the same one-step transition probability. Let \( \{x_n\} \) and \( \{x'_n\} \) be two Markov chains that have probability transition kernel \( P(x, \cdot) \), and let \( C \) be an \((m, \delta, \nu)\)-small set. We use the coupling construction provided by Roberts and Rosenthal [99].

Let \( x_0 = x \) and \( x'_0 \sim \pi(\cdot) \) where \( \pi(\cdot) \) is the invariant distribution of both Markov chains. 1. If \( x_n = x'_n \) then \( x_{n+1} = x'_{n+1} \sim P(x_n, \cdot) \)

2. Else, if \( (x_n, x'_n) \in C \times C \) then with probability \( \delta \), \( x_{n+m} = x'_{n+m} \sim \nu(\cdot) \) with probability \( 1 - \delta \) then independently

\[
x_{n+m} \sim \frac{1}{1-\delta} (P^m(x_n, \cdot) - \delta \nu(\cdot))
\]

\[
x'_{n+m} \sim \frac{1}{1-\delta} (P^m(x'_n, \cdot) - \delta \nu(\cdot))
\]

3. Else, independently \( x_{n+m} \sim P^m(x_n, \cdot) \) and \( x'_{n+m} \sim P^m(x'_n, \cdot) \).
The in-between states \( x_{n+1}, \ldots, x_{n+m-1}, x'_{n+1}, \ldots, x'_{n+m-1} \) are distributed conditionally given \( x_n, x_{n+m}, x'_n, x'_{n+m} \).

By the Coupling Inequality and the previous discussion with Nummelin’s Splitting technique we have \( \|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq P(x_n \neq x'_n) \).

**Remark 4.6.** Through the coupling inequality one can show that \( \pi_0 P^n \to \pi \) in total variation. Furthermore, if Theorem 4.2.3 holds, one can also show that with some further analysis if the initial condition is a fixed deterministic state Theorem 4.2.3 holds, (see Lemma 14 of [99]).

4.3.1 Rates of convergence: Geometric ergodicity

In this section, following [82] and [99], we review results stating that a strong type of ergodicity, geometric ergodicity, follows from a simple drift condition. An irreducible Markov chain is said to satisfy the univariate drift condition. We also note that the univariate drift condition allows us to assume that \( P \) is not necessarily bounded. This does not imply, however, that geometric ergodicity follows from the univariate drift condition with a small set \( C \) such that \( \|P(x, \cdot) - \pi(\cdot)\|_{TV} \leq \lambda V + b1_C \).

Using the coupling inequality, Roberts and Rosenthal [99] prove that geometric ergodicity follows from the univariate drift condition. We also note that the univariate drift condition allows us to assume that \( V \) is bounded on \( C \) without any loss (see Lemma 14 of [99]).

**Theorem 4.3.1 ([99, Theorem 9])** Suppose \( \{x_t\} \) is an aperiodic, irreducible Markov chain with invariant distribution \( \pi \). Suppose \( C \) is a \((1, \epsilon, \nu)\)-small set and \( V : \mathbb{X} \to [1, \infty) \) satisfies the univariate drift condition with constants \( \lambda \in (0, 1) \) and \( b < \infty \). Then \( \{x_t\} \) is geometrically ergodic.

That geometric ergodicity follows from the univariate drift condition with a small set \( C \) is proven by Roberts and Rosenthal by using the coupling inequality to bound the \( TV \)-norm, but an alternate proof is given by Meyn and Tweedie [82] resulting in the following theorem.

**Theorem 4.3.2 ([82, Theorem 15.0.1])** Suppose \( \{x_t\} \) is an aperiodic and irreducible Markov chain. Then the following are equivalent:

(i) \( E_x [\tau_B] < \infty \) for all \( x \in \mathbb{X}, B \in \mathbb{B}^+(\mathbb{X}) \), the invariant distribution \( \pi \) of \( \{x_t\} \) exists and there exists a petite set \( C \), constants \( \gamma < 1, M > 0 \) such that for all \( x \in C \)

\[ |P(x, C) - \pi(C)| < M \gamma^n. \]

(ii) For a petite set \( C \) and for some \( \kappa > 1 \)

\[ \sup_{x \in C} E_x [x^{\kappa C}] < \infty. \]

(iii) For a petite set \( C \), constants \( b > 0, \lambda \in (0, 1) \), and a function \( V : \mathbb{X} \to [1, \infty) \) (finite for some \( x \)) such that

\[ PV \leq \lambda V + b1_C. \]

Any of the conditions imply that there exists \( r > 1, R < \infty \) such that for any \( x \)

\[ \sum_{n=0}^{\infty} r^n \|P^n(x, \cdot) - \pi(\cdot)\|_V \leq RV(x). \]
We note that if (iii) above holds, (ii) holds for all $\kappa \in (1, \lambda^{-1})$.

We now show that under (4.21), Theorem 4.3.2 (ii) holds. If (4.21) holds, the sequence \( \{M_n\} \) is supermartingale, where
\[
M_n = \lambda^{-n} V(x_n) - \sum_{k=0}^{n-1} b_1 C(x_k) \lambda^{-(k+1)},
\]
with $M_0 = V(x_0)$. Then, with (4.21), defining $\tau^N_B = \min\{N, \tau_B\}$ for $B \in B^+(X)$ gives, by Doob’s optional sampling theorem,
\[
E_x \left[ \lambda^{-\tau^N_B} V(x_{\tau^N_B}) \right] \leq V(x) + E_x \left[ \sum_{n=0}^{\tau^N_B-1} b_1 C(x_n) \lambda^{-(n+1)} \right],
\]
for any $B \in B^+(X)$, and $N \in \mathbb{Z}_+$. Since $V$ is bounded above on $C$, we have that $C \subset \{V \leq L_1\}$ for some $L_1$ and thus,
\[
\sup_{x \in C} E_x \left[ \lambda^{-\tau^C_B} V(x_{\tau^C_B}) \right] \leq L_1 + \lambda^{-1} b.
\]
and by the monotone convergence theorem, and the fact that $V$ is bounded from below by 1 everywhere and bounded from above on $C$,
\[
\sup_{x \in C} E_x \left[ \lambda^{-\tau^c} \right] \leq L_1(L_1 + \lambda^{-1} b).
\]

4.3.2 Subgeometric ergodicity

Here, we review the class of subgeometric rate functions (see [56, Sec. 4], [36, Sec. 5], [82], [41], [117]).

Let $\Lambda_0$ be the family of functions $r : \mathbb{N} \to \mathbb{R}_{>0}$ such that
\[
r \text{ is non-decreasing, } \quad r(1) \geq 2
\]
and
\[
\frac{\log r(n)}{n} \downarrow 0 \quad \text{as } n \to \infty.
\]
The second condition implies that for all $r \in \Lambda_0$ if $n > m > 0$ then
\[
n \log r(n + m) \leq n \log r(n) + m \log r(n) \leq n \log r(n) + n \log r(m)
\]
so that
\[
r(m + n) \leq r(m) r(n) \quad \text{for all } m, n \in \mathbb{N}.
\]

The class of subgeometric rate functions $\Lambda$ defined in [117] is the class of sequences $r$ for which there exists a sequence $r_0 \in \Lambda_0$ such that
\[
0 < \liminf_{n \to \infty} \frac{r(n)}{r_0(n)} \leq \limsup_{n \to \infty} \frac{r(n)}{r_0(n)} < \infty.
\]
The main theorem we cite on subgeometric rates of convergence is due to Tuominen and Tweedie [117].

**Theorem 4.3.3 (Theorem 2.1 of [117])** Suppose that $\{x_t\}_{t \in \mathbb{N}}$ is an irreducible and aperiodic Markov chain on state space $X$ with stationary transition probabilities given by $P$. Let $f : X \to [1, \infty)$ and $r \in \Lambda$ be given. The following are equivalent:
(i) there exists a petite set \( C \in \mathcal{B}(X) \) such that

\[
\sup_{x \in C} E_x \left[ \sum_{k=0}^{\tau_C-1} r(k) f(x_k) \right] < \infty
\]

(ii) there exists a sequence \( (V_n) \) of functions \( V_n : X \to [0, \infty] \), a petite set \( C \in \mathcal{B}(X) \) and \( b \in \mathbb{R}_+ \) such that \( V_0 \) is bounded on \( C \),

\[
V_0(x) = \infty \Rightarrow V_1(x) = \infty,
\]

and

\[
P V_{n+1} \leq V_n - r(n) f + b r(n) 1_C, \quad n \in \mathbb{N}
\]

(iii) there exists an \((f, r)\)-regular set \( A \in \mathcal{B}^+(X) \).

(iv) there exists a full absorbing set \( S \) which can be covered by a countable number of \((f, r)\)-regular sets.

**Theorem 4.3.4** [117] If a Markov chain \( \{x_t\} \) satisfies Theorem 4.3.3 for \((f, r)\) then \( r(n) \| P^n(x_0, \cdot) - \pi(\cdot) \|_f \to 0 \) as \( n \) increases.

The conditions of Theorem 4.3.3 may be hard to check, especially (ii), comparing a sequence of Lyapunov functions \( \{V_k\} \) at each time step. We briefly discuss the methods of Douc et al. [41] (see also Hairer [56]) that extend the subgeometric ergodicity results and show how to construct subgeometric rates of ergodicity from a simpler drift condition. [41] assumes that there exists a function \( V : X \to [1, \infty] \), a concave monotone nondecreasing differentiable function \( \phi : [1, \infty] \to (0, \infty] \), a set \( C \in \mathcal{B}(X) \) and a constant \( b \in \mathbb{R} \) such that

\[
P V + \phi V \leq V + b 1_C. \tag{4.24}
\]

If an aperiodic and irreducible Markov chain \( \{x_t\} \) satisfies the above with a petite set \( C \), and if \( V(x_0) < \infty \), then it can be shown that \( \{x_t\} \) satisfies Theorem 4.3.3(ii). Therefore \( \{x_t\} \) has invariant distribution \( \pi \) and is \((\phi V, 1)\)-ergodic so that \( \lim_{n \to \infty} \|P^n(x, \cdot) - \pi(\cdot)\|_{\phi V} = 0 \) for all \( x \) in the set \( \{x : V(x) < \infty\} \) of \( \pi \)-measure 1. The results by Douc et al. build then on trading off \((\phi \circ V, 1)\) ergodicity for \((1, r_\phi)\)-ergodicity for some rate function \( r_\phi \), by carefully constructing the function utilizing concavity; see Propositions 2.1 and 2.5 of [41] and Theorem 4.1(3) of [56].

To achieve ergodicity with a nontrivial rate and norm one can invoke a result involving the class of pairs of ultimately non decreasing functions, defined in [41]. The class \( \mathcal{Y} \) of pairs of ultimately non decreasing functions consists of pairs \( \Psi_1, \Psi_2 : X \to [1, \infty) \) such that \( \Psi_1(x)\Psi_2(y) \leq x + y \) and \( \Psi_i(x) \to \infty \) for one of \( i = 1, 2 \).

**Proposition 4.7.** Suppose \( \{x_t\} \) is an aperiodic and irreducible Markov chain that is both \((1, r)\)-ergodic and \((f, 1)\)-ergodic for some \( r \in \Lambda \) and \( f : X \to [1, \infty) \). Suppose \( \Psi_1, \Psi_2 : X \to [1, \infty) \) are a pair of ultimately non decreasing functions. Then \( \{x_t\} \) is \((\Psi_1 \circ f, \Psi_2 \circ r)\)-ergodic.

Therefore we can show that if \( (\Psi_1, \Psi_2) \in \mathcal{Y} \) and a Markov chain satisfies the condition (4.24), then it is \((\Psi_1 \circ \phi \circ V, \Psi_2 \circ r_\phi)\)-ergodic.

Thus, we observe that the hitting times to a small set is a very important measure in characterizing not only the existence of an invariant probability measure, but also how fast a Markov chain converges to equilibrium. Further results exist in the literature to obtain more computable criteria for subgeometric rates of convergence, see e.g. [41].

**Rates of convergence under random-time state-dependent drift criteria**

The following result builds on and generalizes Theorem 2.1 in [140].
Theorem 4.3.5 Let \( \{x_t\} \) be an aperiodic and irreducible Markov chain with a small set \( C \). Suppose there are functions \( V : X \to (0, \infty) \) with \( V \) bounded on \( C \), \( f : X \to [1, \infty) \), \( \delta : X \to [1, \infty) \), a constant \( b \in \mathbb{R} \), and \( r \in \Lambda \) such that for a sequence of stopping times \( \{\tau_n\} \)

\[
E[V(x_{\tau_n+1}) | x_{\tau_n}] \leq V(x_{\tau_n}) - \delta(x_{\tau_n}) + b1C(x_{\tau_n})
\]

\[
E\left[ \sum_{k=\tau_n}^{\tau_{n+1}-1} f(x_k)r(k) \middle| \mathcal{F}_{\tau_n} \right] \leq \delta(x_{\tau_n}).
\] (4.25)

Then \( \{x_t\} \) satisfies Theorem 4.3.3 and is \((f, r)\)-ergodic.

Further conditions and examples are available in [143].

4.4 Conclusion

This concludes our discussion for controlled Markov chains via martingale methods. We will revisit one more application of martingales while discussing the convex analytic approach to controlled Markov problems. We observed that drift criteria are very powerful tools to establish various forms of stochastic stability and instability.

4.5 Exercises

Exercise 4.5.1 Let \( X \) be an integrable random variable defined on \((\Omega, \mathcal{F}, P)\). Let \( \mathcal{G} = \{\Omega, \emptyset\} \). Show that \( E[X|\mathcal{G}] = E[X] \), and if \( \mathcal{G} = \sigma(X) \) then \( E[X|\mathcal{G}] = X \).

Exercise 4.5.2 Let \( X \) be a random variable such that \( P(|X| < K) = 1 \) for some \( K \in \mathbb{R} \), that is \( X \) takes values from a compact set. Let \( Y_n = X + W_n \) for \( n \in \mathbb{Z}_+ \) and \( W_n \) be an independent noise for every \( n \).

Let \( \mathcal{F}_n \) be the \( \sigma \)-field generated by \( Y_0, Y_1, \ldots, Y_n \).

Does \( \lim_{n \to \infty} E[X|\mathcal{F}_n] \) exist almost surely? Prove your statement.

Can you replace the boundedness assumption on \( X \) with \( E[|X|] < \infty \)?

Exercise 4.5.3 a) Consider a Controlled Markov Chain with the following dynamics:

\[
x_{t+1} = ax_t + bu_t + w_t,
\]

where \( w_t \) is a zero-mean Gaussian noise with a finite variance, \( a, b \in \mathbb{R} \) are the system dynamics coefficients. One controller policy which is admissible (that is, the policy at time \( t \) is measurable with respect to \( \sigma(x_0, x_1, \ldots, x_t) \) and is a mapping to \( \mathbb{R} \)) is the following:

\[
u_t = -a + 0.5 \frac{b}{x_t}.
\]

Show that \( \{x_t\} \), under this policy, has a unique invariant probability measure.

b) Consider a similar setup to the one earlier, with \( b = 1 \):

\[
x_{t+1} = ax_t + u_t + w_t,
\]

where \( w_t \) is a zero-mean Gaussian noise with a finite variance, and \( a \in \mathbb{R} \) is a known number.

This time, suppose, we would like to find a control policy such that there exists an invariant probability measure \( \pi \) for \( \{x_t\} \) and under this invariant probability measure
Further, suppose we restrict the set of control policies to be linear, time-invariant; that is of the form \( u(x_t) = k x_t \) for some \( k \in \mathbb{R} \).

Find the set of all \( k \) values for which there exists an invariant probability measure that has a finite second moment.

**Hint:** Use Foster-Lyapunov criteria.

**Exercise 4.5.4** Suppose that some price process \( \{x_t, t \in \mathbb{Z}_+\} \) is given by the following dynamics:

\[
x_{t+1} = \max(x_t + w_t, 0), \quad t \in \mathbb{Z}_+,
\]

where \( \{w_t\} \) is a sequence of independent and identically distributed \( \{-1, 1\} \)-valued random variables with mean \( \bar{w} > 0 \). Furthermore, \( x_0 \in \mathbb{Z}_+, x_0 > 0 \) is a given initial condition for the process.

Is the price process recurrent in the sense that, \( P_{x_0}(\tau_0 < \infty) = 1 \), where \( \tau_0 = \min(l > 0 : x_l = 0) \)?

**Exercise 4.5.5** Consider a queuing process, with i.i.d. Poisson arrivals and departures, with arrival mean \( \mu \) and service mean \( \lambda \) and suppose the process is such that when a customer leaves the queue, with probability \( p \) (independent of time) it comes back to the queue. That is, the dynamics of the system satisfies:

\[
L_{t+1} = \max(L_t + A_t - N_t + p_t N_t, 0), \quad t \in \mathbb{N}.
\]

where \( E[A_t] = \lambda, E[N_t] = \mu \) and \( E[p_t] = p \).

For what values of \( \mu, \lambda \) is such a system stochastically stable? Prove your statement.

**Exercise 4.5.6** Consider a two server-station network; where a router routes the incoming traffic, as is depicted in Figure 5.7

![Fig. 4.1](image)

Let \( L_1^1, L_1^2 \) denote the number of customers in stations 1 and 2 at time \( t \). Let the dynamics be given by the following:

\[
L_{t+1}^1 = \max(L_t^1 + u_t A_t - N_t^1, 0), \quad t \in \mathbb{N}.
\]

\[
L_{t+1}^2 = \max(L_t^2 + (1 - u_t) A_t - N_t^2, 0), \quad t \in \mathbb{N}.
\]

Customers arrive according to an independent Bernoulli process, \( A_t \), with mean \( \lambda \). That is, \( P(A_t = 1) = \lambda \) and \( P(A_t = 0) = 1 - \lambda \). Here \( u_t \in [0, 1] \) is the router action.

Station 1 has a Bernoulli service process \( N^1_t \) with mean \( n_1 \), and Station 2 with \( n_2 \).

Suppose that a router decides to follow the following algorithm to decide on \( u_t \): If a customer arrives, the router simply sends the incoming customer to the shortest queue.

Find sufficient conditions (on \( \lambda, n_1, n_2 \)) for this algorithm to lead to a stochastically stable system with invariant measure \( \pi \) which satisfies \( E_\pi[L^1 + L^2] < \infty \).
Exercise 4.5.7 Let there be a single server, serving two queues; where the server serves the two queues adaptively in the following sense:

The dynamics of the two queues is expressed as follows:

\[ L_{i+1}^t = \max(L_i^t + A_i^t - N_i^t, 0), \quad i = 1, 2; \quad t \in \mathbb{N} \]

where \( L_i^t \) is the total number of arrivals which are still in the queue at time \( t \) and \( A_i^t \) is the number of customers that have just arrived at time \( t \).

We assume \( \{ A_i^t \} \) to have an independent, identical distribution (i.i.d.) which is Poisson, with mean \( \lambda_i \), that is for all \( k \in \mathbb{Z}_+ \):

\[ P(A_i^t = k) = \frac{\lambda_i^k e^{-\lambda_i}}{k!}, \quad k \in \{0, 1, 2, \ldots, \} \]

\( \{ N_i^t \} \) for \( i = 1, 2 \) is the service process. Suppose that the service process has an i.i.d. Poisson distribution, where the mean at time \( t \) depends on the number of customers such that:

\[ E[N_i^t] = \mu \frac{L_i^t}{L_i^t + L_i^t}; \]
\[ E[N_i^t] = \mu \frac{L_i^t}{L_i^t + L_i^t} \]

Thus, \( \{ N_1^t + N_2^t \} \) is Poisson with mean \( \mu \).

a) When is it that the system is stochastically stable, that is for what values of \( \lambda_1, \lambda_2, \mu \)? Here, by stochastic stability we mean both recurrence and positive (Harris) recurrence (i.e., the existence of an invariant probability measure). Please be explicit.

b) When the system is stable with an invariant probability measure \( \pi \), can you find a bound for \( E_{\pi} [L_1^t + L_2^t] \) as a function of \( \lambda_1, \lambda_2, \mu \)?

Exercise 4.5.8 Consider the following two-server system:

\[ x_{i+1}^t = \max(x_i^t + A_i^t - u_i^t, 0) \]
\[ x_{i+1}^t = \max(x_i^t + A_i^t + u_i^t(1_{[x_i^t \leq x_i^t + A_i^t]} - u_i^t), 0) \]

(4.26)

where \( 1_{(\cdot)} \) denotes the indicator function and \( A_i^t, A_i^t \) are independent and identically distributed (i.i.d.) random variables with geometric distributions, that is, for \( i = 1, 2 \),

\[ P(A_i^t = k) = p_i(1 - p_i)^k \quad k \in \{0, 1, 2, \ldots, \} \]

for some scalars \( p_1, p_2 \) such that \( E[A_i^t] = 1.5 \) and \( E[A_i^t] = 1. \)

Suppose the control actions \( u_1^t, u_2^t \) are such that \( u_1^t + u_2^t \leq 5 \) for all \( t \in \mathbb{Z}_+ \) and \( u_1^t, u_2^t \in \mathbb{Z}_+ \). At any given time \( t \), the controller has to decide on \( u_1^t \) and \( u_2^t \) with knowing \( \{ x_s^t, x_s^t, s \leq t \} \) but not knowing \( A_1^t, A_2^t \).

Is this server system stochastically stabilizable by some policy, that is, does there exist an invariant probability measure under some control policy?

If your answer is positive, provide a control policy and show that there exists a unique invariant distribution.

Exercise 4.5.9 Let there be a single server, serving two queues; where the server serves the two queues adaptively in the following sense. The dynamics of the two queues is expressed as follows:

\[ L_{i+1}^t = \max(L_i^t + A_i^t - N_i^t, 0), \quad i = 1, 2; \quad t \in \mathbb{Z}_+ \]
where $L_t^i$ is the total number of arrivals which are still in the queue at time $t$ and $A_t^i$ is the number of customers that have just arrived at time $t$.

We assume, for $i = 1, 2$, \{A_t^i\} has an independent and identical distribution (i.i.d.) which is Bernoulli so that $P(A_t^i = 1) = \lambda_i = 1 - P(A_t^i = 0)$.

Suppose that the service process is given by:

\[
N_t^1 = 1_{\{L_t^1 \geq L_t^2\}} \quad N_t^2 = 1_{\{L_t^2 > L_t^1\}}
\]

For what values of $\lambda_1, \lambda_2$ is the system stochastically stable? Here, by stochastic stability we mean the existence of an invariant probability measure.

**Exercise 4.5.10** Prove Birkhoff’s Ergodic Theorem for a countable state space; that is the result that for an irreducible Markov chain $\{x_t\}$ living in a countable space $\mathbb{X}$, which has a unique invariant distribution $\mu$, the following applies almost surely:

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} f(x_t) = \sum_i f(i)\mu(i),
\]

for every bounded $f : \mathbb{X} \to \mathbb{R}$.

**Hint:** You may proceed as follows. Define a sequence of empirical occupation measures for every bounded $f$:

\[
v_T(A) = \frac{1}{T} \sum_{t=0}^{T-1} 1_{\{x_t \in A\}}, \quad \forall A \in \mathcal{B}(\mathbb{X}).
\]

Now, define:

\[
F_t(A) = \left( \sum_{s=1}^{t} 1_{\{x_s \in A\}} - t \sum_x P(A|x) v_t(x) \right)
\]

\[
= \left( \sum_{s=1}^{t} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-1} \sum_x P(A|x) 1_{\{x_s = x\}} \right) \quad (4.27)
\]

Let $\mathcal{F}_t = \sigma(x_0, \cdots, x_t)$. Verify that, for $t \geq 2$,

\[
E[F_t(A) | \mathcal{F}_{t-1}]
\]

\[
= E \left[ \left( \sum_{s=1}^{t} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-1} \sum_x P(A|x) 1_{\{x_s = x\}} \right) | \mathcal{F}_{t-1} \right]
\]

\[
= E \left[ \left( 1_{\{x_t \in A\}} - \sum_x P(A|x) 1_{\{x_{t-1} = x\}} \right) | \mathcal{F}_{t-1} \right]
\]

\[
+ \left( \sum_{s=1}^{t-1} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-2} \sum_x P(A|x) 1_{\{x_s = x\}} \right)
\]

\[
= 0 + \left( \sum_{s=1}^{t-1} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-2} \sum_x P(A|x) 1_{\{x_s = x\}} \right) | \mathcal{F}_{t-1} \right] \quad (4.28)
\]

\[
= F_{t-1}(A), \quad (4.29)
\]

where the last equality follows from the fact that $E[1_{\{x_t \in A\}} | \mathcal{F}_{t-1}] = P(x_t \in A | \mathcal{F}_{t-1})$. Furthermore,

\[
|F_t(A) - F_{t-1}(A)| \leq 1.
\]
Now, we have a sequence which is a martingale sequence. We will invoke a martingale convergence theorem; which is applicable for **martingales with bounded increments**. By a version of the martingale stability theorem, it follows that

\[
\lim_{t \to \infty} \frac{1}{t} F_t(A) = 0
\]

You need to now complete the remaining steps.

**Hint:** You can use the Azuma-Hoeffding inequality \([40]\) and the Borel-Cantelli Lemma to complete the steps.

**Exercise 4.5.11** Let \(\tau\) be a stopping time with respect to the filtration \(\mathcal{F}_t\). Let \(X_n\) be a (discrete-time) sequence of random variables so that each \(X_n\) is \(\mathcal{F}_n\)-measurable. Show that \(X_\tau\) is \(\mathcal{F}_\tau\)-measurable.

**Hint:** We need to show that for every real \(a\): \(\{X_\tau \leq a\} \cap \{\tau \leq k\} \in \mathcal{F}_k\). Observe that \(\{X_\tau \leq a\} \cap \{\tau \leq k\} = \bigcup_{m=0}^{k} \{X_\tau \leq a\} \cap \{\tau = m\}\) and that for each \(m\), \(\{X_\tau \leq a\} \cap \{\tau = m\} \in \mathcal{F}_m \subset \mathcal{F}_k\).

**Exercise 4.5.12** (A convergence theorem useful in stochastic approximation) \([89]\) Let \(X_k, \beta_k, Y_k\) be three sequences of non-negative random variables defined on a common probability space and \(\mathcal{F}_k\) be a filtration so that all three random sequences are adapted to it. Suppose that

\[
E[X_{k+1}|\mathcal{F}_k] \leq (1 + \beta_k) X_k + Y_k, \quad k \in \mathbb{N}.
\]

Show that the limit \(\lim_{n \to \infty} X_n\) exists and is finite with probability one conditioned on the event that \(\sum_{k \in \mathbb{N}} \beta_n < \infty\) and \(\sum_{k \in \mathbb{N}} Y_n < \infty\).

**Hint:** Define

\[
M_n = X'_n - \sum_{m=1}^{n-1} Y'_m,
\]

where \(X'_n = \frac{X_n}{\prod_{s=1}^{n} (1 + \beta_s)}\) and \(Y'_n = \frac{Y_n}{\prod_{s=1}^{n} (1 + \beta_s)}\). Define the stopping time:

\[
\tau_n = \min(n : \sum_{m=1}^{n-1} Y'_m > a).
\]

Show first that \(a + M_{\min(\tau_n, n)}\), \(n \in \mathbb{N}\) is a positive supermartingale. Then invoke the supermartingale convergence theorem \([4.7.7]\). Finally, using the fact that \(\sum_{k \in \mathbb{N}} Y_k < \infty\) and that \(\prod_{n} (1 + \beta_n) < \infty\) under the stated conditions, complete the proof.

This theorem is important for a large class of optimization problems (such as the convergence of stochastic gradient descent algorithms) as well as stochastic approximation algorithms. For further reading on stochastic approximation methods, see \([73]\) and \([14]\).

**Exercise 4.5.13** (Another convergence theorem useful in stochastic approximation) Let \(X_k, Y_k, Z_k\) be three sequences of non-negative random variables defined on a common probability space and \(\mathcal{F}_k\) be a filtration so that all three random sequences are adapted to it. Suppose that

\[
E[Y_{k+1}|\mathcal{F}_k] \leq Y_k - X_k + Z_k
\]

(4.30) and \(\sum_{k} Z_k < \infty\). Show that then \(\sum_{k} X_k < \infty\) and \(Y_k\) converges to some random variable \(Y\) almost surely.

**Hint** \([135]\): One proof of this result follows from the previous exercise by noting first that \(E[Y_{k+1}|\mathcal{F}_k] \leq Y_k + Z_k\) with \(\beta_k = 0\). This implies that \(\sum_{k} Y_k\) is finite and that \(Y_k\) converges. Now write \(M_t = Y_t + \sum_{m=1}^{t-1} X_m\) leading to \(E[M_{t+1}|\mathcal{F}_t] \leq M_t + Z_t\). Applying the previous exercise again, it follows that \(M_t\) converges and since \(Y_t\) converges, so does \(\sum_{t} X_t\).

**Exercise 4.5.14** (Application in stochastic optimization) \(\text{This is from Prof. Jerome Le Ny (of Polytechnique Montreal)’s Blog.}\) Consider a convex function \(f : \mathbb{R}^n \to \mathbb{R}\) and denote the set of minima of \(f\) by \(X^*\). We know from convex analysis
that $X^*$ contains, if non-empty, either a single point, or is a convex set. Denote the subdifferential \( \partial f(x) \) of $f$ at $x$, that is the set of subgradients of $f$ at $x$, by $\partial f(x)$ and let $d_i$ be a random variable which is a noisy version of a subgradient of $f$ at $x_i$ at time $t$. A stochastic subgradient algorithm is one with the form:

$$x_{k+1} = x_k - \gamma_k d_{k+1}, \quad x_0 \in \mathbb{R}^n,$$

where $\gamma_k$ is a sequence of non-negative step sizes. We have the following theorem:

**Theorem 4.5.1** Suppose that the set of minima $X^*$ is non-empty and that the stochastic subgradients satisfy that

$$\sup_k E[||d_{k+1}||^2_{\mathcal{F}_k}] < K < \infty,$$

where $\mathcal{F}_k = \sigma(x_0, d_s, s \leq k)$ with the condition that

$$g_{k+1} = E[d_{k+1}|\mathcal{F}_k] \in \partial f(x).$$

Moreover, $\sum_k \gamma_k = \infty$ and $\sum_k \gamma_k^2 < \infty$. Then the sequence of iterates (4.31) converges almost surely to some element $x^* \in X^*$.

**Proof.** $y \in \mathbb{R}^n$, we have that due to the definition of a subgradient,

$$f(y) \geq f(x) + g^T_k (y - x_k).$$

Thus,

$$E[||x_{k+1} - y||^2 |\mathcal{F}_k] = E[||x_k - \gamma_k d_{k+1} - y||^2 |\mathcal{F}_k]$$

$$= E[||x_k - y||^2 - 2\gamma_k (x_k - y)^T d_{k+1} + \gamma_k^2 ||d_{k+1}||^2 |\mathcal{F}_k]$$

$$= E[||x_k - y||^2 |\mathcal{F}_k] - 2\gamma_k (x_k - y)^T E[d_{k+1}|\mathcal{F}_k] + E[\gamma_k^2 ||d_{k+1}||^2 |\mathcal{F}_k]$$

$$\leq E[||x_k - y||^2 |\mathcal{F}_k] - 2\gamma_k (f(x_k) - f(y)) + \gamma_k^2 K$$

(4.32)

Let $y = \bar{x}^* \in X^*$ for some element in $X^*$. Then one obtains through the comparison theorem (Theorem 4.2.2) that

$$E[\sum_k \gamma_k (f(x_k) - f(y))] \leq ||x_0 - y||^2 + \sum_k \gamma_k^2 K.$$ 

In particular, since $f(x_k) - f(y) \geq 0$, through the convergence theorem from the preceding exercises we have that almost surely

$$\sum_k \gamma_k (f(x_k) - f(y)) < \infty.$$ 

Thus, $f(x_k) \to f(y)$. We now show that indeed $x_k \to$ some particular element in $X^*$ (and does not wander in the set). By the convergence result in (4.30) we know that for any $x^* \in X^*$, $||x_k - x^*||$ converges almost surely. This implies that $x_k$ is bounded almost surely. Now, consider a countable dense subset \{ $x_1^*, \ldots, x_n^*, \ldots$ \} of $X^*$. It must be that $||x_k - x_i^*||$ converges for all $i$ through the convergence theorem. On the other hand, since $||x_k||$ is bounded, there exists a converging subsequence for $x_{k_n}$. But the limit of each such subsequence must be identical for otherwise $||x_{k_n} - x_i^*||$ would have different limits. Thus, $x_{k_n}$ must converge to one element in $X^*$. 

\( \diamond \)
Dynamic Programming

In this chapter, we introduce the method of *dynamic programming* for controlled stochastic systems. Let us first recall a few notions discussed earlier.

Recall that a Markov control model is a five-tuple

$$(X, U, \{U(x), x \in X\}, Q, c_t)$$

such that $X$ is the Polish state space, $U$ is the Polish control action space, $U_t(x) \subset U$ is the control action set when the state at time $t$ is $x$, so that

$$\mathcal{K}_t = \{(x, u) : x \in X, u \in U_t(x)\} \subset X \times U,$$

is the set of feasible state-action pairs. Furthermore, $Q$ is a stochastic kernel on $X$ given $\mathcal{K}_t$. Finally $c_t : \mathcal{K}_t \to \mathbb{R}$ is the cost function at time $t$. Often $c_t \equiv c$, that is $c$ does not depend on time. In case $U_t(x)$ and $c_t$ do not depend on $t$, we drop the time index. Then, the controlled Markov model is called a stationary model.

Let, as in Chapter 2, $\Pi_A$ denote the set of all admissible policies. Let $\Pi = \{\gamma_t, 0 \leq t \leq N-1\} \in \Pi_A$ be a policy. Consider the following expected cost:

$$J(x, \Pi) = E_x^H \left[ \sum_{t=0}^{N-1} c(x_t, u_t) + c_N(x_N) \right],$$

where $c_N(\cdot)$ is the terminal cost function. Define

$$J^*(x) := \inf_{\Pi \in \Pi_A} J(x, \Pi)$$

As earlier, let $h_t = \{x_{[0,t]}, u_{[0,t-1]}\}$ denote the history or the information process.

The goal is to find, if there exists one, an admissible policy such that $J^*(x)$ is attained; this will be an optimal policy. We note that the infimum value may not be attained by some policy. In the following, we will present conditions which will ensure the existence of optimal policies.

Before we proceed further, we note that we could express the cost as:

$$J(x, \Pi) = E_x^H \left[ c(x_0, u_0) \right] + E_x^H \left[ c(x_1, u_1) \right] + E_x^H \left[ c(x_2, u_2) \right] + \ldots$$
The above follows from Theorem 4.1.3. Thus, one can inductively obtain:

\[ + E^T[c(x_{N-1}, u_{N-1}) + c_N(x_N)|h_{N-1}] \bigg| h_{N-2} \ldots h_1 \bigg| h_0 \],

\[ = E_x^{\gamma_0 \cdots \gamma_{N-1}} \left[ c(x_0, u_0) \right. \right.

\[ + E^{\gamma_1 \cdots \gamma_{N-1}} \left[ c(x_1, u_1) \right. \right.

\[ + E^{\gamma_2 \cdots \gamma_{N-1}} \left[ c(x_2, u_2) \right. \right.

\[ + \ldots \]

\[ + E^{\gamma_{N-1}} \left[ c(x_{N-1}, u_{N-1}) + c_N(x_N)|h_{N-1}] \bigg| h_{N-2} \ldots h_1 \bigg| h_0 \right]. \]

The above follows from Theorem 4.1.3. Thus, one can inductively obtain:

\[ \inf_I J(x, II) \geq \inf_I E_x^{\gamma_0 \cdots \gamma_{N-1}} \left[ c(x_0, u_0) \right. \right.

\[ + \inf_{\gamma_1 \cdots \gamma_{N-1}} E^{\gamma_1 \cdots \gamma_{N-1}} \left[ c(x_1, u_1) \right. \right.

\[ + \inf_{\gamma_2 \cdots \gamma_{N-1}} E^{\gamma_2 \cdots \gamma_{N-1}} \left[ c(x_2, u_2) \right. \right.

\[ + \ldots \]

\[ + \inf_{\gamma_{N-1}} E^{\gamma_{N-1}} \left[ c(x_{N-1}, u_{N-1}) + c_N(x_N)|h_{N-1}] \bigg| h_{N-2} \ldots h_1 \bigg| h_0 \right]. \]

The discussion above reveals that we can start with the final time stage, obtain a solution for \( t \leq N - 2 \). By a theorem below which will allow us to search over Markov policies for \( \gamma_{N-1} \) and together with the fact that for all \( t \), and measurable functions \( g_t \)

\[ E[g_t(x_{t+1})|h_t, u_t] = E[g_t(x_{t+1})|x_t, u_t], \]

(this follows from the controlled Markov property), we will see that one can restrict the optimal control policies to be Markov. This last step is crucial in identifying a dependence only on the most recent state for an optimal control policy, as we see in the next section. This will allow us to show that, through an inductive argument, policies can be restricted to be Markov without any loss.

### 5.1 Dynamic Programming, Optimality of Markov Policies and Bellman’s Principle of Optimality

#### 5.1.1 Optimality of Deterministic Markov Policies

We will observe that when there is an optimal solution, the optimal solution can be taken to be Markov. Even when an optimal policy may not exist, any measurable policy can be replaced with one which is Markov, under fairly general conditions, as we discuss below. In the following, first, we will follow David Blackwell’s [22] and Hans Witsenhausen [126]’s ideas to obtain a very interesting result.

**Theorem 5.1.1 (Blackwell’s Irrelevant Information Theorem)** Let \( \mathcal{X}, \mathcal{Y}, \mathcal{U} \) be complete, separable, metric spaces, and let \( P \) be a probability measure on \( B(\mathcal{X} \times \mathcal{Y}) \), and let \( c : \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R} \) be a Borel measurable and bounded cost function. Then, for any Borel measurable function \( \gamma : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{U} \), there exists another Borel measurable function \( \gamma^* : \mathcal{X} \rightarrow \mathcal{U} \) such that
\[
\int_X c(x, \gamma^*(x)) P(dx) \leq \int_{X \times Y} c(x, \gamma(x,y)) P(dx, dy)
\]

Thus, policies based only on \(x\) almost surely, are optimal.

**Proof.** We will construct a \(\gamma^*\) given \(\gamma\). Let us write
\[
\int_{X \times Y} c(x, \gamma(x,y)) P(dx, dy) = \int_X \left( \int_U c(x, u) P^\gamma(du|x) \right) P(dx),
\]

where \(P^\gamma(u \in D|x) = \int 1_{\{\gamma(x,y) \in D\}} P(dy|x)\). Consider
\[
h^\gamma(x) := \int_U c(x, u) P^\gamma(du|x)
\]

- Suppose the space \(U\) is countable. In this case, let us enumerate the elements in \(U\) as \(\{u^k, k = 1, 2, \ldots\}\). In this case, we could define:
\[
D_i = \{x \in X : c(x, u^i) \leq h^\gamma(x)\}, i = 1, 2, \ldots
\]

We note that \(X = \bigcup_i D_i\): Suppose not, then \(c(x, u^i) > h^\gamma(x)\) for all \(i\) and thus:
\[
h^\gamma(x) = \left( \int_U c(x, u) P^\gamma(du|x) \right) > h^\gamma(x),
\]

leading to a contradiction. Now define,
\[
\gamma^*(x) = u^k \quad \text{if} \quad x \in D_k \setminus (\bigcup_{i=1}^{k-1} D_i), k = 1, 2, \ldots,
\]

Such a function is measurable, by construction and performs at least as good as \(\gamma\).

- We now provide a proof for the actual statement. Let \(D = \{x, u \in X \times U : c(x, u) \leq h^\gamma(x)\}\). \(D\) is a Borel set since \(c(x, u) - h^\gamma(x)\) is Borel. Define \(D_x = \{u \in U : (x, u) \in D\}\) for all \(x \in X\). Now, for every element \(x\) we can pick a member \(u\) which is in \(D\). The question now is whether the constructed map is Borel measurable. Now, for every \(x\), \(\int 1_{\{\gamma(x,y) \in D\}} P(dy|x) > 0\), since otherwise we would arrive at a contradiction (see e.g. page 41 in [45]). Then, by a measurable selection theorem of Blackwell and Ryll-Nardzewski [24] (see also p. 255 of [45]), there exists a Borel-measurable function \(\gamma^* : X \to U\) such that its graph is contained in \(D\), that is, \(\{(x, \gamma^*(x)) \in D\}\).

\[\Box\]

**Theorem 5.1.2** Let \(\{(x_t, u_t)\}\) be a controlled Markov Chain. Consider the minimization of \(E[\sum_{t=0}^{N-1} c(x_t, u_t) + c_N(x_N)]\), over all control policies which are sequences of causally measurable functions of \(\{x_s, s \leq t\}\), for all \(t \geq 0\). Any measurable policy can be replaced with one which is (deterministic) Markov and which is at least as good as the original policy. In particular, if an optimal control policy exists, there is no loss in restricting policies to be Markov, that is a policy which only uses the current state \(x_t\) and the time information \(t\).

**Proof:** The proof follows from a sequential application of Theorem 5.1.1 starting with the final time stage. For any admissible policy, the cost
\[
E[\{c(x_{N-1}, \gamma_{N-1}(h_{N-1})) + \int_X c_N(z)Q(dz|x_{N-1}, \gamma_{N-1}(h_{N-1}))\}],
\]
can be replaced with a measurable policy \(\gamma^*_t\)
\[
E[\{c(x_{N-1}, \gamma^*_t(x_{N-1})) + \int_X c_N(z)Q(dz|x_{N-1}, \gamma^*_t(x_{N-1}))\}],
\]
which leads a lower cost. Define
Suppose there is a cost $C$, the cost is equal to deterministic Markov. Since, (5.1)

$$J_{N-1}(x_{N-1}) := E[(c(x_{N-1}, \gamma_{N-1}^*(x_{N-1}))) + \int c_N(z)Q(dz|x_{N-1}, \gamma_{N-1}^*(x_{N-1}))],$$

and consider then

$$E[(c(x_{N-2}, \gamma_{N-2}(h_{N-2})) + \int J_{N-1}(z)Q(dz|x_{N-2}, \gamma_{N-2}(h_{N-2}))].$$

This expression can also be lower bounded by a measurable Markov policy $\gamma_{N-2}^*$ so that the expected cost

$$J_{N-2}(x_{N-2}) := E[(c(x_{N-2}, \gamma_{N-2}^*(x_{N-2}))) + \int J_{N-1}(z)Q(dz|x_{N-2}, \gamma_{N-2}^*(x_{N-2}))],$$

is lower than that achieved by the admissible policy. By induction, for all time stages, one can replace the policies with a deterministic Markov policy which leads to a cost which is at least as desirable as the cost achieved by the admissible policy.

Given this result, we have the following important optimality principle.

### 5.1.2 Bellman’s principle of optimality

Let $\{J_t(x_t)\}$ be a sequence of functions on $\mathbb{X}$ defined by

$$J_N(x) = c_N(x)$$

and for $0 \leq t \leq N - 1$

$$J_t(x) = \min_{u \in U_t(x)} \{c(x, u) + \int \int_{\mathbb{X}} J_{t+1}(z)Q(dz|x, u)\}.$$ 

Suppose these functions admit a minimum and are measurable. We will discuss a number of sufficiency conditions on when this is possible. Let there be minimizing selectors which are deterministic and denoted by $\{f_t(x)\}$ and let the minimum expected cost be equal to

$$J_t(x) = c(x, f_t(x)) + \int \int_{\mathbb{X}} J_{t+1}(z)Q(dz|x, f_t(x))$$

Then we have the following:

**Theorem 5.1.3** The policy $\Pi^* = \{f_0, f_1, \ldots, f_{N-1}\}$ is optimal and the cost function is equal to $J^*(x) = J_0(x)$

**Proof:** We compare the cost generated by the above policy, with respect to the cost obtained by any other policy, which can be taken to be deterministic Markov in view of Theorem 5.1.2.

We provide the proof by a backwards induction method in view of (5.1). Consider the time stage $t = N - 1$. For this stage, the cost is equal to

$$J_{N-1}(x_{N-1}) = \min \{c(x_{N-1}, u_{N-1}) + \int c_N(z)Q(dz|x_{N-1}, u_{N-1})\}$$

Suppose there is a cost $C_{N-1}^*(x_{N-1})$, achieved by some policy $\eta = \{\eta_k, k \in \{0, 1, \ldots, N - 1\}\}$, which we take to be deterministic Markov. Since,

$$C_{N-1}^*(x_{N-1})$$

$$= c(x_{N-1}, \eta_{N-1}(x_{N-1})) + \int c_N(z)Q(dz|x_{N-1}, \eta_{N-1}(x_{N-1}))$$

$$\geq J_{N-1}(x_{N-1})$$

$$= \min_{u_{N-1}} \{c(x_{N-1}, u_{N-1}) + \int c_N(z)Q(dz|x_{N-1}, u_{N-1})\}, \tag{5.1}$$
it must be that \( C^*_N(x_{N-1}) \geq J_{N-1}(x_{N-1}) \). Now, we move to time stage \( N - 2 \). In this case, the cost to be minimized is given by

\[
C^*_N(x_{N-2}) = c(x_{N-2}, \eta(x_{N-2})) + \int_X C^*_N(z) Q(dz|x_{N-2}, \eta(x_{N-2})) \\
\geq \min_{u_{N-2}} \{ c(x_{N-2}, u_{N-2}) + \int_X J_{N-1}(z) Q(dz|x_{N-2}, u_{N-2}) \} \\
=: J_{N-2}(x_{N-2})
\]

where the inequality is due to the fact that \( J_{N-1}(x_{N-1}) \leq C^*_N(x_{N-1}) \) and the minimization. We can, by induction show that the recursion holds for all \( 0 \leq t \leq N - 2 \).

\[\diamondsuit\]

5.1.3 Examples

Example 5.1 (Dynamic Programming and Investment). A investor’s wealth dynamics is given by the following:

\[ x_{t+1} = u_t w_t, \]

where \( \{w_t\} \) is an i.i.d. \( \mathbb{R}_+ \)-valued stochastic process. The investor has access to the past and current wealth information and his actions. The goal is to maximize:

\[ J(x_0, \Pi) = E_{x_0} \left[ \sum_{t=0}^{T-1} b(x_t - u_t) \right]. \]

The investor’s action set for any given \( x \) is: \( U(x) = [0, x] \).

For this problem, the state space is \( \mathbb{R} \), the control action space at state \( x \) is \( [0, x] \), the information at the controller is \( I_t = \{ x_{[0,t]}, u_{[0,t-1]} \} \). The kernel is described by the relation \( x_{t+1} = u_t w_t \). Using Dynamic Programming

\[
J_{T-1}(x_{T-1}) = \max_{u_{T-1} \in [0, x_{T-1}]} E[b(x_{T-1} - u_{T-1}) | I_{T-1}] \\
= \max_{u_{T-1} \in [0, x_{T-1}]} b(x_{T-1} - u_{T-1}) = b(x_{T-1}).
\]

Since there is no more future, the investor needs to collect the wealth at time \( T - 1 \), that is \( u_{T-1} = 0 \). For \( t = T - 2 \)

\[
J_{T-2}(x_{T-2}) = \max_{u_{T-2} \in [0, x_{T-2}]} E[b(x_{T-2} - u_{T-2}) + J_{T-1}(x_{T-1}) | I_{T-2}] \\
= \max_{u_{T-2} \in [0, x_{T-2}]} E[b(x_{T-2} - u_{T-2}) + b x_{T-1} | x_{T-2}, u_{T-2}] \\
= \max_{u_{T-2} \in [0, x_{T-2}]} \left( b(x_{T-2} - u_{T-2}) + b E[w_{T-2}] u_{T-2} \right) \\
= \max_{u_{T-2} \in [0, x_{T-2}]} \left( b x_{T-2} + b E[w_{T-2}] - 1 u_{T-2} \right)
\]

It follows then that if \( E[w_{T-2}] > 1 \), \( u_{T-2} = x_{T-2} \) (that is, investment is favourable), otherwise \( u_{T-2} = 0 \). Recursively, one concludes that if \( E[w] > 1 \), \( u_t = x_t \) is optimal until \( t = T - 1 \), at \( t = T - 1 \), \( u_{T-1} = 0 \), leading to \( J_0(x_0) = b E[w] x_0 \). If \( E[w] < 1 \), it is optimal to collect at time 0, that is \( u_0 = 0 \), leading to \( J_0(x_0) = b x_0 \). If \( E[w] = 1 \), both of these policies lead to the same reward.

Example 5.2 (Linear Quadratic Gaussian Systems). Consider the following Linear Quadratic Gaussian (LQG) problem with \( q > 0 \), \( r > 0 \), \( p_T > 0 \):
68 5 Dynamic Programming

\[ \inf_{H} E_T^H \left[ \sum_{t=0}^{T-1} q x_t^2 + r_t^2 + p_T x_T^2 \right] \]

for a linear system:

\[ x_{t+1} = a x_t + u_t + w_t, \]

where \( w_t \) is a zero-mean, i.i.d. Gaussian random variable with variance \( \sigma_w^2 \). We can show, by the method of completing the squares, that:

\[ J_t(x_t) = P_t x_t^2 + \sum_{k=t}^{T-1} P_{t+1} \sigma_w^2 \]

where

\[ P_t = q + P_{t+1} a^2 - \frac{P_{t+1} a^2}{P_{t+1} + r} \]

and the optimal control policy is

\[ u_t = -\frac{P_{t+1} a}{P_{t+1} + r} x_t. \]

Note that, the optimal control policy is Markov (as it uses only the current state). For a more general treatment for such LQG problems, see Section 5.3.

5.2 Existence of Minimizing Selectors and Measurability

The above dynamic programming arguments hold when there exist minimizing control policies (selectors measurable with respect to the Borel field on \( \mathcal{X} \)).

**Measurable Selection Hypothesis:** Given a sequence of functions \( J_t : \mathcal{X} \rightarrow \mathbb{R} \), there exists

\[ J_t(x) = \min_{u_t \in U_t(x)} (c(x_t, u_t) + \int_{\mathcal{X}} J_{t+1}(y) Q(dy|x, u)), \]

for all \( x \in \mathcal{X} \), for \( t \in \{0, 1, 2, \ldots, N-1\} \) with

\[ J_N(x_N) = c_N(x_N). \]

Furthermore, there exist measurable functions \( f_t \) such that

\[ J_t(x) = (c(x_t, f(x_t)) + \int_{\mathcal{X}} J_{t+1}(y) Q(dy|x, f(x_t))), \]

\( \diamond \)

Recall that a set in a normed linear space is (sequentially) compact if every sequence in the set has a converging subsequence.

**Condition 1.** The cost function to be minimized \( c(x_t, u_t) \) is continuous on both \( U \) and \( \mathcal{X} \); \( U_t(x) = U \) is compact; and \( \int_{\mathcal{X}} Q(dy|x, u) v(y) \) is a (weakly) continuous function on \( \mathcal{X} \times U \) for every continuous and bounded \( v \) on \( \mathcal{X} \).

**Condition 2.** For every \( x \in \mathcal{X} \) the cost function to be minimized \( c(x_t, u_t) \) is continuous on \( U \); \( U_t(x) = U \) is compact; and \( \int_{\mathcal{X}} Q(dy|x, u) v(y) \) is a (strongly) continuous function on \( U \) for every bounded, measurable function \( v \) on \( \mathcal{X} \) for every fixed \( x \).

**Theorem 5.2.1** Under Condition 1 or Condition 2, there exists an optimal solution and the Measurable Selection applies, and there exists a minimizing control policy \( f_t : \mathcal{X} \rightarrow U_t(x_t) \).

Furthermore, under Condition 1, the function \( J_t(x) \) is continuous, if \( c_N \) is continuous and bounded.
The result follows from the following three lemma below:

**Lemma 5.2.1** A continuous function $f : \mathbb{X} \to \mathbb{R}$ over a compact set $A \subset \mathbb{X}$ admits a minimum.

**Proof:** Let $\delta = \inf_{x \in A} f(x)$. Let $\{x_i\}$ be a sequence such that $f(x_i)$ converges to $\delta$. Since $A$ is compact $\{x_i\}$ must have a converging subsequence $\{x_{i(n)}\}$. Let the limit of this subsequence be $x_0$. Then, it follows that, $\{x_{i(n)}\} \to x_0$ and thus, by continuity $f(x_{i(n)}) \to f(x_0)$. As such $f(x_0) = \delta$. To see why compactness is important, consider 

$$\inf_{x \in A} \frac{1}{x}$$

for $A = [1, 2)$. How about for $A = \mathbb{R}$? In both cases there does not exist an $x$ value in the specified set which attains the infimum.

**Lemma 5.2.2** Let $U$ be compact, and $c(x, u)$ be continuous on $\mathbb{X} \times U$. Then, $\min_u c(x, u)$ is continuous on $\mathbb{X}$.

**Proof:** Let $x_n \to x$, $u_n$ optimal for $x_n$ and $u$ optimal for $x$. Such optimal action values exist as a result of compactness of $U$ and continuity. Now,

$$\left| \min_u c(x_n, u) - \min_u c(x, u) \right| \leq \max \left( c(x_n, u) - c(x, u), c(x, u_n) - c(x, u) \right)$$

(5.3)

The first term above converges since $c$ is continuous in $x, u$. The second converges also. Suppose otherwise. Then, for some $\epsilon > 0$, there exists a subsequence such that

$$|c(x, u_{k_n}) - c(x, u_{k_n})| \geq \epsilon$$

Consider the sequence $(x_{k_n}, u_{k_n})$. There exists a subsequence such that $(x_{k_n'}, u_{k_n'})$ which converges to $x, u'$ for some $u'$ since $U$ is compact. Hence, for this subsequence, we have convergence of $c(x_{k_n'}, u_{k_n'})$ as well as $c(x, u_{k_n'})$, leading to a contradiction.

**Lemma 5.2.3** Let $c(x, u)$ be a continuous function on $U$ for every $x$, where $U$ is a compact set. Then, there exists a Borel measurable function $f : \mathbb{X} \to U$ such that

$$c(x, f(x)) = \min_u c(x, u)$$

**Proof:** The result builds on [66, Theorem 2], [109] and [71], among others. A sketch is as follows: Let $\tilde{c}(x) := \min_{u \in U} c(x, u)$. The function

$$\tilde{c}(x) := \min_{u \in U} c(x, u),$$

is Borel measurable. This follows from the observation that it is sufficient to prove that $\{x : \tilde{c}(x) > \alpha\}$ is Borel for every $\alpha \in \mathbb{R}$. By continuity of $c$ and compactness of $U$, with a successively refining quantization of the space of control actions $U$ (such a sequence of quantizers map $U$ to a sequence of finite sets (expanding as $n$ increases), so that $\lim_{n \to \infty} |\sup_u Q_n(u) - u| = 0$ and the cardinality $|Q_n(U)| < \infty$ for every $n$)

$$\{x : \tilde{c}(x) > \alpha\} = \bigcap_{n} \bigcap_{Q_n(u), u \in U} \{x : c(x, Q_n(u)) > \alpha\}$$

the result follows since each of $\{x : c(x, Q_n(u)) > \alpha\}$ is Borel. Define $\mathbb{F} := \{(x, u) : c(x, u) = \tilde{c}(x), \ x \in \mathbb{X}\}$. This set is a Borel set. The question is now whether one can construct a measurable (selection) function $\gamma$ in $\mathbb{F}$ so that $\{(x, \gamma(x))\} \subset \mathbb{F}$. We can construct a measurable function which lives in this set, using the property that $U$ is a separable metric space: This builds on measurable selection results due to Schäl [109] and [71] (see Appendix C).
Remark 5.3. The results of Lemma also apply when \( \mathbb{U}(x) \) depends on \( x \) so that it is compact for each \( x \) and \( \{(x, u) : u \in \mathbb{U}(x), x \in \mathbb{X}\} \) is a Borel subset of \( \mathbb{X} \times \mathbb{U} \). If continuity is also to be attained for the value function \( c(x, f(x)) \), a close look at the proof of Lemma reveals that it suffices if \( \mathbb{U} \) is compact and \( \mathbb{U}(x) \) is an upper semi-continuous set-valued function (that is if for any \( x^n \to x \) and \( u^n \in \mathbb{U}(x^n) \) with \( u^n \to u' \) then \( u' \in \mathbb{U}(x) \)) and \( c \) is continuous. For some further discussions, see Proposition D.5 in [59].

We can replace the compactness condition with an \textit{inf-compactness} condition, and modify Condition 1 as below:

**Condition 3.** For every \( x \in \mathbb{X} \) the cost function to be minimized \( c(x, u) \) is continuous on \( \mathbb{X} \times \mathbb{U} \); is non-negative; \( \{u : c(x, u) \leq \alpha\} \) is compact for all \( \alpha > 0 \) and all \( x \in \mathbb{X} \); \( \int_{\mathbb{X}} Q(\mathbb{d}y|x, u)v(y) \) is a continuous function on \( \mathbb{X} \times \mathbb{U} \) for every continuous and bounded \( v \).

**Theorem 5.2.2** Under Condition 3, the Measurable Selection Hypothesis applies.

We could relax the continuity condition and change it with lower semi-continuity. A function is lower semi-continuous at \( x \) if \( \liminf_{x \to x_0} f(x) \geq f(x_0) \). We state the following, see Theorem 3.3.5 in [59].

**Theorem 5.2.3** Suppose that (i) \( \mathbb{U}(x) \) is compact for every \( x \) and \( \{(x, u) : u \in \mathbb{U}(x)\} \) is a Borel subset of \( \mathbb{X} \times \mathbb{U} \), (ii) \( c \) is lower semi-continuous on \( \mathbb{U}(x) \) for every \( x \in \mathbb{X} \) and (iii) \( \int c(x_{t+1}) P(dx_{t+1}|x_t = x, u_t = u) \) is lower semi-continuous on \( \mathbb{U}(x) \) for every \( x \in \mathbb{X} \) and every measurable and bounded \( v \) on \( \mathbb{X} \). Then, the measurable selection hypothesis applies.

For many problems, one can compute an optimal solution directly, without searching for existence.

### 5.3 The Linear Quadratic Gaussian (LQG) Problem

Consider the following linear system

\[
x_{t+1} = Ax_t + Bu_t + w_t,
\]

where \( x \in \mathbb{R}^n, u \in \mathbb{R}^m \) and \( w \in \mathbb{R}^n \). Suppose \( \{w_t\} \) is i.i.d. Gaussian with a given covariance matrix \( E[w_tw_t^T] = W \) for all \( t \geq 0 \).

The goal is to obtain

\[
\inf_{\Pi} J(\Pi, x),
\]

where

\[
J(\Pi, x) = E^\Pi \left[ \sum_{t=0}^{T-1} x_t^T Q x_t + u_t^T R u_t + x_T^T Q_T x_T \right],
\]

with \( R, Q, Q_T > 0 \) (that is, these matrices are positive definite).

In class, we obtained the Dynamic Programming recursion for the optimal control problem.

**Theorem 5.3.1** The optimal control is linear and has the form:

\[
u_t = -(B P_{t+1} B + R)^{-1} B^T P_{t+1} A x_t
\]

where \( P_t \) solves the Discrete-Time Riccati Equation:

\[
P_t = Q + A^T P_{t+1} A - A^T P_{t+1} B ((B P_{t+1} B + R)^{-1} B^T P_{t+1} A,
\]

with final condition \( P_T = Q_T \). The optimal cost is given by

\[
J(x_0) = x_0^T P_0 x_0 + \sum_{t=0}^{T-1} E[w_t^T P_t+1 w_t]
\]
Consider the linear system
\[ x_{t+1} = Ax_t + Bu_t, \quad y_t =Cx_t \] (5.5)
Here, \( y_t \) is a measurement variable and \( x_t \) is an \( \mathbb{R}^n \)-valued state variable. Such a system is said to be controllable, if for any initial \( x_i \) and a final \( x_f \), there exists \( T \in \mathbb{N} \) and a sequence of control actions \( u_0, u_1, \ldots, u_{T-1} \) such that with \( x_0 = x_i, x_T = x_f \). If \( x_f \) is restricted to be \( 0 \in \mathbb{R}^n \), and the above holds, the system is said to be stabilizable. Thus, the only modes in a stabilizable system that are not controllable are the stable ones.

Now let \( B = 0 \) in (5.5). Such a system is said to be observable if by measuring \( y_0, y_1, \ldots, y_T \), for some \( T \in \mathbb{N} \), \( x_0 \) can be uniquely recovered. Such a system is called detectable if a linear if all unstable modes are observable, in the sense that if \( \{ y_t \} \to 0 \), it must be that \( \{ x_t \} \to 0 \).

There are well-known algebraic tests to verify controllability and observability. A very useful result building on the Cayley-Hamilton theorem is that if a system cannot be moved from any initial state to any final state in \( \{t \in \mathbb{N} \} \) → 0, \( x_0 \) can be uniquely recovered. Such a system is called detectable if A linear if all unstable modes are observable, in the sense that if \( \{ y_t \} \to 0 \), it must be that \( \{ x_t \} \to 0 \).

**Theorem 5.3.2** (i) If \((A, B)\) is controllable there exists a solution to the Riccati equation
\[ P = Q + A^T P A - A^T P B ((BPB + R))^{-1} B^T P A. \]

(i) if \((A, B)\) is controllable and, with \( Q = C^T C \), \((A, C)\) is observable; as \( T \to \infty \) in the optimization problem above, the sequence of Riccati recursions,
\[ P_t = Q + A^T P_{t+1} A - A^T P_{t+1} B ((BP_{t+1}B + R))^{-1} B^T P_{t+1} A, \]
converges to some limit \( P \) that satisfies
\[ P = Q + A^T P A - A^T P B ((BPB + R))^{-1} B^T P A. \]
Furthermore, such a \( P \) is unique, and is positive definite. Finally, under the optimal control policy, \( \{ x_t \} \) is stable.

**Remark 5.4.** Part (i) can be relaxed to \((A, B)\) being stabilizable; and part (ii) to \((A, C)\) being detectable for the existence of a unique \( P \) and a stable system under the optimal policy. In this case, however, \( P \) may only be positive semi-definite.

We observe that, if the system has noise, and if we have an average cost optimization problem, the effect of the noise will stay bounded, and the same solution with \( P \) and the induced control policy, will be optimal for the minimization of the expected average cost optimization problem:
\[ \lim_{T \to \infty} \frac{1}{T} E_{\mathcal{F}_T} \left[ \sum_{t=0}^{T-1} x_t^T Q x_t + u_t^T R u_t + x_T^T Q_T x_T \right]. \]
We will discuss average cost optimization problems in **Chapter 7**

### 5.4 Optional: A Strategic Measures Approach

For stochastic control problems, **strategic measures** are defined (see [109], [45] and [48]) as the set of probability measures induced on the product spaces of the state and action pairs by measurable control policies. Given an initial distribution on the state, and a policy, one can uniquely define a probability measure on the product space. Certain properties, such as measurability and compactness, of sets of strategic measures for single decision maker problems were studied in [109], [45], [48] and [23].

We assume, as before, that the spaces considered are standard Borel.
Theorem 5.4.1 (i) Let \( L_R(\mu) \) be the set of strategic measures induced by (possibly randomized) \( \Pi_A \) with \( x_0 \sim \mu \). Then, for any \( P \in L_R(\mu) \), there exists an augmented space \( \Omega \) and a probability measure \( \eta \) on \( B(\Omega) \) such that

\[
P(B) = \int_{\Omega} \eta(d\omega) P^{\Pi(\omega)}(B), \quad B \in B((\mathcal{X} \times \mathcal{U})^T),
\]

where each \( \Pi(\omega) \in \Pi_A \) is deterministic admissible.

(ii) Let \( L_M(\mu) \) be the set of strategic measures induced by \( \Pi_M \) with \( x_0 \sim \mu \). Then, for any \( P \in L_M(\mu) \), there exists an augmented space \( \Omega \) and a probability measure \( \eta \) on \( B(\Omega) \) such that

\[
P(B) = \int_{\Omega} \eta(d\omega) P^{\Pi(\omega)}(B), \quad B \in B((\mathcal{X} \times \mathcal{U})^T),
\]

where each \( \Pi(\omega) \in \Pi_M \) is deterministic Markov.

Proof. Here, we build on Lemma 1.2 in Gikhman and Shorodh [54] and Theorem 1 in [47]. Any stochastic kernel \( P(dx|y) \) can be realized by some measurable function \( x = f(y, v) \) where \( v \) is a uniformly distributed random variable on \([0, 1]\) and \( f \) is measurable (see also [26] for a related argument). One can define a new random variable \( (\omega = (v_0, v_1, \ldots, v_{T-1})) \), for both (i) and (ii). It then follows that both representation results hold. In particular, \( \eta \) can be taken to be the probability measure constructed on the product space \([0, 1]^T\) by the independent variables \( v_k, k \in \{0, 1, \ldots, T-1\} \).

It is interesting to note that, this does not hold for strategic measures induced by \( \Pi_S \).

One implication of this theorem is that if one relaxes the measure \( \eta \) to be arbitrary, a convex representation would be possible. That is, the set

\[
P(B) = \int_{\Omega} \eta(d\omega) P^{\Pi(\omega)}(B), \quad B \in B((\mathcal{X} \times \mathcal{U})^T), \eta \in \mathcal{P}(\Omega)
\]

is convex, when one does not restrict \( \eta \) to be a fixed measure. Furthermore, the extreme points of these convex sets consist of policies which are deterministic. A further implication then is that, since the expected cost function is linear in the strategic measures, one can without any loss consider the extreme points while searching for optimal policies. In particular,

\[
\inf_{\Pi \in \Pi_R M} J(x, \Pi) = \inf_{\Pi \in \Pi M} J(x, \Pi)
\]

and

\[
\inf_{\Pi \in \Pi_A M} J(x, \Pi) = \inf_{\Pi \in \Pi A} J(x, \Pi).
\]

Thus, deterministic policies are as good as any other. This is certainly not surprising in view of Theorem 5.1.1.

We present the following characterization for strategic measures. Let for all \( n \in \mathbb{N}, h_n = \{x_0, u_0, \ldots, x_{n-1}, u_{n-1}, x_n, u_n\} \), and \( P(dx_n|h_{n-1}) = Q(dx_n|x_{n-1}, u_{n-1}) \) be the transition kernel.

Let \( L_A(\mu) \) be the set of strategic measures induced by deterministic policies and let \( L_R(\mu) \) be the set of strategic measures induced by independently provided randomized policies. Such an individual randomized policy can be represented in a functional form: By Lemma 1.2 in Gikhman and Shorodh [54] and Theorem 1 in Feinberg [47], for any stochastic kernel \( \Pi^k \) from \( \mathcal{Y}^k \) to \( \mathcal{U}^k \), there exists a measurable function \( \gamma^k : [0, 1] \times \mathcal{Y}^k \to \mathcal{U}^k \) such that

\[
m\{r : \gamma^k(r, y^k) \in A\} = \Pi^k(u^k \in A|y^k), \quad (5.6)
\]

and \( m \) is the uniform distribution (Lebesgue measure) on \([0, 1]\).

Theorem 5.4.2 A probability measure \( P \in \mathcal{P} \left( \prod_{k=1}^N (\mathcal{X} \times \mathcal{U}) \right) \) is a strategic measure induced by a randomized policy (that is in \( L_R(\mu) \)) if and only if for every \( n \in \mathbb{N} \):
5.5 Infinite Horizon Optimal Discounted Cost Control Problems

\[ \int P(dh_{n-1}, dx_n)g(h_{n-1}, x^n) = \int P(dh_{n-1}) \left( \int X g(h_{n-1}, z)Q(dz|h_{n-1}) \right), \]

(5.7)

and

\[ \int P(dh_n)g(h_{n-1}, x_n, u_n) = \int P(dh_{n-1}, dx_n) \left( \int U_n g(h_{n-1}, x_n, a)\Pi^n(da|h_{n-1}, x_n) \right), \]

(5.8)

for some stochastic kernel \( \Pi^n \) on \( \mathbb{U}^n \) given \( h_n, x_n \), for all continuous and bounded \( g \), with \( P(d\omega_0) = \mu(d\omega_0) \).

**Proof.** The proof follows from the fact that testing the equalities such as (5.7) on continuous and bounded functions implies this property for any measurable and bounded function (that is, continuous and bounded functions form a separating class, see e.g. p. 13 in [19] or Theorem 3.4.5 in [46]).

An implication is the following.

**Theorem 5.4.3** [109] The set of strategic measures induced by admissible randomized policies is compact under the weak convergence topology if \( P(dx_{t+1}|x_t = x, u_t = u) \) is weakly continuous in \( x, u \) and \( X, U \) are compact.

An implication of this result is that optimal policies exist, and are deterministic when the cost function is continuous in \( x, u \).

We note also that Schäl [109] introduces a more general topology, \( w-s \) topology, which requires setwise continuity in the control actions. In this case, one can generalize Theorem 5.4.3 to the setups where Condition 2 applies and existence of optimal policies follows.

**Definition 5.4.1** The \( w-s \) topology on the set of probability measures \( \mathcal{P}(X \times U) \) is the coarsest topology under which \( \int f(x, u)\nu(dx, du) : \mathcal{P}(X \times U) \rightarrow \mathbb{R} \) is continuous for every measurable and bounded \( f \) which is continuous in \( u \) for every \( x \) (but unlike weak topology, \( f \) does not need to be continuous in \( x \)).

**Theorem 5.4.4** [109] The set of strategic measures induced by admissible randomized policies is compact under the \( w-s \) topology if \( P(dx_{t+1}|x_t = x, u_t = u) \) is strongly continuous in \( u \) for every \( x \) and \( X, U \) are compact.

The proofs of Theorems 5.4.3 and 5.4.4 follow from the property that to check whether a conditional independence property, as in (5.7) (5.8), holds testing these on continuous and bounded functions implies this property for any measurable and bounded function. Note that (5.8) holds since there is no conditional independence property condition, and the main issue is to establish that (5.7) holds for any converging sequence of strategic measures. Applying the hypotheses for each of the theorems leads to the desired results.

An implication of Theorem 5.4.4 is that an optimal strategic measure exists under the conditions of the theorem, provided that the cost function \( c \) is continuous in \( u \) for every \( x \). In particular, for any \( w-s \) converging sequence of strategic measures satisfying (5.7)-(5.8) so does the limit. By [109] Theorem 3.7], and the generalization of Portmanteau theorem for the \( w-s \) topology, the lower semi-continuity of of the integral cost over the set of strategic measures leads to the existence of an optimal strategic measure.

Now, we know that an optimal policy will be deterministic as a consequence of Theorem 5.4.4. Thus, a deterministic policy may not make use of randomization. Thus, an optimal team policy among deterministic policies exists.

5.5 Infinite Horizon Optimal Discounted Cost Control Problems

When the time horizon becomes unbounded, we can not directly invoke Dynamic Programming in the form considered earlier. In such settings, we look for conditions for optimality. Infinite horizon problems that we will consider will belong
to two classes: Discounted Cost and Average cost problems. We first discuss the discounted cost problem. The average cost problem is discussed in Chapter 7.

One popular type of problems are ones where the future costs are discounted: The future is less important than today, in part due to the uncertainty in the future, as well as due to economic understanding that current value of a good is more important than the value in the future.

The discounted cost function is given as:

$$J_T^\beta(x_0, \Pi) = E_{x_0}^\Pi \left[ \sum_{t=0}^{T-1} \beta^t c(x_t, u_t) \right],$$

for some $\beta \in (0, 1)$.

If there exists a policy $\Pi^*$ which minimizes this cost, the policy is said to be optimal.

We can consider an infinite horizon problem by taking the limit (when $c$ is non-negative)

$$J_\beta(x_0, \Pi) = \lim_{T \to \infty} E_{x_0}^\Pi \left[ \sum_{t=0}^{T-1} \beta^t c(x_t, u_t) \right],$$

and invoking the monotone convergence theorem:

$$J_\beta(x_0, \Pi) = E_{x_0}^{\Pi^*} \left[ \sum_{t=0}^{\infty} \beta^t c(x_t, u_t) \right].$$

We seek to find

$$J_\beta(x_0) = \inf_{\Pi \in \Pi_A} J_\beta(x_0, \Pi).$$

Define

$$\inf_{\Pi \in \Pi_A} J_T^\beta(x_0, \Pi) = J_T^\beta(x_0)$$

Lemma 5.5.1 Let $\mathbb{A}$ be a set and $\{f_n\}$ be a sequence of measurable mappings from $f_n : \mathbb{A} \to \mathbb{R}$ for all $n \in \mathbb{N}$. Then,

$$\limsup_{n \to \infty} \inf_{x \in \mathbb{A}} f_n(x) \leq \inf_{x \in \mathbb{A}} \limsup_{n \to \infty} f_n(x).$$

Proof. For any $n \in \mathbb{N}$ and $y \in \mathbb{A}$ we have

$$\inf_{x \in \mathbb{A}} f_n(x) \leq f_n(y).$$

This holds for all $n$ we can take the limit superior of both sides, which yields

$$\limsup_{n \to \infty} \inf_{x \in \mathbb{A}} f_n(x) \leq \limsup_{n \to \infty} f_n(y).$$

This inequality holds for all $y \in \mathbb{A}$ and thus

$$\limsup_{n \to \infty} \inf_{x \in \mathbb{A}} f_n(x) \leq \inf_{x \in \mathbb{A}} \limsup_{n \to \infty} f_n(x).$$

By Lemma 5.5.1, we change the order of limit and infimum so that

$$J_\beta(x_0) \geq \limsup_{T \to \infty} J_T^\beta(x_0) \quad (5.10)$$

but since $\lim$ exists for the right-hand side as the expression is monotonically increasing the limit superior becomes an actual limit and thus
$$J_\beta(x_0) \geq \lim_{T \to \infty} J_\beta^T(x_0).$$

We will make use of this relation explicitly in Theorem 5.5.2(iii) below.

Now, observe that (from (5.9))

$$J_\beta^T(x_0, \Pi) = E_{x_0}^\Pi \left[ c(x_0, u_0) + E_{x_0}^\Pi \sum_{t=1}^{T-1} \beta^t c(x_t, u_t) \right],$$

writes as

$$J_\beta^T(x_0, \Pi) = E_{x_0}^\Pi \left[ c(x_0, u_0) + \beta E_x^{\Pi} \sum_{t=1}^{T-1} \beta^{t-1} c(x_t, u_t) \right].$$

Through the controlled Markov property and the fact that without any loss Markov policies are as good as any other for finite horizon problems, it follows that

$$J_\beta^T(x_0) \geq \inf_{u_0} E_{x_0}^\Pi \left[ c(x_0, u_0) + \beta E_x^{\Pi} \left| J_\beta^{T-1}(x_1) \right| \right].$$

The goal is now to take $T \to \infty$ and obtain desirable structural properties. The limit

$$\lim_{T \to \infty} J_\beta^T(x_0)$$

will be a lower bound to $J_\beta(x_0)$ by (5.10). But the inequality will turn out to be an equality under mild conditions to be studied in the following.

The next result is on the exchange of the order of the minimum and limits, which will later show that the inequality above is indeed an equality.

**Lemma 5.5.2** [59] Let $V_n(x, u) \uparrow V(x, u)$ pointwise. Suppose that $V_n$ and $V$ are continuous in $u$ and $u \in \mathbb{U}(x) = \mathbb{U}$ is compact. Then,

$$\lim_{n \to \infty} \min_{u \in \mathbb{U}(x)} V_n(x, u) = \min_{u \in \mathbb{U}(x)} V(x, u)$$

**Proof.** The proof follows from essentially the same arguments as in the proof of Lemma 5.2.2. Let $u^*_n$ solve $\min_{u \in \mathbb{U}(x)} V_n(x, u)$. Note that

$$\left| \min_{u \in \mathbb{U}(x)} V_n(x, u) - \min_{u \in \mathbb{U}(x)} V(x, u) \right| \leq V_n(x, u^*_n) - V_n(x, u^*_n),$$

(5.11)

since $V_n(x, u) \uparrow V(x, u)$. Now, suppose that for some $\epsilon > 0$

$$V(x, u^*_n) - V_n(x, u^*_n) \geq \epsilon,$$

(5.12)

along a subsequence $n_k$. There exists a further subsequence $n'_k$ such that $u^*_n \to \bar{u}$ for some $\bar{u}$. Now, for every $n'_k \geq n$, since $V_n$ is monotonically increasing:

$$V(x, u^*_n) - V_n(x, u^*_n) \leq V(x, u^*_n) - V_n(x, u^*_n).$$

However, $V(x, u^*_n)$ and for a fixed $n$, $V_n(x, u^*_n)$, are continuous hence these two terms converge to: $V(x, \bar{u}) - V_n(x, \bar{u})$. For every fixed $x$ and $\bar{u}$, and every $\epsilon > 0$, we can find a sufficiently large $n$ such that $V(x, \bar{u}) - V_n(x, \bar{u}) \leq \epsilon/2$. Hence (5.12) cannot hold.

Now, recall from dynamic programming equations that with

$$\inf_{\Pi \in \Pi_A} J_\beta^T(x_0, \Pi) = J_\beta^T(x_0)$$

5.5 Infinite Horizon Optimal Discounted Cost Control Problems 75
For some stationary policy

Proof.

If

Lemma 5.5.3

In the following theorem, parts (i) and (ii) are adopted from [59].

Define

This last equation will turn out to be a crucial equation as the next Lemma reveals.

The following result shows that the fixed point equation (5.14) is closely related to optimality.

Define \( \mathbb{T} \) as follows:

In the following theorem, parts (i) and (ii) are adopted from [59].

**Lemma 5.5.3** [Verification Theorem]

(i) If \( v \) is a measurable \( \mathbb{R}_+ \)-valued function under Condition 2 (or continuous and bounded function under Condition 1) with \( v \geq \mathbb{T} v \), then \( v(x) \geq J_\beta(x) \).

(ii) If \( v \leq \mathbb{T} v \) and

for every policy and initial condition, then \( v(x) \leq J_\beta(x) \). As a result, a fixed point to \( \mathbb{T} \) leads to an optimal policy under \( \mathbb{P} \).

(iii) If

is so that \( v = \mathbb{T}(v) \) and \( \mathbb{T}(v)(x) = c(x, f(x)) + \beta E[v(x_1)|x_0 = x, u_0 = f(x)] \) is such that with \( \Pi = \{ f, f, \ldots \} \)

then \( \Pi \) is optimal.

Proof.

(i) For some stationary policy \( f \) that achieves \( \min(c(x, u) + \beta E[v(x_1)|x_0 = x, u_0 = u]) = c(x, f(x)) + \beta E[v(x_1)|x_0 = x, u_0 = f(x)] \), apply repeatedly.
A sufficient condition for (5.15) is that the cost is bounded.

leading to

(ii) If \( T J \)

(iii)(5.10) implies that

\[
E_x^f \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, f(x_k)) \right] + \beta^n E_x^f [v(x_n)]
\]

Since this is correct for every \( n \), it also is correct for the limit. Thus

\[
v(x) \geq \lim_{n \to \infty} E_x^f \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, f(x_k)) \right] \geq J_\beta(x).
\]

leading to

Thus, using the iterated expectations

leading to

If the last term on the right hand size converges to zero, then the result is obtained so that for any fixed policy, \( v \) provides a lower bound on the value function. Taking the infimum over all admissible policies, the desired result \( v(x) \leq J_\beta(x) \) is obtained.

(iii) (5.10) implies that \( J_\beta(x) \geq v(x) \) since \( v \) is the pointwise limit of the discounted cost functions. Now, Part(i) implies that

\[
v(x) = c(x, f(x)) + \beta \int v(y)Q(dy|x, f(x)) = \cdots = E_x^{Hf} \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, f(x_k)) \right] + \beta^n E_x^{Hf} [v(x_n)]
\]

and taking the limit, we have

\[
v(x) = E_x^{Hf} \left[ \sum_{k=0}^{\infty} \beta^k c(x_k, f(x_k)) \right]
\]

implying that \( f \) is optimal.

\[ \diamond \]

It follows then that under (5.14) we obtain

\[
J_\beta(x) \geq v(x) \geq \lim_{n \to \infty} E_x^f \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, f(x_k)) \right] \geq J_\beta(x),
\]

leading to

\[
J_\beta(x) = \lim_{n \to \infty} E_x^f \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, f(x_k)) \right]
\]

A sufficient condition for (5.15) is that the cost is bounded.
Theorem 5.5.1. Suppose the cost function is bounded, non-negative, and one of the the measurable selection conditions (Condition 1 or Condition 2) applies. Then, there exists a solution to the discounted cost problem which solves the fixed point equation.

\[ v(x) = (T(v))(x) := \min_u \{ c(x, u) + \beta \int_X v(y)Q(dy|x, u) \}, \quad \forall x \]  \tag{5.17}

Furthermore, the optimal cost (value function) is obtained by a successive iteration of policies:

\[ v_n(x) = \min_u \{ c(x, u) + \beta \int_X v_{n-1}(y)Q(dy|x, u) \}, \quad \forall x, n \geq 1 \]

with \( v_0(x) = 0 \) and \( v_n \) is a monotonically non-decreasing sequence.

Proof of Theorem 5.5.1. Suppose that Condition 2 holds (we could have considered other conditions as well; see Remark 5.5). We obtain the optimal solution as a limit of discounted cost problems with a finite horizon \( T \). By dynamic programming, we obtain the recursion for every \( T \) as:

\[ J_T^T(x) = 0 \]

\[ J_i^{T}(x) = T(J_{i+1}^{T})(x) = \{ \min_u \{ c(x, u) + \beta \int_X J_{i+1}^{T}(y)Q(dy|x, u) \} \}.  \tag{5.18} \]

This sequence will lead to a solution for a \( T \)-stage discounted cost problem. Since \( J_i^{T}(x) \leq J_{i+1}^{T}(x) \), if there exists some \( J_i^{T}(x) \) such that \( J_i^{T}(x) \uparrow J_i^{\infty}(x) \), we could invoke Lemma 5.5.2 to argue that

\[ J_i^{\infty}(x) = T(J_{i+1}^{\infty})(x) = \{ \min_u \{ c(x, u) + \beta \int_X J_{i+1}^{\infty}(y)Q(dy|x, u) \} \}. \]

Such a limit exists, by the monotone convergence theorem since \( J_i^{\infty}(x) \leq \sum_i \beta^i \sup_x |c(x, u)| < \infty \). Hence, a limit satisfying \( 5.17 \) indeed exists. By \( 5.10 \), a lower bound to an optimal solution will have to satisfy a fixed point equation \( 5.17 \). Building on these conditions, we can ensure that there is a solution to \( 5.17 \) that also satisfies \( 5.15 \) and by Lemma 5.5.3 the solution is the (optimal) value function. This completes the proof.

\[ \square \]

Remark 5.5. The condition that the cost is bounded is not necessary. It suffices that \( 5.13, 5.14 \) hold in addition to \( 5.15 \); that is, \( \int_X J_i^{T+1}(y)Q(dy|x, u) \) is continuous in \( u \) for every \( x \), provided that \( 5.15 \) holds.

We have thus also showed that one can arrive at the solution through a successive approximation method. A stronger result is the following.

Theorem 5.5.2. Suppose the cost function is bounded, non-negative, and one of the measurable selection conditions (Condition 1 or Condition 2) applies. Then, there exists a unique solution to the discounted cost problem which solves the fixed point equation.

\[ v(x) = \min_u \{ c(x, u) + \beta \int_X v(y)Q(dy|x, u) \}, \quad \forall x \]

Furthermore, the optimal cost (value function) is obtained by a successive iteration of policies (known as the Value Iteration Algorithm):

\[ v_n(x) = \min_u \{ c(x, u) + \beta \int_X v_{n-1}(y)Q(dy|x, u) \}, \quad \forall x, n \in \mathbb{N} \]  \tag{5.19}

For any \( v_0 \in L_\infty(\mathbb{X}) \), the sequence converges to a unique fixed point. If \( v_0(x) = 0 \), then \( v_n(x) \uparrow v(x) \) for all \( x \in \mathbb{X} \).

The following result will be useful.
Lemma 5.5.4 (i) The space of measurable functions \( \mathcal{X} \rightarrow \mathbb{R} \) endowed with the \( \| \cdot \|_\infty \) norm is a Banach space, that is

\[
L_\infty(\mathcal{X}) = \{ f : \| f \|_\infty = \sup_x |f(x)| < \infty \}
\]

is a Banach space.

(ii) The space of continuous and bounded functions from \( \mathcal{X} \rightarrow \mathbb{R} \), \( C_b(\mathcal{X}) \), endowed with the \( \| \cdot \|_\infty \) norm is a Banach space.

Proof of Theorem 5.5.2 Depending on the measurable selection conditions, we can take the value functions to be either measurable and bounded, or continuous and bounded. (i) Suppose that we consider the measurable and bounded case. We observe that the vector \( J^\infty \) lives in \( L_\infty(\mathcal{X}) \) (since the cost is bounded, there is a uniform bound for every \( x \)). We will show that the iteration given by

\[
\mathbb{T}(v)(x) = \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} v(y)Q(dy|x, u) \}
\]

is a contraction in \( L_\infty(\mathcal{X}) \). Let

\[
\| \mathbb{T}(v) - \mathbb{T}(v') \|_\infty = \sup_{x \in \mathcal{X}} |\mathbb{T}(v)(x) - \mathbb{T}(v')(x)|
\]

\[
= \sup_{x \in \mathcal{X}} |\{ \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} v(y)Q(dy|x, u) \} - \{ \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} v'(y)Q(dy|x, u) \} \} |
\]

\[
\leq \sup_{x \in \mathcal{X}} \left( A_1 \left\{ c(x, u^*) + \beta \int_{\mathcal{X}} v(y)Q(dy|x, u^*) - c(x, u^*) - \beta \int_{\mathcal{X}} v'(y)Q(dy|x, u^*) \right\} + A_2 \left\{ -c(x, u^{**}) - \beta \int_{\mathcal{X}} v(y)Q(dy|x, u^{**}) + c(x, u^{**}) + \beta \int_{\mathcal{X}} v'(y)Q(dy|x, u^{**}) \right\} \right)
\]

\[
= \sup_{x \in \mathcal{X}} \left( A_1 \left\{ \beta \int_{\mathcal{X}} (v(y) - v'(y))Q(dy|x, u^*) \right\} + A_2 \left\{ \beta \int_{\mathcal{X}} (v'(y) - v(y))Q(dy|x, u^{**}) \right\} \right)
\]

\[
\leq \beta \| v - v' \|_\infty \{ A_1 \int_{\mathcal{X}} Q(dy|x, u^*) + A_2 \int_{\mathcal{X}} Q(dy|x, u^{**}) \}
\]

\[
= \beta \| v - v' \|_\infty \quad (5.20)
\]

Here

\[
A_1 = \left\{ x : \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} v(y)Q(dy|x, u) \} \geq \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} v'(y)Q(dy|x, u) \} \right\},
\]

and \( A_2 \) denotes the complementary event, \( u^{**} \) is the minimizing control for \( \{ c(x, u) + \beta \int_{\mathcal{X}} v(y)Q(dy|x, u) \} \) and \( u^* \) is the minimizer for \( \{ c(x, u) + \beta \int_{\mathcal{X}} v'(y)Q(dy|x, u) \} \).

As a result \( \mathbb{T} \) defines a contraction on the Banach space \( L_\infty(\mathcal{X}) \), and there exists a unique fixed point. Thus, the sequence of iterations in (5.13)

\[
J^T_t(x) = \mathbb{T}(J^T_{t+1})(x) = \{ \min_u \{ c(x, u) + \beta \int_{\mathcal{X}} J^T_{t+1}(y)Q(dy|x, u) \} \}
\]

converges to \( J^\infty_t(x) = J^\infty(\mathcal{X}) \). Thus, if one lets \( v_0(x) = 0 \) for all \( x \in \mathcal{X} \), the iterations increase monotonically and converges to the value function. If one is only interested in convergence (and not the monotone behaviour), any initial function \( v_0 \in L_\infty(\mathcal{X}) \) is sufficient.

(ii) The above discussion also applies by considering a contraction on the space \( C_b(\mathcal{X}) \), if Condition 1 holds; in this case, the value function is continuous.

\[ \Box \]

5.5.1 Discounted cost problems with unbounded per-stage costs

As discussed earlier, one could follow the iteration method for the unbounded case (as in the proof of Theorem 5.5.1), whereas the contraction method in the proof of Theorem 5.5.2 holds for the bounded cost case. The contraction method can also be adjusted for the unbounded case under further conditions: If the cost is not bounded, one can define a weighted
There exist nonnegative real numbers $M$ and $\alpha \in [1, \frac{1}{2})$, and a weight function $w : X \to [1, \infty)$ such that for each $z \in X$, we have
\begin{equation}
\sup_{a \in U} |c(z, u)| \leq M w(z),
\end{equation}
and
\begin{equation}
\sup_{a \in U} \int_X w(y) \eta(dy|x, u) \leq \alpha w(z),
\end{equation}

Assumption 5.5.2 (i) The one stage cost function $c(x, u)$ is nonnegative and continuous in $u$ for every $x$.

(ii) The stochastic kernel $\eta(\cdot|x, u)$ is setwise continuous in $u$ for every $x$, i.e., if $u_k \to u$, then $\int u(y) \eta(dy|x, u_k) \to \int u(y) \eta(dy|x, u)$ for every measurable and bounded function $u$.

(iii) $U$ is compact.

(iv) There exist nonnegative real numbers $M$ and $\alpha \in [1, \frac{1}{2})$, and a weight function $w : X \to [1, \infty)$ such that for each $z \in X$, we have
\begin{equation}
\sup_{a \in U} |c(z, u)| \leq M w(z),
\end{equation}
and
\begin{equation}
\sup_{a \in U} \int_X w(y) \eta(dy|x, u) \leq \alpha w(z),
\end{equation}

Assumption 5.5.3 (i) The one stage cost function $c(x, u)$ is nonnegative and continuous in $(x, u)$.

(ii) The stochastic kernel $\eta(\cdot|x, u)$ is weakly continuous in $(x, u)$, i.e., if $(x_k, u_k) \to (x, u)$, then $\eta(\cdot|x_k, u_k) \to \eta(\cdot|x, u)$ weakly.

(iii) $U$ is compact.

(iv) There exist nonnegative real numbers $M$ and $\alpha \in [1, \frac{1}{2})$, and a continuous weight function $w : X \to [1, \infty)$ such that for each $z \in X$, we have
\begin{equation}
\sup_{a \in U} |c(z, u)| \leq M w(z),
\end{equation}
and
\begin{equation}
\sup_{a \in U} \int_X w(y) \eta(dy|x, u) \leq \alpha w(z),
\end{equation}

Define the operator $T$ on the set of real-valued measurable functions on $X$ as
\begin{equation}
T u(z) = \min_{a \in U} \left[ c(z, a) + \beta \int_X u(y) \eta(dy|z, a) \right].
\end{equation}
It can be proved that $T$ is a contraction operator mapping $B_w(X)$ into itself with modulus $\sigma = \beta \alpha$ (see [60, Lemma 8.5.5]); that is,
\begin{equation}
\|T u - T v\|_w \leq \beta \|u - v\|_w \text{ for all } u, v \in B_w(X).
\end{equation}

Define the function $J^*$ as
\begin{equation}
J^*(z) = \inf_{\varphi \in \Phi} J(\varphi, z).
\end{equation}
We call $J^*$ the discounted value function of the MDP. The theorem below follows from [60, Section 8.5, p.65].

Theorem 5.5.3 [60, Theorem 8.3.6] [60, Lemma 8.5.5]
5.7 Exercises

(i) Suppose Assumption 5.5.1 (or 5.5.2) holds. Then, the value function \( J^* \) is the unique fixed point in \( B_w(\mathcal{X}) \) (or \( B_w(\mathcal{X}) \cap C(\mathcal{X}) \)) under Assumption 5.5.2 of the contraction operator \( T \), i.e.,

\[
J^* = TJ^*.
\] (5.26)

Furthermore, a deterministic stationary policy \( f^* \) is optimal if and only if

\[
J^*(z) = c(z, f^*(z)) + \beta \int_{\mathcal{X}} J^*(y) \eta(dy|z, f^*(z)).
\] (5.27)

Finally, there exists a deterministic stationary policy \( f^* \) which is optimal, so it satisfies (5.27).

(ii) If instead of Assumption 5.5.1, Assumption 5.5.2 holds, the value function \( J^* \) will be the unique fixed point in \( B_w(\mathcal{X}) \cap C(\mathcal{X}) \) of the contraction operator \( T \).

The proof of (i) follows from [60, Theorem 8.3.6]. The proof of item (ii) follows from a minor modification of [60, Lemma 8.5.5].

5.6 Concluding Remarks

5.7 Exercises

Exercise 5.7.1 An investor’s wealth dynamics is given by the following:

\[
x_{t+1} = u_tw_t,
\]

where \( \{w_t\} \) is an i.i.d. \( \mathbb{R}_+ \)-valued stochastic process with \( E[\sqrt{w_t}] = 1 \) and \( u_t \) is the investment of the investor at time \( t \). The investor has access to the past and current wealth information and his previous actions. The goal is to maximize:

\[
J(x_0, \Pi) = E_{x_0}[\sum_{t=0}^{T-1} \sqrt{x_t - u_t}].
\]

The investor’s action set for any given \( x \) is: \( U(x) = [0, x] \). His initial wealth is given by \( x_0 \).

Formulate the problem as an optimal stochastic control problem by clearly identifying the state space, the control action space, the information available at the controller at any time, the transition kernel and a cost functional mapping the actions and states to \( \mathbb{R} \).

Find an optimal policy.

Hint: For \( \alpha \geq 0 \), \( \sqrt{x-u} + \alpha \sqrt{u} \) is a concave function of \( u \) for \( 0 \leq u \leq x \) and its maximum is computed when the derivative of \( \sqrt{x-u} + \alpha \sqrt{u} \) is set to zero.

Exercise 5.7.2 Consider the following linear system:

\[
x_{t+1} = Ax_t + Bu_t + w_t,
\]

where \( x \in \mathbb{R}^n, u \in \mathbb{R}^m \) and \( w \in \mathbb{R}^n \). Suppose \( \{w_t\} \) is i.i.d. Gaussian with a given covariance matrix \( E[w_tw_t^T] = W \) for all \( t \geq 0 \).

The goal is to obtain

\[
\inf_{\Pi} J(\Pi, x),
\]

where
Exercise 5.7.3 (Optimality of threshold type policies)

Consider an inventory-production system given by

\[ x_{t+1} = x_t + u_t - w_t, \]

where \( x_t \) is \( \mathbb{R} \)-valued, with the one-stage cost

\[ c(x_t, u_t, w_t) = bu_t + h \max(0, x_t + u_t - w_t) + p \max(0, w_t - x_t - u_t) \]

Here, \( b \) is the unit production cost, \( h \) is the unit holding (storage) cost and \( p \) is the unit shortage cost; here we take \( p > b \).

At any given time, the decision maker can take \( u_t \in \mathbb{R}_+ \). The demand variable \( w_t \sim \mu \) is a \( \mathbb{R}_+ \)-valued i.i.d. process, independent of \( x_0 \), with a finite mean where \( \mu \) is assumed to admit a probability density function. The goal is to minimize

\[ J(x, \Pi) = E_x^\Pi \left[ \sum_{t=0}^{T-1} c(x_t, u_t, w_t) \right] \]

The controller at time \( t \) has access to \( I_t = \{ x_s, u_s, s \leq t - 1 \} \cup \{ x_t \} \).

Obtain a recursive form for the optimal solution. In particular, show that the solution is of threshold type: There exists a sequence of real-numbers \( s_t \) so that the optimal solution is of the form: \( u_t = 0 \times 1_{\{ s_t \geq s_t \}} + (s_t - x_t) \times 1_{\{ s_t < s_t \}} \). See [15] for a detailed analysis of this problem.

Exercise 5.7.4 ([16]) Consider a burglar who is considering retirement. His goal is to maximize his earnings up to time \( T \). At any time, he can either continue his profession to steal an amount of \( w_t \) which is an i.i.d. \( \mathbb{R}_+ \)-valued random process (he adds this amount to his wealth), or retire.

However, each time he attempts burglary, there is a chance that he gets caught and he loses all of his savings; this happens according to an i.i.d. Bernoulli process so that he gets caught with probability \( \rho \) at each time stage.

Assume that his initial wealth is \( x_0 = 0 \). His goal is to maximize \( E[ x_T ] \). Find his optimal policy for \( 0 \leq t \leq T - 1 \).

Note: Such problems where a decision maker can quit or stop a process are known as optimal stopping problems.

Exercise 5.7.5 A fishery manager annually has \( x_t \) units of fish and sells \( u_t x_t \) of these where \( u_t \in [0, 1] \). With the remaining ones, the next year’s production is given by the following model

\[ x_{t+1} = w_t x_t (1 - u_t) + v_t, \]

with \( x_0 \) is given and \( \{ w_t, v_t \} \) is a sequence of mutually independent, identically distributed sequence of random variables with \( w_t \geq 0, v_t \geq 0 \) for all \( t \) and therefore \( E[w_t] = \bar{w} \geq 0 \) and \( E[v_t] = \bar{v} > 0 \).

At time \( T \), he sells all of the fish. The goal is to maximize the profit over the time horizon \( 0 \leq t \leq T - 1 \).

a) Formulate the problem as an optimal stochastic control problem by clearly identifying the state, the control actions, the information available at the controller, the transition kernel and a cost functional mapping the actions and states to \( \mathbb{R} \).
b) Does there exist an optimal policy? If it does, compute the optimal control policy as a dynamic programming recursion.

Exercise 5.7.6 A common example in mathematical finance applications is the portfolio selection problem where a controller (investor) would like to optimally allocate his wealth between a stochastic stock market and a market with a guaranteed income: Consider a stock with an i.i.d. random return $\sigma_t$ and a bank account with fixed interest rate $r > 0$. These are modeled by:

$$X_{t+1} = X_t u_t (1 + \sigma_t) + X_t (1 - u_t) (1 + r), \quad X_0 = 1$$

and

$$X_{t+1} = X_t (1 + r + u_t (\sigma_t - r))$$

Here, $u_t \in [0, 1]$ denotes the proportion of the money that the investor invests in the stock market. Suppose that the goal is to maximize $E[\log(X_T)]$. Then, we can write:

$$\log(X_T) = \log(\prod_{k=0}^{T-1} \frac{X_{k+1}}{X_k}) = \sum_{k=0}^{T-1} \log((1 + r + u_t(\sigma_t - r)))$$

(F5.28)

Formulate the problem as an optimal stochastic control problem by clearly identifying the state and the control action spaces, the information available at the controller, the transition kernel, and a cost functional mapping the actions and states to $\mathbb{R}$. Find the optimal policy.

Exercise 5.7.7 We will illustrate dynamic programming by considering a simplified version of a paper by B. Hajek (Optimal Control of Two interacting Service Stations; IEEE Trans. Automatic Control, vol. 29, June 1984).

Consider a two server-station network; where a router routes the incoming traffic, as is depicted in Figure 5.1.

![Fig. 5.1](image)

Customers arrive according to a (continuous-time) Poisson process of rate $\lambda$. The router routes to station 1 with probability $u$ and second station with probability $1 - u$. The router has access to the number of customers at both of the queues, while implementing her policy.

Station 1 has a service time distribution which is exponential with rate $\mu_1$, and Station 2 with $\mu_2 = \mu_1$, as well. After some computation, we find out that the controlled transition kernel is given by the following:

$$P(q_{t+1}^1 = q_t^1 + 1, q_{t+1}^2 = q_t^2 | q_t^1, q_t^2) = \lambda \frac{u}{\lambda + 2\mu_1}$$

$$P(q_{t+1}^1 = q_t^1, q_{t+1}^2 = q_t^2 + 1 | q_t^1, q_t^2) = \lambda \frac{(1 - u)}{\lambda + 2\mu_1}$$

$$P(q_{t+1}^1 = \max(q_t^1 - 1, 0), q_{t+1}^2 = q_t^2 | q_t^1, q_t^2) = \frac{\mu_1}{\lambda + 2\mu_1}$$

$$P(q_{t+1}^1 = q_t^1, q_{t+1}^2 = \max(0, q_t^2 - 1) | q_t^1, q_t^2) = \frac{\mu_1}{\lambda + 2\mu_1}$$
There is also a holding cost per unit time. The holding cost at Station 1 is \( c_1 > 0 \) and the cost at Station 2 is \( c_2 > 0 \). That is if there are \( q_1^t \) customers, the cost is \( c_1 q_1^t \) at station 1 at time \( t \) and likewise for Station 2.

The goal of the router is to minimize the expected total holding cost from time 0 to some time \( T \in \mathbb{N} \), where the total cost is

\[
\sum_{t=0}^{T} c_1 q_1^t + c_2 q_2^t.
\]

a) Express the problem as a dynamic programming problem, up until time \( T \). That is; where does the control action live? What is the state space? What is the transition kernel for the controlled Markov Chain?

Write down the dynamic programming recursion, starting from time \( T \) and going backwards.

b) Suppose that \( c_1 = c_2 \). Let \( J_t(q_1^t, q_2^t) \) be the value function at time \( t \) (that is the current cost and the cost to go).

Via dynamic programming, prove the following:

For a given \( t \), if, whenever \( 0 \leq q_1^t \leq q_2^t \) we have that

\[
J_t(q_1^t, q_2^t) \leq J_t(q_1^t - 1, q_2^t + 1),
\]

then the same applies for \( J_{t-1}(\cdot, \cdot) \), for \( t \geq 1 \). With the above, prove that an optimal control policy is given by:

\[
 u_t = 1_{\{q_1^t \leq q_2^t\}},
\]

for all \( t \) values.

**Exercise 5.7.8** Consider a scalar linear system with the following dynamics:

\[
x_{t+1} = ax_t + bu_t + w_t,
\]

where \( \{w_t\} \) is i.i.d Gaussian with zero-mean and unit variance. Suppose that the controller has access to \( I_t = \{x_{[0,t]}, u_{[0,t-1]}\} \) at time \( t \). Suppose that the initial state is \( x_0 = x \) for some \( x \in \mathbb{R} \). We wish to find for some \( \beta \in (0, 1) \):

\[
\inf_{\Pi} J(x_0, \Pi) = E_x^I \left[ \sum_{t=0}^{\infty} \beta^t (qx_t^2 + ru_t^2) \right],
\]

for \( q \geq 0 \) and \( r > 0 \).

Compute the optimal control policy and the optimal cost.

**Hint:** Use Lemma S.5.4. Start with a finite horizon version, and apply dynamic programming, obtain the solution and take the finite horizon to infinity. This is also equivalent to applying value iteration with \( v_0(x) = 0 \) for all \( x \in \mathbb{R} \). You will see that a recursion with \( v_t(x) = C_t x^2 + D_t \) will be obtained and \( C_t \) and \( D_t \) will have limits as \( t \to \infty \), \( C \) and \( D \), respectively. The optimal control will be stationary and deterministic:

\[
u_t = \gamma(x_t) = -(r + \beta C b^2)^{-1} \beta ab C x_t, \quad t \geq 0.
\]

Thus, you need to find \( C \) and \( D \).

Consider a scalar linear system with the following dynamics:

\[
x_{t+1} = ax_t + bu_t + w_t,
\]

where \( \{w_t\} \) is i.i.d Gaussian with zero-mean and unit variance. Suppose that the controller has access to \( I_t = \{x_{[0,t]}, u_{[0,t-1]}\} \) at time \( t \). Suppose that the initial state is \( x_0 = x \) for some \( x \in \mathbb{R} \). We wish to find for some \( \beta \in (0, 1) \):
\[
\inf_{\Pi} J(x_0, \Pi) = E^{\Pi}_x \left[ \sum_{t=0}^{\infty} \beta^t (q x_t^2 + r u_t^2) \right],
\]

for \( q \geq 0 \) and \( r > 0 \).

Compute the optimal control policy and the optimal cost.

Hint: See Lemma 5.5.3 in the lecture notes. Start with a finite horizon version, and apply dynamic programming, obtain the solution and take the finite horizon to infinity. This is also equivalent to applying value iteration with \( v_0(x) = 0 \) for all \( x \in \mathbb{R} \). You will see that a recursion with \( v(x) = C_t x^2 + D_t \) will be obtained and \( C_t \) and \( D_t \) will have limits as \( t \to \infty \), \( C \) and \( D \), respectively. The optimal control will be stationary and deterministic:

\[
u_t = \gamma(x_t) = -(r + \beta C b^2)^{-1} \beta a b c x_t, \quad t \geq 0.
\]

Thus, you need to find \( C \) and \( D \).
Partially Observed Markov Decision Processes, Non-linear Filtering and the Kalman Filter

As discussed earlier in Chapter 2, we consider a system of the form:

\[ x_{t+1} = f(x_t, u_t, w_t), \quad y_t = g(x_t, v_t). \]

Here, \( x_t \) is the state, \( u_t \in U \) is the control, \((w_t, v_t)\) are \((\mathbb{W} \times \mathbb{V})\)-valued i.i.d noise processes where \( w_t \) is independent of \( v_t \). We also assume that the state noise \( w_t \) either has a probability mass function, or admits a probability measure which is absolutely continuous with respect to the Lebesgue measure; this will ensure that the probability measure admits a density function. Hence, the notation \( p(x) \) will denote either the probability mass for discrete-valued spaces or probability density function for uncountable spaces. The controller only has causal access to the second component \( \gamma_t \) of the process. A policy \( \Pi = \{ \gamma_t \} \) is a collection of measurable functions so that \( \gamma_t \) is measurable with respect to \( \sigma(I_t) \) with \( I_t = \{ y(t), u(t-1) \} \) at time \( t \). We denote the observed history space as: \( H_0 := \mathbb{Y}, \quad H_t = H_{t-1} \times \mathbb{Y} \times U \).

In the following \( \mathcal{P}(\mathbb{X}) \) denotes the space of probability measures on \( \mathbb{X} \), which we assume to be a standard Borel (Polish) space. This makes \( \mathcal{P}(\mathbb{X}) \) a Polish space, see Appendix [D]

6.1 Enlargement of the State-Space and the Construction of a Controlled Markov Chain

One could transform a partially observable Markov Decision Problem to a Fully Observed Markov Decision Problem via an enlargement of the state space. In particular, when \( \mathbb{X} \) is countable, we obtain via the properties of total probability the following dynamical recursion

\[
\pi_t(x) := P(x_t = x | y_{[0,t]}, u_{[0,t-1]}) = \frac{\sum_{x_{t-1} \in \mathbb{X}} \pi_{t-1}(x_{t-1}) P(y_t | x_t) P(x_t | x_{t-1}, u_{t-1}) P(u_{t-1} | y_{[0,t-1]}, u_{[0,t-2]})}{\sum_{x_{t-1} \in \mathbb{X}} \sum_{x_t \in \mathbb{X}} \pi_{t-1}(x_{t-1}) P(y_t | x_t) P(x_t | x_{t-1}, u_{t-1}) P(u_{t-1} | y_{[0,t-1]}, u_{[0,t-2])}}.
\]

We will see that the conditional measure process forms a controlled Markov chain in \( \mathcal{P}(\mathbb{X}) \). Note that in the above analysis \( P(u_{t-1} | y_{[0,t-1]}, u_{[0,t-2]} \) is determined by the control policy, and \( P(x_t | x_{t-1}, u_{t-1}) \) is determined by the transition kernel \( \mathcal{T} \) of the controlled Markov chain. Here, the summation needs to be exchanged with an integral in case the variables live in an uncountable space.

The result above leads to the following.

**Theorem 6.1.1** The process \( \{ \pi_t, u_t \} \) is a controlled Markov chain. That is, under any admissible control policy, given the action at time \( t \geq 0 \) and \( \pi_t \), \( \pi_{t+1} \) is conditionally independent from \( \{ \pi_s, u_s, s \leq t-1 \} \).
Proof: Suppose $X$ is countable. Let $D \in \mathcal{B}(\mathcal{P}(X))$. From (6.1)

\[
P(\pi_{t+1} \in D | \pi_s, u_s, s \leq t) \\
= P(F(\pi_t, y_{t+1}, u_t) \in D | \pi_s, u_s, s \leq t) \\
= \sum_{y \in Y} P(F(\pi_t, y_{t+1}, u_t) \in D, y_{t+1} = y | \pi_s, u_s, s \leq t) \\
= \sum_{y \in Y} 1 \left\{ F(\pi_t, y_{t+1} = y, u_t) \in D \right\} P(y_{t+1} = y | \pi_t, u_t) \\
= \sum_{y} 1 \left\{ \frac{\sum_{x \in X} \pi(\pi_t(x))P(y_{t+1} = y | \pi_t(x) = x, u_t)}{\sum_{x \in X} \sum_{y} \pi(\pi_t(x))P(y_{t+1} = y | \pi_t(x) = x, u_t)} \right\} \left( \sum_{x} \sum_{y} P(y_{t+1} = y | x_{t+1})P(x_{t+1} | x_t, u_t) \pi_t(x) \right)
\]

(6.2)

We need to show that the expression $P(d\pi_{t+1} \in D | \pi_t, u_t)$ is a regular conditional probability measure; that is, for every fixed $D$, this is a measurable function on $\mathcal{P}(X) \times \mathcal{U}$ and for every $\pi_t, u_t$, it is a conditional probability measure on $\mathcal{P}(X)$. The rest of the proof follows in Section 6.3.

Let the cost function to be minimized be

\[
E_{x_0}^{II} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right],
\]

where $E_{x_0}^{II} [\cdot]$ denotes the expectation over all sample paths with initial state given by $x_0$ under policy $II$. We can transform the system into a fully observed Markov model as follows. Using the law of the iterated expectation, write the total cost as

Since

\[
E_{x_0}^{II} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right] = E_{x_0}^{II} \left[ \sum_{t=0}^{T-1} E \left[ c(x_t, u_t) | I_t \right] \right].
\]

Given a policy $u_t = \gamma_t(I_t)$, we have that

\[
E \left[ c(x_t, u_t) | I_t \right] = E \left[ E \left[ c(x_t, u_t) | I_t = i_t \right] \right] = E \left[ \sum P(x_t | I_t = i_t)(c(x_t, \gamma_t(i_t))) | I_t = i_t \right],
\]

For each realization $I_t = i_t$, $u_t$ is a fixed number and thus we can view $E \left[ c(x_t, u_t) | I_t = i_t \right] = \sum P(x_t | I_t = i_t)(c(x_t, u_t)) =: \tilde{c}(\pi_t, u_t)$. We can then write the per-stage cost function as

\[
\tilde{c}(\pi, u) = \sum_{x} c(x, u)\pi(x), \quad \pi \in \mathcal{P}
\]

(6.3)

It follows then that $\mathcal{P}$, $\mathcal{U}$, $\mathcal{K}$, $\tilde{c}$ defines a completely observable controlled Markov process. Here $\mathcal{K}$ is the transition kernel defined with (6.2).

Observe that an admissible control policy will select $u_t$ as a function of $I_t$ (that is, its realization $i_t$). However, we know that for a finite horizon problem, by Blackwell’s theorem Theorem 5.1, an optimal policy will only use $\pi_t$ (since $\pi_t, u_t$ forms a controlled Markov chain), and therefore without any loss, we can restrict our search space to policies which are Markov (that is which only use $\pi_t$ and $t$).

Thus, one can obtain the optimal solution by using the filtering equation as a sufficient statistic in a centralized setting, as Markov policies (policies that use the Markov state as their sufficient statistics) are optimal for control of Markov chains, under well-studied sufficiency conditions for the existence of optimal selectors.

We call the control policies which use $\pi$ as their information to generate control as separated control policies. This will be made more explicit in the context of linear Gaussian systems. A Gaussian probability measure can be uniquely identified by knowing the mean and the covariance of the Gaussian random variable. This makes the analysis for estimating a Gaussian
random variable particularly simple to perform, since the conditional estimate of a partially observed (through an additive Gaussian noise) Gaussian random variable is a linear function of the observed variable and the non-linear filtering equation [6.1] becomes significantly simpler. Recall that a Gaussian measure with mean \( \mu \) and covariance matrix \( K_{XX} \) has the following density:

\[
p(x) = \frac{1}{(2\pi)^{2} |K_{XX}|^{1/2}} e^{-1/2((x-\mu)^{T}K_{XX}^{-1}(x-\mu))},
\]

and thus it suffices to compute the mean and the covariance matrix to compute the density.

6.2 The Linear Quadratic Gaussian (LQG) Problem and Kalman Filtering

6.2.1 A supporting result on estimation

**Lemma 6.2.1** Let \( X, Y \) be random variables with finite second moments and \( R > 0 \). The following holds

\[
\inf_{g} E[(X - g(Y))^{T} R(X - g(Y))] = E[(X - G(Y))^{T} R(X - G(Y))],
\]

where \( G(y) = E[X|Y = y] \) almost surely.

**Proof:** Let \( G(y) = E[X|Y = y] + h(y) \), for some measurable \( h \); we then have the following through the law of the iterated expectations:

\[
E[(X - E[X|Y] - h(Y))^{T} R(X - E[X|Y] - h(Y))]
E[E[(X - E[X|Y] - h(Y))^{T} R(X - E[X|Y] - h(Y))|Y = y]]
= E[E[(X - E[X|y])^{T} R(X - E[X|y])|y] + E[h^{T}(Y)Rh(Y)|y] + 2E[(X - E[X|y])^{T} R h(y)|y]]
= E[(X - E[X|y])^{T} R(X - E[X|y])|y] + E[h^{T}(Y)Rh(Y)|y] + 2E[(X - E[X|y])^{T} y Rh(y)]
\geq E[(X - E[X|y])^{T} R(X - E[X|y])|y]
\geq E[(X - E[X|Y])^{T} R(X - E[X|Y])]
\]

\( \Box \)

**Remark 6.1.** We note that the above admits a Hilbert space interpretation or formulation: Let \( H \) denote the space random variables (defined on a probability space) on which an inner product \( \langle X, Y \rangle = E[X^{T}RY] \) is defined; this defines a Hilbert space. Let \( M_{H} \) be a subspace of \( H \), the closed subspace of random variables that are measurable on \( \sigma(Y) \) which have finite second moments. Then, the Projection theorem leads to the observation that, an optimal estimate \( g(Y) \) minimizing \( ||X - g(Y)||_{2}^{2} \), denoted here by \( G(Y) \), is one which satisfies:

\[
\langle X - G(Y), h(Y) \rangle = E[X - G(Y)^{T} Rh(Y)] = 0, \forall h \in M_{H}
\]

The conditional expectation satisfies this since:

\[
E[(X - E[X|Y])^{T} Rh(Y)] = E[E[(X - E[X|y])^{T} Rh(y)|Y = y]] = E[E[(X - E[X|y])^{T} y Rh(y)] = 0,
\]

since \( Pa.s., E[(X - E[X|y])^{T} y] = 0 \).

6.2.2 The Linear Quadratic Gaussian Problem

Consider the following linear system:

\[
x_{t+1} = Ax_{t} + Bu_{t} + w_{t},
\]
Partially Observed Markov Decision Processes, Non-linear Filtering and the Kalman Filter

and

\[ y_t = Cx_t + v_t, \]

where \( x \in \mathbb{R}^n, u \in \mathbb{R}^m \) and \( w \in \mathbb{R}^n, y \in \mathbb{R}^p, v \in \mathbb{R}^p \). Suppose \( \{w_t, v_t\} \) are i.i.d. random Gaussian vectors with given covariance matrices \( E[w_t w_t^T] = W \) and \( E[v_t v_t^T] = V \) for all \( t \geq 0 \).

The goal is to obtain

\[ \inf_{\Pi} J(\Pi, \mu_0), \]

where

\[ J(\Pi, \mu_0) = E_{\mu_0} \left[ \sum_{t=0}^{T-1} x_t^T Q x_t + u_t^T Ru_t + x_t^T Q x_t \right], \]

with \( R > 0 \) and \( Q, Q_T \geq 0 \) (that is, these matrices are positive definite and positive semi-definite).

Building on Lemma 6.2.1, we will show that the optimal control is linear in its expectation and has the form:

\[ u_t = -(BK_{t+1}B + R)^{-1}B^T K_{t+1} A \mathbb{E}[x_t | I_t] \]

where \( K_t \) solves the Discrete-Time Riccati Equation:

\[ K_t = Q + A^T K_{t+1} A - A^T K_{t+1} B ((BK_{t+1}B + R)^{-1}) B^T K_{t+1} A, \]

with final condition \( K_T = Q_T \). The analysis will build on the following observations.

### 6.2.3 Estimation and Kalman Filtering

In this section, we discuss the control-free setup and derive the celebrated Kalman Filter. In the following to make certain computations more explicit and easier to follow, we will use capital letters to denote the random variables and small letters for the realizations of these variables.

For a linear Gaussian system, the state process has a Gaussian probability measure. A Gaussian probability measure can be uniquely identified by knowing the mean and the covariance of the Gaussian random variable. This makes the analysis for estimating a Gaussian random variable particularly simple to perform, since the conditional estimate of a partially observed (through an additive Gaussian noise) Gaussian random variable is a linear function of the observed variable.

Recall that a Gaussian measure with mean \( \mu \) and covariance matrix \( K_{XX} \) has the following density:

\[ p(x) = \frac{1}{(2\pi)^{n/2}|K_{XX}|^{1/2}} e^{-1/2 (x - \mu)^T K_{XX}^{-1} (x - \mu)} \]

**Lemma 6.2.2** Let \( X, Y \) be zero-mean Gaussian processes. Then \( E[X | Y = y] \) is linear in \( y \): With \( \Sigma_{XY} = E[XY^T] \),

\[ E[X | Y = y] = \Sigma_{XY} \Sigma_{YY}^{-1} y, \]

and

\[ E[(X - E[X | Y])(X - E[X | Y])^T] = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T =: D. \]

In particular,

\[ E[(X - E[X | Y])(X - E[X | Y])^T | Y = y] \]

does not depend on the realization \( y \) of \( Y \) and is equal to \( D \).

We note that if the random variables are not-zero mean, one needs to add a constant correction term making the estimate an affine function.
Proof: By Baye’s rule and the fact that the processes admit densities: \( p(x|y) = p(x, y)/p(y) \). Let \( K_{XY} := \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \).

It follows that \( K_{XY}^{-1} \) is also symmetric (since the eigenvectors are the same and the eigenvalues are inverted) and given by:

\[
K_{XY}^{-1} = \begin{bmatrix} \Psi_{XX} & \Psi_{XY} \\ \Psi_{YX} & \Psi_{YY} \end{bmatrix},
\]

Thus, for some normalization constant \( C \),

\[
p(x, y) = Ce^{-1/2(x^T \Psi_{XX} x + 2x^T \Psi_{XY} y + y^T \Psi_{YY} y - y^T \Psi_{YX} x)} = e^{-1/2(y^T K_{YY} y)}.
\]

By the completion of the squares method for the expression, for some matrix \( D \) we obtain

\[
(x^T \Psi_{XX} x + 2x^T \Psi_{XY} y + y^T \Psi_{YY} y - y^T \Psi_{YX} x) = (x - Hy)^T D^{-1}(x - Hy) + Q(y),
\]

it follows that \( H = -\Psi_{XX}^{-1} \Psi_{XY} \) and \( D = \Psi_{YY}^{-1} \). Since \( K_{XY} K_{YY}^{-1} = I \) (and thus \( \Psi_{XX} K_{XY} + \Psi_{XY} K_{YY} = 0 \)), this is also equal to \( \Sigma_{XY} \Sigma_{YY}^{-1} \). Here \( Q(y) \) is a quadratic expression in \( y \). As a result, one obtains

\[
p(x|y) = e^{Q(y)} e^{-1/2(x - Hy)^T D^{-1}(x - Hy)},
\]

Since \( \int p(x|y) dx = 1 \), it follows that \( Q(y) = \frac{1}{(2\pi)^{\frac{T}{2}} |D|^{1/2}} \) and is in fact independent of \( y \). Note also that

\[
E[(X - E[X|Y])(X - E[X|Y])^T] = E[XX^T] - E[(E[X|Y])(E[X|Y])^T] = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T
\]

Remark 6.2. The fact that \( Q(y) \) above does not depend on \( y \) reveals an interesting result that the conditional covariance of \( X - E[X|Y] \) viewed as a Gaussian random variable is equal for all \( y \) values. This is a crucial fact that will be utilized in the derivation of the Kalman Filter.

Remark 6.3. Even if the random variables \( X, Y \) are not Gaussian (but zero-mean), through another Hilbert space formulation and an application of the Projection Theorem, the expression \( \Sigma_{XY} \Sigma_{YY}^{-1} y \) is the best linear estimate, that is the solution to \( \inf_{K} E[(X - KY)^T (X - KY)] \). One can naturally generalize this for random variables with non-zero mean.

We will derive the Kalman filter in the following. The following two lemma are instrumental.

Lemma 6.2.3 If \( E[X] = 0 \) and \( Z_1, Z_2 \) are orthogonal zero-mean Gaussian processes (with \( E[Z_1^T Z_2] = 0 \)), then

\[
E[X|Z_1 = z_1, Z_2 = z_2] = E[X|Z = z_1] + E[X|Z = z_2].
\]

Proof: The proof follows by writing \( z = [z_1, z_2]^T \), noting that \( \Sigma_{ZZ} \) is diagonal and \( E[X|z] = \Sigma_{XZ} \Sigma_{ZZ}^{-1} z \).

Lemma 6.2.4 \( E[(X - E[X|Y])(X - E[X|Y])^T] \) is given by \( D = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T \) above.

Proof: Proof follows from direct computation (as in (6.4)). Note that

\[
\]

since \( X - E[X|Y] \) is orthogonal to \( E[X|Y] \).

Now, we can move on to the derivation of the Kalman Filter.

Consider

\[
x_{t+1} = Ax_t + w_t, \quad y_t = Cx_t + v_t,
\]
with $E[w_tw_t^T] = W$ and $E[v_tv_t^T] = V$ where $\{w_t\}$ and $\{v_t\}$ are mutually independent i.i.d. zero-mean Gaussian processes.

Define
\[
m_t = E[x_t|y_{0,t-1}]
\]
\[
\Sigma_{t|t-1} = E[(x_t - E[x_t|y_{0,t-1}])(x_t - E[x_t|y_{0,t-1}])^T|y_{0,t-1}]
\]
and note that since the estimation error covariance does not depend on the realization $y_{0,t-1}$ (see Remark 6.2), we write also
\[
\Sigma_{t|t-1} = E[(x_t - E[x_t|y_{0,t-1}])(x_t - E[x_t|y_{0,t-1}])^T]
\]

**Theorem 6.2.1** The following holds:
\[
m_{t+1} = Am_t + A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}(y_t - Cm_t)
\]
\[
\Sigma_{t+1|t} = A\Sigma_{t|t-1}A^T + W - (A\Sigma_{t|t-1}C^T)(C\Sigma_{t|t-1}C^T + V)^{-1}(C\Sigma_{t|t-1}A^T)
\]
with
\[
m_0 = E[x_0]
\]
and
\[
\Sigma_{0|0-1} = E[x_0x_0^T]
\]

**Proof:** With $x_{t+1} = Ax_t + w_t$, the following hold:
\[
m_{t+1} = E[Ax_t + w_t|y_{0,t}] = E[Ax_t|y_{0,t}] = E[Am_t + A(x_t - m_t)x_t|y_{0,t}]
\]
\[
= Am_t + E[A(x_t - m_t)|y_{0,t-1}, y_t - E[y_t|y_{0,t-1}]]
\]
\[
= Am_t + E[A(x_t - m_t)|y_{0,t-1}] + E[A(x_t - m_t)|y_t - E[y_t|y_{0,t-1}]]
\]
\[
= Am_t + E[A(x_t - m_t)|y_{0,t-1}] + E[A(x_t - m_t)|y_t - y_{0,t-1}]
\]
(6.5)
\[
= Am_t + E[A(x_t - m_t)|x_t + v_t - E[Cx_t + v_t|y_{0,t-1}]]
\]
\[
= Am_t + E[A(x_t - m_t)|C(x_t + v_t) + v_t]
\]
(6.6)

In the above, (6.5) follows from Lemma 6.2.3. In the above, we also use the fact that $w_t$ is orthogonal to $y_{0,t}$. Let $X = A(x_t - m_t)$ and $Y = y_t - E[y_t|y_{0,t-1}] = y_t - Cm_t = C(x_t - m_t) + v_t$. Then, by Lemma 6.2.2, $E[X|Y] = \Sigma_{XY} \Sigma_{YY}^{-1} Y$ and thus, \[
m_{t+1} = Am_t + A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}(y_t - Cm_t)
\]
Likewise,
\[
x_{t+1} - m_{t+1} = A(x_t - m_t) + w_t - A\Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}(y_t - Cm_t),
\]
leads to, after a few lines of calculations:
\[
\Sigma_{t+1|t} = A\Sigma_{t|t-1}A^T + W - (A\Sigma_{t|t-1}C^T)(C\Sigma_{t|t-1}C^T + V)^{-1}(C\Sigma_{t|t-1}A^T)
\]

The above is the celebrated Kalman filter.

Define now
\[
\tilde{m}_t = E[x_t|y_{0,t}] = m_t + E[x_t - m_t|y_{0,t}] = m_t + E[x_t - m_t|y_t - E[y_t|y_{0,t-1}]].
\]
It follows then that
\[
\tilde{m}_t = A\tilde{m}_{t-1} + \Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}(y_t - Cm_t)
\]
Note that we can also write $m_t = A\tilde{m}_{t-1}$. 

We observe that the zero-mean variable $x_t - \tilde{m}_t$ is orthogonal to $I_t$, in the sense that the error is independent of the information available at the controller, and since the information available is Gaussian, independence and orthogonality are equivalent.

We observe that the recursion in Theorem 6.2.1 is identical to the recursions in Theorem 5.3.2 with writing $A = A^T$, $W = Q$, $C^T = B$. This leads to the following result.

**Theorem 6.2.2** Suppose $(A^T, C^T)$ is controllable (this is equivalent to saying that $(A, C)$ is observable) and $V > 0$. Then, the recursions for the covariance matrices $\Sigma_{\cdot\cdot}$ in Theorem 6.2.1 admit a fixed point. If, in addition, with $W = BB^T$ and that $(A^T, B^T)$ is observable (that is $(A, B)$ is controllable), the fixed point solution is unique, and is positive definite.

**Remark 6.4.** The above suggest that if the observations are sufficiently informative, then the Kalman filter converges to a solution, even in the absence of an irreducibility condition on the original state process $x_t$; under irreducibility, the solution is unique. This observation has been extended in the non-linear filtering context [120].

### 6.2.4 Optimal Control of Partially Observed LQG Systems

With this observation, we can reformulate the quadratic optimization problem as a function of $\tilde{m}_t$, $u_t$ and $x_t - \tilde{m}_t$ as follows. First, in the controlled case, let us define

$$\tilde{m}_t = E[x_t|y_{0,t}, u_{0,t-1}].$$

Let $I_t = \{y_{[0,t]}, u_{[0,t-1]}\}$ Observe now that

$$E[x_t^TQx_t] = E[(x_t - \tilde{m}_t + \tilde{m}_t)^TQ(x_t - \tilde{m}_t + \tilde{m}_t)]$$

$$= E[(x_t - \tilde{m}_t)^TQ(x_t - \tilde{m}_t)] + E[\tilde{m}_t^TQ\tilde{m}_t] + 2E[(x_t - \tilde{m}_t)^TQ\tilde{m}_t]$$

$$= E[(x_t - \tilde{m}_t)^TQ(x_t - \tilde{m}_t)] + E[\tilde{m}_t^TQ\tilde{m}_t] + 2E[E[(x_t - \tilde{m}_t)^TQ\tilde{m}_t|I_t]]$$

$$= E[(x_t - \tilde{m}_t)^TQ(x_t - \tilde{m}_t)] + E[\tilde{m}_t^TQ\tilde{m}_t]$$

(6.8)

since $E[(x_t - \tilde{m}_t)^TQ\tilde{m}_t|I_t] = 0$ by the orthogonality property of the conditional estimation error (recall that $\tilde{m}_t$ is a function of $I_t$). In particular, the cost:

$$J(\Pi, \mu_0) = E^{\Pi}_0 \left[ \sum_{t=0}^{T-1} x_t^TQx_t + u_t^TRu_t + x_T^TQ_Tx_T \right],$$

writes as:

$$E^{\Pi}_0 \left[ \sum_{t=0}^{T-1} \tilde{m}_t^TQ\tilde{m}_t + u_t^TRu_t + \tilde{m}_t^TQ_T\tilde{m}_T \right] + E^{\Pi}_0 \left[ \sum_{t=0}^{T-1} (x_t - \tilde{m}_t)^TQ(x_t - \tilde{m}_t) \right]$$

$$+ E^{\Pi}_0 \left[ (x_T - \tilde{m}_T)^TQ_T(x_T - \tilde{m}_T) \right]$$

(6.9)

for the fully observed system:

$$\tilde{m}_t = A\tilde{m}_{t-1} + Bu_t + \tilde{w}_t,$$

with

$$\tilde{w}_t = \Sigma_{\cdot\cdot}C^T(C\Sigma_{\cdot\cdot}C^T + V)^{-1}(y_t - CA\tilde{m}_{t-1})$$

Furthermore, due to the orthogonality principle, the estimation error in (6.9) does not depend on the control policy $\Pi$ so that the expected cost writes as

$$E^{\Pi}_0 \left[ \sum_{t=0}^{T-1} \tilde{m}_t^TQ\tilde{m}_t + u_t^TRu_t + \tilde{m}_t^TQ_T\tilde{m}_T \right] + E^{\Pi}_0 \left[ \sum_{t=0}^{T-1} (x_t - \tilde{m}_t)^TQ(x_t - \tilde{m}_t) \right]$$

$$+ E^{\Pi}_0 \left[ (x_T - \tilde{m}_T)^TQ_T(x_T - \tilde{m}_T) \right]$$

(6.10)
Thus, the optimal control problem is equivalent to the control of the fully observed state $\tilde{m}_t$, with additive time-varying independent Gaussian noise process. Here, that the error term $(x_t - \tilde{m}_t)$ does not depend on the control policy is a consequence of what is known as the lack of dual effect of control: the control actions up to any time $t$ do not affect the state estimation error process for the future time stages. Using our earlier analysis, it follows then that the optimal control has the following form for all time stages:

$$ u_t = -(BP_{t+1}B^T + R)^{-1}B^TP_{t+1}AE[x_t|I_t], $$

with $P_t$ generated as in Theorem 5.3.1 and $P_T = Q_T$.

**Remark 6.5.** In [13], dual effect is introduced as the property that $(x_t - \tilde{m}_t)$ is (statistically) independent from the past applied control actions. A more general statement would be that $(x_t - \tilde{m}_t)$ does not depend on the past control policies in that the control policies do not alter the realization of the random variable $(x_t - \tilde{m}_t)$, or the applied actions do not affect the realizations of $(x_t - \tilde{m}_t)$. This distinction can be important in certain applications in networked control systems [citation to be included].

Here we observe that the optimal control policy is the same as that in the fully observed setup in Theorem 5.3.1 except that the state is replaced with its estimate. This phenomenon is known as the separation of estimation and control, and for this particular case, a more special version of it, known as the certainty equivalence principle. As observed above, the absence of dual effect plays a key part in this analysis leading the separation of estimation and control principle, in taking $E[(x_t - \tilde{m})^T Q(x_t - \tilde{m})]$ out of the optimization over control policies, since it does not depend on the policy.

Thus, the optimal cost will be

$$ E[\tilde{m}_0^T P_0 \tilde{m}_0] + E[\sum_{k=0}^{T-1} \tilde{w}_k^T P_{k+1} \tilde{w}_k + E(x_t - \tilde{m}_t)^T Q(x_t - \tilde{m}_t)] + E(x_T - \tilde{m}_T)^T Q_T(x_T - \tilde{m}_T) $$

In the above problem, we observed that the optimal control has a separation structure: The controller first estimates the state, and then applies its control action, by regarding the estimate as the state itself. Typically, when the dual effect is absent, separation principle observed above applies. In most general problems, however, the dual effect of the control is present. That is, depending on the control policy, the estimation quality at the controller regarding future states will be affected. As an example, consider a linear system controlled over an erasure channel, where the controller applies a control, but does not know if the packet reaches the destination or not. In this case, the control signal which was intended to be sent, does affect the estimation error quality [69][110].

### 6.3 On the Controlled Markov Construction in the Space of Probability Measures and Extension to General Spaces

In Section 6.1, we observed that we can replace the state with a probability measure valued state. It is important to provide notions of convergence and continuity on the spaces of probability measures to be able to apply the machinery of Chapter 5.

#### 6.3.1 Measurability Issues, Proof of Theorem 6.1.1 and its Extension to Polish Spaces

In 6.1, we need to show that the expression $P(\pi_{t+1} \in D|\pi_t, u_t)$ is a regular conditional probability measure; that is, for every fixed $D$, this is a measurable function on $\mathcal{P}(\mathcal{X}) \times \mathcal{U}$ and for every $\pi_t, u_t$, it is a conditional probability measure on $\mathcal{P}(\mathcal{X})$. The expression $\sum_{\pi_{t-1}(x_{t-1})} \mathbb{P}(y_t|x_t)P(x_t|x_{t-1}, u_{t-1})$ is a measurable function of $\pi_t, u_t$. The measurability follows from the following useful results: A proof of of the first result can be found in [4] (see Theorem 15.13 in [4] or p. 215 in [25]).
Theorem 6.3.1 Let $\mathbb{S}$ be a Polish space and $M$ be the set of all measurable and bounded functions $f : \mathbb{S} \rightarrow \mathbb{R}$. Then, for any $f \in M$, the integral
\[
\int \pi(dx) f(x)
\]
defines a measurable function on $\mathcal{P}(\mathbb{S})$ under the topology of weak convergence.

This is a useful result since it allows us to define measurable functions in integral forms on the space of probability measures when we work with the topology of weak convergence. The second useful result follows from Theorem 6.3.1 and Theorem 2.1 of Dubins and Freedman [42] and Proposition 7.25 in Bertsekas and Shreve [17].

Theorem 6.3.2 Let $\mathbb{S}$ be a Polish space. A function $F : \mathcal{P}(\mathbb{S}) \rightarrow \mathcal{P}(\mathbb{S})$ is measurable on $\mathcal{B}(\mathcal{P}(\mathbb{S}))$ (under weak convergence), if for all $B \in \mathcal{B}(\mathbb{S})$ $(F(\cdot))(B) : \mathcal{P}(\mathbb{S}) \rightarrow \mathbb{R}$ is measurable under weak convergence on $\mathcal{P}(\mathbb{S})$, that is for every $B \in \mathcal{B}(\mathbb{S})$, $(F(\pi))(B)$ is a measurable function when viewed as a function from $\mathcal{P}(\mathbb{S})$ to $\mathbb{R}$.

The discussion in Section 6.1 also applies to settings where $\mathcal{X}, \mathcal{Y}, \mathcal{U}$ are more general Polish spaces. In particular, with $\pi$ denoting a conditional probability measure $\pi_t(A) = P(x_t \in A | I_t)$, we can define a new cost as $c(\pi, u) = \sum_x c(x, u) \pi(x)$, $\pi \in \mathcal{P}$.

In Chapter 5, we had assumed that the state belongs to a complete, separable, metric space; therefore, the existence of optimal policies and the dynamic programming arguments follows the analysis in Chapter 3.

This discussion is useful to establish that under weak convergence topology $(\pi_t, u_t)$ forms a controlled Markov chain for the dynamic programming purposes.

6.3.2 Continuity Properties of Belief-MDP

TO BE ADDED.

6.3.3 Concavity of the Value Function in the Priors

The following theorem establishes concavity of the optimal cost in a single-stage stochastic control problem over the space of initial distributions and this also applies for multi-stage setups. For the special case with $c(x, u) = (x - u)^2$, this result was established in [133] and the case with observation channels was established in Theorem 4.3.1 in [139].

Theorem 6.3.3 Let $\int c(x, \gamma(y)) PQ(dx, dy)$ exist for all $\gamma \in \Gamma$ and $P \in \mathcal{P}(\mathcal{X})$. Then,
\[
J^*(P, Q) = \inf_{\gamma \in \Gamma} E_P^Q \gamma[c(x, \gamma(y))]
\]
is concave in $P$.

Proof. For $a \in [0, 1]$ and $P', P'' \in \mathcal{P}(\mathcal{X})$ we let $P = aP' + (1 - a)P''$. Note that $PQ = aP'Q + (1 - a)P''Q$. We have
\[
J(aP' + (1 - a)P'', Q) = J(P, Q)
\]
\[
= \inf_{\gamma \in \Gamma} E_P^Q \gamma[c(x, \gamma(y))]
\]
\[
= \inf_{\gamma \in \Gamma} \int c(x, \gamma(y)) PQ(dx, dy)
\]
\[
= \inf_{\gamma \in \Gamma} \left( a \int c(x, \gamma(y)) P'Q(dx, dy) + (1 - a) \int c(x, \gamma(y)) P''Q(dx, dy) \right)
\[
\begin{align*}
&\geq \inf_{\gamma \in \Gamma} \left( a \int c(x, \gamma(y)) P' Q(dx, dy) \right) \\
+ &\inf_{\gamma \in \Gamma} \left( (1 - a) \int c(x, \gamma(y)) P'' Q(dx, dy) \right) \\
= &a J(P', Q) + (1 - a) J(P'', Q)
\end{align*}
\]

6.4 Bibliographic Notes

Earlier work on separation results for partially observed Markov Decision Processes include [142], [98].

6.5 Exercises

**Exercise 6.5.1** Consider a linear system with the following dynamics:

\[
x_{t+1} = ax_t + u_t + w_t,
\]

and let the controller have access to the observations given by:

\[
y_t = p_t (x_t + v_t).
\]

Here \( \{w_t, v_t, t \in \mathbb{Z}\} \) are independent, zero-mean, Gaussian random variables, with variances \( E[w^2] \) and \( E[v^2] \). The controller at time \( t \in \mathbb{Z} \) has access to \( I_t = \{y_s, u_s, p_t \mid s \leq t - 1\} \cup \{y_t\} \). Here \( p_t \) is an i.i.d. Bernoulli process such that \( p_t = 1 \) with probability \( p \).

The initial state has a Gaussian distribution, with zero mean and variance \( E[x_0^2] \), which we denote by \( \nu_0 \). We wish to find for some \( r > 0 \):

\[
\inf_{\Pi} J(x_0, \Pi) = E_{\nu_0} \left[ \sum_{t=0}^{3} x_t^2 + ru_t^2 \right],
\]

Compute the optimal control policy and the optimal cost. It suffices to provide a recursive form.

*Hint:* Show that the optimal control has a separation structure. Compute the conditional estimate through a revised Kalman Filter due to the presence of \( p_t \).

**Exercise 6.5.2** Let \( X, Y \) be \( \mathbb{R}^n \) and \( \mathbb{R}^m \) valued zero-mean random vectors defined on a common probability space, which have finite covariance matrices. Suppose that their probability measures are given by \( P_X \) and \( P_Y \) respectively.

Find

\[
\inf_K E[(X - KY)^T (X - KY)],
\]

that is find the best linear estimator of \( X \) given \( Y \) and the resulting estimation error.

*Hint:* You may pose the problem as a Projection Theorem problem.

**Exercise 6.5.3 (Optimal Machine Repair)** Consider a Markov Decision Problem set up as follows. Let there be two possible states that a machine can take: \( \mathcal{X} = \{0, 1\} \), where 0 is the good state and 1 is the bad (‘system is down’) state. Let \( \mathcal{U} = \{0, 1\} \), where 0 is the ’do nothing’ control and 1 is the ’repair’ control. Suppose that the transition probabilities are given by:

\[
\begin{align*}
P(X_{t+1} = 1 | X_t = 1, U_t = 0) &= 1 - \eta, \quad P(X_{t+1} = 0 | X_t = 1, U_t = 0) = \eta > 0 \\
P(X_{t+1} = 1 | X_t = 1, U_t = 1) &= 1, \quad P(X_{t+1} = 0 | X_t = 1, U_t = 0) = 0
\end{align*}
\]
Thus, \( \eta \) is the failure probability and \( \alpha \) is the success probability in the event of a repair.

The controller has access only to measurement variables \( Y_0, \ldots, Y_t \) and \( U_0, \ldots, U_{t-1} \), at time \( t \), where the measurements are generated by a binary symmetric channel: \( P(Y = X) = 1 - \epsilon \) and \( P(Y \neq X) = \epsilon \) for all \( X, Y \) realizations. The per-stage cost function \( c(x, u) \) is given by \( c(0, 0) = 0, c(1, 0) = C, c(0, 1) = c(1, 1) = R \) with \( 0 < R < C \). Show that there exists an optimal control policy for both finite-horizon as well as infinite horizon discounted cost problems. Is the optimal policy of threshold type?

Exercise 6.5.4 (Zero-Delay Source Coding) Let \( \{x_t\}_{t \geq 0} \) be an \( \mathbb{X} \)-valued discrete-time Markov process where \( \mathbb{X} \) can be a finite set or \( \mathbb{R}^n \). Let there be an encoder which encodes (quantizes) the source samples and transmits the encoded versions to a receiver over a discrete noiseless channel with input and output alphabet \( \mathcal{M} = \{1, 2, \ldots, M\} \), where \( M \) is a positive integer. The encoder policy \( \Pi \) is a sequence of functions \( \{\eta_t\}_{t \geq 0} \) with \( \eta_t : \mathcal{M}^t \times (\mathbb{X})^{t+1} \rightarrow \mathcal{M} \). At time \( t \), the encoder transmits the \( M \)-valued message

\[
q_t = \eta_t(I_t)
\]

with \( I_0 = x_0 \) and \( I_t = (q_{[0,t-1]}, x_{[0,t]}) \) for \( t \geq 1 \), where. The collection of all such zero-delay encoders is called the set of admissible quantization policies and is denoted by \( \Pi_A \). A zero-delay receiver policy is a sequence of functions \( \gamma = \{\gamma_t\}_{t \geq 0} \) of type \( \gamma_t : \mathcal{M}^{t+1} \rightarrow \mathbb{U} \), where \( \mathbb{U} \) denotes the finite reconstruction alphabet. Thus

\[
u_t = \gamma_t(q_{[0,t]}), \quad t \geq 0.
\]

For the finite horizon setting the goal is to minimize the average cumulative cost (distortion)

\[
J_{\pi_0}(\Pi, \gamma, T) = E_{\pi_0}^{\Pi, \gamma} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \right], \tag{6.12}
\]

for some \( T \geq 1 \), where \( c_0 : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{R} \) is a nonnegative cost (distortion) function, and \( E_{\pi_0}^{\Pi, \gamma} \) denotes expectation with initial distribution \( \pi_0 \) for \( x_0 \) and under the quantization policy \( \Pi \) and receiver policy \( \gamma \). Show that an optimal encoder uses a sufficient statistic, in particular, it uses \( P(dx_t | q_{[0,t-1]}) \) and the time information, for optimal performance. See [126], [124] and [137] for relevant discussions, among many other references.
The Average Cost Problem

Consider the following average cost problem of finding

$$\inf_{\Pi} J(x, \Pi) = \inf_{\Pi} \lim_{T \to \infty} \frac{1}{T} E_x^{\Pi} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right]$$

This is an important problem in applications where one is concerned about the long-term behaviour, unlike the discounted cost setup where the primary interest is in the short-term time stages.

7.1 Average Cost and the Average Cost Optimality Equation (ACOE) or Inequality (ACOI)

To study the average cost problem, one approach is to establish the existence of an Average Cost Optimality Equation.

**Definition 7.1.1** The collection of functions $g : X \to \mathbb{R}, h : X \to \mathbb{R}, f : X \to U$ is a canonical triplet if for all $x \in X$,

$$g(x) = \inf_{u \in U} \int g(x') P(dx'|x, u)$$

$$g(x) + h(x) = \inf_{u \in U} \left( c(x, u) + \int h(x') P(dx'|x, u) \right)$$

with

$$g(x) = \int g(x') P(dx'|x, f(x))$$

$$g(x) + h(x) = \left( c(x, f(x)) + \int h(x') P(dx'|x, f(x)) \right)$$

**Theorem 7.1.1** [Verification Theorem] Let $g, h, f$ be a canonical triplet. a) If $g$ is a constant and

$$\lim_{n \to \infty} \frac{1}{n} E_x^\Pi [h(x_n)] = 0,$$

(7.1)

for all $x$ and under every policy $\Pi$, then the stationary deterministic policy $\Pi^* = \{ f, f, \cdots \}$ is optimal so that

$$g = J(x, \Pi^*) = \inf_{\Pi \in \Pi_A} J(x, \Pi)$$

where

$$J(x, \Pi) = \lim_{T \to \infty} \frac{1}{T} E_x^{\Pi} \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right].$$
Furthermore,
\[ \frac{1}{n} E_x^H \sum_{t=1}^{n} [c(x_{t-1}, u_{t-1})] - g \leq \frac{1}{n} \left( E_x^H |h(x_n)| + |h(x)| \right) \to 0 \]  \hspace{1cm} (7.2)

b) If \( g \), considered above, is not a constant and depends on \( x \), then under any policy \( \Pi \)
\[ \frac{1}{T} E_x^H \left( \sum_{t=0}^{T-1} g(x_t) \right) \leq \inf_{\Pi} \sup_{H} \frac{1}{N} E_x^H \left( \sum_{t=0}^{T-1} c(x_t, u_t) \right), \]
provided that (7.1) holds. Furthermore, \( \Pi^* = \{ f \} \) is optimal.

**Proof:** We prove (a); (b) follows from a similar reasoning. For any policy \( \Pi \), by the iterated expectations Theorem 4.1.3
\[ E_x^H \left[ \sum_{t=1}^{n} h(x_t) - E_x^H [h(x_t)|x_{0:t-1}, u_{0:t-1}] \right] = 0. \]

Now,
\[ E_x^H [h(x_t)|x_{0:t-1}, u_{0:t-1}] = \int_y h(y) P(x_t \in dy|x_{t-1}, u_{t-1}) \]
\[ = c(x_{t-1}, u_{t-1}) + \int_y h(y) P(dy|x_{t-1}, u_{t-1}) - c(x_{t-1}, u_{t-1}) \]
\[ \geq \min_{u_{t-1} \in \mathcal{U}} \left( c(x_{t-1}, u_{t-1}) + \int_y h(y) P(dy|x_{t-1}, u_{t-1}) \right) - c(x_{t-1}, u_{t-1}) \]
\[ = g + h(x_{t-1}) - c(x_{t-1}, u_{t-1}) \]
\[ \text{The above hold with equality if } \Pi^* = \{ f \} \text{ is adopted since } \Pi^* \text{ provides the pointwise minimum}. \]

Hence, for any policy \( \Pi \)
\[ 0 \leq \frac{1}{n} E_x^H \sum_{t=1}^{n} [h(x_t) - g - h(x_{t-1}) + c(x_{t-1}, u_{t-1})] \]

and
\[ g \leq \frac{1}{n} E_x^H [h(x_n)] - \frac{1}{n} E_x^H [h(x_0)] + \frac{1}{n} E_x^H \sum_{t=1}^{n} c(x_{t-1}, u_{t-1})]. \]

Furthermore, equality holds under \( \Pi^* \) so that
\[ g = E_x^{H*} [h(x_n)]/n - E_x^{H*} [h(x_0)]/n + \frac{1}{n} E_x^{H*} \sum_{t=1}^{n} c(x_{t-1}, u_{t-1})]. \]

Here,
\[ g = \lim_{n \to \infty} \frac{1}{n} E_x^{H*} \sum_{t=1}^{n} c(x_{t-1}, u_{t-1})]. \]

and
\[ \left| \frac{1}{n} E_x^{H*} \sum_{t=1}^{n} c(x_{t-1}, u_{t-1}) - g \right| \leq \frac{1}{n} \left( |E_x^H |h(x_n)| + |h(x)| \right) \to 0 \]

\[ \diamond \]

Remark 7.1. It would be interesting to see if given some conditions which would resemble those in Lemma 5.5.3(iii), the condition (7.1) could be relaxed so that it only is to hold for the policy \( \Pi^* \).
A value iteration method

Define $T(h) = \inf_{u \in U} \left( c(x, u) + \int h(x') P(dx'|x, u) \right)$. Fix $z \in \mathbb{X}$ and consider the space of measurable and bounded functions $h$ with $h(z) = 0$. Let $(g, h, f)$ be a canonical triplet with $g \equiv \rho$ so that

$$\rho + h(x) = \inf_{u \in U} \left( c(x, u) + \int h(x') P(dx'|x, u) \right)$$

Note that if $h^*(z) = 0$, then $\rho = T(h)$ and $h(x) = (T(h))(x) - (T(h))(z) =: (T_z(h))(x)$. Consider the following assumption.

**Assumption 7.1.1** For some $\alpha \in (0, 1)$, and $x, x' \in \mathbb{X}$ and $u, u' \in U$

$$\|P(\cdot|x, u) - P(\cdot|x', u')\|_{TV} \leq 2\alpha$$

Consider the following span norm:

$$\|u\|_{sp} = \sup_x u(x) - \inf_x u(x)$$

The space of measurable bounded functions (that satisfy $h(z) = 0$) with $\|u\|_{sp}$ is a Banach space. Under Assumption 7.1.1 and the measurable selection conditions reviewed in Chapter 5, $T_z$ is a contraction on the space of bounded functions $h$ with $h(z) = 0$, with modulus $\alpha$ [57]. As a result, the iterates given by

$$h_{n+1} = T_z(h_n),$$

with $h_0 \equiv 0$ will converge to a fixed point.

7.1 The Vanishing Discounted Cost Approach to the Average Cost Problem

Average cost emphasizes the asymptotic values of the cost function whereas the discounted cost emphasizes the short-term cost functions. However, under technical restrictions, one can show that the limit as the discounted factor converges to 1, one can obtain a solution for the average cost optimization. We now state one such condition below.

**Theorem 7.1.2** Consider a finite state controlled Markov chain where the state and action spaces are finite, and for every deterministic policy the entire state space is a recurrent set. Let

$$J_\beta(x) = \inf_{\Pi \in \Pi_A} J_\beta(x, \Pi) = \inf_{\Pi \in \Pi_A} E_x^\Pi \left[ \sum_{t=0}^{\infty} \beta^t c(x_t, u_t) \right]$$

and suppose that $\Pi^*_n$ is an optimal deterministic policy for $J_{\beta_n}(x)$. Then, there exists some $\Pi^* \in \Pi_{SD}$ which is optimal for every $\beta$ sufficiently close to 1, and is also optimal for the average cost

$$J(x) = \inf_{\Pi \in \Pi_A} \limsup_{T \to \infty} \frac{1}{T} E_x^\Pi \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right]$$

**Proof.** First note that for every stationary and deterministic policy $f$, $J_\beta(x, f) = (1 - \beta) E_x \left[ \sum_{k=0}^{\infty} \beta^k c(x, f(x_k)) \right]$ is a continuous function on $[0, 1]$ (in $\beta$). Let $\beta_n \uparrow 1$. For each $\beta_n$, $J_{\beta_n}$ is achieved by a stationary and deterministic policy. Since there are only finitely many such policies, there exists at least one policy $f^*$ which is optimal for infinitely many $\beta_n$; call such a sequence $\beta_{n_k}$. We will show that this policy is optimal for the average cost problem also.

It follows that $J_{\beta_{n_k}}(x, f^*) \leq J_{\beta_{n_k}}(x, \Pi)$ for all $\Pi$. Then, infinitely often for every deterministic stationary policy $f$:

$$J_{\beta_{n_k}}(x, f^*) - J_{\beta_{n_k}}(x, f) \leq 0$$
We now claim that for some $\beta^* < 1$, $J_\beta(x, f^*) \leq J_\beta(x, II)$ for all $\beta \in (\beta^*, 1)$. The function $J_{\beta_{n_k}}(x, f^*) - J_{\beta_{n_k}}(x, f)$ is continuous in $\beta$, therefore if the claim were not correct, the function must have infinitely many zeros. On the other hand, one can write the equation

$$J_\beta(x, f) = c(x, f) + \beta \sum_{x'} P(x'|x, f(x)) J_\beta(x', f)$$

in matrix form to obtain $J_\beta(x, f) = (I - \beta P(\cdots|x, f(x)))^{-1} c(x, f(x))$. It follows that, $J_\beta(x, f^*) - J_\beta(x, f)$ is a rational function on the complex region $|z| < 1$; such a function can only have finitely many zeros (unless it is identically zero).

Therefore, it must be that for some $\beta^* < 1$, $J_\beta(x, f^*) \leq J_\beta(x, II)$ for all $\beta \in (\beta^*, 1)$. We note here that such a policy is called a Blackwell-Optimal Policy. Now,

$$(1 - \beta_{n_k})J_{\beta_{n_k}}(x, f^*) \leq (1 - \beta_{n_k})J_{\beta_{n_k}}(x, II) \quad (7.7)$$

for any $II$ and thus,

$$J(x, f^*) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_x^f \left( \sum_{k=0}^{T-1} c(x_k, u_k) \right) \leq \lim_{n_k \to \infty} (1 - \beta_{n_k})J_{\beta_{n_k}}(x, f^*) = \lim_{n_k \to \infty} (1 - \beta_{n_k})J_{\beta_{n_k}}(x, f^*)$$

$$\leq \limsup_{n_k \to \infty} (1 - \beta_{n_k})J_{\beta_{n_k}}(x, II) \leq \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_x^f \left( \sum_{k=0}^{T-1} c(x_k, u_k) \right) \quad (7.8)$$

Here, the sequence of inequalities follows from the following Abelian inequalities (see Lemma 5.3.1 in [59]): Let $a_n$ be a sequence of non-negative numbers and $\beta \in (0, 1)$. Then,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{m=0}^{N-1} a_m \leq \liminf_{\beta \uparrow 1} \sum_{m=0}^{\infty} \beta^m a_m \leq \limsup_{\beta \downarrow 1} \sum_{m=0}^{\infty} \beta^m a_m \leq \limsup_{N \to \infty} \frac{1}{N} \sum_{m=0}^{N-1} a_m \quad (7.9)$$

As a result, $f^*$ is optimal. The optimal cost does not depend on the initial state by the recurrence condition and irreducibility of the chain under the optimal policy.

In the following, we consider more general state spaces.

### 7.1.2 Polish State and Action Spaces, ACOE and ACOI

In the following, we assume that $X$ is a Polish space and $U$ is a compact subset of a Polish space.

Consider the value function for a discounted cost problem as discussed in Section 5.5

$$J_\beta(x) = \min_{u \in U} \left\{ c(x, u) + \beta \int_X J_\beta(y) Q(dy|x, u) \right\}, \quad x \in X.$$

Let $x_0$ be an arbitrary state and for all $x \in X$ consider

$$J_\beta(x) - J_\beta(x_0) = \min_{u \in U} \left( c(x, u) + \beta \int P(dx'|x, u) (J_\beta(x') - J_\beta(x_0)) - (1 - \beta)J_\beta(x_0) \right)$$

As discussed in Section 5.5, this has a solution for every $\beta \in (0, 1)$ under measurable selection conditions.

A family of functions $F$ mapping a metric separable space $S$ to $\mathbb{R}$ is said to be equicontinuous at a point $x_0 \in S$ if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d(x, x_0) \leq \delta \implies |f(x) - f(x_0)| \leq \epsilon$ for all $f \in F$. The family $F$ is said to be equicontinuous if it is equicontinuous at each $x \in S$. 

Now suppose that \( h_{\beta}(x) := J_{\beta}(x) - J_{\beta}(x_0) \) is (over \( \beta \)) equi-continuous and \( X \) is compact. By the Arzela-Ascoli Theorem (Theorem 7.1.3), taking \( \beta \to 1 \) along some sequence, for some subsequence, \( J_{\beta_{n_k}}(x) - J_{\beta_{n_k}}(x_0) \to \eta(x) \). If the optimal average cost is finite, then by the Abelian inequality (7.9)

\[
(1 - \beta_{n_k})J_{\beta_{n_k}}(x_0) \to \zeta^*
\]

(7.10)

will converge along a further subsequence (since the range of these functions would be compact). If we could also exchange the order of the minimum and the limit, one obtains the Average Cost Optimality Equation (ACOE):

\[
\eta(x) = \min_{u \in U} \left( c(x, u) + \int P(dx'|x_t, u_t)\eta(x') - \zeta^* \right)
\]

(7.11)

We now make this observation formal (and relax the compactness assumption on the state space).

**Assumption 7.1.2**

(a) The one stage cost function \( c \) is \( \mathbb{R}_+ \)-valued and is in \( C_b(X \times U) \), that is \( c \) is continuous and bounded.

(b) The stochastic kernel \( \eta(\cdot | x, u) \) is weakly continuous in \( (x, u) \in X \times U \), i.e., if \( (x_k, u_k) \to (x, u) \), then \( \eta(\cdot | x_k, u_k) \to \eta(\cdot | x, u) \) weakly.

(c) \( U \) is compact.

(d) \( X \) is Polish and locally compact so that it is \( \sigma \)-compact, that is, \( X = \cup_n S_n \) where \( S_n \subset S_{n+1} \) and each \( S_n \) is compact.

In addition to Assumption 7.1.2, we impose the following assumption in this section.

**Assumption 7.1.3**

There exists \( \alpha \in (0, 1) \) and \( N \geq 0 \), a nonnegative function \( b \) on \( X \) and a state \( z_0 \in X \) such that,

(e) \( -N \leq h_\beta(z) \leq b(z) \) for all \( z \in X \) and \( \beta \in [\alpha, 1] \), where

\[
h_\beta(z) = J_\beta(z) - J_\beta(z_0),
\]

for some fixed \( z_0 \in X \).

(f) The sequence \( \{h_{\beta(k)}\} \) is equi-continuous, where \( \{\beta(k)\} \) is a sequence of discount factors converging to 1 which satisfies

\[
\lim_{k \to \infty} (1 - \beta(k))J_{\beta(k)}(z) = \rho^* \text{ for all } z \in X \text{ for some } \rho^* \in [0, L].
\]

(g) If \( X \) is not compact, then \( b(z) \) is bounded.

Note that since the one stage cost function \( c \) is bounded by some \( L \geq 0 \), we must have

- \( (1 - \beta)J_\beta(z) \leq L \) for all \( \beta \in (0, 1) \) and \( z \in X \).

Let us recall the Arzela-Ascoli theorem.

**Theorem 7.1.3** Let \( F \) be an equi-continuous family of functions on a compact space \( X \) and let \( h_n \) be a sequence in \( F \) such that the range of \( f_n \) is compact. Then, there exists a subsequence \( h_{n_k} \) which converges uniformly to a continuous function. If \( X \) is \( \sigma \)-compact, that is \( X = \cup_n K_n \) with \( K_n \subset K_{n+1} \) with \( K_n \) compact, the same result holds where \( h_{n_k} \) converges pointwise to a continuous function, and the convergence is uniform on compact subsets of \( X \).

**Theorem 7.1.4** Under Assumptions 7.1.2 and 7.1.3 there exist a constant \( \rho^* \geq 0 \), a continuous and bounded \( h \) from \( X \) to \( \mathbb{R} \) with \( -N \leq h(\cdot) \leq b(\cdot) \), and \( \{f^*\} \in S \) such that \( (\rho^*, h, f^*) \) satisfies the ACOE; that is,

\[
\rho^* + h(z) = \min_{u \in U} \left[ c(z, u) + \int_X h(y)\eta(dy|z, u) \right]
\]
Here, somewhat similar to Lemma 5.2.2, the exchange of limit and minimum follows from writing (using the compactness of $u$)

$$\inf \frac{\partial J(\varphi, z)}{\partial \varphi} = J^*(z) = J\{f^*(z), z\} = \rho^*,$$

for all $z \in \mathcal{X}$.

**Proof.** By (7.10), we have that $(1 - \beta_{n_k})J_\beta(x_0) \to \rho^*$ for some subsequence $n_k$ as $\beta_{n_k} \uparrow 1$. By Assumption 7.1.3 (f) and Theorem 7.1.3 there exists a further subsequence of $h_{n_k}, \{h_{\beta(k)}\}$, which converges (uniformly on compact sets) to a continuous and bounded function $h$. Take the limit in (7.16) along this subsequence, i.e., consider

$$\rho^* + h(z) = \lim \min_U [c(z, u) + \beta(k) \int_{\mathcal{X}} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u)] = \lim \min_U [c(z, u) + \beta(k) \int_{\mathcal{X}} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u)] = \min_U \lim [c(z, u) + \int_{\mathcal{X}} h(y) \eta(\mathrm{d}y | z, u)].$$

Here, somewhat similar to Lemma 5.2.2 the exchange of limit and minimum follows from writing (using the compactness of $U$, the continuity of $[c(z, u) + \beta(k) \int_{\mathcal{X}} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u)]$ on $U$, and the equicontinuity of $\{h_{\beta(k)}\}$):

$$\min_U [c(z, u) + \beta(k) \int_{\mathcal{X}} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u)] = c(z, u_1) + \beta(k) \int_{\mathcal{X}} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u_1)$$

and showing that

$$\max \left| \beta(k) \int_{\mathcal{X}} (h_{\beta(k)}(y) - h(y)) \eta(\mathrm{d}y | z, u_1) \right| \to 0. \quad (7.12)$$

The last item follows from a contra-positive argument to show the convergence of some subsequence, as in Lemma 5.2.2 of $u_n \to u$ for some $u$ and the fact that since for every $\{u_n \to u\}$, the set of probability measures $\eta(\mathrm{d}y | z, u_n)$ is tight, for every $\epsilon > 0$ (here, we use weak continuity condition in Assumption 7.1.2, one can find a compact set $K_n \subset \mathcal{X}$ so that $\int_{\mathcal{X} \setminus K_n} h_{\beta(k)}(y) \eta(\mathrm{d}y | z, u) \leq \epsilon$ (Assumption 7.1.3 (g))). Since on $K_n$, $h_{\beta(k)} \to h$ uniformly and $h$ is bounded, the result follows. If $\mathcal{X}$ is already compact (Assumption 7.1.3 (g)) is not needed.

**Remark 7.2.** One can relax the boundedness conditions on the cost function and use instead Assumptions 4.2.1 and 5.5.1 of [59]; see e.g. Theorem 5.5.4 in in [59]). Further conditions also appear in the literature; see Hernandez-Lerma and Lasserre [60] for an exhaustive list and an analysis for the unbounded cost setup, and [38] for such results and a detailed literature review. Further conditions, which only involve weak continuity, are available in [55] [121], among other references.

In addition, if one cannot verify the equi-continuity assumption, the following holds; note that the condition of strong continuity is required here.

**Theorem 7.1.5 (Theorem 5.4.3 in [59])** Let for every measurable and bounded $g$, the integral $\int g(x_{t+1}) \mathbb{P}(dx_{t+1} | x_t = x, u_t = u)$ be continuous in $u$ and $x$, and there exist $N < \infty$ and a function $b(x)$ with

$$-N \leq h_\beta(x) \leq b(x), \quad \beta \in (0, 1), x \in \mathcal{X} \quad (7.13)$$
and for all $\beta \in [\alpha, 1)$ for some $\alpha < 1$ and $M \in \mathbb{R}_+$:

$$(1 - \beta)J^\beta_\eta(x) \leq M.$$  \hfill (7.14)

**Proof.** By (7.14) and (7.10), we have that $(1 - \beta_{n_k})J_{\beta_{n_k}}(x_0) \rightarrow \zeta^*$ for some subsequence $n_k$ and $\beta_{n_k} \uparrow 1$. On the other hand, once again by (7.10), under any policy $\Pi$, and any sequence $\beta \uparrow 1$

$$\limsup_{\beta \rightarrow 1}(1 - \beta)J^\beta_\eta(x) \leq \limsup_{T \rightarrow \infty} \frac{1}{T}E^H_x \left[ \sum_{k=0}^{T-1} c(x_k, u_k) \right]$$

thus $\zeta^*$ is a lower bound under any admissible policy.

Now consider some arbitrary state $x$,

$$(1 - \beta_{n_k})J_{\beta_{n_k}}(x) = (1 - \beta_{n_k})(J_{\beta_{n_k}}(x) - J_{\beta_{n_k}}(x_0)) + (1 - \beta_{n_k})J_{\beta_{n_k}}(x_0)$$

By (7.13), $(1 - \beta_{n_k})(J_{\beta_{n_k}}(x) - J_{\beta_{n_k}}(x_0)) \rightarrow 0$. Thus, for any $x$, $(1 - \beta_{n_k})J_{\beta_{n_k}}(x) \rightarrow \zeta^*$.

We now show that (7.15) holds. For this, consider again (7.16).

$$(1 - \beta)J^\beta_\eta(x) = J^\beta_\eta(x) - J^\beta_\eta(x_0)$$

$$\geq \min_{u \in U} \left( c(x, u) + \beta \int P(dx'|x, u)[J^\beta_\eta(x') - J^\beta_\eta(x_0)] - (1 - \beta)J^\beta_\eta(x_0) \right)$$

Observe the following along the subsequence $n_k$, with $h_\beta(x) = J^\beta_\eta(x) - J^\beta_\eta(x_0)$, $(1 - \beta_{n_k})J_{\beta_{n_k}}(x_0) \rightarrow \zeta^*$ and

$$= \liminf_{n_k \rightarrow \infty} \min_{u \in U} \left( c(x, u) + \beta_{n_k} \int P(dx'|x, u)h_{\beta_{n_k}}(x') \right) - (1 - \beta_{n_k})J_{\beta_{n_k}}(x_0)$$

$$= \liminf_{n_k \rightarrow \infty} \min_{m > n_k \leq m_{n_k}} \left( c(x, u) + \beta_{n_k} \int P(dx'|x, u)h_{\beta_m}(x') \right) - (1 - \beta_{n_k})J_{\beta_{n_k}}(x_0)$$

$$\geq \liminf_{n_k \rightarrow \infty} \min_{u \in U} \left( c(x, u) + \beta_{n_k} \int P(dx'|x, u)H_k(x') \right) - (1 - \beta_{n_k})J_{\beta_{n_k}}(x_0)$$

$$= \min_{u \in U} \left( c(x, u) + \int P(dx'|x, u)\eta(x') - \zeta^* \right)$$

where $H_k(x) = \inf_{m > n_k} h_{\beta_m}(x)$ and $\eta(x) = \lim_{m \rightarrow \infty} H_k(x)$. The last equality holds if $H_k \uparrow \eta$ as shown in Lemma 5.5.2 and bounded from below and that $\int P(dx'|x, u)H_k(x')$ and $\int P(dx'|x, u)\eta(x')$ are continuous on $U$ (this is where we use the strong continuity property). The conditions in the theorem statement imply these.

We now show that the stationary and deterministic policy $\Pi = \{ \eta, \eta, \xi, \cdots \}$ is optimal. Using (7.15) repeatedly, we have that

$$E^\infty_x \left[ \sum_{k=0}^{T-1} c(x_k, u_k) \right] \leq T\zeta^* + \eta(x) - E^\infty_x \left[ \eta(x_T) \right] \leq T\zeta^* + \eta(x) + N.$$
Dividing by $T$ and taking the $\lim\sup$, leads to the result that

$$
\lim\sup_{T \to \infty} \frac{1}{T} E^{f^*}_{x} \left[ \sum_{k=0}^{T-1} c(x_k, u_k) \right] \leq \zeta^*.
$$

This completes the proof.

We state the following further relaxation.

**Assumption 7.1.4**

(a) The one stage cost function $c$ is nonnegative and continuous.

(b) The stochastic kernel $P(\cdot | x, u)$ is weakly continuous in $(x, u) \in \mathbb{X} \times \mathbb{U}$.

(c) $\mathbb{U}$ is compact.

(d) There exist nonnegative real numbers $M$ and $\alpha \in [1, \frac{1}{\beta})$, and a continuous weight function $w : \mathbb{X} \to [1, \infty)$ such that for each $x \in \mathbb{X}$, we have

\[ \sup_{a \in \mathbb{U}} c(x, a) \leq Mw(x), \]  

(7.16)

\[ \text{and } \int_{\mathbb{X}} w(y)p(dy|x, a) \text{ is continuous in } (x, a). \]  

(7.17)

For any real-valued measurable function $u$ on $\mathbb{X}$, let $Tu : \mathbb{X} \to \mathbb{R}$ is given by

$$
Tu(x) := \min_{a \in \mathbb{U}} \left[ c(x, a) + \beta \int_{\mathbb{X}} u(y)p(dy|x, a) \right].
$$

(7.18)

**Assumption 7.1.5** Suppose Assumption 7.1.4 holds. Moreover, suppose there exist a probability measure $\lambda$ on $\mathbb{X}$ and a continuous function $\phi : \mathbb{X} \times \mathbb{U} \to [0, \infty)$ such that

(e) $\int_{\mathbb{X}} w(y)p(dy|x, a) \leq \alpha w(x) + \lambda(w)\phi(x, a)$ for all $(x, a) \in \mathbb{X} \times \mathbb{U}$, where $\alpha \in (0, 1)$.

(f) $p(D|x, a) \geq \lambda(D)\phi(x, a)$ for all $(x, a) \in \mathbb{X} \times \mathbb{U}$ and $D \in \mathcal{B}(\mathbb{X})$.

(g) The weight function $w$ is $\mu$-integrable.

(h) $\int_{\mathbb{X}} \phi(x, f(x))\lambda(dx) > 0$ for all $f \in \mathcal{F}$.

By [121, Theorem 3.5], there exists a unique fixed point of the following contraction operator with modulus $\alpha$ mapping $B_w(\mathbb{X}) \cap C(\mathbb{X})$ into itself

$$
Fu(x) := \min_{a \in \mathbb{U}} \left[ c(x, a) + \int_{\mathbb{X}} u(y)p(dy|x, a) - \lambda(u)\phi(x, a) \right].
$$

The following theorem is a consequence of [121, Theorems 3.3 and 3.6].

**Theorem 7.1.6** Under Assumption 7.1.5 the following holds. For each $f \in \mathcal{F}$, the stochastic kernel $Q_f(\cdot | x)$ has an unique invariant probability measure $\nu_f$. Furthermore, $w$ is $\nu_f$-integrable, and therefore, $\rho_f := \int_{\mathbb{X}} c(x, f(x))\nu_f(dx) < \infty$. There exist $f^* \in \mathcal{F}$ and $h^* \in B_w(\mathbb{X}) \cap C(\mathbb{X})$ such that the triplet $(h^*, f^*, \rho_{f^*})$ satisfies the average cost optimality equality (ACOE) and therefore, for all $x \in \mathbb{X}$

$$
\inf_{\pi \in \Pi} V(\pi, x) := V^*(x) = \rho_{f^*}.
$$
Remark 7.3. A number of sufficient conditions exist in the literature for the ACOE or the ACOI to hold (see [60], [122]). These conditions typically have the form of Assumption 5.5.1 or 5.5.2 together with geometric ergodicity conditions with condition (5.22) replaced with conditions of the form:

$$\sup_{a \in U} \int_{X} w(y) \eta(dy|x,u) \leq \alpha w(x) + K \phi(x,u),$$

where $$\alpha \in (0,1)$$, $$K < \infty$$ and $$\phi$$ a positive function. In some approaches, $$\phi$$ and $$w$$ needs to be continuous, in others it does not. For example if $$\phi(x,u) = 1_{\{x \in C\}}$$ for some small set $$A$$, then we recover a condition similar to (4.21) leading to geometric ergodicity.

We also note that for the above arguments to hold, there does not need to be a single invariant distribution. Here in (7.16), the pair $$x$$ and $$x_0$$ should be picked as a function of the reachable set under a given sequence of policies. The analysis for such a condition is tedious in general since for every $$\beta$$ a different optimal policy will typically be adopted; however, for certain applications the reachable set from a given point may be independent of the control policy applied.

7.2 The Convex Analytic Approach to Average Cost Markov Decision Problems

The convex analytic approach (typically attributed to Mannne [79] and Borkar [27] (see also [59])) is a powerful approach to the optimization of infinite-horizon problems. It is particularly effective in proving results on the optimality of stationary policies, which can lead to an infinite-dimensional linear program. This approach is particularly effective for constrained optimization problems and infinite horizon average cost optimization problems. It avoids the use of dynamic programming.

We are interested in the minimization

$$\inf_{\Pi \in \Pi_A} \limsup_{T \to \infty} \frac{1}{T} E_{x_0}^{\Pi} \left[ \sum_{t=1}^{T} c(x_t,u_t) \right],$$

where $$E_{x_0}$$ denotes the expectation over all sample paths with initial state given by $$x_0$$ under the admissible policy $$\Pi$$.

7.2.1 Finite State/Action Setup

We first consider the finite space setting where both $$X$$ and $$U$$ are finite sets. We study the limit distribution of the following occupation measures, under any policy $$\Pi$$ in $$\Pi_A$$. Let for $$T \geq 1$$

$$\nu_T(D) = \frac{1}{T} \sum_{t=1}^{T} 1_{\{(x_t,u_t) \in D\}}, \quad D \in \mathcal{B}(X \times U).$$

Average Cost Optimality

Consider any policy $$\Pi$$ in $$\Pi_A$$ and let for $$T \geq 1$$,

$$\mu_T(D) = E[\nu_T(D)] = E_{v_0} \frac{1}{T} \left[ \sum_{t=1}^{T} 1_{\{(x_t,u_t) \in D\}} \right], \quad D \in \mathcal{B}(X \times U),$$

with $$\mu_0(D) = E[\nu_0(D)]$$ for Borel $$D$$.

Then through what is often referred to as a Krylov-Bogoliubov-type argument, for every $$A \subset X$$

$$|\mu_N(A \times U) - \mu_N P(A \times U)|$$
The solution to this problem then gives us the optimal cost (under any policy). Thus, a candidate for an optimal policy can be obtained through the following linear program:

\[
\begin{align*}
\sum_{X \times U} \nu(x, u) c(x, u) = 1/N \left( \mu_0(A \times U) + \cdots + \mu P^{(N-1)}(A \times U) \right) - (v_0 P(A \times U) + \cdots + v P^N(A \times U)) \\
\leq 1/N [\mu_0(A \times U) - \mu_0 P^N(A \times U)] \rightarrow 0.
\end{align*}
\]

Now, if we can ensure that for some subsequence, \(\mu_{t_k} \rightarrow \mu\) for some probability measure \(\mu\), it would follow that \(\mu_{t_k} P(A \times U) \rightarrow \mu P(A \times U)\).

Now define

\[
\mathcal{G}_x = \{ v \in \mathcal{P}(X \times U) : v(B \times U) = \sum_{x, u} P(x_{t+1} \in B \mid x_t = x, u_t = u) v(x, u), \quad B \in \mathcal{B}(X) \}
\]

Further define

\[
\mathcal{G} = \{ v \in \mathcal{P}(X \times U) : \exists \Pi \in \Pi_S, v(A) = \sum_{x, u} P^\Pi ((x_{t+1}, u_{t+1}) \in A \mid x_t = x, u_t = u) v(x, u), \quad A \in \mathcal{B}(X \times U) \}
\]

It is evident that \(\mathcal{G} \subset \mathcal{G}_x\) since there are (seemingly) fewer restrictions for \(\mathcal{G}_x\). We can show that these two sets are indeed equal: For \(v \in \mathcal{G}_x\), if we write: \(v(x, u) = \pi(x) P(u \mid x)\), then, we can construct a consistent \(v \in \mathcal{G}\): \(v(B \times C) = \sum_{x \in B} P(C \mid x)\).

Thus, every (weakly) converging subsequence \(\mu_{t_k}\) will satisfy the above equation (note that, we do not claim that every sequence converges). And hence, any sequence \(\{\mu_{t_k}\}\) will have a converging subsequence and the limit of such a subsequence will be in the set \(\mathcal{G}\). This is where finiteness is helpful: If the state space were countable, there would be no guarantee that every sequence of occupation measures would have a converging subsequence. The following has thus been established.

**Lemma 7.2.1** Under any admissible policy, any converging subsequence \(\{\mu_{t_k}\}\) will converge to the set \(\mathcal{G}\).

Let us define

\[
\gamma^* = \inf_{v \in \mathcal{G}} \sum_{x, u} v(x, u) c(x, u)
\]

Let \(\langle \mu, c \rangle := \sum \mu(x, u) c(x, u)\). Now, we have that

\[
\liminf_{T \to \infty} \langle \mu_T, c \rangle \geq \gamma^*
\]

since for any sequence \(\mu_{T_k}\) which converges to the liminf value, there exists a further subsequence \(\mu_{T_k'}\) (due to the (weak) compactness of the space of expected empirical occupation measures) which has a weak limit, and this weak limit is in \(\mathcal{G}\). We have then that

\[
\lim_{T_k \to \infty} \langle \mu_{T_k}, c \rangle = \langle \lim_{T_k' \to \infty} \mu_{T_k}, c \rangle \geq \gamma^*.
\]

The solution to this problem then gives us the optimal cost (under any policy). Thus, a candidate for an optimal policy can be obtained through the following linear program:

**Linear Program For Finite Models**

Given a cost function \(c\) and transition kernel \(Q\), find the minimum of the linear function

\[
\sum_{X \times U} \nu(x, u) c(x, u).
\]

over all

\[
\nu \in \mathcal{G} = \left\{ \mu : \mu(y, U) = \sum_{X \times U} Q(y \mid (x, u)) \mu(x, u) \right\}.
\]
which can also be written as
\[
\sum_j \mu(y, j) = \sum_{x \times U} P((x, u); y)\mu(x, u) \}
\]
and
\[
\mu(x, y) \geq 0, \quad \sum_{x, y} \mu(x, y) = 1
\]
All of these are linear constraints.

Let \(\mu^*\) be the optimal occupation measure (this exists since the state space is finite, and thus \(G\) is compact, and \(\sum_{x \times U} \mu(x, u)c(x, u)\) is continuous in \(\mu\)). This induces an optimal policy \(\pi(u|x)\) as:
\[
\pi(u|x) = \frac{\mu^*(x, u)}{\sum_u \mu^*(x, u)}
\]
Thus, we can find the optimal policy through a linear program, without using dynamic programming.

**Sample Path Optimality**

The above optimality argument is in the stronger *sample-path* sense, rather than only in expectation. Consider the following:
\[
\inf_{\Pi} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T}[c(x_t, u_t)], \quad (7.21)
\]
where there is no expectation. The above is known as the sample path cost. Let \(F_t\) be the \(\sigma\)-field generated by \(\{x_s, u_s, s \leq t\}\). Define a \(F_t\) measurable process with \(A \in B(\mathbb{X})\):
\[
F_t(A) = \left(\sum_{s=1}^{t} 1_{\{x_s \in A\}} - t \sum_{x \times U} P(A|x, u)v_t(x, u)\right)
\]
Note that this can also be written as
\[
F_t(A) = \left(\sum_{s=1}^{t} \left(1_{\{x_s \in A\}} - \sum_{x \times U} P(A|x, u)1_{\{x_s=x, u_s=u\}}\right)\right)
\]
Thus, for \(t \geq 1\),
\[
E[F_t(A)|F_{t-1}] = E\left[\sum_{s=1}^{t} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-1} P(x_{s+1} \in A|x_s = x, u_s = u)1_{\{(x_s, u_s)=(x, u)\}}|F_{t-1}\right]
\]
\[
= E\left[\left(1_{\{x_t \in A\}} - \sum_{x \times U} P(x_{t+1} \in A|x_t = x, u_t = u)1_{\{(x_{t}, u_{t}, x_{t-1}, u_{t-1})=(x, u)\}}\right)|F_{t-1}\right]
\]
\[
+ \left(\sum_{s=1}^{t-1} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-2} \sum_{x \times U} P(x_{s+1} \in A|x_s = x, u_s = u)1_{\{(x_s, u_s)=(x, u)\}}\right)\]
\[
= 0 \quad (7.22)
\]
\[
+ \left(\sum_{s=1}^{t-1} 1_{\{x_s \in A\}} - \sum_{s=0}^{t-2} \sum_{x \times U} P(x_{s+1} \in A|x_s = x, u_s = u)1_{\{(x_s, u_s)=(x, u)\}}\right)|F_{t-1}\right]\]
\[
= F_{t-1}(A), \quad (7.24)
\]
where \((7.22)\) follows from the fact that 
\[
E[1_{\{x_t \in A\}} | F_{t-1}] = P(x_t \in A | F_{t-1}).
\]
We have then that 
\[
E[F_t(A) | F_{t-1}] = F_{t-1}(A) \quad \forall t \geq 0,
\]
and \(\{F_t(A)\}\) is a martingale sequence.

Furthermore, \(F_t(A)\) is a bounded-increment martingale since 
\[
|F_t(A) - F_{t-1}(A)| \leq 1.
\]
Hence, for every \(T > 2\), \(\{F_1(A), \ldots, F_T(A)\}\) forms a martingale sequence with uniformly bounded increments, and we could invoke the Azuma-Hoeffding inequality \([40]\) to show that for all \(x > 0\)
\[
P(|F_t(A)| \geq x) \leq 2e^{-2x^2 t}
\]
Finally, invoking the Borel-Cantelli Lemma (see Theorem [3.2.1]) for the summability of the estimate above, that is:
\[
\sum_{n=1}^{\infty} 2e^{-2x^2 t} < \infty, \forall x > 0,
\]
we deduce that
\[
\lim_{t \to \infty} \frac{F_t(A)}{t} = 0 \quad \text{a.s.}
\]
Thus,
\[
\lim_{T \to \infty} \left( v_T(A) - \sum_{x \in X} P(A|x,u)v_T(x,u) \right) = 0, \quad A \subset X
\]
Thus, somewhat similar to the arguments in \((7.20)\), every converging subsequence would have to be in the set \(G\).

Let \(\langle v, c \rangle := \sum v(x,u)c(x,u)\). Now, we have that
\[
\lim \inf_{T \to \infty} \langle v_T, c \rangle \geq \gamma^*
\]
since for any sequence \(v_{T_k}\) which converges to the liminf value, there exists a further subsequence \(v_{T_k'}\) (due to the (weak) compactness of the space of occupation measures) which has a weak limit, and this weak limit is in \(G\). Then,
\[
\lim_{T_k \to \infty} \langle v_{T_k}, c \rangle = \left( \lim_{T_k' \to \infty} v_{T_k}, c \right) \geq \gamma^*.
\]
Likewise, for the average cost problem:
\[
\lim \inf_{T \to \infty} E[\langle v_T, c \rangle] \geq E[\lim \inf_{T \to \infty} \langle v_T, c \rangle] \geq \gamma^*.
\]

**Lemma 7.2.2** [Proposition 9.2.5 in \([81]\)] The space \(G\) is convex and closed. Let \(G_e\) denote the set of extreme points of \(G\). A measure is in \(G_e\) if and only if two conditions are satisfied: i) A control policy inducing it is deterministic and ii) under this deterministic policy, the induced Markov chain is irreducible.

**Proof:**
Consider \(\eta\), an invariant measure in \(G\) which is non-extremal. Then, this measure can be written as a convex combination of two measures:
\[
\eta(x,u) = \Theta \eta^1(x,u) + (1 - \Theta) \eta^2(x,u)
\]
\[
\pi(x) = \Theta \pi^1(x) + (1 - \Theta) \pi^2(x)
\]
Let \(\phi^1(u|x), \phi^2(u|x)\) be two policies leading to invariant measures \(\eta_1\) and \(\eta_2\) and invariant measures \(\pi_1\) and \(\pi_2\) respectively. Then, there exists \(\gamma(\cdot)\) such that:
\[
\phi(u|x) = \gamma(x) \phi^1(u|x) + (1 - \gamma(x)) \phi^2(u|x)
\]
where under $\phi$ the invariant measure $\eta$ is attained.

There are two ways this can be satisfied: i) $\phi$ is randomized and ii) If $\phi$ is not randomized with $\phi^1 = \phi^2$ for all $x$ and are deterministic, then, it must be that $\pi^1$ and $\pi^2$ are different measures and have different support sets. In this case, the measures $\eta^1$ and $\eta^2$ do not correspond to ergodic occupation measures (in the sense of positive Harris recurrence).

We can show that for the countable state space case a converse also holds, that is, if a policy is randomized it cannot lead to an extreme point unless it has the same invariant measure under some deterministic policy. That is, extreme points are achieved under deterministic policies. Let there be a policy $P$ which is randomizing between two deterministic policies $P^1$ and $P^2$ at a state $x$ such that with probability $\theta$, $P^1$ is chosen; and with probability $1-\theta$, $P^2$ is chosen. Let $v$, $v^1$, $v^2$ be corresponding invariant measures to $P, P^1, P^2$. Here, $\alpha$ is an accessible atom if $E_\alpha^P[\tau_\alpha] < \infty$ and $P_P^P(x, B) = P_P^P(y, B)$ for all $x, y \in \alpha$, where $\tau_\alpha = \min(k > 0 : \pi_k = \alpha)$ is the return time to $\alpha$. In this case, the expected occupation measure can be shown to be equal to:

$$v(x, u) = \frac{E_\alpha^P[\tau_\alpha] \theta v^1(x, u) + E_\alpha^P[\tau_\alpha](1-\theta)v^2(x, u)}{E_\alpha^P[\tau_\alpha] \theta + E_\alpha^P[\tau_\alpha](1-\theta)},$$

which is a convex combination of $v^1$ and $v^2$.

Hence, a randomized policy cannot lead to an extreme point in the space of invariant occupation measures. We note that, even if one randomizes at a single state, under irreducibility assumptions, the above argument applies.

**Remark 7.4.** The above does not trivially extend to uncountable spaces. In Section 3.2 of [27], there is a discussion for a special setting.

As a result, we can deduce that for such finite state and action spaces an optimal policy is stationary and deterministic:

**Theorem 7.2.1** Let $\mathbb{X}, \mathbb{U}$ be finite spaces. Furthermore, let, for every stationary policy, there be a unique invariant distribution on a single communication class. In this case, a sample path optimal policy and an average cost policy exists. These policies are deterministic and stationary.

**Proof:**
$\mathcal{G}$ is a compact set. There exists an optimal $v$. Furthermore, by some further study, one realizes that the set $\mathcal{G}$ is closed, convex; and its extreme points are obtained by deterministic policies.

As earlier, let $\mu^*$ be the optimal occupation measure (this exists, as discussed above, since the state space is finite, and thus $\mathcal{G}$ is compact, and $\sum_{\mathbb{X} \times \mathbb{U}} \mu(x, u)c(x, u)$ is continuous in $\mu(., .)$). This induces an optimal policy $\pi(u|x)$ as:

$$\pi(u|x) = \frac{\mu^*(x, u)}{\sum_u \mu^*(x, u)}.$$  

Thus, we can find the optimal policy through a linear program, without using dynamic programming.

**7.2.2 General State/Action Spaces**

The argument in Section 7.2.1 applies almost identically. However, for the general case, we need to ensure that the set of expected occupation measures is tight, and that the set $\mathcal{G}$ is closed.

The above holds under a set of technical conditions:

1. (A) The state process takes values in a compact set $\mathbb{X}$. The control space $\mathbb{U}$ is also compact.
2. (A') The cost function satisfies the following condition: $\lim_{K_n \uparrow \mathbb{X}} \inf_{u, x \notin K_n} c(x, u) = \infty$, (here, we assume that the space $\mathbb{X}$ is locally compact Polish space so that it is $\sigma$-compact).
3. (B) There exists a policy leading to a finite cost.
4. (C) The cost function is continuous in \( x \) and \( u \).
5. (D) The transition kernel is weakly continuous in the sense that \( \int Q(dz|x,u)v(z) \) is continuous in both \( x \) and \( u \), for every continuous and bounded function \( v \).

**Theorem 7.2.2** Under the above Assumptions A, B, C, and D (or A', B, C, and D) there exists a solution to the optimal control problem given in (7.21) for almost every initial condition under the optimal invariant probability measure.

The key step is the observation that every occupation measure sequence has a weakly converging subsequence. If this exists, then the limit of such a converging sequence will be in \( \mathcal{G} \) and the analysis presented for the finite state space case will be applicable.

The issue here is that a sequence of measures on the space of invariant measures \( \nu_n \) may have a limiting probability measure, but this limit may not correspond to an invariant measure under a stationary policy. The weak continuity of the kernel, and the separability of the space of continuous functions on a compact set, allow for this generalization.

This ensures that every sequence has a converging subsequence weakly. In particular, there exists an optimal occupation measure.

**Lemma 7.2.3** \( \text{There exists an optimal occupation measure in } \mathcal{G} \text{ under the Assumptions above.} \)

**Proof:** The problem has now reduced to

\[
\inf_{\mu} \int \mu(dx, du)c(x, u),
\]

s.t.

\[\mu \in \mathcal{G}.\]

The set \( \mathcal{G} \) is closed, since if \( \nu_n \to \nu \) and \( \nu_n \in \mathcal{G} \), then for continuous and bounded \( f \in C_b(\mathbb{X}) \), \( \langle \nu_n, f(\cdot) \rangle \to \langle \nu, f \rangle \). By weak-continuity of the kernel \( \int f(x')P(dx'|x,u) \) is also continuous and thus, \( \langle \nu_n, Pf \rangle \to \langle \nu, Pf \rangle = \langle \nu P, f \rangle \). Thus, \( \nu(f) = \nu P(f) \) and \( \nu \in \mathcal{G} \). Therefore, \( \mathcal{G} \) is weakly sequentially compact. Since the integral \( \int \mu(dx, du)c(x, u) \) is lower semi-continuous on the set of measures under weak convergence, and the existence result follows from Weierstrass’ Theorem. \( \diamond \)

Following the above lemma, there exists an optimal expected empirical occupation measure, say \( v \). This defines the optimal stationary control policy by the decomposition:

\[\mu(df|u) = \frac{dv(dx, du)}{d \int u v(dx, du)},\]

\( v \) almost surely, where \( \frac{dv}{du} \) denotes the Radon-Nikodym derivative.

There is a final consideration of reachability; that is, whether from any initial state, or an initial occupation set, the region where the optimal policy is defined is attracted (see [6]).

If the Markov chain is stable and irreducible, then the optimal cost is independent of where the chain starts from, since in finite time, the states on which the optimal occupation measure has support, can be reached.

If this Markov Chain is not irreducible, then, the stationary policy is only optimal when the state process starts at the locations where in finite time the optimal stationary policy is applicable. This is particularly useful for problems where one has control over from which state to start the system.

**Remark 7.5.** Finally, similar results can be obtained when the weak continuity assumption on the transition kernel and continuity assumption on \( c(x, u) \) are eliminated, provided that the set \( \mathcal{G} \) is setwise sequentially compact and \( c \) is measurable.
and bounded. For conditions on the setwise sequential compactness, see e.g. [102] and [65]. A useful result along this
direction is the following.

**Theorem 7.2.3** [111] Let $\mu, \mu_n$ ($n \geq 1$) be probability measures. Suppose $\mu_n \rightarrow \mu$ setwise, $\lim_{n \to \infty} h_n(x) = h(x)$ for all $x \in \mathbb{E}$, and $h, h_n$ ($n \geq 1$) are uniformly bounded. Then, $\lim_{n \to \infty} \int h_n \, d\mu_n = \int h \, d\mu$.

### 7.2.3 Extreme Points and the Optimality of Deterministic Policies

Let $\mu$ be an invariant measure in $G$ which is not extreme. This means that there exists $\kappa \in (0, 1)$ and invariant occupation measures $\mu^1, \mu^2 \in G$ such that:

$$
\mu(dx, du) = \kappa \mu^1(dx, du) + (1 - \kappa) \mu^2(dx, du)
$$

Let $\mu(dx, du) = P(du|x)\mu(dx)$ and for $i = 1, 2$, $\mu^i(dx, du) = P^i(du|x)\mu^i(dx)$. Then,

$$
\mu(dx) = \kappa \mu^1(dx) + (1 - \kappa) \mu^2(dx)
$$

Note that $\mu(dx) = 0 \implies \mu^i(dx) = 0$ for $i = 1, 2$. As a consequence, the Radon-Nikodym derivative of $\mu^i$ with respect to $\mu$ exists. Let $\frac{d\mu^i}{d\mu}(x) = f^i(x)$. Then

$$
P(du|x) = P^1(du|x)\frac{d\mu^1}{d\mu}(x)\kappa + P^2(du|x)\frac{d\mu^2}{d\mu}(x)(1 - \kappa)
$$

is well-defined. Then,

$$
P(du|x) = P^1(du|x)f^1(x)\kappa + P^2(du|x)(1 - \kappa)f^2(x),
$$

such that $f^1(x)\kappa + (1 - \kappa)f^2(x) = 1$. As a result, we have that

$$
P(du|x) = P^1(du|x)\eta(x) + P^2(du|x)(1 - \eta(x)),
$$

with the randomization kernel $\eta(x) = f^1(x)\kappa$. As in Meyn [81], there are two ways where such a representation is possible: Either $P(du|x)$ is randomized, or $P^1(du|x) = P^2(du|x)$ and deterministic but there are multiple invariant probability measures under $P^1$.

The converse direction can also be obtained for the countable state space case: If a measure in $G$ is extreme, than it corresponds to a deterministic policy. The result also holds for the uncountable state space case under further technical conditions. See Borkar [27, Section 3.2] for further discussions.

**Remark 7.6.** If there is an atom (or a pseudo-atom constructed through Nummelin’s splitting technique) which is visited under every stationary policy, then as in the arguments leading to (7.25), one can deduce that on this atom there cannot be a randomized policy which leads to an extreme point on the set of invariant occupation measures.

### 7.2.4 Sample-Path Optimality

As we observed, the discussion in Section 7.2.1 applies to the sample path optimality also. We now discuss a more general setting where the state and action spaces are Polish. Let $\phi : \mathbb{X} \to \mathbb{R}$ be a continuous and bounded function. Define:

$$
v_T(\phi) = \frac{1}{T} \sum_{t=1}^{T} \phi(x, u).
$$

Define a $\mathcal{F}_t$ measurable process, with $\pi$ an admissible control policy (not necessarily stationary or Markov):
The average cost problem

\[ F_t(\phi) = \left( \sum_{s=1}^{t} \phi(x_t) \right) - t \left( \int_{\mathcal{P} \times \mathcal{U}} \int \phi(x'_t) P(dx'_t, du'_t) | x_t(v_t(dx)) \right) \]

(7.26)

We define \( \mathcal{G}_X \) to be the following set in this case.

\[ \mathcal{G}_X = \{ \eta \in \mathcal{P}(X \times U) : \eta(D) = \int_{X \times U} P(D, x) \eta(dx), \ \forall D \in \mathcal{B}(X) \} \]

The optimality argument is in the stronger sample-path sense, rather than only in expectation. Consider the following:

\[ \inf \lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} [c(x_t, u_t)] \]

(7.27)

where there is no expectation. The above is known as the sample path cost. Let \( \langle v, c \rangle := \sum v(x, u) c(x, u) \). If one can guarantee that every sequence of empirical measures \( \{v_t\} \) would have a converging subsequence to some measure \( v \), we would have that

\[ \lim_{T_k \to \infty} \langle v_{T_k}, c \rangle \geq \langle \lim_{T_k \to \infty} v_{T_k}, c \rangle \]

by the fact that for \( c \) continuous, non-negative if \( v_k \to v \),

\[ \liminf_{k \to \infty} (v_k, c) \geq (v, c) \]

Since for any sequence \( v_{T_k} \) which converges to the liminf value, there exists a further subsequence \( v_{T'_k} \) (due to the (weak) compactness of the space of occupation measures) which has a weak limit, and this weak limit is in \( \mathcal{G} \). By Fatou’s Lemma:

\[ \lim_{T_k \to \infty} \langle v_{T_k}, c \rangle = \langle \lim_{T_k \to \infty} v_{T_k}, c \rangle \geq \gamma^* \]

To apply the convex analytic approach, we require that under any admissible policy, the set of sample path occupation measures would be tight, for almost every sample path realization. If this can be established, then the result goes through not only for the expected cost, but also the sample-path average cost, as discussed for the finite state-action setup.

Researchers in the literature have tried to establish conditions which would ensure that the set of empirical occupational measures are tight. These typically follow one of two conditions: Either cost functions are near-monotone type conditions \( \lim_{|x| \to \infty} \inf_{u \in U} c(x, u) = \infty \) or behave like moments \( \lim_{|x| \to \infty} \inf_{u \in U} c(x, u) = \infty \) (when \( X \times U \) is locally compact, there exists a sequence of compact sets \( K_n \) so that \( X \times U = \cup_n K_n \) with \( \lim_{n \to \infty} \inf_{(x, u) \in K_n} c(x, u) = \infty \)), or the Markov chain satisfies strong recurrence properties \[ Chapter 3 \]. Under such conditions, the sequence of empirical occupation measures \( \{v_n\} \) which give rise to a finite cost are almost surely tight, every such sequence has a convergent subsequence and thus the arguments above apply: Every expected average-cost optimal policy is also sample-path optimal provided that the initial condition belongs to the support of the invariant probability measure under an optimal policy.

### 7.3 Constrained Markov Decision Processes

Consider the following average cost problem:

\[ \inf_{\Pi} J(x, \Pi) = \inf_{\Pi} \lim_{T \to \infty} \frac{1}{T} E_x^{\Pi} \sum_{t=0}^{T-1} c(x_t, u_t) \]

(7.28)

subject to the constraints:
7.6 Exercises

\[ \limsup_{T \to \infty} \frac{1}{T} \mathbb{E}_x T^{T-1} \sum_{t=0}^{T-1} d_i(x_t, u_t) \leq D_i \]  
(7.29)

for \( i = 1, 2, \ldots, m \) where \( m \in \mathbb{N} \).

A linear programming formulation leads to the following result.

**Theorem 7.3.1** Let \( X, U \) be countable. Consider \( 7.28 \), \( 7.29 \). An optimal policy will randomize between at most \( m + 1 \) deterministic policies.

Ross also discusses a setup with one constraint where a non-stationary history-dependent policy may be used instead of randomized stationary policies.

Finally, the theory of constrained Markov Decision Processes is also applicable to Polish state and action spaces, but this requires further technicalities. If there is an accessible atom (or an artificial atom as considered earlier in Chapter 3) under any of the policies considered, then the randomizations can be made at the atom.

### 7.4 The Linear Programming Approach to Discounted Cost Problems and Control until a Stopping Time

### 7.5 Bibliographic Notes

We note that the value iteration introduced in Chapter 5 also applies for average cost problems, where instead of the supremum norm in the contraction analysis, one may consider the span of a function; for an extensive review see [39]; see also [58].

### 7.6 Exercises

**Exercise 7.6.1** Consider the convex analytic method we discussed in class. Let \( X, U \) be countable sets and consider the occupation measure:

\[ v_T(A \times B) = \frac{1}{T} \sum_{t=0}^{T-1} 1 \{ (x_t, u_t) \in A \times B \}, \quad A \subset X, B \subset U. \]

While proving that the limit of such a measure process lives in a specific set, the following is used, which you are asked to prove: Let \( F_t \) be the \( \sigma \)-field generated by \( \{ x_s, u_s, s \leq t \} \). Define a \( F_t \) measurable process

\[ F_t(A) = \left( \sum_{s=1}^{t} 1 \{ x_s \in A \} - t \sum_{x \times U} P(X|x) v_t(x, u) \right), \]

where \( \Pi \) is some arbitrary but admissible control policy. Show that, \( \{ F_t(A), \quad t \in \mathbb{Z}_+ \} \) is a martingale sequence.

**Exercise 7.6.2** Let, for a Markov control problem, \( x_t \in X, u_t \in U \), where \( X \) and \( U \) are finite sets denoting the state space and the action space, respectively. Consider the optimal control problem of the minimization of

\[ \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_x \sum_{t=0}^{T-1} c(x_t, u_t) \]

where \( c \) is a bounded function. Further assume that under any stationary control policy, the state transition kernel \( P(x_{t+1} | x_t, u_t) \) leads to an irreducible Markov Chain.
Does there exist an optimal control policy? Can you propose a way to find the optimal policy?

Is the optimal policy also sample-path optimal?

**Exercise 7.6.3** Consider a controlled Markov Chain with state space \( X = \{0, 1\} \), action space \( U = \{0, 1\} \), and transition kernel for \( t \in \mathbb{Z}_+ \):

\[
P(x_{t+1} = 1|x_t = 0, u_t = 1) = \alpha \in (0, 1)
\]
\[
P(x_{t+1} = 1|x_t = 0, u_t = 0) = \beta \in (0, 1)
\]
\[
P(x_{t+1} = 1|x_t = 1, u_t = 0) = P(x_{t+1} = 1|x_t = 1, u_t = 1) = \frac{1}{2}
\]

Let

\[
c(0, 1) = \kappa \in \mathbb{R}_+, \quad c(0, 0) = 1
\]
\[
c(1, 0) = c(1, 1) = 1
\]

Suppose, the goal is to minimize the quantity

\[
\limsup_{T \to \infty} \frac{1}{T} E_0^U T \sum_{t=0}^{T-1} c(x_t, u_t),
\]

over all admissible policies \( \Pi \in \Pi_A \).

Find the optimal policy and the optimal cost, as a function of \( \alpha, \beta, \kappa \). Explain your answer and how you arrived at your solution.

**Exercise 7.6.4** Consider a two-state, controlled Markov Chain with state space \( X = \{0, 1\} \), and transition kernel for \( t \in \mathbb{Z}_+ \):

\[
P(x_{t+1} = 0|x_t = 0) = u_t^0
\]
\[
P(x_{t+1} = 1|x_t = 0) = 1 - u_t^0
\]
\[
P(x_{t+1} = 1|x_t = 1) = u_t^1
\]
\[
P(x_{t+1} = 0|x_t = 1) = 1 - u_t^1.
\]

Here \( u_t^0 \in [0, 2, 1] \) and \( u_t^1 \in [0, 1] \) are the control variables. Suppose, the goal is to minimize the quantity

\[
\limsup_{T \to \infty} \frac{1}{T} E_0^U T \sum_{t=0}^{T-1} c(x_t, u_t),
\]

where

\[
c(0, u^0) = 1 + u^0,
\]
\[
c(1, u^1) = 1.5, \quad \forall u^1 \in [0, 1],
\]

with given \( \alpha, \beta \in \mathbb{R}_+ \).

Find an optimal policy and find the optimal cost.

**Hint:** Consider deterministic and stationary policies and analyze the costs corresponding to such policies.

**Exercise 7.6.5** [Machine repair revisited] Recall Exercise 7.6.5 with an average cost formulation. Show that there exists an optimal control policy and that this policy is stationary.

**Exercise 7.6.6** For an infinite horizon discounted cost partially observed Markov decision problem with finite state, action and measurement spaces, suppose that we wish to restrict the policies to be stationary control policies which only are based
on the most recent observation; that is $u_t = \gamma(y_t)$ for some $\gamma : \mathcal{Y} \to \mathcal{U}$ (clearly, this is suboptimal among all admissible policies, as the analysis in the Chapter shows). Given this restrictive class of policies, can one obtain the optimal policy through linear programming?
8 Learning and Approximation Methods

8.1 Value and Policy Iteration Algorithms

8.1.1 Value Iteration

For discounted cost infinite horizon problems, the value iteration algorithm is an algorithm which is guaranteed to converge to the value function (under an optimal policy). This algorithm follows from the repeated application of the contraction operator discussed earlier in the context of Theorem 5.5.2.

8.1.2 Policy Iteration

We now discuss the Policy Iteration Algorithm. Let $\mathcal{X}$ be countable and $c$ be a bounded cost function. Consider a discounted optimal control problem and let $\gamma_0^\infty$ denote a deterministic stationary policy (which naturally leads to a finite discounted expected cost here), let this expected cost be: $W_0(\cdot)$. That is,

$$ W_0(x) = E_x^{\gamma_0^\infty} \left[ \sum_{k=0}^{\infty} \beta^k c(x_k, \gamma_0(x_k)) \right] $$

Then,

$$ W_0(x) = c(x, \gamma_0(x)) + \beta \int W_0(x_{t+1}) P(dx_{t+1}|x_t = x, u_t = \gamma_0(x)) $$

Let

$$ T(W_0)(x) = \min_{u \in \mathcal{U}} \left( c(x, u) + \beta \int W_0(x_{t+1}) P(dx_{t+1}|x_t = x, u_t = u) \right). $$

Clearly $T(W_0) \leq W_0$ pointwise. Now, let $\gamma_1$ be such that

$$ T(W_0)(x) = c(x, \gamma_1(x)) + \beta \int W_0(x_{t+1}) P(dx_{t+1}|x_t = x, u_t = \gamma_1(x)) \tag{8.1} $$

Now, let

$$ W_1(x) = E_x^{\gamma_1^1} \left[ \sum_{k=0}^{\infty} \beta^k c(x_k, \gamma_1(x_k)) \right] $$

so that,

$$ W_1(x) := c(x, \gamma_1(x)) + \beta \int W_1(x_{t+1}) P(dx_{t+1}|x_t = x, u_t = \gamma_1(x)). $$

Observe that, iterative application of (8.1) one more time implies that
Stochastic approximation methods are used extensively in many application areas. A typical stochastic approximation has the following form:

\[ W_0(x) \geq E^\gamma_x \left[ \sum_{k=0}^{1} \beta^k c(x_k, \gamma_1(x_k)) + \beta^2 \int W_0(x_2)P(dx_2|x_1 = x, u_1 = \gamma_1(x)) \right]. \]

and thus for any \( x \in \mathbb{X}, n \in \mathbb{Z}_+ \):

\[ W_0(x) \geq E^\gamma_x \left[ \sum_{k=0}^{n-1} \beta^k c(x_k, \gamma_1(x_k)) \right] + \beta^n \int W_0(x_n)P(dx_n|x_{n-1} = x, u_{n-1} = \gamma_1(x)). \] (8.2)

This leads to \( W_1(x) \leq T(W_0)(x) \leq W_0(x) \). We can continue this reasoning for \( n \geq 1 \).

**Remark 8.1.** Note that for a finite state action problem, the expression

\[ W_1(x) = E^\gamma_x \left[ \sum_{k=0}^{\infty} c(x_k, \gamma_1(x_k)) = c(x, \gamma^1(x)) + \beta \sum W_1(x_{t+1})P(dx_{t+1}|x_t = x, u_t = \gamma_1(x)) \right] \]

or more generally for a given stationary \( \gamma \):

\[ J_\beta(x, \gamma) = E^\gamma_x \left[ \sum_{k=0}^{\infty} c(x_k, \gamma(x_k)) \right] = c(x, \gamma(x)) + \beta \sum W_1(x_{t+1})P(dx_{t+1}|x_t = x, u_t = \gamma(x)) \]

can be computed by solving the following matrix equation, which can easily be implemented in Matlab:

\[ W = c_\gamma + \beta P^\gamma W, \]

where \( W \) is a column vector consisting of elements \( c(x, \gamma(x)) \) and \( P^\gamma \) is a stochastic matrix with entries \( P(x_{t+1} = x|u_t = x, u_t = \gamma(x)) \).

**Theorem 8.1.1** If there exists \( n \in \mathbb{N} \) such that \( W_n(x) = W_{n+1}(x) \), then \( W = W_n \) solves \( W = T(W) \), and thus it leads to an optimal policy if under some policy \( \Pi \)

\[ E^\Pi_x \left[ \sum_{k=0}^{\infty} \beta^k c(x_k, \Pi_k(x_k)) \right] \leq D(x) \]

for some measurable function \( D(x) \). As a result, one arrives at an optimal policy. For a problem with finite state and action spaces, convergence is guaranteed. With \( n \to \infty \), \( W_n \downarrow W \), and \( W \) again satisfies \( W = T(W) \).

**Proof.** Note that by (8.2) the sequence \( W_n \geq T(W_n) \geq W_{n+1} \), and thus there is a limit \( W \) (since the cost per state is bounded from below). Furthermore, by Lemma 5.5 it follows that the limit \( W \) satisfies \( W = T(W) \) (See Remark 5.5). \( \diamond \)

### 8.2 Stochastic Learning Algorithms

In some Markov Decision Problems, one does not know the true transition kernel, or the cost function and may wish to use past data to obtain an asymptotically optimal solution. In some problems, this may be used as an efficient numerical method to obtain approximately optimal solutions.

There may also be algorithms where a prior probabilistic knowledge on the system dynamics may be used to learn the true system. One may apply Bayesian or non-Bayesian methods.

Stochastic approximation methods are used extensively in many application areas. A typical stochastic approximation has the following form:

\[ x_{t+1} = x_t + \alpha_t \left( F(x_t) - x_t + w_t \right) \] (8.3)
where \( w_t \) is a noise variable. The goal is to arrive at a point \( x^* \) which satisfies \( x^* = F(x^*) \). For a further example, see Exercise 4.5.14.

### 8.2.1 Q-Learning

Q-learning [125, 115, 12] is a stochastic approximation algorithm that does not require the knowledge of the transition kernel, or even the cost (or reward) function for its implementation. In this algorithm, the incurred per-stage cost variable is observed through simulation of a single sample path. When the state and the action spaces are finite, under mild conditions regarding infinitely often hits for all state-action pairs, this algorithm is known to converge to the optimal cost. We now discuss this algorithm.

Consider a Markov Decision Problem with finite state and action sets with the objective of finding

\[
\inf_{\Pi} E_{x_0}^\Pi \left[ \sum_{t=0}^{\infty} \beta^t c(x_t, u_t) \right].
\]

for some \( \beta \in (0, 1) \).

Let \( Q : X \times U \rightarrow \mathbb{R} \) denote the Q-factor of a decision maker. Let us assume that the decision maker uses a stationary random policy \( \Pi : X \mapsto \mathcal{P}(U) \) and updates its Q-factors as: for \( t \geq 0 \),

\[
Q_{t+1}(x, u) = Q_t(x, u) + \alpha_t(x, u) \left( c(x, u) + \beta \min_v Q_t(x_{t+1}, v) - Q_t(x, u) \right)
\]

(8.4)

where the initial condition \( Q_0 \) is given, \( \alpha_t(x, u) \) is the step-size for \( (x, u) \) at time \( t \), \( u_t \) is chosen according to some policy \( \Pi_t \), and the state \( x_t \) evolves according to \( P(\cdot | x_t, u_t) \) starting at \( x_0 \).

**Assumption 8.2.1** For all \( (x, u), t \geq 0 \),

a) \( \alpha_t(x, u) \in [0, 1] \)

b) \( \alpha_t(x, u) = 0 \) unless \( (x, u) = (x_t, u_t) \)

c) \( \alpha_t(x, u) \) is a (deterministic) function of \( (x_0, u_0), \ldots, (x_t, u_t) \). This can also be made only a function of \( t, x \) and \( u \).

d) \( \sum_{t \geq 0} \alpha_t(x, u) \rightarrow \infty \), w.p. 1

e) \( \sum_{t \geq 0} \alpha_t^2(x, u) \leq C \), w.p. 1, for some (deterministic) constant \( C < \infty \).

A common way to pick \( \alpha \) coefficients in the algorithm is to take for every \( x, u \) pair:

\[
\alpha_t(x, u) = \frac{1}{1 + \sum_{k=0}^{t} 1_{(x_k=x, u_k=u)}},
\]

Let \( F \) be an operator on the Q factors defined by:

\[
F(Q)(x, u) = c(x, u) + \beta \sum_{x'} P(x'|x, u) \min_v Q(x', v),
\]

(8.5)

where \( P(x'|x, u) = P(x_1 = x'|x_0 = x, u_0 = u) \) is the transition kernel. Consider the following fixed point equation.

\[
Q^*(x, u) = F(Q^*)(x, u) = c(x, u) + \beta \sum_{x'} P(x'|x, u) \min_v Q^*(x', v)
\]

(8.6)

whose existence follows from the contraction arguments as in Chapter 5, by using the norm \( \|Q\| = \max (x, u)|Q(x, u)| \).

We can write (8.4) as
Our main objective in this section is to find conditions on the components of the MDP under which there exists a sequence of stationary quantizer policies induced by \( \pi \).

Definition 8.2. A quantizer from \( X \) to \( A \) is called a quantizer if the range of \( q \), i.e., \( q(X) = \{ q(x) : x \in X \} \), is finite.

The elements of \( q(X) \) (the possible values of \( q \)) are called the levels of \( q \). The rate \( R = \log_2 |q(X)| \) of a quantizer \( q \) (approximately) represents the number of bits needed to losslessly encode the output levels of \( q \) using binary codewords of equal length. Let \( Q \) denote the set of all quantizers from \( X \) to \( A \). A deterministic stationary quantizer policy is a constant sequence \( \pi = \{ \pi_t \} \) of stochastic kernels on \( A \) given \( X \) such that \( \pi_t(\cdot | x) = \delta_{q(x)}(\cdot) \) for all \( t \) and some \( q \in Q \). For any finite set \( A \subseteq A \), let \( Q(A) \) denote the set of all elements in \( Q \) having range \( A \). Analogous with \( F \), the set of all deterministic stationary quantizer policies induced by \( Q(A) \) will be identified with the set \( Q(A) \).

Our main objective in this section is to find conditions on the components of the MDP under which there exists a sequence of finite subsets \( \{ A_n \}_{n \geq 1} \) of \( A \) for which the following holds:

\[
Q_{t+1}(x,u) = Q_t(x,u) + \alpha_t(x,u) \left( F(Q_t)(x,u) - Q_t(x,u) + \left( c(x,u) + \beta \min_v Q_t(x_{t+1},v) - F(Q_t)(x,u) \right) \right)
\]

which is in the same form as (8.3) since \( c(x,u) + \beta \min_v Q_t(x_{t+1},v) - F(Q_t)(x,u) \) is a zero-mean variable.

Theorem 8.2.1 Under Assumption 8.2.1 the algorithm (8.4) converges almost surely to \( Q^* \). A policy \( \pi \) which satisfies \( \min_u Q^*(x,u) = Q^*(x,f^*(x)) \) is an optimal policy.

Proof Sketch. From (8.7), the process \( Q_t \) satisfies the following form, with \( S_t = Q_t - Q^*_t \):

\[
S_{t+1}(x,u) = (1 - \alpha_t(x)) S_t(x,u) + \alpha_t(x,u) (G(S_t)(x,u) + w_t),
\]

where \( \{ \alpha \} \) satisfies Assumption 8.2.1 \( |G(S_t)| \leq \beta |S_t| \) and \( w_t \) has a uniformly bounded second moment for all realizations of the filtration. In particular, through taking the square of this term, one can first show that, using arguments similar to those in Theorem 4.2.2 that the expectation of \( \sum_t \alpha_t S_t^2 \) is uniformly bounded provided that \( E[S_t^2] \) remains bounded. The condition for the boundedness of \( E[S_t^2] \) can be established using the properties of the evolution. Since \( \alpha_t \) is not summable, there exists a sequence of stopping times so that each summation between the times is bounded from below. Suppose that \( E[S_t^2] \) does not converge to zero. Consider then a subsequence for which \( E[S_t^2] \) is above a given lower bound; through a careful analysis, a contradiction can be obtained. Finally, the uniform boundedness allows for the convergence to be in the almost sure sense. We refer the reader to [115] or to the proof of Theorem 1 in [113] for complete proofs.

Roll-out Algorithms

Roll-out algorithms, also known as sliding-horizon or receding horizon algorithms is an important algorithm which is probably optimal as the horizon length increases. We refer the reader to [108], [102], [39] and [18] among many other papers in this direction.

8.3 Approximation through Quantization of the State and the Action Spaces

This section is based on [104], [102], [103] and [107].

8.3.1 Finite Action Approximation to MDPs

To give a precise definition of the problem we study in this section, we first give the definition of a quantizer from the state to the action space.

Definition 8.2. A measurable function \( q : X \to A \) is called a quantizer from \( X \) to \( A \) if the range of \( q \), i.e., \( q(X) = \{ q(x) : x \in X \} \), is finite.
(P) For any initial point \( x \), we have \( \lim_{n \to \infty} \inf_{q \in Q(A_n)} J(q, x) = \inf_{f \in \mathcal{F}} J(f, x) \) (or \( \lim_{n \to \infty} \inf_{q \in Q(A_n)} V(q, x) = \inf_{f \in \mathcal{F}} V(f, x) \) for the average cost), provided that the set \( \mathcal{F} \) of deterministic stationary policies is an optimal class for the MDP.

In other words, if for each \( n \), MDP\(_n\) is defined as the Markov decision process having the components \( \{X, A_n, p, c\} \), then (P) is equivalent to stating that value function of MDP\(_n\) converges to the value function of the original MDP.

Near Optimality of Quantized Policies Under Strong Continuity

In this section we consider the problem (P) for the MDPs with strongly continuous transition probability. We impose the assumptions below on the components of the Markov decision process.

**Assumption 8.3.1**

(a) The one stage cost function \( c \) is nonnegative and bounded satisfying \( c(x, \cdot) \in C_b(A) \) for all \( x \in X \).

(b) The stochastic kernel \( p(\cdot|x,a) \) is setwise continuous in \( a \in A \).

(c) \( A \) is compact.

Let \( d_A \) denote the metric on \( A \). Since the action space \( A \) is compact and thus totally bounded, one can find a sequence of finite sets \( \Lambda = \{a_{n,1}, \ldots, a_{n,k_n}\} \subset A \) such that for all \( n \),

\[
\min_{i \in \{1, \ldots, k_n\}} d_A(a, a_{n,i}) < 1/n \quad \text{for all } a \in A.
\]

In other words, \( A_n \) is a \( 1/n \)-net in \( A \). In the rest of Section 8.3.1 we assume that the sequence \( \{A_n\}_{n \geq 1} \) is fixed. To ease the notation in the sequel, let us define the mapping \( T_n : \mathcal{F} \to Q(A_n) \) as

\[
T_n(f)(x) = \min_{a \in A_n} d_A(f(x), a),
\]

where ties are broken so that \( T_n(f)(x) \) is measurable.

**Discounted Cost** We consider here problem (P) for the discounted cost with a discount factor \( \beta \in (0, 1) \) under the Assumption [8.3.1]. The following theorem is the main result of this section which states that for any \( f \in \mathcal{F} \), the discounted cost function of \( T_n(f) \in Q(A_n) \) converges to the discounted cost function of \( f \) as \( n \to \infty \). Therefore, it implies that the discounted value function of the MDP\(_n\) converges to the discounted value function of the original MDP.

**Theorem 8.3.** Let \( f \in \mathcal{F} \) and \( \{T_n(f)\} \) be the quantized approximations of \( f \). Then, \( J(T_n(f), x) \to J(f, x) \) as \( n \to \infty \), for all \( x \in X \).

**Average Cost**

In contrast to the discounted cost criterion, the expected average cost is in general not sequentially continuous with respect to strategic measures for the \( ws^\infty \) topology under practical assumptions. Hence, we develop an approach based on the convergence of the sequence of invariant probability measures under quantized stationary policies to solve (P) for the average cost criterion. First, observe that any deterministic stationary policy \( f \) defines a stochastic kernel on \( X \) given \( X \) via

\[
Q_f(\cdot|x) := p(\cdot|x, f(x)).
\]

Let \( Q_f^t \) denote the \( t \)-step transition probability of this Markov chain. If \( Q_f \) admits an ergodic invariant probability measure \( \nu_f \), then by [63] Theorem 2.3.4 and Proposition 2.4.2, there exists an invariant set \( M_f \in B(X) \) with full \( \nu_f \) measure such that for all \( x \) in that set we have

\[
V(f, x) = \int_X c(x, f(x))\nu_f(dx).
\]
The following assumptions will be imposed for the average cost case.

**Assumption 8.3.2** Suppose Assumption 8.3.1 holds. In addition, we have

1. For any \( f \in \mathbb{F} \), \( Q_f \) has a unique invariant probability measure \( \nu_f \).
   
2. The set of invariant probability measures \( \Gamma_F := \{ \nu \in \mathcal{P}(X) : \nu Q_f = \nu \text{ for some } f \in \mathbb{F} \} \) is relatively sequentially compact in the setwise topology.
   
3. There exists \( x \in X \) such that for all \( B \in B(X) \), \( Q^*_f(B|x) \to \nu_f(B) \) uniformly in \( f \in \mathbb{F} \).
4. \( M := \bigcap_{f \in \mathbb{F}} M_f \neq \emptyset \).

The following theorem is the main result of this section. It states that for any \( f \in \mathbb{F} \), the average cost function of \( \Upsilon_n(f) \in Q(\Lambda_n) \) converges to the average cost function of \( f \) as \( n \to \infty \). In other words, the average value function of MDP_n converges to the average value function of the original MDP.

**Theorem 8.4.** Let \( x \in M \) and \( f \in \mathbb{F} \). Then, we have \( V(\Upsilon_n(f), x) \to V(f, x) \) as \( n \to \infty \), under Assumption 8.3.2 with (f1) or (f2).

Note that Assumption 8.3.2 (e),(f2),(g) are satisfied under any of the conditions \( Ri, i \in \{0, 1, 1(a), 1(b), 2, \ldots, 6\} \) in [64]. Moreover, \( M = X \) in Assumption 8.3.2 (g) if at least one of the above conditions holds. Moreover, the condition

\[ p(\cdot | x, a) \leq \zeta(\cdot) \quad \text{for all } x \in X, \ a \in A \quad \text{for some finite measure } \zeta \text{ on } X, \]

implies Assumption 8.3.2 (f1).

**Near Optimality of Quantized Policies Under Weak Continuity**

In this section, we consider (P) for MDPs with weakly continuous transition probability. Specifically, we will show that the value function of MDP_n converges to the value function of the original MDP, which is equivalent to (P).

**Discounted Cost**

Here, we consider the discounted cost case with a discount factor \( \beta \in (0, 1) \). The following assumptions will be imposed for both the discounted cost and the average cost. These assumptions are used in the literature for studying discounted Markov decision processes with unbounded one-stage cost and weakly continuous transition probability.

**Assumption 8.3.3**

1. The one stage cost function \( c \) is nonnegative and continuous.
2. The stochastic kernel \( p(\cdot | x, a) \) is weakly continuous in \((x, a) \in X \times A\).
3. \( A \) is compact.
4. There exist nonnegative real numbers \( M \) and \( \alpha \in \left[1, \frac{1}{\beta}\right) \), and a continuous weight function \( w : X \to [1, \infty) \) such that for each \( x \in X \), we have
   
   \[
   \sup_{a \in A} c(x, a) \leq M w(x), \tag{8.10}
   \]
   
   \[
   \sup_{a \in A} \int_X w(y) p(dy|x, a) \leq \alpha w(x), \tag{8.11}
   \]
   
   and \( \int_X w(y) p(dy|x, a) \) is continuous in \((x, a)\).
8.3 Approximation through Quantization of the State and the Action Spaces

For any real-valued measurable function $u$ on $X$, let $T_u : X \to \mathbb{R}$ is given by

$$
T_u(x) := \min_{a \in A} \left[ c(x, a) + \beta \int_X u(y)p(dy|x, a) \right].
$$

(8.12)

In the literature $T$ is called the Bellman optimality operator for the MDP. Analogously, let us define the Bellman optimality operator $T_n$ of MDP$_n$ as

$$
T_n u(x) := \min_{a \in A_n} \left[ c(x, a) + \beta \int_X u(y)p(dy|x, a) \right].
$$

(8.13)

It can be shown that both $T$ and $T_n$ are contraction operators with modulus $\sigma = \alpha \beta$ mapping $C_u(X)$ into itself. Furthermore, value functions of MDP and MDP$_n$ are fixed points of these operators; that is, $TJ^* = J^*$ and $T_nJ_n^* = J_n^*$. Let us define $v^0 = v^0_n = 0$, and $v^t = T^t$ and $v^{t+1}_n = T_n^t v^t_n$ for $t \geq 1$; that is, $\{v^t\}_{t \geq 1}$ and $\{v^t_n\}_{t \geq 1}$ are successive approximations to the discounted value functions of the MDP and MDP$_n$, respectively.

Lemma 8.5. [106] For any compact set $K \subset X$ and $t \geq 1$, we have

$$
\lim_{n \to \infty} \sup_{x \in K} |v^t_n(x) - v^t(x)| = 0.
$$

(8.14)

The theorem below is the main result of this section which states that the discounted value function of MDP$_n$ converges to the discounted value function of the original MDP. It can be proved by using Lemma 8.5 and taking into account that $\{v^t\}_{t \geq 1}$ and $\{v^t_n\}_{t \geq 1}$ are successive approximations to $J^*$ and $J_n^*$, respectively.

Theorem 8.6. [106] For any compact set $K \subset X$ we have

$$
\lim_{n \to \infty} \sup_{x \in K} |J_n^*(x) - J^*(x)| = 0.
$$

(8.15)

Average Cost

In this section we prove an approximation result analogous to Theorem 8.6 for the average cost case. To do this, some new assumptions are needed on the components of the original MDP in addition to Assumption 8.3.3. A version of these assumptions was used in [121] and [55] to show the existence of the solution to the Average Cost Optimality Equality (ACOE) and Inequality (ACOI).

Assumption 8.3.4 Suppose Assumption 8.3.3 holds with (8.11) replaced by condition (e) below. Moreover, suppose there exist a non-degenerate finite measure $\lambda$ on $\mathbb{X}$ and a positive constant $b$ such that

(e) $\int_{\mathbb{X}} w(y)p(dy|x, a) \leq \alpha w(x) + b$ for all $(x, a) \in \mathbb{X} \times \mathbb{U}$, where $\alpha \in (0, 1)$.

(f) $p(D|x, a) \geq \lambda(D)$ for all $(x, a) \in \mathbb{X} \times \mathbb{U}$ and $D \in (\mathbb{X})$.

(g) The weight function $w$ is $\lambda$-integrable.

Remark 8.7. In the remainder of this section, we let $b = \int w d\lambda$ without loss of generality. Indeed, (i) if $b \leq \int w d\lambda$, we can replace $b$ with $\int w d\lambda$ and the inequality in Assumption 8.3.4(e) is still true. Conversely, (ii) if $\int w d\lambda < b$, then by first increasing the value of $\alpha$ so that $\lambda(\mathbb{X}) + \alpha > 1$, and then adding a constant $k$ to $w$, where

$$
k = \frac{b - \int w d\lambda}{\lambda(\mathbb{X}) + \alpha - 1},
$$

we obtain $b < \int w d\lambda$. Then, as in (i) above, we set $b = \int w d\lambda$ and Assumption 8.3.4(e) now holds for the new $w$ and $\alpha$.

In this section, we suppose that Assumption 8.3.4 holds. The following theorem states that there is a solution to the average cost optimality equation (ACOE) and the stationary policy which minimizes this ACOE is an optimal policy.
**Theorem 8.8.** We have

\[
\lim_{n \to \infty} |V^*_n - V^*| = 0,
\]

where \(V^*\) and \(V^*_n\) \((n \geq 1)\) do not depend on \(x\).

**Remark 8.9.** When one considers partially observed MDPs (POMDPs), it is known that any POMDP can be reduced to a (completely observable) MDP \([142]\) whose states are the posterior state distributions or beliefs of the observer. One can show that setwise continuity of the reduced MDP is not possible even under very strict conditions, whereas weak continuity can be satisfied under reasonable conditions on the transition kernel and the continuity of the measurement channel. Thus the results in this section are applicable to POMDPs; see \([106]\).

One can also obtain rates of convergence results \([1]\).

### 8.3.2 Finite State Approximation to MDPs

In this section our aim is to study the finite-state approximation problem for discrete time Markov decision processes, by reducing it to a finite state MDP obtained through quantization of the state space on a finite grid. In particular, we study the following two problems.

- **(Q1)** Under what conditions on the components of the MDP do the true cost functions of the policies obtained from finite models converge to the optimal value function as the number of grid points goes to infinity?
- **(Q2)** Can we obtain bounds on the performance loss due to discretization in terms of the number of grid points if we strengthen the conditions sufficient in \((Q1)\)?

The approach to solve problem \((Q1)\) can be summarized as follows: (i) first, we obtain approximation results for the compact-state case, (ii) we find conditions under which a compact representation leads to near optimality for non-compact state MDPs, (iii) we obtain the convergence of the finite-state models to non-compact models. A by-product of this analysis, we obtain **compact-state-space approximations** for an MDP with non-compact Borel state space. In particular, our findings directly lead to finite models if the state space is countable.

Here, we consider \((Q1)\) for the MDPs with compact state space. To distinguish compact-state MDPs from non-compact ones, the state space of the compact-state MDPs will be denoted by \(Z\) instead of \(X\). We impose the assumptions below on the components of the Markov decision process.

**Assumption 8.3.5**

\((a)\) The one-stage cost function \(c\) is in \(C_b(Z \times A)\).

\((b)\) The stochastic kernel \(p(\cdot | z, a)\) is weakly continuous in \((z, a)\) and setwise continuous in \(a\).

\((c)\) \(Z\) and \(A\) are compact.

Let \(d_Z\) denote the metric on \(Z\). Since the state space \(Z\) is assumed to be compact and thus totally bounded, one can find a sequence \(\{\{z_{n,i}\}_{i=1}^{k_n}\}_{n \geq 1}\) of finite grids in \(Z\) such that for all \(n\),

\[
\min_{i \in \{1, \ldots, k_n\}} d_Z(z, z_{n,i}) < 1/n \text{ for all } z \in Z.
\]

Let \(Z_n := \{z_{n,1}, \ldots, z_{n,k_n}\}\) and define function \(Q_n\) mapping \(Z\) to \(Z_n\) by

\[
Q_n(z) := \arg \min_{z_{n,i} \in Z_n} d_Z(z, z_{n,i}),
\]
where ties are broken so that \( Q_n \) is measurable. For each \( n, Q_n \) induces a partition \( \{ S_{n,i} \}_{i=1}^{k_n} \) of the state space \( Z \) given by

\[
S_{n,i} = \{ z \in Z : Q_n(z) = z_{n,i} \}.
\]

Let \( \{ \nu_{n,i} \} \) be a sequence of probability measures on \( Z \) satisfying \( \nu_{n,i}(S_{n,i}) > 0 \) for all \( i, n \). We let \( \nu_{n,i} \) be the restriction of \( \nu \) to \( S_{n,i} \) defined by \( \nu_{n,i}(\cdot) := \frac{\nu(\cdot)}{\nu(S_{n,i})} \). The measures \( \nu_{n,i} \) will be used to define a sequence of finite-state MDPs, denoted as MDP\(_n\) (\( n \geq 1 \)), to approximate the original model. To this end, for each \( n \) define the one-stage cost function \( c_n : Z \times A \to [0, \infty) \) and the transition probability \( p_n \) on \( Z_n \) given \( Z \times A \) by

\[
c_n(z_{n,i}, a) := \int_{S_{n,i}} c(z, a) \nu_{n,i}(dz), \quad p_n(\cdot | z_{n,i}, a) := \int_{S_{n,i}} Q_n \ast p(\cdot | z, a) \nu_{n,i}(dz),
\]

where \( Q_n \ast p(\cdot | z, a) \in P(Z_n) \) is the pushforward of the measure \( p(\cdot | z, a) \) with respect to \( Q_n \); that is,

\[
Q_n \ast p(z_{n,j} | z, a) = p(\{ z \in Z : Q_n(z) = z_{n,j} \} | z, a),
\]

for all \( z_{n,j} \in Z_n \). For each \( n \), we define MDP\(_n\) as a Markov decision process with the following components: \( Z_n \) is the state space, \( A \) is the action space, \( p_n \) is the transition probability and \( c_n \) is the one-stage cost function. History spaces, policies and cost functions are defined in a similar way as in the original model.

**Discounted Cost**

Here we consider \((Q1)\) for the discounted cost criterion with a discount factor \( \beta \in (0, 1) \). Recall the Bellman optimality operator \( T \) defined in \((8.12)\). Define also the operator \( T_n \), which is the Bellman optimality operator for MDP\(_n\), by

\[
T_n u(z_{n,i}) = \min_{a \in A} \int_{S_{n,i}} \left[ c(z, a) + \beta \int_{Z} \hat{u}(y)p(dy | z, a) \right] \nu_{n,i}(dz),
\]

where \( u : Z_n \to \mathbb{R} \) and \( \hat{u} \) is the piecewise constant extension of \( u \) to \( Z \) given by \( \hat{u}(z) = u \circ Q_n(z) \). Under Assumption 8.3.5 the fixed point of \( T_n \) is the value function \( J_n^* \) of MDP\(_n\) and there exists an optimal stationary policy \( f_n^* \) for MDP\(_n\). Hence, we have \( J_n^* = T_n J_n^* = T_n J_n(f_n^*, \cdot) = J_n(f_n^*, \cdot) \), where \( J_n \) denotes the discounted cost for MDP\(_n\). Let us extend \( T_n \) to the set of all bounded measurable functions on \( Z \) as follows:

\[
\hat{T}_n u(z) := \min_{a \in A} \int_{S_{n,i}(z)} \left[ c(x, a) + \beta \int_{Z} u(y)p(dy | x, a) \right] \nu_{n,i}(z)(dx), \tag{8.16}
\]

where \( i_n : Z \to \{1, \ldots, k_n\} \) maps \( z \) to the index of the partition \( \{ S_{n,i} \} \) it belongs to. Since \( \hat{T}_n (u \circ Q_n) = (T_n u) \circ Q_n \) for all \( u \in B(Z_n) \), we have

\[
\hat{T}_n (J_n^* \circ Q_n) = (T_n J_n^*) \circ Q_n = J_n^* \circ Q_n.
\]

Hence, the fixed point of \( \hat{T}_n \) is the piecewise constant extension of the fixed point of \( T_n \).

**Remark 8.10**. In the rest of this chapter, when we take the integral of any function with respect to \( \nu_{n,i_n}(z) \), it is tacitly assumed that the integral is taken over all set \( S_{n,i_n}(z) \). Hence, we can drop \( S_{n,i_n}(z) \) in the integral for the ease of notation.

We now define another operator \( F_n \) on \( B(Z) \) by simply interchanging the order of the minimum and the integral in \((8.16)\), i.e.,

\[
F_n u(z) := \int_{A} \min_{a \in A} \left[ c(x, a) + \beta \int_{Z} u(y)p(dy | x, a) \right] \nu_{n,i_n}(z)(dx).
\]

We note that \( F_n \) is the extension (to infinite state spaces) of the operator defined in [101] p. 236 for the proposed approximate value iteration algorithm. However, unlike in [101], \( F_n \) will serve here as an intermediate point between \( T \) and \( T_n \).
converges to the optimal value function $J^*$ (i.e., the value function) of $T$ as $n$ goes to infinity.

**Theorem 8.11.** If $u^*_n$ is the unique fixed point of $F_n$, then $\lim_{n \to \infty} \|u^*_n - J^*\| = 0$.

The next step is to show that the fixed point $\hat{J}^*_n$ of $\hat{T}_n$ converges to the fixed point $J^*$ of $T$. This follows from Theorem 8.11 and the following result: for any $u \in C_b(Z)$, $\|\hat{T}_n u - F_n u\| \to 0$ as $n \to \infty$.

**Theorem 8.12.** [2, 107] The fixed point $\hat{J}^*_n$ of $\hat{T}_n$ converges to the fixed point $J^*$ of $T$. Therefore, the (extended) value function $\hat{J}^*_n \circ Q_n$ of MDP converges to the value function of MDP.

Recall the optimal stationary policy $f^*_n$ for MDP, and extend it to $Z$ by letting $\hat{f}_n(z) = f^*_n \circ Q_n(z)$. Since $\hat{J}^*_n = J^*_n \circ Q_n$, one can show that $\hat{f}_n$ is the optimal selector of $\hat{T}_n \hat{J}^*_n$; that is, $\hat{T}_n \hat{J}^*_n = \hat{T}_{\hat{f}_n} \hat{J}^*_n$, where $\hat{T}_{\hat{f}_n}$ is defined as

$$\hat{T}_{\hat{f}_n} u(z) := \int \left[ c(x, \hat{f}_n(x)) + \beta \int_Z u(y)p(dy|x, \hat{f}_n(x)) \right] \nu_{n, \hat{f}_n}(z)(dx).$$

Define analogously

$$T_{\hat{f}_n} u(z) := c(z, \hat{f}_n(z)) + \beta \int_Z u(y)p(dy|z, \hat{f}_n(z)).$$

It is known that the fixed point of $T_{\hat{f}_n}$ is the true cost function of the stationary policy $\hat{f}_n$ (i.e., $J(\hat{f}_n, z)$).

The following theorem states that the cost function under the policy $\hat{f}_n$ converges to the value function $J^*$ as $n \to \infty$. It follows from Theorem 8.12 and the following result: $\|T_{\hat{f}_n} u - T_{\hat{f}_n} u\| \to 0$ as $n \to \infty$, for any $u \in C_b(Z)$.

**Theorem 8.13.** [2, 107] The discounted cost of the policy $\hat{f}_n$, obtained by extending the optimal policy $f^*_n$ of MDP to $Z$, converges to the optimal value function $J^*$ of the original MDP $\lim_{n \to \infty} \|J(\hat{f}_n, \cdot) - J^*\| = 0$.

Therefore, to find a near optimal policy for the original MDP, it is sufficient to compute the optimal policy of MDP, for sufficiently large $n$, and then extend this policy to the original state space.

**Average Cost**

In this section we impose some new conditions on the components of the original MDP in addition to Assumption 8.3.5 to solve (Q1) for the average cost. A version of first two conditions were imposed in [121] to show the existence of the solution to the Average Cost Optimality Equation (ACOE) and the optimal stationary policy by using the fixed point approach.

**Assumption 8.3.6** Suppose Assumption 8.3.5 holds with item (b) replaced by condition (f) below. In addition, there exist a non-trivial finite measure $\zeta$ on $Z$, a nonnegative measurable function $\theta$ on $Z \times \mathcal{A}$, and a constant $\lambda \in (0, 1)$ such that for all $(z, a) \in Z \times \mathcal{A}$

(d) $p(B|z, a) \geq \zeta(B)\theta(z, a)$ for all $B \in \mathcal{B}(Z),$

(e) $\frac{1}{\lambda} \leq \theta(z, a),$

(f) The stochastic kernel $p(\cdot | z, a)$ is continuous in $(z, a)$ with respect to the total variation distance.

The following theorem is a consequence of [55] Lemma 3.4 and Theorem 2.6 and [121] Theorems 3.3.

**Theorem 8.14.** [2, 107] Under Assumptions 8.3.6 the following holds. For each $f \in \mathcal{F}$, the stochastic kernel $Q_f(\cdot | x)$ has an unique invariant probability measure $\mu_f$. Therefore, we have $V(f, z) = \int_Z c(z, f(z))\mu_f(dz) =: \rho_f$. There exist positive real numbers $R$ and $\kappa < 1$ such that for every $z \in Z$
$\sup_{f \in \mathcal{F}} \|Q_f^\lambda (\cdot | z) - \mu_f\|_{TV} \leq R \kappa \lambda$, 

where $R$ and $\kappa$ continuously depend on $\zeta(Z)$ and $\lambda$. Furthermore, there exist $f^* \in \mathcal{F}$ and $h^* \in \mathcal{B}(Z)$ such that the triplet $(h^*, f^*, \rho_f^*)$ satisfies the average cost optimality inequality (ACOI) and therefore,

$$\inf_{\pi \in \Pi} V(\pi, z) =: V^*(z) = \rho_{f^*}.$$ 

For each $n$, define the one-stage cost function $b_n : Z \times A \to [0, \infty)$ and the stochastic kernel $q_n$ on $Z$ given $Z \times A$ as

$$b_n(z, a) := \int c(x, a) \nu_{n,i_n(z)}(dx),$$

$$q_n(\cdot | z, a) := \int p(\cdot | x, a) \nu_{n,i_n(z)}(dx).$$

Observe that $c_n$ (i.e., the one stage cost function of $\tilde{\text{MDP}}_n$) is the restriction of $b_n$ to $Z_n$, and $p_n$ (i.e., the stochastic kernel of $\tilde{\text{MDP}}_n$) is the pushforward of the measure $q_n$ with respect to $Q_n$; that is, $c_n(z_{n,i}, a) = b_n(z_{n,i}, a)$ for all $i = 1, \ldots, k_n$ and $p_n(\cdot | z_{n,i}, a) = Q_n \ast q_n(\cdot | z_{n,i}, a)$. 

For each $n$, let $\tilde{\text{MDP}}_n$ be defined as a Markov decision process with the following components: $Z$ is the state space, $A$ is the action space, $q_n$ is the transition probability, and $b_n$ is the one-stage cost function. History spaces, policies and cost functions are defined in a similar way as before. We note that a careful analysis of $\tilde{\text{MDP}}_n$ reveals that its Bellman optimality operator is essentially the operator $\tilde{T}_n$. Hence, the value function of $\tilde{\text{MDP}}_n$ is the piecewise constant extension of the value function of $\text{MDP}_n$ for the discounted cost. A similar conclusion will be made for the average cost in Lemma 8.15.

First, notice that if we define

$$\theta_n(z, a) := \int \theta(y, a) \nu_{n,i_n(z)}(dy),$$

$$\zeta_n := Q_n \ast \zeta \text{ (i.e., pushforward of } \zeta \text{ with respect to } Q_n),$$

then it is straightforward to prove that for all $n$, $\tilde{\text{MDP}}_n$ satisfies Assumption 8.3.6(d),(e) when $\theta$ is replaced by $\theta_n$, and $\zeta$ (i.e., Assumption 8.3.6) is true for $\text{MDP}_n$ when $\theta_n$ and $\zeta_n$ are replaced by the restriction of $\theta_n$ to $Z_n$, and $\zeta_n$, respectively. Hence, Theorem 8.14 holds (with the same $R$ and $\kappa$) for $\tilde{\text{MDP}}_n$ and $\text{MDP}_n$ for all $n$. Therefore, we denote by $f_n^*$ and $f_n^*$ the optimal stationary policies of $\tilde{\text{MDP}}_n$ and $\text{MDP}_n$ with the corresponding average costs $\hat{\rho}_f^{n,z}$ and $\rho_f^{n,z}$, respectively. Furthermore, we also write $\tilde{\rho}_f^n$ and $\rho_f^n$ to denote the average cost of any stationary policy $f$ for $\tilde{\text{MDP}}_n$ and $\text{MDP}_n$, respectively. The corresponding invariant probability measures are also denoted in a same manner, with $\mu$ replacing $\rho$.

The following lemma essentially says that $\text{MDP}_n$ and $\tilde{\text{MDP}}_n$ are not very different.

**Lemma 8.15.** The stationary policy given by the piecewise constant extension of the optimal policy $f_n^*$ of $\text{MDP}_n$ to $Z$ (i.e., $f_n^* \circ Q_n$) is optimal for $\tilde{\text{MDP}}_n$ with the same cost function $\rho_f^{n,z}$. Hence, $f_n^* = f_n^* \circ Q_n$ and $\hat{\rho}_f^{n,z} = \tilde{\rho}_f^{n,z}$.

By Lemma 8.15 we can consider $\tilde{\text{MDP}}_n$ in place of $\text{MDP}_n$. The following theorem states that the value function of $\tilde{\text{MDP}}_n$ converges to the value function of $\text{MDP}$ as $n \to \infty$.

**Theorem 8.16.** [107] We have $\sup_{f \in \mathcal{F}} |\tilde{\rho}_f^n - \rho_f| \to 0$ as $n \to \infty$. Therefore, $|\tilde{\rho}_f^{n,z} - \rho_f^{n,z}| \to 0$ as $n \to \infty$.

The following theorem states that if one applies the piecewise constant extension of the optimal stationary policy of $\text{MDP}_n$ to the original MDP, the resulting cost function will converge to the value function of the original MDP. It follows from Lemma 8.15 and Theorem 8.16.

**Theorem 8.17.** [107] The average cost of the optimal policy $\tilde{f}_n^*$ for $\tilde{\text{MDP}}_n$, obtained by extending the optimal policy $f_n^*$ of $\text{MDP}_n$ to $Z$, converges to the optimal value function $J^* = \rho_f^*$ of the original MDP, i.e.,
\[
\lim_{n \to \infty} |\rho f_n - \rho f^*| = 0.
\]

Therefore, to find a near optimal policy for the original MDP, it is sufficient to compute the optimal policy of MDP_n for sufficiently large n, and then extend this policy to the original state space.

[103, 107] have also studied the non-compact state space setups, which we do not overview here.

### Discretization of the Action Space

For computing near optimal policies using well known algorithms, such as value iteration, policy iteration, and Q-learning, the action space must be finite.

It was shown in Theorems 8.3 and 8.4 that any MDP with (infinite) compact action space and with bounded one-stage cost function can be well approximated by an MDP with finite action space under assumptions that are satisfied by c-MDP_n for each n, for both the discounted cost and the average cost cases. Recall the sequence of finite subsets \( \{A_k\} \) of \( A \) from Section 8.3.1. We define c-MDP_n,k as the Markov decision process having the components \( \{X_n, A_k, p_n, c_n\} \) and we let \( \mathcal{F}_n(A_k) \) denote the set of all deterministic stationary policies for c-MDP_n,k. Note that \( \mathcal{F}_n(A_k) \) is the set of policies in \( \mathcal{F}_n \), taking values only in \( A_k \). Therefore, in a sense, c-MDP_n,k and c-MDP_n can be viewed as the same MDP, where the former has constraints on the set of policies. For each n and k, by an abuse of notation, let \( f_n^* \) and \( f_n,k^* \) denote the optimal stationary policies of c-MDP_n and c-MDP_n,k, respectively, for both the discounted and average costs. Then Theorems 8.3 and 8.4 show that for all n, we have

\[
\lim_{k \to \infty} J_n(f_n,k^*, x) = J_n(f_n^*, x) := J_n^*(x)
\]

\[
\lim_{k \to \infty} V_n(f_n,k^*, x) = V_n(f_n^*, x) := V_n^*(x)
\]

for all \( x \in X_n \). In other words, the discounted and average value functions of c-MDP_n,k converge to the discounted and average value functions of c-MDP_n as \( k \to \infty \).

Let us fix \( x \in X \). For n sufficiently large (so \( x \in K_n \)), we choose \( k_n \) such that \( |J_n(f_n,k_n^*, x) - J_n(f_n^*, x)| < 1/n \) (or \( |V_n(f_n,k_n^*, x) - V_n(f_n^*, x)| < 1/n \) for the average cost).

We have \( |\bar{J}_n(f_n,k_n^*, x) - J(f_n,k_n^*, x)| \to 0 \) and \( |\bar{V}_n(f_n,k_n^*, x) - V(f_n,k_n^*, x)| \to 0 \) as \( n \to \infty \), where again by an abuse of notation, the policies extended to \( X \) are also denoted by \( f_n,k_n^* \). Since \( J_n(f_n^*, x) = J_n(f_n,k_n^*, x) \) and \( V_n(f_n^*, x) = V_n(f_n,k_n^*, x) \), it follows that \( J_n(f_n,k_n^*, x) = J_n^*(x) \) and \( V_n(f_n,k_n^*, x) = V_n^*(x) \).

Therefore, before discretizing the state space to compute the near optimal policies, one can discretize, without loss of generality, the action space \( A \) in advance on a finite grid using sufficiently large number of grid points.

### 8.4 Bibliographic Notes

To be added.

### 8.5 Exercises

**Exercise 8.5.1** Consider the following problem: Let \( X = \{1, 2\} \), \( U = \{1, 2\} \), where \( X \) denotes whether a fading channel is in a good state (\( x = 2 \)) or a bad state (\( x = 1 \)). There exists an encoder who can either try to use the channel (\( u = 2 \)) or not use the channel (\( u = 1 \)). The goal of the encoder is send information across the channel.
Suppose that the encoder’s cost (to be minimized) is given by:

\[ c(x, u) = -1_{\{x=2, u=2\}} + \alpha u, \]

for \( \alpha = 1/2 \).

Suppose that the transition kernel is given by:

\[
\begin{align*}
P(x_{t+1} = 2|x_t = 2, u_t = 2) &= 0.8, & P(x_{t+1} = 1|x_t = 2, u_t = 2) &= 0.2 \\
P(x_{t+1} = 2|x_t = 2, u_t = 1) &= 0.2, & P(x_{t+1} = 1|x_t = 2, u_t = 1) &= 0.8 \\
P(x_{t+1} = 2|x_t = 1, u_t = 2) &= 0.5, & P(x_{t+1} = 1|x_t = 1, u_t = 2) &= 0.5 \\
P(x_{t+1} = 2|x_t = 1, u_t = 1) &= 0.9, & P(x_{t+1} = 1|x_t = 1, u_t = 1) &= 0.1
\end{align*}
\]

We will consider either a discounted cost for some \( \beta \in (0, 1) \)

\[
\inf \limits_{\Pi} E^\Pi \left[ \sum_{t=0}^{\infty} \beta^t c(x_t, u_t) \right] \tag{8.17}
\]

or an average cost

\[
\inf \lim \sup \limits_{T \to \infty} \frac{1}{T} E^\Pi \left[ \sum_{t=0}^{T-1} c(x_t, u_t) \right]. \tag{8.18}
\]

a) Using Matlab or some other program, obtain a solution to the problem given above in (8.17) through the following: (i) Q-Learning, (ii) Policy Iteration, or (iii) Value Iteration.

For Policy Iteration, see the discussion in Chapter 8, Section 8.1 of the Lecture Notes on the course web site. For Value Iteration, see Theorem 5.4.2. For Q-Learning, see Section 8.2.1 of the Lecture Notes.

b) Consider the example given in (8.18). Apply your algorithm in part b) to this problem by solving the corresponding linear program. What is the optimal policy? In Matlab, the command linprog can be used to solve linear programming problems.

**Exercise 8.5.2** Let \( c : X \times U \to \mathbb{R}_+ \) be bounded, where \( X \) is the state space and \( U \) is the action space for a controlled stochastic system. Suppose that under a stationary policy \( \gamma \), the expected discounted cost, for \( \beta < 1 \), is given by

\[
J_\beta(x, \gamma) := E^\gamma_x \left[ \sum_{k=0}^{\infty} \beta^k c(x_k, \gamma(x_k)) \right] = c(x, \gamma(x)) + \beta \int J_\beta(x_{t+1}, \gamma) P(dx_{t+1}|x_t = x, u_t = \gamma(x))
\]

Let \( f_1 \) and \( f_2 \) be two stationary policies. Define a third policy, \( g \), as:

\[
g(x) = f_1(x)1_{\{x \in C\}} + f_2(x)1_{\{x \in X \setminus C\}}
\]

where

\[
C = \{ x : J_\beta(x, f_1) \leq J_\beta(x, f_2) \}
\]

and \( X \setminus C \) denotes the complement of this set.

Show that \( J_\beta(x, g) \leq J_\beta(x, f_1) \) and \( J_\beta(x, g) \leq J_\beta(x, f_2) \) for all \( x \in X \).
Decentralized Stochastic Control

In the following, we primarily follow Chapters 2, 3, 4 and 12 of [139] for this topic. For a more complete coverage, the reader may follow [139].

We will consider a collection of decision makers (DMs) where each has access to some local information variable: Such a collection of decision makers who wish to minimize a common cost function and who has an agreement on the system (that is, the probability space on which the system is defined, and the policy and action spaces) is said to be a team. To study such team problems in a systematic fashion, we will obtain classifications of such teams.

9.1 Witsenhausen’s Intrinsic Model

In the following, we consider a decentralized stochastic control model with $N$ stations.

A decentralized control system may either be sequential or non-sequential. In a sequential system, the decision makers (DMs) act according to an order that is specified before the system starts running; while in a non-sequential system the DMs act in an order that depends on the realization of the system uncertainty and the control actions of other DMs.

According to the intrinsic model, any (finite horizon) sequential team problem can be characterized by a tuple \((\Omega, \mathcal{F}), N, \{(\mathcal{U}_i, \mathcal{J}_i), i = 1, \ldots, N\}\) or equivalently by a tuple \((\Omega, \mathcal{F}), N, \{\mathcal{U}_i, i = 1, \ldots, N\}, \{(\mathcal{I}_i, \mathcal{I}_i'), i = 1, \ldots, N\}\) where

- \((\Omega, \mathcal{F})\) is a measurable space representing all the uncertainty in the system. The realization of this uncertainty is called the \textit{primitive variable} of the system. \(\Omega\) denotes all possible realizations of the primitive random variable and \(\mathcal{F}\) is a sigma-algebra over \(\Omega\).
- \(N\) denotes the number of decision makers (DMs) in the system. Each DM takes only one action. If the system has a control station that takes multiple actions over time, it is modeled as a collection of DMs, one for each time instant.
- \\{(\mathcal{U}_i, \mathcal{J}_i), i = 1, \ldots, N\}\ is a collection of measurable spaces representing the action space for each DM. The control action \(\mathcal{u}_i\) of DM \(i\) takes value in \(\mathcal{U}_i\) and \(\mathcal{J}_i\) is a sigma-algebra over \(\mathcal{U}_i\).
- \{\mathcal{J}_i, i = 1, \ldots, N\}\ is a collection of sets in \(\mathcal{F}\) and represents the \textit{information} available to a DM to take an action. Sometimes it is useful to assume that the information is available in terms of an explicit observation that takes values in a measurable space \((\mathcal{I}_i, \mathcal{I}_i')\). Such an observation is generated by a measurable observation function from \(\Omega \times \mathcal{U}_1 \times \cdots \times \mathcal{U}_N\) to \(\mathcal{I}_i'.\) The collection \{\mathcal{J}_i, i = 1, \ldots, N\} or \{(\mathcal{I}_i, \mathcal{I}_i'), i = 1, \ldots, N\}\ is called the \textit{information structure} of the system.
- A \textit{control strategy} (also called a \textit{control policy} or \textit{design}) of a decentralized control system is given by a collection \{\gamma_i, i = 1, \ldots, N\} of functions where \(\gamma_i : (\mathcal{I}_i, \mathcal{I}_i') \rightarrow (\mathcal{U}_i, \mathcal{J}_i)\) (or equivalently, \(\gamma_i : (\Omega, \mathcal{J}_i) \rightarrow (\mathcal{U}_i, \mathcal{J}_i)\)). Let \(\Gamma\) denote the set of all such measurable policies.
Although, there are different ways to define a control objective of a decentralized system, we focus on minimizing a loss function. Other performance measures include minimizing regret, minimizing risk, ensuring safety, and ensuring stability. We will assume that we are given a probability measure \( P \) on \((\Omega, \mathcal{F})\) and a real-valued loss function \( \ell \) on \((\Omega \times \mathbb{R}^1 \times \cdots \times \mathbb{R}^N, \mathcal{F} \otimes \mathcal{U}_1 \otimes \cdots \otimes \mathcal{U}_N) =: (\mathbb{H}, \mathcal{H})\). Any choice \( \gamma = (\gamma^1, \ldots, \gamma^n) \) of the control strategy induces a probability measure \( P' \) on \((\mathbb{H}, \mathcal{H})\). We define the performance \( J(\gamma) \) of a strategy as the expected loss (under probability measure \( P' \)), i.e.,

\[
J(\gamma) = \mathbb{E}^\gamma[\ell(\omega, u^1, \ldots, u^n)]
\]

where \( \omega \) is the primitive variable (or the primitive random variable, since a measure is specified) and \( u^i \) is the control action of DM \( i \).

As an example, consider the following model of a system with two decision makers which is taken from [139]. Let \( \Omega = \{\omega_1, \omega_2, \omega_3\} \), \( \mathcal{F} \) be the power set of \( \Omega \). Let the action space be \( \mathbb{U}^1 = \{U(\text{up}), D(\text{down})\} \), \( \mathbb{U}^2 = \{L(\text{left}), R(\text{right})\} \), and \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \) be the power sets of \( \mathbb{U}^1 \) and \( \mathbb{U}^2 \) respectively. Let the information fields \( \mathcal{J}^1 = \{\emptyset, \{\omega_1\}, \{\omega_2, \omega_3\}, \Omega\} \) and \( \mathcal{J}^2 = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \Omega\} \). (This information corresponds to the non-identical imperfect (quantized) measurement setting considered in [139]).

Suppose the probability measure \( P \) is given by \( P(\omega_i) = p_i, i = 1, 2, 3 \) and \( p_1 = 0.3, p_2 = 0.4, \) and the loss function \( \ell(\omega, u^1, u^2) \) is given by

\[
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
u^1 & | & | \\
\hline
U & 1 & 0 \\
D & 3 & 1 \\
\end{array}
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
\omega_1 \leftrightarrow 0.3 & | & | \\
\omega_2 \leftrightarrow 0.3 & | & | \\
\omega_3 \leftrightarrow 0.4 & | & |
\end{array}
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
U & 1 & 2 \\
D & 2 & 1 \\
\end{array}
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
\omega_1 \leftrightarrow 0.3 & | & | \\
\omega_2 \leftrightarrow 0.3 & | & | \\
\omega_3 \leftrightarrow 0.4 & | & |
\end{array}
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
U & 1 & 2 \\
D & 2 & 1 \\
\end{array}
\begin{array}{c|cc}
   & u^2 & u^2 \\
\hline
\omega_1 \leftrightarrow 0.3 & | & | \\
\omega_2 \leftrightarrow 0.3 & | & | \\
\omega_3 \leftrightarrow 0.4 & | & |
\end{array}
\]

For the above model, the unique optimal control strategy is given by

\[
\gamma^1,y^1(\omega^1) = \begin{cases} U, & y^1 = \{\omega_1\} \\ D, & \text{else} \end{cases}
\]

\[
\gamma^2,y^2(\omega^2) = \begin{cases} R, & y^2 = \{\omega_1, \omega_2\} \\ L, & \text{else} \end{cases}
\]

The development of a systematic solution procedure to a generalized sequential decentralized stochastic control problem is a difficult task. Most of the work in the literature has concentrated on identifying solution techniques for specific subclasses. Typically, these subclasses are characterized on the basis of the information structure of the system.

### 9.2 Classification of information structures

An information structure is called static if the observation of all DMs depends only on the primitive random variable (and not on the control actions of others). Systems that don’t have static information structure are said to have dynamic information structure. In such systems, some DMs influence the observations of others through their actions.

Witsenhausen [127] showed that any dynamic decentralized control system can be converted to a static decentralized control system by an appropriate change of measures. However, very little is known regarding the solution of a non-LQG static system; hence, the above transformation is not practically useful.

#### 9.2.1 Classical, quasiclassical and nonclassical information structures

Centralized control systems are a special case of decentralized control systems; their characterizing feature is centralization of information, i.e., any DM knows the information available to all the DMs that acted before it, or formally, \( \mathcal{J}^i \subseteq \mathcal{J}^{i+1} \) for all \( i \). Such information structures are called classical.
A decentralized system is called quasiclassical or partially nested if the following condition holds: whenever DM $i$ can influence DM $j$ (that is the information at DM $j$ is dependent on the action of DM $i$), then DM $j$ must know the observations of DM $i$, or more formally, $\mathcal{J}^i \subseteq \mathcal{J}^j$. We will use the notation $\text{DM } i \rightarrow \text{DM } j$ to represent this relationship.

Any information structure that is not classical or quasiclassical is called nonclassical.

In a state space model, we assume that the decentralized control system has a state $x_t$ that is evolving with time. The evolution of the state is controlled by the actions of the control stations. We assume that the system has $N$ control stations where each control station $i$ chooses a control action $u^i_t$ at time $t$. The system runs in discrete time, either for finite or infinite horizon.

Let $\mathcal{X}$ denote the space of realizations of the state $x_t$, and $\mathcal{U}$ denote the space of realization of control actions $u^i_t$. Let $\mathcal{T}$ denote the set of time for which the system runs.

The initial state $x_1$ is a random variable and the state of the system evolves as

$$x_{t+1} = f_t(x_t, u^1_t, \ldots, u^N_t; w^0_t), \quad t \in \mathcal{T},$$

where $\{w^0_t, t \in \mathcal{T}\}$ is an independent noise process that is also independent of $x_1$.

We assume that each control station $i$ observes the following at time $t$

$$y^i_t = g^i_t(x_t, w^i_t),$$

where $\{w^i_t, t \in \mathcal{T}\}$ are measurement noise processes that are independent across time, independent of each other, and independent of $\{w^0_t, t \in \mathcal{T}\}$ and $x_1$.

The above evolution does not completely describe the dynamic control system, because we have not specified the data available at each control station. In general, the data $I^i_t$ available at control station $i$ at time $t$ will be a function of all the past system variables $\{x_{[1,t]}, y_{[1,t]}, u_{[1,t-1]}, w_{[1,t]}\}$, i.e.,

$$I^i_t = \eta^i_t(x_{[1,t]}, y_{[1,t]}, u_{[1,t-1]}, w_{[1,t]}),$$

where we use the notation $\mathbf{u} = \{u^1, \ldots, u^N\}$ and $x_{[1,t]} = \{x_1, \ldots, x_t\}$. The collection $\{I^i_t, i = 1, \ldots, N, t \in \mathcal{T}\}$ is called the information structure of the system.

When $\mathcal{T}$ is finite, say equal to $\{1, \ldots, T\}$, the above model is a special case of the sequential intrinsic model presented above. The set $\{x_1, w^0_1, w^1_1, \ldots, w^N_1, t \in \mathcal{T}\}$ denotes the primitive random variable with probability measure given by the product measure of the marginal probabilities; the system has $N \times T$ DMs, one for each control station at each time. DM $(i, t)$ observes $I^i_t$ and chooses $u^i_t$. The information sub-fields $\mathcal{J}^k$ are determined by $\{\eta^i_t, i = 1, \ldots, N, t \in \mathcal{T}\}$.

Some important information structures are

1. **Complete information sharing**: In complete information sharing, each DM has access to present and past measurements and past actions of all DMs. Such a system is equivalent to a centralized system.

   $$I^i_t = \{y_{[1,t]}, u_{[1,t-1]}\}, \quad t \in \mathcal{T}.$$

2. **Complete measurement sharing**: In complete measurement sharing, each DM has access to the present and past measurements of all DMs. Note that past control actions are not shared.

   $$I^i_t = \{y_{[1,t]}\}, \quad t \in \mathcal{T}.$$

3. **Delayed information sharing** In delayed information sharing, each DM has access to $n$-step delayed measurements and control actions of all DMs.
The objective is to choose control laws \( \gamma \). In all the information structures given above, each DM has perfect recall (PR), that is, each DM has full memory of its past information. In general, a DM need not have perfect recall. For example, a DM may only have access to its current observation, in which case the information structure is of an additive form:

\[
I^i_t = \begin{cases} 
  \{y^i_{[t-n+1,t]}, u^i_{[t-n+1,t-1]}y^i_{[t,n]}, u^i_{[t,n]}\}, & t > n \\
  \{y^i_{[1,t]}, u^i_{[1,t-1]}\}, & t \leq n
\end{cases}
\]  

(9.4)

4. **Delayed measurement sharing** In delayed measurement sharing, each DM has access to \( n \)-step delayed measurements of all DMs. Note that control actions are not shared.

\[
I^i_t = \begin{cases} 
  \{y^i_{[t-n+1,t]}, u^i_{[t-n+1,t-1]}y^i_{[t,n]}, u^i_{[t,n]}\}, & t > n \\
  \{y^i_{[1,t]}, u^i_{[1,t-1]}\}, & t \leq n
\end{cases}
\]

5. **Delayed control sharing** In delayed control sharing, each DM has access to \( n \)-step delayed control actions of all DMs. Note that measurements are not shared.

\[
I^i_t = \begin{cases} 
  \{y^i_{[1,t]}, u^i_{[t-n+1,t-1]}, u^i_{[t,n]}\}, & t > n \\
  \{y^i_{[1,t]}, u^i_{[1,t-1]}\}, & t \leq n
\end{cases}
\]

6. **Periodic information sharing** In periodic information sharing, the DMs share their measurements and control periodically after every \( k \) time steps. No information is shared at other time instants.

\[
I^i_t = \begin{cases} 
  \{y^i_{[t/k,k,t]}, u^i_{[t/k,k,t]}, y^i_{[1,t/k,k]}, u^i_{[1,t/k,k]}\}, & t \geq k \\
  \{y^i_{[1,t]}, u^i_{[1,t-1]}\}, & t < k
\end{cases}
\]

7. **Completely decentralized information** In a completely decentralized system, no data is shared between the DMs.

\[
I^i_t = \{y^i_{[1,t]}, u^i_{[1,t-1]}\}, \quad t \in \mathcal{T}.
\]

In all the information structures given above, each DM has perfect recall (PR), that is, each DM has full memory of its past information. In general, a DM need not have perfect recall. For example, a DM may only have access to its current observation, in which case the information structure is

\[
I^i_t = \{y^i_t\}, \quad t \in \mathcal{T}.
\]  

(9.5)

To complete the description of the team problem, we have to specify the loss function. We will assume that the loss function is of an additive form:

\[
\ell (x_{[1,T]}, u_{[1,T]}) = \sum_{t \in \mathcal{T}} c(x_t, u_t)
\]

(9.6)

where each term in the summation is known as the incremental (or stagewise) loss.

The objective is to choose control laws \( \gamma^i_t \) such that \( u^i_t = \gamma^i_t(I^i_t) \) so as to minimize the expected loss (9.6). In the sequel, we will denote the set of all measurable control laws \( \gamma^i_t \) under the given information structure by \( I^i_t \).

Let

\[
\gamma = \{\gamma^1, \cdots, \gamma^N\}
\]

and let a cost function be defined as:

\[
J(\gamma) = E[c(\omega_0, u)] = E[c(\omega_0, \gamma_1(y^1), \cdots, \gamma_N(y^N))],
\]

(9.7)

for some non-negative measurable loss (or cost) function \( c : \Omega \times \prod_k \mathbb{U}^k \to \mathbb{R}_+ \). Here, we have the notation \( u := \{u^i, t \in \mathcal{N}\} \). Here, \( \omega_0 \) may be viewed as the cost function relevant exogenous variable and is contained in \( \omega \).
9.3 Solutions to Static Teams

The material here is largely from [139].

**Definition 9.2.1** For a given stochastic team problem with a given information structure, \( \{J; \Gamma^i, i \in N\} \), a policy (strategy) \( N \)-tuple \( \gamma^* := (\gamma^1, \ldots, \gamma^N) \in \Gamma \) is an optimal team decision rule (team-optimal decision rule or simply team-optimal solution) if

\[
J(\gamma^*) = \inf_{\gamma \in \Gamma} J(\gamma) =: J^*, \tag{9.8}
\]

provided that such a strategy exists. The cost level achieved by this strategy, \( J^* \), is the minimum (or optimal) team cost.

**Definition 9.2.2** For a given \( N \)-person stochastic team with a fixed information structure, \( \{J; \Gamma^i, i \in N\} \), an \( N \)-tuple of strategies \( \gamma^* := (\gamma^1, \ldots, \gamma^N) \) constitutes a Nash equilibrium (synonymously, a person-by-person optimal (php optimal) solution) if, for all \( \beta \in \Gamma^1 \) and all \( i \in N \), the following inequalities hold:

\[
J^* := J(\gamma^*) \leq J(\gamma^{-i*}, \beta), \tag{9.9}
\]

where we have adopted the notation

\[
(\gamma^{-i*}, \beta) := (\gamma^1, \ldots, \gamma^{-i*}, \beta, \gamma^{i+1*}, \ldots, \gamma^N). \tag{9.10}
\]

For notational simplicity, let for any \( 1 \leq k \leq N \), \( \gamma^{-k} := \{\gamma^i, i \in \{1, \cdots, N\} \setminus \{k\}\} \)

In the following, we will denote by bold letters the ensemble of random variables across the DMs; that is \( y = \{y^i, i = 1, \cdots, N\} \) and \( u = \{u^i, i = 1, \cdots, N\} \).

**9.3 Solutions to Static Teams**

The following results are due to Krainak et. al. [70] and [139], generalizing Radner [97]. We follow the presentation in [139], which also contains the proofs of the results.

**Theorem 9.3.1** [70] Let \( \{J; \Gamma^i, i \in N\} \) be a static stochastic team problem where \( U^i = \mathbb{R}^m_i, i \in N \), the loss function \( c(\omega_0, u) \) is convex and continuously differentiable in \( u \) a.s., and \( J(\gamma) \) is bounded from below on \( \Gamma \). Let \( \gamma^* \in \Gamma \) be a policy \( N \)-tuple with a finite cost \( (J(\gamma^*) < \infty) \), and suppose that for every \( \gamma \in \Gamma \) such that \( J(\gamma) < \infty \), the following holds:

\[
\sum_{i \in N} E\{\nabla_{u^i} c(\omega_0; \gamma^*(y)) [\gamma^i(y^i) - \gamma^{i*}(y^i)]\} \geq 0, \tag{9.12}
\]

where \( E\{\cdot\} \) denotes the total expectation. Then, \( \gamma^* \) is a team-optimal policy, and it is unique if \( c \) is strictly convex in \( u \).

Note that the conditions of **Theorem 9.3.1** above do not include the stationarity of \( \gamma^* \), and furthermore inequalities (9.12) may not generally be easy to check, since they involve all permissible policies \( \gamma \) (with finite cost). Instead, either one of the following two conditions will achieve this objective [70] [139]:

(c.1) For all \( \gamma \in \Gamma \) such that \( J(\gamma) < \infty \), the following random variables are integrable

\[
\nabla_{u^i} c(\omega_0; \gamma^*(y)) [\gamma^i(y^i) - \gamma^{i*}(y^i)], \quad i \in N
\]
(c.2) \( \Gamma_i \) is a Hilbert space for each \( i \in N \), and \( J(\gamma) < \infty \) for all \( \gamma \in \Gamma \). Furthermore,
\[
E_{\omega \mid y_i} \{ \nabla_w c(\omega_0; \gamma^*_i(y)) \} \in \Gamma_i, \quad i \in N.
\]

**Theorem 9.3.2** [139] Let \( \{ J; \Gamma_i, i \in N \} \) be a static stochastic team problem which satisfies all the hypotheses of Theorem 9.3.1 with the exception of the inequality (9.12). Instead of (9.12), let either (c.1) or (c.2) be satisfied. Then, if \( \gamma^* \in \Gamma \) is a stationary policy it is also team optimal. Such a policy is unique if \( c(\omega_0; u) \) is strictly convex in \( u \), a.s.

What needs to be shown is that under stationarity, (c.1) or (c.2) implies Theorem 9.3.1; this follows once again from the law of the iterated expectations (Theorem 4.1.3); see [139]. If (c.1) holds, then for all \( i \in N \),
\[
\begin{align*}
E \left[ \nabla_w c(\omega_0; \gamma^*_i(y)) \right| & \gamma^*_i(y) - \gamma^*_i(y) \right] \\
& = E \left[ \left( \nabla_w c(\omega_0; \gamma^*_i(y)) \right| \gamma^*_i(y) - \gamma^*_i(y) \right] |y^i \right] \\
& = 0 \\
& \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad (9.13)
\end{align*}
\]
under stationarity and thus Theorem 9.3.2 holds.

To appreciate some of the fine points of Theorems 9.3.1 and 9.3.2, let us now consider the following example, which was discussed by Radner (1962) [97], and Krainak et al. (1982) [69].

**Example 9.1.** Let \( N = 2 \), \( U^1 = U^2 = \mathbb{R} \), \( \xi = x \) be a Gaussian random variable with zero mean and unit variance \( \sim N(0, 1) \), and the loss functional be given by
\[
L(x; u^1, u^2) = (u^1 - u^2)^2 e^{x^2} + 2u^1 u^2.
\]

Note that \( L \) is strictly convex and continuously differentiable in \( (u^1, u^2) \) for every value of \( x \). Hence, if the true value of \( x \) were known to both agents, the problem would admit a unique team optimal solution: \( u^1 = u^2 = 0 \), which is also stationary. Since this team-optimal solution does not use the precise value of \( x \), it is certainly optimal also under “no-measurement” information at the decision makers. Note, however, that in this case the only pairs that make \( J(\gamma) \) finite, are \( u^1 = u^2 = u \in \mathbb{R} \), since
\[
E[e^{x^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{x^2} dx = \infty.
\]

The set of permissible policies not being an open set, clearly we cannot talk about stationarity in this case. Theorem 9.3.1 (which does not involve stationarity) is applicable here. Note also that for every \( u \in \mathbb{R} \), \( u^1 = u^2 = u \) is anbsp;optimal solution, but only one of these is team optimal.

Now, as a more interesting case consider the measurement scheme:
\[
y^1 = x + w^1; \quad y^2 = x + w^2
\]
where \( w^1 \) and \( w^2 \) are independent random variables uniformly distributed on the interval \([-1, 1]\), which are also independent of \( x \). Note that here the random state of nature, \( \xi \), is chosen as \((x, w^1, w^2)\). Clearly, \( u^1 = u^2 = 0 \) is team-optimal for this case also, but it is not obvious at the outset whether it is stationary or not. Toward this end, let us evaluate (9.11) for \( i = 1 \) and with \( \gamma^2(y^2) = 0 \):
\[
(\partial/\partial u^1) E_{x, y^2|y^1} \{(u^1)^2 e^{x^2}\} = (\partial/\partial u^1)[(u^1)^2 E_{x|y^1} e^{x^2}] = 2u^1 E_{x|y^1} e^{x^2}
\]
where the last step follows because the conditional probability density of \( x \) given \( y^1 \) is nonzero only in a finite interval (thus making the conditional expectation finite). By symmetry, it follows that both derivatives in (9.11) vanish at \( u^1 = u^2 = 0 \), and hence the team-optimal solution is stationary. It is not difficult to see that in fact this is the only pair of stationary policies. Note that all the hypotheses of Theorem 9.3.2 are satisfied here, under condition (c.5).
Quadratic-Gaussian teams

Given a probability space \((\Omega, \mathcal{F}, P_\Omega)\), and an associated vector-valued random variable \(\xi\), let \(\{J; T^i, i \in N\}\) be a static stochastic team problem with the following specifications [139]:

(i) \(U^i = \mathbb{R}^{m_i}, i \in N\); i.e., the action spaces are unconstrained Euclidean spaces.

(ii) The loss function is a quadratic function of \(u\) for every \(\xi\):

\[
L(\xi; u) = \sum_{i,j \in N} u^i R_{ij}(\xi) u^j + 2 \sum_{i \in N} u^i r_i(\xi) + c(\xi)
\]

(9.14)

where \(R_{ij}(\xi)\) is a matrix-valued random variable (with \(R_{ij} = R^i_{ij}\)), \(r_i(\xi)\) is a vector-valued random variable, and \(c(\xi)\) is a random variable, all generated by measurable mappings on the random state of nature, \(\xi\).

(iii) \(L(\xi; u)\) is strictly (and uniformly) convex in \(u\) a.s., i.e., there exists a positive scalar \(\alpha\) such that, with \(R(\xi)\) defined as a matrix comprised of \(N\) blocks, with the \(ij\)'th block given by \(R_{ij}(\xi)\), the matrix \(R(\xi) - \alpha I\) is positive definite a.s., where \(I\) is the appropriate dimensional identity matrix.

(iv) \(R(\xi)\) is uniformly bounded above, i.e., there exists a positive scalar \(\beta\) such that the matrix \(\beta I - R(\xi)\) is positive definite a.s.

(v) \(Y^i = \mathbb{R}^{r_i}, i \in N\), i.e., the measurement spaces are Euclidean spaces.

(vi) \(y^i = \eta^i(\xi), i \in N\), for some appropriate Borel measurable functions \(\eta^i, i \in N\).

(vii) \(I^i\) is the (Hilbert) space of all Borel measurable mappings of \(\gamma^i : \mathbb{R}^{r_i} \rightarrow \mathbb{R}^{m_i}\), which have bounded second moments, i.e., \(E_{\nu^i}\{\gamma^i(y^i)\gamma^i(y^j)\} < \infty\).

(viii) \(E_{\nu}(r_i^j(\xi)r_i(\xi)) < \infty, i \in N; \ E_\xi[c(\xi)] < \infty\).

**Definition 9.3.2** A static stochastic team is quadratic if it satisfies (i)-(viii) above. It is a standard quadratic team if furthermore the matrix \(R\) is constant for all \(\xi\) (i.e., it is deterministic). If, in addition, \(\xi\) is a Gaussian distributed random vector, and \(r_i(\xi) = Q_i \xi, \eta_i(\xi) = H^i \xi, i \in N\), for some deterministic matrices \(Q_i, H^i, i \in N\), the decision problem is a quadratic-Gaussian team (more widely known as a linear-quadratic-Gaussian (LQG) team under some further structure on \(Q_i\) and \(H^i\)).

One class of quadratic teams for which the team-optimal solution can be obtained in closed form are those where the random state of nature \(\xi\) is a Gaussian random vector. Let us decompose \(\xi\) into \(N + 1\) block vectors

\[
\xi = (x', y^1', y^2', \ldots, y^N')
\]

(9.15)
of dimensions \(r_0, r_1, r_2, \ldots, r_N\), respectively. Being a Gaussian random vector, \(\xi\) is completely described in terms of its mean value and covariance matrix, which we specify below:

\[
E[\xi] =: \bar{\xi} = (\bar{x}', \bar{y}^1', \ldots, \bar{y}^N')
\]

(9.16)

\[
\text{cov} (\xi) =: \Sigma, \quad \text{with } [\Sigma]_{ij} =: \Sigma_{ij}, \quad i, j = 0, 1, \ldots, N
\]

(9.17)

\([\Sigma]_{ij}\) denotes the \(ij\)'th block of the matrix \(\Sigma\) of dimension \(r_i \times r_j\), which stands for the cross-variance between the \(i\)'th and \(j\)'th block components of \(\xi\). We further assume (in addition to the natural condition \(\Sigma \geq 0\)) that \(\Sigma_{ii} > 0\) for \(i \in N\), which means that the measurement vectors \(y^i\)'s have nonsingular distributions. To complete the description of the quadratic-Gaussian team, we finally take the linear terms \(r_i(\xi)\) in the loss function (9.14) to be linear in \(x\), which makes \(x\) the “payoff relevant” part of the state of nature:

\[
r_i(\xi) = D_i x, \quad i \in N
\]

(9.18)

where \(D_i\) is an \((r_i \times r_0)\) dimensional constant matrix.
In the characterization of the team-optimal solution for the quadratic-Gaussian team we will need the following important result on the conditional distributions of Gaussian random vectors.

**Lemma 9.3.1** Let \( z \) and \( y \) be jointly Gaussian distributed random vectors with mean values \( \bar{z}, \bar{y} \), and covariance

\[
\text{cov}(z, y) = \begin{pmatrix} \Sigma_{zz} & \Sigma_{zy} \\ \Sigma_{zy} & \Sigma_{yy} \end{pmatrix} \succeq 0, \quad \Sigma_{yy} > 0.
\]

Then, the conditional distribution of \( z \) given \( y \) is Gaussian, with mean

\[
E[z|y] = \bar{z} + \Sigma_{zy} \Sigma_{yy}^{-1}(y - \bar{y})
\]

and covariance

\[
\text{cov}(z|y) = \Sigma_{zz} - \Sigma_{zy} \Sigma_{yy}^{-1} \Sigma_{zy}^T
\]

The complete solution to the quadratic-Gaussian team is now given in the following Theorem.

**Theorem 9.3.3** The quadratic-Gaussian team decision problem as formulated above admits a unique team-optimal solution, that is affine in the measurement of each agent:

\[
\gamma_i^*(y_i) = \Pi_i(y_i^t - \bar{y}^i) + M_i \bar{x}, \quad i \in N.
\]

Here, \( \Pi_i \) is an \((m_i \times r_i)\) matrix \((i \in N)\), uniquely solving the set of linear matrix equations:

\[
R_{ii} \Pi_i \Sigma_{ii} + \sum_{j \in N, j \neq i} R_{ij} \Pi_j \Sigma_{ji} + D_i \Sigma_{0i} = 0,
\]

and \( M_i \) is an \((m_i \times r_0)\) matrix for each \( i \in N \), obtained as the unique solution of

\[
\sum_{j \in N} R_{ij} M_j + D_i = 0, \quad i \in N.
\]

**Remark 9.2.** The proof of this result follows immediately from Theorem 9.3.1. However, a Projection Theorem based concise proof can also be provided exploiting the quadratic nature of the problem [50]. If one can show orthogonality of the estimation error to the space of linear policies, the Gaussianity of the random variables implies orthogonality to the subspace of measurable functions with finite second moment, leading to the desired conclusion.

An important application of the above result is the following static Linear Quadratic Gaussian Problem: Consider a two-controller system evolving in \( \mathbb{R}^n \) with the following description: Let \( x_1 \) be Gaussian and \( x_2 = Ax_1 + B_1 u_1 + B_2 u_2 + w \)

\[
y_1^1 = C_1 x_1 + v_1, \\
y_2^2 = C_2 x_1 + v_2,
\]

with \( w, v^1, v^2 \) zero-mean, i.i.d. disturbances. For \( \rho_1, \rho_2 > 0 \), let the goal be the minimization of

\[
J(\gamma^1, \gamma^2) = \mathcal{E} \left[ ||x_1||^2_2 + \rho_1 ||u_1||^2_2 + \rho_2 ||u_2||^2_2 + ||x_2||^2_2 \right]
\]

over the control policies of the form:

\[
u_i^t = \mu_i^t(y_i^t), \quad i = 1, 2
\]

For such a setting, optimal policies are linear.
9.4 Dynamic Teams Quasi-Classical Information Patterns

Under quasi-classical information, LQG stochastic team problems are tractable by conversion into equivalent static team problems: Consider the following dynamic team with \( N \) agents, where each agent acts only once, with \( A_k, k \in \mathcal{N} \), having the following measurement

\[
y^k = C^k \xi + \sum_{i \rightarrow k} D_{ik} u^i, \quad (9.26)
\]

where \( \xi \) is an exogenous random variable picked by nature, and \( i \rightarrow k \) denotes the precedence relation that the action of \( A_i \) affects the information of \( A_k \) and \( u^i \) is the action of \( A_i \).

If the information structure is quasi-classical, then

\[
\mathcal{I}^k = \{ y^k, \{ \mathcal{I}^i, i \rightarrow k \} \}.
\]

That is, \( A_k \) has access to the information available to all the signaling agents. Such an IS is equivalent to the IS \( \mathcal{I}^k = \{ \tilde{y}^k \} \), where \( \tilde{y}^k \) is a static measurement given by

\[
\tilde{y}^k = \left\{ C^k \xi, \{ C^i \xi, i \rightarrow k \} \right\}. \quad (9.27)
\]

Such a conversion can be done provided that the policies adopted by the agents are deterministic, with the equivalence to be interpreted in the sense that any deterministic policy measurable under the original IS being measurable also under the new (static) IS and vice versa, since the actions are determined by the measurements. The restriction of using only deterministic policies is, however, without any loss of optimality: with policies of all other agents fixed (possibly randomized) no agent can benefit from randomized decisions in such team problems. We discussed this property of irrelevance of random information/actions in optimal stochastic control in Chapter 5 in view of Blackwell’s Irrelevant Information Theorem.

This observation, made by Ho and Chu [67], leads to the following result.

**Theorem 9.4.1** Consider an LQG system with a partially nested information structure. For such a system, optimal solutions are affine (that is, linear plus a constant).

**Remark 9.3.** Another class of dynamic team problems that can be converted into solvable dynamic optimization problems are those where even though the information structure is nonclassical, there is no incentive for signaling because any signaling from say agent \( A_i \) to agent \( A_j \) conveys information to the latter which is “cost irrelevant”, that is it does not lead to any improvement in performance [136] [139].

9.4.1 Non-classical information structures, signaling and its effect on lack of convexity

What makes a large number of problems possessing the nonclassical information structure difficult is the fact that signaling is present: Signaling is the policy of communication through control actions. Under signaling, the decision makers apply their actions to affect the information available at the other decision makers. In this case, the control policies induce a probabilistic map (hence, a channel or a stochastic kernel) from the exogenous random variable space to the observation space of the signaled decision makers. For the nonclassical case, the problem thus also features an information transmission aspect, and the signaling decision maker’s objective also includes the design of an optimal measurement channel.

Consider the following example [136]. Consider a two-controller system evolving in \( \mathbb{R}^n \):

\[
x_{t+1} = Ax_t + B^1 u^1_t + B^2 u^2_t + w_t,
\]

\[
y^1_t = C^1 x_t + v^1_t,
\]

\[
y^2_t = C^2 x_t + v^2_t,
\]
where \( w, v^1, v^2 \) are zero-mean, i.i.d. disturbances, and \( A, B^1, B^2, C^1, C^2 \) matrices of appropriate dimensions. For \( \rho_1, \rho_2 > 0 \), let the objective be the minimization of the cost functional be

\[
J = \mathcal{E} \left[ \left( \sum_{t=1}^{T} |x_t|^2 + \rho_1 |u^1_t|^2 + \rho_2 |u^2_t|^2 \right) + \|x_T\|^2 \right]
\]

over control policies of the form:

\[
u^i_t = \mu^i_t(y^i_0, t, u^i_0, t-1), \quad i = 1, 2; \quad t = 0, 1, \ldots, T - 1.
\]

For a multi-stage problem (say with \( T = 2 \)), unlike \( T = 1 \) in (9.25), the cost is in general no-longer convex in the action variables of the controllers acting in the first stage \( t = 0 \). This is because these actions might affect the estimation quality of the other controller in the future stages, if one DM can signal information to the other DM in one stage. We note that this condition is equivalent to \( C^1 A^l B^2 \neq 0 \) or \( C^2 A^l B^1 \neq 0 \) with \( l + 1 \) denoting the delay in signaling with \( l = 0 \) in the problem considered.

### 9.4.2 Witsenhausen’s Counterexample and its variations

The celebrated Witsenhausen’s counterexample \([128]\) is a dynamic non-classical team problem. Suppose \( x \) and \( w_1 \) are two independent, zero-mean Gaussian random variables with variance \( \sigma^2 \) and 1 so that

\[
y^0 = x, \quad y^1 = u_0 + w_1
\]

\[
u_0 = \gamma_0(x), \quad u_1 = \gamma_1(y^1).
\]

with the performance criterion:

\[
Q_W(x, u_0, u_1) = k(u_0 - x)^2 + (u_1 - u_0)^2, \quad (9.28)
\]

This problem is described by a linear system; all primitive variables are Gaussian and the performance criterion is quadratic, yet linear policies are not optimal. We note that this is a non-convex problem \([141]\) and thus variational methods do not necessarily lead to optimality. In fact, we don’t even have a good lower bound on the optimal cost for Witsenhausen’s counterexample (see \([108]\) for a detailed discussion).

### 9.5 Static Reduction

Following Witsenhausen \([129]\), we say that two information structures are equivalent if: (i) The policy spaces are equivalent/isomorphic in the loose sense that policies under one information structure are realizable under the other information structure, (ii) the costs achieved under equivalent policies are identical almost surely, and (iii) if there are constraints in the admissible policies, the isomorphism among the policy spaces preserves the constraint conditions.

A large class of sequential team problems admit an equivalent information structure which is static. This is called the static reduction of an information structure.

---

**Fig. 9.1: Flow of information in Witsenhausen’s counterexample.**
9.5 Static Reduction

Partially nested case

An important information structure which is not nonclassical, is of the quasi-classical type, also known as partially nested; an IS is partially nested if an agent’s information at a particular stage $t$ can depend on the action of some other agent at some stage $t' \leq t$ only if she also has access to the information of that agent at stage $t'$. For such team problems with partially nested information, one talks about precedence relationships among agents: an agent DM $i$ is precedent to another agent DM $j$ (or DM $i$ communicates to DM $j$), if the former agent’s actions affect the information of the latter, in which case (to be partially nested) DM $j$ has to have the information based on which the action-generating policy of DM $i$ was constructed.

For partially nested (or quasi-classical) information structures, static reduction has been studied by Ho and Chu in the specific context of LQG systems [67] and for a class of non-linear systems satisfying restrictive invertibility properties [68].

Nonclassical case: Witsenhausen’s equivalent model and static reduction of sequential dynamic teams

Witsenhausen shows that a large class of sequential team problems admit an equivalent information structure which is static. This is called the static reduction of an information structure.

An equivalence between sequential dynamics teams and their static reduction is as follows (termed as the equivalent model [129]).

Consider a dynamic team setting according to the intrinsic model where there are $N$ time stages, and each DM observes, for some $t = 1, 2, \ldots, N$,

$$y^t = \eta_t(\omega, u^1, u^2, \ldots, u^{t-1}),$$

and the decisions are generated by $u^t = \gamma_t(y^t)$. Here, as before, $\omega$ is the collection of primitive (exogenous) variables. The resulting cost under a given team policy is, as in (9.7)

$$J(\gamma) = E[c(\omega_0, u)].$$

This dynamic team can be converted to a static team provided that the following absolute continuity condition holds: For every $t \in \mathcal{N}$, there exists a function $f_t$ such that for all Borel $S \subset \mathbb{Y}^t$:

$$P(y^t \in S|\omega, u^1, \ldots, u^{t-1}) = \int_S f_t(\omega, u^1, u^2, \ldots, u^{t-1}, y^t)Q_t(dy^t),$$

(9.29)

where $Q_t$ is some arbitrary reference probability measure on $y^t$. Under any fixed team policy, we can then write

$$P(d\omega, dy) = P(d\omega) \prod_{t=1}^{N} f_t(\omega, u^1, u^2, \ldots, u^{t-1}, y^t)Q_t(dy^t).$$

The cost function $J(\gamma)$ can then be written as

$$J(\gamma) = \int P(d\omega) \prod_{t=1}^{N} (f_t(y_t, \omega, u^1, u^2, \ldots, u^{t-1}, y^t)Q_t(dy^t))c(\omega_0, u),$$

(9.30)

where now the measurement variables can be regarded as independent and by incorporating the $\{f_t\}$ terms into $c$, we can obtain an equivalent static team problem. Hence, the essential step is to appropriately adjust the probability space and the cost function. The new cost function may now explicitly depend on the measurement values, such that

$$c_s(\omega, y, u) = c(\omega_0, u) \prod_{t=1}^{N} (f_t(y_t, \omega, u^1, u^2, \ldots, u^{t-1}, y^t).$$
In this case, we can view $\omega, y$ as the cost-relevant exogenous variable: By an abuse of notation, we will use the same notation $\omega_0$ to denote $\omega, y$ when it is clear that such a cost function comes from a static reduction.

We note that, as Witsenhausen notes in [129], a static reduction always holds when the measurement variables take values from countable set since a reference measure as in $Q_t$ above can be constructed on the measurement variable $y_t$ (e.g., $Q_t(y^t) = \sum_{i \geq 1} 2^{-i} \mathbb{1}_{\{y^t = m_i\}}$ where $\mathcal{Y}^t = \{m_i, i \in \mathbb{N}\}$) so that the absolute continuity condition always holds. On the other hand, for continuous spaces, observe that under a control-sharing pattern with $y^2 = u^1$, the absolute continuity condition required for Witsenhausen’s static reduction may fail: $P(y^2 \in A|u^1) = 1\{u^1 \in A\}$, leading to a delta function supported at $u^1$ and if the reference measure $\mu$ with $y^2 \sim \mu$ admits a density, the absolute continuity condition will not hold. We also note that a continuous-time generalization for static reduction similar to Girsanov’s method has been presented in [33].

9.5.1 Expansion of information Structures: A recipe for identifying sufficient information

We start with a general result on optimum-performance equivalence of two stochastic dynamic teams with different information structures. This is in fact a result which has a very simple proof, but it is quite effective as we will see shortly.

**Proposition 9.5.1** Let $D_1$ and $D_2$ be two stochastic dynamic teams with the same loss function, and differing only in their information structures, $\eta_1$ and $\eta_2$, respectively, with corresponding composite strategy spaces $\Gamma_1$ and $\Gamma_2$, such that $\Gamma_2 \subseteq \Gamma_1$. Let $D_1$ admit a team-optimal solution, denoted by $\gamma^*_1 \in \Gamma_1$, with the further property that $\gamma^*_1 \in \Gamma_2$. Then $\gamma^*_1$ also solves $D_2$.

A recipe for utilizing the result above would be [139]:

Given a team problem, say $D_2$, with IS $\eta_2$, which is presumably difficult to solve, obtain a finer IS $\eta_1$, and solve the team problem under this expanded IS (assuming that this new team problem is easier to solve). Then, if the team-optimal solution here is adapted to the sigma-field generated by the original coarser IS, it solves also the original problem $D_2$.

9.6 Existence of Optimal Solutions

9.7 Convexity Properties and Optimality of Deterministic Policies

9.8 Asymptotic Optimality of Finite Models and Quantized Policies

9.9 Common Knowledge as Information State and the Dynamic Programming Approach to Team Decision Problems

In a team problem, if all the random information at any given decision maker is common knowledge between all decision makers, then the system is essentially centralized. If only some of the system variables are common knowledge, the remaining unknowns may or may not lead to a computationally tractable program generating an optimal solution. A possible approach toward establishing a tractable program is through the construction of a controlled Markov chain where the controlled Markov state may now live in a larger state space (for example a space of probability measures) and the actions are elements in possibly function spaces. This controlled Markov construction may lead to a computation of optimal policies.

Such a dynamic programming approach has been adopted extensively in the literature (see for example, [9], [134], [35], [3], [136], [87] and generalized in [88]) through the use of a team-policy which uses common information to generate partial functions for each DM to generate their actions using local information. Thus, in the dynamic programming approach,
a separation of team decision policies in the form of a two-tier architecture, a higher-level controller and a lower-level controller, can be established with the use of common knowledge.

In the following, we present the ingredients of such an approach, as formalized by Nayyar, Mahajan and Teneketzis [88] and termed the common information approach:

1. Elimination of irrelevant information at the DMs: In this step, irrelevant local information at the DMs, say DM $k$, is identified as follows. By letting the policy at other DMs to be arbitrary, the policy of DM $k$ can be optimized as a best-response function, and irrelevant data at DM $k$ can be removed.

2. Construction of a coordinated system: This step identifies the common information and local/private information at the DMs, after Step 1 above has been carried out. A fictitious coordinator (higher-level controller) uses the common information to generate team policies, which in turn dictates the (lower-level) DMs what to do with their local information.

3. Formulation of the cost function as a Partially Observed Markov Decision Process (POMDP), in view of the coordinator’s optimal control problem: A fundamental result in stochastic control is that the problem of optimal control of a partially observed Markov chain (with additive per-stage costs) can be solved by turning the problem into a fully observed one on a larger state space where the state is replaced by the “belief” on the state.

4. Solution of the POMDP leads to the structural results for the coordinator to generate optimal team policies, which in turn dictates the DMs what actions to take given their local information realizations.

5. Establishment of the equivalence between the solution obtained and the original problem, and translation of the optimal policies. Any coordination strategy can be realized in the original system. Note that, even though there is no real coordinator, such a coordination can be realized implicitly, due to the presence of common information.

We will provide a further explicit setting with such a recipe at work, in the context of the $k$-stage periodic belief sharing pattern in the next section. In particular, Lemma 9.10.1 and Lemma 9.10.2 will highlight this approach. When a given information structure does not allow for the construction of a controlled Markov chain even in a larger, but fixed for all time stages, state space, one question that can be raised is what information requirements would lead to such a structure. We will also investigate this problem in the context of the one-stage belief sharing pattern in the next section.

### 9.10 $k$-Stage Periodic Belief Sharing Pattern

In this section, we will use the term belief for a probability measure-valued random variable. This terminology has been used particularly in the artificial intelligence and computer science communities, which we adopt here. We will, however, make precise what we mean by such a belief process in the following.

#### 9.10.1 $k$-stage periodic belief sharing pattern

As mentioned earlier in Chapter 6, a fundamental result in stochastic control is that the problem of optimal control of a partially observed Markov chain can be solved by turning the problem into a fully observed one on a larger state space where the state is replaced by the belief on the state. Such an approach is very effective in the centralized setting; in a decentralized setting, however, the notion of a state requires further specification. In the following, we illustrate this approach under the $k$-step periodic belief sharing information pattern.

Consider a joint process $\{x_t, y_t, t \in \mathbb{Z}_+\}$, where we assume for simplicity that the spaces where $x_t, y_t$ take values from are finite dimensional real-valued or countable. They are generated by

$$x_{t+1} = f(x_t, u^1_t, \ldots, u^L_t, w_t),$$

$$y^i_t = g(x_t, v^i_t),$$
where \( x_t \) is the state, \( u^i_t \in \mathbb{U} \) is the control action, \((u^i_t, v^i_j, 1 \leq i \leq L)\) are second order, zero-mean, mutually independent, i.i.d. noise processes. We also assume that the state noise, \( u^i_t \), either has a probability mass function, or a probability measure with a density function. To minimize the notational clutter, \( P(x) \) will denote the probability mass function for discrete-valued spaces or probability density function for continuous spaces.

Suppose that there is a common information vector \( I^i_t \) at some time \( t \), which is available to all the decision makers. At times \( ks - 1 \), with \( k > 0 \) fixed, and \( s \in \mathbb{Z}^+ \), the decision makers share all their information: \( I^i_{ks-1} = \{y^{0,ks-1}_{s}, u^{0,ks-1}_{s}\} \) and for \( I^i_{0} = \{P(x_{0})\} \), that is at time 0 the DMs have the same \textit{a priori} belief on the initial state. Hence, at time \( t \), DM \( i \) has access to \( \{y^{i}_{ks,t}, I^i_{ks-1}\} \).

Until the next common observation instant \( t = k(s + 1) - 1 \) we can regard the individual decision functions specific to DM \( i \) as \( \{u^i_t = \gamma^i_s(y^{i}_{ks,t}, I^i_{ks-1})\} \); we let \( \gamma_s \) denote the ensemble of such decision functions and let \( \gamma \) denote the team policy.

It then suffices to generate \( \gamma_s \) for all \( s \geq 0 \), as the decision outputs conditioned on \( y^{i}_{ks,t} \), under \( \gamma^i_s(y^{i}_{ks,t}, I^i_{ks-1}) \), can be generated. In such a case, we can define \( \gamma_s(., I^i_{ks-1}) \) to be the joint team decision rule mapping \( I^i_{ks-1} \) into a space of action vectors: \( \{\gamma^i_s(y^{i}_{ks,t}, I^i_{ks-1})\}, i \in L, t \in \{ks, ks + 1, \ldots, k(s + 1) - 1\}\).

Let \([0, T - 1]\) be the decision horizon, where \( T \) is divisible by \( k \). Let the objective of the decision makers be the joint minimization of

\[
E^\gamma_0 \sum_{t=0}^{T-1} c(x_t, u^1_t, u^2_t, \ldots, u^L_t),
\]

over all policies \( \gamma^1, \gamma^2, \ldots, \gamma^L \), with the initial condition \( x_0 \) specified. The cost function

\[
J_{x_0}(\gamma) = E^\gamma_0 \sum_{t=0}^{T-1} c(x_t, u_t)
\]

can be expressed as:

\[
J_{x_0}(\gamma) = E^\gamma_0 \sum_{s=0}^{k-1} \bar{c}(\gamma_s(., I^i_{ks-1}), \bar{x}_s))
\]

with

\[
\bar{c}(\gamma_s(., I^i_{ks-1}), \bar{x}_s) = E^\gamma_0 \{ \sum_{t=ks}^{k(s+1)-1} c(x_t, u_t) \}
\]

**Lemma 9.10.1** Consider the decentralized system setup above. Let \( I^i_t \) be a common information vector supplied to the DMs regularly every \( k \) time stages, so that the DMs have common memory with a control policy generated as described above. Then, \( \{\bar{x}_s := x_{ks}, \gamma_s(., I^i_{ks-1}), s \geq 0\} \) forms a controlled Markov chain.

In view of the above, we have the following separation result.

**Lemma 9.10.2** Let \( I^i_t \) be a common information vector supplied to the DMs regularly every \( k \) time steps. There is no loss in performance if \( I^i_{ks-1} \) is replaced by \( P(\bar{x}_s|I^i_{ks-1}) \).

An essential issue for a tractable solution is to ensure a common information vector which will act as a sufficient statistic for future control policies. This can be done via sharing information at every stage, or some structure possibly requiring larger but finite delay.

The above motivates us to introduce the following pattern.

**Definition 9.10.1** \( k \)-stage periodic belief sharing pattern An information pattern in which the decision makers share their posterior beliefs to reach a joint belief about the system state is called a belief sharing information pattern. If the belief sharing occurs periodically every \( k \)-stages \( (k > 1) \), the DMs also share the control actions they applied in the
last $k-1$ stages, together with intermediate belief information. In this case, the information pattern is called the $k$-stage periodic belief sharing information pattern.

Remark 9.4. For $k > 1$, it should be noted that, the exchange of the control actions is essential.

9.11 Exercises

Exercise 9.11.1 Consider the following team decision problem with dynamics:

$$x_{t+1} = ax_t + b_1 u_{t1}^1 + b_2 u_{t2}^2 + w_t,$$

$$y_{t1}^1 = x_t + v_{t1}^1,$$

$$y_{t2}^2 = x_t + v_{t2}^2.$$

Here $x_0, v_{t1}^1, v_{t2}^2, w_t$ are mutually and temporally independent zero-mean Gaussian random variables.

Let $\{\gamma_i\}$ be the policies of the controllers so that $u_i^t = \gamma_i^t(y_{i0}^i, y_{i1}^i, \ldots, y_{it}^i)$ for $i = 1, 2$.

Consider:

$$\min_{\gamma_1, \gamma_2} E_{x_0} \gamma_1, \gamma_2 \left[ \left( \sum_{t=0}^{T-1} x_t^2 + \rho_1 (u_{t1}^1)^2 + \rho_2 (u_{t2}^2)^2 \right) + x_T^2 \right],$$

where $\rho_1, \rho_2 > 0$.

Explain if the following are correct or not:

a) For $T = 1$, the problem is a static team problem.

b) For $T = 1$, optimal policies are linear.

c) For $T = 1$, linear policies may be person-by-person-optimal. That is, if $\gamma_1$ is assumed to be linear, then $\gamma_2$ is linear; and if $\gamma_2$ is assumed to be linear then $\gamma_1$ is linear.

d) For $T = 2$, optimal policies are linear.

e) For $T = 2$, linear policies may be person-by-person-optimal.

Exercise 9.11.2 Consider a common probability space on which the information available to two decision makers DM$^1$ and DM$^2$ are defined, such that $I_1$ is available at DM$^1$ and $I_2$ is available at DM$^2$.

R. J. Aumann [11] defines that an information $E$ is common knowledge between two decision makers DM$^1$ and DM$^2$, if whenever $E$ happens, DM$^1$ knows $E$, DM$^2$ knows $E$, DM$^1$ knows that DM$^2$ knows $E$, DM$^2$ knows that DM$^1$ knows $E$, and so on.

Suppose that one claims that an event $E$ is common knowledge if and only if $E \in \sigma(I_1) \cap \sigma(I_2)$, where $\sigma(I_1)$ denotes the $\sigma$-field generated by information $I_1$ and likewise for $\sigma(I_2)$.

Is this argument correct? Provide an answer with precise arguments. You may wish to consult [11], [90], [31] and Chapter 12 of [139].

Exercise 9.11.3 Consider the following static team decision problem with dynamics:

$$x_1 = ax_0 + b_1 u_{01}^1 + b_2 u_{02}^2 + w_0,$$

$$y_{01}^1 = x_0 + v_{01}^1,$$
\[ y_0^2 = x_0 + v_0^2, \]

Here \( v_0, v_2, w_2 \) are independent, Gaussian, zero-mean with finite variances.

Let \( \gamma^i : \mathbb{R} \to \mathbb{R} \) be policies of the controllers: \( u_0^1 = \gamma_0^1(y_0^1), u_0^2 = \gamma_0^2(y_0^2) \).

Find
\[
\min_{\gamma^1, \gamma^2} E_{v_0} y^1, y^2 [x_1^2 + \rho_1(u_1^1)^2 + \rho_2(u_2^2)^2],
\]
where \( v_0 \) is a zero-mean Gaussian distribution and \( \rho_1, \rho_2 > 0 \).

Find an optimal team policy \( \gamma = (\gamma^1, \gamma^2) \).

**Exercise 9.11.4** Let \( X \) be a binary random variable. Suppose two decision makers DM 1 and DM 2 have access to some local random variables \( Y^1 \) and \( Y^2 \), respectively, defined on a common probability space and correlated with \( X \), and exchange their conditional expectations over time. Suppose further that:

- the information \( \sigma \)-fields at each decision maker is increasing: \( F^i_t \subset F^i_{t+1} \), \( i = 1, 2 \), \( t \in \mathbb{Z}_+ \).
- for all \( n \in \mathbb{N} \), there exists \( m > n \) such that \( F^i_m \) contains information on \( E[X|F^j_n] \), \( i, j = 1, 2 \). That is, the decision makers exchange their estimates (but not their raw data - \( Y^i \) is private to DM \( i \), \( i = 1, 2 \)) infinitely often.

State and rigorously justify your answers for the following:

a) [10 Points] Is there a limit for \( \lim_{n \to \infty} E[X|F^i_n] \), \( j = 1, 2 \)? Either argue that the limit exists, or provide a counterexample.

b) [10 Points] For the cases where the limit exists, is it the case that
\[
\lim_{n \to \infty} E[X|F^i_n] = \lim_{n \to \infty} E[X|F^j_n]
\]
Either prove the result, or provide a counterexample.

Hint: See (28) (see also (51) and (116))

**Exercise 9.11.5** Consider a linear Gaussian system with mutually independent and i.i.d. noises:
\[
x_{t+1} = A x_t + \sum_{j=1}^{L} B^j u_t^j + w_t,
\]
\[
y_t^i = C^i x_t + v_t^i, \quad 1 \leq i \leq L,
\]
with the one-step delayed observation sharing pattern.

Construct a controlled Markov chain for the team decision problem: First show that one could have
\[
\{y_t^1, y_t^2, \ldots, y_t^L, P(dx_t|y_{0,t-1}^1, y_{0,t-1}^2, \ldots, y_{0,t-1}^L)\}
\]
as the state of the controlled Markov chain.

Consider the following problem:
\[
E_{\mathbb{P}} \left[ \sum_{t=0}^{T-1} c(x_t, u_1^t, \ldots, u_L^t) \right]
\]
For this problem, if at time \( t \geq 0 \) each of the decision makers (say DM \( i \)) has access to \( P(dx_t|y_{0,t-1}^1, y_{0,t-1}^2, \ldots, y_{0,t-1}^L) \) and their local observation \( y_{0,t}^i \) show that they can obtain a solution where the optimal decision rules only uses
\[
\{P(dx_t|y_{0,t-1}^1, y_{0,t-1}^2, \ldots, y_{0,t-1}^L), y_t^i\}.
\]
What if, they do not have access to \(P(dx(t)|y_1^{[0,t-1]}, y_2^{[0,t-1]}, \ldots, y_L^{[0,t-1]})\), and only have access to \(y_i^{[0,t]}\)? What would be a sufficient statistic for each decision maker for each time stage?

**Exercise 9.11.6** Two decision makers, Alice and Bob, wish to control a system:

\[
\begin{align*}
    x_{t+1} &= ax_t + u^a_t + u^b_t + w_t, \\
    y^a_t &= x_t + v^a_t, \\
    y^b_t &= x_t + v^b_t,
\end{align*}
\]

where \(u^a_t, y^a_t\) are the control actions and the observations of Alice, \(u^b_t, y^b_t\) are those for Bob and \(v^a_t, v^b_t, w_t\) are independent zero-mean Gaussian random variables with finite variance. Suppose the goal is to minimize for some \(T \in \mathbb{Z}_+\):

\[
E_{x_0}^{\Pi^a, \Pi^b} \left[ \sum_{t=0}^{T-1} x^2_t + r_a(u^a_t)^2 + r_b(u^b_t)^2 \right],
\]

for \(r_a, r_b > 0\), where \(\Pi^a, \Pi^b\) denote the policies adopted by Alice and Bob. Let the local information available to Alice be \(I^a_t = \{y^a_s, u^a_s, s \leq t-1\} \cup \{y^a_t\}\), and \(I^b_t = \{y^b_s, u^b_s, s \leq t-1\} \cup \{y^b_t\}\) is the information available at Bob at time \(t\).

Consider an \(n\)--step delayed information pattern: In an \(n\)--step delayed information sharing pattern, the information at Alice at time \(t\) is

\[
I^a_t \cup I^a_{t-n},
\]

and the information available at Bob is

\[
I^b_t \cup I^b_{t-n}.
\]

State if the following are true or false:

a) If Alice and Bob share all the information they have (with \(n = 0\)), it must be that, the optimal controls are linear.

b) Typically, for such problems, for example, Bob can try to send information to Alice to improve her estimation on the state, through his actions. When is it the case that Alice cannot benefit from the information from Bob, that is for what values of \(n\), there is no need for Bob to signal information this way?

c) If Alice and Bob share all information they have with a delay of 2, then their optimal control policies can be written as

\[
\begin{align*}
    u^a_t &= f_a(E[x_t|I^a_{t-2}, I^b_{t-2}, y^a_{t-1}, y^b_{t-1}] ), \\
    u^b_t &= f_b(E[x_t|I^a_{t-2}, I^b_{t-2}, y^a_{t-1}, y^b_{t-1}] ),
\end{align*}
\]

for some functions \(f_a, f_b\). Here, \(E[.|.]\) denotes the expectation.

d) If Alice and Bob share all information they have with a delay of 0, then their optimal control policies can be written as

\[
\begin{align*}
    u^a_t &= f_a(E[x_t|I^a_{t}, I^b_{t}] ), \\
    u^b_t &= f_b(E[x_t|I^a_{t}, I^b_{t}] ),
\end{align*}
\]

for some functions \(f_a, f_b\). Here, \(E[.|.]\) denotes the expectation.
Controlled Stochastic Differential Equations

This chapter introduces the basics of stochastic differential equations. A complete treatment is beyond the scope of these notes, however, the essential tools and ideas will be presented. The reader is referred to e.g. [93] [7] for a more comprehensive treatment.

The construction of a stochastic process on a continuous time interval, such as $[0, T]$ requires more delicacy when compared with a discrete-time stochastic process, as we will observe. In this chapter, we will primarily be concerned with controlled Markov processes $X_t$, each taking values in $\mathbb{R}^n$ for $t \in [0, T]$ or $t \in [0, \infty)$.

10.1 Continuous-time Markov processes

10.1.1 Two ways to construct a continuous-time Markov process

As discussed in Chapter 1 and Section 1.4, one way to define a stochastic process is to view it as a vector valued random variable. However, this requires us to put a proper topology on the set of sample paths, which may be a restrictive task. We will come back to this later.

Another definition would involve defining the process on finitely many time instances: Let $\{X_t(\omega), t \in [0, T]\}$ be stochastic process so that for each $t$, $X_t(\omega)$ is a random variable measurable on some probability space $\Omega, F, P$. We can define the $\sigma$-algebra generated by cylinder sets (as in Chapter 1) of the form:

$$\{\omega \in \Omega : X_{t_1}(\omega) \in A_1, X_{t_2}(\omega) \in A_2, \ldots, X_{t_N}(\omega) \in A_N, A_k \in B(\mathbb{R}^n), k \in \mathbb{N}\}$$

By defining a stochastic process in this fashion and assigning probabilities to such finite dimensional events, Theorem 1.2.3 implies that there exists a unique stochastic process on the $\sigma$-algebra generated by the sets of this form. However, unlike a discrete-time stochastic process, not all properties of the stochastic process is captured by finite dimensional distributions of it and the $\sigma$-field generated by such sets is not a rich set of sets. For example the set of sample paths that satisfy $\sup_{t \in [0, 1]} |X_t(\omega)| \leq 10$ may not be a well-defined event in this $\sigma$-algebra. Likewise, the extension theorem considered in Theorem 1.2.2 requires a probability measure already defined on the cylinder sets; it may not be possible to define such a probability measure by only considering finite dimensional distributions; see [130] for a study of such intricacies. Thus, establishing the existence of a probability measure for the process itself is quite an involved task for such continuous processes.

One may expect that if a stochastic process has continuous sample paths, then by specifying the process on rational time instances will uniquely define the process. Thus, if the process is known to admit certain regularity properties, the technical issues regarding only defining a process on finitely many sample points disappear.
10.1.2 The Brownian motion

Definition 10.1.1 A stochastic process \( B_t \) is called a Wiener process or Brownian motion if (i) the finite dimensional distributions of \( B_t \) are such that \( B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}} \) are independent Gaussian zero mean random variables with \( B_k - B_s \sim N(0, k - s) \), and (ii) \( B_t \) has continuous sample paths.

Such a process exists, see e.g. [119].

Now, going back to the construction we discussed in the previous section, we can define the Brownian motion as a \( C([0, \infty)) \)-valued (that is, a continuous) random variable: The topology on \( C([0, \infty)) \) is the topology of uniform convergence on compact sets (this is a stronger convergence than the topology of pointwise convergence but weaker than the topology of uniform convergence over \( \mathbb{R} \)). This is in agreement with the finite dimensional characterization through which we could define the Brownian motion.

Remark 10.1. A more general topology (than the uniform convergence on compact sets) is the Skorokhod topology defined on the space of functions which are right continuous with left limits: Such a topology defines a separable metric space. However, this discussion is beyond the scope of this chapter [19].

Remark 10.2 (Why Brownian Motion and not any other process?). The Gaussian properties of the continuous limit process is universal in the sense that, any continuous time process with sufficiently regular independent increments (thus with the martingale property) must be the Brownian process (see e.g., [119 Theorem 3.4.1]). In particular, even though typically in the construction of the Brownian motion (or its existence), one considers Gaussian i.i.d. random increments and takes its limit; this is not necessary for the Gaussian properties of the limit. This fact also justifies the use of the Brownian motion for stochastic integration.

10.1.3 The Itô Integral

On White noise

In many physical systems, one encounters models of the form

\[
\frac{dx}{dt} = (f(x_t) + u_t) + n_t,
\]

where \( n_t \) is some noise process. In engineering, one would like to model the noise process to be white, in the sense that \( n_t, n_s \) are independent for \( t \neq s \), and yet, \( n_t \) is zero-mean and with a finite variance for all \( t \).

We call such a process white, because the Fourier transform (and thus the spectrum) of the correlation function defined as \( R(\tau) = E[x_{t_1}x_{t_1+\tau}] \) of such a process is a constant: If the process is a discrete-time process, then this interpretation would be directly applicable since the Fourier transform of a discrete-time impulse would be constant for all frequency values.

For a continuous-time process, however, if \( R(\tau) = E[x_{t_1}x_{t_1+\tau}] = 0 \) for all \( \tau \neq 0 \), then a serious complication arises. If \( R(0) < \infty \), then this signal has zero-energy and its Fourier transform would be identically zero. If \( R(0) = \infty \), then such an \( R \) would have significant irregularities: Such a process would have its correlation function as \( E[n_t n_s] = \delta(t - s) \); where the dirac delta function \( \delta \) (as you may remember from MTHE 335) is a distribution acting on a proper set of test functions (such as the Schwartz signals \( S \)). Such a process is not well-defined Gaussian process since it doesn’t have a well-defined correlation function as \( \delta \) itself is not a function. But just as we discussed in MTHE 335, one can view this process as a distribution, or always work under an integral sign; this way one can make an operational use for such a definition. For example, with \( f, g \) nice functions such as those in \( S \), one could talk about

\[
E \left[ \left( \int f(t) n_t dt \right) \left( \int g(s) n_s ds \right) \right] = \int_{t,s} f(t) g(s) E[n_t n_s] dt ds = \int f(t) g(t) dt,
\]

and obtain operational properties.
Thus, it is evident that \( \{ n_t \} \), as a signal, would not be an ordinary process. On the other hand, we can define an integral of such a process, called \( B_t \), such that \( B_t \) is a continuous process. The derivative of such a process will not exist in the ordinary sense but it will exist in the sense described above.

Thus, while working with \( B_t \), instead of derivatives, we will study integrations. However, it will be evident that we cannot take the ordinary Lebesgue or Riemann integrations for \( B_t \) since \( B_t \) is too irregular. Instead, a creative method to obtain integrations will be introduced: The Itô integral provides a well-defined integration. The properties of integration, differentiation, chain rule etc. for such integrations will be called *stochastic calculus*. Finally, we will add control to the dynamics.

**The Itô integration**

We define a differential equation or a stochastic integral as an appropriate limit to make sense of the expression:

\[
\int_0^t f(s, \omega) dB_s(\omega)
\]

The definition is done in a few steps, as in the construction of the Lebesgue integral briefly discussed in Chapter 1. First, we define simple functions of the form:

\[
f(t, \omega) = \sum_k e_k(\omega) 1_{\{t \in [k2^{-n}, (k+1)2^{-n}]\}}
\]

where \( n \in \mathbb{N} \). We attempt to define

\[
\int_0^T f(t, \omega) dB_t(\omega) = \sum_k e_k(\omega) [B_{t_{k+1}}(\omega) - B_{t_k}(\omega)]
\]

where \( t_k = t_k^{(n)} = k2^{-n}, k = \{0, 1, \cdots, T2^{-n}\} \). In the following, we use this notation.

Now, note that if one defines:

\[
f_1(t, \omega) = \sum_k B_{\{k2^{-n}\}} 1_{\{t \in [k2^{-n}, (k+1)2^{-n}]\}}
\]

it can be shown that

\[
E\left[\int_0^T f_1(t, \omega) dB_t(\omega)\right] = 0,
\]

but with

\[
f_2(t, \omega) = \sum_k B_{\{(k+1)2^{-n}\}} 1_{\{t \in [k2^{-n}, (k+1)2^{-n}]\}}
\]

it can be shown that

\[
E\left[\int_0^T f_1(t, \omega) dB_t(\omega)\right] = T!
\]

Thus, even though both \( f_1 \) and \( f_2 \) look to be reasonable approximations for some function \( f(t, \omega) \), such as \( B_t(\omega) \), the integrals have drastically different meanings.

In particular the variations in the \( B_t \) process is too large to define an integration (in the usual sense of Riemann-Stieltjes), as we discuss further below: It does make a difference on whether one defines \( \int_0^T f(t, \omega) dB_t(\omega) \) as an appropriate limit of a sequence of expressions

\[
\sum_k f(t_k^*, \omega) [B_{\min(T,t_{k+1})}(\omega) - B_{\max(s,t_k)}(\omega)]
\]
for some \( f(t_j^*, \omega) \) with \( t_j^* \in [t_j, t_{j+1}] \). If we take \( t_j^* = t_j \) (the left end point), this is known as the Itô Integral. If we take \( t_j^* = \frac{1}{2}(t_j + t_{j+1}) \), this is known as the Stratonovich integral.

To gain further insight as to why this leads to an issue, see discuss the following.

Using the independent-increments property (that is (i) in Definition \[10.1.1\]) of the Brownian motion, the following can be shown:

**Lemma 10.1.1** In mean-square and hence in probability

\[
\sum_k (B_{t_{k+1}} - B_{t_k})^2 \to b - a.
\]

**Proof.** This follows from uniformly discretizing the time domain, applying the properties of the Brownian motion, and crucially observing that \( E[\sum_k (B_{t_{k+1}} - B_{t_k})^4] \to 0 \) (this follows from the Gaussian nature of the increments). \( \diamond \)

Observe the following \[119\].

**Theorem 10.1.1** Define the total variation of the Wiener process as:

\[
TV(B, a, b) = \sup_{0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq b, k \in \mathbb{N}} \sum_k |B_{t_{k+1}} - B_{t_k}|
\]

Almost surely, \( TV(B, a, b) = \infty \).

**Proof.** By Lemma \[10.1.1\], it follows that there exists some subsequence \( n_m \) so that \( \sum_k (B_{t_{k+1}^m} - B_{t_k^m})^2 \to b - a \) almost surely (see Theorem \[B.3.2\]). Now, if \( TV(B, a, b) < \infty \), this would imply that

\[
\sum_k (B_{t_{k+1}^m} - B_{t_k^m})^2 \leq \sup_k |B_{t_{k+1}^m} - B_{t_k^m}| \sum_k |B_{t_{k+1}^m} - B_{t_k^m}| \to 0,
\]
as \( n \to \infty \), since by continuity \( \sup_k |B_{t_{k+1}^m} - B_{t_k^m}| \to 0 \). This would lead to a contradiction. \( \diamond \)

Now, Itô’s integral will work well, if we restrict the integrand \( f(t, \omega) \) to be such that \( f(t, \omega) \) is measurable on the \( \sigma \)-field generated by \( \{B_s, s \leq t\} \). With this intuition, we define \( \mathcal{F}_t \) to be the \( \sigma \)-algebra generated by \( B_s, s \leq t \). In other words, \( \mathcal{F}_t \) is the smallest \( \sigma \)-algebra containing the sets of the form:

\[
\{\omega : B_{t_1}(\omega) \in A_1, \cdots, B_{t_k}(\omega) \in A_k\}, \quad t_k \leq t,
\]

for Borel \( A_1, \cdots, A_k \). We also assume that all sets of measure zero are included in \( \mathcal{F}_t \) (this process is known as the completion of a \( \sigma \)-field).

**Definition 10.1.2** Let \( \mathcal{N}_t, t \geq 0 \), be an increasing family of \( \sigma \)-algebras of subsets of \( \Omega \). A process \( g(t, \omega) \) is called \( \mathcal{N}_t \)-adapted if for each \( t \), \( g(t, \cdots) \) is \( \mathcal{N}_t \)-measurable.

**Definition 10.1.3** Let \( \mathcal{V}(S, T) \) be the class of functions:

\[
f(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}
\]

such that (i) \( f(t, \omega) \) is \( \mathcal{B}([0, \infty)) \times \mathcal{F} \)-measurable, (ii) \( f(t, \omega) \) is \( \mathcal{F}_t \)-adapted and (iii) \( E[\int_S^T f^2(t, \omega)dt] < \infty \).

We will often take \( S = 0 \) in the following.

For functions in \( \mathcal{V} \), the Itô integral is defined as follows: A function \( \phi \) is called elementary if it has the form:

\[
\phi(t, \omega) = \sum_k e_k(\omega)1_{\{t \in [t_k, t_{k+1})\}}
\]
with \( e_k \) being \( \mathcal{F}_{t_k} \)-measurable. For elementary functions, we define the \( \text{Itô} \) integral as:

\[
\int_0^T \phi(t, \omega) dB_t(\omega) = \sum_k e_k(\omega) \left( B_{t_{k+1}}(\omega) - B_{t_k}(\omega) \right) \tag{10.1}
\]

With this definition, it follows that for a bounded and elementary \( \phi \),

\[
E \left[ \left( \int_0^T \phi(t, \omega) dB_t(\omega) \right)^2 \right] = E \left[ \int_0^T \phi^2(t, \omega) dt \right].
\]

This property is known as the \( \text{Itô} \) isometry. The proof follows from expanding the summation in (10.1) and using the properties of the Brownian motion.

Now, the remaining steps to define the \( \text{Itô} \) integral are as follows:

- **Step 1**: Let \( g \in \mathcal{V} \) and \( g(\cdot, \omega) \) continuous for each \( \omega \). Then, there exist elementary functions \( \phi_n \in \mathcal{V} \) such that

\[
E \left[ \left( \int_0^T (g - \phi_n)^2 dt \right) \right] \to 0,
\]

as \( n \to \infty \). The proof here follows from the dominated convergence theorem.

- **Step 2**: Let \( h \in \mathcal{V} \) be bounded. Then there exist \( g_n \in \mathcal{V} \) such that \( g_n(\cdot, \omega) \) is continuous for all \( \omega \) and \( n \) and

\[
E \left[ \left( h - g_n \right)^2 dt \right] \to 0.
\]

- **Step 3**: Let \( f \in \mathcal{V} \). Then, there exists a sequence \( h_n \in \mathcal{V} \) such that \( h_n \) is bounded for each \( n \) and

\[
E \left[ \left( f - h_n \right)^2 dt \right] \to 0.
\]

Here, we use truncation and then the dominated convergence theorem.

**Definition 10.1.4 (The \( \text{Itô} \) Integral)** Let \( f \in \mathcal{V}(S, T) \). The \( \text{Itô} \) integral of \( f \) is defined by

\[
\int f(t, \omega) dB_t(\omega) = \lim_{n \to \infty} \int_0^T \phi_n(t, \omega) dB_t(\omega),
\]

where the limit is in \( L_2(P) \) in the sense that

\[
\lim_{n \to \infty} E \left[ \left( \int_0^T \phi_n(t, \omega) dB_t(\omega) - \int f(t, \omega) dB_t(\omega) \right)^2 \right] = 0
\]

The existence of a limit is established through the construction of a Cauchy sequence and the completeness of \( L_2(P) \), the space of measurable functions with a finite second moment under \( P \), with the norm as defined.

A computationally useful result is the following.

**Corollary 10.1.1** For all \( f \in \mathcal{V}(S, T) \)

\[
E \left[ \left( \int_0^T f(t, \omega) dB_t(\omega) \right)^2 \right] = E \left[ \int_0^T f^2(t, \omega) dt \right].
\]

And thus, if \( f, f_n \in \mathcal{V}(S, T) \) and
\[ E\left[ \int_0^T (f_n - f)^2dt \right] \to 0 \]

then in \( L_2(P) \)
\[ \int_0^T f_n(t,\omega)dB_t(\omega) \to \int_0^T f(t,\omega)dB_t(\omega) \]

**Example 10.3.** Let us show that
\[ \int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t. \] (10.2)

If we define the elementary function to be:
\[ \phi_n(\omega) = \sum B_j(\omega)1_{t \in [t_j, t_{j+1})}, \]
it follows that \( E\left[ \int_0^t (\phi_n - B_s)^2ds \right] \to 0 \). Therefore, the limit of the integrals of \( \phi_n(\omega) \), that is the \( L_2(P) \) limit of \( \sum_j B_j(B_{j+1} - B_j) \) will be the integral. Observe now that
\[ -(B_{j+1} - B_j)^2 = 2B_j(B_{j+1} - B_j) + B_j^2 - B_{j+1}^2 \]
and thus summing over \( j \), we obtain
\[ \sum_j -(B_{j+1} - B_j)^2 = \sum_j 2B_j(B_{j+1} - B_j) + B_j^2 - B_{j+1}^2, \]
leading to
\[ B_t^2 - \sum_j (B_{j+1} - B_j)^2 = \sum_j 2B_j(B_{j+1} - B_j) \]

Now, taking the intervals \([j, j + 1]\) arbitrarily small, we see that the first term converges to \( B_t^2 - t \) and the term on the right hand side converges to \( 2 \int_0^t B_s dB_s \), leading to the desired result. We will derive the same result using Itô’s formula shortly. The message of this example is to highlight the computational method: Find a sequence of elementary function which converges in \( L_2(P) \) to \( f \), and then compute the integrals, and take the limit as the intervals shrink.

**Remark 10.4.** An important extension of the Itô integral is to a setup where \( f_t \) is \( \mathcal{H}_t \)-measurable, where \( \mathcal{H}_t \subset \mathcal{F}_t = \sigma(B_s, s \leq t) \). In applications, this is important to let us apply the integration to settings where the process that is integrated is measurable only on a subset of the filtration generated by the Brownian process. This allows one to define multi-dimensional Itô integrals as well. This is particularly useful for controlled stochastic differential equations, where the control policies are measurable with respect to a filtration that does not contain that generated by the Brownian motion, but the controller policy cannot depend on the future realizations of the Brownian motion either.

**Remark 10.5.** As explained in [93] and many other texts, a curious student will question the selection of choosing the Itô integral over any other, and in particular the Stratonovich, integral. There is no convincing answer that we could state. Different applications are more suitable for either interpretation. In stochastic control, measurability aspects (of admissible controls) are the most crucial ones. If one appropriately defines the functional or stochastic dependence between a function to be integrated or a noise process, the application of either will come naturally: If the functions are to not look at the future (where the future is that of the Brownian motion or the filtration \( \mathcal{H}_t \) as discussed in Remark 10.4), then Itô’s formula is appropriate.

### 10.2 Itô Formula

Itô’s formula allows us to take integrations of functions of processes and it generalizes the chain rule in classical calculus.

**Definition 10.2.1** We say \( v(S, T) \in \mathcal{W}_\mathcal{H} \) if
\[ v(t, \omega) : [S, T] \times \Omega \to \mathbb{R} \]
such that (i) \( v(t, \omega) \) is \( B([0, \infty)) \times \mathcal{F} \)-measurable, (ii) \( v(t, \omega) \) is \( \mathcal{H}_t \)-adapted where \( \mathcal{H}_t \) is as in Remark 10.2 and (iii) \( P(\int_0^T f^2(t, \omega)dt < \infty) = 1 \).

**Definition 10.2.2 (Itô Process)** Let \( B_t \) be a one-dimensional Brownian motion on \( (\Omega, \mathcal{F}, P) \). A (one-dimensional) Itô process is a stochastic process \( X_t \) on \( (\Omega, \mathcal{F}, P) \) of the form

\[
X_t = X_0 + \int_0^t u(s, \omega) + \int_0^t v(s, \omega) dB_s \tag{10.3}
\]

where \( v \in \mathcal{W}_H \) so that \( v \) is \( \mathcal{H}_t \)-adapted and \( P(\int_0^t v^2(s, \omega)ds < \infty) = 1 \) for all \( t \geq 0 \). Likewise, \( u \) is also \( \mathcal{H}_t \)-adapted and \( P(\int_0^t u^2(s, \omega)ds < \infty) = 1 \) for all \( t \geq 0 \).

Instead of the integral form in (10.3), we may use the differential form:

\[
dX_t = u dt + v dB_t,
\]

with the understanding that this means the integral form.

**Theorem 10.2.1 [Itô Formula]** Let \( X_t \) be an Itô process given by

\[
dX_t = u dt + v dB_t.
\]

Let \( g(t, x) \in C^{1,2}([0, \infty) \times \mathbb{R}) \) (that is, \( g \) is \( C^1 \) in \( t \) and \( C^2 \) in \( x \)). Then,

\[
Y_t = g(t, X_t),
\]

is again an Itô process and

\[
dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(t, X_t)(dX_t)^2
\]

where

\[
(dX_t)^2 = (dX_t)(dX_t)
\]

with \( dt dt = dt dB_t = dB_t dt = 0 \) and \( dB_t dB_t = dt \).

Before we proceed to the proof, let us note that if instead of \( dB_t \), we only had a differentiable function \( m_t \) so that \( dX_t = u dt + vdm_t \), the regular chain rule would lead to:

\[
dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)(udt + vdm_t).
\]

**Remark 10.6.** Note that Itô’s Formula is a generalization of the ordinary chain rule for derivatives. The difference is the presence of the quadratic term that appears in the formula.

**Proof of Theorem 10.2.1** The proof follows from Taylor’s expansion with a particular attention to the second order remainder term, this is why we require the continuity in the second partial derivative with respect to the state variable. The goal is to show that

\[
g(t, X_t) = g(0, X_0) + \int_0^t \left( \frac{\partial g}{\partial s}(s, X_s) + u(s, \omega) \frac{\partial g}{\partial x}(s, X_s) + \frac{1}{2} v(s, \omega) \frac{\partial^2 g}{\partial x^2}(s, X_s) \right) ds \\
+ \int_0^t v(s, \omega) \frac{\partial g}{\partial x}(s, X_s)dB_s
\]

(10.4)

We first assume that \( g, \frac{\partial g}{\partial t}, \frac{\partial g}{\partial x} \) and \( \frac{\partial^2 g}{\partial x^2} \) are bounded. Once one can show the result under such conditions, the general case holds by the fact that any \( C^{1,2} \) function can be approximated by such bounded functions so that the approximating functions
converge to \( g, \frac{\partial g}{\partial t}, \frac{\partial g}{\partial x} \) and \( \frac{\partial^2 g}{\partial x^2} \) uniformly on compact subsets of \([0, \infty) \times \mathbb{R}\). Furthermore, from our earlier analysis, we can restrict \( u \) and \( v \) to be elementary functions. Using Taylor’s theorem we obtain

\[
g(t, X_t) = g(0, X_0) + \sum_j (g(t_{j+1}, X_{t_{j+1}}) - g(t_j, X_{t_j}))
\]

\[
= g(0, X_0) + \sum_j \frac{\partial g}{\partial t} (t_{j+1} - t_j) + \sum_j \frac{\partial g}{\partial x} (X_{t_{j+1}} - X_{t_j})
\]

\[
+ \frac{1}{2} \sum_j \frac{\partial^2 g}{\partial x^2} (X_{t_{j+1}} - X_{t_j})^2
\]

(10.5)

where \( \frac{\partial g}{\partial x} \) are computed at the points \( t_j, X_{t_j} \), and \( \frac{\partial^2 g}{\partial x^2} \) are computed on some point between \( t_j, t_{j+1} \) and \( X_{t_j}, X_{t_{j+1}} \) following Taylor’s expansion with a remainder term. Now, if \( (t_{j+1} - t_j)^2 \to 0 \),

\[
\sum_j \frac{\partial g}{\partial t} (t_j, X_j) (t_{j+1} - t_j) \to \int_0^t \frac{\partial g}{\partial t} (t, X_t) ds
\]

\[
\sum_j \frac{\partial g}{\partial x} (t_j, X_j) (X_{t_{j+1}} - X_{t_j}) \to \int_0^t \frac{\partial g}{\partial x} (t, X_t) dX_t
\]

Since \( u, v \) are elementary,

\[
\sum_j \frac{\partial^2 g}{\partial x^2} (X_{t_{j+1}} - X_{t_j})^2
\]

\[
= \sum_j \frac{\partial^2 g}{\partial x^2} u^2 (t_{j+1} - t_j)^2 + 2 \sum_j \frac{\partial^2 g}{\partial x^2} u(t_j, \omega) v(t_j, \omega) (t_{j+1} - t_j) (B_{t_{j+1}} - B_{t_j})
\]

\[
+ \sum_j \frac{\partial^2 g}{\partial x^2} v^2 (t_{j+1} - t_j)^2
\]

(10.6)

It follows that as \( t_{j+1} - t_j \to 0 \), the first two terms go to zero and the last term tends to \( \int \frac{\partial^2 g}{\partial x^2} v^2 (s, \omega) ds \) in \( L_2(P) \) (note that this leads to the interpretation that \( dB_t^2 = dt \)).

**Example 10.7.** Compute:

\[
\int_0^t B_s dB_s
\]

View \( Y_t = \frac{1}{2} B_t^2 \). Then,

\[
dY_t = B_t dB_t + \frac{1}{2} dt
\]

and thus

\[
\int dY_s = Y_t - Y_0 = \frac{1}{2} B_t^2 = \int B_s dB_s + \frac{1}{2} t.
\]

Note that this is in agreement with \( (10.2) \).

A useful application of Itô’s formula is the following integration by parts formula, obtained by taking \( Y_t = f(t) B_t \) and \( X_t = B_t \):

**Theorem 10.2.2** Let \( f(s, \omega) \) be continuous and have bounded variation with respect to \( s \in [0, t] \) almost surely. Then,

\[
\int_0^t f(s) dB_s = f(t) B_t - \int_0^t B_s df_s.
\]
Itô’s Formula can be extended to higher dimensions.

**Theorem 10.2.3** Let $dX_t = u dt + v dB_t$, be an $n$-dimensional Itô process. Let $g(t, x) = (g_1(t, x), \cdots, g_p(t, x))$ be $C^{1,2}$ mapping from $([0, \infty) \times \mathbb{R}^n) \to \mathbb{R}^p$. Then the process $Y(t, \omega) = g(t, X_t)$ is also an Itô process and its $k$th component satisfies

$$dY_k = \frac{\partial g_k}{\partial t}(t, X_t)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t, X_t)dX^i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t) dX^i dX^j,$$

where $dB_t^i dB_t^j = 1_{i\neq j} dt, dB_t^i dt = dt dB_t^i = dt dt = 0$

### 10.3 Stochastic Differential Equations

Consider now an equation of the form:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$$

(10.7)

with the interpretation that this means

$$X_t = X_0 + \int_0^t b(t, X_t)dt + \int_0^t \sigma(t, X_t)dB_t$$

Three natural questions are as follows: (i) Does there exist a solution to this differential equation? (ii) Is the solution unique? (iii) How can one compute the solution?

**Theorem 10.3.1 (Existence and Uniqueness Theorem)** Let $T > 0$ and $b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions satisfying:

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|), \quad t \in [0, T], x \in \mathbb{R}^n$$

for some $C \in \mathbb{R}$ with $|\sigma|^2 = \sum_{i,j} \sigma_{ij}^2$, and

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D(|x - y|), \quad t \in [0, T], x, y \in \mathbb{R}^n$$

for some constant $D$. Let $X_0 = Z$ be a random variable which is independent of the $\sigma$-algebra generated by $B_s$, $s \geq 0$ with $E[|Z|^2] < \infty$. Then, the stochastic differential equation (10.7) has a unique solution $X_t(\omega)$ that is continuous in $t$ with the property that $X_t$ is adapted to the filtration generated by $\{Z, B_s, s \leq t\}$ and $E[\int_0^T |X_t|^2] < \infty$.

**Proof Sketch.** The proof of existence follows from a similar construction for the existence of solutions to ordinary differential equations: One defines a sequence of iterations:

$$dY^{k+1}_t = X_0 + \int_0^t b(s, Y^k_t)ds + \int_0^t \sigma(s, Y^k_t)dB_s$$

Then, the goal is to obtain a bound on the $L^2$-errors so that

$$\lim_{m,n \to \infty} E[|Y^m_t - Y^n_t|^2 dt] \to 0,$$

so that $Y^n_t$ is a Cauchy sequence under the $L^2(P)$ norm. Call the limit $X$. The next step is to ensure that $X$ indeed satisfies the equation and that there can only be one solution. Finally, one proves that $X_t$ can be taken to be continuous. Let us appreciate some of the conditions stated.

**Remark 10.8.** Consider the following deterministic differential equations:
10.3 Controlled Stochastic Differential Equations

- The differential equation
  \[ \frac{dx}{dt} = \frac{4}{t} x_t \]
  with \( x(1) = 1 \) does not admit a unique solution on the interval \([-1, 1]\).

- The differential equation
  \[ \frac{dx}{dt} = x^2 \]
  with \( x(0) = 1 \) admits the solution \( x_t = \frac{1}{1 - t} \) and as \( t \uparrow 1 \), the solution *explodes* in finite time.

The solution discussed above is what is called a strong solution. Such a solution is such that \( X_t \) is measurable on the filtration generated by the Brownian motion and the initial variable. Such a solution has an important engineering/control appeal in that the solution is completely specified once the realizations of the Brownian motion are together with the initial state are specified.

In many applications, however, the conditions of Theorem 10.3.1 do not hold. In this case, one cannot always find a strong solution. However, in this case, one may be able to find a solution which satisfies the probabilistic flow in the system so that the evolution of the probabilities are satisfied: Note, however that, this solution is no longer adapted to the filtration generated by the Brownian motion and the initial state. Such a solution is called a weak solution or a martingale solution.

This is related to the solution to the Fokker-Planck equation that we will discuss further in the chapter in Section 10.4.1.

10.4 Some Properties of SDEs

**Definition 10.4.1** An Itô diffusion is a stochastic process \( X_t(\omega) \) satisfying a stochastic differential equation of the form:

\[ dX_t = b(X_t)dt + \sigma(X_t)dB_t, \quad t \geq s, \, X_s = x \]

where \( B_t \) is \( m \)-dimensional Brownian motion and \( b, \sigma \) satisfy the conditions of Theorem 10.3.1 so that

\[ |b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|. \]

Note that here \( b, \sigma \) only depend on \( x \) and not on \( t \). Thus, the process here is time-homogenous.

**Theorem 10.4.1** Let \( f \) be Borel. Then, for \( t, h \geq 0 \):

\[ E_x[f(X_{t+h})|\mathcal{F}_t](\omega) = E_{X_t(\omega)}[f(X_h)] \]

**Theorem 10.4.2** (Strong Markov Property) Let \( f \) be Borel and \( \tau \) be a stopping time with respect to \( \mathcal{F}_t \). Then, for \( h \geq 0 \), conditioned on the event that \( \tau < \infty \):

\[ E_x[f(X_{t+h})|\mathcal{F}_{\tau}](\omega) = E_{X_{\tau}(\omega)}[f(X_h)] \]

**Definition 10.4.2** Let \( X_t \) be a time-homogenous Itô diffusion in \( \mathbb{R}^n \). The infinitesimal generator \( A \) of \( X_t \) is defined by:

\[ Af(x) = \lim_{t \to 0} \frac{E_x[f(X_t)] - f(x)}{t}, \quad x \in \mathbb{R}^n. \]

**Lemma 10.4.1** Let \( Y_t = Y_t^x \) be an Itô process in \( \mathbb{R}^n \) of the form:

\[ Y_t^x(\omega) = x + \int_0^t u(s, \omega) + \int_0^t v(s, \omega)dB_s(\omega). \]
Let $f \in C^2_0(\mathbb{R}^n)$, that is $f$ is twice continuously differentiable and has compact support and $\tau$ be a stopping time with respect to $\mathcal{F}_t$ with $E_x[\tau] < \infty$. Assume that $u,v$ are bounded. Then,

$$E[f(Y_\tau)] = f(x) + E_x\left[ \int_0^\tau \left( \sum_i u^i(s,\omega) \frac{\partial f}{\partial x^i}(Y_s) + \frac{1}{2} \sum_{i,j} (vv^T)_{ij}(s,\omega) \frac{\partial^2 f}{\partial x^i \partial x^j}(Y_s) \right) ds \right].$$

This lemma, combined with Definition \[10.4.2\] gives us the following result:

**Theorem 10.4.3** Let $dX_t = b(X_t)dt + \sigma(X_t)dB_t$. If $f \in C^2_0(\mathbb{R}^n)$, then,

$$Af(X_s) = \left( \sum_i b^i(x) \frac{\partial f}{\partial x^i}(X_s) + \frac{1}{2} \sum_{i,j} (\sigma \sigma^T)_{ij}(s,\omega) \frac{\partial^2 f}{\partial x^i \partial x^j}(X_s) \right)$$

A very useful result follows.

**Theorem 10.4.4 (Dynkin’s Formula)** Let $f \in C^2_0(\mathbb{R}^n)$ and $\tau$ be a stopping time with $E_x[\tau] < \infty$. Then,

$$E_x[f(X_\tau)] = f(x) + E_x\left[ \int_0^\tau Af(X_s)ds \right]$$

**Remark 10.9.** The conditions for Dynkin’s Formula can be generalized. As in Theorem \[4.1.5\] if the stopping time $\tau$ is bounded by a fixed constant, the conditions on $f$ can be relaxed. Furthermore if $\tau$ is the exit time from a bounded set, then it suffices that the function is $C^2$ and does not have compact support.

**Remark 10.10.** Consider a stochastic differential equation: $dX_t = b(X_t)dt + \sigma(X_t)dB_t$. A probability measure $P$ on the sample path space (or its stochastic realization $X_t$) is said to be a weak solution if

$$f(X_t) - \int_0^t Af(X_s)ds$$

is a martingale with respect to $\mathcal{M}_t = \sigma(X_s, s \leq t)$, for any $C^2$ function $f$ with bounded first and second order partial derivatives. Every strong solution is a weak solution, but not every weak solution is a strong solution; the stochastic realization may not be defined on the original probability space as a measurable function of the original Brownian motion. For example, if $X_t$ can be defined to be randomized, where the randomization variables are independent noise processes, one could embed the noise terms into a larger filtration; this will lead to a weak solution but not a strong solution since there is additional information required (that is not contained in the original Brownian process).

### 10.4.1 Fokker-Planck equation

The discussion on the generator function (and Dynkin’s formula) suggests that one can compute the evolution of the probability measure $p_t(\cdot) = P(X_t \in \cdot)$, by considering for a sufficiently rich class of functions $f \in \mathcal{D}$

$$E[f(X_t)] = \int p_t(dx) f(x).$$

Suppose that we assume that $p_t$ admits a density and this is denoted by the same letter. Furthermore, let $p(x, t) := p_t(x)$. By taking $\mathcal{D}$ to be the space of smooth signals with compact support, which is a dense subset of the space of square integrable functions on $\mathbb{R}$, using the generator function, and applying integration by parts twice, we obtain that for sources of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t$$

the following equation:
\[
\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x}(b(x)p(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2}(\sigma(x)p(x, t))
\]
with \(p(x, 0) = f(x)\) being the initial probability density. This is the celebrated Fokker-Planck equation.

The Fokker-Planck equation is a partial differential equation whose existence for a solution requires certain technical conditions. As we discussed earlier, this is related to having a weak solution to a stochastic differential equation and in fact they typically imply one another. Of course, the Fokker-Planck equation may admit a density as a solution, but it may also admit a solution in a further weaker sense in that the evolution of the solution measure \(P(X_t \in \cdot)\) may not admit a density. We refer the reader to [95] for a detailed discussion on the Fokker-Planck equation.

10.5 Controlled Stochastic Differential Equations and the Hamilton-Jacobi-Bellman Equation

Suppose now that we have a controlled system:

\[
dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dB_t,
\]
where \(u_t \in U\) is the control action variable. We assume that \(u\) is measurable at least on \(\mathcal{F}_t\) (but we can restrict this further so that it is measurable on a strictly smaller sigma field). Thus, the differential equation is well defined as an Itô integral. We will assume that a solution exists.

As we discussed extensively throughout the notes, the selection of the control actions need to be measurable with respect to some information at the controller. If \(u_t\) is measurable on the filtration generated by \(X_t\), then the policy is called admissible. If it is only a function of \(X_t\) and \(t\), then it is called a Markov policy. If it only depends on \(X_t\), then it is stationary. Randomization also is possible, but this requires a more careful treatment when compared with the discrete-time counterpart (see [7]).

Let us first restrict the policies to be Markov. We will see that under a verification theorem, this is without loss, under certain conditions.

The reader is encouraged to take a look at the verification theorems for discrete-time problems: These are Theorem 5.1.3 for finite horizon problems, Theorem 5.5.3 for discounted cost problems, and Theorem 7.1.1 for average cost problems. You can see that the essential difference is to express the expectations through Dynkin’s formula and the differential generators.

10.5.1 Finite Horizon Problems

Suppose that given a control policy, the goal is to minimize

\[
E[\int_0^T c(s, X_s, u_s)ds + c_T(X_T)],
\]
where \(c_T\) is some terminal cost function.

As in Chapter 5, if a policy is optimal, it is optimal in every possible state:

\[
V_r(X_r) = \min_u E[\int_r^T c(s, X_s, u_s)ds + V_t(X_t)|X_r].
\]

In the following, we provide an informal discussion. Let us assume that \(V_r(x) = V(s, x)\) is \(C^2\). Then,

\[
0 = \min_u \left( E[\int_r^T c(s, X_s, u_s)ds + V_t(X_t)|X_r] - V_r(X_r) \right)
\]

In particular,
0 = \lim_{t \to r} \min_u \left( E\left[ \int_t^r c(s, X_s, u_s) ds + V_t(X_t) \right] - V_r(X_r) \right)

and by changing the order (when can we do this, what are some conditions on \(b\) and \(\sigma\) through Dynkin’s lemma?)

0 = \min_u \lim_{t \to r} \left( E\left[ \int_t^r c(s, X_s, u_s) ds + V_t(X_t) \right] - V_r(X_r) \right)

and using Itô’s Formula leading to

\[ \min_u (c(s, x, u) + AV(x) + \frac{\partial V}{\partial s}(x) = 0 \] (10.8)

Thus, if a policy is optimal, it needs to satisfy the above property provided that \(V\) satisfies the necessary regularity conditions. In fact, this is also sufficient, as the following verification theorem shows. We note that, the verification theorem shows that, an optimal Markov policy that satisfies the verification is optimal over all possible admissible policies.

**Theorem 10.5.1 (Verification Theorem)** Consider: \(dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dB_t\), and let

\[ \mathcal{L}_t^u g(x) = \sum_i b^i(t, x, u) \frac{\partial g}{\partial x^i}(x) + \frac{1}{2} \sum_{i,j} \sigma^i(t, x) \sigma^j(t, x, u) \frac{\partial^2 g}{\partial x^i \partial x^j}(x) \]

Suppose that \(V_t\) is \(C^2(t, x)\) and

\[ \frac{\partial V_t}{\partial t}(x) + \min_u \left( \mathcal{L}_t^u V_t(x) + c(t, x, u) \right) = 0, \quad V_T(x) = c_T(x), \] (10.9)

and \(|E[V_0(X_0)]| < \infty\). Then, an admissible control policy which achieves the minimum for every \((t, x)\) is optimal.

**Proof.** For any admissible control policy, using Itô’s rule

\[ E[V_0(X_0)] = E\left[ \int_0^T \left( -\frac{\partial V_s}{\partial s}(X_s^u) - \mathcal{L}_s^u V_s(X_s^u) \right) ds \right] + E[V_T(X_T^u)] \]

The equation \(\frac{\partial V_t}{\partial t}(x) + \min_u \mathcal{L}_t^u V_t(x) + c(t, x, u) = 0\) implies that for any admissible control: \(\frac{\partial V_t}{\partial t}(x) + \min_u \mathcal{L}_t^u V_t(x) + c(t, x, u) \geq 0\) and

\[-\frac{\partial}{\partial s} V_s(X_s^u) - \mathcal{L}_s^u V_s(X_s^u) \leq c(s, X_s^u, u_s),\]

and thus for any admissible control:

\[ E[V_0(X_0)] \leq E\left[ \int_0^T (c(s, X_s^u, u_s)ds + c_T(X_T^u)) \right] = J(u). \]

On the other hand, a policy \(u^*\) which satisfies the equality in (10.9), leads to an equality and hence \(J(u^*) \leq J(u)\) for any other admissible control.

**Example 10.11.** (Optimal portfolio selection) A common example in finance applications is the portfolio selection problem where a controller (investor) would like to optimally allocate his wealth between a stochastic stock market and a market with a guaranteed income (see [119]). Consider a stock with an average return \(\mu > 0\) and volatility \(\sigma > 0\) and a bank account with interest rate \(r > 0\). These are modeled by:

\[ dS_t = \mu S_t dt + \sigma S_t dB_t, \quad S_0 = 1 \]

\[ dR_t = r R_t dt, \quad R_0 = 1 \]

Suppose that the investor can only use his own money to invest and let \(u_t \in [0, 1]\) denote the proportion of the money that he invests in the stock. This implies that at any given time, his wealth dynamics is given by:
\[
dX_t = \mu u X_t \, dt + \sigma u X_t \, dW_t + r(1 - u_t) X_t \, dt,
\]
or
\[
dX_t = \left( \mu u + r(1 - u) \right) X_t \, dt + \sigma u X_t \, dB_t,
\]
Suppose that the goal is to maximize \( E[\log(X_T)] \) for a fixed time \( T \) (or minimize \( -E[\log(X_T)] \)). In this case, the Bellman equation writes as:
\[
0 = \frac{\partial V_t(x)}{\partial t} + \min_u \left( \frac{\sigma^2 u^2 x^2}{2} \frac{\partial^2 V_t(x)}{\partial x^2} + (\mu u + r(1 - u) x) \frac{\partial V_t(x)}{\partial x} \right),
\]
with \( V_T(x) = -\log(x) \). With a guess of the value function of the form \( V_t(x) = -\log(x) + b_t \), one obtains an ordinary differential equation for \( b_t \) with terminal condition \( b_T = 0 \). It follows that the optimal control is \( u_t(x) = \frac{\mu - r}{\sigma^2} \) and leading to \( V_t(x) = -\log(x_t) - C(T - t) \), for some constant \( C \).

Example 10.12 (The Linear Regulator). Consider a continuous-time counterpart of the LQG problem: Let
\[
dX^u_t = AX^u_t \, dt + Bu_t \, dt + CdW_t,
\]
where \( X^u_t, u_t \) are multi-dimensional, with the goal of minimizing:
\[
J(\gamma) = E[\int_0^T e^{-\lambda s} c(X_s, u_s) \, ds],
\]
where \( R > 0, P_T > 0, Q \geq 0 \). Using the verification theorem, by guessing \( V_t(x) = x^T P_t x \), we obtain through the Bellman’s equation
\[
u_t^* = -R^{-1} B^T P_t X_t,
\]
and
\[
\frac{d}{dt} P_t + AP + PA - PBR^{-1} B^T P + P = 0,
\]
where the last equation is known as the continuous-time Riccati equation.

10.5.2 Discounted Infinite Horizon Problems

Suppose that given a control policy, the goal is to minimize
\[
E[\int_0^\infty e^{-\lambda s} c(X_s, u_s) \, ds].
\]
In this section, we will consider a time-homogenous setup: \( dX_t = b(X_t, u_t) \, dt + \sigma(X_t, u_t) \, dB_t \), and let
\[
\mathcal{L}^u g(x) = \sum_i b^i(x, u) \frac{\partial g}{\partial x^i}(x) + \frac{1}{2} \sum_{i,j} \sigma^i(x, u) \sigma^j(x, u) \frac{\partial^2 g}{\partial x^i \partial x^j}(x)
\]
In this case, we have the following result:

Theorem 10.5.2 (Verification Theorem) Suppose that \( V \) is \( C^2(x) \)
\[
\min_u \mathcal{L}^u V(x) - \lambda V(x) + c(x, u) = 0,
\]
and \( |E[V_0(X_0)]| < \infty \). Then, an admissible control policy which achieves the minimum for every \( (t, x) \) is optimal.

Proof. For any admissible control policy, using Itô’s rule for \( V(X_t)e^{-\lambda t} \) one obtains:
\[ E[V(X_0)] - e^{-\lambda t}E[V(X_t^u)] = E\left[ \int_0^t e^{-\lambda s}(-\mathcal{L}^u V(X_s) + \lambda V(X_s))ds \right] \]

Proceeding as before in Theorem 10.5.1 leads to the desired result.

10.5.3 Average-Cost Infinite Horizon Problems

Suppose that given a control policy, the goal is to minimize

\[ \limsup_{T \to \infty} \frac{1}{T} E\left[ \int_0^T c(X_s, u_s) ds \right]. \]

Once again we consider a time-homogenous setup: \( dX_t = b(X_t, u_t)dt + \sigma(X_t, u_t)dB_t \), and let

\[ \mathcal{L}^u g(x) = \sum_i b_i(x, u) \frac{\partial g}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j} \sigma^i(x) \sigma^j(x, u) \frac{\partial^2 g}{\partial x_i \partial x_j}(x) \]

**Theorem 10.5.3 (Verification Theorem)** Suppose that \( V \) is \( C^2(x) \) and \( \eta \in \mathbb{R} \) so that

\[ \min_u \mathcal{L}^u V(x) - \eta + c(x, u) = 0, \quad (10.11) \]

with the set of admissible strategies that satisfies:

\[ \limsup_{T \to \infty} \frac{E[V(X_0) - V(X_T^u)]}{T} = 0 \]

and \( |E[V_0(X_0)]| < \infty \). Then, an admissible control policy which achieves the minimum for every \((t, x)\) is optimal.

**Proof.** For any admissible control policy, using Itô’s rule for \( V(X_t) \) one obtains:

\[ \frac{E[V(X_0) - V(X_T^u)]}{T} + \eta = E\left[ \frac{1}{T} \int_0^T (\eta - \mathcal{L}^u V(X_s^u))ds \right] \]

Proceeding as before in Theorem 10.5.1 through the use of the Bellman equation leads to the desired result.

10.5.4 The convex analytic method

The analysis we made in Chapter 5 and 7 applies to the diffusion setting as well. In particular, a discounted HJB equation plays the role of the discounted cost optimality equation. For the average cost problems, one can apply either a vanishing discount approach or an convex-analytic approach. We refer the reader to [7], [29] and [112].

10.6 Stochastic Stability of Diffusions

Continuous-time counterparts of Foster-Lyapunov criteria considered in Chapters 3 and 4 exist and are well-developed. We refer the reader to [72], [80], [84]. Dynkin’s formula plays a key role in obtaining the counterparts using the martingale criteria leading to Foster-Lyapunov theorems. For functions \( V \) that are properly defined, as in the Foster-Lyapunov criteria studied in Chapter 4, conditions of the form

\[ AV(x) \leq b 1_{x \in S} \]
will lead to recurrence, positive Harris recurrence and finite expectations, respectively. However, the conditions needed on both $V$ and the Markov process need to be carefully addressed. For example, one needs to ensure that the processes are non-explosive, that is, they do not become unbounded in finite time; and one needs to establish conditions for the strong Markov property.

A useful technique in arriving at stochastic stability is to sample the process to obtain a discrete-time Markov chain, whose stability will imply the stability of the original process through a careful construction of invariant probability measures, similar to the sampled chains in Chapter 3. Furthermore, the analysis of small sets through sampled chains require a careful analysis.

Further stochastic stability results have found applications. We state a few below.

**Theorem 10.6.1** [72] [119] Suppose that there exists a function $V : \mathbb{R}^n \rightarrow \mathbb{R}_+$ which is $C^2$ and satisfies

$$AV(x) \leq -\alpha V(x) + b$$

for some $\alpha, b > 0$. Then,

$$E[V(X_t)] \leq e^{-\alpha t} E[V(X_0)] + \frac{b}{\alpha},$$

provided that $E[V(X_0)] < \infty$.

Let $X_0 = x^*$ be deterministic. We say that

- $x^*$ is stable if for every $\epsilon > 0$ and $\alpha \in (0, 1)$, there exists $\delta > 0$ such that

$$\|X_0 - x^*\| \leq \delta \implies P(\sup_{t \geq 0} \|X_t - x^*\| < \epsilon) > \alpha$$

- $x^*$ is asymptotically stable if it is stable and for every $\alpha \in (0, 1)$ there exists $\delta > 0$ such that whenever $\|X_0 - x^*\| < \delta$, $P(X_t \rightarrow x^*) > \alpha$

- $x^*$ is globally stable if it is stable and $X_t \rightarrow x^*$ almost surely for any $X_0$.

**Theorem 10.6.2** [72] [119]

(i) Let $V : \mathbb{R}^n \rightarrow \mathbb{R}_+$ be $C^2$ and satisfy $AV(x) \leq 0$ for all $x$ in some neighborhood of $x^*$, $V(x^*) = 0$ and $V(x) > 0$ if $x \neq x^*$. Then, $x^*$ is a stable equilibrium.

(ii) Let $V : \mathbb{R}^n \rightarrow \mathbb{R}_+$ be $C^2$ and satisfy $AV(x) < 0$ for all $x$ in some neighborhood of $x^*$, $V(x^*) = 0$ and $V(x) > 0$ if $x \neq x^*$. Then, $x^*$ is asymptotically stable.

- If in addition $V(x) \rightarrow \infty$ and $AV(x) \rightarrow \infty$ as $\|x\| \rightarrow \infty$, then $x^*$ is globally stable.

### 10.7 Exercises

**Exercise 10.7.1**

a) Solve

$$dX_t = \mu X_t + \sigma dB_t$$

Hint: Multiply both sides with the integrating factor $e^{-\mu t}$ and work with $d(e^{-\mu t} X_t)$.

b) Solve

$$dX_t = \mu dt + \sigma X_t dB_t$$

Hint: Multiply both sides with the integrating factor $e^{-\sigma B_t + \frac{1}{2} \sigma^2 t}$. 

Exercise 10.7.2 (White Noise Property of the Brownian Noise) Define the Fourier transform of $dB_t$ as:

$$a_k(\omega) = \int_0^1 d\Omega(\omega)e^{i2\pi k t} dt.$$ 

Show that $a_k, k \in \mathbb{Z}$ is Gaussian, and i.i.d. Now, compute $E[dB_t dB_s]$ and show that this expectation leads to a delta function, through the expression $\sum_k e^{-i2\pi k (t-s)}$, which leads to the delta function in the sense of distributions. Hint: To show the i.i.d. property, it suffices to show that $E[a_k a_j] = 0$ for $k \neq j$, since for Gaussian processes uncorrelatedness is equivalent to independence. This example reveals that even though $dB_t$ does not exists as a well-defined process, under the integral sign, meaningful properties can be exhibited.

Exercise 10.7.3 Complete the details for the solution to the optimal portfolio selection problem given in Example 10.11.

Exercise 10.7.4 Solve an average-cost version of the linear quadratic regulator problem and identify conditions on the cost function that leads to a cost that is independent of the initial condition.
Metric Spaces

Definition A.0.1 A linear space is a space which is closed under addition and multiplication by a scalar.

Definition A.0.2 A normed linear space $X$ is a linear vector space on which a functional (a mapping from $X$ to $\mathbb{R}$, that is a member of $\mathbb{R}^X$) called norm is defined such that:

- $||x|| \geq 0 \quad \forall x \in X$, $||x|| = 0$ if and only if $x$ is the null element (under addition and multiplication) of $X$.
- $||x + y|| \leq ||x|| + ||y||$
- $||\alpha x|| = ||\alpha|| ||x||, \quad \forall \alpha \in \mathbb{R}, \quad \forall x \in X$

Definition A.0.3 In a normed linear space $X$, an infinite sequence of elements $\{x_n\}$ converges to an element $x$ if the sequence $\{||x_n - x||\}$ converges to zero.

Definition A.0.4 A metric defined on a set $X$, is a function $d : X \times X \rightarrow \mathbb{R}$ such that:

- $d(x, y) \geq 0, \quad \forall x, y \in X$ and $d(x, y) = 0$ if and only if $x = y$.
- $d(x, y) = d(y, x), \quad \forall x, y \in X$.
- $d(x, y) \leq d(x, z) + d(z, y), \quad \forall x, y, z \in X$.

Definition A.0.5 A metric space $(X, d)$ is a set equipped with a metric $d$.

A normed linear space is also a metric space, with metric

$$d(x, y) = ||x - y||.$$

An important class of normed spaces that are widely used in optimization and engineering problems are Banach spaces:

A.0.1 Banach Spaces

Definition A.0.6 A sequence $\{x_n\}$ in a normed space $X$ is Cauchy if for every $\epsilon$, there exists an $N$ such that $||x_n - x_m|| \leq \epsilon$, for all $n, m \geq N$.

The important observation on Cauchy sequences is that, every converging sequence is Cauchy, however, not all Cauchy sequences are convergent: This is because the limit might not live in the original space where the sequence elements take values in. This brings the issue of completeness:
Definition A.0.7 A normed linear space $X$ is complete, if every Cauchy sequence in $X$ has a limit in $X$. A complete normed linear space is called Banach.

Banach spaces are important for many reasons including the following one: In optimization problems, sometimes we would like to see if a sequence converges, for example if a solution to a minimization problem exists, without knowing what the limit of the sequence could be. Banach spaces allow us to use Cauchy sequence arguments to claim the existence of optimal solutions. If time allows, we will discuss how this is used by using contraction and fixed point arguments for transformations.

An example is the following: consider the solutions to the equation $Ax = b$ for $A$ a square matrix and $b$ a vector. In class we will identify conditions on an iteration of the form $x_{k+1} = (I - A)x_k - b$ to form a Cauchy sequence and converge to a solution through the contraction principle.

In applications, we will also discuss completeness of a subset. A subset of a Banach space $X$ is complete if and only if it is closed. If it is not closed, one can provide a counterexample sequence which does not converge. If the set is closed, every Cauchy sequence in this set has a limit in $X$ and this limit should be a member of this set, hence the set is complete.

Exercise A.0.1 The space of bounded functions $\{ x : [0, 1] \to \mathbb{R}, \sup_{t \in [0,1]} |x(t)| < \infty \}$ is a Banach space.

The above space is also denoted by $L_{\infty}([0,1]; \mathbb{R})$ or $L_{\infty}([0,1])$.

Theorem A.0.1 $l_p(\mathbb{Z}_+; \mathbb{R}) := \{ x \in f(\mathbb{Z}_+; \mathbb{R}) : ||x||_p = \left( \sum_{i \in \mathbb{N}_+} |x(i)|^p \right)^{\frac{1}{p}} < \infty \}$ is a Banach space for all $1 \leq p \leq \infty$.

Sketch of Proof: The proof is completed in three steps.

(i) Let $\{x_n\}$ be Cauchy. This implies that for every $\epsilon > 0$, $\exists N$ such that for all $n, m \geq N \|x_n - x_m\| \leq \epsilon$. This also implies that for all $n > N$, $\|x_n\| \leq \|x_N\| + \epsilon$. Now let us denote $x_n$ by the vector $\{x_1^n, x_2^n, x_3^n, \ldots\}$. It follows that for every $k$ the sequence $\{x^n_k\}$ is also Cauchy. Since $x^n_k \in \mathbb{R}$, and $\mathbb{R}$ is complete, $x^n_k \to x_k$ for some $x_k$. Thus, the sequence $x_n$ pointwise converges to some vector $x_*$.

(ii) Is $x \in l_p(\mathbb{Z}_+; \mathbb{R})$? Define $x_{n,K} = \{x_1^n, x_2^n, \ldots, x_{K-1}^n, x_K^n, 0, 0, \ldots\}$, that is vector which truncates after the $K$th coordinate. Now, it follows that

$$\|x_{n,K}\| \leq \|x_N\| + \epsilon,$$

for every $n \geq N$ and $K$ and

$$\lim_{n \to \infty} \|x_{n,K}\|^p = \lim_{n \to \infty} \sum_{i=1}^{K} |x_i^n|^p = \sum_{i=1}^{K} |x_i|^p,$$

since there are only finitely many elements in the summation. The question now is whether $\|x_\infty\| \in l_p(\mathbb{Z}_+; \mathbb{R})$. Now,

$$\|x_{n,K}\| \leq \|x_N\| + \epsilon,$$

and thus

$$\lim_{n \to \infty} \|x_{n,K}\| = \|x_K\| \leq \|x_N\| + \epsilon,$$

Let us take another limit, by the monotone convergence theorem (Recall that this theorem says that a monotonically increasing sequence which is bounded has a limit).

$$\lim_{K \to \infty} \|x_{*,K}\|^p = \lim_{K \to \infty} \sum_{i=1}^{K} |x_i|^p = ||x_\infty||_p^p \leq \|x_N\| + \epsilon.$$

(iii) The final question is: Does $\|x_n - x_*\| \to 0$? Since the sequence is Cauchy, it follows that for $n, m \geq N$

$$\|x_n - x_m\| \leq \epsilon.$$
Thus, \[ ||x_nK - x_mK|| \leq \epsilon \]
and since \( K \) is finite
\[ \lim_{m \to \infty} ||x_nK - x_mK|| = ||x_nK - x_sK|| \leq \epsilon \]
Now, we take another limit
\[ \lim_{K \to \infty} ||x_nK - x_sK|| \leq \epsilon \]
By the monotone convergence theorem again,
\[ \lim_{K \to \infty} ||x_nK - x_sK|| = ||x_n - x|| \leq \epsilon \]
Hence, \( ||x_n - x|| \to 0 \).

The above spaces are also denoted \( l_p(\mathbb{Z}_+) \), when the range space is clear from context.

The following is a useful result.

**Theorem A.0.2 (Hölder’s Inequality)**
\[ \sum x(t)y(t) \leq ||x||_p||y||_q, \]
with \( 1/p + 1/q = 1 \) and \( 1 \leq p, q \leq \infty \).

**Remark:** A brief remark for notations: When the range space is \( \mathbb{R} \), the notation \( l_p(\Omega) \) denotes \( l_p(\Omega; \mathbb{R}) \) for a discrete-time index set \( \Omega \) and likewise for a continuous-time index set \( \Omega \), \( L_p(\Omega) \) denotes \( L_p(\Omega; \mathbb{R}) \).

### A.0.2 Hilbert Spaces

We first define pre-Hilbert spaces.

**Definition A.0.8** A pre-Hilbert space \( X \) is a linear vector space where an inner product is defined on \( X \times X \). Corresponding to each pair \( x, y \in X \) the inner product \( \langle x, y \rangle \) is a scalar (that is real-valued or complex-valued). The inner product satisfies the following axioms:

1. \( \langle x, y \rangle = \langle y, x \rangle^* \) (the superscript denotes the complex conjugate) (we will also use \( \overline{\langle y, x \rangle} \) to denote the complex conjugate)
2. \( \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \)
3. \( \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \)
4. \( \langle x, x \rangle \geq 0 \), equals 0 iff \( x \) is the null element.

The following is a crucial result in such a space, known as the Cauchy-Schwarz inequality, the proof of which was presented in class:

**Theorem A.0.3** For \( x, y \in X \),
\[ \langle x, y \rangle \leq \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle}, \]
where equality occurs if and only if \( x = \alpha y \) for some scalar \( \alpha \).

**Exercise A.0.2** In a pre-Hilbert space \( \langle x, x \rangle \) defines a norm: \( ||x|| = \sqrt{\langle x, x \rangle} \)
The proof for the result requires one to show that $\sqrt{\langle x, x \rangle}$ satisfies the triangle inequality, that is

$$||x + y|| \leq ||x|| + ||y||,$$

which can be proven by an application of the Cauchy-Schwarz inequality.

Not all spaces admit an inner product. In particular, however, $l_2(N_+; \mathbb{R})$ admits an inner product with

$$\langle x, y \rangle = \sum_{t \in N_+} x(n)y(n)$$

for $x, y \in l_2(N_+; \mathbb{R})$. Furthermore, $||x|| = \sqrt{\langle x, x \rangle}$ defines a norm in $l_2(N_+; \mathbb{R})$.

The inner product, in the special case of $\mathbb{R}^N$, is the usual inner vector product; hence $\mathbb{R}^N$ is a pre-Hilbert space with the usual inner-product.

**Definition A.0.9** A complete pre-Hilbert space, is called a Hilbert space.

Hence, a Hilbert space is a Banach space, endowed with an inner product, which induces its norm.

**Proposition A.0.1** The inner product is continuous: if $x_n \to x$, and $y_n \to y$, then $\langle x_n, y_n \rangle \to \langle x, y \rangle$ for $x_n, y_n$ in a Hilbert space.

**Proposition A.0.2** In a Hilbert space $X$, two vectors $x, y \in X$ are orthogonal if $\langle x, y \rangle = 0$. A vector $x$ is orthogonal to a set $S \subset X$ if $\langle x, y \rangle = 0$ $\forall y \in S$.

**Theorem A.0.4 (Projection Theorem:)** Let $H$ be a Hilbert space and $B$ a closed subspace of $H$. For any vector $x \in H$, there is a unique vector $m \in B$ such that

$$||x - m|| \leq ||x - y||$, $\forall y \in B.$$

A necessary and sufficient condition for $m \in B$ to be the minimizing element in $B$ is that, $x - m$ is orthogonal to $B$.

**A.0.3 Separability**

**Definition A.0.10** Given a normed linear space $X$, a subset $D \subset X$ is dense in $X$, if for every $x \in X$, and each $\epsilon > 0$, there exists a member $d \in D$ such that $||x - d|| \leq \epsilon$.

**Definition A.0.11** A set is countable if every element of the set can be associated with an integer via an ordered mapping.

Examples of countables spaces are finite sets and the set $\mathbb{Q}$ of rational numbers. An example of uncountable sets is the set $\mathbb{R}$ of real numbers.

**Theorem A.0.5** a) A countable union of countable sets is countable. b) Finite Cartesian products of countable sets is countable. c) Infinite Cartesian products of countable sets may not be countable. d) $[0, 1]$ is not countable.

Cantor’s diagonal argument and the triangular enumeration are important steps in proving the theorem above.

Since rational numbers are the ratios of two integers, one may view rational numbers as a subset of the product space of countable spaces; thus, rational numbers are countable.

**Definition A.0.12** A space $X$ is separable, if it contains a countable dense set.

Separability basically tells that it suffices to work with a countable set, when the set is uncountable. Examples of separable sets are $\mathbb{R}$, and the set of continuous and bounded functions on a compact set metrised with the maximum distance between the functions.
On the Convergence of Random Variables

B.1 Limit Events and Continuity of Probability Measures

Given \( A_1, A_2, \ldots, A_n \in \mathcal{F} \), define:
\[
\limsup_n A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k
\]
\[
\liminf_n A_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k
\]

For the superior limit, an element is in this set, if it is in infinitely many \( A_n \)s. For the inferior case, an element is in the limit, if it is in almost except for a finite number of \( A_n \)s. The limit of a sequence of sets exists if the above limits are equal.

We have the following result:

**Theorem B.1.1** For a sequence of events \( A_n \):
\[
P\left(\liminf_n A_n\right) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P\left(\limsup_n A_n\right)
\]

We have the following regarding continuity of probability measures:

**Theorem B.1.2** (i) For a sequence of events \( A_n, A_n \subset A_{n+1} \),
\[
\lim_{n \to \infty} P(A_n) = P(\bigcup_{n=1}^\infty A_n)
\]

(ii) For a sequence of events \( A_n, A_{n+1} \subset A_n \),
\[
\lim_{n \to \infty} P(A_n) = P(\bigcap_{n=1}^\infty A_n)
\]

B.2 Borel-Cantelli Lemma

**Theorem B.2.1** (i) If \( \sum_n P(A_n) \) converges, then \( P(\limsup_n A_n) = 0 \). (ii) If \( \{A_n\} \) are independent and if \( \sum P(A_n) = \infty \), then \( P(\limsup_n A_n) = 1 \).

**Exercise B.2.1** Let \( \{A_n\} \) be a sequence of independent events where \( A_n \) is the event that the \( n \)th coin flip is head. What is the probability that there are infinitely many heads if \( P(A_n) = 1/n^2 \)?

An important application of the above is the following:
Theorem B.2.2 Let \( Z_n, n \in \mathbb{N} \) and \( Z \) be random variables and for every \( \epsilon > 0 \),
\[
\sum_{n} P(|Z_n - Z| \geq \epsilon) < \infty.
\]
Then,
\[
P(\{\omega : Z_n(\omega) = Z(\omega)\}) = 1.
\]
That is \( Z_n \) converges to \( Z \) with probability 1.

B.3 Convergence of Random Variables

B.3.1 Convergence almost surely (with probability 1)

Definition B.3.1 A sequence of random variables \( X_n \) converges almost surely to a random variable \( X \) if \( P(\{\omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}) = 1 \).

B.3.2 Convergence in Probability

Definition B.3.2 A sequence of random variables \( X_n \) converges in probability to a random variable \( X \) if \( \lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0 \) for every \( \epsilon > 0 \).

B.3.3 Convergence in Mean-square

Definition B.3.3 A sequence of random variables \( X_n \) converges in the mean-square sense to a random variable \( X \) if \( \lim_{n \to \infty} E[|X_n - X|^2] = 0 \).

B.3.4 Convergence in Distribution

Definition B.3.4 Let \( X_n \) be a random variable with cumulative distribution function \( F_n \), and \( X \) be a random variable with cumulative distribution function \( F \). A sequence of random variables \( X_n \) converges in distribution (or weakly) to a random variable \( X \) if \( \lim_{n \to \infty} F_n(x) = F(x) \) for all points of continuity of \( F \).

Theorem B.3.1 a) Convergence in almost sure sense implies in probability. b) Convergence in mean-square sense implies convergence in probability. c) If \( X_n \to X \) in probability, then \( X_n \to X \) in distribution.

We also have partial converses for the above results:

Theorem B.3.2 a) If \( P(|X_n| \leq Y) = 1 \) for some random variable \( Y \) with \( E[Y^2] < \infty \), and if \( X_n \to X \) in probability, then \( X_n \to X \) in mean-square. b) If \( X_n \to X \) in probability, there exists a subsequence \( X_{n_k} \) which converges to \( X \) almost surely. c) If \( X_n \to X \) and \( X_n \to Y \) in probability, mean-square or almost surely. Then \( P(X = Y) = 1 \).

A sequence of random variables is uniformly integrable if:
\[
\lim_{K \to \infty} \sup_n E[X_n 1_{|X_n| \geq K}] = 0.
\]
Note that, if \( \{X_n\} \) is uniformly integrable, then \( \sup_n E[|X_n|] < \infty \).
**Theorem B.3.3** Under uniform integrability, convergence in almost sure sense implies convergence in mean-square.

**Theorem B.3.4** If $X_n \rightarrow X$ in probability, there exists some subsequence $X_{n_k}$ which converges to $X$ almost surely.

**Theorem B.3.5 (Skorohod’s representation theorem)** Let $X_n \rightarrow X$ in distribution. Then, there exists a sequence of random variables $Y_n$ and $Y$ such that, $X_n$ and $Y_n$ have the same cumulative distribution functions; $X$ and $Y$ have the same cumulative distribution functions and $Y_n \rightarrow Y$ almost surely.

With the above, we can prove the following result.

**Theorem B.3.6** The following are equivalent: i) $X_n$ converges to $X$ in distribution. ii) $E[f(X_n)] \rightarrow E[f(X)]$ for all continuous and bounded functions $f$. iii) The characteristic functions $\Phi_n(u) := E[e^{iuX_n}]$ converge pointwise for every $u \in \mathbb{R}$. 
Some Remarks on Measurable Selections

As we observe in Chapter 5, in stochastic control measurability issues arise extensively both for the measurability of control policies as well as that of value functions/optimal costs.

Theorem 5.1.1 and Theorem 5.2.3 are two examples where these were crucially utilized. In addition, we observed that the theory of martingales and filtration, the measurability properties are essential.

One particular aspect is to ensure that maps of the form:

$$J(x) := \inf_{u \in U} c(x,u) \quad \text{(C.1)}$$

are measurable or at least Lebesgue-integrable.

**Theorem C.0.1 (Kuratowski Ryll-Nardzewski Measurable Selection Theorem)\textsuperscript{[71]}** Let $X, U$ be Polish spaces and $\Gamma = (x, \psi(x))$ where $\psi(x) \subset U$ be such that, $\psi(x)$ is closed for each $x \in X$ and $\Gamma$ be a Borel measurable set in $X \times U$. Then, there exists at least one measurable function $f : X \rightarrow U$ such that $\{(x, f(x)), x \in X\} \subset \Gamma$.

This result was utilized in Chapter 5.

A function $f$ is $\mu$-measurable if there exists a Borel measurable function $g$ which agrees with $f$ $\mu$-a.e. A function that is $\mu$-measurable for every probability measure is called universally measurable.

**Fact C.0.1** The image of a Borel set under a measurable, or even a continuous, function may not be measurable.

The integral of a universally measurable function is well-defined and is equal to the integral of a Borel measurable function which is $\mu$-almost equal to that function. While applying dynamic programming, we often seek to establish the existence of measurable functions through the operation:

$$J_t(x_t) = \inf_{u \in U(x_t)} \left( c(x,u) + \int J_{t+1}(x_{t+1})Q(dx_{t+1}|x_t,u) \right)$$

However, we need a stronger condition that universal measurability for the recursions to be well-defined. A function $f$ is called lower semi-analytic if $\{x : f(x) < c\}$ is analytic for each scalar $c$.

**Theorem C.0.2\textsuperscript{[45]}** Let $i : X \rightarrow 2^\mathbb{S}$ (that is, $i$ maps $X$ to subsets of $\mathbb{S}$) be such that $i^{-1}$ is Borel measurable, and $f : \mathbb{S} \rightarrow \mathbb{R}$ be measurable. Then:

$$v(x) = \inf_{z : z \in i(x)} f(z)$$

is lower semi-analytic.

Observe that (see p. 85 of \textsuperscript{[45]})
The set \( \{ z : f(z) < c \} \) is Borel, and thus if \( i^{-1} \) is also Borel, it follows that \( v \) is lower semi-analytic. We require then that \( i^{-1} : S \to X \) to be Borel.

**Theorem C.0.3** Lower semi-analytic functions are universally measurable.

**Theorem C.0.4** Consider \( G = \{ (x, u) : u \in U(x) \} \) which is a Borel measurable set. The map,

\[
v(x) = \inf_{(x, z) \in G} v(x, z),
\]

is lower semi-analytic (and thus universally measurable).

**Proof.** The graph \( G \) is measurable. It follows that \( \{ x : v(x) < c \} = i^{-1}(\{ (x, z) : v(x, z) < c \}) \), where \( i^{-1} \) is the projection of \( G \) onto \( X \), which is a continuous operation; the image may not be measurable but as a measurable mapping of a Borel set, it is analytic. As a result \( v \) is lower semi-analytic. \( \Box \)

Implication: Dynamic programming can be carried out for such expressions. In particular, the following is due to Bertsekas and Shreve (see Chapter 7):

**Theorem C.0.5**

(i) Let \( E_1, E_2 \) be Borel and \( g : E_1 \times E_2 \to \mathbb{R} \) be lower semi-analytic. Then,

\[
h(e_1) = \inf_{e_2 \in E_2} g(e_1, e_2)
\]

is lower semi-analytic.

(ii) Let \( E_1, E_2 \) be Borel and \( g : E_1 \times E_2 \to \mathbb{R} \) be lower semi-analytic. Let \( Q(de_2|e_1) \) be a stochastic kernel. Then,

\[
f(e_1) := \int g(e_2) Q(de_2|e_1)
\]

is lower semi-analytic.

We note that the second result would not be correct if \( g \) is only taken to be universally measurable. The result above ensures that we can follow the dynamic programming arguments in an inductive manner under conditions that are far less restrictive than the conditions stated in the measurable selection conditions.

Building on this discussion, and the material in Chapter 5, we summarize three useful results in the following.

**Fact C.0.2** Consider (C.1).

(i) If \( c \) is continuous on \( X \times U \) and \( U \) is compact, then \( J \) is continuous.

(ii) If \( c \) is continuous on \( U \) for every \( x \), and \( U \) is compact, then \( J \) is measurable (Prop D.5 in [59] and Himmelberg and Schäl [109]); see Theorem 5.2.3.

(iii) If \( c \) is measurable on \( X \times U \) and \( U \) is Borel, then \( J \) is lower semi-analytic.

Compactness of \( U \) is a crucial component for these results. However, \( U(x) \) may be allowed to depend on \( x \), for item (ii) under the assumption that the graph \( G = \{ (x, u) : u \in U(x) \} \) defined above is Borel, and \( U(x) \) is compact for every \( x \); see p. 182 in [59] (and also [66], [109] and [71], among others).
On Spaces of Probability Measures

Let $X$ be a Polish space and let $P(X)$ denote the family of all probability measures on $(X, \mathcal{B}(X))$. Let $\{\mu_n, n \in \mathbb{N}\}$ be a sequence in $P(X)$.

A sequence $\{\mu_n\}$ is said to converge to $\mu \in P(X)$ weakly if

$$\int_{\mathbb{R}^N} c(x)\mu_n(dx) \to \int_{\mathbb{R}^N} c(x)\mu(dx)$$

for every continuous and bounded $c : X \to \mathbb{R}$.

On the other hand, $\{\mu_n\}$ is said to converge to $\mu \in P(X)$ setwise if

$$\int_X c(x)\mu_n(dx) \to \int_X c(x)\mu(dx)$$

for every measurable and bounded $c : X \to \mathbb{R}$. Setwise convergence can also be defined through pointwise convergence on Borel subsets of $X$ (see, e.g., [61]), that is

$$\mu_n(A) \to \mu(A), \quad \text{for all } A \in \mathcal{B}(X)$$

since the space of simple functions are dense in the space of bounded and measurable functions under the supremum norm.

For two probability measures $\mu, \nu \in P(X)$, the total variation metric is given by

$$\|\mu - \nu\|_{TV} := 2 \sup_{B \in \mathcal{B}(X)} |\mu(B) - \nu(B)|$$

for every measurable and bounded $c : X \to \mathbb{R}$ such that $\|f\|_{\infty} = \sup_{x \in X}|f(x)| \leq 1$. A sequence $\{\mu_n\}$ is said to converge to $\mu \in P(X)$ in total variation if $\|\mu_n - \mu\|_{TV} \to 0$.

Setwise convergence is equivalent to pointwise convergence on Borel sets whereas total variation requires uniform convergence on Borel sets. Thus these three convergence notions are in increasing order of strength: convergence in total variation implies setwise convergence, which in turn implies weak convergence.

\footnote{It is important to emphasize that what typically is studied in probability as weak convergence is not the exact weak convergence notion used in functional analysis: The topological dual space of the set of probability measures does not only consist of expectations of continuous and bounded functions. However, the dual space of the space of continuous and bounded functions with the supremum norm does admit a representation in terms of expectations \cite{78}; hence, the weak convergence here is in actuality the weak" convergence in analysis and distribution theory.}
On the other hand, total variation is a stringent notion for convergence. For example a sequence of discrete probability measures never converges in total variation to a probability measure which admits a density function with respect to the Lebesgue measure and such a space is not separable. Setwise convergence also induces a topology on the space of probability measures and channels which is not easy to work with since the space under this convergence is not metrizable (53, p. 59).

However, the space of probability measures on a complete, separable, metric (Polish) space endowed with the topology of weak convergence is itself a complete, separable, metric space [19]. The Prohorov metric, for example, can be used to metrize this space. Since the space of continuous functions on a compact set is separable, one could obtain a countable collection of semi-norms which can then be used to define a metric on this space. As a more practical metric, the Wasserstein metric can also be used (for compact ).

**Definition D.0.1 (Wasserstein metric)** The Wasserstein metric of order 1 for two distributions \( \mu, \nu \in \mathcal{P}(\mathbb{X}) \) is defined as

\[
W_1(\mu, \nu) = \inf_{\eta \in \mathcal{H}(\mu, \nu)} \int_{\mathbb{X} \times \mathbb{X}} \eta(x, y) \| x - y \|, \\
\]

where \( \mathcal{H}(\mu, \nu) \) denotes the set of probability measures on \( \mathbb{X} \times \mathbb{X} \) with first marginal \( \mu \) and second marginal \( \nu \) and \( \| \cdot \| \) is a norm (such as the Euclidean norm).

As noted, for compact \( \mathbb{X} \), the Wasserstein distance of order 1 metrizes the weak topology on the set of probability measures on \( \mathbb{X} \) (see [123, Theorem 6.9]). For non-compact \( \mathbb{X} \), weak convergence combined with convergence of moments (that is of \( \int \mu_n(x)|x| \to \int \mu(dx)|x| \)) is equivalent to convergence in \( W_1 \). Finally, the bounded-Lipschitz metric \( \rho_{BL} \) [123, p.109] can also be used to metrize weak convergence:

\[
\rho_{BL}(\mu, \nu) = \sup_{\|f\|_{BL} \leq 1} \left| \int_{\mathbb{X}} f(x) \mu(dx) - \int_{\mathbb{X}} f(x) \nu(dx) \right|, \\
\]

where

\[
\|f\|_{BL} = \|f\|_\infty + \sup_{e \neq e'} \frac{f(e) - f(e')}{d_{\mathbb{X}}(e, e')},
\]

and \( d_{\mathbb{X}} \) is the metric on \( \mathbb{X} \).

Weak convergence is very important in applications of stochastic control and probability in general. Prohorov’s theorem [43] provides a way to characterize compactness properties under weak convergence. Furthermore, such a convergence notion has very important measurability properties; recall the following from Chapter 5:

**Theorem D.0.1** Let \( \mathbb{S} \) be a Polish space and \( M \) be the set of all measurable and bounded functions \( f : \mathbb{S} \to \mathbb{R} \). Then, for any \( f \in M \), the integral

\[
\int \pi(dx) f(x)
\]

defines a measurable function on \( \mathcal{P}(\mathbb{S}) \) under the topology of weak convergence.

This is a useful result since it allows us to define measurable functions in integral forms on the space of probability measures when we work with the topology of weak convergence. The second useful result follows from Theorem 6.3.1 and Theorem 2.1 of Dubins and Freedman [42] and Proposition 7.25 in Bertsekas and Shreve [17].

**Theorem D.0.2** Let \( \mathbb{S} \) be a Polish space. A function \( F : \mathcal{P}(\mathbb{S}) \to \mathcal{P}(\mathbb{S}) \) is measurable on \( \mathcal{B}(\mathcal{P}(\mathbb{S})) \) (under weak convergence), if for all \( B \in \mathcal{B}(\mathbb{S}) \) (\( (F(\cdot))(B) : \mathcal{P}(\mathbb{S}) \to \mathbb{R} \) is measurable under weak convergence on \( \mathcal{P}(\mathbb{S}) \), that is for every \( B \in \mathcal{B}(\mathbb{S}) \), \( (F(\pi))(B) \) is a measurable function when viewed as a function from \( \mathcal{P}(\mathbb{S}) \) to \( \mathbb{R} \).
References


