Chapter 2
Theory of Finite Horizon Markov Decision Processes

In this chapter we will establish the theory of Markov Decision Processes with a finite time horizon and with general state and action spaces. Optimization problems of this kind can be solved by a backward induction algorithm. Since state and action space are arbitrary, we will impose a structure assumption on the problem in order to prove the validity of the backward induction and the existence of optimal policies. The chapter is organized as follows.
Section 2.1 provides the basic model data and the definition of policies. The precise mathematical model is then presented in Section 2.2 along with a sufficient integrability assumption which implies a well-defined problem. The solution technique for these problems is explained in Section 2.3. Under structure assumptions on the model it will be shown that Markov Decision Problems can be solved recursively by the so-called *Bellman equation*. The next section summarizes a number of important special cases in which the structure assumption is satisfied. Conditions on the model data are given such that the value functions are upper semicontinuous, continuous, measurable, increasing, concave or convex respectively. Also the monotonicity of the optimal policy under some conditions is established. This is an essential property for computations. Finally the important concept of upper bounding functions is introduced in this section. Whenever an upper bounding function for a Markov Decision Model exists, the integrability assumption is satisfied. This concept will be very fruitful when dealing with infinite horizon Markov Decision Problems in Chapter 7. In Section 2.5 the important case of stationary Markov Decision Models is investigated. The notion 'stationary' indicates that the model data does not depend on the time index. The relevant theory is here adopted from the non-stationary case. Finally Section 2.6 highlights the application of the developed theory by investigating three simple examples. The first example is a special card game, the second one a cash balance problem and the last one deals with the classical stochastic LQ-problems. The last section contains some notes and references.

N. Bäuerle and U. Rieder, *Markov Decision Processes with Applications to Finance*, Universitext, DOI 10.1007/978-3-642-18324-9_2,

2.1 Markov Decision Models

After having discussed the scope of Markov Decision Models informally in Chapter 1 we will now give a precise definition of a Markov Decision Model. This can be done by defining the ingredients or input data of the model in mathematical terms.

Definition 2.1.1. A (non-stationary) *Markov Decision Model* with planning horizon $N \in \mathbb{N}$ consists of a set of data $(E, A, D_n, Q_n, r_n, g_N)$ with the following meaning for $n = 0, 1, \ldots, N-1$:

- E is the *state space*, endowed with a σ-algebra $\mathfrak{E}$. The elements (states) are denoted by $x \in E$.
- A is the *action space*, endowed with a σ-algebra $\mathfrak{A}$. The elements (actions) are denoted by $a \in A$.
- $D_n \subset E \times A$ is a measurable subset of $E \times A$ and denotes the set of possible state-action combinations at time n. We assume that D_n contains the graph of a measurable mapping $f_n : E \to A$, i.e. $(x, f_n(x)) \in D_n$ for all $x \in E$. For $x \in E$, the set $D_n(x) = \{a \in A \mid (x, a) \in D_n\}$ is the set of *admissible actions* in state x at time n.
- Q_n is a stochastic transition kernel from D_n to E, i.e. for any fixed pair $(x, a) \in D_n$, the mapping $B \mapsto Q_n(B|x, a)$ is a probability measure on $\mathfrak{E}$ and $(x, a) \mapsto Q_n(B|x, a)$ is measurable for all $B \in \mathfrak{E}$. The quantity $Q_n(B|x, a)$ gives the probability that the next state at time $n+1$ is in B if the current state is x and action a is taken at time n. Q_n describes the *transition law*.
- $r_n : D_n \to \mathbb{R}$ is a measurable function. $r_n(x, a)$ gives the (discounted) *one-stage reward* of the system at time n if the current state is x and action a is taken.
- $g_N : E \to \mathbb{R}$ is a measurable mapping. $g_N(x)$ gives the (discounted) *terminal reward* of the system at time N if the state is x.

Remark 2.1.2. a) In many applications the state and action spaces are Borel subsets of Polish spaces (i.e. complete, separable, metric spaces) or finite or countable sets. The σ-algebras $\mathfrak{E}$ and $\mathfrak{A}$ are then given by the σ-algebras $\mathcal{B}(E)$ and $\mathcal{B}(A)$ of all Borel subsets of E and A respectively. Often in applications E and A are subsets of $\mathbb{R}^d$ or $\mathbb{R}^d_+$.

b) If the one-stage reward function r'_n also depends on the next state, i.e. $r'_n = r'_n(x, a, x')$, then define

$$r_n(x, a) := \int r'_n(x, a, x') Q_n(dx'|x, a).$$

c) Often D_n and Q_n are independent of n and $r_n(x, a) := \beta^n r(x, a)$ and $g_N(x) := \beta^N g(x)$ for a (discount) factor $\beta \in (0, 1]$. In this case the Markov Decision Model is called *stationary* (see Section 2.5). ◊