Section 11: Stochastic Control

Contents

11.1 Deriving Linear Systems via Martingale Methods

We have described earlier in the course the use of "first transition analysis" to derive the linear systems associated with computing certain probabilities and expectations. We now use martingale methods to both re-derive these linear systems, and analyze their uniqueness properties.

To fix the ideas, consider the specific problem of computing

$$u^*(x) \triangleq \mathrm{E}_x \sum_{j=0}^{T-1} r(X_j),$$

where $r : S \to \mathbb{R}_+$ and $T = \inf\{n \geq 0 : X_n \in C^c\}$ with $C^c \neq \emptyset$. Suppose that $u^*(x) < \infty$; then

$$Y = \sum_{j=0}^{T-1} r(X_j)$$

is a P_x-integrable rv (i.e. $\mathrm{E}_x|Y| < \infty$). We now take advantage of the following general result.

Proposition 11.1.1 Let W be an integrable rv, and $(Z_j : j \geq 0)$ be an arbitrary sequence of rv's. Then,

$$M_n \triangleq \mathrm{E}[W|Z_0, \ldots, Z_n]$$

is a martingale adapted to $(Z_n : n \geq 0)$.

Proof: The only non-trivial issue to verify is that

$$\mathrm{E}[M_{n+1}|Z_0, \ldots, Z_n] = M_n.$$

But

$$\mathrm{E}[\mathrm{E}[W|Z_0, \ldots, Z_{n+1}]|Z_0, \ldots, Z_n] = \mathrm{E}[W|Z_0, \ldots, Z_n];$$

this is the well-known *tower property* of conditional expectations. □

Since $u^*(x) < \infty$, evidently Y is P_x-integrable, so that Proposition 11.1.1 applies, yielding the fact that

$$M_n = \mathrm{E}_x[Y|X_0, \ldots, X_n]$$

is a P_x-martingale adapted to $(X_n : n \geq 0)$. But

$$\mathrm{E}_x\left[\sum_{j=0}^{T-1} r(X_j)|X_0, \ldots, X_n\right] = \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)u^*(X_n),$$

and hence we conclude that

$$M_n = \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)u^*(X_n)$$

is a P_x-martingale (adapted to $(X_n : n \geq 0)$). Consequently,

$$\mathrm{E}_x[M_n|X_0, \ldots, X_{n-1}] = M_{n-1}, \quad \mathrm{P}_x \text{ a.s.} \tag{11.1.1}$$

But

$$\begin{aligned}
&\mathrm{E}_x\left[\sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)u^*(X_n)|X_0, \ldots, X_{n-1}\right] \\
&= \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)\mathrm{E}_x[u^*(X_n)|X_0, \ldots, X_{n-1}] \\
&= \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)(Pu^*)(X_{n-1})
\end{aligned}$$

and

$$M_{n-1} = \sum_{j=0}^{(T\wedge(n-1))-1} r(X_j) + I(T \geq n-1)u^*(X_{n-1}).$$

In order that (11.1.1) be valid, it follows that on $\{T > n-1\}$,

$$r(X_{n-1}) + (Pu^*)(X_{n-1}) = u^*(X_{n-1}) \quad \mathrm{P}_x \text{ a.s.} \tag{11.1.2}$$

But $X_{n-1} \in C$ on $\{T > n-1\}$, so one way to guarantee the validity of (11.1.2) is to demand that

$$u^*(y) = r(y) + (Pu^*)(y) \tag{11.1.3}$$

for $y \in C$. Of course, it is evident that $u^*(y) = 0$ for $y \in C^c$. The linear system (11.1.3) is precisely the linear system for u^* derived earlier in the course via first-transition analysis.

We now turn to analysis of the linear system

$$u = r + Pu \quad \text{on } C, \tag{11.1.4}$$

subject to $u = 0$ on C^c, using martingale methods. Consider

$$\widetilde{M}_n = \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n)u(X_n)$$

and assume that $(\widetilde{M}_n : n \geq 0)$ is a P_x-integrable sequence of rv's. Since (11.1.4) guarantees that

$$r(X_{n-1}) + (Pu)(X_{n-1}) = u(X_{n-1})$$

on $\{T > n-1\}$, it is trivial to see that $(\widetilde{M}_n : n \geq 0)$ is now a P_x-martingale adapted to $(X_n : n \geq 0)$. It follows that

$$\mathrm{E}_x \widetilde{M}_n = \mathrm{E}_x \widetilde{M}_0,$$

and hence

$$\mathrm{E}_x \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n) u(X_n) = u(x).$$

Since

$$\mathrm{E}_x \sum_{j=0}^{(T\wedge n)-1} r(X_j) \nearrow \mathrm{E}_x \sum_{j=0}^{T-1} r(X_j)$$

by the Monotone Convergence Theorem, we can conclude that $u(x) = u^*(x)$, provided that

$$\mathrm{E}_x[u(X_n) I(T \geq n)] \to 0$$

as $n \to \infty$. We now have the following result, proved using martingale methods.

Proposition 11.1.2 Suppose that u satisfies (11.1.4) subject to $u = 0$ on C^c and that:

- $(\widetilde{M}_n : n \geq 0)$ is a P_x-integrable sequence of rv's
- $\mathrm{E}_x[u(X_n) I(T \geq n)] \to 0$ as $n \to \infty$.

Then, $u(x) = u^*(x)$.

Remark 11.1.1 If r and u are bounded and $T < \infty$ P_x-a.s., then $(\widetilde{M}_n : n \geq 0)$ is a P_x-integrable sequence of rv's and $\mathrm{E}_x[u(X_n) I(T \geq n)] \to 0$ as $n \to \infty$, so $u(x) = u^*(x)$.

Remark 11.1.2 If $u \geq 0$, the martingale convergence theorem guarantees that either $T < \infty$ P_x-a.s. or $u(X_n) \to 0$ P_x-a.s. on $\{T = \infty\}$.

This identical approach can be used to derive the appropriate linear systems satisfied by all the probabilities and expectations analyzed earlier in the course using first-transition analysis. As just indicated, this martingale approach allows one to also characterize the probabilistically meaningful solution of the linear system.

11.2 Lyapunov Bounds: A Super-martingale Perspective

For a given $g : S \to \mathbb{R}_+$, consider

$$\underset{\sim}{M}_n = \sum_{j=0}^{(T\wedge n)-1} r(X_j) + I(T \geq n) g(X_n)$$

and assume that $(\underset{\sim}{M}_n : n \geq 0)$ is a super-martingale adapted to $(X_n : n \geq)$. Then, the super-martingale property implies that

$$E_x \underset{\sim}{M}_n \leq E_x \underset{\sim}{M}_0$$

for $n \geq 0$, so

$$E_x \sum_{j=0}^{(T \wedge n)-1} r(X_j) + I(T \geq n) g(X_n) \leq g(x).$$

Since g is non-negative, we find that

$$E_x \sum_{j=0}^{(T \wedge n)-1} r(X_j) \leq g(x),$$

so that the Monotone Convergence Theorem allows us to conclude that

$$E_x \sum_{j=0}^{T-1} r(X_j) \leq g(x).$$

Thus, we end up with an upper bound on $u^*(x)$, provided that g is chosen so that $(\underset{\sim}{M}_n : n \geq 0)$ is a super-martingale.

We now need to identify conditions on g so that $(\underset{\sim}{M}_n : n \geq 0)$ is a P_x-super-martingale. As in Section 11.1, this reduces to

$$(Pg)(X_{n-1}) + r(X_{n-1}) \leq g(X_{n-1}) \quad P_x - \text{a.s.}$$

on $\{T > n-1\}$. A sufficient (and close to necessary) condition is that

$$(Pg)(x) + r(x) \leq g(x)$$

for $x \in C$. In other words,

$$Pg \leq g - r \tag{11.2.1}$$

on C. But (11.2.1) is exactly the Lyapunov inequality discussed earlier in the course for bounding the minimal non-negative solution to

$$u = r + Pu.$$

To rigorously verify that any non-negative finite-valued solution to (11.2.1) does indeed produce a super-martingale $(\underset{\sim}{M}_n : n \geq 0)$ yielding the bound $u^*(x) \leq g(x)$, the only real remaining step to verify is that $(\underset{\sim}{M}_n : n \geq 0)$ is indeed automatically a P_x-integrable sequence of rv's. But this follows from (11.2.1) via an easy induction. Because the relation (11.2.1) is exactly what is needed to make $(\underset{\sim}{M}_n : n \geq 0)$ a super-martingale, we then end up with the bound $u^*(x) \leq g(x)$. This gives us a super-martingale proof of the following (Lyapunov bound) result.

Proposition 11.2.1 Suppose that $g : S \to \mathbb{R}_+$ satisfies $Pg \leq g - r$ on C. Then, $u^* \leq g$ on C.

11.3 Stochastic Control: Martingales and the Value Function

We consider now a *controlled Markov chain* in which the system is subject to a control that impacts both the dynamics of the process and the rate at which reward accrues. Such sequential decision-making problems arise in many different applications settings:

- optimal consumption/investment in economics
- admission control/scheduling/routing for queues
- inventory management for supply chains

Here is the classical formulation:

S: state space

$\mathcal{A}(x)$: set of actions available in $x \in S$

$P_a(x, \mathrm{d}y)$: probability of going from x to $\mathrm{d}y$ under choice of action $a \in \mathcal{A}(x)$

$r(x, a)$: reward earned by using action $a \in \mathcal{A}(x)$ in state x

The *control* or *policy* $(A_n : n \geq 0)$ used by the decision-maker must satisfy:

i.) $A_n \in \mathcal{A}(X_n)$ for $n \geq 0$;

ii.) $(A_n : n \geq 0)$ is adapted to $(X_n : n \geq 0)$

Remark 11.3.1 If one wishes to consider randomized controls, one can adapt ii.) to:

ii'.) $(A_n : n \geq 0)$ is adapted to $((X_n, U_n) : n \geq 0)$, where $(U_n : n \geq 0)$ is a sequence of iid uniform(0,1) rv's for which U_n is also independent of $(X_j : 0 \leq j \leq n)$ for $n \geq 0$.

To fix our ideas, suppose we consider a controlled version of the problem considered in Sections 11.1 and 11.2 of these notes. Specifically, suppose that we wish to find an optimal control $(A_n^* : n \geq 0)$ maximizing the expected reward

$$\mathrm{E}\left[\sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x\right]$$

over all policies $(A_n : n \geq 0)$. Set

$$V^*(x) = \sup_{(A_n : n \geq 0)} \mathrm{E}\left[\sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x\right];$$

we call $V^* = (V^*(x) : x \in S)$ the value function for the control problem.

It seems reasonable to presume that $V^*(x)$ is attained by a Markov policy $A^* = (A_n^* : n \geq 0)$ for which $A_n^* = a^*(X_n)$ for $n \geq 0$; computing an optimal control then amounts to calculating $a^*(\cdot)$. Assuming that $V^*(\cdot)$ is indeed attained by such a Markov policy,

$$V^*(x) = \mathrm{E}_x^* \sum_{j=0}^{T-1} r(X_j, c^*(X_j)), \tag{11.3.1}$$

where $\mathrm{E}_x^*(\cdot)$ is the expectation operator associated with the Markov chain $X = (X_n : n \geq 0)$ having one-step transition probabilities given by

$$\mathrm{P}(x, \mathrm{d}y) = P_{a^*(x)}(x, \mathrm{d}y).$$

We have already seen in Section 11.1 that V^* should then satisfy

$$\int_S P_{a^*(X_{n-1})}(X_{n-1}, \mathrm{d}y)V^*(y) + r(X_{n-1}, a^*(X_{n-1})) = V^*(X_{n-1}) \quad \mathrm{P}_x^* \text{ a.s.} \tag{11.3.2}$$

on $\{T > n-1\}$. On the other hand, we will also want

$$V^*(x) \geq \mathrm{E}\left[\sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x\right]$$

for all adapted controls $(A_n : n \geq 0)$, in order to establish the optimality of $(A_n^* : n \geq 0)$. But our Lyapunov discussion of Section 11.2 suggests that this will follow if

$$\sum_{j=0}^{(T \wedge n)-1} r(X_j, A_j) + V^*(X_n)I(T > n-1)$$

is a super-martingale adapted to $(X_n : n \geq 0)$ under the probability associated with (11.3.1). This super-martingale property requires that

$$\int_S P_{A_{n-1}}(X_{n-1}, \mathrm{d}y)V^*(y) + r(X_{n-1}, A_{n-1}) \leq V^*(X_{n-1}) \tag{11.3.3}$$

on $\{T > n-1\}$. A sufficient (and close to necessary) condition is that

$$\int_S P_a(x, \mathrm{d}y)V^*(y) + r(x, a) \leq V^*(x) \tag{11.3.4}$$

for $x \in C$, $a \in \mathcal{A}(x)$, while a sufficient (and close to necessary) condition for (11.3.2) is

$$\int_S P_{a^*(x)}(x, \mathrm{d}y)V^*(y) + r(x, a^*(x)) \leq V^*(x) \tag{11.3.5}$$

for $x \in C$. The two sets of linear inequalities/equalities (11.3.4) and (11.3.5) can be combined into a single equation, namely

$$V^*(x) = \max_{a \in \mathcal{A}(x)} \left[r(x, a) + \int_S P_a(x, \mathrm{d}y)V^*(y)\right] \tag{11.3.6}$$

for $x \in C$, with $V(x) = 0$ for $x \in C^c$; the equation (11.3.6) is called the *optimality equation* or the *Hamilton-Jacobi-Bellman* (HJB) equation for the stochastic control problem.

As in our discussions of linear systems earlier in the course, the HJB equation can typically have multiple solutions. We now need to identify conditions uniquely determining the probabilistically meaningful solution V^* to (11.3.6). We will deal with this in the next section.

Remark 11.3.2 Once V^* has been computed, an optimal policy can be calculated as $A_n^* = a^*(X_n)$, where $a^*(x)$ is any maximizer of

$$\max_{a \in \mathcal{A}(x)} \left[r(x, a) + \int_S P_a(x, \mathrm{d}y)V^*(y)\right].$$

11.4 Proof of Optimality

Suppose now that $r(x,a) \geq 0$ for $x \in S$, $a \in \mathcal{A}(x)$. Then, V^* is clearly non-negative. Note that if V is any non-negative solution to (11.3.6), then

$$\sum_{j=0}^{(T\wedge n)-1} r(X_j, A_j) + V(X_n) I(T > n-1)$$

is a non-negative super-martingale adapted to $(X_n : n \geq 0)$ for any adapted policy $(A_n : n \geq 0)$, so that

$$\mathrm{E}_x \left[\sum_{j=0}^{(T\wedge n)-1} r(X_j, A_j) + V(X_n) | X_0 = x \right] \leq V(x),$$

and hence

$$\mathrm{E}_x \left[\sum_{j=0}^{(T\wedge n)-1} r(X_j, A_j) | X_0 = x \right] \leq V(x),$$

from which it follows that

$$V^*(x) \leq V(x). \tag{11.4.1}$$

In other words, any non-negative solution to (11.3.6) forms an upper bound on $V^*(x)$.

Remark 11.4.1 Actually, all that is required for an upper bound on $V^*(x)$ is that there exists a function $g : S \to \mathbb{R}_+$ such that

$$\mathrm{E}_x \left[\sum_{j=0}^{(T\wedge n)-1} r(X_j, A_j) + g(X_n) \right] \tag{11.4.2}$$

is a non-negative super-martingale adapted to $(X_n : n \geq 0)$ for any adapted policy $(A_n : n \geq 0)$. As in the steps leading to (11.4.1), we can then conclude that $g(x) \geq V^*(x)$. In order that (11.4.2) be a non-negative super-martingale, we require that

$$\int_S P_{A_{n-1}}(X_{n-1}, \mathrm{d}y) g(y) + r(X_{n-1}, A_{n-1}) \leq g(X_{n-1}) \tag{11.4.3}$$

on $\{T > n-1\}$. A sufficient (and close to necessary) condition for (11.4.3) is that

$$g(x) \geq \max_{a \in \mathcal{A}(x)} \left[r(x,a) + \int_S P_a(x, \mathrm{d}y) g(y) \right] \tag{11.4.4}$$

for $x \in C$. In other words, if $g : S \to \mathbb{R}_+$ satisfies (11.4.4), we obtain an upper bound on V^*, namely $V^*(x) \leq g(x)$. Because (11.4.4) involves an inequality solution, it is generally (much) easier to solve than (11.3.6).

Remark 11.4.2 Lower bounds on V^* are often easy to compute. In particular, any policy $(A_j : j \geq 0)$ provides a lower bound on $V^*(x)$:

$$V^*(x) \geq \mathrm{E} \left[\sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x \right].$$

Suppose now that $(r(x,a) : x \in S, a \in \mathcal{A}(x))$ is bounded, and that $\widetilde{\mathrm{P}}_x(T < \infty) = 1$, where $\widetilde{\mathrm{P}}_x(\cdot)$ is the probability associated with the Markov policy $(\widetilde{A}_n : n \geq 0)$, $\widetilde{A}_n = \widetilde{a}(X_n)$ for $n \geq 0$, and $\widetilde{a}(x)$ is any maximizer of

$$\max_{a \in \mathcal{A}(x)} \left[r(x,a) + \int_S P_a(x, \mathrm{d}y) V(y) \right];$$

here, V is a non-negative bounded solution of the optimality equation (11.3.6). Then,

$$\sum_{j=0}^{(T \wedge n)-1} r(X_j, \widetilde{A}_j) + V(X_n) I(T > n-1)$$

is a $\widetilde{\mathrm{P}}_x$-martingale, so that

$$\widetilde{\mathrm{E}}_x \sum_{j=0}^{(T \wedge n)-1} r(X_j, \widetilde{A}_j) + V(X_n) I(T > n-1) = V(x).$$

Since $V(\cdot)$ is bounded and $\widetilde{\mathrm{P}}_x(T > n) \to 0$ as $n \to \infty$,

$$\widetilde{\mathrm{E}}_x V(X_n) I(T > n-1) \to 0$$

as $n \to \infty$. As usual, the Monotone Convergence Theorem ensures that

$$\widetilde{\mathrm{E}}_x \sum_{j=0}^{(T \wedge n)-1} r(X_j, \widetilde{A}_j) \nearrow \widetilde{\mathrm{E}}_x \sum_{j=0}^{T-1} r(X_j, \widetilde{A}_j).$$

We conclude that the policy $(\widetilde{A}_n : n \geq 0)$ attains $V(x)$, so that

$$\widetilde{\mathrm{E}}_x \sum_{j=0}^{T-1} r(X_j, \widetilde{A}_j) = V(x).$$

Consequently, $V^*(x) = V(x)$ and $(\widetilde{A}_n : n \geq 0)$ is an optimal stochastic control. We therefore can summarize this discussion with the following result.

Theorem 11.4.1 i.) Suppose that $r(x,a) \geq 0$ for $x \in S$ and $a \in \mathcal{A}(x)$, and that V is a non-negative (finite-valued) solution of (11.3.6). Then,

$$V(x) \geq \sup_{(A_n : n \geq 0)} \mathrm{E} \left[\mathrm{E}_x \sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x \right]$$

for $x \in C$.

ii.) Suppose that, in addition, $(r(x,a) : x \in S, a \in \mathcal{A}(x))$ is bounded and that V is, in addition, bounded. If $\widetilde{\mathrm{P}}_x(T < \infty) = 1$ under the policy $(\widetilde{A}_n : n \geq 0)$ associated with V, then

$$V(x) = \sup_{(A_n : n \geq 0)} \mathrm{E} \left[\mathrm{E}_x \sum_{j=0}^{T-1} r(X_j, A_j) | X_0 = x \right]$$

and $(\widetilde{A}_n : n \geq 0)$ is an optimal policy.

In other words, if ii.) is satisfied, then V must equal the probabilistically meaningful solution of (11.3.6), namely V^*.

Remark 11.4.3 The above theorem does not address the question of existence of solutions to (11.3.6). This can be dealt with in different ways. When $|S| < \infty$ and $|\mathcal{A}(x)| < \infty$, note that (11.3.4) and (11.3.5) are a set of linear inequalities. The existence of a feasible solution then comes down to feasibility of a linear program (LP). One possible LP that can be used here is:

$$\begin{aligned} \min \quad & \sum_{x \in C} V(x) \\ \text{s/t} \quad & V(x) \geq r(x,a) + \sum_{y} P_a(x,y)V(y), \quad x \in C, a \in \mathcal{A}(x). \end{aligned}$$

Another approach is to use the method of *successive approximations* (known in the dynamic programming literature as *value iteration*). This involves setting

$$V_0(x) = \sup_{a \in \mathcal{A}(x)} [r(x,a)],$$

and defining $(V_n(x) : n \geq 1)$ via the iteration

$$V_{n+1}(x) = \sup_{a \in \mathcal{A}(x)} \left[r(x,a) + \int_C P_a(x, \mathrm{d}y) V_n(y) \right] = (\mathscr{R} V_n)(x),$$

where

$$(\mathscr{R} u)(x) \triangleq \sup_{a \in \mathcal{A}(x)} \left[r(x,a) + \int_C P_a(x, \mathrm{d}y) u(y) \right].$$

Under appropriate assumptions on the problem data, $\mathscr{R}$ is either a monotone operator (i.e. for $u_1 \leq u_2$, $\mathscr{R} u_1 \leq \mathscr{R} u_2$) or a contraction. In either case, $(V_n : n \geq 0)$ is a convergent sequence, and the limit V_∞ should be a solution to (11.3.6).

11.5 Optimal Stopping: Martingales and the Value Function

Consider now the following variant of a stochastic control problem. In particular, suppose that our goal is to compute an optimal stopping time T^* (adapted to $(X_n : n \geq 0)$) which maximizes

$$\mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty)$$

over all stopping times T adapted to the Markov chain $X = (X_n : n \geq 0)$.

As in our previous discussion, let

$$V^*(x) = \sup_T \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty);$$

we assume here that $r : S \to \mathbb{R}_+$ and $\alpha > 0$. Note that this is a control problem in which there are only two possible actions in each state $x \in S$: stop or continue.

Again, we expect that $V^*(x)$ will be attained by a Markov policy T^*; such an optimal stopping time T^* will take the form $T^* = \inf\{n \geq 0 : X_n \in \mathscr{S}\}$, where $\mathscr{S}$ is the *stopping region* of the corresponding stopping time T^*; $\mathscr{C} = S - \mathscr{S}$ is then the *continuation region*. Assuming that $V^*(x)$ is attained by such a stopping time T^*, then

$$V^*(x) = \mathrm{E}_x e^{-\alpha T^*} r(X_{T^*}) I(T^* < \infty). \tag{11.5.1}$$

Once again,

$$\mathrm{E}_x\left[e^{-\alpha T^*} r(X_{T^*}) I(T^* < \infty) | X_0, \ldots, X_n\right] = e^{-\alpha T^*} r(X_{T^*}) I(T^* < n) + e^{-\alpha n} I(T^* \geq n) V^*(X_n) \tag{11.5.2}$$

must be a martingale adapted to $(X_n : n \geq 0)$. In order that (11.5.2) be a martingale, we require that

$$V^*(X_{n-1}) = e^{-\alpha}(PV^*)(X_{n-1}) \quad \text{a.s.} \tag{11.5.3}$$

on $\{T^* > n-1\}$. Since $X_{n-1} \in \mathscr{C}$ on $\{T^* > n-1\}$, we can guarantee (11.5.3) by demanding that

$$V^*(x) = e^{-\alpha}(PV^*)(x)$$

for $x \in \mathscr{C}$. Of course, $V^*(x) = r(x)$ for $x \in \mathscr{S}$, and V^* clearly must satisfy $V^*(x) \geq r(x)$ for $x \in \mathscr{S}$. On the other hand, if

$$e^{-\alpha T} r(X_T) I(T < n) + e^{-\alpha n} I(T \geq n) V^*(X_n) \tag{11.5.4}$$

is a super-martingale adapted to $(X_n : n \geq 0)$, then

$$\mathrm{E}_x\left[e^{-\alpha T} r(X_T) I(T < n) + e^{-\alpha n} I(T \geq n) V^*(X_n)\right] \leq V^*(x).$$

Since V^* is non-negative,

$$\mathrm{E}_x e^{-\alpha T} r(X_T) I(T < n) \leq V^*(x),$$

so that the Monotone Convergence Theorem guarantees that

$$V^*(x) \geq \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty).$$

Hence, $V^*(x)$ is an upper bound on

$$\sup_T \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty); \tag{11.5.5}$$

in conjunction with (11.5.1), this will then ensure that

$$V^*(x) = \sup_T \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty).$$

A sufficient (and close to necessary) condition for (11.5.4) to be a super-martingale for all stopping times is that

$$V(x) \geq e^{-\alpha}(PV)(x)$$

for $x \in S$. Thus, we arrive at the conclusion that V^* should satisfy:

$$\begin{aligned}
V^*(x) &= e^{-\alpha}(PV^*)(x), \quad x \in \mathscr{C} \\
V^*(x) &\geq e^{-\alpha}(PV^*)(x), \quad x \in S \\
V^*(x) &= r(x), \quad x \in \mathscr{S} \\
V^*(x) &\geq r(x), \quad x \in S.
\end{aligned}$$

The four above relations can be summarized via a single optimality (or HJB) equation: V^* should be a solution of

$$V(x) = \max(r(x), e^{-\alpha}(PV)(x)) \tag{11.5.6}$$

for $x \in S$.

Remark 11.5.1 $\mathscr{S} = \{x \in S : V^*(x) = r(x)\}$.

Remark 11.5.2 If $g : S \to \mathbb{R}_+$ is a solution to

$$g(x) \geq \max(r(x), e^{-\alpha}(Pg)(x))$$

for $x \in S$, then

$$g(x) \geq \sup_T \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty).$$

11.6 Proof of Optimality

We study here the question of characterizing conditions under which a given solution V of (11.5.6) coincides with the probabilistically meaningful value function V^*. Suppose that $r : S \to \mathbb{R}_+$ is a bounded function; then V^* is clearly bounded, so it makes sense to consider only bounded solutions V of (11.5.6).

If V is then a non-negative solution of (11.5.6), $\mathscr{S} \triangleq \{x \in S : V(x) = r(x)\}$, and $T^* \triangleq \inf\{n \geq 0 : X_n \in \mathscr{S}\}$, then

$$\mathrm{E}_x \left[e^{-\alpha T^*} r(X_{T^*}) I(T^* < n) + e^{-\alpha n} I(T^* > n) V(X_n) \right] = V(x).$$

But $\mathrm{E}_x e^{-\alpha n} I(T^* > n) V(X_n) \to 0$ (since V is bounded) and Monotone Convergence ensures

$$\mathrm{E}_x e^{-\alpha T^*} r(X_{T^*}) I(T^* < n) \nearrow \mathrm{E}_x e^{-\alpha T^*} r(X_{T^*}) I(T^* < \infty)$$

and hence

$$V(x) = \mathrm{E}_x e^{-\alpha T^*} r(X_{T^*}) I(T^* < \infty).$$

We conclude that:

Theorem 11.6.1 Suppose that $r : S \to \mathbb{R}_+$ is non-negative and bounded, and V is a non-negative bounded solution of (11.5.6). Then,

$$V(x) = V^*(x) (= \sup_T \mathrm{E}_x e^{-\alpha T} r(X_T) I(T < \infty)).$$