

5 Negative Programming

We address the special theory of minimizing positive costs, (noting that the action that extremizes the right hand side of the optimality equation gives an optimal policy), and stopping problems and their solution.

5.1 Stationary policies

A **Markov policy** is a policy that specifies the control at time t to be simply a function of the state and time. In the proof of Theorem 4.1 we used $u_t = f_t(x_t)$ to specify the control at time t . This is a convenient notation for a Markov policy, and we write $\pi = (f_0, f_1, \dots)$. If in addition the policy does not depend on time, it is said to be a **stationary Markov policy**, and we write $\pi = (f, f, \dots) = f^\infty$.

5.2 Characterization of the optimal policy

Negative programming concerns minimizing non-negative costs, $c(x, u) \geq 0$. The name originates from the equivalent problem of maximizing non-positive rewards, $r(x, u) \leq 0$.

The following theorem gives a necessary and sufficient condition for a stationary policy to be optimal: namely, it must choose the optimal u on the right hand side of the optimality equation. Note that in the statement of this theorem we are requiring that the infimum over u is attained as a minimum over u .

Theorem 5.1 *Suppose D or N holds. Suppose $\pi = f^\infty$ is the stationary Markov policy such that*

$$\begin{aligned} c(x, f(x)) + \beta E[F(x_1) \mid x_0 = x, u_0 = f(x)] \\ = \min_u [c(x, u) + \beta E[F(x_1) \mid x_0 = x, u_0 = u]]. \end{aligned}$$

Then $F(\pi, x) = F(x)$, and π is optimal.

Proof. Suppose this policy is $\pi = f^\infty$. Then by substituting the optimality equation into itself and using the fact that π specifies the minimizing control at each stage,

$$F(x) = E_\pi \left[\sum_{t=0}^{s-1} \beta^t c(x_t, u_t) \mid x_0 = x \right] + \beta^s E_\pi [F(x_s) \mid x_0 = x]. \quad (5.1)$$

In case N we can drop the final term on the right hand side of (5.1) (because it is non-negative) and then let $s \rightarrow \infty$; in case D we can let $s \rightarrow \infty$ directly, observing that this term tends to zero. Either way, we have $F(x) \geq F(\pi, x)$. ■

A corollary is that if assumption F holds then an optimal policy exists. Neither Theorem 5.1 or this corollary are true for positive programming (c.f., the example in Section 4.1).

5.3 Optimal stopping over a finite horizon

One way that the total-expected cost can be finite is if it is possible to enter a state from which no further costs are incurred. Suppose u has just two possible values: $u = 0$ (stop), and $u = 1$ (continue). Suppose there is a termination state, say 0, that is entered upon choosing the stopping action. Once this state is entered the system stays in that state and no further cost is incurred thereafter.

Suppose that stopping is mandatory, in that we must continue for no more than s steps. The finite-horizon dynamic programming equation is therefore

$$F_s(x) = \min\{k(x), c(x) + E[F_{s-1}(x_1) \mid x_0 = x, u_0 = 1]\}, \quad (5.2)$$

with $F_0(x) = k(x)$, $c(0) = 0$.

Consider the set of states in which it is at least as good to stop now as to continue one more step and then stop:

$$S = \{x : k(x) \leq c(x) + E[k(x_1) \mid x_0 = x, u_0 = 1]\}.$$

Clearly, it cannot be optimal to stop if $x \notin S$, since in that case it would be strictly better to continue one more step and then stop. The following theorem characterises all finite-horizon optimal policies.

Theorem 5.2 *Suppose S is closed (so that once the state enters S it remains in S .) Then an optimal policy for all finite horizons is: stop if and only if $x \in S$.*

Proof. The proof is by induction. If the horizon is $s = 1$, then obviously it is optimal to stop only if $x \in S$. Suppose the theorem is true for a horizon of $s - 1$. As above, if $x \notin S$ then it is better to continue for more one step and stop rather than stop in state x . If $x \in S$, then the fact that S is closed implies $x_1 \in S$ and so $F_{s-1}(x_1) = k(x_1)$. But then (5.2) gives $F_s(x) = k(x)$. So we should stop if $s \in S$. ■

The optimal policy is known as a **one-step look-ahead rule** (OSLA).

5.4 Example: optimal parking

A driver is looking for a parking space on the way to his destination. Each parking space is free with probability p independently of whether other parking spaces are free or not. The driver cannot observe whether a parking space is free until he reaches it. If he parks s spaces from the destination, he incurs cost s , $s = 0, 1, \dots$. If he passes the destination without having parked the cost is D . Show that an optimal policy is to park in the first free space that is no further than s^* from the destination, where s^* is the greatest integer s such that $(Dp + 1)q^s \geq 1$.

Solution. When the driver is s spaces from the destination it only matters whether the space is available ($x = 1$) or full ($x = 0$). The optimality equation gives

$$\begin{aligned} F_s(0) &= qF_{s-1}(0) + pF_{s-1}(1), \\ F_s(1) &= \min \begin{cases} s, & \text{(take available space)} \\ qF_{s-1}(0) + pF_{s-1}(1), & \text{(ignore available space)} \end{cases} \end{aligned}$$

where $F_0(0) = D$, $F_0(1) = 0$.

Suppose the driver adopts a policy of taking the first free space that is s or closer. Let the cost under this policy be $k(s)$, where

$$k(s) = ps + qk(s-1),$$

with $k(0) = qD$. The general solution is of the form $k(s) = -q/p + s + cq^s$. So after substituting and using the boundary condition at $s = 0$, we have

$$k(s) = -\frac{q}{p} + s + \left(D + \frac{1}{p}\right)q^{s+1}, \quad s = 0, 1, \dots$$

It is better to stop now (at a distance s from the destination) than to go on and take the first available space if s is in the stopping set

$$S = \{s : s \leq k(s-1)\} = \{s : (Dp+1)q^s \geq 1\}.$$

This set is closed (since s decreases) and so by Theorem 5.2 this stopping set describes the optimal policy. ■

If the driver parks in the first available space past his destination and walk backs, then $D = 1 + qD$, so $D = 1/p$ and s^* is the greatest integer such that $2q^s \geq 1$.

5.5 Optimal stopping over the infinite horizon

Let us now consider the stopping problem over the infinite-horizon. As above, let $F_s(x)$ be the infimal cost given that we are required to stop by time s . Let $F(x)$ be the infimal cost when all that is required is that we stop eventually. Since less cost can be incurred if we are allowed more time in which to stop, we have

$$F_s(x) \geq F_{s+1}(x) \geq F(x).$$

Thus by monotone convergence $F_s(x)$ tends to a limit, say $F_\infty(x)$, and $F_\infty(x) \geq F(x)$.

Example: we can have $F_\infty > F$

Consider the problem of stopping a symmetric random walk on the integers, where $c(x) = 0$, $k(x) = \exp(-x)$. The policy of stopping immediately, π , has $F(\pi, x) = \exp(-x)$, and this satisfies the infinite-horizon optimality equation,

$$F(x) = \min\{\exp(-x), (1/2)F(x+1) + (1/2)F(x-1)\}.$$

However, π is not optimal. A symmetric random walk is recurrent, so we may wait until reaching as large an integer as we like before stopping; hence $F(x) = 0$. Inductively, one can see that $F_s(x) = \exp(-x)$. So $F_\infty(x) > F(x)$.

(Note: Theorem 4.2 says that $F_\infty = F$, but that is in a setting in which there is no terminal cost and for different definitions of F_s and F than we take here.)

Example: Theorem 4.1 is not true for negative programming

Consider the above example, but now suppose one is allowed never to stop. Since continuation costs are 0 the optimal policy for all finite horizons and the infinite horizon is never to stop. So $F(x) = 0$ and this satisfies the optimality equation above. However, $F(\pi, x) = \exp(-x)$ also satisfies the optimality equation and is the cost incurred by stopping immediately. Thus it is not true (as for positive programming) that a policy whose cost function satisfies the optimality equation is optimal.

The following lemma gives conditions under which the infimal finite-horizon cost does converge to the infimal infinite-horizon cost.

Lemma 5.3 *Suppose all costs are bounded as follows.*

$$(a) K = \sup_x k(x) < \infty \quad (b) C = \inf_x c(x) > 0. \quad (5.3)$$

Then $F_s(x) \rightarrow F(x)$ as $s \rightarrow \infty$.

Proof. (*starred*) Suppose π is an optimal policy for the infinite horizon problem and stops at the random time τ . Then its cost is at least $(s+1)CP(\tau > s)$. However, since it would be possible to stop at time 0 the cost is also no more than K , so

$$(s+1)CP(\tau > s) \leq F(x) \leq K.$$

In the s -horizon problem we could follow π , but stop at time s if $\tau > s$. This implies

$$F(x) \leq F_s(x) \leq F(x) + KP(\tau > s) \leq F(x) + \frac{K^2}{(s+1)C}.$$

By letting $s \rightarrow \infty$, we have $F_\infty(x) = F(x)$. ■

Note that the problem posed here is identical to one in which we pay K at the start and receive a terminal reward $r(x) = K - k(x)$.

Theorem 5.4 *Suppose S is closed and (5.3) holds. Then an optimal policy for the infinite horizon is: stop if and only if $x \in S$.*

Proof. By Theorem 5.2 we have for all finite s ,

$$F_s(x) = \begin{cases} k(x) & x \in S, \\ < k(x) & x \notin S. \end{cases}$$

Lemma 5.3 gives $F(x) = F_\infty(x)$. ■