Convex Analytic Methods in Markov Decision Processes

Vivek S. Borkar
Department of Computer Science and Automation
Indian Institute of Science
Bangalore 560 012
Email: borkar@csa.iisc.ernet.in

1 Introduction

1.1 Background

Markov decision processes optimize phenomena evolving with time and thus are intrinsically dynamic optimization problems. Nevertheless, they can be cast as abstract 'static' optimization problems over a closed convex set of measures. They then become convex programming (in fact, infinite dimensional linear programming) problems for which the enormous machinery of the latter fields can be invoked and used to advantage. Logically, these are extensions of the linear programming approach to finite state finite action space problems due to Manne [23]. (Further references are given in the 'bibliographical note' at the end.) The attraction of this approach lies in the following:

(i) It leads to elegant alternative derivations of known results, sometimes under weaker hypotheses, from a novel perspective.

(ii) It brings to the fore the possibility of using convex/linear programming techniques for computing near-optimal strategies.

(iii) It allows one to handle certain unconventional problems (such as control under additional constraints on secondary 'costs') where traditional dynamic programming paradigm turns out to be infeasible or awkward.

Our primary focus will be on countable state space and ergodic (or 'long run average') cost, for which the theory is the most elegant. This is done in the next section. Section 3 sketches extensions, first to other cost criteria and then to general state spaces. Section 4 considers multiobjective problems, such as the problem of control under constraints. Section 5 concludes with a brief discussion, followed by a 'bibliographical note'.

1.2 Notation

We shall consider a controlled Markov chain $X_n, n \geq 0$, taking values in a state space S. Initially, we shall consider only countable S, without loss of generality identified with the set $\{0, 1, 2, \cdots\}$. With each $i \in S$ is associated an action space U_i, which is a copy of a fixed compact metric space U. The transition probability $P(j/i, u)$ for $i, j \in S$, $u \in U_i$ will denote the probability of going from i to j, when action u is chosen. Thus $P(j/i, u) \in [0, 1]$ $\forall i, j, u$

with

$$\sum_j P(j/i,u) = 1.$$

Note that there is no loss of generality in letting U_i's be replicas of a fixed compact metric space U. If not, we could always set $U = \Pi_i U_i$ and replace $P(j/i,\cdot)$ by its composition with the projection $U \to U_i$ for each i, j. Let $\{Z_n\}$ denote the U-valued control process. Thus $\{X_n\}$ evolves as per

$$P(X_{n+1} = j/X_m, Z_m, m \leq n) = P(j/X_n, Z_n), n \geq 0.$$

We call $\{Z_n\}$ a stationary strategy if $Z_n = v(X_n)\ \forall n$ for some measurable $v : S \to U$, and a stationary randomized strategy if for each n, Z_n is conditionally independent of $X_m, Z_m, m < n$, given X_n and its regular conditional law given X_n is $f(X_n)$ for some measurable $f : S \to \mathcal{P}(U)$. (Here and elsewhere, $\mathcal{P}(X)$ for a Polish space X is the Polish space of probability measures on X with the Prohorov topology. See, e.g., [12], Chapter 2.) By abuse of terminology, we may identify a stationary (resp., stationary randomized) strategy with the corresponding map $v(\cdot)$ (resp., $f(\cdot)$). Denote by Π_S (resp., Π_{SR}) the set of stationary (resp., stationary randomized) strategies. Note that under either, $\{X_n\}$ is a time-homogeneous Markov chain.

Let $k : S \times U \to R$ be a continuous 'running cost' function, which we assume to be bounded from below. The various 'classical' control problems then are:

(i) Finite horizon contol

Minimize $E\left[\sum_{n=0}^{N} k(X_n, Z_n)\right]$ for some $N \geq 0$.

(ii) Infinite horizon discounted cost control

Minimize $E\left[\sum_{n=0}^{\infty} \beta^n k(X_n, Z_n)\right]$ for some 'disscount factor' $\beta \in (0,1)$.

(iii) Control up to exit time

Minimize $E\left[\sum_{n=0}^{\tau} k(X_n, Z_n)\right]$ for some $\tau = \min\{n \geq 0 : X_n \notin O\}$, O being a prescribed subset of S. ($\tau = \infty$ if $X_n \in O\ \forall n$.)

(iv) Ergodic or long run average cost control

Minimize 'almost surely'

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} k(X_n, Z_n). \tag{1.1}$$

We shall focus primarily on (iv). Note that if $f \in \Pi_{SR}$ is used and X_0 is in the support of an ergodic distribution $\pi \in \mathcal{P}(S)$ for the corresponding time-homogeneous Markov chain $\{X_n\}$, then (1.1) will a.s. equal

$$\sum_i \pi(i)\bar{k}(i, f(i))$$

where $\bar{k} : S \times \mathcal{P}(U) \to R$ is defined by

$$\bar{k}(i, u) = \int k(i, y)u(dy), \ (i, u) \in S \times \mathcal{P}(U).$$

This will be the starting point of our convex analytic formulation of (iv), which we take up next.

2 Ergodic Control

2.1 Ergodic occupation measures

If the chain controlled by $f \in \Pi_{SR}$ has an invariant probability measure $\pi \in \mathcal{P}(S)$, we associate with the pair (π, f) the *ergodic occupation measure* $\hat{\pi}_f$ defined by:

$$\int_{S\times U} g d\hat{\pi}_f = \sum_{i\in S} \pi(i) \ \int_U g(i, y)f(i, dy),$$

for $g \in C_b(S \times U)(\triangleq$ space of bounded continuous maps $S \times U \to R$. Analogous notation will be used throughout.) Let $\mathcal{C}$ be a countable subset of $C_b(S)$ satisfying : for any $f \in C_b(S)$, there exist $\{g_n\}$ in the linear span of $\mathcal{C}$ such that $g_n \to f$ in a bounded, pointwise fashion.

Lemma 2.1 $\hat{\pi}_f$ satisfies : $\forall g \in C_b(S)$,

$$\sum_{i\in S} g(i)\hat{\pi}_f(\{i\}, U) = \int_{S\times U} \left(\sum_{j\in S} P(j/i, a)g(j) \right) d\hat{\pi}_f.$$

Conversely, if $\eta \in \mathcal{P}(S \times U)$ satisfies

$$\sum_{i\in S} g(i)\eta(\{i\}, U) = \int_{S\times U} \left(\sum_{j\in S} P(j/i, a)g(j) \right) d\eta \tag{2.1}$$

for $g \in \mathcal{C}$, then $\eta = \hat{\pi}_f$ for some (π, f).

Proof The first claim is simply the invariance of π under f. For the second, note that (2.1) holds for all $g \in C_b(S)$ if it does so for $g \in \mathcal{C}$. Disintegrate η as $\eta(\{i\}, dy) = \pi(i)f(i, dy)$ with

$\pi \in \mathcal{P}(S)$ and $f : S \to \mathcal{P}(U)$ the appropriate regular conditional law defined π-a.s. uniquely. Identify f with an element of Π_{SR}, whence (2.1) implies that π is invariant under it. The claim follows. □

Often $\mathcal{C} = \{I_{\{i\}}, i \in S\}$ is a convenient choice for $\mathcal{C}$. Let G denote the set of all ergodic occupation measures.

Lemma 2.2 G is closed convex in $\mathcal{P}(S \times U)$.

Proof (2.1) is preserved under convergence in $\mathcal{P}(S \times U)$. In view of Lemma 2.1, this implies that G is closed. Let $\nu_j \in G$, $1 \leq j < n$, with $\nu_j(\{i\}, dy) = \pi_j(i) f_j(i, dy)$ for $f_j \in \Pi_{SR}$ and $\pi_j \in \mathcal{P}(S)$ invariant under respective f_j's. Let $a_j \in (0,1), 1 \leq j \leq n$, with $\sum_j a_j = 1$. Set $\nu = \sum_j a_j \nu_j, \pi = \sum_j a_j \pi_j$, and define $f \in \Pi_{SR}$ by:

$$f(i, dy) = \sum_{j=1}^{n} \left[a_j \pi_j(i) / \left(\sum_{l=1}^{n} a_l \pi_l(i) \right) \right] f_j(i, dy),$$

for $i \in$ support (π), arbitrary otherwise. Then $\nu(\{i\}, dy) = \pi(i) f(i, dy)$, $i \in S$. Now write (2.1) separately for each $\eta = \nu_j$, $1 \leq j \leq n$, multiply it through by a_j, and sum over j. Rearranging terms, one recovers (2.1) for ν. Thus $\nu \in G$. □

We now establish a technical lemma for later use. Let $\nu = \hat{\pi}_f \in G$. Now suppose that for some $j \in$ support (π) (say, $j = 1$), we have $f(j, dy) = f(1, dy) = a\varphi_1(dy) + (1-a)\varphi_2(dy)$ for some $a \in (0,1)$ and $\varphi_1 \neq \varphi_2$ in $\mathcal{P}(U)$. Define $f', f'' \in \Pi_{SR}$ by:

$$f'(i) = f''(i) = f(i), \quad i \neq 1$$
$$f'(1) = \varphi_1, f''(1) = \varphi_2.$$

Lemma 2.3 Both f', f'' above admit invariant probability measures containing '1' in their support.

Proof Changing f to f' or f'' affects only the probabilities of transitions out of 1. Let T, T', T'' denote the mean return times to 1 under f, f', f'' resp. Then clearly $T = aT' + (1-a)T''$. Since $T < \infty$, $1 > a > 0$, we have $T', T'' < \infty$, implying the claim. □

Let $G_e \subset G$ denote the set of extreme points of G.

Lemma 2.4 Every $\nu \in G_e$ is of the form $\nu = \hat{\pi}_f$ for some $f \in \Pi_{SS}$ and π ergodic under f.

Proof Let $\nu \in G_e$ with $\nu(\{i\}, dy) = \pi(i) f(i, dy)$. For $i \notin$ support (π), set $f(i) =$ some Dirac measure, without affecting ν. Let $i \in$ support (π), say, $i = 1$. Suppose $f(1, dy) =$

$a\varphi_1(dy) + (1-a)\varphi_2(dy)$ for some $a \in (0,1), \varphi_1 \neq \varphi_2$ in $\mathcal{P}(U)$. Define f', f'' as above and let π_1, π_2 be ergodic probability measures under f', f'' resp. containing 1 in their supports. Pick $b \in (0,1)$ such that

$$a = b\pi_1(1)/(b\pi_1(1) + (1-b)\pi_2(1)),$$

which is possible because $\pi_1(1), \pi_2(1) > 0$. Let $\pi' = b\pi_1 + (1-b)\pi_2, \nu'(\{i\}, dy) = \pi'(i)f(i, dy)$. A computation similar to that in Lemma 2.2 shows that ν' satisfies (2.1) and hence $\nu' = \hat{\pi}'_f \in G$. Also, support (π') = support $(\pi_1) \cup$ support (π_2) and for $i, j \in S$,

$$\int P(j/i,y)f(i,dy) = a\int P(j/i,y)f'(i,dy) \\ +(1-a)\int P(j/i,y)f''(i,dy). \tag{2.2}$$

Since π_1 (resp., π_2) is ergodic under f' (resp., f''), any two states in its support communicate under f' (resp., f'') and therefore under f in view of (2.2) above. Since 1 is in support $(\pi_1) \cap$ support (π_2), it follows that support (π') is a single communicating class under f. Thus π' is an ergodic probability measure under f. If π is also ergodic, we must have $\pi = \pi'$ and $\nu = \nu'$. As in the proof of Lemma 2.2, one can then verify that $\nu' = b\nu_1 + (1-b)\nu_2$ where $\nu_1(\{i\}, dy) = \pi_1(i)f'(i, dy)$ and $\nu_2(\{i\}, dy) = \pi_2(i)f''(i, dy)$. Since $0 < b < 1$ and $\nu_1 \neq \nu_2, \nu \notin G_e$, a contradiction. Suppose π is not ergodic under f. Then $\pi = c\pi^1 + (1-c)\pi^2$ for $c \in (0,1)$ and $\pi^1, \pi^2 \in \mathcal{P}(S)$ which are distinct invariant probability measures under f. Then $\nu = c\nu^1 + (1-c)\nu^2$ where $\nu^j(\{i\}, dy) = \pi^j(i)f(i, dy), i = 1, 2$ are clearly distinct elements of G. Thus $\nu \notin G_e$, a contradiction. So, (i) π must be ergodic under f, and (ii) (2.2) is impossible, i.e., $f(i, dy)$ is Dirac for $i \in$ support (π). This completes the proof. □

Let $\bar{S} = S \cup \{\infty\}$ denote the one point compactification of S and view $S \subset \bar{S}$ via the natural embedding. Likewise, $\mathcal{P}(S \times U) \subset \mathcal{P}(\bar{S} \times U)$ via the natural embedding. Let $\bar{G}$ be the closure of G in $\mathcal{P}(\bar{S} \times U)$ and $\bar{G}_e$ the set of its extreme points. The statement of the next lemma requires familiarity with Choquet's theorem, which we recall here.

Let E be a Haussdorff locally convex topological vector space and $X \subset E$ a convex compact metrizable subset. Given $\mu \in \mathcal{P}(X)$, call x its barycenter if $f(x) = \int f d\mu$ for all continuous affine $f : X \to R$.

Theorem 2.1 (Choquet) Each $x \in X$ is the barycenter of some $\mu \in \mathcal{P}(X)$ supported on the extreme points of X.

See [14], pp.140-141, for a proof. Metrizability of X ensures that the set of its extreme points is F_σ, hence measurable ([14], p.138), whereas compactness of X ensures that it is nonempty ([14], pp.105).

Lemma 2.5 $G_e \subset \bar{G}_e$ and any $\nu \in G$ is the barycenter of a probability measure on G_e.

Proof Any $\nu \in G_e \backslash \bar{G}_e$ must be a convex combination of two distinct elements of $\bar{G}$ at least one of which must assign strictly positive probability to $\{\infty\} \times U$. But then so will ν, a contradiction. Thus $G_e \subset \bar{G}_e$. Now, if G is compact, $G_e \neq \phi$ and $G_e = \bar{G}_e$, whence the second claim follows from Theorem 2.1. If not, apply the theorem to $\bar{G}$. Then $\bar{G}_e \neq \phi$ and ν is the barycenter of a probability measure Φ on $\bar{G}_e$. If $\Phi(\bar{G}_e \backslash G_e) > 0$, we must have $\nu(\{\infty\} \times U) > 0$, a contradiction. Thus $\Phi(G_e) = 1$. □

Finally, we have:

Lemma 2.6 Each $\nu \in \bar{G}$ is of the form : For $A \subset \bar{S} \times U$ Borel,

$$\nu(A) = \delta\nu'(A \cap (S \times U)) + (1-\delta)\nu''(A \cap (\{\infty\} \times U)) \tag{2.3}$$

with $\delta \in [0,1], \nu' \in G$ and $\nu'' \in \mathcal{P}(\{\infty\} \times U)$.

Proof Clearly, (2.3) holds for some $\nu' \in \mathcal{P}(S \times U)$. The claim is trivial for $\delta = 0$. For $\delta = 1, \nu' = \nu \in G$. Thus let $\delta \in (0,1)$. Let $\nu_n \in G, n \geq 1$, be such that $\nu_n \to \nu$ in $\bar{G}$. Suppose

$$\nu_n(i, dy) = \pi_n(i) f_n(i, dy), n \geq 1.$$

Then for $j \in S$,

$$\int P(j/\cdot, \cdot) d\nu_n = \nu(\{j\} \times U), \ \ n \geq 1.$$

Since $\{j\} \times U$ is both open and closed in $\bar{S} \times U$, we have

$$\nu_n(\{j\} \times U) \to \nu(\{j\} \times U) = \delta\nu'(\{j\} \times U).$$

Letting $S(N) = \{0, 1, \ldots, N\} \subset S$,

$$\begin{aligned} \liminf_{n\to\infty} \int P(j/\cdot, \cdot) d\nu_n &\geq \lim_{n\to\infty} \int_{S(N)\times U} P(j/\cdot, \cdot) d\nu_n \\ &= \int_{S(N)\times U} P(j/\cdot, \cdot) d\nu \\ &= \delta \int_{S(N)\times U} P(j/\cdot, \cdot) d\nu' \\ &\to \delta \int P(j/\cdot, \cdot) d\nu' \end{aligned}$$

as $N \to \infty$. Combining the two,

$$\int P(j/\cdot, \cdot) d\nu' \leq \nu'(\{j\} \times U).$$

Both sides add up to one when summed over j, so equality must hold for all j. Therefore $\nu' \in G$ by Lemma 2.1. □

These lemmas form the backdrop for the convex programming problem we describe next.

2.2 The convex programming problem

Recall that if $f \in \Pi_{SR}$ has an ergodic probability measure π and $X_0 \in$ support (π) a.s., then the ergodic cost a.s. equals $\int kd\hat{\pi}_f$. This suggests the convex programming problem:

Minimize $\int kd\mu$ over $\mu \in G$.

Equivalently:

Minimize $\int kd\mu$ over $\{\mu \in \mathcal{P}(S \times U)$: (2.1) holds $\}$.

This displays it as an infinite dimensional linear program.

Let $\mu^* \in G$ be such that

$$\alpha \triangleq \inf_{\mu \in G} \int kd\mu \leq \int kd\mu^* \leq \alpha + \epsilon$$

for some $\epsilon \geq 0$. (Note that such a μ^* is guarranteed to exist for $\epsilon > 0$.) Such a μ^* is said to be ϵ-optimal (or optimal if $\epsilon = 0$).

Lemma 2.7 For $\epsilon \geq 0$, if there is an ϵ-optimal $\mu^* \in G$, then there is an ϵ-optimal $\overline{\mu} \in G_e$.

Proof By Choquet's theorem, there exists a $\Phi \in \mathcal{P}(G_e)$ such that

$$\alpha \leq \int kd\mu^* = \int_{G_e} \left(\int kd\mu \right) \Phi(d\mu) \leq \alpha + \epsilon.$$

Thus for some $\overline{\mu} \in$ support (Φ), $\alpha \leq \int kd\overline{\mu} \leq \alpha + \epsilon$. □

We next consider two conditions under which an optimal μ^* exists.

Case 1: The near-monotone case

Call $k(\cdot, \cdot)$ near-monotone if

$$\liminf_{i \to \infty} \min_u k(i, u) > \alpha. \tag{2.4}$$

Lemma 2.8 Under (2.4), an optimal $\mu^* \in G_e$ exists.

Proof Let $\mu_n \in G, n \geq 1$, be such that $\int kd\mu_n \downarrow \alpha$. Viewing $G \subset \mathcal{P}(\overline{S} \times U)$ as before, drop to a subsequence if necessary and suppose that $\mu_n \to \overline{\mu}$ in $\mathcal{P}(\overline{S} \times U)$. Write $\overline{\mu} = \delta\mu' + (1-\delta)\mu''$ as in (2.3). Pick $N \geq 1,\ \epsilon > 0$ such that $\inf_n k(i, u) \geq \alpha + \epsilon$ for $i \geq N$. For $n \geq 1$, define

$$k_n(i, u) = k(i, u)I\{i \leq N + n\} + (\alpha + \epsilon)I\{i > N + n\}.$$

Then $\forall\ j$,

$$\alpha \geq \liminf_{n\to\infty} \int k_j d\mu_n = \delta \int k_j d\mu' + (1-\delta)(\alpha+\epsilon).$$

Let $j \uparrow \infty$ on the right to obtain

$$\alpha \geq \delta \int k d\mu' + (1-\delta)(\alpha+\epsilon) \geq \delta\alpha + (1-\delta)(\alpha+\epsilon).$$

Thus $\delta = 1$ and $\int k d\mu' = \alpha$. The claim now follows from Lemma 2.7. □

Case 2: The stable case

Here we assume G to be compact. Since $\mu \to \int k d\mu$ is lower semicontinuous, it attains a minimum on G, hence, by Lemma 2.7, on G_e.

We still need to show that the 'optimum' is an optimum with respect to all admissible strategies. In 'Case 1', this is true:

Lemma 2.9 In the near-monotone case, under any admissible strategy,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} k(X_m, Z_m) \geq \alpha \text{ a.s.} \tag{2.5}$$

To prove this, we introduce $\mathcal{P}(S\times U)$-valued process $\{\nu_n\}$ of 'empirical measures', defined by

$$\int f d\nu_n = \frac{1}{n}\sum_{m=0}^{n-1} f(X_m, Z_m),\ f \in C_b(X\times U).$$

Lemma 2.10 $\nu_n \to \overline{G}$ a.s.

Proof Let $\mathcal{C} = \{I_{\{i\}}, i \in S\} \subset C_b(S)$. Then by the strong law of large number for martingales ([12]), pp. 53-54),

$$\sum_i g(i)\nu_n(\{i\}, U) - \int_{S\times U}\left(\sum_j P(j/i,a)g(j)\right) d\nu_n \to 0 \quad \text{a.s.} \tag{2.6}$$

Since $\mathcal{C}$ is countable, this holds for all sample points outside a common zero probability set. Fix one such sample point. Let $\overline{\nu}$ be a limit point of $\{\nu_n\}$ in $\mathcal{P}(\overline{S}\times U)$ and write $\overline{\nu} = \delta\nu' + (1-\delta)\nu''$ with $\delta \in [0,1]$ and $\nu', \nu'' \in \mathcal{P}(\overline{S}\times U)$ supported on $S\times U$ and $\{\infty\}\times U$ resp. By (2.6), ν' satisfies (2.1) whenever $\delta > 0$. Identifying ν' with its restriction to $S\times U$ by abuse of notation, we have $\nu' \in G$, i.e., $\overline{\nu} \in \overline{G}$. □

Proof of Lemma 2.9 Consider a sample point for which Lemma 2.4 holds and let $\overline{\nu}$ be a limit point of $\{\nu_n\}$ as above, say, $\nu_{n(l)} \to \overline{\nu}$. Then for k_n's as in Lemma 2.8 above,

$$\liminf_{l\to\infty} \int k d\nu_{n(l)} \geq \int k_j d\overline{\nu} \geq \delta \int k_j d\nu' + (1-\delta)(\alpha+\epsilon),\ j \geq 1.$$

Let $j \uparrow \infty$ on the right to obtain

$$\liminf_{l \to \infty} \int k d\nu_{n(l)} \geq \delta \int k d\nu' + (1-\delta)(\alpha+\epsilon) \geq \delta\alpha + (1-\delta)(\alpha+\epsilon),$$

in view of the preceding lemma. The claim follows. □

Case 2 needs more work. For simplicity, we assume that there is a single communicating class under any $f \in \Pi_{SR}$. Let $\mathcal{F}_n = \sigma(X_m, Z_m, m \leq n), n \geq 0$, for an arbitrary admissible control sequence $\{Z_n\}$ and the corresponding chain $\{X_n\}$. Let $\tau(i) = \min\{n > 0 : X_n = i\}, i \in S$.

Lemma 2.11 For any $\{\mathcal{F}_n\}$-stopping time τ, the regular conditional law of $X_{\tau+n}, n \geq 0$, on $\{\tau < \infty\}$ is a.s. the law of a controlled Markov chain on S with action space U and transition probabilities $P(\cdot/\cdot,\cdot)$.

Proof This is a straighforward consequence of the easily verified fact

$$E\left[\left(I_{\{X_{\tau+n+1}=i\}} - P\left(i/X_{\tau+n}, Z_{\tau+n}\right)\right)/\mathcal{F}_\tau\right] = 0 \text{ a.s.}$$

on $\{\tau < \infty\}$ for $i \in S, n \geq 0$. □

Impose the additional condition:

$$(\dagger) \qquad \sup E[\tau(0)^2/X_0 = 0] < \infty,$$

where the supremum is over all admissible control strategies.

Remark To see that (†) is not implied by stability alone, conside a chain on $\{(0,0),(0,1),(1,1),(0,2),(1,2),(2,2),(0,3),\ldots\}$ with transition probabilities: for $j \geq 1$,

$$\begin{aligned} P((i,j)/(i+1,j)) &= 1, 0 \leq i < j, \\ P((0,0)/(0,j)) &= 1, \\ P((j,j)/(0,0)) &= Cj^{-3}, \text{ where } C = \left(\sum_n n^{-3}\right)^{-1}. \end{aligned}$$

Then $E[\tau(0)/X_0 = 0] < \infty$, but $E[\tau(0)^2/X_0 = 0] = \infty$.

Lemma 2.12 Under (†), $\{\nu_n\}$ are a.s. tight in $\mathcal{P}(S \times U)$.

Proof Define stopping times $\tau_0 = 0, \tau_{n+1} = \min\{m > \tau_n : X_m = 0\}, n \geq 0$. By (†) and Lemma 2.11, it follows that

$$\sup_n E[(\tau_{n+1} - \tau_n)^2] < \infty.$$

By the strong law of large numbers for martingales (valid because of (†)), we then have, for $N \geq 1$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left(\sum_{m=\tau_i}^{\tau_{i+1}-1} I_{\{X_m \geq N\}} - E\left[\sum_{m=\tau_i}^{\tau_{i+1}-1} I_{\{X_m \geq N\}} / \mathcal{F}_{\tau_i} \right] \right) = 0 \text{ a.s.}$$

Thus

$$\begin{aligned} \limsup_{n\to\infty} \nu_n(\{N, N+1, \ldots\} \times U) &= \limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} I_{\{X_m \geq N\}} \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left(\sum_{m=\tau_i}^{\tau_{i+1}-1} I_{\{X_m \geq N\}} \right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} E\left[\sum_{m=\tau_i}^{\tau_{i+1}-1} I_{\{X_m \geq N\}} / \mathcal{F}_{\tau_i} \right] \\ &\leq \sup E\left[\sum_{m=0}^{\tau(0)-1} I_{\{X_m \geq N\}} / X_0 = 0 \right], \end{aligned}$$

where the supremum is over all admissible strategies and the last inequality holds by Lemma 2.12. Fix a sample point for which the foregoing holds true for all $N \geq 1$. A standard dynamic programming argument allows us to replace the above supremum by the supremum over Π_{SR}, which is turn equals

$$\sup_{f\in\Pi_{SR}} \left(E[\tau(0)/X_0 = 0] \left(\sum_{i\geq N} \pi_f(i) \right) \right),$$

with π_f = the unique invariant probability under $f \in \Pi_{SR}$. By (†), this is bounded by a constant times $\sup_{f\in\Pi_{SR}} \left(\sum_{i\geq N} \pi_f(i) \right)$. This can be made arbitrarily small by increasing N in view of the compactness (hence tightness) of G. The claim follows. □

Corollary 2.1 For the stable case, under (†), (2.5) holds.

Proof This follows as in Lemma 2.9 on observing that, outside a zero probability set, any limit point $\overline{\nu} = \delta\nu' + (1-\delta)\nu''$ of $\{\nu_n\}$ as in Lemma 2.9 must have $\delta = 1$ by virtue of the above lemma. □

3 Extensions

3.1 Other cost criteria

Recall the infinite horizon discounted cost control problem wherein we seek to minimize $E\left[\sum_{m=0}^{\infty} \beta^m k(X_m, Z_m)\right]$ for a discount factor $\beta \in (0,1)$. This is equivalent to minimizing

$\int kd\mu$ for the 'discounted occupation measure' $\mu \in \mathcal{P}(S \times U)$ defined by

$$\int hd\mu = (1-\beta)E\left[\sum_{m=0}^{\infty} \beta^m h(X_m, Z_m)\right], h \in C_b(S \times U).$$

In particular, if $f \in \Pi_{SR}$ is being used, one can disintegrate μ as $\mu(\{i\}, du) = \pi(i)f(i, du)$. Note that $\pi \in \mathcal{P}(S)$ is given by

$$\pi = (1-\beta)\nu\left(\sum_{m=0}^{\infty} \beta^m P(f)^m\right) \tag{3.1}$$

where $P(f)$ is the transition matrix corresponding to f and ν the law of X_0. (Both π and ν are displayed here as row vectors.) From (3.1), it follows that π satisfies

$$\pi = (1-\beta)\nu + \beta\pi P(f). \tag{3.2}$$

The corresponding μ then satisfies

$$\sum_i h(i)\mu(\{i\}, U) = (1-\beta)\sum_i h(i)\nu(i) + \beta\sum_{i,j}\int h(j)P(j/i, y)\mu(\{i\}, dy). \tag{3.3}$$

That π is the unique solution in $\mathcal{P}(S)$ to (3.2) for given ν follows by iterating (3.2) to recover (3.1). Similarly, a $\mu \in \mathcal{P}(S \times U)$ satisfying (3.3) can be disintegrated as $\mu(\{i\}, dy) = \pi(i)f(i, dy)$ to deduce (3.2) from (3.3), whence it follows that (3.3) characterizes μ as the discounted occupation measure under the corresponding $f \in \Pi_{SR}$ and initial law ν. An immediate consequence of this is:

Lemma 3.1 For any admissible $\{Z_n\}$, the corresponding discounted occupation measure is also the discounted occupation measure for some $f \in \Pi_{SR}$ for the same initial law.

Proof Let ν denote the initial law and π the corresponding discounted occupation measure under $\{Z_n\}$. Disintegrate π as

$$\pi(\{i\}, du) = \nu(i)\varphi_i(du), i \in S,$$

and identify the map $\Phi : i \in S \to \varphi_i(du) \in \mathcal{P}(U)$ with an element of Π_{SR}, denoted by Φ again. Let $\{\tilde{X}_n\}$ be a chain governed by Φ and $\{\tilde{Z}_n\}$ the corresponding control sequence. Let $f \in C_b(S \times U)$ and define

$$g(i) = E\left[\sum_{m=0}^{\infty} \beta^m f(\tilde{X}_m, \tilde{Z}_m)/\tilde{X}_n = i\right], \quad i \in S.$$

Define $Y_0 = g(X_0)$ and for $n \geq 1$,

$$Y_n = \sum_{m=0}^{n-1} \beta^m f(X_m, Z_m) + \beta^n g(X_n),$$

$$\begin{aligned} W_n &= Y_{n+1} - Y_n \\ &= \beta^n f(X_n, Z_n) + \beta^{n+1} g(X_{n+1}) - \beta^n g(X_n). \end{aligned}$$

Then

$$\begin{aligned} &E \left[\sum_{m=0}^{n} \beta^m f(X_m, Z_m)\right] - E[g(X_0)] \\ &\quad = E\left[\sum_{m=0}^{n} W_m\right] - \beta^{n+1} E[g(X_{n+1})] \\ &\quad = E\left[\sum_{m=0}^{n} E[W_m / X_j, Z_j, j \leq m]\right] - \beta^{n+1} E[g(X_{n+1})] \\ &\quad = E\left[\sum_{m=0}^{n} \beta^m \left[f(X_m, Z_m) + \beta \sum_j P(j/X_m, Z_m) g(j) - g(X_m)\right]\right] - \beta^{n+1} E[g(X_{m+1})]. \end{aligned}$$

Let $n \to \infty$ to obtain

$$\begin{aligned} &E \left[\sum_{m=0}^{\infty} \beta^m f(X_m, Z_m)\right] - E[g(X_0)] \\ &\quad = E\left[\sum_{m=0}^{\infty} \beta^m \left[f(X_m, Z_m) + \beta \sum_j P(j/X_m, Z_m) g(j) - g(X_m)\right]\right] \\ &\quad = E\left[\sum_{m=0}^{\infty} \beta^m \left[\int f(X_m, u) \varphi_{X_m}(du) + \beta \sum_j g(j) \int P(j/X_m, u) \varphi_{X_m}(du) - g(X_m)\right]\right] \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(by our construction of } \Phi) \\ &\quad = 0, \end{aligned}$$

because $g(i)$ satisfies

$$g(i) = \int f(i, u) \varphi_i(du) + \beta \sum_j g(j) \int P(j/i, u) \varphi_i(du), \quad i \in S.$$

Therefore, if we choose the law of $\tilde{X}_0$ to be ν, we have

$$\begin{aligned} E\left[\sum_{m=0}^{\infty} \beta^m (X_m, Z_m)\right] &= E[g(X_0)] \\ &= E[g(\tilde{X}_0)] \\ &= E\left[\sum_{m=0}^{\infty} \beta^m f(\tilde{X}_m, \tilde{X}_m)\right]. \end{aligned}$$

Multiplying both sides by $(1 - \beta)$, the claim follows in view of our arbitrary choice of $f \in C_b(S \times U)$. □

This allows us to restrict our attention to strategies in Π_{SR} a priori. Fix $\nu \in \mathcal{P}(S)$. Let μ_f denote the discounted occupation measure corresponding to $f \in \Pi_{SR}$, written as $\mu_f(\{i\}, dy) = \pi_f(i) f(i, dy)$, and let $G = \{\mu_f : f \in \Pi_{SR}\}$.

Lemma 3.2 G is compact convex in $\mathcal{P}(S \times U)$.

Proof That it is closed convex follows along the lines of Lemma 2.2. Compactness follows from tightness, which is a straightforward consequence of the easily verified tightness of the laws of $[X_0, X_1, \ldots]$ in $\mathcal{P}(S^\infty)$ as $\{Z_n\}$ varies over all admissible strategies. (See, e.g., [9], pp. 26-27). □

Now, let $j \in$ support (π_f) be such that for some $f' \neq f''$ in Π_{SR} and $a \in (0,1), f'(j) \neq f''(j), f'(i) = f''(i)\ \forall\, i \neq j$ and $f(i) = af'(i) + (1-a)f''(i)\ \forall\, i$. Since the passage from f to f' or f'' changes only the transition probabilities for leaving state j, one still has $j \in$ support $(\pi_{f'}) \cap$ support $(\pi_{f''})$. This is the counterpart of Lemma 2.3, which allows us to establish the following along the lines of Lemma 2.4.

Lemma 3.3 Extreme points of G correspond to $f \in \Pi_{SS}$.

The equivalent convex programming problem is to minimize $\int k d\mu$ over $\mu \in G = \{\mu :$ (3.3) holds $\}$. As in the 'stable case' above, existence of an optimal $f \in \Pi_{SS}$ is now ensured.

Similar treatments for the finite horizon and exit time problems are possible, with the latter requiring a suitable additional hypothesis to ensure that the mean exit time is bounded. There is, however, one important difference with the finite horizon problem: One has to replace Π_{SS} by $\Pi_{MS} \triangleq$ the set of Markov strategies in Lemma 3.3 and the remark that follows it, where a Markov strategy is a strategy of the form $X_n = v(X_n, n)$ for a measurable $v : S \times \{0, 1, \ldots, N\} \to U$. See [9], Chapter IV, for details.

An important caveat here is that in the foregoing, it is not obvious that an optimal $f \in \Pi_{SS}$ (or Π_{MS} as the case may be) can be found which is optimal regardless of the choice of ν. That it can indeed be falls out naturally from the traditional dynamic programming approach, but not in the convex analytic formulation above.

There is also an alternative approach to these criteria that treats them as special cases of the ergodic problem. For the discounted problem, (3.3) turns out to be the counterpart of (2.1) for the ergodic control of a modified chain that is reset to law ν at i.i.d. times $\{\tau_i\}$ distributed geometrically with parameter β. The intricacies of this approach do not seem warranted in discrete time control problems, where the above treatment is much easier. In continuous time problems like controlled diffusions, however, such embedding into ergodic control for discounted, finite horizon and exit time problems have been effectively used in [5], [21].

Finally, note that $\beta \to 1$ in (3.3) leads to (2.1). This leads to the conclusion that any limit point in $\mathcal{P}(\overline{S} \times U)$ of a sequence $\{\mu_n\}$ of discounted occupation measures corresponding to $\{\beta(n)\} \subset (0,1)$ with $\beta(n) \to 1$, must be of the form $\delta\nu' + (1-\delta)\nu''$, where ν' is an ergodic occupation measure, ν'' is supported on $\{\infty\} \times U$ and $\delta \in [0,1]$. In both near-monotone and stable cases, one can then use familiar arguments (from section 2) to conclude that as $\beta \to 1$, optimal discounted occupation measures tend to the set of optimal ergodic occupation measures. (The latter case also needs (†), which ensures that the laws of $\{X_n\}$ and therefore the discounted occupation measures, remain in a tight set for a prescribed initial law, regardless of the choice of $\{Z_n\}$ and as $\beta \to 1$.) A similar development is possible for the finite horizon problem in the limit as the horizon becomes infinite.

3.2 General state spaces

In this section, we briefly outline extensions to more general, viz., Polish S, concentrating on the ergodic control problem. Denote by $p(x,u,dy) \in \mathcal{P}(S)$ the transition kernel for $\{X_n\}$ with $x \in S, u \in U$. This is assumed to be continuous in (x,u). A $q \in \Pi_{SR}$ will be given by a probability kernel $x \in S \to q(x,dy) \in \mathcal{P}(U)$, with the corresponding transition kernel given by

$$\overline{p}(x,q,dy) \triangleq \int_U q(x,du)p(x,u,dy)$$

If $\eta(dx) \in \mathcal{P}(S)$ is invariant under q, define the corresponding ergodic occupation measure by

$$\nu(dx,du) = \eta(dx)q(x,du). \tag{3.4}$$

The set G of ergodic occupation measures in $\mathcal{P}(S \times U)$ is then characterized by the fact that their disintegration (3.4) satisfies

$$\int_S \int_U \eta(dx)q(x,du)p(x,u,dy) = \eta(dy). \tag{3.5}$$

It is easy to deduce from this that G is closed. Also, if $\nu_i(dx,du) = \eta_i(dx)q_i(x,dy) \in G, i = 1,2$, then

$$\nu(dx,dy) = a\nu_1(dx,dy) + (1-a)\nu_2(dx,dy), a \in (0,1),$$

satisfies (3.5) with $\eta(dx) = a\eta_1(dx) + (1-a)\eta_2(dx)$ and

$$q(x,dy) = \Lambda(x)q_1(x,dy) + (1-\Lambda(x))q_2(x,dy)$$

for $\Lambda(x) = a\frac{d\eta_1}{d\eta}(x)$. Thus G is convex.

Assume as before that $k : S \times U \to R$ is continuous and bounded from below. The definition of 'stable case' above extends to general S in the obvious manner. As for near-monotonicity, the definition can be retained if S is locally compact. If not, one must allow

$k(\cdot,\cdot)$ to be extended real valued and suppose that the set $D_r = \{x : \inf_a k(x,a) \leq r\}$ is compact for each $r > 0$. The reason $+\infty$ must be allowed as a value for $k(\cdot,\cdot)$ is as follows: By Baire category theorem, S that is not locally compact cannot be σ-compact and thus cannot equal $\cup_r D_r$.

Under either condition, a cost-minimizing sequence in G can be shown to be relatively sequentially compact and one can mimick the arguments of section 2 to conclude that the map $\mu \in G \rightarrow \int k d\mu$ attains its minimum on G, hence on G_e.

If S is not compact but locally compact, we can embed it homeorphically into its one point compactification $\overline{S} = S \cup \{\infty\}$. Define empirical measures $\nu_n \in \mathcal{P}(\overline{S} \times U)$ by

$$\nu_n(A \times B) = \frac{1}{n} \sum_{m=0}^{n-1} I_{\{X_m \in A, Z_m \in B\}}, \quad n \geq 1,$$

where A, B are Borel in $\overline{S}, U$ resp. Let $\overline{G} = \{\mu \in \mathcal{P}(\overline{S} \times U) : \mu = a\mu_1 + (1-a)\mu_2$ for some $a \in [0,1], \mu_1 \in G, \mu_2(\{\infty\} \times A) = 1\}$.

Lemma 3.4 $\nu_n \rightarrow \overline{G}$ a.s.

This is proved as in Lemma 2.10, with $\mathcal{C}$ any countable convergence determining class in $C_b(S)$ (see [12], Chapter 2). For the near-monotone case, Lemma 2.2 now follows as before. The stable case is harder. Assume S to be locally compact as before and let B_1, B_2 be concentric closed balls of radii $r_1 < r_2$ resp. in S. Define

$$\begin{aligned}
\sigma_1 &= \min\{m \geq 0 : X_m \in \overline{B_2^c}\}, \\
\tau_1 &= \min\{m > \sigma_1 : X_m \in B_1\}, \\
\sigma_{n+1} &= \min\{m > \tau_n : X_m \in \overline{B_2^c}\}, \\
\tau_{n+1} &= \min\{m > \sigma_n : X_m \in B_1\},
\end{aligned}$$

for $n \geq 1$. Assume that for any open $B \subset S, P(\cup_{n \geq 1}\{X_n \in B\}) = 1$ regardless of the initial condition or control strategy. (This is a 'recurrence' condition.) Then $\tau_n, \sigma_n < \infty$ a.s. $\forall\ n$. The counterpart of (†) now is

$$\sup_{x \in B_1} \sup_{\Pi} E[\tau_1^2 / X_0 = x] < \infty, \tag{3.6}$$

where the first supremum is over $\Pi \triangleq$ the set of all control strategies.

Let $D_N, N \geq 1$, be a concentric family of closed balls in S with radii $N \geq 1$ respectively. Now mimick the proof of Lemma 2.12 to conclude that

$$\limsup_{n \rightarrow \infty} \nu_n(D_N) \leq \sup_{x \in B_1} \sup_{\Pi} E\left[\sum_{m=0}^{\tau_1} I_{\{X_m \notin D_N\}} / X_0 = x\right] \text{ a.s.}$$

$$= \sup_{x \in B_1} \sup_{\Pi_{SR}} E\left[\sum_{m=0}^{\tau_1} I_{\{X_m \notin D_N\}} / X_0 = x\right] \tag{3.7}$$

where the equality follows by a standard dynamic programming argument. Now let $f_m \to f_\infty$ in $\Pi_{SR}, x_m \to x_\infty$ in B_1 and let $X_n^m, n \geq 1$, denote the chain governed by f_m with $X_0^m = x_m, m = 1, 2, \cdots, \infty$. Standard arguments along the lines of [9], pp. 26-28, show that $\{X_n^m\} \to \{X_n^\infty\}$ in law. By Skorohod's theorem ([12], pp.23-24), we may view these processes as being defined on a common probability space and the convergence as being a.s. Define stopping times $\{\tau_m^n, n \geq 1\}, m \geq 1$, correspondingly.

Then

$$\liminf_{m \to \infty} \sum_{\ell=0}^{\tau_1^m} I_{\{X_l^m \notin D_N\}} \geq \sum_{\ell=0}^{\tau_1^\infty} I_{\{X_\ell^\infty \notin D_N\}}$$

a.s. and therefore

$$\liminf_{m \to \infty} E\left[\sum_{\ell=0}^{\tau_1^m} I_{\{X_\ell^m \notin D_N\}}\right] \geq E\left[\sum_{\ell=0}^{\tau_1^\infty} I_{\{X_\ell^\infty \notin D_N\}}\right].$$

Thus the map

$$(x, f) \to E_f\left[\sum_{m=0}^{\tau_1} I_{\{X_\ell^m \notin D_N\}} / X_0 = x\right],$$

with $E_f[\cdot]$ denoting the expectation under $f \in \Pi_{SR}$, is lower semicontinuous. As $N \uparrow \infty$, the r.h.s. $\downarrow 0$. By Dini's theorem, this convergence is uniform in $(x, f) \in B_1 \times \Pi_{SR}$. Thus the r.h.s. of (3.7) can be made arbitrarily small by choosing N sufficiently large. We have proved:

Lemma 3.5 Under (3.6), $\{\nu_n\}$ remains tight a.s. under arbitrary control strategies.

Corollary 3.1 The conclusion of Lemma 2.10 holds in the present set-up.

This follows exactly as for countable S in view of the foregoing. Given that the optimum, when it exists, is attained on G_e = the set of extreme points of G, we shall characterize G_e next, albeit for a special case. We assume in addition to local compactness the following condition: There exists a σ-finite nonnegative measure λ on S such that $p(x, u, dy) << \lambda(dy) \, \forall \, x, u$ and furthermore, if $\varphi(x, u, y)$ denote the corresponding density (i.e., $p(x, u, dy) = \varphi(x, u, y)\lambda(dy)$), then $\varphi(\cdot, \cdot, \cdot)$ is continuous, and $\{\varphi(x, u, \cdot) : x \in S, \; u \in U\}$ bounded equicontinuous and bounded away from zero from below uniformly an compacts. (These conditions are satisfied, e.g., for several stochastic systems in R^d with additive Gaussian white noise).

This has the following important consequence: If η is an invariant probability distribution under $q \in \Pi_{SR}$, then

$$\eta(dy) = \int_S \eta(dx)q(x,du)\varphi(x,u,y)dy,$$

implying that η has a density w.r.t. λ given by

$$\psi(y) = \int_S \eta(dx)q(x,du)\varphi(x,u,y) > 0.$$

Then any two ergodic probability distributions under q are mutually absolutely continuous, hence identical. That is, if q admits an invariant probability distribution, it is unique and the corresponding $\{X_n\}$ ergodic. Suppose now that G is compact.

Lemma 3.6 $Q \triangleq \{\eta \in \mathcal{P}(S) : \eta$ invariant under some $q \in \Pi_{SR}\}$ is compact in the total variation norm topology.

Proof Let $\eta_n \in Q$ be invariant under $q_n \in \Pi_{RS}, n \geq 1$. Then $\nu_n(dx,dy) \triangleq \eta_n(dx)q_n(x,dy) \in G \,\forall\, n$. Since G is compact, $\nu_n \to \nu_\infty$ (say) in $\mathcal{P}(S \times U)$ along a subsequence, denoted by $\{\nu_n\}$ again by abuse of notation. Then $\eta_n \to \eta_\infty$ in $\mathcal{P}(S)$. Let $\psi_n = d\eta_n/d\lambda$ for $n = 1, 2, \ldots, \infty$. Under our assumptions, $\{\psi_n(\cdot)\}$ are equicontinuous and bounded. By the Arzela-Ascoli theorem, we may drop to a subsequence if necessary and suppose that $\psi_n(\cdot) \to \overline{\psi}(\cdot)$ in $C(S)$. Thus for compactly supported $f \in C(S)$,

$$\int f(y)\psi_n(y)\lambda(dy) \to \int f(y)\overline{\psi}(y)\lambda(dy).$$

But

$$\int f(y)\psi_n(y)\lambda(dy) = \int f d\eta_n \to \int f d\eta_\infty.$$

Thus $\overline{\psi} = \psi_\infty$ and $\psi_n \to \psi_\infty$ in $C(S)$. By Scheffe's theorem ([12], pp.26), $\eta_n \to \eta_\infty$ in the total variation norm. This completes the proof. □

We shall now characterize the set G_e for the stable case, i.e., when all $f \in \Pi_{SR}$ are stable and the corresponding G compact. Let $q_i \in \Pi_{SR}, i = 0, 1, 2$, be such that

$$q_0(x,dy) = aq_1(x,dy) + (1-a)q_2(x,dy), q_1 \neq q_2,$$

with $a \in (0,1)$. Let $\eta_i, \nu_i, i = 0, 1, 2$, denote the corresponding invariant distributions and ergodic occupation measures resp., with $\psi_i = d\eta_i/d\lambda$, $i = 0, 1, 2$. Finally, let $\tilde{Q} = \{\psi : S \to R : \psi = d\eta/d\lambda$ for some $\eta \in Q\}$.

Lemma 3.7 $\nu_0 \notin G_e$.

Proof By setting $q_i(x,dy) = q_0(x,dy), i = 1, 2$, for x outside a ball $B_R \subset S$ of a sufficiently large radius $R > 0$, we may assume without loss of generality that $q_1(x,dy) = q_2(x,dy)$ for

$x \notin B_R$. Let $\tilde{q} \in \Pi_{SR}$, with $\tilde{\eta}(dx) = \tilde{\psi}(x)\lambda(dx)$ the corresponding invariant measure and $\tilde{\nu}(dx, dy) = \tilde{\eta}(dx)\tilde{q}(x, dy)$ the corresponding ergodic occupation measure. Define $\hat{q} \in \Pi_{SR}$ by

$$q_0(x, dy) = \frac{b\psi_1(x)q_1(x, dy) + (1-b)\tilde{\psi}(x)\hat{q}(x, dy)}{b\psi_1(x) + (1-b)\tilde{\psi}(x)} \tag{3.8}$$

with $b \in (0, 1)$. In order that this indeed define a $\tilde{q} \in \Pi_{SR}, b$ must be chosen so that

$$\frac{b\psi_1(x)}{b\psi_1(x) + (1-b)\tilde{\psi}(x)} \leq a \quad \forall x \in B_R.$$

(For $x \notin B_R, q_0(x) = q_1(x)$ and any choice of b will do.) For $x \in B_R$, such a choice of b is possible provided

$$\inf_{\Pi_{SR}, x \in B_R} \{\psi(x) : \psi \in \tilde{Q}\} > 0.$$

This is indeed so by our assumptions: If not, we have $\psi_n \in \tilde{Q}, x_n \in B_R, n \geq 1$, with $\psi_n(x_n) \to 0$. By dropping to a subsequence if necessary and using the Arzela-Ascoli theorem, let $\psi_n(\cdot) \to \psi_\infty(\cdot)$ in $C(S), x_n \to x_\infty$ in B_R. Then $\psi_\infty \in \tilde{Q}$, implying $\psi_\infty(x_\infty) > 0$, a contradiction.Thus we can choose a $b \in (0, 1)$ as desired. Fix one such b. Then (3.8) defines a map $\tilde{q} \to \hat{q}$, or equivalently, $\tilde{\eta} \in Q \to \hat{\eta} \in Q, \hat{\eta}$ being the invariant distribution under $\hat{q}$. This map is continuous in the total variation norm: If $\eta_n(dx) = \psi_n(dx)\lambda(dx) \to \eta_\infty(dx) = \psi_\infty(x)\lambda(dx)$ in total variation, $\psi_n(\cdot) \to \psi_\infty(\cdot)$ in $C(S)$ as in the proof of Lemma 3.6. Letting $\hat{\psi}_n$ denote the image of ψ_n under the above map for $n = 1, 2, \ldots, \infty$, it is easily verified from (3.8) that $\hat{\psi}_n \to \hat{\psi}_\infty$ pointwise and hence, by Scheffe's theorem, $\hat{\psi}_n(x)\lambda(dx) \to \hat{\psi}_\infty(x)\lambda(dx)$ in total variation. Since Q is compact convex, Schauder fixed point theorem guarantees a fixed point for this map. That is, there exists a $q^* \in \Pi_{SR}$ with associated invariant distribution $\eta^*(dx) = \psi^*(x)\lambda(dx)$ such that

$$q_0^*(x, dy) = \frac{b\psi_1(x)q_1(x, dy) + (1-b)\psi^*(x)q^*(x, dy)}{b\psi_1(x) + (1-b)\psi^*(x)}$$

Since $q_0 \neq q_1, q_0 \neq q^*$. It is easily deduced from this that for $\nu^*(dx, dy) = \eta^*(dx)q^*(x, dy)$,

$$\nu_0 = b\nu_1 + (1-b)\nu^*, \ \nu_1 \neq \nu^*,$$

i.e., $\nu_0 \notin G_e$. □

If G is not compact, we need the following additional condition, called the 'stability under local perturbations': If $q(x, dy), q'(x, dy) \in \Pi_{SR}$ agree for x outside a bounded subset of S and one of them is stable, so is the other. Note that G is closed for the same reasons as before. Let $q_i \in \Pi_{SR}, i = 0, 1, 2,$ be as before with $\nu_i, i = 0, 1, 2,$ defined correspondingly as before.

Theroem 3.1 $\nu_0 \notin G_e$.

The proof is essentially the same as before : Recall from the proof of Theorem 3.1 that for any stable $q \in \Pi_{SR}$, it suffices to look at the subset of G corresponding to the ergodic occupation measures for $\tilde{q} \in \Pi_{SR}$ that agree with q outside a ball B_R of radius $R > 0$ sufficiently large. By our assumption of 'stability under local perturbations', all such $\tilde{q}$ are stable, so the same proof works.

Corollary 3.2 The conclusion of Lemma 3.3 holds under the conditions of Lemma 3.7, 3.8.

4 Multiobjective problems

4.1 Constrained control: preliminaries

The first mutliobjective problem we consider is that of minimizing one cost with other secondary costs being required to satisfy prescribed bounds. Again we stick to the 'ergodic' set up of section 2 and shall use the notation therein. We start with some convex analytic preliminaries.

Recall the set G of ergodic occupation measures. View it as a subset of the space of finite signed measures on S. Let H be its intersection with $m \geq 1$ closed half spaces thereof. Then H is closed convex. Let H_e be the set of its extreme points. The following is a special case of a result of Dubins [17].

Lemma 4.1 Any $\nu_0 \in H_e$ can be expressed as a convex combination of k points in G_e for some $k \leq m + 1$.

Proof Suppose G is compact. Suppose the claim is false and ν_0 is expressible as a convex combination of $k = m + 2$ points in G, but not less. (A similar proof works for higher k.) Then ν_0 lies in the interior of an $(m + 2)$-simplex B formed by these points in G_e. Let $M =$ the $(m+1)$-dimensional affine space (i.e., translate of a linear subspace) generated by B, and C an open ball in M centered at ν_0 and contained in the interior of B. Then $C \subset B \subset G$. Now consider the intersections with M of the m hyperplanes that form the boundaries of the half spaces defining H. Since at most m of them can intersect at a time, any intersection with M of the intersections of these hyperplanes must have a codimension of at most m in M and thus cannot have a corner in the interior of C. Thus $\nu_0 \notin H_e$, a contradition. For noncompact G, argue as above with $\bar{G}_e$ in place of G_e and observe that if ν_0 is a strict convex combination of points in $\bar{G}_e$, the latter must be in G_e, because ν_0 would otherwise assign a strictly positive probability to $\{\infty\} \times U$, a contradiction. The rest is as before.

In case ν_0 cannot be expressed as a convex combination of finitely many extreme points of G, a simple adaptation of the above proof works. We claim that for any $j \geq 1$, we can find j linearly independent finite line segments in G which have ν_0 at their center. If this were not so for, say, $j = j_0 + 1$, then ν_0 would be in a j_0-dimensional face G' of G and therefore expressible as a convex combiantion of j_0+1 extreme points of G' by Caratheodory's theorem ([14], p.106), therefore of G. This goes against the hypothesis, proving the claim. Now take $j \geq m + 2$, consider the polytope generated by the end points of these line segments, and argue as before. □

Write ν_0 as

$$\nu_0(\{i\}, dy) = \pi_0(i) f(i, du), \ i \in S,$$

corresponding of $f \in \Pi_{SR}$. By the above lemma, ν_0 is a linear combination of $\nu_1, \cdots, \nu_k, k \leq m + 1$, in G_e with strictly positive weights. Letting $\delta_x(du)$ denote the Dirac measure at x, we may disintegrate ν_j's as

$$\nu_j(\{i\}, du) = \pi_j(i)\delta_{\xi_{ij}}(du), i \in S, 1 \leq j \leq k.$$

By abuse of notation, let $\xi_j : i \to \delta_{\xi_{ij}}$ denote the corresponding stationary strategy. Also, by the ergodic decomposition of the chain, $\pi_0 = \sum_{i=1}^{n} a_i \tilde{\pi}_i$ where $a_i \in [0, 1]$ with $\sum a_i = 1$ and $\{\tilde{\pi}_i\}$ are ergodic invariant probability measure under f. In particular, $\tilde{\pi}_i$'s have disjoint supports. We denote these by $\{S_i\}$ respectively.

Lemma 4.2 For $1 \leq j \leq k$, support $(\pi_j) \subset S_i$ for some i.

Proof If not, two states in two distinct S_i's would communicate with each other under the stationary policy ξ_j, hence under f(cf. the proof of Lemma 2.4), a contradiction. □

For $i \in S$ with $\pi_0(i) > 0$, let $N(i) = \{u \in U : u = \xi_{ij}$ for some $j, 1 \leq j \leq k$, satisfying $\pi_j(i) > 0\}$. Let $n(i) = |N(i)| - 1$. Then for each i, $n(i)$ is the 'number of randomizations at i' for f. To understand this terminology, observe that f_i will be of the form

$$f(i, dy) = \sum_{\ell=1}^{|N(i)|} \left(a'_\ell \pi'_\ell(i) \xi'_\ell(i, dy) / \left[\sum_{n=1}^{|N(i)|} a_n \pi'_n(i) \right] \right) \tag{4.1}$$

where $\xi'_l(i, dy)$ are Dirac (cf. proof of Lemma 2.2) with ergodic distributions π'_ℓ and $\{a'_j\}$ satisfy $a'_j > 0$, $\sum_\ell a'_\ell = 1$. Thus $|N(i)| = 2$ corresponds to randomizing between two choices, i.e., a single randomization, and so on. By convention, $n(i) = 0$ if $N(i) = \phi$.

Pick $i \in S$ (if any) such that $\pi_0(i) > 0, n(i) > 0$. Let $N(i) = \{u(1), u(2), \cdots, u(n(i)+1)\}$. Then $f(i, dy)$ is a strict convex combination of the Dirac measures at $u(j)$'s. Define $\psi^{ij} \in \Pi_{SR}$ by:

$$\begin{aligned} \psi^{ij}(\ell, dy) &= f(\ell, dy), \ \ell \neq i, \\ &= \delta_{u(j)}(dy), \ \ell = i, \end{aligned}$$

Lemma 4.3 ν_0 is a strict convex combination of distinct elements $\mu_1, \cdots, \mu_{n(i)+1}$ of G, where μ_j is an ergodic occupation measure under ψ^{ij} for each j. Furthermore, $\mu_1, \cdots, \mu_{n(i)+1}$ form the corners of an $(n(i)+1)$ simplex.

Proof If π_0 is ergodic, the first claim follows by iterating the argument used in the proof of Lemma 2.4. If not, pick the $\tilde{\pi}_\ell$ for which $\tilde{\pi}_\ell(i) > 0$ and apply the same argument. Suppose the second claim is false. Then ν_0 can be expressed as a strict convex combination of elements from $\{\mu_1, \cdots, \mu_{n(i)+1}\}$ in at least two distinct ways. But then (4.1) applied to the present situation allows us to write a finitely supported probability measure $f(i, dy)$ as a strict convex combination of distinct Dirac measures in two different ways, an impossibility. The claim follows. □

Call this $(n(i)+1)$-simplex the 'perturbution simplex' at i, denoted by $Q(i)$.

Lemma 4.4 Let $\nu_1 \neq \nu_2$ in $Q(i)$, with

$$\nu_j(\{l\}, du) = \pi_j(l)\varphi_{jl}(du), \quad l \in S, j = 1, 2.$$

Let $\pi_j(l) > 0$. Then $\varphi_{1l} = \varphi_{2l}$ for $l \neq i$ and $\varphi_{1i} \neq \varphi_{2i}$.

Proof The claim for $l \neq i$ is immediate from (4.1) and our definition of $Q(i)$. That $\varphi_{1i} \neq \varphi_{2i}$ follows from (4.1) and the easily verifiable fact: for $n > 1, b_1, \ldots, b_n > 0$, the map

$$\begin{aligned} &[a_s, \ldots, a_n] \in \{[x_1, \ldots, x_n] : 0 < x_i < 1 \ \forall \ i, \sum_i x_i = 1\} \rightarrow \\ &[a_1 b_1 / c, a_2 b_2 / c, \ldots, a_n b_n / c] \in (0, 1)^n, \end{aligned}$$

with $c = \sum_{i=1}^n a_i b_i$, is one-one. □

Let $Y(i) =$ the $n(i)$-dimensional affine space, in the space of finite signed measures on $S \times U$, spanned by $Q(i)$ for i as above.

Lemma 4.5 If $j \neq i$ satisfies $\pi_0(j) > 0$, $n(j) > 0$, then $Y(i) \cap Y(j) = \{\nu_0\}$.

Proof Suppose not. Then there must exist a $\overline{\nu} \neq \nu_0, \overline{\nu} \in Q(i) \cap Q(j)$. Let Z be the line segment joining $\nu_0, \overline{\nu}$. Then $Z \subset Q(i) \cap Q(j)$. Write a typical $\nu \in Z$ as

$$\nu(\{l\}, du) = \pi(l)\varphi_l(du), l \in S.$$

Note that for $\nu \neq \overline{\nu}$ in Z, support $(\pi) \subset$ support (π_0). As ν moves along Z, Lemma 4.4 and the fact that $Z \subset Q(i)$ implies $\varphi_j(\cdot) = f(j, \cdot)$ all along. Interchange the roles of i, j to conclude $\varphi_i(\cdot) = f(i, \cdot)$ all along, a contradiction to $\overline{\nu} \neq \nu_0$ in view of Lemma 4.4. The claim follows. $\square$

Lemma 4.6 Let $i_1, i_2, \ldots, i_{r+1}$ be different states in S with $n(i_j) > 0$ for all j and $\alpha_1, \alpha_2, \ldots, \alpha_r \in [0, 1]$ with $\sum_{i=1}^r \alpha_i = 1$. Then

$$\{\alpha_1 Q(1) + \alpha_2 Q(2) + \cdots + \alpha_1 Q(r)\} \cap Q(i_{r+1}) = \{\nu_0\}.$$

Proof For $r = 1$, this reduces to the preceding lemma. The general claim follows by induction using an argument similar to that above at each state. $\square$

Let $L(i) = Y(i) - \nu_0$ when $n(i) > 0$. (That is, $L(i)$ is $Y(i)$ translated by ν_0.)

Corollary 4.1 $dim(L(i_1) + \cdots + L(i_r)) = \sum_{n=1}^r dim(L(i_n))$.

This is immediate from the preceding lemma. Thus $L(i_1) + L(i_2) + \cdots + L(i_r)$ is a direct sum.

Lemma 4.7 $\sum_i n(i) \leq m$.

Proof If not, for some $\{i_1, \ldots, i_l\} \subset S$, $\sum_{j=1}^{\ell} n(i_\ell) \geq m + 1$. By the foregoing, corners of $Q(i_1), \ldots, Q(i_l)$ together form a polytope with $\sum_{m=1}^{l} n(i_m) \geq m + 1$ dimensional relative interior that contains ν_0. Now argue as in the proof of Lemma 4.1 to get a contradiction. $\square$

This completes the convex analytic preliminaries for the constrained control problem.

4.2 Constrained control : main results

The constrained control problem is

$$\text{Minimize} \int k_0 d\nu \tag{4.2}$$

subject to

$$\nu \in G, \beta_i \leq \int k_i d\nu \leq \alpha_i, 1 \leq i \leq m, \tag{4.3}$$

where $k_i : S \times U \to [0, \infty)$ are prescribed continuous functions and α_i, β_i prescribed nonnegative scalars for $0 \leq i \leq m$. Let $H = \{\nu \in G : \beta_i \leq \int k_i d\nu \leq \alpha_i, 1 \leq i \leq m\}$, assumed nonempty. Also assume that

$$\alpha_0 \triangleq \inf_{\nu \in H} \int k_0 d\nu < \infty.$$

As before, we consider two cases:

Case 1: Near-monotone case

$$\beta_i = 0 \ \forall \ i, \quad \text{and } \liminf_{j \to \infty} \inf_u k_i(j, u) > \alpha_i, 0 \leq i \leq m$$

Case 2: Stable case

G is compact and H closed, hence compact.

Since the upper inequality in (4.3) is preserved anyway under convergence in $\mathcal{P}(S \times U)$, the closedness of H is the requirement that the lower inequality also do so. This is the case, e.g., if β_i's are zero or if k_i's are bounded, $1 \leq i \leq m$.

Lemma 4.8. In either case, the minimization problem above has a solution in H_e.

Proof Existence of a minimizer in H follows by its compactness in Case 2 and as in Lemma 2.8 in Case 1: One repeats those arguments for each k_j to conclude that if $\{\nu_n\} \subset H$ satisfy $\int k_0 d\nu_n \downarrow \alpha_0$, then $\{\nu_n\}$ is tight and every limit point ν thereof satisfies.

$$\int k_0 d\nu = \alpha_0, \quad \int k_j d\nu \leq \alpha_j, \quad 1 \leq j \leq m.$$

The existence of a minimizer in H_e then follows from Choquet's theorem as in Lemma 2.7. □

Suppose we revert to our original 'almost sure' formulation of the ergodic control problem and consider the constrained version thereof. That is, we seek to minimize almost surely, over all admissible $\{Z_n\}$, the quantity

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} k_0(X_m, Z_m),$$

subject to: for $1 \leq i \leq m$,

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} k_i(X_m, Z_m) \leq \alpha_i \quad \text{a.s.},$$

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} k_i(X_m, Z_m) \geq \beta_i \quad \text{a.s.}.$$

As in section 2, this can be reduced to the problem (4.2)-(4.3) without any loss of generality by using our characterization of limit points of empirical measures in Lemma 2.10, with the proviso that we add condition (†) of section 2 to 'Case 2' above. We refer to this modification as 'Case 2′'.

Theorem 4.1 In both Case 1 and Case 2′, there exists an optimal $f \in \Pi_{SR}$ that requires at most m randomizations. Furthermore, if H has a nonempty interior in G, then these exist $\lambda_1, \cdots, \lambda_{2m} \geq 0$ such that for all $\nu \in G$ and $\eta_1, \cdots, \eta_{2m} \geq 0$, we have : If $\nu_0 \in H_e$ is the optimal point, then

$$\begin{aligned}
&\int k_0 d\nu + \sum_{i=1}^{m} \lambda_i(\alpha_i - \int k_i d\nu) + \sum_{i=1}^{m} \lambda_{m+i}\left(\int k_i d\nu - \beta_i\right) \\
&\geq \int k_0 d\nu_0 + \sum_{i=1}^{m} \lambda_i(\alpha_i - \int k_i d\nu_0) + \sum_{i=1}^{m} \lambda_{m+i}\left(\int k_i d\nu_0 - \beta_i\right) \\
&\geq \int k_0 d\nu_0 + \sum_{i=1}^{m} \eta_i(\alpha_i - \int k_i d\nu_0) + \sum_{i=1}^{m} \eta_{m+i}\left(\int k_i d\nu_0 - \beta_i\right).
\end{aligned}$$

Proof The existence of an optimal $f \in H_e$ is argued above. That it requires at most m randomizations follows from Lemma 4.7 and the observation that only one of the inequalities in (4.3) can be active at a time for each i (except in the degenerate case $\alpha_i = \beta_i$, which is handled analogously). The last claim follows from standard Lagrange multiplier theory ([22], pp.216-219). □

A similar treatment is possible for constrained problems with other (e.g., discounted) cost criteria. An important difference there is the role played by the initial distribution, usually held fixed, which cannot be wished away. Any claims of optimality will be relative to a specific initial distribution (or more generally, a convex compact set thereof).

4.3 A general multiobjective problem

The foregoing had one 'primary' cost function k_0, whose average was to be minimized subject to constraints on several 'secondary' costs $k_1, \cdots, k_m$. In traditional 'multiobjective' framework, one wants to treat them on a comparable footing. Since all of them cannot in general be minimized simultaneously, one seeks a 'Pareto point', defined next. We stick to the ergodic control framework of the preceding section.

Let $\hat{k}(\nu) = [\int k_0 d\nu, \cdots, \int k_m d\nu] \in R^{m+1}$ denote the cost vector corresponding to $\nu \in G$ and define $W = \{\hat{k}(\nu) : \nu \in G\}$, which will be a closed convex subset of R^{m+1}. On W, define a partial order '<' by : $x = [x_0, \cdots, x_m] < y = [y_0, \cdots, y_m]$ if $x_i \leq y_i$ for all i, with a strict inequality for at least one i. Call $x^* \in W$ a Pareto point if there is no $z \in W$ for which $z < x^*$, It is easy to see that a minimizer of $\sum_{i=0}^{m} \lambda_i x_i$ over $x = [x_0, \cdots, x_n] \in W$ for any choice of $\lambda_i > 0$, $0 \leq i \leq m$, will be a Pareto point. In fact, if W is a polytope (i.e., convex hull of finitely many extreme points), then all Pareto points can be obtained thus. By our characterizaton of G_e, it follows that if S and U are finite, then G and hence W will have finitely many extreme points and the foregoing remark applies. More generally, a weaker claim holds, viz., the Pareto points obtained thus are dense in the set of all Pareto points. All this is a consequence of the celebrated Arrow-Barankin-Blackwell theorem [4].

More generally, Pareto points can be obtained by minimizing a '*utility function*' $U : W \to R$, which is any continuous function with the property : $U(y) < U(x)$ wherever $y < x$. The function $x \to \sum_i \lambda_i x_i$ with $\alpha_i > 0 \ \forall i$ is just a special case of this. Another important case is as follows : Let $k^* = [\min_G \int k_0 d\nu, \cdots, \min_G \int k_m d\nu] \in R^{m+1}$, the so called 'ideal point'. The nontrivial case is $k^* \notin W$. Let $\tilde{k} \in W$ be the unique point in W such that $||k^* - \tilde{k}|| = \min_{k \in W} ||k^* - k||$, corresponding to $U(x) = ||x - k^*||$. It is easy to see that this will be a Pareto point. Finding $\tilde{k}$ is an abstract quadratic programming problem.

There is no reason why $\tilde{k}$ should correspond to any element of G_e. The following 'approximation theorem', however, justifies a search for good strategies among those that correspond to a randomization between finitely many points of G_e. Assume that $k_0, \cdots, k_m$ are bounded.

Theroem 4.2 For any $\nu \in G, n \geq 1$, there exists a $\nu^n \in G$ such that ν^n is a convex combination of at most n points of G_e and

$$||\hat{k}(\nu) - \hat{k}(\nu^n)|| = O\left(\frac{1}{n}\right).$$

Proof Let ν be the barycenter of $\Phi \in \mathcal{P}(G_e)$ and $\{\xi(i)\}$ i.i.d. G_e-valued random variables with law Φ. Let $Y(i) = [\int k_0 d\xi(i), \ldots, \int k_m d\xi(i)], i \geq 1$. Then $\{Y(i)\}$ are i.i.d. with mean $\hat{k}(\nu)$ and

$$E\left[\left|\left|\frac{1}{n}\sum_{i=1}^{n} Y(i) - \hat{k}(\nu)\right|\right|^2\right] = \frac{1}{n^2} \cdot n \cdot E[||Y(1) - \hat{k}(\nu)||^2] = \frac{C}{n}$$

for a constant $C > 0$. Thus for at least one sample point,

$$\left|\left|\frac{1}{n}\sum_{i=1}^{n} Y(i) - \hat{k}(\nu)\right|\right|^2 < \frac{C}{n}.$$

The claim follows. □

Note, however, that this argument is not a constructive argument.

5 Concluding Remarks

The foregoing gives a reasonably self-contained account of the key aspects of the convex analytic approach to Markov decision processes. Nevertheless, it is by no means exhaustive. We briefly comment upon few important aspects thereof that were not addressed above.

(i) Dual problems

The 'dual' linear program to the infinite dimensional linear program over occupation measures establishes a link with the traditional dynamic programming approach. To see how this comes by, we first recall the relevant results from the theory of infinite dimensional linear programming [3].

Two topological vector spaces X, Y are said to form a dual pair if there exists a bilinear form $\langle \cdot, \cdot \rangle : X \times Y \to R$ such that the functions $x \to \langle x, y \rangle$ for $y \in Y$ separate points of X and the functions $y \to \langle x, y \rangle$ for $x \in X$ separate points of Y. Endow X with the coarsest topology required to render the former family of maps continuous, and Y with the dual topology. Let P be the positive cone in X and $P^* \subset Y$ the dual cone: $P^* = \{y \in Y : \langle x, y \rangle \geq 0, x \in P\}$.

Let Z, W be another dual pair topologized in a similar manner and $F : X \to Z$ a continuous linear map. Define $F^* : W \to X^*$ by $\langle Fx, w \rangle = \langle x, F^* w \rangle, x \in X, w \in W$. The primal LP problem is:

$$\text{Minimize } \langle x, c \rangle \text{ s.t.} \quad Fx = b, \quad x \in P,$$

with $b \in Z, c \in Y$ prescribed. Let α denote the infimum of $\langle x, c \rangle$ subject to these constraints. The dual problem is

$$\text{Max } \langle b, w \rangle \text{ s.t.} \quad -F^* x + c \in P^*, \quad w \in W.$$

Let α' denote the supremum of $\langle b, w \rangle$ subject to these constraints. It is known that $\alpha \geq \alpha'$. Let $C = \{x \in P : Fx = b\}$, $D = \{(Fx, \langle x, c \rangle) : x \in P\}$. The following results gives conditions for the absence of a 'duality gap'.

Theorem 5.1 ([3], p.53) If $C \neq \phi, D$ is closed and $x \to \langle x, c \rangle$ attains its minimum on C, then $\alpha = \alpha'$.

Applying this to the linear programming formulation of ergodic control in section II, the dual problem and Theorem 5.1 lead to the following characterization of the optimal cost α:

Corollary 5.1 $\alpha = \sup \Big\{ a : \min_u \Big(k(i, u) + f(i) - \sum_j P(j/i, u) f(j) \Big) \geq a, f : S \to R \Big\}$.

Likewise, dualising the primal problem for discounted cost problem leads, in view of Theorem 5.1, to the following characterization of the 'value function' of dynamic programming, defined as

$$V(i) = \inf_{\text{admissible controls}} E\left[\sum_{m=0}^{\infty} \beta^m k(X_m, Z_m)/X_0 = i\right], i \in S:$$

Corollary 5.2 $V(i) = \sup\{f(i) : \inf_u(k(i,u) - f(i) \ + \beta \sum_j P(j/i,u)f(j)) \geq 0, \ f : S \to R\}$.

Similar results are possible for the finite horizon and exit time problems, as well as for general state spaces. See [20] for an extensive account of the latter.

(ii) Computational aspects

An important motivation for linear/convex programming approaches has been the prospect of using to advantage the computational machinery of these fields. Linear programming has always been a 'third' approach to computation in Markov decision processes, after value and policy iteration. Despite the proven convergence results and ease of formulation for the latter, linear programming remains an attractive alternative because it facilitates the use of commercial LP packages which are primed to exploit any extra structure the problem may have to offer. For constrained problems, dynamic programming approach proves quite awkward, while in the linear programming approach the constraints merely translate into additional constraints for the linear program. See [24] for an extensive account of the traditional linear programming algorithms for Markov decision processes and [1] for the corresponding development for constrained problems.

(iii) Mixed problems

There has been some interest in Markov decision processes with mixed criteria, such as criteria that combine ergodic and discounted costs. (See, e.g., [18].)

(iv) Game theoretic formulations

An interesting formulation of multiobjective problem under uncertainty treats it as a game against an antagonist, taking the 'worst case' approach, and seeks to drive the vector ergodic cost to a prescribed set of 'acceptable values' [27]. Though similar in spirit to section 4 above, this approach differs significantly in the kind of techniques used.

We conclude with a bibliographical note. Giving an extensive, even representative bibliography is a major task I do not wish to undertake. The following therefore is confined only to a few key antecedents and the immediate resources for the material presented here.

Further references can be found in [1], [9], [20], [24].

The convex/linear programming approach, as already mentioned, goes back to Manne [23]. The approach to ergodic control taken in section 2 first appeared in [7] for the irreducible case. Related results also appear in [2]. The fully general case presented here appears in [11].

The corresponding results for other cost criteria are from [6], [9]. For the general state space case, the treatment that appears here is new in principle, but closely mimicks the corresponding theory for controlled diffusions developed in [10], [13].

The constrained control problem dates back to [15], [16]. The convex analytic approch presented here is from [11], which improves upon earlier results from [8], [9], [25]. The 'sample path' or 'a.s.' variant was studied in [26], though our treatment is different. See [18] for results in the discounted cost framework. The multiobjective problem studied in subsection 4.2 is from [19], except for Theorem 4.2 which is new. Reference [19] also considers computational issues. Interesting classes of Pareto points have been identified in the discounted case in [18].

References

[1] ALTMAN, E., Constrained Markov decision processes, Reseach Report No. 2574, INRIA Sophia-Antipolis, 1994.

[2] ALTMAN, E.; SHWARTZ, A., Markov decision problems and state-action frequencies, SIAM J. Control and Optim. 29 (1991), pp.786-809.

[3] ANDERSON, E.J.; NASH, P., Linear Programming in Infinite Dimensional Spaces, John Wiley, Chichester, 1987.

[4] ARROW, K.J.; BARANKIN, E.W.; BLACKWELL, D., Admissible points of convex sets, in 'Contributions to the Theory of Games', H.W. Kuhn and A.W. Tucker (eds.), Princeton Uni. Press, Princeton, NJ, 1950, pp.87-91.

[5] BHATT, A.G.; BORKAR, V.S., Occupation measures for controlled Markov processes: characterization and optimality, The Annals of Prob. 24 (1996), pp.1531-1562.

[6] BORKAR, V.S., A convex analytic approach to Markov decision processes, Prob. Theory and Related Fields 78 (1988), pp.583-602.

[7] BORKAR, V.S., Control of Markov chains with long-run average cost criterion : the dynamic programming equations, SIAM J. Control and Optim. 27 (1989), pp.642-657.

[8] BORKAR, V.S., Controlled Markov chains with constraints, *Sadhana* : Indian Academy of Sciences Proc. in Engg. Sciences 15 (1990), pp.405-413.

[9] BORKAR, V.S., Topics in Controlled Markov Chains, Pitman Research Notes in Maths. No.240, Longman Scientific and Technical, Harlow, England, 1991.

[10] BORKAR, V.S., Controlled diffusions with constraints II, J. Math. Analysis and Appl. 176 (1993), pp.310-321.

[11] BORKAR, V.S., Ergodic control of Markov chains with constraints - the general case, SIAM J. Control and Optim. 32 (1994), pp. 176-186.

[12] BORKAR, V.S., Probability Theory : An Advanced Course, Springer Verlag, New York, 1995.

[13] BORKAR, V.S.; GHOSH, M.K.; Controlled diffusions with constraints, J. Math. Analysis and Appl. 152 (1990) pp.88-108.

[14] CHOQUET, G., Lectures on Analysis, Vol.II : Representation Theory, W.A. Benjamin, Inc., Reading, Mass., 1969.

[15] DERMAN, C., Finite State Markovian Decision Processes, Academic Press, New York, 1970.

[16] DERMAN, C.; KLEIN, M.; Some remarks on finite horizon Markovian decision models, Op. Research 13 (1965), pp. 272-278.

[17] DUBINS, L.; On extreme points of convex sets, J. Math. Analysis and Appl. 5 (1962), pp. 237-244.

[18] FEINBERG, E.; SHWARTZ, A.; Constrained discounted dynamic programming, Maths. of Op. Research 21 (1996), pp. 922-945.

[19] GHOSH, M.K.; Markov decision processes with multiple costs, Op. Research Letters 9 (1990), pp. 257-260.

[20] HERNANDEZ-LERMA, O.; LASSERRE, J.B.; Discrete-Time Markov Control Processes, Springer Verlag, New York, 1996.

[21] KURTZ, T.G., STOCKBRIDGE, R.; Existence of Markov controls and characterization of optimal Markov controls, SIAM J. Control and Optim. 36 (1998), pp. 609-653.

[22] LUENBERGER, D.G.; Optimization by Vector Space Methods, John Wiley, New York, 1969.

[23] MANNE, A., Linear programming and sequential decisions, Management Sci. 6 (1960), pp. 259-267.

[24] PUTERMAN, M.; Markov Decision Processes, John Wiley, New York, 1994.

[25] ROSS, K.W.; Randomized and past-dependent policies for Markov decision problems with multiple constraints, Op. Research 37 (1989), pp. 474-477.

[26] ROSS, K.W.; VARADARAJAN, R.; Markov decision processes with sample path constraints: the communicating case, Op. Research 37 (1989), pp. 780-790.

[27] SHIMKIN, N.; SHWARTZ, A.; Guarranteed performance regions in Markovian systems with competing decision makers, IEEE Trans. on Automatic Control 38 (1993), pp. 84-95.