Extreme occupation measures in Markov decision processes with an absorbing state

Alexey Piunovskiy[*] Yi Zhang [†]

December 1, 2023

Abstract. In this paper, we consider a Markov decision process (MDP) with a Borel state space $\mathbf{X} \cup \{\Delta\}$, where Δ is an absorbing state (cemetery), and a Borel action space $\mathbf{A}$. We consider the space of finite occupation measures restricted on $\mathbf{X} \times \mathbf{A}$, and the extreme points in it. It is possible that some strategies have infinite occupation measures. Nevertheless, we prove that every finite extreme occupation measure is generated by a deterministic stationary strategy. Then, for this MDP, we consider a constrained problem with total undiscounted criteria and J constraints, where the cost functions are nonnegative. By assumption, the strategies inducing infinite occupation measures are not optimal. Then, our second main result is that, under mild conditions, the solution to this constrained MDP is given by a mixture of no more than $J + 1$ occupation measures generated by deterministic stationary strategies.

Keywords. Markov decision process. Total cost. Occupation measure. Mathematical programming. Extreme point. Mixture

AMS 2000 subject classification: Primary 90C40; Secondary 90C25

1 Introduction

Perhaps the first paper, where the discounted Markov decision process (MDP) was reformulated as a linear program, is [7]. The modern so called 'convex analytic approach' originates from the works by V.Borkar [4, 5]. It is applied to the models with total cost (discounted or not) as well as with the long-run average cost: let us only mention the book treatments [1, 28, 31] and the survey [6]. This approach proved to be especially fruitful in dealing with problems with constraints, see the survey [32] and the authoritative monograph [29] on finite MDPs, i.e., MDPs with finite state and action spaces. For convex analytic approach to continuous-time MDPs, see e.g., [24, 25, 33] and the monograph [34].

The convex analytic approach is based on the reformulation of the constrained MDP problem as a convex optimization problem in the space of occupation measures with affine objective functions and inequality constraints, where the occupation measures are defined in accordance with the performance criteria of the MDP problem. The space of occupation measures is a convex space (i.e., a convex subset of a cone, not necessarily of a vector space). Thus, here, the relevant notions, such as convex optimization problem, affine

[*] Corresponding author. Department of Mathematical Sciences, University of Liverpool, Liverpool, U.K.. E-mail: piunov@liv.ac.uk.

[†] School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.. Email: y.zhang.29@bham.ac.uk

functions and extreme points, are understood with respect to (wrt) the underlying convex space, see [37]. An important target to show is the existence of an optimal strategy for the MDP problem, whose occupation measure is the convex combination of finitely many extreme points in the space of occupation measures, which we call extreme occupation measures. If the number of constraints in the MDP problem is J, the mixture is over at most $J+1$ extreme occupation measures. Such a strategy is called a $(J+1)$-mixed optimal strategy. Then a key ingredient in the convex analytic approach to MDPs is the characterization of such extreme occupation measures. This task is easier when the state is discrete (finite or countable), as considered in [1, 4, 5, 29], but our consideration in this paper is a Borel MDP model, by which we mean an MDP with Borel state and action spaces.

Let us concentrate on the literature for Borel MDP models. For discounted MDPs, the most relevant recent works include [12, 16, 21], where by using the convex analytic approach, optimal stationary strategies were proved. While mixed strategies were not considered in [12, 21], in establishing the existence of a so called optimal chattering strategy in [16] (see also [22, 23]), the existence of an optimal $(J+1)$-mixed strategy was observed, see the proof of [16, Theorem 2]. For discounted MDPs but under more restrictive conditions, this result appeared in [31, 42]. In an absorbing MDP, there is a costless absorbing state, called 'cemetery' for brevity, and, given the initial state, under each strategy, the expected time until the state process reaches the cemetery is finite. In fact this is equivalent to the expected absorbing time say T being bounded in the set of all strategies, see [18, p.132]. The expected absorbing time can be written as the series of the tail probabilities of $T > m$ over $m \geq 0$. If this series converges uniformly over all strategies, then the MDP is called uniformly absorbing. This definition appeared in [17]. It was observed in [17] that discounted MDPs are special cases of uniformly absorbing MDPs, by viewing the discount factor as the parameter of a geometrically distributed external killing time. For uniformly absorbing MDPs, the existence of a $(J+1)$-mixed optimal strategy was obtained by Feinberg and Rothblum, see [18, Theorem 9.2], as well as that each extreme occupation measure is generated by a deterministic stationary strategy, see [18, Lemma 4.6]; see also [38]. The convex analytic approach was also developed for optimal stopping problems in discrete time, see [10] and the references therein.

In the present paper, we consider an MDP with a Borel state space $\mathbf{X} \cup \{\Delta\}$ and a Borel action space $\mathbf{A}$. The point Δ is a single absorbing state. We call such a model an MDP with an absorbing state or with a cemetery, though it is also known under other names such as the stochastic shortest path problem, see [2], where unconstrained MDP problems were considered and the main interest was the characterization of the optimal value function out of the class of so called proper strategies in terms of the solution to the optimality equation. It is without loss of generality that we consider a fixed initial state rather than a fixed initial distribution. We also assume that the cemetery Δ is costless. For this reason, we consider occupation measures as the total expected state-action frequencies restricted on $\mathbf{X} \times \mathbf{A}$. If a strategy has a finite occupation measure, we call it an absorbing strategy with the given initial state. Proper strategies as considered in [2] can be viewed as special types of absorbing strategies. If the occupation measure of each strategy is finite, our model becomes the absorbing model. Nevertheless, similarly to [9], here we do allow that some strategies have infinite occupation measures. This is the main novelty compared with the aforementioned works [12, 16, 18, 21], see more comments on this below.

Our contributions are as follows. First, we show that every extreme point of the space of finite occupation measures is generated by a deterministic stationary strategy. Then, we consider a constrained problem with total undiscounted criteria and J constraints, where the cost functions are nonnegative. We formulate the problem as a convex program in the space of occupation measures (see (5)). Under mild conditions, we show that there exists an optimal strategy whose occupation measure is in the form of a mixture of no more than $J + 1$ occupation measures of deterministic stationary strategies.

For the latter result, we make the assumption, which in particular, implies that strategies inducing infinite occupation measures are not optimal. Under this assumption, for the MDP problem, instead of dealing with the whole space of occupation measures, it is sufficient to work with the space of finite occupation measures. However, restricting an MDP to absorbing strategies is not the same as considering an absorbing MDP. Firstly, in an absorbing MDP, the total values of all occupation measures are bounded above, whereas if the MDP is not absorbing, then the total values of all finite occupation measures can be unbounded. This can be seen by considering an optimal stopping problem as in [10]: for the set of strategies, stopping at step $n = 1, 2, \dots$, the values of their (finite) occupation measures are unbounded. In this connection, we mention that, for discounted MDPs, see e.g., [31], it is convenient to endow the space of occupation measures with the weak topology generated by bounded continuous functions. The same was done in [18] for absorbing MDPs. To deal with infinite occupation measures, we endow that space with the final topology generated by the projection mapping from the space of strategic measures to occupation measures. These features require new proofs of the key theorems on the characterization of the extreme finite occupation measures (see Theorem 1) and on the sufficiency of mixtures of (occupation measures of) deterministic strategies (see Theorem 2).

In terms of other relevant works, we mention the following. First, constrained total undiscounted Borel MDPs with non-negative cost functions were also studied in [9]. Although it was not assumed a priori in [9] that there is a costless cemetery in the state space, it was shown under some conditions that one can always construct a costless cemetery set, after modifying the admissible action spaces on that set. By merging this set as a costless cemetery, we may view the model in [9] in the framework of the present paper, and can apply to it our first result on the characterization of extreme finite occupation measures. Except for special cases, our second result concerning the optimal mixed strategies are not applicable to the model in [9] because no assumption was made in [9] that strategies with infinite occupation measures were suboptimal or infeasible. On the other hand, neither the extreme occupation measures nor the mixed strategies were considered in [9]. The paper here can be viewed as a complement to it. Second, we note that the results in this paper are also relevant to the studies in continuous-time MDPs, see [26, 35, 36], because the problems considered therein were eventually reduced to an MDP model, see more details in the book [34].

Allowing the cost functions to be negative leads to a more complicated theory. The convex analytic approach to such constrained MDPs was developed in [8, 11], but mixtures of occupation measures were not considered there.

The rest of this paper is organized as follows. The MDP model under study is described in Section 2. Several necessary auxiliary statements are given in Section 3, including the known results on the solvability of the formulated problem. Sections 4 and 5 present the main results: characterization of the extreme occupation measures and sufficiency of the finite mixtures of deterministic stationary strategies. The paper ends with a conclusion in Section 6. The proofs of the main statements are postponed to the appendix.

2 Description of the model

The primitives of an MDP are the following.

- The state space is $\mathbf{X}_\Delta = \mathbf{X} \cup \{\Delta\}$, where $\mathbf{X}$ is a nonempty topological Borel space, endowed with the σ-algebra $\mathcal{B}(\mathbf{X})$, and Δ is the isolated absorbing state (cemetery).
- The action space $\mathbf{A}$ is a nonempty topological Borel space, endowed with the σ-algebra $\mathcal{B}(\mathbf{A})$.
- The transition probability $p(dy|x,a)$ is a stochastic kernel from $\mathbf{X}_\Delta \times \mathbf{A}$ to $\mathcal{B}(\mathbf{X}_\Delta)$; $p(\{\Delta\}|\Delta, a) \equiv 1$.
- The $[-\infty, +\infty]$-valued one-step cost functions $r_j(\cdot,\cdot)$ on $\mathbf{X}_\Delta \times \mathbf{A}$, $j = 0, 1, \dots, J$, where $J \in \{0, 1, \dots\}$ is a fixed integer; $r_j(\Delta, a) \equiv 0$.

Usually, the initial state $x_0 \in \mathbf{X}$ is fixed, but sometimes we consider other arbitrarily fixed initial states $x \in \mathbf{X}$. (See, e.g., Lemma 3.)

Regarding terminology, we often refer to $\{\mathbf{X}_\Delta, \mathbf{A}, p, \{r_j\}_{j=0}^J\}$ as a MDP model or simply a MDP. We may also consider the 'cost-free' MDP model $\{\mathbf{X}_\Delta, \mathbf{A}, p\}$ because several definitions and properties presented below do not involve the properties of the cost functions.

Definition 1 (Strategy) *Consider the MDP* $\{\boldsymbol{X}_\Delta, \boldsymbol{A}, p\}$.

(a) A strategy $\sigma = \{\sigma_n\}_{n=1}^\infty$ *is a sequence of stochastic kernels such that for each* $n = 1, 2, \dots,$ $\sigma_n(da|x_0, a_1, \dots, x_{n-1})$ *is a stochastic kernel from* $(\boldsymbol{X}_\Delta \times \boldsymbol{A})^{n-1} \times \boldsymbol{X}_\Delta$ *to* $\mathcal{B}(\boldsymbol{A})$, *where* $(\boldsymbol{X}_\Delta \times \boldsymbol{A})^0 \times \boldsymbol{X}_\Delta := \boldsymbol{X}_\Delta$.

(b) A strategy $\sigma = \{\sigma_n\}_{n=1}^\infty$ *is Markov if for each* $n = 1, 2, \dots,$ *there is a stochastic kernel* $\sigma_n^M(da|x_{n-1})$ *from* $\boldsymbol{X}_\Delta$ *to* $\mathcal{B}(\boldsymbol{A})$ *such that*

$$\sigma_n^M(da|x_{n-1}) = \sigma_n(da|x_0, a_1, \dots, x_{n-1})$$

for each $(x_0, a_1, \dots, x_{n-1}) \in (\boldsymbol{X}_\Delta \times \boldsymbol{A})^{n-1} \times \boldsymbol{X}_\Delta$.

(c) A strategy $\sigma = \{\sigma_n\}_{n=1}^\infty$ *is called stationary if there is a stochastic kernel* $\sigma^s(da|x)$ *from* $\boldsymbol{X}_\Delta$ *to* $\mathcal{B}(\boldsymbol{A})$ *such that*

$$\sigma^s(da|x_{n-1}) = \sigma_n(da|x_0, a_1, \dots, x_{n-1})$$

for each $n = 1, 2, \dots,$ *and* $(x_0, a_1, \dots, x_{n-1}) \in (\boldsymbol{X}_\Delta \times \boldsymbol{A})^{n-1} \times \boldsymbol{X}_\Delta$. *Below, a stationary strategy is usually identified with* σ^s.

(d) If $\sigma^s(da|x)$ *is concentrated on* $\{\varphi(x_{n-1})\}$, *where* φ *is an* $\mathbf{A}$*-valued measurable mapping, then the stationary strategy is called deterministic stationary. With conventional abuse of notations, we often signify a deterministic stationary strategy by* φ.

(e) We always assume that $\sigma_n(\{\hat{a}\}|x_0, a_1, \dots, \Delta) = 1$ *whenever* $x_{n-1} = \Delta$. *Here* $\hat{a} \in \mathbf{A}$ *is an arbitrarily fixed action.*

As is well known, for each control strategy σ and initial state $x_0 \in \mathbf{X}$, there is a unique strategic measure on the sample space $\Omega := (\mathbf{X}_\Delta \times \mathbf{A})^\infty$, denoted as $\mathsf{P}^\sigma_{x_0}$, which is specified by the following conditions:

$$\mathsf{P}^\sigma_{x_0}(X_0 \in dy) = \delta_{x_0}(dy); \tag{1}$$

and for each $n = 1, 2, \dots$, $\Gamma^{\mathbf{X}}_i \in \mathcal{B}(\mathbf{X}_\Delta)$ $(i = 0, 1, \dots, n)$ and $\Gamma^{\mathbf{A}}_i \in \mathcal{B}(\mathbf{A})$ $(i = 1, 2, \dots, n)$,

$$\begin{aligned} &\mathsf{P}^\sigma_{x_0}(X_0 \in \Gamma^{\mathbf{X}}_0,\ A_1 \in \Gamma^{\mathbf{A}}_1,\ \dots,\ X_{n-1} \in \Gamma^{\mathbf{X}}_{n-1},\ A_n \in \Gamma^{\mathbf{A}}_n) \\ = &\int\limits_{\Gamma^{\mathbf{X}}_0 \times \Gamma^{\mathbf{A}}_1 \times \cdots \times \Gamma^{\mathbf{X}}_{n-1}} \sigma_n(\Gamma^{\mathbf{A}}_n | x_0, a_1, \dots, x_{n-1}) \mathsf{P}^\sigma_{x_0}(X_0 \in dx_0, A_1 \in da_1, \dots, X_{n-1} \in dx_{n-1}); \end{aligned} \tag{2}$$

and

$$\begin{aligned} &\mathsf{P}^\sigma_{x_0}(X_0 \in \Gamma^{\mathbf{X}}_0,\ A_1 \in \Gamma^{\mathbf{A}}_1, \dots,\ X_n \in \Gamma^{\mathbf{X}}_n) \\ = &\int\limits_{\Gamma^{\mathbf{X}}_0 \times \Gamma^{\mathbf{A}}_1 \times \cdots \times \Gamma^{\mathbf{X}}_{n-1} \times \Gamma^{\mathbf{A}}_n} p(\Gamma^{\mathbf{X}}_n | x_{n-1}, a_n) \\ &\times \mathsf{P}^\sigma_{x_0}(X_0 \in dx_0, A_1 \in da_1, \dots, X_{n-1} \in dx_{n-1}, A_n \in da_n). \end{aligned} \tag{3}$$

For details, see [14, 27, 31]. Denote by Σ the set of all strategies, and by $\mathcal{P} := \{\mathsf{P}^\sigma_{x_0} : \sigma \in \Sigma\}$ the set of all strategic measures (with the initial state $x_0 \in \mathbf{X}$). The expectation taken with respect to $\mathsf{P}^\sigma_{x_0}$ is denoted by $\mathsf{E}^\sigma_{x_0}$. We equip the space of probability measures on $\mathcal{B}(\Omega)$, denoted as $\mathcal{P}(\Omega)$, with the weak topology generated by bounded continuous functions on Ω, and fix its trace τ on the space $\mathcal{P}$ of all strategic measures. Then $\mathcal{P}(\Omega)$ is a Borel space, see [3, Corollary 7.25.1], and we endow $\mathcal{P}(\Omega)$ with its Borel σ-algebra.

The constrained optimal control problem for the MDP model $\{\mathbf{X}_\Delta, \mathbf{A}, p, \{r_j\}_{j=0}^J\}$ is

$$\begin{aligned} &\text{Minimize over all strategies } \sigma\text{:} && \mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_0(X_n, A_{n+1})\right] \\ &\text{subject to} && \mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_j(X_n, A_{n+1})\right] \le d_j, \quad j = 1, 2, \dots, J, \end{aligned} \tag{4}$$

where, for $j \in \{0, 1, \dots, J\}$,

$$\mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_j(X_n, A_{n+1})\right] := \mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_j^+(X_n, A_{n+1})\right] - \mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_j^-(X_n, A_{n+1})\right]$$

with $r_j^+(\cdot,\cdot)$ and $r_j^-(\cdot,\cdot)$ being the positive part and the negative part of the function $r_j(\cdot,\cdot)$ so that $r_j(\cdot,\cdot) = r_j^+(\cdot,\cdot) - r_j^-(\cdot,\cdot)$. We accept that $\infty - \infty := \infty$ concerning the definition of $\mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_j(X_n, A_{n+1})\right]$.

If $J = 0$, then the problem is called unconstrained.

Definition 2 (Feasible and optimal strategies) *A strategy is called feasible if all the constraints in (4) are satisfied; it is called feasible with a finite value if, additionally,* $\mathsf{E}^\sigma_{x_0}\left[\sum_{n=0}^\infty r_0(X_n, A_{n+1})\right] \in \mathbb{R} := (-\infty, \infty)$*; it is called optimal if it solves problem (4).*

Definition 3 (Semicontinuous MDP) *An MDP* $\{\boldsymbol{X}_\Delta, \boldsymbol{A}, p, \{r_j\}_{j=0}^J\}$ *is called semicontinuous if* *(a) The action space $\boldsymbol{A}$ is compact.*

(b) For each bounded continuous function $f(\cdot)$ on $\boldsymbol{X}$, $\int_{\boldsymbol{X}} f(y)p(dy|x,a)$ is continuous in $(x,a) \in \boldsymbol{X} \times \boldsymbol{A}$.

(c) For each $j = 0, 1, \ldots, J$, the function $r_j(\cdot,\cdot)$ is lower semicontinuous on $\boldsymbol{X} \times \boldsymbol{A}$.

3 Preliminaries

In this section, we collect some preliminary results, which will be needed in proving the main results of this paper. Several of them are known, or follow from well known results. They will be called propositions. We thus skip the proofs of most of them but always refer to relevant literature.

Proposition 1 *(a) The set $\mathcal{P}$ of all strategic measures, for a fixed initial state $x_0 \in \mathbf{X}$, is a measurable and convex subset of $\mathcal{P}(\Omega)$. (Recall the notations introduced below (3).)*

(b) Suppose conditions (a) and (b) in Definition 3 are satisfied. Then the space $\mathcal{P}$, endowed with the weak topology, is compact.

Proof. For the first statement, see Theorem 8 of [31] and Chapter 5,§5 of [14]. For the second statement, see e.g., [39]. □

Unless stated otherwise, we always endow the space of strategic measures with the weak topology.

The next result is known as the Derman-Strauch Lemma. It asserts that the marginal distributions of each strategy can be replicated by a Markov strategy.

Proposition 2 *For each strategy σ, there is a Markov strategy $\sigma^M = \{\sigma_n^M\}_{n=1}^\infty$ such that*

$$\mathsf{P}_{x_0}^{\sigma}(X_{n-1} \in dx, A_n \in da) = \mathsf{P}_{x_0}^{\sigma^M}(X_{n-1} \in dx, A_n \in da)$$

for each $n = 1, 2, \ldots$. Here σ_n^M is the stochastic kernel from $\boldsymbol{X}$ to $\boldsymbol{A}$ such that

$$\mathsf{P}_{x_0}^{\sigma}(X_{n-1} \in dx, A_n \in da) = \mathsf{P}_{x_0}^{\sigma}(X_{n-1} \in dx)\sigma_n^M(da|x).$$

One can take an arbitrarily fixed version of the stochastic kernel σ_n^M.

Proof. See Lemma 2 of [31]. □

Now it is clear that one can restrict to Markov strategies when investigating problem (4).

Next we introduce occupation measures of strategies.

Definition 4 (Occupation measures) *The occupation measure $\mathsf{M}_{x_0}^{\sigma}$ of a strategy σ in the MDP $\{\boldsymbol{X}_\Delta, \boldsymbol{A}, p\}$ with the initial state $x_0 \in \boldsymbol{X}$ is defined by*

$$\begin{aligned}\mathsf{M}_{x_0}^{\sigma}(\Gamma_X \times \Gamma_A) &:= \mathsf{E}_{x_0}^{\sigma}\left[\sum_{n=1}^{\infty} \mathbb{I}\{X_{n-1} \in \Gamma_X,\ A_n \in \Gamma_A\}\right] \\ &= \sum_{n=1}^{\infty} \mathsf{E}_{x_0}^{\sigma}\left[\mathbb{I}\{X_{n-1} \in \Gamma_X,\ A_n \in \Gamma_A\}\right]\end{aligned}$$

for each $\Gamma_X \in \mathcal{B}(\boldsymbol{X})$ and $\Gamma_A \in \mathcal{B}(\boldsymbol{A})$. The set of all occupation measures is denoted as $\mathcal{D}$; $\mathcal{D}^f := \{\mathsf{M}_{x_0}^{\sigma} :\ \mathsf{M}_{x_0}^{\sigma}(\mathbf{X} \times \mathbf{A}) < \infty\}$ is the set of all finite occupation measures on $\mathcal{B}(\mathbf{X} \times \mathbf{A})$.

Now for all $j = 0, 1, \dots, J$,

$$\mathsf{E}^{\sigma}_{x_0}\left[\sum_{n=0}^{\infty} r_j(X_n, A_{n+1})\right] = \int_{\mathbf{X}\times\mathbf{A}} r_j(x,a)\mathsf{M}^{\sigma}_{x_0}(dx \times da).$$

Accordingly, one can reformulate problem (4) as follows:

$$\begin{aligned} \text{Minimize over } \mathcal{D}: \ R_0(\mathsf{M}) &:= \int_{\mathbf{X}\times\mathbf{A}} r_0(x,a)\mathsf{M}(dx \times da) \qquad (5)\\ \text{subject to } \ R_j(\mathsf{M}) &:= \int_{\mathbf{X}\times\mathbf{A}} r_j(x,a)\mathsf{M}(dx \times da) \le d_j \quad j = 1, 2, \dots, J. \end{aligned}$$

Proposition 3 *The set of all occupation measures $\mathcal{D}$ with the initial state x_0 is a convex set in the cone of $[0,\infty]$-valued measures on $\mathcal{B}(\mathbf{X}\times\mathbf{A})$. The set $\mathcal{D}^f$ of finite occupation measures is a convex subset of the linear space of finite signed measures on $\mathcal{B}(\mathbf{X}\times\mathbf{A})$. It is a (convex) face of $\mathcal{D}$.*

Proof. It follows from Proposition 1 that the convex combination of two measures in $\mathcal{D}$ is still in $\mathcal{D}$. This justifies the first assertion. The second assertion follows from the first assertion and the observation that if $\mathsf{M}_1, \mathsf{M}_2$ are in $\mathcal{D}^f$, then so is their convex combination. For the last assertion, note that if, for some $\alpha \in (0,1)$ and $\mathsf{M}_1, \mathsf{M}_2 \in \mathcal{D}$, $\mathsf{M} = \alpha\mathsf{M}_1 + (1-\alpha)\mathsf{M}_2$ is in $\mathcal{D}^f \subseteq \mathcal{D}$, then it is necessary that $\mathsf{M}_1(\mathbf{X}\times\mathbf{A}) < \infty$ and $\mathsf{M}_2(\mathbf{X}\times\mathbf{A}) < \infty$, meaning that $\mathsf{M}_1, \mathsf{M}_2 \in \mathcal{D}^f$. Hence, $\mathcal{D}^f$ is a face in $\mathcal{D}$. □

The next two results provide some relations satisfied by occupation measures of a strategy (respectively, a stationary strategy).

Proposition 4 *The occupation measure $\mathsf{M}^{\sigma}_{x_0}$ of a strategy σ satisfies the following equation:*

$$\mu(\Gamma \times \boldsymbol{A}) = \delta_{x_0}(\Gamma) + \int_{\boldsymbol{X}\times\boldsymbol{A}} p(\Gamma|y,a)\mu(dy \times da), \ \forall\ \Gamma \in \mathcal{B}(\boldsymbol{X}). \qquad (6)$$

Proof. See Lemma 9.4.3 of [28]. □

Proposition 5 *Suppose σ^s is a stationary strategy. Then*

$$\mathsf{M}^{\sigma^s}_{x_0}(\Gamma_X \times \Gamma_A) = \int_{\Gamma_X} \sigma^s(\Gamma_A|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \boldsymbol{A}), \ \Gamma_X \in \mathcal{B}(\boldsymbol{X}), \ \Gamma_A \in \mathcal{B}(\boldsymbol{A}) \qquad (7)$$

and $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \boldsymbol{A})$ is the (setwise) minimal measure on $\mathcal{B}(\boldsymbol{X})$ satisfying the equation

$$\mu(\Gamma) = \delta_{x_0}(\Gamma) + \int_{\boldsymbol{X}}\int_{\boldsymbol{A}} p(\Gamma|y,a)\sigma^s(da|y)\mu(dy), \ \Gamma \in \mathcal{B}(\boldsymbol{X}). \qquad (8)$$

Proof. See [34, pp.563-564]. □

As was mentioned in Section 1, for discounted MDPs as well as absorbing MDPs, see e.g., [18, 31], the space of occupation measures was often endowed with the weak topology generated by bounded continuous functions. To deal with infinite occupation measures, it is more convenient to endow $\mathcal{D}$ with the final topology generated by the projection mapping from the space of strategic measures to occupation measures. See the next definition.

Definition 5 *ρ is the final topology on $\mathcal{D}$ associated with the mapping $O: \mathcal{P} \to \mathcal{D}$ defined by*

$$\mathsf{M}(dx \times da) = \sum_{n=1}^{\infty} \mathsf{P}(X_{n-1} \in dx, A_n \in da).$$

That is the finest topology for which the mapping O is continuous. A subset $\Gamma \subseteq \mathcal{D}$ is open (wrt ρ) if and only if $O^{-1}(\Gamma)$ is open in $\mathcal{P}$. Recall that $\mathcal{P}$ was endowed with the weak topology.

Lemma 1 *Consider the MDP model $\{\mathbf{X}_\Delta, \mathbf{A}, p\}$.*

(a) Under conditions (a) and (b) of Definition 3 the topological space $(\mathcal{D}, \rho)$ is compact.

(b) For each non-negative lower semicontinuous function $r(\cdot,\cdot): \mathbf{X} \times \mathbf{A} \to [0,\infty]$, the mapping $R(\cdot): \mathcal{D} \to [0,\infty]$ defined by

$$R(\mathsf{M}) := \int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{M}(dx \times da)$$

is lower semicontinuous.

The proofs of all the lemmas and theorems can be found in the appendix.

Corollary 1 *If the MDP $\{\mathbf{X}_\Delta, \mathbf{A}, p, \{r_j\}_{j=0}^J\}$ is semicontinuous and $r_j(\cdot,\cdot) \geq 0$, $j = 0, 1, \dots, J$, then the constrained problem (4) has an optimal solution, provided that there exists a feasible solution.*

Proof. Since the equivalent problems (4) and (5) have feasible solutions, the space $(\mathcal{D}, \rho)$ is compact, and the functions $R_j(\cdot)$ are lower semicontinuous, we see that the set

$$\{\mathsf{M} \in \mathcal{D}: \; R_j(\mathsf{M}) \leq d_j, \quad j = 1, 2, \dots, J\}$$

is nonempty and compact. Thus, the lower semicontinuous function $R_0(\cdot)$ attains its minimum thereon. □

Alternatively, the above corollary also follows from Proposition 1, see also [39], but its proof was given here in the hope of improving readability.

4 Extreme finite occupation measures

In this section we present our first main result concerning the characterization of extreme finite occupation measures. We emphasize that this result does not require any extra conditions on the MDP model; in particular, the MDP does not need to be semicontinuous.

Definition 6 (Induced strategy) *For $\mathsf{M} \in \mathcal{D}^f$, the stationary strategy σ^s, coming form the decomposition*

$$\mathsf{M}(dx \times da) = \sigma^s(da|x)\mathsf{M}(dx \times \mathbf{A})$$

on $\mathcal{B}(\mathbf{X} \times \mathbf{A})$, is called induced (by M). Here one can take an arbitrarily fixed version of the stochastic kernel σ^s, as the following lemma is valid.

The next result asserts that any finite occupation measure is generated by a stationary strategy.

Lemma 2 *Suppose* $\mathsf{M} \in \mathcal{D}^f$ *and* σ^s *is the stationary strategy induced by* M. *(One can take an arbitrary version of the stochastic kernel* σ^s.*) Then* $\mathsf{M} = \mathsf{M}_{x_0}^{\sigma^s}$.

Lemma 2 is known for countable-state MDPs [1, Theorem 8.1]. See also [20], which also provided examples showing that Lemma 2 does not hold for $\mathsf{M} \in \mathcal{D}$ in general.

The next result plays an important role in Step 1 in the proof of Theorem 1.

Lemma 3 *Let a stationary strategy* σ^s *be such that* $\mathsf{M}_{x_0}^{\sigma^s} \in \mathcal{D}^f$. *(E.g.,* σ^s *is the strategy, induced by* $\mathsf{M} \in \mathcal{D}^f$.*) Then the following assertions hold.*

(a)

$$\mathsf{M}_x^{\sigma^s}(\mathbf{X} \times \mathbf{A}) = \mathsf{E}_x^{\sigma^s}\left[\sum_{n=1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\right] < \infty$$

for $\mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})$*-almost all* $x \in \mathbf{X}$.

(b) For a bounded $\mathbb{R}$*-valued function* $f(\cdot)$ *on* $\mathbf{X}$ *with* $f(\Delta) = 0$, *the function*

$$v(x) := \mathsf{E}_x^{\sigma^s}\left[\sum_{n=1}^{\infty} f(X_{n-1})\right] := \mathsf{E}_x^{\sigma^s}\left[\sum_{n=1}^{\infty} f^+(X_{n-1})\right] - \mathsf{E}_x^{\sigma^s}\left[\sum_{n=1}^{\infty} f^-(X_{n-1})\right], \qquad x \in \mathbf{X}$$

is measurable and with finite values for $\mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})$*-almost all* $x \in \mathbf{X}$. *Here the convention of* $\infty - \infty := \infty$ *is in use.*

(c) The function $v(\cdot)$ *in (b) satisfies equation*

$$v(x) = f(x) + \int_{\mathbf{A}} \int_{\mathbf{X}} v(y) p(dy|x,a) \sigma^s(da|x) \qquad \mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})\text{-a.s.} \tag{9}$$

If a measurable bounded function $w(\cdot) : \mathbf{X} \to \mathbb{R}$ *satisfies equation (9), then* $w(x) = v(x)$ *for* $\mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})$*-almost all* $x \in \mathbf{X}$.

We note that the function $v(\cdot)$ in parts (b,c) of the previous lemma may be not finite everywhere, even though the function $f(\cdot)$ was bounded.

Theorem 1 *An occupation measure* $\mathsf{M} \in \mathcal{D}^f$ *is extreme in* $\mathcal{D}^f$ *if and only if* $\mathsf{M} = \mathsf{M}_{x_0}^{\varphi}$ *for some deterministic stationary strategy* φ.

5 Form of the optimal control strategy

In this section, we present our second main result, concerning the existence of an optimal $(J+1)$-mixed strategy to the constrained MDP problem. For this, we will impose further conditions, which, in particular, guarantee that strategies whose occupation measures are infinite are not optimal or feasible, see Theorem 2.

Definition 7 (**$(J+1)$-mixed strategy**) *According to Propositions 1 and 3, if* $\sigma^1, \sigma^2, \ldots, \sigma^{\mathbf{L}}$ *is a finite collection of strategies, then, for a set* $\alpha_1, \alpha_2, \ldots, \alpha_{\mathbf{L}}$ *of non-negative numbers with* $\sum_{l=1}^{\mathbf{L}} \alpha_l = 1$, $\sum_{l=1}^{\mathbf{L}} \alpha_l \mathsf{P}_{x_0}^{\sigma^l}$ *is a strategic measure and* $\sum_{l=1}^{\mathbf{L}} \alpha_l \mathsf{M}_{x_0}^{\sigma^l}$ *is an occupation measure for some strategy* σ. *We call it a mixture of strategies* $\sigma^1, \sigma^2, \ldots, \sigma^{\mathbf{L}}$, *or for brevity, a* $(J+1)$*-mixed strategy.*

Theorem 2 *Suppose the MDP $\{\mathbf{X}_\Delta, \mathbf{A}, p, \{r_j\}_{j=0}^J\}$ with initial state $x_0 \in \mathbf{X}$ is semicontinuous, $r_j(\cdot,\cdot) \geq 0$, $j = 0, 1, \ldots, J$, and there exists a feasible strategy σ with a finite value. Furthermore, assume that, for each strategy σ such that $\mathsf{M}_{x_0}^\sigma \notin \mathcal{D}^f$, there is some $\tilde{j} \in \{0, 1, \ldots, J\}$, possibly depending on σ, satisfying $\int_{\mathbf{X}\times\mathbf{A}} r_{\tilde{j}}(x,a)\mathsf{M}_{x_0}^\sigma(dx \times da) = \infty$.*

Then there exists an optimal strategy in problem (4) in the form of a mixture of $J+1$ deterministic stationary strategies.

The above theorem asserts the existence of an optimal strategy in the form of a mixture of $J+1$ deterministic stationary strategies. It does not claim that every optimal strategy can be represented as a mixture of finitely many deterministic stationary strategies. For completeness, we adapt [29, Example 3.3.3] to demonstrate this.

Example 1 *Consider the MDP with $\boldsymbol{X} = \{0, 1, 2\}$, $\boldsymbol{A} = \{0, 1\}$, $p(\{1\}|1, 0) = 1$, $p(\{2\}|1, 1) = 1$, $p(\{2\}|2, 0) = 1$, $p(\{2\}|2, 1) = p(\{\Delta\}|2, 1) = \frac{1}{2}$, and $p(\{1\}|0, a) = p(\{2\}|0, a) = \frac{1}{2}$ for $a \in \boldsymbol{A}$. The state Δ is a costless cemetery. The state and action spaces are endowed with their discrete topologies.*

Let $x_0 = 0$. Let $r_0(x, a) \equiv 0$, and $r_1(x, a) = 1$ for $x = 1, 2$ and $r_1(0, a) \equiv 0$. Let $J = 1$ and $d_1 = 3$. So any feasible strategy will be optimal, and any non-absorbing strategy σ will be infeasible with $\mathsf{E}_{x_0}^\sigma \left[\sum_{n=0}^\infty r_1(X_n, A_{n+1})\right] = \infty$. All the conditions in Theorem 2 are satisfied.

The class Φ of absorbing deterministic stationary strategies is specified by $\varphi(1) = \varphi(2) = 1$: the state 0 is essentially uncontrolled, and $\varphi(0)$ is immaterial. We put $\hat{a} = 0$. Clearly, $\mathcal{D}^f$ is a proper subset of $\mathcal{D}$, and any strategy that selects 0 at state 1 with probability 1 will be non-absorbing.

Consider a stationary strategy defined by $\sigma^s(\{0\}|1) = \sigma^s(\{1\}|1) = \frac{1}{2}$. Then

$$\mathsf{E}_0^\varphi\left[\sum_{n=0}^\infty r_1(X_n, A_{n+1})\right] =: W_1(0, \varphi) = \frac{1}{2}(1+2) + \frac{1}{2}2 = \frac{5}{2}, \ \forall \varphi \in \Phi;$$

$$\mathsf{E}_0^{\sigma^s}\left[\sum_{n=0}^\infty r_1(X_n, A_{n+1})\right] =: W_1(0, \sigma^s) = \frac{1}{2}(2+2) + \frac{1}{2}2 = 3.$$

Therefore, σ^s and all strategies $\varphi \in \Phi$ are feasible and thus optimal.

On the other hand, $W_1(0, \sigma^s) > W_1(0, \varphi)$ for all $\varphi \in \Phi$, so that the occupation measure of σ^s cannot be represented as the convex combination of the occupation measures of strategies from Φ. Of course, the occupation measure of σ^s cannot be represented as a mixture of occcupation measures of non-absorbing deterministic stationary strategies together with the ones from Φ.

It is well known that, if the MDP is semicontinuous and the cost functions $r(\cdot,\cdot)$ are non-negative, then there exists an optimal solution to the unconstrained problem (4) (i.e., with $J = 0$), which is deterministic stationary: see Corollary 9.17.2 of [3] or Theorems 15.2 and 16.2 of [40]. Therefore, we will assume that $J \geq 1$.

If there are feasible strategies in problem (4), but for all of them $R_0(\mathsf{M}_{x_0}^\sigma) = +\infty$, then all feasible strategies are equally optimal. In this case, the only problem is to find a feasible strategy. To do so, we choose an arbitrary positive index, e.g., $j = 1$, and investigate the problem

$$\begin{aligned} &\text{Minimize over all strategies } \sigma\text{:} && \mathsf{E}^{\sigma}_{x_0}\left[\sum_{n=0}^{\infty} r_1(X_n, A_{n+1})\right] \\ &\text{subject to} && \mathsf{E}^{\sigma}_{x_0}\left[\sum_{n=0}^{\infty} r_j(X_n, A_{n+1})\right] \le d_j, \quad j = 2, 3, \dots, J. \end{aligned}$$

Clearly, after re-enumerating the indices j, we obtain the standard problem (4) with the reduced number of constraints (or just the unconstrained problem in case J was equal to 1). In such situations there is no need to require that the cost function $r_0(\cdot,\cdot)$ exhibits any further properties (semicontinuity etc) except for measurability. After solving the modified problem, we obtain the desired feasible strategy. Clearly, the modified problem has a feasible strategy with a finite value (because the original problem had a feasible strategy). If all the other requirements of Theorem 2 are satisfied for the modified problem, then Theorem 2 remains valid for it. As the result, in such a case, there exists an optimal strategy in the original problem (4) in the form of a mixture of J deterministic stationary strategies.

Let us consider the special case of optimal stopping like in [10]: the action space is $\mathbf{A}_\Delta := \mathbf{A} \cup \{\Delta\}$, where the isolated action Δ means stopping the process: for all $x \in \mathbf{X}$, $p(\{\Delta\}|x, \Delta) = 1$ and $p(\mathbf{X}|x, a) = 1$ for all $a \in \mathbf{A}$. If this MDP $\{\mathbf{X}_\Delta, \mathbf{A}_\Delta, p, \{r_j\}_{j=0}^J\}$ is semicontinuous, $r_j(\cdot,\cdot) \ge 0$, $j = 0, 1, \dots, J$, there exists a feasible strategy with a finite value, and, for some $\tilde{j} \in \{0, 1, \dots, J\}$, $r_{\tilde{j}}(a, x) \ge \delta > 0$ for all $x \in \mathbf{X}$, $a \in \mathbf{A}$, then all the conditions of Theorem 2 are satisfied. $\mathsf{M}^{\sigma}_{x_0} \notin \mathcal{D}^f$ means that the process is never stopped, hence $\int_{\mathbf{X}\times\mathbf{A}} r_{\tilde{j}}(x, a)\mathsf{M}^{\sigma}_{x_0}(dx \times da) = \infty$. According to the above paragraph, one can omit the requirement that the feasible strategy has a finite value.

6 Conclusion

The main results of the current work are Theorems 1 and 2, where we prove that every extreme finite occupation measure is generated by a deterministic stationary strategy, and, under mild conditions, show that the solution to the constrained problem is given by a finite mixture of such strategies. All the similar statements in [1, 4, 5, 18, 16, 22, 31], where the discounted or absorbing models were studied, follow from Theorems 1 and 2.

Appendix

Proof of Lemma 1. Some of the enlisted statements were presented in [9, Lemma 4.1].

(a) The mapping O is continuous, since $\mathcal{D}$ is endowed with the final topology ρ. Thus, $\mathcal{D} = O(\mathcal{P})$ is compact as the continuous image of the compact $\mathcal{P}$, see [13, Chapter I, §5, Lemma 7].

(b) According to Lemma 7.14(a) of [3] , $r(\cdot,\cdot) = \lim_{i\to\infty} r_i(\cdot,\cdot)$, where $r_i(\cdot,\cdot)$ are pointwise increasing bounded continuous functions on $\mathbf{X}\times\mathbf{A}$. For each $i = 1, 2, \dots$ the mapping

$$\mathsf{P}^{\sigma}_{x_0} \to \int_{\mathbf{X}\times\mathbf{A}} r_i(x, a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta \times \mathbf{A})^t \times dx \times da \times (\mathbf{X}_\Delta \times \mathbf{A})^\infty)$$

is continuous for each $t = 0, 1, \dots$ because τ is the weak topology in $\mathcal{P}$. Therefore, the mapping

$$\begin{aligned}\mathsf{P}^{\sigma}_{x_0} &\to \sum_{t=0}^{n}\int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta\times\mathbf{A})^t\times dx\times da\times(\mathbf{X}_\Delta\times\mathbf{A})^\infty)\\ &= \lim_{i\to\infty}\sum_{t=0}^{n}\int_{\mathbf{X}\times\mathbf{A}} r_i(x,a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta\times\mathbf{A})^t\times dx\times da\times(\mathbf{X}_\Delta\times\mathbf{A})^\infty)\end{aligned}$$

is non-negative and lower semicontinuous again due to Lemma 7.14(a) of [3]. The monotone convergence theorem was in use here. Since $r(\cdot,\cdot)\geq 0$, Lemma B.1.1 and Proposition B.1.17 of [34] imply that the mapping

$$\begin{aligned}\mathsf{P}^{\sigma}_{x_0} &\to \sum_{t=0}^{\infty}\int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta\times\mathbf{A})^t\times dx\times da\times(\mathbf{X}_\Delta\times\mathbf{A})^\infty)\\ &= \sup_{n=1,2,\ldots}\sum_{t=0}^{n}\int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta\times\mathbf{A})^t\times dx\times da\times(\mathbf{X}_\Delta\times\mathbf{A})^\infty)\\ &= \int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{M}^{\sigma}_{x_0}(dx\times da) = R(\mathsf{M}^{\sigma}_{x_0}) = R(O(\mathsf{P}^{\sigma}_{x_0}))\end{aligned}$$

is lower semicontinuous. Now, for an arbitrarily fixed $c\in\mathbb{R}$

$$\begin{aligned}&O^{-1}\left(\left\{\mathsf{M}\in\mathcal{D}:\ R(\mathsf{M})=\int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{M}(dx\times da)>c\right\}\right)\\ =&\left\{\mathsf{P}\in\mathcal{P}:\sum_{t=0}^{\infty}\int_{\mathbf{X}\times\mathbf{A}} r(x,a)\mathsf{P}^{\sigma}_{x_0}((\mathbf{X}_\Delta\times\mathbf{A})^t\times dx\times da\times(\mathbf{X}_\Delta\times\mathbf{A})^\infty)>c\right\}.\end{aligned}$$

The set on the right-hand side is open in the topology τ. Hence the set $\{\mathsf{M}\in\mathcal{D}:\ R(M)>c\}$ is open in the topology ρ. □

Proof of Lemma 2. The both measures $\mathsf{M}(dx\times\mathbf{A})$ and $\mathsf{M}^{\sigma^s}_{x_0}(dx\times\mathbf{A})$ are finite and satisfy equation

$$\mu(\Gamma_X)=\delta_{x_0}(\Gamma_X)+\int_{\mathbf{X}}\int_{\mathbf{A}} p(\Gamma_X|y,a)\sigma^s(da|y)\mu(dy),\quad \forall\Gamma_X\in\mathcal{B}(\mathbf{X}): \tag{10}$$

see Proposition 4 and Proposition 5.

Let us show that the measure $\mathsf{M}(dx\times\mathbf{A})$ is absolutely continuous wrt $\mathsf{M}^{\sigma^s}_{x_0}(dx\times\mathbf{A})$ on $\mathcal{B}(\mathbf{X})$.

Suppose for contradiction that $\mathsf{M}(\Gamma\times\mathbf{A})>0$ and $\mathsf{M}^{\sigma^s}_{x_0}(\Gamma\times\mathbf{A})=0$ for some $\Gamma\in\mathcal{B}(\mathbf{X})$. Denote $\Gamma_0:=\Gamma$ and, for $n=0,1,\ldots$, put

$$\begin{aligned}\tilde{\Gamma}_{n+1} &:= \left\{y\in\mathbf{X}:\ \int_{\mathbf{A}} p(\Gamma_n|y,a)\sigma^s(da|y)>0\right\};\\ \Gamma_{n+1} &:= \Gamma_n\cup\tilde{\Gamma}_{n+1}.\end{aligned}$$

Intuitively, Γ_{n+1} is the set of states, starting from which, the state process under σ^s visits Γ with positive probability within $n+1$ steps. We will prove by induction that, for all $n=0,1,\ldots$,

$$\begin{aligned}&\mathsf{M}(\Gamma_n\times\mathbf{A})>0;\\ &\int_{\mathbf{X}\setminus\Gamma_{n+1}}\int_{\mathbf{A}} p(\Gamma_n|y,a)\sigma^s(da|y)\mathsf{M}(dy\times\mathbf{A})=0;\\ &\mathsf{M}^{\sigma^s}_{x_0}(\Gamma_n\times\mathbf{A})=0.\end{aligned}$$

When $n = 0$, these assertions obviously hold because $\int_{\mathbf{A}} p(\Gamma_0|y,a)\sigma^s(da|y) = 0$ for all $y \in \mathbf{X} \setminus \tilde{\Gamma}_1$. Suppose they hold for some $n \geq 0$ and consider the case of $n+1$.

Since $\Gamma_{n+1} \supseteq \Gamma_n$, $\mathsf{M}(\Gamma_{n+1} \times \mathbf{A}) > 0$. Suppose $\mathsf{M}^{\sigma^s}_{x_0}(\tilde{\Gamma}_{n+1} \times \mathbf{A}) > 0$. Then, by (10),

$$\mathsf{M}^{\sigma^s}_{x_0}(\Gamma_n \times \mathbf{A}) \geq \int_{\tilde{\Gamma}_{n+1}} \int_{\mathbf{A}} p(\Gamma_n|y,a)\sigma^s(da|y)\mathsf{M}^{\sigma^s}_{x_0}(dy \times \mathbf{A}) > 0,$$

which contradicts the inductive supposition. Thus, $\mathsf{M}^{\sigma^s}_{x_0}(\tilde{\Gamma}_{n+1} \times \mathbf{A}) = \mathsf{M}^{\sigma^s}_{x_0}(\Gamma_{n+1} \times \mathbf{A}) = 0$. Finally,

$$\int_{\mathbf{X} \setminus \Gamma_{n+2}} \int_{\mathbf{A}} p(\Gamma_{n+1}|y,a)\sigma^s(da|y)\mathsf{M}(dy \times \mathbf{A}) = 0$$

because $\int_{\mathbf{A}} p(\Gamma_{n+1}|y,a)\sigma^s(da|y) = 0$ for all $y \in \mathbf{X} \setminus \tilde{\Gamma}_{n+2} \supseteq \mathbf{X} \setminus \Gamma_{n+2}$.

Therefore, for the increasing sequence $\{\Gamma_n\}_{n=0}^{\infty}$, after we denote $\hat{\Gamma} := \bigcup_{n=0}^{\infty} \Gamma_n$, we have $\mathsf{M}(\hat{\Gamma} \times \mathbf{A}) > 0$ and, by the monotone convergence theorem,

$$\begin{aligned}
&\int_{(\mathbf{X}\setminus\hat{\Gamma})\times\mathbf{A}} p(\hat{\Gamma}|y,a)\mathsf{M}(dy \times da) = \int_{\mathbf{X}\setminus\hat{\Gamma}} \int_{\mathbf{A}} p(\hat{\Gamma}|y,a)\sigma^s(da|y)\mathsf{M}(dy \times \mathbf{A}) \\
= &\lim_{n\to\infty} \int_{\mathbf{X}\setminus\hat{\Gamma}} \int_{\mathbf{A}} p(\Gamma_n|y,a)\sigma^s(da|y)\mathsf{M}(dy \times \mathbf{A}) \\
\leq &\lim_{n\to\infty} \int_{\mathbf{X}\setminus\Gamma_{n+1}} \int_{\mathbf{A}} p(\Gamma_n|y,a)\sigma^s(da|y)\mathsf{M}(dy \times \mathbf{A}) = 0. \qquad (11)
\end{aligned}$$

Note also that $x_0 \notin \hat{\Gamma}$ because $\mathsf{M}^{\sigma^s}_{x_0}(\hat{\Gamma} \times \mathbf{A}) = \lim_{n\to\infty} \mathsf{M}^{\sigma^s}_{x_0}(\Gamma_n \times \mathbf{A}) = 0$.

Recall that $\mathsf{M} \in \mathcal{D}^f$. According to Proposition 2, $\mathsf{M} = \mathsf{M}^{\sigma^M}_{x_0}$ for some Markov strategy σ^M. Since $\mathsf{M}(\hat{\Gamma} \times \mathbf{A}) > 0$ and $x_0 \notin \hat{\Gamma}$, there exists the minimal $n > 0$ such that $\mathsf{P}^{\sigma^M}_{x_0}(X_n \in \hat{\Gamma}) > 0$, for which we have the following equalities:

$$\begin{aligned}
0 < \mathsf{P}^{\sigma^M}_{x_0}(X_n \in \hat{\Gamma}) &= \mathsf{P}^{\sigma^M}_{x_0}(\mathsf{P}^{\sigma^M}_{x_0}(X_n \in \hat{\Gamma}|X_{n-1})) \\
&= \int_{\mathbf{X}} \int_{\mathbf{A}} p(\hat{\Gamma}|y,a)\sigma^M_n(da|y)\mathsf{P}^{\sigma^M}_{x_0}(X_{n-1} \in dy) \\
&= \int_{\mathbf{X}\setminus\hat{\Gamma}} \int_{\mathbf{A}} p(\hat{\Gamma}|y,a)\sigma^M_n(da|y)\mathsf{P}^{\sigma^M}_{x_0}(X_{n-1} \in dy) \\
&= \int_{(\mathbf{X}\setminus\hat{\Gamma})\times\mathbf{A}} p(\hat{\Gamma}|y,a)\mathsf{P}^{\sigma^M}_{x_0}(X_{n-1} \in dy, A_n \in da),
\end{aligned}$$

where the third equality holds by the definition of the integer n. Hence,

$$\int_{(\mathbf{X}\setminus\hat{\Gamma})\times\mathbf{A}} p(\hat{\Gamma}|y,a)\mathsf{M}^{\sigma^M}_{x_0}(dy \times da) = \int_{(\mathbf{X}\setminus\hat{\Gamma})\times\mathbf{A}} p(\hat{\Gamma}|y,a)\mathsf{M}(dy \times da) > 0,$$

which contradicts (11). We have proved that $\mathsf{M}(dx \times \mathbf{A}) \ll \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$.

Let us define the following substochastc kernels on $\mathcal{B}(\mathbf{X})$ given $x \in \mathbf{X}$:

$$\begin{aligned}
\mathsf{P}^0(\Gamma|x) &:= \delta_x(\Gamma); \\
\mathsf{P}^{n+1}(\Gamma|x) &:= \int_{\mathbf{X}} \int_{\mathbf{A}} p(\Gamma|y,a)\sigma^s(da|y)\mathsf{P}^n(dy|x), \quad n = 0,1,\ldots \\
&\quad\ \Gamma \in \mathcal{B}(\mathbf{X}).
\end{aligned}$$

Then $\mathsf{P}^i(\Gamma|x_0) = \mathsf{P}^{\sigma^s}_{x_0}(X_i \in \Gamma)$, $i = 0, 1, 2, \ldots$. Now, for any finite measure μ on $\mathcal{B}(\mathbf{X})$, satisfying equation (10), we have the following iterations of this equation:

$$\begin{aligned}
\mu(\Gamma) &= \delta_{x_0}(\Gamma) + \int_{\mathbf{X}}\int_{\mathbf{A}} p(\Gamma|x_0)\sigma^s(da|x_0) \\
&\quad + \int_{\mathbf{X}}\int_{\mathbf{A}} p(\Gamma|y,a)\sigma^s(da|y)\left(\int_{\mathbf{X}}\int_{\mathbf{A}} p(dy|x,a)\sigma^s(da|x)\mu(dx)\right) \\
&= \mathsf{P}^0(\Gamma|x_0) + \mathsf{P}^1(\Gamma|x_0) + \int_{\mathbf{X}} \mathsf{P}^2(\Gamma|x)\mu(dx) \\
&= \mathsf{P}^0(\Gamma|x_0) + \mathsf{P}^1(\Gamma|x_0) + \mathsf{P}^2(\Gamma|x_0) + \int_{\mathbf{X}} \mathsf{P}^3(\Gamma|x)\mu(dx) \\
&= \ldots = \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{i=1}^{n} \mathbb{I}\{x_{i-1} \in \Gamma\}\right] + \int_{\mathbf{X}} \mathsf{P}^n(\Gamma|x)\mu(dx), \qquad (12) \\
&\quad n = 1, 2, \ldots.
\end{aligned}$$

Here the Fubini Theorem was in use, and the last equality holds because

$$\mathsf{P}^i(\Gamma|x_0) = \mathsf{P}^{\sigma^s}_{x_0}(X_i \in \Gamma), \quad i = 0, 1, 2, \ldots.$$

Since $p(\mathbf{X}|y,a) \leq 1$, for each $x \in \mathbf{X}$ the sequence $\{\mathsf{P}^i(\mathbf{X}|x)\}_{i=0}^{\infty}$ is monotonically non-increasing, so that there exists the limit $\mathsf{P}^\infty(\mathbf{X}|x) := \lim_{i\to\infty} \mathsf{P}^i(\mathbf{X}|x)$, and the function $\mathsf{P}^\infty(\mathbf{X}|\cdot): \mathbf{X} \to [0,1]$ is obviously measurable. By the dominated convergence theorem,

$$\lim_{n\to\infty} \int_{\mathbf{X}} \mathsf{P}^n(\mathbf{X}|x)\mu(dx) = \int_{\mathbf{X}} \mathsf{P}^\infty(\mathbf{X}|x)\mu(dx).$$

Therefore, if we substitute $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$ for $\mu(dx)$ in (12), we obtain

$$\mathsf{M}^{\sigma^s}_{x_0}(\mathbf{X} \times \mathbf{A}) = \lim_{n\to\infty} \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{i=1}^{n} \mathbb{I}\{X_{i-1} \in \mathbf{X}\}\right] + \int_{\mathbf{X}} \mathsf{P}^\infty(\mathbf{X}|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A}),$$

leading to the equation $\int_{\mathbf{X}} \mathsf{P}^\infty(\mathbf{X}|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A}) = 0$, because $\mathsf{M}^{\sigma^s}_{x_0}(\mathbf{X} \times \mathbf{A}) = \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{i=1}^{\infty} \mathbb{I}\{X_{i-1} \in \mathbf{X}\}\right] < \infty$: the both measures $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$ and $\mathsf{M}(dx \times \mathbf{A})$ satisfy equation (8), and $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A}) \leq \mathsf{M}(dx \times \mathbf{A})$ by Proposition 5. Recall that $\mathsf{M} \in \mathcal{D}^f$. Since $\mathsf{P}^\infty(\mathbf{X}|x) \geq 0$, we conclude that $\mathsf{P}^\infty(\mathbf{X}|x) = 0$ $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$-a.s. and $\mathsf{P}^\infty(\mathbf{X}|x) = 0$ $\mathsf{M}(dx \times \mathbf{A})$-a.s. because $\mathsf{M}(dx \times \mathbf{A}) \ll \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$. Hence, for each $\Gamma \in \mathcal{B}(\mathbf{X})$,

$$\begin{aligned}
0 &\leq \limsup_{n\to\infty} \int_{\mathbf{X}} \mathsf{P}^n(\Gamma|x)\mathsf{M}(dx \times \mathbf{A}) \leq \lim_{n\to\infty} \int_{\mathbf{X}} \mathsf{P}^n(\mathbf{X}|x)\mathsf{M}(dx \times \mathbf{A}) \\
&= \int_{\mathbf{X}} \mathsf{P}^\infty(\mathbf{X}|x)\mathsf{M}(dx \times \mathbf{A}) = 0,
\end{aligned}$$

and thus $\lim_{n\to\infty} \int_{\mathbf{X}} \mathsf{P}^n(\Gamma|x)\mathsf{M}(dx \times \mathbf{A}) = 0$. After we substitute $\mathsf{M}(dx \times \mathbf{A})$ for $\mu(dx)$ in (12), we obtain

$$\begin{aligned}
\mathsf{M}(\Gamma \times \mathbf{A}) &= \lim_{n\to\infty} \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{i=1}^{n} \mathbb{I}\{X_{i-1} \in \Gamma\}\right] \\
&\quad + \lim_{n\to\infty} \int_{\mathbf{X}} \mathsf{P}^n(\Gamma|x)\mathsf{M}(dx \times \mathbf{A}) = \mathsf{M}^{\sigma^s}_{x_0}(\Gamma \times \mathbf{A}).
\end{aligned}$$

Finally,

$$\begin{aligned}\mathsf{M}(\Gamma_X \times \Gamma_A) &= \int_{\mathbf{X}} \sigma^s(\Gamma_A|x)\mathsf{M}(dx \times \mathbf{A}) \\ &= \int_{\mathbf{X}} \sigma^s(\Gamma_A|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A}) = \mathsf{M}^{\sigma^s}_{x_0}(\Gamma_X \times \Gamma_A), \\ &\qquad \forall \Gamma_X \in \mathcal{B}(\mathbf{X}),\ \Gamma_A \in \mathcal{B}(\mathbf{A}).\end{aligned}$$

□

Proof of Lemma 3. Note that, if a statement $S(X_m)$ is valid $\mathsf{P}^{\sigma^s}_{x_0}$-a.s. for all $m = 0, 1, 2, \ldots$, then the statement $S(x)$ is valid for $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$ and vice versa.

(a) If the formulated statement does not hold, then there is a set $\Gamma \in \mathcal{B}(\mathbf{X})$ such that, for some $m \geq 0$, $\mathsf{P}^{\sigma^s}_{x_0}(X_m \in \Gamma) > 0$ and

$$\mathsf{E}^{\sigma^s}_x \left[\sum_{n=1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\right] = \infty \quad \forall x \in \Gamma.$$

Now

$$\begin{aligned}\mathsf{M}^{\sigma^s}_{x_0}(\mathbf{X} \times \mathbf{A}) &\geq \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=m+1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\right] \\ &= \mathsf{E}^{\sigma^s}_{x_0}\left[\mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=m+1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\middle| X_m\right]\right] \\ &\geq \int_{\Gamma} \mathsf{E}^{\sigma^s}_x \left[\sum_{n=1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\right] \mathsf{P}^{\sigma^s}_{x_0}(X_m \in dx) = +\infty.\end{aligned}$$

Here

$$\mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=m+1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\middle| X_m\right] = \mathsf{E}^{\sigma^s}_{X_m}\left[\sum_{n=1}^{\infty} \mathbb{I}\{X_{n-1} \in \mathbf{X}\}\right]$$

because the controlled process $\{X_n\}_{n=0}^{\infty}$, under the strategy σ^s, is Markov and time-homogeneous.

The obtained contradiction with the assumption $\mathsf{M}^{\sigma^s}_{x_0} \in \mathcal{D}^f$ proves the statement.

(b) Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be the natural filtration $\mathcal{F}_n := \sigma\{X_0, X_1, \ldots, X_n\}$ of the Markov time-homogeneous process $\{X_n\}_{n=0}^{\infty}$ under the control strategy σ^s and with the initial state $x_0 \in \mathbf{X}$. For the positive and negative parts of $f(\cdot)$, we have the following relations for each fixed $m \geq 0$:

$$\begin{aligned}0 &\leq \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=m+1}^{\infty} f^{\pm}(X_{n-1})\middle| \mathcal{F}_m\right] \\ &= \mathsf{E}^{\sigma^s}_{X_m}\left[\sum_{n=1}^{\infty} f^{\pm}(X_{n-1})\right] < \infty \qquad \mathsf{P}^{\sigma^s}_{x_0}\text{-a.s.}\end{aligned}$$

The very last inequality holds by (a) and the boundedness of the function $f(\cdot)$. Therefore, the function $v(\cdot)$ is with finite values for $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$.

The measurability of $v(\cdot)$ follows from [15, Theorem 3.1]. Statement (b) is proved.

(c) Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be the natural filtration $\mathcal{F}_n := \sigma\{X_0, X_1, \ldots, X_n\}$ of the Markov time-homogeneous process $\{X_n\}_{n=0}^{\infty}$ under the control strategy σ^s and with the initial state $x \in \mathbf{X}$. According to (b),

$$\begin{aligned} v(x) &= f(x) + \mathsf{E}_x^{\sigma^s}\left[\mathsf{E}_x^{\sigma^s}\left[\left.\sum_{n=2}^{\infty} f(X_{n-1})\right| \mathcal{F}_1\right]\right] \\ &= f(x) + \mathsf{E}_x^{\sigma^s}[v(X_1)] \\ &= f(x) + \int_{\mathbf{A}}\int_{\mathbf{X}} v(y)p(dy|x,a)\sigma^s(da|x) \quad \mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})\text{-a.s.} \end{aligned}$$

Here, all the terms are finite $\mathsf{M}_{x_0}^{\sigma^s}(dx \times \mathbf{A})$-a.s. by (b). Equation (9) is proved.

Let us fix an arbitrary $i \in \{0, 1, \ldots\}$ and the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ corresponding to the initial state $x_0 \in \mathbf{X}$. For the function $w(\cdot)$, we have the following obvious equation

$$\mathsf{E}_{x_0}^{\sigma^s}[w(X_{i+1})|\mathcal{F}_i] = \int_{\mathbf{A}}\int_{\mathbf{X}} w(y)p(dy|X_i,a)\sigma^s(da|X_i) \quad \mathsf{P}_{x_0}^{\sigma^s}\text{-a.s.} \tag{13}$$

We are going to prove by induction the following equality

$$w(X_i) = \mathsf{E}_{x_0}^{\sigma^s}\left[\left.\sum_{j=0}^{k} f(X_{i+j})\right| \mathcal{F}_i\right] + \mathsf{E}_{x_0}^{\sigma^s}[w(X_{i+k+1})|\mathcal{F}_i] \quad \mathsf{P}_{x_0}^{\sigma^s}\text{-a.s.} \tag{14}$$

for $k = 0, 1, \ldots$.

When $k = 0$, equality (14) follows from equation (9) for $w(\cdot)$ and (13). Suppose it holds for some $k \geq 0$. Then

$$\begin{aligned} w(X_i) &= \mathsf{E}_{x_0}^{\sigma^s}\left[\left.\sum_{j=0}^{k} f(X_{i+j})\right| \mathcal{F}_i\right] + \mathsf{E}_{x_0}^{\sigma^s}[f(X_{i+k+1})|\mathcal{F}_i] \\ &\quad + \mathsf{E}_{x_0}^{\sigma^s}\left[\left.\int_{\mathbf{A}}\int_{\mathbf{X}} w(y)p(dy|X_{i+k+1},a)\sigma^s(da|X_{i+k+1})\right| \mathcal{F}_i\right] \\ &= \mathsf{E}_{x_0}^{\sigma^s}\left[\left.\sum_{j=0}^{k+1} f(X_{i+j})\right| \mathcal{F}_i\right] + \mathsf{E}_{x_0}^{\sigma^s}[\mathsf{E}_{x_0}^{\sigma^s}[w(X_{i+k+2})|\mathcal{F}_{i+k+1}]|\mathcal{F}_i] \\ &= \mathsf{E}_{x_0}^{\sigma^s}\left[\left.\sum_{j=0}^{k+1} f(X_{i+j})\right| \mathcal{F}_i\right] + \mathsf{E}_{x_0}^{\sigma^s}[w(X_{i+k+2})|\mathcal{F}_i] \quad \mathsf{P}_{x_0}^{\sigma^s}\text{-a.s.} \end{aligned}$$

Here, the first equality is by (9), and the second equality is by (13). Equality (14) is proved.

When $k \to \infty$, since the process $\{X_n\}_{n=0}^{\infty}$ is Markov and time-homogeneous,

$$\mathsf{E}_{x_0}^{\sigma^s}\left[\left.\sum_{j=0}^{k} f(X_{i+j})\right| \mathcal{F}_i\right] = \mathsf{E}_{X_i}^{\sigma^s}\left[\sum_{n=1}^{k+1} f(X_{n-1})\right] \to v(X_i) \quad \mathsf{P}_{x_0}^{\sigma^s}\text{-a.s.}$$

by the definition of the function $v(\cdot)$. According to (a),

$$\mathsf{E}_{x_0}^{\sigma^s}[\mathbb{I}\{X_{i+k+1} \in \mathbf{X}\}|\mathcal{F}_i] = \mathsf{E}_{X_i}^{\sigma^s}[\mathbb{I}\{X_{k+1} \in \mathbf{X}\}] \to 0 \text{ as } k \to \infty \ \mathsf{P}_{x_0}^{\sigma^s}\text{-a.s.}$$

Therefore, since the function $w(\cdot)$ is bounded, $w(X_i) = v(X_i)$ $\mathsf{P}_{x_0}^{\sigma^s}$-a.s., as required. □

Proof of Theorem 1. We assume that $\mathcal{D}^f \neq \emptyset$ and, according to Lemma 2, consider only the occupation measures $\mathsf{M} = \mathsf{M}_{x_0}^{\sigma^s} \in \mathcal{D}^f$ coming from stationary strategies σ^s.

(a) The 'if' part. We will prove a little more general statement: if $\mathsf{M}_{x_0}^{\varphi} \in \mathcal{D}^f$ is the occupation measure generated by a deterministic stationary strategy φ, then $\mathsf{M}_{x_0}^{\varphi}$ is extreme in $\mathcal{D}$ (and certainly in $\mathcal{D}^f$, too).

Suppose $\mathsf{M}_{x_0}^{\varphi} = \alpha \mathsf{M}_1 + (1-\alpha)\mathsf{M}_2$ with $\alpha \in (0,1)$ and $\mathsf{M}_{1,2} \in \mathcal{D}$. Then $\mathsf{M}_{1,2} \in \mathcal{D}^f$ because $\mathsf{M}_{x_0}^{\varphi} \in \mathcal{D}^f$ and, according to Lemma 2, $\mathsf{M}_{1,2} = \mathsf{M}_{x_0}^{\sigma_{1,2}^s}$ for the induced stationary strategies $\sigma_{1,2}^s$. Therefore,

$$\mathsf{M} = \mathsf{M}_{x_0}^{\varphi} = \alpha \mathsf{M}_{x_0}^{\sigma_1^s} + (1-\alpha)\mathsf{M}_{x_0}^{\sigma_2^s}. \tag{15}$$

The goal is to show that $\mathsf{M}_{x_0}^{\sigma_1^s} = \mathsf{M}_{x_0}^{\sigma_2^s} = \mathsf{M}_{x_0}^{\varphi}$.

The both marginal measures $\mathsf{M}_{x_0}^{\sigma_1^s}(dx \times \mathbf{A})$ and $\mathsf{M}_{x_0}^{\sigma_2^s}(dx \times \mathbf{A})$ are absolutely continuous wrt $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$; the Radon-Nikodym derivatives are denoted as $h_1(\cdot)$ and $h_2(\cdot)$ correspondingly. From (15) we have

$$\alpha h_1(x) + (1-\alpha)h_2(x) = 1 \text{ for } \mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})\text{-almost all } x \in \mathbf{X}. \tag{16}$$

Now, using (7), we have equalities

$$\begin{aligned}
\mathsf{M}_{x_0}^{\varphi}(dx \times da) &= \mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})\delta_{\varphi(x)}(da) \\
&= \alpha \mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A}) h_1(x) \sigma_1^s(da|x) \\
&\quad + (1-\alpha)\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A}) h_2(x) \sigma_2^s(da|x) \\
&= \mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})[\alpha h_1(x)\sigma_1^s(da|x) + (1-\alpha)h_2(x)\sigma_2^s(da|x)].
\end{aligned}$$

The expression in the square brackets is the Dirac measure $\delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$. Note that any Dirac measure on $\mathcal{B}(\mathbf{A})$ is extreme in $\mathcal{P}(\mathbf{A})$. Therefore, using (16), we conclude that

- on the set $\mathbf{I}_0 := \{x \in \mathbf{X} : \ \alpha h_1(x) \in (0,1)\}$, $\sigma_1^s(da|x) = \sigma_2^s(da|x) = \delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$-almost all $x \in \mathbf{I}_0$;
- on the set $\mathbf{I}_1 := \{x \in \mathbf{X} : \ \alpha h_1(x) = 1\}$, $\sigma_1^s(da|x) = \delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$-almost all $x \in \mathbf{I}_1$, and the stochastic kernel $\sigma_2^s(da|x)$ may be arbitrary, but on the set $\mathbf{I}_1$, since $(1-\alpha) > 0$, $h_2(x) = 0$ for $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$-almost all $x \in \mathbf{I}_1$, i.e., $\mathsf{M}_{x_0}^{\sigma_2^s}(\mathbf{I}_1 \times \mathbf{A}) = 0$, and the values of $\sigma_2^s(da|x)$ are of no importance for the measure $\mathsf{M}_{x_0}^{\sigma_2^s}(dx \times da)$ on $\mathcal{B}(\mathbf{I}_1 \times \mathbf{A})$;
- symmetrically, on the set $\mathbf{I}_2 := \{x \in \mathbf{X} : \ (1-\alpha)h_2(x) = 1\}$, $\sigma_2^s(da|x) = \delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$-almost all $x \in \{\mathbf{I}_2$ and $\mathsf{M}_{x_0}^{\sigma_1^s}(\mathbf{I}_2 \times \mathbf{A}) = 0$.

Recall that the measures $\mathsf{M}_{x_0}^{\sigma_1^s}(dx \times \mathbf{A})$ and $\mathsf{M}_{x_0}^{\sigma_2^s}(dx \times \mathbf{A})$ are absolutely continuous wrt $\mathsf{M}_{x_0}^{\varphi}(dx \times \mathbf{A})$. Thus, $\sigma_1^s(da|x) = \delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\sigma_1^s}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$ and $\sigma_2^s(da|x) = \delta_{\varphi(x)}(da)$ for $\mathsf{M}_{x_0}^{\sigma_2^s}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$. Hence, φ is the stationary strategy induced by $\mathsf{M}_{x_0}^{\sigma_{1,2}^s}$. and

$$\mathsf{M}_{x_0}^{\sigma_1^s} = \mathsf{M}_{x_0}^{\sigma_2^s} = \mathsf{M}_{x_0}^{\varphi}.$$

by Lemma 2.

(b) The 'only if' part. According to Lemma 2, an extreme point M in $\mathcal{D}^f$ satisfies the equalities $\mathsf{M}(dx \times da) = \mathsf{M}^{\sigma^s}_{x_0}(dx \times da) = \sigma^s(da|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$ on $\mathcal{B}(\mathbf{X} \times \mathbf{A})$ for the induced stationary strategy σ^s.

Step 1. Suppose that

$$\sigma^s(da|x) = \alpha\sigma^s_1(da|x) + (1-\alpha)\sigma^s_2(da|x),$$

where $\alpha \in (0,1)$ and σ^s_1 and σ^s_2 are two essentially different stochastic kernels on $\mathbf{A}$ given $\mathbf{X}$. To be precise, we assume that, for some $\hat{\Gamma}^A \in \mathcal{B}(\mathbf{A})$ and $\hat{\Gamma}^X \in \mathcal{B}(\mathbf{X})$,

$$\mathsf{M}^{\sigma^s}_{x_0}(\hat{\Gamma}^X \times \mathbf{A}) > 0 \text{ and } \sigma^s_2(\hat{\Gamma}^A|x) > \sigma^s_1(\hat{\Gamma}^A|x) \text{ for all } x \in \hat{\Gamma}^X.$$

The stochastic kernels $\sigma^s_{1,2}$ define the corresponding stationary strategies, again denoted as $\sigma^s_{1,2}$. We will show that, in this case, the measure $\mathsf{M}^{\sigma^s}_{x_0}$ is not extreme in $\mathcal{D}^f$.

If $\mathsf{M}^{\sigma^s_1}_{x_0}(dx \times \mathbf{A}) = \mathsf{M}^{\sigma^s_2}_{x_0}(dx \times \mathbf{A}) = \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$, then, by Lemma 5,

$$\begin{aligned}
\mathsf{M}^{\sigma^s}_{x_0}(dx \times da) &= \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})\sigma^s(da|x) \\
&= \alpha\mathsf{M}^{\sigma^s_1}_{x_0}(dx \times \mathbf{A})\sigma^s_1(da|x) + (1-\alpha)\mathsf{M}^{\sigma^s_2}_{x_0}(dx \times \mathbf{A})\sigma_2(da|x) \\
&= \alpha\mathsf{M}^{\sigma^s_1}_{x_0}(dx \times da) + (1-\alpha)\mathsf{M}^{\sigma^s_2}_{x_0}(dx \times da).
\end{aligned}$$

Therefore, the measure $\mathsf{M}^{\sigma^s}_{x_0}$ is not extreme in $\mathcal{D}^f$, as $\mathsf{M}^{\sigma^s_1}_{x_0} \neq \mathsf{M}^{\sigma^s_2}_{x_0}$ and $\mathsf{M}^{\sigma^s_1}_{x_0}, \mathsf{M}^{\sigma^s_2}_{x_0} \in \mathcal{D}^f$. (Recall that $\mathcal{D}^f$ is a face of $\mathcal{D}$.)

Suppose now that $\mathsf{M}^{\sigma^s_1}_{x_0}(dx \times \mathbf{A}) \neq \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$ or $\mathsf{M}^{\sigma^s_2}_{x_0}(dx \times \mathbf{A}) \neq \mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$. There exists the first moment $\tau > 0$ such that

$$\text{either } \mathsf{P}^{\sigma^s_1}_{x_0}(X_\tau \in dx) \neq \mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx) \text{ or } \mathsf{P}^{\sigma^s_2}_{x_0}(X_\tau \in dx) \neq \mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx).$$

Without loss of generality we assume that the first inequality holds. What actually happens is that the both inequalities hold simultaneously at the moment τ: see (19).

Since

$$\mathsf{P}^{\sigma^s_1}_{x_0}(X_{\tau-1} \in dx) = \mathsf{P}^{\sigma^s_2}_{x_0}(X_{\tau-1} \in dx) = \mathsf{P}^{\sigma^s}_{x_0}(X_{\tau-1} \in dx),$$

we have equality

$$\mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx) = \alpha\mathsf{P}^{\sigma^s_1}_{x_0}(X_\tau \in dx) + (1-\alpha)\mathsf{P}^{\sigma^s_2}_{x_0}(X_\tau \in dx) \tag{17}$$

because

$$\sigma^s(da|x) = \alpha\sigma^s_1(da|x) + (1-\alpha)\sigma^s_2(da|x).$$

Note, the controlled process $\{X_n\}_{n=0}^\infty$ is Markov and time-homogeneous under all strategies σ^s and $\sigma^s_{1,2}$, with the transition probabilities

$$\int_{\mathbf{A}} p(dy|x,a)\sigma^s(da|x) \quad \text{and} \quad \int_{\mathbf{A}} p(dy|x,a)\sigma^s_{1,2}(da|x)$$

correspondingly.

Let us introduce the Markov (non-stationary) strategies σ^{M_1} and σ^{M_2} by the formulae

$$\sigma^{M_{1,2}}_n(da|x) = \mathbb{I}\{n \neq \tau\}\sigma^s(da|x) + \mathbb{I}\{n = \tau\}\sigma^s_{1,2}(da|x).$$

The combination of strategic measures $\alpha\mathsf{P}^{\sigma^{M_1}}_{x_0} + (1-\alpha)\mathsf{P}^{\sigma^{M_2}}_{x_0}$ satisfies the key properties of the strategic measure $\mathsf{P}^{\sigma^s}_{x_0}$: see (1),(2),(3), or formula (1.7) in [31], or [27, §2.2.3]. Thus,

$$\mathsf{P}^{\sigma^s}_{x_0} = \alpha\mathsf{P}^{\sigma^{M_1}}_{x_0} + (1-\alpha)\mathsf{P}^{\sigma^{M_2}}_{x_0} \Longrightarrow \mathsf{M}^{\sigma^s}_{x_0} = \alpha\mathsf{M}^{\sigma^{M_1}}_{x_0} + (1-\alpha)\mathsf{M}^{\sigma^{M_2}}_{x_0} \tag{18}$$

and, like previously, $\mathsf{M}^{\sigma^{M_1}}_{x_0}, \mathsf{M}^{\sigma^{M_2}}_{x_0} \in \mathcal{D}^f$ because $\mathsf{M}^{\sigma^s}_{x_0} \in \mathcal{D}^f$. We aim to show that $\mathsf{M}^{\sigma^{M_1}}_{x_0}(dx \times \mathbf{A}) \neq \mathsf{M}^{\sigma^{M_2}}_{x_0}(dx \times \mathbf{A})$ on $\mathcal{B}(\mathbf{X})$, leading to the desired assertion that $\mathsf{M}^{\sigma^s}_{x_0}$ is not extreme in $\mathcal{D}^f$.

Since the strategies σ^s, $\sigma^s_{1,2}$ and $\sigma^{M_{1,2}}$ are Markov and $\sigma^s = \alpha\sigma^s_1 + (1-\alpha)\sigma^s_2$, we have the following relations:

$$\begin{aligned}
\mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx) &= \alpha\mathsf{P}^{\sigma^{M_1}}_{x_0}(X_\tau \in dx) + (1-\alpha)\mathsf{P}^{\sigma^{M_2}}_{x_0}(X_\tau \in dx)\\
&= \alpha\mathsf{P}^{\sigma^s_1}_{x_0}(X_\tau \in dx) + (1-\alpha)\mathsf{P}^{\sigma^s_2}_{x_0}(X_\tau \in dx);\\
\mathsf{P}^{\sigma^{M_1}}_{x_0}(X_\tau \in dx) &= \mathsf{P}^{\sigma^s_1}_{x_0}(X_\tau \in dx) \neq \mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx);\\
\mathsf{P}^{\sigma^{M_2}}_{x_0}(X_\tau \in dx) &= \mathsf{P}^{\sigma^s_2}_{x_0}(X_\tau \in dx) \neq \mathsf{P}^{\sigma^{M_1}}_{x_0}(X_\tau \in dx),\ \ \mathsf{P}^{\sigma^s}_{x_0}(X_\tau \in dx).
\end{aligned} \tag{19}$$

The first three lines here are according to the definitions of τ and of the strategies $\sigma^{M_{1,2}}$ (see also (17)), and inequalities (19) follow from them.

Let $\Gamma \in \mathcal{B}(\mathbf{X})$ be such that $\mathsf{P}^{\sigma^{M_1}}_{x_0}(X_\tau \in \Gamma) \neq \mathsf{P}^{\sigma^{M_2}}_{x_0}(X_\tau \in \Gamma)$. We fix the following bounded functions on $\mathbf{X}_\Delta$, equal to zero on Δ:

$$\begin{aligned}
h(x) &:= \mathbb{I}\{x \in \Gamma\};\\
f(x) &:= h(x) - \int_{\mathbf{A}}\int_{\mathbf{X}} h(y)p(dy|x,a)\sigma^s(da|x).
\end{aligned}$$

According to Lemma 3(b,c), the function

$$v(x) := \mathsf{E}^{\sigma^s}_x\left[\sum_{n=1}^{\infty} f(X_{n-1})\right], \qquad x \in \mathbf{X}$$

is measurable and equals $h(x)$ for $\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$-almost all $x \in \mathbf{X}$ because $h(\cdot)$ satisfies equation (9).

Let $\{\mathcal{F}\}_{t=0}^\infty$ be the natural filtration of the process $\{X_n\}_{n=0}^\infty$, i.e., $\mathcal{F}_t := \sigma\{X_0, X_1, \dots, X_t\}$. According to the definition of the strategies $\sigma^{M_{1,2}}$, $\sigma^{M_{1,2}}_n = \sigma^s$ for $n > \tau$, so

$$\begin{aligned}
\mathsf{E}^{\sigma^{M_{1,2}}}_{x_0}\left[\sum_{n=\tau+1}^{\infty} f(X_{n-1})\middle|\mathcal{F}_\tau\right] &= \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=\tau+1}^{\infty} f(X_{n-1})\middle|\mathcal{F}_\tau\right]\\
= \mathsf{E}^{\sigma^s}_{X_\tau}\left[\sum_{n=1}^{\infty} f(X_{n-1})\right] &= h(X_\tau) \quad \mathsf{P}^{\sigma^s}_{x_0}\text{-a.s., and thus } \mathsf{P}^{\sigma^{M_{1,2}}}_{x_0}\text{-a.s.}
\end{aligned}$$

The second equality holds because the controlled process $\{X_n\}_{n=0}^\infty$ under the stationary strategy σ^s is Markov and time-homogeneous. Note also that $\mathsf{P}^{\sigma^{M_{1,2}}}_{x_0} \ll \mathsf{P}^{\sigma^s}_{x_0}$ by (18). Now, since $\sigma^{M_{1,2}}_n = \sigma^s$ for $n < \tau$,

$$\begin{aligned}
\int_{\mathbf{X}} f(x)\mathsf{M}^{\sigma^{M_{1,2}}}_{x_0}(dx \times \mathbf{A}) &= \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=1}^{\tau} f(X_{n-1})\right]\\
&\quad + \mathsf{E}^{\sigma^{M_{1,2}}}_{x_0}\left[\mathsf{E}^{\sigma^{M_{1,2}}}_{x_0}\left[\sum_{n=\tau+1}^{\infty} f(X_{n-1})\middle|\mathcal{F}_\tau\right]\right]\\
&= \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=1}^{\tau} f(X_{n-1})\right] + \mathsf{E}^{\sigma^{M_{1,2}}}_{x_0}[h(X_\tau)]\\
&= \mathsf{E}^{\sigma^s}_{x_0}\left[\sum_{n=1}^{\tau} f(X_{n-1})\right] + \mathsf{P}^{\sigma^{M_{1,2}}}_{x_0}(X_\tau \in \Gamma).
\end{aligned}$$

As the result, by the definition of the subset Γ,

$$\begin{aligned}&\int_{\mathbf{X}} f(x)\mathsf{M}_{x_0}^{\sigma^{M_1}}(dx\times\mathbf{A})\neq\int_{\mathbf{X}} f(x)\mathsf{M}_{x_0}^{\sigma^{M_2}}(dx\times\mathbf{A})\\ \Longrightarrow\quad &\mathsf{M}_{x_0}^{\sigma^{M_1}}(dx\times\mathbf{A})\neq\mathsf{M}_{x_0}^{\sigma^{M_2}}(dx\times\mathbf{A})\quad\text{on}\ \ \mathcal{B}(\mathbf{X}).\end{aligned}$$

Hence, the measure $\mathsf{M}_{x_0}^{\sigma^s}$ is not extreme in $\mathcal{D}^f$.

The further steps in fact repeat the arguments in the proof of Theorem 10 of [31], but we provide the details for completeness.

Step 2. We will show that, if $\mathsf{M}_{x_0}^{\sigma^s}$ is an extreme point in $\mathcal{D}^f$, then, for each $\Gamma^A\in\mathcal{B}(\mathbf{A})$, $\Gamma^X\in\mathcal{B}(\mathbf{X})$, $\alpha\in(0,1)$, in case $\mathsf{M}_{x_0}^{\sigma^s}(\Gamma^X\times\mathbf{A})>0$, there is $x\in\Gamma^X$ such that either $\sigma^s(\Gamma^A|x)<\alpha$ or $\sigma^s(\Gamma^A|x)>1-\alpha$.

This statement is trivial for $\alpha>1/2$: if $\sigma^s(\Gamma^A|x)\geq\alpha>1/2$, then $1-\sigma^s(\Gamma^A|x)<1/2<\alpha$. Thus, below we assume that $\alpha\in(0,1/2]$.

The proof is by contradiction. Namely, suppose there exist $\hat\Gamma^A\in\mathcal{B}(\mathbf{A})$, $\hat\Gamma^X\in\mathcal{B}(\mathbf{X})$ and $\alpha\in(0,1/2]$ such that $\mathsf{M}_{x_0}^{\sigma^s}(\hat\Gamma^X\times\mathbf{A})>0$ and $\sigma^s(\hat\Gamma^A|x)\in[\alpha,1-\alpha]$ for all $x\in\hat\Gamma^X$. Consider the following stochastic kernels:

$$\begin{aligned}\sigma_1^s(\Gamma^A|x)&=\begin{cases}\sigma^s(\Gamma^A|x), & \text{if } x\notin\hat\Gamma^X;\\ \frac{\sigma^s(\Gamma^A\cap(\hat\Gamma^A)^c|x)}{\sigma^s((\hat\Gamma^A)^c|x)}, & \text{if } x\in\hat\Gamma^X;\end{cases}\\ \sigma_2^s(\Gamma^A|x)&=\begin{cases}\sigma^s(\Gamma^A|x), & \text{if } x\notin\hat\Gamma^X;\\ \frac{\sigma^s(\Gamma^A\cap\hat\Gamma^A|x)}{1-\alpha}+\sigma^s(\Gamma^A\cap(\hat\Gamma^A)^c|x)\frac{\sigma^s((\hat\Gamma^A)^c|x)-\alpha}{(1-\alpha)\sigma^s((\hat\Gamma^A)^c|x)}, & \\ & \text{if } x\in\hat\Gamma^X;\end{cases}\\ &\Gamma^A\in\mathcal{B}(\mathbf{A}),\end{aligned}$$

which are well defined because $\sigma^s((\hat\Gamma^A)^c|x)\geq\alpha>0$.

Clearly, $\alpha\sigma_1^s(\Gamma^A|x)+(1-\alpha)\sigma_2^s(\Gamma^A|x)=\sigma^s(\Gamma^A|x)$ for all $\Gamma^A\in\mathcal{B}(\mathbf{A})$ and $x\in\mathbf{X}$; $\mathsf{M}_{x_0}^{\sigma^s}(\hat\Gamma^X\times\mathbf{A})>0$, and, for all $x\in\hat\Gamma^X$,

$$\sigma_2^s(\hat\Gamma^A|x)-\sigma_1^s(\hat\Gamma^A|x)=\frac{\sigma^s(\hat\Gamma^A|x)}{1-\alpha}-0\geq\frac{\alpha}{1-\alpha}>0.$$

Therefore, the measure $\mathsf{M}_{x_0}^{\sigma^s}$ is not extreme in $\mathcal{D}^f$ according to the statement in Step 1.

Step 3. We will show that, if $\mathsf{M}_{x_0}^{\sigma^s}$ is an extreme point in $\mathcal{D}^f$, then, for each $\Gamma^A\in\mathcal{B}(\mathbf{A})$,

$$\sigma^s(\Gamma^A|x)\in\{0,1\}\text{ for }\mathsf{M}_{x_0}^{\sigma^s}(dx\times\mathbf{A})\text{-almost all }x\in\mathbf{X}.$$

Let $\Gamma^A\in\mathcal{B}(\mathbf{A})$ be arbitrarily fixed and introduce the sets

$$\Gamma_{\Gamma^A}^X(i):=\left\{x\in\mathbf{X}:\ \sigma^s(\Gamma^A|x)\in\left[\left(\frac{1}{2}\right)^i,1-\left(\frac{1}{2}\right)^i\right]\right\}\in\mathcal{B}(\mathbf{X}),\ i=1,2,\dots.$$

For eah $i=1,2,\dots$, $\mathsf{M}_{x_0}^{\sigma^s}(\Gamma_{\Gamma^A}^X(i)\times\mathbf{A})=0$ because, otherwise, for $\Gamma^A,\Gamma^X:=\Gamma_{\Gamma^A}^X(i)$, and $\alpha:=(\frac{1}{2})^i$, we would have $\mathsf{M}_{x_0}^{\sigma^s}(\Gamma^X\times\mathbf{A})>0$ and, for all $x\in\Gamma^X$, $\sigma^s(\Gamma^A|x)\in[\alpha,1-\alpha]$, which contradicts the statement proved at Step 2.

Note that $\Gamma_{\Gamma^A}^X(i)\subset\Gamma_{\Gamma^A}^X(i+1)$ for all $i=1,2,\dots$. Now, for

$$\Gamma_{\Gamma^A}^X:=\bigcup_{i=1}^{\infty}\Gamma_{\Gamma^A}^X(i)=\{x\in\mathbf{X}:\ \sigma^s(\Gamma^A|x)\in(0,1)\},$$

we have $\mathsf{M}^{\sigma^s}_{x_0}(\Gamma^X_{\Gamma^A}\times\mathbf{A}) = \lim_{i\to\infty}\mathsf{M}^{\sigma^s}_{x_0}(\Gamma^X_{\Gamma^A}(i)\times\mathbf{A}) = 0$, and the desired statement is proved.

Step 4. Finally, we proceed to construct the measurable mapping $\varphi:\ \mathbf{X}\to\mathbf{A}$ such that

$$\sigma^s(da|x) = \delta_{\varphi(x)}(da) \text{ for } \mathsf{M}^{\sigma^s}_{x_0}(dx\times\mathbf{A})\text{-almost all } x\in\mathbf{X}.$$

As a separable metrizable space, $\mathbf{A}$ has a totally bounded metrization κ [3, Corollary 7.6.1]:

$$\forall\varepsilon>0\ \ \exists\{a_1,a_2,\ldots,a_n\}\subset\mathbf{A}:\ \ \mathbf{A}=\bigcup_{i=1}^n O(a_i,\varepsilon),$$

where $O(a_i,\varepsilon):=\{a\in\mathbf{A}:\ \kappa(a,a_i)<\varepsilon\}$.

Let $\varepsilon_k:=\frac{1}{2^k}$ $(k=1,2,\ldots)$ and let $\{a^k_1,a^k_2,\ldots,a^k_{n_k}\}$ be the corresponding ε_k-net in $\mathbf{A}$. Denote

$$S^k_i := \{x\in\mathbf{X}:\ \sigma^s(O(a^k_i,\varepsilon_k)|x)\notin\{0,1\}\}$$

and $S:=\bigcup_{k,i}S^k_i$. These sets are obviously measurable, and, for all $k=1,2,\ldots$, $i=1,2,\ldots,n_k$, $\mathsf{M}^{\sigma^s}_{x_0}(S^k_i\times\mathbf{A})=0$ according to the statement proved on Step 3; $\mathsf{M}^{\sigma^s}_{x_0}(S\times\mathbf{A})=0$, too.

We are going to construct the desired mapping φ on $\mathbf{X}\setminus S$ and then put $\varphi(x)\equiv\hat a$ for all $x\in S$, for an arbitrarily fixed $\hat a\in\mathbf{A}$.

Let $x\in\mathbf{X}\setminus S$ be fixed. Since $\mathbf{A}=\bigcup_{i=1}^{n_1}O(a^1_i,\varepsilon_1)$ and, for each $i\in\{1,2,\ldots,n_1\}$, $\sigma^s(O(a^1_i,\varepsilon_1)|x)\in\{0,1\}$, there is the minimal index $i_1\in\{1,2,\ldots,n_1\}$ such that $\sigma^s(O(a^1_{i_1},\varepsilon_1)|x)=1$. We denote $\bar O^1:=O(a^1_{i_1},\varepsilon_1)$. Suppose that we have constructed a set $\bar O^k\in\mathcal{B}(\mathbf{A})$ for $k=1,2,\ldots$ such that $\sigma^s(\bar O^k|x)=1$. Then we put

$$\bar O^{k+1}:=\bar O^k\cap\hat O^{k+1},$$

where $\hat O^{k+1}=O(a^{k+1}_{i_{k+1}},\varepsilon_{k+1})$ is the first one among the neighbourhoods $\{O(a^{k+1}_i,\varepsilon_{k+1})\}_{i=1}^{n_{k+1}}$ on which $\sigma^s(\cdot|x)$ takes the value 1; thus $\sigma^s(\hat O^{k+1}|x)=1$. Note, $\sigma^s(\bar O^{k+1}|x)=1$ because

$$\begin{aligned}1 &= \sigma^s(\bar O^k\cup\hat O^{k+1}|x)=\sigma^s(\bar O^k|x)+\sigma^s(\hat O^{k+1}|x)-\sigma^s(\bar O^k\cap\hat O^{k+1}|x)\\ &= 2-\sigma^s(\bar O^{k+1}|x).\end{aligned}$$

For the sequence $\{\bar O^k\}_{k=1}^\infty$, we have the following assertions.

- $\sigma^s(\bar O^k|x)=1$ for all $k=1,2,\ldots$ and $\bar O^1\supseteq\bar O^2\supseteq\ldots$. Thus $\sigma^s\left(\bigcap_{k=1}^\infty\bar O^k|x\right)=1$, and hence $\bigcap_{k=1}^\infty\bar O^k\neq\emptyset$.
- $\bigcap_{k=1}^\infty\bar O^k=\{b\}$ is a singleton because, if $b_1,b_2\in\bigcap_{k=1}^\infty\bar O^k$, then, for each $k\geq1$, $b_1,b_2\in O(a^k_{i_k},\varepsilon_k)$ for some $i_k\in\{1,2,\ldots,n_k\}$ leading to the inequalities

$$\kappa(b_1,b_2)\leq\kappa(b_1,a^k_{i_k})+\kappa(a^k_{i_k},b_2)\leq2\varepsilon_k.$$

As the result, $\kappa(b_1,b_2)\leq\lim_{k\to\infty}2\varepsilon_k=0$.

We put $\varphi(x):=b$ for that preliminarily fixed $x\in\mathbf{X}\setminus S$ and for $b\in\mathbf{A}$ such that $\bigcap_{k=1}^\infty\bar O^k=\{b\}$. As was shown above, $\sigma^s(\{\varphi(x)\}|x)=\sigma^s(\bigcap_{k=1}^\infty\bar O^k|x)=1$; so $\sigma^s(da|x)=\delta_{\varphi(x)}(da)$ for all $x\in\mathbf{X}\setminus S$, that is, for $\mathsf{M}^{\sigma^s}_{x_0}(dx\times\mathbf{A})$-almost all $x\in\mathbf{X}$ because $\mathsf{M}^{\sigma^s}_{x_0}(S\times\mathbf{A})=0$. The mapping $\varphi:\ \mathbf{X}\to\mathbf{A}$ is measurable because, for all $\Gamma^A\in\mathcal{B}(\mathbf{A})$,

$$\{x\in\mathbf{A}:\ \varphi(x)\in\Gamma^A\}=\left\{\begin{array}{ll}\{x:\ \sigma^s(\Gamma^A|x)=1\}\setminus S, & \text{if } \hat a\notin\Gamma^A;\\ \{x:\ \sigma^s(\Gamma^A|x)=1\}\cup S, & \text{if } \hat a\in\Gamma^A.\end{array}\right.$$

(Recall that the stochastic kernel σ^s is measurable and $S \in \mathcal{B}(\mathbf{X})$.)

As the result,

$$\mathsf{M}(dx \times da) = \mathsf{M}^{\sigma^s}_{x_0}(dx \times da) = \sigma^s(da|x)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A}) = \delta_{\varphi(x)}(da)\mathsf{M}^{\sigma^s}_{x_0}(dx \times \mathbf{A})$$

on $\mathcal{B}(\mathbf{X} \times \mathbf{A})$, and $\mathsf{M} = \mathsf{M}^{\varphi}_{x_0}$ according to Lemma 2 because the deterministic stationary strategy φ is induced by $\mathsf{M} \in \mathcal{D}^f$. □

Before proving Theorem 2, we present several statements on mathematical programs.

Suppose $\mathcal{X}$ is a convex compact space and $\hat{\mathcal{C}}$ is the space of $(-\infty, +\infty]$-valued bounded from below lower semicontinuous affine functions on $\mathcal{X}$. Let $R_0(\cdot), R_1(\cdot), \dots, R_J(\cdot) \in \hat{\mathcal{C}}$ and consider the following constrained problem

$$\text{Minimize over } x \in \mathcal{X}\text{: } R_0(x) \text{ subject to } R_j(x) \le d_j,\ j = 1, 2, \dots, J, \tag{20}$$

where $d_j \in \mathbb{R}$ are fixed constants and $J \ge 1$. Here by a convex space we mean a convex subset of a cone. This definition does not involve a linear space to embed the given space in. The terms of affine functions and extreme points are understood wrt convex spaces, or say convex sets in a cone. See the definitions in [37], where further relevant literature can be found. For all our applications here, it is sufficient to remember that the space of occupation measures is a convex subset of the cone of $[0, \infty]$-valued measures. When some occupation measures can take infinite values, it is difficult to embed the space of occupation measures into a convex subset of any linear space, given that we use the usual notions of addition and scalar multiplication for measures.

Proposition 6 *Consider problem (20) as was described in the above paragraph. Suppose problem (20) is non-degenerate, i.e., there is at least one point $\hat{x} \in \mathcal{X}$ satisfying all the inequalities in (20). Assume also that $\hat{\mathcal{C}}$ separates points in $\mathcal{X}$. Then there exists a solution to problem (20) in the form $\sum_{k=1}^{J+1} \alpha_k x_k$, where $\alpha_k \in [0, 1]$, $\sum_{k=1}^{J+1} \alpha_k = 1$, and x_k is extreme in $\mathcal{X}$ for each $k = 1, 2, \dots, J+1$.*

Proof. See [37, Theorem 2.1]. □

If $\mathbb{E}$ is a nonempty convex subset of $\mathbb{R}^n$ and $u \in \mathbb{E}$, then $G(u)$ denotes the minimal face of $\mathbb{E}$ containing the point u. A point $u \in \mathbb{E}$ is called Pareto optimal if, for each $v \in \mathbb{E}$, the componentwise inequality $v \le u$ implies that $v = u$. The collection of all Pareto optimal points is denoted by $Par(\mathbb{E})$. The following result taken from [19] reveals the structure of $G(u)$.

Proposition 7 *Suppose $\mathbb{E}$ is a fixed nonempty convex subset of $\mathbb{R}^n$, and $u \in Par(\mathbb{E})$. Then the following assertions are valid.*

(a) $G(u) \subseteq Par(\mathbb{E})$.

(b) For some $1 \le k \le n$, there exist hyperplanes

$$\mathbb{H}^i = \{x \in \mathbb{R}^n :\ \langle x, b^i \rangle = \beta^i\},\ i = 1, 2, \dots, k$$

with the following properties:

(i) $b^i \ge 0$ *for* $i = 1, 2, \dots, k-1$ *and* $b^k > 0$. *Here all the inequalities are componentwise.*

(ii) $\mathbb{H}^1$ *is supporting to* $\mathbb{E}^0 := \mathbb{E}$ *at* u; *for* $i = 1, 2, \dots, k-1$, $\mathbb{E}^i := \mathbb{E}^{i-1} \cap \mathbb{H}^i$ *and* $\mathbb{H}^{i+1}$ *is supporting to* $\mathbb{E}^i$ *at* u;

(iii) $G(u) = \mathbb{E}^k := \mathbb{E}^{k-1} \cap \mathbb{H}^k$.

Proof. See Lemmas 3.1 and 3.2 of [19]. □

Now we are ready to present the proof of Theorem 2.

Proof of Theorem 2. The idea is to make use of Proposition 6 and Theorem 1.

In the space $\mathcal{D}$ of occupation measures, we fix the topology ρ as in Definition 5. According to Corollary 1, there is a solution $\mathsf{M}^* \in \mathcal{D}$ to problem (5), equivalent to (4), which has the form of problem (20):

- the space $\mathcal{X} = \mathcal{D}$ is convex compact due to Proposition 3 and Lemma 1(a);
- the mappings $R_j(\cdot)$, $j = 0, 1, \ldots, J$, are non-negative, affine and lower semicontinuous by Lemma 1(b).

According to Step 1 in the proof of [37, Theorem 2.1], one can accept that the point

$$\vec{R}^* := (R_0(\mathsf{M}^*), R_1(\mathsf{M}^*), \ldots, R_J(\mathsf{M}^*)) \in \mathbb{R}^{J+1}$$

belongs to $Par(\mathbb{O} \cap \mathbb{R}^{J+1})$, where

$$\mathbb{O} := \{\vec{R}(\mathsf{M}) = (R_0(\mathsf{M}), R_1(\mathsf{M}), \ldots, R_J(\mathsf{M})),\ \mathsf{M} \in \mathcal{D}\}$$

is the (convex) objective space.

We denote $\mathbb{E}^0 = \mathbb{E} := \mathbb{O} \cap \mathbb{R}^{J+1}$ and emphasize that

$$\begin{aligned} \vec{R}(\mathsf{M}) \in \mathbb{E}^0 \Longleftrightarrow \mathsf{M} \in \mathbb{F}^0 &:= \{\mathsf{M} \in \mathcal{D} :\ \vec{R}(\mathsf{M}) \in \mathbb{R}^{J+1}\} \\ &= \{\mathsf{M} \in \mathcal{D}^f :\ \vec{R}(\mathsf{M}) \in \mathbb{R}^{J+1}\}. \end{aligned}$$

The equality holds because, due to the imposed conditions, the component $R_{\tilde{j}}(\mathsf{M})$ cannot be finite if $\mathsf{M}(\mathbf{X} \times \mathbf{A}) = +\infty$. The set $\mathbb{F}^0$, the full pre-image of $\mathbb{E}^0$ wrt the mapping $\vec{R}(\cdot) :\ \mathcal{D} \to (\mathbb{R} \cup \{\infty\})^{J+1}$, is a face of $\mathcal{D}^f$: recall that the mapping $\vec{R}(\cdot)$ is affine.

Consider the sets $\mathbb{E}^i$, $i = 0, 1, \ldots, k \le J+1$ and the hyperplanes $\mathbb{H}^i = \{x \in \mathbb{R}^{J+1} :\ \langle x, b^i \rangle = \beta^i\}$, $i = 1, 2, \ldots, k$, as in Proposition 7 applied to $n = J+1$, $\mathbb{E}$ and $u = \vec{R}^*$. Let $\mathbb{F}^i$ be the full pre-image of $\mathbb{E}^i$ wrt the mapping $\vec{R}(\cdot)$. Note that $\mathsf{M}^* \in \mathbb{F}^i$ for all $i = 0, 1, \ldots, k$ because $\vec{R}^* = \vec{R}(\mathsf{M}^*) \in \mathbb{E}^i$ for all $i = 0, 1, \ldots, k$.

Firstly, let us prove that, for each $i = 0, 1, \ldots, k$, $\mathbb{F}^i$, is a (nonempty) face of $\mathcal{D}^f$. Roughly speaking, $\mathbb{F}^{i+1}$ is a face of $\mathbb{F}^i$ because $\mathbb{E}^{i+1}$ is the exposed face of $\mathbb{E}^i$. The statement to be proved is valid for $i = 0$. Suppose it holds for some $i = 0, 1, \ldots, k-1$. Then

$$\begin{aligned} \mathbb{F}^{i+1} &= \mathbb{F}^i \cap \{\mathsf{M} \in \mathcal{D}^f :\ \vec{R}(\mathsf{M}) \in \mathbb{R}^{J+1},\ \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle = \beta^{i+1}\} \\ &= \{\mathsf{M} \in \mathbb{F}^i :\ \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle = \beta^{i+1}\} \end{aligned}$$

because $\mathbb{E}^{i+1} = \mathbb{E}^i \cap \mathbb{H}^{i+1}$. For each $\mathsf{M} \in \mathbb{F}^i$, $\vec{R}(\mathsf{M}) \in \mathbb{E}^i$, so $\langle \vec{R}(\mathsf{M}), b^{i+1} \rangle \ge \beta^{i+1}$ because the hyperplane $\mathbb{H}^{i+1}$ supports $\mathbb{E}^i$ at $\vec{R}^*$. Therefore, if $\mathsf{M} = \alpha \mathsf{M}_1 + (1-\alpha)\mathsf{M}_2 \in \mathbb{F}^{i+1}$ for $\alpha \in (0,1)$ and $\mathsf{M}_1, \mathsf{M}_2 \in F^i$, then

$$\langle \vec{R}(\mathsf{M}_{1,2}), b^{i+1} \rangle = \beta^{i+1}, \text{ and hence } \mathsf{M}_1, \mathsf{M}_2 \in \mathbb{F}^{i+1}.$$

Thus, $\mathbb{F}^{i+1}$ is a face of $\mathbb{F}^i$ and, consequently, a face of $\mathcal{D}^f$ because $\mathbb{F}^i$ is a face of $\mathcal{D}^f$ by the induction supposition.

We have proved that $\mathbb{F}^k$ is a nonempty face of $\mathcal{D}^f$ and $\mathsf{M}^* \in \mathbb{F}^k$. In fact, $\mathbb{F}^k$ is the full pre-image of $G(\vec{R}^*) = \mathbb{E}^k$, the minimal face of $\mathbb{E} = \mathbb{O} \cap \mathbb{R}^{J+1}$ containing $\vec{R}^*$: see Proposition 7.

Secondly, let us show that the face $\mathbb{F}^k$ is closed and hence compact. Since the hyperplanes $\mathbb{H}^{i+1} = \{x \in \mathbb{R}^{J+1} : \langle x, b^{i+1} \rangle = \beta^{i+1}\}$ are supporting $\mathbb{E}^i$ at $u = \vec{R}^*$ ($i = 0, 1, \dots, k-1$), one can also write

$$\mathbb{F}^{i+1} \quad = \quad \{\mathsf{M} \in \mathbb{F}^i : \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle \le \beta^{i+1}\} = \mathbb{F}^i \cap \{\mathsf{M} \in \mathcal{D} : \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle \le \beta^{i+1}\},$$

so that

$$\begin{aligned} \mathbb{F}^k &= \mathbb{F}^0 \cap \left(\bigcap_{i=0}^{k-1} \{\mathsf{M} \in \mathcal{D} : \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle \le \beta^{i+1}\} \right) \\ &= \bar{\mathbb{F}}^0 \cap \left(\bigcap_{i=0}^{k-2} \{\mathsf{M} \in \mathcal{D} : \langle \vec{R}(\mathsf{M}), b^{i+1} \rangle \le \beta^{i+1}\} \right), \end{aligned} \tag{21}$$

where

$$\bar{\mathbb{F}}^0 \quad := \quad \mathbb{F}^0 \cap \{\mathsf{M} \in \mathcal{D} : \langle \vec{R}(\mathsf{M}), b^k \rangle \le \beta^k\} = \{\mathsf{M} \in \mathcal{D} : \langle \vec{R}(\mathsf{M}), b^k \rangle \le \beta^k\}.$$

The second equality holds because $\vec{R}(\mathsf{M}) \ge 0$ and $b^k > 0$: if $\langle \vec{R}(\mathsf{M}), b^k \rangle \le \beta^k$, then $\vec{R}(\mathsf{M}) \in \mathbb{R}^{J+1}$; so that $\mathsf{M} \in \mathbb{F}^0$. Note, $\bar{\mathbb{F}}^0$ is not necessarily a face of $\mathcal{D}$. The space $(\mathcal{D}, \rho)$ is compact and all the mappings $\langle \vec{R}(\cdot), b^{i+1} \rangle : \mathcal{D} \to [0, \infty]$, $i = 0, 1, \dots, k-1$, are lower semicontinuous by Lemma 1. Therefore, the set $\bar{\mathbb{F}}^0$ is closed, and the face $\mathbb{F}^k$ of $\mathcal{D}^f$ is closed by (21), hence, compact, as the closed subset of the compact $\mathcal{D}$.

The space $\hat{\mathcal{C}}$ of $(-\infty, +\infty]$-valued bounded from below lower semicontinuous affine functions on $\mathcal{D}^f$ separates points in $\mathcal{D}^f$. Indeed, if $\mathsf{M}_1 \ne \mathsf{M}_2$ are two measures from $\mathcal{D}^f$, then

$$C(\mathsf{M}_1) := \int_{\mathbf{X} \times \mathbf{A}} c(x, a) \mathsf{M}_1(dx \times da) \ne \int_{\mathbf{X} \times \mathbf{A}} c(x, a) \mathsf{M}_2(dx \times da) =: C(\mathsf{M}_2)$$

for some non-negative bounded continuous function $c(\cdot, \cdot) : \mathbf{X} \times \mathbf{A} \to \mathbb{R}$. (See Lemma 2.3 of [41], Theorem 5.9 of [30] and Proposition 7.18 of [3].) The mapping $C(\cdot)$ is non-negative, lower semicontinuous and affine by Lemma 1(b), so that the desired assertion follows.

Since the compact face $\mathbb{F}^k \subseteq \mathcal{D}^f$ contains M^*, one can consider problem (5), on $\mathbb{F}^k$, not on $\mathcal{D}$. This problem

$$\begin{aligned} R_0(\mathsf{M}) &:= \int_{\mathbf{X} \times \mathbf{A}} r_0(x, a) \mathsf{M}(dx \times da) \to \min_{\mathsf{M} \in \mathbb{F}^k} \qquad (22) \\ \text{s.t. } R_j(\mathsf{M}) &:= \int_{\mathbf{X} \times \mathbf{A}} r_j(x, a) \mathsf{M}(dx \times da) \le d_j, \quad j = 1, 2, \dots, J, \end{aligned}$$

satisfies all the conditions of Proposition 6:

- the space $\mathbb{F}^k$ is convex compact;
- the mappings $R_j(\cdot) : \mathbb{F}^k \to [0, \infty]$, $j = 0, 1, \dots, J$, are non-negative, lower semicontinuous by Lemma 1(b), and affine;
- problem (22) is non-degenerate, as the measure $\mathsf{M}^* \in \mathbb{F}^k$ satisfies all the constraints;

- The space $\hat{\mathcal{C}}$ separates points in $\mathbb{F}^k \subseteq \mathcal{D}^f$.

According to Proposition 6, there exists a solution to problem (22) (hence, to problem (5)) in the form $\sum_{l=1}^{J+1} \alpha_l \mathsf{M}_l$, where $\alpha_l \in [0,1]$, $\sum_{l=1}^{J+1} \alpha_l = 1$, and M_l is extreme in $\mathbb{F}^k$ for each $l = 1, 2, \ldots, J+1$. Since $\mathbb{F}^k$ is a face of $\mathcal{D}^f$, M_l is extreme also in $\mathcal{D}^f$ for each $l = 1, 2, \ldots, J+1$ and equals $\mathsf{M}_{x_0}^{\varphi_l}$ for some deterministic stationary strategy φ_l in accordance with Theorem 1. Therefore, the mixture $\hat{\mathsf{M}} := \sum_{l=1}^{J+1} \alpha_l \mathsf{M}_{x_0}^{\varphi_l}$ solves problem (5), and the corresponding strategic measure $\hat{\mathsf{P}} := \sum_{l=1}^{J+1} \alpha_l \mathsf{P}_{x_0}^{\varphi_l}$ solves problem (4). $\square$

Acknowledgment

We would like to thank the referees for their valuable comments, which improved the presentation of this paper.

References

[1] Altman, E. (1999). *Constrained Markov Decision Processes.* Chapman and Hall/CRC, Boca Raton.

[2] Bertsekas, D. (2018). Proper policies in infinite-state stochastic shortest path problems. *IEEE Trans. Automat. Contr.* **63**, 3787–3792.

[3] Bertsekas, D. and Shreve, S. (1978). *Stochastic Optimal Control.* Academic Press, New York.

[4] Borkar, V. (1988). A convex analytic approach to Markov decision processes. *Probab. Th. Rel. Fields* **78**, 583–602.

[5] Borkar, V. (1991). *Topics in Controlled Markov Chains.* Longman Scientific and Technical, England.

[6] Borkar. V. (2002). Convex analytic methods in Markov decision processes. In *Handbook of Markov Decision Processes*, Feinberg, E.A. and Shwartz, A. (eds). Srpinger, New York.

[7] D'Epenoux, F. (1960). Sur un probleme de production et de stockage dans l'Aetoire. *Rev. Franc. Rech. Operationelle* **14**, 3–16.

[8] Dufour, F. and Genadot, A. (2020). A convex programming approach for discrete-time Markov decision processes under the expected total reward criterion. *SIAM J. Control Optim.* **58**, 2535–2566.

[9] Dufour, F., Horiguchi, M. and Piunovskiy, A. (2012). The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach. *Adv. Appl. Probab.* **44**, 774–793.

[10] Dufour, F. and Piunovskiy, A. (2010). Multiobjective stopping problem for discrete-time Markov processes: convex analytic approach. *J. Appl. Probab.* **47**, 947–966.

[11] Dufour, F. and Piunovskiy, A. (2013). The expected total cost criterion for Markov decision processes under constraints. *Adv. Appl. Probab.* **45**, 837–859.

[12] Dufour, F. and Prieto-Rumeau, T. (2016). Conditions for the solvability of the linear programming formulation for constrained discounted Markov decision processes. *App. Math. Optim.* **74**, 27–51.

[13] Dunford, N. and Schwartz, J.T. (1958). *Linear Operators. Part I: General Theory.* John Wiley & Sons, Inc., New York.

[14] Dynkin, E.B. and Yushkevich, A.A. (1979). *Controlled Markov Processes.* Springer, New York.

[15] Fainberg, E.A. (1982). Controlled Markov processes with arbitrary numerical criteria. *Theory Probab. Appl.* **27**, 486–503.

[16] Feinberg, E.A., Jaskiewicz, A. and Nowak, A.S. (2020). Constrained discounted Markov decision processes with Borel state spaces. *Automatica,* **111**, N.108582.

[17] Feinberg, E.A. and Piunovskiy, A. (2019). Sufficiency of deterministic policies for atomless discounted and uniformly absorbing MDPs with multiple criteria. *SIAM J. Control Optim.* **57**, 163–191.

[18] Feinberg, E.A. and Rothblum, U. (2012). Splitting randomized stationary policies in total-reward Markov decision processes. *Math. Oper. Res.* **37**, 129–153.

[19] Feinberg, E.A. and Shwartz, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.* **21**, 922–944.

[20] Feinberg, E.A. and Sonin, I.M. (1996) Notes on equivalent stationary policies in Markov decision processes with total rewards. *Math. Meth. Oper. Res.* **44**, 205–221.

[21] Gonzalez-Hernandez, J. and Hernández-Lerma, O. (2000). Constrained Markov control processes in Borel spaces: the discounted case. *Math. Methods Oper. Res.* **52**, 271–285.

[22] Gonzalez-Hernandez, J. and Hernández-Lerma, O. (2005). Extreme points of the sets of randomized strategies in constrained optimization and control problems. *SIAM J. on Optim.* **15**, 1085–1104.

[23] Gonzalez-Hernandez, J. and Villarreal, C.E. (2011). Optimal policies for constrained average-cost Markov decision processes. *TOP* **19**, 107–120.

[24] Guo, X.P., Huang, Y.H. and Zhang, Y. (2017). Constrained continuous-time Markov decision processes on the finite horizon. *Appl. Math. Optim.* **75**, 317–341.

[25] Guo, X.P. and Zhang, Y. (2016). Optimality of mixed policies for average continuous-time Markov decision processes with constraints. *Math. Oper. Res.* **41**, 1276–1296.

[26] Guo, X.P. and Zhang, Y. (2017). Constrained total undiscounted continuous-time Markov decision processes. *Bernoulli* **23**, 1694–1736.

[27] Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[28] Hernández-Lerma, O. and Lasserre, J. (1999). *Further Topics in Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[29] Kallenberg, L.C.M. (1983). *Linear Programming and Finite Markovian Control Problems.* Math. Centre Tracts. **148**, Amsterdam.

[30] Parthasarathy, K.R. (2005). *Probability Measures on Metric Spaces.* AMS Chelsea Publishing, Rhode Island.

[31] Piunovskiy, A. (1997). *Optimal Control of Random Sequences in Problems with Constraints.* Kluwer, Dordrecht.

[32] Piunovskiy, A. (1998). Controlled random sequences: methods of convex analysis and problems with functional constraints. *Russ. Math. Surveys* **53**, 1233–1293.

[33] Piunovskiy, A. and Zhang, Y. (2011). Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.* **49**, 2032–2061.

[34] Piunovskiy, A. and Zhang, Y. (2020). *Continuous-Time Markov Decision Processes.* Springer, Cham.

[35] Piunovskiy, A. and Zhang, Y. (2020). On reducing a constrained gradual-impulsive control problem for a jump Markov model to a model with gradual control only. *SIAM J. Control Optim.* **58**, 192–214.

[36] Piunovskiy, A. and Zhang, Y. (2022). Gradual-impulsive control for continuous-time Markov decision processes with total undiscounted costs and constraints: linear programming approach via a reduction method. *SIAM J. Control Optim.* **60**, 1892–1917.

[37] Piunovskiy, A. and Zhang, Y. (2023). On the structure of optimal solutions in a mathematical programming problem in a convex space. *Oper. Res. Lett.* **51**, 488–493.

[38] Piunovskiy, A. and Zhang, Y. (2023). On the continuity of the projection mapping from strategic measures to occupation measures in absorbing Markov decision processes. Preprint. Available at arXiv:2311.14043

[39] Schäl, M. (1975). On dynamic programming: compactness of the space of policies. *Stoch. Process. Their Appl.* **3**, 345–364.

[40] Schäl, M. (1975). Conditions for optimality in dynamic programming and for the limit of n-stage policies to be optimal. *Z. Wahrscheinlichkeitstheorie verw. Gebite.* **32**, 179–196.

[41] Varadarajan, V. (1958). Weak convergence of measures on separable metric spaces. *Sankhyā* **19**, 15–22.

[42] Zhang, Y. (2013). Convex analytic approach to constrained discounted Markov decision processes with non-constant discount factors. *TOP* **21**, 378–408.