INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Robustness of convex optimization with application to controled Markov chains

Mabel M. TIDBALL and Eitan ALTMAN

N° 2933

July 1996

THÈME 4

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Robustness of convex optimization with application to controled Markov chains

Mabel M. TIDBALL [*] and Eitan ALTMAN [**]

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet MIAOU

Rapport de recherche n ° 2933 — July 1996 — 22 pages

Abstract: We present two stability results in this paper. We first obtain sufficient conditions for the continuity of optimal values and solutions of convex programs in general vector spaces, as well as some types of robustness of some sub-optimal solutions. We then use these results in order to establish a new result in stochastic dynamic control of discrete event systems (known as constrained Markov Decision Processes): the convergence of the value and optimal policies of the problem with discounted costs, to the ones for the problem with expected average cost.

Key-words: Convex optimization. Sensivility analysis. Constrained control of Markov chains

(Résumé : tsvp)

[*] Department of Mathematics. University of Rosario. Pellegrini 250, 2000 Rosario. Argentina
[**] INRIA, Centre Sophia-Antipolis, 2004 Route des Lucioles, B.P.93, 06902 Sophia-Antipolis Cedex, France

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Robustesse de l'optimisation convexe avec applications au contrôle de chaînes de Markov

Résumé : Nous présentons dans ce papier deux résultats de stabilité. D'abord, nous trouvons des conditions suffisantes pour avoir la continuité de la valeur optimale et des solutions des programmes convexes dans des espaces vectoriels généraux, ainsi qu'un type de robustesse pour certaines solutions sub-optimales.

Ensuite, nous utilisons ces résultats pour établir un nouveau résultat en contrôle dynamique stochastique de systèmes à événements discrets (connu sous le nom de Processus de Décision Markoviens): la convergence de la valeur et des politiques optimales du problème avec coût pénalisé vers celles du problème avec coût moyen.

Mots-clé : Optimisation convexe. Analyse de sensibilité. Chaînes de Markov controlées avec contraintes

1 Introduction

We consider a sequence $\mathbf{MP}_n$, $n = 1, 2, ...$ of convex programs and a "limit" one, denoted by $\mathbf{MP}_\infty$, or simply by $\mathbf{MP}$. These are defined on some vector spaces, possibly infinite dimensional ones. $\mathbf{MP}$ is assumed to be feasible (it has at least one solution). However, for any given n, $\mathbf{MP}_n$ need not be feasible, and even if it is, it need not possess an optimal solution (i.e., it may only have ϵ-optimal solutions). We are interested in the following questions:

- (i) Do the values of $\mathbf{MP}_n$ converge to the value of $\mathbf{MP}$?
- (ii) Do optimal (or almost optimal) policies converge in some sense?
- (iii) Given an (almost) optimal policy for $\mathbf{MP}_n$, will it be an almost optimal policy for $\mathbf{MP}$, if n is sufficiently large?
- (iv) Conversely, given an optimal policy for $\mathbf{MP}$, will it be an almost optimal policy for $\mathbf{MP}_n$, for all n sufficiently large?

We provide in this paper sufficient conditions for having convergence in the sense of (i) and (ii) above. It turns out that the answers for (iii) and for (iv) is in general negative, unlike the unconstrained case. The reason is that an optimal policy for $\mathbf{MP}_n$ may be unfeasible for $\mathbf{MP}$, and vice versa. We shall, however, establish sufficient conditions for the following slightly weaker version of (iii) and (iv):

- (iii') Given an optimal policy for $\mathbf{MP}_n$, can we perturb it "slightly" so that it becomes almost optimal for $\mathbf{MP}$, if n is sufficiently large?
- (iv') Given an optimal policy for $\mathbf{MP}$, can we perturb it "slightly" so that it becomes almost optimal for $\mathbf{MP}_n$, for all n sufficiently large?

We apply our results to establish convergence properties in constrained Markov Decision Processes (MDPs). These type of control models have many applications in telecommunications and other fields [4, 16, 18, 21, 22, 24]. We present sufficient conditions for the convergence of the values and of optimal policies, as well as some robustness properties of sub-optimal policies. Related results were already obtined in [2], but had a strong restriction on the ergodic structure of the controlled model. It was required to have a single ergodic class under any stationary policy. This condition enabled us to restrict to stationary policies, for which some general theorem on approximation [1] could be used. In the present paper we make no assumption on the ergodic structure, thus allowing a multi-chain situation. For such ergodic structure, it is known that stationary policies need not be optimal (nor even ϵ-optimal) for the expected average cost criterion, and one has to use either Markov policies (see [19, 20]) or mixed-stationary policies (this term was introduced by Feinberg [15]; it refers to policies that are highly non-stationary). We use the latter approach to establish, with the help of the results from the first part of the paper, the convergence of the values and policies.

We briefly mention some related work on the continuity and sensitivity analysis of mathematical programs, and of control problems. Many papers and books studied similar problems in the case of finite dimensional state, e.g. [13, 17, 23]. Several special issues of scientific journals were devoted to these questions, as well as other related sensitivity, stabaility and parametric analysis: *Mathematical Programming* **21**, 1984, *Annals of Operations Research* **27**, 1990. Convergence results for constrained dynamic control problems were obtained in [1, 2, 3, 6, 7]. Conditions were obtained there for the convergence in the transition probabilities, in the horizon and in the immediate cost. These results were applied to adaptive control problems [6] and to problems of finite state approximations of constrained MDPs.

2 The model

Let $U \subset \mathbf{U}$, where $\mathbf{U}$ is a topological vector space, $V_n = (V_n^1, ..., V_n^K) \in \mathbb{R}^{\mathrm{K}}$, $n = 1, 2, ..., \infty$, $(V_\infty = V)$ and

$$C_n : U \to \mathbb{R} \qquad \mathrm{n} = 1, ..., \infty \qquad (\mathrm{C}_\infty = \mathrm{C})$$

$$D_n : U \to \mathbb{R}^{\mathrm{K}} \qquad \mathrm{n} = 1, ..., \infty \qquad (\mathrm{D}_\infty = \mathrm{D})$$

$$\Delta_n = \{u \in U : D_n(u) \leq V_n\} \qquad n = 1, 2..., \infty \qquad (\Delta_\infty = \Delta)$$

We define the values of the constrained problems:

$$R_n = inf_{u \in \Delta_n} C_n(u); \qquad R = inf_{u \in \Delta} C(u)$$

and we assume: there exists $M_1 > 0$ such that

$$\lim_{n \to \infty} D_n = D, \qquad \text{uniformly in} \quad \bigcup_{k \geq M_1} \Delta_k \cup \Delta \tag{1}$$

$$\lim_{n \to \infty} C_n = C, \qquad \text{uniformly in} \quad \bigcup_{k \geq M_1} \Delta_k \cup \Delta \tag{2}$$

$$U \quad \text{is a convex set} \tag{3}$$

$$D : U \to \mathbb{R}^{\mathrm{K}} \quad \text{is a convex function} \tag{4}$$

$$C : U \to \mathbb{R} \quad \text{is a convex function} \tag{5}$$

$$\exists M > 0 \quad \text{such that} \qquad -M \leq C(u) \qquad \forall u \in U \tag{6}$$

$$\exists v \in U, \exists \eta > 0 \quad \text{such that} \qquad D^k(v) \leq V^k - \eta, \ \forall k = 1, ..., K \tag{7}$$

$$V_n \to V, \qquad n \to \infty \tag{8}$$

We shall denote by $(\mathbf{H})$ the set of hypothesis (1) - (8).

We want to answer the following questions:
(i) Does $R_n \to R$ when $n \to \infty$?
(ii) Convergence of policies: Let $\pi : U \to \Delta$, fix some $\epsilon \geq 0$. Let ϵ_n be a sequence of positive real numbers such that $\overline{\lim}_{n\to\infty} \epsilon_n \leq \epsilon$. Assume that u_n^* is an ϵ_n-optimal policy for the nth optimal cost function R_n. Is $\pi(u_n^*)$ "almost" optimal for the limit optimal cost function R, for n large enough?
(iii) Robustness of the optimal policy: If u^* is ϵ-optimal for the limit optimal cost function, can we derive of it an "almost" optimal policy for the nth approximating optimal cost function, for all n large enough?
(iv) Let $\overline{u} \in U$ be some limit point of u_n^*, defined above. Is $\overline{u}$ ϵ-optimal for the limit optimal cost function?

Remark 2.1 *1- Hypothesis* (1) *and* (2) *can be relaxed, as shown in Section* 3.3.
2- In many applications U depends on n. We will deal with this case in Section 3.4.
3- The fact that U is a topological space is only used for establishing statement (iv).

For any vector $\mathcal{V} \in \mathbb{R}^{\mathrm{K}}$ and any constant $v \in \mathbb{R}$, we shall understand below $\mathcal{V} + v$ to mean the vector in $\mathbb{R}^{\mathrm{K}}$ obtained by adding the constant v to each of the components of $\mathcal{V}$. We shall say that $u \in U_n$ is ϵ-optimal for R_n if $u \in \Delta_n$ and $C_n(u) \leq R_n + \epsilon$.

3 Key Theorems for approximations

We begin by introducing some definitions in order to present the main results of this paper. Let δ be such that $\eta > \delta > 0$ (where η is defined in (7)). Define:

$$\epsilon_\delta(u) = \min_{\lambda\in[0,1]} \{\lambda D(v) + (1-\lambda)D(u) \leq V - \delta\} \tag{9}$$

where v is defined in (7). Define $\pi_\delta : U \to \Delta$ in the following way:

$$\pi_\delta(u) = \epsilon_\delta(u)v + (1 - \epsilon_\delta(u))u \tag{10}$$

Remark 3.1 *$\epsilon_\delta(u)$ is well defined because by (7) we have:*

$$\{\lambda : \lambda D(v) + (1-\lambda)D(u) \leq V - \delta\} \neq \emptyset$$

so $\pi_\delta(u)$ is well defined. Hence, $\pi_\delta(u) \in \Delta$, because by (3) we have $\pi_\delta(u) \in U$ and by (4) and the definition of $\epsilon_\delta(u)$, we have:

$$D(\pi_\delta(u)) = D(\epsilon_\delta(u)v + (1-\epsilon_\delta(u))u) \leq \epsilon_\delta(u)D(v) + (1-\epsilon_\delta(u))D(u) \leq V - \delta \tag{11}$$

In this section we shall prove the following approximation theorems. The first establishes the convergence of costs:

Theorem 3.1 *If* (**H**) *holds then $R_n \to R$ as $n \to \infty$*

The second one deals with the different types of policy convergence.

Theorem 3.2 *Let δ be sucht that $\eta > \delta > 0$. If (**H**) holds then:*
i) Let u_n be ϵ_n-optimal for R_n, $\limsup_n \epsilon_n \leq \epsilon$, then there exists $N(\epsilon,\delta)$ such that $\forall n \geq N(\epsilon,\delta)$ $\pi_\delta(u_n)$ is $O(\epsilon+\delta)$-optimal for R.
ii) Let u be ϵ-optimal for R then there exists $N(\epsilon,\delta)$ such that $\forall n \geq N(\epsilon,\delta)$, $\pi_\delta(u)$ is $O(\epsilon+\delta)$-optimal for R_n.
Moreover if there exists $M \geq 0$ such that C and D are lower semicontinuous in $\bigcup_{n\geq M} \Delta_n \cup \Delta$ then:
iii) Let u_n be ϵ_n-optimal for R_n such that $u_n \to u$ when $n \to \infty$ then u is $O(\epsilon)$-optimal for R.

In order to prove these theorems we are going to prove some auxiliary lemmas.

3.1 Auxiliary properties

Lemma 3.1 *Let δ be such that $\eta > \delta > 0$. If (1), (7) and (8) hold, then:*

$$\limsup_{n\to\infty} \left\{ \sup_{u\in\Delta_n} \epsilon_\delta(u) \right\} \leq \frac{\delta}{\eta}$$

Proof. By definition of Δ_n we have that $\forall u \in \Delta_n$, $D_n(u) \leq V_n$. By (1) and (8) we have:

$$\forall \hat{\epsilon} > 0 \quad \forall u \in \Delta_n \quad \exists N(\hat{\epsilon})/ \quad \forall n \geq N(\hat{\epsilon}) \quad D(u) \leq V + \hat{\epsilon}. \tag{12}$$

By (7) and (12) we have:

$$\lambda D(v) + (1-\lambda)D(u) \leq \lambda(V-\eta) + (1-\lambda)(V+\hat{\epsilon}) \quad \forall \hat{\epsilon} > 0 \quad \forall u \in \Delta_n \quad \forall \lambda \in [0,1] \tag{13}$$

But:

$$\lambda(V-\eta) + (1-\lambda)(V+\hat{\epsilon}) \leq V - \delta \Longleftrightarrow \lambda \geq \frac{\hat{\epsilon}+\delta}{\eta+\hat{\epsilon}} \quad \forall \hat{\epsilon} > 0 \quad \forall u \in \Delta_n \quad \forall \lambda \in [0,1] \tag{14}$$

By (13) and (14) we obtain:

$$\left[\frac{\hat{\epsilon}+\delta}{\eta+\hat{\epsilon}}, 1\right] \subseteq \{\lambda \in [0,1] : \lambda D(v) + (1-\lambda)D(u) \leq V - \delta\} \quad \forall \hat{\epsilon} > 0 \quad \forall u \in \Delta_n$$

and then, by definition of $\epsilon_\delta(u)$, we have:

$$\epsilon_\delta(u) \leq \frac{\hat{\epsilon}+\delta}{\eta+\hat{\epsilon}} \quad \forall u \in \Delta_n, \quad n \geq N(\hat{\epsilon}), \quad \forall \hat{\epsilon} > 0.$$

From this last inequality we deduce:

$$\sup_{u\in\Delta_n} \epsilon_\delta(u) \leq \frac{\hat{\epsilon}+\delta}{\eta+\hat{\epsilon}} \quad n \geq N(\hat{\epsilon}) \quad \forall \hat{\epsilon}$$

and this last inequality implies

$$\limsup_{n\to\infty}\left[\sup_{u\in\Delta_n}\epsilon_\delta(u)\right]\leq\frac{\delta}{\eta}$$

■

Remark 3.2 *Let δ be such that $\eta>\delta>0$. In the same way of Lemma 3.1 and if (7) holds we can prove that:*

$$\sup_{u\in\Delta}\epsilon_\delta(u)\leq\frac{\delta}{\eta}$$

for the proof it is only necessary to remark that (13) and (14) become:

$$\lambda D(v)+(1-\lambda)D(u)\leq\lambda(V-\eta)+(1-\lambda)V$$

$$\lambda(V-\eta)+(1-\lambda)V\leq V-\delta\Longleftrightarrow\lambda\geq\frac{\delta}{\eta}$$

Lemma 3.2 *Let δ be such that $\eta>\delta>0$. If (1), (5), (6), (7) and (8) hold, then:*

$$\limsup_{n\to\infty}\left[\sup_{u\in\Delta_n}\{C(\pi_\delta(u))-C(u)\}\right]\leq(C(v)+M)\frac{\delta}{\eta}$$

Proof. $\forall u\in\Delta_n$ by (5) and (10) we obtain:

$$C(\pi_\delta(u))-C(u)\leq\epsilon_\delta(u)C(v)+(1-\epsilon_\delta(u))C(u)-C(u)=\epsilon_\delta(u)[C(v)-C(u)]\qquad(15)$$

then by (15), (6) and Lemma 3.1 we have:

$$\limsup_{n\to\infty}\sup_{u\in\Delta_n}\epsilon_\delta(u)[C(v)-C(u)]\leq(C(v)+M)\limsup_{n\to\infty}\sup_{u\in\Delta_n}\epsilon_\delta(u)\leq(C(v)+M)\frac{\delta}{\eta}$$

■

Remark 3.3 *Let δ be such that $\eta>\delta>0$. By Remark 3.2 and if (5), (3) and (7) hold we can obtain:*

$$\sup_{u\in\Delta}\{C(\pi_\delta(u))-C(u)\}\leq(C(v)+M)\frac{\delta}{\eta}$$

3.2 Proof of the main results

The following two propositions prove the convergence of the approximate optimal cost functions when n tends to infinity.

Proposition 3.1 *If (1), (2), (5), (6), (7) and (8) hold, then:*

$$\forall\epsilon>0\quad\exists N(\epsilon)/\ n\geq N(\epsilon)\Longrightarrow R-R_n\leq\epsilon$$

Proof. Let δ de such that $\eta > \delta > 0$ and let u_n be ϵ-optimal for R_n, that means:

$$R_n + \epsilon = \inf_{u \in \Delta_n} C_n(u) + \epsilon \geq C_n(u_n)$$

then:

$$R - R_n \leq C(\pi_\delta(u_n)) - C_n(u_n) + \epsilon = C(\pi_\delta(u_n)) - C(u_n) + C(u_n) - C_n(u_n) + \epsilon \qquad (16)$$

but by Lemma 3.2 there exists $N_1(\epsilon)$ such that for all $n \geq N_1(\epsilon)$ we have:

$$C(\pi(u_n)) - C(u_n) \leq (C(v) + M)\frac{\delta}{\eta}, \qquad \forall \delta < \eta$$

and by (2) there exists $N_2(\epsilon)$ such that for all $n \geq N_2(\epsilon)$ we have $C(u_n) - C_n(u_n) \leq \epsilon$, by these two inequalities (16) becomes:

$$R - R_n \leq 2\epsilon + (C(v) + M)\frac{\delta}{\eta} \quad \forall \epsilon, \quad \forall \delta, \quad n \geq \max\left(N_1(\epsilon), N_2(\epsilon)\right)$$

■

Remark 3.4 *Let δ be such that $\eta > \delta > 0$. If (1), (6), (7) and (8) hold, there exists $\overline{N}(\delta)$ such that $\pi_\delta(U) \in \Delta_n$ if $n \geq \overline{N}(\delta)$, $\forall u \in U$. In fact by (1) there exists $N(\delta)$ such that if $n \geq N(\delta)$, $D_n(\pi_\delta(u)) \leq D(\pi_\delta(u)) + \delta/2 \ \ \forall u \in U$, and by (11)*

$$D_n(\pi_\delta(u)) \leq D(\pi_\delta(u)) + \frac{\delta}{2} \leq V - \delta + \frac{\delta}{2} \leq V - \frac{\delta}{2}$$

then, by (8) we have that there exists $\hat{N}$ such that $V - \delta/2 \leq V_n$ if $n \geq \hat{N}$, so, we have $D_n(\pi_\delta(u)) \leq V_n$ for all $n \geq \overline{N}(\delta) = \max\left(N(\delta), \hat{N}\right)$.

Proposition 3.2 *If (1), (2), (4), (7) and (8) hold, then:*

$$\forall \epsilon > 0 \ \ \exists N(\epsilon) \text{ such that } n \geq N(\epsilon) \Longrightarrow R_n - R \leq \epsilon$$

Proof. Let δ be sucht that $\eta > \delta > 0$ and let $\overline{u}$ be ϵ-optimal for R, that means:

$$R + \epsilon = \inf_{u \in \Delta} C(u) + \epsilon \geq C(\overline{u})$$

then by Remark 3.4, exist $\overline{N}(\delta)$ such that if $n \geq \overline{N}(\delta)$ (in order to insure $\pi_\delta(\overline{u}) \in \Delta_n$) we have:

$$\begin{aligned} R_n - R &= \inf_{u \in \Delta_n} C_n(u) - \inf_{u \in \Delta} C(u) \\ &\leq C_n(\pi_\delta(\overline{u})) - C(\overline{u}) + \epsilon = \\ &\quad C_n(\pi_\delta(\overline{u})) - C(\pi_\delta(\overline{u})) + C(\pi_\delta(\overline{u})) - C(\overline{u}) + \epsilon \end{aligned} \qquad (17)$$

but by Remark 3.3 there exists $N_1(\epsilon)$ such that for all $n \geq N_1(\epsilon)$ we have:

$$C(\pi_\delta(\overline{u})) - C(\overline{u}) \leq (C(v) + M)\frac{\delta}{\eta}, \qquad \forall \delta < \eta$$

and by (2) there exists $N_2(\epsilon)$ such that for all $n \geq N_2(\epsilon)$ we have $C_n(\pi_\delta(\overline{u})) - C(\pi_\delta(\overline{u})) \leq \epsilon$, by these two inequalities (17) becomes:

$$R_n - R \leq 2\epsilon + (C(v) + M)\frac{\delta}{\eta} \quad \forall \epsilon, \quad \forall \delta \quad n \geq \max\left(\overline{N}(\delta), N_1(\epsilon), N_2(\epsilon)\right)$$

■

The above two propositions prove theorem 3.1.

Proof of Theorem 3.2: Let u_n be ϵ_n-optimal for R_n, then:

$$C(\pi_\delta(u_n)) - R = C(\pi_\delta(u_n)) - C(u_n) + C(u_n) - C_n(u_n) + C_n(u_n) - R \tag{18}$$

but by Lemma 3.2 there exists $N_1(\epsilon)$ such that

$$C(\pi_\delta(u_n)) - C(u_n) \leq (C(v) + M)\frac{\delta}{\eta}$$

if $n \geq N_1(\epsilon)$, $\forall \delta \leq \eta$; then by (2) and Theorem 3.1 there exists $N(\epsilon)$ such that (18) becomes:

$$C(\pi_\delta(u_n)) - R \leq (C(v) + M)\frac{\delta}{\eta} + 2\epsilon + \epsilon_n \quad \forall n \geq N(\epsilon). \quad \forall \epsilon, \quad \forall \delta$$

This establishes i). The proof of ii) is similar.
iii) Let u_n be ϵ-optimal for R_n such that $u_n \to u$ when $n \to \infty$. By the lower semicontinuity of D, (1) and (8) we have that for all $\epsilon > 0$ there exists N such that $D(u) \leq D(u_n) + \epsilon \leq D_n(u_n) + 2\epsilon \leq V_n + 2\epsilon \leq V$ if $n \geq N$, that implies $u \in \Delta$.
As u_n is ϵ-optimal for R_n we have that $C_n(u_n) - R_n \leq \epsilon_n \quad n \geq N_1$, then:

$$C(u) - R = C(u) - C(u_n) + C(u_n) - C_n(u_n) + C_n(u_n) - R_n + R_n - R \tag{19}$$

by the lower semicontinuity of C we have that there exist N_2 such that $\forall n \geq N_2$, $C(u) - C(u_n) \leq \epsilon$, then by (2) and Theorem 3.1 (19) becomes:

$$C(u) - R \leq 3\epsilon + \epsilon_n \quad n \geq \max(N, N_1, N_2)$$

■

3.3 Weaker conditions for convergence

In this section we show that we can assume a weaker hypothesis. In particular we are going to prove that hypothesis (1) and (2) can be replaced by the weaker one:

$$D_n \to D \text{ and } C_n \to C \qquad \text{uniformly in}$$

$$\overline{\Delta} = \{u \in U : D(u) \leq V + \hat{\epsilon};\, C(u) \leq C(v) + \hat{\epsilon}\} \tag{20}$$

where $\hat{\epsilon} > 0$ is an arbitrary constant. We call **(H')** the set of hypothesis (4)-(8) and (20). We can prove the following convergence Theorem:

Theorem 3.3 *If* **(H')** *holds then* $R_n \to R$ *when* $n \to \infty$

To prove this theorem we define:

$$\overline{\Delta}_n = \{u \in U : D(u) \leq V + \hat{\epsilon},\, C(u) \leq C(v) + \hat{\epsilon},\, D_n(u) \leq V_n\},$$

we consider the following auxiliary problem:

$$\text{Find} \quad \overline{R}_n = \inf_{u \in \overline{\Delta}_n} C_n(u), \qquad \overline{R} = \inf_{u \in \overline{\Delta}} C(u)$$

and we prove the following to lemmas:

Lemma 3.3 *If* **(H')** *holds then* $\overline{R}_n \to \overline{R}$ *when* $n \to \infty$

Proof. The proof of the convergence follows the same arguments of the above sections, we must only change the definition of function ϵ_δ, in this case we define:

$$\pi_\delta : U \longrightarrow \overline{\Delta}; \qquad \pi_\delta(u) = \epsilon_\delta(u)v + (1 - \epsilon_\delta(u))u$$

with:
$\epsilon_\delta(u) =$

$$\min_{\lambda \in [0,1]} \{\lambda D(v) + (1-\lambda)D(u) \leq V - \delta \text{ and } \lambda C(v) + (1-\lambda)C(u) \leq C(v) + \hat{\epsilon} - \delta\}$$

Where $\hat{\epsilon}$ is given in (20) and δ is fixed such that $0 < \delta < \min(\eta, \hat{\epsilon})$.
With these assumptions the set
$\{\lambda D(v) + (1-\lambda)D(u) \leq V - \delta \text{ and } \lambda C(v) + (1-\lambda)C(u) \leq C(v) + \hat{\epsilon} - \delta\} \neq \emptyset$.
Now it is easy to prove $\pi_\delta : U \to \overline{\Delta}_n$ if n is large enough and:

$$\sup_{u \in \overline{\Delta}} \epsilon_\delta(u) \leq \max\left(\frac{\delta}{\eta}, \frac{\delta}{\hat{\epsilon}}\right); \qquad \limsup_{n \to \infty} \sup_{u \in \overline{\Delta}_n} \epsilon_\delta(u) \leq \max\left(\frac{\delta}{\eta}, \frac{\delta}{\hat{\epsilon}}\right)$$

and the proof follows straightforward as in the previous sections.
■

Now we are going to prove that our two problems are equivalents.

Lemma 3.4 *If* **(H')** *holds then* $R = \overline{R}$

Proof. Let $u \in \Delta$ ϵ-optimal for R, so $C(u) \leq C(v) + \epsilon$, moreover as $u \in \Delta$, $D(u) \leq V$ then $u \in \overline{\Delta}$ that implies $\overline{R} \leq C(u)$, then we have:

$$\overline{R} \leq C(u) \leq R + \epsilon \qquad \forall \epsilon$$

then we have proved $\overline{R} \leq R$. By definition of Δ_n and $\overline{\Delta}_n$ it is easy to see that $R_n \leq \overline{R}_n, \quad \forall n$, so we have:

$$R_n \to R; \quad R_n \leq \overline{R}_n; \quad \overline{R}_n \to \overline{R}; \quad \overline{R} \leq R$$

that obviously implies $R = \overline{R}$.

■

The above lemmas now imply Theorem 3.3.

3.4 The case where U depends on n

Let be

$$C_n : U_n \to \mathbb{R} \qquad \mathrm{n} = 1, ..., \infty; \qquad \mathrm{D_n} : \mathrm{U_n} \to \mathbb{R}^{\mathrm{K}} \qquad \mathrm{n} = 1, ..., \infty$$

$$\Delta_n = \{u \in U_n : D_n(u) \leq V_n\} \quad n = 1, 2... \quad \Delta = \{u \in U : D(u) \leq V\}, \quad (U = U_\infty)$$

We define:

$$R_n = inf_{u \in \Delta_n} C_n(u); \qquad R = inf_{u \in \Delta} C(u)$$

We are going to assume some hipothesis ($\mathbf{H}''$) in order to obtain $U_n \to U$ in "some sense", so we require that there exist functions $\sigma_1^n : U_n \to U$, $\sigma_2^n : U \to U_n$ such that:

$$\limsup_{n \to \infty} \left\{ \sup_{u \in \Delta_n} [D(\sigma_1^n(u)) - D_n(u)] \right\} \leq 0 \tag{21}$$

$$\limsup_{n \to \infty} \left\{ \sup_{u \in \Delta} [D_n(\sigma_2^n(u)) - D(u)] \right\} \leq 0 \tag{22}$$

$$\limsup_{n \to \infty} \left\{ \sup_{u \in \Delta_n} [C(\sigma_1^n(u)) - C_n(u)] \right\} \leq 0 \tag{23}$$

$$\limsup_{n \to \infty} \left\{ \sup_{u \in \Delta} [C_n(\sigma_2^n(u)) - C(u)] \right\} \leq 0 \tag{24}$$

We are going to call ($\mathbf{H}''$) the set of hypothesis (4)-(8) with (21)-(24). Note that (21)-(24) replace references (1)-(2) in the case $U = U_n$, $\forall n$. We want to prove:
(i) $R_n \to R$ when $n \to \infty$.
(ii) u_n^* being ϵ_n-optimal policy for the nth optimal cost function R_n implies $\pi_\delta(\sigma_1^n(u_n^*))$ "almost" optimal for the limit optimal cost function R, for n large enough.
iii) Let u be ϵ-optimal for R, then $\sigma_2^n(\pi_\delta(u))$ is almost optimal for R_n for n large enough .
And under more restrictive assumptions:
(iv) Let $u \in U$ be some limit point of $\sigma_1^n(u_n^*)$, $u_n^* \in U_n$, defined above, then $\sigma_1^n(u)$ is $O(\epsilon)$-optimal for the limit optimal cost function.

In this subsection we are going to prove the following theorems:

Theorem 3.4 *If* $(\mathbf{H}'')$ *holds then* $R_n \to R$ *when* $n \to \infty$

Theorem 3.5 *Let* δ *be such that* $\eta > \delta > 0$. *If* $(\mathbf{H}'')$ *holds then:*
i) Let u_n *be* ϵ_n*-optimal for* R_n, $\limsup_n \epsilon_n \leq \epsilon$, *then there exists* $n \geq N(\epsilon, \delta)$ *such that* $\forall n \geq N(\epsilon, \delta)$, $\pi_\delta(\sigma_1^n(u_n))$ *is* $O(\epsilon + \delta)$*-optimal for* R.
ii) Let u *be* ϵ*-optimal for* R *then there exists* $N(\epsilon, \delta)$ *such that* $\forall n \geq N(\epsilon, \delta)$, $\sigma_2^n(\pi_\delta(u))$ *is* $O(\epsilon + \delta)$*-optimal for* R_n.
Moreover if there exists $M \geq 0$ *such that* C *and* D *are lower semicontinuous in* $\bigcup_{n \geq M} \Delta_n \cup \Delta$ *then:*
iii) Let u_n *be* ϵ_n*-optimal for* R_n, $\limsup_n \epsilon_n \leq \epsilon$, *such that* $\sigma_1^n(u_n) \to u$ *when* $n \to \infty$ *then* u *is* $O(\epsilon)$*-optimal for* R.

To prove these theorems we prove first the following lemmas:

Lemma 3.5 *Let* δ *be such that* $\eta > \delta > 0$. *If (7), (8) and (21) hold, then:*

$$\limsup_{n \to \infty} \left\{ \sup_{u \in \Delta_n} \epsilon_\delta(\sigma_1^n(u)) \right\} \leq \frac{\delta}{\eta}$$

Proof. Let $u \in \Delta_n$, then $D_n(u) \leq V_n$; by (21) and (8) we have:

$$\forall \hat{\epsilon} > 0 \quad \exists N(\hat{\epsilon}) / \quad \forall n \geq N(\hat{\epsilon}) \quad D(\sigma_1^n(u)) \leq D_n(u) \leq V + \hat{\epsilon} \tag{25}$$

by (7) and (25) we have:

$$\lambda D(v) + (1 - \lambda) D(\sigma_1^n(u)) \leq \lambda(V - \eta) + (1 - \lambda)(V + \hat{\epsilon}) \tag{26}$$

then as in Lemma 2.1 we obtain:

$$\left[\frac{\hat{\epsilon} + \delta}{\eta + \hat{\epsilon}}, 1 \right] \subseteq \{ \lambda \in [0,1] : \lambda D(v) + (1 - \lambda) D(\sigma_1^n(u)) \leq V - \delta \} \qquad \forall \hat{\epsilon} > 0$$

and then by definition of $\epsilon_\delta(\sigma_1^n(u))$ we have:

$$\epsilon_\delta(\sigma_1^n((u)) \leq \frac{\hat{\epsilon} + \delta}{\eta + \hat{\epsilon}} \qquad \forall u \in \Delta_n, \quad n \geq N(\hat{\epsilon}), \quad \forall \hat{\epsilon} > 0.$$

and with this last inequality we deduce the thesis.
■

Analogously to Lemma 2.2 we can prove:

Lemma 3.6 *Let* δ *be such that* $\eta > \delta > 0$. *If (5), (6), (7), (8) and (21) hold, then:*

$$\limsup_{n \to \infty} \left[\sup_{u \in \Delta_n} \{ C(\pi_\delta(\sigma_1^n(u))) - C(\sigma_1^n(u)) \} \right] \leq (C(v) + M) \frac{\delta}{\eta}$$

Wiht the following two propositions can easily prove theorem 3.4.

Proposition 3.3 *If (5), (6), (7), (21) and (23) hold, then:*

$$\forall \epsilon > 0 \quad \exists N(\epsilon) / \; n \geq N(\epsilon) \Longrightarrow R - R_n \leq \epsilon$$

Proof. Let δ be such that $\eta > \delta > 0$ and let u_n be ϵ-optimal for R_n then

$$R - R_n \leq C(\pi_\delta(\sigma_1^n(u_n))) - C_n(u_n) + \epsilon =$$

$$C(\pi_\delta(\sigma_1^n(u_n))) - C(\sigma_1^n(u_n)) + C(\sigma_1^n(u_n)) - C_n(u_n) + \epsilon \tag{27}$$

but by Lemma 3.6 there exists $N_1(\epsilon)$ such that for all $n \geq N_1(\epsilon)$ we have:

$$C(\pi_\delta(\sigma_1^n(u_n))) - C(\sigma_1^n(u_n)) \leq (C(v) + M)\frac{\delta}{\eta}, \qquad \forall \delta < \eta$$

and by (23) there exists $N_2(\epsilon)$ such that for all $n \geq N_2(\epsilon)$ we have
$C(\sigma_1^n(u_n)) - C_n(u_n) \leq \epsilon$, by this two inequalities (27) becomes:

$$R - R_n \leq 2\epsilon + (C(v) + M)\frac{\delta}{\eta} \quad \forall \epsilon, \quad \forall \delta, \quad n \geq \max(N_1(\epsilon), N_2(\epsilon))$$

■

Remark 3.5 *If (6), (7), (8) and (22) hold, there exists N such that $(\sigma_2^n \circ \pi_\delta)(U) \subset \Delta_n$ if $n \geq N$. In effect by (22) there exists N such that if $n \geq N$ then $D_n(\sigma_2^n(\pi_\delta(u))) \leq D(\pi_\delta(u)) + \frac{\delta}{2}$ $\forall u \in U$, and by (11)*

$$D_n(\sigma_2^n(\pi_\delta(u))) \leq D(\pi_\delta(u)) + \frac{\delta}{2} \leq V - \delta + \frac{\delta}{2} \leq V - \frac{\delta}{2}$$

then by (8) we have that there exists $\hat{N}$ such that $D_n(\sigma_2^n(\pi_\delta(u))) \leq V_n$ for all $n \geq \hat{N}$.

With this Remark we can prove the following proposition as in Proposition 2.2

Proposition 3.4 *If (4), (7), (8), (22) and (24) hold, then:*

$$\forall \epsilon > 0 \quad \exists N(\epsilon) / \; n \geq N(\epsilon) \Longrightarrow R_n - R \leq \epsilon$$

Now, theorem 3.4 follows from propositions 3.3 and 3.4.

Proof of Theorem 3.5: Let u_n be ϵ_n-optimal for R_n, then:

$$C(\pi_\delta(\sigma_1^n(u_n))) - R =$$

$$C(\pi_\delta(\sigma_1^n(u_n))) - C(\sigma_1^n(u_n)) + C(\sigma_1^n(u_n)) - C_n(u_n) + C_n(u_n) - R_n + R_n - R \tag{28}$$

but by Lemma 3.6 there exists $N_1(\epsilon)$ such that

$$C(\pi_\delta(\sigma_1^n(u_n))) - C(\sigma_1^n(u_n)) \leq (C(v) + M)\frac{\delta}{\eta}$$

if $n \geq N_1(\epsilon)$, $\forall \delta \leq \eta$; then by (23) and Theorem 3.1 there exists $N(\epsilon)$ such that (28) becomes:

$$C(\pi_\delta(\sigma_1^n(u_n))) - R \leq (C(v) + M)\frac{\delta}{\eta} + 2\epsilon + \epsilon_n \qquad \forall n \geq N(\epsilon). \qquad \forall \epsilon, \qquad \forall \delta$$

With this we prove i). The proof of ii) is similar.
iii) By the lower semicontinuity of D we have that $u \in \Delta$.
Let u_n be ϵ-optimal for R_n, then we have:

$$C(u) - R = C(u) - C(\sigma_1^n(u_n)) + C(\sigma_1^n(u_n)) - C_n(u_n) + C_n(u_n) - R_n + R_n - R \qquad (29)$$

the lower semicontinuity of C implies that there exists N such that $\forall n \geq N$, $C(u) - C(\sigma_1^n(u_n)) \leq \epsilon$, then by (23) and Theorem 3.4 (29) becomes:

$$C(u) - R \leq 3\epsilon + \epsilon_n$$

■

Remark 3.6 *In iii) of Theorem 3.5 we need the lower semicontinuity of D to prove that $u \in \Delta$. If we do not have this property we can still conclude that $\pi_\delta(u)$ is $(\epsilon + \delta)$-optimal for R, $\forall \eta > \delta > 0$.*

4 Constrained MDPs: the convergence in the discount factor

In this Section we consider constrained Markov Decision Processes (MDP), known also as controlled Markov chains, with a general ergodic structure (multi-chain). We consider the discounted cost and the average cost. We shall obtain new results on the convergence of the values and optimal policies of the discounted cost, as the discount factor tends to one, to the value and to optimal policies corresponding to the expected average cost. A similar result for the special unichain case was obtained in [2].

Consider an MDP with a finite state space $\mathbf{X} = \{0, 1, ..., N\}$ and a finite *action space* $\mathbf{A}$. Without loss of generality, we assume that in any state x all actions in $\mathbf{A}$ are available. The probability to go from state x to state y given that action a is used, is given by the transition probability P_{xay}. A policy u in the *policy space* $\mathbf{U}_h$ is described as $u = \{u_1, u_2, ...\}$, where u_t, applied at time epoch t, is a probability measure over $\mathbf{A}$ conditioned on the whole history of actions and state prior to t, as well as the state at time t. Given an initial distribution β on $\mathbf{X}$, each policy u induces a probability measure denoted by P_β^u on the space of sample paths of states and actions (which serves as the canonical sample space Ω). The corresponding expectation operator is denoted by E_β^u. On this probability space are defined the state and action processes, X_t, A_t, $t = 1, 2, ...$

A *Markov policy* $u \in \mathbf{U}_M$ is characterized by the dependence of u_{t+1} on the current state and the time only. A *stationary policy* $g \in \mathbf{U}_S$ is characterized by a single conditional

probability measure $p^g_{\cdot|x}$ over $\mathbf{A}$, so that $p^g_{\mathbf{A}|x} = 1$; under g, X_t becomes a Markov chain with stationary transition probabilities, given by $P^g_{xy} = \sum_{a\in\mathbf{A}} p^g_{a|x} P_{xay}$. The class of *stationary deterministic policies* $\mathbf{U}_D$ is a subclass of $\mathbf{U}_S$, and every $g \in \mathbf{U}_D$ is characterized by a mapping g:$\mathbf{X} \to \mathbf{A}$, so that $p^g_{\cdot|x} = \delta_{g(x)}(\cdot)$ is concentrated at the point $g(x)$ for each x. Let L be the number of stationary deterministic policies among $\mathbf{U}_D$, and enumerate the policies in $\mathbf{U}_D$ such that $\mathbf{U}_D = \{u^1, ..., u^L\}$.

It will often be usefull to extend the definition of a policy $u = (u_1, u_2, ...)$ so as to allow u_t to depend not only on the history, but also on some additional randomizing mechanism. In particular, for any finite class of policies $G \subset \mathbf{U}_h$, we define $\overline{M}(G)$ to be the class of mixed policies generated by G, we call these mixed-G policies. A mixed-G policy $\hat{q}$ is identified with a distribution q over G; the controller first uses q to choose some policy $u \in G$, and then proceeds with that policy from time 1 onwards. Define $\mathcal{U} := \overline{M}(U_D)$. Finally, we denote $\mathbf{U} = \mathbf{U}_h \cup \mathcal{U}$.

Definition 4.1 *For any initial distribution q over the set $\mathbf{U}_D$, we shall identify the policy $m(q) \in \mathcal{U}$ to be the the one that chooses initially the policy u^j with probability q_j.*

Any given distribution β for the initial state (at time 1) and a policy u define a probability measure P^u_β on which the stochastic processes X_t and A_t of the states and actions are defined. When β is concentrated on some state x (i.e. $\beta = \delta_x$), we shall use the notation P^u_x instead of P^u_β.

Let $c : \mathbf{X}\times\mathbf{A} \to \mathbb{R}$, and $d : \mathbf{X}\times\mathbf{A} \to \mathbb{R}^{\mathrm{K}}$ be immediate cost functions, $d = (d^1, d^2, ..., d^k)$.

Fix some discount factor $\alpha \in [0, 1)$, and defined the normalized discounted costs corresponding to an initial distribution β and a policy u by

$$\begin{aligned} C_\alpha(\beta, u) &= (1-\alpha)\sum_{t=1}^{\infty} E^u_\beta \alpha^{t-1} c(X_t, A_t) \\ D^k_\alpha(\beta, u) &= (1-\alpha)\sum_{t=1}^{\infty} E^u_\beta \alpha^{t-1} d^k(X_t, A_t), \qquad k = 1, ..., K. \end{aligned}$$

Define the average costs associated with a policy u and with an initial distribution β on $\mathbf{X}$:

$$\begin{aligned} C_{ea}(\beta, u) &= \overline{\lim_{t\to\infty}} \frac{1}{t} E^u_\beta \left[\sum_{s=1}^{t} c(X_s, A_s)\right] \\ D^k_{ea}(\beta, u) &= \overline{\lim_{t\to\infty}} \frac{1}{t} E^u_\beta \left[\sum_{s=1}^{t} d^k(X_s, A_s)\right], \qquad k = 1, ..., K. \end{aligned}$$

Given a vector $V \in \mathbb{R}^{\mathrm{K}}$, we consider the subset $\Pi_V \subset \mathbf{U}$ of policies satisfying the constraints

$$D(\beta, u) \le V. \tag{30}$$

A policy $u \in \Pi_V$ is called feasible. We introduce the following Constrained Problem COP: find a policy u^* that achieves

$$C(\beta) := \inf_{u \in \Pi_V} C(\beta, u). \tag{31}$$

In (30) and (31), the costs stand for either the discounted or the average cost. COP is said to be feasible if Π_V is nonempty.

We are now ready to state the first main result:

Theorem 4.1 *(Convergence of the value) Assume that there exists some policy $v \in \mathbf{U}$ such that*

$$D^k(\beta, v) \leq V^k - \eta, \qquad \forall k = 1, ..., K, \tag{32}$$

for some $\eta > 0$. Then, the value converges in the discount factor:

$$\lim_{\alpha \to 1} C_\alpha(\beta) = C_{ea}(\beta).$$

The proof of this Theorem, as well as the convergence of optimal policies, are delayed to the next section.

5 Convergence of the values and the policies

In order to be able to define the convergence of optimal policies, we shall show that one may restrict the search of optimal policies to the simple subclasses of policies $\mathcal{U}$, without loss of optimality. Moreover, we should relate the solutions of COP to solutions of mathematical programming, in order to be able to apply the tools that we developed. Note that the control problem is already of the form of a Mathematical program, but the cost is not convex in the policies. We shall show that when restricting to $\mathcal{U}$, the costs are convex functions.

There are several ways to solve (31). For the discounted cost, the solution was given by Kallenberg in [20] using a LP approach. For the expected average cost, there are several possible LP approaches: the one by Hordijk and Kallenberg [19, 20], the one by Feinberg [15], and a related one by Altman and Shwartz [8]; for a definition slightly different than in (31), an efficient LP method for computing ϵ-optimal solutions was obtained by Ross and Varadarajan, see [25, 26]. Lagrangian techniques have also been used to solve constrained MDPs with a single constraint, see Beutler, Ross and Sennott [10, 11, 27, 28]. The relation between the Lagrange and the LP approaches was pointed out in [9].

Remark 5.1 *All the references above considered the solution of the constrained MDP among the policies $\mathbf{U}_h$. However, a standard argument due to Derman and Strauch [14] shows (in a constructive way) that for any policy in $u \in \mathbf{U}$, there exist an equivalent policy in $\chi \in \mathbf{U}_M$ under which the marginal probabilities of the states and actions are the same as those under u, and in particular, both the discounted and the expected average costs are the same. Thus below, whenever we obtain an optimal policy among $\mathcal{U}$, one may consider the policy χ instead without loss of optimality.*

We would like to prove the convergence results by showing that there is a correspondence between values and optimal solutions of the control problem, and values and optimal solutions of related LPs, and then use the general results from the previous section. While this is possible for the uni-chain case, it turns out that for the multi-chain case, the LP introduced by Kallenberg [20] for the discounted cost is completely different than any of the LPs for the expected average cost, (e.g. the number of decision variables is different). Therefore, as a first step, we shall introduce a new LP method for computing the value and optimal policies for the discounted cost problem, which is an adaptation of the one for the expected average cost in Feinberg [15] and Altman and Shwartz [8]. This will allow us to have the same type of LP for both the discounted and the average cost.

Denote the simplex $S(L) := \{\gamma \in R^L : \gamma_i \geq 0, i = 1, ..., L, \ \sum_{i=1}^L \gamma_i = 1\}$. Introduce the following $\mathbf{LP}_\alpha$: Find $\gamma^* \in S(L)$ that achieves:

$$\mathcal{C}_\alpha^* := \min_{\gamma \in S(L)} \sum_{i=1}^L \gamma_i C_\alpha(\beta, u^i), \qquad s.t. \tag{33}$$

$$\mathcal{D}_\alpha^k(\gamma) := \sum_{i=1}^L \gamma_i D_\alpha^k(\beta, u^i) \leq V^k, \ k = 1, ..., K \tag{34}$$

Define $\mathcal{C}_\alpha(\gamma) := \sum_{i=1}^L \gamma_i C_\alpha(\beta, u^i)$. We say that the LP is feasible if the subset of $S(L)$ satisfying the constraints (34) is nonempty.

Theorem 5.1 *(Relation between LP and the constrained MDP, the discounted cost).*
(i) For any $\gamma \in S(L)$, the policy $m(\gamma) \in \mathcal{U}$ (see Definition 4.1) satisfies

$$C_\alpha(\beta, m(\gamma)) = \mathcal{C}_\alpha(\gamma), \qquad D_\alpha^k(\beta, m(\gamma)) = \mathcal{D}_\alpha^k(\gamma), \qquad k = 1, ..., K.$$

(ii) For any vector of costs

$$\{C_\alpha(\beta, u), D_\alpha^k(\beta, u), k = 1, ..., K\}$$

achievable by some policy $u \in \mathbf{U}$, there exists some $v \in \mathcal{U}$ achieving the same vector of costs.
(iii) COP_α is feasible if and only if LP_α is, and the optimal values are the same: $\mathcal{C}_\alpha^* = C_\alpha(\beta)$. Moreover, if γ^* is optimal for LP_α, then $m(\gamma^*)$ is optimal for COP_α.

Proof. Denote

$$f_\alpha(\beta, u; y, a) := (1 - \alpha) \sum_{t=1}^\infty P_\beta^u(X_t = y, A_t = a), \qquad x \in \mathbf{X}, a \in \mathbf{A},$$

and let $f_\alpha(\beta, u)$ be the vector whose $(y, a)th$ elements are given by $f_\alpha(\beta, x; y, a)$. Then (see e.g. [12])

$$C_\alpha(\beta, u) = c \cdot f_\alpha(\beta, u) = \sum_{y \in \mathbf{X}} \sum_{a \in \mathbf{A}} c(y, a) f_\alpha(\beta, u; y, a), \tag{35}$$

with a similar representation for the costs $D^k_\alpha(\beta,u)$. For any class of policies $\overline{U}$, denote $\mathbf{L}_\alpha(\beta,\overline{U}) = \cup_{u\in\overline{U}} f_\alpha(\beta,u)$. It is known that the set $\mathbf{L}_\alpha(\beta,\mathbf{U}_h)$ is convex, compact, and its extreme points are $\{f_\alpha(\beta,u), u\in\mathbf{U}_D\}$, see [12, 20]. For any probability γ over $\mathbf{U}_D$, we clearly have

$$f_\alpha(\beta,m(\gamma)) = \sum_{i=1}^{L}\gamma_i f_\alpha(\beta,u^i). \tag{36}$$

Consequently, $\mathbf{L}_\alpha(\beta,\mathbf{U})$ is convex, compact, and its extreme points are $\{f_\alpha(\beta,u), u\in\mathbf{U}_D\}$.

Combining this with (35), we conclude that the set of achievable costs

$$\cup_{u\in\mathbf{U}}\{C_\alpha(\beta,u), D^k_\alpha(\beta,u), k=1,...,K\} \tag{37}$$

is also is convex, compact, and its extreme points are

$$\{C_\alpha(\beta,u), D^k_\alpha(\beta,u), k=1,...,K, u\in\mathbf{U}_D\}. \tag{38}$$

By combining (35) with (36), we get for any probability γ over $\mathbf{U}_D$

$$C_\alpha(\beta,m(\gamma)) = \sum_{i=1}^{L}\gamma_i C_\alpha(\beta,u^i), \qquad D^k_\alpha(\beta,m(\gamma)) = \sum_{i=1}^{L}\gamma_i C_\alpha(\beta,u^i),\ k=1,...,K.$$

Hence, the set of performance measures achievable by $u\in\mathcal{U}$ is also convex, compact, with the extreme points in the set (38), and thus, equal to the set (37) achievable by all policies. This establishes (i) and (ii), and implies (iii). ∎

The LP method corresponding to (33) for the expected average cost, due to Feinberg [15], is the same: $\mathbf{LP_{ea}}$: Find $\gamma^*\in S(K)$ that achieves:

$$\mathcal{C}^*_{ea} := \min_{\gamma\in S(L)}\sum_{i=1}^{L}\gamma_i C_{ea}(\beta,u^i), \qquad s.t. \tag{39}$$

$$\mathcal{D}^k_{ea}(\gamma) := \sum_{i=1}^{L}\gamma_i D^k_{ea}(\beta,u^i) \le V^k,\ k=1,...,K \tag{40}$$

Define $\mathcal{C}_{ea}(\gamma) := \sum_{i=1}^{L}\gamma_i C_{ea}(\beta,u^i)$.

Theorem 5.2 *(Relation between LP and the constrained MDP, the expected average cost). (i) For any $\gamma\in S(L)$, the policy $m(\gamma)\in\mathcal{U}$ (see Definition 4.1) satisfies*

$$C_{ea}(\beta,m(\gamma)) = \mathcal{C}_{ea}(\gamma), \qquad D^k_{ea}(\beta,m(\gamma)) = \mathcal{D}^k_{ea}(\gamma), \qquad k=1,...,K.$$

(ii) For any vector of costs

$$\{C_{ea}(\beta,u), D^k_{ea}(\beta,u), k=1,...,K\}$$

achievable by some policy $u \in \mathbf{U}$, there exists a dominating $v \in \mathcal{U}$, i.e. such that

$$C_{ea}(\beta, v) \leq C_{ea}(\beta, u), \quad D^k_{ea}(\beta, v) \leq D^k_{ea}(\beta, u), k = 1, ..., K.$$

(iii) COP_{ea} is feasible if and only if LP_{ea} is, and the optimal values are the same: $\mathcal{C}^*_{ea} = C_{ea}(\beta)$. Moreover, if γ^* is optimal for LP_{ea}, then $m(\gamma^*)$ is optimal for COP_{ea}.

Proof. Denote

$$f^t_{ea}(\beta, u; y, a) := t^{-1} \sum_{s=1}^{t} P^u_\beta(X_s = y, A_s = a), \qquad x \in \mathbf{X}, a \in \mathbf{A},$$

and let $f^t_{ea}(\beta, u)$ be the vector whose $(y, a)th$ elements are given by $f^t_{ea}(\beta, x; y, a)$. For any $u \in U_D$, since the setate process is a Markov chain, it is known that

$$f_{ea}(\beta, u) = \lim_{t \to \infty} f^t_{ea}(\beta, u) \quad \text{exist}, \tag{41}$$

and it is straight-forward to show that

$$C_{ea}(\beta, u) = c \cdot f_{ea}(\beta, u) = \sum_{y \in \mathbf{X}} \sum_{a \in \mathbf{A}} c(y, a) f_{ea}(\beta, u; y, a), \tag{42}$$

with a similar representation for the costs $D^k_{ea}(\beta, u)$. It is then clear that (41) and (42) hold in fact for any $u \in \mathcal{U}$, which establishes (i).

For a fixed initial distribution β, and for any policy $v \in \mathbf{U}_h$, and any accumulation point f of the sequence $f^t_{ea}(\beta, u)$, there exists some $u \in \mathcal{U}$ such that $f_{ea}(\beta, u) = f$. This is a direct consequence of Theorem 2 in [19], and is a special case of the result in [15] (who studies the Semi-Markov case). Combining this, with the fact that (41) and (42) hold for any $u \in \mathcal{U}$, establishes (ii), by using Corollary 2.5 in [5]. Finally, (iii) is a consequence of (i) and (ii). ■

For a given $u \in \mathcal{U}$ we shall understand below $\pi_\delta(u) = m(\pi_\delta(\gamma))$ where γ is such that $u = m(\gamma)$ and π_δ is defined in (10). We are now ready to state the second main result for MDPs:

Theorem 5.3 *Assume that the Slater conditions (32) hold. Consider a sequence α_n converging to 1, and let COP_n be the constrained optimal control problem corresponding to the discount factor α_n. Let δ be such that $\eta > \delta > 0$. Then*
i) Let $u_n \in \mathcal{U}$ be ϵ_n-optimal for COP_n, $\limsup_n \epsilon_n \leq \epsilon$, then there exist $N(\epsilon, \delta)$ such that $\forall n \geq N(\epsilon, \delta)$, $\pi_\delta(u_n)$ is $O(\epsilon + \delta)$-optimal for COP_{ea}.
ii) Let $u \in \mathcal{U}$ be optimal for COP_{ea}. Then there exist $N(\epsilon, \delta)$ such that $\forall n \geq N(\epsilon, \delta)$, $\pi_\delta(u)$ is $O(\epsilon + \delta)$-optimal for COP_n.
iii) Let $u_n \in \mathcal{U}$ be optimal for COP_n and let $\gamma_n \in S(L)$ be such that $u_n = m(\gamma_n)$. Assume that γ_n converges to some γ. Then $m(\gamma)$ is optimal for COP_{ea}.

Proof of Theorems 4.1 and 5.3:
We apply below Theorems 3.1 and 3.2 to obtain the convergence of the optimal values and the convergence and robustness of policies LP_α to the optimal value of LP_{ea}, and consequently, by Theorems 5.1 (ii), (iii) and 5.2 (ii), (iii), the convergence for the original constrained optimal control problems. It remains to show that the required conditions in Theorems 3.1 and 3.2 hold.

It is well known that for any $u \in U_D$,

$$\lim_{\alpha \to 1} C_\alpha(\beta, u) = C_{ea}(\beta, u), \qquad \lim_{\alpha \to 1} D^k_\alpha(\beta, u) = D^k_{ea}(\beta, u), \ k = 1, ..., K. \tag{43}$$

Choose an arbitrary $u \in \mathcal{U}$, and let γ be such that $u = m(\gamma)$. Then, due to Theorems 5.1 (i) and 5.2 (i),

$$\begin{aligned} & |C_\alpha(\beta, u) - C_{ea}(\beta, u)| \\ \leq & \left| \sum_{j=1}^{L} \gamma_j \left[C_\alpha(\beta, u^j) - C_{ea}(\beta, u^j) \right] \right| \\ \leq & \max_{1 \leq j \leq L} |C_\alpha(\beta, u^j) - C_{ea}(\beta, u^j)|. \end{aligned}$$

which does not depends on u.
With the same applied to the costs D^k, this implies that (43) holds for any $u \in \mathcal{U}$, and that the convergence is uniform over $\mathcal{U}$. Equivalently, for any $\gamma \in S(L)$,

$$\lim_{\alpha \to 1} \mathcal{C}_\alpha(\gamma) = \mathcal{C}_{ea}(\gamma), \qquad \lim_{\alpha \to 1} \mathcal{D}^k_\alpha(\gamma) = \mathcal{D}^k_{ea}(\gamma), \ k = 1, ..., K \tag{44}$$

uniformly in γ. This establishes conditions (1) and (2).

Since the set U in (3) is given by $S(L)$, it is clearly convex, which establishes condition (3). As the costs are linear in γ, conditions (4) and (5) hold. Since γ is bounded in a simplex, this implies condition (6).

It follows from condition (32) and from Theorem 5.2, that there exists some $\eta > 0$ and some $\gamma \in S(L)$, such that the policy $m(\gamma)$ satisfies:

$$\mathcal{D}^k_{ea}(\gamma) = D^k_{ea}(\beta, m(\gamma)) \leq V^k - \eta, \qquad k = 1, ..., K.$$

This establishes condition (7). Finally, condition (8) trivially holds, as $V = V_n$ do not depend on n. ∎

References

[1] E. Altman, "Denumerable Constrained Markov Decision Problems and Finite Approximations", *Math. of Operations Research*, **19**, No. 1, pp. 169-191, 1994.

[2] E. Altman, "Asymptotic Properties of Constrained Markov Decision Processes", *Zeitschrift für Operations Research*, Vol. 37, Issue 2, pp. 151-170, 1993.

[3] E. Altman and V. A. Gaitsgory, "Stability and Singular Perturbations in Constrained Markov Decision Problems", *IEEE Trans. Auto. Control*, **38**, No. 6, pp. 971-975, 1993.

[4] E. Altman and A. Shwartz, "Optimal priority assignment: a time sharing approach", *IEEE Transactions on Automatic Control* Vol. AC-34 No. 10, pp. 1089-1102, 1989.

[5] E. Altman and A. Shwartz, "Markov decision problems and state-action frequencies," *SIAM J. Control and Optimization.* **29**, No. 4, pp. 786-809, 1991

[6] E. Altman and A. Shwartz, "Adaptive control of constrained Markov chains", *IEEE Transactions on Automatic Control*, **36**, No. 4, pp. 454-462, 1991.

[7] E. Altman and A. Shwartz, "Sensitivity of constrained Markov Decision Problems", *Annals of Operations Research*, **32**, pp. 1-22, 1991.

[8] E. Altman and A. Shwartz, "Time-sharing policies for controlled Markov chains", *Operations Research*, **41**, No. 6, pp. 1116-1124, 1993.

[9] E. Altman and F. Spieksma, "The Linear Program approach in Markov Decision Problems revisited", to appear in *Zeitschrift für Operations Research*, **42**, Issue 2, 1995.

[10] F. J. Beutler and K. W. Ross, "Optimal policies for controlled Markov chains with a constraint", *J. Mathematical Analysis and Applications* **112**, 236-252, 1985.

[11] F. J. Beutler and K. W. Ross, "Time-Average Optimal Constrained Semi-Markov Decision Processes", *Advances of Applied Probability* **18**, No. 2, pp. 341-359, 1986.

[12] V. S. Borkar, "A convex analytic approach to Markov decision processes", *Probab. Th. Rel. Fields*, Vol. 78, pp. 583-602, 1988.

[13] G. B. Dantzig, J. Folkman and N. Shapiro, "On the continuity of the minimum set of a continuous function", *J. Math. Anal. and Applications,* **17**, 519-548, 1967.

[14] C. Derman and R. E. Strauch, "On memoryless rules for controlling sequential control processes", *Ann. Math. Stat>* **37**, pp. 276-278, 1966.

[15] E. A. Feinberg, "Constrained Semi-Markov Decision Processes With Average Rewards", submitted to *ZOR.*

[16] E. A. Feinberg and M. I. Reiman, "Optimality of randomized trunk reservation", submitted to *Probability in the Engineering and Informational Sciences.*

[17] V. A. Gaitsgory and A. A. Pervozvanskii, "Perturbation Theory for Mathematical Programming Problems", *JOTA*, 389-410, 1986.

[18] A. Hordijk and F. Spieksma, "Constrained admission control to a queuing system" *Advances of Applied Probability* Vol. 21, pp. 409-431, 1989.

[19] A. Hordijk and L. C. M. Kallenberg, "Constrained undiscounted stochastic dynamic programming", *Mathematics of Operations Research,* **9**, No. 2, May 1984.

[20] L. C. M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Mathematical Centre Tracts 148, Amsterdam, 1983.

[21] A. Lazar, "Optimal flow control of a class of queuing networks in equilibrium", *IEEE Transactions on Automatic Control*, Vol. 28 no. 11, pp. 1001-1007, 1983.

[22] P. Nain and K. W. Ross, "Optimal Priority Assignment with hard Constraint," *Transactions on Automatic Control,* Vol. 31 No. 10, pp. 883-888, October 1986.

[23] Pervozvanskii A. A. and V. A. Gaitsgory, *Theory of suboptimal Decision: Decomposition and Aggregation*, Kluwer Academic Publisher, Dordrecht, 1988.

[24] K. W. Ross and B. Chen, "Optimal scheduling of interactive and non interactive traffic in telecommunication systems", *IEEE Trans. on Auto. Control,* Vol. 33 No. 3 pp. 261-267, 1988.

[25] K. Ross and R. Varadarajan, 'Markov Decision Processes with Sample path constraints: the communicating case", *Operations Research*, **37**, No. 5, pp. 780-790, 1989.

[26] K. Ross and R. Varadarajan, "Multichain Markov Decision Processes with a Sample Path Constraint: A Decomposition Approach", *MOR*, Vol. 16 No. 1, pp. 195-207, 1991.

[27] L. I. Sennott, "Constrained discounted Markov decision chains", *Probability in the Engineering and Informational Sciences*, **5**, pp. 463-475, 1991.

[28] L. I. Sennott, "Constrained average cost Markov decision chains", to appear in *Probability in the Engineering and Informational Sciences.*

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399