General limit value in Dynamic Programming

Jérôme Renault[*]

Revised version, May 2013

Abstract

We consider a dynamic programming problem with arbitrary state space and bounded rewards. Is it possible to uniquely define a limit value for the problem, when the "patience" of the decision-maker tends to infinity ? We consider, for each evaluation θ (a probability distribution over positive integers) the value function v_θ of the problem where the weight of any stage t is given by θ_t, and we investigate the uniform convergence of a sequence $(v_{\theta^k})_k$ when the "impatience" of the evaluations vanishes, in the sense that $\sum_t |\theta_t^k - \theta_{t+1}^k| \to_{k \to \infty} 0$. We prove that this uniform convergence happens if and only if the metric space $\{v_{\theta^k}, k \geq 1\}$ is totally bounded. Moreover there exists a particular function v^*, independent of the particular chosen sequence $(\theta^k)_k$, such that any limit point of such sequence of value functions is precisely v^*. The result applies in particular to discounted payoffs when the discount factor vanishes, as well as to average payoffs where the number of stages goes to infinity, and extends to models with stochastic transitions.

Keywords: dynamic programming, average payoffs, discounted payoffs, general evaluations, limit value, vanishing impatience, uniform convergence of the values.

1 Introduction

In a dynamic programming problem with arbitrary state space Z and bounded payoffs, is it possible to define a unique possible limit value when the "patience" of the decision-maker tends to infinity ?

For each evaluation (probability distribution over positive integers) $\theta = (\theta_t)_{t \geq 1}$, we consider the value function v_θ of the problem where the initial state is arbitrary in Z and the weight of any stage t is given by θ_t. The *total variation* of θ, that we also call the *impatience* of θ, is defined by $TV(\theta) = \sum_{t=1}^\infty |\theta_{t+1} - \theta_t|$. For instance, for each positive integer n the evaluation $\theta = (1/n, ..., 1/n, 0, ..., 0, ...)$ induces the

value function $\bar{v}_n$ corresponding to the maximization of the mean payoff for the first n stages; and for any λ in $(0, 1]$ the evaluation $\theta = (\lambda(1-\lambda)^{t-1})_t$ induces the discounted value function v_λ.

A well known theorem of Hardy and Littlewood (see e.g. Lippman, 1969) implies that for an *uncontrolled* problem, the pointwise convergence of $(\bar{v}_n)_n$, when n goes to infinity, and of $(v_\lambda)_\lambda$, when λ goes to 0, are equivalent, and that in case of convergence both limits are the same. However, Lehrer and Sorin (1992) provided an example of a dynamic programming problem where $(\bar{v}_n)_n$ and $(v_\lambda)_\lambda$ have different pointwise limits. But they also proved that the *uniform* convergence of $(\bar{v}_n)_n$ and of $(v_\lambda)_\lambda$ are equivalent, with equality of the limit in case of convergence. And Sorin and Monderer (1993) extended this result to families of evaluations satisfying some conditions. Mertens and Neyman (1982) proved that when the family $(v_\lambda)_\lambda$ uniformly converges and has bounded variation, then the dynamic programming problem has a *uniform value*, in the sense that for all initial state z and $\varepsilon > 0$, there exists a play with mean payoffs from stage 1 to stage T at least $v - \varepsilon$ provided T is large enough (see also Lehrer Monderer 1994 and Sorin Moderer 1993 for proofs that the uniform convergence of $(v_\lambda)_\lambda$ or $(\bar{v}_n)_n$ does not imply the existence of the *uniform value* of the problem). And if the uniform value exists, one can show that all value functions v_θ are close to v^*, whenever θ is a *non increasing* evaluation with small θ_1. The reason is that whenever θ is non increasing, the θ-payoff of a play can be expressed as a convex combination of the Cesàro values $(\bar{v}_n)_n$.

In the present paper, we investigate the uniform convergence of sequences $(v_{\theta^k})_k$ when the "impatience" of the evaluations vanishes, in the sense that $\sum_t |\theta_t^k - \theta_{t+1}^k| \to_{k \to \infty} 0$. We will prove in Theorem 2.5 that this uniform convergence happens if and only if the metric space $\{v_{\theta^k}, k \geq 1\}$ (with the distance between functions given by the sup of their differences), is totally bounded. Moreover the uniform limit, whenever it exists, can only be the following function, which is independent of the particular chosen sequence $(\theta^k)_k$:

$$v^* = \inf_{\theta \in \Theta} \sup_{m \geq 0} v_{m,\theta},$$

where for each evaluation $\theta = (\theta_t)_{t \geq 1}$, the function $v_{m,\theta}$ is the value corresponding to the evaluation with weight 0 for the first m stages and with weight θ_{t-m} for stages $t > m$. Consequently, while speaking of uniform convergence of the value functions when the impatience of the decision-maker goes to zero, v^* can be considered as the unique possible limit value. We also give simple conditions on the state space, the payoffs and the transitions (mainly compactness, continuity and non expansiveness) implying the uniform convergence of such value functions.

The paper is organized as follows: section 2 contains the model and the main results, which are shown to extend to the case of stochastic transitions. Section 3 contains a few examples and counterexamples and section 4 contains the proof

of Theorem 2.5. In the last section we formulate the following conjecture, which is shown to be true for uncontrolled problems: does the uniform convergence of $(\bar{v}_n)_n$, or equivalently of $(v_\lambda)_\lambda$, implies the general convergence of the value functions, in the sense that: $\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta$ s.t. $TV(\theta) \leq \alpha, \ \|v_\theta - v^*\| \leq \varepsilon$?

2 Model and results

2.1 General values in dynamic programming problems

We consider a dynamic programming problem given by a non empty set of states Z, a correspondence F with non empty values from Z to Z, and a mapping r from Z to $[0, 1]$. Z is called the set of states, F is the transition correspondence and r is the reward (or payoff) function. An initial state z_0 in Z defines the following dynamic programming problem: a decision-maker, also called player, first has to select a new state z_1 in $F(z_0)$, and is rewarded by $r(z_1)$. Then he has to choose z_2 in $F(z_1)$, has a payoff of $r(z_2)$, etc... Intuitively, the decision-maker is interested in maximizing his "long-term" payoffs, for whatever it means. From now on we fix $\Gamma = (Z, F, r)$, and for every state z_0 we denote by $\Gamma(z_0) = (Z, F, r, z_0)$ the corresponding problem with initial state z_0. For z_0 in Z, a play at z_0 is a sequence $s = (z_1, ..., z_t, ...) \in Z^\infty$ such that: $\forall t \geq 1, z_t \in F(z_{t-1})$. We denote by $S(z_0)$ the set of plays at z_0, and by $S = \cup_{z_0 \in Z} S(z_0)$ the set of all plays. The set of bounded functions from Z to $\mathbb{R}$ is denoted by $\mathcal{V}$, and for v and v' in $\mathcal{V}$ we use the distance $d_\infty(v, v') = \sup_{z \in Z} |v(z) - v'(z)|$.

Cesàro values. For $n \geq 1$ and $s = (z_t)_{t \geq 1} \in S$, the average payoff of the play s up to stage n is defined by: $\bar{\gamma}_n(s) = \frac{1}{n} \sum_{t=1}^n r(z_t)$. And the n-stage average value of $\Gamma(z_0)$ is: $\bar{v}_n(z_0) = \sup_{s \in S(z_0)} \bar{\gamma}_n(s)$. By the Bellman-Shapley recursive formula, for all n and z we have: $n \, \bar{v}_n(z) = \sup_{z' \in F(z)} (r(z') + (n-1) \, \bar{v}_{n-1}(z'))$. We also have $|\bar{v}_n(z) - \sup_{z' \in F(z)} \bar{v}_n(z')| \leq \frac{2}{n}$, and a pointwise limit of $(\bar{v}_n)_n$ should satisfy $v(z) = \sup_{z' \in F(z)} v(z')$ for all z.

Discounted values. Given $\lambda \in (0, 1]$, the λ-discounted payoff of a play $s = (z_t)_{t \geq 1}$ is $\gamma_\lambda(s) = \lambda \sum_{t=1}^\infty (1 - \lambda)^{t-1} r(z_t)$, and the λ-discounted value at the initial state z_0 is $v_\lambda(z_0) = \sup_{s \in S(z_0)} \gamma_\lambda(s)$. It is easily proved that v_λ is the unique mapping in $\mathcal{V}$ satisfying the fixed point equation : $\forall z \in Z, v_\lambda(z) = \sup_{z' \in F(z)} (\lambda \, r(z') + (1 - \lambda) \, v_\lambda(z'))$. It implies $|v_\lambda(z) - \sup_{z' \in F(z)} v_\lambda(z')| \leq \lambda$, and a pointwise limit of $(v_\lambda)_\lambda$ should also satisfy $v(z) = \sup_{z' \in F(z)} v(z')$ for all z.

General values. We denote by Θ the set of probability distributions over positive integers. An element $\theta = (\theta_t)_{t \geq 1}$ in Θ is called an evaluation.

Definition 2.1.

The θ-payoff of a play $s = (z_t)_{t \geq 1}$ is $\gamma_\theta(s) = \sum_{t=1}^\infty \theta_t r(z_t)$,

and the θ-value of $\Gamma(z_0)$ is $v_\theta(z_0) = \sup\limits_{s \in S(z_0)} \gamma_\theta(s)$.

For each stage t we denote by δ_t the Dirac mass on stage t, so that an evaluation $\theta = (\theta_t)_{t \geq 1}$ is also written $\theta = \sum_{t=1}^\infty \theta_t \delta_t$. We denote by $\bar{n}$ the Cesàro evaluation $(1/n, ..., 1/n, 0, ..., 0, ...) = (1/n) \sum_{t=1}^n \delta_t$, so that the notation v_θ for $\theta = \bar{n}$ coincides with the Cesàro value $\bar{v}_n = v_{\bar{n}}$ (this value is often simply denoted v_n, but we will avoid this notation here to prevent any confusion with the value function with evaluation δ_n).

It is easy to see that for each evaluation θ, the Bellman recursive formula can be written as follows:

$$v_\theta(z) = \sup_{z' \in F(z)} \left(\theta_1 r(z') + (1 - \theta_1) \, v_{\theta^+}(z') \right),$$

where if $\theta_1 < 1$, the "shifted" evaluation θ^+ is defined as $(\frac{\theta_{t+1}}{1 - \theta_1})_{t \geq 1}$.

Lemma 2.2. *For all evaluations θ in Θ and all states z in Z,*

$$|v_\theta(z) - \sup_{z' \in F(z)} v_\theta(z')| \leq \theta_1 + \sum_{t \geq 2} |\theta_t - \theta_{t-1}|.$$

Proof: Consider any $z_1 \in F(z)$, and for $\varepsilon > 0$ a play $s = (z_2, z_3, ...,)$ in $S(z_1)$ such that $\gamma_\theta(s) \geq v_\theta(z_1) - \varepsilon$. We have:

$$
\begin{aligned}
v_\theta(z) &\geq \theta_1 r(z_1) + \sum_{t=2}^\infty \theta_t r(z_t) \\
&\geq \theta_1 r(z_1) + \sum_{t=2}^\infty \theta_{t-1} r(z_t) + \sum_{t=2}^\infty (\theta_t - \theta_{t-1}) r(z_t) \\
&\geq v_\theta(z_1) - \varepsilon - \sum_{t=2}^\infty |\theta_t - \theta_{t-1}|.
\end{aligned}
$$

Conversely, choose $s = (z_1, z_2, ...)$ in $S(z)$ such that $\gamma_\theta(s) \geq v_\theta(z) - \varepsilon$.

$$
\begin{aligned}
v_\theta(z) &\leq \varepsilon + \theta_1 r(z_1) + \sum_{t=2}^\infty \theta_{t-1} r(z_t) + \sum_{t=2}^\infty (\theta_t - \theta_{t-1}) r(z_t) \\
&\leq \varepsilon + \theta_1 + v_\theta(z_1) + \sum_{t=2}^\infty |\theta_t - \theta_{t-1}|.
\end{aligned}
$$

$\square$

Definition 2.3. *The total variation of an evaluation $\theta = (\theta_t)_{t \geq 1}$ is*

$$TV(\theta) = \sum_{t=1}^\infty |\theta_{t+1} - \theta_t|.$$

We have $\sup_t \theta_t \leq TV(\theta) \leq 2$. In the case of a Cesàro evaluation $\theta = (1/n, ..., 1/n, 0, 0, ...)$, we have $TV(\theta) = 1/n$. For a discounted evaluation $\theta = (\lambda(1 - \lambda)^{t-1})_{t \geq 1}$, we have $TV(\theta) = \lambda$. A small $TV(\theta)$ corresponds to a patient evaluation, and sometimes we will refer to $TV(\theta)$ as the *impatience* of θ. We will consider here limits when $TV(\theta)$ goes to zero, generalizing the cases where $n \longrightarrow \infty$ or $\lambda \longrightarrow 0$. Notice that if an evaluation θ is non increasing, i.e. satisfies $\theta_{t+1} \leq \theta_t$ for all t, we have that $TV(\theta) = \theta_1$. In the case of a sequence of non increasing evaluations $(\theta^k)_k$, the condition $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$ is equivalent to the condition $\sup_{t \geq 1} \theta_t^k \xrightarrow[k \to \infty]{} 0$.

We always have:

$$(1 - \theta_1) \sum_{t \geq 1} |\theta_t - \theta_t^+| \leq \theta_1 + TV(\theta),$$

so if $TV(\theta)$ is small, the L^1-distance between θ and the shifted evaluation θ^+ is also small. Notice also the following inequalities.

For any given T, denote by $\bar{\theta}(T)$ the arithmetic mean of $\theta_1, ..., \theta_T$. We have for all $t = 1, ..., T$:

$$|\theta_t - \bar{\theta}(T)| \leq \sum_{t'=1}^{T-1} |\theta_{t'} - \theta_{t'+1}| \leq TV(\theta).$$

So if $TV(\theta)$ is small, then for all T and $t \leq T$, the weight θ_t is close to the average $\bar{\theta}(T)$.

Given an evaluation θ and $m \geq 0$, we write $v_{m,\theta}$ for the value function associated to the evaluation $\theta' = \sum_{t=1}^{\infty} \theta_t \delta_{m+t}$. Lemma 2.2 implies that $d_\infty(v_\theta, v_{1,\theta}) \leq \theta_1 + \sum_{t \geq 2} |\theta_t - \theta_{t-1}| \leq 2TV(\theta)$. The following function will play a very important role in the sequel:

Definition 2.4. *Define for all z in Z,*

$$v^*(z) = \inf_{\theta \in \Theta} \sup_{m \geq 0} v_{m,\theta}(z).$$

2.2 Main results

We now state the main result of this paper. Recall that a metric space is totally bounded (or precompact) if for all $\varepsilon > 0$ it can be covered by finitely many balls with radius ε.

Theorem 2.5. *Let $(\theta^k)_{k \geq 1}$ be a sequence of evaluations such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$. We have for all z in Z:*

$$v^*(z) = \inf_{k \geq 1} \sup_{m \geq 0} v_{m,\theta^k}(z).$$

5

Moreover, the sequence $(v_{\theta^k})_k$ uniformly converges if and only if the metric space $(\{v_{\theta^k}, k \geq 1\}, d_\infty)$ is totally bounded. And in case of convergence, the limit value is v^*.

This theorem generalizes Theorem 3.10 in Renault, 2011, which was only dealing with Cesàro evaluations[1]. In particular, there is a unique possible limit point for all sequences $(v_{\theta^k})_k$ such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$, and consequently any (uniform) limit of such sequence is v^*. Notice that this is not true if we replace uniform convergence by pointwise convergence: even for uncontrolled problems, it may happen that several limit points are possible. As an immediate corollary of Theorem 2.5, when Z is finite the sequence $(v_{\theta^k})_k$ is bounded and has a unique limit point, so it converges to v^*.

Corollary 2.6. *Assume that Z is endowed with a distance d such that: a) (Z, d) is a precompact metric space, and b) the family $(v_\theta)_{\theta \in \Theta}$ is uniformly equicontinuous. Then there is general uniform convergence of the value functions to v^*, i.e.*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta \text{ s.t. } TV(\theta) \leq \alpha, \ \|v_\theta - v^*\| \leq \varepsilon.$$

The proof of Corollary 2.6 from Theorem 2.5 follows from 1) Ascoli's theorem, and 2) the fact that the convergence of $(v_{\theta^k})_k$ to v^* for each sequence of evaluations such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$ is enough to have the general uniform convergence of the value functions to v^*.

Whenever Z is endowed with a distance d, we say that F is non expansive (with respect to d) if: $\forall z \in Z, \forall z' \in Z, \forall z_1 \in F(z), \exists z_1' \in F(z')$ s.t. $d(z_1, z_1') \leq d(z, z')$.

Corollary 2.7. *Assume that Z is endowed with a distance d such that: a) (Z, d) is a precompact metric space, b) r is uniformly continuous, and c) F is non expansive. Then we have the same conclusions as Corollary 2.6, there is general uniform convergence of the value functions to v^*, i.e.*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta \text{ s.t. } TV(\theta) \leq \alpha, \ \|v_\theta - v^*\| \leq \varepsilon.$$

Proof of Corollary 2.7. One can proceed as in the proof of corollary 3.9 in Renault, 2011. Given two states z and z', one can construct inductively from each play $s = (z_t)_{t \geq 1}$ at z a play $s' = (z_t')_{t \geq 1}$ at z' such that $d(z_t, z_t') \leq d(z, z')$ for all t. Regarding payoffs, we introduce the modulus of continuity $\hat{\varepsilon}$ of r by:
$$\hat{\varepsilon}(\alpha) = \sup_{z, z' \text{s.t.} d(z, z') \leq \alpha} |r(z) - r(z')| \text{ for each } \alpha \geq 0.$$
So $|r(z) - r(z')| \leq \hat{\varepsilon}(d(z, z'))$ for each pair of states z, z', and $\hat{\varepsilon}$ is continuous at 0. Using the previous construction, we obtain that for z and z' in Z, for all $k \geq 1$, $|v_{\theta^k}(z) - v_{\theta^k}(z')| \leq \hat{\varepsilon}(d(z, z'))$. In particular, the family $(v_{\theta^k})_{k \geq 1}$ is uniformly equicontinuous, and Corollary 2.6 gives the result. $\qquad\square$

---

[1]In this paper it is also proved that if the Cesàro values $(v_{\bar{n}})_n$ uniformly converge then the limit can only be $\inf_{n \geq 1} \sup_{m \geq 0} v_{m, \bar{n}}$ which in this case is also equal to $\sup_{m \geq 0} \inf_{n \geq 1} v_{m, \bar{n}}$.

A completely different proof of Corollary 2.7, with another expression for the limit value v^*, can be found in Theorem 3.9 of Renault Venel 2012.

2.3 Extension to stochastic transitions

We generalize here Theorem 2.5 to the case of stochastic transitions. We will only consider transitions with finite support, and given a set X we denote by $\Delta_f(X)$ the set of probabilities with finite support over X. We consider now stochastic dynamic programming problems of the following form. There is an arbitrary non empty set of states X, a transition given by a multi-valued mapping $F : X \rightrightarrows \Delta_f(X)$ with non empty values, and a payoff (or reward) function $r : X \to [0,1]$. The interpretation is that given an initial state x_0 in X, a decision-maker has to choose a probability with finite support u_1 in $F(x_0)$, then x_1 is selected according to u_1 and there is a payoff $r(x_1)$. Then the player has to select u_2 in $F(x_1)$, x_2 is selected according to u_1 and the player receives the payoff $r(x_2)$, etc...

Following Maitra and Sudderth (1996), we say that $\Gamma = (X, F, r)$ is a Gambling House. We identify an element x in X with its Dirac measure δ_x in $\Delta(X)$, we write $Z = \Delta_f(X)$ and an element in Z is written $u = \sum_{x \in X} u(x)\delta_x$. In case the values of F only consist of Dirac measures on X, we are in the previous case of a dynamic programming problem.

We linearly extend r and F to $\Delta_f(X)$ by defining for each u in Z, the payoff $r(u) = \sum_{x \in X} r(x)u(x)$ and the transition $F(u) = \{\sum_{x \in X} u(x)f(x), s.t. \ f : X \to Z$ and $f(x) \in F(x) \ \forall x \in X\}$. A play at x_0 is a sequence $\sigma = (u_1, ..., u_t, ...) \in Z^\infty$ such that $u_1 \in F(x_0)$ and $u_{t+1} \in F(u_t)$ for each $t \geq 1$, and we denote by $\Sigma(x_0)$ the set of plays at x_0. Given an evaluation θ, the θ-payoff of a play $\sigma = (u_1, ..., u_t, ...)$ is defined as: $\gamma_\theta(\sigma) = \sum_{t \geq 1} \theta_t r(u_t)$, and the θ-value at x_0 is:

$$v_\theta(x_0) = \sup_{\sigma \in \Sigma(x_0)} \gamma_\theta(\sigma).$$

v_θ is by definition a mapping from X to $[0, 1]$, and we define as before, for all x in X:

$$v^*(x) = \inf_{\theta \in \Theta} \sup_{m \geq 0} v_{m,\theta}(x).$$

Theorem 2.5 easily extends to this context.

Theorem 2.8. *Let $(\theta^k)_{k \geq 1}$ be a sequence of evaluations with vanishing total variation, i.e. such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$. We have:*

$$\forall x \in X, \ \ v^*(x) = \inf_{k \geq 1} \sup_{m \geq 0} v_{m,\theta^k}(x).$$

Moreover, the sequence $(v_{\theta^k})_k$ uniformly converges if and only if the metric space $(\{v_{\theta^k}, k \geq 1\}, d_\infty)$ is totally bounded. And in case of convergence, the limit value is v^*.

Proof. Consider the deterministic dynamic programming problem $\Gamma = (Z, F, r)$. For any evaluation θ, the associated θ-value function $\tilde{v}_\theta : Z \longrightarrow [0,1]$ is the affine extension of $v_\theta : X \longrightarrow [0,1]$. We put, as in Definition 2.4, for all z in Z:

$$\tilde{v}^*(z) = \inf_{\theta \in \Theta} \sup_{m \geq 0} \tilde{v}_{m,\theta}(z).$$

Notice that as an "infsup" of affine functions, there is no reason a priori for $\tilde{v}^*$ to be affine. However, the restriction of $\tilde{v}^*$ to X is v^*.

Consider now a sequence $(\theta^k)_{k \geq 1}$ of evaluations with vanishing total variation. Applying Theorem 2.5 to Γ, we first obtain that for all x in X:

$$v^*(x) = \inf_{k \geq 1} \sup_{m \geq 0} v_{m,\theta^k}(x).$$

Moreover, given two evaluations θ and θ', we have (using the same notation d_∞ for the distances on $[0,1]^X$ and on $[0,1]^Z$):

$$
\begin{aligned}
d_\infty(\tilde{v}_\theta, \tilde{v}_{\theta'}) &= \sup_{z \in Z} |\tilde{v}_\theta(z) - \tilde{v}_{\theta'}(z)|, \\
&= \sup_{z \in Z} \left| \int_{p \in X} v_\theta(p) - v_{\theta'}(p) du(p) \right|, \\
&= d_\infty(v_\theta, v_{\theta'}).
\end{aligned}
$$

Consequently, $(\{v_{\theta^k}, k \geq 1\}, d_\infty)$ is totally bounded if and only if $(\{\tilde{v}_{\theta^k}, k \geq 1\}, d_\infty)$ is, and this completes the proof. $\qquad \square$.

3 Examples

The first very simple example shows that, even when the set of states is finite, it is not possible to obtain the conclusions of Theorem 2.5 or Corollaries 2.6 and 2.7 with sequences of evaluations satisfying the weaker convergence condition: $\sup_{t \geq 1} \theta_t^k \longrightarrow_{k \to \infty} 0$.

Example 3.1. Consider the following dynamic programming problem with 2 states: $Z = \{z_0, z_1\}$, $F(z_0) = \{z_1\}$, $F(z_1) = \{z_0\}$, with payoffs $r(z_0) = 0$ and $r(z_1) = 1$. We have a deterministic Markov chain, so that any play alternates forever between z_0 and z_1. Define for each k the evaluations $\theta^k = \frac{1}{k} \sum_{t=1}^{k} \delta_{2t-1}$ and $\theta'^k = \frac{1}{k} \sum_{t=1}^{k} \delta_{2t}$. We have $v_{\theta^k}(z_0) = v_{\theta'^k}(z_1) = 1$, and $v_{\theta^k}(z_1) = v_{\theta'^k}(z_0) = 0$ for all k. Define now ν^k as θ^k when k is even, and θ'^k when k is odd. The evaluation ν^k satisfies $\sup_t \nu_t^k = \frac{1}{k} \longrightarrow_{k \to \infty} 0$, however $(v_{\nu^k}(z_0))_k$ and $(v_{\nu^k}(z_1))_k$ do not converge. $\qquad \square$

Lehrer and Sorin (1992) proved that the uniform convergence of the Cesàro values $(\bar{v}_n)_{n \geq 1}$ was equivalent to the uniform convergence of the discounted values $(v_\lambda)_{\lambda \in (0,1]}$. The following example shows that this property does not extend to general evaluations: given 2 sequences of TV-vanishing evaluations $(\theta^k)_{k \geq 1}$ and $(\theta'^k)_{k \geq 1}$, the uniform convergence of $(v_{\theta^k})_k$ and $(v_{\theta'^k})_k$ are not equivalent.

Example 3.2. In this example, $(\bar{v}_n)_n$ will pointwise converge to the constant $1/2$ whereas for a particular sequence of evaluations $(\theta^k)_k$ with total variation going to zero, we will have $v_{\theta^k}(z) = 1$ for all k and z.

We construct a dynamic programming problem defined via a rooted tree T without terminal nodes (as in Sorin Monderer 1992 or Lehrer Monderer 1994). T has countably many nodes, and the payoff attached to each node is either 0 or 1.

We first construct a tree T_1, with countably many nodes and root z_0. Each node has an outgoing degree one, except the root which has countably many potential successors $z_1, z_2,..., z_n...$ On the nth branch starting from z_n, each node has a unique successor and the payoffs starting from z_n are successively 0 for n stages, then 1 for n stages, then 0 until the end of the play.
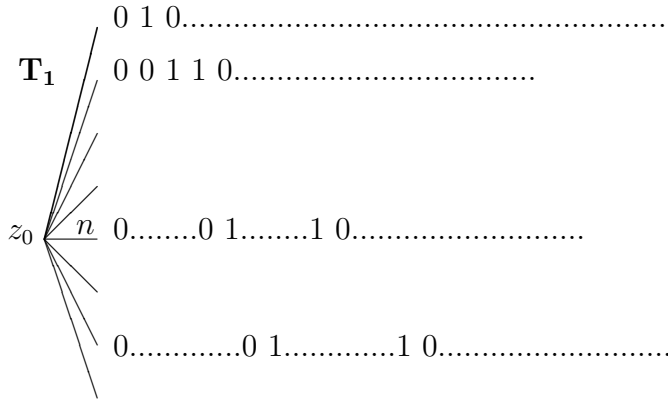
We now define T inductively from T_1. T_2 is obtained from T_1 by attaching the tree T_1 to each node of $T_1\backslash\{z_0\}$. This means that for each node z of $T_1\backslash\{z_0\}$ we add a copy of the tree T_1 where z plays the role of the root of T_1. And for each l, the tree T_l is obtained by attaching the tree T_1 to each node of $T_{l-1}\backslash T_{l-2}$. Finally, T is defined as the union $\bigcup_{l\geq 1} T_l$.

Starting from z_0, any sequence of n consecutive payoffs of 1 has to be preceeded by n consecutive payoffs of 0, so $\bar{v}_n(z_0) \leq 1/2$ for each $n \geq 1$, and for each node z and even integer n it is possible to get exactly $n/2$ payoffs of 0 followed by $n/2$ payoffs of 1. Consequently one can deduce that $(\bar{v}_n(z))_n$ converges to $1/2$ for each state z. But $\sup_{z\in Z} \bar{v}_n(z) = 1$ for each n, and the convergence is not uniform.

Consider now for any k, the evaluation $\theta^k = (0, ...0, \frac{1}{K}, ..., \frac{1}{K}, 0, ...) = \frac{1}{K}\sum_{t=1}^{K} \delta_{t+K}$. We have $v_{\theta^k}(z) = 1$ for all k and z, so $(v_{\theta^k})_k$ uniformly converges to $v^* = 1$. $\quad\square$

Example 3.3. The condition $(\{v_\theta, \theta \in \Theta\}, d_\infty)$ totally bounded is satisfied with the hypotheses of Corollary 2.6 or Corollary 2.7, and is sufficient to obtain the general uniform convergence of the value functions. This condition turns out to be stronger than having $(\{v_{\theta^k}, k \geq 1\}, d_\infty)$ totally bounded for every sequence of evaluations with vanishing TV.

In the following example, there is no control and the state space Z is the set of all integers, with transition given by the shift: $F(z) = \{z+1\}$. The payoffs are given by $r(0) = 1$ and $r(z) = 0$ for all $z \neq 0$.

For all evaluations $\theta = (\theta_t)_{t \geq 1}$, we have $\sup_{z \in Z} v_\theta(z) = \sup_t \theta_t$, so we have general uniform convergence of the value functions to $v^* = 0$.

For all positive t, we can consider the evaluation given by the Dirac measure on t. We have $v_{\delta_t}(-t) = 1$, and $v_{\delta_t}(z) = 0$ if $z \neq -t$. The set $\{v_{\delta_t}, t \geq 1\}$ is not totally bounded. $\qquad\square$

4 Proof of Theorem 2.5

We start with a few notations and definitions. We define inductively a sequence of correspondences $(F^n)_n$ from Z to Z, by $F^0(z) = \{z\}$ for every state z, and $\forall n \geq 0$, $F^{n+1} = F^n \circ F$ (the composition being defined by $G \circ H(z) = \{z" \in Z, \exists z' \in H(z), z" \in G(z')\}$). $F^n(z)$ represents the set of states that the decision-maker can reach in n stages from the initial state z. We also define for every state z, $G^m(z) = \bigcup_{n=0}^{m} F^n(z)$ and $G^\infty(z) = \bigcup_{n=0}^{\infty} F^n(z)$. The set $G^\infty(z)$ is the set of states that the decision-maker, starting from z, can reach in a finite number of stages.

For all θ in Θ, $m \geq 0$ and initial state z, we clearly have:

$$v_{m,\theta}(z) = \sup_{z' \in F^m(z)} v_\theta(z') = \sup_{s \in S(z)} \sum_{t=1}^{\infty} \theta_t r(z_{m+t}).$$

In the sequel, we fix a sequence of evaluations $(\theta^k)_{k \geq 1}$ such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$.

Lemma 4.1. *For all $m_0 \geq 0$ and z in Z,*

$$\liminf_k \sup_{m \leq m_0} v_{m,\theta^k}(z) = \liminf_k v_{\theta^k}(z).$$

Proof: For each k, we have $v_{\theta^k}(z) \geq v_{1,\theta^k}(z) - \theta_1^k - TV(\theta^k)$ by Lemma 2.2, so $v_{\theta^k}(z) \geq v_{1,\theta^k}(z) - 2TV(\theta^k)$. Iterating, we obtain that:
$v_{\theta^k}(z) \geq \sup_{m \leq m_0} v_{m,\theta^k}(z) - 2m_0 TV(\theta^k)$. $\qquad\square$

A key result is the following proposition, which is true for all evaluations θ.

Proposition 4.2. *For all evaluations θ in Θ and initial state z in Z,*

$$\sup_{z' \in G^\infty(z)} v_\theta(z') \geq \limsup_k v_{\theta^k}(z).$$

Proof of Proposition 4.2 z and θ being fixed, put $\beta = \sup_{z' \in G^\infty(z)} v_\theta(z')$. Fix $\varepsilon \in (0,1]$, there exists T_0 such that $\sum_{t=T_0+1}^{\infty} \theta_t \leq \varepsilon$, and fix $T_1 \geq T_0/\varepsilon$.

For any play $s = (z_1, ..., z_t, ...)$ in $S(z)$, we have by definition of β that for all T, $\sum_{t=T+1}^{\infty} \theta_{t-T} r(z_t) \leq \beta$. Let m be a non negative integer, we define:

$$A_m = \sum_{T=mT_1}^{(m+1)T_1-1} \sum_{t=T+1}^{\infty} \theta_{t-T} r(z_t) \leq T_1 \beta.$$

$$
\begin{aligned}
A_m &= \sum_{t=mT_1+1}^{\infty} r(z_t) \sum_{T=mT_1}^{\min\{(m+1)T_1-1,t-1\}} \theta_{t-T}, \\
&\geq \sum_{t=mT_1+1}^{(m+1)T_1} r(z_t)(\theta_1 + ... + \theta_{t-mT_1}), \\
&\geq (1-\varepsilon) \sum_{t=T_0+mT_1}^{(m+1)T_1} r(z_t), \\
&\geq (1-\varepsilon)\left(\sum_{t=1+mT_1}^{(m+1)T_1} r(z_t) - (T_0 - 1) \right).
\end{aligned}
$$

We obtain:

$$T_1 \beta \geq (1-\varepsilon) T_1 \overline{\gamma}_{mT_1, T_1}(s) - (1-\varepsilon)(T_0 - 1),$$

and

$$\overline{\gamma}_{mT_1, T_1}(s) \leq \frac{\beta}{1-\varepsilon} + \varepsilon. \tag{1}$$

We now consider $\gamma_{\theta^k}(s)$ for k large. We compute $\sum_{t=1}^{\infty} \theta_t^k r(z_t)$ by dividing the stages into blocks of length T_1. For each $m \geq 0$, let $\overline{\theta^k}(m)$ be the Cesàro average of θ_t^k, where t ranges from $mT_1 + 1$ to $(m+1)T_1$. Notice that for all such t, we have $|\theta_t^k - \overline{\theta^k}(m)| \leq \sum_{t'=mT_1+1}^{(m+1)T_1-1} |\theta_{t'} - \theta_{t'+1}|$. We have:

$$
\begin{aligned}
\sum_{t=mT_1+1}^{(m+1)T_1} \theta_t^k r(z_t) &\leq \sum_{t=mT_1+1}^{(m+1)T_1} \overline{\theta^k}(m) r(z_t) + \sum_{t=mT_1+1}^{(m+1)T_1} |\theta_t^k - \overline{\theta^k}(m)| r(z_t), \\
&\leq \overline{\theta^k}(m) \overline{\gamma}_{mT_1, T_1}(s) + T_1 \sum_{t=mT_1+1}^{(m+1)T_1-1} |\theta_t^k - \theta_{t+1}^k|, \\
&\leq \overline{\theta^k}(m) \left(\frac{\beta}{1-\varepsilon} + \varepsilon \right) + T_1 \sum_{t=mT_1+1}^{(m+1)T_1-1} |\theta_t^k - \theta_{t+1}^k|.
\end{aligned}
$$

where the last inequality follows from Equation (1). Summing up over m, we obtain:

$$\gamma_{\theta^k}(s) \leq \frac{\beta}{1-\varepsilon} + \varepsilon + T_1\, TV(\theta^k).$$

Consequently, $\limsup_k v_{\theta^k}(z) \leq \frac{\beta}{1-\varepsilon} + \varepsilon$, and this is true for all ε. □

11

Corollary 4.3.

$$\inf_{\theta \in \Theta} \sup_{m \geq 0} v_{m,\theta} = \inf_{k \geq 1} \sup_{m \geq 0} v_{m,\theta^k}.$$

Proof: Consider an initial state z, and write $\alpha = \inf_k \sup_m v_{m,\theta^k}(z)$. It is clear that $\alpha \geq \inf_{\theta \in \Theta} \sup_{m \geq 0} v_{m,\theta}(z)$. Now for each $k \geq 1$ there exists $m(k)$, such that $v_{m(k),\theta^k}(z) \geq \alpha - 1/\overline{k}$, and we define the evaluation $\theta'^k = \sum_{t=m(k)+1}^{\infty} \theta_{t-m(k)}^k \delta_t$. We have $TV(\theta'^k) = TV(\theta^k) \xrightarrow[k \to \infty]{} 0$, so by Proposition 4.2 we obtain that for all evaluations θ, $\sup_{z' \in G^\infty(z)} v_\theta(z') \geq \limsup_k v_{\theta^{m(k)},k}(z) \geq \alpha$. $\qquad\square$

From Lemma 4.1 and Proposition 4.2, one can easily deduce the following corollary.

Corollary 4.4. *For all $m_0 \geq 0$ and z in Z,*

$$\inf_{k \geq 1} \sup_{m \leq m_0} v_{m,\theta^k}(z) \leq \liminf_k v_{\theta^k}(z) \leq \limsup_k v_{\theta^k}(z) \leq \inf_{k \geq 1} \sup_{m \geq 0} v_{m,\theta^k}(z).$$

And we can now conclude the proof of the Theorem 2.5, proceeding as in the proof of Theorem 3.10 in Renault, 2011. For all states z and z', we define $d(z,z') = \sup_{k \geq 1} |v_{\theta^k}(z) - v_{\theta^k}(z')|$. The space (Z,d) is now a pseudometric space (may not be Hausdorff).

End of the proof of Theorem 2.5. By assumption, there exists a finite set of indices I such that for all $k \geq 1$, there exists i in I satisfying $d_\infty(v_{\theta^k}, v_{\theta^i}) \leq \varepsilon$. Consider now the set $\{(v_{\theta^i}(z))_{i \in I}, z \in Z\}$, it is a subset of the compact metric space $[0,1]^I$ with the uniform distance, so it is itself precompact and we obtain the existence of a finite subset C of states in Z such that:

$$\forall z \in Z, \exists c \in C, \forall i \in I, |v_{\theta^i}(z) - v_i(c)| \leq \varepsilon.$$

We have obtained that for each $\varepsilon > 0$, there exists a finite subset C of Z such that for every z in Z, there is $c \in C$ with $d(z,c) \leq \varepsilon$. The pseudometric space (Z,d) is itself precompact. Equivalently, any sequence in Z admits a Cauchy subsequence for d. Notice that all value functions v_{θ^k} are clearly 1-Lipschitz for d.

Fix z in Z, and consider now the sequence of sets $(G^m(z))_{m \geq 0}$. For all m, $G^m(z) \subset G^{m+1}(z)$ so using the precompacity of (Z,d) it is not difficult to show (see, e.g. step 2 in the proof of Theorem 3.7 in Renault, 2011) that $(G^m(z))_{m \geq 0}$ converges to $G^\infty(z)$, in the sense that:

$$\forall \varepsilon > 0, \exists m \geq 0, \forall z' \in G^\infty(z), \exists z'' \in G^m(z), \ d(z',z'') \leq \varepsilon. \tag{2}$$

We now use Corollary 4.4 to conclude. We have for all m :

$$\inf_{k \geq 1} \sup_{z' \in G^m(z)} v_{\theta^k}(z') \leq \liminf_k v_{\theta^k}(z) \leq \limsup_k v_{\theta^k}(z) \leq \inf_{k \geq 1} \sup_{z' \in G^\infty(z)} v_{\theta^k}(z').$$

Fix finally $\varepsilon > 0$, and consider $k \geq 1$ and $m \geq 0$ given by assertion (2). Let z' in $G^\infty(z)$ be such that $v_{\theta^k}(z') \geq \sup_{z' \in G^\infty(z)} v_{\theta^k}(z') - \varepsilon$. Let z'' in $G^m(z)$ be such that $d(z', z'') \leq \varepsilon$. Since v_{θ^k} is 1-Lipschitz for d, we obtain $v_{\theta^k}(z'') \geq \sup_{z' \in G^\infty(z)} v_{\theta^k}(z') - 2\varepsilon$. Consequently, $\sup_{z' \in G^m(z)} v_{\theta^k}(z') \geq \sup_{z' \in G^\infty(z)} v_{\theta^k}(z') - 2\varepsilon$ for all k, so

$$\inf_{k \geq 1} \sup_{z' \in G^m(z)} v_{\theta^k}(z') \geq \inf_{k \geq 1} \sup_{z' \in G^\infty(z)} v_{\theta^k}(z') - 2\varepsilon.$$

We obtain $\liminf_{k \geq 1} v_{\theta^k}(z) \geq \limsup_{k \geq 1} v_{\theta^k}(z) - 2\varepsilon$, and so $(v_{\theta^k}(z))_k$ converges. Since (Z, d) is precompact and all v_{θ^k} are 1-Lipschitz, the convergence is uniform. $\square$

5   An open question

We know since Lehrer and Sorin (1992) that the uniform convergence of the Cesàro values $(v_{\bar{n}})_{n \geq 1}$ is equivalent to the uniform convergence of the discounted values $(v_\lambda)_{\lambda \in (0,1]}$. Example 3.2 shows that is possible to have no uniform convergence of the Cesàro values (or equivalently of the discounted values) but uniform convergence for a particular sequence of evaluations with vanishing TV. Could it be the case that the Cesàro values and the discounted values have the following "universal" property ?

Assuming uniform convergence of the Cesàro values, do we have general uniform convergence of the value functions, i.e. is it true that $(v_{\theta^k})_k$ uniformly converges for every sequence of evaluations $(\theta^k)_{k \geq 1}$ such that $TV(\theta^k) \xrightarrow[k \to \infty]{} 0$?

The above property is true in case of an uncontrolled problem, i.e. when the transition F is single-valued.

Proposition 5.1. *For an uncontrolled problem, the uniform convergence of the Cesàro values implies the general uniform convergence of the value functions:*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta \ s.t. \ TV(\theta) \leq \alpha, \ \|v_\theta - v^*\| \leq \varepsilon.$$

Proof: Fix $\varepsilon > 0$. By assumption there exists N such that for all states z in Z, $|\bar{v}_N(z) - v^*(z)| \leq \varepsilon$. Consider an arbitrary evaluation θ and an initial state z_0. For each positive t we denote by z_t the state reached from z_0 in t stages, we have $v_\theta(z_0) = \sum_{t=1}^\infty \theta_t r(z_t)$ and $v^*(z_0) = v^*(z_t)$ for all t.

Divide the set of stages into consecutive blocks of length N: $B^0 = \{1, ..., N\}$,..., $B^m = \{mN + 1, ..., (m+1)N\}$,... Denote by $\bar{\theta}(m)$ the mean of θ over B^m, we have $\sum_{m=0}^\infty N\bar{\theta}(m) = 1$. We also write $\bar{r}(m)$ for the mean $\frac{1}{N} \sum_{t \in B^m} r(z_t)$. We have $\bar{r}(m) = \bar{v}_N(z_{mN})$, so $|\bar{r}(m) - v^*(z_0)| \leq \varepsilon$ for all m.

Computing payoffs by blocks, we have

$$
\begin{aligned}
v_\theta(z_0) &= \sum_{m=0}^\infty \sum_{t \in B^m} \theta_t r(z_t), \\
&= \sum_{m=0}^\infty \sum_{t \in B^m} (\theta_t - \bar{\theta}(m)) r(z_t) + \sum_{m=0}^\infty N \bar{\theta}(m) \bar{r}(m).
\end{aligned}
$$

So we obtain:

$$
v_\theta(z_0) - v^*(z_0) = \sum_{m=0}^\infty \sum_{t \in B^m} (\theta_t - \bar{\theta}(m)) r(z_t) + \sum_{m=0}^\infty N \bar{\theta}(m)(\bar{r}(m) - v^*(z_0)),
$$

and

$$
|v_\theta(z_0) - v^*(z_0)| \leq \sum_{m=0}^\infty N \sum_{t \in B^m} |\theta_{t+1} - \theta_t| + \varepsilon \leq N \ TV(\theta) + \varepsilon.
$$

If $TV(\theta) \leq \frac{\varepsilon}{N}$, we get $|v_\theta(z_0) - v^*(z_0)| \leq 2\varepsilon$, hence the result. $\qquad\square$

6   References

Blackwell, D. (1962): Discrete dynamic programming. The Annals of Mathematical Statistics, 33, 719-726.

Lehrer, E. and D. Monderer (1994): Discounting versus Averaging in Dynamic Programming. Games and Economic Behavior, 6, 97-113.

Lehrer, E. and S. Sorin (1992): A uniform Tauberian Theorem in Dynamic Programming. Mathematics of Operations Research, 17, 303-307.

Lippman, S. (1969): Criterion Equivalence in Discrete Dynamic Programming. Operations Research 17, 920-923.

Maitra, A.P. and Sudderth, W.D. (1996): Discrete gambling and stochastic games, Springer Verlag.

Mertens, J-F. and A. Neyman (1981): Stochastic games. International Journal of Game Theory, 1, 39-64.

Monderer, D. and S. Sorin (1993): Asymptotic properties in Dynamic Programming. International Journal of Game Theory, 22, 1-11.

Renault, J. (2011): Uniform value in Dynamic Programming. Journal of the European Mathematical Society, vol. 13, p.309-330.

Renault, J. and X. Venel (2012): A distance for probability spaces, and long-term values in Markov Decision Processes and Repeated Games. preprint hal-00674998