On the limit values in dynamic optimization

Jérôme Renault

Université Toulouse 1, TSE-GREMAQ

Optimization, Games and Dynamics, IHP, November 29, 2011

Introduction

A dynamic programming problem $\Gamma(z_0) = (Z, F, r, z_0)$ given by a non empty set of states Z, an initial state z_0, a transition correspondence F from Z to Z with non empty values, and a reward mapping r from Z to $[0,1]$. (bounded payoffs)

A player chooses z_1 in $F(z_0)$, has a payoff of $r(z_1)$, then he chooses z_2 in $F(z_1)$, etc...
Admissible plays: $S(z_0) = \{s = (z_1, ..., z_t, ...) \in Z^\infty, \forall t \geq 1, z_t \in F(z_{t-1})\}$.

n-stage problem, for $n \geq 1$:
$v_n(z) = \sup_{s \in S(z)} \frac{1}{n} \left(\sum_{t=1}^n r(z_t)\right)$.

λ-discounted pb, for $\lambda \in (0,1]$:
$v_\lambda(z) = \sup_{s \in S(z)} \left(\lambda \sum_{t=1}^\infty (1-\lambda)^{t-1} r(z_t)\right)$.

More generally, for each proba $\theta = (\theta_t)_{t \geq 1}$ on positive integers, define the θ-value by $v_\theta(z) = \sup_{s \in S(z)} \left(\sum_{t \geq 1} \theta_t r(z_t)\right)$.

Introduction

A dynamic programming problem $\Gamma(z_0) = (Z, F, r, z_0)$ given by a non empty set of states Z, an initial state z_0, a transition correspondence F from Z to Z with non empty values, and a reward mapping r from Z to $[0,1]$. (bounded payoffs)

A player chooses z_1 in $F(z_0)$, has a payoff of $r(z_1)$, then he chooses z_2 in $F(z_1)$, etc...
Admissible plays: $S(z_0) = \{s = (z_1, ..., z_t, ...) \in Z^\infty, \forall t \geq 1, z_t \in F(z_{t-1})\}$.

n-stage problem, for $n \geq 1$:
$v_n(z) = \sup_{s \in S(z)} \frac{1}{n} \left(\sum_{t=1}^n r(z_t)\right)$.

λ-discounted pb, for $\lambda \in (0,1]$:
$v_\lambda(z) = \sup_{s \in S(z)} \left(\lambda \sum_{t=1}^\infty (1-\lambda)^{t-1} r(z_t)\right)$.

More generally, for each proba $\theta = (\theta_t)_{t \geq 1}$ on positive integers, define the θ-value by $v_\theta(z) = \sup_{s \in S(z)} \left(\sum_{t \geq 1} \theta_t r(z_t)\right)$.

We have:

$$v_n(z) = \sup_{z' \in F(z)} \left(\frac{1}{n} r(z') + \frac{n-1}{n} v_{n-1}(z') \right), \text{ so } |v_n(z) - \sup_{z' \in F(z)} v_n(z')| \leq \frac{2}{n}$$

$$v_\lambda(z) = \sup_{z' \in F(z)} \left(\lambda r(z') + (1-\lambda) v_\lambda(z') \right), \text{ so } |v_\lambda(z) - \sup_{z' \in F(z)} v_\lambda(z')| \leq \lambda$$

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

Example 0:

Limit value at z_0: $v^*(z_0) = 1/2$.

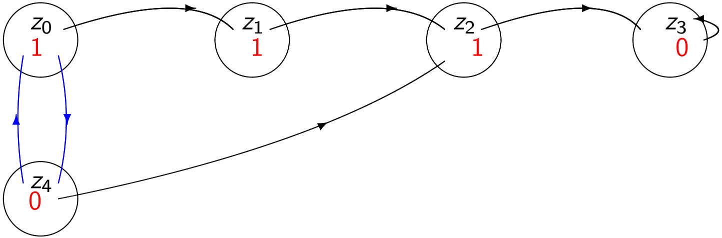
We have:

$$v_n(z) = \sup_{z' \in F(z)} \left(\frac{1}{n} r(z') + \frac{n-1}{n} v_{n-1}(z') \right), \text{ so } |v_n(z) - \sup_{z' \in F(z)} v_n(z')| \leq \frac{2}{n}$$

$$v_\lambda(z) = \sup_{z' \in F(z)} \left(\lambda r(z') + (1-\lambda) v_\lambda(z') \right), \text{ so } |v_\lambda(z) - \sup_{z' \in F(z)} v_\lambda(z')| \leq \lambda$$

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

Example 0:



Limit value at z_0: $v^*(z_0) = 1/2$.

Questions: 1) General uniform convergence: existence and equality of the uniform limits of v_n, v_λ and v_θ when the "length" becomes large: $n \to \infty$, $\lambda \to 0$, $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \to 0$?

Ex: - Cesàro $\theta = (1/n, ..., 1/n, 0, ..., 0, ...)$ with n large.
- Discounted: $\theta = (\lambda(1-\lambda)^{t-1})_{t\geq 1}$, with $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| = \lambda$ small.
- $\theta = (\theta_t)_{t\geq 1}$, with $\theta_{t+1} \leq \theta_t$ and $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| = \theta_1$ small.
- Shifted Cesàro: $\theta = (0, ..., 0, 1/n, ...1/n, 0, ..., 0, ...)$ with arbitrary many early zeros, and n large.

Say that there is general uniform CV if : for each $\varepsilon > 0$ there exists $\alpha > 0$ such that if $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Characterization of the limit v^* ?

0 player (ie. F single-valued): $(v_n(z))_n$ converges iif $(v_\lambda(z))_\lambda$ converges, and in case of CV both limits are the same (Hardy-Littlewood).

1-player: $\lim_{n\to\infty} v_n(z)$ and $\lim_{\lambda\to 0} v_\lambda(z)$ may exist and differ.
$(v_n)_n$ converges uniformly iif $(v_\lambda)_\lambda$ converges uniformly, and in case of CV both limits are the same (Lehrer-Sorin 1992). Same for particular families (v_θ) satisfying $\theta_{t+1} \leq \theta_t$ for each t + extra conditions (Sorin Monderer 1993).

Questions: 1) General uniform convergence: existence and equality of the uniform limits of v_n, v_λ and v_θ when the "length" becomes large: $n \to \infty$, $\lambda \to 0$, $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \to 0$?

Ex: - Cesàro $\theta = (1/n, ..., 1/n, 0, ..., 0, ...)$ with n large.
- Discounted: $\theta = (\lambda(1-\lambda)^{t-1})_{t\geq 1}$, with $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| = \lambda$ small.
- $\theta = (\theta_t)_{t\geq 1}$, with $\theta_{t+1} \leq \theta_t$ and $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| = \theta_1$ small.
- Shifted Cesàro: $\theta = (0, ..., 0, 1/n, ...1/n, 0, ..., 0, ...)$ with arbitrary many early zeros, and n large.

Say that there is general uniform CV if : for each $\varepsilon > 0$ there exists $\alpha > 0$ such that if $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Characterization of the limit v^* ?

0 player (ie. F single-valued): $(v_n(z))_n$ converges iif $(v_\lambda(z))_\lambda$ converges, and in case of CV both limits are the same (Hardy-Littlewood).

1-player: $\lim_{n\to\infty} v_n(z)$ and $\lim_{\lambda\to 0} v_\lambda(z)$ may exist and differ.
$(v_n)_n$ converges uniformly iif $(v_\lambda)_\lambda$ converges uniformly, and in case of CV both limits are the same (Lehrer-Sorin 1992). Same for particular families (v_θ) satisfying $\theta_{t+1} \leq \theta_t$ for each t + extra conditions (Sorin Monderer 1993).

Questions: 2) Uniform and general uniform value.
Large *unknown* horizon: when is it possible to play ε-optimally simultaneously in *any* "long" enough problem ?

Say that $\Gamma(z)$ has a (Cesàro-)uniform value if $(v_n(z))_n$ has a limit $v^*(z)$, and one can guarantee this limit: $\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists n_0$ $\forall n \geq n_0, \frac{1}{n}\left(\sum_{t=1}^n r(z_t)\right) \geq v^*(z) - \varepsilon.$

If $\Gamma(z)$ has a (Cesàro-)uniform value, it has a discounted uniform value.
The uniform CV of (v_n) does not imply the existence of the uniform value (Monderer Sorin 93, Lehrer Monderer 94).
Sufficient conditions for the existence of the uniform value given by Mertens and Neyman 1982, from stochastic games (convergence of $(v_\lambda)_\lambda$ with a BV condition).

Say that $\Gamma(z)$ has a general uniform value if $(v_\theta(z))_\theta$ has a limit $v^*(z)$, and one can guarantee this limit:
$\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists \alpha > 0,$

$$\sum_{t=1}^{\infty} \theta_t r(z_t) \geq v^*(z) - \varepsilon \text{ whenever } \sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$$

Questions: 2) Uniform and general uniform value.
Large *unknown* horizon: when is it possible to play ε-optimally simultaneously in *any* "long" enough problem ?

Say that $\Gamma(z)$ has a (Cesàro-)uniform value if $(v_n(z))_n$ has a limit $v^*(z)$, and one can guarantee this limit: $\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists n_0$ $\forall n \geq n_0, \frac{1}{n}(\sum_{t=1}^{n} r(z_t)) \geq v^*(z) - \varepsilon$.

If $\Gamma(z)$ has a (Cesàro-)uniform value, it has a discounted uniform value.
The uniform CV of (v_n) does not imply the existence of the uniform value (Monderer Sorin 93, Lehrer Monderer 94).
Sufficient conditions for the existence of the uniform value given by Mertens and Neyman 1982, from stochastic games (convergence of $(v_\lambda)_\lambda$ with a BV condition).

Say that $\Gamma(z)$ has a general uniform value if $(v_\theta(z))_\theta$ has a limit $v^*(z)$, and one can guarantee this limit:
$\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists \alpha > 0,$

$$\sum_{t=1}^{\infty} \theta_t r(z_t) \geq v^*(z) - \varepsilon \text{ whenever } \sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$$

Questions: 2) Uniform and general uniform value.
Large *unknown* horizon: when is it possible to play ε-optimally simultaneously in *any* "long" enough problem ?

Say that $\Gamma(z)$ has a (Cesàro-)uniform value if $(v_n(z))_n$ has a limit $v^*(z)$, and one can guarantee this limit: $\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists n_0$ $\forall n \geq n_0, \frac{1}{n}(\sum_{t=1}^n r(z_t)) \geq v^*(z) - \varepsilon.$

If $\Gamma(z)$ has a (Cesàro-)uniform value, it has a discounted uniform value. The uniform CV of (v_n) does not imply the existence of the uniform value (Monderer Sorin 93, Lehrer Monderer 94).
Sufficient conditions for the existence of the uniform value given by Mertens and Neyman 1982, from stochastic games (convergence of $(v_\lambda)_\lambda$ with a BV condition).

Say that $\Gamma(z)$ has a general uniform value if $(v_\theta(z))_\theta$ has a limit $v^*(z)$, and one can guarantee this limit:
$\forall \varepsilon > 0, \exists s = (z_1, ..., z_t, ...) \in S(z), \exists \alpha > 0,$

$$\sum_{t=1}^{\infty} \theta_t r(z_t) \geq v^*(z) - \varepsilon \text{ whenever } \sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$$

2. Examples

3. General Results

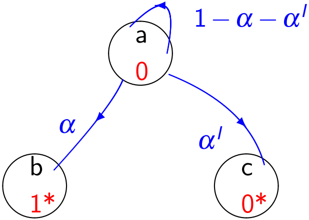
3.a) The auxiliary functions $v_{m,n}$ and uniform CV of (v_n)
3.b) Uniform convergence for (v_θ)
3.c) The auxiliary functions $w_{m,n}$ and existence of the uniform value
3.d) The compact non expansive case: characterizing the limit value v^*(with X. Venel)
3.e) On computing v^* and the speed of convergence

4. Applications

4.a) Standard Markov Decision Processes with finitely many states
4.b) Non expansive control problems (with M. Quincampoix)
4.c) MDP with imperfect observation with finitely many states.
4.d) Repeated games with an informed controller

2. Examples

Ex 1: A Markov decision process
$K = \{a, b, c\}$. b and c are absorbing with payoffs 1 and 0. Start at a, choose $\alpha \in [0, 1/2]$, and move to b with proba α and to c with proba α^l, with $l > 1$.

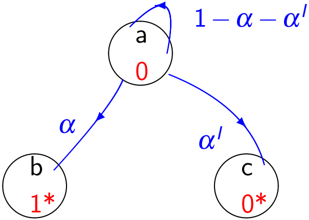


$\rightarrow$ Dynamic Programming Pb with $Z = \Delta(K)$, $r(z) = z^b$, $z_0 = \delta_a$ and $F(z) = \{(z^a(1-\alpha-\alpha^l), z^b + z^a\alpha, z^c + z^a\alpha^l), \alpha \in [0, 1/2]\}$.

The uniform value exists and $v^*(z_0) = 1$. *no ergodicity*
We have $v_\lambda(a) = 1 - C\lambda^{(l-1)/l} + o(\lambda^{(l-1)/l})$, with $C = \frac{l}{(l-1)^{\frac{l-1}{l}}}$.

2. Examples

Ex 1: A Markov decision process
$K = \{a, b, c\}$. b and c are absorbing with payoffs 1 and 0. Start at a, choose $\alpha \in [0, 1/2]$, and move to b with proba α and to c with proba α^l, with $l > 1$.



$\rightarrow$ Dynamic Programming Pb with $Z = \Delta(K)$, $r(z) = z^b$, $z_0 = \delta_a$ and $F(z) = \{(z^a(1-\alpha-\alpha^l), z^b + z^a\alpha, z^c + z^a\alpha^l), \alpha \in [0, 1/2]\}$.

The uniform value exists and $v^*(z_0) = 1$. *no ergodicity*
We have $v_\lambda(a) = 1 - C\lambda^{(l-1)/l} + o(\lambda^{(l-1)/l})$, with $C = \frac{l}{(l-1)^{\frac{l-1}{l}}}$.

Ex 2: $Z = \{z \in \mathbb{C}, |z| = 1\}$, $F(e^{i\alpha}) = e^{i(\alpha+1)}$ for all α. Then

$$v^*(z_0) = \frac{1}{2\pi}\int_0^{2\pi} r(e^{i\alpha})d\alpha$$

Ex 3: (Aumann Maschler) A finite family $(G^k)_{k\in K}$ of payoff matrices in $[0,1]^{I\times J}$, and $p \in \Delta(K)$ define a zero-sum repeated game where: first, some k is selected according to p and told to player 1 only, then G^k is repeated over and over.

$$v_n(p) = \sup_{x\in\Delta(I)^K}\left(\frac{1}{n}g(p,x) + \frac{n-1}{n}\sum_{i\in I} x(p)(i)v_{n-1}(\hat{p}(x,i))\right).$$

where $p \in \Delta(K)$, $g(p,x) = \min_j(\sum_k p^k G^k(x^k,j))$ and $\hat{p}(x,i)$ is the conditional belief on $\Delta(K)$ given p, x, i.
Can be written as a "standard" dynamic programming problem with state space $\Delta_f(\Delta(K)) \times [0,1]$.
Well known: the limit value exists. Define $u(p) = Val\left(\sum_k p^k G^k\right)$, then

$$v^* = \mathrm{cav}\, u = \inf\{v : \Delta(K) \to [0,1], v \text{ concave } v \geq u\}$$

Ex 2: $Z=\{z\in\mathbb{C},|z|=1\}$, $F(e^{i\alpha})=e^{i(\alpha+1)}$ for all α. Then

$$v^*(z_0)=\frac{1}{2\pi}\int_0^{2\pi} r(e^{i\alpha})d\alpha$$

Ex 3: (Aumann Maschler) A finite family $(G^k)_{k\in K}$ of payoff matrices in $[0,1]^{I\times J}$, and $p\in\Delta(K)$ define a zero-sum repeated game where: first, some k is selected according to p and told to player 1 only, then G^k is repeated over and over.

$$v_n(p)=\sup_{x\in\Delta(I)^K}\left(\frac{1}{n}g(p,x)+\frac{n-1}{n}\sum_{i\in I}x(p)(i)v_{n-1}(\hat{p}(x,i))\right).$$

where $p\in\Delta(K)$, $g(p,x)=\min_j(\sum_k p^k G^k(x^k,j))$ and $\hat{p}(x,i)$ is the conditional belief on $\Delta(K)$ given p, x, i.
Can be written as a "standard" dynamic programming problem with state space $\Delta_f(\Delta(K))\times[0,1]$.
Well known: the limit value exists. Define $u(p)=Val\left(\sum_k p^k G^k\right)$, then

$$v^*=\operatorname{cav}u=\inf\{v:\Delta(K)\to[0,1], v \text{ concave } v\geq u\}$$

Ex 2: $Z = \{z \in \mathbb{C}, |z| = 1\}$, $F(e^{i\alpha}) = e^{i(\alpha+1)}$ for all α. Then

$$v^*(z_0) = \frac{1}{2\pi}\int_0^{2\pi} r(e^{i\alpha})d\alpha$$

Ex 3: (Aumann Maschler) A finite family $(G^k)_{k\in K}$ of payoff matrices in $[0,1]^{I\times J}$, and $p \in \Delta(K)$ define a zero-sum repeated game where: first, some k is selected according to p and told to player 1 only, then G^k is repeated over and over.

$$v_n(p) = \sup_{x\in\Delta(I)^K}\left(\frac{1}{n}g(p,x) + \frac{n-1}{n}\sum_{i\in I} x(p)(i)v_{n-1}(\hat{p}(x,i))\right).$$

where $p \in \Delta(K)$, $g(p,x) = \min_j(\sum_k p^k G^k(x^k, j))$ and $\hat{p}(x,i)$ is the conditional belief on $\Delta(K)$ given p, x, i.
Can be written as a "standard" dynamic programming problem with state space $\Delta_f(\Delta(K)) \times [0,1]$.
Well known: the limit value exists. Define $u(p) = Val\left(\sum_k p^k G^k\right)$, then

$$v^* = \mathrm{cav}\, u = \inf\{v : \Delta(K) \to [0,1], v \text{ concave } v \geq u\}$$

3a. The auxiliary functions $v_{m,n}$ and the uniform CV of (v_n)

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t\geq 1}$, define:

$$\gamma_{m,n}(s) = \frac{1}{n}\sum_{t=1}^{n} r(z_{m+t}) \text{ and } v_{m,n}(z) = \sup_{s\in S(z)}\gamma_{m,n}(s).$$

The player first makes m moves in order to reach a "good initial state", then plays n moves for payoffs.

Write $v^-(z) = \liminf_n v_n(z)$, $v^+(z) = \limsup_n v_n(z)$,
$v^* = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(z)$.

Lemma 1: $v^-(z) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(z)$.

Lemma 2: $\forall m_0$,

$$\inf_{n\geq 1}\sup_{m\leq m_0} v_{m,n}(z) \leq v^-(z) \leq v^+(z) \leq \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(z).$$

can be restated as:

$$\inf_{n\geq 1}\sup_{z'\in G^{m_0}(z)} v_n(z') \leq v^-(z) \leq v^+(z) \leq v^*(z) = \inf_{n\geq 1}\sup_{z'\in G^{\infty}(z)} v_n(z').$$

where $G^{m_0}(z)$ is the set of states that can be reached from z in at most m_0 stages, and $G^\infty(z) = \cup_m G^m(z)$.

3a. The auxiliary functions $v_{m,n}$ and the uniform CV of (v_n)

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t \geq 1}$, define:

$$\gamma_{m,n}(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_{m+t}) \text{ and } v_{m,n}(z) = \sup_{s \in S(z)} \gamma_{m,n}(s).$$

The player first makes m moves in order to reach a "good initial state", then plays n moves for payoffs.

Write $v^-(z) = \liminf_n v_n(z)$, $v^+(z) = \limsup_n v_n(z)$,

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

Lemma 1: $v^-(z) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z)$.

Lemma 2: $\forall m_0$,

$$\inf_{n \geq 1} \sup_{m \leq m_0} v_{m,n}(z) \leq v^-(z) \leq v^+(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z).$$

can be restated as:

$$\inf_{n \geq 1} \sup_{z' \in G^{m_0}(z)} v_n(z') \leq v^-(z) \leq v^+(z) \leq v^*(z) = \inf_{n \geq 1} \sup_{z' \in G^{\infty}(z)} v_n(z').$$

where $G^{m_0}(z)$ is the set of states that can be reached from z in at most m_0 stages, and $G^{\infty}(z) = \cup_m G^m(z)$.

3a. The auxiliary functions $v_{m,n}$ and the uniform CV of (v_n)

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t \geq 1}$, define:

$$\gamma_{m,n}(s) = \frac{1}{n} \sum_{t=1}^{n} r(z_{m+t}) \text{ and } v_{m,n}(z) = \sup_{s \in S(z)} \gamma_{m,n}(s).$$

The player first makes m moves in order to reach a "good initial state", then plays n moves for payoffs.

Write $v^-(z) = \liminf_n v_n(z)$, $v^+(z) = \limsup_n v_n(z)$,
$v^* = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z)$.

Lemma 1: $v^-(z) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z)$.

Lemma 2: $\forall m_0$,

$$\inf_{n \geq 1} \sup_{m \leq m_0} v_{m,n}(z) \leq v^-(z) \leq v^+(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z).$$

can be restated as:

$$\inf_{n \geq 1} \sup_{z' \in G^{m_0}(z)} v_n(z') \leq v^-(z) \leq v^+(z) \leq v^*(z) = \inf_{n \geq 1} \sup_{z' \in G^\infty(z)} v_n(z').$$

where $G^{m_0}(z)$ is the set of states that can be reached from z in at most m_0 stages, and $G^\infty(z) = \cup_m G^m(z)$.

Define $V=\{v_n, n\geq 1\}\subset\{v: Z\longrightarrow[0,1]\}$, endowed with $d_\infty(v,v')=\sup_z|v(z)-v'(z)|$.

Thm 1 (R, JEMS 2011): $(v_n)_n$ CVU iff V is precompact.
And the uniform limit v^* can only be:

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

Sketch of proof:
1) Define $d(z,z')=\sup_{n\geq 1}|v_n(z)-v_n(z')|$. Prove that (Z,d) is pseudometric precompact. Clearly, each v_n is 1-Lipschitz for d.
2) Fix z. Prove that: $\forall\varepsilon>0$, $\exists m_0$, $\forall z'\in G^\infty(z), \exists z''\in G^{m_0}(z)$ s.t. $d(z',z'')\leq\varepsilon$.
3) Use $\inf_{n\geq 1}\sup_{z'\in G^{m_0}(z)} v_n(z')\leq v^-(z)\leq v^+(z)\leq \inf_{n\geq 1}\sup_{z'\in G^\infty(z)} v_n(z')$, and conclude.

Define $V = \{v_n, n \geq 1\} \subset \{v : Z \longrightarrow [0,1]\}$, endowed with $d_\infty(v, v') = \sup_z |v(z) - v'(z)|$.

Thm 1 (R, JEMS 2011): $(v_n)_n$ CVU iff V is precompact.
And the uniform limit v^* can only be:

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

Sketch of proof:
1) Define $d(z,z') = \sup_{n\geq 1} |v_n(z) - v_n(z')|$. Prove that (Z,d) is pseudometric precompact. Clearly, each v_n is 1-Lipschitz for d.
2) Fix z. Prove that: $\forall \varepsilon > 0$, $\exists m_0$, $\forall z' \in G^\infty(z), \exists z'' \in G^{m_0}(z)$ s.t. $d(z', z'') \leq \varepsilon$.
3) Use $\inf_{n\geq 1} \sup_{z' \in G^{m_0}(z)} v_n(z') \leq v^-(z) \leq v^+(z) \leq \inf_{n\geq 1} \sup_{z' \in G^\infty(z)} v_n(z')$, and conclude.

Define $V = \{v_n, n \geq 1\} \subset \{v : Z \longrightarrow [0,1]\}$, endowed with $d_\infty(v, v') = \sup_z |v(z) - v'(z)|$.

Thm 1 (R, JEMS 2011): $(v_n)_n$ CVU iff V is precompact.
And the uniform limit v^* can only be:

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

Sketch of proof:
1) Define $d(z,z') = \sup_{n\geq 1} |v_n(z) - v_n(z')|$. Prove that (Z,d) is pseudometric precompact. Clearly, each v_n is 1-Lipschitz for d.
2) Fix z. Prove that: $\forall \varepsilon > 0$, $\exists m_0$, $\forall z' \in G^\infty(z), \exists z'' \in G^{m_0}(z)$ s.t. $d(z', z'') \leq \varepsilon$.
3) Use
$\inf_{n\geq 1} \sup_{z' \in G^{m_0}(z)} v_n(z') \leq v^-(z) \leq v^+(z) \leq \inf_{n\geq 1} \sup_{z' \in G^\infty(z)} v_n(z')$, and conclude.

3b. Uniform CV of$(v_\theta)_\theta$

Let $(\theta^k)_{k\geq 1}$ be a family of probas s.t. $\sum_{t\geq 1}|\theta^k_{t+1}-\theta^k_t|\to 0$. Write $v^k=v_{\theta^k}$ and for each m put $v^{m,k}(z)=\sup_{s\in S(z)}\sum_{t\geq 1}\theta^k_t r(z_{m+t})$.

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

Lemma 3: $\forall m_0$,

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

Theorem (11-2011): $(v^k)_k$ CVU iff $\{v^k, k\geq 1\}$ is precompact.
And the uniform limit can only be:

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

All sequences $(v^k)_k$ have a unique limit point which is v^*.

3b. Uniform CV of$(v_\theta)_\theta$
Let $(\theta^k)_{k\geq 1}$ be a family of probas s.t. $\sum_{t\geq 1}|\theta^k_{t+1}-\theta^k_t| \to 0$. Write $v^k = v_{\theta^k}$ and for each m put $v^{m,k}(z) = \sup_{s\in S(z)} \sum_{t\geq 1}\theta^k_t r(z_{m+t})$.

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

Lemma 3: $\forall m_0$,
$\inf_k \sup_{m\leq m_0} v^{m,k}(z) \leq \liminf_k v^k(z) \leq \limsup_k v^k(z) \leq \inf_{k\geq 1}\sup_{m\geq 0} v^{m,k}(z)$.

Theorem (11-2011): $(v^k)_k$ CVU iff $\{v^k, k\geq 1\}$ is precompact.
And the uniform limit can only be:

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

All sequences $(v^k)_k$ have a unique limit point which is v^*.

A counterexample: Z countable, (v_n) pointwise CV to $1/2$, $(v^k)_k$ CVU to 1.

Corollary 1: In the following cases, we have general uniform convergence: for each $\varepsilon > 0$ there exists $\alpha > 0$ such that:
if $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.

a) Z is endowed with a distance d such that (Z, d) is precompact, and the family $(v_\theta)_\theta$ is uniformly equicontinuous.

b) Z is endowed with a distance d such that (Z, d) is compact, r is continuous and F is non expansive:
$\forall z \in Z, \forall z' \in Z, \forall z_1 \in F(z), \exists z'_1 \in F(z') \; s.t. \; d(z_1, z'_1) \leq d(z, z')$.

c) Z is finite (Blackwell, 1962).

3.c. The auxiliary functions $w_{m,n}$ and the Cesàro-uniform value

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t \geq 1}$, we define:

$$\gamma_{m,n}(s) = \frac{1}{n}\sum_{t=1}^{n} r(z_{m+t}), \text{ and } v_{m,n}(z) = \sup_{s \in S(z)} \gamma_{m,n}(s).$$

$\mu_{m,n}(s) = \min\{\gamma_{m,t}(s), t \in \{1, ..., n\}\}$, and $w_{m,n}(z) = \sup_{s \in S(z)} \mu_{m,n}(s)$.

$w_{m,n}$: the player first makes m moves in order to reach a "good initial state", but then his payoff only is the minimum of his next n average rewards.

Lemma 3:

$$v^+(z) \leq \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z) :=_{def} v^*(z).$$

Consider $W = \{(w_{m,n})_{m \geq 0, n \geq 1}\}$, endowed with the metric $d_\infty(w, w') = \sup\{|w(z) - w'(z)|, z \in Z\}$.

Thm 2 (R, JEMS 2011): Assume that W is precompact.
Then for every initial state z in Z, the pb has a Cesàro-uniform value which is: $v^*(z) = \sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(z) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z)$. And $(v_n)_n$ uniformly converges to v^*.

3.c. The auxiliary functions $w_{m,n}$ and the Cesàro-uniform value

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t\geq 1}$, we define:

$$\gamma_{m,n}(s) = \frac{1}{n}\sum_{t=1}^{n} r(z_{m+t}), \text{ and } v_{m,n}(z) = \sup_{s\in S(z)} \gamma_{m,n}(s).$$

$\mu_{m,n}(s) = \min\{\gamma_{m,t}(s), t \in \{1, ..., n\}\}$, and $w_{m,n}(z) = \sup_{s\in S(z)} \mu_{m,n}(s)$.

$w_{m,n}$: the player first makes m moves in order to reach a "good initial state", but then his payoff only is the minimum of his next n average rewards.

Lemma 3:

$$v^+(z) \leq \inf_{n\geq 1} \sup_{m\geq 0} w_{m,n}(z) = \inf_{n\geq 1} \sup_{m\geq 0} v_{m,n}(z) :=_{def} v^*(z).$$

Consider $W = \{(w_{m,n})_{m\geq 0, n\geq 1}\}$, endowed with the metric $d_\infty(w, w') = \sup\{|w(z) - w'(z)|, z \in Z\}$.

Thm 2 (R, JEMS 2011): Assume that W is precompact.
Then for every initial state z in Z, the pb has a Cesàro-uniform value which is: $v^*(z) = \sup_{m\geq 0} \inf_{n\geq 1} w_{m,n}(z) = \sup_{m\geq 0} \inf_{n\geq 1} v_{m,n}(z)$. And $(v_n)_n$ uniformly converges to v^*.

3.c. The auxiliary functions $w_{m,n}$ and the Cesàro-uniform value

For $m \geq 0$ and $n \geq 1$, $s = (z_t)_{t\geq 1}$, we define:

$$\gamma_{m,n}(s) = \frac{1}{n}\sum_{t=1}^{n} r(z_{m+t}), \text{ and } v_{m,n}(z) = \sup_{s\in S(z)}\gamma_{m,n}(s).$$

$\mu_{m,n}(s) = \min\{\gamma_{m,t}(s), t \in \{1,...,n\}\}$, and $w_{m,n}(z) = \sup_{s\in S(z)}\mu_{m,n}(s)$.

$w_{m,n}$: the player first makes m moves in order to reach a "good initial state", but then his payoff only is the minimum of his next n average rewards.

Lemma 3:

$$v^+(z) \leq \inf_{n\geq 1}\sup_{m\geq 0} w_{m,n}(z) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(z) :=_{def} v^*(z).$$

Consider $W = \{(w_{m,n})_{m\geq 0, n\geq 1}\}$, endowed with the metric $d_\infty(w,w') = \sup\{|w(z) - w'(z)|, z \in Z\}$.

Thm 2 (R, JEMS 2011): Assume that W is precompact. Then for every initial state z in Z, the pb has a Cesàro-uniform value which is: $v^*(z) = \sup_{m\geq 0}\inf_{n\geq 1} w_{m,n}(z) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(z)$. And $(v_n)_n$ uniformly converges to v^*.

Corollary 2: W is precompact, and thus the previous theorem applies in the following cases:

a) Z is endowed with a distance d such that (Z, d) is precompact, and the family $(w_{m,n})_{m \geq 0, n \geq 1}$ is uniformly equicontinuous.

b) Z is endowed with a distance d such that (Z, d) is compact, r is continuous and F is non expansive.

c) Z is finite.

3.d. The compact non expansive case; characterizing v^* (with X. Venel)

Fix Z compact metric, and F non expansive, and put $E = \{r : Z \longrightarrow [0,1], r\ C^0\}$. For each r in E, there is a limit value $\Phi(r)$. We have $\Phi(r) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}[r]$.

What are the properties of $\Phi : E \longrightarrow E$?

Ex: 0 player, ergodic Markov chain on a finite set: $\Phi(r) = < m^*, r >$, with m^* the unique invariant measure.

Define $A = \{r \in E, \Phi(r) = 0\}$, and $B = \{x \in E, \forall z\ \ x(z) = \sup_{z' \in F(z)} x(z')\}$. For each r, $\Phi(r) \in B$.

Proposition:

1) B is the set of fixed points of Φ, and $\Phi \circ \Phi = \Phi$.
2) for each r, $r - \Phi(r) \in A$. Hence we have $r = v + w$, with $v = \Phi(r) \in B$, and $w = r - \Phi(r) \in A$.
3) There exists a smallest function v in B such that $r - v \in A$, and this function is $\Phi(r)$.

$$\Phi(r) = \min\{v, v \in B \text{ and } r - v \in A\}.$$

3.d. The compact non expansive case; characterizing v^* (with X. Venel)

Fix Z compact metric, and F non expansive, and put $E = \{r : Z \longrightarrow [0,1], r\ C^0\}$. For each r in E, there is a limit value $\Phi(r)$. We have $\Phi(r) = \inf_{n\geq 1} \sup_{m\geq 0} v_{m,n}[r]$.

What are the properties of $\Phi : E \longrightarrow E$?

Ex: 0 player, ergodic Markov chain on a finite set: $\Phi(r) = < m^*, r >$, with m^* the unique invariant measure.

Define $A = \{r \in E, \Phi(r) = 0\}$, and $B = \{x \in E, \forall z\ \ x(z) = \sup_{z' \in F(z)} x(z')\}$. For each r, $\Phi(r) \in B$.

Proposition:
1) B is the set of fixed points of Φ, and $\Phi \circ \Phi = \Phi$.
2) for each r, $r - \Phi(r) \in A$. Hence we have $r = v + w$, with $v = \Phi(r) \in B$, and $w = r - \Phi(r) \in A$.
3) There exists a smallest function v in B such that $r - v \in A$, and this function is $\Phi(r)$.

$$\Phi(r) = \min\{v, v \in B \text{ and } r - v \in A\}.$$

3.d. The compact non expansive case; characterizing v^* (with X. Venel)

Fix Z compact metric, and F non expansive, and put
$E = \{r : Z \longrightarrow [0,1], r\ C^0\}$. For each r in E, there is a limit value $\Phi(r)$.
We have $\Phi(r) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}[r]$.

What are the properties of $\Phi : E \longrightarrow E$?

Ex: 0 player, ergodic Markov chain on a finite set: $\Phi(r) = < m^*, r >$, with m^* the unique invariant measure.

Define $A = \{r \in E, \Phi(r) = 0\}$, and
$B = \{x \in E,\ \forall z\ \ x(z) = \sup_{z' \in F(z)} x(z')\}$. For each r, $\Phi(r) \in B$.

Proposition:
1) B is the set of fixed points of Φ, and $\Phi \circ \Phi = \Phi$.
2) for each r, $r - \Phi(r) \in A$. Hence we have $r = v + w$, with $v = \Phi(r) \in B$, and $w = r - \Phi(r) \in A$.
3) There exists a smallest function v in B such that $r - v \in A$, and this function is $\Phi(r)$.

$$\Phi(r) = \min\{v, v \in B \text{ and } r - v \in A\}.$$

3.d. The compact non expansive case; characterizing v^* (with X. Venel)

Fix Z compact metric, and F non expansive, and put
$E = \{r : Z \longrightarrow [0,1], r\ C^0\}$. For each r in E, there is a limit value $\Phi(r)$.
We have $\Phi(r) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}[r]$.

What are the properties of $\Phi : E \longrightarrow E$?

Ex: 0 player, ergodic Markov chain on a finite set: $\Phi(r) = < m^*, r >$, with m^* the unique invariant measure.

Define $A = \{r \in E, \Phi(r) = 0\}$, and
$B = \{x \in E, \forall z\ \ x(z) = \sup_{z' \in F(z)} x(z')\}$. For each r, $\Phi(r) \in B$.

Proposition:
1) B is the set of fixed points of Φ, and $\Phi \circ \Phi = \Phi$.
2) for each r, $r - \Phi(r) \in A$. Hence we have $r = v + w$, with $v = \Phi(r) \in B$, and $w = r - \Phi(r) \in A$.
3) There exists a smallest function v in B such that $r - v \in A$, and this function is $\Phi(r)$.

$$\Phi(r) = \min\{v, v \in B \text{ and } r - v \in A\}.$$

Particular cases:

1) If the problem is ergodic ($\Phi(r)$ is constant for each r), then the decomposition $r = v + w$ with v in B and w in A is unique: Φ is the projection onto B along A .

2) Assume the game is *leavable*, i.e. $z \in \Gamma(z)$ for each z. Then $B = \{x \in E, \forall z \;\; x(z) \geq \sup_{z' \in F(z)} x(z')\}$ (*excessive functions*) is convex, and

$$\Phi(r) = \min\{v, v \in B, v \geq r\}$$

(Gambling Fundamental Theorem, Dubins Savage 1965)

Particular cases:

1) If the problem is ergodic ($\Phi(r)$ is constant for each r), then the decomposition $r = v + w$ with v in B and w in A is unique: Φ is the projection onto B along A .

2) Assume the game is *leavable*, i.e. $z \in \Gamma(z)$ for each z. Then $B = \{x \in E, \forall z \;\; x(z) \geq \sup_{z' \in F(z)} x(z')\}$ (*excessive functions*) is convex, and

$$\Phi(r) = \min\{v, v \in B, v \geq r\}$$

(Gambling Fundamental Theorem, Dubins Savage 1965)

X is a metric compact space, $F : X \rightrightarrows \Delta_f(X)$ is non expansive (for the W distance), $r : X \to [0,1]$ is continuous (and linearly extended to $\Delta(X)$).

Define $\hat{F} : \Delta_f(X) \rightrightarrows \Delta_f(X)$ the mixed extension of F by:

$$\hat{F}(u) = \left\{ \int_{p\in X} f(p)du(p), \text{where } f(p) \in \text{conv} F(p) \text{for all } p \right\}.$$

We now have a dynamic programming problem $(\Delta_f(X), \hat{F}, r)$ where for each θ, v_θ is affine. Put:

$$R = \{u \in \Delta(X), (u,u) \in \overline{Graph\ \hat{F}}\}, \text{and}$$

$$v^*(p) = \inf\{w(p), w : \Delta(X) \to [0,1] \text{ affine } C^0\ s.t.$$
$$(1) \forall p' \in X, w(p') \geq \sup\nolimits_{u \in F(p')} w(u)$$
$$(2) \forall u \in R, w(u) \geq r(u)\}.$$

Theorem 3 (R-Venel 11-2011): For each $\varepsilon > 0$ there exists $\alpha > 0$ such that if θ satisfies $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Moreover, for each u in $\Delta_f(X)$ and $\varepsilon > 0$, there exists a play σ in $\hat{S}(u)$ and $\alpha > 0$ such that: $(\sum_{t=1}^\infty \theta_t r(u_t)) \geq v^*(u) - \varepsilon$ if $\sum_{t\geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$

X is a metric compact space, $F : X \rightrightarrows \Delta_f(X)$ is non expansive (for the W distance), $r : X \to [0,1]$ is continuous (and linearly extended to $\Delta(X)$).

Define $\hat{F} : \Delta_f(X) \rightrightarrows \Delta_f(X)$ the mixed extension of F by:

$$\hat{F}(u) = \left\{ \int_{p \in X} f(p) du(p), \text{where } f(p) \in \text{conv} F(p) \text{for all } p \right\}.$$

We now have a dynamic programming problem $(\Delta_f(X), \hat{F}, r)$ where for each θ, v_θ is affine. Put:

$$R = \{u \in \Delta(X), (u,u) \in \overline{Graph\ \hat{F}}\}, \text{and}$$

$$\begin{aligned} v^*(p) = \inf\{w(p), w : \Delta(X) \to [0,1] \text{ affine } C^0\ s.t. \\ (1) \forall p' \in X, w(p') \geq \sup_{u \in F(p')} w(u) \\ (2) \forall u \in R, w(u) \geq r(u)\}. \end{aligned}$$

Theorem 3 (R-Venel 11-2011): For each $\varepsilon > 0$ there exists $\alpha > 0$ such that if θ satisfies $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Moreover, for each u in $\Delta_f(X)$ and $\varepsilon > 0$, there exists a play σ in $\hat{S}(u)$ and $\alpha > 0$ such that: $(\sum_{t=1}^{\infty} \theta_t r(u_t)) \geq v^*(u) - \varepsilon$ if $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$

X is a metric compact space, $F : X \rightrightarrows \Delta_f(X)$ is non expansive (for the W distance), $r : X \to [0,1]$ is continuous (and linearly extended to $\Delta(X)$).

Define $\hat{F} : \Delta_f(X) \rightrightarrows \Delta_f(X)$ the mixed extension of F by:

$$\hat{F}(u) = \left\{ \int_{p \in X} f(p) du(p), \text{where } f(p) \in \text{conv} F(p) \text{for all } p \right\}.$$

We now have a dynamic programming problem $(\Delta_f(X), \hat{F}, r)$ where for each θ, v_θ is affine. Put:

$$R = \{u \in \Delta(X), (u,u) \in \overline{Graph\ \hat{F}}\}, \text{and}$$

$$v^*(p) = \inf\{w(p), w : \Delta(X) \to [0,1] \text{ affine } C^0 \ s.t.$$
$$(1) \forall p' \in X, w(p') \geq \sup_{u \in F(p')} w(u)$$
$$(2) \forall u \in R, w(u) \geq r(u)\}.$$

Theorem 3 (R-Venel 11-2011): For each $\varepsilon > 0$ there exists $\alpha > 0$ such that if θ satisfies $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Moreover, for each u in $\Delta_f(X)$ and $\varepsilon > 0$, there exists a play σ in $\hat{S}(u)$ and $\alpha > 0$ such that: $(\sum_{t=1}^{\infty} \theta_t r(u_t)) \geq v^*(u) - \varepsilon$ if $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$

X is a metric compact space, $F : X \rightrightarrows \Delta_f(X)$ is non expansive (for the W distance), $r : X \to [0,1]$ is continuous (and linearly extended to $\Delta(X)$).

Define $\hat{F} : \Delta_f(X) \rightrightarrows \Delta_f(X)$ the mixed extension of F by:

$$\hat{F}(u) = \left\{ \int_{p \in X} f(p) du(p), \text{where } f(p) \in \text{conv} F(p) \text{for all } p \right\}.$$

We now have a dynamic programming problem $(\Delta_f(X), \hat{F}, r)$ where for each θ, v_θ is affine. Put:

$$R = \{u \in \Delta(X), (u,u) \in \overline{Graph\ \hat{F}}\}, \text{and}$$

$$v^*(p) = \inf\{w(p), w : \Delta(X) \to [0,1] \text{ affine } C^0\ s.t.$$
$$(1) \forall p' \in X, w(p') \geq \sup_{u \in F(p')} w(u)$$
$$(2) \forall u \in R, w(u) \geq r(u)\}.$$

Theorem 3 (R-Venel 11-2011): For each $\varepsilon > 0$ there exists $\alpha > 0$ such that if θ satisfies $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$, then $\|v_\theta - v^*\| \leq \varepsilon$.
Moreover, for each u in $\Delta_f(X)$ and $\varepsilon > 0$, there exists a play σ in $\hat{S}(u)$ and $\alpha > 0$ such that: $(\sum_{t=1}^{\infty} \theta_t r(u_t)) \geq v^*(u) - \varepsilon$ if $\sum_{t \geq 1} |\theta_{t+1} - \theta_t| \leq \alpha$

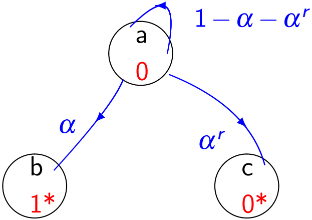
3.e. Computing v^* and the speed of convergence (with X. venel)
Markov Decision Processes with finite state and actions: in a neighborhood of zero, v_λ is a rational function. So $v_\lambda(z) = v^*(z) + O(\lambda)$, and also $v_n(z) = v^*(z) + O(1/n)$.

Untrue with infinitely many actions: example 2 with $r > 1$

We have $v_\lambda(a) = 1 - C\lambda^{(r-1)/r} + o(\lambda^{(r-1)/r})$, with $C = \frac{r}{(r-1)^{\frac{r-1}{r}}}$.

3.e. Computing v^* and the speed of convergence (with X. venel)
Markov Decision Processes with finite state and actions: in a neighborhood of zero, v_λ is a rational function. So $v_\lambda(z) = v^*(z) + O(\lambda)$, and also $v_n(z) = v^*(z) + O(1/n)$.

Untrue with infinitely many actions: example 2 with $r > 1$



We have $v_\lambda(a) = 1 - C\lambda^{(r-1)/r} + o(\lambda^{(r-1)/r})$, with $C = \frac{r}{(r-1)^{\frac{r-1}{r}}}$.

Pb: compute $\lim_\lambda v_\lambda$, where:

$$v_\lambda(z) = \sup_{z' \in F(z)} \lambda r(z') + (1-\lambda) v_\lambda(z').$$

One has: $v^*(z) = \sup_{z' \in F(z)} v^*(z')$, but r has disappeared.

Assume ergodicity, with an expansion $v_\lambda(z) = v^* + \lambda V(z) + o(\lambda)$, for some function V. Then the Average Cost Optimality Equation holds:

$$v^* + V(z) = \sup_{z' \in F(z)} r(z') + V(z').$$

What if no ergodicity, or if the speed of CV is different ?

Idea: write $\lambda r(z') + (1-\lambda) v_\lambda(z') \sim v_\lambda(z') + \lambda r(z') - \lambda v^*(z')$, and consider an (approximate) solution of:

$$h_\lambda(z) = \sup_{z' \in F(z)} h_\lambda(z') + \lambda(r(z') - v^*(z'))).$$

Pb: compute $\lim_\lambda v_\lambda$, where:

$$v_\lambda(z) = \sup_{z' \in F(z)} \lambda r(z') + (1-\lambda) v_\lambda(z').$$

One has: $v^*(z) = \sup_{z' \in F(z)} v^*(z')$, but r has disappeared.

Assume ergodicity, with an expansion $v_\lambda(z) = v^* + \lambda V(z) + o(\lambda)$, for some function V. Then the Average Cost Optimality Equation holds:

$$v^* + V(z) = \sup_{z' \in F(z)} r(z') + V(z').$$

What if no ergodicity, or if the speed of CV is different ?

Idea: write $\lambda r(z') + (1-\lambda) v_\lambda(z') \sim v_\lambda(z') + \lambda r(z') - \lambda v^*(z')$, and consider an (approximate) solution of:

$$h_\lambda(z) = \sup_{z' \in F(z)} h_\lambda(z') + \lambda(r(z') - v^*(z')).$$

Pb: compute $\lim_\lambda v_\lambda$, where:

$$v_\lambda(z) = \sup_{z' \in F(z)} \lambda r(z') + (1-\lambda) v_\lambda(z').$$

One has: $v^*(z) = \sup_{z' \in F(z)} v^*(z')$, but r has disappeared.

Assume ergodicity, with an expansion $v_\lambda(z) = v^* + \lambda V(z) + o(\lambda)$, for some function V. Then the Average Cost Optimality Equation holds:

$$v^* + V(z) = \sup_{z' \in F(z)} r(z') + V(z').$$

What if no ergodicity, or if the speed of CV is different ?

Idea: write $\lambda r(z') + (1-\lambda) v_\lambda(z') \sim v_\lambda(z') + \lambda r(z') - \lambda v^*(z')$, and consider an (approximate) solution of:

$$h_\lambda(z) = \sup_{z' \in F(z)} h_\lambda(z') + \lambda(r(z') - v^*(z')).$$

Verification principle :

Assume that $(h_\lambda)_\lambda$ uniformly converges to some $h_0 : Z \to [0,1]$, and that $\frac{1}{\lambda}\|h_\lambda - \tilde{h}_\lambda\| \longrightarrow 0$, where $\tilde{h}_\lambda(z) = \sup_{z' \in F(z)} h_\lambda(z') + \lambda(r(z') - h_0(z'))$.

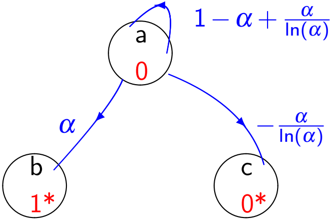
Then $(v_\lambda)_\lambda$ also uniformly converges to h_0, and

$$\|v_\lambda - h_0\| \leq 2\|h_\lambda - h_0\| + \frac{1}{\lambda}\|h_\lambda - \tilde{h}_\lambda\| \longrightarrow_{\lambda \to 0} 0.$$

And if v_λ UCV to h_0, then v_λ itself satisfies $\frac{1}{\lambda}\|v_\lambda - \tilde{v}_\lambda\| \longrightarrow 0$.

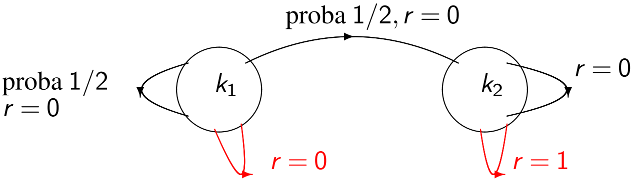
Rem: a similar principle holds for $\lim_n v_n$.

Ex:



We have $v_\lambda(a) = 1 + \frac{1}{\ln(\lambda)} + O(\lambda)$.

Ex: a *blind* MDP with 2 states and 2 actions where $\|v_\lambda - 1\| \sim C\lambda\ln(\lambda)$.



- The value is difficult to compute. $K = \{a, b\}$, $p = (1/2, 1/2)$,
$M = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix}$, $G^a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $G^b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.
If $\alpha = 1$, the value is $1/4$ (Aumann Maschler setup).
If $\alpha \in [1/2, 2/3]$, the value is $\frac{\alpha}{4\alpha - 1}$ (Hörner *et al.* 2006, Marino 2005 for $\alpha = 2/3$).
What is the value for $\alpha = 0.9$?

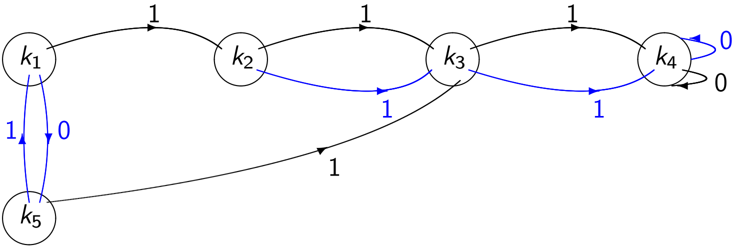
- The value is difficult to compute. $K = \{a, b\}$, $p = (1/2, 1/2)$,

$$M = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix},\ G^a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } G^b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $\alpha = 1$, the value is $1/4$ (Aumann Maschler setup).
If $\alpha \in [1/2, 2/3]$, the value is $\frac{\alpha}{4\alpha - 1}$ (Hörner *et al.* 2006, Marino 2005 for $\alpha = 2/3$).
What is the value for $\alpha = 0.9$?

4.a. Standard Markov Decision Processes with a finite set of states

Controlled Markov chains

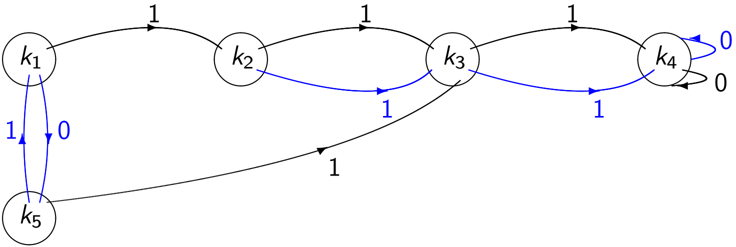


MDP $\Psi(p_0)$: A finite set of states K, a non empty set of actions A, a transition function q from $K \times A$ to $\Delta(K)$, a reward function $g : K \times A \longrightarrow [0,1]$, and an initial probability p_0 on K.

k_1 in K is selected according to p_0 and told to the player, then he selects a_1 in A and receives a payoff of $g(k_1, a_1)$. A new state k_2 is selected according to $q(k_1, a_1)$ and told to the player, etc...

4.a. Standard Markov Decision Processes with a finite set of states

Controlled Markov chains



MDP $\Psi(p_0)$: A finite set of states K, a non empty set of actions A, a transition function q from $K \times A$ to $\Delta(K)$, a reward function $g : K \times A \longrightarrow [0,1]$, and an initial probability p_0 on K.

k_1 in K is selected according to p_0 and told to the player, then he selects a_1 in A and receives a payoff of $g(k_1, a_1)$. A new state k_2 is selected according to $q(k_1, a_1)$ and told to the player, etc...

A pure strategy: $\sigma = (\sigma_t)_{t\geq 1}$, with $\forall t$, $\sigma_t : (K\times A)^{t-1}\times K \longrightarrow A$ defines the action to be played at stage t. (p_0,σ) generates a proba on plays, one can define the expected payoffs and the n-stage values.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z=\Delta(K)\times[0,1]$, a new initial state $z_0=(p_0,0)$, new payoff function $r(p,y)=y$ for all (p,y) in Z, a transition correspondence such that for every $z=(p,y)$ in Z,

$$F(z)=\left\{\left(\sum_{k\in K} p^k q(k,a_k), \sum_{k\in K} p^k g(k,a_k)\right), a_k\in A\ \forall k\in K\right\}.$$

Put $d((p,y),(p',y'))=\max\{\|p-p'\|_1,|y-y'|\}$.

Apply theorem 3 to obtain the UCV of $(v_\theta)_\theta$ (*for any set A*).
Well known for the Cesàro limit when A finite (Blackwell 1962), and for A compact and q, g continuous in a (Dynkin Yushkevich 1979).

A pure strategy: $\sigma = (\sigma_t)_{t\geq 1}$, with $\forall t$, $\sigma_t : (K \times A)^{t-1} \times K \longrightarrow A$ defines the action to be played at stage t. (p_0, σ) generates a proba on plays, one can define the expected payoffs and the n-stage values.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z = \Delta(K) \times [0,1]$, a new initial state $z_0 = (p_0, 0)$, new payoff function $r(p,y) = y$ for all (p,y) in Z, a transition correspondence such that for every $z = (p,y)$ in Z,

$$F(z) = \left\{ \left(\sum_{k\in K} p^k q(k,a_k), \sum_{k\in K} p^k g(k,a_k) \right), a_k \in A \, \forall k \in K \right\}.$$

Put $d((p,y),(p',y')) = \max\{\|p-p'\|_1, |y-y'|\}$.

Apply theorem 3 to obtain the UCV of $(v_\theta)_\theta$ (*for any set A*).
Well known for the Cesàro limit when A finite (Blackwell 1962), and for A compact and q, g continuous in a (Dynkin Yushkevich 1979).

A pure strategy: $\sigma = (\sigma_t)_{t\geq 1}$, with $\forall t$, $\sigma_t : (K\times A)^{t-1}\times K \longrightarrow A$ defines the action to be played at stage t. (p_0,σ) generates a proba on plays, one can define the expected payoffs and the n-stage values.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z=\Delta(K)\times[0,1]$, a new initial state $z_0=(p_0,0)$, new payoff function $r(p,y)=y$ for all (p,y) in Z, a transition correspondence such that for every $z=(p,y)$ in Z,

$$F(z)=\left\{\left(\sum_{k\in K} p^k q(k,a_k), \sum_{k\in K} p^k g(k,a_k)\right), a_k\in A\,\forall k\in K\right\}.$$

Put $d((p,y),(p',y'))=\max\{\|p-p'\|_1, |y-y'|\}$.

Apply theorem 3 to obtain the UCV of $(v_\theta)_\theta$ (*for any set A*).
Well known for the Cesàro limit when A finite (Blackwell 1962), and for A compact and q, g continuous in a (Dynkin Yushkevich 1979).

A pure strategy: $\sigma = (\sigma_t)_{t\geq 1}$, with $\forall t$, $\sigma_t : (K\times A)^{t-1}\times K \longrightarrow A$ defines the action to be played at stage t. (p_0,σ) generates a proba on plays, one can define the expected payoffs and the n-stage values.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z=\Delta(K)\times[0,1]$, a new initial state $z_0=(p_0,0)$, new payoff function $r(p,y)=y$ for all (p,y) in Z, a transition correspondence such that for every $z=(p,y)$ in Z,

$$F(z)=\left\{\left(\sum_{k\in K} p^k q(k,a_k), \sum_{k\in K} p^k g(k,a_k)\right), a_k\in A\,\forall k\in K\right\}.$$

Put $d((p,y),(p',y'))=\max\{\|p-p'\|_1,|y-y'|\}$.

Apply theorem 3 to obtain the UCV of $(v_\theta)_\theta$ (*for any set* A).
Well known for the Cesàro limit when A finite (Blackwell 1962), and for A compact and q, g continuous in a (Dynkin Yushkevich 1979).

And there is general uniform value if we allow for mixed strategies. The expression for v^* becomes:

$$v^* = \inf\{v : \Delta(K) \to [0,1] \text{ affine } s.t.$$

$$(1)\ \forall k \in K, v(k) \geq \sup_{a \in A} v(q(k,a))$$

$$(2)\ \forall (p,y) \in R, \sum_k p^k v(k) \geq y\}.$$

where $R = \{(p,y) \in \Delta(K) \times [0,1], (p,y) \in \overline{\text{conv}}\{(\sum_k p^k q(k,a_k), \sum_k p^k g(k,a_k)), \forall k, a_k \in A\}$.

4.b. Application to non expansive control problems (with M. Quincampoix)

We consider a control problem of the following form:

$$V_t(x_0) = \sup_{u \in \mathscr{U}} \frac{1}{t} \int_{s=0}^{t} g(x_{x_0,u}(s), u(s))ds, \qquad (1)$$

where $t > 0$, U is a non empty measurable set of controls (subset of a Polish space), $\mathscr{U} = \{u : I\!R_+ \longrightarrow U \; measurable\}$,
$g : I\!R^n \times U \longrightarrow [0,1]$ is measurable, and $x_{x_0,u}$ is the solution of:

$$\dot{x}(s) = f(x(s), u(s)), \quad x(0) = x_0. \qquad (2)$$

x_0 is an initial state in $I\!R^n$, $f : I\!R^n \times U \longrightarrow I\!R^n$ is measurable, Lipschitz in x uniformly in u, and s.t. $\exists a > 0, \forall x, u, \|f(x,u)\| \leq a(1 + \|x\|)$.

Say the problem has a Cesàro-uniform value if it has a limit value $V^*(x_0) = \lim_{t \to \infty} V_t(x_0)$ and:

$$\forall \varepsilon > 0, \exists u \in \mathscr{U}, \exists t_0, \forall t \geq t_0, \frac{1}{t} \int_{s=0}^{t} g(x_{x_0,u}(s), u(s))ds \geq V^*(x_0) - \varepsilon.$$

4.b. Application to non expansive control problems (with M. Quincampoix)

We consider a control problem of the following form:

$$V_t(x_0) = \sup_{u\in\mathscr{U}} \frac{1}{t}\int_{s=0}^{t} g(x_{x_0,u}(s), u(s))ds, \quad (1)$$

where $t > 0$, U is a non empty measurable set of controls (subset of a Polish space), $\mathscr{U} = \{u : I\!R_+ \longrightarrow U \; measurable\}$,
$g : I\!R^n \times U \longrightarrow [0,1]$ is measurable, and $x_{x_0,u}$ is the solution of:

$$\dot{x}(s) = f(x(s), u(s)), \quad x(0) = x_0. \quad (2)$$

x_0 is an initial state in $I\!R^n$, $f : I\!R^n \times U \longrightarrow I\!R^n$ is measurable, Lipschitz in x uniformly in u, and s.t. $\exists a > 0, \forall x, u, \|f(x,u)\| \leq a(1+\|x\|)$.

Say the problem has a Cesàro-uniform value if it has a limit value $V^*(x_0) = \lim_{t\to\infty} V_t(x_0)$ and:

$$\forall \varepsilon > 0, \exists u \in \mathscr{U}, \exists t_0, \forall t \geq t_0, \frac{1}{t}\int_{s=0}^{t} g(x_{x_0,u}(s), u(s))ds \geq V^*(x_0) - \varepsilon.$$

No ergodicity condition here (Arisawa-Lions 98, Bettiol 2005,...).
The limit value may depend on the initial state.

Example 1: in the complex plane, $f(x,u) = ix$.
if $g(x,u) = g(x)$, then

$$V_t(x_0) \xrightarrow[t \to \infty]{} \frac{1}{2\pi|x_0|} \int_{|z|=|x_0|} g(z)dz.$$

Example 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset I\!R$.
$g(x,u) = g(x)$ continuous.

Example 3: $f(x,u) = -x + u$, with $u \in U$ compact subset of $I\!R^n$.
$g(x,u) = g(x)$ continuous.

No ergodicity condition here (Arisawa-Lions 98, Bettiol 2005,...).
The limit value may depend on the initial state.

Example 1: in the complex plane, $f(x,u) = ix$.
if $g(x,u) = g(x)$, then

$$V_t(x_0) \xrightarrow[t\to\infty]{} \frac{1}{2\pi|x_0|}\int_{|z|=|x_0|} g(z)dz.$$

Example 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset IR$.
$g(x,u) = g(x)$ continuous.

Example 3: $f(x,u) = -x + u$, with $u \in U$ compact subset of IR^n.
$g(x,u) = g(x)$ continuous.

No ergodicity condition here (Arisawa-Lions 98, Bettiol 2005,...).
The limit value may depend on the initial state.

Example 1: in the complex plane, $f(x,u) = ix$.
if $g(x,u) = g(x)$, then

$$V_t(x_0) \xrightarrow[t\to\infty]{} \frac{1}{2\pi|x_0|}\int_{|z|=|x_0|} g(z)dz.$$

Example 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset I\!R$.
$g(x,u) = g(x)$ continuous.

Example 3: $f(x,u) = -x + u$, with $u \in U$ compact subset of $I\!R^n$.
$g(x,u) = g(x)$ continuous.

Example 4: in $I\!R^2$, $x(0) = (0,0)$, control set $U = [0,1]$,
$\dot{x} = f(x,u) = \begin{pmatrix} u(1-x_1) \\ u^2(1-x_1) \end{pmatrix}$, and $g(x) = x_1(1-x_2)$.

if $u = \varepsilon$ constant, then $x_1(t) = 1 - exp(-\varepsilon t)$ and $x_2(t) = \varepsilon x_1(t)$.

Uniform value $V(0,0) = 1$.
$V(x_1,x_2) = 1 - x_2$. *no ergodicity*

Notations: for every $t > 0$, $m \geq 0$, $x_0 \in I\!R^n$ and $u \in \mathcal{U}$, we define the average payoff induced by u between time m and time $m+t$ by:

$$\gamma_{m,t}(x_0,u) = \frac{1}{t}\int_m^{m+t} g(x_{x_0,u}(s),u(s))ds,$$

and the value of the problem where the time interval $[0,m]$ can be devoted to reach a good initial state, is denoted by:

$$V_{m,t}(x_0) = \sup_{u\in\mathcal{U}} \gamma_{m,t}(x_0,u).$$

Theorem (R- Quincampoix SICON 2011) Assume that:
(H1) $g = g(x)$ is continuous on IR^n.
(H2) $G(x_0)$ is bounded.
(H3) $\forall x \in K$, $\forall y \in K$, $\sup_{u\in U}\inf_{v\in U} < x-y, f(x,u)-f(y,v) > \leq 0$.

Then $V_t(x_0) \xrightarrow[t\to\infty]{} V^*(x_0)$. The convergence is uniform over $G(x_0)$, and $V^*(x_0) = \inf_{t\geq 1}\sup_{m\geq 0} V_{m,t}(x_0) = \sup_{m\geq 0}\inf_{t\geq 1} V_{m,t}(x_0)$. And the value is Cesàro-uniform.

11-2011: moreover we have general uniform convergence

$$\sup_{u\in\mathscr{U}} \int_{s=0}^{+\infty} \theta_s g(x_{x_0,u}(s))ds \to V^*(x_0) \text{ when } \int_{s=0}^{+\infty} |\theta(s+1)-\theta(s)|ds \to 0.$$

(and general uniform value if we allow for random controls)

• example 1 & 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset IR$.
• example 3: $f(x,u) = -x+u$, whith $u \in U$ compact subset of IR^n.
• example 4: H3 not satisfied (but conclusions satisfied). -> generalization of the theorem to deal with more general distances.

Theorem (R- Quincampoix SICON 2011) Assume that:
(H1) $g = g(x)$ is continuous on IR^n.
(H2) $G(x_0)$ is bounded.
(H3) $\forall x \in K$, $\forall y \in K$, $\sup_{u\in U}\inf_{v\in U} < x-y, f(x,u)-f(y,v) > \leq 0$.

Then $V_t(x_0) \xrightarrow[t\to\infty]{} V^*(x_0)$. The convergence is uniform over $G(x_0)$, and $V^*(x_0) = \inf_{t\geq 1}\sup_{m\geq 0} V_{m,t}(x_0) = \sup_{m\geq 0}\inf_{t\geq 1} V_{m,t}(x_0)$. And the value is Cesàro-uniform.
11-2011: moreover we have general uniform convergence

$$\sup_{u\in\mathscr{U}} \int_{s=0}^{+\infty} \theta_s g(x_{x_0,u}(s))ds \to V^*(x_0) \text{ when } \int_{s=0}^{+\infty} |\theta(s+1)-\theta(s)|ds \to 0.$$

(and general uniform value if we allow for random controls)

• example 1 & 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset IR$.
• example 3: $f(x,u) = -x+u$, whith $u \in U$ compact subset of IR^n.
• example 4: H3 not satisfied (but conclusions satisfied). -> generalization of the theorem to deal with more general distances.

Theorem (R- Quincampoix SICON 2011) Assume that:
(H1) $g = g(x)$ is continuous on $I\!R^n$.
(H2) $G(x_0)$ is bounded.
(H3) $\forall x \in K$, $\forall y \in K$, $\sup_{u \in U} \inf_{v \in U} < x - y, f(x,u) - f(y,v) > \leq 0$.

Then $V_t(x_0) \underset{t \to \infty}{\longrightarrow} V^*(x_0)$. The convergence is uniform over $G(x_0)$, and $V^*(x_0) = \inf_{t \geq 1} \sup_{m \geq 0} V_{m,t}(x_0) = \sup_{m \geq 0} \inf_{t \geq 1} V_{m,t}(x_0)$. And the value is Cesàro-uniform.
11-2011: moreover we have general uniform convergence

$$\sup_{u \in \mathscr{U}} \int_{s=0}^{+\infty} \theta_s g(x_{x_0,u}(s))ds \to V^*(x_0) \text{ when } \int_{s=0}^{+\infty} |\theta(s+1) - \theta(s)| ds \to 0.$$

(and general uniform value if we allow for random controls)

• example 1 & 2: in the complex plane, $f(x,u) = ixu$, with $u \in U \subset I\!R$.
• example 3: $f(x,u) = -x + u$, whith $u \in U$ compact subset of $I\!R^n$.
• example 4: H3 not satisfied (but conclusions satisfied). ->
generalization of the theorem to deal with more general distances.

4.c. MDPs with partial observation. Hidden controlled Markov chain

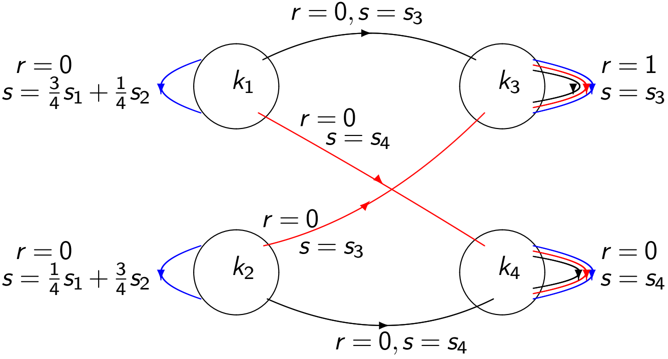
More general model where the player may not perfectly observe the state.

States $K=\{k_1,k_2,k_3,k_4\}$, Actions $\curvearrowright, \curvearrowright, \curvearrowright$, Signals: $\{s_1,s_2,s_3,s_4\}$.
$p_0 = 1/2\,\delta_{k_1} + 1/2\,\delta_{k_2}$.

Playing $\curvearrowright$ for a large number of stages, and then $\curvearrowright$ or $\curvearrowright$ depending on the stream of signals received, is ε-optimal. $v^*(p_0)=1$, the uniform value exists, but non existence of 0-optimal strategies.

4.c. MDPs with partial observation. Hidden controlled Markov chain

More general model where the player may not perfectly observe the state.

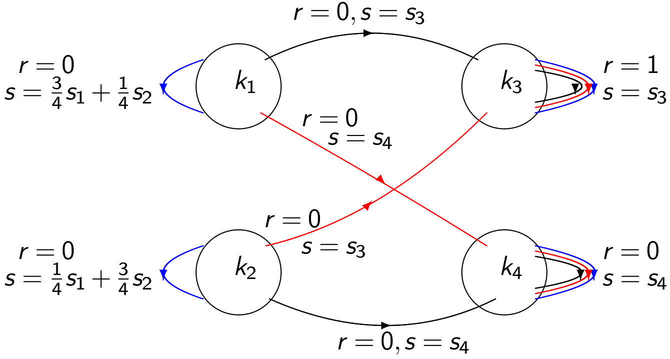


States $K = \{k_1, k_2, k_3, k_4\}$, Actions $\curvearrowright, \curvearrowright, \curvearrowright$, Signals: $\{s_1, s_2, s_3, s_4\}$.
$p_0 = 1/2\, \delta_{k_1} + 1/2\, \delta_{k_2}$.

Playing $\curvearrowright$ for a large number of stages, and then $\curvearrowright$ or $\curvearrowright$ depending on the stream of signals received, is ε-optimal. $v^*(p_0) = 1$, the uniform value exists, but non existence of 0-optimal strategies.

4.c. MDPs with partial observation. Hidden controlled Markov chain

More general model where the player may not perfectly observe the state.



States $K = \{k_1, k_2, k_3, k_4\}$, Actions $\curvearrowright, \curvearrowright, \curvearrowright$, Signals: $\{s_1, s_2, s_3, s_4\}$.
$p_0 = 1/2\,\delta_{k_1} + 1/2\,\delta_{k_2}$.

Playing $\curvearrowright$ for a large number of stages, and then $\curvearrowright$ or $\curvearrowright$ depending on the stream of signals received, is ε-optimal. $v^*(p_0) = 1$, the uniform value exists, but non existence of 0-optimal strategies.

Finite set of states K, initial probability p_0 on K, non empty set of actions A, and also a non empty set of signals S. Transition $q : K \times A \to \Delta_f(S \times K)$, and reward function $g : K \times A \to [0,1]$.

k_1 in K is selected according to p_0 and is not told to the player. At stage t the player selects an action $a_t \in A$, and has a (unobserved) payoff $g(k_t, a_t)$. Then a pair (s_t, k_{t+1}) is selected according to $q(k_t, a_t)$, and s_t is told to the player. The new state is k_{t+1}, and the play goes to stage $t+1$.
Rosenberg Solan Vieille 2002: for K, A and S finite the Cesàro uniform value exists.

Write $X = \Delta(K)$. Assume that the state of some stage has been selected according to p in X and the player plays some action a in A. This defines a probability $\hat{q}(p,a)$ on the future belief of the player on the state of the next stage. $\hat{q}(p,a) \in \Delta_f(X)$.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z = \Delta_f(X) \times [0,1]$, new initial state $z_0 = (\delta_{p_0}, 0)$, new payoff function $r(u,y) = y$ for all (u,y) in Z, transition correspondence such that for every $z = (u,y)$ in Z,

$$F(z) = \{(H(u,f), R(u,f)), f : X \longrightarrow \Delta_f(A)\},$$

where $H(u,f) = \sum_{p \in X} u(p) \left(\sum_{a \in A} f(p)(a) \hat{q}(p,a)\right) \in \Delta_f(X)$,
and $R(u,f) = \sum_{p \in X} u(p) \left(\sum_{k \in K, a \in A} p^k f(p)(a) g(k,a)\right)$.

Use $\|.\|_1$ on X. $\Delta(X)$: Borel probabilities over X, with the weak-* topology. Topology metrized by the Wasserstein distance :

$$\forall u \in \Delta(X), \forall v \in \Delta(X), \ \ d(u,v) = \sup_{f \in E_1} |u(f) - v(f)|.$$

Write $X = \Delta(K)$. Assume that the state of some stage has been selected according to p in X and the player plays some action a in A. This defines a probability $\hat{q}(p,a)$ on the future belief of the player on the state of the next stage. $\hat{q}(p,a) \in \Delta_f(X)$.

$\to$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z = \Delta_f(X) \times [0,1]$, new initial state $z_0 = (\delta_{p_0}, 0)$, new payoff function $r(u,y) = y$ for all (u,y) in Z, transition correspondence such that for every $z = (u,y)$ in Z,

$$F(z) = \{(H(u,f), R(u,f)), f : X \longrightarrow \Delta_f(A)\},$$

where $H(u,f) = \sum_{p \in X} u(p) \left(\sum_{a \in A} f(p)(a) \hat{q}(p,a)\right) \in \Delta_f(X)$, and $R(u,f) = \sum_{p \in X} u(p) \left(\sum_{k \in K, a \in A} p^k f(p)(a) g(k,a)\right)$.

Use $\|.\|_1$ on X. $\Delta(X)$: Borel probabilities over X, with the weak-* topology. Topology metrized by the Wasserstein distance :

$$\forall u \in \Delta(X), \forall v \in \Delta(X), \;\; d(u,v) = \sup_{f \in E_1} |u(f) - v(f)|.$$

Write $X = \Delta(K)$. Assume that the state of some stage has been selected according to p in X and the player plays some action a in A. This defines a probability $\hat{q}(p,a)$ on the future belief of the player on the state of the next stage. $\hat{q}(p,a) \in \Delta_f(X)$.

$\rightarrow$ Auxiliary deterministic Pb $\Gamma(z_0)$: new set of states $Z = \Delta_f(X) \times [0,1]$, new initial state $z_0 = (\delta_{p_0}, 0)$, new payoff function $r(u,y) = y$ for all (u,y) in Z, transition correspondence such that for every $z = (u,y)$ in Z,

$$F(z) = \{(H(u,f), R(u,f)), f : X \longrightarrow \Delta_f(A)\},$$

where $H(u,f) = \sum_{p \in X} u(p) \left(\sum_{a \in A} f(p)(a)\hat{q}(p,a)\right) \in \Delta_f(X)$,
and $R(u,f) = \sum_{p \in X} u(p) \left(\sum_{k \in K, a \in A} p^k f(p)(a) g(k,a)\right)$.

Use $\|.\|_1$ on X. $\Delta(X)$: Borel probabilities over X, with the weak-* topology. Topology metrized by the Wasserstein distance :

$$\forall u \in \Delta(X), \forall v \in \Delta(X), \;\; d(u,v) = \sup_{f \in E_1} |u(f) - v(f)|.$$

Z is precompact metric and all the values v_θ are 1-Lipschitz. Apply corollary a to obtain the general CV of $(v^\theta)_\theta$

And use the distance d^* and theorem 3 to get the existence of the general uniform value (R-Venel 11-2011).

Z is precompact metric and all the values v_θ are 1-Lipschitz. Apply corollary a to obtain the general CV of $(v^\theta)_\theta$

And use the distance d^* and theorem 3 to get the existence of the general uniform value (R-Venel 11-2011).

Let K be finite, $X = \Delta(K)$ endowed with $\|.\|_1$.
We define:
$D = \{f : X \to I\!R, \forall p\, f(p) = Val(\sum_k p^k G^k)$ for some matrices $G^1, ..., G^K$ with values in $[-1,1]\}$,
and

$$D' = \{f : X \to I\!R, \forall a, b \geq 0, \forall x, y \in X, af(x) - b(y) \leq \|ax - by\|\}.$$

We have $D \subset D' \subset Lip_1$.

$$\begin{aligned} d^*(u,v) \quad &=_{def} \quad \sup_{f \in D} |u(f) - v(f)| \\ &= \quad \sup_{f \in D'} |u(f) - v(f)| \\ &= \quad \inf_{(P,Q) \in R(u,v)} \left(\int \int \|P(x,y)x - Q(x,y)y\| du(x) dv(y) \right) \end{aligned}$$

where $R(u,v) =$

$$\left\{ (P,Q) : X^2 \to [0,1], \int_y P(x,y) dv(y) = 1 \; u \; a.s. \text{ and } \int_x Q(x,y) d(ux) = 1 \; v \; a.s \right\}.$$

Then for each finite S, the map $\Psi : \Delta(K \times S) \to \Delta_f(X)$ is non expansive.

Let K be finite, $X = \Delta(K)$ endowed with $\|.\|_1$.
We define:
$D = \{f : X \to I\!R, \forall p\, f(p) = Val(\sum_k p^k G^k)$ for some matrices $G^1,\ldots,G^K$ with values in $[-1,1]\}$,
and

$$D' = \{f : X \to I\!R, \forall a, b \geq 0, \forall x, y \in X, af(x) - b(y) \leq \|ax - by\|\}.$$

We have $D \subset D' \subset Lip_1$.

$$\begin{aligned} d^*(u,v) &=_{def} \sup_{f\in D}|u(f) - v(f)| \\ &= \sup_{f\in D'}|u(f) - v(f)| \\ &= \inf_{(P,Q)\in R(u,v)}\left(\int\int \|P(x,y)x - Q(x,y)y\| du(x) dv(y)\right) \end{aligned}$$

where $R(u,v) =$

$$\left\{(P,Q) : X^2 \to [0,1], \int_y P(x,y)dv(y) = 1 \; u \; a.s. \text{ and } \int_x Q(x,y)d(ux) = 1 \; v \; a.s\right\}.$$

Then for each finite S, the map $\Psi : \Delta(K \times S) \to \Delta_f(X)$ is non expansive.

4.d. Application to repeated games with an informed controller

General zero-sum repeated game. $\Gamma(\pi)$

- Five non empty and finite sets

a set of states: K,
sets of actions: I for player 1, and J for player 2,
sets of signals: C for player 1, and D for player 2.

- an initial distribution $\pi \in \Delta(K \times C \times D)$,

a payoff function g from $K \times I \times J$ to $[0,1]$,
and a transition q from $K \times I \times J$ to $\Delta(K \times C \times D)$.

At stage 1: (k_1, c_1, d_1) is selected according to π, player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses i_1 in I and player 2 chooses j_1 in J. The payoff for player 1 is $g(k_1, i_1, j_1)$.

At any stage $t \geq 2$: (k_t, c_t, d_t) is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, player 1 learns c_t and player 2 learns d_t. Simultaneously, player 1 chooses i_t in I and player 2 chooses j_t in J. The stage payoff for player 1 is $g(k_t, i_t, j_t)$.

4.d. Application to repeated games with an informed controller

General zero-sum repeated game. $\Gamma(\pi)$

• Five non empty and finite sets
a set of states: K,
sets of actions: I for player 1, and J for player 2,
sets of signals: C for player 1, and D for player 2.

• an initial distribution $\pi \in \Delta(K \times C \times D)$,
a payoff function g from $K \times I \times J$ to $[0,1]$,
and a transition q from $K \times I \times J$ to $\Delta(K \times C \times D)$.

At stage 1: (k_1, c_1, d_1) is selected according to π, player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses i_1 in I and player 2 chooses j_1 in J. The payoff for player 1 is $g(k_1, i_1, j_1)$.

At any stage $t \geq 2$: (k_t, c_t, d_t) is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, player 1 learns c_t and player 2 learns d_t. Simultaneously, player 1 chooses i_t in I and player 2 chooses j_t in J. The stage payoff for player 1 is $g(k_t, i_t, j_t)$.

4.d. Application to repeated games with an informed controller

General zero-sum repeated game. $\Gamma(\pi)$

- Five non empty and finite sets

a set of states: K,
sets of actions: I for player 1, and J for player 2,
sets of signals: C for player 1, and D for player 2.

- an initial distribution $\pi \in \Delta(K \times C \times D)$,

a payoff function g from $K \times I \times J$ to $[0,1]$,
and a transition q from $K \times I \times J$ to $\Delta(K \times C \times D)$.

At stage 1: (k_1, c_1, d_1) is selected according to π, player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses i_1 in I and player 2 chooses j_1 in J. The payoff for player 1 is $g(k_1, i_1, j_1)$.

At any stage $t \geq 2$: (k_t, c_t, d_t) is selected according to $q(k_{t-1}, i_{t-1}, j_{t-1})$, player 1 learns c_t and player 2 learns d_t. Simultaneously, player 1 chooses i_t in I and player 2 chooses j_t in J. The stage payoff for player 1 is $g(k_t, i_t, j_t)$.

A pair of behavioral strategies (σ,τ) induces a probability over plays. The n-stage payoff for player 1 is:

$$\gamma_n^\pi(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=1}^{n} g(k_t,i_t,j_t)\right).$$

The n-stage value exists:

$$v_n(\pi) = \sup_\sigma \inf_\tau \gamma_n^\pi(\sigma,\tau) = \inf_\tau \sup_\sigma \gamma_n^\pi(\sigma,\tau).$$

Definition The repeated game $\Gamma(\pi)$ has a uniform value if:
• $(v_n(\pi))_n$ has a limit $v(\pi)$ as n goes to infinity,
• Player 1 can uniformly guarantee this limit:
$\forall \varepsilon > 0$, $\exists \sigma$, $\exists n_0$, $\forall n \geq n_0$, $\forall \tau$, $\gamma_n^\pi(\sigma,\tau) \geq v(\pi) - \varepsilon$,
• Player 2 can uniformly guarantee this limit:
$\forall \varepsilon > 0$, $\exists \tau$, $\exists n_0$, $\forall n \geq n_0$, $\forall \sigma$, $\gamma_n^\pi(\sigma,\tau) \leq v(\pi) + \varepsilon$.

A pair of behavioral strategies (σ, τ) induces a probability over plays. The n-stage payoff for player 1 is:

$$\gamma_n^{\pi}(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=1}^{n} g(k_t, i_t, j_t)\right).$$

The n-stage value exists:

$$v_n(\pi) = \sup_{\sigma}\inf_{\tau}\, \gamma_n^{\pi}(\sigma,\tau) = \inf_{\tau}\sup_{\sigma}\, \gamma_n^{\pi}(\sigma,\tau).$$

Definition The repeated game $\Gamma(\pi)$ has a uniform value if:

- $(v_n(\pi))_n$ has a limit $v(\pi)$ as n goes to infinity,
- Player 1 can uniformly guarantee this limit:

$\forall \varepsilon > 0$, $\exists \sigma$, $\exists n_0$, $\forall n \geq n_0$, $\forall \tau$, $\gamma_n^{\pi}(\sigma,\tau) \geq v(\pi) - \varepsilon$,

- Player 2 can uniformly guarantee this limit:

$\forall \varepsilon > 0$, $\exists \tau$, $\exists n_0$, $\forall n \geq n_0$, $\forall \sigma$, $\gamma_n^{\pi}(\sigma,\tau) \leq v(\pi) + \varepsilon$.

Hypothesis HX: Player 1 is informed, in the sense that he can always deduce the state and player 2's signal from his own signal.

(formally, there exists $\hat{k} : C \longrightarrow K$ and $\hat{d} : C \longrightarrow D$ such that:
$\pi(E) = 1$, and $q(k,i,j)(E) = 1$, $\forall (k,i,j) \in K \times I \times J$,
where $E = \{(k,c,d) \in K \times C \times D, \hat{k}(c) = k$ and $\hat{d}(c) = d\}$.)

HX does not imply that P1 knows the actions played by P2.

Hypothesis HY: Player 1 controls the transition, in the sense that the marginal of the transition q on $K \times D$ does not depend on player 2's action.

HX and HY are satisfied in the models of - Repeated games with lack of information on one side (Aumann Maschler 1966), - Markov chain games with lack of information on one side (Renault 2006),- Stochastic games with a single controller and incomplete information on the side of his opponent (Rosenberg Solan Vieille 2004).

Given $m \geq 0$ and $n \geq 1$, define the payoffs and auxiliary value functions:

$$\gamma^{\pi}_{m,n}(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=m+1}^{m+n} g(k_t, i_t, j_t)\right),$$

$$v_{m,n}(\pi) = \sup_{\sigma}\inf_{\tau}\, \gamma^{\pi}_{m,n}(\sigma,\tau) = \inf_{\tau}\sup_{\sigma}\, \gamma^{\pi}_{m,n}(\sigma,\tau).$$

Thm (R, MOR 2011): Under HX and HY, the repeated game $\Gamma(\pi)$ has a Cesàro-uniform value, which is:

$$v^*(\pi) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(\pi).$$

And $(v_n)_n$ uniformly converges to v^* on $\{\pi, \pi(E)=1\}$. Player 1 has ε-optimal strategies. Player 2 has 0-optimal strategies.
And there is general uniform convergence of the value functions, and general uniform value (R-Venel 11-2011).

Given $m \geq 0$ and $n \geq 1$, define the payoffs and auxiliary value functions:

$$\gamma^{\pi}_{m,n}(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=m+1}^{m+n} g(k_t, i_t, j_t)\right),$$

$$v_{m,n}(\pi) = \sup_\sigma \inf_\tau \gamma^{\pi}_{m,n}(\sigma,\tau) = \inf_\tau \sup_\sigma \gamma^{\pi}_{m,n}(\sigma,\tau).$$

Thm (R, MOR 2011): Under HX and HY, the repeated game $\Gamma(\pi)$ has a Cesàro-uniform value, which is:

$$v^*(\pi) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(\pi).$$

And $(v_n)_n$ uniformly converges to v^* on $\{\pi, \pi(E)=1\}$. Player 1 has ε-optimal strategies. Player 2 has 0-optimal strategies.
And there is general uniform convergence of the value functions, and general uniform value (R-Venel 11-2011).

Given $m \geq 0$ and $n \geq 1$, define the payoffs and auxiliary value functions:

$$\gamma^{\pi}_{m,n}(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=m+1}^{m+n} g(k_t, i_t, j_t)\right),$$

$$v_{m,n}(\pi) = \sup_{\sigma}\inf_{\tau}\,\gamma^{\pi}_{m,n}(\sigma,\tau) = \inf_{\tau}\sup_{\sigma}\,\gamma^{\pi}_{m,n}(\sigma,\tau).$$

Thm (R, MOR 2011): Under HX and HY, the repeated game $\Gamma(\pi)$ has a Cesàro-uniform value, which is:

$$v^*(\pi) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(\pi).$$

And $(v_n)_n$ uniformly converges to v^* on $\{\pi, \pi(E)=1\}$. Player 1 has ε-optimal strategies. Player 2 has 0-optimal strategies.
And there is general uniform convergence of the value functions, and general uniform value (R-Venel 11-2011).

Given $m \geq 0$ and $n \geq 1$, define the payoffs and auxiliary value functions:

$$\gamma^{\pi}_{m,n}(\sigma,\tau) = I\!E_{I\!P_{\pi,\sigma,\tau}}\left(\frac{1}{n}\sum_{t=m+1}^{m+n} g(k_t, i_t, j_t)\right),$$

$$v_{m,n}(\pi) = \sup_{\sigma}\inf_{\tau} \gamma^{\pi}_{m,n}(\sigma,\tau) = \inf_{\tau}\sup_{\sigma} \gamma^{\pi}_{m,n}(\sigma,\tau).$$

Thm (R, MOR 2011): Under HX and HY, the repeated game $\Gamma(\pi)$ has a Cesàro-uniform value, which is:

$$v^*(\pi) = \inf_{n\geq 1}\sup_{m\geq 0} v_{m,n}(\pi) = \sup_{m\geq 0}\inf_{n\geq 1} v_{m,n}(\pi).$$

And $(v_n)_n$ uniformly converges to v^* on $\{\pi, \pi(E) = 1\}$. Player 1 has ε-optimal strategies. Player 2 has 0-optimal strategies.
And there is general uniform convergence of the value functions, and general uniform value (R-Venel 11-2011).

Aumann, R.J. and M. Maschler (1995):
Repeated games with incomplete information. With the collaboration of R. Stearns.
Cambridge, MA: MIT Press.

A. Araposthathis, V. Borkar, E. Fernández-Gaucherand, M. Ghosh and S. Marcus.
Discrete-time controlled Markov Processes with average cost criterion: a survey. *SIAM Journal of Control and Optimization*, 31, 282–344, 1993.

M. Arisawa and P.L. Lions
On ergodic stochastic control.
Com. in partial differential equations, 23, 2187–2217, 1998.

P. Bettiol
On ergodic problem for Hamilton-Jacobi-Isaacs equations
ESAIM: Cocv, 11, 522–541, 2005.

D. Blackwell.
Discrete dynamic programming,

Annals of Mathematical Statistics, 33, 719–726, 1962.

Coulomb, J.M. (2003):
Games with a recursive structure. based on a lecture of J-F. Mertens.
Chapter 28, Stochastic Games and Applications, A. Neyman and S. Sorin eds, Kluwer Academic Publishers.

L. Dubins and L. Savage.
How to gamble if you must: inequalities for stochastic porcesses.
McGraw-Hill, 1965. 2nd edition 1976 Dover, New York.

E.B. Dynkin and A.A. Yushkevich.
Controlled Markov Processes,
Springer, 1979.

O. Hernández-Lerma, et J.B. Lasserre.
Long-Run Average-Cost Problems.
Discrete-Time Markov Control Processes, Ch. 5, 75–124, 1996.

E. Lehrer et D. Monderer.
Discounting versus Averaging in Dynamic Programming.

Games and Economic Behavior, 6, 97–113, 1994.

E. Lehrer et S. Sorin.
A uniform Tauberian Theorem in Dynamic Programming.
Mathematics of Operations Research, 17, 303–307, 1992.

S. Lippman.
Criterion Equivalence in Discrete Dynamic Programming.
Operations Research, 17, 920–923, 1969.

J.-F. Mertens.
Repeated games.
Proceedings of the International Congress of Mathematicians, Berkeley 1986, 1528–1577. *American Mathematical Society*, 1987.

J.-F. Mertens et A. Neyman.
Stochastic games,
International Journal of Game Theory, 1, 39-64, 1981.

D. Monderer et S. Sorin.
Asymptotic properties in Dynamic Programming.
International Journal of Game Theory, 22, 1–11, 1993.

M. Quincampoix and J. Renault.
On the existence of a limit value in some non expansive optimal control problems.SICON 49, pp 2118-2132, October 2011.

M. Quincampoix and F. Watbled
Averaging methods for discontinuous Mayer's problem of singularly perturbed control systems.
Nonlinar analysis, 54, 819–837, 2003.

J. Renault.
The value of Markov chain games with lack of information on one side.
Mathematics of Operations Research, 3, 490–512, 2006.

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

J. Renault.
The value of Repeated Games with an informed controller.

arXiv : 0803.3345. to appear in MOR.

D. Rosenberg, E. Solan et N. Vieille.
Blackwell Optimality in Markov Decision Processes with Partial Observation.
The Annals of Statistics, 30, 1178–1193, 2002.

Rosenberg, D., Solan, E. and N. Vieille (2004):
Stochastic games with a single controller and incomplete information.
SIAM Journal on Control and Optimization, 43, 86-110.

Sorin, S. (1984):
Big match with lack of information on one side (Part I),
International Journal of Game Theory, 13, 201-255.

Sorin, S. and S. Zamir (1985):
A 2-person game with lack of information on 1 and 1/2 sides.
Mathematics of Operations Research, 10, 17-23.

S. Sorin.
A First Course on Zero-Sum Repeated Games.

Mathématiques et Applications, Springer, 2002.