Long-term values in Markov Decision Processes and Repeated Games, and a new distance for probability spaces.

Jérôme Renault[*], Xavier Venel [†]

Abstract

We study long-term Markov Decision Processes and Gambling Houses, with applications to any partial observation MDPs with finitely many states and zero-sum repeated games with an informed controller. We consider a decision-maker which is maximizing the weighted sum $\sum_{t\geq 1}\theta_t r_t$, where r_t is the expected reward of the t-th stage. We prove the existence of a very strong notion of long-term value called *general uniform value*, representing the fact that the decision-maker can play well independently of the evaluations $(\theta_t)_{t\geq 1}$ over stages, provided the total variation (or impatience) $\sum_{t\geq 1}|\theta_{t+1}-\theta_t|$ is small enough. This result generalizes previous results of Rosenberg, Solan and Vieille [35] and Renault [31] that focus on arithmetic means and discounted evaluations. Moreover, we give a variational characterization of the general uniform value via the introduction of appropriate invariant measures for the decision problems, generalizing the fundamental theorem of gambling or the Aumann-Maschler cav$u$ formula for repeated games with incomplete information.

Apart the introduction of appropriate invariant measures, the main innovation in our proofs is the introduction of a new metric d_* such that partial observation MDP's and repeated games with an informed controller may be associated to auxiliary problems that are non-expansive with respect to d_*. Given two Borel probabilities over a compact subset X of a normed vector space, we define $d_*(u,v)=\sup_{f\in D_1}|u(f)-v(f)|$, where D_1 is the set of functions satisfying: $\forall x,y\in X, \forall a,b\geq 0,\ af(x)-bf(y)\leq \|ax-by\|$. The particular case where X is a simplex endowed with the L^1-norm is particularly interesting: d_* is the largest distance over the probabilities with finite support over X which makes every disintegration non-expansive. Moreover, we obtain a Kantorovich-Rubinstein type duality formula for $d_*(u,v)$ involving couples of measures (α,β) over $X\times X$ such that the first marginal of α is u and the second marginal of β is v.

Keywords: Gambling Houses, Partial observation Markov decision processes, Repeated games, Wasserstein metric, Uniform value, Characterization of the value.

1 Introduction

The standard model of Markov Decision Processes (or Controlled Markov chains) was introduced by Bellman [6] in the 1950s and has been extensively studied since then. In this model, a decision-maker perfectly observes at the beginning of every stage what is the current state, and chooses an action accordingly. The current state and the selected action induce a stage payoff and the law of the next state. The most common ways to aggregate the infinite stream of expected payoff into a global evaluation are the discounted evaluation (where the stream of payoffs is evaluated with a discount factor) and the finite evaluation (where the stream of payoffs is evaluated by the arithmetic average of the payoffs of the first T stages). More generally,

[*] TSE (Université Toulouse 1 Capitole), 21 allée de Brienne, 31000 Toulouse, France. E-mail: jerome.renault@ut-capitole.fr

[†] Paris School of Economics - University Paris 1, 106-110 Boulevard de de l'hôpital, 75647 Paris Cedex 13, France E.mail: xavier.venel@univ-paris1.fr

for any sequence $\theta = (\theta_t)_{t\geq 1}$, with $\theta_t \geq 0$ and $\sum_{t\geq 1} \theta_t = 1$, one can define the θ-evaluation as the problem where the payoff at stage t has a weight θ_t. The supremum over the strategies of the decision maker of the expected payoff under a θ-evaluation is then called the θ-value.

There are two traditional ways to study long term values in MDP. The first one called the asymptotic approach focuses on the convergence to the same limit of the discounted values when the discount factor goes to 0 and of the finite values when the number of stages goes to ∞. Whenever this joint limit exists, it is called the limit value. The second one called the uniform approach is stronger than the limit value and focuses on the possibility for the decision maker to guarantee the asymptotic value in any sufficiently long game. The value is then called the uniform value. When the sets of states and actions are both finite, Blackwell [9] proved the existence of a strategy which is optimal for all discount factors close to 0 (sufficiently patient). It implies the existence of the limit value and of the uniform value. Denardo and Fox [15] and Hordjik and Kallenberg [19] (see Remark 4.7) later provided a characterization of the limit value by the introduction of the "Average Cost Optimality Equation".

This model was generalized into several directions. Some authors tried to relax the finiteness assumption under some ergodicity conditions (Runggaldier and Stettner [37], Borkar [10],[11], see Arapostathis *et al.* [1] for a survey of the different results and techniques). Under such conditons, the limit value is independent of the initial state and satisfies the "Average Cost Optimailty Criterion". Contrary to a large part of the literature, we won't assume any ergodicity condition. The model was also generalized to POMDPs where the decision-maker no longer observes the current state. At the beginning of each stage the decision maker receives a signal which depends on the previous and current states and on his previous action. In order to study such problem a natural approach is to go back to the standard model of MDPs with full observation on the new state, with new state space the space of beliefs on the original state (see Astrom, K.J. [3], Sawaragi and Yoshikawa [38] and Rhenius [34]). Using the structure of the belief state, Rosenberg *et al.* [35] proved the existence of the uniform value in POMDPs when the sets of states, actions and signals are finite. Renault [31] gave another proof and extended their result by removing the finiteness assumption on signals and actions.

Shapley [39] introduced an extension of standard MDPs to 2-player called stochastic games: the state variable is now simultaneously controlled by 2 players having opposite interests. Both the notion of limit value and uniform value can be defined in this new framework. Whenever states and actions are finite, the existence of the limit value is due to Bewley and Kohlberg [7]. A few years later, Mertens and Neyman [24] proved in this setup the existence of the uniform value. Naturally the model of stochastic games also generalizes to partial information (see Mertens [26]), but the existence of possible private information for the different players implies a very complex structure on the auxiliary state space. Following Harsanyi [18], Mertens and Zamir [25] introduced the universal belief space which synthesizes all the information for both players in a general repeated game: their beliefs about the state, their beliefs about the beliefs of the other player, etc. So far, the results of the literature always concern some subclasses of games where we can explicitly write the auxiliary game in a "small" tractable set. A lot of work has especially been done about games with one fully informed player and one player with partial information, and we will only consider such games here. In the simplest model introduced by Aumann and Maschler (see reference from 1995), a state is initially chosen and remains fixed for the rest of the game. Renault [30] extended the analysis to a general underlying Markov chain on the state space (see also Neyman, [28]). Renault [32] proved the existence of the uniform value when the informed player can additionally control the evolution of the state variable (see also Rosenberg *et al.* [36]).

In this paper, we study the existence of long-term values with respect to a more general set of evaluations than finite and discounted evaluations. We will consider the case of a "patient" decision-maker optimizing in the long term, in the sense that given an evaluation $\theta = (\theta_t)_t$, its total variation, $TV(\theta) = \sum_{t\geq 1} |\theta_{t+1} - \theta_t|$ is sufficiently small. We say that the MDP has a general limit value if v_θ converges uniformly as $TV(\theta)$ goes to 0. We say that it has a general uniform value if the limit value exists and for each $\varepsilon > 0$, there exists a strategy which is ε-optimal for each evaluation θ such that $TV(\theta)$ is sufficiently small (so that if $TV(\theta)$ is small, it is possible to play well without knowing exactly θ). For MDP with finitely many states and actions, Blackwell's result [9] immediatly implies the existence of these stronger notions of values. For POMDPs and repeated games with an informed controller, Rosenberg *et al.* [35], Renault [31] and Renault [33] only prove positive results concerning discounted and finite evaluations. We will study here four different models: Gambling Houses introduced by Dubins and Savage [16], Markov Decision Processes (where contrary to gambling houses, there is an explicit set of actions), Partial Observation Decision Processes and Repeated Games. Moreover, we will provide a new characterization of the limit value with the introduction of appropriate invariant measures.

In Section 1, we first consider Gambling Houses defined by $\Gamma = (X, F, r)$, where X is the state space, $r : X \to [0,1]$ is the running payoff and $F : X \rightrightarrows \Delta_f(X)$ is the transition multifunction. Given an initial state x_0 in X, a decision-maker, or player, has to choose u_1 in $F(x_0)$, then x_1 is selected according to u_1 and there is a payoff $r(x_1)$, etc. We show in Theorem 2.10 that if X is metric compact, r is continuous and F is non-expansive with respect to the Kantorovitch-Rubinstein metric, then the problem has a general uniform value v^* characterized by:

$$\begin{aligned}\forall x \in X,\ v^*(x) \quad &= \inf \quad \big\{ w(x), w : \Delta(X) \to [0,1] \text{ affine continuous s.t.} \\ &\qquad (1)\ \forall y \in X, w(y) \geq \sup_{u \in F(y)} w(u) \text{ and } (2)\ \forall u \in R, w(u) \geq r(u)\ \big\}.\end{aligned}$$

where R is a suitably defined set of invariant measures for the Gambling House Γ (see Definition 2.9). Hence in addition of strengthening the usual existence result, we also provide a characterization of the limit value. Note that no ergodicity condition is assumed, and in general the limit value v^* does depend on the initial state.

Unfortunately, this result does not allow us to go beyond and to study Partial Observation Markov Decision Processes or Repeated Games with an informed controller, where the basic state space is a finite set K. The natural state space to study these problems becomes the simplex $X = \Delta(K)$ of probabilities over K, and one can try to apply Theorem 2.10 to the induced Gambling House. However it turns out that the KR distance is not small enough to make the transitions non-expansive (see example 3.12 later), so even in simple cases of incomplete information the hypotheses of Theorem 2.10 won't be satisfied. Section 2 is dedicated to the introduction and the study of a new metric, which will be well adapted to the study of POMDPs and Repeated games with an informed controller in Section 3 and Section 4. Given X a subset of a normed vector space, we introduce the pseudo-metric over $\Delta(X)$ given by:

$$d_*(u, v) = \sup_{f \in D_1} |u(f) - v(f)|,$$

where D_1 is the set of functions satisfying: $\forall x, y \in X, \forall a, b \geq 0,\ af(x) - bf(y) \leq \|ax - by\|$. This metric is smaller than the KR metric and induces the same topology. We prove in Theorem 3.5 a first duality theorem expressing d_* as the infimum on some couplings. When $X = \Delta(K)$ is a simplex endowed with the L^1-norm, we show that d_* is a metric metrizing the weak-* topology and describe several equivalent formulations of this metric. In Theorem 3.10, we provides a duality theorem for probabilities with finite support: given $u, v \in \Delta_f(X)$,

$$d_*(u, v) = \min_{(\alpha,\beta) \in \mathcal{M}_4(u,v)} \sum_{(x,y) \in U \times V} \|x\alpha(x, y) - y\beta(x, y)\|,$$

where $\mathcal{M}_4(u, v)$ is the set of couples (α, β) of probability measures on supp $(u) \times$ supp (v) such that the first marginal of α is u and the second marginal of β is v. Finally, one can characterize the metric d_* by introducing disintegrations of probabilities. For any finite set S, define the disintegration mapping ψ_S on $\Delta(K \times S)$ by $\psi_S(\pi) = \sum_{s\in S} \pi(s)\delta_{p(s)}$ where for each s, $p(s)$ is the posterior on K given s. The mapping ψ_S is 1-Lipschitz from $(\Delta(K \times S), \|.\|_1)$ to $(\Delta_f(X), d_*)$. Moreover, d_* exactly is the largest distance such that any of these mappings is 1-Lipschitz (Theorem 3.13). This is a desirable property (not shared by the KR metric), since π contains some information on s, and certainly more information than $\psi_S(\pi)$.

In Section 3, we use the metric d_* in order to study standard MDPs that we will later use in Section 4. A standard MDPs Ψ is given by a set of states X, a non empty set of actions A, a mapping $q : X \times A \to \Delta_f(X)$ and a payoff function $g : X \times A \to [0, 1]$. At each stage, the player learns the current state x and chooses an action a. He then receives the payoff $g(x, a)$, a new state is drawn accordingly to $q(x, a)$ and the game proceeds to the next stage. We assume in Theorem 4.5 that X is a compact subset of a simplex $\Delta(K)$, and moreover that $\forall x \in X, \forall y \in X, \forall a \in A, \forall f \in D_1, \forall\alpha \geq 0, \forall\beta \geq 0$, $|\alpha f(q(x, a)) - \beta f(q(y, a))| \leq \|\alpha x - \beta y\|_1$ and $|\alpha g(x, a) - \beta g(y, a)| \leq \|\alpha x - \beta y\|_1$. Then we prove that Ψ has a general uniform value v^* characterized by: for all x in X,

$$\begin{aligned} v^*(x) \quad &= \inf \quad \big\{ w(x), w : \Delta(X) \to [0, 1] \text{ affine continuous } s.t. \\ &\qquad (1)\ \forall x' \in X, w(x') \geq \sup_{a\in A} w(q(x', a)) \text{ and } (2)\ \forall (u, y) \in RR, w(u) \geq y \ \big\}, \end{aligned}$$

where RR is a suitably defined set of *invariant couples* for the MDP Ψ (see Definition 4.4). This result can in particular be applied when X is finite and provides an alternative formulation for the limit value in the simplest case (see Remark 4.7). To prove Theorem 4.5, we use the properties of the metric d_* introduced in Section 2.

Finally, in Section 4 we apply the result of Section 3 to any POMDP with a finite set of states (without assumptions on the set of actions), and to any repeated game with an informed controller with finitely many states and actions. Hence proving the existence of the general uniform value in these models. Finally, we recall an open problem showing the difficulty to compute v^* in general.

2 Long-term values for Gambling Houses

In this section, we study Gambling Houses that are non-expansive for the Kantorovitch-Rubinstein metric and we prove the existence of the general uniform value.

Given X a compact metric set, we denote by $\mathcal{C} = \mathcal{C}(X)$ the set of continuous functions from X to the reals, and by $\mathcal{C}_1$ the set of 1-Lipschitz functions in $\mathcal{C}$. We denote by $\Delta(X)$ the set of Borel probability measures on X, by $\Delta_f(X)$ the set of Borel probability measure with finite supports and for each x in X we write δ_x for the Dirac probability measure on x. It is well known that $\Delta(X)$ is compact for the weak-* topology, and this topology can be metrizable by the (Wasserstein) Kantorovich-Rubinstein distance:

$$\forall u, v \in \Delta(X),\ \ d_{KR}(u, v) = \sup_{f\in\mathcal{C}_1}\ u(f) - v(f).$$

The standard Kantorovich duality formula reads (see e.g. Villani 2003, p.207):

$$d_{KR}(u, v) = \sup_{f\in\mathcal{C}_1}\ |u(f) - v(f)| = \min_{\gamma\in\Pi(u,v)} \int_{(x,y)\in X\times X} d(x, y)\ d\gamma(x, y),$$

where $\Pi(u, v)$ denotes the set of transference plans, or couplings, of u and v, that is the set of probability distributions over $X \times X$ with first marginal u and second marginal v.

2.1 Model

The model is the following. There is a 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 (or player) 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 choose u_2 in $F(x_1)$, x_2 is selected according to u_2 and the player receives the payoff $r(x_2)$, etc. Note that there is no explicit action set here, and that the transitions take values in $\Delta_f(X)$ and hence have finite support.

We say that $\Gamma = (X, F, r)$ is a Gambling House. We identify the elements in X with their Dirac measures in $\Delta(X)$, and in case the values of F only consist of Dirac measures on X, we view F as a correspondence from X to X and say that Γ is a *deterministic* Gambling House (or a Dynamic Programming problem). An element of $\Delta_f(X)$ is written $u = \sum_{x\in X} u(x)\delta_x$. The set of stages is $I\!N^* = \{1, ..., t, ...\}$, and a probability distribution over stages is called an evaluation. Given an evaluation $\theta = (\theta_t)_{t\geq 1}$ and an initial stage x_0 in X, the θ-problem $\Gamma_\theta(x_0)$ is the optimization problem defined by a decision-maker starting from x_0 and maximizing the expectation of $\sum_{t\geq 1} \theta_t r(x_t)$.

Formally, we first linearly extend r and F to $\Delta_f(X)$ by defining for each $u = \sum_{x\in X} u(x)\delta_x$ in $\Delta_f(X)$, the payoff $r(u) = \sum_{x\in X} r(x)u(x)$ and the transition by

$$F(u) = \left\{ \sum_{x\in X} u(x)f(x), s.t.\ f : X \to \Delta_f(X) \text{ and } f(x) \in F(x),\ \forall x \in X \right\}.$$

Definition 2.1. *The mixed extension of F is the correspondence from $\Delta_f(X)$ to itself which associates to every $u = \sum_{x\in X} u(x)\delta_x$ in $\Delta_f(X)$ the image:*

$$\hat{F}(u) = \left\{ \sum_{x\in X} u(x)f(x),\ s.t.\ f : X \to \Delta_f(X) \text{ and } f(x) \in \operatorname{conv}F(x)\ \forall x \in X \right\}.$$

The graph of $\hat{F}$ is the convex hull of the graph of F. Moreover $\hat{F}$ is an affine correspondence, as shown by the lemma below whose proof can be found in the appendix.

Lemma 2.2. $\forall u, u' \in \Delta_f(X)$, $\forall \alpha \in [0,1]$, $\hat{F}(\alpha u + (1-\alpha)u') = \alpha\hat{F}(u) + (1-\alpha)\hat{F}(u')$.

Definition 2.3. *A pure play, or deterministic play, at x_0 is a sequence $\sigma = (u_1, ..., u_t, ...) \in \Delta_f(X)^\infty$ such that $u_1 \in F(x_0)$ and $u_{t+1} \in F(u_t)$ for each $t \geq 1$. A play, or mixed play, at x_0 is a sequence $\sigma = (u_1, ..., u_t, ...) \in \Delta_f(X)^\infty$ such that $u_1 \in \operatorname{conv}F(x_0)$ and $u_{t+1} \in \hat{F}(u_t)$ for each $t \geq 1$. We denote by $\Sigma(x_0)$ the set of mixed plays at x_0.*

A pure play is a particular case of a mixed play. Mixed plays corresponds to situations where the decision-maker can select, at every stage t and state x_{t-1}, *randomly* the law u_t of the new state. A mixed play at x_0 naturally induces a probability distribution over the set $(X \times \Delta_f(X))^\infty$ of sequences $(x_0, u_0, x_1, u_1, ...)$, where X and $\Delta_f(X)$ are endowed with the discrete σ-algebra and $(X \times \Delta_f(X))^\infty$ is endowed with the product σ-algebra. We do not need any measurability assumption since the range of F is $\Delta_f(X)$, hence any strategy generates only a countable set of states.

Definition 2.4. *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:*

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

It is easy to see that the supremum in the definition of v_θ can be taken over the set of pure plays at x_0. We extend linearly v_θ to $\Delta_f(X)$ by defining for each $u = \sum_{x\in X} u(x)\delta_x$, $v_\theta(u) =$

$\sum_{x\in X} u(x)v_\theta(x)$. We have the following recursive formula. For each evaluation $\theta = (\theta_t)_{t\geq 1}$ such that $\theta_1 < 1$, denote by θ^+ the "shifted" evaluation $\left(\frac{\theta_{t+1}}{1-\theta_1}\right)_{t\geq 1}$ then the recursive formula reads:

$$\forall \theta \in \Delta(I\!N^*), \forall x \in X,\ v_\theta(x) = \sup_{u\in \mathrm{conv} F(x)} \left(\theta_1 r(u) + (1-\theta_1) v_{\theta^+}(u)\right).$$

By linearity the supremum can be taken over $F(x)$. It is also easy to see that for all evaluations θ and initial states x, we have the inequality:

$$|v_\theta(x) - \sup_{u\in F(x)} v_\theta(u)| \leq \theta_1 + \sum_{t\geq 2} |\theta_t - \theta_{t-1}|. \tag{1}$$

In this paper, we are interested in the limit behavior when the decision-maker is very patient. Given an evaluation θ, we define the total variation of θ by:

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

The decision-maker is considered as patient whenever $TV(\theta)$ is small, so $TV(\theta)$ may be seen as the impatience of θ (see Renault [33] and Sorin [40] p. 105). When $\theta = (\theta_t)_{t\geq 1}$ is non increasing, then $TV(\theta)$ is just θ_1. A classic example is when $\theta = \frac{1}{n}\sum_{t=1}^{n} \delta_t$, the value v_θ is just denoted v_n and the evaluation corresponds to the average payoff from stage 1 to stage n. In this case $TV(\theta) = 1/n \xrightarrow[n\to\infty]{} 0$. We also have $TV(\theta) = 1/n$ if $\theta = \sum_{t=m+1}^{m+n} \frac{1}{n}\delta_t$ for some non-negative m. Another example is the case of discounted payoffs where $\theta = (\lambda(1-\lambda)^{t-1})_{t\geq 1}$ for some discount factor $\lambda \in (0,1]$. In this case the value v_θ is denoted v_λ and $TV(\theta) = \lambda \xrightarrow[\lambda\to 0]{} 0$.

Definition 2.5. *The Gambling House $\Gamma = (X, F, r)$ has a* general limit value v^* *if (v_θ) uniformly converges to v^* when $TV(\theta)$ goes to zero, i.e.:*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta,\ \left(\ TV(\theta) \leq \alpha \implies (\forall x \in X, |v_\theta(x) - v^*(x)| \leq \varepsilon)\ \right).$$

The existence of the general limit value implies in particular that $(v_n)_n$ and $(v_\lambda)_\lambda$ converge to the same limit when n goes to $+\infty$ and λ goes to 0. This is coherent with the result of Lehrer and Sorin [21], which states that the uniform convergence of $(v_n)_n$ and $(v_\lambda)_\lambda$ are equivalent. A recent characterization of the uniform convergence of a sequence of value functions $(v_{\theta^k})_k$, when $TV(\theta^k) \xrightarrow[k\to\infty]{} 0$, can be found in Renault [33], and it is shown that all such sequences have a unique possible limit point given by $v^* = \inf_{\theta\in\Delta(I\!N^*)} \sup_{m\geq 0} v_{m,\theta}$, where $v_{m,\theta}$ is the value corresponding to the evaluation with weight 0 for the first m stages and with weight θ_{t-m} for stages $t > m$.

In the definition of the general limit value, we require all value functions to be close to v^* when the patience is high, but the plays used may depend on the precise expression of θ. In the following definition, we require the same play to be simultaneously optimal for all θ patient enough.

Definition 2.6. *The Gambling House $\Gamma = (X, F, r)$ has a* general uniform value *if it has a general limit value v^* and moreover for each $\varepsilon > 0$ one can find $\alpha > 0$ and for each initial state x a mixed play $\sigma(x)$ at x satisfying:*

$$\forall \theta,\ \left(\ TV(\theta) \leq \alpha \implies (\forall x \in X, \gamma_\theta(\sigma(x)) \geq v^*(x) - \varepsilon)\ \right).$$

The literature has mainly focused on the evaluations $\theta = \sum_{t=1}^{n} \frac{1}{n}\delta_t$ and $\theta = (\lambda(1-\lambda)^{t-1})_{t\geq 1}$. The standard (Cesàro)-uniform value can be defined by restricting the evaluations to be Cesàro means: for each $\varepsilon > 0$ one can find n_0 and for each initial state x a mixed play $\sigma(x)$ at x satisfying: $\forall n \geq n_0, \forall x \in X, \gamma_n(\sigma(x)) \geq v^*(x) - \varepsilon$. Recently, Renault [31] considered deterministic Gambling Houses and characterized the uniform convergence of the value functions $(v_n)_n$. The existence of the standard Cesàro-uniform value is proved under some assumptions, including the case where the set of states X is metric precompact, the transitions are non-expansive and the payoff function is uniformly continuous. As a corollary, the existence of the uniform value is shown in Partial Observation Markov Decision Processes with finite set of states (after each stage the decision-maker just observes a stochastic signal possibly correlated to the new state).

2.2 Result

We now present our main theorem for Gambling Houses. Equation (1) implies that the general limit value v^* necessarily has to satisfy some rigidity property. The linear extension of the function v^* to $\Delta_f(X)$ can only be an "excessive function" in the terminology of potential theory [14] and Gambling Houses (Dubins and Savage [16], Maitra and Sudderth [22]).

Definition 2.7. *An affine function w defined on $\Delta_f(X)$ (or $\Delta(X)$) is said to be excessive if for all x in X, $w(x) \geq \sup_{u \in F(x)} w(u)$.*

Example 2.8. Let us consider the splitting transition given by a finite set K, $X = \Delta(K)$ and for each x in X, $F(x) = \{u \in \Delta(X), \ \sum_{p \in X} u(p)\ p = x\}$ is the set of probabilities on X centered at x. Then the function w from $\Delta_f(X)$ or $\Delta(X)$ to $I\!R$ is excessive if and only if the restriction of w to X is concave. Moreover given $u, u' \in \Delta(X)$, $u' \in \hat{F}(u)$ if and only if u' is a sweeping of u as defined by Choquet [14]: for all continuous concave functions f from X to $I\!R$, $u'(f) \leq u(f)$.

Assume now that X is a compact metric space. Any continuous function f on X can be extended naturally into an affine continuous function on $\Delta(X)$ by $f(u) = \int_{x \in X} f(x) du(x)$ for all Borel probabilities on X. In particular the payoff function r is naturally extended to an affine continuous function on $\Delta(X)$ that we will still denote by r. In the following definition, we consider the closure of the graph of $\hat{F}$ within the (compact) set $\Delta(X \times X)$.

Definition 2.9. *An element u in $\Delta(X)$ is said to be an invariant measure of the Gambling House $\Gamma = (X, F, r)$ if $(u, u) \in \mathrm{cl}(Graph\ \hat{F})$. The set of invariant measures of Γ is denoted by R, so that:*

$$R = \{u \in \Delta(X), (u, u) \in \mathrm{cl}(Graph\ \hat{F})\}.$$

R is a convex compact subset of $\Delta(X)$. Even when Γ is deterministic, we still need to work in the space $\Delta(X)$ of probabilities over X to define invariant measures. Morally, we have replaced the time averages by the space averages.

Theorem 2.10. *Consider a Gambling House $\Gamma = (X, F, r)$ such that X is a compact metric space, r is continuous from X to $[0,1]$ and F is non-expansive with respect to the Kantorovich-Rubinstein distance, i.e. $\forall x \in X, \forall x' \in X, \forall u \in F(x), \exists u' \in F(x')$ s.t. $d_{KR}(u, u') \leq d(x, x')$.*

Then the Gambling House Γ has a general uniform value v^* characterized by:

$$\begin{aligned} \forall x \in X,\ v^*(x) \quad = \inf \quad & \big\{ w(x), w : \Delta(X) \to [0,1] \text{ affine continuous s.t.} \\ & (1)\ \forall y \in X, w(y) \geq \sup_{u \in F(y)} w(u) \text{ and } (2)\ \forall u \in R, w(u) \geq r(u)\ \big\}. \end{aligned}$$

That is, the affine extension of v^* to $\Delta(X)$ is the smallest continuous affine function which is 1) *excessive and* 2) *above the running payoff r on invariant measures.*

The proof of Theorem 2.10 is in the Appendix. Notice that:
1) when $\Gamma = (X, F, r)$ is deterministic, the hypotheses are satisfied as soon as X is metric compact for some metric d, r is continuous and F is non-expansive for d.
2) when X is finite, one can use the distance $d(x, x') = 2$ for all $x \neq x'$ in X, so that for u and u' in $\Delta(X)$, $d_{KR}(u, u') = \|u - u'\|_1 = \sum_{x \in X} |u(x) - u'(x)|$, and the hypotheses are automatically satisfied. We will prove later a more general result for a model of MDP with finite state space, allowing for explicit actions influencing transitions and payoffs (see Corollary 4.6).

Remark 2.11. The non-expansivity condition in order to prove the existence of the limit value and/or of the uniform value was already introduced in Renault [31] for Gambling Houses. Such a condition was also used in a continuous time framework by Quincampoix and Renault [29] and Buckdahn et al. [13].

Remark 2.12. The formula also holds when there is no decision-maker, i.e. when F is single-valued, and there are some similarities with the Von Neumann's ergodic theorem [41]. Let Z be a Hilbert space and Q be a linear isometry on Z, this theorem states that for all $z \in Z$, the sequence $z_n = \frac{1}{n}\sum_{t=1}^{n} Q^t(z)$ converges to the projection z^* of z on the set R of fixed points of Q. Using the linearity and the non-expansiveness leads to a characterization by the set of fixed points. In particular, having in mind linear payoff functions of the form $(z \mapsto < l, z >)$, we have that the projection z^* of z on R is characterized by:

$$\forall l \in \Delta_f(X), < l, z^* >=< l^*, z >= \inf\{< l', z >, l' \in R \text{ and } < l', r > \geq < l, r > \forall r \in R\}.$$

Example 2.13. We consider here a basic periodic sequence of 0 and 1. Let $X = \{0, 1\}$ and for all $x \in X$, $F(x) = \{1 - x\}$ and $r(x) = x$. There is a unique invariant measure $u = 1/2\delta_0 + 1/2\delta_1$, and the general uniform value exists and satisifes $v^*(x) = \frac{1}{2}$ for all states x. Notice that considering evaluations $\theta = (\theta_t)_t$ such that θ_t is small for each t without requiring $TV(\theta)$ small, would not necessarily lead to v^*. Consider for instance $\theta^n = \sum_{t=1}^{n} \frac{1}{n}\delta_{2t}$ for each n, we have $v_{\theta^n}(x) = x$ for all x in X.

Example 2.14. The state space is the unit circle, let $X = \{x \in \mathbb{C}, |x| = 1\}$ and $F(e^{i\alpha}) = e^{i(\alpha+1)}$ for all real α. If we denote by μ the uniform distribution (Haar probability measure) on the circle, the mapping F is μ-ergodic and μ is F-invariant. By Birkhoff's theorem [8], we know that the time average converges to the space average μ-almost surely. Here μ is the unique invariant measure, and we obtain that the general uniform value is the constant:

$$\forall x \in X, \ v^*(x) = \frac{1}{2\pi}\int_0^{2\pi} r(e^{i\alpha})d\alpha.$$

Notice that we obtain the convergence of the value $v_\theta(x)$ to $v^*(x)$ for all x in X, and not only for μ-almost all x in X.

Example 2.15. Let $\Gamma = (X, F, r)$ be an MDP satisfying the hypotheses of Theorem 2.10, and such that for all $x \in X$, $\delta_x \in F(x)$. Therefore the set R is equal to $\Delta(X)$. In the terminology of Gambling Theory (see Maitra Sudderth, [22], Γ is called a *leavable* Gambling House since at each stage the player can stay at the current state. The limit value v^* is here characterized by:

$$v^* = \inf\{v : X \to [0, 1] \text{ continuous}, v \text{ is excessive and } v \geq r\}.$$

In the above formula, v excessive means: $\forall x \in X, v(x) \geq \sup_{u \in F(x)} \mathbb{E}_u(v)$. This is a variant of the *Fundamental Theorem of Gambling Theory* (see section 3.1 in Maitra Sudderth [22]).

Example 2.16. The following deterministic Gambling House, which is an extension of Example 1.4.4. in Sorin [40] and of Example 5.2 of Renault [31], shows that the assumptions of Theorem 2.10 allow for many speeds of convergence to the limit value v^*. Here $l > 1$ is a fixed parameter, X is the simplex $\{x = (p^a, p^b, p^c) \in \mathbb{R}^3_+, p^a + p^b + p^c = 1\}$ and the initial state is $x_0 = (1, 0, 0)$. The payoff is $r(p^a, p^b, p^c) = p^b - p^c$, and the transition is defined by: $F(p^a, p^b, p^c) = \{((1 - \alpha - \alpha^l)p^a, p^b + \alpha p^a, p^c + \alpha^l p^a), \alpha \in [0, 1/2]\}$.

The probabilistic interpretation is the following: there are 3 points a, b and c, and the initial point is a. The payoff is 0 at a, it is +1 at b, and -1 at c. At point a, the decision-maker has to choose $\alpha \in [0, 1/2]$: then b is reached with probability α, c is reached with probability α^l, and the play stays in a with the remaining probability $1 - \alpha - \alpha^l$. When b (resp. c) is reached, the play stays at b (resp. c) forever. So the decision-maker starting at point a wants to reach b and to avoid c. By playing at each stage $\alpha > 0$ small enough, he can get as close to b as he wants.

Back to our deterministic setup, we use norm $\|.\|_1$ and obtain that X is compact, F is non-expansive and r is continuous, so that Theorem 2.10 applies, and the limit value is given by $v^*(p^a,p^b,p^c)=p^a+p^b$. Notice that even though the data are very smooth, there is no 0-optimal strategy here, in the sense that there is no mixed play σ at a such that $\lim_{TV(\theta)\to 0}\gamma_\theta(\sigma)=1$, any good strategy requires to take a positive risk.

If we denote by x_λ the value $v_\lambda(x_0)$, we have for all $\lambda\in(0,1]$: $x_\lambda=\phi_\lambda(x_\lambda)$, where for all $x\in \mathbb{R}$, $\phi_\lambda(x)=\max_{\alpha\in[0,1/2]}(1-\lambda)(1-\alpha-\alpha^l)x+\alpha$. Since $x_\lambda\in(0,1)$, the first order condition gives $(1-\lambda)(-1-l\alpha^{l-1})x_\lambda+1=0$ and we can obtain: $x_\lambda=\frac{1}{(1-\lambda)}\left(l\left(\frac{\lambda}{(1-\lambda)(l-1)}\right)^{\frac{l-1}{l}}+1\right)^{-1}$. Finally we compute an equivalent of $x_\lambda-1$ as λ goes to 0, and obtain: $1-v_\lambda(x_0)\ \sim\ C\lambda^{\frac{l-1}{l}}$ with $C=\frac{l}{(l-1)^{\frac{(l-1)}{l}}}$.

3 A distance for belief spaces

In the previous section, we proved an existence result for Gambling Houses that are non-expansive for the KR-metric. We now want to go beyond and to study Partial Observation Markov Decision Processes or Repeated Games with an informed controller, where the basic state space is a finite set K. The natural state space becomes the simplex $X=\Delta(K)$ of probabilities over K, and one can try to apply Theorem 2.10 to the induced Gambling House. However it turns out that the KR distance is not small enough to make the transitions non-expansive (see example 3.12 later), so even in simple cases of incomplete information the hypotheses of Theorem 2.10 won't be satisfied.

In this section, we introduce and study a new metric, which will be well adapted to our problems. We first introduce a pseudo-metric on the set of probabilities over a compact subset X of a real normed space, and prove a first duality theorem (Theorem 3.5). Then, we focus on the special case where X itself is a probability space over a finite set K and prove our main duality theorem (Theorem 3.10). Finally, we provide a fundamental characterization of our metric as the largest metric compatible with disintegrations over finite sets (Theorem 3.13).

3.1 A pseudo-distance for probabilities on a compact subset of a normed vector space

Let X be a compact subset of a real normed vector space V. Recall that $\mathcal{C}=\mathcal{C}(X)$ denotes the set of continuous functions on X and $\mathcal{C}_1$ the set of 1-Lipschitz functions.

We introduce here a new pseudo-distance on $\Delta(X)$, which is not greater than d_{KR} and in some cases also metrizes the weak-* topology. We start with several definitions, which will turn out to be equivalent. Let u and v be in $\Delta(X)$.

Definition 3.1. $d_1(u,v)=\sup_{f\in D_1}\ u(f)-v(f)$,
where $D_1=\{f\in\mathcal{C},\forall x,y\in X,\forall a,b\geq 0,\ af(x)-bf(y)\leq\|ax-by\|\}$.

Note that any linear functional on X with norm 1 induces an element of D_1. If f is in D_1 then $-f$ is also in D_1, so $d_1(u,v)=\sup_{f\in D_1}\ |u(f)-v(f)|$ and d_1 is a pseudo-distance on $\Delta(X)$. We also have $D_1\subset\mathcal{C}_1$, so that $d_1(u,v)\leq d_{KR}(u,v)$ and the supremum in the definition of $d_1(u,v)$ is achieved. Given x and y in X, it is known that $d_{KR}(\delta_x,\delta_y)=\|x-y\|$ and the supremum in the definition of d_{KR} is reached by a linear functional on X. It follows that $d_1(\delta_x,\delta_y)=\|x-y\|$.

Notice that $D_1=\{f\in\mathcal{C},\forall x,y\in X,\forall a,b\geq 0,\ |af(x)-bf(y)|\leq\|ax-by\|\}$. If f, g are in D_1 then $\sup\{f,g\}$ and $\inf\{f,g\}$ also are in D_1, and D_1 is a convex lattice with greatest element $(x\mapsto\|x\|)$ and smallest element $(x\mapsto-\|x\|)$. If $0\in X$, then all f in D_1 satisfy $f(0)=0$.

Example 3.2. First consider the particular case where $X = [0,1]$ endowed with the usual norm. Then all f in D_1 are linear. As a consequence, $d_1(u,v) = 0$ for $u = 1/2\,\delta_0 + 1/2\,\delta_1$ and $v = \delta_{1/2}$. We do not have the separation property and d_1 is not a distance in this case[1].

Let us modify the example. X now is the set of probability distributions over 2 elements, viewed as $X = \{(x, 1-x), x \in [0,1]\}$. We use the norm $\|.\|_1$ to measure the distance between $(x, 1-x)$ and $(y, 1-y)$, so that $V = I\!R^2$ is endowed with $\|(x_1,x_2)-(y_1,y_2)\| = |x_1-y_1|+|x_2-y_2|$. Consider f in $\mathcal{C}$ such that $f((x,1-x)) = x(1-x)$ for all x. f now belongs to D_1, and $d_1(u,v) \geq 1/4 > 0$ for $u = 1/2\,\delta_0 + 1/2\,\delta_1$ and $v = \delta_{1/2}$. One can show that $(\Delta(X), d_1)$ is a compact metric space in this case (see Proposition 3.7 later), and for applications in this paper d_1 will be a particularly useful distance whenever X is a simplex $\Delta(K)$ endowed with $\|x-y\| = \sum_{k\in K} |x^k - y^k|$. □

Notice also that the Kantorovich-Rubinstein metric on $\Delta(X)$ only depends on the restriction of the norm $\|.\|$ on the set X. Especially if X is finite and if $\|x - x'\| = 2$ for all $x \neq x' \in X$, then for all $u, v \in \Delta(X)$ we have $d_{KR}(u,v) = \|u - v\|_1$. This is not the case when considering the metric d_1, where two norms on V giving the same metric on X may lead to different pseudo-metrics on $\Delta(X)$.

We now give other expressions for the pseudo-distance d_1.

Definition 3.3. $d_2(u,v) = \sup_{(f,g)\in D_2}\ u(f) + v(g)$,
where $D_2 = \{(f,g) \in \mathcal{C}\times\mathcal{C}, \forall x,y \in X, \forall a,b \geq 0,\ af(x) + bg(y) \leq \|ax - by\|\}$.

Definition 3.4. $d_3(u,v) = \inf_{\gamma\in\mathcal{M}_3(u,v)} \int_{X^2\times[0,1]^2} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu)$,
where $\mathcal{M}_3(u,v)$ *is the set of finite positive measures on* $X^2 \times [0,1]^2$ *such that:* $\forall f \in \mathcal{C}$,

$$\int_{(x,y,\lambda,\mu)\in X^2\times[0,1]^2} \lambda f(x) d\gamma(x,y,\lambda,\mu) = u(f) \text{ and } \int_{(x,y,\lambda,\mu)\in X^2\times[0,1]^2} \mu f(y) d\gamma(x,y,\lambda,\mu) = v(f).$$

Theorem 3.5. *For all* u *and* v *in* $\Delta(X)$, $d_1(u,v) = d_2(u,v) = d_3(u,v)$.

The proof of Theorem 3.5 uses the Hahn-Banach theorem and is an involved elaboration on the proof of the standard Kantorovich duality formula in Dudley ([17], see Lemma 11.8.5 p.423), it can be found in the appendix. Compared to the formula for d_{KR}, our proof requires to handle 2 extra variables λ and μ corresponding to a and b in the definition of D_1 (see Remark 6.9). This will lead to the introduction of the intermediary quantity d_2^+.

We will now use in Subsection 3.2 the duality result $d_1 = d_3$ to prove our main duality formula (Theorem 3.10) in the case where X is a simplex. This duality formula Theorem 3.10, as well as our universal characterization via disintegrations (Theorem 3.13), will only apply to probabilities with finite support (see Remark 3.14).

3.2 The case of probabilities over a simplex

The case where X itself is a probability space is interesting for applications, and we assume here that $X = \Delta(K)$ is a simplex, where K is a non empty finite set. We use $\|p\| = \sum_k |p^k|$ for every vector $p = (p^k)_{k\in K}$ in $I\!R^K$, and view X as the set of vectors in $I\!R_+^K$ with norm 1: $X = \{p = (p^k)_{k\in K} \in I\!R_+^K,\ \sum_{k\in K} p^k = 1\}$.

We now prove that d_1 is a distance on $\Delta(X)$ metrizing the weak-* topology.

Lemma 3.6. *The linear span of* D_1 *is dense in* $\mathcal{C}(X)$.

The proof of Lemma 3.6 is in the appendix. Notice that D_1 itself is not dense in $\mathcal{C}(x)$.

[1] More generally one can show that if there exists $x \neq 0$ such that the segment $[0,x]$ is included in X, then all f in D_1 are linear on $[0,x]$ and $d_1(\delta_{x/2}, 1/2\,\delta_0 + 1/2\,\delta_x) = 0$, so d_1 is not a distance. In the case $X = [-1,1]$ with the usual norm, one can show that $D_1 = \{f \in \mathcal{C}, \exists\ \alpha,\beta \geq 0 \text{ s.t.} f(x) = \alpha x$ for $x \in [-1,0]$ and $f(x) = \beta x$ for $x \in [0,1]\}$.

Proposition 3.7. *d_1 is a distance on $\Delta(X)$ metrizing the weak-* topology.*

Proof: Because the linear span of D_1 is dense in $\mathcal{C}(X)$, we obtain the separation property and d_1 is a distance on $\Delta(X)$. Because $D_1 \subset \mathcal{C}_1$, we have $d_1 \leq d_{KR}$. Since $(\Delta(X), d_{KR})$ is a compact metric space, the identity map from $(\Delta(X), d_{KR})$ to $(\Delta(X), d_1)$ is bicontinuous, and we obtain that $(\Delta(X), d_1)$ is a compact metric space and d_1 and d_{KR} are equivalent. (See for instance proposition 2 page 138 [4]). Therefore d_1 is a distance on $\Delta(X)$ metrizing the weak-* topology.

Remarks 3.8.

1) One can also here give another definition of d_1 by using functions introduced by Aumann and Machler [5] for repeated games with incomplete information.

Given a collection of matrices $(G^k)_{k\in K}$ (all of the same finite size $I \times J$) indexed by K and with values in $[-1, 1]$, we define the "non revealing function" f in $\mathcal{C}(X)$ by:

$$\forall p \in X,\ f(p) = \mathrm{Val}\left(\sum_{k\in K} p^k G^k\right) = \max_{x\in\Delta(I)} \min_{y\in\Delta(J)} \sum_{i\in I, j\in J} x(i)y(j)\left(\sum_{k\in K} p^k G^k(i,j)\right)$$
$$= \min_{y\in\Delta(J)} \max_{x\in\Delta(I)} \sum_{i\in I, j\in J} x(i)y(j)\left(\sum_{k\in K} p^k G^k(i,j)\right).$$

Here Val denotes the minmax value of a matrix, and $f(p)$ is the minmax value of the average matrix $\sum_k p^k G^k$. The set of all such non revealing functions f, where I, J and $(G^k)_{k\in K}$ vary, is denoted by D_0.

By construction, we have $D_0 \subset D_1$. Moreover, one can show that the closure of D_0 is D_1, so that restricting the supremum in Definition 3.1 to functions in D_0 defines the same distance. One can also show that allowing for infinite sets I, J in the definition of D_0 (still assuming that all games $\sum_k p^k G^k$ have a value) would not change the distance. The interest for this type of distances defined through games previously appeared while doing research on Markov Decision Processes with partial observation and repeated games with an informed controller (see Proposition 5.1. chapter VI p. 357 in Mertens *et al.* [27], Renault [31] or [32]).

2) Let LF be the set of linear forms on $(I\!R^K, \|.\|_1)$ with norm at most 1. In the definition of d_1, one can also replace D_1 by the lattice generated by the restrictions to $\Delta(K)$ of the elements of LF.

From now on, we just write $d_*(u, v)$ for the distance $d_1 = d_2 = d_3$ on $\Delta(X)$. Elements of X can be viewed as elements of $\Delta(X)$, and for p, q in X, we have: $d_{KR}(\delta_p, \delta_q) = d_*(\delta_p, \delta_q) = \|p - q\|$. We now present a dual formulation for our distance, in the spirit of the Kantorovich duality formula from optimal transport. We will concentrate on probabilities on X with finite support.

Definition 3.9. *Let u and v be in $\Delta_f(X)$ with respective supports U and V. We define the set $\mathcal{M}_4(u, v) =$*

$$\left\{(\alpha, \beta) \in (I\!R_+{}^{U\times V})^2, s.t.\ \forall x \in U, \sum_{y'\in V} \alpha(x, y') = u(x) \text{ and } \forall y \in V, \sum_{x'\in U} \beta(x', y) = v(y)\right\}.$$

We put $d_4(u, v) = \inf_{(\alpha,\beta)\in\mathcal{M}_4(u,v)} \sum_{(x,y)\in U\times V} \|x\alpha(x, y) - y\beta(x, y)\|.$

$\mathcal{M}_4(u, v)$ is the set of couples (α, β) of probability measures on $U \times V$ such that the first marginal of α is u and the second marginal of β is v. Notice that diagonal elements in $\mathcal{M}_4(u, v)$ coincide with elements of $\Pi(u, v)$, i.e. with probability distributions over $X \times X$ with first marginal u and second marginal v. The set $\mathcal{M}_4(u, v)$ is simply a polytope in the Euclidean space $(I\!R^{U\times V})^2$, so the infimum in the definition of $d_4(u, v)$ is achieved. The next theorem is the main result of this section.

Theorem 3.10. *(Duality formula) Let u and v be in $\Delta_f(X)$ with respective supports U and V.*

$$d_*(u,v) = \sup_{f \in D_1} |u(f) - v(f)| = \min_{(\alpha,\beta)\in \mathcal{M}_4(u,v)} \sum_{(x,y)\in U\times V} \|x\alpha(x,y) - y\beta(x,y)\|.$$

The proof can be found in the appendix. We conclude this part by a simple but universal property of the distance d_*.

Definition 3.11. *Given finite sets S and K, the disintegration, or posterior mapping, ψ_S from $\Delta(K\times S)$ to $\Delta_f(X)$ is defined by:*

$$\psi_S(\pi) = \sum_{s\in S} \pi(s)\delta_{p(s)}$$

where for each s, $\pi(s) = \sum_k \pi(k,s)$ and $p(s) = (p^k(s))_{k\in K} \in X$ is the posterior on K given s (defined arbitrarily if $\pi(s)=0$) : for each k in K, $p^k(s) = \frac{\pi(k,s)}{\pi(s)}$.

$\psi_S(\pi)$ is a probability with finite support over X. Intuitively, think of a joint variable (k,s) being selected according to π, and an agent just observes s. His knowledge on K is then represented by $p(s)$ and $\psi_S(\pi)$ represents the ex-ante information that the agent will know about the variable k.

$\Delta(K\times S)$ is endowed as usual with the $\|.\|_1$ norm. One can show that ψ_S is continuous whenever X is endowed with the weak-* topology. Intuitively, $\psi_S(\pi)$ has less information than π, because the agent does not care about s itself but just on the information about k given by s. So one may hope that the mapping ψ_S is 1-Lipschitz (non-expansive) for a well chosen distance on $\Delta(X)$. The following example shows that the Kantorovich-Rubinstein distance d_{KR} is not appropriate for this.

Example 3.12. Consider the case where $K=\{a,b,c\}$ and $S=\{\alpha,\beta\}$. We denote by π and π' the following laws on $\Delta(K\times S)$:

$$\begin{array}{cc} & S \\ K & \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \\ \frac{1}{4} & 0 \end{pmatrix} \\ & \pi \end{array} \quad \text{and} \quad \begin{array}{c} S \\ \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{1}{4} \end{pmatrix} \\ \pi' \end{array}.$$

Their disintegrations are respectively $\psi_S(\pi) = \frac{1}{2}\ (1/2,0,1/2) + \frac{1}{2}\ (0,1,0)$ and $\psi_S(\pi') = \frac{1}{4}(1,0,0) + \frac{3}{4}\,(0,2/3,1/3)$.

We define the test function $f : \Delta(K) \to [-1,1]$ by: $f(0,1,0) = \frac{1}{3}$, $f\,(1/2,0,1/2) = -\frac{1}{3}$, $f(1,0,0) = \frac{2}{3}$, $f\,(0,2/3,1/3) = 1$. f can be extended to a 1-Lispchitz function on the simplex $\Delta(K)$ by the McShane-Whitney extension theorem (McShane, [23]), or directly by $f(p_a,p_b,p_c) = \max\{1-p_a-|p_b-2/3|-|p_c-1/3|, -1/3+|p_a-p_c|\}$). We have $\|\pi-\pi'\| = \frac{1}{2}$ and $d_{KR}(\psi_S(\pi),\psi_S(\pi')) \geq \psi_S(\pi')(f) - \psi_S(\pi)(f) = \frac{11}{12} - 0 > \frac{1}{2}$. So the posterior mapping ψ_S is not 1-Lipschitz from $(\Delta(K\times S),\|.\|_1)$ to $(\Delta(X),d_{KR})$.

The next theorem shows that our distance d_* has the fundamental property to make all disintegrations ψ_S non-expansive, and is the largest distance to do so.

Theorem 3.13. *For each finite set S, the mapping ψ_S is 1-Lipschitz from $(\Delta(K\times S),\|.\|_1)$ to $(\Delta_f(X),d_*)$. Moreover, d_* is the largest distance on $\Delta_f(X)$ having this property:*

$$\forall u,u' \in \Delta_f(X),\ d_*(u,u') = \min\{\|\pi-\pi'\|_1,\ s.t.\ \exists S \textit{ finite},\ \psi_S(\pi)=u,\ \psi_S(\pi')=u'\}.$$

The proof simply uses the duality formula.

Proof: For clarity, we will precise the space on which each norm is considered. First fix S and π, π' in $\Delta(K \times S)$. Write $u = \psi_S(\pi) = \sum_{s\in S} \pi(s)\delta_{p(s)}$ and $u' = \psi_S(\pi') = \sum_{s\in S} \pi'(s)\delta_{p'(s)}$. Recall that for every $s \in S$ and every $k \in K$, $\pi(k,s) = \pi(s)p^k(s)$ and $\pi'(k,s) = \pi'(s)p'^k(s)$. For any f in D_1, we have

$$u(f) - u'(f) \;=\; \sum_{s\in S} (\pi(s)f(p(s)) - \pi'(s)f(p'(s))\,.$$

By definition of D_1, for every $s \in S$ we have

$$\pi(s)f(p(s)) - \pi'(s)f(p'(s)) \leq \|\pi(s)p(s) - \pi(s)p(s)\|_{1,K}.$$

Therefore,

$$\begin{aligned} u(f) - u'(f) &\leq \sum_{s\in S} \|\pi(s)p(s) - \pi(s)p(s)\|_{1,K} \\ &= \sum_{s\in S} \|(\pi(k,s))_k - (\pi'(k,s))_k\|_{1,K}, \\ &= \sum_{s\in S}\sum_{k\in K} |\pi(k,s) - \pi'(k,s)|, \\ &= \|\pi - \pi'\|_{1,K\times S}. \end{aligned}$$

So $d_*(u,u') \leq \|\pi - \pi'\|_{1,K\times S}$, and ψ_S is 1-Lipschitz.

Let now u and v be in $\Delta_f(X)$ with respective supports U and V. Using the duality formula of Theorem 3.10, one can find $(\alpha,\beta) \in \mathcal{M}_4(u,v)$ such that

$$d_*(u,v) = \sum_{(x,y)\in U\times V} \|\alpha(x,y)x - \beta(x,y)y\|.$$

Define $S = U \times V$ and $\pi, \pi' \in \Delta(K \times S)$ by $\pi(k,(x,y)) = x(k)\alpha(x,y)$ and $\pi'(k,(x,y)) = y(k)\beta(x,y)$. By definition of $\mathcal{M}_4(u,v)$, π and π' are probabilities and

$$\begin{aligned} \|\pi - \pi'\|_{1,K\times S} &= \sum_{k\in K,(x,y)\in U\times V} |x(k)\alpha(x,y) - y(k)\beta(x,y)| \\ &= \sum_{(x,y)\in U\times V} \|\alpha(x,y)x - \beta(x,y)y\|. \end{aligned}$$

□

Finally notice that considering an infinite set S dramatically changes the picture, as shown by the following simple example (mentioned by F. Santambrogio).

Remark 3.14. Fix $K = \{a,b\}$ and $S = [0,1]$, and define for each π in $\Delta(K \times S)$, the image $\psi_S(\pi)$ in $\Delta(X)$ by: $\psi_S(\pi)(f) = \int_{k,s} f(p(s))d\pi(k,s)$ for all f in $\mathcal{C}(X)$ (here again, $X = \Delta(K)$ and $p(s)$ is the posterior on K given S). $\Delta(K \times S)$ and $\Delta(X)$ are endowed with weak-* topologies.

Consider the uniform probability π over $K \times S$, then $\psi_S(\pi) = \delta_{\frac{1}{2}a+\frac{1}{2}b}$ is the Dirac measure on $\frac{1}{2}a + \frac{1}{2}b$. We approximate π by considering finer and finer grids of the unit interval. For each positive integer n, partition $[0,1)$ into $A_n = \cup_{k=0}^{n-1}[\frac{2k}{2n}, \frac{2k+1}{2n})$ and $B_n = \cup_{k=0}^{n-1}[\frac{2k+1}{2n}, \frac{2k+2}{2n})$. We can define π_n by first choosing s in S according to the Lebesgue measure, then set $k = a$ if $s \in A_n$ and set $k = b$ if $s \in B_n$. Knowing s perfectly determines k here, and $\psi_S(\pi_n) = \frac{1}{2}\delta_a + \frac{1}{2}\delta_b$. However π_n converges to π, so ψ_S is not even continuous.

4 Long-term values for standard MDPs

In Section 2, we studied Gambling Houses that are non-expansive for the Kantorovitch-Rubinstein metric. Unfortunately, this model is not adapted to the study of Partial Observation Markov Decision Process since the associated transitions are not 1-Lipschitz for the KR metric.

We consider here a variant of the model of Gambling House that we call standard Markov Decision Processes, where the decision-maker has an explicit action set and where the payoffs may depend both on the current state and action. The main result of this section is Theorem 4.5 which uses the metric d_* defined in Section 3. We will show in section 5 that this theorem applies to the standard MDPs associated to a POMDP or to a repeated game with an informed controller.

A standard Markov Decision Problem Ψ is given by a non empty set of states X, a non empty set of actions A, a mapping $q : X \times A \to \Delta_f(X)$ and a payoff function $g : X \times A \to [0, 1]$. At each stage, the player learns the current state x and chooses an action a. He then receives the payoff $g(x, a)$, a new state is drawn accordingly to $q(x, a)$ and the game proceeds to the next stage.

Definition 4.1. *A pure, or deterministic, strategy is a sequence of mappings $\sigma = (\sigma_t)_{t\geq 1}$ where $\sigma_t : (X \times A)^{t-1} \to A$ for each t. A strategy (or behavioral strategy) is a sequence of mappings $\sigma = (\sigma_t)_{t\geq 1}$ where $\sigma_t : (X \times A)^{t-1} \to \Delta_f(A)$ for each t. We denote by Σ the set of strategies.*

$(X \times A)^0$ is a singleton, so σ_1 is viewed as an element of $\Delta_f(A)$ representing the lottery on actions played at the first stage (or simply if σ is pure, the action in A played at the first stage). A pure strategy is a particular case of strategy. An initial state x_1 in X and a strategy σ naturally induce a probability distribution with finite support over the set of finite histories $(X \times A)^n$ for all n, which can be uniquely extended to a probability over the set $(X \times A)^\infty$ of infinite histories.

Definition 4.2. *Given an evaluation θ and an initial state x_1 in X, the θ-payoff of a strategy σ at x_1 is defined as $\gamma_\theta(x_1, \sigma) = I\!E_{x_1,\sigma}\left(\sum_{t\geq 1} \theta_t g(x_t, a_t)\right)$, and the θ-value at x_1 is:*

$$v_\theta(x_1) = \sup_{\sigma\in\Sigma} \gamma_\theta(x_1, \sigma).$$

As for Gambling Houses, it is easy to see that the supremum can be taken over the smaller set of pure strategies, and one can derive a recursive formula linking the value functions. General limit and uniform values are defined as in Subsection 2.

Definition 4.3. *Let $\Psi = (X, A, q, g)$ be a standard MDP.*

Ψ has a general limit value v^* if (v_θ) uniformly converges to v^* when $TV(\theta)$ goes to zero, i.e. for each $\varepsilon > 0$ one can find $\alpha > 0$ such that:

$$\forall\theta,\ \ (\ TV(\theta) \leq \alpha \implies (\forall x \in X, |v_\theta(x) - v^*(x)| \leq \varepsilon)\ ).$$

Ψ has a general uniform value if it has a general limit value v^*, and if for each $\varepsilon > 0$ one can find $\alpha > 0$ and a behavior strategy $\sigma(x)$ for each initial state x satisfying:

$$\forall\theta,\ \ (TV(\theta) \leq \alpha \implies (\forall x \in X, \gamma_\theta(x, \sigma(x)) \geq v^*(x) - \varepsilon)\ ).$$

We now present a notion of invariance for the MDP Ψ. The next definition will be similar to Definition 2.9, however one needs to be slightly more sophisticated here to incorporate the payoff component. First, we define for every $a \in \Delta_f(A)$ and for every $x \in X$,

$$g(x, a) = \sum_{i\in \text{supp }(a)} a(i)g(x, i).$$

and $q(x,a) \in \Delta(X)$ is the unique probability distribution such that for every $f \in \mathcal{C}(X,[0,1])$,

$$q(x,a)(f) = \sum_{i \in \text{supp }(a)} a(i) q(x,i)(f).$$

Assume now that X is a compact metric space, and define for each (u,y) in $\Delta_f(X) \times [0,1]$,

$$\hat{F}(u,y) = \left\{ \left(\sum_{x \in X} u(x) q(x,a(x)), \sum_{x \in X} u(x) g(x,a(x)) \right), \textit{ where } a : X \to \Delta_f(A) \right\}.$$

We have defined a correspondence $\hat{F}$ from $\Delta_f(X) \times [0,1]$ to itself. It is easy to see that $\hat{F}$ always is an affine correspondence (see Lemma 6.16 later). In the following definition we consider the closure of the graph of $\hat{F}$ within the compact set $(\Delta(X) \times [0,1])^2$, with the weak topology.

Definition 4.4. *An element (u,y) in $\Delta(X) \times [0,1]$ is said to be an invariant couple for the MDP Ψ if $((u,y),(u,y)) \in cl(Graph(\hat{F}))$. The set of invariant couples of Ψ is denoted by RR.*

Our main result for standard MDPs is the following theorem, where X is assumed to be a compact subset of a simplex $\Delta(K)$, with K a finite set. Denote $D_1 = \{f \in \mathcal{C}(\Delta(K)), \forall x, y \in \Delta(K), \forall a, b \geq 0,\ af(x) - bf(y) \leq \|ax - by\|_1\}$, and any f in D_1 is linearly extended to $\Delta(\Delta(K))$. Notice that the set D_1 is a subset of the set of 1-Lipschitz function.

Theorem 4.5. *Let $\Psi = (X, A, q, g)$ be a standard MDP where X is a compact subset of a simplex $\Delta(K)$ with K finite, and such that:*

$$\forall x \in X, \forall y \in X, \forall a \in A, \forall f \in D_1, \forall \alpha \geq 0, \forall \beta \geq 0,$$

$$|\alpha f(q(x,a)) - \beta f(q(y,a))| \leq \|\alpha x - \beta y\|_1 \text{ and } |\alpha g(x,a) - \beta g(y,a)| \leq \|\alpha x - \beta y\|_1.$$

Then Ψ has a general uniform value v^* characterized by: for all x in X,

$$\begin{aligned} v^*(x) \quad = \inf \quad & \big\{ w(x), w : \Delta(X) \to [0,1] \text{ affine continuous } s.t. \\ & (1)\ \forall x' \in X, w(x') \geq \sup_{a \in A} w(q(x',a)) \text{ and } (2)\ \forall (u,y) \in RR, w(u) \geq y\ \big\}. \end{aligned}$$

The proof can be found in the appendix. The key argument is the use of the new metric d_* over probability space on the simplex defined in Section 3 and Subsection 3.2. Under our assumptions, the standard MDP is equivalent to a Gambling House which is non-expansive for this metric d_*. Moreover d^* satisfies the duality Theorem 3.10, and we can adapt the proof of Theorem 2.10.

When the state space is finite, we have an immediate corollary of the above theorem.

Corollary 4.6. *Consider a standard MDP (K, A, q, g) with a finite set of states K. Then it has a general uniform value v^*, and for each state k:*

$$\begin{aligned} v^*(k) \quad = \inf \quad & \big\{ w(k), w : \Delta(K) \to [0,1] \text{ affine } \ s.t. \\ & (1)\ \forall k' \in K, w(k') \geq \sup_{a \in A} w(q(k',a)) \text{ and } (2) \forall (p,y) \in RR, w(p) \geq y\ \big\}. \end{aligned}$$

$$\textit{with } \ RR = \{(p,y) \in \Delta(K) \times [0,1], ((p,y),(p,y)) \in \mathrm{cl}(conv(Graph(F)))\}$$

and $F(k,y) = \{(q(k,a), g(k,a)), a \in A\}$.

Proof of Corollary 4.6: K is viewed as a subset of the simplex $\Delta(K)$, endowed with the L^1-norm. Fix k, k' in K, a in A, $\alpha \geq 0$ and $\beta \geq 0$. We have $\|\alpha k - \beta k'\| = |\alpha - \beta|$ if $k = k'$, and $\|\alpha k - \beta k'\| = \alpha + \beta$ otherwise.

First we have

$$|\alpha g(k,a) - \beta g(k',a)| \leq \begin{cases} |\alpha - \beta| g(k,a) \text{ if } k = k' \\ \alpha + \beta \text{ otherwise} \end{cases} \quad ,$$

so in all cases $|\alpha g(k,a) - \beta g(k',a)| \leq \|\alpha k - \beta k'\|$. Secondly, consider $f \in D_1$. f takes values in $[-1,1]$, so similarly we have: $|\alpha f(q(k,a)) - \beta f(q(k',a))| \leq \|\alpha k - \beta k'\|$.

So we can apply Theorem 4.5, and the graph of $\hat{F}$ is the convex hull of the graph of F.

Remark 4.7. When both sets of states and actions are finite, we are in the simple setting of Blackwell [9]. In this case, one can deduce the existence of the general uniform value from the proof of Blackwell. Our theorem yields in addition a characterization. This characteriation is the dual formulation of a result of Denardo and Fox [15]. We say that a couple $(w,h) \in I\!R^K \times I\!R^K$ is superharmonic, in the sense of Hordjik and Kallenberg [19], if

$$\forall k \in K, \forall a \in A, \ w(k) + h(k) \geq g(k,a) + \sum_{k' \in K} q(k,a)(k')h(k'). \tag{2}$$

Denardo and Fox [15] showed that the value v^* is the smallest (pointwise) excessive function that can be completed by a function h such that (v^*, h) is superharmonic.

The existence of a function h such that (w,h) is superharmonic is equivalent to condition (2) of Corollary 4.6. Given a function w, solving equation (2) is a linear programming problem with $K \times A$ inequalities. By Farkas' lemma, it has a solution if and only if the following linear programming problem (D_w), with unknown $\pi \in I\!R^{K \times A}$, has no solution:

$$\begin{array}{rrl} \forall (k,a) \in K \times A & \pi(k,a) & \geq 0 \\ \forall k \in K & \sum_{a' \in A} \pi(k,a') & = \sum_{k' \in K, a' \in A} \pi(k',a') q(k',a')(k) \\ & \sum_{k' \in, a' \in A} \pi(k',a') g(k',a') & > \sum_{k' \in K, a' \in A} \pi(k',a') w(k'). \end{array}$$

Fix w a function on K. We prove that (D_w) has a solution if and only if condition (2) of Corollary 4.6 is not satisfied. Therefore both conditions are equivalent.

Let π be a solution of (D_w). We can assume without loss of generality that $\sum_{k,a} \pi(k,a) = 1$. We denote by p the marginal of π on K and put for all $k \in K$, $\sigma(k) = \left(\frac{\pi(k,a')}{p(k)}\right)_{a' \in A} \in \Delta(A)$ if $p(k) > 0$, and define arbitrarily $\sigma(k)$ if $p(k) = 0$. Denote by σ the strategy which plays $\sigma(k)$ if the state is k for each k. The payoff obtained by playing σ from distribution p is $y = \sum_{k' \in, a' \in A} \pi(k',a') g(k',a')$. The second line of equations of (D_w) implies that p is invariant by σ, so $(p,y) \in RR$ and the last equation implies that $y > w(p)$. The function w does not satisfy condition (2) of Corollary 4.6.

Conversally if condition (2) of Corollary 4.6 is not satisfied, there exists $(p,y) \in RR$ such that $y > w(p)$. By compactness of the set of probabilities over K and the set of payoffs, there exists a strategy $\sigma \in \Delta(A)^K$, such that p is invariant under σ and y is the payoff obtained by playing σ from distribution p. The probability π defined by: $\forall k \in K, \forall a \in A, \pi(k,a) = p(k)\sigma(k,a)$, is then a solution of (D_w).

Notice that Denardo and Fox also use duality theory but they study directly the minimization problem with unknown w and h and deduce a dual maximization problem.

5 Applications to partial observation and games

In this section, we use Theorem 4.5 to prove the existence of the general uniform value in any POMDP with finite set of states (without assumptions on the set of actions), and in any repeated game with an informed controller with finitely many states and actions. Subsection 5.1 is dedicated to POMDPs and Subsection 5.2 to repeated games with an informed controller. In both cases, we apply Theorem 4.5.

5.1 POMDP with finitely many states

We now consider a more general model of MDP with actions where after each stage, the decision-maker does not perfectly observe the state. An MDP with partial observation, or POMDP, $\Gamma = (K, A, S, q, g)$ includes a finite set of states K, a non empty set of actions A and a non empty set of signals S. The transition q now goes from $K \times A$ to $\Delta_f(S \times K)$ and the payoff function g still goes from $K \times A$ to $[0, 1]$. Given an initial probability p_1 on K, the POMDP $\Gamma(p_1)$ is played as follows. An initial state k_1 in K is selected according to p_1 and is not told to the decision-maker. At every stage $t \geq 1$, the decision-maker selects an action $a_t \in A$. If the current state is k_t, he has a (unobserved) payoff $g(k_t, a_t)$ and a pair (s_t, k_{t+1}) is drawn according to $q(k_t, a_t)$. Then the player learns s_t, and the play proceeds to stage $t + 1$ with new state k_{t+1}. A behavioral strategy is now a sequence $(\sigma_t)_{t\geq 1}$ of applications with for each t, $\sigma_t : (A \times S)^{t-1} \to \Delta_f(A)$. As usual, an initial probability on K and a behavior strategy σ induce a probability distribution over $(K \times A \times S)^\infty$ and we can define the θ-values and the notions of general limit and uniform values accordingly.

Theorem 5.1. *A POMDP with finitely many states has a general uniform value, i.e. there exists $v^* : \Delta(K) \to I\!R$ with the following property: for each $\varepsilon > 0$ one can find $\alpha > 0$ and for each initial probability p a behavior strategy $\sigma(p)$ such that for each evaluation θ with $TV(\theta) \leq \alpha$,*

$$\forall p \in \Delta(K), |v_\theta(p) - v^*(p)| \leq \varepsilon \text{ and } \gamma_\theta(\sigma(p)) \geq v^*(p) - \varepsilon.$$

Proof of Theorem 5.1: It is natural to introduce an auxiliary MDP with state variable the belief of the decision-maker on the state in K. We define Ψ the standard MDP on $X = \Delta(K)$ with the same set of actions A and the following payoff and transition functions:

- $r : X \times A \longrightarrow [0, 1]$ s.t. $r(p, a) = \sum_{k\in K} p(k)g(k, a)$ for all p, a,
- $\hat{q} : X \times A \to \Delta_f(X)$ such that $\hat{q}(p, a) = \sum_{s\in S} q(p, a)(s)\delta_{\chi(p,a,s)}$, where $q(p, a)(s) = \sum_k p^k q(k, a)(s)$ and $\chi(p, a, s) \in \Delta(K)$ is the belief on the new state after playing a at p and observing the signal s: $\forall k' \in K, \chi(p, a, s)(k') = \frac{q(p,a)(k',s)}{q(p,a)(s)} = \frac{\sum_k p^k q(k,a)(k',s)}{\sum_k p^k q(k,a)(s)}$.

The POMDP $\Gamma(p_1)$ and the standard MDP $\Psi(p_1)$ have the same value for all θ-evaluations. For each strategy σ in $\Psi(p_1)$, the player can guarantee the same payoff in the original game $\Gamma(p_1)$ by mimicking the strategy σ. So if we prove that Ψ has a general uniform value it will imply that the POMDP Γ has a general uniform value.

To conclude the proof, we will simply apply Theorem 4.5 to the MDP Ψ. We need to check the assumptions on the payoff and on the transition.

Consider any p, p' in X, $a \in A$, $\alpha \geq 0$ and $\beta \geq 0$. We have:

$$|\alpha r(p, a) - \beta r(p', a)| = \left| \sum_k (\alpha p(k) - \beta p'(k))g(k, a) \right| \leq \|\alpha p - \beta p'\|.$$

Moreover for any $f \in D_1$, we have:

$$\begin{aligned}
|\alpha\hat{q}(p, a)(f) - \beta\hat{q}(p', a)(f)| &= \left| \sum_{s\in S} \alpha q(p, a)(s)f(\chi(p, a, s)) - \sum_{s\in S} \beta q(p', a)(s)f(\chi(p', a, s)) \right| \\
&\leq \sum_s \|\alpha q(p, a)(., s) - \beta q(p', a)(., s)\| \\
&\leq \sum_{s,k,k'} |\alpha p(k')q(k', a)(k, s) - \beta p'(k')q(k', a)(k, s)| \\
&\leq \sum_{s,k,k'} q(k', a)(k, s)|\alpha p(k') - \beta p'(k')| = \|\alpha p - \beta p'\|.
\end{aligned}$$

where the first inequality comes from the definition of D_1.

By Theorem 4.5, the MDP Ψ has a general uniform value and we deduce that the POMDP Γ has a general uniform value. □

Example 5.2. Let $\Gamma = (K, A, S, q, g, p_1)$ be a POMDP where $K = \{k_1, k_2\}$, $A = \{a, b\}$, $S = \{*\}$ and $p_1 = \delta_{k_1}$. The initial state is k_1 and since there is only one signal, the decision-maker will obtain no additional information on the state. We say that he is in the dark. The payoff is given by $g(k_1, a) = g(k_1, b) = g(k_2, b) = 0$ and $g(k_2, a) = 1$, and the transition by $q(k_1, a) = q(k_1, b) = \delta_{*,k_1}$, $q(k_2, a) = \delta_{*,k_2}$ and $q(k_1, b) = \frac{1}{2}\delta_{*,k_1} + \frac{1}{2}\delta_{*,k_2}$. On one hand if the decision-maker plays a then the state stays the same, and he receives a payoff of 1 if and only if the state is k_2. On the other hand if he plays b then he receives a payoff of 0 but the probability to be in state k_2 increases.

We define the function r from $X \times A = \Delta(K) \times A$ to $[0, 1]$ by $r((p, 1-p), a) = 1 - p$ and $r((p, 1-p), b) = 0$ for all $p \in [0, 1]$, and we define the transition $\hat{q}$ from $X \times A$ to $\Delta_f(X)$ by :

$$\hat{q}((p, 1-p), a) = \delta_{(p,1-p)} \text{ and } \hat{q}((p, 1-p), b) = \delta_{(p/2,1-p/2)}.$$

Then the standard MDP $\Psi = (\Delta(K), A, \hat{q}, r)$ is the MDP associated in the previous proof to Γ. This MDP is here deterministic, because the decision is in the dark.

The existence of a general uniform value is immediate here. Given $n \geq 1$, the strategy $\sigma = b^n a^\infty$ which plays n times b and then a for the rest of the game, guarantees a stage payoff of $(1 - \frac{1}{2^n})$ from stage $n + 1$ on, so the game has a general uniform value equal to 1. Finally if we consider the discounted evaluations, one can show that the speed of convergence of v_λ is here slower than λ :

$$v_\lambda(p_1) = 1 - \frac{\ln(\lambda)}{\ln(2)}\lambda + O(\lambda).$$

The partial observation allows for a speed of convergence slower than λ contrary to the perfect observation case where it is well known that the convergence is in $O(\lambda)$.

Remark 5.3. It is here unknown if the uniform value exists in pure strategies, i.e. if the behavior strategies $\sigma(p)$ of Theorem 5.1 can be chosen with values in A. This was already an open problem for the Cesàro-uniform value, that is when only evaluations of the form $\theta = \frac{1}{n}\sum_{t=1}^n \delta_t$ are considered (see Rosenberg *et al.* [35] and Renault [31] for different proofs requiring the use of behavioral strategies). In the present proof, there are two related places where the use of lotteries on actions is important. First in the proof of the convergence of the function $h_{T,n}$ (within the proof of Theorem 2.10), we used Sion's theorem in order to exchange a supremum and an infimum, and to do so the convexity of the set of strategies was required. Secondly when we prove that the extended transition is 1-Lipschitz (see Lemma 6.16), the coupling between the two distributions u and u' requires some randomization.

5.2 Zero-sum repeated games with an informed controller

We finally consider zero-sum repeated games with an informed controller. We start with a general model $\Gamma = (K, I, J, C, D, q, g)$ of zero-sum repeated game, where we have 5 non empty finite sets: a set of states K, two sets of actions I and J and two sets of signals C and D, and we also have a transition mapping q from $K \times I \times J$ to $\Delta(K \times C \times D)$ and a payoff function g from $K \times I \times J$ to $[0, 1]$. Given an initial probability π on $\Delta(K \times C \times D)$, the game $\Gamma(\pi) = \Gamma(K, I, J, C, D, q, g, \pi)$ is played as follows: at stage 1, a triple (k_1, c_1, d_1) is drawn according to π, player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses an action i_1 in I and player 2 chooses an action j_1 in J. Player 1 gets a (unobserved) payoff $r(k_1, i_1, j_1)$ and player 2 the opposite payoff. Then a new triple (k_2, c_2, d_2) is drawn accordingly to $q(k_1, i_1, j_1)$. Player 1 observes c_2, player 2 observes d_2 and the game proceeds to the next stage, etc.

A (behavioral) strategy for player 1 is a sequence $\sigma = (\sigma_t)_{t\geq 1}$ where for each $t \geq 1$, σ_t is a mapping from $(C \times I)^{t-1} \times C$ to $\Delta(I)$. Similarly a strategy for player 2 is a sequence of mappings $\tau = (\tau_t)_{t\geq 1}$ where for each $t \geq 1$,τ_t is a mapping from $(D \times J)^{t-1} \times D$ to $\Delta(J)$. We denote respectively by Σ and $\mathcal{T}$ the set of strategies of player 1 and player 2. An initial distribution π and a couple of strategies (σ, τ) define for each t a probability on the possible histories up to stage t, which can be uniquely extended to a probability on the set of infinite histories $(K \times C \times D \times I \times J)^{+\infty}$.

Given an evaluation θ, we define the θ-payoff of (σ, τ) in $\Gamma(\pi)$ as the expectation under $\mathbb{P}_{\pi,\sigma,\tau}$ of the payoff function,

$$\gamma_\theta(\pi, \sigma, \tau) = \mathbb{E}_{\pi,\sigma,\tau} \left(\sum_t \theta_t \, r(k_t, i_t, j_t) \right).$$

By Sion's theorem the game with θ-payoff has a value: $v_\theta(\pi) = \max_{\sigma\in\Sigma} \min_{\tau\in\mathcal{T}} \gamma_\theta(\pi, \sigma, \tau) = \min_{\tau\in\mathcal{T}} \max_{\sigma\in\Sigma} \gamma_\theta(\pi, \sigma, \tau)$, and we can define the general limit value as in the MDP framework. Note that we do not ask the convergence to be uniform for all π in $\Delta(K \times C \times D)$, because we will later make some assumptions, in particular on the initial distribution.

Definition 5.4. *The repeated game $\Gamma(\pi) = (K, I, J, C, D, q, g, \pi)$ has a general limit value $v^*(\pi)$ if $v_\theta(\pi)$ converges to $v^*(\pi)$ when $TV(\theta)$ goes to zero, i.e.:*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta, \ \left(TV(\theta) \le \alpha \implies (|v_\theta(\pi) - v^*(\pi)| \le \varepsilon) \right).$$

Definition 5.5. *The repeated game $\Gamma(\pi)$ has a general uniform value if it has a general limit value $v^*(\pi)$ and for each $\varepsilon > 0$ one can find $\alpha > 0$ and a couple of strategies σ^* and τ^* such that for all evaluations θ with $TV(\theta) \le \alpha$:*

$$\forall \tau \in \mathcal{T}, \gamma_\theta(\pi, \sigma^*, \tau) \ge v^*(\pi) - \varepsilon \ \text{ and } \ \forall \sigma \in \Sigma, \gamma_\theta(\pi, \sigma, \tau^*) \le v^*(\pi) + \varepsilon.$$

Without further assumption, the general values may fail to exist. We will focus here on the case of a repeated game with an informed controller, as introduced in Renault [32]. The first assumption concerns the information of the first player. We assume that he can always reconstruct the current state and the signal of the second player from his own signal:

Assumption 5.6. *There exist two mappings $\widetilde{k} : C \to K$ and $\widetilde{d} : C \to D$ such that, if E denotes $\{(k, c, d) \in K \times C \times D, \ \widetilde{k}(c) = k, \ \widetilde{d}(c) = d\}$, we have: $\forall (k, i, j) \in K \times I \times J$, $q(k, i, j)(E) = 1$, and $\pi(E) = 1$.*

Moreover we will assume that only player 1 has a meaningful influence on the transitions, in the following sense.

Assumption 5.7. *The marginal of the transition on $K \times D$ is not influenced by player 2's action. For k in K, i in I and j in J, we denote by $\bar{q}(k, i)$ the marginal of $q(k, i, j)$ on $K \times D$.*

The second player may influence the signal of the first player but he cannot prevent him either from learning his state or from learning his own signal. Moreover he cannot influence his own information, thus he has no influence on his beliefs about the state or about the beliefs of player 1 about his beliefs. A repeated game satisfying assumptions 5.6 and 5.7 is called a repeated game with an informed controller. It was proved in Renault [32] that for such games the Cesàro-uniform value (that is, when only evaluations of the form $\theta = \frac{1}{n}\sum_{t=1}^n \delta_t$ are considered) exists and we will extend it here to the general uniform value.

Example 5.8. We consider the simplest case of zero-sum repeated game with incomplete information introduced by Aumann and Maschler in the sixties (see reference [5]). It is defined by a finite family $(G^k)_{k\in K}$ of payoff matrices in $[0,1]^{I\times J}$ and $p \in \Delta(K)$ an initial probability. At the first stage, some state k is selected according to p and told to player 1 only. The second player knows the initial distribution p but not the realization of the state. Then the matrix game G^k is repeated over and over. At each stage the players observe past actions but not their payoff (notice that player 1 can always reconstruct the payoff from the actions and the state). Formally it is a zero-sum repeated game $\Gamma = (K, I, J, C, D, q, g)$ as defined previously, with $C = K \times I \times J$ and $D = I \times J$, and for all $(k, i, j) \in K \times I \times J$, $g(k, i, j) = G^k(i, j)$ and $q(k, i, j) = \delta_{k,(k,i,j),(i,j)}$. For all $p \in \Delta(K)$, we denote by $\Gamma(p)$ the game where the initial probability $\pi \in \Delta(K \times C \times D)$ is given by $\pi = \sum_{k \in K} p(k)\delta_{k,(k,i_0,j_0),(i_0,j_0)}$ with $(i_0, j_0) \in I \times J$ fixed.

For each n, we denote by $v_n(p)$ the value of the n-stage game with initial probability p, where the payoff is the expected average of the first n payoffs. The value satisfies the standard recursive formula:

$$v_n(p) = \sup_{a \in \Delta(I)^K} \left(\frac{1}{n} R(p, a) + \frac{n-1}{n} \sum_{i \in I} a(p)(i) v_{n-1}(\chi(p, a, i)) \right),$$

where $a^k \in \Delta(I)$ represents the lottery on actions played by player 1 if the state is k, $a(p)(i) = \sum_{k \in K} p^k a^k(i)$ is the probability that player 1 plays i, $R(p, a) = \min_j(\sum_k p^k G^k(a^k, j))$ is the minimal expected payoff for player 1, and $\chi(p, a, i)$ is the conditional belief on $\Delta(K)$ given p, a, i:

$$\chi(p, a, i) = \left(\frac{p(k)a^k(i)}{a(p)(i)} \right)_k .$$

Starting from a belief p about the state, if player 2 observes action i and knows that the distribution of actions of player 1 is a, then he updates his beliefs to $\chi(p, a, i)$. Aumann and Maschler have proved that the limit value exists and is characterized by

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

where $f^*(p) = Val\left(\sum_k p^k G^k\right)$ for all $p \in \Delta(K)$. The function f^* is the value of the game, called the non-revealing game, where player 1 is forbidden to use his information.

Theorem 5.9. *A zero-sum repeated game with an informed controller has a general uniform value.*

The proof, to be found in the appendix, will consist of 5 steps. First we introduce an auxiliary standard Markov Decision Process $\Psi(\hat{\pi})$ on the state space $X = \Delta(K)$. Then we show that for all evaluations θ, the repeated game $\Gamma(\pi)$ and the MDP $\Psi(\hat{\pi})$ have the same θ-value. In step 3 we check that the MDP satisfies the assumption of Theorem 4.5 so it has a general limit value and a general uniform value v^*. As a consequence the repeated game has a general limit value $v^*(\pi)$. Then we prove that player 1 can use an ϵ-optimal strategy of the auxilliary MDP in order to guarantee $v^*(\pi) - \epsilon$ in the original game. Finally we prove that Player 2 can play by blocks in the repeated game in order to guarantee $v^*(\pi) + \epsilon$. We obtain that $v^*(\pi)$ can be guaranteed by both players in the repeated game, so it is the general uniform value of $\Gamma(\pi)$.

Example 5.10. The computation of the value is in general a difficult problem, as shown by the next example introduced in Renault [30] and studied by Hörner *et al.* [20]. In this example the value exists but has been computed only for some values of the parameter. The set of states is $K = \{k_1, k_2\}$, the set of actions of player 1 is $I = \{T, B\}$, the set of actions of player 2 is $J = \{L, R\}$, and the payoff of player 1 is given by:

$$\underset{k_1}{\begin{array}{c|cc} & L & R \\ T & 1 & 0 \\ B & 0 & 0 \end{array}} \quad \text{and} \quad \underset{k_2}{\begin{array}{c|cc} & L & R \\ T & 0 & 0 \\ B & 0 & 1 \end{array}}.$$

The sequence of states follows an exogeneous Markov chain, with initial probability $(1/2, 1/2)$ and transition matrix $\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$, where p is a given parameter. At the beginning of every stage, only player 1 oberves the current state in K, and at the end of each stage the actions played are observed. (with the previous notations $C = K \times I \times J$, $D = I \times J$, and $q(k,i,j) = p\ \delta_{k,(k,i,j),(i,j)} + (1-p)\ \delta_{k',(k',i,j),(i,j)}$ for all k in K, $k' \in K \backslash \{k\}$, i in I and j in J).

For each value of the parameter $p \in [0,1]$, we have a repeated game Γ^p and by symmetry it is sufficient to study the case $p \in [1/2, 1]$. If $p = 1$ we are in the setup of Example 5.8 and the value is clearly 1/4. Hörner *et al.* [20] proved that for $p \in [1/2, 2/3)$, the value is $v_p = \frac{p}{4p-1}$. Bressaud and Quas [12] showed recently that for $p \in [2/3, .73]$ the value satisfies the following equation $\frac{1}{v_p} = u_0 + u_0 u_1 + u_0 u_1 u_2 + ...$, where (u_n) is defined by $u_0 = 1$ and $u_{n+1} = \max\{\psi(u_n), 1 - \psi(u_n)\}$ with $\psi(u) = 3p - 1 - \frac{2p-1}{u}$. What is the value for $p = 0.9$?

6 Appendix

6.1 Proof of Theorem 2.10

In this section we consider a compact metric space (X, d), and we use the Kantorovich-Rubinstein distance $d = d_{KR}$ on $\Delta(X)$. We start with a lemma.

Lemma 6.1. *Let $F : X \rightrightarrows \Delta_f(X)$ be non-expansive for d_{KR}. Then the mixed extension of F is 1-Lipschitz from $\Delta_f(X)$ to $\Delta_f(X)$ for d_{KR}.*

Proof of Lemma 6.1. We first show that the mapping $(p \mapsto \text{conv}F(p))$ is non-expansive from X to $\Delta_f(X)$. Indeed, consider p and p' in X, and $u = \sum_{i \in I} \alpha_i u_i$, with I finite, $\alpha_i \geq 0$, $u_i \in F(p)$ for each i, and $\sum_{i \in I} \alpha_i = 1$. By assumption for each i one can find u'_i in $F(p')$ such that $d_{KR}(u_i, u'_i) \leq d(p, p')$. Define $u' = \sum_{i \in I} \alpha_i u'_i$ in $\text{conv}F(p')$. We have:

$$\begin{aligned} d_{KR}(u, u') &= \sup_{f \in \mathcal{C}_1} \left(\sum_i \alpha_i u_i(f) - \sum_i \alpha_i u'_i(f) \right), \\ &= \sup_{f \in \mathcal{C}_1} \sum_{i \in I} \alpha_i (u_i(f) - u'_i(f)), \\ &\leq \sum_{i \in I} \alpha_i\ d_{KR}(u_i, u'_i), \\ &\leq d(p, p'). \end{aligned}$$

We now prove that $\hat{F}$ is 1-Lipschitz from $\Delta_f(X)$ to $\Delta_f(X)$. Let u_1, u_2 be in $\Delta_f(X)$ and $v_1 = \sum_{p \in X} u_1(p) f_1(p)$, where $f_1(p) \in \text{conv}F(p)$ for each p. By the Kantorovich duality formula, there exists a coupling $\gamma = (\gamma(p,q))_{(p,q) \in X \times X}$ in $\Delta_f(X \times X)$ with first marginal u_1 and second marginal u_2 satisfying:

$$d_{KR}(u_1, u_2) = \sum_{(p,q) \in X \times X} \gamma(p,q) d(p,q).$$

For each p, q in X by the first part of this proof there exists $f^p(q) \in \text{conv}F(q)$ such that $d_{KR}(f^p(q), f_1(p)) \leq d(p,q)$. We define:

$$f_2(q) = \sum_{p \in X} \frac{\gamma(p,q)}{u_2(q)} f^p(q) \in \text{conv}F(q), \text{ and } v_2 = \sum_{q \in X} u_2(q) f_2(q) \in \hat{F}(u_2).$$

We now conclude.

$$\begin{aligned}
d_{KR}(v_1, v_2) &= d_{KR}\left(\sum_{p\in X} u_1(p) f_1(p), \sum_{q\in X} u_2(q) f_2(q)\right)\\
&= d_{KR}\left(\sum_{p,q} \gamma(p,q) f_1(p), \sum_{q,p} \gamma(p,q) f^p(q)\right)\\
&\leq \sum_{p,q} \gamma(p,q) d_{KR}(f_1(p), f^p(q))\\
&\leq \sum_{p,q} \gamma(p,q) d(p,q) = d_{KR}(u_1, u_2).
\end{aligned}$$

The mixed extension of F is 1-lipschitz. □

We now consider a Gambling House $\Gamma = (X, F, r)$ and assume the hypotheses of Theorem 2.10 are satisfied. We will work with the deterministic Gambling House $\hat{\Gamma} = (\Delta_f(X), \hat{F}, r)$. Recall that r is extended to an affine and continuous mapping on $\Delta(X)$ whereas $\hat{F}$ is an affine non-expansive correspondence from $\Delta_f(X)$ to $\Delta_f(X)$.

For p in X, the pure plays in $\hat{\Gamma}$ at the initial state δ_p coincide with the mixed plays in Γ at the initial state p. As a consequence, the θ-value for Γ at p coincides with the θ-value for $\hat{\Gamma}$ at δ_p, which is written $v_\theta(p) = v_\theta(\delta_p)$. Because $\hat{F}$ and r are affine on $\Delta_f(X)$, the θ-value for $\hat{\Gamma}$, as a function defined on $\Delta_f(X)$, is the affine extension of the original v_θ defined on X. So we have a unique value function v_θ which is defined on $\Delta_f(X)$ and is affine. Because $\hat{F}$ is 1-Lipschitz and r is uniformly continuous, all the value functions v_θ have the same modulus of continuity as r, so $(v_\theta)_\theta$ is an equicontinuous family of mappings from $\Delta_f(X)$ to $[0,1]$. Consequently, we extend v_θ to an affine mapping on $\Delta(X)$ with the same modulous of continuity, and the family $(v_\theta)_\theta$ now is an equicontinuous family of mappings from $\Delta(X)$ to $[0,1]$. $\Delta_f(X)$ being precompact, this is enough to obtain the existence of a general limit value, see Renault [33]. Here we will moreover obtain a characterization of this value and the existence of the general uniform value.

We define R and v^* as in the statements of Theorem 2.10, so that for all x in X,

$$\begin{aligned}
v^*(x) \quad = \inf \quad &\big\{ w(x), w : \Delta(X) \to [0,1] \text{ affine } \textit{continuous s.t.}\\
&(1)\ \forall y \in X, w(y) \geq \sup_{u\in F(y)} w(u) \text{ and } (2) \forall u \in R, w(u) \geq r(u) \ \big\}.
\end{aligned}$$

We start with a lemma based on the non-expansiveness of $\hat{F}$.

Lemma 6.2. *1) Given (u, u') in $\text{cl}(Graph(\hat{F}))$, v in $\Delta_f(X)$ and $\varepsilon > 0$, there exists $v' \in \hat{F}(v)$ such that $d(u', v') \leq d(u, v) + \varepsilon$.*

2) Given a sequence $(z_t)_{t\geq 0}$ of elements of $\Delta(X)$ such that $(z_t, z_{t+1}) \in \text{cl}(Graph(\hat{F}))$ for all $t \geq 1$, and given $\varepsilon > 0$, one can find a sequence $(z'_t)_{t\geq 0}$ of elements of $\Delta_f(X)$ such that $(z'_t)_{t\geq 1}$ is a play at z'_0, and $d(z_t, z'_t) \leq \varepsilon$ for each $t \geq 0$.

Proof of Lemma 6.2: 1) For all $\varepsilon > 0$ there exists $(z, z') \in Graph(\hat{F})$ such that $d(z, u) \leq \varepsilon$ and $d(z', u') \leq \varepsilon$. Because $\hat{F}$ is non-expansive, one can find v' in $\hat{F}(v)$ such that $d(z', v') \leq d(z, v)$. Consequently, $d(v', u') \leq d(v', z') + d(z', u') \leq d(z, v) + \varepsilon \leq d(u, v) + 2\varepsilon$.

2) It is first easy to construct (z'_0, z'_1) in the graph of $\hat{F}$ such that $d(z'_0, z_0) \leq \varepsilon$ and $d(z'_1, z_1) \leq \varepsilon$. $(z_1, z_2) \in \text{cl}(Graph(\hat{F}))$ so by 1) one can find (z'_2) in $\hat{F}(z'_1)$ such that $d(z_2, z'_2) \leq d(z_1, z'_1) + \varepsilon^2 \leq \varepsilon + \varepsilon^2$. Iterating, we construct a play $(z'_t)_{t\geq 1}$ at z'_0 such that $d(z_t, z'_t) \leq \varepsilon + \varepsilon^2 + ... + \varepsilon^t$ for each t.

We now prove[2] Theorem 2.10 with the two following propositions.

Proposition 6.3. *Γ has a general limit value given by v^*.*

Proof of Proposition 6.3: By Ascoli's theorem, it is enough to show that any limit point of $(v_\theta)_\theta$ (for the uniform convergence) coincides with v^*. We thus assume that $(v_{\theta^k})_k$ uniformly converges to v on $\Delta(X)$ when k goes to ∞, for a family of evaluations satisfying:

$$\sum_{t\geq 1} |\theta^k_{t+1} - \theta^k_t| \longrightarrow_{k\to\infty} 0.$$

We need to show that $v = v^*$.

A) We first show that $v \geq v^*$.

It is plain that v can be extended to an affine function on $\Delta(X)$ and has the modulus of continuity of r. Because $\sum_{t\geq 1} |\theta^k_{t+1} - \theta^k_t| \longrightarrow_{k\to\infty} 0$, we have by Equation (1) of Section 2 that: $\forall y \in X, v(y) = \sup_{u\in F(y)} v(u)$.

Let now u be in R. By Lemma 6.2 for each ε one can find u_0 in $\Delta_f(X)$ and a play $(u_1, u_2, ..., u_t, ...)$ such that $u_t \in \hat{F}(u_{t-1})$ and $d(u, u_t) \leq \varepsilon$ for all $t \geq 0$. Because r is uniformly continuous, we get $v(u) \geq r(u)$.

By definition of v^* as an infimum, we obtain: $v^* \leq v$.

B) We show that $v^* \geq v$.

Let w be a continous affine mapping from $\Delta(X)$to $[0,1]$ satisfying (1) and (2) of the definition of v^*. It is enough to show that $w(p) \geq v(p)$ for each p in X. Fix p in X and $\varepsilon > 0$.

For each k, let $\sigma^k = (u^k_1, ..., u^k_t, ...) \in \Delta_f(X)^\infty$ be a play at δ_p for $\hat{\Gamma}$ which is almost optimal for the θ^k-value, in the sense that $\sum_{t\geq 1} \theta^k_t r(u^k_t) \geq v_{\theta^k}(p) - \varepsilon$. Define:

$$u(k) = \sum_{t=1}^{\infty} \theta^k_t u^k_t \in \Delta(X), \text{ and } u'(k) = \sum_{t=1}^{\infty} \theta^k_t u^k_{t+1} \in \Delta(X).$$

$u(k)$ and $u'(k)$ are well-defined limits of normal convergent series in the Banach space $\mathcal{C}(X)'$. Because $\hat{F}$ is affine, its graph is a convex set and $(u(k), u'(k)) \in \text{cl}(Graph(\hat{F}))$ for each k.

Moreover, we have $d(u(k), u'(k)) \leq \text{diam}(X)(\theta^k_1 + \sum_{t=2}^{\infty} |\theta^k_t - \theta^k_{t-1}|)$, where $\text{diam}(X)$ is the diameter of X. Consequently, $\sum_{t\geq 1} |\theta^k_{t+1} - \theta^k_t| \longrightarrow_{k\to\infty} 0$ implies $d(u(k), u'(k)) \longrightarrow_{k\to\infty} 0$. Considering a limit point of the sequence $(u(k), u'(k))_k$, we obtain some u in R. By assumption on w, $w(u) \geq r(u)$. Moreover, for each k we have $r(u(k)) = \sum_{t\geq 1} \theta^k_t r(u^k_t) \geq v_{\theta^k}(p) - \varepsilon$, so $r(u) \geq v(p) - \varepsilon$.

Because w is excessive, we obtain that for each k the sequence $(w(u^k_t))_t$ is non increasing, so $w(u(k)) = \sum_{t\geq 1} \theta^k_t w(u^k_t) \leq w(p)$. So we obtain:

$$w(p) \geq w(u) \geq r(u) \geq v(p) - \varepsilon.$$

This is true for all ε, so $w \geq v$. □

Proposition 6.4. *Γ has a general uniform value.*

Proof of Proposition 6.4: First we can extend the notion of mixed play to $\Delta_f(X)$. A mixed play at $u_0 \in \Delta_f(X)$, is a sequence $\sigma = (u_1, ..., u_t, ...) \in \Delta_f(X)^\infty$ such that $u_{t+1} \in \hat{F}(u_t)$ for

[2] A variant of the proof would be to consider the Gambling House on $\Delta(X)$ where the transition correspondence is defined so that its graph is the closure of the graph of $\hat{F}$. Part 1) of Lemma 6.2 shows this correspondence is also non-expansive.

each $t \geq 0$, and we denote by $\Sigma(u_0)$ the set of mixed plays at u_0. Given t, T in $I\!N$, $n \in I\!N^*$ and $u_0 \in \Delta_f(X)$, we define for each mixed play $\sigma = (u_t)_{t\geq 1} \in \Sigma(u_0)$ the auxiliary payoff:

$$\gamma_{t,n}(\sigma) = \frac{1}{n}\sum_{l=t+1}^{t+n} r(u_l), \text{ and } \beta_{T,n}(\sigma) = \inf_{t\in\{0,\dots,T\}} \gamma_{t,n}(\sigma).$$

We also define the auxiliary value function: for all u in $\Delta_f(X)$,

$$h_{T,n}(u_0) = \sup_{\sigma\in\Sigma(u_0)} \beta_{T,n}(\sigma).$$

Clearly, $\beta_{T,n}(\sigma) \leq \gamma_{0,n}(\sigma)$ and $h_{T,n}(u_0) \leq v_n(u_0)$. We can write:

$$\begin{aligned} h_{T,n}(u_0) &= \sup_{\sigma\in\Sigma(u_0)} \inf_{\theta\in\Delta(\{0,\dots,T\})} \frac{1}{n}\sum_{t=0}^{T}\theta_t \sum_{l=t+1}^{t+n} r(u_l) \\ &= \sup_{\sigma\in\Sigma(u_0)} \inf_{\theta\in\Delta(\{0,\dots,T\})} \sum_{l=1}^{T+n} \beta_l(\theta,n) r(u_l). \end{aligned}$$

where for each l in $1, ..., T+n$, $\beta_l(\theta, n) = \frac{1}{n}\sum_{t=Max\{0,l-n\}}^{\text{Min}\{T,l-1\}} \theta_t$. By construction, $\hat{F}$ is affine, so $\Sigma(u_0)$ is a convex subset of $\Delta_f(X)^\infty$. $\Delta(\{0,...,T\})$ is convex compact and the payoff $\sum_{l=1}^{T+n} \beta_l(\theta,n) r(u_l)$ is affine both in θ and in σ. We can apply a standard minmax theorem to get:

$$h_{T,n}(u_0) = \inf_{\theta\in\Delta(\{0,\dots,T\})} \sup_{\sigma\in\Sigma(u_0)} \sum_{l=1}^{T+n} \beta_l(\theta,n) r(u_l).$$

We write $\theta_t = 0$ for $t > T$ and for each $l \geq 0$: $\beta_l(n,\theta) = \frac{1}{n}(\theta_0 + ... + \theta_{l-1})$ if $l \leq n$, $\beta_l(\theta, n) = \frac{1}{n}(\theta_{l-n} + ... + \theta_{l-1})$ if $n+1 \leq l \leq n+T$, $\beta_l(n,\theta) = 0$ if $l > n+T$. The evaluation $\beta(\theta, n)$ is a particular probability on stages and $h_{T,n}(u_0) = \inf_{\theta\in\Delta(\{0,...,T\})} v_{\beta(\theta,n)}(u_0)$. It is easy to bound the total variation of $\beta(\theta, n)$:

$$\sum_{l\geq 0} |\beta_{l+1}(\theta,n) - \beta_l(\theta,n)| = \sum_{l=0}^{n-1} \frac{\theta_l}{n} + \sum_{l\geq n} \frac{1}{n}|\theta_l - \theta_{l-n}| \leq \frac{3}{n} \longrightarrow_{n\to\infty} 0.$$

The impatience of $\beta(\theta, n)$ goes to zero as n goes to infinity, uniformly in θ. So we can use the previous Proposition 6.3 to get:

$$\forall \varepsilon > 0, \exists n_0, \forall n \geq n_0, \forall \theta \in \Delta(I\!N), \forall u_0 \in \Delta_f(X),\ |v_{\beta(\theta,n)}(u_0) - v^*(u_0)| \leq \varepsilon.$$

This implies that $h_{\infty,n}(u_0) :=_{def} \inf_{\theta\in\Delta(I\!N)} v_{\beta(\theta,n)}(u_0) = \inf_{T\geq 0} h_{T,n}(u_0)$ converges to $v^*(u_0)$ when $n \to \infty$, and the convergence is uniform over $\Delta_f(X)$. Consequently, if we fix $\varepsilon > 0$ there exists n_0 such that for all u_0 in $\Delta_f(X)$, for all $T \geq 0$, there exists a play $\sigma^T = (u_t^T)_{t\geq 1}$ in $\Sigma(u_0)$ such that the average payoff is good on every interval of n_0 stages starting before $T+1$: for all $t = 0, ..., T$, $\gamma_{t,n_0}(\sigma^T) \geq v^*(u_0) - \varepsilon$.

We fix u_0 in $\Delta_f(X)$ and consider, for each T, the play $\sigma^T = (u_t^T)_{t\geq 1}$ in $\Sigma(u)$ as above. By a diagonal argument we can construct for each $t \geq 1$ a limit point z_t in $\Delta(X)$ of the sequence $(u_t^T)_{T\geq 0}$ such that for all t we have $(z_t, z_{t+1}) \in \text{cl}(Graph(\hat{F}))$, with $z_0 = u_0$. For each $m \geq 0$, we have $\frac{1}{n_0}\sum_{t=m+1}^{m+1+n_0} r(u_t^T) \geq v^*(u_0) - \varepsilon$ for T large enough, so at the limit we get: $\frac{1}{n_0}\sum_{t=m+1}^{m+1+n_0} r(z_t) \geq v^*(u_0) - \varepsilon$.

r being uniformly continuous, there exists α such that $|r(z) - r(z')| \leq \varepsilon$ as soon as $d(z, z') \leq \alpha$. By Lemma 6.2, one can find a mixed play $\sigma' = (z'_1, ..., z'_t, ...)$ at $\Sigma(z_0)$ such that for each t, $d(z_t, z'_t) \leq \alpha$. We obtain that for each $m \geq 0$, $\frac{1}{n_0}\sum_{t=m+1}^{m+1+n_0} r(z'_t) \geq v^*(u) - 2\varepsilon$.

Consequently we have proved: $\forall \varepsilon > 0$, there exists n_0 such that for all initial state p in X, there exists a mixed play $\sigma' = (z'_t)_t$ at p such that: $\forall m \geq 0$, $\frac{1}{n_0}\sum_{t=m+1}^{m+1+n_0} r(z'_t) \geq v^*(p) - 2\varepsilon$.

Let $\theta \in \Delta(I\!N^*)$ be an evaluation, it is now easy to conclude. First if $v^*(p) - 2\epsilon < 0$, then any play is 2ϵ-optimal. Otherwise, for each $j \geq 1$, denote by $\overline{\theta_j}$ the maximum of θ on the block $B^j = \{(j-1)n_0 + 1, ..., jn_0\}$. For all $t \in B^j$, we have: $\overline{\theta_j} \geq \theta_t \geq \overline{\theta_j} - \sum_{t' \in \{(j-1)n_0+1,...jn_0-1\}} |\theta_{t'+1} - \theta_{t'}|$. As a consequence, we have for all j:

$$\begin{aligned} \sum_{t=(j-1)n_0+1}^{jn_0} \theta_t r(z'_t) &\geq \overline{\theta_j} \sum_{t=(j-1)n_0+1}^{jn_0} r(z'_t) \;-\; n_0 \sum_{t' \in \{(j-1)n_0+1,...,jn_0-1\}} |\theta_{t'+1} - \theta_{t'}| \\ &\geq \sum_{t=(j-1)n_0+1}^{jn_0} \theta_t (v^*(p) - 2\varepsilon) \;-\; n_0 \sum_{t' \in \{(j-1)n_0+1,...,jn_0-1\}} |\theta_{t'+1} - \theta_{t'}| \end{aligned}$$

and by summing over j we get: $\gamma_\theta(x_0, \sigma') \geq v^*(p) - 2\epsilon - n_0 TV(\theta) \geq v^*(p) - 3\epsilon$ as soon as $TV(\theta)$ is small enough. □

6.2 Proof of Theorem 3.5

We first introduce another pseudo-distance on $\Delta(X)$.

Definition 6.5. $d_2^+(u,v) = \inf_{\varepsilon>0} d_2^\varepsilon(u,v)$, where $d_2^\varepsilon(u,v) = \sup_{(f,g)\in D_2^\varepsilon} u(f) + v(g)$ *and for* $\varepsilon > 0$, $D_2^\varepsilon = \{(f,g) \in \mathcal{C}\times\mathcal{C}, \forall x, y \in X, \forall a, b \in [0,1],\ af(x) + bg(y) \leq \varepsilon + \|ax - by\|\}$.

We will show that $d_1 = d_2 = d_2^+ = d_3$. The proof is split into several parts.

Proposition 6.6. $d_1 = d_2 = d_2^+$.

It is plain that $d_1 \leq d_2 \leq d_2^+$, so all we have to prove is $d_2^+ \leq d_1$. We start with a lemma.

Lemma 6.7. *Fix* $\varepsilon > 0$*, and let* f *in* $\mathcal{C}$ *be such that:* $\forall x \in X$, $\forall a \in [0,1]$, $af(x) \leq \varepsilon + a\|x\|$. *Define* $\hat{f}$ *by:*

$$\forall y \in X,\ \hat{f}(y) = \inf_{a\in[0,1], b\in(0,1], x\in X} \frac{1}{b}\left(\varepsilon + \|ax - by\| - af(x)\right).$$

Then for each y *in* X, $-\|y\| \leq \hat{f}(y) \leq -f(y) + \varepsilon$. *Moreover* $\hat{f} \in \mathcal{C}_1$*, and:*

$$\forall x \in X, \forall y \in X, \forall a \in [0,1], \forall b \in [0,1],\ a\hat{f}(x) - b\hat{f}(y) \leq a\varepsilon + \|by - ax\|.$$

Proof of Lemma 6.7: By assumption on f, we have for all y in X, a in $[0,1]$, b in $(0,1]$, x in X: $\frac{1}{b}\left(\varepsilon + \|ax - by\| - af(x)\right) \geq \frac{1}{b}\left(-a\|x\| + \|ax - by\|\right) \geq -\|y\|$. In the definition of $\hat{f}(y)$, considering $a = b = 1$ and $x = y$ yields $\hat{f}(y) \leq -f(y) + \varepsilon$.

Fix x and y in X, a and b in $[0,1]$. We have:

$$\begin{aligned} a\hat{f}(x) - b\hat{f}(y) &= a \inf_{a',b',x'} \frac{1}{b'}\left(\varepsilon + \|a'x' - b'x\| - a'f(x')\right) \\ &\quad -b \inf_{a'',b'',x''} \frac{1}{b''}\left(\varepsilon + \|a''x'' - b''y\| - a''f(x'')\right). \end{aligned}$$

If $a = 0$, then the inequality $\hat{f}(y) \geq -\|y\|$ leads to $-b\hat{f}(y) \leq b\|y\|$. If $b = 0$, choose $a' = 0$, $b' = 1$ and $x' = x$ to get $a\hat{f}(x) \leq a\varepsilon + \|ax\|$.

If $ab > 0$, given $\eta > 0$, choose a'', b'', x'' η-optimal in the second infimum. We can define $x' = x''$, and choose $a' \in [0,1]$ and $b' \in (0,1]$ such that $\frac{a'}{b'} = \frac{b}{a}\frac{a''}{b''}$. We obtain:

$$\begin{aligned} a\hat{f}(x) - b\hat{f}(y) &\leq b\eta + (\frac{a}{b'} - \frac{b}{b''})\varepsilon + (\|\frac{a''}{b''}bx'' - ax\| - \|\frac{a''}{b''}bx'' - by\|) \\ &\leq b\eta + (\frac{a}{b'} - \frac{b}{b''})\varepsilon + \|ax - by\|. \end{aligned}$$

If $a = b > 0$, choose $a' = a''$ and $b' = b''$ to obtain: $\hat{f}(x) - \hat{f}(y) \leq \|x - y\|$ and therefore $\hat{f}$ is 1-Lipschitz.

Otherwise, we distinguish two cases. If $\frac{a}{b}b'' \leq 1$, we define $b' = \frac{a}{b}b''$ and $a' = a''$ and we get $a\hat{f}(x) - b\hat{f}(y) \leq b\eta + \|ax - by\|$. If $\frac{a}{b}b'' > 1$, we define $b' = 1$ and $a' = \frac{a''b}{b''a} \in [0,1]$ and obtain $a\hat{f}(x) - b\hat{f}(y) \leq b\eta + a\varepsilon + \|ax - by\|$. Thus for all $\eta > 0$, we have

$$a\hat{f}(x) - b\hat{f}(y) \quad \leq \quad b\eta + a\varepsilon + \|ax - by\|,$$

and therefore $a\hat{f}(x) - b\hat{f}(y) \leq a\varepsilon + \|ax - by\|$.

Proof of Proposition 6.6: Fix u and v in $\Delta(X)$, and consider $\varepsilon > 0$. For each (f,g) in D_2^ε, we have $-f + \varepsilon \geq \hat{f} \geq g$ and $(f, \hat{f})$ in D_2^ε. We also have $(\hat{f}, f) \in D_2^\varepsilon$ so iterating the construction, we get $(\hat{f}, \hat{\hat{f}}) \in D_2^\varepsilon$, and $-\hat{f} + \varepsilon \geq \hat{\hat{f}} \geq f$.

Now, $u(f) + v(g) \leq u(\hat{\hat{f}}) + v(\hat{f}) \leq -u(\hat{f}) + \varepsilon + v(\hat{f})$. Hence we have obtained:

$$d_2^\varepsilon(u,v) \leq \varepsilon + \sup_{f \in C_{\varepsilon(u,v)}} -u(f) + v(f),$$

where $C_{\varepsilon(u,v)}$ is the set of functions f in $\mathcal{C}_1$ satisfying:

$$\forall x \in X, \forall y \in X, \forall a \in [0,1], \forall b \in [0,1],\ af(x) - bf(y) \leq a\varepsilon + \|ax - by\| \textit{ and } f(y) \geq -\|y\|.$$

For each positive k, one can choose f_k in $\mathcal{C}_1$ achieving the above supremum for $\varepsilon = 1/k$. Taking a limit point of $(f_k)_k$ yields a function f in D_1 such that: $-u(f) + v(f) \geq d_2^+(u,v)$. The function $f^* = -f$ is in D_1 and satisfies $u(f^*) - v(f^*) \geq d_2^+(u,v)$, and the proof of Proposition 6.6 is complete. □

Proposition 6.8. $d_2^+ \geq d_3$.

Proof of Proposition 6.8: The proof is based on (a corollary of) Hahn-Banach theorem. Define: $H = \mathcal{C}(X^2 \times [0,1]^2)$ and

$$L = \{\varphi \in H, \exists f, g \in \mathcal{C}(X) \ s.t. \ \forall x, y \in X, \forall \lambda, \mu \in [0,1], \varphi(x,y,\lambda,\mu) = \lambda f(x) + \mu g(y)\}.$$

H is endowed with the uniform norm and L is a linear subspace of H. Note that the unique constant mapping in L is 0. Fix u and v in $\Delta(X)$, and let r be the linear form on L defined by $r(\varphi) = u(f) + v(g)$, where $\varphi(x,y,\lambda,\mu) = \lambda f(x) + \mu g(y)$ for all x, y, λ, μ.

Fix now $\varepsilon > 0$, and put:

$$U_\varepsilon = \{\varphi \in H, \forall x, y \in X, \forall \lambda, \mu \in [0,1], \varphi(x,y,\lambda,\mu) \leq \|\lambda x - \mu y\| + \varepsilon\}.$$

We have:

$$\sup_{\varphi \in L \cap U_\varepsilon} r(\varphi) = d_2^\varepsilon(u,v).$$

U_ε is a convex subset of H which is radial at 0, in the sense that: $\forall \varphi \in H$, $\exists \delta > 0$ such that $t\varphi \in U_\varepsilon$ as soon as $|t| \leq \delta$. By a corollary of Hahn-Banach theorem (see theorem 6.2.11 p.202 in Dudley, [17]), r can be extended to a linear form on H such that:

$$\sup_{\varphi \in U_\varepsilon} r(\varphi) = d_2^\varepsilon(u,v).$$

Given $\varphi \in H$, we have $\varepsilon\varphi/\|\varphi\|_\infty \in U_\varepsilon$, which implies that $r(\varphi) \leq \|\varphi\|_\infty d_2^\varepsilon(u,v)/\varepsilon$, so that r belongs to H'. If $\varphi \geq 0$, we have $t\varphi \in U_\varepsilon$ if $t \leq 0$, so that $r(\varphi) \geq d_2^\varepsilon(u,v)/t$ for all $t \leq 0$ and $r(\varphi) \geq 0$. By Riesz Theorem, r can be represented by a positive finite measure γ on $X^2 \times [0,1]^2$.

Given f in $\mathcal{C}$, one can consider $\varphi_f \in L$ defined by $\varphi_f(x,y,\lambda,\mu) = \lambda f(x)$. Then $r(\varphi_f) = \gamma(\varphi_f)$ gives: $u(f) = \int_{(x,y,\lambda,\mu)\in X^2\times[0,1]^2} \lambda f(x) d\gamma(x,y,\lambda,\mu)$, and similarly

$$v(f) = \int_{(x,y,\lambda,\mu)\in X^2\times[0,1]^2} \mu f(y) d\gamma(x,y,\lambda,\mu),$$

and we obtain that $\gamma \in \mathcal{M}_3(u,v)$.

Since $\gamma \geq 0$, we have $\sup_{\varphi\in U_\varepsilon} r(\varphi) = r(\varphi^*)$ where $\varphi^*(x,y,\lambda,\mu) = \|\lambda x - \mu y\| + \varepsilon$. We get $d_2^\varepsilon(u,v) = \int_{X^2\times[0,1]^2} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) + \varepsilon\gamma(X^2 \times [0,1]^2)$, so

$$d_2^\varepsilon(u,v) \geq \int_{X^2\times[0,1]^2} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) \geq d_3(u,v). \qquad \square$$

Remark 6.9. *The proof given here uses elements of the proof of the standard Kantorovich duality formula in Dudley ([17], see Lemma 11.8.5 p.423). However the arguments need to be more sophisticated here. In particular, there is no need in the standard duality for the extra variables λ and μ, and the analogs of our sets H and L are $\hat{H} = \mathcal{C}(X^2)$ and $\hat{L} = \{\varphi \in \hat{H}, \exists f,g \in \mathcal{C}(X)$ s.t. $\forall x,y \in X, \varphi(x,y) = f(x) + g(y)\}$. It is enough to define $\hat{U} = \{\varphi \in \hat{H}, \forall x,y \in X, \varphi(x,y) < \| x - y\|\}$. $\hat{U}$ is convex and open, so radial at any of his element, and $\hat{U} \cap \hat{L}$ is not empty. In the present setup, if we simply define $U = \{\varphi \in H, \forall x,y \in X, \forall \lambda,\mu \in [0,1], \varphi(x,y,\lambda,\mu) < \|\lambda x - \mu y\|\}$, we have $U \cap L = \emptyset$, hence a problem. These considerations have led to the introduction of the sets U_ε and the intermediate distance d_2^+ beforehand.*

It is now easy to conclude the proof of Theorem 3.5.

Lemma 6.10. *$d_3 \geq d_2$.*

Proof of Lemma 6.10: Fix $(f,g) \in D_2$ and $\gamma \in \mathcal{M}_3(u,v)$.

$$\begin{aligned} u(f) + v(g) &= \int_{X^2\times[0,1]^2} \lambda f(x) d\gamma(x,y,\lambda,\mu) + \int_{X^2\times[0,1]^2} \mu g(y) d\gamma(x,y,\lambda,\mu) \\ &= \int_{X^2\times[0,1]^2} (\lambda f(x) + \mu g(y)) d\gamma(x,y,\lambda,\mu) \\ &\leq \int_{X^2\times[0,1]^2} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu). \quad \square \end{aligned}$$

6.3 Proof of Lemma 3.6

We will use the Stone-Weierstrass theorem (see for instance Lemma A7.2 in Ash [2] p. 392) in order to show that the linear span of D_1 is dense in $\mathcal{C}(X)$. We denote span(D_1) the linear span of D_1.

If f and g belong to D_1 and $\lambda \in [0,1]$, then $-f$, $\sup\{f,g\}$, $\inf\{f,g\}$ and $\lambda f + (1-\lambda)g$ are in D_1. It follows that the linear span of D_1 is stable by sup and inf operations.

We now show that for every $x,y \in X$ such that $x \neq y$ and every function $f : X \to I\!R$, there exists a function $h \in$ span(D_1) such that $f(x) = h(x)$ and $f(y) = h(y)$. Fix f a function and $x,y \in X$. We define Y the linear span of x and y. Define the function φ from Y to $I\!R$ by

$$\forall \lambda,\mu \in I\!R,\ \varphi(\lambda x + \mu y) = \lambda f(x) + \mu f(y).$$

By Hahn-Banach theorem, φ can be extended to a linear mapping on $I\!\!R^K$ denoted g with the same operator norm. It follows that there exists $C > 0$ such that for every $p, q \in X$, for every $a, b \geq 0$,

$$|ag(p) - bg(q)| = |g(ap - bq)| \leq C\|ap - bq\|.$$

Denote by h the restriction of g to X. It is continuous and affine, therefore it is in $\text{span}(D_1)$. By Stone-Weierstrass theorem, we deduce that D_1 is dense in $\mathcal{C}(X)$.

6.4 Proof of Theorem 3.10

Let u and v be in $\Delta(X)$, and denote by U and V the respective supports of u and v. We write $S = X^2 \times [0,1]^2$, and we start with a lemma, where no finiteness assumption on U or V is needed. Recall that $\mathcal{M}_3(u,v)$ is the set of finite positive measures on S satisfying for each f in $\mathcal{C}$:

$$\int_{(x,y,\lambda,\mu)\in S} \lambda f(x) d\gamma(x,y,\lambda,\mu) = u(f), \text{ and } \int_{(x,y,\lambda,\mu)\in S} \mu f(y) d\gamma(x,y,\lambda,\mu) = v(f).$$

Lemma 6.11. *For each $\gamma \in \mathcal{M}_3(u,v)$, we have:*

$$\int_S \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) = 2 + \int_{U\times V\times[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu)\, d\gamma(x,y,\lambda,\mu).$$

Proof of Lemma 6.11: Write $A(\gamma) = \int_S \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu)$. Using the definition of $\mathcal{M}_3(u,v)$, we can obtain that $1 = \int_S \lambda \mathbf{1}_{x\in U} d\gamma = \int_S \mu \mathbf{1}_{y\in V} d\gamma$. This implies: $\int_S \lambda \mathbf{1}_{x\notin U} d\gamma = \int_S \mu \mathbf{1}_{y\notin V} d\gamma = 0$, so that $\lambda \mathbf{1}_{x\notin U} = \mu \mathbf{1}_{y\notin V} = 0 \;\; \gamma$. a.s. We can write:

$$A(\gamma) = \int_S \mathbf{1}_{x\in U, y\in V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) + \int_S \mathbf{1}_{x\in U, y\notin V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu)$$
$$+ \int_S \mathbf{1}_{x\notin U, y\in V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) + \int_S \mathbf{1}_{x\notin U, y\notin V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu)$$
$$= \int_S \mathbf{1}_{x\in U, y\in V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) + \int_S \mathbf{1}_{x\in U, y\notin V} \lambda d\gamma(x,y,\lambda,\mu) + \int_S \mathbf{1}_{x\notin U, y\in V} \mu d\gamma(x,y,\lambda,\mu) + 0.$$

We now use $1 = \int_S \mathbf{1}_{x\in U, y\in V} \lambda d\gamma + \int_S \mathbf{1}_{x\in U, y\notin V} \lambda d\gamma$ and $1 = \int_S \mathbf{1}_{x\in U, y\in V} \mu d\gamma + \int_S \mathbf{1}_{x\notin U, y\in V} \mu d\gamma$ to obtain:

$$A(\gamma) = 2 + \int_S \mathbf{1}_{x\in U, y\in V} \|\lambda x - \mu y\| d\gamma(x,y,\lambda,\mu) - \int_S \mathbf{1}_{x\in U, y\in V} \lambda d\gamma - \int_S \mathbf{1}_{x\in U, y\in V} \mu d\gamma. \quad \square$$

We assume in the sequel that U and V are finite, and define $d_5(u,v)$ as follows:

Definition 6.12. *Define*

$$\mathcal{M}_5(u,v) = \big\{(\alpha,\beta) = (\alpha(x,y), \beta(x,y))_{(x,y)\in U\times V} \in {I\!\!R_+}^{U\times V} \times {I\!\!R_+}^{U\times V}, s.t.$$
$$\forall x \in U, \sum_{y'\in V} \alpha(x,y') \leq u(x) \text{ and } \forall y \in V, \sum_{x'\in U} \beta(x',y) \leq v(y)\big\},$$

$$\textit{and } d_5(u,v) = \inf_{(\alpha,\beta)\in\mathcal{M}_5(u,v)} \left(2 + \sum_{(x,y)\in U\times V} \big(\|x\alpha(x,y) - y\beta(x,y)\| - \alpha(x,y) - \beta(x,y)\big)\right).$$

$\mathcal{M}_5(u,v)$ is a polytope in the Euclidean space $(I\!\!R^{U\times V})^2$, so the infimum in the definition of $d_5(u,v)$ is achieved.

Lemma 6.13. $d_3(u,v) \geq d_5(u,v)$.

Proof of Lemme 6.13: Let γ be in $\mathcal{M}_3(u,v)$. Fix for a while (x,y) in $U \times V$, and assume that $\gamma(x,y) > 0$. We define $\gamma(.|x,y)$ the conditional probability on $[0,1]^2$ given (x,y) by: for all $\varphi \in C([0,1]^2)$,

$$\int_{[0,1]^2} \varphi(\lambda,\mu) d\gamma(\lambda,\mu|x,y) = \frac{1}{\gamma(x,y)} \int_{(x',y',\lambda,\mu)\in S} \mathbf{1}_{x'=x,y'=y} \varphi(\lambda,\mu) d\gamma(x',y',\lambda,\mu).$$

So that

$$\gamma(x,y) \int_{[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu) d\gamma(\lambda,\mu|x,y) = \int_{(\lambda,\mu)\in[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu) d\gamma(x,y,\lambda,\mu).$$

Define $P(x,y) = \int_{(\lambda,\mu)\in[0,1]^2} \lambda d\gamma(\lambda,\mu|x,y)$ and $Q(x,y) = \int_{(\lambda,\mu)\in[0,1]^2} \mu d\gamma(\lambda,\mu|x,y)$. The mapping $\Psi : (\lambda,\mu) \mapsto \|\lambda x - \mu y\| - \lambda - \mu$ is convex so by Jensen's inequality we get:

$$\int_{(\lambda,\mu)\in[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu) d\gamma(\lambda,\mu|x,y) \geq \|xP(x,y) - yQ(x,y)\| - P(x,y) - Q(x,y).$$

Now, by Lemma 6.11,

$$\begin{aligned} A(\gamma) &= 2 + \sum_{x\in U, y\in V} \int_{(\lambda,\mu)\in[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu)\, d\gamma(x,y,\lambda,\mu) \\ &= 2 + \sum_{x\in U, y\in V, \gamma(x,y)>0} \int_{(\lambda,\mu)\in[0,1]^2} (\|\lambda x - \mu y\| - \lambda - \mu)\, d\gamma(x,y,\lambda,\mu) \\ &\geq 2 + \sum_{x\in U, y\in V, \gamma(x,y)>0} \gamma(x,y) \left(\|xP(x,y) - yQ(x,y)\| - P(x,y) - Q(x,y)\right). \end{aligned}$$

For (x,y) in $U \times V$, define $\alpha(x,y) = \gamma(x,y)P(x,y) \geq 0$ *and* $\beta(x,y) = \gamma(x,y)Q(x,y) \geq 0$ (with $\alpha(x,y) = \beta(x,y) = 0$ if $\gamma(x,y) = 0$). We get:

$$A(\gamma) \geq 2 + \sum_{x\in U, y\in V} \left(\|x\alpha(x,y) - y\beta(x,y)\| - \alpha(x,y) - \beta(x,y)\right).$$

We have, for each x in U:

$$\begin{aligned} \sum_{y\in V} \alpha(x,y) &= \sum_{y\in V, \gamma(x,y)>0} \int_{(\lambda,\mu)\in[0,1]^2} \lambda d\gamma(x,y,\lambda,\mu) \\ &\leq \int_{(y,\lambda,\mu)\in X\times[0,1]^2} \lambda d\gamma(x,y,\lambda,\mu) = u(x). \end{aligned}$$

where the last equality comes from the definition of $\mathcal{M}_3(u,v)$. Similarly, for each y in V we can show that $\sum_{x\in U} \beta(x,y) \leq v(y)$, and Lemma 6.13 is proved. □

Lemma 6.14. $d_5(u,v) \geq d_4(u,v)$.

Proof of Lemme 6.14: Consider (α^*, β^*) achieving the minimum in the definition of $d_5(u,v)$. Assume that there exists x^* such that $\sum_{y\in V} \alpha(x^*,y) < u(x^*)$. For any x in X and z in $\mathbb{R}_+^K$, it is easy to see that the mapping $l : (\alpha \mapsto \|x\alpha - z\| - \alpha)$ is nonincreasing from $\mathbb{R}_+$ to $\mathbb{R}$ (as the sum of the mappings $l_k : (\alpha \mapsto |\alpha x^k - z^k| - \alpha x^k)$, each l^k being non increasing in α). As a consequence, one can choose any y^* in V and increase $\alpha(x^*,y^*)$ in order to saturate the constraint without increasing the objective. So we can assume without loss of generality that $\sum_{y\in V} \alpha(x^*,y) = u(x^*)$ for all x^* and similarly $\sum_{x\in U} \beta(x,y^*) = v(y^*)$ for all y^*.

Consequently,

$$\begin{aligned} d_5(u,v) &= 2 + \sum_{(x,y)\in U\times V} \left(\|x\alpha^*(x,y) - y\beta^*(x,y)\| - \alpha^*(x,y) - \beta^*(x,y)\right) \\ &= \sum_{(x,y)\in U\times V} \|x\alpha^*(x,y) - y\beta^*(x,y)\| \geq d_4(u,v). \quad \square \end{aligned}$$

Lemma 6.15. $d_4(u,v) \geq d_2(u,v)$.

Proof of Lemma 6.15: Fix $(f,g) \in D_2$ and $(\alpha,\beta) \in \mathcal{M}_4(u,v)$.

$$\begin{aligned} u(f) + v(g) &= \sum_{x\in U} f(x)u(x) + \sum_{y \in Y} g(y)v(y) \\ &= \sum_{(x,y)\in U\times V} f(x)\alpha(x,y) + g(y)\beta(x,y) \\ &\leq \sum_{(x,y)\in U\times V} \|\alpha(x,y)x - \beta(x,y)y\| \leq d_4(u,v). \end{aligned}$$

We have shown that $d_3(u,v) \geq d_5(u,v) \geq d_4(u,v) \geq d_2(u,v) = d_3(u,v) = d_1(u,v)$. This ends the proof of Theorem 3.10.

6.5 Proof of Lemma 2.2

We prove that

$$\forall u, u' \in \Delta_f(X),\ \forall \alpha \in [0,1],\ \hat{F}(\alpha u + (1-\alpha)u') = \alpha \hat{F}(u) + (1-\alpha)\hat{F}(u').$$

The $\subset$ part is clear. To see the reverse inclusion, let $v = \alpha \sum_{x\in X} u(x)f(x) + (1-\alpha)\sum_{x\in X} u'(x)f'(x)$ be in $\alpha\hat{F}(u) + (1-\alpha)\hat{F}(u')$, with transparent notations. Define

$$g(x) = \frac{\alpha u(x)f(x) + (1-\alpha)u'(x)f'(x)}{\alpha u(x) + (1-\alpha)u'(x)},$$

for each x such that the denominator is positive. Then $g(x) \in \text{conv}F(x)$, and

$$v = \sum_{x\in X} \left(\alpha u(x) + (1-\alpha)u'(x)\right) g(x) \in \hat{F}(\alpha u + (1-\alpha)u').$$

6.6 Proof of Theorem 4.5

Assume that X is a compact subset of a simplex $\Delta(K)$, and let $\Psi = (X, A, q, g)$ be a standard MDP such that: $\forall x \in X, \forall y \in X, \forall a \in A, \forall f \in D_1, \forall \alpha \geq 0, \forall \beta \geq 0$,

$$|\alpha f(q(x,a)) - \beta f(q(y,a))| \leq \|\alpha x - \beta y\|_1 \text{ and } |\alpha g(x,a) - \beta g(y,a)| \leq \|\alpha x - \beta y\|_1.$$

We write $Z_c = \Delta_f(X) \times [0,1]$, and $\overline{Z}_c = \Delta(X) \times [0,1]$. We will use the metric d_* introduced previously and its restriction to $\Delta(X)$, so that $\overline{Z}_c$ is a compact metric space. For all $(u,y), (u',y') \in \Delta_f(X) \times [0,1]$, we put $d((u,y),(u',y')) = \max(d_*(u,u'), |y-y'|)$ so that (Z_c, d) is a precompact metric space. Recall we have defined the correspondence $\hat{F}$ from Z_c to itself such that for all (u,y) in Z_c,

$$\hat{F}(u,y) = \left\{(Q(u,\sigma), G(u,\sigma)) \ s.t. \ \sigma : X \to \Delta_f(A)\right\},$$

with the notations $Q(u,\sigma) = \sum_{x\in X} u(x)q(x,\sigma(x))$ and $G(u,\sigma) = \sum_{x\in X} u(x)g(x,\sigma(x))$. We simply define the payoff function r from Z_c to $[0,1]$ by $r(u,y) = y$ for all (u,y) in Z_c. We start with a crucial lemma, which shows the importance of the duality formula of Theorem 3.10.

Lemma 6.16. *$\hat{F}$ is an affine and non-expansive correspondence from Z_c to itself.*

Proof of Lemma 6.16. We first show that: $\forall u, u' \in \Delta_f(X)$, $\forall \alpha \in [0,1]$, $\forall y, y' \in [0,1]$, $\hat{F}(\alpha u + (1-\alpha)u', \alpha y + (1-\alpha)y') = \alpha\hat{F}(u,y) + (1-\alpha)\hat{F}(u',y')$. First the transition does not depend on the second coordinate so we can forget it for the rest of the proof. The $\subset$ part is clear. To see the reverse inclusion, consider $\sigma : X \to \Delta_f(A)$, $\sigma' : X \to \Delta_f(A)$ and $v = \alpha \sum_{x\in X} u(x)q(x,\sigma(x)) + (1-\alpha)\sum_{x\in X} u'(x)q(x,\sigma'(x))$ in $\alpha\hat{F}(u) + (1-\alpha)\hat{F}(u')$. Define

$$\sigma^*(x) = \frac{\alpha u(x)\sigma(x) + (1-\alpha)u'(x)\sigma'(x)}{\alpha u(x) + (1-\alpha)u'(x)},$$

for each x such that the denominator is positive. Then $v = \sum_{x\in X}(\alpha u + (1-\alpha)u'(x))q(x,\sigma^*(x))$, and $\hat{F}$ is affine.

We now prove that $\hat{F}$ is non-expansive. Let $z = (u, y)$ and $z' = (u', y')$ be in Z_c. We have $d((u, y),(u', y')) \geq d_*(u, u')$ and denote by U and U' the respective supports of u and u'. By the duality formula of Theorem 3.10, there exists $\alpha = (\alpha(p, p'))_{(p,p')\in U\times U'}$ and $\beta = (\beta(p, p'))_{(p,p')\in U\times U'}$ with non-negative coordinates satisfying: $\sum_{p'\in U'} \alpha(p, p') = u(p)$ for all $p \in U$, $\sum_{p\in U} \beta(p, p') = u'(p')$ for all $p' \in U'$, and

$$d_*(u, u') = \sum_{(p,p')\in U\times U'} \|p\ \alpha(p, p') - p'\ \beta(p, p')\|_1.$$

Consider now $v = Q(u, \sigma) = \sum_{p\in U} u(p)q(p, \sigma(p))$ for some $\sigma : X \to \Delta_f(A)$. We define for all p' in U':

$$\sigma'(p') = \sum_{p\in U} \frac{\beta(p, p')}{u'(p')}\sigma(p),$$

and $v' = Q(u', \sigma') = \sum_{p'\in U'} u'(p')q(p', \sigma'(p'))$. Then $v' \in \hat{F}(u', y')$, and for each test function φ in D_1 we have:

$$\begin{aligned}
|\varphi(v) - \varphi(v')| &= |\sum_{p,p'} \alpha(p, p')\varphi(q(p, \sigma(p))) - \beta(p, p')\varphi(q(p', \sigma(p)))| \\
&= |\sum_{p,p',a} \alpha(p, p')\sigma(p)(a)\varphi(q(p, a)) - \beta(p, p')\sigma(p)(a)\varphi(q(p', a))| \\
&\leq \sum_{p,p'} \|\alpha(p, p')p - \beta(p, p')p'\|_1 = d_*(u, u'),
\end{aligned}$$

and therefore $d_*(v, v') \leq d_*(u, u')$. In addition we have a similar result on the payoff,

$$\begin{aligned}
|G(u, \sigma) - G(u', \sigma')| &= |\sum_{p,p'} \alpha(p, p')g(p, \sigma(p)) - \beta(p, p')g(p', \sigma(p))| \\
&\leq \sum_{p,p'} \|\alpha(p, p')p - \beta(p, p')p'\|_1 \\
&\leq d_*(u, u').
\end{aligned}$$

Thus we have $d((Q(u, \sigma), G(u, \sigma)),(Q(u', \sigma'), G(u', \sigma'))) \leq d_*(u, u') \leq d(z, z')$. □

Recall that the set of invariant couples of the MDP Ψ is:

$$RR = \{(u, y) \in \overline{Z}_c, ((u, y),(u, y)) \in cl(Graph(\hat{F}))\},$$

and the function $v^* : X \longrightarrow I\!R$ is defined by:

$$\begin{aligned}
v^*(x) \quad = \inf \quad &\big\{w(x), w : \Delta(X) \to [0, 1] \text{ affine } \textit{continuous s.t.} \\
&(1)\ \forall y \in X, w(y) \geq \sup_{a\in A} w(q(y, a)) \text{ and } (2)\ \forall (u, y) \in RR, w(u) \geq y\ \big\}.
\end{aligned}$$

We now consider the deterministic Gambling gouse $\hat{\Gamma} = (Z_c, \hat{F}, r)$. Z_c is precompact metric, $\hat{F}$ is affine non-expansive and r is obviously affine and uniformly continuous. Given an evaluation θ, the θ-value of $\hat{\Gamma}$ at $z_0 = (u, y)$ is denoted by $\hat{v}_\theta(u, y) = \hat{v}_\theta(u)$ and does not depend on y. The recursive formula of Section 2 yields:

$$\begin{aligned}\forall (u, y) \in Z, \ \ \hat{v}_\theta(u) &= \sup_{(u', y') \in \hat{F}(u)} \theta_1 y' + (1 - \theta_1)\hat{v}_{\theta^+}(u') \\ &= \sup_{\sigma \in X \to \Delta_f(A)} \left(\theta_1 G(u, \sigma) + (1 - \theta_1)\hat{v}_{\theta^+}(Q(u, \sigma))\right).\end{aligned}$$

Because $\hat{F}$ and r are affine, $\hat{v}_\theta$ is affine in u and the supremum in the above expression can be taken over functions from X to A. Because $\hat{F}$ is non-expansive and r is 1-Lipschitz, each $\hat{v}_\theta$ is 1-Lipschitz.

We denote by v_θ the θ-value of the MDP Ψ and linearly extend it to $\Delta_f(X)$. It turns out that the recursive formula satisfied by v_θ is similar to the above recursive formula for $\hat{v}_\theta$, so that $v_\theta(u) = \hat{v}_\theta(u, y)$ for all u in $\Delta_f(X)$ and y in $[0, 1]$. As a consequence, the existence of the general limit value in both problems $\hat{\Gamma}$ and Ψ is equivalent. Moreover, a deterministic play in $\hat{\Gamma}$ induces a strategy in Ψ, so that the existence of the general uniform value in $\hat{\Gamma}$ will imply the existence of the general uniform value in Ψ (note that deterministic and mixed plays in $\hat{\Gamma}$ are equivalent since $\hat{F}$ has convex values).

It is thus sufficient to show that $\hat{\Gamma}$ has a general uniform value given by v^*, and we can mimic the end of the proof of Theorem 2.10. Lemma 6.2 applies word for word. Finally, one can proceed almost exactly as in Propositions 6.3 and 6.4 to show that $\hat{\Gamma}$, hence Ψ, has a general uniform value given by v^*.

6.7 Proof of Theorem 5.9

Assume that $\Gamma(\pi) = (K, I, J, C, D, q, g, \pi)$ is a repeated game with an informed controller, i.e. that assumptions 5.6 and 5.7 are satisfied. The proof will consist of 5 steps. First we introduce an auxiliary standard Markov Decision Process $\Psi(\hat{\pi})$ on the state space $X = \Delta(K)$. Then we show that for all evaluations θ, the repeated game $\Gamma(\pi)$ and the MDP $\Psi(\hat{\pi})$ have the same θ-value. In step 3 we check that the MDP satisfies the assumption of Theorem 4.5 so it has a general limit value and a general uniform value v^*. As a consequence the repeated game has a general limit value $v^*(\pi)$. Then we prove that player 1 can use an ϵ-optimal strategy of the auxiliary MDP in order to guarantee $v^*(\pi) - \epsilon$ in the original game. Finally we prove that Player 2 can play by blocks in the repeated game in order to guarantee $v^*(\pi) + \epsilon$. We obtain that $v^*(\pi)$ can be guaranteed by both players in the repeated game, so it is the general uniform value of $\Gamma(\pi)$.

For every $P \in \Delta(K \times C \times D)$, we denote by $\overline{P}$ the marginal of P on $K \times D$. We denote by ψ_D the disintegration on $\Delta(K \times D)$ with respect to D (recall Theorem 3.13): for all $\mu \in \Delta(K \times D)$, $\psi_D(\mu) = \sum_{d \in D} \mu(d)\delta_{\mu(.|d)}$.

Step 1: We put $X = \Delta(K)$ and $A = \Delta(I)^K$ and for every p in X, a in A and b in $\Delta(J)$, we define:

$$\begin{aligned}
r(p,a,b) &= \sum_{(k,i,j)\in K\times I\times J} p^k a^k(i) b(j) g(k,i,j) \in [0,1],\\
R(p,a) &= \inf_{b\in\Delta(J)} r(p,a,b) = \inf_{j\in J} r(p,a,j),\\
\overline{q}(p,a) &= \sum_{(k,i)\in K\times I} p^k a^k(i) \overline{q}(k,i) \in \Delta(K\times D),\\
Q(p,a) &= \psi_D(\overline{q}(p,a)) = \sum_{d\in D} \overline{q}(p,a)(d)\delta_{\chi(p,a,d)} \in \Delta_f(X).
\end{aligned}$$

Since $\overline{q}(p,a)$ is a probability distribution over $\Delta(K\times D)$. For every $d\in S$, $\overline{q}(p,a)(d)$ denotes the probability to observe d and $\chi(p,a,d)$ denotes the conditional probability on K knowing d, i.e. the belief of the second player on the new state after observing the signal d and knowing that player 1 has played a at p:

$$\forall k'\in K, \chi(p,a,d)(k') = \frac{\overline{q}(p,a)(k',d)}{\overline{q}(p,a)(d)} = \frac{\sum_k p^k q(k,a(k))(k',d)}{\sum_k p^k q(k,a(k))(d)}.$$

We define the auxiliary MDP $\Psi = (X,A,Q,R)$, and denote the θ-value in the MDP by $\hat{v}_\theta$. The MDP with initial state $\hat{\pi}$ has strong links with the repeated game $\Gamma(\pi)$.

Step 2: By proposition 4.23, part b) in Renault [32], we have for all evaluations θ with finite support:

$$v_\theta(\pi) = \hat{v}_\theta(\hat{\pi}).$$

The proof relies on the same recursive formula satisfied by v and $\hat{v}$, and the equality can be easily extended to any evaluation θ.

$$\forall\theta\in\Delta(\mathbb{N}^*), \forall p\in X,\ v_\theta(p) = \sup_{a\in A}\inf_{b\in B}\left(\ \theta_1 r(p,a,b) + (1-\theta_1)v_{\theta^+}(Q(p,a))\ \right).$$

where v_{θ^+} is naturally linearly extended to $\Delta_f(X)$. As a consequence if $\Psi(\hat{\pi})$ has a general limit value so does the repeated game $\Gamma(\pi)$.

Step 3: Let us check that Ψ satisfies the assumption of Theorem 4.5. Consider p, p' in X, a in A, and $\alpha\geq 0$ and $\beta\geq 0$. We have:

$$\begin{aligned}
|\alpha R(p,a) - \beta R(p',a)| &\leq \sup_{b\in\Delta(J)} |\alpha r(p,a,b) - \beta r(p',a,b)|\\
&\leq \sup_{b\in\Delta(J)} \left|\sum_{k\in K} \alpha p^k g(k,a^k,b) - \beta p'^k g(k,a^k,b)\right|\\
&\leq \sup_{b\in\Delta(J)} \sum_{k\in K} \left|\alpha p^k - \beta p'^k\right| = \|\alpha p - \beta p'\|_1.
\end{aligned}$$

Moreover, let $\varphi:\Delta(K)\longrightarrow\mathbb{R}$ be in D_1.

$$\begin{aligned}
|\alpha\varphi(Q(p,a)) - \beta\varphi(Q(p',a))| &= \sum_{d\in D} \left(\alpha\overline{q}(p,a)(d)\varphi(\chi(p,a,d)) - \beta\overline{q}(p',a)(d)\varphi(\chi(p',a,d))\right)\\
&\leq \sum_{d\in D} \|\alpha\ \overline{q}(p,a)(d)\ \chi(p,a,d) - \beta\ \overline{q}(p',a)(d)\ \chi(p',a,d)\|_1\\
&\leq \sum_{d\in D} \|\alpha\ (\bar{q}(p,a)(k',d))_{k'} - \beta\ (\bar{q}(p',a)(k',d))_{k'}\|_1\\
&\leq \sum_{d\in D}\sum_{k\in K} \|\alpha p^k\ (\bar{q}(k,a)(k',d))_{k'} - \beta p'^k\ (\bar{q}(k,a)(k',d))_{k'}\|_1\\
&\leq \sum_{d\in D}\sum_{k'\in K}\sum_{k\in K} \bar{q}(k,a)(k',d)\left|\alpha p^k - \beta p'^k\right| = \|\alpha p - \beta p'\|_1.
\end{aligned}$$

So $\Psi = (X, A, Q, R)$ has a general limit value and a general uniform value that we denote by v^*. As a consequence, $\Gamma(\pi)$ has a general limit value $v^*(\pi)$.

Step 4: Given $\varepsilon > 0$, there exist $\alpha > 0$ and a strategy σ in the MDP $\Psi(\hat{\pi})$ such that the θ-payoff in the MDP is large: $\hat{\gamma}_\theta(\hat{\pi}, \sigma) \geq v^*(\pi) - \varepsilon$ whenever $TV(\theta) \leq \alpha$. Moreover if we look at the end of the proof of Theorem 4.5 we can choose σ to be induced by a deterministic play in the Gambling gouse $\hat{\Gamma}$ with state space $Z_c = \Delta_f(X) \times [0, 1]$. As a consequence one can mimic σ to construct a strategy σ^* in the original repeated game $\Gamma(\pi)$ such that: $\forall \tau \in \mathcal{T}, \gamma_\theta(\pi, \sigma^*, \tau) \geq v^*(\pi) - \varepsilon$ whenever $TV(\theta) \leq \alpha$.

Step 5: Finally we show that player 2 can also guarantee the value v^* in the repeated game Γ. Note that in the repeated game he cannot compute the state variable in $\Delta(K)$ without knowing the strategy of player 1. Nevertheless he has no influence on the transition function so playing independently by large blocks will be sufficient for him in order to guarantee $v^*(\pi)$. We use the following characterization of the value proved in Renault [32]:

$$v^*(\pi) = \inf_n \sup_m v_{m,n}(\pi).$$

where $v_{m,n}$ is the value of the game with payoff function the Cesàro mean of the stage payoffs between stages $m+1$ and $m+n$. We proceed as in proposition 4.22 of Renault [32]. Fix $n_0 \geq 1$, then we consider the strategy τ^* which for each $j \in I\!N$, plays optimally in the game with evaluation the Cesàro mean of the payoffs on the block of stages $B^j = \{n_0(j-1)+1, ..., n_0 j\}$. Since player 2 does not influence the state, τ^* is well defined and guarantees $\sup_{t \geq 0} v_{t,n_0}(z)$ on each block B^j.

Let θ be an evaluation and σ be a strategy of player 1. For each $j \geq 1$, denote by $\underline{\theta_j}$ the minimum of θ on the block B^j. We have

$$\begin{aligned}
\gamma_\theta(\pi, \sigma, \tau^*) &= \sum_{j=1}^{+\infty} I\!E_{\pi,\sigma,\tau^*} \left(\sum_{t=(j-1)n_0+1}^{jn_0} \theta_t \, g(k_t, a_t, b_t) \right) \\
&\leq \sum_{j=1}^{+\infty} n_0 \, \underline{\theta}_j \sup_{t \geq 0} v_{t,n_0}(\pi) + n_0 \sum_{t=1}^{+\infty} |\theta_{t+1} - \theta_t| \\
&\leq \sup_{t \geq 0} v_{t,n_0}(\pi) + n_0 TV(\theta).
\end{aligned}$$

Given ϵ, there exists n_0 such that $\sup_{t \geq 0} v_{t,n_0}(\pi) \leq v^*(\pi) + \epsilon$. Fix $\alpha = \frac{\epsilon}{n_0}$ and τ^* defined as before then for all θ such that $TV(\theta) \leq \alpha$, we have

$$\sup_{\sigma \in \Sigma} \gamma_\theta(\pi, \sigma, \tau^*) \leq v^*(\pi) + 2\epsilon,$$

and this concludes the proof of Theorem 5.9.

Acknowledgements. The authors gratefully acknowledge the support of the Agence Nationale de la Recherche, under grant ANR JEUDY, ANR-10-BLAN 0112, as well as the PEPS project Interactions INS2I "Propriétés des Jeux Stochastique de Parité à Somme Nulle avec Signaux". The second author acknowledges the support of the Israel Science Foundation under Grant ♯1517/11. The authors also thank two anonymous referees for their comments.

References

[1] Arapostathis A., Borkar V., Fernández-Gaucherand E., Ghosh M., Marcus S.: Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J. Control Opt. 31(2), 282–344 (1993)

[2] Ash R.: Real analysis and probability, vol. 239. Academic Press New York (1972)

[3] Aström K.: Optimal control of Markov processes with incomplete state information. J. Math. Anal. Appl. 10, 174–205 (1965)

[4] Aubin J.: Applied abstract analysis. John Wiley & Sons (1977)

[5] Aumann R., Maschler M., Stearns R.: Repeated games with incomplete information. The MIT press (1995)

[6] Bellman R.: A Markovian decision problem. J. Math. Mech. (1957)

[7] Bewley T., Kohlberg E.: The asymptotic theory of stochastic games. Math. Oper. Res. 1(3), 197–208 (1976)

[8] Birkhoff G.: Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA 17(12), 656-660 (1931)

[9] Blackwell D.: Discrete dynamic programming. Ann. Math. Statist. 33, 719–726 (1962)

[10] Borkar V.: Average cost dynamic programming equations for controlled Markov chains with partial observations. SIAM J. Control Opt. 39(3), 673-681 (2000)

[11] Borkar V.: Dynamic programming for ergodic control of Markov chains under partial observations: a correction. SIAM J. Control Opt. 45(6), 2299–2304 (2007)

[12] Bressaud X., Quas A.: Asymmetric warfare. Arxiv:1403.1385 (2014)

[13] Buckdahn, R., Goreac, D., Quincampoix, M.: Existence of asymptotic values for nonexpansive stochastic control systems. Appl. Math. Optim. 70(1), 1–28. Springer (2014)

[14] Choquet G.: Existence et unicité des représentations intégrales au moyen des points extrémaux dans les cônes convexes. Sém. Bourbaki 4, 33–47 (1956)

[15] Denardo E., Fox B.: Multichain Markov renewal programs. SIAM J. Appl. Math. 16(3), 468–487 (1968)

[16] Dubins L., Savage L.: How to gamble if you must: Inequalities for stochastic processes. McGraw-Hill New York (1965)

[17] Dudley R. M.: Real analysis and probability, vol. 74. Cambridge University Press (2002)

[18] Harsanyi J.: Games with incomplete information played by "Bayesian" players, I-III. part I. The basic model. Manag. Sci. 14(3), 159–182 (1967)

[19] Hordijk A., Kallenberg L.: Linear programming and Markov decision chains. Manag.Sci. 25(4), 352–362 (1979)

[20] Hörner J.,Rosenberg D.,Solan E.,Vieille N.: On a Markov game with one-sided information. Oper. Res. 58(4), 1107–1115 (2010)

[21] Lehrer E., Sorin S.: A uniform Tauberian theorem in dynamic programming, Math. Methods Oper. Res. 17(2), 303–307 (1992)

[22] Maitra A., Sudderth W.: Discrete gambling and stochastic games, vol. 32. Springer Verlag (1996)

[23] McShane E.J.: Extension of range of functions. Bull. Amer. Math. Soc. (N.S.) 40.12, 837–842 (1934)

[24] Mertens J.-F., Neyman A.: Stochastic games. Internat. J. Game Theory 10(2), 53–66 (1981)

[25] Mertens J.-F., Zamir S.: Formulation of Bayesian analysis for games with incomplete information, International J. Game Theory 14(1), 1–29 (1985)

[26] Mertens J.-F.: Repeated games, in Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986), (Providence, RI), pp. 1528–1577, Amer. Math. Soc. (1987)

[27] Mertens J.F., Sorin S., Zamir S.: Repeated games, CORE Discussion Papers, vol. 9420,9421,9422 (1994)

[28] Neyman A.: Existence of optimal strategies in markov games with incomplete information. Internat. J. Game Theory 37(4), 581–596 (2008)

[29] Quincampoix, M. and Renault, J.: On the existence of a limit value in some nonexpansive optimal control problems. SIAM J. Control Optim. 49(5), 2118–2132 (2011)

[30] Renault J.: The value of Markov chain games with lack of information on one side. Math. Oper. Res. 31(3), 490–512 (2006)

[31] Renault J.: Uniform value in dynamic programming. J. Eur. Math. Soc. 13(2), 309–330 (2011)

[32] Renault J.: The value of repeated games with an informed controller. Math. Oper. Res. 37(1), 154–179 (2012)

[33] Renault J.: General long-term values in dynamic programming. J. Dynam. Games 3(1), 471–484 (2014)

[34] Rhenius D.: Incomplete information in Markovian decision models. Ann. Statist. 2(6), 1327–1334 (1974)

[35] Rosenberg D., Solan E., Vieille N.: Blackwell optimality in Markov decision processes with partial observation. Ann. Statist. 30(4), 1178–1193 (2002)

[36] Rosenberg D., Solan E., Vieille N.: Stochastic games with a single controller and incomplete information. SIAM J. Control Optim. 43(1), 86–110 (electronic) (2004)

[37] Runggaldier W. J., Stettner Ł.: On the construction of nearly optimal strategies for a general problem of control of partially observed diffusions, Stochastics: Intern. J.Prob. Stochastic Proc. 37(1-2), 15–47 (1991)

[38] Sawaragi Y., Yoshikawa T.: Discrete-time Markovian decision processes with incomplete state observation. Ann. Math. Statist. 41(1), 78–86 (1970)

[39] Shapley L.: Stochastic games. Proc. Natl. Acad. Sci. U. S. A. 39(10), 1095–1100 (1953)

[40] Sorin S.: A first course on zero-sum repeated games, vol. 37. Springer (2002)

[41] Von Neumann J.: Proof of the quasi-ergodic hypothesis. Proc. Nat. Acad. Sci. U.S.A. 18(1), 70–82 (1932)