On the Structure of Expected Payoff Sets in Multi-Objective Markov Decision Processes

James C. A. Main Mickaël Randour

F.R.S.-FNRS and UMONS – Université de Mons, Belgium

MoVe Seminar – October 4, 2024

Random strategies and multiple objectives

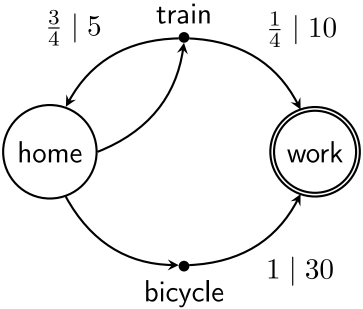
- We study **Markov decision processes** with **multiple payoffs**.
- In general, the satisfaction of multi-objective queries requires **randomised strategies**.

Main questions

- What is the relationship between expected payoffs of **pure strategies** and expected payoffs of **general strategies**?
- What **type of randomisation** do we need for multi-objective queries?

$\rightarrow$ **Goal**: results for the **broadest possible class of payoffs**.

Markov decision processes



Markov decision process $\mathcal{M}$

- Finite state space S
- Finite action space A
- Randomised transitions

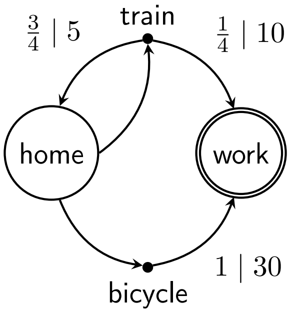
Plays are sequences in $(SA)^\omega$ coherent with transitions.

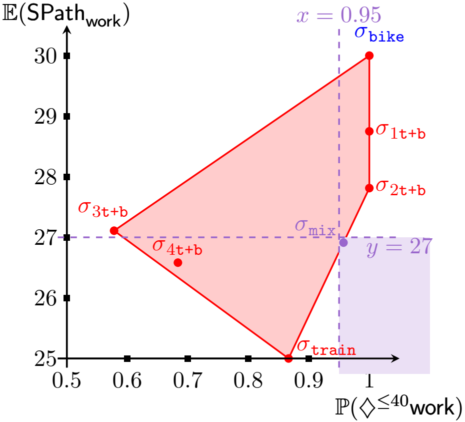
- A **strategy** is a function $\sigma\colon (SA)^*S \to \mathcal{D}(A)$
- A strategy σ is **pure** if it is not randomised.
- A strategy σ and initial state s induce a **distribution** $\mathbb{P}_s^\sigma$ over plays.
- A **payoff** is a measurable function $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$.
- We let $\mathbb{E}_s^\sigma(f) = \int_{\pi \in \mathsf{Plays}(\mathcal{M})} f(\pi) \mathrm{d}\mathbb{P}_s^\sigma(\pi)$.

Multi-objective Markov decision processes

We consider **two goals**:

- reaching work under 40 minutes with **high probability**;
- minimising the **expectancy** of the time to reach work.





What are good payoffs?

To provide formal results, we need to **constrain considered payoffs**.

$\rightsquigarrow \mathbb{E}_s^\sigma(f)$ should be well-defined **for all strategies**.

Good payoff functions

Three types of **good** payoffs:

- **non-negative** payoffs: $f \geq 0$;
- **non-positive** payoffs: $f \leq 0$;
- **universally integrable** payoffs: $\mathbb{E}_s^\sigma(|f|) \in \mathbb{R}$ for all strategies σ and all $s \in S$.

For a **multi-dimensional payoff** $\bar{f} = (f_1, \ldots, f_d)$ and $s \in S$, we let:

- $\mathsf{Pay}_s(\bar{f}) = \{\mathbb{E}_s^\sigma(\bar{f}) \mid \sigma \text{ strategy}\}$;
- $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}) = \{\mathbb{E}_s^\sigma(\bar{f}) \mid \sigma \text{ pure strategy}\}$.

Universally integrable payoffs

In the introductory example, we had $\mathsf{Pay}_{\mathsf{home}}(\bar{f}) = \mathrm{conv}(\mathsf{Pay}^{\mathsf{pure}}_{\mathsf{home}}(\bar{f}))$.

When does this generalise?

Theorem

Let $\bar{f} = (f_1, \dots, f_d)$ be ***universally integrable****. Then, for all states s,*

$$\mathsf{Pay}_s(\bar{f}) = \mathrm{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})).$$

In particular, to match the expected payoff of any strategy, it suffices to:

- ***mix $d+1$ pure strategies****;*
- *consider strategies use* ***randomisation at most d*** *along any play.*

Sequel: proof of a weaker result

If $\bar{f}$ is universally integrable, then $\mathrm{cl}(\mathsf{Pay}_s(\bar{f})) = \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}^{\mathsf{pure}}_s(\bar{f})))$.

Universally integrable payoffs

A simpler proof

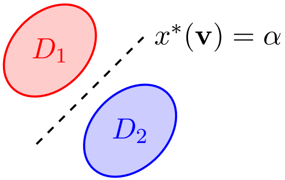
Non-direct inclusion: $\mathsf{Pay}_s(\bar{f}) \subseteq \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$.

Let σ be a strategy and $\mathbf{q} = \mathbb{E}_s^\sigma(\bar{f})$. Assume $\mathbf{q} \notin \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$.

Main idea: reduction to a **one-dimensional** payoff.

Theorem (Hyperplane separation theorem)

*Let D_1, $D_2 \subseteq \mathbb{R}^d$ be **disjoint convex** sets. If D_1 is **closed** and D_2 is **compact**, then there exists a **linear form** $x^* \colon \mathbb{R}^d \to \mathbb{R}$ and $\varepsilon > 0$ such that for all $\mathbf{p}_1 \in D_1$ and $\mathbf{p}_2 \in D_2$, $x^*(\mathbf{p}_1) + \varepsilon < x^*(\mathbf{p}_2)$.*



Universally integrable payoffs

A simpler proof

Non-direct inclusion: $\mathsf{Pay}_s(\bar{f}) \subseteq \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$.

Let σ be a strategy and $\mathbf{q} = \mathbb{E}_s^\sigma(\bar{f})$. Assume $\mathbf{q} \notin \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})))$.

Main idea: reduction to a **one-dimensional** payoff.

- There exists a linear form x^* such that, for all **pure strategies** τ,

$$x^*(\mathbb{E}_s^\tau(\bar{f})) < x^*(\mathbf{q})$$

- By linearity, we obtain that for all pure strategies τ,

$$\mathbb{E}_s^\tau(x^*(\bar{f})) < \mathbb{E}_s^\sigma(x^*(\bar{f}))$$

Lemma

Let f be ***universally integrable****. For all strategies σ, there exists a* ***pure*** *strategy τ such that $\mathbb{E}_s^\sigma(f) \leq \mathbb{E}_s^\tau(f)$.*

Universally integrable payoffs

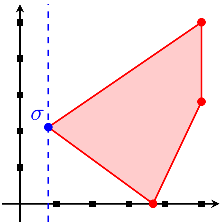
General argument

What about the general case?

1. If $\mathsf{Pay}_s(\bar{f})$ is **compact**, we have $\mathsf{Pay}_s(\bar{f}) = \mathrm{conv}(\mathrm{extr}(\mathsf{Pay}_s(\bar{f})))$.
   $\rightsquigarrow$ The argument can be adapted to **compact polyhedral sets**.

   BUT

   Good hyperplanes do not generally exist for all extreme points.
2. $\mathsf{Pay}_s(\bar{f})$ need not be closed.



- Consider a **vertex $\mathbf{q}$** obtained by σ.
- There is a **hyperplane** intersecting $\mathsf{Pay}_s(\bar{f})$ only at $\mathbf{q}$.
- There exists a **linear form** x^* such that σ is **optimal** from s for $x^* \circ \bar{f}$.

Universally integrable payoffs

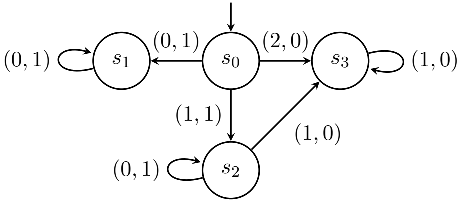
General argument

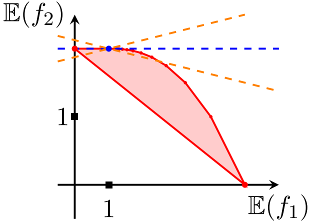
What about the general case?

1. If $\mathsf{Pay}_s(\bar{f})$ is **compact**, we have $\mathsf{Pay}_s(\bar{f}) = \mathrm{conv}(\mathrm{extr}(\mathsf{Pay}_s(\bar{f})))$.
   $\rightsquigarrow$ The argument can be adapted to **compact polyhedral sets**.

   BUT

   Good hyperplanes do not generally exist for all extreme points.
2. $\mathsf{Pay}_s(\bar{f})$ need not be closed.





$$f_1(s_0a_0s_1\dots) = \sum_{\ell=0}^{\infty}(\tfrac{3}{4})^\ell w_1(a_\ell);\ f_2(s_0a_0s_1\dots) = \sum_{\ell=0}^{\infty}(\tfrac{1}{2})^\ell w_2(a_\ell).$$

Universally integrable payoffs

General argument: sketch

Proof goal

For all strategies σ, $\mathbf{q} = \mathbb{E}_s^\sigma(\bar{f}) \in \operatorname{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\mathbf{q}))$.

- Construct linear map $L_\mathbf{q}$ such that:
  - a σ is **lexicographically optimal** for $L_\mathbf{q} \circ \bar{f}$;
  - b $\mathbf{q} \in \operatorname{ri}(\mathsf{Pay}_s(\bar{f}) \cap V)$ where $V = \{\mathbf{p} \in \mathbb{R}^d \mid L_\mathbf{q}(\mathbf{p}) = L_\mathbf{q}(\mathbf{q})\}$.
- Show that $\operatorname{ri}(\mathsf{Pay}_s(\bar{f}) \cap V) = \operatorname{ri}(\operatorname{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})) \cap V)$, i.e.,

$$\operatorname{cl}(\mathsf{Pay}_s(\bar{f}) \cap V) = \operatorname{cl}(\operatorname{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})) \cap V)$$

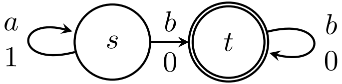
Key lemma

If $\bar{f}$ is universally integrable, then for all strategies σ and all $s \in S$, there exists a pure strategy τ such that

$$\mathbb{E}_s^\sigma(\bar{f}) \leq_{\mathsf{lex}} \mathbb{E}_s^\tau(\bar{f}).$$

Beyond universally integrable payoffs

Example



Payoffs

1. reaching $t \rightsquigarrow f_1 = \mathbb{1}_{\Diamond t}$;
2. sum of weights $\rightsquigarrow f_2 = \sum_{\ell=0}^{\infty} w(c_\ell)$.

- $\mathbb{E}_s^{\sigma_a}(f_2) = +\infty \implies f_2$ is **not universally integrable**.
- $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}) = \{(0, +\infty)\} \cup \{(1, \ell) \mid \ell \in \mathbb{N}\}$.
  $\implies \mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})) = (\{1\} \times \mathbb{R}_{\geq 0}) \cup ([0, 1[\times \{+\infty\})$.
- We have $(1, +\infty) \in \mathsf{Pay}_s(\bar{f})$ via σ such that for all $\ell \in \mathbb{N}$:

$$\sigma(s(as)^\ell)(a) = \begin{cases} \frac{1}{2} & \text{if } \ell \in 2^{\mathbb{N}} \\ 1 & \text{if } \ell \notin 2^{\mathbb{N}} \end{cases}$$

$\rightarrow$ **The theorem and the key lemma do not generalise.**

Beyond universally integrable payoffs

Theorem

Let $\bar{f} = (f_1, \ldots, f_d)$ be a good payoff and $s \in S$.
For all strategies σ, all $\varepsilon > 0$ and all $M \in \mathbb{R}$, there exist finitely many pure strategies $\tau_1, \ldots, \tau_n$ and coefficients $\alpha_1, \ldots, \alpha_n \in [0, 1]$ such that $\sum_m^n \alpha_m = 1$ and for all $1 \leq j \leq d$:

- *if $\mathbb{E}_s^\sigma(f_j) = +\infty$, then $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \geq M$,*
- *if $\mathbb{E}_s^\sigma(f_j) = -\infty$, then $\sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq -M$, and,*
- *otherwise, if $\mathbb{E}_s^\sigma(f_j) \in \mathbb{R}$,*
  $\mathbb{E}_s^\sigma(f_j) - \varepsilon \leq \sum_{m=1}^n \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq \mathbb{E}_s^\sigma(f_j) + \varepsilon$.

Informally, we have

$$\mathrm{cl}(\mathsf{Pay}_s(\bar{f})) = \mathrm{cl}(\mathrm{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f}))).$$

Thank you for your attention !

A set of expected payoffs that is not closed

For $j \in \{1, 2\}$, we consider the payoff f_j such that, for all plays $s_0 a_0 s_1 \ldots$,

$$f_j(s_0 a_0 s_1 \ldots) = \mathbb{1}_{\mathsf{Reach}(\{t\})}(s_0 a_0 s_1 \ldots) \cdot \sum_{\ell=0} \left(\frac{3}{4}\right)^{\ell} w_j(a_\ell).$$