Mixing Any Cocktail with Limited Ingredients: On the Structure of Payoff Sets in Multi-Objective POMDPs and its Impact on Randomised Strategies

James C. A. Main[1] Mickael Randour[2]

[1]University of Oxford, UK

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

LICS 2026 – July 22, 2026

What is this talk about?

On the Structure of **Payoff Sets** in **Multi-Objective POMDPs** and its Impact on **Randomised Strategies**

- **POMDPs**: partially observable Markov decision processes.
- **Multi-objective**: the goal is to optimise several payoff functions at once.
- **Payoff sets**: the set of expected payoffs that can be obtained using some strategy.
- **Randomised strategies**: randomised strategies are necessary in general

Motivation and contributions

Main motivation: what is the **complexity of strategies** required for multi-objective constraints in POMDPs?

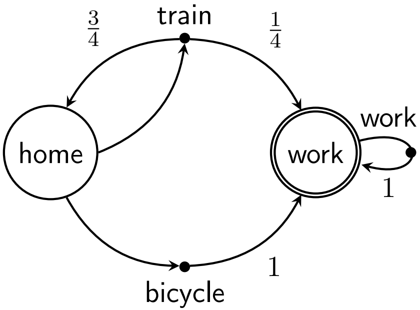
- Strategy complexity is typically measured by **memory requirements**.

What can we say about randomisation requirements in multi-objective POMDPs?

We investigate the relationship between expected payoffs of **pure strategies** and expected payoffs of **general strategies** and the necessary **type of randomisation** for multi-objective queries.

We only consider **fully observable MDPs** in the remainder of this talk → our results generalise to POMDPs directly.

Countable Markov decision processes



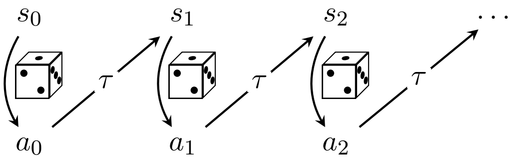
Markov decision process $\mathcal{M}$:

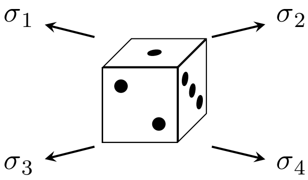
- **Countable** state space S
- **Countable** action space A
- **Randomised** transitions

- **Plays** are sequences in $(SA)^\omega$ coherent with transitions.
- A **payoff** is a measurable function $f\colon \mathsf{Plays}(\mathcal{M}) \to \bar{\mathbb{R}}$.

Strategies

A **strategy** describes how to play. A **pure strategy** is a function $\sigma\colon (SA)^*S \to A$.



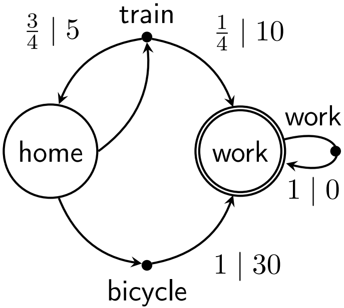


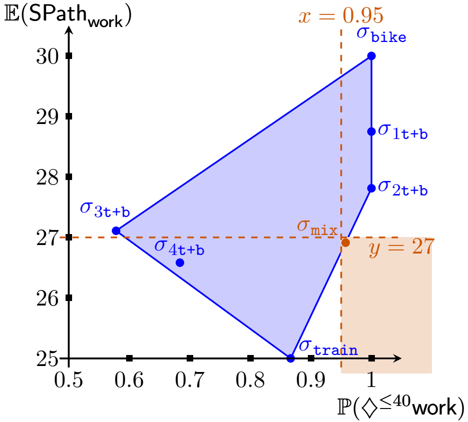
Behavioural strategy: $\sigma\colon (SA)^*S \to \mathcal{D}(A)$

Mixed strategy: $\mathcal{D}(\sigma\colon (SA)^*S \to A)$

- A strategy σ and initial state s induce a **distribution** $\mathbb{P}_s^\sigma$ over **plays**.
- We let $\mathbb{E}_s^\sigma(f)$ be the **expectation** with respect to $\mathbb{P}_s^\sigma$.

Multi-objective Markov decision processes





We consider **two goals**:

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

What are good payoffs?

We **constrain payoffs** to ensure that **all expectations are well-defined**.

For a given initial state s, we consider two types of **good payoffs**:

- **universally integrable payoffs**: $\mathbb{E}_s^\sigma(|f|) \in \mathbb{R}$ for all strategies σ.
- **universally unambiguously integrable payoffs**: $\mathbb{E}_s^\sigma(\max(f, 0)) \in \mathbb{R}$ or $\mathbb{E}_s^\sigma(\max(-f, 0)) \in \mathbb{R}$ for all strategies σ.

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

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

Question: What is the relationship between $\mathsf{Pay}_s(\bar{f})$ and $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$?

Universally integrable payoffs

Main theorem

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

When does this generalise?

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

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

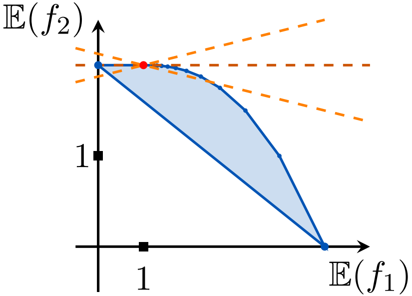
In particular, to match the expected payoff of any strategy, it suffices to:

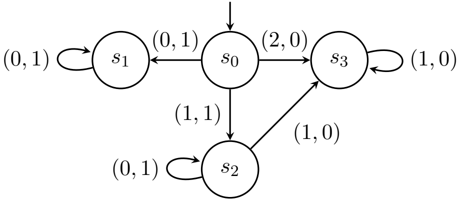
- **mix $d+1$ pure strategies**;
- consider strategies that use **randomisation at most d** along any play.

Universally integrable payoffs

A compact example

For **compact payoff sets**, it is sufficient to show that **extreme points** can be achieved by **pure strategies**.





$$f_1(s_0 a_0 s_1 \dots) = \sum_{\ell=0}^{\infty} \left(\tfrac{3}{4}\right)^{\ell} w_1(a_\ell)$$

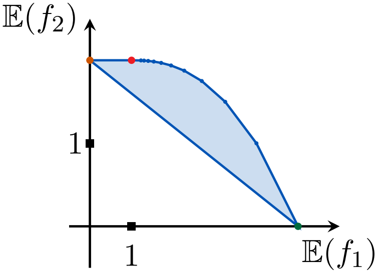
$$f_2(s_0 a_0 s_1 \dots) = \sum_{\ell=0}^{\infty} \left(\tfrac{1}{2}\right)^{\ell} w_2(a_\ell).$$

Lexicographic optimisation

We rely on a reduction to **lexicographic optimisation** to show our main theorem.

Lemma. Let $\bar{f} = (f_1, \ldots, f_d)$ be **universally integrable**. For all states s and strategies σ, there exists a **pure strategy** τ such that

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

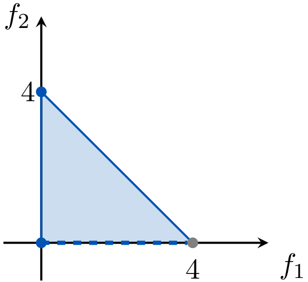


How to transform payoffs to make points **lexicographically optimal**?

- For $(4, 0)$, consider f_1;
- for $(0, 2)$, consider $-f_1$;
- for $(1, 2)$, consider (f_2, f_1).

General payoff sets

In general, expected payoff sets of universally integrable payoffs may **not be compact**.

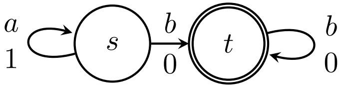


- We reason **face by face** of the payoff set.
- We show that each face is the set of **lexicographic optima** for some **linear function**.
- We combine this function with **separating hyperplanes** to prove the general theorem.

Beyond universally integrable payoffs

Example

The previous theorem does not generalise to payoffs that are **not universally integrable**, even in **one-dimensional settings**.



$$f = \begin{cases} 0 & \text{if } t \text{ is not reached} \\ \sum_{\ell=0}^{\infty} w(c_\ell) & \text{otherwise.} \end{cases}$$

- We have $\mathsf{Pay}_s^{\mathsf{pure}}(f) = \mathbb{N}$, hence $\mathsf{conv}(\mathsf{Pay}_s^{\mathsf{pure}}(f)) = \mathbb{R}_{\geq 0}$.
- However, there exists a **randomised strategy** σ such that $\mathbb{E}_s^{\sigma}(f) = +\infty \notin \mathbb{R}_{\geq 0}$.

What can we say about **universally unambiguously integrable payoffs**?

Universally unambiguously integrable payoffs

Theorem. Let $s \in S$ and $\bar{f} = (f_1, \ldots, f_d)$ be **universally unambiguously integrable**. We have

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

In other words, 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 $\sum_{m=1}^{n} \alpha_m = 1$ 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 $\mathbb{E}_s^{\sigma}(f_j) \in \mathbb{R}$, $\mathbb{E}_s^{\sigma}(f_j) - \varepsilon \leq \sum_{m=1}^{n} \alpha_m \mathbb{E}_s^{\tau_m}(f_j) \leq \mathbb{E}_s^{\sigma}(f_j) + \varepsilon$.

Main idea: **approximating expectation** under general mixed strategies by the expectation of a strategy that mixes over a **countable set of strategies**.

Continuous payoff functions

Another question. When do Pareto-optimal strategies exist?

Sufficient criterion: have **closed payoffs sets**.

Definition. A payoff is **continuous** if, informally, plays that share **long common prefixes** have close payoffs.

Continuity can be used to provide a sufficient criterion for closed payoff sets.

Theorem. Assume that the POMDP is **finite**. Let $\bar{f} = (f_1, \dots, f_d)$ be a continuous payoff such that $(f_1^2, \dots, f_d^2)$ is universally integrable. Then for all states s, $\mathsf{Pay}_s(\bar{f})$ and $\mathsf{Pay}_s^{\mathsf{pure}}(\bar{f})$ are **closed subsets of $\mathbb{R}^d$**.

Idea. The **finiteness assumption** ensures that the set of plays of the POMDP is **compact**.

Conclusion

Summary:

- For **universally integrable payoffs**, **randomisation is not useful** for lexicographic optimisation and **finite-support mixed strategies** can match the expectation vector of any strategy.
- For **universally unambiguously integrable payoffs**, **finite-support mixed strategies** can approximate the expectation vector of any strategy.
- In finite **POMDPs**, **universally square integrable continuous payoffs** admit closed sets of expected payoffs.

Future work

- Characterisation of payoff functions for which **finite-memory strategies** suffice to achieve any payoff in MDPs and in POMDPs.
- Understanding when payoff functions have **closed payoff sets**.