HOW TO FIND THE EXACT PARETO FRONT FOR MULTI-OBJECTIVE MDPS?

Yining Li, Peizhong Ju, Ness Shroff

THE OHIO STATE UNIVERSITY

EDGE Institute

Introduction

Multi-objective MDPs with conflicting objectives

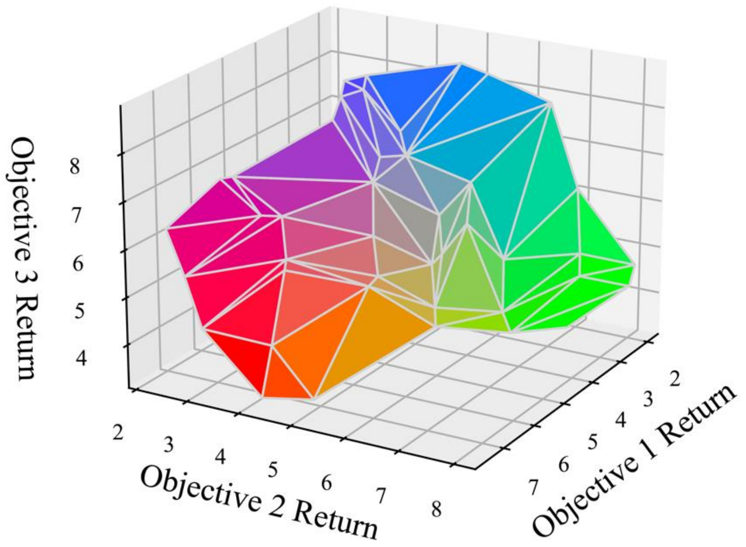
Transform the multi-objective RL to a single-objective RL problem: Capturing the true preference vector is challenging.
- balancing between objectives with different scales
- changing preference vectors

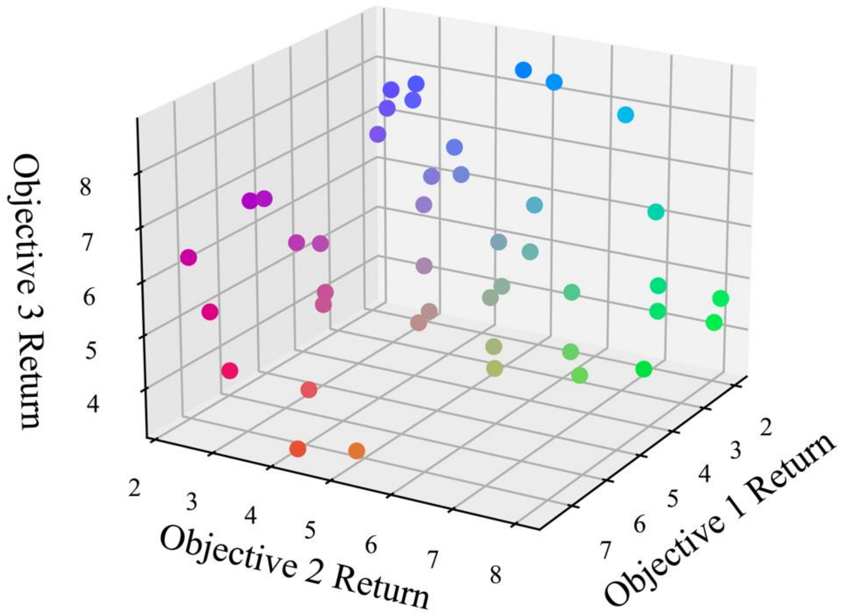
How to find the Pareto front

- Iteratively finding the preference that improves the current Pareto front the most to avoid blind traversing the preference space: constrained to find deterministic Pareto optimal policies
- Solving single-objective problems by scalarizing the multi-objective and combining Pareto optimal policies: require sampling through the continuous preference space

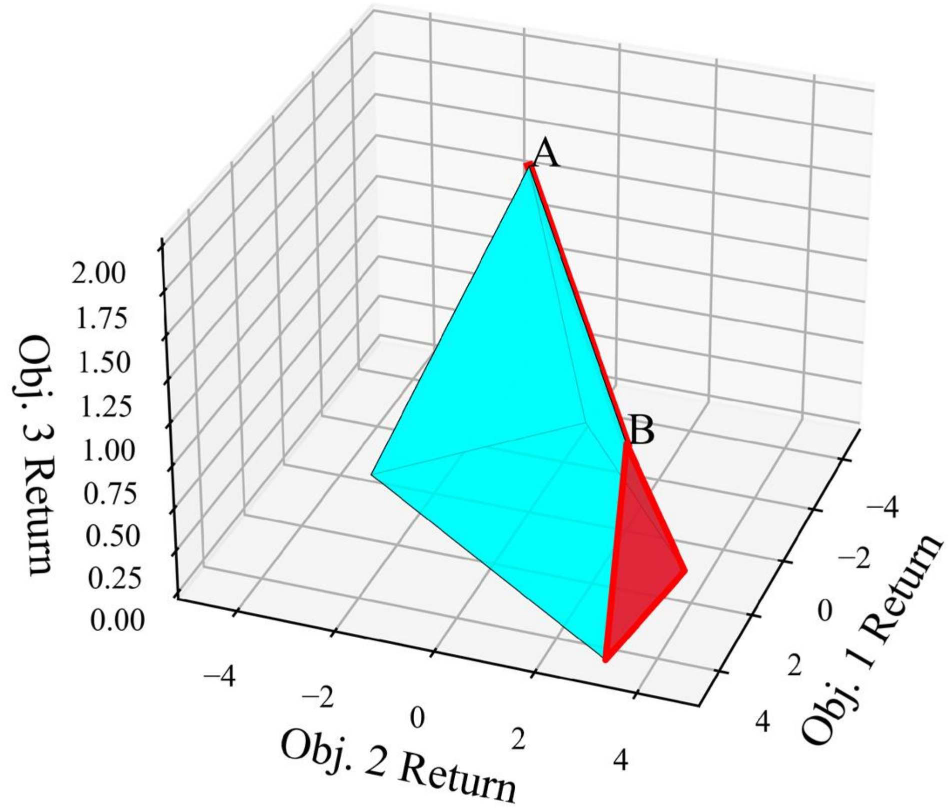
Multi-objective MDP

Deterministic Pareto-Optimal Policies





Multi-objective MDP



An edge (lower-dim face) can be part of the Pareto Front

Introduction

Question:

How can we efficiently obtain the full exact Pareto front in MO-MDPs?

Multi-objective MDP

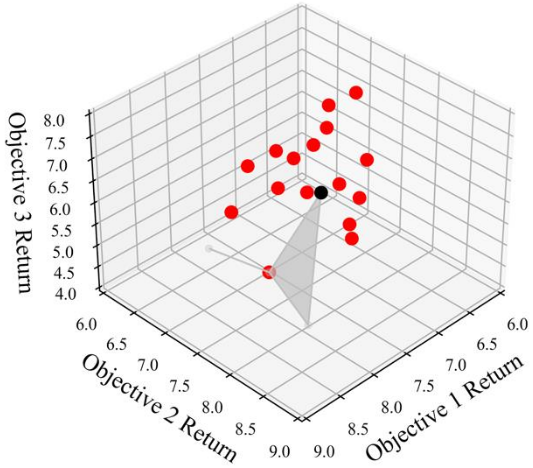
- A discounted muti-objective MDP $(S, A, \mathbf{P}, \boldsymbol{r}, \gamma)$
  - S: states
  - A: actions
  - $\mathbf{P}$: transition probability
  - $\boldsymbol{r} \in \mathrm{R}^{|S|\times|A|\times|D|}$: reward tensor, where $\boldsymbol{r}(s, a)$ is a D-dimensional reward vector whose different elements correspond to different objectives
  - γ: discount factor
- Every objective is a long-term return.
  - $\mathbf{V}^{\pi}(s) = E[\sum_t \gamma^t \boldsymbol{r}(s_t, a_t) \mid s_0 = s, a_t \sim \pi, s_{t+1} \sim \mathbf{P}]$
- Pareto optimal policies: No objective can be improved without sacrificing others.
- Pareto front: the set of all Pareto optimal policies

Our Approach: Search along the Pareto front

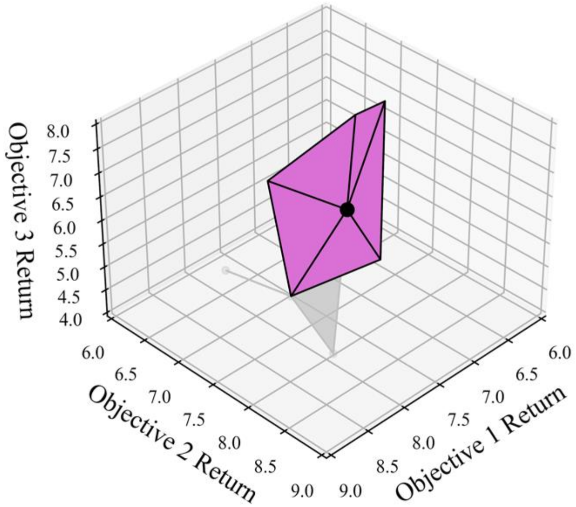
- Key geometrical characteristics:
  - Pareto front lies on the boundary of a convex polytope, with its vertices corresponding to deterministic policies.
  - Any neighboring policies on this boundary differ by only one state-action pair
- Benefits:
  - Solve the MDP only once
  - Can find the exact Pareto front
  - Much more efficient

Our Approach: Search along the Pareto front

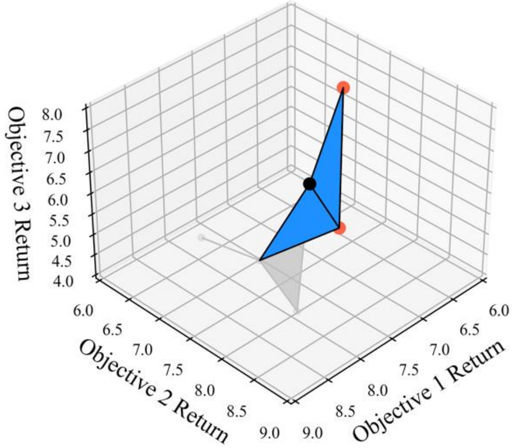
1. Initialization: solve the single-objective optimal policy with an arbitrarily chosen preference vector.
2. Loop: the total iteration number is the same as the number of vertices on the Pareto front.
   - Steps per iteration:



Step 1: Neighboring Policies Identification



Step 2: Incident Faces Calculation



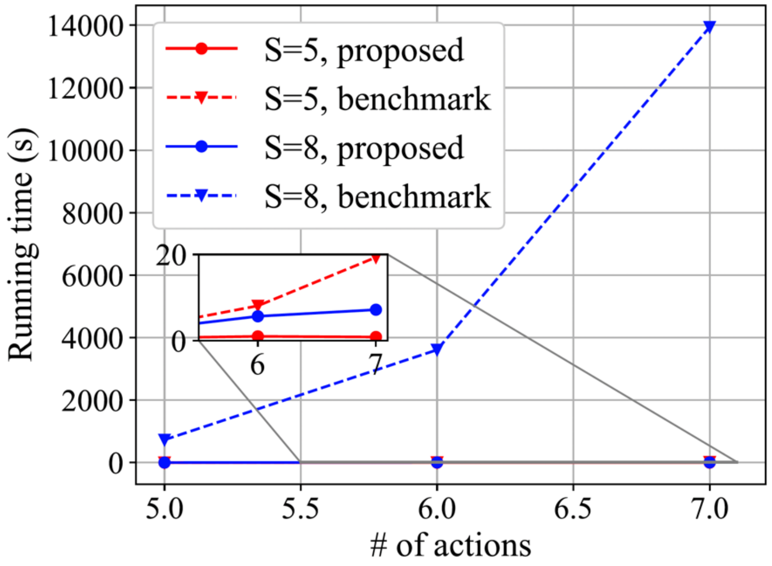
Step 3: Pareto Front Extraction

Our Approach: Search along the Pareto front

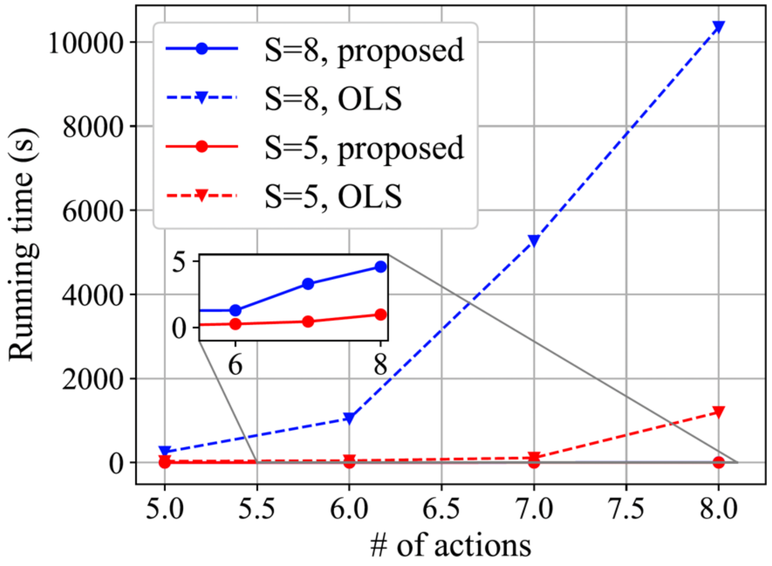
Intuitions behind theoretical analysis:

- Distance-one Property on Boundary of Pareto front: Any neighboring policies on this boundary differ by only one state-action pair.
- Sufficiency of Traversing Over Edges: The Pareto front is on the surface of the convex polytope, so the Pareto front is continuous.
- Locality Property of the Pareto front: the faces of the Pareto front intersecting at a deterministic policy can be found by computing the convex hull of the returns of this deterministic policy and non-dominated deterministic policies that differ by one state-action pair.

Efficiency Comparison



3 objectives



4 objectives