Optimistic planification of deterministic systems

Jean-François Hren, Rémi Munos

INRIA Nord-Europe SequeL Team

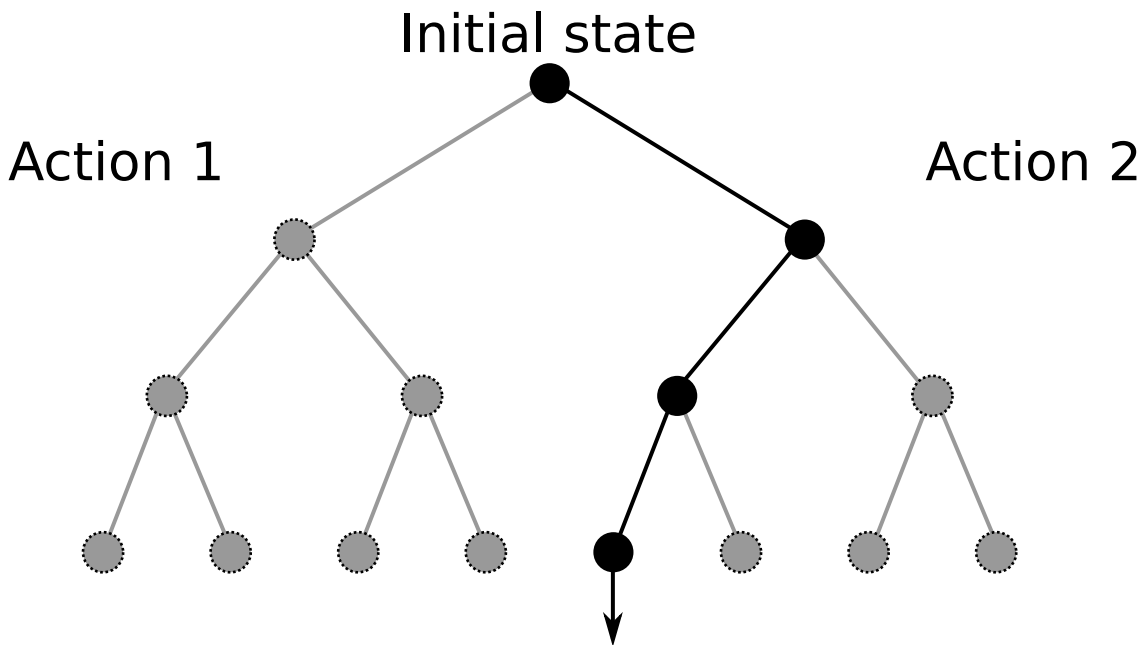
15 april 2009

Introduction

- Sequential decision making problems
- Deterministic dynamics
- Deterministic rewards ($\in [0, 1]$)
- Large state space
- Finite action space
- Finite numerical resources not know ahead of time
- Access to a generative model

Planning under finite numerical resources

- Let x_t be the current state of a real system
  - Using a generative model to create paths and rewards, build a look-ahead tree starting from x_t
- Select the action to apply to x_t on the real system
- Restart with x_{t+1}



Planning under finite numerical resources

- Finite numerical resources (number of expended nodes n)
- Anytime algorithms
- Minimization of the regret resulting from choosing the action returned by $\mathcal{A}$ instead of the optimal one

Regret of the action-selection Algorithm $\mathcal{A}$

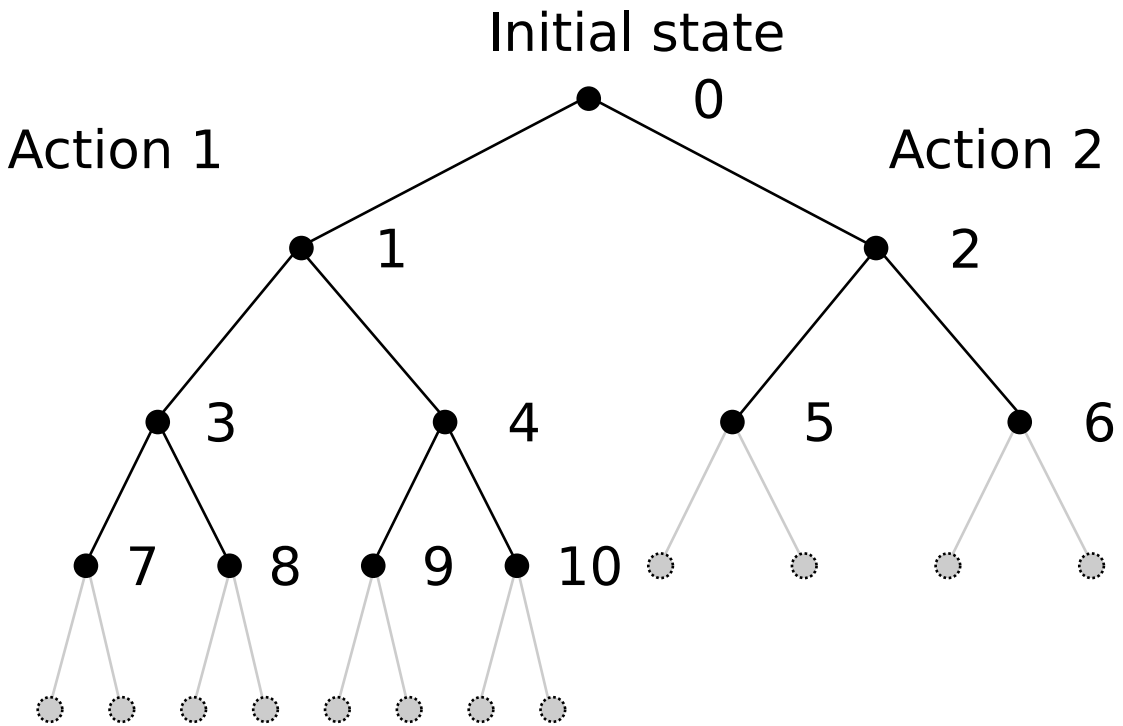
$$R_{\mathcal{A}}(n) \stackrel{\text{def}}{=} \max_{a \in A} Q^*(x,a) - Q^*(x,\mathcal{A}(n))$$

$$Q^*(x_0,a) = r(x_0,a) + \gamma \max_{b \in A} Q^*(x_1 = f(x_0,a), b)$$

Performance of the resulting policy

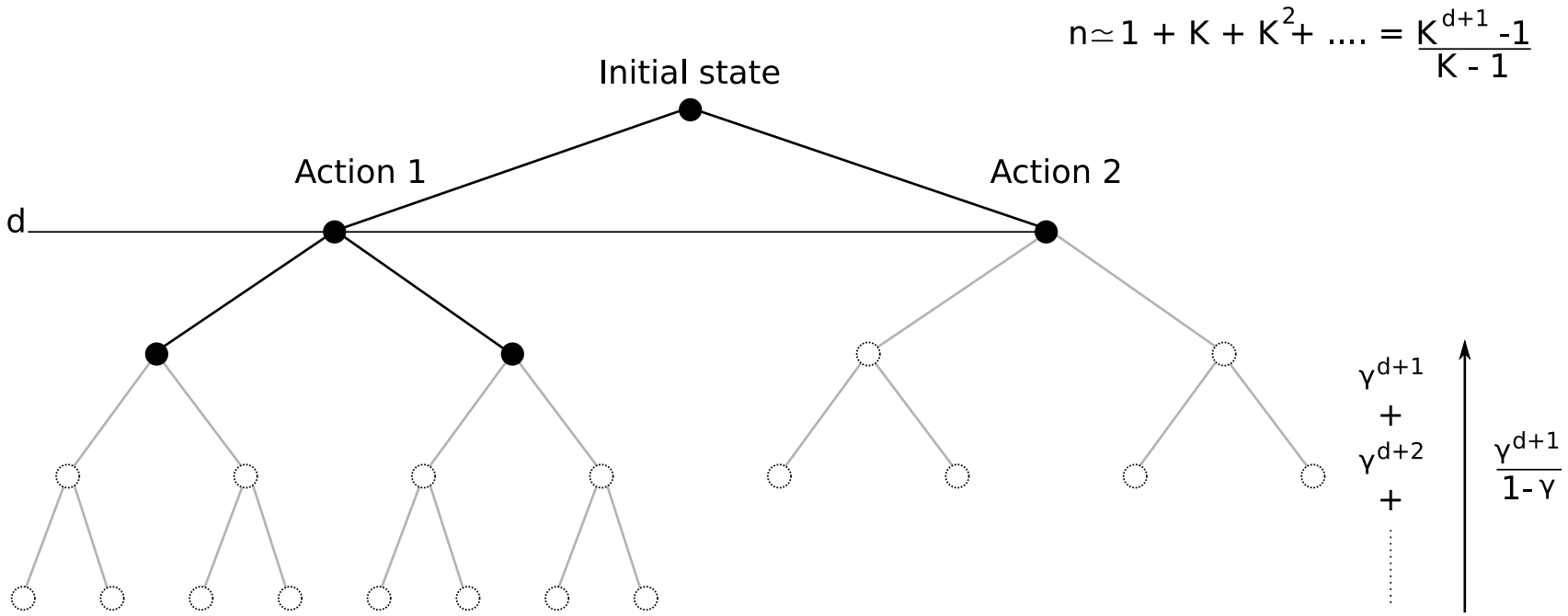
$$V^*(x) - V^{\pi_{\mathcal{A}}}(x) \leq \frac{R_{\mathcal{A}}(n)}{1-\gamma}$$

- Build the look-ahead tree in a breadth-first order while numerical resources available
- Select the action a that maximise $Q(x_0, a)$



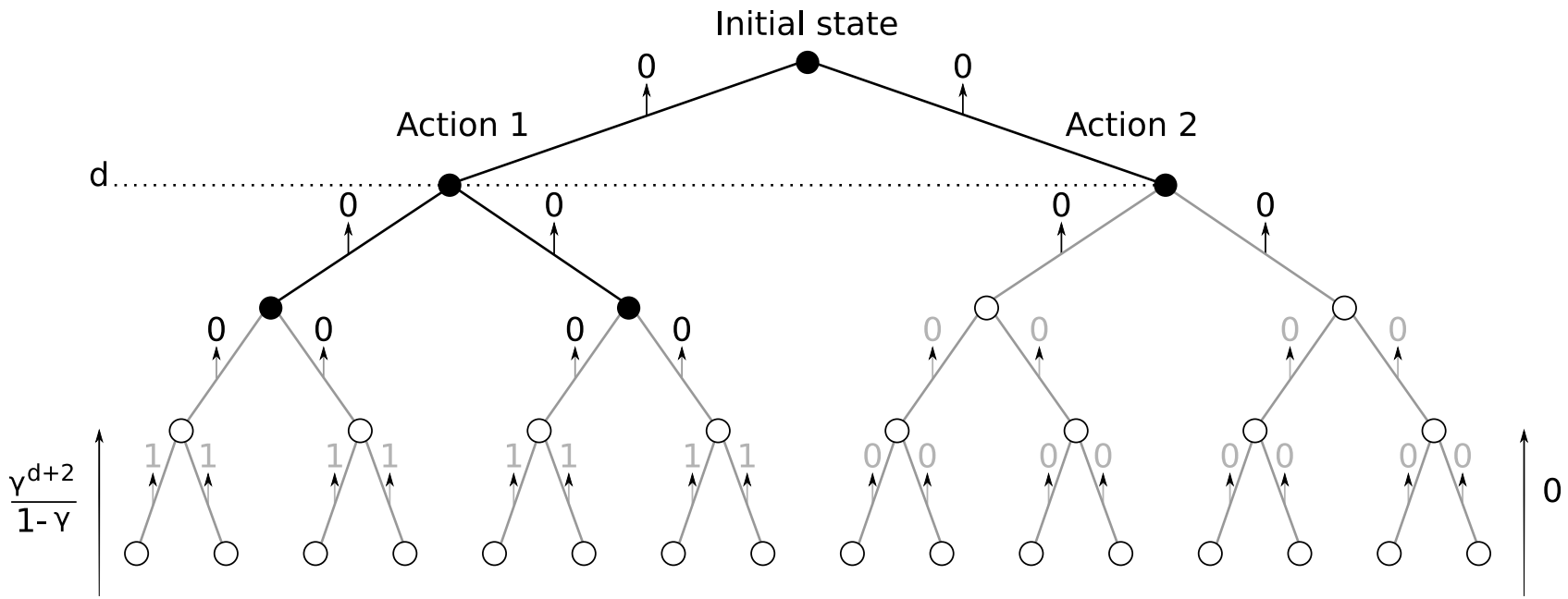
Upper bound on the regret

$$R_{\mathcal{A}_U}(n) \leq \frac{\gamma^{d+1}}{1-\gamma} = O\left(n^{-\frac{\log(1/\gamma)}{\log K}}\right)$$



Lower bound on the regret

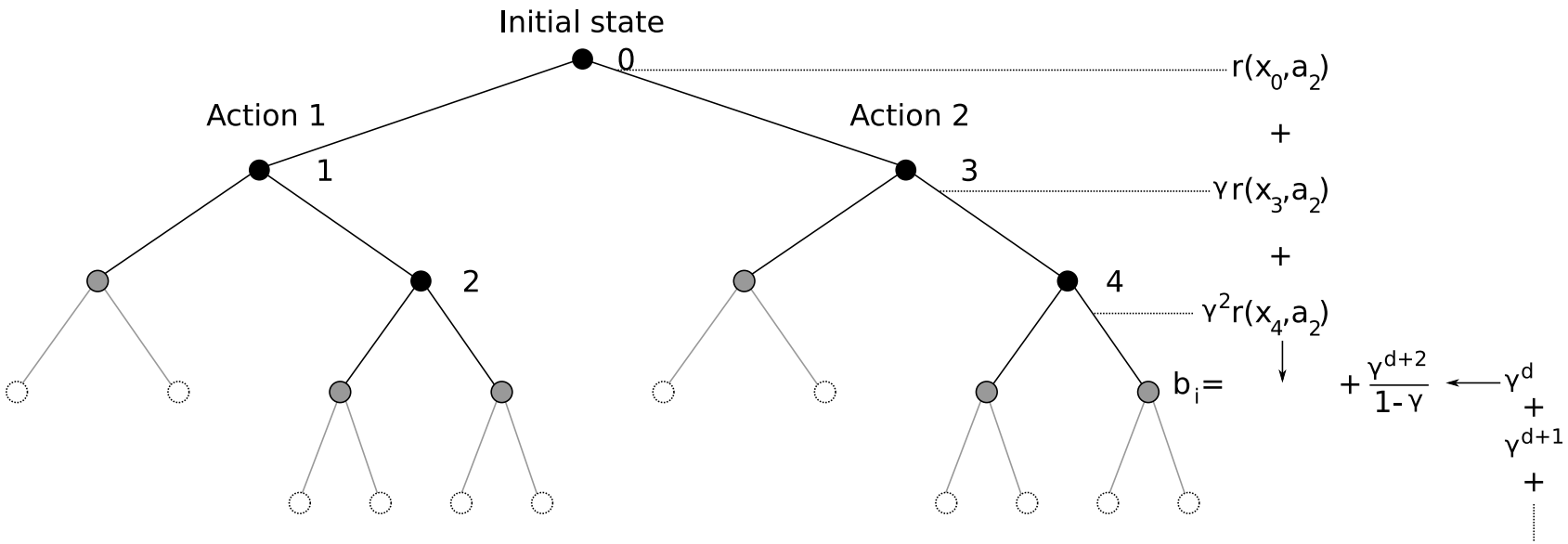
$$R_{\mathcal{A}_U}(n) = \Omega\left(n^{-\frac{\log(1/\gamma)}{\log K}}\right)$$



- Select the node i with the highest bound

$$b_i \stackrel{\text{def}}{=} \sum_{t=0}^{d-1} \gamma^d r_t + \frac{\gamma^d}{1-\gamma} \text{ with } d \text{ the depth of the node } i$$

- Select the action a that maximise $Q(x_0, a)$

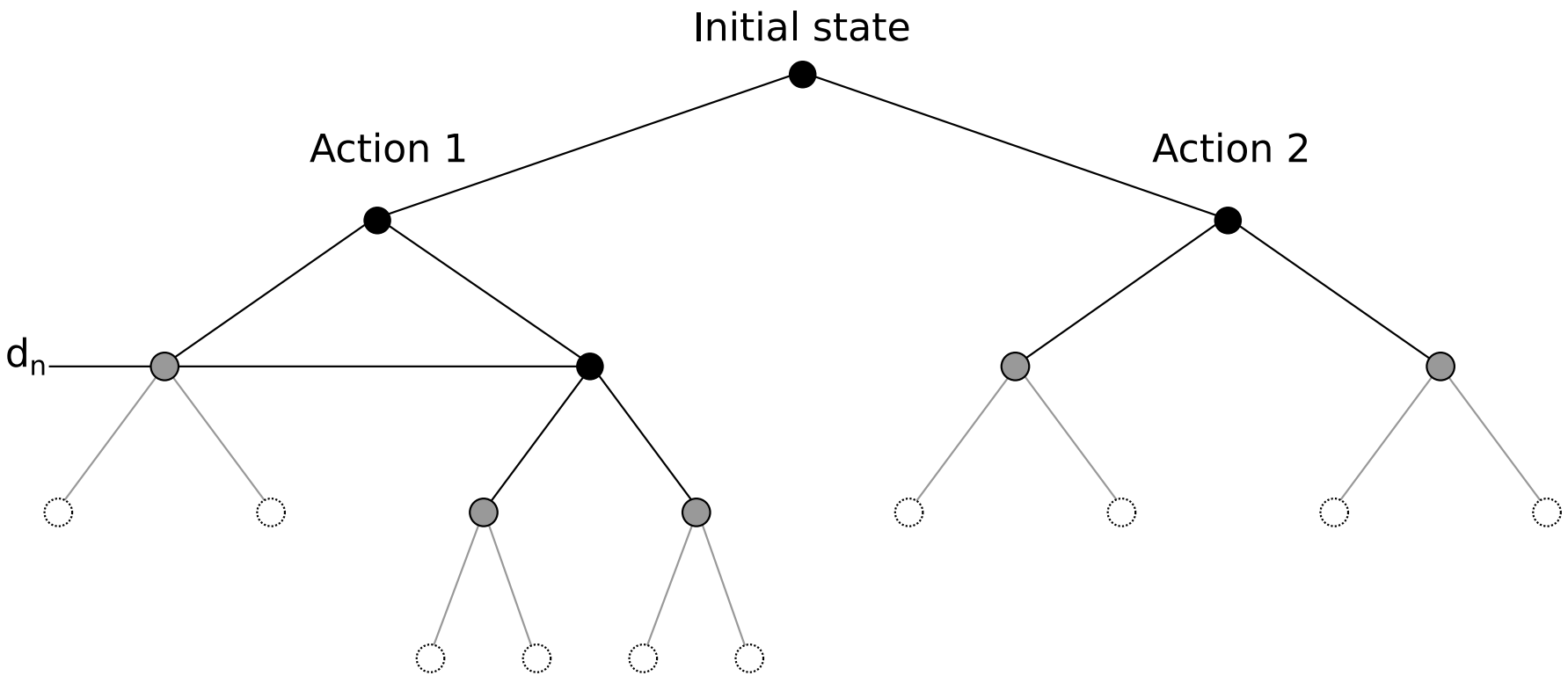


Optimistic planning

Upper bound on the regret

$$R_{\mathcal{A}_O}(n) \leq \frac{\gamma^{d_n}}{1-\gamma}$$

with d_n the maximal depth of the expanded tree



Definitions

Let $p(\epsilon)$, the proportion of ϵ-optimal paths in the tree
Let β , a coefficient such that $p(\epsilon) = O(\epsilon^\beta)$
Let $\kappa \stackrel{\text{def}}{=} K\gamma^\beta \in [1, K]$

- κ close to 1, few near-optimal paths
- κ close to K, lot of near-optimal paths

Upper bound on the regret

$$R_{\mathcal{A}_O}(n) = O\left(n^{-\frac{\log(1/\gamma)}{\log \kappa}}\right)$$

κ is the branching factor of the subtree of nodes that need to be expended

Demo

- cart-pole
- double cart-pole with a spring
- cart-pole with obstacles