Institute of Science and Technology Austria

HARVARD UNIVERSITY

Lower Bound on Howard's Policy Iteration for Deterministic Markov Decision Processes

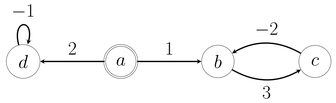
Ali Asadi, Krishnendu Chatterjee, Jakob de Raaij

Abstract

We study Howard's Policy Iteration for Deterministic Markov Decision Processes. The best known upper bound is exponential and the current known lower bound is as follows: For the input size I, the algorithm requires $\widetilde{\Omega}(\sqrt{I})$ iterations, i.e., the current lower bound on iterations is sub-linear with respect to the input size. Our main result is an improved lower bound for this fundamental algorithm where we show that for the input size I, the algorithm requires $\widetilde{\Omega}(I)$ iterations.

Deterministic Markov Decision Processes

- A finite directed weighted graph $G = (V, E, w)$;
- Weight function $w\colon E \to \mathbb{Z}$ assigns a weight to each edge;
- n is the number of vertices, m is the number of edges, and W is the maximum absolute weight.



Mean-payoff Objectives

- Mean-payoff for an infinite path $\omega = \langle v_0, v_1, \ldots \rangle$

$$\text{MeanPayoff}(\omega) := \liminf_{T\to\infty} \frac{1}{T} \sum_{i=0}^{T-1} w(v_i, v_{i+1})$$

The controller wants to maximize the payoff.

Howard's Policy Iteration

The algorithm starts with an arbitrary policy σ_0. In iteration k, the algorithm locally improves the current policy:

- The algorithm computes the payoff and the potential of the policy σ_k;
- it then obtains the policy σ_{k+1} by locally maximizing first the payoff and second the potential.

The algorithm terminates if $\sigma_{k+1} = \sigma_k$.

Problem

What is the lower bound on the number of iterations required by Howard's Policy Iteration algorithm in deterministic Markov Decision Processes?

Motivation

- Although Howard's policy iteration runs fast in practice [1], the theoretical guarantees are a mystery.
- Lower and upper bounds for Howard's policy iteration have deep theoretical impacts, e.g., establishing lower bounds for pivoting methods in linear programming [2].

Our Result

Theorem. *There exists a family of graphs with $\mathcal{O}(n)$ vertices, $\mathcal{O}(n^2)$ edges, and weights of size up to $W = \mathcal{O}(n^2)$ on which Howard's Policy Iteration takes $\Omega(n^2)$ iterations.*

Comparison to Previous Results

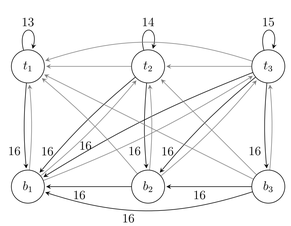
Upper bound on number of iterations: $\mathcal{O}(n^3W)$ [4, 6] & exponential non-parametric bound [5]

Lower bound on number of iterations:

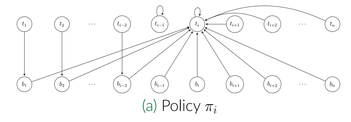
	$\lvert V\rvert$	$\lvert E\rvert$	W	Input Size I	# Iterations	Upper Bound
Prev. best known [3]	$2n$	m	$\mathcal{O}(n^{n^2})$	$\mathcal{O}(mn^2 \log n)$	$m - n + 1$	exponential
Ours	$2n$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2 \log n)$	$\Omega(n^2)$	$\mathcal{O}(n^5)$

Our example

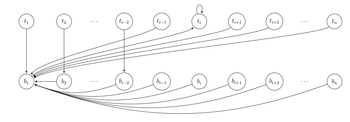
In our example, the top vertex t_i is connected to all vertices with indices less than or equal to i. The bottom vertex b_i is connected to all top vertices and to all bottom vertices with indices less than i. The weights of the self-loops increase with i, unlabeled (gray) edges have weight 0.



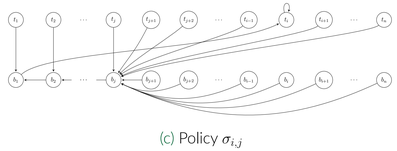
Sequence of policies



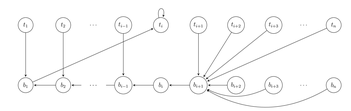
(a) Policy π_i



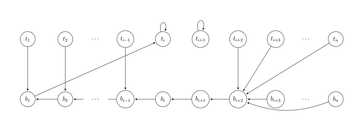
(b) Policy $\sigma_{i,1}$.



(c) Policy $\sigma_{i,j}$



(d) Policy $\sigma_{i,i+1}$



(e) Policy τ_i

Open Problems

- It is an open conjecture by [3] that m is an upper bound on the number of iterations.
- Howard's Policy Iteration algorithm performs outstandingly well in practice even for (stochastic) MDPs or stochastic games, in spite of the theoretical exponential lower bounds. Some analysis of the smoothed complexity has been published, but a comprehensive explanation is still missing.

References

[1] Ali Dasdan.
Experimental analysis of the fastest optimum cycle ratio and mean algorithms.
ACM Transactions on Design Automation of Electronic Systems (TODAES), 9(4):385–418, 2004.

[2] Oliver Friedmann, Thomas Dueholm Hansen, and Uri Zwick.
Subexponential lower bounds for randomized pivoting rules for the simplex algorithm.
In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 283–292, 2011.

[3] Thomas Dueholm Hansen and Uri Zwick.
Lower bounds for howard's algorithm for finding minimum mean-cost cycles.
In *International Symposium on Algorithms and Computation*, pages 415–426. Springer, 2010.

[4] Ronald A Howard.
Dynamic programming and markov processes.
MIT Press google schola, 2:39–47, 1960.

[5] Martin L. Puterman.
Markov Decision Processes.
John Wiley and Sons, 1994.

[6] Uri Zwick and Mike Paterson.
The complexity of mean payoff games on graphs.
Theoretical Computer Science, 158(1-2):343–359, 1996.