A polynomial algorithm for deterministic mean-payoff games

Bruno Scherrer*

March 15, 2022

Abstract

We describe a polynomial algorithm for solving deterministic mean payoff games. Our algorithm solves a mean payoff game with n vertices and integer edge-costs between $-W$ and W in time ... This in particular implies that a parity game with n vertices and d priorities can be solved in time .... This answers positively the long-standing open problem whether these problems are in P.

Consider a mean payoff game played by two players, MAX and MIN, on a graph with n vertices $X = \{1, 2, \dots, n\} = X_+ \sqcup X_-$ and directed edges E. For any vertex x, we write $f(x) = \{y; (x, y) \in E\}$ the set of vertices that can be reached from x by following one edge. An integer cost $-R \leq r(x) \leq R$ is associated to each node x. The vertices of X_+ (resp. X_-) belong to MAX (resp. MIN). The game starts in some vertex x_0. The player who owns the current vertex chooses a next vertex by following an edge. So on and so forth, these choices generate an infinitely long trajectory $(x_0, x_1, \dots)$. The goal of MAX is to maximize

$$\lim\inf_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} r(x_t),$$

while that of MIN is to minimize

$$\lim\sup_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} r(x_t).$$

A practical way to work on mean payoff games is to consider their γ-discounted variant for some $0 \leq \gamma < 1$, that only differs by the objective. In such a game, MAX tries to maximize the value

$$\sum_{t=0}^{\infty} \gamma^t r(x_t)$$

while MIN tries to minimize it. For any pair of positional strategies $\mu : X_+ \to X$ for MAX and $\nu : X_-$ for MIN (mappings such that for all x, $\mu(x) \in f(x)$ and $\nu(x) \in f(x)$), let us write $P_{\mu,\nu}$ for the transition matrix: for all $(i, j) \in \{1, 2, \dots, n\}^2$, $P_{\mu,\nu}(i, j)$ equals 1 if and only if (μ, ν) induce a transition $i \to j$ and 0 else. Let us introduce the following Bellman operators:

$$\begin{aligned} T_{\mu,\nu} v &= g + \gamma P_{\mu,\nu} v, \\ T_\mu v &= \min_\nu T_{\mu,\nu} v, \\ T v &= \max_\mu T_\mu v, \end{aligned}$$

that are all γ-contraction with respect to the max norm. The discounted value $v_{\mu,\nu}$ when MAX and MIN respectively use μ and ν satisfies

$$v_{\mu,\nu} = \sum_{t=0}^{\infty} (\gamma P_{\mu,\nu})^t r = (I - \gamma P_{\mu,\nu})^{-1} r$$

*INRIA, bruno.scherrer@inria.fr

and is the fixed point of $T_{\mu,\nu}$. The optimal value v_* is the fixed point of T, and any pair of positional strategies μ_*, ν_* that satisfy $T_{\mu_*,\nu_*} v_* = T v_* = v_*$ are known to be optimal [**?**].

In Section **??**, we shall describe and analyse an algorithm for computing optimal cycles in γ-discounted payoff. In Section **??**, we shall describe a scaling approach that allows to reduce the complexity dependency of the first algorithm with respect to W from W to $\log W$. Finally, in Section **??**, we will explain that this approach allows to solve the Mean Payoff game in polynomial time if one chooses γ sufficiently close to 1.

1 An algorithm to compute optimal cycles

We consider the following iterative algorithm:

1. (Initialization) Set $k = 0$ and take an arbitrary policy μ_0 for MAX.

2. (Evaluation) Compute the value v_k and an optimal counter-policy ν_k of MIN in the 1-player problem where MAX uses μ_k:

$$v_k = T v_k = T_{\mu_k} v_k = T_{\mu_k,\nu_k} v_k$$

3. (Computation of the n-step advantage) Compute the advantage δ_n and a pair of n-horizon strategies $(\mu_k^{(n)}, \ldots, \mu_k^{(1)})$ and $(\nu_k^{(n)}, \ldots, \nu_k^{(1)})$ such that:

$$\delta_k = T^n v_k - v_k = T_{\mu_k^{(n)},\nu_k^{(n)}} T_{\mu_k^{(n-1)},\nu_k^{(n-1)}} \ldots T_{\mu_k^{(1)},\nu_k^{(1)}} v_k - v_k.$$

4. (Identification of converged nodes) Compute the set

$$Z_k = \{x \in X \ ; \ \delta_k(x) = 0\}.$$

   If $Z_k \neq \emptyset$: 1) remove the nodes of the MIN-attractor A_k set of Z_k from the game (along with the transitions that go to A_k). If the game still has nodes, increment k by 1 and go to step 2 (otherwise stop).

5. (Computation of a stationary policy) Compute the values $w_k^{(n)}, \ldots, w_k^{(1)}$ in the 1-player problems for MIN where MAX uses the n-periodic strategies $\sigma_k^{(n)} = (\mu_k^{(n)}, \ldots, \mu_k^{(1)})$, $\sigma_k^{(n-1)} = (\mu_k^{(n-1)}, \ldots, \mu_k^{(1)}, \mu_k^{(n)})$, $\ldots$, $\sigma_k^{(1)} = (\mu_k^{(1)}, \mu_k^{(n)}, \ldots, \mu_k^{(2)})$:

$$\begin{aligned}
w_k^{(n)} &= T_{\mu_k^{(n)}} \ldots T_{\mu_k^{(1)}} w_k^{(n)}, \\
w_k^{(n-1)} &= T_{\mu_k^{(n-1)}} \ldots T_{\mu_k^{(1)}} T_{\mu_k^{(n)}} w_k^{(n-1)}, \\
\vdots \quad & \qquad\qquad \vdots \\
w_k^{(1)} &= T_{\mu_k^{(1)}} T_{\mu_k^{(n)}} \ldots T_{\mu_k^{(2)}} w_k^{(1)}.
\end{aligned}$$

   Compute the pointwise maximum $w_k = \max_i w_k(i)$, and identify a policy μ_{k+1} that satisfies:

$$T_{\mu_k} w_k = T w_k$$

   Increment k by 1 and go to step 2.

2 A scaling approach

3 Application to the Mean Payoff game

4 Conclusion

We have shown that the problem "Mean Payoff Game" is in P. To our knowledge, this problem was only previously known to be in $NP \cap co-NP$ [**?**].

It was shown in [**?**] that any parity game (a game that is central to μ-calculus model checking) on a graph with n vertices and d priorities can be reduced to a mean payoff game on the same graph with edge costs bounded in absolute value by $W = n^d$. As a consequence, Theorem **??** implies that:

Theorem 1. *A parity game with n vertices and d priorities can be solved in time ...*

Though "Parity Game" was long thought to be in $NP \cap co-NP$ and has recently be shown to be quasi-polynomial [**?**], it is in fact in P.