Problem Set no. 2 – With solutions

Given: November 29, 2020

Credit: Ishay Golinsky

Exercise 2.1 (a) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the *value iteration operator* of a discounted TBSG on n states, i.e., $T(\mathbf{y})_i = \min_{a \in A_i} c_a + \lambda \sum_{j \in S} p_{a,j} y_j$, if $i \in S_0$, and $T(\mathbf{y})_i = \max_{a \in A_i} c_a + \lambda \sum_{j \in S} p_{a,j} y_j$, if $i \in S_1$, for $\mathbf{y} = (y_i) \in \mathbb{R}^n$, where $0 < \lambda < 1$ is the *discount factor*. Prove that T is a *contraction*, i.e., that $||T(\mathbf{x}) - T(\mathbf{y})||_\infty \leq \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(b) Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be the *value iteration operator* of a *stopping* TBSG on n states, i.e., $T(\mathbf{y})_i = \min_{a \in A_i} c_a + \sum_{j \in S} p_{a,j} y_j$, if $i \in S_0$, and $T(\mathbf{y})_i = \max_{a \in A_i} c_a + \sum_{j \in S} p_{a,j} y_j$, if $i \in S_1$, for $\mathbf{y} = (y_i) \in \mathbb{R}^n$. Prove that $T^{(n)}$, i.e., T iterated n times, is a *contraction*, i.e., there exists $0 < \lambda < 1$ such that $||T^{(n)}(\mathbf{x}) - T^{(n)}(\mathbf{y})||_\infty \leq \lambda ||\mathbf{x} - \mathbf{y}||_\infty$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Solution 2.1 (a) Denote $p_a = (p_{a,1}, \ldots, p_{a.n})^T$ and

$$a^{i,x} = \begin{cases} \arg\min_{a \in A_i} \{c_a + \lambda \langle p_a, x \rangle\} & , i \in S_0 \\ \arg\max_{a \in A_i} \{c_a + \lambda \langle p_a, x \rangle\} & , i \in S_1 \end{cases}.$$

Assume WLOG that $T(x)_i \geq T(y)_i$. If $i \in S_0$ we get

$$\begin{aligned} T(x)_i - T(y)_i &= (c_{a^{i,x}} + \lambda \langle p_{a^{i,x}}, x \rangle) - (c_{a^{i,y}} + \lambda \langle p_{a^{i,y}}, y \rangle) \\ &\leq (c_{a^{i,y}} + \lambda \langle p_{a^{i,y}}, x \rangle) - (c_{a^{i,y}} + \lambda \langle p_{a^{i,y}}, y \rangle) \\ &= \lambda \langle p_{a^{i,y}}, x - y \rangle \\ &\leq \lambda \|p_{a^{i,y}}\|_1 \cdot \|x - y\|_\infty \\ &\leq \lambda \|x - y\|_\infty . \end{aligned}$$

Similarly, if $i \in S_1$ we get

$$\begin{aligned} T(x)_i - T(y)_i &\leq (c_{a^{i,x}} + \lambda \langle p_{a^{i,x}}, x \rangle) - (c_{a^{i,x}} + \lambda \langle p_{a^{i,x}}, y \rangle) \\ &= \lambda \langle p_{a^{i,x}}, x - y \rangle \\ &\leq \lambda \|p_{a^{i,x}}\|_1 \cdot \|x - y\|_\infty \\ &\leq \lambda \|x - y\|_\infty . \end{aligned}$$

It follows that

$$\begin{aligned} \|T(x) - T(y)\|_\infty &= \max_i |T(x)_i - T(y)_i| \\ &\leq \lambda \|x - y\|_\infty , \end{aligned}$$

as required.

(b) Consider a game of n steps and a terminal cost vector x. For a strategy profile (σ, τ) and a starting state i, let $C_{\text{act}}^{\sigma,\tau,i}$ be the total cost of all actions taken, and $C_{\text{end}}^{\sigma,\tau,i}(x)$ be the termination cost incurred. By induction on the number of steps, one can show that

$$T^{(n)}(x)_i = \mathbb{E}\left[C_{\text{act}}^{\sigma_x,\tau_x,i} + C_{\text{end}}^{\sigma_x,\tau_x,i}(x)\right]$$

for optimal strategies σ_x, τ_x (that depend on x). For a profile σ, τ and a starting state i we denote $p^{\sigma,\tau,i} = \left(p_1^{\sigma,\tau,i}, \ldots, p_n^{\sigma,\tau,i}\right)^T$, where $p_j^{\sigma,\tau,i}$ is the probability of reaching state j after exactly n steps. Now, using the optimality of (σ_x, τ_x) and (σ_y, τ_y), we have

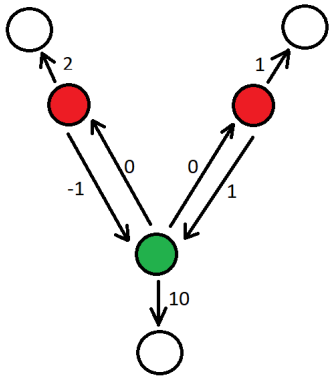
$$\begin{aligned}
T^{(n)}(x)_i - T^{(n)}(y)_i &= \mathbb{E}\left[C_{\text{act}}^{\sigma_x,\tau_x,i} + C_{\text{end}}^{\sigma_x,\tau_x,i}(x)\right] - \mathbb{E}\left[C_{\text{act}}^{\sigma_y,\tau_y,i} + C_{\text{end}}^{\sigma_y,\tau_y,i}(y)\right] \\
&\leq \mathbb{E}\left[C_{\text{act}}^{\sigma_y,\tau_x,i} + C_{\text{end}}^{\sigma_y,\tau_x,i}(x)\right] - \mathbb{E}\left[C_{\text{act}}^{\sigma_y,\tau_x,i} + C_{\text{end}}^{\sigma_y,\tau_x,i}(y)\right] \\
&= \mathbb{E}\left[C_{\text{end}}^{\sigma_y,\tau_x,i}(x) - C_{\text{end}}^{\sigma_y,\tau_x,i}(y)\right] \\
&= \left\langle p^{\sigma_y,\tau_x,i}, x - y \right\rangle \\
&\leq \left\| p^{\sigma_y,\tau_x,i} \right\|_1 \cdot \|x - y\|_\infty \\
&\leq (1 - \epsilon) \|x - y\|_\infty ,
\end{aligned}$$

where ϵ is as obtained in exercise 1.3. It follows that

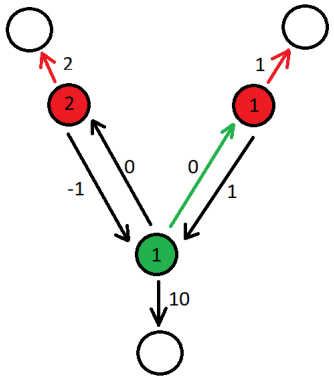
$$\|T(x) - T(y)\|_\infty \leq (1 - \epsilon) \|x - y\|_\infty .$$

Exercise 2.2 Construct a stopping TBSG on which there is a sequence of alternating improving switches by both players that cycles.

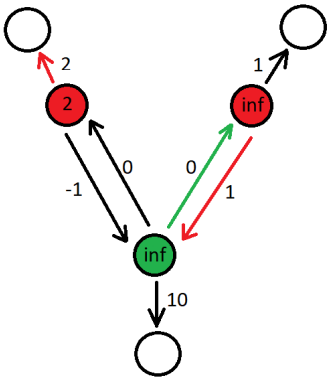
Solution 2.2 Consider the following game:



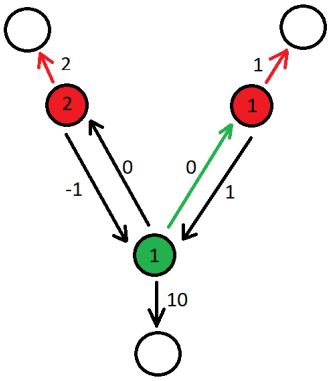
Sinks are represented by white nodes. Given a positional strategy profile, on each node we write the total cost incurred in a game that starts from that node, e.g.:

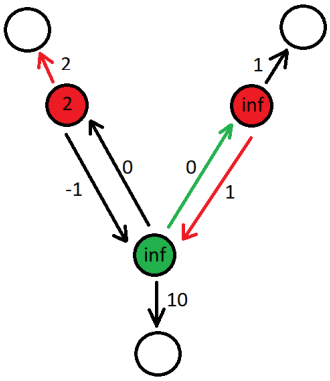


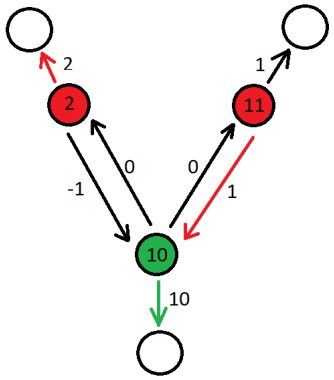
For the sake of simplicity we will show a sequence of improving switches on a non-stopping game. A stopping game can be obtained by adding a termination probability λ to all transitions (equivalently, adding a discount factor λ). When $\lambda \to 1^-$, the discounted cost tends to the non-discounted cost (which may be infinite. Note that under positional strategies, the cost of each cycle is either strictly positive or strictly negative, hence the total cost is well defined). Therefore we can set λ close enough to 1 such that all the improving switches are justified in the discounted game as well. Initial strategies:

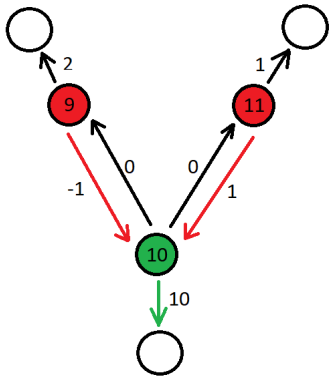


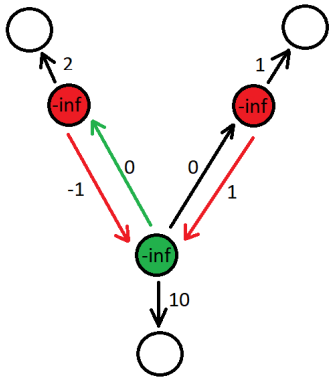
max updates first:



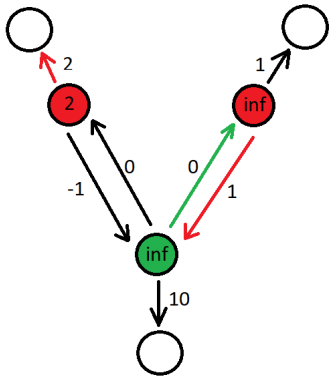








Back to initial strategies:



Exercise 2.3 Prove Lemma 3 in slide 28 of lecture 3: Let $\pi_0 = (\sigma_0, \tau_0)$, where τ_0 is a best response to σ_0. Let π_1 be a profile obtained by one iteration of Howard's algorithm. Then $\mathbf{y}^{\pi_1} \leq T(\mathbf{y}^{\pi_0})$.

Solution 2.3 Let $i \in S_0$, and let $a = \arg\min_{a \in A_i} c_a + \sum_j p_{a,j} y_j^{\pi_0}$. we get

$$y_i^{\pi_1} = c_a + \sum_j p_{a,j} y_j^{\pi_1} \leq c_a + \sum_j p_{a,j} y_j^{\pi_0} = T(y^{\pi_0})_i,$$

where the first inequality holds since $y^{\pi_1} \leq y^{\pi_0}$. If $i \in S_1$, we get

$$y_i^{\pi_1} = \max_{a \in A_i} c_a + \sum_j p_{a,j} y_j^{\pi_1} \leq \max_{a \in A_i} c_a + \sum_j p_{a,j} y_j^{\pi_0} = T(y^{\pi_0})_i.$$

where the first equality holds since max plays according to a best response to min.

Exercise 2.4 Give a full proof of the theorem on slide 34 of lecture 3: Let Γ be a TBSG. Then, there exists positional strategies σ^* and τ^* of the two players, and $\lambda^* = \lambda^*_\Gamma$, such that σ^* and τ^* are optimal for every discount factor $\lambda^* \leq \lambda < 1$.

Solution 2.4 Denote $\lambda_n = 1 - \frac{1}{n}$. Let $\pi_n = (\sigma_n, \tau_n)$ be an optimal positional strategy profile for the λ_n-discounted game. Since there are only finitely many positional profiles, there is a $\pi^* = (\sigma^*, \tau^*)$ such that $\pi_n = \pi^*$ infinitely many times. As stated in class, for any profile π, the function $\lambda \mapsto y_\lambda^\pi$ is rational. Let $0 < \lambda^* < 1$ be any number greater than all the (finitely many) poles and zeros in $(0, 1)$ of all the non-zero functions of the form $\lambda \mapsto y_\lambda^\pi - y_\lambda^{\pi^*}$. We will show that π^* is optimal for the λ-discounted game for any $\lambda \in [\lambda^*, 1)$, or equivalently, that for any positional σ, τ it holds that

$$y_\lambda^{\sigma^*,\tau} \leq y_\lambda^{\pi^*} \leq y_\lambda^{\sigma,\tau^*}.$$

By negation, assume WLOG that $y_\lambda^{\sigma^*,\tau} > y_\lambda^{\pi^*}$ for some $\lambda \in [\lambda^*, 1)$ and τ. From the choice of π^* there is some $\lambda' > \lambda$ for which π^* is optimal, and in particular, $y_{\lambda'}^{\sigma^*,\tau} \leq y_{\lambda'}^{\pi^*}$. Now, since $y_\lambda^{\sigma^*,\tau} - y_\lambda^{\pi^*}$ has no poles in $[\lambda, \lambda']$, it follows from the Intermediate Value Theorem that $y_\lambda^{\sigma^*,\tau} - y_\lambda^{\pi^*}$ has a zero in $[\lambda, \lambda']$, contradiction.

Exercise 2.5 Show that the value of a MPG played on a graph $G = (V, E, c)$, where $V = V_0 \cup V_1$ and $c : E \to \mathbb{R}$, that starts at $s \in V$ is at *most* μ if and only if there exists a subset $U \subseteq V$ such that $s \in U$ and a potential function $h : U \to \mathbb{R}$ such that (1) If $u \in U \cap V_0$, then there exists an edge $(u, v) \in E$ such that $v \in U$ and $h(u) + c(u, v) \leq h(v) + \mu$. (2) If $u \in U \cap V_1$, then for every edge $(u, v) \in E$ we have $v \in U$ and $h(u) + c(u, v) \leq h(v) + \mu$.

Solution 2.5 $\Longleftarrow$: Assume that such U and h exist. Let σ be the following positional strategy for player 0: whenever in $u \in U \cap V_0$, move to one of the v's assumed to exists in (1), otherwise move arbitrarily. We will show that for every strategy τ of player 1, the resulting limiting average is at most μ. Set some τ and let $\text{play}^{\sigma,\tau}(s) = (v_1, v_2, \ldots)$ be the path taken throughout the game. Note that the path lies within U, thus

$$\frac{1}{k} \sum_{i=1}^{k} c\left(v_i, v_{i+1}\right) \leq \frac{1}{k} \sum_{i=1}^{k} \left(h\left(v_{i+1}\right) - h\left(v_i\right) + \mu\right) = \mu + \frac{h\left(v_{k+1}\right) - h\left(v_1\right)}{k} \overset{k \to \infty}{\longrightarrow} \mu,$$

and therefore

$$\limsup_{k \to \infty} \frac{1}{k} \sum_{i=1}^{k} c\left(v_i, v_{i+1}\right) \leq \mu.$$

$\Longrightarrow$: Assume that the value of a game starting from s is at most μ. Let σ^* be an optimal positional strategy for player 0. As we saw in class this strategy forces all cycles reachable from s to be of mean weight at most μ.
Consider the sub-graph $G' = (U, E')$ that contains all possible trajectories of the game when player 0 plays σ^* (U are all vertices reachabel from s when player-0 plays σ^*). Consider the modified cost $c' : E' \longrightarrow \mathbb{R}$ defined by

$$c'(u, v) = c(u, v) - \mu.$$

For a vertex $u \in U$, let $h(u)$ be the maximum weight of a path (nder c') from s to u (this is well defined since there are no positive cycles in G'). Let $u, v \in U$ such that $(u, v) \in E'$, then $h(v) \geq h(u) + c'(u, v)$ is exactly the triangle inequality, as required of h.

Exercise 2.6 Let $G = (V, E, c)$, where $V = V_0 \cup V_1$ and $c : E \to [-W, W]$ be an MPG. Let $Val(u)$ be the value of the MPG that starts at u. Let $Val_k(u)$ be the value of the game that starts at u and goes on for exactly k steps. (The game always goes on for k steps, even if a cycle is formed.) Show that for every $u \in V$, we have $Val(u) - \frac{2nW}{k} \leq \frac{Val_k(u)}{k} \leq Val(u) + \frac{2nW}{k}$. (Hint: show that using an optimal positional strategy for player 0 in the infinite game insures that $Val_k(u) \leq (k - n)\,Val(u) + nW$, where $n = |V|$. Recall that such an optimal positional strategy for player 0 ensures that every cycle formed has an average cost of at most $Val(u)$.)

Solution 2.6 Let σ^* be an optimal positional strategy of player 0 for the infinite game and let τ be any strategy of player 1. By traversing the trajectory of the play and removing cycles whenever formed, we can decompose $\text{play}_k^{\sigma^*,\tau}(u)$ to a sum of cycles and a path of length $\leq n - 1$:

$$\text{play}_k^{\sigma^*,\tau}(u) = P + C_1 + \cdots C_r.$$

(To be precise, by "sum" we mean that the edges included in the walks $P, C_1, \ldots, C_r$ are exactly those of $\text{play}_k^{\sigma^*,\tau}(u)$, considering multiplicities.) Since every cycle has an average cost of at most $\text{Val}(u)$, we have

$$\text{cost}_{Tot}\left(\text{play}_k^{\sigma^*,\tau}(u)\right) = \text{cost}_{Tot}(P) + \sum_{i=1}^{r} \text{cost}_{Tot}(C_i) \leq |P|\,W + \sum_{i=1}^{r} |C_i|\,\text{Val}(u) =$$
$$= |P|\,W + (k - |P|)\,\text{Val}(u) \leq nW + |P| \cdot |\text{Val}(u)| + k \cdot \text{Val}(u) \leq 2nW + k \cdot \text{Val}(u).$$

Since this holds for every τ, it follows that

$$\frac{\text{Val}_k(u)}{k} \leq \text{Val}(u) + \frac{2nW}{k}.$$

A similar argument, using an optimal positional τ^* for player 1 instead of σ^*, shows that

$$\text{Val}(u) - \frac{2nW}{k} \leq \frac{\text{Val}_k(u)}{k}.$$

Exercise 2.7 Use the result of the previous exercise to show that value iteration can be used to obtain an $O(mn^3W)$ to obtain the values of all vertices in a MPG $G = (V, E, c)$ where $c : E \to \{-W, \ldots, 0, \ldots, 1\}$. (I.e., all edge costs are integers of absolute value at most W.) (Hint: the value of each vertex is a rational number with denominator at most n.)

Solution 2.7 We use the following claim which we shall prove later.

Claim: Let r be a rational number with denominator at most n and p be any real number such that $|r-p| < \frac{1}{2n^2}$. Then r can be recovered from p in $O(n)$ time.

Since $\mathrm{Val}(u)$ is the average cost of some cycle in (V,E), and since costs are integers, $\mathrm{Val}(u)$ is a rational number with denominator at most n. Using the claim above, it is enough to approximate $\mathrm{Val}(u)$ up to $\frac{1}{2n^2}$. In exercise 2.6 we showed the following approximation by a k-steps game

$$\left|\mathrm{Val}(u) - \frac{\mathrm{Val}_k(u)}{k}\right| \leq \frac{2nW}{k}.$$

Setting $k = 4n^3W + 1$, we get

$$\left|\mathrm{Val}(u) - \frac{\mathrm{Val}_k(u)}{k}\right| < \frac{1}{2n^2},$$

as needed. Computing $\mathrm{Val}_k(u)$ for all vertices u amounts to computing

$$T(0), T^{(2)}(0), \ldots, T^{(k)}(0),$$

where each is done in a straight-forward way in $O(m)$ time. The total time required is thus

$$O(mk) = O\left(mn^3W\right).$$

Proof of claim: If $\frac{x}{y}, \frac{z}{w}$ are rationals with $|y|, |w| \leq n$ (here $x, y, z, w \in \mathbb{Z}$) such that

$$\left|\frac{x}{y} - \frac{z}{w}\right| < \frac{1}{n^2},$$

then

$$|xw - zy| < \frac{|yw|}{n^2} \leq 1,$$

and therefore $xw = zy$, meaning $\frac{x}{y} = \frac{z}{w}$. It follows that there exist at most one such rational number within the segment $\left(p - \frac{1}{2n^2}, p + \frac{1}{2n^2}\right)$. To find that number we loop through all possible denominators $y = 1, \ldots, n$ and check for the existence of an integer x such that

$$x \in \left(yp - \frac{y}{2n^2}, yp + \frac{y}{2n^2}\right),$$

which can be done in $O(1)$. Once such x, y have been found, we conclude $r = \frac{x}{y}$.