Infinite Horizon problems

The infinite value function of policy $u$ for the infinite horizon problem with discount factor $\alpha$ and initial state $i$ is

$$W(i, u) = E[\sum_{n=0}^{\infty} \alpha^n r(X_n, u(X_n)), X_0 = i].$$

The optimal value function is $W(i) = \max_u W(i, u)$. $u^*$ is an optimal policy iff each initial state $i$, $W(i, u^*) = W(i)$. The recursion equation is

$$W(i, u) = r(i, u(i)) + \alpha \sum_j T_{ij} W(j).$$

**Optimal stopping times**

**Rules for initial known optimal values.**
- If $i$ is a dead-end state (you can’t leave), $W(i) = r(i)$
- If $r(i)$ is $\geq$ all rewards reachable from $i$, $W(i) = r(i)$
- If $i$ is in an irreducible class, $W(i) = r(j)$ where $r(j)$ is the max reward in the irreducible class.

If $W(i, u)$ satisfies this dynamic programming equation, then $W(i, u) = W(i)$ and $u$ is an optimal policy.

The following process gives the successive approximations $w_n(i)$. Let $w_0 = [0, 0, 0, ...]^T$. Define $w_{n+1}$ by applying the dynamic equation for $W(i)$ to $w_n(i)$:

$$w_{n+1}(i) = \max_a [r(i, a) + \alpha \sum_j T_{ij} w_n(j)].$$

**Lemma.** Given $u$, with transition matrix $T_{ij}(u)$, let $r_u = [r(0, u(0)), r(1, u(1)), r(2, u(2)), ...]^T$. Let $X = [x_0, x_1, x_2, ...]^T = [W(0, u), W(1, u), W(2, u), ...]^T$. Then $X$ is the solution to the matrix equation $(I - \alpha T_u)X = r_u$.

**Finite horizon portfolios.** The value function for the finite horizon problem with final time $T$ and policy $u$ and initial state $X_{init} = (x, y, l)$ is:

$$V(x, y, l, u) = E[R(x_T, l_T) + \sum_{n=0}^{T-1} r(X_n, U_n^a, U_n^b) | X_0 = X_{init}].$$

If $U_{n-1}^a = a$ and $U_{n-1}^b = b$, then the recursion equation is

$$V_{n-1}(x, y, l, u) = ax + by + \frac{1}{T} \sum_{i=1}^{T} r_i(x_i, x_i^* V_{n}(x_i^*, k-a, 1+\beta y, l-b, u).$$

The optimal value $V(x, y, l) = \max_u V(x, y, l, u)$ has the usual dynamic programming equation:

$$V_{n-1}(x, y, l, u) = \max_{a,b} \{ ax + by + \frac{1}{T} \sum_{i=1}^{T} r_i(x_i, x_i^* V_{n}(x_i^*, k-a, 1+\beta y, l-b, u).$$

Two person zero-sum games

**Lemma.** The amount $I$ can expect to win by consistently playing the strategy for a given row is the minimum payoff for that row.

**Maximin Theorem.** The strategy for $I$ with the maximum expected win is the strategy whose row has the maximum minimum. This is the maximin strategy for $I$.

**Definition.** A game is strictly determined if the entry in the maximin row and the minimax column

= the payoff for $I$’s maximin strategy

= the payoff for $II$’s minimax strategy.

This entry is the saddle or equilibrium point of the game.

**Theorem.** If there is an equilibrium point, then both players can optimize their earnings by always playing their simple (unmixed) maximin and minimax strategies.