MATH 3826 Assignment 3

In a Markov decision problem, another criterion often used, different than the expected average return per unit time, is that of the expected discounted return. In this criterion we choose a number $\alpha, 0 < \alpha < 1$, and try to choose a policy so as to maximize $\mathrm{E}[\sum_{i=0}^{\infty} \alpha^i R(X_i, a_i)]$ (that is, rewards at time n are discounted at rate α_n). Suppose that the initial state is chosen according to the probability b_l. That is

$$P\{X_0 = i\} = b_i, i = 1, \ldots, n$$

For a given policy $\boldsymbol{\beta}$, let y_{ja} denote the expected discounted time that the process is in state j and action a is chosen. That is

$$y_{ja} = \mathrm{E}_{\boldsymbol{\beta}} \left[\sum_{n=0}^{\infty} \alpha^n I\{X_n = j, a_n = a\} \right]$$

where for any event A the indicator variable I_A is defined by

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$$

(a) Show that

$$\sum_a y_{ja} = \mathrm{E} \left[\sum_{n=0}^{\infty} \alpha^n I\{X_n = j\} \right]$$

or, in other words, $\sum_a y_{ja}$ is the expected discounted time in state j under $\boldsymbol{\beta}$.

(b) Show that

$$\sum_j \sum_a y_{ia} = \frac{1}{1-\alpha}$$

$$\sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a)$$

Hint: For the second equation, use the identity

$$I\{X_{n+1} = j\} = \sum_i \sum_a I\{X_n = i, a_n = a\} I\{X_{n+1} = j\}$$

Take expectations of the preceding to obtain

$$\mathrm{E}[I\{X_{n+1} = j\}] = \sum_i \sum_a \mathrm{E}[I\{X_n = i, a_n = a\}] P_{ij}(a)$$

(c) Let $\{y_{ja}\}$ be a set of numbers satisfying

$$\sum_j \sum_a y_{ja} = \frac{1}{1-\alpha},$$

$$\sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a)$$

Argue that y_{ja} can be interpreted as the expected discounted time that the process is in state j and action a is chosen when the initial state is chosen according to the probabilities b_j and the policy $\boldsymbol{\beta}$, given by

$$\beta_i(a) = \frac{y_{ia}}{\sum_a y_{ia}}$$

is employeed.

Hint: Derive a set of equations for the expected discounted times when policy $\boldsymbol{\beta}$ is used and show that they are equivalent to the equation above.

(d) Argue that an optimal policy with respect to the expected discounted return criterion can be obtained by first solving the linear program

$$\text{maximize } \sum_j \sum_a y_{ja} R(j, a),$$

$$\text{such that } \sum_j \sum_a y_{ja} = \frac{1}{1-\alpha},$$

$$\sum_a y_{ja} = b_j + \alpha \sum_i \sum_a y_{ia} P_{ij}(a),$$

$$y_{ja} \geq 0, \text{ all } j, a;$$

and then defining the policy $\boldsymbol{\beta}^*$ by

$$\boldsymbol{\beta}_i^*(a) = \frac{y_{ia}^*}{\sum_a y_{ia}^*}$$

where the y_{ia}^* are the solutions of the linear program.