2.997 Decision-Making in Large-Scale Systems
MIT, Spring 2004 Handout #5

Solution for PS 1

1. (a) Let w_t be the wealth at time t and also the state of this problem, $t = 0, 1, \ldots, T$. Let a_t denote the decision made at time t, i.e., the portion of the wealth in the risky asset. Let $s = 1.03$ and r_t denote a stationary random variable with the mass function $P(r_t = r_u) = P(r_t = r_d) = 0.5$, which corresponds to the fluctuation of the stock price. Then the wealth evolves as follows.

$$\begin{aligned} w_{t+1} &= s(1-a_t)w_t + (1+r_t)a_t w_t \\ &= [s + (1+r_t-s)a_t]w_t \end{aligned}$$

Let $U(\cdot)$ be the utility function of the investor. Then we have

$$\begin{aligned} J_T(w_T) &= U(w_T) \\ J_t(w_t) &= \max_{a_t \in [0,1]} \mathrm{E}\left\{ J_{t+1}((s + (1+r_t-s)a_t)w_t) | x(0) = x_0 \right\}, \ \forall t = 1, 2, \ldots, T-1. \end{aligned}$$

We first discuss the case of $U(x) = x$. Then

$$\begin{aligned} a_{T-1} &= \arg\max_{a \in [0,1]} \left[\frac{1}{2}(s + (1+r_u-s)a)w_{T-1} + \frac{1}{2}(s + (1-r_u-s)a)w_{T-1} \right] \\ &= \arg\max_{a\in[0,1]} \left[s + (1 - s + \frac{r_u - r_d}{2})a \right] \\ &= \begin{cases} 0, & \text{if } 1 - s + \frac{r_u - r_d}{2} \leq 0 \Rightarrow r_u - r_d \leq 0.06 \\ 1, & \text{otherwise.} \end{cases} \end{aligned}$$

Notice that the greedy policy at the $(T-1)$ stage is independent of the state and the following backward computations are in the same structure, we know that the optimal policy for maximizing $\mathrm{E}(w_T)$ is

$$a^* = \begin{cases} 0, & \text{if } r_u - r_d \leq 0.06, \\ 1, & \text{otherwise.} \end{cases}$$

The optimal policy implies that the upward trend of the stock price must be strong enough, i.e., $r_u - r_d > 0.06$, for the investor to buy stocks.

Next, we consider the case of $U(x) = \log x$. Similarly,we have

$$\begin{aligned} a_{T-1} &= \arg\max_{a \in [0,1]} \left\{ \frac{1}{2} \log\left[(s + (1+r_u-s)a)w_{T-1}\right] + \frac{1}{2}\log\left[(s + (1-r_d-s)a)w_{T-1}\right] \right\} \\ &= \arg\max_{a\in[0,1]} \left\{ \log w_{T-1} + \frac{1}{2}\log\left[(s + (1+r_u-s)a)(s+(1-r_d-s)a)\right] \right\} \end{aligned}$$

Since $\log(\cdot)$ is a strictly concave function and $[0,1]$ is a compact set, we know that there exists a unique solution for the above optimization problem. By differentiating the above function and setting to zero, we obtain the greedy policy at stage $T-1$

$$\text{if } r_u \leq 0.03, \text{ then } a_{T-1} = 0,$$

$$\text{if } r_u > 0.03, \text{ then } a_{T-1} = \begin{cases} 0, & \text{if } r_u - r_d < 0.06, \\ 1, & \text{if } r_u(0.94 - 2r_d) > 0.94r_d - 0.00018 \\ \frac{1.03(r_u - r_d - 0.06)}{2(r_u - 0.03)(r_d + 0.03)}, & \text{otherwise.} \end{cases}$$

One can also observe that the greedy policy is independent of the state and the structure of the iteration functions are identical; therefore the optimal policy is the following stationary policy.

$$\text{if } r_u \leq 0.03, \text{ then } a^* = 0,$$

$$\text{if } r_u > 0.03, \text{ then } a^* = \begin{cases} 0, & \text{if } r_u - r_d < 0.06, \\ 1, & \text{if } r_u(0.94 - 2r_d) > 0.94r_d - 0.00018 \\ \frac{1.03(r_u - r_d - 0.06)}{2(r_u - 0.03)(r_d + 0.03)}, & \text{otherwise.} \end{cases}$$

Comparing the result for the cases $U(x) = x$ and $U(x) = log(x)$, one can conclude that the concavity of the utility function $U(\cdot)$ implies that the investor is more conservative.

(b) Since the transaction cost is proportional to the amount of the buying and selling the stocks, we need to track the amount of the wealth in the saving account and in the stock for each time period. We use (x_t, y_t) to denote the wealth in the saving account and in the stock at time t, respectively. Moreover, let $\theta = 0.05$ denote the proportional rate of transaction cost. Then, we have the transition equation

$$\begin{aligned} x_{t+1} &= s(x_t - a_t - \theta|a_t|) \\ y_{t+1} &= (1 + r)(y_t + a_t), \forall t = 0, 1, \ldots, T - 1, \end{aligned}$$

where a_t is the decision made at time t and, to make this problem sensible, we have $a_t \in A_t = [-y_t, \frac{x_t}{1+c}]$ so that the investor will not have negative wealth. Note that $a_t > 0$ can be interpreted as buying stocks, $a_t < 0$ can be interpreted as selling stocks, and $a_t = 0$ denotes no trading. Then one can write the Bellman's equation as follows.

$$\begin{aligned} J_T(x_T, y_T) &= U(x_T, y_T) \; (= (x_T + y_T) \text{ or } = \log(x_T + y_T)) \\ J_t(x_t, y_t) &= \max_{a_t \in A_t} \mathrm{E}_r \{J_{t+1}(s(x_t - a_t - \theta|a_t|), (1 + r)(y_t + a_t))\}, \forall t = T - 1, T - 2, \ldots, 0. \end{aligned}$$

(c) When $r_u = 1.2$ and $r_d = 0.9$, we have $\mathrm{E}[w_T] = 16.2845$ and $\mathrm{E}[\log w_T] = 0.827$. Notice that, in the latter case, we have the ratio of the wealth in the stock to the wealth in the saving account converge to a constant, i.e., 0.1568, fast. For the case $r_u = 1.4$ and $r_d = 0.7$, we have $\mathrm{E}[w_T] = 402.242$ and $\mathrm{E}[\log w_T] = 2.5629$. Again, when $U(x) = \log x$, the ratio of the wealth in the stock to the wealth in the saving account converges to a constant, i.e., 0.7521, fast. This is a property of the optimal policy for the logarithm utility.

As we discuss before, the larger the value $(r_u - r_d)$ is, the more the investor should invest in the stock market, since the mean rate of return of the risky asset becomes higher. This property is shown in our numerical results. Moreover, our conjecture in (a) that the concavity of the utility function makes the investor more conservative is proved in our numerical result as well. In both scenario, the investor invests less wealth in the stock than in the saving account, and the profit adopting the concave utility function is less than that from the linear utility function.

2. (a) Let $[p, q]$ denote the state that the multiplication of matrix $A_{p+1} \cdots A_q$ is taken place for $1 \leq p \leq q \leq n$. Let $J([p, q])$ denote the minimum cost of computing the matrix multiplication. We first consider the backward induction at state $[0, n]$. Considering the matrix multiplication $(A_1 \cdots A_n)$ is a

consequence of multiplication of two matrices $(A_1 \cdots A_k)$ and $(A_{k+1} \cdots A_n)$ for some integer $1 \leq k < n$, we have

$$J([0,n]) = min_{1 \leq k < n} \left\{ J([0,k]) + J([k,n]) + r_0 r_k r_n \right\}.$$

Then, the original problem of finding $J([0,n])$ is divided into two subproblems of finding $J([0,k])$ and $J([k,n])$, which have the same structure as $J([0,n])$. Hence, we can put it into the framework of MDP. Let $g_k([m,n]) = r_m r_k r_n$ denote the cost of computing the multiplication of two matrices $(A_{m+1} A_{m+2} \cdots A_k)$ and $(A_{k+1} \cdots A_n)$. Then we have

$$\begin{aligned} J^* &= J([0,n]), \quad \text{and for all} 0 \leq p \leq q \leq n, \\ J([p,q]) &= \begin{cases} 0, & \text{if } p+1 \geq q, \\ \min_{p<k\leq q} \left\{ \underbrace{g_k([p,q])}_{=r_p r_k r_q} + J([m,k]) + J([k,n]) \right\}, & \text{otherwise.} \end{cases} \end{aligned}$$

Since the number of possible combinations of p, q, k is $\Theta(n^3)$ and each combination is examined once, we have that this computation requires $\mathcal{O}(n^3)$ computations.

(b) The recursive computation of the value $J([0,4])$ can be presented as follows.

				[0,4]					
		↙ 3000		↓ 70000		↘ 2000			
	[0,1][1,4]			[0,2][2,4]			[0,3][3,4]		
↙ 14000	↓ 4200		↙ 21000		↘ 14000		↓ 600	↘ 1400	
[1,2][2,4]	[1,3][3,4]		[0,2]		[2,4]		[0,1][1,3]		[0,2][2,3]
							↓ 4200		↓ 21000
							[1,3]		[0,2]

The computation shows that the optimal order is $(A_2 A_3) \rightarrow (A_1 A_2 A_3) \rightarrow (A_1 A_2 A_3)$, which needs 6800 computations.

(c) The optimal order becomes $(A_1 A_2) \rightarrow (A_3 A_4) \rightarrow (A_1 A_2)(A_3 A_4)$, which requires 105000 computations. The ratio of the maximum to minimum number of multiplications is 15.44.

3. Let J_k denote the value of the cost-to-go function at iteration k. Denote by x_k the state whose value is updated at iteration k. We first show that

$$\begin{aligned} |J_{k+1}(x_k) - J^*(x_k)| &\leq \|J_k - J^*\|_\infty & (1) \\ \|J_{k+1} - J^*\|_\infty &\leq \|J_k - J^*\|_\infty. & (2) \end{aligned}$$

The inequality (1) follows immediately from $J_{k+1}(x_k) = (TJ_k)(x_k)$ and the contraction property of operator T. We also have, for all other x, $J_{k+1}(x) = J_k(x)$, and (2) follows.

Now let $1 = k_1 < k_2 < \ldots$ be iteration numbers such that each state has its value updated at least once during iterations $k_i, \ldots, k_{i+1} - 1$. Note that there are infinitely many such k_i since each state is updated infinitely many times. We will show that

$$\|J_{k_{i+1}} - J^*\|_\infty \leq \alpha \|J_{k_i} - J^*\|_\infty, \forall i = 1, 2, \ldots, \quad (3)$$

so that $\lim_{i\to\infty} J_{k_i} = J^*$. This combined with (2) implies that $\lim_{k\to\infty} J_k = J^*$.

Take an arbitrary i, and for each state x let k_x denote the last iteration in the set $\{k_i, \ldots, k_{i+1}-1\}$ when the value at state x is updated. Then we have

$$\begin{aligned} |J_{k_{i+1}}(x) - J^*(x)| &= |J_{k_x+1}(x) - J^*(x)| \\ &\le \alpha \|J_{k_x} - J^*\|_\infty \\ &\le \alpha \|J_{k_i-1} - J^*\|_\infty. \end{aligned}$$

Since this is true for all x, (3) follows.

4. (a) By observing

$$(I - \alpha P_u)\sum_{t=0}^{\infty} \alpha^t P_u^t = \sum_{t=0}^{\infty} \alpha^t P_u^t - \sum_{t=1}^{\infty} \alpha^t P_u^t = I$$

we conclude that $(I-\alpha P_u)^{-1} = \sum_{t=0}^{\infty} \alpha^t P_u^t$.

(b) We first show that $\mu_u = (1-\alpha)c^T(I-\alpha P_u)^{-1}$ is a probabilistic measure. We observe that $\mu_u \ge 0$, and

$$(1-\alpha)c^T(I-\alpha P_u)^{-1}e = (1-\alpha)c^T\sum_{t=0}^{\infty}\alpha^t P_u^t e = (1-\alpha)c^T ce \sum_{t=0}^{\infty}\alpha^t = ce = 1$$

where e is a vector of unit elements with the appropriate dimension.

Suppose $\mu_u(x) = 0$ for some $x \in \mathcal{S}$. Then, by $\mu_u = (1-\alpha)c^T\sum_{t=0}^{\infty}(I-\alpha P_u)^{-1}$, since $c > 0$, one column of matrix $(I-\alpha P_u)^{-1}$ must be a null vector, which contradicts to our conclusion in (a). Hence, $\mu_u(x) > 0$, $\forall x \in \mathcal{S}$.

(c) For every $\nu \in (0,\infty)^{|\mathcal{S}|}$, define

$$||J||_{1,\nu} = \sum_{x\in\mathcal{S}} \nu(x)|J(x)|.$$

We will show that $||\cdot||_{1,c}$ is a norm.

First, we observe that $||J||_{1,c} \ge 0$ by definition, and, for any real number α, we have

$$||\alpha J||_{1,c} = \sum_{x\in\mathcal{S}} c(x)|\alpha J(x)| = |\alpha|\sum_{x\in\mathcal{S}} c(x)|J(x)| = |\alpha|||J||_{1,c}.$$

It is trivial to see that $||J||_{1,c} = 0$ if and only if $J = 0$. Last, for any two vectors J and K,

$$\begin{aligned} ||J+K||_{1,c} &= \sum_{x\in\mathcal{S}} c(x)|J(x)+K(x)| \\ &\le \sum_{x\in\mathcal{S}} c(x)(|J(x)|+|K(x)|) \\ &= ||J||_{1,c} + ||K||_{1,c} \end{aligned}$$

Therefore, $||\cdot||_{1,c}$ is a norm.

(e) Since u_J is a greedy policy with respect to J, we have

$$\begin{aligned} J_{u_J} - J &= (I-\alpha P_{u_J})^{-1} g_{u_J} - J \\ &= (I-\alpha P_{u_J})^{-1}(g_{u_J} + \alpha P_{u_J} J - J) \\ &= (I-\alpha P_{u_J})^{-1}(TJ - J) \end{aligned}$$

Then, since $J \leq J^*$, we have

$$\begin{aligned} ||J_{u_J} - J^*||_{1,c} &= c^T(J_{u_J} - J^*) \\ &\leq c^T(J_{u_J} - J) \\ &= c^T(I - \alpha P_{u_J})^{-1}(TJ - J) \\ &= \frac{1}{1-\alpha}\mu_{u_J}^T(TJ - J). \end{aligned}$$

Therefore, $||J_{u_J} - J^*||_{1,c} \leq \frac{1}{1-\alpha}\mu_{u_J}^T(TJ - J)$. Next, we have

$$\begin{aligned} \frac{1}{1-\alpha}\mu_{u_J}^T(TJ - J) &\leq \frac{1}{1-\alpha}\mu_{u_J}^T|TJ - J| \\ &= \frac{1}{1-\alpha}||TJ - J||_{1,\mu_{u_J}} \end{aligned}$$

(f) Given $J_0 \leq J^*$, by monotonicity, we have $J_1 = TJ_0 \leq TJ^* = J^*$ and thus $J_k \leq J^*$ for all integer $k \geq 1$. Since $J_k \leq J^*$ for all integer $k \geq 1$, we can apply (e) and have

$$\begin{aligned} ||J_{u_{J_k}} - J^*||_{1,c} &\leq \frac{1}{1-\alpha}\mu_{u_{J_k}}(TJ_k - J_k) \\ &\leq \frac{1}{1-\alpha}||TJ_k - J_k||_{1,u_{J_k}} \\ &\leq \frac{1}{1-\alpha}||TJ_k - J_k||_\infty \\ &= \frac{1}{1-\alpha}||TJ_k - TJ_{k-1}||_\infty \\ &\leq \frac{\alpha}{1-\alpha}||J_k - J_{k-1}||_\infty \\ &= \frac{\alpha}{1-\alpha}||TJ_{k-1} - J_{k-2}||_\infty \\ &\vdots \\ &\leq \frac{\alpha^k}{1-\alpha}||TJ_0 - J_0||_\infty. \end{aligned}$$

Hence, we have $||J_{u_{J_k}} - J^*||_{1,c} \leq \frac{\alpha^k}{1-\alpha}||TJ_0 - J_0||_\infty$.

(g) We have

$$\begin{aligned} \|J_{u_{J_k}} - J^*\|_{1,c} &= \sum_x c(x)|J_{u_{J_k}}(x) - J^*(x)| \\ &\leq \sum_x c(x)\|J_{u_{J_k}} - J^*\|_\infty \\ &= \|J_{u_{J_k}} - J^*\|_\infty. \end{aligned}$$

Hence $\|J_{u_{J_k}} - J^*\|_{1,c} \leq \epsilon$ is a weaker stopping criterion than $\|J_{u_{J_k}} - J^*\|_\infty < \epsilon$, leading the algorithm to stop first. Note that, if the latter criterion is satisfied, we know that policy u_{J_k} has a cost-to-go that is close to the optimal cost-to-go function J^* for all initial states in the system. In the case of the weighted 1-norm criterion, u_{J_k} has a cost-to-go function that is close to J^* in *expected value*, for states distributed according to distribution c; the cost-to-go could still be very large at states x corresponding to very small $c(x)$. Hence $\|J_{u_{J_k}} - J^*\|_\infty < \epsilon$ offers stronger guarantees. However, a situation when one may consider using the weaker criterion $\|J_{u_{J_k}} - J^*\|_{1,c} \leq \epsilon$ is when certain states are known not to be very important, for instance if they are visited extremely rarely. Especially if state spaces are large, requiring $\|J_{u_{J_k}} - J^*\|_\infty < \epsilon$ may be overly restrictive and even infeasible, and it becomes more important to distinguish between states that are more or less relevant.