100 %

MATH4406 – Assignment 6

Patrick Laub (ID: 42051392)

October 25, 2014

1 Advanced Regular Finite Markov Chains

Exercise 1

As $P^* = \lim_{T\to\infty} P^T$ then

$$PP^* = P \lim_{T\to\infty} P^T = \lim_{T\to\infty} P^{T+1} = P^* .$$

and equally $P^*P = P^*$. The last equality is

$$P^*P^* = \left(\lim_{T\to\infty} P^T\right)\left(\lim_{T\to\infty} P^T\right) = \lim_{T\to\infty} P^{2T} = P^* .$$

$$\therefore PP^* = P^*P = P^*P^* = P^* .$$

□

Exercise 4

As Theorem 2 Part 2 states that $\exists\, W$ s.t.

$$P = W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} W$$

then

$$P^2 = W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} W \times W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} W = W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix}^2 W = W^{-1}\begin{bmatrix} Q^2 & 0 \\ 0 & 1 \end{bmatrix} W .$$

Iterating n times gives

$$P^n = W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix}^n W = W^{-1}\begin{bmatrix} Q^n & 0 \\ 0 & 1 \end{bmatrix} W .$$

So therefore

$$\begin{aligned}\frac{1}{T}\sum_{t=0}^{T-1} P^t = \frac{1}{T}\sum_{t=0}^{T-1} W^{-1}\begin{bmatrix} Q^t & 0 \\ 0 & 1 \end{bmatrix} W &= W^{-1}\left[\frac{1}{T}\sum_{t=0}^{T-1}\begin{bmatrix} Q^t & 0 \\ 0 & 1 \end{bmatrix}\right] W \\ &= W^{-1}\begin{bmatrix} \frac{1}{T}\sum_{t=0}^{T-1} Q^t & 0 \\ 0 & T/T \end{bmatrix} W .\end{aligned}$$

$$\therefore \frac{1}{T}\sum_{t=0}^{T-1} P^t = W^{-1}\begin{bmatrix} \frac{1}{T}\sum_{t=0}^{T-1} Q^t & 0 \\ 0 & 1 \end{bmatrix} W .$$

□

Ex 6 $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} P^t = \lim_{T\to\infty} W^{-1}\begin{bmatrix} \frac{1}{T}\sum_{t=0}^{T-1} Q^t & 0 \\ 0 & 1 \end{bmatrix} W$ by Ex. 4

$= W^{-1}\begin{bmatrix} \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q^t & 0 \\ 0 & 1 \end{bmatrix} W = W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W$ by Ex. 5

$=: P^*$ by Thm. 2. Part 3.

$\therefore \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} P^t = W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W = P^*$. $\square$ ✓

Ex 7 1) $B^\# B B^\# = W^{-1}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W W^{-1} \begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} W W^{-1} \begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W$

$= W^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = W^{-1}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = B^\#$. ✓

$\therefore B^\# B B^\# = B^\#$.

2) $BB^\# = W^{-1}\begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} W W^{-1}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = W^{-1}\begin{bmatrix} CC^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = W^{-1}\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} W$

$= W^{-1}\begin{bmatrix} C^{-1}C & 0 \\ 0 & 0 \end{bmatrix} W = W^{-1}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix} W W^{-1}\begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix} = B^\# B$. ✓

$\therefore BB^\# = B^\# B$.

3) Same as 1), replace $C^{-1} \leftrightarrow C$ everywhere and $B^\# \leftrightarrow B$. ✓

Ex 8 $I - P + P^* = W^{-1}\begin{bmatrix} I & 0 \\ 0 & 1 \end{bmatrix} W - W^{-1}\begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} W + W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W$ by Thm 2.

$= W^{-1}\begin{bmatrix} I-Q & 0 \\ 0 & 0 \end{bmatrix} W + W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W = W^{-1}\begin{bmatrix} I-Q & 0 \\ 0 & 1 \end{bmatrix} W$ ✓

Thm 2 Part 2(b) tells us $(I-Q)^{-1}$ exists so therefore

$\therefore (I-P+P^*)^{-1} = Z_P = W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 1 \end{bmatrix} W$ exists so $(I-P+P^*)$ is non-singular ✓

$\square$

Ex 9 $(I-P) = W^{-1}\begin{bmatrix} (I-Q) & 0 \\ 0 & 0 \end{bmatrix} W$ as per Ex. 8 so

$$H_P = (I-P)^{\#} = W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = \underbrace{W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 1 \end{bmatrix} W}_{Z_P \text{ (inverse of Ex 8)}} - \underbrace{W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W}_{P^* \text{ (Thm 2 Part 3)}}$$

$= Z_P - P^*$ and as $P^* = Z_P P^*$ (from Ex 10 Part 4)

$= Z_P - Z_P P^*$

$= Z_P(I - P^*)$. $\therefore H_P = Z_P(I-P^*)$. $\square$.

Ex 11 1) $(I-\lambda P)(1+\rho)R^\rho = (I-\lambda P)(1+\rho)(\rho I + (I-P))^{-1}$

$= (1+\rho)^{-1}(\rho I + (I-P))\ (1+\rho)(\rho I + (I-P))^{-1}$

$= \dfrac{1+\rho}{1+\rho}(\rho I + (I-P))(\rho I + (I-P))^{-1} = I$.

$\therefore (I-\lambda P)^{-1} = (1+\rho)R^\rho$. $\square$. ✓

2) Substitute $\lambda = (1+\rho)^{-1}$ into part 1) to get

$(I-\lambda P)^{-1} = \frac{1}{\lambda}R^\rho$

$\therefore R^\rho = \lambda(I-\lambda P)^{-1}$. $\square$. ✓

Ex 13 Know that $W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W = P^*$ from Thm 2 Part 3, so the first term: $\therefore \rho^{-1}W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W = \rho^{-1}P^*$. $\square$.

Ex 20 $P^* + (I-P)H_P = I$, $\underset{\sim}{h} = H_P\underset{\sim}{r}$, $\underset{\sim}{g} = P^*\underset{\sim}{r}$.

$\Rightarrow \underset{\sim}{g} + (I-P)\underset{\sim}{h} = \underset{\sim}{r}$ by right-multiplying by $\underset{\sim}{r}$. $\square$

Ex 12 $(R^{\rho})^{-1} = \rho I + I - P = W^{-1}\begin{bmatrix} \rho I + B & 0 \\ 0 & \rho \end{bmatrix} W$ is given so

taking inverse gives $R^{\rho} = W^{-1}\begin{bmatrix} (\rho I+B)^{-1} & 0 \\ 0 & \rho^{-1} \end{bmatrix} W$ and splitting the block matrices apart gives

See below for proof this exists!

$$\therefore R^{\rho} = \rho^{-1} W^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} W + W^{-1}\begin{bmatrix} (\rho I+B)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W. \quad \square$$

Ex 14 Consider Laurent series expansion of $(\rho I+B)^{-1}$, first factor out B^{-1}

$$(\rho I + B)^{-1} = (I + \rho B^{-1})^{-1} B^{-1}$$ as $B^{-1} = (I-Q)^{-1}$ exists by Thm 2 Part 2(b).

See over page for justification of this line.

$$= \left[\sum_{n=0}^{\infty} (-1)^n (\rho B^{-1})^n\right] B^{-1}$$ by geometric series expansion when $\sigma(\rho B^{-1}) = \rho\,\sigma((I-Q)^{-1}) = \frac{\rho}{\sigma(I-Q)} < 1$

$$= \sum_{n=0}^{\infty} (-\rho)^n B^{-(n+1)}$$ i.e. $\rho < \sigma(I-Q)$ or equiv. $\rho < \sigma(I-P)$ (by Thm 2 Part 2(c)).

$$\Rightarrow W^{-1}\begin{bmatrix} (\rho I+B)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W = W^{-1}\begin{bmatrix} \sum_{n=0}^{\infty}(-\rho)^n B^{-(n+1)} & 0 \\ 0 & 0 \end{bmatrix} W$$

$$= \sum_{n=0}^{\infty} (-\rho)^n W^{-1}\begin{bmatrix} (I-Q)^{-(n+1)} & 0 \\ 0 & 0 \end{bmatrix} W$$ by rearrang. & subst. $B = I - Q$.

$$= \sum_{n=0}^{\infty} (-\rho)^n \left(W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W \right)^{n+1}$$ since the WW^{-1} all cancel out

$\therefore$ Second term $= \sum_{n=0}^{\infty} (-\rho)^n H_P^{\,n+1}$ when $\rho \in (0, \sigma(I-P))$ $\square$ $\leftarrow$ by Ex 9. $\left(H_P = (I-P)^{\#} = W^{-1}\begin{bmatrix} (I-Q)^{-1} & 0 \\ 0 & 0 \end{bmatrix} W \right)$

Ex 12 Continued $\rho I + B = \rho I + (I - Q)$ and remember $\sigma(Q) < 1$ by Thm. 2 Part 2(a).

$$= (1+\rho) I - Q.$$

Consider $I - (1+\rho)^{-1} Q$, as $\sigma\left(\frac{1}{1+\rho} Q\right) = \frac{1}{1+\rho}\sigma(Q) < 1$ then

$(I - (1+\rho)^{-1}Q)^{-1}$ exists,

and as such any (non-zero) scalar multiple of $I - (1+\rho)^{-1}Q$ is invertible,

$\Rightarrow ((1+\rho)I - Q)^{-1}$ exists

$\therefore (\rho I + B)^{-1}$ exists.

Ex 14 continued To show $(\rho I + B)^{-1} = (I + \rho B^{-1})^{-1} B^{-1}$

$$(\rho I + B)^{-1}\left[(I+\rho B^{-1})^{-1}B^{-1}\right]^{-1} = (\rho I + B)^{-1} B (I + \rho B^{-1})$$

$$= (\rho I + B)^{-1}(B + \rho I) = I. \quad \checkmark$$

Ex 15 $\quad h = H_P r = \left(\sum_{t=0}^{\infty}(P^t - P^*)\right) r$ by Thm. 3 Part 3

$$= \sum_{t=0}^{\infty} P^t r - P^* r = \sum_{t=0}^{\infty} P^t r - P^t P^* r \qquad \text{as } PP^* = P^* \text{ so } P^t P^* = P^*$$

$$= \sum_{t=0}^{\infty} P^t (r - P^* r) = \sum_{t=0}^{\infty} P^t (r - g) \qquad \text{as } g = P^* r. \quad \checkmark$$

At $t=0$ then $P^0(r-g) = I(r-g)$, in component form if starting at $s \in S$ then $r(s) - g(s) = r(X_0) - g(X_0)$, and at $t=1$ then $P^1(r-g) = \mathbb{E}_s[r(X_1) - g(X_1)]$, so adding together gives $h(s) = \sum_{t=0}^{\infty} \mathbb{E}_s[r(X_t) - g(X_t)]$

$$\therefore h(s) = \mathbb{E}_s\left[\sum_{t=0}^{\infty} r(X_t) - g(X_t)\right]. \quad \square$$

, so right Hmmm... Okay.

Ex 18 $\quad v_\lambda = \frac{1}{1-\lambda} g + h + o(1-\lambda)$

So

$$\lim_{\lambda \uparrow 1}(1-\lambda)v_\lambda = \lim_{\lambda \uparrow 1} \underbrace{g}_{\text{constant w.r.t } \lambda} + \underbrace{h(1-\lambda)}_{\to 0} + \underbrace{(1-\lambda)}_{\to 0}\underbrace{o(1-\lambda)}_{\to 0}$$

$$= g.$$

$$\therefore g = \lim_{\lambda \uparrow 1}(1-\lambda) v_\lambda. \quad \square$$

Ex 16 $\quad h = \sum_{t=0}^{\infty} P^t (r - g) = \underbrace{\sum_{t=0}^{T-1} P^t r}_{V_{T+1}} - \underbrace{\sum_{t=0}^{T-1} P^t g}_{P^t g = g \text{ as } P^t P^* = P^*} + \underbrace{\sum_{t=T}^{\infty} P^t (r - g)}_{\to 0 \text{ as } T \to \infty, \text{ so } o(1)}$

(as eventually no terms in the summation).

$$\Rightarrow h = V_{T+1} - \overbrace{\sum_{t=0}^{T-1} g}^{Tg} + o(1)$$

$$\therefore V_{T+1} = Tg + h + o(1). \quad \square$$

Ex 5. $$\sum_{t=0}^{T-1} Q^t(I-Q) = \sum_{t=0}^{T-1} Q^t - Q^{t+1} = Q^0 - Q^T = I - Q^T$$

$$\Rightarrow \sum_{t=0}^{T-1} Q^t = (I-Q^T)(I-Q)^{-1}$$ as $(I-Q)$ is non-singular.

So $$\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q^t = \lim_{T\to\infty} \frac{1}{T} (I-Q^T)(I-Q)^{-1}$$

$$= \left(\lim_{T\to\infty} \frac{I-Q^T}{T}\right)(I-Q)^{-1}$$ As $\sigma(Q)<1$ then $Q^T \to 0$

$$= 0\cdot(I-Q)^{-1} = 0.$$

$$\therefore \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} Q^t = 0.$$ $\square$

Part 2 Examining policies whose stationary distributions spent $\geq 10\%$ of time "heavily worn" had a pattern, if the decision rule is of the form

[0, 0, x, x],
[0, x, 0, x], or
[x, x, x, 0]

where x is a space for any value in A_s.

So skipping these "patterns" in the argmax does constrained optimization.

$\therefore$ optimal policy is $d(s) \begin{cases} 0 & , s=0 \\ 1 & , s=1 \\ 1 & , s=2 \\ 1 & , s=3 \end{cases}$. or

```
function d = policy_iteration()
% Author: Patrick Laub
    % Problem data.
    S = 0:3; numS = numel(S);
    A = 0:2; numA = numel(A);
    P = zeros(numS, numA, numS);
    P(1,:,:) = [0.1, 0.6, 0.2, 0.1; 0.8, 0.2, 0, 0; 0.95, 0.05, 0, 0];
    P(2,:,:) = [0, 0.1, 0.6, 0.3; 0.7, 0.2, 0.1, 0; 0.85, 0.1, 0.05, 0];
    P(3,:,:) = [0, 0, 0.2, 0.8; 0.3, 0.4, 0.2, 0.1; 0.65, 0.2, 0.1, 0.05];
    P(4,:,:) = [0, 0, 0, 1; 0, 0.6, 0.2, 0.2; 0.5, 0.5, 0, 0];

    % Start with any guess decision rule.
    d = ones(numS, 1);
    nextD = d;

    % Construct decision rule at each step.
    for iters = 1:1e5
        % Construct reward and transition vectors.
        r = -d;
        Pd = zeros(numS);

        for s = 1:numS
            Pd(s, :) = P(s, d(s)+1, :);
        end

        % Do policy evalutation
        % i.e. solve 0 = r_{d_n} - g + (P_{d_n} - I) h.
        % Start by solving r = Q W
        Q = (eye(numS) - Pd);
        Q(:,1) = 1;
        W = Q \ r;

        % Extract the gain and bias.
        g = W(1)
        h = [0; W(2:end)]

        r - g*ones(numS, 1) + (Pd - eye(numS))*h

        % Do policy improvement.
        for i = 1:numS
            m = -Inf; bestA = -1;
            for a = 1:numA
                if i==2 && nextD(1) == 0 && a == 1
                    continue;
                end
                if i==3 && nextD(1) == 0 && a == 1
                    continue;
                end
                if i==4 && a == 1
                    continue
                end
                num = -a + reshape(P(i, a, :), 1, numS) * h;
                if num > m
                    m = num;
                    bestA = a-1;
```

```
                end
            end

            if bestA == -1
                error('Something went wrong!');
            end
            nextD(i) = bestA;
        end

        % If the decision rule has converged then stop iterating.
        if all(nextD == d)
            break;
        end

        d = nextD;
    end

    if iters == 1e5
        warning('Did not converge!');
    end
end
```