MATH4406 – Assignment 5*

Patrick Laub (ID: 42051392)

October 7, 2014

1 The machine replacement model

1.1 Real-world motivation

Consider the machine to be the entire world. Over time the creator has running costs — forgiving sinners and such — or it can call judgment day sending everyone up or down (with cost R of effort and paperwork involved) then start over again with a new world. The state of the world $i \in S$ is the number of sins committed in each year, and the world moves randomly over $S = \{0, 1, \dots\}$ but in generate trends upwards (due to the general increasing trend of wickedness and population/sinner growth). To be explicit: $R = 1$ day (i.e. it days an entire day to judge everybody), $C(i) = 10^{-100} i$ (assuming 10^{-100} days per prayer answered), and say $T_i \sim \text{TN}(1000i, 1000)$ (where TN is truncated normal (since everything is normal), scaled and discretized to match the support S), i.e.

$$P_{i,j} \propto \exp\left\{-\frac{(j-1000i)^2}{2 \times 1000^2}\right\}.$$

1.2 Stochastic ordering

This makes sense as machines are, on the whole, expected to deteriorate over time. Even though there may be local fluctuations (e.g. my computer may work slightly better today than it did yesterday) the overall trend is into hopeless machine oblivion.

1.3 Markov decision process

Discount factor:

$$\lambda \in (0, 1)\,.$$

State space:

$$S = \{0, 1, \dots\}\,.$$

Action set:

$$\forall s \in S \quad \mathcal{A}_s = \{0, 1\}\,.$$

Transition probabilities:

$$p(j \mid s, a) = \begin{cases} 1\,, & j = 0, s \in S, a = 1, \\ P_{s,j}\,, & j, s \in S, a = 0 \\ 0\,, & \text{otherwise} \end{cases}.$$

Rewards:

$$r(s, a) = \begin{cases} -C(s)\,, & s \in S, a = 0 \\ -C(s) - R\,, & s \in S, a = 1 \end{cases}.$$

*I call false advertising on the naming of these as "homework"!

Objective: for all $s \in S$ then (using X_t, Y_t as per notes)

$$\min_{\pi \in \Pi^{\rm MS}} \mathbb{E}_s^{\pi} \left[\sum_{t=1}^{\infty} \lambda^{t-1} r(X_t, Y_t) \right] .$$

1.4 Optimality equation

Equation (6.2.2) is, $\forall i \in S$,

$$v(i) = \sup_{a \in \mathcal{A}_i} \left\{ r(i,a) + \sum_{j \in S} \lambda \, p(j \,|\, i, a) \, v(j) \right\} . \tag{1}$$

As our action sets are finite the supremum is achieved,

$$v(i) = \max_{a \in \{0,1\}} \left\{ r(i,a) + \sum_{j \in S} \lambda \, p(j \,|\, i, a) \, v(j) \right\} .$$

However our rewards are negative, so this is equivalent to the following (using the minimisation criterion)

$$\begin{aligned} v(i) &= \min_{a \in \{0,1\}} \left\{ -r(i,a) + \sum_{j \in S} \lambda \, p(j \,|\, i, a) \, v(j) \right\} \\ &= \min \left\{ C(i) + R + \lambda v(0) \ , \ C(i) + \sum_{j=0}^{\infty} \lambda \, P_{i,j} \, v(j) \right\} \\ &= C(i) + \min \left\{ R + \lambda v(0) \ , \ \lambda \sum_{j=0}^{\infty} P_{i,j} \, v(j) \right\} . \end{aligned}$$

□

1.5 Optimality equation increasing

Lemma: The (intermediate) values functions $v^n(\cdot)$ created by value iteration algorithm — starting with $v^0(i) = C(i)$ — are all increasing functions.

Proof: Show by induction. At $n = 0$ then $v^0(i) = C(i)$ is increasing by assumption in the given problem.

Assume at step $k \geq 0$ then $v^k(\cdot)$ is an increasing function, and consider step $k+1$. Take any $s, t \in S$ with $s < t$ then look at the sign of

$$v^{k+1}(t) - v^{k+1}(s) = C(t) + \min\{R + \lambda v^k(0) \, , \, \lambda \sum_{j=0}^{\infty} P_{t,j} v^k(j)\} - C(s) - \min\{R + \lambda v^k(0) \, , \, \lambda \sum_{j=0}^{\infty} P_{s,j} v^k(j)\} .$$

To show that $v^{k+1}(\cdot)$ is increasing, i.e. that $v^{k+1}(t) - v^{k+1}(s) > 0$, then it will be demonstrated that

$$\underbrace{C(t) - C(s)}_{>0} + \overbrace{\min\{R + \lambda v^k(0) \, , \, \lambda \sum_{j=0}^{\infty} P_{t,j} v^k(j)\} - \min\{R + \lambda v^k(0) \, , \, \lambda \sum_{j=0}^{\infty} P_{s,j} v^k(j)\}}^{\geq 0} . \tag{2}$$

The first equality is true by the assumption of the problem (i.e. $C(\cdot)$ is increasing, as noted earlier). To show the second inequality is true then consider the general problem of showing $\min\{a,b\}-\min\{c,d\} \geq 0$ for any $a,b,c,d \in \mathbb{R}$. One approach is to show that for every element in the first set, there exists an element in the second which is not larger, i.e.

$$(a \geq c \vee a \geq d) \wedge (b \geq c \vee b \geq d) \Rightarrow \min\{a,b\} \geq \min\{c,d\} \Leftrightarrow \min\{a,b\} - \min\{c,d\} \geq 0.$$

To apply this argument to the second inequality of (2) then it needs to be shown that

$$\left(R + \lambda v^k(0) \geq R + \lambda v^k(0)\right) \vee \left(R + \lambda v^k(0) \geq \lambda \sum_{j=0}^{\infty} P_{s,j} v^k(j)\right) \tag{3}$$

and

$$\left(\lambda \sum_{j=0}^{\infty} P_{t,j} v^k(j) \geq R + \lambda v^k(0)\right) \vee \left(\lambda \sum_{j=0}^{\infty} P_{t,j} v^k(j) \geq \lambda \sum_{j=0}^{\infty} P_{s,j} v^k(j)\right). \tag{4}$$

Statement (3) is trivially true by the first logical statement ($x \geq x$ always!). Next (4) will be shown true by proving the second logical statement. Say that T_s and T_t are defined as per the question description (i.e. as the random next state visited after being in states s and t resp.), then

$$\sum_{j=0}^{\infty} P_{t,j} v^k(j) = \mathbb{E}[v^k(T_t)] \quad \text{and} \quad \sum_{j=0}^{\infty} P_{s,j} v^k(j) = \mathbb{E}[v^k(T_s)]\,.$$

As $v^k(\cdot)$ is an increasing function — this is the inductive assumption — then we have that $\mathbb{E}[v^k(T_t)] \geq \mathbb{E}[v^k(T_s)]$ due to the stochastic ordering induced by P. So

$$\lambda \sum_{j=0}^{\infty} P_{t,j} v^k(j) - \lambda \sum_{j=0}^{\infty} P_{s,j} v^k(j) = \lambda \left(\mathbb{E}[v^k(T_t)] - \mathbb{E}[v^k(T_t)]\right) \geq 0\,.$$

$\Rightarrow v^k(\cdot)$ is an increasing function implies that $v^{k+1}(\cdot)$ is an increasing function.

$\therefore$ For all $n \in \mathbb{N}_0$ then $v^n(\cdot)$ is an increasing function.

□

Corollary: $v(\cdot)$ (i.e. the solution to (1)) is an increasing function.

Remember that value iteration converges to the true value function, i.e.

$$\lim_{n\to\infty} v^n(\cdot) = v^*_\lambda(\cdot)$$

and the left-hand side is a sequence of increasing functions (by the lemma) so therefore $v^*_\lambda(\cdot)$ is also an increasing function. Since value iteration converges to an optimal policy then $v^*_\lambda(\cdot)$ is a solution to (1)) hence $v(\cdot)$ is an increasing function. □

1.6 Optimal policy

The optimality equation derived earlier was

$$v(i) = C(i) + \min\left\{R + \lambda v(0)\ ,\ \lambda \sum_{j=0}^{\infty} P_{i,j}\, v(j)\right\}$$

and hence the optimal policy satisfies

$$d(i) \in \arg\min_{a\in\{0,1\}} \left\{ C(i) + \underbrace{\left(\lambda \sum_{j=0}^{\infty} P_{i,j}\, v(j)\right)}_{=f(i)} \mathbb{1}(a=0) + \overbrace{\left(R + \lambda v(0)\right)}^{=\kappa} \mathbb{1}(a=1) \right\}$$
$$\in \arg\min_{a\in\{0,1\}} \left\{ f(i)\mathbb{1}(a=0) + \kappa \mathbb{1}(a=1) \right\}$$

where $\kappa \in \mathbb{R}$ and $f : S \to \mathbb{R}$. As done earlier, $f(i)$ can be rewritten as $f(i) = \lambda \mathbb{E}[v(T_i)]$. Since $v(\cdot)$ is an increasing function (from previous corollary) then the stochastic ordering applies, so $\mathbb{E}[v(T_i)] \leq \mathbb{E}[v(T_{i+1})]$, hence for $i, j \in S$ then

$$i < j \Rightarrow f(i) \leq f(j)\,.$$

As $f(\cdot)$ is non-decreasing and κ is constant then this leaves three cases:

1. $f(\cdot) \leq \kappa$ always,
2. $f(\cdot) \leq \kappa$ up until some point then afterwards $f(\cdot) \geq \kappa$,
3. $f(\cdot) \geq \kappa$ always.

A succinct summary is to say $\exists\, \bar{i} \in S \cup \{\infty\}$ such that

$$i < \bar{i} \Rightarrow f(\cdot) \leq \kappa, \quad i \geq \bar{i} \Rightarrow f(\cdot) \geq \kappa\,.$$

Hence the optimal policy is

$$d(i) = \begin{cases} 0, & i < \bar{i} \\ 1, & i \geq \bar{i} \end{cases}$$

i.e. that the machine is only replaced when in states $i \geq \bar{i}$. □

1.7 Never replace policy

A somewhat trivial case were $\bar{i} = \infty$ (which corresponds to the optimal policy of never replacing the machine) is when $R = \infty$. If $f(i)$ is always finite and R infinite then $d(i) = 0$ for all $i \in S$.

1.8 Algorithm to find replacement point

Consider a sequence of restricted state spaces $s_1, s_2, \ldots$ where $s_i = \{0, 1, \ldots, 10^i\}$. Use policy iteration on each of these finite MDPs to get optimal policies $d_i^* : s_i \to \{0, 1\}$ until some $\bar{i} = \min\{s : d_i^*(s) = 1\}$ exists. If s_i gets too large before $d_i^*(s) = 1$ occurs then estimate $\bar{i} = \infty$.

2 Contraction mappings and rates of convergence

2.1 Rates of convergence of value iteration

Definition: Say V is the set of bounded real valued functioned on S with componentwise partial order and norm $||v|| \stackrel{\text{def}}{=} \sup_{s\in S} |v(s)|$.

Nomenclature: Say that for $v \in V$ that $||v - v_\lambda^*||$ is the **error** of v.

Theorem 6.3.3.:[1] Let $v^0 \in V$ and let $\{v^n\}$ denote the iterates of value iteration. Then the following global convergence rate properties hold for the value iteration algorithm :

[1] All text in italics are my additions/explanations, the rest is a direct copy of the theorem statement.

a) convergence is linear at rate λ, *i.e. for each* $n = 0, 1, \ldots$ *then*

$$\left|\left|v^{n+1} - v^*_\lambda\right|\right| \leq \lambda \left|\left|v^n - v^*_\lambda\right|\right|,$$

Meaning: the error at step $n+1$ *is decreasing linearly at rate* λ*. E.g. say* $\lambda = 0.5$ *then the error of* v_{n+1} *is no greater than half the error of* v_n*.*

b) its asymptotic average rate of convergence equals λ, *i.e.*

$$\limsup_{n\to\infty} \left[\frac{\left|\left|v^n - v^*_\lambda\right|\right|}{\left|\left|v^0 - v^*_\lambda\right|\right|}\right]^{1/n} = \lambda,$$

Meaning: for any $\epsilon > 0$ *there exists an* N *such that, for* $n \geq N$*,*

$$\left|\left|v^n - v^*_\lambda\right|\right| \leq (\lambda + \epsilon)^n \left|\left|v^0 - v^*_\lambda\right|\right|.$$

Say we wish to know the number of iterations n *required to reduce the error of the* n*-th step (i.e.* $||v^n - v^*_\lambda||$*) by a fraction* ϕ *of the initial error (i.e.* $||v^0 - v^*_\lambda||$*). Then* $n \approx \log(\phi)/\log(\lambda)$*.*

c) it converges $\mathcal{O}(\lambda^n)$, *i.e.*

$$\limsup_{n\to\infty} \frac{\left|\left|v^n - v^*_\lambda\right|\right|}{\lambda^n} < \infty,$$

Meaning: the error of v^n *converges geometrically, with rate* λ*, to* v^*_λ*.*

d) for all n

$$\left|\left|v^n - v^*_\lambda\right|\right| \leq \frac{\lambda^n}{1-\lambda} \left|\left|v^1 - v^0\right|\right|,$$

Meaning: The result considers the n*-th value iterate* v^n *and gives an upper bound to its error.*

e) for any $d_n \in \arg\max_{d\in D}\{r_d + \lambda P_d v^n\}$,

$$\left|\left|v_\lambda^{(d_n)^\infty} - v^*_\lambda\right|\right| \leq \frac{2\lambda^n}{1-\lambda} \left|\left|v^1 - v^0\right|\right|.$$

Meaning: An optimal decision rule (i.e. optimal stationary policy) is constructed by

$$d^* \in \arg\max_{d\in D}\{r_d + \lambda P_d v^*_\lambda\}$$

and so the decision rule d_n *constructed by*

$$d_n \in \arg\max_{d\in D}\{r_d + \lambda P_d v^n\}$$

can be seen as the n*-th approximation to the optimal decision rule. Policy evaluation for this stationary policy will give a value* $v_\lambda^{(d_n)^\infty}$ *which is different from the value iterate* v^n *that generated it. This statement says the error of* $v_\lambda^{(d_n)^\infty}$ *cannot be greater than double the error of* v^n*.*

2.2 Proof

a) For any $v^0 \in V$, the iterates of value iteration satisfy

$$\left|\left|v^{n+1} - v^*_\lambda\right|\right| = \left|\left|Lv^n - Lv^*_\lambda\right|\right| \leq \lambda \left|\left|v^n - v^*_\lambda\right|\right|. \tag{5}$$

This means that value iteration converges linearly with rate no greater than λ, but is this bound achievable? If one chooses a sequence starting with v^0 as given below, then the upper bound is

attained.

Let $e \in V$ is the unit function, i.e. $e(s) = 1, \forall s \in S$. Choosing $v^0 = v^*_\lambda + ke$, where $k \in \mathbb{R} \setminus \{0\}$, gives

$$v^1 - v^*_\lambda = \lambda(v^0 - v^*_\lambda)\,.$$

Thus for this sequence, (5) holds with equality.

$$\therefore \text{ Value iteration converges linearly at rate } \lambda\,.$$

□

b) Iterating (5) gives, for $n = 1, \ldots,$ that

$$\Rightarrow ||v^n - v^*_\lambda|| \leq \lambda^n ||v^0 - v^*_\lambda||\,, \tag{6}$$

$$\Rightarrow \frac{||v^n - v^*_\lambda||}{||v^0 - v^*_\lambda||} \leq \lambda^n\,,$$

$$\Rightarrow \left[\frac{||v^n - v^*_\lambda||}{||v^0 - v^*_\lambda||}\right]^{1/n} \leq \lambda\,.$$

As this holds for all n then it also holds for the lim sup of the sequence:

$$\limsup_{n\to\infty} \left[\frac{||v^n - v^*_\lambda||}{||v^0 - v^*_\lambda||}\right]^{1/n} = \lim_{n\to\infty} \sup_{m\geq n} \left\{\left[\frac{||v^n - v^*_\lambda||}{||v^0 - v^*_\lambda||}\right]^{1/m}\right\} \leq \lim_{n\to\infty} \lambda = \lambda.$$

$$\Rightarrow \limsup_{n\to\infty} \left[\frac{||v^n - v^*_\lambda||}{||v^0 - v^*_\lambda||}\right]^{1/n} \leq \lambda\,.$$

That equality holds follows by again choosing $v^0 = v^*_\lambda + ke$.

$$\therefore \text{ The asymptotic average rate of convergence equals } \lambda\,.$$

□

c) Returning to (6) and dividing by λ^n gives

$$\frac{||v^n - v^*_\lambda||}{\lambda^n} \leq ||v^0 - v^*_\lambda||\,.$$

Choosing $v^0 = v^*_\lambda + ke$ as above shows that equality holds.

$$\therefore \text{ Value iteration converges } \mathcal{O}(\lambda^n).$$

□

d) From a) we have that

$$||v^n - v^*_\lambda|| \leq \lambda ||v^{n-1} - v^*_\lambda|| = \lambda ||v^{n-1} - v^n + v^n - v^*_\lambda||\,,$$

and by the triangle inequality

$$||v^n - v^*_\lambda|| \leq \lambda ||v^n - v^*_\lambda|| + \lambda ||v^n - v^{n-1}||\,,$$

$$\Rightarrow ||v^n - v^*_\lambda||(1 - \lambda) \leq \lambda ||v^n - v^{n-1}||\,,$$

$$\Rightarrow \left|\left|v^n - v^*_\lambda\right|\right| \leq \frac{\lambda}{1-\lambda}\left|\left|v^n - v^{n-1}\right|\right|, \tag{7}$$

Note that

$$\left|\left|v^n - v^{n-1}\right|\right| = \left|\left|Lv^{n-1} - Lv^{n-2}\right|\right| \leq \lambda\left|\left|v^{n-1} - v^{n-2}\right|\right|$$

and iterating this (i.e. using the contraction mapping property many times) gives

$$\left|\left|v^n - v^{n-1}\right|\right| \leq \lambda^{n-1}\left|\left|v^1 - v^0\right|\right|. \tag{8}$$

Substituting (8) into (7) gives the result

$$\therefore \left|\left|v^n - v^*_\lambda\right|\right| \leq \frac{\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|.$$

□

e) First a lemma will be proved in order to complete this larger proof.

Lemma:

$$\left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| \leq \frac{\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|.$$

Proof: Again using the triangle inequality

$$\left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| \leq \left|\left|v_\lambda^{(d_n)^\infty} - Lv^n\right|\right| + \left|\left|Lv^n - v^n\right|\right|.$$

Theorem 6.2.5 tells us that $v_\lambda^{(d_n)^\infty}$ is a (actually the unique) fixed point of $L_{(d_n)^\infty}$; i.e. $v_\lambda^{(d_n)^\infty} = L_{(d_n)^\infty} v_\lambda^{(d_n)^\infty}$. Also $Lv^n = L_{(d_n)^\infty} v^n$ by the fact that d_n was chosen to maximise the policy evaluation step. So substituting these statements (and $v^n = Lv^{n-1}$) into the last inequality gives:

$$\begin{aligned}\left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| &\leq \left|\left|L_{(d_n)^\infty} v_\lambda^{(d_n)^\infty} - L_{(d_n)^\infty} v^n\right|\right| + \left|\left|Lv^n - Lv^{n-1}\right|\right| \\ &\leq \lambda\left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| + \lambda\left|\left|v^n - v^{n-1}\right|\right|.\end{aligned}$$

Collecting terms:

$$\Rightarrow \left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right|(1-\lambda) \leq \lambda\left|\left|v^n - v^{n-1}\right|\right|.$$

The right-hand side is bounded by (8) so

$$\left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right|(1-\lambda) \leq \lambda\,\lambda^{n-1}\left|\left|v^1 - v^0\right|\right|$$

$$\therefore \left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| \leq \frac{\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|.$$

□

RTP:

$$\left|\left|v_\lambda^{(d_n)^\infty} - v^*_\lambda\right|\right| \leq \frac{2\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|.$$

Proof: Using the triangle inequality, the previous lemma, and d):

$$\begin{aligned}\left|\left|v_\lambda^{(d_n)^\infty} - v^*_\lambda\right|\right| &\leq \left|\left|v_\lambda^{(d_n)^\infty} - v^n\right|\right| + \left|\left|v^n - v^*_\lambda\right|\right| \\ &\leq \frac{\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right| + \frac{\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|\end{aligned}$$

$$\therefore \left|\left|v_\lambda^{(d_n)^\infty} - v^*_\lambda\right|\right| \leq \frac{2\lambda^n}{1-\lambda}\left|\left|v^1 - v^0\right|\right|.$$

□

2.3 Conditions for policy iteration

Firstly, it must be noted that policy iteration (PI) is not directly applicable to infinite-state or infinite-action models (modified PI may address this). So consider only finite-state and finite-action MDPs. If for every $s \in S$ then: $\mathcal{A}_s$ is compact and convex, $p(j \mid s, a)$ is affine in a, and $r(s, a)$ is strictly concave and twice continuously differentiable in a then PI converges quadratically. In this case PI is preferred to VI. The conditions arise as PI can be viewed as a form of Newton's method, which has superlinear convergence for "nice" problems (something like Lipschitz continuous bla bla).