Caleb Logemann
AERE 504 Intelligent Air Systems
Final Take-Home Exam

#1 What is the size of the state space?

The state space is the product of the number of options of r, h and t. So the size of the state space is $2 \times 21 \times 11 = 462$.

#2 What is the size of the observation space?

After each action we know that t decreases by one, and we can't directly observe if aircraft B is responsive or not. We can only observe the new h. So the size of the observation space is 21.

#3 What is the dimensionality of our belief state?

The only unobservable part of the state space is the responsiveness r of aircraft B, so the dimensionality of the belief space is 1. The belief state will be a single parameter p which will represent the probability that $r = 0$, and thus $1 - p$ is the probability that $r = 1$.

#4 Assume our initial belief is uniform over all states with $t = 10$. After the first observation, how many components of the belief vector will be non-zero?

After the first observation we know h and t exactly, so the only unknown is r. If we started with an uniform belief that would mean that intially we assign each possibility probability 0.5. One observation won't change this probability to zero. This means that the only component of the belief vector will be non-zero.

#5 Suppose we have a belief b that assigns probability 1 to state $[1, 10, 1]^T$; what is $Q^*(b, a_{+1})$ (assume $\lambda = -0.5$)? Provide an exact numerical value and explain.

From this state and taking this action there are two possible future states either $[1, 10, 0]^T$ with probability .75 or $[1, 9, 0]^T$ with probability .25. In either case the aircraft will not collide and the best action to take is a_0, so we know that

$$U([1, 10, 0]^T) = U([1, 9, 0]^T) = 0$$

Thus we can compute $Q^*(b, a_{+1})$ as

$$Q^*(b, a_{+1}) = R(s, a) + \sum_{s'} (T(s'|s, a)U(s'))$$

$$
\begin{aligned}
&= R([1,10,1]^T) + R(a_{+1}) + 0.75U([1,10,0]^T) + 0.25U([1,9,0]^T) \\
&= 0 - 0.5 + 0.75 \times 0 + 0.25 \times 0 \\
&= -0.5
\end{aligned}
$$

This makes sense as the cost of a_{+1} is -0.5 and a collision is not possible so nothing else affects the Q value.

#6 Suppose we have a belief b that assigns probability 1 to state $[0,10,1]^T$; what is $U^*(b)$ (assume $\lambda \leq 0$)? Provide an exact numerical value and explain.

In this case there is one step left before the model terminates. Given that $h = 10$, any action that is executed will result in $t = 0$ with $h \neq 0$, so $U^*(s') = 0$ for any subsequent state. If we execute the action a_{+1} there are

$$
\begin{aligned}
U^*(b) &= \max_a \left\{ \sum_s \left(b(s)R(s,a) + b(s) \sum_{s'} \left(T(s'|s,a)U^*(s') \right) \right) \right\} \\
&= \max_a \left\{ R([0,10,1]^T, a) + \sum_{s'} \left(T(s'|[0,10,1]^T, a)U^*(s') \right) \right\} \\
&= \max_a \left\{ R([0,10,1]^T, a) \right\} \\
&= \max \left\{ R([0,10,1]^T, a_{+1}), R([0,10,1]^T, a_{-1}), R([0,10,1]^T, a_0) \right\} \\
&= \max\{R(a_{+1}), R(a_{-1}), R(a_0)\} \\
&= \max\{\lambda, \lambda, 0\} \\
&= 0
\end{aligned}
$$

The value of this state is 0, because the best action to take is a_0 since h is large enough and t is small enough to guaruntee that a collision will not happen.

#7 Is it possible for $U^*\left([r,h,t]^T\right) \neq U^*\left([r,-h,t]^T\right)$ for some λ, r, h, and t? If so provide an example. If not, provide a simple explanation.

I will assume that the belief state assigns probability 1 to both respective states. In this case it is not possible for the values to be different. The model doesn't care if A is h above or below B. Is equally probable in both cases, also the cost of going up or going down is the same, so U^* will be the same in both cases.

#8 As $\lambda \to -\infty$, what is $\min_s\{U^*(s)\}$? Why?

For a state s where $h = 0$ and $t = 0$ with probability p, then

$$U^*(s) = -p$$

For all other states

$$U^*(s) = 0.$$

As $R(a_0)$ will be greater than λ for all actions and states. Therefore, if we let $p = 1$, then

$$\min_s\{U^*(s)\} = -1.$$

This makes sense as the cost of changing altitude becomes too great, either the system will collide or it will not. So the minimum will be the case were the aircraft collide.

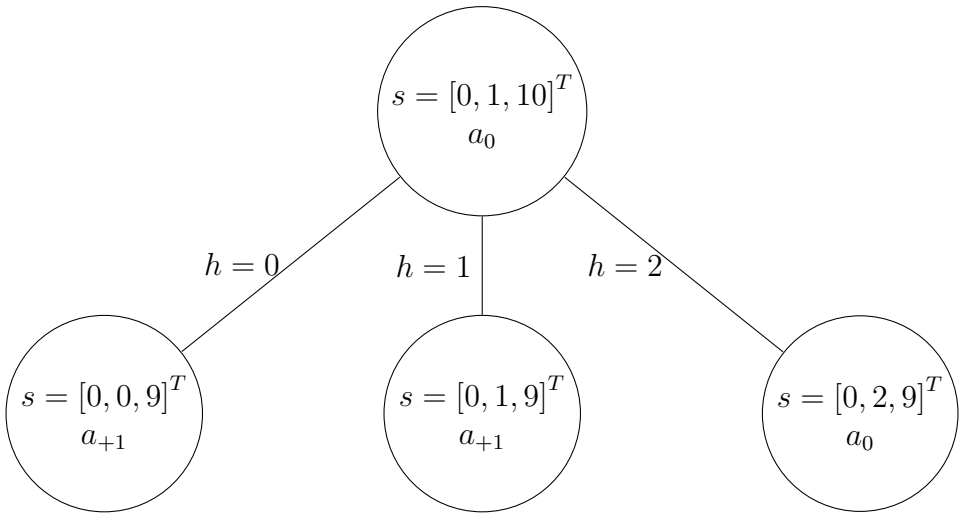
#9 Suppose we have a belief b that assigns probability 1 to state $[0, 9, 0]^T$. State an action that will maximize $Q^*(b, a)$ when $\lambda = 5$. Is it unique?

Either the action a_{+1} or the action a_{-1} will maximize $Q^*(b, a)$ to be 5. In either case

$$Q^*\Big([0, 9, 0]^T, a\Big) = R(a) = 5$$

Thus this is not a unique choice.

#10 Draw a two-step conditional plan from the state $[0, 1, 10]^T$ where the action associated with the root node is a_0. Only show the observation branches that have a non-zero probability of occuring.



#11 If we are using the fast informed bound (FIB) to approximate the optimal value function, how many alpha vectors will there be?

In the FIB method, there is one alpha vector for every action. Therefore there will be three alpha vectors for this model.

#12 If α_{QMDP} is an alpha vector generated by QMDP and α_{FIB} is an alpha vector generated by FIB, can there exist a b, then such that $b^T\alpha_{QMDP} < b^T\alpha_{FIB}$? Why or why not?

No this is not possible. Both QMDP and FIB provide upper bounds for $U^*(b)$, however FIB uses information about the partial observability so $b^T\alpha_{FIB}$ will be a tighter bound than $b^T\alpha_{QMDP}$. This means that

$$U^*(b) \leq b^T\alpha_{FIB} \leq b^T\alpha_{QMDP}$$

for all belief states. This shows that it is not possible for $b^T\alpha_{QMDP} < b^T\alpha_{FIB}$.

#13 Suppose we have a belief state b that assigned probability 0.5 to $[0,0,1]^T$ and probability 0.5 to $[1,0,1]^T$. What is the value for $U^*(b)$ in terms of λ (which may take on any negative value)?

$$U^*(b) = \max_a \left\{ \sum_s \left(b(s) \left(R(s,a) + \sum_{s'} \left(T(s'|s,a) U^*(s') \right) \right) \right) \right\}$$

First consider a_{+1}, then aircraft B is either responsive or not.

$$\begin{aligned}
Q\left([1,0,1]^T, a_{+1}\right) &= R\left([1,0,1]^T, a_{+1}\right) + \sum_{s'} \left(T(s'|s,a_{+1}) U^*(s') \right) \\
&= \lambda + 0.25U^*\left([1,3,0]^T\right) + 0.5U^*\left([1,2,0]^T\right) + 0.25U^*\left([1,1,0]^T\right) \\
&= \lambda + 0.25 \times 0 + 0.5 \times 0 + 0.25 \times 0 \\
&= \lambda \\
Q\left([0,0,1]^T, a_{+1}\right) &= R\left([0,0,1]^T, a_{+1}\right) + \sum_{s'} \left(T(s'|s,a_{+1}) U^*(s') \right) \\
&= \lambda + 0.25U^*\left([0,2,0]^T\right) + 0.5U^*\left([0,1,0]^T\right) + 0.25U^*\left([0,0,0]^T\right) \\
&= \lambda + 0.25 \times 0 + 0.5 \times 0 + 0.25 \times -1 \\
&= \lambda - 0.25
\end{aligned}$$

$$
\begin{aligned}
Q(b, a_{+1}) &= \sum_s \left(b(s) \left(R(s, a_{+1}) + \sum_{s'} \left(T(s'|s, a_{+1}) U^*(s') \right) \right) \right) \\
&= \frac{1}{2}\lambda + \frac{1}{2}(\lambda - 0.25) \\
&= \lambda - \frac{1}{8}
\end{aligned}
$$

Second consider a_{-1}, then

$$
\begin{aligned}
Q\left([1,0,1]^T, a_{-1}\right) &= R\left([1,0,1]^T, a_{-1}\right) + \sum_{s'} \left(T(s'|s, a_{-1}) U^*(s') \right) \\
&= \lambda + 0.25U^*\left([1,-3,0]^T\right) + 0.5U^*\left([1,-2,0]^T\right) + 0.25U^*\left([1,-1,0]^T\right) \\
&= \lambda + 0.25 \times 0 + 0.5 \times 0 + 0.25 \times 0 \\
&= \lambda \\
Q\left([0,0,1]^T, a_{-1}\right) &= R\left([0,0,1]^T, a_{-1}\right) + \sum_{s'} \left(T(s'|s, a_{-1}) U^*(s') \right) \\
&= \lambda + 0.25U^*\left([0,-2,0]^T\right) + 0.5U^*\left([0,-1,0]^T\right) + 0.25U^*\left([0,0,0]^T\right) \\
&= \lambda + 0.25 \times 0 + 0.5 \times 0 + 0.25 \times -1 \\
&= \lambda - 0.25 \\
Q(b, a_{-1}) &= \sum_s \left(b(s) \left(R(s, a_{-1}) + \sum_{s'} \left(T(s'|s, a_{-1}) U^*(s') \right) \right) \right) \\
&= \frac{1}{2}\lambda + \frac{1}{2}(\lambda - 0.25) \\
&= \lambda - \frac{1}{8}
\end{aligned}
$$

Finally consider a_0, then

$$
\begin{aligned}
Q\left([1,0,1]^T, a_0\right) &= R\left([1,0,1]^T, a_0\right) + \sum_{s'} \left(T(s'|s, a_{-1}) U^*(s') \right) \\
&= 0 + 0.25U^*\left([1,1,0]^T\right) + 0.5U^*\left([1,0,0]^T\right) + 0.25U^*\left([1,-1,0]^T\right) \\
&= 0 + 0.25 \times 0 + 0.5 \times -1 + 0.25 \times 0 \\
&= -0.5 \\
Q\left([0,0,1]^T, a_0\right) &= R\left([0,0,1]^T, a_0\right) + \sum_{s'} \left(T(s'|s, a_{-1}) U^*(s') \right) \\
&= 0 + 0.25U^*\left([0,1,0]^T\right) + 0.5U^*\left([0,0,0]^T\right) + 0.25U^*\left([0,-1,0]^T\right)
\end{aligned}
$$

$$
\begin{aligned}
&= 0 + 0.25 \times 0 + 0.5 \times -1 + 0.25 \times 0 \\
&= -0.5 \\
Q(b, a_0) &= \sum_s \left(b(s) \left(R(s, a_0) + \sum_{s'} \left(T(s'|s, a_0) U^*(s') \right) \right) \right) \\
&= \frac{1}{2}\left(-\frac{1}{2}\right) + \frac{1}{2}\left(-\frac{1}{2}\right) \\
&= -\frac{1}{2}
\end{aligned}
$$

Now

$$
\begin{aligned}
U^*(b) &= \max_a \left\{ \sum_s \left(b(s) \left(R(s, a) + \sum_{s'} \left(T(s'|s, a) U^*(s') \right) \right) \right) \right\} \\
&= \max\left\{ \lambda - \frac{1}{8}, \lambda - \frac{1}{8}, -\frac{1}{2} \right\} \\
&= \begin{cases} \lambda - \frac{1}{8} & \lambda > -\frac{3}{8} \\ -\frac{1}{2} & \lambda \leq -\frac{3}{8} \end{cases}
\end{aligned}
$$

This makes sense because if the cost of taking rising or lowering is too high, then it is better to take a_0 and hope that $h \neq 0$ through chance in the transition.

#14 Why would you not use a particle filter to update your belief for this problem?

You would not use a particle filter to update your belief for this problem, because the state space is not particularly large or continuous. A particle filter is sampling approach that uses particle to sample the state space. In this case the state space is small enough to be enumerated and so sampling isn't necessary.

#15 Suppose your initial belief is uniform over the state space and then you observe that aircraft A is 3 units above aircraft b after executing a_0. What probability would an exact Bayesian update of your belief state assign to aircraft B being non-responsive? Why?

Selecting action a_0 will not provide any more information about the responsiveness of aircraft B. If B is responsive then $\dot{h}_B = 0$ and if B is non-responsive then $\dot{h}_B = 0$. In either case aircraft B continues level, so any fluctuation to h is through chance. Thus an exact Bayesian update would give the same probability, in this case $p = 0.5$ as an uniform intial belief was assumed.

An exact Bayesian update

#16 Write a little paragraph about what you learned in this class.

In this class I learned about Markov Processes and how they are used to describe reinforcement learning models. I learned about algorithms to formulate policies for both observable Markov Processes and partially observable processes. The main thing that stuck out to me was the process used to formulate these problems in an accessible manner. This will help me in the future, when I would like to solve similar problems.