Practice Final H

1 The forest and the trees

1. (a) Write the following sentences in FOL using the following predicates:

$inforest(a)$	True if a is in the forest, False otherwise
$tree(a)$	True if a is a tree, False otherwise
$alive(a)$	True if a is alive, False otherwise
$eq(a, b)$	True if a and b are equal, False otherwise

i. (2 points) The forest contains at least two trees.

ii. (2 points) Every tree in the forest is alive.

(b) (2 points) You have two sentences, α and β. What is a strategy for showing that α *does not* entail β?

(c) (3 points) Consider the sentence "every tree is alive." It is not entailed by the sentences from part (a). Prove this using the strategy you outlined in part (b). Hint: This should not be a lot of work.

(d) Prove the following claim:

$$\forall a.\ \mathit{alive}(a) \land \mathit{inforest}(a) \implies \mathit{tree}(a)$$

$$\exists a.\ \mathit{alive}(a) \land \mathit{inforest}(a)$$

entails

$$\exists a.\ \mathit{tree}(a) \land \mathit{inforest}(a)$$

using a proof by refutation.

i. (1 point) When using refutation, what are we proving? Your answer should be of the form "x entails y" where x is a set of FOL sentences and y is either True or False.

ii. (2 points) Convert to CNF. Put your work in the box, and the resulting proof lines in the table.

line #	statement

iii. (3 points) Complete the proof.

line #	statement	line #s used	unification

2 Telescope scheduling (CSP formulations)

You are trying to schedule observations on the space telescope. The telescope has three instruments, each of which can be aimed at a different objective. You need to make a schedule that lasts for a fixed number k of observation time periods. Meanwhile, there are m scientists. Each of them has submitted a list of n requests for observations that they would like to make. (An observation is specified by an objective, an instrument, and a time period). The greedy scientists cannot all be satisfied, so we will try to find a schedule that satisfies the following constraints:

A Two observations from each scientist's list will be made.

B At most one observation per instrument per time slot is scheduled.

C The observations scheduled for a single time slot must be consistent. (Assume that you're given a function that tests for consistency of a set of observations, based on their positioning requirements for the telescope, etc.)

We will formulate this problem as a constraint satisfaction problem (CSP). We will consider 4 formulations of the problem and for each one, ask the same 4 questions.

2. (4 points) The variables are the $3k$ instrument/time slots.

   i. What is the value domain for the variables?

   ii. What is the size of the whole space of assignments (in terms of k, m, and n)?

   iii. Which of the constraints are automatically satisfied because of the formulation? Mark all that apply.
   ◯ A ◯ B ◯ C

   iv. Can all constraints (A, B, and C) be formulated as binary constraints in this formulation? If they can, explain how. If not, provide a counterexample.

3. (4 points) The variables are the k time slots.
   i. What is the value domain for the variables?

   ii. What is the size of the whole space of assignments (in terms of k, m, and n)? (Big O is fine)

   iii. Which of the constraints are automatically satisfied because of the formulation? Mark all that apply.
   ◯ A ◯ B ◯ C

   iv. Can all constraints (A, B, and C) be formulated as binary constraints in this formulation? If they can, explain how. If not, provide a counterexample.

4. (4 points) The variables are the m scientists.
   i. What is the value domain for the variables?

   ii. What is the size of the whole space of assignments (in terms of k, m, and n).

   iii. Which of the constraints are automatically satisfied because of the formulation? Mark all that apply.
   ◯ A ◯ B ◯ C

iv. Can all constraints (A, B, and C) be formulated as binary constraints in this formulation? If they can, explain how. If not, provide a counterexample.

5. (4 points) The variables are the mn scientists' requests.

i. What is the value domain for the variables?

ii. What is the size of the whole space of assignments (in terms of k, m, and n).

iii. Which of the constraints are automatically satisfied because of the formulation? Mark all that apply.

◯ A ◯ B ◯ C

iv. Can all constraints (A, B, and C) be formulated as binary constraints in this formulation? If they can, explain how. If not, provide a counterexample.

3 LineLand

6. Consider a fantastic creature, living on the infinite integer number line $-\infty, \ldots, \infty$. Its state is described by (x, r) where x is the coordinate of its center on the line and r is its radius. So, it occupies the interval $[x - r, x + r]$. It can move around on the line (change x), and depending on where it is and on what actions it takes, it may grow or shrink (change r). Generally, when it moves, it grows, but if it gets into the shrink zone, it will shrink.

   In general, its action can be described with an integer value dx. In any given problem, there is a special "shrink" zone, $[Z_{\text{lo}}, Z_{\text{hi}}]$, which is an interval of the line. The transition model is as follows:

$$T((x,r),dx) = \begin{cases} (x+dx, \max(0, r-1)) & \text{if } Z_{\text{lo}} \leq x \leq Z_{\text{hi}} \\ (x+dx, r+|dx|) & \text{otherwise} \end{cases}$$

   The creature's goal is specified by a goal region $[g_{\text{lo}}, g_{\text{hi}}]$, and the goal is for the *entire* creature (the whole interval $[x - r, x + r]$) to be contained in the goal region, so that

$$g_{\text{lo}} <= x - r \qquad \text{and} \qquad x + r <= g_{\text{hi}} \ .$$

   All actions have a cost of 1.

   We'll consider a specific instance of this type of problem in which:

   - The agent's actions (possible values of dx) are $(-2, -1, 0, 1, 2)$.
   - The initial state, s_0, is $(0, 1)$.
   - The goal region is $[3, 7]$ (so $g_{\text{lo}} = 3$ and $g_{\text{hi}} = 7$).
   - The shrink zone is $[4, 20]$ (so $Z_{\text{lo}} = 4$ and $Z_{\text{hi}} = 20$).

   (a) (6 points) What is an optimal solution to this problem? If there is none, explain why. If there are multiple optimal solutions, any one is fine. Provide the list of actions.

   (b) (5 points) Which of the following are relaxations of this problem, which could be applied for *any initial state, goal, and shrink zone*? In each case, assume that the rules for transitions and goals stay the same except for the stated change. (Mark all that apply)

   - ◯ The creature's shape never grows.
   - ◯ The creature's shape never shrinks.
   - ◯ Only the creature's center needs to be inside the goal.
   - ◯ The creature can move at most distance 1 per step.

◯ The creature can move any number of squares per step.

(c) (5 points) Provide heuristic that is admissible for this problem instance, and has the property that $h(s) > 1$ for at least one state s.

(d) Pat suggests adding a search-pruning rule, so that if uniform cost search (UCS) has visited a state (x, r) and then it reaches a new state (x, r'), if $r' > r$, we ignore the new state and do not add it to the frontier.

i. (3 points) Given that UCS has visited (x, r) before (x, r'), what can we infer about the cost of the lowest-cost path from the root to reach (x, r) compared to the cost of the lowest-cost path from the root to reach (x, r')?

ii. (3 points) Would using Pat's pruning rule prevent UCS from finding an optimal solution to this problem?

(e) (3 points) There's another interpretation of the model we have been studying. Rather than a creature that shrinks and grows, we could think of our agent as a point robot that has set-based uncertainty about its position on the line (initial uncertainty set is s_0), and a non-deterministic transition function that might have any of a set of possible states as its result. We would like to find a conformant plan (a plan that is guaranteed to reach the goal no matter what happens) for this agent, where the goal for the point robot to be anywhere inside the interval $[g_{\text{lo}}, g_{\text{hi}}]$.

Explain what, if anything, we would need to change about this planning problem formulation to find a conformant plan.

4 Expectoration

7. (15 points) Consider a family of discrete MDPs, each with 1000 states. (In the following, use modular arithmetic, so state -1 is the same as state 999, and state 1000 is the same as state 0.)

   - There are two actions, A and B.
   - The reward for entering entering state 0 is $+99$.
   - The reward for entering state 999 is -99.
   - All other rewards are 0.
   - The discount factor is 1.0.

   MDPs in this family are parameterized by an integer $k \in \{0, \ldots, 100\}$, which governs the transition model as follows:

   - Action A: State i transitions with probability $1/(2k+1)$ to each state in $\{i-k+1, \ldots, i+k+1\}$.
   - Action B: State i transitions with probability $1/(2k+1)$ to each state in $\{i-k-1, \ldots, i+k-1\}$.

   So, action A produces an uniform distribution centered at $i+1$ and action B produces an uniform distribution centered at $i-1$, both of width $2k+1$. For example,

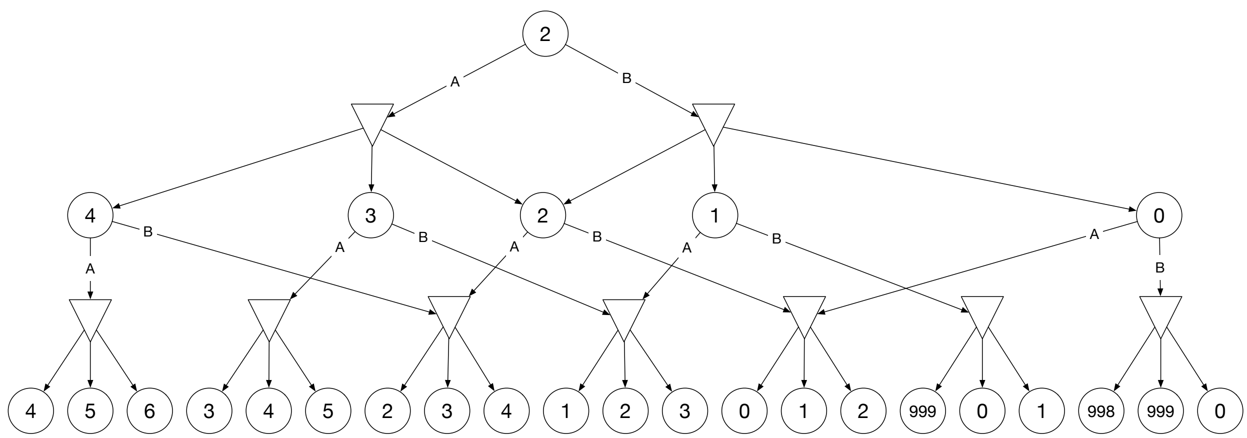
   - When $k=0$, action A moves deterministically one step "up" and action B moves deterministically one step "down".
   - When $k=1$, action A moves with equal probability "up" 2, "up" 1, or stays in the original state; action B moves with equal probability "down" 2, "down" 1, or stays in the original state.
   - When $k=2$, action A moves with equal probability "up" 3, "up" 2, "up" 1, stays in the original state, or moves "down" 1; action B moves with equal probability "down" 3, "down" 2, "down" 1, stays in the original state, or moves "up" 1.

   (a) (2 points) How many leaves are there in the expectimax tree for horizon 3 in the MDP with $k=1$? (It is fine to write an unevaluated expression).

   (b) (2 points) How many leaves are there in the expectimax tree for horizon 3 in the MDP with $k=100$? (It is fine to write an unevaluated expression)

   (c) (2 points) How many distinct leaves are there in the expectimax AODAG (that is, graph in which states that are reachable in the same level of the tree are represented as a single node) for horizon h in the MDP? (Provide your answer in terms of h and k).

(d) (3 points) The following figure illustrates the horizon-2 expectimax AODAG starting at state 2 for $k = 1$. (Note that we did not coalesce the leaf nodes in this figure, because it becomes very difficult to read.)



Provide the horizon-2 Q values $Q(2, A)$ and $Q(2, B)$.

(e) (2 points) What is the optimal horizon 2 policy for the $k = 1$ MDP, starting at state 2?

(f) (2 points) Consider the case of $k = 100$ and horizon 10. We might prefer to use a sampling based method (sparse sampling or MCTS). Explain why.

(g) (2 points) Treesa wants to use sparse sampling with 50 samples at each node. In the case when $k = 100$, horizon is just 1 and the state is 101, what difficulty might Treesa's method face?

5 Charge!

8. (14 points) You may recall the example POMDP from homework 9. If not, here's all the info about it.

> Our robot has lost its charger! The robot lives in a world with $K = 3$ locations but it needs to have the charger in location $L0$ (where the power outlet is) in order to plug it in and charge. The robot doesn't have any uncertainty about its own location (and we won't model that) but it does have uncertainty about the location of the charger. When the robot looks in location `loc`, by executing a `look(loc)` action, it will receive an observation in $\{0, 1\}$.
>
> - If the charger is in the location `loc`, it will get observation 1 with probability 0.9.
> - If the charger is not in location `loc`, it will get observation 1 with probability 0.4.
>
> The robot can also try to execute an action `move(loc1, loc2)` (which moves the charger):
>
> - If the charger is actually in `loc1`, then with probability 0.8, the charger will be moved to `loc2`, otherwise, it will stay in `loc1`.
> - If the charger is not in `loc1`, then nothing will happen.
>
> Because of the constraints of the robot's arm, we can't do all possible move actions. Specifically, the only valid movements are `move(0, 1)`, `move(1, 2)`, and `move(2, 0)`. The robot has two more actions, `charge` and `nop`:
>
> - If it executes the `charge` action when the charger is in location 0, then it gets a reward of 10 and the game terminates.
> - If it executes the `charge` action in any other state, it gets reward -100 and the game terminates.
> - The `nop` action has reward 0, does not supply a useful observation, and does not affect the world state.
>
> It gets reward -0.1 for every look action and -0.5 for every move action in all other states.
> We are not discounting: $\gamma = 1$.
> For actions that don't yield useful information about the environment (`move`, `charge`, and `nop`), we assume that they get observation 0 with probability 1, independent of the state.

In some belief states, the optimal horizon-2 action was to look in location 0 and in others it was to look in location 1. We are going to explore this asymmetry.

(a) (2 points) Starting with belief $b = (.9, .1, 0)$ over the charger being in locations `(loc0, loc1, loc2)`, what is the likelihood of seeing the object (getting observation 1) if you look in location 0?

(b) (3 points) What is the posterior belief (distribution over all three possible charger locations) if you look in location 0 and see the object?

(c) (3 points) What is the posterior belief (distribution over all three charger locations) if you look in location 0 and don't see the object?

In addition to the values you computed above, we have computed some additional ones:

- The likelihood of seeing the object if you look in location 1 is .45.
- The posterior if you look in location 1 and see the object is $(.8, .2, 0)$.
- The posterior if you look in location 1 and don't see the object is $(.98, .02, 0)$

(d) (3 points) At horizon 1, if $b(\texttt{loc0}) >\approx .91$, then it is worthwhile to charge, otherwise not. What is the optimal horizon-2 policy tree if you start with belief $b = (.9, .1, 0)$?

(e) (3 points) What is this policy tree's expected value in $b = (.9, .1, 0)$?

6 The leftovers are sus

9. You're hungry at night and you open your fridge to find some questionable-looking leftovers. You're not exactly sure if it's a good idea to eat them.

   As a student of 6.4110, you decide to model this problem as a POMDP and try to solve it to make your decision for you!

   Specifically, you use an HMM to model the state of the leftovers over time. The state space is $\{good, mid, toxic\}$. We denote a state via S_t, where t is in days ($t = 1, 2, 3, \ldots$).

   Initially (i.e., when you open the fridge), you're sure they're good: $P(S_0 = good) = 1$.

   There are four actions available: $\{eat, wait, boil,$ and $sniff\}$. The *eat* and *sniff* actions do not affect the state (i.e. $S_t = S_{t+1}$ if either of these two actions is executed).

   The state transition model for the *wait* action is:

		S_{t+1}		
		good	mid	toxic
	good	0.5	0.5	0
S_t	mid	0	0.5	0.5
	toxic	0	0	1

   The state transition model for the *boil* action is:

		S_{t+1}		
		good	mid	toxic
	good	0	1	0
S_t	mid	0	1	0
	toxic	0	0.9	0.1

   The reward function is:

Action	**good**	**mid**	**toxic**
wait	0	0	0
boil	-1	-1	-1
eat	$+10$	$+5$	-100
sniff	-1	-1	-5

   The only action that yields any observations is *sniff*. You observe "fine" or "bad".

$$P(\text{"fine"} \mid good) = 1$$
$$P(\text{"fine"} \mid mid) = .9$$
$$P(\text{"fine"} \mid toxic) = 0.1$$

(a) (4 points) What is your belief about the leftovers if you are initially sure they're good, but then wait two days? Provide a numerical answer.

(b) (3 points) If you start with belief $(0.3, 0.3, 0.4)$ (corresponding to $(P(good), P(mid), P(toxic))$) and you sniff and observe "fine", what is your new belief? Provide a numerical answer.

(c) (3 points) If you start with belief $(0.5, 0.5, 0)$ (corresponding to $(P(good), P(mid), P(toxic))$) and you boil your leftovers, what is your new belief? Provide a numerical answer.

7 To eat or not to eat

10. Now, let's think about what to do! Consider the problem setup from Question 6 above **(make sure to read everything before part(a) in Question 6 carefully!)**

    (a) (2 points) What is the alpha vector corresponding to the the horizon-1 policy of taking the *sniff* action? Specify it as a vector of numbers corresponding to states (good, mid, toxic).

    (b) (5 points) Consider the policy tree: '*sniff*(*eat*, *wait*)'. Recall that this means that we take the action *sniff*, and eat the leftovers if the resulting observation is 'fine', and we *wait* if the resulting observation is 'bad'.

    What is the alpha vector representing the value of this policy tree? Specify it as a vector of numbers corresponding to states (good, mid, toxic).

(c) (3 points) Here are the alpha vectors for three policy trees (assuming the process terminates after an *eat* action):

$$\begin{aligned} eat &: [10.0, 5.0, -100.0] \\ boil(eat, eat) &: [4.0, 4.0, -6.5] \\ sniff(eat, boil(eat, eat)) &: [9.0, 3.9, -20.85] \end{aligned}$$

Each vector is formatted as the alpha values for [good, mid, toxic].
Provide a belief state (formatted as $(P(good), P(mid), P(toxic))$) for which the first strategy is the best.

(d) (3 points) If the three policy trees in the previous question were your only options, what is the first action you should take in belief state $[0.333, 0.333, 0.334]$?

8 Contagion!

11. Consider the problem setup from Question 6 above **(make sure to read everything before part(a) in Question 6 carefully!)**

Now, suppose you have four other housemates (call them V_1, V_2, V_3, and V_4), and they have leftovers in the fridge, as well as you (V_0). The leftovers are arranged in the fridge in a ring, in order V_0, V_1, V_2, V_3, and V_4, so that V_4 is adjacent to both V_3 and V_0. Let's model the contagion as a factor graph. Each variable node in this graph will correspond to a person's leftovers (i.e., we will have a separate variable node V_i for $i \in \{0, 1, 2, 3, 4\}$). Each variable can either be in the state (good, mid, toxic). Between each neighboring pair of leftovers is a factor:

		V_{i+1}		
		good	mid	toxic
	good	20	1	1
V_i	mid	1	10	1
	toxic	1	1	10

(a) (2 points) Without any further evidence, what is the maximum likelihood assignment? If there is a tie, list all of the maxima.

(b) (3 points) You decide to use variable elimination to determine the marginal distribution on V_0 (the state of your leftovers). Draw the factor graph that results after eliminating variable V_1.

(c) (3 points) Let's call the new factor created as a result of the elimination ψ_{new}. An (incomplete) factor table for this new factor is

ψ_{new}	good	mid	toxic
good	???	31	31
mid	31	???	21
toxic	31	21	???

Provide the missing values $\psi_{\text{new}}(\text{good, good})$, $\psi_{\text{new}}(\text{mid, mid})$, and $\psi_{\text{new}}(\text{toxic, toxic})$).

Roommate V_2 eats their leftovers and dies. They were definitely toxic. Conditioned on this information, you are interested in figuring out the state of your leftovers.

(d) (3 points) Returning to our original factor graph, but taking this new information into account, can you use sum-product belief propagation to figure this out? If so, draw the factor graph you would operate on, and indicate any factors that are different from the one we began with. If not, explain why we cannot use sum-product in this case.

- ◯ We can use sum-product
- ◯ We cannot use sum-product