Proof for Equivalency

1 Properties

Here we assume the MDP is a 2 action MDP i.e. the action at any state in any policy is either 0 or 1. For any policy p, of all the states belonging to it, the states, switching will improve the policy is called **Improvement Set** of Policy p. It is represented by $IS(p)$. Since we are using Howard's Policy Iteration, all the states in $IS(p)$ are switched. The improvement set can also be represented as a n sized bit-vector with it i^{th} element as 1 if $s_i \in IS(p)$ and as 0 if $s_i \notin IS(p)$.

Any new policy we get by switching only in $IS(p)$(not necessarily Howard's switch) dominates current policy p. We are calling this set **Improved Policy Set**, represented by $IPS(p)$. Also any policy we get by switching only in $S \setminus IS(p)$ is dominated by p itself, where S is the set of all states. We are calling the set **Deproved Policy Set**, represented by $DPS(p)$.

1.1 Subset Condition

Let $P =< p_1, p_2, p_3....p_m >$ be the sequence of policies we get from policy iteration. So, for any $i, j \in \{1, 2...m\}$ if $i < j$ then, $IS(p_i) \not\subseteq IS(p_j)$. This can be written as.

$$\forall i < j(IS(p_i) \not\subseteq IS(p_j))$$

1.2 Intersection Condition

Let $P =< p_1, p_2, p_3....p_m >$ be the sequence of policies we get from policy iteration. So, for any $i, j \in \{1, 2...m\}$ if $i < j$ then, $DPS(p_i)$, should not have any element from $IPS(p_j)$. This can be written as.

$$\forall i < j(DPS(p_i) \cap IPS(p_j) = \emptyset)$$

1.3 OR Matrix Condition

A sequence of m policy over n state MDP can be represented as $n \times m$ matrix. Such a matrix A is called Order Regular(OR) Matrix if

$$\forall i < j \in [m], \exists k \in [n], (A_{i,k} \neq A_{i+1,k} = A_{j,k} = A_{j+1,k})$$

Here $[m]$ is the set $\{1, 2, 3...m\}$. $A_{m+1,k} = A_{m,k}$, because no $m+1$ exists in the matrix.

2 Proofs of Equivalency

Here we prove that Intersection Condition is equivalent to OR Matrix Condition and both of these imply the subset condition.

2.1 Intersection Condition = OR Matrix Condition

For any $i < j, DPS(p_i) \cap IPS(p_j) \neq \emptyset$ occurs iff for every $k \in [n]$ one of the following is true :

- $p_i(s_k) = p_j(s_k)$ **OR**
- $s_k \in IS(p_j)$ **OR**
- $s_k \in S \setminus IS(p_i)$

As shown in figure 1.

If some state s_k of policy p_j is in Improvement set of p_j, then due to Howard's switching s_k will not have the same action in p_{j+1}. Similarly, if s_k of policy p_i is in Deprovement set of p_i, then s_k will have the same action in p_{i+1}. The above conditions can be written in OR matrix format as

- $A_{i,k} = A_{j,k}$ **OR**
- $A_{j,k} \neq A_{j+1,k}$ **OR**
- $A_{i,k} = A_{i+1,k}$

Hence,

$$DPS(p_i) \cap IPS(p_j) \neq \emptyset \Leftrightarrow (\forall k, A_{i,k} = A_{j,k} \vee A_{j,k} \neq A_{j+1,k} \vee A_{i,k} = A_{i+1,k})$$

Taking Contapositive,

$$DPS(p_i) \cap IPS(p_j) = \emptyset \Leftrightarrow (\exists k, A_{i,k} \neq A_{j,k} \wedge A_{j,k} = A_{j+1,k} \wedge A_{i,k} \neq A_{i+1,k})$$

Or,

$$\forall i < j, (DPS(p_i) \cap IPS(p_j) = \emptyset \Leftrightarrow A_{i,k} \neq A_{i+1,k} = A_{j,k} = A_{j+1,k})$$

This proves that Intersection Condition is equivalent to OR Matrix Condition.

2.2 OR matrix Condition $\Rightarrow$ Subset Condition

Since for some $k \in [n]$ $A_{i,k} \neq A_{i+1,k}$ and $A_{j,k} = A_{j+1,k}$ this means $s_k \in IS(p_i)$ and $s_k \notin IS(p_j)$. Hence, $IS(p_i) \not\subseteq IS(p_j)$. This shows that OR matrix Condition $\Rightarrow$ Subset Condition.

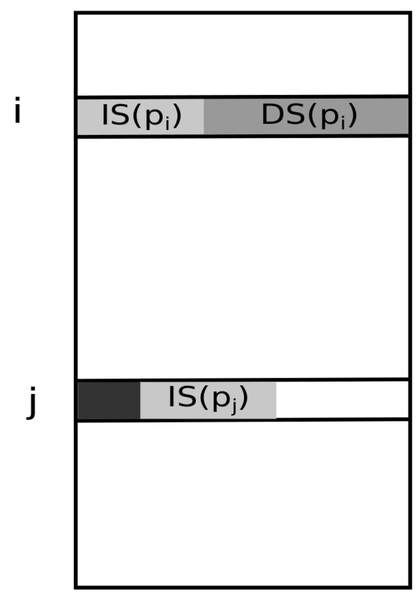

Figure 1: Light grey region shows Improvement set of policies, medium grey region shows Deprovement set of region of policy i. The DSP of p_i will have fixed action values of states in its IS. Similarly The ISP of p_j will have fixed action values of states not in its IS.Dark grey region is the intersection of $IS(p_i)$ and $DS(p_j)$.So for the intersection of ISP and DSP to be non empty, the Dark grey region should have same values in p_i and p_j as all other places can be matched (it can be changed either in p_i or p_j).