A. Proofs

First we state some lemmas that will be used in proving the main theorems.

A.1. Proper MDP Models

Example 1. Let $J_{\text{MaxEnt}}$ be the MaxEntRL objective. Then, the MaxEnt MDP model $\mathcal{P}_{\text{MDP}}[R; d, T, J_{\text{MaxEnt}}]$ is proper.

Proof. Let $R$ be any set of reward functions. We need to show that $\forall r, \hat{r} \in R, r \cong_{\tau} \hat{r} \Rightarrow p_r = p_{\hat{r}}$. If $r \cong_{\tau} \hat{r}$, then $r$ and $\hat{r}$ have trajectory level rewards shifted by a constant, i.e., for all $x_0 \in X^0$ there exists a constant $c_{x_0}$ such that $\forall \tau \in \Omega[x_0, d, T], \hat{r}(\tau) = r(\tau) + c_{x_0}$. It suffices to show that the optimal policies for $r, \hat{r}$ are the same. For any policy family $\Pi$,

$$\arg\max_{\pi \in \Pi} \mathbb{E}_{r \sim \pi}[\hat{r}(\tau)] + \mathcal{H}(\pi) = \arg\max_{\pi \in \Pi} \left( \sum_{x_0 \in X^0} \sum_{\tau \in \Omega[x_0, d, T]} p(\tau; \pi)\hat{r}(\tau) \right) + \mathcal{H}(\pi)$$

$$= \arg\max_{\pi \in \Pi} \left( \sum_{x_0 \in X^0} \sum_{\tau \in \Omega[x_0, d, T]} p(\tau; \pi)(r(\tau) + c_{x_0}) \right) + \mathcal{H}(\pi)$$

$$= \arg\max_{\pi \in \Pi} \sum_{x_0 \in X^0} \sum_{\tau \in \Omega[x_0, d, T]} p(\tau; \pi)r(\tau) + \left( \sum_{x_0 \in X^0} P_0(x_0)c_{x_0} \right) + \mathcal{H}(\pi)$$

$$= \arg\max_{\pi \in \Pi} \sum_{x_0 \in X^0} \sum_{\tau \in \Omega[x_0, d, T]} p(\tau; \pi)r(\tau) + \sum_{x_0 \in X^0} P_0(x_0)c_{x_0} + \mathcal{H}(\pi)$$

where $\mathcal{H}(\pi) := \mathbb{E}_\pi[-\sum_{t=0}^T \gamma^t \log \pi(a_t|s_t)]$ is the $\gamma$-discounted causal entropy. The last step holds since $\mathbb{E}_{x_0 \sim P_0}[c_{x_0}]$ is constant with respect to $\pi$. \hfill $\square$

A.2. Weak Identifiability

Lemma 1. For all reward families $R, r, \hat{r} \in R$, and any $(d, T)$, $(r \cong_{x,a} \hat{r}) \Rightarrow (r \cong_{\tau} \hat{r})$

Proof. Let $r, \hat{r} \in R$ be rewards such that $r \cong_{x,a} \hat{r}$. For all $\tau, \tau' \in \Omega[d, T]$, where $\tau = (x_t, a_t)_{0 \leq t \leq T}, \tau' = (x'_t, a'_t)_{0 \leq t \leq T}$,

$$\hat{r}(\tau) - r(\tau) = \sum_{t=0}^T \gamma^t(\hat{r}(x_t, a_t) - r(x_t, a_t))$$

$$= \sum_{t=0}^T \gamma^t(\hat{r}(x'_t, a'_t) - r(x'_t, a'_t))$$

$$= \hat{r}(\tau') - r(\tau')$$

where $6 \Rightarrow 7$ holds since for all $0 \leq t \leq T$, $(x_t, a_t), (x'_t, a'_t) \in X \times A$ and so $\hat{r}(x_t, a_t) - r(x_t, a_t) = \hat{r}(x'_t, a'_t) - r(x'_t, a'_t)$. Thus, $r \cong_{\tau} \hat{r}$ \hfill $\square$

Proposition 1. A proper MDP model is strongly identifiable only if it is weakly identifiable

Proof. We prove the contrapositive: if a proper MDP model is not weakly identifiable it is also not strongly identifiable. Let $\mathcal{P}_{\text{MDP}}[R; d, T, J]$ be a proper MDP model that is not weakly identifiable. Since the model is not weakly identifiable, there exists $r, \hat{r} \in R$ such that either $(r \cong_{\tau} \hat{r}$ and $p_r \neq p_{\hat{r}})$ or $(r \ncong_{\tau} \hat{r}$ and $p_r = p_{\hat{r}}).$ Since the model is proper the former cannot be true. Thus it must be that there exists $r, \hat{r} \in R$ such that $r \ncong_{\tau} \hat{r}$ and $p_r = p_{\hat{r}}$. Then, by the contrapositive of Lemma 1, $r \ncong_{x,a} \hat{r}$. Thus, $p_r = p_{\hat{r}} \neq r \cong_{x,a} \hat{r} \Rightarrow \mathcal{P}_{\text{MDP}}[R; d, T, J]$ is not strongly identifiable as desired. \hfill $\square$
We prove the contrapositive. Suppose there exists a MaxEnt MDP model and a reward function such that the solution to the linear system is non-unique. This implies that the trajectory distribution of the MaxEnt optimal policy is not unique. Thus, one solution to the linear system is not the unique solution.

\[ p_r(\tau) = \frac{e^{r(\tau)}}{Z_r} \]

where \( Z_r = \int_{\Omega[d,T]} e^{r(\tau')} d\tau' \) is the partition function. Then, for all \( r \in \Omega[d,T] \)

\[ p_r(\tau) = p_{\hat{r}}(\tau) \]

\[ \log p_r(\tau) = \log p_{\hat{r}}(\tau) \]

\[ r(\tau) - \log Z_r = \hat{r}(\tau) - \log Z_{\hat{r}} \]

\[ r(\tau) = \hat{r}(\tau) + \log \frac{Z_{\hat{r}}}{Z_r} \]

Since, \( \log \frac{Z_{\hat{r}}}{Z_r} \) is a constant w.r.t \( r \), we have \( r \equiv_{\tau} \hat{r} \) as desired.

**Proof.** Let \( P_{\text{MDP}}[R; d, T, J] \) be an MDP model that is weakly identifiable. Then, it is strongly identifiable if and only if for all \( r, \hat{r} \in R \), \( (r \equiv_{\tau} \hat{r}) \Rightarrow (r \equiv_{x,a} \hat{r}) \).

We prove the contrapositive. Suppose there exists \( r, \hat{r} \in R \) such that \( r \equiv_{\tau} \hat{r} \) but \( r \not\equiv_{x,a} \hat{r} \). By weak identifiability, \( (r \equiv_{\tau} \hat{r}) \Rightarrow (p_r = p_{\hat{r}}) \), so \((r \equiv_{x,a} \hat{r}) \Rightarrow (p_r = p_{\hat{r}}) \). Thus, \( P_{\text{MDP}}[R; d, J, T] \) is not strongly identifiable.

**Corollary 1.** Let \( P_{\text{MDP}}[R; d, T, J] \) be an MDP model that is weakly identifiable, \( R \) be the set of all rewards, \( |X^0| = 1 \), and \( \gamma = 1 \). Then, it is strongly identifiable if and only if \( \text{rank}(A[d,T]) = |X \times A| \)

**Proof.** Let \( P_{\text{MDP}} \) be weakly identifiable, \( \gamma = 1 \).

(Sufficiency) We seek to show that if \( \text{rank}(A[d,T]) = |X \times A| \), then \( P_{\text{MDP}} \) is strongly identifiable. By Proposition 2, it suffices to show that for all \( r, \hat{r} \in R \), \( (r \equiv_{\tau} \hat{r}) \Rightarrow (r \equiv_{x,a} \hat{r}) \), i.e trajectory equivalence implies state-action equivalence.

Since \( A[d,T] \) is full rank, the solution to the linear system \( A[d,T]r_{x,a} = r_x \) is unique for any \( r_x \). Let \( r_x, \hat{r}_x \) be two trajectory equivalent rewards such that \( r_x = \hat{r}_x + c \) for some constant vector \( c = (c, ..., c) \in \mathbb{R}^{|\Omega[d,T]|} \). Then,

\[ A[d,T]r_{x,a} - A[d,T]\hat{r}_{x,a} = r_x - \hat{r}_x \]

\[ A[d,T](r_{x,a} - \hat{r}_{x,a}) = c \]

Since \( A[d,T] \) is a trajectory matrix, \( \forall i, \sum_j A_{ij}d[T] = T + 1 \), i.e all feasible trajectories are of the same length and hence visit the same number of (not necessarily distinct) nodes. Thus, one solution to \( A[d,T](r_{x,a} - \hat{r}_{x,a}) = c \) is to let...
r_{x,a} - \hat{r}_{x,a} = (\frac{c}{T+1}, \ldots, \frac{c}{T+1}). In fact, since \( A[d,T] \) is full rank, \((\frac{c}{T+1}, \ldots, \frac{c}{T+1})\) is the only solution and thus \( r_{x,a}, \hat{r}_{x,a} \) are trajectory equivalent, \( \forall x, a \in \mathcal{X} \times \mathcal{A}, r(x,a) = \hat{r}(x,a) + \frac{c}{T+1} \) implying \( r \equiv_{x,a} \hat{r} \).

(Necessity) We show that if \( \mathcal{P}_{\text{MDP}} \) be strongly identifiable, then \( \text{rank}(A[d,T]) = |\mathcal{X} \times \mathcal{A}|. \) By strong identifiability, \( \forall r, \hat{r} \in \mathcal{R}, (r \equiv_{r} \hat{r}) \implies (r \equiv_{x,a} \hat{r}) \) and thus general solutions to
\[
A[d,T]r_{x,a} = r_{\tau}
\]
must only be constant shifts of a particular solution. Equivalently, \( \ker(A[d,T]) \) must only contain constant vectors. We then claim that in fact \( \ker(A[d,T]) \) only contains the zero vector and thus \( A[d,T] \) is full rank.

Suppose for contradiction that \( e \in \ker(A[d,T]) \) for some non-zero constant vector \( e \). Then, for any scalar \( k \in \mathbb{R} \), \( ke \in \ker(A[d,T]) \). Thus the kernel must contain all constant vectors. Pick a strictly positive constant vector \( c^+ = (c^+, \ldots, c^+) \) where \( c^+ > 0 \). Then, \( c^+ \in \ker(A[d,T]) \) \( \implies A[d,T]c^+ = 0 \), so \( \forall i, \sum_j A_{ij}[d,T]c^+ = c^+ \sum_j A_{ij}[d,T] = 0 \). Since \( A[d,T] \) is a trajectory (path) matrix, its entries represent visitation counts of a state-action pair and thus are all non-negative, i.e \( \forall i, j, A_{ij}[d,T] \geq 0 \). Therefore, \( \forall i, \sum_j A_{ij}[d,T] = 0 \) \( \implies \forall i, j, A_{ij}[d,T] = 0 \), so \( A[d,T] \) is the zero-matrix. Then, \( \ker(A[d,T]) = \mathbb{R}^{1 \times |\mathcal{A}|} \) which contradicts strong identifiability. Therefore, \( \ker(A[d,T]) \) can only contain the zero vector and \( A[d,T] \) is full rank, i.e \( \text{rank}(A[d,T]) = |\mathcal{X} \times \mathcal{A}|. \)

A.3. Properties of Domain Graphs

Lemma 2. Let \( G_d = (V_d, E_d, V_0) \) be a domain graph.

1. (Commutative) For all \( V \subseteq V_d \) and \( t, t' \geq 0 \), \( L_{t'}(L_t(V)) = L_{t+t'}(V) \)

2. (Monotonic) For all \( V, V' \subseteq V_d \) such that \( V \subseteq V' \) and \( t \geq 0 \), \( L_t(V) \subseteq L_t(V') \)

Proof. • (Commutative) We first prove that \( L_{t'}(L_0(V)) \subseteq L_{t+t'}(V) \). Let \( v \in L_{t'}(L_0(V)) \), then by Definition 6, \( \exists \zeta' = (v'_0)_{0 \leq t' \leq t'} \) such that \( v'_0 = v \) and \( v'_0 \in L_0(V) \). Then, \( v \in L_{t+t'}(V) \) since there exists a path \( \zeta \oplus \zeta'_t = (v_0, \ldots, v_t, v'_t, \ldots, v'_{t'}) \) such that \( v'_t = v \) and \( v_0 \in V \).

Next, we prove that \( L_{t+t'}(V) \subseteq L_{t'}(L_t(V)) \). If \( v \in L_{t+t'}(V) \) then \( \exists \zeta'' = (v''_0)_{0 \leq t+t' \leq t'} \) such that \( v''_0 = v \) and \( v''_0 \in V \). Then, there exists paths \( \zeta = (v_0, \ldots, v_t, v'_t, \ldots, v'_{t+t'}) \) and \( \zeta' = (v'_0, \ldots, v'_t, \ldots, v'_{t+t'}) \) which can be joined to form \( \zeta'' \). Therefore, \( v \in L_{t'}(L_t(V)) \) since there exists a path \( \zeta ' \) such that \( v'_0 = v \) and \( v_0 \in V \). Since \( V \subseteq V' \), \( v_0 \in V' \) as well. Therefore, \( v \in L_t(V') \).

• (Monotonic) Let \( V, V' \subseteq V_d \) satisfy \( V \subseteq V' \). If \( v \in L_t(V) \), then by Definition 6, \( \exists \zeta = (v_i)_{0 \leq i \leq t} \) such that \( v_t = v \) and \( v_0 \in V \). Since \( V \subseteq V' \), \( v_0 \in V' \) as well. Therefore, \( v \in L_t(V') \).

Lemma 3. If \( G_d \) is coverable, then \( \cup(x,a) \in \mathcal{X} \times \mathcal{A} \text{supp}(P(\cdot|x,a)) = \mathcal{X} \)

Proof. Since \( G_d \) is coverable, there exists \( v \in V_0^d \) and \( t \geq 0 \) such that \( L_t(v) = V_d \). If \( G_d \) is 0-coverable, i.e \( L_0(v) = \{v\} = V_0 = \mathcal{X} \times \mathcal{A} \), then \( |\mathcal{X} \times \mathcal{A}| = 1 \) and thus \( \text{supp}(P(\cdot|x,a)) = \{v\} = \mathcal{X} \). For \( t \geq 1 \), since \( L_t(v) = L_1(L_{t-1}(v)) = V_d \) and \( L_{t-1}(v) \subseteq V_d \), by Lemma 2 monotonicity, we have \( L_1(L_{t-1}(v)) = V_d \subseteq L_1(V_0) \). Since \( L_1(V_0) \subseteq V_d \), it must be that \( L_1(V_d) = V_d = \mathcal{X} \times \mathcal{A} \). By definition of layers, \( L_1(V_d) = (\cup(x,a) \in \mathcal{X} \times \mathcal{A} \text{supp}(P(\cdot|x,a))) \times \mathcal{A} \) and thus \( \cup(x,a) \in \mathcal{X} \times \mathcal{A} \text{supp}(P(\cdot|x,a)) = \mathcal{X} \).

Lemma 4. Let \( G_d \) be a domain graph and \( v \in V_d \) be \( t \)-covering. Then for all \( t' \geq t \), \( L_{t'}(v) = V_d \).

Proof. We prove by induction.

• Base \( t' = t \): trivially holds since \( L_t(v) = V_d \) by definition of a covering vertex.

• For \( t' \geq t \): \( L_{t'}(v) = V_d \Rightarrow L_{t'+1}(v) = V_d \).
\[
L_{t'+1}(v) = L_1(L_{t'}(v)) = L_1(V_d)
\]
\[
L_t(v) = L_1(L_{t-1}(v)) = V_d \quad \text{and} \quad L_{t'-1}(v) \subseteq V_d, \quad \text{we have that} \quad L_1(L_{t-1}(v)) = V_d \subseteq L_1(V_d) \quad \text{by Lemma 2 monotonicity.}
\]
Since \( L_1(V_d) \subseteq V_d \), it must be that \( L_1(V_d) = V_d \).
Proposition 3. Let $G_d$ be strongly connected. Then, $G_d$ is aperiodic if and only if it is coverable.

Proof. (aperiodic $\Rightarrow$ coverable) If $G_d$ is aperiodic, there exists two cycles $C = (v_i)_{0 \leq i \leq k}$, $C' = (v'_i)_{0 \leq i \leq k'}$ of coprime length $k, k'$ for any $v \in V^0_d$ and any destination vertex $\tilde{v} \in V_d$ consider paths that start from $v$, reaches $v_0$ via a shortest path $\zeta^v_{v_0}$, loops $n$ times around cycle $C'$ back to $v_0$, reaches $v'_0$ via a shortest path $\zeta^v_{v'_0}$, loops $n'$ times around cycle $C''$ back to $v'_0$, and finally reaches $\tilde{v}$ via a shortest path $\zeta^v_{v'_0} \tilde{v}$, i.e.

$$\zeta^v_{v_0} = \zeta^v_{v_0} + n \cdot C_d + \zeta^v_{v_0} + n' \cdot C'_d + \zeta^v_{v'_0} \tilde{v}$$

The paths $\zeta^v_{v_0}, \zeta^v_{v'_0}, \zeta^v_{v'_0} \tilde{v}$ exist by strong connectivity of $G_d$. We let $|\zeta|$ denote the length of a path. Then,

$$|\zeta^v_{v_0}| = nk + nk' + |\zeta^v_{v'_0}| + |\zeta^v_{v_0} + \zeta^v_{v'_0} \tilde{v}|$$

Since $k, k'$ are coprime, for all $|\zeta^v_{v_0}| \geq (k - 1)(k' - 1) + |\zeta^v_{v_0} + \zeta^v_{v'_0} \tilde{v}|$, there exists $n, n'$ such that Eq. 8 holds. (Corollary 2 of Denardo (1977)) Furthermore, since $|\zeta^v_{v_0}|, |\zeta^v_{v_0} + \zeta^v_{v'_0} \tilde{v}| \leq |V_d|$ since they are shortest paths. Thus, for any destination vertex $\tilde{v} \in V_d$ and all lengths $t \geq (k - 1)(k' - 1) + 3|V_d|$, there exists a path $\zeta^v_{v_0} \tilde{v}$ such that $|\zeta^v_{v_0}| = T$. Therefore, $G_d$ is coverable.

(coverable $\Rightarrow$ aperiodic) If $G_d$ is coverable, there exists $v \in V^0_d$ and $t \geq 0$ such that $L_t(v) = V_d$. If $t = 0$, then $V_d = \{v\}$ and there must be an edge $(v, v) \in E_d$. Therefore, there exists cycles $(v, v), (v, v, v)$ which have coprime lengths 1 and 2, respectively. For $t \geq 1$, by Lemma 4, $L_{t+1}(v) = V_d$. Since $v \in L_t(v)$ and $v \in L_{t+1}(v)$, there exists cycles of coprime length $t, t + 1$ that start and end at $v$. Thus, $G_d$ is aperiodic.

\[\square\]

A.4. Strong Identifiability

Theorem 2. (Strong Identification Condition) For all $(d, r, T, J)$ such that the MDP model $P_{MDP}[R; d, T, J]$ is proper and $G_d$ is strongly connected,

- (Sufficiency) $P_{MDP}[R; d, T, J]$ is weakly identifiable, $G_d$ is $T_0$-coverable, and $T \geq 2T_0 \Rightarrow P_{MDP}[R; d, T, J]$ is strongly identifiable
- (Necessity) $P_{MDP}[R; d, T, J]$ is strongly identifiable $\Rightarrow P_{MDP}[R; d, T, J]$ is weakly identifiable, $G_d$ is coverable.

Proof. Let $P_{MDP}[R; d, T, J]$ be proper and $G_d$ be strongly connected.

- (Sufficiency) Let $P_{MDP}[R; d, T, J]$ be proper and weakly identifiable, $G_d$ be strongly connected and $T_0$-covering, and $T \geq 2T_0$. By Proposition 2 it suffices to show that

$$\forall x \in A^0, \forall \tau, \tau' \in \Omega[x, d, T], \tilde{r}(\tau) - r(\tau') = \tilde{r}(\tau') - r(\tau')$$

In the language of domain graphs, this statement translates to:

$$\forall v \in V^0_d, v, \zeta, \zeta' \in Z[v, d, T], \tilde{r}(\zeta) - r(\zeta) = \tilde{r}(\zeta') - r(\zeta') \Rightarrow \forall v, v' \in V_d, \tilde{r}(v) - r(v) = \tilde{r}(v') - r(v')$$

Let $r, \tilde{r}$ be any two rewards such that, $\forall v \in V^0_d, \forall \zeta, \zeta' \in Z[v, d, T], \tilde{r}(\zeta) - r(\zeta) = \tilde{r}(\zeta') - r(\zeta')$ or equivalently $\tilde{r}(\zeta) - r(\zeta') = r(\zeta) - r(\zeta')$. Let $v^0_0 \in V^0_d$ be any vertex that is $T_0$-cover and for any integer $t \geq 0$ let $H_t$ be the statement that,

$$\forall v, v' \in L_t(v^0_0), \tilde{r}(v') - \tilde{r}(v) = r(v') - r(v)$$

Since $v^0_0$ is $T_0$-covering, we have that $L_{T_0}(v^0_0) = V_d$, so it suffices to prove $H_{T_0}$. We prove by strong induction.

$H_0$: Trivially true, since $L_0(v^0_0) = \{v^0_0\}$ only has one element.
Thus, by strong induction

**Proof.**

By definition, 

\[ r(v_{t+1}) - r(v_t) = r(v_{t+1}) - r(v_t) + r(v_{t+1}) - r(v_{t+1}) + r(v_{t+1}) - r(v_{t+1}) + r(v_{t+1}) - r(v_{t+1}) \]

since for all \( 0 \leq t' < t \), \( v_{t-v_{t'}} \in L_{t-v_{t'}}(v_0) \) and by the inductive hypothesis \( H_{<t} \), it holds that, for all \( 0 \leq t' \leq t \), \( r(v_{t-v_{t'}}) - r(v_{t-v_{t'}}) = r(v_{t-v_{t'}}) - r(v_{t-v_{t'}}) \).

By definition, \( Z_t \subseteq Z[v_0, d, T] \), and thus by weak identifiability, \( r(v) - r(v') = r(v) - r(v') \). Combining with Eq. 9, 10, 11, we get that for all \( v, v' \in \{ v \mid \exists (v_i)_{0 \leq i \leq T} \in Z_t \ s.t. v = v_T \} = L_t(v_0) \),

\[ r(v_T) - r(v_T') = r(v_T) - r(v_T') \]

Thus, by strong induction \( H_t \) is true for \( 0 \leq t \leq T_0 \), which includes \( H_{T_0} \).

• (Necessity) Next we prove necessity. To do so, we will first prove some useful properties of layer sequences.

**Lemma 5.** Let \( G_d \) be strongly connected. Then, for all \( v, v' \in V_d \), there exists \( t \geq 1 \) such that \( v' \in L_t(v) \).

**Proof.** Pick any \( v, v' \in V_d \). Since \( G_d \) is strongly connected, there exists a path \( \zeta \) of length \( |\zeta| \geq 1 \) between \( v, v' \). Thus \( v' \in L_{|\zeta|}(v) \).

**Lemma 6.** Let \( G_d \) be strongly connected. Then for all \( v, v' \in V_d \) and \( T \geq 0 \), there exists \( t \geq T \) such that \( v' \in L_t(v) \).

**Proof.** If \( v' \in L_T(v) \), then we are done. If \( v' \notin L_T(v) \), then choose any vertex \( v_T \in L_T(v) \). There exists a path \( \zeta^{v \rightarrow v_T} \) that starts from \( v \) and reaches \( v_T \). Since \( G_d \) is strongly connected there exists a path \( \zeta^{v_T \rightarrow v'} \) that starts from \( v_T \) and reaches \( v' \). Thus \( \zeta^{v \rightarrow v'} = \zeta^{v \rightarrow v_T} \oplus \zeta^{v_T \rightarrow v'} \) is a path that starts from \( v \) and reaches \( v' \) in \( |\zeta^{v \rightarrow v'}| \geq T \) steps and \( v' \in L_{|\zeta^{v \rightarrow v'}|}(v) \).

**Lemma 7.** Let \( G_d \) be strongly connected. Let \( T_v \geq 1 \) denote the smallest positive horizon such that \( v \in L_{T_v}(v) \). Then, for all \( v \in V_d \), the sequence \( (L_{nT_v}(v))_{n \geq 0} \) converges to a limiting layer \( L(v) \subseteq V_d \), i.e., for all \( v \in V_d \) there exists \( \bar{n}_v \geq 0 \) such that, for all \( n \geq \bar{n}_v \), \( L_{nT_v}(v) = L(v) \).

**Proof.** Since \( G_d \) is connected, \( v \) must be able to reach itself again and so there indeed exists a \( T_v \geq 1 \) such that \( v \in L_{T_v}(v) \).
We now see that sequence \( (L_{nT_v}(v))_{n \geq 0} \) is "growing", i.e. \( L_{nT_v}(v) \subseteq L_{(n+1)T_v}(v) \) for all \( n \geq 0 \) by induction. The base case when \( n = 0 \) holds trivially by how we've defined \( L_{T_v}(v) \) since \( L_0(v) = \{v\} \subseteq L_{T_v}(v) \). Now assume for induction that \( L_{nT_v}(v) \subseteq L_{(n+1)T_v}(v) \). Then,
\[
L_{(n+1)T_v}(v) = L_{T_v}(L_{nT_v}(v)) \subseteq L_{T_v}(L_{(n+1)T_v}(v)) = L_{(n+2)T_v}(v)
\]
by Lemma 2, monotonicity.

We now see that sequence \( \{L_{nT_v}(v)\}_n \) is growing and bounded above, i.e \( L_{nT_v}(v) \subseteq L_{(n+1)T_v}(v) \) and \( L_{nT_v}(v) \subseteq V_d \) for all \( n \geq 0 \). Thus the sequence must converge to some fixed set \( \hat{L}(v) \subseteq V_d \), i.e there exists \( \bar{n}_v \geq 0 \) such that \( L_{nT_v}(v) = \hat{L}(v) \) for all \( n \geq \bar{n}_v \).

**Lemma 8.** Let \( G_d \) be connected. Then, for all \( v \in V_d \), the sequence \( \{L_t(v)\}_{t \geq 0} \) is **eventually periodic**, i.e. for all \( v \in V_d \), there exist \( T_v \geq 0, \delta_v \geq 1 \) such that, for all \( t \geq T_v \), \( L_t(v) = L_{t+\delta_v}(v) \).

**Proof.** By Lemma 7, since \( G_d \) is connected, for all \( v \in V_d \), \( (L_{nT_v}(v))_{n \geq 0} \) converges to a limiting layer \( \hat{L}(v) \), i.e. for all \( v \in V_d \), there exists \( \bar{n}_v \geq 0 \) such that, for all \( n \geq \bar{n}_v \), \( L_{nT_v}(v) = \hat{L}(v) \).

Set \( T_v = \bar{n}_vT_v \) and \( \delta_v = T_v \). Then we see that for all \( t \geq T_v = \bar{n}_vT_v \), it holds that
\[
L_{t+\delta_v}(v) = L_{t-\bar{n}_vT_v} + \bar{n}_vT_v + T_v(v) = L_{t-\bar{n}_vT_v} + (\bar{n}_v+1)T_v(v)
\]
(12)
\[
= L_{t-\bar{n}_vT_v}(L_{(\bar{n}_v+1)T_v}(v)) = L_{t-\bar{n}_vT_v}(L_{\bar{n}_vT_v}(v))
\]
(13)
\[
= L_{t-\bar{n}_vT_v}(\hat{L}(v)) = L_{t-\bar{n}_vT_v+\bar{n}_vT_v}(v)
\]
(14)
\[
= L_t(v)
\]
where 12 \( \rightarrow \) 13 holds since \( \bar{n}_v + 1 \) \( T_v \) \( \geq 0 \) and \( t - \bar{n}_vT_v \geq 0 \). Furthermore, 13 \( \rightarrow \) 14 holds by Lemma 7 since \( L_{(\bar{n}_v+1)T_v}(v) = \hat{L}(v) = L_{\bar{n}_vT_v}(v) \).

In words, Lemma 8 states that the layers induced by starting at any vertex always converge to a periodic sequence.

**Definition 8.** Let \( (a_t)_{t \geq 0} \) be a sequence. We say that a sequence \( (b_t)_{t \geq 0} \) is a **tail** of the sequence \( (a_t)_{t \geq 0} \) if and only if there exists an index \( N \geq 0 \) such that \( b_t = a_t+N \). Let \( (a_t)_{t \geq 0} \) be an eventually periodic sequence. We say that a sequence \( (b_t)_{t \geq 0} \) is a **periodic tail** of the sequence \( (a_t)_{t \geq 0} \) if and only if \( (b_t)_{t \geq 0} \) is a periodic sequence and a tail of \( (a_t)_{t \geq 0} \).

We now prove some characteristics of the periodic tail.

**Lemma 9.** Let \( G_d \) be strongly connected. Let us denote \( \hat{L}_{t}(v) := L_{t}(\hat{L}(v)) \). Then, the sequence \( \{\hat{L}_{t}(v)\}_{t \geq 0} \) is a periodic tail of the sequence \( \{L_{t}(v)\}_{t \geq 0} \).

**Proof.** From Lemma 7, \( (L_{nT_v}(v))_{n \geq 0} \) converges to \( \hat{L}(v) \), so there exists \( \bar{n}_v \) such that \( L_{\bar{n}_vT_v}(v) = \hat{L}(v) \). Therefore, \( L_{t+\bar{n}_vT_v}(v) = \hat{L}(v) \) and \( \{L_{t}(v)\}_{t \geq 0} \) is a tail of the sequence \( \{L_{t}(v)\}_{t \geq 0} \). It is left to show that \( \{\hat{L}_{t}(v)\}_{t \geq 0} \) is periodic.

\( L_0(v) = L_{T_v}(v) \) \( \Rightarrow \) (\( \forall t \geq 0 \), \( L_t(v) = L_{T_v+t(v)} \)), therefore \( \{\hat{L}_{t}(v)\}_{t \geq 0} \) is periodic.

**Lemma 10.** Let \( G_d \) be strongly connected. Let \( T_v \geq 1 \) denote the smallest horizon \( t \geq 1 \) such that \( v \in L_t(v) \). Let \( \delta_v \geq 1 \) denote the period of the tail sequence \( \{\hat{L}_{t}(v)\}_{t \geq 0} \) so that \( \hat{L}_{t}(v) = \hat{L}_{t'(v)} \) for \( 0 \leq t < t' \) if and only if \( (t' - t) \mod \delta_v = 0 \). Then \( T_v \mod \delta_v = 0 \).

**Proof.** We first know that \( T_v \geq \delta_v \) trivially holds since \( \hat{L}_0(v) = \hat{L}_{T_v}(v) \). Since \( T_v \geq \delta_v > 0 \) are integers, \( T_v \) admits a unique quotient \( q \geq 1 \) and remainder \( m \geq 0 \) by Euclid's lemma, i.e \( T_v = q\delta_v + m \). Assume for contradiction that \( m > 0 \). Then, \( q\delta_v < T_v \) and \( m < \delta_v \). But then we have \( L_{q\delta_v}(v) = \hat{L}_{T_v}(v) \) and \( T_v - q\delta_v = m < \delta_v \) and there does not exist an integer \( n > 0 \) such that \( T_v - q\delta_v = n\delta_v \) which is a contradiction. Thus it must be that \( m = 0 \) as desired.
Lemma 11. Let $G_d$ be strongly connected. Let $\delta_v \geq 1$ denote the period of the tail sequence $(\bar{L}_t(v))_{t \geq 0}$ so that $\bar{L}_t(v) = \bar{L}_{t'}(v)$ for $0 \leq t < t'$ if and only if $(t' - t) \mod \delta_v = 0$. Then, for all $v \in V_d$ and $0 \leq t < t'$ such that $t' - t \mod \delta_v \neq 0$, $\bar{L}_t(v) \cap \bar{L}_t(v) = \emptyset$, i.e., limiting layers within a period are all disjoint sets regardless of the starting vertex. Equivalently, for all $v \in V_d$ and $0 \leq t < t'$, $\bar{L}_t(v) = \bar{L}_{t'}(v)$ if $t' - t \mod \delta_v = 0$ and $\bar{L}_t(v) \cap \bar{L}_{t'}(v) = \emptyset$ otherwise.

Proof. Since $(\bar{L}_t(v))_{t \geq 0}$ is periodic with period $\delta_v$, it suffices to prove that for all $v \in V_d$ and $0 \leq t < t' \leq \delta_v$ such that $t' - t < \delta_v$, $\bar{L}_t(v) \cap \bar{L}_{t'}(v) = \emptyset$. We first prove the following claim:

Claim 1. For all $v \in V_d$ and $t \geq 0$, if $t \mod \delta_v \neq 0$, then $v \notin \bar{L}_t(v)$.

Proof. Again, due to periodicity, it suffices to prove that for all $v \in V_d$ and $0 < t < \delta_v$, $v \notin \bar{L}_t(v)$. Assume for contradiction that there exist $v \in V_d$ and $0 < t < \delta_v$ such that $v \in \bar{L}_t(v)$.

- We then claim that $\bar{L}_0(v) \subseteq \bar{L}_t(v)$. Assume, again, for contradiction that $\bar{L}_0(v) \not\subseteq \bar{L}_t(v)$. Let $T_v \geq 1$ denote the smallest horizon $t \geq 1$ such that $v \in \bar{L}_t(v)$. Since $G_d$ is connected, by Lemma 10, $T_v = q\delta_v$ for some quotient integer $q \geq 1$. Then for all $n \geq 0$

$$L_{nT_v}(v) \subseteq L_{nT_v}(\bar{L}_t(v)) = L_{nT_v}(L_t(v)) = L_{t+nT_v}(L_t(v)) = L_{t+n\delta_v}(v)$$

where the inclusion relation holds by monotonicity since $v \in \bar{L}_t(v)$ by outer assumption and the second equality holds by commutativity. (Lemma 2)

Since $G_d$ is connected, by Lemma 7, there exists $\bar{n}_v \geq 0$ such that, for all $n \geq \bar{n}_v$, $L_{nT_v}(v) = L(v) = L_0(v)$. Combining this result with Eq. 15, there exists $\bar{n}_v \geq 0$ such that, for all $n \geq \bar{n}_v$

$$L_{nT_v}(v) = L_0(v) \subseteq L_{t+n\delta_v}(v)$$

Then, since $\bar{L}_0(v) \not\subseteq \bar{L}_t(v)$, there exists $\bar{n}_v \geq 0$ such that $\bar{L}_t(v) \neq \bar{L}_{t+n\delta_v}(v)$ for $n \geq \bar{n}_v$ which contradicts the assumption that $(\bar{L}_t(v))_{t \geq 0}$ is periodic with period $\delta_v$. Thus, by contradiction, we have shown $L_0(v) \subseteq \bar{L}_t(v)$.

- Now we enumerate all cases for $\bar{L}_t(v)$ that satisfy $L_0(v) \subseteq \bar{L}_t(v)$.

If $\bar{L}_0(v) = \bar{L}_t(v)$, then this contradicts the assumption that $(\bar{L}_t(v))_{t \geq 0}$ is periodic with period $\delta_v$.

If $\bar{L}_0(v) \subset \bar{L}_t(v)$, then for all $n \geq 1$,

$$\bar{L}_{nt}(v) = L_{nt}(\bar{L}(v)) = L_{nt}(\bar{L}_0(v)) \subseteq L_{nt}(\bar{L}_t(v)) = L_{nt}(L_t(v)) = L_{(n+1)t}(L_t(v)) = L_{(n+1)t}(v)$$

where the inclusion relation holds by monotonicity since we’ve just assumed $\bar{L}_0(v) \subset \bar{L}_t(v)$ and the fourth equality holds by commutativity. (Lemma 2) By transitivity this implies that for all $1 \leq n \leq n'$,

$$\bar{L}_{nt}(v) \subseteq \bar{L}_{n't}(v)$$

Choosing $n = 1$ and $n' = \delta_v$ we have $\bar{L}_0(v) \subset \bar{L}_t(v) \subseteq \bar{L}_{\delta_v}(v)$ and so $\bar{L}_0(v) \neq \bar{L}_{\delta_v}(v)$. Since $t > 0$ this again contradicts the periodicity of $(\bar{L}_t(v))_{t \geq 0}$. Thus we have shown, by contradiction, for all $v \in V_d$ and $0 < t < \delta_v$, $v \notin \bar{L}_t(v)$.

Now to prove the original lemma, assume for contradiction that there exists $0 \leq t < t' \leq \delta_v$ such that $0 < t' - t < \delta_v$, and a shared vertex $v_{t',v} \in V_d$ such that $v_{t',v} \in \bar{L}_t(v)$ and $v_{t',v} \in \bar{L}_{t'}(v)$. Since $G_d$ is strongly connected $v_{t',v}$ can reach $v$ and so there exists a $l$ such that $v \in \bar{L}_{t+l}(v)$ and $v \in \bar{L}_{t'+l}(v)$ by trivial extension of Lemma 6. We now enumerate all cases for the value of $t + l$.

If $t + l \mod \delta_v \neq 0$, this contradicts Claim 1 since $v \in \bar{L}_{t+l}(v)$.

If $t + l \mod \delta_v = 0$, then $t' + l \mod \delta_v \neq 0$ since $(t' + l) - (t + l) = t' - t < \delta_v$. this contradicts Claim 1 since $v \in \bar{L}_{t'+l}(v)$.

Lemma 12. Let $G_d$ be strongly connected. Then, for all $v, v' \in V_d$, the sequence $(\bar{L}_t(v'))_{t \geq 0}$ is a periodic tail of the sequence $(\bar{L}_t(v'))_{t \geq 0}$ i.e. vertex layers all converge to the same periodic sequence regardless of the starting vertex.
Proof. Pick any \( v, v' \in V_d \) and consider their corresponding periodic tail sequences \((\hat{L}_t(v'))_{t \geq 0}, (\hat{L}_t(v'))_{t \geq 0}\). (which exists by Lemma 8) Without loss of generality, we will let the first layer of the periodic tails be those containing the initial vertex, i.e. \( v \in L_0(v), v' \in L_0(v') \). Such layers exist in the periodic tail by Lemma 6.

Let \( t_v, t_{v'} \geq 0 \) denote the horizons at which \( v \in L_{t_v}(v), v' \in L_{t_{v'}}(v) \). Again, such layers exist by Lemma 6. Then, we claim \( L_0(v) \subseteq L_{t_v}(v') \). To see this, first note that the sequence \((L_{nT_v}(v))_{n \geq 0}\), where \( T_v \geq 1 \) is the shortest time horizon at which \( v \in L_{T_v}(v) \), converges to \( L_0(v) \) by Lemma 7. Furthermore, \((L_{t_v+nT_v}(v'))_{n \geq 0} = (L_{t_{v'}}(v'))_{n \geq 0}\) since \((v \in L_{t_v}(v'), L_{t_v+nT_v}(v')) \Rightarrow (L_{t_v+nT_v}(v') = L_{t_v}(v')) \) by Lemma 11. Since \( \{v \subseteq L_{t_v}(v') \), it follows from monotonicity (Lemma 2) that \( L_{nT_v}(v) \subseteq L_{nT_v}(L_{t_v}(v')) = L_{t_v+nT_v}(v') \) for all \( n \geq 0 \). Since there exists an \( \bar{n}_{v} \geq 0 \) such that \( L_{\bar{n}_{v}}(v) = L_{\bar{n}_{v}}(v) \) we thus have \( L_{\bar{n}_{v}}(v) 
subseteq L_{t_v}(v') \). The same argument can be applied to obtain \( L_0(v') \nsubseteq L_{t_{v'}}(v) \).

We now consider the two different cases. If \( \hat{L}_0(v') = \hat{L}_0(v) \), then it trivially follows that the sequences \((\hat{L}_t(v'))_{t \geq 0}, (\hat{L}_t(v'))_{t \geq 0}\) are the same. For the second case if \( \hat{L}_0(v') \neq \hat{L}_0(v) \) then \( L_t(v) = L_t(\hat{L}_0(v)) \subseteq L_t(\hat{L}_{t_v}(v')) = \hat{L}_{t_v+t_{v'}}(v') \) for all \( t \geq 0 \) and \( L_t(v') = L_t(\hat{L}_0(v')) \subseteq L_t(\hat{L}_{t_{v'}}(v)) = \hat{L}_{t_{v'}+t_{v}}(v) \) for all \( t \geq 0 \). Thus \( L_0(\hat{L}_0(v')) \subseteq \hat{L}_{t_{v'}}(v) \subseteq L_{t_{v'}+t_{v}}(v) \). Then, \( v \in L_0(v) \Rightarrow v \in L_{t_{v'}+t_{v}}(v') \) and it follows that \( L_0(v) = \hat{L}_{t_{v'}+t_{v}}(v') \) by Lemma 11. Thus, \( L_0(v) = \hat{L}_{t_{v'}}(v') \) which implies that \( L_t(v) = L_{t_{v'}+t_{v}}(v') \) for all \( t \geq 0 \) and so \((L_t(v))_{t \geq 0} \) is a tail of \((L_t(v'))_{t \geq 0} \).

From Lemma 12, we see that the layer sequence converges to the same periodic sequence regardless of the starting vertex. Thus, we shall henceforth denote a periodic tail of \( G_d \) as \((L_t)_{t \geq 0} \) dropping the dependence on initial vertex.

Lemma 13. Let \( G_d \) be strongly connected and let \((\bar{L}_t)_{t \geq 0} \) be a periodic tail of the layer sequences in \( G_d \). For all \( v \in V_d \) and \( t, t' \geq 0 \), \((\bar{L}_t(v) \cap \bar{L}_{t'}(v) \neq \emptyset) \Rightarrow (\bar{L}_t(v) \subseteq \bar{L}_{t'}(v)) \).

Proof. Suppose for contradiction that there exists \( v \in V_d \) and \( t, t' \geq 0 \) such that \((L_t(v) \cap \bar{L}_{t'}(v) \neq \emptyset) \), but \((L_t(v) \nsubseteq \bar{L}_{t'}(v)) \). Let \( v^- \in L_0(v) - \bar{L}_{t'}(v) \) and \( v^\ominus \in L_0(v) \cap L_{t'}(v) \).

Let \( T_v \geq 1 \) denote the smallest positive horizon such that \( v \in L_{T_v}(v) \). Then, \( L_{nT_v}(v) = L_{nT_v}(L_{T_v}(v)) \subseteq L_{nT_v}(L_{T_v}(v)) = L_{(n+1)T_v}(v) = L_{(n+1)T_v}(v) \) for all \( n \geq 0 \) by Lemma 2 since \((L_{nT_v}(v))_{n \geq 0} \subseteq (L_{(n+1)T_v}(v))_{n \geq 0} \). Thus, the sequence \((L_{nT_v}(v))_{n \geq 0} \) must converge to some fixed set \( L_v \) since the sequence is growing and bounded above, i.e. \( L_{nT_v}(v) \subseteq L_{(n+1)T_v}(v) \) and \( L_{nT_v}(v) \subseteq V_d \) for all \( n \geq 0 \). Thus, \( \bar{L}_{t'} \) is an element of the tail \((\bar{L}_t)_{t \geq 0} \). Since \( v^\ominus \in L_{t'}(v) \), we have \( v^\ominus, v' \in L_{t'}(v) \). This contradicts Lemma 11 since \( L_{t'}, L_{t'} \) are two tail layers that are not the same but also not disjoint.

We now prove the necessary direction of the main theorem. We show the contrapositive, i.e. if either \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is not weakly identifiable or not coverable, it is not strongly identifiable. By Proposition 1, \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) must be weakly identifiable to be strongly identifiable. Thus, consider \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) that is weakly identifiable but not coverable. By Proposition 2 it suffices to show that \( \exists \bar{r}, \bar{r} \in R \) such that \( r \neq x \neq \bar{r} \) but \( r \simeq \bar{r} \).

Let \((\bar{L}_t)_{t \geq 0} \) be a periodic tail of the layer sequences in \( G_d \). Let \( r, r' \) be two rewards such that \( \forall v \notin L_0, r(v) = r(v) \) and \( \forall v \in L_0, r(v) = r(v) + c \) for some constant \( c \in \mathbb{R} \). Since there does not exist a covering initial state, clearly, \( L_0 \subseteq V_d \) and thus \( r \neq x \neq \bar{r} \). We will show that \( r \simeq \bar{r} \) to conclude that the \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is not strongly identifiable.

For all \( v \in V_d \) and for all paths \( \zeta = (v^0)_{0 \leq t \leq T}, \zeta' = (v'^0)_{0 \leq t \leq T} \) such that \( \zeta, \zeta' \in Z[v, d, T] \), we claim that \( \hat{r}(v^t) - \hat{r}(v'^t) = r(v^t) - r(v'^t) \) for all \( 0 \leq t \leq T \). To see this, first note that \( v^t, v'^t \in L_d(v) \) for all \( t \geq 0 \). We consider two cases: (1) If \( v^t \in L_0, v'^t \in L_0 \) since \( v^t, v'^t \in L_d(v) \) and, by Lemma 13, \((L_t(v) \cap L_{t'}(v) \neq \emptyset) \Rightarrow (L_t(v) \subseteq L_{t'}) \). Thus, \( \hat{r}(v^t) - \hat{r}(v'^t) = r(v^t) + c - v'^t - c = r(v^t) - r(v'^t) \). (2) If \( v^t \notin L_0, v'^t \notin L_0 \) since \( v^t, v'^t \in L_d(v) \) and, by the contrapositive of Lemma 13, \((L_t(v) \subseteq L_{t'}) \Rightarrow (L_t(v) \cup L_{t'} = \emptyset) \). Thus, \( \hat{r}(v^t) - \hat{r}(v'^t) = r(v^t) - r(v'^t) \)

Then,

\[
\hat{r}(\zeta') - \hat{r}(\zeta) = \sum_{t=0}^{T} \gamma^t (r(v'^t) - r(v^t)) \\
= \sum_{t=0}^{T} \gamma^t (\hat{r}(v'^t) - \hat{r}(v^t)) \\
= \hat{r}(\zeta') - \hat{r}(\zeta)
\]
Therefore, \( r, \hat{r} \) are two trajectory equivalent rewards which are not state-action equivalent. Hence \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is not strongly identifiable.

**Corollary 2.** (Strong Identification Condition) For all \((d, r, T, J)\) such that \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is a proper MDP model and \( G_d \) is strongly connected,

- (Sufficiency) \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is weakly identifiable, \( G_d \) aperiodic \(\Rightarrow\) \( \exists T_0 \geq 0 \) such that \( \forall T \geq T_0, \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable
- (Necessity) \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable \(\Rightarrow\) \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is weakly identifiable, \( G_d \) aperiodic.

**Proof.** Let \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) be proper and \( G_d \) be strongly connected. • (Sufficiency) Since \( G_d \) is strongly connected and aperiodic, it is covering by Proposition 3, i.e. there exists an initial vertex \( v_0 \in V^0_d \) that is \( \tau^* \)-covering for some \( \tau^* \). Let \( T_0 = 2\tau^* \), \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable for all \( T \geq T_0 \) by Theorem 2.

• (Necessity) If \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is proper, strongly identifiable, and \( G_d \) is strongly connected, by Theorem 3, it is weakly identifiable and \( G_d \) is coverable. Since, \( G_d \) is strongly connected and coverable, it is aperiodic by Proposition 3.

### A.5. Strong Identifiability Test Algorithms

**Theorem 3.** Let \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) be a weakly identifiable MDP model and \( G_d \) be strongly connected. Then,

- (Correctness) \( \text{MDPIDTest}(\mathcal{P}_{\text{MDP}}[R; d, T, J]) \) returns 1 (True) if and only if \( \exists T \) such that \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable.

- (Efficiency) \( \text{MDPIDTest} \) runs with time and space complexity \( O(|E_d|) \)

**Proof.** • (Correctness) \( \text{MDPIDTest} \) returns 1 (True) if and only if the directed graph \( G_d \) is aperiodic as shown in (Denardo, 1977; Jarvis and Shier, 1999). Since \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is weakly identifiable and \( G_d \) is strongly connected, \( G_d \) is aperiodic if and only if \( \exists T \) such that \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable by Corollary 2.

• (Efficiency) Graph aperiodicity testing can be done in \( O(|E_d|) \) space and time as shown in (Denardo, 1977; Jarvis and Shier, 1999).

**Corollary 3.** (Strong Identification Condition) For all \((d, r, T, J)\) such that the MDP model \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is proper.

- (Sufficiency) \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is weakly identifiable, \( G_d \) is \( T_0 \)-coverable, and \( T \geq 2T_0 \) \(\Rightarrow\) \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable.

**Proof.** This result immediately follows from the proof of the Sufficiency direction for Theorem 2.

**Theorem 4.** Let \( \mathcal{P}_{\text{MDP}}[R; d, T, J] \) be a weakly identifiable MDP model. Then,

- (Correctness) If \( \text{MDPCoverTest}(\mathcal{P}_{\text{MDP}}[R; d, T, J]) \) returns 1 (True) then, \( \exists T_0 \) such that \( \forall T \geq T_0, \mathcal{P}_{\text{MDP}}[R; d, T, J] \) is strongly identifiable.

- (Efficiency) \( \text{MDPCoverTest} \) runs with time complexity \( O(|V_d|^3 \log |V_d|) \) and space complexity \( O(|V_d|^2) \)

**Proof.** • Since \( M \) is the transition matrix, i.e \( M_{ij} = \hat{P}(v^{(j)}|v^{(i)}) \) where \( \hat{P}(x', a'|x, a) = P(x'|x, a) \), it is clear that \( M_{ij}^{V_d^2} \neq 0 \) if and only if \( v^{(j)} \in L_{V_d^2}(v^{(i)}) \). If \( \text{MDPCoverTest} \) returns 1 (True), then there exists \( v^{(i)} \in V^0_d \) that has a fully non-zero row \( M_{ij}^{V_d^2} \), i.e \( L_{V_d^2}(v^{(i)}) = V_d \). Thus, \( G_d \) is \( |V_d|^2 \)-coverable by \( v^{(i)} \). Let \( T_0 = 2|V_d|^2 \) and the result follows from Corollary 3. Therefore, \( \text{MDPCoverTest} \) returns 1 (True) if and only if

• (Efficiency) It is well known that computing matrix powers \( A^m \) (where the matrix \( A \) has size \( n \times n \)) can be done in \( O(n \log m) \) time and \( O(n^2) \) space (Cormen et al., 2009). Since \( M \) has size \( |V_d| \times |V_d| \), computing \( M^{V_d^2} \) has time complexity \( O(|V_d|^3 \log |V_d|) = O(|V_d|^3 \log |V_d|) \) and space complexity \( O(|V_d|^2) \). A naive approach to checking for rows

---

**Reward Identification in Inverse Reinforcement Learning**

---
with only non-zero entries requires enumerating over all elements of $M$ which can be done in $O(|V_a|^2)$ time and $O(1)$ space, thus not affecting the overall efficiency of the algorithm.