Appendices

A AUSO instance from Section 4.1

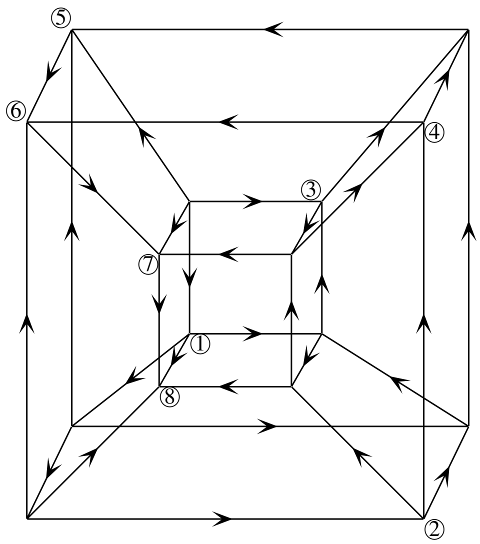

Figure 5: The only 4-AUSO (up to an isomorphism) on which HPI performs 8 vertex evaluations. The 8 vertices are numbered in sequence. This AUSO does not satisfy the Holt-Klee conditions. Notice, for example, that the inner 3-AUSO does not have 3 vertex-disjoint paths from source to sink.

B Proofs from Section 5

We provide a proof of Theorem 9, which uses the MDP designed by Melekopoglou and Condon [1994], shown in Figure 3. Recall from Section 5 that we only consider states $s \in \{1, 2, \dots, n\}$ as a part of our analysis.

For this proof, we find it convenient to consider a slight modification to RPI. If a policy π has $m > 1$ improvable states, note that RPI obtains $\pi' \succ \pi$ by picking uniformly at random among the $2^m - 1$ improving policies in $I(\pi)$. We consider an algorithm RPI1 that instead picks π' uniformly at random from $I(\pi) \cup \{\pi\}$. The reason for so doing is that RPI1 can be implemented by independently switching each improvable state with probability $1/2$, which simplifies our analysis. The consequence, though, is that RPI1 is not strictly a PI algorithm, since with a finite probability, we can get $\pi' = \pi$. This probability is at most $1/2$, and therefore, the expected number of policies visited by RPI1 (which might contain repetitions) is at most twice the expected number of policies visited by RPI. To prove the theorem, we show below that the former quantity is at least $n + 1$.

Building on Melekopoglou and Condon [1994], first we obtain a simple rule to check if a state s is switchable.

Lemma 10. *For a policy π for M_n, a state s is switchable if and only if*

$$\sum_{s' \leq s} \pi(s) \equiv 0 \mod 2.$$

Proof. For states $s \in \{1, 2, \dots, n\}$, Melekopoglou and Condon [1994] define

$$a(1) = -\frac{1}{2} \text{ and } a(s+1) = a(s)\left(\frac{1}{2} - \pi(s)\right).$$

It is easy to verify from the definition that $a(s+1)$ is negative if and only if $\sum_{s' \leq s} \pi(s) \equiv 0 \mod 2$ [Melekopoglou and Condon, 1994, see Corollary 2.3]. Since $a(s+1) = a(s)(\frac{1}{2} - \pi(s))$, $a(s+1)$ is negative if and only if $\pi(s) = 0$ and $a(s) < 0$, or $\pi(s) = 1$ and $a(s) > 0$. Based on the structure of M_n, Melekopoglou and Condon[1994, see Corollary 2.4] show that the latter condition is equivalent to s being switchable. □

The crucial step in our proof is to define a *progress* function f on the policy space, which is then shown to be non-increasing with respect to PI updates.

Definition 11. *For a policy π for M_n,*

$$f(\pi) \stackrel{\text{def}}{=} \min\left(\textit{states}(T^{\pi}) \cup \{n+1\}\right).$$

In other words, $f(\pi)$ is defined to be the smallest switchable state if π is not optimal, and $n + 1$ if it is $\pi^\star$. The lemma below establishes the monotonicity of f.

Lemma 12. *If RPI1 visits the policies $\pi^0, \pi^1, \dots, \pi^m$ in sequence, then for $1 \leq i \leq m$, $f(\pi^{i-1}) \leq f(\pi^i)$.*

Proof. Since we stop when there are no improvable states, $f(\pi^{m-1}) \leq f(\pi^m) = n + 1$. Otherwise assume that $i < m$. Let $f(\pi^{i-1}) = s$. Since vertex s is the smallest switchable state in π^{i-1}, any state s' will not be switched in π^{i-1} for $1 \leq s' < s$, and hence $\pi^i(s') = \pi^{i-1}(s')$. It follows from Lemma 10 that states $1, 2, \dots, s-1$ are not switchable in π^i. Thus, $f(\pi^i) \geq s = f(\pi^{i-1})$. □

Next we show that as RPI1 proceeds, with sufficiently high probability f increases quite slowly. It follows thereafter that at least $n + 1$ policy evaluations must be made in expectation if $\pi^0 = 0^n$ is the initial policy ($f(\pi^0)$ and $f(\pi^\star)$ differ by n).

Lemma 13. *If RPI1 visits the policies $\pi^0, \pi^1, \dots, \pi^m$ in sequence, then for $1 \leq i \leq m$, $t \geq 0$,*

$$\mathbb{P}\{f(\pi^i) - f(\pi^{i-1}) \geq t\} \leq \frac{1}{2^t}.$$

Proof. If $t = 0$, the RHS is 1 and the result trivial. Henceforth we assume $t > 0$. The proof splits into cases $f(\pi^{i-1}) = 1$ and $f(\pi^{i-1}) > 1$, which we consider in turn. Let $[x]$ denote the set $\{1, 2, \ldots, x\}$.

If $f(\pi^{i-1}) = 1$, $s = 1$ is switchable in π^{i-1}. From Lemma 10, $\pi^{i-1}(1) = 0$. Thus, let $\pi^{i-1} = 0^{s'}x$ for some $1 \leq s' \leq n$, $x \in \{0,1\}^{n-s'}$, and x starts with 1 or x is empty. Applying Lemma 10 for states $1, 2, \ldots, s'+1$, we get that states $1, 2, \ldots, s'$ are switchable in π^{i-1} and $s'+1$ is not switchable in π^{i-1}, if $s'+1 \in [n]$. If $f(\pi^i) \geq t+1$, the states $1, 2, \ldots, t$ are not switchable in π^i. Applying Lemma 10 for states $1, 2, \ldots, t$, we get that $\pi^i = 10^{t-1}y$ where $y \in \{0,1\}^{n-t}$. If $s' = n$, $t \leq n = s'$. t cannot be greater than s' if $s'+1 \in [n]$ as that will imply $\pi^i(s'+1) = 0 \neq \pi^{i-1}(s'+1)$, despite $s'+1$ not being switchable in π^{i-1}. Hence, if $t > s'$, $\mathbb{P}\{f(\pi^i) \geq t+1\} = 0 \leq \frac{1}{2^t}$. Otherwise $t \leq s'$. Therefore, states $1, 2, \ldots, t$ are switchable in π^{i-1}. To get to π^i from π^{i-1}, the state 1 must be switched and the states $2, 3, \ldots, t$ must not be switched. As each state is switched with probability $\frac{1}{2}$ by RPI1, the probability of this event happening is exactly $\frac{1}{2^t}$.

If $f(\pi^{i-1}) = s > 1$, s is switchable in π^{i-1} and $1, 2, \ldots, s-1$ are not switchable in π^{i-1}. Applying Lemma 10 for states $1, 2, \ldots, s$, we get $\pi^{i-1} = 10^{s-2}10^{s'}x$ for some $0 \leq s' \leq n-s$, $x \in \{0,1\}^{n-s-s'}$, and x starts with 1 or x is empty. Applying Lemma 10 for states $s+1, s+2, \ldots, s+s'$, we get that states $s+1, s+2, \ldots, s+s'$ are also switchable in π^{i-1} and $s+s'+1$ is not switchable in π^{i-1}, if $s+s'+1 \in [n]$. Note that since $i-1 < m$, $\pi^{i-1} \neq \pi^*$ and hence $s \leq n$. If $f(\pi^i) \geq s+t$, the states $1, 2, \ldots, s+t-1$ are not switchable in π^i. Applying Lemma 10 for states $1, 2, \ldots, s+t-1$, we get that $\pi^i = 10^{s+t-2}y$ where $y \in \{0,1\}^{n-s-t+1}$. If $s+s' = n$, $s+t-1 \leq n = s+s'$. $s+t-1$ cannot be greater than $s+s'$ if $s+s'+1 \in [n]$ as that will imply $\pi^i(s+s'+1) = 0 \neq \pi^{i-1}(s+s'+1)$, despite $s+s'+1$ not being switchable in π^{i-1}. Hence, if $s+t-1 > s+s'$, $\mathbb{P}\{f(\pi^i) \geq s+t\} = 0 \leq \frac{1}{2^t}$. Otherwise $s+t-1 \leq s+s'$. Therefore, states $s, s+1, \ldots, s+t-1$ are switchable in π^{i-1}. To get to π^i from π^{i-1}, the state s must be switched and the states $s+1, s+2, \ldots, s+t-1$ must not be switched. As each state is switched with probability $\frac{1}{2}$ by RPI1, the probability of this event happening is exactly $\frac{1}{2^t}$. □

Definition 14. *We define $L : \Pi \to \mathbb{R}_{\geq 0}$, where $L(\pi)$ is the expected number of policies evaluated by RPI1 starting from π.*

Note that even if we start from $\pi^0 = \pi^*$, we need to evaluate π^0 to know that it is optimal. Hence $L(\pi^*) = 1$.

Definition 15. *We define $N : [n+1] \to \mathbb{R}_{\geq 0}$, where*

$$N(s) = \min_{\pi \in \Pi, f(\pi) = s} L(\pi).$$

It directly follows from the definition that $N(f(\pi)) \leq L(\pi)$ for any $\pi \in \Pi$.

Theorem 16. *For $s \in [n+1]$, $N(s) \geq n+2-s$.*

Proof. If $s = n+1$, $f(\pi) = n+1$ is true only for $\pi = \pi^*$. Hence $N(n+1) = L(\pi^*) = 1 \geq n+2-(n+1)$.

Now, let $s \in [n]$. Let π be a policy such that $N(s) = L(\pi)$. Hence $f(\pi) = s$. Since $f(\pi^*) = n+1$, π is not optimal. Let π' be obtained from π by an RPI1 update.

First we upper-bound the expectation of $f(\pi')$. Since $f(\pi')$ is a non-negatively valued random variable, we can use the following expression for its expectation.

$$\begin{aligned}
\mathbb{E}[f(\pi')] &= \sum_{n+1 \geq s' \geq 1} \mathbb{P}\{f(\pi') \geq s'\} \\
&= \sum_{s \geq s' \geq 1} 1 + \sum_{n+1 \geq s' > s} \mathbb{P}\{f(\pi') \geq s'\} \\
&\leq s + \sum_{n+1 \geq s' > s} \frac{1}{2^{s'-s}} \\
&\leq s + \sum_{k=1}^{\infty} \frac{1}{2^k} \\
&= s + 1.
\end{aligned}$$

Now, assuming inductively that $N(s') \geq n+2-s'$ for $s < s' \leq n+1$, we can lower-bound $N(s) = L(\pi)$ as

$$\begin{aligned}
N(s) = &\quad 1 + \sum_{\pi'' \in \Pi} L(\pi'')\mathbb{P}\{\pi' = \pi''\} \\
\geq &\quad 1 + \sum_{\pi'' \in \Pi} N(f(\pi''))\mathbb{P}\{\pi' = \pi''\} \\
= &\quad 1 + \sum_{n+1 \geq s' \geq 1} \left[\sum_{\pi'' \in \Pi, f(\pi'') = s'} N(s')\mathbb{P}\{\pi' = \pi''\} \right] \\
= &\quad 1 + \sum_{n+1 \geq s' \geq 1} N(s') \left[\sum_{\pi'' \in \Pi, f(\pi'') = s'} \mathbb{P}\{\pi' = \pi''\} \right] \\
= &\quad 1 + \sum_{n+1 \geq s' \geq 1} N(s')\mathbb{P}\{f(\pi') = s'\} \\
= &\quad 1 + \sum_{n+1 \geq s' \geq s} N(s')\mathbb{P}\{f(\pi') = s'\},
\end{aligned}$$

since $\mathbb{P}\{f(\pi') < s = f(\pi)\} = 0$. We rearrange terms in a convenient form, and apply $\mathbb{E}[f(\pi')] \leq s+1$, to get

$$\begin{aligned}
N(s) \geq\ & 1 + \sum_{n+1 \geq s' \geq s} (N(s') - n - 2 + s')\mathbb{P}\{f(\pi'\} = s'\} \\
& + \sum_{n+1 \geq s' \geq s} (n + 2 - s')\mathbb{P}\{f(\pi') = s'\} \\
=\ & 1 + \sum_{n+1 \geq s' \geq s} (N(s') - n - 2 + s')\mathbb{P}\{f(\pi'\} = s'\} \\
& + n + 2 - \sum_{n+1 \geq s' \geq s} s'\mathbb{P}\{f(\pi'\} = s'\} \\
=\ & \sum_{n+1 \geq s' \geq s} (N(s') - n - 2 + s')\mathbb{P}\{f(\pi') = s'\} \\
& + n + 3 - \mathbb{E}[f(\pi')] \\
\geq\ & \sum_{n+1 \geq s' \geq s} (N(s') - n - 2 + s')\mathbb{P}\{f(\pi') = s'\} \\
& + n + 2 - s.
\end{aligned}$$

By the induction hypothesis, $N(s') - n - 2 + s'$ is non-negative for $s' > s$. Therefore, after removing terms corresponding to $s' > s$, we get

$$N(s) \geq n + 2 - s + (N(s) - n - 2 + s)\mathbb{P}\{f(\pi') = s\},$$

which rearranges into

$$(N(s) - n - 2 + s)(1 - \mathbb{P}\{f(\pi') = s\}) \geq 0.$$

Now, $\mathbb{P}\{f(\pi') = s\}$ cannot be 1 because there is a policy $\pi'' = \mathsf{modify}(\pi, \{(s, a)\}) \in \Pi$, where $a \in \{0, 1\}$ and $a \neq \pi(s)$, such that $\mathbb{P}\{\pi' = \pi''\} > 0$ and $f(\pi'') > s$ (since s is not switchable in π''). Hence, we must have $N(s) \geq n + 2 - s$. $\square$

At this point, Theorem 9 follows as a corollary; the statement of the theorem is reproduced below.

Corollary 17. *Starting from $\pi^0 = 0^n$, the expected number of policies RPI evaluates on M_n before terminating is at least $\frac{n+1}{2}$.*

Proof. For $\pi^0 = 0^n$, $f(\pi^0) = 1$. Thus

$$L(\pi^0) \geq N(1) \geq n + 2 - 1 = n + 1.$$

In other words, RPI1 evaluates at least $n + 1$ policies in expectation, which implies RPI evaluates at least half that number of policies in expectation. $\square$