# Finding Options that Minimize Planning Time (Appendix)

**Yuu Jinnai** [1] **David Abel** [1] **D Ellis Hershkowitz** [2] **Michael L. Littman** [1] **George Konidaris** [1]

## A. Appendix: Inapproximability of MOMI

In this section we prove Theorem 4:

**Theorem 4.**

1. *MOMI is $\Omega(\log n)$ hard to approximate even for deterministic MDPs unless $P = NP$.*
2. *MOMI$_{gen}$ is $2^{\log^{1-\epsilon} n}$-hard to approximate for any $\epsilon > 0$ even for deterministic MDPs unless $NP \subseteq DTIME(n^{poly \log n})$.*
3. *MOMI is $2^{\log^{1-\epsilon} n}$-hard to approximate for any $\epsilon > 0$ unless $NP \subseteq DTIME(n^{poly \log n})$.*

For Theorems 4.2 and 4.3 we reduce our problem to the Min-Rep, problem, originally defined by (Kortsarz, 2001). Min-Rep is a variant of the better studied label cover problem (Dinur & Safra, 2004) and has been integral to recent hardness of approximation results in network design problems (Dinitz et al., 2012; Bhattacharyya et al., 2012). Roughly, Min-Rep asks how to assign as few labels as possible to nodes in a bipartite graph such that every edge is "satisfied."

**Definition 1** (Min-Rep):
**Given** *a bipartite graph $G = (A \cup B, E)$ and alphabets $\Sigma_A$ and $\Sigma_B$ for the left and right sides of $G$ respectively. Each $e \in E$ has associated with it a set of pairs $\pi_e \subseteq \Sigma_A \times \Sigma_B$ which satisfy it.*
**Return** *a pair of assignments $\gamma_A : A \to \mathcal{P}(\Sigma_A)$ and $\gamma_B : B \to \mathcal{P}(\Sigma_B)$ such that for every $e = (A_i, B_j) \in E$ there exists an $(a, b) \in \pi_e$ such that $a \in \gamma_A(A_i)$ and $b \in \gamma_B(B_j)$. The objective is to minimize $\sum_{A_i \in A} |\gamma_A(A_i)| + \sum_{B_j \in B} |\gamma_B(B_j)|$.*

We illustrate a feasible solution to an instance of Min-Rep in Figure 1.

The crucial property of Min-Rep we use is that no polynomial-time algorithm can approximate Min-Rep well. Let $\tilde{n} = |A| + |B|$.

**Lemma 1** (Kortsarz 2001)**.** *Unless NP $\subseteq$ DTIME($n^{poly \log n}$), Min-Rep admits no $2^{\log^{1-\varepsilon} \tilde{n}}$ polynomial-time approximation algorithm for any $\varepsilon > 0$.*

As a technical note, we emphasize that all relevant quantities in Min-Rep are polynomially-bounded. In Min-Rep we have $|\Sigma_A|, |\Sigma_B| \leq \tilde{n}^{c'}$ for constant $c'$. It immediately follows that $\sum_e |\pi_e| \leq n^c$ for constant $c$.

### A.1. Hardness of Approximation of MOMI with Deterministic MDP

*Theorem 4.1 Proof.* The optimization version of the set-cover problem cannot be approximated within a factor of $c \cdot \ln n$ by a polynomial-time algorithm unless P = NP (Raz & Safra, 1997). The set-cover optimization problem can be reduced to MOMI with a similar construction for a reduction from SetCover-DEC to OI-DEC. Here, the targeted minimization values of the two problems are equal: $P(\mathcal{C}) = |\mathcal{O}|$, and the number of states in OI-DEC is equal to the number of elements in the set cover on transformation. Assume there is a polynomial-time algorithm within a factor of $c \cdot \ln n$ approximation for MOMI where $n$ is the number of states in the MDP. Let SetCover($\mathcal{U}, \mathcal{X}$) be an instance of the set-cover problem. We can convert the instance into an instance of MOMI($M, 0, 2$). Using the approximation algorithm, we get a solution $\mathcal{O}$ where $|\mathcal{O}| \leq c \ln n |\mathcal{O}^*|$, where $\mathcal{O}^*$ is the optimal solution. We construct a solution for the set cover $\mathcal{C}$ from the solution to the MOMI $\mathcal{O}$ (see the construction in the proof of Theorem 1). Because $|\mathcal{C}| = |\mathcal{O}|$ and $|\mathcal{C}^*| = |\mathcal{O}^*|$, where $\mathcal{C}^*$ is the optimal solution for the set cover, we get $|\mathcal{C}| = |\mathcal{O}| \leq c \ln n |\mathcal{O}^*| = c \ln n |\mathcal{C}^*|$. Thus, we acquire a $c \cdot \ln n$ approximation solution for the set-cover problem within polynomial time, something only possible if P=NP. Thus, there is no polynomial-time algorithm with a factor of $c \cdot \ln n$ approximation for MOMI, unless P=NP. □

### A.2. Hardness of Approximation of MOMI$_{gen}$

We now show our hardness of approximation of $2^{\log^{1-\epsilon} n}$ for MOMI$_{gen}$, Theorem 4.2.[1]

We start by describing our reduction from an instance of Min-Rep to an instance of MOMI$_{gen}$. The intuition behind our reduction is that we can encode choosing a label for a

[1] We assume that $\mathcal{O}'$ is a "good" set of options in the sense that there exists some set $\mathcal{O}^* \subseteq \mathcal{O}'$ such that $L_{\epsilon,V_0}(\mathcal{O}^*) \leq \ell$. We also assume, without loss of generality, that $\varepsilon < 1$ throughout this section; other values of $\varepsilon$ can be handled by re-scaling rewards in our reduction.

vertex in Min-Rep as choosing an option in our MOMI$_{gen}$ instance. In particular, we will have a state for each edge in our Min-Rep instance and reward will propagate quickly to that state when value iteration is run only if the options corresponding to a satisfying assignment for that edge are chosen.

More formally, our reduction is as follows. Consider an instance of Min-Rep, MR, given by $G = (A \cup B, E)$, $\Sigma_A$, $\Sigma_B$ and $\{\pi_e\}$. Our instance of MOMI$_{gen}$ is as follows where $\gamma = 1$ and $l = 3$.[2]

- **State space** We have a single goal state $S_g$ along with states $S'_g$ and $S''_g$. For each edge $e$ we create a state $S_e$. Let $\text{Sat}_A(e)$ consist of all $a \in \Sigma_A$ such that $a$ is in some assignment in $\pi_e$. Define $\text{Sat}_B(e)$ symmetrically. For each edge $e \in E$ we create a set of $2 \cdot |\text{Sat}_A(e)|$ states, namely $S_{ea}$ and $S'_{ea}$ for every $a \in \text{Sat}_A(e)$. We do the same for $b \in \text{Sat}_B(e)$.
- **Actions and Transitions** We have a single action from $S'_g$ to $S_g$, a single action from $S''_g$ to $S'_g$. For each edge $e$ we have the following deterministic actions: Every $S'_{ea}$ has a single outgoing action to $S_{ea}$ for $a \in \text{Sat}_A(e)$; Every $S_{eb}$ has a single outgoing action to $S_{eb'}$ for $b \in \text{Sat}_B(e)$; Every $S_{ea}$ has an outgoing action to $S_{eb}$ if $(a, b) \in \pi_e$ and every $S'_{eb}$ has a single outgoing action to $S_g$; Lastly, we have a single action from $S'_{ea}$ to $S''_g$ for every $a \in \text{Sat}_A(e)$.
- **Reward** The reward of arriving in $S_g$ is 1. The reward of arriving in every other state is 0.
- **Option Set** Our option set $\mathcal{O}'$ is as follows. For each vertex $A_i \in A$ and each $a \in \Sigma_A$ we have an option $O(A_i, a)$: The initiation set of this option is every $S_e$ where $e$ is incident to $A_i$; The termination set of this option is every $S_{ea}$ where $A_i$ is incident to $e$; The policy of this option takes the action from $S'_{ea}$ to $S_{ea}$ when in $S'_{ea}$ and the action from $S_e$ to $S'_{ea}$ when in $S_e$.

  Symmetrically, for every vertex $B_j \in B$ and each $b \in \Sigma_B$ we have an option $O(B_j, b)$: The initiation set of this option is every $S_{eb}$ where $e$ is incident to $B_j$; The termination set of this option is $S_g$; The policy of this option takes the action from $S_{eb}$ to $S'_{eb}$ when in $S_{eb}$ and from $S'_{eb}$ to $S_g$ when in $S'_{eb}$.

[2] It is easy to generalize these results to $l \geq 4$ by replacing certain edges with paths.

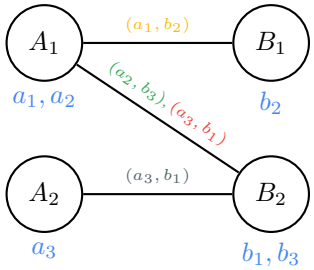



Figure 1: An instance of Min-Rep with $\Sigma_A = \{a_1, a_2, a_3\}$ and $\Sigma_B = \{b_1, b_2, b_3\}$. Edge $e$ is labeled with pairs in $\pi_e$. Feasible solution $(\gamma_A, \gamma_b)$ illustrated where $\gamma_A(A_i)$ and $\gamma_B(B_j)$ below $A_i$ and $B_j$ in blue. Constraints colored to coincide with stochastic action colors in Figure 3.

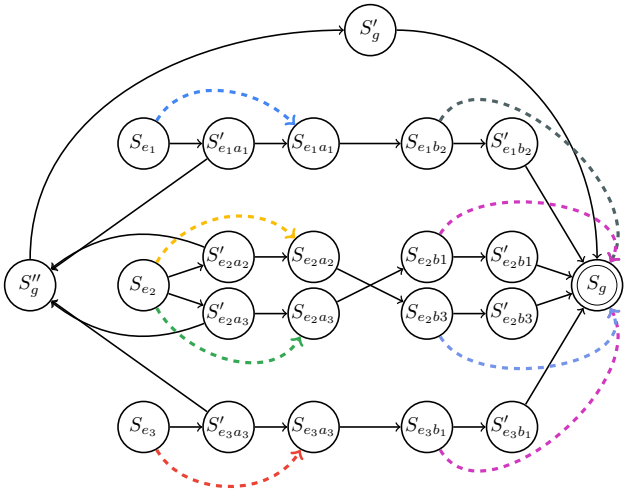



Figure 2: Our MOMI$_{gen}$ reduction applied to the Min-Rep problem in Figure 1. $e_1 = (A_1, B_1)$, $e_2 = (A_1, B_2)$, $e_3 = (A_2, B_2)$. Actions given in solid lines and each option in $\mathcal{O}'$ represented in its own color as a dashed line from initiation to termination states. Notice that a single option goes from $S_{e3b1}$ and $S_{e_2b_1}$ to $S_g$.

One should think of choosing option $O(v, x)$ as corresponding to choosing label $x$ for vertex $v$ in the input Min-Rep instance. Let MOMI$_{gen}$(MR) be the MDP output given instance MR of Min-Rep and see Figure 3 for an illustration of our reduction.

Let $\text{OPT}_{\text{MOMI}_{gen}}$ be the value of the optimal solution to MOMI$_{gen}$(MR) and let $\text{OPT}_{\text{MR}}$ be the value of the optimal Min-Rep solution to MR. The following lemmas demonstrates the correspondence between a MOMI$_{gen}$ and Min-Rep solution.

**Lemma 2.** *$OPT_{\text{MOMI}_{gen}} \leq OPT_{MR}$*

*Proof.* Given a solution $(\gamma_A, \gamma_B)$ to MR, define $\mathcal{O}_{\gamma_A,\gamma_B} := \{O(v, x) : v \in V(G) \wedge (\gamma_A(v) = x \vee \gamma_B(v) = x)\}$ as the corresponding set of options. Let $\gamma^*_A$ and $\gamma^*_B$ be the optimal solutions to MR which is of cost $\text{OPT}_{\text{MR}}$.

We now argue that $\mathcal{O}_{\gamma^*_A,\gamma^*_B}$ is a feasible solution to MOMI$_{gen}$(MR) of cost $\text{OPT}_{\text{MR}}$, demonstrating that the op-

timal solution to $\text{MOMI}_{gen}(\text{MR})$ has cost at most $\text{OPT}_{\text{MR}}$. To see this notice that by construction the $\text{MOMI}_{gen}$ cost of $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ is exactly the Min-Rep cost of $(\gamma_A^*, \gamma_B^*)$.

We need only argue, then, that $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ is feasible for $\text{MOMI}_{gen}(\text{MR})$ and do so now. The value of every state in $\text{MOMI}_{gen}(\text{MR})$ is 1. Thus, we must guarantee that after 3 iterations of value iteration, every state has value 1. However, without any options every state except each $S_e$ has value 1 after 3 iterations of value iteration. Thus, it suffices to argue that $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ guarantees that every $S_e$ will have value 1 after 3 iterations of value iteration. Since $(\gamma_A^*, \gamma_B^*)$ is a feasible solution to MR we know that for every $e = (A_i, B_j)$ there exists an $\bar{a} \in \gamma_A^*(A_i)$ and $\bar{b} \in \gamma_B^*(B_j)$ such that $(\bar{a}, \bar{b}) \in \pi_e$; correspondingly there are options $O(A_i, \bar{a}), O(B_j, \bar{b}) \in \mathcal{O}_{\gamma_A^*,\gamma_B^*}$. It follows that, given options $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ from, $S_e$ one can take option $O(A_i, \bar{a})$ then the action from $S_{e\bar{a}}$ to $S_{e\bar{b}}$ and then option $O(B_j, \bar{b})$ to arrive in $S_g$; thus, after 3 iterations of value iteration the value of $S_e$ is 1. Thus, we conclude that after 3 iterations of value iteration every state has converged on its value. □

We now show that a solution to $\text{MOMI}_{gen}(\text{MR})$ corresponds to a solution to MR. For the remainder of this section $\gamma_A^{\mathcal{O}}(A_i) := \{a : O(A_i, a) \in \mathcal{O}\}$ and $\gamma_B^{\mathcal{O}}(B_j) := \{b : O(B_j, b) \in \mathcal{O}\}$ is the Min-Rep solution corresponding to option set $\mathcal{O}$.

**Lemma 3.** *For a feasible solution to* $\text{MOMI}_{gen}(\text{MR})$*, $\mathcal{O}$, we have $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution to MR of cost $|\mathcal{O}|$.*

*Proof.* Notice that by construction the Min-Rep cost of $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is exactly $|\mathcal{O}|$. Thus, we need only prove that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution for MR.

We do so now. Consider an arbitrary edge $e = (A_i, B_j) \in E$; we wish to show that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ satisfies $e$. Since $\mathcal{O}$ is a feasible solution to $\text{MOMI}_{gen}(\text{MR})$ we know that after 3 iterations of value iteration every state must converge on its value. Moreover, notice that the value of every state in $\text{MOMI}_{gen}(\text{MR})$ is 1. Thus, it must be the case that for every $S_e$ there exists a path of length 3 from $S_e$ to $S_g$ using either options or actions. The only such paths are those that take an option $O(A_i, a)$, then an action from $S_{ea}$ to $S_{eb}$ then option $O(B_j, b)$ where $(a, b) \in \pi_e$. It follows that $a \in \gamma_A^{\mathcal{O}}(A_i)$ and $b \in \gamma_B^{\mathcal{O}}(B_j)$. But since $(a, b) \in \pi_e$, we then know that $e$ is satisfied. Thus, every edge is satisfied and so $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution to MR. □

*Theorem 4.2 Proof.* Assume NP $\not\subseteq$ DTIME($n^{\text{poly}\log n}$) and for the sake of contradiction that there exists an $\varepsilon > 0$ for which polynomial-time algorithm $\mathcal{A}_{\text{MOMI}_{gen}}$ can $2^{\log^{1-\varepsilon} n}$-approximate $\text{MOMI}_{gen}$. We use $\mathcal{A}_{\text{MOMI}_{gen}}$ to $2^{\log^{1-\varepsilon'} \tilde{n}}$ approximate Min-Rep for a fixed constant $\varepsilon' > 0$ in polynomial-time, thereby contradicting Lemma 1. Again, $\tilde{n}$ is the number of vertices in the graph of the Min-Rep instance.

We begin by noting that the relevant quantities in $\text{MOMI}_{gen}(\text{MR})$ are polynomially-bounded. Notice that the number of states $n$ in the MDP in $\text{MOMI}_{gen}(\text{MR})$ is at most $O(\tilde{n}^2|\Sigma_A||\Sigma_B|) = \tilde{n}^c$ for some fixed constant $c$ by the aforementioned assumption that $\Sigma_A$ and $\Sigma_B$ are polynomially-bounded in $\tilde{n}$.[3]

Our polynomial-time approximation algorithm to approximate instance MR of Min-Rep is as follows: Run $\mathcal{A}_{\text{MOMI}_{gen}}$ on $\text{MOMI}_{gen}(\text{MR})$ to get back option set $\mathcal{O}$. Return $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ as defined above as our solution to MR.

We first argue that our algorithm is polynomial-time in $\tilde{n}$. However, notice that for each vertex, we create a polynomial number of states. Thus, the number of states in $\text{MOMI}_{gen}(\text{MR})$ is polynomially-bounded in $\tilde{n}$ and so $\mathcal{A}_{\text{MOMI}_{gen}}$ runs in time polynomial in $\tilde{n}$. A polynomial runtime of our algorithm immediately follows.

We now argue that our algorithm is a $2^{\log^{1-\varepsilon'} \tilde{n}}$-approximation for Min-Rep for some $\varepsilon' > 0$. Applying Lemma 3, the approximation of $\mathcal{A}_{\text{MOMI}_{gen}}$ and then Lemma 2, we have that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution for MR with cost

$$\begin{aligned}\text{cost}_{\text{Min-Rep}}(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}}) &= |\mathcal{O}| \\ &\leq 2^{\log^{1-\varepsilon} n} \text{OPT}_{\text{MOMI}_{gen}} \\ &\leq 2^{\log^{1-\varepsilon} n} \text{OPT}_{\text{MR}}\end{aligned}$$

Thus, $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a $2^{\log^{1-\varepsilon} n}$ approximation for the optimal Min-Rep solution where $n$ is the number of states in the MDP of $\text{MOMI}_{gen}(\text{MR})$. Now recalling that $n \leq \tilde{n}^c$ for fixed constant $c$. We therefore have that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a $2^{\log^{1-\varepsilon} \tilde{n}^c} = 2^{c^{1-\varepsilon} \log^{1-\varepsilon} \tilde{n}} \leq c' \cdot 2^{\log^{1-\varepsilon} \tilde{n}}$ approximation for a constant $c'$. Choosing $\varepsilon$ sufficiently small, we have that $c' \cdot 2^{\log^{1-\varepsilon} \tilde{n}} \leq 2^{\log^{1-\varepsilon'} \tilde{n}}$ for sufficiently large $\tilde{n}$.

Thus, our polynomial-time algorithm is a $2^{\log^{1-\varepsilon'} \tilde{n}}$-approximation for Min-Rep for $\varepsilon' > 0$, thereby contradicting Lemma 1. We conclude that $\text{MOMI}_{gen}$ cannot be $2^{\log^{1-\epsilon} n}$-approximated. □

### A.3. Hardness of Approximation of MOMI with Stochastic MDP

We now show our hardness of approximation of $2^{\log^{1-\epsilon} n}$ for MOMI, Theorem 4.3. We will notably use the stochasticity

[3] It is also worth noticing that since we create at most $O(\tilde{n}|\Sigma_A| + \tilde{n}|\Sigma_B|)$ options, the total number of options in $\mathcal{O}'$ is at most polynomial in $\tilde{n}$.

of the input MDP to show this result.[4]

We begin by describing our reduction from an instance of Min-Rep to an instance of MOMI. The intuition behind our reduction is as follows. As in our reduction for MOMI$_{gen}$ we will have vertex for each edge in our Min-Rep instance and reward will propagate quickly to that vertex when value iteration is run only if the options corresponding to a satisfying assignment for that edge are chosen. The challenge, however, is that since our options are now only point options (whereas in MOMI$_{gen}$ they were arbitrary options) it seems that we can no longer constrain a solution to choose options exactly corresponding to a feasible Min-Rep solution.

To solve this issue we critically use stochasticity. Whether or not a given edge in a Min-Rep is satisfied is an or of ands: A fixed edge is satisfied when *one* of its satisfying assignments is met (an or) and a given satisfying assignment is met when both endpoints have the right labels (an and). We will exploit the fact that the value of a state in an MDP is a max over actions to encode the "or" in Min-Rep and we will use the fact that in a stochastic MDP the value of a (state, action) pair is the sum over states to encode the "and" in Min-Rep.

More formally, our reduction is as follows. Consider instance MR of Min-Rep given by $G = (A \cup B, E)$, $\Sigma_A$, $\Sigma_B$ and $\{\pi_e\}$. Our instance of MOMI is as follows where $\gamma = 1$ and $l = 2$.[5]

- **State space** We have a goal state $S_i$ for each $A_i \in A$. Again, let $\text{Sat}_A(e)$ consist of all $a \in \Sigma_A$ such that $a$ is in some assignment in $\pi_e$. For each $A_i \in A$ and $a \in \text{Sat}_A(e)$ we will we add to our MDP states $S_{ia}$ and $S'_{ia}$. We symmetrically do the same for all states in $\Sigma_B$. For each $e \in E$ we will also add a state $S_e$.[6]
- **Actions and Transitions** Every $S_{ia}$ state has a single action to $S'_{ia}$ and every $S_{ia}$ state has a single action to $S_i$. The same symmetrically holds for states from a $B_j \in B$. Every $S_e$ for $e = (A_i, B_j)$ has $|\pi_{(A_i,B_j)}|$ actions associated with it, namely $\{\alpha_{(a,b)}\}$ where $(a, b) \in \pi_{(A_i,B_j)}$. Action $\alpha_{(a,b)}$ has a probability .5 of transitioning to state $S_{ia}$ and a probability .5 of transitioning to state $S_{jb}$.
- **Reward** The reward of arriving in any $S_i$ or $S_j$ for $A_i \in A$ or $B_j \in B$ is 1 and 0 for every other state.

[4]We may assume without loss of generality $\varepsilon < .5$ throughout this section; rewards in our reduction can be re-scaled to handle larger $\varepsilon$.

[5]It is easy to generalize these results to $l \geq 3$ by replacing edges with paths.

[6]It is not hard to see that this construction can be modified so that we have only a single goal state if need be; we need only set every $S_i$ and $S_j$ to be the same state. We assume multiple goal states for ease of exposition.

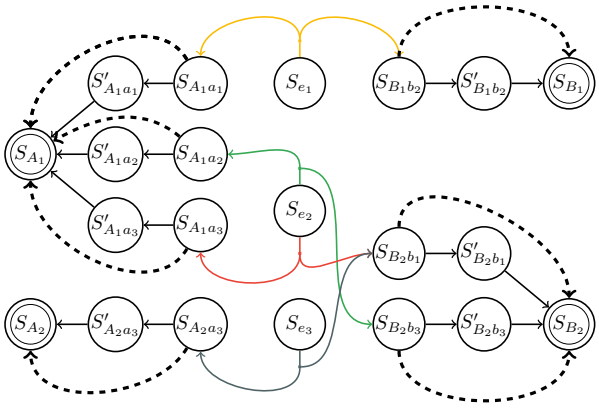



Figure 3: Our MOMI reduction applied to the Min-Rep problem in Figure 1. $e_1 = (A_1, B_1)$, $e_2 = (A_1, B_2)$, $e_3 = (A_2, B_2)$. Stochastic options colored according to the pair in $\pi_e$ to which they correspond, branching into the two states in which they arrive with equal probability. Deterministic action given as solid black arcs. Possible point options given as dashed arcs.

Notice that no point options have $S_e$ as an initialization state since any such option would have a .5 probability of never terminating (and we assume our options always terminate). See Figure 3 for an illustration of our reduction. One should think of choosing a point option from $S_{ia}$ to $S_i$ as corresponding to choosing label $a$ for $A_i$ in the input Min-Rep instance. The same holds for label $b$ for $B_j$ and choosing a point option from $S_{jb}$ to $S_j$. Let MOMI(MR) be the MOMI instance output by our reduction given instance MR of Min-Rep.

We now demonstrate that our reduction allows us to show that MOMI cannot be $2^{\log^{1-\varepsilon} n}$-approximated for any $\varepsilon > 0$. Let $\text{OPT}_{\text{MOMI}}$ be the value of the optimal solution to MOMI(MR) and let $\text{OPT}_{\text{MR}}$ be the value of the optimal Min-Rep solution to MR. The following lemmas demonstrates the correspondence between a MOMI and Min-Rep solution.

**Lemma 4.** *$OPT_{\text{MOMI}} \leq OPT_{MR}$*

*Proof.* Our proof translates between point options in our reduction and assignments in the input Min-Rep instance in the natural way. Given a solution $(\gamma_A, \gamma_B)$ to MR, define $\mathcal{O}_{\gamma_A,\gamma_B}$ as consisting of all point options from $S_{ia}$ to $S_i$ if $a \in \gamma_A(A_i)$ and all points options from $S_{jb}$ to $S_j$ if $b \in \gamma_B(B_j)$. Let $\gamma^*_A$ and $\gamma^*_B$ be the optimal solutions to MR which is of cost $\text{OPT}_{\text{MR}}$.

We claim that $\mathcal{O}_{\gamma^*_A,\gamma^*_B}$ is a feasible solution to MOMI(MR) of cost $\text{OPT}_{\text{MR}}$, demonstrating that the optimal solution to MOMI(MR) has cost at most $\text{OPT}_{\text{MR}}$. To see this notice that by construction the MOMI cost of $\mathcal{O}_{\gamma^*_A,\gamma^*_B}$ is exactly

the Min-Rep cost of $\gamma_A^*, \gamma_B^*$.

We need only argue, then, that $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ is feasible for MOMI(MR) and do so now. Notice that the value of every state in MOMI is 1. Thus, we must guarantee that after 2 iterations of value iteration, every state has value 1. However, without any options every state except for $S_e$ where $e \in E$ has value 1 after 2 iterations of value iteration. Thus, it suffices to argue that $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ guarantees that every $S_e$ will have value 1 after 2 iterations of value iteration. Since $(\gamma_A^*, \gamma_B^*)$ is a feasible solution to MR we know that for every $e = (A_i, B_j)$ there exists $\bar{a} \in \gamma_A^*(A_i)$ and $\bar{b} \in \gamma_B^*(B_j)$ such that $(\bar{a}, \bar{b}) \in \pi_e$; correspondingly there is some action from $S_e$ with a .5 probability of resulting in state $S_{i\bar{a}}$ and a .5 probability of resulting in state $S_{j\bar{b}}$ where $\mathcal{O}_{\gamma_A^*,\gamma_B^*}$ has a point option from $S_{i\bar{a}}$ to $S_i$ and a point options from $S_{j\bar{b}}$ to $S_j$. That is, $V_1(S_{i\bar{a}}) = 1$ and $V_1(S_{j\bar{b}}) = 1$. Thus, after one iteration of value iteration the values of $S_{i\bar{a}}$ and $S_{j\bar{b}}$ are both 1 and so after two iterations of value iteration the value of $S_e$ is

$$\begin{aligned} V_2(S_e) &= \max_{\alpha_{(a,b)}} .5 \cdot (V_1(S_{ia})) + .5 \cdot (V_1(S_{jb})) \\ &\geq .5 \cdot (V_1(S_{i\bar{a}})) + .5 \cdot (V_1(S_{j\bar{b}})) \\ &= 1. \end{aligned}$$

Thus, $V_2(S_e) = 1$ for every $S_e$ and so we conclude that after two iterations of value iteration every state has converged on its value. □

We now show that a solution to MOMI(MR) corresponds to a solution to MR. For the remainder of this section let $\gamma_A^{\mathcal{O}}(A_i) := \{a : O(S_{ja}, S_j) \in \mathcal{O}\}$ and $\gamma_B^{\mathcal{O}}(B_j) := \{b : O(S_{jb}, S_j) \in \mathcal{O}\}$ where for the remainder of this section $O(S, S')$ stands for a point option with initiation state $S$ and termination state $S'$.

**Lemma 5.** *For any feasible solution $\mathcal{O}$ to* MOMI(*MR*) *we have $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution to MR of cost $|\mathcal{O}|$.*

*Proof.* Notice that by construction the Min-Rep cost of $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is exactly $|\mathcal{O}|$. Thus, we need only prove that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution for MR.

We do so now. Consider an arbitrary edge $e = (A_i, B_j) \in E$; we wish to show that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ satisfies $e$. Since $\mathcal{O}$ is a feasible solution we know that after two iterations of value iteration every state must converge on its value (up to an $\epsilon$ factor which we can ignore by our above assumption that $\varepsilon < .5$). Moreover, notice that the value of every state in MOMI(MR) is 1. Thus, it must be the case that for every $S_e$ we have $V_2(S_e) = 1$ for $e = (A_i, B_j)$. It follows, then, that there is some action $\alpha_{(\bar{a},\bar{b})}$ where $(\bar{a}, \bar{b}) \in \pi_{(A_i,B_j)}$ such that

$$1 = V_2(S_e) = .5 \cdot (V_1(S_{i\bar{a}})) + .5 \cdot (V_1(S_{j\bar{b}})).$$

Since the value of every state is at most 1, it follows that $V_1(S_{i\bar{a}}) = V_1(S_{j\bar{b}}) = 1$. However, since $V_1(S_{i\bar{a}})$ and $V_1(S_{j\bar{b}})$ are both two hops from the only goal reachable from them ($S_i$ and $S_j$ respectively) it must be the case that there is some point option from $S_{i\bar{a}}$ to $S_i$ and $S_{j\bar{b}}$ to $S_j$. Thus, by definition of $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ we then have $\bar{a} \in \gamma_A^{\mathcal{O}}$ and $\bar{b} \in \gamma_B^{\mathcal{O}}$. Since $(\bar{a}, \bar{b}) \in \pi_{(A_i,B_j)}$ it follows that arbitrary edge $e = (A_i, B_j)$ is satisfied. Thus, every edge in $E$ is satisfied and so $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a feasible solution for MR. □

Finally, we conclude the hardness of approximation of MOMI.

*Theorem 4.3 Proof.* Assume NP $\not\subseteq$ DTIME($n^{\text{poly}\log n}$) and for the sake of contradiction that there exists an $\varepsilon > 0$ for which a polynomial-time algorithm $\mathcal{A}_{\text{MOMI}}$ can $2^{\log^{1-\varepsilon} n}$-approximate MOMI. We use $\mathcal{A}_{\text{MOMI}}$ to $2^{\log^{1-\varepsilon'} \tilde{n}}$ approximate Min-Rep for a fixed constant $\varepsilon' > 0$ in polynomial-time in $\tilde{n}$, thereby contradicting Lemma 1. Again, $\tilde{n}$ is the number of vertices in the graph of the Min-Rep instance.

We begin by noting that the relevant quantities in MOMI(MR) are polynomially-bounded. Let $\tilde{n} := |A| + |B|$ be the number of vertices in our MR instance. Notice that the number of states in the MDP, $n$, in our MOMI(MR) instance is at most $O(\tilde{n} + 2|A||\Sigma_A| + |B||\Sigma_B| + |E|) \leq \tilde{n}^c$ for some fixed constant $c$ by the aforementioned assumption that $\Sigma_A$ and $\Sigma_B$ are polynomially-bounded in $\tilde{n}$.[7]

Our polynomial-time approximation algorithm to approximate instance MR of Min-Rep is as follows: Run $\mathcal{A}_{\text{MOMI}}$ on MOMI(MR) to get back option set $\mathcal{O}$. Return $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ as defined above as our solution to MR.

We first argue that our algorithm is polynomial time in $\tilde{n}$. For each vertex in MR, we create a polynomial number of states and actions. Thus, the number of states in MOMI(MR) is polynomially-bounded in $\tilde{n}$ and so $\mathcal{A}_{\text{MOMI}}$ runs in time polynomial in $\tilde{n}$. A polynomial runtime of our algorithm immediately follows.

We now argue that our algorithm is a $2^{\log^{1-\varepsilon'} \tilde{n}}$-approximation for Min-Rep for some $\varepsilon' > 0$. Applying Lemma 5, the approximation of $\mathcal{A}_{\text{MOMI}}$ and then Lemma 4, we have that the Min-Rep cost of $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is

$$\begin{aligned} \text{cost}_{\text{Min-Rep}}(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}}) &= |\mathcal{O}| \\ &\leq 2^{\log^{1-\varepsilon} n} \text{OPT}_{\text{MOMI}} \\ &\leq 2^{\log^{1-\varepsilon} n} \text{OPT}_{\text{MR}} \end{aligned}$$

Thus, $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a $2^{\log^{1-\varepsilon} n}$ approximation for the opti-

[7] It is worth noting, also, that since we create at most $\sum_e |\pi_e|$ actions for any state, the number of total actions in our MDP is at most polynomial in $\tilde{n}$.

mal Min-Rep solution where $n$ is the number of states in the MDP of MOMI(MR). Now recalling that $n \leq \tilde{n}^c$ for fixed constant $c$. We therefore have that $(\gamma_A^{\mathcal{O}}, \gamma_B^{\mathcal{O}})$ is a $2^{\log^{1-\varepsilon} \tilde{n}^c} = 2^{c^{1-\varepsilon} \log^{1-\varepsilon} \tilde{n}} \leq c' \cdot 2^{\log^{1-\varepsilon} \tilde{n}}$ approximation for a constant $c'$. Choosing $\varepsilon$ sufficiently small, we have that $c' \cdot 2^{\log^{1-\varepsilon} \tilde{n}} \leq 2^{\log^{1-\varepsilon'} \tilde{n}}$ for sufficiently large $\tilde{n}$.

Thus, our polynomial-time algorithm is a $2^{\log^{1-\varepsilon'} \tilde{n}}$-approximation for Min-Rep for $\varepsilon' > 0$, thereby contradicting Lemma 1. We conclude that MOMI cannot be $2^{\log^{1-\epsilon} n}$-approximated. □

### A.4. A-MIMO

In this subsection we show the following theorem (we show Theorem 5 later):

**Theorem 6.** A-MIMO *has following properties:*

1. A-MIMO *runs in polynomial time.*
2. *If the MDP is deterministic, it has a bounded suboptimality of* $O(\log^* k)$.
3. *The number of iterations to solve the MDP using the acquired options is upper bounded by* $P(\mathcal{C})$.

**Theorem 6.1.** A-MIMO *runs in polynomial time.*

*Proof.* Each step of the procedure runs in polynomial time.

(1) Solving an MDP takes polynomial time. To compute $d$ we need to solve MDPs at most $|\mathcal{S}|$ times. Thus, it runs in polynomial time.

(2) The approximation algorithm we deploy for solving the asymmetric-k center which runs in polynomial time (Archer, 2001). Because the procedure by Archer (2001) terminates immediately after finding a set of options which guarantees the suboptimality bounds, it tends to find a set of options smaller than $k$. In order to use the rest of the options effectively within polynomial time, we use a procedure Expand to greedily add a few options at once until it finds all $k$ options. We enumerate all possible set of options of size $r = \lceil \log k \rceil$ (if $|\mathcal{O}| + \log k > k$ then we set $r = k - |\mathcal{O}|$) and add a set of options which minimizes $\ell$ (breaking ties randomly) to the option set $\mathcal{O}$. We repeat this procedure until $|\mathcal{O}| = k$. This procedure runs in polynomial time. The number of possible option set of size $r$ is ${}_rC_n = O(n^r) = O(k)$. We repeat this procedure at most $\lceil k / \log k \rceil$ times, thus the total computation time is bounded by $O(k^2 / \log k)$.

(3) Immediate.

Therefore, A-MIMO runs in polynomial time. □

Before we show that it is sufficient to consider a set of options with its terminal state set to the goal state of the MDP.

**Lemma 6.** *There exists an optimal option set for MIMO and MOMI with all terminal state set to the goal state.*

*Proof.* Assume there exists an option with terminal state set to a state other than the goal state in the optimal option set $\mathcal{O}$. By triangle inequality, swapping the terminal state to the goal state will monotonically decrease $d(s, g)$ for every state. By swapping every such option we can construct an option set $\mathcal{O}'$ with $L_{\epsilon, V_0}(\mathcal{O}') \leq L_{\epsilon, V_0}(\mathcal{O})$. □

Lemma imply that discovering the best option set among option sets with their terminal state fixed to the goal state is sufficient to find the best option set in general. Therefore, our algorithms seek to discover options with termination state fixed to the goal state.

Using the option set acquired, the number of iterations to solve the MDP is bounded by $P(\mathcal{C})$. To prove this we first generalize the definition of the distance function to take a state and a set of states as arguments $d_\epsilon : \mathcal{S} \times 2^{\mathcal{S}} \to \mathbb{N}$. Let $d_\epsilon(s, \mathcal{C})$ the number of iterations for $s$ to converge $\epsilon$-optimal if every state $s' \in \mathcal{C}$ has converged to $\epsilon$-optimal: $d_\epsilon(s, \mathcal{C}) := \min(d'_\epsilon(s), 1 + d'_\epsilon(s, \mathcal{C})) - 1$. As adding an option will never make the number of iterations larger,

**Lemma 7.**

$$d(s, \mathcal{C}) \leq \min_{s' \in \mathcal{C}} d(s, s'). \tag{1}$$

Using this, we show the following proposition.

**Theorem 6.2.** *The number of iterations to solve the MDP using the acquired options is upper bounded by* $P(\mathcal{C})$.

*Proof.* $P(\mathcal{C}) = \max_{s \in S} \min_{c \in \mathcal{C}} d(s, c) \geq \max_{s \in S} d(s, \mathcal{C}) = L_{\epsilon, V_0}(\mathcal{O})$ (using Equation 1). Thus $P(\mathcal{C})$ is an upper bound for $L_{\epsilon, V_0}(\mathcal{O})$. □

The reason why $P(\mathcal{C})$ does not always give us the exact number of iterations is because adding two options starting from $s_1, s_2$ may make the convergence of $s_0$ faster than $d(s_0, s_1)$ or $d(s_0, s_2)$. Example: Figure 4 is an example of such an MDP. From $s_0$ it may transit to $s_1$ and $s_2$ with probability 0.5 each. Without any options, the value function converges to exactly optimal value for every state with 3 steps. Adding an option either from $s_1$ or $s_2$ to $g$ does not shorten the iteration for $s_0$ to converge. However, if we add two options from $s_1$ and $s_2$ to $g$, $s_0$ converges within 2 steps, thus the MDP is solved with 2 steps.

The equality of the statement 1 holds if the MDP is deterministic. That is, $d(s, \mathcal{C}) = \min_{s' \in \mathcal{C}} d(s, s')$ for deterministic MDP.

**Theorem 6.3.** *If the MDP is deterministic, it has a bounded suboptimality of* $O(\log^* k)$.

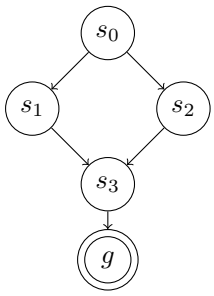



Figure 4: An example of an MDP where $d(s, \mathcal{C}) < \min_{s' \in \mathcal{C}} d(s, s')$. Here the transition induced by the optimal policy is stochastic, thus from $s_0$ one may go to $s_1$ and $s_2$ by probability 0.5 each. Either adding an option from $s_1$ or $s_2$ to $g$ does not make the convergence faster, but adding both makes it faster.

*Proof.* First we show $P(\mathcal{C}^*) = L_{\epsilon,V_0}(\mathcal{O}^*)$ for deterministic MDP. From $d(s, \mathcal{C}) = \min_{s' \in \mathcal{C}} d(s, s')$, $P(\mathcal{C}^*) = \max_{s \in \mathcal{S}} \min_{c \in \mathcal{C}^*} d(s, c) = \max_{s \in \mathcal{S}} d(s, \mathcal{C}^*) = L_{\epsilon,V_0}(\mathcal{O}^*)$.

The asymmetric $k$-center solver guarantees that the output $\mathcal{C}$ satisfies $P(\mathcal{C}) \leq c(\log^* k + O(1))P(\mathcal{C}^*)$ where $n$ is the number of nodes (Archer, 2001). Let MIMO$(M, \epsilon, k)$ be an instance of MIMO. We convert this instance to an instance of asymmetric $k$-center AsymKCenter$(\mathcal{U}, d, k)$, where $|\mathcal{U}| = |\mathcal{S}|$. By solving the asymmetric $k$-center with the approximation algorithm, we get a solution $\mathcal{C}$ which satisfies $P(\mathcal{C}) \leq c(\log^* k + O(1))P(\mathcal{C}^*)$. Thus, the output of the algorithm $\mathcal{O}$ satisfies $L_{\epsilon,V_0}(\mathcal{O}) = P(\mathcal{C}) \leq c(\log^* k + O(1))P(\mathcal{C}^*) = c(\log^* k + O(1))L_{\epsilon,V_0}(\mathcal{O}^*)$. Thus, $L_{\epsilon,V_0}(\mathcal{O}) \leq c(\log^* k + O(1))L_{\epsilon,V_0}(\mathcal{O}^*)$ is derived. □

**Proposition 1** (Greedy Strategy)**.** *Let an option set $\mathcal{O}$ be a set of point option constructed by greedily adding one point option which minimizes the number of iterations. An improvement $L_{\epsilon,V_0}(\emptyset) - L_{\epsilon,V_0}(\mathcal{O})$ by the greedy algorithm can be arbitrary small (i.e. 0) compared to the optimal option set.*

*Proof.* We show by the example in a shortest-path problem in Figure 5. The MDP can be solved within 4 iterations without options: $L_{\epsilon,V_0}(\emptyset) = 4$. With an optimal option set of size $k = 2$ the MDP can be solved within 2 iterations: $L_{\epsilon,V_0}(\mathcal{O}^*) = 2$ (an initiation state of each option in optimal option set is denoted by $*$ in the Figure). On the other hand, a greedy strategy may not improve $L$ at all. No single point option does not improve $L$. Let's say we picked a point option from $s_1$ to $g$. Then, there is no single point option we can add to that option to improve $L$ in the second iteration. Therefore, the greedy procedure returns $\mathcal{O}$ which has $L_{\epsilon,V_0}(\emptyset) - L_{\epsilon,V_0}(\mathcal{O}) = 0$. Therefore, $(L_{\epsilon,V_0}(\emptyset) - L_{\epsilon,V_0}(\mathcal{O}))/(L_{\epsilon,V_0}(\emptyset) - L_{\epsilon,V_0}(\mathcal{O}^*))$ can be arbitrary small non-negative value (i.e. 0).

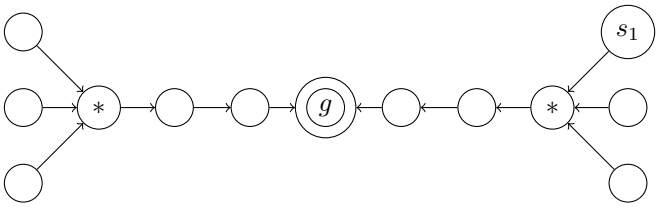



Figure 5: Example of MIMO where the improvement of a greedy strategy can be arbitrary small compared to the optimal option set.

□

### A.5. A-MOMI

In this subsection we show the following theorem:

**Theorem 5.** A-MOMI *has the following properties:*

1. A-MOMI *runs in polynomial time.*
2. *It guarantees that the MDP is solved within $\ell$ iterations using the option set acquired by* A-MOMI *$\mathcal{O}$.*
3. *If the MDP is deterministic, the option set is at most $O(\log n)$ times larger than the smallest option set possible to solve the MDP within $\ell$ iterations.*

**Theorem 5.1.** A-MOMI *runs in polynomial time.*

*Proof.* Each step of the procedure runs in polynomial time.

(1) Solving an MDP takes polynomial time (Littman et al., 1995). To compute $d$ we need to solve MDPs at most $|\mathcal{S}|$ times. Thus, it runs in polynomial time.

(4) We solve the set cover using a polynomial time approximation algorithm (Chvatal, 1979) which runs in $O(n^3)$, thus run in polynomial time.

(2), (3), and (5) Immediate. □

**Theorem 5.2.** A-MOMI *guarantees that the MDP is solved within $\ell$ iterations using the option set $\mathcal{O}$.*

*Proof.* A state $s \in X_g^+$ reaches optimal within $\ell$ steps by definition. For every state $s \in \mathcal{S} \setminus X_g^+$, the set cover guarantees that we have $X_{s'} \in \mathcal{C}$ such that $d(s, s') < \ell$. As we generate an option from $s'$ to $g$, $s'$ reaches to optimal value with 1 step. Thus, $s$ reaches to $\epsilon$-optimal value within $d(s, s') + 1 \leq \ell$. Therefore, every state reaches $\epsilon$-optimal value within $\ell$ steps. □

**Theorem 5.3.** *If the MDP is deterministic, the option set is at most $O(\log n)$ times larger than the smallest option set possible to solve the MDP within $\ell$ iterations.*

*Proof.* Using a suboptimal algorithm by Chvatal (1979) we get $\mathcal{C}$ such that $|\mathcal{C}| \leq O(\log n)|\mathcal{C}^*|$. Thus, $|\mathcal{O}| = |\mathcal{C}| \leq O(\log n)|\mathcal{C}^*| = O(\log n)|\mathcal{O}^*|$. □

## Appendix: Experiments

We show the figures for experiments. Figure 6 shows the options found by solving MIMO optimally/suboptimally in four room domain. Figure 7 shows the options in 9x9 grid domain.

## References


Archer, A. Two O(log* k)-approximation algorithms for the asymmetric k-center problem. In *International Conference on Integer Programming and Combinatorial Optimization*, pp. 1–14, 2001.

Bhattacharyya, A., Grigorescu, E., Jung, K., Raskhodnikova, S., and Woodruff, D. P. Transitive-closure spanners. *SIAM Journal on Computing*, 41(6):1380–1425, 2012.

Chvatal, V. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3):233–235, 1979.

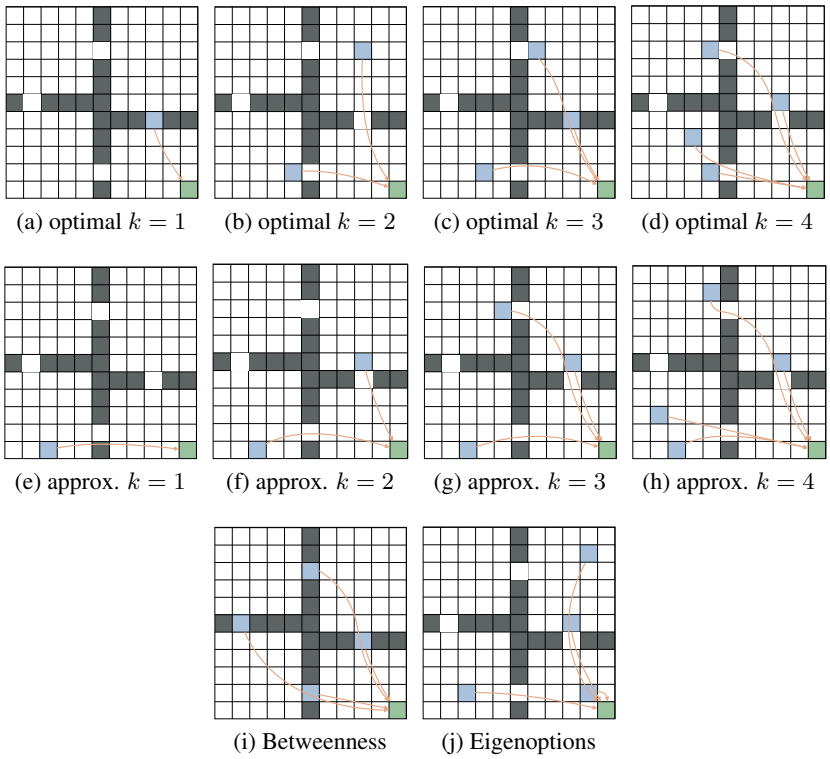
Dinitz, M., Kortsarz, G., and Raz, R. Label cover instances with large girth and the hardness of approximating basic k-spanner. In *International Colloquium on Automata, Languages, and Programming*, pp. 290–301. Springer, 2012.

Dinur, I. and Safra, S. On the hardness of approximating label-cover. *Information Processing Letters*, 89(5):247–254, 2004.

Kortsarz, G. On the hardness of approximating spanners. *Algorithmica*, 30(3):432–450, 2001.

Littman, M. L., Dean, T. L., and Kaelbling, L. P. On the complexity of solving Markov decision problems. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 394–402, 1995.

Raz, R. and Safra, S. A sub-constant error-probability low-degree test, and a sub-constant error-probability pcp characterization of np. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pp. 475–484. ACM, 1997.

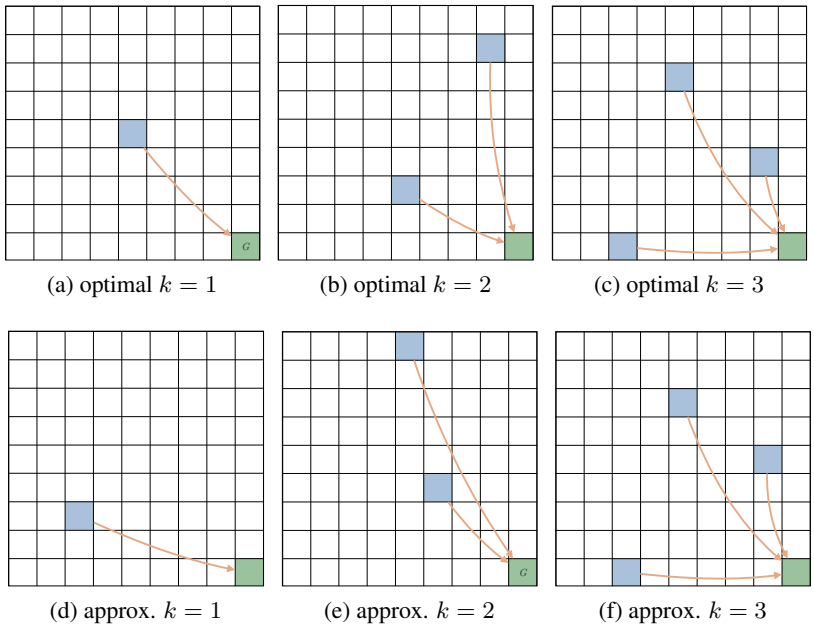


(a) optimal $k = 1$ (b) optimal $k = 2$ (c) optimal $k = 3$ (d) optimal $k = 4$

(e) approx. $k = 1$ (f) approx. $k = 2$ (g) approx. $k = 3$ (h) approx. $k = 4$

(i) Betweenness (j) Eigenoptions

Figure 6: Comparison of the optimal point options vs. options generated by the approximation algorithm A-MIMO. We observed that the approximation algorithm is similar to that of optimal options. Note that optimal option set is not unique: there can be multiple optimal option set, and we are visualize one of them returned by the solver.




(a) optimal $k = 1$ (b) optimal $k = 2$ (c) optimal $k = 3$

(d) approx. $k = 1$ (e) approx. $k = 2$ (f) approx. $k = 3$

Figure 7: Comparison of the optimal point options for planning vs. bottleneck options proposed for reinforcement learning in the four room domain. Initiating conditions are shown in blue, the goal in green.