Cost Partitioning Heuristics for Stochastic Shortest Path Problems Technical Report CS-2022-001

Thorsten Klößner[1], Florian Pommerening[2], Thomas Keller[2], Gabriele Röger[2]

[1]Saarland University, Saarland Informatics Campus, Germany
[2]University of Basel, Switzerland
kloessner@cs.uni-saarland.de, {florian.pommerening, tho.keller, gabriele.roeger}@unibas.ch

This technical report contains full proofs for claims in our ICAPS 2022 paper "Cost Partitioning Heuristics for Stochastic Shortest Path Problems". Notation and equation references refer to their definitions in the full paper.

Admissibility of Abstraction Heuristics

In the paper, we make the claim that the abstraction heuristic $h^\alpha(s, c')$ is a generally admissible heuristic for a class of abstraction mappings $\alpha : S \to S_\alpha$. We assume $c' = c$ in the following and prove admissibility, otherwise we can simply consider the SSP where the original cost function is changed to c'. Let $\tilde{s} \in S$ in the following.

To make the notation easier, we denote the version of (LP 1) where states and actions irrelevant for $\tilde{s}$ are pruned by (LP 1'). Recall that this LP computes $J^\star(\tilde{s})$, even in presence of negative-cost transition cycles. To prove admissibility, we transform a solution to (LP 1') for the abstract SSP (computing $J^\star(\alpha(\tilde{s}))$) to a solution of (LP 1') for the original SSP with equal objective value. This shows the claim because both LPs are maximization problems and therefore $J^\star(\alpha(\tilde{s})) \leq J^\star(\tilde{s})$.

Theorem 1 *Let α be an abstraction mapping. Let y be a solution to (LP 1') for the abstraction. Then $\mathsf{y}'_s := \mathsf{y}_{\alpha(s)}$ is a solution to (LP 1') for the original state space with equal objective value.*

Proof. We first need to argue that this assignment is well-defined, i.e., if s is relevant in the original state space, then $\alpha(s)$ is relevant in the abstract state space. Let $s \in S$ be a relevant state. By definition of relevance, there is a non-stationary $\tilde{s}$-proper policy π which has a possible path $\vec{p} = \langle s_1, \pi(s_1), s_2, \pi(s_2), \ldots, \pi(s_n), s_{n+1}\rangle$ from $s_1 = \tilde{s}$ to $s_{n+1} = s$. From $\vec{p}$, we can construct the corresponding abstract path $\vec{p}_\alpha = \langle \alpha(s_1), \pi(s_1), \alpha(s_2), \pi(s_2), \ldots, \pi(s_n), \alpha(s_{n+1})\rangle$ from $\alpha(s_1) = \alpha(\tilde{s})$ to $\alpha(s_{n+1}) = \alpha(s)$. Using the definition of the transition relation in the abstraction, it is easy to see that $\vec{p}_\alpha$ is possible in the abstraction. We now construct a non-stationary $\alpha(\tilde{s})$-proper abstract policy π_α which may execute $\vec{p}_\alpha$. This will show that $\alpha(s)$ is relevant for $\alpha(\tilde{s})$.

First, we choose any abstract stationary policy which maximizes the probability to reach the abstract goal for every abstract state. Note that such a policy always exists (Puterman 1994). Next, we make this policy non-stationary and obtain π_α by instead choosing $\pi(s_i)$ in $\alpha(s_i)$ $(1 \leq i \leq n)$ for the first n steps. Obviously, π_α may execute $\vec{p}_\alpha$. To see that π_α is $\alpha(\tilde{s})$-proper, note that the maximal goal probability of an state $s \in S$ is less or equal to the maximal goal probability of its abstract state $\alpha(s)$ in the abstraction (Klößner et al. 2021). The concrete goal can be reached with certainty from s_i since π is $\tilde{s}$-proper, so this must also be the case for $\alpha(s_i)$ with respect to the abstract goal, where $1 \leq i \leq n+1$. Similarly, this holds for the immediate successors of $\alpha(s_i)$ when applying $\pi(s_i)$. Hence, this policy can only reach abstract states from which the abstract goal is reachable with certainty during the first n steps. After $n+1$ steps, the policy maximizes the goal probability by construction, so all in all the policy will reach the abstract goal with certainty. Thus the policy is $\alpha(\tilde{s})$-proper and $\alpha(s)$ is relevant.

With well-definedness of y'_s ensured, we have for all relevant states s and actions $a \in A(s)$:

$$\mathsf{y}'_s = \mathsf{y}_{\alpha(s)} \leq c_\alpha(\alpha(s), a) + \sum_{\tau \in S_\alpha} T_\alpha(\alpha(s'), a, \tau)\mathsf{y}_\tau$$

due to constraints (2) of (LP 1') for the abstraction α. We continue by applying the definition of c_α to obtain

$$\cdots = \left(\min_{\substack{s' \in \alpha^{-1}(\alpha(s)) \\ \text{s.t. } a \in A(s')}} c(s', a) \right) + \sum_{\tau \in S_\alpha} T_\alpha(\alpha(s), a, \tau)\mathsf{y}_\tau \leq c(s, a) + \sum_{\tau \in S_\alpha} T_\alpha(\alpha(s), a, \tau)\mathsf{y}_\tau.$$

Lastly, we apply the definition of T_α and get

$$\cdots = c(s, a) + \sum_{\tau \in S_\alpha} \left(\sum_{t \in \alpha^{-1}(\tau)} T(s, a, t) \right) \mathsf{y}_\tau = c(s, a) + \sum_{t \in S} T(s, a, t)\mathsf{y}_{\alpha(t)} = c(s, a) + \sum_{t \in S} T(s, a, t)\mathsf{y}'_t.$$

Putting all things together, we have $\mathsf{y}'_s \leq c(s,a) + \sum_{t \in S} T(s,a,t)\mathsf{y}'_t$, so constraints (2) are satisfied. For constraint (1), we have $\mathsf{y}'_{s_\star} = \mathsf{y}_{\alpha(s_\star)} = 0$. Finally, the objective value remains the same with $\mathsf{y}_{\tilde{s}} = \mathsf{y}_{\alpha(\tilde{s})}$. □

Corollary 1 *For any abstraction mapping α subject to restrictions defined in the paper, h^α is generally admissible.*

Approximate Linear Programming & Potential Heuristics

Guestrin et al. define Approximate Linear Programming (ALP) on infinite-horizon discounted-reward MDPs. Such an MDP is a tuple $\langle \mathcal{X}, A, \mathcal{R}, \mathcal{P} \rangle$, where $\mathcal{X}$ is a finite set of states, A is a finite set of actions, $\mathcal{R} : \mathcal{X} \times A \to \mathbb{R}$ is a reward function, and $\mathcal{P}(t \mid s, a)$ specifies the probability of ending in state t when executing action a in state s. The reward is discounted with a factor γ and bounded by a maximal reward $R_{\max}$, which will not be important in the following.
MDPs can be compiled into SSPs. Let $M = \langle \mathcal{X}, A, \mathcal{R}, \mathcal{P} \rangle$ be an MDP and s_0 be one of its states. We construct the SSP $\Pi(M, s_0) = \langle S, A, T, s_0, s_\star, c \rangle$ with

- $S = \mathcal{X} \cup \{s_g\}$ where s_g is a fresh artificial goal state
- $T(s,a,t) = \begin{cases} \gamma \cdot \mathcal{P}(t \mid s, a) & \text{if } s \in \mathcal{X} \text{ and } t \in \mathcal{X} \\ 1 - \gamma & \text{if } s \in \mathcal{X} \text{ and } t = s_g \\ 0 & \text{if } s = s_g \end{cases}$
- $s_\star = s_g$
- $c(s,a) = -\mathcal{R}(s,a)$

Note that because of the discounted reward, the expected value of all states in the MDP is finite. In the compiled SSP every action has a non-zero chance of ending in the goal state, so there cannot be any cycles of negative costs that can be repeated arbitrarily often. This means that even with negative action costs, the SSP cannot have a value of $-\infty$.
We will now first introduce approximate linear programming as defined for MDPs by Guestrin et al. and then define it for SSPs. We will then show that our definition for SSPs is a proper generalization of the definition for MDPs, in the sense that ALP computed on an MDP M gives an equivalent result for state s_0 as ALP computed on the SSP $\Pi(M, s_0)$.

ALP for MDPs. Approximate linear programming for an MDP $\langle \mathcal{X}, A, \mathcal{R}, \mathcal{P} \rangle$ is defined over a set of *basis functions* $\mathcal{F} = \{f_1, \ldots, f_n\}$ and a *state relevance function* ρ. For consistency with our paper we slightly deviate from the notation of Guestrin et al. here, who use h for the basis functions and α for the state relevance function. ALP optimizes variables w_f with the following LP.

$$\text{Minimize } \sum_{s \in \mathcal{X}} \rho(s) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) \text{ subject to}$$

$$\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) \geq \mathcal{R}(s,a) + \gamma \sum_{t \in \mathcal{X}} \mathcal{P}(t \mid s, a) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(t) \qquad \text{for all } s \in \mathcal{X} \text{ and } a \in A \tag{17}$$

where all variables $(\mathsf{w}_f)_{f \in \mathcal{F}}$ are unrestricted.

ALP for SSPs. We now specify a similar LP based on an SSP $\langle S, A, T, s_0, s_\star, c \rangle$.

$$\text{Maximize } \sum_{s \in S} \rho(s) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) \text{ subject to}$$

$$\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s_\star) = 0 \tag{18}$$

$$\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) \leq c(s,a) + \sum_{t \in S} T(s,a,t) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(t) \qquad \text{for all } s \in S \text{ and } a \in A(s) \tag{19}$$

where all variables $(\mathsf{w}_f)_{f \in \mathcal{F}}$ are unrestricted.

Theorem 2 *Let M be an MDP, $\mathcal{F}$ a set of basis functions defined on the states of M, and ρ a state-relevance function. Further, let s_0 be a state of M and $s_\star$ the artificial goal state used in $\Pi(M, s_0)$. Consider the set $\mathcal{F}'$ of basis functions that are the functions in $\mathcal{F}$ extended with $\{s_\star \mapsto 0\}$ and ρ' as ρ extended with $\{s_\star \mapsto 0\}$.*
Then w is an optimal solution for ALP on M with basis functions $\mathcal{F}$ and state-relevance function ρ if and only if $-\mathsf{w}$ is an optimal solution for ALP on $\Pi(M, s_0)$ with basis functions $\mathcal{F}'$ and state-relevance function ρ'.

Proof. Consider a solution w that satisfies (17). Constraint (18) is satisfied by any solution as $f'(s_\star) = 0$ for all $f' \in \mathcal{F}'$. To see that $\mathsf{w}' = -\mathsf{w}$ satisfies (19), first note that no action is applicable in $s_\star$ and all actions are applicable in all other states,

so constraints (17) and (19) quantify over the same states and actions. We now plug in the definitions from the compilation $\Pi(M, s_0)$ for an action a and a state s.

$$\begin{aligned}\sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(s) = -\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) &\overset{(17)}{\leq} -\mathcal{R}(s,a) - \gamma \sum_{t \in \mathcal{X}} \mathcal{P}(t \mid s, a) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(t) \\ &= c(s,a) + \sum_{t \in \mathcal{X}} T(s,a,t) \sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(t) \\ &\overset{(18)}{=} c(s,a) + \sum_{t \in S} T(s,a,t) \sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(t)\end{aligned}$$

For the other direction, consider a solution w' that satisfies (18)–(19). To see that $\mathsf{w} = -\mathsf{w}'$ satisfies (17), we again plug in the definitions from the compilation $\Pi(M, s_0)$ for an action a and a state s.

$$\begin{aligned}\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s) = -\sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(s) &\overset{(19)}{\geq} -c(s,a) - \sum_{t \in S} T(s,a,t) \sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(t) \\ &\overset{(18)}{=} -c(s,a) - \sum_{t \in \mathcal{X}} T(s,a,t) \sum_{f' \in \mathcal{F}'} \mathsf{w}'_{f'} f'(t) \\ &= \mathcal{R}(s,a) + \gamma \sum_{t \in \mathcal{X}} \mathcal{P}(t \mid s, a) \sum_{f \in \mathcal{F}} \mathsf{w}_f f(t)\end{aligned}$$

In summary, we have shown that w satisfies (17) iff $-\mathsf{w}$ satisfies (18)–(19). As the objective functions of the two LPs are identical, w is a minimal solution iff $-\mathsf{w}$ is a maximal one. □

Simplifying ALP for Indicator Functions. In the paper, we focused our attention to basis functions that are indicator functions of abstract states, i.e., for an abstraction $\alpha : S \to S^\alpha$, we consider one basis function f_σ for each abstract state $\sigma \in S^\alpha$ that is defined as

$$f_\sigma(s) = \begin{cases} 1 & \text{if } \alpha(s) = \sigma \\ 0 & \text{otherwise.} \end{cases}$$

We consider a set of abstractions $\mathcal{A}$ and assume that the sets of abstract states are pairwise disjoint so an abstract state uniquely identifies its abstraction. For each state s and each abstraction α, exactly one of the basis functions associated with α has the value 1 while all others have the value 0. In the paper, we claim that this simplifies the LP computed for ALP as follows.

$$\begin{aligned}&\text{Maximize } \sum_{s \in S} \rho(s) \sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(s)} \text{ subject to} \\ &\quad \mathsf{w}_{\alpha(s_\star)} = 0 && \text{for all } \alpha \in \mathcal{A} && (20) \\ &\quad \sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(s)} \leq c(s,a) + \sum_{t \in S} T(s,a,t) \sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(t)} && \text{for all } s \in S \text{ and } a \in A(s) && (21)\end{aligned}$$

where all variables are unrestricted.

It is easy to see that $\sum_{f \in \mathcal{F}} \mathsf{w}_f f(s)$ simplifies to $\sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(s)}$ for our choice of basis functions and thus (19) simplifies to (21). Likewise, it is easy to see that (20) implies (18) because all elements in the sum of (18) will be 0 either because $\mathsf{w}_\sigma = 0$ (for abstract goal states $\sigma = \alpha(s_\star)$) or because $f_\sigma(s_\star) = 0$ (for all other abstract states σ). It remains to show that using the stronger constraint (20) instead of (18) does not exclude any optimal solution. For this, we consider a solution w that satisfies (18)–(19). As shown above, constraint (19) can be rewritten to

$$\sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(s)} \leq c(s,a) + \sum_{t \in S} T(s,a,t) \sum_{\alpha \in \mathcal{A}} \mathsf{w}_{\alpha(t)}$$

which is equivalent to

$$\sum_{\alpha \in \mathcal{A}} \left(\mathsf{w}_{\alpha(s)} - \sum_{t \in S} T(s,a,t) \mathsf{w}_{\alpha(t)} \right) \leq c(s,a).$$

Shifting all weights of one abstraction by a constant b has no effect in this constraint:

$$\begin{aligned}\sum_{\alpha\in\mathcal{A}}\left((\mathsf{w}_{\alpha(s)}+b)-\sum_{t\in S}T(s,a,t)(\mathsf{w}_{\alpha(t)}+b)\right) &= \sum_{\alpha\in\mathcal{A}}\left(\mathsf{w}_{\alpha(s)}+b-\sum_{t\in S}T(s,a,t)\mathsf{w}_{\alpha(t)}-b\sum_{t\in S}T(s,a,t)\right)\\ &= \sum_{\alpha\in\mathcal{A}}\left(\mathsf{w}_{\alpha(s)}+b-\sum_{t\in S}T(s,a,t)\mathsf{w}_{\alpha(t)}-b\right)\\ &= \sum_{\alpha\in\mathcal{A}}\left(\mathsf{w}_{\alpha(s)}-\sum_{t\in S}T(s,a,t)\mathsf{w}_{\alpha(t)}\right)\end{aligned}$$

Thus, for every solution w that satisfies (18)–(19), we can construct the solution w' as $\mathsf{w}'_{\alpha(s)} = \mathsf{w}_{\alpha(s)} - \mathsf{w}_{\alpha(s_\star)}$. This still satisfies (19) as shown above and it satisfies (20) which implies (18). The objective value of the new solution w' is the same as for w:

$$\sum_{s\in S}\rho(s)\sum_{\alpha\in\mathcal{A}}\mathsf{w}'_{\alpha(s)} = \sum_{s\in S}\rho(s)\left(\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(s)}-\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(s_\star)}\right)\overset{(20)}{=}\sum_{s\in S}\rho(s)\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(s)}$$

Equivalence of ALP and Transition Cost Partitioning. The next step in the paper shows that using ALP for indicator functions of abstract states as in (20)–(21) computes a transition cost partitioning optimized according to state-relevance function ρ. The LP optimizing this cost partitioning is:

$$\begin{aligned}&\text{Maximize } \sum_{s\in S}\rho(s)\sum_{\alpha\in\mathcal{A}}\mathsf{y}_{\alpha(s)} \text{ subject to}\\ &\quad \mathsf{y}_{\alpha(s_\star)} = 0 && \text{for all } \alpha\in\mathcal{A} && (22)\\ &\quad \mathsf{y}_{\alpha(s)} \le \mathsf{c}_{\alpha sa} + \sum_{t\in S}T(s,a,t)\mathsf{y}_{\alpha(t)} && \text{for all } \alpha\in\mathcal{A},\ s\in S \text{ and } a\in A(s) && (23)\\ &\quad \sum_{\alpha\in\mathcal{A}}\mathsf{c}_{\alpha sa} \le c(s,a) && \text{for all } s\in S \text{ and } a\in A(s) && (24)\\ &\text{where all variables are unrestricted.}\end{aligned}$$

Theorem 3 *The LPs* (20)–(21) *and* (22)–(24) *are equivalent.*

Proof. For any solution w satisfying constraints (20)–(21) consider $\mathsf{y} = \mathsf{w}$ and $\mathsf{c}_{\alpha sa} = \mathsf{w}_{\alpha(s)} - \sum_{t\in S}T(s,a,t)\mathsf{w}_{\alpha(t)}$. With these values, constraint (20) implies (22) and constraint (23) trivializes to $\mathsf{w}_{\alpha(s)} \le \mathsf{w}_{\alpha(s)}$. We can show that constraint (24) is also satisfied:

$$\sum_{\alpha\in\mathcal{A}}\mathsf{c}_{\alpha sa} = \sum_{\alpha\in\mathcal{A}}(\mathsf{w}_{\alpha(s)}-\sum_{t\in S}T(s,a,t)\mathsf{w}_{\alpha(t)}) = \sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(s)}-\sum_{t\in S}T(s,a,t)\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(t)}\overset{(21)}{\le} c(s,a)$$

For the other direction, consider a solution y,c satisfying constraints (22)–(24) and define $\mathsf{w} = \mathsf{y}$. Clearly, (20) is implied by (22). To see that (21) is satisfied, first we sum (23) for all abstractions and then use (24):

$$\begin{aligned}\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(s)} = \sum_{\alpha\in\mathcal{A}}\mathsf{y}_{\alpha(s)} &\overset{(23)}{\le} \sum_{\alpha\in\mathcal{A}}(\mathsf{c}_{\alpha sa}+\sum_{t\in S}T(s,a,t)\mathsf{y}_{\alpha(t)})\\ &= \sum_{\alpha\in\mathcal{A}}\mathsf{c}_{\alpha sa}+\sum_{t\in S}T(s,a,t)\sum_{\alpha\in\mathcal{A}}\mathsf{y}_{\alpha(t)}\\ &\overset{(24)}{\le} c(s,a)+\sum_{t\in S}T(s,a,t)\sum_{\alpha\in\mathcal{A}}\mathsf{w}_{\alpha(t)}\end{aligned}$$

In both cases, the objective value of the transformed solution is the same as the one for the original solution. □

References

Klößner, T.; Torralba, Á.; Steinmetz, M.; and Hoffmann, J. 2021. Pattern Databases for Goal-Probability Maximization in Probabilistic Planning. In Goldman, R. P.; Biundo, S.; and Katz, M., eds., *Proceedings of the Thirty-First International Conference on Automated Planning and Scheduling (ICAPS 2021)*, 80–89. AAAI Press.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc.