# Bridging the Gap between Abstractions and Critical-Path Heuristics via Hypergraphs (Technical Report)

**Marcel Steinmetz**
CISPA Helmholtz Center for Information Security
Saarland Informatics Campus
Saarbrücken, Germany
marcel.steinmetz@cispa.saarland

**Álvaro Torralba**
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
torralba@cs.uni-saarland.de

### Abstract

Abstractions and critical-path heuristics are among the most important families of admissible heuristics in classical planning. In this paper, we present a new family of heuristics, which we name *hyperabstractions*, given by the combination of the principal ideas underlying abstractions and critical-path heuristics. Hyperabstractions approximate goal distances through a mapping from states to *sets* of abstract states. The abstract transition behavior forms a relation between abstract states and sets of abstract states, and is formally represented via the notion of hypergraphs. We show that both abstractions and critical-path heuristics can naturally be expressed as members of this family. Moreover, we devise a method to construct hyperabstractions, using either a set of abstractions or a critical-path heuristic as a seed, in a way that guarantees that the resulting distance estimations dominate those of the input heuristics, sometimes even strictly. By finding suitable cost partitionings for hyperabstraction heuristics, this dominance result is preserved even in comparison to the additive combination of the input heuristics. Our experiments indicate the potential of this new class of heuristics, opening a wide range of future research topics.

## Introduction

Heuristic search is a very popular method in classical planning. There is a huge body of research on how to automatically obtain good heuristic functions. Relating different families of heuristics is important to deepen the understanding of existing and to derive new heuristics. Helmert and Domshlak (2009) studied the relation between abstraction (Edelkamp 2001; 2006; Helmert et al. 2014), critical-path (Haslum and Geffner 2000), delete-relaxation and landmark heuristics (Bonet and Geffner 2001; Hoffmann, Porteous, and Sebastia 2004), showing that some of them are closely connected. However, the exact connection between general abstraction and critical-path heuristics is yet unclear.

We devise a new type of heuristics, *hyperabstractions*, that naturally generalizes abstraction and critical-path heuristics. Hyperabstractions are functions that map states to sets of *abstract concepts*. Heuristic values are computed as distances between sets of concepts in hypergraphs, concrete *interpretations* of hyperabstractions, with particular properties. Each hyperabstraction comes with multiple interpretations, differing in size and quality of the distance estimates. We show that every critical-path heuristic can be compiled into a dominating hyperabstraction heuristic in polynomial time. Moreover, for every set of abstraction heuristics, there exist polynomially constructable hyperabstraction heuristics dominating their combination via the maximum. Both results directly carry over to additive ensembles via the application of *cost partitionings*. While there exist polynomial algorithms computing the optimal cost partitioning for sets of abstractions (Katz and Domshlak 2008a), the complexity of optimal cost partitioning for critical-path heuristics is not known. In this paper, we show that finding the optimal cost-partitioning for critical-path heuristics is **NP**-hard. This result generalizes directly to hyperabstraction heuristics.

The experiments confirm the potential improvements in informativeness compared to abstraction and critical-path heuristics. Nevertheless, better construction methods for hyperabstraction heuristics remain an important question.

## Preliminaries

We introduce the planning formalisms, the heuristic functions considered in this paper, as well as hypergraphs.

### FDR Planning

A *FDR* planning task (Bäckström 1995) is a tuple $\Pi = \langle \mathcal{V}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{G} \rangle$ of *variables* $\mathcal{V}$, each $v \in \mathcal{V}$ with finite domain $\mathcal{D}_v$; *actions* $\mathcal{A}$; a complete variable assignment $s_{\mathcal{I}}$, called the *initial state*; and a partial variable assignment $\mathcal{G}$, the *goal*. Each *action* $a \in \mathcal{A}$ defines a *precondition* $pre_a$ and an *effect* $eff_a$, both partial variable assignments to $\mathcal{V}$, and a non-negative cost $c_a \in \mathbb{R}_0^+$. A *fact* is a variable value pair $\langle v, d \rangle$ where $v \in \mathcal{V}$ and $d \in \mathcal{D}_v$. We use conjunctions of facts and partial variable assignments interchangeably. Both are treated as sets of facts, using usual set operations for manipulation and comparison. Two conjunctions of facts $C, C'$ are *compatible*, written $C \parallel C'$, if $C(v) = C'(v)$ for every $v \in \mathcal{V}$ where both are defined.

We denote by $\mathcal{S}$ the set of all states of $\Pi$. An action $a$ is applicable in state $s$ if $pre_a \parallel s$. $\mathcal{A}(s)$ gives the set of all actions applicable in $s$. For $a \in \mathcal{A}(s)$, the result of applying $a$ in $s$, written $s[\![a]\!]$, is given by the variable assignments in $s$, overwritten by those in $eff_a$. The *state space* of $\Pi$ is given by the labeled transition system $\Theta^{\Pi} = \langle \mathcal{S}, \mathcal{T}, \mathcal{L}, s_{\mathcal{I}}, \mathcal{S}_{\mathcal{G}} \rangle$ where the set of *transitions* contains $\langle s, a, s[\![a]\!] \rangle \in \mathcal{T}$ for every $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$; *transition labels* $\mathcal{L} = \mathcal{A}$; and $\mathcal{S}_{\mathcal{G}}$ gives all *goal states*, i. e., $s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}$ if $\mathcal{G} \parallel s_{\mathcal{G}}$. A *plan* for $s$ is a sequence of actions $\pi = \langle a_1, \dots, a_n \rangle$ that labels a path from $s$ to some $s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}$. $\pi$ is *optimal* if its summed-up action cost is minimal among all plans for $s$.

## Heuristics

A *heuristic (function)* is a function $h \colon \mathcal{S} \mapsto \mathbb{R}^+_0 \cup \{\infty\}$ that maps states to an approximation of the cost-to-go to reach a goal state. The *perfect heuristic* $h^*$ assigns each state $s$ to the cost of an optimal plan for $s$, $h^*(s) = \infty$ if no plan for $s$ exists. A heuristic $h$ is *admissible* if $h(s) \leq h^*(s)$ for every $s \in \mathcal{S}$. $h$ is *consistent* if $h(s) - h(t) \leq c(a)$ for every $\langle s, a, t \rangle \in \mathcal{T}$. $h$ is *goal-aware* if $h(s_{\mathcal{G}}) = 0$ for all $s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}$. Every goal-aware and consistent heuristic is admissible. A heuristic $h$ *dominates* another $h'$, written $h \geq h'$, if $h(s) \geq h'(s)$ for all $s \in \mathcal{S}$. $h$ *strictly dominates* $h'$, written $h > h'$, if $h \geq h'$, and $h(s) > h'(s)$ for at least one state $s \in \mathcal{S}$.

In this paper, we will consider two families of admissible heuristics: *abstraction*, and *critical-path heuristics*.

**Abstractions** Abstraction heuristics compute the goal-distance in an *abstract state space*, dropping the distinction between some of the original states so to make the computation of $h^*$ feasible. Formally, an *abstraction* is a surjective function $\alpha \colon \mathcal{S} \mapsto \mathcal{S}^{\alpha}$ of states to *abstract states*. $\alpha$ implicitly introduces an equivalence relation between states, where $s \sim_{\alpha} t$ iff $\alpha(s) = \alpha(t)$. For an abstract state $s^{\alpha} \in \mathcal{S}^{\alpha}$, we denote by $[s^{\alpha}]$ all states $s \in \mathcal{S}$ such that $\alpha(s) = s^{\alpha}$.

The *abstract state space* of $\Theta^{\Pi}$ induced by $\alpha$ is given by the labeled transition system $\Theta^{\alpha} = \langle \mathcal{S}^{\alpha}, \mathcal{T}^{\alpha}, \mathcal{A}, s^{\alpha}_{\mathcal{I}}, \mathcal{S}^{\alpha}_{\mathcal{G}} \rangle$ where $s^{\alpha}_{\mathcal{I}} = \alpha(s_{\mathcal{I}})$, $\mathcal{S}_{\mathcal{G}} = \{\alpha(s_{\mathcal{G}}) \mid s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}\}$, and $\mathcal{T}^{\alpha} = \{\langle \alpha(s), a, \alpha(t) \rangle \mid \langle s, a, t \rangle \in \mathcal{T}\}$. We denote by $h^*_{\Theta^{\alpha}}$ the function assigning every abstract state $s^{\alpha}$ in $\Theta^{\alpha}$ to the minimal cost to reach any state in $\mathcal{S}^{\alpha}_{\mathcal{G}}$ from $s^{\alpha}$. The *abstraction heuristic* associated with $\alpha$ is given by $h^{\alpha}(s) := h^*_{\Theta^{\alpha}}(\alpha(s))$.

Various techniques have been proposed to automatically construct the abstraction function $\alpha$. Different approaches aim to be as accurate as possible while still practical to compute, e. g., symbolic representations of $\Theta^{\alpha}$ (Edelkamp 2002; Torralba et al. 2017), or designing $\alpha$ so to compute $h^*_{\Theta^{\alpha}}$ without actually constructing $\Theta^{\alpha}$ (Katz and Domshlak 2008b). In this paper, however, we consider only abstractions where the abstraction mapping $\alpha$ and its abstract state space can be computed and stored in an *explicit* form. Examples are pattern databases (PDBs) (Edelkamp 2001; Haslum et al. 2007), merge-and-shrink (MS) (Helmert et al. 2014), and Cartesian abstractions (Seipp and Helmert 2018).

**Critical-Path Heuristics** Critical-path heuristics estimate goal distance by breaking reasoning down to *atomic conjunctions*. Which conjunctions are treated as atomic, and thus how accurate the estimations are going to be, is controlled via the parameter $\mathcal{C}$. Let $\mathcal{C}$ be any set of conjunctions of facts. The *regression* of a conjunction $C$ by an action $a$ is defined if $C \parallel \mathit{eff}_a$ and $(C \setminus \mathit{eff}_a) \parallel \mathit{pre}_a$. If defined, the regression is given by $\mathit{Regr}(C, a) = (C \setminus \mathit{eff}_a) \cup \mathit{pre}_a$. Otherwise, we write $\mathit{Regr}(C, a) = \bot$. The *critical-path heuristic* over $\mathcal{C}$ is defined as $h^{\mathcal{C}}(s) = h^{\mathcal{C}}(s, \mathcal{G})$ where $h^{\mathcal{C}}(s, C) =$

$$\begin{cases} 0 & C \subseteq s \\ \min\limits_{a \in \mathcal{A}, \mathit{Regr}(C,a) \neq \bot} (c_a + h^{\mathcal{C}}(s, \mathit{Regr}(C, a))) & C \in \mathcal{C} \\ \max\limits_{C' \in \mathcal{C}, C' \subseteq C} h^{\mathcal{C}}(s, C') & \text{otherwise} \end{cases} \tag{1}$$

The most common method to choose $\mathcal{C}$ is enumerating all conjunctions of size up to $m$, where $m \in \mathbb{N}^+$ is a parameter. The resulting heuristic is denoted $h^m$ (Haslum and Geffner 2000; Haslum 2009). Recent works used the flexibility of critical-path heuristics to refine the heuristic online, during search, by incrementally adding conjunctions to $\mathcal{C}$ (Steinmetz and Hoffmann 2017; Fickert and Hoffmann 2017).

**Cost Partitioning** Multiple admissible heuristics can trivially be combined in a way that preserves admissibility by taking the maximum. Summing up the individual estimations dominates the maximum, but is admissible only under particular conditions. A popular method satisfying such condition by construction is *cost partitioning* (Katz and Domshlak 2008a). Cost partitioning does not only allow the admissible combination of multiple heuristics, but can be used also to improve a single heuristic. A prominent example for the latter is the LM-cut heuristic (Helmert and Domshlak 2009).

Formally, let $c' \colon \mathcal{A} \mapsto \mathbb{R}^+_0$ be any cost function. We denote by $h[\![c']\!]$ the heuristic function $h$ computed in the copy of $\Pi$ whose cost function is replaced by $c'$. A *cost partitioning* is a tuple of cost functions $\mathbf{c} = \langle c_1, \dots, c_n \rangle$ such that for every action $a \in \mathcal{A}$, it holds that $\sum_{i=1}^{n} c_i(a) \leq c(a)$. Using cost partitionings, an ensemble of heuristics $h_1, \dots, h_n$ can be additively combined through $h_{1,\dots,n}[\![\mathbf{c}]\!] := \sum_{i=1}^{n} h_i[\![c_i]\!]$. $h_{1,\dots,n}[\![\mathbf{c}]\!]$ is admissible if all heuristics $h_1, \dots, h_n$ are admissible (Katz and Domshlak 2008a). A cost partitioning is applied to a single heuristic $h$ by setting $h_i = h$ for all $i$. We denote the additive combination by $h[\![\mathbf{c}]\!]$ in that case.

Clearly, the estimates of $h_{1,\dots,n}[\![\mathbf{c}]\!]$ depend crucially on the distribution of the action costs over the individual heuristics. We say that a cost partitioning $\mathbf{c} = \langle c_1, \dots, c_n \rangle$ is *optimal* for $h_1, \dots, h_n$ in a state $s$, if $h_{1,\dots,n}[\![\mathbf{c}]\!](s) \geq h_{1,\dots,n}[\![\mathbf{c}']\!](s)$ for all other cost partitionings $\mathbf{c}'$ of size $n$. For a single heuristic $h$, the size of $\mathbf{c}$ is not fixed a priori. Hence, in the definition of optimality for single heuristics, the size restriction is dropped. For an ensemble of abstraction heuristics, an optimal cost partitioning can be found in polynomial time (Katz and Domshlak 2008a). For critical-path heuristics it was an open question whether it is possible to compute an optimal cost partitioning in polynomial time.

## Hypergraphs

A *labeled weighted directed hypergraph* (Gallo, Longo, and Pallottino 1993) is given by a triple $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w \rangle$ consisting of a finite set of *nodes* $\mathcal{N}$, a finite set of *labels* $\mathcal{L}$, each $l \in \mathcal{L}$ associated with a *weight* $w(l) \in \mathbb{R}^+_0$, and a finite set of *hyperedges* $\mathcal{E} \subseteq 2^{\mathcal{N}} \times \mathcal{L} \times 2^{\mathcal{N}}$. A hyperedge $e$ is a tuple $\langle T_e, l_e, H_e \rangle$ where $T_e$ is the *tail* of $e$, $H_e$ is the *head* $e$, and $l_e$ is the *label*. $e$ is called a *backward (*B*) hyperedge* if $|H_e| \leq 1$. $e$ is a *strict* B*-hyperedge* if $|H_e| = 1$. $e$ is a *(strict) forward (*F*) hyperedge* if $|T_e| \leq 1$ ($|T_e| = 1$). $\mathcal{H}$ is a *(strict)* B*-hypergraph* if it contains only (strict) B-hyperedges, and similarly, $\mathcal{H}$ is an *(strict)* F*-hypergraph* if all its edges are (strict) F-hyperedges. Hypergraphs satisfying the B- and F-conditions at the same time are called BF*-hypergraphs*. The *symmetric image* of $\mathcal{H}$ is given by the hypergraph $\widehat{\mathcal{H}} = \langle \widehat{\mathcal{N}}, \widehat{\mathcal{E}}, \widehat{\mathcal{L}} \rangle$ where $\widehat{\mathcal{N}} = \mathcal{N}$, $\widehat{\mathcal{L}} = \mathcal{L}$, and, for every $e \in \mathcal{E}$, $\widehat{\mathcal{E}}$ contains $\widehat{e} = \langle H_e, l_e, T_e \rangle$. Note that $\widehat{\mathcal{H}}$ is an F-hypergraph iff $\mathcal{H}$ is a B-hypergraph, and a B-hypergraph iff $\mathcal{H}$ is an F-hypergraph.

For the rest of the paper, we will exclusively consider B- and F-hypergraphs. Let $\mathcal{H}$ be a B-hypergraph, and $N, N' \subseteq \mathcal{N}$ be two subsets of nodes. The *minimal distance*, *distance* for short, from $N$ to $N'$ in $\mathcal{H}$ is given by the point-wise maximal function satisfying[1] $d^{\mathrm{B}}_{\mathcal{H}}(N, N') =$

$$\begin{cases} 0 & \text{if } N' \subseteq N \\ \min_{e \in \mathcal{E}, n' \in H_e} (w(l_e) + d^{\mathrm{B}}_{\mathcal{H}}(N, T_e)) & \text{if } N' = \{n'\} \\ \max_{n' \in N'} d^{\mathrm{B}}_{\mathcal{H}}(N, \{n'\}) & \text{otherwise} \end{cases} \quad (2)$$

Note the similarities between Equations (1) and (2). Both are indeed closely related. We will come back to this comparison when we later show how critical-path heuristics can be phrased in terms of this hypergraph notation.

In F-hypergraphs the distance $d^{\mathrm{F}}_{\mathcal{H}}$ is defined symmetrically. Using the symmetric image operator, it holds that $d^{\mathrm{F}}_{\mathcal{H}}(N, N') = d^{\mathrm{B}}_{\widehat{\mathcal{H}}}(N', N)$. We will omit the F and B superscript if the type is clear, and omit the sub- and superscripts all together if also $\mathcal{H}$ is clear from the context. Gallo et al. (1993) have shown a polynomial (in $|\mathcal{H}|$) algorithm to compute minimal distances in hypergraphs.

## Hyperabstractions

In this section we introduce *hyperabstractions* and show how admissible heuristics can be derived from this notion. Consider any planning task $\Pi = \langle \mathcal{V}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{G} \rangle$, and let $\Theta^{\Pi} = \langle \mathcal{S}, \mathcal{T}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{S}_{\mathcal{G}} \rangle$ be its associated state space.

**Definition 1 (Hyperabstraction)** *A* hyperabstraction *is a function $\rho \colon \mathcal{S} \mapsto 2^{\mathcal{P}}$ mapping states to* abstract concepts. *$\mathcal{P}$ is the set of abstract concepts associated with $\rho$.*

Abstract concepts describe features of particular states. Propositional formulae over facts are natural variants of such features, but more complex structures such as for example those underlying abstract states qualify as well. Similarly to abstractions, for any $p \in \mathcal{P}$, we denote by $[p]$ the set of all states $s \in \mathcal{S}$ such that $p \in \rho(s)$. Note, however, that in contrast to abstract states, the same state may be represented by multiple abstract concepts. We chose the term *abstract concept* instead of *abstract state* to make this difference explicit.

Each state being possibly associated with multiple concepts, abstract transitions should no longer connect different concepts individually, but form a relation between *sets* of them. In principle, hypergraphs provide the possibility to define complex transition behavior between arbitrary sets of concepts. However, the distance metric for such general hypergraphs gets difficult to define, or expensive to compute (see e. g. Ausiello and Laura 2017), and are thus not particularly suited for our purpose: heuristic computation. In the following, we thus consider specifically the F- and B-hypergraph special cases. We propose two different *interpretations* of hyperabstractions accordingly. Let $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w \rangle$ by a hypergraph with nodes $\mathcal{N} = \mathcal{P}$, labels $\mathcal{L} = \mathcal{A}$, and weights $w = c$. We say that:

**Definition 2 (F-interpretation)** *$\mathcal{H}$ is an* F-interpretation *of $\rho$ if $\mathcal{H}$ is an* F*-hypergraph, and for every transition $\langle s, a, t \rangle \in \mathcal{T}$ and for every $p_s \in \rho(s)$, $\mathcal{H}$ has a hyperedge $\langle \{p_s\}, a, P_t \rangle \in \mathcal{E}$ such that $P_t \subseteq \rho(t)$.*

**Definition 3 (B-interpretation)** *$\mathcal{H}$ is a* B-interpretation *of $\rho$ if $\mathcal{H}$ is a* B*-hypergraph, and for every transition $\langle s, a, t \rangle \in \mathcal{T}$ and for every $p_t \in \rho(t)$, $\mathcal{H}$ has a hyperedge $\langle P_s, a, \{p_t\} \rangle \in \mathcal{E}$ such that $P_s \subseteq \rho(s)$.*

To ensure that distances obtained from hyperabstractions are admissible, all transitions in the original state space must be preserved by the abstract transition relation of the interpretations. Both definitions, however, leave open the exact choice of the sets $P_t$, respectively $P_s$, to do so for any particular transition. Defining *the*, unique, (F or B) interpretation associated with a hyperabstraction raises certain complications, as we will see below. The restriction of hyperedges to tails of size $\leq$1 (F-interpretation) and heads of size $\leq$1 (B-interpretation) leads to a very fundamental difference in how $\Theta^{\Pi}$'s transitions are reflected in the hypergraphs. Given an abstract concept $p \in \mathcal{P}$ and action $a \in \mathcal{A}$, the hyperedges of F-interpretations enumerate consequences of the application of $a$ in the context of $p$, i. e., possible effects of applying $a$ in the states $[p]$. In contrast, the hyperedges of B-interpretations enumerate different conditions under which the application of $a$ makes true $p$, i. e., results in a state in $[p]$. Phrased in common search terms, F-interpretations provide information in terms of *progression*, whereas B-interpretations provide information on the *regression*.

Heuristic functions associated with a hyperabstraction $\rho$ are obtained directly from the distance metrics of the F- and B-interpretations of $\rho$. To make apparent the symmetric nature of F- and B-hyperabstraction heuristics, we define both heuristics as functions $h_\rho : (2^{\mathcal{S}} \times 2^{\mathcal{S}}) \mapsto \mathbb{R}^+_0 \cup \{\infty\}$, $h_\rho(S, T)$ yielding an approximation of the minimal cost of all paths in $\Theta^{\Pi}$ between any state $s \in S$ and any $t \in T$. Abusing the notation, we use the same heuristic symbol to denote $h_\rho(s) := h_\rho(\{s\}, \mathcal{S}_{\mathcal{G}})$. Let $\mathcal{H}^{\mathrm{F}}$ be any F-interpretation of $\rho$, and $\mathcal{H}^{\mathrm{B}}$ be any B-interpretation $\mathcal{H}^{\mathrm{B}}$ of $\rho$. We define:

**Definition 4 (F-hyperabstraction heuristic)** *The* F-hyperabstraction heuristic *associated with $\rho$ and $\mathcal{H}^{\mathrm{F}}$ is given by $h^{\mathrm{F}}_\rho[\mathcal{H}^{\mathrm{F}}](S, T) = d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\bigcap_{s \in S} \rho(s), \bigcup_{t \in T} \rho(t))$.*

**Definition 5 (B-hyperabstraction heuristic)** *The* B-hyperabstraction heuristic *associated with $\rho$ and $\mathcal{H}^{\mathrm{B}}$ is given by $h^{\mathrm{B}}_\rho[\mathcal{H}^{\mathrm{B}}](S, T) = d^{\mathrm{B}}_{\mathcal{H}^{\mathrm{B}}}(\bigcup_{s \in S} \rho(s), \bigcap_{t \in T} \rho(t))$.*

[1]We assume that $\min(\emptyset) = \infty$ and $\max(\emptyset) = 0$.

We omit the hypergraph parameter if it is clear from the context, or unimportant for the discussion. The difference in F- versus B-hypergraph distance function requires to swap the set operators in the definitions. In the forward case, for any two sets $P_S, P_T \subseteq \mathcal{P}$, $d^{\mathrm{F}}(P_S, P_T)$ measures the maximal distance $d^{\mathrm{F}}(\{p_s\}, P_T)$ of any $p_s \in P_S$. Because of this maximization, to guarantee that $h^{\mathrm{F}}_{\rho}(S, T)$ admissibly estimates of the distance from any $s \in S$ to the states in $T$, $P_S$ may only contain abstract concepts representing all states in $S$. On the other hand, $P_T$ must contain every abstract concept representing some state in $T$ so that $d^{\mathrm{F}}(\{p_s\}, P_T)$ admissibly approximates the cost to reach from $[p_s]$ any state in $T$. Symmetrically, $d^{\mathrm{B}}(P_S, P_T)$ is the maximum over $d^{\mathrm{B}}(P_S, \{p_t\})$ for all $p_t \in P_T$. To ensure the admissibility of $h^{\mathrm{B}}_{\rho}$, $P_T$ may hence only contain abstract concepts that represent all states in $T$, while $P_S$ must contain every abstract concept representing any state in $S$.

Moreover, observe that, for any fixed set $P_T \subseteq \mathcal{P}$, $d^{\mathrm{F}}(\{p\}, P_T)$ can be precomputed individually for every abstract concept $p$. Since the heuristic computations $h^{\mathrm{F}}_{\rho}(s)$ will generate $d^{\mathrm{F}}(P_S, P_T)$ calls, changing only the $P_S$ part, $h^{\mathrm{F}}_{\rho}(s)$ can be computed just based on lookups of the precomputed distances. This optimization is not possible for $h^{\mathrm{B}}_{\rho}$, since the distance estimate $d^{\mathrm{B}}(P_S, P_T)$ requires the consideration of $P_S$ as a whole, which however varies from state to state.

Before going into formal claims, consider the following example to get an intuitive understanding of the two hyperabstraction variants.

**Example 1 ("Robot in a china shop")** *Consider the following task. There are three variables: whether the robot* $R$ *has entered the shop (*$T$*) or not (*$F$*), and the state of two vases* $V_1$ *and* $V_2$ *(clean* $C$*, held by the robot* $R$*, broken* $B$*). The (unit-cost) actions are:* enter *changing* $R$ *from* $F$ *to* $T$*,* $\mathrm{pickup}(V_i)$ *with precondition* $\{V_i = C, R = T\}$ *and effect* $\{V_i = R\}$*,* $\mathrm{drop}(V_i)$ *requiring that* $\{V_i = R\}$ *and setting* $\{V_i = B, R = F\}$ *in its effect, and* $\mathrm{smash}(V_i, V_j)$ *with precondition* $\{R = T, V_i = R, V_j = C\}$ *and effect* $\{V_j = B\}$*. The initial state is* $s_{\mathcal{I}} = \{R = F, V_1 = C, V_2 = C\}$*. The goal is* $\mathcal{G} = \{R = T, V_1 = B, V_2 = B\}$*. An optimal plan for* $s_{\mathcal{I}}$ *is given by the action sequence* $\langle \mathrm{enter}, \mathrm{pickup}(V_1), \mathrm{smash}(V_1, V_2), \mathrm{drop}(V_1), \mathrm{enter}\rangle$.

*Let* $\mathcal{P}$ *be the set of all facts of* $\Pi$*. Consider the hyperabstraction* $\rho : \mathcal{S} \mapsto 2^{\mathcal{P}}$ *that maps every state to the facts true in it. Figure 1 depicts a* F*- and a* B*-interpretation of* $\rho$*, omitting edge labels and self-loops for the sake of readability. Ignore the red part for this example.*

*Both interpretations are constructed by connecting for every action the precondition and effect facts accordingly. Consider* $\mathrm{drop}(V_1)$ *along with its hyperedges (those marked in blue).* $\mathrm{drop}(V_1)$ *affects* $V_1$ *and* $R$*. In* $\mathcal{H}^{\mathrm{F}}$*, applying this action in any state with* $R{=}T$ *will change the value of* $R$ *to* $F$*. The hyperedge* $\langle \{R{=}T\}, \mathrm{drop}(V_1), \{R{=}F\}\rangle$ *represents every such transition.* $V_1{=}B$ *could be added to the head as well, which would strengthen the hyperedge, further constraining the set of successor states, and thus possibly increasing the distance estimates. Adding* $V_1{=}B$ *is however not required to satisfy Definition 2. Similarly, the hyperedge*

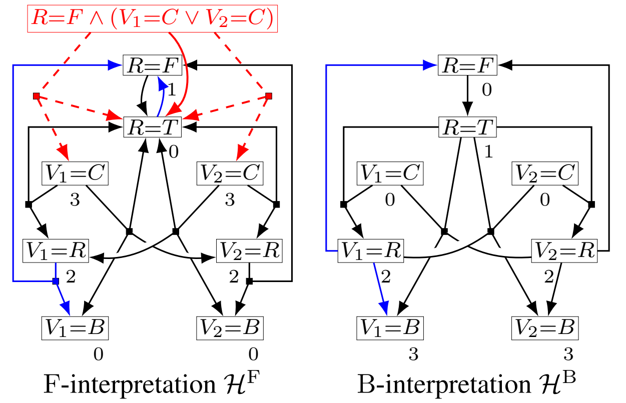


Figure 1: F- and B-interpretations for the task in Example 1.

$\langle \{V_1{=}R\}, \mathrm{drop}(V_1), \{R{=}F, V_1{=}B\}\rangle$ *covers the application of* $\mathrm{drop}(V_1)$ *in every state where* $V_1{=}R$*. The remaining abstract concepts either don't represent any state where* $\mathrm{drop}(V_1)$ *is applicable, e. g.,* $V_1{=}C$*, or remain invariant, e. g., the* $V_2$ *facts. In* $\mathcal{H}^{\mathrm{B}}$*, every application of* $\mathrm{drop}(V_1)$ *requires* $V_1{=}R$*. The action always makes true the facts* $R{=}F$ *and* $V_1{=}B$*, encoded by the two depicted hyperedges.*

*The resulting heuristic estimates are* $h^{\mathrm{F}}_{\rho}[\mathcal{H}^{\mathrm{F}}](s_{\mathcal{I}}) = h^{\mathrm{B}}_{\rho}[\mathcal{H}^{\mathrm{B}}](s_{\mathcal{I}}) = 3$*. There is only a single goal state, so both measure the distance from the set of facts* $P_{\mathcal{I}} := s_{\mathcal{I}}$ *to* $P_{\mathcal{G}} := \mathcal{G}$*. Figure 1 annotates every* $p \in \mathcal{P}$ *with the distances* $d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{p\}, P_{\mathcal{G}})$ *and* $d^{\mathrm{B}}_{\mathcal{H}^{\mathrm{B}}}(P_{\mathcal{I}}, \{p\})$*. Consider the fact* $p = \langle V_1, R\rangle$*. The heuristic values of both heuristics rely on* $d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{p\}, P_{\mathcal{G}}) = d^{\mathrm{B}}_{\mathcal{H}^{\mathrm{B}}}(P_{\mathcal{I}}, \{p\}) = 2$*.* $p$*'s only outgoing hyperedge in* $\mathcal{H}^{\mathrm{F}}$ *leads to* $R{=}F$ *and* $V_1{=}B$*. The estimated distance for the former fact is* $1$*, the latter is contained in* $P_{\mathcal{G}}$*. The maximum over both distances is* $1$*, resulting in* $d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(p, P_{\mathcal{G}}) = 2$*. In* $\mathcal{H}^{\mathrm{B}}$*, the only hyperedge leading to* $p$ *has* $R{=}T$ *and* $V_1{=}C$ *in its tail. The distance from* $P_{\mathcal{I}}$ *to the former is* $1$*, the latter is* $0$*. The maximum is again* $1$*, yielding* $d^{\mathrm{B}}_{\mathcal{H}^{\mathrm{B}}}(P_{\mathcal{I}}, \{p\}) = 2$*. Notably, the* F *and* B *distance functions account for different applications of* enter*:* $d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}$ *observes that* enter *must be executed after* drop*;* $d^{\mathrm{B}}_{\mathcal{H}^{\mathrm{B}}}$ *sees that* enter *must be executed prior to* pickup.

As pointed out earlier, there might be different hypergraphs satisfying Definitions 2 or 3, differing in the design of the transition relation. Since the hyperabstraction heuristics compute their estimations based on the chosen hyperedges, the quality of the heuristics is strongly affected by those choices. Consider the red part of Figure 1, which introduces a new abstract concept $p$ representing all states where the robot has not entered the shop and at least one of the vases is clear. Figure 1 shows two possible ways to represent the enter transitions leaving the states $[p]$: via the single solid hyperedge, or via the two dashed ones. Choosing the former, the information is lost that one of the vases was clear beforehand. In this case, $h^{\mathrm{F}}_{\rho}(s_{\mathcal{I}})$ stays 3. The latter also satisfies the F-interpretation definition: the application of enter in any state $s \in [p]$ changes the value of $R$ to $T$. Moreover, one of $s(V_1){=}C$ and $s(V_2){=}C$ must have been true beforehand, both variables are not affected by $\mathrm{enter}$. For the latter choice, we however obtain $h^{\mathrm{F}}_{\rho}(s_{\mathcal{I}}) = 4$.

Of course, one would like to always choose the hypergraph yielding the most accurate estimations relative to the abstract concepts at hand. Unfortunately this not is not feasible in general. Let $h^{\mathrm{F}*}_{\rho}$ and $h^{\mathrm{B}*}_{\rho}$ denote the "best" hyperabstraction heuristics one can obtain for $\rho$. This is, dropping the distinction between F and B for the remainder of this section, $h^{\mathrm{X}*}_{\rho}$ is defined such that, for every X-interpretation $\mathcal{H}$ of $\rho$, it holds that $h^{\mathrm{X}*}_{\rho} \geq h^{\mathrm{X}}_{\rho}[\mathcal{H}]$, and there is an X-interpretation $\mathcal{H}^*$ of $\rho$ such that $h^{\mathrm{X}*}_{\rho} = h^{\mathrm{X}}_{\rho}[\mathcal{H}^*]$. Then

**Theorem 1** *(i) $h^{\mathrm{X}*}_{\rho}$ is well-defined, i. e., for every $\rho$ there exists an* X*-interpretation $\mathcal{H}^*$ of $\rho$ as defined in the text, and (ii) given an* X*-interpretation $\mathcal{H}$ of $\rho$ and state $s$, deciding whether $h^{\mathrm{X}}_{\rho}[\mathcal{H}](s) = h^{\mathrm{X}*}_{\rho}(s)$ is **NP**-hard.*

*Proof (sketch).* Consider the forward case. For (i), a desired F-interpretation $\mathcal{H}^*$ can be constructed by creating for every transition $\langle s, a, t\rangle$ and for every $p_s \in \rho(s)$, a hyperedge $\langle \{p_s\}, a, \rho(t)\rangle$. Regarding (ii), similar to the example given above, hyperedges can be used to resolve disjunctive conditions in atomic concepts. One can design $\Pi_n$ and $\rho_n$ such that such disjunctions are turned into transition nondeterminism, and thus exponentially many hyperedges must be considered to obtain $h^{\mathrm{X}*}_{\rho}$.

Despite this rather disappointing result, hyperabstraction heuristics remain admissible regardless of the interpretation. Moreover, we will see in the next sections that polynomially constructable interpretations suffice to show dominance over various existing heuristics from the literature.

**Theorem 2** *For every task $\Pi$, hyperabstraction $\rho$, and* X*-interpretation $\mathcal{H}$ of $\rho$, $h^{\mathrm{X}}_{\rho}[\mathcal{H}]$ is consistent and goal-aware.*

*Proof (sketch).* Follows from the requirements on the structure of the hyperedges $\mathcal{E}$ imposed by Definitions 2 and 3.

## Relation to Existing Techniques

Both abstraction and critical-path heuristic can naturally be written as hyperabstractions. The former boils down to a special case of hyperabstractions where every state is mapped to exactly one abstract concept. The latter can be seen as a hyperabstraction that maps every state to the set of atomic conjunctions true in that state.

This section goes beyond this simple observation, showing that critical-path heuristics can indeed be seen as one particular interpretation of a hyperabstraction. Moreover, we will establish the connection between abstractions and hyperabstractions, regarding hyperabstractions as a new way to admissibly combine multiple abstractions.

In both comparisons, we will provide dominance results for the non-additive case. Those results directly generalize to the additive combination of abstraction and critical-path heuristics through the use of appropriate cost partitionings. This promotes the quest for computing optimal cost partitionings for hyperabstractions. We close the section with a negative result, showing that finding optimal cost partitionings for hyperabstractions is **NP**-hard in general.

### Critical-Path Heuristics

That critical-path heuristics are related to hypergraphs has been observed before, e. g., (Haslum 2006). However, this relation has so far not been spelled out formally and explicitly. We do this here using our notions of hyperabstractions. More specifically, we show how to construct, from $h^{\mathcal{C}}$, a hyperabstraction $\rho^{\mathcal{C}}$ and B-interpretation $\mathcal{H}^{\mathcal{C}}$ of $\rho^{\mathcal{C}}$ such that the equations underlying $h^{\mathcal{C}}$ and $h^{\mathrm{B}}_{\rho}[\mathcal{H}^{\mathcal{C}}]$ turn into the same.

Let $\mathcal{C}$ be any set of atomic conjunctions. Consider the hyperabstraction $\rho^{\mathcal{C}}$ with abstract concepts $\mathcal{P} = \mathcal{C}$, mapping every state to the atomic conjunctions satisfied in it. The hypergraph $\mathcal{H}^{\mathcal{C}}$ underlying $h^{\mathcal{C}}$'s distance computation is constructed as follows. $\mathcal{H}^{\mathcal{C}}$ contains one node for every atomic conjunction. For every conjunction $C \in \mathcal{C}$ and action $a \in \mathcal{A}$, $\mathcal{H}^{\mathcal{C}}$ contains a hyperedge $e^{C,a}$ iff *Regr*$(C, a)$ is defined, and $e^{C,a} = \langle T_{e^{C,a}}, a, \{C\}\rangle$ where $T_{e^{C,a}} = \{C' \in \mathcal{C} \mid C' \subseteq \mathit{Regr}(C, a)\}$. It is straightforward to verify that $d^{\mathrm{B}}_{\mathcal{H}^{\mathcal{C}}}$ is indeed equivalent to the recursive definition of $h^{\mathcal{C}}$. Observe that $\mathcal{H}^{\mathcal{C}}$ also satisfies Definition 3. Consider any transition $\langle s, a, t\rangle \in \mathcal{T}$, and any conjunction satisfied in $t$, $C_t \in \rho^{\mathcal{C}}(t)$. Clearly, *Regr*$(C_t, a)$ must be defined and *Regr*$(C_t, a) \subseteq s$. By the construction of $\mathcal{H}^{\mathcal{C}}$, there is an edge $e^{C_t,a} \in \mathcal{E}$ labeled with $a$ and whose tail contains exactly the atomic conjunctions satisfied in *Regr*$(C_t, a)$. Thus $T_{e^{C_t,a}} \subseteq \rho^{\mathcal{C}}(s)$.

**Theorem 3** *For every task $\Pi$, and set of atomic conjunctions $\mathcal{C}$, one can construct a* B*-interpretation $\mathcal{H}$ of $\rho^{\mathcal{C}}$ in polynomial time such that $h^{\mathrm{B}}_{\rho^{\mathcal{C}}}[\mathcal{H}] \geq h^{\mathcal{C}}$.*

This result immediately carries over to the comparison to additive critical-path heuristics (Haslum, Bonet, and Geffner 2005; Helmert and Domshlak 2009), i. e., is orthogonal to the application of cost-partitionings.

The definition of $\mathcal{H}^{\mathcal{C}}$ above is loosely related to the $\Pi^m$-compilation (Haslum 2009). The $\Pi^m$ compilation explicitly encodes the satisfaction of atomic conjunctions $C$ via new state variables $\pi_C$. To determine the satisfaction of $C$, $\Pi^m$ introduces copies of actions $a^f$ augmenting the precondition of $a$ by additional context information, sets of facts $f$, very similar to the construction of $\mathcal{H}^{\mathcal{C}}$'s hyperedges. The $\Pi^m$ compilation differs from the $\mathcal{H}^{\mathcal{C}}$ construction in that we consider arbitrary conjunctions. Moreover, $\Pi^m$ introduces an action copy $a^f$ for every action $a$ and set of facts $f$ of size of at most $m-1$, but sets true $\pi_C$ for all conjunctions $C$ guaranteed to be satisfied after the application of $a$ in the context of $f$. In contrast, $\mathcal{H}^{\mathcal{C}}$ considers individually the satisfaction of each $C$ by $a$ via separate hyperedges $e^{C,a}$.

B-hyperabstractions generalize critical-path heuristics in two aspects: (1) they natively support more complex abstract states; (2) critical-path heuristics represent only one particular instantiation of the associated hyperabstraction. How to exploit (1) will be the topic of an upcoming section. Regarding (2), Theorem 1 already indicates that $\mathcal{H}^{\mathcal{C}}$ might not be the *optimal* B-interpretation for $\rho^{\mathcal{C}}$. In fact, under particular circumstances, it is possible to consider different B-interpretations, leading to higher heuristic estimates, while still staying within the polynomial (in $|\mathcal{C}|$) size bound. For instance, assume there are actions that affect variables without imposing preconditions on them. Such variables will be left unspecified in every regression over this action. Hence, conjunctions that contain assignments to these variables will never be part of the tails of the hyperedges generated for this action. To avoid this shortcoming, one can generate multiple hyperedges for an atomic conjunction action pair, instead of just a single one. Then in each of those hyperedges, one can augment the regression with an additional context, allowing to include more atomic conjunctions in the tails of the hyperedges, and thus may lead to higher heuristic estimates.

**Theorem 4** *There exist* $\Pi$*,* $\mathcal{C}$*, and* $\mathrm{B}$*-interpretations* $\mathcal{H}$ *of* $\rho^{\mathcal{C}}$ *where* $|\mathcal{H}|$ *is polynomially bounded in* $|\mathcal{C}|$ *but* $h^{\mathrm{B}}_{\rho^{\mathcal{C}}}[\mathcal{H}] > h^{\mathcal{C}}$*.*

The idea above is similar to *context-splitting* (Röger, Pommerening, and Helmert 2014). The latter is a technique that, given a propositional formula $\phi$ over facts and an action $a$, constructs a new task where $a$ is split into two actions that are identical to $a$ but one having the precondition $pre_a \wedge \phi$ and the other one $pre_a \wedge \neg\phi$. This split can be helpful to derive (admissible) additive heuristics, e. g., in the construction of cost partitionings the costs of both splits can be distributed independently to different heuristics. However, in contrast to context-splitting, hyperabstraction interpretations allow to split hyperedges much more specifically, independently of other hyperedges of the same action. Moreover, and much more crucially, hyperedges can be split over non-disjoint contexts. In the example in Figure 1, the two dashed hyperedges are not mutually exclusive: both are simultaneously representing the application of $\mathrm{enter}$ to the initial state.

## Abstractions

Let $\alpha\colon \mathcal{S} \mapsto \mathcal{S}^{\alpha}$ be any abstraction, and $\Theta^{\alpha}$ be its induced abstract state space. Labeled transition systems can be seen as a special case of labeled hypergraphs whose hyperedges have heads and tails with size of exactly one, i. e., satisfy the strict $\mathrm{BF}$-hypergraph criterion. This has two immediate consequences: (1) $\Theta^{\alpha}$ satisfies Definitions 2 and 3, by construction. (2) For strict $\mathrm{BF}$-hypergraphs the forward distance $d^{\mathrm{F}}(\{n\}, N')$ boils down to the minimal graph distance from $n$ to any $n' \in N'$. Therefore, $h^{\mathrm{F}}_{\alpha}[\Theta^{\alpha}](s) = d^{\mathrm{F}}_{\Theta^{\alpha}}(\{\alpha(s)\}, \mathcal{S}^{\alpha}_{\mathcal{G}}) = h^{*}_{\Theta^{\alpha}}(\alpha(s)) = h^{\alpha}(s)$, i. e., abstraction heuristics can be interpreted directly as F-hyperabstraction heuristics. Moreover, in the construction of any F-interpretation $\mathcal{H}$ of $\alpha$, there are exactly two possible hyperedges to represent any transition $\langle s, a, t\rangle \in \mathcal{T}$: either $\langle\{\alpha(s)\}, a, \{\alpha(t)\}\rangle$ or $\langle\{\alpha(s)\}, a, \emptyset\rangle$. The former is selected by $\Theta^{\alpha}$, and clearly carries more information than the latter, i. e., $h^{\alpha} = h^{\mathrm{F}*}_{\alpha}$. However, although $\Theta^{\alpha}$ being a $\mathrm{B}$-interpretation of $\alpha$, $h^{\alpha}$ is in general not equivalent to $h^{\mathrm{B}}_{\alpha}[\Theta^{\alpha}]$. In strict $\mathrm{BF}$-hypergraphs, $d^{\mathrm{B}}(\{n\}, N')$ gives the *maximum* of the minimal graph distances from $n$ to any $n' \in N'$. To account for that, $h^{\mathrm{B}}$ considers in $N'$ only those abstract concepts representing *all* goal states. In case of the abstraction $\alpha$, this means that $h^{\mathrm{B}}_{\alpha}[\Theta^{\alpha}] = h^{\alpha}$ only if (*) $|\mathcal{S}^{\alpha}_{\mathcal{G}}| = 1$, but $h^{\mathrm{B}}_{\alpha}[\Theta^{\alpha}](s) = d^{\mathrm{B}}_{\mathcal{H}}\Theta^{\alpha}(\{\alpha(s)\}, \emptyset) = 0 \le h^{\alpha}$ if $\alpha$ maps any two goal states to different abstract states. Note that (*) necessarily holds for tasks in transition normal form (TNF) (Pommerening and Helmert 2015).

Single abstractions hence constitute trivial instantiations of our concepts of F-hyperabstraction heuristics, and under certain circumstances also $\mathrm{B}$-hyperabstraction heuristics. The converse is however not true. Helmert et al. (2014) have shown planning tasks where the critical-path heuristic $h^2$ is perfect, but, unless **P** $=$ **NP**, it is not possible to construct any M&S abstraction $\alpha$ with $h^{\alpha} = h^{*}$ in polynomial time. Given the observations from the previous section, this results directly carries over to $\mathrm{B}$-hyperabstraction heuristics in general. It turns out that the hypergraph characteristics are actually not required to show this relation:

**Theorem 5** *There exist families of tasks* $\Pi_n$ *and polynomially size-bounded forward and backward hyperabstraction heuristics such that* $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}] = h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}] = h^{*}$ *and both* $\mathcal{H}^{\mathrm{F}}$ *and* $\mathcal{H}^{\mathrm{B}}$ *are* $\mathrm{BF}$*-hypergraphs, but unless* ***P*** $=$ ***NP****, it is not possible to construct any M&S abstraction* $\alpha$ *such that* $h^{\alpha} = h^{*}$ *in polynomial time.*

*Proof (sketch).* The proof is based on a family of tasks representing CNF formulas. Given a formula $\phi$, the corresponding task is designed so that states represent possible Boolean variable assignments, $h^{*}(s) = 1$ iff $s$ does not satisfy the formula, and $h^{*}(s) = 0$ otherwise. Using this, the evaluation of $\phi$ can be encoded as F and B hyperabstraction heuristics, even restricted to $\mathrm{BF}$-hypergraphs. However, if it were possible to construct a M&S abstraction $\alpha$ such that $h^{\alpha} = h^{*}$, for arbitrary $\phi$, in time polynomial in the size of $\phi$, then we could solve SAT for any CNF formula in polynomial time through an inspection of $\alpha$.

Note that even though this claim considers only M&S abstractions, it also applies to less general cases like PDBs.

So far, we have seen that hyperabstractions are strictly more powerful than single abstractions, even restricted to $\mathrm{BF}$-hypergraphs, and thus without making use of the possibility to define actual hyperedges. We next show that single hyperabstractions also dominate the combination of collections of abstractions. Unfortunately, we cannot report any result for the other direction, i. e., it remains unclear whether a single hyperabstraction can be always compiled into an equivalent collection of abstractions, in polynomial time. Let $\alpha_1, \dots, \alpha_n$ be any set of abstractions. Consider the hyperabstraction whose abstract concepts $\mathcal{P}$ are given by the union of all $\mathcal{S}^{\alpha_i}$, combining all the abstraction functions into the single hyperabstraction function: $\rho(s) = \{\alpha_1(s), \dots, \alpha_n(s)\}$. Consider the $\mathrm{BF}$-hypergraph $\mathcal{H}$, given by the union of all induced abstract state spaces $\Theta^{\alpha_1}, \dots, \Theta^{\alpha_n}$. $\mathcal{H}$ satisfies Definitions 2 and 3 similarly to the single abstraction case. Moreover, since abstract states of different abstractions are not connected, the forward distance in $\mathcal{H}$ from any abstract state to the abstract goal states remains the same as in the abstract state's corresponding abstraction, i. e., it holds that $d^{\mathrm{F}}_{\mathcal{H}}(\{s^{\alpha_i}\}, \bigcup_i \mathcal{S}^{\alpha_i}_{\mathcal{G}}) = d^{\mathrm{F}}_{\mathcal{H}}(\{s^{\alpha_i}\}, \mathcal{S}^{\alpha_i}_{\mathcal{G}})$, and therefore $h^{\mathrm{F}}_{\rho}[\mathcal{H}](s) = d^{\mathrm{F}}_{\mathcal{H}}(\{\alpha_1(s), \dots, \alpha_n(s)\}, \bigcup_i \mathcal{S}^{\alpha_i}_{\mathcal{G}}) = \max_{1 \le i \le n} d^{\mathrm{F}}_{\mathcal{H}}(\{\alpha_i(s)\}, \mathcal{S}^{\alpha_i}_{\mathcal{G}}) = \max_{1 \le i \le n} h^{\alpha_i}(s)$.

In the construction of $\mathcal{H}$, we still did not make use of the expressiveness of hypergraphs. The consideration of hypergraphs instead of simple transition systems allows drawing connections *between* the different abstractions. Consider for example two projections, one on variable $v$, one on $u$. If there is an action that modifies both $v$ and $u$, then every hyperedge labeled by this action can be extended by connections to states in both projections. Moreover, adding more abstract states to the tail (in the $\mathrm{B}$ case) and to the head (in the $\mathrm{F}$ case), can never cause a decrease in distance. Thus adding such connections can only be beneficial for the resulting heuristic. In other words, hyperabstractions can be seen as a new method to admissibly combine a set of abstractions, which dominates just taking the maximum:

**Theorem 6** *Let $\alpha_1, \dots, \alpha_n$ be any abstractions. Consider $\rho$ as described in the text. One can always construct an $\mathrm{F}$-interpretation $\mathcal{H}$ of $\rho$ in time polynomial in the size of $\Theta^{\alpha_1}, \dots, \Theta^{\alpha_n}$ such that $h^{\mathrm{F}}_{\rho,\mathcal{H}} \geq \max_{1 \leq i \leq n} h^{\alpha_i}$. There are cases where this dominance holds strictly. If $\Pi$ is in TNF, the same holds for backward hyperabstraction heuristics.*

This result applies to the additive combination of abstractions as well. In particular, given any set of abstraction heuristics that are additively combined via a cost partitioning $\mathbf{c}$, Theorem 6 can easily be extended to show the existence of a *single* hyperabstraction heuristic that, under the application of the same $\mathbf{c}$, dominates the additive ensemble:

**Corollary 1** *For every ensemble of abstractions $\alpha_1, \dots, \alpha_n$, a single $\mathrm{F}$-hyperabstraction heuristic $h^{\mathrm{F}}_{\rho}[\mathcal{H}]$ can be constructed in polynomial time (in the size of the abstract state space) such that it holds, for every cost partitioning $\mathbf{c}$, that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}]\!] \geq h_{\alpha_1,\dots,\alpha_n}[\![\mathbf{c}]\!]$. If $\Pi$ is in TNF, the same holds for $\mathrm{B}$ hyperabstraction heuristics.*

## Optimal Cost Partitioning

Plugging multiple cost functions into a single hyperabstraction heuristic provides a very powerful formalism able to dominate arbitrary additive ensembles of abstraction heuristics as well as the application of cost partitionings to critical-path heuristics. This raises the question of whether we can do anything more with additive hyperabstractions. To answer this question to at least some extent, we next show that in contrast to additive abstractions, it is hard to compute the optimal cost partitioning for hyperabstraction heuristics. The proof works via a detour to the critical-path heuristic $h^1$:

**Theorem 7** *Optimal cost partitioning for $h^1$ is **NP**-hard.*

Since critical-path heuristics can be seen as particular $\mathrm{B}$-interpretations of a hyperabstraction, this result immediately carries over to backward hyperabstraction heuristics. Moreover, due to the symmetric nature of the definitions of $\mathrm{B}$- and $\mathrm{F}$-hyperabstraction heuristics, it is not difficult to extend the proof for $h^1$ to work also for $\mathrm{F}$-interpretations. We obtain:

**Corollary 2** *Optimal cost partitioning for hyperabstraction heuristics in general is **NP**-hard.*

**Algorithm 1:** Generic algorithm for generating all $\mathrm{B}$-hyperedges with head $t^{\alpha_i} \in \mathcal{S}^{\alpha_i} \subseteq \mathcal{P}$, and action $a \in \mathcal{A}$. To obtain $\mathrm{F}$-interpretations, the regression operation must be substituted by progression, and the head and the tail of the hyperedges must be swapped.

**Input:** Abstractions $\alpha_1, \dots, \alpha_k$,
abstract state $t^{\alpha_i} \in \mathcal{S}^{\alpha_i}$, action $a \in \mathcal{A}$
**Output:** B-hyperedges $\mathcal{E}(t^{\alpha_i}, a)$

$X \leftarrow \texttt{SelectSplit}(\{\alpha_1, \dots, \alpha_k\}, t^{\alpha_i}, a)$ ;
$Y \leftarrow \texttt{SelectOthers}(\{\alpha_1, \dots, \alpha_k\}, t^{\alpha_i}, a)$ ;
$\mathcal{E}(t^{\alpha_i}, a) \leftarrow \emptyset$ ;
**if** $X = \emptyset$ **then**
  $T_e \leftarrow \{s^\alpha \in \mathcal{S}^\alpha \mid \alpha \in Y, \textit{Regr}([t^{\alpha_i}], a) \subseteq [s^\alpha]\}$;
  $\mathcal{E}(t^\alpha, a) \leftarrow \{\langle T_e, a, \{t^{\alpha_i}\}\rangle\}$
**foreach** $S_X \in \prod_{\alpha \in X} \mathcal{S}^\alpha$ **do**
  $S_{ctxt} \leftarrow \textit{Regr}([t^{\alpha_i}], a) \cap \bigcap_{s^\alpha \in S_X} [s^\alpha]$ ;
  **if** $S_{ctxt} \neq \emptyset$ **then**
    $S_Y \leftarrow \{s^\alpha \in \mathcal{S}^\alpha \mid \alpha \in Y, S_{ctxt} \subseteq [s^\alpha]\}$ ;
    $T_e \leftarrow S_X \cup S_Y$;
    $\mathcal{E}(t^\alpha, a) \leftarrow \mathcal{E}(t^\alpha, a) \cup \{\langle T_e, a, \{t^\alpha\}\rangle\}$ ;
**return** $\mathcal{E}(t^\alpha, a)$ ;

## Practical Construction of Hyperabstractions

Hyperabstraction heuristics consist of two components: (1) the hyperabstraction function $\rho$, and (2) an interpretation of $\rho$ as a hypergraph. Regarding (1), given the vast literature in the automatic construction of abstraction heuristics in classical planning, it is natural to define $\rho$ as the union of abstraction functions $\alpha_1, \dots, \alpha_k$. Each individual abstraction $\alpha_i$ can for example be a PDB, Cartesian, or M&S abstraction. The considered abstractions do not have to be of the same type, though. The set of abstract concepts are given by $\mathcal{P} = \bigcup_{i \in [1,k]} \mathcal{S}^{\alpha_i}$. Each state $s$ is mapped to its corresponding abstract state in each abstraction, i. e., $\rho(s) := \{\alpha_1(s), \dots, \alpha_k(s)\}$.

Regarding (2), we next show a generic method to generate $\mathrm{B}$-hyperedges so to obtain different $\mathrm{B}$-interpretations for $\rho$. The construction of $\mathrm{F}$-interpretations is omitted for the sake of brevity, but works analogously. To provide a unified algorithm that supports all the different abstraction variants, we treat abstract states as the equivalence relations they induce. The algorithm operates directly on these sets of states. In an actual implementation, these sets can of course not be enumerated explicitly, each abstract state possibly representing an exponential number of real states. However, the specific operations used by the algorithm can be implemented efficiently for various kinds of abstractions (particularly Cartesian and PDBs), taking into account the concrete structure and representation of abstract states. For computing the tail of the $\mathrm{B}$-hyperedges, a notion for the regression of sets of states is required. Let $T \subseteq \mathcal{S}$ be any set of states, and $a \in \mathcal{A}$ be any action. We define the regression of $T$ over $a$ as $\textit{Regr}(T, a) = \{s \in \mathcal{S} \mid \langle s, a, t\rangle \in \mathcal{T}, t \in T\}$.

Consider any action $a \in \mathcal{A}$. We next describe the method, Algorithm 1, to generate the $\mathrm{B}$-hyperedges for this action. To satisfy Definition 3, we must add, for every abstract state

| | `SelectSplit` | `SelectOthers` |
|---|---|---|
| (a) $h^m$ | $\emptyset$ | $\{\alpha_1, \ldots, \alpha_k\}$ |
| (b) $\max_{j\in[1,k]} h^{\alpha_j}$ | $\{\alpha_i\}$ | $\emptyset$ |
| (c) combination | $\{\alpha_i\}$ | $\{\alpha_1, \ldots, \alpha_k\}$ |

Table 1: Different instantiations of Algorithm 1.

$t^{\alpha_i} \in \mathcal{P}$ where $Regr([t^{\alpha_i}], a) \neq \emptyset$, at least one representative hyperedge $e = \langle T_e, a, \{t^{\alpha_i}\}\rangle$. Different interpretations may be constructed differing in how many such edges $e$ to consider. There are two extremes:

(i) A single hyperedge that represents all transitions into any state in $[t^{\alpha_i}]$. In this case, $T_e = \{s^\alpha \in \mathcal{P} \mid Regr([t^{\alpha_i}], a) \subseteq [s^\alpha]\}$ includes the maximal number of abstract states, satisfying Definition 3.

(ii) One hyperedge for every concrete transition into any state in $[t^{\alpha_i}]$, i. e., for every state $s \in Regr([t^{\alpha_i}], a)$, the hyperedge $e_{s,a,t^{\alpha_i}}$ with tail $T_{e_{s,a,t^{\alpha_i}}} = \rho(s)$.

(i) gives the simplest "reasonable" choice of a B-interpretation of $\rho$, yet pays this simplicity through possible information loss. The red part in Figure 1 shows an example. On the contrary, (ii) provides the most informative hypergraph possible, but the number of hyperedges is worst-case exponential in the number of abstract states.

To trade-off between heuristic accuracy and computational cost, Algorithm 1 allows to interpolate between the two extremes by means of the methods `SelectSplit` and `SelectOthers`, both choosing a subset of the input abstractions. The abstractions given by `SelectSplit` are used to determine transitions for which to generate separate hyperedges as in (ii). To do so, the hyperedges are built so that each $e \in \mathcal{E}(t^{\alpha_i}, a)$ contains $S_X \subseteq T_e$ for some choice $S_X$ of abstract states for all abstractions selected by `SelectSplit`. Each $e$ is supposed to cover only those transitions $s[\![a]\!] \in [t^{\alpha_i}]$ where $s$ is jointly represented by all abstract states in $S_X$. To ensure that Definition 3 is satisfied, Algorithm 1 computes the collection $\mathcal{E}(t^{\alpha_i}, a)$ in an exhaustive manner, considering, and possibly creating a hyperedge, for every combination of abstract states $S_X \in \prod_{\alpha\in X} \mathcal{S}^\alpha$. For the remaining abstractions $\alpha \notin X$, in order to preserve that every generated hyperedge $e$ with set $S_X$ indeed covers all applications $s[\![a]\!]$ relevant to $S_X$, we may include in the tail of $e$ an abstract state $s^\alpha \in \mathcal{S}^\alpha$ only if $s^\alpha$ represents *all* the states $s$ with such transition (line 10). The method `SelectOthers` determines the abstractions for which to do this analysis, and thus which abstractions to take into account in the computation of the tails in addition to $X$.

Table 1 shows three particular choices of `SelectSplit` and `SelectOthers`, resulting in: (a) the critical-path heuristic $h^m$, considering as abstractions the projections onto all variable subsets of size up to $m$; (b) the maximum over multiple abstractions, and (c) a new combination, which guarantees to dominate both of them.

## Experiments

Given the extensive research on the construction of state-of-the-art abstractions and critical-path heuristics, e. g., (Seipp and Helmert 2018; Helmert et al. 2014; Franco et al. 2017; Steinmetz and Hoffmann 2017), we do not expect to beat those configurations. The goal of our evaluation is to verify whether the theoretical dominance results also show in practice. For that, we investigate whether applying the hyperabstraction construction on top of a set of abstractions is more informative than taking their maximum. Our implementation is in Fast Downward (Helmert 2006). We are using all IPC STRIPS benchmarks of the optimal tracks. The experiments were performed on a cluster of Intel E5-2660 machines running at 2.20 GHz, restricting CPU time to 30 minutes and memory to 4 GB. We use $PDB(m)$: pattern databases of size up to $m \in \{1, 2, 3\}$; as well as Cartesian abstractions constructed via the CEGAR approach (Seipp and Helmert 2018) as seed abstractions. The hyperabstraction interpretations are computed using Algorithm 1 with the parameters as shown in Table 1(c).

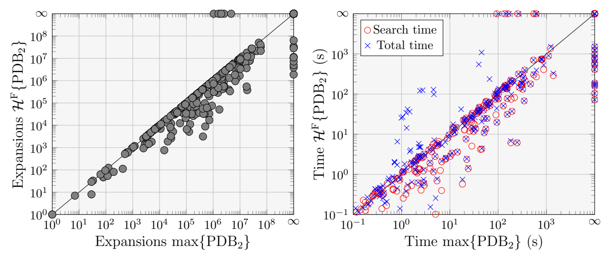


Figure 2: Comparison of the F-hyperabstraction based on $PDB(2)$ versus taking their maximum in terms of expansions until the last $f$-layer and runtime.

Figure 2 compares the systematic PDBs of size 2 against the F-interpretation of the corresponding hyperabstraction. The F-interpretation of abstraction heuristics can provide more informative estimates, reducing the number of expansions in some cases compared to taking the maximum among them. This result can be expected to carry over to the additive scenario where the same cost-partitioning is used for abstractions and hyperabstractions. Moreover, F-interpretations do not have a large overhead over abstraction heuristics in evaluation time, though the comparison in total time shows that the construction is more expensive.

Applying the hyperabstraction construction on the CEGAR abstractions did not have a considerable effect on the heuristic accuracy, regardless of considering F- or B-hyperabstraction heuristics. An explanation is that the construction of the individual Cartesian abstractions focuses on different parts of the planning task.

## Conclusion

In this paper we introduced hyperabstractions, a new family of heuristics that generalizes abstractions and critical-path heuristics, dominating most admissible heuristics in the literature. Hyperabstractions map each state to a set of abstract states. This is related to more general notions of abstractions used in model-checking, where the same state may be mapped into multiple abstract predicates (Cousot and Cousot 1977; Ball, Podelski, and Rajamani 2001). Previous multimapping abstractions for planning resulted in less informed estimates than taking the maximum over several abstraction heuristics (Pang and Holte 2011). We circumvent this by computing distances in an hypergraph instead, ensuring that hyperabstractions dominate the corresponding abstraction heuristics, sometimes strictly.

Hyperabstractions have potential for obtaining stronger heuristics than previous families of heuristics, e.g., as a novel method to combine different abstractions, or by explicitly reasoning about the hypergraphs underlying the computation of critical-path heuristics. However, it is yet unknown how to exploit their full potential. Promising lines for future research include finding automatic methods to derive hyperabstractions (i.e., identifying cases where there may be oportunities to combine abstractions via an hyperabstraction), and additive ensembles thereof.

**Acknowledgments.** Thanks to the anonymous reviewers for their very helpful comments. This work has been partially supported by German Federal Ministry of Education and Research (BMBF) through funding for CISPA (grant no. 16KIS0656), and the German Research Foundation (DFG), under grant HO 2169/5-1. The FAI research group at Saarland University has received support by DFG grant 389792660 as part of TRR 248 (see https://perspicuous-computing.science).

## References

Ausiello, G., and Laura, L. 2017. Directed hypergraphs: Introduction and fundamental algorithms—a survey. *Theoretical Computer Science* 658:293 – 306.

Bäckström, C. 1995. Expressive equivalence of planning formalisms. *Artificial Intelligence* 76(1–2):17–34.

Ball, T.; Podelski, A.; and Rajamani, S. K. 2001. Boolean and Cartesian abstraction for model checking C programs. In Margaria, T., and Yi, W., eds., *Proceedings of the 7th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'01)*, 268–283. Springer.

Bonet, B., and Geffner, H. 2001. Planning as heuristic search. *Artificial Intelligence* 129(1–2):5–33.

Cousot, P., and Cousot, R. 1977. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *Proceedings of the 5th ACM Symposium on Principles of Programming Languages (POPL'77)*, 238–252.

Edelkamp, S. 2001. Planning with pattern databases. In Cesta, A., and Borrajo, D., eds., *Proceedings of the 6th European Conference on Planning (ECP'01)*, 13–24. Springer-Verlag.

Edelkamp, S. 2002. Symbolic pattern databases in heuristic search planning. In Ghallab, M.; Hertzberg, J.; and Traverso, P., eds., *Proceedings of the 6th International Conference on Artificial Intelligence Planning and Scheduling (AIPS'02)*, 274–283. Toulouse, France: AAAI Press, Menlo Park.

Edelkamp, S. 2006. Automated creation of pattern database search heuristics. In *Proceedings of the 4th Workshop on Model Checking and Artificial Intelligence (MoChArt 2006)*, 35–50.

Fickert, M., and Hoffmann, J. 2017. Complete local search: Boosting hill-climbing through online heuristic-function refinement. In *Proceedings of the 27th International Conference on Automated Planning and Scheduling (ICAPS'17)*. AAAI Press.

Franco, S.; Torralba, A.; Lelis, L. H.; and Barley, M. 2017. On creating complementary pattern databases. In Sierra, C., ed., *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17)*. AAAI Press/IJCAI.

Gallo, G.; Longo, G.; and Pallottino, S. 1993. Directed hypergraphs and applications. *Discrete Applied Mathematics* 42(2):177–201.

Haslum, P., and Geffner, H. 2000. Admissible heuristics for optimal planning. In Chien, S.; Kambhampati, R.; and Knoblock, C., eds., *Proceedings of the 5th International Conference on Artificial Intelligence Planning Systems (AIPS'00)*, 140–149. Breckenridge, CO: AAAI Press, Menlo Park.

Haslum, P.; Botea, A.; Helmert, M.; Bonet, B.; and Koenig, S. 2007. Domain-independent construction of pattern database heuristics for cost-optimal planning. In Howe, A., and Holte, R. C., eds., *Proceedings of the 22nd National Conference of the American Association for Artificial Intelligence (AAAI'07)*, 1007–1012. Vancouver, BC, Canada: AAAI Press.

Haslum, P.; Bonet, B.; and Geffner, H. 2005. New admissible heuristics for domain-independent planning. In Veloso, M. M., and Kambhampati, S., eds., *Proceedings of the 20th National Conference of the American Association for Artificial Intelligence (AAAI'05)*, 1163–1168. Pittsburgh, Pennsylvania, USA: AAAI Press.

Haslum, P. 2006. Improving heuristics through relaxed search - an analysis of TP4 and HSP*a in the 2004 planning competition. *Journal of Artificial Intelligence Research* 25:233–267.

Haslum, P. 2009. $h^m(P) = h^1(P^m)$: Alternative characterisations of the generalisation from $h^{\max}$ to $h^m$. In Gerevini, A.; Howe, A.; Cesta, A.; and Refanidis, I., eds., *Proceedings of the 19th International Conference on Automated Planning and Scheduling (ICAPS'09)*, 354–357. AAAI Press.

Helmert, M., and Domshlak, C. 2009. Landmarks, critical paths and abstractions: What's the difference anyway? In Gerevini, A.; Howe, A.; Cesta, A.; and Refanidis, I., eds., *Proceedings of the 19th International Conference on Automated Planning and Scheduling (ICAPS'09)*, 162–169. AAAI Press.

Helmert, M.; Haslum, P.; Hoffmann, J.; and Nissim, R. 2014. Merge & shrink abstraction: A method for generating lower bounds in factored state spaces. *Journal of the Association for Computing Machinery* 61(3):16:1–16:63.

Helmert, M. 2006. The Fast Downward planning system. *Journal of Artificial Intelligence Research* 26:191–246.

Hoffmann, J.; Porteous, J.; and Sebastia, L. 2004. Ordered landmarks in planning. *Journal of Artificial Intelligence Research* 22:215–278.

Katz, M., and Domshlak, C. 2008a. Optimal additive composition of abstraction-based admissible heuristics. In Rintanen, J.; Nebel, B.; Beck, J. C.; and Hansen, E., eds., *Proceedings of the 18th International Conference on Automated Planning and Scheduling (ICAPS'08)*, 174–181. AAAI Press.

Katz, M., and Domshlak, C. 2008b. Structural patterns heuristics via fork decomposition. In Rintanen, J.; Nebel, B.; Beck, J. C.; and Hansen, E., eds., *Proceedings of the 18th International Conference on Automated Planning and Scheduling (ICAPS'08)*, 182–189. AAAI Press.

Pang, B., and Holte, R. C. 2011. State-set search. In Borrajo, D.; Likhachev, M.; and Linares López, C., eds., *Proceedings of the 4th Annual Symposium on Combinatorial Search (SOCS'11)*. AAAI Press.

Pommerening, F., and Helmert, M. 2015. A normal form for classical planning tasks. In Brafman, R.; Domshlak, C.; Haslum, P.; and Zilberstein, S., eds., *Proceedings of the 25th International Conference on Automated Planning and Scheduling (ICAPS'15)*, 188–192. AAAI Press.

Röger, G.; Pommerening, F.; and Helmert, M. 2014. Optimal planning in the presence of conditional effects: Extending lm-cut with context splitting. In Schaub, T., ed., *Proceedings of the 21st European Conference on Artificial Intelligence (ECAI'14)*, 765–770. Prague, Czech Republic: IOS Press.

Seipp, J., and Helmert, M. 2018. Counterexample-guided Cartesian abstraction refinement for classical planning. *Journal of Artificial Intelligence Research* 62:535–577.

Steinmetz, M., and Hoffmann, J. 2017. State space search nogood learning: Online refinement of critical-path dead-end detectors in planning. *Artificial Intelligence* 245:1–37.

Torralba, Á.; Alcázar, V.; Kissmann, P.; and Edelkamp, S. 2017. Efficient symbolic search for cost-optimal planning. *Artificial Intelligence* 242:52–79.

## Proofs

**Theorem 1** *(i) $h^{\mathrm{X}*}_\rho$ is well-defined, i. e., for every $\rho$ there exists an* X*-interpretation $\mathcal{H}^*$ of $\rho$ as defined in the text, and (ii) given an* X*-interpretation $\mathcal{H}$ of $\rho$ and state $s$, deciding whether $h^{\mathrm{X}}_\rho[\mathcal{H}](s) = h^{\mathrm{X}*}_\rho(s)$ is **NP**-hard.*

*Proof.*

**Part (i)**

B-*interpretation* Let $\Pi = \langle \mathcal{V}, \mathcal{A}, s_\mathcal{I}, \mathcal{G}\rangle$ be any task, and $\rho$ be any hyperabstraction with associated abstract concepts $\mathcal{P}$. Consider the B-hypergraph $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w\rangle$ where $\mathcal{N} = \mathcal{P}$, $\mathcal{L} = \mathcal{A}$, $w = c$, and for every transition $\langle s, a, t\rangle$ in state space of $\Pi$, and for every abstract concept $p_t \in \rho(t)$, $\mathcal{E}$ contains the hyperedge $\langle \rho(s), a, \{p_t\}\rangle$. Observe that (1) $\mathcal{H}$ is a B-interpretation of $\rho$, and (2) that for every other B-interpretation $\mathcal{H}'$ of $\rho$, it holds that $h^{\mathrm{B}}_\rho[\mathcal{H}] \geq h^{\mathrm{B}}_\rho[\mathcal{H}']$. (1) directly follows from the definition of $\mathcal{E}$. For (2), we show that $d^{\mathrm{B}}_{\mathcal{H}}(P, P') \geq d^{\mathrm{B}}_{\mathcal{H}'}(P, P')$ for every $P, P' \subseteq \mathcal{P}$. Due to the maximization in the last case of Equation (2), it suffices to show, for every $P \subseteq \mathcal{P}$ and $p' \in \mathcal{P}$, that $d^{\mathrm{B}}_{\mathcal{H}}(P, \{p'\}) \geq d^{\mathrm{B}}_{\mathcal{H}'}(P, \{p'\})$. If $p' \in P$, the claim follows directly. Otherwise, note that the edges in $\mathcal{H}$ subsume those in $\mathcal{H}'$, i. e., for every $\langle P_s, a, \{p_t\}\rangle \in \mathcal{E}$, $\mathcal{H}'$ must have an edge $\langle P'_s, a, \{p_t\}\rangle$ such that $P'_s \subseteq P_s$. Assume this was not true for some hyperedge $e = \langle P_s, a, \{p_t\}\rangle \in \mathcal{E}$. Let $\langle s, a, t\rangle$ be the transition corresponding to $e$. Then, for every hyperedge $\langle P'_s, a, \{p_t\}\rangle \in \mathcal{E}'$, it holds that $P'_s \not\subseteq P_s = \rho(s)$, i. e., $\mathcal{H}'$ does not contain any hyperedge representing the transition $\langle s, a, t\rangle$ with $p_t \in \rho(t)$. This is a contradiction to the assumption, so such $e \in \mathcal{E}$ cannot exist. Clearly, for every $P'' \subseteq P' \subseteq \mathcal{P}$, it holds that $d^{\mathrm{B}}_{\mathcal{H}}(P, P'') \leq d^{\mathrm{B}}_{\mathcal{H}}(P, P')$. Now, consider any hyperedge $\langle P_s, a, \{p'\}\rangle \in \mathcal{E}$ that minimizes the middle in the definition of $d^{\mathrm{B}}_{\mathcal{H}}(P, \{p'\})$. Since $\mathcal{H}'$ must contain an hyperedge $\langle P'_s, a, \{p'\}\rangle \in \mathcal{E}'$ with $P'_s \subseteq P_s$, $d^{\mathrm{B}}_{\mathcal{H}'}(P, \{p'\})$ can hence not be larger than $d^{\mathrm{B}}_{\mathcal{H}}(P, \{p'\})$. (1) and (2) together show the claim.

F-*interpretation* For every $\Pi$ and $\rho$, construct a hypergraph $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w\rangle$ as follows: $\mathcal{N} = \mathcal{P}$, $\mathcal{L} = \mathcal{A}$, $w = c$, and $\mathcal{E}$ contains $\langle \{p_s\}, a, \rho(t)\rangle$ for every transition $\langle s, a, t\rangle$ in $\Theta^\Pi$ and $p_s \in \rho(s)$. Proofing that this hypergraph $\mathcal{H}$ is indeed a *perfect* F-interpretation of $\rho$ is done symmetrically to the B case.

**Part (ii)** For arbitrary sets of abstract states (e.g. defined via arbitrary CNF formulas over the state variables), computing whether an action induces a transition between two abstract states is already **NP**-hard, since it requires to compute the regression/progression of an action over an arbitrary formula. However, our claim holds even restricted to conjunctive abstract concepts only, where progression and regression can be easily computed in polynomial time.

The proof is by reduction from SAT. We next describe a family of tasks containing one planning task for every CNF formula with non-empty clauses. Let $\phi$ be any CNF formula, then $\neg\phi$ is a DNF formula with non-empty clauses $C_1, \dots, C_m$ and variables $X_1, \dots, X_n$. The corresponding task $\Pi_{\neg\phi} = \langle \mathcal{V}, \mathcal{A}, s_\mathcal{I}, \mathcal{G}\rangle$ has Boolean variables $\mathcal{V} = \{x_1, \dots, x_n, y, g\}$; initial state $s_\mathcal{I}(v) = 0$ for all $v \in \mathcal{V}$; goal $\mathcal{G} = \{g = 1\}$; and the following two unit-cost actions:

- Action $\mathrm{setY}$ with precondition $y = 0$ and effect $y = 1$.
- Action $\mathrm{setG}$ with precondition $\{g = 0, y = 1\}$ and effect $g = 1$.

An optimal plan for $s_\mathcal{I}$ is given by $\langle \mathrm{setY}, \mathrm{setG}\rangle$, i. e., $h^*(s_\mathcal{I}) = 2$. Below, We provide two hyperabstractions $\rho^{\mathrm{F}}, \rho^{\mathrm{B}}$ such that $h^{\mathrm{F}*}_{\rho^{\mathrm{F}}}(s_\mathcal{I}) = 2$ and $h^{\mathrm{B}*}_{\rho^{\mathrm{B}}}(s_\mathcal{I}) = 2$ if and only if $\neg\phi$ is a tautology (and hence, if and only if $\phi$ is satisfiable).

For the sake of simplicity, for any literal $L$ and state $s$, we write $s \models L$ if $s(x_j) = 0$ and $L = \neg X_j$, or $s(x_j) = 1$ and $L = X_j$. Otherwise $s \not\models L$. Similarly, for every state $s$ and clause $C_i$, we write $s \models C_i$ if $s \models L$ for all literals $L \in C_i$, and $s \not\models C_i$ otherwise.

F-*interpretation* Let $\mathcal{P}^{\mathrm{F}} = \{\mathsf{C}_1, \ldots, \mathsf{C}_\mathsf{m}, \mathsf{I}, \mathsf{Y}, \mathsf{G}\}$, and consider the following hyperabstraction function:

$$\begin{aligned}\rho^{\mathrm{F}}(s) :=& \{\mathsf{C}_\mathsf{j} \mid j \in [1,m], s(g)=0, s(y)=1, s \models C_j\} \\ & \cup \{\mathsf{I} \mid s(y)=0 \wedge s(g)=0\} \\ & \cup \{\mathsf{Y} \mid s(y)=1\} \\ & \cup \{\mathsf{G} \mid s(g)=1\}\end{aligned}$$

The key is that the set of concepts $\mathsf{C}_1, \ldots, \mathsf{C}_\mathsf{m}$ represent the set of all states such that $s(y) = 1 \wedge s(g) = 0$ if and only if $\neg\phi$ is a tautology. Only in that case, one can avoid creating a hyperedge from $\mathsf{I}$ to $\mathsf{Y}$, which increases the heuristic value because some goal states are mapped into $\mathsf{Y}$.

**$\phi$ is not satisfiable:** Then, $\neg\phi$ is a tautology. We show that $h^{\mathrm{F}*}_{\rho}(s_\mathcal{I}) = 2$.

Consider the following F-hypergraph $\mathcal{H}^{\mathrm{F}} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w\rangle$ where $\mathcal{N} = \mathcal{P}^{\mathrm{F}}$, $\mathcal{L} = \mathcal{A}$, $w = c$, and $\mathcal{E}$ contains the following hyperedges:

(1) $\langle\{\mathsf{I}\}, \mathrm{setY}, \{\mathsf{C}_\mathsf{j}\}\rangle$ for every $j \in [1,m]$
(2) $\langle\{\mathsf{G}\}, \mathrm{setY}, \{\mathsf{G}\}\rangle$,
(3) $\langle\{\mathsf{C}_\mathsf{j}\}, \mathrm{setG}, \{\mathsf{G}\}\rangle$ for $j \in [1,m]$,
(4) $\langle\{\mathsf{Y}\}, \mathrm{setG}, \{\mathsf{Y}\}\rangle$

This satisfies the definition of F-interpretation because:

- $\mathsf{I}$: setG is not applicable in any state $s$ with $\mathsf{I} \in \rho(s)$. Consider any transition $\langle s, \mathrm{setY}, t\rangle$ with $\mathsf{I} \in \rho(s)$. Because $\mathsf{I} \in \rho(s)$, and due to the definition of setY, it holds that $t(g) = 0$ and $t(y) = 1$. Moreover, since $\neg\phi$ is a tautology, there must exist some $j \in [1,m]$ such that $t \models C_j$. Therefore, $\mathsf{C}_\mathsf{j} \in \rho(t)$. By construction, (1), $\langle\{\mathsf{I}\}, \mathrm{setY}, \{\mathsf{C}_\mathsf{j}\}\rangle \in \mathcal{E}$, i. e., $\langle s, \mathrm{setY}, t\rangle$ is indeed represented in our F-hypergraph.
- $\mathsf{G}$: setG is not applicable in any state $s$ with $\mathsf{G} \in \rho(s)$. Moreover, $\mathsf{G}$ is invariant under the application of setY, i. e., (2) indeed covers all transitions $\langle s, \mathrm{setY}, t\rangle$ with $\mathsf{G} \in \rho(s)$.
- $\mathsf{C}_\mathsf{j}$: setY is not applicable in any state $s$ with $\mathsf{C}_\mathsf{j} \in \rho(s)$. Moreover, every application of setG necessarily makes true $\mathsf{G}$. In particular, (3) covers all transitions $\langle s, \mathrm{setG}, t\rangle$ with $\mathsf{C}_\mathsf{j} \in \rho(s)$.
- $\mathsf{Y}$: setY is not applicable in $s$ if $\mathsf{Y} \in \rho(s)$. Moreover, $\mathsf{Y}$ is invariant under the application of setG.

With this interpretation,

$$\begin{aligned}2 &= h^*(s_\mathcal{I}) \\ &\geq h^{\mathrm{F}*}_{\rho}(s_\mathcal{I}) \\ &\geq h^{\mathrm{F}}_{\rho}[\mathcal{H}^{\mathrm{F}}](s_\mathcal{I}) \\ &= d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\rho(s_\mathcal{I}), \bigcup_{s_\mathcal{G} \in \mathcal{S}_\mathcal{G}} \rho(s_\mathcal{G})) \\ &= d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\mathsf{I}\}, \{\mathsf{G}, \mathsf{Y}\}) \\ &\overset{(1)}{=} \min_{j \in [1,m]} 1 + d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\mathsf{C}_\mathsf{j}\}, \{\mathsf{G}, \mathsf{Y}\}) \\ &\overset{(3)}{=} 1 + 1 + d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\mathsf{G}\}, \{\mathsf{G}, \mathsf{Y}\}) = 2\end{aligned}$$

and therefore $h^{\mathrm{F}*}_{\rho^{\mathrm{F}}}(s_\mathcal{I}) = 2$.

**$\phi$ is satisfiable:** We show that $h^{\mathrm{F}*}_{\rho}(s_\mathcal{I}) < 2$.

Let $V : \{X_1, \ldots, X_n\} \mapsto \{0,1\}$ be a satisfying assignment to $\phi$. Then, for all clauses $C_i$ in $\neg\phi$, $V$ does not satisfy the clause. Let $\mathcal{H}^{\mathrm{F}}$ be any F-interpretation of $\rho$. Consider the state $s$ where $s(g) = 0$, $s(y) = 0$ and $s(x_i) = V(X_i)$ for all $i \in [1,n]$. Clearly, $\mathsf{I} \in \rho(s)$, and setY is applicable in $s$. Consider the transition $\langle s, \mathrm{setY}, t\rangle$. Due to Definition 2, $\mathcal{H}^{\mathrm{F}}$ must contain a hyperedge $e_\mathsf{Y} = \langle\{\mathsf{I}\}, \mathrm{setY}, P_t\rangle$ such that $P_t \subseteq \rho(t)$. Since $s \not\models C_j$ for every $j \in [1,m]$ by construction, and setY does not change the value of any $x_i$ variable, it follows that $t \not\models C_j$ for every $j \in [1,m]$. Moreover, from $s(g) = 0$ it follows that $t(g) = 0$. In conclusion, $P_t \subseteq \rho(t) = \{\mathsf{Y}\}$. We obtain

$$\begin{aligned}h^{\mathrm{F}}_{\rho}[\mathcal{H}^{\mathrm{F}}](s_\mathcal{I}) &= d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\mathsf{I}\}, \{\mathsf{Y}, \mathsf{G}\}) \\ &\overset{e_\mathsf{Y}}{\leq} 1 + d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\mathsf{Y}\}, \{\mathsf{Y}, \mathsf{G}\}) \\ &= 1 < 2\end{aligned}$$

and hence $h^{\mathrm{F}*}_{\rho}(s_\mathcal{I}) < 2$ as desired.

B-*interpretation* The proof for B-hyperabstractions is similar. Let $\mathcal{P}^{\mathrm{B}} = \{\mathsf{C}_1, \ldots, \mathsf{C}_\mathsf{m}, \mathsf{I}, \overline{\mathsf{G}}, \mathsf{G}\}$, and consider the following hyperabstraction function:

$$\begin{aligned}\rho^{\mathrm{B}}(s) :=& \{\mathsf{C}_\mathsf{j} \mid j \in [1,m], s(g)=0, s(y)=1, s \models C_j\} \\ & \cup \{\mathsf{I} \mid s(y)=0 \wedge s(g)=0\} \\ & \cup \{\overline{\mathsf{G}} \mid s(g)=0\} \\ & \cup \{\mathsf{G} \mid s(g)=1\}\end{aligned}$$

**$\phi$ is not satisfiable:** Consider the B-hypergraph $\mathcal{H}^{\mathrm{B}} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, s_\mathcal{I}\rangle$ with $\mathcal{N} = \mathcal{P}^{\mathrm{B}}$, $\mathcal{L} = \mathcal{A}$, $w = c$, and the following hyperedges: (1) $\langle\{\mathsf{I}\}, \mathrm{setY}, \{\mathsf{C}_\mathsf{j}\}\rangle$ for $j \in [1,m]$; (2) $\langle\{\mathsf{G}\}, \mathrm{setY}, \{\mathsf{G}\}\rangle$; (3) $\langle\{\mathsf{C}_\mathsf{j}\}, \mathrm{setG}, \{\mathsf{G}\}\rangle$ for $j \in [1,m]$; and (4) $\langle\{\overline{\mathsf{G}}\}, \mathrm{setY}, \{\overline{\mathsf{G}}\}\rangle$. Showing that $\mathcal{H}^{\mathrm{B}}$ is indeed a B-interpretation of $\rho^{\mathrm{B}}$ can be done as in the F case. The most interesting case is given by the transitions $\langle s, \mathrm{setG}, t\rangle$ (thus $\mathsf{G} \in \rho(t)$). Due to the precondition of setG, $s(y) = 1$ and $s(g) = 0$. Since $\neg\phi$ is a tautology, there must exist some $j \in [1,m]$ such that $s \models C_j$. Therefore, $\mathsf{C}_\mathsf{j} \in \rho(s)$, and with (3) we thus guarantee to represent all those transitions. Plugging in the heuristic and distance definitions, we obtain

$$\begin{aligned}
2 &= h^*(s_{\mathcal{I}}) \\
&\geq h_\rho^{\mathrm{B}*}(s_{\mathcal{I}}) \\
&\geq h_\rho^{\mathrm{B}}[\mathcal{H}^{\mathrm{B}}](s_{\mathcal{I}}) \\
&= d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\rho(s_{\mathcal{I}}), \bigcap_{s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}} \rho(s_{\mathcal{G}})) \\
&= d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\{\mathsf{I}, \overline{\mathsf{G}}\}, \{\mathsf{G}\}) \\
&\overset{(2)+(3)}{=} \min_{j \in [1,m]} 1 + d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\{\mathsf{I}, \overline{\mathsf{G}}\}, \{\mathsf{C_j}\}) \\
&\overset{(1)}{=} 1 + 1 + d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\{\mathsf{I}, \overline{\mathsf{G}}\}, \{\mathsf{I}\}) \\
&= 2
\end{aligned}$$

and thus $h_\rho^{\mathrm{B}*}(s_{\mathcal{I}}) = 2$.

**$\phi$ is satisfiable:** Let $V : \{X_1, \dots, X_n\} \mapsto \{0, 1\}$ be a satisfying assignment to $\phi$. Then, for all clauses $C_j$ in $\neg\phi$, $V$ does not satisfy the clause. Let $\mathcal{H}^{\mathrm{B}}$ be any B-interpretation of $\rho$. Consider the state $s$ where $s(g) = 0$, $s(y) = 1$ and $s(x_i) = V(X_i)$ for all $i \in [1, n]$, and consider the transition $\langle s, \mathrm{setG}, t\rangle$. From the selection of $s$, it follows that $s \not\models C_j$ for every $j \in [1, m]$, and thus $\rho(s) = \{\overline{\mathsf{G}}\}$. Due to setG, it holds that $\mathsf{G} \in \rho(t)$. Definition 3 therefore requires $\mathcal{H}^{\mathrm{B}}$ to contain a hyperedge $e_{\mathsf{G}} = \langle P_s, \mathrm{setG}, \{\mathsf{G}\}\rangle$ with $P_s \subseteq \rho(s) = \{\overline{\mathsf{G}}\}$. We get

$$\begin{aligned}
h_\rho^{\mathrm{B}}[\mathcal{H}^{\mathrm{B}}](s_{\mathcal{I}}) &= d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\{\mathsf{I}, \overline{\mathsf{G}}\}, \{\mathsf{G}\}) \\
&\overset{e_{\mathsf{G}}}{\leq} 1 + d_{\mathcal{H}^{\mathrm{B}}}^{\mathrm{B}}(\{\mathsf{I}, \overline{\mathsf{G}}\}, \{\overline{\mathsf{G}}\}) \\
&= 1 < 2
\end{aligned}$$

Hence, $h_\rho^{\mathrm{B}*}(s_{\mathcal{I}}) < 2$.

□

In a previous version of the paper, to prove that it is hard to compute the optimal interpretation of an hyperabstraction, we claimed that there exist families of planning tasks whose optimal interpretation is exponential in size in the number of abstract concepts (see Theorem 8 below). This is only the case for B-interpretations, and does not hold for F-interpretations. However, as Theorem 1 shows, finding the optimal F-interpretation is still hard even if it has polynomial size.

**Theorem 8** *There are families of tasks $\Pi_n$ and hyperabstractions $\rho_n$ with abstract concepts $\mathcal{P}_n$ where every* B*-interpretation $\mathcal{H}$ of $\rho$ with $h_\rho^{\mathrm{B}}[\mathcal{H}] = h_\rho^{\mathrm{B}*}$ is exponentially large in $|\mathcal{P}_n|$.*

*Proof.* Consider the family of planning tasks $\Pi_n$ with Boolean variables $\mathcal{V} = \{g, u_1, \dots, u_n, v_1, \dots, v_n\}$, goal $\mathcal{G} = \{g = 1\}$, and actions $\mathcal{A} = \{\mathrm{set}(1), \dots, \mathrm{set}(n), \mathrm{finish}\}$ where $\mathrm{set}(i)$ has precondition $u_i = 0$ and effect $u_i = 1$, and finish has precondition $\{g = 0, u_1 = 1, \dots, u_n = 1\}$ and effect $g = 1$. All actions have cost 1. The initial state is not important since, here, we are particularly interested in hyperabstractions $\rho$ and B-interpretations $\mathcal{H}$ where $h_\rho^{\mathrm{B}}[\mathcal{H}] = h_\rho^{\mathrm{B}*}$, i. e., where $h_\rho^{\mathrm{B}}[\mathcal{H}](s) = h_\rho^{\mathrm{B}*}(s)$ for *every* state $s$.

The variables $v_1, \dots, v_n$ are irrelevant in this task, since neither of them is affected by any action or appears in the goal. We however make use of them in the design of the hyperabstraction function. Namely, consider as abstract concepts the fact conjunctions $\mathcal{P} = \{\{g = 1\}\} \cup \{\{u_i = 1, v_i = d\} \mid i \in [1, n], d \in \{0, 1\}\}$. We define the hyperabstraction function accordingly: $\rho(s) = \{C \in \mathcal{P} \mid C \subseteq s\}$. For simplicity, we will refer by $C_i^d$ to the conjunction $\{u_i = 1, v_i = d\}$, where $d \in \{0, 1\}$. $C_{\mathcal{G}}$ is used to denote the conjunction $\{g = 1\}$. Note that these abstract concepts only allow to make statements over the $u_i$ value when distinguishing between the assignments to $v_i$ at the same time.

Consider the set of all non-goal states $\hat{\mathcal{S}}$ where $s \in \hat{\mathcal{S}}$ iff $s(g) = 0$ and $s(u_i) = 0$ for at least one $i \in [1, n]$. We first show that (1) $h_\rho^{\mathrm{B}*}(s) \geq 2$ for all such states. We then show (2) for every B-interpretation $\mathcal{H}$ of $\rho$, if $h_\rho^{\mathrm{B}}[\mathcal{H}](s) = h_\rho^{\mathrm{B}*}(s) \geq 2$ holds for all $s \in \hat{\mathcal{S}}$, then $\mathcal{H}$ needs to contain (at least) $2^n$ many hyperedges.

(1) We give the construction of a B-interpretation $\mathcal{H}$ of $\rho$ such that $h_\rho^{\mathrm{B}}[\mathcal{H}](s) \geq 2$ for all $s \in \hat{\mathcal{S}}$. directly. $\mathcal{H}$'s edges are defined as follows (omitting self-loops for the sake of brevity): for every $i \in [1, n]$ and $d \in \{0, 1\}$, $\langle \emptyset, \mathrm{set}(i), \{C_i^d\}\rangle \in \mathcal{E}$; for every $d_1, \dots, d_n \in \{0, 1\}^n$, $\langle \{C_1^{d_1}, \dots, C_n^{d_n}\}, \mathrm{finish}, \{C_{\mathcal{G}}\}\rangle \in \mathcal{E}$. Note that $\mathcal{H}$ is indeed a B-interpretation of $\rho$: $\mathrm{set}(i)$ is the only action which can possibly achieve $C_j^d$, i. e., the only action labeling a transition $\langle s, a, t\rangle$ with $C_j^d \in \rho(t)$ but $C_j^d \notin \rho(s)$, and since $\emptyset \subseteq \rho(s)$, this hyperedge clearly covers all such transitions; finish is the only action achieving $C_{\mathcal{G}}$, and $\mathcal{H}$ contains a separate finish hyperedge for every transition in $\Theta^\Pi$ where $C_{\mathcal{G}}$ is made true.

Let $s \in \hat{\mathcal{S}}$ be arbitrary, and consider any conjunction $C_j^d \not\subseteq s$, i. e., $C_j^d$ such that $s(u_j) = 0$ or $s(v_j) \neq d$. Hence, by construction, $d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), \{C_j^d\}) = 1 + d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), \emptyset) = 1$. Furthermore, since $s(g) = 0$, it holds that $d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), \{C_{\mathcal{G}}\}) = 1 + \min_{\langle T, \mathrm{finish}, \{C_{\mathcal{G}}\}\rangle \in \mathcal{E}} d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), T)$. Every edge $e$ in this minimization contains in $T$, for every $k \in [1, n]$, one of the two $C_k^0$ and $C_k^1$. In particular, $T$ must contain $C_j^0$ or $C_j^1$ for the $j$ where $s(u_j) = 0$. Therefore, $d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), T) \geq 1$, i. e., $h_\rho^{\mathrm{B}*}(s) \geq h_\rho^{\mathrm{B}}[\mathcal{H}](s) = d_{\mathcal{H}}^{\mathrm{B}}(\rho(s), \{C_{\mathcal{G}}\}) \geq 1 + 1 = 2$.

(2) Let $\mathcal{H}$ be any B-interpretation of $\rho$ such that $h_\rho^{\mathrm{B}}[\mathcal{H}](s) \geq 2$ for all $s \in \hat{\mathcal{S}}$. We first show (*) for every hyperedge $e = \langle T_e, \mathrm{finish}, \{C_{\mathcal{G}}\}\rangle$ in $\mathcal{H}$ where (a) $C_{\mathcal{G}} \notin T_e$, and (b) $T_e \subseteq \rho(s)$ for some $s \in \mathcal{S}$, it holds for every $i \in [1, n]$ that either $C_i^0 \in T_e$ or $C_i^1 \in T_e$. Condition (a) filters out self-loops. Condition (b) removes from consideration those hyperedges that do not represent any transition. Assume that (*) was not true, and let $e = \langle T_e, \mathrm{finish}, \{C_{\mathcal{G}}\}\rangle$ be an hyperedge of $\mathcal{H}$ such that (a) and (b) are satisfied, but $C_i^0, C_i^1 \notin T_e$ for some $i$. We choose the following

state $s$ from $\hat{\mathcal{S}}$: $s(g) = s(u_i) = 0$; $s(u_j) = 1$ where $j \neq i$; $s(v_j) = d$ if $C_j^d \in T_e$ (assumption (b) makes sure that $d$ is unique); and arbitrary value $s(v_j)$ otherwise. Therefore, $T_e \subseteq \rho(s)$, i. e., $d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), \{C_{\mathcal{G}}\}) \leq 1 + d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), T_e) = 1 + 0 = 1 < 2$, a contradiction to the assumption.

Finally observe that, for every possible combination of values $d_1, \dots, d_n \in \{0, 1\}^n$, $\mathcal{H}$ needs to contain a finish-hyperedge $e$ with tail $T_e = \{C_1^{d_1}, \dots, C_n^{d_n}\}$. Consider any state $s$ where $s(g) = 0$ and $s(u_i) = 1$ for all $i \in [1, n]$. Let $d_i := s(v_i)$ for every $i \in [1, n]$. Clearly, finish is applicable in $s$, leading to a state where $C_{\mathcal{G}}$ is satisfied. Hence, by Definition 3, $\mathcal{H}$ must contain a hyperedge $e_s = \langle T_{e_s}, \text{finish}, \{C_{\mathcal{G}}\}\rangle$ such that $T_{e_s} \subseteq \rho(s) = \{C_1^{d_1}, \dots, C_n^{d_n}\}$. Via (*), it follows that $\{C_1^{d_1}, \dots, C_n^{d_n}\} \subseteq T_{e_s}$. $T_{e_s}$ cannot contain any other conjunction without violating $T_{e_s} \subseteq \rho(s)$. Therefore, $\mathcal{H}$ needs to contain a separate hyperedge $e_{s'}$ for every non-goal state $s'$ where $s'(u_i) = 1$ for every $1 \in [1, n]$. There are $2^n$ such states. This hence shows that $\mathcal{H}$ has to contain at least $2^n$ many hyperedges in order to encode $h^{\mathrm{B}*}_\rho$ in this particular task, and for this particular choice of $\rho$.

□

Theorem 8 does not hold for F-interpretations. In a nutshell, for every $\rho$, we can find a F-interpretation $\mathcal{H}$ such that $\mathcal{H}$ is a BF-hypergraph and $h^{\mathrm{F}}_\rho[\mathcal{H}] = h^{\mathrm{F}*}_\rho$: Denote by $\mathcal{P}_{\mathcal{G}}$ the goal concepts, i. e., $\mathcal{P}_{\mathcal{G}} = \bigcup_{s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}} \rho(s_{\mathcal{G}})$. Given $\mathcal{H}^*$, construct $\mathcal{H}$ as follows: for every hyperedge $\langle\{p\}, a, H_e\rangle \in \mathcal{E}^*$, select some $q \in H_e$ such that $d^{\mathrm{F}}_{\mathcal{H}^*}(\{q\}, \mathcal{P}_{\mathcal{G}}) = d^{\mathrm{F}}_{\mathcal{H}^*}(H_e, \mathcal{P}_{\mathcal{G}})$, and add to $\mathcal{H}$ the edge $\langle\{p\}, a, \{q\}\rangle$. Obviously, $h^{\mathrm{F}*}_\rho = h^{\mathrm{F}}_\rho[\mathcal{H}]$, but the number of different edges $\langle\{p\}, a, \{q\}\rangle$ is at most quadratic in the number of abstract concepts.

**Theorem 2** *For every task* $\Pi$*, hyperabstraction* $\rho$*, and* X*-interpretation* $\mathcal{H}$ *of* $\rho$*,* $h^{\mathrm{X}}_\rho[\mathcal{H}]$ *is consistent and goal-aware.*

*Proof.*

**B-interpretations** Let $\rho$ be any hyperabstraction, and $\mathcal{H}$ be any B-interpretation of $\rho$. Let $P$ denote the set of abstract concepts $\bigcap_{s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}} \rho(s_{\mathcal{G}})$. Obviously, it holds for every goal state $s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}$ that $P \subseteq \rho(s_{\mathcal{G}})$, and thus $h^{\mathrm{B}}_\rho[\mathcal{H}](s_{\mathcal{G}}) = d^{\mathrm{B}}_{\mathcal{H}}(\rho(s_{\mathcal{G}}), P) = 0$. In other words, $h^{\mathrm{B}}_\rho[\mathcal{H}]$ is goal-aware.

Regarding consistency, consider an arbitrary transition $\langle s, a, t\rangle$ in the state space of $\Pi$. For every abstract concept $q_t \in \rho(t)$, it holds that $d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), \{q_t\}) \leq c(a)$: If $q_t \in \rho(s)$ this follows directly. Otherwise, due to Definition 3, $\mathcal{H}$ must contain a hyperedge $\langle T_e, a, \{q_t\}\rangle$ such that $T_e \subseteq \rho(s)$. It directly follows that $d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), \{q_t\}) \leq d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), T_e) + c(a) = 0 + c(a) = c(a)$. That $d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), \{q\}) \leq d^{\mathrm{B}}_{\mathcal{H}}(\rho(t), \{q\}) + c(a)$, for every $q \in \mathcal{P}$, now follows from the definition of the distance function.

Spelling out the details, the distance function $d^{\mathrm{B}}_{\mathcal{H}}$ can be computed via an iterative procedure known as the Bellman-Ford algorithm. The algorithm computes a sequence of monotonically decreasing functions $d^{\mathrm{B}}_0, d^{\mathrm{B}}_1, \dots$ eventually converging to $d^{\mathrm{B}}_{\mathcal{H}}$, i. e., there exists some $i$ such that $d^{\mathrm{B}}_i = d^{\mathrm{B}}_{\mathcal{H}}$. The functions are defined as follows:

$$d^{\mathrm{B}}_0(P, Q) := \begin{cases} 0 & \text{if } Q \subseteq P \\ \infty & \text{otherwise} \end{cases}$$

and

$$d^{\mathrm{B}}_i(P, Q) := \begin{cases} 0 & \text{if } Q \subseteq P \\ \min\limits_{\langle T, a', \{q\}\rangle \in \mathcal{E}} c(a') + d^{\mathrm{B}}_{i-1}(P, T) & \text{if } Q = \{q\} \\ \max\limits_{q \in Q} d^{\mathrm{B}}_i(P, \{q\}) & \text{otherwise} \end{cases}$$

Observe that if (*) there are $i$ and $j$ such that, for every $Q \subseteq \mathcal{P}$, $d^{\mathrm{B}}_i(\rho(s), Q) \leq d^{\mathrm{B}}_j(\rho(t), Q) + c(a)$, it follows for every offset $k$ that $d^{\mathrm{B}}_{i+k}(\rho(s), Q) \leq d^{\mathrm{B}}_{j+k}(\rho(t), Q) + c(a)$. In particular, by choosing $k$ large enough to hit the fix-point, this shows that $d^{\mathrm{B}}_{\mathcal{H}}(\rho(s), Q) \leq d^{\mathrm{B}}_{\mathcal{H}}(\rho(t), Q) + c(a)$. This observation can be proven by induction on $k$. The induction hypothesis, $k = 0$, holds by assumption. For the induction step $k+1$, let $q \in \mathcal{P}$ be arbitrary. For $q \in \rho(s)$, the inequality holds trivially. If $q \in \rho(t)$, then, due to Definition 3, $\mathcal{H}$ contains an hyperedge $\langle T, a, \{q\}\rangle$ with $T \subseteq \rho(s)$, i. e., $d^{\mathrm{B}}_{i+k+1}(\rho(s), \{q\}) \leq c(a) + d^{\mathrm{B}}_{i+k}(\rho(s), T) = c(a)$. Finally, consider $q \in \mathcal{P}$ such that $q \notin \rho(s)$ and $q \notin \rho(t)$. Therefore,

$$\begin{aligned} d^{\mathrm{B}}_{i+k+1}(\rho(s), \{q\}) &= \min_{\langle T, a', \{q\}\rangle \in \mathcal{E}} c(a') + d^{\mathrm{B}}_{i+k}(\rho(s), T) \\ &\overset{\text{(IH)}}{\leq} c(a) + \min_{\langle T, a', \{q\}\rangle \in \mathcal{E}} c(a') + d^{\mathrm{B}}_{j+k}(\rho(t), T) \\ &= c(a) + d^{\mathrm{B}}_{j+k+1}(\rho(t), \{q\}) \end{aligned}$$

Finally, note that $i = 1$ and $j = 0$ satisfy the condition (*): It holds that $d^{\mathrm{B}}_0(\rho(t), \{q\}) = 0$ iff $q \in \rho(t)$, and $d^{\mathrm{B}}_0(\rho(t), \{q\}) = \infty$ otherwise. From Definition 3, it follows that, for every $q_t \in \rho(t)$, there is a hyperedge $\langle T, a, \{q_t\}\rangle$ such that $T \subseteq \rho(s)$, i. e., $d^{\mathrm{B}}_1(\rho(s), \{q_t\}) \leq c(a) + d^{\mathrm{B}}_0(\rho(s), T) = c(a)$.

**F-interpretations** Let $\rho$ be any hyperabstraction, and $\mathcal{H}$ be any F-interpretation of $\rho$. Let $P_{\mathcal{G}}$ denote the set of abstract concepts $\bigcup_{s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}} \rho(s_{\mathcal{G}})$. Obviously, for every goal state $s_{\mathcal{G}} \in \mathcal{S}_{\mathcal{G}}$, it holds that $\rho(s_{\mathcal{G}}) \subseteq P_{\mathcal{G}}$, i. e., $h^{\mathrm{F}}_\rho[\mathcal{H}](s_{\mathcal{G}}) = d^{\mathrm{F}}_{\mathcal{H}}(\rho(s_{\mathcal{G}}), P_{\mathcal{G}}) = 0$. Hence, $h^{\mathrm{F}}_\rho[\mathcal{H}]$ is goal-aware.

To show consistency, consider again any transition $\langle s, a, t\rangle \in \mathcal{T}$. If $\rho(s) = \emptyset$, then $h^{\mathrm{F}}_\rho[\mathcal{H}](s) = 0$, and the claim follows directly. Otherwise, let $p_s \in \rho(s)$ be such that $h^{\mathrm{F}}_\rho[\mathcal{H}](s) = d^{\mathrm{F}}_{\mathcal{H}}(\{p_s\}, P_{\mathcal{G}})$. If $p_s \in P_{\mathcal{G}}$, then $h^{\mathrm{F}}_\rho[\mathcal{H}](s) = 0$, and the claim again follows directly. Finally, assume that $p_s \notin P_{\mathcal{G}}$. Definition 2 guarantees the existence of a hyperedge $\langle\{p_s\}, a, Q_t\rangle \in \mathcal{E}$ such that $Q_t \subseteq \rho(t)$. We obtain:

$$\begin{aligned} h^{\mathrm{F}}_\rho[\mathcal{H}](s) &= d^{\mathrm{F}}_{\mathcal{H}}(\{p_s\}, P_{\mathcal{G}}) \\ &\leq d^{\mathrm{F}}_{\mathcal{H}}(Q_t, P_{\mathcal{G}}) + c(a) \\ &\leq d^{\mathrm{F}}_{\mathcal{H}}(\rho(t), P_{\mathcal{G}}) + c(a) \\ &= h^{\mathrm{F}}_\rho[\mathcal{H}](t) + c(a) \end{aligned}$$

This concludes the proof.

□

**Theorem 4** *There exist* $\Pi$, $\mathcal{C}$, *and* B*-interpretations* $\mathcal{H}$ *of* $\rho^{\mathcal{C}}$ *where* $|\mathcal{H}|$ *is polynomially bounded in* $|\mathcal{C}|$ *but* $h^{\mathrm{B}}_{\rho^{\mathcal{C}}}[\mathcal{H}] > h^{\mathcal{C}}$.

*Proof.* Consider the planning task $\Pi$ with two Boolean variables $\mathcal{V} = \{u, v\}$, initial state $s_{\mathcal{I}} = \{u = 0, v = 0\}$, goal $\mathcal{G} = \{u = 1, v = 1\}$, and one action $\mathcal{A} = \{a\}$ with precondition $pre_a = \{u = 1\}$, effect $eff_a = \{v = 1\}$, and cost of 1. Clearly, this task is unsolvable. Consider the conjunctions $\mathcal{C} = \{C_1, C_2\}$ where $C_1 = \{u = 1, v = 0\}$ and $C_2 = \{u = 1, v = 1\}$.

Observe that $h^{\mathcal{C}}(s_{\mathcal{I}}) < \infty$, i. e., $h^{\mathcal{C}}$ does not recognize $s_{\mathcal{I}}$ as dead-end (for this particular set $\mathcal{C}$):

$$\begin{aligned} h^{\mathcal{C}}(s_{\mathcal{I}}) &= \max_{C \in \mathcal{C}: C \subseteq \mathcal{G}} h^{\mathcal{C}}(s_{\mathcal{I}}, C) \\ &= h^{\mathcal{C}}(s_{\mathcal{I}}, C_2) \\ &= c(a) + h^{\mathcal{C}}(s_{\mathcal{I}}, pre_a) \\ &= c(a) + h^{\mathcal{C}}(s_{\mathcal{I}}, \{u = 1\}) \\ &= c(a) + \max \emptyset = c(a) = 1 \\ &< \infty \end{aligned}$$

On the other hand, however, we can construct a B-interpretation $\mathcal{H}$ of the hyperabstraction $\rho^{\mathcal{C}}$ corresponding to $\mathcal{C}$ so that $h^{\mathrm{B}}_{\rho}[\mathcal{H}](s_{\mathcal{I}}) = \infty$: Let $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w \rangle$ with $\mathcal{N} = \mathcal{C}$, $\mathcal{L} = \mathcal{A}$, $w = c$, and two hyperedges $\mathcal{E} = \{e_1, e_2\}$. Both hyperedges have the same head and label, i. e., $H_{e_1} = H_{e_2} = \{C_2\}$ and $l_{e_1} = l_{e_2} = a$. The tails are given by $T_{e_1} = \{C_1\}$ and $T_{e_2} = \{C_2\}$. Intuitively, we split the application of $a$ into two contexts: one where $v$ is already 1, and one where it is 0. When assuming $v = 1$, then $a$ becomes a noop action, since the conjunction $C_2$ must have been true before the application, and remains true afterwards. On the other hand, $v = 0$ implies that $C_1$ must be true before the application of $a$. Since $C_1$ cannot be satisfied after the application of $a$, and since $e_1$ and $e_2$ are exhaustive in the sense that every transition $\langle s, a, t \rangle$ where $C_2 \in \rho(t)$ is represented by either of them, $\mathcal{H}$ hence satisfies Definition 3. Moreover, neither $C_1$ nor $C_2$ is true in initial state. We obtain $h^{\mathrm{B}}_{\rho}[\mathcal{H}](s_{\mathcal{I}}) = d^{\mathrm{B}}_{\mathcal{H}}(\emptyset, \{C_2\}) = \infty$ as desired.

$\Pi$ can be easily extended to handle also the case where $\mathcal{C}$ additionally contains all singleton conjunctions, i. e., all facts, as it is usually the case when using $h^{\mathcal{C}}$.

□

**Theorem 5** *There exist families of tasks* $\Pi_n$ *and polynomially size-bounded forward and backward hyperabstraction heuristics such that* $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}] = h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}] = h^*$ *and both* $\mathcal{H}^{\mathrm{F}}$ *and* $\mathcal{H}^{\mathrm{B}}$ *are* BF*-hypergraphs, but unless **P** = **NP**, it is not possible to construct any M&S abstraction* $\alpha$ *such that* $h^{\alpha} = h^*$ *in polynomial time.*

*Proof.* Consider the family of planning tasks containing a task for every CNF formula. Let $\phi$ be any CNF formula with variables $X_1, \dots, X_n$ and clauses $C_1, \dots, C_m$. The corresponding FDR planning task $\Pi_{\phi} = \langle \mathcal{V}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{G} \rangle$ has Boolean variables $\mathcal{V} = \{x_1, \dots, x_n, c_1, \dots, c_m\}$, goal $\mathcal{G} = \{c_1 = 1, \dots, c_m = 1\}$, and the following actions

- For every clause $C_i$, $i \in [1, m]$, and for every literal $L \in C_i$: $a_{i,L} \in \mathcal{A}$ with

$$pre_{a_{i,L}} = \begin{cases} \{x_j = 0\} & \text{if } L = \neg X_j \\ \{x_j = 1\} & \text{if } L = X_j \end{cases}$$

  $eff_{a_{i,L}} = \{c_i = 1\}$, and cost of 0.

- $b \in \mathcal{A}$ with $pre_b = \emptyset$ and $eff_b = \{c_1 = 1, \dots, c_m = 1\}$ and cost of 1.

The initial state is not important for this proof. For the sake of simplicity, for any literal $L$ and state $s$, we write $s \models L$ if $s(x_j) = 0$ and $L = \neg X_j$, or $s(x_j) = 1$ and $L = X_j$. Otherwise $s \not\models L$. Similarly, for every state $s$ and clause $C_i$, we write $s \models C_i$ if there exists a literal $L \in C_i$ such that $s \models L$, and $s \not\models C_i$ otherwise. In other words, $\Pi_{\phi}$ is designed so that $h^*(s) = 1$ iff there exists a clause $C_i$ such that $s(c_i) = 0$ and $s \not\models C_i$. If no such clause exists, it holds that $h^*(s) = 0$.

The remainder of this proof is structured as follows: We will first show the existence of a B-hyperabstraction heuristic $h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}]$ such that $h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}] = h^*$, $\mathcal{H}^{\mathrm{B}}$ is a BF-hypergraph, and the construction and computation of $h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}]$ is polynomial in $|\phi|$. Afterwards, we will show that the same holds for some F-hyperabstraction heuristic $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}]$. We will conclude the proof, showing that unless **P=NP**, it is not possible to construct any M&S abstraction $\alpha$ in polynomial time such that $h^{\alpha} = h^*$.

**B-Hyperabstraction** We consider $h^1$ interpreted as B-hyperabstraction heuristic $h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}]$. For that, $\rho^{\mathrm{B}}$ uses $\Pi$'s facts as abstract concepts, and maps every state to the set of facts true in it, i. e., $\rho^{\mathrm{B}}(s) = s$. $\mathcal{H}^{\mathrm{B}}$'s hyperedges are defined as follows: for every action $a_{i,L} \in \mathcal{A}$, $\mathcal{H}^{\mathrm{B}}$ contains the hyperedge $\langle pre_{a_{i,L}}, a_{i,L}, \{c_i = 1\} \rangle$, and for every $i \in [1, m]$, $\mathcal{H}^{\mathrm{B}}$ contains the hyperedge $\langle \emptyset, b, \{c_i = 1\} \rangle$. Moreover, to completely satisfy Definition 3, $\mathcal{H}^{\mathrm{B}}$ additionally contains the following self-loop edges: $\langle \{x_i = d\}, a, \{x_i = d\} \rangle$ for every $d \in \{0, 1\}$ and $a \in \mathcal{A}$; and $\langle \{c_i = d\}, a, \{c_i = d\} \rangle$ for all $d \in \{0, 1\}$ and $a \in \mathcal{A}$ such that $a \neq a_{i,L}$ for every $L \in C_i$. Since the heads and tails of all hyperedges contain at most one abstract concept at a time, $\mathcal{H}^{\mathrm{B}}$ is indeed a BF-hypergraph. It is straightforward to show that $\mathcal{H}^{\mathrm{B}}$ is a B-interpretation of $\rho^{\mathrm{B}}$. $h^{\mathrm{B}}_{\rho^{\mathrm{B}}}[\mathcal{H}^{\mathrm{B}}] = h^1$ holds by construction.

We finally show that $h^1 = h^*$. Let $s$ be any state. If $h^*(s) = 0$, the admissibility of $h^1$ directly implies that $h^1(s) = 0$. Assume that $h^*(s) = 1$. Hence, there exists some clause $C_i$ such that $s(c_i) = 0$ and $s \not\models C_i$. In other words, $1 = h^*(s) \geq h^1(s) \geq h^1(s, \{c_i = 1\})$. Hence, $h^1(s) = 0$ only if $h^1(s, pre_{a_{i,L}}) = 0$ for some action setting $c_i$ to 1. This can obviously be only the case if $s \models L$ for some $L \in C_i$, i. e., if $s \models C_i$. Since we assumed that $s \not\models C_i$, we therefore get $h^1(s, \{c_i = 1\}) = c(b) + h^1(s, \emptyset) = 1$.

**F-Hyperabstraction** Let $\mathcal{P} = \{\neg C_1, \dots, \neg C_m, \mathcal{G}\}$, and consider the following hyperabstraction function:

$$\rho^{\mathrm{F}}(s) := \{\neg C_i \mid i \in [1,m], s(c_i) = 0, s \not\models C_i\} \cup \begin{cases} \{\mathcal{G}\} & \text{if } \forall i \in [1,m] : s(c_i) = 1 \\ \emptyset & \text{otherwise} \end{cases}$$

Consider the BF-hypergraph $\mathcal{H}^{\mathrm{F}} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w \rangle$ with nodes $\mathcal{N} = \mathcal{P}$, labels $\mathcal{L} = \mathcal{A}$, weights $w = c$, and the following BF-hyperedges: for every $i \in [1,m]$, $\langle \{\neg C_i\}, b, \{\mathcal{G}\} \rangle \in \mathcal{E}$; for every pair of clauses, $i, j \in [1,m]$ such that $i \neq j$, and for every literal $L \in C_i$, $\mathcal{H}^{\mathrm{F}}$ contains the self-loop $\langle \{\neg C_j\}, a_{i,L}, \{\neg C_j\} \rangle$; and for every action $a \in \mathcal{A}$, $\langle \{\mathcal{G}\}, a, \{\mathcal{G}\} \rangle \in \mathcal{E}$. Clearly, for any given state $s$, $\rho^{\mathrm{F}}(s)$ can be computed in polynomial time. Moreover, $\mathcal{H}^{\mathrm{F}}$ can be constructed in polynomial time, the number of nodes in $\mathcal{H}^{\mathrm{F}}$ is $m + 1$, the number of hyperedges is bounded by $m^2 \cdot n$. We next show that (1) $\mathcal{H}^{\mathrm{F}}$ is a F-interpretation of $\rho^{\mathrm{F}}$, and (2) that $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}] = h^*$.

We start with (1). $\mathcal{G}$ is invariant under the application of every action. $\mathcal{H}^{\mathrm{F}}$ accounts for that via separate self-loop edges for every action. Let $\neg C_i \in \mathcal{P}$ be arbitrary, and consider any state $s$ with $\neg C_i \in \rho^{\mathrm{F}}(s)$, i. e., $s(c_i) = 0$ and $s \not\models C_i$. Clearly, none of the actions $a_{i,L}$ is applicable in $s$, and thus there is no transition from $s$ for any of them. It remain the case of the $a_{j,L}$ actions, for $j \neq i$, and the $b$ action. The former actions do not affect $\neg C_i$, i. e., in every transition $\langle s, a_{j,L}, t \rangle$, it holds that $\neg C_i \in \rho^{\mathrm{F}}(t)$. $\mathcal{H}^{\mathrm{F}}$ contains a self-loop edge for every such action. On the other hand, applying $b$ necessarily makes $\mathcal{G}$ true, i. e., for every transition $\langle s, b, t \rangle$, it holds that $\mathcal{G} \in \rho^{\mathrm{F}}(t)$. The hyperedge $\langle \{\neg C_i\}, b, \{\mathcal{G}\} \rangle$ represents all those transitions. In conclusion, $\mathcal{H}^{\mathrm{F}}$ indeed satisfies Definition 2.

Finally, note that $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}] = h^*$. Let $s$ be any state. If $h^*(s) = 0$, then due to the admissibility of $h^{\mathrm{F}}_{\rho}$, $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}](s) = 0$ holds as well. Assume that $h^*(s) = 1$, and let $C_i$ be such that $s(c_i) = 0$ and $s \not\models C_i$. Hence, $\neg C_i \in \rho^{\mathrm{F}}(s)$, and $h^{\mathrm{F}}_{\rho^{\mathrm{F}}}[\mathcal{H}^{\mathrm{F}}](s) = d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\rho^{\mathrm{F}}(s), \{\mathcal{G}\}) \geq d^{\mathrm{F}}_{\mathcal{H}^{\mathrm{F}}}(\{\neg C_i\}, \{\mathcal{G}\}) = 1$.

**Merge-and-Shrink Abstraction** Now, assume that we could construct a merge-and-shrink abstraction $\alpha$ in polynomial time (and thus also space) such that $h^\alpha(s) = h^*$. Consider the set of states $\hat{S}$ where $s \in \hat{S}$ iff $s(c_i) = 0$ for all $i \in [1,m]$. Deciding whether $\phi$ is satisfiable or not can be done by iterating once over all abstract states $s^\alpha$ with abstract goal distance of 0, and checking for each of them whether the intersection $[s^\alpha] \cap \hat{S}$ is empty. If we have found an abstract state $s^\alpha$ with goal distance 0 that represents a state $s \in \hat{S}$, i. e., $h^*(s) = h^\alpha(s) = 0$, then $s \models C_i$ for all $i \in [1,m]$, and therefore $\phi$ is satisfiable. On the other hand, $\phi$ is unsatisfiable if every $s \in \hat{S}$ has $h^*(s) = h^\alpha(s) = 1$. We show below that $[s^\alpha] \cap \hat{S} \neq \emptyset$ can be tested in polynomial time in the size of $\phi$. Since the number of abstract states is polynomial in the size of $\Pi$ (and thus $\phi$), the overall algorithm runs hence runs in polynomial time in the size $\phi$. Unless **P=NP**, we can therefore conclude hat it is not possible to compute any M&S abstraction $\alpha$ in polynomial time such that $h^\alpha = h^*$.

To decide for an abstract state $s^\alpha$ whether $[s^\alpha] \cap \hat{S} \neq \emptyset$, we make a minor modification to the M&S procedure. Namely, for every abstraction $\alpha'$ constructed during M&S, we maintain Boolean flags $f_{s^{\alpha'}}$ indicating whether this is satisfied or not for abstract state $s^{\alpha'}$. This requires additional space which is polynomial in the size of the abstraction. The value of $f_{s^{\alpha'}}$ is determined along the execution of merge-and-shrink procedure:

- For every atomic abstraction: if the corresponding variable is $v = x_i$ for some $i \in [1,n]$, then $f_{x_i=0} = f_{x_i=1} = 1$; if $v = c_i$ for some $i \in [1,m]$, then $f_{c_i=0} = 1$ and $f_{c_i} = 0$.
- Merge of two abstractions $\alpha_1$ and $\alpha_2$: every abstract state $s^{\alpha_1 \otimes \alpha_2}$ of the merge corresponds to a pair of abstract states $s^{\alpha_1} \in \mathcal{S}^{\alpha_1}$ and $s^{\alpha_2} \in \mathcal{S}^{\alpha_2}$. For the abstract states of $\alpha_1$ and $\alpha_2$, the flags have already been computed. Thus, the value of $f_{s^{\alpha_1 \otimes \alpha_2}}$ can be simply determined from $f_{s^{\alpha_1}} \wedge f_{s^{\alpha_2}}$.
- Shrinking an abstraction $\alpha'$: merging $k$ abstract states $s_1^{\alpha'}, \dots, s_k^{\alpha'}$ into a single abstract state $s^{\alpha'}$. The values $f_{s_1^{\alpha'}}, \dots, f_{s_k^{\alpha'}}$ have be computed before. The value of $f_{s^{\alpha'}}$ can be computed by $f_{s_1^{\alpha'}} \vee \dots \vee f_{s_k^{\alpha'}}$.

Obviously the computation of the flags only adds a polynomial overhead to the merge-and-shrink construction. Moreover, it is straightforward to proof that $f_{s^\alpha} = 1$ iff $[s^\alpha] \cap \hat{S} \neq \emptyset$.

□

**Theorem 7** *Optimal cost partitioning for $h^1$ is **NP**-hard.*

*Proof.* The proof is by reduction from SAT. We next describe a family of tasks containing one planning task for every CNF formula. Let $\phi$ be any CNF formula with non-empty clauses $C_1, \dots, C_m$ and variables $X_1, \dots, X_n$. The corresponding task $\Pi_\phi = \langle \mathcal{V}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{G} \rangle$ has Boolean variables $\mathcal{V} = \{x_1^f, \dots, x_n^f, x_1^t, \dots, x_n^t, g^t, g^f\}$; initial state $s_{\mathcal{I}}(v) = 0$ for all $v \in \mathcal{V}$; goal $\mathcal{G} = \{g^t = 1, g^f = 1\}$; and the following $3n + 2m$ actions

$$\mathcal{A} = \{\text{setF(i)}, \text{setT(i)}, \text{achieve}\mathcal{G}\text{(i)} \mid i \in [1,n]\} \cup \{\text{evalF(j)}, \text{evalT(j)} \mid j \in [1,m]\}$$

where

| Action | Precondition | Effect | Cost |
|---|---|---|---|
| setF(i) | $\{x_i^f = 0\}$ | $\{x_i^f = 1\}$ | 1 |
| setT(i) | $\{x_i^t = 0\}$ | $\{x_i^t = 1\}$ | 1 |
| achieve$\mathcal{G}$(i) | $\{x_i^t = 1, x_i^f = 1\}$ | $\{g^t = 1, g^f = 1\}$ | 0 |
| evalF(j) | $\{x_i^f = 1 \mid X_i \in C_j\}$ $\cup \{x_i^t = 1 \mid \neg X_i \in C_j\}$ | $\{g^f = 1\}$ | 0 |
| evalT(j) | $\{x_i^t = 1 \mid X_i \in C_j\}$ $\cup \{x_i^f = 1 \mid \neg X_i \in C_j\}$ | $\{g^t = 1\}$ | 0 |

Let $\mathbf{c}^*$ be any cost partitioning for which $h^1[\![\mathbf{c}^*]\!](s_{\mathcal{I}})$ is maximal. We next show that $h^1[\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$ if $\phi$ is satisfiable, and $h^1[\![\mathbf{c}^*]\!](s_{\mathcal{I}}) < 2$ otherwise. For every CNF formula

$\phi, \Pi_\phi$ can be constructed in polynomial time and space in the size of $\phi$. Hence, if it was possible to compute $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I})$ for all tasks $\Pi_\phi$, in time polynomial in $|\Pi_\phi|$, then we could decide SAT in polynomial time.

$\phi$ **is satisfiable:** We show that $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) = 2$.

Let $V : \{X_1, \dots, X_n\} \mapsto \{0, 1\}$ be a satisfying assignment to $\phi$. Consider the cost partitioning $\langle c_1, c_2\rangle$ with $c_1(a) = 0$ and $c_2(a) = 0$ for all actions $a$, but:

- For every $i \in [1, n]$: $c_1(\text{setT(i)}) = 1$ if $V(X_i) = 1$ and $c_1(\text{setF(i)}) = 1$ if $V(X_i) = 0$.
- For every $i \in [1, n]$: $c_2(\text{setT(i)}) = 1$ if $V(X_i) = 0$ and $c_2(\text{setF(i)}) = 1$ if $V(X_i) = 1$.

In other words, $c_2 = c - c_1$, i.e., $\langle c_1, c_2\rangle$ is indeed a cost partitioning of $c$. We next show that (1) $h^*(s_\mathcal{I}) \le 2$, (2) $h^1[\![c_1]\!](s_\mathcal{I}) \ge 1$, and (3) $h^1[\![c_2]\!](s_\mathcal{I}) \ge 1$. Given those observations, we obtain

$$\begin{aligned} 2 &\overset{(1)}{\ge} h^*(s_\mathcal{I}) \\ &\ge h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) \\ &\ge h^1[\![\langle c_1, c_2\rangle]\!](s_\mathcal{I}) \\ &\overset{(2)+(3)}{\ge} 1 + 1 = 2 \end{aligned}$$

and therefore $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) = 2$ as desired.

Regarding (1), the action sequence $\langle \text{setF}(1), \text{setT}(1), \text{achieve}\mathcal{G}(1)\rangle$ is a valid plan for $s_\mathcal{I}$. It has cost 2 and hence $h^*(s_\mathcal{I}) \le 2$.

Regarding (2), $c_1$ assigns every action to either 0 or 1, i.e., $h^1[\![c_1]\!](s_\mathcal{I})$ can only take an integer value. Since $h^1$ is admissible, and due to (1), it thus follows that $h^1[\![c_1]\!](s_\mathcal{I}) \in \{0, 1, 2\}$. Assume for contradiction that $h^1[\![c_1]\!](s_\mathcal{I}) = 0$. This particularly implies that $h^1[\![c_1]\!](s_\mathcal{I}, \{g^t{=}1\}) = 0$. $g^t{=}1$ is not true in $s_\mathcal{I}$, so must be achieved via one of achieve$\mathcal{G}$(i) or evalT(j). For every $i \in [1, n]$, it holds that $c_1(\text{setT(i)}) = 1$ or $c_1(\text{setF(i)}) = 1$. It follows $h^1[\![c_1]\!](s_\mathcal{I}, pre_{\text{achieve}\mathcal{G}\text{(i)}}) = 1$. To obtain $h^1[\![c_1]\!](s_\mathcal{I}, \{g^t{=}1\}) = 0$, there must hence be some $j \in [1, m]$ such that $h^1[\![c_1]\!](s_\mathcal{I}, pre_{\text{evalT(j)}}) = 0$. But this can be only the case if, for every $X_i \in C_j$, it holds that $c_1(\text{setT(i)}) = 0$ and for every $\neg X_i \in C_j$ that $c_1(\text{setF(i)}) = 0$. Plugging in the definition of $c_1$, it therefore holds for every $X_i \in C_j$ that $V(X_i) = 0$, and for every $\neg X_i \in C_j$ that $V(X_i) = 1$. But this means that $V$ does actually not satisfy $\phi$, a contradiction to the assumption. In conclusion, we obtain $h^1[\![c_1]\!](s_\mathcal{I}) \ge 1$.

The proof of (3) is symmetric. Similar to $c_1$, it holds that $h^1[\![c_2]\!](s_\mathcal{I}) \in \{0, 1, 2\}$. Assume again for contradiction that $h^1[\![c_2]\!](s_\mathcal{I}) = 0$. This in particular implies that $h^1[\![c_2]\!](s_\mathcal{I}, \{g^f{=}1\}) = 0$. By construction, for every $i \in [1, n]$, it holds that either $c_2(\text{setT(i)}) = 1$ or $c_2(\text{setF(i)}) = 1$. To obtain $h^1[\![c_2]\!](s_\mathcal{I}, \{g^f{=}1\}) = 0$ there must thus exist some $j \in [1, m]$ such that $h^1[\![c_2]\!](s_\mathcal{I}, pre_{\text{evalF(j)}}) = 0$. Let $j$ be any such index. Then, for every $X_i \in C_j$: $c_2(\text{setF(i)}) = 0$, and for every $\neg X_i \in C_j$: $c_2(\text{setT(i)}) = 0$. It follows from that definition of $c_2$ that, for every $X_i \in C_j$: $V(X_i) = 0$, and for every $\neg X_i \in C_j$: $V(X_i) = 1$. This is again a contradiction to the assumption that $V$ is satisfying assignment to $\phi$. This completes the proof for the satisfiable case.

$\phi$ **is unsatisfiable:** We show that $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) < 2$.

Consider any cost function $c'$ such that $c'(a) \le c(a)$ for all actions $a \in \mathcal{A}$. Denote by $\text{h}_{c'} := h^1[\![c']\!](s_\mathcal{I})$. Below, we prove the following two statements:

(1) For every $i \in [1, n]$, it holds that $\max\{c'(\text{setT(i)}), c'(\text{setF(i)})\} \ge \text{h}_{c'}$. Since all action costs are non-negative, this particularly shows that $c'(\text{setT(i)}) + c'(\text{setF(i)}) \ge \text{h}_{c'}$.

(2) There is at least one $i \in [1, n]$ such that $c'(\text{setT(i)}) \ge \text{h}_{c'}$ and $c'(\text{setF(i)}) \ge \text{h}_{c'}$.

Consider any optimal cost partitioning $\mathbf{c}^* = \langle c_1, \dots, c_K\rangle$ for $h^1$ and $s_\mathcal{I}$. If $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) = 0$, then the claim would follow immediately. For the rest of the proof, we assume that $h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) > 0$. Thus, $\mathbf{c}^*$ must contain some cost function $c'$ such that $\text{h}_{c'} > 0$. Without loss of generality, we assume that $c_1$ is such a cost function. Let $i \in [1, n]$ be any index that satisfies property (2) for $c_1$. By using the observations from above, we obtain:

$$\begin{aligned} h^1[\![\mathbf{c}^*]\!](s_\mathcal{I}) &\overset{\text{Def } \mathbf{c}^*}{=} \text{h}_{c_1} + \sum_{k=2}^{K} \text{h}_{c_k} \\ &\overset{(1)}{\le} \text{h}_{c_1} + \sum_{k=2}^{K} \big(c_k(\text{setT(i)}) + c_k(\text{setF(i)})\big) \\ &= \text{h}_{c_1} + \sum_{k=2}^{K} c_k(\text{setT(i)}) + \sum_{k=2}^{K} c_k(\text{setF(i)}) \\ &\overset{\text{Def } \mathbf{c}^*}{\le} \text{h}_{c_1} + \big(c(\text{setT(i)}) - c_1(\text{setT(i)})\big) \\ &\qquad + \big(c(\text{setF(i)}) - c_1(\text{setF(i)})\big) \\ &\overset{(2)}{\le} \text{h}_{c_1} + \big(c(\text{setT(i)}) - \text{h}_{c_1}\big) + \big(c(\text{setF(i)}) - \text{h}_{c_1}\big) \\ &= \text{h}_{c_1} + (1 - \text{h}_{c_1}) + (1 - \text{h}_{c_1}) \\ &= 2 - \text{h}_{c_1} \\ &\overset{\text{h}_{c_1} > 0}{<} 2 \end{aligned}$$

which shows the claim. It remains to show that (1) and (2) are actually true. Let $c'$ be any cost function such that $c'(a) \le c(a)$ for all actions $a$.

Regarding (1), assume for contradiction there was some $i \in [1, n]$ such that $c'(\text{setT(i)}) < \text{h}_{c'}$ and $c'(\text{setF(i)}) < \text{h}_{c'}$. Then, $h^1[\![c']\!](s_\mathcal{I}) \le h^1[\![c']\!](s_\mathcal{I}, pre_{\text{achieve}\mathcal{G}\text{(i)}}) = \max\{c'(\text{setT(i)}), c'(\text{setF(i)})\} < \text{h}_{c'}$, which obviously contradicts the definition of $\text{h}_{c'}$. Hence, (1) must be satisfied.

Assume for contradiction that (2) is not satisfied, i.e., it holds for every $i \in [1, n]$ that $c'(\text{setF(i)}) < \text{h}_{c'}$ or $c'(\text{setT(i)}) < \text{h}_{c'}$. We next show that this allows to construct a satisfying assignment $V$ to $\phi$, contradicting the assumption that $\phi$ is unsatisfiable. By definition, $h^1[\![c']\!](s_\mathcal{I})$ is the maximum of $h^1[\![c']\!](s_\mathcal{I}, \{g^t{=}1\})$ and $h^1[\![c']\!](s_\mathcal{I}, \{g^f{=}1\})$. Distinguish between the two cases:

- Case $h^1[\![c']\!](s_{\mathcal{I}}, \{g^t{=}1\}) = \mathrm{h}_{c'}$. Consider the following variable assignment $V : \{X_1, \dots, X_n\} \mapsto \{0, 1\}$:

$$V(X_i) = \begin{cases} 1 & \text{if } c'(\text{setF(i)}) < \mathrm{h}_{c'} \\ 0 & \text{if } c'(\text{setT(i)}) < \mathrm{h}_{c'} \end{cases}$$

  Note that $V$ is uniquely defined: Given our assumption, at least one of the two cases must apply to every variable $X_i$. Due to (1), it is however not possible that $c'(\text{setT(i)}) < \mathrm{h}_{c'}$ and $c'(\text{setF(i)}) < \mathrm{h}_{c'}$.

  Since $h^1[\![c']\!](s_{\mathcal{I}}, \{g^t{=}1\}) = \mathrm{h}_{c'}$, it follows for every $j \in [1, m]$ that $h^1[\![c']\!](s_{\mathcal{I}}, pre_{\text{evalT(j)}}) \geq \mathrm{h}_{c'}$. All setF(i) and setT(i) actions are applicable initially, so the $h^1$ values of their preconditions are 0. Hence, $h^1[\![c']\!](s_{\mathcal{I}}, pre_{\text{evalT(j)}})$ is given by the maximum cost of all $c'(\text{setT(i)})$ for $X_i \in C_j$ and $c'(\text{setF(i)})$ for $\neg X_i \in C_j$. In other words, every clause $C_j \in \phi$ must either contain (a) $X_i \in C_j$ such that $c'(\text{setT(i)}) \geq \mathrm{h}_{c'}$, or (b) $\neg X_i \in C_j$ such that $c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$. For (a), it follows that $c'(\text{setF(i)}) < \mathrm{h}_{c'}$, i. e., $V(X_i) = 1$, and thus $V$ satisfies $C_j$. Vice versa, in (b), it follows that $c'(\text{setT(i)}) < \mathrm{h}_{c'}$, i. e., $V(X_i) = 0$, and thus $V$ also satisfies $C_j$. In conclusion, $V \models \phi$.

- The case $h^1[\![c']\!](s_{\mathcal{I}}, \{g^f{=}1\}) = \mathrm{h}_{c'}$ works symmetrically. Consider the following variable assignment:

$$V(X_i) = \begin{cases} 1 & \text{if } c'(\text{setT(i)}) < \mathrm{h}_{c'} \\ 0 & \text{if } c'(\text{setF(i)}) < \mathrm{h}_{c'} \end{cases}$$

  Again $c'$ uniquely and properly defines $V$.

  Since $h^1[\![c']\!](s_{\mathcal{I}}, \{g^f{=}1\}) = \mathrm{h}_{c'}$, it follows for every $j \in [1, m]$ that $h^1[\![c']\!](s_{\mathcal{I}}, pre_{\text{evalF(j)}}) \geq \mathrm{h}_{c'}$. Similarly as above, it follows that every clause $C_j \in \phi$ contains (a) $X_i \in C_j$ such that $c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$, or (b) $\neg X_i \in C_j$ such that $c'(\text{setT(i)}) \geq \mathrm{h}_{c'}$. For (a), it follows that $c'(\text{setT(i)}) < \mathrm{h}_{c'}$, i. e., $V(X_i) = 1$, and thus $V$ satisfies $C_j$. Vice versa, in (b), it follows that $c'(\text{setF(i)}) < \mathrm{h}_{c'}$, i. e., $V(X_i) = 0$, and thus $V$ also satisfies $C_j$. In summary, $V \models \phi$.

  This concludes the proof of (2).

□

**Corollary 2** *Optimal cost partitioning for hyperabstraction heuristics in general is **NP**-hard.*

*Proof.* Critical-path heuristics being a special case of B-hyperabstraction heuristics, Theorem 7 leads directly to the desired result in the B case. For the remainder of this proof, we thus focus on F-hyperabstraction heuristics. The proof is symmetric to the proof of Theorem 7. We spell out the details for the sake of completeness.

We next describe a family of planning tasks, containing one task for every CNF formula, that will be used for the reduction of SAT to $h^{\mathrm{F}}_{\rho}$ optimal cost partitioning. Let $\phi$ be any CNF formula with variables $X_1, \dots, X_n$ and clauses $C_1, \dots, C_m$. The corresponding task $\Pi_\phi = \langle \mathcal{V}, \mathcal{A}, s_{\mathcal{I}}, \mathcal{G} \rangle$ has Boolean variables $\mathcal{V} = \{x^f_1, \dots, x^f_n, x^t_1, \dots, x^t_n, g^t, g^f\}$; initial state $s_{\mathcal{I}}(v) = 1$ for all $v \in \mathcal{V}$ but $s_{\mathcal{I}}(g^t) = s_{\mathcal{I}}(g^f) = 0$; goal $\mathcal{G}(v) = 1$ for all variables $v \in \mathcal{V}$; and the following $3n + 2m$ actions:

$$\begin{aligned} \mathcal{A} = &\{\text{setF(i), setT(i), achieve}\mathcal{G}\text{(i)} \mid i \in [1, n]\} \\ &\cup \{\text{evalF(j), evalT(j)} \mid j \in [1, m]\} \end{aligned}$$

with slightly different semantics than the actions in the proof of Theorem 7:

| Action | Precondition | Effect | Cost |
|---|---|---|---|
| setF(i) | $\{x^f_i{=}0\}$ | $\{x^f_i{=}1\}$ | 1 |
| setT(i) | $\{x^t_i{=}0\}$ | $\{x^t_i{=}1\}$ | 1 |
| achieve$\mathcal{G}$(i) | $\{g^t{=}0, g^f{=}0\}$ | $\{g^t{=}1, g^f{=}0\}$ $\cup \{x^t_i = 0, x^f_i = 0\}$ | 0 |
| evalF(j) | $\{g^f = 0\}$ | $\{g^f = 1\}$ $\cup \{x^f_i{=}0 \mid X_i \in C_j\}$ $\cup \{x^t_i{=}0 \mid \neg X_i \in C_j\}$ | 0 |
| evalT(j) | $\{g^t = 0\}$ | $\{g^t = 1\}$ $\cup \{x^t_i{=}0 \mid X_i \in C_j\}$ $\cup \{x^f_i{=}0 \mid \neg X_i \in C_j\}$ | 0 |

We next give a (polynomially sized) hyperabstraction function $\rho$ and F-interpretation $\mathcal{H}$ of $\rho$ such that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$ iff $\phi$ is satisfiable, and $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) < 2$ otherwise. Since for every CNF formula $\phi$, $\Pi_\phi$ can be constructed in time polynomial in $|\phi|$, checking $\phi$'s satisfiability can hence be polynomially reduced to checking whether $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$.

A desired hyperabstraction function $\rho$ is given by $\rho(s) = s$, using $\Pi_\phi$'s facts as abstract concepts. The F-interpretation $\mathcal{H} = \langle \mathcal{N}, \mathcal{E}, \mathcal{L}, w \rangle$ of $\rho$ is defined as follows: $\mathcal{N}$ is the set of $\Pi_\phi$'s facts; $\mathcal{L} = \mathcal{A}$; and $w = c$. The hyperedges $\mathcal{E}$ are obtained by connecting for every action, every precondition fact with the effect:

- For every $i \in [1, n]$

($\mathcal{E}1$) $\langle \{x^f_i{=}0\}, \text{setF(i)}, \{x^f_i{=}1\} \rangle \in \mathcal{E}$,

($\mathcal{E}2$) $\langle \{x^t_i{=}0\}, \text{setT(i)}, \{x^t_i{=}1\} \rangle \in \mathcal{E}$,

($\mathcal{E}3$) $\langle \{g^t{=}0\}, \text{achieve}\mathcal{G}\text{(i)}, \{g^t{=}1, x^t_i{=}0, x^f_i{=}0\} \rangle \in \mathcal{E}$,

($\mathcal{E}4$) $\langle \{g^f{=}0\}, \text{achieve}\mathcal{G}\text{(i)}, \{g^f{=}1, x^t_i{=}0, x^f_i{=}0\} \rangle \in \mathcal{E}$,

- For every $i \in [1, m]$

($\mathcal{E}5$) $\langle \{g^f{=}0\}, \text{evalF(i)}, \{g^f{=}1\} \cup \{x^f_j{=}0 \mid X_j \in C_i\} \cup \{x^t_j{=}0 \mid \neg X_j \in C_i\} \rangle \in \mathcal{E}$,

($\mathcal{E}6$) $\langle \{g^t{=}0\}, \text{evalT(i)}, \{g^t{=}1\} \cup \{x^t_j{=}0 \mid X_j \in C_i\} \cup \{x^f_j{=}0 \mid \neg X_j \in C_i\} \rangle \in \mathcal{E}$,

To completely satisfy Definition 2, $\mathcal{E}$ also needs to contain some self-loop edges and edges for the goal facts. We omit their specification for the sake of brevity. Note however that both kinds of hyperedges do not affect the goal distance estimations (goal facts have by default a goal distance of 0, and similar to regular graphs, self-loops can be never part of a shortest path). Moreover, the number of self loops is bounded by the number of nodes, and for every goal fact and action, we must add at most one hyperedge to cover all corresponding transitions. Hence, the size of $\mathcal{H}$ is polynomial in $\Pi_\phi$, and therefore polynomial in the size of $\phi$. To simplify the write-up, in the following, we use $d^{\mathrm{F}}_{\mathcal{H}}(p)$ to denote the goal distance $d^{\mathrm{F}}_{\mathcal{H}}(\{p\}, \mathcal{G})$ for any fact $p$.

Let $\mathbf{c}^*$ be any optimal cost partitioning for $h^{\mathrm{F}}_{\rho}[\mathcal{H}]$ and $s_{\mathcal{I}}$. It is left to show that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$ if $\phi$ is satisfiable, and $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) < 2$ otherwise. We split the proof again in the two cases:

**$\phi$ is satisfiable** We need to show that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$.

Let $V : \{X_1, \ldots, X_n\} \mapsto \{0, 1\}$ be any satisfying assignment to $\phi$. We construct a cost partitioning $\langle c_1, c_2\rangle$ from $V$ as follows:

$$c_1(a) := \begin{cases} 1 & \text{if } V(X_i){=}1 \text{ and } a = \text{setT(i)} \\ 1 & \text{if } V(X_i){=}0 \text{ and } a = \text{setF(i)} \\ 0 & \text{otherwise} \end{cases}$$

and

$$c_2(a) := \begin{cases} 1 & \text{if } V(X_i){=}1 \text{ and } a = \text{setF(i)} \\ 1 & \text{if } V(X_i){=}0 \text{ and } a = \text{setT(i)} \\ 0 & \text{otherwise} \end{cases}$$

Obviously, $c_1 + c_2 = c$, i. e., $\langle c_1, c_2\rangle$ is indeed a cost partitioning of $c$. Note that (1) $h^*(s_{\mathcal{I}}) \leq 2$, (2) $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}}) \geq 1$, and (3) $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_2]\!](s_{\mathcal{I}}) \geq 1$. Hence

$$\begin{aligned} 2 &\overset{(1)}{\geq} h^*(s_{\mathcal{I}}) \\ &\geq h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) \\ &\geq h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}}) + h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_2]\!](s_{\mathcal{I}}) \\ &\overset{(2)+(3)}{\geq} 2 \end{aligned}$$

which shows that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 2$ as desired.

For (1), the action sequence $\langle \text{achieve}\mathcal{G}(1), \text{setF}(1), \text{setT}(1)\rangle$ is a plan for $s_{\mathcal{I}}$ and has cost 2.

Since both $c_1$ and $c_2$ only use integer values, $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}})$ and $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}})$ must be integer values as well. Assume for contradiction that (2) was not satisfied, i. e., $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}}) < 1$ and thus $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c_1]\!](s_{\mathcal{I}}) = 0$. This in particular implies that $d^{\mathrm{F}}_{\mathcal{H}}(g^t{=}0) = 0$. Due to the design of $c_1$, it holds for every $i \in [1, n]$ that $d^{\mathrm{F}}_{\mathcal{H}}(\{x^t_i{=}0, x^f_i{=}0\}) = \max\{c_1(\text{setT(i)}), c_1(\text{setF(i)})\} = 1$. In other words, the hyperedge of type $(\mathcal{E}3)$ induce a goal distance of 1. To obtain $d^{\mathrm{F}}_{\mathcal{H}}(g^t{=}0) = 0$, there must thus be some $j \in [1, m]$ such that $d^{\mathrm{F}}_{\mathcal{H}}(\mathit{eff}_{\text{evalT(j)}}) = 0$. It follows for every (a) $X_i \in C_j$ that $d^{\mathrm{F}}_{\mathcal{H}}(x^t_i{=}0) = 0$, and for every (b) $\neg X_i \in C_j$ that $d^{\mathrm{F}}_{\mathcal{H}}(x^f_i{=}0) = 0$. For (a), $(\mathcal{E}2)$ is the only goal-leading hyperedge for $x^t_i{=}0$. Therefore, it must hold that $c_1(\text{setT(i)}) = 0$, i. e., $V(X_i) = 0$. Similarly, for (b), $(\mathcal{E}1)$ is the only goal-leading hyperedge for $x^f_i{=}0$. This implies that $c_1(\text{setF(i)}) = 0$, and hence $V(X_i) = 1$. Both cases together show that $V \not\models C_j$, and therefore $V \not\models \phi$, which is a contradiction to the assumption. We conclude that (2) must hold. The proof of (3) is symmetric. This completes the part for satisfiable $\phi$.

**$\phi$ is unsatisfiable** We must show that $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) < 2$.

Let $c' \leq c$ be any cost function. We denote by $\mathrm{h}_{c'} := h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![c']\!](s_{\mathcal{I}})$. Below, we will show the following two statements:

(1) For every $i \in [1, n]$, it holds that $c'(\text{setT(i)}) \geq \mathrm{h}_{c'}$ or $c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$. By definition, $c'(a) \geq 0$ for every action $a \in \mathcal{A}$. Therefore, $c'(\text{setT(i)}) + c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$.

(2) There exists an index $i \in [1, n]$ such that $c'(\text{setT(i)}) \geq \mathrm{h}_{c'}$ and $c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$.

Let $\mathbf{c}^* = \langle c_1, \ldots, c_K\rangle$ be any optimal cost partitioning. If $h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) = 0$, the claim would follow directly. Otherwise, there must exist some $c' \in \mathbf{c}^*$ such that $\mathrm{h}_{c'} > 0$. Without loss of generality, assume that $c_1$ has $\mathrm{h}_{c_1} > 0$. Let $i \in [1, n]$ be the index for which (2) for $c_1$ is satisfied. We get:

$$\begin{aligned} h^{\mathrm{F}}_{\rho}[\mathcal{H}][\![\mathbf{c}^*]\!](s_{\mathcal{I}}) &\overset{\text{Def } \mathbf{c}^*}{=} \mathrm{h}_{c_1} + \sum_{k=2}^{K} \mathrm{h}_{c_k} \\ &\overset{(1)}{\leq} \mathrm{h}_{c_1} + \sum_{k=2}^{K} \big(c_k(\text{setT(i)}) + c_k(\text{setF(i)})\big) \\ &= \mathrm{h}_{c_1} + \sum_{k=2}^{K} c_k(\text{setT(i)}) + \sum_{k=2}^{K} c_k(\text{setF(i)}) \\ &\overset{\text{Def } \mathbf{c}^*}{\leq} \mathrm{h}_{c_1} + (1 - c_1(\text{setT(i)})) + (1 - c_1(\text{setF(i)})) \\ &\overset{(2)}{\leq} \mathrm{h}_{c_1} + (1 - \mathrm{h}_{c_1}) + (1 - \mathrm{h}_{c_1}) \\ &= 2 - \mathrm{h}_{c_1} \\ &\overset{\mathrm{h}_{c_1} > 0}{<} 2 \end{aligned}$$

We finally need to show that (1) and (2) are indeed true. Let $c' : \mathcal{A} \mapsto \mathbb{R}^+_0$ be any cost function such that $c'(a) \leq c(a)$ for every $a \in \mathcal{A}$. Regarding (1), let $i \in [1, n]$ be arbitrary. It holds that

$$\begin{aligned} \mathrm{h}_{c'} &\overset{(\mathcal{E}3)+(\mathcal{E}4)}{\leq} d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](\mathit{eff}_{\text{achieve}\mathcal{G}\text{(i)}}) \\ &= d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](\{x^t_i{=}0, x^f_i{=}0, g^t{=}1, g^f{=}1\}) \\ &= \max\{d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](x^t_i{=}0), d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](x^f_i{=}0), 0\} \\ &\overset{(\mathcal{E}1)+(\mathcal{E}2)}{=} \max\{c'(\text{setT(i)}), c'(\text{setF(i)})\} \end{aligned}$$

what shows the claim.

Assume for contradiction (2) was not true, i. e., assume for every $i \in [1, n]$ that $c'(\text{setT(i)}) < \mathrm{h}_{c'}$ or $c'(\text{setF(i)}) < \mathrm{h}_{c'}$. We show that $c'$ allows to construct a satisfying assignment $V$ to $\phi$. By definition of $h^{\mathrm{F}}_{\rho}$, $\mathrm{h}_{c'} = \max\{d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](g^f{=}0), d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](g^t{=}0)\}$. Distinguish between the two cases:

- $d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](g^f{=}0) = \mathrm{h}_{c'}$. We define $V$ as follows:

$$V(X_i) := \begin{cases} 1 & \text{if } c'(\text{setT(i)}) < \mathrm{h}_{c'} \\ 0 & \text{if } c'(\text{setF(i)}) < \mathrm{h}_{c'} \end{cases}$$

Note that for every $i \in [1, n]$, exactly one of the two cases must apply, i. e., $V$ is well-defined. Let $C_j \in \phi$ be an arbitrary clause. It holds that

$$\begin{aligned}\mathrm{h}_{c'} &= d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](g^f{=}0)\\ &\overset{(\mathcal{E}5)}{\leq} d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](\mathit{eff}_{\text{evalF(j)}})\\ &= \max \begin{cases} c'(\text{setF(i)}) & \text{for } X_i \in C_j \\ c'(\text{setT(i)}) & \text{for } \neg X_i \in C_j \end{cases}\end{aligned}$$

Therefore, there must exist (a) some $X_i \in C_j$ such that $c'(\text{setF(i)}) \geq \mathrm{h}_{c'}$, or (b) some $\neg X_i \in C_j$ such that $c'(\text{setT(i)}) \geq \mathrm{h}_{c'}$. In case of (a), it follows that $c'(\text{setT(i)}) < \mathrm{h}_{c'}$ from our assumption, and therefore $V(X_i) = 1$, i. e., $V \models C_j$. For (b), $c'(\text{setF(i)}) < \mathrm{h}_{c'}$, and hence $V(X_i) = 0$, i. e., $V \models C_j$. We conclude that $V$ satisfies all clauses $C_j$, so $V \models \phi$. This is a contradiction to the assumption that $\phi$ is unsatisfiable.

- $d^{\mathrm{F}}_{\mathcal{H}}[\![c']\!](g^t{=}0) = \mathrm{h}_{c'}$. The proof works symmetrically, using the following variable assignment:

$$V(X_i) := \begin{cases} 1 & \text{if } c'(\text{setF(i)}) < \mathrm{h}_{c'} \\ 0 & \text{if } c'(\text{setT(i)}) < \mathrm{h}_{c'} \end{cases}$$

Since both cases lead to a contradiction, (2) must hence be satisfied. This completes the proof.

□