PPCP: The Proofs

Maxim Likhachev
Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
maximl@seas.upenn.edu

Anthony Stentz
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
axs@rec.ri.cmu.edu

1 Notations and Assumptions

In this section we introduce some notations and formalize mathematically the class of problems our algorithm is suitable for. We assume that the environment is fully deterministic and can be modeled as a graph. That is, if we were to know the true value of each variable that represents the missing information about the environment then there would be no uncertainty in an outcome of any action. There are certain elements of the environment, however, whose status we are uncertain about and which affect the outcomes (and/or possible costs) of one or more actions. In the following we re-phrase this mathematically.

Let X be a full state-vector (a belief state). We assume it can be split into two sets of variables, $S(X)$, $H(X)$: $X = [S(X); H(X)]$. $S(X)$ is the finite set of variables whose values are always observed and the number of possible values is also finite. $H(X)$ is the set of (hidden) variables that initially represented the missing information about the environment. The variables in $H(X)$ are never moved to $S(X)$. X_{start} is used to denote the start state, all the values of the variables in $H(X_{\text{start}})$ are unknown. The goal of the planner is to construct a policy that reaches any state X such that $S(X) = S_{\text{goal}}$, where S_{goal} is given, while minimizing the expected cost of execution.

We assume perfect sensing. For the sake of easier notation let us introduce an additional value u for each variable $h_i \in H$. The setting $h_i(X) = u$ at state X will represent the fact that the value of h_i is unknown at X. If $h_i(X) \neq u$, then the true value of h_i is known at X since sensing is perfect. We restrict that all the variables that make up X can take only a finite number of distinct values.

We assume at most one hidden variable per action. Let $A(S(X))$ to denote the finite set of actions available at any state Y whose $S(Y) = S(X)$. Each action $a \in A(S(X))$ taken at state X may have one or more outcomes. If the execution of the action does not depend on any of the variables h_i whose values are not yet known, then there is only one outcome of a. Otherwise, there can be more than one outcome. We assume that each such action can not be controlled by more than one hidden variable. (The value of one hidden variable can affect more than one action though.) We use $h^{S(X),a}$ to represent the hidden variable that controls the outcomes and costs of action a taken at state X. By $h^{S(X),a} = \mathbf{null}$ we denote the case when

there was never any uncertainty about the outcome of action a taken at state X. The set of possible outcomes of action a taken $S(X)$ is notated by $succ(S(X), a)$, whereas $c(S(X), a, S(Y))$ such that $S(Y) \in succ(S(X), a)$ denotes the cost of the action and the outcome $S(Y)$. The costs are assumed to be bounded from below by a (small) positive constant. Sometimes, we will need to refer to the set of successors in the belief state-space. In these cases we will use the notation $succ(X, a)$ to denote the set of belief states Y such that $S(Y) \in succ(S(X), a)$ and $H(Y)$ is the same as $H(X)$ except for $h^{S(X),a}$ which also remains the same if it was known at X and is different otherwise. The function $P_{X,a}(succ(X, a))$, the probability distribution of outcomes of a executed at X, follows the probability distribution of $h^{S(X),a}$, $P(h^{S(X),a})$. Once action a was executed at state X the actual value of $h^{S(X),a}$ can be deduced since we assumed the sensing is perfect and the environment is deterministic.

We assume independence of the hidden variables. For the sake of efficient planning we assume that the variables in H can be considered independent of each other and therefore $P(H) = \prod_{i=1}^{|H|} P(h_i)$.

We assume clear preferences on the values of the hidden variables are available. We require that for each variable $h \in H$ we are given its preferred value, denoted by b (i.e., best). This value must satisfy the following property. Given any state X and any action a such that $h^{S(X),a}$ is not known (that is, $h^{S(X),a}(X) = u$), there exists a successor state X' such that $h^{S(X),a}(X') = b$ and $X' = \text{argmin}_{Y \in succ(X,a)} c(S(X), a, S(Y)) + v^*(Y)$, where $v^*(Y)$ is the expected cost of executing an optimal policy at state Y (**Def. 1**). We will use the notation $succ(X, a)^b$ (i.e., the best successor) to denote the state X' whose $h^{S(X),a}(X') = b$ if $h^{S(X),a}(X) = u$ and whose $h^{S(X),a}(X') = h^{S(X),a}(X)$ otherwise.

A Appendix: The Proofs

The pseudocode below assumes the following:

1. Every state $\tilde{X}$ in the search state space initially is assumed to have $v(\tilde{X}) = g(\tilde{X}) = \infty$ and $besta(\tilde{X}) = \mathbf{null}$;

procedure ComputePath$(X_{\rm pivot})$
$\tilde{X}_{\rm searchgoal} = GetStateinSearchGraph([S(X_{\rm pivot}); H(X_{\rm pivot})])$;
$g(\tilde{X}_{\rm searchgoal}) = v(\tilde{X}_{\rm searchgoal}) = \infty$;
OPEN $= \emptyset$;
for every H whose every element h_i satisfies:
  $[(h_i = u \bigvee h_i = b) \bigwedge h_i(X_{\rm pivot}) = u]$ OR $[h_i = h_i(X_{\rm pivot}) \bigwedge h_i(X_{\rm pivot}) \neq u]$
  $\tilde{X} = GetStateinSearchGraph([S_{\rm goal}; H])$;
  $v(\tilde{X}) = \infty$, $g(\tilde{X}) = 0$, $besta(\tilde{X}) = \mathbf{null}$;
  insert $\tilde{X}$ into *OPEN* with $g(\tilde{X}) + h(\tilde{X})$;
while($g(\tilde{X}_{\rm searchgoal}) > \min_{\tilde{X}' \in OPEN} g(\tilde{X}') + h(\tilde{X}')$)
  remove $\tilde{X}$ with the smallest $g(\tilde{X}) + h(\tilde{X})$ from *OPEN*;
  $v(\tilde{X}) = g(\tilde{X})$;
  for each action a and $X' = [S(X'); H(X'); \underline{H}^u(X_{\rm pivot})]$ s.t. $\tilde{X} = [S(succ(X', a)^b); H(succ(X', a)^b)]$
    $\tilde{X}' = GetStateinSearchGraph([S(X'); H(X')])$;
    $Q_a = \sum_{Y \in succ(X', a)} P(X', a, Y) \cdot \max(c(S(X'), a, S(Y)) + w(Y), c(S(X'), a, S(Y)) + v(\tilde{X}))$;
    if $g(\tilde{X}') > Q_a$
      $g(\tilde{X}') = Q_a$;
      $besta(\tilde{X}') = a$;
      insert/update $\tilde{X}'$ in *OPEN* with the priority equal to $g(\tilde{X}') + h(\tilde{X}')$;

Figure 1: ComputePath function

The pseudocode below assumes the following:

1. Every state X initially has $0 \leq w(X) \leq w^b(X)$ and $besta(X) = \mathbf{null}$.

procedure UpdateMDP$(X_{\rm pivot})$
$X = X_{\rm pivot}$; $\tilde{X} = GetStateinSearchGraph([S(X_{\rm pivot}); H(X_{\rm pivot})])$;
while ($S(X) \neq S_{\rm goal}$)
  $w(X) = g(\tilde{X})$; $w([S(X); H(X); \underline{H}^u(X_{\rm pivot})]) = g(\tilde{X})$; $besta(X) = besta(\tilde{X})$;
  if ($besta(X) = \mathbf{null}$) break;
  $X = succ(X, besta(X))^b$; $\tilde{X} = GetStateinSearchGraph([S(X); H(X)])$;

procedure Main()
$X_{\rm pivot} = X_{\rm start}$;
while ($X_{\rm pivot}$! $= \mathbf{null}$)
  ComputePath($X_{\rm pivot}$);
  UpdateMDP($X_{\rm pivot}$);
  find state X on the current policy that has
    $w(X) < E_{X' \in succ(X, besta(X))}(c(S(X), besta(X), S(X')) + w(X'))$;
  if found set $X_{\rm pivot}$ to X;
  otherwise set $X_{\rm pivot}$ to $\mathbf{null}$;

Figure 2: Main function

Let us first define several variables that we will use during the proofs. Let $\underline{H}^b$ be defined as $\underline{H}$ with each $\underline{h}_i$ equal to u replaced by b. X^b is then defined as $[S(X); H(X); \underline{H}^b(X)]$. Let $\underline{H}^u(X)$ be $\underline{H}(X)$ but with each $\underline{h}_i = \underline{h}_i^b$ replaced by u. For every state X we then define X^u state as the following state: $X^u = [S(X); H(X); \underline{H}^u(X)]$.

We now introduce optimistic Q-values. Every state-action pair X and $a \in A(S(X))$ has a $Q_{f,w}(X,a) > 0$ associated with it that is calculated from the action costs $c(S(X), a, S(Y))$ for all states $Y \in succ(X,a)$, the non-negative f-value for state $X' = succ(X,a)^b$ and the non-negative values $w(Y)$ for all states $Y \in succ(X,a)$. $Q_{f,w}(X,a)$ is defined as follows:

$$Q_{f,w}(X,a) = \sum_{Y \in succ(X,a)} P(X,a,Y) \cdot \max(c(S(X),a,S(Y)) + w(Y), c(S(X),a,S(X')) + f(X')) \quad (1)$$

We now define an optimistic path from X_n to X_0 whose $S(X_0) = S_{\text{goal}}$ as follows: $\pi = [\{X_n, a_n, X_{n-1}\}, \ldots, \{X_1, a_1, X_0\}]$, where every time a_i is stochastic, an outcome $X_{k-1} = succ(X_k, a_k)^b$. We define an optimistic cost of an optimistic path $\pi = [\{X_n, a_n, X_{n-1}\}, \ldots, \{X_1, a_1, X_0\}]$ under a non-negative value function w recursively as follows:

$$\phi_\pi(X_i, X_0) = \begin{cases} 0 & \text{if } i = 0 \\ Q_{f(X_{i-1}) = \phi_\pi(X_{i-1}, X_0), w}(X_i, a_i) & \text{if } i > 0 \end{cases} \quad (2)$$

We call a path defined by $besta$ pointers from X_n to X_0 as follows: $\pi_{best} = [\{X_n, a_n, X_{n-1}\}, \ldots, \{X_1, a_1, X_0\}]$, where $a_i = besta(X_i)$ and $X_{i-1} = succ(X_i, a_i)^b$.

We define a greedy path $\pi_{\text{greedy},f,w}(X_n, X_0) = [\{X_n, a_n, X_{n-1}\}, \ldots, \{X_1, a_1, X_0\}]$ with respect to functions f and w that map each state X onto non-negative real-values. It is defined as a path π from X_n to X_0 where for every $1 \leq i \leq n$ $a_i = \text{argmin}_{a \in A(S(X_i))} Q_{f,w}(X_i, a)$ and the outcome $X_{i-1} = succ(X_i, a_i)^b$.

We also define w^b values of states as costs of reaching a goal state under the assumption that the values of the missing variables are all set to b:

$$w^b(X) = \begin{cases} 0 & \text{if } S(X) = S_{\text{goal}} \\ \min_{a \in S(X)} (c(S(X), a, succ(X,a)^b) + w^b(succ(X,a)^b)) & \text{otherwise} \end{cases} \quad (3)$$

A.1 ComputePath Function

In this section we will prove theorems that mainly concern the ComputePath function. We will consider a single execution of ComputePath function. We will take the following convention: the search state-space at any particular execution of ComputePath will be denoted by $\tilde{S}$, any state in $\tilde{S}$ will be denoted by a letter with ˜ above it. The states in the original MDP will not use ˜ sign above it. Thus, if X is a full state, then

$\tilde{X} = [S(X), H(X)]$. We will also reserve the notation $X^{\tilde{X}+\underline{H}^u}$ to denote a full state $[S(\tilde{X}); H(\tilde{X}); \underline{H}^u(X_{\text{pivot}})]$.

Similarly to the definition of π in a full state-space, an optimistic path from $\tilde{X}_n$ to $\tilde{X}_0$ is defined as $\tilde{\pi} = [\{X_n^{\tilde{X}+\underline{H}^u}, a_n, succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b\}, \{X_{n-1}^{\tilde{X}+\underline{H}^u}, a_{n-1}, succ(X_{n-1}^{\tilde{X}+\underline{H}^u}, a_{n-1})^b\}, \ldots, \{X_1^{\tilde{X}+\underline{H}^u}, a_1, succ(X_1^{\tilde{X}+\underline{H}^u}, a_1)^b\}]$, where for every $1 \leq i \leq n$ $\tilde{X}_{i-1} = [S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b); H(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)]$.

Similarly to the definition of π_{best}, a path $\tilde{\pi}_{best}$ from $\tilde{X}_n$ to $\tilde{X}_0$ in a full state-space is defined as $\tilde{\pi}$ from $\tilde{X}_n$ to $\tilde{X}_0$ where for every $1 \leq i \leq n$ $a_i = best a(\tilde{X}_i)$.

In addition, we define a greedy path $\tilde{\pi}_{\text{greedy},f,w}$ with respect to functions f and w that map each state $\tilde{X}$ onto non-negative real-values. It is defined as a path $\tilde{\pi}$ from $\tilde{X}_n$ to $\tilde{X}_0$ where for every $1 \leq i \leq n$ $a_i = \text{argmin}_{a \in A(S(\tilde{X}_i))} Q_{f,w}(X_i^{\tilde{X}+\underline{H}^u}, a)$.

We define goal distances, g^*-values under a function w recursively as follows:

$$g^*(\tilde{X}) = \begin{cases} 0 & \text{if } S(\tilde{X}) = S_{\text{goal}} \\ \min_{a \in A(S(\tilde{X}))} Q_{f(Y=succ(X^{\tilde{X}+\underline{H}^u}, a)^b)=g^*(\tilde{Y}),w}(X^{\tilde{X}+\underline{H}^u}, a) & \text{otherwise} \end{cases} \quad (4)$$

Finally, we require that the heuristics are consistent in the following sense: $h(\tilde{X}_{\text{searchgoal}}) = 0$ and for every other state $\tilde{X}$, $a \in A(S(\tilde{X}))$ and $\tilde{Y}$ s.t. $Y = succ(X^{\tilde{X}+\underline{H}^u}, a)^b$, $h(\tilde{Y}) \leq h(\tilde{X}) + c(S(\tilde{X}), a, S(\tilde{Y}))$.

A.1.1 Low-level Correctness

Lemma 1 *Given a non-negative function w, for any state $\tilde{X}_n$ $g^*(\tilde{X}_n) = \phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) = \min_{\tilde{\pi}}$ from $\tilde{X}_n$ to $\tilde{X}_0$ $\phi_{\tilde{\pi}}(\tilde{X}_n, \tilde{X}_0)$ where $\tilde{X}_0$ is the only state on $\tilde{\pi}_{greedy,g^*,w}(\tilde{X}_n, \tilde{X}_0)$ that has $S(\tilde{X}_0) = S_{\text{goal}}$. In addition, it holds that $H(\tilde{X}_0)$ satisfies the equation on line 5.*

Proof: Let us first prove that $g^*(\tilde{X}_n) = \phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0)$. Let us write out the formula for $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0)$. If $n = 0$, $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) = g^*(\tilde{X}_n) = 0$ since $S(\tilde{X}_n) = S(\tilde{X}_0) = S_{\text{goal}}$.

Suppose now, $n \neq 0$. Then

$$\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) = \min_{a \in A(S(\tilde{X}_n))} Q_{f(X_{n-1})=\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_{n-1}, \tilde{X}_0),w}(X_n^{\tilde{X}+\underline{H}^u}, a)$$

According to the definition of an optimistic path $\tilde{\pi}$, $X_{n-1} = succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b$. It is thus the exact same formula as for g^*-values (equation 4).

Let us now prove that $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) = \min_{\tilde{\pi}}$ from $\tilde{X}_n$ to $\tilde{X}_0$ $\phi_{\tilde{\pi}}(\tilde{X}_n, \tilde{X}_0)$. Let us denote argmin$_{\tilde{\pi}}$ from $\tilde{X}_n$ to $\tilde{X}_0$ $\phi_{\tilde{\pi}}(\tilde{X}_n, \tilde{X}_0)$ by $\tilde{\pi}^*(\tilde{X}_n, \tilde{X}_0)$ and $\min_{\tilde{\pi}}$ from $\tilde{X}_n$ to $\tilde{X}_0$ $\phi_{\tilde{\pi}}(\tilde{X}_n, \tilde{X}_0)$ by $\phi_{\tilde{\pi}^*}(\tilde{X}_n, \tilde{X}_0)$.

Since $\tilde{\pi}^*(\tilde{X}_n, \tilde{X}_0)$ is an optimal optimistic path, $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) \geq \phi_{\tilde{\pi}^*}(\tilde{X}_n, \tilde{X}_0)$. We therefore need to show that $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) \leq \phi_{\tilde{\pi}^*}(\tilde{X}_n, \tilde{X}_0)$ also.

The proof is a simple proof by contradiction. Let us assume that $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_n, \tilde{X}_0) > \phi_{\tilde{\pi}^*}(\tilde{X}_n, \tilde{X}_0)$. This implies that $\phi_{\tilde{\pi}^*}(\tilde{X}_n, \tilde{X}_0)$ is finite and therefore path $\tilde{\pi}^*(\tilde{X}_n, \tilde{X}_0)$ is finite (since $\phi_{\tilde{\pi}^*}(\tilde{X}_i, \tilde{X}_0) > \phi_{\tilde{\pi}^*}(\tilde{X}_{i-1}, \tilde{X}_0)$ for all $n \geq i > 0$ and $\phi_{\tilde{\pi}^*}(\tilde{X}_0, \tilde{X}_0) = 0$).

Consider a pair of states $\tilde{X}_i$ and $\tilde{X}_{i-1}$ on the path $\tilde{\pi}^*(\tilde{X}_n, \tilde{X}_0)$ such that $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_i, \tilde{X}_0) > \phi_{\tilde{\pi}^*}(\tilde{X}_i, \tilde{X}_0)$ but $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_{i-1}, \tilde{X}_0) \leq \phi_{\tilde{\pi}^*}(\tilde{X}_{i-1}, \tilde{X}_0)$. Such pair must exist since at least for X_0, $\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_0, \tilde{X}_0) = \phi_{\tilde{\pi}^*}(\tilde{X}_0, \tilde{X}_0) = 0$. Then we get the following contradiction.

$$\begin{aligned}
\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_i, \tilde{X}_0) &\leq Q_{f(X_{i-1})=\phi_{\tilde{\pi}_{greedy,g^*,w}}(\tilde{X}_{i-1},\tilde{X}_0),w}(X_i^{\tilde{X}+\underline{H}^u}, a) \\
&\leq Q_{f(X_{i-1})=\phi_{\tilde{\pi}^*}(\tilde{X}_{i-1},\tilde{X}_0),w}(X_i^{\tilde{X}+\underline{H}^u}, a) \\
&= \phi_{\tilde{\pi}^*}(\tilde{X}_i, \tilde{X}_0)
\end{aligned}$$

We now show that it holds that $H(\tilde{X}_0)$ satisfies the equation on line 5. Consider any h_i. Until path $\tilde{\pi}$ involves executing an action whose outcomes depend on h_i, any state $\tilde{X}_i$ on the path will have $h_i(\tilde{X}_i) = h_i(\tilde{X}_{\text{pivot}})$. Suppose now at state $\tilde{X}_i$ an action a is executed whose outcomes depend on h_i. Then, if $h_i(X_{\text{pivot}}) \neq u$, the action is deterministic and $h_i(\tilde{X}_{i-1}) = h_i(\tilde{X}_{\text{pivot}})$, which is consistent with the equation on line 5. $h_i(\tilde{X}_{i-1})$ remains to be such until the end of the path. On the other hand, if $h_i(X_{\text{pivot}}) = u$, then action a may have multiple outcomes, but an optimistic path always chooses the preferred outcome: $X_{i-1} = succ(X_{i-1}^{\tilde{X}+\underline{H}^u}, a)^b$. Therefore, $h_i(\tilde{X}_{i-1}) = b$ and remains such until the end of the path. This is again consistent with the equation on line 5. Finally, if path $\tilde{\pi}$ does not involve executing an action whose outcomes depend on h_i, then $h_i(\tilde{X}_0) = h_i(\tilde{X}_{\text{pivot}})$, which is also consistent with the equation on line 5. ■

Lemma 2 *Given a non-negative function w and a path $\tilde{\pi}_{greedy,g^*,w}$ from $\tilde{X}_n$ to any state $\tilde{X}_0$ with $S(\tilde{X}_0) = S_{\text{goal}}$ it holds that $g^*(\tilde{X}_n) \geq \sum_{j=n}^{i+1} c(S(\tilde{X}_j), a_j, S(\tilde{X}_{j-1})) + g^*(\tilde{X}_i)$ for any $0 \leq i \leq n$*

Proof: The following is the proof that the theorem holds for $i = n - 1$.

$$\begin{aligned}
g^*(\tilde{X}_n) &= \min_{a \in A(S(\tilde{X}_n))} Q_{f(Y=succ(X_n^{\tilde{X}+\underline{H}^u}, a)^b)=g^*(\tilde{Y}),w}(X_n^{\tilde{X}+\underline{H}^u}, a) \\
&= Q_{f(Y=succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b)=g^*(\tilde{Y}),w}(X_n^{\tilde{X}+\underline{H}^u}, a_n) \\
&= \sum_{Y \in succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)} P(X_n^{\tilde{X}+\underline{H}^u}, a_n, Y) \cdot \\
&\quad \max(c(S(\tilde{X}), a_n, S(\tilde{Y})) + w(Y), c(S(\tilde{X}), a_n, S(succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b)) + g^*(succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b)) \\
&\geq c(S(\tilde{X}), a_n, S(succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b)) + g^*(succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b) \\
&= c(S(\tilde{X}), a_n, S(\tilde{X}_{n-1})) + g^*(\tilde{X}_{n-1})
\end{aligned}$$

The proof for $0 \leq i \leq n-1$ holds by induction on i. ■

Lemma 3 *At any point in time, for any state $\tilde{X}$ it holds that $v(\tilde{X}) \geq g(\tilde{X})$.*

Proof: The theorem clearly holds before line 9 was executed for the first time since for each state $\tilde{X}$ $v(\tilde{X}) = \infty$. Afterwards, the g-values can only decrease (lines 15-16). For any state $\tilde{X}$, on the other hand, $v(\tilde{X})$ only changes on line 11 when it is set to $g(\tilde{X})$. Thus, it is always true that $v(\tilde{X}) \geq g(\tilde{X})$. ■

Lemma 4 *Assuming function w is non-negative, at line 9, the following holds:*

- $g(\tilde{X}) = 0$, $besta(\tilde{X}) = \mathbf{null}$ *for every state $\tilde{X}$ whose $S(\tilde{X}) = S_{\text{goal}}$ and $H(\tilde{X})$ satisfies the equation on line 5*
- $g(\tilde{X}) = Q_{f(Y)=v(\tilde{Y}),w}(X^{\tilde{X}+\underline{H^u}}, besta(\tilde{X}))$ *and* $besta(\tilde{X}) = argmin_{a \in A(S(\tilde{X}))} Q_{f(Y)=v(\tilde{Y}),w}(X^{\tilde{X}+\underline{H^u}}, a)$*, for every other state $\tilde{X}$*
- *if $g(\tilde{X}) = \infty$, then $besta(\tilde{X}) = \mathbf{null}$*

Proof: The theorem holds the first time line 9 is executed. This is so because every state $\tilde{X} \in \tilde{S}$ has $v(\tilde{X}) = \infty$. As a result, the right-hand side of the equation 1 evaluated under function $f = \infty$ is equal to ∞, independently of action a. This is correct, since after the initialization every state $\tilde{X}$ with $S(\tilde{X}) \neq S_{\text{goal}}$ or whose $H(\tilde{X})$ does not satisfy the equation on line 5 has $g(\tilde{X}) = \infty$, $besta(\tilde{X}) = \mathbf{null}$ and every state $\tilde{X}$ with $S(\tilde{X}) = S_{\text{goal}}$ and $H(\tilde{X})$ satisfying the equation on line 5 has $g(\tilde{X}) = 0$, $besta(\tilde{X}) = \mathbf{null}$.

The only place where g- and v-values are changed afterwards is on lines 11 and 16. If $v(s)$ is changed in line 11, then it is decreased according to Lemma 3. Thus, it may only decrease the g-values of its successors. The test on line 15 checks this and updates the g-values and $besta$ pointers as necessary. Since all costs are positive and never change, g-value of a state $\tilde{X}$ with $S(\tilde{X}) = S_{\text{goal}}$ and $H(\tilde{X})$ satisfying the equation on line 5 can never be changed: it will never pass the test on line 15, and thus is always 0. Also, since g-values do not increase, it continues to hold that if $g(\tilde{X}) = \infty$, then $besta(\tilde{X}) = \mathbf{null}$.

■

Lemma 5 *At line 9, OPEN contains all and only states $\tilde{X}$ whose $v(\tilde{X}) \neq g(\tilde{X})$.*

Proof: The first time line 9 is executed the theorem holds since after the initialization the only states in *OPEN* are the states $\tilde{X}$ with $v(\tilde{X}) = \infty \neq 0 = g(\tilde{X})$. The rest of the states have infinite values.

During the following execution whenever we decrease $g(\tilde{X})$ (line 16), and as a result make $g(\tilde{X}) < v(\tilde{X})$ (Lemma 3), we insert it into *OPEN*; whenever we remove $\tilde{X}$ from *OPEN* (line 10) we set $v(\tilde{X}) = g(\tilde{X})$ (line 11) making the state consistent. We never modify $v(\tilde{X})$ or $g(\tilde{X})$ elsewhere. ■

Lemma 6 *Assuming function w is non-negative, suppose $\tilde{X}$ is selected for expansion on line 10. Then the next time line 9 is executed $v(\tilde{X}) = g(\tilde{X})$, where $g(\tilde{X})$ before and after the expansion of $\tilde{X}$ is the same.*

Proof: Suppose $\tilde{X}$ is selected for expansion. Then on line 11 $v(\tilde{X}) = g(\tilde{X})$, and it is the only place where a v-value changes. We, thus, only need to show that $g(\tilde{X})$ does not change. It could only change if $\tilde{X}' = \tilde{X}$ and $g(\tilde{X}') > Q_a$ at one of the executions of line 15. The former condition means that there exists a such that $\tilde{X} = [S(succ(X^{\tilde{X}+\underline{H}^u}, a)^b); H(succ(X^{\tilde{X}+\underline{H}^u}, a)^b)]$. The later condition means that $g(\tilde{X}) > Q_{f(Y)=v(\tilde{Y}),w}(X^{\tilde{X}+\underline{H}^u}, a)$.

Since $\tilde{X} = [S(succ(X^{\tilde{X}+\underline{H}^u}, a)^b); H(succ(X^{\tilde{X}+\underline{H}^u}, a)^b)]$, $f(succ(X^{\tilde{X}+\underline{H}^u}, a)^b) = v(\tilde{X}) = g(\tilde{X})$. Hence, $g(\tilde{X}) > Q_{f(succ(X^{\tilde{X}+\underline{H}^u},a)^b)=g(\tilde{X}),w}(X^{\tilde{X}+\underline{H}^u}, a)$. This means that $g(\tilde{X}) > c(S(\tilde{X}), a, S(\tilde{X})) + g(\tilde{X})$ which is impossible since costs are positive. ■

Lemma 7 *Assuming function w is non-negative, at line 9, for any state $\tilde{X}$, an optimistic cost of a path defined by $besta$ pointers, $\tilde{\pi}_{best}$, from $\tilde{X}$ to a state $\tilde{X}_0$ whose $S(\tilde{X}_0) = S_{\text{goal}}$ is no larger than $g(\tilde{X})$, that is, $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) \leq g(\tilde{X})$. In addition, $v(\tilde{X}) \geq g(\tilde{X}) \geq g^*(\tilde{X})$.*

Proof: $v(\tilde{X}) \geq g(\tilde{X})$ holds according to Lemma 3. We thus need to show that $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) \leq g(\tilde{X})$, and $g(\tilde{X}) \geq g^*(\tilde{X})$. The statement follows if $g(\tilde{X}) = \infty$. We thus can restrict our proof to a finite g-value.

Consider a path $\tilde{\pi}_{best}$ from $\tilde{X} = \tilde{X}_n$ to a state $\tilde{X}_0$: $\tilde{\pi}_{best} = [\{X_n^{\tilde{X}+\underline{H}^u}, a_n, succ(X_n^{\tilde{X}+\underline{H}^u}, a_n)^b\}, \{X_{n-1}^{\tilde{X}+\underline{H}^u}, a_{n-1}, succ(X_{n-1}^{\tilde{X}+\underline{H}^u}, a_{n-1})^b\}, \ldots, \{X_1^{\tilde{X}+\underline{H}^u}, a_1, succ(X_1^{\tilde{X}+\underline{H}^u}, a_1)^b\}]$, where $a_i = besta(\tilde{X}_i)$ and $\tilde{X}_{i-1} = [S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b); H(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)]$.

We now show that $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) \leq g(\tilde{X})$ by contradiction. Suppose it does not hold. Let us then pick a state $\tilde{X}_k$ on the path that is closest to $\tilde{X}_0$ and for which $\phi_{\tilde{\pi}_{best}}(\tilde{X}_k, \tilde{X}_0) > g(\tilde{X}_k)$. $S(X_k) \neq S_{\text{goal}}$ because otherwise $\phi_{\tilde{\pi}_{best}}(\tilde{X}_k, \tilde{X}_0) = 0$ from the definition of ϕ-values. Consequently, $\phi_{\tilde{\pi}_{best}}(\tilde{X}_k, \tilde{X}_0) = Q_{f(succ(X_k^{\tilde{X}+\underline{H}^u},a_k)^b)=\phi_{\tilde{\pi}_{best}}(\tilde{X}_{k-1},\tilde{X}_0),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k)$. According to Lemma 4 $g(\tilde{X}_k) = Q_{f(Y)=v(\tilde{Y}),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k)$, where $a_k = besta(\tilde{X}_k)$. From Lemma 3 it then also follows that $g(\tilde{X}_k) \geq Q_{f(Y)=g(\tilde{Y}),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k)$. Hence, $g(\tilde{X}_k) \geq Q_{f(succ(X_k^{\tilde{X}+\underline{H}^u},a_k)^b)=g(\tilde{X}_{k-1}),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k)$.

Finally, because of the way we picked state $\tilde{X}_k$, $\phi_{\tilde{\pi}_{best}}(\tilde{X}_{k-1}, \tilde{X}_0) \leq g(\tilde{X}_{k-1})$. As a result, $g(\tilde{X}_k) \geq Q_{f(succ(X_k^{\tilde{X}+\underline{H}^u},a_k)^b)=g(\tilde{X}_{k-1}),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k) \geq Q_{f(succ(X_k^{\tilde{X}+\underline{H}^u},a_k)^b)=\phi_{\tilde{\pi}_{best}}(\tilde{X}_{k-1},\tilde{X}_0),w}(X_k^{\tilde{X}+\underline{H}^u}, a_k) = \phi_{\tilde{\pi}_{best}}(\tilde{X}_k, \tilde{X}_0)$. This is a contradiction to the assumption that $\phi_{\tilde{\pi}_{best}}(\tilde{X}_k, \tilde{X}_0) > g(\tilde{X}_k)$.

Since $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) \leq g(\tilde{X})$ the proof that $g(\tilde{X}) \geq g^*(\tilde{X})$ follows directly from Lemma 1. ■

A.1.2 Main theorems

Theorem 1 *Assuming function w is non-negative, at line 9, for any state $\tilde{X}$ with ($h(\tilde{X}) < \infty \wedge g(\tilde{X}) + h(\tilde{X}) \leq g(\tilde{U}) + h(\tilde{U})\ \forall \tilde{U} \in$ OPEN), it holds that $g(\tilde{X}) = g^*(\tilde{X})$.*

Proof: We prove by contradiction. Suppose there exists $\tilde{X}$ such that $h(\tilde{X}) < \infty \wedge g(\tilde{X}) + h(\tilde{X}) \leq g(\tilde{U}) + h(\tilde{U})\ \forall \tilde{U} \in$ *OPEN*, but $g(\tilde{X}) \neq g^*(\tilde{X})$. According to Lemma 7 it then follows that $g(\tilde{X}) > g^*(\tilde{X})$. This also implies that $g^*(\tilde{X}) < \infty$. We also assume that $S(\tilde{X}) \neq S_{\text{goal}}$ or $H(\tilde{X})$ does not satisfy the equation on line 5 since otherwise $g(\tilde{X}) = 0 = g^*(\tilde{X})$ from Lemma 4.

Consider a path $\tilde{\pi}_{greedy,g^*,w}$ from $\tilde{X} = \tilde{X}_n$ to a state $\tilde{X}_0$ whose $S(\tilde{X}_0) = S_{\text{goal}}$. According to Lemma 1, the cost of this path is $g^*(\tilde{X})$ and $H(\tilde{X}_0)$ satisfies the equation on line 5. Such path must exist since $g^*(\tilde{X}) < \infty$ and from equation 4 it is clear that $g^*(\tilde{X}_i) > g^*(\tilde{X}_{i-1})$ for each $1 \leq i \leq n$ on the path.

Our assumption that $g(\tilde{X}) > g^*(\tilde{X})$ means that there exists at least one $\tilde{X}_i$ on the path $\tilde{\pi}_{greedy,g^*,w}$, namely $\tilde{X}_{n-1}$, whose $v(\tilde{X}_i) > g^*(\tilde{X}_i)$. Otherwise,

$$
\begin{aligned}
g(\tilde{X}) = g(\tilde{X}_n) &= \text{(Lemma 4)} \\
\min_{a \in A(S(\tilde{X}_n))} Q_{f(Y)=v(\tilde{Y}),w}(X^{\tilde{X}+\underline{H}^u}, a) &\leq \\
Q_{f(Y)=v(\tilde{Y}),w}(X^{\tilde{X}+\underline{H}^u}, a_i) &= \text{(def. of } \tilde{\pi}) \\
Q_{f(Y)=v(\tilde{X}_{n-1}),w}(X^{\tilde{X}+\underline{H}^u}, a_i) &\leq \\
Q_{f(Y)=g^*(\tilde{X}_{n-1}),w}(X^{\tilde{X}+\underline{H}^u}, a_i) &= \text{(def. of } g^*) \\
g^*(\tilde{X}_n) = g^*(\tilde{X}) &
\end{aligned}
$$

Let us now consider $\tilde{X}_i$ on the path with the smallest index $i \geq 0$ (that is, closest to $\tilde{X}_0$) such that $v(\tilde{X}_i) > g^*(\tilde{X}_i)$. We will first show that $g^*(\tilde{X}_i) \geq g(\tilde{X}_i)$. It is clearly so when $i = 0$ according to Lemma 4 which says that $g(\tilde{X}_i) = 0$ whenever $S(\tilde{X}_i) = S_{\text{goal}}$ and $H(\tilde{X}_i)$ satisfies the equation on line 5. For $i > 0$ we use the fact that $v(\tilde{X}_{i-1}) \leq g^*(\tilde{X}_{i-1})$ from the way $\tilde{X}_i$ was chosen,

$$
\begin{aligned}
g(\tilde{X}_i) &= \text{(Lemma 4)} \\
\min_{a \in A(S(\tilde{X}_i))} Q_{f(Y)=v(\tilde{Y}),w}(X_i^{\tilde{X}+\underline{H}^u}, a) &\leq \\
Q_{f(Y)=v(\tilde{Y}),w}(X_i^{\tilde{X}+\underline{H}^u}, a_i) &= \text{(def. of } \tilde{\pi}) \\
Q_{f(Y)=v(\tilde{X}_{i-1}),w}(X_i^{\tilde{X}+\underline{H}^u}, a_i) &\leq \\
Q_{f(Y)=g^*(\tilde{X}_{i-1}),w}(X_i^{\tilde{X}+\underline{H}^u}, a_i) &= \text{(def. of } g^*)
\end{aligned}
$$

$$g^*(\tilde{X}_i)$$

We thus have $v(\tilde{X}_i) > g^*(\tilde{X}_i) \geq g(\tilde{X}_i)$, which implies that $\tilde{X}_i \in$ *OPEN* according to Lemma 5.

We will now show that $g(\tilde{X}) + h(\tilde{X}) > g(\tilde{X}_i) + h(\tilde{X}_i)$, and finally arrive at a contradiction. According to our assumption $g(\tilde{X}) > g^*(\tilde{X})$ and $h(\tilde{X}) < \infty$, therefore

$$\begin{aligned}
g(\tilde{X}) + h(\tilde{X}) &= \\
g(\tilde{X}_n) + h(\tilde{X}_n) &> \\
g^*(\tilde{X}_n) + h(\tilde{X}_n) &\geq \text{(Lemma 2)} \\
\sum_{j=n}^{i+1} c(S(\tilde{X}_j), a_j, S(\tilde{X}_{j-1})) + g^*(\tilde{X}_i) + h(\tilde{X}_n) &\geq \text{(property of } h\text{)} \\
\sum_{j=n-1}^{i+1} c(S(\tilde{X}_j), a_j, S(\tilde{X}_{j-1})) + g^*(\tilde{X}_i) + h(\tilde{X}_{n-1}) &\geq \\
\dots & \\
g^*(\tilde{X}_i) + h(\tilde{X}_i) &\geq \\
g(\tilde{X}_i) + h(\tilde{X}_i) &
\end{aligned}$$

This inequality, however, implies that $\tilde{X}_i \notin$ *OPEN* since according to the conditions of the theorem $g(\tilde{X}) + h(\tilde{X}) \leq g(\tilde{U}) + h(\tilde{U})\ \forall \tilde{U} \in$ *OPEN*. But this contradicts to what we have proven earlier. ■

A.1.3 Correctness

The corollaries in this section show how the theorems in the previous section lead quite trivially to the correctness of ComputePath. We also show that each state is expanded at most once, similar to the guarantee that A* makes for deterministic graphs whenever heuristics are consistent.

Corollary 1 *When the ComputePath function exits the following holds for any state $\tilde{X}$ with $h(\tilde{X}) < \infty \wedge g(\tilde{X}) + h(\tilde{X}) \leq \min_{\tilde{X}' \in OPEN}(g(\tilde{X}') + h(\tilde{X}'))$: an optimistic cost of a path defined by $besta$ pointers, $\tilde{\pi}_{best}$, from $\tilde{X}$ to a state $\tilde{X}_0$ whose $S(\tilde{X}_0) = S_{\text{goal}}$ is equal to $g^*(\tilde{X})$, that is, $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) = g^*(\tilde{X})$.*

Proof: According to Theorem 1 the condition $h(\tilde{X}) < \infty \wedge g(\tilde{X}) + h(\tilde{X}) \leq \min_{\tilde{X}' \in OPEN}(g(\tilde{X}') + h(\tilde{X}'))$ implies that $g(\tilde{X}) = g^*(\tilde{X})$. From Lemma 7 it then follows that $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) \leq g^*(\tilde{X})$. Since $g^*(\tilde{X})$ is an optimistic cost of a least-cost path from $\tilde{X}$ to $\tilde{X}_0$ according to Lemma 1, $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0) = g^*(\tilde{X})$. ■

Corollary 2 *When the ComputePath function exits the following holds: an optimistic cost of a path defined by $besta$ pointers, $\tilde{\pi}_{best}$, from $\tilde{X}_{\text{searchgoal}}$ to a state $\tilde{X}_0$ whose $S(\tilde{X}_0) = S_{\text{goal}}$ is equal to $g^*(\tilde{X}_{\text{searchgoal}})$, that is, $\phi_{\tilde{\pi}_{best}}(\tilde{X}_{\text{searchgoal}}, \tilde{X}_0) = g^*(\tilde{X}_{\text{searchgoal}})$. The length of this path is finite.*

Proof: According to the termination condition of the ComputePath function, upon its exit $g(\tilde{X}_{\text{searchgoal}}) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$. Since $h(\tilde{X}_{\text{searchgoal}}) = 0$ the proof that the cost of the path is equal to $g^*(\tilde{X}_{\text{searchgoal}})$ then follows directly from Corollary 1

To prove that the path defined by $besta$ pointers is always finite, first consider the case of $g(\tilde{X}_{\text{searchgoal}}) = \infty$. According to lemma 4 then, $besta(\tilde{X}_{\text{searchgoal}}) = \textbf{null}$ and the path defined by $besta$ pointers is therefore empty. Suppose now $g(\tilde{X}_{\text{searchgoal}}) \neq \infty$. Since $g(\tilde{X}_{\text{searchgoal}}) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$ and $h(\tilde{X}_{\text{searchgoal}}) = 0$, theorem 1 applies and therefore $\infty > g(\tilde{X}_{\text{searchgoal}}) = g^*(\tilde{X}_{\text{searchgoal}})$. As a result, the optimistic cost of the path defined by $besta$ pointers is also finite according to lemma 7. Considering that the costs are bounded from below by a positive constant, it shows that the path is of finite length. ■

Corollary 3 *When the ComputePath function exits the following holds for each state $\tilde{X}$ on the path $\tilde{\pi}_{best}(\tilde{X}_{\text{searchgoal}}, \tilde{X}_0)$: $g(\tilde{X}) = g^*(\tilde{X})$.*

Proof: At the time ComputePath terminates $g(\tilde{X}_{\text{searchgoal}}) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$. and $h(\tilde{X}_{\text{searchgoal}}) = 0$. Thus, according to theorem 1, $g(\tilde{X}_{\text{searchgoal}}) = g^*(\tilde{X}_{\text{searchgoal}})$.

We now prove that the theorem holds for the rest of the states on the path defined by $besta$ pointers. The case when $g(\tilde{X}_{\text{searchgoal}}) = \infty$ is trivially proven by noting that in this case $besta(\tilde{X}_{\text{searchgoal}}) = \textbf{null}$ according to lemma 4. We therefore consider the case when $g(\tilde{X}_{\text{searchgoal}}) \neq \infty$. We prove the theorem for this case by induction. Suppose $g(\tilde{X}_i) = g^*(\tilde{X}_i)$, $g(\tilde{X}_i) + h(\tilde{X}_i) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$ and $h(\tilde{X}_i) < \infty$. This is true at least for the first state on the path, namely, $\tilde{X}_{\text{searchgoal}}$. We will show that $g(\tilde{X}_{i-1}) = g^*(\tilde{X}_{i-1})$, $g(\tilde{X}_{i-1}) + h(\tilde{X}_{i-1}) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$ and $h(\tilde{X}_{i-1}) < \infty$. This induction step will prove the statement of the theorem.

The property $h(\tilde{X}_{i-1}) < \infty$ follows from the consistency of heuristics and the fact that $h(\tilde{X}_i) < \infty$. By consistency $h(\tilde{X}_{i-1}) \leq h(\tilde{X}_i) + c(S(\tilde{X}_i), besta(\tilde{X}_i), S(\tilde{X}_{i-1}))$. $h(\tilde{X}_i)$ is finite according to our induction assumption, whereas the costs are finite because $\infty > g(\tilde{X}_{\text{searchgoal}}) = g^*(\tilde{X}_{\text{searchgoal}})$. Thus, $h(\tilde{X}_{i-1}) < \infty$.

To prove that $g(\tilde{X}_{i-1}) + h(\tilde{X}_{i-1}) \leq \min_{\tilde{X}' \in \textit{OPEN}}(g(\tilde{X}') + h(\tilde{X}'))$ we will show that $g(\tilde{X}_{i-1}) + h(\tilde{X}_{i-1}) \leq g(\tilde{X}_i) + h(\tilde{X}_i)$ as follows:

$$
\begin{aligned}
g(\tilde{X}_{i-1}) + h(\tilde{X}_{i-1}) &\leq && \text{consistency of heuristics} \\
g(\tilde{X}_{i-1}) + h(\tilde{X}_i) + c(S(\tilde{X}_i), besta(\tilde{X}_i), S(\tilde{X}_{i-1})) &\leq && \text{lemma 3} \\
v(\tilde{X}_{i-1}) + c(S(\tilde{X}_i), besta(\tilde{X}_i), S(\tilde{X}_{i-1})) + h(\tilde{X}_i) &\leq && \\
\sum_{Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, besta(\tilde{X}_i))} P(X_i^{\tilde{X}+\underline{H}^u}, besta(\tilde{X}_i), Y) \cdot & && \\
\max(c(S(\tilde{X}_i), besta(\tilde{X}_i), S(\tilde{Y})) + w(Y), c(S(\tilde{X}), besta(\tilde{X}_i), S(\tilde{X}_{i-1})) + v(\tilde{X}_{i-1})) + h(\tilde{X}_i) &= && \text{eq. 1} \\
Q_{f(Y)=v(\tilde{Y}), w}(X_i^{\tilde{X}+\underline{H}^u}, besta(\tilde{X}_i)) + h(\tilde{X}_i) &= && \text{lemma 4} \\
g(\tilde{X}_i) + h(\tilde{X}_i) &\leq && \text{inductive assumption} \\
\min_{\tilde{X}' \in OPEN} (g(\tilde{X}') + h(\tilde{X}')) & &&
\end{aligned}
$$

Finally, the fact that $g(\tilde{X}_{i-1}) = g^*(\tilde{X}_{i-1})$ now comes directly from theorem 1. ■

Theorem 2 *No state is expanded more than once during the execution of the ComputePath function.*

Proof: Suppose a state $\tilde{X}$ is selected for expansion for the first time during the execution of the ComputePath function. Then, it is removed from *OPEN* set on line 10. According to theorem 1 its g-value at this point is equal to $g^*(\tilde{X})$. On line 11 the state is made consistent by setting its v-value to its g-value. The only way how $\tilde{X}$ can be chosen for expansion again is if it is inserted into *OPEN*, but this only happens if its g-value is decreased. This however is impossible since $g(\tilde{X})$ is already equal to $g^*(\tilde{X}) = \min_{\tilde{\pi}}$ from $_{\tilde{X}_n}$ to $_{\tilde{X}_0}\, \phi_{\tilde{\pi}}(\tilde{X}_n, \tilde{X}_0)$ where $\tilde{X}_0$ has $S(\tilde{X}_0) = S_{\text{goal}}$ (according to Lemma 1) and $g(\tilde{X})$ must always remain an upper bound on $\phi_{\tilde{\pi}_{best}}(\tilde{X}, \tilde{X}_0)$ (according to Lemma 7). ■

A.2 Main Function

In this section we present the theorems about the main function of the algorithm. All references to line numbers are for the figure 2 unless explicitly specified otherwise.

By $w^*(X)$ we denote a minimum expected cost of a policy for reaching a goal state from state X. We also introduce w^u-values defined recursively as follows:

$$w^u(X) = \begin{cases} 0 & \text{if } S(X) = 0 \\ \min_{a \in A(S(X))} Q_{w^u, w^u}(X^u, a) & \text{otherwise} \end{cases} \tag{5}$$

We also define goal distances for full states g^*-values under a function w recursively as follows:

$$g^*(X) = \begin{cases} 0 & \text{if } S(X) = S_{\text{goal}} \\ \min_{a \in A(S(X))} Q_{f(Y = succ(X^u, a)^b) = g^*(Y), w}(X^u, a) & \text{otherwise} \end{cases} \tag{6}$$

Lemma 8 *For each X, $w^u(X) = w^u(X^u)$*

Proof: According to equation 5, if $S(X) = S(X^u) = S_{\text{goal}}$ then $w^u(X) = w^u(X^u) = 0$. Otherwise, $w^u(X) = \min_{a \in A(S(X))} Q_{w^u, w^u}(X^u, a) = \min_{a \in A(S(X))} Q_{w^u, w^u}((X^u)^u, a) = w^u(X^u)$. ■

Lemma 9 *For each X, $g^*(X) = g^*(X^u)$*

Proof: According to the definition, $X^u = [S(X); H(X); \underline{H}^u(X)]$, and therefore $S(X) = S(X^u)$. Suppose first $S(X) = S_{\text{goal}}$. Then, according to equation 6, $g^*(X) = 0$ and $g^*(X^u) = 0$.

Now suppose $S(X) = S(X^u) \neq S_{\text{goal}}$. Then, according to equation 6, $g(X) = \min_{a \in A(S(X))} Q_{f(Y=succ(X^u,a)^b)=g^*(Y),w}(X^u, a)$ and $g(X^u) = \min_{a \in A(S(X^u))} Q_{f(Y=succ((X^u)^u,a)^b)=g^*(Y),w}((X^u)^u, a)$.

$(X^u)^u = X^u$ because $\underline{H}^u(X)$ does not contain any $\underline{h}_i$ elements equal to b and therefore $\underline{H}^u(X^u) = \underline{H}^u(X)$. Also, $S(X) = S(X^u)$. Consequently, $g(X^u) = \min_{a \in A(S(X))} Q_{f(Y=succ(X^u,a)^b)=g^*(Y),w}(X^u, a) = g(X)$. ■

Lemma 10 *For each X and $a \in A(S(X))$, $h^{S(X),a}(succ(X,a)^b) = h^{S(X),a}(succ(X^u,a)^b)$ and $g^*(succ(X^u,a)^b) = g^*(succ(X,a)^b)$*

Proof: We consider all possible cases for $h^{S(X),a}(X)$. Suppose first $h^{S(X),a}(X) = \textbf{null}$. That is, action a is (and always was) deterministic. Then $h^{S(X),a}(X^u) = \textbf{null}$ also and therefore $h^{S(X),a}(succ(X,a)^b) = h^{S(X),a}(succ(X^u,a)^b) = \textbf{null}$. Also, $succ(X^u,a)^b = (succ(X,a)^b)^u$ because h-values are not affected by action a and therefore $g^*(succ(X^u,a)^b) = g^*(succ(X,a)^b)$ according to lemma 9.

Suppose now $h^{S(X),a}(X) \neq b$. Then again $h^{S(X),a}(X^u) = h^{S(X),a}(X)$ and therefore $h^{S(X),a}(succ(X,a)^b) = h^{S(X),a}(succ(X^u,a)^b)$. Also, $succ(X^u,a)^b = (succ(X,a)^b)^u$ because h-values are not affected by action a and therefore $g^*(succ(X^u,a)^b) = g^*(succ(X,a)^b)$ according to lemma 9.

Now suppose $h^{S(X),a}(X) = b$. If $h^{S(X),a} \in H$, then $h^{S(X),a}(X^u) = b$, whereas if $h^{S(X),a} \in \underline{H}$, then $h^{S(X),a}(X^u) = u$. In either case, however, $h^{S(X),a}(succ(X,a)^b) = h^{S(X),a}(succ(X^u,a)^b) = b$. Also, $g^*(succ(X,a)^b) = g^*((succ(X,a)^b)^u)$ and $g^*(succ(X^u,a)^b) = g^*((succ(X^u,a)^b)^u)$ according to lemma 9. But $(succ(X,a)^b)^u = (succ(X^u,a)^b)^u$ and therefore $g^*(succ(X,a)^b) = g^*(succ(X^u,a)^b)$ as stated in the theorem. ■

Theorem 3 *Suppose that before line 10 is executed for every state X it is true that $0 \leq w(X) \leq w(X^u)$. Then after line 11 is executed for each state X on π_{best} from X_{pivot} to a goal state it holds that $w(X) \geq E_{X' \in succ(X,besta(X))}(c(S(X), besta(X), S(X')) + w(X'))$ if $S(X) \neq S_{\text{goal}}$ and $w(X) = 0$ otherwise.*

Proof: We first prove that after line 11 is executed for each state X_i on π_{best} from $X_{\text{pivot}} = X_n$ to a goal state X_0 it is true that $\tilde{X}_i = [S(X_i); H(X_i)]$, where $\tilde{X}_i$ is the ith state on $\tilde{\pi}_{best}$ from $\tilde{X}_{\text{pivot}} = \tilde{X}_n$ to a goal state $\tilde{X}_0$. We prove this by induction. It certainly holds for $i = n$ since $\tilde{X}_n = [S(X_{\text{pivot}}); H(X_{\text{pivot}})] = [S(X_n); H(X_n)]$. We now prove that it continues to hold for $i - 1$.

On line 6 we pick X_{i-1} to be equal to $succ(X_i, a_i)^b$, where $a_i = besta(X_i) = besta(\tilde{X}_i)$. We thus need to show that $[S(succ(X_i, a_i)^b); H(succ(X_i, a_i)^b)]$ is the $i-1$th state on $\tilde{\pi}_{best}$. According to the definition of $\tilde{\pi}_{best}$ the $i-1$th state on it is defined as: $\tilde{X}_{i-1} = [S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b); H(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)]$. We thus need to show that $S(succ(X_i, a_i)^b) = S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)$ and $H(succ(X_i, a_i)^b) = H(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)$, where $X_i = [S(X_i); H(X_i); \underline{H}(X_i)]$ and $X_i^{\tilde{X}+\underline{H}^u} = [S(X_i); H(X_i); \underline{H}^u(X_{\text{pivot}})]$.

Since according to the definition of π_{best} $X_i = succ(X_{i+1}, a_{i+1})^b = succ(succ(X_{i+2}, a_{i+2})^b, a_{i+1})^b$ and so on, $\underline{h}^{S(X_i),a_i}(X_i)$ can only be different from $\underline{h}^{S(X_i),a_i}(X_i^{\tilde{X}+\underline{H}^u})$ if $\underline{h}^{S(X_i),a_i}(X_i^{\tilde{X}+\underline{H}^u}) = u$ and $\underline{h}^{S(X_i),a_i}(X_i) = b$. Consequently, the same preferred outcome of action a_i exists for both X_i and $X_i^{\tilde{X}+\underline{H}^u}$, namely the one that has $\underline{h}^{S(X_i),a_i} = b$. In other words, $[S(succ(X_i, a_i)^b); H(succ(X_i, a_i)^b)] = [S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b); H(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)]$. That is, $[S(X_{i-1}); H(X_{i-1})] = \tilde{X}_{i-1}$

We now prove the statement of the theorem itself. Consider an arbitrary X_i on π_{best} from $X_{\text{pivot}} = X_n$ to a goal state X_0. Because of the statement we have just proven and the execution of line 4, $w(X_i) = g(\tilde{X}_i)$, where $\tilde{X}_i = [S(X_i); H(X_i)]$ is the ith state on $\tilde{\pi}_{best}$ from $\tilde{X}_{\text{pivot}} = \tilde{X}_n$ to a goal state $\tilde{X}_0$.

If $i = 0$ then $w(X_i) = g(\tilde{X}_i) = 0$ according to Lemma 4.

Suppose now $i > 0$. According to Lemma 4 then $w(X_i) = g(\tilde{X}_i) = Q_{f(Y)=v(\tilde{Y}),w_{old}}(X_i^{\tilde{X}+\underline{H}^u}, a_i)$, where w_{old} is w-function before the execution of the ComputePath function. In addition, the w-value of each $X' \in succ(X_i, a_i)$ such that $X' \neq X_{i-1}$ remains the same as that before the ComputePath function was called. This is so because UpdateMDP does not update w-values of states with at least one $\underline{h}_j$-value that is neither equal to $\underline{h}_j(X_i^{\tilde{X}+\underline{H}^u})$ nor equal to b. Moreover, from Lemma 3 $v(\tilde{X}_{i-1}) \geq g(\tilde{X}_{i-1}) = w(X_{i-1})$. Hence $w(X_i) \geq Q_{f(Y)=w(X_{i-1}),w}(X_i^{\tilde{X}+\underline{H}^u}, a_i)$.

Thus,

$$w(X_i) \geq \sum_{Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)} P(X_i^{\tilde{X}+\underline{H}^u}, a_i, Y) \cdot \max(c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(Y)) + w(Y), c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)) + w(X_{i-1}))$$

We distinguish two cases. First, suppose $\underline{h}^{S(X_i),a_i}(X_i)$ is different from $\underline{h}^{S(X_i),a_i}(X_i^{\tilde{X}+\underline{H}^u})$. This is only possible if $\underline{h}^{S(X_i),a_i}(X_i^{\tilde{X}+\underline{H}^u}) = u$ and $\underline{h}^{S(X_i),a_i}(X_i) = b$. The latter implies that there is only one outcome of $succ(X_i, a_i)$, namely, X_{i-1}. Hence,

$$\begin{aligned} w(X_i) &\geq \sum_{Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)} P(X_i^{\tilde{X}+\underline{H}^u}, a_i, Y) \cdot (c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)) + w(X_{i-1})) \\ &= (c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)) + w(X_{i-1})) \\ &= c(S(X_i), a_i, S(X_{i-1})) + w(X_{i-1})) \\ &= c(S(X_i), besta(X_i), S(X_{i-1})) + w(X_{i-1})) \\ &= E_{X' \in succ(X_i, besta(X_i))}(c(S(X_i), besta(X), S(X')) + w(X')) \end{aligned}$$

Now suppose $\underline{h}^{S(X_i),a_i}(X_i) = \underline{h}^{S(X_i),a_i}(X_i^{\tilde{X}+\underline{H}^u})$. Then the probability distribution is the same for $succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)$ and $succ(X_i, a_i)$. Hence,

$$
\begin{aligned}
w(X_i) \quad &\geq \\
P(X_i^{\tilde{X}+\underline{H}^u}, a_i, succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b) \cdot (c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b)) + w(X_{i-1})) + \\
\sum_{Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, a_i) s.t. Y \neq succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b} P(X_i^{\tilde{X}+\underline{H}^u}, a_i, Y) \cdot (c(S(X_i^{\tilde{X}+\underline{H}^u}), a_i, S(Y)) + w(Y)) \quad = \\
P(X_i, a_i, X_{i-1}) \cdot (c(S(X_i), a_i, S(X_{i-1})) + w(X_{i-1})) + \\
\sum_{Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, a_i) s.t. Y \neq succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b} (P(X_i^{\tilde{X}+\underline{H}^u}, a_i, Y) \cdot (c(S(X_i), a_i, S(Y)) + w(Y)))
\end{aligned}
$$

Consider now $Y \in succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)$ such that $Y \neq succ(X_i^{\tilde{X}+\underline{H}^u}, a_i)^b$. Consider also $Z \in succ(X_i, a_i)$ such that $\underline{h}^{S(X_i),a_i}(Y) = \underline{h}^{S(X_i),a_i}(Z)$ (that is, Y and Z are corresponding outcomes). Then $Y = [S(Z); H(Z); \underline{H}^u(Z)] = Z^u$. Consequently, $w(Y) = w_{old}(Y) \geq w_{old}(Z) = w(Z)$ according to the assumptions of the theorem. As a result,

$$
\begin{aligned}
w(X_i) \quad &\geq \\
\sum_{Y \in succ(X_i, a_i)} (P(X_i, a_i, Y) \cdot (c(S(X_i), a_i, S(Y)) + w(Y))) \quad &= \\
E_{X' \in succ(X_i, besta(X_i))} (c(S(X_i), besta(X), S(X')) + w(X'))
\end{aligned}
$$

■

Theorem 4 *For each state X, it holds that $w^b(X) \leq g^*(X)$.*

Proof:

The case of $g^*(X) = \infty$ is trivial. We therefore assume that $g^*(X)$ is finite and prove by induction. Suppose there exist X such that $w^b(X) > g^*(X)$. It could not have been a state whose $S(X) = S_{\text{goal}}$ since according to the definitions of $w^b(X)$ and $g^*(X)$, they are both equal to 0. We therefore assume that $S(X) \neq S_{\text{goal}}$. Then

$$
w^b(X) \quad = \quad \min_{a \in S(X)} (c(S(X), a, succ(X, a)^b) + w^b(succ(X, a)^b))
$$

and,

$$
\begin{aligned}
g^*(X) \quad &= \quad \min_{a \in A(S(X))} Q_{f(Y = succ(X^u, a)^b) = g^*(Y), w}(X^u, a) \\
&= \quad \min_{a \in A(S(X))} \sum_{Y \in succ(X^u, a)} P(X^u, a, Y) \cdot \max(c(S(X^u), a, S(Y)) + w(Y), \\
&\qquad\qquad c(S(X^u), a, S(succ(X^u, a)^b)) + g^*(succ(X^u, a)^b)) \\
&\geq \quad \min_{a \in A(S(X))} c(S(X^u), a, S(succ(X^u, a)^b)) + g^*(succ(X^u, a)^b) \\
&\geq \quad \min_{a \in A(S(X))} c(S(X), a, S(succ(X, a)^b)) + g^*(succ(X, a)^b)
\end{aligned}
$$

.

The last line is due to the fact that $S(X) = S(X^u)$, $S(succ(X^u, a)^b) = S(succ(X, a)^b)$ and $g^*(succ(X^u, a)^b) = g^*(succ(X, a)^b)$ according to lemma 10. Let us consider a path $\pi = [\{X_n, a_n, X_{n-1}\}, \ldots, \{X_1, a_1, X_0\}]$, where $X_n = X$, $S(X_0) = S_{\text{goal}}$ and for every tuple $\{X_i, a_i, X_{i-1}\}$, $X_{i-1} = succ(X_i, a_i)^b$ and $a_i = \min_{a \in A(S(X_i))} c(S(X_i), a, succ(X_i, a)^b) + g^*(succ(X_i, a)^b)$. Since $g^*(X)$ is finite and all costs are bounded from below by a positive constant, the path π is finite.

Because $w^b(X_n) > g^*(X_n)$ and $w^b(X_0) = g^*(X_0) = 0$, there must be a tuple $\{X_i, a_i, X_{i-1}\} \in \pi$ such that $w^b(X_i) > g^*(X_i)$ whereas $w^b(X_{i-1}) \leq g^*(X_{i-1})$. But then we get the following contradiction

$$\begin{aligned} g^*(X_i) &= c(S(X_i), a_i, succ(X_i, a_i)^b) + g^*(succ(X_i, a_i)^b) \\ &\geq c(S(X_i), a_i, succ(X_i, a_i)^b) + w^b(succ(X_i, a_i)^b) \\ &\geq \min_{a \in A(S(X_i))} c(S(X_i), a, succ(X_i, a)^b) + w^b(succ(X_i, a)^b) \\ &= w^b(X_i) \end{aligned}$$

. ■

Theorem 5 *After each execution of the UpdateMDP function, for each state X it holds that $w_{old}(X) \leq w(X) \leq g^*_{old}(X) \leq g^*(X)$, where w_{old}-values are w-values before the execution of UpdateMDP, $g^*_{old}(X)$ are g^*-values under w_{old}-values and $g^*(X)$ are g^*-values under w-values.*

Proof:

First, let us show that before the first execution of UpdateMDP for every X it holds that $w(X) \leq g^*(X)$. It holds because according to the assumptions about state initialization before the main function is executed, $w(X) \leq w^b(X)$. On the other hand, according to theorem 4, $w^b(X) \leq g^*(X)$.

We now prove by induction. Suppose $w_{old}(X) \leq g^*_{old}(X)$ before the call to UpdateMDP. We need to show that after UpdateMDP function returns, for each state X we have $w_{old}(X) \leq w(X) \leq g^*(X)$.

Let us first prove that $w_{old}(X) \leq w(X)$. We only need to consider the states updated by UpdateMDP function since w-values of all other states remain unchanged. We first prove by induction on the execution of line 4 that for each state X updated by UpdateMDP it holds that $X^u = X^{\tilde{X}+\underline{H}^u}$. Consider the first time, line 4 is executed. Then $X = X_{\text{pivot}}$. Therefore, $X^u = [S(X); H(X); \underline{H}^u(X_{\text{pivot}})] = X^{\tilde{X}+\underline{H}^u}$. UpdateMDP also updates directly $[S(X); H(X); \underline{H}^u(X_{\text{pivot}})] = X^{\tilde{X}+\underline{H}^u}$. Now consider the ith execution of line 4, whereas on all previous executions it held that $X^u = X^{\tilde{X}+\underline{H}^u}$. At ith execution, state X is a state which is equal to some $succ(Y, besta(Y))^b$, where Y is a state that was updated during (i-1)th execution of line 4. Thus, $\underline{H}^u(X) = \underline{H}^u(Y)$ and therefore $X^u = [S(X); H(X); \underline{H}^u(X_{\text{pivot}})] = X^{\tilde{X}+\underline{H}^u}$. Once again, UpdateMDP also updates $[S(X); H(X); \underline{H}^u(X_{\text{pivot}})] = X^{\tilde{X}+\underline{H}^u}$.

Thus, for each state updated by UpdateMDP, it holds that $X^u = [S(X); H(X); \underline{H}^u(X_{\text{pivot}})]$. As a result, if $S(X) = S_{\text{goal}}$, then according to the definition of g-values, $g^*_{old}(\tilde{X}) = g^*_{old}(X) = 0$. On the other hand, if $S(X) \neq S_{\text{goal}}$, then

$$\begin{aligned} g^*_{old}(\tilde{X}) &= \min_{a \in A(S(\tilde{X}))} Q_{f(Y=succ(X^{\tilde{X}+\underline{H}^u},a)^b)=g^*(\tilde{Y}),w_{old}}(X^{\tilde{X}+\underline{H}^u}, a) \\ &= \min_{a \in A(S(X))} Q_{f(Y=succ(X^u,a)^b)=g^*(\tilde{Y}),w_{old}}(X^u, a) \end{aligned}$$

Because $g^*(\tilde{Y}) = g^*(Y) = 0$ if $S(Y) = s_{\text{goal}}$, it then holds that

$$\begin{aligned} g^*_{old}(\tilde{X}) &= \min_{a \in A(S(X))} Q_{f(Y=succ(X^u,a)^b)=g^*(Y),w_{old}}(X^u, a) \\ &= g^*_{old}(X) \end{aligned}$$

Thus, for each state X updated by UpdateMDP, it holds that $g^*_{old}(\tilde{X}) = g^*_{old}(X)$. Also, according to corollary 3, $g(\tilde{X}) = g^*_{old}(\tilde{X})$ and from induction assumption $w_{old}(X) \leq g^*_{old}(X)$. Thus when UpdateMDP executes $w(X) = g(\tilde{X})$ on line 4, then

$$w(X) = g(\tilde{X}) = g^*_{old}(\tilde{X}) = g^*_{old}(X) \geq w_{old}(X)$$

Suppose now UpdateMDP executes $w([S(X); H(X); \underline{H}^u(X_{\text{pivot}})]) = g(\tilde{X})$ on line 4. According to lemma 9 $g^*_{old}(X) = g^*_{old}(X^u) = g^*_{old}(X^{\tilde{X}+\underline{H}^u})$ and since $w_{old}(X^{\tilde{X}+\underline{H}^u}) \leq g^*_{old}(X^{\tilde{X}+\underline{H}^u})$, it follows that:

$$w(X^{\tilde{X}+\underline{H}^u}) = g(\tilde{X}) = g^*_{old}(\tilde{X}) = g^*_{old}(X) = g^*_{old}(X^{\tilde{X}+\underline{H}^u}) \geq w_{old}(X^{\tilde{X}+\underline{H}^u})$$

We now prove that for every state X $w(X) \leq g^*_{old}(X) \leq g^*(X)$. We first note that since as we have just proved none of w-values decreased, it holds that for every state X $g^*(X) \geq g^*_{old}(X)$.

Suppose X was not updated by UpdateMDP, that is, $w(X) = w_{old}(X)$. Then $w(X) = w_{old}(X) \leq g^*_{old}(X) \leq g^*(X)$.

Now suppose $w(X)$ was updated by UpdateMDP. Once again suppose the update is $w(X) = g(\tilde{X})$ on line 4. Then

$$w(X) = g(\tilde{X}) = g^*_{old}(\tilde{X}) = g^*_{old}(X) \leq g^*(X)$$

Now suppose the update is $w([S(X); H(X); \underline{H}^u(X_{\text{pivot}})]) = g(\tilde{X})$ on line 4. Then

$$w(X^{\tilde{X}+\underline{H}^u}) = g(\tilde{X}) = g^*_{old}(\tilde{X}) = g^*_{old}(X) = g^*_{old}(X^{\tilde{X}+\underline{H}^u}) \leq g^*(X^{\tilde{X}+\underline{H}^u})$$

■

Theorem 6 *For a non-negative function $w \leq w^u$, the following holds: for each state X, $g^*(X)$ is bounded from above by $w^u(X)$.*

Proof:

This certainly holds for X whose $S(X) = S_{\text{goal}}$. Now suppose $S(X) \neq S_{\text{goal}}$. We prove by contradiction and assume that there exists one or more states X whose $g^*(X) > w^u(X)$, which implies that $w^u(X)$ is finite. Let us consider a path $\pi_{\text{greedy},w^u,w^u}(X_n, X_0)$ where $X_n = X$ and $S(X_0) = s_{\text{goal}}$. According to its definition, for every pair of states X_i, X_{i-1} on this path it holds that $X_{i-1} = succ(X_i, a_i)^b$, where $a_i = \arg\min_{a' \in A(S(X_i))} Q_{w^u,w^u}(X_i^u, a')$. Since all costs are positive and $w^u(X)$ is finite, $w^u(X_i) > w^u(X_{i-1})$. Also, since the costs are bounded from below by a positive constant, $w^u(X_0) = 0$ and $w^u(X)$ is finite, it holds that the path $\pi_{\text{greedy},w^u,w^u}(X_n, X_0)$ is finite.

This means that there must exist such X_i on the path that $g^*(X_i) > w^u(X_i)$, while $g^*(X_{i-1}) \leq w^u(X_{i-1})$ where $X_{i-1} = succ(X_i, a_i)^b$ Then we arrive at the following contradiction

$$
\begin{aligned}
g^*(X_i) &\leq \\
Q_{f(X_{i-1}=succ(X_i^u,a_i)^b)=g^*(X_{i-1}),w^u}(X_i^u, a_i) &\leq \\
Q_{f(X_{i-1}=succ(X_i^u,a_i)^b)=g^*(X_{i-1}),w^u}(X_i^u, a_i) &= \\
\sum_{Y \in succ(X_i^u,a_i)} P(X_i^u, a_i, Y) \cdot \max(c(S(X_i), a_i, S(Y)) + w^u(Y), c(S(X_i), a_i, S(X_{i-1})) + g^*(X_{i-1})) &\leq \\
\sum_{Y \in succ(X_i^u,a_i)} P(X_i^u, a_i, Y) \cdot \max(c(S(X_i), a_i, S(Y)) + w^u(Y), c(S(X_i), a_i, S(X_{i-1})) + w^u(X_{i-1})) &= \\
Q_{w^u,w^u}(X_i^u, a_i) &= \\
\min_{a' \in A(S(X_i))} Q_{w^u,w^u}(X_i^u, a') &= \\
w^u(X) &
\end{aligned}
$$

■

Theorem 7 *For each state X, $w(X)$ is bounded from above by $w^u(X)$.*

Proof:

Before the first execution of UpdateMDP, according to the initialization assumptions, for every state X, $0 \leq w(X) \leq w^b(X)$. Also, according to theorem 4, for each state X, $w^b(X) \leq g^*(X)$ and according to theorem 6, $g^*(X) \leq w^u(X)$. Thus, $0 \leq w(X) \leq w^u(X)$.

We now prove the theorem by induction. Suppose before the ith execution of UpdateMDP, it holds that $0 \leq w_{old}(X) \leq w^u(X)$, where w_{old}-values are w-values right before the ith execution of UpdateMDP function. We need to show that after the ith execution of UpdateMDP function, the inequality $0 \leq w(X) \leq w^u(X)$ holds.

According to theorem 6, for every state X, $g^*_{old}(X) \leq w^u(X)$. At the same time, according to theorem 6, $w(X) \leq g^*_{old}(X)$. Thus, $w(X) \leq w^u(X)$.

The inequality $0 \leq w(X)$ follows from the fact that $0 \leq w_{old}(X)$ and theorem 6, according to which, $w_{old}(X) \leq w(X)$. ■

Theorem 8 *PPCP terminates, and at that time,* $w(X_{\text{start}}) \leq w^u(X_{\text{start}})$ *and the expected cost of the policy of always taking action* $besta(X)$ *at any state* X *starting at* X_{goal} *until state* X_0 *whose* $S(X_0) = S_{\text{goal}}$ *is reached is no more than* $w(X_{\text{start}})$.

Proof:

We will first show that the algorithm terminates.

For this let us first show that the set of all possible policies $\pi(X_{\text{start}})$ ever considered by PPCP is guaranteed to be finite. To prove this we need to show that any policy considered by PPCP is acyclic. Then, the fact that the set of all such policies is finite will be due to the belief state-space itself being finite. Any policy PPCP has at any point of time is acyclic because after each stochastic action a at any state X, the corresponding $h^{S(X),a}$ is set to a value not equal to u in the outcome states and remains such in all of their descendants, whereas all the ancestors of X and X itself had $h^{S(X),a} = u$. The deterministic paths in between any two stochastic actions on the policy or in between X_{start} and the first stochastic action on the policy, on the other hand, are all segments of the paths returned by ComputePath function and these paths are finite according to corollary 2.

Thus, the set of all possible policies considered by PPCP is finite. The termination criterion for the algorithm is that all states on its current policy are have w-values at least as large as the expectation over the action cost plus w-values of the successors of their action defined by $besta$ pointer, except for the goal states, whose w-values are 0 because they are bounded by w^u-values according to theorem 7 and these are zeroes for goal states. In other words, for every X on the current policy s.t. $S(X) \neq S_{\text{goal}}$ it holds that

$$w(X) \quad \geq \quad E_{X' \in succ(X, besta(X))}(c(S(X), besta(X), S(X')) + w(X')) \tag{7}$$

At each iteration, PPCP fixes at least one state X on the policy to satisfy this equation. While fixing the equation, PPCP may change $besta$ action for state X and/or change w-value of X. There is a finite number of possible subtrees below X that PPCP can consider since the set of all possible policies considered by PPCP is finite.

The change in the w-value of X may potentially affect other states, but since the policy is acyclic it can not affect the states that are descendants of X. The number of ancestors of X, on the other hand, is finite since the policy is acyclic and the belief state-space is finite. Therefore, PPCP is bound to arrive in a finite number of iterations at a policy for which all of the states that belong to it satisfy the equation 7. Each iteration is also guaranteed to be finite for the following reasons. First, the ComputePath function is guaranteed to return because each state is expanded no more than once per search according to theorem 2. Second, the UpdateMDP function is guaranteed to return, because the path it processes is guaranteed to be of finite length according to corollary 2.

We now show that after PPCP terminates, $w(X_{\text{start}}) \leq w^u(X_{\text{start}})$ and the expected cost of the policy of always taking action $besta(X)$ at any state X starting at

$X_{\rm goal}$ until state Y whose $S(Y) = S_{\rm goal}$ is reached is no more than $w(X_{\rm start})$. The first part comes directly from theorem 7. The second part can be proved as follows.

Consider the following potential function that we maintain while executing the policy defined by $besta$ actions starting with $X_{\rm start}$: $F(t) = costsofar(t) + w(X_t)$, where t is the current time-step. So, initially $F(t=0) = 0 + w(X_{\rm start}) = w(X_{\rm start})$. We execute the policy until we reach a state Y such that $S(Y) = S_{\rm goal}$. Suppose it happens at timestep $t = k$. That is, $Y = X_t$. Then $F(t=k) = costsofar(k) + 0 = costsofar(k)$. We need to show that the expected value of $F(t=k)$ is bounded above by $w(X_{\rm start})$.

Initially, $E\{F(t=0)\} = w(X_0)$, where $X_0 = X_{\rm start}$. Now consider the expectation at the ith step:

$$\begin{aligned} E\{F(t=i)\} &= E\{costsofar(i) + w(X_i)\} \\ &= E\{costsofar(i-1) + cost(i) + w(X_i)\} \\ &= E\{costsofar(i-1)\} + E\{cost(i) + w(X_i)\} \end{aligned}$$

Since all states on the policy (except for the goal states) satisfy the equation 7, we have $w(X_{i-1}) \geq E\{cost(i) + w(X_i)\}$. After taking an additional expectation we have $E\{w(X_{i-1}) - cost(i)\} \geq E\{w(X_i)\}$. Hence,

$$\begin{aligned} E\{F(t=i)\} &= E\{costsofar(i-1)\} + E\{cost(i) + w(X_i)\} \\ &\leq E\{costsofar(i-1)\} + E\{cost(i) + (w(X_{i-1}) - cost(i))\} \\ &= E\{costsofar(i-1) + w(X_{i-1})\} \\ &= E\{F(t=i-1)\} \end{aligned}$$

By induction then $E\{F(t=k)\} \leq E\{F(t=0)\} = w(X_{\rm start})$. ■

Theorem 9 *Suppose there exists a minimum expected cost policy ρ^* that satisfies the following condition: for every pair of states $X_1 \in \rho^*$ and $X_2 \in \rho^*$ such that X_2 can be reached with a non-zero probability from X_1 when following policy ρ^* it holds that either $h^{S(X_1),\rho^*(X_1)} \neq h^{S(X_2),\rho^*(X_2)}$ or $h^{S(X_1),\rho^*(X_1)} = h^{S(X_2),\rho^*(X_2)} = \mathbf{null}$. Then the policy defined by $besta$ pointers at the time PPCP terminates is also a minimum expected cost policy.*

Proof: Let us assume that there exists a minimum expected cost policy ρ^* that satisfies the conditions of the theorem. That is, for every pair of states $X_1 \in \rho^*$ and $X_2 \in \rho^*$ such that X_2 can be reached with a non-zero probability from X_1 when following policy ρ^* it holds that either $h^{S(X_1),\rho^*(X_1)} \neq h^{S(X_2),\rho^*(X_2)}$ or $h^{S(X_1),\rho^*(X_1)} = h^{S(X_2),\rho^*(X_2)} = \mathbf{null}$. Since ρ^* is an optimal policy, its expected cost is $w^*(X_{\rm start})$.

We will show that $w^u(X_{\rm start}) \leq w^*(X_{\rm start})$. This will prove the theorem since the expected cost of the policy returned by PPCP is bounded from above by $w^u(X_{\rm start})$. The expected cost the policy will then be exactly equal to $w^*(X_{\rm start})$ since ρ^* is already an optimal policy.

Let us prove by contradiction and assume that $w^u(X_{\text{start}}) > w^*(X_{\text{start}})$. This also means that $w^*(X_{\text{start}})$ is finite and therefore all branches on the policy ρ^* end up at states X whose $S(X) = S_{\text{goal}}$ since an optimal policy when sensing is perfect is acyclic.

Let us now pick a state $X \in \rho^*$ such that $w^u(X) > w^*(X)$, but all the successor states Y of action $\rho^*(X)$ executed at state X have $w^u(Y) \leq w^*(Y)$. Such state X must exist because at least for $X = X_{\text{start}}$ it holds that $w^u(X) > w^*(X)$ and all branches of the policy end up at states Y whose $S(Y) = S_{\text{goal}}$ and for these states $w^u(Y) = w^*(Y) = 0$ according to the definition of w^u and w^* values.

By definition,

$$\begin{aligned} w^u(X) &= \min_{a \in A(S(X))} Q_{w^u,w^u}(X^u, a) \\ &\leq Q_{w^u,w^u}(X^u, \rho^*(X)) \\ &= \sum_{Z \in succ(X^u,\rho^*(X))} P(X^u, a, Z) \cdot \max(c(S(X), a, S(Z)) + w^u(Z), \\ &\qquad c(S(X), a, S(Z)) + w^u(succ(X^u, \rho^*(X))^b)) \end{aligned}$$

Let us now consider $h^{S(X),\rho^*(X)}(X)$. It must be the case that either $h^{S(X),\rho^*(X)}(X) = u$ or $h^{S(X),\rho^*(X)}(X) = \mathbf{null}$ since, according to the assumptions of the theorem, no action whose outcome depends on $h^{S(X),\rho^*(X)}$ could have been executed before. Thus, $h^{S(X),\rho^*(X)}(X^u) = h^{S(X),\rho^*(X)}(X)$.

This property has an important implication that we will use. For any pair of states $Y \in succ(X, \rho^*(X))$ and $Z \in succ(X^u, \rho^*(X))$ such that $h^{S(X),\rho^*(X)}(Z^u) = h^{S(X),\rho^*(X)}(Y^u)$ (in other words, Y and Z are corresponding outcomes of action $\rho^*(X)$ executed at X and X^u respectively), it holds that $P(X, \rho^*(X), Y) = P(X^u, \rho^*(X), Z)$ and $Y^u = Z^u$.

Using this fact and lemma 8 we can derive the following.

$$\begin{aligned} w^u(X) &\leq \sum_{Z \in succ(X^u,\rho^*(X))} P(X^u, a, Z) \cdot \max(c(S(X), a, S(Z)) + w^u(Z), \\ &\qquad c(S(X), a, S(Z)) + w^u(succ(X^u, \rho^*(X))^b)) \\ &= \sum_{Z \in succ(X^u,\rho^*(X))} P(X^u, a, Z) \cdot \max(c(S(X), a, S(Z)) + w^u(Z^u), \\ &\qquad c(S(X), a, S(Z)) + w^u((succ(X^u, \rho^*(X))^b)^u)) \\ &= \sum_{Y \in succ(X,\rho^*(X))} P(X, a, Y) \cdot \max(c(S(X), a, S(Y)) + w^u(Y^u), \\ &\qquad c(S(X), a, S(Y)) + w^u((succ(X, \rho^*(X))^b)^u)) \\ &= \sum_{Y \in succ(X,\rho^*(X))} P(X, a, Y) \cdot \max(c(S(X), a, S(Y)) + w^u(Y), \end{aligned}$$

$$c(S(X), a, S(Y)) + w^u(succ(X, \rho^*(X))^b))$$

According to the way we picked X, $w^u(Y) \leq w^*(Y)$ for every $Y \in succ(X, \rho^*(X))$. Moreover, from the definition of clear preferences it follows that $c(S(X), a, S(Y)) + w^u(succ(X, \rho^*(X))^b) \leq c(S(X), a, S(Y)) + w^u(Y)$ for all $Y \in succ(X, \rho^*(X))$. Hence, we obtain the following contradiction

$$\begin{aligned} w^u(X) &\leq \sum_{Y \in succ(X,\rho^*(X))} P(X, a, Y) \cdot \max(c(S(X), a, S(Y)) + w^*(Y), \\ &\qquad\qquad c(S(X), a, S(Y)) + w^*(succ(X, \rho^*(X))^b)) \\ &= \sum_{Y \in succ(X,\rho^*(X))} P(X, a, Y) \cdot (c(S(X), a, S(Y)) + w^*(Y)) \\ &= w^*(X) \end{aligned}$$

■