Efficiently Computing Transitions in Cartesian Abstractions

Jendrik Seipp
Linköping University, Sweden
jendrik.seipp@liu.se

Abstract
Counterexample-guided Cartesian abstraction refinement yields strong heuristics for optimal classical planning. The approach iteratively finds a new abstract solution, checks where it fails for the original task and refines the abstraction to avoid the same failure in subsequent iterations. The main bottleneck of this refinement loop is the memory needed for storing all abstract transitions. To address this issue, we introduce an algorithm that efficiently computes abstract transitions on demand. This drastically reduces the memory consumption and allows us to solve tasks during the refinement loop and during the search that were previously out of reach.

Introduction
A common approach to solving classical planning tasks optimally is A∗ search (Hart, Nilsson, and Raphael 1968) with an admissible heuristic (e.g., Helmert and Domshlak 2009; Karpas and Domshlak 2009; Katz and Domshlak 2010; Pommerening et al. 2015; Sievers and Helmert 2021). Heuristics based on abstractions of the planning task have been particularly successful (e.g., Franco et al. 2017; Seipp 2019; Drexler, Seipp, and Speck 2021; Kreft et al. 2023).

Counterexample-guided abstraction refinement (CEGAR) is a prominent way of generating such abstractions (Clarke et al. 2003). Since the introduction of CEGAR for classical planning in the context of Cartesian abstractions (Ball, Podelski, and Rajamani 2001; Seipp and Helmert 2013), the method has also been adapted to pattern databases (PDBs; Culberson and Schaeffer 1998; Edelkamp 2001; Rovner, Sievers, and Helmert 2019) and domain abstractions (Hernádvölgyi and Holte 2000; Kreft et al. 2023). Furthermore, CEGAR has been used to create PDBs and Cartesian abstractions for probabilistic planning tasks (Klößner et al. 2022; Klößner, Seipp, and Steinmetz 2023).

CEGAR starts with a very coarse abstraction and then iteratively finds a cheapest abstract solution, checks where it fails for the original task and refines the abstraction to avoid the same flaw in subsequent iterations by splitting the state that caused the flaw into two new states. If CEGAR finds a solution for the original task during the refinement process, it is guaranteed to be optimal. Otherwise, the resulting abstraction can be used as a heuristic for an A∗ search.

Since domain abstractions and especially PDBs do not allow for fine-grained refinement, it is infeasible to solve non-trivial tasks while refining these types of abstractions. Therefore, existing approaches for these abstraction types mainly create collections of abstractions focusing on different aspects of the task (e.g., Haslum et al. 2007; Pommerening, Röger, and Helmert 2013; Franco et al. 2017; Seipp 2019).

Cartesian abstractions, however, allow for fine-grained refinements since each iteration only adds one additional state. Consequently, Cartesian CEGAR is able to solve large tasks during the refinement loop. Previously, the main bottlenecks of the refinement loop in the classical planning setting were the times for finding the next cheapest solution and the next flaw in it, but these two bottlenecks have been addressed recently by incrementally revising all cheapest paths (Seipp, von Allmen, and Helmert 2020) and by finding and addressing batches of flaws (Speck and Seipp 2022).

Now, the main bottleneck of the refinement loop is the memory needed for storing the abstract transitions. It is well known that storing abstract transitions, not abstract states, is the limiting factor for abstractions. In Cartesian abstractions the problem is especially severe since we need access to both the incoming and outgoing transitions of a state in order to efficiently rewire the transition system after a refinement step. Merge-and-shrink abstractions address the problem by using label reduction (Sievers and Helmert 2021). PDBs and domain abstractions circumvent the issue by computing abstract transitions on demand (Rovner, Sievers, and Helmert 2019; Kreft et al. 2023). To do this efficiently, they use perfect hashing (Sievers, Ortlieb, and Helmert 2012) and the successor generator data structure (Helmert 2006).

Cartesian abstractions are too general to allow for perfect hashing. However, they are specific enough that a successor generator can efficiently compute the operators o applicable in abstract state a. To efficiently compute which abstract states b can be reached from a by applying o, we turn to the abstraction’s refinement hierarchy, which records all splits during the refinement loop in a tree data structure.

In our experiments, we show that computing transitions on demand drastically reduces the memory footprint and thus increases the number of tasks solved during the refinement loop. For the remaining tasks, we obtain much better heuristic estimates than before and consequently solve many additional tasks in the ensuing A∗ search.
Background

A SAS* planning task (Bäckström and Nebel 1995) is a tuple \( \Pi = (V, O, s_0, g) \), where \( V = \{ v_1, \ldots, v_n \} \) is a finite set of state variables, each with an associated finite domain \( dom(v_i) \). An atom is a pair \( \langle v, d \rangle \) with \( v \in V \) and \( d \in dom(v) \). A partial state \( s \) maps a subset \( \mathcal{V}(s) \) of \( V \) to values \( s[v] \in dom(v) \) for \( v \in \mathcal{V}(s) \). If \( \mathcal{V}(s) = V \), we call \( s \) a state. The set of all states in \( \Pi \) is \( S(\Pi) \). We often treat partial states as sets of atoms. Updating partial state \( p \) with partial state \( q \) results in partial state \( r = p \oplus q \), with \( r[v] = q[v] \) for all \( v \in \mathcal{V}(q) \), and \( r[v] = p[v] \) for all \( v \in \mathcal{V}(p) \setminus \mathcal{V}(q) \).

Each operator \( o \in O \) is a pair \( \langle pre(o), eff(o) \rangle \), where \( pre(o) \) and \( eff(o) \) are partial states specifying the precondition and effect of \( o \). The postcondition of \( o \) is \( post(o) = pre(o) \oplus eff(o) \). Operator \( o \) is applicable in state \( s \) if \( pre(o) \subseteq s \) and applying \( o \) in \( s \) results in state \( s \oplus eff(o) \). The cost of \( o \) is \( cost(o) \in \mathbb{R}_+^+ \). The initial state \( s_0 \) is a partial state. Solving \( \Pi \) optimally implies finding a cheapest iteratively-applicable sequence of operators that transforms \( s_0 \) into a state \( s \) with \( s \subseteq s \).

A task \( \Pi \) induces a transition system \( T \) which is a directed, labeled graph with states \( S(T) = S(\Pi) \), labels \( L(T) = O \), transitions \( T(T) = \{ s \xrightarrow{o} s' \mid o \in O, s \in S(T), pre(o) \subseteq s \} \), initial state \( s_0(T) = s_0 \) and goal states \( s_f(T) = \{ s \mid s \in S(T), s \subseteq s_f \} \).

An abstraction \( \approx \) of \( T \) is an equivalence relation over \( S(T) \) (Seipp and Helmhert 2018). It induces an abstract transition system \( T' \) with states \( S(T') = \{ \lfloor s \rfloor \mid s \in S(T) \} \), labels \( L(T') = L(T) \), transitions \( T'(T') = \{ \lfloor s \rfloor \xrightarrow{o} \lfloor s' \rfloor \mid o \in O, s \in S(T), pre(o) \subseteq s \} \), initial state \( s_0(T') = \lfloor s_0 \rfloor \) and goal states \( \{ \lfloor s \rfloor \mid s \in S_f(T) \} \). An abstract state \( s \) is Cartesian if it has the form \( A_1 \times \ldots \times A_n \), where \( A_i = dom(v_i, a) \subseteq dom(v_i) \) for all \( 1 \leq i \leq |V| \). An abstraction is Cartesian if all its states are Cartesian. A partial state \( p \) induces the Cartesian set \( \mathcal{C}(p) = A_1 \times \ldots \times A_n \), with \( A_i = \lfloor p[v_i] \rfloor \) if \( v_i \in \mathcal{V}(p) \) and \( A_i = dom(v_i) \) otherwise.

The intersection of two Cartesian sets \( a = A_1 \times \ldots \times A_n \) and \( b = B_1 \times \ldots \times B_n \) is \( ab = (A_1 \cap B_1) \times \ldots \times (A_n \cap B_n) \).

The projection of a Cartesian set \( b \) over operator \( o \in O \) is \( regr(b, o) = A_1 \times \ldots \times A_n \) with
\[
A_i = \begin{cases} B_i & \text{if } v_i \notin \mathcal{V}(post(o)) \\ \emptyset & \text{if } v_i \in \mathcal{V}(post(o)) \text{ and } post(o)[v_i] \notin B_i \\ \{ pre(o)[v_i] \} & \text{if } v_i \in \mathcal{V}(pre(o)) \text{ and } post(o)[v_i] \in B_i \\ dom(v_i) & \text{otherwise} \end{cases}
\]

Similarly, the progression of a Cartesian set \( a = A_1 \times \ldots \times A_n \) over operator \( o \in O \) is \( progr(a, o) = B_1 \times \ldots \times B_n \) with
\[
B_i = \begin{cases} A_i & \text{if } v_i \notin \mathcal{V}(post(o)) \\ \emptyset & \text{if } v_i \in \mathcal{V}(pre(o)) \text{ and } pre(o)[v_i] \notin A_i \\ \{ post(o)[v_i] \} & \text{otherwise} \end{cases}
\]

Efficiently Computing Transitions

For Cartesian CEGAR, we need to efficiently obtain both the incoming and outgoing transitions of a given abstract state.\(^1\)

Traditionally, this has been done by storing all transitions and rewiring them after each refinement step. Since this approach often quickly consumes huge amounts of memory, we now present an approach that computes the transitions on demand. We begin by stating under which conditions a Cartesian abstraction contains a given transition.

Proposition 1. Cartesian abstraction \( T' \) contains a transition from state \( a \in S(T') \) to state \( b \in S(T') \) via operator \( o \in O \) if \( a \cap C(pre(o)) \neq \emptyset \) and \( progr(o, a) \cap b \neq \emptyset \).

Proof. Let \( [x] \) abbreviate \( x \neq \emptyset \). Then \( a \xrightarrow{o} b \in T(T') \)
\[\exists s : s \in a \land pre(o) \subseteq s \land s \oplus post(o) \in b \]
\[\exists s : C(s) \land C(pre(o)) \cap C(s) \land C(s \oplus post(o)) \cap b \]
\[C(pre(o)) \cap a \neq \emptyset \land progr(o, a) \cap b \neq \emptyset \]

Step 1 follows from the definition of abstract transition systems. Step 2 converts all (partial) concrete states into Cartesian sets. Step 3 uses the definition of Cartesian sets and progression. \( \square \)

Note that \( progr(o, a) \cap b \neq \emptyset \) implies \( a \cap C(pre(o)) \neq \emptyset \). By making the two conditions explicit, it becomes apparent that we can divide the task of computing the outgoing transitions of an abstract state \( a \) into two steps: first we compute the set of operators \( o \) that are applicable in \( a \), then we compute the set of abstract states \( b \) reachable from \( a \) via transitions labeled with \( o \).

Outgoing Operators

The set \( C_{out}(a) \) of operators applicable in abstract state \( a \) is \( \{ o \in O \mid a \cap C(pre(o)) \neq \emptyset \} \). The naive way of computing \( C_{out}(a) \) is to iterate over \( O \) and checking each operator for applicability. We can improve over this computation by exploiting the fact that \( a \) is a Cartesian set, which allows us to feed it into the successor generator data structure, developed for efficiently enumerating all operators applicable in a given concrete state (Helmert 2006).

In its original form, the successor generator is a tree data structure, where each internal node \( n \) branches over a subset of the values \( d \in dom(n.var) \) of a variable \( n.var \). Additionally, internal nodes have a child node for the “don’t care” value \( T \). When querying a successor generator for a given concrete state \( s \), at each node \( n \) we follow the “don’t care” child and the child for \( s[n.var] \) if it is defined. The traversal stops at the leaf nodes, which store the sets of applicable operators. A successor generator only needs space \( O(|\sum_{o \in O} |pre(o)| + eff(o)|) \) and querying it is usually sub-linear in the number of operators, and in the best case only linear in the number of applicable operators (Sievers, Ortlieb, and Helmhert 2012).

Since our abstract states are Cartesian, we can reuse the successor generator data structure with minimal adaptation. We can construct the successor generator in exactly the same way as for concrete states. Only the querying needs to be altered: instead of testing only the single value \( s[n.var] \) in search, which requires access to both incoming and outgoing transitions (Seipp, von Allmen, and Helmhert 2020).

\(^1\)While a plain forward search would only need outgoing transitions, it is much faster to find cheapest paths with incremental
We know from Proposition 1 that the set of abstract states over operator postconditions instead of preconditions. This set by looping over all operators, we use the successor set of operators that can reach an abstract state definitions of Cartesian progression and regression.

Proof sketch. Then prog

A Cartesian state $a$ is a Cartesian set of concrete states $S$. We can compute the set of operators applicable in at least one $s$ in $S$ by looping over $S$, querying the successor generator for $s$ and collecting all reported operators. Algorithm 1 interleaves these traversals by considering all states $s$ in $S$ at the same time.

Proof sketch. By case distinction over the conditions in the definitions of Cartesian progression and regression.

Since $\text{regr}(b, o) \cap a \neq \emptyset$ implies $C(\text{post}(o)) \cap b \neq \emptyset$, the set of operators that can reach an abstract state $b$ is $\mathcal{O}_m(b) = \{ o \in \mathcal{O} | C(\text{post}(o)) \cap b \neq \emptyset \}$. Again, instead of computing this set by looping over all operators, we use the successor generator data structure. This time, however, we let it branch over operator postconditions instead of preconditions.

## Outgoing Transitions

We know from Proposition 1 that the set of abstract states $b$ that can be reached from $a$ via an operator $o \in \mathcal{O}(a)$ is $\mathcal{T}(a, o) = \{ a \xrightarrow{o} b \ | \ b \in S(T'), \text{prog}(a, o) \cap b \neq \emptyset \}$. The naive computation of this set loops over all states $b \in S(T')$ and checks whether $\text{prog}(a, o)$ overlaps with $b$. Since each iteration of the refinement loop adds another abstract state, this computation will run slower and slower over time.

To compute $\mathcal{T}(a, o)$ efficiently, we turn to another tree data structure, the refinement hierarchy, which holds a record of all refinements (Seipp and Helmert 2018). Each

\^2To simplify the presentation, we assume that each refinement splits off a single atom. To account for splitting off multiple atoms, our implementation uses a directed acyclic graph instead of a tree.

node in this binary tree represents a Cartesian set and the leaf nodes are the abstract states in the current abstraction. Each non-leaf node $n$ holds the variable $n$.var for which the associated Cartesian set was split and pointers to the two resulting child nodes $n.left$ and $n.right$.

Algorithm 2 shows the $\text{INTERSECT}$ function which uses the refinement hierarchy with root node $n$ to compute the set of abstract states that intersect with a given Cartesian set $c$. We use the function to obtain $\mathcal{T}(a, o)$ as $\text{INTERSECT}(\text{prog}(a, o), n)$.

## Proposition 4

For Cartesian set $c$ and root node $n$ of a refinement hierarchy for abstraction $T'$, $\text{INTERSECT}(c, n)$ computes the set of abstract states in $T'$ that overlap with $c$.

Proof sketch. When intersecting two Cartesian sets, we can consider each variable independently of the others. $\text{INTERSECT}$ uses this to compute the overlapping states recursively, at each node $n$ checking for which of the children the intersection for the split variable $n$.var is non-empty.

Even though we need to follow at least one child node at each internal node, the fact that the depth of the refinement hierarchy is bounded by the number $N$ of atoms in $\Pi$ makes $\text{INTERSECT}$ an appealing alternative to looping over all $O(2^N)$ states in the abstraction.

## Incoming Transitions

Proposition 3 shows that the transitions induced by operator $o$ that lead into state $b$ are $\mathcal{T}_n(a, o) = \{ a \xrightarrow{o} b \ | \ a \in S(T'), \text{regr}(b, o) \cap a \neq \emptyset \}$. To compute this set efficiently, we call $\text{INTERSECT}(\text{regr}(b, o), n)$.

## Caching Optimal Transitions

There is a middle ground between storing all transitions and storing no transitions: we can store only optimal transitions. A transition $a \xrightarrow{o} b$ is optimal iff $h_{T'}(a) = \text{cost}(a) + h_{T'}(b)$, where $h_{T'}(x)$ is the cost of a cheapest path from $x$ to a goal state in $S_*(T')$. The CEGAR algorithm uses incremental search (Seipp, von Allmen, and Helmert 2020) to maintain for each state $a$ a transition $a \xrightarrow{o} b$ that starts a cheapest path from $a$. In several places of the algorithm the incremental search only needs access to the optimal transitions, so by caching them, we can often avoid computing all transitions.
to a ten-fold speedup for some commonly solved tasks, but the refinement loop slows down drastically, and transitions naively on demand (NAIVE) we never run out of memory, but the refinement loop slows down drastically, leading to solving only 352 tasks during refinement. Using successor generators for computing operators (SG) incurs up to a ten-fold speedup for some commonly solved tasks, but this only translates to solving one extra task during refinement (353 tasks in total). In contrast, computing transitions using the refinement hierarchy (RH), while computing operators naively, leads to solving 582 tasks during refinement, a 65% increase over SG. Using the tree data structures for both computations (SGRH) leads to solving 600 tasks during refinement, while still almost never running out of memory. Finally, caching all optimal transitions (SGRHC) hits the sweet spot between memory usage and runtime and solves 637 tasks during refinement, 57 tasks more than STORE.

Figure 1 compares our strongest variants, SGRH and SGRHC, to STORE in terms of runtime and memory consumption during the refinement loop. The plots visualize the time vs. memory trade-off: while SGRH is slightly slower than STORE, it uses much less memory. SGRHC uses more memory than SGRH but still less memory than STORE for most tasks. As a result, SGRHC is roughly as fast as STORE.

Regarding heuristic accuracy, Table 1 shows that all algorithm variants suffer from diminishing returns: solving additional tasks during the refinement becomes harder and harder and all variants benefit from switching from the refinement loop to an A* search eventually. We also see that all resulting heuristics are so fast to evaluate that runtime almost never becomes a bottleneck. Our strongest algorithm variants solve more tasks overall (up to 848 tasks) than the previous state of the art (STORE: 835 tasks). This is the case not only since more tasks are solved during refinement, but also since the resulting heuristics are more accurate. SGRHC computes a higher lower bound than STORE for 633 tasks, while the opposite is only true for 152 tasks. Also, SGRHC needs fewer expansions than STORE until the last f layer for 217 tasks, while the opposite only holds for 15 tasks.

**Conclusions**

Our algorithms for efficiently computing transitions in Cartesian abstractions drastically reduce the memory usage during the refinement loop, while only slowing it down slightly. If we store all optimal transitions, we can trade a bit of memory for faster runtime and solve even more tasks.

In future work, we want to evaluate whether the benefits of our algorithms for single abstractions carry over to the setting where we compute multiple Cartesian abstractions.

**Experiments**

We implemented our algorithms in the Scorpion planning system, which is an extension of Fast Downward (Helmert 2006) and used the Downward Lab toolkit (Seipp et al. 2017) for running experiments. Our benchmark set consists of all 1827 tasks without conditional effects from the optimal sequential tracks of the International Planning Competitions 1998–2018. We limit runtime to 30 minutes and memory to 4 GiB. When the refinement loop exhausts the internal time limit of 20 minutes or tries to use more than 3.5 GiB of memory, we stop refining and use the resulting heuristic in an A* search. All benchmarks, code and experiment data are available online (Seipp 2024).

We compare the previous state of the art (STORE) to five variants of our algorithms in an ablation study. Before inspecting heuristic quality, we evaluate the effects on the refinement loop, which is our main focus. Table 1 shows that if we store all transitions in memory (STORE), we solve 580 tasks during refinement, but run out of memory for the vast majority of the remaining tasks. By computing all operators and transitions naively on demand (NAIVE) we never run out of memory, but the refinement loop slows down drastically, leading to solving only 352 tasks during refinement. Using successor generators for computing operators (SG) incurs up to a ten-fold speedup for some commonly solved tasks, but...
Acknowledgments

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation and by TAILOR, a project funded by the EU Horizon 2020 research and innovation programme under grant agreement no. 952215. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973. We thank Daniel Gnadt and Malte Helmert for their helpful comments on a draft of this paper.

References


