arXiv:1710.05061v2 [cs.FL] 17 Oct 2017

A New Technique for Reachability of States in Concatenation Automata

Sylvie Davies

University of Waterloo
Department of Pure Mathematics
sldavies@uwaterloo.ca

Abstract

We present a new technique for demonstrating the reachability of states in deterministic finite automata representing the concatenation of two languages. Such demonstrations are a necessary step in establishing the state complexity of the concatenation of two languages, and thus in establishing the state complexity of concatenation as an operation. Typically, ad-hoc induction arguments are used to show particular states are reachable in concatenation automata. We prove some results that seem to capture the essence of many of these induction arguments. Using these results, reachability proofs in concatenation automata can often be done more simply and without using induction directly.

1 Introduction

Formal definitions are postponed until Section 2.

The *state complexity* of a regular language L, denoted $\mathrm{sc}(L)$, is the least number of states needed to recognize the language with a deterministic finite automaton. The *state complexity of an operation* on regular languages is the worst-case state complexity of the result of the operation, expressed as a function of the maximal allowed state complexity of the input languages. For example, suppose L is a language of state complexity at most m, and K is a language of state complexity at most n. It is known that the intersection $L \cap K$ has state complexity at most mn, and that this upper bound can be attained. Thus, we say that the *state complexity of intersection* is the function $(m, n) \mapsto mn$.

To establish the state complexity of an operation, there are two steps. First, one derives an upper bound. For example, in the case of intersection, if the input languages L and K have state complexity at most m and at most n respectively, then the standard direct product construction gives an automaton for $L \cap K$ with mn states, leading to the aforementioned upper bound of mn. Next, one searches for witnesses to the upper bound, that is, languages which attain the upper bound for each value of m and n. In the case of intersection,

this means for each pair (m, n), one must find a pair of languages (L_m, K_n) with $\mathrm{sc}(L_m) \leq m$ and $\mathrm{sc}(K_n) \leq n$ such that $\mathrm{sc}(L_m \cap K_n) = mn$.

One must not only find these witnesses but also *prove* that the desired state complexity bound is reached. Such proofs are the subject of this paper. We are interested in the case where the operation is *concatenation* of languages. We assume that one is working within some subclass of the regular languages, and has derived an upper bound $f(m, n)$ for the worst-case state complexity of concatenation within this subclasses. We also assume one has found (by computer search or some other means) candidate witnesses for this upper bound, in the form of two sequences of languages $(L_m : m \geq 1)$ and $(K_n : n \geq 1)$ such that $\mathrm{sc}(L_m) \leq m$ and $\mathrm{sc}(K_n) \leq n$. The goal is to prove that for each pair (m, n), the concatenation $L_m K_n$ has state complexity $f(m, n)$. We may divide such a proof into three steps:

1. Construct an automaton $\mathcal{A}$ for $L_m K_n$ in the standard way.
2. Show that $\mathcal{A}$ contains at least $f(m, n)$ reachable states.
3. Show exactly $f(m, n)$ reachable states in $\mathcal{A}$ are pairwise distinguishable.

We present a new technique for dealing with step (2) of this process. The standard way to construct a deterministic finite automaton $\mathcal{A}$ for the concatenation of two languages yields an automaton in which the states are *sets*; to show a particular set is reachable, one typically proceeds by induction on the size of the set. We prove a result that seems to generalize many of these ad-hoc induction arguments, and can be used to establish reachability of sets without directly using induction. Additionally, we prove some helpful lemmas that make our main result easier to apply.

We demonstrate our technique by applying it to a variety of concatenation witnesses taken from the literature. The state complexity of concatenation has been studied in the class of all regular languages, as well as many subclasses. Table 1 lists some examples of subclasses that have been studied, and the state complexity of concatenation in each subclass. See the cited papers for definitions of each subclass and derivations/proofs of each complexity. The complexities listed are "restricted" complexities, that is, they are computed under the assumption that both inputs to the concatenation operation share the same alphabet. "Unrestricted" state complexity of concatenation (where the inputs may be languages over different alphabets) has also been studied, and will be discussed later, but is not included in the table.

If the state complexity of concatenation grows exponentially with n (indicated in Table 1 by **bold** type), it is typical to use an induction argument to prove the desired number of states is reachable. It is cases like this in which our technique is most likely to be useful. We selected 16 concatenation witnesses, all from subclasses in which the state complexity of concatenation is exponential in n, and tried to apply our technique to these witnesses. In many cases we were able to produce shorter and simpler proofs than the original authors, and we only found two cases in which our technique did not work or was not useful.

Subclass	Complexity	Subclass	Complexity
Regular [1, 9, 15, 19]	$\mathbf{(m-1)2^n + 2^{n-1}}$	Prefix-closed [6, 9]	$\mathbf{(m+1)2^{n-2}}$
Unary [16, 17, 19]	$\sim mn$ (asymptotically)	Prefix-free [9, 13, 14]	$m+n-2$
Finite unary [10, 18]	$m+n-2$	Suffix-closed [6, 8]	$mn-n+1$
Finite binary [10]	$\mathbf{(m-n+3)2^{n-2}-1}$	Suffix-free [8, 12]	$\mathbf{(m-1)2^{n-2}+1}$
Star-free [7]	$\mathbf{(m-1)2^n + 2^{n-1}}$	Right ideal [4, 5, 9]	$\mathbf{m+2^{n-2}}$
Non-returning [3, 11]	$\mathbf{(m-1)2^{n-1}+1}$	Left ideal [4, 5, 8]	$m+n-1$

Table 1: Subclasses of regular languages and the state complexity of the concatenation operation within each subclass. **Bold** type indicates that the complexity grows exponentially in terms of n.

This suggests that our technique is widely applicable and should be considered as an viable alternative to the traditional induction argument when attempting reachability proofs in concatenation automata.

The rest of the paper is structured as follows. Section 2 contains background material and definitions needed to understand the paper. Section 3 describes our new technique and proves the relevant results. Section 4 contains examples of our technique applied to numerous concatenation witnesses from the literature. Section 5 concludes the paper.

2 Preliminaries

2.1 Relations and Functions

A *binary relation* ρ between X and Y is a subset of $X \times Y$. If $\rho \subseteq X \times Y$ and $\tau \subseteq Y \times Z$, the *composition* of ρ and τ is the relation

$$\rho\tau = \{(x,z) \in X \times Z : \text{ there exists } y \in Y \text{ such that } (x,y) \in \rho \text{ and } (y,z) \in \tau\}.$$

For $x \in X$ and $\rho \subseteq X \times Y$, the *image* of x under ρ is the set $x\rho = \{y \in Y : (x,y) \in \rho\}$. For $x \notin X$ we define $x\rho = \emptyset$. The *converse* of a binary relation $\rho \subseteq X \times Y$ is the relation $\rho^{-1} = \{(y,x) : (x,y) \in \rho\} \subseteq Y \times X$. The set $y\rho^{-1} = \{x \in X : (x,y) \in \rho\}$ is called the *preimage* of y under ρ. Elements of this set are callled *preimages* of y; for example, if $x \in y\rho^{-1}$ we say that x is a preimage of y.

If we write $\mathcal{P}(S)$ for the power set of a set S (that is, the set of all subsets of S), then we can view ρ as a map $\rho\colon X \to \mathcal{P}(Y)$. We may also *extend ρ by union* to a map $\rho\colon \mathcal{P}(X) \to \mathcal{P}(Y)$ as follows: for $S \subseteq X$, we define

$$S\rho = \bigcup_{s \in S} s\rho.$$

We thus have two ways to make sense of an expression like $x\rho\tau$: it is the image of x under the composite relation $\rho\tau \subseteq X \times Z$, and it is also the image of the set $x\rho \subseteq Y$ under the map $\tau\colon \mathcal{P}(Y) \to \mathcal{P}(Z)$. Additionally, we have a way to make sense of a composition $\rho\tau\colon X \to \mathcal{P}(Z)$ of maps $\rho\colon X \to \mathcal{P}(Y)$ and $\tau\colon Y \to \mathcal{P}(Z)$: take the composition of the corresponding relations.

A *function* $f\colon X \to Y$ is a binary relation $f \subseteq X \times Y$ such that $|xf| = 1$ for all $x \in X$. Following our notation for binary relations, we write functions to the *right* of their arguments. Composition of functions is defined by composing the corresponding relations. Thus the order of composition is *left-to-right*; in a composition fg, first f is applied and then g.

A *transformation* of a set X is a function $t\colon X \to X$, that is, a function from X into itself. We say t is a *permutation* of X if $Xt = X$. We say t *acts as a permutation* on $S \subseteq X$ if $St = S$. If t acts as a permutation on S, then every element of S has at least one preimage under t, that is, for all $s \in S$, the set $st^{-1} = \{x \in X : xt = s\}$ is non-empty.

A *cyclic permutation* of a set $\{x_1, \ldots, x_k\} \subseteq X$ is a permutation p such that $x_i p = x_{i+1}$ for $1 \leq i < k$, $x_k p = x_1$, and $xp = x$ for all $x \in X \setminus \{x_1, \ldots, x_k\}$. We denote such a permutation as $(x_1, \ldots, x_k)$. A cyclic permutation of a two-element set is called a *transposition*. The identity transformation is denoted id.

The notation $(S \to x)$ for $S \subseteq X$ and $x \in X$ denotes a transformation that sends every element of S to x and fixes every element of $S \setminus X$. For example, $(\{i\} \to j)$ denotes a transformation that maps i to j and fixes everything else. The transformation $(X \to x)$ is a constant transformation that maps every element of X to x.

In the case where $X = \{1, 2, \ldots, n\}$, the notation $({}_i^j x \to x + 1)$ denotes a transformation such that for each x with $i \leq x \leq j$, the transformation sends x to $x + 1$, and every other x is fixed. For example, the transformation $({}_2^{n-1} x \to x + 1)$ fixes 1, sends x to $x + 1$ for $2 \leq x \leq n - 1$, and fixes n. The notation $({}_i^j x \to x - 1)$ is defined similarly.

2.2 Automata

A *finite automaton* (FA) is a tuple $\mathcal{A} = (Q, \Sigma, T, I, F)$ where Q is a finite set of *states*, Σ is a finite set of *letters* called an *alphabet*, $T \subseteq Q \times \Sigma \times Q$ is a set of *transitions*, $I \subseteq Q$ is a set of *initial states*, and $F \subseteq Q$ is a set of *final states*.

We now define a binary relation $T_w \subseteq Q \times Q$ for each $w \in \Sigma^*$. Define $T_\varepsilon = \{(q, q) : q \in Q\}$; in terms of maps, this is the identity map on Q. For $a \in \Sigma$, define $T_a = \{(p, q) \in Q \times Q : (p, a, q) \in T\}$. For $w = a_1 \cdots a_k$ with $a_1, \ldots, a_k \in \Sigma$, define $T_w = T_{a_1} \cdots T_{a_k}$. The relation T_w is called the *relation induced by* w. The set $\{T_w : w \in \Sigma^*\}$ is a monoid under composition, called the *transition monoid* of $\mathcal{A}$. For technical reasons, if w is a word but is *not* a word over Σ, we define T_w to be the empty relation: $qT_w = \emptyset$ for all $q \in Q$.

Observe that the set of transitions of an FA is determined by the relations T_a. Furthermore, each relation T_a is determined by the set of images qT_a, where $q \in Q$. Hence we often define the transitions of an FA by specifying the images qT_a for each $q \in Q$ and $a \in \Sigma$.

If $\mathcal{A} = (Q, \Sigma, T, I, F)$ is a finite automaton such that $|I| = 1$ and T_a is a *function* for each $a \in \Sigma$, we say $\mathcal{A}$ is *deterministic*. We abbreviate "deterministic finite automaton" to DFA.

Let $\mathcal{A} = (Q, \Sigma, T, I, F)$ be an FA. A word $w \in \Sigma^*$ is *accepted* by $\mathcal{A}$ if we have $IT_w \cap F \neq \emptyset$. If $\mathcal{A}$ is a DFA with $I = \{i\}$, this condition becomes $iT_w \in F$. The *language* of $\mathcal{A}$, denoted $L(\mathcal{A})$, is the set of all words it accepts. Languages of FAs are called *regular languages*. A sequence of transitions $(q_0, a_1, q_1), (q_1, a_2, q_2) \ldots, (q_{k-1}, a_k, q_k)$ with $w = a_1 \cdots a_k$ is called a *path* from q_0 to q_k with label w, and the path is *accepting* if $q_0 \in I$ and $q_k \in F$. The FA $\mathcal{A}$ accepts a word w if and only if there is an accepting path with label w. We write $p \xrightarrow{w} q$ to mean that there is a path from p to q with label w.

Given two regular languages L and K with DFAs $\mathcal{A} = (Q^{\mathcal{A}}, \Sigma^{\mathcal{A}}, T^{\mathcal{A}}, i^{\mathcal{A}}, F^{\mathcal{A}})$ and $\mathcal{B} = (Q^{\mathcal{B}}, \Sigma^{\mathcal{B}}, T^{\mathcal{B}}, i^{\mathcal{B}}, F^{\mathcal{B}})$, we may construct an FA $\mathcal{AB} = (Q, \Sigma, T, I, F)$ that accepts the concatenation LK as follows:

- $Q = Q^{\mathcal{A}} \cup Q^{\mathcal{B}}$. We assume without loss of generality that $Q^{\mathcal{A}} \cap Q^{\mathcal{B}} = \emptyset$.
- $\Sigma = \Sigma^{\mathcal{A}} \cup \Sigma^{\mathcal{B}}$.
- $T = T^{\mathcal{A}} \cup T^{\mathcal{B}} \cup \{(q, a, i^{\mathcal{B}}) : qT_a^{\mathcal{A}} \in F^{\mathcal{A}}, a \in \Sigma^{\mathcal{A}}\}$.
- $I = \{i^{\mathcal{A}}\}$ if $i^{\mathcal{A}} \notin F^{\mathcal{A}}$, and otherwise let $I = \{i^{\mathcal{A}}, i^{\mathcal{B}}\}$.
- $F = F^{\mathcal{B}}$.

Proposition 1. *The FA $\mathcal{AB}$ accepts the concatenation LK.*

Proof. Suppose $w \in LK$; we want to show that w is accepted by $\mathcal{AB}$. We can write $w = uv$ with $u \in L$ and $v \in K$. There are two cases: u can be empty or non-empty. If $u = \varepsilon$ then $\varepsilon \in L$, so $i^{\mathcal{A}} \in F^{\mathcal{A}}$. Thus $I = \{i^{\mathcal{A}}, i^{\mathcal{B}}\}$. It follows that $IT_v \supseteq i^{\mathcal{B}}T_v \ni i^{\mathcal{B}}T_v^{\mathcal{B}}$, which is final since $v \in K$. If $u \neq \varepsilon$ we can write $u = xa$ for some word x and letter a. Then $IT_{xa} \supseteq i^{\mathcal{A}}T_{xa}^{\mathcal{A}}$, which contains an element of $F^{\mathcal{A}}$ since $xa \in L$. It follows that $IT_x \supseteq i^{\mathcal{A}}T_x^{\mathcal{A}}$ contains some state q such that $qT_a^{\mathcal{A}} \in F^{\mathcal{A}}$. Thus $IT_{xa} \supseteq qT_a \ni i^{\mathcal{B}}$, and so $IT_{xav} = IT_w \ni i^{\mathcal{B}}T_v^{\mathcal{B}}$, which is final since $v \in K$.

Conversely, let w be accepted by $\mathcal{AB}$. Choose an accepting path for w. There are two cases: either this accepting path starts from $i^{\mathcal{B}}$, or it contains exactly one transition leading from a state of $Q^{\mathcal{A}}$ to $i^{\mathcal{B}}$. In the first case we must have $I = \{i^{\mathcal{A}}, i^{\mathcal{B}}\}$, and so $i^{\mathcal{A}} \in F^{\mathcal{A}}$. This means $\varepsilon \in L$. If an accepting path starts from $i^{\mathcal{B}}$, all transitions on the path must belong to $T^{\mathcal{B}}$. Thus we have $i^{\mathcal{B}}T_w^{\mathcal{B}} \in F^{\mathcal{B}}$. It follows that $w \in K$, and so $w \in LK$. In the second case, where the path contains exactly one transition from $Q^{\mathcal{A}}$ to $i^{\mathcal{B}}$, note that this transition must be of the form $(q, a, i^{\mathcal{B}})$ where $qT_a^{\mathcal{A}} \in F^{\mathcal{A}}$. Note also that every transition before this one lies in $T^{\mathcal{A}}$, and every transition after lies in $T^{\mathcal{B}}$. Write $w = uav$; then $i^{\mathcal{A}}T_{ua}^{\mathcal{A}} = qT_a^{\mathcal{A}} \in F^{\mathcal{A}}$, so $ua \in L$. Also, $i^{\mathcal{B}}T_v^{\mathcal{B}}$ must be final, or our path would not be accepting. Thus $v \in K$, and $uav = w \in LK$. □

We are interested in the deterministic state complexity of concatenation, so we convert the FA $\mathcal{AB} = (Q, \Sigma, T, I, F)$ to a DFA recognizing the same language. The DFA we use is $\mathcal{C} = (\mathcal{P}(Q), \Sigma, T, I, F_0)$, where $S \subseteq Q$ is in F_0 if $S \cap F \neq \emptyset$. Since each relation T_a can be viewed as a function from $\mathcal{P}(Q)$ to itself, and there is a unique initial state $I \in \mathcal{P}(Q)$, this automaton is indeed deterministic.

Since $IT_w \in F_0$ if and only if $IT_w \cap F \neq \emptyset$, we see that $\mathcal{C}$ recognizes the same language as $\mathcal{AB}$. We call $\mathcal{C}$ the *concatenation DFA* for $\mathcal{A}$ and $\mathcal{B}$.

We make some observations and introduce some conventions to make it easier to work with the concatenation DFA.

- Since we are assuming $\mathcal{A}$ and $\mathcal{B}$ are DFAs, the only reachable states in $\mathcal{C}$ have the form $S^{\mathcal{A}} \cup S^{\mathcal{B}}$, where $S^{\mathcal{A}} \subseteq Q^{\mathcal{A}}$, $S^{\mathcal{B}} \subseteq Q^{\mathcal{B}}$, and $|S^{\mathcal{A}}| \leq 1$. Without loss of generality, we can assume the state set of $\mathcal{C}$ consists of states of this form, rather than all of $\mathcal{P}(Q)$.

- We mark the states of $\mathcal{A}$ with primes so they can be distinguished from the states of $\mathcal{B}$. So a variable named p or q generally means a element of $Q^{\mathcal{B}}$, while p' or q' means an element of $Q^{\mathcal{A}}$.

- We identify the set $S^{\mathcal{A}} \cup S^{\mathcal{B}}$ with the ordered pair $(S^{\mathcal{A}}, S^{\mathcal{B}})$. Hence we can view the states of $\mathcal{C}$ as these ordered pairs. Reachable states are either of the form $(\emptyset, S)$ or $(\{q'\}, S)$ with $q' \in Q^{\mathcal{A}}$, $S \subseteq Q^{\mathcal{B}}$.

- For convenience, we frequently make no distinction between singleton sets and the elements they contain, and so write (q', S) rather than $(\{q'\}, S)$.

- Rather than T_w, $T_w^{\mathcal{A}}$ and $T_w^{\mathcal{B}}$, we simply write w when it is clear from context which relation is meant. For example, $(q', S)w$ means $(q', S)T_w$ since (q', S) is a state of $\mathcal{C}$, and thus T_w is the natural relation to apply. From our convention for marking the states of $\mathcal{A}$ and $\mathcal{B}$ with primes, one can infer that $q'w$ means $q'T_w^{\mathcal{A}}$ and qw means $qT_w^{\mathcal{B}}$.

- Rather than $i^{\mathcal{A}}$ and $i^{\mathcal{B}}$, let $1'$ denote the initial state of $\mathcal{A}$ and let 1 denote the initial state of $\mathcal{B}$. We also assume without loss of generality that $Q^{\mathcal{A}} = \{1', 2', \ldots, m'\}$ and $Q^{\mathcal{B}} = \{1, 2, \ldots, n\}$ for some m and n.

Under these conventions, the transitions of $\mathcal{C}$ can be described as follows:

$$(q', S)a = \begin{cases} (\emptyset, Sa), & \text{if } a \in \Sigma^{\mathcal{B}} \setminus \Sigma^{\mathcal{A}}; \\ (q'a, \emptyset), & \text{if } a \in \Sigma^{\mathcal{A}} \setminus \Sigma^{\mathcal{B}} \text{ and } q'a \notin F^{\mathcal{A}}; \\ (q'a, 1), & \text{if } a \in \Sigma^{\mathcal{A}} \setminus \Sigma^{\mathcal{B}} \text{ and } q'a \in F^{\mathcal{A}}; \\ (q'a, Sa), & \text{if } a \in \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}} \text{ and } q'a \notin F^{\mathcal{A}}; \\ (q'a, Sa \cup 1), & \text{if } a \in \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}} \text{ and } q'a \in F^{\mathcal{A}}. \end{cases}$$

Recall that $T_w^{\mathcal{A}}$ is the empty relation if w is not a word over $\Sigma^{\mathcal{A}}$, and similarly for $\mathcal{B}$. Thus the transitions admit a simpler description:

$$(q', S)a = \begin{cases} (q'a, Sa \cup 1), & \text{if } a \in \Sigma^{\mathcal{A}} \text{ and } q'a \in F^{\mathcal{A}}; \\ (q'a, Sa), & \text{otherwise.} \end{cases}$$

2.3 State Complexity

We say a DFA $\mathcal{A}$ is *minimal* if it has the least number of states among all DFAs that recognize $L(\mathcal{A})$. It is well known that each regular language has a unique minimal DFA (up to renamings of the states). The *state complexity* of a regular language L, denoted $\mathrm{sc}(L)$, is the number of states in its minimal DFA.

There is a subtlety in this definition, arising from the fact that there are two common ways to define equality of functions. One way is to say that functions are simply certain sets of ordered pairs, and are equal if they are equal as sets. The other way is to say that functions are triples (f, D, C), where D is the domain of the function and C is the codomain, and thus two functions are equal if they are equal as sets *and* have the same domain and codomain. Since words over alphabets are formally defined as functions, the first viewpoint implies that two words over distinct alphabets can be equal, while the second viewpoint implies two words over distinct alphabets are always distinct. We call the first viewpoint the *unrestricted viewpoint*, and the second the *restricted viewpoint*, since the second viewpoint has more restrictive conditions for function equality.

Now, consider how this affects the state complexity of the language $L = \{a\}^*$ over alphabet $\{a, b\}$. The smallest DFA with alphabet $\{a, b\}$ that recognizes L has two states; a second state is necessary to exclude the words that contain b. Thus in the restricted viewpoint, the state complexity of L is two. But in the unrestricted viewpoint, L is equal to the language $\{a\}^*$ over alphabet $\{a\}$, which is recognized by a one-state DFA; thus the state complexity of L is one.

The following characterization of minimality is useful. Let $\mathcal{D} = (Q, \Sigma, T, i, F)$ be a DFA. A state $q \in Q$ is *reachable* if $iw = q$. For $p, q \in Q$, we say q is *reachable from* p if $pw = q$. Two states $p, q \in Q$ are *indistinguishable* if they are equivalent under the following equivalence relation: $p \sim q$ if for all $w \in \Sigma^*$, we have $pw \in F \iff qw \in F$. Otherwise they are *distinguishable* by some word w such that $pw \in F \iff qw \notin F$. In the restricted viewpoint, a DFA is minimal if and only if all of its states are reachable and pairwise distinguishable. In the unrestricted viewpoint, we also require that the DFA has an alphabet of minimal size.

Let $\circ$ be a binary operation on regular languages. The *state complexity of the operation* $\circ$ is the following function, where m and n are positive integers:

$$(m, n) \mapsto \max\{\mathrm{sc}(L \circ K) : \mathrm{sc}(L) \leq m, \mathrm{sc}(K) \leq n\}.$$

This is the worst-case state complexity of the result of the operation, expressed as a function of the maximal allowed state complexities of the input languages. As with state complexity of languages, this definition differs depending on whether we adopt the restricted or unrestricted viewpoint, but the consequences are farther-reaching.

In the restricted viewpoint, we must assume that the inputs to the binary operation are languages over a *common alphabet*. The restricted viewpoint considers words over different alphabets to be always distinct, so it generally does not make sense to perform binary operations on languages over different alphabets. For example, if we take the language $L = \{ab\}$ over alphabet $\{a, b\}$, and

the language $K = \{ab\}$ over alphabet $\{a, b, c\}$, the union $L \cup K$ contains two distinct elements both representing the word ab. This set $L \cup K$ is arguably not a language at all, since it cannot be written as a set of words over a single alphabet. Thus when computing the *restricted state complexity* of binary operations, we only consider inputs with the same alphabet.

In the unrestricted viewpoint, there is no issue in allowing the input languages to have different alphabets. Thus when computing the *unrestricted state complexity* of binary operations, we consider all possible inputs to the operation, including pairs of languages with different alphabets. Allowing for different alphabets makes unrestricted state complexity slightly more complicated to compute. In fact, for many years, papers on operational state complexity only considered restricted state complexity. Unrestricted state complexity was first studied by Brzozowski [2] in 2016, who pointed out that the restriction to common alphabets is artificial and can be removed.

Let us derive an upper bound for the restricted and unrestricted state complexities of the concatenation operation. We begin with two DFAs $\mathcal{A}$ and $\mathcal{B}$ that have m and n states respectively. The number of reachable states in the concatenation DFA $\mathcal{C}$ for $\mathcal{A}$ and $\mathcal{B}$ gives an upper bound for the state complexity of $L(\mathcal{A})L(\mathcal{B})$. Recall that reachable states have the form $(S^{\mathcal{A}}, S^{\mathcal{B}})$, where $S^{\mathcal{A}} \subseteq Q^{\mathcal{A}}$, $S^{\mathcal{B}} \subseteq Q^{\mathcal{B}}$ and $|S^{\mathcal{A}}| \leq 1$. Since $|Q^{\mathcal{A}}| = m$, there are $m + 1$ possible values for $S^{\mathcal{A}}$ (each of the singletons and the empty set). Since $|Q^{\mathcal{B}}| = n$, there are 2^n possible values for $S^{\mathcal{B}}$. However, if $S^{\mathcal{A}} = \{f\}$ for a final state $f \in F^{\mathcal{A}}$, the transition structure of $\mathcal{C}$ tells us that we must have $1 \in S^{\mathcal{B}}$. Thus if $|F^{\mathcal{A}}| = k$, then there are at most $(m + 1 - k)2^n$ states with a non-final state or the empty set in the first component, and $k2^{n-1}$ states with a final state in the first component. It follows there are at most $(m + 1 - k)2^n + k2^{n-1}$ reachable states in $\mathcal{C}$. This is maximized by taking $k = 1$, giving an upper bound of $m2^n + 2^{n-1}$ in the unrestricted case. For the restricted case, note that we cannot get the empty set in the first component, since this requires using a letter in $\Sigma^{\mathcal{B}} \setminus \Sigma^{\mathcal{A}}$. Thus we get an upper bound of $(m-1)2^n + 2^{n-1}$ in the restricted case. We will see later that both of these bounds are tight.

3 Results

Let $\mathcal{A} = (Q^{\mathcal{A}}, \Sigma^{\mathcal{A}}, T^{\mathcal{A}}, 1', F^{\mathcal{A}})$ and $\mathcal{B} = (Q^{\mathcal{B}}, \Sigma^{\mathcal{B}}, T^{\mathcal{B}}, 1, F^{\mathcal{B}})$ be DFAs, with $Q^{\mathcal{A}} = \{1', 2', \ldots, m'\}$ and $Q^{\mathcal{B}} = \{1, 2, \ldots, n\}$ for positive integers m and n. Let $\mathcal{C} = (Q, \Sigma, T, I, F)$ denote the concatenation DFA of $\mathcal{A}$ and $\mathcal{B}$ as defined in Section 2.2.

Remark. Let $p', q' \in Q^{\mathcal{A}}$, let $X, Y, Z \subseteq Q^{\mathcal{B}}$, and let $w \in \Sigma^*$. Then:

$$\text{In } \mathcal{C}, \text{ if } (p', X)w = (q', Y), \text{ then } (p', X \cup Z)w = (q', Y \cup Zw).$$

Indeed, recall that the pair (p', X) stands for the set $\{p'\} \cup X$. Thus $(\{p'\} \cup X)w = \{p'w\} \cup Xw = \{q'\} \cup Y$. It follows that $p'w = q'$ and $Xw = Y$. Hence $(\{p'\} \cup X \cup Z)w = \{p'w\} \cup Xw \cup Zw = \{q'\} \cup Y \cup Zw$, which in our pair notation is $(q', Y \cup Zw)$. We will readily use this basic fact in proofs.

Before stating our main result formally, we give some motivating exposition. Fix a state $s' \in Q^{\mathcal{A}}$ and a subset B of $Q^{\mathcal{B}}$. The state s' is called the *focus state* or simply *focus*; it is often taken to be the initial state $1'$ but in general can be any state. The subset B is called the *base*. Fix a set T with $B \subseteq T \subseteq Q^{\mathcal{B}}$, called the *target*. Our goal is to give sufficient conditions under which starting from (s', B), we can reach (s', S) for all sets S with $B \subseteq S \subseteq T$. That is, we can reach any state of the concatenation DFA $\mathcal{C}$ in which the first component is the focus and the second component lies between the base and the target.

The idea is to first assume we can reach (s', B), the state consisting of the focus and the base. Now, for $q \in Q$, define a *q-word* to be a word w such that $(s', B)w = (s', B \cup q)$. We can think of this as a word that "adds" the state q to the base B. Our next assumption is that we have a q-word for each state q in the target T. To reach a set S with $B \subseteq S \subseteq T$, we will repeatedly use q-words to add each missing element of S to the base B.

There is a problem with this idea, which we illustrate with an example. Suppose w_p is a p-word and w_q is a q-word, and we want to reach $(s', B \cup \{p, q\})$. Starting from (s', B) we may apply w_p to reach $(s', B \cup p)$. But now if we apply w_q, we reach $(s', B \cup \{pw_q, q\})$. There is no guarantee that we have $pw_q = p$, and in many cases we will not. What we should really do is find a state r such that $rw_q = p$, use an r-word to reach $(s', B \cup r)$, and then apply w_q to reach $(s', B \cup \{p, q\})$. But this idea only works if p has a preimage under w_q, which may not be the case.

We resolve this by making a technical assumption, which ensures that preimages will always exist when we attempt constructions like the above. First, define a *construction set* for the target T to be a set of words consisting of exactly one q-word for each $q \in T$. If W is a construction set for T, we write $W[q]$ for the unique q-word in W.

We say a construction set is *complete* if there is a total order $\prec$ on the target T such that for all $p, q \in T$ with $p \prec q$, the state q has at least one preimage under the unique p-word $W[p]$, and at least one of these preimages lies in T. More formally, whenever $p \prec q$, the set $qW[p]^{-1} = \{s \in Q^{\mathcal{B}} : sW[p] = q\}$ intersects T non-trivially. Our final assumption is that we have a complete construction set for T.

Note that the definition of a q-word depends not only on q, but also on s' and B. Since a construction set for T is a set of q-words, the definition of construction set also depends on s' and B. For simplicity, we omit this dependence on s' and B from the notation for q-words and construction sets.

We summarize the definitions that have just been introduced:

- Fix a state $s' \in Q^{\mathcal{A}}$, called the *focus*, and a set $B \subseteq Q^{\mathcal{B}}$ called the *base*.
- For $q \in Q^{\mathcal{B}}$, a *q-word* is a word w such that $(s', B)w = (s', B \cup q)$.
- Given a *target* set T with $B \subseteq T \subseteq Q^{\mathcal{B}}$, a *construction set* for T is a set of words that contains exactly one q-word for each $q \in T$.
- The unique q-word in a construction set W is denoted by $W[q]$.

- A construction set for T is *complete* if there exists a total order $\prec$ on T such that for all $p, q \in T$ with $p \prec q$, we have

$$qW[p]^{-1} \cap T = \{s \in Q^{\mathcal{B}} : sW[p] = q\} \cap T \neq \emptyset.$$

Now, we state our main theorem, which gives the formal version of the construction described above.

Theorem 1. *Fix a state $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. If there is a complete construction set for T, then all states of the form (s', S) with $B \subseteq S \subseteq T$ are reachable from (s', B) in $\mathcal{C}$. In particular, if (s', B) itself is reachable, then all states (s', S) with $B \subseteq S \subseteq T$ are reachable.*

Proof. Note that if $B \subseteq S \subseteq T$, we can write $S = R \cup B$ with $R \cap B = \emptyset$ and $R \subseteq T$. Thus it suffices to show that all states of the form $(s', R \cup B)$ with $R \cap B = \emptyset$ and $R \subseteq T$ are reachable from (s', B). We proceed by induction on $|R|$. When $|R| = 0$, the only state of this form is (s', B) itself.

Now suppose every state $(s', R \cup B)$ with $R \cap B = \emptyset$, $R \subseteq T$ and $0 \leq |R| < k$ is reachable from (s', B). We want to show this also holds for $|R| = k$. Let W be a complete construction set for T and let $\prec$ be the corresponding total order on T. Let p be the minimal element of R under $\prec$. Let w be $W[p]$, the unique p-word in W. For all $q \in R \setminus p$, we have $p \prec q$ and thus qw^{-1} contains an element of T (since W is complete).

Construct sets X and Y as follows: starting with $X = \emptyset$, for each $q \in R \setminus p$, choose an element of $qw^{-1} \cap T$ and add it to X. Then set $Y = X \setminus B$. Observe that X is a subset of T of size $|R \setminus p| = k - 1$. Hence Y is a subset of T of size at most $k - 1$ such that $Y \cap B = \emptyset$. It follows by the induction hypothesis that $(s', Y \cup B)$ is reachable from (s', B). But $Y \cup B = X \cup B$, so $(s', X \cup B)$ is reachable from (s', B). By the definition of X, we have $Xw = R \setminus p$. Since w is a p-word, we have $(s', B)w = (s', B \cup p)$, and thus

$$(s', X \cup B)w = (s', Xw \cup B \cup p) = (s', (R \setminus p) \cup B \cup p) = (s', R \cup B).$$

Hence $(s', R \cup B)$ is reachable from (s', B), as required. □

The definition of completeness is somewhat complicated, which makes it difficult to use Theorem 1. Thus, we next prove some results giving useful sufficient conditions for a construction set to be complete. Before stating our first such result, we introduce some notation.

Define $\Sigma_0 = \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}}$. We call Σ_0 the *shared alphabet* of $\mathcal{A}$ and $\mathcal{B}$. The following remark shows that when $\Sigma^{\mathcal{A}} \neq \Sigma^{\mathcal{B}}$, it is important to work exclusively with the shared alphabet when looking for complete construction sets. Of course, if $\Sigma^{\mathcal{A}} = \Sigma^{\mathcal{B}}$ then the shared alphabet is just the common alphabet of both automata, and there is nothing to worry about.

Remark. A construction set for a non-empty target cannot be complete unless it is a subset of Σ_0^*. To see this, suppose W is a construction set and let $w \in W$. If w contains a letter from $\Sigma^{\mathcal{A}} \setminus \Sigma^{\mathcal{B}}$, then w is not a word over $\Sigma^{\mathcal{B}}$. Recall that if

w is not a word over $\Sigma^{\mathcal{B}}$, then $T_w^{\mathcal{B}}$ is defined to be the empty relation. Thus the converse relation $(T_w^{\mathcal{B}})^{-1}$ is also empty, which means qw^{-1} is empty for all q. It follows W cannot be complete. On the other hand, suppose w contains a letter from $\Sigma^{\mathcal{B}} \setminus \Sigma^{\mathcal{A}}$. Then $(s', B)w = (\emptyset, Bw)$. Hence w is not a q-word for any q, and so w cannot be an element of a construction set, which is a contradiction. It follows that all words in a complete construction set are words over the shared alphabet Σ_0.

Lemma 1. *Fix $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. Let $x_1, \ldots, x_j$ be words over Σ_0 that act as permutations on T, and let y be an arbitrary word over Σ_0. Choose $x_0 \in \{\varepsilon, x_1, \ldots, x_j\}$. Define*

$$W = \{x_1, x_2, \ldots, x_j\} \cup \{x_0 y, x_0 y^2, \ldots, x_0 y^k\}.$$

If W is a construction set for T, then it is complete.

Proof. For $1 \leq i \leq j$, let $w_i = x_i$. For $1 \leq i \leq k$, let $w_{j+i} = x_0 y^i$. Let $\ell = j + k$. Then we have $W = \{w_1, \ldots, w_\ell\}$. Let q_i be the state in T such that $(s', B)w_i = (s', B \cup q_i)$. Define an order $\prec$ on T so that $q_1 \prec q_2 \prec \cdots \prec q_\ell$. We claim this order makes W complete. Notice that $w_r = W[q_r]$, the unique q_r-word in W. Thus we must show that whenever $q_r \prec q_s$, we have $q_s w_r^{-1} \cap T \neq \emptyset$.

Suppose $r < s$ and $r \leq j$. Then $w_r = x_r$ acts as a permutation on T. Thus $q_s w_r^{-1} \cap T$ is non-empty, since $q_s \in T$.

Suppose $r < s$ and $r > j$. Since $s - r > 0$, we can write $w_s = x_0 y^{s-j} = x_0 y^{s-r} y^{r-j} = w_{j+s-r} y^{r-j}$. Thus $(s', B)w_s = (s', B \cup q_{j+s-r})y^{r-j} = (s', B \cup q_s)$. There are two possibilities: $q_{j+s-r} y^{r-j} = q_s$, or $q y^{r-j} = q_s$ for some $q \in B$.

In either case, $q_s (y^{r-j})^{-1} \cap T$ is non-empty. That is, there exists $q \in T$ such that $q y^{r-j} = q_s$. Since x_0 acts as a permutation on T, there exists $p \in T$ such that $p x_0 = q$. Thus $p x_0 y^{r-j} = p w_r = q_s$. It follows that $q_s w_r^{-1} \cap T$ is non-empty, as required. □

Usually, we will use one of the following corollaries instead of Lemma 1 itself.

Corollary 1. *Fix $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. Let x and y be words over Σ_0 such that x acts as a permutation on T. Suppose W is one of the following sets:*

1. $\{y, y^2, \ldots, y^k\}$.
2. $\{\varepsilon, y, y^2, \ldots, y^k\}$.
3. $\{x, xy, xy^2, \ldots, xy^k\}$.
4. $\{\varepsilon, x, xy, xy^2, \ldots, xy^k\}$.

If W is a construction set for T, then it is complete.

Proof. All statements follow easily from Lemma 1:

1. Set $j = 0$.

2. Set $j = 1$ and $x_0 = x_1 = \varepsilon$.

3. Set $j = 1$ and $x_0 = x_1 = x$.

4. Set $j = 2$, $x_1 = \varepsilon$ and $x_0 = x_2 = x$. □

Corollary 2. *Fix $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. Let $W \subseteq \Sigma_0^*$ be a construction set for T.*

1. *If every word in W acts as a permutation on T, then W is complete.*

2. *If there is a word $w \in W$ such that every word in $W \setminus w$ acts as a permutation on T, then W is complete.*

Proof. Both statements follow easily from Lemma 1:

1. Set $k = 0$ in Lemma 1.

2. Set $k = 1$, $x_0 = \varepsilon$ and $y = w$ in Lemma 1. □

In the special case where W contains ε, Corollary 2 admits the following generalization, which we found occasionally useful.

Lemma 2. *Fix $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. Let $W = \{\varepsilon, w_1, \ldots, w_k\}$ be a construction set for T, where $w_1, \ldots, w_k$ are non-empty words over Σ_0. Suppose that for every word $w \in W$, there exists a set S with $T \setminus B \subseteq S \subseteq T$ such that w acts as a permutation on S. Then W is complete.*

Proof. Write $B = \{q_1, \ldots, q_j\}$. Note that ε is a q_i-word for $1 \leq i \leq j$. Thus by the definition of a construction set, ε is the unique q_i-word in W for each $q_i \in B$, that is, $W[q_i] = \varepsilon$ for $1 \leq i \leq j$. In particular, each non-empty word in W is a q-word for some $q \in T \setminus B$. For $1 \leq i \leq k$, let q_{j+i} be the state such that $(s', B)w_i = (s', B \cup q_{j+i})$. Then $T = \{q_1, \ldots, q_{j+k}\}$. Note that $W[q_i] = \varepsilon$ if $1 \leq i \leq j$, and $W[q_i] = w_{i-j}$ if $j + 1 \leq i \leq j + k$.

Define an order $\prec$ on T by $q_1 \prec q_2 \prec \cdots \prec q_{j+k}$. We claim this order makes W complete. Choose $q_r, q_s \in T$ with $q_r \prec q_s$; we want to show that $q_s W[q_r]^{-1} \cap T \neq \emptyset$. Suppose $q_r \in B$. Then $W[q_r] = \varepsilon$, and we have $q_s\varepsilon^{-1} \cap T$ non-empty as required. Now if $q_r \notin B$, then since $q_r \prec q_s$ we also have $q_s \notin B$. In this case, $W[q_r] = w_{r-j}$, which acts as a permutation on some superset S of $T \setminus B$. Since $q_s \in T \setminus B$, it follows that q_s has a preimage under w_{r-j}, and furthermore this preimage lies in T, since S is a subset of T. Thus $q_s w_{r-j}^{-1} \cap T \neq \emptyset$ as required. This proves that W is complete. □

Note that *all words* referred to in the above lemmas and corollaries are words over Σ_0, the shared alphabet of $\mathcal{A}$ and $\mathcal{B}$. When working with automata that have different alphabets, it is important to only use words over the shared alphabet when trying to find a complete construction set.

The following "master theorem" summarizes all the results of this section. We have attempted to state this theorem in a form such that it can be cited without having to first define all the notions introduced in this section, such as q-words and construction sets and completeness.

Theorem 2. *Let $\mathcal{A} = (Q^{\mathcal{A}}, \Sigma^{\mathcal{A}}, T^{\mathcal{A}}, i^{\mathcal{A}}, F^{\mathcal{A}})$ and $\mathcal{B} = (Q^{\mathcal{B}}, \Sigma^{\mathcal{B}}, T^{\mathcal{B}}, i^{\mathcal{B}}, F^{\mathcal{B}})$ be DFAs. Let $\mathcal{C} = (Q, \Sigma, T, I, F)$ denote the concatenation DFA of $\mathcal{A}$ and $\mathcal{B}$, as defined in Section 2.2. Let $\Sigma_0 = \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}}$.*

Fix a state $s' \in Q^{\mathcal{A}}$ and sets $B \subseteq T \subseteq Q^{\mathcal{B}}$. Suppose that for each $q \in T$, there exists a word $w_q \in \Sigma_0^*$ such that $(s', B) \xrightarrow{w_q} (s', B \cup q)$ in $\mathcal{C}$. Let $W = \{w_q : q \in T\}$. Suppose that one of the following conditions holds:

1. *There exist words $x, y \in \Sigma_0^*$, where x acts as a permutation on T, such that W can be written in one of the following forms:*
   - $W = \{y, y^2, \ldots, y^k\}$.
   - $W = \{\varepsilon, y, y^2, \ldots, y^k\}$.
   - $W = \{x, xy, xy^2, \ldots, xy^k\}$.
   - $W = \{\varepsilon, x, xy, xy^2, \ldots, xy^k\}$.
2. *Every word in W acts as a permutation on T.*
3. *There exists $w \in W$ such that every word in $W \setminus w$ acts as a permutation on T.*
4. *W contains ε, and for every non-empty word $w \in W$, there exists a set S such that $T \setminus B \subseteq S \subseteq T$ and w acts as a permutation on S.*
5. *There exists a total order $\prec$ on T such that for all $p, q \in T$ with $p \prec q$, the set $qw_p^{-1} = \{s \in Q^{\mathcal{B}} : s \xrightarrow{w_p} q\}$ contains an element of T.*

If one of the above conditions holds, then every state of the form (s', X) with $B \subseteq X \subseteq T$ is reachable from (s', B) in $\mathcal{C}$.

4 Examples

We now demonstrate our technique by applying it to various concatenation witnesses from the literature.

Theorem 3 (Regular Language Witness. Brzozowski and Sinnamon, 2017 [9])**.** *Let $t\colon Q^{\mathcal{A}} \to Q^{\mathcal{A}}$ be a transformation such that $j't = 1'$. Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	*Final States*
$\mathcal{A}$:	$(1', \ldots, m')$	t	$\{m'\}$
$\mathcal{B}$:	$(1, \ldots, n)$	$(2 \to 1)$	$\{n\}$

If $\gcd(j-1, n) = 1$, then $\mathcal{C}$ has $(m-1)2^n + 2^{n-1}$ reachable states. In particular, transformations t with $2't = 1'$ work for all m and n.

The authors of [9] proved this result with $t = (1', 2')$, but we prove a slightly more general statement.

Proof. The initial state of $\mathcal{C}$ is $(1', \emptyset)$. Set $x = a^m$ and $y = a^{j-1}b$. We have

$$(1', \emptyset) \xrightarrow{x} (1', 2) \xrightarrow{y^k} (1', 2 + k(j-1)).$$

(Addition in the second component is performed modulo n.) Since $j-1$ and n are coprime, it follows from elementary number theory that $W = \{x, xy, \ldots, xy^{n-1}\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = 1'$ and $B = \emptyset$). By Corollary 1, it is complete. Hence $(1', S)$ is reachable for all $S \subseteq Q^{\mathcal{B}}$. To reach (q', S) for q' non-final, first reach $(1', Sa^{-(q-1)})$ and then apply a^{q-1}. To reach $(m', S \cup 1)$ for $S \subseteq Q^{\mathcal{B}} \setminus 1$, first reach $((m-1)', Sa^{-1})$ and then apply a. □

Note that the above theorem only gives conditions for $(m-1)2^n + 2^{n-1}$ states to be *reachable*; it is not necessarily true that all the reachable states are pairwise distinguishable. For example, if t is the constant transformation $(Q^{\mathcal{A}} \to 1')$ then $(p', Q^{\mathcal{B}})$ and $(q', Q^{\mathcal{B}})$ are indistinguishable. However, in [9] the authors take t to be the transposition $(1', 2')$ and find that all reachable states are pairwise distinguishable.

In the remainder of our examples, all of the states we show are reachable will also be pairwise distinguishable. Since the focus of this paper is reachability, we refer to the original authors for distinguishability proofs in most cases. In cases where the original authors did not provide a distinguishability proof, we give a brief argument for completeness.

The next example involves two DFAs with different alphabets: we have $\Sigma^{\mathcal{A}} = \{a, b, c\}$ and $\Sigma^{\mathcal{B}} = \{a, b, d\}$. Our construction set will consist of words over the shared alphabet $\Sigma_0 = \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}} = \{a, b\}$.

Theorem 4 (Regular Language Witness. Brzozowski, 2016 [2])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	d	*Final States*
$\mathcal{A}$:	$(1', \ldots, m')$	$(1', 2')$	$(m' \to 1')$		$\{m'\}$
$\mathcal{B}$:	$(1, 2)$	$(1, \ldots, n)$		id	$\{n\}$

Then $\mathcal{C}$ has $m2^n + 2^{n-1}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. If n is odd, we have

$$(1', \emptyset) \xrightarrow{a^m} (1', 2) \xrightarrow{bb} (1', 4) \xrightarrow{bb} \cdots \xrightarrow{bb} (1', n-1),$$

$$(1', n-1) \xrightarrow{bb} (1', 1) \xrightarrow{bb} (1', 3) \xrightarrow{bb} \cdots \xrightarrow{bb} (1', n).$$

Thus $\{a^m, a^m bb, a^m (bb)^2, \ldots, a^m (bb)^{n-1}\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = 1'$ and $B = \emptyset$). By Corollary 1, it is complete (taking $x = a^m$ and $y = bb$).

If n is even, we have

$$(1', \emptyset) \xrightarrow{a^m} (1', 2) \xrightarrow{bb} (1', 4) \xrightarrow{bb} \cdots \xrightarrow{bb} (1', n),$$

$$(1', n) \xrightarrow{ab} (1', 1) \xrightarrow{bb} (1', 3) \xrightarrow{bb} \cdots \xrightarrow{bb} (1', n-1).$$

The words used to reach each state $(1', q)$ form a construction set for $Q^{\mathcal{B}}$ (with $s' = 1$ and $B = \emptyset$). We cannot use Corollary 1 to show it is complete (since the appearance of ab breaks the pattern), but notice that all the words in the construction set are words over $\{a, b\}$, and a and b both act as permutations on $Q^{\mathcal{B}}$. Thus all words in the construction set are permutations of $Q^{\mathcal{B}}$, and so by Corollary 2 it is complete.

In either case, we have a complete construction set for $Q^{\mathcal{B}}$ and so $(1', S)$ is reachable for all $S \subseteq Q^{\mathcal{B}}$. We can reach (q', S) for $q' \neq m'$ and $(m', S \cup 1)$ by words in a^*, as in Theorem 3. This gives $(m-1)2^n + 2^{n-1}$ reachable states. Additionally, from (q', S) we can reach $(\emptyset, S)$ by d, for an extra 2^n states.

For distinguishability of the reached states, see [2]. □

The main differences in reachability proofs between the different-alphabet case (unrestricted state complexity) and the same-alphabet case (restricted state complexity) are as follows:

- When looking for a complete construction set, we are restricted to using words over the shared alphabet $\Sigma_0 = \Sigma^{\mathcal{A}} \cap \Sigma^{\mathcal{B}}$.
- Usually some additional states can be reached using letters in $\Sigma^{\mathcal{A}} \setminus \Sigma^{\mathcal{B}}$ or $\Sigma^{\mathcal{B}} \setminus \Sigma^{\mathcal{A}}$, e.g., the states of the form $(\emptyset, S)$ in the previous example.

As these differences are not too significant, we will stick to the same-alphabet case for the remainder of our examples.

Theorem 5 (Regular Language Witness. Brzozowski, 2013 [1])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	*Final States*
$\mathcal{A}$:	$(1', \ldots, m')$	$(1', 2')$	$(m' \to 1')$	$\{m'\}$
$\mathcal{B}$:	$(1, \ldots, n)$	$(1, 2)$	$(n \to 1)$	$\{n\}$

Then $\mathcal{C}$ has $(m-1)2^n + 2^{n-1}$ reachable and pairwise distinguishable states.

Proof. The initial state of $\mathcal{C}$ is $(1', \emptyset)$. For $0 \leq k \leq n-2$ we have

$$(1', \emptyset) \xrightarrow{a^m} (1', 2) \xrightarrow{(ab)^k} (1', 2+k).$$

Also, $(1', n) \xrightarrow{c} (1', 1)$. Thus $\{a^m, a^m ab, a^m (ab)^2, \ldots, a^m (ab)^{n-2}, a^m (ab)^{n-2} c\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = 1'$ and $B = \emptyset$).

This construction set does not quite have the right form to apply Corollary 1, due to the last word $a^m (ab)^{n-2} c$. However, notice that all words in W except for $a^m (ab)^{n-2} c$ are in fact permutations of $Q^{\mathcal{B}}$, so Corollary 2 shows that W is complete. Hence all states $(1', S)$ with $S \subseteq Q^{\mathcal{B}}$ are reachable. We can reach (q', S) for $q' \neq m'$ and $(m', S \cup 1)$ by words in a^*, as in Theorem 3.

For distinguishability of the reached states, see [1]. □

Theorem 6 (Regular Language Witness. Yu, Zhuang and Salomaa, 1994 [19])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	*Final States*
$\mathcal{A}$:	$(1', \ldots, m')$	$(Q^{\mathcal{A}} \to 1')$	id	$\{m'\}$
$\mathcal{B}$:	id	$(1, \ldots, n)$	$(Q^{\mathcal{B}} \to 2)$	$\{n\}$

Then $\mathcal{C}$ has $(m-1)2^n + 2^{n-1}$ reachable and pairwise distinguishable states.

Proof. The initial state of $\mathcal{C}$ is $(1', \emptyset)$. For $k \leq n-2$ we have $(1', \emptyset) \xrightarrow{a^m} (1', 2) \xrightarrow{b^k} (1', 2+k)$, and $(1', n) \xrightarrow{b} (1', 1)$. It follows that $\{a^m, a^m b, \ldots, ab^{n-1}\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = 1'$ and $B = \emptyset$). By Corollary 1, it is complete (taking $x = a^m$ and $y = b$). Hence all states $(1', S)$ with $S \subseteq Q^{\mathcal{B}}$ are reachable. We can reach (q', S) for $q' \neq m'$ and $(m', S \cup 1)$ by words in a^*.

Let (p', S) and (q', T) be distinct states of $\mathcal{C}$. If $S \neq T$, let r be a state in the symmetric difference of S and T. Then b^{n-r} distinguishes the states. If $S = T$ and $p' < q'$, then $ca^{m-q}b^{n-2}$ distinguishes the states. □

Theorem 7 (Regular Language Witness. Maslov, 1970 [15])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	*Final States*
$\mathcal{A}$:	$(1', \ldots, m')$	id	$\{m'\}$
$\mathcal{B}$:	$(n-1, n)$	$({}_1^{n-1}q \to q+1)$	$\{n\}$

Then $\mathcal{C}$ has $(m-1)2^n + 2^{n-1}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. We have

$$(1', \emptyset) \xrightarrow{a^m} (1', 1) \xrightarrow{b^k} (1', 1+k).$$

Thus $\{a^m, a^m b, a^m b^2, \ldots, a^m b^{n-1}\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = 1'$ and $B = \emptyset$). By Corollary 1, it is complete. Hence $(1', S)$ is reachable for all $S \subseteq Q^{\mathcal{B}}$. We can reach (q', S) for $q' \neq m'$ and $(m', S \cup 1)$ by words in a^*, as in Theorem 3.

Let (p', S) and (q', T) be distinct states of $\mathcal{C}$. If $S \neq T$, let r be a state in the symmetric difference of S and T. Then b^{n-r} distinguishes the states. If $S = T$ and $p' < q'$, by b^n we reach (p', n) and (q', n). Then by a^{m-q} we reach $((p+m-q)', na^{m-q})$ and $(m', na^{m-q} \cup 1)$. These states differ in their second component, so they are distinguishable. □

Theorem 8 (Star-Free Witness. Brzozowski and Liu, 2012 [7])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	d
$\mathcal{A}$:	$({}_1^{m-1}q' \to (q+1)')$	$({}_2^{m}q' \to (q-1)')$	id	$(Q^{\mathcal{A}} \to m')$
$\mathcal{B}$:	$({}_2^{n-1}q \to q+1)$	id	$({}_1^{n-1}q \to q+1)$	$({}_2^{n}q \to q-1)$

and let $F^{\mathcal{A}} = \{m'\}$ and $F^{\mathcal{B}} = \{n-1\}$. Then $\mathcal{C}$ has $(m-1)2^n + 2^{n-1}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. We have

$$(1', \emptyset) \xrightarrow{a^m} (m', 1) \xrightarrow{c^k} (m', \{1, 1+k\}).$$

Hence $\{\varepsilon, c, c^2, \ldots, c^{n-1}\}$ is a construction set for $Q^{\mathcal{B}}$ (with $s' = m'$ and $B = \{1\}$). By Corollary 1, it is complete. Thus $(m', S \cup 1)$ is reachable for all $S \subseteq Q^{\mathcal{B}}$.

To reach (q', S) for non-final $q' \in Q^{\mathcal{A}}$ and $S \subseteq Q^{\mathcal{B}}$, proceed as follows. If $1 \in S$, first reach $(m', S \cup 1)$ then apply b^{m-q}. If $1 \notin S$, let i be the smallest element of S. Set $T = \{q - (i-1) : q \in S \setminus i\}$ and reach $(m', T \cup 1)$. Then $(m', T \cup 1) \xrightarrow{b^{m-q}} (q', T \cup 1) \xrightarrow{c^{i-1}} (q', (S \setminus i) \cup i) = (q', S)$.

For distinguishability of the reached states, see [7]. □

Theorem 9 (Non-Returning Witness. Brzozowski and Davies, 2017 [3])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	*Final States*
$\mathcal{A}$:	$(2', \ldots, m')(1' \to 2')$	$(2', 3')(1' \to 3')$	$\{m'\}$
$\mathcal{B}$:	$(2, \ldots, n)(1 \to 2)$	$(3, \ldots, n)(2 \to 3)(1 \to 2)$	$\{n\}$

Then $\mathcal{C}$ has $(m-1)2^{n-1} + 1$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. Let $x = a^{m-1}$ and $y = ab$. If n is even,

$$(1', \emptyset) \xrightarrow{a} (2', \emptyset) \xrightarrow{x} (2', 2) \xrightarrow{y} (2', 4) \xrightarrow{y} (2', 6) \xrightarrow{y} \cdots \xrightarrow{y} (2', n),$$

$$(2', n) \xrightarrow{y} (2', 3) \xrightarrow{y} (2', 5) \xrightarrow{y} \cdots \xrightarrow{y} (2', n-1).$$

If n is odd,

$$(1', \emptyset) \xrightarrow{a} (2', \emptyset) \xrightarrow{x} (2', 2) \xrightarrow{y} (2', 4) \xrightarrow{y} (2', 6) \xrightarrow{y} \cdots \xrightarrow{y} (2', n-1),$$

$$(2', n-1) \xrightarrow{y} (2', 3) \xrightarrow{y} (2', 5) \xrightarrow{y} \cdots \xrightarrow{y} (2', n).$$

In both cases, Corollary 1 implies that $\{x, xy, \ldots, xy^{n-2}\}$ is a complete construction set for $Q^{\mathcal{B}} \setminus 1$ (with $s' = 2'$ and $B = \emptyset$). It follows that $(2', S)$ is reachable for all $S \subseteq Q^{\mathcal{B}} \setminus 1$. This gives 2^{n-1} reachable states.

To reach (q', S) for non-final $q' \in Q^{\mathcal{A}} \setminus 1$ and $S \subseteq Q^{\mathcal{B}} \setminus 1$, note that a acts as a permutation on $Q^{\mathcal{B}} \setminus 1$, and so there exists $T \subseteq Q^{\mathcal{B}} \setminus 1$ such that $Ta^{q-2} = S$. Thus we can first reach $(2', T)$ and then apply a^{q-2}. To reach $(m', S \cup 1)$ for $S \subseteq Q^{\mathcal{B}} \setminus 1$, reach $((m-1)', T)$ where $Ta = S$ and apply a. Counting the initial state $(1, \emptyset)$, we get $(m-1)2^{n-1} + 1$ reachable states.

For distinguishability of the reached states, see [3]. □

Theorem 10 (Non-Returning Witness. Eom, Han and Jirásková, 2016 [11])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c
$\mathcal{A}$:	$(2', \ldots, m')(1' \to 2')$	$(1' \to 2')$	$(1' \to 2')$
$\mathcal{B}$:	$(1 \to 2)$	$(2, \ldots, n)(1 \to 2)$	$\binom{n-1}{3} q \to q+1)(1 \to 2)(n \to 2)$

and let $F^{\mathcal{A}} = \{m'\}$ and $F^{\mathcal{B}} = \{n\}$. Then $\mathcal{C}$ has $(m-1)2^{n-1} + 1$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. We have

$$(1', \emptyset) \xrightarrow{a} (2', \emptyset) \xrightarrow{a^{m-1}} (2', 2) \xrightarrow{b^k} (2', 2+k).$$

Hence by Corollary 1, $\{a^{m-1}, a^{m-1}b, \ldots, a^{m-1}b^{n-2}\}$ is a complete construction set for $Q^{\mathcal{B}} \setminus 1$ (with $s' = 2'$ and $B = \emptyset$). It follows that $(2', S)$ is reachable for all $S \subseteq Q^{\mathcal{B}} \setminus 1$. To reach (q', S) for q' non-final, reach $(2', S)$ and apply a^{q-2}. For $(m', S \cup 1)$, reach $((m-1)', S)$ and apply a.

For distinguishability of the reached states, see [11]. □

Theorem 11 (Prefix-Closed Witness. Brzozowski, Jirásková and Zou, 2014 [6])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	*Final States*
$\mathcal{A}$:	id	id	$\binom{m-1}{1}q' \to (q+1)'$	$\{1', \ldots, (m-1)'\}$
$\mathcal{B}$:	$(1, \ldots, n-1)$	$\binom{n-1}{2}q \to q+1$	id	$\{1, \ldots, n-1\}$

Then $\mathcal{C}$ has $(m+1)2^{n-2}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', 1)$. For $k \leq n-2$ we have $(1', 1) \xrightarrow{a^k} (1', \{1, 1+k\})$. Thus by Corollary 1 the set $\{\varepsilon, a, a^2, \ldots, a^{n-2}\}$ is a complete construction set for $Q^{\mathcal{B}} \setminus n$, with $s' = 1'$ and $B = \{1\}$. Hence $(1', S \cup 1)$ is reachable for each $S \subseteq Q^{\mathcal{B}} \setminus n$. From $(1', S \cup 1)$ with $S \subseteq Q^{\mathcal{B}} \setminus n$, we reach $(q', S \cup 1)$ for $2 \leq q \leq m$ by c^{q-1}. This gives $m2^{n-2}$ reachable states.

To reach (m', S) with $S \subseteq Q^{\mathcal{B}} \setminus n$, set S non-empty, and $1 \notin S$, let p be the smallest element of S. Let $T = Sa^{-(p-1)}$; then $1 \in T$ since $1a^{p-1} = p$. Reach (m', T) and apply a^{p-1} to reach (m', S). There are $2^{n-2} - 1$ non-empty sets that exclude 1 and n, and we can reach an additional state (m', n) from $(m', n-1)$ by b. This gives another 2^{n-2} reachable states, for a total of $(m+1)2^{n-2}$ states.

For distinguishability of the reached states, see [6]. □

Theorem 12 (Suffix-Free Witness. Brzozowski and Sinnamon, 2017 [8])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c
$\mathcal{A}$:	$(1' \to m')(2', \ldots, (m-1)')$	$(1' \to m')(2', 3')$	$(2', m')(1' \to 2')$
$\mathcal{B}$:	$(1 \to n)(2, 3)$	$(2, n)(1 \to 2)$	$(1 \to n)(2, \ldots, n-1)$

and let $F^{\mathcal{A}} = \{(m-1)'\}$ and $F^{\mathcal{B}} = \{n-1\}$. Then $\mathcal{C}$ has $(m-1)2^{n-2} + 1$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. We have

$$(1', \emptyset) \xrightarrow{c} (2', \emptyset) \xrightarrow{a^{m-3}} ((m-1)', 1) \xrightarrow{c} ((m-1)', \{1, n\}).$$

Then for $k \leq n-3$ we have

$$((m-1)', \{1, n\}) \xrightarrow{bb} ((m-1)', \{1, 2, n\}) \xrightarrow{c^k} ((m-1)', \{1, 2+k, n\}).$$

Thus $W = \{\varepsilon, bb, bbc, bbc^2, \ldots, bbc^{n-3}\}$ is a construction set for $Q^{\mathcal{B}}$, with $s' = (m-1)'$ and $B = \{1, n\}$. In fact, W is complete by Lemma 2 since b and c act as permutations on $Q^{\mathcal{B}} \setminus 1$.

It follows that $((m-1)', S \cup \{1, n\})$ is reachable for all $S \subseteq Q^{\mathcal{B}}$. To reach $(q', S \cup n)$ for $2 \le q \le m-2$ and $1 \notin S$, note that a acts as a permutation on $Q^{\mathcal{B}} \setminus 1$. Thus we first reach $((m-1)', Sa^{-(q-1)} \cup \{1, n\})$ then apply a^{q-1}. To reach $(m', S \cup n)$ with $1 \notin S$, first reach $(2', Sc^{-1} \cup n)$ then apply c. Since there are 2^{n-2} subsets of $Q^{\mathcal{B}} \setminus \{1, n\}$, this gives $(m-1)2^{n-2}$ reachable states. Adding one for the initial state $(1', \emptyset)$ gives $(m-1)2^{n-2} + 1$.

For distinguishability of the reached states, see [8]. Note that the authors of [8] use a different concatenation DFA from our $\mathcal{C}$: they first delete the sink states m' from $\mathcal{A}$ and n from $\mathcal{B}$, and then form the concatenation of these modified DFAs. However, the same words used for distinguishing states in [8] can be used to distinguish states of $\mathcal{C}$. □

Theorem 13 (Suffix-Free Witness. Han and Salomaa, 2009 [12])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	$\mathcal{A}$:	$\mathcal{B}$:
a	$(2', \ldots, (m-1)')(1' \to m')$	$(1 \to n)$
b	$(1' \to m')$	$(2, \ldots, n-1)(1 \to n)$
c	$((Q^{\mathcal{A}} \setminus 1') \to m')(1' \to 2')$	$(1 \to n)$
d	$((Q^{\mathcal{A}} \setminus 2') \to m')$	$(1 \to 2)$

and let $F^{\mathcal{A}} = \{2'\}$ and $F^{\mathcal{B}} = \{2\}$. Then $\mathcal{C}$ has $(m-1)2^{n-2} + 1$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. For $k \le n-3$ we have

$$(1', \emptyset) \xrightarrow{cb} (2', \{1, n\}) \xrightarrow{d} (2', \{1, 2, n\}) \xrightarrow{b^k} (2', \{1, 2+k, n\}).$$

Thus $W = \{\varepsilon, d, db, \ldots, db^{n-3}\}$ is a construction set for $Q^{\mathcal{B}}$, with $s' = 2'$ and $B = \{1, n\}$. By Lemma 2, W is complete, since d and b act as permutations on $Q^{\mathcal{B}} \setminus \{1, n\}$.

There are 2^{n-2} states of the form $(2', S \cup \{1, n\})$ with $S \subseteq Q^{\mathcal{B}}$ and $S \cap \{1, n\} = \emptyset$. For each of these states, we reach $(q', S \cup n)$ for $3 \le q \le m-1$ by a^{q-2}, and $(m', S \cup n)$ by c. Adding in the initial state $(1', \emptyset)$ gives a total of $(m-1)2^{n-2}+1$ reachable states.

For distinguishability of the reached states, see [12]. Note that the authors of [12] work with a reduced concatenation DFA obtained by identifying, for each q' and S, the indistiguishable states (q', S) and $(q', S \cup n)$. Thus, for example, they write that $(q', \emptyset)$ is reachable for $3 \le q \le m-1$; these states are not reachable in our DFA $\mathcal{C}$, but states (q', n) for $3 \le q \le m-1$ are reachable. □

Theorem 14 (Right Ideal Witness. Brzozowski and Sinnamon, 2017 [9])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	*Final States*
$\mathcal{A}$:	$(1', \ldots, (m-1)')$	$(2' \to 1')$	$(\binom{m-1}{1} q' \to (q+1)')$	$\{m'\}$
$\mathcal{B}$:	$(1, \ldots, n-1)$	$(2 \to 1)$	$(\binom{n-1}{1} q \to q+1)$	$\{n\}$

Then $\mathcal{C}$ has $m + 2^{n-2}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. Note that $(1', \emptyset) \xrightarrow{a^{q-1}} (q', \emptyset)$ for $1 \leq q \leq m-1$, so these $m-1$ states are reachable. For $0 \leq k \leq n-3$ we have

$$((m-1)', \emptyset) \xrightarrow{c} (m', 1) \xrightarrow{a} (m', \{1, 2\}) \xrightarrow{(ab)^k} (m', \{1, 2+k\}).$$

Hence $\{\varepsilon, a, aab, a(ab)^2, \ldots, a(ab)^{n-3}\}$ is a construction set for $Q^{\mathcal{B}} \setminus n$, with $s' = m'$ and $B = \{1\}$. By Corollary 1, it is complete. Hence $(m', S \cup 1)$ is reachable for all $S \subseteq Q^{\mathcal{B}} \setminus n$.

We have reached $(m-1) + 2^{n-2}$ states so far. Additionally, we have $(m', \{1, n-1\}) \xrightarrow{cb} (m', \{1, n\})$, giving $m + 2^{n-2}$.

For distinguishability of the reached states, see [9]. □

Theorem 15 (Right Ideal Witness. Brzozowski, Davies and Liu, 2016 [4])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	c	*Final States*
$\mathcal{A}$:	$(1', \ldots, (m-1)')$	$(2', \ldots, (m-1)')$	$((m-1)' \to m')$	$\{m'\}$
$\mathcal{B}$:	$(1, \ldots, n-1)$	$(2, \ldots, n-1)$	$(n-1 \to n)$	$\{n\}$

Then $\mathcal{C}$ has $m + 2^{n-2}$ reachable and pairwise distinguishable states.

Proof. The initial state is $(1', \emptyset)$. Note that $(1', \emptyset) \xrightarrow{a^{q-1}} (q', \emptyset)$ for $1 \leq q \leq m-1$, so these $m-1$ states are reachable. For $0 \leq k \leq n-3$ we have

$$((m-1)', \emptyset) \xrightarrow{c} (m', 1) \xrightarrow{a} (m', \{1, 2\}) \xrightarrow{b^k} (m', \{1, 2+k\}).$$

Hence $\{\varepsilon, a, ab, ab^2, \ldots, ab^{n-3}\}$ is a construction set for $Q^{\mathcal{B}} \setminus n$, with $s' = m'$ and $B = \{1\}$. By Corollary 1, it is complete. Hence $(m', S \cup 1)$ is reachable for all $S \subseteq Q^{\mathcal{B}} \setminus n$.

We have reached $(m-1) + 2^{n-2}$ states so far. Additionally, we have $(m', \{1, n-1\}) \xrightarrow{c} (m', \{1, n\})$, giving $m + 2^{n-2}$.

For distinguishability of the reached states, see [4]. □

Theorem 16 (Right Ideal Witness. Brzozowski, Jirásková and Li, 2013 [5])**.** *Define $\mathcal{A}$ and $\mathcal{B}$ as follows:*

	a	b	*Final States*
$\mathcal{A}$:	$(\binom{m-1}{1} q' \to (q+1)')$	$(\binom{m-1}{1} q' \to (q+1)')$	$\{m'\}$
$\mathcal{B}$:	$(1, \ldots, n-1)$	$(\binom{n-1}{2} q \to q+1)$	$\{n\}$

Then $\mathcal{C}$ has $m + 2^{n-2}$ reachable and distinguishable states.

Proof. The initial state is $(1', \emptyset)$. Note that $(1', \emptyset) \xrightarrow{a^{q-1}} (q', \emptyset)$ for $1 \leq q \leq m-1$, so these $m-1$ states are reachable. For $0 \leq k \leq n-3$ we have

$$((m-1)', \emptyset) \xrightarrow{a} (m', 1) \xrightarrow{a} (m', \{1, 2\}) \xrightarrow{b^k} (m', \{1, 2+k\}).$$

Hence $\{\varepsilon, a, ab, ab^2, \ldots, ab^{n-3}\}$ is a construction set for $Q^{\mathcal{B}} \setminus n$, with $s' = m'$ and $B = \{1\}$. By Corollary 1, it is complete. Hence $(m', S \cup 1)$ is reachable for all $S \subseteq Q^{\mathcal{B}} \setminus n$.

We have reached $(m-1) + 2^{n-2}$ states so far. Additionally, we have $(m', \{1, n-1\}) \xrightarrow{b} (m', \{1, n\})$, giving $m + 2^{n-2}$.

For distinguishability of the reached states, see [5]. Note that the authors of [5] use a different concatenation DFA, constructed by removing state m' from $\mathcal{A}$ and then forming the concatenation in the usual way. However, the same words used in [5] can be used to distinguish states in $\mathcal{C}$. □

We now give two examples where our method of proof does not seem applicable or helpful. When attempting concatenation state complexity proofs, it seems best to consider both traditional techniques and the technique we present in this paper, switching between the two options if one does not yield an easy argument.

Example 1 (Prefix-Closed Witness. Brzozowski and Sinnamon, 2017 [9])**.** Our technique does not seem to work well with the following witness languages. Define $\mathcal{A}$ and $\mathcal{B}$ as follows:

	$\mathcal{A}$:	$\mathcal{B}$:
a	$(1', \ldots, (m-1)')$	$(1, \ldots, n-1)$
b	$(1', 2')$	$(2 \to 1)$
c	$(2' \to 1')$	$(\binom{n-1}{1} q \to q-1)$
d	$(\binom{m-1}{1} q' \to (q-1)')$	$(1, 2)$

and let $F^{\mathcal{A}} = \{1', \ldots, (m-1)'\}$ and $F^{\mathcal{B}} = \{1, \ldots, n-1\}$.

The inductive proof given by the authors of [9] has a different structure from the type of argument captured by Theorem 1. To reach a state (q', S), in Theorem 1 we start from some state (q', B) and apply a word that fixes the first component q'. In [9] the authors instead start from a state (p', B) and apply a word w such that $p'w = q'$. The proof in [9] is short and clean, whereas a proof in the style of Theorem 1 seems to require complicated arguments. It is possible that Theorem 1 could be generalized to cover arguments of the form used in [9], but we have not found such a generalization.

Example 2 (Finite Binary Witness. Câmpeanu, Culik, Salomaa and Yu, 2001 [10])**.** Our technique does not apply to the following witness languages. Define $\mathcal{A}$ and $\mathcal{B}$ as follows:

	a	b	Final States
$\mathcal{A}$:	$(\binom{m-1}{1} q' \to (q+1)')$	$(\binom{m-1}{1} q' \to (q+1)')$	$\{1', \ldots, (m-1)'\}$
$\mathcal{B}$:	$(\binom{n-1}{2} q \to q+1)(1 \to n)$	$(\binom{n-1}{1} q \to q+1)$	$\{n-1\}$

Additionally, assume that $m + 1 \geq n > 2$. Then $\mathcal{C}$ has $(m-n+3)2^{n-2} - 1$ reachable and pairwise distinguishable states. This is the maximum for finite languages over a binary alphabet when $m + 1 \geq n > 2$.

Let us consider why Theorem 1 cannot be used here. The point of Theorem 1 is to build up states (s', S) by starting from (s', B) and using words that fix the focus state s'. But in this witness, no state of $\mathcal{A}$ is fixed by any word except for the non-final sink state m'. So to use Theorem 1, the focus state must be m'. But from a state of the form (m', S), we can only reach sets (m', T) with $|T| \leq |S|$, since m' is a non-final sink state. So there is no way to start from some base state (m', B) and build up larger sets, which is the strategy of Theorem 1.

5 Conclusions

We have introduced a new technique for demonstrating the reachability of states in DFAs for the concatenation of two regular languages, and provided evidence that this technique is useful in a wide variety of cases. However, we found two cases (Examples 1 and 2) where our technique does not seem applicable. Example 1 in particular suggests that Theorem 1 may admit a generalization that covers more types of inductive proofs. We leave this as an open problem.

Acknowledgements

I thank Jason Bell and Janusz Brzozowski for proofreading and helpful comments. This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871.

References

[1] J. A. Brzozowski. In search of most complex regular languages. *Int. J. Found. Comput. Sc.*, 24(06):691–708, 2013.

[2] J. A. Brzozowski. Unrestricted state complexity of binary operations on regular languages. In C. Câmpeanu, F. Manea, and J. Shallit, editors, *DCFS 2016*, volume 9777 of *LNCS*, pages 60–72. Springer, 2016.

[3] J. A. Brzozowski and S. Davies. Most complex non-returning regular languages. In G. Pighizzini and C. Câmpeanu, editors, *DCFS 2017*, volume 10316 of *LNCS*, pages 89–101. Springer, 2017.

[4] J. A. Brzozowski, S. Davies, and B. Y. V. Liu. Most complex regular ideal languages. *Discrete Math. Theoret. Comput. Sc.*, 18(3), 2016. Paper #15.

[5] J. A. Brzozowski, G. Jirásková, and B. Li. Quotient complexity of ideal languages. *Theoret. Comput. Sci.*, 470:36–52, 2013.

[6] J. A. Brzozowski, G. Jirásková, and C. Zou. Quotient complexity of closed languages. *Theory Comput. Syst.*, 54:277–292, 2014.

[7] J. A. Brzozowski and B. Liu. Quotient complexity of star-free languages. *Int. J. Found. Comput. Sc.*, 23(06):1261–1276, 2012.

[8] J. A. Brzozowski and C. Sinnamon. Complexity of left-ideal, suffix-closed and suffix-free regular languages. In F. Drewes, C. Martín-Vide, and B. Truthe, editors, *LATA 2017*, volume 10168 of *LNCS*, pages 171–182. Springer, 2017.

[9] J. A. Brzozowski and C. Sinnamon. Complexity of right-ideal, prefix-closed, and prefix-free regular languages. *Acta Cybernetica*, 23(1):9–41, 2017.

[10] C. Câmpeanu, K. Culik, K. Salomaa, and S. Yu. State complexity of basic operations on finite languages. In O. Boldt and H. Jürgensen, editors, *WIA 1999*, volume 2214 of *LNCS*, pages 60–70. Springer, 2001.

[11] H.-S. Eom, Y.-S. Han, and G. Jirásková. State complexity of basic operations on non-returning regular languages. *Fund. Inform.*, 144:161–182, 2016.

[12] Y.-S. Han and K. Salomaa. State complexity of basic operations on suffix-free regular languages. *Theoret. Comput. Sci.*, 410(27-29):2537–2548, 2009.

[13] Y.-S. Han, K. Salomaa, and D. Wood. Operational state complexity of prefix-free regular languages. In Z. Ésik and Z. Fülöp, editors, *AFL 2009*, pages 99–115. Institute of Informatics, University of Szeged, Hungary, 2009.

[14] G. Jirásková and M. Krausová. Complexity in prefix-free regular languages. In I. McQuillan, G. Pighizzini, and B. Trost, editors, *DCFS 2010*, pages 236–244. University of Saskatchewan, 2010.

[15] A. N. Maslov. Estimates of the number of states of finite automata. *Dokl. Akad. Nauk SSSR*, 194:1266–1268 (Russian), 1970. English translation: Soviet Math. Dokl. 11(1970) 1373–1375.

[16] C. Nicaud. Average state complexity of operations on unary automata. In M. Kutyłowski, L. Pacholski, and T. Wierzbicki, editors, *MFCS 1999*, pages 231–240. Springer, 1999.

[17] G. Pighizzini and J. Shallit. Unary language operations, state complexity and Jacobsthal's function. *International Journal of Foundations of Computer Science*, 13(01):145–159, 2002.

[18] S. Yu. State complexity of regular languages. *J. Autom. Lang. Comb.*, 6:221–234, 2001.

[19] S. Yu, Q. Zhuang, and K. Salomaa. The state complexities of some basic operations on regular languages. *Theor. Comput. Sci.*, 125(2):315–328, 1994.