# Determinization of Integral Discounted-Sum Automata is Decidable***

Shaull Almagor and Neta Dafni (✉)

Technion, Haifa, Israel
shaull@technion.ac.il
netad@campus.technion.ac.il

**Abstract.** Nondeterministic Discounted-Sum Automata (NDAs) are nondeterministic finite automata equipped with a discounting factor $\lambda > 1$, and whose transitions are labelled by weights. The value of a run of an NDA is the discounted sum of the edge weights, where the $i$-th weight is divided by $\lambda^i$. NDAs are a useful tool for modelling systems where the values of future events are less influential than immediate ones.
While several problems are undecidable or open for NDA, their deterministic fragment (DDA) admits more tractable algorithms. Therefore, determinization of NDAs (i.e., deciding if an NDA has a functionally-equivalent DDA) is desirable.
Previous works establish that when $\lambda \in \mathbb{N}$, then every *complete* NDA, namely an NDA whose states are all accepting and its transition function is complete, is determinizable. This, however, no longer holds when the completeness assumption is dropped.
We show that the problem of whether an NDA has an equivalent DDA is decidable when $\lambda \in \mathbb{N}$ (in particular, it is in EXPSPACE and is PSPACE−hard).

**Keywords:** Discounted Sum Automata · Determinization · Quantitative Automata

## 1 Introduction

Traditional methods of modelling systems rely on Boolean automata, where every word is assigned a Boolean value (i.e., accepted or rejected). This setting is often generalized into a richer, quantitative one, where every word is assigned a numerical value, and thus the Boolean concept of a language, i.e., a set of words, is lifted to a more general function, namely a function from words to values.

A particular instance of quantitative automata is that of *discounted-sum automata*. There, the weight function sums the weights along the run, but discounts the future. Discounting as a general notion is a well studied concept in game theory and various social choice models [9]. Computational models with discounting, such as discounted-payoff games [21, 3, 1], discounted-sum Markov

* This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 989/22)

** The full version can be found on https://arxiv.org/abs/2310.09115


N. Kobayashi and J. Worrell (Eds.): FoSSaCS 2024, LNCS 14574, pp. 191–211, 2024.
https://doi.org/10.1007/978-3-031-57228-9_10

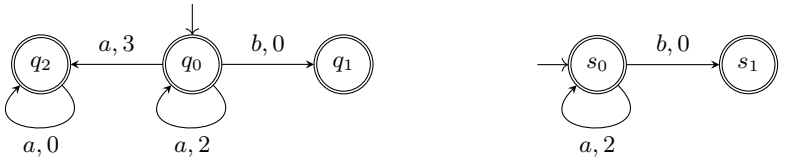
Decision Processes [17, 19, 14] and discounted-sum automata [15, 12, 13, 11], are therefore useful to model settings where the far future has less influence than the immediate future.

In this work we focus on non-deterministic discounted-sum automata (NDAs). An NDA is a quantitative automaton equipped with a *discounting factor* $\lambda > 1$. The value of a run is the discounted sum of the transitions along the run, where the value of transition $i$ is divided by $\lambda^i$. The value of a word is then the value of the minimal accepting run on it. We also allow *final weights* that are added to the run at its end (with appropriate discounting).

Unlike Boolean automata, NDAs are strictly more expressive than their deterministic counterpart (DDAs) [11]. In particular, certain decision problems for NDAs are undecidable, but become decidable for DDAs [5]. There is, however, a subclass of NDAs that always admit an equivalent DDA: the *complete integral NDAs* [6]. An automaton is *complete* if its transition function is total and all its states are accepting with final weight 0. This means that runs never "die", and that all runs are accepting. An NDA is *integral* if its discounting factor $\lambda$ is an integer. It is further shown in [6] that if the completeness requirement is removed then for every discounting factor there is an integral NDA that is not determinizable.

The existence of NDAs that are not determinizable implies that the determinization problem is not trivial. However, its decidability and complexity have not been studied. In this work, we show that determinization of integral NDAs is decidable. Specifically, we show that determinization is in EXPSPACE and is PSPACE − hard.

*Example 1.* We demonstrate the determinization problem, as well as some intricacies involved in its analysis. Consider the NDA in Figure 1a. Intuitively, the NDA either reads only $a$'s, or reads a word of the form $a^*b$. However, it guesses in $q_0$ whether it is going to read many $a$'s, in which case it may be worthwhile incurring weight 3 to $q_2$ in order to read the remaining $a$'s at cost 0.




(a) NDA $\mathcal{A}$. (b) Equivalent DDA for $\lambda = 3$.

Fig. 1: The NDA $\mathcal{A}$ on the left is determinizable with $\lambda = 3$, with an equivalent DDA depicted on the right. However, $\mathcal{A}$ is not determinizable with $\lambda = 2$.

We now ask if this NDA has a deterministic equivalent. As it turns out, this is dependent on the discounting factor. Indeed, consider the discounting factor $\lambda = 3$, then when reading the word $a^k$, the run that remains in $q_0$ has weight $\sum_{i=0}^{k} 2 \cdot 3^{-i} = 3 - 3^{-k} < 3$, whereas a run that moves to $q_2$ at step $j \geq 0$ has weight $\sum_{i=0}^{j-1} 2 \cdot 3^{-i} + 3 \cdot 3^{-j} = 3 - 3^{-j} + 3^{-j} = 3$. Thus, it is always

beneficial to remain in $q_0$. In this case, we do have a deterministic equivalent, depicted in Figure 1b. We remark that the fact that this deterministic equivalent is obtained by removing transitions is not a standard behaviour, and typically determinization involves a blowup.

Next, consider the discounting factor $\lambda = 2$. Similar analysis shows that for the word $a^k$, the weight of the run that stays in $q_0$ is $\sum_{i=0}^{k} 2 \cdot 2^{-i} = 4 - 2 \cdot 2^{-k}$, whereas leaving to state $q_2$ at step 0 yields cost 3, so the latter is preferable for large $k$. Intuitively, this means that nondeterminism is necessary in this setting, since the NDA does not "know" whether $b$ will be seen. Indeed, for $\lambda = 2$ this NDA is not determinizable.

Observe that in the case of $\lambda = 2$, the two "extreme" runs on $a^k$, namely the one that stays in $q_0$ and the one that immediately leaves to $q_2$, create a "gap" between their values that tends toward 1 as $k$ increases. Keeping in mind that for large $k$ the transition value is multiplied by $2^{-k}$, intuitively this gap becomes huge. As we show in this work, this concept of gaps exactly characterizes whether an NDA can be determinized. □

We remark that for non-integral NDAs, many problems, including the determinization problem, are open due to number-theoretic difficulties [8]. Therefore, it is unlikely that progress is made there, pending breakthroughs in number theory.

*Related Work* Discounted-sum automata have been studied in various contexts. Specifically, certain algorithmic problems for them are still open, and are closely related to longstanding open problems [8]. In addition, they are not closed under standard Boolean operations [7] (which is often the case in quantitative models, due to the "minimum" semantics which conflicts with notions of conjunction).

Recently, discounted sum automata were also studied in the context of two-player games [10]. Of particular interest are "regret-minimizing strategies", where the concept of regret minimization is closely related to determinization of automata [16].

An extension of discounted-sum automata to multiple discounting factors (NMDAs) was studied in [4, 5], where NMDAs are NDAs where every transition is allowed a different discounting factor. NMDAs are generally non-determinizable, but imposing certain restrictions on the choice of discounting factors can ensure determinizability [4]. We remark that the study of NMDAs is still only with respect to complete automata.

Determinization of other quantitative models has also received some attention in recent years. A major open problem is the decidability of determinization for weighted automata over the tropical semiring (for some subclasses it is known to be decidable [20, 18]). Interestingly, a tropical weighted automaton can be seen as the "limit" of NDAs where $\lambda \to 1$. This, however, does not seem to help in resolving the decidability of the former.

In [2], the determinization problem for one-counter nets (OCNs) is studied. OCNs are automata equipped with a *counter* that cannot decrease below zero. They can be thought of as pushdown automata with a singleton stack alphabet.

Most notions of determinizability introduced in [2] are undecidable, with one case being open (and seemingly related to the setting of weighted automata).

Due to space constraints, some proofs appear in the full version.

## 2 Preliminaries

A *nondeterministic integral discounted-sum automaton* (NDA) is a tuple $\mathcal{A} = (\Sigma, Q, Q_0, \alpha, \delta, \mathtt{val}, \mathtt{fval}, \lambda)$, where $\Sigma$ is a finite alphabet, $Q$ is a finite set of states, $Q_0 \subseteq Q$ is a set of initial states, $\alpha \subseteq Q$ is a set of accepting states, $\delta \subseteq Q \times \Sigma \times Q$ is a transition relation, $\mathtt{val} : \delta \to \mathbb{Z}$ is a *weight function* that assigns to each transition $(p, \sigma, q) \in \delta$ a *weight* $\mathtt{val}((p, \sigma, q)) \in \mathbb{Z}$, $\mathtt{fval} : \alpha \to \mathbb{Z}$ is a *final weight function* that assigns a *final weight*[1] to every accepting state, and $1 < \lambda \in \mathbb{N}$ is an integer *discounting factor*.

The existence of a transition $(p, \sigma, q) \in \delta$ means that when $\mathcal{A}$ is in state $p$ and reads the letter $\sigma$ it can move to state $q$. If there exists $q$ such that $(p, \sigma, q) \in \delta$, we say that $p$ has a *$\sigma$-transition*. If $p$ does not have a $\sigma$-transition, that means that when in state $p$ and reading the letter $\sigma$, $\mathcal{A}$'s run cannot continue.

Consider a word $w = w_1 \cdots w_n \in \Sigma^*$. A *run* of $\mathcal{A}$ on $w$ is a sequence of states $\rho = \rho_0, \rho_1, \ldots, \rho_n$ such that $\rho_0 \in Q_0$ and $(\rho_{i-1}, w_i, \rho_i) \in \delta$ for every $1 \leq i \leq n$. The run is *accepting* if $\rho_n \in \alpha$. The *weight* of $\rho$ is the discounted sum $\mathtt{val}(\rho) = \Sigma_{i=0}^{n-1} \lambda^{-i} \mathtt{val}(\rho_i, w_{i+1}, \rho_{i+1})$.

The *value* of $w$ by $\mathcal{A}$, denoted $\mathcal{A}^*(w)$, is $\min\{\mathtt{val}(\rho) + \lambda^{-n}\mathtt{fval}(\rho_n) \mid \rho = \rho_0, \ldots, \rho_n$ is an accepting run on $w\}$, that is, the minimal weight of a run on $w$ including final weights, or $\infty$ if no such run exists. Two NDAs $\mathcal{A}, \mathcal{B}$ are *equivalent* if $\mathcal{A}^*(w) = \mathcal{B}^*(w)$ for every $w \in \Sigma^*$.

We say that $\mathcal{A}$ is *deterministic* (DDA, for short) if $|Q_0| = 1$ and $\{q \in Q | (p, \sigma, q) \in \delta\} \leq 1$ for every $p \in Q, \sigma \in \Sigma$. Note that if $\mathcal{A}$ is deterministic then for every word there is at most one run starting in each state. For a DDA we define the partial function $\delta^* : Q \times \Sigma^* \hookrightarrow Q$ such that $\delta^*(q, w)$ is the final state in the run on $w$ starting in $q$, if such a run exists. We say that an NDA $\mathcal{A}$ is *determinizable* if it has an equivalent DDA.

It will also be useful to consider non-accepting runs and runs that start and end in specific states of $\mathcal{A}$. For sets of states $P, P' \subseteq Q$ we define $\mathcal{A}_{[P \to P']}(w)$ to be the weight of a minimal run of $\mathcal{A}$ on $w$ from some state in $P$ to some state in $P'$. Similarly, $\mathcal{A}_{[P \to_f P']}(w)$ is the minimal weight of an accepting run including final weights. When $P$ or $P'$ is a singleton $\{p\}$ we omit the parenthesis. When $P = Q_0$ and $P' = Q$ (or $\alpha$, for the setting of including final weights) we omit the sets and write e.g., $\mathcal{A}(w)$ instead of $\mathcal{A}_{[Q_0 \to Q]}(w)$, and $\mathcal{A}^*(w)$ instead of $\mathcal{A}_{[Q_0 \to_f \alpha]}(w)$. Under these notations, if a run does not exist, the assigned value is $\infty$. For the remainder of the paper, fix an integral NDA $\mathcal{A}$.

[1] In some works, the weights are assumed to be rational. For determinizability we can assume all weights are integers, since we can always multiply every weight by a common denominator.

## 3 Gaps and Separation of Runs

In this section we lay down the basic definitions we use throughout the paper, concerning the ways several runs on the same word accumulate different weights.

Denote by $m_{\mathcal{A}}$ the maximal absolute value of a weight of a transition or a final weight in $\mathcal{A}$. Recall that the geometric sum (for $\lambda > 1$) satisfies $\sum_{i=0}^{\infty} \lambda^{-i} = \frac{\lambda}{1-\lambda}$. Therefore, $\frac{\lambda}{\lambda-1} m_{\mathcal{A}}$ is an upper bound on $|\mathtt{val}(\rho)|$ for any run $\rho$. Indeed, we have

$$|\mathtt{val}(\rho_0, \ldots, \rho_n)| = |\Sigma_{i=0}^{n-1} \lambda^{-i} \mathtt{val}(\rho_i, w_{i+1}, \rho_{i+1})| \leq \Sigma_{i=0}^{n-1} \lambda^{-i} m_{\mathcal{A}} < \frac{\lambda}{\lambda - 1} m_{\mathcal{A}}$$

Clearly, the same bound holds when including final weights.

Let $\mathcal{M} = 2\frac{\lambda}{\lambda-1} m_{\mathcal{A}}$, then for every two runs $\rho^1, \rho^2$ we have $|\mathtt{val}(\rho^1) - \mathtt{val}(\rho^2)| < \frac{\lambda}{\lambda-1} m_{\mathcal{A}} - (-\frac{\lambda}{\lambda-1} m_{\mathcal{A}}) = \mathcal{M}$. The constant $\mathcal{M}$ is central in our study of gaps between runs.

Consider a word $w \in \Sigma^*$. The run attaining the minimal value $\mathcal{A}^*(w)$ might not be minimal while reading prefixes of $w$. The *gap* between the value of an eventually-minimal run and minimal runs on prefixes of $w$ is central to characterizing determinizability of NDAs [6]. This gap is captured by the following definition.

**Definition 1 (Recoverable gap).** *Consider words $w, z \in \Sigma^*$ and states $q_u, q_l \in Q$. the tuple $(w, q_u, q_l)$ is called a* recoverable gap with respect to $z$, *or simply a* recoverable gap, *if the following hold:*

1. $\mathcal{A}_{[Q_0 \to q_l]}(w) \leq \mathcal{A}_{[Q_0 \to q_u]}(w)$, *and*
2. $\mathcal{A}_{[Q_0 \to q_u]}(w) + \lambda^{-|w|} \mathcal{A}_{[q_u \to_f \alpha]}(z) = \mathcal{A}^*(wz) < \infty$.

Intuitively, in a recoverable gap $(w, q_u, q_l)$ there are runs $\rho_1$ and $\rho_2$ of $\mathcal{A}$ on $w$ that end in $q_u$ and $q_l$, respectively, where $\rho_1$ attains a higher value than $\rho_2$, but there is a suffix $z$ that "recovers" this gap: when reading $z$ from $q_u$ starting with weight $\mathtt{val}(\rho_1)$, the resulting minimal run including final weight attains the minimal value of a run of $\mathcal{A}$ on $w \cdot z$. This is depicted in Figure 2.

For a recoverable gap $(w, q_u, q_l)$ we define $\mathtt{gap}(w, q_u, q_l) = \lambda^{|w|}(\mathcal{A}_{[Q_0 \to q_u]}(w) - \mathcal{A}_{[Q_0 \to q_l]}(w))$. The normalizing factor $\lambda^{|w|}$ eliminates the effect of the length of $w$ on the gap, allowing us to study gaps independently of the length of their corresponding words.

We say that $\mathcal{A}$ has *finitely/infinitely many recoverable gaps* if the set $\{\mathtt{gap}(w, q_u, q_l) \mid w \in \Sigma^*, q_u, q_l \in Q\}$ is finite/infinite, respectively. Note that since $\mathcal{A}$ is integral, $\lambda^{|w|}(\mathcal{A}_{[Q_0 \to q_u]}(w) - \mathcal{A}_{[Q_0 \to q_l]}(w))$ is always an integer and so the existence of infinitely many recoverable gaps is equivalent to the existence of unboundedly large recoverable gaps.

While gaps refer to two distinct runs, we sometimes need a more global view of gaps. To this end, we lift the definition to all the reachable states, as follows.

**Definition 2 ($n$-separation).** *For a word $w$ and $n \in \mathbb{N}$, we say that $w$ has the* $n$-separation property *if there exists a partition of $Q$ into two non-empty sets of states $U, L$ such that the following holds:*

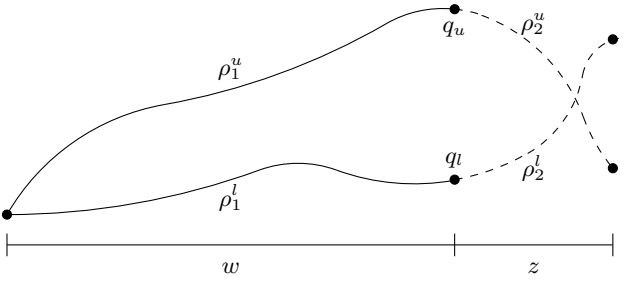



Fig. 2: The run $\rho_1^l$, ending in state $q_l$, is the minimal run of $\mathcal{A}$ on $w$. The higher run $\rho_1^u$ is the minimal run on $w$ that ends in state $q_u$, thus creating a gap between $q_u$ and $q_l$. However, the concatenation $\rho_1^u \cdot \rho_2^u$ is the minimal run on the concatenated word $wz$, while the concatenation $\rho_1^l \cdot \rho_2^l$, where $\rho_2^l$ is the minimal run on $z$ starting in $q_l$, is not smaller. Therefore, the gap is recoverable. Note that here the final weights are zero.

1. *For every* $q_u \in U$ *and* $q_l \in L$, $\lambda^{|w|}(\mathcal{A}_{[Q_0 \to q_u]}(w) - \mathcal{A}_{[Q_0 \to q_l]}(w)) > n$.
2. *There exist* $q_u \in U$ *and* $z \in \Sigma^*$ *such that for every* $q_l \in L$, $(w, q_u, q_l)$ *is a recoverable gap with respect to* $z$.

*We sometimes explicitly specify that* $w$ has the $n$-separation property with respect to $(U, L, q_u)$, *or* with respect to $(U, L, q_u, z)$. *If there exists* $w$ *with the* $n$*-separation property, we say that* $\mathcal{A}$ *has the* $n$*-separation property.*

See Figure 3 for a depiction of $n$-separation.

## 4 Determinizability of Integral NDAs is Decidable – Proof Overview

Recall that our goal is to show the decidability of the determinization problem.

As showed in [6], determinizability is closely related to recoverable gaps. More precisely, a DDA $\mathcal{D}$ that "attempts" to be equivalent to $\mathcal{A}$ must keep track of all the relevant runs of $\mathcal{A}$. If two runs end in the same state, it is clearly enough to track only the minimal one. However, this may still require keeping track of runs that attain unboundedly high values (when normalized). Therefore, in order for $\mathcal{D}$ to be finite, it must discard information on runs that get too high. The main issue is whether we can give a bound above which runs are no longer relevant.

For *complete* integral NDAs, there are always finitely many recoverable gaps, and this is used to show that complete NDAs are always determinizable [6]. For a general integral NDA $\mathcal{A}$, we similarly show in Section 5 that if there are only finitely many recoverable gaps, then $\mathcal{A}$ is determinizable.

There are now two main challenges: First, to show that if $\mathcal{A}$ has infinitely many recoverable gaps, then it is not determinizable, and second, that it is decidable whether $\mathcal{A}$ has finitely many recoverable gaps.

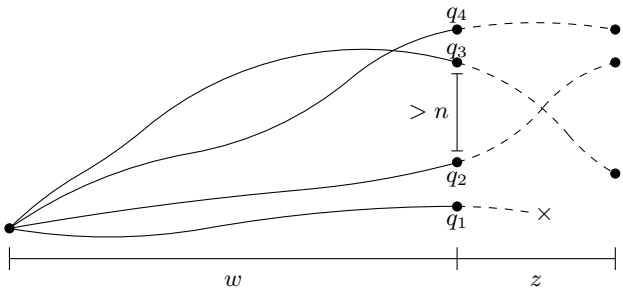



Fig. 3: Depicted are minimal runs of an NDA on a word $w$ that end in each of four states, $q_1, q_2, q_3, q_4$, and minimal runs on $z$ starting in each of them. The run from $q_1$ (lowest) "gets stuck", i.e., such a run from $q_1$ on $z$ does not exist. The states are partitioned into two sets $L = \{q_1, q_2\}$ and $U = \{q_3, q_4\}$, with a gap larger than $n$ between them after reading $w$; additionally, one of the upper runs then becomes minimal after reading $z$, since each of the lower runs either ends higher or "gets stuck". This means that the word $w$ has the $n$-separation property with respect to $(U, L, q_3, z)$.

We start by showing in Section 6.1 that we can compute a bound $\mathcal{N}$ such that $\mathcal{A}$ has infinitely many recoverable gaps if and only if it has a recoverable gap larger than $\mathcal{N}$. Next, in Section 6.2, we show that the existence of a gap larger than $\mathcal{N}$ is also equivalent to some word having the $\mathcal{N}$-separation property.

We then turn to exhibit a small-model property on witnesses for $\mathcal{N}$-separation. Specifically, we show in Section 7 that if there exist $w, z$ such that $w$ has the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z)$, then we can bound the length of the shortest $w, z$.

Using the above, we obtain the decidability of whether $\mathcal{A}$ has infinitely many recoverable gaps. In addition, we use these results to prove (in Lemma 11) that if $\mathcal{A}$ has infinitely many recoverable gaps, then it is not determinizable. This allows us to conclude the decidability of determinization in Theorem 1.

Conceptually, our approach can be viewed as a "standard" one when treating determinization of quantitative models, in the sense that considering gaps between runs generally characterizes when a deterministic equivalent exists [16, 2]. The crux is showing that this condition is decidable. To this end, our work greatly differs from other works on weighted automata in that we establish the decidability of the condition. Technically, this involves careful analysis of the behaviours of runs under discounting.

## 5 Finitely Many Recoverable Gaps Imply Determinizability

The main result of this section is an adaptation of the determinization techniques in [6] from complete to general automata. While the construction itself is similar, the correctness proof requires finer analysis. We remark that in the case that $\mathcal{A}$ is a complete NDA and all final weights are zero, the construction obtains a complete DDA with all final weights zero, thus generalizing the result in [6].

**Lemma 1.** *If an NDA $\mathcal{A}$ has finitely many recoverable gaps, then it is determinizable.*

*Proof.* Let $\mathcal{A} = (\Sigma, Q, Q_0, \alpha, \delta, \texttt{val}, \texttt{fval}, \lambda)$ be an NDA with finitely many recoverable gaps. We construct a DDA $\mathcal{D} = (\Sigma, Q_D, \{v_0\}, \alpha_D, \delta_D, \texttt{val}_D, \texttt{fval}_D, \lambda)$ that is equivalent to $\mathcal{A}$.

Since $\mathcal{A}$ has finitely many recoverable gaps, there exists a bound $B \in \mathbb{N}$ on the size of those gaps. The states of $\mathcal{D}$ are then $Q_D = \{0, \ldots, B, \infty\}^Q$. Intuitively, a run of $\mathcal{D}$ tracks, for each $q \in Q$, the gap between the minimal run of $\mathcal{A}$ on $w$ ending in $q$ and the minimal run on $w$ overall. When this gap becomes too large to be recoverable, the states corresponding to the higher run are assigned $\infty$. For $v \in Q_D$ and $q \in Q$, we denote by $(v_q)$ the entry in $v$ corresponding to $q$. The initial state is therefore $(v_0)_q = \begin{cases} 0 & q \in Q_0 \\ \infty & q \notin Q_0 \end{cases}$, assigning for each $q \in Q$ the weight of the minimal run of $\mathcal{A}$ on the empty word ending in $q$.

We now turn to define $\delta_D$. Intuitively, when taking a transition, $\mathcal{D}$ D first updates the vector entry of every state with the value of the minimal run on the new word ending in it, using the values specified in the last vector. Then, if the minimal entry is not 0, the entries are shifted so that it becomes 0, and the value subtracted from every entry is assigned to the transition weight. Finally, the entries are all multiplied by $\lambda$ to account for the word length. Thus, the actual value of the minimal run is exactly the value attained by $\mathcal{D}$, and the vector entries correctly represent the normalized gaps. The construction is demonstrated in Figure 4. Formally:

- For every $v \in Q_D$, and for every $\sigma \in \Sigma$ such that there exists $q \in Q$ with $v_q < \infty$ and $q$ has a $\sigma$-transition, define $u \in \{0, \ldots, B, \infty\}^Q$ as follows.
  - Define the intermediate vector $u'$: For every $q \in Q$, $u'_q = \min_{q' \in Q}(v_{q'} + \texttt{val}(q', \sigma, q))$, where $\texttt{val}(q', \sigma, q)$ is regarded as $\infty$ if $(q', \sigma, q) \notin \delta$.
  - Define $r = \min_{q \in Q} u'_q$, the offset of the vector from 0. Note that $r$ is finite due to the requirement that there exists $q \in Q$ with $v_q < \infty$ and $q$ has a $\sigma$-transition.
  - For every $q \in Q$, $u_q = \begin{cases} \lambda(u'_q - r) & \lambda(u'_q - r) \leq B \\ \infty & \text{otherwise} \end{cases}$

  Where $\infty$ is handled using the standard semantics. Note that $u \in \{0, \ldots, B, \infty\}^Q$ as $\mathcal{A}$ is integral. The manipulations done on the intermediate vector $u'_q$ when defining $u_q$ should be viewed as normalization – first subtracting $r$

so that the gap represented by $u_q$ is with respect to the minimal run overall over $w$; then multiplying by $\lambda$ to account for the length of $w$. Note that the subtraction of $r$ also implies that $\min_{q \in Q} u_q = 0$, as is expected since $\min_{q \in Q} \mathcal{A}_{[Q_0 \to q]}(w) = \mathcal{A}(w)$.
- We now introduce the transition $(v, \sigma, u) \in \delta_D$.
- We set $\mathtt{val}_D(v, \sigma, u) = r$. This can be viewed, together with the subtraction of $r$ from every entry of $u'$, as transferring the weight from each entry of $u'$ to the transition.


(a) an NDA $\mathcal{A}$. (b) An equivalent DDA $\mathcal{D}$.

Fig. 4: An example of an NDA $\mathcal{A}$ (on the left) and the resulting DDA $\mathcal{D}$ (on the right), with $\lambda = 2$. The name of each state of $\mathcal{D}$ corresponds to a vector whose first entry tracks $q_0$ and the second $q_1$. We demonstrate the construction using the $a$-transition from $(2, 0)$ to itself. First we construct the intermediate vector $u'$: $(u')_{q_0} = \min(2 + \mathtt{val}(q_0, a, q_0), 0 + \mathtt{val}(q_1, a, q_0)) = \min(2 + 1, 0 + \infty) = 3$ and $(u')_{q_1} = \min(2 + \mathtt{val}(q_0, a, q_1), 0 + \mathtt{val}(q_1, a, q_1)) = \min(2 + 0, 0 + \infty) = 2$, and so $u' = (3, 2)$. We then have $r = 2$, which is assigned to the weight of the transition, and $u = 2(3 - 2, 2 - 2) = (2, 0)$.

We next define $\alpha_D$ and $\mathtt{fval}_D$. We set $\alpha_D$ to include every vector $v$ such that $v_q < \infty$ for some $q \in \alpha$. We note that the construction can be viewed as a generalization of the standard subset construction, where for a vector $v$, the states $q$ that satisfy $v_q < \infty$ represent the states that can be reached by $\mathcal{A}$ when reading $w$, ignoring those states whose gap is unrecoverable. For $v \in \alpha_D$, we set $\mathtt{fval}_D(v) = \min_{q \in \alpha}(v_q + \mathtt{fval}(q))$. Figure 4 depicts an example for an NDA and the DDA constructed from it (with no final weights). Note that we do not yet actually provide an algorithm for constructing $\mathcal{D}$ from $\mathcal{A}$, since that requires computing $B$.

The correctness of this construction is proved in the full version. □

## 6 Recoverable Gaps and $n$-separation

### 6.1 A Large Gap is Equivalent to Infinitely Many Gaps

In this section we show that the existence of infinitely many recoverable gaps is characterized by the existence of a (computable) large-enough recoverable gap.

Consider a run $\rho = \rho_0 \dots \rho_n$, and recall that $\mathtt{val}(\rho)$ is the weight of $\rho$ and that $\mathcal{M} = 2\frac{\lambda}{\lambda-1}m_{\mathcal{A}}$, where $m_{\mathcal{A}}$ is the maximal absolute value of a weight of a transition or a final weight in $\mathcal{A}$. We denote by $\Gamma(\rho) = \lambda^n \mathtt{val}(\rho)$ the normalized "un-discounted" value of $\rho$. For two runs $\rho^1, \rho^2$ on the same word $w$, we are interested in the value $\Gamma(\rho^1) - \Gamma(\rho^2)$, as it captures how far the runs are from each other, in the sense of how difficult it is to recover their gap. We claim that if two runs get too far from each other, the gap between them from that point on can only increase. Intuitively, this is because at each step the value is multiplied by $\lambda$, and so beyond a certain gap size, this multiplication separates the runs further even if their added values pull them closer before multiplying by $\lambda$.

**Lemma 2.** *Let $\rho^1 = \rho^1_0, \dots, \rho^1_{n+1}$ and $\rho^2 = \rho^2_0, \dots, \rho^2_{n+1}$ be two runs of $\mathcal{A}$, such that $\Gamma(\rho^1_0, \dots, \rho^1_n) - \Gamma(\rho^2_0, \dots, \rho^2_n) > \mathcal{M}$. Then $\Gamma(\rho^1_0, \dots, \rho^1_{n+1}) - \Gamma(\rho^2_0, \dots, \rho^2_{n+1}) > \Gamma(\rho^1_0, \dots, \rho^1_n) - \Gamma(\rho^2_0, \dots, \rho^2_n)$.*

In particular, once the gap between $\rho^1, \rho^2$ is larger than $\mathcal{M}$, concatenating any runs to $\rho^1, \rho^2$ can only increase the gap and therefore cannot result in $\rho^1$ "bypassing" $\rho^2$.

**Corollary 1.** *Let $\rho^1 = \rho^1_0, \dots, \rho^1_n$ and $\rho^2 = \rho^2_0, \dots, \rho^2_n$ be two runs such that $\mathtt{val}(\rho^1) \leq \mathtt{val}(\rho^2)$. Then for every $0 \leq i \leq n$, it holds that $\Gamma(\rho^1_0, \dots, \rho^1_i) - \Gamma(\rho^2_0, \dots, \rho^2_i) \leq \mathcal{M}$.*

On the other hand, the gap between two runs cannot increase too much within a small number of steps. We capture the contra-positive of this, by showing that if two runs reach a large enough gap, then the runs have been far from each other for a long suffix.

**Lemma 3.** *Consider $n_{steps}, n_{gap} \in \mathbb{N}$, there exists an effectively computable number $N$ such that for any two runs $\rho^1 = \rho^1_0, \dots, \rho^1_n$ and $\rho^2 = \rho^2_0, \dots, \rho^2_n$, if $\Gamma(\rho^1) - \Gamma(\rho^2) > N$ then $n > n_{steps}$ and $\Gamma(\rho^1_0, \dots, \rho^1_{n-n_{steps}}) - \Gamma(\rho^2_0, \dots, \rho^2_{n-n_{steps}}) > n_{gap}$.*

We also need a version of Corollary 1 where the inequality between the weights of the runs includes final weights. We claim that concatenating any runs to runs that are far from each other cannot result in the lower run "bypassing" the upper run, including final weights:

**Lemma 4.** *Let $\rho^u, \rho^l$ be two runs of $\mathcal{A}$ on $w$, ending in states $q_u, q_l$ respectively, such that $\Gamma(\rho^u) - \Gamma(\rho^l) > \mathcal{M}$. Let $\rho^{u_f}, \rho^{l_f}$ be accepting runs on $z$ starting in $q_u, q_l$ respectively and ending in $q_{u_f}, q_{u_l}$ respectively. Then $val(\rho^u \rho^{u_f}) + \lambda^{-|wz|} \boldsymbol{fval}(q_{u_f}) > val(\rho^l \rho^{l_f}) + \lambda^{-|wz|} \boldsymbol{fval}(q_{l_f})$.*

In particular, once a gap becomes too large, the only way to recover from it is if the lower run cannot continue at all.

**Lemma 5.** *Consider a recoverable gap $(w, q_u, q_l)$ with respect to $z$ such that $\mathsf{gap}(w, q_u, q_l) > \mathcal{M}$, then $\mathcal{A}_{[q_u \to_f \alpha]}(z) < \infty$ and $\mathcal{A}_{[q_l \to_f \alpha]}(z) = \infty$.*

*Proof.* From the second condition in the definition of recoverable gap (Definition 1), we have $\mathcal{A}_{[q_u \to_f \alpha]}(z) < \infty$. Let $\rho^u, \rho^l$ be minimal runs on $|w|$ ending in $q_u, q_l$ respectively. Assume by way of contradiction that $\mathcal{A}_{[q_l \to_f \alpha]}(z) < \infty$, that is, there exists an accepting run $\rho^{l\prime}$ on $z$ starting in $q_l$. Let $\rho^{u\prime}$ be a minimal accepting run on $z$ starting in $q_u$. Since $\lambda^{|w|}(\mathtt{val}(\rho^u) - \mathtt{val}(\rho^l)) = \mathtt{gap}(w, q_u, q_l) > \mathcal{M}$, Lemma 4 contradicts the fact that $\rho^1\rho^{1\prime}$ is a minimal accepting run on $wz$ by the definition of recoverable gap. □

We can now prove the main result of this section.

**Lemma 6.** *There exists an effectively computable number $N$ (depending on $\mathcal{A}$) such that $\mathcal{A}$ has infinitely many recoverable gaps if and only if there exists a recoverable gap $(w, q_u, q_l)$ such that $\mathtt{gap}(w, q_u, q_l) > N$.*

*Proof Overview* We start with an overview of the more complex direction – the existence of a large recoverable gap implies the existence of infinitely many recoverable gaps. Assume that $(w, q_u, q_l)$ is a large recoverable gap with respect to $z$. We consider two minimal runs $\rho^{q_u}, \rho^{q_l}$ on $w$ ending in $q_u$ and $q_l$, respectively. These two runs end "far" from each other, so we can use Lemma 3 to claim that for a large enough $N$, they have already been far from each other for a while. Specifically, for the last $n_{\text{steps}}$ steps the gap between the runs was at least $n_{\text{gap}}$ for some large $n_{\text{steps}}, n_{\text{gap}}$ that we choose to fit our needs.

We now look for two indices $i < j$ among the last $n_{\text{steps}}$ indices of $w$ such that pumping the infix of $w$ between $i$ and $j$ generates words that induce unboundedly large recoverable gaps. To do so, we choose $n_{\text{steps}}$ such that $Q$ can be partitioned into two sets of states – an upper set $U$ and a lower set $L$, that are far from each other and "separate" the runs $\rho^{q_u}, \rho^{q_l}$ at step $i$. In particular, pumping the infix does not interleave the runs, and maintains the growing gap. The above is depicted in Figure 5. We require the following properties:

1. Every two runs on the prefix $w_1 \cdots w_i$ of $w$ ending in $U$ and in $L$, respectively, that are minimal runs ending in their respective states, are far enough from each other to satisfy the condition of Lemma 2;
2. Every run on $w$ that is minimal among the runs ending in $q_u$ has to visit $U$ at the $i$'th step;
3. Every run on $w$ that is minimal among the runs ending in $q_l$ has to visit $L$ at the $i$'th step;

As we show, finding such a partition is possible by choosing $n_{\text{gap}} = (|Q| - 1)\mathcal{M}$.

Next, we show that in fact, $U$ and $L$ induce a certain separation of the run *trees* emanating from them on the pumped words. Specifically, we show that:

(i) There exist runs of $\mathcal{A}$ on the pumped words (denoted $w^{(*)}$) ending in $q_u, q_l$.
(ii) Every run on $w^{(*)}$ (ending in any state) that is a prefix of a minimal run on $w^{(*)}z$ has to visit $U$ at the $i$'th step. That is, a variant of Condition (2), where instead of $q_u$ we consider any state $p$ reached after reading $w^{(*)}$ along a minimal run on $w^{(*)}z$.

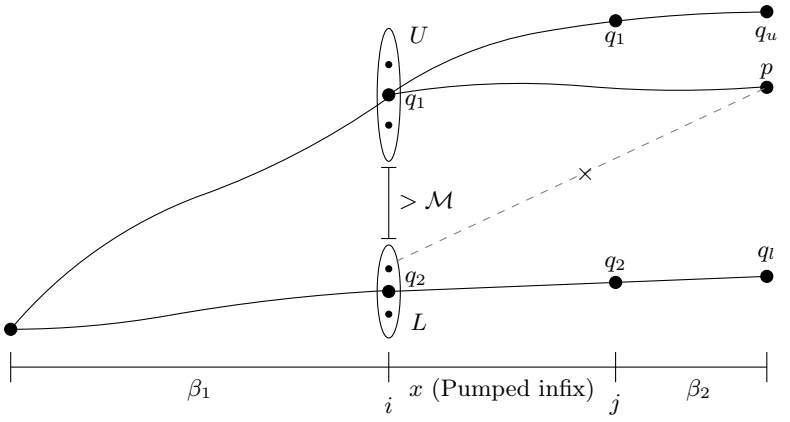



Fig. 5: At step $i$ of $\mathcal{A}$'s run on $w$, the states are partitioned into an upper set $U$ and a lower set $L$ that are separated by a large gap. The runs $\rho^{q_u}, \rho^{q_l}$ visit $U, L$ respectively, meaning the gap between them can only grow after step $i$. The indices $i, j$ are chosen such that both runs $\rho^{q_u}, \rho^{q_l}$ repeat states and the sets of ancestors $\mathrm{Anc}_q(i)$,$\mathrm{Anc}_q(j)$ are identical for each $q \in Q$. The state $p$, which is visited after reading a pumped word $w^{(*)}$ by a minimal run on $w^{(*)}z$, is not reachable from $L$ on any of the pumped suffixes.

(iii) Condition (3) above holds not only for $w$ but for the pumped words $w^{(*)}$ as well.

Note that (ii) and (iii) imply that runs on $w^{(*)}$ also induce a recoverable gap.

From this, it follows from Condition 1 and Lemma 2 that the pumped words induce unboundedly large recoverable gaps.

In order to ensure (i), we require $n_{\text{steps}} \geq |Q|^2$ (which is the length of the "large gaps" suffix) such that both runs $\rho^{q_u}, \rho^{q_l}$ must repeat their pair of respective states at some indices $i, j$. Consequently, the runs $\rho^{q_u}, \rho^{q_l}$ can be pumped to achieve the desired runs.

To ensure (ii), it follows from Corollary 1 and the fact that $\rho^{q_u}$ is a prefix of a minimal run on $wz$ that any state $p$ reached along a run on $w^{(*)}z$ after reading $w^{(*)}$ is not reachable from $L$ when reading $w_{i+1} \cdots w_{|w|}$, and we want to ensure that $p$ is not reachable from $L$ when reading the pumped suffix as well. For that, for each state $q$ and for each index of $w$ we consider the set of states $\mathrm{Anc}_q(i)$ from which $q$ is reachable when reading the respective suffix (from index $i$), called the *ancestors* of $q$ at index $i$, and it is enough to require that for each state $q$ this set is identical for indices $i$ and $j$. This, in turn, requires to increase $n_{\text{steps}}$ by a factor of $2^{|Q|^2}$. Combined with the previous requirement on $i, j$, we choose $n_{\text{steps}} = |Q|^2 2^{|Q|^2}$.

Finally, for condition (3) in (iii), we use the fact that there exists a run ending in $q_l$ that visits $L$ at the $i$'th step (namely the pumped run) and apply

Corollary 1. Indeed, any run that does not visit $L$ at the $i$'th step must visit $U$ instead, and by Corollary 1 and the gap between $U$ and $L$, it must be larger than the run we have that visits $L$ and therefore not minimal. □

*Proof (of Lemma 6).* Consider runs $\rho_0^1, \ldots, \rho_n^1$ and $\rho_0^2, \ldots, \rho_n^2$. From Lemma 3, we can effectively compute $N$ such that if $\Gamma(\rho_0^1, \ldots, \rho_n^1) - \Gamma(\rho_0^2, \ldots, \rho_n^2) > N$, then $n > |Q|^2 2^{|Q|^2}$ and $\Gamma(\rho_0^1, \ldots, \rho_{n-|Q|^2 2^{|Q|^2}}^1) - \Gamma(\rho_0^2, \ldots, \rho_{n-|Q|^2 2^{|Q|^2}}^2) > (|Q|-1)\mathcal{M}$.

Assume that $(w, q_u, q_l)$ is a recoverable gap with respect to $z$ and $\mathsf{gap}(w, q_u, q_l) > N$. Let $\rho^{q_u} = \rho_0^{q_u} \ldots \rho_{|w|}^{q_u}$ be a run on $w$ ending in $q_u$ that is minimal among the runs on $w$ ending in $q_u$, and similarly $\rho^{q_l} = \rho_0^{q_l} \ldots \rho_{|w|}^{q_l}$ for $q_l$. Since these runs are minimal runs ending in their respective states, it holds that $\Gamma(\rho^{q_u}) - \Gamma(\rho^{q_l}) = \mathsf{gap}(w, q_u, q_l) > N$, and so we have $|w| > |Q|^2 2^{|Q|^2}$ and $\Gamma(\rho_0^1, \ldots, \rho_{|w|-|Q|^2 2^{|Q|^2}}^1) - \Gamma(\rho_0^2, \ldots, \rho_{|w|-|Q|^2 2^{|Q|^2}}^2) > (|Q|-1)\mathcal{M}$.

For every $q \in Q$ and $1 \leq i \leq |w|$, let $\mathrm{Anc}_q(i) = \{q' \in Q \mid \mathcal{A}_{[q' \to q]}(w_{i+1} \cdots w_{|w|}) < \infty\}$ be the set of *ancestors* of $q$ at step $i$, i.e., states from which $q$ is reachable when reading the input $w_{i+1}, \cdots, w_{|w|}$. By the pigeonhole principle there exist $|w| - |Q|^2 2^{|Q|^2} \leq i < j \leq |w|$ such that

- For every $q \in Q$, $\mathrm{Anc}_q(i) = \mathrm{Anc}_q(j)$,
- $\rho_i^{q_u} = \rho_j^{q_u}$ and $\rho_i^{q_l} = \rho_j^{q_l}$.

Write $w = \beta_1 x \beta_2$ where $\beta_1 = w_1 \cdots w_i$, $x = w_{i+1} \cdots w_j$ and $\beta_2 = w_{j+1} \cdots w_{|w|}$. We now turn to show that by pumping $x$, we can obtain unboundedly large recoverable gaps.

Let $k \in \mathbb{N}$. We can easily show that $\beta_1 x^k \beta_2$ induces unboundedly large gaps between $q_u$ and $q_l$, but that would not be sufficient: We also need those gaps to be recoverable with respect to $z$, that is, the minimal run on $\beta_1 x^k \beta_2 z$ has to visit $q_u$ after reading $\beta_1 x^k \beta_2$. However, this is not necessarily true: It can visit a different state, and we need to show that that state is also far enough from $q_l$. The runs $\rho_0^{q_u} \ldots \rho_i^{q_u} (\rho_{i+1}^{q_u} \ldots \rho_j^{q_u})^k \rho_{j+1}^{q_u} \ldots \rho_{|w|}^{q_u}$ and $\rho_0^{q_l} \ldots \rho_i^{q_l} (\rho_{i+1}^{q_l} \ldots \rho_j^{q_l})^k \rho_{j+1}^{q_l} \ldots \rho_{|w|}^{q_l}$ are runs on $\beta_1 x^k \beta_2$ ending in $q_u, q_l$ respectively. In particular, since there exists a run on $z$ starting in $q_u$, we have that $\mathcal{A}$ has a run on $\beta_1 x^k \beta_2 z$. Let $\rho$ be minimal among those runs, and let $q_{\min_k}$ be the state $\rho$ visits after reading $\beta_1 x^k \beta_2$. Let $\rho^{q_{\min_k}, k}, \rho^{q_l, k}$ be runs on $\beta_1 x^k \beta_2$ that are minimal among the runs ending in $q_{\min_k}, q_l$ respectively. Note that $\rho^{q_{\min_k}, k}$ can be obtained as a prefix of $\rho$, since $\rho$ is minimal. Then we have $\mathsf{gap}(\beta_1 x^k \beta_2, q_{\min_k}, q_l) = \Gamma(\rho^{q_{\min_k}, k}) - \Gamma(\rho^{q_l, k})$, and it remains to show that $\Gamma(\rho^{q_{\min_k}, k}) - \Gamma(\rho^{q_l, k})$ can get unboundedly large for a large enough $k$.

We already know that the runs $\rho^{q_u}, \rho^{q_l}$ are far enough from each other at their $i$'th step to satisfy the condition of Lemma 2, and we want to show that the same is true for $\rho^{q_{\min_k}, k}, \rho^{q_l, k}$.

To do so, we intuitively show that after reading $\beta_1$, the runs $\rho^{q_u}, \rho^{q_l}$ have become so far apart that they now stem from disjoint sets of states with a large

gap between them. Formally, consider the sets

$$U' = \{q \in Q \mid \text{there exists a run } \rho \text{ on } w \text{ with } \mathtt{val}(\rho) = \mathcal{A}_{[Q_0 \to q_u]}(w) \text{ and } \rho_i = q\}$$
$$L' = \{q \in Q \mid \text{there exists a run } \rho \text{ on } w \text{ with } \mathtt{val}(\rho) = \mathcal{A}_{[Q_0 \to q_l]}(w) \text{ and } \rho_i = q\}$$

That is, $U'$ (resp. $L'$) is the set of states that appear at step $i$ in a minimal run to $q_u$ (resp. $q_l$). For every $q \in Q$, let $v(q) = \lambda^i \mathcal{A}_{[Q_0 \to q]}(\beta_1)$ be the "undiscounted" value of a minimal run of $\mathcal{A}$ on $\beta_1$ ending in $q$. Then, from Lemma 3 and the constants we chose, for every $q'_u \in U', q'_l \in L'$ we have $v(q'_u) - v(q'_l) > (|Q|-1)\mathcal{M}$.

In particular, there is a partition of $Q$ into two sets $U, L$ such that $U' \subseteq U, L' \subseteq L$ and $v(p) - v(q) > \mathcal{M}$ for every $p \in U$ and $q \in L$. Indeed, otherwise the maximal gap between two states is less than $(|Q|-1)\mathcal{M}$. We next show that (i) $\rho_i^{q_{\min_k},k} \in U$, and (ii) $\rho_i^{q_l,k} \in L$.

For (i), we note that $\mathcal{A}_{[L \to q_{\min_k}]}(x\beta_2) = \infty$: Otherwise, since $\mathcal{A}_{[q_{\min_k} \to_f \alpha]}(z) < \infty$, there exists an accepting run on $wz$ that visits $L$ after reading $\beta_1$ and $q_{\min_k}$ after reading $w$. By Lemma 4, such a run must be of lower weight than any run that visits $U$ after reading $\beta_1$, in contradiction to the fact that $\rho^{q_u}$ is a prefix of a minimal run on $wz$ by the definition of recoverable gap. Since $\mathrm{Anc}_{q_{\min_k}}(i) = \mathrm{Anc}_{q_{\min_k}}(j)$, we also have that $\mathcal{A}_{[L \to q_{\min_k}]}(x^k\beta_2) = \infty$. In particular, $\rho_i^{q_{\min_k},k} \in U$.

For (ii), the run $\rho_0^{q_l} \dots \rho_i^{q_l}(\rho_{i+1}^{q_l} \dots \rho_j^{q_l})^k \rho_{j+1}^{q_l} \dots \rho_{|w|}^{q_l}$ satisfies $\rho_i^{q_l} \in L$ (since it is in $L'$), and in particular $\mathcal{A}_{[L \to q_l]}(x^k\beta_2) < \infty$. By Corollary 1, any run whose $i$'th state is in $U$ results in a higher weight than any run whose $i$'th state is in $L$, and so since $\rho^{q_l,k}$ is minimal we have $\rho_i^{q_l,k} \in L$.

It remains to show that the runs $\rho^{q_{\min_k},k}, \rho^{q_l,k}$, being far from each other at the $i$'th step, get unboundedly far from each other as $k$ increases. Let $f_u, f_l : \{i, i+1, \dots\} \to \mathbb{N}$ be defined as follows:

- $f_u(i) = \min_{q \in U} v(q)$
- For $m \geq i$, $f_u(m+1) = \lambda(f_u(m) - m_{\mathcal{A}})$
- $f_l(i) = \max_{q \in L} v(q)$
- For $m \geq i$, $f_l(m+1) = \lambda(f_l(m) + m_{\mathcal{A}})$

Intuitively, $f_u$ (resp. $f_l$) represents a lower (resp. upper) bound on the "undiscounted" weight of runs visiting $U$ (resp. $L$) in their $i$'th step. That is, for every $m \geq i$ we have $\Gamma(\rho_0^{q_{\min_k},k} \dots \rho_m^{q_{\min_k},k}) \geq f_u(m)$, and $\Gamma(\rho_0^{q_l,k} \dots \rho_m^{q_l,k}) \leq f_l(m)$. Additionally, $f_u(i) - f_l(i) > \mathcal{M}$ and so the function $f_u(m) - f_l(m)$ increases with $m$. Thus, for every $M \in \mathbb{N}$, taking a large enough $k$, we can obtain $\Gamma(\rho^{q_{\min_k},k}) - \Gamma(\rho^{q_l,k}) \geq f_u(|\beta_1 x^k \beta_2|) - f_l(|\beta_1 x^k \beta_2|) > M$. This concludes the proof that if $\mathcal{A}$ has a large recoverable gap, then it has infinitely many recoverable gaps.

For the converse direction, assume $\mathcal{A}$ has infinitely many recoverable gaps. Since $\mathcal{A}$ is integral, the term $\lambda^{|w|}(\mathcal{A}_{[Q_0 \to p]}(w) - \mathcal{A}_{[Q_0 \to q]}(w))$ is always an integer, therefore infinitely many recoverable gaps imply the existence of unboundedly large recoverable gaps, and in particular one larger than $N$. □

*Remark 1.* Following the arguments in the proofs of Lemmas 3 and 6, the number $N$ provided by Lemma 6 equals $\lambda^{|Q|^2 2^{|Q|^2}}((|Q|-1)\mathcal{M}-\mathcal{M})+\mathcal{M}$. We denote this value by $\mathcal{N}$.

### 6.2 A Large Gap is Equivalent to Separation

Recall that a gap refers to minimal runs that end in two specific states, but ignores the remaining states (to an extent). A more "holistic" view of gaps is via separations (Definition 2). In this section we show that the two views are equivalent, and that both characterize when $\mathcal{A}$ has infinitely many gaps.

**Lemma 7.** *$\mathcal{A}$ has a recoverable gap larger than $\mathcal{N}$ if and only if $\mathcal{A}$ has the $\mathcal{N}$-separation property.*

*Proof.* Assume that $\mathcal{A}$ has a recoverable gap larger than $\mathcal{N}$. By Lemma 6, there exist unboundedly large recoverable gaps, and in particular there exists a recoverable gap $(w, q_u, q_l)$ with $\mathsf{gap}(w, q_u, q_l) > (|Q|-1)\mathcal{N}$.

Intuitively, when ordering the states by the weight of the minimal run that reaches each state, such a gap implies a gap of at least $\mathcal{N}$ between two successive states, leading to the desired partition. We then claim that the sets are separated by the same suffix $z$ that separates the states from the original gap.

Write $Q = \{q_1, \ldots, q_{|Q|}\}$ such that $\mathcal{A}_{[Q_0 \to q_1]}(w) \leq \ldots \leq \mathcal{A}_{[Q_0 \to q_{|Q|}]}(w)$ (recall that if there is no run on $w$ ending in $q$, then $\mathcal{A}_{[Q_0 \to q]}(w) = \infty$), and let $i_l < i_u$ be indices such that $q_l = q_{i_l}, q_u = q_{i_u}$. Then there exists $j \in \{i_l, \ldots, i_{u-1}\}$ such that $\mathcal{A}_{[Q_0 \to q_{j+1}]}(w) - \mathcal{A}_{[Q_0 \to q_j]}(w) > \mathcal{N}$. Let $U = \{q_{j+1}, \ldots, q_{|Q|}\}$ and $L = \{q_1, \ldots, q_j\}$. Then for every $q'_u \in U, q'_l \in L$ we have $\lambda^{|w|}(\mathcal{A}_{[Q_0 \to q'_u]}(w) - \mathcal{A}_{[Q_0 \to q'_l]}(w)) > \mathcal{N}$.

Consider $z \in \Sigma^*$ such that the gap $(w, q_u, q_l)$ is recoverable with respect to $z$. Note that $\mathcal{N} > \mathcal{M}$, and so it follows from Lemma 4 that $\mathcal{A}_{[q'_l \to_f \alpha]}(z) = \infty$ for every $q'_l \in L$. Indeed, if $\mathcal{A}$ had a run on $z$ starting in $q'_l$, concatenating it to a minimal run on $w$ ending in $q'_l$ would result in a run of lower weight than any run on $wz$ that visits $q_u$ after reading $w$, contradicting the fact that $(w, q_u, q_l)$ is a recoverable gap. Additionally, it follows from $(w, q_u, q_l)$ being a recoverable gap that there exists a minimal run on $wz$ that visits $q_u$ after reading $w$. Then $(w, q_u, q'_l)$ is a recoverable gap with respect to $z$, and so $w$ has the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z)$.

For the converse direction, assume that $w$ has the $\mathcal{N}$-separation property with respect to some $(U, L, q_u, z)$. In particular, $\mathcal{A}_{[Q_0 \to q_u]}(w) + \lambda^{-|w|}\mathcal{A}_{[q_u \to_f \alpha]}(z) < \infty$. Let $q'_u \in U$ be such that $\mathcal{A}_{[Q_0 \to q'_u]}(w) + \lambda^{-|w|}\mathcal{A}_{[q'_u \to_f \alpha]}(z)$ is minimal. Let some $q'_l \in L$. Then $(w, q'_u, q'_l)$ is a recoverable gap with respect to $z$, and it is larger than $\mathcal{N}$, as needed. □

## 7 Bounding the Witnesses for Separation

In Section 6 we show that $\mathcal{A}$ has infinitely many recoverable gaps if and only if there exists a word $w$ with the $\mathcal{N}$-separation property. Expanding Definition 2,

this happens if and only if there exist a partition of $Q$ into two sets $U, L$ and there exist words $w, z$ that "separate" $U$ from $L$. In this section we can bound the length of such minimal $w, z$. We start with $w$ (see the full version for the detailed proof).

**Lemma 8.** *Let $C = \frac{\lambda}{\lambda-1}(\mathcal{N}|Q| + 2m_{\mathcal{A}})$. Assume that $w$ has the $\mathcal{N}$-separation property for some $w \in \Sigma^*$. Then there exists $w'$ such that $w'$ has the $\mathcal{N}$-separation property and $|w'| \leq (C+2)^{|Q|}$.*

*Proof (Sketch).* Assume that $w$ has the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z)$.

We start by using an identical construction to that of Lemma 1, with bound $C$, in order to define a sequence of vectors $v_0, \ldots, v_{|w|}$ with $v_i \in \{0, \ldots, C, \infty\}^Q$ for every $0 \leq i \leq |w|$ that, intuitively, keep track of the runs of $\mathcal{A}$ on $w$, as follows.

- For every $q \in Q$ set $(v_0)_q = \begin{cases} 0 & q \in Q_0 \\ \infty & \text{otherwise} \end{cases}$
- For every $i > 0, q \in Q$ let $v'_{i,q} = \min_{q' \in Q}((v_{i-1})_{q'} + \mathtt{val}(q', w_i, q))$, where $\mathtt{val}(q', \sigma, q)$ is regarded as $\infty$ if $(q', \sigma, q) \notin \delta$ (the $v'_{i,q}$ are "intermediate" values).
- For every $i > 0$ let $r_i = \min_{q \in Q} v'_{i,q}$ (the $r_i$ are the offset of the vector from 0).
- For every $i > 0, q \in Q$ set $(v_i)_q = \begin{cases} \lambda(v'_{i,q} - r_i) & \lambda(v'_{i,j} - r_i) \leq C \\ \infty & \text{otherwise} \end{cases}$

Recall that intuitively, $(v_i)$ tracks, for each $q \in Q$, the gap between the minimal run on $w_1 \cdots w_i$ ending in $q$ and the minimal run on this prefix overall. When this gap becomes large enough that recovering from it implies the existence of $\mathcal{N}$-separation, it is denoted $\infty$.

Denote the normalized difference $\lambda^i(\mathcal{A}_{[Q_0 \to q]}(w_1 \cdots w_i) - \mathcal{A}(w_1 \cdots w_i))$ by $\Delta_{q,i}(w)$. It is easy to show that $v_i$ keeps the correct weight of runs whose gap from the minimal one remains always under $C$. However, if a gap of a run goes over $C$ but then comes back down, then $v_i$ no longer tracks it correctly. To account for this, we claim that since $w$ has the $\mathcal{N}$-separation property, for every $q, i$ at least one of the following must hold:

- $(v_i)_q = \begin{cases} \Delta_{q,i}(w) & \Delta_{q,i}(w) \leq C \\ \infty & \text{otherwise} \end{cases}$.
- There exists $i' < i$ such that $w_1 \cdots w_{i'}$ has the $\mathcal{N}$-separation property.

That is, either $v_i$ tracks the runs correctly, or there is some shorter prefix that already has the $\mathcal{N}$-separation property.

The proof is by induction on $i$, with the only problematic case arising when $(v_{i-1})_{q'} = \infty$, and so the information about the exact value of the gap represented by $(v_{i-1})_{q'}$ is gone. We consider the normalization value $r_i$ (i.e., the offset of the minimal run from 0): if $r_i$ is small, then the gap represented by $(v_i)_q$ is

still very large, and we show that marking it as $\infty$ is sound. Otherwise, if $r_i$ is large, then the above gap might indeed be wrongly marked as $\infty$. However, we show that in this case, $r_i$ is so large that we can actually obtain an $\mathcal{N}$-separation property "below" $r_i$, using a shorter witness. More precisely:

- If $(v_{i-1})_{q'} = \infty$ and $r_i \leq C\frac{\lambda-1}{\lambda} - m_{\mathcal{A}}$, then since $(v_{i-1})_{q'} = \infty$, we have $(v_i)_q = \infty$. It remains to show that $\lambda^i(\mathcal{A}_{[Q_0 \to q]}(w_1 \cdots w_i) - \mathcal{A}(w_1 \cdots w_i)) > C$. Indeed,

$$\begin{aligned}
&\lambda^i(\mathcal{A}_{[Q_0 \to q]}(w_1 \cdots w_i) - \mathcal{A}(w_1 \cdots w_i)) \\
\geq &\lambda^i(\mathcal{A}_{[Q_0 \to q']}(w_1 \cdots w_{i-1}) - \mathcal{A}(w_1 \cdots w_{i-1}) - (m_{\mathcal{A}} + r_i) \cdot \lambda^{-(i-1)}) \\
= &\lambda(\lambda^{i-1}(\mathcal{A}_{[Q_0 \to q']}(w_1 \cdots w_{i-1}) - \mathcal{A}(w_1 \cdots w_{i-1})) - r_i - m_{\mathcal{A}}) \\
> &\lambda(C - (C\frac{\lambda-1}{\lambda} - m_{\mathcal{A}}) - m_{\mathcal{A}}) = \lambda(\lambda^{-1}C + m_{\mathcal{A}} - m_{\mathcal{A}}) > C
\end{aligned}$$

  where the first transition follows by observing that when reading $w_i$, in the worst case, the weight of a specific run can decrease by $\lambda^{-(i-1)}m_{\mathcal{A}}$, and the overall weight of the word can increase by $\lambda^{-(i-1)}r_i$.
- $r_i > C\frac{\lambda-1}{\lambda} - m_{\mathcal{A}}$. This is only possible if for every $q_l$ such that $(v_{i-1})_{q_l} < C\frac{\lambda-1}{\lambda} - 2m_{\mathcal{A}} = \mathcal{N}|Q|$, $q_l$ has no $w_i$-transition. Let $L'' = \{q_l \in Q \mid (v_{i-1})_{q_l} < \mathcal{N}|Q|\}$. Write $Q = q_1, \ldots, q_{|Q|}$ such that $(v_{i-1})_{q_1} \leq \ldots \leq (v_{i-1})_{q_{|Q|}}$, and so $L'' = \{q_1, \ldots, q_{|L''|}\}$. Since $w$ has the $\mathcal{N}$-separation property, in particular $\mathcal{A}$ has a run on $w$ and so $L'' \subsetneq Q$. Then, there exists $1 \leq r \leq |L''|$ such that $(v_{i-1})_{q_{r+1}} - (v_{i-1})_{q_r} > \mathcal{N}$. Let $U' = \{q_{r+1}, \ldots, q_{|Q|}\}, L' = \{q_1, \ldots, q_r\}$, and note that for every $q'_l \in L'$, $q'_l$ has no $w_i$-transition. For every $q'_l \in L', q'_u \in U'$, we have $\lambda^{i-1}(\mathcal{A}_{[Q_0 \to q'_u]}(w_1 \cdots w_{i-1}) - \mathcal{A}_{[Q_0 \to q'_l]}(w_1 \cdots w_{i-1}) = (v_{i-1})_{q'_u} - (v_{i-1})_{q'_l} > \mathcal{N}$. Let $q'_u \in U'$ be such that $\mathcal{A}_{[Q_0 \to q'_u]}(w_1 \cdots w_{i-1}) + \lambda^{-(i-1)}\mathcal{A}_{[q'_u \to_f \alpha]}(w_i)$ is minimal. Then for every $q'_l \in L'$, $(w_1 \cdots w_{i-1}, q'_u, q'_l)$ is a recoverable gap with respect to $w_i$, and so $w_1 \cdots w_{i-1}$ has the $\mathcal{N}$-separation property with respect to $(U', L', q'_u, w_i)$, and we are done.

Now, it remains to show that if $|w| > (C+2)^{|Q|}$, there exists $w'$ such that $|w'| < |w|$ and $w'$ has the $\mathcal{N}$-separation property.

To this end, we use the induction hypothesis and the pigeonhole principle to remove an infix of $w$, and argue that the resulting word $w'$ also has the $\mathcal{N}$-separation property with respect to some $(U', L', q'_u)$: Either all of the minimal runs ending in the states of $L$ have values far enough (below) of $C$, in which case $U', L'$ can be chosen to be $U, L$ respectively; or some state of $L$ attains a high value, in which case there must be a large gap between two consecutive states of $L$, and the resulting lower set can be chosen as $L'$. As for $q'_u$, it is simply enough to consider the state in $U'$ that the minimal run on $w'z$ visits after reading $w'$ (see the full version for the details). □

Next, we give a bound on the length of the minimal separating suffix $z$ from Definition 2. Recall that by Lemma 5, a large gap can only be recoverable if the smaller runs cannot continue at all. Following that, we can now limit the search

to suffixes that separate runs in a Boolean sense (i.e., making one accept and another reject). This yields a bound from standard arguments about Boolean automata, as follows.

**Lemma 9.** *Consider a word $w$ that has the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z)$. Then there exists $z'$ such that $w$ has the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z')$ and $|z'| \leq 2^{2|Q|}$.*

## 8 Determinizability of Integral NDAs is Decidable

In this section we establish the decidability of determinization. To this end, we start by completing the characterization of determinizable NDAs by means of gaps, and then use the results from previous sections to conclude the decidability of this characterization.

Recall that in Lemma 1 we show that finitely many recoverable gaps imply determinizability. In this section we show the converse, thus completing the characterization of determinizable integral NDAs as exactly those that have finitely many recoverable gaps.

We first need the following lemma which is proved in [6, Lemma 5].

**Lemma 10.** *Consider an NDA $\mathcal{A}$ for which there is an equivalent DDA $\mathcal{D}$. If there is a state $q$ of $\mathcal{A}$ and words $w, w', z$ such that:*

- *$\mathcal{A}$ has runs on $w$ and $w'$ ending in $q$;*
- *$\mathsf{gap}(w, q, p) \neq \mathsf{gap}(w', q, p')$, where $p, p'$ are the last states of some minimal runs of $\mathcal{A}$ on $w, w'$ respectively;*
- *both gaps $(w, q, p)$ and $(w', q, p')$ are recoverable with respect to $z$;*

*then the runs of $\mathcal{D}$ on $w$ and $w'$ end in different states.*

We now show the converse of Lemma 1.

**Lemma 11.** *If an NDA $\mathcal{A}$ has infinitely many recoverable gaps, it is not determinizable.*

*Proof.* Assume by way of contradiction that $\mathcal{A}$ has an equivalent DDA $\mathcal{D}$ and infinitely many recoverable gaps. For every $q \in Q$, let

$$G_q = \{w \mid \mathcal{A} \text{ has a recoverable gap of the form } (w, q, p) \text{ for some } p\}$$

Since $Q$ is finite and $\mathcal{A}$ has infinitely many recoverable gaps, there exists $q \in Q$ such that $G_q$ is infinite. By Lemma 9, there is a finite collection $Z$ of words such that every recoverable gap is recoverable with respect to some word in $Z$. Therefore there exist $z \in Z$ and an infinite subset $G'_q \subseteq G_q$ such that for every $w \in G'_q$, the gap $(w, q, p)$ is recoverable with respect to $z$ for some $p$. By Lemma 10, for every two words $w, w' \in G'_q$, the runs of $\mathcal{D}$ on $w$ and $w'$ end in different states, in contradiction to the fact that $\mathcal{D}$ has finitely many states. □

Consider an NDA $\mathcal{A}$. By Lemmas 1, 6 to 9 and 11 we have that $\mathcal{A}$ has an equivalent DDA if and only if for every $w, z$ such that $|w| \leq (\frac{\lambda}{\lambda-1}(\mathcal{N}|Q|+2m_{\mathcal{A}})+2)^{|Q|}$ and $|z| \leq 2^{2|Q|}$, it holds that $w$ does not have the $\mathcal{N}$-separation property with respect to $(U, L, q_u, z)$ for every $U, L, q_u$. Since the latter condition can be checked by traversing finitely many words and simulating the runs of $\mathcal{A}$ on each of them, we can conclude our main result.

**Theorem 1.** *The problem of whether an integral NDA has a deterministic equivalent is decidable.*

*Remark 2 (Complexity of Determinization).* Using the bounds on $w, z$, one can guess $w, z$ on-the-fly, while keeping track of the weights of minimal runs to all states, discarding those that go above $C$ as per Lemma 8, to check whether $\mathcal{A}$ has the $\mathcal{N}$-separation property. Since $\mathcal{N}$ is double exponential in the size of $\mathcal{A}$, this procedure can be done in $\mathsf{NEXPSPACE} = \mathsf{EXPSPACE}$. Thus, determinizability is in $\mathsf{EXPSPACE}$. For a lower bound, determinizability is also $\mathsf{PSPACE} - \mathsf{hard}$ by a standard reduction from NFA universality. Tightening this gap is left open. Note that for lowering the upper bound, we would need a refined application of the pigeonhole principle in Lemma 6, which seems somewhat out of reach for the pumping argument. Conversely, for increasing the lower bound, we would need to show that using discounting we can somehow force a double-exponential blowup in determinization. While this might be within reach, no such example are known for e.g., tropical weighted automata, suggesting that this may be very difficult.

**Acknowledgments** The authors thank Guy Raveh for fruitful discussion regarding Lemma 3.

## References

1. de Alfaro, L., Henzinger, T.A., Majumdar, R.: Discounting the future in systems theory. In: Automata, Languages and Programming. pp. 1022–1037 (2003)
2. Almagor, S., Yeshurun, A.: Determinization of one-counter nets. In: 33rd International Conference on Concurrency Theory, CONCUR 2022 (Sep 2022)
3. Andersson, D.: An improved algorithm for discounted payoff games. In: ESSLLI Student Session. pp. 91–98. Citeseer (2006)
4. Boker, U., Hefetz, G.: Discounted-Sum Automata with Multiple Discount Factors. In: 29th EACSL Annual Conference on Computer Science Logic (CSL 2021). vol. 183, pp. 12:1–12:23 (2021)
5. Boker, U., Hefetz, G.: On the comparison of discounted-sum automata with multiple discount factors. Foundations of Software Science and Computation Structures LNCS 13992 p. 371 (2023)
6. Boker, U., Henzinger, T.A.: Determinizing Discounted-Sum Automata. In: Computer Science Logic (CSL'11) - 25th International Workshop/20th Annual Conference of the EACSL. Leibniz International Proceedings in Informatics (LIPIcs), vol. 12, pp. 82–96 (2011)

7. Boker, U., Henzinger, T.A.: Exact and approximate determinization of discounted-sum automata. Logical Methods in Computer Science **10** (2014)
8. Boker, U., Henzinger, T.A., Otop, J.: The target discounted-sum problem. In: 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science. pp. 750–761. IEEE (2015)
9. Broome, J.: Discounting the future. Philosophy & Public Affairs **23**(2), 128–156 (1994)
10. Cadilhac, M., Pérez, G.A., Van Den Bogaard, M.: The impatient may use limited optimism to minimize regret. In: Foundations of Software Science and Computation Structures: 22nd International Conference, FOSSACS 2019. pp. 133–149. Springer (2019)
11. Chatterjee, K., Doyen, L., Henzinger, T.A.: Quantitative languages. In: Computer Science Logic. pp. 385–400 (2008)
12. Chatterjee, K., Doyen, L., Henzinger, T.A.: Alternating weighted automata. In: Fundamentals of Computation Theory. pp. 3–13 (2009)
13. Chatterjee, K., Doyen, L., Henzinger, T.A.: Expressiveness and closure properties for quantitative languages. In: 2009 24th Annual IEEE Symposium on Logic In Computer Science. pp. 199–208 (2009). https://doi.org/10.1109/LICS.2009.16
14. Chatterjee, K., Forejt, V., Wojtczak, D.: Multi-objective discounted reward verification in graphs and mdps. In: Logic for Programming, Artificial Intelligence, and Reasoning. pp. 228–242 (2013)
15. Droste, M., Kuske, D.: Skew and infinitary formal power series. Theoretical Computer Science **366**(3), 199–227 (2006)
16. Filiot, E., Jecker, I., Lhote, N., Pérez, G.A., Raskin, J.F.: On delay and regret determinization of max-plus automata. In: 2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). pp. 1–12. IEEE (2017)
17. Gimbert, H., Zielonka, W.: Limits of multi-discounted markov decision processes. In: 22nd Annual IEEE Symposium on Logic in Computer Science (LICS 2007). pp. 89–98 (2007). https://doi.org/10.1109/LICS.2007.28
18. Kirsten, D.: A burnside approach to the termination of mohri's algorithm for polynomially ambiguous min-plus-automata. RAIRO - Theoretical Informatics and Applications **42**(3), 553–581 (2008). https://doi.org/10.1051/ita:2008017
19. Madani, O., Thorup, M., Zwick, U.: Discounted deterministic markov decision processes and discounted all-pairs shortest paths. ACM Trans. Algorithms **6**(2) (apr 2010)
20. Mohri, M.: Finite-state transducers in language and speech processing. Computational Linguistics **23**(2), 269–311 (1997)
21. Zwick, U., Paterson, M.: The complexity of mean payoff games on graphs. Theoretical Computer Science **158**(1), 343–359 (1996)