Properties of Multiset Language Classes Defined by Multiset Storage Automata

Manfred Kudlek, Patrick Totzke, and Georg Zetzsche
{kudlek,3totzke,3zetzsch}@informatik.uni-hamburg.de

Department Informatik, MIN Fakultät, Universität Hamburg
http://www.informatik.uni-hamburg.de/TGI

Abstract. The previously introduced multiset language classes defined by multiset storage automata are being explored with respect to their closure properties and alternative characterizations.

1 Introduction

Multiset languages and their characterizations play an important role in various areas of Theoretical Computer Science, as in Concurrency Theory or Membrane Computing. Generative models have been investigated in [5]. For word languages it is well known that they can be characterized by grammars as well as by automata. Analogous automata can be introduced for multiset languages.

In addition to multiset finite automata (MFA) [1, 6], multiset storage automata (MSA), the commutative counterparts of pushdown automata have been introduced recently [7]. In this paper we continue our work on MSA especially considering closure properties of the corresponding multiset language classes and alternative characterizations.

After a short review of the necessary definitions in section 2 we will give an overview of the closure properties that are known at present. An algebraic characterization of the stated classes will be given in the following chapter.

2 Definitions

We will recapitulate the definitions of *multisets* together with some elementary operations as well as *multiset storage automata* in two variants: one with, one without the ability for 0-testing. We define the language classes that should form our primary concern in this paper and recall from [7] how they relate to each other and some other classes.

Definition 1. (Multisets) *If M is a set then $\alpha : M \to \mathbb{N}$ is a* multiset *over M, where $\mathbb{N}$ denotes the set of natural numbers including 0. The set of all multisets over M is denoted by $M^{\oplus}$. If $|M| < \infty$ then $M^{\oplus}$ can be identified with $\mathbb{N}^{|M|}$. $M^{\oplus}$ is a commutative monoid with operation $\oplus$ and neutral element $\mathbf{0}_M$, defined by $\forall x \in M \;:\; (\alpha \oplus \beta)(x) = \alpha(x) + \beta(x)$ and $\forall x \in M \;:\; \mathbf{0}_M(x) = 0$.*

The operation $\ominus$ is defined by $\forall x \in M \ : \ (\alpha \ominus \beta)(x) = \max(0, \alpha(x) - \beta(x))$.

The operations $\sqcap$ und $\sqcup$ are defined as $(\alpha \sqcup \beta)(x) = \max(\alpha(x), \beta(x))$ and $(\alpha \sqcap \beta)(x) = \min(\alpha(x), \beta(x))$ for all $x \in M$, where $\alpha, \beta \in M^{\oplus}$. These operations are carried over to subsets of $M^{\oplus}$. For $A, B \subseteq M^{\oplus}$, let

$$A \sqcup B = \{\alpha \sqcup \beta \mid \alpha \in A,\ \beta \in B\}, \quad A \sqcap B = \{\alpha \sqcap \beta \mid \alpha \in A,\ \beta \in B\}.$$

A partial order $\sqsubseteq$ is given by $\alpha \sqsubseteq \beta \Leftrightarrow \forall x \in M \ : \ \alpha(x) \leq \beta(x)$.
Futhermore, $\alpha \sqsubset \beta \Leftrightarrow \alpha \sqsubseteq \beta \wedge \alpha \neq \beta$.

For $A, B \subseteq M^{\oplus}$, let

$$A \oplus B = \{\alpha \oplus \beta \mid \alpha \in A,\ \beta \in B\}$$

and $A^{\oplus} = \cup_{i=0}^{\infty} A^i$ with $A^0 = \{\mathbf{0}_M\}$ and $A^{i+1} = A^i \oplus A$ for $i \in \mathbb{N}$. $A^{\oplus}$ is also the smallest submonoid of $M^{\oplus}$ that contains A.

If Σ is an alphabet, and $u \in \Sigma^*$ then $\Psi(u) \in \Sigma^{\oplus}$ denotes the Parikh *vector of u given by $|u|_x$ for $x \in \Sigma$. If $L \subseteq \Sigma^*$ then*

$$\Psi(L) = \{\Psi(u) \mid u \in L\} \subseteq \Sigma^{\oplus}$$

denotes the set of all Parikh vectors of L.

We also use the notation $\langle y \rangle$ for singleton multisets i.e. $\langle y \rangle(x) = 0$ for $y \neq x$ and $\langle y \rangle(y) = 1$. For any set B let $\bar{B} = \{\langle b \rangle \mid b \in B\}$.

Definition 2. (multiset storage automaton with 0-test)
A (nondeterministic) multiset storage automaton with 0-test (MSA^0) *is a structure of the form $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \bot)$, where*
Q *is a finite set of states,*
Σ *a finite set of input symbols,*
Γ *a finite set of stack symbols,*
$K_1 \subseteq Q \times \Sigma^{\oplus} \times \Gamma^{\oplus} \times \Gamma^{\oplus} \times Q$,
$K_2 \subseteq Q \times \Sigma^{\oplus} \times \{\langle \bot \rangle\} \times ((\Gamma^{\oplus} \oplus \{\langle \bot \rangle\}) \cup \{\mathbf{0}_{\Gamma \cup \{\bot\}}\}) \times Q$,
$K \ = K_1 \cup K_2$,
$Q_F \subseteq Q$ *the set of final states,*
$Q_0 \subseteq Q$ *the set of initial states,*
$\bot \ \notin \Gamma$ *the* bottom *symbol*

A transition *is a quintuple $t = (q, \alpha, \gamma, \delta, q') \in K$.*

A configuration *is a triple $c = (q, \mu, \sigma)$ where $q \in Q$, $\mu \in \Sigma^{\oplus}$, and $\sigma \in (\Gamma^{\oplus} \oplus \{\langle \bot \rangle\}) \cup \{\mathbf{0}_{\Gamma \cup \{\bot\}}\}$. Here σ represents the current content of the memory in which the symbol $\bot$ occurs at most once.*
An initial configuration has the form $c_0 = (q_0, \mu_0, \langle \bot \rangle)$ with $q_0 \in Q_0$ and $\mu_0 \in \Sigma^{\oplus}$. A terminal configuration with final state has the form $c_f = (q_f, \mathbf{0}_{\Sigma}, \sigma)$ with $q_f \in Q_F$ and $\sigma \in (\Gamma^{\oplus} \oplus \{\langle \bot \rangle\}) \cup \{\mathbf{0}_{\Gamma \cup \{\bot\}}\}$.

A configuration of the form $c_f = (q, \mathbf{0}_\Sigma, \mathbf{0}_{\Gamma\cup\{\perp\}})$ *with* $q \in Q$ *is called a terminal configuration with empty stack.*

A step $\vdash_0$ *is a relation defined by* $c \vdash_0 c'$ *iff* $c = (q, \mu, \sigma)$, $c' = (q', \mu', \sigma')$, *where* $t = (q, \alpha, \gamma, \delta, q') \in K$, $\alpha \sqsubseteq \mu$, $\sigma' = \sigma \ominus \gamma \oplus \delta$ *and* $\mu' = \mu \ominus \alpha$.
Additionally, if $t \in K_1$ *we require* $\gamma \sqsubseteq \sigma$ *and if* $t \in K_2$, $\gamma = \langle\perp\rangle = \sigma$ *must hold. Note that the* bottom *symbol* $\perp$ *can only be read if there are no other symbols in the stack thus being a test for 0. Let* $\vdash_0^*$ *denote the reflexive and transitive closure of* $\vdash_0$.

The multiset language *accepted by a MSA*^{0} *A with final state is defined by*

$$M_f^0(A) = \{\mu \in \Sigma^\oplus \mid (q_0, \mu, \langle\perp\rangle) \vdash_0^* (q_f, \mathbf{0}_\Sigma, \sigma), \\ q_0 \in Q_0,\ q_f \in Q_F,\ \sigma \in (\Gamma^\oplus \oplus \{\langle\perp\rangle\} \cup \{\mathbf{0}_{\Gamma\cup\{\perp\}}\})\}$$

The multiset language *accepted by a MSA*^{0} *A with empty stack is defined by*

$$M_0^0(A) = \{\mu \in \Sigma^\oplus \mid (q_0, \mu, \langle\perp\rangle) \vdash_0^* (q_f, \mathbf{0}_\Sigma, \mathbf{0}_{\Gamma\cup\{\perp\}}),\ q_0 \in Q_0\}$$

Let $\mathbf{mSA_f^0}$ *(*$\mathbf{mSA_0^0}$*) denote the class of multiset languages accepted by an MSA*^{0} *with final state (empty stack), respectively.*

Definition 3. (multiset storage automaton)
A (nondeterministic) multiset storage automaton *(MFA) is defined similarly to MSA*^{0} *except for the parts where the bottom symbol* $\perp$ *is used. Therefore* K, *the set of edge labels in an MSA is reduced to the set* K_1 *and the notion of the current memory state in a configuration is* $\sigma \in \Gamma^\oplus$.

In order to allow filling the memory at the beginning of a computation with an arbitrary but fixed content we additionally specify the special symbol $\triangle \in \Gamma$ *that is the only initial content of a MSA's memory. Note that contrary to* $\perp$ *in a MSA*^{0}*, a MSA can not test for the presence of* $\triangle$*. So an MSA is a tuple* $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \triangle)$*, where* Q *is a set of states,* Σ, Γ *are disjoint alphabets,* $K \subseteq Q \times \Sigma^\oplus \times \Gamma^\oplus \times \Gamma^\oplus \times Q$ *is the set of edges,* Q_F *and* Q_0 *are the final and the initial states, respectively, and* $\triangle \in \Gamma$ *is a special symbol that is present in the initial configuration. This leaves us at:*

The multiset language *accepted by a MSA A with final state is defined by*

$$M_f(A) = \{\mu \in \Sigma^\oplus \mid (q_0, \mu, \langle\triangle\rangle) \vdash^* (q_f, \mathbf{0}_\Sigma, \sigma),\ q_0 \in Q_0,\ q_f \in Q_F,\ \sigma \in \Gamma^\oplus\}.$$

The multiset language *accepted by a MSA A with empty stack is defined by*

$$M_0(A) = \{\mu \in \Sigma^\oplus \mid (q_0, \mu, \langle\triangle\rangle) \vdash^* (q_f, \mathbf{0}_\Sigma, \mathbf{0}_\Gamma),\ q_0 \in Q_0\}.$$

Let $\mathbf{mSA_f}$ *and* $\mathbf{mSA_0}$ *denote the classes of multiset languages accepted by MSA with final state and empty stack, respectively.*

As already shown in [7], the following relations hold. Please refer to [2], [4] or [7] for notations of Petri net languages.

Theorem 1. $\Psi(\mathcal{L}^\lambda) \subset \mathbf{mSA_f} \subseteq \mathbf{mARB} = \Psi(\mathcal{L}_0^\lambda) = \mathbf{mSA_0}$
$\subseteq \mathbf{mSA_f^0} = \mathbf{mSA_0^0} \subset \mathbf{PsRE}.$ □

3 Closure Properties

Lemma 1. $\mathbf{mSA_f}$ *and* $\mathbf{mSA_0^0}$ *are closed under multiset addition* $\oplus$.

Proof. In order to combine two automata into one that accepts the sum ($\oplus$) of the accepted languages, one can simply run them in sequence after making sure that their memory alphabets Γ are disjoint. In the case of $\mathbf{mSA_f}$, one has to add silent transitions from the final states of the first to the starting state of the second automaton. In the case of $\mathbf{mSA_0^0}$, one needs to add silent transitions that read $\perp$ to make sure that the storage is empty, from every state of the first automaton to the starting state of the second. □

$\mathbf{mSA_0} = \mathbf{mARB}$ has been known to be closed under $\oplus$ as shown in [6].

Lemma 2. $\mathbf{mSA_f}$ *and* $\mathbf{mSA_0^0}$ *are closed under union* $\cup$.

Proof. As for DFA any two MSA$^0 A = (Q_A, \Sigma_A, \Gamma_A, K_A, Q_{FA}, q_{0A}, \perp)$ and $B = (Q_B, \Sigma_B, \Gamma_B, K_B, Q_{FB}, q_{0B}, \perp)$ accepting L_A and L_B respectively, can be combined to an automaton C of the same type accepting $L_A \cup L_B$ by introducing two silent transitions, one for each original automaton, from a newly added starting state (q_{0C}): $C = (Q_A \uplus Q_B, \Sigma_A \cup \Sigma_B, \Gamma_A \cup \Gamma_B, K_C, Q_{FA} \cup Q_{FB}, q_{0C}, \perp)$ $K_C = K_A \cup K_B \cup \{(q_{0C}, \mathbf{0}_\Sigma, \mathbf{0}_\Gamma, \mathbf{0}_\Gamma, q_{0A}), (q_{0C}, \mathbf{0}_\Sigma, \mathbf{0}_\Gamma, \mathbf{0}_\Gamma, q_{0B})\}$. Since this construction does not use the 0-test it also works for MSA . □

Again, for $\mathbf{mSA_0}$ the corresponding result has been known as shown in [6].

Lemma 3. $\mathbf{mSA_f}$*,* $\mathbf{mSA_0}$ *and* $\mathbf{mSA_0^0}$ *are closed under finite substitution* σ.

Proof. In a quasi Σ lettering MSA^0 (MSA) that accepts L, replace every transition (q, α, r, w, q') by the set of transitions $\{(q, a, r, w, q') | a \in \sigma(\alpha)\}$. The resulting automaton accepts $\sigma(L)$. □

Lemma 4. $\mathbf{mSA_0^0}$ *is closed under iteration* $^\oplus$.

Proof. Let $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \perp)$ be a MSA^0 with $M_0^0(A) = M$. A accepts only by a final transition $t = (q, \alpha, \langle\perp\rangle, \mathbf{0}_{\Gamma\cup\{\perp\}}, q')$ because $\perp$ must be the last remaining symbol in the memory. For every such transition and every $q_0 \in Q_0$, add to K a transition $(q, \alpha, \langle\perp\rangle, \langle\perp\rangle, q_0)$ and let A' be the resulting MSA^0. Then $M_0^0(A') = M^\oplus$. □

Lemma 5. *Let $\mathcal{C}$ be a class of multiset languages closed under homomorphisms, sum and finite substitution. Let $\mathcal{C}$ further contain the semilinear sets. Then the following conditions are equivalent:*

a) $\mathcal{C}$ is closed under intersection with sets of the form $\{\mu \in \Sigma^\oplus \mid |\mu|_a = |\mu|_b\}$.
b) $\mathcal{C}$ is closed under intersection with semilinear sets,
c) $\mathcal{C}$ is closed under intersection of any of its members,
d) $\mathcal{C}$ is closed under inverse homomorphisms.

Proof. We will show the implications d)⇒a)⇒c)⇒b)⇒d) and thereby prove the equivalence of all four conditions.

"d)⟹a)". Let $L \subseteq \Sigma^{\oplus}$ be a member of $\mathcal{C}$ and $E = \{\langle a \rangle \oplus \langle b \rangle\}^{\oplus}$, where $a, b \in \Sigma$. Let $\Gamma := \Sigma \setminus \{b\}$ and $h : \Gamma^{\oplus} \to \Sigma^{\oplus}$ be the homomorphism defined by $h(a) = \langle a \rangle \oplus \langle b \rangle$, $h(x) = x$ for $x \in \Gamma$, $x \neq a$. Then

$$L \cap E = h(h^{-1}(L)).$$

Therefore, the implication d)⟹a) holds.

"a)⟹c)". Let $L, K \subseteq \Sigma^{\oplus}$, where $L, K \in \mathcal{C}$. Let $\hat{\Sigma} := \{\hat{x} \mid x \in \Sigma\}$ be a set of fresh symbols. Let $h : (\Sigma \cup \hat{\Sigma})^{\oplus} \to \Sigma^{\oplus}$ and $g : \Sigma^{\oplus} \to \hat{\Sigma}^{\oplus}$ be the homomorphisms defined by $h(x) = x$, $h(\hat{x}) = \mathbf{0}_{\Sigma}$, $g(x) = \hat{x}$ for $x \in \Sigma$. Furthermore, for $x \in \Sigma$, let $E_x = \{\mu \in (\Sigma \cup \hat{\Sigma})^{\oplus} \mid |\mu|_x = |\mu|_{\hat{x}}\}$. Then obviously,

$$L \cap K = h\left((L \oplus g(K)) \cap \bigcap_{x \in \Sigma} E_x\right).$$

This proves a)⟹c).

"c)⟹b)". Since $\mathcal{C}$ contains the semilinear sets, the implication c)⟹b) is clear.

"b)⟹d)". Let $h : \Sigma^{\oplus} \to \Gamma^{\oplus}$ be a homomorphism and $L \subseteq \Gamma^{\oplus}$ be a member of $\mathcal{C}$. Let $\Lambda := \{x \in \Sigma \mid h(x) = \mathbf{0}_{\Gamma}\}$, and $\Delta := \Sigma \setminus \Lambda$. Write $\Delta = \{x_1, \ldots, x_n\}$. Furthermore, let $h(x_i) = y_i \oplus z_i$ be an arbitrary decomposition such that $y_i \in \Gamma$ and $z_i \in \Gamma^{\oplus}$, $i = 1, \ldots, n$. Let $\Xi = \{\xi_1, \ldots, \xi_n\}$ be a set of fresh symbols, $\tilde{\Gamma} := \Gamma \cup \Xi$,

$$S := \{\langle \xi_i \rangle \oplus z_i \mid 1 \leq i \leq n\}^{\oplus}.$$

Let $\sigma : \Gamma^{\oplus} \to 2^{\tilde{\Gamma}^{\oplus}}$ be the substitution defined by

$$\sigma(x) = \{x\} \cup \{\xi_i \in \Xi \mid x = y_i, 1 \leq i \leq n\}.$$

To complete the construction, let $g : \tilde{\Gamma}^{\oplus} \to \Delta^{\oplus}$ the homomorphism defined by $g(\xi_i) = x_i$ for $1 \leq i \leq n$ and $g(x) = \mathbf{0}_{\Delta}$ for every $x \in \Gamma$. Now we have

$$h^{-1}(L) = g(\sigma(L) \cap S) \oplus \Lambda^{\oplus}.$$

Since S is semilinear and $\Lambda^{\oplus}$ is semilinear and therefore contained in $\mathcal{C}$, this proves b)⟹d). □

Table 1 gives an overview over the closure properties. Since all three classes are closed under homomorphisms and sum, lemma 5 yields the equivalence of closure under arbitrary intersection and closure under semilinear intersection. We therefore omit the entry for semilinear intersection. The abbreviations $h(L)$, $\sigma(L)$ stand for arbitrary homomorphisms and finite substitution, respectively. + means that the class is closed under the respective operator. An empty entry means that we don't know yet whether the closure holds or not. The results for $\mathbf{mSA_0}$, except for iteration $^{\oplus}$ were obtained in [6].

	$\mathbf{mSA_f}$	$\mathbf{mSA_0}$	$\mathbf{mSA_0^0}$
$L_1 \oplus L_2$	+	+	+
$L_1 \ominus L_2$		+	
$L_1 \cup L_2$	+	+	+
$L_1 \cap L_2$		+	
$L_1 \sqcup L_2$		+	
$L_1 \sqcap L_2$		+	
$\sigma(L)$	+	+	+
$L^{\oplus}$			+
$h(L)$	+	+	+

Table 1. Closure properties

4 Algebraic characterizations

In [7], we proved the inclusion $\Psi(\mathcal{L}^{\lambda}) \subset \mathbf{mSA_f} \subseteq \mathbf{mSA_0} \subseteq \mathbf{mSA_0^0}$. In this paper, we will give descriptions of $\mathbf{mSA_0}$ and $\mathbf{mSA_0^0}$ in terms of $\mathbf{mSA_f}$ and $\mathbf{mSA_0}$, respectively. This way, we obtain conditions that are equivalent to the strictness of the inclusions.

The classes $\mathbf{mSA_f}$ and $\mathbf{mSA_0}$ can be compared using the following theorem.

Theorem 2. *For every $L \in \mathbf{mSA_0}$, $L \subseteq \Sigma^{\oplus}$, there is a $K \in \Psi(\mathcal{L}^{\lambda})$, $K \subseteq \Gamma^{\oplus}$, a semilinear set $S \subseteq \Gamma^{\oplus}$ and a homomorphism $h : \Gamma^{\oplus} \to \Sigma^{\oplus}$ such that $L = h(K \cap S)$.*

Proof. Let $L \subseteq \Sigma^{\oplus}$ be a member of $\mathbf{mSA_0} = \Psi(\mathcal{L}_0^{\lambda})$ and $N = (P, T, \tau, \Sigma, \rho, \mu_0, R_f)$ be a Petri net such that $\Psi(L_f(N)) = L$. Write $T = \{t_1, \ldots, t_n\}$, let $\Gamma = \{\xi_1, \ldots, \xi_n\}$ be a set of fresh symbols and $\tilde{\rho} : T \to \Gamma$ the mapping with $\tilde{\rho}(t_i) = \xi_i$ for $1 \leq i \leq n$. Furthermore, let $h : \Gamma^{\oplus} \to \Sigma^{\oplus}$ the homomorphism satisfiying $h(\xi_i) = \rho(t_i)$ for $1 \leq i \leq n$. Let N' be the Petri net $(P, T, \tau, \Gamma, \tilde{\rho}, \mu_0, R_f)$ and $K := \Psi(L(N'))$.

Analogous to the free commutative monoid, we will need the free abelian group generated by a certain set. For a set A, let $A^{\otimes}$ denote the set of mappings $A \to \mathbb{Z}$. Equipped with the operation $+$, $(\mu + \nu)(a) = \mu(a) + \nu(a)$, $A^{\otimes}$ is an abelian group satisfiying $A^{\oplus} \subseteq A^{\otimes}$.

Now we can define the homomorphism $\delta : T^{\oplus} \to P^{\otimes}$. For every $t \in T$, we set $\delta(t) := \mu_1 - \mu_0$, where $\tau(t) = (\mu_0, \mu_1)$.

Furthermore, let

$$S := \{\nu \in \Gamma^{\oplus} \mid \mu_0 + \delta(\tilde{\rho}^{-1}(\nu)) \in R_f\}.$$

Since $\delta \circ \tilde{\rho}^{-1}$ is a homomorphism $\Gamma^{\oplus} \to P^{\otimes}$ and R_f is finite, Corollary 3 implies that S is a semilinear set.

To complete the proof, we show that $L = h(K \cap S)$, so let $\nu \in L$. Then there is a firing sequence $\sigma \in T^*$ in the net N, such that $\mu_0 \to_{\sigma} \mu$ for some $\mu \in R_f$ and $\rho(\Psi(\sigma)) = \nu$. $\mu_0 \to_{\sigma} \mu$ implies that $\tilde{\rho}(\sigma) \in L(N')$ and therefore

$\nu' := \Psi(\tilde{\rho}(\sigma)) \in K$. On the other hand, $\mu_0 \to_\sigma \mu$ implies $\mu_0 + \delta(\Psi(\sigma)) = \mu \in R_f$, so

$$\mu_0 + \delta(\tilde{\rho}^{-1}(\nu')) = \mu_0 + \delta(\Psi(\sigma)) = \mu \in R_f$$

and therefore $\nu' \in S$. Since $\Psi \circ \rho = h \circ \Psi \circ \tilde{\rho}$, we have

$$\nu = \Psi(\rho(\sigma)) = h(\Psi(\tilde{\rho}(\sigma))) = h(\nu'),$$

so $\nu \in h(K \cap S)$.

Now let $\nu \in h(K \cap S)$, say $\nu = h(\nu')$ where $\nu' \in K \cap S$. The condition $\nu' \in K = \Psi(L(N'))$ yields a firing sequence $\sigma \in T^*$ such that $\nu' = \Psi(\tilde{\rho}(\sigma))$ and $\mu_0 \to_\sigma \mu$ for some $\mu \in P^\oplus$. Since $\nu' \in S$, we have

$$\mu = \mu_0 + \delta(\Psi(\sigma)) = \mu_0 + \delta(\tilde{\rho}^{-1}(\nu')) \in R_f$$

and therefore $\rho(\sigma)) \in L_f(N)$. This however, implies

$$\nu = h(\nu') = h(\Psi(\tilde{\rho}(\sigma))) = \Psi(\rho(\sigma)) \in \Psi(L_f(N)) = L.$$

□

This can be used to describe the relation between the classes $\mathbf{mSA_f}$ and $\mathbf{mSA_0}$. Since $\mathbf{mSA_f}$ is closed under homomorphisms and $\Psi(\mathcal{L}^\lambda) \subset \mathbf{mSA_f}$, it is clear that $\mathbf{mSA_f} = \mathbf{mSA_0}$ if $\mathbf{mSA_f}$ is closed under intersection with semilinear sets. Furthermore, $\mathbf{mSA_0}$ is closed under intersection with semilinear sets, this is Lemma 7 in [6]. Therefore, the next theorem follows immediately.

Theorem 3. $\mathbf{mSA_f} = \mathbf{mSA_0}$ *iff* $\mathbf{mSA_f}$ *is closed under intersection with semilinear sets.*

For the classes $\mathbf{mSA_0}$ and $\mathbf{mSA_0^0}$, an analogous theorem can be established. In order to do that, we will prove the following.

Theorem 4. *For every* $L \in \mathbf{mSA_0^0}$*, there exist* $L_i, K_i \in \mathbf{mSA_0}$*,* $i = 1, \dots, n$*, such that* $L = \bigcup_{i=1}^n L_i \oplus K_i^\oplus$.

The proof however, requires some preparation.

Definition 4. *Let* $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \bot)$ *be an MSA*^{0}*. Then the relation* $\Longrightarrow_A$ *on the configurations is defined as follows.* $(q, \mu, \sigma) \Longrightarrow_A (q', \mu', \sigma')$ *iff there is an* $(q, \alpha, \gamma, \delta, q') \in K$ *such that* $\gamma, \delta \in \Gamma^\oplus$*,* $\gamma \sqsubseteq \sigma$*,* $\mu = \mu' \oplus \alpha$ *and* $\sigma' = \sigma \ominus \gamma \oplus \delta$*. In other words,* $\Longrightarrow_A$ *describes exactly the steps not involving* $\bot$*-edges. Now for* $q, q' \in Q$ *and* $\sigma \in \Gamma^\oplus$*, let*

$$N_A(q, \sigma, q') := \{\mu \in \Sigma^\oplus \mid (q, \mu, \sigma) \Longrightarrow_A^* (q', \mathbf{0}_\Sigma, \mathbf{0}_\Gamma)\}.$$

Lemma 6. *Let* $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \bot)$ *be an MSA*^{0} *and* $q, q' \in Q$*,* $\sigma \in \Gamma^\oplus$*. Then* $N_A(q, \sigma, q') \in \mathbf{mSA_0}$.

Proof. Construct an MSA $B = (Q, \Sigma, \Gamma, K', \{q'\}, \{q\}, \triangle)$, where

$$K' := \{(q_1, \alpha, \gamma, \delta, q_2) \in K \mid \gamma \in \Gamma^\oplus\} \cup \{(q, \mathbf{0}_\Sigma, \langle\triangle\rangle, \mathbf{0}_\Gamma, q)\}.$$

Then obviously, $N_A(q, \sigma, q') = M_0(B)$. □

Lemma 7. *For every* $L \in \mathbf{mSA_0^0}$*, there is an* $R \in \mathbf{mREG}$ *and a* $\mathbf{mSA_0}$*-substitution*[1] σ *such that* $L = \sigma(R)$*.*

Proof. Let $L = M_0^0(A)$ with $A = (Q, \Sigma, \Gamma, K, Q_F, Q_0, \perp)$. Every calculation of A can be decomposed into *phases*, seperated by $\perp$-steps. A $\perp$-step is a step involving a $\perp$-edge and a phase is a sequence of non-$\perp$-steps. In the MFA $B = (Q', \Sigma', Q'_F, q_0, K')$, every state-transition either represents a phase or a $\perp$-step. Therefore, it contains two kinds of states and two kinds of edges.

The set of states is constructed as follows. Every state in Q'_1 either corresponds to the initial configuration of A or to a configuration reached after a $\perp$-step:

$$\begin{aligned} Q'_1 := &\{(q', \delta) \mid \exists q \in Q, \mu \in \Sigma^{\oplus} : (q, \mu, \langle\perp\rangle, \delta, q') \in K\} \\ &\cup\ \{(q, \langle\perp\rangle) \mid q \in Q_0\}. \end{aligned}$$

The states in Q'_2 correspond to configurations obtained by executing a phase and having stack content $\langle\perp\rangle$ or $\mathbf{0}_\Gamma$. Thus, the states in Q'_2 correspond to those configurations entered by $\perp$-edges:

$$Q'_2 := \{(q, \mathbf{0}_\Gamma), (q, \langle\perp\rangle) \mid q \in Q\}.$$

Now $Q' := Q'_1 \cup Q'_2$ and $Q'_F := \{(q, \mathbf{0}_\Gamma) \mid q \in Q_F\}$.

To account for the different types of actions simulated by the edges, the input alphabet of the MFA is also split into two parts. Σ_1 contains the symbols written on phase-edges:

$$\Sigma_1 := \{(q, \delta, q') \mid (q, \delta) \in Q'_1, q' \in Q\},$$

The edges that simulate $\perp$-steps are labeled with symbols from Σ. Therefore $\Sigma' := \Sigma_1 \cup \Sigma$.

The set K' of edges is $K' = K'_1 \cup K'_2$. Thereby, K'_1 contains the edges simulating the phases, while those in K'_2 simulate the $\perp$-steps:

$$K'_1 := \{((q, \langle\perp\rangle \oplus \delta),\ (q, \delta, q'),\ (q', \langle\perp\rangle)) \mid (q, \delta) \in Q'_1,\ \ q' \in Q\},$$

$$K'_2 := \{((q, \langle\perp\rangle),\ \mu,\ (q', \delta)) \mid (q, \mu, \langle\perp\rangle, \delta, q') \in K\}.$$

To complete the construction, the substitution σ is defined. The symbols in Σ_1 are mapped to sets of multisets that can be read while executing a phase (i.e. languages of the form $N_A(q, \delta, q')$): $\sigma((q, \delta, q')) := N_A(q, \delta, q')$ for all $(q, \delta, q') \in \Sigma_1$. Since the edges simulating $\perp$-steps are labeled with the symbols read by the corresponding edge in A, this can be carried over to the substitution: $\sigma(a) = \{a\}$ for all $a \in \Sigma$. Now for every $x \in \Sigma'$, it is $\sigma(x) \in \mathbf{mSA_0}$, so σ is a $\mathbf{mSA_0}$-substitution. Furthermore, it is obvious that $\sigma(R) = L$. $\square$

Now we are ready to prove Theorem 4.

[1] a substitution $\sigma : \Sigma'^{\oplus} \to 2^{\Sigma^{\oplus}}$ with $\sigma(x) \in \mathbf{mSA_0}$ for all $x \in \Sigma'$, if $L \subseteq \Sigma^{\oplus}$ and $R \subseteq \Sigma'^{\oplus}$

Proof (Theorem 4). Apply Lemma 7 and write $L = \sigma(R)$ where $R \in \mathbf{mREG}$, $R \subseteq \Gamma^{\oplus}$, and σ is an $\mathbf{mSA_0}$-substitution. Since R is semilinear, it is

$$R = \bigcup_{i=1}^{n} \{x_i\} \oplus \{y_{i,1}, \ldots, y_{i,k_i}\}^{\oplus}$$

for certain $x_i, y_{i,j} \in \Gamma^{\oplus}$, $1 \leq i \leq n$, $1 \leq j \leq k_i$. The closure of $\mathbf{mSA_0}$ against addition and union implies that there are $K_i, L_i \in \mathbf{mSA_0}$, $1 \leq i \leq n$, such that $\sigma(x_i) = K_i$ and $\sigma(\{y_{i,1}, \ldots, y_{i,k_i}\}) = L_i$. With these, it follows

$$L = \sigma(R) = \bigcup_{i=1}^{n} K_i \oplus L_i^{\oplus}.$$

□

Since $\mathbf{mSA_0}$ is closed under sum and union, it is clear that $\mathbf{mSA_0} = \mathbf{mSA_0^0}$ if $\mathbf{mSA_0}$ is closed under iteration. Furthermore, $\mathbf{mSA_0^0}$ is closed unter iteration and therefore the converse also holds.

Corollary 1. $\mathbf{mSA_0} = \mathbf{mSA_0^0}$ *iff* $\mathbf{mSA_0}$ *is closed under iteration.*

Theorem 4 also yields a useful property of $\mathbf{mSA_0^0}$. Since every language over one symbol in $\mathbf{mSA_0} = \Psi(\mathcal{L}_0^{\lambda})$ is regular, as shown in [3], and the class of regular languages is closed under iteration, sum and union, we obtain the following corollary.

Corollary 2. *Every language in* $\mathbf{mSA_0^0}$ *over one symbol is regular.*

5 Conclusion

So far we did not consider the deterministic versions of multiset automata. Several different variants of such are possible. The investigation of such deterministic multiset finite and storage automata will be given in another paper.

References

1. E. Csuhaj-Varjú, C. Martín Vide, V. Mitrana (2001): *Multiset Automata.* In: Multiset Processing, eds. C. Calude, Gh. Păun, G. Rozenberg, A. Salomaa, *LNCS* **2235**, 69-83.
2. M. Hack (1976): *Petri Net Languages.* MIT Technical Report 159.
3. D. Hauschildt and M. Jantzen. Petri net algorithms in the theory of matrixgrammars, Acta Informatica, vol. 31 (1994) 719-728.
4. M. Jantzen (1979): *On the Hierarchy of Petri Net Languages. RAIRO* ITTI, **13**, nr. 1, 19-30.
5. M. Kudlek, C. Martín Vide, Gh. Păun (2001): *Toward FMT (Formal Macroset Theory).* In: Multiset Processing, eds. C. Calude, Gh. Păun, G. Rozenberg, A. Salomaa, *LNCS* **2235**, 123-134.
6. M. Kudlek, V. Mitrana (2001): *Closure Properties of Multiset Language Families. Fund. Inf.*, **49**, 191-203.
7. M. Kudlek, P. Totzke, G. Zetzsche (2008): *Multiset Storage Automata. CS&P 2008*

A Commutative Monoids

This section contains some information about commutative monoids needed in section 4. For a commutative monoid M, we use $+$ as the symbol for its operation, if there is no danger of confusion. Furthermore, for a subset $T \subseteq M$, by $T^{\oplus}$ we denote the submonoid of M generated by T (i.e. the smallest submonoid of M containing T). The relation $\sqsubseteq$ is also generalized to commutative monoids: If M is a commutative monoid and $m, m' \in M$, we write $m \sqsubseteq m'$ iff there is a $n \in M$ such that $m' = m + n$.

Definition 5. *Let M be a commutative monoid. A set $S \subseteq M$ is called* lower bound *for $T \subseteq M$, iff $S \subseteq T \setminus \{0\}$ and for every $t \in T \setminus \{0\}$ there is an $s \in S$ with $s \sqsubseteq t$.*

Lemma 8. *Every set $S \subseteq \mathbb{N}^k$ has a finite lower bound with respect to $\sqsubseteq$.*

Proof. We construct the finite lower bound by an induction on k. The case $k = 1$ is trivial, so let $k > 1$. Choose an element $x = (x_1, \ldots, x_k) \in S \setminus \{0\}$. We will need the index set

$$I := \{(i, j) \in \mathbb{N}^2 \mid 1 \leq i \leq k,\ 0 \leq j < x_i\}.$$

Now let

$$S_{i,j} := \{t \in S \mid t_i = j\}, \quad (i, j) \in I.$$

and consider the mapping

$$\begin{aligned} \varphi_{i,j} : S_{i,j} &\longrightarrow \mathbb{N}^{k-1}, \quad (i, j) \in I \\ \varphi_{i,j}(t_1, \ldots, t_k) &= (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k). \end{aligned}$$

Obviously, $\varphi_{i,j}$ is an injective homomorphism that satisfies

$$t \sqsubseteq t' \Longleftrightarrow \varphi_{i,j}(t) \sqsubseteq \varphi_{i,j}(t')$$

for all $t, t' \in S_{i,j}$. By induction hypothesis, for every set

$$T_{i,j} := \varphi_{i,j}(S_{i,j}) \subseteq \mathbb{N}^{k-1}$$

we can find a finite lower bound $T'_{i,j}$. We will show that the finite set

$$S' := \{x\} \cup \bigcup_{(i,j) \in I} \varphi_{i,j}^{-1}(T'_{i,j})$$

is a lower bound of S. Therefore, $s \in S$ is arbitrarily chosen with $x \not\sqsubseteq s$. Then, there is an index i with $j := s_i < x_i$, i.e. $s \in S_{i,j}$. $T'_{i,j}$ now contains an element s' with $s' \sqsubseteq \varphi_{i,j}(s)$, hence we have $s \sqsupseteq \varphi_{i,j}^{-1}(s') \in \varphi_{i,j}^{-1}(T'_{i,j}) \subseteq S'$. ◻

Definition 6. *Let M be a commutative monoid. A subset $S \subseteq M$ is called* quasi-invertible *iff $b \in S$ holds whenever $a, b \in M$ satisfy $a \in S$ and $a + b \in S$.*

For example, every kernel of a homomorphism is a quasi-invertible submonoid. Or, slightly more general, if $h : M \to M'$ is a homomorphism of commutative monoids and $S \subseteq M'$ is quasi-invertible, then $h^{-1}(S) \subseteq M$ is quasi-invertible as well.

Lemma 9. *Let M be a finitely generated commutative monoid and $S \subseteq M$ a quasi-invertible submonoid. Then S is finitely generated.*

Proof. Since M is finitely generated, there exists a surjective homomorphism $h : \mathbb{N}^k \to M$ for some k. Therefore, it suffices to show that $T := h^{-1}(S)$ is finitely generated.

It is easily seen that T is quasi-invertible. Furthermore, with Lemma 8 we can find a finite lower bound T' for T. We will show that T' generates T. Let $t \in T$ and $k = |t|$. An induction on k will prove $t \in T'^{\oplus}$. In the case $k = 1$ we have $t \in T'$ and for $k > 1$ there is a $t' \in T'$ with $t' \sqsubseteq t$. Thus there is a $t'' \in \mathbb{N}^k$ with $t = t' + t''$ and $|t''| < k$. Since T is quasi-invertible, we have $t'' \in T$. Hence, the induction hypothesis implies $t'' \in T'^{\oplus}$, and therefore $t = t' + t'' \in T'^{\oplus}$. Thus $T = T'^{\oplus}$. □

Theorem 5. *Let M be a finitely generated commutative monoid, G an abelian group and $\varphi : M \to G$ a homomorphism. Then for every $g \in G$, the set $S = \varphi^{-1}(g)$ is semilinear.*

Proof. Let S' be a finite lower bound for S. Since the kernel of φ is a quasi-invertible submonoid of M, it is finitely generated: $\ker\varphi = F^{\oplus}$ for some finite set $F \subseteq \ker\varphi$. We claim that

$$S = \bigcup_{x \in S'} \{x\} + F^{\oplus}. \tag{1}$$

Let $s \in S$ be arbitrary. Since S' is lower bound, there is an $x \in S'$ such that $x \sqsubseteq s$, so there is a $y \in M$ such that $s = x + y$. Since $x \in S$, we have $\varphi(x) = b$ which implies

$$b + \varphi(y) = \varphi(x + y) = \varphi(s) = b,$$

so $\varphi(y) = 0$ and $y \in F^{\oplus}$.

For $x \in S'$ and $y \in F^{\oplus}$, we have

$$\varphi(x + y) = \varphi(x) + \varphi(y) = b,$$

so $x + y \in S$. Therefore equation (1) holds and S is semilinear. □

Corollary 3. *Let M be a finitely generated commutative monoid, G an abelian group and $\varphi : M \to G$ a homomorphism. Then for every $g_0 \in G$ and every finite subset $H \subseteq G$, the set $S = \{m \in M \mid g_0 + \varphi(m) \in H\}$ is semilinear.*

Proof. Let H' be the finite set $\{g - g_0 \mid g \in H\}$. Then $\varphi^{-1}(h)$ is semilinear for every $h \in H'$ and therefore $S = \bigcup_{h \in H'} \varphi^{-1}(h)$ is semilinear as a finite union of semilinear sets. □