The Support of Unranked Weighted Tree Automata

Manfred Droste   Doreen Götze

Universität Leipzig

27.09.2013

- 1980s: theory of tree automata and tree languages (Gécseg, Steinby)
    - used to analyze and specify systems

- quantitative aspects, like maximum costs, became interesting

    $\curvearrowright$ weighted ranked tree automata over semirings

    - results summarized by Fülöp, Vogler in a chapter of "Handbook of Weighted Automata" (eds. Droste, Kuich, Vogler)

- for some application unranked trees are more convenient, e.g. (fully structured) XML-documents

$\to$ Droste, Vogler: weighted unranked tree automata (2009)

Non-Emptyness Problem

Is $supp(\mathcal{M}) \neq \emptyset$ for a given weighted unranked tree automaton $\mathcal{M}$?

Definition (Support)

$\mathcal{M}$ weighted unranked tree automaton.

$$supp(\mathcal{M}) = \{\xi \in U_\Sigma \mid (r_\mathcal{M}, \xi) \neq 0\}$$

- since the 80's: support of probabilistic automata has been studied e.g. in books by Paz, Bukharaev
- $\exists$ weighted automata with non-recognizable support
- effective construction of support automata for weighted automata over commutative, zero-sum free semirings (Kirsten)

Goals

- define weighted unranked tree automata over strong bimonoids
- effective construction of support automata for
  weighted unranked tree automata
  over commutative, zero-sum free strong bimonoids

Strong Bimonoids

$\mathbb{K} = (K, +, \cdot, \mathbb{0}, \mathbb{1})$:

- semiring without distributivity

$\rightarrow$ $(K, +, \mathbb{0})$ and $(K, \cdot, \mathbb{1})$ monoids s.th.

  - $(K, +, \mathbb{0})$ commutative & $\mathbb{0}$ is multiplicative zero

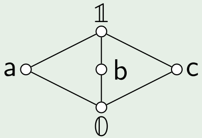
- called commutative if $(K, \cdot, \mathbb{1})$ commutative
- called zero-sum free if $k + l = \mathbb{0} \Rightarrow k = l = \mathbb{0}$

Example

$(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ with $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ and $x \cdot_4 y = (x \cdot y) \mod 4$

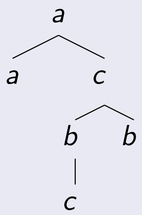
Example

from multi-valued logics:
all bounded lattices $(\mathcal{L}, \vee, \wedge, \mathbb{0}, \mathbb{1})$,
e.g.:



Trees

finite tree over unranked alphabet Σ



- U_Σ: set of all trees over Σ

Weighted Tree Automata over Strong Bimonoids

Definition (weighted unranked tree automata (wuta) over $\mathbb{K}$, Σ)

$\mathcal{M} = (Q, \mathcal{A}, \gamma)$ such that

- Q: non-empty, finite set of states
- $\gamma : Q \to \mathbb{K}$: root weight function
- $\mathcal{A}$: family of weighted word automata $\mathcal{A}_{q,a}$ ($q \in Q, a \in \Sigma$) over $\mathbb{K}$, Q

Definition (behavior of wuta)

tree series $r_{\mathcal{M}} : U_\Sigma \to \mathbb{K}$ with

$$(r_{\mathcal{M}}, \xi) = \sum_{\substack{(q,s,t) \text{ is valid} \\ \text{extended run on } \xi}} \mathrm{wt}(q, s, t) \cdot \gamma(q)$$

Valid Extended Runs (Droste, Vogler 2011)

Main Idea of Valid Extended Run (q, s, t) on tree ξ

- imagine an assignment from each position w of ξ to a state q_w s.th.

  $\rightarrow$ $q_\varepsilon = q$

- $\mathcal{A}_{q_w,\xi(w)}$ is called on $q_{w.1} \dots q_{w.\mathrm{rk}_\xi(w)}$

  - if w no leaf: s chooses one run $(p^w_{i-1}, q_{w.i}, p^w_i)_{1 \le i \le \mathrm{rk}_\xi(w)}$

  $\rightarrow$ $s(w.i) = (p^w_{i-1}, q_{w.i}, p^w_i)$

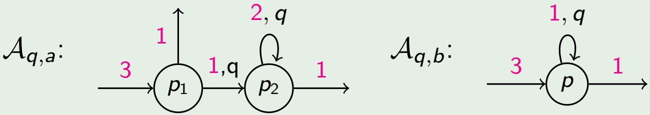
  $\rightarrow$ $s(\varepsilon)$ undefined

  - if w leaf: t chooses state p^w

  $\rightarrow$ $t(w) = p^w$

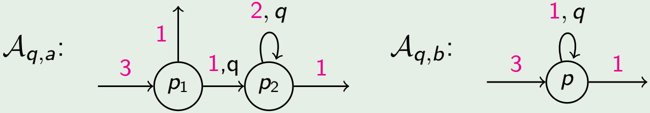
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and



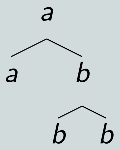
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and



- valid extended run (q, s, t) on tree ξ_{ex}:

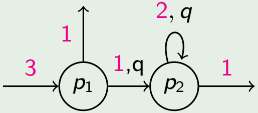
tree ξ_{ex}



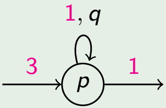
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and

$\mathcal{A}_{q,a}$:

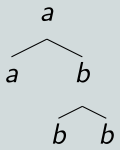


$\mathcal{A}_{q,b}$:

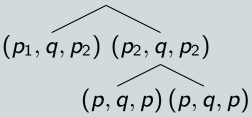


- valid extended run (q, s, t) on tree ξ_{ex}:

tree ξ_{ex}

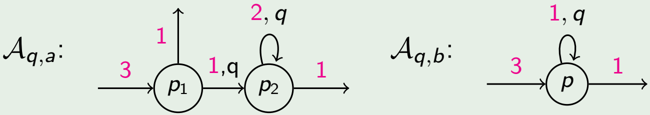


s

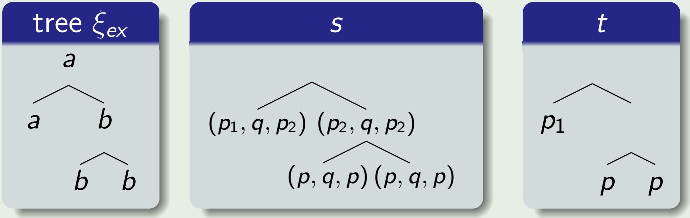


Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and



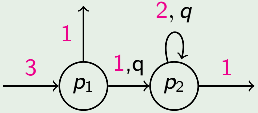
- valid extended run (q, s, t) on tree ξ_{ex}:



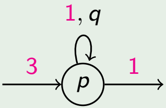
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and

$\mathcal{A}_{q,a}$:

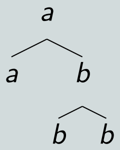


$\mathcal{A}_{q,b}$:

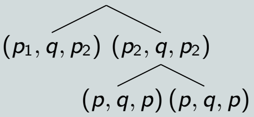


- valid extended run (q, s, t) on tree ξ_{ex}:

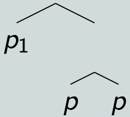
tree ξ_{ex}



s



t

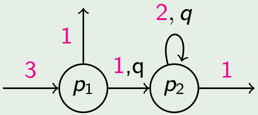


- if w no leaf: $\mathrm{wt}(q, s, t)_w = \mathrm{wt}(s(w.i)_{1 \leq i \leq \mathrm{rank}_\xi(w)})$
- if w leaf: $\mathrm{wt}(q, s, t)_w = \lambda(t(w)) \cdot \nu(t(w))$

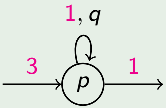
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and

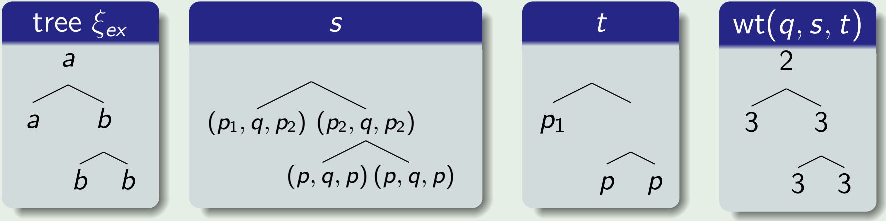
$\mathcal{A}_{q,a}$:



$\mathcal{A}_{q,b}$:



- valid extended run (q, s, t) on tree ξ_{ex}:

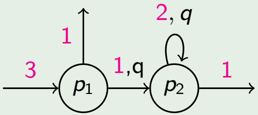


- if w no leaf: $\mathrm{wt}(q, s, t)_w = \mathrm{wt}(s(w.i)_{1 \le i \le \mathrm{rank}_\xi(w)})$
- if w leaf: $\mathrm{wt}(q, s, t)_w = \lambda(t(w)) \cdot \nu(t(w))$

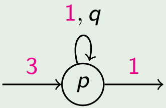
Example for a wuta over $(\mathbb{Z}_4, \max, \cdot_4, 0, 1)$ and $\Sigma = \{a, b\}$

- wuta $\mathcal{M}_{ex}$ with $Q = \{q\}$, $\gamma(q) = 2$ and

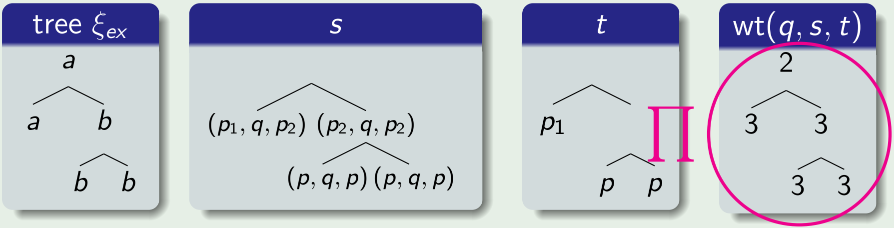
$\mathcal{A}_{q,a}$:



$\mathcal{A}_{q,b}$:



- valid extended run (q, s, t) on tree ξ_{ex}:



$= 2$

- if w no leaf: $\mathrm{wt}(q, s, t)_w = \mathrm{wt}(s(w.i)_{1 \le i \le \mathrm{rank}_\xi(w)})$
- if w leaf: $\mathrm{wt}(q, s, t)_w = \lambda(t(w)) \cdot \nu(t(w))$

Support Automata Construction

$\mathbb{K}$ zero-sum free, commutative strong bimonoid,
$\mathcal{M}$ wuta over $\mathbb{K}$,Σ.

Goal

- build unranked tree automaton $\mathcal{M}_s$ with

$$\mathcal{L}(\mathcal{M}_s) = supp(r_{\mathcal{M}})$$

- unranked tree automaton (uta) over Σ: $\mathcal{M}_s = (Q, \delta, F)$ s. th.
  - Q: non-empty, finite set of states
  - $F \subseteq Q$: set of final states
  - $\delta : Q \times \Sigma \to \{A_{q,a} \text{ over Q} \mid a \in \Sigma, q \in Q\}$ s.th. $\delta(q, a) = \mathcal{A}_{q,a}$
- successful run on ξ: $\kappa : \mathrm{pos}(\xi) \to Q$ with
  - $\kappa(w.1) \dots \kappa(w.\mathrm{rk}_{\xi(w)}) \in \mathcal{L}(\delta(\kappa(w), \xi(w)))$ for all positions w
  - $\kappa(\varepsilon) \in F$
- $\mathcal{L}(\mathcal{M}_s) = \{\xi \in U_\Sigma \mid \exists$ successful run of $\mathcal{M}_s$ on $\xi\}$

Construction of $\mathcal{M}_s$:

- $\mathcal{M}_s$ runs should simulates valid extended runs
- $(r_{\mathcal{M}}, w) \in \langle c_1, \ldots, c_n \rangle \subseteq \mathbb{K}$

$\rightarrow$ states of $\mathcal{M}_s$ counts use of $c_1, \ldots, c_n$ in $\mathcal{M}$

- Kirsten (with Dickson's lemma):
  $\exists y \in \mathbb{N}$: only have to count until $y \Rightarrow$ finitely many states
  $\rightarrow$ called $\mathrm{dg}(C)$
- state final if generators counted do not generate $\mathbb{0}$

for all sub-automata $A_{q,a} = (Q_{q,a}, \lambda_{q,a}, \mu_{q,a}, \nu_{q,a})$:

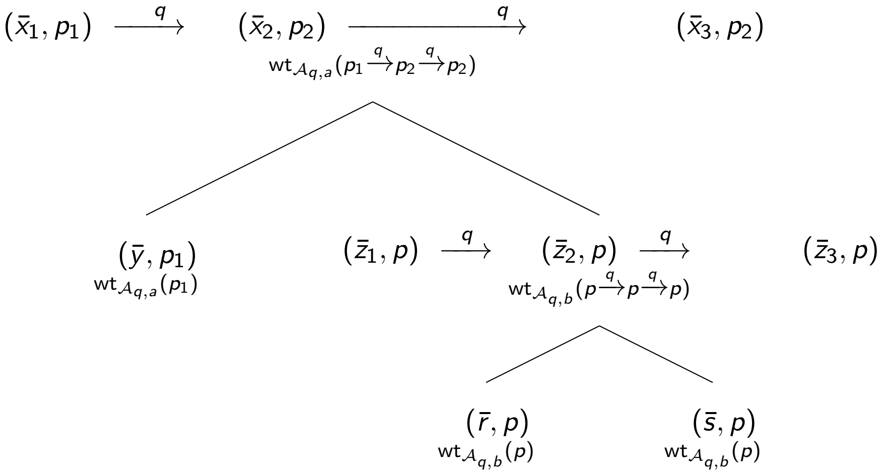
- adopting Kirsten's construction
- $\mathcal{A}_{q,a,s} = (Q_{q,a,s}, T_{q,a,s}, I_{q,a,s}, F_{q,a,s})$
  - $Q_{q,a,s} = \{0, \ldots, \mathrm{dg}(\mathrm{C})\}^n \times Q_{q,a}$
  - $(\bar{x}, p) \in I_{q,a,s}$ iff $x_i = \begin{cases} 1 & \text{if } \lambda_{q,a}(p) = c_i, \\ 0 & \text{if } \lambda_{q,a}(p) \neq c_i \end{cases}$
  - $((\bar{x}, p), b, (\bar{y}, p')) \in T_{q,a,s}$ iff $\bar{y} = \bar{x} \oplus \mu_{q,a}(p, b, p')$
  - $(\bar{x}, p) \in F_{q,a,s}$ iff $[\![\bar{x} \oplus \nu_{q,a}(p)]\!] \neq \mathbb{0}$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p_1) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p_1)$$


$(\bar{x}_1, p_1) \xrightarrow{q} (\bar{x}_2, p_2) \xrightarrow{q} (\bar{x}_3, p_2)$
$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2)$
$(\bar{y}, p_1)$
$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1)$
$(\bar{z}_1, p) \xrightarrow{q} (\bar{z}_2, p) \xrightarrow{q} (\bar{z}_3, p)$
$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p)$
$(\bar{r}, p)$
$\mathrm{wt}_{\mathcal{A}_{q,b}}(p)$
$(\bar{s}, p)$
$\mathrm{wt}_{\mathcal{A}_{q,b}}(p)$

$$(\bar{x}_1, p_1) \xrightarrow{(\quad ,q)} (\bar{x}_2, p_2) \xrightarrow{(\qquad\qquad\qquad ,q)} (\bar{x}_3, p_2)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \quad (\qquad\qquad\qquad ,q)$$

$$(\bar{y}, p_1)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \quad (\quad ,q)$$

$$(\bar{z}_1, p) \xrightarrow{(\quad ,q)} (\bar{z}_2, p) \xrightarrow{(\quad ,q)} (\bar{z}_3, p)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \quad (\qquad\qquad\qquad ,q)$$

$$(\bar{r}, p) \qquad\qquad (\bar{s}, p)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\quad ,q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\quad ,q)$$

$(\bar{x}_1, p_1) \xrightarrow{(\quad ,q)} (\bar{x}_2, p_2) \xrightarrow{(\qquad\qquad ,q)} (\bar{x}_3, p_2)$

$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \quad (\qquad\qquad ,q)$

$(\bar{y}, p_1)$

$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \quad (\bar{y}',q)$

$(\bar{z}_1, p) \xrightarrow{(\quad ,q)} (\bar{z}_2, p) \xrightarrow{(\quad ,q)} (\bar{z}_3, p)$

$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \quad (\qquad\qquad ,q)$

$(\bar{r}, p)$

$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{r}',q)$

$(\bar{s}, p)$

$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{s}',q)$

$$(\bar{x}_1, p_1) \xrightarrow{(\quad ,q)} (\bar{x}_2, p_2) \xrightarrow{(\qquad\qquad ,q)} (\bar{x}_3, p_2)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \quad (\qquad\qquad ,q)$$

$$(\bar{y}, p_1)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \quad (\bar{y}',q)$$

$$(\bar{z}_1, p) \xrightarrow{(\bar{r}',q)} (\bar{z}_2, p) \xrightarrow{(\bar{s}',q)} (\bar{z}_3, p)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \quad (\qquad\qquad ,q)$$

$$(\bar{r}, p) \qquad\qquad (\bar{s}, p)$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{r}',q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{s}',q)$$

$$(\bar{x}_1, p_1) \xrightarrow{(\quad ,q)} (\bar{x}_2, p_2) \xrightarrow{(\qquad\qquad ,q)} (\bar{x}_3, p_2)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \qquad (\qquad\qquad ,q)$$

$$(\bar{0}, (\bar{y}, p_1)) \qquad (\quad , (\bar{z}_1, p)) \xrightarrow{(\bar{r}',q)} (\quad , (\bar{z}_2, p)) \xrightarrow{(\bar{s}',q)} (\qquad\qquad , (\bar{z}_3, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \quad (\bar{y}',q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \quad (\qquad\qquad ,q)$$

$$(\bar{0}, (\bar{r}, p)) \qquad (\bar{0}, (\bar{s}, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{r}',q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{s}',q)$$

$$(\bar{x}_1, p_1) \xrightarrow{(\quad ,q)} (\bar{x}_2, p_2) \xrightarrow{(\qquad\qquad ,q)} (\bar{x}_3, p_2)$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \quad (\qquad\qquad ,q)$$

$$(\bar{0}, (\bar{y}, p_1))$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \quad (\bar{y}', q)$$

$$(\bar{0}, (\bar{z}_1, p)) \xrightarrow{(\bar{r}', q)} (\bar{r}', (\bar{z}_2, p)) \xrightarrow{(\bar{s}', q)} (\bar{r}' \mathbin{\overline{\oplus}} \bar{s}', (\bar{z}_3, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \quad (\bar{r}' \mathbin{\overline{\oplus}} \bar{s}' \mathbin{\overline{\oplus}} z_3', q)$$

$$(\bar{0}, (\bar{r}, p)) \qquad (\bar{0}, (\bar{s}, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{r}', q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{s}', q)$$

$$(\bar{0}, (\bar{x}_1, p_1)) \xrightarrow{(\bar{y}', q)} (\bar{y}', (\bar{x}_2, p_2)) \xrightarrow{(\overline{rsz_3}', q)} (\bar{y}' \mathbin{\overline{\oplus}} \overline{rsz_3}', (\bar{x}_3, p_2))$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1 \xrightarrow{q} p_2 \xrightarrow{q} p_2) \qquad (\bar{y}' \mathbin{\overline{\oplus}} \overline{rsz_3}' \mathbin{\overline{\oplus}} \bar{x}_3', q)$$

$$(\bar{0}, (\bar{y}, p_1))$$

$$\mathrm{wt}_{\mathcal{A}_{q,a}}(p_1) \qquad (\bar{y}', q)$$

$$(\bar{0}, (\bar{z}_1, p)) \xrightarrow{(\bar{r}', q)} (\bar{r}', (\bar{z}_2, p)) \xrightarrow{(\bar{s}', q)} (\bar{r}' \mathbin{\overline{\oplus}} \bar{s}', (\bar{z}_3, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p \xrightarrow{q} p \xrightarrow{q} p) \qquad (\bar{r}' \mathbin{\overline{\oplus}} \bar{s}' \mathbin{\overline{\oplus}} z_3', q)$$

$$(\bar{0}, (\bar{r}, p)) \qquad (\bar{0}, (\bar{s}, p))$$

$$\mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{r}', q) \qquad \mathrm{wt}_{\mathcal{A}_{q,b}}(p) \quad (\bar{s}', q)$$

$\mathcal{M}_s = (Q_s, \Sigma, \delta, F)$ s.th.

- $Q_s = \{0, \dots, \mathrm{dg(C)}\}^n \times Q$
- $(\bar{x}, q) \in F$ iff $[\![\bar{x} \oplus \gamma(q)]\!] \neq \mathbb{0}$
- $\delta((\bar{x}, q), a) = \mathcal{A}^{\bar{x}}_{q,a,s}$

$\rightarrow$ derive $\mathcal{A}^{\bar{x}}_{q,a,s}$ over Q_s from $\mathcal{A}_{q,a,s}$ by:

- extending state set to $\{0, \dots, \mathrm{dg(C)}\}^n \times (\{0, \dots, \mathrm{dg(C)}\}^n \times Q_{q,a})$
- transition $(\bar{y}_1, (\bar{x}_1, p_1)) \xrightarrow{(\bar{z},q)} (\bar{y}_2, (\bar{x}_2, p_2))$ iff:

$$((\bar{x}_1, p_1), q, (\bar{x}_2, p_2)) \in T_{q,a,s} \;\wedge\; \bar{y}_2 = \bar{y}_1 \overline{\oplus} \bar{z}\,,$$

- those states $((\bar{y}_1, (\bar{x}_1, p_1))$ are final with

$$((\bar{x}_1, p_1)) \in F_{q,a,s} \;\wedge\; \bar{x} = \bar{y}_1 \overline{\oplus} (\bar{x}_1 \oplus \nu_{q,a}(p_1))$$

Main Theorem

Theorem

Σ alphabet, $\mathbb{K}$ zero-sum free, commutative strong bimonoid.

1. *$\forall \mathcal{M} \in$ wuta over Σ, $\mathbb{K}$: $supp(r_{\mathcal{M}})$ is recognizable.*
2. *$\exists$ effective construction of support automaton $\Leftrightarrow$ ZGP decidable in $\mathbb{K}$*

Definition ZGP (Kirsten, 2009)

Given $s_i, s'_j \in \mathbb{K}$, decides if $\mathbb{0} \in s_1 \cdots s_m \cdot \langle s'_1, \ldots, s'_{m'} \rangle$

use ZGP to:

- decide if a state is final
- calculate bound for counting in states

$\rightarrow$ bimonoid operations do not need to be computable

Summary

Theorem

Σ alphabet, $\mathbb{K}$ zero-sum free, commutative strong bimonoid.

1. *$\forall \mathcal{M} \in$ wuta over Σ, $\mathbb{K}$: $supp(r_{\mathcal{M}})$ is recognizable.*
2. *$\exists$ effective construction of support automaton $\Leftrightarrow$ ZGP decidable in $\mathbb{K}$*

$\rightarrow$ extends analogous result of Kirsten for weighted automata to weighted unranked tree automata and to strong bimonoids

Future Work:

- investigate connection of weighted unranked tree automata and nested weighted automata