Semirings, Semimodules, and Weighted Automata

Uli Fahrenberg

2022

Contents

1 Foreword

These notes constitute the feeble beginnings to a book on weighted automata which I started writing in Spring 2020 and then abandoned in favor of other projects. It was motivated by some frustration about the available literature on the subject; specifically, but not exclusively, I felt that far too much emphasis has been put on formal power series, and that the definitions and results in weighted automata theory are often too far removed from the standard automata intuition.

I have myself learned the basics of weighted automata from Manfred Droste and the late Zoltán Ésik, and I continue to use Droste, Kuich and Vogler's excellent *Handbook of Weighted Automata* [DKV09] for reference. Otherwise, Pin's *Handbook of Automata Theory* [Pin21] contains a very readable chapter on weighted automata [DK21], and Sakarovitch's classic work *Elements of Automata Theory* [Sak09] (or [Sak03] in French) contains many things which I have yet to discover.

These notes are work in progress, and I would very much like to receive any criticism, suggestions for improvements, etc., at

`uli@lrde.epita.fr`

Rennes, April 2022
Uli Fahrenberg

2 Semirings

2.1 Monoids

2.1 Definition. A *monoid* $(S, \otimes, 1)$ consists of a set S, an operation $\otimes : S \times S \to S$, and an element $1 \in S$ such that

- $x \otimes (y \otimes z) = (x \otimes y) \otimes z$ for all $x, y, z \in S$ (*associativity*);
- $x \otimes 1 = 1 \otimes x = x$ for all $x \in S$ (*unit law*).

The element 1 is called the *unit* of the monoid. We will often write S instead of $(S, \otimes, 1)$ when operation and unit are clear from the context, and we will generally omit the multiplication symbol $\otimes$ from terms.

We record the easy fact that the unit of a monoid is uniquely determined.

2.2 Lemma. *Let S be a monoid and $y, z \in S$ such that $xy = x$ and $zx = x$ for all $x \in S$. Then $y = z = 1$.*

Proof: We have $y = 1 \otimes y = 1$ and $z = z \otimes 1 = 1$. □

2.3 Definition. A monoid S is *commutative* if $x \otimes y = y \otimes x$ for all $x, y \in S$.

In commutative monoids, one traditionally denotes the operation by $\oplus$ and the unit by 0.

2.4 Definition. Two monoids $(S, \otimes, 1)$ and $(\tilde{S}, \tilde{\otimes}, \tilde{1})$ are *isomorphic* if there exists a bijection $\varphi : S \to \tilde{S}$ such that $\varphi(1) = \tilde{1}$ and $\varphi(x \otimes y) = \varphi(x) \,\tilde{\otimes}\, \varphi(y)$ for all $x, y \in S$.

2.5 Definition. Let A be a set. The *free monoid* on A is $(A^*, ., \varepsilon)$, where $A^* = A^0 \cup A^1 \cup A^2 \cup \cdots$ is the set of finite sequences of elements of A, ε is the empty sequence (hence $A^0 = \{\varepsilon\}$), and the operation . is concatenation of sequences defined by $(x_1, \ldots, x_n).(y_1, \ldots, y_m) = (x_1, \ldots, x_n, y_1, \ldots, y_m)$.

Note that for any one-element set $A = \{a\}$, the free monoid $\{a\}^*$ is isomorphic to the commutative monoid $(\mathbb{N}, +, 0)$ of natural numbers.

2.2 Semirings

2.6 Definition. A *semiring* $(S, \oplus, \otimes, 0, 1)$ consists of a set S, operations $\oplus, \otimes : S \times S \to S$, and elements $0, 1 \in S$ such that

- $(S, \oplus, 0)$ is a commutative monoid;
- $(S, \otimes, 1)$ is a monoid;
- $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$ and $(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$ for all $x, y, z \in S$ (*distributivity*);
- $x \otimes 0 = 0 \otimes x = 0$ for all $x \in S$ (*annihilation law*).

We follow the standard convention to mostly omit $\otimes$ from terms and to declare that $\otimes$ binds stronger than $\oplus$. The distributivity laws may now be written as $x(y \oplus z) = xy \oplus xz$ and $(x \oplus y)z = xz \oplus yz$.

2.7 Lemma. *If a semiring S contains more than one element, then* $0 \neq 1$.

Proof: Assume that $0 = 1$ and let $x \in S$. By the unit and annihilation laws, $x = 1x = 0x = 0$. $\square$

2.8 Definition. The *center* of a semiring S is the set

$$C(S) = \{x \in S \mid \forall y \in S : xy = yx\}\,.$$

2.9 Definition. Two semirings $(S, \oplus, \otimes, 0, 1)$ and $(\tilde{S}, \tilde{\oplus}, \tilde{\otimes}, \tilde{0}, \tilde{1})$ are *isomorphic* if there exists a bijection $\varphi : S \to \tilde{S}$ such that $\varphi(0) = \tilde{0}$, $\varphi(1) = \tilde{1}$, and $\varphi(x \oplus y) = \varphi(x) \mathbin{\tilde{\oplus}} \varphi(y)$ and $\varphi(x \otimes y) = \varphi(x) \mathbin{\tilde{\otimes}} \varphi(y)$ for all $x, y \in S$.

2.3 Examples

2.3.1 The natural numbers

The semiring of natural numbers consists of the set $\mathbb{N} = \{0, 1, 2, \dots\}$ together with the usual operations of addition and multiplication; in the notation of Definition 2.6, it is $(\mathbb{N}, +, \cdot, 0, 1)$.

Any ring or field is also a semiring, so the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, and the real numbers $\mathbb{R}$ all form semirings under the usual operations of addition and multiplication.

2.3.2 One- and two-element semirings

Note that by the requirement $0, 1 \in S$, any semiring must have at least one element.

Up to isomorphism there is precisely one semiring with only one element, the *trivial* semiring $(\{0\}, \square, \square, 0, 0)$ in which $0 = 1$ and $0 \,\square\, 0 = 0$. Some authors would deny this structure the right to be called a semiring, by requiring that $0 \neq 1$ should always hold.

The *integers modulo* 2 form a semiring with the usual addition and multiplication. This is usually denoted by $(\{0, 1\}, +, \cdot, 0, 1\}$, where it is understood that 0 and 1 are the residue classes modulo 2, that is, $1 + 1 = 0$. (Note that this is, in fact, a field.)

The *boolean semiring* is $(\{\mathbf{ff}, \mathbf{tt}\}, \vee, \wedge, \mathbf{ff}, \mathbf{tt})$, where $\mathbf{ff}$ and $\mathbf{tt}$ denote the logical values of *false* and *true*, respectively, $\vee$ is logical disjunction, and $\wedge$ is logical conjunction.

2.10 Lemma. *Up to isomorphism, the integers modulo* 2 *and the boolean semiring are the only semirings with two elements.*

Proof: Let $(S, \oplus, \otimes, 0, 1)$ be a two-element semiring. By Lemma 2.7 we must have $0 \neq 1$, hence $S = \{0, 1\}$. Using the semiring axioms, we deduce the following equations:

$$\begin{array}{llll} 0 \oplus 0 = 0 & 0 \oplus 1 = 1 & 1 \oplus 0 = 1 & \\ 0 \otimes 0 = 0 & 0 \otimes 1 = 0 & 1 \otimes 0 = 0 & 1 \otimes 1 = 1 \end{array}$$

We see that the only choice we have is whether $1 \oplus 1$ should be 0 or 1. In the first case, S is isomorphic to the integers modulo 2; in the second case, S is isomorphic to the boolean semiring. $\square$

Also note that the boolean semiring is *self-dual*: it is isomorphic to the semiring $(\{\mathbf{ff}, \mathbf{tt}\}, \wedge, \vee, \mathbf{tt}, \mathbf{ff})$ where $\mathbf{ff}$ and $\mathbf{tt}$ (and disjunction and conjunction) have switched roles.

2.3.3 The tropical semirings

The name *tropical* has been given to the semirings $(\mathbb{R} \cup \{-\infty\}, \max, +, -\infty, 0)$ and $(\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$; sometimes only one of them is referred to as tropical and the other as *arctical.*[1] Hence the elements are real numbers, addition is given by maximum or minimum, and multiplication is addition. The two semirings are isomorphic by the function mapping x to $-x$.

The tropical semirings have found wide application in control theory. Sometimes only subsets of the real numbers are used, hence also the semiring $(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$, for example, is tropical.

[1] The origin of the name “tropical” appears to be an attempt to honor Brazilian mathematician Imre Simon, which makes the use of “arctical” rather doubtful. Using *sub*tropical for the min-plus and *sup*tropical for the max-plus variant might be preferable.

2.3.4 The max-min semiring

The max-min semiring is $(\mathbb{R}_{\geq 0} \cup \{\infty\}, \max, \min, 0, \infty)$, with maximum as addition and minimum as multiplication; sometimes the natural numbers $\mathbb{N}$ are used instead of the non-negative reals $\mathbb{R}_{\geq 0}$.

The restriction of the max-min semiring to $(\{0, \infty\}, \max, \min, 0, \infty)$ is isomorphic to the boolean semiring.

2.3.5 Language semirings

Let Σ be a finite set (an *alphabet*). A *language* is a subset $L \subseteq \Sigma^*$. Extending the concatenation operation in Σ^* to languages by $A.B = \{x.y \mid x \in A, y \in B\}$, the power set of Σ^* forms the *language semiring* $(2^{\Sigma^*}, \cup, ., \emptyset, \{\varepsilon\})$.

Let $\mathsf{RegL}(\Sigma) \subseteq 2^{\Sigma^*}$ be the set of *regular* languages, that is, the languages recognized by finite automata. Given that unions and concatenations of regular languages are again regular, it is easy to see that $(\mathsf{RegL}(\Sigma), \cup, ., \emptyset, \{\varepsilon\})$ again forms a semiring.

2.4 Properties

The following properties of semirings are all quite elementary, but they will be useful later on. The proofs are routine exercises.

2.11 Lemma (Subsemiring). *Let* $(S, \oplus, \otimes, 0, 1)$ *be a semiring and* $A \subseteq S$ *such that* $0, 1 \in A$, $A \oplus A \subseteq A$, *and* $A \otimes A \subseteq A$, *where these sets are given by* $A \oplus A = \{x \oplus y \mid x, y \in A\}$ *and* $A \otimes A = \{x \otimes y \mid x, y \in A\}$. *Let* $\oplus_A$ *and* $\otimes_A$ *be the restrictions of* $\oplus$ *and* $\otimes$ *to* A, *then* $(A, \oplus_A, \otimes_A, 0, 1)$ *is again a semiring.*

As an example, the center $C(S)$ of a semiring S is its largest commutative subsemiring.

2.12 Lemma (Product semiring). *Let* $(S_1, \oplus_1, \otimes_1, 0_1, 1_1)$ *and* $(S_2, \oplus_2, \otimes_2, 0_2, 1_2)$ *be semirings. Define* $S = S_1 \times S_2$ *and the operations* $\oplus$ *and* $\otimes$ *on* S *by* $(x_1, x_2) \oplus (y_1, y_2) = (x_1 \oplus_1 y_1, x_2 \oplus_2 y_2)$ *and* $(x_1, x_2) \otimes (y_1, y_2) = (x_1 \otimes_1 y_1, x_2 \otimes_2 y_2)$. *Let* $0 = (0_1, 0_2)$ *and* $1 = (1_1, 1_2)$, *then* $(S, \oplus, \otimes, 0, 1)$ *is again a semiring.*

2.13 Lemma (Function semiring). *Let* A *be a set and* $(S, \oplus, \otimes, 0, 1)$ *a semiring. Denote by* S^A *the set of functions from* A *to* S *and define operations* $\tilde{\oplus}$ *and* $\tilde{\otimes}$ *on* S^A *by* $(f \mathbin{\tilde{\oplus}} g)(x) = f(x) \oplus g(x)$ *and* $(f \mathbin{\tilde{\otimes}} g)(x) = f(x) \otimes g(x)$. *Let* $\tilde{0}, \tilde{1} \in S^A$ *be the constant functions which map every* x *to* 0 *and to* 1, *respectively, then* $(S^A, \tilde{\oplus}, \tilde{\otimes}, \tilde{0}, \tilde{1})$ *is again a semiring.*

Using the standard equivalence between subsets of a given set and functions to the boolean values $\mathbf{ff}$ and $\mathbf{tt}$, we now see that the language semiring 2^{Σ^*} is isomorphic to the semiring of functions $\Sigma^* \to \{\mathbf{ff}, \mathbf{tt}\}$.

3 Reachability in Weighted Automata

3.1 Weighted Automata

3.1 Definition. A *weighted automaton* over a semiring S is a structure (Q, I, K, T), where Q is a finite set of *states*, $I, K \subseteq Q$ are the subsets of *initial* and *accepting* states, and $T \subseteq Q \times S \times Q$ is a finite *transition relation.*

Figure 3.1 displays a simple example of a weighted automaton with six states The single initial state is marked with an incoming arrow, the single accepting state is marked with a double border.

3.2 Reachability

3.2 Definition. A (finite) *path* in a weighted automaton (Q, I, K, T) is a finite sequence $\pi = ((q_1, x_1, q'_1), \dots, (q_n, x_n, q'_n)) \in T^*$, for some $n \geq 0$, such that $q'_i = q_{i+1}$ for all $i = 1, \dots, n-1$.

That is, the transitions in a path are glued at their start and end states: the end state of the transition i is the start state of transition $i + 1$. We will often denote a path as above as

$$(q_1, x_1, \dots, x_n, q'_n) \quad \text{or} \quad q_1 \xrightarrow{x_1} \cdots \xrightarrow{x_n} q'_n \,.$$

The *empty* path contains no transitions (the case $n = 0$ above) and is denoted ε.

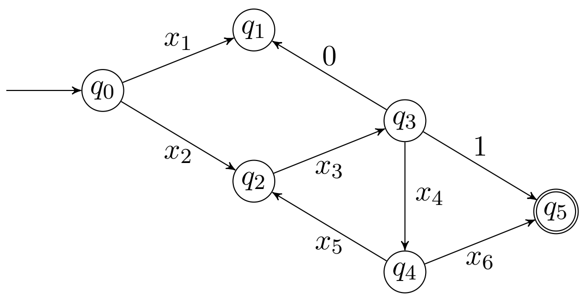

Figure 3.1: A weighted automaton

3.3 Definition. A path $(q_1, x_1, \ldots, x_n, q'_n)$ in a weighted automaton (Q, I, K, T) is *accepting* if $q_1 \in I$ and $q'_n \in K$.

The example automaton in Fig. 3.1 contains several accepting paths, for example

$$
\begin{aligned}
&q_0 \xrightarrow{x_2} q_2 \xrightarrow{x_3} q_3 \xrightarrow{1} q_5 \,, \\
&q_0 \xrightarrow{x_2} q_2 \xrightarrow{x_3} q_3 \xrightarrow{x_4} q_4 \xrightarrow{x_6} q_5 \,, \\
&q_0 \xrightarrow{x_2} q_2 \xrightarrow{x_3} q_3 \xrightarrow{x_4} q_4 \xrightarrow{x_5} q_2 \xrightarrow{x_3} q_3 \xrightarrow{1} q_5 \,, \\
&q_0 \xrightarrow{x_2} q_2 \xrightarrow{x_3} q_3 \xrightarrow{x_4} q_4 \xrightarrow{x_5} q_2 \xrightarrow{x_3} q_3 \xrightarrow{x_4} q_4 \xrightarrow{x_6} q_5 \,.
\end{aligned}
$$

3.4 Definition. The *value* of a path $\pi = (q_1, x_1, \ldots, x_n, q'_n)$ is

$$|\pi| = x_1 \cdots x_n \,.$$

Hence the transition weights along a path are *multiplied* to compute its value. We are now ready to define *reachability* in weighted automata; below $\{\!\{\cdot\}\!\}$ denotes a *multiset*.

3.5 Definition. The *reachability value* of a weighted automaton A is

$$|A| = \bigoplus \{\!\{ |\pi| \mid \pi \text{ accepting path in } A \}\!\}$$

if the sum exists, where $\bigoplus \emptyset = 0$ by convention.

That is, to compute $|A|$ we add the values of all accepting paths.

The provision that the value $|A|$ is defined *if it exists* leads us directly into the first trouble with our formalism: as the example of Fig. 3.1 shows, there may indeed exist *infinitely* many accepting paths in a weighted automaton, so that the expression defining $|A|$ becomes an infinite sum. We will need to be concerned with the existence of such infinite sums.

3.3 Semirings with infinite sums

It will be sufficient for us to introduce *countably infinite* sums in semirings. For this we need some notation regarding infinite sequences. A function $A : \mathbb{N} \to S$ into a monoid $(S, \otimes, 1)$ has *finite support* if the set $\{x \in \mathbb{N} \mid A(x) \neq 1\}$ is finite. The *shift* operator $\Delta : \mathbb{N} \to \mathbb{N}$ is given by $\Delta(n) = n+1$, and the *unit sequence* is $\mathbb{1} : \mathbb{N} \to S$ given by $\mathbb{1}(n) = 1$ (the multiplicative unit in the semiring $S^{\mathbb{N}}$).

Note that $A : \mathbb{N} \to S$ has finite support iff there is $k \geq 0$ such that the k-shift $A \circ \Delta^k = \mathbb{1}$.

3.6 Definition. A *partial (countably) infinite product* in a monoid $(S, \otimes, 1)$ is a partial function $\bigotimes : S^{\mathbb{N}} \rightharpoonup S$ which is defined for all $A : \mathbb{N} \to S$ with finite support and which satisfies that $\bigotimes \mathbb{1} = 1$;

1. if $\bigotimes A$ is defined, then so is $\bigotimes(A \circ \Delta)$, and $\bigotimes A = A(0) \otimes \bigotimes(A \circ \Delta)$; and

2. for any $A : \mathbb{N} \to S$ for which $\bigotimes A$ is defined and any sequence $0 = n_0 \le n_1 \le \cdots$ which increases without a bound, also $\bigotimes B$, where $B(k) = A(n_k) \cdots A(n_{k+1} - 1)$, is defined, and $\bigotimes A = \bigotimes B$.

The monoid S is *(countably) complete* if $\bigotimes A$ is defined for all $A : \mathbb{N} \to S$.

Here the first conditions imposes that the infinite product extends the finite one, and the second that recombining an infinite product into an infinite product of finite products does not change the value. Note that this implies associativity of infinite products.

For any monoid S, $a, c \in S$ and $B : \mathbb{N} \to S$, let $aBc : \mathbb{N} \to S$ be the sequence given by $(aBc)(n) = aB(n)c$.

3.7 Definition. A *partial (countably) infinite sum* in a semiring $(S, \oplus, \otimes, 0, 1)$ is a partial infinite product $\bigoplus$ in the monoid $(S, \oplus, 0)$ which satisfies that

- if $\bigoplus A$ is defined, then so is $\bigoplus(A \circ \sigma)$ for any permutation $\sigma : \mathbb{N} \to \mathbb{N}$, and $\bigoplus A = \bigoplus(A \circ \sigma)$; and
- if $\bigoplus B$ is defined, then so is $\bigoplus(aBc)$ for any $a, c \in S$, and $\bigoplus(aBc) = a(\bigoplus B)c$.

The semiring S is *(countably) complete* if $\bigoplus A$ is defined for all $A : \mathbb{N} \to S$.

The conditions impose that the infinite sum is commutative, and that the product distributes over infinite sums. We will generally omit mentioning countability from now. It is possible to define more general notions of infinite sums and completeness, but we will not need this here.

It will be convenient to also use infinite sums defined on countable *multisets* instead of infinite sequences. An infinite sequence $A : \mathbb{N} \to S$ corresponds to the multiset $M(A) = \{\!\{A(n) \mid n \ge 0\}\!\}$, and by commutativity, $\bigoplus A = \bigoplus A'$ for any A, A' with $M(A) = M(A')$. Hence we may define, for any multiset B which is at most countably infinite, $\bigoplus B = \bigoplus A$, where A is any infinite sequence such that $M(A) = B$.

3.8 Lemma. *For any weighted automaton A over a complete semiring S, $|A|$ is defined.*

Proof: As A is finite, the set of paths in A is at most countably infinite, hence the sum

$$|A| = \bigoplus \{\!\{ |\pi| \mid \pi \text{ accepting path in } A \}\!\}$$

exists. □

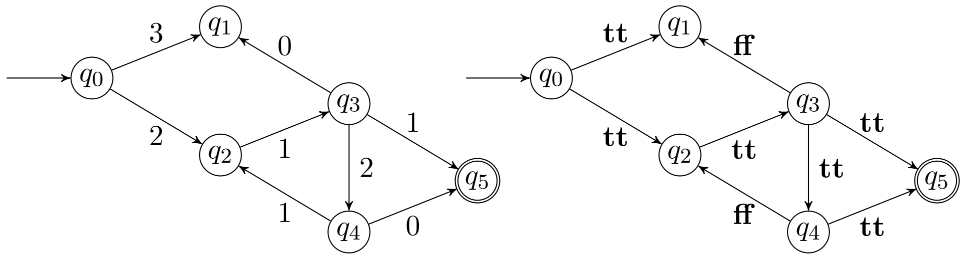

Figure 3.2: Weighted automata over $\mathbb{N}$ and $\{\mathbf{ff}, \mathbf{tt}\}$, respectively

3.4 Examples

3.4.1 Automata over the natural numbers

The left side of Fig. 3.2 displays a weighted automaton A over the semiring $\mathbb{N}$ of natural numbers. In this semiring, the value of a path is the product of all its transition weights, and the definition of the reachability value of A specializes to

$$|A| = \sum \{\!\{ |\pi| \mid \pi \text{ accepting path in } A \}\!\} \,.$$

$\mathbb{N}$ is not complete, and for our example automaton, $|A|$ is in fact undefined: Let $\pi_1, \pi_2, \dots$ be the sequence of paths given as

$$\begin{aligned}
\pi_1 &= q_0 \xrightarrow{2} q_2 \xrightarrow{1} q_3 \xrightarrow{1} q_5 \,, \\
\pi_2 &= q_0 \xrightarrow{2} q_2 \xrightarrow{1} q_3 \xrightarrow{2} q_4 \xrightarrow{1} q_2 \xrightarrow{1} q_3 \xrightarrow{1} q_5 \,, \\
\pi_3 &= q_0 \xrightarrow{2} q_2 \xrightarrow{1} q_3 \xrightarrow{2} q_4 \xrightarrow{1} q_2 \xrightarrow{1} q_3 \xrightarrow{2} q_4 \xrightarrow{1} q_2 \xrightarrow{1} q_3 \xrightarrow{1} q_5 \,,
\end{aligned}$$

etc. That is, π_i is the path which traverses the loop $q_2 \to q_3 \to q_4$ precisely $i-1$ times. Then $|\pi_i| = 2^i$, so that the infinite sum in the definition of $|A|$ tends to infinity.

3.4.2 Automata over the boolean semiring

The right side of Fig. 3.2 shows a weighted automaton A over the boolean semiring $\{\mathbf{ff}, \mathbf{tt}\}$. Here, the value of a path is $\mathbf{ff}$ iff one of its transitions has the label $\mathbf{ff}$, and $|A| = \mathbf{tt}$ iff there exists an accepting path π in A with value $|\pi| = \mathbf{tt}$.

The boolean semiring is complete, thus $|A|$ exists for any weighted automaton over $\{\mathbf{ff}, \mathbf{tt}\}$. Furthermore, if we take the convention to only take $\mathbf{tt}$-labeled transitions into account (so that a label $\mathbf{ff}$ on a transition means "disabled"), then $|A| = \mathbf{tt}$ iff an accepting state is *reachable* from an initial state through a finite path. For our example, $|A| = \mathbf{tt}$.

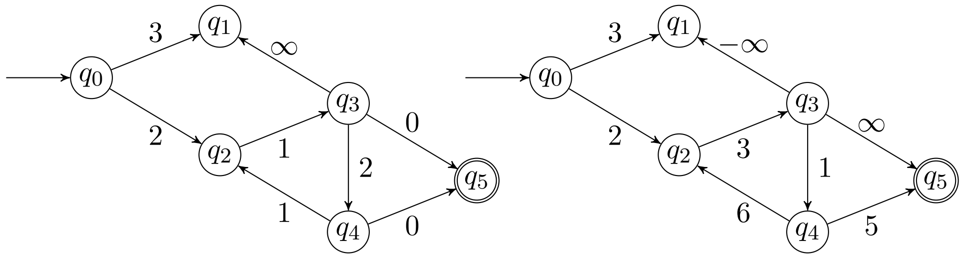

Figure 3.3: Weighted automata over the min-plus and the max-min semiring

3.4.3 Automata over the min-plus semiring

The left side of Fig. 3.3 displays a weighted automaton A over the min-plus semiring $(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$. In this semiring, the value of a path is the sum of its transition labels, and the definition of the reachability value of A specializes to

$$|A| = \inf \{\!\{ |\pi| \mid \pi \text{ accepting path in } A \}\!\} .$$

The semirings $(\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$ and $(\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$ are complete, hence $|A|$ exists for any weighted automaton over them. If we take the convention that the labels on transitions denote lengths, then $|A|$ is, in fact, the length of a *shortest path* from an initial to an accepting state in $|A|$. In our example, $|A| = 3$, as witnessed by the path

$$q_0 \xrightarrow{2} q_2 \xrightarrow{1} q_3 \xrightarrow{0} q_5 .$$

3.4.4 Automata over the max-min semiring

The right side of Fig. 3.3 shows a weighted automaton A over the semiring $(\mathbb{N} \cup \{\infty\}, \max, \min, 0, \infty)$. Here, the value of a path is the maximum of its transition weights, and then $|A|$ is the infimum of all weights of accepting paths.

The semirings $(\mathbb{R}_{\geq 0} \cup \{\infty\}, \max, \min, 0, \infty)$ and $(\mathbb{N} \cup \{\infty\}, \max, \min, 0, \infty)$ are complete, thus $|A|$ always exists in these cases. If we interpret the transition weights as *capacities*, for example of pipes, then the value of a path is the path's capacity; the minimum size of a pipe along the path. $|A|$ is then the size of the biggest pipe from an initial to a final state. In our example, $|A| = 2$.

3.4.5 Standard finite automata

As a last example, Fig. 3.4 depicts a weighted automaton over the language semiring $(2^{\Sigma^*}, \cup, ., \emptyset, \{\varepsilon\})$. The value of a path is now the concatenation of the languages along its transitions, and $|A|$ is the union of the languages of all accepting paths. That is, $|A|$ is the *language* of A interpreted as a standard

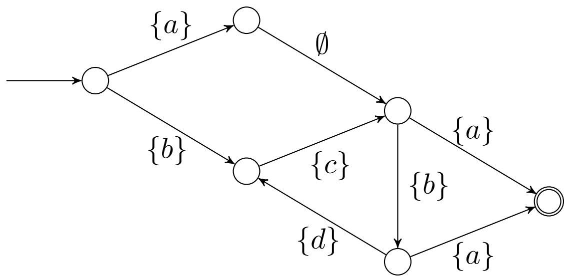

Figure 3.4: A weighted automaton over the language semiring

finite automaton over Σ. The language semiring is complete, so $|A|$ always exists. In our example,

$$|A| = \{b\}.\{cbd\}^*.\{ca, cba\}\,.$$

3.5 Variants of Weighted Automata

3.5.1 Weighted automata with input

In other literature, for example [DKV09], weighted automata (Q, I, K, T) over a semiring S are sometimes defined with an extra input alphabet Σ, and then the transitions are of type $T \subseteq Q \times \Sigma \times S \times Q$. (Hence these automata can be understood as *transducers* from Σ to S.)

Reachability values of such automata are functions $\Sigma^* \to S$ defined as follows. First, for a path $\pi = (q_1, a_1, x_1, \dots, a_n, x_n, q_n')$, define its *word* as $|\pi|_{\Sigma^*} = a_1 \cdots a_n \in \Sigma^*$ and its *value* as $|\pi|_S = x_1 \cdots x_n \in S$. Then the reachability value of A is $|A|_\Sigma : \Sigma^* \to S$, defined by letting

$$|A|_\Sigma(w) = \bigoplus \{\!\{ |\pi|_S \mid \pi \text{ accepting path in } A, |\pi|_{\Sigma^*} = w \}\!\}$$

for every $w \in \Sigma^*$.

Note that contrary to our earlier definition of $|A|$, $|A|_\Sigma$ always exists: any word $w \in \Sigma^*$ is of finite length, so the sum above is finite for every w.

We proceed to show that our simpler setting and the one above are essentially the same. First, weighted automata with input are a generalization of our setting, as follows. Let $A = (Q, I, K, T)$ be a weighted automaton (without input) over a semiring S as of Def. 3.1. Let $\Sigma = \{a\}$ be any one-letter alphabet and define a weighted automaton with input $\tilde{A} = (Q, I, K, \tilde{T})$ by

$$\tilde{T} = \{(q, a, x, q') \mid (q, x, q') \in T\}\,.$$

Now define the *length* of any path $\pi = (q_1, x_1, \dots, x_n, q_n')$ in A by $\mathsf{len}(\pi) = n$, then $|\tilde{A}|_\Sigma$ is essentially a stratification of $|A|$ by the lengths of paths:

$$|\tilde{A}|_\Sigma(a^n) = \bigoplus \{\!\{ |\pi|_S \mid \pi \text{ accepting path in } A, \mathsf{len}(\pi) = n \}\!\}$$

Hence, using the recombination property 2 of Def. 3.6, $|A|$, if it exists, is the infinite sum of the values $|\tilde{A}|_\Sigma(a^n)$:

$$|A| = \bigoplus_{n \geq 0} |\tilde{A}|_\Sigma(a^n)$$

In order to show the equivalence in the other direction, we need to introduce another product between functions $\Sigma^* \to S$, the so-called *Cauchy product*.

3.9 Definition. For a finite set Σ and a semiring $(S, \oplus, \otimes, 0, 1)$, the semiring of *formal power series* on Σ with coefficients in S is $\big(S\langle\langle\Sigma^*\rangle\rangle, \tilde{\oplus}, \tilde{\otimes}, \mathbf{0}, \mathbf{1}\big)$, where $S\langle\langle\Sigma^*\rangle\rangle = S^{\Sigma^*}$ is the set of functions $\Sigma^* \to S$ and the operations and units are given as follows.

$$(f \mathbin{\tilde{\oplus}} g)(w) = f(w) \oplus g(w) \qquad (f \tilde{\otimes} g)(w) = \bigoplus \{\!\{ f(u)g(v) \mid w = u.v \}\!\}$$

$$\mathbf{0}(w) = 0 \qquad \mathbf{1}(w) = \begin{cases} 1 & \text{if } w = \varepsilon \\ 0 & \text{if } w \neq \varepsilon \end{cases}$$

We will always denote the semiring of functions $\Sigma^* \to S$ with the standard product of Lemma 2.13 by $\Sigma^* \to S$ (or S^{Σ^*}) and the semiring of formal power series on the same set of functions, with the Cauchy product as above, by $S\langle\langle\Sigma^*\rangle\rangle$. In this context, the product on $\Sigma^* \to S$ is also called the *Hadamard product*.

Now let $A = (Q, I, K, T)$ be a weighted automaton over a semiring S and with input in Σ. Define a weighted automaton $\tilde{A} = (Q, I, K, \tilde{T})$, without input and over the semiring $S\langle\langle\Sigma^*\rangle\rangle$, by

$$\tilde{T} = \big\{ (q, \langle a, x\rangle, q') \mid (q, a, x, q') \in T \big\} .$$

Here $\langle a, x\rangle \in S\langle\langle\Sigma^*\rangle\rangle$ denotes the function given by

$$\langle a, x\rangle(w) = \begin{cases} x & \text{if } w = a\,, \\ 0 & \text{if } w \neq a\,. \end{cases}$$

3.10 Lemma. *$|\tilde{A}|$ exists, and $|A|_\Sigma = |\tilde{A}|$.*

Proof: Note that $|\tilde{A}|$ is an element of $S\langle\langle\Sigma^*\rangle\rangle$, so the types in the equation agree.

By construction of $\tilde{A}$, we have a bijection $\Phi : T \to \tilde{T}$ given by $\Phi(q, a, x, q') = (q, \langle a, x\rangle, q')$. This extends to a bijection, also denoted Φ, from paths in A to paths in $\tilde{A}$. For any path π in A, $|\Phi(\pi)| : \Sigma^* \to S$ is given by

$$|\Phi(\pi)|(w) = \begin{cases} |\pi|_S & \text{if } w = |\pi|_{\Sigma^*}\,, \\ 0 & \text{if } w \neq |\pi|_{\Sigma^*}\,. \end{cases} \tag{3.1}$$

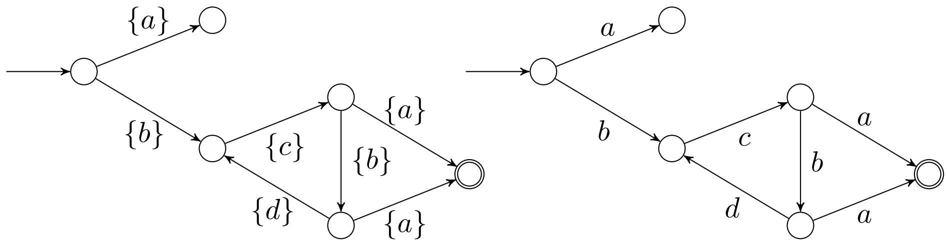

Figure 3.5: A finite automaton as a weighted automaton over 2^{Σ^*} (left) and as a weighted automaton over $\{\mathbf{ff}, \mathbf{tt}\}$ with input Σ (right)

Now let $w \in \Sigma^*$; we show that $|A|_\Sigma(w) = |\tilde{A}|(w)$. We have

$$\begin{aligned} |A|_\Sigma(w) &= \bigoplus \{\!\{ |\pi|_S \mid \pi \text{ accepting path in } A, |\pi|_{\Sigma^*} = w \}\!\} \\ &= \bigoplus \{\!\{ |\Phi(\pi)|(w) \mid \pi \text{ accepting path in } A, |\pi|_{\Sigma^*} = w \}\!\} \\ &= \bigoplus \{\!\{ |\Phi(\pi)|(w) \mid \pi \text{ accepting path in } A \}\!\} \\ &= \bigoplus \{\!\{ |\tilde{\pi}|(w) \mid \tilde{\pi} \text{ accepting path in } \tilde{A} \}\!\} \\ &= \bigoplus \{\!\{ |\tilde{\pi}| \mid \tilde{\pi} \text{ accepting path in } \tilde{A} \}\!\}(w) = |\tilde{A}|(w)\,, \end{aligned}$$

where the second and third equality are applications of (3.1) and the fourth is due to bijectivity of Φ on paths. ☐

3.11 Example. Using the correspondence between subsets of Σ^* and functions $\Sigma^* \to \{\mathbf{ff}, \mathbf{tt}\}$, we now see that standard finite automata over Σ can be understood *both* as weighted automata over the language semiring 2^{Σ^*}, as in Section 3.4.5, *and* as weighted automata over $\{\mathbf{ff}, \mathbf{tt}\}$ with input Σ. Figure 3.5 shows that the difference is purely syntactic.

4 Matrix Semirings

4.1 Weighted Automata in Matrix Form

In order to motivate the developments in the forthcoming chapters, we introduce a very convenient *matrix* representation of weighted automata.

Let $A = (Q, I, K, T)$ be a weighted automaton over a semiring S and define a matrix $\tau : Q \times Q \to S$ by letting

$$\tau(q, q') = \bigoplus \{\!\{ x \mid (q, x, q') \in T \}\!\}$$

for all $q, q' \in Q$. Note that the sum is defined because we have assumed T to be a *finite* transition relation, and recall that $\bigoplus \emptyset = 0$.

Conversely, when we are given a transition matrix $\tau : Q \times Q \to S$, we can define $T = \{(q, \tau(q, q'), q') \mid q, q' \in Q\}$. Note that the so-defined automaton has precisely one transition between any two states.

We show that using τ instead of T for computations does not affect the reachability value of A. To this end, let $A = (Q, I, K, T)$ be any weighted automaton over a semiring S, define $\tilde{A} = (Q, I, K, \tau)$ with $\tau : Q \times Q \to S$ as above, and define a *path* in $\tilde{A}$ to be any sequence $\tilde{\pi} = (q_1, \ldots, q_{n+1})$, for $n \geq 0$ and all $q_i \in Q$. The *value* of such path is given by $|\tilde{\pi}|_\sim = \tau(q_1, q_2) \cdots \tau(q_n, q_{n+1})$, and $|\tilde{A}|_\sim$ is the infinite sum over the values of all accepting paths as before.

4.1 Lemma. *$|A|$ exists iff $|\tilde{A}|_\sim$ exists, and in the affirmative case, $|A| = |\tilde{A}|_\sim$.*

Proof: Define a mapping Φ from paths in A to paths in $\tilde{A}$ by

$$\Phi(q_1, x_1, \ldots, q_n, x_n, q_n') = (q_1, \ldots, q_n, q_n') \,.$$

Then $\Phi^{-1}(\tilde{\pi})$ is finite for all paths $\tilde{\pi}$ in $\tilde{A}$, and by definition of $\tilde{T}$ and distributivity,

$$|\tilde{\pi}|_\sim = \bigoplus \{\!\{ |\pi| \mid \pi \in \Phi^{-1}(\tilde{\pi}) \}\!\}$$

for any path $\tilde{\pi}$ in $\tilde{A}$. Now Φ^{-1} induces a *partition* on the set of paths in A, and any path π is accepting in A iff $\Phi(\pi)$ is accepting in $\tilde{A}$, so that if $|A|$

exists, then

$$\begin{aligned}|A| &= \bigoplus \{\!\{ |\pi| \mid \pi \text{ accepting path in } A \}\!\} \\ &= \bigoplus \{\!\{ \bigoplus \{\!\{ |\pi| \mid \pi \in \Phi^{-1}(\tilde\pi) \}\!\} \mid \tilde\pi \text{ accepting path in } \tilde A \}\!\} \\ &= \bigoplus \{\!\{ |\tilde\pi|_\sim \mid \tilde\pi \text{ accepting path in } \tilde A \}\!\} = |\tilde A|_\sim\,,\end{aligned}$$

also showing existence of $|\tilde A|_\sim$. Similarly, reading the equations backwards, existence of $|\tilde A|_\sim$ implies existence of $|A|$.

Note that we have used the recombination property 2 of Def. 3.6 for the second equality above. □

We can hence represent the transitions of a weighted automaton (Q, I, K, T) over a semiring S as a matrix $\tau : Q \times Q \to S$; but it is convenient to go even further. The set Q of states is finite, so we can set $Q = \{1, \dots, n\}$ for some $n \geq 0$. Then τ becomes an element in the *matrix semiring* $S^{[n]\times[n]}$ which we shall introduce in the next section.

Instead of the subsets $I, K \subseteq Q$ we can specify initial and accepting *vectors* $\iota, \kappa : Q \to \{0, 1\}$ with the convention that $\iota(q) = 1$ iff $q \in I$ and $\kappa(q) = 1$ iff $q \in K$. As a generalization, it will be convenient to let $\iota, \kappa : Q \to S$ be general vectors. Putting things together, we arrive at the following definition which generalizes Def. 3.1 and uses the notation $[n] = \{1, \dots, n\}$; by convention, $[0] = \emptyset$ is the empty set.

4.2 Definition. A *weighted automaton (in matrix form)* over a semiring S is a structure (n, ι, κ, τ), where $n \geq 0$ is the (finite) number of *states*, $\iota, \kappa \in S^{[n]}$ are the *initial* and *accepting vectors*, and $\tau \in S^{[n]\times[n]}$ is the *transition matrix*.

For good measure we also update Definitions 3.2, 3.4 and 3.5. The notion of accepting path from Def. 3.3 can be omitted now because of the definitions of ι and κ.

4.3 Definition. Let $A = (n, \iota, \kappa, \tau)$ be a weighted automaton over a semiring S.

- A (finite) *path* in A is a finite sequence $\pi = (q_1, \dots, q_{m+1})$ of states $q_1, \dots, q_{m+1} \in [n]$ for some $m \geq 0$.
- For a path $\pi = (q_1, \dots, q_{m+1})$, $\mathsf{first}(\pi) = q_1$ and $\mathsf{last}(\pi) = q_{m+1}$ denote its first and last states.
- The *value* of a path $\pi = (q_1, \dots, q_{m+1})$ is

$$|\pi| = \tau_{q_1 q_2} \cdots \tau_{q_m q_{m+1}}\,.$$

- The *reachability value* of A is

$$|A| = \bigoplus \{\!\{ \iota_{\mathsf{first}(\pi)} |\pi| \kappa_{\mathsf{last}(\pi)} \mid \pi \text{ path in } A \}\!\}$$

  if the sum exists.

4.2 Partial Semirings

In order to properly introduce matrices in semirings below, we need to relax the definition of a semiring to allow for *partial* operations and multiple units.

4.4 Definition. A *partial monoid* $(S, D, \otimes, U)$ consists of a set S, a subset $D \subseteq S \times S$, an operation $\otimes : D \to S$, and a subset $U \subseteq S$ such that

- for all $x, y, z \in S$, $(y, z) \in D$ and $(x, yz) \in D$ iff $(x, y) \in D$ and $(xy, z) \in D$, and in the affirmative case, $x(yz) = (xy)z$;
- for all $x \in S$ and $y, z \in U$, $(x, y) \in D$ implies $xy = x$ and $(z, x) \in D$ implies $zx = x$.

That is, D denotes the *domain* of the composition $\otimes$: the set of pairs of elements for which their composition is defined. Associativity of $\otimes$ is required to hold whenever one side of the equation $x(yz) = (xy)z$ is defined, and elements of U act as units in all composable pairs.

4.5 Definition. A partial monoid $(S, D, \otimes, U)$ is *commutative* if it holds for all $x, y \in S$ that $(x, y) \in D$ iff $(y, x) \in D$, and in the affirmative case, $xy = yx$.

4.6 Definition. A *partial semiring* $(S, D_\oplus, D_\otimes, \oplus, \otimes, U_\oplus, U_\otimes)$ consists of a set S, subsets $D_\oplus, D_\otimes \subseteq S \times S$, operations $\oplus : D_\oplus \to S$, $\otimes : D_\otimes \to S$, and subsets $U_\oplus, U_\otimes \subseteq S$ such that

- $(S, D_\oplus, \oplus, U_\oplus)$ is a commutative partial monoid;
- $(S, D_\otimes, \otimes, U_\otimes)$ is a partial monoid;
- for all $x, y, z \in S$, $(y, z) \in D_\oplus$ and $(x, y \oplus z) \in D_\otimes$ iff $(x, y), (x, z) \in D_\otimes$ and $(xy, xz) \in D_\oplus$, and in the affirmative case, $x(y \oplus z) = xy \oplus xz$;
- for all $x, y, z \in S$, $(x, y) \in D_\oplus$ and $(x \oplus y, z) \in D_\otimes$ iff $(x, z), (y, z) \in D_\otimes$ and $(xz, yz) \in D_\oplus$, and in the affirmative case, $(x \oplus y)z = xz \oplus yz$;
- for all $x \in S$ and $y, z \in U_\oplus$, $(x, y) \in D_\otimes$ implies $xy \in U_\oplus$ and $(z, x) \in D$ implies $zx \in U_\oplus$.

Note that the annihilation law needs to take into account that the product of an element x and a zero element might be *another* zero element.

4.3 Matrix Semirings

Recall that for $n \geq 0$, $[n] = \{1, \dots, n\}$; by convention, $[0] = \emptyset$ is the empty set.

4.7 Definition. Let $(S, \oplus, \otimes, 0, 1)$ be a semiring. The *matrix semiring* over S is the partial semiring $\mathcal{M}(S) = (\bigcup_{n,m \geq 0} S^{[n]\times[m]}, D_{\tilde{\oplus}}, D_{\tilde{\otimes}}, \tilde{\oplus}, \tilde{\otimes}, U_{\tilde{\oplus}}, U_{\tilde{\otimes}})$, where the domains, operations and units are given as follows.

$$
\begin{gathered}
D_{\tilde{\oplus}} = \{(A : [n] \times [m] \to S, B : [n] \times [m] \to S) \mid n, m \geq 0\} \\
D_{\tilde{\otimes}} = \{(A : [n] \times [m] \to S, B : [m] \times [k] \to S) \mid n, m, k \geq 0\} \\
(x \mathbin{\tilde{\oplus}} y)_{ij} = x_{ij} \oplus y_{ij} \qquad (x \mathbin{\tilde{\otimes}} y)_{ij} = \bigoplus \{\!\{ x_{ik}\, y_{kj} \mid k \in [m] \}\!\} \\
U_{\tilde{\oplus}} = \{A : [n] \times [m] \to S \mid n, m \geq 0, A_{ij} = 0 \text{ for all } i, j\} \\
U_{\tilde{\otimes}} = \{A : [n] \times [n] \to S \mid n \geq 0, A_{ij} = 1 \text{ if } i = j; 0 \text{ if } i \neq j\}
\end{gathered}
$$

Implicitly the above entails that the types of the operations are $\tilde{\oplus} : S^{[n]\times[m]} \times S^{[n]\times[m]} \to S^{[n]\times[m]}$ and $\tilde{\otimes} : S^{[n]\times[m]} \times S^{[m]\times[k]} \to S^{[n]\times[k]}$. Note that the formulas for addition and multiplication are the usual ones from linear algebra.

The subsets $S^{[n]\times[m]}$ of $\mathcal{M}(S)$ for each $n, m \geq 0$ constitute commutative monoids. For $n = m$, $S^{[n]\times[n]}$ is a (total) semiring. Note how the product on $S^{[n]\times[n]}$ is a mix between the pointwise Hadamard product and the convolution-like Cauchy product. Hence $S^{[n]\times[n]}$ is the same as $S^{[n^2]}$ as sets, but not as semirings.

The subsets $S^{[1]\times[n]}$ and $S^{[n]\times[1]}$ consist of row and column *vectors* over S. As monoids, they are both isomorphic to $S^{[n]}$, the monoid of n-tuples of elements of S. We will make no notational distinction between row and column vectors.

The following is the fundamental theorem on weighted automata, showing that reachability values may be computed in the matrix semiring.

4.8 Theorem. *Let $A = (n, \iota, \kappa, \tau)$ be a weighted automaton over a semiring S. If $|A|$ exists, then*

$$|A| = \bigoplus \{\!\{ \iota \tau^m \kappa \mid m \geq 0 \}\!\} .$$

Proof: Assume that $|A|$ exists. Then

$$
\begin{aligned}
|A| &= \bigoplus \{\!\{ |\pi| \mid \pi \text{ path in } A \}\!\} \\
&= \bigoplus_{m \geq 0} \bigoplus \{\!\{ |\pi| \mid \pi \text{ path in } A, \mathsf{len}(\pi) = m \}\!\} .
\end{aligned}
$$

Let $m \geq 0$; we show that $\iota \tau^m \kappa = \bigoplus \{\!\{ |\pi| \mid \pi \text{ path in } A, \mathsf{len}(\pi) = m \}\!\}$. We have $\iota\kappa = \bigoplus_{i=1}^n \iota_i \kappa_i$,

$$\iota \tau \kappa = \bigoplus_{i=1}^{n} \bigoplus_{j=1}^{n} \iota_i \tau_{ij} \kappa_j ,$$

and in general,

$$\begin{aligned}
\iota\tau^m\kappa &= \bigoplus \{\!\{\iota_{q_1}\tau_{q_1q_2}\cdots\tau_{q_mq_{m+1}}\kappa_{q_{m+1}} \mid q_1,\dots,q_{m+1}\in\{1,\dots,n\}\}\!\} \\
&= \bigoplus \{\!\{\iota_{q_1}\tau_{q_1q_2}\cdots\tau_{q_mq_{m+1}}\kappa_{q_{m+1}} \mid \pi=(q_1,\dots,q_{m+1}) \text{ path in } A\}\!\} \\
&= \bigoplus \{\!\{|\pi| \mid \pi=(q_1,\dots,q_{m+1}) \text{ path in } A\}\!\}\,. \qquad \square
\end{aligned}$$

4.4 Semimodules and Linear Transformations

4.9 Definition. A *left semimodule* (V,μ) over a semiring $(S,\oplus_S,\otimes_S,0_S,1_S)$ consists of a commutative monoid $(V,\oplus_V,0_V)$ together with a left action $\mu : S\times V\to V$ which satisfies the following for all $x,y\in S$ and $u,v\in V$:

$$\begin{aligned}
\mu(x\oplus_S y,v) &= \mu(x,v)\oplus_V \mu(y,v) & \mu(1_S,v) &= v\\
\mu(x,u\oplus_V v) &= \mu(x,u)\oplus_V \mu(x,v) & \mu(0_S,v) &= 0_V\\
\mu(xy,v) &= \mu(x,\mu(y,v)) & \mu(x,0_V) &= 0_V
\end{aligned}$$

A *right* semimodule over S is defined analogously, with action $\mu : V\times S\to V$. If S is *commutative*, then the two notions agree.

We will usually make no notational distinction between the operations in S and in V and also write μ as simple multiplication. Using these conventions, the above properties can be written in a more familiar form:

$$\begin{aligned}
(x\oplus y)v &= xv\oplus yv & 1\otimes v &= v\\
x(u\oplus v) &= xu\oplus xv & 0\otimes v &= 0\\
(xy)v &= x(yv) & x\otimes 0 &= 0
\end{aligned}$$

4.10 Definition. Let U and V be left semimodules over a semiring S. A function $f : U\to V$ is a *left-linear transformation* if the following hold for all $u,v\in U$ and $x\in S$:

$$f(u\oplus v)=f(u)\oplus f(v) \qquad f(xv)=xf(v)$$

Right-linear transformations of right semimodules are defined analogously; for a *commutative* semiring S, the two notions agree.

4.11 Definition. A *bisemimodule* (V,μ,ν) over a semiring $(S,\oplus_S,\otimes_S,0_S,1_S)$ consists of left and right semimodules (V,μ) and (V,ν) over S such that

$$(xv)y=x(vy)$$

for all $x,y\in S$, $v\in V$.

4.12 Definition. Let $(S, \otimes, \oplus, 0, 1)$ be a semiring and $n \geq 0$. The *free bisemimodule of dimension* n over S consists of the commutative monoid $(S^{[n]}, \tilde{\oplus}, \tilde{0})$ and the left and right actions $\mu : S \times S^{[n]} \to S^{[n]}$, $\nu : S^{[n]} \times S \to S^{[n]}$ given as follows:

$$(u \,\tilde{\oplus}\, v)_i = u_i \oplus v_i \qquad \tilde{0}_i = 0 \qquad \mu(x, v)_i = x v_i \qquad \nu(v, x)_i = v_i x$$

Note that $\tilde{\oplus}$ and $\tilde{0}$ are the standard operation and unit in the monoid of functions $[n] \to S$.

We proceed to show that there is a bijective correspondence between right-linear transformations $S^{[m]} \to S^{[n]}$ and matrices in $S^{[n]\times[m]}$. For a matrix $M \in S^{[n]\times[m]}$ define $f^M : S^{[m]} \to S^{[n]}$ by $f^M(v) = Mv$, where we use the isomorphism between $S^{[m]\times[1]}$ and $S^{[m]}$.

Conversely, for a linear transformation $f : S^{[m]} \to S^{[n]}$ define the matrix $M^f \in S^{[n]\times[m]}$ by $M^f_{ij} = f(e^j)_i$, where $e^j \in S^{[m]}$, for $j \in [m]$, are the *unit vectors* given by

$$e^j_i = \begin{cases} 1 & \text{if } j = i\,, \\ 0 & \text{if } j \neq i\,. \end{cases}$$

4.13 Lemma. *The mappings $M \mapsto f^M$ and $f \mapsto M^f$ are inverse to each other and hence constitute a bijection between right-linear transformations $S^{[m]} \to S^{[n]}$ and matrices in $S^{[n]\times[m]}$.*

Proof: Let $M \in S^{[n]\times[m]}$, $f = f^M$, and $\tilde{M} = M^f$. Then for all $i \in [n]$, $j \in [m]$,

$$\tilde{M}_{ij} = f(e^j)_i = (Me^j)_i = \bigoplus \{\!\{ M_{ik} e^j_k \mid k \in [m] \}\!\} = M_{ij}\,.$$

For the other direction, we first show that $v = \bigoplus \{\!\{ e^k v_k \mid k \in [m] \}\!\}$ for all $v \in S^{[m]}$: for all $i \in [m]$,

$$\begin{aligned} \left(\bigoplus \{\!\{ e^k v_k \mid k \in [m] \}\!\} \right)_i &= \bigoplus \{\!\{ (e^k v_k)_i \mid k \in [m] \}\!\} \\ &= \bigoplus \{\!\{ e^k_i v_k \mid k \in [m] \}\!\} = v_i\,. \end{aligned}$$

Now let $f : S^{[m]} \to S^{[n]}$ be a linear transformation, let $M = M^f$, and $\tilde{f} = f^M$. Let $v \in S^{[m]}$ and $i \in [n]$, then

$$\begin{aligned} f(v)_i = f\left(\bigoplus \{\!\{ e^k v_k \mid k \in [m] \}\!\} \right)_i &= \left(\bigoplus \{\!\{ f(e^k) v_k \mid k \in [m] \}\!\} \right)_i \\ &= \bigoplus \{\!\{ f(e^k)_i v_k \mid k \in [m] \}\!\} \end{aligned}$$

and also

$$\tilde{f}(v)_i = (Mv)_i = \bigoplus \{\!\{ M_{ik} v_k \mid k \in [m] \}\!\} = \bigoplus \{\!\{ f(e^k)_i v_k \mid k \in [m] \}\!\}\,. \qquad \square$$

Note that the proof only used the right-semimodule structure of $S^{[n]}$. An analogous result holds for the equivalence between left-linear transformations and mappings of the form $v \mapsto vM$.

We have seen that left- and right-linearity are the same in *commutative* semirings. The following proposition shows that these are the only examples of transformations which are both left- and right-linear; recall that $C(S) \subseteq S$ denotes the center of S.

4.14 Proposition. *Let S be a semiring, $m, n \geq 0$, and $f : S^{[m]} \to S^{[n]}$ a mapping which is both left- and right-linear. Let L and R be the associated matrices so that $f(v) = Lv = vR$ for all $v \in S^{[m]}$, then $L_{ji} = R_{ij} \in C(S)$ for all $i \in [n], j \in [m]$.*

Proof: Setting $v = e^j$, for $j \in [m]$, the equation $Le^j = e^j R$ implies that $L_{ji} = R_{ij}$ for all $i \in [n], j \in [m]$. Now let $x \in S$ and $v = e^j x = xe^j$, then $Lv = vR$ implies that $L_{ji}x = xL_{ji}$ for all $i \in [n], j \in [m]$. □

4.5 Weighted Automata as Affine Transformations

4.15 Definition. Let U and V be left semimodules over a semiring S. A function $f : U \to V$ is an *left-affine transformation* if there exist a left-linear transformation $g : U \to V$ and a constant $v \in V$ so that $f(u) = g(u) \oplus v$ for all $u \in U$.

Right-affine transformations of right semimodules are defined analogously; for a *commutative* semiring S, the two notions agree.

Between free bisemimodules, a right-affine transformation $f : S^{[m]} \to S^{[n]}$ is thus given by a matrix $A \in S^{[n]\times[m]}$ and a vector $v \in S^{[n]}$ such that for all $u \in S^{[m]}$, $f(u) = Au \oplus v$.

4.16 Definition. Let $A = (n, \iota, \kappa, \tau)$ be a weighted automaton over a semiring S. The *left* and *right reachability vectors* of A are $|A\rangle, \langle A| \in S^{[n]}$ given by

$$|A\rangle_q = \bigoplus \left\{\!\!\left\{ \iota_{\mathsf{first}(\pi)}|\pi| \;\middle|\; \pi \text{ path in } A, \mathsf{last}(\pi) = q \right\}\!\!\right\},$$
$$\langle A|_q = \bigoplus \left\{\!\!\left\{ |\pi|\kappa_{\mathsf{last}(\pi)} \;\middle|\; \pi \text{ path in } A, \mathsf{first}(\pi) = q \right\}\!\!\right\}$$

if the involved sums exist.

4.17 Lemma. *If $|A\rangle$ exists, then $|A|$ exists and $|A| = |A\rangle\kappa$. If $\langle A|$ exists, then $|A|$ exists and $|A| = \iota\langle A|$.*

Proof: If $|A\rangle$ exists, then

$$\begin{aligned}
|A\rangle\kappa &= \bigoplus \{\!\{ |A\rangle_q \kappa_q \mid q \in [n] \}\!\} \\
&= \bigoplus \left\{\!\!\left\{ \bigoplus \left\{\!\!\left\{ \iota_{\mathsf{first}(\pi)}|\pi|\kappa_q \;\middle|\; \pi \text{ path in } A, \mathsf{last}(\pi) = q \right\}\!\!\right\} \;\middle|\; q \in [n] \right\}\!\!\right\} \\
&= \bigoplus \left\{\!\!\left\{ \iota_{\mathsf{first}(\pi)}|\pi|\kappa_{\mathsf{last}(\pi)} \;\middle|\; \pi \text{ path in } A \right\}\!\!\right\} = |A|\,.
\end{aligned}$$

Similarly, if $\langle A|$ exists, then

$$\begin{aligned}\iota\langle A| &= \bigoplus\{\!\{\iota_q\langle A|_q \mid q\in[n]\}\!\} \\ &= \bigoplus\Big\{\!\!\Big\{\bigoplus\big\{\!\!\big\{\iota_q|\pi|\kappa_{\mathsf{last}(\pi)} \bigm| \pi \text{ path in } A, \mathsf{first}(\pi)=q\big\}\!\!\big\} \bigm| q\in[n]\Big\}\!\!\Big\} \\ &= \bigoplus\big\{\!\!\big\{\iota_{\mathsf{first}(\pi)}|\pi|\kappa_{\mathsf{last}(\pi)} \bigm| \pi \text{ path in } A\big\}\!\!\big\} = |A|\,. \end{aligned}$$

□

4.18 Lemma. *If $|A\rangle$ exists, then*

$$|A\rangle = |A\rangle\tau \oplus \iota\,.$$

If $\langle A|$ exists, then

$$\langle A| = \tau\langle A| \oplus \kappa\,.$$

Proof: Let $q\in[n]$. For the first equation,

$$\begin{aligned}\big(|A\rangle\tau\oplus\iota\big)_q &= \bigoplus\big\{\!\!\big\{|A\rangle_r\tau_{rq} \bigm| r\in[n]\big\}\!\!\big\}\oplus\iota_q \\ &= \bigoplus\big\{\!\!\big\{\iota_{\mathsf{first}(\pi)}|\pi| \bigm| \pi \text{ path in } A, \mathsf{last}(\pi)=q, \mathsf{len}(\pi)\geq 1\big\}\!\!\big\} \\ &\quad \oplus\bigoplus\big\{\!\!\big\{\iota_{\mathsf{first}(\pi)}|\pi| \bigm| \pi \text{ path in } A, \mathsf{last}(\pi)=q, \mathsf{len}(\pi)= 0\big\}\!\!\big\} \\ &= |A\rangle_q\,. \end{aligned}$$

Similarly, for the second equation,

$$\begin{aligned}\big(\tau\langle A|\oplus\kappa\big)_q &= \bigoplus\big\{\!\!\big\{\tau_{qr}\langle A|_r \bigm| r\in[n]\big\}\!\!\big\}\oplus\kappa_q \\ &= \bigoplus\big\{\!\!\big\{|\pi|\kappa_{\mathsf{last}(\pi)} \bigm| \pi \text{ path in } A, \mathsf{first}(\pi)=q, \mathsf{len}(\pi)\geq 1\big\}\!\!\big\} \\ &\quad \oplus\bigoplus\big\{\!\!\big\{|\pi|\kappa_{\mathsf{last}(\pi)} \bigm| \pi \text{ path in } A, \mathsf{first}(\pi)=q, \mathsf{len}(\pi)= 0\big\}\!\!\big\} \\ &= \langle A|_q\,. \end{aligned}$$

□

We may thus view a weighted automaton $A=(n,\iota,\kappa,\tau)$ over a semiring S as a left-affine transformation $v\mapsto v\tau\oplus\iota$ or a right-affine transformation $v\mapsto \tau v\oplus\kappa$ on $S^{[n]}$, and then the reachability vectors of A are *fixed points* of these transformations. This motivates our treatment of fixed points in the next chapters.

Bibliography

[DK21] Manfred Droste and Dietrich Kuske. Weighted automata. In *Handbook of Automata Theory* [Pin21], chapter 4.

[DKV09] Manfred Droste, Werner Kuich, and Heiko Vogler, editors. *Handbook of Weighted Automata*. EATCS Monographs in Theoretical Computer Science. Springer, 2009.

[Pin21] Jean-Eric Pin. *Handbook of Automata Theory*. European Mathematical Society Publishing House, 2021.

[Sak03] Jacques Sakarovitch. *Eléments de théorie des automates*. Vuibert, 2003.

[Sak09] Jacques Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, 2009.