Automata as Linear Operators

1 Introduction

The goal of this sketch is to view view automata as homomorphisms between linear spaces.

2 Preliminaries

2.1 Notation

(descriptive text here)

2.2 Algebra

Definition 1 (Monoid). *A monoid* $(M, \cdot, \mathbf{1})$ *consists of a set* M *for which:*

(a) Monoid multiplication $\cdot \colon M \times M \to M$ *is associative.*

(b) The element $\mathbf{1} \in M$ *is the unique identity of multiplication.*

Definition 2 (Semiring). *A semiring* $(R, +, \cdot, \mathbf{0}, \mathbf{1})$ *consists of a set* R *for which:*

(a) Semiring multiplication $\cdot \colon R \times R \to R$ *distributes over semiring addition* $+ \colon R \times R \to R$.

(b) $(R, +, \mathbf{0})$ *is a commutative monoid.*

(c) $(R, \cdot, \mathbf{1})$ *is a monoid.*

(d) Multiplication by zero $\mathbf{0}$ *annihilates* R.

Definition 3 (Semimodule). *For a semiring* R*, a* R*-semimodule* $(M, +, \times, \mathbf{0})$ *is such that:*

(a) $(M, +, \mathbf{0})$ *is a commutative group.*

(b) *Scalar multiplication* $\cdot\colon R \times M \to M$ *is a semiring action on* M.

Definition 4 (Group)**.** *A group* $(G, \cdot, \mathbf{1})$ *consists of a set* G *such that:*

(a) *Group multiplication* $\cdot\colon G \times G \to G$ *is associative.*

(b) $\mathbf{1}$ *is the identity under multiplication.*

(c) *If* $a \in G$*, then it has an inverse* $a^{-1} \in G$.

Definition 5 (Field)**.** *A field* $(K, +, \cdot, \mathbf{0}, \mathbf{1})$ *is a set* K *such that:*

(a) *Field multiplication* $\cdot\colon K \times K \to K$ *distributes over field addition* $+\colon K \times K \to K$.

(b) $(K, +, \mathbf{0})$ *is a commutative group.*

(c) $(K, \cdot, \mathbf{1})$ *is a commutative group.*

(d) *Multiplication by zero* $\mathbf{0}$ *annihilates* K.

Definition 6 (Linear Space)**.** *For a field* K*, a* K*-linear space* $(V, +, \cdot, \mathbf{0})$ *over a field* K *is a set such that:*

(a) $(V, +, \mathbf{0})$ *is a commutative group.*

(b) *Scalar multiplication* $\cdot\colon \mathbb{K} \times V \to V$ *is a field action on* V.

Definition 7 (Linear Operator)**.** *For* K*-linear spaces* V *and* W*, a linear operator* $T\colon V \to W$ *is a homomorphism.*

Definition 8 (Non-Deterministic Finite Automata)**.** *A non-deterministic finite automata (NFA)* $(\Sigma, Q, \delta, S, F)$ *is a finite state machine that accepts or rejects strings. A NFA reads from a tape of finite tape of cells in which each cell has a letter from the finite alphabet* Σ*. At any point, the NFA is in some state* Q*, and upon reading the current tape cell's contents the transition function* $\delta\colon \Sigma \times Q \to 2^Q$ *will transition the NFA non-deterministically to any one of the possible next states in addition to advancing to read the next cell. The set of start states* $S \subseteq Q$ *is the set of states that the NFA may begin at, while the set of final states* $F \subseteq Q$ *is the state of states that the NFA must be in in order to accept a string after reading the last letter on the tape.*

Definition 9 (Metric Space)**.** *A metric space* (M, d) *is a space* M *equipped with a metric function* $d\colon M \times M \to \mathbb{R}_{\geq 0}$ *such that:*

(a) $d(x, y) = 0$ *if and only if* $x = y$.

(b) $d(x, y) = d(y, x)$.

(c) *For all* $x, y, z \in M$, $d(x, z) \leq d(x, y) + d(x, z)$.

Definition 10 (Normed Space)**.** *A normed space* $(V, \|\cdot\|)$ *is a* K*-linear space* V *equipped with a norm function* $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ *such that:*

(a) *For all* $x, y \in V$, $\|x + y\| \leq \|x\| + \|y\|$.

(b) *For all* $a \in K$ *and* $x \in V$, $\|ax\| = |a| \cdot \|x\|$.

(c) *For* $\mathbf{0} \in V$, $\|\mathbf{0}\| = 0$.

3 Automata as Operators

We leverage some of the algebraic sections previously discussed to construct representations relevant to us.

3.1 Algebraic Structures of Automata

For a NFA A, with $A = (\Sigma, Q, \delta, S, F)$, the goal is to view A acting as a linear operator between spaces. In particular, the approach we take is to see δ as a linear operator between spaces: this is perhaps the most obvious approach, because to begin with, δ is already the transition function.

But the questions are then: what are the appropriate linear spaces, and what are the actions of the linear operator? One initial thought is that the transition function can, in some way, be seen as a directed acyclic graph on the states Q, where each edge is weighted by the letters in the alphabet Σ that induce the transition. The representation as an adjacency matrix does appear in literature [1]. Roughly, if M is the transition matrix, then $M_{i,j} \in 2^{\Sigma}$ denotes the states that will transition state q_i to q_j.

While such a matrix representation is useful, it is not immediately obvious how such a matrix does indeed correspond to a linear operator, especially on what linear spaces. One possible interpretation is to see the linear spaces as $|Q|$-dimensional, where each dimension of the space corresponds to one member of Q. The objects of the space is then sets of strings over Σ.

Definition 11 (String Space)**.** *The string space of* Σ *is a semiring* $\left(2^{\Sigma^\star}, \cup, \cdot, \mathbf{0}, \mathbf{1}\right)$ *such that:*

(a) The semiring addition is the set union $\cup \colon 2^{\Sigma^\star} \times 2^{\Sigma^\star} \to 2^{\Sigma^\star}$.

(b) The semiring multiplication is the string concatenation $\cdot \colon 2^{\Sigma^\star} \times 2^{\Sigma^\star} \to 2^{\Sigma^\star}$ *such that:*

$$A \cdot B = \{a \cdot b \ : a \in A, b \in B\}$$

Here the string space is the power set of all strings generated by Σ through monoid multiplication (string concatenation). Defining the string space like this allows us to equip it with a semiring structure by viewing addition as set union. For convenience, we may write R instead of $2^{\Sigma^\star}$ to denote the set corresponding to the string space.

Observe that R has the structure of a 1-dimensional linear space. The big difference, however, is that the scalar elements are elements of a semiring rather than a field. Nevertheless, R is still closed under linear operations, and is therefore a linear space.

Theorem 1. *A string space R is a linear space.*

Proof. **(descriptive text here)** □

The natural extension of a 1-dimensional linear space is a n-dimensional linear space.

Definition 12 (n-String Space)**.** *For a string space R and $n \in \mathbb{Z}_+$, the n-dimensional string space R^n is then the free semimodule isomorphic to n copies of R.*

A natural representation of R^n is as a n-dimensional vector, and in this case we prefer row vectors to column vectors. We abuse notation to identify elements of R^n with their row vector representation. Furthermore, we demonstrate that this is indeed still a linear space.

Theorem 2. *A string space R^n is a linear space.*

Proof. **(descriptive text)** □

In particular, it would be nice to have a linear operator between string spaces.

Definition 13 (Linear String Space Operator)**.** *A linear string space operator is a linear operator $A\colon R^n \to R^m$.*

In particular, we are interested in a matrix representation.

Definition 14 (Matrix Representation of Linear String Space Operator)**.** *The matrix representation of a linear string space operator $A\colon R^n \to R^m$ is a matrix $A \in M_{n\times m}(R)$ that acts on row vectors of R^n by right multiplication.*

Here each entry of the matrix denotes the sets strings that are concatenated during transition. As with the n-dimensional string space R^n, we abuse notation for A to stand in for both the linear operator and its matrix representation. We now show that this is indeed a linear operator.

Theorem 3. *The matrix representation of $A\colon R^n \to R^m$ is a linear operator.*

Proof. **(descriptive text)** □

Observe that the elements for the matrix of A are drawn from R (which is just $2^{\Sigma^\star}$) rather than 2^{Σ}, which is what we would expect for an NFA. In other words, transitions in A are given by sets of potentially long strings, rather than just single alphabets. The definition provided here is intended to be slightly more general, with the NFA case of single-letter transitions being a special case. Nevertheless, given a NFA, we are now ready to describe a particular matrix representation as a linear operator between n-string spaces.

Definition 15 (Matrix Representation of δ)**.** *For a NFA $(\Sigma, Q, \delta, S, F)$ with string space R generated by Σ and $n = |Q|$, the matrix representation of δ as a linear operator is a matrix $A \in M_{n\times n}(R)$ where:*

$$A_{i,j} = \{a \,:\, ((a, q_i), B) \in \delta, q_j \in B\}$$

Of course, this is a linear string space operator simply because every entry of the transition matrix will be a set of singleton strings.

3.2 Examples

Example 1. *Consider $\Sigma = \{a, b, c\}$ and the NFA below:*

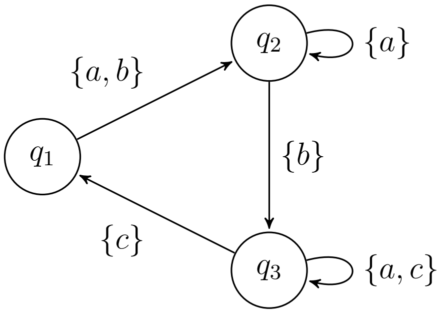

Figure 1: NFA Graph

The most important distinction between this and a typical NFA is that we implicitly assume every state to be both starting and accepting. In other words, accepting strings is very permissive, which simplifies things for now. However, to embed this into our model there are several steps.

First, we have the string space $\left(2^{\Sigma^\star}, \bigcup, \cdot, \mathbf{0}, \mathbf{1}\right)$.

Next, for the matrix A that we will construct, take $A_{i,j}$ to denote the transition from state q_i to state q_j. The matrix is then:

$$A = \begin{bmatrix} \mathbf{0} & \{a, b\} & \mathbf{0} \\ \mathbf{0} & \{a\} & \{b\} \\ \{c\} & \mathbf{0} & \{a, c\} \end{bmatrix}$$

In order to perform string concatenation towards the right, transition matrices act by right-matrix multiplication. That is, if $v \in R^3$ is the initial n-dimensional string space, then the subsequent string space is vA.

To briefly demonstrate, two transitions of the matrix A appears as follows:

$$A^2 = \begin{bmatrix} \mathbf{0} & \{a, b\} & \mathbf{0} \\ \mathbf{0} & \{a\} & \{b\} \\ \{c\} & \mathbf{0} & \{a, c\} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \{a, b\} & \mathbf{0} \\ \mathbf{0} & \{a\} & \{b\} \\ \{c\} & \mathbf{0} & \{a, c\} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \{aa, ba\} & \{ab, bb\} \\ \{bc\} & \{aa\} & \{ab, ba, bc\} \\ \{ac, cc\} & \{ca, cb\} & \{aa, ac, ca, cc\} \end{bmatrix}$$

In general, from graph theory, A^k denotes the kth consecutive transition using A, and each entry $A^k_{i,j}$ is the set of strings that will get from state q_i to q_j in k steps.

4 Automata Metrics

Having established a method for embedding NFA in linear spaces, our attention now turns to comparing them. In short, we are interested in finding how similar the strings recognized by two NFAs are.

4.1 Measures and Metrics

We leverage previous work in defining a measure, and only provide some definitions here. We borrow heavily from previous work in [2].

Definition 16 (String Space Measure)**.** *Let $(\Sigma^\star, \sigma(\Sigma^\star))$ be a measurable space defined on the strings $\Sigma^\star$. If $\lambda\colon \sigma(\Sigma^\star) \to \mathbb{R}_{\geq 0}$ is a measure, then we call it a string space measure, and the measure space $(\Sigma^\star, \sigma(\Sigma^\star), \lambda)$ a string measure space.*

The elements of a string measure space consists of strings finite, while the σ-algebra generated consists of sets of finite strings. A measure $\lambda\colon \sigma(\Sigma^\star) \to \mathbb{R}_{\geq 0}$ is also equipped, and may be defined.

We have also shown previously [2] that measures are able to define a metric.

Definition 17 (String Space Metric)**.** *Let $(\Sigma^\star, \sigma(\Sigma^\star), \lambda)$ be a string measure space. The string metric space $\left(2^{\Sigma^\star}, \zeta\right)$ is a metric space with a metric function $\zeta\colon 2^{\Sigma^\star} \times 2^{\Sigma^\star} \to \mathbb{R}_{\geq 0}$ defined as:*

$$\zeta(A, B) = \lambda(A \triangle B)$$

Theorem 4. *The string space metric function $\zeta\colon 2^{\Sigma^\star} \times 2^{\Sigma^\star} \to \mathbb{R}_{\geq 0}$ is a metric function.*

Proof. **(descriptive text)** □

Recall that although we write $2^{\Sigma^\star}$ here, they are synonymous with R. Having defined a metric on string spaces, we are interested in examining how we may compare two linear operators between string spaces. In other words, for two operators $A_1, A_2\colon R^n \to R^m$, it is possible that the two are isomorphic up to some permutation. Note that we want to consider the special case of $n = m$ first, for which theories of permutation matrix related to graph isomorphism are well-developed. After all, we are focused on NFAs here, which have nice an straightforward graph representations.

Theorem 5 (Square Operator Permutation)**.** *Let $P_n \subseteq M_{n\times n}(R)$ denote the family of n-dimensional permutation matrices. Two linear operators $A_1, A_2\colon R^n \to R^n$ are isomorphic if there is a permutation matrix $P \in P_n$ such that:*

$$A_1 = P A_2 P^T$$

Proof. **(descriptive text)** □

Permutation matrices give us the ability to talk about isomorphisms up to permutation. However they lack the ability to about how similar two operators might be. To do this, for two operators of the same dimensions, we may be interested in seeing how far the image of the operators in each dimension differ. Without permutation matrices for two linear operators A, B, we may try something like this:

$$\sup_{1\leq i,j\leq n} \zeta(A_{i,j}, B_{i,j})$$

In other words, we find entries of the matrix that yield the most distance with respect to the ζ string space metric function. With permutation matrices, we may try something like this:

$$\inf_{P \in P_n} \sup_{1 \leq i,j \leq n} \zeta\left(A_{i,j}, B_{i,j}\right)$$

This is fairly straightforward, and more or less just amounts to saying to take the best permutation of one of the operators with respect to the other, and then take the distance with respect to ζ.

However, there are still several concerns here. First, we need to generalize this to cases where A and B are not the same dimension. In addition, we are not completely sure if taking the infimum of $P \in P_n$ retains a metric structure, which means that such an operation may need to be defined with respect to a fixed permutation matrix.

We first consider how square operators of different dimensions may be compared.

Definition 18 (Linear Operator Dimension). *Let $A \colon R^n \to R^n$ be a square linear string space operator. For $m > n$, the dimension extension of A is a square linear operator $\overline{A} \colon R^m \to R^m$ whose matrix representation has form:*

$$\overline{A} = \begin{bmatrix} A & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Although the newly appended dimensions of operator action effectively have trivial action, this still acts as an embedding into a higher dimension. With this, we then attempt to define a metric between two square linear operators of different dimensions:

Definition 19 (Non-Uniform Dimension Square Linear Operator Metric). *For two square linear operators $A \colon R^n \to R^n$ and $B \colon R^m \to R^m$ where $m > n$, fix the permutation matrix $P \in P_m$, define the operator metric $Z_{p,n,m} \colon (R^n \to R^n) \times (R^m \to R^m) \to \mathbb{R}_{\geq 0}$ as:*

$$Z_{p,n,m}\left(A, B\right) = \sup_{1 \leq i,j \leq m} \zeta\left(\overline{A}_{i,j}, \left(PBP^T\right)_{i,j}\right)$$

In short, when given two square linear operators A and B where A is a smaller dimension than B, the goal of the metric function Z is to first stretch A into $\overline{A}$ to match the dimension of B, and then apply the supremum coordinate-wise metric on $\overline{A}$ and B.

Theorem 6. *The function $Z_{p,n,m} \colon (R^n \to R^n) \times (R^m \to R^m) \to \mathbb{R}_{\geq 0}$ defined above is a metric.*

Proof. Once the permutation matrix $P \in P_m$ is fixed, this reduces to applying the supremum over a finite set of metric functions for each entry of the matrix. □

The challenge, of course, is finding the correct permutation matrix P. We note that although we used the supremum here, another viable option could be general p-norm type summations, which also do induce a metric because they satisfy the triangle inequality.

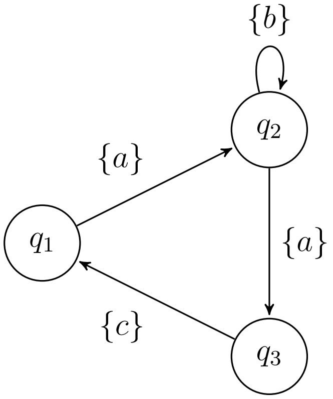

Figure 2: NFA A

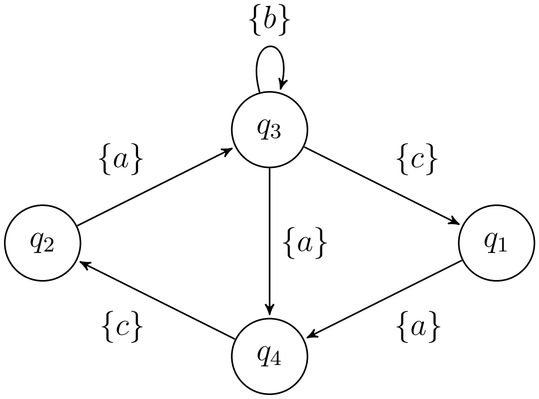

Figure 3: NFA B

4.2 Examples

Example 2. *Consider $\Sigma = \{a, b, c\}$ and two NFAs defined below:*

The respective transition matrices are then:

$$A = \begin{bmatrix} \mathbf{0} & \{a\} & \mathbf{0} \\ \mathbf{0} & \{b\} & \{a\} \\ \{c\} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \overline{A} = \begin{bmatrix} \mathbf{0} & \{a\} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \{b\} & \{a\} & \mathbf{0} \\ \{c\} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad B = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \{a\} \\ \mathbf{0} & \mathbf{0} & \{a\} & \mathbf{0} \\ \{c\} & \mathbf{0} & \{b\} & \{a\} \\ \mathbf{0} & \{c\} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The corresponding permutation matrix is:

$$P = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

Indeed, we have:

$$PBP^T = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \{a\} \\ \mathbf{0} & \mathbf{0} & \{a\} & \mathbf{0} \\ \{c\} & \mathbf{0} & \{b\} & \{a\} \\ \mathbf{0} & \{c\} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \{a\} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \{b\} & \{a\} & \{c\} \\ \{c\} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \{a\} & \mathbf{0} \end{bmatrix}$$

If we proceed with the metric ζ calculation (assuming a counting measure) at each entry. We find the most significant differences occur at row-column index pairs $(2, 4)$ and $(4, 3)$:

$$\zeta_{(2,4)}\left(\overline{A}_{2,4}, \left(PBP^T\right)_{2,4}\right) = \zeta_{(2,4)}\left(\mathbf{0}, \{c\}\right) = 1 \qquad \zeta_{(4,3)}\left(\overline{A}_{4,3}, \left(PBP^T\right)_{4,3}\right) = \zeta_{(4,3)}\left(\mathbf{0}, \{a\}\right) = 1$$

References

[1] John E Savage. *Models of computation*. Vol. 136. Addison-Wesley Reading, MA, 1998.

[2] Anton Xue. *Language Measures*. URL: `https://antonxue.github.io/sketches/lang-measures/lang-measures.pdf`.