One-dimensional topological quantum field theories with zero-dimensional defects and finite state automata

Mee Seong Im
United States Naval Academy
Annapolis, MD

Maurice Auslander Distinguished Lectures and International Conference
Woods Hole Oceanographic Institution, Quissett Campus
Woods Hole, Massachusetts, USA

Joint with Mikhail Khovanov, Columbia University, New York, NY
arXiv:2202.13398
\( \mathbb{B} = \{0, 1 : 1 + 1 = 1\} \) Boolean semiring.

\( \Sigma \): alphabet (a finite set of letters). \( \Sigma^* \): free monoid on the letters \( \Sigma \).

Example: \( \Sigma = \{a, b\} \). Words \( aaa, ababbbba, bbaaab \), etc. Empty word \( \emptyset \) is unit element.

FSA (Finite State Automaton): words in \( \Sigma \) are inputs; finitely many states \( Q \) and

transitions between the states \( \Sigma \times Q \rightarrow Q \) according to the letters read. Has initial

(starting) state \( q_{\text{in}} \) and terminating (accepting) states \( Q_t \). Example:

Language \( L = (a + b)^* b(a + b) \).

Second from last letter is \( b \). Four states.

Initial state given by the empty word \( q_{\text{in}} = x \).

Accepting states \( Q_t = \{z, y + z\} \).

The states \( z \) and \( y + z \) correspond to

the words \( (a + b)^* ba \) and \( (a + b)^* bb \), respectively.

Notation \( y + z \) comes from relation to \( \mathbb{B} \)-modules.
Regular language: one recognized by an FSA.

A word can be viewed as an interval with dots (defects) labelled by letters of the language \( L_I \). Reading a sequence along oriented interval gives a word \( \omega = a_1 a_2 \cdots a_n \).

Evaluation \( \alpha_I : \Sigma^* \rightarrow \mathbb{B} \) of decorated intervals is the same as an interval language \( L_I \): \( \omega \in L_I \iff \alpha_I(\omega) = 1 \).

Add a circular language \( L_\circ \) (for words on a circle \( \omega_1 \omega_2 \in L_\circ \iff \omega_2 \omega_1 \in L_\circ \)).

With pair \( L = (L_I, L_\circ) \), associate a \( \mathbb{B} \)-valued multiplicative evaluation \( \alpha \) of decorated 1-manifolds (defects labelled by letters in \( \Sigma \)).
\[ \alpha : \text{closed 1-dimensional manifolds} \rightarrow \mathcal{B} \text{ which satisfies} \]
\[ \alpha(M_1 \sqcup M_2) = \alpha(M_1)\alpha(M_2), \]
\[ \alpha(\emptyset_1) = 1 \text{ since } m \text{ is multiplicative}, \]
\[ \alpha(M_1) = \alpha(M_2) \text{ if } M_1 \cong M_2. \]

View interval as a “closed” 1-manifold.
\( \alpha = (\alpha_I, \alpha_\circ) \) is determined by its values \( \alpha_I(\omega) \) on decorated intervals and values \( \alpha_\circ(\omega) \) on decorated circles:
\[ \alpha_I(\omega) = 1 \iff \omega \in L_I \quad \text{and} \quad \alpha_\circ(\omega) = 1 \iff \omega \in L_\circ. \]

Universal construction starts with a (multiplicative) evaluation of closed \( n \)-dimensional objects and produces state spaces for \( (n-1) \)-dimensional objects and maps for \( n \)-cobordisms between these objects.

Use universal construction to define state spaces of oriented 0-dimensional manifolds (sign sequences \( \varepsilon = (-, -, +) \), for example).
Sign sequence: \( \varepsilon = (- - +) \). Sign sequences are objects of our category of 1-dim cobordisms with 0-dim defects in \( \Sigma \).

From \( \alpha \), one can define state spaces \( A(\varepsilon) \) for 0-dimensional objects \( \varepsilon \), by starting with a free \( \mathbb{B} \)-semimodule \( \text{Fr}(\varepsilon) \) with a basis \( \{[[M]] \partial M \cong \varepsilon\} \) given by formal symbols \( [M] \) of all 1-dimensional objects \( M \) which have \( \varepsilon \) as outer boundary (with a fixed diffeomorphism \( \partial M \cong \varepsilon \)).

A state in the state space \( A(\varepsilon) \):

![Diagram](image-url)
On $Fr(\varepsilon)$, introduce a bilinear pairing $(\ , \ )_{\varepsilon}$ given on basis elements $[M_1], [M_2]$ with $\partial M_1 \cong \varepsilon \cong \partial M_2$ by coupling $M_1, M_2$ along the boundary and evaluating the resulting closed object $M_1 \cup_\varepsilon M_2$ via $\alpha$:

$$([M_1], [M_2])_{\varepsilon} := \alpha(M_1 \cup_\varepsilon \overline{M_2}).$$

Note that $A(+) \cong A(-)^* = \text{Hom}(A(-), \mathbb{B})$ via $\omega \mapsto (\omega' \mapsto \alpha(\omega' \omega)) \in \mathbb{B}$. 

Mee Seong Im

TQFTs and automata

October 28, 2022 6 / 27
Now define the state space $A(\varepsilon)$ as the quotient of $\text{Fr}(\varepsilon)$ by an equivalence relation,

$$A(\varepsilon) := \text{Fr}(\varepsilon)/\sim,$$

where $\sum_i [M_i] \sim \sum_j [M'_j]$ if for any $M$ with $\partial M = \varepsilon$,

$$\sum_i \alpha(M_i \cup_{\varepsilon} \overline{M}) = \sum_j \alpha(M'_j \cup_{\varepsilon} \overline{M}) \in \mathbb{B} = \{0, 1 : 1 + 1 = 1\}.$$

State space $A(\varepsilon)$ is spanned by $\mathbb{B}$-linear combinations of 1-manifolds $M$ with $\partial M \cong \varepsilon$, modulo relations: two linear combinations are equal if for any way to close them up and evaluate using $\alpha$, the result is the same.

One of the relations for the language $L_I = (a + b)^* b (a + b)$:

$$\begin{bmatrix} \downarrow \\downarrow \end{bmatrix} \sim \begin{bmatrix} \downarrow \downarrow \end{bmatrix} a^n \iff \alpha \begin{bmatrix} \cdot \downarrow \omega' \end{bmatrix} = \alpha \begin{bmatrix} \cdot \downarrow \omega' \end{bmatrix} a^n$$

for any $\omega' \in \Sigma^*$. 

Mee Seong Im

TQFTs and automata

October 28, 2022
If $\omega' = ba$, then

\[
\alpha \begin{pmatrix} a \\ b \end{pmatrix} = \alpha \begin{pmatrix} a \\ b \\ a^n \end{pmatrix} = 1
\]

If $\omega' = ab$, then

\[
\alpha \begin{pmatrix} b \\ a \end{pmatrix} = \alpha \begin{pmatrix} b \\ a \\ a^n \end{pmatrix} = 0
\]

State spaces $A(-)$, $A(+)$ depend only on the interval language $L_I$, not on the circular language $L_\circ$ (spaces $A(+-)$, etc. depend on both).
An evaluation table of the language $L = (a + b)^* b(a + b)$ to compute the bilinear form on our spanning sets for $A(\cdot) + A(\cdot)$ with values in $\mathbb{B}$. The matrix is not symmetric.

Defining relations:

\[
\begin{align*}
x + y &= y \\
x + z &= z
\end{align*}
\]

\[
A(-) = \frac{\mathbb{B}x \oplus \mathbb{B}y \oplus \mathbb{B}z}{\langle x + y = y, x + z = z \rangle}
\]

Consists of 5 elements:

\{0, x, y, z, y + z\},

with $x, y, z$ irreducible.
State space of $A(+-)$ is spanned by:

A 1-manifold $M$ with $\partial M = \varepsilon' \sqcup -\varepsilon$ induces a map $A(\varepsilon) \to A(\varepsilon')$ by concatenation.

Get a functor from category of $\Sigma$-decorated oriented 1-dim cobordisms to $B$-semimodules. No subtraction in $B$-semimodules; can add only.

A $B$-semimodule $V$ is a commutative idempotented monoid under addition:

$$a + a = a \text{ for } a \in V \text{ since } 1 + 1 = 1.$$  Also $0 + a = a$,

$$a + b = b + a, \quad (a + b) + c = a + (b + c).$$

Such $V$ correspond to sup-semilattices, with join (least upper bound) $a \lor b := a + b$, and $a \leq b$ iff $a + b = b$.

0 is the minimal element, i.e., $0 \leq a$ for any $a$.

Any finite sup-semilattice is a finite lattice, with meet $a \land b := \sum_{c \leq a, b} c$ and $1 = \sum_{c \in V} c$. 
In lattices:

Join: idempotent \( a \lor a := a + a = a \) (since \( 1 + 1 = 1 \in \mathbb{B} \)), commutative, associative.

Meet: idempotent \( a \land a := \sum_{c \leq a} c = c + \ldots + a = a \), commutative, associative.

Meet and join satisfy \( a \lor (a \land b) = a, \quad a \land (a \lor b) = a \).

Why?

\[
a \lor (a \land b) = a + \left( \sum_{c \leq a,b} c \right) = a \quad \text{and} \quad a \land (a \lor b) = a \land (a + b) = \sum_{c \leq a,a+b} c = a.
\]
We mostly use $\mathbb{B}$-semimodule structure (join, not meet).

\[
\text{$\mathbb{B}$-semimodules} \iff \text{comm. idemp. monoids} \iff \text{sup-semilattices (with 0)}
\]

finite (sup)-semilattices $\iff$ finite lattices

$\mathbb{B}$-semimodules constitute a category; morphisms are semimodule homomorphisms

\[f : V \rightarrow W, \ f(0) = 0, \ f(a + b) = f(a) + f(b)\].

$\text{Hom}(V, W)$ is a $\mathbb{B}$-semimodule (category $\mathbb{B}$–mod has internal homs). But $\mathbb{B}$–mod is not a rigid category (cannot “bend” objects and morphisms).

Subcategory of finite projective $\mathbb{B}$-semimodules (finite distributive (semi)lattices) is rigid.

Categories of cobordisms in the universal construction that we build from evaluations are rigid.
Any cobordism $C$ between $\varepsilon, \varepsilon'$ induces a semimodule homomorphism $A(\varepsilon) \to A(\varepsilon')$ of concatenation with $C$:

A cobordism from $(- - + - ++)$ to $(- - + + - ++)$. 
A cobordism from $\varepsilon$ to $\varepsilon'$ can be viewed as an element in the state space $A(\varepsilon' \sqcup -\varepsilon)$, i.e., a cobordism $C : \varepsilon = (+-) \rightarrow \varepsilon' = (+-+)$ corresponds to a state in the state space $A(+---)$:

\[
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
+ \quad - \quad + \quad + \quad - \\
\end{array} \\
\begin{array}{c}
a \\
\end{array} \\
\begin{array}{c}
c \quad b \\
\end{array} \\
\begin{array}{c}
+ \quad - \\
\end{array}
\end{array}
\end{array}
\quad \Leftrightarrow \quad
\begin{array}{c}
\begin{array}{c}
\begin{array}{c}
+ \quad - \quad + \quad + \quad - \\
\end{array} \\
\begin{array}{c}
a \quad b \\
\end{array} \\
\begin{array}{c}
b \quad c \\
\end{array} \\
\end{array}
\end{array}
\]
Recall the language \( L = (a + b)^* b (a + b) \). The module \( A(-) \) is spanned by \( x, y, z \), and has relations \( x + y = y \) and \( x + z = z \). This module is not free. We’ll encounter its free cover later in the construction of minimal NFA (nondeterministic FA) for \( L \).

The semimodule consists of 5 elements: \( \{0, x, y, z, y + z\} \). The lattice corresponding to this language is:

The finite topological space associated to this example:

Lattices that come from finite topological spaces are distributive.
If a lattice contains either as a sublattice,

\[ x_i + x_j = x_i + x_k = x_j + x_k \]

\[ x_i \cap x_j = x_i \cap x_k = x_j \cap x_k \]

\[ x_i \cap x_j + x_j = x_j \]

\[ x_i \cap x_j < x_i, x_j, x_k \]

then the lattice is not distributive.

In such a case, there is no finite topological space associated to the language.
Example: for the language $L_I = \{a, a^2\}$, lattices $A(-), A(\cdot)$ are not distributive.

\[
\begin{array}{c}
\begin{array}{c}
\bullet 3 = 0 \\
\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\
1 \\
2
\end{array}
\end{array}
\]

\[
\begin{array}{cccc}
x_0 & + & x_1 & + \\
\downarrow & 0 & 1 & 1 \\
x_1 & + & 1 & 1 & 0 \\
x_2 & + & 1 & 0 & 0 \\
\end{array}
\]

\[
x_0 + x_1 = x_0 + x_2
\]

\[
x_1 + x_2 = x_1
\]

\[
x_0 + x_1 = x_0 + x_2
\]

\[
x_0 + x_1 = x_0 + x_2
\]

\[
\begin{array}{c}
\begin{array}{c}
\bullet 0 \\
\bullet 1 \\
\bullet x_1 \\
\bullet x_2 \\
\end{array}
\end{array}
\]

\[
N_5
\]
For the language $L_I = \{a, a^2\}$, how should we draw the finite topological space associated to $L_I$?

But $x_0 \neq x_0 + x_1$. So the open set containing $x_0$ cannot be the entire space.

But since $x_0 \neq x_0 + x_2$, this finite topological space does not correspond to $L_I$ as well.
**Theorem.** Languages $L_I, L_\circ$ are regular iff the state space $A(\varepsilon)$ is a finite $B$-semimodule for all sequences $\varepsilon$.

Get a $B$-valued topological theory with finite hom spaces for any such pair of languages.

To recover minimal automaton for $L_I$, consider the state space $A(-)$. It consists of $B$-linear combinations of diagrams below on the left, modulo equivalence relations coming from the pairing

$$A(-) \times A(+) \longrightarrow B.$$
How do we build the minimal deterministic FSA and nondeterministic FSA for $L_I$ from $A(\cdot)$?

Free monoid $\Sigma^*$ generated by $\Sigma$ (monoid of words) acts on $A(\cdot)$, by composing with dots at the end of the strand.

State space $A(\cdot)$ contains the subset $Q^-=\{\langle \omega \rangle \}$ of pure states. $Q^-$ is then the set of states of the minimal deterministic FSA for $L_I$. Action of $\Sigma$ comes from restriction of its action on $A(\cdot)$ (action by concatenation with dots at the top).

Initial state $q_{in}=\langle \emptyset \rangle$. A state $\langle \omega \rangle$ is accepting iff $\alpha_I(\omega)=1$. Nondeterministic FSA for $L_I$ come from coverings of $A(\cdot)$ by free $\mathbb{B}$-modules with lifted action of $\Sigma$ and unit, trace $\alpha$ maps.
\[ \tilde{m}_a \subseteq B^J \] free semimodule cover; minimal NFA for \( L_I \), where \( J = \text{irr}(A(-)) \) (irreducible if \( a \neq b + c \), where \( b \neq a \), \( c \neq a \))

\[ m_a \subseteq A(-) \] state space of 0-manifold

\[ m_a \subseteq Q_- \] minimal DFA for \( L_I \)

Every word gives a diagram in \( A(-) \).

Start with a state \( \omega \) and take images of all \( \omega \in A(-) \) under the action by \( \Sigma^* \), i.e.,

\[ \omega = \omega = \langle \omega \rangle \in A(-) \quad \Rightarrow \quad a \omega = \langle \omega \rangle = \langle \omega \rangle \Rightarrow \begin{cases} 1 & \text{if } \omega a \in L_I, \\ 0 & \text{if } \omega a \notin L_I. \end{cases} \]

\[ q_{\text{in}} = \langle \emptyset \rangle \mapsto \langle a_1 \rangle \mapsto \langle a_1 a_2 \rangle \mapsto \ldots \mapsto \langle a_1 a_2 \ldots a_n \rangle = \begin{cases} 1 & \text{if } a_1 \ldots a_n \in L_I, \\ 0 & \text{if } a_1 \ldots a_n \notin L_I. \end{cases} \]
In general, there could be more than 1 minimal NFA.

Two minimal nondeterministic automata on 3 states that accept the language $L = (a + b)^* b(a + b)$.

The second automaton has an additional $b$ arrow from $y$ to $x$ and an additional $b$ loop at $x$.

Multiple minimal NFA for $L$ appear due to several ways of lifting action of $\Sigma^*$ from $A(-)$ to $B^J$. 
Some regular languages allow decomposition of identity

\[
\alpha \left( \begin{array}{c}
\uparrow \\
\bullet x \\
\downarrow \\
y
\end{array} \right) = \sum_{i=1}^{m} \alpha \left( \begin{array}{c}
\uparrow \\
\bullet x \\
\downarrow \\
u_i
\end{array} \right) \alpha \left( \begin{array}{c}
\uparrow \\
\bullet v_i \\
\downarrow \\
y
\end{array} \right)
\]

for some set of pairs of words \((u_i, v_i), 1 \leq i \leq m\).

That is, for any \(x, y \in \Sigma^*\),

\[
\alpha_I(xy) = \sum_{i=1}^{m} \alpha_I(xu_i)\alpha_I(v_iy).
\]
Returning to our example $L = (a + b)^* b(a + b)$,

\[
\begin{align*}
\alpha_1(xy) &= \alpha_1(x)\alpha_1(bay) + \alpha_1(xb)\alpha_1(by) + \alpha_1(xba)\alpha_1(y).
\end{align*}
\]
For $L_I$ with a decomposition of the identity, there is a unique associated circular language such that the decomposition still holds:

\[
\alpha_o \left( \begin{array}{c}
\circlearrowleft \\
\omega
\end{array} \right) := \alpha_I \left( \begin{array}{c}
\circlearrowleft \\
\omega
\end{array} \right) = \sum_{i=1}^{m} \alpha_I \left( \begin{array}{c}
u_i \\
v_i
\end{array} \right) = \sum_{i=1}^{m} \alpha_I (v_i \omega u_i).
\]
This gives a $\mathbb{B}$-valued TQFT: $A(\varepsilon)$ is the tensor product of $A(\cdot)$ for the sequence of signs in $\varepsilon$.

For example, $A(\cdot\cdot\cdot) \cong A(\cdot) \otimes A(\cdot) \otimes A(\cdot)$.

This is a TQFT for oriented 1-manifolds with 0-dimensional $\Sigma$-labelled defects, valued in the Boolean semiring $\mathbb{B}$.

**Proposition.** A regular language $L$ has a decomposition of the identity if and only if $A(\cdot)$ is a projective $\mathbb{B}$-semimodule (equivalently, a distributive lattice).

A finite semimodule $P$ is projective if it’s a retract of a free semimodule:

$$ P \xrightarrow{\iota} \mathbb{B}^n \xrightarrow{p} P, \quad p\iota = \text{id}_P. $$

Note that $\iota \circ p$ is an idempotent.

Such semimodules correspond to finite topological spaces $X$, with elements of the semimodule given by open subsets $U \subset X$ and $U + V := U \cup V$. 
Thank you!