Interpreting finite state automata and regular languages via one-dimensional Boolean TQFTs and topological theories with defects

Geometric/Topological Quantum Field Theories and Cobordisms

Mee Seong Im (USNA)

March 16, 2023

TQFTs in dimensions 1, 2, 3+.

Topological Quantum Field Theories (TQFTs) play a fundamental role in modern mathematics and mathematical physics.

Fancy TQFTs appear in dimensions 3 and higher:

- Witten–Reshetikhin–Turaev 3D TQFT.
- Heegaard–Floer homology of 3-manifolds and its extension to 4-cobordisms and its many incarnations deliver an example or a 4-dimensional TQFT.
- Several link homology theories are 4-dimensional TQFTs restricted to links in $\mathbb{R}^3$ and link cobordisms in $\mathbb{R}^3 \times [0,1]$ (bigraded Khovanov and Khovanov–Rozansky theories).

TQFTs in dimension 2 are described by commutative Frobenius algebras (B, ε). Here B is a commutative algebra and $\varepsilon : B \longrightarrow \mathrm{k}$ a nondegenerate trace, with k the ground field.

Cobordisms.

Let Cob(1) be the category of one-dimensional oriented cobordisms between 0-manifolds:

> *Objects:* closed oriented 0-manifolds, represented by finite sequences of signs $+, -$. For instance, $M = (+, -, +, +)$. The empty sequence $\emptyset_0$ is the identity object.
>
> *Morphisms:* $\mathrm{Hom}(M, N) \simeq \{B : \partial B \simeq \overline{M} \sqcup N\} / \mathrm{diffeom}$, where $\overline{M}$ is M with opposite orientation.

Two cobordisms are the same in Cob(d) if they are diffeomorphic relative to their boundary. Composition is given by concatenating the cobordisms along boundaries.

The definition of a one-dimensional TQFT.

Definition

TQFT of dimension 1 is a symmetric, monoidal functor

$$Z : \mathrm{Cob}(1) \longrightarrow \mathbb{C}\mathrm{-vect}.$$

In particular, it preserves tensor products $\otimes$.

The $\otimes$ in Cob(1) is given by disjoint union of manifolds while $\otimes$ in $\mathbb{C}\mathrm{-vect}$ is given by the tensor product of vector spaces:

$$Z(M \sqcup N) \simeq Z(M) \otimes Z(N), \quad Z(\varnothing) \simeq \mathbb{C},$$

where $\mathbb{C}$ is a unit with respect to the tensor product on $\mathbb{C}$-vector spaces.

For flexibility, skip Atiyah's unitarity condition (compatibility of orientation reversal with hermitian structure on state spaces), *not requiring* $\overline{Z(M)} = Z(\overline{M})$.

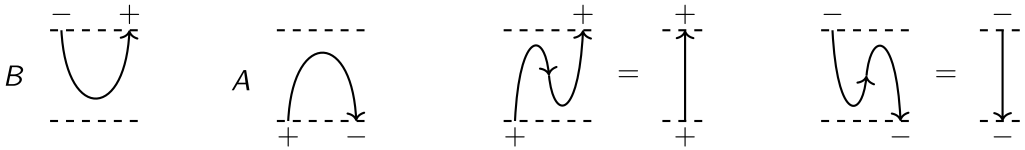
Denote category Cob(1) also by $\mathcal{C}$.

Objects are 0-dimensional manifolds with orientation: $\overset{+}{\bullet}$, $\overset{-}{\bullet}$.

$Z\left(\overset{+}{\bullet}\right) = X$ a finite dimensional $\mathbb{C}$-vector space, where $\overset{+}{\bullet}$ has positive orientation.

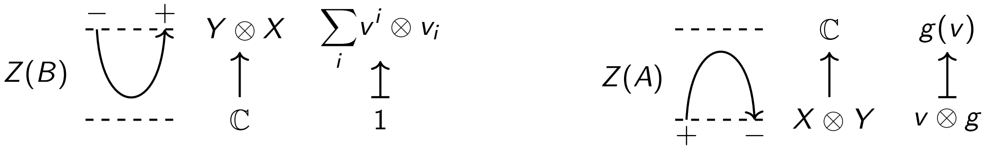
$Z\left(\overset{-}{\bullet}\right) = Y$ a finite dimensional $\mathbb{C}$-vector space, where $\overset{-}{\bullet}$ has negative orientation.

These spaces are related by maps induced by cobordisms A, B below on the left, with the isotopy relations on them shown on the right:



Cobordisms A, B induce maps $Z(A), Z(B)$, and the isotopy relations imply that X, Y are vector space duals of each other, $Y \simeq X^{\vee}$.

Given this isomorphism, maps $Z(A), Z(B)$ can be written as evaluation and coevaluation maps on these dual vector spaces:



Lines can intersect (virtual intersections), corresponding to transposition morphisms $X \otimes X \longrightarrow X \otimes X$ (or for $X \otimes Y$, etc.). The TQFT functor is symmetric monoidal.

Classification.

Conclusion: A 1-dimensional TQFT $Z : \mathcal{C} \longrightarrow \mathbb{C}\text{−vect}$ is determined by its value on a point, i.e., by a single finite-dimensional vector space X. Isomorphic vector spaces produce isomorphic TQFTs.

Our cobordisms consist of circles and arcs connecting pairs of boundary points.

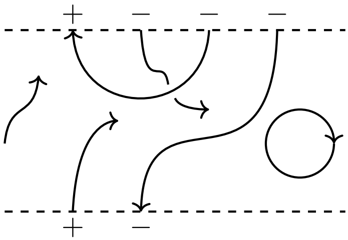
A circle, which is the only closed connected 1-manifold, evaluates to $\dim(X) \in \mathbb{N} \subset \mathbb{C}$.

Let us *enhance* this simple setup by adding:

inner boundary points and *0-dimensional defects (dots) with labels*.

I. Allow a cobordism to have components that may end in the middle, with *floating* boundary points. We separate boundary points into *outer*, that is, at the top or bottom boundary 0-manifold of the cobordism, and *inner* or *floating*.

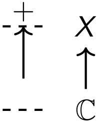
An interval component in a cobordism may have 0, 1, or 2 *floating* boundary points.



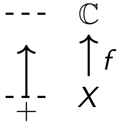
Get a category $\mathcal{C}_{\text{in}}$ of such cobordisms with inner endpoints and consider TQFT functors $\mathcal{C}_{\text{in}} \longrightarrow \mathbb{C}-\text{vect}$.

Half-intervals.

Half-interval has one *outer* and one *floating* endpoint. Depending on whether it points *in* or *out* of the outer endpoint, it defines a *vector* or a *covector* in X.



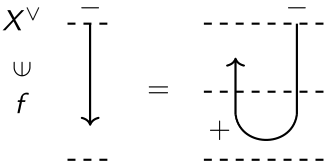
$v \in X$ vector



$f : X \to \mathbb{C}$ linear map

A TQFT for $\mathcal{C}_{\text{in}}$ assigns a vector space X to the $+$ endpoint and a vector v to a half-interval that enters a $+$ point. To a half-interval that exits from a $+$ point, we assign a covector f.

Half-intervals that enter or exit $-$ endpoints are obtained from those for $+$ via duality:



$$\sum_i f(v_i)v^i = f \in X^\vee$$

$$\uparrow$$

$$\sum_i v_i \otimes v^i$$

$$\uparrow$$

$$1$$

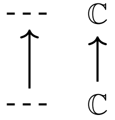
because

$$\sum_i f(v_i)v^i(v_j) = \sum_i \delta_{ij} f(v_i) = f(v_j)$$

Floating intervals.

Floating interval has two inner endpoints and can be obtained as the composition of two half-intervals. It evaluates to $f(v)$.

floating interval



$f(v) \in \mathbb{C}$

Thus, to our TQFT, there are now assigned *two* numbers: $\dim(X) \in \mathbb{N}$ and $\lambda = f(v) \in \mathbb{C}$ (or in the ground field). These are the invariants of a *circle* and of a *floating interval* (two possible connected floating components, up to diffeomorphism).

Such TQFTs are still easy to classify:

If $\lambda \neq 0$, then up to isomorphism such a TQFT is *unique*, with $v = (1, 0, \ldots, 0)^T$ and $f = (\lambda, 0, \ldots, 0)$.

If $\lambda = 0$, there are *four* isomorphism classes:

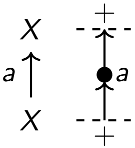
(1) $v = 0, f = 0$, (2) $v = 0, f \neq 0$, (3) $v \neq 0, f = 0$, (4) $v \neq 0, f \neq 0$.

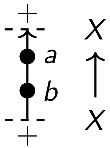
For case (4), need $\dim(X) \geqslant 2$ and can take $v = (1, 0, \ldots, 0)^T$, $f = (0, 1, 0, \ldots, 0)$.

To add more parameters, we introduce 0-dimensional defects.

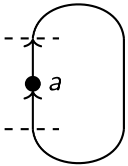
II. Place *dots (0-dimensional defects)* on a cobordism and *label* them by elements of a set $\Sigma = \{a, b, \ldots\}$ of *letters*. Dots can slide along a cobordism but cannot change order (on an interval) or cyclic order (on a circle).

Get a symmetric monoidal category $\mathcal{C}_\Sigma$ of *Σ-decorated* oriented one-cobordisms with *inner endpoints*. A TQFT functor $\mathcal{C}_\Sigma \longrightarrow \mathbb{C}\mathrm{-vect}$ associates a *linear operator* $X \xrightarrow{a} X$ to a vertical upward line with a defect labelled a. Concatenation of defects corresponds to composition of linear operators. A circle with a dot (or a product of dots) is taken as the trace of the corresponding operator(s) by the TQFT functor.



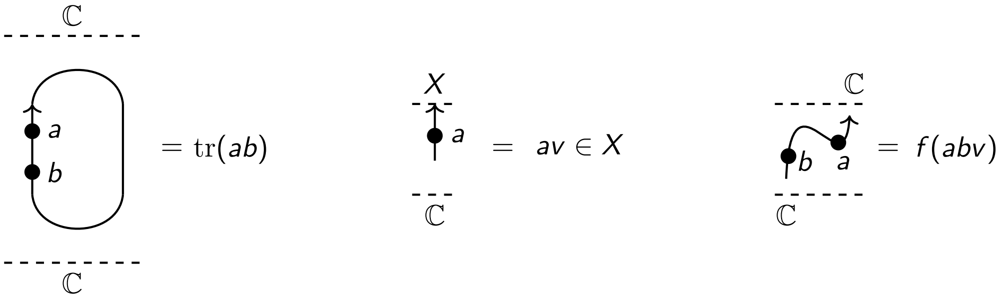


$a, b : X \to X$ linear operators
(no relations)



$\mathrm{tr}(a)$

Combine inner boundary points and defects.



Left: circle with a sequence of defects computes the trace of the corresponding product of operators.

Center: a defect near the floating endpoint applies the operator to the vector associated to the endpoint for "in" oriented endpoints.

Right: for "out" endpoint, the operator acts on a functional f.

Summary.

We've added two refinements to 1-cobordisms resulting in a chain of categories and (inclusion) functors

$$\mathrm{Cob}(1) = \mathcal{C} \subset \mathcal{C}_{\mathrm{in}} \subset \mathcal{C}_{\Sigma}.$$

A TQFT for the category $\mathcal{C}_{\Sigma}$ is described by a *f.d. vector space* X together with a *vector* $v \in X$, a *covector* $f \in X^{\vee}$, and *linear operators* $a : X \longrightarrow X$ for each $a \in \Sigma$. Classification problem for such TQFTs then reduces to classifying such data on $\mathbb{C}^n$ up to conjugation by $\mathrm{GL}(n)$, a *wild problem* (no classification available) once $|\Sigma| \geqslant 2$.

Studying 1D TQFTs with defects is *essentially linear algebra*: we get operators (for labelled dots), vectors and covectors (for inner endpoints).

Another possible extension: Labeling intervals between dots by different colors allows to introduce multiple vector spaces X_i, one for each color i, and linear maps $X_i \longrightarrow X_j$ for dots separating these intervals.

Extending from $\mathbb{C}$ to a commutative ring R.

One can, more generally, replace $\mathbb{C}$ by a *commutative ring* R and look for TQFTs $Z : \mathrm{Cob}(1) \longrightarrow R\text{–mod}$, that is, valued in the tensor category of R-modules.

A vector space X assigned to $+$ is then replaced by a *finite rank projective* R-module $V := Z(+)$. These conditions on the module V are needed to have duality maps for cap and cup cobordisms (note that $Z(-) \cong V^* = \mathrm{Hom}_R(V, R)$)

$$V^* \otimes_R V \longrightarrow R, \qquad R \longrightarrow V \otimes_R V^*$$

that satisfy the isotopy relations on slide 5. As before, we enrich this setup by adding inner endpoints and Σ-valued labels to cobordism for the category $\mathcal{C}_\Sigma$.

A tensor functor $Z : \mathcal{C}_\Sigma \longrightarrow R\text{–}\mathrm{mod}$ is then determined by a vector $v \in V$, a covector $f : V \longrightarrow R$, $f \in V^*$, and endomorphisms $a : V \longrightarrow V$ for each letter $a \in \Sigma$.

Ring R must be commutative. Decorated floating intervals and circles evaluate to elements of R and they can freely float past each other. Hence, their evaluations in R must commute, and a restriction to a commutative ring R is natural.

Replacing R by the Boolean semiring $\mathbb{B}$.

One can go further and replace commutative ring R by a commutative *semiring*. A semiring has multiplication and addition but no subtraction, in general.

It turns out that replacing $\mathbb{C}$ by a commutative semiring (for example, Boolean semiring $\mathbb{B}$) adds a twist and a different kind of complexity to the theory. As we'll see now, this replacement relates 1D TQFTs with defects and inner endpoints to regular languages and automata. Here,

$\mathbb{B} = \{0, 1 : 1 + 1 = 1\}$ is the Boolean semiring, replacing the ground field $\mathbb{C}$ of a TQFT.

We now consider this nonlinear case and its connection to formal languages and automata.

Finite State Automata (FSA).

Finite State Automata (FSA) are a basic structure in computer science. They are *memoryless* machines on *finitely many states* that, given a word ω, decide whether ω belongs to a particular *regular language* L, that is, a language recognized by a *regular expression*.

Setup. A finite set Σ is called an *alphabet* (consists of a finite set of letters).

Σ^* is the free monoid on the letters in Σ. Empty word $\emptyset$ is the unit element.

Example: a two-letter alphabet $\Sigma = \{a, b\}$. Words aaa, $ababbb$, $bbaaab$, etc. are in Σ^*.

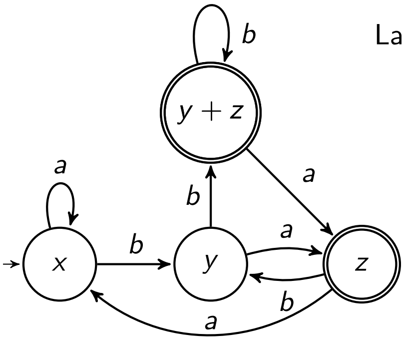
Example of a regular language: $L = (a+b)^* b(a+b)$. Here $a+b$ means *either a or b*. *Star* means *any number of times (perhaps none)*. This regular language L consists of words where 2nd from the last letter is b.

Another example: $L = b^2(aa^* + b^2)^*$. Language of words that start with b and even number of b's appears in each batch between a's. For example $b^2a^3b^4ab^2a^2b^6a \in L$.

Last example: Language $L = \{a^n b^n\}_{n \geqslant 0}$ is not regular (need to remember n when half-way across the word).

Finite State Automata (FSA).

FSA (Finite State Automaton): words in Σ are inputs; consists of finitely many states Q and transitions between the states given by a function $\delta : \Sigma \times Q \to Q$ according to the letters read. Has initial (starting) state q_{in} and terminating (accepting) states Q_t.
Example:



Language $L = (a+b)^* b(a+b)$ from earlier.

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 $y+z$ are reached by words $(a+b)^* ba$ and $(a+b)^* bb$, respectively.

Notation $y+z$ comes from relation to $\mathbb{B}$-modules.

Nondeterministic finite state automata.

Regular language (equivalent definition): one recognized by an FSA.

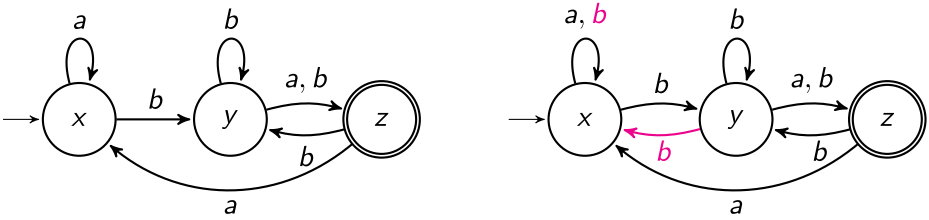
The above are called *deterministic* FSA since for each letter $a \in \Sigma$ and each state $q \in Q$ there is at most one arrow labelled a out of q. So from each state, there is at most one path ω for each word $\omega = a_1 \cdots a_n$.

A *nondeterministic* FSA $\widetilde{Q}$ has a transition function $\Sigma \times Q \longrightarrow P(Q)$. To state q and letter a, there is associated a subset of Q – all states to which one can go from q if the letter a is next in the word ω.

For convenience, we also allow more than one initial state in an NFA (nondeterministic FA). A word ω is in the language L associated to automaton $\widetilde{Q}$ if there exists a path ω in the automaton that starts in some initial state and ends in some accepting state.

Nondeterministic FA are more efficient than deterministic FA but describe the same set of regular languages. In a NFA you need to make an effort to decide whether $\omega \in L$, in contrast with a DFA where the initial state is unique and a path ω is unique (or does not exist).

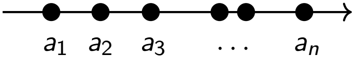
Examples of minimal nondeterministic automata on 3 states that accept our language $L = (a+b)^* b(a+b)$, with the difference in transition functions shown in red.



Minimal DFA for a regular language L is unique. Minimal NFA for L are not unique, in general. In the above examples, Q_{in} consists of a single state, but multiple initial states are allowed in NFA.

Let us interpret automata and regular languages via Boolean-valued TQFTs in one-dimension with inner endpoints and defects:

(A) A word can be viewed as an interval with dots (defects) labelled by letters in Σ. Reading a sequence of defect labels along an oriented interval gives a word.



$$\omega = a_1 a_2 \cdots a_n$$

Fix a nondeterministic FA $\widetilde{Q}$. It recognizes a regular language L.

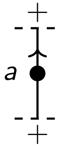
We want to evaluate a word ω (an interval with a word ω written on it) to 1 if it is in the regular language L and to 0 if it is not in L. Here $0, 1 \in \mathbb{B}$, the Boolean semiring.

$\mathbb{B} = \{0, 1 | 1 + 1 = 1\}$. No subtraction available.

Take an automaton $\widetilde{Q}$ and form $\mathbb{B}Q$, the free Boolean semimodule with a basis Q of states. Elements of $\mathbb{B}Q$ are Boolean linear combinations, that is, finite subsets of Q. Sum $q_1 + \ldots + q_m$ corresponds to the subset $\{q_1, \ldots, q_m\} \subset Q$. Note that Q is a finite set.

To a letter $a \in \Sigma$ assign a map of semimodules $\mathbb{B}Q \xrightarrow{a} \mathbb{B}Q$ taking q to the sum of states to which there is a-arrow from q:

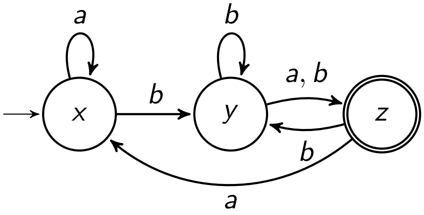
$$a(q) = \sum_{q \xrightarrow{a} q'} q'.$$



This is the map we assign to
an upward-oriented vertical interval
with a defect labelled $a \in \Sigma$.

Example.

For one of our minimal NFA for $L = (a+b)^* b(a+b)$, Boolean-valued matrices of a, b in the basis $\{x, y, z\}$ of states are



$$a = \begin{array}{c} \\ x \\ y \\ z \end{array}\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\end{array} \qquad b = \begin{array}{c} \\ x \\ y \\ z \end{array}\begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}\end{array}.$$

Cup and cap maps in the TQFT for $\widetilde{Q}$.

To a $-$ endpoint, associate the dual free $\mathbb{B}$-module $\mathbb{B}Q^*$. There is a perfect pairing

$$\mathbb{B}Q^* \times \mathbb{B}Q \longrightarrow \mathbb{B}, \quad q_1^*(q_2) = \delta_{q_1,q_2}$$

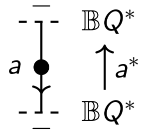
and the coevaluation map

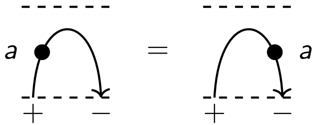
$$\mathbb{B} \longrightarrow \mathbb{B}Q \otimes \mathbb{B}Q^*, \; 1 \longmapsto \sum_{q \in Q} q \otimes q^*.$$

These maps are represented by the cup and cap diagrams in our Boolean TQFT for the automaton $\widetilde{Q}$. Isotopy relations hold.

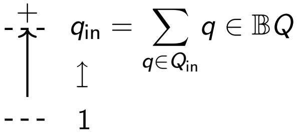
Dotted down arrows and half-intervals.

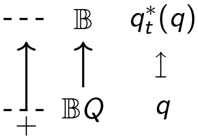
To a downward arrow labelled $a \in \Sigma$ associate the dual map $\mathbb{B}Q^* \xrightarrow{a^*} \mathbb{B}Q^*$ given by the transposed matrix of a. Isotopy relations on dots hold.





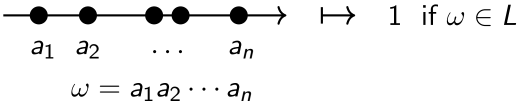
To half-intervals, assign elements $q_{\text{in}} = \sum_{q \in Q_{\text{in}}} q \in \mathbb{B}Q$ and $q_t^* = \sum_{q \in Q_t} q^* \in \mathbb{B}Q^*$, the sum of initial states and the sum of delta functions over accepting states.

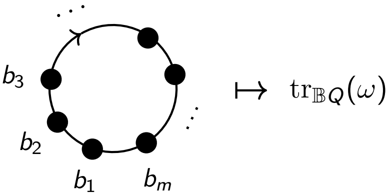




Summary.

An automaton $\widetilde{Q}$ gives a Boolean-valued TQFT that to a $+$ point assigns the free $\mathbb{B}$-module $\mathbb{B}Q$. A floating interval $I(\omega)$ with a word ω on it evaluates to 1 iff $\omega \in L$, the language recognized by the automaton. Otherwise it evaluates to 0.





A circle with a circular word $\omega = b_1 b_2 \cdots b_m$ on it evaluates to 1 iff for some state q there is a path ω that starts and ends at q. This can be written as $\mathrm{tr}_{\mathbb{B}Q}(\omega) = 1$.

Get a circular language $L_\circ = \{\omega | \mathrm{tr}_{\mathbb{B}Q}(\omega) = 1\}$. A language L is *circular* iff $\omega_1\omega_2 \in L \Leftrightarrow \omega_2\omega_1 \in L$. Necessarily, $L_\circ$ associated to $\widetilde{Q}$ is regular.

Theorem (P.Gustafson, M.S.Im, R.Kaldawy, M.Khovanov, Z.Lihn, arXiv:2301.00700)

A nondeterministic automaton $\widetilde{Q}$ on alphabet Σ defines a Boolean one-dimensional TQFT

$$\mathcal{F}_{\widetilde{Q}} \ : \ \mathcal{C}_{\Sigma} \longrightarrow \mathbb{B}\mathrm{-mod}$$

with Σ-defects and inner endpoints. Regular language L recognized by $\widetilde{Q}$ corresponds to floating intervals $I(\omega)$ that evaluate to 1. Circular language $L_{\circ}$ (the trace language of $\widetilde{Q}$) describes words placed on circles that $\mathcal{F}_{\widetilde{Q}}$ evaluates to 1.

In particular, to $\widetilde{Q}$ there is assigned a pair of languages $(L, L_{\circ})$, with the 2nd language circular.

Furthermore, there is a bijection between nondeterministic automata $\widetilde{Q}$ as above and Boolean one-dimensional TQFTs $\mathcal{F}$ for $\mathcal{C}_{\Sigma}$ such that $\mathcal{F}(+)$ is a *free* $\mathbb{B}$-module. States of an automaton are elements of the unique basis of $\mathcal{F}(+)$.

Boolean TQFTs: prospects.

TQFTs over a field have seen a phenomenal development over the past several decades, since the pioneering work of Witten, Atiyah, Donaldson, Floer, Jones, Reshetikhin–Turaev, Turaev–Viro and many others in late 80s and early 90s. Mathematical structure of these TQFTs is remarkably rich in dimensions 3, 4 and higher, with TQFTs in dimension two described by commutative Frobenius algebras.

Boolean TQFTs are a novelty. Our joint paper [1] shows that already in the *toy dimension one* and allowing defects on one-manifolds Boolean TQFTs interpret canonical structures in computer science: *regular languages and nondeterministic finite state automata*.

Nothing is known about Boolean TQFTs in dimensions two and higher. They are worth investigating for possible connections to *higher-dimensional and cellular automata*, *polycategory theory* and *topoi* (boolean semimodules and semilattices relate to topoi).

[1] Paul Gustafson, Mee Seong Im, Remy Kaldawy, Mikhail Khovanov, Zachary Lihn, *Automata and one-dimensional TQFTs with defects*, arXiv:2301.00700.

Evaluations on floating 1-manifolds.

A pair of regular languages $(L_I, L_\circ)$, with the second language circular, does not always come from a Boolean TQFT. However, it does come from a Boolean *topological theory*, which is a weak (lax) form of a TQFT.

Assume that a Boolean-valued evaluation α is given, determined by the pair of languages $(L_{\mathsf{I}}, L_\circ)$ as above. The evaluation α is defined on floating (closed) diagrams, which are floating intervals and circles with defects.

A *floating interval* with defects $a_1 \cdots a_n$ evaluates to $\alpha_{\mathsf{I}}(a_1 \cdots a_n) \in \mathbb{B}$. Evaluation α_{I} is determined by the language L_{I}, with $\omega \in L_{\mathsf{I}} \Leftrightarrow \alpha_{\mathsf{I}}(\omega) = 1$.

A *circle* with defects $b_1 \cdots b_m$ evaluates to $\alpha_\circ(b_1 \cdots b_m) \in \mathbb{B}$. Evaluation $\alpha_\circ$ is determined by the language $L_\circ$, with $\omega \in L_\circ \Leftrightarrow \alpha_\circ(\omega) = 1$.

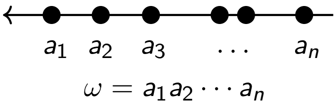
$\Rightarrow \alpha_{\mathsf{I}}, \alpha_\circ$ are functions from words (resp. circular words) to $\mathbb{B}$.

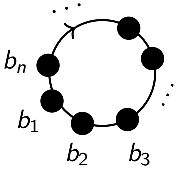
Evaluations on floating 1-manifolds.

Σ^* is the monoid of all words in the alphabet Σ. Then

$$\alpha_{\mathsf{I}} : \Sigma^* \longrightarrow \mathbb{B}, \quad \alpha_{\circ} : \Sigma^*/_{\sim} \longrightarrow \mathbb{B}$$

are the two functions above, where $\sim$ is the equivalence relation on words: $\omega_1\omega_2 \sim \omega_2\omega_1$ for words ω_1, ω_2. They are determined by languages L_{I} and $L_{\circ}$, respectively, and tell us how to evaluate decorated intervals and circles to elements of $\mathbb{B}$ (defects labelled by letters in Σ).





Universal construction over $\mathbb{C}$ vs over Boolean semiring $\mathbb{B}$.

Given such a pair $\alpha = (\alpha_{\mathsf{I}}, \alpha_{\circ})$, we will build a generalized 1D topological theory (this is weaker than a TQFT).

First, extend evaluation α to unions of decorated circles and floating intervals via multiplicativity condition.

In a universal construction of topological theories, one starts with a multiplicative evaluation of closed objects (such as closed d-manifolds) and builds a vector space for each $(d-1)$-manifold N via a linear combination of d-manifolds M with boundary N, $\partial M \cong N$. A linear combination $\sum_i \lambda_i M_i = 0$ with each $\partial M_i \cong N$ if for any M with $\partial M \cong N$, the evaluation

$$\sum_i \lambda_i \alpha(\overline{M} \cup_N M_i) = 0.$$

Add defects to manifolds $\Rightarrow$ one-dimensional $(d = 1)$ case becomes nontrivial.

Our Boolean case:

$$\begin{array}{rcll} \alpha \ : \ \text{closed 1-dimensional manifolds} & \longrightarrow & \mathbb{B} & \text{which satisfies} \\ \alpha(M_1 \sqcup M_2) & = & \alpha(M_1)\alpha(M_2), & \\ \alpha(\emptyset_1) & = & 1 \quad \text{since } m \text{ is multiplicative}, & \\ \alpha(M_1) & = & \alpha(M_2) \ \text{ if } M_1 \cong M_2. & \end{array}$$

View interval as a "closed" 1-manifold. $\alpha = (\alpha_{\mathsf{I}}, \alpha_{\circ})$ is determined by its values $\alpha_{\mathsf{I}}(\omega)$ on decorated floating intervals and values $\alpha_{\circ}(\omega)$ on decorated circles:

$$\alpha_{\mathsf{I}}(\omega) = 1 \Leftrightarrow \omega \in L_{\mathsf{I}} \quad \text{and} \quad \alpha_{\circ}(\omega) = 1 \Leftrightarrow \omega \in L_{\circ}.$$

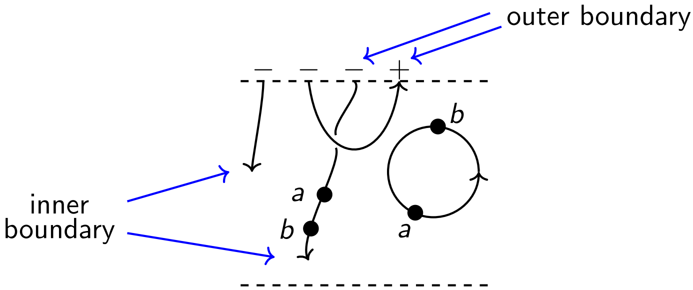
Universal construction starts with a (multiplicative) evaluation of closed n-dimensional objects and produces state spaces for $(n-1)$-dimensional objects and maps for n-cobordisms between these objects.

For us, $n = 1$. We now 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 $\mathcal{C}_\Sigma$ of 1-dim cobordisms with 0-dim defects in Σ.

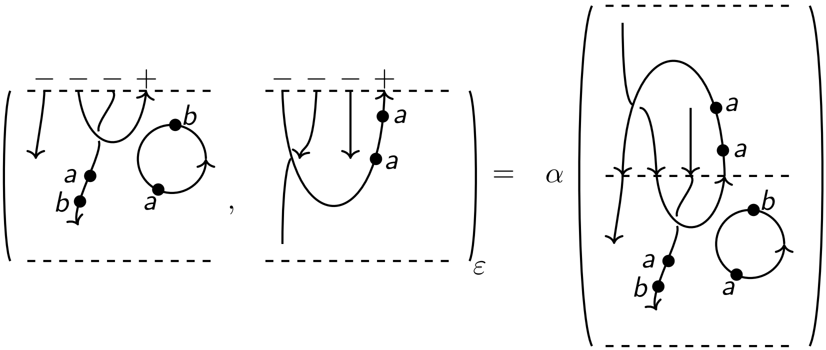
From α, one can define state spaces $A(\varepsilon)$ for 0-dimensional objects ε, by starting with a free $\mathbb{B}$-semimodule $\mathrm{Fr}(\varepsilon)$ with a basis $\{[M]\}_{\partial M \cong \varepsilon}$ given by formal symbols $[M]$ of all 1-dimensional objects M which have ε as outer boundary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

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



On $\mathrm{Fr}(\varepsilon)$, 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 α:

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



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

Now define the state space $A(\varepsilon)$ as the quotient of $\mathrm{Fr}(\varepsilon)$ by an equivalence relation,

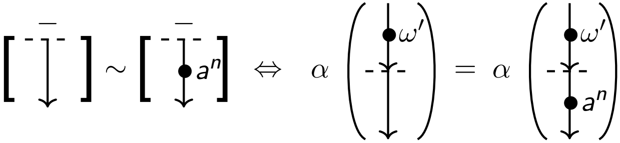
$$A(\varepsilon) \ := \ \mathrm{Fr}(\varepsilon)/\sim,$$

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

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

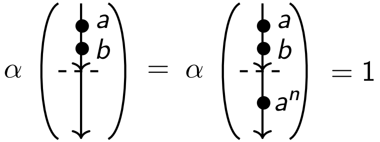
State space $A(\varepsilon)$ is spanned by $\mathbb{B}$-linear combinations of 1-manifolds M with $\partial M \cong \varepsilon$, modulo relations: two linear combinations are equal if for any way to close them up and evaluate using α, the result is the same.

One of the relations for the language $L_{\mathsf{I}} = (a+b)^* b(a+b)$:

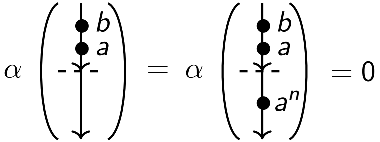


for any $\omega' \in \Sigma^*$.

If $\omega' = ba$, then



If $\omega' = ab$, then



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 $A(+)$ and $A(-)$ with values in $\mathbb{B}$. The matrix is not symmetric.

spanning elmt		x	x	y	x	z	y	$y+z$
spanning elmt			a	b	a a	a b	b a	b b
x'		0	0	0	0	1	0	1
y'	a	0	0	1	0	0	1	1
y'	b	0	0	1	0	0	1	1
0	a a	0	0	0	0	0	0	0
0	a b	0	0	0	0	0	0	0
z'	b a	1	1	1	1	1	1	1
z'	b b	1	1	1	1	1	1	1

Defining relations:

$$x + y = y$$

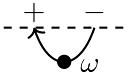
$$x + z = z$$

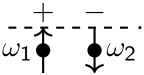
$$A(-) = \frac{\mathbb{B}x \oplus \mathbb{B}y \oplus \mathbb{B}z}{\langle x+y=y, x+z=z \rangle}$$

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 $A(\varepsilon) \longrightarrow A(\varepsilon')$ by concatenation.

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

A $\mathbb{B}$-semimodule V is a commutative idempotented monoid under addition:

$$x + x = x \text{ for } x \in V \text{ since } 1 + 1 = 1. \text{ Also } 0 + x = x,$$
$$x + y = y + x, \qquad (x + y) + z = x + (y + z).$$

Such V correspond to sup-semilattices, with join (least upper bound) $x \vee y := x + y$, and $x \leqslant y$ iff $x + y = y$.

0 is the minimal element, i.e., $0 \leqslant x$ for any x.

Any finite sup-semilattice is a finite lattice, with meet $x \wedge y := \sum_{z \leqslant x, y} z$ and $1 = \sum_{z \in V} z$.

We mostly use $\mathbb{B}$-semimodule structure (join, not meet).

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

$$\text{finite (sup)-semilattices} \Leftrightarrow \text{finite lattices}$$

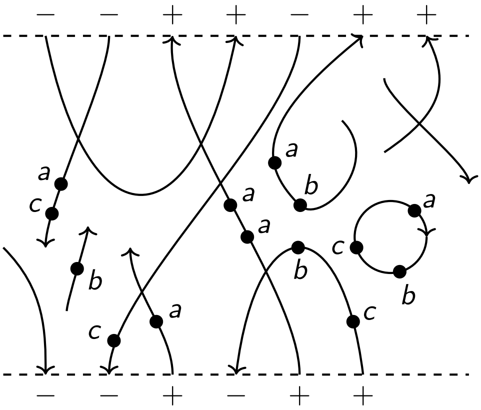
$\mathbb{B}$-semimodules constitute a category; morphisms are semimodule homomorphisms $f : V \longrightarrow W$, $f(0) = 0$, $f(x+y) = f(x) + f(y)$.

$\mathrm{Hom}(V, W)$ 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.

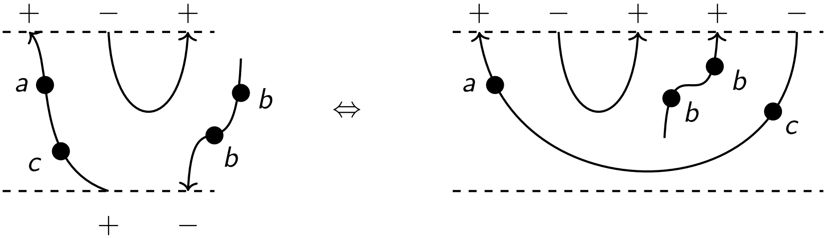
Categories of cobordisms in the universal construction that we build from evaluations are rigid.

Any cobordism C between $\varepsilon, \varepsilon'$ induces a semimodule homomorphism $A(\varepsilon) \to A(\varepsilon')$ of concatenation with C:



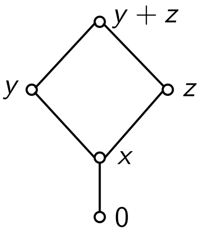
A cobordism from $(- - + - + +)$ to $(- - + + - + +)$.

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

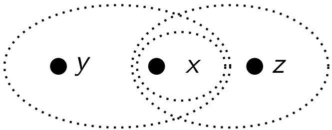


Recall the language $L = (a+b)^* b(a+b)$. The module $A(-)$ is spanned by x, y, z, and has relations $x + y = y$ and $x + z = z$. This module is not free. We'll encounter its free cover later in the construction of minimal NFA (nondeterministic FA) for L.

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



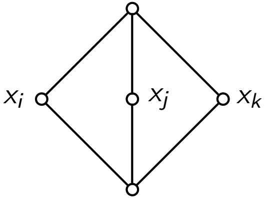
The finite topological space associated to this example:



Lattices that come from finite topological spaces are distributive.

If a lattice contains either as a sublattice,

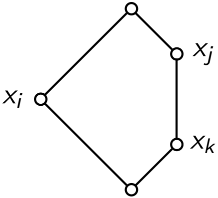
$$x_i + x_j = x_i + x_k = x_j + x_k$$



$$x_i \cap x_j = x_i \cap x_k = x_j \cap x_k$$

$$x_i \cap x_j + x_j = x_j$$

$$x_i \cap x_j < x_i, x_j, x_k$$



$$x_i + x_j = x_i + x_k \text{ for } j < k$$

$$x_j + x_k = x_j \text{ for } j < k$$

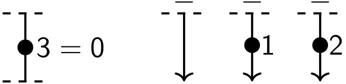
$$x_i \cap x_k < x_i, x_k$$

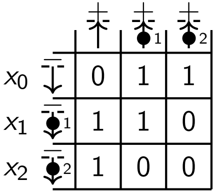
$$x_i, x_j < x_i + x_j$$

then the lattice is not distributive.

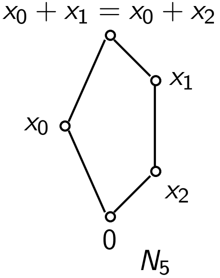
In such a case, there is no finite topological space associated to the language.

Example: for the language $L_{\mathrm{I}} = \{a, a^2\}$, lattices $A(-), A(+)$ are not distributive.

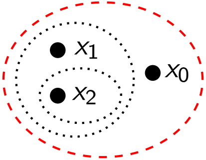




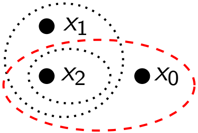
$$x_1 + x_2 = x_1$$
$$x_0 + x_1 = x_0 + x_2$$



For the language $L_{\text{I}} = \{a, a^2\}$, 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.



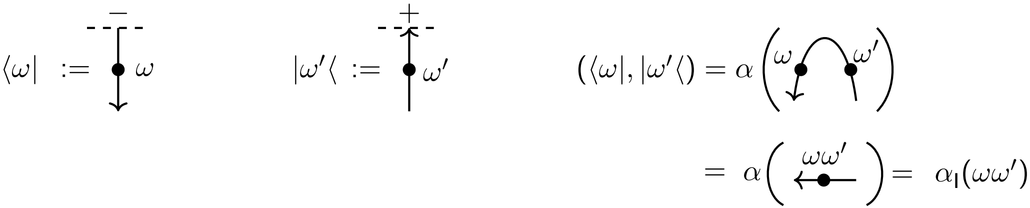
But since $x_0 \neq x_0 + x_2$, this finite topological space does not correspond to L_{I} as well.

Theorem. Languages $L_{\mathsf{I}}, L_{\circ}$ are regular $\Leftrightarrow$ the state space $A(\varepsilon)$ is a finite $\mathbb{B}$-semimodule for all sequences ε.

Get a $\mathbb{B}$-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 $\mathbb{B}$-linear combinations of diagrams below on the left, modulo equivalence relations coming from the pairing

$$A(-) \times A(+) \longrightarrow \mathbb{B}.$$



How do we build the minimal deterministic FSA and nondeterministic FSA for L_{I} from $A(-)$?

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

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

Initial state $q_{\mathrm{in}} = \langle\emptyset|$. A state $\langle\omega|$ is accepting iff $\alpha_{\mathsf{I}}(\omega) = 1$. *Nondeterministic* FSA for L_{I} come from coverings of $A(-)$ by free $\mathbb{B}$-modules with lifted action of Σ and unit, trace α maps.

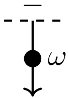
$\widetilde{m}_a \circlearrowright \mathbb{B}^J$ free semimodule cover; minimal NFA for L_I, where $J = \mathrm{irr}(A(-))$ (irreducible if $x \neq y + z$, where $y \neq x$, $z \neq x$)

$\downarrow$

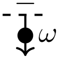
$m_a \circlearrowright A(-)$ state space of 0-manifold $-$

$\uparrow$

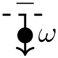
$m_a \circlearrowright Q_-$ minimal DFA for L_I

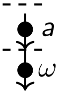


Every word gives a diagram in $A(-)$.

Start with a state 

and take images of all $\omega \in A(-)$ under the action by Σ^*, i.e.,

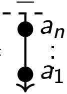


$= \omega = \langle \omega | \in A(-) \mapsto$


$=$

$\omega a \in A(-) \mapsto \begin{cases} 1 & \text{if } \omega a \in L_\mathsf{I}, \\ 0 & \text{if } \omega a \notin L_\mathsf{I}. \end{cases}$

$q_{\mathrm{in}} =$

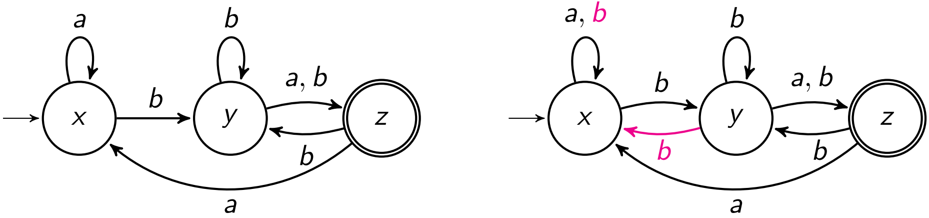
$= \langle \varnothing | \overset{a_1}{\mapsto} \langle a_1 | \overset{a_2}{\mapsto} \langle a_1 a_2 | \mapsto \ldots \overset{a_n}{\mapsto} \langle a_1 a_2 \cdots a_n | =$


$\mapsto \begin{cases} 1 & \text{if } a_1 \cdots a_n \in L_\mathsf{I}, \\ 0 & \text{if } a_1 \cdots a_n \notin L_\mathsf{I}. \end{cases}$

In general, there could be more than 1 minimal NFA.

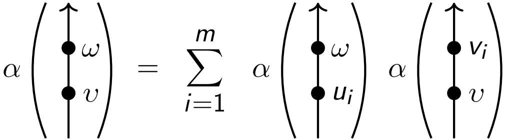
Two minimal nondeterministic automata on 3 states that accept the language $L = (a+b)^* b(a+b)$.

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 Σ^* from $A(-)$ to $\mathbb{B}^J$.

Some regular languages allow decomposition of identity

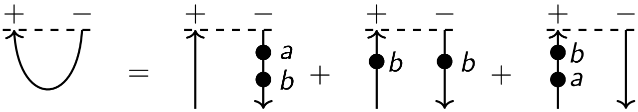


for some set of pairs of words (u_i, v_i), $1 \leqslant i \leqslant m$.

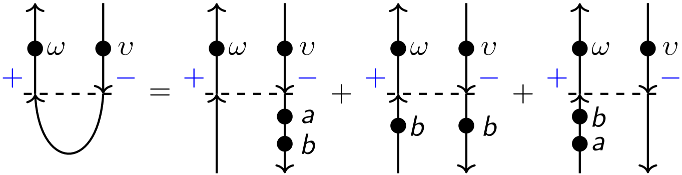
That is, for any $\omega, \upsilon \in \Sigma^*$,

$$\alpha_I(\omega\upsilon) \;=\; \sum_{i=1}^{m} \alpha_I(\omega u_i)\, \alpha_I(v_i \upsilon).$$

Returning to our example $L = (a+b)^* b (a+b)$,

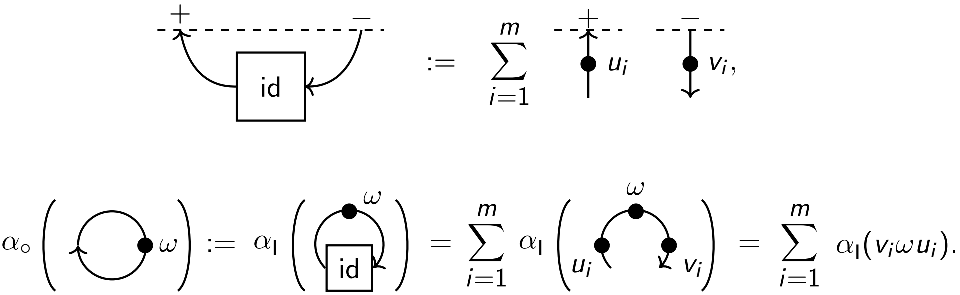


So



$$\alpha_{\mathsf{I}}(\omega\upsilon) = \alpha_{\mathsf{I}}(\omega)\,\alpha_{\mathsf{I}}(ba\upsilon) + \alpha_{\mathsf{I}}(\omega b)\,\alpha_{\mathsf{I}}(b\upsilon) + \alpha_{\mathsf{I}}(\omega ba)\,\alpha_{\mathsf{I}}(\upsilon).$$

For L_I with a decomposition of the identity, there is a unique associated circular language such that the decomposition still holds:



This gives a $\mathbb{B}$-valued TQFT: $A(\varepsilon)$ is the tensor product of $A(+), A(-)$ for the sequence of signs in ε.

For example, $A(++-) \cong A(+) \otimes A(+) \otimes A(-)$.

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

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

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

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

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 $U + V := U \cup V$.

Summary.

A pair $\alpha = (\alpha_{\mathsf{I}}, \alpha_{\circ})$ gives rise to a Boolean topological theory (state spaces are finite) iff $\alpha_{\mathsf{I}}, \alpha_{\circ}$ are regular languages.

Such a theory is *a weakly (=lax) monoidal functor* from the category of oriented 1D cobordisms with Σ-defects to the category of finite (semi-)modules over $\mathbb{B}$.

State spaces $A(-)$, $A(+) \cong A(-)^*$ are determined by the interval language α_{I} only.

If $A(-)$ is a projective $\mathbb{B}$-semimodule (comes from a finite topological space X), there is a unique circular language $\alpha_{\circ}$ making the pair of languages $(\alpha_{\mathsf{I}}, \alpha_{\circ})$ into a Boolean 1D TQFT with defects (maps $A(\varepsilon) \otimes A(\varepsilon') \longrightarrow A(\varepsilon\varepsilon')$ are isomorphisms of state spaces). Language $\alpha_{\circ}$ is given by traces of action of $\omega \in \Sigma^*$ on $A(-)$.

1D Σ-defect TQFTs are more general than automata. Boolean combinations of states are replaced by open sets in X, and a letter $a \in \Sigma$ takes open sets to open sets respecting unions of sets.

Work in progress.

1. Distributivity of $A(-)$ is a subtle property of a regular language α_{I}, even for $\Sigma = \{a\}$ (single letter). Study distributivity of regular languages (joint with R. Kaldawy, M. Khovanov, Z. Lihn).

2. Allow defects to accumulate towards inner endpoints. Evaluation of infinite words. Resulting topological theories relate to sofic systems and symbolic dynamics (joint with M. Khovanov, P. Gustafson).

3. Automata with boundary. Boolean evaluations beyond automata.

4. Interpretation of pseudocharacters (character-like functions on groups, useful in number theory) via 1D topological theories with defects (joint with M.Khovanov, V. Ostrik; on the arXiv:2303.02696).

5. Boolean two-dimensional topological theories and TQFTs. Ultimately hope to study these topological theories in dimension three as well.

Thank you!