SIMPLICIAL REPRESENTATION OF SET-MEMBERSHIP

DAVID I. SPIVAK

1. Introduction

This note was inspired by a question of Peter Gates. Set-membership is a relation: if \( S \) is a set of sets and \( E \) is a set of elements then membership can be represented by an injective function

\[
i: M \hookrightarrow E \times S.
\]

Typically, one wants to analyze patterns in how elements are shared between different sets \( s \in S \). Atkins has suggested that one analyzes these types of situations in terms of simplicial sets. I agree, and I’ll write down how that is done. I also believe that fuzzy simplicial sets can be useful in analysis; I’ll describe this as well.

The data of set-membership can be represented as a simplicial set, a system of vertices, line segments, solid triangles, solid tetrahedra, and higher-dimensional triangular analogues. These are called hulls; every hull has \( n \) vertices for some \( n \geq 1 \) and a hull with \( n \) vertices is called an \( n \)-hull. A simplicial set is a system of such hulls which are not free-standing but instead share sub-hulls with each other. Going back to set membership, each set has a number of elements, and these sets are not free-standing but instead can share elements with each other.

In our conversion of a set-membership scenario \( i: M \hookrightarrow E \times S \) into a simplicial set \( S_i \), the vertices in \( S_i \) will represent elements \( e \in E \). Each set \( s \in S \) will become a hull in \( S_i \) consisting of a vertex for each member of that set – thus if \( s \) consists of \( n \) elements it will be represented by an \( n \)-hull.

Suppose that \( s = \{a, b, c, d\} \) and \( t = \{a, b, e\} \) are sets. As a simplicial set, we’ll see this as the union of a 4-hull (tetrahedron) \( s \mapsto \Delta^4 \{a,b,c,d\} \) and a 3-hull (triangle) \( t \mapsto \Delta^3 \{a,b,c\} \), glued together along a 2-hull (edge) \( (s \cap t) \mapsto \Delta^2 \{a,b\} \). While \( t \) represents a filled in triangle, one can have a different set-membership scenario that resulted in an “empty triangle.” For example let \( x = \{a, b\}, y = \{b, c\}, z = \{a, c\} \); this would be represented as an empty triangle.

The result of all this is to connect a collection of elements if that collection is contained in one of the sets. But what if a collection of elements is found in multiple sets – we may want to say that this collection of elements are more strongly connected. For example if we are recording diagnoses, each person may have an associated set of diagnoses. We may want to connect a collection of diseases more strongly if there are more people who have had each of these diseases. For this, we should perhaps use fuzzy simplicial sets. Connections have a “fuzziness” or “strength”. One may turn a fuzzy simplical set into a metric space in which tight connections are closer together. See [Spivak. Metric realization of fuzzy simplicial sets].

This project was supported by ONR grant: N000140910466.
In this note, when I say simplicial sets I do not mean to include the possibility of "degeneracies." Those readers who do not know what degeneracies are do not need to know, but one includes them if having one instance of something is considered to imply arbitrarily many instances of that thing (as a disease diagnosed once in a patient could be diagnosed again for every moment it inhabits the patient). For those readers who do know what degeneracies are, I am simply using the category of ordered finite non-empty sets and inclusions between them rather than the category of ordered finite non-empty sets and all maps between them. I also blur the distinction between simplicial sets and symmetric simplicial sets. The first corresponds to lists, whereas the second corresponds to sets. Anything said about one in this paper could be applied to the other with the appropriate changes.

2. Simplicial representation of set-membership

Fix a set-membership scenario \( i: M \hookrightarrow E \times S \). We call elements of \( S \) sets, elements of \( E \) elements and elements of \( M \) memberships.

The membership complex associated to \( i \) is the simplicial set \( X(i) \) with \( n \)-hulls

\[
X(i)_n := \{ M \times S M \times S M \times \cdots \times S M \}
\]

where there are \( n \) copies of \( M \) being multiplied. So \( X(i)_0 = M \) is the set of memberships and \( X(i)_1 = M \times S M \) represents two members of the same set.

The simplicial set representing all possible sets on elements \( E \) is a big hull; if \( \#(E) = m \) then it is an \( m \)-hull, which we denote \( \check{E} \). The map \( M \to E \) induces a map \( M \times \cdots \times M \to E \times \cdots \times E \), which we can compose with the inclusion \( M \times S \cdots \times S M \to M \times \cdots \times M \) to give a map

\[
X(i) \to \check{E}.
\]

The image of this map is our other object of interest; we denote it by \( Y(i) \) and call it the reduced membership complex associated to \( i \).

Consider again the example of diagnoses. Given diagnosis data \( i: M \hookrightarrow E \times S \), the membership complex \( X(i) \) consists of a disjoint union of different hulls, one for each patient. The reduced membership complex \( Y(i) \) is a (not necessarily disjoint) union of hulls – we gather diseases into a hull if they have all been diagnosed in some patient. So each patient has an associated hull of diseases, and as different patients may share some subset of diagnoses, different hulls can be connected.

As mentioned above, if we want to more strongly associate diseases that are more commonly shared by a patient then we need to use the map \( \pi: X(i) \to Y(i) \) (the image map of \( X(i) \to \check{E} \)) to create a fuzzy simplicial set.

3. Fuzzy simplicial representation of set-membership

Let \( i: M \hookrightarrow E \times M \) be a membership scenario, and let \( \pi: X(i) \to Y(i) \) be as above. To every \( n \)-hull in \( y \in Y(i)_n \), let \( \eta(y) = \frac{\#\pi^{-1}(y)}{\#S} \in (0, 1] \). In our running application, \( y \) represents a set of diseases and \( \eta(y) \) represents the number of people who share every disease in \( y \) divided by the total number of people.

If \( y_1 \) is a face of \( y_2 \) then we easily check that \( \eta(y_1) \geq \eta(y_2) \); thus \((Y, \eta)\) satisfies the axioms of a fuzzy simplicial set. To draw this fuzzy simplicial set, imagine that each vertex, edge, triangle, etc. was drawn on a grey scale. The more people who have a disease or collection of diseases, the darker we should draw the shape representing that disease or collection.