Mathematics of Classifications, Chu spaces and the Continuum

DANIEL PARROCHIA
Institute of philosophical research (IRPHIL)
University Jean Moulin - Lyon III
1C avenue des Frères Lumière CS 78242 69372 Lyon Cedex 08
FRANCE

Abstract: We call *classification* the operation consisting of distributing objects in classes or groups which are, in general, less numerous than them. The relations between these classes may be (or not) partially or totally ordered. So there exist many kinds of classification schemes. Formally speaking, a classification may be a lattice, a semilattice, a chain, a hypergraph, a matroid, a tree, etc. Our purpose is to find the underlying mathematical structure of all these classifications. We explain how we can represent them in a unique way, constituting what has been called before a ”metaclassification” and which is, in fact, a Chu space. Thus, category theory can describe in terms of morphisms different operations on classifications that, according to Barwise and Seligman, we report in the following. Finally, we show that such a ”metaclassification” is the fundamental brick from which we can get some information about the mathematical continuum.

Key-Words: Partition, partition lattice, chains of partitions, Chu spaces, ellipsoids, continuum.

Received: May 19, 2022. Revised: February 13, 2023. Accepted: March 21, 2023. Published: April 13, 2023.

1 Introduction

In this article, we first define what should be an *invariant* for a classification, as well as a *classification function*. Then we give different examples of increasingly strong classificatory structures (hypergraphs, covers, class systems, partitions). In the following, we study more particularly the lattice of the partitions of a finite set. We show that this model generalizes to the case of minimum covers and to other situations. We then introduce the project of a general theory of classifications with its related concepts (metaclassification, Chu spaces, categories and infomorphisms). Finally, failing to be able to construct a real algebra of classifications, we introduce a conjecture of convergence for the ellipsoids of the space of classifications. We end on the question of the continuum, already mentioned in [19].

2 Invariants and classification functions

Let E be a nonempty finite or infinite set. The various elements of E are usually compared by the means of some *invariant* (see, [13]).

For example, partitioning $\mathbb{N}$, the set of natural numbers, into odd and even numbers, supposes you take for invariant their classes modulo 2. Now, if you want to classify the abstract sets in general, then you will have to take for invariant their cardinals. In experimental sciences (physics, chemistry, natural sciences...), there are more complex invariants, such as discrete groups, Lie groups and so on.

Invariants are criteria that allow us to tell whether the objects we compare are similar or not. However, in practical domains, the invariants are not easy to find. Generally, we must previously construct the notions of ”likeness”or ”proximity”of two objects according to the phenomenology, and use, in order to express them rigorously, a mathematical coefficient of similarity, which is a pairwise ”distance” among the objects of the basic set. These ones are points of the space that we identify with the set $E = \{a, b, ..., x...\}$.

We then define a function $d : E \times E \to \mathbb{R}$ with particular properties (generally, it satisfies the axioms of some *distance* or *dissimilarity coefficient*) that we will define later.

Now a *classification function* is a function f that takes this distance function d on E and returns some organization Γ of E (hypergraph, system of classes, partition, etc.). Γ is a ”classification” and the sets in Γ will be called ”classes”. There may be some problem to define such a function and some yet reasonable requirements (such as scale invariance, richness and consistency) lead to a contradiction. One or the other must be weakened (see, [14]).

3 Examples of different structures

Definition 3.1 (Hypergraph)**.** Let $P(E)$ be the powerset of E. A hypergraph is a pair $H = (E, P)$, where E is a set of vertices (or nodes) and P a set of nonempty subsets called (hyper)edges or links. Therefore, P is a subset of $P(E) \setminus \emptyset$. In such a structure, the set of edges, does not "cover" the set E, because some node may have a degree zero, i.e. may have no link to some edge (see, [6]).

Definition 3.2 (Covers)**.** Assume Definition 3.1 and suppose now we add the following conditions (see, [18]):

(C_0) $\qquad E \in P.$

In this case, we say that P is a cover (or covering) of the set of vertices E.

Definition 3.3 (Systems of classes)**.** Assume now Definition 3.1 and 3.2. and suppose that, for every element, its singleton is in P. In symbols:

(C_1) $\qquad \forall x \in P,\ \{x\} \in P.$

then, we get a system of classes, in the sense of [7].

Definition 3.4 (Partitions)**.** Assume 3.1, 3.2 and 3.3 and let us add now the new following condition. For every c_i belonging to P:

(C_2) $\qquad c_i \cap c_j = \emptyset.$

(C_3) $\qquad \cup c_i = E.$

Then P is a partition of E and the c_i are the classes of the partition P.

4 The lattice of partitions

Call now xPy, the relation "x belongs to the same class as y" and denote $\mathbf{P}(E)$ the set of partitions of a set E. A partition P' is finer than a partition P if $xPy \implies xP'y$. This relation allows us to define a partial order on $P(E)$ that we shall denote $P' \leq P$. We can see immediately that $(\mathbf{P}(E), \leq)$ is a lattice: 1) it is a partial order; 2) moreover, every pair (P, P') has a greatest lower bound $P \wedge P'$ and a least upper bound $P \vee P'$. One proves that $P(E)$ is complemented, semi-modular and atomic (if the initial data E is a non atomic set, we can, under reasonable conditions, reduce the data to the atomic elements of E) (see, [15]).

Example : The lattice of partitions for |E| = 3 (see Fig. 1) : as usual, we write (a, b) for $\{\{a\}, \{b\}\}$:

$$P_0 = (a, b, c); \quad P_1 = (ab, c); \quad P_2 = (ac, b);$$

$$P_3 = (a, bc); \quad P_4 = E = (abc).$$

P_0 is the discrete partition, whose classes are singletons. P_4 is only one class, say: E.

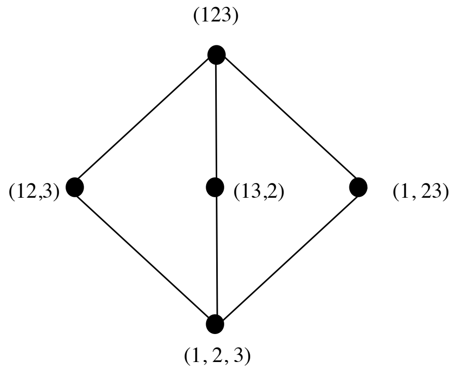


Figure 1: The partition lattice for $|E| = 3$

5 Hierarchical classification

In this context, a hierarchical classification, i.e a chain C of partitions of the lattice $\mathbf{P}(E)$, is a totally ordered subset of $\mathbf{P}(E)$. We have:

$$C = \{P_1, P_2, ..., P_n\} \quad \text{with} \quad P_1 < P_2 < ... < P_n,$$
$$\text{and } P_i \in \mathbf{P}(E).$$

Example : the four chains of $\mathbf{P}(E)$ for $|E| = 3$ are:

$$C_0 = \{E, P_0\}; \quad C_1 = \{E, P_1, P_0\};$$
$$C_2 = \{E, P_2, P_0\}; \quad C_3 = \{E, P_3, P_0\}.$$

Of course, there is a one-to-one correspondence between chains and usual presentation of hierarchical classifications, that are trees or dendograms (see Fig. 2):

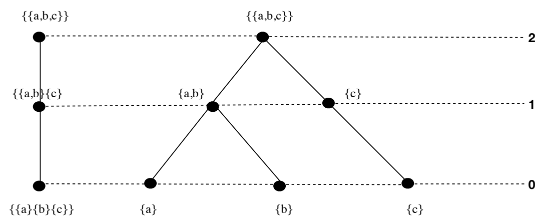


Figure 2: Correspondence between chain and hierarchical classification

Note that the whole set of chains $\mathbf{C}(E)$ has itself a mathematical structure : it is a semilattice for set intersection. This model allows us to get all the possible partitions of $\mathbf{P}(E)$ and all the possible chains of $\mathbf{C}(E)$, the set of chains of the partition lattice $\mathbf{P}(E)$.

6 Distance and ultrametric matrices

Given a tree or dendogram T corresponding to a chain of partitions, we introduce now (as in Fig.2, right) an index $i \in I(I \subset \mathbb{N})$ associated to the levels of this structure, so it becomes an indexed hierarchy H.

Let us now define on the product set $E \times E$ a mapping $d : E \times E \to \mathbb{R}^+$, such that, for all $e, e' \in E$, we have the following properties:

1. $d(a, b) \geq 0$;

2. $d(e, e') = 0$ iff $e = e'$;

3. $d(e, e') = d(e', e)$.

d is a *dissimilarity coefficient*. Now if it satisfies also:

4. $\forall e, e', e'' \in E, d(e, e') \leq \text{Max}(d(e, e''), d(e', e''))$, then, d is an ultrametrics.

By setting, for all $e, e' \in E,\ d(e, e') = i(He, e')$, we can deduce from the indexed hierarchy H a dissimilarity d on E, which is in fact an ultrametrics on E. Clearly, $d(e, e')$ is the smallest $i \in \mathbf{R}^+$, so that e and e' are in the same class of $P_i,\ i$ being the level of the classification.

Example : Consider Fig. 2 (right) : 0 is the level of discrete partition, 1 the intermediary level and 2 the level of the coarse partition, the set E itself, as a whole part of itself. We can see that:

$d(a, a) = 0$ because level 0 is the level where 1 belongs to the same class as itself. $d(a, b) = 1$ because level 1 is the level where 1 belongs to the same class as 2. $d(a, c) =$ 2 because level 2 is the level where 1 belongs to the same class as 3.

So we can write the corresponding ultrametric matrix:

	a	b	c
a	0	1	2
b	1	0	2
c	2	2	0

The problem is that there is no chance the objects of the world be spontaneously ordered according to some ultrametric distance. If there any, we have to extract it from the empirical data, with some reasonable hypotheses.

7 A generalization

To solve the problem, clustering scientists propose algorithms to construct classifications by bottom-up or top-down methods which aggregate the objects to be classified by comparing the pairwise distances defined on them. There is a lot of possibilities.

Since, a new problem arise : there is no consensus about the type of distances we can choose to put empirical datas in order, whether it concerns the distances between the objects or the distances between the subsets of objects.

But in the fundational point of view which is ours, it is clear that we must choose the most general distance d that we can find.

As the relations between objects and classes are not necessarily exclusive but yield overlapping situations, some searchers like Jardine and Sibson (see, [12]) have proposed to weaken the notions of partition and classification, changing them, respectively, in the notions of k-partitions and k-classifications.

Let us explain, first, the meaning of such generalizations. The authors begin with a definition of a dissimilarity coefficient like the previous one. They just precise that it is not necessary that the function d satisfies (4), the ultrametric inequality. On the contrary, it may satisfy a weaker condition such that:

5. $d(a, c) \leq d(a, b) + d(b, c)$ for all $a, b, c \in E$ (this inequality defines in fact what is called now a ”linear dissimilarity).

On this basis, Jardine and Sibson introduce the following definitions:

Definition 7.1 (k-partition)**.** A k-partition is a partition which makes possible the existence of a maximum of $k - 1$ objects in the overlaps between the classes (i.e. in the intersection of them).

Definition 7.2 (k-classification)**.** A k-classification is a nested sequence of k-partitions. The system will be hierarchic in case $k = 1$, and overlapping in case $k > 1$.

Definition 7.3 (k-dendogram)**.** By a corresponding generalization, we may define the notion of a k-dendogram, i.e. a dendogram where clusters, at a given level, may overlap to the extent of $k - 1$ objects.

In this view, the nearest neighbor algorithm aggregates to an object e_m its nearest neighbor e_n and then iterate the same method with e_n, etc.). The result, noted as B_1 in the hierarchical case, can be generalized in a sequence of methods denoted by (B_k). B_1 is the ”single-link” method, which leads to a hierarchical dendogram (1-dendogram). The second member of the sequence (B_2) may be called the ”double-link”method, and leads to a dendogram in which clusters may overlap to the extent of one object (a 2-dendogram). And so on.

Let E be the set of objects to be classified, with Card

$E = p$. Then $B_p(1)$ gives an exact representation of the dissimilarity coefficient. It can be shown that the family of measures of distortion is monotonic decreasing with increasing k, becoming zero in case $k = p - 1$.

The sequence of cluster methods (B_k) can be given a simple graph-theoretic description which generalizes the one given for the single-link method. The clusters at level h in $(B_k(d))$ may be obtained as follows. A graph is drawn, whose vertices represent the objects and whose edges join just those pairs of points which represent objects, with dissimilarity less than or equal to a given number h.

The maximal complete subgraphs (i.e. the maximal subsets of the set of vertices in which all possible edges are present) are marked, and wherever the vertex sets of two such subgraphs intersect in at least k vertices, further edges are drawn in to make the union of the two vertex sets into a complete subgraph. The process is repeated until there is no further alteration. We just indicate the result in Fig. 3:

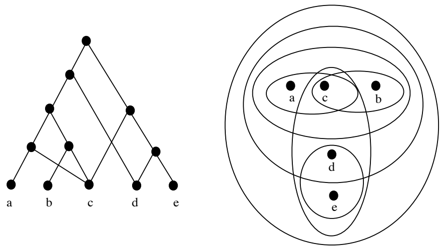


Figure 3: A k-classification

As Jardine and Sibson said, this algorithm is not suitable for computation, and graphic representations are not necessarily very meaningful. But it does not matter for our purpose.

The main point is the comment of J.-P. Benzecri (see, [5]). This one shows that the generalization of the ultrametric inequality, which is carried out by the authors and leads to their hierarchy of overlapping clusters, is based on a system of pseudo-bowls defined as follows:

Definition 7.4 (Pseudo-bowl)**.** We call "pseudo-bowl"with diameter δ, compared to the index of distance d, a subset B of E, the set of objects, which is maximal among those such that:

$$\forall i, i' \in B, d(i, i') \leq \delta.$$

As Jardine and Sibson show, B is an "ultimate cluster"for the binary relation $d(i, i') \leq \delta$. This relation is an equivalence relation only in the case d is ultrametric.

On the contrary, if d satisfies the inequality of order k, two pseudo-bowls of the same diameter (compared with d) coincide if the cardinality of their intersection is superior or equal to $k - 2$. Obviously, in the case $k = 3$, d is ultrametric.

Of course, as Benzecri writes, Jardine and Sibson's approach incurs some criticism:

1. As the authors recognized, there is no simple algorithm for generating the system of all pseudo-bowls with any diametrer δ associated to a given distance d.

2. Now if we want the condition of order k very different from the condition $p = 3$, the number $k - 3$ (i.e. the order of covering which is allowed between the clusters) must be negligible, compared with the cardinality of the set E of objects. Unhappily, it does not seem that the algorithm of distance changes could be still practicable for $k > 6$.

3. The tree-representation of this kind of classification is not very clear.

Anyway, this proves that the most general type of dissimilarity coefficient that we can choose is the one which yields this space of pseudo-bowls.

8 Other situations

In fact, the case of Fig. 3, as the following example of a more complex semilattice (see Fig. 4), where clusters, at different levels, may overlap to the extent of 2 - 1 = 1 object (so we have a 2-dendogram in the terminology of Jardine and Sibson), already give ultrametrics :

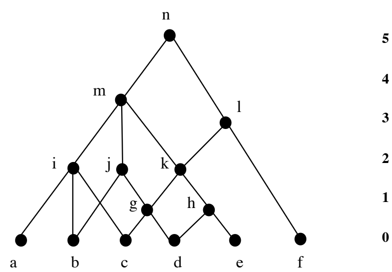


Figure 4: A more complex semilattice

Indeed, this semilattice is an overlapping classification. Pose:

$$i = \{a, b, c\}; \quad j = \{b, d\}; \quad g = \{c, d\}; \quad h = \{d, e\};$$
$$k = \{g, h\} = \{c, d, e\}; \quad l = \{k, f\} = \{c, d, e, f\};$$
$$m = \{i, j, k\} = \{a, b, c, d, e\};$$
$$n = \{m, l\} = \{a, b, c, d, e, f\}.$$

Its ultrametric matrix of distances appears down below:

	a	b	c	d	e	f
a	0	2	2	4	4	5
b	2	0	2	2	4	5
c	2	2	0	1	2	3
d	4	2	1	0	1	3
e	4	4	2	1	0	3
f	5	5	3	3	3	0

So, to some k-classifications may correspond ultrametric distances. These ones are included in the set of all ultrametrics distances whose structure has been discovered by Gondran ([10]) in the following theorem:

Theorem 8.1 (Gondran 1976)**.** *The matrix of ultrametric distances between the elements of a set E has a structure of semiring over $\mathbb{R} \cup +\{\infty\}$.*

Proof. There exists an associative law $\oplus$, which may be interpreted as "Min".

$$d_i \oplus d_j = \mathrm{Min}(d_i, d_j).$$

The unit element of $\oplus$ is $(+\infty)$, and we have:

$$d_i \oplus +\infty = +\infty \oplus d_i = \mathrm{Min}(d_i, +\infty) = d_i.$$

There exists also an associative law *, which is distributive over $\oplus$. This law * is interpreted as "Max", and its unit element is 0, since:

$$d_i * d_j = \mathrm{Max}(d_i, d_j)$$

and:

$$0 * d_i = d_i * 0 = \mathrm{Max}(0, d_i) = d_i.$$

This structure may be extended to M_n, the set of all the matrices of the chains of $\mathbf{C}(E)$. □

In the same way, as the set of minimal covers form a lattice, the chains of minimal covers are the chains of this lattice and their ultrametric matrices are in the semiring $(\mathbb{R} \cup \{+\infty\}, \oplus, *)$ of Gondran. However, as the set of covers is only a preorder, we cannot get easily the chains of covers, and so, in this case, the previous good property is lost.

This means also that many overlapping classifications have matrices of distances that are not ultrametric. The situation may be the same for fuzzy classifications.

9 Towards a general theory of classifications

This situation suggests to present every type of quasi-taxonomic organization defined on a set of objects (system of classes, partitions, k-partitions, classifications, k-classifications, etc.) as a system of intersecting bowls (see, [19]).

In order to get the most general point of view, we may replace (as we have done in Fig. 3 right) bowls by ellipsoids, knowing that, when some "good" distance, if there is any, has been defined on the set of objects for some classification, the ellipsoids may be interpreted as ellipsoids of inertia which give information about the aspects of classes.

Proposition 9.1. *Classes and every type of quasi-taxonomic organization of classes on a set of objects E may be represented as ellipsoids of the space, i.e., if we project it on a surface of dimension 2, as simple ellipses of the plane.*

Proof. Let E be the set of objects to be classified. As we know, every object e_i may be represented as a point of the space. Let S_k a subset of points. Every subset S_j may be circumbscribed by an ellipse. Suppose now that an S_j belongs to some S_k. Then, S_k will be also an ellipsoid. And so on, until we reach the whole space of objets E. □

We use ellipsoids or ellipses, rather than simple circles or n-spheres, because in factorial analysis, we can represent classes by ellipsoids of the space. In the case when classifications are constructed by bottom-up methods from a table of data, we can get information on the direction of extension of the ellipses associated to the classes in relation to the respective weight of their elements (we call that kind of ellipses "ellipses of inertia"). In computer science, there exist some programs to get that automatically (see, [11]).

When we do not get such an information, it is always possible to have a qualitative representation of classes in 2-dimension by ellipses. In those cases, the direction and the extension of ellipses are purely conventional.

Partitions and chains of partitions give very clear hypergraphs. Covers, chains of covers, lattices of covers or lattices of minimal covers, crossed partitions (like the Mendeleiev table of Elements), etc. would yield more complicated representations. Indeed, it is always possible to draw these ones.

DOI: 10.37394/23209.2023.20.14

10 Metaclassification

Let us recall now the following sets that can be defined on a nonempty finite set E:

1. $P(E)$, power set of E;

2. $\mathbf{P}(E)$, set of all partitions of E.

3. $\mathbf{C}(E)$, set of all chains of partitions on E.

Let us define now : $\mathcal{U}(E) = \{P(E), \mathbf{P}(E), \mathbf{C}(E)\}$. $\mathcal{U}(E)$ is the whole universe of classifications.

Even if it is a bit complicated, we can give an idea of this universe by representing all the possible pseudo-bowls, in fact ellipsoids (Neuville's ellipsoids) and sequences of ellipsoids, in one and the same figure. Intersecting or non intersecting classes are at the periphery or the biggest circle. Then, partitions may be found in the middle circle and chains just around the center. Pierre Neuville called this figure a "metaclassification"(see Fig. 5).

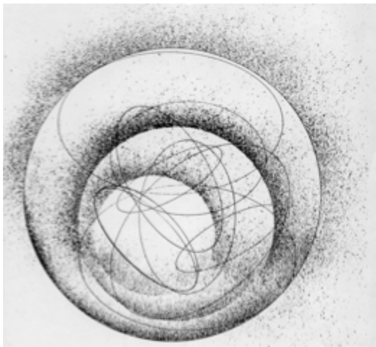


Figure 5: Metaclassification

When we ask : what is such a structure (from a mathematical point of view)? The answer is that we have got, at the most general level, the set of all possible open subsets of a set. Why do we speak of "open subsets"? Because we must observe that the boundaries of the ellipsoids are just there to realize the classes. But only the elements that they contain are to be considered as actually forming the classes. So, the classes themselves must be identified only with the *interiors* of the ellipsoids.

Now a well-known result of topology is that the algebra of the set of open subsets of a set is a Heyting algebra (see, [21]) . By the way, the previous structure is also a relatively pseudo-complemented lattice (see, [22]).

11 Chu spaces

The foregoing developments are already included in [18] and [19]. But we think we have now a bit more. In fact, more complete presentation of metaclassification makes it appear as a *Chu space*.

The Chu space concept originated with Michael Barr in 1979 (see, [1]) but the details were developed by his student Po-Hsiang Chu, whose master's thesis formed the appendix of the 1979 paper[1].

Statically, a Chu space (A, r, X) over a set K consists of a set A of points, a set X of states, and a function $r : A \times X \to K$. This makes it an $A \times X$ matrix with entries drawn from K, or equivalently a K-valued binary relation between A and X (ordinary binary relations being 2-valued).

More dynamically, Chu spaces transform in the manner of topological spaces, with A as the set of points, X as the set of open sets, and r as the membership relation between them, where K is the set of all possible degrees of membership of a point in an open set.

It is easy to recognize these elements in the above metaclassification, so we can state the following proposition:

Proposition 11.1. *The metaclassification $M(\mathcal{U})$ is the Chu space $(E, r, \mathcal{C})$ where E is the set of elements to be classified, $\mathcal{C}$ the set of classes, and r the degree of membership of an element to a class.*

If $K = \{0, 1\}$, we get the situation described in [18] or [19] and, for every $e \in E$ and $C \in \mathcal{C}$, we have $e \in C$ or $e \notin C$, which yields formal classifications. But if $K = [0, 1]$, we may integrate in the structure fuzzy classifications in the sense of Zadeh (see, [25], [4]) or in the sense of Lerman (see, [16]).

One of the advantages of using Chu's space is to get information about what is in the metaclassification structure and how, precisely, it is there. This amounts to comparing two different metaclassifications, i.e. two particular Chu spaces, therefore to constructing a continuous function from one to the next.

Now we can observe that the counterpart of a continuous function from (A, r, X) to (B, s, Y) is a pair (f, g) of functions $f : A \to B$, $g : Y \to X$ satisfying the adjointness condition $s(f(a), y) = r(a, g(y))$ for all $a \in A$ and $y \in Y$. That is, f maps points forwards at the same time as g maps states backwards. The adjointness condition makes g the inverse image function f^{-1}, while the choice of X for the codomain of g corresponds to the requirement for continuous functions that the inverse image of open sets be open. Such a pair is called a *Chu transform or morphism of Chu spaces*[2].

As it is known, the category of Chu spaces over K

[1] For the history of the construction, see, [2].

[2] A replacement of the adjointness condition $(s(f(a), y) = r(a, g(y))$ by the more general one $(s(f(a), y) \leq r(a, g(y))$ leads to a variant of Chu spaces, called "dialectica spaces" and due to de Paiva (see, [17]).

and their maps is denoted by $\mathbf{Chu}(\mathbf{Set}, K)$. As is clear from the symmetry of the definitions, it is a self-dual category: it is equivalent (in fact isomorphic) to its dual, the category obtained by reversing all the maps. It is furthermore a $*$-autonomous category with dualizing object $(K, \lambda, \{*\})$ where $\lambda : K \times \{*\} \to K$ is defined by $\lambda(k, *) = k$, as indicated in [1][3].

Usually, Chu spaces arise as the case $V = \mathbf{Set}$, that is, when the monoidal category V is specialized to the cartesian closed category **Set** of sets and their functions. But a more general enriched category, The category $\mathbf{Chu}(V, k)$ appeared for the first time in the appendix to [1], which constitutes the master's thesis of Po-Hsiang Chu.

Chu spaces are very important for our foundational point of view in classification theory[4].

From the metaclassification, understood as a Chu space, we can deduce a lot of mathematical structures. For example, Chu spaces over 2 realize both topological spaces and coherent spaces introduced by J.-Y. Girard, while Chu spaces over K realize any category of vector spaces over a field whose cardinality is at most that of K. This was extended by Vaughan Pratt (see, [20]) to the realization of k-ary relational structures by Chu spaces over 2^k. For example, the category **Grp** of groups and their homomorphisms is realized by $\mathbf{Chu}(\mathbf{Set}, 8)$ since the group multiplication can be organized as a ternary relation (see, [8]), and $\mathbf{Chu}(\mathbf{Set}, 2)$ realizes a wide range of "logical" structures such as semilattices, distributive lattices, complete and completely distributive lattices, Boolean algebras, complete atomic Boolean algebras, etc. Of course, the semiring of Gondran (which corresponds to hierarchical classifications) may be realized in the same way.

So Chu spaces not only formalize our metaclassification structure but give all required accesses to a classification of mathematical structures. But we can say more.

12 Chu spaces and informorphisms

The application of the formalism of Chu spaces in classification theory was carried out by Jon Barwise and Jerry Seligman in their reflection on information theory (see, [3]). As one can easily understand, the Chu spaces defined above formalize not only the notion of "meclassification", but that of "classification" itself. In addition, they allow us to define in an elegant and unified way a certain number of operations on classifications.

According to Barwise and Seligman (see, [3], 69), a classification $A = (\text{tok}(A), \text{typ}(A), \models_A)$ is defined by the following objects:

1. a set tok(A) of objects to classify, called "tokens" of A;
2. a set, typ(A) of objects used to classify tokens, and called "types" of A;
3. a binary relation, $\models_A$ between tok(A) and typ(A).

It is clear that such a definition is in fact only a paraphrase of the mathematical notion of Chu space. We now introduce some terminology.

12.1 Terminology

1. If $a \models_A \alpha$, then a is said to be *of type* α in A.
2. For any token $a \in \text{tok}(A)$, *the type-set* of a is the set:
$$\text{typ}(a) = \{\alpha \in \text{typ}(A) | a \models_A \alpha\}.$$
3. Similarly, for any type $\alpha \in \text{typ}(A)$, *the set of tokens* of α is the set:
$$\text{tok}(a) = \{a \in \text{typ}(A) | a \models_A \alpha\}.$$
4. α_1 and α_2 *in* typ(A) are *coextensive* in A - which is written $\alpha_1 \sim_A \alpha_2$ - if tok α_1) = tok(α_2).
5. The tokens a_1 and a_2 are *indistinguishable* in A - which is written $a_1 \sim_A a_2$ - if typ(α_1) = typ(α_2).
6. Finally, the A classification is *separated* if there is no distinct indistinguishable token, i.e. if $\alpha_1 \sim_A \alpha_2$ implies $a_1 \sim_A a_2$.
7. The classification A is *extensional* if all the coextensive types are identical, i.e. if $\alpha_1 \sim_A \alpha_2$ implies that $\alpha_1 \sim \alpha_2$.

Like any Chu space, a classification, of course, can be both extensional and separate.

[3] Incidentally, one knows that, as such, it is a model of Jean-Yves Girard's linear logic (see, [9]).

[4] Chu space were not studied in their own right until more than a decade after the appearence of the more general enriched notion. In [19], we quoted the book of Barwise and Seligman, who used some application of them in information theory (see, [3], 33, 92). We even mention ([19], 38) that they defined a classification (A, B, R) as consisting of a set A of tokens, a set B of types and a relation R between tokens and types (such that $R \subseteq A \times B$), but this language was not very clear and lacked generality. For the authors, tokens could be objects and types could be attributes describing the objects. But, as this structure was essentially defined for the study of a channel in information theory and has been only applied elsewere in formal concept analysis, it was very difficult to see the real importance of Chu spaces for a foundational viewpoint in classification theory. It was not obvious, indeed, to see that they could realize a lot of mathematical structures and that using categories and morphisms to formalize classification theory was very useful, even if the authors introduced some trivial operations on classifications as, for example, sums, construction of invariants or quotients, and some dualizing of these last (see, [3], 81-88).

12.2 Informorphisms

Barwise and Seligman, following the model of morphisms between Chu spaces, introduce the notion of *infomorphism*. It is indeed necessary to be able to think of the relations between classifications defined on the same objects or, on the contrary, on different objects. To do this, given two classifications to be compared, one then defines $f = \langle f^\wedge, f^\vee \rangle$ pairs of functions such as $f^\wedge$ is a function of the types of one of the classifications in the types of the other, and $f^\vee$ is a function of the tokens of one of the classifications in the tokens of the other.

Definition 12.1. We say that f, from A to B, is a contravariant pair, and we write: $f : A \rightleftarrows B$ if $f^\wedge$: typ (A) to typ (B) and $f^\vee$: tok (B) to tok (A).

Definition 12.2. We say that f, from A to B, is a covariant pair, and we write: $f : A \rightrightarrows B$ if $f^\wedge$: typ$(A) \to$ typ(B) and $f^\vee$: tok$(A) \to$typ (B).

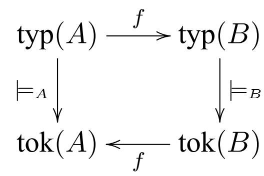
Definition 12.3. A *infomorphism* $f : A \rightleftarrows B$ from A to B is a contravariant pair of functions $f = \langle f^\wedge, f^\vee \rangle$ satisfying the following fundamental property of informorphisms:

$$f^\vee(\text{b}) \models_A \alpha \quad \text{ssi} \quad b \models_B f^\wedge(\alpha),$$

for each token $b \in$tok(B) and each type $\alpha \in$typ(A).

The accents will be omitted to the right of f if there is no confusion, knowing that the circumflex puts f on types and that the reverse accent puts it on tokens.

In terms of a diagram, we can then write:



12.3 Duality type-tokens

When including a subsection you must use, for its heading, small letters, 12pt, left justified, bold, Times New Roman as here.

Definition 12.4. For any classification A, the *flip* of A is the classification $A^\perp$ whose tokens are the types of A, whose types are the tokens of A and such that $\alpha \models_{A^\perp} a$ if and only if $a \models_A \alpha$.

Given a pair of functions $f : A \rightleftarrows B$, we define $f^\perp : B^\perp \rightleftarrows A^\perp$ by $f^{\perp\wedge} = f^\vee$and $f\perp^\vee = f^\wedge$. When classifications are defined by double-entry data matrices, this operation is equivalent to swapping rows and columns.

We can prove three propositions concerning infomorphisms:

Proposition 12.1. *$f : A \rightleftarrows B$ is an infomorphism if and only if $f^\perp : B^\perp \rightleftarrows A^\perp$ is an informorphism.*

Proposition 12.2. *1. $(A^\perp)^\perp = A$ and $(f^\perp)^\perp = f$.*

2. $(fg)^\perp = g^\perp f^\perp$.

Proposition 12.3. *Given an informorphism $f : A \rightleftarrows B$, if a type α is coextensive with a type α' in A, then, $f(\alpha)$ is coextensive with $f(\alpha')$ in B.*

12.4 Operations on classifications

There are obviously many operations which, from a given classification, make it possible to generate another classification. Barwise and Seligman have described essentially two of them, as well as their interactions with informorphisms.

Definition 12.5. Given the classifications A and B the *sum* $A+B$ of A and B is the classification defined as follows:

1. The set tok$(A + B)$ is the Cartesian product of tok(A) and tok (B). Specifically, the tokens $(A + B)$ are the pairs (a, b) of tokens such as $a \in$ tok$A)$ and $b \in$ tok(B).
2. The set typ$(A+B)$ is the disjoint union of typ(A) and typ(B). Concretely, the types of $A + B$ are pairs (i, α) where $i = 0$ and $\alpha \in$ typ(A) or else $i = 1$ and $\alpha \in$ typ(B).
3. The classification relation $\models_{A+B}$ of $A+B$ is defined by:

$$(a, b) \models_{A+B} (0, \alpha) \quad \text{iff} \quad a \models_A \alpha;$$

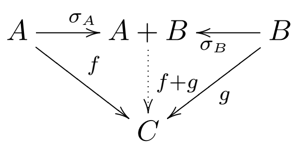
$$(a, b) \models_{A+B} (1, \beta) \quad \text{iff} \quad b \models_B \beta;$$

Definition 12.6. There are natural infomorphisms $\sigma_A : A \rightleftarrows A + B$ and $\sigma_B : B \rightleftarrows A + B$ defined as follows:

1. $\sigma_A(\alpha) = \langle 0, \alpha \rangle$ for each α in typ(A);
2. $\sigma_B(\beta) = \langle 1, \beta \rangle$ for each $\beta \in$typ(B);

3. For each pair $\langle a, b\rangle \in$ tok $(A + B), \sigma_A(\langle a, b\rangle) = a$ and $\sigma_B(\langle a, b\rangle) = b$.

Proposition 12.4 (Universal application for sums)**.** *Given the infomorphisms $f : A \rightleftarrows C$ and $g : B \rightleftarrows C$, there exists a unique informorphism $f + g$ such that the following diagram is commutative.*



The notions of sum and infomorphism can then be generalized to families of classifications without any problem (see, [3], 83).

12.5 Invariants and quotients

Definition 12.7 (Invariant)**.** Given a classification A, an *invariant* is a pair $I = (\Sigma, R)$ composed of a set $\Sigma \subseteq$ typ(A) of types of A and of a binary relation R between the tokens of A such that, if aRb, then, for each $\alpha \in \Sigma, a \models_A \alpha$ if and only if $b \models_A \alpha$.

Definition 12.8 (Quotient)**.** Let $I = (\Sigma, R)$ be an invariant of the classification A. The *quotient* of A by I, denoted by A/I, is the classification having types Σ, whose tokens are the R -classes of equivalence of tokens of A, with the condition:

$$[a]_R \models_{A/I} \alpha \quad \text{if and only if} \quad a \models_A \alpha.$$

We also introduce the following definitions:

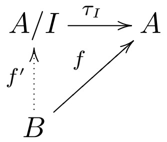
Definition 12.9. Given an invariant $I = (\Sigma, R)$ over A, the it canonical quotient infomorphism τ_I : $A/I \rightleftarrows A$ is the function inclusion defined on the types and who, on the tokens. applies each token of A to its equivalence class.

Definition 12.10. Given an invariant $I = (\Sigma, R)$ on A, an infomorphism $f : B \rightleftarrows A$ respects I if:

1. For each $\beta\ in$ typ(B), $f(\beta) \in \Sigma$.
2. If a_1Ra_2, then, $f(a_1) = f(a_2)$.

We can also formulate the following propositions and definitions:

Proposition 12.5. *Let I be an invariant of A. Given any infomorphism $f : B \rightleftarrows A$ which respects I, there exists a unique informorphism $f' : B \rightleftarrows A/I$ such that the following diagram is commutative:*



Definition 12.11. An informorphism $f : A \rightleftarrows B$ is said to be ”token-identical” if tok(A) = tok(B) and if $f^\vee$is the identity function on this set. Dually, f is said to be ”type-identical” if typ(A) = typ(B) and if f^ is, this time, the identity.

Proposition 12.6. *Given the classifications A and B, if A is separated, then there is at most one type-identical infomorphism from A to B.*

We can still show, as Barwise and Seligman do (see, [3], 87-88), that there exist dual notions of invariants and quotients, hence also a dual proposition to Proposition 11.6.

In conclusion, the immersion of the theory of classifications, via the notion of Chu space, in a categorical formalism, undoubtedly has the merit of unifying the formalisms ordinary used to describe classifications and the operations that can be done on them. In this sense, it presents a stylistic improvement in the presentation of the theory. However, basically, nothing is really fundamentally new, and category theory, as Barwise and Seligman apply it, tends instead to complicate and obscure the simpler operations (embedding of one classification into another one, sum of two (or more) classifications, etc.). The formalism does not allow, in particular, to identify new classes of morphisms which would make it possible to go further than what was proposed in [19] or [18]. It remains that the notion of ”Chu space” makes it possible to treat classifications of objects and classification of mathematical structures in the same way. It also shows that the ”metaclassification” (or set of Neuville ellipsoids) defined in [19] is in fact a Chu space.

13 A conjecture of convergence

As we may see, Fig. 5, representing what we have called ”metaclassification”, shows classes and sequences of classes that are partitions or classifications, overlapping or not, formal or fuzzy, and so on. This includes of course the pseudo-partitions and pseudo-classifications in the sense of Bruckner.

Now an observation : knowing that the aim of a partition or of a classification is always to locate an element in one or more than one class, if we have successive partitions or classifications, it means that all the sequences of classes contained in this graph are decreasing sequences. René Thom (see, [24]) have already suggested that a decreasing sequence of embedded open sets could converge on this minimal element which is the point.

So we can propose the following conjecture:

Conjecture 13.1. *The universe $\mathcal{U}(E)$ of all classifications possibly defined on a set E effectively presents itself as a decreasing sequence of open subsets converging on this minimal element which is the point, i.e., as it happens, the index of the metaclassification.*

It follows from the previous observations that $\mathcal{U}(E)$, equipped with variable distances d only defined by the fact that $d < \delta$ is not a complete metric space. So we cannot apply to this space the good theorems of topology like, for instance, the theorem of Baire.

Moreover, let T be the topology defined on $\mathcal{U}(E)$. We know that a sequence $(u_n)_{n\in\mathbb{N}} \in U^{\mathbb{N}}$ converges on $\ell \in U$ if, for every open set O of T containing the element ℓ, there exists a natural number N such that all the u_n, for $n \geq N$, belong to O. We know also that it would be sufficient that the space $\mathcal{U}(E)$ were a separable space to assert that this limit is unique. But $\mathcal{U}(E)$ is not a separable space and we do not know how to fix this N.

However, this conjecture - the convergence of the Neuville's ellipsoids - if it comes to be verified - would mean that we have in fact a proof of existence of a ”metaclassification”, that is, a proof of existence of a kind of ”classification of classifications”, even if we cannot be clearer about what it contains. Taking things back to front, this may also suggest that a point is something which is capable of generating a metaclassification.

Of course, this result would be a relatively poor one: because we have not very much information about this overlapping classification.

In particular, we do not know its laws of composition, we have no metric on it and we do not know the relations between the classifications it contains and how we could go from one to another: in fact, we lack a true algebra of classifications. But this existence would prove that one can think of such a structure, even if it is not actually accessible in details.

$\lvert E\rvert$	1	2	3	4	5	6	7	8
$\lvert\mathbf{P}(E)\rvert$	1	2	5	15	52	203	877	4140

Table 1: Number of partitions in $\mathbf{P}(E)$ according to the cardinal of E

Another result, connected to it, is the following one:

Proposition 13.2. *The set of all classifications on an infinite set generates the continuum.*

Proof. To see that, it is sufficient to remember that, from Cantor, the cardinality of [0,1] is the same than the cardinality of $\mathbb{R}$ or $\mathbb{R}^n$ and even of $\mathbb{R}^{\mathbb{N}}$. Now if we associate a pair (w_1, w_2) of elements of $\mathbb{C}^*$ to the semi-axes of any ellipses, we can prove (see, [23]) that the set of all the sequences of ellipses has the same cardinality than $\mathbb{R}^2$, i.e the same cardinality than [0,1]. □

14 Concrete problems: the search for an algebra of classifications

There are many modes of organization on a set and it is sometimes very difficult to choose between them.

As we know, partitions are very numerous (see table 1).

So it is not very easy to examine which classification is the best one among, say, several thousands of them. The situation is worse with weaker structures like covers or even minimal covers. Let's quickly explain what it is.

Recall that a family $\mathcal{F}$ of nonempty subsets of a set E, whose union contains the given set E (and which contains no duplicated subsets) is called a *cover* (or a *covering*) of E.

A particular kind of cover is the *minimal cover*. A minimal cover is a cover for which the removal of one member destroys the covering property.

Of course, we can make orderings on covers and build hierarchies of covers or minimal covers (see, [19]). But if the set $\mathbf{L}(E)$ of minimal covers is a lattice for the refinement relation, the set $\mathbf{R}(E)$ of all covers has no interesting properties : it is only a preorder (or a quasi-order) for the refinement relation (that we define in the same way as for the partition ordering). Moreover, computing the covers of a set leads immediately to big numbers (see table 2). So it becomes rapidly impossible to examine the very numerous possible chains of covers.

$\|E\|$	1	2	3	4
$\|\mathbf{R}(E)\|$	1	5	109	32297

Table 2: Number of covers in $\mathbf{R}(E)$ according to the cardinal of E

Worse again would be the situation if we examine the number of all possible fuzzy partitions, classifications or chains of covers that we can define on a set.

One way to escape would be to get some algebra of classifications, which can explain clearly the links between their structures. Such an algebra is difficult to get because classifications are commutative but nonassociative structures, and almost all mathematical structures are associative. But there are some candidates like, for instance, what we call "dendriform algebras"(for more information on them, see, [19]).

15 Conclusion

Let us resume the main point of these little introduction to the mathematics of classification. According to us, the whole set of classifications realizes a new construction of the continuum. There are already many constructions of the continuum in mathematics but this one does not need a particular hypothesis about the infinite. One just needs Cantor theorem about the bijection between [0,1] and $\mathbb{R}$ or, more generally, $\mathbb{R}^n$. We must hope now that the metaclassification and its construction of the continuum will be explored in the future by mathematicians, so that we get much more information about the set of ellipsoids or sequences of ellipsoids contained in it.

References

[1] Barr, M., **-Autonomous categories*, Springer-Verlag, Berlin, 1979.

[2] Barr, M., "The Chu Construction: history of an idea", *Theory and Applications of Categories*, Vol. 17, No. 1, 10?16, 2006.

[3] Barwise, J. and Seligman J., *Information Flow, the Logic of Distributed Systems*, Cambridge University Press, Cambridge, 1997, 2008.

[4] Bellacicco, A., "Fuzzy Classifications", *Synthese*, Vol. 33, No. 2/4, 273-281, Sept. 1976.

[5] Benzécri, J.-P., et alii., *L'Analyse des données*, tome 2, correspondances, Dunod, Paris, 1973.

[6] Berge, C., *Graphes et hypergraphes*, Dunod, Paris, 1970.

[7] Brucker, F., Barthélemy, J.-P., Eléments de classifications, Hermès-Lavoisier, Paris, 2007.

[8] Certaine, J., "The ternary operation $(abc) = ab^{-1}c$ of a group, *Bull. Amer. Math. Soc.* 49, 869-877, 1943.

[9] Girard, J.-Y., "Linear logic", *Theoretical Computer Science*, 50, 1-102, 1987.

[10] Gondran, M., Minoux, M., *Graphes et Algorithmes*, Eyrolles, Paris, 1979.

[11] Jambu, M., *Classification automatique pour l'analyse des données*, 2 tomes, Bordas-Dunod, Paris, 1978.

[12] Jardine, N., Sibson, R., "A model for taxonomy", *Mathematical Bio-sciences* 2 (1968), 465-482.

[13] Kechris, A. S., *Actions of Polish groups and classification problems*, Analysis and Logic, London Math. Soc. Lecture Note Series, Cambridge, Cambridge University Press, 2001.

[14] Kleinberg, J., "An impossibility theorem for Clustering", *Advances in Neural Information Processing Systems (NIPS)*, Proceedings of the 2002 Conference held in British Columbia, Canada, 15, 2002, 529-536.

[15] Lerman, I., *Les bases de la classification automatique*, Gauthier-Villars, Paris, 1970.

[16] Lerman, I., *Classification automatique et analyse ordinale des données*, Dunod, Paris, 1981.

[17] Paiva (de), V., "A dialectica-like model of linear logic", *Proc. Conf. on Category Theory and Computer Science*, Springer-Verlag Lecture Notes in Computer Science, 389, 341-356, Manchester, September 1989.

[18] Parrochia, D., "Mathematical Theory of Classification", *Knowledge Organization* 45(2):184-201, January 2018.

[19] Parrochia D, Neuville, P., *Towards a general theory of classifications*, Birkhäuser, Basel, 2013.

[20] Pratt, V. R., "The Stone gamut: A coordinatization of mathematics", *Proc. 10th Annual IEEE Symp. on Logic in Computer Science*, 444-454, Montreal, 1995.

[21] Rasiowa, H., Sikorski, R., *The Mathematics of Metamathematics*, Paṅstwowe Wydawnictwo Naukowe, Varsovie, 1963, 3rd ed., 1970.

[22] Rasiowa, H., *An Algebraic Approach to Non-Classical Logics*, North Holland, Amsterdam, London, 1974.

[23] Serre, J.-P., *Course in Arithmetics*, Springer-Verlag, Berlin, 1973.

[24] Thom, R., ”Une définition continue du Nombre”, *Les rencontres physiciens-mathématiciens de Strasbourg - RCP25*, ” Conférences de M. Chaperon, A. Chenciner, R. Lozi, J. Martinet et J.P. Ramis, P. Moussa, R. Moussu, F. Pham, R. Thom, exp. no 6, p. 163-189, tome 41, 1990.

[25] Zadeh, L. A., ”Fuzzy sets”, *Information and Control*, 8, 338-353, 1965.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

Daniel Parrochia is the sole author of this article and all that it contains.

Conflict of interest

The author declares no conflict of interest.

Sources of funding for research presented in a scientific article or scientific article itself

No source of funding.