Topological forms of information, Daniel Bennequin, 15/09/2014

This short note summarizes results obtained in the last years with Pierre Baudot (Max Planck, Leipzig). The last part, on optimal discrimination, comes from a joint research of P.B. and me with Guillaume Marrelec (LIF, Pitié Salpétrière, Paris). However the present author has the entire responsibility of the following summary. We thank MaxEnt14 for the opportunity to present these researches to the information science community.

Introduction

What is information ? is a question that has received several answers according to the different problems investigated ; the best known definition was given by Shannon [S], using random variables and a probability law, for the problem of optimal message compression ; but the first definition was given by Fisher, as a metric associated to a smooth family of probability laws, for optimal discrimination by statistical tests ; it is a limit of the Kullback-Leibler divergence, which was introduced to estimate the accuracy of a statistical model of empirical data. More generally Kolmogorov considered that the concept of information is the heart of probability theory. However, Evariste Galois saw the application of group's theory for discriminating solutions of an algebraic equation as a first step toward a general theory of ambiguity, that was developed further by Riemann, Picard, Vessiot, Lie, Poincaré, Cartan, for systems of differential equations ; it is also a theory of information. In another direction René Thom claimed that information must have a topological content (see [T]) ; he gave the example of the unfolding of the coupling of two dynamical systems, but he had in mind the whole domain of algebraic or differential topology. These approaches have in common to define secondary objects, either functions, groups or homology cycles, for measuring in what sense a couple departs from independency. For instance, in the case of Shannon the mutual information is $I(X;Y) = H(X) + H(Y) - H(X,Y)$, and for Galois it is the quotient set $IGal(L_1; L_2|K) = (Gal(L_1|K) \times Gal(L_2|K))/Gal(L|K)$, where L_1, L_2 are two fields containing a field K in an algebraic closure Ω of K, and L is the field generated by L_1 and L_2 in Ω.

We suggest that all information quantities are of co-homological nature, in a setting which depends of a pair of categories (cf. [$McL1, 2$]), one for the data on a system, like random variables or functions of solutions of an equation, and one for the parameters of this system, like probability laws or coefficients of equations ; the first one generates an algebraic structure like a monoid, or more generally a monad, and the second one generates a representation of this structure, as do for instance conditioning, or adding new numbers ; then

information quantities are co-cycles associated to this module.
We will see that, given a set of random variables on a finite set Ω and a simplicial subset of probabilities on Ω, the *entropy* appears as the only one universal co-homology class of degree one. The higher mutual information functions that were defined by Shannon are co-cycles (or twisted co-cycles for even orders), and they correspond to higher homotopical constructions. In fact this description is equivalent to the theorem of Hu Kuo Ting $[HKT]$, that gave a set theoretical interpretation of the mutual information decomposition of the total entropy of a system. Then we can use information co-cycles to describe forms of the information distribution between a set of random data ; figures like links, chains and borromean links appear in this context, giving rise to a new kind of topology.

Information homology

Here we consider random variables (say r.v) on a finite set Ω as *congruent* when they define the same partition ; the join r.v YZ corresponds to the less fine partition that is finer than Y and Z. This defines a monoid structure on the set $\Pi(\Omega)$ of partitions of Ω, with 1 as a unit, and where each element is idempotent, i.e. $\forall X, XX = X$. An *information category* is a set $\mathcal{S}$ of r.v such that, for any $Y, Z \in \mathcal{S}$ less fine than $U \in \mathcal{S}$, the join YZ belongs to $\mathcal{S}$, cf. [1]. An ordering on $\mathcal{S}$ is given by $Y \leq Z$ when Z refines Y, which also defines the morphisms $Y \to Z$ in $\mathcal{S}$. In what follows we always assume that 1 belongs to $\mathcal{S}$. The simplex $\Delta(\Omega)$ parameterizes all probability laws on Ω ; we choose a simplicial sub-complex $\mathcal{P}$ in $\Delta(\Omega)$, that is stable by all conditioning operations by elements of $\mathcal{S}$. By definition, for $N \in \mathbb{N}$, an information N-cochain is a family of measurable numerical functions of $P \in \mathcal{P}$ indexed by the sequences $(S_1; ...; S_N)$ in $\mathcal{S}$ majored by an element of $\mathcal{S}$, whose values depend only of the image law $(S_1, ..., S_N)_* P$. This condition is natural from a topos point of view, cf. $[McL1], [1]$. For $N = 0$ this gives only the constants. We denote by $\mathcal{C}^N$ the vector space of N-cochains. The following formula corresponds to the *averaged conditioning* of Shannon $[S]$:

$$S_0.F(S_1; ...; S_N; \mathbb{P}) = \sum \mathbb{P}(S_0 = v_j) F(S_1; ...; S_N; \mathbb{P}|S_0 = v_j), \qquad (1)$$

where the sum is taken over all values of S_0, and the vertical bar is ordinary conditioning. It satisfies the associativity condition $(S'_0 S_0).F = S'_0.(S_0.F)$. The coboundary operator δ is defined by

$$\delta F(S_0; ...; S_N; \mathbb{P}) = S_0.F(S_1; ...; S_N; \mathbb{P}) + \sum_0^{N-1} (-1)^i F(...; (S_i, S_{i+1}); ...; S_N; \mathbb{P}) + (-1)^N F(S_0; ...; S_{N-1}; \mathbb{P}), \quad (2)$$

It corresponds to a standard non-homogeneous bar complex (cf. [McL2]). Another co-boundary operator on $\mathcal{C}^N$ is δ_t (t for twisted or trivial action or topological complex) that is defined by the above formula with the first term $S_0.F(S_1;...;S_N;\mathbb{P})$ replaced by $F(S_1;...;S_N;\mathbb{P})$. The corresponding co-cycles are defined by the equations $\delta F = 0$ or $\delta_t F = 0$ respectiveley. We easily verify that $\delta \circ \delta = 0$ and $\delta_t \circ \delta_t = 0$; co-homology $H^*(\mathcal{S};\mathcal{P})$ resp. $H_t^*(\mathcal{S};\mathcal{P})$ is defined by taking co-cycles modulo the elements of the image of δ resp. δ_t, named co-boundaries. The fact that classical entropy $H(X;\mathbb{P}) = -\sum_i p_i \log_2 p_i$ is a 1-co-cycle is the fundamental equation $H(X,Y) = H(X) + X.H(Y)$.

Theorem 1 (cf. [1]) for the full simplex $\Delta(\Omega)$, and every structure $\mathcal{S}$, if Ω has at least four elements and if $\mathcal{S}$ can induce all partitions of these four elements, the information co-homology group of degree one is one-dimensional and generated by the classical entropy.

Problem 1 : compute the homology of higher degrees. We conjecture that for binary variables it is zero, but that in general non-trivial classes appear, deduced from polylogarithms. This could require to connect with the works of Dupont, Bloch, Goncharov, Elbaz-Vincent, Gangl et al. on motives (cf. $[EVG]$), that started from the discovery of Cathelineau (1988) that entropy appears in the computation of the degree one homology of the discrete group SL_2 over $\mathbb{C}$ with coefficients in the adjoint action (cf. $[C]$).

Suppose $\mathcal{S}$ is the monoid $\Pi(\Omega)$ of all partitions. The higher mutual informations were defined by Shannon as alternating sums :

$$I_N(S_1;...;S_N;\mathbb{P}) = \sum_{k=1}^{k=N} (-1)^{k-1} \sum_{I \subset [N]; card(I)=k} H(S_I;\mathbb{P}), \tag{3}$$

where S_I denotes the join of the S_i such that $i \in I$. We have $I_1 = H$ and $I_2 = I$ is the usual mutual information : $I(S;T) = H(S) + H(T) - H(S,T)$.

Theorem 2 (cf. [1]) $\quad I_{2m} = \delta_t \delta \delta_t ... \delta \delta_t H, \quad I_{2m+1} = -\delta \delta_t \delta \delta_t ... \delta \delta_t H.$

Thus odd information quantities are information co-cycles, because they are in the image of δ, and even information quantities are topological co-cycles, because they are in the image of δ_t.

In [1] we show that this description is equivalent to the theorem of Hu Kuo Ting (1962) $[HKT]$, giving a set theoretical interpretation of the mutual information decomposition of the total entropy of a system : mutual information, join and averaged conditioning correspond respectively to intersection, union and difference $A \backslash B = A \cap B^c$. In special cases we can interpret I_N as homotopical algebraic invariants ; for instance for $N = 3$, suppose that $I(X;Y) = I(Y;Z) = I(Z;X) = 0$, then $I_3(X;Y;Z) = -I((X,Y);Z)$ can be defined as a Milnor and Massey invariant for links, as presented in $[LV]$ (cf. page 284), through the 3-ary obstruction to associativity of products in

a subcomplex of a differential algebra, cf. [1]. The absolute minima of I_3 correspond to Borromean links, interpreted as synergy, cf. $[M]$.

Extension to quantum information

Positive hermitian $n \times n$-matrices ρ, normalized by $Tr(\rho) = 1$, are named density of states and are considered as quantum probabilities on $E = \mathbb{C}^n$. Real quantum observables are $n \times n$-matrices hermitian matrices Z, and, by definition, the amplitude, or expectation, of the observable Z in the state ρ is given by the formula $\mathbb{E}(Z) = Tr(Z\rho)$. See for instance $[NC]$. Two real observables Y, Z are *congruent* if their eigenspaces are the same, thus orthogonal decomposition of E are quantum analogs of partitions. The join is well defined for commuting observables. An information structure $\mathcal{S}_Q$ is given by a subset of observables, such that if Y, Z have common refined eigenspaces decomposition in $\mathcal{S}_Q$, their join (Y, Z) belongs to $\mathcal{S}_Q$. We assume that $\{E\}$ belongs to $\mathcal{S}_Q$. We define information N-cochains as for the classical case.
The image of a density ρ by an observable Y is $\rho_Y = \sum_A E_A^* \rho E_A$, where the E_A's are the spectral projectors of the observable Y.
The action of a variable on the cochains space $\mathcal{C}_Q^*$ is given by the *quantum averaged conditioning* :

$$Y.F(Y_0; ...; Y_m; \rho) = \sum_A Tr(E_A^* \rho E_A) F(Y_0; ...; Y_m; E_A^* \rho E_A) \tag{4}$$

From here we define coboundary operators δ_Q and δ_{Qt} by the formula (2), then notions of co-cycles, co-boundaries and co-homology classes follow. We have $\delta_Q \circ \delta_Q = 0$ and $\delta_{Qt} \circ \delta_{Qt} = 0$; cf. [1].
The *Von-Neumann entropy* of ρ is $S(\rho) = \mathbb{E}_\rho(-\log_2(\rho)) = -Tr(\rho \log_2(\rho))$, then the entropy of Y in state ρ is $S(Y; \rho) = S(\rho_Y)$, and the classical entropy is $H(Y; \rho) = -\sum_A Tr(E_A^* \rho E_A) \log_2(Tr(E_A^* \rho E_A))$. It is well known that $S((X, Y); \rho) = H(X; \rho) + X.S(Y; \rho)$ when X, Y commute, cf. $[NC]$. In particular, by taking $Y = 1_E$ we see that classical entropy measures the default of equivariance of the quantum entropy, i.e. $H(X; \rho) = S(X; \rho) - (X.S)(\rho)$. Then, if we define the *reduced quantum entropy* by $s(X; \rho) = S(X; \rho) - S(\rho)$, we get a 1-cocycle of quantum information. In fact H is also a 1-cocycle and it is co-homologous to s by the following lemma : $\delta_Q(S) = s - H$.
Theorem 3 (cf. [1]) : as soon as $n \geq 4$ and $\mathcal{S}_Q$ can induce all orthogonal decompositions less fine than the canonical basis on a four dimensional subspace, s or H generates the 1 dimensional co-homology of δ_Q.

Concavity and convexity properties of information quantities

The simplest classical information structure $\mathcal{S}$ is the monoid generated by a family of "elementary" binary variables $S_1, ..., S_n$, it is remarkable that in

this case, the information functions $I_{N,J} = I_N(S_{j_1}; ...S_{j_N})$ over all the subsets $J = \{j_1, ..., j_N\}$ of $[n] = \{1, ..., n\}$, different from $[n]$ itself, give algebraically independent functions on the probability simplex $\Delta(\Omega)$ of dimension $2^n - 1$. They form coordinate on the quotient of $\Delta(\Omega)$ by a finite group.
Let $\mathcal{L}_d$ denotes the Lie derivative with respect to $d = (1, ..., 1)$ in the vector space $\mathbb{R}^{2^n}$, and $\triangle$ the Euclidian Laplace operator on $\mathbb{R}^{2^n}$, then $\Delta = \triangle - 2^{-n}\mathcal{L}_d \circ \mathcal{L}_d$ is the Laplace operator on the simplex $\Delta(\Omega)$ defined by equating the sum of coordinates to 1.
Theorem 4 (cf. [2]) on the affine simplex $\Delta(\Omega)$ the functions $I_{N,J}$ with N odd (resp. even) satisfies the inequality $\Delta I_N \geq 0$ (resp. $\Delta I_N \leq 0$).
In other terms, for N odd the $I_{N,J}$ are super-harmonic which is a kind of weak concavity and for N even they are sub-harmonic which is a kind of weak convexity. In particular when N is even (resp. odd) $I_{N,J}$ has no local maximum (resp. minimum) in the interior of $\Delta(\Omega)$.
Problem 2 : What can be said of the other critical points of $I_{N,J}$? What can be said of the restriction of one information function on the intersection of levels of other information functions ? Information topology depends on the shape of these intersections and on the Morse theory for them.

Monadic cohomology of information

Now we consider the category $\mathcal{S}_*$ of *ordered partitions* of Ω over $\mathcal{S}$, i.e. pairs (π, ω) where $\pi \in \mathcal{S}$ and ω is a bijection from $\{1, ..., l(\omega)\}$ with Ω/π, where $l(\omega)$ is the length of π, i.e. the number of pieces of Ω given by π. The indices of these pieces are the values of the r.v associated to (π, ω). A *rooted tree decorated* by $\mathcal{S}_*$ is an oriented finite tree Γ, with a marked initial vertex s_0, named the *root* of Γ, where each vertex s is equipped with an element F_s of $\mathcal{S}_*$, such that edges issued from s correspond to the values of F_s.
The notation $\mu(m; n_1, ..., n_m)$ denotes the operation which associates to an ordered partition (π, ω) of length m and m ordered partitions (π_i, ω_i) of respective lengths n_i, the ordered partition that is obtained by cutting the pieces of π using the π_i and respecting the order. An evident unit element for this operation is π_0. The symbol μ_m denotes the collection of those operations for m fixed. Be careful that in general the result of $\mu(m; n_1, ..., n_m)$ is not a partition of length $n_1 + ... + n_m$, thus the μ_m do not define what is named an *operad* ; cf. [LV], [F]. However they allow the definition of a filtered version of operad, with unity, associativity and covariance for permutations, cf. [3]. We apply the Schur construction (cf. [F])to the μ_m to get a monad (cf. [$McL1$], [F]) : take for V is the real vector space freely generated by $\mathcal{S}_*$; it is graded by the partition length as direct sum of spaces $V(m)$. As for ordinary operads we introduce $\mathcal{V} = \bigoplus_{m\geq 0} V(m) \otimes_{\mathfrak{S}_m} V^{\otimes m}$; Schur composition is defined by $\mathcal{V} \circ \mathcal{V} = \bigoplus_{m\geq 0} V(m) \otimes_{\mathfrak{S}_m} \mathcal{V}^{\otimes m}$. It is easy to verify that the collection

$(\mu_m; m \in \mathbb{N})$ defines a linear map $\mu : \mathcal{V} \circ \mathcal{V} \to \mathcal{V}$, and the trivial partition π_0 defines a linear map $\epsilon : \mathbb{R} \to \mathcal{V}$, that satisfied to the axioms of a monad, i.e. $\mu \circ (Id \circ \mu) = \mu \circ (\mu \circ Id), \quad \mu \circ (Id \circ \epsilon) = Id = \epsilon \circ (\eta \circ Id)$.

Let $\mathcal{F}$ be the vector space of real measurable functions on the set $\mathcal{P}$ of probability laws, considered as a $\mathcal{S}_*$ module of pure degree 1, in such a manner that $\mathcal{F} \circ \mathcal{V}^{\circ m}$ coincides with $\mathcal{F} \otimes \mathcal{V}^{\circ m}$. For a r.v S of length m and m decorated trees $(S_1^s; S_2^s; ...; S_k^s); 1 \leq s \leq m$ of level k, we pose

$$F_S(S_1; S_2; ..., S_k; \mathbb{P}) = \sum_s \mathbb{P}(S = s) F(S_1^s; S_2^s; ..., S_k^s; \mathbb{P}|(S = s)); \qquad (5)$$

this is a function of the decorated tree $(S; S_1; S_2; ...; S_k)$ of level $k+1$ that roots in S and S_i^s is placed at the end of the edge $S = v$. This formula extending (1) defines a map $\theta : \mathcal{F} \circ \mathcal{V} \to \mathcal{F}$, that is an action to the right in the sense of monads, i.e. $\theta \circ (Id \circ \mu) = \theta \circ (\theta \circ Id); \quad \theta \circ (Id \circ \epsilon) = Id$.
Say that $F(S; S_1; S_2; ..., S_{k-1}; \mathbb{P})$ is *local* if its value depends only of the images of $\mathbb{P}$ by the join of the decorating variables of the corresponding tree. Then copy the formalism of Beck (see $[F]$) with this locality condition, to get monadic information co-homology : a cochain of degree k is an element of $\mathcal{F} \circ \mathcal{V}^{\circ k}$ whose components are local ; the operator δ comes from the simplicial structure associated to θ and μ :

$$\delta F(S; S_1; ...; S_k; \mathbb{P}) = F_S(S_1; ...; S_k; \mathbb{P}) + (-1)^{k+1} F(S; ...; S_{k-1}; \mathbb{P}) + \sum_{i=1}^{i=k} (-1)^i F(S; ...; \mu(S_{i-1} \circ S_i); S_{i+1}; ...; S_k; \mathbb{P}) \qquad (6)$$

This gives co-homology groups $H_\tau^*(\mathcal{S}, \mathcal{P})$, τ for tree. The fact that entropy $H(S_*\mathbb{P}) = H(S; \mathbb{P})$ defines a 1-cocycle expresses an equation of Fadeev, generalized by Baez, Fritz and Leinster $[BLF]$, who gave another interpretation, based on true operad structure over the set of all finite probability laws.
Theorem 5 (cf. [3]) : as soon as Ω has more than four points, $H_\tau^1(\Pi(\Omega), \Delta(\Omega))$ is the one dimensional vector space generated by the entropy.
Another right action of $\mathcal{V}$ on $\mathcal{F}$ is given by (5) where on the right side $\mathbb{P}|(S = s)$ is replaced by $\mathbb{P}$ itself. From here and the simplicial structure associated to θ and μ, we define an operator δ_t, then a twisted information co-homology. This allows us to define higher information quantities for strategies : for $N = 2M + 1$ odd $I_{\tau,N} = -(\delta\delta_t)^M H$ and for $N = 2M + 2$ even $I_{\tau,N} = \delta_t(\delta\delta_t)^M H$.
This gives for $N = 2$ a notion of mutual information between a variable S of length m and a collection T of m variables $T_1, ..., T_m$:

$$I_\tau(S; T; \mathbb{P}) = \sum_{i=1}^{i=m} \mathbb{P}(S = i)(H(T_i; \mathbb{P}) - H(T_i; \mathbb{P}|S = i)). \qquad (7)$$

When all the T_i are equals we recover the ordinary mutual information of Shannon.

The forms of information strategies

A rooted tree Γ decorated by $\mathcal{S}_*$ can be seen as a strategy to discriminate between points in Ω. For each vertex s there is a minimal set of chained edges $\alpha_1, ..., \alpha_k$ connecting s_0 to s ; the cardinal k is named the *level* of s ; this chain defines a sequence $(F_0, v_0; F_1, v_1; ...; F_{k-1}, v_{k-1})$ of observables and values of them ; then we can associate to s the subset Ω_s of Ω where each F_j takes the value v_j. At a given level k the sets Ω_s form a partition π_k of Ω ; the first one π_0 is the unit partition of length 1, and π_l is finer than π_{l-1} for any l. By recurrence over k it is easy to deduce from the orderings of the values of F_s an embedding in the Euclidian plane of the subtrees $\Gamma(k)$ at level k such that the values of the variables issued from each vertex are oriented in the direct trigonometric sense, thus π_k has a canonical ordering ω_k. Remark that many branches of the tree gives the empty set for Ω_s after some level ; we name them dead branches. It is easy to prove that the set $\Pi(\mathcal{S})_*$ of ordered partitions that can be obtained as a (π_k, ω_k) for some tree Γ and some level k is closed by the natural ordered join operation, then as it contains π_0 it forms a monoid, which contains the monoid $M(\mathcal{S}_*)$ generated by $\mathcal{S}_*$.
Complete discrimination of Ω by $\mathcal{S}_*$ exists when the final partition of Ω by singletons is attainable as a π_k ; optimal discrimination correspond to minimal level k. When the set Ω is a subset of the set of words $x_1, ..., x_N$ with letters x_i belonging to given sets M_i of respective cardinalities m_i, the problem of optimal discrimination by observation strategies Γ decorated by $\mathcal{S}_*$ is equivalent to a problem of minimal rewriting by words of type $(F_0, v_0), (F_1, v_1), ..., (F_k, v_k)$; it is a variant of optimal coding, where the alphabet is given. The topology of the poset of discriminant strategies can be computed in terms of the free Lie algebra on Ω, cf. [F].
Probabilities $\mathbb{P}$ in $\mathcal{P}$ correspond to *a priori* knowledge on Ω. In many problems $\mathcal{P}$ is reduced to one element, that is the uniform law. Let s be a vertex in a strategic tree Γ, the set $\mathcal{P}_s$ of probability laws that are obtained by conditioning through the equations $F_i = v_i; i = 0, ..., k-1$ of a minimal chain leading from s_0 to s measures the evolution of knowledge when applying the strategy. The entropy $H(F; \mathbb{P}_s)$ for F in $\mathcal{S}_*$ and $\mathbb{P}_s$ in $\mathcal{P}_s$ gives a measure of information we hope when applying F at s in the state $\mathbb{P}_s$. The maximum entropy algorithm consists in choosing at each vertex s a variable that has the maximal conditioned entropy $H(F; \mathbb{P}_s)$.
Theorem 6 (cf. [3]) : to find one false piece of different weight among N pieces for $N \geq 3$, when knowing the false piece is unique, by the minimal numbers of weighing, the maximal entropy algorithm successes.

However we have another measure of information of the resting ambiguity at s, by taking for *Galois group* G_s the set of permutations of Ω_s that respect globally the set $\mathcal{P}_s$ and the set of restrictions of elements of $\mathcal{S}_*$ to Ω_s, and that preserve one by one the equations $F_i = v_i$. Along branches of Γ this gives a decreasing sequence of groups, whose successive quotients measure the evolution of acquired information in an algebraic sense.

Problem 3 : Generalize Th.6. Can we use algorithms based on the Galoisian measure of information ? Can we use higher information quantities associated to trees for optimal discrimination ?

Bibliography

[1], [2], [3] designate three preprints : P. Baudot, D.Bennequin, Information Topology, I, II, III. Also a Springer book is in preparation.

[BFL], J.B. Baez, T. Fritz, T. Leinster, A Characterization of Entropy in Terms of Information Loss, *Entropy* 13 (2011) pp 1945-57

[B] D. Bennequin, Information et dualité, in *Complexité-simplexité*, éditeurs A. Berthoz, J.L. Petit, OpenEdition, Collège-de-France, (2014).

[C] J.L. Cathelineau, Sur l'homologie de SL2 à coefficients dans l'action adjointe, *Math. Scand.* 63 (1988), pp. 51-86.

[CT] T. Cover, J. Thomas, Elements of Information Theory, Wiley (2006).

[F] B.Fresse, Koszul duality of operads and homology of partitionn posets, *Contemp. Math.* 346, Amer. Math. Soc. (2004), pp. 115-215

[EVG] P. Elbaz-Vincent, H. Gangl, On Poly(ana)logs I, *Compositio Mathematica* 130 (2002) pp 161-210

[HKT] Hu Kuo Ting, On the amount of Information, *Theory Probab. Appl.* 7-4 (1962), pp. 439-447

[K] A.Y. Khinchin, *Mathematical Foundations of Information Theory*, Dover (1957).

[LV] J-L. Loday, B. Valette, *Algebraic operads*, Springer (2012)

[$McL1$] S. Mac Lane, *Categories for the working mathematician*, Graduate Texts in Mathematics 5, Springer, 2nd Ed. (1997).

[$McL2$] S. Mac Lane, *Homology*, Classic in Mathematics, Springer, Reprint of the 1975 edition.

[M] H. Matsuda, Physical nature of higher order mutual information : Intrinsic correlations and frustration, *Phys. Review E*, 62 (2000), pp 3096-3102.

[NC] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge Series on Information and the Natural Sciences (2000).

[T] R. Thom, *Stabilité struturelle et morphogénèse*, deuxième édition, Inter-Edition, Paris (1977).