Workshop on Rigidity and Symmetry

Fields Institute, Toronto, October 17-21, 2011

Chirality and Symmetry Measures: some Open Problems

Michel Petitjean

MTi, UMR-S 973, INSERM, University Paris 7

http://petitjeanmichel.free.fr/itoweb.petitjean.html

SYMMETRY MEASURES DEFINITION

Step 1.
Space of objects: $E = \{X, X \text{ is an object }\}$
Define a relation of equivalence "=" over E.

Step 2.
Space of transformations (operators): $\Theta = \{T, TX \in E\}$
Composition of transformations: $\Theta \otimes \Theta \mapsto \Theta$
Θ is a group for this product. This group is not assumed to be Abelian.
I is the neutral element of the group. The inverse of T is T^{-1}.

The group Θ is acting on E:
- Associativity (not the one of the group): $T_1(T_2X) = (T_1T_2)X$
- Identity transformation: $IX = X$

Symmetry definition: An object $X \in E$ is symmetric when it exists $T \neq I$ such that $TX = X$.

Immediate properties:
(a) A symmetric object X is such that $T^kX = X$ for any $k \in \mathbb{Z}$.
(b) When X is symmetric for the transformations T_1 and T_2, thus $T_1T_2X = T_2T_1X = X$.

Remark:
The full definition of symmetry let us to discover naturally five groups rather than setting one arbitrarily.
(M. Petitjean. *A Definition of Symmetry*. Symm. Cult. Sci., **2007**, 18[2-3], 99-119.)

SYMMETRY MEASURES DEFINITION

Step 3.
Metrisation of E: we define a distance d between objects.

Characteristic property: X is symmetric if and only if it exists $T \neq I$ such that $d(TX, X) = 0$.

Step 4.
We consider the quantity $S(X) = Inf_{\{T \in \Theta, T \neq I\}} d(TX, X)$. Immediate property: X symmetric $\Rightarrow S(X) = 0$.

Symmetry measure definition: $S(X)$ is called a symmetry measure of X.

Quasi-symmetry: S(X) "small", but compared to what ?
When possible, we normalize $S(X)$.

Reciprocity ? $S(X) = 0 \Rightarrow X$ symmetric

Reciprocity FALSE, in general: we can get $S(X) = 0$ even when there is no $T \neq I$ such that $TX = X$.

Reciprocity TRUE, e.g. when $card(\Theta)$ is finite.

Other example: we look for the infinum over an adequate subset of Θ (case of chirality, i.e. indirect symmetry)

EXAMPLE OF THE FINITE SETS. NOTATIONS

Set of n points $x_1,...,x_n$ in R^d, $n > 1$, $d > 1$.

For clarity, we assume w.l.o.g., that the finite set is centered: $\sum_{i=1}^{n} x_i = 0$.

X : matrix of the coordinates of the n points (n lines, d columns).

The object associated to X is the class of equivalence of the matrices deduced from X by lines permutations.

P : matrix of permutation of order n.

t : translation vector in R^d.

R : rotation matrix in R^d.

Q : orthogonal matrix in R^d associated to an odd number of mirror inversions $(det(Q) = -1)$.

I : identity matrix (the dimension will be clear from the context).

$\mathbf{1}$: vector of which all components are equal to 1 (the dimension will be clear from the context).

Transposed matrices and vectors: X', P', t', R', Q', $\mathbf{1}'$ etc...

T : inertia of the distribution of the n points, i.e. trace of the covariance matrix.

$T = Tr(X'X)/n$, and should not be confused with a transformation.

We assume that $T \neq 0$.

ATTEMPT TO QUANTIFY DIRECT SYMMETRY: THE DIRECT SYMMETRY INDEX

Direct symmetry index: $DSI \overset{def}{=} \frac{1}{2T} Min_{\{P \neq I, R, t\}} D^2, \quad D^2 = Tr(X-Y)'(X-Y)/n, \quad Y = P(X - \mathbf{1}.t')R'$

The set of transformations is the set of direct isometries $E^+(d)$.

$Tr(X'Y)/n$ is a scalar product inducing a matrix norm (the Frobenius norm, up to the factor $\sqrt{n}$), then D is a distance between the matrices X and Y.

- All $n! - 1$ permutations P are allowed (but constraints due to "colors" can be set: see further).
- The optimal translation is $t = 0$. The optimal rotation is known for $d \leq 3$ (see further).
- DSI takes values in $[0..1]$, and is insensitive to scaling and mirror inversions.
- "Symmetric set" $\Rightarrow DSI = 0$.

Additional assumtion: "*distinct points*": $\forall i \in \{1..n\}, \forall j \in \{1..n\}, i \neq j \Rightarrow x_i \neq x_j$

- Reciprocity TRUE: $DSI = 0 \Rightarrow$ "symmetric set", but what about two arbitrarily close points ?

Relaxing the condition "*distinct points*" would have lead to $DSI = 0$
(reached for $R = I$, P permuting two identical points)

In this situation, even adding the condition $R \neq I$ does not suffice: $Inf_{\{P \neq I, R \neq I, t\}} D^2 = 0$
because we can set P permuting two identical points, and R arbitrarily close to I.

We fail to extend the direct symmetry index to continuous distributions!

I call any geometrical figure, or group of points, *chiral*, and say that it has *chirality* if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself.

Lord Kelvin, 1904

Remark: despite what is usually believed, the orientability of space is **NOT** required to define chirality.
(M. Petitjean. *Chirality in Metric Spaces*. Symm. Cult. Sci., **2010**, 21[1-3], 27-36.)

Chiral index: $CHI \overset{def}{=} \frac{d}{4T} Min_{\{P,R,t\}} D^2$, $\quad D^2 = Tr(X-Y)'(X-Y)/n$, $\quad Y = P(X - \mathbf{1}.t')Q'R'$

- The inversion Q is constant and arbitrary
- All $n!$ permutations P are allowed (but constraints due to "colors" can be set: see further).
- The optimal translation is $t = 0$. The optimal rotation is known for $d \leq 3$ (see further).
- CHI takes values in $[0..1]$, and is insensitive to scaling and mirror inversions.
- Achiral set $\Rightarrow CHI = 0$.
- Reciprocity TRUE: $CHI = 0 \Rightarrow$ achiral set.

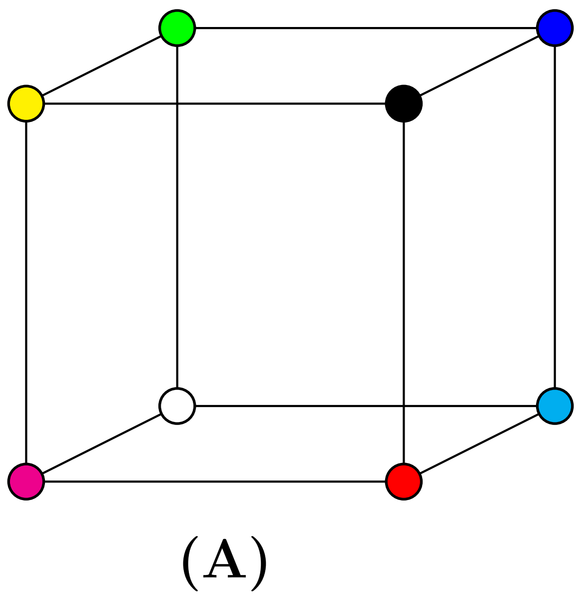
The "*distinct points*" assumption is no more needed.

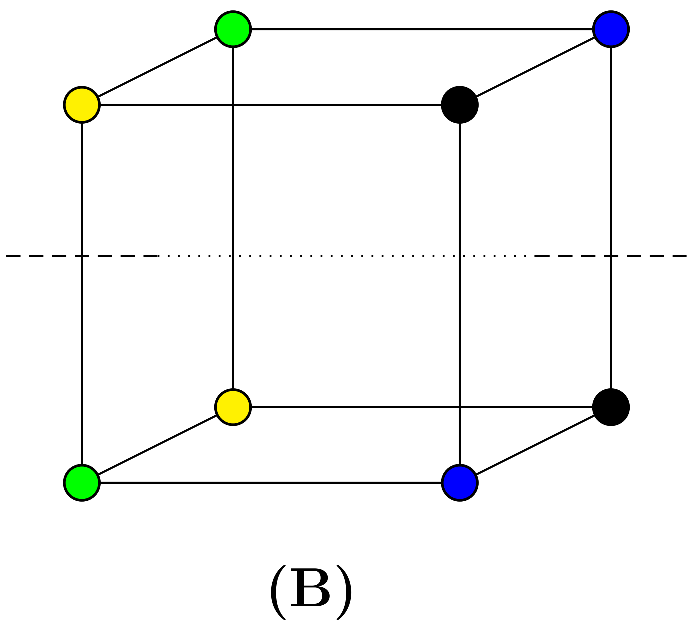
Clearly, restricting the transformations to indirect isometries solved the problem we encountered in step 4 of our definition of symmetry measures.

We are able to extend the chiral index to continuous distributions!

Before that, we need to handle the "pairwise correspondence" constraint.

EUCLIDEAN SYMMETRY AND CHIRALITY
ARE NOT STRICTLY GEOMETRIC CONCEPTS





The cube (A) is NOT symmetric; the cube (B) has a C2 rotation axis; both are CHIRAL

We do need a mechanism to associate points.

We are going to build it in a general framework,
not necessarily involving symmetry or chirality.

WASSERSTEIN METRIC

Random vectors X et Y (values in R^d), respective distributions $\mathcal{P}_x$ and $\mathcal{P}_y$ (no symmetry here).
W: joint distribution of (X, Y) with fixed marginals $\mathcal{P}_x$ and $\mathcal{P}_y$.
$\{W\}$: set of all joint distributions W.

L^2 Wasserstein distance: $D^2(\mathcal{P}_x, \mathcal{P}_y) = Inf_{\{W\}} E(X - Y)'(X - Y)$

<u>Case where both $\mathcal{P}_x$ and $\mathcal{P}_y$ are discrete and finite with n equiprobable values:</u>

W is represented by square matrix also noted W, nW being bistochastic.

$$D^2(\mathcal{P}_x, \mathcal{P}_y) = Min_{\{W\}} \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{i,j}(x_i - y_j)'(x_i - y_j)$$

Standard linear programme: the set $\{W\}$ is convex, closed, and bounded.
The lower bound is reached for one or several joint distributions $W = P/n$, where P is a permutation matrix.

Let p be the permutation of order n associated to P.

For a fixed p, each point x_i is bijectively associated to the point y_j, with $j = p(i)$: $P_{i,j} = \delta_{i,p(i)}$

$$D^2(\mathcal{P}_x, \mathcal{P}_y) = \frac{1}{n} Min_{\{P\}} \sum_{i=1}^{i=n} (x_i - y_{p(i)})'(x_i - y_{p(i)})$$

For each of the $n!$ pairwise correspondences between the n points x_i and the n points $y_{p(i)}$,
we calculate the sum of the n squared distances (or their quadratic mean),
then we search the pairwise correspondence(s) minimizing this sum.

FORMAL LINK BETWEEN WASSERSTEIN METRIC AND LEAST SQUARES METHODS

The problem of the optimal superposition of two sets of n points:

The sum of the n squared distances is minimized for a class of transformations of the second set (usual transformations: affine, orthogonal, rotation and translation, etc...)

• **Procrustes** methods under free correspondence, encountered in descriptive statistics.

• **RMS** or **RMSD** method under free correspondence: "Root Mean Square", "Root Mean Square Deviation", encountered in chemistry and structural biology (case of isometries in R^3).

In the case of a free correspondence, the Procrustes distance and the RMS distance are L^2 Wasserstein distances.

Fixed correspondence: these methods were many times rediscovered in sciences, and are by far more used than the free correspondence ones.

We need a method able to "fix a pairwise correspondence" in the discrete case, while still operating on continuous distributions.

A method permitting to "fix a pairwise correspondence" in the case of continuous distributions is welcome, too!

COLORED MIXTURES

We do an unusual presentation of ***mixture*** of distributions.

The process has two steps:
(1) Pick a "color"
(2) The d-variate distribution is determined by the choice of the color

Measurable space $(C \times R^d, A \otimes B)$:
C: space of colors, not empty. Examples: $C = R$; $C = \{red, green, blue\}$.
A: σ-algebra on C
B: Borel σ-algebra on R^d

K is a random variable defined on the probability space (C, A, P), P being the probability distribution of K.

To each color $c \in C$ we associate a probability distribution $\tilde{P}_c$ of R^d via the function $\Phi : c \mapsto \tilde{P}_c = \Phi(c)$.

The value of the distribution function of $\tilde{P}_c$ at the point x is a conditional probability noted $\tilde{F}(x|c)$.

We consider the random variable (K, X), X being a random vector taking values in R^d.

X is called a ***colored mixture*** when its distribution function F is: $F(x) = \int\limits_{c \in C} \tilde{F}(x|c) P(dc)$

COUPLES OF COLORED MIXTURES

Now we consider the couple of random variables $((K_x, X), (K_y, Y))$, where X and Y are colored mixtures.

P_{xy} : joint probability distribution of (K_x, K_y)

$\tilde{W}$: conditional joint distribution function associated to (Φ_x, Φ_y); in general, $\Phi_x \neq \Phi_y$

The joint distribution function W of the couple (X, Y) is got by integration:

$$W(x, y) = \int_{c_x \in C} \int_{c_y \in C} \tilde{W}(x, y | c_x, c_y) P_{xy}(dc_x, dc_y)$$

Nothing new until now about mixtures and couples of mixtures, except that we bother with a slightly more complicated presentation.

NEW: THE "COLORED" MIXTURES MODEL: HANDLING THE CONSTRAINTS ON CORRESPONDENCES

Additional assumption $\boxed{K_x \overset{a.s.}{=} K_y}$

Consequences:

$P_{xy}(dc_x, dc_y) = P(dc_x)\delta_{[c_x=c_y]}dc_y$ (δ is the Dirac function)

$W(x,y) = \int_{c \in C} \tilde{W}(x,y|c)P(dc)$ (P is marginal of P_{xy}, i.e. the distribution of K_x or K_y).

$\boxed{\textit{The dependancy between } K_x \textit{ and } K_y \textit{ in the space of colors induces a dependancy between the r.v. } X \textit{ and } Y.}$

Application: we define the colored L^2 Wasserstein distance between distributions of colored mixtures:

$$D^2 \overset{def}{=} Inf_{\{W\}} E(X-Y)'(X-Y)$$

(the concept is applicable to any L^p Wasserstein distance, $p \in \mathbb{N}^*$)

The set $\{W\}$ above is a non empty subset of the one used in the classical Wasserstein distance definition.

NOT all joint distributions are "permitted": e.g., the one of the convolution product can be outside $\{W\}$!!

When C contains only one color (i.e. K_x and K_y are an a.s. constant random variable), the colored Wasserstein distance is the ordinary Wasserstein distance (no space of colors).

EXAMPLE: THE COLORED BERNOUILLI DISTRIBUTION

$$C = \{red, green\} \qquad\qquad P(red) = P(green) = 1/2$$

X: we associate to $c = red$ the distribution such that $Prob(0) = 1$.
and we associate to $c = green$ the distribution such that $Prob(1) = 1$.

$Prob(X = 0|red) = 1 \quad Prob(X = 1|red) = 0$
$Prob(X = 0|green) = 0 \quad Prob(X = 1|green) = 1$

Here we choose Y distributed as X (in this particular example, $\Phi_2 = \Phi_1$).

General expression of the joint distribution W: $\quad W(x, y) = \sum_{c \in \{red, green\}} \tilde{W}(x, y|c)P(c)$

When $Prob(X = x|c) \neq 0$: $\quad \tilde{W}(x, y|c) = Prob(X = x|c)Prob((Y = y|c)|(X = x|c))$

When $Prob(Y = y|c) \neq 0$: $\quad \tilde{W}(x, y|c) = Prob(Y = y|c)Prob((X = x|c)|(Y = y|c))$

When $x \neq y$:
one (but only one) of the quantities $Prob(X = x|c)$ or $Prob(Y = y|c)$ is zero. Then, respectively, either $Prob((X = x|c)|(Y = y|c)) = 0$, or $Prob((Y = y|c)|(X = x|c)) = 0$. Then: $\tilde{W}(x, y|c) = 0$.

When $x = y$, $Prob(X = x|c)$ and $Prob(Y = y|c)$ take both together either the value 0 or the value 1.

We get $\tilde{W}(x, y|c)$ equal either to 0 or to 1.

There is only one possible joint distribution W:

	$Y=0$	$Y=1$	
$X=0$	$1/2$	0	$1/2$
$X=1$	0	$1/2$	$1/2$
	$1/2$	$1/2$	1

Both marginals X and Y have indeed the Bernouilli distribution of parameter $1/2$.

Dependancy between X and Y: $Cov(X,Y) = E(XY) - E(X)E(Y) = 1/4$

Thus the correlation coefficient is $r = +1$.

COLORED MODEL: PARTICULAR CASES

(a) Only one color: $P(c) = 1$.

All possible joint distributions of (X, Y) are "permitted".

This situation is equivalent to the one of the non colored model.

For physicists: << Model of ***indiscernibles particles*** >> (does not assume a finite number of particles).

(b) Almost surely constant random variables:

$\forall c \in C, \exists (x, y)$ such that $Prob(X = x|c) = 1$ and $Prob(Y = y|c) = 1$

Thus it exists only one possible joint distribution W.

There are 2 sets of "colored points" (common set of colors).

In each set of points, not two points have the same color.

For physicists: << Model of ***discernibles particles*** >> (does not assume a finite number of particles).

There is a bijection between these two sets: the points are pairwise associated.

The two sets have the same cardinality, but this latter may be finite or not.

Case of two samples of size n:

Only one joint distribution, i.e. one allowed permutation matrix $P = nW$

We rediscover the Procrustes methods and the RMS methods with fixed correspondence (permutation p): the sum D^2 of the n squared distances is minimized for some class of transformations of the second set (e.g. isometry, rotation, etc.)

$$D^2 = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - y_{p(i)})'(x_i - y_{p(i)})$$

The Procrustes distance D is a "colored" Wasserstein distance.

More general case for two "colored" samples of size n:

Each of the two sets of n points is partitioned into k subsets, of sizes $n_1, n_2, ..., n_k$.
The correspondence is free for each pair of subsets associeted to a color.

There are $\prod_{i=1}^{i=k} n_i!$ allowed permutations.

COLORED MODEL: MAIN INTEREST

We are now able to fix the "pairwise correspondence", even for infinite sets. The most general form of this "pairwise correspondence" is a joint distribution.

THE CHIRAL INDEX: GENERAL DEFINITION

X et Y are colored mixtures such that the distribution of Y is a mirror image of the distribution of X, this image being submitted to an arbitrary rotation R and translation t.
The inertia T of X or Y is assumed to be finite and not null.

$$\boxed{\text{Definition:} \quad \chi \stackrel{def}{=} \tfrac{d}{4T} Inf_{\{W,R,t\}} E(X-Y)'(X-Y)}$$

- χ takes value in $\{0;1\}$.
- χ is insensitive to translation, rotation, mirror inversion, and scaling.
- $\chi = 0 \Leftrightarrow X$ is achiral.
- $\chi = 1 \Rightarrow V = \sigma^2 I$ (the variance matrix V is proportional to I)

The optimal translation t is such that $EY = EX$, i.e. $t = 0$ when $EX = EY = 0$.
The optimal rotation R is known for $d \leq 3$.
The optimal joint distribution W exists and it is known for $d = 1$:
see the solution of the Monge-Kantorovitch transportation problem for a quadratic cost.

Discernible particles model (finite cardinality is not assumed):

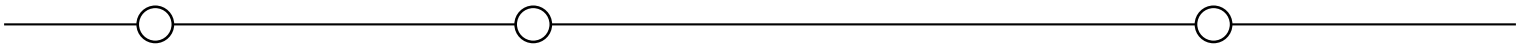
$\chi = d\sigma_d^2/T$ (σ_d^2 is the smaller eigenvalue of V)

$\chi = 1 \Leftrightarrow V = \sigma^2 I$

(**e.g., a cube of which not two vertices bear the same color, is of maximal chirality**)

Two "colored samples": $\chi = CHI$

THREE POINTS SETS ON THE REAL LINE



α : ratio of the lengths of the two adjacent segments.

$$\boxed{\chi = \frac{(1-\alpha)^2}{4(1+\alpha+\alpha^2)}}$$

Many scientists defined chirality measures, but few have the desired properties

(M. Petitjean, *Chirality and Symmetry Measures: A Transdisciplinary Review.* Entropy, **2003**, 5[3], 271-312.)

In the simplest case (above), a safe chirality measure should satisfy to:

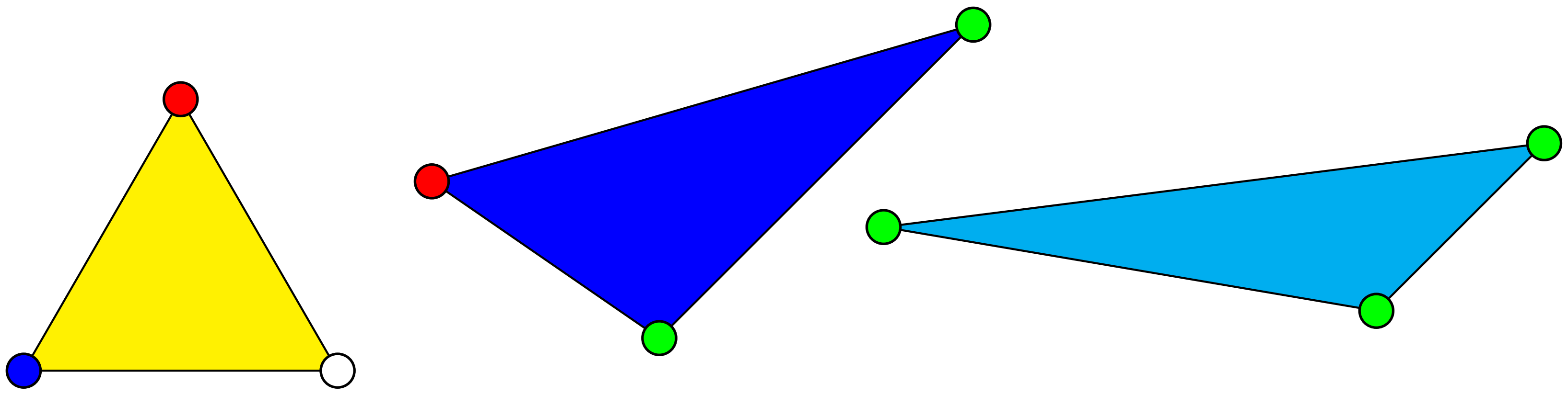
- χ is a function of α and ONLY of α, the unique parameter of the set (invariance through isometries).
- $\chi = 0$ IF AND ONLY IF the set is "symmetric" (i.e. achiral: α=1).
- χ is a continuous function of α.
- $\chi(\alpha) = \chi(1/\alpha)$ (invariance through scaling).

MAXIMAL CHIRALITY THREE POINTS SETS IN THE PLANE

(A) All non-equivalent vertices: $\chi = 1$, reached for the equialteral triangle.
This result generalizes in any dimension: the most chiral simplex with all non-equivalent vertices is regular.

(B) Two equivalent vertices. Distances ratio: $\sqrt{1-\sqrt{6}/4} : 1 : \sqrt{1+\sqrt{6}/4}$; $\chi = 1 - \sqrt{2}/2$

(C) Three equivalent vertices. Distances ratio: $1 : \sqrt{4+\sqrt{15}} : \sqrt{(5+\sqrt{15})/2}$; $\chi = 1 - 2\sqrt{5}/5$



(A) (B) (C)

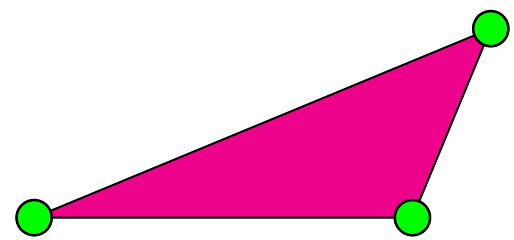
LEAST DIRECT SYMMETRIC THREE POINTS SETS

(D) All non-equivalent vertices: **direct symmetry is impossible.**

(E) Two equivalent vertices. Abscissas: $(-1-\sqrt{3})/2$, $(-1+\sqrt{3})/2$, 1 (aligned points); $DSI = 1$, $\forall d$.

(F) Three equivalent vertices. Angles: $\pi/4, \pi/8, 5\pi/8$; $DSI = 1 - \sqrt{2}/2$

(E)



(F)

GEOMETRIC PROPERTY OF THE EXTREMAL TRIANGLES (A), (B), (C), (E), (F)

The squared lengths of the sides are equal to three times the squared distances vertex-barycenter.

$$\|x_2 - x_3\|^2 = \|x_1 - g\|^2, \qquad \|x_1 - x_2\|^2 = \|x_2 - g\|^2, \qquad \|x_3 - x_1\|^2 = \|x_3 - g\|^2, \qquad g = (x_1 + x_2 + x_3)/3$$

CARE: The relation is symmetric for two points only

GRAPHS

$\chi_G \stackrel{def}{=} \frac{d}{4T} Min_{\{P,R,t\}} Tr(X - PXQ'R')'(X - PXQ'R')/n$ (n: number of nodes of the graph realized in R^d).

<u>Each permutation is bijectively associated to one graph automorphism.</u>

E.g.: A ring of n equivalent nodes has $2n$ permutations, not $n!$ permutations.

When the graph is complete, $\chi_G = CHI$.

<u>Application to molecular graphs</u> (realized in R^3):

The nodes are the atoms and the edges are the chemical bonds.
Molecular graphs are simple, undirected, no loops on nodes, nodes are labelled with atom types (include atomic symbol, atomic mass, charge etc.), and eges are labelled with colors (chemical bond type).
In general, molecular graphs are not connected: the connex components are treated separately (**QCM** freeware).

• The water molecule H-O-H has 3 nodes and 2 edges. Its graph has 2 automorphisms. $\chi_G = 0$ (planarity).

• Br-CHF-Cl has 5 nodes with all different colors, and 4 edges with the same color.
There is only one automorphism, corresponding to $P = I$.
Assuming that the carbon atom lies at the center of a regular tetrahedron, we get $\chi_G = 1$.
If there are no colors on the extremal vertices (e.g. as in the methane CH_4: 24 automorphisms), we get $\chi_G = 0$.

Both χ and χ_G generalize CHI.

Open problem: can we have a unique definition for χ and χ_G ?

CONTINUITY, CONVERGENCE

We would be happy with something like: *"similar" objects* $\Rightarrow$ *"close" chiral indices*

• In the finite discrete case, with colors and/or graphs,
χ and χ_G are indeed a continuous functions of the coordinates x_1, ..., x_n.

• In the case of colored mixtures, when there is only one color (or equivalently, no color at all):

$\mathcal{P}_n$ is the distribution of the random vector X_n
$\mathcal{P}$ is the distribution of the random vector X (of inertia T).

Convergence theorem:
If the sequence $(\mathcal{P}_n)$ of probability distributions converge to $\mathcal{P}$ and $E[X_n' X_n] \to E[X'X] < \infty$, and $T > 0$,
then $\chi(\mathcal{P}_n) \to \chi(\mathcal{P})$.

Remark: *Weaker is the convergence criteria in the space of objects is, stronger is the convergence theorem.*
The convergence in distribution (i.e. convergence in law) is the weakest usual convergence encountered for r.v.

Particular situation of interest: sequence of samples.

The chiral index of a parent distribution can be estimated from the sample chiral index.

ASYMETRY COEFFICIENT. SKEWNESS

• Pearson 1895: $S_k = E(X-\mu)^3/\sigma^3$, $\mu = EX$, $\sigma = \sqrt{E(X-\mu)^2}$

Symmetric distribution (i.e. achiral distribution) $\Rightarrow S_k = 0$. **RECIPROCITY FALSE**

• Chiral index (no color, $d=1$): $\chi = (1 + Inf_{\{W\}} r)/2$

W is the joint distribution of (X_1, X_2), X_1 and X_2 being **identically** distributed

r is the correlation coefficient between X_1 and X_2 ($\chi \leq 1/2$ because $Inf_{\{W\}} r \leq 0$)

Symmetric (i.e. achiral) distribution $\Leftrightarrow \chi = 0$ **(RECIPROCITY IS NOW TRUE)**

Sample of size n, no color: the minimal correlation is reached when the observations sorted in increasing values are correlated with the observations sorted in decreasing values: ***easy with a pocket calculator***

Expressions of χ from embedded interval midranges and lengths of the order statistics $x_{i:n}, i = 1, ..., n$:

$$\chi = [\sum_{i=1}^{i=n} (\tfrac{x_{i:n} + x_{n+1-i:n}}{2})^2 - n \cdot \bar{x}^2]/(ns^2) \qquad \chi = 1 - [\sum_{i=1}^{i=n} (\tfrac{x_{i:n} - x_{n+1-i:n}}{2})^2]/(ns^2)$$

Open problem: build symmetry tests for some classes of the parent distribution (normality, uniformity, etc.) and find the asymptotic distribution of χ_n or of some simple function of χ_n so that we can use standard tables.

RANDOM VECTORS: WHAT ARE THE MAXIMAL CHIRALITY DISTRIBUTIONS ?

X_1 and X_2 are identically distributed. For clarity, $EX_1 = EX_2 = 0$, wlog. $T = Tr(EX_1'X_1) = Tr(EX_2'X_2)$

$d = 1$: $\chi = (1 + Inf_{\{W\}}r)/2$ $\chi \leq 1/2$

Upper bound asymptotically reached by the Bernouili law of parameter tending to 0 or to 1.

$d = 2$: $\chi = 1 - Sup_{\{W\}}|\eta_1 - \eta_2|/T$ η_1, η_2: eigenvalues of $E(X_1X_2' + X_2X_1')/2$

or: $\chi = 1 - Sup_{\{W\}}|Ez_1z_2|/T$ z_1, z_2: complex r.v. associated respectively to X_1 and X_2

Sample of size n, matrix X (n lines, 2 columns, centered: $\mathbf{1}'X = 0$), permutation matrix P.

$\chi = 1 - Max_{\{P\}}|\eta_1 - \eta_2|/T$ η_1, η_2: eigenvalues of $X'(P + P')X/2n$

or: $\chi = 1 - Max_{\{P\}}|z'Pz|/\|z\|^2$ $z \in C^n$, $\mathbf{1}'z = 0$

Theorem: $Max_{\{P\}}|z'Pz|$ is reached for $P = P'$ (still true with the colored model for $d \leq 2$).

Open problem: find $Inf_{\{z \neq 0, \mathbf{1}'z=0\}} Max_{\{P\}}|z'Pz|/\|z\|^2$

(then use the stochastic convergence theorem $\chi_n \to \chi$ to get the upper bound of χ)

Theorem: $\chi \in [1 - 1/\pi; 1 - 1/2\pi]$

Adding the assumption $z'z = 0$ (i.e. variance matrix V proportional to I), we get $\chi \leq 1 - 2/3\pi$

We know a family of samples so that χ is arbitrarily close to $1 - 1/\pi$ (and so that $z'z$ tends to 0)

Conjecture: the upper bound $1 - 1/\pi$ is optimal.

$d \geq 3$: The upper bound of χ is in $[1/2; 1]$.

Open problem: find the upper bound of χ and characterize the associated distributions.

$d = 3$: W being fixed, the optimal quaternion q^* is the unit eigenvector of the largest eigenvalue of

$$B = \left[\begin{array}{c|c} 0 & E(X_1 \wedge X_2)' \\ \hline E(X_1 \wedge X_2) & I \cdot Tr(Z + Z') - (Z + Z') \end{array}\right] \qquad Z = E(X_1 X_2')$$

$$D^2 = D_0^2 - 2q^{*\prime} B q^*, \qquad D_0^2 = E(X_2 + X_1)'(X_2 + X_1)$$

(D_0^2 and the elements of $E(X_1 \wedge X_2)$ are linear combinations of the elements of Z)

Remark: $\boxed{\text{To get the optimal } q^* \text{ when } X_1 \text{ and } X_2 \text{ have different distributions, replace } X_2 \text{ by } -X_2.}$

$d > 3$: Open problem: find the optimal Procrustes rotation.

Remark: As previously mentioned, for any d, when all vertices have different colors, $V = \sigma^2 I \Leftrightarrow \chi = 1$
(it is the case for the set of vertices of any regular d-polytope)

Other open problems:

- No color: does $V = \sigma^2 I$ characterize maximal chirality distributions ?
- No color: can the upper bound of χ be reached, or is it only asymptotical ?
- Starting from a chiral distribution, can we define its associated closest achiral distribution ?
  (some sufficient conditions are known in the finite discrete case)
- Translational symmetry, infinite mass: the colored mixtures model fails. What about lattices, helices, etc. ?

BIBLIOGRAPHY

D. Coppersmith and M. Petitjean. *About the Optimal Density Associated to the Chiral Index of a Sample from a Bivariate Distribution*. Compt. Rend. Acad. Sci. Paris, Série I, **2005**, 340[18], 599-604.

M. Petitjean. *About the Upper Bound of the Chiral Index of Multivariate Distributions*. AIP Conf. Proc. **2008**, 1073, 61-66.

M. Petitjean. *A propos de la référence achirale*. Compt. Rend. Chim., **2006**, 9[10], 1249-1251. *(in French)*

M. Petitjean. *From Shape Similarity to Shape Complementarity: toward a Docking Theory*. J. Math. Chem. **2004**, 35[3], 147-158.

M. Petitjean. *Chirality and Symmetry Measures: A Transdisciplinary Review*. Entropy, **2003**, 5[3], 271-312.

M. Petitjean. *Chiral mixtures*. J. Math. Phys., **2002**, 43[8], 4147-4157.

M. Petitjean. *Chiralité quantitative: le modèle des moindres carrés pondérés*. Compt. Rend. Acad. Sci. Paris, serie IIc, **2001**, 4[5], 331-333. *(in French)*

M. Petitjean. *On the Root Mean Square Quantitative Chirality and Quantitative Symmetry Measures*. J. Math. Phys., **1999**, 40[9], 4587-4595.

M. Petitjean. *Calcul de chiralité quantitative par la méthode des moindres carrés*. Compt. Rend. Acad. Sci. Paris, serie IIc, **1999**, 2[1], 25-28. *(in French)*

M. Petitjean. *About Second Kind Continuous Chirality Measures. 1. Planar Sets*. J. Math. Chem., **1997**, 22[2-4], 185-201.

S.T. Rachev. *Probability Metrics and the Stability of Stochastic Models*. Wiley, **1991**.

S.T. Rachev and L. Rüschendorf. *Mass Transportation Problems*. vol. I: *Theory*. Springer-Verlag, **1998**.

C. Villani. *Topics in Optimal Transportation*. AMS Graduate Studies in Mathematics, Vol. 58, **2003**.

C. Villani. *Optimal Transport. Old and New*. Springer-Verlag, **2009**.