

SOR'13

The 12th International Symposium on
Operations Research in Slovenia

Dolenjske Toplice, Slovenia, September 25-27, 2013

**THE CHIRAL INDEX: APPLICATIONS TO MULTIVARIATE
DISTRIBUTIONS AND TO 3D MOLECULAR GRAPHS**

Michel Petitjean

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

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

WHAT IS CHIRALITY ?

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

Isometry: conservation of distances: combination of translations, rotations, and mirror inversions.

Direct isometry: only translations and rotations.

Indirect isometry: translations and rotations combined with an **ODD** number of mirror inversions.

Achirality: an achiral object is *identical* to one of its images through an indirect isometry.

Chirality: an object which is not achiral is chiral.

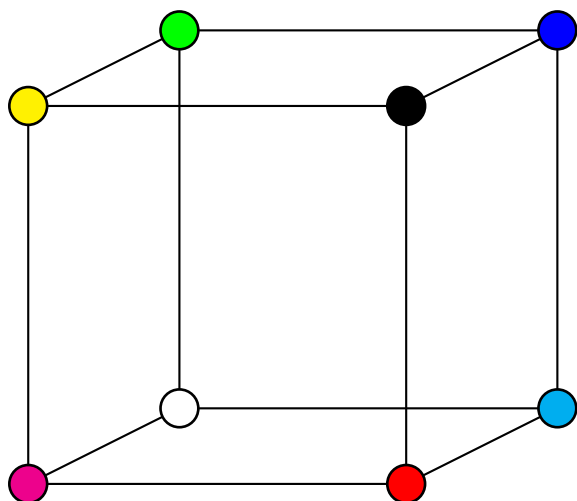
Chirality measure: quantitative measure of the deviation from achirality.

Remark 1: the full definitions of symmetry and chirality are more complex (group theory).

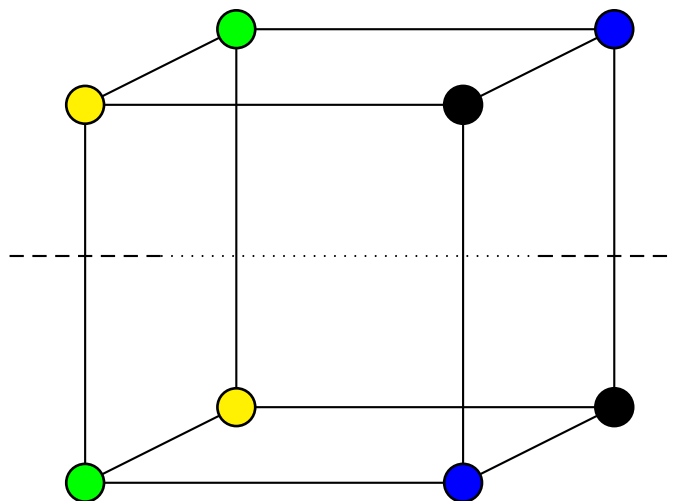
Remark 2: despite a common belief, the orientability of space is **NOT** required to define chirality.

PROBLEM

EUCLIDEAN SYMMETRY AND CHIRALITY ARE NOT STRICTLY GEOMETRIC CONCEPTS



(A)



(B)

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

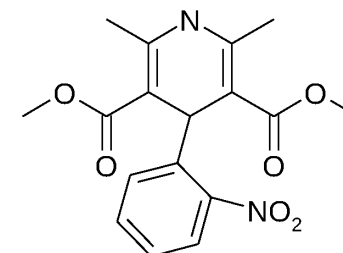
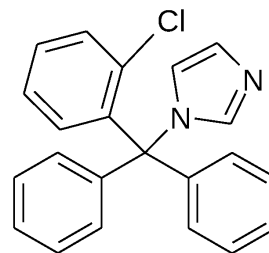
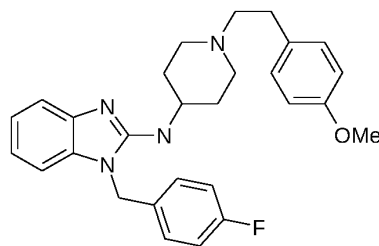
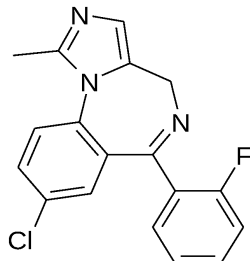
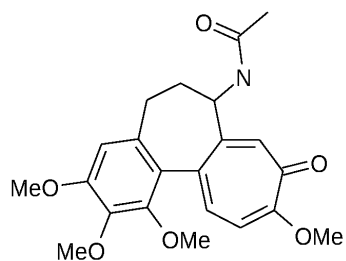
A mechanism to associate points is needed.

We built it in a general framework, not necessarily involving symmetry or chirality.

MOLECULAR GRAPHS

- Molecular graphs are realized in \mathbb{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 colors: atom type (atomic symbol), atomic mass, charge etc.
- Edges are labelled with colors: chemical bond type.
- In general, molecular graphs are not connected: the connex components are treated separately.

Some examples of molecular graphs:



The water molecule H-O-H has 3 nodes and 2 edges. Its graph has 2 automorphisms.

Br-CHF-Cl has 5 nodes with all different colors, and 4 edges with the same color.

There is only one automorphism.

The molecular graph of the cyclohexane ring C₆ has 12 automorphisms, not 6! automorphisms.

MEASURING THE ASYMMETRY OF DISTRIBUTIONS

Symmetric (i.e. **achiral**) distributions: binomial of parameter $1/2$, Gauss, etc.

Asymmetric (i.e. **chiral**) distributions: exponential, Poisson, etc.

Remark: for univariate distributions, there is **NO** direct symmetry.

X is a random variable of expectation μ and variance σ^2

Coefficient of asymmetry: e.g. Pearson's skewness (1895): $E\left(\frac{X-\mu}{\sigma}\right)^3$

But: it exists asymmetric distributions with null skewness !

*We need an asymmetry coefficient for multivariate distributions
which takes the value 0 if and only if the distribution is **indirect** symmetric.*

We must handle colors, molecular graphs, and multivariate distributions.

We first need a method able to fix a pairwise correspondence (permutation) in the discrete case, while able to operate in the continuous case (fix a bijection).

CLASSICAL MIXTURES OF DISTRIBUTIONS

A mixture distribution function is a convex linear combination of distribution functions.

The coefficients of the convex linear combination are the probabilities associated to the mixing distribution. This latter is usually discrete and finite.

The mixed distributions (called *components*) are usually those of random vectors taking values in \mathbb{R}^d . These components should not be confused with the d components of some vector of \mathbb{R}^d .

We are going to do an unusual presentation of mixtures of distributions.

COLORED MIXTURES

We consider a two steps process:

- (1) Pick a "color"
- (2) The d -variate distribution is determined by the choice of the color

Measurable space $(C \times \mathbb{R}^d, A \otimes B)$:

C : space of colors, not empty. Examples: $C = R$; $C = \{\text{red, green, blue}\}$.

A : σ -algebra on C

B : Borel σ -algebra on \mathbb{R}^d

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

To each color $c \in C$ we associate a probability distribution \tilde{P}_c of \mathbb{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 \mathbb{R}^d .

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

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 complicated presentation.

THE COLORED MIXTURES MODEL: HANDLING THE CONSTRAINTS ON CORRESPONDENCES

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

Consequences:

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

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

The dependancy between K_x and K_y in the space of colors induces a dependancy between the r.v. X and Y .

In other words, with probability 1 the two picked colors are the same, so that the d -variate distributions associated resp. to X and Y are "constrained" to be selected together.

E.g., when these two d -variate distributions are those of a.s. constant random vectors, we have put "two colored points in correspondence".

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.

(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 two sets of "colored points" (common set of colors).

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

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.

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

THE COLORED WASSERSTEIN DISTANCE

We define the **colored Wasserstein distance** D_c between distributions of colored mixtures X and Y .

\mathcal{P}_x and \mathcal{P}_y : respective distributions of X and Y

W : joint distribution of (X, Y) with fixed marginals \mathcal{P}_x and \mathcal{P}_y

\mathcal{W}_c : set of all joint distributions W

$$D_c^2(\mathcal{P}_x, \mathcal{P}_y) \stackrel{\text{def}}{=} \text{Inf}_{\{W \in \mathcal{W}_c\}} E(X - Y)'(X - Y)$$

(can be generalized to colored L^p Wasserstein distances, $p \in \mathbb{N}^*$)

\mathcal{W}_c is not empty but NOT all joint distributions between X and Y are possible, because in general X and Y cannot be independent.

When C contains only one color, the colored Wasserstein distance is the usual Wasserstein distance (i.e. no color) between X and Y , as encountered in the Monge-Kantorovitch transportation problem for a quadratic cost.

\mathcal{W}_c is a subset of the set of all joint distributions between X and Y which is used to define the usual Wasserstein distance.

Case where both \mathcal{P}_x and \mathcal{P}_y are discrete and finite with n equiprobable values (samples)

W is represented by a square matrix that we denote also by W , nW being bistochastic.

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

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 $j = p(i)$: $P_{i,j} = \delta_{i,p(i)}$

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

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.

Case of two *colored* samples of size n

Two sets of n points pairwise associated (n colors):

There is only one joint distribution, i.e. only one permutation matrix $P = nW$

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

$\sqrt{n}D_c$ is the Frobenius norm of the difference between the array of the x_i and the array of the $y_{p(i)}$.

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, \dots, n_k .

The correspondence is free for each pair of subsets associated to a color.

There are $\prod_{i=1}^{i=k} n_i!$ permutations to enumerate to find D .

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** (Root Mean Square Deviation) method under free correspondence: they are encountered in chemistry and structural biology in the case of isometries in \mathbb{R}^3 .

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

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

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

SOME OPTIMIZATION RESULTS FOR PROCRUSTES PROBLEMS

Analytical results on the optimal joint distribution W^* exist for $d = 1$ in the non colored case: see the solution of the Monge-Kantorovitch transportation problem for a quadratic cost.

Assuming that some W is fixed we look for the optimal Procrustes transformations.

- Optimal translation t^* : it is such that $EY = EX$, i.e. $t^* = 0$ when $EX = EY = 0$

- Optimal affine transformation: $E(XY') \cdot [E(YY')]^{-1}$

- Optimal orthogonal transformation: UV'

U is the orthonormal matrix of eigenvectors of $E(XY')E(YX')$

V is the associated orthonormal matrix of eigenvectors of $E(YX')E(XY')$

Remark: in the discrete and finite case, this solution using the singular values decomposition was known since more than one half century. It was rediscovered many times, although it was often erroneously proposed as a solution for the optimal rotation.

Optimal rotations

• $d = 2$: $R^* = \cos(r^*) \cdot I + \sin(r^*) \cdot \Pi$, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\Pi = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

$$\cos(r^*) = E(X'Y)/\mathcal{E}, \quad \sin(r^*) = E(X'\Pi Y)/\mathcal{E}$$

$$\mathcal{E} = [(E(X'Y))^2 + (E(X'\Pi Y))^2]^{1/2}, \quad D^{*2} = E(X'X) + E(Y'Y) - 2\mathcal{E}$$

• $d = 3$: the optimal quaternion q^* is the unit eigenvector of the largest eigenvalue of B

$$B = \left[\begin{array}{c|c} 0 & E(Y \wedge X)' \\ \hline E(Y \wedge X) & (Z + Z') - \text{Tr}(Z + Z') \cdot I \end{array} \right], \quad I \text{ is the identity matrix,}$$

$$Z = E(YX'), \quad D^{*2} = D_0^2 - 2q^{*'} B q^*, \quad D_0^2 = E(X - Y)'(X - Y)$$

Remark: D_0^2 and the elements of $E(X \wedge Y)$ are linear combinations of the elements of Z

• $d > 3$: it is an open problem.

THE CHIRAL INDEX: DEFINITION

X and \bar{X} are colored mixtures. The distribution of \bar{X} is a mirror image of the distribution of X , this image being submitted to an arbitrary rotation R and translation t .

W is a joint distribution of (X, Y) .

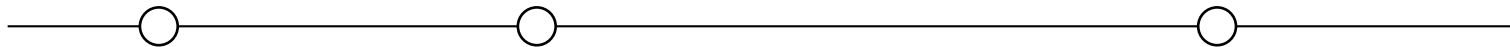
The inertia T of X or \bar{X} , i.e. the trace of the variance matrix, is assumed to be finite and not null.

The chiral index is the colored Wasserstein distance between the distributions of X and \bar{X} , minimized for all rotations R and translations t , and divided by $4T/d$.

$$\text{Definition: } \chi \stackrel{\text{def}}{=} \frac{d}{4T} \text{Inf}_{\{W, R, t\}} E(X - \bar{X})'(X - \bar{X})$$

- χ depends only on the distribution of X .
- 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 the identity matrix I)

BUILDING CHIRALITY MEASURES: A TEST SET
THREE NON COLORED POINTS ON THE REAL LINE



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

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

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).
- $\chi(1) = 0$ (the set is achiral $\Rightarrow \chi$ is null).
- $\chi = 0 \Rightarrow \alpha = 1$ (χ is null \Rightarrow the set is achiral).
- χ is a continuous function of α .
- $\chi(\alpha) = \chi(1/\alpha)$ (invariance through scaling).

Many scientists defined chirality measures, but most failed to satisfy the properties above.

FINITE SETS OF EQUALLY WEIGHTED POINTS

Set of n points, x_1, \dots, x_n in \mathbb{R}^d , colored or not: $X = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix}$, $Z = (I - \mathbf{1} \cdot \mathbf{1}'/n) \cdot X$

$\mathbf{1}$ is a vector of n components, all equal to 1, and $(I - \mathbf{1} \cdot \mathbf{1}'/n)$ is the centering operator. Thus $EZ = 0$, the covariance matrix is $V = Z'Z/n$, and $T = Tr(V)$.

Q : user arbitrary fixed orthogonal matrix in \mathbb{R}^d with $det(Q) = -1$.

R : unknown rotation matrix of order d .

P : unknown permutation matrix of order n , acting on the lines of X (or Z).

$$\chi = \frac{d}{4T} Min_{\{P,R\}} Tr(Z - PZQ'R)'(Z - PZQ'R)/n$$

Remark: the points have weights $1/n$, so that the inertia T is indeed $Tr(Z'Z/n)$, not $Tr(Z'Z)$.

The chiral (or achiral) object associated to X is the class of equivalence of the matrices deduced from X by the lines permutations "permitted" by the colors of the points.

CHIRALITY AND MOLECULAR GRAPHS

Same formula than for finite sets of colored points:

$$\chi = \frac{d}{4T} \text{Min}_{\{P,R\}} \text{Tr}(Z - PZQ'R')(Z - PZQ'R')/n$$

...except that the permutations P are those associated to the molecular graph automorphisms.

This set of automorphisms depends not only on the colors of the nodes, but also on the color of the edges and of the full graph (assumed to be simple and undirected, connected, with no loops on nodes)

Reminder: the nodes are the atoms and the edges are the chemical bonds.

- The water molecule H-O-H has 3 nodes and 2 edges.

Its graph has 2 automorphisms. $\chi = 0$ (planar molecule).

- Methane CH₄: there are 24 automorphisms.

The carbon atom lies at the center of a regular tetrahedron with the H as vertices: $\chi = 0$.

- 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, $\chi = 1$.

WHAT ABOUT DIRECT SYMMETRY MEASURES ?

For clarity, we assume that the colored mixture X is of null expectation.

Starting from the expression of the chiral index: $\frac{d}{4T} \text{Inf}_{\{W,R,t\}} E(X - Y)'(X - Y)$
then requiring Y to have the the same distribution than X through an arbitrary direct isometry...

...we would get an index always equal to 0.

E.g., when the arbitrary direct isometry is the identity,
just set $R = I$, $t = 0$, and W such that $d^2W(x, y) = dF(x) \cdot \delta_{[y=x]} dy$,
where F is the distribution function of X ,
and $\delta[y = x]$ is the Dirac delta function of $y \in \mathbb{R}^d$ at the point $y = x$.

How to solve that problem ?

- Removing $R = I$ and $t = 0$ from the set of unknown isometries fails: the lower bound is unchanged!
- Removing the optimal W fails, except in the finite case: just set $P \neq I$.

That solution is implemented in the QCM freeware: it computes χ and the direct symmetry index DSI for molecular graphs. The normalizing factor for DSI is $2T$ in any dimension d .

- Other solution: define a direct symmetry measure for a finite subgroup of direct isometries.

MAXIMAL CHIRALITY SETS AND MOLECULAR GRAPHS

We know that: $\chi = 1 \Rightarrow V = \sigma^2 I$.

In the case where there is only one possible joint distribution of (X, \bar{X}) , $\chi = d\sigma_d^2/T$

(σ_d^2 is the smaller eigenvalue of V)

It is the case for n equally weighted colored points, not two of them having the same color.

Under the "all different colors" assumption, the set of the vertices of the regular d -simplex, the set of the vertices of the d -cube, etc., are maximal chirality sets: $\chi = 1$,

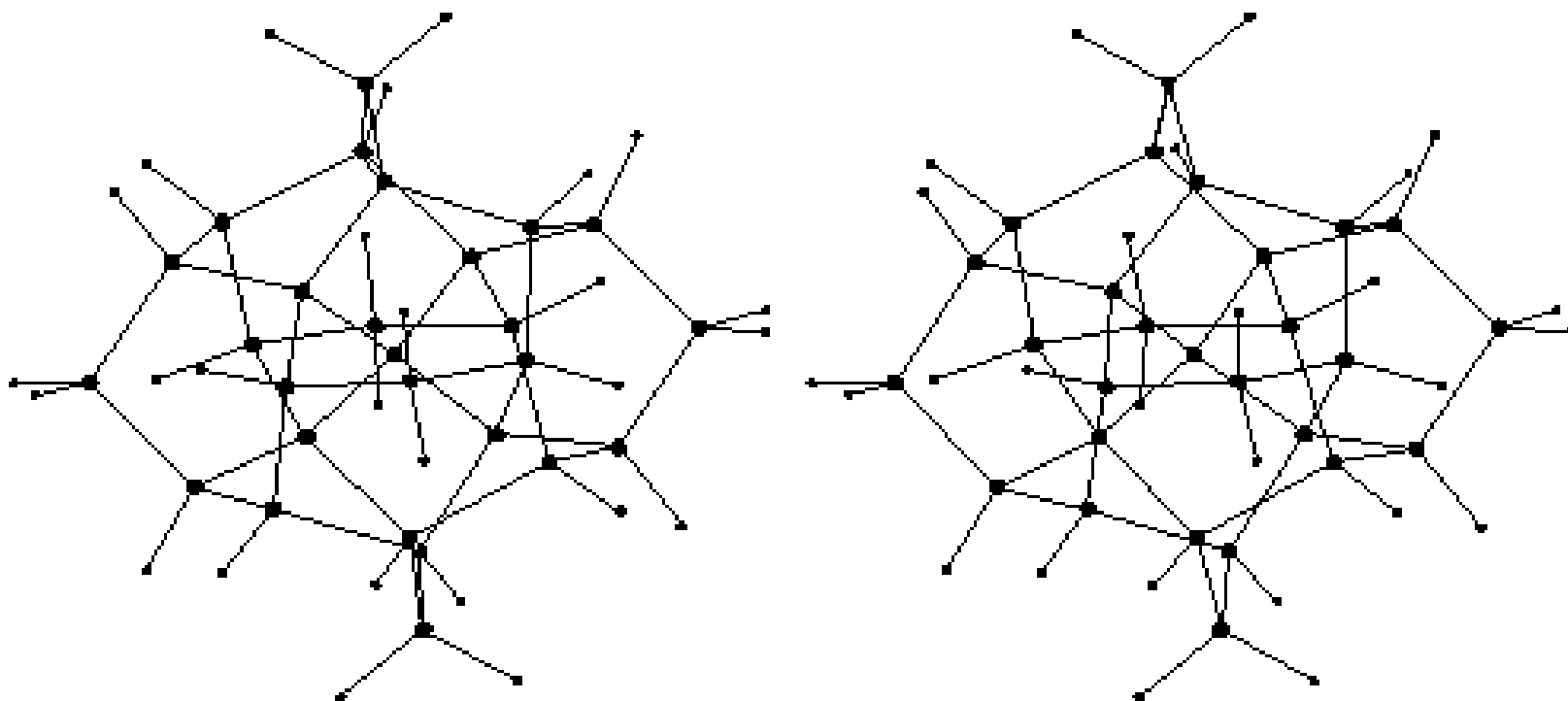
...although we would have $\chi = 0$ if the colors were discarded !

Maximal chirality molecular graphs:

Br-CHF-Cl, assuming that C is at the center of a regular tetrahedron, has $\chi = 1$.

Remarks:

- Chirality in chemistry is a complex concept involving dynamic properties and deformability. The present framework either assumes rigidity, or applies to a fixed conformation of the molecule.
- An efficient enumeration of molecular graphs automorphisms is required.



Two views of the [6.6]chiralane, $C_{27}H_{28}$, in configuration S.

This chiral and symmetric molecule is rigid. It was designed by A. Schwartz in 2004.

The full molecular graph has 768 automorphisms. $\chi = 0.9824$

The hydrogen suppressed graph (only carbons) has 12 automorphisms. $\chi = 1.0000$

MAXIMAL CHIRALITY THREE POINTS SETS IN THE PLANE

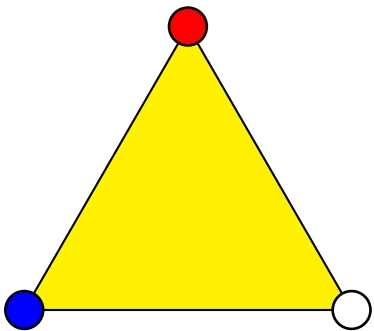
(A) All non-equivalent vertices: $\chi = 1$, reached for the equilateral triangle.

This result generalizes in any dimension:

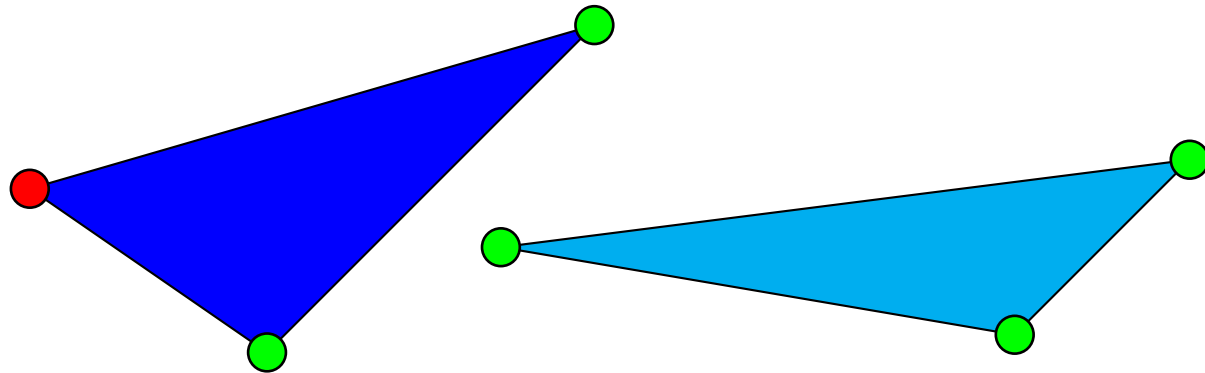
the most chiral simplex with all non-equivalent vertices is regular.

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

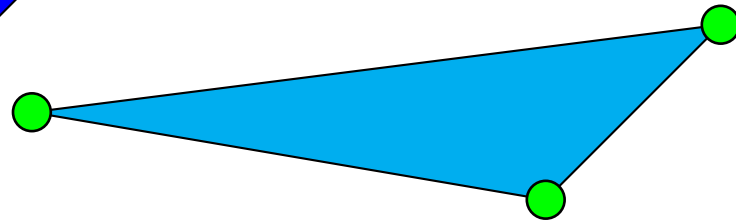
(C) Three equivalent vertices. Distances ratio: $1 : \sqrt{4 + \sqrt{15}} : \sqrt{(5 + \sqrt{15})/2}$; $\chi = 1 - 2\sqrt{5}/5$



(A)



(B)



(C)

LEAST DIRECT SYMMETRIC THREE POINTS SETS IN THE PLANE

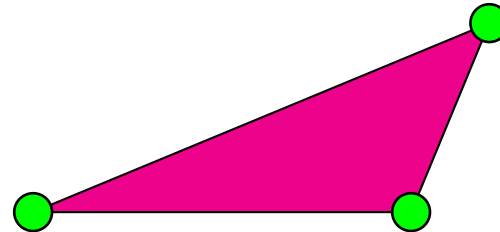
(D) All non-equivalent vertices: direct symmetry is impossible because we set $P \neq I$.

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

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



(E)



(F)

Geometric property of the five extremal triangles (A), (B), (C), and (E), (F):

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

$$\|x_2 - x_3\|^2 = 3\|x_1\|^2, \quad \|x_1 - x_2\|^2 = 3\|x_2\|^2, \quad \|x_3 - x_1\|^2 = 3\|x_3\|^2, \quad x_1 + x_2 + x_3 = 0$$

CARE: The relation is symmetric for two points only.

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 is a continuous function of the coordinates x_1, \dots, x_n .
- In the case of colored mixtures, when there is only one color (or equivalently, no color at all):

Convergence theorem:

\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 .

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

Remark:

*Weaker is the convergence criteria in the space of objects is, stronger is the convergence theorem.
The convergence in distribution (i.e. 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.

The upper bound of χ for a class of distributions can be sought from samples of increasing sizes.

THE CHIRAL INDEX AS AN ASYMETRY COEFFICIENT FOR DISTRIBUTIONS (NO COLOR)

• Achiral distribution \Rightarrow null Pearson's skewness **BUT RECIPROCITY FALSE**

• Achiral distribution \Rightarrow null chiral index **AND RECIPROCITY IS TRUE**

• Null chiral index \Rightarrow achiral distribution

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

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 $Inf_{\{W\}}r \leq 0$)

Sample of size n : 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**

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

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

Open problem: build symmetry tests for some class of the parent distribution (normal, uniform, etc.), find the asymptotic distribution of χ_n or of a simple function of χ_n so that we can use standard tables.

DISTRIBUTIONS OF RANDOM VECTORS: SOME RESULTS ON THE CHIRAL INDEX

X_1 and X_2 are identically distributed. For clarity, $EX_1 = EX_2 = 0$.

$d = 1$: $\chi = (1 + \text{Inf}_{\{W\}} r)/2$ Upper bound of the chiral index: $\chi_m = 1/2$
 Upper bound asymptotically reached by the Bernoulli law of parameter tending to 0 or to 1.

$d = 2$: $\chi = 1 - \text{Sup}_{\{W\}} |\eta_1 - \eta_2|/T$ η_1, η_2 : eigenvalues of $E(X_1X_2' + X_2X_1')/2$
 or: $\chi = 1 - \text{Sup}_{\{W\}} |Ez_1z_2|/T$ z_1, z_2 : complex r.v. associated respectively to X_1 and X_2
 $\chi_m \in [1 - 1/\pi; 1 - 1/2\pi]$

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

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

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

Theorem: The optimal permutation is symmetric (true with the colored model for $d \leq 2$).

$d \geq 3$: $\chi_m \in [1/2; 1]$

$d = 3$: The optimal Procrustes rotation is known.

SOME OPEN PROBLEMS

- No color: for $d \geq 2$, find χ_m and characterize the associated distributions.
- No color: can the upper bound χ_m be reached, or is it only asymptotical ?
- No color: does $V = \sigma^2 I$ characterizes (asymptotically) maximal chirality distributions ?
- For $d = 2$: find $\text{Inf}_{\{z \neq 0, z'z=0\}} \text{Max}_{\{P\}} |z'Pz| / \|z\|^2$
then use the stochastic convergence theorem to get χ_m
Adding the assumption $z'z = 0$ (i.e. V proportional to I), we get $\chi \leq 1 - 2/3\pi$
We know a family of samples with χ arbitrarily close to $1 - 1/\pi$ (it is so that $z'z$ tends to 0)
Conjecture: for $d = 2$, $\chi_m = 1 - 1/\pi$
- For $d > 3$: find the optimal Procrustes rotation.
- With or without color: given a chiral distribution, can we define its closest achiral distribution ?
(some sufficient conditions are known in the finite discrete case)
- Translational symmetry, infinite mass: the colored mixtures model fails (lattices, helices, etc.)

Family of sets conjectured to be asymptotically of maximal chirality: $\text{Lim Sup } (\chi) = 1 - 1/\pi$

Fix $\epsilon > 0$ then choose even integer $m > 1/\epsilon$.

$$\omega = e^{(2\pi)i/(2m)} \quad (\omega^{2m} = 1)$$

Select an integer $r > m^4/\epsilon^2$ then

select an even integer $k > r^{m-1}/\epsilon$

$z \in C^n$; z has $m + 3$ blocks of elements.

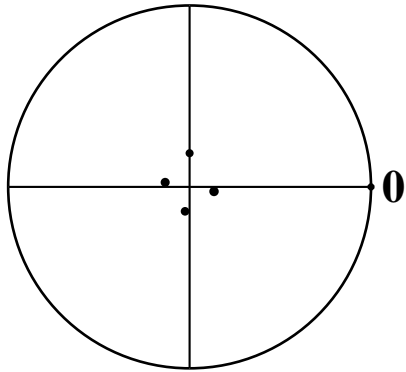
Each block $j, j = 0..m + 2$, contains identical elements.

$$n = 1 + r + r^2 + \dots + r^{m-1} + k + \frac{k}{2} + \frac{k}{2}$$

$$S = \sum_{j=0}^{j=m-1} \omega^j r^{j/2}$$

z is such that $z' \mathbf{1} = 0$ and $z' z = 0$

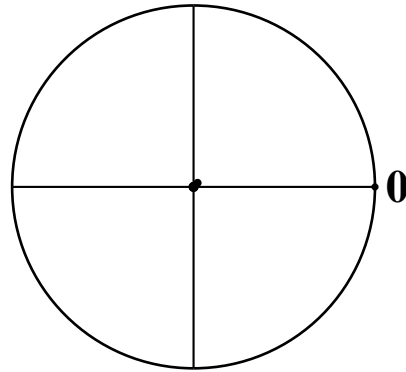
<i>block</i>	z_j	<i>multiplicity</i>
0	1	1
1	$\omega/r^{1/2}$	r
2	ω^2/r	r^2
\vdots	\vdots	\vdots
j	$\omega^j/r^{j/2}$	r^j
\vdots	\vdots	\vdots
$m - 1$	$\omega^{m-1}/r^{(m-1)/2}$	r^{m-1}
m	$-S/k$	k
$m + 1$	iS/k	$k/2$
$m + 2$	$-iS/k$	$k/2$



$\epsilon = 0.750$

$m = 2 ; m+3 = 5 ; r = 29$

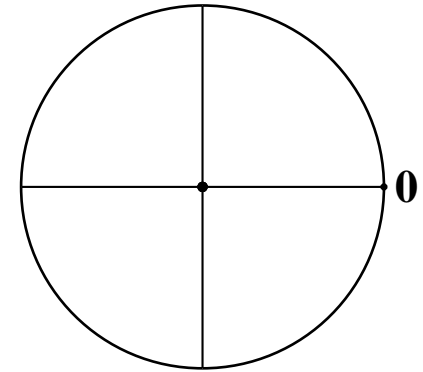
$k = 0.400E+02 ; n = 0.110E+03$



$\epsilon = 0.500$

$m = 4 ; m+3 = 7 ; r = 1025$

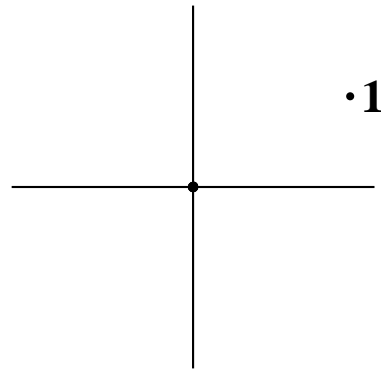
$k = 0.215E+10 ; n = 0.539E+10$



$\epsilon = 0.250$

$m = 6 ; m+3 = 9 ; r = 20737$

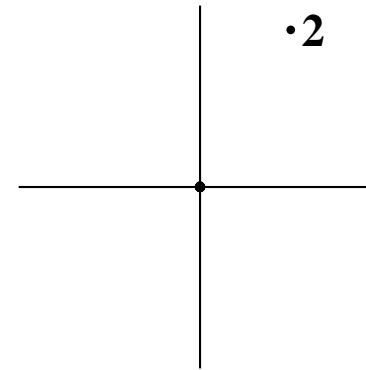
$k = 0.153E+23 ; n = 0.345E+23$



$\epsilon = 0.250 ;$ deleted block: 0 ; scaling: 144

$m = 6 ; m+3 = 9 ; r = 20737$

$k = 0.153E+23 ; n = 0.345E+23$



$\epsilon = 0.250 ;$ deleted blocks: 0,1 ; scaling: 20737

$m = 6 ; m+3 = 9 ; r = 20737$

$k = 0.153E+23 ; n = 0.345E+23$