Chiral mixtures

Michel Petitjean

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 43, NUMBER 8 AUGUST 2002

Chiral mixtures

Michel Petitjean[a)]
ITODYS (CNRS, ESA 7086), 1 rue Guy de la Brosse, 75005 Paris, France

(Received 22 November 2001; accepted for publication 11 April 2002)

An index evaluating the amount of chirality of a mixture of colored random variables is defined. Properties are established. Extreme chiral mixtures are characterized and examples are given. Connections between chirality, Wasserstein distances, and least squares Procrustes methods are pointed out. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1484559]

I. INTRODUCTION

Classifying a set as symmetric or not has been viewed as a dichotomic yes–no decision process for centuries. Attempts to evaluate the amount of symmetry has received little attention. Grünbaum (1963) noticed the difficulty to elaborate a rational approach of this problem. Physicists and chemists proposed various measures of the amount of chirality: see, for instance, Harris *et al.* (1999), Le Guennec (2000) or references cited by Petitjean (1997). Most methods handle only homogeneous solids, or only discrete sets. Many methods are limited to planar or spatial sets, and continuity properties are often ignored. E.g., for a homogeneous solid, the chiral index of Gilat (1989) is the normalized volume of the symmetric difference between the solid and its inverted image. The volume of the symmetric difference is the distance introduced by Dinghas (1957), this distance being itself the square of the $L2$-norm induced distance between the indicator functions of the solids. In this situation, continuity fails when the set becomes subdimensional. Clearly, functional distances applied to a set and its inverted image have no adequate continuity property because they are applied to densities rather than to distribution functions.

Thus, evaluating the degree of chirality of a random vector X in R^d is possible from some probability metric between the distribution of X and the distribution of its translated and rotated inverted image. The translation and the rotation are denoted respectively t and R. We consider now any two random vectors X and Y in R^d, and we look for a probability metric. For example, F being the distribution function of X, and G being the distribution function of Y, the quantity μ_K [Eq. (1.1)] is issued from the Kolmogorov metric:

$$\mu_K = \mathrm{Inf}_{\{R,t\}}(\mathrm{Sup}_{\{x\}}|F(x)-G(x)|). \tag{1.1}$$

But it was noticed previously (Petitjean, 1997, and 1999a), that some applications require us to consider colored mixtures, i.e., mixtures of colored random variables (see definition in the next section). An example is the algebraic charge density of a molecule or ion, which may be viewed as a mixture of two charge densities, namely the positive one and the negative one. As shown below, the quantity μ_K is not adequate for colored mixtures, because it is not sensitive to colors, and an extension of the Wasserstein distance will be preferred.

II. COLORED MIXTURES AND WASSERSTEIN DISTANCES

The assumption that Y is distributed as a translated and rotated inverted image of X is not used in this section.

A reason to work with the colored model is that, when evaluating the degree of chirality, Y has the distribution of a rotated inverted image of X, and therefore Y is a mixture such that each

a)Electronic mail: petitjean@itodys.jussieu.fr

component $\tilde{Y}$ retains the color of its associated component $\tilde{X}$ and is distributed as the rotated inverted image of $\tilde{X}$. In other words, the mirror (in fact, the inversion operator) sees the colors, e.g., the eight vertices of a cube constitute indeed a chiral figure in R^3 when they have all different colors. Another application needing a probability metric sensitive to colors is the optimal superposition problem (see Sec. II B), the quantitative chirality evaluation being just an instance of this problem.

A. Colored mixtures

When X is a mixture of colored random variables $\tilde{X}$, the more general formulation of its distribution is written in Eq. (2.1) with the mixing distribution P_1, and, similarly, the mixture Y of colored random variables $\tilde{Y}$ is written in Eq. (2.2) with the mixing distribution P_2:

$$F(x) = \int \tilde{F}(x,c) \cdot dP_1(c), \tag{2.1}$$

$$G(y) = \int \tilde{G}(y,c) \cdot dP_2(c). \tag{2.2}$$

When all the components of a mixture have the same color, it means that there is in fact only one colored component, and the colored mixture is an ordinary random vector in R^d. A colored mixture may be viewed as an ordinary mixture of random vectors, for which a supplementary axis has been added (the space of colors), this axis not being of numeric nature.

The joint distribution W of X and Y is expressed with the mixing distribution P operating on the mixed distributions $\tilde{W}$ [Eq. (2.3)]:

$$W(x,y) = \int \int \tilde{W}(x,y,c_1,c_2) \cdot d^2P(c_1,c_2). \tag{2.3}$$

In Eqs. (2.1)–(2.3), the summations are performed over the spaces of the colors. Now, we assume that the two colored mixtures X and Y are defined on the same space of colors. Moreover, the distribution of the colors is assumed to be the same for X and Y, i.e., the respective marginal distributions of X and Y on the space of colors are identical, and therefore can be fully correlated. This correlation is indeed assumed now: $P(c_1,c_2)$ is null when $c_1 \neq c_2$, i.e., $d^2P(c_1,c_2)$ is expressed with the Dirac delta function in Eq. (2.4), and integration over c_2 is performed in (2.3) to give the expression of W in Eq. (2.5), in which P_1 is renamed P and c_1 is renamed C:

$$d^2P(c_1,c_2) = dP_1(c_1) \cdot \delta_{[c_2=c_1]} dc_2, \tag{2.4}$$

$$W(x,y) = \int \tilde{W}(x,y,C) \cdot dP(C). \tag{2.5}$$

Clearly, the independence of the mixtures X and Y cannot be assumed now, except if X has only one colored component. This ''colored model'' is such that coupling the colors of a couple of mixtures X and Y induces constraints on the existence of their joint distributions W [Eq. (2.4)], and the set of joint distributions satisfying Eq. (2.5) is a nonempty subset of the set of the joint distributions of the same couple of mixtures discarding colors.

Equations (2.4) and (2.5) are assumed to stand further.

B. Colored Wasserstein distance and Procrustes methods

A probability metric depending on the joint density is sensitive to the constraints arising from colors [see Eq. (2.5)]. The well known Wasserstein metric (Dobrushin, 1970; Dudley, 1989;

Rachev, 1991) is so. The Wasserstein metric is itself an instance of the Kantorovich functional, which is encountered in the transportation problem (see equation 1.1.25 in Rachev and Rüschendorf, 1998).

The distributions associated respectively to X and Y are P^1 and P^2, and the matricial transposition operator is denoted by the quote. We name here colored Wasserstein distance $C(P^1,P^2)$ the extension of the $L2$ Wasserstein distance μ to colored mixtures, for which the lower bound of the expectation $E[(X-Y)'\cdot(X-Y)]$ is taken for all rotations R, translations t, and joint distributions W satisfying (2.5), i.e., such that each $\tilde{W}$ belongs to the class of all joint distributions of $\tilde{X}$ and $\tilde{Y}$:

$$D^2(W)=E[(X-Y)'\cdot(X-Y)], \tag{2.6}$$

$$\mu(P^1,P^2)=\mathrm{Inf}_{\{W\}}D(W), \tag{2.7}$$

$$C(P^1,P^2)=\mathrm{Inf}_{\{R,t\}}\mu(P^1,P^2). \tag{2.8}$$

In Eq. (2.6), it should be noticed that the expectation is defined through a $2d$-dimensional Lebesgue–Stiltjes integral, rather than a d-dimensional one. On the other hand, for any joint distribution W, computing $E(X'\cdot X)$ with the $2d$-dimensional integral leads to the same value that $E(X'\cdot X)$ computed with the d-dimensional one. The same remark is valid for $E(X)$, $E(Y)$, and $E(Y'\cdot Y)$.

Data analysis methods performing an optimal superposition of a set on another one via a least squares method were named Procrustes methods by Hurley and Cattell (1962), and the sum of the least squares is named the Procrustes distance. These methods are classified with the type of transformation allowed to superpose the moving set on the fixed set: general linear transformation, orthogonal transformation, or pure rotation. The 3D instance of this latter is usually encountered in physics, chemistry and bioinformatics: see references on the RMS algorithm cited in Petitjean (1998). The translation is optional, and it is always shown that the optimal translation is obtained when the mean points are superposed at the origin before further optimizations. The null expectation condition is not required in Procrustes methods.

Clearly, minimizing the $L2$ Wasserstein distance $\mu(P^1,P^2)$ [Eq. (2.7)], for some class of affine transformations of Y, generalizes the Procrustes method, because the usual one is its instance when X and Y are finite mixtures of n colored almost constant vectors, such that there is a one to one correspondence between the n vectors $\tilde{X}_i$ and $\tilde{Y}_i$. In this discrete situation, the unique feasible joint distribution is a bistochastic matrix equal to I/n, I being the identity matrix (colors are supposed to be enumerated in the same order for X and Y), and the Procrustes distance $\mathrm{Min}(D^2)$ is just the minimized sum of the squared distances between the n pairs of vectors. The Procrustes distance is the minimum of the distance induced by the norm itself induced by the scalar product $Tr(Z_X'\cdot Z_Y)$, where Z_X and Z_Y are two (n,d) rectangular matrices. The optimal rotation is analytically known when $d=2$ (see Section 3 in Petitjean, 1997), and when $d=3$ (see appendix in Petitjean, 1999b). The optimal orthogonal transformation is analytically known for any d (Golub and van Loan 1985).

For the noncolored model, i.e., when the n colors are identical, we get the Procrustes method without prefixed correspondence, for which the minimization of D^2 involves the enumeration of at most $n!$ possible correspondences between the two sets. Looking at the probabilistic formulation, the optimal joint distribution exists and is a bistochastic matrix equal to $1/n$ times a permutation matrix, because it is an extreme point of the convex polytope of the feasible solutions of the associated linear programming problem.

To summarize, the Procrustes distance becomes an instance of the $L2$ Wasserstein distance when this latter is extended to colored mixtures and minimized for a class of affine transformations of Y. Using the colored Wasserstein distance C [Eq. (2.8)] assumes that we work in the space of finite inertia colored mixtures, but the finite inertia condition could be relaxed if other adequate Wasserstein distances (see Rachev, 1991) are extended to colored mixtures. For clarity,

we restrict the affine transformations to rotations. In this situation, C is in fact a metric for classes of equivalence of colored mixtures, the colored mixtures being in the same class when their distributions are rotated (and optionally translated) images of one of them. It is pointed out that the colored Wasserstein distance is not defined when the mixtures have different marginal distributions in the space of colors. In this situation, an attempt to work with the ''maximal common substructure'' concept rather than with distances has been done for finite discrete sets (Petitjean, 1998). Of course, when the mixture Y is distributed as $\phi(X)$, ϕ being any transform leaving unchanged the marginal of X in the space of colors, C is indeed defined.

Some immediate properties of $C(P^1,P^2)$ follow.

Let m_{X_i} and m_{Y_i} be the respective expectations of X and Y attached to the i axis, $i \in [1,...,d]$, and σ_{X_i} and σ_{Y_i} be their respective standard deviations. The covariance attached to the i axis is c_i, and the respective inertia are T_X and T_Y. Equation (2.6) is now expandable as

$$D^2 = \sum_i \left[(\sigma_{X_i}^2 + m_{X_i}^2) + (\sigma_{Y_i}^2 + m_{Y_i}^2) - 2(c_i + m_{X_i} m_{Y_i})\right].$$

And, after rearrangement,

$$D^2 = T_X + T_Y + \sum_i \left[(m_{X_i} - m_{Y_i})^2 - 2c_i\right]. \tag{2.9}$$

The inertias and the covariances do not depend on the expectations. Thus the optimal translation t is such that $E(X) = E(Y)$, and the expression of D^2 becomes

$$D^2 = T_X + T_Y - 2\sum c_i. \tag{2.10}$$

Although the optimal joint distribution is not ensured to exist (Rachev and Rüschendorf, 1998), the optimal rotation is shown to exist, but may be not unique (Appendix A). The optimal general transformation and the optimal orthogonal transformation are known (Appendix A).

III. PROPERTIES OF THE CHIRAL INDEX

Let X and Y be colored mixtures in R^d, Y having the distribution of a translated and rotated inverted image of X. W is the joint distribution of the couple X, Y and T is the inertia of X or Y, i.e., $T = E[(X - E(X))' \cdot (X - E(X))]$ and $T = E[(Y - E(Y))' \cdot (Y - E(Y))]$. We define the chiral index χ as follows:

$$\chi = \frac{d}{4T} C^2(P^1,P^2). \tag{3.1}$$

In Eq. (3.1), P^2 being function of P^1, χ depends only on the law of X. In other words, χ is the normalized squared colored Wasserstein distance between the mixtures X and Y, Y being distributed as a translated and rotated inverted image of X. The chiral index is restricted to finite non-null inertia distributions. The situation $T = 0$ arises when X is almost surely equal to some constant x_0, and offers little interest. We neglect it. The chiral index is insensitive to isometries and is size free. As noticed in the preceding section, the optimal translation is obtained when $E(X) = E(Y)$, meaning that X and Y should be centered.

For clarity, we assume without loss of generality that the condition $E(X) = E(Y) = 0$ is satisfied in all this section.

The correlation coefficient attached to the i axis is r_i. Assuming the existence of the correlation coefficients, we get from Eqs. (2.10), (3.1) and (2.8):

$$D^2 = 2T - 2\sum c_i, \tag{3.2}$$

$$\chi = \frac{d}{2}\left(1 - \frac{\mathrm{Sup}_{\{R,W\}}(\Sigma c_i)}{T}\right). \tag{3.3}$$

When $d=1$, $R=1$, Y is distributed as $-X$, and there is only one standard deviation σ, and one correlation coefficient r. Equations (3.2) and (3.3) become

$$D^2 = 2\sigma^2(1-r), \tag{3.4}$$

$$\chi = \frac{1 - \mathrm{Sup}_{\{W\}}(r)}{2}. \tag{3.5}$$

In Eq. (3.5) the chiral index depends on one parameter only. For the noncolored model, this parameter is the maximal correlation of Gebelein (1952), applied to X and $-X$.

Now we return back to the d-dimensional space, and we look for a joint distribution ensured to exist. As noticed in the previous section, the independence of the mixtures X and Y cannot be assumed, except if X has only one colored component. The chiral index is proportional to the colored Wasserstein distance between the colored mixtures X and Y, Y being distributed as a rotated inverted image X (which does not mean that Y is a rotated inverted image of X). When Y is indeed the image of X through rotation R and inversion Q, the joint distribution of (X,Y) expressed from the mixed joint distributions $\tilde{W}(x,y,C)$ in Eq. (3.6) is ensured to exist:

$$d^2\tilde{W}(x,y,C) = d\tilde{F}(x,C)\cdot h_{[y=R\cdot Q\cdot x]}dy. \tag{3.6}$$

In Eq. (3.6), $h_{[y=y_0]}$ denotes the product of the d Dirac delta functions associated to the point y_0. Expression (3.6) is reported in (2.5) for integration over C, and, using Eq. (2.1), the final expression of the joint distribution is, as for a noncolored model:

$$d^2W(x,y) = dF(x)\cdot h_{[y=R\cdot Q\cdot x]}dy. \tag{3.7}$$

Equation (2.6) is expanded for this particular joint distribution to get Eq. (3.8), in which the expectation is calculated through a d-dimensional integral:

$$D^2 = 2T - 2E(X'\cdot R\cdot Q\cdot X). \tag{3.8}$$

The chiral index being insensitive to isometries, we assume now that the covariance matrix of X is diagonal, and that Y is the image of X through the inversion of the coordinate associated to the smallest variance axis. We take $R=I$. The inertia being the sum of the variances, Eq. (3.8) becomes

$$D^2 = 4\sigma_d^2. \tag{3.9}$$

The ratio of the smallest variance to the inertia is upper bounded by $1/d$, thus χ is upper bounded by 1. This bound is the best possible because it is reached for some particular random variables, as shown in Sec. V (see also the colored Bernoulli distribution in Appendix B):

$$0 \leqslant \chi \leqslant 1. \tag{3.10}$$

We consider now the finite discrete situation. The joint distribution is expressed with the square bistochastic matrix of the probabilities W_{ij} of each couple of values $\{x_i, y_j\}$. Using Eq. (2.6), the chiral index is rewritten

$$D^2=\sum_i \sum_j W_{ij}\cdot(x_i-y_j)'\cdot(x_i-y_j), \tag{3.11}$$

$$\chi=\frac{d}{4T}\mathrm{Inf}_{\{W_{ij},R,t\}}D^2. \tag{3.12}$$

Equations (3.11) and (3.12) were proposed previously to evaluate the amount of chirality of a fnite d-dimensional set, and thus our present approach generalizes the previous one [see Eqs. (3) and (4) in Petitjean (2001)]. It was also shown that, for the mass-uniform discrete case, the bistochastic matrix associated to the joint distribution is a permutation matrix. This particular situation means that there is a one-to-one mapping between the points of the set and those of its inverted image. In general, this mapping is not symmetric.

IV. CONVERGENCE

Obtaining the convergence of the chiral index from the convergence of the random variables is desirable to ensure some kind of continuity property of the chiral index. The weakest usual type of convergence possible for random variables is the convergence in law (in distribution), e.g., convergence of densities is a stronger assumption because this latter implies convergence in law [see Scheffé's theorem in Billingsley (1995)].

We consider the noncolored model. Let X_n be a sequence of random vectors converging to X in law. We assume also the convergence of $E[X_n'\cdot X_n]$ to $E[X'\cdot X]$, this latter quantity being finite. Apart from when X is almost surely constant, the convergence properties of the chiral index will arise from the convergence of $\mathrm{Inf}_{\{R\}}(\mu^2(P_n^1,P_n^2))=\mathrm{Inf}_{\{W_n,R\}}(D_n^2)$ to $\mathrm{Inf}_{\{R\}}(\mu^2(P^1,P^2))=\mathrm{Inf}_{\{W,R\}}(D^2)$, where μ denotes the Wasserstein distance [see Eq. (2.7)]. We use the triangle inequality to write

$$\begin{aligned}
&\mu(P^1,P^2)\leqslant\mu(P^1,P_n^1)+\mu(P_n^1,P_n^2)+\mu(P_n^2,P^2),\\
&\mu(P_n^1,P_n^2)\leqslant\mu(P_n^1,P^1)+\mu(P^1,P^2)+\mu(P^2,P_n^2),\\
&|\mu(P_n^1,P_n^2)-\mu(P^1,P^2)|\leqslant\mu(P_n^1,P^1)+\mu(P_n^2,P^2).
\end{aligned} \tag{4.1}$$

The inversion matrix Q being constant, inequation (4.1) stands for any rotation R common to Y_n and Y. For clarity, we name ϵ_n the second member of inequation (4.1). Obviously, ϵ_n does not depend on R. We note respectively $\mu_n(R)=\mu(P_n^1,P_n^2)$ and $\mu_\infty(R)=\mu(P^1,P^2)$. Inequation (4.1) is rewritten

$$|\mu_n(R)-\mu_\infty(R)|\leqslant\epsilon_n. \tag{4.2}$$

Let R_n and R_∞ be optimal rotations (which are shown to exist in Appendix A), respectively associated to D_n^2 and D^2. Inequation (4.2) stands for any R, and then stands for R_n and R_∞:

$$|-\mu_n(R_n)+\mu_\infty(R_n)|\leqslant\epsilon_n, \tag{4.3}$$

$$|\mu_n(R_\infty)-\mu_\infty(R_\infty)|\leqslant\epsilon_n. \tag{4.4}$$

We deduce from addition of (4.3) and (4.4)

$$|[\mu_n(R_\infty)-\mu_n(R_n)]+[\mu_\infty(R_n)-\mu_\infty(R_\infty)]|\leqslant 2\epsilon_n. \tag{4.5}$$

We know from optimality of rotations that each of the two quantities in brackets is nonnegative. Thus both quantities are upper bounded by $2\epsilon_n$:

$$|\mu_n(R_\infty)-\mu_n(R_n)|\leqslant 2\epsilon_n, \tag{4.6}$$

$$|\mu_\infty(R_n)-\mu_\infty(R_\infty)|\leqslant 2\epsilon_n. \tag{4.7}$$

Then, adding (4.3) and (4.7),

$$|\mu_n(R_n)-\mu_\infty(R_\infty)|=|\mu_n(R_n)-\mu_\infty(R_n)+\mu_\infty(R_n)-\mu_\infty(R_\infty)|\leqslant 3\epsilon_n .$$

This inequation is rewritten in terms of Wasserstein distances:

$$|C(P_n^1,P_n^2)-C(P^1,P^2)|\leqslant 3\epsilon_n . \tag{4.8}$$

It was assumed that X_n is converging to X in distribution, and that there was convergence of $E[X_n'\cdot X_n]$ to $E[X'\cdot X]$, with $E[X'\cdot X]<\infty$. These convergences are preserved through affine transformations. Thus, the distribution of Y_n is also converging to that of Y, discarding or not the common rotation used in inequation (4.2), and $E[Y_n'\cdot Y_n]$ is converging to $E[Y'\cdot Y]$ We know from theorem 6.2.1 in Rachev (1991) that the $L2$ Wasserstein distances $\mu(P_n^1,P^1)$ and $\mu(P_n^2,P^2)$ are tending to zero. Then, $\epsilon_n\to 0$, and we get from (4.8) the convergence of $C(P_n^1,P_n^2)$ to $C(P^1,P^2)$.

Looking to the definition of the chiral index in Eq. (3.1) shows that we need also to establish the convergence of the inertia, i.e., the centered two-order moment. The convergence of the two-order moment was assumed, thus it suffices to get the convergence of $E[X_n]$ to $E[X]$. Let A be any almost surely constant random vector, and P^A its distribution. We have from the triangle inequality:

$$|\mu(P_n^1,P^A)-\mu(P^1,P^A)|\leqslant\mu(P_n^1,P^1)$$

and therefore

$$|E[X_n'\cdot X_n]-E[X'\cdot X]-2E[A]'\cdot(E[X_n]-E[X])|\to 0 .$$

Setting the constant successively equal to each of the d canonical base vectors lead to get the desired convergence for each of the d components of the first order moment.

The convergence theorem follows now for the chiral index:

Theorem: *If the sequence* (P_n) *of probability distributions converges to* P *and* $E[X_n'\cdot X_n]\to E[X'\cdot X]<\infty$, *and* $E[(X-E[X])'\cdot(X-E[X])]>0$, *then* $\chi(P_n)\to\chi(P)$.

V. EXTREME CHIRALITY RANDOM VARIABLES

The chiral index maps X onto the interval $[0;1]$. Assuming $E(X)=E(Y)=0$, we look first to the minimum of the chiral index. Let us define a mixture X as achiral when it has the distribution of one of its rotated and inverted images. In this situation, X and Y can be identically distributed, and thus they can be fully correlated, i.e., $E(X'\cdot Y)=E(X'\cdot X)=E(Y'\cdot Y)$, and $\chi=0$. Conversely, when $\chi=0$, X is almost surely equal to Y, Y having the distribution of a rotated inverted image of X, meaning that X is achiral.

Now we look to the maximum of the chiral index. We assume that X has a diagonal covariance matrix, and that Y is the image of X through inversion of the coordinate associated to the smallest variance axis. We reuse the joint distribution in Eqs. (3.7) and (3.8), and $R=I$ is set, such that Eq. (3.9) stands. The ratio of the smallest variance to the inertia being upper bounded by $1/d$; χ cannot be equal to 1 unless all the d variances are equal. Therefore, the covariance matrix of X is proportional to the identity matrix. This covariance matrix is insensitive to isometries, and any rotation R is optimal for the joint distribution. Equation (5.1) expresses thus a necessary condition to get $\chi=1$:

$$E(X\cdot X')=\sigma^2\cdot I. \tag{5.1}$$

The d-dimensional finite mixture of n almost surely constant equiprobable colored variables is such that the joint distribution in Eqs. (3.7) and (3.8) is the only one feasible when all colors are different. It has been shown (Petitjean, 1999b) that the lower bound of D^2 in Eq. (3.8) is indeed that of Eq. (3.9), and the chiral index of the mixture is d times the percentage of inertia associated to the smallest eigenvalue of the covariance matrix of X:

$$\chi = d \cdot \sigma_d^2 \Big/ \sum_i \sigma_i^2. \tag{5.2}$$

Thus, $\chi = 0$ when the set of the n equiprobable points is subdimensional, and $\chi = 1$ when Eq. (5.1) is satisfied. Well-known figures satisfy Eq. (5.1), including regular planar polygons, cube and hypercubes, octahedron and higher dimensional analogs. Regular simplices fall also in this category. It should be pointed out that when the n colors are identical, these mixtures have a null chiral index because there is a symmetry $(d-1)$-hyperplane.

Some maximal chirality random variables can be exhibited for the noncolored model. The joint distribution of the convolution product always exists, and from Eq. (3.3), it comes that the chiral index is upper bounded by $d/2$. When $d = 1$, this bound is optimal, because it cannot be lowered for the Bernoulli distribution (see Appendix B). When $d \geqslant 2$, finding the upper bound for the noncolored model is an open problem. The distribution of three equiprobable points in the plane maximizing χ has been exhibited (Petitjean, 1997).

VI. DISCUSSION AND CONCLUSION

In the definition of χ, the division by the inertia T was needed to get a size free chiral index. Thus the degenerate random variable X with a null inertia has no chiral index, because both D^2 and T are null. Viewing this degenerate situation via the limit of a family of parametrized random variables makes no sense, in general, because the result depends on how the parameters are used to get a null inertia, and because no convergence exists around the singularity $T = 0$.

Conditions under which the convergence theorem (Sec. IV) could be extended to any colored mixture are to be investigated. A consequence of this convergence theorem is that the chiral index associated to the sample converges to that of the random variable. This could be used to get Monte Carlo approximations of χ when the analytical solution is unreachable, but building consistent estimators is outside the scope of this article. Computing the chiral index of a sample is equivalent to compute it in the finite discrete mass-uniform distribution. For the latter, the unidimensional case is solved analytically, and suitable numerical techniques have been built when $d = 2$ and $d = 3$ (Petitjean, 1997, 1999a, b). Computing χ for a general finite discrete distribution is a non linearly constrained optimization problem [see Eqs. (3.11) and (3.12)]. Constraints arising from the joint distribution are linear equalities and inequalities, because the matrix associated to the joint distribution is bistochastic. Constraints arising from the rotation are quadratic, i.e., $R' \cdot R = I$, and there is the polynomial constraint on the determinant of R.

For the noncolored model, when the rotation is fixed, our optimization problem is an instance of the transportation problem, which is a linear programming one. For the latter, solving algorithms and existence conditions of optimal joint distributions have been recently reviewed in Rachev and Rüschendorf (1998) (see also Anderson and Nash, 1987), and numerous results are available in the monodimensional case.

Compared to the noncolored model, the colored model introduces additional constraints on W. These constraints are handled by the $L2$ Wasserstein metric. Extending our present approach to other color sensitive probability metrics potentially gives rise to a family of similarity measures between colored mixtures, which seems not yet to be investigated, and from which the associated family of chiral indices could be derived.

It should be noticed that monodimensional distributions, such as the Gaussian, are confusingly called symmetric in most books. They are in fact achiral. Evaluating the amount of chirality is a different concept from evaluating the amount of direct symmetry. How to extend the present approach to direct symmetry is an open problem.

ACKNOWLEDGMENTS

The author is very grateful to the reviewer, who has done a careful reading and has suggested pertinent corrections, particularly about notations and about the formulation of the convergence theorem.

APPENDIX A: OPTIMAL PROCRUSTES TRANSFORMATIONS

The results in this appendix are valid for colored mixtures, and therefore stand for random vectors. We consider the colored Wasserstein distance $C(P^1,P^2)$ [Eqs. (2.6)–(2.8)], and we look for the lower bound of D^2 when the mixture Y is submitted to a linear transformation A and a translation t:

$$D^2=E[(X-(A\cdot Y+t))'\cdot(X-(A\cdot Y+t))], \tag{A1}$$

$$C^2=\mathrm{Inf}_{\{A,W,t\}}D^2. \tag{A2}$$

The gradient in t is null when $t=E(X)-A\cdot E(Y)$. It means that both mixtures should be centered before looking to the optimal value of A. The optional translation is further ignored, such that all results listed in this appendix remain valid, whether or not X and Y are centered prior any optimization. Now we look to the lower bound of D^2 for A. We have a quadratic expression of A, except if A is orthogonal:

$$D^2=E[(X-A\cdot Y)'\cdot(X-A\cdot Y)], \tag{A3}$$

$$C^2=\mathrm{Inf}_{\{A,W\}}D^2. \tag{A4}$$

1. The optimal general linear transformation

Derivating in (A3), we get:

$$E[2\cdot A\cdot Y\cdot Y'-2\cdot X\cdot Y']=0, \tag{A5}$$

$$A=E[X\cdot Y']\cdot(E[Y\cdot Y'])^{-1}. \tag{A6}$$

When the noncentered covariance matrix of Y is not inversible, we can try to solve by interchanging X and Y. If both noncentered covariance matrices are singular, the problem is in fact subdimensional.

2. The optimal orthogonal transformation

The solution given by Golub and van Loan (1985) is restricted to finite sets of equiprobable points (in a nonprobabilistic context). It is extended here to colored mixtures. For clarity, we set $A=Q$. Equation (A7) shows that D^2 is an affine expression of Q:

$$D^2=E[X'\cdot X+Y'\cdot Y-2\cdot X'\cdot Q\cdot Y]. \tag{A7}$$

Now we look for the upper bound of:

$$E[X'\cdot Q\cdot Y]=Tr(E[Y\cdot X']\cdot Q). \tag{A8}$$

Let us write in Eq. (A9) the singular value decomposition of the square matrix $E[Y\cdot X']$, i.e., S being the diagonal matrix containing the singular values, U being the orthonormal matrix of eigenvalues of $E[X\cdot Y']\cdot E[Y\cdot X']$, and V being the associated orthonormal matrix of eigenvalues of $E[Y\cdot X']\cdot E[X\cdot Y']$, we have

$$E[Y\cdot X']=V\cdot S\cdot U'. \tag{A9}$$

We look for the upper bound of $Tr(V\cdot S\cdot U'\cdot Q)=Tr(U'\cdot Q\cdot V\cdot S)$. The coefficients of the diagonal matrix S are non-negative, thus the trace is maximized when the coefficients of the orthogonal matrix $U'\cdot Q\cdot V$ are all equal to 1, meaning that $U'\cdot Q\cdot V=I$. The optimal matrix Q is

$$Q=U\cdot V'. \tag{A10}$$

When S is nonsingular, the determinant of Q is obtained from (A9) and (A10):

$$\det(Q) = \mathrm{sign}(\det(E[Y \cdot X'])). \tag{A11}$$

The sense of the eigenvectors of U and V are not independant, because the non-normalized matrix of eigenvalues of $E[Y \cdot X'] \cdot E[X \cdot Y']$ (which becomes V after normalization) is equal to $E[Y \cdot X'] \cdot U$. Thus, changing the sense of any eigenvector of U is still possible, but does not affect Q.

The optimal Q is unique, except when S has at least one null diagonal element.

3. The optimal *d*-dimensional rotation

As for the general orthogonal transformation [see Eqs. (A7) and (A8) in which we set $Q = R$ for clarity], we look to the upper bound of $Tr(E[Y \cdot X'] \cdot R)$, which is a linear expression of the unknown rotation. The set of rotations is closed and bounded in R^{d^2}. Our constrained maximization problem of a linear form in R^{d^2} has indeed a solution, but it may be not unique. The general expression of the solution is unknown, except in some particular situations. When $\det(E[Y \cdot X']) > 0$, the optimal rotation is given in Eq. (A10).

4. The optimal planar rotation

The planar rotation matrix is parametrized with the angle r:

$$R = I \cdot \cos(r) + \Pi \cdot \sin(r), \tag{A12}$$

where I is the identity matrix, and Π the antisymmetric matrix associated to the rotation of angle $\pi/2$. Reporting (A12) in (A3) and derivating for r gives the minimum and the maximum of D^2. The minimum is

$$\cos(r) = E[X' \cdot Y]/\mathbf{E}, \tag{A13}$$

$$\sin(r) = E[X' \cdot \Pi \cdot Y]/\mathbf{E}, \tag{A14}$$

$$\mathbf{E} = [(E[X' \cdot Y])^2 + (E[X' \cdot \Pi \cdot Y])^2]^{1/2}, \tag{A15}$$

$$D^2 = E[X' \cdot X] + E[Y' \cdot Y] - 2 \cdot \mathbf{E}. \tag{A16}$$

5. The optimal spatial rotation

The spatial rotation R is parametrized with the unit quaternion q. Its first component is the cosinus of the half rotation angle, and its other three components are the rotation axis, with length equal to the sinus of the half rotation angle. The quaternions q and $-q$ are associated to the same rotation. The optimal quaternion maximizes the quadratic form $q' \cdot B \cdot q$ in Eq. (A17) and the proof is essentially that established in the appendix of Petitjean (1999b) for finite sets of equiprobable points (in a nonprobabilistic context). It is extended here to colored mixtures. The optimal quaternion is the unit eigenvector associated to the highest eigenvalue of the symmetric matrix B:

$$D^2 = D_0^2 - 2 \cdot q' \cdot B \cdot q, \tag{A17}$$

$$D_0^2 = E[(X - Y)' \cdot (X - Y)], \tag{A18}$$

$$B = \begin{pmatrix} 0 & c' \\ c & (Z + Z' - I \cdot Tr(Z + Z')) \end{pmatrix}, \tag{A19}$$

$$Z = E[Y \cdot X'], \tag{A20}$$

$$c = E[Y \wedge X]. \tag{A21}$$

Note that the elements of c are computable from those of Z.

APPENDIX B: THE BERNOULLI DISTRIBUTION

The Bernouilli distribution is translated here to get a null expectation, i.e., the value $1-m$ has probability $P(1-m)=m$ and the value $-m$ has probabilty $P(-m)=1-m$. The rotation $R=1$, and the joint distributions between X and Y distributed as $-X$, are conveniently parametrized by only one parameter $p=P(X=-m\cap Y=m-1)$. Therefore, $P(X=1-m\cap Y=m-1)=m-p$, $P(X=-m\cap Y=m)=1-m-p$, and $P(X=1-m\cap Y=m)=p$. The covariance is $c=p-m(1-m)$, and the maximal correlation coefficient is reached for $p=m$ when $m\in[0;\frac{1}{2}]$, and for $p=1-m$ when $m\in[\frac{1}{2};1]$, i.e., $r=m/(1-m)$ and $r=(1-m)/m$, respectively. According to Eq. (2.1), $\chi=1-(\frac{1}{2})/(1-m)$ when $m\in]0;\frac{1}{2}]$, and $\chi=1-(1/2)/m$ when $m\in[\frac{1}{2};1[$. The chiral index is null when $m=\frac{1}{2}$, and is tending to $\frac{1}{2}$ when m is tending to 0 or to 1. The line $m=\frac{1}{2}$ is a symmetry axis for the graph of the function $\chi(m)$.

The colored Bernoulli distribution is, as for the noncolored one, a mixture of two random variables almost surely constant, with mixing proportions m and $1-m$. As previously, the mixture is translated to get a null expectation. However, the two components of the mixture are colored, and thus $P(X=-m\cap Y=m-1)=0$ and $P(X=1-m\cap Y=m)=0$. Setting now $p=P(X=-m\cap Y=m)$, the covariance is $c=-p\cdot m^2-(1-p)\cdot(1-m)^2$, and the maximal correlation coefficient is reached for $p=1$ when $m\in[0;\frac{1}{2}]$, and for $p=0$ when $m\in[\frac{1}{2};1]$, i.e., $r=-m/(1-m)$ and $r=(m-1)/m$, respectively. According to Eq. (2.1), $\chi=(\frac{1}{2})/(1-m)$ when $m\in]0;\frac{1}{2}]$, and $\chi=1/2m$ when $m\in[\frac{1}{2};1[$. The chiral index is equal to 1 when $m=\frac{1}{2}$, and is tending to $\frac{1}{2}$ when m is tending to 0 or to 1. The line $m=\frac{1}{2}$ is a symmetry axis for the graph of the function $\chi(m)$. This graph is the image of the previous one through the symmetry axis $\chi=\frac{1}{2}$.

Anderson, E. J. and Nash, P., *Linear Programming in Infinite Dimensional Spaces. Theory and Applications* (Wiley Interscience, Chichester, UK 1987), Chap. 5.

Billingsley, P., *Probability and Measure* (Wiley, New York, 1995), Chap. 3, Sec. 16.

Dinghas, A., "Über das Verhalten der Entfernung zweier Punktmengen bei gleichzeigtiger Symmetrisierung derselben," Arch. Math. **8**, 46–51 (1957).

Dobrushin, R. L., "Prescribing a system of random variables by conditional distributions," Theor. Probab. Appl. **15**, 458–486 (1970).

Dudley, R. M., *Real Analysis and Probability* (Wadsworth, Pacific Grove, CA, 1989), Chap. 11.8.

Gebelein, H., "Maximalkorrelation und Korrelationsspectrum," Z. Angew. Math. Mech. **32**, 9–19 (1952).

Gilat, G., "Chiral coefficient-a measure of the amount of structural chirality," J. Phys. A **22**, L545–L550 (1989).

Golub, G. H. and van Loan, C. F., *Matrix Computations* (John Hopkins University Press, Baltimore, MD, 1985), Sec. 12.4.

Grünbaum, B., *Measures of Symmetry for Convex Sets, Proceedings of Symposia in Pure Mathematics, Vol. VII*, Convexity, edited by V. L. Klee (American Mathematical Society, Providence, RI, 1963), pp. 233–270.

Harris, A. B., Kamien, R. D., and Lubensky, T. C., "Molecular chirality and chiral parameters," Rev. Mod. Phys. **71**, 1745–1757 (1999).

Hurley J. R. and Cattell, R. B., "The Procrustes Program: producing direct rotation to test a hypothesized factor structure," Behav. Sci. **7**, 258–262 (1962).

Le Guennec, P., "Two-dimensional theory of chirality," J. Math. Phys. **41**, 5954–5985, 5986–6006 (2000).

Petitjean, M., "About second kind continuous chirality measures. 1. Planar sets," J. Math. Chem. **22**, 185–201 (1997).

Petitjean, M., "Interactive maximal common 3D substructure searching with the combined SDM/RMS algorithm," Comput. Chem. (Oxford) **22**, 463–465 (1998).

Petitjean, M., "Calcul de chiralité quantitative par la méthode des moindres carrés," C.R. Acad. Sci., Ser. IIc: Chim **2**, 25–28 (1999a).

Petitjean, M., "On the root mean square quantitative chirality and quantitative symmetry measures," J. Math. Phys. **40**, 4587–4595 (1999b).

Petitjean, M., "Chiralité quantitative: le modèle des moindres carrés pondérés," C.R. Acad. Sci., Ser. IIc: Chim **4**, 331–333 (2001).

Rachev, S. T., *Probability Metrics and the Stability of Stochastic Models* (Wiley, New York, 1991), Chap. 6.

Rachev, S. T. and Rüschendorf, L., *Mass Transportation Problems* (Springer-Verlag, New York, 1998), Vol. I, Chap. 1.