REGULAR MULTIDISTANCES

Javier Martín **Gaspar Mayor**

Department of Mathematics and Computer Science
University of the Balearic Islands
{javier.martin,gmayor}@uib.es

Abstract

The conventional definition of a distance (metric) can be extended to apply to collections of more than two elements. In this paper we introduce a new class of multidistances that we call regular. Regularity, which can be considered in-between weakness and strongness properties, is studied for some remarkable families of multidistances.

Keywords: distance, multidistance, Fermat multidistãnce, OWA–based multidistance, Sum–based multidistance.

1 INTRODUCTION

The conventional definition of distance over a space specifies properties that must be obeyed by any measure of "how separated" two points in that space are. However often one wants to measure "how separated" the members of a collection of more than two elements are. The usual way to do it is to combine the distance values for all pairs of elements in the collection, into an aggregate measure. Thus, given an Euclidean triangle (A, B, C) we can combine the distances AB, AC, BC using, for instance, a 3-dimensional OWA operator, say W. Then, we calculate the distance of A, B, C by means of the formula $D(A, B, C) = W(AB, AC, BC)$. It is clear that we have to choose the weighting vector of W such that the multi-argument distance function D satisfies a group of axioms that extend to some degree those for ordinary distance functions. Of course we can consider other procedures to measure how separated the vertices A, B, C are: in Euclidean geometry the Fermat point of a triangle (A, B, C) is the point F for which the sum of the distances from F to the vertices is as small as possible; i.e. it is the point F such that $FA + FB + FC$ is minimized. Then we can define $D(A, B, C) = FA+FB+FC$. For pairwise distances and related distance matrix see for example [1].

A recent paper [3] deals with the problem of aggregating pairwise distance values in order to construct a multidistance function [8, 5].

In addition to their intrinsic mathematical interest, multidistances have many potential applications. They can be directly incorporated into many domains: indistinguishability measures [4], distance–based clustering, pattern recognition, etc., where the extension of ordinary (binary) distances to multidimensional collections can be of interest. This is also the case of the so-called Jensen-Shannon divergence (JSD) which is a distance for probability distributions that have been used to treat different problems such as annalysis of symbolic sequences, examination of texts in literature or separation of quantum states [6, 2].

In this paper we introduce a class of multi-argument distance functions that we call regular multidistances. After presenting some basic properties, regularity is studied for three relevant classes of multidistances: OWA–based multidistances, Sum–based multidistances and Fermat's multidistances.

2 PRELIMINARIES

Let us begin with the definition of multidistance, given in [5].

Definition 1 *A function* $D\colon \bigcup_{n\geq 1} X^n \to [0,\infty]$ *is a* multidistance *on a set* X *if it fulfills the following properties for all* n*, for all* $x_1, \ldots, x_n, y \in X$*:*

(m1) $D(x_1, \ldots, x_n) = 0$ *if and only if* $x_1 = \ldots = x_n$*,*

(m2) $D(x_1, \ldots, x_n) = D(x_{\pi(1)}, \ldots, x_{\pi(n)})$ *for any permutation* π *of* $1, \ldots, n$*,*

(m3) $D(x_1, \ldots, x_n) \leq D(x_1, y) + \ldots + D(x_n, y)$.

We say that D *is a* strong multidistance *if it fulfills (m1), (m2) and:*

(m3') $D(x_1,\ldots,x_n) \le D(\mathbf{x_1},\mathbf{y})+\ldots+D(\mathbf{x_k},\mathbf{y})$ *for any partition*[1] $\{\mathbf{x_1},\ldots,\mathbf{x_k}\}$ *of* $\mathbf{x}=(x_1,\ldots,x_n)$*, for all* $\mathbf{y}\in\bigcup_{n\ge1}X^n$*.*

Distances or multidistances taking values in the interval $[0,1]$ are called *normalized*.

Remark 1

1) *If D is a multidistance on X, then the restriction of D to X^2, $D|_{X^2}$, is an ordinary distance on X.*
2) *An ordinary distance d on X can be extended in order to obtain a multidistance. For example, we can define the following function* $D_M\colon\bigcup_{n\ge1}X^n\to[0,\infty]$*:*

$$D_M(\mathbf{x}) = \begin{cases} 0 & \text{if } n=1, \\ \max_{i<j}\{d(x_i,x_j)\} & \text{if } n\ge 2. \end{cases} \qquad (1)$$

Conditions (m1), (m2) and (m3) are easy to verify and then D_M is a multidistance such that $D_M|_{X^2}=d$. It will be called maximum multidistance *and it will be treated in Section 3.1.*

We will denote by $\mathcal{D}_d$ the set of multidistances D such that $D|_{X^2}=d$.

Multidistances allow us to define balls centered at lists [5].

Definition 2 *Given a multidistance D and a list $\mathbf{x}=(x_1,\ldots,x_n)\in\bigcup_{n\ge1}X^n$, the* closed ball *of center $\mathbf{x}$ and radius $r\in R$ is the set:*

$$B(\mathbf{x},r)=\{y\in X\colon D(\mathbf{x},y)\le D(\mathbf{x})+r\}. \qquad (2)$$

This definition is an extension of the usual one for balls centered at points.

3 REGULAR MULTIDISTANCES

This section deals with the property of regularity for multidistances. First we give the definition and some properties and then, the regularity of several classes of multidistances is studied.

Definition 3 *A multidistance D is* regular *if the following property holds, for all $\mathbf{x}\in X^n$ and $y\in X$:*

$$D(\mathbf{x})\le D(\mathbf{x},y). \qquad (3)$$

That is, the multidistance between the elements of a list can not decrease if we add an element.

[1] A partition of $\mathbf{x}=(x_1,\ldots,x_n)\in X^n$ is a set of lists $\{\mathbf{x_i}=(x_{i_1},\ldots,x_{i_{n_i}})\in X^{n_i}; i=1,\ldots,k\}$ chosen in such a way that the set of subindexes $\{\{i_1,\ldots,i_{n_i}\}, i=1,\ldots,k\}$ is a partition of $\{1,\ldots,n\}$.

Remark 2 *Regularity has a precise meaning in terms of the balls centered at lists: a multidistance is regular if and only if the balls with negative radius are empty.*

Regularity can be seen as a first step of strongness: if we consider a one–part partition, condition (m3') reduces to:

$$D(\mathbf{x})\le D(\mathbf{x},\mathbf{y}), \qquad (4)$$

but this is equivalent to (3).

As a consequence, we have the following result, relating these two properties.

Proposition 1 *Any strong multidistance is regular.*

A particular case of interest for strong multidistances is when the added element is one of the list.

Proposition 2 *If D is a strong multidistance on X, then*

$$D(\mathbf{x})=D(\mathbf{x},x_i) \qquad (5)$$

for all $\mathbf{x}\in X^n$, for any element x_i of the list $\mathbf{x}$.

Proof: Apply condition (m3') to the list $(\mathbf{x},x_i)$, the partition $\{(x_i,x_i),(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)\}$ and the added element x_i. As $D(x_i,x_i,x_i)=0$ and $D(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n,x_i)=D(\mathbf{x})$, we have $D(\mathbf{x},x_i)\le D(\mathbf{x})$.

The condition of regularity $D(\mathbf{x})\le D(\mathbf{x},x_i)$ completes the proof. ■

Therefore, repeated elements are superfluous when dealing with strong multidistances.

Basic properties of regularity are the following.

Proposition 3 *Let D and D' be regular multidistances on X.*

1) *$D+D'$ is also a regular multidistance on X.*
2) *If $k>0$, then kD is also a regular multidistance on X.*
3) *$\frac{D}{1+D}$ and $\min\{1,D\}$ are normalized regular multidistances on X.*

From now on we consider a metric space (X,d) with cardinality $|X|\ge 2$. We are going to study the regularity of some remarkable classes of multidistances belonging to $\mathcal{D}_d$.

3.1 Regularity of the OWA–based multidistances

Let $\mathcal{W}=\{W_n; n\ge 2\}$ be a family of OWAs [7], where the weights of the $\binom{n}{2}$–dimensional OWA W_n

$\omega_1^n, \ldots, \omega_{\binom{n}{2}}^n$, with $\omega_1^n + \ldots + \omega_{\binom{n}{2}}^n = 1$, are applied to the list of the $\binom{n}{2}$ pairwise distances arranged in an increasing order.

We can define a function $D_W\colon \bigcup_{n\geq 1} X^n \to [0,\infty]$ in this way:

$$D_W(\mathbf{x}) = \begin{cases} 0 & \text{if } n = 1, \\ W_n(\overbrace{d(x_1,x_2), \ldots, d(x_{n-1},x_n)}^{\binom{n}{2}}) & \\ & \text{if } n \geq 2, \end{cases} \tag{6}$$

for all $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$.

A special case is when $W_n = \max$, with lists of weights $(1, 0, \ldots, 0)$ for all n, obtaining the maximum multidistance D_M given in (1).

Lemma 1 *The multidistance D_M is strong.*

Proof. Let $\{\mathbf{x_1}, \ldots, \mathbf{x_k}\}$ be a partition of $\mathbf{x}$. We can suppose that the maximum distance is reached at x_1, x_2, that is, $D_M(\mathbf{x}) = d(x_1, x_2)$, with $x_1 \in \mathbf{x_r}$, $x_2 \in \mathbf{x_s}$ and $\mathbf{x_r} \neq \mathbf{x_s}$. So, for any component y of the list $\mathbf{y}$,

$$\begin{aligned} D_M(\mathbf{x}) = d(x_1,x_2) &\leq d(x_1,y) + d(x_2,y) \\ &\leq D_M(\mathbf{x_r},\mathbf{y}) + D_M(\mathbf{x_s},\mathbf{y}) \\ &\leq \textstyle\sum_{i=1}^k D_M(\mathbf{x_i},\mathbf{y}). \end{aligned}$$

And if x_1, x_2 belong to the same part, say $\mathbf{x_r}$, then

$$\begin{aligned} D_M(\mathbf{x}) = d(x_1,x_2) &\leq D_M(\mathbf{x_r},\mathbf{y}) \\ &\leq \textstyle\sum_{i=1}^k D_M(\mathbf{x_i},\mathbf{y}). \end{aligned}$$

Therefore condition (m3') is fulfilled and hence the multidistance D_M is strong. ■

The following result gives a necessary and sufficient condition for these OWA–based functions to be multidistances.

Proposition 4 *The function D_W is a multidistance if and only if, for all $n \geq 3$:*

$$\omega_1^n + \ldots + \omega_{n-1}^n > 0. \tag{7}$$

Proof: Condition (m2): d and W_n are symmetric and then D_W also is.

As D_M is a multidistance (lemma 1),

$$D_W(\mathbf{x}) \leq D_M(\mathbf{x}) \leq \sum_{i=1}^n d(x_i, y)$$

for all $y \in X$, and so (m3) holds.

Finally, condition (m1). Let us suppose that D_W fulfills it. If we take the list $(a, \overbrace{b, \ldots, b}^{n-1})$, with $d(a,b) = l > 0$, then:

$$\begin{aligned} 0 < D(a, \overbrace{b, \ldots, b}^{n-1}) &= W_n(\overbrace{l, \ldots, l}^{n-1}, \overbrace{0, \ldots, 0}^{\binom{n-1}{2}}) \\ &= l \cdot (\omega_1^n + \ldots + \omega_{n-1}^n), \end{aligned}$$

that is, $\omega_1^n + \ldots + \omega_{n-1}^n > 0$.

Reciprocally, let us suppose now (7).

Obviously, if $x_1 = \ldots = x_n$ then $D(\mathbf{x}) = 0$. And if there exist i, j such that $x_i \neq x_j$, then $d(x_i, x_j) > 0$. In this case, for all $k \neq i, j$ either $d(x_i, x_k) > 0$ or $d(x_j, x_k) > 0$; that is, there are at least $n-1$ non-zero pairwise distances and so $D(\mathbf{x}) \neq 0$. ■

The next proposition gives a necessary condition for an OWA–based multidistance to be regular.

Proposition 5 *If a function D_W defined by a family of OWAs $\mathcal{W} = \{W_n; n \geq 2\}$ is regular, then*

$$\omega_1^n + \ldots + \omega_{n-1}^n = 1, \tag{8}$$

for all $n \geq 3$.

Proof: We will prove it by induction. Let $a, b \in X$ such that $d(a,b) = l > 0$.

The base case is $n = 3$. For the lists (a,b) and (a,b,b) the condition of regularity is $l \leq \omega_1^3 l + \omega_2^3 l$, that is, $\omega_1^3 + \omega_2^3 = 1$.

Now the inductive step. Suppose that (8) holds for n. Consider the lists $(a, \overbrace{b, \ldots, b}^{n-1})$ and $(a, \overbrace{b, \ldots, b}^{n})$, whose lists of ordered pairwise distances are $(\overbrace{l, \ldots, l}^{n-1}, \overbrace{0, \ldots, 0}^{\binom{n-1}{2}})$ and $(\overbrace{l, \ldots, l}^{n}, \overbrace{0, \ldots, 0}^{\binom{n}{2}})$ respectively. The condition of regularity in this case is:

$$\omega_1^n + \ldots + \omega_{n-1}^n \leq \omega_1^{n+1} + \ldots + \omega_n^{n+1},$$

and so $\omega_1^{n+1} + \ldots + \omega_n^{n+1} = 1$. ■

Remark 3 *Condition* (8) *is necessary but it does not suffice. Let us consider, for example, the OWAs*

$$\begin{aligned} W_3 &= (1,0,0), \\ W_4 &= (0,0,1,0,0,0), \end{aligned}$$

fulfilling it and belonging to a family $\mathcal{W}$ which defines a multidistance D_W.

For the points $(0,1), (1,0), (0,0)$ *in the Euclidean plane we have the following values, according to (6):*

$$\begin{aligned} D_W((0,1),(1,0),(0,0)) &= W_3(\sqrt{2},1,1) \\ &= \sqrt{2}, \end{aligned}$$

$$\begin{aligned} D_W((0,1),(1,0),(0,0),(0,0)) &= W_4(\sqrt{2},1,1,1,1,0) \\ &= 1, \end{aligned}$$

and so D_W is not regular.

In fact, condition (8) is sufficient if and only if the range of the distance d consists exactly of two values.

Let us conclude with an example.

Example 1 *As we have seen in lemma 1, D_M, with OWAs $(1,0,\dots,0)$, is a strong multidistance. On the opposite side, the OWAs $(0,\dots,0,1)$ define the function D_m:*

$$D_m(\mathbf{x}) = \begin{cases} 0 & \text{if } n=1, \\ \min_{i<j}\{d(x_i,x_j)\} & \text{if } n \geq 2, \end{cases} \tag{9}$$

which is not a multidistance.

Between them, the multidistance $\frac{D_M+D_m}{2}$, corresponding to the OWAs $(\frac{1}{2},0,\dots,0,\frac{1}{2})$, $n \geq 2$, which is not regular.

Finally, if we take the OWAs $(\frac{1}{2},\frac{1}{2},0,\dots,0)$, $n \geq 2$, then the result is a multidistance D_W: the semisum of the two greatest pairwise distances. It is regular but in general it is not strong: if there exist three points $a,b,c \in X$ such that $d(a,b) < d(a,c) < d(b,c)$, then

$$\begin{aligned} D_W(a,b,c) &= \tfrac{d(a,c)+d(b,c)}{2}, \\ D_W(a,b,c,c) &= d(b,c), \end{aligned}$$

and D_W is not strong because it does not fulfill (5).

3.2 Regularity of the sum–based multidistances

The following multidistances have been defined in [5]. They are based on the sum of the pairwise distance values for all pairs of elements of the list, multiplied by a factor λ which depends on its length.

Definition 4 *The sum–based multidistances are the functions $D_\lambda\colon \bigcup_{n\geq 1} X^n \to [0,\infty]$ defined by*

$$D_\lambda(\mathbf{x}) = \begin{cases} 0 & \text{if } n=1, \\ \lambda(n)\sum_{i<j} d(x_i,x_j), & \text{if } n \geq 2, \end{cases} \tag{10}$$

where:

(i) $\lambda(2) = 1$,

(ii) $0 < \lambda(n) \leq \frac{1}{n-1}$ *for any $n > 2$.*

With respect to the regularity of these multidistances, we have the following result.

Proposition 6 *The multidistances D_λ are regular if and only if $\lambda(n) = \frac{1}{n-1}$ for all $n \geq 2$.*

Proof: First we prove by induction that if the multidistance D_λ is regular, then $\lambda(n) = \frac{1}{n-1}$ for all $n \geq 2$.

The base case holds by definition: $\lambda(2) = 1$.

Suppose now that $\lambda(n-1) = \frac{1}{n-2}$. If we take lists of the form $(a,\dots,a,b)$ we have:

$$D_\lambda(\overbrace{a,\dots,a}^{n-2},b) = \lambda(n-1)\cdot(n-2)l = l$$

and

$$D_\lambda(\overbrace{a,\dots,a}^{n-1},b) = \lambda(n)\cdot(n-1)l.$$

The condition of regularity (3) is

$$l \leq \lambda(n)\cdot(n-1)l,$$

that is, $\lambda(n) \geq \frac{1}{n-1}$ or, taking into account (ii) in definition 4, $\lambda(n) = \frac{1}{n-1}$.

Let us see now that the multidistance:

$$D_\lambda(\mathbf{x}) = \begin{cases} 0 & \text{if } n=1, \\ \frac{\sum_{i<j} d(x_i,x_j)}{n-1}, & \text{if } n \geq 2, \end{cases} \tag{11}$$

is regular. The condition of regularity (3) is:

$$\frac{\sum_{i<j} d(x_i,x_j)}{n-1} \leq \frac{\sum_{i<j} d(x_i,x_j) + \sum_{i=1}^{n} d(x_i,y)}{n}$$

for any $y \in X$, that is,

$$\textstyle\sum_{i<j} d(x_i,x_j) \leq (n-1)\sum_{i=1}^{n} d(x_i,y).$$

But this is fulfilled:

$$\begin{aligned} \textstyle\sum_{i<j} d(x_i,x_j) &\leq \textstyle\sum_{i<j}(d(x_i,y)+d(x_j,y)) \\ &= (n-1)\textstyle\sum_{i=1}^{n} d(x_i,y). \end{aligned}$$

■

Remark 4

i) The arithmetic mean of the pairwise distances corresponds to the multidistance D_λ with $\lambda(n) = \frac{1}{\binom{n}{2}}$, $n \geq 2$. It coincides with the multidistance D_W that has the OWAs $W_n = (\frac{1}{\binom{n}{2}},\dots,\frac{1}{\binom{n}{2}})$, $n \geq 2$, and it is not regular.

ii) The multidistance D_λ, with $\lambda = \frac{1}{n-1}$, is not strong. Consider for example the lists (a,a,b) and (a,a,b,b), with $d(a,b) > 0$.

$$\begin{aligned} D_\lambda(a,a,b) &= d(a,b), \\ D_\lambda(a,a,b,b) &= \tfrac{4}{3}d(a,b), \end{aligned}$$

but this means, see proposition 2, that D_λ is not strong.

iii) There are not multidistances D_λ fulfilling (5): $D(\mathbf{x}) = D(\mathbf{x}, x_i)$. For example, if $d(a,b) > 0$ then it should be

$$D_\lambda(a,a,b) = D_\lambda(a,a,a,b) = D_\lambda(a,a,b,b),$$

but

$$D_\lambda(a,a,a,b) = \lambda(4) \cdot 3d(a,b),$$
$$D_\lambda(a,a,b,b) = \lambda(4) \cdot 4d(a,b),$$

that is, $\lambda(4) = 0$ and D_λ is not a multidistance.

3.3 Regularity of the multidistance of Fermat

This multidistance was defined in [5] and it is based on the idea of the Fermat point explained in the introduction.

Definition 5 *The multidistance of Fermat is the function $D_F\colon \bigcup_{n\geq 1} X^n \to [0,\infty]$ defined by:*

$$D_F(x_1,\ldots,x_n) = \min_{x\in X}\{\sum_{i=1}^{n} d(x_i,x)\}. \qquad (12)$$

The following result states the regularity of this multidistance.

Proposition 7 *The multidistance of Fermat D_F is regular.*

Proof:

$$\begin{aligned} D_F(x_1,\ldots,x_n) &= \min_{x\in X}\{\textstyle\sum_{i=1}^{n} d(x_i,x)\} \\ &\leq \min_{x\in X}\{\textstyle\sum_{i=1}^{n} d(x_i,x) + d(y,x)\} \\ &= D_F(x_1,\ldots,x_n,y). \end{aligned}$$

■

Once the regularity of the multidistances of Fermat has been established, let us deal with the strongness.

Proposition 8 *The multidistance of Fermat D_F is not strong.*

Proof: Consider the lists (a,a,b) and (a,a,b,b) with $d(a,b) > 0$.

$$\begin{aligned} D_F(a,a,b) &= \min_{x\in X}\{2d(a,x) + d(b,x)\} \\ &= d(a,b), \end{aligned}$$

because the minimum is reached at $x = a$, and

$$\begin{aligned} D_F(a,a,b,b) &= 2\min_{x\in X}\{d(a,x) + d(b,x)\} \\ &\geq 2d(a,b), \end{aligned}$$

for all $x \in X$, which means, taking into account proposition 2, that D_F is not strong. ■

4 CONCLUSIONS

The property of regularity for multidistances has been introduced. This property, which is weaker than strongness, has been study for different families of multidistances, obtaining in each case complete characterizations.

Acknowledgements

The authors acknowledge the support of the Govern Balear grant PCTIB2005GC1–07 and the Spanish DGI grant MTM2009–10962.

References

[1] L.M. Blumenthal (1970). Theory and applications of distance geometry, 2nd ed. New York, Chelsea Publishing Co.

[2] P.W. Lamberti, A.P. Majtey, A. Borras, M. Casas, A. Plastino (2008). Metric character of the quantum Jensen-Shannon divergence. Physical Review 77.

[3] J. Martín, G. Mayor (2009). Aggregating pairwise distance values. Proccedings of the EUROFUSE 2009, Pamplona (Spain), pp. 147–152.

[4] J. Martín, G. Mayor (2009). An axiomathical approach to T–multi-indistinguishabilities. Proccedings of the IFSA–EUSFLAT 2009, Lisbon (Portugal), pp. 1723–1728.

[5] J. Martín, G. Mayor (2009). How separated Palma, Inca and Manacor are? Proccedings of the AGOP 2009, Palma de Mallorca (Spain), pp. 195–200.

[6] M.A. Ré, D.P. Prato, P.W. Lamberti (2007). Discriminación de distribuciones de probabilidad usando distancias entrópicas. Anales AFA 19, pp. 36–39.

[7] R.R. Yager (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18, pp. 183–190.

[8] D.H. Wolpert (2004). Metrics for more than two points at once. International Conference on Complex Systems ICCS 2004, Y. Bar-Yam (Ed.), Perseus books, in press.