An extension of the concept of distance as functions of several variables

Gergely Kiss, Jean-Luc Marichal, and Bruno Teheux

Mathematics Research Unit, University of Luxembourg
Luxembourg, Luxembourg
{gergely.kiss,jean-luc.marichal,bruno.teheux@}@uni.lu

Abstract. Extensions of the concept of distance to more than two elements have been recently proposed in the literature to measure to which extent the elements of a set are spread out. Such extensions may be particularly useful to define dispersion measures for instance in statistics or data analysis. In this note we provide and discuss an extension of the concept of distance, called n-distance, as functions of n variables. The key feature of this extension is a natural generalization of the triangle inequality. We also provide some examples of n-distances that involve geometric and graph theoretic constructions.

1 Introduction

The notion of metric space is one of the key ingredients in many areas of pure and applied mathematics, particularly in analysis, topology, and statistics.

Denote the half line $[0,+\infty[$ by $\mathbb{R}_+$. Recall that a *metric space* is a pair (X,d), where X is a nonempty set and d is a distance on X, that is a function $d\colon X^2 \to \mathbb{R}_+$ satisfying the following properties:

(i) $d(x_1,x_2)=0$ if and only if $x_1=x_2$,
(ii) $d(x_1,x_2)=d(x_2,x_1)$ for all $x_1,x_2\in X$,
(iii) $d(x_1,x_2)\leqslant d(x_1,z)+d(z,x_2)$ for all $x_1,x_2,z\in X$.

Property (iii) is often refereed to as *triangle inequality*.

It is natural to generalize the concept of metric space by considering a notion of "distance" among more than two elements of X. The idea behind such a notion is to measure in some sense how spread out the elements of X are. Several attempts in this line have been proposed for instance in [2–4,6,8,9]. For example, Martín and Mayor [6] recently introduced the concept of multidistance as follows. Let S_n denote the set of all permutations on $\{1,\ldots,n\}$. A *multidistance* on a nonempty set X is a function $d\colon \cup_{n\geqslant 1} X^n \to \mathbb{R}_+$ satisfying the following properties for every integer $n\geqslant 1$:

(i) $d(x_1,\ldots,x_n)=0$ if and only if $x_1=\cdots=x_n$,
(ii) $d(x_1,\ldots,x_n)=d(x_{\pi(1)},\ldots,x_{\pi(n)})$ for all $x_1,\ldots,x_n\in X$ and all $\pi\in S_n$,
(iii) $d(x_1,\ldots,x_n)\leqslant\sum_{i=1}^n d(x_i,z)$ for all $x_1,\ldots,x_n,z\in X$.

Properties of multidistances as well as instances such as Fermat multidistance and smallest enclosing ball multidistances have been investigated for example in [1, 5–7]

In this short note we introduce and discuss the following alternative generalization of the concept of metric space by considering the underlying distance as a function of $n \geqslant 2$ variables.

Definition 1. *Let $n \geqslant 2$ be an integer. We say that an n-*metric space *is a pair (X, d), where X is a nonempty set and d is an n-*distance *on X, that is a function $d\colon X^n \to \mathbb{R}_+$ satisfying the following properties:*

(i) $d(x_1, \ldots, x_n) = 0$ *if and only if* $x_1 = \cdots = x_n$,
(ii) $d(x_1, \ldots, x_n) = d(x_{\pi(1)}, \ldots, x_{\pi(n)})$ *for all* $x_1, \ldots, x_n \in X$ *and all* $\pi \in S_n$,
(iii) There exists $K \in [0, 1]$ *such that* $d(x_1, \ldots, x_n) \leqslant K \sum_{i=1}^{n} d(x_1, \ldots, x_n)|_{x_i=z}$ *for all* $x_1, \ldots, x_n, z \in X$.

We denote by K^* the smallest constant K for which (iii) holds. For $n = 2$, we assume that $K^* = 1$.

Clearly, Definition 1 gives an extension of the concept of metric space. Indeed, a function $d\colon X^2 \to \mathbb{R}_+$ is a distance if and only if it is a 2-distance.

We observe that an important feature of n-distances is that they have a fixed number of arguments, contrary to multidistances (see Martín and Mayor [6]), which have an indefinite number of arguments. In particular, an n-distance can be defined without referring to any given 2-distance.

Example 1 (Drastic n-distance). The function $d\colon X^n \to \mathbb{R}_+$ defined by $d(x_1, \ldots, x_n) = 0$, if $x_1 = \cdots = x_n$, and $d(x_1, \ldots, x_n) = 1$, otherwise, is an n-distance, called the *drastic n-distance*, for which the best constant K^* is given by $\frac{1}{n-1}$ for every $n \geqslant 2$. The function $d'\colon X^n \to \mathbb{R}_+$ defined by $d'(x_1, \ldots, x_n) = |\{x_1, \ldots, x_n\}| - 1$ is an n-distance for which the best constant is $K^* = 1$.

Proposition 1. *Let d and d' be n-distances on X and let $\lambda > 0$. The following assertions hold.*

(1) $d + d'$ *and* $\lambda\, d$ *are n-distance on X.*
(2) $\frac{d}{1+d}$ *is an n-distance on X, with value in* $[0, 1]$.

2 A generalization of n-distances

Condition (iii) in Definition 1 can be generalized as follows.

Definition 2. *Let $g\colon \mathbb{R}_+^n \to \mathbb{R}_+$ be a symmetric function. We say that a function $d\colon X^n \to \mathbb{R}^+$ is a g-*distance *if it satisfies conditions (i) and (ii) in Definition 1 as well as the condition*

$$d(x_1, \ldots, x_n) \leqslant g\big(d(x_1, \ldots, x_n)|_{x_1=z}\,, \ldots,\, d(x_1, \ldots, x_n)|_{x_n=z}\big)$$

for all $x_1, \ldots, x_n, z \in X$.

In view of Proposition 1, it is natural to ask that $d+d'$, $\lambda\, d$, and $\frac{d}{1+d}$ be g-distances whenever so are d and d'. The following proposition provides sufficient conditions on g for these properties to hold. We observe that these conditions are rather strong.

Proposition 2. *Let $g\colon \mathbb{R}_+^n \to \mathbb{R}_+$ be a symmetric function. The following assertions hold.*

(1) If g is positively homogeneous, i.e., $g(\lambda\,\mathbf{r}) = \lambda\, g(\mathbf{r})$ for all $\mathbf{r}\in\mathbb{R}_+^n$ and all $\lambda>0$, then for every $\lambda>0$, $\lambda\, d$ is a g-distance whenever so is d.
(2) If g is superadditive, i.e., $g(\mathbf{r}+\mathbf{s}) \geqslant g(\mathbf{r})+g(\mathbf{s})$ for all $\mathbf{r},\mathbf{s}\in\mathbb{R}_+^n$, then $d+d'$ is a g-distance whenever so are d and d'.
(3) If g is both positively homogeneous and superadditive, then it is concave.
(4) If g is bounded from below and additive, that is, $g(\mathbf{r}+\mathbf{s}) = g(\mathbf{r})+g(\mathbf{s})$ for all $\mathbf{r},\mathbf{s}\in\mathbb{R}_+^n$, then and only then there exist $\lambda_1,\ldots,\lambda_n\geqslant 0$ such that

$$g(\mathbf{r}) \;=\; \sum_{i=1}^{n}\lambda_i\, r_i \tag{1}$$

(5) Suppose that g has the form (1) with $\lambda_i \geqslant 1$ for $i=1,\ldots,n$. Then $\frac{d}{1+d}$ is a g-distance whenever so is d.

3 Examples

We end this note by considering a few examples of n-distances that arise in different fields of pure and applied mathematics.

Example 2 (Basic examples). Given a metric space (X,d) and an integer $n\geqslant 2$, the maps $d_{\max}\colon X^n\to\mathbb{R}_+$ and $d_{\Sigma}\colon X^n\to\mathbb{R}_+$ defined by

$$d_{\max}(x_1,\ldots,x_n) = \max_{1\leqslant i<j\leqslant n} d(x_i,x_j)$$
$$d_{\Sigma}(x_1,\ldots,x_n) = \sum_{1\leqslant i<j\leqslant n} d(x_i,x_j)$$

are n-distances for which the best constants are given by $K^* = \frac{1}{n-1}$.

Example 3 (Geometric constructions). Let $x_1,\ldots,x_n$ be $n\geqslant 2$ arbitrary points in $\mathbb{R}^k$ ($k\geqslant 2$) and denote by $B(x_1,\ldots,x_n)$ the smallest closed ball for the Euclidean distance containing $x_1,\ldots,x_n$. It can be shown that this ball always exists, is unique, and can be determined in linear time.

(1) The radius of $B(x_1,\ldots,x_n)$ is an n-distance whose best constant K^* satisfies $K^*\geqslant\frac{1}{n-1}$ and we conjecture that $K^*=\frac{1}{n-1}$.
(2) The k-dimensional volume of $B(x_1,\ldots,x_n)$ is an n-distance and we conjecture that the best constant K^* is given by $K^*=\frac{1}{n-1-(1/2)^k}$. Actually this value for K^* is correct for $k=2$.

Example 4 (Fermat point based n-distances). Given a metric space (X, d), and an integer $n \geqslant 2$, the *Fermat set* F_Y of any n-element subset $Y = \{y_1, \ldots, y_n\}$ of X, is defined as

$$F_Y = \Big\{ x \in X : \sum_{i=1}^{n} d(x_i, x) \leqslant \sum_{i=1}^{n} d(x_i, z) \ \text{ for all } z \in X \Big\}.$$

Since the function $h\colon X \to \mathbb{R}_+$ defined by $h(x) = \sum_{i=1}^{n} d(x_i, x)$ is continuous and bounded from below by 0, the Fermat set of an n-element subset of X is never empty. Hence, we can define a function $d_F\colon X^n \to \mathbb{R}_+$ by setting

$$d_F(x_1, \ldots, x_n) = \min \Big\{ \sum_{i=1}^{n} d(x_i, x) : x \in X \Big\}.$$

Thus defined, the map $d_F\colon X^n \to \mathbb{R}_+$ is an n-distance on X for which the best constant K^* satisfies $K^* \leqslant \frac{1}{\lceil (n-1)/2 \rceil}$.

4 Further research

In this note, we have introduced and discussed an extension of the concept of distance, called n-distance, as functions of n-variables. The key feature of this extension is a natural generalization of the triangle inequality. Finding the best constant for various classes of n-distances and studying their topological properties are topics of current research.

References

1. I. Aguiló, J. Martín, G. Mayor, J. Suñer, and O. Valero. Smallest enclosing ball multidistance. *Communications in Information and Systems* 12(3):185–194,2012.
2. L. M. Blumenthal. *Theory and applications of distance geometry*, (2nd edition). Chelsea Publishing Company, New York, 1970.
3. A. Kołacz and P. Grzegorzewski. Measures of dispersion for multidimensional data. Working paper, 2015.
4. L. Mao. On multi-metric spaces. *Scientia Magna* 2(1):87–94, 2006.
5. J. Martín and G. Mayor. Some properties of multiargument distances and Fermat multidistance. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Theory and Methods. 13th Int. Conf., IPMU 2010, Dortmund, Germany, June 28-July 2, 2010. Proceedings, Part I. Volume 80 of the series Communications in Computer and Information Science. Springer, pp. 703-711, 2010.
6. J. Martín and G. Mayor. Multi-argument distances. *Fuzzy Sets and Systems* 167:92–100, 2011.
7. J. Martín, G. Mayor, and O. Valero. Functionally Expressible Multidistances. Proc. EUSFLAT-LFA Conf. Juillet 2011, Aix-les-Bains, France. Series: Advances in Intelligent Systems Research. Atlantis Press, pp. 41–46, 2011.
8. J. Rintanen. Distance estimates for planning in the discrete belief space. In AAAI'04 Proc. of the 19th national conf. on Artifical intelligence, pp. 525–530, 2004.
9. D. H. Wolpert. Metrics for more than two points at once. In Int. Conf. on Complex Systems 2004, Y. Bar-Yam (Ed.), Perseus books, in press.