SOME ISSUES IN AGGREGATION OF MULTIDIMENSIONAL DATA

Marek Gagolewski
(1) Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
(2) Faculty of Mathematics and Information Science, Warsaw University of Technology
ul. Koszykowa 75, 00-661 Warsaw, Poland
gagolews@ibspan.waw.pl

Summary

The aggregation theory usually takes an interest in summarizing a predefined number of points in the real line. In many applications, like in statistics, data analysis, and mining, the notion of a mean – a nondecreasing, internal, and symmetric fusion function – plays a key role. Nevertheless, when it comes to aggregating a set of points in higher dimensional spaces, the componentwise extension of monotonicity and internality might not be the best choice. Instead, the invariance to certain classes of geometric transformations seems to be crucial in such a case.

Keywords: Aggregation, centroid, Tukey median, 1-center, 1-median, convex hull, affine invariance, orthogonalization.

1 INTRODUCTION

For fixed d, let us consider a fusion function $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ that takes a set of n vectors $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ aggregates them into one vector in $\mathbb{R}^d$. In other words, F is such that:

$$\mathsf{F}\left(\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_d^{(1)} \end{bmatrix}, \dots, \begin{bmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_d^{(n)} \end{bmatrix} \right) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{bmatrix}.$$

From now on we assume that all vectors are column vectors. Hence, we may conceive F as a function acting on a $d \times n$ real matrix,

$$\mathbf{X} = [\mathbf{x}^{(1)}\ \mathbf{x}^{(2)}\ \cdots\ \mathbf{x}^{(n)}].$$

Note that in data analysis, $\mathbf{x}^{(i)}$ is often called an *observation* (and represent an object or experimental unit), whereas $x_j^{(i)}$ denotes the results of measuring the jth variable or feature (like temperature, weight, etc.) of the ith observation (e.g., a person).

Example 1. Let us take any three non-colinear points in $\mathbb{R}^2$. Even in such a simple case many ways to aggregate a triad exist in the literature, see [16]. The notion of a triangle center function, cf. [3], when rewritten in terms of vertex coordinates, leads us to a fusion function which is – among others – rotation and scale invariant (see below). Among the most well-known triangle centers we find the centroid, in-, circum-, and orthocenter. What is interesting, C. Kimberling's *Encyclopedia of Triangle Centers*[1] as of May 1, 2015 lists, names, and characterizes over 7373 such aggregation methods.

Identifying sine qua non conditions that F should fulfill in order to be useful in a particular application area is important, as the class of all fusion functions is of course too broad. The aim of this short contribution is to attract the aggregation theoreticians' attention to the multidimensional data fusion task, which is from time to time explored in the fields like computational statistics and computational geometry. It should be noted that the in-depth study of unidimensional data aggregation methods successfully led to numerous interesting results which allowed to understand many data fusion processes much better, see [1, 13]. We believe that it might also be the case when we assume that $d > 1$.

In Section 2 we recall the typical axiomatization of a *mean* (aggregation function), which – in the $d = 1$ case – is the classic object of interest of the aggregation theory. It turns out that componentwise extensions of two important properties, namely the monotonicity and internality, is not a good choice in a higher di-

[1] Available online at http://faculty.evansville.edu/ck6/encyclopedia/.

mension space. Instead of them, in Section 3 we focus our attention to invariance to particular classes of geometric transformations. We indicate, among others, a quite general way to construct an orthogonal invariant fusion function.

In Section 4 we present a few ideas on how to symmetrize multidimensional fusion functions. As we will see, it is not a trivial task, as for $d > 1$ we do not have a natural linear ordering of the input values. Finally, Section 5 concludes the paper and presents some other issues one may encounter in multidimensional data fusion tasks.

2 MULTIDIMENSIONAL EXTENSIONS OF MONOTONICITY AND INTERNALITY

As it was already noted, in the aggregation theory, we mostly focus on the $d = 1$ case. The notion of a mean (internal aggregation function) $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}$, which may be used to determine the "most typical observation" among a given set of values typically requires the fulfillment of the three following properties (see [1, 13] and also [4, 18, 25]):

- symmetry, i.e. for any permutation σ of the set $\{1, \ldots, n\}$ it holds:
$$\mathsf{F}(x_1^{(1)}, \ldots, x_1^{(n)}) = \mathsf{F}(x_1^{(\sigma(1))}, \ldots, x_1^{(\sigma(n))}),$$
- nondecreasingness, which requires that whenever $x_1^{(i)} \le x_1'^{(i)}$ for all $i = 1, \ldots n$, we have:
$$\mathsf{F}(x_1^{(1)}, \ldots, x_1^{(n)}) \le \mathsf{F}(x_1'^{(1)}, \ldots, x_1'^{(n)}),$$
- internality, that is:
$$\mathsf{F}(x_1^{(1)}, \ldots, x_1^{(n)}) \in \left[\bigwedge_{i=1}^n x_1^{(i)}, \bigvee_{i=1}^n x_1^{(i)}\right].$$

Note that, under monotonicity, internality is equivalent to idempotence, i.e., $(\forall x \in \mathbb{R})$ it holds $\mathsf{F}(x, x, \ldots, x) = x$. Also please notice that the choice of the above properties, as well as the set $\mathbb{R}$ in which the vectors' elements reside is particularly sound in statistics and data analysis. From now on we assume in this paper that only these two practical domains attract our interest. This is because, e.g., a decision making or fuzzy reasoning task is of a much different nature.

Let us extend the above properties in such a way that they are valid for any d. Symmetry is the least problematic one: we may simply assume that $\mathsf{F}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathsf{F}(\mathbf{x}^{(\sigma(1))}, \ldots, \mathbf{x}^{(\sigma(n))})$ must hold for any σ. The easiest and perhaps the most natural approach to extend the other two is by applying them in a componentwise manner.

First of all, note that the ordering structure on $\mathbb{R}$ may easily be extended to $\mathbb{R}^d$ by determining the so-called *product order*. The partial order $\le_d$ is defined so as to for $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ we have $\mathbf{x} \le_d \mathbf{x}'$ if and only if $(\forall i \in \{1, \ldots, d\})$ $x_i \le x_i'$. This leads to the concept of $\le_d$ (componentwise)-nondecreasingness. Such an approach is often used when the topic of aggregation on products of lattices/chains is explored, see e.g. [5, 7, 19].

On the other hand, componentwise internality may be defined by requiring that $\mathsf{F}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$ belongs to the set:
$$\left[\bigwedge_{i=1}^n x_1^{(i)}, \bigvee_{i=1}^n x_1^{(i)}\right] \times \cdots \times \left[\bigwedge_{i=1}^n x_d^{(i)}, \bigvee_{i=1}^n x_d^{(i)}\right],$$
which is basically the bounding (hyper)rectangle of a given set of input points, compare also [27].

Here are two exemplary fusion functions that fulfill these three properties.

Example 2. The componentwise extension of the arithmetic mean,
$$\mathsf{CwMean}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_1^{(i)} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n x_d^{(i)} \end{bmatrix},$$
also called the *centroid* (barycenter, geometric center) of a set of points, is a point such that
$$\mathbf{y} = \arg\min_{\mathbf{y}\in\mathbb{R}^d} \sum_{i=1}^n \mathfrak{d}_2(\mathbf{x}^{(i)}, \mathbf{y})^2,$$
where $\mathfrak{d}_2$ is the Euclidean distance. This notion is crucial e.g. in the definition of the k-means [22] clustering algorithm.

Example 3. The componentwise extension of the sample median, Med, namely:
$$\mathsf{CwMed}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \begin{bmatrix} \mathsf{Med}\left(x_1^{(1)}, \ldots, x_1^{(n)}\right) \\ \vdots \\ \mathsf{Med}\left(x_d^{(1)}, \ldots, x_d^{(n)}\right) \end{bmatrix},$$
is sometimes used, cf. [30], as a robust estimate of a multidimensional probability distribution's median.

Yet, the following fusion functions are not $\le_d$-nondecreasing.

Example 4. The *Euclidean 1-center* (smallest enclosing ball, bounding sphere) problem aims at finding:

$$\mathsf{1center}_{\mathfrak{d}_2}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}) = \arg\min_{\mathbf{y}\in\mathbb{R}^d} \bigvee_{i=1}^{n} \mathfrak{d}_2(\mathbf{x}^{(i)},\mathbf{y}).$$

It was first proposed by James Sylvester in 1857 [31]. Such a formulation is used in many real-world applications, see e.g. [11], like patter recognition (finding reference points), computational biology (proteine analysis), graphics (ray tracing, culling), data mining (e.g., support vector machines, high-dimensional clustering, nearest neighbor search). This is also the case of the facility location problem, which aims to seek for a location of a distribution center that minimizes the distance to a customer that is situated the farthest away. Unfortunately, there is no analytic solution to the Euclidean 1-center problem, cf. [12] for a discussion and an algorithm.

Euclidean 1-center is not componentwise monotone. Consider $n = 3$ and $d = 2$ with $\mathbf{x}^{(1)} = [1,-1]^T$, $\mathbf{x}^{(2)} = [-1,1]^T$, $\mathbf{x}^{(3)} = [-\sqrt{2},0]^T$. We have $\mathsf{1center}_{\mathfrak{d}_2}(\dots) = [0,0]^T$. Letting $\mathbf{x}'^{(1)} = \mathbf{x}^{(1)} + [3,0]^T$ we get $\mathsf{1center}_{\mathfrak{d}_2}(\dots) \approx [1.3,-0.5]^T \not\geq_2 [0,0]^T$.

Example 5. Ca. 1650, Evangelista Torricelli proposed a solution to a problem posed by Pierre de Fermat in the early 17th century: given three points in a plane, find the fourth point for which the sum of its distances to the three given points is as small as possible, cf. [20]. This task can be formulated for arbitrary number of points as follows. Find $\mathbf{y}$ such that:

$$\mathsf{1median}_{\mathfrak{d}_2}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}) = \arg\min_{\mathbf{y}\in\mathbb{R}^d} \sum_{i=1}^{n} \mathfrak{d}_2(\mathbf{x}^{(i)},\mathbf{y}).$$

Such a point, called in the literature the *1-median*, geometric median, spatial median, L_1-median, Fermat-Weber, or Torricelli point, generalizes the concept of a one-dimensional median (assuming that the minimum is unique). In statistics, these are also known as L_1 estimators. Again, no analytic formula is known here.

Also 1-median is not componentwise monotone. Take $d = 2$, $n = 3$, and $\mathbf{x}^{(1)} = [0,0]^T$, $\mathbf{x}^{(2)} = [0,1]^T$, $\mathbf{x}^{(3)} = [1,0]^T$. We have $\mathsf{1median}_{\mathfrak{d}_2}(\dots) \simeq [0.211,0.211]^T$. However, when we take $\mathbf{x}'^{(3)} = \mathbf{x}^{(3)} + [0,2]^T$, then we get $\mathsf{1median}_{\mathfrak{d}_2}(\dots) = [0,1]^T \not\geq_2 [0.211,0.211]^T$.

Example 6. Tukey [32] introduced the concept of the *halfplane location depth* of $\mathbf{y}$ relative to a given set of points in $\mathbb{R}^d$. It is the smallest number of $\mathbf{x}^{(i)}$'s contained in any closed halfhyperplane with boundary line through $\mathbf{y}$. In other words:

$$\begin{aligned}&\mathrm{tdepth}_d(\mathbf{y};\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})\\ =\ &\min\nolimits_{\mathbf{u}\in\mathbb{R}^d,|\mathbf{u}|=1} |\{i : \mathbf{u}^T\mathbf{x}^{(i)} \geq \mathbf{u}^T\mathbf{y}\}|.\end{aligned}$$

As the deepest point in $d = 1$ generalizes the concept of a median, a deepest value in higher dimensions can be thought of as a multidimensional median: the center of gravity of the deepest depth region is called the *Tukey median*, TkMed. In fact, a bagplot, a bivariate version of the box-and-whisker plot, bases on such a notion [29]. For other multidimensional generalizations of the median, like the Oja or the Liu medians, please refer e.g. to [28, 30].

Tukey median is not componentwise monotone. Consider $n = 4$ and $d = 2$ with $\mathbf{x}^{(1)} = [0,0]^T$, $\mathbf{x}^{(2)} = [1,0]^T$, $\mathbf{x}^{(3)} = [1,1]^T$, and $\mathbf{x}^{(4)} = [0,1]^T$. We have $\mathsf{TkMed}(\dots) = [0.5,0.5]^T$. Letting $\mathbf{x}'^{(4)} = \mathbf{x}^{(4)} + [1,0]^T$ we get $\mathsf{TkMed}(\dots) = [2/3,1/3]^T \not\geq_2 [0.5,0.5]^T$.

What is more, monotonicity is not the only property that is somehow problematic. It may be observed that the above extension of internality is not a necessarily nice generalization of ordinary internality. Even though all the above-presented fusion functions fulfill it, it seems to be too weak. Let $d = 2$, $n = 3$ and consider $\mathbf{x}^{(1)} = [1,0]^T$, $\mathbf{x}^{(2)} = [0,0]^T$, $\mathbf{x}^{(3)} = [0,1]^T$. If a fusion function CwG is a componentwise extension of, e.g., $\mathsf{G}(y_1,\dots,y_n) = \bigvee_{i=1}^n y_i$, then $\mathsf{CwG}(\mathbf{x}^{(1)},\mathbf{x}^{(2)},\mathbf{x}^{(3)}) = [1,1]^T$. Wee see that this is rather not plausible.

Remark 7. The Euclidean 1-center, 1-median, centroid, and componentwise median may be expressed as minimizers of some penalty function. A generalization of the latter two – componentwise cases – have been studied in a product lattice framework in [5].

3 TRANSLATION, SCALE, ORTHOGONAL, AND AFFINE INVARIANCE

Instead of considering monotonicity and internality, researchers in fields like computational statistics and geometry most often focus on invariances with respect to specific classes of geometrical transformations, see e.g. [9]. Namely, one might be interested in finding a fusion function F which fulfills for all input vectors:

- translation invariance: for all $\mathbf{t}\in\mathbb{R}^d$,
$$\mathsf{F}(\mathbf{x}^{(1)}+\mathbf{t},\dots,\mathbf{x}^{(n)}+\mathbf{t}) = \mathsf{F}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}) + \mathbf{t},$$
- scale invariance: for all $s\in\mathbb{R}$,
$$\mathsf{F}(s\mathbf{x}^{(1)},\dots,s\mathbf{x}^{(n)}) = s\mathsf{F}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}),$$
- orthogonal invariance: for all orthogonal matrices $\mathbf{A}\in\mathbb{R}^{d\times d}$, i.e. matrices such that $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$, or equivalently $\mathbf{A}^T = \mathbf{A}^{-1}$,
$$\mathsf{F}(\mathbf{A}\mathbf{x}^{(1)},\dots,\mathbf{A}\mathbf{x}^{(n)}) = \mathbf{A}\mathsf{F}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}),$$

and/or

- affine invariance: for all matrices $\mathbf{A} \in \mathbb{R}^{d\times d}$ of full rank and all $\mathbf{t} \in \mathbb{R}^d$,

$$\mathsf{F}(\mathbf{A}\mathbf{x}^{(1)}+\mathbf{t},\dots,\mathbf{A}\mathbf{x}^{(n)}+\mathbf{t}) = \mathbf{A}\mathsf{F}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})+\mathbf{t}.$$

Note that orthogonal invariance implies invariance to all possible rotations of input points (and reflections against the axes). On the other hand, affine invariance implies translation, scale, and orthogonal invariance. Moreover, it also covers the case of scaling the coordinates differently in each direction. It is quite useful, as in the practice of data analysis, one often standardizes the variables:

$$x_i \mapsto \frac{x_i - \bar{x}_i}{\mathsf{sd}_i},$$

where $\bar{x}_i$ and sd_i denotes the arithmetic mean and standard deviation, respectively, with respect to the ith coordinate, $i = 1,\dots,d$. Also, some machine learning methods (like the Principal Component Analysis) assume that the data points may freely be rotated.

It can be noted that, e.g., the centroid and the Tukey median (as well as the Oja and Liu medians) are affine invariant. On the other hand, e.g., the componentwise median is not even orthogonal invariant, see Table 1. In fact, from [13, Proposition 2.116] it follows that the only continuous, symmetric, and componentwise fusion function is CwMean.

Table 1: Exemplary fusion functions and the properties they fulfill: M – componentwise monotonicity, T – translation, S – scale, O – orthogonal, and A – affine invariance.

Function	M	T	S	O	A
CwMean	✓	✓	✓	✓	✓
CwMed	✓	✓	✓		
TkMed		✓	✓	✓	✓
1center		✓	✓	✓	
1median		✓		✓	
OrMed		✓	✓	✓	

3.1 Orthogonalization

Orthomedian [14] is a nice example of orthogonalization of the componentwise median. It is defined as an averaged median of all orthogonally transformed input data sets. As the group $\mathcal{O}(d)$ of orthogonal $d \times d$ matrices is compact, we may define:

$$\begin{aligned}&\mathsf{OrMed}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})\\ &= \int_{\mathcal{O}(d)} \mathbf{A}^{-1}\mathsf{CwMed}(\{\mathbf{A}\mathbf{x}^{(i)}\}_{i=1,\dots,n})\,d\mathbf{A}.\end{aligned}$$

This fusion function is orthogonal invariant (by construction). Interestingly, it is no more $\le_d$-nondecreasing, so this new property is introduced at some cost.

The above construction is general and can be applied on any fusion function CwG that is a componentwise extension of $\mathsf{G} : \mathbb{R}^n \to \mathbb{R}$. However, it is not easy to compute numerically (e.g., Monte Carlo methods may be used for this purpose, see [8, Sec. 3] for a random uniform – with respect to the Haar measure, see [8] – orthogonal matrix generation algorithm). Thus, here we propose another approach which is valid if $(\forall x_i)$ $\mathsf{G}(x_1,\dots,x_n) = -\mathsf{G}(-x_1,\dots,-x_n)$.

Let $\mathbf{X} = [\mathbf{x}^{(1)}\ \mathbf{x}^{(2)}\ \cdots\ \mathbf{x}^{(n)}]$ and assume that: $\mathbf{X}_c = \mathbf{X} - \mathsf{CwMean}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})$ is a centered version of $\mathbf{X}$. Let us consider the singular value decomposition of $\mathbf{X}_c^T = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where $\mathbf{U}$ is an $n\times n$ orthogonal matrix, $\mathbf{D}$ is a $n \times d$ diagonal matrix, and $\mathbf{V}$ is a $d \times d$ orthogonal matrix. The eigenvectors $\mathbf{v}^{(i)}$ are called *principal component* directions of $\mathbf{X}_c$, see e.g. [15, Sec. 3.4 and Sec. 14.5]. The first principal component direction $\mathbf{v}^{(1)}$ has the property that $\mathbf{z}^{(1)} = \mathbf{X}_c^T\mathbf{v}^{(1)}$ is of the largest sample variance, d_1^2/n among all normalized linear combinations of $\mathbf{X}_c$'s rows. Subsequent principal components have maximum variance subject to being orthogonal to the earlier ones. Thus,

$$\begin{aligned}&\mathsf{OrG}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})\\ =&\mathbf{V}^{-1^T}\mathsf{CwG}\left(\mathbf{V}^T(\mathbf{X} - \mathsf{CwMean}(\mathbf{X}))\right) + \mathsf{CwMean}(\mathbf{X})\end{aligned}$$

is surely orthogonal and translation invariant.

As for the monotonicity, if G is nondecreasing, then OrG is nondecreasing with respect to the direction that has the maximal variance (and other directions that orthogonal to it and also maximize the remaining variance). Also, if G is internal, then G fulfills the bounding box-based internality.

3.2 A note on internality

Let us recall the notion of the convex hull $\mathrm{CH}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)})$ of a finite set of points. It is the smallest convex set (polytope) that includes all the provided points. Equivalently, it is the set of all convex combinations of $\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}$:

$$\mathrm{CH}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}) = \Big\{ \textstyle\sum_{i=1}^n w_i\mathbf{x}^{(i)} : \text{for all vectors } \mathbf{w} \ge \mathbf{0} \text{ with } \sum_{i=1}^n w_i = 1\Big\}.$$

Having in mind that for $d = 1$ the convex hull is a real interval, the definition of internality may be extended to $d \ge 2$ by using this very notion, which seems far more adequate. Here, we could require that

$$\mathsf{F}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}) \in \mathrm{CH}(\mathbf{x}^{(1)},\dots,\mathbf{x}^{(n)}).$$

Interestingly, it might be shown that if $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ is rotation and translation invariant, then the bounding-box based internality and the convex hull-based one coincide. This is because the convex hull is invariant to rotations and translations and that it is a subset of the bounding box. Moreover, the convex hull may be expressed as the intersection of appropriate halfspaces [10]. The points may always be rotated so that any convex hull's face is aligned within the axes. Then the hyperplane that includes such a face coincides with the hyperplane including the bounding box's face.

4 ORDERED FUSION FUNCTIONS

Given a non-symmetric unidimensional function, one may easily symmetrize it by referring to the notion of an order statistic, i.e. the ith smallest value among a set of input elements. It is because, cf. [13, Thm. 2.34], $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}$ is symmetric if and only if there exists a function $\mathsf{G} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mathsf{F}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = \mathsf{G}(\mathbf{x}^{(\sigma(1))}, \dots, \mathbf{x}^{(\sigma(n))}),$$

where σ is an ordering permutation the input values. In such a way, e.g., a weighted arithmetic mean becomes the OWA operator. Such a construction is only valid, however, in the $d = 1$ case, as here a natural linear order $\le$ is defined.

If $d > 1$, then it is not easy to determine which values are "small" or "large", especially if we allow a set of points to be orthogonally transformed.

For this purpose, one may order the input values with respect to increasing distances from a fixed point, e.g., the set's componentwise mean. More elaborate approaches may base on the concept of the so-called data depth. For instance, the affine invariant *Oja depth* [26] (or the simplical volume depth) for any given $\mathbf{y} \in \mathbb{R}^d$ is given by:

$$\begin{aligned} &\mathrm{odepth}(\mathbf{y}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) \\ = &\ \frac{1}{1 + \sum_{(i_1,\dots,i_d)} \mathrm{volume}(\mathrm{CH}(\mathbf{y}, \mathbf{x}^{(i_1)}, \dots, \mathbf{x}^{(i_d)}))}. \end{aligned}$$

Other affine invariant data depth measures include the already mentioned Tukey depth, simplical depth [21]:

$$\begin{aligned} &\mathrm{sdepth}(\mathbf{y}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) \\ = &\ \left|\left\{(i_1, \dots, i_{d+1}) : \mathbf{y} \in \mathrm{CH}(\mathbf{x}^{(i_1)}, \dots, \mathbf{x}^{(i_{d+1})})\right\}\right| \end{aligned}$$

or the zonoid data depth [9]:

$$\begin{aligned} &\mathrm{zdepth}(\mathbf{y}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) \\ = &\ \sup\{\alpha \in [0, 1] : \mathbf{y} \in D_\alpha(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})\}, \end{aligned}$$

where D_α is the α*-trimmed region* of the empirical distribution generated by $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$, i.e.:

$$\begin{aligned} &D_\alpha(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) \\ = &\ \left\{\textstyle\sum_{i=1}^n \lambda_i \mathbf{x}^{(i)} : \sum_{i=1}^n \lambda_i = 1, (\forall i) \alpha\lambda_i \le 1/n\right\}. \end{aligned}$$

With them, the points $\mathbf{x}^{(i)}$, $i = 1, \dots, n$, may be ordered with respect to their decreasing depths. In other words, we may make use of a permutation σ of $\{1, \dots, n\}$ such that $\sigma(i) \le \sigma(j)$ implies that for $i < j$:

$$\mathrm{d}(\mathbf{x}^{(\sigma(i))}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) \le \mathrm{d}(\mathbf{x}^{(\sigma(j))}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}),$$

where d is some data depth measure.

Having an ordered version of the input set of points, one may easily define, e.g., multidimensional versions of trimmed or winsorized means, cf. [24].

5 CONCLUSIONS

We briefly explored a few issues in aggregation of multidimensional data. We recalled some interesting fusion functions used in computational statistics and geometry.

It turns out that some of the main ideas of unidimensional aggregation cannot be simply extended to the multidimensional case. The most important one concerns the componentwise monotonicity. Notably, recently the necessity of the notion of monotonicity is being put into question in the classical framework too, see e.g. [2, 6]. In the $d = 1$ case it seems quite natural and moreover it simplifies the way the analytic results are derived. Thus, further studies concerning this notion for $d > 1$ shall be conducted.

Instead of focusing on monotonicity, it seems that researchers in computational statistics and geometry rather focus on invariances to particular classes of geometric transformations. In this contribution, for example, we presented a simple way to guarantee translation and orthogonal invariance. Also note that a recent contribution on measures of dispersion of multidimensional data [17] (some of them base on the notion of multidistances' [23] minimizers) also focuses on translation and rotation invariance.

We additionally noted that some other concepts, like internality or a fusion function's symmetrization also need some more elaborate approaches.

Another problem with multidimensional fusion functions is that many of the tools encountered in the literature cannot be expressed by analytic formulas. This drastically complicates the theoretical studies on them. Even though it is possible to characterize somehow all unidimensional fusion functions that fulfill some of the properties discussed here, things definitely get more complicated in higher dimensions.

REFERENCES

[1] G. Beliakov, A. Pradera, and T. Calvo (2007). *Aggregation functions: A guide for practitioners.* Springer-Verlag.

[2] G. Beliakov, T. Calvo, and T. Wilkin (2014). Three types of monotonicity of averaging functions. *Knowledge-Based Systems* **72**, pp. 114–122.

[3] O. Bottema (1981). Het begrip "merkwaardig" met betrekking tot punten in de driehoeksmeetkunde. *Nieuw Tijdschr. Wisk.* **69**, pp. 2–7.

[4] P. Bullen (2003). *Handbook of means and their inequalities.* Springer Science, Dordrecht.

[5] H. Bustince, E. Barrenechea, T. Calvo, S. James, and G. Beliakov (2014). Consensus in multi-expert decision making problems using penalty functions defined over a cartesian product of lattices. *Information Fusion* **17**, pp. 56–64.

[6] H. Bustince, J. Fernandez, A. Kolesárová, and R. Mesiar (2014). Fusion functions and directional monotonicity. *Communications in Computer and Information Science* **444**, pp. 262–268.

[7] B. De Baets and R. Mesiar (1999). Triangular norms on product lattices. *Fuzzy Sets and Systems* **104**, pp. 61–75.

[8] P. Diaconis and M. Shahshahani (1987). The subgroup algorithm for generating uniform random variables. *Probability In Engineering And Information Sciences* **1**, pp. 15–32.

[9] R. Dyckerhoff, G. Koshevoy, and K. Mosler (1996). Zonoid data depth: Theory and computation. In A. Prat et al. (Eds.), *Proc. COMPSTAT 1996*, Heidelberg, Physica-Verlag, pp. 235–240.

[10] H. Edelsbrunner (1987). *Algorithms in Combinatorial Geometry.* Springer-Verlag, Heidelberg.

[11] B. Gärtner (1999). Fast and robust smallest enclosing balls. *Lecture Notes in Computer Science* **1643**, pp. 325–338.

[12] B. Gärtner and S. Schönherr (2000). An efficient, exact, and generic quadratic programming solver for geometric optimization. In *Proc. 16th ACM Symposium on Computational Geometry*, pp. 110–118.

[13] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap (2009). *Aggregation functions.* Cambridge University Press.

[14] R. Grübel (1996). Orthogonalization of multivariate location estimators: The orthomedian. *The Annals of Statistics* **24**, pp. 1457–1473.

[15] T. Hastie, R. Tibshirani, and J. Friedman (2013). *The Elements of Statistical Learning.* Springer.

[16] C. Kimberling (1998). Triangle centers and central triangles. *Congressus Numerantium* **129**, pp. 1–295.

[17] A. Kołacz and P. Grzegorzewski (2015). Measures of dispersion for multidimensional data, submitted paper.

[18] A. Kolmogorov (1930). Sur la notion de la moyenne. *Atti della R. Academia nazionale dei Lincei* **12**, pp. 388–391.

[19] M. Komorníková and R. Mesiar (2011). Aggregation functions on bounded partially ordered sets and their classification. *Fuzzy Sets and Systems* **175**, pp. 48–56.

[20] J. Krarup and S. Vajda (1997). On Torricelli's geometrical solution to a problem of Fermat. *IMA Journal of Management Mathematics* **8**, pp. 215–223.

[21] R. Y. Liu (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18**, pp. 405–414.

[22] J. B. MacQueen (1967). Some methods for classification and analysis of multivariate observations. In *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, vol. 1, Berkeley, University of California Press, pp. 281–297.

[23] J. Martin and G. Mayor (2011). Multi-argument distances. *Fuzzy Sets and Systems* **167**, pp. 92–100.

[24] J.-C. Massé (2009). Multivariate trimmed means based on the Tukey depth. *Journal of Statistical Planning and Interference* **139**, pp. 366–384.

[25] M. Nagumo (1930). Über eine Klasse der Mittelwerte. *Japanese Journal of Mathematics* **7**, pp. 71–79.

[26] H. Oja (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters* **1**, pp. 327–332.

[27] S. Ovchinnikov (1996). Means on ordered sets. *Mathematical Social Sciences* **32**, pp. 39–56.

[28] P. J. Rousseeuw and A. Struyf (2004). Computation of robust statistics: Depth, median, and related measures. In J. E. Goodman and J. ORourke (Eds.), *The Handbook of Discrete and Computational Geometry*, Chapman & Hall/CRC, Boca Raton, pp. 1279–1292.

[29] P. J. Rousseeuw, I. Ruts, and J. W. Tukey (1999). The bagplot: A bivariate boxplot. *The American Statistician* **53**, pp. 382–387.

[30] C. G. Small (1990). A survey of multidimensional medians. *International Statistical Review* **58**, pp. 263–277.

[31] J. J. Sylvester (1857). A question in the geometry of situation. *Quarterly Journal of Mathematics* **1**, p. 79.

[32] J. W. Tukey (1974). Mathematics and the picturing of data. *Proc. Intl. Congress of Mathematicians*, pp. 523–531.