Penalty-based aggregation of multidimensional data

Marek Gagolewski[a,b]

[a]*Systems Research Institute, Polish Academy of Sciences*
ul. Newelska 6, 01-447 Warsaw, Poland
[b]*Faculty of Mathematics and Information Science, Warsaw University of Technology*
ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract

Research in aggregation theory is nowadays still mostly focused on algorithms to summarize tuples consisting of observations in some real interval or of diverse general ordered structures. Of course, in practice of information processing many other data types between these two extreme cases are worth inspecting. This contribution deals with the aggregation of lists of data points in $\mathbb{R}^d$ for arbitrary $d \geq 1$. Even though particular functions aiming to summarize multidimensional data have been discussed by researchers in data analysis, computational statistics and geometry, there is clearly a need to provide a comprehensive and unified model in which their properties like equivariances to geometric transformations, internality, and monotonicity may be studied at an appropriate level of generality. The proposed penalty-based approach serves as a common framework for all idempotent information aggregation methods, including componentwise functions, pairwise distance minimizers, and data depth-based medians. It also allows for deriving many new practically useful tools.

Keywords: multidimensional data aggregation, penalty functions, data depth, centroid, median

1. Introduction

Aggregation theory [5, 9, 32, 38] focuses on a formal analysis of functions that, given a set of objects of the same kind, output a single item which is (in some sense) representative of all the inputs. Till very recently, functions like $\mathsf{F} : \mathbb{I}^n \to \mathbb{I}$ for some $\mathbb{I} = [a, b]$ and $n \geq 2$, fulfilling key application-specific

---

properties like nondecreasingness, internality, conjunctivity, etc., were the most common objects of interest in this field. It is important to stress that particular aggregation functions were of course known long before aggregation theory became a genuine branch of applied mathematics and information science. In particular, *means* or *averaging functions* [5, 11] – that is functions that are at least idempotent – include the famous arithmetic mean, median, quasi-arithmetic means, and OWA [74] operators, to name just a few. However, the emergence of this domain allowed to seek common patterns bracketing diverse ways to handle information overload as well as understand data aggregation processes much better. Apart from aggregation on bounded posets, e.g., [21, 44, 56], from quite recently, we finally start to observe a growing interest in aggregation *of* and *on* other practically useful structures, see, e.g., the papers [58, 60] concerning the problem of combining rankings.

In this paper we are interested in functions that aim to aggregate a sequence of n numeric lists $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ for a fixed $d \geq 1$. Each such $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ can be written as:

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

Equivalently, we may conceive F as a function acting on a $d \times n$ matrix like:

$$
\mathbf{X} = [\mathbf{x}^{(1)}\, \mathbf{x}^{(2)}\, \cdots\, \mathbf{x}^{(n)}] = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_d \end{bmatrix}.
$$

Note that in data analysis, $\mathbf{x}^{(i)} \in \mathbb{R}^d = \mathbb{R}^{d \times 1}$ is typically called an observation – it designates an object or experimental unit (e.g., a person, autonomous vehicle, spatial location). On the other hand, $\mathbf{x}_j \in \mathbb{R}^{1 \times n}$ denotes the j-th variable or feature (such as temperature, weight, or velocity).

The concept of a penalty function based on data in the real line, i.e., with $d = 1$, was first introduced by Yager [73], and then extended in numerous works, see, e.g., [8, 15, 16, 75]; in particular, [13] gives the most recent summary together with a critical historical overview. Its aim is to measure the amount of "disagreement" between the inputs and the output being computed. Such a frame-

2

work provides a very appealing way to define new *means*: it can be shown that minimizers of some penalty coincide with the class of idempotent functions [13].

Our aim here is to propose a penalty-based framework for idempotent functions that act on observations in $\mathbb{R}^d$ for an arbitrary $d \geq 1$, significantly extending our previous preliminary studies [32, 33]. In the next section the notion of a penalty function generalizing the classical one is proposed. Basic desired properties of penalty-based mappings are discussed in Sect. 3. In Sect. 4, 5, and 6 we present three noteworthy classes of aggregation methods, respectively: componentwise extensions of unidimensional functions, those constructed upon pairwise distances between observations, and those defined by means of the notion of data depth. Section 7 concludes the paper and discusses some concepts which are frequently considered in classical aggregation theory, but in the extended setting are difficult to maintain. This includes, among others, the notion of monotonicity and orness measures.

2. Penalty-based framework

Given an arbitrary $x \in \mathbb{R}$, with $(d * x)$ we denote a d-tuple $(x, x, \ldots, x) \in \mathbb{R}^d$. Binary operations like $+, -, \cdot, /, \wedge$ (minimum), and $\vee$ (maximum) on vectors of equal lengths d are applied elementwise and thus output a vector of length d too. On the other hand, if one of the operands is a scalar, then it is extended to a vector of length d in such a way that, e.g., $\mathbf{x} + t = \mathbf{x} + (d * t)$. If $\mathbf{A} \in \mathbb{R}^{d \times n}$ is a matrix with d rows and n columns and $\mathbf{t} \in \mathbb{R}^d$, then by, e.g., $\mathbf{A} + \mathbf{t}$ we mean $\mathbf{A} + [\mathbf{t} \, \mathbf{t} \, \cdots \, \mathbf{t}]$, i.e., $\mathbf{t}$ is treated as a column vector. Moreover, $\mathbf{A} + t = \mathbf{A} + (d * t) = \mathbf{A} + [(d * t) \, \cdots \, (d * t)]$. Finally, for some $n \in \mathbb{N}$ we denote with $[n]$ the set $\{1, 2, \ldots, n\}$ and with $\| \cdot \|_p$ the L^p-norm on $\mathbb{R}^d$.

Having established the notation convention used throughout the paper, let us introduce the notion of a penalty function relative to given $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$.

Definition 1. We call $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ a *penalty function*, whenever:

(a) $P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = 0$ if and only if $\mathbf{x}^{(i)} = \mathbf{y}$ for all $i \in [n]$;

(b) for every fixed $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ the minimum set of $P(\cdot; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$, i.e., $\left\{ \mathbf{y} \in \mathbb{R}^d : P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \inf_{\mathbf{y}' \in \mathbb{R}^d} P(\mathbf{y}'; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \right\}$, is nonempty, bounded, and convex.

In the $d = 1$ case we of course allow for the minimum set of $P(\cdot; \mathbf{X})$ to be either a singleton or a real interval. Therefore, we can say that the above definition generalizes the classical one (as given in [8]; see also [13]).

In order to be able to generate data aggregation methods based on the notion of a penalty function, let us denote with $\mathrm{CoG}(M) = (m_1, \ldots, m_d)$ the center of gravity (centroid, geometric center) of a nonempty, bounded, and convex set $M \subseteq \mathbb{R}^d$. For instance, if M is full-dimensional, then $m_i = \int_M x_i \, d\mathbf{x} / \int_M d\mathbf{x}$. In particular, it is worth noting that centroids of convex polytopes are quite convenient to compute numerically for small d (this, however, depends on their representation, see [30, 59]), and that for $d = 1$ it holds $\mathrm{CoG}([a, b]) = (a + b)/2$, $a \leq b$. Moreover, it always holds $\mathrm{CoG}(M) \in M$.

Definition 2. Given a penalty function $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$, the corresponding *P-based function* is defined as:

$$\mathsf{F}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \arg^* \min_{\mathbf{y} \in \mathbb{R}^d} P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}),$$

where $\arg^* \min_{\mathbf{y} \in \mathbb{R}^d} P(\mathbf{y}; \mathbf{X}) := \mathrm{CoG}\left(\left\{\mathbf{y} \in \mathbb{R}^d : P(\mathbf{y}; \mathbf{X}) = \inf_{\mathbf{y}' \in \mathbb{R}^d} P(\mathbf{y}'; \mathbf{X})\right\}\right)$ denotes the center of gravity of the minimum set of $P(\cdot; \mathbf{X})$.

Let us indicate some results which will be useful in the sequel. First of all, we should point out that penalty functions are defined up to – among others – some strictly increasing transforms.

Lemma 3. *Let $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ be a penalty function. Associate with each $\mathbf{X} \in (\mathbb{R}^d)^n$ a strictly increasing bijection $\varphi_{\mathbf{X}} : [0, \infty] \to [0, \infty]$ and define $P' : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ in such a way that for any $\mathbf{y}, \mathbf{X}$ it holds $P'(\mathbf{y}; \mathbf{X}) = \varphi_{\mathbf{X}}(P(\mathbf{y}; \mathbf{X}))$. Then P' is also a penalty function and the P-based function is equivalent to the P'-based one.*

Proof. The satisfaction of condition (a) in Def. 1 is evident. Thus, now fix $\mathbf{X} \in (\mathbb{R}^d)^n$. We have $\varphi_{\mathbf{X}}\left(\inf_{\mathbf{y}} P(\mathbf{y}; \mathbf{X})\right) = \inf_{\mathbf{y}} \varphi_{\mathbf{X}}(P(\mathbf{y}; \mathbf{X}))$. Therefore, $\left\{\mathbf{y} : P(\mathbf{y}; \mathbf{X}) = \inf_{\mathbf{y}'} P(\mathbf{y}'; \mathbf{X})\right\} = \left\{\mathbf{y} : \varphi_{\mathbf{X}}(P(\mathbf{y}; \mathbf{X})) = \inf_{\mathbf{y}'} \varphi_{\mathbf{X}}(P(\mathbf{y}'; \mathbf{X}))\right\}$. □

What is more, we shall see that some penalty functions can easily be obtained by studying mappings which are – among others – convex with respect to $\mathbf{y}$. Observe that, by definition, for any penalty function P and all $\mathbf{X} \in (\mathbb{R}^d)^n$, the effective domain of $P(\cdot; \mathbf{X})$, that is $\left\{\mathbf{y} \in \mathbb{R}^d : P(\mathbf{y}; \mathbf{X}) < \infty\right\}$, is not empty. As its codomain does not include $-\infty$, each $P(\cdot; \mathbf{X}) \geq 0$ is what we call in convex analysis a *proper* function, see [49, 61]. Yet, we should keep in mind that this time we act on an unbounded domain, whilst when aggregating unidimensional data one typically

assumes that all observations are in a bounded real interval, e.g., $[0, 1]$ or $[-1, 1]$. Therefore, apart from convexity, we need some additional, but very natural, conditions. An exemplary set of such properties is given in the following lemma.

Lemma 4. *Let* $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ *be such that* $P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = 0$ *if and only if* $\mathbf{x}^{(i)} = \mathbf{y}$ *for all* $i \in [n]$. *For every* $\mathbf{X} \in (\mathbb{R}^d)^n$ *if:*

- $P(\cdot; \mathbf{X})$ *is convex;*

- $P(\mathbf{y}; \mathbf{X}) < \infty$ *for all* $\mathbf{y} \in \mathbb{R}^d$;

- $\lim_{\|\mathbf{y}\| \to \infty} P(\mathbf{y}; \mathbf{X}) = \infty$;

then P *is a penalty function.*

Proof. Fix $\mathbf{X} \in (\mathbb{R}^d)^n$. By [49, Corollary 2.1.3], $P(\cdot; \mathbf{X})$ is everywhere continuous and hence surely lower semicontinuous. By [49, Theorem 4.1.3], the minimum set of $P(\cdot; \mathbf{X})$ is not empty and based on [61, Theorem 4.6] – convex. As $\inf_{\mathbf{y}} P(\mathbf{y}; \mathbf{X}) < \infty$ and $\lim_{\|\mathbf{y}\| \to \infty} P(\mathbf{y}; \mathbf{X}) = \infty$, we conclude that it must be bounded too, which completes the proof. $\qquad \square$

3. Desired properties

3.1. Idempotence

It is straightforward to see that each P-based function F is *idempotent*, i.e., such that for all $\mathbf{x} \in \mathbb{R}^d$ it holds $\mathsf{F}(\mathbf{x}, \ldots, \mathbf{x}) = \mathbf{x}$. What is more, just like in the $d = 1$ case (see [13]), we have the following much stronger result.

Theorem 5. $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ *is a P-based function for some penalty* $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ *if and only if it is idempotent.*

Proof. ($\Longrightarrow$) Follows from condition (a) in Def. 1 and the fact that $\mathrm{CoG}(\{\mathbf{x}\}) = \mathbf{x}$.

($\Longleftarrow$) Assume that F is idempotent and consider $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ defined for any $\mathbf{y}, \mathbf{X}$ as $P(\mathbf{y}; \mathbf{X}) = \|\mathsf{F}(\mathbf{X}) - \mathbf{y}\|_2 + \|\mathsf{V}(\mathbf{X})\|_2$, where $\mathsf{V} : (\mathbb{R}^d)^n \to [0, \infty]^d$ denotes the componentwise extension of the sample variance, $\mathsf{V}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathbf{v}$ with $n^2 v_i = \sum_{j=1}^{n} \sum_{k=j}^{n} \left(x_i^{(j)} - x_i^{(k)} \right)^2$ for all $i \in [d]$. For every fixed $\mathbf{X}$, the minimum set of $P(\cdot; \mathbf{X})$ is the singleton $\{\mathsf{F}(\mathbf{X})\}$, with $\inf_{\mathbf{y}} P(\mathbf{y}; \mathbf{X}) = \|\mathsf{V}(\mathbf{X})\|_2$. As $\mathsf{V}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathbf{0}$ if and only if $\mathbf{x}^{(j)} = \mathbf{x}^{(k)}$ for all $j, k \in [n]$ we have that $P(\mathbf{y}; \mathbf{X}) = 0$ if and only if $\mathbf{y} = \mathbf{x}^{(j)}$ for all $j \in [n]$, and the proof is complete. $\qquad \square$

On a side note, the $\|\mathsf{V}(\mathbf{X})\|_2$ term (which was inspired by the one in [13]) could be used to relax the assumption about P in Lemma 4.

5

3.2. *Equivariances to transforms*

As noted in [33], researchers in data analysis, computational statistics and geometry – fields from which most well-known multidimensional idempotent aggregation tools originate – are typically interested in properties that deal with *equivariances* to different geometrical transformations, see, e.g., [66]. Intuitively, equivariance means that by transforming a data set we obtain an aggregated value which follows the same transformation[1].

For instance, one frequently applies various transformations at the data cleansing or data wrangling step. In particular, standardization involves translation by a vector of componentwise means and scaling by the reciprocal of the square root of componentwise variances. On the other hand, principal component analysis (PCA) relies on some data set rotation.

Before we present some general results which deal with such types of behavior, we should emphasize that the center of gravity of a given set is equivariant to affine transformations.

Lemma 6. *Given a convex set $M \subseteq \mathbb{R}^d$, for all matrices $\mathbf{A} \in \mathbb{R}^{d \times d}$ of full rank and $\mathbf{t} \in \mathbb{R}^d$ it holds $\mathrm{CoG}(\mathbf{A}M + \mathbf{t}) = \mathrm{CoG}(\{\mathbf{A}\mathbf{x} + \mathbf{t} : \mathbf{x} \in M\}) = \mathbf{A}\,\mathrm{CoG}(M) + \mathbf{t}$.*

Proof. Assume that $\mathrm{CoG}(M) = (m_1, \ldots, m_d)$ and $\mathrm{CoG}(M') = (m'_1, \ldots, m'_d)$, where the convex set $M' = \mathbf{A}M + \mathbf{t}$. We shall only prove that the lemma holds for a full-dimensional M, as a similar reasoning is required in cases in which M lies on a d'-dimensional hyperplane for some $d' < d$. We have:

$$
\begin{aligned}
m'_i &= \frac{\int_{M'} x_i \, dx_1 \cdots dx_d}{\int_{M'} dx_1 \cdots dx_d} = \frac{\int_M |\det(\mathbf{A})| \left(\sum_{j=1}^n a_{ij} x_j + t_i \right) dx_1 \cdots dx_d}{\int_M |\det(\mathbf{A})| \, dx_1 \cdots dx_d} \\
&= \sum_{j=1}^n a_{ij} \frac{\int_M x_j \, dx_1 \cdots dx_d}{\int_M dx_1 \cdots dx_d} + t_i = \sum_{j=1}^n a_{ij} m_j + t_i.
\end{aligned}
$$

Thus, $\mathrm{CoG}(\mathbf{A}M + \mathbf{t}) = \mathbf{A}\,\mathrm{CoG}(M) + \mathbf{t}$, and the lemma follows. $\qquad\square$

Proposition 7. *Let a penalty function $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ be such that for all matrices $\mathbf{A} \in \mathbb{R}^{d \times d}$ of full rank and $\mathbf{t} \in \mathbb{R}^d$ there exists a strictly increasing*

---

[1]On the other hand, in our case *in*variance would mean that the aggregated value is insensitive to transformations of a particular kind. Note that the term *equi*variance is frequently used in data analysis literature (e.g., [66]), whereas aggregation theory (e.g., [38]) would refer to the very same property as an *in*variance. Yet, in this paper we need a clear distinction between these two.

bijection $\varphi_{\mathbf{A},\mathbf{t}} : [0,\infty] \to [0,\infty]$ *such that for every* $\mathbf{y} \in \mathbb{R}^d$ *and* $\mathbf{X} \in (\mathbb{R}^d)^n$ *it holds* $\varphi_{\mathbf{A},\mathbf{t}}(P(\mathbf{A}\mathbf{y}+\mathbf{t};\mathbf{A}\mathbf{X}+\mathbf{t})) = P(\mathbf{y};\mathbf{X})$. *Then the P-based function is* affine equivariant, *i.e.,* $\mathsf{F}(\mathbf{A}\mathbf{X}+\mathbf{t}) = \mathbf{A}\mathsf{F}(\mathbf{X}) + \mathbf{t}$.

Proof. Take any $\mathbf{A}, \mathbf{t}$, and $\mathbf{X}$. By Lemma 3 and the fact that affine transforms of convex sets preserve convexity, we have:

$$\mathbf{t} + \mathbf{A}\mathsf{F}(\mathbf{X}) = \mathbf{t} + \mathbf{A}\arg^*\min_{\mathbf{y}\in\mathbb{R}^d} P(\mathbf{y};\mathbf{X}) = \mathbf{t} + \mathbf{A}\arg^*\min_{\mathbf{y}\in\mathbb{R}^d}\varphi_{\mathbf{A},\mathbf{t}}(P(\mathbf{A}\mathbf{y}+\mathbf{t};\mathbf{A}\mathbf{X}+\mathbf{t}))$$

$$= \mathbf{t} + \mathbf{A}\arg^*\min_{\mathbf{y}\in\mathbb{R}^d} P(\mathbf{A}\mathbf{y}+\mathbf{t};\mathbf{A}\mathbf{X}+\mathbf{t}) = \arg^*\min_{\mathbf{y}\in\mathbb{R}^d} P(\mathbf{y};\mathbf{A}\mathbf{X}+\mathbf{t}) = \mathsf{F}(\mathbf{t}+\mathbf{A}\mathbf{X}).$$

$\square$

Remark 8. The above proposition also works if we restrict ourselves to some specific families of affine transforms, including:

- *translation equivariance* – for all $\mathbf{t} \in \mathbb{R}^d$, $\mathsf{F}(\mathbf{X}+\mathbf{t}) = \mathsf{F}(\mathbf{X}) + \mathbf{t}$;

- *uniform-scale equivariance* – for all $s > 0$, $\mathsf{F}(s\mathbf{X}) = s\mathsf{F}(\mathbf{X})$;

- *d-scale equivariance* – for all $\mathbf{s} \in]0,\infty[^d$, $\mathsf{F}(\mathbf{s}\mathbf{X}) = \mathbf{s}\mathsf{F}(\mathbf{X})$;

- *orthogonal equivariance* – for all orthogonal matrices $\mathbf{O} \in \mathbb{R}^{d\times d}$, $\mathsf{F}(\mathbf{O}\mathbf{X}) = \mathbf{O}\mathsf{F}(\mathbf{X})$;

as well as their combinations.

Note that the class of orthogonal transformations consists of all rotations around the point $\mathbf{0} \in \mathbb{R}^d$ (whenever $\det\mathbf{O} = 1$), reflections against the axes, and their compositions. Each affine equivariant function is also equivariant to all *similarity transforms*, that is uniform-scale, translation, and orthogonal transformations.

3.3. Various internalities

As far as unidimensional functions are concerned, some authors (see, e.g., [5, 11]) require a stronger version of idempotence in order to call F an *averaging function* or a *mean*. Namely, *internality*[2] states that a function's output value must lie "somewhere in-between" the input values. According to [38], already Cauchy in 1821 considered under a name mean an internal function; similarly, Gini in the 1950s needed only this very property when discussing various means; see [5].

---

[2]Internality, see [38, Definition 2.53], is sometimes referred to as compensativity.

In the $d = 1$ case, internality is defined by requiring for any $\mathbf{X}$ that $\mathsf{F}(\mathbf{X}) \in \left[\bigwedge_{i=1}^{n} x_1^{(i)}, \bigvee_{i=1}^{n} x_1^{(i)} \right] = [\mathsf{Min}(\mathbf{x}_1), \mathsf{Max}(\mathbf{x}_1)]$. For arbitrary $d \geq 1$, this property can be generalized in a componentwise manner. First, let us introduce the notion of the bounding box of $\mathbf{X}$:

$$\mathrm{BB}(\mathbf{X}) = \left[\bigwedge_{i=1}^{n} x_1^{(i)}, \bigvee_{i=1}^{n} x_1^{(i)} \right] \times \cdots \times \left[\bigwedge_{i=1}^{n} x_d^{(i)}, \bigvee_{i=1}^{n} x_d^{(i)} \right].$$

Definition 9. We call $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ a *BB-internal* function, whenever $\mathsf{F}(\mathbf{X}) \in \mathrm{BB}(\mathbf{X})$ for all $\mathbf{X} \in (\mathbb{R}^d)^n$.

However, unidimensional internality can also be generalized as follows. Denote with $\mathrm{CH}(\mathbf{X})$ the convex hull of a point set $\mathbf{X}$, i.e., the set generated by all convex combinations of observations in $\mathbf{X}$:

$$\mathrm{CH}(\mathbf{X}) = \left\{ \sum_{i=1}^{n} w_i \mathbf{x}^{(i)} : (w_1, \ldots, w_n) \in [0, 1]^n, \sum_{i=1}^{n} w_i = 1 \right\}.$$

Definition 10. We call $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ a *CH-internal* function, whenever $\mathsf{F}(\mathbf{X}) \in \mathrm{CH}(\mathbf{X})$ for all $\mathbf{X} \in (\mathbb{R}^d)^n$.

As for $d = 1$ it holds that $\mathrm{CH}(\mathbf{X}) = \mathrm{BB}(\mathbf{X})$ is the smallest real interval that contains all the points in $\mathbf{X}$, BB- and CH-internality are both indeed generalizations of the ordinary one. Moreover, for arbitrary d, it is easily seen that CH-internality (which is particularly appealing in data analysis) implies BB-internality, which in turn implies idempotence. Interestingly, under rotation equivariance, BB- and CH-internality coincide.

Lemma 11. *Let* $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ *be rotation equivariant and such that for any* $\mathbf{X} \in (\mathbb{R}^d)^n$ *it holds* $\mathsf{F}(\mathbf{X}) \in \mathrm{BB}(\mathbf{X})$. *Then* $\mathsf{F}(\mathbf{X}) \in \mathrm{CH}(\mathbf{X})$.

Proof. Fix $\mathbf{X} \in (\mathbb{R}^d)^n$ and let $\mathcal{R}$ denote the set of all d-dimensional rotation matrices. It is well-known that the convex hull of a set of points may be expressed as an intersection of appropriately constructed closed halfspaces, see [29]. In fact, as any bounding box can be written as an intersection of some $2d$ closed halfspaces and we may always find $\mathbf{R} \in \mathcal{R}$ such that an arbitrarily chosen face of $\mathrm{CH}(\mathbf{X})$ is a subset of some face of $\mathbf{R}^{-1}\mathrm{BB}(\mathbf{RX})$, we have $\mathrm{CH}(\mathbf{X}) = \bigcap_{\mathbf{R} \in \mathcal{R}} \mathbf{R}^{-1}\mathrm{BB}(\mathbf{RX})$. For any $\mathbf{R} \in \mathcal{R}$, by rotation equivariance and bounding-box internality, we have that $\mathsf{F}(\mathbf{X}) = \mathbf{R}^{-1}\mathsf{F}(\mathbf{RX}) \in \mathbf{R}^{-1}\mathrm{BB}(\mathbf{RX})$. Thus, $\mathsf{F}(\mathbf{X}) \in \mathrm{CH}(\mathbf{X})$, which is precisely the assertion of the proposition. $\square$

3.4. Other

Apart from idempotence, internalities, and equivariances we should point out a few other properties that may be desirable in certain applications. First of all, the classical *symmetry* is easily generalizable to a case of $d \geq 1$: we say that F obeys such a property, whenever for all $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)} \in \mathbb{R}^d$ and every permutation σ of the set $[n]$ it holds that $\mathsf{F}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathsf{F}(\mathbf{x}^{\sigma(1)}, \ldots, \mathbf{x}^{\sigma(n)})$. The same is true as far as, e.g., *continuity* is concerned.

We should point out that orthogonal equivariance implies different kinds of symmetries too – one with respect to a permutation of rows of an input matrix (the original one refers to arbitrary rearrangements of columns) and one about the point $\mathbf{0}$.

On the other hand, in Sect. 7.1 we shall provide a few remarks about monotonicity, which is very frequently assumed when $d = 1$, but – as we will see – is problematic for $d > 1$. Until then, we shall say that F is *componentwise-nondecreasing*, whenever for all $\mathbf{X}, \mathbf{X}' \in (\mathbb{R}^d)^n$ and any $j \in [d]$, if for every $i \in [n]$ it holds $x_j^{(i)} \leq x_j^{(i)'}$, then $\mathsf{F}(\mathbf{X})_j \leq \mathsf{F}(\mathbf{X}')_j$.

In the three consecutive sections we review different noteworthy classes of idempotent functions.

4. Componentwise extensions of unidimensional functions

The first class of penalty-based functions studied in this paper consists of *componentwise*[3] extensions of d classical, unidimensional functions – each of them is applied on a separate data dimension (variable). Here is how we can combine d classical penalty functions in order to obtain one that acts on $\mathbb{R}^d$.

Proposition 12. *Let* $\mathsf{A} : [0, \infty]^d \to [0, \infty]$ *be a function that is strictly increasing in each variable and fulfills* $\mathsf{A}(d * 0) = 0$. *For any d unidimensional penalty functions* $P_1, \ldots, P_d : \mathbb{R}^1 \times (\mathbb{R}^1)^n \to [0, \infty]$, *the mapping* $P : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, \infty]$ *given by:*

$$P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathsf{A}\left(P_1(y_1; x_1^{(1)}, \ldots, x_1^{(n)}), \ldots, P_d(y_d; x_d^{(1)}, \ldots, x_d^{(n)})\right)$$

---

[3] Such functions are referred to as of *decomposable* type in [44].

9

is a penalty function. Moreover, denoting the P_i-based unidimensional function with $\mathsf{F}_i : \mathbb{R}^n \to \mathbb{R}$, $i \in [d]$, the corresponding P-based function is such that:

$$\mathsf{F}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = \begin{bmatrix} \mathsf{F}_1(x_1^{(1)}, \dots, x_1^{(n)}) \\ \vdots \\ \mathsf{F}_d(x_d^{(1)}, \dots, x_d^{(n)}) \end{bmatrix}.$$

Proof. By strict increasingness of A, $P(\mathbf{y}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = 0$ if and only if for all $i \in [d]$ it holds $P_i(y_i; x_i^{(1)}, \dots, x_i^{(n)}) = 0$; by Def. 1 in the $d = 1$ case, this is equivalent to stating that $\mathbf{y} = \mathbf{x}^{(j)}$ for every $j \in [n]$.

For any fixed $\mathbf{X} \in (\mathbb{R}^d)^n$, let $M_i \subseteq \mathbb{R}$ be the minimum set of $P_i(\cdot; \mathbf{x}_i)$, $i \in [d]$, and $M \in \mathbb{R}^d$ the minimum set of $P(\cdot; \mathbf{X})$. First of all, we shall show that $\inf_{\mathbf{y} \in \mathbb{R}^d} P(\mathbf{y}; \mathbf{X}) = \mathsf{A}(\inf_{y_1 \in \mathbb{R}} P_1(y_1; \mathbf{x}_1), \dots, \inf_{y_d \in \mathbb{R}} P_d(y_d; \mathbf{x}_d))$. Assume otherwise: given $\mathbf{u} \in [0, \infty]^d$ such that $\inf_{\mathbf{y} \in \mathbb{R}^d} P(\mathbf{y}; \mathbf{X}) = \mathsf{A}(u_1, \dots, u_d)$ and $v_i = \inf_{y_i \in \mathbb{R}} P_i(y_i; \mathbf{x}_i)$, let $\mathsf{A}(u_1, \dots, u_d) < \mathsf{A}(v_1, \dots, v_d)$. Due to the strict increasingness of A, we have that for some $k \in [d]$ it holds $u_k < v_k$. However, this would imply that $u_k < \inf_{y_k \in \mathbb{R}} P_k(y_k; \mathbf{x}_k)$, a contradiction.

Now for any $y_1, \dots, y_d$ with $y_i \in M_i$, $i \in [d]$, we clearly have that $(y_1, \dots, y_d) \in M$. Moreover, given any $\mathbf{y}' \in M$, for all i it holds $y_i' \in M_i$. Therefore, $M = M_1 \times \cdots \times M_d$. As, by Def. 1, each M_i is either a singleton or a real interval, M is obviously a hyperrectangle (which is a particular nonempty, bounded, and convex set) and therefore P is a penalty function. In this case it holds that $\mathrm{CoG}(M) = (\mathrm{CoG}(M_1), \dots, \mathrm{CoG}(M_d))$. Thus, we have that $\mathsf{F}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = (\mathsf{F}_1(\mathbf{x}_1), \dots, \mathsf{F}_d(\mathbf{x}_d))$, which completes the proof. $\square$

Here, the resulting idempotent function does not depend on the actual choice of A. Therefore, from now on we can call it a $(P_1, \dots, P_d)$-based function or even simpler – a *componentwise* one.

Example 13. Among noteworthy instances of the discussed mappings we find:

- the *componentwise extension of the median* defined as:

$$\mathrm{CwMedian}(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = \begin{bmatrix} \mathrm{Median}\left(x_1^{(1)}, \dots, x_1^{(n)}\right) \\ \vdots \\ \mathrm{Median}\left(x_d^{(1)}, \dots, x_d^{(n)}\right) \end{bmatrix} = \begin{bmatrix} \mathrm{Median}(\mathbf{x}_1) \\ \vdots \\ \mathrm{Median}(\mathbf{x}_d) \end{bmatrix},$$

where $\mathrm{Median}(u_1, \dots, u_n) = (u_{\lfloor (n+1)/2 \rfloor} + u_{\lceil (n+1)/2 \rceil})/2$. As Median is a P_i-based function with, e.g., $P_i(y_i; \mathbf{x}_i) = \sum_{j=1}^{n} |x_i^{(j)} - y_i|$, $i \in [d]$, we may obtain $\mathrm{CwMedian}$ by considering, e.g., $P(\mathbf{y}; \mathbf{X}) = \sum_{i=1}^{d} \sum_{j=1}^{n} |x_i^{(j)} - y_i|$;

- the *componentwise extension of the arithmetic mean* generated by, e.g., a penalty function $P(\mathbf{y}; \mathbf{X}) = \sum_{i=1}^{d} \sum_{j=1}^{n} (x_i^{(j)} - y_i)^2$.

More generally, componentwise weighted medians and arithmetic means may easily be introduced.

In the componentwise setting, some of the already known results on unidimensional penalty-based functions, see, e.g., [6–8, 10, 12, 14–16, 51, 71, 75] can be extrapolated to our extended framework. Nevertheless, we should be careful, as many of the facts in classical aggregation theory are limited to tuples with elements in [0, 1] or [0, ∞[– in our case we did not introduce such limits on the input elements' domain as they do not get along well with geometric transformations listed above.

For instance, it is known (see, e.g., [15, 51]) that a weighted quasi-arithmetic mean with a continuous and strictly-monotone generator function $\varphi : \mathbb{R} \to \bar{\mathbb{R}}$ and a weighting vector $\mathbf{w} \in [0, 1]^n$, $\sum_{i=1}^{n} w_i = 1$, that is:

$$\mathsf{QAM}_{\varphi, \mathbf{w}}(x_1^{(1)}, \ldots, x_1^{(n)}) = \varphi^{-1}\left(\sum_{i=1}^{n} w_i \varphi(x_1^{(i)})\right),$$

is a P_1-based function with, e.g., $P_1(y; \mathbf{x}_1) = \sum_{i=1}^{n} w_i (\varphi(x_1^{(i)}) - \varphi(y))^2$. We note that some of the "classical" generators [11] are only valid whenever aggregated values are nonnegative. In particular, this is how we may extend the well-known results showed in [54]:

- If $\varphi(x) = \exp(\gamma x)$ for some $\gamma \neq 0$, we obtain an exponential mean. Exponential means and the arithmetic mean are the only translation equivariant quasi-arithmetic means.

- If $\varphi(x) = \operatorname{sign}(x)|x|^p$ for some $p \neq 0$ (a signed power function), then we get a power mean. Power means are the only scale equivariant quasi-arithmetic means (originally, Nagumo's result includes the geometric mean too, but this function is not valid for elements generally in $\mathbb{R}$).

Taking these facts into account, we may easily obtain componentwise extensions of weighted quasi-arithmetic means that are translation or d-scale equivariant. If we require both types of equivariances, we must restrict ourselves to componentwise extensions of weighted arithmetic means (or consider, e.g., weighted medians).

Let us also note that in the componentwise setting, d-scale equivariance is equivalent to the uniform-scale one. Moreover, if we are interested in componentwise and – among others – orthogonal equivariant functions, then we are restricted to aggregation methods based on weighted arithmetic means.

Proposition 14. *A BB-internal and continuous componentwise function is orthogonal equivariant if and only if it is a componentwise extension of a weighted arithmetic mean.*

Proof. ($\Longleftarrow$) Obvious.

($\Longrightarrow$) Let us express $\mathsf{F}(\mathbf{X})$ as $(\mathsf{F}_1(\mathbf{x}_1), \ldots, \mathsf{F}_d(\mathbf{x}_d))$. By the definition of orthogonal equivariance, we have that:

$$\mathsf{F}_i\left(\sum_{j=1}^{d} o_{i,j} x_j^{(1)}, \ldots, \sum_{j=1}^{d} o_{i,j} x_j^{(n)}\right) = \sum_{j=1}^{d} o_{i,j} \mathsf{F}_j(x_j^{(1)}, \ldots, x_j^{(n)}),$$

holds for all orthogonal $\mathbf{O} \in \mathbb{R}^{d \times d}$, $i \in [d]$, as well as $\mathbf{X} \in (\mathbb{R}^d)^n$.

First of all, let us note that the set of all orthogonal matrices includes those which permute coordinate axes, i.e., which are formed by permuting the rows of the identity matrix. Hence, for arbitrary $i \in [d]$, considering $(o_{i,1}, \ldots, o_{i,d}) = (1, 0, \ldots, 0)_\sigma$ for every permutation σ of the set $[d]$, we conclude that:

$$\mathsf{F}_i\left(x_j^{(1)}, \ldots, x_j^{(n)}\right) = \mathsf{F}_j\left(x_j^{(1)}, \ldots, x_j^{(n)}\right),$$

necessarily holds for all $i, j \in [d]$ and $\mathbf{X}$. Hence, under our assumptions, F must be a componentwise extension of a single unidimensional, internal, and continuous function F_1.

Take any $i \in [d]$ and consider $\mathbf{X}$ such that $x_j^{(k)} = 0$ for all $k \in [n]$ and $j \neq i$. Orthogonal equivariance implies that:

$$\mathsf{F}_1\left(o_{i,i} x_i^{(1)}, \ldots, o_{i,i} x_i^{(n)}\right) = o_{i,i} \mathsf{F}_1\left(x_i^{(1)}, \ldots, x_i^{(n)}\right),$$

holds for every $o_{i,i} \in [-1, 1]$ (recall that rows and columns of an orthogonal matrix are orthogonal unit vectors) and all $x_i^{(1)}, \ldots, x_i^{(n)}$.

Next, again for some fixed $i \in [d]$, we consider an orthogonal matrix $\mathbf{O}$ such that $o_{i,j} = 1/\sqrt{d}$ for all $j \in [d]$. From the equivariance of our interest and the fact

that $1/\sqrt{d} \in [-1, 1]$, we have:

$$\mathsf{F}_1\left(\sum_{j=1}^{d} x_j^{(1)}/\sqrt{d}, \dots, \sum_{j=1}^{d} x_j^{(n)}/\sqrt{d}\right) = \sum_{j=1}^{d} \mathsf{F}_1(x_j^{(1)}, \dots, x_j^{(n)})/\sqrt{d}$$

$$= \sum_{j=1}^{d} \mathsf{F}_1(x_j^{(1)}/\sqrt{d}, \dots, x_j^{(n)}/\sqrt{d}).$$

Denoting with $y_j^{(k)} = x_j^{(k)}/\sqrt{d}$, we have that:

$$\mathsf{F}_1\left(\sum_{j=1}^{d} y_j^{(1)}, \dots, \sum_{j=1}^{d} y_j^{(n)}\right) = \sum_{j=1}^{d} \mathsf{F}_1(y_j^{(1)}, \dots, y_j^{(n)}),$$

which implies that F_1 is necessarily additive. Together with continuity, by [38, Proposition 2.116], we have that $\mathsf{F}_1(\mathbf{x}) = \sum_{i=1}^{n} c_i x_i$ with $c_1, \dots, c_n \in \mathbb{R}$. Taking into account the internality condition, we have that $\sum_{i=1}^{n} c_i = 1$ (because $\mathsf{F}_1(x, \dots, x) = \sum_{i=1}^{n} c_i x = x$) and $c_1, \dots, c_n \in [0, 1]$ (from the fact that $\mathsf{F}_1(0, \dots, 0, x_i, 0, \dots, 0) = c_i x_i \in [0, x_i]$). Thus, F_1 is a weighted arithmetic mean. $\square$

Hence, the only symmetric, BB-internal, continuous, and orthogonal equivariant componentwise function is formed by extending the (unweighted) arithmetic mean.

On a side note, there exist various orthogonalization methods, i.e., transformations that make a given function orthogonal equivariant, compare the notion of the orthomedian [40] as well as the idea of the singular value decomposition-based orthogonalization [32, 33]. Unfortunately, results of such transformations typically lead to functions that are no longer componentwise ones. Moreover, similarly to orthogonalization, affinitization of a function can be applied if needed. Its special case for a componentwise extension of the Median function was first proposed in [18], see also [53].

Concerning componentwise-nondecreasingness, we have what follows.

Proposition 15. $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ *is componentwise-nondecreasing if and only if it is a componentwise extension of some d nondecreasing in each variable unidimensional functions.*

Proof. ($\Longleftarrow$) Straightforward.

($\Longrightarrow$) Let us express $\mathsf{F}(\mathbf{X})$ as $(\mathsf{F}_1(\mathbf{X}), \dots, F_d(\mathbf{X}))$ with $\mathsf{F}_1, \dots, \mathsf{F}_d : (\mathbb{R}^d)^n \to \mathbb{R}$. For an arbitrary $i \in [d]$, we shall show that F_i is in fact only a function of $x_i^{(1)}, \dots, x_i^{(n)}$.

For any fixed $x_i^{(1)}, \ldots, x_i^{(n)}$ and each $\mathbf{X}', \mathbf{X}''$ with $\mathbf{x}_i = \mathbf{x}_i' = \mathbf{x}_i''$, the definition of componentwise-nondecreasingness immediately implies that $\mathsf{F}_i(\mathbf{X}') = \mathsf{F}_i(\mathbf{X}'') =: \tilde{\mathsf{F}}_i(x_i^{(1)}, \ldots, x_i^{(n)})$. Moreover, $\tilde{\mathsf{F}}_i$ is clearly nondecreasing in each variable, and the proof is complete. □

Also please notice that in Proposition 14, if we consider componentwise-nondecreasingness together with idempotence instead of BB-internality and continuity, then by [38, Proposition 2.116] we will reach exactly the same conclusion as in the original setting. Componentwise-nondecreasingness with orthogonal equivariance implies nondecreasingness in each possible direction, but in the case of idempotent functions this can only be obtained when we apply a componentwise extension of a weighted arithmetic mean.

5. Distance-based penalties and corresponding functions

In the case of componentwise functions, there are no interactions between the variables. This can be undesirable in certain applications. Thus, let us investigate a second noteworthy class of penalty functions which is formed by a proper aggregation of distances from a point $\mathbf{y}$ to each observation in $\mathbf{X}$.

Here, by the term *distance* [22] on $\mathbb{R}^d$ we mean a function $\eth : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty[$ for which it holds at least that $\eth(\mathbf{y}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$ and $\eth(\mathbf{y}, \mathbf{x}) = \eth(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. If a distance fulfills the triangle inequality, it is called a metric.

Definition 16. Let $\eth$ be an arbitrary distance on $\mathbb{R}^d$. Given a function $\mathsf{D} : [0, \infty[^n \to [0, \infty]$ that is nondecreasing in each variable and such that $\mathsf{D}(\mathbf{d}) = 0$ if and only if $\mathbf{d} = \mathbf{0}$, we shall call:

$$P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = \mathsf{D}\left(\eth(\mathbf{y}, \mathbf{x}^{(1)}), \ldots, \eth(\mathbf{y}, \mathbf{x}^{(n)})\right) \tag{2}$$

a *distance*-based penalty function, whenever for each $\mathbf{X} \in (\mathbb{R}^d)^n$ the minimum set of $P(\cdot; \mathbf{X})$ is nonempty, bounded, and convex.

Clearly, not all distances as well as aggregation functions D lead to proper penalties: this is the case, e.g., of the sums of Hamming metrics. Here we shall be interested in a particularly well-behaving class of distances generated by a norm (note that if $\eth$ is convex and fulfills the triangle inequality, it must be generated by some norm, see [41, 72] for a proof).

Proposition 17. *Take any norm* $\| \cdot \| : \mathbb{R}^d \to [0, \infty[$ *and let* $\eth(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. *If a function* $\mathsf{D} : [0, \infty[^n \to [0, \infty]$ *enjoys the following properties:*

- *nondecreasingness in each variable;*

- $D(\mathbf{d}) = 0$ *if and only if* $\mathbf{d} = \mathbf{0}$*;*

- *convexity;*

- $D(\mathbf{d}) < \infty$ *for all* $\mathbf{d} \in [0, \infty[^n$*;*

- $\lim_{\|\mathbf{d}\| \to \infty} D(\mathbf{d}) = \infty$*;*

then $P(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) = D\left(\mathfrak{d}(\mathbf{y}, \mathbf{x}^{(1)}), \ldots, \mathfrak{d}(\mathbf{y}, \mathbf{x}^{(n)})\right)$ *is a penalty function.*

Proof. Condition (a) in Def. 1 is straightforward. By homogeneity of a norm, we have that $\lim_{\|\mathbf{y}\| \to \infty} \mathfrak{d}(\mathbf{y}, \mathbf{x}^{(i)}) = \infty$ for any fixed $\mathbf{x}^{(i)}$. Convexity of $\mathfrak{d}(\cdot, \mathbf{x}^{(i)})$ as well as convexity and nondecreasingness of D immediately imply the convexity of $P(\cdot; \mathbf{X})$. By applying Lemma 4, we approach the desired conclusion about the minimum set of P. $\qquad\square$

Typical choices of D include:

- $D(\mathbf{d}) = \bigvee_{i=1}^{n} d_i$ (the maximum);

- $D(\mathbf{d}) = \sum_{i=1}^{n} w_i d_i$, where $w_1, \ldots, w_n > 0$ and $\sum_{i=1}^{n} w_i = 1$ (weighted arithmetic means strictly increasing in each variable);

and their generalizations:

- $D(\mathbf{d}) = \sum_{i=1}^{n} w_i d_{(i)}$ with $0 \le w_i \le \cdots \le w_n \le 1$ and $\sum_{i=1}^{n} w_i = 1$ (convex OWA operators);

- convex weighted quasi-arithmetic means with strictly positive weights, e.g., p-power means with $p \ge 1$ and exponential means.

Here are a few examples of distance-based functions.

Example 18. If $D(d_1, \ldots, d_n) = \sqrt{\sum_{i=1}^{n} \frac{1}{n} d_i^2}$ (the quadratic mean) and $\mathfrak{d}(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2$ (the Euclidean metric), then the corresponding function is known as the *centroid*, which is exactly the aforementioned componentwise extension of the arithmetic mean, CwAMean. More generally, a weighted quadratic mean leads to a componentwise extension of a weighted arithmetic mean.

15

If $\mathsf{D}(d_1, \ldots, d_n) = \sum_{i=1}^{n} \frac{1}{n} d_i$ (the arithmetic mean) and $\eth(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_1$ (the Manhattan metric), then in turn we get the *Manhattan 1-median*, which corresponds to the componentwise median, $\mathsf{CwMedian}$.

On the other hand, with $\mathsf{D}(d_1, \ldots, d_n) = \bigvee_{i=1}^{n} d_i$ (the sample maximum) and $\eth(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_\infty$ (the Chebyshev metric) we obtain the *Chebyshev 1-center*, which is the componentwise extension of the midrange, $\mathsf{F}_1(\mathbf{x}) = (\mathsf{Max}(\mathbf{x}) + \mathsf{Min}(\mathbf{x}))/2$, i.e., it is the center of the points' bounding rectangle.

Example 19. Let $\mathsf{D}(d_1, \ldots, d_n) = \sum_{i=1}^{n} \frac{1}{n} d_i$ (arithmetic mean) and $\eth(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2$ (the Euclidean metric). Here, the corresponding function is known as the *Euclidean 1-median*, geometric median, spatial median, mediancenter, L^1-median, Fermat-Weber, or Torricelli point. In the unidimensional case, it is equivalent to the sample median and thus the minimizer of the penalty might not be unique. However, for $d \geq 2$ and $\mathbf{X}$ such that it is not concentrated on a line, the minimum set of the corresponding distance-based penalty function is always a singleton [52].

A weighted spatial median can be considered too, i.e., with D being a weighted arithmetic mean: $\mathsf{1median_w}(\mathbf{X}) = \arg^* \min_{\mathbf{y} \in \mathbb{R}^d} \sum_{i=1}^{n} w_i \|\mathbf{y} - \mathbf{x}^{(i)}\|_2$. Unfortunately, the analytic solution is only known in trivial cases like those in which one weight is equal to 1. Yet, it may be shown, see [69], that a weighted 1-median is a point $\mathbf{y}$ such that:

$$\sum_{i=1}^{n} \frac{w_i \mathbf{y}}{\|\mathbf{y} - \mathbf{x}^{(i)}\|_2} = \sum_{i=1}^{n} \frac{w_i \mathbf{x}^{(i)}}{\|\mathbf{y} - \mathbf{x}^{(i)}\|_2}.$$

This equation stands for a basis of the iterative Weiszfeld (Vázsonyi) algorithm [70] which in practice is frequently applied in order to compute this function.

Example 20. Let $\mathsf{D}(d_1, \ldots, d_n) = \bigvee_{i=1}^{n} d_i$ (the sample maximum) and again $\eth(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2$ (the Euclidean metric). The corresponding function – giving the center of the smallest ball enclosing all input points – is called the *Euclidean 1-center*, see [67]. Notably, such an aggregation tool is used in many real-world applications, see, e.g., [36], which include: pattern recognition (finding reference points), computational biology (protein analysis), graphics (ray tracing, culling, object collision detection), and nearest neighbor search. What is more, for $d = 2$ we have an important operational research application, known as the facility location problem, which aims to seek the location of the distribution center that minimizes the distance to a customer that is situated farthest away.

It may be shown that the Euclidean 1-center can be expressed as $\mathbf{y} = \sum_{i=1}^{n} v_i \mathbf{x}^{(i)}$, where $\mathbf{v}$ is the solution to the quadratic programming task minimize $\mathbf{v}^T \mathbf{X}^T \mathbf{X} \mathbf{v} -$

$(\text{diag}(\mathbf{X}^T\mathbf{X}))^T\mathbf{v}$ w.r.t. $\mathbf{v}$ subject to $\mathbf{1}^T\mathbf{v} = 1$ and $\mathbf{v} \geq \mathbf{0}$, see [37].

The following regularities concerning equivariances to geometrical transformations may be observed in the case of norm-generated distances.

Proposition 21. *Take any norm* $\|\cdot\|$ *and function* D *fulfilling assumptions listed in Proposition 17. Denote with P the corresponding distance-based penalty function and with* F *the P-based idempotent function. Then:*

(a) F *is translation equivariant;*

(b) if D *is homogeneous, then* F *is uniform-scale equivariant;*

(c) if $\|\cdot\|$ *is the Euclidean norm, then* F *is orthogonal equivariant;*

(d) if D *is strictly increasing in each variable as well as for any* $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ *such that if* $x_i' \geq x_i \geq 0$ *or* $x_i' \leq x_i \leq 0$ *for all* $i \in [d]$ *and additionally* $x_j' > x_j \geq 0$ *or* $x_j' < x_j \leq 0$ *for some* j *it holds* $\|\mathbf{x}\| < \|\mathbf{x}'\|$, *then* F *is BB-internal.*

Proof. Fix $\mathbf{X} \in (\mathbb{R}^d)^n$.

(a) For all $\mathbf{t}$ we have $P(\mathbf{y} + \mathbf{t}; \mathbf{X} + \mathbf{t}) = \mathsf{D}\left(\|\mathbf{y} - \mathbf{t} - \mathbf{x}^{(1)} + \mathbf{t}\|, \ldots, \|\mathbf{y} - \mathbf{t} - \mathbf{x}^{(n)} + \mathbf{t}\|\right) = P(\mathbf{y}; \mathbf{X})$. By Proposition 7 and Remark 8, this implies that F is translation equivariant.

(b) As each norm is absolutely homogeneous, for all $s > 0$ we have $P(s\mathbf{y}; s\mathbf{X}) = \mathsf{D}(s\|\mathbf{y} - \mathbf{x}^{(1)}\|, \ldots, s\|\mathbf{y} - \mathbf{x}^{(n)}\|) = s^\alpha P(\mathbf{y}; \mathbf{X})$ for some α. Applying Proposition 7 and Remark 8, we have that F is uniform-scale equivariant.

(c) In this case we have that $\|\mathbf{y} - \mathbf{x}\| = \sqrt{(\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x})}$. If $\mathbf{O} \in \mathbb{R}^{d \times d}$, then $\|\mathbf{Oy} - \mathbf{Ox}\| = \sqrt{(\mathbf{Oy} - \mathbf{Ox})^T(\mathbf{Oy} - \mathbf{Ox})} = \sqrt{(\mathbf{y} - \mathbf{x})^T\mathbf{O}^T\mathbf{O}(\mathbf{y} - \mathbf{x})}$. If $\mathbf{O}$ is orthogonal, then $\mathbf{O}^T = \mathbf{O}^{-1}$, and so $P(\mathbf{Oy}; \mathbf{OX}) = P(\mathbf{y}; \mathbf{X})$. Hence, F is orthogonal equivariant.

(d) Assume otherwise, let $\mathbf{y}$ be in the minimum set of $\mathsf{D}(\|\mathbf{y} - \mathbf{x}^{(1)}\|, \ldots, \|\mathbf{y} - \mathbf{x}^{(n)}\|)$, but $\mathbf{y} \notin \mathrm{BB}(\mathbf{X})$. Take $\mathbf{y}'$ with:

$$y_i' = \left(\bigwedge_{j=1}^n x_i^{(j)}\right) \vee y_i \wedge \left(\bigvee_{j=1}^n x_i^{(j)}\right)$$

for $i \in [d]$. Obviously, $\mathbf{y}' \in \mathrm{BB}(\mathbf{X})$. Moreover, for all $i \in [d]$ and $j \in [n]$ it holds that $x_i^{(j)} - y_i \geq x_i^{(j)} - y_i' \geq 0$ or $x_i^{(j)} - y_i \leq x_i^{(j)} - y_i' \leq 0$ and if $y_i' \neq y_i$, then the inequality is strict. This implies $\mathsf{D}(\|\mathbf{y} - \mathbf{x}^{(1)}\|, \ldots, \|\mathbf{y} - \mathbf{x}^{(n)}\|) > \mathsf{D}(\|\mathbf{y}' - \mathbf{x}^{(1)}\|, \ldots, \|\mathbf{y}' - \mathbf{x}^{(n)}\|))$, a contradiction. Thus, $\mathsf{F}(\mathbf{X}) \in \mathrm{BB}(\mathbf{X})$. $\qquad\square$

17

In particular, among scale equivariant (and hence homogeneous) D functions, we have power means (see above). Moreover, all L^p norms, $p \geq 1$, fulfill condition (d). From this and Lemma 11 we can imply that if D is strictly increasing, then for the Euclidean norm we surely obtain a CH-internal function.

Remark 22. Distance-based penalty minimizers can be applied in some clustering and prototype-based classification algorithms. For instance, a generalized k-means algorithm (see, e.g., [17, 45]) aims to find $k \geq 2$ cluster centers $\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(k)} \in \mathbb{R}^d$ that partition input data points into k disjoint groups. The i-th point's membership to one of the clusters, $c(i) \in [k]$, is expressed in terms of its proximity to a cluster center. This algorithm aims to minimize the total distance between all input points and their corresponding cluster centers:

$$\text{minimize} \sum_{i \in [n]} \eth(\mathbf{x}^{(i)}, \boldsymbol{\mu}^{(c(i))}) \quad \text{w.r.t. } c : [n] \to [k] \text{ (onto)},$$

where:

$$\boldsymbol{\mu}^{(i)} = \arg{}^* \min_{\mathbf{y} \in \mathbb{R}^d} \mathsf{D}\{\eth(\mathbf{y}, \mathbf{x}^{(j)}) : j \in [n], c(j) = i\} \tag{3}$$

for some extended (defined for any arity) aggregation function D and distance $\eth$. Moreover, we may similarly consider hierarchical clustering procedures based on a generalized centroid linkage criterion, see, e.g., [35].

Remark 23. Some authors (see, e.g., [44, 56, 57]) define internality as:

$$\mathsf{F}(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \in \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}. \tag{4}$$

If a function fulfilling this property is needed, we may restrict the search domain and seek a *set exemplar*, $\arg \min_{\mathbf{y} \in \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}} P(\mathbf{y}; \mathbf{X})$ (note that the solution might be ambiguous, but then we may return the input vector minimizing the penalty and, e.g., having the smallest index) instead of $\arg^* \min_{\mathbf{y} \in \mathbb{R}^d} P(\mathbf{y}; \mathbf{X})$. For instance, if we take P generated by D being the arithmetic mean and $\eth$ being the Euclidean metric, such a function is called the *medoid* – it can be used as a crude approximation of the 1-median.

6. Data depth-based penalties and corresponding medians

The purpose of the notion of data depth is to measure of how "deep" or "central" a point $\mathbf{y}$ is with respect to a point cloud $\mathbf{X}$. It may be used, e.g., to visualize (mostly bivariate) data sets [48], detect outliers, compute statistical hypotheses tests [20, 46], design control charts, and even support decision making [65].

It is assumed that the depth of a point $\mathbf{y} \in \mathbb{R}^d$ relative to $\mathbf{X} \in (\mathbb{R}^d)^n$ is quantified via a bounded from above function $\mathsf{depth} : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, b]$ for some $b > 0$. Zuo and Serfling in [77] list some desirable properties that this notion should fulfill, namely, for any $\mathbf{X}$ and $\mathbf{y}$ they require conditions like:

(zs1) affine invariance: for all $\mathbf{A} \in \mathbb{R}^{d \times d}$ of full rank and $\mathbf{t} \in \mathbb{R}^d$:

$$\mathsf{depth}(\mathbf{A}\mathbf{y} + \mathbf{t}; \mathbf{A}\mathbf{X} + \mathbf{t}) = \mathsf{depth}(\mathbf{y}; \mathbf{X});$$

(zs2) monotonicity relative to the deepest point: if $\mathsf{depth}(\mathbf{y}; \mathbf{X}) = \sup_{\mathbf{y}} \mathsf{depth}(\mathbf{y}; \mathbf{X})$, then for all $\mathbf{z}$ and $\alpha \in [0, 1]$ it holds:

$$\mathsf{depth}(\mathbf{z}; \mathbf{X}) \le \mathsf{depth}(\alpha\mathbf{y} + (1 - \alpha)\mathbf{z}; \mathbf{X});$$

(zs3) vanishing at infinity: $\lim_{\|\mathbf{y}\| \to \infty} \mathsf{depth}(\mathbf{y}; \mathbf{X}) = 0$.

For the purpose of our discussion, we propose the following set of *sine qua non* requirements. Then we shall relate the above ones to those discussed in Sect. 3.

Definition 24. A *depth measure* bounded from above by $b > 0$ is a function $\mathsf{depth} : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, b]$ such that:

(a) $\mathsf{depth}(\mathbf{y}; \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}) = b$ if and only if $\mathbf{y} = \mathbf{x}^{(i)}$ for all $i \in [n]$;

(b) for any $\mathbf{X} \in (\mathbb{R}^d)^n$, the set $\left\{ \mathbf{y} \in \mathbb{R}^d : \mathsf{depth}(\mathbf{y}; \mathbf{X}) = \sup_{\mathbf{y}' \in \mathbb{R}^d} \mathsf{depth}(\mathbf{y}; \mathbf{X}) \right\}$ is nonempty, bounded, and convex.

Based on this definition, we may introduce the notion of a depth-based penalty.

Proposition 25. *Let* $\mathsf{depth} : \mathbb{R}^d \times (\mathbb{R}^d)^n \to [0, b]$ *be a depth measure bounded from above by b. Then P given for any* $\mathbf{y}, \mathbf{X}$ *by* $P(\mathbf{y}; \mathbf{X}) = b/\mathsf{depth}(\mathbf{y}; \mathbf{X}) - 1$ *is a penalty function.*

Clearly, in the spirit of Lemma 3, we may transform both the depth function as well as P by some strictly increasing bijection and obtain another well-defined depths and penalty measures.

Of course, if depth is affine invariant (compare (zs1)), then the corresponding idempotent function (referred to as a depth-based median in the literature) is affine equivariant (by Proposition 7). Conditions (zs2) and (zs3), on the other hand, are clearly connected to, e.g., those in Lemma 4.

Let us indicate a few exemplary depths.

Example 26. The *halfplane location depth* [68] of **y** relative to **X** is defined as the smallest number of points in **X** contained in any closed halfhyperplane with boundary line through **y**. In other words, it is a depth measure bounded from above by *n* such that:

$$\mathsf{tdepth}(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \quad = \quad \min_{\mathbf{u} \in \mathbb{R}^d, \|\mathbf{u}\|=1} \left| \left\{ i : \mathbf{u}^T \mathbf{x}^{(i)} \geq \mathbf{u}^T \mathbf{y} \right\} \right|.$$

In the $d = 1$ case, the Tukey depth is related to the observations' ranking. The sample minimum and maximum are the points of depth 1, the median is of depth $n/2$ (the "deepest" value), and the first and the third quartiles are of depth $n/4$. As noted in [24], one can define trimmed means by, say, averaging points of depth $\geq n/10$.

It is easily seen that the set of all points of depth $\geq \delta$ (a δ-depth contour), for any given $\delta > 0$, is either empty or is a convex polytope. What is more, a point outside the convex hull of **X** is always of depth 0 [63]. The Tukey depth is affine equivariant, see [24, Lemma 2.1].

The *Tukey median*, TkMedian, is defined as the center of gravity of the deepest Tukey depth region. For $d = 1$, the Tukey median generalizes the concept of the median. Thus, in higher dimensions this function can be thought of as a multidimensional median. For algorithms to compute the Tukey depth and the corresponding median see [3, 19, 62–64].

Example 27. The Liu *simplical depth* [47] of $\mathbf{y} \in \mathbb{R}^d$ with respect to $\mathbf{X} \in (\mathbb{R}^d)^n$ is a depth measure bounded from above by $\binom{n}{d+1}$ defined as:

$$\mathsf{sdepth}(\mathbf{y}; \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}) \quad = \quad \left| \left\{ \{i_1, \ldots, i_{d+1}\} : 1 \leq i_1 < \cdots < i_{d+1} \leq n \right. \right.$$
$$\left. \left. \text{and } \mathbf{y} \in \mathrm{CH}\left(\mathbf{x}^{(i_1)}, \ldots, \mathbf{x}^{(i_{d+1})} \right) \right\} \right|.$$

In the bivariate case the Liu depth is given as the number of triangles formed by any three elements in **X** that contain **y**. It is easily seen that the Liu depth is affine invariant.

The *simplical median* is defined similarly to the Tukey median – as a point with the greatest simplical depth or the center of gravity of the deepest Liu depth region. For appropriate algorithms refer to [3, 4, 62].

Example 28. The convex hull peeling depth (see [28]; according to [42] the idea was proposed by Tukey) is determined by consecutively computing a convex hull

of a set of points and removing values lying on its boundary. The corresponding median may be constructed by computing the center of gravity of the "last" convex hull. According to [2], convex hull peeling may be done in $O(n \log^2 n)$ time for $d = 2$.

Numerous other depths may be found in the literature, e.g., the Oja depth [55], the L^1 depth [69], projection depth [76], perihedral depth [26], Delaunay depth [1, 39], zonoid data depth [27], and many others, see [48] for a review.

7. Final remarks

We presented a penalty-based model which is a common framework for all the idempotent multidimensional aggregation methods. What is more, some results can be easily generalized to the case of aggregation in vector spaces (e.g., in Sect. 5 – normed vector spaces). Please note that further functions may be obtained, e.g., by compositions of mappings. For instance, given $\mathsf{G} : (\mathbb{R}^d)^k \to \mathbb{R}^d$ and $\mathsf{F}_1, \dots, \mathsf{F}_k : (\mathbb{R}^d)^n \to \mathbb{R}^d$, which are idempotent, CH-, BB-internal, translation, uniform-scale, orthogonal, affine equivariant etc., then an idempotent function given by $\mathsf{F}(\mathbf{X}) = \mathsf{G}(\mathsf{F}_1(\mathbf{X}), \dots, \mathsf{F}_k(\mathbf{X}))$ is also, respectively, idempotent, CH-, BB-internal, and so forth.

Within the new framework, the properties of idempotent aggregation tools (means, averaging functions) may be studied at an appropriate level of generality. In this contribution we mostly put an emphasis on the properties that have a clear geometrical interpretation: those that deal with equivariances to particular transformations or different notions of internality. Yet, we should stress that some of the multidimensional functions – especially those listed in Sec. 6 – can be much more difficult to study formally, as they might not be defined with closed-form analytic formulas (compare, e.g., the notion of the convex-hull peeling-based median).

What is more, let us mention that idempotent aggregation methods (means) can be supplemented with an appropriate spread or dispersion measure, see [34, 43, 50] which can summarize a quite different aspect of a given data sample.

7.1. Notes on monotonicity

Aggregation theory very often assumes a kind of monotonicity with respect to the inputs. Such an assumption is not only quite natural in the $d = 1$ case, but also useful in deriving new facts. Nevertheless, in the multidimensional case, this notion is somehow problematic.

We have indicated that componentwise-nondecreasingness inevitably leads to componentwise functions (see Proposition 15), which are unable to take into account any possible dependencies between data dimensions.

Therefore, we may consider another natural generalization of the unidimensional monotonicity – a product lattice extension of the ordinary $\leq$ on $\mathbb{R}$. We shall say that F is $\leq_d$-*nondecreasing*, whenever for every $\mathbf{X}$ and $\mathbf{X}'$ if for all $i \in [n]$ it holds $\mathbf{x}^{(i)} \leq_d \mathbf{x}'^{(i)}$, then $\mathsf{F}(\mathbf{X}) \leq_d \mathsf{F}(\mathbf{X}')$, where, e.g., $\mathbf{x}^{(i)} \leq_d \mathbf{x}'^{(i)}$ if and only if for all $j \in [d]$ we have $x_j^{(i)} \leq x_j'^{(i)}$. Clearly, if F is componentwise-nondecreasing, then it is $\leq_d$-nondecreasing.

However, as far as some practical applications are concerned, this kind of monotonicity may cause a function to behave in an undesirable fashion: an increment on one of the variables can yield an increment on all the output dimensions. Moreover, if combined with, e.g., orthogonal equivariance, we get what follows.

Proposition 29. *Let* $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$ *be equivariant to reflections. Then* F *is* $\leq_d$-*nondecreasing if and only if* F *is componentwise-nondecreasing.*

Proof. ($\Longleftarrow$) Direct.

($\Longrightarrow$) For an arbitrary $j \in [d]$, let us take any $\mathbf{X}, \mathbf{X}' \in (\mathbb{R}^d)^n$ such that $\mathbf{x}_j \leq_n \mathbf{x}'_j$ and $\mathbf{x}_k = \mathbf{x}'_k$ for all $[d] \ni k \neq j$. Denote with $\mathbf{y} = \mathsf{F}(\mathbf{X})$ and $\mathbf{y}' = \mathsf{F}(\mathbf{X}')$. As for all $i \in [n]$ it holds $\mathbf{x}^{(i)} \leq_d \mathbf{x}'^{(i)}$, by $\leq_d$-nondecreasingness we surely have that $y_k \leq_d y'_k$, $k \in [d]$. However, by Proposition 15 we must show that $y_k = y'_k$ for all $k \neq j$. Consider a reflection matrix $\mathbf{R} \in \mathbb{R}^{d \times d}$ with $r_{u,v} = 0$ for $u \neq v$, $r_{k,k} = -1$ for $k \neq j$, and $r_{j,j} = 1$. As for all $i \in [n]$ it holds $\mathbf{R}\mathbf{x}^{(i)} \leq_d \mathbf{R}\mathbf{x}'^{(i)}$ and F is reflection equivariant, we have $\mathbf{R}\mathsf{F}(\mathbf{X}) \leq_d \mathbf{R}\mathsf{F}(\mathbf{X}')$. Therefore, for every $k \neq j$ we have $-y_k \leq_d -y'_k$ and $y_k \leq_d y'_k$ at the same time, QED. $\square$

Based on results given in Sect. 4, the only orthogonal equivariant, $\leq_d$-nondecreasing, and idempotent functions are componentwise extensions of some weighted arithmetic mean.

One could thus think of relaxing the above requirements and assume that if $\mathbf{x}^{(i)} \leq_d \mathbf{x}'^{(i)}$ for all $i \in [n]$, then we expect that $\mathsf{F}(\mathbf{X}) \not\geq_d \mathsf{F}(\mathbf{X}')$ (note that $\leq_d$ is merely a partial order). But this is not the case of, e.g., the Euclidean 1-median. For instance, take $d = 2$, $n = 3$, and $\mathbf{x}^{(1)} = [0, 0]^T$, $\mathbf{x}^{(2)} = [1, -5]^T$, $\mathbf{x}^{(3)} = [20, 1]^T$. We have $\mathsf{1median}_{\eth_2}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}) \simeq [1.961, -2.305]^T$. However, when we take $\mathbf{x}'^{(3)} = \mathbf{x}^{(3)} + [1980, 1]^T$, then we get $\mathsf{1median}_{\eth_2}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}'^{(3)}) \simeq [1.946, -3.351]^T <_2 [1.961, -2.305]^T$.

If monotonicity of some sort is necessary, we can, however, extend the concept of weak-nondecreasingness, see [71]. We say that a function $\mathsf{F} : (\mathbb{R}^d)^n \to \mathbb{R}^d$

is *weakly-monotone*, whenever $\mathsf{F}(\mathbf{X} + \mathbf{t}) \geq_d \mathsf{F}(\mathbf{X})$ for any $\mathbf{t} \geq_d \mathbf{0}$ and $\mathbf{X} \in (\mathbb{R}^d)^n$. Surely, every translation equivariant function is weakly-monotone, but the converse is not necessarily true.

7.2. Choosing a function

There are various ways that can aid in choosing a function for practical use. One of them may be based on the set of useful properties that an aggregation method should fulfill, such as a particular type of equivariance. Another idea is to rely on a function's numerical characteristics.

In the unidimensional case and under the assumption that $x_1^{(i)} \in \mathbb{I} = [0, 1]$, $i \in [n]$, notions like orness and andness are often considered, see [25, 31]. Unfortunately, they cannot be easily generalized to functions like $\mathsf{F} : (\mathbb{I}^d)^n \to \mathbb{I}^d$ when $d > 1$. A possible idea to overcome this limitation would be to consider, e.g., a quite different measure based on a properly normalized expected Euclidean distance between the outputs generated by F and the boundary of the $\mathbb{I}^d$ set under the assumption that input data are independent and uniformly distributed on $\mathbb{I}^d$. Alternatively, taking into account the fact that if $X \sim \mathrm{U}(\mathbb{I}^d)$, then $\mathbb{E} X = (d * 0.5)$, we may consider a mean squared error-like measure $\mathbb{E} \|\mathsf{F}(X_1, \dots, X_n) - (d * 0.5)\|_2^2$. Nevertheless, these characteristics will only work if $\mathbb{I}$ is bounded.

We can also consider the notion of a breakdown value [23] which is meant to serve as a measure of a function's robustness to the presence of potential outliers. Its aim is to express "the smallest amount of contamination which can cause the estimator to give an arbitrarily bad answer". Formally, $\mathsf{breakval}(\mathsf{F}) = \inf_{\mathbf{X} \in (\mathbb{R}^d)^n} \varepsilon_{\mathsf{F}}(\mathbf{X}) \in \left[\frac{1}{n}, 1\right]$, where:

$$\varepsilon_{\mathsf{F}}(\mathbf{X}) = \min_{m \in [n]} \left\{ \frac{m}{n} : \sup_{\mathbf{Y}_m \in (\bar{\mathbb{R}}^d)^n} \|\mathsf{F}(\mathbf{X}) - \mathsf{F}(\mathbf{Y}_m)\| = \infty \right\},$$

where the supremum is over all possible data sets $\mathbf{Y}_m$ obtained from $\mathbf{X}$ in such a way that exactly m points are replaced with arbitrary values. In particular, for translation equivariant functions we have that $\mathsf{breakval}(\mathsf{F}) \leq 0.5$, see [23].

Acknowledgments

References

[1] M. Abellanas, M. Claverol, F. Hurtado, Point set stratification and Delaunay depth, Computational Statistics & Data Analysis 51 (2007) 2513–2530.

[2] G. Aloupis, Geometric measures of data depth, in: DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 2006, pp. 147–158.

[3] G. Aloupis, C. Cortés, F. Gómez, M. Soss, G. Toussaint, Lower bounds for computing statistical depth, Computational Statistics & Data Analysis 40 (2002) 223–229.

[4] G. Aloupis, S. Langerman, M. Soss, G. Toussaint, Algorithms for bivariate medians and a Fermat-Torricelli problem for lines, Computational Geometry: Theory and Applications 26 (2003) 69–79.

[5] G. Beliakov, H. Bustince, T. Calvo, A Practical Guide to Averaging Functions, Springer, 2016.

[6] G. Beliakov, T. Calvo, S. James, On penalty-based aggregation functions and consensus, in: E. Herrera-Viedma, et al. (Eds.), Consensual Processes, STUDFUZZ 267, 2011, pp. 23–40.

[7] G. Beliakov, S. James, Stability of weighted penalty-based aggregation functions, Fuzzy Sets and Systems 226 (2013) 1–18.

[8] G. Beliakov, S. James, A penalty-based aggregation operator for non-convex intervals, Knowledge-Based Systems 70 (2014) 335–344.

[9] G. Beliakov, A. Pradera, T. Calvo, Aggregation functions: A guide for practitioners, Springer-Verlag, 2007.

[10] G. Beliakov, T. Wilkin, On some properties of weighted averaging with variable weights, Information Sciences 281 (2014) 1–7.

[11] P. Bullen, Handbook of means and their inequalities, Springer Science+Business Media, Dordrecht, 2003.

[12] H. Bustince, E. Barrenechea, M. Pagola, Relationship between restricted dissimilarity functions, restricted equivalence functions and normal e_N-functions: Image thresholding invariant, Pattern Recognition Letters 29 (2008) 525–536.

[13] H. Bustince, G. Beliakov, G.P. Dimuro, B. Bedregal, R. Mesiar, On the definition of penalty functions in data aggregation, Fuzzy Sets and Systems (2016). In press, doi:10.1016/j.fss.2016.09.011.

[14] H. Bustince, J. Fernandez, R. Mesiar, A. Pradera, G. Beliakov, Restricted dissimilarity functions and penalty functions, in: S. Galichet, et al. (Eds.), Proc. Eusflat/LFA'11, 2011, pp. 79–85.

[15] T. Calvo, G. Beliakov, Aggregation functions based on penalties, Fuzzy Sets and Systems 161 (2010) 1420–1436.

[16] T. Calvo, R. Mesiar, R.R. Yager, Quantitative weights and aggregation, IEEE Transactions on Fuzzy Systems 12 (2004) 62–69.

[17] A. Cena, M. Gagolewski, Fuzzy *k*-minpen clustering and *k*-nearest-minpen classification procedures incorporating generic distance-based penalty minimizers, in: J. Carvalho, et al. (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, Part II, volume 611 of *Communications in Computer and Information Science*, Springer, 2016, pp. 445–456.

[18] B. Chakraborty, P. Chaudhuri, On a transformation and re-transformation technique for constructing an affine equivariant multivariate median, Proceedings of the American Mathematical Society 124 (1996) 2539–2547.

[19] T.M. Chan, An optimal randomized algorithm for maximum Tukey depth, in: Proc. 15th ACM-SIAM Symp. Discrete Algorithms (SODA), 2004, pp. 430–436.

[20] S. Chenouri, C.G. Small, A nonparametric multivariate multisample test based on data depth, Electronic Journal of Statistics 6 (2012) 760–782.

[21] M. Demirci, Aggregation operators on partially ordered sets and their categorical foundations, Kybernetika 42 (2006) 261–277.

[22] M.M. Deza, E. Deza, Encyclopedia of Distances, Springer, 2014.

[23] D. Donoho, Breakdown properties of multivariate location estimates, Ph.D. thesis, Department of Statistics, Harvard University, 1982.

[24] D.L. Donoho, M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, The Annals of Statistics 20 (1992) 1803–1827.

[25] J.J. Dujmović, Weighted conjunctive and disjunctive means and their application in system evaluation, Publikacije Elektrotehničkog Fakulteta Univerziteta u Beogradu 461–497 (1974) 147–158.

[26] S. Durocher, R. Fraser, A. Leblanc, J. Morrison, M. Skala, On combinatorial depth measures, in: Proc. 26th Canadian Conf. Computational Geometry, 2014, pp. 206–211.

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

[28] W. Eddy, Convex hull peeling, in: Proc. COMPSTAT'82, Physica-Verlag, Vienna, 1982, pp. 42–47.

[29] H. Edelsbrunner, Algorithms in Combinatorial Geometry, Springer-Verlag, Heidelberg, 1987.

[30] K. Elbassioni, H.R. Tiwary, Complexity of approximating the vertex centroid of a polyhedron, Theoretical Computer Science 421 (2012) 56–61.

[31] J. Fernández Salido, S. Murakami, Extending Yager's orness concept for the OWA aggregators to other mean operators, Fuzzy Sets and Systems 139 (2003) 515–542.

[32] M. Gagolewski, Data Fusion: Theory, Methods, and Applications, Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, 2015.

[33] M. Gagolewski, Some issues in aggregation of multidimensional data, in: M. Baczynski, B. De Baets, R. Mesiar (Eds.), Proc. 8th International Summer School on Aggregation Operators (AGOP 2015), University of Silesia, Katowice, Poland, 2015, pp. 127–132.

[34] M. Gagolewski, Spread measures and their relation to aggregation functions, European Journal of Operational Research 241 (2015) 469–477.

[35] M. Gagolewski, A. Cena, M. Bartoszuk, Hierarchical clustering via penalty-based aggregation and the Genie approach, in: V. Torra, et al. (Eds.), Modeling Decisions for Artificial Intelligence, volume 9880 of *Lecture Notes in Artificial Intelligence*, Springer, 2016, pp. 191–202.

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

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

[38] M. Grabisch, J.L. Marichal, R. Mesiar, E. Pap, Aggregation Functions, Cambridge University Press, 2009.

[39] P. Green, Peeling bivariate data, in: V. Barnett (Ed.), Interpreting multivariate data, Wiley, New York, 1981.

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

[41] L. Guerrini, An extension of Witzgall's result on convex metrics, Divulgaciones Matematicás 13 (2005) 83–89.

[42] P.J. Huber, The 1972 Wald lecture robust statistics: A review, Annals of Mathematical Statistics 42 (1972) 1041–1067.

[43] A. Kołacz, P. Grzegorzewski, Measures of dispersion for multidimensional data, European Journal of Operational Research 251 (2016) 930–937.

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

[45] F. Leisch, A toolbox for K-centroids cluster analysis, Computational Statistics & Data Analysis 51 (2006) 526–544.

[46] J. Li, R.Y. Liu, New nonparametric tests of multivariate locations and scales using data depth, Statistical Science 19 (2004) 686–696.

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

[48] R.Y. Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: Descriptive statistics, graphics and inference, The Annals of Statistics 27 (1999) 783–858.

[49] R. Lucchetti, Convexity and Well-Posed Problems, CMS Books in Mathematics, 2006.

[50] J. Martín, G. Mayor, Dispersion measures and multidistances on $\mathbb{R}^k$, in: M.B. Ferraro, et al. (Eds.), Soft Methods for Data Science, Springer, 2017, pp. 347–354.

[51] R. Mesiar, Fuzzy set approach to the utility, preference relations, and aggregation operators, European Journal of Operational Research 176 (2007) 414–422.

[52] P. Milasevic, G. Ducharme, Uniqueness of the spatial median, The Annals of Statistics 15 (1987) 1332–1333.

[53] J. Möttönen, K. Nordhausen, H. Oja, Asymptotic theory of the spatial median, Nonparametrics and Robustness in Modern Statistical Inference and Time Series 7 (2010) 182–193.

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

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

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

[57] S. Ovchinnikov, Invariant functions on simple orders, Order 14 (1998) 365–371.

[58] R. Pérez-Fernández, M. Rademaker, B. De Baets, Monometrics and their role in the rationalisation of ranking rules, Information Fusion 34 (2017) 16–27.

[59] L. Rademacher, Approximating the centroid is hard, in: Symposium on Computational Geometry, 2007, pp. 302–305.

[60] M. Rademaker, B. De Baets, A ranking procedure based on a natural monotonicity constraint, Information Fusion 17 (2014) 74–82.

[61] R.T. Rockafellar, Convex Analysis, Princeton University Press, New Jersey, 1970.

[62] P.J. Rousseeuw, I. Ruts, Algorithm AS 307: Bivariate location depth, Applied Statistics 45 (1996) 516–526.

[63] P.J. Rousseeuw, I. Ruts, Constructing the bivariate Tukey median, Statistica Sinica 8 (1998) 827–839.

[64] P.J. Rousseeuw, A. Struyf, Computing location depth and regression depth in higher dimensions, Statistics and Computing 8 (1998) 193–203.

[65] P.J. Rousseuw, I. Ruts, The depth function of a population distribution, Metrika 49 (1999) 213–244.

[66] C.G. Small, A survey on multidimensional medians, International Statistical Review 58 (1990) 263–277.

[67] J.J. Sylvester, A question in the geometry of situation, Quarterly Journal of Pure and Applied Mathematics 1 (1857) 79.

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

[69] Y. Vardi, C.H. Zhang, The multivariate L_1-median and associated data depth, Proceedings of the National Academy of Sciences 97 (2000) 1423–1426.

[70] E. Weiszfeld, Sur le point par lequel la somme des distances de *n* points donnés est minimum, Tohoku Mathematics Journal 43 (1937) 355–386.

[71] T. Wilkin, G. Beliakov, Weakly monotonic averaging functions, International Journal of Intelligent Systems 30 (2015) 144–169.

[72] C. Witzgall, On convex metrics, Journal of Research of the National Bureau of Standards – B. Mathematics and Mathematical Physics 69B (1965) 175–177.

[73] R.R. Yager, Toward a general theory of information aggregation, Information Sciences 68 (1993) 191–206.

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

[75] R.R. Yager, A. Rybalov, Understanding the median as a fusion operator, International Journal of General Systems 26 (1997) 239–263.

[76] Y. Zuo, Projection-based depth functions and associated medians, The Annals of Statistics 31 (2003) 1460–1490.

[77] Y. Zuo, R. Serfling, General notions of statistical depth function, The Annals of Statistics 28 (2000) 461–482.