Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 323 (2017) 1–18
FUZZY sets and systems
www.elsevier.com/locate/fss

On the definition of penalty functions in data aggregation

Humberto Bustince [a,*], Gleb Beliakov [b], Graçaliz Pereira Dimuro [a,c], Benjamín Bedregal [d], Radko Mesiar [e,f]

[a] Depto of Automática y Computación and Institute of Smart Cities, Universidad Publica de Navarra, Campus Arrosadía, Navarra, 31006, Spain
[b] School of Engineering and Information Technology, Deakin University, 221 Burwood Hwy, Burwood, 3125, Australia
[c] Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Av. Itália km 08, Campus Carreiros, Rio Grande, 96201-900, Brazil
[d] Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal, 59072-970, Brazil
[e] Slovak University of Technology, Radlinskeho 11, 810 05 Bratislava, Slovakia
[f] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 18208 Prague, Czech Republic

Received 18 April 2016; received in revised form 13 September 2016; accepted 16 September 2016

Available online 21 September 2016

Abstract

In this paper, we point out several problems in the different definitions (and related results) of penalty functions found in the literature. Then, we propose a new standard definition of penalty functions that overcomes such problems. Some results related to averaging aggregation functions, in terms of penalty functions, are presented, as the characterization of averaging aggregation functions based on penalty functions. Some examples are shown, as the penalty functions based on spread measures, which happen to be continuous. We also discuss the definition of quasi-penalty functions, in order to deal with non-monotonic (or weakly/directionally monotonic) averaging functions.

Keywords: Penalty functions; Aggregation functions; Averaging aggregation function; Spread measures; Quasi-penalty functions

1. Introduction

Aggregation functions [1,2] are mainly used for obtaining a single output value from several input values. This procedure is indispensable in many applications, such as fuzzy rule based systems and classification systems [3–5], pattern recognition, image processing [6] or decision making [7,8]. Examples of aggregation functions are t-norms and t-conorms, uninorms, overlap and grouping functions, weighted quasi-arithmetic means, ordered weighted averaging (OWA) functions, Choquet and Sugeno integrals (see, e.g., [3,4,9–11]).

* Corresponding author.
E-mail addresses: bustince@unavarra.es (H. Bustince), gleb@deakin.edu.au (G. Beliakov), gracaliz@furg.br, gracaliz@gmail.com (G. Pereira Dimuro), bedregal@dimap.ufrn.br (B. Bedregal), radko.mesiar@stuba.sk (R. Mesiar).

http://dx.doi.org/10.1016/j.fss.2016.09.011
0165-0114/© 2016 Published by Elsevier B.V.

In particular, averaging aggregation functions [1,12] provide output values that are bounded by the minimum and maximum of the inputs, representing a consensus value of the inputs. They include a large class of functions (e.g., quasi-arithmetic means, medians, OWA functions and fuzzy integrals), which are often used in preference aggregation, aggregation of expert opinions, judgments in sports competitions. See [12] and the references therein.

Penalty functions have been defined as a measure of deviation from such a consensus value, or a penalty for not having a consensus, and have been studied by several authors, e.g., [8,11–20]. Examples of functions that minimize some penalties, called penalty-based functions, are the weighted arithmetic and geometric means, the median and the mode.

After the early works by Yager [13], Yager and Rybalov [14] and Calvo et al. [15], Calvo and Beliakov [11] provided a general definition of penalty function (and the corresponding penalty-based function) in 2010 and tried to show that every averaging function can be represented as a penalty-based function, studying a large class of such averaging aggregation functions. This definition did not encompass all the required features to agree with the intuition carried by the concept of penalty, and for this reason, other different definitions of penalty functions appeared in the literature (see, e.g., [8,16–18]).

However, we have found several problems in those different definitions, as well as in their related results, which prevent their use in real world applications and also in further theoretical research. Then, the objectives of this paper are:

- To shortly describe the evolution of the idea of using penalty functions in aggregation/fusion processes, by presenting the different definitions of penalty functions found in the literature, pointing out several problems we have encountered;
- To analyze and discuss such problems, proposing a new standard definition of penalty functions that overcomes all of them, that is, encompassing the most important compatible properties of the existing definitions, but suppressing the controversial ones;
- To present some important results related to averaging aggregation functions, as the characterization of penalty-based averaging aggregation functions;
- To present some examples, as the continuous penalty functions based on spread measures [21], which formalize the measures of absolute spread known from statistics, exploratory data analysis and data mining (e.g., the sample variance, the standard deviation and the range), and are able to measure the deviation that exists among the different inputs (and so, they are conceptually related to penalty functions);
- To sketch the definition of quasi-penalty functions, in order to deal with non-monotonic (or weakly/directionally monotonic) averaging functions, (e.g., the mode, which is weakly monotone, robust estimators of location and the least median of squares estimators) [12,18,22].

The paper is organized as follows. Section 2 presents some preliminary concepts, including new results that are necessary for the development of the paper. Section 3 recovers the evolution of the definition of penalty functions since it was initially proposed, analyzing and discussing the problems we have encountered. Our proposal for a new standard definition of penalty functions, and the related results, are introduced in Section 4, which also discusses the use of spread measures to obtain continuous penalty functions and mentions the role of quasi-penalty functions. Finally, Section 5 is the Conclusion.

2. Preliminaries

For the sake of completeness, we recall here some mathematical notions and results that will be useful for our subsequent developments. We also introduce some results that are necessary for the development of the paper, and fix some notations.

We denote by $\mathbb{I}$ a closed subinterval of the extended real line, i.e., $\mathbb{I} = [a, b] \subseteq \mathbb{R}$. We start by recalling the notion of convex and quasi-convex function.

Definition 2.1. A function $f : \mathbb{I} \to \mathbb{R}$ is convex if for every $x, y \in \mathbb{I}$ and for every $\lambda \in [0, 1]$ the inequality

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

holds.

Definition 2.2. A function $f : \mathbb{I} \to \mathbb{R}$ is quasi-convex if for every $x, y \in \mathbb{I}$ and for every $\lambda \in [0, 1]$ the inequality

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$$

holds.

Proposition 2.1. *Consider $\mathbb{I} = [a, b] \subseteq \mathbb{R}$ and a function $g : \mathbb{I} \to \mathbb{R}$. If there exists $c \in \mathbb{I}$ such that g is decreasing on $[a, c]$ and increasing on $[c, b]$, then g is quasi-convex.*

Proof. Consider $c \in \mathbb{I}$ such that g is decreasing on $[a, c]$ and increasing on $[c, b]$, $x, y \in \mathbb{I}$, $\lambda \in [0, 1]$ and define $z = \lambda x + (1-\lambda)y$. Then, one has the following cases:

1. If $x < c < y$ then we have that $x \leq z \leq y$ and:
   (a) If $z \leq c$ then, since g is decreasing on $[a, c]$, we have that $g(z) \leq g(x)$ and, then,
   $$g(\lambda x + (1-\lambda)y) = g(z) \leq \max\{g(x), g(y)\}.$$
   (b) If $c \leq z$ then, since g is increasing on $[c, b]$, we have that $g(z) \leq g(y)$ and, then,
   $$g(\lambda x + (1-\lambda)y) = g(z) \leq \max\{g(x), g(y)\}.$$
2. If $\max\{x, y\} \leq c$ then, since g is decreasing on $[a, c]$, we have that $g(z) \leq g(x)$ and $g(z) \leq g(y)$. It follows that
   $$g(\lambda x + (1-\lambda)y) = g(z) \leq \max\{g(x), g(y)\}.$$
3. If $c \leq \min\{x, y\}$ then, since g is increasing on $[c, b]$, we have that $g(z) \leq g(x)$ and $g(z) \leq g(y)$. It follows that
   $$g(\lambda x + (1-\lambda)y) = g(z) \leq \max\{g(x), g(y)\}.$$

□

Proposition 2.2. *Consider the function $g : \mathbb{I} \to \mathbb{R}$. Then it holds that:*

(i) *If g is monotonic then g is quasi-convex;*
(ii) *If g is convex then g is quasi-convex;*
(iii) *If g is convex then g has a minimizer in $\mathbb{I}$.*

Proof. To prove (i), consider Proposition 2.1 and observe that whenever g is increasing then $c = a$, and if g is decreasing, then $c = b$. The proofs of (ii) and (iii) are immediate [23,24]. □

Example 2.1. Consider $c \in \mathbb{R}$. The constant function $g_c : \mathbb{I} \to \mathbb{R}$, defined, for all $x \in \mathbb{I}$, by

$$g_c(x) = c, \tag{1}$$

is convex, and thus, quasi-convex. Additionally, whenever $h : \mathbb{I} \to \mathbb{R}$ is a quasi-convex function, then, for all $k \in \mathbb{R}$, the function $g_{h,k} : \mathbb{I} \to \mathbb{R}$, defined, for all $x \in \mathbb{I}$, by

$$g_{h,k}(x) = h(x) + k, \tag{2}$$

is also quasi-convex. In fact, for all $x_1, x_2 \in \mathbb{I}$ and for each $\lambda \in [0, 1]$, one has that:

$$\begin{aligned} g_{h,k}(\lambda x_1 + (1-\lambda)x_2) &= h(\lambda x_1 + (1-\lambda)x_2) + k \\ &\leq \max\{h(x_1), h(x_2)\} + k \\ &= \max\left\{g_{h,k}(x_1), g_{h,k}(x_2)\right\}, \end{aligned}$$

which shows that $g_{h,k}$ is quasi-convex.

Example 2.2. Consider $\theta \in \mathbb{R}$. The function $g_{\|\theta} : \mathbb{I} \to \mathbb{R}$, defined, for all $x \in \mathbb{I}$, by

$$g_{\|\theta}(y) = | \theta - y | \tag{3}$$

is convex, and, hence, by Proposition 2.2(ii), it is quasi-convex.

From the point of view of the present work, the most relevant fact about quasi-convex functions is the following property shown in [24]:

Proposition 2.3. *Let $f : \mathbb{I} \to \mathbb{R}$ be a quasi-convex function. Then, the set of minimizers of f is either empty or a connected set.*

Observe that on the real line, a connected set must be an interval which can be closed, open or semiopen.

Along with the definition of quasi-convexity, and since we are going to require that the set of minimizers of the considered quasi-convex functions is not empty, we also need the notion of lower semicontinuity [23].

Definition 2.3. A function $f : \mathbb{I} \to \mathbb{R}$ is lower semicontinuous at $x_0 \in \mathbb{I}$ if

$$\lim \inf_{x \to x_0} f(x) \geq f(x_0).$$

Analogously one defines upper semicontinuity. Observe that a function is continuous at $x_0 \in \mathbb{I}$ if and only if it is upper and lower semicontinuous there [23].

Lower semicontinuity on a compact domain ensures that the set of minimizers of a function is not empty [25]. It follows that:

Proposition 2.4. *Let $f : \mathbb{I} \to \mathbb{R}$ be a lower semicontinuous function, with $\mathbb{I}$ bounded. Then, the set of minimizers of f is not empty.*

Combining Propositions 2.3 and 2.4, we get the following result, which is key for our subsequent developments.

Corollary 2.1. *Let $f : \mathbb{I} \to \mathbb{R}$ be a quasi-convex and lower semicontinuous function, with $\mathbb{I}$ bounded. Then, the set of minimizers of f is a connected non-empty set.*

In the rest of this section, we consider n-ary functions, with $n \geq 2$.

Now, we recall here the definition of idempotent function.

Definition 2.4. A function $F : \mathbb{I}^n \to \mathbb{R}$ is idempotent if, for every $x \in \mathbb{I}$ it holds that

$$F(x, \ldots, x) = x.$$

We also review the notions of aggregation and averaging aggregation functions.

Definition 2.5. [1,2] A function $A : [0, 1]^n \to [0, 1]$ is said to be an n-ary aggregation function if:

(A1) A is increasing[1] in each argument: for each $i \in \{1, \ldots, n\}$, if $x_i \leq y$, then $A(x_1, \ldots, x_n) \leq A(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)$;

(A2) A satisfies the boundary conditions: $A(0, \ldots, 0) = 0$ and $A(1, \ldots, 1) = 1$.

Definition 2.6. An aggregation function $f : [0, 1]^n \to [0, 1]$ is called averaging if it is bounded by the minimum and maximum of its arguments, that is, for all $(x_1, \ldots, x_n) \in [0, 1]^n$, it holds that:

$$\min\{x_1, \ldots, x_n\} \leq f(x_1, \ldots, x_n) \leq \max\{x_1, \ldots, x_n\}.$$

[1] In this paper, an increasing (decreasing) function does not need to be strictly increasing (decreasing).

Due to the monotonicity of aggregation functions f, the averaging behavior is equivalent to the idempotency property.

Finally, we present the notion of spread measures, introduced by Gagolewski in [21], which are not aggregation functions in general. Spread measures formalize the measures of absolute spread.

Definition 2.7. The mapping $\mathrm{V} : \mathbb{I}^n \to \mathbb{R}^+$ is a spread measure if:

(V1) For each $\vec{x} \preccurlyeq_n \vec{x}'$ it holds that $\mathrm{V}(\vec{x}) \leq V(\vec{x}')$;
(V2) For any $x \in \mathbb{I}$ it holds that $\mathrm{V}(x, \ldots, x) = 0$;

where $\preccurlyeq_n$ is a preorder on $\mathbb{I}^n$ defined by

$$\vec{x} \preccurlyeq_n \vec{x}' \Leftrightarrow \forall i, j = 1, \ldots, n : (x_i - x_j)(x'_i - x'_j) \geq 0 \ \wedge \ | x_i - x_j | \leq | x'_i - x'_j |, \tag{4}$$

and meaning that $\vec{x}$ does not have greater absolute spread than $\vec{x}'$.[2]

When the property **(V2)** is replaced with a stronger property **(V2')** $\mathrm{V}(\vec{x}) = 0$ if and only if $\vec{x} = (x, \ldots, x)$, the spread measure is said to be strong.

Define

$$\vec{x} \prec_n \vec{x}' \Leftrightarrow (\forall i, j = 1, \ldots, n : (x_i - x_j)(x'_i - x'_j) \geq 0 \wedge | x_i - x_j | \leq | x'_i - x'_j |) \\ \wedge (\exists r, s \in \{1, \ldots, n\} : (x_r - x_s)(x'_r - x'_s) \geq 0 \wedge | x_r - x_s | < | x'_r - x'_s |),$$

meaning that $\vec{x}$ has strictly smaller absolute spread than $\vec{x}'$.

When $\vec{x} \prec_n \vec{x}'$ and this implies the strict inequality $\mathrm{V}(\vec{x}) < V(\vec{x}')$, we say that this spread measure is strict.

Remark 2.1. Consider $c \in \mathbb{I}$ and $\vec{x} = (x_1, \ldots, x_n), \vec{c} = (c, \ldots, c), \vec{x_c} = (x_1 + c, \ldots, x_n + c) \in \mathbb{I}^n$. Then, it is easy to check that $\vec{c} \preccurlyeq_n \vec{x}$. However, whenever there exist $r, s \in \{1, \ldots, n\}$, such that $x_r \neq x_s$, then one has that $\vec{c} \prec_n \vec{x}$. On the other hand, since $\vec{x} \preccurlyeq_n \vec{x_c}$ and $\vec{x_c} \preccurlyeq_n \vec{x}$, then, for any spread measure $\mathrm{V} : \mathbb{I}^n \to \mathbb{R}^+$, it holds that $V(\vec{x_c}) = V(\vec{x})$, that is, V is shift invariant in the sense of Gagolewski [21] (a constant change in every input does not result in a change of the output).[3]

Proposition 2.5. *Any strict spread measure* $\mathrm{V} : \mathbb{I}^n \to \mathbb{R}^+$ *is strong.*

Proof. Let V be a strict spread measure. From **(V2)**, it is immediate that if $\vec{x} = (x, \ldots, x) \in \mathbb{I}^n$, then $\mathrm{V}(\vec{x}) = 0$. Now, consider $\vec{x}' = (c, \ldots, c) \in \mathbb{I}^n$ and $\vec{x} \in \mathbb{I}^n$, such that $\vec{x} \neq \vec{x}'$. Suppose that $\mathrm{V}(\vec{x}) = 0$. By Remark 2.1, one has that $\vec{x}' \prec \vec{x}$. However, by **(V2)**, one has that

$$\mathrm{V}(\vec{x}') = \mathrm{V}(c, \ldots, c) = 0 \not< 0 = \mathrm{V}(\vec{x}),$$

which is a contradiction, since V is strict. Then, one concludes that $\vec{x} = (c, \ldots, c)$. Therefore, V is strong. □

Example 2.3. Examples of spread measures include the variance and the standard deviation. The variance can be computed as:

$$V(\vec{x}) = \frac{1}{n^2} \sum_{i=1}^{n-1} \sum_{j>i}^{n} (x_i - x_j)^2.$$

Another useful spread measure is the mean absolute deviation MD, given by:

[2] In the definition of the preorder $\preccurlyeq_n$ (Eq. (4)), the condition $\forall i, j = 1, \ldots, n : (x_i - x_j)(x'_i - x'_j) \geq 0$ means that $\vec{x}$ and $\vec{x}'$ should be comonotonic [2].

[3] Observe that there is another notion of shift invariance related to aggregation functions given by Lázaro et al. [26] (a constant change in every input should result in a corresponding change of the output), which also appears in [2], but, in this case, not every spread measure is shift invariant.

$$MD(\vec{x}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|.$$

Both spread measures are strict.

3. Different definitions of penalty functions

Yager [13] in 1993 has attempted to present the initial ideas related to the use of a penalty function to help aggregation processes. In that paper, Yager introduced the notion of a penalty cost that may be attributed to some input datum x_i, whenever one disregards it by concluding as a result of the aggregation process some value y that conflicts with x_i.

Later, in 1997, Yager and Rybalov [14] have considered the idea of using the minimization of a penalty function as a method for obtaining a fused value of a collection of n observations,[4] organized in an input vector $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. We formalized their approach here, as follows:

Definition 3.1. The function $LP : \mathbb{R}^2 \to \mathbb{R}^+$ is said to be a local penalty function if, for any $x_i, x_j, y \in \mathbb{R}$ and $i, j = 1, \dots, n$, it satisfies:

(LP3.1-1) $LP(x_i, y) = 0$, if $x_i = y$;
(LP3.1-2) $LP(x_i, y) > 0$, if $x_i \neq y$;
(LP3.1-3) $LP(x_i, y) \geq LP(x_j, y)$, if $| x_i - y |>| x_j - y |$,

where y is called the fused value related to each observation in $\vec{x}$.

Remark 3.1. Note that the conditions **(LP3.1-1)** and **(LP3.1-2)** are equivalent to:

$$LP(x_i, y) = 0 \text{ if and only if } x_i = y.$$

Definition 3.2. Let $LP : \mathbb{R}^2 \to \mathbb{R}^+$ be a local penalty function. A penalty function $P : \mathbb{R}^{n+1} \to \mathbb{R}^+$ is defined, for any $\vec{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ as:

$$P(\vec{x}, y) = \sum_{i=1}^{n} LP(x_i, y), \tag{5}$$

where $y \in \mathbb{R}$ is called the fused value of $\vec{x} \in \mathbb{R}^n$.

The function P measures the total penalty incurred, whenever y is the fused value of the vector $\vec{x}$. The best (optimum) fused value of the elements in $\vec{x}$ is the value y^* that minimizes the penalty function P. Thus, the fused value of $\vec{x}$, namely, $f(x_1, \dots, x_n)$, is obtained as the value y^* such that

$$P(\vec{x}, y^*) = \min_y P(\vec{x}, y).$$

The function f, in more recent works, is called a penalty-based function, or P-function, for short.

In the Problem 1 bellow, we point out an important issue related to Yager and Rybalov's approach:

Problem 1. It is clear that the output y^* minimizing the total penalization $P(\vec{x}, \cdot)$ in general need not exist, or there can exist more than one solution [14].

By providing different forms for the penalty function, Yager and Rybalov [14] presented different forms of fusion methods. For example, in one of these methods the penalty function is described by the absolute difference between an input x_i and the fused value y, which resulted in the median type aggregation.

[4] In [14], the observations may be of any type, even of non numeric type. For the sake of simplicity, in this paper we consider just numerical values.

When assuming that each of the observations has an associated weight function (e.g., the weight can be a measure of the importance or reliability of a certain sensor), the weighted penalty function is obtained as follows:

Definition 3.3. Let $LP : \mathbb{R}^2 \to \mathbb{R}^+$ be a local penalty function. The weighted penalty function $wP : \mathbb{R}^{n+1} \to \mathbb{R}^+$ is defined, for any $\vec{x} \in \mathbb{R}^n$ and $y \in \mathbb{R}$ as:

$$wP(\vec{x}, y) = \sum_{i=1}^{n} w_i LP(x_i, y), \tag{6}$$

where $w_i \geq 0$ is the weight associated to the observation x_i, with $i = 1, \dots, n$.

Remark 3.2. Observe that there are no conditions imposed on the weights w_i associated to the observations other than $w_i \geq 0$ and $\sum_{i=1}^{n} w_i > 0$.

Let $\mathbb{I} = [a, b] \subseteq [-\infty, +\infty]$. In 2004, Calvo et al. [15] have slightly modified the approach proposed in [14], introducing the following definition[5]:

Definition 3.4. The function $LP : \mathbb{I}^2 \to [0, \infty]$ is a local penalty function on $\mathbb{I}$ if and only if, for any $x, y \in \mathbb{I}$, it satisfies:

(LP3.4-1) $LP(x, y) = 0$ if $x = y$, and
(LP3.4-2) $LP(x, y) \geq LP(z, y)$, whenever $x \geq z \geq y$ or $x \leq z \leq y$.

Considering the local penalty $LP : \mathbb{I}^2 \to [0, \infty]$, a penalty function $P : \mathbb{I}^{n+1} \to [0, \infty]$ is defined similarly to Equation (5).

To avoid Problem 1, Calvo et al. [15] restricted their considerations to local penalty functions satisfying

$$LP = K \circ f, \quad \text{that is,} \quad LP(x, y) = K(f(x), f(y)), \tag{7}$$

where $f : \mathbb{I} \to [-\infty, +\infty]$ is a continuous strictly monotonic function and $K : [-\infty, +\infty]^2 \to [0, \infty]$ is a local penalty function, which is convex in each component. Such local penalty function is called a faithful local penalty function.

Adopting faithful local penalty functions $LP : \mathbb{I}^2 \to [0, \infty]$, a penalty function $P : \mathbb{I}^{n+1} \to [0, \infty]$ is defined similarly to Equation (5). Then, Calvo et al. [15] introduced the following definition:

Definition 3.5. Let $LP : \mathbb{I}^2 \to [0, \infty]$ be a faithful local penalty function. A function $f_P : \bigcup_{n \in \mathbb{N}} \mathbb{I}^n \to \mathbb{I}$, defined for all $\vec{x} \in \mathbb{I}^n$ and $n \in \mathbb{N}$, by

$$f_P(\vec{x}) = \frac{l_{\vec{x}} + r_{\vec{x}}}{2},$$

where

$$l_{\vec{x}} = \inf\{u \in \mathbb{I} \mid \forall v \in \mathbb{I} : P(\vec{x}, u) \leq P(\vec{x}, v)\}$$
$$r_{\vec{x}} = \sup\{u \in \mathbb{I} \mid \forall v \in \mathbb{I} : P(\vec{x}, u) \leq P(\vec{x}, v)\}$$

is called a penalty-based function, or P-function, for short.

Analogously to Definition 3.3 and considering faithful local penalty functions, Calvo et al. [15] have incorporated quantitative weights into the definition of penalty functions.

[5] In [15], the local penalty function of Definition 3.4 was called simply penalty function. We decided to maintain the nomenclature introduced in [14] in order to facilitate the comparison between the approaches.

Remark 3.3. Equation (5) may be rewritten as:

$$P(\vec{x}, y) = \sum_{i=1}^{n} D(x_i, y),$$

where the dissimilarity D is given in the form

$$D(x, y) = K(f(x) - f(y)),$$

for a convex function $K : [-\infty, +\infty] \to [0, \infty]$ and a continuous strictly monotonic function $f : \mathbb{I} \to [-\infty, +\infty]$. By Proposition 2.2(iii), K has a minimum. Observe that convex real functions having a unique minimum $K(0) = 0$ can be constructed easily. Moreover, the penalty function P is, in fact, related to a one-variable functions, which can be seen as a distance function. In this case, one always obtains an idempotent aggregation function, which in case that f is linear then is also shift invariant in the sense of Gagolewski [21] (see Remark 2.1). This approach is related to the so-called Daróczy means, where one does not look for minimizers, but for solutions of $\sum_{i=1}^{n} \phi(x_i, y) = 0$, where ϕ, for each fixed y, is continuous and strictly decreasing in x and $\phi(y, y) = 0$ (e.g., in case of smooth κ-means, $\phi(x, y) = \kappa(f(x) - f(y))$).

In fact, although the approach by Calvo et al. [15] is very simple and sound, one can point out an important problem, which has motivated the researchers to the search for other complementary approaches:

Problem 2. The weakness of the approach of Calvo et al. [15] is that not all idempotent aggregation functions can be obtained by this method, as idempotent uninorms[6] and nullnorms.[7] This is the case of the idempotent uninorm $U : [0, 1]^2 \to [0, 1]$, defined by

$$U(x, y) = \begin{cases} \max\{x, y\} & \text{if } x, y \in [0, e] \\ \min\{x, y\} & \text{otherwise,} \end{cases}$$

where $e \in]0, 1[$ is the neutral element of U [27].

In order to have a more general definition, and towards to have solution for Problem 2, in [11], Calvo and Beliakov introduced the following definition of a penalty function:

Definition 3.6. The function $P : \mathbb{I}^{n+1} \to [0, \infty]$ is a penalty function if and only if it satisfies:

(P3.6-1) $P(\vec{x}, y) \geq 0$, for all $\vec{x}$,y;
(P3.6-2) $P(\vec{x}, y) = 0$ if $\vec{x} = \vec{y}$, and[8]
(P3.6-3) For every fixed $\vec{x}$, the set of minimizers of $P(\vec{x}, y)$ is either a singleton or an interval.

Definition 3.7. A penalty-based function $f : \mathbb{I}^n \to \mathbb{I}$ is defined, for all $\vec{x} \in X^n$, by

$$f(\vec{x}) = \arg\min_{y} P(\vec{x}, y), \tag{8}$$

if y is the unique minimizer, and $y = \frac{a+b}{2}$ if the set of minimizers is the interval $]a, b[$ (or $[a, b]$).

We abbreviate penalty-based function as P-function.

[6] A uninorm is a bivariate aggregation function $U : [0, 1]^2 \to [0, 1]$ that is associative, commutative, and with neutral element $e \in]0, 1[$. [1]

[7] A nullnorm is a bivariate aggregation function $V : [0, 1]^2 \to [0, 1]$ that is associative, commutative, and such that there exists an element $a \in]0, 1[$ verifying: (i) $\forall t \in [0, a] : V(t, 0) = t$, (ii) $\forall t \in [a, 1] : V(t, 1) = t$ [1]

[8] Observe that the vector $\vec{y}$ should be defined as $\vec{y} = (\underbrace{y, \ldots, y}_{n \text{ times}})$, although the authors did not mentioned that in [11].

Remark 3.4. In Definition 3.6, observe that:

1. The condition **(P3.6-1)** is redundant, since this is a consequence of the fact that the range of the function P is $[0, \infty]$.
2. A singleton can be seen as a degenerated interval $\{k\} = [k, k]$.

Definition 3.6 presents some problems identified below.

Problem 3. Consider $\mathbb{I} = [-\infty, +\infty]$ and the penalty function $P(\vec{x}, y) = 0$. Then, the P-function is $f(\vec{x}) = \frac{-\infty+\infty}{2}$. However, the paper does not provide an algebra for $[-\infty, +\infty]$, so the value is not properly defined.

Problem 4. Calvo and Beliakov claim that "any P-function is idempotent" [11, page 1424, paragraph before Note 3], which is a reasonable property for developing some applications. However, we found that their definition of penalty function may give rise to a P-function that is not idempotent. For example, consider $P : [0, 1]^{n+1} \to [0, \infty]$ defined by

$$P(x_1, \ldots, x_n, y) = \begin{cases} 0 & \text{if } y \le x_1; \\ 1 & \text{otherwise.} \end{cases}$$

Clearly, P is well defined and satisfies **(P3.6-1)** and **(P3.6-2)**. P also satisfies **(P3.6-3)**, since the set of minimizers of $P(x_1, \ldots, x_n, y)$ is $[0, x_1]$. Therefore, according to Definition 3.6, P is a penalty function. However, the P-function $f : [0, 1]^n \to [0, 1]$ is defined by $f(\vec{x}) = \frac{x_1}{2}$, for each $x_1, \ldots, x_n \in [0, 1]$, and, thus, f is not idempotent.

In [17], Beliakov and James discussed the problem of aggregating some special kind of discontinuous intervals, which they called non-convex intervals, providing a different definition of penalty function, as follows:

Definition 3.8. The function $P : [0, 1]^{n+1} \to [0, \infty]$ is a penalty function if and only if it satisfies:

(P3.8-1) $P(\vec{x}, y) = 0$ if $x_i = y$, for all i;
(P3.8-2) $P(\vec{x}, y) > 0$ if $x_i \neq y$ for some i, and
(P3.8-3) For every fixed $\vec{x}$, the set of minimizers of $P(\vec{x}, y)$ is either a singleton or an interval.

The concept of penalty-based function, for a penalty function in the sense of Definition 3.8, is defined analogously to Definition 3.7.

Definition 3.8 is not equivalent to Definition 3.6. For instance, the function $P \equiv 0$ considered in the discussion of Problem 3 above is not a penalty function in the sense of Definition 3.8.

Remark 3.5. In Definition 3.8, the Problems 3 and 4 are solved. In particular, conditions **(P3.8-1)** and **(P3.8-2)** guarantee that any P-function is idempotent.

Observe that the restriction of the domain to $[0, 1]^{n+1}$ is not very significant, since the results can be extended to any other bounded interval in a straightforward way.

As pointed in [17, Remark 1], the conditions (P3.8-1) and (P3.8-2) of Definition 3.8 are equivalent to the condition (P3.8-1) substituting the "if" by "if and only if" (as made, more recently, in [18]). With this in mind, we can infer that the P-functions based on the notion of penalty function provided in Definition 3.8 are necessarily idempotent.

Now observe that there are many averaging functions that are P-functions. For example, the arithmetic mean AM can be generated from

$$P(\vec{x}, y) = \frac{\sum_{i=1}^{n} (x_i - y)^2}{n},$$

which is a penalty function in the sense of Definition 3.8.

A question arises: "Is each averaging aggregation function a P-function?" To answer this question, Calvo and Beliakov in [11, Theorem 1] provided the following theorem:

Theorem 3.1. *[11, Theorem 1] Let $f : \mathbb{I}^n \to \mathbb{I}$ be an idempotent function. Then there exists a penalty function $P : \mathbb{I}^{n+1} \to [0, \infty]$, such that*

$$f(\vec{x}) = \arg\min_y P(\vec{x}, y).$$

Remark 3.6. The proof of [11, Theorem 1] is not valid when one considers the Definition 3.8. In fact, if $A : [0, 1]^n \to [0, 1]$ is an averaging aggregation function, then the function $P_A : [0, 1]^{n+1} \to [0, 1]$, defined by $P_A(\vec{x}, y) = | A(\vec{x}) - y |$, is not a penalty function in the sense of Definition 3.8, since, whenever one takes $y = A(\vec{x})$, it holds that $P_A(\vec{x}, A(\vec{x})) = 0$, for each $\vec{x} \in [0, 1]^n$, and, thus, the property **(P3.8-2)** fails.

In this paper, we answer the question discussed in Remark 3.6 providing a stronger theorem than the (non-proved) Theorem 1 in [11] in the context of Definition 3.8:

Theorem 3.2. *A function $f : [0, 1]^n \to [0, 1]$ is a P-function, for a penalty function in the sense of Definition 3.8, if and only if f is idempotent.*

Proof. ($\Rightarrow$) If $P : [0, 1]^{n+1} \to [0, \infty]$ is a penalty function in the sense of Definition 3.8, then, by the conditions **(P3.8-1)** and **(P3.8-2)**, the set of minimizers of $P(\vec{x}, \cdot)$ when $\vec{x} = (x, \ldots, x)$ is $\{x\}$. Therefore, one has that

$$f(x, \ldots, x) = \arg\min_y P(x, \ldots, x, y) = x,$$

and, thus, f is idempotent.

($\Leftarrow$) Consider an idempotent function $f : [0, 1]^n \to [0, 1]$, $\epsilon > 0$, and the function $P_f : [0, 1]^{n+1} \to [0, +\infty]$, defined by

$$P_f(\vec{x}, y) = \begin{cases} 0 & \text{if } x_i = y \text{ for each } i \\ | f(\vec{x}) - y | + \epsilon & \text{otherwise.} \end{cases}$$

Clearly, $P(\vec{x}, y) > 0$ if $x_i \neq y$ for some i **(P3.8-2)**, and, additionally, $P(\vec{x}, y) > 0$ if $x_i \neq y$ for some i **(P3.8-1)**. Now observe that the set of minimizers of $P_f(\vec{x}, \cdot)$ is the singleton $\{f(\vec{x})\}$ **(P3.8-3)**, and, thus, we have that P_f is a penalty function, in the sense of Definition 3.8, and $f(\vec{x}) = \arg\min_y P_f(\vec{x}, y)$. □

Note that P_f can be defined for any function f, and the corresponding P_f-function $g : [0, 1]^n \to [0, 1]$, for $c \in [0, 1]$, is given by

$$g(x) = \begin{cases} c & \text{if } x = (c, \ldots, c); \\ f(x) & \text{otherwise,} \end{cases}$$

and hence $g = f$ if and only if f is idempotent.

It follows that each averaging aggregation function is a P-function, for a penalty function in the sense of Definition 3.8.

Remark 3.7. The proof of Theorem 3.2 uses the same argument as that in [11], but with an extra term, which makes the penalty function discontinuous. Later we show that one can define a continuous penalty function, e.g., by using a suitable spread measure [21].

More recently, Wilkin and Beliakov [18] discussed the non-monotonic averaging problem by using weakly monotonic aggregation functions (see also [3,22]). For example, the mode is non-monotonic but weakly monotonic idempotent function. They introduced the following definition of a penalty function [18, Definition 10], which overcomes some restrictions:

Definition 3.9. For any closed, nonempty interval $\mathbb{I} \subseteq [-\infty, +\infty]$, the function $P : \mathbb{I}^{n+1} \to \mathbb{R}$ is a penalty function if and only if it satisfies:

(P3.9-1) $P(\vec{x}, y) \geq c$, for all $\vec{x} \in \mathbb{I}^n, y \in \mathbb{I}$, for some constant $c \in \mathbb{R}$;
(P3.9-2) $P(\vec{x}, y) = c$ if and only if all $x_i = y$, for all $i = 1 \ldots n$, and
(P3.9-3) P is quasi-convex in y for any $\vec{x}$.

Remark 3.8. In this definition, negative penalties are allowed, although one may consider $c = 0$. Observe that there exists in the literature the use of some negative "penalty" parameters, but in other different contexts, such as (i) partial differential equations, in order to keep the error due to the violation of the constraints within a desired tolerance [19], and (ii) linear systems of equations, for the modeling of tyings (constraints that relate multiple degrees of freedom) [20]. In the context considered in this paper, as discussed in Section 2, negative penalties have no meaning. We also observe that, in [8], Bustince et al. used a definition of penalty function that is a particular case of Definition 3.9 by considering only non-negative penalties.

Remark 3.9. Observe that, whenever one considers $P(\vec{x}, y) \geq c$ in **(P3.9-1)**, then one has that $P - c \geq 0$ and $P - c$ is also a penalty function giving the same output. Although one can point out that there is no need to consider such general lower bound c, we remark that this is not a problem of Definition 3.9.

Remark 3.10. In this definition, Problem 4 is solved. In particular, condition **(P3.9-2)** guarantees that any P-function is idempotent. However, Problem 3 remains. In fact, consider $c \in \mathbb{R}$ and the function $P : [-\infty, +\infty]^{n+1} \to \mathbb{R}$, defined by:

$$P(\vec{x}, y) = \begin{cases} c & \text{if } x_i = y, \text{ for all } i = 1 \ldots n \\ 2c & \text{otherwise.} \end{cases}$$

It is immediate that P satisfies **(P3.9-1)** and **(P3.9-2)**. Moreover, for a fixed $\vec{x}$, one has the following cases:

Case 1: There does not exist $z \in [-\infty, +\infty]$ such that $\vec{x} = (z, \ldots, z)$. In this case, it holds that:

$$P(\vec{x}, \lambda y_1 + (1 - \lambda)y_2) = 2c = \max\{P(\vec{x}, y_1), P(\vec{x}, y_2)\}.$$

Case 2: There exists $z \in [-\infty, +\infty]$ such that $\vec{x} = (z, \ldots, z)$. Thus, whenever $\lambda y_1 + (1 - \lambda)y_2 = z$, then

$$P(\vec{x}, \lambda y_1 + (1 - \lambda)y_2) = c \leq \max\{P(\vec{x}, y_1), P(\vec{x}, y_2)\}.$$

Otherwise, if $\lambda y_1 + (1 - \lambda)y_2 \neq z$, then one has that either $y_1 \neq z$ or $y_2 \neq z$. Suppose $y_1 \neq z$. Then it holds that

$$P(\vec{x}, \lambda y_1 + (1 - \lambda)y_2) = 2c = \max\{2c, P(\vec{x}, y_2)\}.$$

Analogously, one shows the same result for $y_2 \neq z$.

This proves that P is quasi-convex in y, for any $\vec{x}$. Thus, P is a penalty function in the sense of Definition 3.9. Then, the P-based function $f : [-\infty, +\infty]^n \to [-\infty, +\infty]$ is such that $f(\vec{x}) = \frac{-\infty+\infty}{2}$, when $\vec{x}$ satisfies the **Case 1** above.

Remark 3.11. In [18, Theorem 1], the authors state that any idempotent function can be represented as a P-function, and, in [18, Corollary 1], they affirm that any averaging aggregation function can be expressed as a P-function. However, the proofs are not presented. In particular, the problem discussed in Remark 3.6 in the context of Definition 3.9 remains.

Problem 5. Contrary to what is claimed in [18], quasi-convexity does not guarantee non-empty sets of minimizers (see Corollary 2.1). Since an empty set is trivially convex, quasi-convex functions may have an empty set of minimizers.

Example 3.1. Consider the function $P : [0,1]^3 \to \mathbb{R}$, defined by

$$P(x_1, x_2, y) = \begin{cases} 0 & \text{if } x_1 = x_2 = y; \\ y & \text{if } y > \frac{x_1+x_2}{2}; \\ 1 & \text{if } y \leq \frac{x_1+x_2}{2} \wedge \{x_1, x_2\} \neq \{y\}. \end{cases}$$

Trivially, P satisfies **(P3.9-1)** and **(P3.9-2)** for $c = 0$. Consider $y_1, y_2 \in [0,1]$ and $\lambda \in [0,1]$. Without loss of generality, we assume that $y_1 \leq y_2$. Consider $y = \lambda y_1 + (1-\lambda) y_2$. Then, we have that $y_1 \leq y \leq y_2$. It follows that:

1. If $x_1 = x_2 = y$ then $P(x_1, x_2, y) = 0 \leq \max\{P(x_1, x_2, y_1), P(x_1, x_2, y_2)\}$.
2. If $y > \frac{x_1+x_2}{2}$ then

$$\begin{aligned} P(x_1, x_2, y) &= y \\ &\leq y_2 \\ &\leq \max\{P(x_1, x_2, y_1), y_2\} \\ &= \max\{P(x_1, x_2, y_1), P(x_1, x_2, y_2)\}. \end{aligned}$$

3. If $y \leq \frac{x_1+x_2}{2}$ and $\{x_1, x_2\} \neq y$ then

$$\begin{aligned} P(x_1, x_2, y) &= 1 \\ &= \max\{1, P(x_1, x_2, y_2)\} \\ &= \max\{P(x_1, x_2, y_1), P(x_1, x_2, y_2)\}. \end{aligned}$$

Therefore, one concludes that P satisfies **(P3.9-3)**, that is, P is quasi-convex in y for any (x_1, x_2). Now, consider the P-function $f(x_1, x_2) = \arg\min_y P(x_1, x_2, y)$. For $x_1 = 0.4$ and $x_2 = 0.6$, the set of minimizers is empty.

4. A proposal for the definition of penalty functions and related results

Taking into account all the problems in the definitions of penalty function existing in the literature, as we discussed in the previous section, in the following we propose a new definition that encompasses all conditions for it to satisfy the desired properties of the notion of penalty, allowing for its use in applications.

Denote by $\mathbb{I} \subseteq \mathbb{R}$ any closed real interval.

Definition 4.1. For any closed interval $\mathbb{I} \subseteq \mathbb{R}$, the function $P : \mathbb{I}^{n+1} \to \mathbb{R}^+$ is a penalty function if and only if there exists $c \in \mathbb{R}^+$ such that:

(P4.1-1) $P(\vec{x}, y) \geq c$, for all $\vec{x} \in \mathbb{I}^n, y \in \mathbb{I}$;
(P4.1-2) $P(\vec{x}, y) = c$ if and only if $x_i = y$, for all $i = 1 \ldots n$, and
(P4.1-3) P is quasi-convex lower semi-continuous in y for each $\vec{x} \in \mathbb{I}^n$.

Remark 4.1. Observe that, differently from Definition 3.9, Definition 4.1 considers only non-negative c values (see Remark 3.8).

Definition 4.2. Let $f : \mathbb{I}^n \to \mathbb{I}$ be a function and P be a penalty function in the sense of Definition 4.1. Then f is said to be a P-function if, for each $\vec{x} \in \mathbb{I}^n$, one has that

$$f(\vec{x}) = \frac{a+b}{2} \tag{9}$$

where

$$[a, b] = cl(Minz(P(\vec{x}, \cdot)))$$

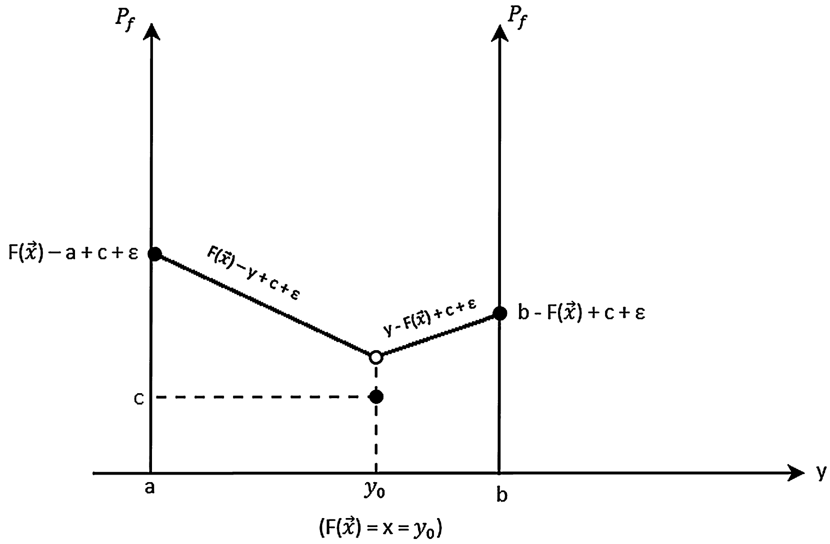

Fig. 1. A penalty function $P_f : [a, b]^{n+1} \to \mathbb{R}^+$.

and $Minz(P(\vec{x}, \cdot))$ is the set of minimizers of $P(\vec{x}, \cdot)$, that is:

$$Minz(P(\vec{x}, \cdot)) = \{y \in \mathbb{I} \mid P(\vec{x}, y) \leq P(\vec{x}, z), \text{ for each } z \in \mathbb{I}\}, \tag{10}$$

and $cl(S)$ is the closure of $S \subseteq \mathbb{I}$.

Remark 4.2. Regarding Definition 4.2, observe that, from Corollary 2.1, the properties of quasi-convexity and lower semicontinuity of a penalty function P imply that the set of minimizers of $P(\vec{x}, \cdot)$ is either a singleton or an interval (see Definition 3.6 and 3.8).

Theorem 4.1. *A function $f : \mathbb{I}^n \to \mathbb{I}$ is a P-function in the sense of Definition 4.2 if and only if f is idempotent.*

Proof. ($\Rightarrow$) If $P : \mathbb{I}^{n+1} \to \mathbb{R}^+$ is a penalty function in the sense of Definition 4.1, then, considering the conditions **(P4.1-1)** and **(P4.1-2)**, the proof is analogous to the proof of Theorem 3.2 ($\Rightarrow$).

($\Leftarrow$) Consider an idempotent function $f : \mathbb{I}^n \to \mathbb{I}$, $\epsilon > 0$, $c \geq 0$ and the function $P_f : \mathbb{I}^{n+1} \to \mathbb{R}^+$, defined, for all $\vec{x} \in \mathbb{I}^n$ and $y \in \mathbb{I}$ by:

$$P_f(\vec{x}, y) = \begin{cases} c & \text{if } x_i = y \text{ for each } i \\ \mid f(\vec{x}) - y \mid + c + \epsilon & \text{otherwise.} \end{cases} \tag{11}$$

Fig. 1 helps to follow the proof. It is immediate that: **(P4.1-1)** $P(\vec{x}, y) \geq c$, for all $\vec{x} \in \mathbb{I}^n$, $y \in \mathbb{I}$, and **(P4.1-2)** $P(\vec{x}, y) = c$ if and only if $x_i = y$, for all $i = 1 \ldots n$. To prove **(P4.1-3)**, we firstly show that P_f is quasi-convex in y for each fixed $\vec{x} \in \mathbb{I}^n$. For that, observe that one of the following cases holds:

1. $P_f(\vec{x}, y) = c$. In this case, one has that $P_f(\vec{x}, y) = g_c(y)$, where g_c is the constant function defined in Equation (1), with $c \geq 0$, which is quasi-convex as shown in Example 2.1.
2. $P_f(\vec{x}, y) = \mid f(\vec{x}) - y \mid + c + \epsilon$. In this case it holds that:

$$P_f(\vec{x}, y) = g_{\| f(\vec{x})}(y) + c + \epsilon = g_{g_{\| f(\vec{x})}, (c+\epsilon)}(y),$$

where $g_{\| f(\vec{x})}$ is defined by Equation (3), with $\theta \in [0, 1]$, and $g_{g_{\| f(\vec{x})}, (c+\epsilon)}$ is defined by Equation (2), with $k = c + \epsilon > 0$. By Corollary 2.2, one has that $g_{\| f(\vec{x})}$ is quasi-convex. Then, as shown in Example 2.1, $g_{g_{\| f(\vec{x})}, (c+\epsilon)}$ is also quasi-convex.

It remains to show that P_f is lower semi-continuous in y. Since f is idempotent, then, for all $y_0 \in \mathbb{I}$ and $\vec{x} \in \mathbb{I}^n$, whenever $x_i = y_0$, for all i, then one has that $f(\vec{x}) = f(y_0, \ldots, y_0) = y_0$. It follows that:

$$\liminf_{y \to y_0} P_f(\vec{x}, y) = \mid f(\vec{x}) - y_0 \mid + c + \epsilon \geq c = P_f(\vec{x}, y_0),$$

where lim inf is the limit inferior (of the function P_f at point y_0) for a fixed $\vec{x}$. This concludes the proof that P_f is a penalty function in the sense of Definition 4.1. In addition, since the set of minimizers of $P_f(\vec{x}, \cdot)$ is the singleton $\{f(\vec{x})\}$, we have that $f(\vec{x}) = \arg\min_y P_f(\vec{x}, y)$. □

Observe that Theorem 4.1 provides a method for the construction of a penalty function P_f for each considered idempotent f. However, such penalty function P_f is not continuous (see Remark 3.7). In the following example, we show that using spread measures it is possible to define a continuous penalty function P_f, for an idempotent function f.

Example 4.1. Consider an idempotent function $f : \mathbb{I}^n \to \mathbb{I}$, $c > 0$ and the function $P_f : \mathbb{I}^{n+1} \to \mathbb{R}^+$, defined, for all $\vec{x} \in \mathbb{I}^n$ and $y \in \mathbb{I}$ by:

$$P_f(\vec{x}, y) = \mid f(\vec{x}) - y \mid + V(\vec{x}) + c, \tag{12}$$

where V is a strict continuous spread measure. We show that P_f is a continuous penalty function:

(P4.1-2) If $P_f(\vec{x}, y) = c$ then one has that $f(\vec{x}) - y = 0$ and $V(\vec{x}) = 0$. Since V is strict, by Proposition 2.5, V is strong, and thus, $\vec{x} = (k, \ldots, k)$, for some $k \in \mathbb{I}$. Since f is idempotent, it follows that $x_i = k = y$, for all i. On the other hand, whenever $x_i = y$, for all i, then, since f is idempotent, it holds that:

$$P_f(\underbrace{y, \ldots, y}_{n}, y) = |f(y, \ldots, y) - y| + V(y, \ldots, y) + c = c.$$

(P4.1-1) Clearly, $P(\vec{x}, y) > c$ if $x_i \neq y$ for some i.
(P4.1-3) The function P_f is clearly continuous (and, thus, lower semicontinuous), and, for each fixed $\vec{x}$, it is defined on a bounded, closed interval, so it attains a minimum. Furthermore, considering Example 2.1 and Corollary 2.2, one has that

$$P_f(\vec{x}, y) = g_{g_{\| f(\vec{x})}, (V(\vec{x})+c)}(y),$$

and, then, P_f is quasi-convex in y (note that the quasi-convexity in y does not depend on V). Thus, by Corollary 2.1, the set of minimizers is in fact a non-empty connected set. Then, P_f is a continuous penalty function (in all variables), in the sense of Definition 4.1, and $f(\vec{x}) = \arg\min_y P_f(\vec{x}, y)$.

The next example shows another way of constructing penalty functions.

Example 4.2. Consider $\mathbb{I} = [a, b] \subseteq \mathbb{R}$, $a \neq b$. Let $W = (w_1, \ldots, w_n)$ be a weight vector, with $\sum_{i=1}^{n} w_i = 1$, and $A : \mathbb{I}^n \to \mathbb{I}$ be an aggregation function such that

$$A(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i x_i \Leftrightarrow x_1 = \ldots = x_n.$$

Then the function $P_A : \mathbb{I}^{n+1} \to \mathbb{R}^+$, defined, for all $\vec{x} \in \mathbb{I}^n$ and $y \in \mathbb{I}$, by:

$$P_A(x_1, \ldots, x_n, y) = \begin{cases} \mid y - A(x_1, \ldots, x_n) \mid & \text{if } y = \sum_{i=1}^{n} w_i x_i, \\ b - a & \text{otherwise} \end{cases}$$

is a penalty function. It follows that, considering $c = 0$, trivially, **(P4.1-1)** holds for all $\vec{x} \in \mathbb{I}^n$ and $y \in \mathbb{I}$. To show **(P4.1-2)**, observe that whenever $P_A(x_1, \ldots, x_n, y) = 0$ then $y = A(x_1, \ldots, x_n)$. Thus, one has that $y = \sum_{i=1}^{n} w_i x_i =$

$A(x_1, \dots, x_n)$, and, then, it holds that $x_1 = \dots = x_n = x$, that is, $y = \sum_{i=1}^{n} w_i x_i = x \sum_{i=1}^{n} w_i = x$. On the other hand, if $x_1 = \dots = x_n = y$, then one has that $y = \sum_{i=1}^{n} w_i x_i = A(x_1, \dots, x_n)$, and, thus, $P_A(x_1, \dots, x_n, y) = 0$.

Before we prove the property **(P4.1-3)**, observe that, since A is clearly idempotent (and, then, it is an averaging function), for each $\vec{x} \in \mathbb{I}^n$ it holds that

$$\min\{x_1, \dots, x_n\} \le A(x_1, \dots, x_n) \le \max\{x_1, \dots, x_n\}.$$

Thus, for each $y \in \mathbb{I}$, one has that

$$b - a \ge \max\{|\, y - \min(x_1, \dots, x_n)\,|, |\, y - \max(x_1, \dots, x_n)\,|\} \ge |\, y - A(x_1, \dots, x_n)\,|\,.$$

Therefore, it holds that $P_A(x_1, \dots, x_n, y) \le b - a$.

Now, suppose that $y = \lambda y_1 + (1 - \lambda) y_2$, for some $\lambda \in [0, 1]$. Then, one has the following cases:

1. $y_1 \neq y$ or $y_2 \neq y$: then either

   $$P_A(x_1, \dots, x_n, y_1) = b - a$$

   or

   $$P_A(x_1, \dots, x_n, y_2) = b - a,$$

   and, therefore, it is immediate that

   $$P_A(x_1, \dots, x_n, y) \le b - a = \max\{P_A(x_1, \dots, x_n, y_1), P_A(x_1, \dots, x_n, y_2)\}.$$

2. $y = y_1 = y_2$: then trivially

   $$P_A(x_1, \dots, x_n, y) = \max\{P_A(x_1, \dots, x_n, y_1), P_A(x_1, \dots, x_n, y_2)\}.$$

Hence, the conclusion is that P_A is quasi-convex.

To show the lower semi-continuity of P_A in y, fix $\vec{x} \in \mathbb{I}^n$ and consider $y_0 \in \mathbb{I}$. If $y_0 \neq \sum_{i=1}^{n} w_i x_i$ then it holds that

$$\liminf_{y \to y_0} P_A(\vec{x}, y) = b - a = P_A(\vec{x}, y_0).$$

On the other hand, if $y_0 = \sum_{i=1}^{n} w_i x_i$ then one has the following cases:

1. $y_0 = x_1 = \dots = x_n$: in this case, we have that

   $$\liminf_{y \to y_0} P_A(\vec{x}, y) = b - a > 0 = P_A(\vec{x}, y_0).$$

2. $y_0 \neq x_i$ for some i: here, one has that

   $$\liminf_{y \to y_0} P_A(\vec{x}, y) = b - a \ge |\, y_0 - A(x_1, \dots, x_n)\,| = P_A(\vec{x}, y_0).$$

It follows that P_A is lower semi-continuous. Thus, **(P4.1-3)** is valid. In addition, since the set of minimizers of $P_A(\vec{x}, \cdot)$ is $\{\sum_{i=1}^{n} w_i x_i\}$, we have that $\sum_{i=1}^{n} w_i x_i = \arg\min_y P_A(\vec{x}, y)$. Moreover, in case that $x_1 = \dots = x_n$, then one has that $A(\vec{x}) = \arg\min_y P_A(\vec{x}, y)$.

From Theorems 3.2 and 4.1, one has that a function $f : [0, 1]^n \to [0, 1]$ is a P-function in the sense of Definition 4.2 if and only if f is a P-function in the sense of Definition 3.8.

Corollary 4.1. *Let $A : \mathbb{I}^n \to \mathbb{I}$ be an increasing function. A is an averaging aggregation function if and only if A is a P-function in the sense of Definition 4.2.*

Proof. ($\Rightarrow$) Since A is an averaging aggregation function, than it is idempotent. The result follows from Theorem 4.1. ($\Leftarrow$) It is immediate. $\square$

Observe that Corollary 4.1 says that each *increasing* P-function is an averaging aggregation function.

Lemma 4.1. *Let $f : \mathbb{I}^n \to \mathbb{I}$ be an idempotent function. Then there is a penalty function P such that f is a P-function and, for each $\vec{x} \in \mathbb{I}^n$, P has just one minimizer, that is, $Minz(P(\vec{x'}, \cdot))$ is a degenerate interval.*

Proof. It is straightforward, since P_f in Equation (11) is a penalty function, in the sense of Definition 4.1, that has just one minimizer for each $\vec{x} \in \mathbb{I}^n$. $\square$

In what follows, denote the set of closed sub-intervals of $\mathbb{I}$ by $\mathbb{II} = \{[x, y] \subseteq \mathbb{I} \mid x \leq y\}$.

Theorem 4.2 *(Characterization of averaging aggregation functions based on penalty functions). A function $A : \mathbb{I}^n \to \mathbb{I}$ is an averaging aggregation function if and only if A is a P-function in the sense of Definition 4.2 for a penalty function P satisfying:*

$$\forall \vec{x}, \vec{x'} \in \mathbb{I}^n : \vec{x} \leq \vec{x'} \Rightarrow cl(Minz(P(\vec{x}, \cdot))) \leq_{KM} cl(Minz(P(\vec{x'}, \cdot))), \tag{13}$$

where $Minz$ is defined in Equation (10) and $\leq_{KM}$ is the Kulisch–Miranker interval order restricted to $\mathbb{II}$ [28], defined, for all $[x, y], [x', y'] \in \mathbb{II}$, by

$$[x, y] \leq_{KM} [x', y'] \Leftrightarrow x \leq x' \text{ and } y \leq y'.$$

Proof. ($\Rightarrow$) Let $A : \mathbb{I}^n \to \mathbb{I}$ be an averaging aggregation function. It follows from Lemma 4.1 that there is a penalty function P_A in the sense of Definition 4.1, such that A is a P_A-function and the sets of minimizers of P_A are always degenerate intervals. Consider $\vec{x}, \vec{x'} \in \mathbb{I}^n$ such that $\vec{x} \leq \vec{x'}$. By Lemma 4.1, one has that $Minz(P_A(\vec{x}, \cdot)) = [a, a]$ and $Minz(P_A(\vec{x'}, \cdot)) = [a', a']$, for some $a, a' \in \mathbb{I}$. Then, since A is increasing, it holds that $A(\vec{x}) \leq A(\vec{x'})$ and, therefore, by Equation (9), it follows that $a = \frac{a+a}{2} \leq \frac{a'+a'}{2} = a'$. Hence, it holds that $Minz(P_A(\vec{x}, \cdot)) \leq_{KM} Minz(P_A(\vec{x'}, \cdot))$, and, thus, $cl(Minz(P_A(\vec{x}, \cdot))) \leq_{KM} cl(Minz(P_A(\vec{x'}, \cdot)))$. Therefore, P_A satisfies Equation (13).

($\Leftarrow$) Denote $[a, b] = cl(Minz(P(\vec{x}, \cdot)))$ and $[a', b'] = cl(Minz(P(\vec{x'}, \cdot)))$ and suppose that A is a P-function in the sense of Definition 4.2 for a penalty function P satisfying (13). Then, for all $\vec{x}, \vec{x'} \in \mathbb{I}^n$ such that $\vec{x} \leq \vec{x'}$ it holds that $[a, b] \leq_{KM} [a', b']$, that is, $a \leq a'$ and $b \leq b'$. Then one has that $A(\vec{x}) = \frac{a+b}{2} \leq \frac{a'+b'}{2} = A(\vec{x'})$, and, therefore, A is increasing. The result follows from Corollary 4.1. $\square$

Remark 4.3. Recently non-monotone averaging functions were discussed in [12,18]. The mode is a prominent example of a non-monotone averaging idempotent function, which is also weakly monotone. Other examples include robust estimators of location such as the least trimmed squares and the least median of squares estimators [12,18,29]. Non-monotone idempotent functions can also be expressed as minimizers of certain expressions. The mode, for example, is the minimizer of the following function

$$P(\vec{x}, y) = \sum_{i=1}^{n} p(x_i, y) \qquad \text{where} \quad p(x_i, y) = \begin{cases} 0, & \text{if } x_i = y, \\ 1 & \text{otherwise.} \end{cases} \tag{14}$$

Note that this function is not quasi-convex, and therefore is not a penalty function in the sense of Definition 4.1. Beliakov and Wilkin [18, Definition 12] proposed the following definition suitable for many non-monotone averaging functions:

Table 1
Comparison among the different definitions of penalty functions.

Property	Definition 3.6	Definition 3.8	Definition 3.9	Definition 4.1
$P(\vec{x}, y) \geq 0$	X	X		
$x_i \neq y$ for some $i \Rightarrow P(\vec{x}, y) > 0$		X		
$x_i = y \Rightarrow P(\vec{x}, y) = 0$	X	X		
$P(\vec{x}, y) = 0 \Leftrightarrow x_i = y$		X		
$P(\vec{x}, y) \geq c, c \in \mathbb{R}$			X	
$P(\vec{x}, y) \geq c, c \in \mathbb{R}^+$				X
$P(\vec{x}, y) = c \Leftrightarrow x_i = y$			X	X
Quasi-convexity			X	X
Lower semi-continuity				X
The set of minimizers is either a singleton or an interval	X	X		X

For all $\vec{x} \in \mathbb{I}^n$, $y \in \mathbb{I}$, $i = 1 \ldots n$.

For any closed interval $\mathbb{I} \subseteq \mathbb{R}$, the function $P : \mathbb{I}^{n+1} \to \mathbb{R}^+$ is a quasi-penalty function if it satisfies:

(i) $P(\vec{x}, y) \geq c \quad$ for all $\vec{x} \in \mathbb{I}^n,\ y \in \mathbb{I}$ and some constant $c \in \mathbb{R}$;
(ii) $P(\vec{x}, y) = c$ if and only if all $x_i = y$, for all $i = 1 \ldots n$, and
(iii) $P(\vec{x}, y)$ is lower semi-continuous in y for any $\vec{x}$.

We note that the values of c can be restricted to non-negative numbers. Also note that this definition differs from Definition 4.1 by the absence of quasi-convexity of P. As a consequence of lower semi-continuity, P necessarily has a non-empty set of minimizers (see Proposition 2.4), but it is not necessarily an interval. For example, for any $\vec{x}$ with distinct components, the function (14) has multiple minimizers at x_i, $i = 1, \ldots, n$, which all can be taken as the mode of such an input vector.

5. Conclusion

In this paper, we recovered the evolution of the definition of penalty functions since it was initially proposed, analyzing different definitions as well their related results, and discussing their properties and problems. We then provided a new definition of penalty function, which requires the conditions that are compatible with what one means intuitively by penalty function, but suppressing the theoretical gaps we have found in the previous definitions, so allowing its application in real-world problems, and providing the basis for further theoretical developments.

We also gave an example of penalty functions based on spread measures, and, to deal with non-monotonic (or weakly/directionally monotonic) averaging functions, we discussed the definition of quasi-penalty functions, so providing more flexibility in their use in different applications. Several results were presented, as the characterization of averaging functions based on penalty functions.

In order to summarize the discussion provided in this paper, Table 1 presents a general comparison among the main properties of the different definitions of penalty functions after Calvo and Beliakov [11] (Definition 3.6).

In the future, we intend to study the incorporation of weights to our proposed approach.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science and Technology under the project TIN2013-40765-P, by the Brazilian funding agency CNPq under Processes 481283/2013-7, 306970/2013-9, 232827/2014-1 and 307681/2012-2, and by APVV-14-0013.

References

[1] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions: A Guide for Practitioners, Springer, Berlin, 2007.
[2] M. Grabisch, J. Marichal, R. Mesiar, E. Pap, Aggregation Functions, Cambridge University Press, Cambridge, 2009.

[3] G. Lucca, J. Sanz, G. Pereira Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H. Bustince Sola, Pre-aggregation functions: construction and an application, IEEE Trans. Fuzzy Syst. 24 (2) (2016) 260–272.

[4] G. Lucca, G.P. Dimuro, V. Mattos, B. Bedregal, H. Bustince, J.A. Sanz, A family of Choquet-based non-associative aggregation functions for application in fuzzy rule-based classification systems, in: 2015 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, IEEE, Los Alamitos, 2015, pp. 1–8.

[5] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, H. Bustince, Enhancing multi-class classification in FARC-HD fuzzy classifier: on the synergy between n-dimensional overlap functions and decomposition strategies, IEEE Trans. Fuzzy Syst. 23 (5) (2015) 1562–1580.

[6] D. Paternain, J. Fernandez, H. Bustince, R. Mesiar, G. Beliakov, Construction of image reduction operators using averaging aggregation functions, Fuzzy Sets Syst. 261 (2015) 87–111.

[7] E. Barrenechea, J. Fernandez, M. Pagola, F. Chiclana, H. Bustince, Construction of interval-valued fuzzy preference relations from ignorance functions and fuzzy preference relations. Application to decision making, Knowl.-Based Syst. 58 (2014) 33–44.

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

[9] E.P. Klement, R. Mesiar, E. Pap, Triangular norms. Position paper II: general constructions and parameterized families, Fuzzy Sets Syst. 145 (3) (2004) 411–438.

[10] B.C. Bedregal, G.P. Dimuro, H. Bustince, E. Barrenechea, New results on overlap and grouping functions, Inf. Sci. 249 (2013) 148–170.

[11] T. Calvo, G. Beliakov, Aggregation functions based on penalties, Fuzzy Sets Syst. 161 (10) (2010) 1420–1436.

[12] G. Beliakov, H. Bustince, T. Calvo, A Practical Guide to Averaging Functions, Springer, Berlin, New York, 2016.

[13] R.R. Yager, Toward a general theory of information aggregation, Inf. Sci. 68 (3) (1993) 191–206.

[14] R.R. Yager, A. Rybalov, Understanding the median as a fusion operator, Int. J. Gen. Syst. 26 (3) (1997) 239–263.

[15] T. Calvo, R. Mesiar, R.R. Yager, Quantitative weights and aggregation, IEEE Trans. Fuzzy Syst. 12 (1) (2004) 62–69.

[16] G. Beliakov, T. Calvo, S. James, Consensus measures constructed from aggregation functions and fuzzy implications, Knowl.-Based Syst. 55 (2014) 1–8.

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

[18] T. Wilkin, G. Beliakov, Weakly monotonic averaging functions, Int. J. Intell. Syst. 30 (2) (2015) 144–169.

[19] S. Ilanko, A. Tucker, The use of negative penalty functions in solving partial differential equations, Commun. Numer. Methods Eng. 21 (3) (2005) 99–106.

[20] H. Askes, S. Piercy, S. Ilanko, Tyings in linear systems of equations modelled with positive and negative penalty functions, Commun. Numer. Methods Eng. 24 (11) (2008) 1163–1169.

[21] M. Gagolewski, Spread measures and their relation to aggregation functions, Eur. J. Oper. Res. 241 (2) (2015) 469–477.

[22] H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, Directional monotonicity of fusion functions, Eur. J. Oper. Res. 244 (1) (2015) 300–308.

[23] A.J. Kurdila, M. Zabarankin, Convex Functional Analysis, Springer, Berlin, 2005.

[24] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.

[25] R.A. Gordon, The Integrals of Lebesgue, Denjoy, Perron, and Henstock, American Mathematics Society, Rhode Island, 1994.

[26] J. Lázaro, T. Rückschlossová, T. Calvo, Shift invariant binary aggregation operators, Fuzzy Sets Syst. 142 (1) (2004) 51–62.

[27] B. De Baets, Idempotent uninorms, Eur. J. Oper. Res. 118 (3) (1999) 631–642.

[28] G.P. Dimuro, B.C. Bedregal, R.H.N. Santiago, R.H.S. Reiser, Interval additive generators of interval t-norms and interval t-conorms, Inf. Sci. 181 (18) (2011) 3898–3916.

[29] P. Rousseeuw, A. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 2003.