Weakly monotone averaging functions

Tim Wilkin, Gleb Beliakov
School of Information Technology, Deakin University,
221 Burwood Hwy, Burwood 3125, Australia
tim.wilkin@deakin.edu.au, gleb@deakin.edu.au

Abstract

Monotonicity with respect to all arguments is fundamental to the definition of aggregation functions. It is also a limiting property that results in many important non-monotonic averaging functions being excluded from the theoretical framework. This work proposes a definition for weakly monotonic averaging functions, studies some properties of this class of functions and proves that several families of important non-monotonic means are actually weakly monotonic averaging functions. Specifically we provide sufficient conditions for weak monotonicity of the Lehmer mean and generalised mixture operators. We establish weak monotonicity of several robust estimators of location and conditions for weak monotonicity of a large class of penalty-based aggregation functions. These results permit a proof of the weak monotonicity of the class of spatial-tonal filters that include important members such as the bilateral filter and anisotropic diffusion. Our concept of weak monotonicity provides a sound theoretical and practical basis by which (monotone) aggregation functions and non-monotone averaging functions can be related within the same framework, allowing us to bridge the gap between these previously disparate areas of research.

Keywords: aggregation functions, monotonicity, means, penalty-based functions, non-monotonic functions

1. Introduction

The aggregation of several input variables into a single representative output arises naturally as a problem in many practical applications and domains. The research effort has been disseminated throughout various fields including economics, computer science, mathematics and engineering, with the subsequent mathematical formulation of aggregation problems having coalesced into a significant body of knowledge concerning aggregation functions. A wide range of aggregation functions are presented in the literature, including the weighted quasi-arithmetic means, ordered weighted averages, triangular norms and co-norms, Choquet and Sugeno integrals and many more. Several recent books provide a comprehensive overview of this field of study (Beliakov et al. [5], Grabisch et al. [17], Torra [32]).

Aggregation functions are commonly used within fuzzy logic, where logical connectives are typically modeled using triangular norms and triangular co-norms. Beyond this field the averaging functions - more commonly known as *means* - that are frequently applied in decision problems, statistical analysis and in image and signal processing. Means have been an important tool and topic of study for over two millennia, with examples such as the arithmetic, geometric and harmonic means known to the Greeks (Rubin [28]). Each of these means shares a fundamental property with the broader class of aggregation functions; that of monotonicity with respect to all arguments (Beliakov et al. [5], Grabisch et al. [17], Torra [32]). There are though many means of significant practical and theoretical importance that are non-monotonic and hence not classified as aggregation functions. For example, a non-monotonic average of pixel intensities within an image subset is used to perform tasks such as image reduction (Wilkin [35]), filtering (van den Boomgaard and van de Weijer [34], Sylvain et al. [30]) or smoothing (Barash and Comaniciu [3]). Within statistics, robust estimators of location are used to estimate the central tendency of a data set and the mode, an average possibly known to the Greeks (Rubin [29]), is a classic example of a non-monotonic average.

Monotonicity with respect to all arguments has an important interpretation in decision making problems: an increase in one criterion should not lead to the decrease of the overall score or utility. However, in image processing an increase in only one pixel value above its neighbours may be due to noise or corruption and should not necessarily increase the intensity value that represents that region. Accordingly, the averaging functions used in such applications do not fit within the established theories regarding aggregation functions and are typically dealt with only from the signal processing perspective.

There are also many non-monotonic means appearing in the literature, with the mode, Gini means, Lehmer means, Bajraktarevic means (Beliakov et al. [5], Bullen [8]) and mixture functions (Ribeiro and Marques Pereira [24], Marques Pereira and Ribeiro [20]) being particularly well known cases. Ideally we would like a formal framework for averaging functions that encompasses non-monotonic means and places them in context with existing monotonic aggregation functions, enabling us to better understand the relationships within this broad class of functions. In so doing we are then able to broaden our understanding of non-monotonic averaging as an aggregation problem.

We achieve this aim herein by relaxing the monotonicity requirement for averaging aggregation functions and propose a new definition that encompasses many non-monotonic averaging functions. We justify this approach by the following interpretation of averaging: while we accept that an increase in one input, or coordinate, may lead to a decrease of the aggregate value, we argue that the same increase coincident in all inputs should only lead to an increase of the aggregate value. This is akin to the property of shift-invariance, which along with homogeneity is one of the basic requirements of the non-monotonic location estimators (Rousseeuw and Leroy [27]). We do not impose shift-invariance though, as that would severely limit the range of averaging functions that fall under our definition averaging functions (for instance, the only shift invariant quasi-arithmetic means are weighted arithmetic means). Rather we consider the property of directional monotonicity in the direction of the vector $(1, 1, \ldots, 1)$, which is obviously implied by shift-invariance as well as by the standard definition of monotonicity. We call this property *weak monotonicity* within the context of aggregation functions and we investigate it herein.

The remainder of this article is structured as follows. In Section 2 we provide the necessary mathematical foundations that underpin the subsequent material. Section 3 provides the main definitions and presents various properties of weakly monotone aggregation functions. Within Section 4 we examine several non-monotonic means and prove that they are, in fact, weakly monotonic. In Section 5 we draw our conclusions and discuss future research directions arising as a result of this investigation.

2. Preliminaries

2.1. Aggregation functions

In this article we make use of the following notations and assumptions. Without loss of generality we assume that the domain of interest is any closed, non-empty interval $\mathbb{I} \subseteq \bar{\mathbb{R}} = [-\infty, \infty]$ and that tuples in $\mathbb{I}^n$ are defined as $\mathbf{x} = (x_{i,n} | n \in \mathbb{N}, i \in \{1, ..., n\})$. We write x_i as the shorthand for $x_{i,n}$ such that it is implicit that $i \in \{1, ..., n\}$. Furthermore, $\mathbb{I}^n$ is ordered such that for $\mathbf{x}, \mathbf{y} \in \mathbb{I}^n$, $\mathbf{x} \leq \mathbf{y}$ implies that each component of $\mathbf{x}$ is no greater than the corresponding component of $\mathbf{y}$. Unless otherwise stated, a constant vector given as $\mathbf{a}$ is taken to mean $\mathbf{a} = a(\underbrace{1, 1, \ldots, 1}_{n \text{ times}}) = a\mathbf{1}$, where $a \in \mathbb{R}$ is a constant and n is implicit within the context of use.

The vector $\mathbf{x}_{\nearrow}$ denotes the result of permuting the vector $\mathbf{x}$ such that its components are in non-decreasing order, that is, $\mathbf{x}_{\nearrow} = \mathbf{x}_{\sigma}$, where σ is the permutation such that $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \ldots \leq x_{\sigma(n)}$. Similarly, the vector $\mathbf{x}_{\searrow}$ denotes the result of permuting $\mathbf{x}$ such that $x_{\sigma(1)} \geq x_{\sigma(2)} \geq \ldots \geq x_{\sigma(n)}$. We will make use of the common shorthand notation for a sorted vector, being $\mathbf{x}_{()} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$. In such cases the ordering will be stated explicitly and then $x_{(k)}$ represents the kth largest or smallest element of $\mathbf{x}$ accordingly.

Consider now the following definitions:

Definition 1. A function $F : \mathbb{I}^n \to \bar{\mathbb{R}}$ is **monotonic** (non-decreasing) if and only if, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{I}^n, \mathbf{x} \leq \mathbf{y}$ then $F(\mathbf{x}) \leq F(\mathbf{y})$.

Definition 2. A function $F : \mathbb{I}^n \to \mathbb{I}$ is an **aggregation function** in $\mathbb{I}^n$ if and only if F is monotonic non-decreasing in $\mathbb{I}$ and $F(\mathbf{a}) = a$, $F(\mathbf{b}) = b$, with $\mathbb{I}^n = [a, b]^n$.

Definition 3. A function F is called **idempotent** if for every input $\mathbf{x} = (t,\, t,\, \ldots,\, t)$, $t \in \mathbb{I}$ the output is $F(\mathbf{x}) = t$.

The functions of most interest in this article are those that have averaging behaviour.

Definition 4. A function F has **averaging behaviour** (or is averaging) if for every $\mathbf{x}$ it is bounded by $\min(\mathbf{x}) \leq F(\mathbf{x}) \leq \max(\mathbf{x})$.

Aggregation functions that have averaging behaviour are idempotent, whereas idempotency and monotoicity imply averaging behaviour.

Definition 5. A function is called **internal** if its value coincides with one of the arguments.

Of particular relevance is the notion of shift-invariance Calvo et al. [10], Lázaro et al. [19] (which is also called difference scale invariance Grabisch et al. [17]). A constant change in every input should result in a corresponding change of the output.

Definition 6. A function $F : \mathbb{I}^n \to \mathbb{I}$ is **shift-invariant** *(stable for translations) if* $F(\mathbf{x} + a\mathbf{1}) = F(\mathbf{x}) + a$ whenever $\mathbf{x}, \mathbf{x} + a\mathbf{1} \in \mathbb{I}^n$.

Definition 7. A function F is **homogeneous** (with degree one) if $F(a\mathbf{x}) = aF(\mathbf{x})$ for all $a\mathbf{x} \in \mathbb{I}^n$.

Aggregation functions that are shift-invariant and homogeneous are known as *linear aggregation functions*. The canonical example of a linear aggregation function is the arithmetic mean.

2.2. Means

The term *mean* is used synonymously with averaging aggregation functions. Chisini's definition of a mean as an average states that the mean of n independent variables $(x_1, ..., x_n)$, with respect to a function F, is a value M for which replacement of each value x_i in the input by M, results in the output M (Chisini [12], stated in Grabisch et al. [17]). I.e.,

$$F(x_1, ..., x_n) = F(M, ..., M) = M$$

As was noted by de Finetti (de Finetti [13], stated in Grabisch et al. [17]), Chisini's definition does not necessarily satisfy Cauchy's requirement that a mean be an internal value (Cauchy [11]). However, by assuming that F is a non-decreasing, idempotent function, then existence, uniqueness and internality of M are restored to Chisini's definition. Gini (Gini [16], p.64), writes that an average of several quantities is a value obtained as a result of a certain procedure, which equals to either one of the input quantities, or a new value that lies in between the smallest and the largest input. The requirement that F be non-decreasing is too strict given the aims of this article and as such, following many authors (e.g., Gini [16], Bullen [8]), we take the definition of a mean to be any averaging (and hence idempotent) function.

Definition 8. A function $M : \mathbb{I}^n \to \mathbb{I}$ is called a **mean** if and only if it is averaging.

The basic examples of (monotonic) means found within the literature include weighted arithmetic mean, weighted quasi-arithmetic mean, ordered weighted average (OWA), order statistic $S_k(\mathbf{x}) = x_{(k)}$, and the median. Less known examples include Choquet and Sugeno integrals and their special cases; the logarithmic means, Heronian means, Bonferroni means and others Bullen [8], Grabisch et al. [17].

In continuing, we wish to consider a broader class of means to include those that are not necessarily monotonic. A classic example is the mode, being the most frequent input, which is routinely used in statistics. [1] The mode is not monotonic as the following example shows. Taking the vectors $\mathbf{x} = (1,1,2,2,3,3,3), \mathbf{y} = (1,1,0,0,0,0,0)$, and $\mathbf{z} = (1,1,1,1,1,1,1)$, then $Mode(\mathbf{x}) = 3, Mode(\mathbf{x}+\mathbf{y}) = 2$ and $Mode(\mathbf{x}+\mathbf{z}) = 4$.

An important class of means that are not always monotonic are those expressed by the Mean of Bajraktarevic, which is a generalisation of the weighted quasi-arithmetic means.

Definition 9. Mean of Bajraktarevic. Let $\mathbf{w}(t) = (w_1(t), ..., w_n(t))$ be a vector of weight functions $w_i : \mathbb{I} \to [0, \infty)$, and let $g : \mathbb{I} \to \bar{\mathbb{R}}$ be a strictly monotonic function. The mean of Bajraktarevic is the function

$$M^g_{\mathbf{w}}(\mathbf{x}) = g^{-1}\left(\frac{\sum_{i=1}^{n} w_i(x_i)g(x_i)}{\sum_{i=1}^{n} w_i(x_i)}\right) \tag{2.1}$$

When $g(x_i) = x_i$, and all weight functions are the same, the Bajraktarevic mean is called a *mixture function* (or *mixture operator*) and is given by

$$M_w(\mathbf{x}) = \frac{\sum_{i=1}^{n} w(x_i)x_i}{\sum_{i=1}^{n} w(x_i)} \tag{2.2}$$

[1] In general the mode is multivalued, so in order to make it a single-variate function, a convention is needed to select one of the multiple outputs, e.g. the smallest.

For the case where the weight functions are are distinct $w_i(x_i)$, the operator $M_{\mathbf{w}}(\mathbf{x})$ is a *generalised mixture function*. A particularly interesting sub-class of Bajraktarevic means are Gini means, obtained by setting $w_i(t) = w_i t^q$ and $g(t) = t^p$ when $p \neq 0$, or $g(t) = \log(t)$ if $p = 0$.

$$G_w(\mathbf{x}; p, q) = \left(\frac{\sum_{i=1}^{n} w_i x_i^{p+q}}{\sum_{i=1}^{n} w_i x_i^q} \right)^{\frac{1}{p}} \tag{2.3}$$

Gini means generalise the (weighted) power means (for $q = 0$) and hence include the minimum, maximum and the arithmetic mean as special cases. Another special case of the Gini mean is the Lehmer, or counter-harmonic mean, obtained when $p = 1$. The contra-harmonic mean is the Lehmer mean with $q = 1$. We will investigate the Lehmer mean and its properties further in Section 4.

2.3. Penalty based functions

In Calvo and Beliakov [9] it was demonstrated that averaging aggregation functions can be expressed as the solution of a minimisation problem of the form

$$F(\mathbf{x}) = \arg\min_{y} \mathcal{P}(\mathbf{x}, y) \tag{2.4}$$

where $\mathcal{P}(\mathbf{x}, y)$ is a penalty function satisfying the following definition:

Definition 10. Penalty function. The function $\mathcal{P} : \mathbb{I}^{n+1} \to \mathbb{R}$ is a penalty function if and only if it satisfies:

1. $\mathcal{P}(\mathbf{x}, y) \geq c \quad \forall\, \mathbf{x} \in \mathbb{I}^n,\ y \in \mathbb{I}$;
2. $\mathcal{P}(\mathbf{x}, y) = c$ if and only if all $x_i = y$; and,
3. $\mathcal{P}(\mathbf{x}, y)$ is quasi-convex in y for any $\mathbf{x}$,

for some constant $c \in \mathbb{R}$ and any closed, non-empty interval $\mathbb{I}$.

A function P is quasi-convex if all its sublevel sets are convex, that is $S_\alpha(P) = \{x | P(x) \leq \alpha\}$ are convex sets for all α, see Rockafellar [25]. The first two conditions ensure that $\mathcal{P}$ has a strict minimum and that a consensus of inputs ensures minimum penalty, providing idempotence of $F(\mathbf{x})$. The third condition implies a unique minimum (but possibly many minimisers that form a convex set). Since multiplication by, or addition of a constant to $\mathcal{P}$ will not change the minimisation, $\mathcal{P}$ may be shifted (if desired) so that $c = 0$. One can think of $\mathcal{P}$ as describing the dissimilarity or disagreement between the inputs $\mathbf{x}$ and the value y. It follows that F is a function that minimises the chosen dissimilarity. It is not necessary to explicitly state F, provided a suitable penalty function is given and the optimisation problem solvable. Subsequently it is sufficient to solve (2.4) to obtain the aggregate $\mu = F(\mathbf{x})$.

Non-monotonic averaging functions can also be represented by a penalty function. For penalty-based functions we have the following results due to Calvo and Beliakov [9].

Theorem 1. *Any idempotent function $F : \mathbb{I}^n \to \mathbb{I}$ can be represented as a penalty based function $\mathcal{P} : \mathbb{I}^{n+1} \to \mathbb{I}$ such that*

$$F(\mathbf{x}) = \arg\min_y \mathcal{P}(\mathbf{x}, y).$$

Corollary 1. *Any averaging function can be expressed as a penalty based function.*

As mentioned in Mesiar et al. [22], mixture functions can be written as a penalty function with

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} w(x_i)(x_i - y)^2.$$

Clearly the necessary condition of the minimum is

$$\mathcal{P}_y(\mathbf{x}, y) = -2\left(\sum_{i=1}^{n} w(x_i)x_i - y\sum_{i=1}^{n} w(x_i)\right) = 0.$$

Hence $\mathcal{P}(\mathbf{x}, y)$ defines a mixture function. A representation of a function as a penalty based function sometimes can simplify technical proofs, as we shall see later in the paper.

It is apparent given the examples presented that many means are non-monotonic and thus not aggregation functions according to Definition 2. In the next section we introduce weak monotonicity and consider some properties of weakly monotonic averaging functions. We subsequently investigate several important examples and show that they are indeed weakly monotonic functions, allowing us to place them in a new framework with existing averaging aggregation functions.

3. Weak monotonicity

3.1. Main definition

As mentioned in Section 1 we are motivated by two important issues. The first one is that there exist many means that are not generally monotonic and hence not aggregation functions, while the second one is that there are many practical applications in which non-monotonic means have shown to provide good aggregate values commensurate with the objectives of the aggregation. To encapsulate these non-monotonic means within the framework of aggregation functions we aim to relax the monotonicity condition and present the class of *weakly monotonic averaging functions*. The definition of weak monotonicity provided herein is prompted by applications and intuition, which suggests that it is reasonable to expect that a representative value of the inputs does not decrease if all the inputs are increased by the same amount (or shifted uniformly) as the relative positions of the inputs are not changed. A formal definition that conveys this property is as follows.

Definition 11. *A function f is called* **weakly monotonic** *non-decreasing (or directionally monotonic) if $F(\mathbf{x}+a\mathbf{1}) \geq F(\mathbf{x})$ for any $a > 0$, $\mathbf{1} = (\underbrace{1, 1, \ldots, 1}_{n\ times})$, such that $\mathbf{x}, \mathbf{x} + a\mathbf{1} \in \mathbb{I}^n$.*

Remark 1. If F is directionally differentiable in its domain then weak monotonicity is equivalent to non-negativity of the directional derivative $D_{\mathbf{1}}(F)(\mathbf{x}) \geq 0$.

Remark 2. Evidently monotonicity implies weak monotonicity, hence all aggregation functions are weakly monotonic. By Definition 6 all shift-invariant functions are also weakly monotonic. It is self evident that weakly monotonic non-decreasing functions form a cone in the linear vector space of weakly monotonic (increasing or decreasing) functions.

3.2. Properties

Let us establish some useful properties of weakly monotonic averages. Consider the function $F : \mathbb{I}^n \to \mathbb{I}$ formed by the composition $F(\mathbf{x}) = A(B_1(\mathbf{x}), B_2(\mathbf{x}))$, where A, B_1 and B_2 are means.

Proposition 1. *If A is monotonic and B_1, B_2 are weakly monotonic, then F is weakly monotonic.*

Proof. By weak monotonicity $B_i(\mathbf{x} + a\mathbf{1}) \geq B_i(\mathbf{x})$ implies that $\exists \delta_i \geq 0$ such that $B_i(\mathbf{x} + a\mathbf{1}) = B_i(\mathbf{x}) + \delta_i$, with $\mathbf{x}, \mathbf{x} + a\mathbf{1} \in \mathbb{I}^n$. Thus $F(\mathbf{x} + a\mathbf{1}) = A(b_1 + \delta_1, b_2 + \delta_2)$, where $b_i = B_i(\mathbf{x})$. The monotonicity of A ensures that $A(b_1 + \delta_1, b_2 + \delta_2) \not< A(b_1, b_2)$ and hence $F(\mathbf{x} + a\mathbf{1}) \geq F(\mathbf{x})$ and F is weakly monotonic. □

By trivial extension, since all monotonic functions are also weakly monotonic, then if either of B_1 or B_2 is monotonic, then F is again weakly monotonic.

Proposition 2. *If A is weakly monotonic and B_1, B_2 are shift invariant, then F is weakly monotonic.*

Proof. Shift invariance implies that $\forall a : B_i(\mathbf{x} + a\mathbf{1}) = B_i(\mathbf{x}) + a$, with $\mathbf{x}, \mathbf{x} + a\mathbf{1} \in \mathbb{I}^n$. Thus $F(\mathbf{x} + a\mathbf{1}) = A(b_1 + a, b_2 + a)$, where $b_i = B_i(\mathbf{x})$. The weak monotonicity of A ensures that $A(b_1 + a, b_2 + a) \not< A(b_1, b_2)$ and hence $F(\mathbf{x} + a\mathbf{1}) \geq F(\mathbf{x})$ and F is weakly monotonic. □

Consider functions of the form $\varphi(\mathbf{x}) = (\varphi(x_1), \varphi(x_2), ..., \varphi(x_n))$.

Proposition 3. *If A is weakly monotonic and $\varphi(x)$ is a linear function then the $\varphi-$transform $A_\varphi(\mathbf{x}) = F(\mathbf{x}) = \varphi^{-1}(A(\varphi(\mathbf{x})))$ is weakly monotonic.*

Proof. $\varphi(x) = \alpha x + \beta$ and hence $\varphi(x + a) = \alpha(x + a) + \beta = \alpha x + \beta + \alpha a = \varphi(x) + c$. Hence

$$\begin{aligned} F(\mathbf{x} + a\mathbf{1}) &= \varphi^{-1}\left(A(\varphi(x_1 + a), ..., \varphi(x_n + a))\right) \\ &= \varphi^{-1}\left(A\left(\varphi(\mathbf{x}) + c\mathbf{1}\right)\right) \\ &= \frac{A\left(\varphi(\mathbf{x}) + c\mathbf{1}\right) - \beta}{\alpha} \\ &\geq \frac{A\left(\varphi(\mathbf{x})\right) - \beta}{\alpha} \\ &= \varphi^{-1}\left(A(\varphi(\mathbf{x}))\right) \end{aligned}$$

by weak monotonicity of A. Hence $F(\mathbf{x} + a\mathbf{1}) \geq F(\mathbf{x})$ and F is weakly monotonic. □

Note that unlike in the case of standard monotonicity, a nonlinear φ-transform does not always preserve weak monotonicity.

Corollary 2. *The dual A^d of a weakly monotonic function A is weakly monotonic under standard negation.*

The following result is relevant to an application of weakly monotonic averages in image processing, discussed in Section 4.3.

Theorem 2. *Let $f : \mathbb{I}^n \to \mathbb{I}$ be a shift invariant function, and g be a function. Let F be a penalty based averaging function with the penalty $\mathcal{P}$ depending on the terms $g\left(x_i - f(\mathbf{x})\right)(x_i - y)^2$. Then F is shift-invariant and hence weakly monotonic.*

Proof. Let

$$\mu = \arg\min_y \mathcal{P}\left(g\left(x_1 - f(\mathbf{x})\right)(x_1 - y)^2, ..., g\left(x_n - f(\mathbf{x})\right)(x_n - y)^2\right).$$

Then

$$\arg\min_y \mathcal{P}(\mathbf{x} + a\mathbf{1}, y) = \arg\min_y \mathcal{P}\left(g\left(x_1 + a - f(\mathbf{x} + a\mathbf{1})\right)(x_1 + a - y)^2, ...\right.$$
$$\left. ..., g\left(x_n + a - f(\mathbf{x} + a\mathbf{1})\right)(x_n + a - y)^2\right) =$$

(by shift invariance)

$$= \arg\min_y \mathcal{P}\left(g\left(x_1 - f(\mathbf{x})\right)(x_1 + a - y)^2, \ldots, g\left(x_n - f(\mathbf{x})\right)(x_n + a - y)^2\right)$$
$$= \mu + a.$$

□

Remark 3. Indeed we need not restrict ourselves to penalty functions with terms depending on $(x_i - y)^2$. Functions D that depend on the differences $x_i - y$ with the minimum $D(0)$ will satisfy the above proof and satisfy the conditions on $\mathcal{P}$ with regards to the existence of solutions to (2.4). In particular, Huber type functions used in robust regression can replace the squares of the differences.

3.3. Counter-cases

For the $\varphi-$transform, if φ is nonlinear then F may or may not be weakly monotonic for all $\mathbf{x}$, which can be observed by example.

Example 1. Take $\mathbf{x} = (1, 8, 16, 35, 47.9)$ and $\varphi(t) = \sqrt{t}$, then $\varphi(\mathbf{x}) = (1, 2\sqrt{2}, 4, \sqrt{35}, \sqrt{47.9})$ and $\varphi(\mathbf{x} + \mathbf{1}) = (\sqrt{2}, 3, \sqrt{17}, 6, \sqrt{48.9})$. If A is the shorth (we prove the weak monotonicity of the shorth in Section 4) then $A(\varphi(\mathbf{x})) = 5.61$ and $A(\varphi(\mathbf{x}+\mathbf{1})) = 2.84$. As $\varphi^{-1}(t) = t^2$ clearly $5.61^2 > 2.84^2$ and $F = A_\varphi$ is not weakly monotonic.

Internal means are not necessarily weakly monotonic, as illustrated by the following example.

Example 2. Take $\mathbf{x} = (x_1, x_2) \in [0, 1]^2$ and

$$F(\mathbf{x}) = \begin{cases} \min(\mathbf{x}) & \text{if } x_1 + x_2 \geq 1 \\ \max(\mathbf{x}) & \text{otherwise} \end{cases}$$

which is internal with values in the set $\{\min(\mathbf{x}), \max(\mathbf{x})\}$. Consider the points $\mathbf{x} = (1/4, 0)$ and $\mathbf{y} = (3/4, 0)$, then $F(\mathbf{x}) = 1/4$ and $F(\mathbf{y}) = 3/4$. It follows that $F(\mathbf{x} + 1/4\mathbf{1}) = 1/2 > F(\mathbf{x})$, however $F(\mathbf{y} + 1/4\mathbf{1}) = 1/4 < F(\mathbf{y})$. Hence this F is not weakly monotonic for all $\mathbf{x} \in \mathbb{I}^2$.

4. Examples of weakly monotonic means

In this section we look at several examples of weakly monotonic, but not necessarily monotonic averaging functions. We begin by considering several of the robust estimators of location, then move on to mixture functions and some interesting cases of means from the literature. While some of the examples involve shift-invariant functions, many of their nonlinear φ-transforms yield proper weakly monotonic functions.

The functions presented below are defined through penalties that are not quasi-convex, therefore we need to drop the condition that $\mathcal{P}$ is quasi-convex from Definition 10.

Definition 12. Quasi-penalty function. The function $\mathcal{P} : \mathbb{I}^{n+1} \to \mathbb{R}$ is a quasi-penalty function if and only if it satisfies:

1. $\mathcal{P}(\mathbf{x}, y) \geq c \quad \forall \mathbf{x} \in \mathbb{I}^n$, $y \in \mathbb{I}$;
2. $\mathcal{P}(\mathbf{x}, y) = c$ if and only if all $x_i = y$; and,
3. $\mathcal{P}(\mathbf{x}, y)$ is lower semi-continuous in y for any $\mathbf{x}$,

for some constant $c \in \mathbb{R}$ and any closed, non-empty interval $\mathbb{I}$.

Note that the third condition ensures the existence of the minimum and a non-empty set of minimisers. In the case where the set of minimisers of $\mathcal{P}$ is not an interval, we need to adopt a reasonable rule for selecting the value of the penalty-based function F. We suggest stating in advance that in such cases we choose the infimum of the set of minimisers of $\mathcal{P}$. From now one $\mathcal{P}$ will refer to quasi-penalty functions.

4.1. Estimators of Location

Perhaps the most widely used estimator of location is the mode, being the most frequent input.

Example 3. Mode: The mode is the minimiser of the (quasi)penalty function

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

It follows that $F(\mathbf{x} + a\mathbf{1}) = \arg\min_y \mathcal{P}(\mathbf{x} + a\mathbf{1}, y) = \arg\min_y \sum_{i=1}^{n} p(x_i + a, y)$, which is minimised for the value $y = F(\mathbf{x}) + a$. Hence, $F(\mathbf{x} + a\mathbf{1}) = F(\mathbf{x}) + a$ and thus the mode is shift invariant. By Definition 6 the mode is weakly monotonic.

Remark 4. Note that the mode may not be uniquely defined, e.g., the mode of $(1, 1, 2, 2, 3, 4, 5)$, in which case we use a suitable convention. The quasi-penalty $\mathcal{P}$ associated with the mode is not quasi-convex, and as such it may have several minimisers. A convention is needed as to which minimiser is selected, e.g., the smallest or the largest. Other examples of non-monotonic means that follow also involve quasi-penalties, and the same convention as for the mode is adopted. Then also discrete scales can be considered, compare to, e.g., the paper of Kolesárová et al. [18].

The Least Trimmed Squares estimator (Rousseeuw and Leroy [27]) rejects up to 50% of the data values as outliers and minimises the squared residual using the remaining data.

Example 4. Least Trimmed Squares (LTS): The LTS uses the (quasi)penalty function

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{h} r_{(i)}^2$$

where $r_{(i)} = S_i(\mathbf{r})$ is the ith order statistic of $\mathbf{r}$, $r_k = x_k - y$ and $h = \left\lfloor \frac{n}{2} \right\rfloor + 1$. If σ is the order permutation of $\{1, ..., n\}$ such that $\mathbf{r}_\sigma = \mathbf{r}_\nearrow$, then the minima of $\mathcal{P}$ occur when $P_y = -2\sum_{i=1}^{h}(x_{\sigma(i)} - y) = 0$, which implies that the minimum value is $\mu = \frac{1}{h}\sum_{i=1}^{h} x_{\sigma(i)}$. Since $S_k(\mathbf{x})$ is shift invariant then $S_i(\mathbf{r}+a\mathbf{1}) = r_{\sigma(i)}+a$ and thus

$$\mathcal{P}(\mathbf{x} + a\mathbf{1}, y) = \sum_{i=1}^{h} v_{\sigma(i)}^2$$

where $v_k = ((x_k + a) - y)$. It follows that the value y that minimises $\mathcal{P}(\mathbf{x} + a\mathbf{1}, y)$ is $y = \mu + a$, hence the LTS is shift invariant and thus weakly monotonic.

The remaining estimators of location presented compute their value using the shortest contiguous sub-sample of $\mathbf{x}$ containing at least half of the values. The candidate sub-samples are the sets $X_k = \{x_j : j \in \{k, k + 1, ..., k + \left\lfloor \frac{n}{2} \right\rfloor\},\ k = 1, ..., \left\lfloor \frac{n+1}{2} \right\rfloor$. The length of each set is taken as $\|X_k\| = \left| x_{k+\left\lfloor \frac{n}{2} \right\rfloor} - x_k \right|$ and thus the index of the shortest sub-sample is

$$k^* = \arg\min_i \|X_i\|,\ i = 1, ..., \left\lfloor \frac{n+1}{2} \right\rfloor.$$

Under the translation $\bar{\mathbf{x}} = \mathbf{x} + a\mathbf{1}$ the length of each sub-sample is unaltered since $\left\|\bar{X}_k\right\| = \left|\bar{x}_{k+\left\lfloor \frac{n}{2} \right\rfloor} - \bar{x}_k\right| = \left|(x_{k+\left\lfloor \frac{n}{2} \right\rfloor} + a) - (x_k + a)\right| = \left|x_{k+\left\lfloor \frac{n}{2} \right\rfloor} - x_k\right| = \|X_k\|$ and thus k^* remains the same.

Consider now the Least Median of Squares estimator (Rousseeuw [26]), which is the midpoint of X_{k^*}.

Example 5. Least Median of Squares (LMS): The LMS can be computed by minimisation of the (quasi)penalty function

$$\mathcal{P}(\mathbf{x}, y) = median\left\{(x_i - y)^2 \,|y \in \mathbb{I},\ x_i \in X_{k^*}\right\}$$

The value y minimises the penalty $\mathcal{P}(\mathbf{x} + a\mathbf{1}, y)$, given by

$$\min_y \mathcal{P}(\mathbf{x} + a\mathbf{1}, y) = \min_y\ median\left\{(x_j + a - y)^2 \,|y \in \mathbb{I}, x_j \in X_{k^*}\right\} = \mathcal{P}(\mathbf{x}, \mu),$$

is clearly $y = \mu + a$. Hence, $F(\mathbf{x} + a\mathbf{1}) = F(\mathbf{x}) + a$ and the LMS is shift invariant and weakly monotonic.

The Shorth (Andrews et al. [1]) is the arithmetic mean of X_{k^*}

Example 6. Shorth: The shorth is given by

$$F(\mathbf{x}) = \frac{1}{h}\sum_{i=1}^{h} x_i,\ \ x_i \in X_{k^*},\ \ h = \left\lfloor \frac{n}{2} \right\rfloor + 1$$

Since the set X_{k^*} is unaltered under translation and the arithmetic mean is shift invariant, then the shorth is shift invariant and hence weakly monotonic.

Example 7. OWA Penalty Functions: Penalty functions having the form

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} w_i S_i\left((\mathbf{x} - y\mathbf{1})^2\right)$$

define regression operators, $F(\mathbf{x})$ (Yager and Beliakov [37]). Consider the following results dependent on the weight vector $\Delta = (w_1, ..., w_n)$.

1. $\Delta = \mathbf{1}$ generates Least Squares regression and F is monotonic and hence weakly monotonic;
2. $\Delta = (0, ..., 0, 1)$ generates Chebyshev regression and F is monotonic and hence weakly monotonic;
3. Since all the terms $S_i\left((\mathbf{x} - y\mathbf{1})^2\right)$ are constant under transformation $(\mathbf{x}, y) \rightarrow (\mathbf{x} + a\mathbf{1}, y + a)$ (cf Theorem 2), the OWA regression operators are shift-invariant for any choice of the weight vector Δ.
4. For $\Delta = \begin{cases} (0, ..., 0_{k-1}, 1/2, 1/2, 0, ...0) & n = 2k \textit{ is even} \\ (0, ..., 0_{k-1}, 1, 0, ..., 0) & n = 2k - 1 \textit{ is odd} \end{cases}$ then F is the Least Median of Squares operator and hence shift invariant and weakly monotonic; and
5. For $\Delta = (1, ..., 1_h, 0, ..., 0), h = \left\lfloor \frac{n}{2} \right\rfloor + 1$ then F is the Least Trimmed Squares operator and hence is shift invariant and weakly monotonic.

In the cases 3-5 the OWA regression operators are not monotonic.

Example 8. Density based means: The density based means were introduced in Angelov and Yager [2]. Let d_{ij} denote the distance between inputs x_i and x_j. The *density based mean* is defined as

$$y = \sum_{i=1}^{n} w_i(\mathbf{x}) x_i, \tag{4.1}$$

where

$$w_i(\mathbf{x}) = \frac{u_i(\mathbf{x})}{\sum_{j=1}^n u_j(\mathbf{x})} = \frac{K_C(\frac{1}{n}\sum_{j=1}^n d_{ij}^2)}{\sum_{k=1}^n K_C(\frac{1}{n}\sum_{j=1}^n d_{kj}^2)}, \tag{4.2}$$

and where K_C is Cauchy kernel given by

$$K_C(t) = (1+t)^{-1}. \tag{4.3}$$

As shown in Beliakov and Wilkin [6] density based means are shift-invariant and hence weakly monotonic. Extensions of the formulas (4.2), (4.3) are also presented.

It may appear that the class of weakly monotonic averages consists mostly of shift-invariant functions, as the above examples illustrate. This impression is due to the fact that such examples came from robust regression, where the very definition of robust estimators of location involve shift-invariance Rousseeuw and Leroy [27]. However, the class of weakly monotonic functions is reacher, as various (but not all) $\varphi-$transforms of shift-invariant functions (with non-linear φ) are weakly monotonic but not shift-invariant. Some results on the conditions on φ which preserve weak monotonicity are presented in Wilkin et al. [36]. A few more examples are presented in the sequel.

4.2. Mixture Functions

The mixture functions were given by Eqn. (2.2), which we recall here for clarity

$$M_w(\mathbf{x}) = \frac{\sum_{i=1}^{n} w(x_i)x_i}{\sum_{i=1}^{n} w(x_i)}.$$

Mesiar et al. [22] have shown that under the constraint that w is non-decreasing and differentiable, if $w(x) \geq w'(x)\cdot(b-x), x \in [a,b] = \mathbb{I}$, then M_w is an aggregation function and hence monotonic (and by extension, also weakly monotonic). Additionally, M_w is invariant to scaling of the weight functions (i.e., $M_{\alpha w} = M_w \ \forall \alpha \in \mathbb{R}\backslash\{0\}$). In Mesiar and Spirkova [21], it was shown that the dual, M_w^d, of M_w is generated by $w(1-x)$.

As mentioned in Section 2, a special case of the Gini means (with $p=1$) are the Lehmer means, which are generally not monotonic. Lehmer means are mixture functions with weight function $w(t) = t^q$, which is neither increasing for all $q \in \mathbb{R}$ nor shift invariant. Note that for $q < 0$ the value of Lehmer means at $\mathbf{x}$ with at least one component $x_i = 0$ is defined as the limit when $x_i \to 0^+$, so that L_q is continuous on $[0, \infty)^n$

We begin by establishing some general properties of Lehmer means.

Lemma 1. *The Lehmer mean $L_q : [0, \infty)^n \to [0, \infty)$, given by*

$$L_q(\mathbf{x}) = \frac{\sum_{i=1}^{n} x_i^{q+1}}{\sum_{i=1}^{n} x_i^{q}}, \qquad q \in \mathbb{R} \tag{4.4}$$

is

1. *homogeneous;*
2. *monotonic (and linear) along the rays emanating from the origin;*
3. *averaging;*
4. *idempotent;*
5. *not generally monotonic in* $\mathbf{x}$*;*
6. *has neutral element* 0 *for* $q > 0$*; and,*
7. *has absorbing element* 0 *for* $q < 0$*.*

The proof is presented in the Appendix.

We now establish a sufficient condition for weak monotonicity of Lehmer means, which depends on both q and the number of arguments n. We provide a relation between these two quantities.

Theorem 3. *The Lehmer mean of n arguments, is weakly monotonic on $[0, \infty)^n$ if $n \le 1 + \left(\frac{q+1}{q-1}\right)^{q-1}$, $q \in \mathbb{R} \setminus (0, 1)$.*

Proof. The Lehmer mean for $q \in [-1, 0]$ is known to be monotonic (Farnsworth and Orr [15]) and hence weakly monotonic in that parameter range. In the range $q \in (0, 1)$ the Lehmer mean is not weakly monotonic, because it's partial derivative at $\mathbf{x} = (a, b)$ when $a \to 0^+$ tends to $-\infty$. Hence we focus on the cases $q \ge 1$ and $q < -1$. The proof is easier to present in penalty-based representation, as the partial derivatives have more compact form. As stated in Section 2.3, $L_q(\mathbf{x})$ can be written as a penalty-based function (2.4) with penalty $\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} x_i^q (x_i - y)^2$. Differentiation w.r.t y yields

$$\mathcal{P}_y(\mathbf{x}, y) = -2 \sum_{i=1}^{n} \left(x_i^{q+1} - x_i^q y\right).$$

At the minimum we have the implicit equation $\mathcal{P}_y = F(\mathbf{x}, y) = 0$, with the necessary condition that yields $y = L_q(\mathbf{x})$. We remind that for any $x_i = 0$ the Lehmer mean is defined in the limit as $x_i \to 0^+$. The partial derivatives $\frac{\partial L_q(\mathbf{x})}{\partial x_i}$ are given by the implicit derivative $\frac{\partial y}{\partial x_i} = -\frac{F_{x_i}}{F_y}$, with

$$F(\mathbf{x}, y) = \sum_{i=1}^{n} x_i^{q+1} - y \sum_{i=1}^{n} x_i^q = 0.$$

By differentiation $F_y(\mathbf{x}, y) = -\sum_{i=1}^{n} x_i^q \leq 0,\ \forall x_i \in [0, \infty)$ and thus the sign of the partial derivatives depends on the sign of F_{x_i}, which is given by

$$F_{x_i}(\mathbf{x}, y) = (q+1)x_i^q - q x_i^{q-1} y.$$

These derivatives can be either positive or negative. To establish weak monotonicity we require that the directional derivative of $L_q(\mathbf{x})$ in the direction $(1, 1, ..., 1)$ be non-negative. We have that $(\mathrm{D}_{\mathbf{1}} L_q)(\mathbf{x}) = \frac{1}{\sqrt{n}} \nabla L_q(\mathbf{x}) \cdot \mathbf{1} = \frac{n}{\sqrt{n} F_y(\mathbf{x}, y)} \sum_{i=1}^{n} F_{x_i}(\mathbf{x}, y)$ and thus the sign of the directional derivative is determined only by the sign of $\sum_{i=1}^{n} F_{x_i}(\mathbf{x}, y)$. We will henceforth work with the sorted inputs, $\mathbf{x}_{()} = \mathbf{x}_{\searrow}$ such that $\mathbf{x}_{(1)}$ is thus the largest input and $x_{(n)}$ the smallest.

Consider first the case: $q \geq 1$.

We examine the term $F_{x_{(1)}}$ and note that $y \leq x_{(1)}$ for any input $\mathbf{x}$ since $L_q(\mathbf{x})$ is averaging (condition 3 of Lemma 1). Then it follows that

$$F_{x_{(1)}} = (q+1)x_{(1)}^q - q x_{(1)}^{q-1} y \geq (q+1)x_{(1)}^q - q x_{(1)}^{q-1} x_{(1)} = x_{(1)}^q \geq 0.$$

For the remaining x_i we compute the smallest possible value of F_{x_i} by selecting the point of minimum value, which is attained for

$$\frac{\partial F_{x_i}}{\partial x_i} = q\,(q+1)\,x_i^{q-1} - q\,(q-1)\,x_i^{q-2} y = 0.$$

At the optimum either $x_i^* = 0$ or

$$\begin{aligned} q(q+1)\,(x_i^*)^{q-1} - q(q-1)\,(x_i^*)^{q-2}\,y = 0& \\ \Rightarrow x_i^* = \left(\frac{q-1}{q+1}\right) y \geq 0.& \end{aligned}$$

At $x_i^* = 0$ we have that $F_{x_i} = 0$ (for $q > 1$) and $F_{x_i} = -y$ (for $q = 1$), and at $x_i^* = \left(\frac{q-1}{q+1}\right) y$ we have that

$$\begin{aligned} F_{x_i}(x_i^*) &= (q+1)\left(\left(\frac{q-1}{q+1}\right) y\right)^q - q\left(\left(\frac{q-1}{q+1}\right) y\right)^{q-1} y \\ &= (q-1)\left(\frac{q-1}{q+1}\right)^{q-1} y^q - q\left(\frac{q-1}{q+1}\right)^{q-1} y^q \\ &= y^q\left(\frac{q-1}{q+1}\right)^{q-1}(q-1-q) \\ &= -y^q\left(\frac{q-1}{q+1}\right)^{q-1} \\ &\geq -x_{(1)}^q\left(\frac{q-1}{q+1}\right)^{q-1}. \end{aligned}$$

Since $(\mathrm{D}_{\mathbf{1}} L_q)(\mathbf{x}) \propto \sum_{i=1}^{n} F_{x_i}$ then

$$(\mathrm{D}_{\mathbf{1}} L_q)(\mathbf{x}) = c\left(F_{x_{(1)}} + \sum_{i=2}^{n} F_{x_{(i)}}\right),$$

and since each $F_{x_{(i)}} \geq -x_{(1)}^q \left(\frac{q-1}{q+1}\right)^{q-1}$ then

$$
\begin{aligned}
(\mathrm{D}_{\mathbf{1}} L_q)(\mathbf{x}) &\geq c\left(F_{x_{(1)}} + (n-1)\left(-x_{(1)}^q\left(\frac{q-1}{q+1}\right)^{q-1}\right)\right) \\
&= c\left(x_{(1)}^q - (n-1)\left(\frac{q-1}{q+1}\right)^{q-1} x_{(1)}^q\right) \\
&= c x_{(1)}^q \left(1-(n-1)\left(\frac{q-1}{q+1}\right)^{q-1}\right).
\end{aligned}
$$

This expression is non-negative and hence $L_q(\mathbf{x})$ is weakly monotonic provided that

$$
(n-1)\left(\frac{q-1}{q+1}\right)^{q-1} \leq 1 \qquad \text{or} \qquad n \leq 1 + \left(\frac{q+1}{q-1}\right)^{q-1}, q > 1.
$$

For $q = 1$ we get $n \leq 2$.

Now consider the case: $q < -1$. We have that

$$
F_{x_i} = \frac{(1-p)x_i + py}{x_i^{p+1}}, \qquad p = |q| > 1
$$

and note that these derivatives are defined in the limit for the case where $x_i = 0$. I.e., $F_{x_i}^+\big|_{x_i=0} = \lim\limits_{x_i \to 0^+} F_{x_i}$. We now examine the term $F_{x_{(n)}}$ and note that $y \geq x_{(n)}$ since $L_q(\mathbf{x})$ is averaging. Thus

$$
\begin{aligned}
F_{x_{(n)}} &= \frac{(1-p)x_{(n)} + py}{x_{(n)}^{p+1}} \\
&\geq \frac{(1-p)x_{(n)} + px_{(n)}}{x_{(n)}^{p+1}} = \frac{1}{x_{(n)}^p}.
\end{aligned}
$$

Again we consider the remaining x_i by seeking the minimum of F_{x_i}, given by

$$
\frac{\partial F_{x_i}}{\partial x_i} = -\frac{p(1-p)}{x_i^{p+1}} - \frac{p(p+1)}{x_i^{p+2}} y = 0.
$$

This attains a minimum at $x_i = \left(\frac{p+1}{p-1}\right) y$ and substitution into F_{x_i} gives

$$
\begin{aligned}
F_{x_i}\left(\frac{p+1}{p-1}y\right) &= \frac{(1-p)\left(\frac{p+1}{p-1}y\right)+py}{\left(\frac{p+1}{p-1}y\right)^{p+1}} \\
&= \frac{-1}{y^p}\left(\frac{p+1}{p-1}\right)^{-(p+1)} \\
&\geq \frac{-1}{x_{(n)}^p}\left(\frac{p+1}{p-1}\right)^{-(p+1)}.
\end{aligned}
$$

The directional derivative of $L_q(\mathbf{x})$ can be written as

$$
\begin{aligned}
(\mathrm{D}_{\mathbf{1}}L_q)(\mathbf{x}) &= c\left(F_{x_{(n)}} + \sum_{i=1}^{n-1} F_{x_{(i)}}\right) \\
&\geq c\left(\frac{1}{x_{(n)}^p} - \frac{n-1}{x_{(n)}^p}\left(\frac{p+1}{p-1}\right)^{-(p+1)}\right) \\
&= \frac{c}{x_{(n)}^p}\left(1-(n-1)\left(\frac{p+1}{p-1}\right)^{-(p+1)}\right).
\end{aligned}
$$

We note that the sign of this derivative does not change in the limit as $x_{(n)} \to 0^+$ and is non-negative for

$$
\begin{aligned}
n &\leq 1 + \left(\frac{p+1}{p-1}\right)^{p+1} \\
&= 1 + \left(\frac{q-1}{q+1}\right)^{-q+1}, \quad q=-p.
\end{aligned}
$$

Hence, in both cases $(q < -1, q \geq 1)$ we obtain the requirement for a non-negative directional derivative - and hence weak monotonicity of $L_q(\mathbf{x})$ - as being $n \leq 1 + \left(\frac{q+1}{q-1}\right)^{q-1}$. For the case $-1 \leq q \leq 0$ this remains a sufficient condition for weak monotonicity, although clearly overly restrictive. □

Remark 5. As suggested to us, as

$$
\left(\frac{q+1}{q-1}\right)^{q-1} = \left(\left(1+\frac{2}{q-1}\right)^{\frac{q-1}{2}}\right)^2,
$$

and the right hand side is increasing (with q) and approaches e^2 as $q \to \infty$, we have a restriction that for all $q > 1$ weak monotonicity holds for at most $n < 9$ arguments.

A further suggestion was to show that Lehmer means are weakly monotonic for any number of arguments for negative powers q. This can be achieved by examining the directional derivative $(\mathrm{D}_{\mathbf{1}} L_q)(\mathbf{x})$ directly, and in the near future we shall formalise this result.

So while Lehmer means are an interesting example of mixture functions (with familiar power functions as the weights) their usefulness in many applications would be limited, as they are weakly monotonic only for a restricted range of (positive) powers and the number of arguments.

Corollary 3. *The contra-harmonic mean ($q = 1$) is weakly monotonic only for two arguments.*

4.3. Spatial-Tonal Filters

The well known class of spatial-tonal filters includes the mode filter (van de Weijer and van den Boomgaard [33]), bilateral filter (Tomasi and Manduchi [31]) and anisotropic diffusion (Perona and Malik [23]) among others. This is an important class of filters developed to preserve edges within images when performing tasks such as filtering or smoothing. While these filters are commonly expressed in integral notation over a continuous space, they are implemented in discrete form over a finite set of pixels that take on finite values in a closed interval. It can be shown that the class of functions is given (in discrete form) by the averaging function

$$F_{\Delta}^{g}(\mathbf{x}; x_1) = \frac{\sum_{i=1}^{n} w_i g(|x_i - x_1|) x_i}{\sum_{i=1}^{n} w_i g(|x_i - x_1|)}, \tag{4.5}$$

where the weights w_i are nonlinear and non-convex functions of the locations of the pixels, which have intensity x_i. In all practical problems the locations are constant and hence can be pre-computed to produce the constant weight vector $\Delta = (w_1, w_2, ..., w_n)$. The pixel x_1 is the pixel to be filtered/smoothed such that its new value is $\bar{x}_1 = F_{\Delta}^{g}(\mathbf{x}; x_1)$.

The function F^g_Δ is nonlinear and not monotonic. It is trivially shown to be expressed as a penalty-based function with penalty

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} w_i g(|x_i - x_1|)(x_i - y)^2.$$

In image filtering applications it is known that this penalty minimises the mean squared error between the filtered image and the noisy source image (Elad [14]). By Proposition 2 it follows directly that the filter F^g_Δ is shift invariant and hence weakly monotonic. Furthermore, Theorem 2 permits us to generalise this class of filters to be those penalty based averaging functions having penalty function

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} w_i g(|x_i - f(\mathbf{x})|)(x_i - y)^2 \tag{4.6}$$

or even further using other bivariate function $D : \mathbb{I}^2 \to \mathbb{R}$ (as discussed in Remark 3)

$$\mathcal{P}(\mathbf{x}, y) = \sum_{i=1}^{n} w_i g(|x_i - f(\mathbf{x})|)D(x_i, y) \tag{4.7}$$

The implication of replacing x_1 with $f(\mathbf{x})$ in the scaling function g is that we may use any shift-invariant aggregation of $\mathbf{x}$, which allows us to account for the possibility that x_1 is itself an outlier within the local region of the image. For example, we could use the median, the mode or the shorth for $f(\mathbf{x})$. This provides an interesting result and invites further research in the application of weakly monotonic means to spatial-tonal filtering and smoothing problems.

5. Conclusion

In this article we have introduced the concept of weakly monotonic averaging functions and examined some of the properties of these functions. We have studied several families of means previously considered to be simply non-monotonic, and shown them to be weakly monotonic. Specifically we have established a sufficient condition for the weak monotonicity of the Lehmer mean - which is an important subclass of the Mean of Bajraktarevic - and shown that several important non-monotonic regression operators are actually weakly monotonic. Additionally we have proven that a large class of penalty-based functions are also weakly monotonic, which admits a very large class of aggregation functions. This has permitted a simple proof that the class of image processing filters known as spatial-tonal filters are weakly monotonic averaging functions. This class subsumes the class of spatial averaging filters, such as the Gaussian blur filter. Most importantly, given the definition of weak monotonicity, all aggregation functions are weakly monotonic and thus we have not needed to redefine monotonic aggregation in order to relate it to weakly monotonic averaging.

This study was prompted by two issues. First, that there exist several important classes of means that fall outside of the current definition of aggregation functions, which requires monotonicity in all arguments. These include the examples presented in Section 4: the robust estimators of location (such as the mode and the shorth), mixture functions and the spatial-tonal filters used extensively in image processing. It appears reasonable to treat these functions within the same framework that includes the monotonic means.

The second issue is that applications such as image processing and robust statistics require non-monotone averaging, where the main concern is that of noise (outliers) within the data. While the average must be a representative value of the inputs, we wish to avoid the possibility that one or more erroneous inputs drives the value of the output. A small increase above the average may be reasonable, however a large increase should permit that input to be discounted or ignored and the average to possibly decrease.

The concept of weakly monotone aggregation addresses both of these issues, bringing the existing (monotone) aggregation functions and many of the non-monotone means into the same framework. Our proposal then is to redefine the class of averaging aggregation functions to be not those functions that are monotonic, but rather the class of weakly monotonic functions that are averaging.

Proposal: A function $F : \mathbb{I}^n \to \mathbb{I}$ is an averaging aggregation function on $\mathbb{I}^n$ if and only if it is weakly monotonic non-decreasing on $\mathbb{I}$ and averaging.

It remains to be seen whether or not weak monotonicity is the minimal requirement for defining averaging aggregation functions and whether this weaker definition is justified for other types of aggregation, such as conjunctive and disjunctive functions. Furthermore, are there other possibilities for the relaxation of monotonicity that provide for a unified framework of aggregation theory and practice? We leave these questions for future work.

References

References

[1] Andrews, D., Bickel, P., Hampel, F., Huber, P., Rogers, W., Tukey, J., 1972. Robust Estimates of Location: Surveys and Advances. Princeton University Press, Princeton, N.J.

[2] Angelov, P., Yager, R., 2013. Density-based averaging - a new operator for data fusion. Information Sciences 222, 163–174.

[3] Barash, D., Comaniciu, D., 2004. A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift. Image and Vision Computing 22, 73–81.

[4] Beliakov, G., James, S., Li, G., 2011. Learning Choquet-integral-based metrics for semisupervised clustering. IEEE Trans. on Fuzzy Syst. 19 (3), 562–574.

[5] Beliakov, G., Pradera, A., Calvo, T., 2007. Aggregation Functions: A Guide for Practitioners. Vol. 221 of Studies in Fuzziness and Soft Computing. Springer-Verlag, Berlin.

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

[7] Blumenson, L. E., 1960. A derivation of n-dimensional spherical coordinates. The American Mathematical Monthly 67 (1), 63–66.

[8] Bullen, P., 2003. Handbook of Means and Their Inequalities. Kluwer, Dordrecht.

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

[10] Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R., 2002. Aggregation operators: properties, classes and construction methods. In: Calvo, T., Mayor, G., Mesiar, R. (Eds.), Aggregation Operators. New Trends and Applications. Physica-Verlag, Heidelberg, New York, pp. 3–104.

[11] Cauchy, A. L., 1821. Cours d'analyse de l'ecole royale polytechnique. Analyse algébrique 1.

[12] Chisini, O., 1929. Sul concetto di media. Periodico di Matematiche 4, 106–116.

[13] de Finetti, B., 1931. Sul concetto di media. Giorn. Ist. Ital. Attuari 2 (3), 369–396.

[14] Elad, M., 2002. On the origin of the bilateral filter and ways to improve it. IEEE Trans. on Image Processing 11 (10), 1141–1151.

[15] Farnsworth, D., Orr, R., 1986. Gini means. American Mathematical Monthly 93 (8), 603–607.

[16] Gini, C., 1958. Le Medie. Unione Tipografico-Editorial Torinese, Milan (Russian translation, Srednie Velichiny, Statistica, Moscow, 1970).

[17] Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E., 2009. Aggregation Functions. Encyclopedia of Mathematics and Its Foundations. Cambridge University Press.

[18] Kolesárová, A., Mayor, G., Mesiar, R., 2007. Weighted ordinal means. Inf. Sciences 177, 3822–3830.

[19] Lázaro, J., Rückschlossová, T., Calvo, T., 2004. Shift invariant binary aggregation operators. Fuzzy Sets and Systems 142, 51–62.

[20] Marques Pereira, R., Ribeiro, R., 2003. Aggregation with generalized mixture operators using weighting functions. Fuzzy Sets and Systems 137, 43–58.

[21] Mesiar, R., Spirkova, J., 2006. Weighted means and weighting functions. Kybernetika 42 (2), 151–160.

[22] Mesiar, R., Spirkova, J., Vavrikova, L., 2008. Weighted aggregation operators based on minimization. Inform. Sci. 178, 113–1140.

[23] Perona, P., Malik, J., 1990. Scale space and edge detection using anisotropic diffusion. IEEE Trans. on Pattern Analysis and Machine Intelligence 12 (7), 629–639.

[24] Ribeiro, R., Marques Pereira, R., 2003. Generalized mixture operators using weighting functions. Europ. J. Oper. Research 145, 329–342.

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

[26] Rousseeuw, P., 1984. Least median of squares regression. Journal of the American Statistical Association 79 (388), 871–880.

[27] Rousseeuw, P., Leroy, A., 1987. Robust Regression and Outlier Detection. John Wiley and Sons.

[28] Rubin, E., 1968. The statistical world of herodotus. The American Statistician 22 (1), 31–33.

[29] Rubin, E., 1971. Quantitative commentary on thucydides. The American Statistician 25 (4), 52–54.

[30] Sylvain, P., Kornprobst, P., Tumblin, J., Durand, F., 2008. Bilateral filtering: Theory and applications. Foundations and Trends in Computer Graphics and Vision 4 (1), 1–73.

[31] Tomasi, C., Manduchi, R., January 1998. Bilateral filtering for gray and color images. In: Proc. of the 6th Int. Conf. on Computer Vision. Bombay, India, pp. 839–846.

[32] Torra, V. Narukawa, Y., 2007. Modeling Decisions. Information Fusion and Aggregation Operators. Springer, Berlin, Heidelberg.

[33] van de Weijer, J., van den Boomgaard, R., 2001. Local mode filtering. In: Computer Vision and Pattern Recognition. Vol. 2. IEEE Computer Society, pp. 428–433.

[34] van den Boomgaard, R., van de Weijer, J., 2002. On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. In: Proc. of the 16th Int. Conf. on Pattern Recognition. Vol. 3. pp. 927–930.

[35] Wilkin, T., 2013. Image reduction operators based on non-monotonic averaging functions. In: Proceedings of the 10th IEEE Int. Conf. on Fuzzy Systems (to appear). Hyderabad, India.

[36] Wilkin, T., Beliakov, G., Calvo, T., 2014. Weakly monotone averaging functions. In: Proceedings of IPMU'2014 Conference. Montpellier, France.

[37] Yager, R., Beliakov, G., 2010. OWA operators in regression problems. IEEE Trans. on Fuzzy Syst. 18, 106–113.

Appendix

Proof of Lemma 1.

Proof. Consider each of the following:

1. **Homogeneous**: Set $\mathbf{x} = \lambda \mathbf{u}$ then

$$L_q(\mathbf{x}) = L_q(\lambda \mathbf{u}) = \frac{\sum_{i=1}^{n} (\lambda u_i)^{q+1}}{\sum_{i=1}^{n} (\lambda u_i)^{q}} = \lambda \frac{\sum_{i=1}^{n} u_i^{q+1}}{\sum_{i=1}^{n} u_i^{q}} = \lambda L_q(\mathbf{u}).$$

Hence L_q is homogeneous with degree 1.

2. **Monotonic (and linear) along the rays**: Consider the generalised spherical coordinates (Blumenson [7]) $(r, \theta, \phi_1, ..., \phi_{n-2})$, $r \geq 0$, $0 \leq \theta \leq 2\pi$, $0 \leq \phi_i \leq \pi$ for the hypersphere $S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = r\}$. We will restrict the angle variables so that $x_i \in [0, \infty)$. The transformation to an orthonormal Euclidean basis E_n produces the vector $\mathbf{x}$ of length r having components

$$\begin{aligned} x_1 &= r \cos(\phi_1) \\ x_j &= r \cos(\phi_j) \prod_{k=1}^{j-1} \sin(\phi_k), \quad j = 2, ..., n-1 \\ x_n &= r \prod_{k=1}^{n-1} \sin(\phi_k) \end{aligned}$$

where $\phi_{n-1} = \theta$ and $0 \leq \phi_i \leq \pi/2,\ i = 1, ..., n-1,\ r \geq 0$. The Lehmer mean of the Euclidean vector $\mathbf{x}$ is therefore

$$
\begin{aligned}
L_q(\mathbf{x}) &= \frac{[r\cos(\phi_1)]^{q+1} + \sum_{j=2}^{n-1}\left[r\cos(\phi_j)\prod_{k=1}^{j-1}\sin(\phi_k)\right]^{q+1} + \left[r\prod_{k=1}^{n-1}\sin(\phi_k)\right]^{q+1}}{[r\cos(\phi_1)]^{q} + \sum_{j=2}^{n-1}\left[r\cos(\phi_j)\prod_{k=1}^{j-1}\sin(\phi_k)\right]^{q} + \left[r\prod_{k=1}^{n-1}\sin(\phi_k)\right]^{q}} \\
&= r\left[\frac{[\cos(\phi_1)]^{q+1} + \sum_{j=2}^{n-1}\left[\cos(\phi_j)\prod_{k=1}^{j-1}\sin(\phi_k)\right]^{q+1} + \left[\prod_{k=1}^{n-1}\sin(\phi_k)\right]^{q+1}}{[\cos(\phi_1)]^{q} + \sum_{j=2}^{n-1}\left[\cos(\phi_j)\prod_{k=1}^{j-1}\sin(\phi_k)\right]^{q} + \left[\prod_{k=1}^{n-1}\sin(\phi_k)\right]^{q}}\right] \\
&= rf(\phi_1, ..., \phi_{n-1})
\end{aligned}
$$

Along rays emanating from the origin each $f(\cdot)$ is constant and hence $L_q(\mathbf{x}) = \alpha_\phi r$ is linear.

3. **Averaging**: Let $\mathbf{x}_\sigma = \mathbf{x}_\searrow$ and take $a = x_{(1)}$ and $b = x_{(m)}$ denote the value of the largest and the smallest non-zero elements of $\mathbf{x}$ respectively. By homogeneity

$$
L_q(\mathbf{x}) = a\frac{1 + \sum_{i=2}^{n}\left(\frac{x_{(i)}}{a}\right)^{q+1}}{1 + \sum_{i=2}^{n}\left(\frac{x_{(i)}}{a}\right)^{q}} = a\frac{1+\alpha}{1+\beta}
$$

Since $x_{(i)} \leq a$ for all $i = 1, ..., n$ then $\alpha \leq \beta$. Hence $L_q(\mathbf{x}) \leq a$. Similarly,

$$
L_q(\mathbf{x}) = b\frac{1 + \sum_{i=1}^{m-1}\left(\frac{x_{(i)}}{b}\right)^{q+1}}{1 + \sum_{i=1}^{m-1}\left(\frac{x_{(i)}}{b}\right)^{q}} = b\frac{1+\gamma}{1+\delta}
$$

Since $x_{(i)} \geq b$ for all $i = 1, ..., m-1$ and $x_{(i)} = 0$ for all $i = m+1, ..., n$ then $\delta \leq \gamma$. Hence $L_q(\mathbf{x}) \geq b$. Thus, $\min(\mathbf{x}) \leq L_q(\mathbf{x}) \leq \max(\mathbf{x})$ and $L_q(\mathbf{x})$ is averaging.

4. **Idempotent**: For any vector $\mathbf{x} = (t, t, ..., t)$ we have that

$$L_q(\mathbf{x}) = \frac{t \sum_{i=1}^{n} t^q}{\sum_{i=1}^{n} t^q} = t$$

and hence L_q is idempotent.

5. **Not generally monotonic in x**: Take $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (1, 1/2)$, then for $q > 0$, $L_q(\mathbf{x}) = 1$ and $L_q(\mathbf{y}) = \frac{1+(1/2)^{q+1}}{1+(1/2)^q} = \frac{\left(2^{q+1}+1\right)}{(2^{q+1}+2)} < 1$. Thus $\mathbf{x} < \mathbf{y}$ and $L_q(\mathbf{x}) > L_q(\mathbf{y})$, hence $L_q(\mathbf{x})$ is not generally monotonic in $\mathbf{x}$ for all $q \in \mathbb{R}$.
6. **Has neutral element of 0 for** $q > 0$: Consider $\mathbf{x} = (a, 0)$ then $L_q(\mathbf{x}) = \lim_{x_2 \to 0^+} \frac{a^{q+1}+x_2^{q+1}}{a^q+x_2^q} = a$ for $q > 0$.
7. **Has absorbing element of** 0 **for** $q < 0$: Consider $\mathbf{x} = (a, 0)$ then $L_q(\mathbf{x}) = \lim_{x_2 \to 0^+} L_q(1, x_2) = \lim_{x_2 \to 0^+} \frac{a+x_2^{q+1}}{a+x_2^q} = \frac{\frac{a}{x_2^q}+x_2}{\frac{a}{x_2^q}+1} = 0.$

$\square$