

This is the authors' final peer reviewed (post print) version of the item published as:

Beliakov, G and James, S 2014, A penalty-based aggregation operator for non-convex intervals, Knowledge-based systems, vol. 70, pp. 335-344.

## Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30067739

Reproduced with the kind permission of the copyright owner

Copyright: 2014, Elsevier

# A penalty-based aggregation operator for non-convex intervals 

Gleb Beliakov and Simon James ${ }^{1}$<br>Deakin University, School of Information Technology, 221 Burwood Hwy, Burwood, Victoria 3125, Australia


#### Abstract

In the case of real-valued inputs, averaging aggregation functions have been studied extensively with results arising in fields including probability and statistics, fuzzy decision-making, and various sciences. Although much of the behavior of aggregation functions when combining standard fuzzy membership values is well established, extensions to interval-valued fuzzy sets, hesitant fuzzy sets, and other new domains pose a number of difficulties. The aggregation of non-convex or discontinuous intervals is usually approached in line with the extension principle, i.e. by aggregating all real-valued input vectors lying within the interval boundaries and taking the union as the final output. Although this is consistent with the aggregation of convex interval inputs, in the non-convex case such operators are not idempotent and may result in outputs which do not faithfully summarize or represent the set of inputs. After giving an overview of the treatment of non-convex intervals and their associated interpretations, we propose a novel extension of the arithmetic mean based on penalty functions that provides a representative output and satisfies idempotency.


Keywords: Aggregation functions, penalty-based functions, interval-valued fuzzy sets, hesitant fuzzy sets, averaging operators

[^0]
# A penalty-based aggregation operator for non-convex intervals 

Gleb Beliakov and Simon James ${ }^{1}$<br>Deakin University, School of Information Technology, 221 Burwood Hwy, Burwood, Victoria 3125, Australia

## 1. Introduction

The arithmetic mean is the standard "go to" operator employed in various sciences, statistics, economics and fuzzy decision making for aggregating a set of inputs into a single representative value. For example, a number of small errors may arise naturally when we conduct repeated experiments or in our data collection, so we use the arithmetic mean to average the results, providing a reasonable estimate of what might be the true value. In fuzzy decision making, the arithmetic mean of the expert evaluations or membership values can be used to compare potential alternatives, allowing us to choose the best overall option. On the other hand, we may simply be interested in a summary statistic that tells us in some way what is normal or expected for a particular set of inputs, e.g. to describe a population in terms of the average life expectancy.

More broadly, the arithmetic mean is one example of an averaging aggregation function [5, 14, 23]. Aggregation functions have been studied for various practical applications and a number of alternatives to the arithmetic mean have been proposed that may perform more reliably for certain types of data. In the face of uncertainty pertaining to the inputs, either arising from linguistic descriptions or data collection methods, the need has also been identified to extend aggregation functions to deal with inputs expressed as intervals [11], pairs of positive and negative information such as Atanassov

[^1]orthopairs [1] or other multiple-valued inputs [2, 3].
Although a number of results have been established for these data types (especially in the field of fuzzy sets aggregation), more recently some researchers have tackled the problem of aggregating inputs provided as nonconvex or discontinuous intervals, i.e. intervals that contain gaps or are comprised of a sequence of disjoint intervals. In statistics research, such inputs can occur through censoring where the data cannot be observed over particular intervals. In probability theory, the study of random sets also gives rise to non-convex sets and the need to calculate their expecatation (see [20] for a detailed overview). For examples in the fuzzy research community, we can mention the hesitant fuzzy sets of Torra and Narukawa [22, 24] - where the input usually denotes a discrete set of possible evaluations between 0 and 1, the generalized grey numbers of Yang and John [29, 30] - where an input is known to lie within a potentially non-convex range of values, and the discontinuous intervals of Wagner et al. [26].

These are inputs of the form

$$
\begin{equation*}
A_{i}=\bigcup_{j=1}^{m_{i}}\left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right] \tag{1}
\end{equation*}
$$

where $\left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right]$denotes the $j$-th interval with $a_{i_{j}}^{+}<a_{i_{j+1}}^{-}, j=1,2, \ldots, m_{i}-1$. It may also be convenient to represent such intervals as a sequence of intervals (as in [16]), i.e. for $A_{i}=\left[a_{i_{1}}^{-}, a_{i_{1}}^{+}\right] \cup, \ldots, \cup\left[a_{i_{m_{i}}}^{-}, a_{i_{m_{i}}}^{+}\right]$we will simply write $A_{i}=\left\langle\left[a_{i_{1}}^{-}, a_{i_{1}}^{+}\right], \ldots,\left[a_{i_{m_{i}}}^{-}, a_{i_{m_{i}}}^{+}\right]\right\rangle$.

We can consider the following situations where it may be useful to work with non-convex intervals.

Example 1 (Uncertainty with travel times [30]). Two trains are scheduled to depart 5 minutes apart however both could be up to 2 minutes late. The earlier train is sometimes full and it takes 3 minutes to travel to the next station. The time it will take a passenger (arriving in time for the first train) to reach the next station can be represented with the input $\langle[3,5],[8,10]\rangle$.

Example 2 (Species population recovery intervals [21]). An ecology expert is asked to provide her estimation of when a species will reach healthy population levels following a forest fire. A species may increase in population immediately following the fire, then decrease as other species start to recover, before increasing again to its pre-disturbance levels. The expert can use the
non-convex interval $\langle[1,2],[8,20]\rangle$ to indicate that the population is predicted to be at healthy levels 1-2 years and 8-20 years after the fire, but below the threshold at other times.
Example 3 (Analysis of recurring health problems [15]). After a patient is treated and released from hospital, they may make a full recovery or sometimes their health will deteriorate and they will be readmitted. In some cases, the patient may be readmitted a number of times. In order to investigate contributing factors, experts represent each patient's time in hospital with non-convex intervals, e.g. an input 〈[12, 13], [16, 18], [24, 27]〉 would indicate a patient was readmitted for 3 periods after 12, 16 and 24 months, and stayed in hospital for various lengths of time.

In each of the examples above, the inputs represent temporal data [16], with an event (or the uncertainty pertaining to an event) taking duration over discontinuous time periods. However there may also be cases where we need to aggregate non-convex intervals that represent spatial data, e.g. in the fusion of sensor readings observing a non-continuous space, measurements that are uncertain because they lie outside an observable range (censored data), or even evaluations in fuzzy decision making [24]. The authors in [26] also note example applications in forensic science, hazard detection, agreement-based modeling and computing with respect to linguistic descriptions.

In research areas such as statistics, a common approach for handling either standard interval or non-convex interval inputs is to represent them with single values, e.g. a mid-point or the most probable value. While this may be effective under certain conditions, working with the inputs in their original form can allow for robust analyses and inferences which are free of assumptions pertaining to the source of uncertainty [11].

The extension of averaging aggregation functions to non-convex inputs in the fuzzy domain has thus far been approached in a manner consistent with the extension principles applied for fuzzy operations (e.g. in [10] for operations on fuzzy numbers) and interval arithmetic [11]. All possible realvalued input vectors $\mathbf{x}$ with $x_{i} \in A_{i} \forall i$ are aggregated and the union of these aggregated values is taken as the output. We contend that whilst this approach is suitable for most situations when the intervals are convex, the non-convex case presents two unique problems:

1. These resulting aggregation functions are not idempotent, i.e. it does not necessarily hold that $f(t, t, \ldots, t)=t$ if $t$ is a non-convex interval. Idempotency is a key property for averaging functions when it is
desired that the output gives a representative or typical value. If we wanted to consider the "average" evaluation from 10 ecology experts in Example 2 above and all 10 provided the same interval $\langle[1,2],[8,20]\rangle$, then we would expect the same non-convex interval to be returned as the output.
2. As the number of inputs grows large, aggregating non-convex inputs in this fashion converges towards aggregating their convex hulls or envelopes (the intervals defined by the lower and upper bounds). This raises the question of whether anything is gained by using non-convex intervals to represent the uncertainty of the model in the first place.

In this article, we approach the problem of aggregating non-convex intervals in the framework of penalty-based functions. We show that existing methods can be recovered with the choice of various penalties and and then propose a new penalty which results in an extension of the arithmetic mean which is idempotent and can more faithfully represent a "typical input".

The article will be structured as follows. In Section 2 we will give an overview of the concepts that underlie the proposed operators. In particular, we look at aggregation functions, penalty-based methods for constructing them, and various types of inputs for which they have been defined. In Section 3, we recall the definitions of functions which have been used in various settings to aggregate non-convex interval valued inputs, noting their relationship to penalty functions defined for intervals. We then consider the problem of defining penalties between non-convex intervals in Section 4 and in Section 5, we propose a new penalty for non-convex inputs and define our new operator. We present some numeric examples in Section 6 to help illustrate differences between the existing and proposed methods, before discussing some other potential approaches in Section 7 and concluding in Section 8.

## 2. Preliminaries

We will firstly provide the basic definitions relating to aggregation functions and show how they can be defined with respect to penalty functions. We then give an overview of various input types which have extended the use of real inputs to incorporate uncertainty into decision processes and modeling applications.

We will consider aggregation functions (see [5, 14, 23]) defined over the unit interval.

Definition 1 (Aggregation function). An aggregation function $f:[0,1]^{n} \rightarrow$ $[0,1]$ is a function non-decreasing in each argument and satisfying $f(0, \ldots, 0)=$ 0 and $f(1, \ldots, 1)=1$.

The monotonicity of aggregation functions is important when used for decision making to ensure that an increase to one of the criteria should not result in a decrease in the overall evaluation. Here we are interested particularly in averaging aggregation functions.

Definition 2 (Averagaing aggregation function). An aggregation function $f$ is considered to be averaging where for $\mathbf{x} \in[0,1]^{n}$,

$$
\min (\mathbf{x}) \leq f(\mathbf{x}) \leq \max (\mathbf{x})
$$

Due to the monotonicity of aggregation functions, averaging behavior is equivalent to idempotency, i.e. $f(t, t, \ldots, t)=t$.

Typical examples include the arithmetic mean (sometimes referred to simply as "the average") and the median. For an input vector $\mathbf{x}$ consisting of $n$ values, the arithmetic mean $A M: \mathbf{x} \in[0,1]^{n} \rightarrow[0,1]$ is given by

$$
\begin{equation*}
A M(\mathbf{x})=\sum_{i=1}^{n} \frac{1}{n} x_{i} \tag{2}
\end{equation*}
$$

If it is desired that the contributions of some inputs affect the overall value more than others, the inputs can also be weighted by replacing the $\frac{1}{n}$ by a weight $w_{i}, \sum_{i=1}^{n} w_{i}=1$.

The median $\operatorname{Med}(\mathbf{x})$, on the other hand is given by the middle value when the inputs are arranged in descending or ascending order. If the number of inputs is even, the median can be any point in the interval bounded by the middle two values, however usually the arithmetic mean or midpoint of these is used.

An aggregation function is said to be internal if its output always coincides with one of its inputs. Internal functions will always be idempotent, however, clearly the reverse is not true. Examples of internal functions include the minimum and maximum functions, and the median when $n$ is odd.

Both the arithmetic mean and median can be expressed as the value which minimizes the sum of differences between the inputs and output, i.e. we have,

$$
W A M(\mathbf{x})=\arg \min _{y} \sum_{i=1}^{n} w_{i}\left(x_{i}-y\right)^{2}
$$

and,

$$
\operatorname{Med}_{\mathbf{w}}(\mathbf{x})=\arg \min _{y} \sum_{i=1}^{n} w_{i}\left|x_{i}-y\right|
$$

For $w_{i}=\frac{1}{n}$ we recover the standard arithmetic mean $A M(\mathbf{x})$ and median $\operatorname{Med}(\mathbf{x})$.

Where functions can be defined in this fashion, we can refer to them as penalty-based functions $[28,8,7,6]$. For inputs $x_{i} \in[0,1]$ we have the following definition.

Definition 3 (Penalty function). A penalty function $P:[0,1]^{n+1} \rightarrow[0, \infty]$ satisfies:
i) $P(\mathbf{x}, y)=0$ if $x_{i}=y \forall i$;
ii) $P(\mathbf{x}, y)>0$ if $x_{i} \neq y$ for some $i$;
iii) For every fixed $\mathbf{x}$, the set of minimizers of $P(\mathbf{x}, y)$ is either a singleton or an interval ${ }^{2}$.

The penalty-based function is then given by

$$
f(\mathbf{x})=\arg \min _{y} P(\mathbf{x}, y),
$$

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

Remark 1. Conditions i) and ii) imply that $P(\mathbf{x}, y)=0$ if and only if all $x_{i}=y$, hence if all $x_{i}$ are the same then the only minimizer of $P$ is $y=x_{i}$ and the function will be idempotent. We can relax the second condition (as in $[7,6])$ and instead require $\left.i i^{\prime}\right) P(\mathbf{x}, y) \geq 0$ if $x_{i} \neq y$ for some $i$, however this has the potential to lead to operators which are not idempotent. Condition $i$ ) can also be relaxed such that a unique minimum is reached (since adding a constant will not change the minimizer).

[^2]In particular we can focus on the class of penalty-based functions expressed as the sum of all individual penalties $p\left(x_{i}, y\right)$,

$$
\begin{equation*}
f(\mathbf{x})=\arg \min _{y} \sum_{i=1}^{n} w_{i} p\left(x_{i}, y\right) \tag{3}
\end{equation*}
$$

The weighted or unweighted arithmetic mean then corresponds with the penalty $p\left(x_{i}, y\right)=\left(x_{i}-y\right)^{2}$ while the median is related to the absolute differences $p\left(x_{i}, y\right)=\left|x_{i}-y\right|$.

To deal with various forms of uncertainty in decision making contexts, some research has looked to extend the definitions of aggregation functions to other types of inputs. In the theory of fuzzy sets, examples such as intervalvalued fuzzy membership, Atanassov orthopairs [1] (also referred to as intuitionistic fuzzy values) and hesitant fuzzy sets can all be considered as special cases of fuzzy multisets studied by Yager and Miyamoto [18, 19, 27]. Where we might usually assign a degree of membership to each element in a set, fuzzy multisets allow multiple membership values to be assigned.

We recall the following definitions.
Definition 4 (Interval-valued input). An interval valued input $\bar{a}_{i} \subseteq[0,1]$ is given by $\bar{a}_{i}=\left[a_{i}^{-}, a_{i}^{+}\right]$where $a_{i}^{-}$and $a_{i}^{+}$denote the lower and upper endpoints respectively with $0 \leq a_{i}^{-} \leq a_{i}^{+} \leq 1$.

Uncertainty resulting in interval-valued data can arise in a number of ways, from precision of measurement, rounding and missing data, to what is referred to in statistics as censoring, i.e. when measurements lie outside a window of observation or detectability (See [11] for an overview of intervalvalued data and methods for analysis).

Although arising independently and studied in different contexts, there exist many similarities between the so-called grey numbers and intervalvalued fuzzy sets. Grey numbers are used in the theory of grey sets [9] and essentially denote an uncertain value by the subset of potential candidates, over which the probabilistic likelihood is considered to be equal. Although typically grey numbers have been denoted by intervals, the concept of generalized grey numbers has been considered in [29, 30].

Definition 5 (Generalized grey number). Let $g_{i}^{ \pm} \in \Re$ be an unknown
real number within a union set of $m_{i}$ intervals $^{3}$,

$$
\begin{equation*}
g_{i}^{ \pm} \in \bigcup_{j=1}^{m_{i}}\left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right] \tag{4}
\end{equation*}
$$

where $\sup \left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right]<\inf \left[a_{i_{j+1}}^{-}, a_{i_{j+1}}^{+}\right]$for $j=1,2, \ldots, m-1$. We denote the probability that $g_{i}^{ \pm} \in\left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right]$by $p_{j}$ and if it holds that $p_{j}>0$ and $\sum_{i=1}^{m_{i}} p_{j}=1$, then we call $g_{i}^{ \pm}$a generalized grey number.

Remark 2. The variable $g_{i}^{ \pm}$itself is not considered to be an interval or collection of intervals, but rather a definite value that lies within the (possibly non-convex) range of values (see [30] for a discussion of key differences in interpretation). Grey numbers none-the-less have a representation with equivalent notation to a non-convex interval. For example, a grey number $g_{i}^{ \pm}=\langle[0.2,0.3],[0.45,0.8]\rangle$ is some real value that lies either between 0.2 and 0.3 or between 0.45 and 0.8 , however it is equally likely that it could be any of these.

Research into the operations for grey numbers denoted by convex intervals has been conducted in [29] and some applications to decision making have been studied (e.g. [17]). The theory of more generalized grey numbers, in particular the extension of aggregation operators to non-convex intervalvalued grey numbers, was considered in [30].

Another extension of fuzzy sets with the possibility of multiple membership values are the hesitant fuzzy sets of Torra and Narukawa [22, 24]. Hesitant fuzzy sets are essentially fuzzy multisets with a different underlying interpretation. We present the definition as relating to fuzzy membership.

Definition 6 (Hesitant fuzzy membership). For a given element of a reference set, the set of hesitant membership degrees will be denoted $H_{i}$ and can be any convex or non-convex subset of the unit interval, i.e. $H_{i} \subseteq[0,1]$.

Torra notes that while interval values and Atanassov orthopairs can be considered as special cases, the typical case of a hesitant fuzzy set is one where the possible membership values are expressed as a finite subset of

[^3]the unit interval. For example $H_{i}=\langle\{0.4\},\{0.5\},\{0.6\}\rangle$ indicates that the membership value of the $i$-th input could be $0.4,0.5$ or 0.6 . This would be the case in any setting where only a finite set of evaluations is permitted, i.e. the decision maker here believes that any score from 0.4 to 0.6 is possible, however values such as 0.45 or 0.423 are not permissible evaluations.

Remark 3. Although the notation of hesitant fuzzy values may give the impression of them representing sets of sets (intervals and or single values), they are more precisely the union of multiple membership functions. This means the intervals or singletons should be non-overlapping as is the case with Eqs. (1) and (4).

## 3. Existing averaging functions defined for convex and non-convex interval inputs

As we have discussed, functions such as the arithmetic mean can be conceived from different perspectives: as an overall evaluation, as a representative value or summary statistic and as the minimizer of squared differences between the output and inputs. In some research areas, a common approach for handling interval or subset-valued inputs is to represent them with single values, e.g. a mid-point or centroid. While this may be effective under certain conditions, working with the inputs in their original form can allow for robust analyses and inferences which acknowledge the uncertainty inherent to the context [11]. In this section we will look at existing extensions of averaging aggregation functions, taking note of how their properties and behavior transfer to other types of inputs.

The straightforward approach for extending existing functions and arithmetic operations to interval-valued inputs has been to take the set of all possible outputs that result when real values are supplied for each interval [11], i.e. for a multivariate function $f$ defined on the reals and intervals $\bar{a}_{1}, \bar{a}_{2}, \ldots, \bar{a}_{n}$ we have,

$$
\begin{equation*}
f^{I}\left(\bar{a}_{1}, \bar{a}_{2}, \ldots, \bar{a}_{n}\right)=\left\{f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \mid x_{1} \in \bar{a}_{1}, x_{2} \in \bar{a}_{2}, \ldots, x_{n} \in \bar{a}_{n}\right\} \tag{5}
\end{equation*}
$$

For three or more inputs, and for non-monotone functions, this will lead to a set of (potentially) overlapping intervals. If the output is to be used as the basis for decision-making or is intended to give a representative summary of the inputs, it is preferable that it be defined over the same space as the inputs. From Eq. (5) it is then common to take the union (or convex
hull obtained from the minimum and maximum results). This leads to the following expression for the arithmetic mean.

$$
\begin{equation*}
A M^{I}\left(\bar{a}_{1}, \bar{a}_{2}, \ldots, \bar{a}_{n}\right)=\left[A M\left(a_{1}^{-}, a_{2}^{-}, \ldots, a_{n}^{-}\right), A M\left(a_{1}^{+}, a_{2}^{+}, \ldots, a_{n}^{+}\right)\right] . \tag{6}
\end{equation*}
$$

Where the output can be determined from separate functions acting on the lower and upper bounds respectively, we say the extended function is representable. The penalty-based expression of such functions is straightforward, since we can simply add the penalties associated with each input in the framework of Eq. (3). In other words, we consider the penalty or distance between intervals to be given by $p\left(\bar{a}_{i}, \bar{y}\right)=\left(a_{i}^{-}-y^{-}\right)^{2}+\left(a_{i}^{+}-y^{+}\right)^{2}$, which leads to the following penalty-based expression for $A M^{I}$.

$$
\begin{equation*}
A M^{I}\left(\bar{a}_{1}, \bar{a}_{2}, \ldots, \bar{a}_{n}\right)=\arg \min _{\bar{y}} \sum_{i=1}^{n}\left(a_{i}^{-}-y^{-}\right)^{2}+\left(a_{i}^{+}-y^{+}\right)^{2} \tag{7}
\end{equation*}
$$

Remark 4. Alternative penalties or measures of distance between intervals could also be used. If we use the squared differences between the mid-points for $p\left(\bar{x}_{i}, \bar{y}\right)$, then the resulting function will reduce to the standard arithmetic mean of these values. Other options include the Wasserstein distance, which weights the squared difference of the mid-points with the half-length of the radius of the intervals [25].

As an extension of the arithmetic mean, $A M^{I}$ exhibits a number of desirable properties. Whether it is expressed in terms of Eq. (5), (6) or (7), it will coincide with the arithmetic mean if each of the interval inputs are equivalent to singletons (i.e. $a_{i}^{-}=a_{i}^{+}, \forall i$ ), is monotone, idempotent and "averaging" in the sense of being bounded by the infimum and supremum of the inputs ${ }^{4}$, and minimizes a penalty based on squared differences. Furthermore, we note that the output, which will be a convex interval in general, serves well to summarize the inputs with a representative value.

The problem of aggregating non-convex inputs has been considered in various contexts. For hesitant fuzzy sets, generalized grey sets and the nonconvex intervals in [26], averaging operators have been defined in a way consistent with Eq. (5), taking the union of all outputs that could be obtained

[^4]from the possible combinations of real-valued inputs. The extension of the arithmetic mean can again be simplified by considering the union of all intervals $\left\{A M^{I}\left(\bar{a}_{1}, \bar{a}_{2}, \ldots, \bar{a}_{n}\right) \mid \bar{a}_{1} \in A_{1}, \bar{a}_{2} \in A_{2}, \ldots, \bar{a}_{n} \in A_{n}\right\}$ where $\bar{a}_{i} \in A_{i}$ indicates that $\bar{a}_{i}$ is one of the distinct convex intervals comprising $A_{i}$. We hence obtain the following definition.

Definition 7 (Arithmetic mean $A M^{U}$ for non-convex intervals). We consider a set of (potentially) non-convex interval inputs $A_{i}=\left\langle\left[a_{i_{1}}^{-}, a_{i_{1}}^{+}\right], \ldots\right.$, $\left.\left[a_{i_{m_{i}}}^{-}, a_{i_{m_{i}}}^{+}\right]\right\rangle$. The arithmetic mean $A M^{U}$ is given by

$$
\begin{align*}
& A M^{U}\left(A_{1}, A_{2}, \ldots, A_{n}\right)= \\
& \quad \bigcup_{j_{1}=1}^{m_{1}} \ldots \bigcup_{j_{n}=1}^{m_{n}} A M^{I}\left(\left[a_{1_{j_{1}}}^{-}, a_{1_{j_{1}}}^{+}\right],\left[a_{2_{j_{2}}}^{-}, a_{2_{j_{2}}}^{+}\right], \ldots,\left[a_{n_{j_{n}}}^{-}, a_{n_{j_{n}}}^{+}\right]\right), \tag{8}
\end{align*}
$$

where $A M^{I}$ is defined as it is in Eq. (6).
The following example helps illustrate the practical calculation of $A M^{U}$.
Example 4. Let $A_{1}=\langle[0.6,0.8],[0.9,1]\rangle$ and $A_{2}=\langle[0,0.3],[0.7,1]\rangle$ be two non-convex interval-valued inputs. Their arithmetic mean is calculated:

$$
\begin{aligned}
A M^{U}\left(A_{1}, A_{2}\right)= & \bigcup_{j_{1}=1}^{2} \bigcup_{j_{2}=1}^{2} A M^{I}\left(\left[a_{1_{j_{1}}}^{-}, a_{1_{j_{1}}}^{+}\right],\left[a_{2_{j_{2}}}^{-}, a_{{j_{2}}_{2}}^{+}\right]\right) \\
= & \left\{A M^{I}([0.6,0.8],[0,0.3])\right\} \cup\left\{A M^{I}([0.6,0.8],[0.7,1])\right\} \cup \\
& \left\{A M^{I}([0.9,1],[0,0.3])\right\} \cup\left\{A M^{I}([0.9,1],[0.7,1])\right\} \\
= & {[0.3,0.55] \cup[0.65,0.9] \cup[0.45,0.65] \cup[0.8,1] } \\
= & \langle[0.3,1]\rangle .
\end{aligned}
$$

When the aggregation of non-convex intervals is intended to model the range of possible values that the output could take with respect to the uncertainty pertaining to the inputs, e.g we want to know the possible average travel times for multiple passengers in Example 1, then this method of aggregation seems appropriate. However we see that although each of the $m_{1} \times m_{2}$ sub-aggregations (before taking the union) could be expressed in terms of an associated penalty function as in Eq. (7), there is no real
sense of some underlying penalty $p\left(A_{i}, Y\right)$ between each of the non-convex inputs and the output $Y=A M^{U}$. In this case, the output is equal to what would be obtained if the convex hulls $[0.6,1]$ and $[0,1]$ were aggregated, while in other cases, a non-convex interval could be obtained, e.g. $A M^{U}(\langle[0.3,0.5],[0.9,1]\rangle,\langle[0.9,1]\rangle)=\langle[0.6,0.75],[0.9,1]\rangle$.

Furthermore, we note that such aggregation is not idempotent, e.g. for $A=\langle[0,0.2],[0.8,1]\rangle, A M^{U}(A, A)=A M^{U}(\langle[0,0.2],[0.8,1]\rangle,\langle[0,0.2],[0.8,1]\rangle)$ $=\langle[0,0.2],[0.4,0.6],[0.8,0.1]\rangle \neq A$. We have the following proposition.

Proposition 1. For an averaging function $f$ defined on the reals and a set of (potentially) non-convex inputs $A_{i}$, we denote by $f^{U}$ the operator defined according to the extension principle, i.e.

$$
f^{U}\left(A_{1}, A_{2}, \ldots, A_{n}\right)=\bigcup_{a_{1} \in A_{1}} \ldots \bigcup_{a_{n} \in A_{n}} f\left(a_{1}, a_{2}, \ldots, a_{n}\right) .
$$

Then, $f^{U}$ will be idempotent if and only if $f$ is internal.
Proof. For sufficiency, we note for all $a \in A$ the value $f(a, a, \ldots, a)=a$ will be present in the output. Furthermore, internality means that no value $a^{\prime} \notin A$ can be obtained from $f\left(a_{1}, a_{2}, \ldots, a_{n}\right)$ when $a_{1}, a_{2}, \ldots, a_{n} \in A$. On the other hand to show necessity, we consider functions of two variables and note that without internality there exists a triplet $a_{1}<a^{\prime}<a_{2}, a_{1}, a_{2} \in A, a^{\prime} \notin A$ such that $f\left(a_{1}, a_{2}\right)=a^{\prime}$ and hence if there is a non-convex interval $A_{i}$ with a gap $] a_{1}, a_{2}\left[\right.$, the aggregate will have $a^{\prime}$ included in the output of $f^{U}$ and hence $f^{U}(A, A) \neq A$.

The framework of penalty functions provides us with an intuitive way of ensuring that operators produce outputs which are faithfully representative of or as "close" as possible to the set of inputs. Penalty functions can often be interpreted in terms of consensus amongst the inputs, i.e. if all the inputs are the same value $A$, then the output should be $A$ and the penalty is zero. If one of the inputs deviates however, we impose some penalty for this and the output will be the value which minimizes the total disagreement between the inputs and the final result. The property of idempotency is hence usually a natural by-product ${ }^{5}$. The question that arises then is how to appropriately

[^5]define a penalty between two non-convex intervals.

## 4. Defining penalties over non-convex intervals

In [13], the authors consider the weighted Frechet mean $F_{d}$ [12], which can be expressed similarly to Eq. (3) with a distance $d$ replacing $p$. For their paper, the distance between a subset $A_{i}$ and an element $y$ is given by

$$
d\left(A_{i}, y\right)=\inf \left\{\left(a_{i}-y\right)^{2} \mid a_{i} \in A_{i}\right\}
$$

A key thing to note here is that $y$ is an element in the metric space (rather than a set or interval). If there are multiple $y$ which minimize the total distance, then the set of all such minimizers can be referred to as the Frechet mean set. For sets of non-convex or convex intervals which overlap, all values $y$ in the intersection will result in $d\left(a_{i}, y\right)=0$ and hence this intersection would be the output $F_{d}$. On the other hand, for non-overlapping intervals, the Frechet mean will either be a single value, e.g. $F_{d}(\langle[0,0.3]\rangle,\langle[0.7,1]\rangle)=0.5$, or set of points, e.g. where one of the inputs is non-convex we could have $F_{d}(\langle[0,0.2],[0.7,1]\rangle,\langle[0.4,0.5]\rangle)=\langle\{0.3\},\{0.6\}\rangle$. Whilst the behavior over intersections ensures that $F_{d}$ will be idempotent ${ }^{6}$, the behavior for nonintersecting inputs means that the output will not always reflect the uncertainty associated with the inputs.

If it is desired that the output take the same form as the inputs, i.e. the aggregation of intervals results in an interval, the aggregation of non-convex intervals generally leads to a non-convex interval and so on, then we require a penalty that models each aspect of the input space. The problem is that the number of distinct sub-intervals may vary from input to input. If the the value of $m_{i}=m$ is fixed for all $A_{i}$, then penalties can be considered separately for each $j=1,2, \ldots, m$ i.e. the penalty associated with two nonconvex intervals is given by

$$
p^{L}\left(A_{i}, Y\right)=\sum_{j=1}^{m}\left(a_{i_{j}}^{-}-y_{j}^{-}\right)^{2}+\left(a_{i_{j}}^{+}-y_{j}^{+}\right)^{2}
$$

[^6]and we have a special case of the aggregation functions defined over a Cartesian product of lattices in [6]. We will denote the extension of the arithmetic mean obtained in this way by $A M^{L}$,

Definition 8 (Arithmetic mean $A M^{L}$ for non-convex intervals). We consider a set of (potentially) non-convex interval inputs $A_{i}=\left\langle\left[a_{i_{1}}^{-}, a_{i_{1}}^{+}\right], \ldots\right.$, $\left.\left[a_{i_{m}}^{-}, a_{i_{m}}^{+}\right]\right\rangle$where $m$ is fixed for all $i$. The arithmetic mean $A M^{L}$ is given by

$$
\begin{equation*}
A M^{L}\left(A_{1}, A_{2}, \ldots, A_{n}\right)=\bigcup_{j=1}^{m} A M^{I}\left(\left[a_{1_{j}}^{-}, a_{1_{j}}^{+}\right], \ldots,\left[a_{n_{j}}^{-}, a_{n_{j}}^{+}\right]\right) \tag{9}
\end{equation*}
$$

where $A M^{I}$ is defined as it is in Eq. (6).
The monotonicity of $A M^{I}$ ensures that the output will be a non-convex interval with $m$ distinct sub-intervals, and this combined with the idempotency of $A M^{I}$ means that $A M^{L}$ can also be considered idempotent, monotone with respect to each $j$ and averaging. The composition of $A M^{L}$ and the penalty by which it can be defined, however imply that there is some relationship between each of the $j$-th sub-intervals. For example, for the two non-convex intervals $A_{1}=\langle[0.3,0.5],[0.7,0.8]\rangle$ and $A_{2}=\langle[0,0.1],[0.3,0.5]\rangle$, it could be possible that the sub-interval $[0.3,0.5]$ of both inputs relates to the same information or uncertain aspect in the data while the additional sub-intervals are related to aspects which are not common to both intervals.

If indeed we have a set of inputs corresponding to each $j$ (which might happen if we want to fuse a number of observations that are both left- and right-censored ${ }^{7}$ ), then it may not be a reasonable restriction that the intervals be non-overlapping. Furthermore, such data may be better handled separately with no need to represent inputs as they are in Eq. (1) or for operators to be defined on the extended space.

In the following section we propose a novel penalty between non-convex inputs which leads to a penalty-based aggregation function with the aim of providing a faithful summary of the inputs, taking the same form of the inputs but also with the ability to handle differing values for $m_{i}$ and satisfying idempotency.

[^7]
## 5. A new operator for non-convex sets based on the closest penalty

For the operator $A M^{U}$, we note that idempotency is lost because for a non-convex input, e.g. $A=\langle[0,0.2],[0.8,1]\rangle$, the operator includes the average of the distinct sub-intervals $A M^{I}([0,0.2],[0.8,1])$ as well as the identical sub-intervals $A M^{I}([0,0.2],[0,0.2])$ and $A M^{I}([0.8,1],[0.8,1])$. In order to avoid this, we need to define a penalty that in this case would only incorporate the penalties between the identical sub-intervals. The operator $A M^{L}$ exhibits this behavior, however it can not handle inputs with a varying number of sub-intervals, so $A M^{L}(\langle[0,0.2],[0.8,1]\rangle,\langle[0,0.2]\rangle)$ is undefined. Furthermore, it will always assume that sub-intervals across inputs are related based on the value of $j$, however this may not necessarily be the case.

Here we consider a penalty function which for each convex sub-interval $\left[a_{i_{j}}^{-}, a_{i_{j}}^{+}\right]$, only counts the penalty to the closest sub-interval in $Y$ and vice versa. The essential idea is that we wish to minimize the extent to which two inputs disagree, accounting for the possibility that separate intervals may refer to different aspects of the evaluations.

We let the penalty between any two non-convex intervals $A_{i}$ and $Y$ be defined as,

$$
\begin{align*}
p^{N C}\left(A_{i}, Y\right)= & \sum_{j_{i}=1}^{m_{i}} \min _{j=1, \ldots, m}\left(\left(a_{i_{j_{i}}}^{-}-y_{j}^{-}\right)^{2}+\left(a_{i_{j_{i}}}^{+}-y_{j}^{+}\right)^{2}\right) \\
& +\sum_{j=1}^{m} \min _{j_{i}=1, \ldots, m_{i}}\left(\left(a_{i_{j_{i}}}^{-}-y_{j}^{-}\right)^{2}+\left(a_{i_{j_{i}}}^{+}-y_{j}^{+}\right)^{2}\right) . \tag{10}
\end{align*}
$$

The following proposition allows us to use $p^{N C}$ in order to define idempotent operators for non-convex inputs.

Proposition 2. The penalty $p^{N C}$ defined between non-convex intervals satisfies the following properties:
i) $p^{N C}(A, Y)=0$ if $A=Y$;
ii) $p^{N C}(A, Y)>0$ if $A \neq Y$.

Proof. For each sub-interval in $A$, a partial penalty is included for the closest sub-interval in $Y$ while for each sub-interval in $Y$, a penalty
is calculated to the closest sub-interval of $A$. If $A=Y$, then only the penalties between pairs of identical sub-intervals will be included and the overall penalty will be zero. On the other hand, if $A \neq Y$ then there will either be a sub interval $\left[a_{i_{j_{i}}}^{-}, a_{i_{j_{i}}}^{+}\right]$or a sub-interval $\left[y_{j}^{-}, y_{j}^{+}\right]$for which the minimum penalty is not zero.

We therefore have the first two corresponding properties from Definition 3. In the case of non-convex intervals, we can relax the third requirement, since it is not necessary for the minimizer to be a convex interval or single value.

Remark 5. Although in this case we have used squared differences so that the resulting penalty-based operator will exhibit behavior consistent with the arithmetic mean, clearly other choices are possible if we wish to extend functions such as the median.

From this penalty, we can instantiate the following extension of the arithmetic mean, which will be idempotent for non-convex intervals and coincides with the arithmetic means defined for intervals or real-valued inputs if the inputs are all convex intervals or real values respectively.

Definition 9 (Arithmetic mean $A M^{N C}$ for non-convex intervals). For $a$ set of (potentially) non-convex inputs $A_{i}=\left\langle\left[a_{i_{1}}^{-}, a_{i_{1}}^{+}\right], \ldots,\left[a_{i_{m_{i}}}^{-}, a_{i_{m_{i}}}^{+}\right]\right\rangle$, the arithmetic mean $A M^{N C}$ is given by,

$$
\begin{equation*}
A M^{N C}\left(A_{1}, A_{2}, \ldots, A_{n}\right)=\arg \min _{Y} \sum_{i=1}^{n} p^{N C}\left(A_{i}, Y\right) \tag{11}
\end{equation*}
$$

Remark 6. Since small changes in the input may mean the penalty $p^{N C}$ can incorporate differing compositions of sub-intervals (see Example 5 and some of the figures below), and in particular since the value of $m$ can vary, there will be some sets of inputs for which $Y$ may not be unique. In such cases, a selection criterion can be adopted in order to make the function well defined, such as one which minimizes the number of sub-intervals or which favors outputs closer to the lower sub-intervals than the higher ones.

To simplify our notation, from here on we will denote the $j$-th sub-interval of a non-convex input $A_{i}$ by $\left[\bar{a}_{i_{j_{i}}}\right]$ while for the output $Y$ we will use $\left[\bar{y}_{j}\right]$.

We will denote the partial penalty between the sub-intervals calculated from squared differences by $p^{I}\left(\left[\bar{a}_{i_{j}}\right],\left[\bar{y}_{j}\right]\right)$ and so we have

$$
p^{N C}\left(A_{i}, Y\right)=\sum_{j_{i}=1}^{m_{i}} \min _{j=1, \ldots, m} p^{I}\left(\left[\bar{a}_{i_{j}}\right],\left[\bar{y}_{j}\right]\right)+\sum_{j=1}^{m} \min _{j_{i}=1, \ldots, m_{i}} p^{I}\left(\left[\bar{a}_{i_{j}}\right],\left[\bar{y}_{j}\right]\right) .
$$

Figure 1 helps illustrate how $A M^{N C}$ is calculated and the associated penalties that are incorporated.


Figure 1: Shows the output $A M^{N C}$ for two non-convex inputs. The penalties used denoting the closest input sub-intervals to the sub-intervals of $Y$ are indicated by the black dotted lines, while the closest $\left[\bar{y}_{j}\right]$ to each of the sub-intervals in $A_{1}$ and $A_{2}$ are indicated by the dotted grey lines.

In Figure 1(a), the inputs are seen to each comprise two separate evaluations and so the output $Y=\left(\left[\bar{y}_{1}\right],\left[\bar{y}_{2}\right]\right)$ uses the penalties in both directions (grey and black lines) between the first intervals $\left[\bar{a}_{1_{1}}\right],\left[\bar{a}_{2_{1}}\right]$ and second intervals $\left[\bar{a}_{1_{2}}\right],\left[\bar{a}_{2_{2}}\right]$ respectively. So for $\left[\bar{y}_{1}\right]$ we will have,

$$
\begin{aligned}
{\left[\bar{y}_{1}\right] } & =\arg \min _{[\bar{y}]} \sum_{i=1}^{2}\left(p^{I}\left(\left[\bar{a}_{i_{1}}\right],[\bar{y}]\right)+p^{I}\left([\bar{y}],\left[\bar{a}_{i_{1}}\right]\right)\right) \\
& =\arg \min _{[\bar{y}]} \sum_{i=1}^{2} 2 \cdot p^{I}\left(\left[\bar{a}_{i_{1}}\right],[\bar{y}]\right) \\
& =\arg \min _{[\bar{y}]} \sum_{i=1}^{2} p^{I}\left(\left[\bar{a}_{i_{1}}\right],[\bar{y}]\right),
\end{aligned}
$$

and similarly for $\left[\bar{y}_{2}\right]$. The sub-intervals of $Y$ will hence be minimized respectively by the arithmetic means $A M^{I}\left(\left[\bar{a}_{1_{1}}\right],\left[\bar{a}_{2_{1}}\right]\right)$ and $A M^{I}\left(\left[\bar{a}_{1_{2}}\right],\left[\bar{a}_{2_{2}}\right]\right)$, so $A M^{N C}$ coincides with $A M^{L}$ for these inputs.

On the other hand, in Figure 1(b), we see that both sub-intervals of $A_{2}$ could be seen to form a group with the first sub-interval of $A_{1}$. The penalties included between $A_{1}$ and $Y$ are the same as they are for Figure 1(a), however for the penalties used between $A_{2}$ and $Y$, we see that $\left[\bar{y}_{1}\right]$ is the closest subinterval in $Y$ for both sub-intervals in $A_{2}$ (represented by dotted grey lines), while the closest sub-interval of $A_{2}$ to $\left[\bar{y}_{1}\right]$ is $\left[\bar{a}_{2_{1}}\right]$ and the closest to $\left[\bar{y}_{2}\right]$ is $\left[\bar{a}_{2_{2}}\right]$ (dotted black lines). So in this case, for $\left[\bar{y}_{1}\right]$ we have

$$
\begin{aligned}
{\left[\bar{y}_{1}\right] } & =\arg \min _{[\bar{y}]} \sum_{i=1}^{2}\left(\left(\sum_{\substack{j_{1}=1 \\
j_{2}=1,2}} p^{I}\left(\left[\bar{a}_{i_{j_{i}}}\right],[\bar{y}]\right)\right)+p^{I}\left([\bar{y}],\left[\bar{a}_{i_{1}}\right]\right)\right) \\
& =2 \cdot p^{I}\left(\left[\bar{a}_{1_{1}}\right],[\bar{y}]\right)+2 \cdot p^{I}\left(\left[\bar{a}_{2_{1}}\right],[\bar{y}]\right)+p^{I}\left(\left[\bar{a}_{2_{2}}\right],[\bar{y}]\right) .
\end{aligned}
$$

The sub-interval $\left[\bar{y}_{1}\right]$ incorporates 5 partial penalties in total - in both directions from $\left[\bar{a}_{1_{1}}\right]$ and $\left[\bar{a}_{2_{1}}\right]$ and from just one direction from $\left[\bar{a}_{2_{2}}\right]$ It will hence be minimized by a weighted arithmetic mean of these sub-intervals, with $\left[\bar{a}_{1_{1}}\right],\left[\bar{a}_{2_{1}}\right]$ having twice the weight of $\left[\bar{a}_{2_{2}}\right]$, i.e.

$$
\left[\bar{y}_{1}\right]=\frac{2}{5}\left[\bar{a}_{1_{1}}\right]+\frac{2}{5}\left[\bar{a}_{2_{1}}\right]+\frac{1}{5}\left[\bar{a}_{2_{2}}\right] .
$$

Rather than solving the penalty expression in Eq. (11) using optimization, we therefore note that once the set of closest inputs to each $\left[\bar{y}_{j}\right]$ is given, the analytic solution will be a weighted mean of these inputs. Calculation hence requires taking all possible subset combinations and determining
which of these minimizes the penalty ${ }^{8}$. For each $\left[\bar{y}_{j}\right]$, we require the closest sub-interval from each of the inputs (corresponding with the black dotted penalties indicated in Figure 1), and then all input sub-intervals are partitioned between the $\left[\bar{y}_{j}\right]$. For example, in Figure 1 (a) the partition is

$$
\left\{\left[\bar{a}_{1_{1}}\right],\left[\bar{a}_{2_{1}}\right]\right\} /\left\{\left[\bar{a}_{1_{2}}\right],\left[\bar{a}_{2_{2}}\right]\right\},
$$

while in Figure 1 (b) the partition is

$$
\left\{\left[\bar{a}_{1_{1}}\right],\left[\bar{a}_{2_{1}}\right],\left[\bar{a}_{2_{2}}\right]\right\} /\left\{\left[\bar{a}_{1_{2}}\right]\right\} .
$$

The computational cost of finding all of these partitions can be reduced by limiting ourselves to feasible options, i.e. if an evaluation included 3 subintervals, we couldn't partition the first and third sub-intervals to $\left[\bar{y}_{1}\right]$ and the second sub-interval to $\left[\bar{y}_{2}\right]$.

We note that in some cases, the number of distinct sub-intervals in $Y$ could differ to that of the inputs, even where $m_{i}$ is the same for all $A_{i}$. For example, consider the inputs $A_{1}=\langle[0.4,0.6],[0.8,0.9]\rangle$ and $A_{2}=$ $\langle[0.1,0.2],[0.3,0.4]\rangle$. In this case, the total penalty using $p^{N C}$ if $Y$ is a single interval is 0.5567 , with the resulting output $Y=\langle[0.3833,0.5167]\rangle$, while the penalty when $Y$ is comprised of two intervals is 1.0175 and the output is $Y=\langle[0.275,0.4],[0.475,0.625]\rangle$ (Visual representations of these inputs and the penalties used are shown in Figure 2). Even if we account for the higher number of penalties considered, a single interval is still better at minimizing the given penalty. On the other hand, in cases such as that shown previously in Figure 1 (b), use of a single interval results in a larger penalty (even though only 6 penalties are used rather than 8 ).

Another example is shown in Figure 3. Here, $A_{1}=\langle[0.2,0.3],[0.8,0.9]\rangle$ is comprised of 2 sub-intervals while $A_{2}=\langle[0.1,0.25]\rangle$ is a single interval. When $m=1$ for the output $Y$, we have a minimizer $Y=\langle[0.28,0.4]\rangle$ and a total penalty of 0.663 . For $m=2$, the output is $Y=\langle[0.15,0.275],[0.5667,0.6833]\rangle$ and the total penalty is 0.6208 . Again, even though more penalties are taken into account for $m=2$, the overall penalty is lower.

Since the calculation of each of the $\left[\bar{y}_{j}\right]$ depends on the closest inputs, $A M^{N C}$ will not be monotone in general. Consider the following example.

[^8]

Figure 2: Shows an example of the difference between the penalties used for a single interval as opposed to 2 distinct intervals. In this case, the overall penalty is minimized by using a single interval.

Example 5. Suppose we have the inputs $A_{1}=\langle[0.1,0.3],[0.4,0.5]\rangle$ and $A_{2}=$ $\langle[0.2,0.4],[0.7,0.9]\rangle$. This gives an output

$$
Y=A M^{N C}\left(A_{1}, A_{2}\right)=\langle[0.2,0.38],[0.6,0.7667]\rangle
$$

with $\left[\bar{y}_{1}\right]$ taking the average $\frac{2}{5}\left[\bar{a}_{1_{1}}\right]+\frac{2}{5}\left[\bar{a}_{2_{1}}\right]+\frac{1}{5}\left[\bar{a}_{1_{2}}\right]$. If $\left[\bar{a}_{1_{2}}\right]$ is increased from $[0.4,0.5]$ to $[0.5,0.6]$, the penalty is no longer minimized with the current partition, and $\left[\bar{y}_{1}\right]$ will now only take the average of $\left[\bar{a}_{1_{1}}\right]$ and $\left[\bar{a}_{2_{1}}\right]$ (in both directions). As a result, the interval $\left[\bar{y}_{1}\right]$ will decrease. Furthermore, $\left[\bar{y}_{2}\right]$ will also decrease since it now considers the penalty $p^{I}\left(\left[\bar{y}_{2}\right],\left[\bar{a}_{1_{2}}\right]\right)$ in both directions. The resulting output is $Y^{\prime}=\langle[0.15,0.35],[0.6,0.75]\rangle$ which intuitively should be ordered less than $Y$.

Ensuring monotonicity in this example would require us to fix the intervals used in the calculation of each $\left[\bar{y}_{j}\right]$ as is the case for $A M^{L}$. If the number of sub-intervals of the inputs differs, then there is no consistent way of achieving this. Fixing each $\left[\bar{y}_{j}\right]$ to be obtained with respect to corresponding intervals of the inputs also infers a priori knowledge on relationships that exist between each of the inputs and their sub-intervals. In some of these cases, it may make more sense to treat each of these aspects separately and aggregate accordingly, or alternatively apply pre-processing to the data in


Figure 3: Shows an example of the difference between the penalties used for a single interval as opposed to 2 distinct intervals. In this case, the overall penalty is minimized by using a single interval.
order to ensure the number of sub-intervals are fixed and relate to the same aspect of the evaluations.

## 6. Numeric examples

The use of each of the operators for non-convex interval inputs $A M^{N C}, A M^{L}$ and $A M^{U}$ all make assumptions on the type of data with properties that should be taken into account depending on the application. For example, $A M^{U}$ may not be useful for providing a representative value from expert opinions because it is not idempotent. On the other hand, whilst $A M^{L}$ is idempotent, it is only defined where the number of distinct subsets is fixed for all of the inputs. Finally, while $A M^{N C}$ is idempotent and makes no assumptions about the relationship between each of the sub-intervals, it is not monotone and is not appropriate if we are interested in the expectation or range of possible values from uncertain inputs in a probabilistic sense.

A number of illustrative input sets, each with 4 inputs comprised of 2 distinct sub-intervals are shown in Figure 4 with their values given in Tables 1-2.

The set depicted in Figure 4(a) helps illustrate the idempotency of $A M^{N C}$ and $A M^{L}$. We note that for $A M^{U}$ here, the output is 5 sub-intervals spread between the two sub-intervals common to all inputs.


Figure 4: Outputs for $A M^{N C}, A M^{L}, A M^{U}$ for example input sets of 4 non-convex subsets of $[0,1]$. Example (a) illustrates the idempotency of $A M^{N C}$ and $A M^{L}$. In examples (b)-(d), the output of $A M^{U}$ is an interval coinciding with the upper and lower bounds of $A M^{L}$.

Table 1: Aggregation of subsets with two sub-intervals

| Example | Figure 4 (a) | Figure 4 (b) |
| :--- | :---: | :---: |
| $A_{1}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.5,0.8]\rangle$ |
| $A_{2}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.7,0.9]\rangle$ |
| $A_{3}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ |
| $A_{4}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.5,0.7]\rangle$ |
|  |  |  |
| $A M^{N C}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.6,0.8]\rangle$ |
| $A M^{L}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.6,0.8]\rangle$ |
|  |  |  |
| $A M^{U}$ | $\langle[0.2,0.3],[0.325,0.425]$, | $\langle[0.2,0.8]\rangle$ |
|  | $[0.45,0.55],[0.575,0.675]$, |  |
|  | $[0.7,0.8]\rangle$ |  |

For the set in Figure $4(\mathrm{~b}), A M^{N C}$ and $A M^{L}$ coincide once again as the inputs form two distinct groups, which are hence aggregated separately by these operators. $A M^{U}$ in this case is a single convex interval which encompasses these evaluations.

The sets in Figures 4(c) and 4(d), also show inputs that could be partitioned into two distinct groups, however in the case of some inputs, both sub-intervals are aligned with a single group. In these cases, the upper interval of $A M^{L}$ is brought down by these evaluations to a greater degree than the upper interval of $A M^{N C}$.

For these particular examples, there are only slight differences between the aggregated results for $A M^{L}$ and $A M^{N C}$. It should be noted that while the calculation of $A M^{U}$ or $A M^{N C}$ are both suitable for inputs with a different number of distinct sub-intervals, they require much more computation time.

Remark 7. We can also mention the outputs that would result if using the Frechet mean and the distance d defined between non-convex intervals and real values. For all input sets, there is at least one point of intersection, so we obtain the outputs: (a) $\langle[0.2,0.3],[0.7,0.8]\rangle$; (b) $\langle[0.2,0.3],\{0.7\}\rangle$; (c)

Table 2: Aggregation of subsets with two sub-intervals

| Example | Figure 4 (c) | Figure 4 (d) |
| :--- | :---: | :---: |
| $A_{1}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.1,0.5],[0.7,0.8]\rangle$ |
| $A_{2}$ | $\langle[0.2,0.3],[0.7,0.8]\rangle$ | $\langle[0.2,0.3],[0.75,0.85]\rangle$ |
| $A_{3}$ | $\langle[0.1,0.2],[0.3,0.4]\rangle$ | $\langle[0.1,0.2],[0.3,0.4]\rangle$ |
| $A_{4}$ | $\langle[0.1,0.2],[0.3,0.4]\rangle$ | $\langle[0.3,0.5],[0.8,0.9]\rangle$ |
| $A M^{N C}$ | $\langle[0.18,0.28],[0.5667,0.6667]\rangle$ | $\langle[0.2111,0.4],[0.6857,0.7857]\rangle$ |
| $A M^{L}$ | $\langle[0.15,0.25],[0.5,0.6]\rangle$ | $\langle[0.175,0.375],[0.6375,0.7375]\rangle$ |
| $A M^{U}$ | $\langle[0.15,0.6]\rangle$ | $\langle[0.175,0.7375]\rangle$ |

$\langle\{0.2\},\{0.3\}\rangle$; and (d) 0.3 .
We will now briefly mention some other potential approaches for addressing this issue before concluding the paper.

## 7. Alternative aggregations and future research

The extensions of the arithmetic mean $A M^{N C}, A M^{L}$ and $A M^{U}$ all coincide with the arithmetic mean $A M^{I}$ for interval valued inputs (and in turn, real valued inputs). The following approaches could also be developed.

### 7.1. Clustering

It is important to note that for each of the given arithmetic mean extensions, separate intervals of the same input are treated as related to one another in the aggregation process. However, the problem can also be approached by treating these sub-intervals as independent evaluations. For example, we can take the average of all inputs,

$$
Y=\frac{1}{\sum_{i=1}^{n} m_{i}} \sum_{i=1}^{n} \sum_{j_{i}=1}^{m_{i}}\left[\bar{a}_{i_{j_{i}}}\right]
$$

Further, we can partition the sub-intervals as we do for $A M^{N C}$ and define $Y$ such that it is composed of $m$ sub-intervals, each calculated as the arithmetic mean of one of the partitions. This can be framed as an $m$-means clustering problem. We start with a randomly assigned non-convex subset $Y$ with $m$ distinct intervals and then calculate the closest intervals to each of the $j$ sub-intervals. We then update $Y$ such that each of the $\left[\bar{y}_{j}\right]$ are calculated from these intervals (leaving $\left[\bar{y}_{j}\right]$ unchanged if there are no intervals in its cluster. For the input sets from Figure 4 (a) and (b), the partition of the two clusters will be as it is for $A M^{N C}$ and $A M^{L}$ and the outputs will be the same. For the sets in Figure 4(c) and (d), the two clusters will again match the partitioning used by $A M^{N C}$, however now we calculate the $\left[\bar{y}_{1}\right]$ and $\left[\bar{y}_{2}\right]$ solely from the separated clusters. So in (c), $\left[\bar{y}_{2}\right]$ is calculated as the mean of $\left[\bar{a}_{1_{2}}\right]$ and $\left[\bar{a}_{2_{2}}\right]$ and in (d), $\left[\bar{y}_{2}\right]$ is the average of $\left[\bar{a}_{1_{2}}\right],\left[\bar{a}_{2_{2}}\right]$ and $\left[\bar{a}_{4_{2}}\right]$. In general, this clustering approach has the potential to get caught in a local optimum, however with 10 random initializations and 10 iterations, the outputs for the input sets from Figure 4 (c) and (d) are determined to be $\langle[0.2,0.3],[0.7,0.8]\rangle$ and $\langle[0.2,0.38],[0.75,0.85]\rangle$ respectively (which differs to $A M^{N C}$ and $A M^{L}$ ). However, this approach is not a true generalization of the arithmetic mean since for standard real-values, the output will not coincide with the standard arithmetic mean. For example, providing two real inputs $a_{1}$ and $a_{2}$ could result in the output comprised of the non-convex set $\left\langle\left[\bar{a}_{1}\right],\left[\bar{a}_{2}\right]\right\rangle$.

### 7.2. Internal functions

To restore idempotency but also ensure that the output is as close as possible to all sub-intervals of the inputs, we could also define $Y$ to be the $A_{i}$ (taken from the set of inputs) that minimizes the overall penalty, i.e.

$$
\begin{equation*}
f^{I N T}\left(A_{1}, A_{2}, \ldots, A_{n}\right)=\arg \min _{Y=A_{1}, A_{2}, \ldots, A_{n}} \sum_{i=1}^{n} \sum_{j_{i}=1}^{m_{i}} \sum_{j=1}^{m} p\left(\left[\bar{a}_{i_{j_{i}}}\right],\left[\bar{y}_{j}\right]\right) . \tag{12}
\end{equation*}
$$

We hence choose the input which disagrees least with the remaining inputs as our representative. If we use the squared differences for our penalty, this method could still be seen as an extension of the arithmetic mean, however we note once again that it will not be a true generalization since the standard arithmetic mean is not internal for real inputs.

## 8. Conclusion

Here we have considered existing extensions of averaging aggregation functions to inputs expressed as non-convex intervals and proposed an alternative aggregation operator by using a novel penalty function. We compared this operator to existing approaches with some numerical examples. The operators coincide if the inputs are standard intervals or real-valued inputs, however the new operator $A M^{N C}$ is idempotent for non-convex interval valued inputs and is able to handle a differing number of sub-intervals. It provides a useful way to aggregate such inputs into an output defined over the same space that is representative, giving some indication of what is "typical".

We noted that whereas for interval inputs and real inputs the properties of averaging behavior and idempotency are equivalent, the difficulty in defining monotonicity for inputs composed of different numbers of distinct intervals has implications on the relationship between these properties and their usefulness for practical contexts would need to be considered separately.

Key to the selection of which operator is most suitable for a given context is an understanding of the purpose served by the aggregation operator whether it is providing a score for comparison, giving an idea of possible outputs relative to uncertainty around the inputs, or summarizing the data with a representative value.

## Acknowledgement

This work was supported by the Deakin University Central Research Grants Scheme, project number RM25299.

## References

[1] K. Atanassov. Intuitionistic fuzzy sets. Fuzzy Sets and Systems, 80:8796, 1986.
[2] K. Atanassov and G. Gargov. Interval valued intuitionistic fuzzy sets. Fuzzy Sets and Systems, 31(3):343-349, 1989.
[3] B. Bedregal, G. Beliakov, H. Bustince, T. Calvo, R. Mesiar, and D. Paternain. A class of fuzzy multisets with a fixed number of memberships. Information Sciences, 189:1-17, 2012.
[4] G. Beliakov and S. James. Defining Bonferroni means over lattices. In IEEE International Conference on Fuzzy Systems, pages 1-8, Brisbane, Australia, 2012.
[5] G. Beliakov, A. Pradera, and T. Calvo. Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg, 2007.
[6] H. Bustince, E. Barrenechea, T. Calvo, S. James, and G. Beliakov. Consensus in multi-expert decision making problems using penalty functions defined over a Cartesian product of lattices. Information Fusion, 17:5664, 2014.
[7] T. Calvo and G. Beliakov. Aggregation functions based on penalties. Fuzzy Sets and Systems, 161:1420-1436, 2010.
[8] T. Calvo, R. Mesiar, and R. R. Yager. Quantitative weights and aggregation. IEEE Transactions on Fuzzy Systems, 12(1):62-69, 2004.
[9] J.-L. Deng. Control problems of grey systems. Systems $\& 3$ Control Letters, 1(5):288-294, 1982.
[10] D. Dubois and H. Prade. Operations on fuzzy numbers. International Journal of Systems Science, 9(6):613-626, 1978.
[11] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg. Experimental Uncertainty Estimation and Statistics for Data having Interval Uncertainty. Sandia National Laboratory Technical Report SAND2007-0939, Setauket, New York, 2007.
[12] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. Annales de L'Institut Henri Poincaré, 10(4):215-310, 1948.
[13] C. E. Ginestet, A. Simmons, and E. D. Kolaczyk. Weighted Frechet means as convex combinations in metric spaces: Properties and generalized median inequalities. Statistics and Probability Letters, 82:18591863, 2012.
[14] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. Aggregation Functions. Cambridge University press, Cambridge, 2009.
[15] Z. Guo, T. M. Gill, and H. G. Allore. Modeling repeated time-to-event health conditions with discontinuous risk intervals. an example of a longitudinal study of functional disability among older persons. Methods of Information in Medicine, 47(2):107-116, 2008.
[16] G. Ligozat. Qualitative Spatial and Temporal Reasoning. John Wiley \& Sons, Hoboken, NJ, USA, 2013.
[17] Y.-H. Lin, P.-C. Lee, and H.-I. Ting. Dynamic multi-attribute decision making model with grey number evaluations. Expert Systems with Applications, 35:1638-1644, 2008.
[18] S. Miyamoto. Multisets and fuzzy multisets. In Liu Z.-Q. and S Miyamoto, editors, Soft computing and human-centered machines, pages 9 - 33. Springer, 2000 .
[19] S. Miyamoto. Remarks on basics of fuzzy sets and fuzzy multisets. Fuzzy Sets and Systems, 156:427-431, 2005.
[20] I. Molchanov. Theory of random sets. Springer, 2005.
[21] D. G. Nimmo, L. T. Kelly, L. M. Spence-Bailey, S. J. Watson, A. Haslem, J. G. White, M. F. Clarke, and A. F. Bennett. Predicting the centurelong post-fire responses of reptiles. Global Ecology and Biogeography, 21:1062-1073, 2012.
[22] V. Torra. Hesitant fuzzy sets. International Journal of Intelligent Systems, 25:529 - 539, 2010.
[23] V. Torra and Y. Narukawa. Modeling Decisions. Information Fusion and Aggregation Operators. Springer, Berlin, Heidelberg, 2007.
[24] V. Torra and Y. Narukawa. On hesitant fuzzy sets and decision. In The 18th IEEE International Conference on Fuzzy Systems, pages 1378 - 1382, Jeju Island, Korea, 2009.
[25] R. Verde and A. Irpino. A new interval data distance based on the wasserstein metric. In Data Analysis, Machine Learning and Applications, pages $705-712$. Springer Berlin Heidelberg, 2008.
[26] C. Wagner, D. T. Anderson, and T. C. Havens. Generalization of the fuzzy integral for discontinuous interval- and non-convex interval fuzzy set-valued inputs. In IEEE International Conference on Fuzzy Systems, pages 1-8, 2013.
[27] R. R. Yager. On the theory of bags. International Journal of General Systems, 13:23-37, 1986.
[28] R. R. Yager and A. Rybalov. Understanding the median as a fusion operator. International Journal of General Systems, 26(3):239-263, 1997.
[29] Y. Yang and R. John. Grey sets and greyness. Information Sciences, 185:249-264, 2012.
[30] Y. Yang, S.-F. Liu, and R. John. Uncertainty representation of grey numbers and grey sets. IEEE Transactions on Cybernetics, under review, 2013.


[^0]:    ${ }^{1}$ Corresponding author. Tel.: +61 392517481.
    E-mail addresses: gleb@deakin.edu.au.au (G. Beliakov), sjames@deakin.edu.au (S. James).

[^1]:    ${ }^{1}$ Corresponding author. Tel.: +61 392517481.
    E-mail addresses: gleb@deakin.edu.au.au (G. Beliakov), sjames@deakin.edu.au (S. James).

[^2]:    ${ }^{2}$ The third property, which ensures the function is well defined, has been alternatively expressed as requiring quasi-convexity [6] or that $P(\mathbf{x}, y) \geq P(\mathbf{z}, y)$ whenever $x_{i} \geq z_{i} \geq y$ or $x_{i} \leq z_{i} \leq y \forall i$, i.e. a kind of monotonicity with respect to each argument [8].

[^3]:    ${ }^{3}$ The intervals can be open or closed.

[^4]:    ${ }^{4}$ It was shown in [4] that for lattices, idempotency and averaging behavior are equivalent for monotone functions (analogously to real inputs).

[^5]:    ${ }^{5}$ This is provided that the penalty between any input and the output is strictly greater than zero if they are note the same.

[^6]:    ${ }^{6}$ We note that with this choice of $d$, we can have $d(A, y)=0$ even though technically $A \neq y$, however the union of all such values will result in a Frechet mean set $Y=A$.

[^7]:    ${ }^{7}$ That is, an observation lies either below or above a given observable interval range, usually due to the limits of the way in which it is measured.

[^8]:    ${ }^{8}$ An implementation of the function in Eq. (11), as well as the other functions considered here can be found at http://aggregationfunctions.wordpress.com .

