An undesirable behaviour of a recent extension of OWA operators to the setting of multidimensional data

Ral Pérez-Fernández[1,2,⋆]

[1] Department of Statistics and O.R. and Mathematics Didactics, University of Oviedo, Spain, perezfernandez@uniovi.es
[2] KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium

Abstract. OWA operators have been ubiquitous in many disciplines since they were introduced by Yager in 1988. Aside of some other intuitive properties (e.g. monotonicity and idempotence), OWA operators are known to be continuous and, for some carefully constructed weighing vectors, very robust in the presence of outliers. In a recent paper, a natural extension of OWA operators to the setting of multidimensional data has been proposed based on the use of a linear extension of the product order by means of several weighted arithmetic means. Unfortunately, OWA operators constructed in such a way focus too strongly on the level sets of one of the weighted arithmetic means. It is here shown that this focus ultimately results in a forfeit of the properties of continuity and robustness in the presence of outliers.

Keywords: OWA operator · Multidimensional data · Linear extension · Weighted arithmetic mean.

1 Introduction

Back in 1988, Yager introduced OWA operators in the context of decision making [12] as a family of functions that lie in between the 'and' and the 'or' operators. Formally, an OWA operator simply is the result of applying a symmetrization process to a weighted arithmetic mean [7] in which the weighted arithmetic mean is applied to the order statistics of the values to be aggregated rather than to the values themselves. Aside of symmetry, OWA operators satisfy very natural properties such as (increasing) monotonicity, idempotence and continuity.

Several families of OWA operators have been studied in the literature [13], probably centered OWA operators being the most prominent family [14]. Interestingly, some centered OWA operators have been studied in the field of statistics due to their robustness in the presence of outliers, e.g., the median, trimmed means and winsorized means.

⋆ This research has been partially supported by the Spanish MINECO project (TIN2017-87600-P) and the Research Foundation of Flanders (FWO17/PDO/160).

The field of multivariate statistics has studied for a long time how to extend the notion of median to the multidimensional setting [10]. The field of aggregation theory is also experimenting an increasing interest in this multidimensional setting. For instance, one can find recent works on penalty-based aggregation of multidimensional data [5] and on the property of monotonicity for multidimensional functions [6, 9]. It is no surprise then that an extension of OWA operators to the multidimensional setting has been recently proposed by De Miguel et al. [2] by making use of a linear extension of the product order. Unfortunately, as we shall see in the upcoming sections, the consideration of a linear extension of the product order extends OWA operators to the multidimensional setting at the cost of losing continuity and, if applicable, the robustness in the presence of outliers.

The remainder of the paper is structured as follows. In Section 2, it is discussed how a linear extension of the product order can be defined by means of several weighted arithmetic means. Section 3 presents the extension of OWA operators to the multidimensional setting by De Miguel et al. The strange behaviour of such extension is discussed in Section 4. We end with some concluding remarks in Section 5.

2 Linear extensions of the product order by means of weighted arithmetic means

Consider n points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$. The j-th component of the point $\mathbf{x}_i$ is denoted by $\mathbf{x}_i(j)$. The product order $\leq_m$ on $\mathbb{R}^m$ is defined as $\mathbf{x}_{i_1} \leq_m \mathbf{x}_{i_2}$ if $\mathbf{x}_{i_1}(j) \leq \mathbf{x}_{i_2}(j)$ for any $j \in \{1, \ldots, m\}$. Obviously, $\leq_m$ is not a linear order on $\mathbb{R}^m$.

As discussed in [2] (see Proposition 2), a linear extension of $\leq_m$ can be defined by means of m linearly independent weighted arithmetic means $M_1, \ldots, M_m : \mathbb{R}^m \to \mathbb{R}$. More precisely, the linear extension $\preceq_{\mathbf{M}}$ of $\leq_m$ based on $\mathbf{M} = (M_1, \ldots, M_m)$ is defined by $\mathbf{x}_{i_1} \preceq_{\mathbf{M}} \mathbf{x}_{i_2}$ if $\mathbf{x}_{i_1} = \mathbf{x}_{i_2}$ or there exists $k \in \{1, \ldots, m\}$ such that

$$\begin{aligned} &M_j(\mathbf{x}_{i_1}) \leq M_j(\mathbf{x}_{i_2}) \,, \text{ for any } j \in \{1, \ldots, k-1\} \,, \\ &M_k(\mathbf{x}_{i_1}) < M_k(\mathbf{x}_{i_2}) \,. \end{aligned}$$

The most prominent such linear extensions of $\mathbb{R}^m$ are the lexicographic orders $\preceq_\sigma$ [4], where a permutation σ of $\{1, \ldots, n\}$ serves for establishing a sequential order in which the different components are considered. Formally, for any $j \in \{1, \ldots, k\}$, M_j is defined as $M_j(\mathbf{x}_i) = \mathbf{x}_i(\sigma(j))$.

In the two-dimensional case, Xu and Yager's linear order $\preceq_{\mathrm{XY}}$ on $\mathbb{R}^2$ [11] (induced by $M_1(\mathbf{x}_i) = \frac{1}{2}\mathbf{x}_i(1) + \frac{1}{2}\mathbf{x}_i(2)$ and $M_2(\mathbf{x}_i) = \mathbf{x}_i(2)$) is also very well-known[3] in the context of intervals and intuitionistic fuzzy sets.

[3] It is admittedly more common to find an equivalent definition of the order in which M_2 is defined as $M_2(\mathbf{x}_i) = \mathbf{x}_i(2) - \mathbf{x}_i(1)$. This equivalent definition is here abandoned in order to guarantee M_2 to be monotone increasing.

Example 1. Consider $\mathbf{x}_1 = (3,1)$, $\mathbf{x}_2 = (1,3)$ and $\mathbf{x}_3 = (3,3)$. It obviously holds that $\mathbf{x}_1 \leq_2 \mathbf{x}_3$ and $\mathbf{x}_2 \leq_2 \mathbf{x}_3$, however, $\mathbf{x}_1$ and $\mathbf{x}_2$ are not comparable with respect to $\leq_2$.

If one considers Xu and Yager's linear order $\preceq_{\mathrm{XY}}$ on $\mathbb{R}^2$, it holds that $\mathbf{x}_1 \preceq_{\mathrm{XY}} \mathbf{x}_2$ due to the fact that

$$M_1(\mathbf{x}_1) = \frac{1}{2}\mathbf{x}_1(1) + \frac{1}{2}\mathbf{x}_1(2) = 2 \leq 2 = \frac{1}{2}\mathbf{x}_2(1) + \frac{1}{2}\mathbf{x}_2(2) = M_1(\mathbf{x}_2)\,,$$
$$M_2(\mathbf{x}_1) = \mathbf{x}_1(2) = 1 < 3 = \mathbf{x}_2(2) = M_2(\mathbf{x}_2)\,.$$

It is concluded that $\mathbf{x}_1 \preceq_{\mathrm{XY}} \mathbf{x}_2 \preceq_{\mathrm{XY}} \mathbf{x}_3$.

An illustration of this procedure is given in Figure 1.

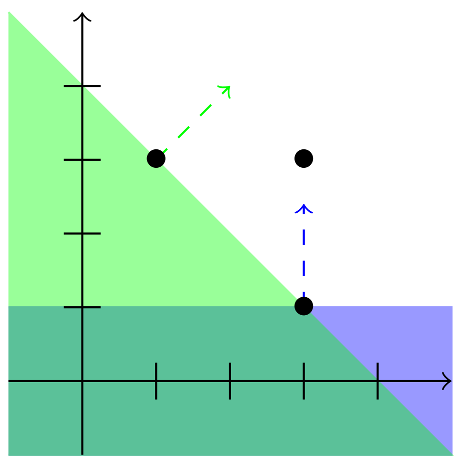

Fig. 1. Graphical representation of the linear extension of the product order based on $M_1(\mathbf{x}_i) = \frac{1}{2}\mathbf{x}_i(1) + \frac{1}{2}\mathbf{x}_i(2)$ and $M_2(\mathbf{x}_i) = \mathbf{x}_i(2)$ for the points $\mathbf{x}_1 = (3,1)$, $\mathbf{x}_2 = (1,3)$ and $\mathbf{x}_3 = (3,3)$. The green area represents the points that lead to values of M_1 smaller than those given by $\mathbf{x}_1$ and $\mathbf{x}_2$. The blue area represents the points that lead to values of M_2 smaller than that given by $\mathbf{x}_2$. The green and blue dashed arrows respectively represent the direction in which M_1 and M_2 increase.

3 Extending OWA operators to the setting of multidimensional data by means of a linear extension of the product order

OWA operators as defined by Yager [12] for the aggregation of unidimensional data are characterized by a weighing vector $\mathbf{w} = (w_1, \ldots, w_n)$ with $w_i \geq 0$ for any

$i \in \{1, \ldots, n\}$ and $\sum_{i=1}^{n} w_i = 1$. In particular, the OWA operator $f_{\mathbf{w}} : \mathbb{R}^n \to \mathbb{R}$ associated with $\mathbf{w}$ is defined as

$$f_{\mathbf{w}}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i \, x_{(i)} \, ,$$

where $x_{(i)}$ denotes the i-th largest value among $x_1, \ldots, x_n$.

Typical examples of OWA operators are

- the minimum, where $\mathbf{w} = (0, \ldots, 0, 1)$;
- the mid-range, where $\mathbf{w} = (\frac{1}{2}, 0, \ldots, 0, \frac{1}{2})$;
- the arithmetic mean, where $\mathbf{w} = (\frac{1}{n}, \ldots, \frac{1}{n})$;
- the median, where,
  - if n is odd, $\mathbf{w} = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 appears at the middle position, or,
  - if n is even, $\mathbf{w} = (0, \ldots, 0, \frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$, where the two $\frac{1}{2}$ appear at the middle positions;
- and the maximum, where $\mathbf{w} = (1, 0, \ldots, 0)$.

When moving to the setting of multidimensional data, since $\leq_m$ is not a linear order on $\mathbb{R}^m$, it is possible that one cannot simply identify the i-th largest point among $\mathbf{x}_1, \ldots, \mathbf{x}_n$. De Miguel et al. proposed to consider a linear extension of $\mathbb{R}^m$ in order to straightforwardly extend OWA operators to the setting of multidimensional data. More precisely, given a linear extension $\preceq$ of the product order $\leq_m$ on $\mathbb{R}^m$, the OWA operator $f_{\mathbf{w},\preceq} : (\mathbb{R}^m)^n \to \mathbb{R}^m$ associated with a weighing vector $\mathbf{w} = (w_1, \ldots, w_n)$ and $\preceq$ is defined as follows:

$$f_{\mathbf{w},\preceq}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} w_i \, \mathbf{x}_{(i)} \, ,$$

where $\mathbf{x}_{(i)}$ denotes the i-th largest point among $\mathbf{x}_1, \ldots, \mathbf{x}_n$ according to $\preceq$.

Example 2. Continue with Example 1. Consider $\mathbf{w} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Due to the fact that $\mathbf{x}_1 \preceq_{\text{XY}} \mathbf{x}_2 \preceq_{\text{XY}} \mathbf{x}_3$, it holds that

$$f_{\mathbf{w},\preceq_{\text{XY}}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \frac{1}{4}\mathbf{x}_3 + \frac{1}{2}\mathbf{x}_2 + \frac{1}{4}\mathbf{x}_1 = \frac{1}{4}(3,3) + \frac{1}{2}(1,3) + \frac{1}{4}(3,1) = \left(2, \frac{5}{2}\right) .$$

4 A strange behaviour

A further look reflects that the computation of the OWA operator $f_{\mathbf{w},\preceq_{\mathbf{M}}} : (\mathbb{R}^m)^n \to \mathbb{R}^m$ associated with a weighing vector $\mathbf{w} = (w_1, \ldots, w_n)$ and a linear extension $\preceq_{\mathbf{M}}$ of the product order by means of m weighted arithmetic means $\mathbf{M}$ works as follows. At first $\mathbb{R}^m$ is reduced into the unidimensional quotient space spanned by the level sets of M_1. The unidimensional OWA operator $f_{\mathbf{w}} : \mathbb{R}^n \to \mathbb{R}$ associated with $\mathbf{w}$ is computed within this unidimensional space and it is only in case the order of the points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ is not uniquely determined that $M_2, \ldots, M_m$ are further considered.

This oversimplification of an m-dimensional space into a unidimensional space leads to two main issues. Firstly, unlike in the unidimensional setting, OWA operators as defined in the previous section are no longer continuous functions.

Example 3. Continue with Example 2. Consider now $\mathbf{x}_2' = (1-\varepsilon, 3)$ for some $\varepsilon > 0$. It then holds that $\mathbf{x}_2' \preceq_{\mathrm{XY}} \mathbf{x}_1 \preceq_{\mathrm{XY}} \mathbf{x}_3$, and, thus,

$$\begin{aligned} f_{\mathbf{w},\preceq_{\mathrm{XY}}}(\mathbf{x}_1, \mathbf{x}_2', \mathbf{x}_3) &= \frac{1}{4}\mathbf{x}_3 + \frac{1}{2}\mathbf{x}_1 + \frac{1}{4}\mathbf{x}_2' = \frac{1}{4}(3,3) + \frac{1}{2}(3,1) + \frac{1}{4}(1-\varepsilon,3) \\ &= \left(\frac{5}{2} - \frac{\varepsilon}{4}, 2\right) . \end{aligned}$$

It is concluded that $f_{\mathbf{w},\preceq_{\mathrm{XY}}}$ is not continuous.

Even worse, points have such an undesirable freedom of movement within its own level set of the first weighted arithmetic mean that even very robust OWA operators in the presence of outliers in the unidimensional setting become non-robust in higher dimensions.

Example 4. Continue with Example 2. Consider now the weighing vector $\mathbf{w}' = (0,1,0)$ associated with the median. Note that, for any $a > 0$, it holds that $\mathbf{x}_1 \preceq_{\mathrm{XY}} \big(\mathbf{x}_2 + (-a,a)\big) \preceq_{\mathrm{XY}} \mathbf{x}_3$. For instance, let $a = 100$, it holds that

$$\begin{aligned} f_{\mathbf{w}',\preceq_{\mathrm{XY}}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) &= \mathbf{x}_2 = (1,3) \ , \\ f_{\mathbf{w}',\preceq_{\mathrm{XY}}}\big(\mathbf{x}_1, \big(\mathbf{x}_2 + (-a,a)\big), \mathbf{x}_3\big) &= \mathbf{x}_2 + (-a,a) = (-99, 103) \ . \end{aligned}$$

It is concluded that $f_{\mathbf{w}',\preceq_{\mathrm{XY}}}$ is not robust. Actually, this example implies that the finite sample breakdown point [3, 8] of the median is $\frac{1}{n}$ in the multidimensional setting, rather than $\frac{1}{2}$ as in the unidimensional setting. This lack of robustness might not be a big deal in the context of De Miguel et al. [2] since the use of these OWA operators is restricted to a unit hypercube, however, it definitely becomes a major problem if dealing with an unbounded domain (as typically is the case in multivariate statistics [5]).

5 Concluding remarks

In this paper, a recent extension of OWA operators to the setting of multidimensional data is discussed. As natural as said extension sounds, it is proven to lead to functions that are neither continuous, nor robust. This is due to the fact that the use of a linear extension of the product order is inherently linked to a unidimensional behaviour, and should definitely be abandoned in the multidimensional setting. The use of geometric quantiles [1] instead of linear extensions of the product order in the construction of OWA operators for multidimensional data is encouraged by the author. This direction will be further explored in future work.

References

1. Chaudhuri, P.: On a geometric notion of quantiles for multivariate data. Journal of the American Statistical Association **91**(434), 862–872 (1996)
2. De Miguel, L., Sesma-Sara, M., Elkano, M., Asiain, M., Bustince, H.: An algorithm for group decision making using n-dimensional fuzzy sets, admissible orders and OWA operators. Information Fusion **37**, 126–131 (2017)
3. Donoho, D.L., Huber, P.J.: The notion of breakdown point. In: A Festschrift for Erich L. Lehmann, pp. 157–184. Wadsworth, Belmont, USA (1983)
4. Fishburn, P.C.: Lexicographic orders, utilities and decision rules: A survey. Management Science **20**(11), 1442–1471 (1974)
5. Gagolewski, M.: Penalty-based aggregation of multidimensional data. Fuzzy Sets and Systems **325**, 4–20 (2017)
6. Gagolewski, M., Pérez-Fernández, R., De Baets, B.: An inherent difficulty in the aggregation of multidimensional data. IEEE Transactions on Fuzzy Systems, in press. DOI: 10.1109/TFUZZ.2019.2908135 (2019)
7. Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E.: Aggregation functions: Means. Information Sciences **181**, 1–22 (2011)
8. Hampel, F.R.: A general qualitative definition of robustness. The Annals of Mathematical Statistics **42**, 1887–1896 (1971)
9. Pérez-Fernández, R., De Baets, B., Gagolewski, M.: A taxonomy of monotonicity properties for the aggregation of multidimensional data. Information Fusion **52**, 322–334 (2019)
10. Small, C.G.: A survey of multidimensional medians. International Statistical Review **58**(3), 263–277 (1990)
11. Xu, Z., Yager, R.R.: Some geometric aggregation operators based on intuitionistic fuzzy sets. International Journal of General Systems **35**(4), 417–433 (2006)
12. Yager, R.R.: On Ordered Weighted Averaging aggregation operators in multicriteria decisionmaking. IEEE Transactions on Systems, Man, and Cybernetics **18**(1), 183–190 (1988)
13. Yager, R.R.: Families of OWA operators. Fuzzy Sets and Systems **59**, 125–148 (1993)
14. Yager, R.R.: Centered OWA operators. Soft Computing **11**, 631–639 (2007)