Hierarchical Clustering with OWA-based Linkages, the Lance–Williams Formula, and Dendrogram Inversions

Marek Gagolewski[a,b,c,*], Anna Cena[b], Simon James[a] and Gleb Beliakov[a]

[a]Deakin University, School of IT, Geelong, VIC 3220, Australia
[b]Warsaw University of Technology, Faculty of Mathematics and Information Science, ul. Koszykowa 75, 00-662 Warsaw, Poland
[c]Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447 Warsaw, Poland

ARTICLE INFO

Keywords:
OWA operators
hierarchical clustering
dendrogram
inversion
the Lance–Williams formula

Abstract

Agglomerative hierarchical clustering based on Ordered Weighted Averaging (OWA) operators not only generalises the single, complete, and average linkages, but also includes intercluster distances based on a few nearest or farthest neighbours, trimmed and winsorised means of pairwise point similarities, amongst many others. We explore the relationships between the famous Lance–Williams update formula and the extended OWA-based linkages with weights generated via infinite coefficient sequences. Furthermore, we provide some conditions for the weight generators to guarantee the resulting dendrograms to be free from unaesthetic inversions.

Please cite this paper as: Gagolewski M., Cena A., James S., Beliakov G., Hierarchical clustering with OWA-based linkages, the Lance–Williams formula, and dendrogram inversions, Fuzzy Sets and Systems ***473****, 108740, 2023, DOI:10.1016/j.fss.2023.108740*

1. Introduction

Cluster analysis (e.g., [18]) is a statistical and machine learning task whose aim is to discover interesting or otherwise useful partitions of a given dataset in a purely unsupervised way.

Hierarchical agglomerative clustering algorithms (e.g., [16]) allow for partitioning the datasets for which merely a pairwise distance function (e.g., a metric) is defined. Most importantly, the number of clusters is not set in advance – a whole hierarchy of nested partitions can be generated with ease, and then depicted on a tree-like diagram called a *dendrogram*.

Hierarchical agglomerative clustering revolves around one simple idea: in each step, we merge the pair of *closest* clusters. To measure the proximity between two point sets, the intracluster distance is defined as an extension of the point-pairwise distance called a *linkage function*.

For instance, in the single linkage approach, the distance between a cluster pair is given by the distance between the closest pair of points, one from the first cluster, the other from the second one. In complete linkage, we take the farthest-away pair. In average linkage, we compute the arithmetic mean between all the pairwise distances.

In Section 3, we recall two wide classes of linkage functions that generalise these three cases. The first group consists of linkages generated by the well-known Lance–Williams formula [13, 15]. The second class considers convex combinations of ordered pairwise distances between clusters, i.e., the OWA operators (ordered weighted averages [19]). Such OWA-based linkages were introduced by Yager in [20] (see also [17] where they were re-invented) and include many linkage generators that are not covered by the classical Lance–Williams setting.

From the practical side, their usefulness has been thoroughly evaluated in [4], where also some further tweaks were proposed to increase the quality of the generated results, e.g., by including in Genie correction for cluster size inequality [7].

However, OWA-based linkages have not been studied thoroughly from the theoretical perspective. This paper aims to fill this gap.

*Corresponding author
Email addresses: m.gagolewski@deakin.edu.au (M. Gagolewski); anna.cena@pw.edu.pl (A. Cena); s.james@deakin.edu.au (S. James); gleb@deakin.edu.au (G. Beliakov)
URL: https://www.gagolewski.com (M. Gagolewski)
ORCID(s): 0000-0003-0637-6028 (M. Gagolewski)

In particular, after recalling the basic definitions in Section 2 and Section 3, Section 4 characterises the relationship between the Lance–Williams linkage update formula and OWA linkages with two different weight generators. Then, Section 5 gives some conditions for the linkages to yield clusterings which can be represented aesthetically on dendrograms (without the so-called inversions). Section 6 concludes the paper.

2. Hierarchical Agglomerative Clustering

Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be the input data set comprised of n points in $\mathbb{R}^d$, where $\boldsymbol{x}_i = (x_{i,1}, \dots, x_{i,d})$ for each i. From now on, we assume that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ is the Euclidean distance between two given points. However, let us note that most of the results presented below hold for any set of n objects equipped with a semimetric.

A *k-partition* of $\mathbf{X}$, $k \geqslant 1$, is defined as $\mathcal{C} = \{C_1, \dots, C_k\}$, where $\emptyset \neq C_u \subseteq \mathbf{X}$, $C_u \cap C_v = \emptyset$ for $u \neq v$ and $\bigcup_{u=1}^{k} C_u = \mathbf{X}$.

Here is the most basic scheme in which the general *agglomerative hierarchical clustering* procedure can be formalised:

1. Start with $\mathcal{C}^{(0)} = \{C_1^{(0)}, \dots, C_n^{(0)}\}$, $C_i^{(0)} = \{\boldsymbol{x}_i\}$, where each cluster is a singleton.
2. For $j = 1, \dots, n-1$:
   2.1. Select clusters to merge:
   $$(u, v) = \underset{(u,v),\ u<v}{\arg\min}\, d(C_u^{(j-1)}, C_v^{(j-1)});$$
   2.2. Merge clusters $C_u^{(j-1)}$ and $C_v^{(j-1)}$, i.e.:
      2.2.1. $C_u^{(j)} = C_u^{(j-1)} \cup C_v^{(j-1)}$;
      2.2.2. $C_i^{(j)} = C_i^{(j-1)}$ for $u \neq i < v$;
      2.2.3. $C_i^{(j)} = C_{i+1}^{(j-1)}$ for $i > v$.

Thus, an agglomerative hierarchical clustering algorithm forms a hierarchy of nested $(n, n-1, \dots, 2, 1)$-partitions $\mathcal{C}^* = \left(\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \dots, \mathcal{C}^{(n-1)}\right)$.

In the algorithm above, $d : 2^{\mathbf{X}} \times 2^{\mathbf{X}} \to [0, \infty]$ denotes a chosen *linkage function*, whose aim is to measure the distance between two point clusters, fulfilling at least $d(U, U) = 0$, $d(U, V) = d(V, U)$, and:

$$d(\{\boldsymbol{x}_i\}, \{\boldsymbol{x}_j\}) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|$$

for any $U, V \subseteq \mathbf{X}$ and $\boldsymbol{x}_i, \boldsymbol{x}_j \in \mathbf{X}$.

Classical choices of d include (see, e.g., [16]):

- the single linkage:
$$d_{\mathrm{MIN}}(U, V) = \min_{\boldsymbol{u}\in U, \boldsymbol{v}\in V} \|\boldsymbol{u} - \boldsymbol{v}\|,$$
- the complete linkage:
$$d_{\mathrm{MAX}}(U, V) = \max_{\boldsymbol{u}\in U, \boldsymbol{v}\in V} \|\boldsymbol{u} - \boldsymbol{v}\|,$$
- the average linkage (UPGMA; *Unweighted Pair Group Method with Arithmetic Mean*):
$$d_{\mathrm{AMean}}(U, V) = \frac{1}{|U||V|} \sum_{\boldsymbol{u}\in U, \boldsymbol{v}\in V} \|\boldsymbol{u} - \boldsymbol{v}\|,$$

- the Ward linkage:

$$
\begin{aligned}
d_{\text{Ward}}(U,V) &= \frac{2}{|U|+|V|}\sum_{\boldsymbol{u}\in U,\boldsymbol{v}\in V}\|\boldsymbol{u}-\boldsymbol{v}\|^2 \\
&- \frac{|V|}{|U|(|U|+|V|)}\sum_{\boldsymbol{u},\boldsymbol{u}'\in U}\|\boldsymbol{u}-\boldsymbol{u}'\|^2 \\
&- \frac{|U|}{|V|(|U|+|V|)}\sum_{\boldsymbol{v},\boldsymbol{v}'\in V}\|\boldsymbol{v}-\boldsymbol{v}'\|^2,
\end{aligned}
$$

- the centroid linkage (UPGMC; *Unweighted Pair Group Method Centroid*):

  $$d_{\text{Cent}}(U,V) = \|\boldsymbol{\mu}_u - \boldsymbol{\mu}_v\|,$$

  where $\boldsymbol{\mu}_u$, $\boldsymbol{\mu}_v$ are the respective clusters' centroids (componentwise arithmetic means),

- weighted average linkage (WPGMA; *Weighted Pair Group Method with Arithmetic Mean*):

  $$d_{\text{WAMean}}(U,V) = \frac{|W|}{|W|+|Z|}d_{\text{WAMean}}(U,W) + \frac{|Z|}{|W|+|Z|}d_{\text{WAMean}}(U,Z),$$

  assuming that $V = W \cup Z$ in one of the previous iterations,

- the median linkage (WPGMC; *Weighted Pair Group Method Centroid*) given by:

  $$d_{\text{Median}}(U,V) = \frac{1}{2}d_{\text{Median}}(U,W) + \frac{1}{2}d_{\text{Median}}(U,Z) - \frac{1}{4}d_{\text{Median}}(W,Z),$$

  assuming that $V = W \cup Z$ in one of the previous iterations.

Some of the above cases can be generalised through the Lance–Williams formula [13, 15] or the OWA-based linkages [20], which we shall discuss next.

3. Linkage Classes

Let us recall two noteworthy linkage classes, based respectively on the Lance–Williams formula and the OWA operator-based intercluster distances.

3.1. Lance–Williams Linkages

In [13], G. Lance and W. Williams proposed an iterative formula that generalises many common linkages and allows for a fast update of the intercluster distances after each cluster pair merge.

Assuming that in the j-th step of the procedure we are about to merge $C_u^{(j-1)}$ and $C_v^{(j-1)}$, then for every other (intact) cluster $C_z^{(j)} = C_z^{(j-1)}$, $z \notin \{u,v\}$, the new distances are:

$$
\begin{aligned}
d\left(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}\right) &= \alpha_u\, d(C_z^{(j-1)}, C_u^{(j-1)}) \\
&+ \alpha_v\, d(C_z^{(j-1)}, C_v^{(j-1)}) \\
&+ \beta\;\, d(C_u^{(j-1)}, C_v^{(j-1)}) \\
&+ \gamma\;\, \left| d(C_z^{(j-1)}, C_u^{(j-1)}) - d(C_z^{(j-1)}, C_v^{(j-1)}) \right|,
\end{aligned}
\tag{1}
$$

for some α_u, α_v, β, and γ that might depend on $n_u = |C_u^{(j-1)}|$, $n_v = |C_v^{(j-1)}|$, and $n_z = |C_z^{(j)}|$.

Table 1 gives some common choices of the above coefficients.

Note that the Lance–Williams formula only utilises the information about $d(C_u^{(j-1)}, C_v^{(j-1)})$, $d(C_z^{(j-1)}, C_u^{(j-1)})$, and $d(C_z^{(j-1)}, C_v^{(j-1)})$, as well as the cardinalities of the clusters. Further, as we would like the linkage to be symmetric, it is required that $\alpha_u(n_u,n_v,n_z) = \alpha_v(n_v,n_u,n_z)$, $\beta(n_u,n_v,n_z) = \beta(n_v,n_u,n_z)$, and $\gamma(n_u,n_v,n_z) = \gamma(n_v,n_u,n_z)$.

Table 1
Common coefficients in the Lance–Williams formula [13, 15]

linkage	α_u	α_v	β	γ
single	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
complete	$\frac{1}{2}$	$\frac{1}{2}$	0	$+\frac{1}{2}$
average (UPGMA)	$\frac{n_u}{n_u+n_v}$	$\frac{n_v}{n_u+n_v}$	0	0
weighted average (WPGMA)	$\frac{1}{2}$	$\frac{1}{2}$	0	0
centroid (UPGMC)	$\frac{n_u}{n_u+n_v}$	$\frac{n_v}{n_u+n_v}$	$-\frac{n_u n_v}{(n_u+n_v)^2}$	0
median (WPGMC)	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
Ward	$\frac{n_u+n_z}{n_u+n_j+n_z}$	$\frac{n_v+n_z}{n_u+n_v+n_z}$	$-\frac{n_z}{n_u+n_v+n_z}$	0

3.2. OWA-based linkages

OWA operators, i.e., convex combinations of order statistics, were introduced in the aggregation and decision making context by Yager in [19]. Let us introduce their version that acts on element sequences of any length.

Definition 1. An *extended* [3, 14] *OWA operator* is defined as:

$$\mathrm{OWA}_{\triangle}(d_1, d_2, \dots, d_m) = \sum_{i=1}^{m} c_{i,m} d_{(i)}$$

where a given weighting triangle is $\triangle = (c_{i,m} \in [0,1],\ m \in \mathbb{N}, i = 1, \dots, m \,:\, (\forall m) \sum_{i=1}^{m} c_{i,m} = 1)$, and $d_{(1)} \geqslant d_{(2)} \geqslant \dots \geqslant d_{(m)}$.

By definition, $\mathrm{OWA}_{\triangle}$ is symmetric, idempotent, and nondecreasing in each variable; compare [1, 5, 10]. Hence, it is also internal, i.e., $d_{(1)} \geqslant \mathrm{OWA}_{\triangle}(d_1, \dots, d_m) \geqslant d_{(m)}$ for all possible inputs $d_1, \dots, d_m$ of any cardinality.

In the clustering context, OWA-based linkage was proposed by Yager in [20]; see also [4, 17].

Definition 2. For a given weighting triangle $\triangle$, the $\mathrm{OWA}_{\triangle}$*-based linkage* is defined as:

$$d_{\triangle}(C_u, C_v) = \mathrm{OWA}_{\triangle}(\{\|\boldsymbol{u} - \boldsymbol{v}\| \,:\, \boldsymbol{u} \in C_u, \boldsymbol{v} \in C_v\})$$

for any point sets C_u and C_v, i.e., it is the OWA operator applied on all the $m = |C_u||C_v|$ pairwise distances between the two sets.

Remark 3. In particular, for weights fulfilling:

- $c_{i,m} = \frac{1}{m}$: we obtain the average linkage,
- $c_{m,m} = 1$ and $c_{i,m} = 0$ for $i < m$: we get the single linkage,
- $c_{1,m} = 1$ and $c_{i,m} = 0$ for $i > 1$: we enjoy the complete linkage.

Note that numerous new linkages that do not fit the Lance–Williams formula are now possible. This includes, e.g., distance aggregates that correspond to the median and any other quantiles, trimmed and winsorised means, the arithmetic mean of a few first smallest observations, a fuzzified/smoothed minimum, and so forth; see [4] for many example classes.

There are a few generic ways to generate the weighting triangles known in the literature; see [11, 19] and [2, 9]. In this paper, we will be interested in studying:

- $c_{i,m} = \frac{c_i}{\sum_{j=1}^{m} c_j}$, where $(c_1, c_2, \dots)$ is such that $c_i \geqslant 0$ for all $i = 1, 2, \dots$ and $c_1 = 1$,
- $c_{m-i+1,m} = \frac{c_i}{\sum_{j=1}^{m} c_j}$, where $(c_1, c_2, \dots)$ is such that $c_i \geqslant 0$ for all $i = 1, 2, \dots$ and $c_1 = 1$.

The assumption that $c_1 = 1$ is with no loss in generality provided that we disallow the ill-defined case $c_1 = 0$, where we could have utilised the convention $0/0 = 1$. We do not enable it to keep the presentation simple.

Hence, for a given coefficients vector $\boldsymbol{c} = (c_1, c_2, \dots)$ with $c_1 = 1$ and $c_2, c_3, \dots, \geqslant 0$, we can define two extended OWA operators:

$$\mathrm{OWA}_{\boldsymbol{c}}(d_1, \dots, d_m) = \frac{\sum_{i=1}^{m} c_i d_{(i)}}{\sum_{i=1}^{m} c_i}, \tag{2}$$

$$\mathrm{OWA}'_{\boldsymbol{c}}(d_1, \dots, d_m) = \frac{\sum_{i=1}^{m} c_i d_{(m-i+1)}}{\sum_{i=1}^{m} c_i}, \tag{3}$$

for any m and $d_1, \dots, d_m$, where $d_{(i)}$ denotes the i-th greatest value in the sequence.

4. OWA linkages and the Lance–Williams formula

In this section, we characterise which OWA-based linkages (defined via $\mathrm{OWA}_{\boldsymbol{c}}$ and $\mathrm{OWA}'_{\boldsymbol{c}}$) can be expressed by the Lance–Williams formula, and vice versa. It turns out that under the two assumed weight generation schemes, these only include the single, average, and complete linkages.

Theorem 4. Assume that d is generated by the Lance–Williams formula, i.e., for any C_u, C_v, C_z it holds:

$$\begin{aligned} d\Big(C_z, C_u \cup C_v\Big) &= \alpha_u\, d(C_z, C_u) \\ &+ \alpha_v\, d(C_z, C_v) \\ &+ \beta\ \, d(C_u, C_v) \\ &+ \gamma\ \Big|d(C_z, C_u) - d(C_z, C_v)\Big|, \end{aligned}$$

with $\alpha_u(|C_u|, |C_v|, |C_z|) = \alpha_v(|C_v|, |C_u|, |C_z|)$, $\beta(|C_u|, |C_v|, |C_z|) = \beta(|C_v|, |C_u|, |C_z|)$, and $\gamma(|C_u|, |C_v|, |C_z|) = \gamma(|C_v|, |C_u|, |C_z|)$. Then there exists $\boldsymbol{c} = (c_1, c_2, \dots)$ with $c_1 = 1$ and $c_2, c_3 \geqslant 0$ such that for every C_u, C_v it holds:

$$d(C_u, C_v) = \mathrm{OWA}_{\boldsymbol{c}}(\{\|\boldsymbol{u} - \boldsymbol{v}\| : \boldsymbol{u} \in C_u, \boldsymbol{v} \in C_v\})$$

if and only if $\boldsymbol{c}$ is either:

- $\boldsymbol{c} = (1, 0, 0, \dots)$ (complete linkage – the maximum) or
- $\boldsymbol{c} = (1, 1, 1, \dots)$ (average linkage – the arithmetic mean).

Proof. That these conditions are sufficient is evident.

Hence, assume that d is generated by the Lance–Williams formula and that:

$$d\Big(C_z, C_u \cup C_v\Big) = \frac{\sum_{i=1}^{n+m} c_i w_{(i)}}{\sum_{i=1}^{n+m} c_i}$$

for $\boldsymbol{w} = (u_1, \dots, u_n, v_1, \dots, v_m)$, where u_i is the distance between an i-th pair of points in $C_u \times C_z$ and v_j is the distance between a j-th pair of points in $C_v \times C_z$. Due to the symmetry of OWAs, it does not matter how we enumerate the pairs, hence we assume $u_1 \geqslant \dots \geqslant u_n$ and $v_1 \geqslant \dots \geqslant v_m$, where $n = |C_u||C_z|$, $m = |C_v||C_z|$.

It is apparent that $d\left(C_z, C_u \cup C_v\right)$ cannot depend on $d(C_u, C_v)$, and hence it necessarily holds that $\beta = 0$. Therefore, let us note that if $\mathrm{OWA}_{\boldsymbol{c}}(u_1, \dots, u_n) \geqslant \mathrm{OWA}_{\boldsymbol{c}}(v_1, \dots, v_m)$, then we have:

$$\mathrm{OWA}_c(u_1, \dots, u_n, v_1, \dots, v_m) = (\alpha_u + \gamma)\mathrm{OWA}_c(u_1, \dots, u_n) + (\alpha_v - \gamma)\mathrm{OWA}_c(v_1, \dots, v_m).$$

As mentioned earlier, each OWA operator is internal. Hence, it necessarily holds $\alpha_u + \alpha_v = 1$, a condition which we obtain by considering $n = m = 1$ and $u_1 = v_1 = 1$, because it yields:

$$\frac{c_1 \cdot 1 + c_2 \cdot 1}{c_1 + c_2} = (\alpha_u + \gamma) \cdot 1 + (\alpha_v - \gamma) \cdot 1.$$

From now on assume that $\boldsymbol{c} \neq (1, 0, 0, \dots)$ (i.e., it is not the complete linkage), i.e., $\alpha_u + \gamma < 1$ and $\alpha_v - \gamma > 0$.

In the case where $n = m$ and $u_1 = v_1 = 1$ and $u_2 = v_2 = u_3 = v_3 = \cdots = 0$ we have:

$$\frac{(c_1 + c_2) \cdot 1}{\sum_{i=1}^{2n} c_i} = \frac{c_1 \cdot 1}{\sum_{i=1}^{n} c_i}.$$

Under the assumption that $c_1 = 1$, this implies:

$$c_2 = \frac{\sum_{i=n+1}^{2n} c_i}{\sum_{i=1}^{n} c_i}$$

for all $n \geqslant 1$.

Now consider the case $n = 1$, $m > 1$ with $u_1 \geqslant v_1 \geqslant v_2 \geqslant \dots$. If $u_1 = 1$ and $v_1 = v_2 = \cdots = 0$, this breeds:

$$\frac{c_1 \cdot 1}{\sum_{i=1}^{m+1} c_i} = (\alpha_u + \gamma) \cdot 1.$$

Therefore, $(\alpha_v - \gamma) = \frac{\sum_{i=2}^{m+1} c_i}{\sum_{i=1}^{m+1} c_i}$. If $u_1 = v_1 = 1$ and $v_2 = v_3 = \cdots = 0$, then:

$$\frac{c_1 \cdot 1 + c_2 \cdot 1 + c_3 \cdot 0 + \cdots + c_{m+1} \cdot 0}{\sum_{i=1}^{m+1} c_i} = \frac{c_1}{\sum_{i=1}^{m+1} c_i} \cdot 1 + \frac{\sum_{i=2}^{m+1} c_i}{\sum_{i=1}^{m+1} c_i} \frac{c_1 \cdot 1 + c_2 \cdot 0 + \cdots + c_t \cdot 0}{\sum_{i=1}^{m} c_i}.$$

Thus,

$$c_2 = c_1 \frac{\sum_{i=2}^{m+1} c_i}{\sum_{i=1}^{m} c_i}.$$

Assuming $c_1 = 1$ and $m = 2$, this yields:

$$c_3 = c_2^2.$$

With $c_1 = 1$ and studying further $m > 2$, we get each time that $c_m = c_2^{m-1}$. But from the previous equation we have that $c_2 = \frac{c_3 + c_4}{c_1 + c_2} = \frac{c_2^2 + c_2^3}{1 + c_2}$, Hence, $1 + c_2 = c_2 + c_2^2$ and thus $c_2 = 1$, which implies that necessarily for all i it must hold $c_i = 1$, which corresponds to the average linkage.

As we have already considered the complete linkage case separately, the proof is complete. □

Theorem 5. Assume that d is generated by the Lance–Williams formula, just like above. Then there exists $\boldsymbol{c} = (c_1, c_2, \dots)$ with $c_1 = 1$ and $c_2, c_3 \geqslant 0$ such that for every C_u, C_v we have:

$$d(C_u, C_v) = \mathrm{OWA}'_{\boldsymbol{c}}(\{\|\boldsymbol{u} - \boldsymbol{v}\| : \boldsymbol{u} \in C_u, \boldsymbol{v} \in C_v\})$$

if and only if $\boldsymbol{c}$ is either:

- $\boldsymbol{c} = (1, 0, 0, \dots)$ (single linkage – the minimum) or
- $\boldsymbol{c} = (1, 1, 1, \dots)$ (average linkage – the arithmetic mean).

Proof. Sufficiency of the above is obvious. The reasoning required to show the necessary part is very similar to the one we have conveyed in the proof of Theorem 4.

As it necessarily holds that $\beta = 0$, let us note that if $\mathrm{OWA}'_{\boldsymbol{c}}(u_1, \dots, u_n) \geqslant \mathrm{OWA}'_{\boldsymbol{c}}(v_1, \dots, v_m)$, then we have:

$$\mathrm{OWA}'_{\boldsymbol{c}}(u_1, \dots, u_n, v_1, \dots, v_m) = (\alpha_u + \gamma)\mathrm{OWA}'_{\boldsymbol{c}}(u_1, \dots, u_n) + (\alpha_v - \gamma)\mathrm{OWA}'_{\boldsymbol{c}}(v_1, \dots, v_m),$$

for some $\alpha_u + \alpha_v = 1$.

From now on assume that $\boldsymbol{c} \notin (1, 0, 0, \dots)$ (i.e., it is not the single linkage), i.e., $\alpha_u + \gamma > 0$ and $\alpha_v - \gamma < 1$.

Consider the case where $n = m$ and $u_1 = v_1 = u_2 = v_2 = \dots = u_{n-1} = v_{m-1} = 1$ and $u_n = v_m = 0$. This implies:

$$c_2 = \frac{\sum_{i=n+1}^{2n} c_i}{\sum_{i=1}^{n} c_i}$$

for all $n \geqslant 1$.

Next, we study $n > 1$ and $m = 1$ with $u_1 \geqslant \dots \geqslant u_n \geqslant v_1$. If $v_1 = 0$ and $u_1 = \dots = u_n = 1$, then we obtain:

$$\frac{\sum_{i=2}^{n+1} c_i}{\sum_{i=1}^{n+1} c_i} = (\alpha_u + \gamma) \cdot 1.$$

Therefore, $(\alpha_v - \gamma) = \frac{c_1}{\sum_{i=1}^{n+1} c_i}$. If $u_1 = 1$ and $u_2 = \dots = u_n = v_1 = 0$, then:

$$c_{n+1} = c_n \frac{\sum_{i=2}^{n+1} c_i}{\sum_{i=1}^{n} c_i},$$

which, similarly as in the previous proof, implies that $c_i = 1$ for all i (average linkage), QED. □

5. Inversion-free dendrograms

We can depict a hierarchy of nested partitions, $\mathcal{C}^* = \{\mathcal{C}^{(0)}, \mathcal{C}^{(1)}, \dots, \mathcal{C}^{(n-1)}\}$, on a *dendrogram*, which is an undirected tree whose leaves represent the initial singleton clusters $C_1^{(0)}, \dots, C_n^{(0)}$. The root corresponds to $C_1^{(n-1)} = \mathbf{X}$, and the inside nodes depict the clusters which are merged in each step of the procedure.

Let $h_d : \mathcal{C}^* \to [0, \infty)$ be a function which assigns each hierarchy level $\mathcal{C}^{(j)}$ a specific height, given by:

$$h_d(\mathcal{C}^{(j)}) = \begin{cases} 0 & j = 0, \\ \min_{u,v} d(C_u^{(j-1)}, C_v^{(j-1)}) & j \geqslant 1. \end{cases}$$

In other words, it is the distance (as given by the chosen linkage d) between the two clusters merged in the j-th step.

Remark 6. For example, Figure 1 depicts two cluster dendrograms of an example dataset. On both dendrograms, we see the merging of singleton clusters based on $d(\{\boldsymbol{x}_6\}, \{\boldsymbol{x}_{14}\}) = 0.14142$ (in the first stage) and $d(\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_{13}\}) = 0.5831$ (at different stages). In both cases, the final merge is done between $\{\boldsymbol{x}_{15}, \boldsymbol{x}_4, \boldsymbol{x}_9\}$ and $\boldsymbol{X} \setminus \{\boldsymbol{x}_{15}, \boldsymbol{x}_4, \boldsymbol{x}_9\}$, however, these are at different heights as specified by the linkage functions in use (complete on the lefthand side and centroid on the right).

Unfortunately, as noted in [12, 15], the height function h_d is not necessarily increasing for every linkage d, i.e., it does not always hold that:

$$h_d(\mathcal{C}^{(j-1)}) \leqslant h_d(\mathcal{C}^{(j)}) \tag{4}$$

for all j. This may lead to dendrogram pathologies known as *inversions*.

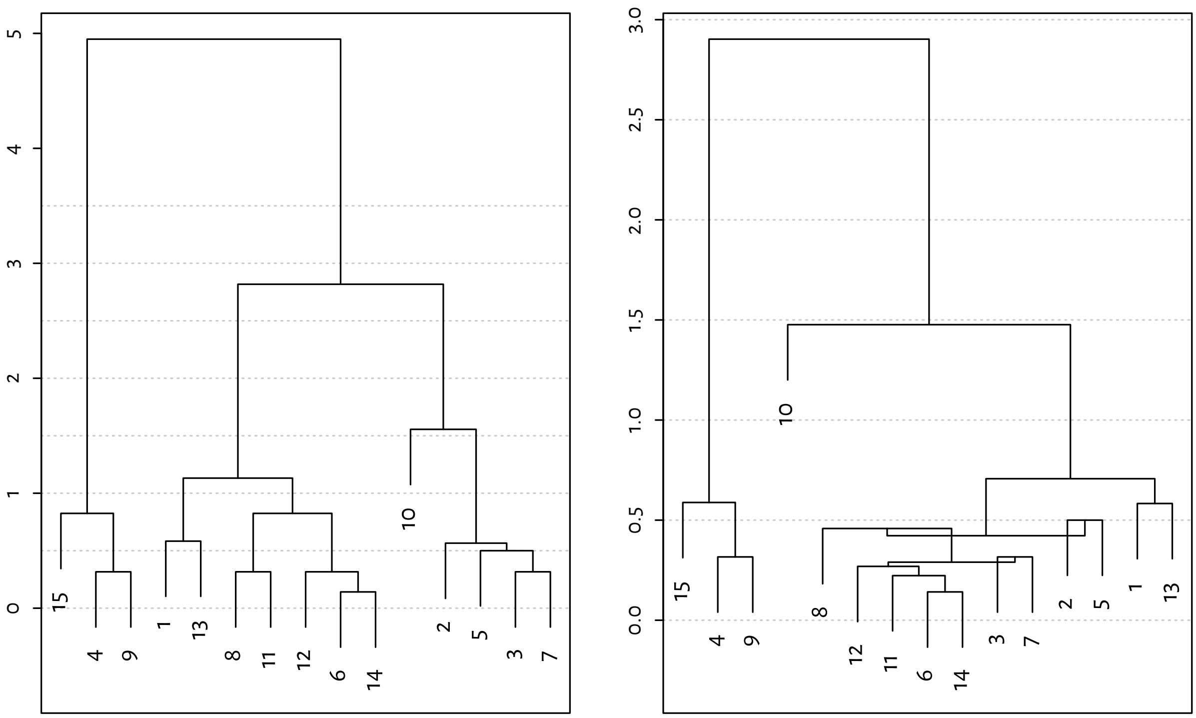

Figure 1: Dendrograms for complete (lefthand side) and centroid (righthand side) linkage-based clustering of the same example subset of the `iris` dataset, as plotted by the `stats:::plot.hclust` function in R. Note the inversions.

Remark 7. The right side of Figure 1 depicts the clustering of an example dataset based on the centroid linkage. Note the two dendrogram inversions, which are due to the fact that the intercluster distances at some higher levels of the hierarchy are smaller than the ones computed in some previous merge steps. This happens when we merge, e.g., the cluster that features the points $\boldsymbol{x}_3$ and $\boldsymbol{x}_7$ with the one containing $\boldsymbol{x}_6$, $\boldsymbol{x}_{11}$, $\boldsymbol{x}_{12}$, and $\boldsymbol{x}_{14}$. In this case, we have $d(\{\boldsymbol{x}_3\},\{\boldsymbol{x}_7\}) > d(\{\boldsymbol{x}_3,\boldsymbol{x}_7\},\{\boldsymbol{x}_6,\boldsymbol{x}_{11},\boldsymbol{x}_{12},\boldsymbol{x}_{14}\})$.

Theorem 8. h_d is increasing if and only if $(\forall j)$ $(\forall z \in \{1,\dots,n-j\})$, $z \neq u$ and for $u, v \in \{1,\dots,n-j+1\}$ such that $(u,v) = \arg\min_{u,v} d(C_u^{(j-1)}, C_v^{(j-1)})$ it holds:

$$d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}) \geqslant d(C_u^{(j-1)}, C_v^{(j-1)}). \tag{5}$$

Proof. ($\Longrightarrow$) Let us fix $j \in \{1,\dots,n-1\}$ and let $(u,v) = \arg\min_{u,v} d(C_u^{(j-1)}, C_v^{(j-1)})$. By the definition of h_d, we have $h_d(\mathcal{C}^{(j)}) = d(C_u^{(j-1)}, C_v^{(j-1)})$. Without loss of generality, we can assume that $u < v$, and therefore $C_u^{(j)} = C_u^{(j-1)} \cup C_v^{(j-1)}$. From this it follows that:

$$\begin{aligned} d(C_u^{(j-1)}, C_v^{(j-1)}) = h_d(\mathcal{C}^{(j)}) \leqslant h_d(\mathcal{C}^{(j+1)}) = d(C_q^{(j)}, C_w^{(j)}) \\ \leqslant d(C_z^{(j)}, C_u^{(j)}) = d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}) \end{aligned}$$

for all $z \in \{1,\dots,n-j\}$ where $d(C_q^{(j)}, C_w^{(j)}) = \min_{q,w} d(C_q^{(j)}, C_w^{(j)})$.

($\Longleftarrow$) Let us assume that:

$$d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}) \geqslant d(C_u^{(j-1)}, C_v^{(j-1)}) = \min_{u<v} d(C_u^{(j-1)}, C_v^{(j-1)}),$$

but at the same time:

$$\min_{u<v} d(C_u^{(j-1)}, C_v^{(j-1)}) > \min_{z,w} d(C_z^{(j)}, C_w^{(j)}).$$

We thus get $C_w^{(j)} \neq C_u^{(j-1)} \cup C_v^{(j-1)} \neq C_z^{(j)}$ and also $C_w^{(j)} = C_w^{(j-1)}$ and $C_z^{(j)} = C_z^{(j-1)}$. This contradicts our assumption, since $\min_{z,w} d(C_z^{(j)}, C_w^{(j)}) = \min_{z,w} d(C_z^{(j-1)}, C_w^{(j-1)})$. Therefore, h_d is increasing and the proof is complete. □

As far as the Lance–Williams formula is concerned, we have the following result [15].

Theorem 9. If:

(1) $\alpha_u + \alpha_v + \beta \geqslant 1$,
(2) $\alpha_u \geqslant 0$, $\alpha_v \geqslant 0$
(3) γ is such that:
  (a) $\gamma \geqslant 0$ or
  (b) $\gamma < 0$ and $|\gamma| \leqslant \min\{\alpha_u, \alpha_v\}$,

then for d given by (1), it holds:

$$d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}) \geqslant d(C_u^{(j-1)}, C_v^{(j-1)}),$$

for every $j = 1, \ldots, n-1$, $(u, v) = \arg\min_{(u,v)} d(C_u^{(j-1)}, C_v^{(j-1)})$, $u \neq v$, $z \neq u$, and $z \in \{1, \ldots, n-j\}$, i.e., the corresponding h_d is increasing.

See [15] for a proof.
Hence, single, complete, average, weighted average, and Ward linkages yield increasing h_d.

Let us move on to the OWA linkage case. We note that already even very simple weighting triangles can fail to produce increasing OWA-based h_d.

Remark 10. If $d_\triangle(d_1, \ldots, d_m) = 0.5(d_{(m)} + d_{(m-1)})$ for $m > 1$, i.e., the arithmetic mean of the two smallest values, then h_d is not increasing. For example, assume that $\mathbf{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}$ and let the distance matrix be such that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ are given by:

$$\mathbf{D} = \begin{bmatrix} 0.0 & 0.4 & 0.6 & 0.9 \\ 0.4 & 0.0 & 0.9 & 0.6 \\ 0.6 & 0.9 & 0.0 & 0.7 \\ 0.9 & 0.6 & 0.7 & 0.0 \end{bmatrix}.$$

Then the agglomerative clustering algorithm yields what follows.

1. Start with 4 singletons $\{\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$;
   the closest pair of clusters is such that $d(\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}) = 0.4$;
   merge this pair.
2. Current partition is $\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}\}$;
   now the closest pair of clusters has $d(\{\boldsymbol{x}_3\}, \{\boldsymbol{x}_4\}) = 0.7$;
   merge them.
3. Current partition is $\{\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}\}$;
   the closest (and the only) pair now fulfils $d(\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3, \boldsymbol{x}_4\}) = 0.6$ (which is a smaller distance than in the previous iteration);
   merge this pair.

Thus, when depicted on a dendrogram, we would get an inversion.

Unfortunately, for OWA-based linkages, $d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)})$ does not depend on $d(C_u^{(j-1)}, C_v^{(j-1)})$. Therefore, we need to consider a different condition that will be suitable in the dendrogram plotting context.

By the construction of the agglomerative hierarchical clustering algorithm, we know that $d(C_u^{(j-1)}, C_v^{(j-1)}) \leqslant d(C_z^{(j-1)}, C_u^{(j-1)})$ and $d(C_u^{(j-1)}, C_v^{(j-1)}) \leqslant d(C_z^{(j-1)}, C_v^{(j-1)})$ for all $z \notin \{u, v\}$, i.e.,

$$d(C_u^{(j-1)}, C_v^{(j-1)}) \leqslant \min\{d(C_z^{(j-1)}, C_u^{(j-1)}), d(C_z^{(j-1)}, C_v^{(j-1)})\},$$

as otherwise, C_z would have been merged with either C_u or C_j in the j-th step, and not C_u with C_v.

Therefore, if:

$$d(C_z^{(j)}, C_u^{(j-1)} \cup C_v^{(j-1)}) \geqslant \min\{d(C_z^{(j-1)}, C_u^{(j-1)}), d(C_z^{(j-1)}, C_v^{(j-1)})\}, \tag{6}$$

then *necessarily* the corresponding h_d is increasing.

Focusing on the first type of extended OWA, we shall thus be interested in identifying coefficient vectors $\boldsymbol{c}$ such that:

$$\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}, \boldsymbol{v}) \geqslant \min\{\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}), \mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{v})\} \tag{7}$$

for all $\boldsymbol{u}, \boldsymbol{v}$ of any cardinalities, where $(\boldsymbol{u}, \boldsymbol{v})$ denotes vector concatenation.

The first result we present toward this will help us establish some necessary conditions.

Theorem 11. Let $\boldsymbol{c} = (c_1, c_2, \dots)$ be a coefficient vector with $c_1 = 1$ and $c_2, c_3, \dots \geqslant 0$. Then for any n, m and $u_1, \dots, u_n, v_1, \dots, v_m$, if it holds:

$$\mathrm{OWA}_{\boldsymbol{c}}(u_1, \dots, u_n, v_1, \dots, v_m) \geqslant \min\{\mathrm{OWA}_{\boldsymbol{c}}(u_1, \dots, u_n), \mathrm{OWA}_{\boldsymbol{c}}(v_1, \dots, v_m)\},$$

then for all k, l such that: $\sum_{i=1}^n c_i \sum_{i=1}^l c_i \geqslant \sum_{i=1}^m c_i \sum_{i=1}^k c_i$, we must have:

$$\sum_{i=k+1}^{k+l} c_i \sum_{i=1}^{m} c_i \geqslant \sum_{i=n+1}^{n+m} c_i \sum_{i=1}^{l} c_i. \tag{8}$$

Proof. For any k and n, with $k \leqslant n$, let $\bar{\boldsymbol{e}}_k^n = (\bar{e}_1, \dots, \bar{e}_n)$ denote a sequence such that $\bar{e}_1 = \cdots = \bar{e}_k = \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^k c_i}$ and $\bar{e}_{k+1} = \cdots = \bar{e}_n = 0$. It follows that $\mathrm{OWA}_{\boldsymbol{c}}(\bar{\boldsymbol{e}}_k^n) = 1$.

For any $k \leqslant n$, $l \leqslant m$, consider $\boldsymbol{u} = \bar{\boldsymbol{e}}_k^n$ and $\boldsymbol{v} = \bar{\boldsymbol{e}}_l^m$. As a necessary condition, we must have:

$$\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}, \boldsymbol{v}) \geqslant 1 = \min\{\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}), \mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{v})\}.$$

If $u_1 \geqslant v_1$, i.e., if $\sum_{i=1}^n c_i \sum_{i=1}^l c_i \geqslant \sum_{i=1}^m c_i \sum_{i=1}^k c_i$, the above is equivalent to stating that:

$$\begin{aligned}
\frac{\sum_{i=1}^k c_i u_i + \sum_{i=k+1}^{k+l} c_i v_{i-k}}{\sum_{i=1}^{n+m} c_i} &\geqslant 1 \iff \\
\sum_{i=1}^k c_i u_i + \sum_{i=k+1}^{k+l} c_i v_{i-k} &\geqslant \sum_{i=1}^{n+m} c_i \iff \\
\sum_{i=1}^k c_i \frac{\sum_{i=1}^n c_i}{\sum_{i=1}^k c_i} + \sum_{i=k+1}^{k+l} c_i \frac{\sum_{i=1}^m c_i}{\sum_{i=1}^l c_i} &\geqslant \sum_{i=1}^{n+m} c_i \iff \\
\sum_{i=1}^n c_i \sum_{i=1}^l c_i + \sum_{i=k+1}^{k+l} c_i \sum_{i=1}^m c_i &\geqslant \sum_{i=1}^{n+m} c_i \sum_{i=1}^l c_i \iff
\end{aligned}$$

$$\sum_{i=k+1}^{k+l} c_i \sum_{i=1}^{m} c_i \quad \geqslant \quad \sum_{i=n+1}^{n+m} c_i \sum_{i=1}^{l} c_i.$$

□

Particular cases of the above allow us to establish the following.

Corollary 12. For $c_1 = 1$, Eq. (7) necessarily implies that:

$$c_2 \geqslant c_3 \geqslant c_4 \geqslant \ldots \geqslant 0. \tag{9}$$

and:

$$\frac{\sum_{i=1}^{l} c_i}{\sum_{i=1}^{m} c_i} \leqslant \frac{\sum_{i=l+1}^{2l} c_i}{\sum_{i=m+1}^{2m} c_i}, \qquad l \leqslant m. \tag{10}$$

Equation 9 is obtained from the condition in Theorem 11 for $n \geqslant 2$, $l = m = 1$, $k = n - 1$, while Eq. 10 follows from the special case $k = l$ and $n = m$.

While these two conditions are necessary, the following example illustrates that they are not sufficient.

Example 13. The sequence $\boldsymbol{c} = (1, 1/2, 3/8, 3/8, 9/32, 9/32, 9/32, 9/32)$ satisfies the necessary conditions in Eq. (9) and Eq. (10), however fails to satisfy Eq. (7) for e.g., $\boldsymbol{u} = (1.875, 0, 0)$, $\boldsymbol{v} = (1.6875, 1.6875, 0, 0, 0)$. We have $\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}) = 1$, $\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{v}) = 1$, and $\mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}, \boldsymbol{v}) = 0.99306$.

Hence, we look towards establishing a sufficient condition, which we formulate as follows.

Theorem 14. Let $\boldsymbol{c} = (c_1, c_2, \ldots)$ be a coefficient vector with $c_1 = 1$ and $c_2 \geqslant c_3 \geqslant \ldots \geqslant 0$. Then a sufficient condition for (7) is:

$$\sum_{i=n+1}^{n+l} c_i \sum_{i=1}^{m} c_i \geqslant \sum_{i=n+1}^{n+m} c_i \sum_{i=1}^{l} c_i. \tag{11}$$

for any n, $l < m$ and $u_1, \ldots, u_n, v_1, \ldots, v_m$.

The proof is based on several lemmas presented below.

Lemma 15. Let $\mathbf{c}, \mathbf{d}$ be weighting vectors and $\mathbf{u} \in [0, \infty)^n$. Then $\mathrm{OWA}_{\mathbf{c}}(\mathbf{u}) \leqslant \mathrm{OWA}_{\mathbf{d}}(\mathbf{u})$ if and only if, for $k = 1, \ldots, n$, $\mathrm{OWA}_{\mathbf{c}}(\mathbf{e}_k) \leqslant \mathrm{OWA}_{\mathbf{d}}(\mathbf{e}_k)$, where $\mathbf{e}_k = (1, 1, \ldots, 1, 0, 0 \ldots, 0)$ is the vector with k ones and $n - k$ zeros.

Proof. It is sufficient to show this for $\mathbf{u} \in \{\mathbf{x} \in [0, 1] : x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n\}$ using symmetry and homogeneity of OWA functions. The graph of OWA on that subdomain is a fragment of a plane, and the inequalities at all the vertices are necessary and sufficient for one graph dominating the other. □

Note that the above relationship can also be stated in terms of the cumulative sum of OWA weights, and related to the idea of stochastic dominance, i.e., for $\sum_{i=1}^{n} c_i = \sum_{i=1}^{n} d_i$ it holds that:

$$\sum_{i=1}^{k} c_i \leqslant \sum_{i=1}^{k} d_i, \quad k = 1, \ldots, n.$$

Let us introduce notation $\mathbf{c}_{+k} = (c_{k+1}, c_{k+2}, \ldots, c_{k+n+1})$.

Lemma 16. If a weighting vector $\mathbf{c}$ satisfies:

$$\sum_{i=k+1}^{k+m} c_i \sum_{i=1}^{l} c_i \leqslant \sum_{i=1}^{m} c_i \sum_{i=k+1}^{k+l} c_i, \tag{12}$$

then $\mathrm{OWA}_{\mathbf{c}_{+0}}(\mathbf{u}) \leqslant \mathrm{OWA}_{\mathbf{c}_{+k}}(\mathbf{u})$ for all $\mathbf{u}$ and $k = 1, 2, \ldots$.

Sketch of the proof. Apply Lemma 15 with $\mathbf{c} = \mathbf{c}_{+0}$ and $\mathbf{d} = \mathbf{c}_{+l}$ and note that $\mathrm{OWA}_{\mathbf{c}_{+k}}(\mathbf{e}_l) = \sum_{i=k+1}^{k+l} c_i / \sum_{i=k+1}^{k+m} c_i$. The trivial cases of the sums in (12) being 0 are considered separately. □

Another piece of notation, let $\widetilde{\mathrm{OWA}}_{\mathbf{c}}(\mathbf{u}, \mathbf{v}) = \frac{1}{\sum_{i=1}^{n+m} c_i}(\sum_{i=1}^{n} c_i u_i + \sum_{i=n+1}^{n+m} c_i v_{i-n})$, whose difference to OWA is that no sorting step after concatenation takes place (it is assumed that $\mathbf{u}, \mathbf{v}$ are sorted separately).

Lemma 17. Let $\bar{\mathbf{u}} = (\bar{u}, \bar{u}, \ldots, \bar{u}) \in [0, \infty)^m$ and $\bar{u} = \mathrm{OWA}_{\mathbf{c}}(\mathbf{u})$. Then for any $m \geqslant 1$:

$$\widetilde{\mathrm{OWA}}_{\mathbf{c}}(\mathbf{u}, \bar{\mathbf{u}}) = \bar{u}. \tag{13}$$

Proof. By definition, $\widetilde{\mathrm{OWA}}_{\mathbf{c}}(\boldsymbol{u}, \bar{\boldsymbol{u}}) = \alpha \cdot \mathrm{OWA}_{\boldsymbol{c}}(\boldsymbol{u}) + (1-\alpha) \cdot \mathrm{OWA}_{\boldsymbol{c}_{+m}}(\boldsymbol{u})$, with $\alpha = \sum_{i=1}^{m} c_i / \sum_{i=1}^{m+n} c_i$. Hence, due to idempotency of the OWA and since we are taking a convex combination, we have:

$$\widetilde{\mathrm{OWA}}_{\mathbf{c}}(\boldsymbol{u}, \bar{\boldsymbol{u}}) = \alpha \cdot \bar{u} + (1-\alpha)\bar{u} = \bar{u}.$$

□

We can now formulate the proof of the above theorem.

Proof of Theorem 14. Assume without loss in generality that in (7), $\mathrm{OWA}_{\boldsymbol{c}}(\mathbf{u}) = \mathrm{OWA}_{\boldsymbol{c}}(\mathbf{v}) = \bar{u}$, as any increase in, say, $\mathbf{v}$ will lead to an increase on the left but not on the right, as well as $u_1 \geqslant v_1$. Then:

$$\begin{aligned} \bar{u} =& \widetilde{\mathrm{OWA}}_{\mathbf{c}}(\mathbf{u}, \bar{\mathbf{u}}) \quad \text{(Lemma 17)} \\ \leqslant& \widetilde{\mathrm{OWA}}_{\mathbf{c}}(\mathbf{u}, \mathbf{v}) \quad \text{(Lemma 16)} \\ \leqslant& \mathrm{OWA}_{\mathbf{c}}(\mathbf{u}, \mathbf{v}) \quad \text{(decreasing weights starting from the second one).} \end{aligned}$$

Note that v_1 will not end up in the first position in the sorted list $(\mathbf{u}, \mathbf{v})$, hence even if $c_1 < c_{n+1}$, it does not negate the last inequality. □

Note that, as long as there are no zeroes in the denominators, we can write our sufficient condition as:

$$\frac{\sum_{i=1}^{l} c_i}{\sum_{i=1}^{m} c_i} \leqslant \frac{\sum_{i=k+1}^{k+l} c_i}{\sum_{i=k+1}^{k+m} c_i} \tag{14}$$

for all $k > 1$, $m > l$. This amounts to all effective weighting vectors obtained from the first m arguments of $\boldsymbol{c}$ being stochastically dominated by all weighting vectors of the same length starting from a different index.

Note two differences to the necessary conditions in Theorem 11. First, there is a condition attached to (8). Second, we have $k < n$. Thus, the condition in (14) is unfortunately stronger than the necessary conditions previously established. We can weaken the sufficient conditions (14) by modifying Theorem 14 to:

$$\frac{\sum_{i=1}^{l} c_i}{\sum_{i=1}^{m} c_i} \leqslant \frac{c_k + \sum_{i=n+2}^{n+l} c_i}{\sum_{i=n+1}^{n+m} c_i}, \tag{15}$$

for some $k < n$ (owing to the fact that v_1 will be positioned ahead of some u_i, $i = k, \ldots, n$ in the sorted list $(\mathbf{u}, \mathbf{v})$).

Example 18. It is easily seen that $\boldsymbol{c} = (1, 0, 0, \dots)$ and $\boldsymbol{c} = (1, 1, 1, \dots)$, corresponding to complete and average linkage, respectively, fulfil the conditions in Theorem 14.

Example 19. All sequences of the form $\boldsymbol{c} = \boldsymbol{e}_k$ satisfy (14).

It is worthy to note that, in line with the condition that $\boldsymbol{c}$ need only be decreasing from c_2 onwards, sequences such as $(1, 2, 1, 1, 0, 0, \dots, 0)$ and $(1, 2, 2, 1, 0, \dots, 0)$ can be verified as satisfying the sufficient condition. One can also see that $(0, 1, 0, 0, \dots, 0)$ will satisfy (7). It can either be viewed separately from the framework that fixes $c_1 = 1$ or as a limiting case ($c_2 \to \infty$). In other words, general decreasingness is not a requirement on $\boldsymbol{c}$.

On the other hand, we observe that a fairly simple rule that will ensure satisfaction of (14) is for ratios between sequential values to be increasing, i.e., $c_i/c_{i+1} \leqslant c_{i+1}/c_{i+2}$. However, this is stronger than the sufficient condition. While Examples 18 and 19 above adhere to this rule, the following example shows that decreasing ratios also cannot be framed as a necessary requirement.

Example 20. A sequence that satisfies Theorem 14 is $\boldsymbol{c} = (1, 1/2, 1/5, 7/75, 0, 0, \dots)$, however this does not have an increasing sequence of ratios, i.e. we have $c_1/c_2 = 2, c_2/c_3 = 5/2$, however $c_3/c_4 = 15/7 < 5/2$.

Thus, the degree of the allowed tightening of the necessary and loosening of the sufficient conditions is still an open problem.

6. Conclusion and Future Work

OWA-based linkages were proposed in [20] and were reinvented in [17]. In [4], the practical usefulness of OWA-based clustering was evaluated thoroughly on numerous benchmark datasets from the suite described in [6]. It was noted that adding the Genie correction for cluster size inequality [7] leads to high-quality partitions, especially based on linkages that rely on a few closest point pairs (e.g., the single linkages and fuzzified/smoothened minimum). These papers provide many examples of practically useful OWA weight generators.

In this paper, we have presented some previously missing theoretical results concerning the OWA-based linkages. First, we have shown that the OWA-based linkages and the Lance–Williams formula only have three instances in common: the single, average, and complete linkages. Both classes enable a very fast (linear-time) update between iterations and thus are of potential practical interest.

Then, we gave some necessary and sufficient conditions for the coefficient generating schemes to guarantee the resulting dendrograms being free from unaesthetic inversions (note that the mentioned Genie correction might additionally introduce inversions by itself). How to tighten these condition sets in the form of "if and only if" statements is still an open problem: follow-up research is welcome.

In the future, we suggest considering similar results concerning different weighting triangle generating schemes, e.g., $w_{i,z} = w\left(\frac{i}{z}\right) - w\left(\frac{i-1}{z}\right)$, where $w : [0, 1] \to [0, 1]$ is a monotone bijection (compare [19]).

Furthermore, in [4], a generalised, three-stage OWA linkage scheme was introduced. There are also generalisations of the Lance–Williams formula, e.g., [8]. Inspecting the relationships between them could also be conveyed.

Acknowledgments

This research was supported by the Australian Research Council Discovery Project ARC DP210100227 (MG, GB, SJ).

Conflict of interest

All authors certify that they have no affiliations with nor involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

CRediT authorship contribution statement

Marek Gagolewski: Conceptualisation, Methodology, Formal analysis, Writing – Original Draft. **Anna Cena:** Formal analysis, Visualisation, Writing – Original Draft. **Simon James:** Formal analysis,Writing – Original Draft. **Gleb Beliakov:** Formal analysis,Writing – Original Draft.

References

[1] Beliakov, G., Bustince, H., Calvo, T., 2016. A practical guide to averaging functions. Springer.
[2] Beliakov, G., James, S., 2013. Stability of weighted penalty-based aggregation functions. Fuzzy Sets and Systems 226, 1–18.
[3] Calvo, T., Mayor, G., Torrens, J., Suner, J., Mas, M., Carbonell, M., 2000. Generation of weighting triangles associated with aggregation functions. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems 8, 417–451.
[4] Cena, A., Gagolewski, M., 2020. Genie+OWA: Robustifying hierarchical clustering with OWA-based linkages. Information Sciences 520, 324–336. doi:10.1016/j.ins.2020.02.025.
[5] Gagolewski, M., 2015. Data Fusion: Theory, Methods, and Applications. Institute of Computer Science, Polish Academy of Sciences, Warsaw. doi:10.5281/zenodo.6960306.
[6] Gagolewski, M., 2022. A framework for benchmarking clustering algorithms. SoftwareX 20, 101270. URL: https://clustering-benchmarks.gagolewski.com, doi:10.1016/j.softx.2022.101270.
[7] Gagolewski, M., Bartoszuk, M., Cena, A., 2016. Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. Information Sciences 363, 8–23. doi:10.1016/j.ins.2016.05.003.
[8] Gan, G., Ma, C., Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. ASA-SIAM Series on Statistics and Applied Probability, Philadelphia, Alexandria.
[9] Gomez, D., Rojas, K., Montero, J., Rodriguez, J., Beliakov, G., 2014. Consistency and stability in aggregation operators: An application to missing data problems. International Journal of Computational Intelligence Systems 7, 595–604.
[10] Grabisch, M., Marichal, J.L., Mesiar, R., Pap, E., 2009. Aggregation functions. Cambridge University Press.
[11] Jamison, B., Orey, S., Pruitt, W., 1965. Convergence of weighted averages of independent random variables. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 4, 40–44.
[12] Johnson, S.C., 1967. Hierarchical clustering schemes. Psychometrica 32, 241–254.
[13] Lance, G., Williams, W., 1967. A general theory of classification sorting strategies: 1. Hierarchical systems. Computer Journal , 373–380.
[14] Mayor, G., Calvo, T., 1997. On extended aggregation functions, in: Proc. IFSA 1997. Academia, Prague. volume 1, pp. 281–285.
[15] Milligan, G.W., 1979. Ultrametric hierarchical clustering algorithms. Psychometrika 44, 343–346.
[16] Müllner, D., 2011. Modern hierarchical, agglomerative clustering algorithms. ArXiv:1109.2378 [stat.ML] URL: http://arxiv.org/abs/1109.2378.
[17] Nasıbov, E., Kandemır-Cavas, C., 2011. OWA-based linkage method in hierarchical clustering: Application on phylogenetic trees. Expert Systems with Applications 38, 12684–12690.
[18] Wierzchoń, S., Kłopotek, M., 2018. Modern Algorithms for Cluster Analysis. Springer. doi:10.1007/978-3-319-69308-8.
[19] Yager, R.R., 1988. On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Transactions on Systems, Man, and Cybernetics 18, 183–190.
[20] Yager, R.R., 2000. Intelligent control of the hierarchical agglomerative clustering process. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 30, 835–845.