# Space Distortion and Monotone Admissibility in Agglomerative Clustering

Akinobu Takeuchi*, Hiroshi Yadohisa**, and Koichi Inada**

* College of Social Relations, Rikkyo (St. Paul's) University,
  Nishi-Ikebukuro 3-34-1, Tokyo 171-8501, JAPAN,
  *E-mail:* akitake@rikkyo.ac.jp

** Department of Mathematics and Computer Science, Kagoshima University,
  Korimoto 1-21-35, Kagoshima 890-0065, JAPAN,
  *E-mail:* (yado, inada)@sci.kagoshima-u.ac.jp

## Abstract

This paper discusses the admissibility of agglomerative hierarchical clustering algorithms with respect to space distortion and monotonicity, as defined by Yadohisa et al. and Batagelj, respectively. Several admissibilities and their properties are given for selecting a clustering algorithm. Necessary and sufficient conditions for an updating formula, as introduced by Lance and Williams, are provided for the proposed admissibility criteria. A detailed explanation of the admissibility of eight popular algorithms is also given.

Keywords and phrases: admissibility, AHCA (agglomerative hierarchical clustering algorithm), monotonicity, space distortion.

## 1.   Introduction

Agglomerative hierarchical clustering algorithms (AHCAs) were carefully studied by Lance and Williams (1967). If cluster $I$ and $J$ fuse to form a new cluster $I \cup J$, then Lance and Williams propose a dissimilarity between cluster $I \cup J$ and another cluster $K$ could be calculated by a recursive updating formula:

$$d_{(IJ)K} = \alpha_i d_{IK} + \alpha_j d_{JK} + \beta d_{IJ} + \gamma \left| d_{IK} - d_{JK} \right| \tag{1.1}$$

where parameters $\alpha_i, \alpha_j, \beta$, and $\gamma$ are either some constants or functions of the number of the object belonging to clusters $I, J$, and $K$; $d_{IJ}, d_{IK}$, and $d_{JK}$ are the dissimilarities between clusters $I$ and $J$, $I$ and $K$, and $J$ and $K$, respectively. The parameters characterize and determine the clustering algorithm. The myriad of available AHCAs can be defined by determining the parameters. This paper addresses the problem of choosing a clustering algorithm from available agglomerative algorithms. Many researchers have attacked this problem using the concept of admissibility. Fisher and Van Ness (1971) first introduced the concept of admissibility for clustering algorithms. They defined nine types of clustering algorithms admissibility and indicated the relationships between these particular admissibilities and some popular AHCAs. Space distortion was defined by Lance and Williams (1967) and this concept was converted into "Space-distorting admissibility" by Chan and Van Ness (1993, 1994b, and 1996). Space distortion is a well-known phenomenon in clustering and this is discussed by many researchers (e.g. Williams et al. (1966), Watson et al. (1966)). Mirkin (1996) proposed an admissibility with respect to the monotonicity of the updating formula. Mirkin (1996; p 249) discussed the characterization of monotone admissibility which gives the resulting cluster hierarchy a proper graphical representation. This property is usually considered as a necessary condition for an AHCA to be a good one. Mirkin also provided the relationship between monotone admissibility and the parameters in (1.1). In the papers cited above, space distortion and monotonicity depended *only* on the clustering algorithm that was used. On the other hand, we belive space distortion and monotonicity should be defined depending on *both* the algorithm selected *and* the data to be analyzed. Yadohisa et al. (1999) propose space distortion criteria, that they call the "distortion ratio at stage $m$" and the "total distortion

ratio", which are defined depending on both the algorithm selected and the data to be analyzed. Using these criteria, they also proposed several new space distortion admissibilities.

In this paper, we give the necessary and sufficient conditions between Yadohisa et al.'s admissibilities and the parameters in (1.1). In addition, we propose new admissibilities with respect to monotonicity, which are defined by Batagelj (1981). We also provide the necessary and sufficient conditions between the these admissibilities and the parameters in (1.1). By using these conditions, we may choose a proper clustering algorithm, in the sense of distortion and monotonicity.

We now define the notations used throughout this paper. Cluster $I$ at stage $m$ ($1 \leq m < N$) is denoted as $C_I(m)$. We denote the dissimilarity between objects $p$ and $q$ by $d_{pq}$, the dissimilarity between $C_I(m)$ and $C_J(m)$ by $d_{IJ}$, and the number of objects to be clustered by $N$. The notation $p \in C_I(m)$ indicates that object $p$ belongs to $C_I(m)$; the number of objects belonging to $C_I(m)$ is denoted by $n_I$.

When $C_T(m)$ and $C_K(m)$ are combined at stage $m$ and $C_T(m)$ is not a singleton, it is assumed that $C_T(m)$ was formed from $C_I(t)$ and $C_J(t)$, which were combined at stage $t$ ($1 < t < m$), and that $C_K(m)$ is a singleton or was formed from $C_{I'}(t')$ and $C_{J'}(t')$, which were combined at stage $t'$ ($1 \leq t' < t$). Hereafter, we assume this relationship between the two combined clusters, without loss of generality, and we assume $d_{IJ} < d_{IK} \leq d_{JK}$.

We abbreviate the single linkage algorithm as SL, the complete linkage algorithm as CL, the weighted average algorithm (WPGMA) as WA, the median algorithm (WPGMC) as MD, the group average algorithm (UPGMA) as GA, the centroid algorithm (UPGMC) as CE, the minimum variance algorithm (Ward's method) as WD, and the flexible algorithm with $\beta = -0.25$ as FX.

## 2. Space Distortion Ratio

Yadohisa et al. (1999) defined several new space distortion criteria, not just for the clustering algorithm but for each combination of objects. Each criterion is represented by a numerical value that allows its use in selecting a particular clustering algorithm in applications. In this section, we introduce the combined distance at stage $m$ and the distortion measures proposed by Yadohisa et al. (1999).

**Definition 2.1**: The distance between $C_T(m)$ and $C_K(m)$, which combine at stage $m$ ($1 \leq m \leq N-1$), is defined by the formula:

$$d_{TK}^m = \begin{cases} d_{pq} \ (p \in C_T(m), q \in C_K(m)), & (n_T = n_K = 1), \\ \alpha_i d_{IK} + \alpha_j d_{JK} + \beta d_{IJ} + \gamma |d_{IK} - d_{JK}|, & \text{(otherwise)}, \end{cases} \tag{2.1}$$

and is referred to as the "combined distance at stage $m$", where $\alpha_i, \alpha_j, \beta$, and $\gamma$ are either constants or functions in (1.1). We should note that clusters $C_T(m)$ and $C_K(m)$ have been uniquely determined by using (1.1) and assumptions in Section 1 before we calculate the $d_{TK}^m$. Hereafter a clustering algorithm for which a combined distance can be defined by (2.1) is called an AHCA.

Yadohisa et al. (1999) proposed two criteria for space distortion. The two criteria measure space distortion at each stage and in the entire clustering, respectively. Additionally, these criteria are defined depending on *both* the algorithm selected *and* the data to be analyzed.

**Definition 2.2**: If $C_T(m)$ and $C_K(m)$ combine at stage $m$ ($1 \leq m \leq N-1$), then the distortion ratio at stage $m$ is defined by:

$$R_m(d_{TK}^m) = \begin{cases} \kappa, & (n_T = n_K = 1), \\ \frac{\kappa(d_{TK}^m - d_{IJ}) - \eta(d_{TK}^m - d_{MK})}{d_{MK} - d_{IJ}}, & \text{(otherwise)}, \end{cases} \tag{2.2}$$

where $d_{MK} = (d_{IK} + d_{JK})/2$ and will be called the "standard distance". The values $\kappa$ and $\eta$ ($\kappa > \eta$) are different predetermined real constants or functions of $n_I$, $n_J$, and $n_K$. The distortion ratio measures the difference between the combined distance and the standard distance. If the combined distance equals the standard distance, then the distortion ratio is $\kappa$. If the combined distance coincides with $d_{IJ}$, which is the boundary of a monotone algorithm (see Section 4), then the distortion ratio is $\eta$. Both $\kappa$ and $\eta$ are determined by the judgment of the data analyst. Large values of $|\kappa - \eta|$ make the ratio sensitive to differences between the combined distance and the standard distance. Yadohisa et al. (1999) recommend that $\kappa = 1$

and $\eta = 0$. Then, we can judge the monotonicity from the magnitude positivity of (2.2) and can understand the distortion by comparing (2.2) with unity.

Another criterion for representing the total distortion in a whole clustering can be defined as follows.

**Definition 2.3**: Total distortion ratio is defined as:

$$TR_L = \frac{1}{N-L} \sum_{m=1}^{N-L} R_m(d_{TK}^m),\tag{2.3}$$

where $L$ is a number of clusters selected.

Here we note that the total distortion ratio means the overall average of the distortion ratio in the clustering processes.

# 3.    Space Distortion Admissibility

Yadohisa et al. (1999) proposed several new space distortions and admissibilities for clustering algorithms. In this section, we give the relationships between these admissibilities and the parameters in (1.1).

## 3.1    $\varepsilon$-Space Distortion Admissible

Here, we give a brief summary of $\varepsilon$-space distortion admissible proposed by Yadohisa et al. (1999). This approach is motivated by decision theory and has been studied by many researchers (Fisher and Van Ness, 1971; Van Ness, 1973; Chen and Van Ness, 1994a, 1994b). The admissibilities defined in this paper contain those considered by Chen and Van Ness (1993, 1994b) as a special case.

**Definition 3.1** (Yadohisa et al., 1999): For a given $\varepsilon > 0$ and a standard distance $d_{MK}$, a clustering algorithm is called $\varepsilon$-conserving admissible if the inequality:

$$|R_m(d_{TK}^m) - R_m(d_{MK})| < \varepsilon(R_m(d_{JK}) - R_m(d_{IK}))\tag{3.1}$$

holds for all $m$, except when two clusters are singletons that combine, or when $d_{IK} = d_{JK}$.

**Definition 3.2** (Yadohisa et al., 1999): For a given $\varepsilon > 0$ and a standard distance $d_{MK}$, a clustering algorithm is called $\varepsilon$-dilating admissible if the inequality:

$$R_m(d_{TK}^m) - R_m(d_{MK}) \geq \varepsilon(R_m(d_{JK}) - R_m(d_{IK}))\tag{3.2}$$

holds for all $m$, except when two clusters are singletons that combine, or when $d_{IK} = d_{JK}$.

**Definition 3.3** (Yadohisa et al., 1999): For a given $\varepsilon > 0$ and a standard distance $d_{MK}$, a clustering algorithm is called $\varepsilon$-contracting admissible if the inequality:

$$R_m(d_{MK}) - R_m(d_{TK}^m) \geq \varepsilon(R_m(d_{JK}) - R_m(d_{IK}))\tag{3.3}$$

holds for all $m$, except when two clusters are singletons that combine, or when $d_{IK} = d_{JK}$.

If $d_{MK} = (d_{IK} + d_{JK})/2$ and $\varepsilon = 1/2$, then $\varepsilon$-conserving admissibility, $\varepsilon$-dilating admissibility, and $\varepsilon$-contracting admissibility conform to the space-conserving admissibility proposed by Chen and Van Ness (1993), and to the space-dilating admissibility and the space-contracting admissibility proposed by Chen and Van Ness (1994b), respectively.

As well as these, Yadohisa et al. (1999) proposed further admissibilities in a whole clustering by using the total distortion ratio; the total distortion admissibilities of the space, $\varepsilon$-total space conserving admissible, $\varepsilon$-total space dilating admissible, and $\varepsilon$-total space contracting admissible. For details we refer to Yadohisa et al. (1999).

## 3.2 The Relationship between $\varepsilon$-Space Distortion Admissibility and Parameters of the Updating Formula

We give three theorems with respect to the relationship between $\varepsilon$-space distortion admissibility and the parameters in (1.1).

**Theorem 3.1**: When $d_{MK} = (d_{IK} + d_{JK})/2$, an AHCA is $\varepsilon$-space conserving admissible for any dataset if and only if the parameters in (1.1) satisfy the following three conditions:

$$\text{(i) } \frac{1}{2} - \varepsilon < \alpha_j + \gamma < \frac{1}{2} + \varepsilon, \text{ (ii) } \alpha_i + \alpha_j = 1, \text{ and (iii) } \alpha_i + \alpha_j + \beta = 1. \tag{3.4}$$

**Theorem 3.2**: When $d_{MK} = (d_{IK} + d_{JK})/2$, an AHCA is $\varepsilon$-space dilating admissible for any dataset if and only if the parameters in (1.1) satisfy the following three conditions:

$$\text{(i) } \alpha_j + \gamma \geq \frac{1}{2} + \varepsilon, \text{ (ii) } \alpha_i + \alpha_j \geq 1, \text{ and (iii) } \alpha_i + \alpha_j + \beta \geq 1. \tag{3.5}$$

**Theorem 3.3**: When $d_{MK} = (d_{IK} + d_{JK})/2$, an AHCA is $\varepsilon$-space contracting admissible for any dataset if and only if the parameters in (1.1) satisfy the following three conditions:

$$\text{(i) } \alpha_j + \gamma \leq \frac{1}{2} - \varepsilon, \text{ (ii) } \alpha_i + \alpha_j \leq 1, \text{ and (iii) } \alpha_i + \alpha_j + \beta \leq 1. \tag{3.6}$$

When the parameters in (1.1) do not depend on the number of clusters at the combined stage, we can determine whether the algorithm is $\varepsilon$-space distortion admissible before analyzing the data.

## 4.  Monotone Admissibility

Here, we propose new criteria with respect to "monotonicity", first discussed by Batagelj (1981). Batagelj's definition with respect to the monotonicity of the combined distance is as follows.

**Definition 4.1** (Batagelj, 1981): The AHCA is monotone if

$$d_{(IJ)K} \geq d_{IJ}$$

where $d_{(IJ)K}$ is the updating formula when $C_{I \cup J}(m)$ and $C_K(m)$ combine.

We transcribe Batagelj's definition using the distortion ratio as the following equation:

$$R_m(d_{TK}^m) \geq R_m(d_{IJ}) = \eta.$$

We can measure the degree of monotonicity of the combined distance at stage $m$ by using the difference between $R_m(d_{TK}^m)$ and $\eta$. Similarly, we can define monotonicity in a whole clustering using the total distortion ratio. In this section, we propose some concepts of monotonicity including Batagelj's monotonicity as a special case. Using these concepts, we define new admissibilities and also give the relationship between these admissibilities and Lance and Williams's updating formula.

## 4.1  $\delta$-Monotone Admissible

We define new criteria for the monotonicity of the combined distance by using the distortion ratio and the total distortion ratio. Additionally, we propose a new admissibility for the monotonicity of a clustering algorithm.

**Definition 4.2**: For a given $\delta(\in R)$ and a standard distance $d_{MK}$, the combined distance is $\delta$-monotone at stage $m$ if the inequality:

$$R_m(d_{IJ}) + \delta(R_m(d_{MK}) - R_m(d_{IJ})) \leq R_m(d_{TK}^m) \tag{4.1}$$

4

holds, except when two clusters are singletons that combine, or when $d_{IK} = d_{JK}$.

The $\delta$-monotonicity of the combined distance is judged in each merge stage by using the distortion ratio. Another criterion for representing $\delta$-monotonicity in the entire clustering can be defined as follows.

**Definition 4.3**: For a given $\delta(\in R)$ and a standard distance $d_{MK}$, the combined distance is $\delta$-total monotone in the entire clustering if the inequality:

$$\eta + \delta(\kappa - \eta) \leq TR_L \tag{4.2}$$

holds, where $L$ is the number of clusters selected.

From these definitions, we can define the admissibility of a clustering algorithm with respect to the monotonicity of the combined distance. Here we also note that the "total" means the "overall average" in the same manner as the total distortion ratio.

**Definition 4.4**: For a given $\delta(\in R)$ and a standard distance $d_{MK}$ if the inequality:

$$R_m(d_{IJ}) + \delta(R_m(d_{MK}) - R_m(d_{IJ})) \leq R_m(d_{TK}^m) \tag{4.3}$$

holds for all $m$, except when two clusters are singletons that combine or when $d_{IK} = d_{JK}$, then the clustering algorithm is called $\delta$-monotone admissible.

**Definition 4.5**: For a given $\delta(\in R)$ and a standard distance $d_{MK}$ if the inequality:

$$\eta + \delta(\kappa - \eta) \leq TR_L \tag{4.4}$$

holds, where $L$ is the number of clusters selected, then the clustering algorithm is called $\delta$-total monotone admissible.

These admissibilities include as a special case the monotone admissibility proposed by Mirkin (1996). The difference between two combined distances in neighboring stages is $d_{TK}^m - d_{T'K'}^{m-1}$, for which the $\delta$-monotone admissible clustering algorithm with large $\delta$ is relatively large when compared to those with a small $\delta$. A large value of $\delta$ may indicate the robustness of the result of a clustering. That is, a $\delta$-monotone admissible algorithm with a relatively large $\delta$ does not modify the result of the clustering, in spite of a small changes to the data.

Here, we describe some propositions for $\delta$-monotone and $\delta$-total monotone admissible.

**Proposition 4.1**: For a given $\delta \leq 1$, if an AHCA is $\delta$-monotone admissible, then the algorithm is $\delta$-total monotone admissible.

**Proposition 4.2**: For a given $\delta > 1$ and $\tau \leq 1$, if an AHCA is $\delta$-monotone admissible, then the algorithm is $\tau$-total monotone admissible.

**Proposition 4.3**: For a given $\delta < \delta'$, if an AHCA is $\delta'$-monotone admissible, then the algorithm is $\delta$-monotone admissible. Analogously, this proposition is satisfied for $\delta$-total monotone admissible algorithm.

**Proposition 4.4**: $\delta$-monotone admissible with $\delta = 0$ conforms to monotone admissible as proposed by Mirkin (1996).

**Proposition 4.5**: When $d_{MK} = (d_{IK} + d_{JK})/2$, FX is $\delta$-monotone admissible if the parameter $\beta$ of the algorithm satisfies $\delta \leq 1 - \beta$ and $\beta < 1$. The algorithm is $\delta$-total monotone admissible if $\beta$ satisfies $\delta \leq 1 - \beta$ and $0 \leq \beta < 1$.

## 4.2 The Relationship between $\delta$-Monotone Admissibility and the Parameters of the Updating Formula

We present a theorem with respect to the relationship between $\delta$-monotone admissibility and the parameters in (1.1).

**Theorem 4.1**: When $d_{MK} = (d_{IK} + d_{JK})/2$, an AHCA is $\delta$-monotone admissible for any dataset if and only if the parameters in (1.1) satisfy the following three conditions:

$$\text{(i) } \alpha_j + \gamma \geq \frac{1}{2}\delta, \text{ (ii) } \alpha_i + \alpha_j \geq \delta, \text{ and (iii) } \alpha_i + \alpha_j + \beta \geq 1. \tag{4.5}$$

# 5.  Data Dependence of the Admissibility of Clustering Algorithms

The concepts of space distortion and monotonicity were defined based on the updating formula in DuBien and Warde (1979) and Batagelj (1981), respectively. These concepts did not depend the data analyzed. On the other hand, Yadohisa, et al. (1999) pointed out that these concepts should be dependent on the data, since the phenomena of space distortion and monotonicity, which were discussed in Williams, et al. (1966) and Watson, et al. (1966), strongly depended on the data, even if only a single algorithm were used. Admissibilities proposed in this paper are based on extended space distortion ($\varepsilon$-space distortion) and extended monotonicity ($\delta$-monotone), which are defined as dependent on the data. In other words, these admissibilities inevitably depend on the data analyzed and can not be determined only by using an algorithm.

# 6.  Numerical Example

In this section, we analyze an artificial dataset in two-dimensional space using the algorithms listed in Table 1. This data, forming a so-called "touching cluster", is typical of the "chain" cluster described by Williams et al. (1966). The coordinates are listed in Table 2 and plotted in Figure 1.

Table 1: List of AHCAs

| Algorithm* | $\alpha_i$ | $\alpha_j$ | $\beta$ | $\gamma$ |
|---|---|---|---|---|
| SL | $1/2$ | $1/2$ | $0$ | $-1/2$ |
| CL | $1/2$ | $1/2$ | $0$ | $1/2$ |
| GA | $n_I/(n_I + n_J)$ | $n_J/(n_I + n_J)$ | $0$ | $0$ |
| WA | $1/2$ | $1/2$ | $0$ | $0$ |
| CE | $n_I/(n_I + n_J)$ | $n_J/(n_I + n_J)$ | $-n_I n_J/(n_I + n_J)^2$ | $0$ |
| MD | $1/2$ | $1/2$ | $-1/4$ | $0$ |
| WD | $(n_I + n_K)/(n_I + n_J + n_K)$ | $(n_J + n_K)/(n_I + n_J + n_K)$ | $-n_K/(n_I + n_J + n_K)$ | $0$ |
| FX | $(1 - \beta)/2$ | $(1 - \beta)/2$ | $\beta < 1$ | $0$ |

*Reference for each method can be found in Cormack (1971).

Table 2: Coordinates of 17 objects in two-dimensional space

| No | Coordinates | No | Coordinates | No | Coordinates |
|---|---|---|---|---|---|
| 1 | 3.999, 10.007 | 7 | 7.499,  7.502 | 13 | 13.502, 11.003 |
| 2 | 5.005,  8.502 | 8 | 7.800, 10.500 | 14 | 14.002,  8.000 |
| 3 | 4.999, 12.000 | 9 | 8.997, 11.000 | 15 | 13.997, 12.997 |
| 4 | 6.003,  7.002 | 10 | 9.999, 10.000 | 16 | 16.004,  9.998 |
| 5 | 6.998,  9.793 | 11 | 11.997,  9.002 | 17 | 17.004, 11.998 |
| 6 | 7.302, 12.502 | 12 | 12.005, 12.003 | | |

The distortion ratio and the total distortion ratio are both calculated with $\kappa = 1$, $\eta = 0$, and $d_{MK} = (d_{IK} + d_{JK})/2$ in Table 3. The $\gamma_m$ statistic at stage $m$ developed by Goodman and Kruskal (1954) is also calculated. The notation 11 – 12 at stage 12 of CL in Table 3 means that cluster 11, which contains object 14, and cluster 12, which contains objects 13 and 15, are combined at this stage. $TR_2$, $TG$, and C. corr. in Table 3 denote the total distortion ratio with $L = 2$, the average of the Goodman and Kruskal's statistic (1954), and the cophenetic correlation of Sokal and Rohlf (1962), respectively. As $L$ is the number of clusters selected, we consider that the $TR_L$ represents a distortion of the whole clustering. In table 3, we calculated $TR_2$ as an example.

By comparing the value of the distortion ratio with zero, we can ascertain whether the algorithm is monotone. For a more accurate assessment, we should use the concepts of $\varepsilon$-space distortion and the $\delta$-monotonicity of the combined distance at a given stage. We can use the same approach to assess the total
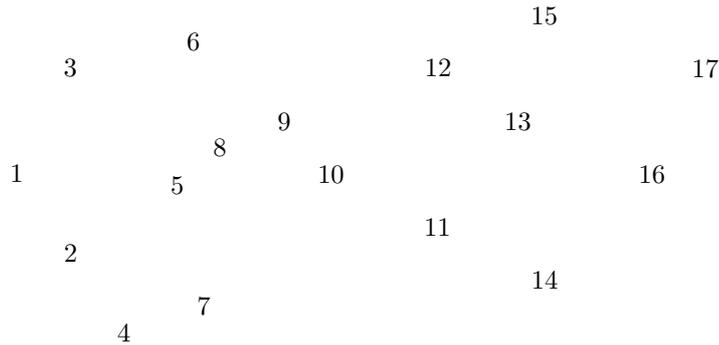
Figure 1: Scatter diagram of two-dimensional space

distortion ratio. Since we use the combined distance of WA as the standard distance, the ratio at all stages and the total distortion ratio by WA are equal to unity in Table 3.

The $\varepsilon$-space distortion admissibilities for the cases of $\varepsilon = 1/2, 1/3$, and $1/4$ and the $\delta$-monotone admissibilities for the cases of $\delta = 1, 0$ and $-1$ are indicated in Tables 4 and 5, respectively. Generally, from the definitions, $\varepsilon$-space distortion and $\delta$-monotone admissibilities are sensitive concepts in contrast to the total admissibilities (see, Table 4 and 5). By changing the value of $\varepsilon$ and $\delta$ we can select an algorithm which satisfies $\varepsilon$-(total) space distortion admissibilities and/or $\delta$-(total) monotone admissibility according to need.

## 7.   Concluding Remarks

Many research fields, including psychology and sociology, use AHCAs. However, we belive that the characteristics of these AHCAs with respect to space distortion and monotonicity have not yet been sufficiently investigated. In this paper, we provide the necessary and sufficient conditions of the admissibilities proposed by Yadohisa et al. (1999) for Lance and Williams's updating formula. By comparing the relationship among the parameters, we can select AHCAs adapted both to the data being analyzed and to the expected space distortion. Moreover, we extend the concept of monotonicity and provide several theorems concerning the relationship between the new concepts and the AHCAs. This is useful not only for selecting an algorithm, but also for assessing the robustness of the result of a clustering from a slight change in the data analyzed. Knowing the general properties of AHCAs is important as criteria for selecting an algorithm from AHCAs. The results given here should considerably reduce time and labor required in choosing a clustering algorithm from the profusion of algorithms. Finally, we would like to note that the admissibilities defined in this paper can be extended by changing the definitions of $d_{TK}^m$. For example, we may extend such admissibilities by using the updating formula of Jambu (1978) instead of Lance and Williams's formula in the definition of $d_{TK}^m$.

## References

[1] Batagelj, V. (1981). Note on ultrametric hierarchical clustering algorithms, *Psychometrika*, **46**, 351–353.

[2] Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society, (Series A)*, **38**, 47–62.

[3] Chen, Z. & Van Ness, J. (1993). *Space-conserving agglomerative algorithms*. Technical Report #224 Mathematical Sciences, University of Texas.

[4] Chen, Z. & Van Ness, J. (1994a). Metric admissibility and agglomerative clustering. *Communication in Statistics: Simulation & Computation*, **23**, 833–845.

[5] Chen, Z. & Van Ness, J. (1994b). Space-contracting, space-dilating, and positive admissible clustering algorithms. *Pattern Recognition*, **27**, 853–857.

Table 3: Distortion ratios and total distortion ratios of 8 AHCAs

| | SL | | | CL | | | WA | | | MD | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stage | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ |
| 1 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 |
| 2 | $5-9$ | 0.96 | 0.31 | $9-10$ | 1.00 | 1.00 | $9-10$ | 1.00 | 1.00 | $9-10$ | 1.00 | 1.00 |
| 3 | $5-10$ | 0.91 | 0.22 | $4-7$ | 1.00 | 1.00 | $4-7$ | 1.00 | 1.00 | $4-7$ | 1.00 | 1.00 |
| 4 | $4-7$ | 0.91 | 1.00 | $12-13$ | 1.00 | 1.00 | $12-13$ | 1.00 | 1.00 | $5-9$ | 0.91 | 0.35 |
| 5 | $12-13$ | 0.92 | 1.00 | $1-2$ | 1.00 | 1.00 | $1-2$ | 1.00 | 1.00 | $12-13$ | 0.92 | 1.00 |
| 6 | $2-4$ | 0.90 | 0.34 | $12-15$ | 0.99 | 1.25 | $12-15$ | 0.99 | 1.00 | $12-15$ | 0.91 | $-0.32$ |
| 7 | $1-2$ | 0.77 | 0.01 | $16-17$ | 0.97 | 1.00 | $5-9$ | 0.92 | 1.00 | $1-2$ | 0.92 | 1.00 |
| 8 | $12-15$ | 0.78 | 0.75 | $11-14$ | 0.95 | 1.00 | $16-17$ | 0.92 | 1.00 | $5-6$ | 0.87 | 0.48 |
| 9 | $5-6$ | 0.76 | 0.44 | $3-6$ | 0.89 | 1.00 | $11-14$ | 0.92 | 1.00 | $16-17$ | 0.88 | 1.00 |
| 10 | $1-3$ | 0.68 | 0.40 | $5-9$ | 0.93 | 1.27 | $3-6$ | 0.93 | 1.00 | $11-14$ | 0.89 | 1.00 |
| 11 | $5-11$ | 0.65 | 0.09 | $1-4$ | 0.80 | 1.48 | $1-4$ | 0.80 | 1.00 | $1-4$ | 0.80 | 0.50 |
| 12 | $16-17$ | 0.65 | 1.00 | $11-12$ | 0.70 | 1.11 | $3-5$ | 0.73 | 1.00 | $3-5$ | 0.73 | 0.39 |
| 13 | $5-14$ | 0.49 | 0.01 | $3-5$ | 0.72 | 1.87 | $12-16$ | 0.66 | 1.00 | $1-3$ | 0.59 | $-0.15$ |
| 14 | $1-5$ | 0.44 | 0.05 | $11-16$ | 0.71 | 2.00 | $1-3$ | 0.63 | 1.00 | $12-16$ | 0.63 | 0.30 |
| 15 | $1-12$ | 0.52 | 0.08 | $1-3$ | 0.86 | 1.40 | $11-12$ | 0.86 | 1.00 | $11-12$ | 0.86 | $-0.04$ |
| $TR_2$ | 0.445 | | | 1.224 | | | 1.000 | | | 0.566 | | |
| $TG$ | 0.771 | | | 0.908 | | | 0.897 | | | 0.869 | | |
| C.corr. | 0.572 | | | 0.755 | | | 0.757 | | | 0.745 | | |
| | GA | | | CE | | | WD | | | FX $(\beta=-0.25)$ | | |
| Stage | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ | Combine | $\gamma_m$ | $R_m$ |
| 1 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 | $5-8$ | 1.00 | 1.00 |
| 2 | $9-10$ | 1.00 | 1.00 | $9-10$ | 1.00 | 1.00 | $9-10$ | 1.00 | 1.00 | $9-10$ | 1.00 | 1.00 |
| 3 | $4-7$ | 1.00 | 1.00 | $4-7$ | 1.00 | 1.00 | $4-7$ | 1.00 | 1.00 | $4-7$ | 1.00 | 1.00 |
| 4 | $12-13$ | 1.00 | 1.00 | $5-9$ | 0.91 | 0.35 | $12-13$ | 1.00 | 1.00 | $12-13$ | 1.00 | 1.00 |
| 5 | $1-2$ | 1.00 | 1.00 | $12-13$ | 0.92 | 1.00 | $1-2$ | 1.00 | 1.00 | $1-2$ | 1.00 | 1.00 |
| 6 | $12-15$ | 0.99 | 1.00 | $12-15$ | 0.91 | $-0.32$ | $16-17$ | 0.97 | 1.00 | $12-15$ | 0.99 | 1.25 |
| 7 | $5-9$ | 0.92 | 1.00 | $1-2$ | 0.92 | 1.00 | $11-14$ | 0.92 | 1.00 | $16-17$ | 0.97 | 1.00 |
| 8 | $16-17$ | 0.92 | 1.00 | $5-6$ | 0.87 | 0.48 | $12-15$ | 0.95 | 1.33 | $11-14$ | 0.95 | 1.00 |
| 9 | $11-14$ | 0.92 | 1.00 | $16-17$ | 0.88 | 1.00 | $3-6$ | 0.89 | 1.00 | $3-6$ | 0.89 | 1.00 |
| 10 | $3-6$ | 0.93 | 1.00 | $11-14$ | 0.89 | 1.00 | $5-9$ | 0.93 | 1.50 | $5-9$ | 0.93 | 1.25 |
| 11 | $1-4$ | 0.80 | 1.00 | $1-4$ | 0.80 | 0.50 | $1-4$ | 0.80 | 1.50 | $1-4$ | 0.80 | 1.25 |
| 12 | $3-5$ | 0.73 | 1.00 | $1-5$ | 0.37 | 0.01 | $3-5$ | 0.73 | 1.33 | $3-5$ | 0.73 | 1.25 |
| 13 | $11-12$ | 0.72 | 1.00 | $1-3$ | 0.59 | 0.04 | $11-12$ | 0.72 | 1.36 | $12-16$ | 0.66 | 1.25 |
| 14 | $11-16$ | 0.71 | 0.81 | $11-12$ | 0.74 | 0.33 | $11-16$ | 0.71 | 1.16 | $11-12$ | 0.71 | 1.25 |
| 15 | $1-3$ | 0.86 | 0.91 | $11-16$ | 0.86 | $-0.64$ | $1-3$ | 0.86 | 1.45 | $1-3$ | 0.86 | 1.25 |
| $TR_2$ | 0.981 | | | 0.516 | | | 1.176 | | | 1.117 | | |
| $TG$ | 0.906 | | | 0.854 | | | 0.905 | | | 0.906 | | |
| C.corr. | 0.757 | | | 0.747 | | | 0.746 | | | 0.751 | | |

Note: $TG = (\sum_{m=1}^{15} \gamma_m)/15$

Table 4: Space distortion admissibilities of 8 AHCAs: Determined after analysis

| admissible | SL | CL | WA | MD | GA | CE | WD | FX |
|---|---|---|---|---|---|---|---|---|
| 1/2 conserving | No | No | Yes* | No | Yes* | No | No | No |
| 1/2 dilating | No | Yes* | No | No | No | No | No | No |
| 1/2 contracting | Yes* | No | No | No | No | No | No | No |
| 1/3 conserving | No | No | Yes* | No | Yes | No | No | No |
| 1/3 dilating | No | Yes* | No | No | No | No | No | No |
| 1/3 contracting | Yes* | No | No | No | No | No | No | No |
| 1/4 conserving | No | No | Yes* | No | Yes | No | No | No |
| 1/4 dilating | No | Yes* | No | No | No | No | No | No |
| 1/4 contracting | Yes* | No | No | Yes | No | Yes | No | No |
| 1/2 total conserving | No | No | Yes* | No | Yes* | Yes | No | No |
| 1/2 total dilating | No | Yes* | No | No | No | No | Yes | Yes |
| 1/2 total contracting | Yes* | No | No | Yes | No | No | No | No |
| 1/3 total conserving | No | No | Yes* | No | Yes | No | No | No |
| 1/3 total dilating | No | Yes* | No | No | No | No | Yes | Yes |
| 1/3 total contracting | Yes* | No | No | Yes | No | Yes | No | No |
| 1/4 total conserving | No | No | Yes* | No | Yes | No | No | No |
| 1/4 total dilating | No | Yes* | No | No | No | No | Yes | Yes |
| 1/4 total contracting | Yes* | No | No | Yes | No | Yes | No | No |

* Theoretically determined

Table 5: $\delta$-monotone admissibilities of 8 AHCAs: Determined after analysis

| admissible | SL | CL | WA | MD | GA | CE | WD | FX |
|---|---|---|---|---|---|---|---|---|
| 1 monotone | No | Yes* | Yes* | No | No | No | Yes | Yes |
| 0 monotone | Yes* | Yes* | Yes* | No | Yes* | No | Yes* | Yes* |
| −1 monotone | Yes* | Yes* | Yes* | Yes | Yes* | Yes | Yes* | Yes* |
| 1 total monotone | No | Yes* | Yes* | No | No | No | Yes | Yes |
| 0 total monotone | Yes* | Yes* | Yes* | No | Yes* | Yes | Yes* | Yes* |
| −1 total monotone | Yes* | Yes* | Yes* | Yes | Yes* | Yes | Yes* | Yes* |

* Theoretically determined

9

[6] Chen, Z. & Van Ness, J. (1996). Space-conserving agglomerative algorithms. *Journal of Classification*, **13**, 157–168.

[7] DuBien, J. L & Warde, W. D. (1979). A mathematical comparison of the members of an infinite family of agglomerative clustering algorithms, *Canadian Journal of Statistics*, **7**, 29–38.

[8] Fisher, L. & Van Ness, J. (1971). Admissible cluster procedures. *Biometrika*, **58**, 91–104.

[9] Goodman, L. A. & Kruskal, W. H. (1954). Measures of association for cross-classifications, *Journal of the American Statistical Association*, **49**, 732–764.

[10] Jambu, M. (1978). *Classification Automatique pour l'Analyse des Données*, North-Holland, Amsterdam.

[11] Lance, G. N. & Williams, W. T. (1967). A general theory of classificatory sorting strategies, 1. hierarchical systems. *The Computer Journal*, **9**, 373–380.

[12] Mirkin, B. (1996). *Mathematical Classification and Clustering.* London: Kluwer Academic Publishers.

[13] Sokal, R. R. & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. *Taxon*, **11**, 33–40.

[14] Van Ness, J. (1973). Admissible cluster procedures II. *Biometrika*, **60**, 422–424.

[15] Watson, L., Williams, W. T., & Lance, G. N. (1966). Angiosperm taxonomy: A comparative study of some novel numerical techniques, *Journal Linnaean Society*, **59**, 491–501.

[16] Williams, W. T., Lambert, J. M., & Lance, G. N. (1966). Multivariate methods in plant ecology, V. *Journal Ecology*, **54**, 427–445.

[17] Yadohisa, H., Takeuchi, A., & Inada, K. (1999). Developing criteria for measuring space distortion in combinatorial cluster analysis and methods for controlling the distortion. *Journal of Classification*, **16**, 45–62.

# Appendix

**Proof of Theorem 3.1**:

Here we provide notation to help prove the theorems. Note that $d_{IK} \neq d_{JK}$ and $d_{MK} = (d_{IK} + d_{JK})/2$ will be assumed in the theorems. From these assumptions, we denote

$$d_{IK} = d_{IJ} + \Delta \quad \text{and} \quad d_{JK} = d_{IK} + \Delta'$$

where $\Delta$ and $\Delta'$ are positive numbers that are determined at each combining stage. We denote

$$A = \varepsilon(d_{JK} - d_{IK}) - (d_{TK}^m - d_{MK}), \quad B = \varepsilon(d_{JK} - d_{IK}) - (d_{MK} - d_{TK}^m).$$

**Lemma A**: If $A > 0$ and $B > 0$ hold for all $m$, the clustering algorithm is $\varepsilon$-space conserving admissible.

*Proof.* We assume that $A > 0$ and $B > 0$ hold for all $m$, then,

$$\varepsilon(d_{JK} - d_{IK}) > d_{TK}^m - d_{MK} \tag{A1}$$

and

$$\varepsilon(d_{JK} - d_{IK}) > d_{MK} - d_{TK}^m \tag{A2}$$

are satisfied for all $m$. Due to the monotonicity of the distortion ratio, if $d_{TK}^m \geq d_{MK}$, then $R_m(d_{TK}^m) \geq R_m(d_{MK})$, and hence,

$$R_m(\varepsilon(d_{JK} - d_{IK})) > R_m(d_{TK}^m - d_{MK}) \tag{A3}$$

and

$$R_m(\varepsilon(d_{JK} - d_{IK})) > R_m(d_{MK} - d_{TK}^m). \tag{A4}$$

From the definition of the distortion ratio, we obtain

$$R_m(d_{TK}^m + d_{TK}^{m'}) = R_m(d_{TK}^m) + R_m(d_{TK}^{m'}) + \omega, \tag{A5}$$

$$R_m(k d_{TK}^m) = k R_m(d_{TK}^m) + (k-1)\omega \quad (k \in R), \tag{A6}$$

and then

$$R_m(d_{TK}^m - d_{TK}^{m'}) = R_m(d_{TK}^m) - R_m(d_{TK}^{m'}) - \omega, \tag{A7}$$

where

$$\omega = \frac{\kappa d_{IJ} - \eta d_{MK}}{d_{MK} - d_{IJ}}.$$

Now, we rewrite $R_m(\varepsilon(d_{JK} - d_{IK})), R_m(d_{TK}^m - d_{MK})$, and $R_m(d_{MK} - d_{TK}^m)$ by (A6) and (A7) as follows;

$$R_m(\varepsilon(d_{JK} - d_{IK})) = \varepsilon(R_m(d_{JK}) - R_m(d_{IK})) - \omega, \tag{A8}$$

$$R_m(d_{TK}^m - d_{MK}) = R_m(d_{TK}^m) - R_m(d_{MK}) - \omega, \tag{A9}$$

and

$$R_m(d_{TK}^m - d_{MK}) = R_m(d_{TK}^m) - R_m(d_{MK}) - \omega.$$

Thus, we get

$$\varepsilon(R_m(d_{JK}) - R_m(d_{IK})) > R_m(d_{TK}^m) - R_m(d_{MK})$$

and

$$\varepsilon(R_m(d_{JK}) - R_m(d_{IK})) > R_m(d_{MK}) - R_m(d_{TK}^m)$$

from (A1) and (A2), respectively. Hence, if $A > 0$ and $B > 0$ hold for all $m$, then the clustering algorithm is $\varepsilon$-space conserving admissible. $\square$

First, we consider the case where the parameters of an algorithm satisfy conditions (i), (ii), and (iii). Then we get $A > 0$ and $B > 0$ from

$$
\begin{aligned}
A &= \varepsilon(d_{JK} - d_{IK}) - (d_{TK}^m - d_{MK}) \\
&= (-\varepsilon - \alpha_i + \gamma + \frac{1}{2})d_{IK} + (\varepsilon - \alpha_j - \gamma + \frac{1}{2})d_{JK} - \beta d_{IJ} \\
&= (-\alpha_i - \alpha_j + 1 - \beta)d_{IJ} + (-\alpha_i - \alpha_j + 1)\Delta + (\varepsilon - \alpha_j - \gamma + \frac{1}{2})\Delta', \\
B &= \varepsilon(d_{JK} - d_{IK}) - (d_{MK} - d_{TK}^m) \\
&= (-\varepsilon + \alpha_i - \gamma - \frac{1}{2})d_{IK} + (\varepsilon + \alpha_j + \gamma - \frac{1}{2})d_{JK} + \beta d_{IJ} \\
&= (\alpha_i + \alpha_j - 1 + \beta)d_{IJ} + (\alpha_i + \alpha_j - 1)\Delta + (\varepsilon + \alpha_j + \gamma - \frac{1}{2})\Delta'.
\end{aligned}
$$

Therefore, from Lemma A, the algorithm is $\varepsilon$-space conserving admissible.

Conversely, we consider the case where an algorithm is $\varepsilon$-space conserving admissible. First, we assume that (i) and (ii) are satisfied, but (iii) is not. That is,

$$1/2 - \varepsilon < \alpha_j + \gamma < 1/2 + \varepsilon, \alpha_i + \alpha_j = 1, \alpha_i + \alpha_j + \beta > 1$$

or

$$1/2 - \varepsilon < \alpha_j + \gamma < 1/2 + \varepsilon, \alpha_i + \alpha_j = 1, \alpha_i + \alpha_j + \beta < 1$$

holds. For the first case, we assume $d_{TK}^m \geq d_{MK}$ and choose a $\Delta'$ small. Since the coefficients of the first and second terms of $A$ are nonpositive and we choose a $\Delta'$ small, $A < 0$. This contradicts the assumption the fact that the algorithm is $\varepsilon$-space conserving admissible. For the second case, we also obtain a contradiction in the same manner when we assume $d_{TK}^m < d_{MK}$ and choose a $\Delta'$ small.

In the same way, we can obtain contradictions for cases when at least one of the conditions (i), (ii), and (iii) does not hold. In Table 6, we summarize all possible cases that result in a contradiction. Therefore, if an algorithm is $\varepsilon$-space conserving admissible then conditions (i), (ii), and (iii) hold by reduction to absurdity.
$\square$

Theorem 3.2, 3.3, and 4.1 are proved similarly.

Table 6: Cases of contradiction for $\varepsilon$-space conserving admissible

| (i) or not (i) | (ii) or not (ii) | (iii) or not (iii) | case | contradiction |
|---|---|---|---|---|
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 = 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}, \Delta' \to 0$ | $A < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 = 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}, \Delta' \to 0$ | $B < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 > 1$ | $c_3 = 1$ | $d_{TK}^m \geq d_{MK}, \Delta' \to 0$ | $A < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 < 1$ | $c_3 = 1$ | $d_{TK}^m < d_{MK}, \Delta' \to 0$ | $B < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 > 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}, \Delta' \to 0$ | $A < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 > 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}, \Delta \to 0, \Delta' \to 0$ | $B < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 < 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}, \Delta \to 0, \Delta' \to 0$ | $A < 0$ |
| $1/2 - \varepsilon < c_1 < 1/2 + \varepsilon$ | $c_2 < 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}, \Delta' \to 0$ | $B < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 = 1$ | $c_3 = 1$ | $d_{TK}^m < d_{MK}$ | $B < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 = 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}$ | $A < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 = 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}$ | $B < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 > 1$ | $c_3 = 1$ | $d_{TK}^m \geq d_{MK}$ | $A < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 < 1$ | $c_3 = 1$ | $d_{TK}^m < d_{MK}$ | $B < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 > 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}$ | $A < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 > 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}, \Delta \to 0$ | $B < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 < 1$ | $c_3 > 1$ | $d_{TK}^m \geq d_{MK}, \Delta \to 0$ | $A < 0$ |
| $1/2 + \varepsilon \leq c_1$ or $c_1 \leq 1/2 - \varepsilon$ | $c_2 < 1$ | $c_3 < 1$ | $d_{TK}^m < d_{MK}$ | $B < 0$ |

Note: $c_1 = \alpha_j + \gamma, c_2 = \alpha_i + \alpha_j, c_3 = \alpha_i + \alpha_j + \beta$; $\Delta \to 0$ stands for choosing small $\Delta$