A Mathematical Model for Classification and Identification

D. Leuschner

Dresden, Germany

Abstract: A unified model for classification and identification is presented by means of the theory of mathematical structures. The methodologies for classification and identification differ. For classification, an interchange scalar mechanism (ISCL) is proposed which improves the capacity of classifications irrespective of character weighting, the number of characters used, and the form of the clustering algorithm. It modifies a distance between two OTUs by considering their distances to their nearest neighbors. Formulas are developed for reliability, velocity, and capability of taxonomic keys.

Keywords: Classification; Identification; Interchange scalar classifications; Numerical taxonomy; Taxonomic keys.

1. Introduction

The investigation of theoretical and practical problems in numerical taxonomy (Sneath and Sokal 1973) by means of mathematical (algebraical, analytical, and numerical) models leads to a conceptual framework in which both classification and identification may be performed. I use the term numerical taxonomy (NT) as an abbreviated term for mathematical models in taxonomy. The elementary taxa (e.g., varieties, species, genera) $t_1, t_2, \ldots, t_N$ represent a set $T: t_i \in T$ where N is the number of taxa. These

taxa define a space in which it is possible to perform both classification and identification. It is the purpose of this paper to provide effective and comprehensive examples of classification and identification by applying the theory of mathematical structures (Leuschner 1985).

2. Numerical Taxonomic Methods

In this section I furnish concise but sufficient definitions of some taxonomic terms. Taxonomic methodology is concerned with reducing the number of subsets in a given *taxonomic set*. The methods of numerical taxonomy are applications of suitable lattice reduction algorithms. Reduction of the lattice in Figure 1 to the classification in Figure 2 is possible by calculating a distance $d(p_1 \, p_2)$ and applying an algorithm to reduce the number of connections and points in the *set*. The exact definition of the taxonomic set and a number of useful axioms concerning it are given in Appendix 1. All classifications considered here will be hierarchic with nested sets of taxonomic units.

These axioms in the Appendix 1 are necessary but not sufficient criteria by which to create classifications or taxonomic keys for identifying individual taxonomic units. Elementary taxa (synonymous with operational taxonomic units; Sneath and Sokal 1973) possess empirical descriptors, characters, and a number of questions concerning these come to mind. Which characters are we to choose? How are we to measure distances from these characters? What is the appropriate reduction algorithm?

To answer these questions, I shall use a utilitarian approach. The elementary taxa t_i, $i = 1,2,\ldots,N$, possess the characters m_j, $j = 1,2,\ldots,M$. The characters m_j assume values $m_{jl(j)}$ where $l(j)$ denotes the vector of states of character m_j. When $l(j)$ may assume an infinite number of values, it is termed a continuous character and when it may assume a finite number of values, it is termed discrete (see Leuschner 1974). The number of values of $l(j)$ is L_j or L, neglecting the index of character j.

Define a character space C_M of dimension M. Each axis is defined by a character and the axes are orthogonal. Points in the character space represent elementary taxa. In this space every taxon possesses M coordinates. A two-dimensional character space is shown in Figure 3, where m_1 is continuous and m_2 is discrete. A classification of the taxa (Figure 4) and a structure of a key to the four taxa (Figure 5) are not the same, i.e., while taxonomic classifications and taxonomic keys may be isomorphic, they need not be so, even though they have the same underlying number of taxa N. One obvious difference is that some of the terminal points of a key may be empty sets. Thus, the methodologies for classification and identification are different: their distance measures and lattice reduction algorithms should be
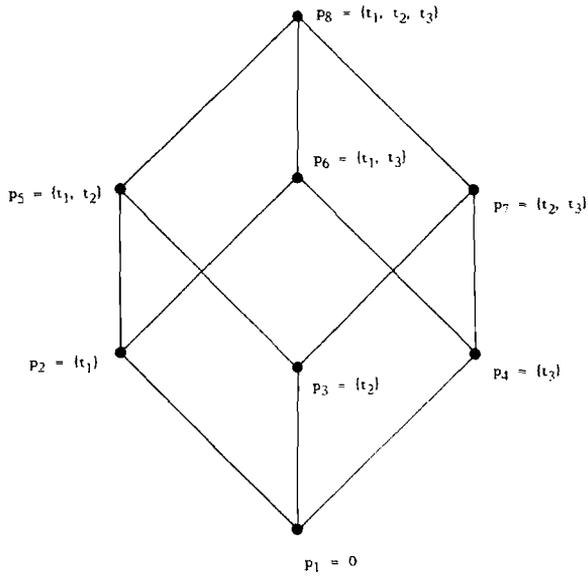
Figure 1. Representation of a taxonomic space $P(T)$, with a set T of three taxa, $T = \{t_1, t_2, t_3\}$ and possible links between the elements of T.
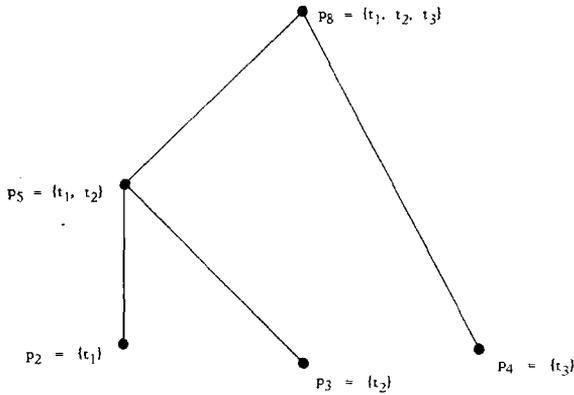
Figure 2. An example of a taxonomic lattice (dendrogram) derived from Figure 1. A lattice reduction algorithm will omit links (p_i) to form the lattice.
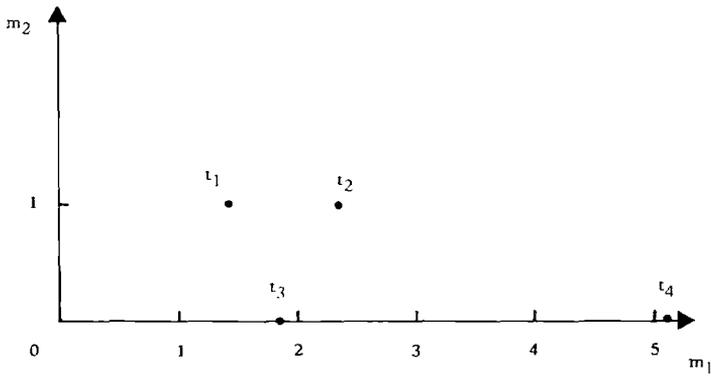
Figure 3. A two-dimensional character space with four elementary taxa (t_1 to t_4). Character m_1 is continuous and character m_2 is discrete.
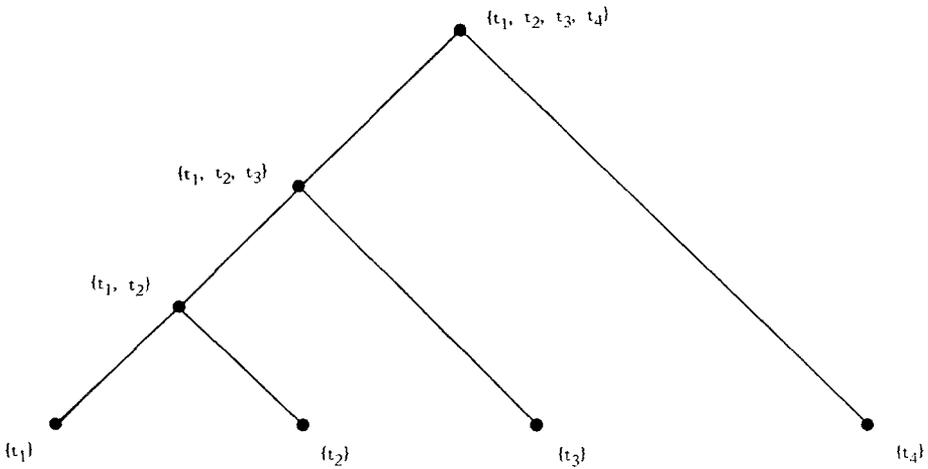
Figure 4. A possible taxonomic classification for the 4 taxa in Figure 3.
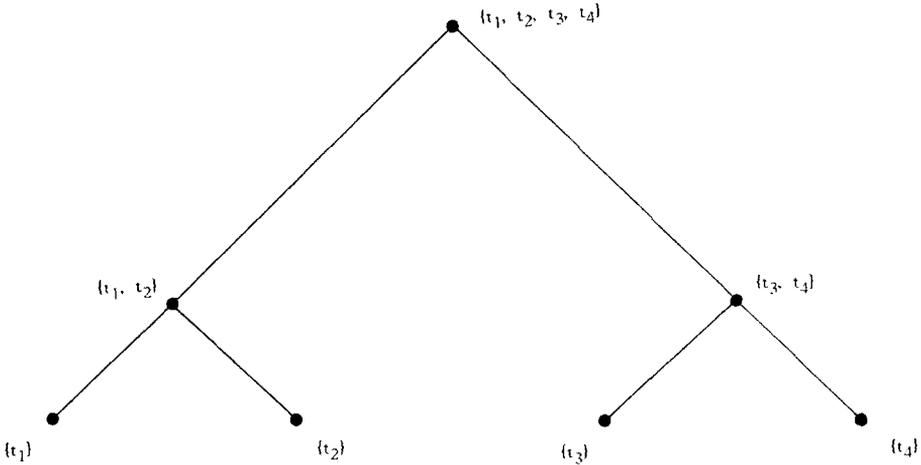
Figure 5. A possible taxonomic key for the 4 taxa in Figure 3. Compare with Figure 4.

appropriate. The capabilities of both classifications and keys will be functions not only of their purposes, but of the types of taxa and characters, of the numbers of taxa N, of the numbers of characters M, of the number of character states L, and of the structure of the classification or key. Moreover, as already accepted by virtually all systematists, the algorithm for creating a structure has an influence on the capability of the latter. The formulation of classifications and keys are discussed below under separate headings.

3. Classification

A taxonomic classification should be natural (in phenetics, the best reflection of overall similarity; in cladistics, the best reflection of evolutionary branching patterns), stable, predictive, and of high information content (see Sokal 1986). Most numerical methods define a distance $d(p_1 p_2)$ in character space (C_M), using axiom system III (see Appendix 1 and Sneath and Sokal 1973). This procedure is equivalent to a map from the M-dimensional character space to a distance space, here termed R, which is formally one-dimensional. In R, every pair of taxa possesses one coordinate value.

To realize the desired qualities of a classification, most phenetic classification procedures employ simple, scalar classification procedures or SAHN methods (Sneath and Sokal 1973), the most frequently used being the UPGMA algorithm. In effect, a mathematical model operating on distances in this manner, neglects the location and direction of taxa in character space, and assumes that character space is homogeneous and isotropic (Figure 6A).
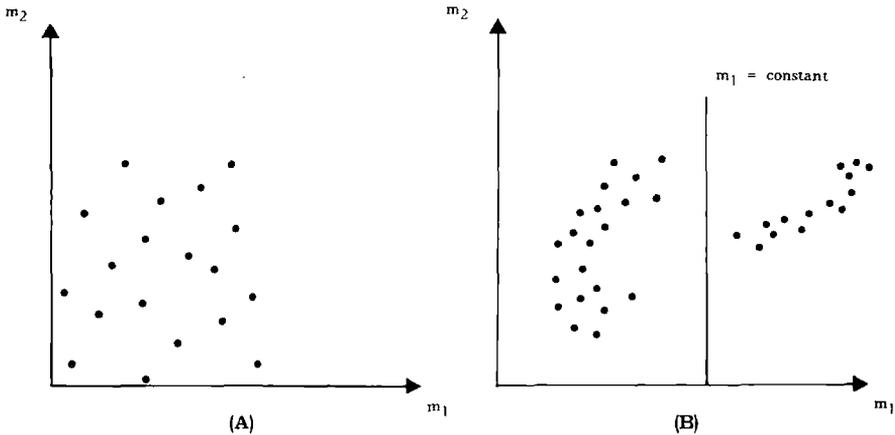
Figure 6. Taxonomic structure. A. An example of amorphous structure in the distribution of taxa in a two-dimensional character space. The points represent the locations of the taxa. B. An example of inhomogeneous and anisotropic structure in a two-dimensional character space. The plane represented by $m_1 = a$ constant serves to subdivide the taxa and is equivalent to a dichotomy in a taxonomic key.

In other words, all locations and all directions in the space are equivalent. There is rarely justification for such an assumption and Figure 6B shows a character space that is distinctly inhomogeneous and anisotropic (e.g., Ostrander 1969), i.e., there are areas and directions showing points (taxa) and others showing no points. The biological manifold manifold can be described as an inhomogeneous and anisotropic discontinuum in reference to the density of taxa in the character space.

There are methods that improve on simple scalar classification by accounting for the nature of character space. Character weighting (Sneath and Sokal 1973) is one such procedure. Another possibility, the approach to be developed here, is to use algorithms that interchange taxa in a classification. These are termed interchange scalar classifications (ISCL's). Under scalar interchange, Axiom system III of Appendix 1 is valid, but the distances are changed by interchanging taxa with surrounding taxa. The procedure may be termed thermodynamic taxonomy where the process of interchange itself is one "thermodynamic property." Another "thermodynamic property" is the increase in information gained by regarding taxa of different order than the one under consideration (Leuschner 1981, 1984; Zotin 1981, 1982). An algorithmic description of the procedure, suitable for continuous or ordered multistate characters, follows:

(1) Zero order ISCL (no interchanges permitted)

The distances are constant: follow the UPGMA algorithm
(2) First order ISCL (nearest-neighbor interchanges permitted)
- (a) Take two taxa, p_1, p_2
- (b) The nearest neighbors of p_1 and p_2 are termed p_{1n}^1 and p_{2n}^1, respectively.
- (c) Construct vectors representing the averages of all character states for the pairs of nearest neighbors, p_{1n}^1 and p_{2n}^1, giving a hypothetical auxiliary taxon p_{12n}^1.

- (d) Compute the distance between p_1 and p_{12n}^1, and between p_2 and p_{12n}^1.
- (e) Call these distances $d(p_1 \, p_a)$ and $d(p_2 \, p_a)$, respectively.
- (f) If $(d(p_1 \, p_a) + d(p_2 \, p_a)) > 2d(p_1 \, p_2)$ (Figure 7 and Appendix 2), then do not change $d(p_1 \, p_2)$. Select two taxa $\neq p_1, p_2$ and go to (b)
- (g) Else, reduce the distance $d(p_1 \, p_2)$ in R by a factor $f = (d(p_1 \, p_a) + d(p_2 \, p_a)) / 2d(p_1 \, p_2)$ (Figure 7).
- (h) Select two taxa $\neq p_1, p_2$ and go to (b)
- (i) When all distances have been transformed, follow the UPGMA algorithm. The locations of the clusters must be computed from the original values of the elementary taxa, not from the means of the transformed distances.

(3) Second order ISCL (nearest-neighbor and second-nearest-neighbor interchanges permitted).
- (a) For each taxon p_i considered, take into account both the nearest neighbor and the second nearest neighbor, $p_{in,1}, p_{in,2}$.
- (b) Calculate the averages for the M character states for $p_{1n,1}, p_{1n,2}$ and for $p_{2n,1}, p_{2n,2}$.
- (c) Find the average for these two vectors of all M character states to obtain a hypothetical auxiliary taxon p_{12n}.
- (d) Test and reduce the distance $d(p_1 \, p_2)$ using the factor f in the same way as for first order ISCL. Transform all distances and follow the UPGMA algorithm.

Third and higher order ISCL is performed in a fashion analogous to second order ISCL. In the case where the number of the surrounding taxa excluding p_1 and p_2 is odd, the last taxon, the residual neighbor taxon when all other taxa are attributed to p_1 or p_2, respectively, should be attributed to both p_1 and p_2.

The interchange procedure described above is an attempt to improve the capability of classifications irrespective of character weighting, the

number of characters used, the form of clustering algorithm used, and so on (Smirnov 1968, 1969). The model is an agglomerative procedure that results in a hierarchy with rank order, and phylogenetic interpretation is possible (see Sokal 1984). This thermodynamic approach may be improved by employing distance functions based on information theory (Schwartz 1963; Leuschner and Heine 1973; Leuschner 1974). Further properties of ISCL and applications will be furnished in a subsequent paper.

Using ISCL procedures, it is possible to measure the capability of a classification as:

$$c_{TC} = a \ln n \tag{1}$$

where a is a constant and n is the order of the ISCL procedure. Note the logarithm. It is of similar significance as in Boltzmann's formula in thermodynamics (Leuschner 1979). The measure in equation (1) is comparable only for classifications of the same set of elementary taxa and ISCL procedures of different orders. Moreover, high capability matches high information content and high information content matches high reliability. Note that when there are an even number of taxa, the maximum value of c_{TC} is:

$$c_{TC,max} = a \ln n_{max}$$
$$= a \ln ((N - 2) / 2)$$
$$= a \ln (N / 2 - 1) . \tag{2}$$

Here, n_{max} is only possible in the first step of an ISCL procedure.

4. Identification

If a determination must be made at all taxonomic ranks then a taxonomic key must be equivalent to a taxonomic classification. One problem of identification is to find a suitable set of characters at all dichotomies of a classification. Another approach is to find only those characters necessary to determine the lowest taxonomic rank of a specimen. There is conflict between the certainty of identification (best served by as many characters as possible) and ease of identification (best served by as few characters as possible). The velocity and reliability of identification are complementary (Leuschner and Sviridov 1986; Sviridov and Leuschner 1986). Identification differs from classification in that new, unknown taxa, not used in the original construction of a classification are tested against it.

The classification procedures described above are agglomerative (cf. divisive clustering, Sneath and Sokal 1973), but one approach to

identification proceeds from the set $\{t_1, t_2, \ldots, t_N\}$ to the elementary taxa $\{t_1\}, \{t_2\}, \ldots, \{t_N\}$, a divisive process. Assume that the number of characters $M \geq$ the number of taxa N. In character space, the equation $m_j = constant$ defines a hyperplane that divides the space (e.g., Figure 6B). This hyperplane is equivalent to a set of distances that are the projections of the taxa $\{t_i\}$ onto m_j. Using this definition, one may create structures that employ distances that fulfill Axiom system III of Appendix 1. One may proceed to find a set of hyperplanes that separate all of the individual taxa, resulting in a monothetic, divisive key. If no single character will unambiguously define a branching point, further characters may be used, resulting in a partially polythetic key. Other identification approaches need not be divisive. There are extensive Bayesian and distance-based identification systems where an unknown taxon u is identified with the closest taxon of an established classification (see Sneath and Sokal, 1973, for examples).

The formulae for certainty and ease of identification may be represented symbolically (see also Appendix 3): Define reliability Z, as

$$Z = \sum_{i=1}^{N} p_i \, (w_j \, (mw)^{\log N'})_i \qquad (3)$$

and velocity V as

$$V = (\sum_{i=1}^{N} p_i (\tau_j \, (m\tau)^{\log N'})_i \,)^{-1} \qquad (4)$$

where N is the number of taxa, p_i is the probability that elementary taxon i will be identified correctly, w_j is the probability of errorless passage through j, j is the character for the first branching point in the key, mw is the average probability of errorless passage through the remaining characters (excluding character j), N' is the number of taxa of the group in which elementary taxon i lies after its errorless passage through j, τ_j is the cost (e.g., time) spent working with character j, $m\tau$ is the average cost for the remaining characters, and log is $\log_2$.

Further, we may define the capability of a taxonomic key

$$c_{TK} = bZ + dV \qquad (5)$$

where b and d are constants. To maximize c_{TK} one may apply Lagrangian formalism, the mathematical method of seeking an extremum under an additional condition. For example, where V is a constant, let $b = 1$ and d is the Lagrangian multiplier. Ignoring Z, $c_{TK} = dV$, and from this it is possible to calculate a maximum value for c_{TK}. V is maximized when τ is minimized for all elementary taxa, and τ is minimized if all paths from the set $\{t_1, t_2, \ldots, t_N\}$ to the elementary taxa are equally short (Figure 5). The

number of characters q that must be tested is minimized. When the number of taxa is a power of 2, the minimum value of q is an integer because

$$q_{min} = \log_2 N \tag{6}$$

The maximum value of c_{TK} is:

$$c_{TK,max} = d^{'} \left(1 / \log_2 N\right) \tag{7}$$

where $d^{'}$ is a constant.

5. Discussion and Conclusions

There is no general theory for seeking a best lattice reduction algorithm and there is no general law that will describe the capability of a classification or taxonomic key as a function of N, M, L, and parameters describing the operations in classification and identification. Only for special cases is it possible to compare different classifications and identifications. There are important differences between classifications and keys that define the best ways in which to build them. Classification often is agglomerative, while identification is inherently a divisive process that must account for both efficiency and reliability. The use of computers greatly reduces the need for efficiency in the construction of keys, in which case they may become similar to the classifications on which they are based. Identification based on classifications will furnish considerably more information at each node in the search procedure than would taxonomic keys, whose branch points are frequently based on single characters only.

In numerical taxonomy, critical considerations relate to pair-group clustering cycles, dichotomous branching patterns, and the complementary nature of reliability and velocity. Progress in numerical taxonomy may be identified by the discovery of general laws for finding suitable lattice reduction algorithms, or the computation of indices for the capability of classifications and keys that makes the comparison of algorithms possible. Such progress is conditional on the training of taxonomists (Heywood 1975) to take advantage of mathematical and operational advances.

Appendix 1.

Define a set $P = P(T)$, the set of subsets of T, termed the power set of T, where $p_k \in P(T)$. Then, for example, when $T = \{t_1, t_2, t_3\}$, $P(T) = \{0, \{t_1\}, \{t_2\}, \{t_3\}, \{t_1, t_2\}, \{t_1, t_3\}, \{t_2, t_3\}, \{t_1, t_2, t_3\}\}$. When these subsets fulfill an additional criterion--their members have close relationships to one another--they are called taxa or clusters.

The notations { } and 0 mean set and empty set, respectively. Two operations are declared in P: set union, denoted $\cup$ and set intersection, denoted $\cap$. The notation .AND. and .OR. indicate the logical "and" and the logical "or", respectively, and := means "equal by definition". The term *iff* means "if and only if" and $\Rightarrow$ means "implies". The following axioms shall be valid for $p_1, p_2, p_3 \in P(T)$:

Axiom System I: Forms of Clustering

(1)   both intersection and union in $P(T)$ are unique:

$$p_1 \cap p_2 := \{t \in T: t \in p_1 \text{ .AND. } t \in p_2\}$$

$$p_1 \cup p_2 := \{t \in T: t \in p_1 \text{ .OR. } t \in p_2\}$$

(2)   the commutative law:

$$p_1 \cap p_2 = \{t \in T: t \in p_1 \text{ .AND. } t \in p_2\}$$

$$= \{t \in T: t \in p_2 \text{ .AND. } t \in p_1\}$$

$$= p_2 \cap p_1$$

$$p_1 \cup p_2 = p_2 \cup p_1$$

(3)   the associative law:

$$p_1 \cap (p_2 \cap p_3) = (p_1 \cap p_2) \cap p_3$$

$$p_1 \cup (p_2 \cup p_3) = (p_1 \cup p_2) \cup p_3$$

(4)   the absorption law:

$$p_1 \cap (p_1 \cup p_2) = p_1$$

$$p_1 \cup (p_1 \cap p_2) = p_1$$

Further, $p_1 \leq p_2$ *iff* $(t \in p_1 \Rightarrow t \in p_2)$

Axiom System II: Hierarchical Clustering

(1')   the law of reflexivity:

$$p_1 \leq p_1$$

(2')   the law of transivity:

$$p_1 \leq p_2 \text{ .AND. } p_2 \leq p_3 \Rightarrow p_1 = p_3$$

(3')   the law of antisymmetry:

$$p_1 \leq p_2 \; .AND. \; p_2 \leq p_1 \Rightarrow p_1 = p_2$$

Lastly, let $d(p_1 p_2)$ be the distance between p_1 and p_2, where $d(p_1 p_2) > 0$ for all $p_1 \neq p_2$. (We shall restrict d to being positive semidefinite).

Axiom System III: Sequential and Nonoverlapping Clustering

(1'')   the law of identity:

$$d(p_1 p_2) = 0 \; iff \; p_1 = p_2$$

(2'')   the law of symmetry:

$$d(p_1 p_2) \equiv d(p_2 p_1)$$

(3'')   the triangle inequality:

$$d(p_1 p_2) \leq d(p_1 p_3) + d(p_3 p_2)$$

In summary, Axiom System I represents a lattice (a specific algebraic structure), Axiom System II sets up a partial ordering structure (a specific order structure), and Axiom System III defines a metric space (a specific topological structure). The example in Figure 1, $T = \{t_1, t_2, t_3\}$, shows that Axiom Systems I and II are compatible. A lattice can be represented by a graph where 0 is the empty set or zero element and $\{t_1, t_2, t_3\}$ is the unit element of the lattice.

Axiom System III may be represented by an algorithm that reduces the number of connections in the lattice. A lattice such as the one in Figure 1 may be reduced to a taxonomic lattice, that is, a classification (Figure 2) or a key to be used for identification. The type of taxonomic lattice, a classification or key, and the form of the reduction algorithm, depend on the choice of $d(p_1 p_2)$ in Axiom System III and on the limit for the distance in $P(T)$.

Appendix 2.

A two-dimensional paradigm for some step of a first order ISCL-procedure is shown in Figure 7. We consider the two taxa p_1 and p_2 with their nearest neighbors p_{1n}^1 and p_{2n}^1, respectively. We consider the following distance transformation. The distance $d(p_1 p_2)$ is claimed to be equal to the semi-major axis a of an ellipse for which by definition $e^2 = a^2 - b^2$ holds. By setting $e = a/2 = d(p_1 p_2)/2$ one obtains the semi-minor axis $b = \sqrt{2} \, a / 2$. Thus we have an ellipse characterized by $g_k + h_k = g + h =$
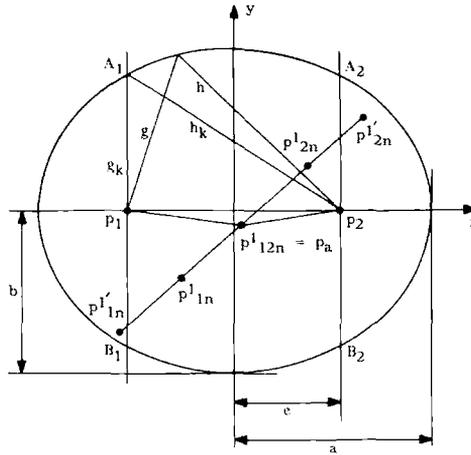
Figure 7. Geometric relationships of nearest neighbors in first order ISCL. In this example, p_1 and p_2 are the two taxa being considered. Their nearest neighbors are p^1_{1n} and p^1_{2n} respectively. The average vector for these two points is shown as p^1_{12n}. The distances $g + h = 2d(p_1p_2)$ and the curve transcribed by g and h is an ellipse.

$2\,d(p_1p_2)$ as in Figure 7. We can compute the auxiliary taxon $p_{12n} = p_a$. If $A_{ES} \equiv area\ A(A_1A_2B_2B_1A_1)$, then

$$p^1_{1n}, p^1_{2n} \in A_{ES} = \{p \mid -d(p_1p_2)/2 < x < +d(p_1p_2)/2 ,$$
$$-\sqrt{d(p_1p_2)^2 - x^2 /2} < y < +\sqrt{d(p_1p_2)^2 - x^2 /2} \} \qquad (a)$$

and it follows that:

$$d(p_1p_a) + d(p_2p_a) < 2d(p_1p_2) \qquad (b)$$

We define d_{cr} as the critical agglomeration distance in the cluster step under consideration. Furthermore, we set in the example $d(p_1p_2) = 1$, $d(p_1p^1_{1n}) = 0.38, d(p_2p^1_{2n}) = 0.23, d_{cr} = 3/4$. Therefore we obtain

$$d(p_1p_a) + d(p_2p_a) \cong 1.03 > 0.75$$

$$f = (d(p_1p_a) + d(p_2p_a))/2d(p_1p_2) \cong 1.03/2 \cong 0.52 < 0.75.$$

The transformed distance d_{tr} is

$$d_{tr}(p_1p_2) = fd(p_1p_2) \cong 0.52d(p_1p_2) = 0.52 < d_{cr} < d(p_1p_2) .$$

Because $d_{tr}(p_1p_2) < d_{cr}$ an agglomeration takes place in this cluster step, although $d(p_1p_2) > d_{cr}$. In the case of p^1_{1n} and p^1_{2n} we obtain the same

p_a. Though condition (a) is valid no transformation is made, because condition (b) is violated. The procedure given above can be generalized to many dimensions and other character types.

Appendix 3.

The following is a justification for formulas (3) and (4). The fundamental idea is that the decision at the first branching point is the most important decision of the key. Two examples will make this clear. The first is a case where the two taxa following the first branching point are an elementary taxon on one branch and an important taxon of high rank containing the balance of elementary taxa on the other branch. As a second example we imagine the case of two taxa of relatively high rank following the first branching point, both containing approximately the same number of elementary taxa. In these two cases and others we obtain much information after passing the first branching point. We, therefore, prefer the first branching point in formulas (3) and (4).

References

HEYWOOD, V. H. (1975), ''Contemporary Objectives in Systematics,'' in *Proceedings of the Eighth International Conference on Numerical Taxonomy, Oeiras, Portugal*, ed. G. F. Estabrook, San Francisco: W. H. Freeman, 258-283.

LEUSCHNER, D. (1974), *Einführung in die numerische Taxonomie*, Gustav Fischer, Jena.

LEUSCHNER, D. (1979), *Grundbegriffe der Thermodynamik*, Berlin: Akademie-Verlag.

LEUSCHNER, D. (1981), ''A Specification of Numerical Taxonomy: Thermodynamic Taxonomy,'' *Biometrical Journal, 23*, 611-620.

LEUSCHNER, D. (1984), ''Information and Cybernetic Aspects of Biological Thermodynamics,'' in *Thermodynamics and Regulation of Biological Processes*, eds. I. Lamprecht and A. I. Zotin, Berlin: De Gruyter, 3-18.

LEUSCHNER, D. (1985), ''Modellierung - Rucksack oder Schnürsenkelprinzip,'' *Vortrag im Seminar: Statistische Analyse und Modellierung von Zusammenhängen, 24-29/March/1985*, Reinhardsbrunn, GDR.

LEUSCHNER, D. and HEINE, R. (1973), ''Informationstheoretische Kriterien in der numerischen Taxonomie,'' *Biometrische Zeitschrift, 15*, 393-401.

LEUSCHNER, D. and SVIRIDOV, A. V. (1986), ''The Mathematical Theory of Taxonomic Keys,'' *Biometrical Journal, 28*, 109-113.

OSTRANDER, C. C. (1969), ''A Mathematical Study of the Genus *Pentremites*, in *Numerical Taxonomy*, ed. A. J. Cole, London: Academic Press, 165-180.

SCHWARTZ, L. S. (1963), *Principles of Coding, Filtering and Information Theory*, Baltimore: Spartan Books.

SMIRNOV, E. S. (1968), ''Taxonomische Analyse als Mittel zum Aufbau eines natürlichen Systems,'' *Beiträge zur Entomologie, 18*, 347-376.

SMIRNOV, E. S. (1969), *Taxonomical Analysis* (in Russian), Moscow: Moscow University Press.

SNEATH, P. H. A. and SOKAL, R. R. (1973), *Numerical Taxonomy*, San Francisco: W.H. Freeman.

SOKAL, R. R. (1984), "Die Caminalcules als taxonomische Lehrmeister," *Studien zur Klassifikation, 15*, 15-31.

SOKAL, R. R. (1986), "Phenetic Taxonomy: Theory and Methods," *Annual Review of Ecology and Systematics, 17*, 423-442.

SVIRIDOV, A. V. and LEUSCHNER, D. (1986), "Optimization of Taxonomic Keys by Means of Probabilistic Modelling," *Biometrical Journal, 28*, 609-616.

ZOTIN, A. I. (1981), "Bioenergetical trend of the evolutionary process of organisms," (in Russian), *Scientific Centre of Biological Investigation, Academy of Sciences, Pushkino, USSR*, 3-10.

ZOTIN, A. I. (1982), "Velocity and Direction of the Evolutionary Progress of Organisms," (in Russian), *Zhurnal Obzhenye Biologii, 43*, 3-13.