Pattern Recognition. II. Linear and Nonlinear Methods for Displaying Chemical Data[1a]

B. R. Kowalski[1b] and C. F. Bender*

Contribution from the Lawrence Livermore Laboratory, University of California, Livermore, California. *Received July 21, 1972*

Abstract: Pattern recognition is a developing discipline within the field of artificial intelligence that can be used to solve chemical problems. Possibly the most useful of its four major branches is visual display. Linear and nonlinear display methods can be used to represent multivariant chemical data in two dimensions, thereby allowing the chemist an approximative visual examination of his data. This examination often determines the course of further action in a pattern recognition application and, in some cases, can actually allow the chemists to solve the problem using only the display. Several display methods are presented for comparison with the aid of two synthetically generated data sets and two chemical examples.

In an earlier paper,[2] the field of pattern recognition, a newly emerging discipline within artificial intelligence, was introduced to the chemical literature. The paper used two examples to demonstrate that the methods of pattern recognition, when used in combination, provide a general approach to solving a class of data processing problems commonly encountered in experimental chemistry. A statement of the general problem is: Can an obscure property of a collection of objects (elements, compounds, mixtures, etc.) be detected and/or predicted using indirect measurements, made on the objects, that are known to be related to the property *via* some unknown relationship? Obviously, the problem is not only to find and predict the property, but also to try to find the mathematical relationship that links the measurements to the property. Therefore, the problem can be considered as a mapping of objects from measurement space into property space. Screening chemical compounds to find the best performer for a particular process (medical, agricultural, production, etc.) is an example that can clarify the above point. If a chemist knows of a few compounds that perform well in the process of interest, and a few compounds that perform poorly, he might want to find other compounds that show promise. The empirical method often used is to make a number of chemical and/or physical measurements on each of the known (previously tested) compounds to see if a discriminating measurement can be found that separates the good performers from the poor performers. In some cases, the measurements can be found in data handbooks and need not actually be measured. The obscure property, in this case, is the relative performance of each compound in the process. Hence, the problem reduces to mapping candidate compounds from measurement space to performance space.

All too often, a single discriminating measurement cannot be found, and only a combination of measurements can provide the necessary information. When the number of measurements is small (≤ 3), the human is the best pattern recognizer. However, when the number of measurements is large (>3) and the objects many ($>\sim 3$), the techniques of pattern recognition can be most beneficial.

(1) (a) This work was performed under the auspices of the U. S. Atomic Energy Commission. (b) Department of Chemistry, Colorado State University, Fort Collins, Colo. 80521.

(2) B. R. Kowalski and C. F. Bender, *J. Amer. Chem. Soc.*, **94**, 5632 (1972).

Pattern Recognition Approach Display

If the measurements for a particular application have been selected properly, it is reasonable to assume that "like" objects will have similar measurements. (This is, of course, in the context of the particular property under study.) One method of representing the likeness among objects is to consider the objects as points in an n-dimensional hyperspace, where n is the number of measurements made on each object. The values of the measurements are the coordinates that position each point in n-space. The objects can be said to occupy a position in measurement space. Now as mentioned earlier, the goal is to map the points from this measurement space to some property or outcome space. If the property is known, and if some objects with known property are available, the application is termed *supervised learning*[2] and is best solved by methods that are not within the scope of this paper. If, however, the sought-for property is not exactly known, or if examples are nonexistent the application is one of *unsupervised learning*.[2] The practical distinction is whether or not the computer is given information about the sought-for property. There are several methods to apply in the latter case; the class of methods examined in this paper is called *display methods*. The common goal of display methods is to represent the data structure of points in n-space by the same number of points in m-space where ($n > m$). In most cases, m is equal to 2. Clearly, this cannot always be done exactly and in most cases information loss is inevitable. However, most display methods seek to preserve information and, therefore, minimize the information loss according to some criterion. Within the applications experience of the authors, display methods are the most useful tools in pattern recognition. The reason for this, as mentioned above, is that the human is the best pattern recognizer in the familiar two- or three-dimensional space. If the data structure is *not* so complex as to be totally nonrepresentable in two-space, then many applications can be solved by the scientist with the aid of display methods.

Basically, there are two general approaches to displaying n-space (n will be assumed to be greater than 2) in two-space: linear and nonlinear. The two-

coordinate axes of the resulting two-space display can either be linear or nonlinear combinations of the n coordinates of the original n measurements. Popular display methods utilizing these two approaches will be discussed in some detail following a description of the data used to compare these methods.

Data. Four different sets of data are used to aid the discussion of display methods. Two sets are composed of synthetically generated data and the other two sets use real chemical data.

The first set of data consists of points on the surface and one point in the center of a three-dimensional cube. This set will be called CUBE. CUBE is a unit cube in the positive quadrant of three-space with one corner at the origin. Its 27 points are located at the corners, the center faces, half-way along each side, and one point at the center of this unit cube.

The second set, called SPHERE, consists of 62 points on the surface of a sphere of unit radius and centered at the origin. Two points lie at each pole, 12 points are equally spaced at the equator, and 12 points lie at each of four latitude equally spaced between the poles and the equator.

Data set three is an improved version of the second data set in the introductory paper[2] and will be referred to as CHEM. CHEM consists of 64 points representing 64 elements in six-space. The six measurements are (1) most important valence, (2) melting point, (3) covalent radius, (4) ionic radius, (5) electronegativity, and (6) ΔH of fusion. The object of the first study[2] using these data was to try to separate those elements with basic higher valence oxides from those with predominantly acidic higher valence oxides, and then to classify the amphoteric oxides as being slightly more acidic or basic.

The sought-for property connected with CHEM is a continuous value as opposed to the sought-for property in the forth data set, ARROW, which is discrete (class membership). ARROW consists of 74 points in ten-space. The data are taken from an application of pattern recognition to the source verification and artifact identification study of archaeological obsidian.[3] Forty-five samples of obsidian were collected from four known sources in a particular geographical region. Twenty-nine obsidian artifacts (arrowheads, tools, etc.) strongly suspected as coming from these four sources were also collected. The problem, which was solved by pattern recognition,[3] was to verify the existence of four sources and then to classify the unknown artifacts using the concentrations of ten trace elements measured on each sample.

CUBE and SPHERE are common three-dimensional geometrical figures and are used to demonstrate the effect of the display methods on known data structures. The data were used as is and were not scaled or weighted.

CHEM and ARROW represent chemical data sets of two often encountered types. Both of these data sets were autoscaled.[2] In CHEM, the property (acidity and basicity) is a continuous value, whereas in ARROW, the property is class membership, which is discrete. The latter set is really a classical type of application for pattern recognition but the former demonstrates the flexibility of pattern recognition. Display methods are quite useful in determining which type of property (continuous or discrete) is inherent in the data as will be demonstrated.

(3) B. R. Kowalski, T. F. Schatzki, and F. H. Stross, *Anal. Chem.*, 44, 2176 (1972).

Figure 1. Projection of CUBE on two axes.

Linear Projections

Variable by Variable Plotting. Probably the most used data-projection scheme is the variable by variable plot. These two-space plots are useful because they eliminate a considerable amount of verbiage that would be necessary to convey the same amount of information. Although it seems almost trivial at first, the variable by variable (VXV) plot is a linear projection and is a logical starting point that must be included in a comparison of display methods. It should be understood that for an application involving n measurements (n-space), there are $^1/_2n(n-1)$ different plots that must be examined. The plots contain no projection errors and if all are examined, no information loss results. However, only a one by one comparison is possible and $^1/_2n(n-1)$ can get to be a large number. In a real sense, the amount of information that is realized can be minimal.

Because of the orientation of CUBE, the three possible VXV plots are all the same (Figure 1). Of the three possible VXV plots of SPHERE, only two are unique as shown in Figures 2 and 3. It is important to note that Figure 1 gives no hint of a three-dimensional structure. The situation is somewhat improved in Figures 2 and 3. These figures will be useful for studying the more sophisticated methods.

There are 15 independent VXV plots for CHEM and 45 VXV plots for ARROW. In the interest of brevity, only a few are presented here. Figure 4 is the CHEM data plotting covalent radius *vs.* ΔH of fusion. Figure 5 is the CHEM data plotting melting point *vs.* ionic radius. While some separation is evident between the acids and bases in the two plots neither plot is adequate for the separation. A similar condition exists for the ARROW data. When iron is plotted *vs.* zirconium (Figure 6), sources two and four are nicely separated but one and three overlap badly. In Figure 7, titanium is plotted *vs.* barium and, while classes three and two are separated, one and four now overlap. The need for a better projection is evident from Figures 4–7.

Rotation and Projection. The variable by variable plot is a true linear projection where the two coordinates

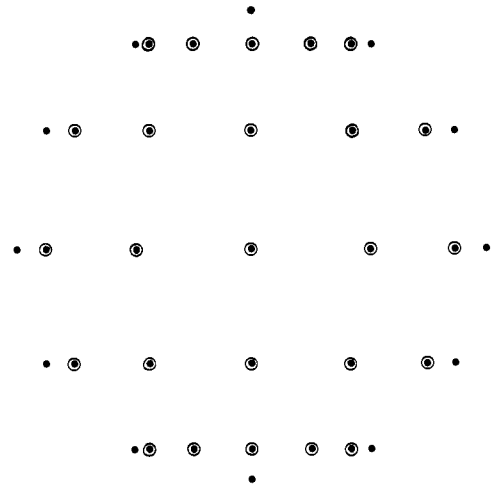

Figure 2. Projection of SPHERE on axes one and two (circles indicate overlap).

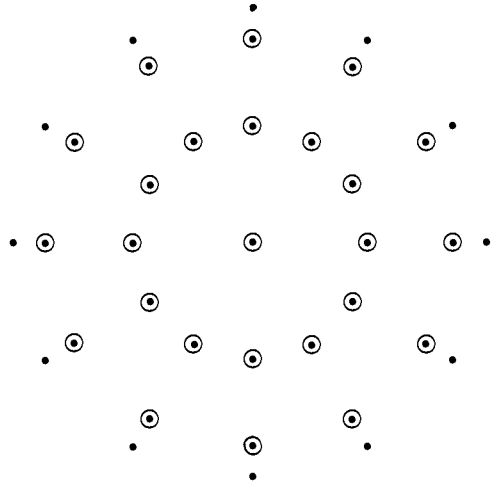

Figure 3. Projection of SPHERE on axes one and three (circles indicate overlap).

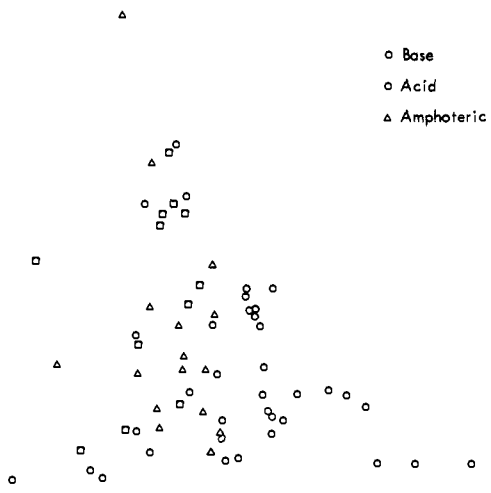

Figure 4. CHEM data projected on covalent radius and ΔH of fusion (overlapping points not shown).

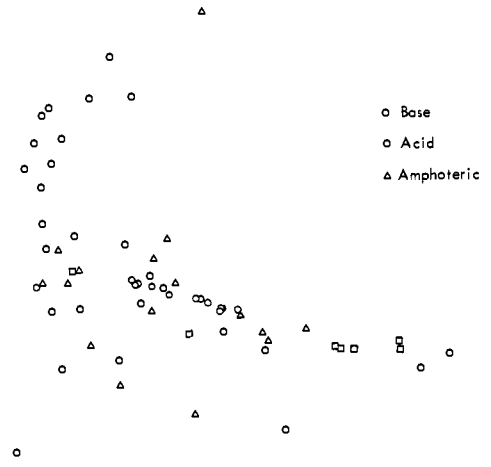

Figure 5. CHEM data projected on melting point and ionic radius.

have a trivial linear relationship to only two of the variables. Other types of linear projections are possible. Noll[4] has shown that multidimensional hyperobjects can be rotated through any angle in any dimension and then projected onto a plane. By using two-dimensional rotation matrices $R_{ij}(\alpha)$, points in n-space are rotated by angle α in the i,j plane. A full rotation in n-dimensional space can be considered as $^1/_2n(n-1)$ two-dimensional rotations and the final rotation matrix R, in a four-dimensional space for example, is written as

$$R = R_{12}(\alpha_1)R_{13}(\alpha_2)R_{14}(\alpha_3)R_{23}(\alpha_4)R_{24}(\alpha_5)R_{34}(\alpha_6) \quad (1)$$

Once the rotation is made, the points can be viewed as before, variable by variable, but now each new variable is a linear combination of the original variables.

Unless some information is known that indicates a certain rotation to be beneficial, it is most difficult to know actually how to start. However, if an interactive computer graphics terminal is available, all rotations can be made in real time. The effect is remarkable and an excellent three-dimensional perspective is possible. Figure 8 is the result of a slight rotation in two planes in the direction shown by the arrows. A comparison to Figure 1 gives a hint as to how real-time rotation and projection can give a three-space perspective. Still, only three variables at a time can really be viewed which is only a slight improvement over two. The need to display points in higher dimensional spaces must rely on more sophistication.

Eigenvector Projection. The Karhunen-Loeve transformation[5] is quite useful for feature selection in pattern recognition. As a special case of this transformation a data display can be obtained. The Karhunen–Loeve method creates new variables as *linear* combinations of the original variables and can be thought of as automatic multidimensional rotation. A unique ordering is the result of this transformation. The first new variable contains the greatest amount of variance and each successive new variable contains the next greatest amount of the residual variance. In this way, redundancies in the data can be eliminated by truncating the last few variables if their variance is zero or near zero. Thus, the transformation is optimal (in

(4) A. M. Noll, *Commun. ACM (Ass. Comput. Mach.)*, **10**, 469 (1967).

(5) K. Fukunaga and W. L. G. Koontz, *IEEE Trans. Comput.*, **C-19**, 311 (1970).

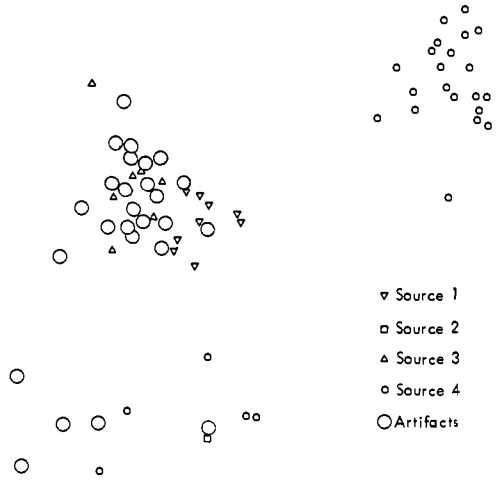

Figure 6. ARROW data projected on the iron and zirconium measurements (overlapping points not shown).

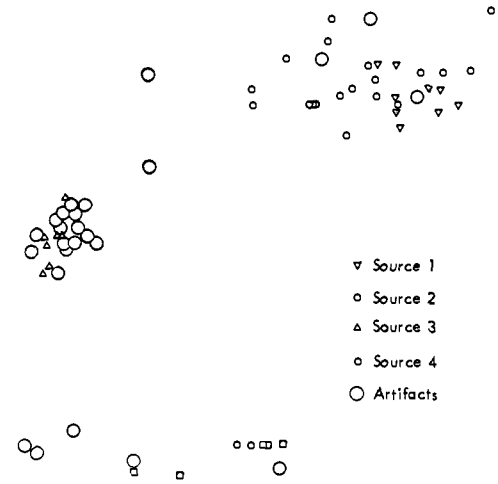

Figure 7. ARROW data projected on the titanium and barium measurements (overlapping points not shown).

the sense that variance is preserved) for feature ordering and selection.

The method starts by calculating the n variable (measurement) means $\bar{X}_k (k = 1 \text{ to } n)$ as

$$\bar{X}_k = \frac{1}{m} \sum_{l=1}^{m} X_{lk} \tag{2}$$

where n is the number of variables and m is the number of points (objects). Next the covariance matrix $\mathbf{C}$ is generated, each element C_{ij} of which compares variables i and j as

$$C_{ij} = \sum_{l=1}^{m} (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j) \tag{3}$$

Next, the eigenvalues λ_k and eigenvectors μ_k for $k = 1$ to n are calculated by solving

$$\mathbf{C}\mu_k = \lambda_k \mu_k \tag{4}$$

For display purposes in two dimensions, the basis vectors μ_1 and μ_2 corresponding to the two largest eigenvalues λ_1 and λ_2 would be used as a projection plane. The eigenvector plot is a linear projection because the new two-space coordinates are linear combinations of all the original coordinates. Additionally, the projection is the best linear projection that can be obtained (minimum mean-squared error of variance).

Now since each eigenvalue is proportional to the variance along its corresponding eigenvector, a measure is available of the per cent variance retained by the eigenvector projection. This value, % V, calculated as

$$\% \text{ V} = (\lambda_1 + \lambda_2)100 \Big/ \sum_{i=1}^{n} \lambda_i \tag{5}$$

is useful for determining the reliability of interpretations made using a projection. If % V is equal to 50, for instance, only one-half of the variance has been retained and interpretation will have a high risk.

Figure 9 is the eigenvector plot (% V = 87) of the 27 points in CUBE. Although Figure 8 looks more like a cube, Figure 9 is actually a better representation of the information. The circled points indicate overlapping points which can cause problems in interpretation. Figure 10 is the SPHERE data projected on the two dominant eigenvectors (% V = 71). The

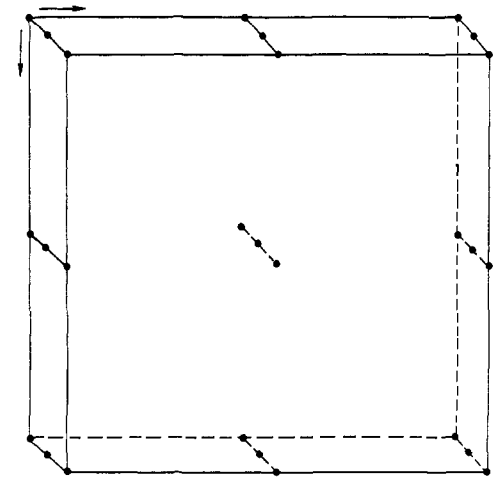

Figure 8. Rotation of CUBE followed by projection on two axes.

problem of overlapping points (again circled) is much more severe in this case and, even though the plot represents the best that can be done using a linear projection, interpretation would involve high risk. Figures 11 and 12 are the eigenvector plots of CHEM (% V = 72) and ARROW (% V = 73), respectively. Although the primary goals of each application are met, namely, the different classes are separated, assigning classifications to the unknowns would again involve a high risk since almost 30% of the variance is lost in each projection. One of the useful facts obtained by examining Figures 11 and 12 is that there must be a large amount of redundant information in the measurements. This is so because the classes are linearly separable in two-space even though CHEM used six measurements and ARROW ten. The plots indicate that certain measurements might be eliminated and that dimensionality reduction[5] studies should be made.

Before proceeding to nonlinear data display methods, two important points should be made. First, data containing axes of symmetry (CUBE and SPHERE) do not give unique projections. CHEM and ARROW do not suffer from this problem as most applications using nonsynthetic data do not. Second, all of the techniques described in this paper can be used to reduce n-space to m-space where $n > m$. Two-dimensional plots are

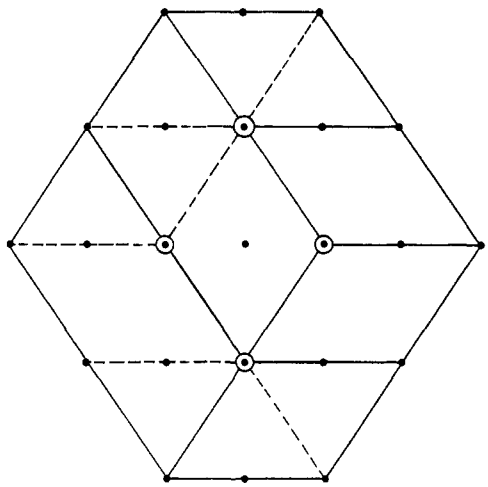

Figure 9. Eigenvector projection of CUBE (circles show overlap).

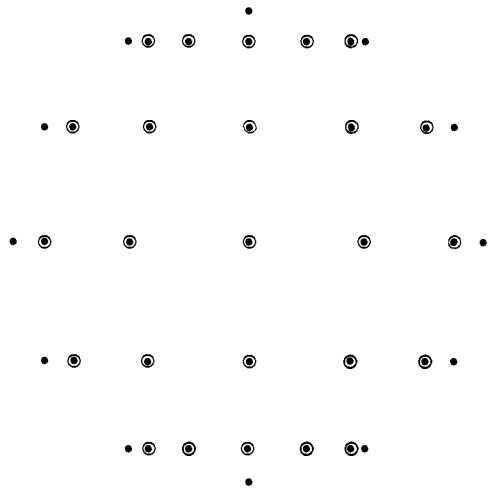

Figure 10. Eigenvector projection of SPHERE (circles show overlap).

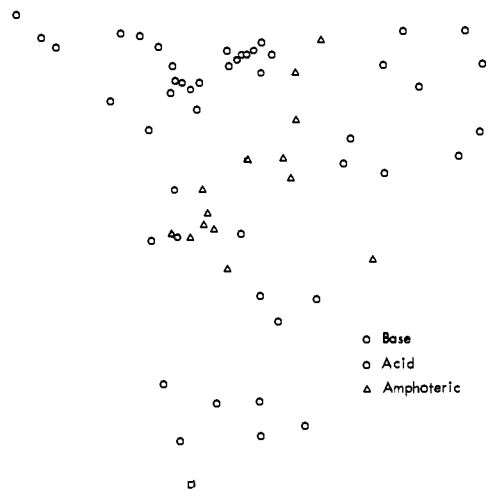

Figure 11. Eigenvector projection of CHEM (overlapping points not shown).

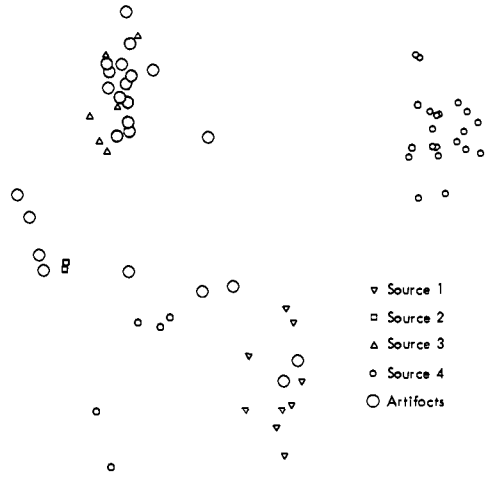

Figure 12. Eigenvector projection of ARROW (overlapping points not shown).

naturally most useful for this presentation and, therefore, most of the examples have m equal to 2. Using computer graphics,[6] however, three-space plots can be quite useful.

Nonlinear Mappings

In the above section, linear display methods were referred to as projections. In this section, methods that produce displays with coordinates that are *not* linear combinations of the original n-space coordinates are discussed. The resulting displays are called maps.

Specialized Maps. As might be expected, there are many criteria that can be used to map n-space to m-space ($n > m$) in a nonlinear manner. Some of these methods closely resemble clustering and even classification methods. Since these specialized techniques can utilize any criteria that can come to mind, only a few will be mentioned to give a flavor of what can be done.

Patrick, *et al.*,[7] describes two particular mappings, called dovetail mapping and column mapping, that can be used to display multidimensional space in one dimension. The two methods are really indexing schemes and break up the entire n-space into finite sized n-dimensional regions. These regions are then given an identification index. If a point falls in a particular region, it is given the corresponding index. The index relates to a particular position on the real line. The two types of mappings are not unique but they do show logical approaches to display.

Fukunaga and Olsen[8] have presented an interesting technique (again nonlinear) for the two-dimensional display of multivariate data. The method uses one of three procedures to normalize the data, and then displays the points on a two-dimensional display whose coordinates are the Euclidean distances from two particular points in the n-space. In the two-class, supervised learning mode, the two points are the geometric centers of the two classes. The method preserves some geometric structure while putting heavy weight on class separability. In the unsupervised learning mode, the method closely resembles a cluster analysis[2] tool. The procedure starts by selecting a random line and, assuming that the points on either side of the line form two classes, the pseudo-class means are calculated and the map displayed. The line is then

(6) J. W. Sammon, Jr., A. H. Proctor, and D. F. Roberts, *Pattern Recog.*, 3, 37 (1971).

(7) E. A. Patrick, D. R. Anderson, and F. K. Bechtel, *IEEE Trans. Comput.*, C-17, 949 (1968).

(8) K. Fukunaga and D. R. Olsen, *ibid.*, 917 (1971).

changed and the process repeated. Iteration continues until the class separation is reasonable.

A number of other nonlinear methods have been suggested by Kruskal.[9] These methods are not discussed in order to keep the size of this paper manageable.

Nonlinear Mapping by Error Minimization. Possibly the most useful display of multidimensional data has been presented by Sammon[10] and first applied to chemical information by Kowalski and Bender.[2] The method is called nonlinear mapping (NLM) and uses a logical criterion (preservation of interpoint distances). The procedure suggested by Sammon[10] has been significantly modified and is presented in the following.

The eigenvector plot, described above, is used as the starting configuration for the NLM map. All of the n-space interpoint distances, d_{ij}^*, are calculated as

$$d_{ij}^* = \left[\sum_{k=1}^{n} (X_{ik} - X_{jk})^2\right]^{1/2} \quad (6)$$

and all of the two-space interpoint distances, d_{ij}, are calculated as

$$d_{ij} = [(Y_{i1}^* - Y_{j1})^2 + (Y_{i2} - Y_{j2})^2]^{1/2} \quad (7)$$

where the Y's are found by the rotation matrix that diagonalizes C in eq 4. The object here is iteratively to change the two coordinates (Y_{i1} and Y_{i2}) for each point Y_i so as to minimize an error function E, defined as

$$E(\rho) = \sum_{i<j} \frac{(d_{ij}^* - d_{ij})^2}{(d_{ij}^*)^\rho} \quad (8)$$

The minimization attempts to *preserve interpoint distances* by finding d_{ij}'s that are as close as possible to d_{ij}^*'s. Here, $(d_{ij}^*)^\rho$ is a weighting factor whose effect is examined later. Since the d_{ij}^*'s are constants for any given application, the unknowns in this error function are the two-space coordinates from eq 7.

The distance function used in this paper is the common Euclidean distance but any distance function can be used. The generalized Mahalanobis distance[11] is

$$D_{ij} = \left[\sum_{k=1}^{m} (Y_{ik} - Y_{ik})^P\right]^{1/P} \quad (9)$$

and provides a selection of possible functions. White[12] has suggested the use of the Hamming metric

$$H_{ij} = \sum_{k=1}^{m} |Y_{ik} - Y_{jk}| \quad (10)$$

which has certain computational advantages.

In order to iteratively change the two-space coordinates and minimize E, a gradient method should be used. Sammon[10] suggests the method of steepest descent. Since the minimization may involve considerable computation, careful selection of a technique is important. The method used here is the Polak–Ribiere[13] method which is similar to the well-known Fletcher–Reeves[14] conjugate gradient method. Since this method is adequately described and a computer program is outlined in the reference cited, the details will not be given here.

(9) J. B. Kruskal, *Psychometrika*, **29**, 115 (1954).
(10) J. W. Sammon, Jr., *IEEE Trans. Comput.*, **C-18**, 401 (1969).
(11) P. C. Mahalanobis, *Proc. Nat. Inst. Sci. India*, **122**, 49 (1936).
(12) I. White, *IEEE Trans. Comput.*, 220 (1972).
(13) E. Polak, "Computational Methods in Optimization," Academic Press, New York, N. Y., 1971, p 53.
(14) Reference 13, p 52.

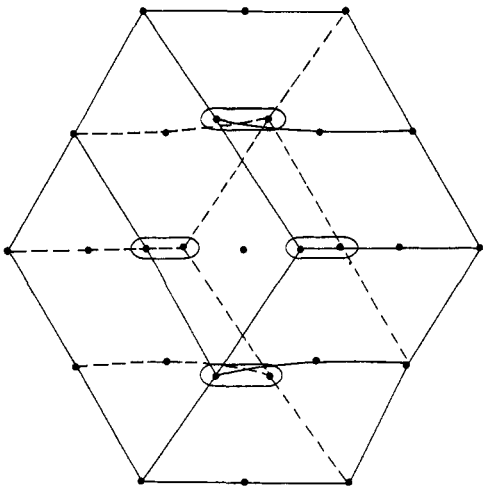

Figure 13. Nonlinear map of CUBE.

NLM would be a difficult method to understand if it was not for a very useful analogy in physics. The analogy comes from the study of forces exerted on springs. Let us assume that we have a collection of points in three-space and we wish to use NLM to map them to two-space. Now, if *every* point is connected to *every other* point by a tensionless spring, the total energy in the springs would be zero in three-space. Equation 8 can be used to estimate the total energy in the springs where d_{ij}^* is the length of the spring between points i and j in the original three-space and d_{ij} is the length of the compressed spring after mapping to two-space. A linear projection method simply rotates the collection of points in three-space and then collapses them onto a plane. The points are not allowed to move laterally during the projection and a possible result is overlapping points as is the case in Figures 9 and 10. The energy in the springs is quite high in these cases. NLM on the other hand allows the points to move about freely so that overlap is eliminated and the squeezed, two-space configuration has the minimum tension (energy) in the springs. Thus, finding this lowest energy two-space configuration is equivalent to minimizing eq 8 with respect to the two-space coordinates.

Figure 13 is the NLM of CUBE after allowing a certain number of iterations. This figure should be compared to Figure 9. The four elipses enclose points that overlapped in Figure 9.

Figure 14 is the NLM of SPHERE. In this case, as in Figure 13, the value of ρ in eq 8 is 2. This corresponds to an equal weighting of small and large distances. In other words, the same amount of effort is spent preserving distances of all magnitudes. Using the spring model, this means that all of the springs have the same spring constant. When ρ is equal to -2, as in Figure 15, the large distances are preserved at the expense of the small distances. The distortion is clearly shown by comparing Figures 14 and 15. In order to see the improvement of NLM over eigenvector projection, Figures 10 and 14 can be compared. The problem of overlapping points is solved by NLM.

Figure 16 is the NLM of CHEM. Again, as in Figure 11, the acidic oxide elements are easily separated

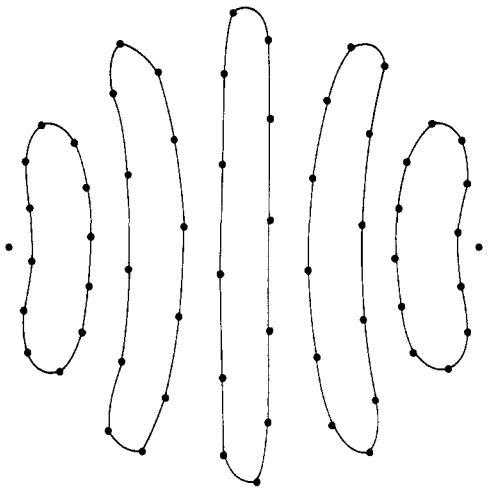

Figure 14. Nonlinear map of SPHERE ($\rho = 2$ in eq 8).

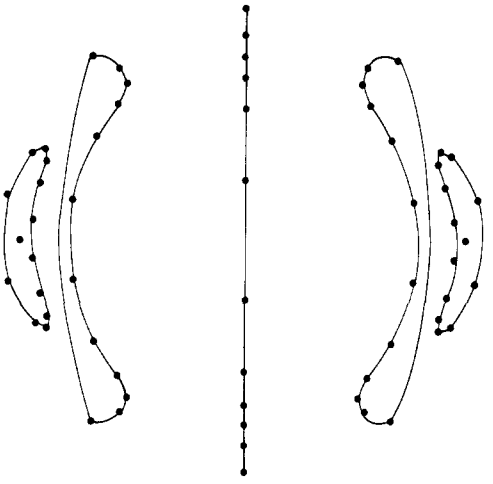

Figure 15. Nonlinear map of SPHERE ($\rho = -2$ in eq 8).

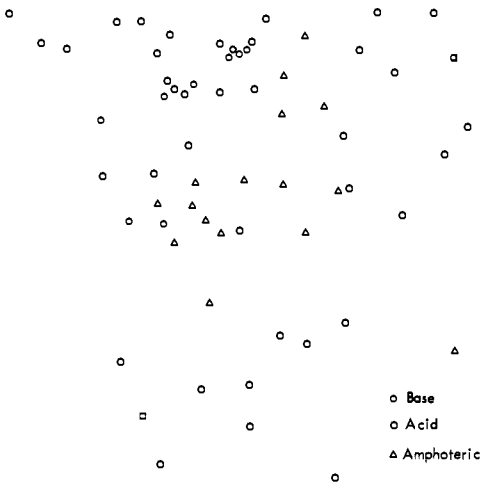

Figure 16. Nonlinear map of CHEM.

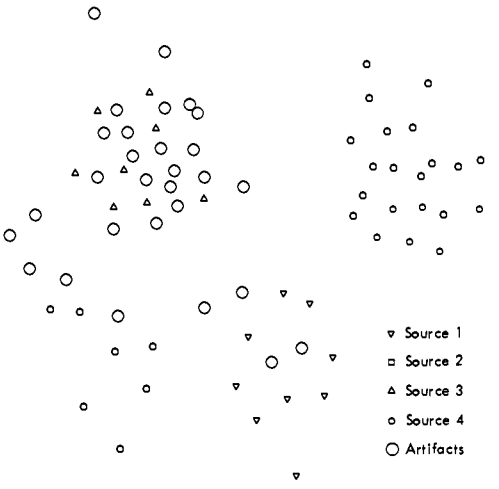

Figure 17. Nonlinear map of ARROW.

from the basic oxide elements. However, classification of the unknowns (amphoterics) is considerably more reliable using Figure 16. In fact, the error (eq 8) drops by more than four orders of magnitude going from Figure 11 to Figure 16.

The eigenvector plot of ARROW (Figure 12) can be compared to the NLM (Figure 17) of ARROW. In both figures, the four classes are separated, but classification of unknowns is more reliable using Figure 17 because the actual data structure is more faithfully represented in the latter. Actually, the real value of these figures does not lie in classifying unknowns. Rather, the advantage stems from allowing the scientist to better understand the *type* of data he must analyze. Looking only at the knowns in Figures 16 and 17 it is clear that the two classes in CHEM are separable but not distinct. This suggests some kind of scale or continuous value moving from acids to bases as indeed is the case. It is also clear that ARROW is not of the same type, but rather consists of discrete classes.

Conclusion

Display methods are usually the first step in a pattern recognition application. They are the first contact that the scientist has with the information and, since a study of the display often determines the course of further action, they form a very important branch of pattern recognition. If the data have a simple structure, quite often the scientist can complete the application using only the display. If the data structure is complex, pattern recognition methods from the preprocessing, cluster analysis, and classification branches can be applied. If a large amount of redundancy is present, as in CHEM and ARROW, preprocessing will almost certainly involve dimensionality reduction. If known classes overlap in the display, they may still be separated in n-space by any one of a number of classification methods. Hence, display methods are most useful because they allow the scientist an approximative "look" at his data.

It is worth emphasizing a point made in the introductory paper.[2] The true value of pattern recognition is realized when several methods are used in combination as a system. Each method affords a different approach to processing the multivariant information. Thus, since the computer programs are relatively inexpensive to run, display methods nicely complement the other branches of pattern recognition. In this way, they can be used, most effectively, for displaying the results of cluster analysis and classification. All of the methods described herein are part of a collection of pattern recognition methods used at Lawrence Livermore Laboratory.

Acknowledgments. We express our thanks to J. W. Frazer and R. A. Anderson of Lawrence Livermore Laboratory for stimulating and encouraging discussions during the course of these studies. We also appreciate the helpful suggestions made by James L. Booker of the Bureau of Investigational Services, California Department of Justice, that greatly improved the CHEM data.

Molecular Orbital Theory of the Electronic Structure of Organic Compounds. XVII. Internal Rotation in 1,2-Disubstituted Ethanes

Leo Radom, William A. Lathan, Warren J. Hehre, and John A. Pople*[1]

Contribution from the Department of Chemistry, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213. *Received June 9, 1972*

Abstract: *Ab initio* molecular orbital theory is used to study internal rotation in the complete set of 1,2-disubstituted ethanes XCH_2CH_2Y (X, Y = CH_3, NH_2, OH, or F). Conformational predictions are in agreement with available experimental data. Factors which are found to influence the conformational preferences include steric, dipolar, and hyperconjugative interactions and intramolecular hydrogen bonding.

It is well known that the most stable conformations of ethane (CH_3CH_3) and monosubstituted ethanes (XCH_2CH_3) are staggered. For these molecules, the three staggered conformations generated in a 360° rotation of the methyl group are equivalent and an energy of approximately 3 kcal mol^{-1} is required for their interconversion. On the other hand for 1,2-disubstituted ethanes (XCH_2CH_2Y) rotation about the central C–C bond leads to nonequivalent staggered arrangements corresponding to a trans and a pair of gauche structures. If the internal rotation potential function has minima near these staggered arrangements, the molecule will have distinct rotational isomeric forms (rotamers). Additional rotational isomers may arise through rotation about C–X and C–Y. Although interconversion of such rotamers in general also requires relatively little energy and is therefore quite rapid at ordinary temperatures, there is substantial evidence for their separate existence. Information on the structures and relative energies of the separate rotamers and the potential barriers between them has been obtained by numerous experimental techniques including infrared, Raman, nuclear magnetic resonance and microwave spectroscopy, dipole moments, electron diffraction, electrical birefringence, ultrasonic absorption, and calorimetry.[2]

Molecular orbital theory has not yet been extensively applied to 1,2-disubstituted ethanes. The only such molecule studied by *ab initio* methods has been *n*-butane.[3,4] For monosubstituted ethanes, on the other hand, a general theoretical study has had some success in describing internal rotation.[5,6] We are therefore encouraged to apply the same method to the disubstituted systems. In this paper we give results for all the distinguishable staggered conformations of the set of saturated molecules XCH_2CH_2Y (X, Y = CH_3, NH_2, OH, F) and compare the calculated energies with experimental results where possible. This corresponds to a complete mapping of the internal rotation potential hypersurface at a 120° grid. This is too coarse to determine positions and numbers of local minima precisely, but it does indicate some of the broad features of the surfaces. For *n*-butane, *n*-propyl fluoride, and 1,2-difluoroethane, a more complete study is made. In addition, for all the molecules we consider the energy of interaction between substituents in terms of bond separation energy concepts developed earlier.[7,8]

Method

Standard LCAO–SCF molecular orbital theory[9] with the extended 4-31G basis set[10] is used. Ideally, complete optimization of bond lengths and bond angles would be desirable. However, for the relatively large set of molecules discussed here, the computation time required to do this would be too great. For one molecule (*n*-butane), we use partially optimized geometries, but in all other cases, bond lengths and angles are given the standard values listed by Pople and Gordon.[11] The results we obtain are clearly subject to the errors inherent to this approximation. All staggered conformations of the molecules XCH_2CH_2Y have been considered. The notation used to specify the rotational

(1) Author to whom correspondence should be addressed.
(2) For reviews, see: (a) S. Mizushima, "Structure of Molecules and Internal Rotation," Academic Press, New York, N. Y., 1954; (b) N. Sheppard, *Advan. Spectrosc.*, **1**, 288 (1959); (c) E. L. Eliel, "Stereochemistry of Carbon Compounds," McGraw-Hill, New York, N. Y., 1962; (d) E. L. Eliel, N. L. Allinger, S. J. Angyal, and G. A. Morrison, "Conformational Analysis," Interscience, New York, N. Y., 1965; (e) M. Hanack, "Conformation Theory," Academic Press, New York, N. Y., 1965; (f) J. P. Lowe, *Progr. Phys. Org. Chem.*, **6**, 1 (1968); (g) E. Wyn-Jones and R. A. Pethrick, *Top. Stereochem.*, **5**, 205 (1970); (h) Symposium on "Energetics of Conformational Changes," *J. Mol. Struct.*, **6**, 1 (1970).
(3) J. R. Hoyland, *J. Chem. Phys.*, **49**, 2563 (1968).
(4) L. Radom and J. A. Pople, *J. Amer. Chem. Soc.*, **92**, 4786 (1970).
(5) L. Radom, W. J. Hehre, and J. A. Pople, *ibid.*, **93**, 289 (1971).
(6) L. Radom, W. J. Hehre, and J. A. Pople, *ibid.*, **94**, 2371 (1972).
(7) R. Ditchfield, W. J. Hehre, J. A. Pople, and L. Radom, *Chem. Phys. Lett.*, **5**, 13 (1970).
(8) W. J. Hehre, R. Ditchfield, L. Radom, and J. A. Pople, *J. Amer. Chem. Soc.*, **92**, 4796 (1970).
(9) C. C. J. Roothaan, *Rev. Mod. Pyys.*, **23**, 69 (1951).
(10) R. Ditchfield, W. J. Hehre, and J. A. Pople, *J. Chem. Phys.*, **54**, 724 (1971).
(11) J. A. Pople and M. S. Gordon, *J. Amer. Chem. Soc.*, **89**, 4253 (1967).