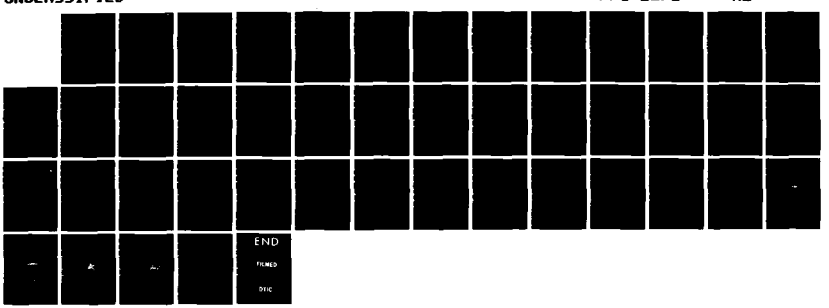
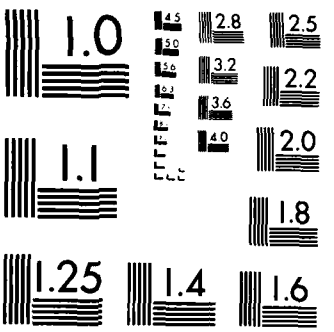
AD-A158 295 OUTER PRODUCT CALCULATIONS(U) AERODYNE RESEARCH INC BILLERICA MA JUL 85 ARI-4662 N00014-84-C-2387 1/1

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1964 A

AD-A158 295

ARI-4662

OUTER PRODUCT CALCULATIONS

MONTHLY PROGRESS REPORTS NOS. 7,8,9 & 10

Prepared By:

Aerodyne Research, Inc.
45 Manning Road
Billerica, MA. 01821

Prepared For:

Naval Research Laboratory
4555 Overlook Avd., SW
Washington, D.C. 20375

Contract No. N000~~14~~-84-C-2387

Period: March 19, 1985 - July 15, 1985

July, 1985

FILE COPY

DTIC
ELECTE
AUG 15 1985
E

85 7 [illegible]

I. PROGRESS

Progress during the mid-March to mid-June time frame was minimal since the anticipated incremental funding was not received until mid-June and since the program was redirected from optical systolic array processor outer product applications to associative memory applications.

Dr. John Gruninger and Mr. Scott Israel of Aerodyne visited NRL on April 2 in order to participate in the redirection of the Aerodyne effort. This redirection process was completed on June 21 with a joint meeting of NRL and Aerodyne personnel at Aerodyne.

It was agreed that Aerodyne efforts for the rest of FY'85 would be focused on the investigation of iterative algorithms for possible optical implementation of nonlinear novelty filters. A first cut discussion of relevant concepts has been forwarded to Dr. Lee and is attached as an addendum to this report. (See Addendum No. 1)

A report on approximate singular value decomposition techniques for ill-conditioned matrices was prepared and presented at the OSA Special Meeting on Optical Computing in Reno, NV. in March. A copy of the vugraph presentation is attached. (See Addendum No. 2)

II. FUTURE WORK

Drs. Gruninger and Putnam will continue algorithm development for novelty filter applications. They will visit NRL during the next reporting period to present progress to that point.

III. FINANCIAL STATUS

Spending for this period (four months) - $14,166

Cumulative Spending (through July 5, 1985) - $59,309.

ADDENDUM NO. 1

DTIC COPY INSPECTED 3

Acc... DTIC ... Dist A-1

AERODYNE RESEARCH, Inc.

45 Manning Road
The Research Center at Manning Park
Billerica, Massachusetts 01821 (617)663-9500

July 15, 1985

Dr. John Lee
Naval Research Laboratory
Washington, D.C. 20375.

Dear John,

Here are my thoughts and materials for further efforts.

I have included work in rough draft form on the incorporation of weights and constraints into projection operators. As I mentioned to you earlier, I have been working on subspace methods and oblique projection methods for some time now. The draft on weights shows that weighting can be used to improve the condition number of a memory matrix. A general method for weight selection may be of considerable use in making recall less sensitive to noise. The connection between constraints, generalized inverses of several varieties and subspace methods is briefly outlined in the material labeled Generalized Inverses and Constrained Projections. There are some algorithms developed that are similar to the Greville algorithm for the Moore Penrose Inverse. These should be implementable in optical hardware. These algorithms have been used for image reconstruction and restoration.

The viewgraph material shows an interesting example of constrained and weighted projection techniques in use as a novelty filter. This work was done some time ago for a different purpose and was never pursued. The example filter expects a spectrum which is a member of one of eight classes of compounds or is a mixture of members of the eight classes. The classes are:

aliphatic hydrocarbons	R
aliphatic alcohols	ROH
aliphatic chlorides	RCl
aliphatic chloro alcohols	RClO H
aromatic hydrocarbons	ϕ
aromatic alcohols	ϕOH
aromatic chlorides	ϕCl
aromatic chloro alcohols	ϕClOH

The example unknown that was testing the filter was a spectrum of 2-Chloro, 4 Nitro Benzyl Alcohol, an aromatic nitro chloro alcohol. The unknown is similar to the classes present but differs in having a nitro group. Hence there should be a novelty.

Three calculations were performed and reported in the Table on the third viewgraph of the package. The first of the three calculations are for a standard Novelty filter using a Moore Penrose pseudo inverse. The second calculation included constraints to keep the components of the estimated spectra positive. The third calculation included these constraints as well as a weighting based on the unknown itself. This last Novelty filter is not only nonlinear but also differs from unknown to unknown.

The first filter finds the Novelty as

$$N = \left(1 - \sum_k O_k\right) U$$

where N is the Novelty, U is the unknown and O_k is an Oblique projection for the kth class. $P = \sum_k O_k$ is an orthogonal projection. This filter is the standard linear novelty filter.

The second filter has constraints

$$N = \left(1 - \sum_k \hat{O}_k\right) U$$

where $$\hat{O}_k U = S_k \geq 0$$

The third filter has constraints and is weighted with the unknown

$$N = \left(1 - \sum_k \hat{O}_k(U)\right) U$$

where $\hat{O}_k(U)$ is a function of the current unknown.

The table of results gives the lengths of the component vectors of the subspaces. The "Noise" is the Novelty. The negative numbers appearing for the unconstrained filter are a result of the filter both adding and subtracting components to attempt to match the novelty.

Both the constrained and the weighted and constrained filter do a much better job. Almost the entire unknown spectrum is recovered by the weighted novelty filter. For the unknown normalized to unit length the length of the noise (novelty) is .93.

Efforts over the next few months will focus on Subspace method of Associative Recall. Key areas of interest include:

- Nonlinear novelty filters.
  Introduction of nonlinearity through constraints both on known and unknown. Introduction of nonlinearity through input pattern weighting of filter.

- Investigate iterative algorithms for possible optical implementation of nonlinear novelty filters.

Efforts in the Fall will include:

- Recall from partial pattern matching.
  Develop analysis for effects of noise and near linear dependencies on partial pattern matching. Investigate effects of weighting on reducing apparent linear dependencies.

I think this is an important area where analysis is needed. Recent work has been based on Monte Carlo models. (See G. S. Stiles and Dong Lih Denq "On the Effect of Noise on the Moore Penrose Generalized Inverse Associative Memory", IEEE TRANS PAM I, 7, 358 (1985)).

I hope this material will be of help when Bob Carter comes to see you. I will be in Eglin, Florida Thursday and Friday. I will give you a call and let you know how to reach me if you have questions then. I will have difficulty arranging a trip to NRL until after the 1st of August. I hope early in August will be alright.

Sincerely,

John Gruninger

John H. Gruninger
Senior Applied Mathematician.

JHG:sf
Encl.

References on Iterative Algorithms

Yair Censor

Finite Series Expansion Reconstruction Method, Proc. IEEE, 71, 409 (1983).

Arnold Lent and Yair Censor

"Extensions of Hildreth's Row Action Method for Quadratic Programming", SIAM J. Control and Optimization, 18, 444 (1980).

Ronald Schafer, Russell Mersereau and Mark Richards

Constrained Iterative Restoration Algorithms, Proc. IEEE 69, 432 (1981).

Tommy Elfving

On Some Methods of Entropy Maximization and Matrix Scaling
Lin. Algebra and its Applications, 34, 321 (1980).

CONSTRAINED PROJECTION OPERATORS

The representation of a class as a subspace of its features does not always contain all of the information that in known about the class. For example with spectral classes, the absorbances are always positive, but the subspaces spanned by a set of spectra include both positive and negative absorbances. In order to avoid nonphysical spectral estimates as well as incorrect classifications, it is useful to constrain the absorbances to be positive. We develop below the projection method for two types of constraints, inequality constraints and equality constraints. The constraint methods are developed in the framework of constrained least squares. The result is a set of operators for constrained oblique projection.

The solution to the least square problem can be expressed as

$$\text{minimize } (b-Ax)^T(b-Ax)$$

with respect to x. The solution $\hat{Ax}$ is given by

$$\hat{Ax} = A(A^TA)^{-1}A^Tb = Pb$$

where P is the orthogonal projector. The inclusion of constraints can be accomplished using Lagrangian techniques. Constraints of the form $Gx = 0$, equality constraints and $Gx \geq 0$ inequality constraints will be considered. The more general constraints relations $Gx \geq h$ can be converted to the above by the transformation $x = y + G^Ih$ where G^I is the pseudoinverse of G.

Equality Constraints

The least square method with equality constraints can be expressed as

$$\text{minimize } (b-Ax)^T (b-Ax)$$

subject to $Gx = 0$.

The associated Lagrangian is

$$L(x,\lambda) = \frac{1}{2} x^T A^T Ax - x^T A^T b - \lambda Gx$$

A saddlepoint solution of $L(x,\lambda)$ is obtained if

$$A^T Ax + A^T b - G^T \lambda = 0$$

$$Gx = 0$$

and $\lambda \geq 0$

λ is the vector of Lagrange Multipliers.

The solution x of the constrained problem can be given in terms of the unconstrained problem as

$$x = \hat{x} + (A^T A)^{-1} G^T \lambda$$

where λ is given by

$$\lambda = - (G(A^T A)^{-1} G^T)^{-1} G\hat{x}$$

The solution Ax can be expressed as

$$Ax = A\, Q(A^T A)^{-1} Ab = P_Q b$$

where $Q = 1 - (A^TA)^{-1}G^T[G(A^TA)^{-1}G^T]^{-1}G$

A bit of algebra will reveal that Q is idempotent and hence a projector.

$$Q^2 = Q \text{ and further } P_Q^2 = P_Q$$

Q projects the unrestricted solution $\hat{x}$ onto the constraint subspace.

The residual includes the unrestricted residual as well as the residual arising from projection onto the

$$R_Q = A(1-Q)(A^TA)^{-1}A^Tb$$

subspace that violates the constraints.

Inequality Constraints

For inequality constraints the optimization problem is

$$\text{Minimize } (b-Ax)^T(b-Ax)$$

$$\text{Subject to } Gx \geq 0$$

Following the method outlined for equality constraints the solution for x and λ are obtained from the solution of

$$(A^TA)X - A^Tb - G^T\lambda = 0$$

$$Gx \geq 0$$

$$\lambda G x = 0$$

and $$\lambda \geq 0$$

The major difference between equality and inequality constraints is that the inequality constraints need not be active. If $\hat{x}$ satisfies the constraints in that

$$G\hat{x} > 0 \quad \text{then} \quad x = \hat{x}$$

$\lambda = 0$ is the solution. In general for every component $\hat{x}_i$ of $\hat{x}$ that is greater than zero, the corresponding Lagrange multiplier λ_i is zero. The nonzero λ_i corresponds to active constraints $x_i = 0$. These correspond to the unrestricted solutions $\hat{x}_i$ violating the constraints that is $G_i\hat{x}_i \leq 0$. The nonzero λ_i are given as before x, λ and P_Q have the same properties. Q however depends on b, if for example

$$G\hat{x} = G(A^TA)^{-1}A^Tb > 0 \quad \text{then} \quad Q = 1 \quad .$$

Without loss of generality we can assume that the first k constraints are inactive and the remaining l-k constraints are active. Then G can be partitioned into $[G_1 G_2]^T$ where G_2 corresponds to the active constraints and the nonzero Lagrangian components solved for from

$$\bar{\lambda} = -\left(G_2(A^TA)^{-1}G_2^T\right)^{-1} G_2^T \hat{x}$$

The projection operates P_Q for the constrained least square problem can be considered as a sum of oblique projection operators, one for each class as in the case of unconstrained least squares.

RELATION BETWEEN GENERALIZED INVERSES AND PROJECTION METHODS

A relationship between certain generalized inverses and constrained oblique projection operators was identified. The variety of generalized inverses suggests that a large variety of potentially useful projection operators can be generated.

Oblique projectors and constraints can be considered in terms of generalized inverses of matrices. A projection operator can be written as $P_w = w\, w^I$ where w^I is the generalized inverse of w. The selection of the Moore Penrose inverse

$$w^I = (w^T w)^I\, w^T = (w^T w)^{-1}\, w^T = \Delta^{-1} w^T$$

leads to the orthogonal projector $P_w = w\Delta^{-1}w^T$. The Moore Penrose inverse is identical to the true inverse for nonsingular matrices. The existence of Δ^{-1} is based on the linear independence of the columns of w. The Moore Penrose inverse X satisfies the following four properties:

$$AXA = A \tag{1}$$

$$XAX = X \tag{2}$$

$$(AX)^T = AX \tag{3}$$

$$(XA)^T = XA \tag{4}$$

An example of a generalized inverse that does not satisfy the four conditions is

$$A^I = (Z^T A)^I\, Z^T$$

This inverse satisfies conditions 1, 2 and 4 but not 3 as

$$0 = A(Z^TA)^I Z^T \neq 0^T = Z(A^TZ)^I A^T$$

Condition (3) is equivalent to requiring $A A^I$ to be an orthogonal projection. The oblique projections themselves contain a generalized inverse. The inverse $A^I = (Z^TA)^I Z^T$ is called a 1, 2, 4 inverse.[1] In this nomenclature the Moore Penrose inverse is a 1, 2, 3, 4 inverse.

The construction of Z as

$$Z = A - B(B^TB)^{-1} (B^TA)$$

insures that Z^TA will be singular only if the subspaces spanned by A and B are not independent. If the subspaces are dependent then one or more of the columns of Z will be of zero length. For nonsingular Z^TA, the generalized inverse $(Z^TA)^I$ can be replaced with the true inverse in A^I and 0 respectively. The Moore Penrose inverse is identical to the true inverse in this case. If we let $(Z^TA)^I = Y^TZ^TA)^I Y^T$ and substitute this inverse into 0 we have

$$0 = A(Z^TA)^I Z^T = A(Y^TZ^TA)^I Y^TZ^T$$

using the Identity $(Z^TA)(Z^TA)^{-1} = 1$ yields

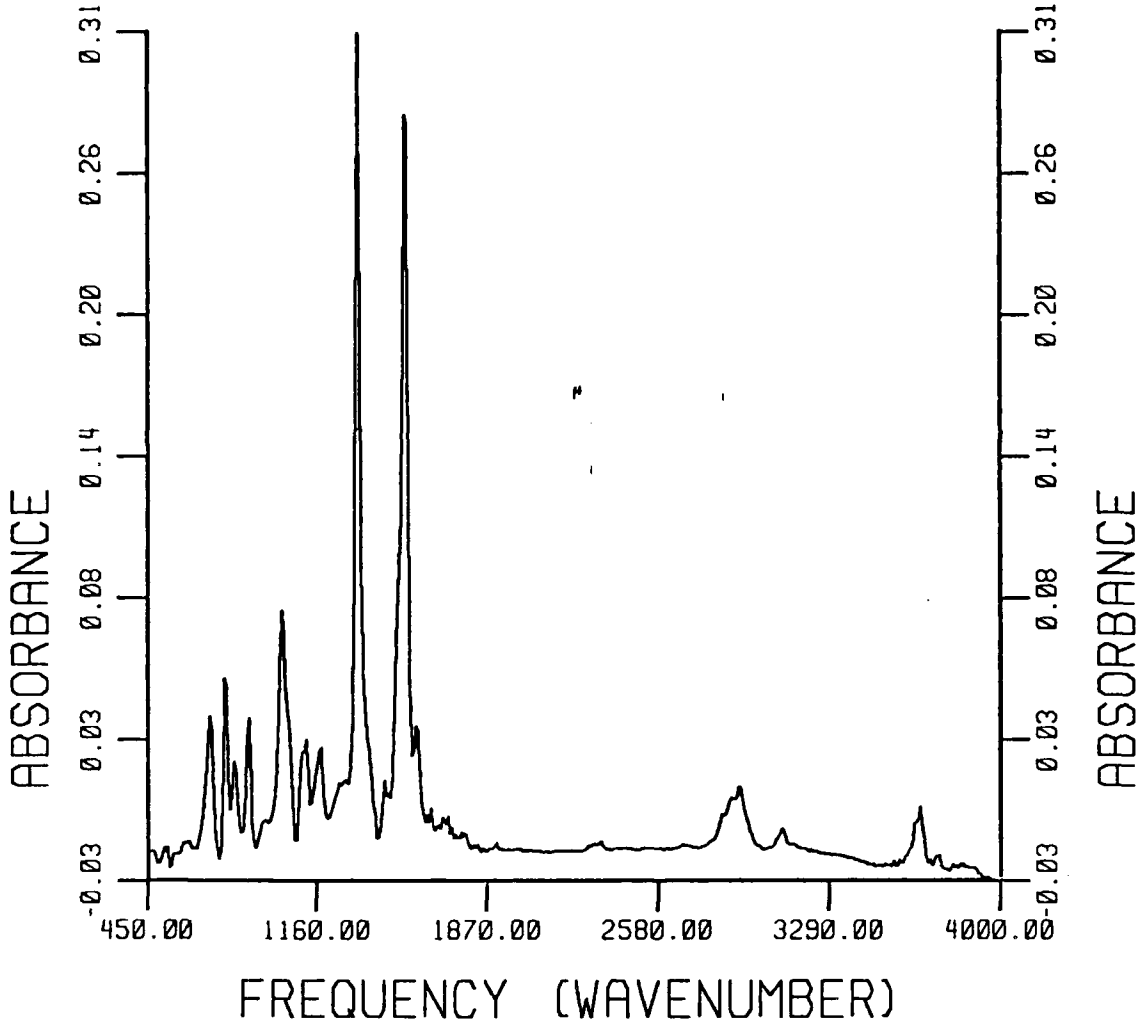
$$0 = A(Y^TZ^TA)^I(Y^TZ^TA)(Z^TA)^{-1} Z^T = AQ(Z^TA)^{-1} Z^T$$

where $Q = (Y^TZ^TA)^I(Y^TZ^TA)$ is a projection operator.

This result is a special case of the following. A generalized inverse of a nonsingular matrix X is equal to the product of an idempotent operator and the true inverse

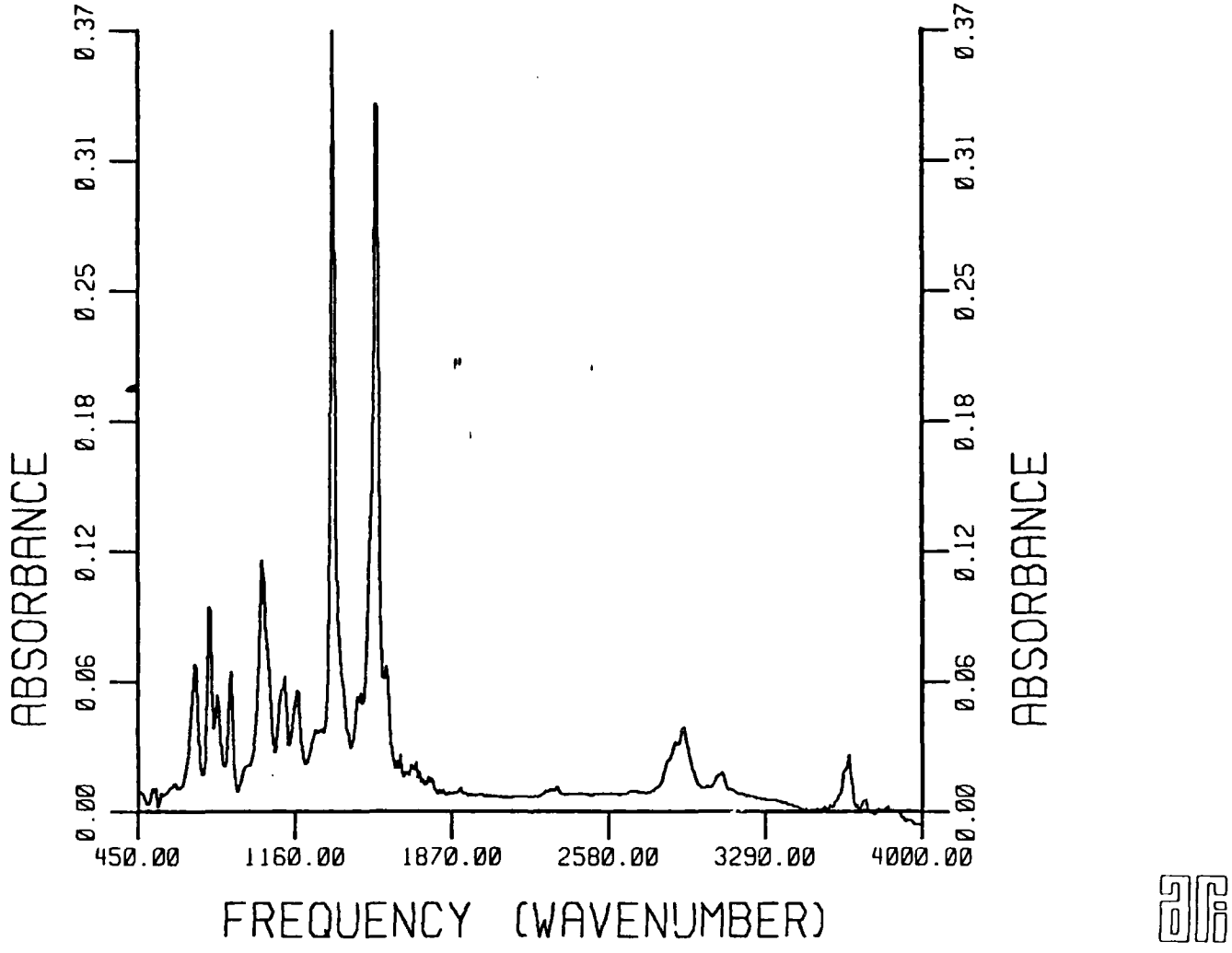
..... ERROR SPECTRUM .....

1 2-CHLORO.4 NITRO BENZYL ALCOHOL



..... UNKNOWN SPECTRUM .....

1 2-CHLORO 4 NITRO BENZYL ALCOHOL



ESTIMATED FRACTIONAL CONTRIBUTIONS TO TOTAL ABSORBANCE

THE UNKNOWN SPECTRUM IS A SPECTRUM OF 2-CHLORO, 4 NITRO BENZYL ALCOHOL

	R	ROH	RCL	RCLOH	ϕ	ϕOH	ϕCL	ϕCLOH	NOISE
NO CONSTRAINTS	-0.25	0.50	0.09	-0.04	0.18	0.01	0.17	-0.07	0.40
CONSTRAINTS	•	•	•	0.09	•	0.13	0.27	0.05	0.44
CONSTRAINTS* WITH WEIGHTING	•	•	•	•	•	•	0.07	•	0.93

1 STANDARD DEVIATION = 0.04

*WEIGHTING EQUALS UNKNOWN SPECTRUM ($Q_{II} = U_I$)

IMPORTANCE OF CONSTRAINTS AND WEIGHTING FOR ANALYSIS OF A SPECTRUM THAT IS NOT IN THE WORKING DATA BASE.

1. FREE MINIMIZATION LEADS TO

    (A) NEGATIVE ABSORBANCE

    (B) ESTIMATES OF NEGATIVE FRACTIONAL ABSORBANCES

2. WEIGHTING LEADS TO IMPROVED ESTIMATES OF FRACTIONAL ABSORBANCES

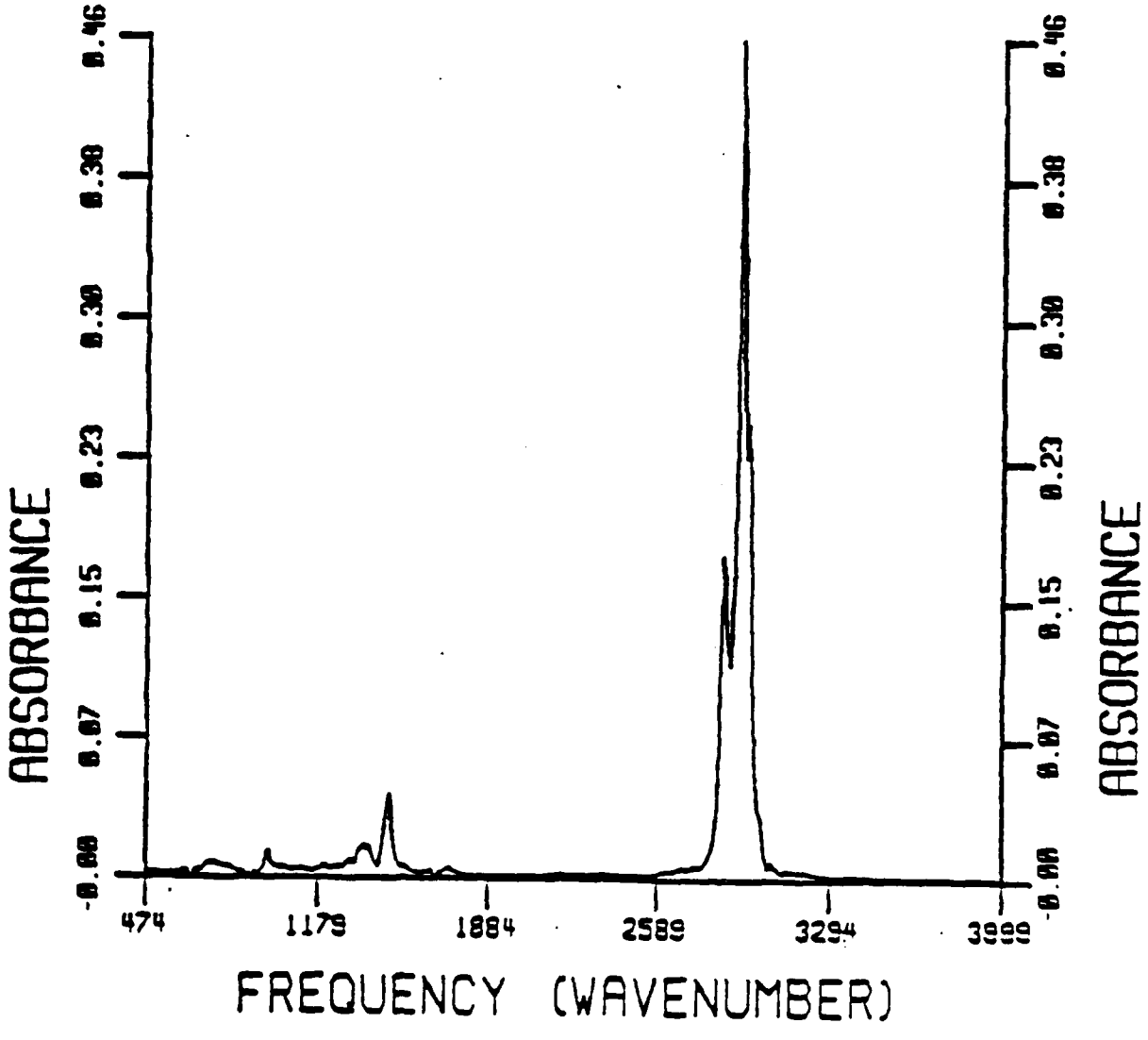
    (A) AMPLIFY IMPORTANT WAVE NUMBER REGIONS

MIXTURE ANALYSIS

- ANALYSIS OF COMPLEX MIXTURES AND DETERMINATION OF FRACTIONAL CONTRIBUTIONS FROM EACH COMPONENT
- COMPONENT SEPARATION CAN BE OPTIMIZED
- CAN BE APPLIED WHEN THERE IS A DISTRIBUTION OF REFERENCE PATTERNS ASSOCIATED WITH EACH COMPONENT
- SELECTS OPTIMAL REPRESENTER(S) FOR EACH COMPONENT
- USES CONSTRAINED OPTIMIZATION TECHNIQUES TO INSURE THAT COMPONENTS CONTRIBUTE POSITIVELY AND THAT THE OCCURRENCE OF COMPONENTS NOT IN REFERENCE SET CAN BE RECOGNIZED

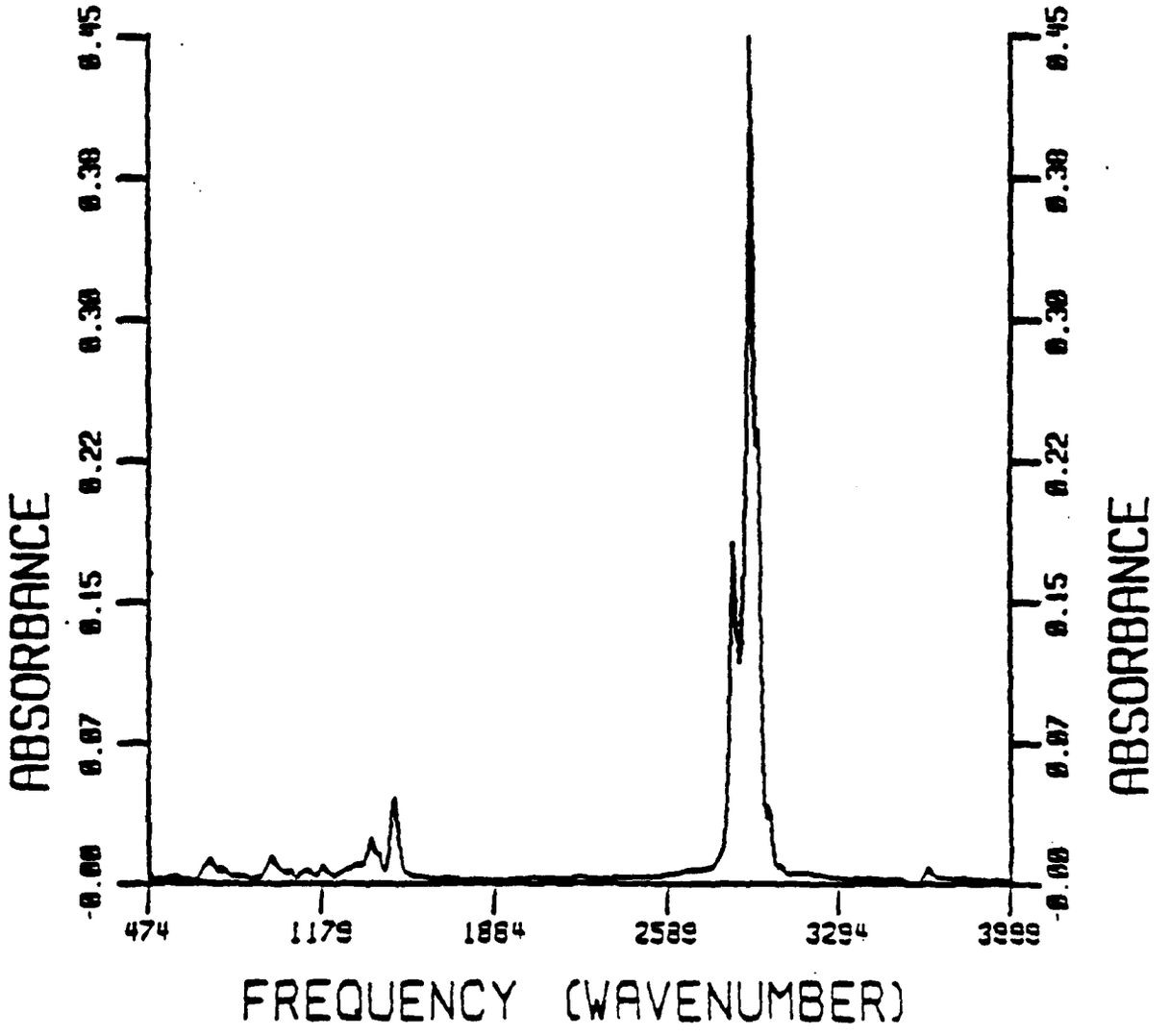
HYDROCARBON NEAREST TO ALCOHOL 1

HYDROCARBONS VERSUS ALCOHOLS



ALCOHOL NEAREST TO HYDROCARBON 1

HYDROCARBONS VERSUS ALCOHOLS



and the weighted projection is given by

$$\hat{O}_A = A(\tilde{Z}^T A)^I \tilde{Z}^T$$

$\hat{O}_A$ works directly on unweighted vectors. Thus no additional burden over unweighted projections occurs once $\hat{Z}$ and $\hat{O}$ are formed.

for I_X as was obtained in Eq. (7). The use of the incorrect weights in this model has not had an adverse effect on the solution variances. The weighting generates a numerically stable model with a well conditioned normal matrix which is easily inverted. This example offers some hope for the use of weighting schemes which minimize apparent correlation or minimize the importance of measurements which are not useful for classification. Such schemes are of obvious use in spectral pattern recognition, where spectra from different classes may differ significantly only in a few bands. This can be illustrated with a comparison of aliphatic hydrocarbons and aliphatic alcohols. Figures 1 and 2 show that several features are essentially identical and hence useless for classification purposes.

The use of weights in oblique projection techniques is attractive. The weighted projection method can be formulated so that once the projector is generated during a learning step, no additional burden will be incurred. The classification step is as simple as in the unweighted case. The form of the oblique projection operator used here is $O_A = A(Z^TA)^{-1}Z^T$ where Z is a basis for the projection onto A which is orthogonal to ϕ_B. That is $Z^TB = 0$ for any vector B that belongs to ϕ_B. When using weighted measurements a Z is sought which is W orthogonal to ϕ_B. In the unweighted case Z is constructed from

$$Z = A - B(B^TB)^{-1}B^TA$$

For the weighted case the desired $\hat{Z}$ is

$$\hat{Z} = WA - WB(B^TWB)^{-1}(A^TWB)$$

Then $$\hat{Z}^TB = A^TWB - (B^TWB)(B^TWB)^{-1}(A^TWB) = 0$$

contribute to an increase in the condition number and to the correlation between class 1 and class 2.

A weighting scheme is used which attempts to minimize the apparent correlation between classes by minimizing the condition number. The following diagonal weight matrix will cause the apparent correlation to be zero.

$$W = \begin{pmatrix} 1 & & & & 0 \\ & \frac{1}{m-1} & & & \\ & & \frac{1}{m-1} & & \\ & & & \ddots & \\ 0 & & & & \frac{1}{m-1} \end{pmatrix}$$

W is a mxm diagonal matrix. The model now uses a weighted least squares calculation. The matrix (A^TWA) is diagonal and class 1 appears uncorrelated with class 2. The condition number $\text{Cond}(A^TWA) = 1$, an absolute minimum. We have

$$(A^TWA) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \qquad (A^TWA)^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

and

$$\begin{vmatrix} X_1 \\ X_2 \end{vmatrix} = (A^TWA)^{-1} A^TWb$$

Since these weights are not the inverse of the covariance of the measurement errors, the covariance of X must be calculated from Eq. (2) rather than Eq. (3) and in principle will not yield the minimum variances. However substitution of the required quantities in Eq. (2) yields the same estimate

average vectors for class 1 and class 2 respectively. The m features are measured with equal precision and the features are uncorrelated. The covariance matrix of the errors can be expressed as $\Sigma_R = \sigma^2 I$.

$$\text{Let } A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 1 \end{pmatrix} \tag{5}$$

Even though the m measurements are of equal precision, only the first will be useful in distinguishing between the two classes. For the minimum distance classifier $X_1 > X_2$ if b belongs to class 1.

For this example

$$A^T A = \begin{pmatrix} m & m-2 \\ m-2 & m \end{pmatrix} \tag{6}$$

the condition number $\text{Cond}(A^T A) = m-1$ and the covariance can be calculated from (4) as

$$\Sigma_X = \sigma^2 (A^T A)^{-1} = \frac{\sigma^2}{4(m-1)} \begin{pmatrix} m & 2-m \\ 2-m & m \end{pmatrix} \tag{7}$$

The uncertainties in X_1 and X_2 are obtained from the diagonal elements

$$\sigma_{ii}^2 = \frac{\sigma^2}{4} \left(\frac{m}{m-1}\right) = \frac{\sigma^2}{4}$$

and are independent of m for large m. The inclusion of a large number of precisely measured but unuseful measurements does not reduce σ_{ii} but does

Gauss' Theorem states an important criteria for selection of weights. The theorem states that the weight which will give minimum variance is the inverse of the covariance of the measurement errors. That is if $W = \Sigma_R^{-1}$ then the σ^2_{jj} will be a minimum. This intuitively makes sense as measurements with large variances will have small weights and measurements with small variances will be weighted heavily. The minimum variance covariance matrix for X is given by

$$\Sigma_X = (A^T \Sigma_R^{-1} A^T)^{-1} \tag{3}$$

If $\Sigma_R = \sigma^2 I$ then

$$\Sigma_X = \sigma^2 (A^T A)^{-1} \tag{4}$$

Σ_R is usually not known and the assumption of it being a constant matrix $\sigma^2 I$ is ubiquitious since it shows no prejudice against any measurements. The unweighted least squares procedure will give an unbiased estimate of X and if the error covariance is a constant matrix, the unweighted least squares will give a minimum variance estimate of X. The penalty for using the wrong weight ($W \neq \Sigma_R^{-1}$) is the loss of minimum variance. A critical question concerning minimum variance is how sensitive is the variance to the wrong weight. (Note: no weight at all W = 1 is the wrong weight if Σ_R is not a constant matrix.) The selection of measurements or weighting of measurements is a standard approach in pattern recognition. The weighting or selection is based more on usefulness in distinguishing classes rather than on concerns about measurement error. The general rather than accidental success of such procedures would require that this type of measurement weighting does not have a large effect on the covariance. A simple example lends support to this idea. Consid - a two class problem and a simple minimum distance classifier based on b = Ax where A is m x 2 whose columns a_1 and a_2 are the

WEIGHTED FEATURES AND SUBSPACE METHODS

The incorporation of feature weighting into oblique projectors was considered. Justification for its use is provided below from a standpoint of least squares theory. The weighting of features is based on the statistical description of the measurement errors. The assumptions are that the average values of errors are zero and that the variances and covariances are known. If the model $R = b - Ax$ is adequate, the errors, R_i, associated with each row or measurement will be unbiased. By this is meant that with an ensemble of repeated measurement of b_i the set of R_i will have zero mean. Using this ensemble the covariance matrix of the errors, Σ_R, can be formed. For obvious reasons the covariance matrix is often not known. In any event the estimate of the errors in the solution vector x is related to the covariance matrix, Σ_R. The ensemble of solution vectors $\{X_i\}$ has a covariance matrix

$$\Sigma_X = A^I \Sigma_R (A^I)^T \tag{1}$$

where A^I is the pseudoinverse of A. For unweighted least squares the inverse is the Moore Penrose inverse, $A^I = (A^TA)^{-1}A^T$. For weighted least squares the pseudoinverse can be written $A^I = (A^TWA)^{-1}A^TW$, where W is the weight matrix. The weight matrix must be positive definite. The covariance matrix for the solution vector, Σ_X is in the most general formulation given by

$$\Sigma_X = (A^TWA)^{-1}(A^TW \Sigma_R WA)(A^TWA)^{-1} \tag{2}$$

It is the diagonal elements of Σ_S, σ^2_{jj}, that give the errors associated with X_j.

$$X^I = (X^I X)X^{-1} = X^{-1}XX^I$$

The left and right idempotent operators are the projectors $(X^I X)$ and (XX^I) respectively. If the Moore Penrose inverse is used the idempotent operators are the identity operators.

The projection operators developed for constrained least squares applications are also expressible in terms of 1, 2, 4 generalized inverses. The problem of minimizing $AX = b$ subject to $GX = 0$ had the solution $P_G b$ where

$$P_G = AQ(A^T A)^{-1} A^T \qquad \text{and}$$

$$Q = 1 - (A^T A)^{-1} G^T [G(A^T A)^{-1} G^T]^{-1} G$$

Substituting $H = (A^T A^{-1}) G^T$ and $S^T = G$ yields $Q = 1 - H(S^T H)^{-1} S^T$ with the 1, 2, 4 generalized inverse $(S^T H^{-1}) S^T$.

Reference

Generalized Inverses: Theory and Applications, Adi Ben-Israel and Thomas N.E. Greville, John Wiley 1974.

AERODYNE RESEARCH, Inc.

ADDENDUM NO. 2

OUTER PRODUCT EXPANSION: APPROXIMATE SINGULAR VALUE DECOMPOSITIONS

PREPARED BY

J. GRUNINGER AND H.J. CAULFIELD
AERODYNE RESEARCH, INC.
BILLERICA, MA 01821

PREPARED FOR

OPTICAL SOCIETY OF AMERICA
TOPICAL MEETING OF OPTICAL COMPUTING
LAKE TAHOE, NV

MARCH 18-20, 1985

45 Manning Road
The Research Center at Manning Park
Billerica, Massachusetts 01821 (617)663-9500

APPLICATIONS IN SPECTRAL PATTERN RECOGNITION

- DATA COMPRESSION
- FEATURE SELECTION

TRUNCATED OUTER PRODUCT EXPANSIONS

$$A = YLZ^T$$

- A MXN MATRIX
- Y AND Z ARE ORTHOGONAL. L IS A K X K NONSINGULAR MATRIX
- IF L IS DIAGONAL THEN EXPANSION IS (SVD)

$$A = Y\lambda Z^T$$

- Y IS A COLUMN BASIS OF A. Z IS A ROW BASIS OF A

SPECTRAL MEASUREMENTS

- LARGE NUMBER OF CANDIDATE FEATURES
- HIGH CORRELATION AMONG FEATURES
- CONSEQUENTLY
  - MORE CANDIDATE FEATURES THAN SPECTRA
  - NO CHARACTERIZATION OF FEATURE DISTRIBUTIONS

GENERALIZED NEAREST NEIGHBOR DIFFERENCE METHOD

- NONPARAMETRIC CLASS DISTRIBUTIONS CAN BE MULTIMODAL
- DEALS WITH UNDERSAMPLING BY EXTENDING CLASS MEMBERSHIP TO INCLUDE MIXTURES
- GENERATES AN ENTIRE SET OF ORTHOGONAL FEATURES
- FEATURES ARE DIFFERENCES BETWEEN GENERALIZED NEAREST NEIGHBOR PAIRS

NEAREST NEIGHBOR PAIRS

FIRST PAIR (A_1B_1)

A_1 IS THE MIXTURE OF CLASS A SPECTRA WHICH HAS THE STRONGEST OVERLAP WITH CLASS B

B_1 IS THE MIXTURE OF CLASS B SPECTRA WHICH HAS THE STRONGEST OVERLAP WITH CLASS A

KTH PAIR (A_K, B_K)

A_K IS THAT COMBINATION OF CLASS A SPECTRA WHICH HAS THE STRONGEST OVERLAP WITH CLASS B AND IS ORTHOGONAL TO ALL $K-1$ PAIR MEMBERS

B_K IS THAT COMBINATION OF CLASS B SPECTRA WHICH HAS THE STRONGEST OVERLAP WITH CLASS A AND IS ORTHOGONAL TO ALL $K-1$ PAIR MEMBERS

GENERALIZED NEAREST NEIGHBOR METHOD

FIND $A = Y_A L_A Z_A^T$ AND $B = Y_B L_B Z_B^T$

FORM $S_{AB} = Y_A^T Y_B$

- S_{AB} IS A $(K_A \times K_B)$ OVERLAP MATRIX

FIND $S_{AB} = U_A \lambda_{AB} U_B^T$

OR

$S_{AB} = U_A L_{AB} U_B^T$

- THE DOMINANT (LEFT/RIGHT) SINGULAR VECTORS YIELD THE DESIRED MIXTURES OF THE COLUMNS OF (A/B)
- IN GENERAL THE SUBSPACE WHERE BOTH CLASSES OVERLAP STRONGLY IS ALL THAT IS NEEDED

GENERALIZED NEAREST NEIGHBOR METHOD - CONTINUED

FIND NEIGHBOR PAIRS $W_A = Y_A U_A$ $W_B = Y_B U_B$

THEN

$$W_A^T W_A = 1 \qquad W_B^T W_B = 1$$

$$W_A^T W_B = \lambda_{AB}$$

FIND FEATURES $F(i) = W_A \cdot (1) - W_B (1)$

APPLICATION

FTIR SPECTRA

A Class of Aliphatic Hydrocarbons (1832,97)

B Class of Aliphatic Alcohols (1832,155)

Pseudo Rank A ~ 16 $A = Y_A L_A Z_A^T$

Pseudo Rank B ~ 20 $B = Y_B L_B Z_B^T$

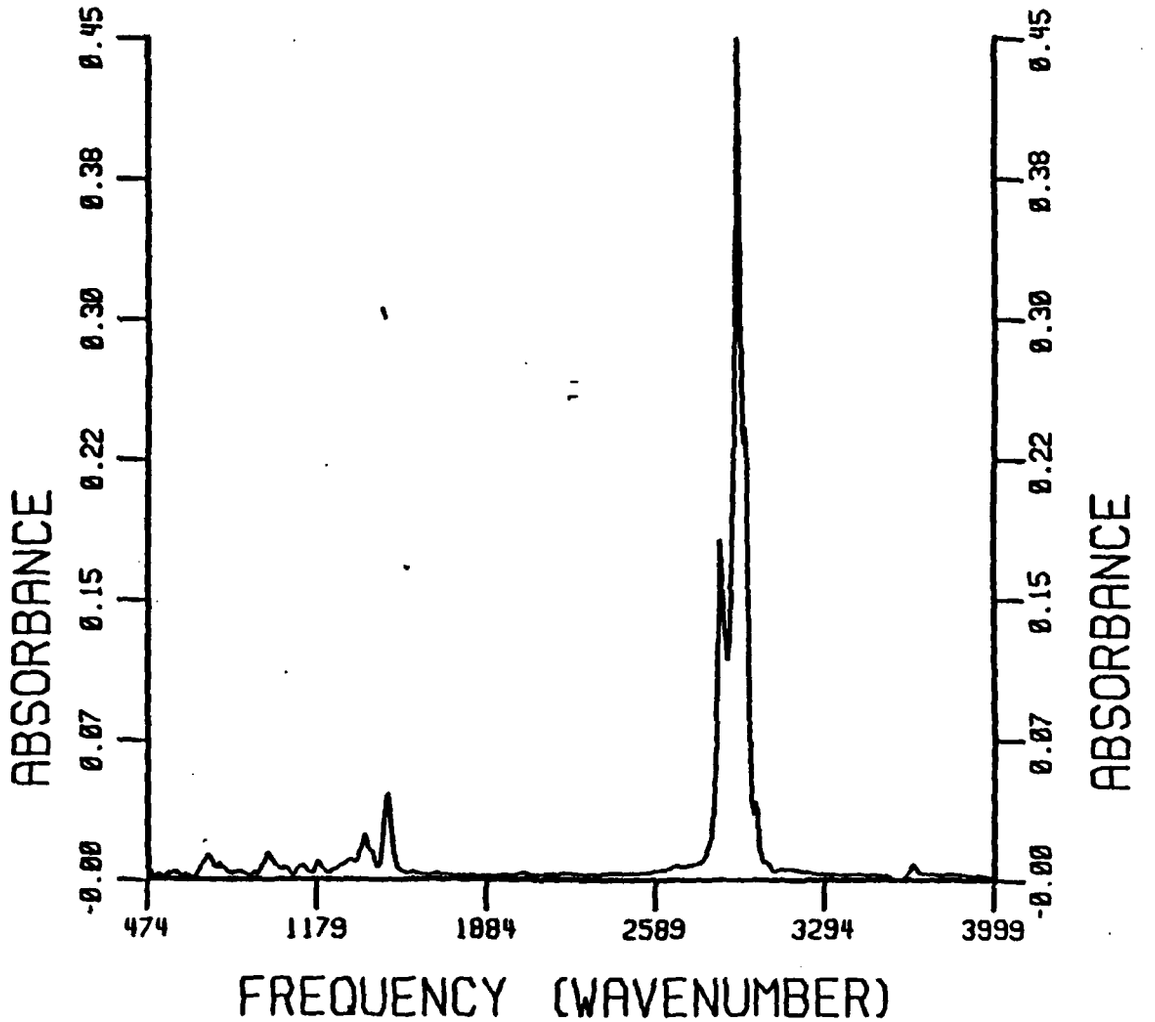
S_{AB} (16,20)

$\lambda_1 = .983$

$\lambda_2 = .978$

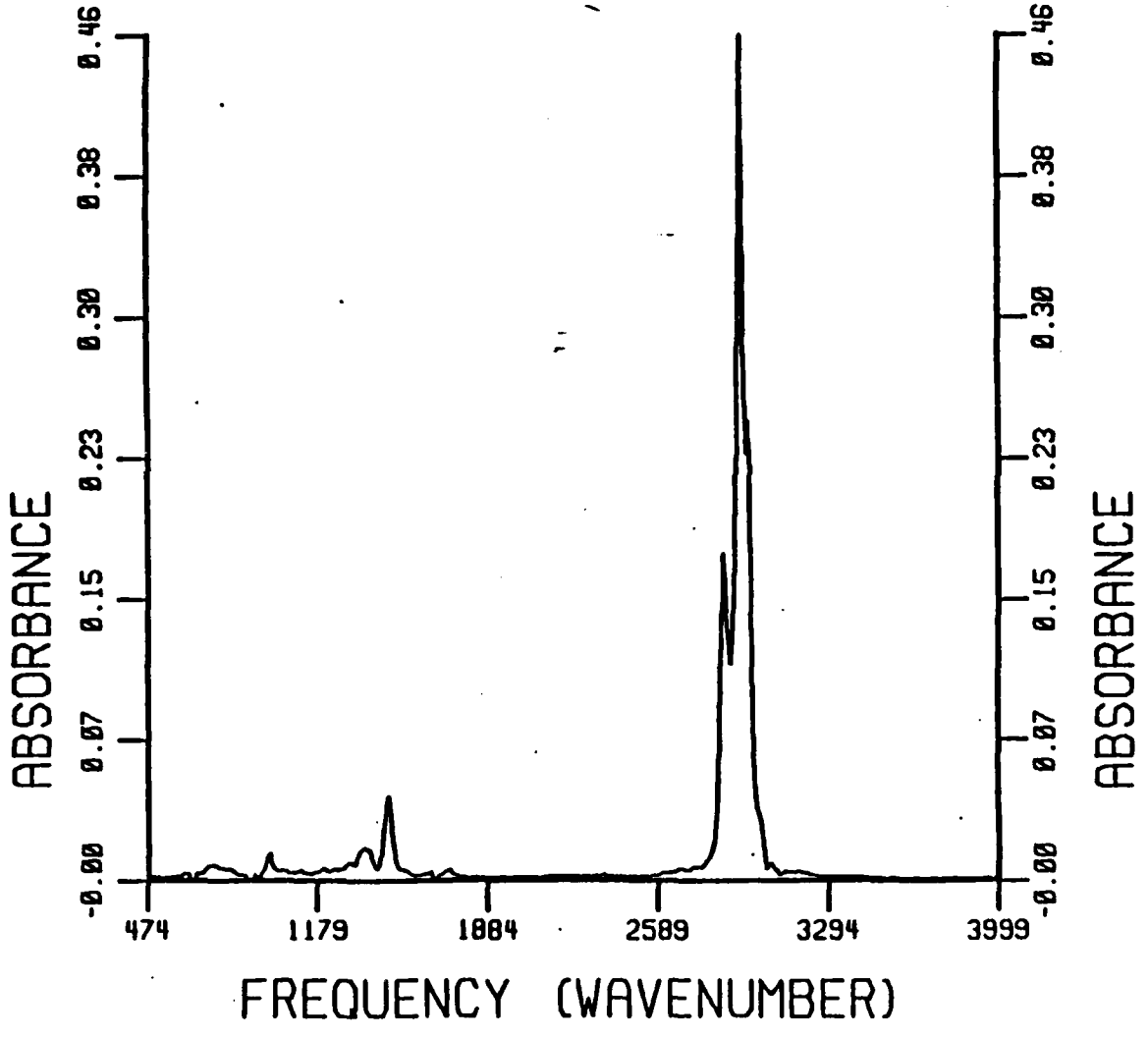
ALCOHOL NEAREST TO HYDROCARBON 1

HYDROCARBONS VERSUS ALCOHOLS



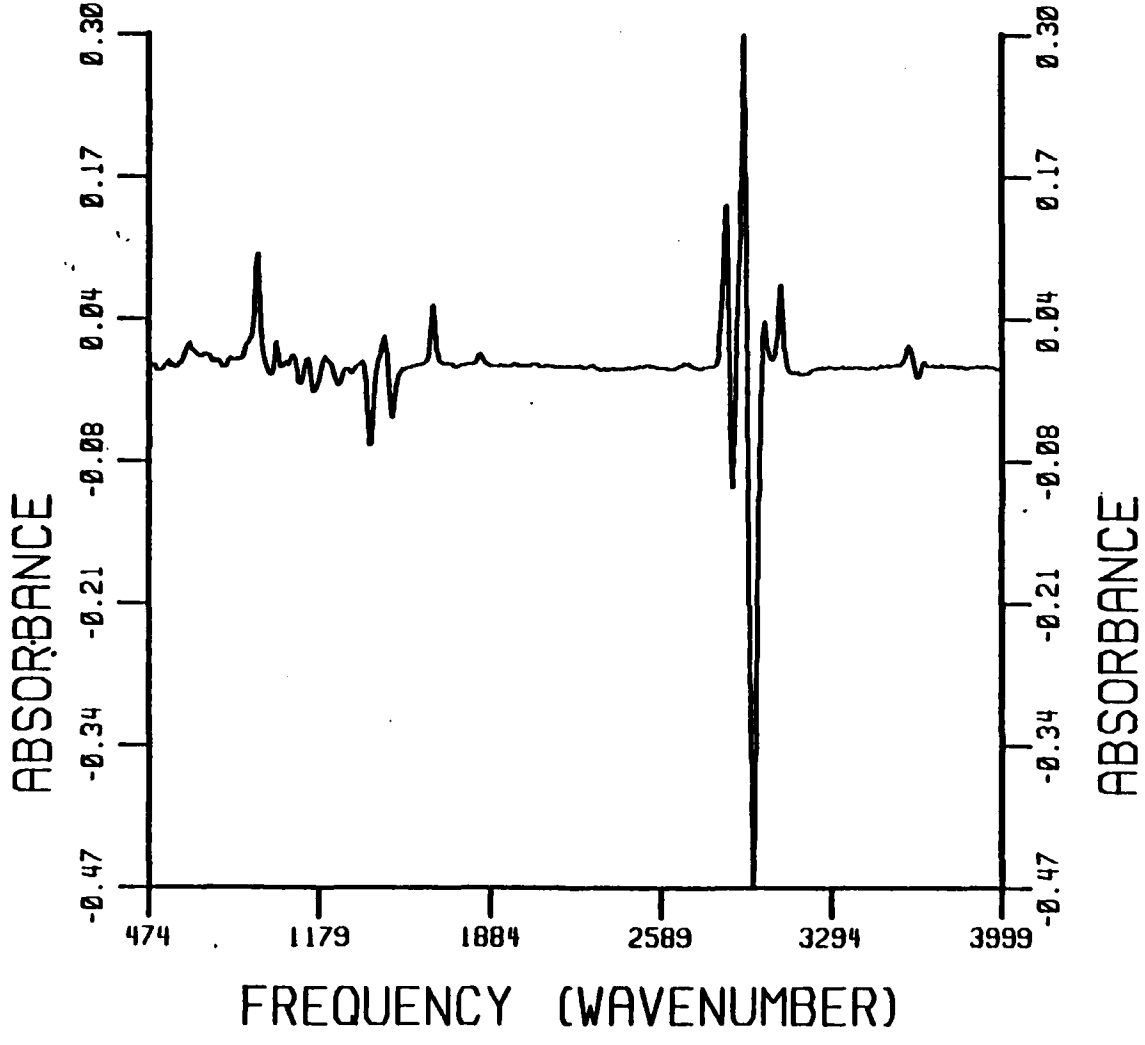
HYDROCARBON NEAREST TO ALCOHOL 1

HYDROCARBONS VERSUS ALCOHOLS



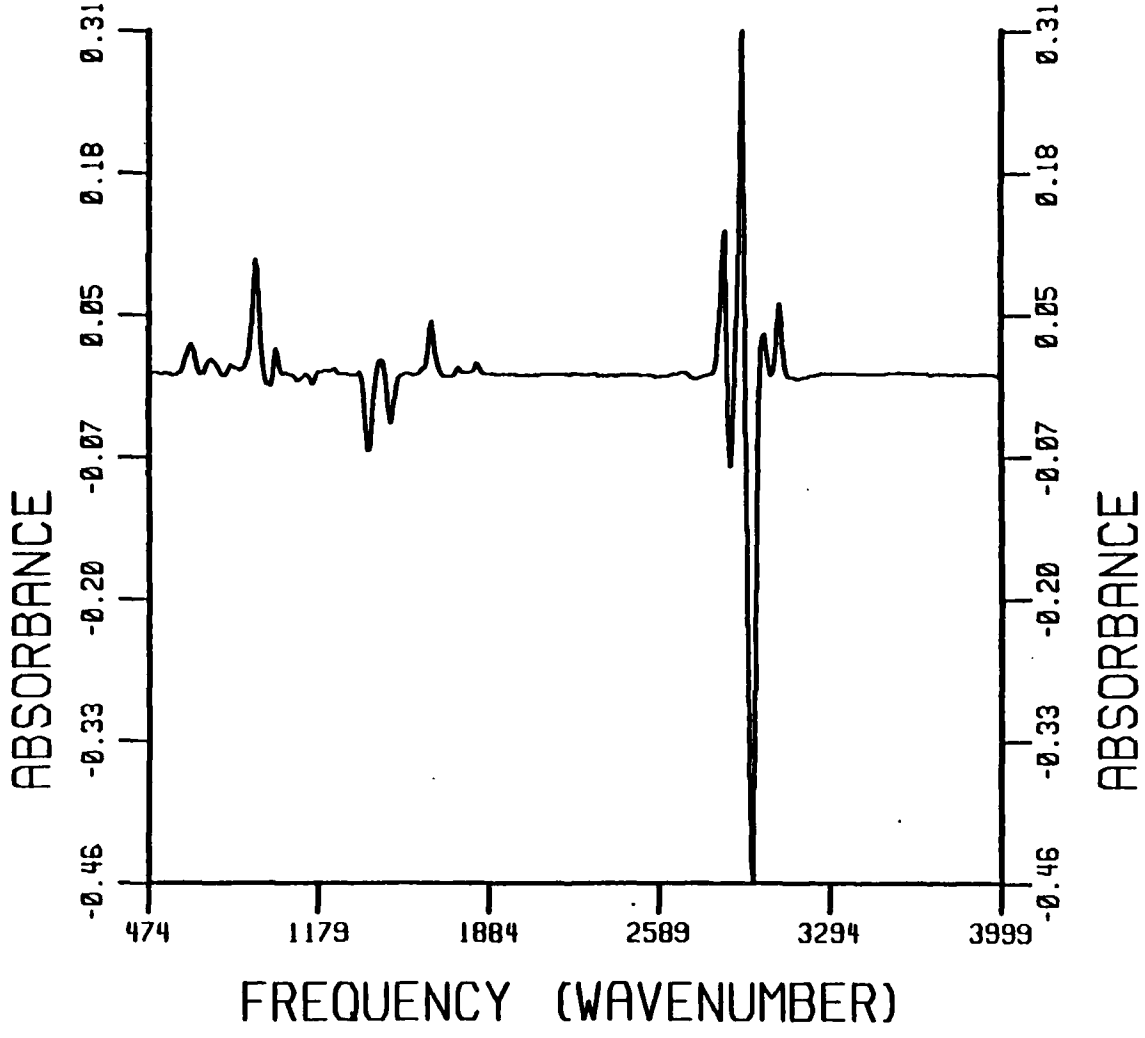
ALCOHOL NEAREST TO HYDROCARBON 2

HYDROCARBONS VERSUS ALCOHOLS



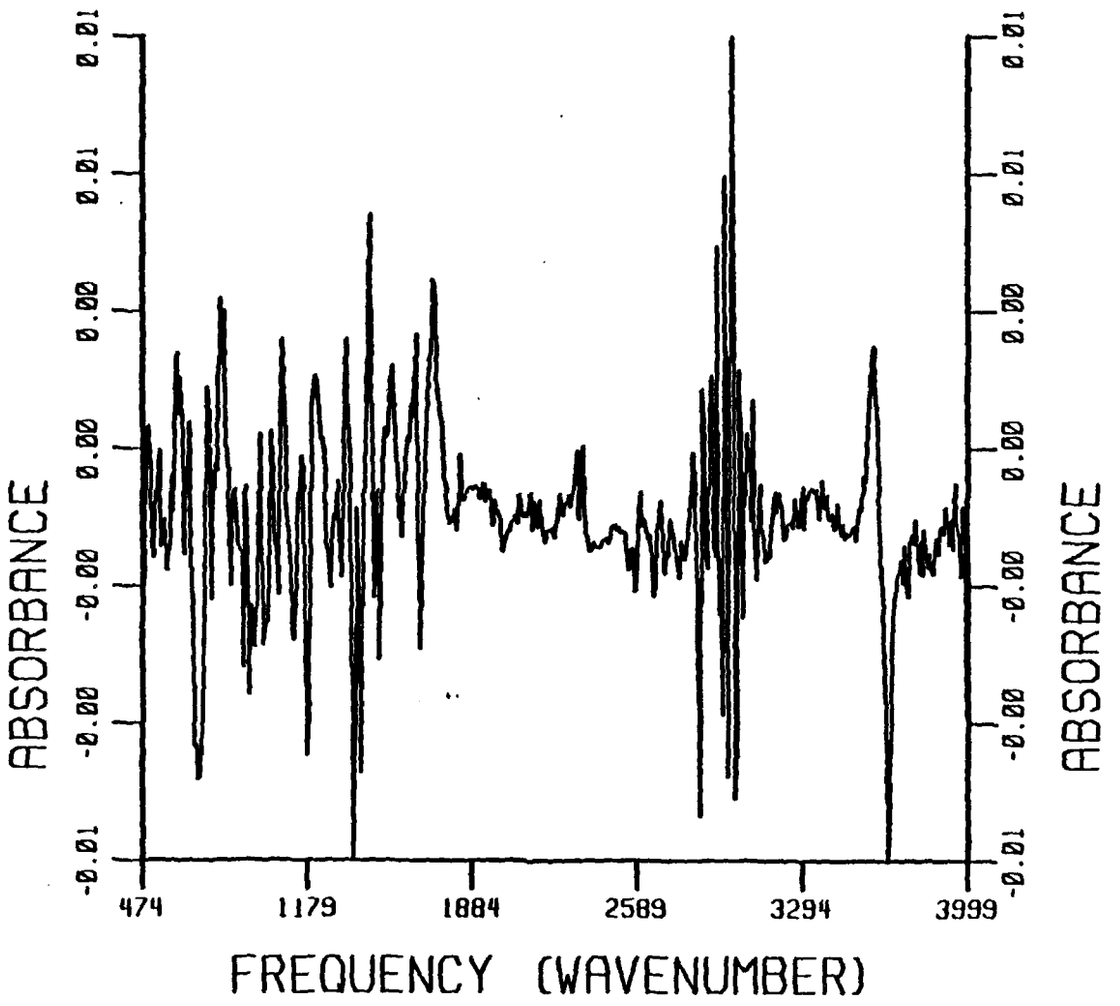
HYDROCARBON NEAREST TO ALCOHOL 2

HYDROCARBONS VERSUS ALCOHOLS



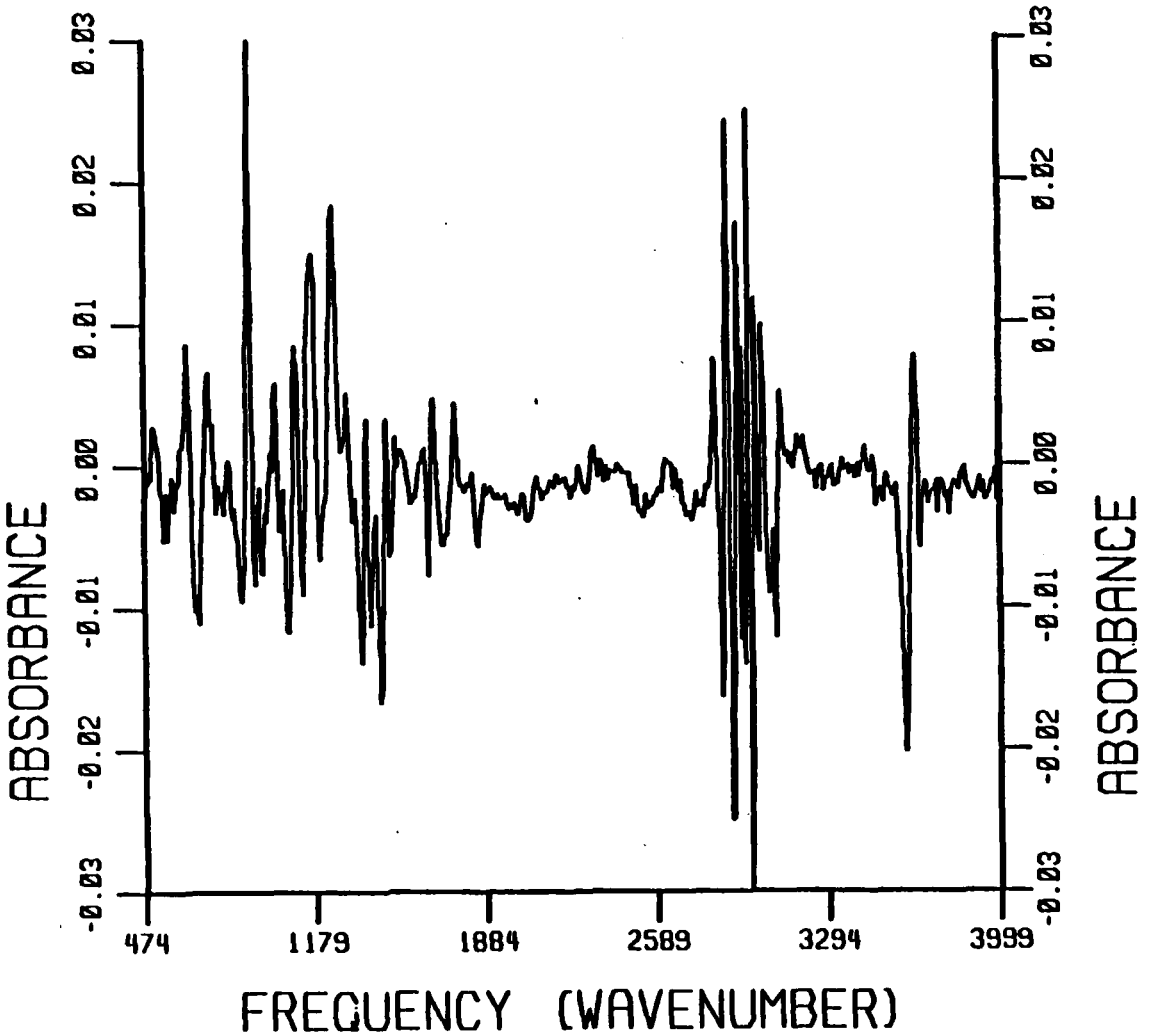
NEAREST NEIGHBOR DISTANCE NUMBER 1

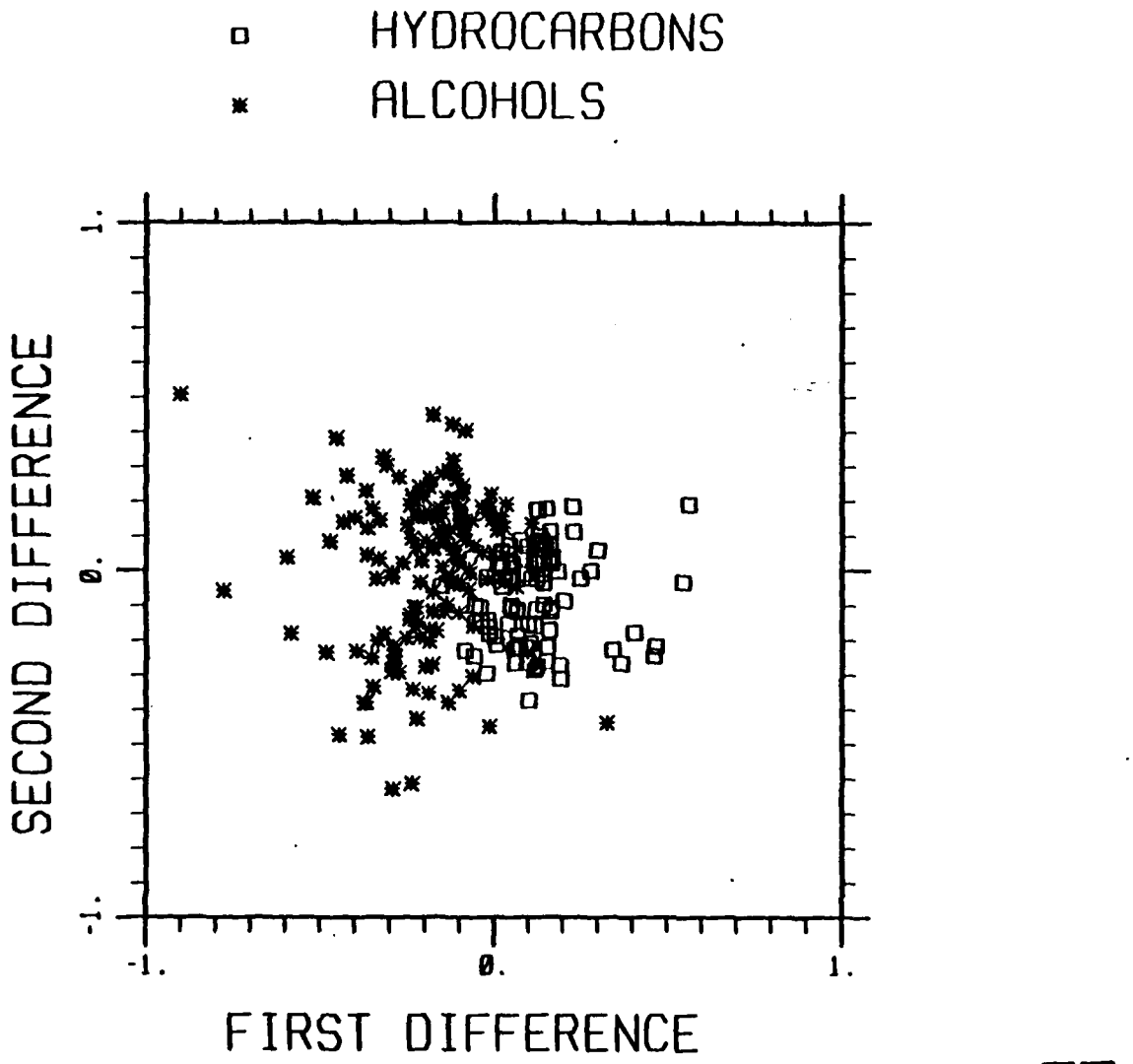
HYDROCARBONS VERSUS ALCOHOLS

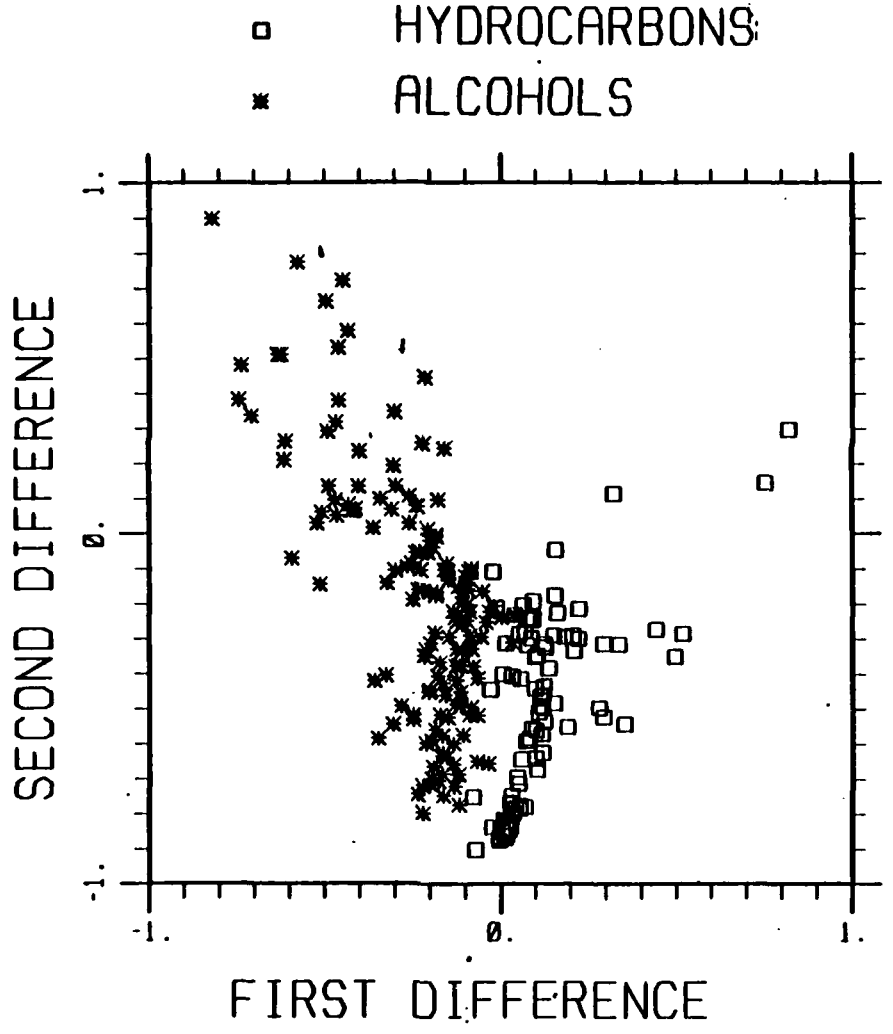


NEAREST NEIGHBOR DISTANCE NUMBER 2

HYDROCARBONS VERSUS ALCOHOLS




HYDROCARBONS
ALCOHOLS
SECOND DIFFERENCE
1.
0.
-1.
-1.
0.
1.
FIRST DIFFERENCE


HYDROCARBONS
ALCOHOLS
SECOND DIFFERENCE
1.
0.
-1.
-1.
0.
1.
FIRST DIFFERENCE

CONCLUSIONS

APPROXIMATE SVD IMPACTS OPTICAL COMPUTING THROUGH

- REPLACEMENT OF GIVEN MATRIX WITH AN APPROXIMATE MATRIX WHICH IS WELL CONDITIONED

AND

- POTENTIALLY VAST COMPUTATIONAL SAVINGS IN MATRIX-MATRIX OUTER PRODUCT MULTIPLICATION

END

FILMED

10-85

DTIC