Proceeding in this first way we will be successful only if $N-1$ is highly composite. If $N-1$ is only modestly composite, as with $N=563$, the savings of the FFT algorithm will be overcome by the fact that more than one DFT must be computed. However, there is another way we can proceed which is not subject to these limitations. The second method is based on the observation that a circular correlation or convolution where the number of points is not highly composite can be computed as a part of a circular convolution with a larger number of points. Letting N' be any highly composite integer greater than $2N-4$, we create an N' point sequence $\{b_i\}$ by inserting $(N'-N+1)$ zeros between the zeroth and first points of $\{a_{((g^{-i}))}\}$ and we create a second N' point sequence $\{c_i\}$ by periodically repeating the $N-1$ point sequence $\{\exp(-j(2\pi/N)g^i)\}$ until N' points are present. Then the inverse DFT of the product of the DFTs of $\{b_i\}$ and $\{c_i\}$ contains $\{A_{((g^k))}-a_0\}$ as a subsequence—the first $N-1$ points. Since N' can be chosen to be highly composite, even a power of two, an FFT algorithm can be used to compute the DFTs.

Using either technique, about one-third of the computation can be saved if the transform of $\{\exp(-j(2\pi/N)g^i)\}$ is precomputed. One method requires a computation proportional to $(N-1)$ times the sum of the factors of $(N-1)$ whereas the second method requires a computation proportional to $N'\log N'$. Furthermore, the summation called for in (4) and the addition of a_0 to each other A_k can each be performed with negligible additional computation by operating on intermediate quantities available when the correlation is done by FFT techniques.

III. Conclusions

While the restriction that N be a highly composite number for FFT techniques to be useful has not proved severe, it is interesting to know that it can be removed. On the other hand, the recognition that a DFT can be expressed as a convolution may be useful in itself, as this implies that a single network with fixed parameters can compute all the points of a DFT. It is expected that such diverse applications as radar beam forming and modem design may profitably use this result.

Charles M. Rader
M.I.T. Lincoln Lab.[2]
Lexington, Mass. 02173

References

[1] J. W. Cooley and J. W. Tukey, "An algorithm for the machine computation of complex Fourier series," *Math. Comput.*, vol. 19, pp. 297–301, April 1965.
[2] "What is the fast Fourier transform?" *Proc. IEEE*, vol. 55, pp. 1664–1674, October 1967.
[3] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. New York: Dover, 1965, p. 827 and pp. 864–869.
[4] T. G. Stockham, "High speed convolution and correlation," *1966 Spring Joint Computer Conf., AFIPS Proc.*, vol. 28. Washington, D.C.: Spartan, 1966, pp. 229–233.

[2] Operated with support from the U. S. Air Force.

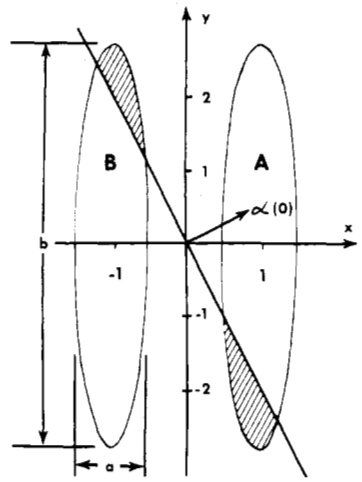

Fig. 1. Two-class ellipsoidal pattern distributions. In this example $a=1.0$ and $b=5.6$. The initial separating hyperplane, which is perpendicular to the weight vector $\alpha(0)$, identifies *incorrectly* one-tenth of the patterns (shaded region), thus $F=90.0$ percent.

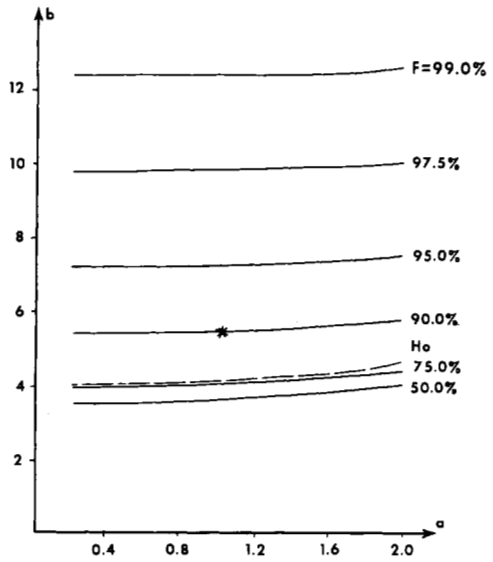

Fig. 2. Critical values of the parameters for two learning algorithms. The region below the dotted line indicates successful convergence of the Ho-Agrawala algorithm. The solid lines mark the upper limits of correct convergence, for a given initial hyperplane, of the Nagy-Shelton algorithm. In the example of Fig. 1, shown with an asterisk, the former would fail, while the latter would suceed provided the initial error rate is lower than 10.0 percent.

On a Theoretical Pattern Recognition Model of Ho and Agrawala

***Abstract*—Two versions of an unsupervised learning algorithm for pattern recognition are compared by means of numerical calculations based on two-dimensional ellipsoidal pattern distributions.**

In a recent letter, Ho and Agrawala[1] describe a theoretical model intended to explain some previously published experimental results in character recognition.[2] They call attention to the introduction of a simplifying assumption, expected to have little effect on the behavior of the recognition algorithm, to render the analysis tractable. Inspired by their observations we have calculated the performance of the algorithm, with and without this modification, for a specific family of distributions also suggested by Ho and Agrawala.

Manuscript received February 19, 1968.

[1] Y. C. Ho and A. K. Agrawala, "On the self-learning scheme of Nagy and Shelton," *Proc. IEEE (Letters)*, vol. 55, pp. 1764–1765, October 1967.
[2] G. Nagy and G. L. Shelton, Jr., "Self-corrective character recognition system," *IEEE Trans. Information Theory*, vol. IT-12, pp. 215–222, April 1966.

Using their notation, the two versions of the unsupervised learning algorithm are

$$\alpha(k+1) = XX^T\alpha(k) \tag{1}$$

$$\alpha(k+1) = X\,\mathrm{sgn}\,(X^T\alpha(k)) \tag{2}$$

where the $\alpha(k)$ are the successive approximations to the weight vector characterizing the hyperplane separating the two classes, and the columns of the X matrix are the pattern vectors to be classified.

Ho and Agrawala show that the procedure described by (1) always converges to the eigenvector associated with the largest eigenvalue of the sample covariance matrix XX^T. It will be seen, however, that the asymptotic behavior of (2) depends strikingly on the initial weight vector $\alpha(0)$.

The family of distributions considered consists of patterns uniformly distributed on two ellipses symmetrically located about the origin, as shown in Fig. 1. Ho and Agrawala's procedure (1) converges to the correct hyperplane (the y-axis) whenever the y component of the variance of the overall distribution is inferior to the x component. This imposes constraints on the relation between the major axis b and the minor axis a of the ellipses, as shown by the dotted line in Fig. 2.

The analysis of (2) may be transformed into the computation of the centroids of circular segments by appropriate dilatation of the x-axis. The parameters of the analysis are a, b, and the fraction F of the patterns correctly identified by the initial hyperplane. As seen from Fig. 2, the weight vector will converge to the x-axis (hyperplane parallel to the y-axis) whenever $b \lesssim 3.5$, regardless of the initial weight vector. Successful convergence in more difficult cases (more elongated ellipses) requires better than chance performance by the initial weight vector.

The difference between the two algorithms is, of course, the weight given to the "margin of security" of each decision in computing the contribution of a pattern to the new weight vector. Comparison of the results shows that even with a good initial weight vector it is dangerous to weight too heavily the most securely classified patterns at the expense of the rank and file. This is in accordance with the authors' experimental results as well as with the findings of Ide and Tunis[3] in modifying their perceptron-type algorithm to work in the bootstrap mode described above.

Since the curves obtained by the two methods are almost parallel, one may also conclude from Fig. 2 that whenever the initial error rate is known to exceed 24 percent it is advantageous to try Ho and Agrawala's procedure, while otherwise one can only gain by using the Nagy-Shelton algorithm.

Though the numerical results presented here hold only for the highly restricted case of two-class patterns with components uniformly distributed over equal and symmetrically located ellipses, it is hoped that this calculation, along with the work of Ho and Agrawala, adds to the store of experience necessary for the eventual use of unsupervised learning-teaching schemes in practical pattern recognition problems.

G. NAGY
N. TUONG
Département d'informatique
Université de Montréal
Montréal, Canada

Remarks by Y. C. Ho and A. K. Agrawala[4]

We believe the numerical results shown by Nagy and Tuong reinforce our original analysis. Three things were claimed in our analysis:

1) Equation (1), $\alpha(k+1) = XX^T\alpha(k)$, will always converge in the direction of maximal eigenvector of the sample covariance matrix XX^T.

2) Equation (1) will fail on a class of problems for which correct α is orthogonal to the direction of the maximal eigenvector of XX^T.

3) Equation (1) reflects the essence of the Nagy-Shelton scheme of (2).

We do not detect any disagreement in the work reported by Nagy and Tuong with our claims 1) and 2), particularly in view of the statement in their fourth paragraph. Their numerical results show that for the geometrical configuration that they chose in Fig. 1, the maximal eigenvector direction of the sample covariance matrix XX^T lies in the y-direction only if $b \lesssim 3.5$. If we accept the authors' dotted line in Fig. 2 as the boundary along which the eigenvalues of the matrix XX^T are equal, the scheme of (1) must fail, by definition, at the asterisk point. Equation (1), after all, is only a "dominant direction seeking algorithm." Although we cannot prove the convergence of (2) we feel that, as stated in claim 3), it does essentially the same thing. The data in Fig. 2 bear out this fact despite the choice of the experiment.[5] Note that the data indicate that if one starts with random initial $\alpha(0)$, then something like 80 percent of the time (2) will fail to converge to the correct hyperplane the minute one crosses the theoretical limit indicated by the dotted line. This percentage goes up rapidly as one goes away from this boundary in the direction of larger b. In other words, we have the situation shown in Fig. 3.

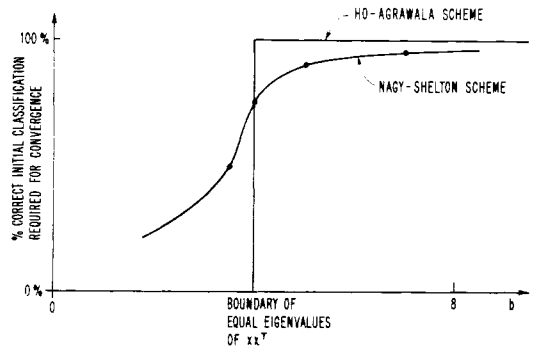

Fig. 3.

Thus, it seems that the data corroborate our predictions. On the other hand we certainly agree with the authors' contention that more experimentation on all proposed algorithms is useful as well as necessary.

Y. C. Ho
A. K. AGRAWALA
Aiken Computation Lab.
Harvard University
Cambridge, Mass. 02138

Optimization of the Frequency Response of a Fabry-Perot Interferometer

***Abstract*—An experimental technique is described which permits direct observation of the mirror irregularity profile of a plane-parallel Fabry-Perot interferometer. Using this profile as a guide, the mirror configuration which minimizes the irregularities can be found and the interferometer performance optimized. Typical results are presented.**

It is well known that the ultimate resolving power of a plane-parallel Fabry-Perot interferometer is determined by the irregularities of the mirror surfaces.[1-3] If only axially directed modes are considered in an interferometer with highly reflecting mirrors, the effect of these irregularities is to detune portions of the optical cavity from the main resonance frequency of a mode, thereby broadening the frequency-response curve of the mode. This detuning effect can easily be observed if the interferometer is excited in an axial mode by a well-collimated normally incident single-mode laser beam of wide diameter. If the beam transmitted by the interferometer is viewed on a ground-glass screen[4] while either the laser or the inteferometer is tuned through the resonance frequency of an axial mode, those portions of the mirror surfaces which transmit light at a particular portion of the frequency-response curve can be directly observed, and thus any asymmetry or poor behavior of the frequency-response curve can be immediately diagnosed. By viewing this transmitted light pattern with the mirrors in different relative positions, the optimum relative mirror setting can be found systematically, and thus the best interferometer performance can be achieved for a particular pair of mirrors.

[3] E. R. Ide and C. J. Tunis, "An experimental investigation of an unsupervised adaptive algorithm," IBM Corp., Systems Develop. Div., Endicott, N. Y., Rept. TRO1.967, July 29, 1966.

[4] Manuscript received March 11, 1968. This work was made possible through partial support to the Div. of Engrg. and Appl. Phys., Harvard University, by the U. S. Army Research Office, the USAF Office of Scientific Research, and the U. S. Office of Naval Research, under Joint Services Electronics Program Contracts Nooo14-67-A-0298-0006; and by NASA Grant NGR-22-007-068.

[5] The choice of the geometry of Fig. 1 is somewhat unfortunate due to the large separation of the centroids. Since the matrix XX^T is the covariance of the samples of *both* classes, a very large ratio of b to a is needed to create a modest ratio of the two eigenvalues. Furthermore, this ratio of the eigenvalues will critically depend upon the number of samples used in the experiment, which the authors did not mention. Thus, we feel that the example, while looking deceptively similar to one proposed in our original analysis (Ho and Agrawala,[1] Fig. 2), is not a particularly good vehicle for a test.

Manuscript received February 13, 1968. This work was supported principally by U. S. Navy Office of Naval Research Contract N00014-67-A-0204-0019.

[1] M. Born and E. Wolf, *Principles of Optics*, 3rd revised ed. Oxford and New York: Pergamon Press, 1965, p. 332.

[2] C. Dufour and R. Picca, "Sur l'interféromètre Fabry-Pérot; importance des imperfections des surfaces," *Rev. Opt.*, vol. 24, p. 19, 1945.

[3] R. Chabbal, "Recherche des meilleures conditions d'utilisation d'un spectromètre photoélectrique Fabry-Perot," *J. Rech. Centre Nat'l Rech. Sci., Labs. Bellevue* (Paris), vol. 24, pp. 138–186, 1953.

[4] Photographs of this transmission pattern made by N. J. Woolf are shown in an article by R. Chabbal, "Finesse limite d'un Fabry-Pérot formé de lames imparfaites," *J. Phys.*, vol. 19, p. 295, 1958.