Weller, A.C., Romney, A.K. (1990). *Metric scaling: Correspondence analysis.* Newsbury Park (CA): Sage.

Wilkinson, J.H. (1965). *The algebraic eigenvalue problem.* New York: Oxford University Press.

Cottrell, G.W., & Fleming, M.K. (1990). Face recognition using unsupervised feature extraction. In *Proceedings of the International Neural Network Conference*, pp 322-325. Paris, France. Dordrecht: Kluwer.

Cottrell, G.W., & Metcalfe , J. (1991). EMPATH: Face, sex and emotion recognition using holons. In R.P. Lippman, J. Moody, & D.S. Touretzky (Eds.), *Advances in neural information processing systems 3*, pp. 564-571. San Mateo, CA :Morgan Kaufmann.

Golomb, B.A., Lawrence, D.T., & Sejnowski, T.J. 1991. Sexnet: a neural network identifies sex from human face. n R.P. Lippman, J. Moody, & D.S. Touretzky (Eds.), *Advances in neural information processing system 3*, pp 572-577. San Mateo, CA :Morgan Kaufman.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* London: Academic press.

Kohonen, T., (1977). *Associative memory: A system theoretic approach.* Berlin: Springer-Verlag.

Levine, D.S. (1991). *Introduction to neural and cognitive modeling.* Hillsdale (N.J.): Lawrence Erlbaum.

Mardia, K.V., Kent, J.T., Bibby, J.M., (1979). *Multivariate analysis.* London: Academic Press.

Minsky, M.,& Papert, S. (1969). *Perceptrons: An introduction to computational geometry.* Cambridge (MA): MIT press.

Nishisato, S. (1994). *Dual scaling: An introduction to practical data analysis.* Hillsdale: Lawrence Erlbaum.

Nosofsky, R.M. (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F.G., Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale (NJ): Lawrence Erlbaum.

O'Toole, A.J., & Abdi. H. (1989). Connectionist approaches to visually based feature extraction. In G. Tiberghien (Ed.) *Advances in cognitive psychology, (Vol 2).* London: John Wiley.

O'Toole, A.J., Abdi. H., Deffenbacher, K.A., & Bartlett, J.C. (1991). Classifying faces by race and sex using an autoassociative memory trained for recognition. In K.J. Hammond, & D. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society.* Hillsdale, N. J.: Erlbaum.

O'Toole, A.J., Abdi. H., Deffenbacher, K.A., & Valentin, D. (1993). A low-dimensional representation of faces in the higher dimensions of the space. *Journal of the Optical Society of America A, 10,* 405-411.

Sirovich, L., & Kirby M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A, 4*, 519-524.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuro-Science, 3*, 71-86.

Valentin, D, Abdi, H., O'Toole, A.J., & Cottrell, G.W. (1994). Connectionist models of face processing: A survey. *Pattern Recognition*, *27*, 1208–1230.

Valentin, D., Abdi, H., & O'Toole (in press). Principal component and neural network analyses of face images: Exploration into the nature of the information available for classifying faces by sex. In C. Dowling, F.S. Roberts, & P. Theuns (Eds.), *Progress in mathematical psychology.* Hillsdale: Erlbaum.

The first step is to compute the standard singular value decomposition of the matrix

$$\mathbf{Y} = \mathbf{B}^{\frac{1}{2}}\mathbf{X}\mathbf{M}^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \quad \text{with } \mathbf{V}^T\mathbf{V} = \mathbf{U}^T\mathbf{U} = \mathbf{I}\ . \tag{46}$$

The generalized singular value decomposition is then derived from the singular value decomposition of $\mathbf{Y}$ as:

$$\mathbf{X} = \mathbf{B}^{-\frac{1}{2}}\mathbf{Y}\mathbf{M}^{-\frac{1}{2}} = \mathbf{B}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T\mathbf{M}^{-\frac{1}{2}} = \widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Delta}}\widetilde{\mathbf{V}}^T \tag{47}$$

with

$$\widetilde{\mathbf{U}} = \mathbf{B}^{-\frac{1}{2}}\mathbf{U}, \quad \widetilde{\mathbf{V}} = \mathbf{B}\mathbf{V}, \quad \widetilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}\ . \tag{48}$$

This satisfies the constraints of Eq. 45:

$$\widetilde{\mathbf{U}}^T\mathbf{M}\widetilde{\mathbf{U}} = \mathbf{U}^T\mathbf{B}^{-\frac{1}{2}}\mathbf{B}\mathbf{B}^{-\frac{1}{2}}\mathbf{U} = \mathbf{U}^T\mathbf{U} = \mathbf{I}$$

and

$$\widetilde{\mathbf{V}}^T\mathbf{B}\widetilde{\mathbf{V}} = \mathbf{V}^T\mathbf{M}^{-\frac{1}{2}}\mathbf{M}\mathbf{M}^{-\frac{1}{2}}\mathbf{V} = \mathbf{V}^T\mathbf{V} = \mathbf{I} \tag{49}$$

References

Abdi, H. (1988). A generalized approach for connectionist auto-associative memories: interpretation, implications and illustration for face processing. In J. Demongeot (Ed.) *Artificial intelligence and cognitive sciences.* Manchester: Manchester University Press.

Abdi, H. (1993). Précis de connexionisme. In J.F. Le Ny, (Ed.), *Intelligence artificielle et intelligence naturelle.* Paris: PUF.

Abdi, H. (1994a). *Les réseaux de neurones.* Grenoble: Presses Universitaires de Grenoble.

Abdi, H. (1994b). A neural network primer. *Journal of Biological Systems, 2*, 247–282.

Abdi, H., & O'Toole A.J. (submitted). *A Widrow-Hoff learning-rule for the generalization of the linear autoassociator.*

Abdi, H., Valentin, D., Edelman, B.G., & O'Toole, A.J. (submitted). *More about the difference between men and women: Evidence from linear neural networks and principal component approaches.*

Anderson, J.A., Silverstein, J.W., Ritz, S.A., & Jones, R.S. (1977) Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84*, 413-451.

Ashby, F.G. (1992). Multidimensional models of categorization. In F.G., Ashby (Ed.), *Multidimensional models of perception and cognition.* Hillsdale (NJ): Lawrence Erlbaum.

Benzécri, J.P., (1973). *L'analyse des données (2 Vol.).* Paris: Dunod.

Bruce, V. Ellis, H. Gibling, F., & Young (1987) Parallel processing of the sex and familiarity of faces. *Canadian Journal of Psychology, 41*, 510-520.

Brunelli, R., & Poggio, T. (1992). HyperBF Networks for sex classification. *Proceedings of the Image Understanding Workshop*, DARPA, San Diego, January 1992.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception, 22*, 153-176.

Table 3 shows that equalizing the first and second eigenvalues does reduce the difference in learning rate observed between the classical and generalized autoassociators. However, the generalized autoassociator is still learning faster than the classical one (56 *vs.* 83 iterations on the average). This result indicates that part of the superiority of the generalized model is indeed due to the difference in the eigenvalue ranges between the two models. However, it is not due to this difference alone.

DISCUSSION

In the present paper we have proposed a generalized version of the classical linear autoassociator. This model is of interest from both a theoretical perspective and a practical reason. First, the model makes interesting theoretical links between the neural network "learning" perspective and the statistical perspective of least-squares approximation. This analysis makes clear that, with a proper choice of constraints, generalized autoassociators can implement all of the techniques of the general linear model, including canonical correlation, discriminant analysis, and correspondence analysis. Second, numerous practical applications require the *a priori* imposition of constraints operating at the level of individual parts of the input code and at the level of individual stimuli. In the present simulations, we applied the generalized model to the task of classifying faces by sex. Our method for imposing constraints made the generalized model equivalent to correspondence analysis, which differentially weights the units of the input code directly as a function of their informational value. While this manipulation did not change the accuracy of the model on the task, it speeded up the learning considerably by biasing individual parts of the code to affect the structure of the feature space. Examples of this kind of pre-wiring constraints are common in many cognitive science applications. The simulations presented indicate that these generalized methods can yield valuable learning benefits when they are adequately utilized. The sex classification task presented here is but one example of the many possible schemes available for imposing linear constraints on a learning task. The many statistical "variations on a theme" that are commonly encountered in the literature are evidence for the ready applicability of other such schemes. The diversity of the neural network literature often makes it difficult to find a common statistical thread through the various models proposed to simulate human information processing. The generalized model we have proposed provides such a thread through the commonly used linear statistical and neural network models.

Acknowledgments. Thanks are due to June Chance and Al Goldstein for providing the faces used in the simulations and to Betty Edelman for helpful comments on a previous version of this paper.

APPENDIX

The generalized singular value decomposition can be computed from the standard singular value decomposition of a matrix. Let $\mathbf{X}$ be a $I \times K$ rectangular matrix, $\mathbf{M}$ be a $K \times K$ positive definite matrix, and $\mathbf{B}$ be a $I \times I$ positive definite matrix. The generalized singular value decomposition of $\mathbf{X}$ under the constraints of $\mathbf{M}$ and $\mathbf{B}$ is given as:

$$\mathbf{X} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Delta}}\widetilde{\mathbf{V}}^T \quad \text{with} \quad \widetilde{\mathbf{V}}^T\mathbf{M}\widetilde{\mathbf{V}} = \widetilde{\mathbf{U}}^T\mathbf{B}\widetilde{\mathbf{U}} = \mathbf{I} \; . \tag{45}$$

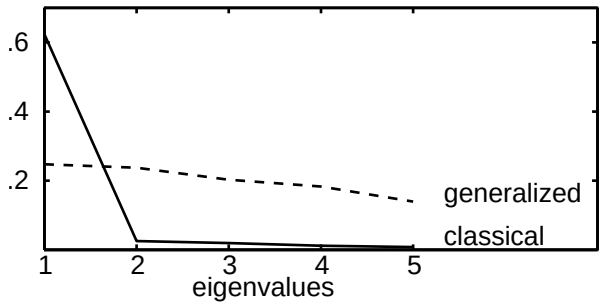

FIG. 2. — Plot of the first five eigenvalues obtained for the classical and the generalized models. Note that the difference between the first two eigenvalues is greater for the classical model than for the generalized model.

model, an average of 1933 iterations were necessary to obtain the same level of performance with the classical model. It is interesting to note that, for both models, the male faces were categorized as "male" much faster than the female faces were categorized as "female". This bias can be explained by the fact that, for the specific set of faces used in these simulations, female faces are more widely scattered around their barycenter than are male faces. In other words, when not perfectly reconstructed by the memory, some female faces are closer to the male barycenter than to the female barycenter, and hence are categorized as "male". Additional iterations are necessary to reconstruct these specific faces to a level that enables differentiation from the male faces.

Observation of the eigenvalues associated with the eigenvectors of $\mathbf{W}$ and $\widetilde{\mathbf{W}}$ shows that the difference between the first and the second eigenvalues is much smaller for the generalized model than for the classical one (cf. Figure 2).

Since, using a Widrow-Hoff learning rule amounts to equalizing iteratively the non-zero eigenvalues of the weight matrix (cf. Eq. 13, and Eq. 24) the faster learning rate of the generalized model might be due to this difference in the pattern of eigenvalues. In other words, the superiority of the generalized model in this application (*i.e.,* with the specific set of constraints used here) might result from the fact that the generalized weight matrix, after simple Hebbian learning, is already almost sphericized. To check this hypothesis, we carried out a second series of simulations in which the first and the second highest eigenvalues were systematically equalized in both models.

	Classical		Generalized	
η	Males	Females	Males	Females
$\frac{3}{2}\lambda_{\max}^{-1}$	20	50	20	30
$\lambda_{\max}^{-1}$	40	70	20	50
$\frac{1}{2}\lambda_{\max}^{-1}$	70	130	50	90

TABLE 3. — Minimum number of iterations necessary to achieve a perfect categorization using Widrow-Hoff learning as a function of the type of model and the learning constant, when the first and the second eigenvalues are equalized.

	Classical		Generalized	
η	Males	Females	Males	Females
$\frac{3}{2}\lambda_{\max}^{-1}$	500	1200	15	30
$\lambda_{\max}^{-1}$	1000	1500	20	50
$\frac{1}{2}\lambda_{\max}^{-1}$	1500	3100	50	90

TABLE 2. — Minimum number of iterations necessary to achieve a perfect categorization using Widrow-Hoff learning as a function of the type of model and the learning constant.

Hoff learning rule used to reconstruct the faces *prior* to sex classification was used as an indication of the learning speed. Specifically, the learning speed was defined as the minimum number of iterations necessary to achieve a perfect sex classification.

Stimuli: The stimuli were the 160 face images used in the first simulation.

Procedure: For both models (classical and generalized), the complete set of faces (80 males and 80 females) was stored using a Widrow-Hoff learning rule (Eq. 13 and 24, respectively) with different values of t. After each iteration, $\mathbf{W}_{[t]}$ was decomposed into its eigenvectors:

$$\begin{aligned} \mathbf{W}_{[t]} &= \mathbf{U}\mathbf{\Phi}_{[t]}\mathbf{U}^T \qquad \text{for the classical associator} \\ \mathbf{W}_{[t]} &= \widetilde{\mathbf{U}}\mathbf{\Phi}_{[t]}\widetilde{\mathbf{U}}^T \qquad \text{for the generalized associator .} \end{aligned} \tag{42}$$

The effect of Widrow-Hoff learning is equivalent to projecting the columns of $\mathbf{X}$ onto the eigenvectors of $\mathbf{W}$ followed by an expansion or dilatation of the projections as indicated by the diagonal matrix $\mathbf{\Phi}_{[t]}^{\frac{1}{2}}$. Specifically, the coordinates of the faces at time t were evaluated as:

$$\mathbf{G}_{[t]} = \mathbf{X}^T\mathbf{U}\mathbf{\Delta}^{-1}\mathbf{\Phi}_{[t]}^{\frac{1}{2}} = \mathbf{V}\mathbf{\Phi}_{[t]}^{\frac{1}{2}} \tag{43}$$

for the classical autoassociator, and as

$$\widetilde{\mathbf{G}}_{[t]} = \mathbf{X}^T\mathbf{B}\widetilde{\mathbf{U}}\widetilde{\mathbf{\Delta}}^{-1}\mathbf{\Phi}_{[t]}^{\frac{1}{2}} = \widetilde{\mathbf{V}}\mathbf{\Phi}_{[t]}^{\frac{1}{2}} \tag{44}$$

for the generalized autoassociator.

The faces were then categorized according to their sex using steps 3 and 4 of the categorization algorithm described in the previous section. This procedure was iteratively repeated until all the faces were perfectly classified. To make sure that the pattern of results we obtained was not due to a specific value of the learning constant η, three simulations were carried out using different values of η: $\frac{3}{2}\lambda_{\max}^{-1}$, $\lambda_{\max}^{-1}$, and $\frac{1}{2}\lambda_{\max}^{-1}$, with $\lambda_{\max}$ representing the highest eigenvalue.

Results: The minimum numbers of iterations necessary to achieve a perfect sex classification with both the classical and the generalized model are presented in Table 2. This Table shows that for all three learning constants, the generalized autoassociator discriminated between male and female faces much faster than the classical one did. Whereas, it took, on the average, 57 iterations to obtain a perfect sex categorization with the generalized

	Classical autoassociator		Generalized autoassociator	
# eigenvectors	Number of faces in the training set: 20			
2	.76	.73	.81	.74
10	.94	.77	.93	.76
18	1	.78	1	.76
# eigenvectors	Number of faces in the training set: 50			
2	.76	.75	.79	.78
10	.85	.78	.87	.78
48	1	.80	1	.79
# eigenvectors	Number of faces in the training set: 80			
2	.78	.80	.80	.79
10	.84	.80	.84	.80
78	1	.84	1	.84
# eigenvectors	Number of faces in the training set: 110			
2	.78	.79	.81	.79
10	.84	.78	.84	.78
108	1	.85	1	.85
	Old	New	Old	New

TABLE 1. — Proportion of correct classifications obtained with a classical autoassociator *versus* proportion of correct sex classification obtained with a generalized autoassociator as a function of the number of eigenvectors used to reconstruct the faces and of the number of faces in the training sets. For each condition, the performance is averaged across 50 trials.

Results: The average proportion of correct sex classification for both models are presented in Table 1. Observation of this Table shows that:

1) The performance of the classical autoassociator is similar to that reported in previous work. Specifically, the accuracy of categorization increases as a function of both the number of eigenvectors used to reconstruct the faces and the number of faces in the training set. The best performance (100% correct classification for the old faces and 85% for the new faces) was obtained with a training set of 110 faces and 108 eigenvectors. In previous work, using a similar categorization algorithm with a classical autoassociator, we found that with a training set of 158 faces 90% of the new faces were correctly classified as male and female (Abdi *et al.*, in preparation).

2) No substantial difference in performance accuracy can be seen between the two models.

In summary, the two models appear to be equally accurate on a sex classification task independently of the training set size and of the number of eigenvectors used to reconstruct the faces. This is not entirely surprising since the performance of the classical model is already impressive and probably difficult to improve. Comparable levels of performance were found using different models such as backpropagation networks (Cottrell & Metcalfe, 1991; Golomb, Lawrence, & Sejnowski, 1991) or HyperBF networks (Brunelli & Poggio, 1992). This general high level of sex categorization performance is probably due to the fact that sex discrimination is a linear problem.

4.2 Learning speed

The purpose of this second series of simulations was to compare the learning speed of the classical and the generalized model. The number of iterations of the Widrow-

for the classical model, and as

$$\widetilde{\mathbf{W}}_{[\infty]} = \widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T \text{ with } \widetilde{\mathbf{U}}^T\mathbf{B}\widetilde{\mathbf{U}} = \mathbf{I} \tag{35}$$

for the generalized model (cf. Eq. 13, and Eq. 24).

- *Step 2.* The projections of all faces (learned and new) onto the eigenvectors of $\mathbf{W}_{[\infty]}$ were computed as

$$\begin{aligned} \mathbf{G}_{[\infty]} &= \mathbf{X}^T\mathbf{U}\boldsymbol{\Delta}^{-1} = \mathbf{V} \quad \text{for the learned faces} \\ \mathbf{G}_{[\infty]} &= \mathbf{X}_{\text{new}}^T\mathbf{U}\boldsymbol{\Delta}^{-1} \qquad \text{for the new faces} \end{aligned} \tag{37}$$

for the classical model (cf. Eq. 9), and

$$\begin{aligned} \widetilde{\mathbf{G}}_{[\infty]} &= \mathbf{X}^T\mathbf{B}\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Delta}}^{-1} = \widetilde{\mathbf{V}} \quad \text{for the learned faces} \\ \widetilde{\mathbf{G}}_{[\infty]} &= \mathbf{X}_{\text{new}}^T\mathbf{B}\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Delta}}^{-1} \qquad \text{for the new faces} \end{aligned} \tag{39}$$

for the generalized model (cf. Eq. 27). Recall that using a Widrow-Hoff learning rule amounts to sphericizing the weight matrix. As a consequence, after complete Widrow-Hoff learning, the variance of the projections onto each eigenvector is equal to 1 and hence Eqs. 9 and 27 reduce to Eqs. 37 and 39.

- *Step 3.* For each model the coordinate vectors of the average male ($\mathbf{m}$) and female ($\mathbf{f}$) faces were computed by taking the mean of the projections of the male and female learned faces onto the L first eigenvectors (*i.e.,* the ones with the largest eigenvalues), respectively:

$$\mathbf{m} = \frac{1}{J}\sum_{j\in\{\text{male faces}\}}^{J}\mathbf{g}_j \quad \text{and} \quad \mathbf{f} = \frac{1}{J'}\sum_{j'\in\{\text{female faces}\}}^{J'}\mathbf{g}_{j'} \ . \tag{40}$$

Where J represents the number of learned male faces, J' the number of learned female faces, $\mathbf{g}_j$ the vector of the projections of the j-th male face on the first L eigenvectors, and $\mathbf{g}_{j'}$ the vector of the projections of the j'-th female face on the first L eigenvectors.

- *Step 4.* The categorization of a given face $\mathbf{x}_k$ was determined on the basis of the euclidean distance between its projection $\hat{\mathbf{x}}_k$ onto the first L eigenvectors and the average faces:

$$d(\hat{\mathbf{x}}_k, \mathbf{m}) = \|\hat{\mathbf{x}}_k - \mathbf{m}\| \quad \text{and} \quad d(\hat{\mathbf{x}}_k, \mathbf{f}) = \|\hat{\mathbf{x}}_k - \mathbf{f}\| \ . \tag{41}$$

Faces closer to the average female face were classified as female, and faces closer to the average male face were classified as male.

The number of faces *per* training set (N) ranged from 20 to 110, and the eigenvectors used to reconstruct the faces (L) varied from 2 to $N-2$ (where N is the rank of $\mathbf{W}$). To ensure that the model performance was not sample dependent, the categorization procedure was repeated 50 times for each condition.

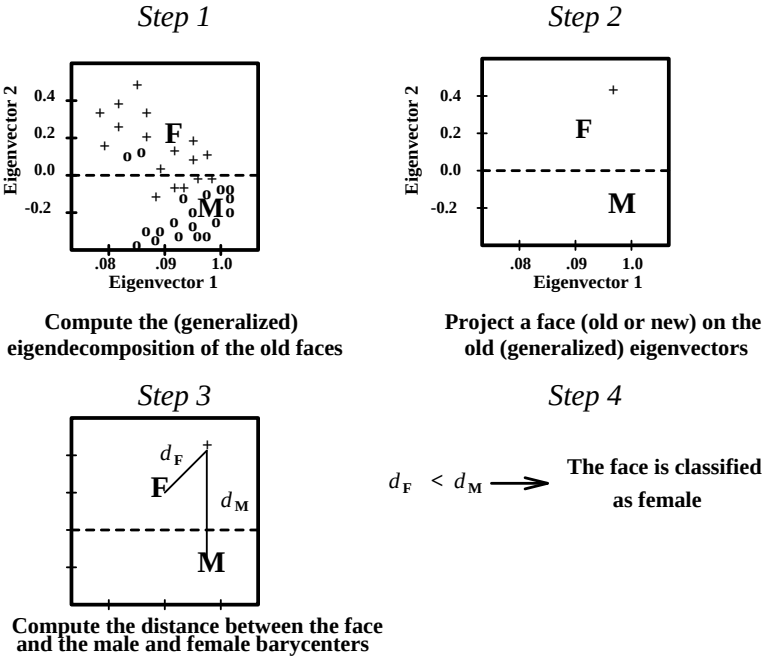

FIG. 1. — Different steps used to classify the faces according to their sex.

there is enough information in 46×31 face images to classify them accurately according to their sex (Abdi *et al.*, in preparation; Valentin *et al.*, submitted).

Procedure: Different samples of N (ranging from 2 to 110) faces were randomly selected (under the constraint that half of the faces were male and the other half female) from the original set of 160 faces and used as input for both a classical and a generalized autoassociator. The remaining faces were used to test the ability of the two models to generalize to new faces. The estimation of the sex of the faces was done by using a perceptron as a categorization network. The perceptron is a very simple neural network, and is equivalent to discriminant analysis (cf. Minsky & Papert, 1969; Levine, 1991). In the specific case of two face categories, the optimal classification procedure is equivalent to computing the coordinates of the barycenter (or center of gravity) of each class (*i.e.,* sex), and then computing the distance to both barycenters for the face to be classified. The face is then classified as belonging to the sex with the closest barycenter. Figure 1 illustrates the different steps of the categorization algorithm used in the following simulation:

- *Step 1.* For each training set, a classical and a generalized autoassociative memory were created from the face images using complete Widrow-Hoff learning and decomposed into eigenvectors. The weight matrix $\mathbf{W}$ at infinity was computed as:

$$\mathbf{W}_{[\infty]} = \mathbf{U}\mathbf{U}^T \text{ with } \mathbf{U}^T\mathbf{U} = \mathbf{I} \qquad (\mathbf{34})$$

analysis" (Benzécri, 1973; Greenacre, 1984; Weller & Romney, 1990) or as "dual scaling" (Nishisato, 1994). Strictly speaking, in correspondence analysis $m_{k,k}$ would be equal to $x_{+,k}/x_{++}$. However, since $x_{+,k} = 1$, our model is a particular case of correspondence analysis (*i.e.,* when all the columns sum to a constant).

The generalized euclidean distance associated with this technique is the so-called *Chi-square* distance. It is essentially an informational distance. When the sum of squared distances of each point to the barycenter or centroid is computed, it produces the usual Chi-square statistic used to analyze a contingency table in elementary statistics. Specifically,

$$\chi^2 = x_{+,+} \sum_k m_{k,k} d^2(\mathbf{x}_k, \mathbf{c}) = x_{+,+} \sum_k m_{k,k} (\mathbf{x}_k - \mathbf{c})^T \mathbf{B} (\mathbf{x}_k - \mathbf{c}) \ , \qquad (\mathbf{33})$$

where $\mathbf{c}$ gives the coordinates of the centroid or average face (cf. Benzécri, 1973; or Greenacre, 1984, for a proof).

To compare the classical and generalized autoassociators, two series of simulations were performed. For each simulation, faces were used as input for both a classical and a generalized autoassociator. The two autoassociators were then used to predict the sex of the faces *via* a perceptron approach. The first series of simulations evaluates the accuracy of the sex classification achieved by both models when full Widrow-Hoff learning was used. Previous work showed that the ability of the linear autoassociator to predict the sex of faces varies as a function of the number of faces in the training set and the number of eigenvectors used to reconstruct the faces (Abdi, Valentin, & O'Toole, in preparation; Valentin, Abdi, & O'Toole, submitted). Thus, the comparison between classical and generalized models was carried out using training sets of different sizes, and faces reconstructed with different numbers of eigenvectors to cover the performance range associated with these variations. The second series of simulations evaluates the number of iterations (*i.e.,* speed of learning) necessary to reconstruct the faces in a way that allows a perfect sex categorization for faces in the training sets.

In the next two sections, we will show that although there is no difference in the accuracy with which faces can be categorized in the standard and generalized autoassociative model (cf. Simulation 1), there is a vast difference in the speed with which the learning takes place in the two models (cf. Simulation 2). The generalized model learns to classify faces by sex much more quickly than does the standard autoassociator.

4.1 Accuracy of sex classification

Stimuli: A set of 160 full face pictures of young Caucasian adults, 80 females, and 80 males, was used in the following simulation. Each face was first digitized from slide using 16 gray levels to give a $151 \times 225 = 33975$ pixel image. The images were roughly aligned along the axis of the eyes so that the eyes of all faces were about the same height. None of the pictured faces had major distinguishing characteristics such as beards, glasses, or jewelry. To save processing time, each face was then compressed to a $46 \times 31 = 1426$ pixel image and coded as a 1426×1 vector $\mathbf{x}_k$ concatenated from the rows of the face image. The compression was done by local averaging using a 5×5 window. This compression technique reduces the number of pixels in an image and as a consequence filters out high frequency information. However, this is not a problem since we demonstrated earlier that

4. Categorizing faces by sex

In recent years, a number of connectionist models have been applied to the problems of face recognition and categorization (for a review see Valentin, Abdi, O'Toole, & Cottrell, in press). These models represent faces explicitly (Turk & Pentland, 1991; Sirovich & Kirby, 1987) or *via* a neural network architecture (Cottrell & Fleming, 1991; O'Toole & Abdi, 1989) in terms of the eigendecomposition of a matrix storing pixel-based descriptions of faces. The eigenvectors, in this framework, can be thought of as a set of features from which the faces are built. Likewise, the projections of the faces onto the eigenvectors can be interpreted as an indication of the extent to which each eigenvector characterizes individual faces. This type of approach suggests that faces can be efficiently represented using tools derived from multivariate statistical analysis. Specifically, previous work showed that complex perceptual discrimination such as the categorization of faces along visually derived dimensions (*e.g.,* sex, race, age, *etc.*) can be achieved by a simple linear autoassociator (O'Toole, Abdi, Deffenbacher, & Bartlett, 1991; O'Toole, Abdi, Deffenbacher, & Valentin, 1993). Among these perceptual categorization problems, sex classification is one of the most biologically important and probably one of the easiest and fastest categorizations made by human beings. For example, Bruce, Ellis, Gibling, and Young (1987), reported an average sex categorization time of 613 ms for unfamiliar faces and 620 ms for familiar faces. In a more recent study, Burton, Bruce, and Dench (1993) reported that human subjects were able to classify photographs of 179 adults with respect to sex with 96% accuracy, even though the hair was concealed by a swimming cap. In this section, a generalized linear autoassociator is applied to the problem of categorizing faces according to their sex. To evaluate the usefulness of the generalized model, the performance of the generalized autoassociator is compared with the performance of a standard autoassociator on the same task.

In this specific application, the main idea was to have the cells of the memory take on differential importance as a function of their use. Specifically, each cell responds as the inverse of its use during the learning period. The rationale, behind this coding scheme, is to make the cells more discriminative. So, for example, if a cell is active for all the faces, it does not provide any information about a subset of specific faces. On the other hand, a cell that is active relatively rarely should be important for the detection of the subset of faces that triggers its activity. Formally, this is equivalent to defining $\mathbf{B}$ as being a $I \times I$ diagonal matrix with:

$$b_{i,i} = \frac{x_{i,+}}{x_{++}} \qquad (\mathbf{32})$$

where $x_{i,+}$ represents the total of the i-th row of the face matrix $\mathbf{X}$ and x_{++} represents the grand total of $\mathbf{X}$.

Since, there was no *a priori* reason to give more importance to some faces than to others, the face vectors were normalized (*i.e.,* pre-processed) so that the sum of the pixels representing each face was equal to one (*i.e.,* $\sum_i x_{i,k} = 1$, and $\mathbf{M} = \mathbf{I}$). In addition, to giving an identical importance to each stimuli in the learning set, this particular pre-processing has the advantage of transforming the matrix $\mathbf{X}$ into a "profile" matrix (*i.e.,* each column of $\mathbf{X}$ adds up to 1). With this specific choice for $\mathbf{B}$ and $\mathbf{M}$, the generalized autoassociator implements the multivariate statistical analysis known as "correspondence

where $\widetilde{\mathbf{U}}$ are the generalized eigenvectors of $\mathbf{W}$.

The generalized eigenvectors and eigenvalues of the weight matrix $\mathbf{W}$ can be obtained directly using a generalization of the singular value decomposition of the matrix of stimuli $\mathbf{X}$. Formally:

$$\mathbf{X} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Delta}}\widetilde{\mathbf{V}}^T \qquad \text{with } \widetilde{\mathbf{V}}^T\mathbf{M}\widetilde{\mathbf{V}} = \widetilde{\mathbf{U}}^T\mathbf{B}\widetilde{\mathbf{U}} = \mathbf{I} \tag{26}$$

where $\widetilde{\mathbf{U}}$ represents the matrix of generalized eigenvectors of $\mathbf{X}\mathbf{X}^T$, $\widetilde{\mathbf{V}}$ represents the matrix of generalized eigenvectors of $\mathbf{X}^T\mathbf{X}$, and $\widetilde{\mathbf{\Delta}}$ is the matrix of generalized singular values. The projections, $\widetilde{\mathbf{G}}$, of the K stimuli of the training set on the R eigenvectors of the weight matrix can be found as:

$$\widetilde{\mathbf{G}} = \mathbf{X}^T\mathbf{B}\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}\widetilde{\mathbf{\Delta}} \; . \tag{27}$$

From Eq. 27, it is easy to derive that the generalized variance of the projections on one eigenvector is equal to the eigenvalue associated with this eigenvector:

$$\widetilde{\mathbf{G}}^T\mathbf{M}\widetilde{\mathbf{G}} = \widetilde{\mathbf{\Delta}}\widetilde{\mathbf{V}}^T\mathbf{M}\widetilde{\mathbf{V}}\widetilde{\mathbf{\Delta}} = \widetilde{\mathbf{\Lambda}} \; . \tag{28}$$

Likewise, the projections of a set of K' new stimuli (*i.e.,* the test set), $\mathbf{X}_{\text{new}}$, on the eigenvectors of $\mathbf{W}$ can be computed as:

$$\widetilde{\mathbf{G}}_{\text{new}} = \mathbf{X}_{\text{new}}^T\mathbf{B}\widetilde{\mathbf{U}} \; . \tag{29}$$

In terms of distances, the generalized autoassociator represents the stimuli using their generalized euclidean distance. The generalized euclidean distance between stimuli k and k' is computed as

$$d^2_{\mathbf{B}}(\mathbf{x}_k, \mathbf{x}_{k'}) = (\mathbf{x}_k - \mathbf{x}_{k'})^T\mathbf{B}(\mathbf{x}_k - \mathbf{x}_{k'}) \; . \tag{30}$$

The distance between stimuli k and k' can be expressed also, through the eigendecomposition of the generalized weight matrix as

$$d^2_{\mathbf{B}}(\mathbf{x}_k, \mathbf{x}_{k'}) = d^2_{\mathbf{B}}(\tilde{\mathbf{g}}_k, \tilde{\mathbf{g}}_{k'}) = (\tilde{\mathbf{g}}_k - \tilde{\mathbf{g}}_{k'})^T(\tilde{\mathbf{g}}_k - \tilde{\mathbf{g}}_{k'}) \tag{31}$$

where $\tilde{\mathbf{g}}_k$ (respectively $\tilde{\mathbf{g}}_{k'}$) is the vector of the $\mathbf{B}$-projections of stimulus k (respectively k') onto the generalized eigenvectors.

Generalized euclidean distances are widely used in a variety of applications. For example, Nosofsky (1992, p365, *ff.*, Eq. 1 and 3, see also Ashby, 1992), represents stimuli in his generalized context model (GCM) with a parameter standing for the strength of a stimulus, and with features weighted by an attentional parameter. Adapting his notation to the present paper, Nosofsky's model can be seen as equivalent to representing the strength of a stimulus by the diagonal terms $m_{k,k}$ of $\mathbf{M}$ ($\mathbf{M}$ being a diagonal matrix in this case), and the attentional weights by the diagonal terms $b_{i,i}$ of $\mathbf{B}$ ($\mathbf{B}$ being diagonal also). Categorization can be then considered to be a function of the generalized distance to the centers of the categories. Another relatively well-known example of a generalized euclidean distance is the "Mahalanobis" distance used in conjunction with discriminant analysis. In this case, the matrix $\mathbf{B}$ is the inverse of the between-features (or dimensions) correlation matrix.

matrix $\mathbf{B}$ before recall by multiplication through $\mathbf{W}$. Specifically, recall of a given stimulus $\mathbf{x}_\ell$ is obtained as:

$$\hat{\mathbf{x}}_\ell = \mathbf{W}\mathbf{B}\mathbf{x}_\ell \ . \tag{19}$$

If the stimuli stored in the memory do not form a $\mathbf{B}$-orthogonal set, recall will not be perfect. The memory will add some noise (or cross-talk) to the original stimulus:

$$\begin{aligned}\hat{\mathbf{x}}_\ell &= \mathbf{W}\mathbf{B}\mathbf{x}_\ell \\ &= \sum_k m_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{B}\mathbf{x}_\ell \\ &= m_\ell \mathbf{x}_\ell {\mathbf{x}_\ell}^T \mathbf{B}\mathbf{x}_\ell + \sum_{k\neq\ell} m_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{B}\mathbf{x}_\ell \\ &= m_\ell \mathbf{x}_\ell {\mathbf{x}_\ell}^T \mathbf{B}\mathbf{x}_\ell + \sum_{k\neq\ell} m_k \cos_{\mathbf{B}}(\mathbf{x}_k \mathbf{x}_\ell)\mathbf{x}_k \ .\end{aligned} \tag{20}$$

The quality of reconstruction of the stimulus can be evaluated using the generalized cosine between $\hat{\mathbf{x}}_\ell$ and $\mathbf{x}_\ell$ (cf. Eq. 17).

If every pair of stimuli in the learning set is $\mathbf{B}$-orthogonal, then the output of the memory will be proportional to the original stimulus:

$$\begin{aligned}\hat{\mathbf{x}}_\ell &= m_\ell \gamma_\ell \mathbf{x}_\ell + \sum_{\ell\neq k} m_k \cos_{\mathbf{B}}(\mathbf{x}_k \mathbf{x}_\ell)\mathbf{x}_k \\ &= m_\ell \gamma_\ell \mathbf{x}_\ell\end{aligned} \tag{21}$$

with γ_ℓ being a scalar equal to ${\mathbf{x}_\ell}^T \mathbf{B}\mathbf{x}_\ell$. When the stimuli stored in the memory are not $\mathbf{B}$-orthogonal, some patterns will be perfectly reconstructed by the memory:

$$\mathbf{W}\tilde{\mathbf{u}}_r = \tilde{\lambda}_k \tilde{\mathbf{u}}_r \quad \text{with: } \tilde{\mathbf{u}}_r^T \mathbf{B}\tilde{\mathbf{u}}_r = 1 \ . \tag{22}$$

The vectors $\tilde{\mathbf{u}}_r$ are the "generalized eigenvectors" of $\mathbf{W}$ (these generalized eigenvectors can be computed using a standard eigendecomposition routine, cf. Wilkinson, 1965, and Appendix). Since the eigenvectors are $\mathbf{B}$-orthogonal, and the eigenvalues are non-negative, the matrix $\mathbf{W}$ can be reconstructed as:

$$\mathbf{W} = \sum_r^R \tilde{\lambda}_r \tilde{\mathbf{u}}_r \tilde{\mathbf{u}}_r^T = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{U}}^T \quad \text{with} \quad \tilde{\mathbf{U}}^T \mathbf{B}\tilde{\mathbf{U}} = \mathbf{I} \ . \tag{23}$$

Similarly, the Widrow-Hoff error correction learning rule (cf. Eq. 13) can be generalized and $\mathbf{W}$ at time $t+1$ can be expressed as:

$$\begin{aligned}\mathbf{W}_{[t+1]} &= \mathbf{W}_{[t]} + \eta(\mathbf{X} - \mathbf{W}_{[t]}\mathbf{B}\mathbf{X})\mathbf{X}^T \\ &= \tilde{\mathbf{U}}[\mathbf{I} - (\mathbf{I} - \eta\tilde{\mathbf{\Lambda}})^{t+1}]\tilde{\mathbf{U}}^T \ .\end{aligned} \tag{24}$$

Abdi and O'Toole (submitted) show that with a learning constant η smaller than $2\lambda_{\max}^{-1}$ this procedure converges toward:

$$\mathbf{W}_{[\infty]} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \tag{25}$$

3. Generalized autoassociator

3.1 Notation and definition

First, the differential importance and non-independence of both the stimuli and the cells of the memory allowed by the generalization is formalized as two sets of weights that correspond to the importance of individual stimuli and individual units (*i.e.,* features describing the stimuli or equivalently memory cells) respectively. Specifically, the set of constraints imposed on the units is represented by a positive-definite matrix of order $I \times I$ denoted $\mathbf{B}$. For example, if we want the importance of a unit to be inversely proportional to its use, $\mathbf{B}$ will be defined as the diagonal matrix of the inverse column margin of matrix $\mathbf{X}$ (*i.e.,* $b_{i,i} = x_{i+}^{-1}$ with x_{i+} representing the total of the i-th row of $\mathbf{X}$, or, equivalently $x_{i+} = \sum_k x_{i,k}$). The set of constraints imposed on the stimuli are represented by a positive-definite matrix of order $K \times K$ denoted $\mathbf{M}$. For example, if we want to give a differential importance to each stimulus according to the value of its general activation, $\mathbf{M}$ will be defined as the diagonal matrix of the row margin of matrix $\mathbf{X}$ (*i.e.,* $m_{k,k} = x_{+k}$ with x_{+k} representing the total of the k-th column of $\mathbf{X}$, or, equivalently $x_{+k} = \sum_i x_{i,k}$). Note that choices other than a diagonal matrix are possible for $\mathbf{B}$ and $\mathbf{M}$.

Second, in order to analyze the properties of the generalized autoassociator, we need to generalize some basic notions of euclidean geometry. The generalized norm of vector $\mathbf{x}_k$, denoted $\mathbf{B}$-norm, is given by:

$$\|\mathbf{x}_k\|_{\mathbf{B}} = \sqrt{\mathbf{x}_k^T \mathbf{B} \mathbf{x}_k} \ . \tag{15}$$

The biased orthogonality of the pair of vectors $\mathbf{x}_k$ and $\mathbf{x}_{k'}$, denoted $\mathbf{B}$-orthogonality, is given by:

$$\mathbf{x}_k \perp_{\mathbf{B}} \mathbf{x}_{k'} = \mathbf{x}_k^T \mathbf{B} \mathbf{x}_{k'} = 0 \ . \tag{16}$$

The generalized cosine, denoted $\mathbf{B}$-cosine, is given by:

$$\cos_{\mathbf{B}}(\mathbf{x}_k, \mathbf{x}_{k'}) = \frac{\mathbf{x}_k^T \mathbf{B} \mathbf{x}_{k'}}{\|\mathbf{x}_k\|_{\mathbf{B}} \|\mathbf{x}_{k'}\|_{\mathbf{B}}} \ . \tag{17}$$

Finally, for convenience, the stimuli, $\mathbf{x}_k$, are normalized in the metric defined by $\mathbf{B}$ (*i.e.,* $\mathbf{x}_k^T \mathbf{B} \mathbf{x}_k = 1$).

3.2 Model description

As in the classical model, the stimuli are stored in the memory by modifiying the intensity of the connections between units, with the exception that during learning a differential importance is given to each stimulus. Formally:

$$\mathbf{W} = \mathbf{X} \mathbf{M} \mathbf{X}^T \ . \tag{18}$$

The effect of the constraints on the units (*i.e.,* the bias matrix $\mathbf{B}$) can be interpreted as a filtering or re-coding scheme for the original stimuli prior to storage in the memory. This effect can be modeled during the recall phase as a pre-multiplication of the stimuli by the

where $\mathbf{g}_k$ (respectively $\mathbf{g}_{k'}$) is the vector of the projections of stimulus k (respectively k') onto the eigenvectors. This suggests the use of principal component analysis to display the stimuli as they are "perceived" by the auto-associative memory.

A final point worth noting is that the eigenvectors and eigenvalues of the weight matrix $\mathbf{W}$ can be obtained directly using the singular value decomposition of the original matrix of stimuli $\mathbf{X}$. Formally:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \qquad \text{with } \mathbf{V}^T\mathbf{V} = \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{8}$$

where $\mathbf{U}$ represents the matrix of eigenvectors of $\mathbf{X}\mathbf{X}^T$, $\mathbf{V}$ represents the matrix of eigenvectors of $\mathbf{X}^T\mathbf{X}$, and $\boldsymbol{\Delta}$ is the matrix of singular values which are equal to the square root of the eigenvalues of $\mathbf{X}\mathbf{X}^T$ or $\mathbf{X}^T\mathbf{X}$ (they are the same). The projections, $\mathbf{G}$, of the K stimuli of the training set on the R eigenvectors of the weight matrix can be found as:

$$\mathbf{G} = \mathbf{X}^T\mathbf{U} = \mathbf{V}\boldsymbol{\Delta} \ . \tag{9}$$

Within the framework of principal component analysis, $\mathbf{G}$ is the matrix of the projections of the stimuli on the principal components. From Eq. 9, it is easy to derive that the variance of the projections on a given eigenvector is equal to the eigenvalue associated with this eigenvector:

$$\mathbf{G}^T\mathbf{G} = \boldsymbol{\Delta}\mathbf{V}^T\mathbf{V}\boldsymbol{\Delta} = \boldsymbol{\Lambda} \ . \tag{10}$$

Likewise, the projections of a set of K' new stimuli (*i.e.,* not learned by the memory), $\mathbf{X}_{\text{new}}$, on the eigenvectors of $\mathbf{W}$ can be computed as:

$$\mathbf{G}_{\text{new}} = \mathbf{X}_{\text{new}}^T\mathbf{U} \ . \tag{11}$$

Within the framework of principal component analysis, $\mathbf{G}_{\text{new}}$ contains the projections of the supplementary elements (*i.e.,* stimuli) on the principal components.

In order to improve the storage capacity of an autoassociative memory, most applications use the Widrow-Hoff learning rule. The Widrow-Hoff learning rule corrects the difference between the response of the system and the expected response by changing iteratively the weights in matrix $\mathbf{W}$ as follows:

$$\mathbf{W}_{[t+1]} = \mathbf{W}_{[t]} + \eta(\mathbf{X} - \mathbf{W}_{[t]}\mathbf{X})\mathbf{X}^T \tag{12}$$

where η is a learning constant. The Widrow-Hoff learning rule can also be analyzed in terms of eigenvectors and eigenvalues (Abdi, 1994). Specifically, $\mathbf{W}$ at time t can be expressed as:

$$\mathbf{W}_{[t]} = \mathbf{U}\boldsymbol{\Phi}_{[t]}\mathbf{U}^T \qquad \text{with} \quad \boldsymbol{\Phi}_{[t]} = [\mathbf{I} - (\mathbf{I} - \eta\boldsymbol{\Lambda})^t] \ . \tag{13}$$

With a learning constant η smaller than $2\lambda_{\max}^{-1}$ ($\lambda_{\max}$ being the largest eigenvalue) this procedure converges toward

$$\mathbf{W}_{[\infty]} = \mathbf{U}\mathbf{U}^T \tag{14}$$

which indicates that using the Widrow-Hoff error correction learning rule amounts to equalizing all the eigenvalues of $\mathbf{W}$ (*i.e.,* to sphericizing the weight matrix).

where $\hat{\mathbf{x}}_k$ represents the response of the memory. The quality of the response of the system can be measured by computing the cosine of the angle between $\mathbf{x}_k$ and $\hat{\mathbf{x}}_k$:

$$\cos(\mathbf{x}_k, \hat{\mathbf{x}}_k) = \frac{\mathbf{x}_k^T \hat{\mathbf{x}}_k}{\|\mathbf{x}_k\| \|\hat{\mathbf{x}}_k\|} \tag{3}$$

where $\|\mathbf{x}_k\|$ is the euclidean norm of the vector $\mathbf{x}_k$ (*i.e.,* $\|\mathbf{x}_k\| = \sqrt{\mathbf{x}_k^T \mathbf{x}_k}$). A cosine of 1 indicates a perfect reconstruction of the stimulus.

When the stimulus set is composed of non-orthogonal stimuli, the associator does not perfectly reconstruct the stimuli that are stored. On the other hand, some new patterns are perfectly reconstructed, creating, in a way, the equivalent of a "false alarm" or "false recognition." These patterns are defined by the equation:

$$\mathbf{W}\mathbf{u}_r = \lambda_r \mathbf{u}_r \qquad \text{with} \quad \mathbf{u}_r^T \mathbf{u}_r = 1 \tag{4}$$

where $\mathbf{u}_r$ denotes the r-th eigenvector of $\mathbf{W}$, and λ_r the eigenvalue associated with the r-th eigenvector.

From Eq. 1, it can be seen that the matrix $\mathbf{W}$ is equivalent to a cross-product matrix, and hence is semi-positive definite (*i.e.,* all its eigenvalues are positive or zero). Consequently, $\mathbf{W}$ can be reconstructed as a weighted sum of its eigenvectors:

$$\mathbf{W} = \sum_{r=1}^{R} \lambda_r \mathbf{u}_r \mathbf{u}_r^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \qquad \text{with} \quad \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{5}$$

where $\mathbf{I}$ stands for the identity matrix, $\mathbf{\Lambda}$ represents the diagonal matrix of eigenvalues and R is the rank of the matrix $\mathbf{W}$. The eigenvectors in $\mathbf{U}$ are usually ordered according to their eigenvalues. This formulation makes clear the close relationship between the classical linear autoassociator and some techniques used in multivariate statistical analysis. Specifically, using an autoassociative memory to store and recall a set of objects is equivalent to performing a principal component analysis on the cross-product matrix of the feature set describing these objects (cf. Anderson *et al.*, 1977).

Associated with the technique of principal component analysis, is the notion of a distance. One way of looking at the eigendecomposition of the matrix $\mathbf{W}$ is to note that the euclidean distance between stimuli as well as the euclidean distance between any stimulus and the average stimulus (*i.e.,* the barycenter, or centroid, of the set of stimuli) is now decomposed orthogonally along the eigenvectors. Specifically, the euclidean distance between stimuli k and k' is computed as:

$$d^2(\mathbf{x}_k, \mathbf{x}_{k'}) = (\mathbf{x}_k - \mathbf{x}_{k'})^T (\mathbf{x}_k - \mathbf{x}_{k'}) \ . \tag{6}$$

The distance can be expressed also, through the eigendecomposition as

$$d^2(\mathbf{x}_k, \mathbf{x}_{k'}) = d^2(\mathbf{g}_k, \mathbf{g}_{k'}) = (\mathbf{g}_k - \mathbf{g}_{k'})^T (\mathbf{g}_k - \mathbf{g}_{k'}) \ . \tag{7}$$

that storing stimuli in an autoassociative memory amounts to creating the cross-product matrix of the stimuli and computing its eigendecomposition (Abdi, 1988, 1993, in press; Anderson *et al.*, 1977; Kohonen, 1977). This is equivalent to computing the principal component analysis of the set of features used to describe the stimuli. One advantage of this type of analysis is that it makes it clear that classical autoassociators implement least-squares approximation (or Wiener filtering, cf. Abdi, 1994). In terms of optimization problems, the interest of the generalized autoassociator described in this paper is that it implements a generalized least-squares approximation or a least-squares approximation under (linear) constraints. This technique is used in various settings. In multivariate statistical analysis, for example, canonical analysis (and hence the complete set of generalized linear models) can be easily derived within this framework (*e.g.,* Mardia, Kent, & Bibby, 1979; Greenacre, 1984). As a consequence, neural networks can be easily shown to be equivalent to traditional statistical and optimization techniques.

This paper is organized as follows. First, the basic features of the classical autoassociative model are briefly presented along with their relationship to the linear model of multivariate analysis. Then, a generalization of this model is described and analyzed in terms of statistical and optimization problems. Specifically, we demonstrate that a generalized linear autoassociator implements the general linear model of multivariate statistics. Finally, we show that a linear generalized autoassociator implementing correspondence analysis (*i.e.,* a specific case of the general linear model) can be used successfully to categorize a set of faces according to their sex. We show, in this specific application, that the generalized version of the autoassociator can learn the task as accurately as the standard version, but does so more quickly.

2. Classical model

Objects to be stored in an autoassociative memory are represented by $I \times 1$ column vectors $\mathbf{x}_k$ whose I components code the values of the I features used to describe the objects. In a neural network implementation, these components represent the activation of the input units (*i.e.,* cells). For convenience, the vectors $\mathbf{x}_k$ are assumed to be normalized so that $\mathbf{x}_k^T \mathbf{x}_k = 1$ (with $\mathbf{x}_k^T$ denoting the transpose of $\mathbf{x}_k$). The set of K stimuli to be stored in the memory is represented by an $I \times K$ matrix $\mathbf{X}$ in which the k-th column is equal to $\mathbf{x}_k$. The autoassociative memory (or weight matrix) is represented by an $I \times I$ matrix $\mathbf{W}$. The values in the weight matrix correspond to the connection strengths between the units of the memory.

The stimuli are stored in the memory by changing the strength of the connections between units. This can be done using a simple Hebbian learning rule:

$$\mathbf{W} = \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^T = \mathbf{X}\mathbf{X}^T \ . \tag{1}$$

Recall of a given stimulus $\mathbf{x}_k$ is given by:

$$\hat{\mathbf{x}}_k = \mathbf{W}\mathbf{x}_k \tag{2}$$

content addressable memory in the sense that it learns to associate inputs to themselves. As such, the model can operate also as a powerful pattern completion device, capable of reconstructing learned input stimuli with memory keys that have been degraded either by adding noise or by ablating parts of the code. Kohonen (1977), for example, showed that an autoassociative memory can be used to store images of human faces and reconstruct the original faces when features have been omitted or degraded.

When a linear autoassociative memory is viewed as a "neural network", the values in the weight matrix correspond to the connection strengths between the cells or units of the memory. Learning, in this framework, amounts to finding a set of connections between input units that minimizes the error in reconstructing the input stimuli. In the "standard" or "classical" autoassociator, all the units composing the memory are equivalent and independent, and all the stimuli to be stored in the memory are of equal importance. While this kind of model is capable of solving many pattern recognition problems, other problems require additional constraints. Specifically, many real pattern recognition problems operate under well-established *a priori* constraints that can function either at the level of differentiating parts of the code and/or at the level of differentiating individual stimuli as a function of their "importance" in building the model representation.

One example of the way in which different parts of a code can be differentially important for solving a problem can be seen in the representational constraints that are implemented in many biological vision systems to enhance luminance contrast. Such constraints are required in order to make optimal use of the strongly limited bandwidth of the optic nerve, which transmits information from the retina to the cortex. The neural scheme operating in these visual systems capitalizes on the fact that individual parts of the retinal code are not equally informative. For example, areas of the retina that contain information about luminance contrast are more important than areas of the retina that are uniformly illuminated. The generalization of the autoassociator that we propose allows for a mechanism by which individual *units* of the memory can be assigned differential importance.

In addition to implementing representational constraints, it can be useful to implement a mechanism that allows for a differential weighting of individual stimuli. For example, to model human memory, it is often necessary to take into account factors that differentiate the importance of individual stimuli in building a representation of the problem. In the learning of lists, temporal interference between successive items is one such factor. This phenomenon can be easily simulated by implementing a differential weighting of the stimuli as a function of their position on the list. For example, in the case of retroactive interference, a more recent stimulus interferes with the memory of previously learned stimuli. More precisely, the importance of any given stimulus is inversely proportional to its position in the learning sequence. In addition to the differential weighting of parts of the code, the generalized autoassociator allows for the implementation of *a priori* biases in the stimulus set. The differential importance (and non-independence) of both the units and the stimuli are defined as a set of constraints expressed *via* positive definite matrices operating on the autoassociator.

The classical linear autoassociator has been often analyzed in terms of the eigendecomposition or singular value decomposition of a matrix. Specifically, it has been shown

Ref: Abdi, H., Valentin. D., O'Toole, A.J., (In press, 1996) A generalized autoassociator model for face processing and sex categorization: From principal component analysis to multivariate analysis. In D. Levine (Ed.) *Optimality in biological and artificial networks.* Hillsdale: Erlbaum.

A generalized autoassociator model for face processing and sex categorization: From principal component analysis to multivariate analysis

Hervé Abdi*, Dominique Valentin*† and Alice J. O'Toole*

* *School of Human Development: The University of Texas at Dallas, MS: GR.4.1., Richardson, TX75083–0688, U.S.A.*
† *Université de Bourgogne à Dijon, Boulevard Gabriel, 21000 Dijon, France*

ABSTRACT

In this paper we propose a generalized version of the classical linear autoassociator that can be shown to implement a generalized least-squares approximation under linear constraints. The standard linear autoassociator is known to implement principal component analysis, whereas the generalized model implements the general linear model (*e.g.,* canonical correlation). In practical terms, this generalization allows for the imposition of *a priori* constraints that enable differential weighting of both individual units of the input code and individual stimuli. As an illustration of the utility of the generalized model, we present simulations comparing the accuracy and learning speed of the standard and generalized versions of the autoassociator for the problem of categorizing faces by sex. We show that while the two models are equally accurate, the generalized model learns the task considerably faster than does the standard model.

1. Introduction

Recent years have witnessed a strong resurgence of interest in the field of neural networks. Some of the earliest models characterizing this resurgence were simple linear associative memory models (Anderson, Silverstein, Ritz, & Jones, 1977; Kohonen, 1977). Associative memories are capable of learning associations between input-output pairs such that the memory produces the appropriate output in response to a learned or "associated" input. The inner workings of these associative models, and other related "neural" network models are reminiscent of the computational character of the brain. Specifically, the computations required to implement the storage and retrieval of information in the network can be carried out in parallel and the representation of individual learned associations is not localized in the memory, but rather, is "distributed" throughout the entire network.

The purpose of the present paper is to propose a generalization of a particular case of a linear associator model known as an "*auto*associator". The autoassociator can act as a