Biol. Cybern. 56, 367–374 (1987)

Biological Cybernetics

Some Algebraic Relations Between Involutions, Convolutions, and Correlations, with Applications to Holographic Memories

P. H. Schönemann

Purdue University, Department of Psychological Sciences, West Lafayette, IN 47906, USA

Abstract. Convolutions $*$ and correlations $\#$ in spaces H of doubly infinite sequences are related by $a \# b = S(a * Sb)$, where S is an involution which reflects the order in the integral domain Z on which the sequences are defined. This relation can be used to represent a non-associative correlation algebra $\langle H, \# \rangle$ by an associative convolution algebra equipped with the involution S which, as is shown, greatly simplifies derivations. Related matrix representations of $\#, *, S$ are given for sequences with finite support in Re^n. Some implications for holographic memory models are discussed.

1 Introduction and Definitions

After a long period of neglect, the so-called "holographic memory models" are now actively discussed in the psychological literature (e.g., Murdock 1982; Metcalfe-Eich 1982). These models rest on a convolution/correlation paradigm which, though well-known to physicists and communication engineers, is relatively new to social scientists. In this paper, some of the basic definitions of the convolution and correlation operation are reviewed. It is shown that some of the algebraic complications induced by the non-commutative and non-associative correlation operation can be eliminated by introducing a simple involution. It is also shown that in the finite case convolutions and correlations can be computed as standard matrix products. Finally, it is noted that the limitation to finite feature vectors weakens the intuitive appeal of such memory models because they imply that the memory task becomes more difficult as the stimuli become simpler and more structured and that perfect recall is only possible for cues with exactly one non-zero feature. These difficulties can be circumvented by replacing correlation with another retrieval mechanism. We begin with a review of the basic definitions needed for a cogent discussion of holographic memories:

An *(anchored) doubly infinite real sequence*, $\{y_k\}$, is a map from the integers Z to the real numbers Re with typical image

$$\{y_k\} := \{\ldots, y_{-m}, \ldots, y_0, \ldots, y_m, \ldots\}. \tag{1.1}$$

The set of all such sequences will be denoted "H". Any finite sequence, e.g., a finite-dimensional vector $\mathbf{y} \in \mathrm{Re}^{p+1}$,

$$\mathbf{y}' := (y_0, y_1, \ldots, y_p) \tag{1.2}$$

can always be interpreted as an element of H for which y_k is identically 0 for $k < 0$ and $k > p$, since then the map

$$m: \mathrm{Re}^{p+1} \to H: \quad m(y_0, \ldots, y_p) = (\ldots, 0, 0, y_0, \ldots, y_p, 0, 0, \ldots) \tag{1.3}$$

is 1:1. In this case we will say "$\{y_k\}$ in H has finite support of order $p+1$" or, "the doubly infinite sequence $\{y_k\}$ has finite support $\mathbf{y}$ in Re^{p+1}". In applications to memory models one often works with finite sequences of the type

$$\mathbf{z}' := (z_{-m}, \ldots, z_0, \ldots, z_m) \in \mathrm{Re}^{2m+1} \tag{1.4}$$

i.e., with elements of H with finite support of odd order $2m+1$. Note that (1.3) is simply a special case of (1.4).

In practice, the injection m of Re^n into H amounts to appending as many zeros as necessary to the finite sequence to perform the computations in H. After dropping the appropriate number of leading and trailing zeros, the result can be retrieved as an element of Re^q (with $q \neq n$, in general). Since most sequences arising in experimental psychology are likely to be finite, this case will be emphasized. The injection m will be used when convenient to go from Re^n to H and back. The restriction to sequences with finite support avoids problems of convergence. Finite vectors in Re^n will be underlined and treated as column vectors if they are not followed by a transposition sign (′).

Under the usual definitions of vector addition and scalar multiplication over Re,

$$+: H^2 \to H: \{y_k\} + \{z_k\} = \{y_k + z_k\}, \tag{1.5a}$$

$$\cdot: \mathrm{Re} x H \to H: s\{y_k\} = \{sy_k\} \tag{1.5b}$$

the algebraic system $\langle H, \mathrm{Re}, +, \cdot \rangle$ becomes a vector space. If it can be equipped with a bilinear "multiplication" $\circ$ for the vector elements, it becomes an algebra. A familiar example is the algebra $\langle \mathrm{Re}^{pxp}, \mathrm{Re}, +, \cdot, \circ \rangle$, where $\circ$ denotes matrix multiplication. However, in general, $\circ$ need not be associative.

The maps

$$B: H \to H: B\{y_k\} = \{y_{k-1}\}$$
and $$\tag{1.6}$$
$$F: H \to H: F\{y_k\} = \{y_{k+1}\}$$

are called "backward shifts" and "forward shifts" in the time series literature. Since both maps preserve linear combinations in H, B, and F are linear maps. They are also bijective since F is the inverse of B. On applying either map, e.g., B, m times in succession, one obtains $B^m\{y_k\} = \{y_{k-m}\}$. B^m and F^m can be interpreted as translations. B^m translates the origin backward m steps and F^m translates it forward m steps (see Fig. 1).

The map

$$S: H \to H: S\{y_k\} = \{y_{-k}\} \tag{1.7}$$

is also a linear bijection. Since $S^2 = I$ (the identity map), but $S \neq I$, it is an involution in H. The effect of S is to reverse the order of the elements of $\{y_k\}$ (see again Fig. 1).

Convolution $*$ is a binary operation in H defined by

$$*: H^2 \to H: \{x_i\} * \{y_i\} = \{\Sigma_i x_i y_{k-i}\} = \{z_k\}, \tag{1.8}$$

where the subscripts i, k range from $-\infty$ to ∞. This operation only exists for all pairs if H is suitably restricted to ensure convergence of the infinite sum. For sequences with finite support $*$ always exists. The definition implies that $*$ is commutative and associative, and that it has a two-sided identity, given by

$$\{u_k\} \quad \text{with} \quad u_0 = 1, u_k = 0 \quad \text{for} \quad k \neq 0. \tag{1.9}$$

Correlation $\#$ is a binary operation in H defined by

$$\#: H^2 \to H: \{x_i\} \# \{y_i\} = \{\Sigma_k x_i y_{i+k}\} = \{z_k\}, \tag{1.10}$$

where the subscripts i, k range from $-\infty$ to ∞. This definition implies that $\#$ is neither commutative nor associative, and that it has only a left-sided, but no right-sided identity. The left-sided identity, u_L, is again given by (1.9). Note that the definitions of $*$, $\#$, and u refer to H, not Re^n.

k:	...	-3	-2	-1	0	1	2	3	4	5	6 ...
$\underline{x}'$					x_0	x_1	x_2	x_3	x_4		
$\{x_k\}$	...	0	0	0	x_0	x_1	x_2	x_3	x_4	0	0 ...
$\{y_k\} = B\{x_k\}$	...	0	0	0	0	x_0	x_1	x_2	x_3	x_4	0 ...
$\{z_k\} = F\{x_k\}$	...	0	0	x_0	x_1	x_2	x_3	x_4	0	0	0 ...
$\{w_k\} = F^2\{x_k\}$	...	0	x_0	x_1	x_2	x_3	x_4	0	0	0	0 ...
$\{w_k\}$	...	0	w_{-2}	w_{-1}	w_0	w_1	w_2	0	0	0	0 ...
$S\{w_k\}$	...	0	w_2	w_1	w_0	w_{-1}	w_{-2}	0	0	0	0 ...

Fig. 1. Embedding of finite sequences in *H*, shifts and involution. (The boxed portion shows the transpose of a coeffient matrix [**a**] used in the matrix representation of $a * b$)

2 Relations Between $*$, $\#$, and S

The definitions of $*$ and $\#$ in (1.8) and (1.10) imply that these operations relate to each other via the involution S in (1.7):

$$a \# b = S(a * Sb). \tag{2.1}$$

Proof. $a * Sb := \{\Sigma_i a_i b_{i-k}\} = S(a \# b)$, $S^2 = I$.

The same reasoning shows:

$$Sa * Sb = S(a * b), \tag{2.2}$$

i.e. the involution S can be factored out of a convolution.

Proof. $Sa * Sb = \{\Sigma_i a_{-i} b_{i-k}\} = \{\Sigma_j a_j b_{-k-j}\} =: S(a * b)$.

Notational note: $Sa := S(a)$. Parenthesis must be retained for $S(a * b)$ and $S(a \# b)$.

Similarly, S can be factored out of a correlation:

$$Sa \# Sb = S(a \# b). \tag{2.3}$$

Proof. $Sa \# Sb = S(Sa * b) = a * Sb = S(a \# b)$, by (2.1), (2.2).

The expression for $\#$ in terms of $*$ and S can now be further simplified:

$$a \# b = b * Sa. \tag{2.4}$$

which is also computationally more efficient than (2.1). Professor Drazin pointed out that (2.4) implies $Sa = a \# u$.

Proof. $a \# b = S(a * Sb) = b * Sa$, by (2.1), (2.2).

Together with the basic properties of $*$ and S:

$$S^2 = I, a * b = b * a, a * (b * c) = (a * b) * c = a * b * c, \tag{2.5}$$

Eq. (2.1), (2.2) can be used to derive numerous other relations between $\#$, $*$, and S.

Table 1. Cayley tables for $*$ and $\#$ in $\langle H, \#, *, S\rangle$ (row element first factor)

$*$	a	b	Sa	Sb	$\#$	a	b	Sa	Sb
a	$a*a$	$a*b$	$a\#a$	$b\#a$	a	$a\#a$	$a\#b$	$S(a*a)$	$S(b*a)$
b	$a*b$	$b*b$	$a\#b$	$b\#b$	b	$b\#a$	$b\#b$	$S(b*a)$	$S(b*b)$
Sa	$a\#a$	$a\#b$	$S(a*a)$	$S(a*b)$	Sa	$a*a$	$a*b$	$a\#a$	$b\#a$
Sb	$b\#a$	$b\#b$	$S(b*a)$	$S(b*b)$	Sb	$b*a$	$b*b$	$a\#b$	$b\#b$

$(a\#c)\#b=(b\#c)\#a=(c\#a)*b=(c\#b)*a=c\#(a*b)$

In particular, these identities can be used to reduce any expression $f(a,\ldots,z;*,\#,S)$ containing $\#$, $*$, and S to a "convolution normal form"

$$f(a,\ldots,z;*,\#,S)=S^{k(a)}a*\ldots*S^{k(z)}z,\ k(x)\in\{0,1\} \quad (2.6)$$

which is free of the non-associative $\#$. This canonical form provides a quick and mechanical check of the identity of two expressions in $\#$, $*$, and S. For example,

$$(a\#c)\#b=S[(a\#c)*Sb]=S[S(a*Sc)*Sb]= \\ =a*b*Sc. \quad (2.7a)$$

Similarly,

$$c\#(a*b)=Sc*(a*b)=a*b*Sc, \quad (2.7b)$$

$$(c\#a)*b=(Sc*a)*b=a*b*Sc, \quad (2.7c)$$

$$(c\#b)*a=(Sc*b)*a=a*b*Sc, \quad (2.7d)$$

$$(b\#c)\#a=(Sb*c)\#a=S(Sb*c)*a=a*b*Sc. \quad (2.7e)$$

Hence,

$$(a\#c)\#b=(b\#c)\#a=(c\#a)*b=(c\#b)*a \\ =c\#(a*b). \quad (2.8)$$

This set of identities is basic for holographic memory theory.

Cayley tables for other products involving $\#$, $*$, and S are given in Table 1.

3 Matrix Representation of $\#$, $*$, and S

For sequences with finite support in Re^n the involution S can be represented by an $n\times n$ matrix S of the form

$$S=(s_{ij}),\ s_{ij}=1 \quad \text{for} \quad j=(n+1)-i \\ =0 \quad \text{elsewhere}, \quad (3.1)$$

i.e. S is a square matrix with ones in the counter diagonal and zeros elsewhere.

The finite support of a convolution of two sequences with finite support, e.g., $\mathbf{a}*\mathbf{b}$ with $\mathbf{a}\in\mathrm{Re}^{p+1}$, $\mathbf{b}\in\mathrm{Re}^{q+1}$, can be computed as a matrix product of the form $[\mathbf{a}]\mathbf{b}$ or, alternatively, of the form $[\mathbf{b}]\mathbf{a}$. This follows on expressing $\mathbf{a}*\mathbf{b}$ as a polynomial in the backward shift operator $B(\ )$ with weights b_i:

$$a*b=b*a=\{\Sigma_i a_{j-i}b_i\}=(\Sigma_i B^i(\mathbf{a})b_i). \quad (3.2)$$

To obtain a rectangular coefficient matrix $[\mathbf{a}]$, leading and trailing zeros are needed. Their use is justified by embedding the $B^i(\mathbf{x})$ in H. We can then construct a coefficient matrix $[\mathbf{a}]$ in terms of the backward shift operator $B(\ )$ as follows:

$$[\mathbf{a}]:=(a, B(a), \ldots, B^q(a)), \quad (3.3)$$

where the first column a of the coefficient matrix $[a]$ is the column vector $\mathbf{a}$ augmented by q trailing zeros, $B(a)$ a column vector of one leading zero followed by $\mathbf{a}$ and $q-1$ trailing zeros, etc., and $B^q(a)$ is a column vector of q leading zeros followed by $\mathbf{a}$. The transpose of such a coefficient matrix is outlined as a box in Fig. 1. The resulting matrix $[\mathbf{a}]$ is of order $(p+q+1)x(q+1)$. It can be used to compute the convolution $\mathbf{a}*\mathbf{b}$ as a standard matrix product

$$\mathbf{a}*\mathbf{b}=[\mathbf{a}]\mathbf{b}. \quad (3.4)$$

To illustrate this, let $\mathbf{a}':=(1\ \ 3\ \ 5)$ and $\mathbf{b}':=(1\ \ 2)$. Then

$$\begin{pmatrix} \mathbf{1} & 0 \\ 3 & \mathbf{1} \\ 5 & 3 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \mathbf{1} \\ 2 \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ 5 \\ 11 \\ 10 \end{pmatrix} \quad (3.5)$$

$(aB(a))\,\mathbf{b}=\mathbf{a}*\mathbf{b}=\mathbf{c}$.

The boldface components are the anchors a_0, b_0, c_0 needed to reembed $\mathbf{c}$ in H, if desired. The definition (1.8) implies that z_0 is the term containing the product x_0y_0. There are, of course, other devices for representing convolutions in terms of matrices (see, e.g., Borsellino and Poggio 1973; Metcalfe Eich 1982, p. 660). The representation suggested here illuminates the connection between convolution and backward shifts and has a number of immediate implications:

(i) If $\mathbf{a}\neq\emptyset$ and $\mathbf{a}*\mathbf{x}=\mathbf{c}$ is consistent (i.e., has at least one solution), then the solution

$$\mathbf{x}=[\mathbf{a}]^+\mathbf{c} \quad (3.6)$$

is unique (where $[\mathbf{a}]^+:=([\mathbf{a}]'[\mathbf{a}])^{-1}[\mathbf{a}]'$ is the Moore-Penrose inverse of $[\mathbf{a}]$ in (3.5)).

Proof. If $\mathbf{a}\neq\emptyset$, $[\mathbf{a}]$ has full column rank by construction.

This means one can always deconvolve a "memory trace" $\mathbf{a}*\mathbf{b}$ for $\mathbf{b}$ exactly, if $\mathbf{a}$, $\mathbf{c}$ are given, and suggests an alternative to correlations for deconvolving memory traces $\mathbf{a}*\mathbf{b}$ (see Sect. 5).

Hence, by commutativity of $*$:

(ii) For sequences of finite support, $*$ is left and right cancellable.

Thus there are no null-divisors in H restricted to sequences with finite support.

We now ask under which conditions $\mathbf{a}*\mathbf{b}=\mathbf{c}$ is consistent. In view of the matrix representation (3.5), well-known results from linear algebra (see, e.g., Ben-Israel and Greville 1974) imply:

(iii) For sequences with finite support, $\mathbf{a}*\mathbf{x}=\mathbf{c}$ is consistent iff $[\mathbf{a}]\,[\mathbf{a}]^{+}\mathbf{c}=\mathbf{c}$.

Although the equation $\mathbf{a}*\mathbf{x}=\mathbf{c}$ need not be consistent in general (consider $\mathbf{a}=\emptyset\neq\mathbf{c}$), we always can solve it in a least squares sense with the expression given in (3.6):

(iv) If $\mathbf{a}*\mathbf{b}=\mathbf{c}$ is not consistent then

$$\hat{\mathbf{a}}=[\mathbf{b}]^{+}\mathbf{c} \quad\text{and}\quad \hat{\mathbf{b}}=[\mathbf{a}]^{+}\mathbf{c} \tag{3.7}$$

are least squares estimates of $\mathbf{a}$, $\mathbf{b}$, respectively.

It should be noted that the fact that $[\mathbf{a}]^{+}$ always exists does not mean that a^{+} always exists in H restricted to finite sequences. On the contrary:

(v) If H is restricted to finite sequences with maximal support n, then the semigroup $\langle H, *\rangle$ is not regular.

Proof. A semigroup is called "regular" if every element a has a generalized inverse a^{-} which satisfies $a*a^{-}*a=a$. In view of the associativity and commutativity of $*$ in H, this is equivalent to the existence of x in $(a*a)*x=a$, i.e., to the consistency of the matrix equation $[\mathbf{a}*\mathbf{a}]\mathbf{x}=\mathbf{a}$. On choosing, e.g., $\mathbf{a}'=(1\ \ 1)$, one finds that this equation need not be consistent.

Hence as long as H is restricted to sequences of finite support n, $\mathbf{a}$ need not have a generalized convolution inverse, let alone a regular convolution inverse. This is relevant for holographic memory models defined on finite feature spaces because in general it rules out exact deconvolution of the basic identity (2.8) via correlations (see Sect. 5).

We now turn to correlations. Equations (2.4), (3.4) imply that the finite support of a correlation of two finite sequences, e.g., $\mathbf{a}\#\mathbf{b}$ with $\mathbf{a}\in\mathrm{Re}^{p+1}$, $\mathbf{b}\in\mathrm{Re}^{q+1}$, can be computed as

$$\mathbf{a}\#\mathbf{b}=[\mathbf{b}]S\mathbf{a}\,. \tag{3.8}$$

The coefficient matrix is the same as (3.3) for computing the convolution of a with b in the order $b*a$. The vector $\mathbf{a}$ is postmultiplied by an involution matrix of order $p+1$. To illustrate this numerically with $\mathbf{a}':=(1\ \ 3\ \ 5)$ and $\mathbf{b}':=(1\ \ 2)$:

$$\begin{pmatrix}1&0&0\\2&1&0\\0&2&1\\0&0&2\end{pmatrix}\begin{pmatrix}&&1\\&1&\\1&&\end{pmatrix}\begin{pmatrix}1\\3\\5\end{pmatrix}=\begin{pmatrix}1&0&0\\2&1&0\\0&2&1\\0&0&2\end{pmatrix}\begin{pmatrix}5\\3\\1\end{pmatrix}=\begin{pmatrix}5\\13\\7\\2\end{pmatrix}$$

$$(b, B(b), B^2(b))S\mathbf{a}=[\mathbf{b}]S\mathbf{a}=a\#b\,. \tag{3.9}$$

Analogous to (3.6) one finds:

If $\mathbf{a}\#\mathbf{b}=\mathbf{c}$ is consistent, then the solutions

$$\mathbf{a}=S_1[\mathbf{b}]^{+}S_2\mathbf{c} \quad\text{and}\quad \mathbf{b}=S_1[\mathbf{a}]^{+}S_2\mathbf{c} \tag{3.10}$$

are unique.

Proof. For the left factor, the argument is identical to that given for convolutions, with (3.7) replacing (3.4). For the right factor, one has $\mathbf{a}\#\mathbf{b}=\mathbf{c}\Leftrightarrow\mathbf{b}\#\mathbf{a}=S_1\mathbf{c}\Leftrightarrow[\mathbf{a}](S_2\mathbf{b})=S_1\mathbf{c}$. Note that the orders of the two involution matrices S_1, S_2 are different.

Since the results for consistency and least squares solutions of equations involving $\#$ are direct analogues to those for $*$ in view of (2.4), they are omitted.

4 Geometric Interpretation of S as a Reflection in H

As long H is restricted to sequences of maximal finite support n, it can always be equipped with the natural scalar product $\mathbf{a}'\mathbf{b}$. Relative to this scalar product, the involution S is an isometry, since $S'S=I$. Since the determinant of S is -1, it is a reflection in an $(n-1)$-dimensional hyperplane in Re^n.

A sequence v in H is called "even" if it satisfies $Sv=v$. A sequence w is called "odd" if it satisfies $Sw=-w$. The orthogonal projector associated with S,

$$P:=(I+S)/2 \tag{4.1}$$

maps all vectors in Re^n into even vectors,

$$\mathbf{v}:=P\mathbf{y}=((y_k+y_{-k})/2)\,, \tag{4.2}$$

i.e., into the subspace $E:=\{\mathbf{v}|S\mathbf{v}=\mathbf{v}\}$, which is invariant under P. Since S and hence P are symmetric, P is the orthogonal projection into E. The complementary projector

$$Q:=I-P=(I-S)/2 \tag{4.3}$$

maps any $\mathbf{y}$ into an odd vector,

$$\mathbf{w}:=Q\mathbf{y}=((y_k-y_{-k})/2)\,, \tag{4.4}$$

so that Q is the orthogonal projector for the complementary subspace E^c of all odd vectors, $\{\mathbf{w}|S\mathbf{w}=-\mathbf{w}\}$ in Re^n. Hence all sequences in H with finite support can be expressed uniquely as sums of pairwise orthogonal even and odd vectors. Orthogonality of $\mathbf{v}$, $\mathbf{w}$ also follows from $\mathbf{v}'\mathbf{w}=(\mathbf{v}'S')\mathbf{w}=-\mathbf{v}'\mathbf{w}\Rightarrow 2\mathbf{v}'\mathbf{w}=0$.

As Borsellino and Poggio (1973) note, the definition of E implies

$$\text{In}\, E := \{\mathbf{v} | S\mathbf{v} = \mathbf{v}\},\ \mathbf{a} \# \mathbf{b} = \mathbf{a} * \mathbf{b}. \tag{4.5}$$

Proof. $a \# b = b * Sa = b * a$.

Hence, if one restricts memory models to E, then correlations become superfluous, because they can always be replaced by convolutions, without having to worry about the involution S.

However, this convenience has its price in terms of redundancy, because one finds from the traces of the projectors for the dimensions of the two complementary subspaces E, of even vectors, and E^c, of odd vectors, depending on whether n is odd or even,

$$\begin{aligned} &\text{if} \quad n = 2m: && \dim(E) = \dim(E^c) = m \\ &\text{if} \quad n = 2m+1: && \dim(E) = m+1,\ \dim(E^c) = m. \end{aligned} \tag{4.6}$$

The reason for the difference in dimensions when n = odd is that $w_0 = 0$ for all vectors $\mathbf{w}$ in E^c, by (4.4).

Equation (4.5) also implies that $\langle E, *, S\rangle$ and $\langle E^c, \#, S\rangle$ are subalgebras of $\langle \text{Re}^n, *, S\rangle$ and $\langle \text{Re}^n, \#, S\rangle$, since E is closed both under $*$ and $\#$. On the other hand, E^c is not closed under either operation.

5 Applications to Holographic Memory Models

Holographic memory models have been studied for some time, mostly in Europe. Almost 30 years ago Reichardt (1957) suggested already autocorrelations as a "functional principle of the CNS" (title of his paper). Gabor himself, the inventor of the holography principle, has commented repeatedly on the analogy of this paradigm with distributed memory storage and recall (e.g., Gabor 1948, 1968, 1969). In the 60's, a number of authors have proposed and evaluated memory paradigms based on the convolution/correlation principle, among them Julesz and Pennington (1965), Longuet-Higgins (1968), Willshaw and Longuet-Higgins (1969), van Heerden (1963) and others. An excellent review of this early literature is given by Willshaw (1981).

Until very recently, this early work had virtually no impact on psychology. For example Murdock's (1974) authoritative review of the memory literature which covers nearly 900 titles does not mention any one of the above references. Yet some of the earlier work on holographic memory models contains valuable lessons which perhaps should not be entirely ignored. For example, Willshaw, one of the early contributors, after reviewing the relevant literature in considerable detail and comparing the holographic memory paradigm with the matrix paradigm, arrives at a more sobering outlook on the prospects for the correlation/convolution paradigm to serve as a building block of a viable theory of human memory than some of the more recent holography converts: "As far as biological applications are concerned, instead of treating biological patterns as strings of random digits it would be worth investigating their structure, that is, the logical relations between their component parts" (Willshaw 1981, p. 103).

Stripped to basics, holographic memory models require at least four assumptions:

(i) the items ("stimuli") can be represented by "feature vectors" in an n-dimensional real space, i.e., by elements from Re^n.

(ii) these feature vectors are "noiselike" in a sense yet to be rendered precise,

(iii) during the learning phase, when subjects are presented with pairs of items, a, b, with associated feature vectors $\mathbf{a}$, $\mathbf{b}$, a "memory trace" is formed which consists of the convolution $\mathbf{a} * \mathbf{b}$ of the two feature vectors, and

(iv) at the time of recall, when a third item c with feature vector $\mathbf{c} \in \text{Re}^n$ is presented as a "cue" (or "probe"), recall of a learned item consists of the correlation of the cue with the memory trace $a * b$. Cue c need not be different from a or b. Thus, schematically:

Learning. Presenting items with feature vectors $\mathbf{a}$, $\mathbf{b}$ $\rightarrow$memory trace $= \mathbf{a} * \mathbf{b}$

Recall. Presenting cue with feature vector $\mathbf{c} \rightarrow$ recall $= \mathbf{c} \# (\mathbf{a} * \mathbf{b})$.

Under these assumption, by (2.8), recall reduces to recall with cue $\mathbf{c} = (\mathbf{c} \# \mathbf{a}) * \mathbf{b} = (\mathbf{c} \# \mathbf{b}) * \mathbf{a}$. (5.1)

This paradigm is motivated by the expectation that the expression in (5.1) will be in some sense "similar" to the feature vector of one of the learned items, a, b, which is thus "recalled", though perhaps imperfectly. While this expectation is justified in optics where the stimuli are complex waves, it may be more problematic in psychology, because Eq. (5.1) only works when assumption *(ii)* is met. Otherwise the recalled feature vector $\hat{\mathbf{a}}$ need not bear much resemblance to the stimulus vector $\mathbf{a}$ even when the cue $\mathbf{c}$ is identical to $\mathbf{b}$.

To see this, let

$$\mathbf{b}' = \mathbf{c}' = (0.3\ \ 0.7\ \ 0.7\ \ 0.3), \tag{5.2}$$

so that

$$\mathbf{b} \# \mathbf{b}' = (0.09\ \ 0.42\ \ 0.91\ \ 1.16\ \ 0.91\ \ 0.42\ \ 0.09), \tag{5.3}$$

which is far from a delta function $\mathbf{u}$. Hence recall of a learned stimulus

$$\mathbf{a}' = (0.3\ \ -0.7\ \ 0.7\ \ -0.3) \tag{5.4}$$

is

$$\hat{\mathbf{a}}' = (0.027\ \ 0.063\ \ 0.042\ \ -0.022\ \ -0.028\ \ 0.028\ \ 0.022\ \ -0.042\ \ -0.063\ \ -0.027) \tag{5.5}$$

which has only cosine 0.194 with **a**, even though cue **c** coincides with **b** in this case.

Assumption *(ii)* is equivalent to requiring that one of the correlations, $c \# a$ or $c \# b$, is "close" in some sense to the convolution identity u in (1.9) to retrieve one of the learned items a or b. In the simplest case, when the cue is identical to one of the learned items ($c=a$ or $c=b$), this means all autocorrelations r_k (the components of $a \# a, b \# b$) must be close to zero except r_0.

Borsellino and Poggio (1973) call vectors with the property

$$x \# x = u \Leftrightarrow x * Sx = u \tag{5.6}$$

"noiselike". They argue this assumption is usually met for complex stimuli: "All complicated patterns (practical examples are printed letters or ground glass surfaces and impulse sequences which are coded with pseudorandom shift register codes) are, in first approximation, noiselike functions" (p. 117). An analogous assumption is also needed for the matrix models, namely that the signal vectors are orthogonal, so as to suppress "cross-talk" (e.g., Kohonen et al. 1981, p. 117f.).

To avoid confusion with other, only approximately noiselike vectors, we shall call vectors with property (5.6) "perfectly noiselike". Similarly, we shall say "item a is perfectly recalled by cue c" if $c \# (a * b) = a$. If b is perfectly noiselike in the sense of (5.6), and used as a cue to recall a, (5.1) reduces to

$$\text{recall with cue } b = (\mathbf{b} \# \mathbf{b}) * \mathbf{a} = \mathbf{u} * \mathbf{a} = \mathbf{a}\,, \tag{5.7}$$

so that, in this case, item a is perfectly recalled.

We now inquire how many perfectly noiselike cues are contained in Re^n. To this end, one has to solve $\mathbf{x} \# \mathbf{x} = \mathbf{x} * S\mathbf{x} = \mathbf{u}$, i.e. the matrix equation $[\mathbf{x}]S\mathbf{x} = \mathbf{u}$ with $\mathbf{x}$ in Re^n and $\mathbf{u}$ of the form (1.9) in Re^{2n+1}. One finds that the seemingly innocuous transition from the infinite-dimensional wave spaces of physics to the finite-dimensional feature spaces of psychology has a drastic implication:

The basic convolution/correlation paradigm rules out perfect recall for the overwhelming majority of finite feature vectors by restricting it to cues with only one nonzero feature.

Proof. Let there be $n+1$ features, and consider the first n rows of $\mathbf{x} \# \mathbf{x}$, which must be zero:

$$\begin{pmatrix} x_0 & & & \\ x_1 & x_0 & & \\ & \cdots & & \\ x_{n-1} & & x_0 & 0 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n-1} \\ \cdots \\ x_0 \end{pmatrix} = \emptyset\,. \tag{5.8}$$

If $x_0 \neq 0$, the first n columns of the coefficient matrix are linearly independent so that, on dropping the last column of the coefficient matrix, the remaining solution vector must be zero, i.e., $x_0 \neq 0 \Rightarrow x_1 = x_2 = \ldots x_n = 0$. On the other hand, if $x_0 = 0$, and x_1 is non-zero, one obtains a reduced full column rank system by striking out the last two columns of the above coefficient matrix, etc.

While in practice one may not insist on perfect recall requiring perfectly noiselike cues, the fact that such models imply quite generally that recall deteriorates as the stimuli become simpler and more clearly structured – i.e., less noiselike – is counterintuitive and would have to be dealt with in some way. One possibility has been suggested by Borsellino and Poggio: "In order to implement an associative memory in a convolution-correlation structure, it seems necessary to induce a suitable isomorphism between the signal space and a noiselike set. That means some "noisecoding" of the input signals which have to be mapped into random or pseudorandom sequences".

Even after the stimuli have been "complexified" in some way, effective retrieval further requires that the dimension of the feature space be large. For most of the complicated patterns cited by Borsellino and Poggio, with the possible exception of printed letters, this requirement seems to be met. Similarly, Kohonen (1984), in his simulations with photographs of human faces, employs feature spaces with dimension n on the order of 3000. As he demonstrates, under these conditions the performance of his matrix model can be quite impressive.

However, it is less obvious why the stimuli of a paired associate learning task of nonsense syllables, for example, should require feature spaces of 3000 dimensions. All this underscores the need to follow up on Borsellino and Poggio's suggestion to explicate the transformation which maps the actual stimuli into the required noiselike feature space if one wants to convert the holography paradigm into a psychological "theory" or "model" of distributed memory.

6 A Revised Associative Recall Mechanism

A second problem which weakens the intuitive appeal of holographic memory models is the convolution/correlation paradigm itself. As Pike (1985), among others, has pointed out, from a psychological point of view the storage by $*$ and the retrieval through $\#$ appears to be no less ad hoc than the noiselike nature of the feature space. This criticism could be blunted by invoking Eq. (2.1) which permits replacement of two implausible operations, $*$, $\#$, by one implausible operation $*$ and a less implausible map, S.

On the other hand, since the assumption that retrieval is accomplished through correlation is entirely ad hoc in the first place, and entails the counterintuitive prediction that recall becomes more difficult as the stimuli become simpler and more clearly structured, a preferable alternative might be to discard the correlation altogether and to replace it with the Moore-Penrose solution (3.6) as a retrieval mechanism for deconvolving the memory trace $\mathbf{a} * \mathbf{b}$. For example, when c is "similar" to a which was paired with b during the learning phase, giving rise to the memory trace $a * b$, then

$$\hat{\mathbf{b}} = [\mathbf{c}]^+ (\mathbf{a} * \mathbf{b}) \tag{6.1}$$

will be the recall of b with cue c. In particular, for $c = a$:

$$\mathbf{b} = [\mathbf{a}]^+ (\mathbf{a} * \mathbf{b}). \tag{6.2}$$

In this case, b will be perfectly recalled because $a * b = c$ is consistent. For all other cues c recall gives a least squares approximation (3.7). The solution vector $\hat{\mathbf{b}}$ could be compared with the feature vectors $\mathbf{b}$ in numerous ways, e.g., in terms of the normalized scalar product.

Since this revised retrieval scheme no longer requires that the feature vectors be noiselike, it dispenses with the need to justify an intervening random map between the actual features of the stimuli and their representations in the storage paradigm. Instead of charging randomness with the unaccustomed task of inducing structure and enhancing discrimination, random vectors $\mathbf{e}$, $\mathbf{d}$ can now be introduced in their more traditional role as perturbations which dilute recall at the initial stages of learning. This can be done by entering $\mathbf{a}+\mathbf{e}$ and $\mathbf{b}+\mathbf{d}$ into the convolution. At the k'th trial the increment added into the stored memory trace will be

$$(\mathbf{a}+\mathbf{e}_k) * (\mathbf{b}+\mathbf{d}_k) = \mathbf{a} * \mathbf{b} + \mathbf{e}_k * \mathbf{b} + \mathbf{d}_k * \mathbf{a} + \mathbf{e}_k * \mathbf{d}_k, \tag{6.3}$$

in view of the bilinearity of $*$. This yields a simple associative learning/memory paradigm with (6.3) as the learning stage and

$$\text{recall} = [\mathbf{c}]^+ \{\Sigma_k (\mathbf{a}+\mathbf{e}_k) * (\mathbf{b}+\mathbf{d}_k)/n\} \tag{6.4}$$

instead of (5.1) as the recall stage after n learning trials. In particular, if a is used as a cue to recall b after n trials, recall will equal

$$\begin{aligned}\text{recall} &= \mathbf{b} = [\mathbf{a}]^+ \Sigma_k (\mathbf{a}+\mathbf{e}_k) * (\mathbf{b}+\mathbf{d}_k)/n \\ &= \mathbf{b} + \Sigma_k \mathbf{a} * \mathbf{d}_k/n + \Sigma_k \mathbf{b} * \mathbf{e}_k/n + \Sigma_k \mathbf{e}_k * \mathbf{d}_k/n.\end{aligned} \tag{6.5}$$

As n increases, the approximation to b tends to improve because the three error terms approach zero. Thus, on repeatedly pairing the same stimuli a, b, recall will improve with the number of learning trials as long as the cue is similar to one of the two learned items. The learning rate can be controlled through the variances of the random components e_i, d_i. For example, to render the paradigm more realistic, the error variances could be made a function of the number of features, so that learning slows down and recall becomes more difficult as the stimuli become more complex, rather than the other way around, as the conventional correlation/convolution paradigm implies.

The revised paradigm (6.3), (6.4) is not offered as one more "theory" of distributed memory, because in terms of neurology it is just as unmotivated and implausible as the conventional convolution/correlation paradigm. However, compared to the latter, the revision does have a number formal points in its favor:

(a) It naturally combines with a simple learning paradigm which represents learning as successive extinction of random noise in the associations, rather than requiring that a separate adaptive mechanism be grafted on the storage/recall paradigm.
(b) There is no need to assume that the feature spaces from which **a** and **b** are selected are noiselike.
(c) There is no restriction that the cues required for perfect recall lie in selected subspaces.
(d) There is no requirement that the dimensions of the feature spaces be large.
(e) The paradigm applies to the feature vectors of the stimuli directly, instead of requiring an intervening transformation.
(f) On suitable choice of var(e) and var(d), recall becomes easier, not more difficult, with simpler stimuli.
(g) It narrows the gap between holographic and matrix models, especially those developed by Kohonen, who also employs Moore-Penrose inverses and the associated projectors.

7 Related Work

Finally, I wish to acknowledge my debt to Borsellino and Poggio's (1973) paper on the formal aspects of the convolution/correlation paradigm. In this important paper, the authors relate earlier results on nonassociative algebras by Albert (1942) to holographic memories. Most of the results stated here are already latent in the papers by Albert, and Borsellino and Poggio. However, the present focus is different and more specific. While Borsellino and Poggio did note the "isotopy" (a kind of weak equivalence between $\langle H, \# \rangle$ and $\langle H, * \rangle$ which in effect involves the involution we denoted S), they did not give any identities which would permit the resolution of expressions involving correlations into products of convolutions and involutions, as was done here. The authors also employed a different matrix representation, which is due to Albert. The representation suggested here capitalizes on the relation between between the shift operators and $\#$ and $*$ and lends itself to convenient computations of correlations as standard matrix products. This convenience has been achieved at the expense of limiting all computations to sequences with finite support in Re^n.

Acknowledgements. I would like to thank Professors Michael P. Drazin, Jerome Busemeyer, and the editor for their suggestions for improving the legibility of this manuscript.

References

Albert AA (1942) Non-associative algebras. I, II. Ann Math 43:685–723

Ben-Israel A, Greville TNE (1974) Generalized inverses: theory and applications. Wiley, New York

Borsellino A, Poggio T (1973) Convolution and correlation algebras. Kybernetik 13:113–122

Gabor D (1948) A new microscopic principle. Nature 161:777–778

Gabor D (1968) Holographic model for temporal recall. Nature 217:1288–1289

Gabor D (1969) Associative holographic memories. IBM J Res Dev 13:156–159

Julesz B, Pennington KS (1965) Equidistributed information maps: an analogy to holograms and memory. J Opt Soc Am 55:604

Kohonen T (1984) Selforganization and associative memory. Springer, Berlin Heidelberg New York Tokio

Kohonen T, Oja E, Lehtiö P (1981) Storage and processing of information in distributed associative memory systems. In: Hinton GE, Anderson JA (eds) Parallel models of associative memory. Erlbaum, Hillsdale

Longuet-Higgins HC (1968) Holographic model of temporal recall. Nature 217:104

Metcalfe Eich J (1982) A composite holographic associative recall model. Psychol Rev 89:627–661

Murdock BB (1974) Human memory: theory and data. Wiley, New York Toronto London Sidney

Murdock BB (1982) A theory for the storage and retrieval of item and associative information. Psychol Rev 89:609–626

Pike R (1984) Comparison of convolution and matrix distributed memory systems for associative recall and recognition. Psychol Rev 91:281–294

Reichardt W (1957) Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. Z Naturforsch 12b:448–457

Heerden PJ van (1963) A new optical method for storing and retrieving information. Appl Opt 2:387–392

Willshaw D (1981) Holography, associative memory, and inductive generalization. In: Hinton GE, Anderson JA (eds) Parallel models of associative memory. Erlbaum, Hillsdale

Willshaw D, Longuet-Higgins HC (1969) The holophone – recent developments. In: Michie D (ed) Machine intelligence 4. University Press, Edinburgh

Received: November 3, 1986

Peter H. Schönemann
Department of Psychological Sciences
Purdue University
West Lafayette, IN 47907
USA