recover the causes (rows of $\mathbf{W}$) have the simple relation: $\mathbf{W} = \mathbf{A}^{-1}$.

All that remains in defining an algorithm to learn $\mathbf{W}$ (and thus also $\mathbf{A}$) is to decide what constitutes a "cause". A number of proposals are considered in the Discussion, however, in the next two sections, we concentrate on algorithms producing causes which are decorrelated, and those attempting to produce causes that are statistically independent.

DECORRELATION AND INDEPENDENCE

The matrix, $\mathbf{W}$, is a decorrelating matrix when the covariance matrix of the output vector, $\mathbf{u}$, satisfies:

$$\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle = \text{diagonal matrix}. \tag{4}$$

In general, there will be many $\mathbf{W}$ matrices which decorrelate. For example, in the case of equation (2), when $\langle \mathbf{u}\mathbf{u}^T \rangle = \mathbf{I}$, then:

$$\mathbf{W}^{\mathrm{T}}\mathbf{W} = \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle^{-1} \tag{5}$$

which clearly leaves freedom in the choice of $\mathbf{W}$. There are, however, several special solutions to equation (5).

The orthogonal (global) solution [$\mathbf{W}\mathbf{W}^T = \mathbf{S}$]

Principal Components Analysis (PCA) is the orthogonal solution to equation (4). The principal components come from the eigenvectors of the covariance matrix, which are the columns of a matrix, $\mathbf{E}$, satisfying:

$$\mathbf{E}\mathbf{D}\mathbf{E}^{-1} = \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle \tag{6}$$

where $\mathbf{D}$ is the diagonal matrix of eigenvalues. Substituting equation (6) into equation (5) and solving for $\mathbf{W}$ gives the PCA solution, $\mathbf{W}_P$:

$$\mathbf{W}_P = \mathbf{D}^{-\frac{1}{2}}\mathbf{E}^{\mathrm{T}}. \tag{7}$$

This solution is unusual in that the filters (rows of $\mathbf{W}_P$) are orthogonal, so that $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{D}^{-1}$, a scaling matrix. These filters thus have several special properties:

1. The PCA filters define orthogonal directions in the vector space of the image.
2. The PCA basis functions (columns of $\mathbf{A}_P$, or rows of $\mathbf{W}_P^{\mathrm{T}}$—see Fig. 1) are just scaled versions of the PCA filters (rows of $\mathbf{W}_P$). This latter property is true because $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{D}^{-1}$ means that $\mathbf{W}^{-\mathrm{T}} = \mathbf{D}\mathbf{W}$.
3. When the image statistics are stationary (Field, 1994), the PCA filters are *global* Fourier filters, ordered according to the amplitude spectrum of the image.

Example PCA filters are shown in Fig. 3(a).

The symmetrical (local) solution [$\mathbf{W}\mathbf{W}^T = \mathbf{W}^2$]

If we force $\mathbf{W}$ to be symmetrical, so that $\mathbf{W}^{\mathrm{T}} = \mathbf{W}$, then the solution, $\mathbf{W}_Z$ to equation (5) is:

$$\mathbf{W}_Z = \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle^{-1/2}. \tag{8}$$

Like most other decorrelating filters, but unlike PCA, the basis functions and the filters coming from $\mathbf{W}_Z$ will be different from each other, and neither will be orthogonal. We might call this solution ZCA, since the filters it produces are zero-phase (symmetrical). ZCA is in several ways the polar opposite of PCA. It produces *local* (centre–surround type) whitening filters, which are ordered according to the phase spectrum of the image. That is, each filter whitens a given pixel in the image, preserving the spatial arrangement of the image and flattening its frequency (amplitude) spectrum. $\mathbf{W}_Z$ is related to the transforms described by Goodall (1960) and Atick & Redlich (1993).

Example ZCA filters and basis functions are shown in Fig. 3(b).

The independent (semi-local) solution [$f_{\mathbf{u}}(\mathbf{u}) = \Pi_i f_{u_i}(u_i)$]

Another way to constrain the solution is to attempt to produce outputs which are not just decorrelated, but statistically independent, the much stronger requirement of Independent Components Analysis, or ICA (Jutten & Hérault, 1991; Comon, 1994). The u_i are independent when their probability distribution, $f_{\mathbf{u}}$, factorizes as follows: $f_{\mathbf{u}}(\mathbf{u}) = \Pi_i f_{u_i} u_i$, equivalently, when there is zero mutual information between them: $I(u_i,u_j) = 0, \forall\, i \neq j$. A number of approaches to ICA have some relations with the one we describe below, notably Cardoso & Laheld (1996), Karhunen *et al.* (1996), Amari *et al.* (1996), Cichocki *et al.* (1994) and Pham *et al.* (1992). We refer the reader to these papers, to the two above, and to Bell & Sejnowski (1995a) for further background on ICA.

As we will show, in the Results, ICA on natural images produces decorrelating filters which are sensitive to both phase (locality) and frequency information, just as in transforms involving oriented Gabor functions (Daugman, 1985) or wavelets.* They are, thus, semi-local,

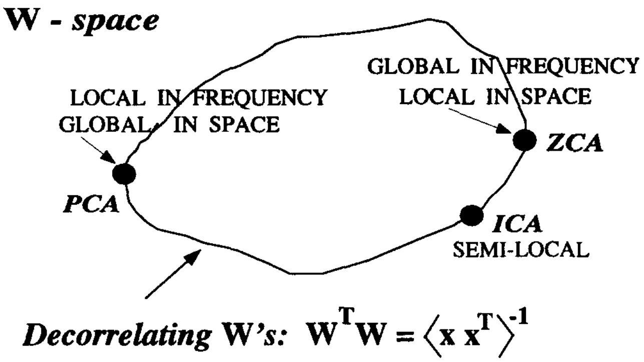

FIGURE 2. A schematic depiction of weight–space. A subspace of all matrices $\mathbf{W}$, here represented by the loop (of course it is a much higher-dimensional closed subspace), has the property of decorrelating the input vectors, $\mathbf{x}$. On this manifold, several special linear transformations can be distinguished: PCA (global in space and local in frequency), ZCA (local in space and global in frequency), and ICA, a privileged decorrelating matrix which, if it exists, decorrelates higher- as well as second-order moments. ICA filters are localized, but not down to the single pixel level, as ZCA filters are (see Fig. 3.)

*See the Proceedings of IEEE, 84, 4, April 1996—a special issue on wavelets.