Maths for Brain Imaging: Lecture 4

W.D. Penny
Wellcome Department of Imaging Neuroscience,
University College, London WC1N 3BG.

October 27, 2006

1 Linear algebra

1.1 Orthogonal Matrices

The set of vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are *orthogonal* if

$$\boldsymbol{q}_j^T \boldsymbol{q}_k = \begin{array}{ll} 0 & j \neq k \\ d_{jk} & j = k \end{array} \tag{1}$$

If these vectors are placed in columns of the matrix $\boldsymbol{Q}$ then

$$\boldsymbol{Q}^T \boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{D} \tag{2}$$

1.2 Orthonormal Matrices

The set of vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are *orthonormal* if

$$\boldsymbol{q}_j^T \boldsymbol{q}_k = \begin{array}{ll} 0 & j \neq k \\ 1 & j = k \end{array} \tag{3}$$

If these vectors are placed in columns of the matrix $\boldsymbol{Q}$ then

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I} \tag{4}$$

Hence, the transpose equals the inverse

$$\boldsymbol{Q}^T = \boldsymbol{Q}^{-1} \tag{5}$$

The vectors $\boldsymbol{q}_1..\boldsymbol{q}_k$ are said to provide an *orthonormal basis*. This means that *any* vector can be written as a linear combination of the basis vectors. A trivial example is the two-dimensional cartesian coordinate system where $\boldsymbol{q}_1 = [1, 0]^T$ (the x-axis) and $\boldsymbol{q}_2 = [0, 1]^T$ (the y-axis). More generally, to represent the vector $\boldsymbol{x}$ we can write

$$\boldsymbol{x} = \tilde{x}_1\boldsymbol{q}_1 + \tilde{x}_2\boldsymbol{q}_2 + ... + \tilde{x}_d\boldsymbol{q}_d \tag{6}$$

To find the appropriate coefficients $\tilde{x}_k$(the co-ordinates in the new basis), multiply both sides by $\boldsymbol{q}_k^T$. Due to the orthonormality property all terms on the right disappear except one leaving

$$\tilde{x}_k = \boldsymbol{q}_k^T \boldsymbol{x} \tag{7}$$

The new coordinates are the projections of the data onto the basis functions (re. definition of projections in earlier lecture, there is no denominator since $\boldsymbol{q}_k^T \boldsymbol{q}_k = 1$). In matrix form, equation 6 can be written as $\boldsymbol{x} = \boldsymbol{Q}\tilde{\boldsymbol{x}}$ which therefore has the solution $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{-1}\boldsymbol{x}$. But given that $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$ we have

$$\tilde{\boldsymbol{x}} = \boldsymbol{Q}^T \boldsymbol{x} \tag{8}$$

So for orthonormal bases, eg. Fourier or wavelets, data can be transformed from data to parameter space and vice-versa without inverse operators (not so for GLM with arbitrary design matrix).

1.3 Determinants

The determinant of a two-by-two matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{9}$$

is given by

$$\det(\boldsymbol{A}) = ad - bc \tag{10}$$

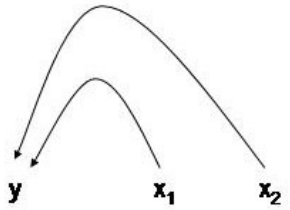
The determinant of a three-by-three matrix

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \tag{11}$$

is given by

$$\det(\boldsymbol{A}) = a \det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b \det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c \det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right) \tag{12}$$

Determinants are important because of their properties. In particular, if two rows of a matrix are equal then the



determinant is zero eg. if

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ a & b \end{bmatrix} \tag{13}$$

then

$$\det(\boldsymbol{A}) = ab - ba = 0 \tag{14}$$

In this case the transformation from $\boldsymbol{x} = [x_1, x_2]^T$ to $\boldsymbol{y} = [y_1, y_2]^T$ given by

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \tag{15}$$

reduces two pieces of information (x_1 and x_2) to one piece of information

$$y = y_1 = y_2 = ax_1 + bx_2 \tag{16}$$

In this case it is not possible to reconstruct $\boldsymbol{x}$ from $\boldsymbol{y}$; the transformation is not invertible - the matrix $\boldsymbol{A}$ does not have an inverse and the value of the determinant provides a test for this: If $\det(\boldsymbol{A}) = 0$ the matrix $\boldsymbol{A}$ is not invertible; it is *singular*. Conversely, if $\det(\boldsymbol{A}) \neq 0$ then $\boldsymbol{A}$ *is* invertible.

Another important property of determinants is that they measure the 'volume' of a matrix. For a 3-by-3 matrix the three rows of the matrix form the edges of a cube. The determinant is the volume of this cube. For a d-by-d matrix the rows form the edges of a 'parallepiped'. Again, the determinant is the volume.

We also write

$$\det(\boldsymbol{A}) = |\boldsymbol{A}| \tag{17}$$

1.4 Eigenanalysis

The square matrix $\boldsymbol{A}$ has eigenvalues λ and eigenvectors $\boldsymbol{q}$ if

$$\boldsymbol{A}\boldsymbol{q} = \lambda\boldsymbol{q} \tag{18}$$

Therefore

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{q} = 0 \tag{19}$$

To satisfy this equation either $\boldsymbol{q} = 0$, which is uninteresting, or the matrix $\boldsymbol{A} - \lambda\boldsymbol{I}$ must reduce $\boldsymbol{q}$ to the null vector (a single point). For this to happen $\boldsymbol{A} - \lambda\boldsymbol{I}$ must be singular. Hence

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0 \tag{20}$$

Eigenanalysis therefore proceeds by (i) solving the above equation to find the eigenvalues λ_i and then (ii) substituting them into equation 18 to find the eigenvectors. For example, if

$$\boldsymbol{A} = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix} \tag{21}$$

then

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = (4 - \lambda)(-3 - \lambda) - (-5)(2) = 0 \quad (22)$$

which can be rearranged as

$$\begin{aligned} \lambda^2 - \lambda - 2 &= 0 \\ (\lambda + 1)(\lambda - 2) &= 0 \end{aligned} \quad (23)$$

Hence the eigenvalues are $\lambda = -1$ and $\lambda = 2$. Substituting back into equation 18 gives an eigenvector $\boldsymbol{q}_1$ which is any multiple of $[1, 1]^T$. Similarly, eigenvector $\boldsymbol{q}_2$ is any multiple of $[5, 2]^T$.

We now note that the determinant of a matrix is also equal to the product of its eigenvalues

$$\det(\boldsymbol{A}) = \prod_k \lambda_k \quad (24)$$

We also define the *Trace* of a matrix as the sum of its diagonal elements

$$Tr(\boldsymbol{A}) = \sum_k a_{kk} \quad (25)$$

and note that it is also equal to the sum of the eigenvalues

$$Tr(\boldsymbol{A}) = \sum_k \lambda_k \tag{26}$$

Eigenanalysis applies only to square matrices.

1.5 Diagonalization

If we put the eigenvectors into the columns of a matrix

$$\boldsymbol{Q} = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & . & \boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \tag{27}$$

then, because, $\boldsymbol{A}\boldsymbol{q}_k = \lambda_k \boldsymbol{q}_k$, we have

$$\boldsymbol{A}\boldsymbol{Q} = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \lambda_1\boldsymbol{q}_1 & \lambda_2\boldsymbol{q}_2 & . & \lambda_d\boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \tag{28}$$

If we put the eigenvalues into the matrix $\boldsymbol{\Lambda}$ then the above matrix can also be written as $\boldsymbol{Q\Lambda}$. Therefore,

$$\boldsymbol{AQ} = \boldsymbol{Q\Lambda} \tag{29}$$

Pre-multiplying both sides by $\boldsymbol{Q}^{-1}$ gives

$$\boldsymbol{Q}^{-1}\boldsymbol{AQ} = \boldsymbol{\Lambda} \tag{30}$$

This shows that any square matrix can be converted into a diagonal form (provided it has distinct eigenvalues; see eg. [4] p. 255).

1.6 Spectral Theorem

For any real *symmetric* matrix all the eigenvalues will be real and there will be d distinct eigenvalues and orthogonal eigenvectors. They can be normalised and placed into the matrix $\boldsymbol{Q}$. Since $\boldsymbol{Q}$ is now orthonormal we have $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$. Hence

$$\boldsymbol{Q}^T\boldsymbol{AQ} = \boldsymbol{\Lambda} \tag{31}$$

Pre-multiplying by $\boldsymbol{Q}$ and post-multiplying by $\boldsymbol{Q}^T$ gives

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \tag{32}$$

which is known as the *spectral theorem*. It says that any real, symmetric matrix can be represented as above where the columns of $\boldsymbol{Q}$ contain the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues, λ_i. Equivalently,

$$\boldsymbol{A} = \begin{bmatrix} | & | & . & | \\ | & | & . & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & . & \boldsymbol{q}_d \\ | & | & . & | \\ | & | & . & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & & \\ & & & \lambda_d \end{bmatrix} \begin{bmatrix} - & - & \boldsymbol{q}_1 & - & - \\ - & - & \boldsymbol{q}_2 & - & - \\ . & . & . & & . \\ - & - & \boldsymbol{q}_d & - & - \end{bmatrix} \tag{33}$$

This can also be written as a summation

$$\boldsymbol{A} = \sum_{k=1}^{d} \lambda_k \boldsymbol{q}_k \boldsymbol{q}_k^T \tag{34}$$

This provides a particularly efficient way to compute powers of matrices

$$\boldsymbol{A}^k = \boldsymbol{Q}\boldsymbol{\Lambda}^k\boldsymbol{Q}^T \tag{35}$$

This is particularly useful for solving multivariate difference and differential equations (see later lecture). Using the above with $k = -1$ shows $\det(A^{-1}) = 1/\det(A)$.

1.7 Quadratic Forms

The quadratic function

$$f(\boldsymbol{x}) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + ... + a_{dd}x_d^2 \quad (36)$$

can be written in matrix form as

$$f(\boldsymbol{x}) = [x_1, x_2, ..., x_d] \begin{bmatrix} a_{11} & a_{12} & a_{1d} \\ a_{21} & a_{22} & a_{2d} \\ & & \\ & & \\ a_{d1} & a_{d2} & a_{dd} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ x_d \end{bmatrix} \quad (37)$$

which is written compactly as

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \quad (38)$$

If $f(\boldsymbol{x}) > 0$ for any non-zero $\boldsymbol{x}$ then $\boldsymbol{A}$ is said to be positive-definite. Similarly, if $f(\boldsymbol{x}) \geq 0$ then $\boldsymbol{A}$ is positive-semi-definite.

If we substitute $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ and $\boldsymbol{x} = \boldsymbol{Q}\boldsymbol{y}$ where $\boldsymbol{y}$ are the projections onto the eigenvectors, then we can write

$$f(\boldsymbol{x}) = \boldsymbol{y}^T \boldsymbol{\Lambda} \boldsymbol{y} \quad (39)$$

$$= \sum_i y_i^2 \lambda_i$$

Hence, for positive-definiteness we must therefore have $\lambda_i > 0$ for all i (ie. positive eigenvalues).

2 Principal Component Analysis

Given a set of data vectors $\{\boldsymbol{x}_n\}$ we can construct a covariance matrix

$$\boldsymbol{C} = \frac{1}{N} \sum_n (\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T \tag{40}$$

or, if we construct a matrix $\boldsymbol{X}$ with rows equal to $\boldsymbol{x}_n - \bar{\boldsymbol{x}}$ then

$$\boldsymbol{C} = \frac{1}{N} \boldsymbol{X}^T \boldsymbol{X} \tag{41}$$

Because covariance matrices are real and symmetric we can apply the spectral theorem

$$\boldsymbol{C} = \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^T \tag{42}$$

If the eigenvectors (columns of $\boldsymbol{Q}$) are normalised to unit length, they constitute an orthonormal basis. If the eigenvalues are then ordered in magnitude such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d$ then the decomposition is known as Principal Component Analysis (PCA). The projection of a data point $\boldsymbol{x}_n$ onto the principal components is

$$\boldsymbol{y}_n = \boldsymbol{Q}^T \boldsymbol{x}_n \tag{43}$$

The mean projection is

$$\bar{\boldsymbol{y}} = \boldsymbol{Q}^T \bar{\boldsymbol{x}} \tag{44}$$

The covariance of the projections is given by the matrix

$$\boldsymbol{C}_y = \frac{1}{N} \sum_n (\boldsymbol{y}_n - \bar{\boldsymbol{y}})(\boldsymbol{y}_n - \bar{\boldsymbol{y}})^T \tag{45}$$

Substituting in the previous two expressions gives

$$\begin{aligned} \boldsymbol{C}_y &= \frac{1}{N} \sum_n \boldsymbol{Q}^T (\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T \boldsymbol{Q} \qquad (46) \\ &= \boldsymbol{Q}^T \boldsymbol{C} \boldsymbol{Q} \\ &= \boldsymbol{\Lambda} \end{aligned}$$

where $\mathbf{\Lambda}$ is the diagonal eigenvalue matrix with entries λ_k ($\sigma_k^2 = \lambda_k$). This shows that the variance of the kth projection is given by the kth eigenvalue. Moreover, it says that the projections are uncorrelated. PCA may therefore be viewed as a linear transform

$$\boldsymbol{y} = \boldsymbol{Q}^T \boldsymbol{x} \tag{47}$$

which produces uncorrelated data.

2.1 The Multivariate Gaussian Density

In d dimensions the general multivariate normal probability density can be written

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{C}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T C^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}})\right) \tag{48}$$

where the mean $\bar{\boldsymbol{x}}$ is a d-dimensional vector, $\boldsymbol{C}$ is a $d \times d$ covariance matrix, and $|\boldsymbol{C}|$ denotes the determinant of $\boldsymbol{C}$. Because the determinant of a matrix is the product of its eigenvalues then for covariance matrices, where the

eigenvalues correspond to variances, the determinant is a single number which represents the total volume of variance. The quantity

$$M(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}})^T C^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}}) \tag{49}$$

which appears in the exponent is called the *Mahalanobis distance* from $\boldsymbol{x}$ to $\bar{\boldsymbol{x}}$.

2.2 Singular Value Decomposition

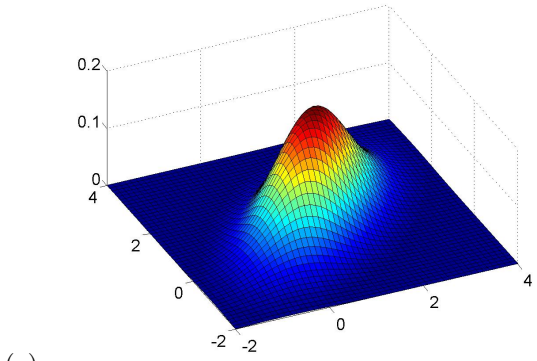
The eigenvalue-eigenvector factorisation (see equation 32)

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \tag{50}$$

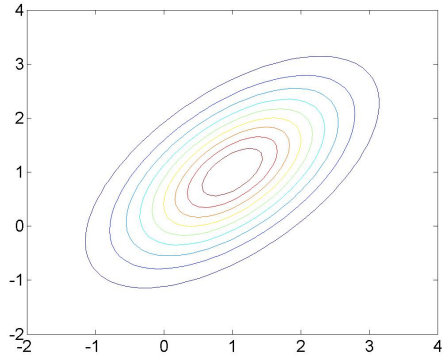
applies to real symmetric matrices only. There is an equivalent factorisation for rectangular matrices, having N rows and d columns, called Singular Value Decomposition (SVD)

$$\boldsymbol{A} = \boldsymbol{Q}_1 \boldsymbol{D} \boldsymbol{Q}_2^T \tag{51}$$

where $\boldsymbol{Q_1}$ is an orthonormal N-by-N matrix, $\boldsymbol{Q_2}$ is an orthonormal d-by-d matrix, $\boldsymbol{D}$ is a diagonal matrix of



(a)



(b)

Figure 1: *(a) 3D-plot and (b) contour plot of Multivariate Gaussian PDF with* $\boldsymbol{\mu} = [1, 1]^T$ *and* $\boldsymbol{C}_{11} = \boldsymbol{C}_{22} = 1$ *and* $\boldsymbol{C}_{12} = \boldsymbol{C}_{21} = 0.6$ *ie a positive correlation of* $r = 0.6$.

dimension N-by-d and the kth diagonal entry in $\boldsymbol{D}$ is known as the kth singular value, σ_k.

If we substitute the SVD of $\boldsymbol{A}$ into $\boldsymbol{A}^T\boldsymbol{A}$, after some rearranging, we get

$$\boldsymbol{A}^T\boldsymbol{A} \;=\; \boldsymbol{Q}_2\boldsymbol{D}^T\boldsymbol{D}\boldsymbol{Q}_2^T \tag{52}$$

which is of the form $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$ where $\boldsymbol{Q} = \boldsymbol{Q}_2$ and $\boldsymbol{\Lambda} = \boldsymbol{D}^T\boldsymbol{D}$. This shows that the columns of $\boldsymbol{Q}_2$ contain the eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ and that $\boldsymbol{D}$ contains the square roots of the corresponding eigenvalues. Similarly, by substituting the SVD of $\boldsymbol{A}$ into $\boldsymbol{A}\boldsymbol{A}^T$ we can show that the columns of $\boldsymbol{Q}_1$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$.

2.2.1 Relation to PCA

Given a data matrix $\boldsymbol{X}$ constructed as before (see PCA section), except that the matrix is scaled by a normalisation factor $\sqrt{1/N}$, then $\boldsymbol{X}^T\boldsymbol{X}$ is equivalent to the covariance matrix $\boldsymbol{C}$. If we therefore decompose $\boldsymbol{X}$ using SVD, the principal components will apear in $\boldsymbol{Q}_2$ and the

square roots of the corresponding eigenvalues will appear in $\boldsymbol{D}$.

Therefore we can implement PCA in one of two ways (i) compute the covariance matrix and perform an eigen-decomposition or (ii) use SVD directly on the (normalised) data matrix.

See eg. `alan_svd.m`.

2.3 PET verbal fluency data

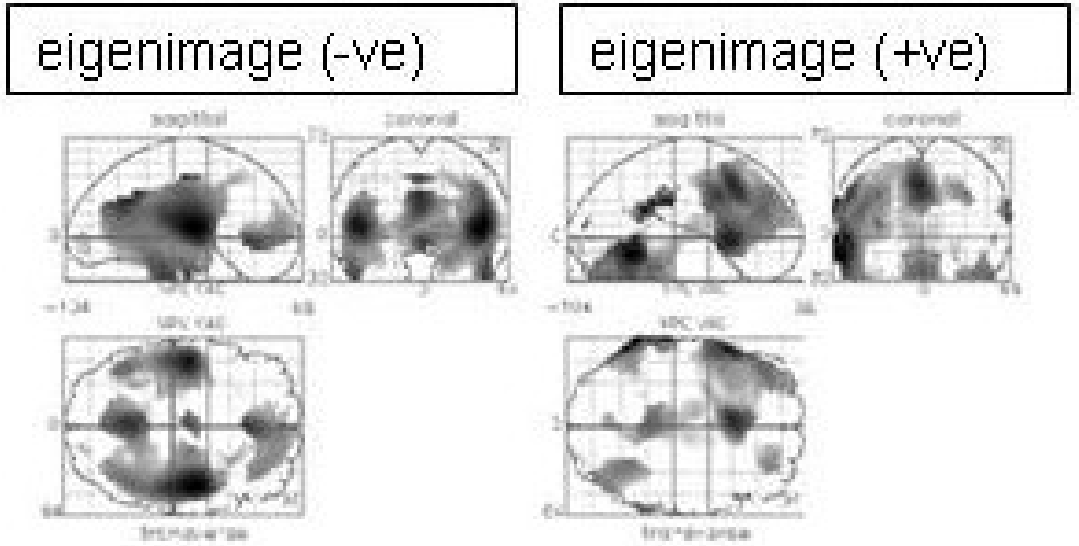
Subject scanned under two alternating conditions (i) word generation and (ii) word shadowing. Six repetitions of each. GLM analysis to select voxels showing significant variation over the 12 scans. Zero mean voxel activities over scans.

Create matrix M of dimension $N_{scans} \times N_{voxels}$. Application of SVD

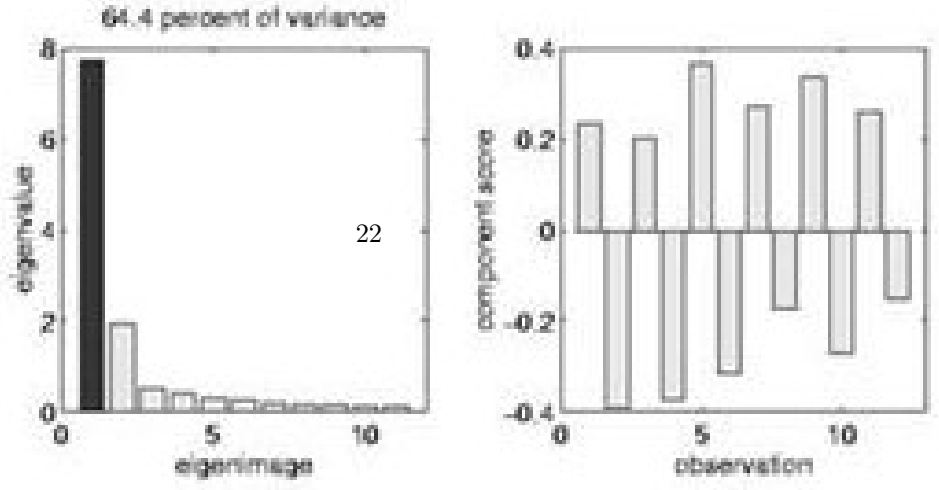
$$USV^T = M \tag{53}$$

places temporal components (eigenvariates) in columns of U and spatial components (eigenimages) in columns of V. Diagonal elements in S show that first mode accounts for 64% variance, second 16%.

First eigenimage has positive loadings in anterior cingulate, left DLPFC, Broca's area, thalamic nuclei and cerebellum (regions showing higher activity in generation than shadowing). Negative loadings bitemporally and in posterior cingulate.

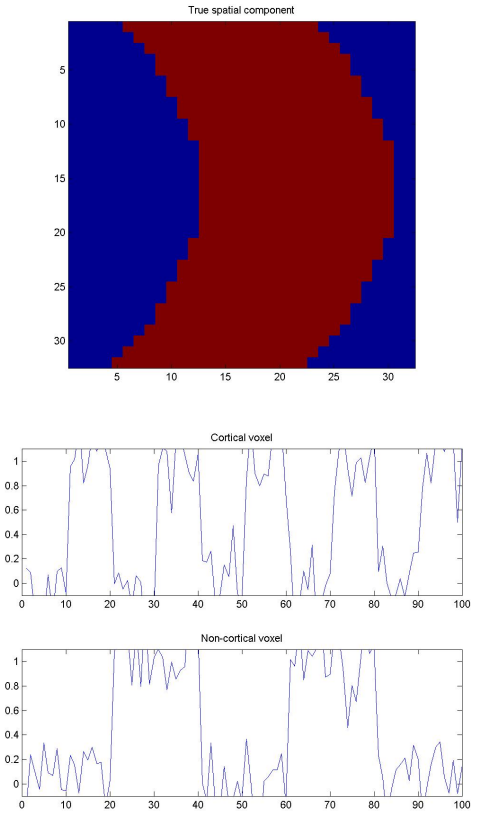


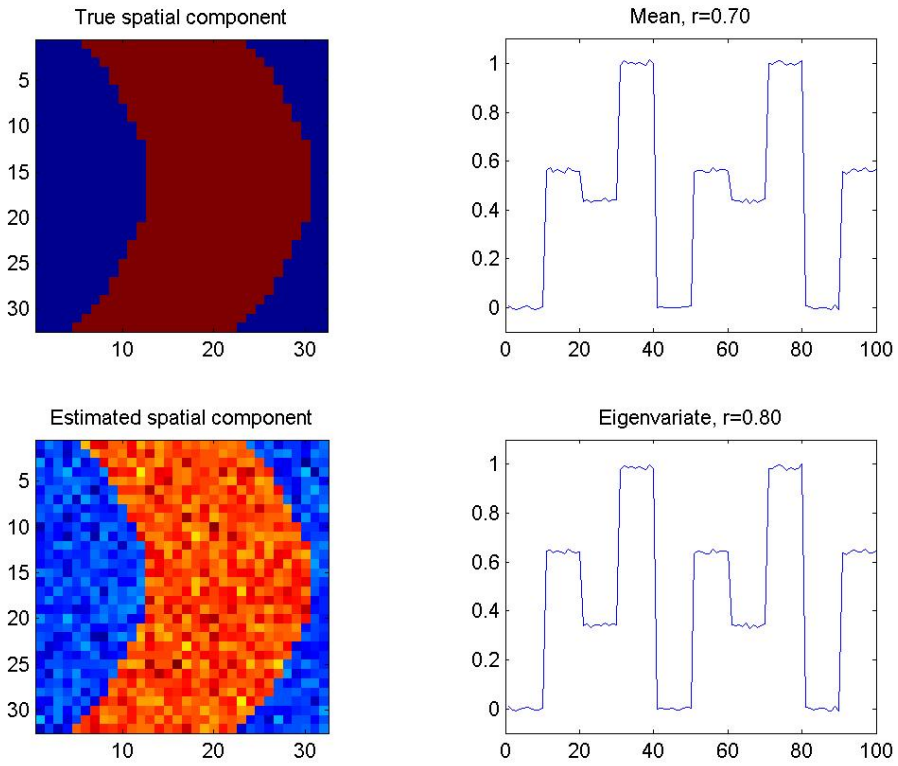
Eigenimage analysis:



2.4 Summarising regional activity

See `region_svd.m`


True spatial component
5
10
15
20
25
30
5
10
15
20
25
30
Cortical voxel
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
80
90
100
Non-cortical voxel
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
80
90
100


True spatial component
Mean, r=0.70
Estimated spatial component
Eigenvariate, r=0.80

3 Structural Equation Modelling

Structural Equation Models (SEMs) comprise a set of regions and a set of directed connections. Importantly, a causal semantics is ascribed to these connections where an arrow from A to B means that A causes B. Causal relationships are thus not inferred from the data but are assumed a-priori [2].

We consider networks comprising N regions in which the activity at time t is given by the $N \times 1$ vector y_t. If there are T time points and Y is an $N \times T$ data matrix comprising $t = 1..T$ such vectors then the likelihood of the data is given by

$$p(Y|\theta) = \prod_{t=1}^{T} p(y_t|\theta) \tag{54}$$

where θ are the parameters of an SEM.

The second SEM equation specifies the generative model at time t

$$p(y_t|\theta) = \mathsf{N}\left(y_t; 0, \Sigma(\theta)\right) \tag{55}$$

which denotes that the activities are zero mean Gaussian variates with a covariance, $\Sigma(\theta)$, that is a function of the connectivity matrix θ. The form of this function is specified implicitly by the regression equation that describes how activity in one area is related to activity in other areas via a set of path coefficients, M, as

$$y_t = My_t + e_t \tag{56}$$

where e_t are zero mean Gaussian innovations or errors of covariance R. Typically R will be a diagonal matrix and we write the error variance in region i as σ_i^2. Regions are connected together via the $N \times N$ path coefficient matrix M where the M_{ij} denotes a connection from region j to region i. The parameters of an SEM, θ, are the unknown elements of M and R. Re-write as

$$y_t = (I_N - M)^{-1} e_t \tag{57}$$

This form is particularly useful as it shows us how to generate data from the model. Firstly, we generate the Gaussian variates e_t and then pre-multiply by $(I_N -$

$M)^{-1}$. This is repeated for each t. This form also allows us to express the covariance of y_t as a function of θ

$$\Sigma(\theta) = (I_N - M)^{-1} R (I_N - M)^{-T} \tag{58}$$

3.1 Estimation

Given a set of parameters θ we can compute the likelihood of a data set from equations 54, 55 and 58. Given a data set one can therefore find the connectivity matrix that maximises the (log) likelihood using standard optimisation methods [3].

$$L(\theta) = -\frac{T}{2} \log |\Sigma(\theta)| - \frac{NT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} y_t^T \Sigma(\theta)^{-1} y_t \tag{59}$$

If we define the sample covariance as

$$S = \frac{1}{T} \sum_{t=1}^{T} y_t y_t^T \tag{60}$$

then, by noting that the last term is a scalar and that the trace of a scalar is that same scalar value, and using the circularity property of the trace operator (that is, $\mathsf{Tr}(AB) = \mathsf{Tr}(BA)$), we can write

$$L(\theta) = -\frac{T}{2}\log|\Sigma(\theta)| - \frac{NT}{2}\log 2\pi - \frac{T}{2}\mathsf{Tr}(S\Sigma(\theta)^{-1}) \quad (61)$$

If we use unbiased estimates of the sample covariance matrix then we replace T's in the above equation by $T-1$'s. If we now also drop those terms that are not dependent on the model parameters we get

$$L(\theta) \;=\; -\frac{T-1}{2}\left(\log|\Sigma(\theta)| + \mathsf{Tr}(S\Sigma(\theta)^{-1})\right) \quad (62)$$

Maximum likelihood estimates can therefore be obtained by maximising the above function.

3.2 Inference

Statistical inference is based on the likelihood ratio for comparing models i and j is

$$R_{ij} = \frac{p(Y|\theta, m = i)}{p(Y|\theta, m = j)} \tag{63}$$

If $L(\theta_i)$ and $L(\theta_j)$ are the corresponding log-likelihoods then the log of the likelihood ratio is

$$\log R_{ij} = L(\theta_i) - L(\theta_j) \tag{64}$$

Under the null hypothesis that the models are identical, and for large T, $-2 \log R_{ij}$ is distributed as a chi-squared variable having degrees of freedom equal to the difference in number of parameters between the models (see p.265 in [1]). This only applies to nested models.

A special case of the above test arises when one wishes to evaluate the goodnees of fit of a single model. We will denote this as 'model 1'. This can be achieved by comparing the likelihood of model 1 to the likelihood

of the least restrictive (most complex) model one could possibly adopt ('model 0') with covariance equal to the sample covariance ie. $\Sigma(\theta) = S$. The has likelihood

$$\begin{aligned} L_0 &= -\frac{T-1}{2}\left(\log|S| + \mathsf{Tr}(SS^{-1})\right) \qquad (65) \\ &= -\frac{T-1}{2}\left(\log|S| + N\right) \qquad (66) \end{aligned}$$

The corresponding (log) likelihood ratio is

$$\log R_{10} = -\frac{T-1}{2}\left(\log|\Sigma(\theta)| + \mathsf{Tr}(S\Sigma(\theta)^{-1}) - \log|S| - N\right) \qquad (67)$$

which in turn has a corresponding chi-squared value

$$\chi^2 = (T-1)F(\theta) \qquad (68)$$

where

$$F(\theta) = \log|\Sigma(\theta)| + \mathsf{Tr}(S\Sigma(\theta)^{-1}) - \log|S| - N \qquad (69)$$

The corresponding degrees of freedom are equal to the degrees of freedom in model 0, k, minus the degrees of

freedom in model 1, q. For an N-dimensional covariance matrix there are k=N(N+1)/2 degrees of freedom. For model 1, q equals the total number of connectivity and variance parameters to be estimated. The associate χ^2 test provides a way of assessing if an SEM fits the data sufficiently.

For more general model comparisons the χ^2 statistic associated with the LR test can be written as

$$\chi^2 = (T-1)(F(\theta_1) - F(\theta_2)) \tag{70}$$

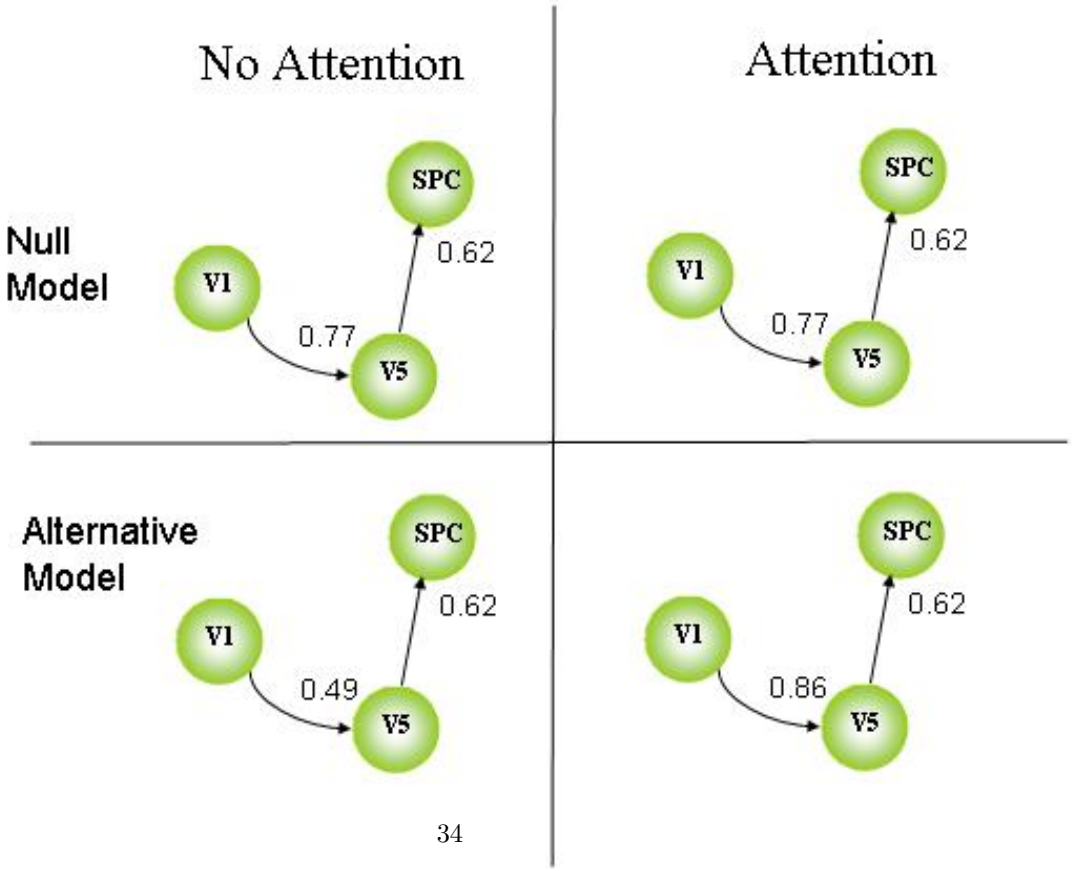
3.3 Attention to visual motion fMRI data

We first use a feedforward architecture. The null model has all parameters fixed between conditions giving $k = 8$ (two path coefficients and six error variance parameters). The alternative model allows V1-V5 to change giving $q = 9$. The alternative model fits better ($p = 0.003$).

But in comparison to the sample covariance, where $k = N(N+1)/2 = 6$ degrees of freedom per data set and two data sets (two conditions) ie. $k = 12$, the alternative

model is sig. different ($p < 1e - 5$). It is therefore not a good model of the data.

Feedforward SEM



Feedforward SEM

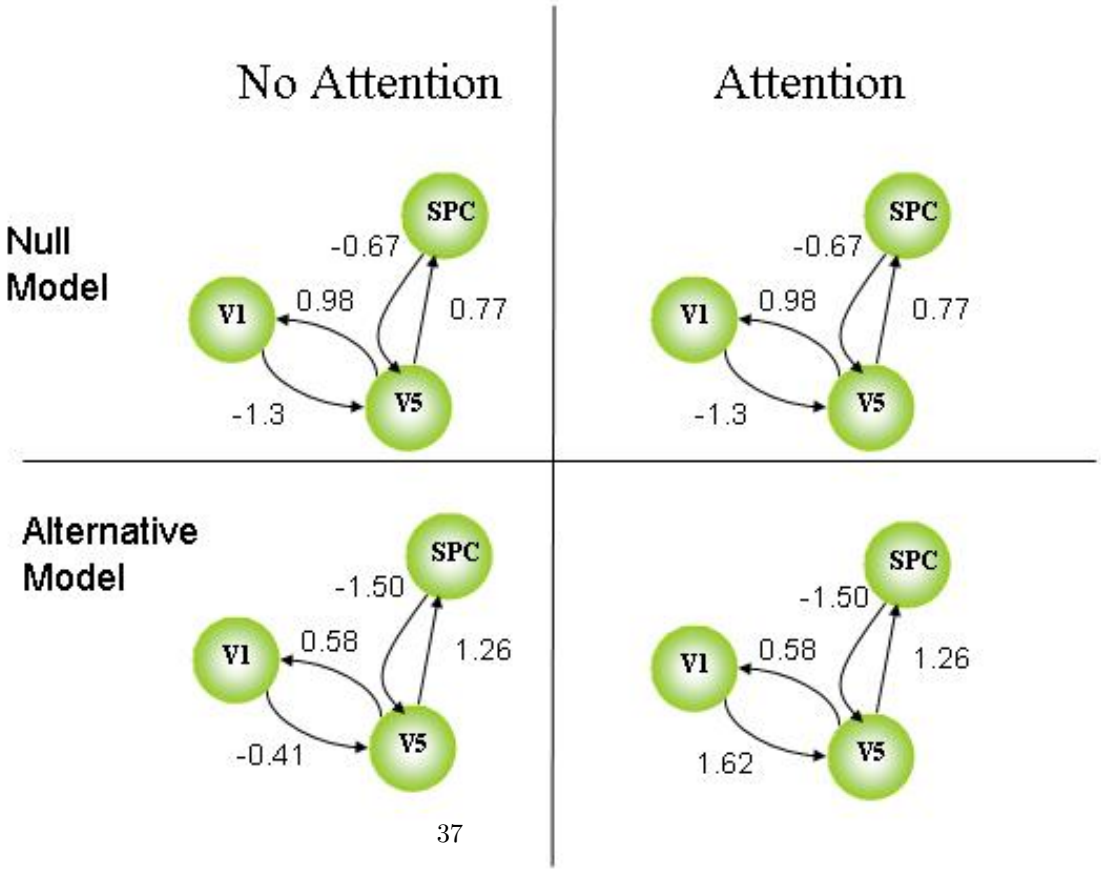
	No Attention	Attention
Null Model	**Sample covariance matrix, S:** 1.0000 0.4881 0.0724 0.4881 1.0000 0.5013 0.0724 0.5013 1.0000	**Sample covariance matrix, S:** 1.0000 0.8605 0.7810 0.8605 1.0000 0.6963 0.7810 0.6963 1.0000
Alternative Model	**Alternative Model covariance matrix, $\Sigma(\theta)$:** 1.0000 0.4881 0.3012 0.4881 1.0000 0.6171 0.3012 0.6171 1.1429	**Alternative Model covariance matrix , $\Sigma(\theta)$:** 1.0000 0.8605 0.5310 0.8605 1.0000 0.6171 0.5310 0.6171 0.9023

We now use a recpirocal architecture. The null model has all parameters fixed between conditions giving $k = 10$. The alternative model allows V1-V5 to change giving $q = 11$. The alternative model fits better ($p = 9e - 6$).

In comparison to the sample covariance, where $k = 12$, the alternative model is not sig. different ($p = 0.05$) (well, its borderline !). It is therefore an acceptable model of the data.

As compared to the feedforward model, the correlations between V1 and SPC are modelled more accurately (at a minor cost of not modelling V1-V5 and V5-SPC correlations quite so accurately).

Reciprocal SEM



Reciprocal SEM

	No Attention	Attention
Null Model	Sample covariance matrix, S : 1.0000 0.4881 0.0724 0.4881 1.0000 0.5013 0.0724 0.5013 1.0000	Sample covariance matrix, S : 1.0000 0.8605 0.7810 0.8605 1.0000 0.6963 0.7810 0.6963 1.0000
Alternative Model	Alternative Model covariance matrix, Σ(θ) : 0.9489 0.4063 0.1419 0.4063 0.8697 0.4615 0.1419 0.4615 1.1061	Alternative Model covariance matrix, Σ(θ) : 1.0346 0.9170 0.7807 0.9170 1.0923 0.7342 0.7807 0.7342 0.9495

References

[1] K.A. Bollen. *Structural Equations with Latent Variables*. John Wiley, New York, 1989.

[2] J. Pearl. Graphs, causality, and structural equation models. *Sociological Methods and Research*, 27:226–284, 1998.

[3] W. H. Press, S.A. Teukolsky, W.T. Vetterling, and B.V.P. Flannery. *Numerical Recipes in C*. Cambridge, 1992.

[4] G. Strang. *Linear algebra and its applications*. Harcourt Brace, 1988.