STATISTICAL SIGNAL PROCESSING

MATRIX METHODS

P.R. WHITE

EIGENVALUES AND EIGENVECTORS

SYMMETRIC MATRICES

POSITIVE DEFINITE MATRICES

RANK

COLUMN AND NULL SPACES

SOLUTION OF SYSTEMS OF EQUATIONS

SINGULAR VALUE DECOMPOSITION

PSEUDO-INVERSE

APPLICATIONS

 © ISVR 1999

EIGENVALUES AND EIGENVECTORS

Consider a **square** matrix A of dimension $L \times L$.

The eigenvalues and eigenvectors are properties of A. They derive from the relation:

$$A\underline{u} = \lambda \underline{u} \qquad (1)$$

where λ is a scalar and $\underline{u}$ is an $L \times 1$ (column) vector.

The scalars λ which satisfy (1) are called the eigenvalues and each is associated with an eigenvector, $\underline{u}$.

Equation (1) can be interpreted by noting that the eigenvectors are vectors such that if one pre-multiplies them by A one obtains the original vector scaled by the value λ.

If $\underline{u}$ satisfies (1) then $\alpha\underline{u}$ will also satisfy (1), where α is any scalar.

 © ISVR 1999

EIGENVALUES

One can (in principle) compute the eigenvalues using the fact that (1) implies that:

$$\left|A-\lambda I\right|=0 \qquad (2)$$

where I is the $L\times L$ identity matrix and |.| denotes the determinant.

The LHS of (2) represents an Lth order polynomial in λ, the eigenvalues are the L roots of this equation.

Most commonly there are L different eigenvalues; they are potentially complex.

These eigenvalues will be written $\lambda_1,\lambda_2, \ldots \lambda_L$, and we shall order them such that

$$\left|\lambda_1\right| \geq \left|\lambda_2\right| \geq \left|\lambda_3\right| \geq .... \geq \left|\lambda_L\right|$$

Note that if the elements of A are real then the eigenvalues must occur in conjugate pairs.

 © ISVR 1999

EIGENVECTORS

For every eigenvalue, λ_k , there is an associated eigenvector $\underline{u}_k$.

The eigenvector can be computed by solving the homogeneous equation

$$\left(A-\lambda_k I\right)\underline{u}_k = 0 \qquad (3)$$

If $\underline{u}_k$ satisfies (3) then so does $c\underline{u}_k$ where c is any scalar.

The eigenvectors are specified in some normalised form, usually one normalises the eigenvectors so that

$$\underline{u}_k{}^H \underline{u}_k = \sum_{p=1}^{L} \left|u_k(p)\right|^2 = 1$$

^{H} denotes the conjugate transpose and $u_k(p)$ is the pth element of $\underline{u}_k$.

This form of normalisation will be used throughout.

 © ISVR 1999

EIGENVALUES AND EIGENVECTORS OF SYMMETRIC MATRICES[1]

If the matrix A is real and symmetric, *i.e.* $A^t=A$, then the eigen representation has a special form:

1) The eigenvalues and eigenvectors are all real.

2) The eigenvectors are orthogonal, *i.e.*

$$\underline{u}_m{}^t \underline{u}_n = \sum_{p=1}^{L} u_m(p)u_n(p) = 0 \qquad \text{for all } m \neq n$$

and are orthonormal since they are normalised such that:

$$\underline{u}_m{}^t \underline{u}_m = 1$$

3) The matrix A can be decomposed as

$$A=U\Lambda U^t$$

where U and Λ are matrices defined as

$$U = \left[\underline{u}_1 \underline{u}_2 \cdots \underline{u}_L\right] \quad \Lambda = diag\left\{\lambda_1, \lambda_2, \cdots, \lambda_L\right\}$$

The columns of U are the eigenvectors of A.

U is referred to as being *orthonormal*, since $U^tU=I$

[1] A symmetric matrix MUST also be square.

 © ISVR 1999

EIGEN DECOMPOSITION OF THE INVERSE OF A SYMMETRIC MATRIX

One can write

$$A = U\Lambda U^t = \sum_{n=1}^{L} \lambda_n \underline{u}_n \underline{u}_n{}^t$$

Further

$$A^{-1} = \sum_{n=1}^{L} \frac{1}{\lambda_n} \underline{u}_n \underline{u}_n{}^t$$

Clearly if any eigenvalue is zero then the inverse of A does not exist.

It can be shown that in general (not just for symmetric matrices) that

$$|A| = \prod_{k=1}^{L} \lambda_k = \lambda_1 \times \lambda_2 \times \cdots \lambda_L$$

Hence if any eigenvalue is zero then $|A|$ is zero, *i.e.* A^{-1} does not exist.

A matrix that can not be inverted is said to be *singular*.

 © ISVR 1999

POSITIVE DEFINITE MATRICES AND QUADRATIC FORMS

A matrix is *positive definite* if all its eigenvalues are greater than 0.

It is *positive semi-definite* if all its eigenvalues are non-negative (allowing some eigenvalues to be zero).

A *quadratic form* is

$$q(\underline{y}) = \underline{y}^t A \underline{y}$$

where $\underline{y}$ is a column vector.

If A is symmetric then since $A = \sum_{k=1}^{L} \lambda_k \underline{u}_k \underline{u}_k^{\,t}$

$$q(\underline{y}) = \sum_{k=1}^{L} \lambda_k \left(\underline{u}_k^{\,t} \underline{y} \right)^2$$

For a positive semi-definite matrix λ_k is non-negative and $\left(\underline{u}_k^{\,t} \underline{y} \right)^2$ is a non-negative scalar, so $q(\underline{y})$ must also be non-negative (for all $\underline{y}$).

The statement that $q(\underline{y})$ is positive all non-zero choices of $\underline{y}$, is an alternative definition of positive definiteness.

 © ISVR 1999

RANK

The *rank* of a square matrix, *r*, is the number of non-zero eigenvalues it possesses.

A square matrix is said to be of *full rank* if all its eigenvalues are non-zero, ($r=L$), otherwise the matrix is said to be *rank deficient* ($r<L$).

Rank deficient matrices are singular (have no inverse).

The rank of a matrix also defines the number of linearly independent columns/rows it possesses.

If the matrix is of full rank then no one column/row can be made by forming linear combinations of the other columns/rows.

If a matrix is rank deficient then there are $L-r$ columns/rows that can be formed by a linear combination of the other columns/rows.

 © ISVR 1999

COLUMN[2] SPACE OF A (SYMMETRIC) MATRIX

A vector $\underline{y}$ is said to be in the *column space* of A if it can be constructed by a linear combination of the columns of A, *i.e.* if

$$\underline{y} = \alpha_1 \underline{a}_1 + \alpha_2 \underline{a}_2 + \cdots + \alpha_L \underline{a}_L = A\underline{\alpha}$$

$\underline{a}_k$ is the kth column of A, α_k is a scalar, $\underline{\alpha}$ is a column vector of the α_k's.

An orthonormal basis for the column space of a symmetric matrix is the set of eigenvectors associated with the non-zero eigenvalues.

If $\underline{y}$ is in the column space then one can also write

$$\underline{y} = \beta_1 \underline{u}_1 + \beta_2 \underline{u}_2 + \cdots + \beta_r \underline{u}_r = U_c \underline{\beta}$$

The matrix U_c is defined as $U_c = \left[\underline{u}_1 \underline{u}_2 \cdots \underline{u}_r\right]$

$$A\underline{\alpha} = \left\{ \sum_{k=1}^{r} \lambda_k \left(\underline{u}_k \underline{u}_k^{\,t} \right) \right\} \underline{\alpha} = \sum_{k=1}^{r} \lambda_k \left(\underline{u}_k^{\,t} \underline{\alpha} \right) \underline{u}_k = U_c \underline{\beta}$$

with $\beta_k = \lambda_k (\underline{u}_k^{\,t} \underline{\alpha})$, $k=1,2, \ldots, r$ or in matrix notation:

$$\underline{\beta} = \Lambda U^t \underline{\alpha}$$

where Λ is truncated to remove the zero eigenvalues.

[2] We shall only refer to columns, but comparable statements can be made for rows.

 © ISVR 1999

NULL SPACE OF A (SYMMETRIC) MATRIX

Consider a matrix that has some eigenvalues which are zero, so that it is singular (has no inverse) and rank deficient.

Then there exists some $\underline{x} \neq \underline{0}$ such that

$$A\underline{x} = \underline{0}$$

For example if $\underline{u}_k$ is an eigenvector associated zero eigenvalue, *i.e.* $\lambda_k = 0$, then from (1) $A\underline{u}_k = \lambda_k \underline{u}_k$ and since $\lambda_k = 0$, $A\underline{u}_k = \underline{0}$.

The set of vectors, $\underline{x}$, for which $A\underline{x} = \underline{0}$ is called the *null space* of A.

The dimension of the null space is L-r.

A basis for the null space is the set of eigenvectors associated with the zero eigenvalues, *i.e.* $\underline{u}_k$, $k = r+1, r+2, \ldots, L$.

So that if $\underline{x}$ is in the null space then:

$$\underline{x} = \gamma_1 \underline{u}_{r+1} + \gamma_2 \underline{u}_{r+2} + \cdots + \gamma_{L-r} \underline{u}_L = U_n \underline{\gamma}$$

where $U_n = \left[\underline{u}_{r+1}\, \underline{u}_{r+2} \cdots \underline{u}_L \right]$ and $\underline{\gamma}$ is a vector of the coefficients γ_k.

Note that $AU_n = 0$ since for each column of U_n, $A\underline{u}_k = \underline{0}$.

 © ISVR 1999

SOLUTION OF SYMMETRIC SYSTEMS OF LINEAR EQUATIONS

I: THE NON-SINGULAR CASE

The set of linear equations

$$a_{11}x_1 + a_{12}x_2 + \cdots a_{1L}x_L = b_1,\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2L}x_L = b_2,\ \ldots.$$

$$a_{L1}x_1 + a_{L2}x_2 + \cdots a_{LL}x_L = b_L$$

Can be expressed in the form

$$A\underline{x}=\underline{b}$$

Where the ith, j^{th} element of A is $a_{i,j}$ and the column vectors $\underline{x}$ and $\underline{b}$ contain the values x_j and b_j.

If an inverse to A exists, *i.e.* it is not singular (A is full rank), then

$$\underline{x} = A^{-1}\underline{b}$$

This represents the unique and exact solution of the original system of equations. Since

$$A^{-1} = \sum_{n=1}^{L} \frac{1}{\lambda_n}\underline{u}_n\underline{u}_n{}^t$$

then we can also write

$$\underline{x} = A^{-1}\underline{b} = \sum_{n=1}^{L} \frac{1}{\lambda_n}\underline{u}_n\left(\underline{u}_n{}^t\underline{b}\right) = \sum_{n=1}^{L} \frac{\underline{u}_n{}^t\underline{b}}{\lambda_n}\underline{u}_n$$

 © ISVR 1999

SOLUTION OF SYMMETRIC SYSTEMS OF LINEAR EQUATIONS
II: THE SINGULAR CASE

Consider a system of equations described by $A\underline{x}=\underline{b}$ such that A is singular.

If $\underline{b}$ is in the column space of A then the system of equation is said to be *consistent* and there are an infinite number of possible solutions $\underline{x}$ which satisfy $A\underline{x}=\underline{b}$.

If $\underline{b}$ is **not** in the column space of A then the system of equation is said to be *inconsistent* and there are no solutions $\underline{x}$ which satisfy $A\underline{x}=\underline{b}$.

Examples

i) A consistent set of undetermined equations-

$$x_1+x_2=2$$
$$2x_1+2x_2=4$$

all solutions of the form $x_1=c$, $x_2=2-c$ satisfy the equations.

ii) An inconsistent set of undetermined equations-

$$x_1+x_2=2$$
$$2x_1+2x_2=3$$

No values of x_1 and x_2 will exactly satisfy the equations.

 © ISVR 1999

SOLUTION OF SYMMETRIC SYSTEMS OF LINEAR EQUATIONS

II: THE SINGULAR CASE (CONT'D)

If the system of equations in consistent then, if $\underline{x}$ is one solution then any vector of the form

$$\underline{x}+U_n\underline{\gamma}$$

will also satisfy the $A\underline{x}=\underline{b}$. Recall that $U_n\underline{\gamma}$ describes any vector in the null space of A.

Proof: $A(\underline{x}+U_n\underline{\gamma})=A\underline{x}+AU_n\underline{\gamma}=A\underline{x}+\underline{0}=\underline{b}$

For an inconsistent system of equations then one can not find an exact solution, however, one can seek the 'best' solution, such that $A\underline{x}$ is as close as possible to $\underline{b}$, *e.g.* find $\underline{x}$ such that

$$\left\|\underline{e}\right\|_2 = \underline{e}^t\,\underline{e} = \sum e(k)^2 \quad \text{where} \quad \underline{e}=\underline{b}-A\underline{x}$$

is minimised.

This solution is given by:

$$\underline{x} = \left(\sum_{k=1}^{r} \frac{1}{\lambda_k} \underline{u}_k \underline{u}_k{}^t \right) \underline{b} = A^{\#}\underline{b}$$

$A^{\#}$ is called the pseudo inverse, below is a comparison of $A^{\#}$ and A^{-1}

$$A^{\#} = \sum_{k=1}^{r} \frac{1}{\lambda_k} \underline{u}_k \underline{u}_k{}^t \qquad\qquad A^{-1} = \sum_{k=1}^{L} \frac{1}{\lambda_k} \underline{u}_k \underline{u}_k{}^t$$

If A is full rank then $r=L$ and $A^{\#}=A^{-1}$.

 © ISVR 1999

SINGULAR VALUE DECOMPOSITION (SVD)

Discussions thus far have relied upon the orthonormal decomposition

$$A=U\Lambda U^t \quad \text{with} \quad U^tU=I$$

This is only valid for symmetric matrices.

SVD provides a generalisation of this result, valid for all matrices (including those that are rectangular or non-symmetric).

The general result is that:

$$A=USV^t, \quad \text{with} \quad V^tV=I \quad \text{and} \quad U^tU=I$$

If A has dimensions $(m\times n)$ then:

U has dimension $(m\times m)$

V has dimension $(n\times n)$

S has dimension $(m\times n)$, is diagonal with non-negative elements:

$\{S\}_{i,j}=0$ for $i\neq j$ and $\{S\}_{i,i}=s_i$, ordered such that $s_1\geq s_2\geq\ldots\geq s_K$

The elements on the diagonal of S (s_i) are the *singular values* of A.
The columns of U are called the *left singular vectors* of A.
The columns of V are called the *right singular vectors* of A.

These concepts are generalisations of eigenvalues and eigenvectors.

All the elements of U, S and V are real if A is real.

 © ISVR 1999

RELATIONSHIP BETWEEN SVD AND EIGEN DECOMPOSITIONS

The matrix A can be used to generate 2 positive definite symmetric matrices[3], M_1 and M_2:

$$M_1=A^tA \text{ and } M_2=AA^t$$

Assuming $A=USV^t$ and $A^t=VS^tU^t$ then:

$$M_1=A^tA=VS^tU^tUSV^t=V(S^tS)V^t$$

$$M_2=AA^t=USV^tVS^tU^t=U(SS^t)U^t$$

Thus one can identify:

- The right singular vectors (V) of A are the eigenvectors of M_1.
- The left singular vectors (U) of A are the eigenvectors of M_2.
- If $n>m$ then there are m singular values which are the square roots of eigenvalues of M_1. The eigenvalues of M_2 being the eigenvalues of M_1 appended with n-m zeros.
- If $m>n$ then there are n singular values which are the square roots of eigenvalues of M_2.

[3] Note that A^tA is positive definite, since the quadratic form $\underline{x}^tA^tA\underline{x}=\underline{y}^t\underline{y}=\Sigma y(k)^2\geq 0$, $\underline{y}=A\underline{x}$

 © ISVR 1999

RANK AND INVERSION (REVISITED)

In general the rank, r, of a matrix is defined as the number of non-zero singular values.

For a square matrix this is also the number of non-zero eigenvalues.

The rank of a matrix is such that:

$$r \leq \min\{m,n\}$$

This is because $\min\{m,n\}$ is the number of singular values that a matrix possesses.

If $r=\min\{m,n\}$ then the matrix is referred to as *full rank* (compare with definition for square matrices).

For A^{-1} to exist then A must be square.

If A is square, but not symmetric, then:

$$A^{-1} = VS^{-1}U^t = \sum_{k=1}^{L} \frac{1}{s_k} \underline{v}_k \underline{u}_k{}^t$$

 © ISVR 1999

PSEUDO-INVERSE

The general form for the pseudo-inverse of a matrix is:

$$A^{\#} = \sum_{k=1}^{r} \frac{1}{s_k} \underline{v}_k \underline{u}_k{}^t$$

This is a matrix of dimension ($n \times m$) {Recall A has dimension ($n \times m$)}.

The properties of $A^{\#}$ are such that

$$A^{\#}AA^{\#}=A \text{ and } AA^{\#}A=A^{\#}$$

Under certain conditions the pseudo inverse can be defined without recourse to SVD:

For rectangular matrices, such that $m>n$, $A^{\#} = \left(A^t A\right)^{-1} A^t$, assuming that $\left(A^t A\right)^{-1}$ exists.

For rectangular matrices, such that $n>m$, $A^{\#} = A^t \left(AA^t\right)^{-1}$, assuming that $\left(AA^t\right)^{-1}$ exists.

 © ISVR 1999

APPLICATION I: SINE WAVES

Consider the following sampled sine wave:

$$x(n)=\sin(2\pi f_0 n\Delta)$$

$$x(n\text{-}1)=\sin(2\pi f_0 n\Delta-2\pi f_0\Delta)$$

By adding weighted versions of these two signals one can generate any sine wave for frequency f_0, since:

$$x(n\text{-}1)=\cos(2\pi f_0\Delta)\sin(2\pi f_0 n\Delta) - \sin(2\pi f_0\Delta)\cos(2\pi f_0 n\Delta)$$

$$\Rightarrow \cos(2\pi f_0 n\Delta)=\cot(2\pi f_0\Delta)x(n) - x(n\text{-}1)/\sin(2\pi f_0\Delta)$$

and

$$A\sin(2\pi f_0 n\Delta+\phi)=B\sin(2\pi f_0\Delta)+C\cos(2\pi f_0\Delta)=D\,x(n)+E\,x(n\text{-}1)$$

where the constants B and C have to be chosen appropriately.

From this observation it follows that if one forms 2 column vectors $\underline{x}_n$ and $\underline{x}_{n\text{-}1}$:

$$\underline{x}_n=\left[x(n),x(n-1),\cdots,x(n-L)\right]^t$$

$$\underline{x}_{n-1}=\left[x(n-1),x(n-2),\cdots,x(n-L-1)\right]^t$$

where $x(n)=\sin(2\pi f_0 n\Delta)$ then:

$$\underline{x}_{n-p}=\alpha_1\underline{x}_n+\alpha_2\underline{x}_{n-1}$$

 © ISVR 1999

APPLICATION I: SINE WAVES (CONT'D)

Consider a time series $x(n)$, n=0,…,N-1 and a data matrix, X, is constructed as:

$$X = \begin{bmatrix} x(0) & x(1) & x(2) & \cdots & x(L-1) \\ x(1) & x(2) & x(3) & \cdots & x(L) \\ x(2) & x(3) & x(4) & \cdots & x(L+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(N-L+1) & x(N-L+2) & x(N-L+3) & \cdots & x(N) \end{bmatrix}$$

where L is a user defined parameter ($<N$).

If the time series consists of a single sine wave component then each column of X can be formed by a linear combination of two other columns, *i.e.* X has rank 2.

So that if SVD is conducted on X then there will only be 2 non-zero singular values.

If the time series is the sum of K sine waves then the rank of X is $2K$, i.e. the SVD will reveal $2K$ non-zero singular values.

 © ISVR 1999

APPLICATION I: SINE WAVES PLUS NOISE

If $x(n)$ consists of

$$x(n) = \sum_{k=1}^{K} A_k \sin(2\pi f_k n\Delta) + w(n)$$

i.e. is the sum of K sine waves plus white noise, of power σ^2.

Then there will be:

- L-2K singular values with mean $\sigma\sqrt{L}$, associated with the noise
- 2K larger singular values associated with the sinusoids

The effect of adding white noise is to increase all of the singular values by an average of $\sigma\sqrt{L}$.

The above result applies regardless of the frequencies of the sinusoids.

One can use such techniques to estimate the number of sinusoidal components in a time series.

 © ISVR 1999

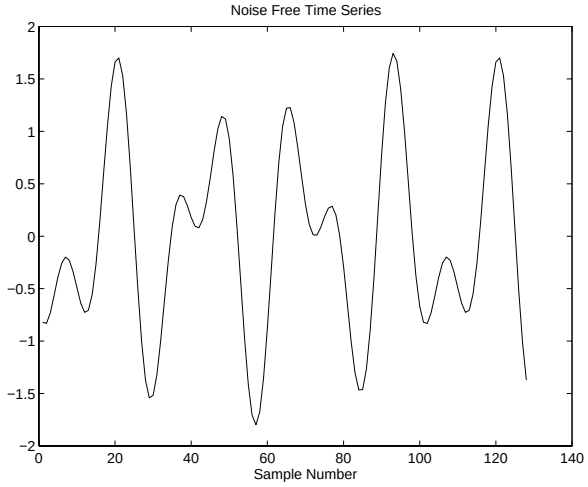
EXAMPLE

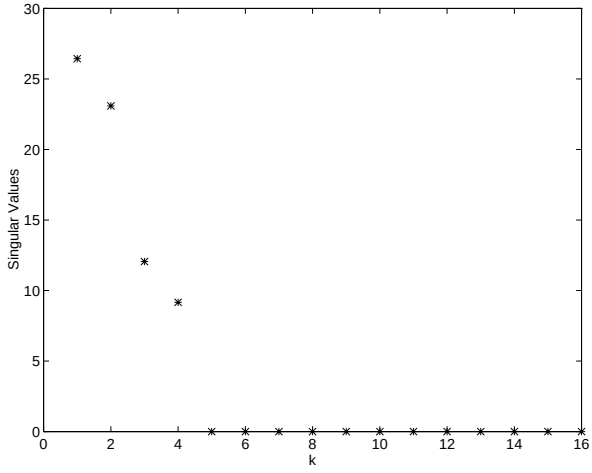
A time series is generated according to

$$x(n)=\sin(2\pi 4n\Delta+\phi_1)+\sin(2\pi 7n\Delta+\phi_2)+\sigma w(n)$$

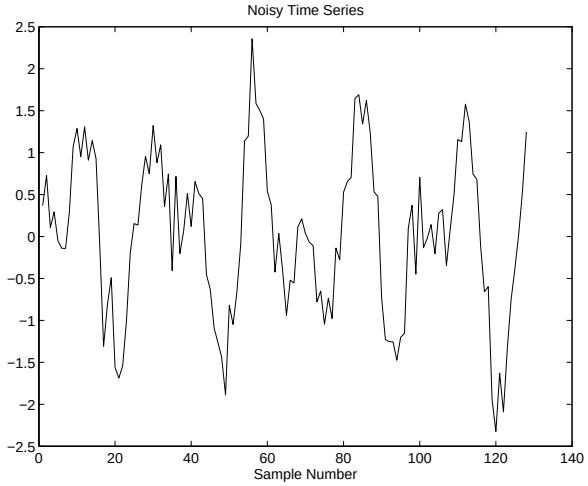
Δ is 0.01, ϕ_1 and ϕ_2 are random phases, $w(n)$ is white noise process with variance 1, so that $\sigma w(n)$ has variance σ^2.

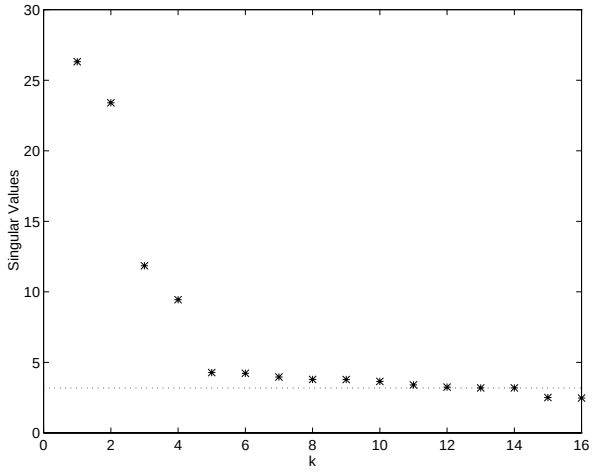
For $\sigma=0$ (no noise)





For $\sigma=0.3$





 © ISVR 1999

APPLICATION II: AUTO-REGRESSIVE MODELLING

One model used for time series data is the auto-regressive (AR) model

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \cdots + a_L y(n-L)$$

This states that the next sample in a sequence is formed by a linear combination of the passed L sample values.

Fitting this model to data entails trying to find a set of parameters $\underline{a}$ that best match the measured data.

Writing out a few terms one obtains

$$y(L) = a_1 y(L-1) + a_2 y(L-2) + \cdots + a_L y(0)$$

$$y(L+1) = a_1 y(L) + a_2 y(L-1) + \cdots + a_L y(1)$$

$$\vdots$$

$$y(N-1) = a_1 y(N-2) + a_2 y(N-3) + \cdots + a_L y(N-L-1)$$

This can be written in matrix form as:

$$\underline{y} = Y\underline{a}$$

where

$\underline{y} = [y(L), y(L+1), \ldots, y(N)]^t$

Y is a data matrix, similar in structure to X, whose k^{th} column is $[y(L-k), y(L-k+1), y(N-k-1)]^t$.

$\underline{a}$ is the vector of unknown coefficients $[a_1, a_2, \ldots, a_L]^t$

 © ISVR 1999

APPLICATION II: AUTO-REGRESSIVE MODELLING (CONT`D)

The system of equations $Y\underline{a}=\underline{y}$ is generally over determined. There are N-L-1 equations in L unknowns[4].

The value of $\underline{a}$ that most closely satisfies this equation is given by

$$\underline{a}=Y^{\#}\underline{y}$$

Since this is an over-determined system of equations then the solution is unlikely to be exact, *i.e.* $Y\underline{a}\neq\underline{y}$. The error in the model fitting one can express as:

$$\underline{e}=\underline{y}-Y\underline{a}$$

The use of a pseudo-inverse results in the minimisation of $\|\underline{e}\|^2$.

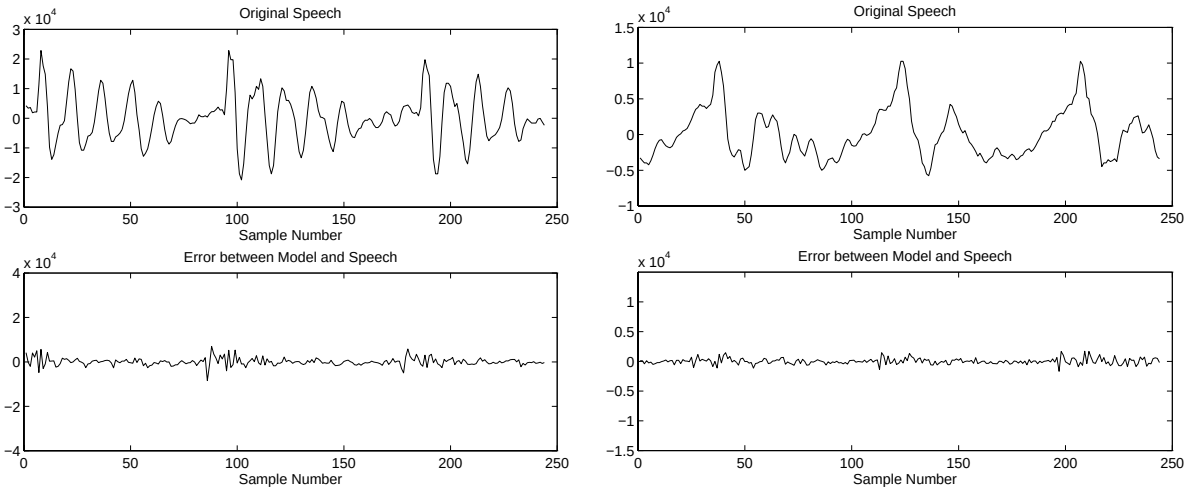
[4] Typical values of L for speech modelling are *8-16* and N might be *128 or 256*.

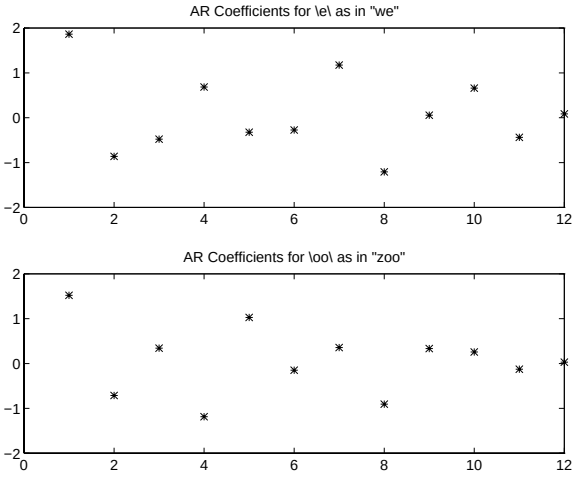
 © ISVR 1999

EXAMPLE

Consider 2 sections of (voiced) speech. Taken from the vowels /e/ in “we” and /oo/ in “zoo”.

The AR parameters are computed for segments of length 256 samples, 32 ms, with L=12.





 © ISVR 1999

APPLICATION III: PRINCIPLE COMPONENT ANALYSIS (PCA)

Consider a set of measurements, $\underline{x}_k$, k=1,2,3, …, K.

These measurements will be assumed to be related to each other in some sense. We shall consider 2 cases:

1) The signals, $\underline{x}_k$, are measurements made at different locations.

2) Signals, $\underline{x}_k$, correspond to features extracted from classification data.

PCA produces a representation of the data in which the energy is maximally concentrated. Further, in the new representation the data channels are orthogonormal.

So the vectors $\underline{x}_k$ are mapped to the vectors $\underline{u}_k$ such that

$$\begin{aligned} \underline{u}_i^t \underline{u}_j &= 0 \quad i \neq j \\ &= 1 \quad i = j \end{aligned}$$

This mapping is achieved using the SVD and the $\underline{u}_k$'s can be recognised as the left singular vectors.

The singular values, s_k, can be regarded as amplitudes of these terms.

The original data, $\underline{x}_q$, can be reconstructed from the $\underline{u}_k$ using

$$\underline{x}_q = \sum_{k=1}^{K} s_k \underline{u}_k v_q(k)$$

 © ISVR 1999

APPLICATION III: PRINCIPLE COMPONENT ANALYSIS (PCA) (CONT'D)

The singular vectors are ordered such that $s_1 \geq s_2 \geq \ldots \geq s_K$

If a few, say K_1, of the singular vectors are significantly larger than the others. Then the first K_1 singular vectors contain essentially the same information as the K vectors $\underline{x}_k$.

It can be shown that $\underline{u}_k$ $k=1,2,..,K_1$ are the optimal (in the mean square sense) orthonormal basis to approximate the data, i.e. they preserve the most energy of any basis.

Thus $\underline{u}_k$ represents a low(er) order approximation to the data.

Thus PCA can be used to reduce the number of features used in a classification problem.

The use of a feature set that has been reduced using PCA has the advantages of:
1) Being based on fewer features, means that less data is required for training.
2) The PCA features are defined to be uncorrelated, so making the modelling easy – implicitly assumes a common covariance matrix and Gaussian statistics.

 © ISVR 1999

CONCLUSIONS

Square symmetrical matrices are naturally decomposed using eigen based methods.

The concept of SVD extends this to non-square, non-symmetric, matrices.

One can use SVD to generalise the concept of an inverse.

The singular values of a suitably constructed data matrix can be used to infer the number of signal components present in a time series.

The problem of estimating AR parameters can be formulated using SVD.

 © ISVR 1999