MATRIX ALGEBRA REVIEW FOR STATISTICS

LESSON 4 NOTES

Rev. 2.1

by

Robert A. LaBudde, Ph.D.

XVIII. EIGENVALUES & EIGENVECTORS OF A REAL SYMMETRIC MATRIX:

The matrices for which eigensets are most useful are the class of symmetric matrices, particularly positive-definite symmetric matrices.

A square matrix **A** is "symmetric" if it is equal to its transpose:

$$\mathbf{A} = \mathbf{A}^t \tag{XVIII.1}$$

This means that

$$a_{i\,j} = a_{j\,i} \qquad \text{all } i, j \tag{XVIII.2}$$

An example of a symmetric 3 x 3 matrix is

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 4 \end{bmatrix} \tag{XVIII.3}$$

All diagonal matrices are necessarily also symmetric. Special efficient algorithms exist for finding the eigenset of a symmetric matrix.

Symmetric matrices have some special properties of their eigenset. In particular a real symmetric matrix **A** has all real eigenvalues. Because **A** is real, and its eigenvalues are real, then all eigenvectors can be specified in real numbers as well. Having an all real-valued eigenset is a great advantage in numerical algorithms and in supplying physical interpretations to modeling. In addition, if **U** is the matrix of eigenvectors of **A** where each column is a normalized eigenvector, then **U** is an "orthogonal" matrix, meaning that its transpose is its inverse. This implies

$$\mathbf{U}^t\,\mathbf{U} = \mathbf{U}\,\mathbf{U}^t = \mathbf{I} \tag{XVIII.4a}$$

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Lambda}\,\mathbf{U}^t \tag{XVIII.4b}$$

$$\mathbf{A}^k = \mathbf{U}\,\mathbf{\Lambda}^k\,\mathbf{U}^t \qquad k = 1,2,3,\ldots \tag{XVIII.4c}$$

Note that eq.(XVIII.4b) implies all real symmetric matrices are diagonalizable. Note also that each eigenvector (column of **U**) is "orthogonal" (i.e., has an inner product zero) to every other column of **U**. The rank of **A** is given by the number of its non-zero eigenvalues.

For the example matrix **A** in eq.(XVIII.3), and using R, the eigenset of **A** is:

```
> A<- matrix(c(3,1,0, 1,2,1, 0,1,4), ncol=3, byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    3    1    0
[2,]    1    2    1
[3,]    0    1    4

> e<- eigen(A, symmetric=TRUE)
> e #show results
$values
[1] 4.532089 3.347296 1.120615

$vectors
          [,1]       [,2]       [,3]
[1,] 0.2931284  0.8440296 -0.4490988
[2,] 0.4490988  0.2931284  0.8440296
[3,] 0.8440296 -0.4490988 -0.2931284

> U<- e$vectors   #get eigenvector matrix
> t(U) %*% U      #show U is orthogonal
              [,1]          [,2]          [,3]
[1,]  1.000000e+00 -9.183192e-17 -8.038004e-17
[2,] -9.183192e-17  1.000000e+00 -6.837250e-17
[3,] -8.038004e-17 -6.837250e-17  1.000000e+00

> t(U) %*% A %*% U    #show inv(U) A U diagonalizes A
              [,1]          [,2]          [,3]
[1,]  4.532089e+00 -2.235625e-16 -7.220786e-17
[2,] -5.746272e-16  3.347296e+00 -1.710627e-15
[3,] -1.246561e-16 -1.703241e-15  1.120615e+00
```

Note that: 1) all eigenvalues are real; 2) all eigenvector elements are real; 3) the eigenvector matrix is orthogonal; and 4) $\mathbf{U}^t\,\mathbf{A}\,\mathbf{U}$ diagonalizes **A** and puts its eigenvalues on the diagonal.

A symmetric matrix **A** is called “positive definite” (denoted “p.d.”) if and only if *all* of its eigenvalues are greater than zero. The example matrix in eq.(XVIII.3) is positive definite.

It is called “positive semi-definite” (denoted “p.s.d.”) if *at least one* eigenvalue is zero and all others are greater than or equal to zero.

It is called “non-negative definite” (denoted “n.n.d.”) if and only if *all* of its eigenvalues are greater than or equal to zero. N.n.d. includes both p.d. and p.s.d. matrices.

Positive definite matrices have some special properties concerning their eigenset.

A positive definite symmetric matrix A must have (necessary conditions):

$a_{i\,i}$	> 0	for all i	(diagonal positive) (XVIII.5a)
$a_{i\,i} + a_{j\,j} > 2 \mid a_{i\,j} \mid$		for $i \neq j$	(diagonally dominant) (XVIII.5b)
The element with largest absolute value lies on the main diagonal			(XVIII.5c)
$\mid \mathbf{A} \mid$	> 0		(determinant positive) (XVIII.5d)

XIX. SPECTRAL DECOMPOSITION OF A REAL SYMMETRIC MATRIX:

Consider any two conformable matrices **A** and **B**. Let $\mathbf{a}_1 \ldots \mathbf{a}_m$ be the column vectors of **A** and $\mathbf{b}_1 \ldots \mathbf{b}_m$ the row vectors of **B**. Then the product **A B** may be written in terms of outer products as

$$\mathbf{A}\,\mathbf{B} = \Sigma\, \mathbf{a}_j \times \mathbf{b}_j \qquad \text{(XIX.1)}$$

For example, suppose

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \text{(XIX.2)}$$

Then

$$\mathbf{A}\,\mathbf{B} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 2 & 2 \end{bmatrix} \qquad \text{(XIX.3)}$$

and

$$\mathbf{A}\,\mathbf{B} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \times [2\ 1] + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \times [0\ 1] + \begin{bmatrix} 2 \\ 0 \end{bmatrix} \times [1\ 0] \qquad \text{(XIX.4a)}$$

$$= \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 2 & 2 \end{bmatrix} \qquad \text{(XIX.5b)}$$

Now suppose **A** is a real symmetric n x n matrix and $\mathbf{u}_1 \ldots \mathbf{u}_n$ are its eigenvectors, which are column vectors in the combined eigenvector matrix **U**.

Because **U** is orthogonal,

$$\mathbf{U}^t\,\mathbf{U} = \mathbf{I} = \mathbf{U}\,\mathbf{U}^t \qquad \text{(XIX.6)}$$

Using eq.(XIX.1),

$$\mathbf{U}\,\mathbf{U}^t = \mathbf{I} = \Sigma\, \mathbf{u}_j \times \mathbf{u}_j^t \qquad \text{(XIX.7)}$$

Multiplying both sides of eq.(XIX.7) on the left by **A**,

$$\mathbf{A} = \Sigma\, \mathbf{A}\,\mathbf{u}_j \times \mathbf{u}_j^t = \Sigma\, \lambda_j\, \mathbf{u}_j \times \mathbf{u}_j^t \qquad \text{(XIX.8)}$$

Eq.(XIX.8) is called the “spectral decomposition” of **A**. It also implies that

$$\mathbf{A}^k = \Sigma \lambda_j^k \mathbf{u}_j \text{ x } \mathbf{u}_j^t \qquad \text{(XIX.9)}$$

The useful property shown in eq.(XIX.8) can be written out as

$$\mathbf{A} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^t + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^t + \ldots + \lambda_n \mathbf{u}_n \mathbf{u}_n^t \qquad \text{(XIX.10)}$$

Note that each $\mathbf{u}_i \mathbf{u}_i^t$ outer product is an n x n matrix, and we've suppressed the "x" notation for outer product.

This "spectral decomposition" of **A** allows us to represent **A** as the linear combination of individual (orthogonal) eigenvector matrices, each with a coefficient equal to the corresponding eigenvalue.

Usually we start with a symmetric matrix **A**, find its eigenset, and then construct the right-hand side of eq.(XIX.10). But the reverse direction also works.

Suppose we wish to find a symmetric matrix **A** that has eigenvalues $\lambda_1 = 2$, $\lambda_2 = 1$ and $\lambda_3 = 0$, with corresponding orthonormal set of eigenvectors we choose to be

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} / \sqrt{2} \qquad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{3} \qquad \mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} / \sqrt{6} \qquad \text{(XIX.11)}$$

We can find such an **A** easily by using eq.(XIX.10). First we find the eigenvector matrices by

$$\mathbf{E}_1 = \mathbf{u}_1 \mathbf{u}_1^t = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} [1 \;\; -1 \;\; 0] / 2 = \begin{bmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(XIX.12a)}$$

$$\mathbf{E}_2 = \mathbf{u}_2 \mathbf{u}_2^t = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \;\; 1 \;\; 1] / 3 = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \qquad \text{(XIX.12b)}$$

$$\mathbf{E}_3 = \mathbf{u}_3 \mathbf{u}_3^t = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} [1 \;\; 1 \;\; -2] / 6 = \begin{bmatrix} 1/6 & 1/6 & -1/3 \\ 1/6 & 1/6 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \qquad \text{(XIX.12c)}$$

Then, using eq.(XIX.10),

$$\mathbf{A} = 2 \mathbf{E}_1 + 1 \mathbf{E}_2 + 0 \mathbf{E}_3 \qquad \text{(XIX.13a)}$$

$$= \begin{bmatrix} 4/3 & -2/3 & 1/3 \\ -2/3 & 4/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \qquad \text{(XIX.13b)}$$

Using R, we now find the eigenset of the **A** in eq.(XIX.13b):

```
> A<- matrix(c(4/3,-2/3,1/3, -2/3,4/3,1/3, 1/3,1/3,1/3), ncol=3, byrow=TRUE)
> A
          [,1]       [,2]      [,3]
[1,]  1.3333333 -0.6666667 0.3333333
[2,] -0.6666667  1.3333333 0.3333333
[3,]  0.3333333  0.3333333 0.3333333

> eigs<- eigen(A)   #find eigenset
> eigs
$values
[1] 2.000000e+00 1.000000e+00 8.881784e-16

$vectors
             [,1]      [,2]       [,3]
[1,]  7.071068e-01 0.5773503 -0.4082483
[2,] -7.071068e-01 0.5773503 -0.4082483
[3,]  5.637851e-18 0.5773503  0.8164966

> eigVals<- eigs$values #strip off eigenvalues
> eigVals
[1] 2.000000e+00 1.000000e+00 8.881784e-16

> eigVects<- eigs$vectors
> eigVects
             [,1]      [,2]       [,3]
[1,]  7.071068e-01 0.5773503 -0.4082483
[2,] -7.071068e-01 0.5773503 -0.4082483
[3,]  5.637851e-18 0.5773503  0.8164966
```

Note that the eigenvalues are, in fact, { 2, 1, 0} as desired, and the eigenvectors are those given in eq.(2), except that $\mathbf{e}_3$ has the opposite sign, which is irrelevant and an artifact of the numerical algorithm used.

We can show that A is equal to its spectral decomposition by using R and the equivalent of eq.(XIX.10):

```
> B<- eigVals[1]*eigVects[,1]%*%t(eigVects[,1]) +
eigVals[2]*eigVects[,2]%*%t(eigVects[,2]) +
+       eigVals[3]*eigVects[,3]%*%t(eigVects[,3])

> B  #reconstruction from the spectral decomposition
          [,1]       [,2]      [,3]
[1,]  1.3333333 -0.6666667 0.3333333
[2,] -0.6666667  1.3333333 0.3333333
[3,]  0.3333333  0.3333333 0.3333333
```

```
> A - B   #show equal (differences all zero)
             [,1]          [,2]          [,3]
[1,] 2.220446e-16  2.220446e-16  1.110223e-16
[2,] 2.220446e-16  1.110223e-15 -2.220446e-16
[3,] 1.110223e-16 -2.220446e-16 -1.776357e-15

> #alternative formula (outer product without transpose)
> B<- eigVals[1]*eigVects[,1]%o%eigVects[,1] +
eigVals[2]*eigVects[,2]%o%eigVects[,2] +
+       eigVals[3]*eigVects[,3]%o%eigVects[,3]

> B    #reconstruction of A from spectral decomposition
           [,1]       [,2]      [,3]
[1,]  1.3333333 -0.6666667 0.3333333
[2,] -0.6666667  1.3333333 0.3333333
[3,]  0.3333333  0.3333333 0.3333333

> A - B   #show equal (differences all zero)
             [,1]          [,2]          [,3]
[1,] 2.220446e-16  2.220446e-16  1.110223e-16
[2,] 2.220446e-16  1.110223e-15 -2.220446e-16
[3,] 1.110223e-16 -2.220446e-16 -1.776357e-15

> #show individual eigenvector matrices
> E1<- eigVects[,1]%*%t(eigVects[,1])   #outer product of 1st eigenvector
> E1
              [,1]          [,2]          [,3]
[1,]  5.000000e-01 -5.000000e-01  3.986563e-18
[2,] -5.000000e-01  5.000000e-01 -3.986563e-18
[3,]  3.986563e-18 -3.986563e-18  3.178537e-35

> E2<- eigVects[,2]%*%t(eigVects[,2])   #outer product of 2nd eigenvector
> E2
          [,1]      [,2]      [,3]
[1,] 0.3333333 0.3333333 0.3333333
[2,] 0.3333333 0.3333333 0.3333333
[3,] 0.3333333 0.3333333 0.3333333

> E3<- eigVects[,3]%*%t(eigVects[,3])   #outer product of 3rd eigenvector
> E3
           [,1]       [,2]       [,3]
[1,]  0.1666667  0.1666667 -0.3333333
[2,]  0.1666667  0.1666667 -0.3333333
[3,] -0.3333333 -0.3333333  0.6666667

> B<- eigVals[1]*E1 + eigVals[2]*E2 + eigVals[3]*E3   #spectral decomposition
of A

> B
           [,1]       [,2]      [,3]
[1,]  1.3333333 -0.6666667 0.3333333
[2,] -0.6666667  1.3333333 0.3333333
[3,]  0.3333333  0.3333333 0.3333333

> A-B   #show equal (differences all zero)
             [,1]          [,2]          [,3]
[1,] 2.220446e-16  2.220446e-16  1.110223e-16
[2,] 2.220446e-16  1.110223e-15 -2.220446e-16
[3,] 1.110223e-16 -2.220446e-16 -1.776357e-15
```

XX. FURTHER COMMENTS ON THE SPECTRAL DECOMPOSITION OF A REAL SYMMETRIC MATRIX

An n x n real symmetric matrix **A** with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ and orthonormal eigenvectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ has the useful property that

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^t + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^t + ... + \lambda_n \mathbf{e}_n \mathbf{e}_n^t \qquad \text{(XX.1)}$$

(Note that each $\mathbf{e}_i \mathbf{e}_i^t$ is an n x n matrix.)

Eq.(XX.1) is called the "spectral decomposition" of **A**, and allows us to represent **A** as the linear combination of individual (orthogonal) eigenvector matrices, each with a coefficient equal to the corresponding eigenvalue.

Usually we start with a symmetric matrix **A**, find its eigenset, and then construct the right-hand side of eq.(XX.1). But the reverse direction also works.

Suppose we wish to find a symmetric matrix **A** that has eigenvalues $\lambda_1 = 2$, $\lambda_2 = 1$ and $\lambda_3 = 0$, with corresponding orthonormal set of eigenvectors we choose to be

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} / \sqrt{2} \qquad \mathbf{e}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} / \sqrt{3} \qquad \mathbf{e}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} / \sqrt{6} \qquad \text{(XX.2)}$$

We can find such an **A** easily by using eq.(XX.1). First we find the eigenvector matrices by

$$\mathbf{E}_1 = \mathbf{e}_1 \mathbf{e}_1^t = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} [1 \ \ -1 \ \ 0] / 2 = \begin{bmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 0.5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{(XX.3a)}$$

$$\mathbf{E}_2 = \mathbf{e}_2 \mathbf{e}_2^t = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \ \ 1 \ \ 1] / 3 = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \qquad \text{(XX.3b)}$$

$$\mathbf{E}_3 = \mathbf{e}_3 \mathbf{e}_3^t = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} [1 \ \ 1 \ \ -2] / 6 = \begin{bmatrix} 1/6 & 1/6 & -1/3 \\ 1/6 & 1/6 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \qquad \text{(XX.3c)}$$

Then, using eq.(XX.1),

$$\mathbf{A} = 2\,\mathbf{E}_1 + 1\,\mathbf{E}_2 + 0\,\mathbf{E}_3 \qquad \text{(XX.4a)}$$

XXI. STATISTICS APPLICATION: PRINCIPAL COMPONENTS ANALYSIS

Eigenvalues and eigenvectors are computed routinely in multivariate statistics, particularly in the methods of principal components analysis, factor analysis and structural equation modeling.

The matrices most often involved in eigenset analysis are the covariance matrix of a collection of variables thought to have a joint multivariate normal distribution, or upon the corresponding correlation matrix.

Principal component analysis is typically used to reduce a large set of covariates into a smaller set of covariates that are linear combinations of the original variables and which have almost the same explanatory power. The method involves keeping the eigenvectors associated with the largest eigenvalues and dropping those associated with the smallest eigenvalues. The covariance matrix is used when the variables are commensurate in location and scale, and the correlation matrix otherwise.

Consider the violent crime data example shown in Section VII. Using R, we read the data and compute the correlation matrix **R**:

```
> crime = read.csv('violent-crimes.csv', header=TRUE)
> crime
             Murder Rape Robbery AggrAssault
Dallas         15.0 53.3   553.9       584.2
Houston        18.2 41.2   548.3       561.7
Las Vegas      11.6 54.6   409.0       507.7
Los Angeles    12.4 27.3   370.0       377.2
New York        7.3 13.1   287.9       329.6
Philadelphia   27.7 65.5   749.1       720.1
Phoenix        15.4 36.2   287.5       398.5
San Antonio     9.2 39.8   179.6       388.7
San Diego       5.4 27.5   170.8       300.8
San Jose        3.2 23.6   111.9       248.2

> R<-cor(crime)     #correlation matrix
> R
               Murder      Rape   Robbery AggrAssault
Murder      1.0000000 0.7649509 0.9172615   0.9190375
Rape        0.7649509 1.0000000 0.7559736   0.8920793
Robbery     0.9172615 0.7559736 1.0000000   0.9586910
AggrAssault 0.9190375 0.8920793 0.9586910   1.0000000
```

Generally all correlations are large, particularly so for Murder, Robbery and Assault. The high mutual correlations indicate a factor in common that may be responsible for most of the variation. This factor, and possibly others, is what we hope to utilize to condense the covariate set. If all correlations were very small, we could not hope to gain much from principal components analysis.

Computing the eigenset of the correlation matrix using R,

```
> e<- eigen(R, symmetric=TRUE)  #eigenset of correlation matrix
> e
$values
[1] 3.608667702 0.289955922 0.091804787 0.009571589

$vectors
           [,1]        [,2]        [,3]        [,4]
[1,] -0.4998382 -0.38689744  0.77401604 -0.03703121
[2,] -0.4715407  0.82205065  0.09177377 -0.30571176
[3,] -0.5045167 -0.41666508 -0.55846953 -0.50986756
[4,] -0.5227562  0.03055013 -0.28388119  0.80324596
```

We note that the first (largest) eigenvalue is 3.6 and much larger than the next nearest value of 0.29. A rule of thumb in principal components analysis is to keep eigenvalues greater than 1.0 for a correlation matrix, and to drop those less than 1.0. Applying this rule here, this means we keep only the first eigenvalue and eigenvector and drop the others as insignificant. Here 3.61 / 4 = 90% of the variation in the scaled variables can be accounted for by using only the first eigenvector. A typical goal is to include at least 80-85% of the variation in the retained set.

The elements of the eigenvector are used to interpret the meaning of the factor. Here, all four elements are roughly equal and of the same sign, suggesting the factor is an average across all types of crime. We might term this factor “Violent Crime”. The second eigenvector has a large coefficient for Rape, and opposite sign coefficients for Murder and Robbery. This factor might correspond to “Sex Motive”.

EXAMPLE: STOCK PRICE INDICES

File “djia-nasdaq-sp500.csv” contains weekly closing prices of the Dow Jones Industrial Average (“DJIA”), NASDAQ Composite (“NASDAQ”) and the Standard and Poor 500 (“SP500”) stock price indices from January 2000 to August 2009.

The study question of interest is: Are these three indices independent, or could just as good predictive results be obtained from a single composite index? The method of solution is principal components analysis (“PCA”).

Using R:

```
> d<- read.csv('djia-nasdaq-sp500.csv', header=TRUE)
> head(d)
       Date    DJIA  Nasdaq   SP500
1 8/24/2009 9544.20 2028.77 1028.93
2 8/17/2009 9505.96 2020.90 1026.13
3 8/10/2009 9321.40 1985.52 1004.09
4  8/3/2009 9370.07 2000.25 1010.48
5 7/27/2009 9171.61 1978.50  987.48
6 7/20/2009 9093.24 1965.96  979.26

> r<- cor(d[,2:4])  #get correlation matrices of indices
```

```
> r
            DJIA    Nasdaq     SP500
DJIA   1.0000000 0.5308666 0.9145645
Nasdaq 0.5308666 1.0000000 0.7926296
SP500  0.9145645 0.7926296 1.0000000
```

The correlation matrix indicates that the DJIA and SP500 indices are highly correlated, suggesting both are measuring the same factor. The NASDAQ is only moderately correlated with the other two, more strongly with the SP500.

```
> eigen(r)
$values
[1] 2.50287574 0.47776224 0.01936201

$vectors
          [,1]        [,2]       [,3]
[1,] 0.5693583  0.61885758  0.5411529
[2,] 0.5317696 -0.77924738  0.3316544
[3,] 0.6269388  0.09893848 -0.7727606
```

The eigenvalues of the correlation matrix suggest one primary factor accounting for 83% (2.503/3.0) of the variation over the 9 year period. The first two eigenvalues account for 99.4% of the variation, with the third eigenvalue close to zero.

The first principal component appears to be a nearly equally weighted average of the three indices, suggesting a simple combined or composite effect.

The second principal component is a difference between the DJIA and the NASDAQ, perhaps a “Non-technology Index”.

Interestingly, PCA has shown that the three commonly reported indices are heavily correlated, and, to a large extent, could be replaced by a single new index

$$z_F = 0.569\, z_{DJIA} + 0.532\, z_{NASDAQ} + 0.629\, z_{SP500}$$

where “z” denotes the scaled value obtained by dividing the difference from the mean by the standard deviation of the variable.

It is well known that stock prices tend to be lognormally distributed, so we repeat the analysis on the ln transform of the data:

```
> lnd<- log(d[,2:4])  #log transform data
> head(lnd)
      DJIA   Nasdaq    SP500
1 9.163689 7.615185 6.936275
2 9.159674 7.611298 6.933550
3 9.140068 7.593636 6.911837
4 9.145276 7.601027 6.918181
5 9.123868 7.590094 6.895156
6 9.115287 7.583736 6.886797

> lnr<- cor(lnd)
> lnr
            DJIA    Nasdaq     SP500
```

```
DJIA   1.0000000 0.6768004 0.9313534
Nasdaq 0.6768004 1.0000000 0.8666828
SP500  0.9313534 0.8666828 1.0000000

> eigen(lnr)
$values
[1] 2.65440126 0.32718342 0.01841532

$vectors
          [,1]        [,2]       [,3]
[1,] 0.5689697  0.64447506  0.5108086
[2,] 0.5520611 -0.75971670  0.3435973
[3,] 0.6095097  0.08650109 -0.7880453
```

The results are very similar, and an even larger (88%) fraction of the variance is accounted for by the first principal component alone.

APPENDIX: SIMULATION OF INDEPENDENT NORMALLY DISTRIBUTED PCA

Principal components analysis ("PCA") presumes that the largest eigenvalues will do a good job of representing the effect of the correlation or covariance matrix. A typical rule of thumb for separating the large eigenvalues from the small ones is to keep all eigenvalues of the correlation matrix which are greater than 1.0, although there are suggestions that 0.8 or even 0.6 might be a better cutoff point. (For the covariance matrix, a similar rule would be $\Sigma \lambda_j / m$, where m is the number of variables involved.)

If we are trying to interpret the principal components, we can question whether or not the pattern found in PCA is real, due to associations among the variables, or whether the principal components found are an artifact of sampling (random error).

To give an illustration of the effect of random error on PCA, a simulation was carried out in which 100 data each were sampled from 5 independent and identically distributed normal distributions. (Assuming identical normal distributions does not restrict generality if we perform PCA on the correlation matrix, which scales out differences in means and variance anyway.)

Here are the results of 20 such trials using the R language:

```
> nReal<- 20
> for (iReal in 1:nReal) {
+   x<- matrix(rnorm(500,0,1), ncol=5) #generate 100 data each for 5 N(0,1)
variables
+   r<- cor(x)  #find correlation matrix
+   cat(iReal,' Eigenvalues:',eigen(r)$values,'\n')
+ }
1  Eigenvalues: 1.282292 1.125415 0.986286 0.8961944 0.7098124
2  Eigenvalues: 1.251738 1.042236 0.9561595 0.9047293 0.8451374
3  Eigenvalues: 1.336952 1.105069 0.9628346 0.836717 0.7584274
4  Eigenvalues: 1.425468 1.275160 0.9075213 0.7454728 0.6463781
5  Eigenvalues: 1.274658 1.053772 1.036961 0.8420212 0.7925878
6  Eigenvalues: 1.257969 1.180749 1.030211 0.8583897 0.6726815
7  Eigenvalues: 1.210724 1.091834 0.9972152 0.9336695 0.766557
8  Eigenvalues: 1.303008 1.129526 0.9705615 0.8681005 0.7288032
9  Eigenvalues: 1.245101 1.203262 0.978541 0.877914 0.6951824
10  Eigenvalues: 1.329935 1.20158 1.032604 0.7855297 0.6503505
11  Eigenvalues: 1.267294 1.074416 1.016560 0.9112212 0.7305078
12  Eigenvalues: 1.362344 1.050021 1.003362 0.847511 0.7367624
13  Eigenvalues: 1.240339 1.105359 1.003854 0.9326096 0.7178384
14  Eigenvalues: 1.202207 1.082255 0.987631 0.9331407 0.7947661
15  Eigenvalues: 1.245698 1.154347 0.9915986 0.8871854 0.7211705
16  Eigenvalues: 1.279396 1.076280 0.9993965 0.9545645 0.6903631
17  Eigenvalues: 1.173535 1.138088 1.012821 0.9296874 0.7458679
18  Eigenvalues: 1.399528 1.189058 0.9213491 0.8885342 0.6015303
19  Eigenvalues: 1.243628 1.096169 0.985683 0.8963932 0.7781269
20  Eigenvalues: 1.204153 1.076639 1.009506 0.8866016 0.8231007
```

Note that the largest eigenvalue ranges from 1.2 to 1.4, and the smallest from 0.6 to 0.8. This is not a very large difference, so this should raise our comfort level with respect to

PCA. The results also indicate why 0.6 might be a better cutoff level than 1.0, at least for m = 5.

You should also note that the ratio of the largest to smallest eigenvalues found will increase with m, so the results here are specific to m = 5.

XXII. GEOMETRIC INTERPRETATION OF VECTORS AND MATRICES

Consider a point P in an n-dimensional space with an orthogonal coordinate system. (Such a space is called a "Euclidean" space, as it is similar to the familiar x,y,z physical system.) We denote the coordinates of P by

$$P = (x_1, x_2, ..., x_n) \tag{XXII.1}$$

The origin of the coordinate system is O,

$$O = (0, 0, ..., 0) \tag{XXII.2}$$

Corresponding to a point P is a vector **p** which is the directed line segment from O to P, with elements equal to the difference in coordinates of O and P, or

$$\mathbf{p} = \begin{bmatrix} x1 \\ x2 \\ ... \\ xn \end{bmatrix} \tag{XXII.3}$$

Now consider an n x n invertible matrix **A** and the product **A p**. This product can be viewed as a transformation of the coordinates of **p** to new coordinates denoted by the vector **q**, where the elements of **q** are the new difference in coordinates of the line segment from O to P. The point O remains the origin of the new coordinate system. So **q** is given by

$$\mathbf{q} = \mathbf{A}\,\mathbf{p} = \begin{bmatrix} x1' \\ x2' \\ ... \\ xn' \end{bmatrix} \tag{XXII.4}$$

Without further restrictions on **A**, it may correspond to stretching or contracting of distances in the new coordinates, or it may result in axes that are not orthogonal (perpendicular).

To maintain orthogonality, we must require that **A** be an "orthogonal" matrix, which preserves the Euclidean distance metric we prize. I.e., the distance from O to any physical point P must remain the same whether we express coordinates in the old (x_1, x_2, ..., x_n) system or in the new (x_1', x_2', ..., x_n') system. This means that

$$x_1^2 + x_2^2 + ... + x_n^2 = x_1'^2 + x_2'^2 + ... + x_n'^2 \tag{XXII.5}$$

for any vectors **p** and **q** related by eq.(XXII.4).

So if **A** is an orthogonal matrix, it defines an orthogonal (distance preserving) transformation of the coordinate system.

There are only a few types of orthogonal transformations possible. One is to simply rotate the coordinate axes about an axis through O, which does not stretch or contract the axes and doesn't alter their perpendicular nature. Another is to reflect the axes with respect to a plane through O. This introduces factors of –1 which don't have an effect in eq.(XXII.5). Another orthogonal transformation is inversion with respect to O. All of these don't affect eq.(XXII.5).

The eigenvalues of an orthogonal matrix can only be +1 or –1, so the determinant is also +1 or –1. The eigenvalues of a real orthogonal matrix are not necessarily real, and typically are complex.

In 1-dimension, the only orthogonal matrices are **I** = [1] and **A** = [-1]. **A** corresponds to inversion with respect to the origin O.

In 2-dimensions, the general rotation matrix is given by

$$\mathbf{R}(t) = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \qquad \text{(XXII.6)}$$

where t is an angle of rotation counterclockwise about an axis through O and perpendicular to the 2-dimensions. Note that $|\mathbf{R}(t)| = 1$ for any t.

In 3-dimensions, there are 3 generic rotations possible about the 3 axes, giving rise to

$$\mathbf{R}_x(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(t) & -\sin(t) \\ 0 & \sin(t) & \cos(t) \end{bmatrix} \qquad \text{(XXII.7a)}$$

$$\mathbf{R}_y(t) = \begin{bmatrix} \cos(t) & 0 & \sin(t) \\ 0 & 1 & 0 \\ \sin(t) & 0 & \cos(t) \end{bmatrix} \qquad \text{(XXII.7b)}$$

$$\mathbf{R}_z(t) = \begin{bmatrix} \cos(t) & -\sin(t) & 0 \\ \sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{(XXII.7c)}$$

Other rotations can be expressed as products of these 3 types. Such rotation matrices are commonly used in physics and aeronautics to represent change in coordinates from space-fixed to body-fixed axes for rigid objects, such as molecules, aircraft and missiles.

XXIII. STATISTICAL APPLICATION OF VECTOR GEOMETRY:

Suppose N subjects are chosen at random (independent), and p response variables are measured on each subject. Let the p measurements on subject k be denoted the column vector $\mathbf{x}_k$. We compute the sample mean $\mathbf{m}$ and sample covariance matrix $\mathbf{S}$ by

$$\mathbf{m} = (\mathbf{x}_1 + \mathbf{x}_2 + ... + \mathbf{x}_N) / N \qquad \text{(XXIII.1a)}$$

$$\mathbf{S} = \{ (\mathbf{x}_1 - \mathbf{m}) (\mathbf{x}_1 - \mathbf{m})^t + ... (\mathbf{x}_N - \mathbf{m}) (\mathbf{x}_N - \mathbf{m})^t \} / (N - 1) \qquad \text{(XXIII.1b)}$$

If the p response variables jointly following a p-dimensional multivariate normal distribution, then a 100(1-α)% confidence region for the true population mean $\boldsymbol{\mu}$ is given by the hyperellipsoid determined by all $\boldsymbol{\mu}$ such that

$$(\mathbf{m} - \boldsymbol{\mu})^t \mathbf{S}^{-1} (\mathbf{m} - \boldsymbol{\mu}) \leq \omega^2 \qquad \text{(XXIII.2)}$$

where

$$\omega^2 = p (N - 1) F_{p, N-p} (1-\alpha) / [N (N - p)] \qquad \text{(XXIII.3a)}$$

$$\approx \chi_p^2(1 - \alpha) \quad \text{for N – p large} \qquad \text{(XXIII.3a)}$$

and $F_{p, N-p} (1-\alpha)$ is the (1- α)-th quantile of the F-distribution for p and N-p degrees of freedom, and $\chi_p^2(1 - \alpha)$ is the (1- α)-th quantile of the chi-square distribution for p degrees of freedom. The quantity ω is called the "Mahalonobis distance" or "statistical distance", as it is weighted by the inverse of the covariance matrix.

The "quadratic form" eq.(XXIII.2) defines a hyperellipsoid in p-dimensions whose axes of symmetry are the eigenvectors of $\mathbf{S}$. For p = 2, the region is the interior of an ellipse. For p = 3, the region is the interior of an ellipsoid.

A "positive-definite" quadratic form has the RHS of eq.(XXIII.2) intrinsically positive, and

$$(\mathbf{m} - \boldsymbol{\mu})^t \mathbf{S}^{-1} (\mathbf{m} - \boldsymbol{\mu}) = 0 \quad \text{if and only if } \mathbf{m} = \boldsymbol{\mu} \qquad \text{(XXIII.4)}$$