Statistics 5041
7. Eigen structure

Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

No fooling this time . . . , we're almost done with the matrix stuff.
Let $\mathbf{X}$ be a $p \times p$ matrix. Suppose that vector u is such that

$$\mathbf{X}u = \lambda u$$

for some scalar λ. Then u is an *eigenvector* of $\mathbf{X}$ and λ is its associated *eigenvalue*.

```
Cmd> X <- matrix(vector(2,1,1,2),2);X
(1,1)            2            1
(2,1)            1            2

Cmd> X %*% vector(1,1)
(1,1)            3
(2,1)            3

Cmd> X %*% vector(1,-1)
(1,1)            1
(2,1)           -1
```

Eigenvalue 3 with eigenvector $(1, 1)'$; eigenvalue 1 with eigenvector $(1, -1)'$.
Note that

$$\mathbf{X}u = \lambda u = \lambda \mathbf{I}_p u$$

so that

$$(\mathbf{X} - \lambda \mathbf{I}_p)$$

is a singular matrix and $|\mathbf{X} - \lambda \mathbf{I}_p| = 0$. The eigenvalues of $\mathbf{X}$ are the values of λ that make $\mathbf{X} - \lambda \mathbf{I}_p$ singular.

```
Cmd> X - 3*I;det(X - 3*I)
(1,1)           -1            1
(2,1)            1           -1
WARNING: argument to det() is singular
(1)            0

Cmd> X - I;det(X - I)
(1,1)            1            1
(2,1)            1            1
WARNING: argument to det() is singular
(1)            0
```

```
Cmd> eigenvals(X)
(1)            3            1

Cmd> eigen(X)
component: values
(1)            3            1
component: vectors
(1,1)     -0.70711      0.70711
(2,1)     -0.70711     -0.70711

Cmd> vector(1,1)/sqrt(2)
(1)      0.70711      0.70711

Cmd> vector(1,-2)/sqrt(2)
(1)      0.70711      -1.4142
```

The eigenvalues of a diagonal matrix are the diagonal elements of the matrix. Euclidean basis vectors (forming $\mathbf{I}_p$) are the eigenvectors.

All eigenvalues are nonzero if a matrix is nonsingular. (Otherwise $\mathbf{X}u = 0u = 0$ for some nonzero u, which means that there is a (nonzero) linear combination of columns that is zero, which means that $\mathbf{X}$ is not full rank.)

The trace of a square matrix equals the sum of its eigenvalues

Note that if $\mathbf{X}u = \lambda u$, then $\mathbf{X}(cu) = c\mathbf{X}u = c\lambda u = \lambda(cu)$ for any scalar c.

By convention, take $||u|| = 1$, but direction (multiplication by 1 or –1) is arbitrary. A *normalized* eigenvector.

By convention, arrange so that eigenvalues are in decreasing order.

Special rules for *symmetric* $\mathbf{X}$. When $\mathbf{X}$ is symmetric:

There are always p *real* (as opposed to complex) eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

Exactly p *linearly independent* eigenvectors $u_1, u_2, \ldots, u_p$ with *real* (as opposed to complex) elements.

Eigenvectors corresponding to *distinct* eigenvalues are *orthogonal*.

It is possible to choose eigenvalues to be mutually orthogonal.

An *orthogonal* matrix is a $p \times p$ matrix $\mathbf{U}$ with orthogonal column of norm 1.

$$\mathbf{U}'\mathbf{U} = \mathbf{I}_p = \mathbf{U}\mathbf{U}'$$

For orthogonal $\mathbf{U}$ and vector y

$$||y|| = (y'y)^{.5} = (y'\mathbf{U}\mathbf{U}'y)^{.5} = ||\mathbf{U}y||$$

Similarly, $||y|| = ||\mathbf{U}'y||$.

Let $\mathbf{X}$ be a symmetric matrix, let Λ be a diagonal matrix with the eigenvalues of $\mathbf{X}$ on the diagonal, and let $\mathbf{U}$ be a matrix with columns containing orthogonal eigenvectors of $\mathbf{X}$ ($\mathbf{U}$ is not unique). Then

$$\begin{aligned}
\mathbf{X}\mathbf{U} &= \mathbf{U}\Lambda \\
\mathbf{X}\mathbf{U}\mathbf{U}' &= \mathbf{U}\Lambda\mathbf{U}' \\
\mathbf{X} &= \mathbf{U}\Lambda\mathbf{U}' \\
\mathbf{X} &= \lambda_1\check{\mathbf{U}}_1\check{\mathbf{U}}_1' + \lambda_2\check{\mathbf{U}}_2\check{\mathbf{U}}_2' + \ldots + \lambda_p\check{\mathbf{U}}_p\check{\mathbf{U}}_p'
\end{aligned}$$

This is the *spectral decomposition* of $\mathbf{X}$.

Let $\Lambda^{.5}$ be a diagonal matrix with $\sqrt{\lambda_i}$ down the diagonal; let $\Lambda^{-.5}$ be a diagonal matrix with $1/\sqrt{\lambda_i}$ down the diagonal; and let Λ^{-1} be a diagonal matrix with $1/\lambda_i$ down the diagonal, the ordinary inverse of Λ. (The negative power matrices require nonzero eigenvalues.)
Let

$$\begin{aligned}\mathbf{A} &= \mathbf{U}\Lambda^{.5}\mathbf{U}' \\ \mathbf{B} &= \mathbf{U}\Lambda^{-1}\mathbf{U}' \\ \mathbf{C} &= \mathbf{U}\Lambda^{-.5}\mathbf{U}'\end{aligned}$$

$\mathbf{X}$, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are all symmetric.

$$\begin{aligned}\mathbf{AA} &= (\mathbf{U}\Lambda^{.5}\mathbf{U}')(\mathbf{U}\Lambda^{.5}\mathbf{U}') = \mathbf{U}\Lambda\mathbf{U}' = \mathbf{X} \\ \mathbf{BX} &= (\mathbf{U}\Lambda^{-1}\mathbf{U}')(\mathbf{U}\Lambda\mathbf{U}') = \mathbf{UU}' = \mathbf{I}_p \\ \mathbf{XB} &= (\mathbf{U}\Lambda\mathbf{U}')(\mathbf{U}\Lambda^{-1}\mathbf{U}') = \mathbf{UU}' = \mathbf{I}_p \\ \mathbf{CXC} &= (\mathbf{U}\Lambda^{-.5}\mathbf{U}')(\mathbf{U}\Lambda\mathbf{U}')(\mathbf{U}\Lambda^{-.5}\mathbf{U}') = \mathbf{UU}' = \mathbf{I}_p\end{aligned}$$

$\mathbf{A}$ is a square root of $\mathbf{X}$; $\mathbf{B}$ is the inverse of $\mathbf{X}$; $\mathbf{C}$ is an inverse square root of $\mathbf{X}$.
The eigenvalues of the inverse are the reciprocals of the eigenvalues.
Consider a $p \times p$ symmetric matrix $\mathbf{H}$ and a p-vector x. The expression

$$x'\mathbf{H}x$$

is called a *quadratic form*. When you expand it out, you get terms in x_i^2 and $x_i x_j$, all second order polynomial terms.
Let $q = x'\mathbf{H}x$.
If $q > 0$ for all x, $\mathbf{H}$ is called *positive definite*.
If $q \geq 0$ for all x, $\mathbf{H}$ is called *positive semi-definite* or *nonnegative definite*.
If $q > 0$ for some values of x and $q < 0$ for other values of x, $\mathbf{H}$ is called *indefinite*.
Let $\mathbf{U}$ be an orthogonal matrix consisting of the eigenvectors of $\mathbf{H}$, and let x be some p-vector. Then

$$x = \mathbf{U}v$$

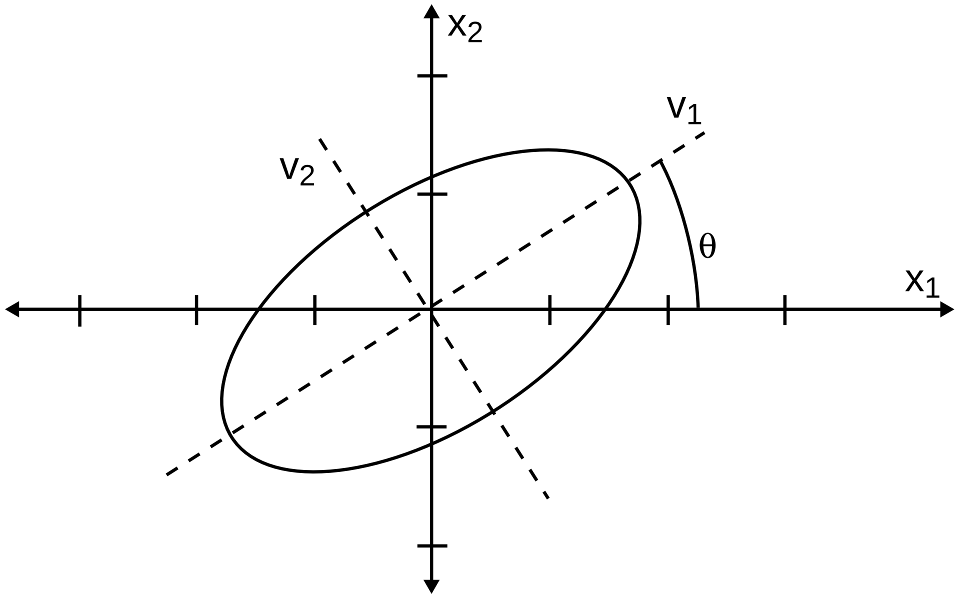
where the coefficients v are the coordinates of x in the new basis formed by $\mathbf{U}$. Note that $v = \mathbf{U}'x$.
Recall that $\mathbf{H} = \mathbf{U}\Lambda\mathbf{U}'$. Then

$$\begin{aligned}x'\mathbf{H}x &= (\mathbf{U}v)'\ \mathbf{U}\Lambda\mathbf{U}'\ (\mathbf{U}v) \\ &= v'\mathbf{U}'\mathbf{U}\Lambda\mathbf{U}'\mathbf{U}v \\ &= v'\Lambda v \\ &= \lambda_1 v_1^2 + \lambda_2 v_2^2 + \ldots + \lambda_p v_p^2\end{aligned}$$

Eigenvalues of $\mathbf{H}$ determine positive definite, semi-definite, or indefinite.
We've seen quadratic forms before.

$$D = x'\mathbf{H}x$$

was the locus of points at constant Mahalanobis distance from the origin and formed an ellipse.



Recall that Mahalanobis distance (squared) is

$$D = x'\mathbf{S}^{-1}x$$

Let $\mathbf{U}\Omega\mathbf{U}'$ be the spectral decomposition of $\mathbf{S}$. $\sqrt{\omega_i}$ is the standard deviation along the ith rotated axis. Let $\mathbf{U}\Lambda\mathbf{U}' = \mathbf{U}\Omega^{-1}\mathbf{U}'$ be the spectral decomposition of $\mathbf{H} = \mathbf{S}^{-1}$. Then

$$D = x'\mathbf{S}^{-1}x = x'\mathbf{H}x = v'\Omega^{-1}v = v'\Lambda v$$

In this picture, the new axes are the eigenvectors of $\mathbf{H}$. The coordinates of a point on the new axes are just $v = \mathbf{U}'x$.

The eigenvalues of $\mathbf{H}$ are the reciprocal squares of the elongation along the new axes. That is, make the identification

$$\lambda_i v_i^2 = \frac{v_i^2}{\omega_i} = \frac{v_i^2}{s_i^2}$$

where s_i is length of the ellipse in the ith new axis direction (corresponds to standard deviation of the point cloud in along the ith new axis).

Maximization for $p \times p$ positive definite $\mathbf{B}$.

$$\overset{\max}{x \neq 0} \frac{x'\mathbf{B}x}{x'x} = \lambda_1$$

$$\overset{\min}{x \neq 0} \frac{x'\mathbf{B}x}{x'x} = \lambda_p$$

The first is achieved with the first eigenvector, the second is achieved with the last eigenvector.

Cauchy-Schwarz inequality: x and y any two p-vectors;

$$(x'y)^2 \leq (x'x)(y'y)$$

Maximum achieved if $x = cy$ for some scalar c.

Extended Cauchy-Scharz inquality: x and y any two p-vectors; $\mathbf{B}$ $p \times p$ positive definite.

$$(x'y)^2 \leq (x'\mathbf{B}x)(y'\mathbf{B}^{-1}y)$$

Maximum achieved if $x = c\mathbf{B}^{-1}y$ for some scalar c.
Maximization lemma: x and y any two p-vectors; $\mathbf{B}$ $p \times p$ positive definite.

$$\overset{\max}{x \neq 0} \frac{x'y}{x'\mathbf{B}x} = y'\mathbf{B}y$$

Maximum is achieved when $x = c\mathbf{B}^{-1}y$ for some scalar c.
One other important case. Let $\mathbf{A} = \mathbf{BC}$ where $\mathbf{B}$ and $\mathbf{C}$ are symmetric. Then the eigenvectors and eigenvalues of $\mathbf{A}$ are real.

```
Cmd> B
(1,1)         18.45       -0.8878        1.8833
(2,1)       -0.8878        9.1963      -0.88071
(3,1)        1.8833      -0.88071        3.3216

Cmd> C
(1,1)        6.2696       -4.5205       -1.4487
(2,1)       -4.5205        13.702        1.2594
(3,1)       -1.4487        1.2594        11.073

Cmd> A <- B%*%C; A
(1,1)        116.96       -93.195       -6.9941
(2,1)       -45.862        128.91        3.1162
(3,1)        10.977       -16.398        32.941

Cmd> v <- rnorm(3)

Cmd> for(i,run(200)) v <- A%*%v
v<-v/sum(v^2)^.5;;
```

Power method gets eigenvector for largest eigenvalue.

```
Cmd> v
(1,1)      -0.78615
(2,1)       0.60628
(3,1)      -0.11995

Cmd> (A %*% v)/v
(1,1)        187.76
(2,1)        187.76
(3,1)        187.76
```

```
Cmd> out<-releigen(C,solve(B));out
component: values
(1)       187.76       57.253       33.795
component: vectors
(1,1)       2.9966       -3.072      0.18176
(2,1)      -2.3109      -1.9634     0.029876
(3,1)      0.45721    -0.062747       1.7631

Cmd> A%*%out$vectors/out$values'
(1,1)       2.9966       -3.072      0.18176
(2,1)      -2.3109      -1.9634     0.029876
(3,1)      0.45721    -0.062747       1.7631
```