Matrix Algebra, part 2

Ming-Ching Luoh

2005.9.12

Characteristic Roots and Vectors

Also called Eigenvalues and Eigenvectors.
The solution to a set of equations

$$Ac = \lambda c$$

where λ is the eigenvalue and c is the eigenvector of a square matrix A.
If c is a solution, then kc is also a solution for any k.
Therefore, it is usually normalized so that $c'c = 1$.

To solve for the solution, the above system implies $Ac = \lambda I_K c$. Therefore,

$$(A - \lambda I)c = 0$$

This has non-zero solution if and only if the matrix $(A - \lambda I)$ is singular or has a zero determinant, or $|A - \lambda I| = 0$. The polynomial in λ is the characteristic equation of A.

For example, if $A = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$, then

$$\begin{aligned} |A - \lambda I| &= \begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} \\ &= (5-\lambda)(5-\lambda) - 2 \\ &= \lambda^2 - 9\lambda + 18 = 0 \\ \lambda &= 6, 3. \end{aligned}$$

For the eigenvectors,

$$\begin{bmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\lambda = 6, \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \pm \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\lambda = 3, \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \pm \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

The general results of eigenvalues and eigenvectors of a symmetric matrix A.

- A $K \times K$ symmetric matrix has K distinct eigenvectors, $c_1, c_2, \cdots, c_K$.
- The corresponding eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_K$ are real, but need not to be distinctive.
- Eigenvectors of a symmetric matrix are orthogonal, i.e., $c_i' c_j = 0, \ i \neq j$.

We can collect the K eigenvectors in the matrix

$$C = [c_1, c_2, \cdots, c_K],$$

and the K eigenvalues in the same order in a diagonal matrix,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}$$

From $Ac_i = \lambda_i c_i$, we have $AC = C\Lambda$. Since eigenvectors are orthogonal and $c_i'c = 1$,

$$C'C = \begin{bmatrix} c_1'c_1 & c_1'c_2 & \cdots & c_1'c_K \\ c_2'c_1 & c_2'c_2 & \cdots & c_2'c_K \\ \vdots & \vdots & \vdots & \vdots \\ c_K'c_1 & c_K'c_2 & \cdots & c_K'c_K \end{bmatrix} = I$$

This implies $C' = C^{-1}$ and $CC' = CC^{-1} = I$.

Diagonalization and Spectral Decomposition

Since $AC = C\Lambda$, pre-multiply both sides by C', we have

$$C'AC = C'C\Lambda = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}$$

On the other hand, if we post-multiply both sides by $C^{-1} = C'$, we have

$$\begin{aligned} A &= C\Lambda C' = [c_1, \cdots, c_K] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix} \begin{bmatrix} c_1' \\ \vdots \\ c_K' \end{bmatrix} \\ &= \sum_{i=1}^{K} \lambda_i c_i c_i' \end{aligned}$$

This is called the **spectral decomposition** of matrix A.

Rank of a Matrix

Rank of a Product. For any matrix A and nonsingular matrices B and C, then rank of BAC is equal to the rank of A.

pf: It is known that if A is $n \times K$ and B is a square matrix of rank K, then rank(AB)=rank(A), and rank(A)=rank(A').

Since C is nonsingular, rank(BAC)=rank((BA)C)=rank(BA). Also we have rank(BA)=rank((BA)$'$)=rank($A'B'$). It follows that rank($A'B'$)=rank(A') since B' is nonsingular if B is nonsingular.

Therefore, rank(BAC)=rank(A')=rank(A).

Using the result of diagonalization that $C'AC = \Lambda$ and C' and C are nonsingular ($|\lambda| = |C||A||C'| \neq 0$), we have,

$$\text{rank}(C'AC) = \text{rank}(A) = \text{rank}(\Lambda)$$

This leads to the following theorem.

Theorem

Rank of a Symmetric Matrix. *The rank of a symmetric matrix is the number of nonzero eigenvalues it contains.*

It appears that the above Theorem applies to symmetric matrix only. However, note the fact that rank(A)=rank($A'A$), $A'A$ will always be a symmetric square matrix. Therefore,

Theorem

Rank of a Matrix. *The rank of any matrix A is the number of nonzero eigenvalues in $A'A$.*

Trace of a Matrix

The trace of a square $K \times K$ matrix is the sum of its diagonal elements, $\text{tr}(A) = \sum_{i=1}^{K} a_{ii}$. Some results about trace are as follows.

- $\text{tr}(cA) = c\text{tr}(A)$, c is a scalar.
- $\text{tr}(A') = \text{tr}(A)$.
- $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$.
- $\text{tr}(I_K) = K$

- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

proof:
Let $A : n \times K$, $B : K \times n$, $C = AB$ and $D = BA$.
Since $c_{ii} = a^i b_i = \sum_{k=1}^{K} a_{ik} b_{ki}$, and
$d_{kk} = b^k a_k = \sum_{i=1}^{n} b_{ki} a_{ik}$,

$$\begin{aligned} \mathrm{tr}(AB) &= \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{K} a_{ik} b_{ki} \\ &= \sum_{k=1}^{K} \sum_{i=1}^{n} b_{ki} a_{ik} = \sum_{k=1}^{K} d_{kk} = \mathrm{tr}(BA) \end{aligned}$$

This permutation rule can be applied to any cyclic permutation in a product:

$$\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA) = \mathrm{tr}(CDAB) = \mathrm{tr}(DABC).$$

Along the diagonalization rule of matrix A, we can derive a rule for the trace of a matrix.

$$\begin{aligned} \mathrm{tr}(C'AC) &= \mathrm{tr}(ACC') = \mathrm{tr}(AI) \\ &= \mathrm{tr}(A) = \mathrm{tr}(\Lambda) = \sum_{k=1}^{K} \lambda_k \end{aligned}$$

Theorem

Trace of a Matrix. *The trace of a matrix equals the sum of its eigenvalues.*

Determinant of a Matrix

One particular simple way of calculating the determinant of a matrix is as follows. Since $C'AC = \Lambda$, $|C'AC| = |\Lambda|$.

$$\begin{aligned} |C'AC| &= |C'| \cdot |A| \cdot |C| = |C'| \cdot |C| \cdot |A| = |C'C| \cdot |A| \\ &= |I| \cdot |A| = |A| = |\Lambda| = \prod_{k=1}^{K} \lambda_k \end{aligned}$$

Theorem

Determinant of a Matrix. *The determinant of a matrix equals the product of its eigenvalues.*

Powers of a Matrix

With the result that $A = C\Lambda C'$, we can show the following theorem.

Theorem

Eigenvalues of a Matrix Power. *For any nonsingular symmetric matrix* $A = C\Lambda C'$, $A^K = C\Lambda^K C'$, $K = \cdots, -2, -1, 0, 1, 2, \cdots$.

In other word, eigenvalues of A^K is λ_i^K while eigenvectors are the same. This is for all intergers.

For real number, we have

Theorem

Real Powers of a Positive Definite Matrix. *For a positive definite matrix* $A = C\Lambda C'$, $A^r = C\Lambda^r C'$, *for any real number,* r.

A positive definite matrix is a matrix with all eigenvalues being positive. A matrix is nonnegative definite if all the eigenvalues are either positive or zero.

Quadratic Forms and Definite Matrices

Many optimization problems involve the quadratic form

$$q = x'Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$$

where A is a symmetric matrix, x is a column vector. For a given A,

- If $x'Ax > (<)0$ for all nonzero x, then A is positive (negative) definite.
- If $x'Ax \geq (\leq)0$ for all nonzero x, then A is nonnegative definite or positive semi-definite (nonpositive definite).

We can use results we have so far to check a matrix for definiteness. Since $A = C\Lambda C'$,

$$x'Ax = x'C\Lambda C'x = y'\Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2$$

where $y = C'x$ is another vector.

Theorem

Definite Matrices. *Let A be a symmetric matrix. If all the eigenvalues of A are positive (negative), then A is positive (negative) definite. If some of them are zeros, then A is nonnegative (nonpositve) if the remainder are positive (negative).*

Nonnegative definite of Matrices

- If A is nonnegative definite, then $|A| \geq 0$.
  Since determinant is the product of eigenvalues, it is nonnegative.
- If A is positive definite, so is A^{-1}.
  The eigenvalues of A^{-1} are the reciprocals of eigenvalues of A which are all positive. Therefore A^{-1} is also positive definite.
- The identity matrix I is positive definite.
  $x'Ix = x'x > 0$ if $x \neq 0$.

- (Very important) If A is $n \times K$ with full rank and $n > K$, then $A'A$ is positive definite and AA' is nonnegative definite.
  By assumption that A is with full rank, $Ax \neq 0$. So

$$x'A'Ax = y'y = \sum_i y_i^2 > 0.$$

For the latter case, since A have more rows than columns, there is an x such that $A'x = 0$. We only have $y'y \geq 0$.

- If A is positive definite amd B is nonsingular, then $B'AB$ is positive definite.
  $x'B'ABx = y'Ay > 0$, where $y = Bx$. But y can not be zero because B is nonsingular.
- For A to be negative definite, all A's eigenvalues must be negative. But te determinant of A is positive if A is in even order and negative if it is in odd order.

Calculus and Matrix Algebra

We have some definitions to begin with.

$$y = f(x), \; x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where y is a scalar and x is a column vector.

Then the first derivative or **gradient** is defined as

$$\underset{n\times 1}{\frac{\partial f}{\partial x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix}$$

And the second derivatives matrix or **Hessian** is,

$$\underset{n \times n}{H} = \frac{\partial^2 f}{\partial x \partial x'} = \begin{bmatrix} \frac{\partial^2 y}{\partial x_1 \partial x_1} & \frac{\partial^2 y}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_1 \partial x_n} \\ \frac{\partial^2 y}{\partial x_2 \partial x_1} & \frac{\partial^2 y}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 y}{\partial x_n \partial x_1} & \frac{\partial^2 y}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 y}{\partial x_n \partial x_n} \end{bmatrix}$$

In the case of a linear function $y = a'x = x'a = \sum_{i=1}^{n} a_i x_i$, where a and x are both column vectors,

$$\underset{n\times 1}{\frac{\partial y}{\partial x}} = \frac{\partial(a'x)}{\partial x} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a.$$

In a set of linear functions,

$$\underset{n\times 1}{y} = \underset{n\times n}{A}\ \underset{n\times 1}{x}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix} = [a_1, a_2, \cdots, a_n]$$

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j = a^i x$$

The derivatives are

$$\underset{n\times 1}{\frac{\partial y_i}{\partial x}} = a^{i\prime}, \underset{1\times n}{\frac{\partial y_i}{\partial x'}} = a^i$$

$$\frac{\partial (AX)}{\partial x'} = \frac{\partial y}{\partial x'} = \begin{bmatrix} \frac{\partial y_1}{\partial x'} \\ \frac{\partial y_2}{\partial x'} \\ \vdots \\ \frac{\partial y_n}{\partial x'} \end{bmatrix} = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^n \end{bmatrix} = A$$

$$\frac{\partial (AX)}{\partial x} = A'$$

A quadratic form is written as

$$
\begin{aligned}
y &= x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij} \\
\underset{n\times 1}{\frac{\partial(x'Ax)}{\partial x}} &= \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} x_j a_{1j} + \sum_{i=1}^{n} x_i a_{i1} \\ \sum_{j=1}^{n} x_j a_{2j} + \sum_{i=1}^{n} x_i a_{i2} \\ \vdots \\ \sum_{j=1}^{n} x_j a_{nj} + \sum_{i=1}^{n} x_i a_{in} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&+ \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= Ax + A'x = (A + A')x \\
&= 2Ax \text{ if } A \text{ is symmetric} .
\end{aligned}
$$

The Hessian matrix is

$$
\underset{n \times n}{\frac{\partial^2 (x'Ax)}{\partial x \partial x'}} = 2A
$$

Optimization

In a single variable case, $y = f(x)$. The first order (necessary) condition for an optimum is $\frac{dy}{dx} = 0$. And the second order (sufficient) condition is $\frac{d^2y}{d^2x} < 0$ for a maximum and $\frac{d^2y}{d^2x} > 0$ for a mimimum.

For the optimization of a function with several variable $y = f(x)$ where x is a vector. The first order condition is

$$\frac{\partial f(x)}{\partial x} = 0$$

And the second order condition is that the Hessian matrix

$$H = \frac{\partial^2 f(x)}{\partial x \partial x'}$$

must be positive definite for a minimum and negative definite for a maximum.

One example, we want to find x to maximize $R = a'x - x'Ax$, where x and a are $n \times 1$, A is symmetric $n \times n$. The first order condition is

$$\underset{n \times 1}{\frac{\partial R}{\partial x}} = a - 2Ax = 0, x = -\frac{1}{2}A^{-1}a.$$

And the Hessian Matrix is

$$\underset{n \times n}{\frac{\partial^2 R}{\partial x \partial x'}} = -2A.$$

If A is in quadratic form and is positive definite, then $-2A$ is negative definite. This ensures a maximum.

An Example: OLS Regression

$$\underset{1\times 1}{y_i} = \underset{1\times K}{x_i} \underset{K\times 1}{\beta} + \underset{1\times 1}{u_i} \quad i = 1, 2, \cdots, n$$

$$\underset{n\times 1}{y} = \underset{n\times K}{x} \underset{K\times 1}{\beta} + \underset{n\times 1}{u} .$$

The OLS estimator mininizes the sum of squared residuals $\sum_{i=1}^{n} u_i^2 = u'u$. In other words,

$$\begin{aligned} \min_{\beta} S(\beta) &= (y - x\beta)'(y - x\beta) \\ &= y'y - \beta'x'y - y'x\beta + \beta'x'x\beta \end{aligned}$$

The first order condition is

$$\underset{K\times 1}{\frac{\partial S(\beta)}{\partial \beta}} = -2x'y + 2x'x\beta = 0$$

$$\hat{\beta}_{OLS} = (x'x)^{-1}x'y = (x'x)^{-1}x'(x\beta + u),$$
$$= \beta + (x'x)^{-1}x'u$$

And the Hessian matrix is

$$\underset{K\times K}{\frac{\partial^2 S(\beta)}{\partial \beta \beta'}} = 2x'x$$

The Hessian matrix is positive definite. This ensures a minimum.