Stat200C

STAT200C: Review of Linear Algebra

Instructor: Zhaoxia Yu

1 Review of Linear Algebra

1.1 Vector Spaces, Rank, Trace, and Linear Equations

1.1.1 Rank and Vector Spaces

Definition A vector whose elements are stacked vertically is known as a column vector, or simply a vector. Its transpose is called a row vector.

Example. $\mu = (\mu_1, \cdots, \mu_p)^T = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_p \end{pmatrix}$.

Definition Vectors $x_1, \cdots, x_k$, where $x_i \in R^n, i = 1, \cdots, n$, are called linearly dependent if there exist numbers $\lambda_1, \cdots, \lambda_k$, not all zero, such that

$$\lambda_1 x_1 + \cdots + \lambda_k x_k = 0$$

In other words, they are linearly dependent if at least one of them can be written as a linear combination of the others.

Definition A set of vectors are called linearly independent if they are not linearly dependent.

We can also use the following definition:

Definition Vectors $x_1, \cdots, x_k$, where $x_i \in R^n, i = 1, \cdots, n$, are said to be linearly independent if $\lambda_1 x_1 + \cdots + \lambda_k x_k = 0$ implies $\lambda_1 = \cdots = \lambda_k = 0$. If $x_1, \cdots, x_k$ are not linearly independent, we say they are linearly dependent.

Example. $x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$\lambda_1 x_1 + \lambda_2 x_2 = 0$ implies that $\begin{pmatrix} \lambda_1 + \lambda_2 \\ \lambda_1 \end{pmatrix} = 0 \Rightarrow \lambda_1 = \lambda_2 = 0$.

Definition We say two vectors $x, y \in R^n$ are orthogonal if $x^T y = \sum_{i=1}^n x_i y_i = 0$.

Example: $x_1 = (1,0)^T$ and $x_2 = (0,1)^T$.

Note: Orthogonality implies linearly independence, but linearly independence does not guarantee orthogonality.

Proof: (orthogonality implies linearly independence) Suppose $x_1, \cdots, x_k$ are orthogonal and linearly dependent. Because they are linearly dependent, $\exists \lambda_1, \cdots, \lambda_k$, such that $\sum \lambda_i x_i = 0$. Then we have

$$\begin{aligned}
\lambda_1 x_1^T x_1 + \cdots + \lambda_k x_k^T x_1 &= \lambda_1 x_1^T x_1 = 0 \\
\lambda_1 x_1^T x_2 + \cdots + \lambda_k x_k^T x_2 &= \lambda_2 x_2^T x_2 = 0 \\
\cdots \quad \cdots & \quad \cdots \\
\lambda_1 x_1^T x_k + \cdots + \lambda_k x_k^T x_k &= \lambda_k x_k^T x_k = 0
\end{aligned}$$

which implies that $\lambda_i = 0$ for $i = 1, \cdots, k$, contradicting with the definition of linearly dependent.

The other part can be proved by providing an counterexample. Let $x_1 = (1,1,0)^T$ and $x_2 = (0,1,1)^T$. Clearly they are linearly independent but not orthogonal. Suppose they are linearly dependent, then there exist a_1, a_2 such that $a_1 = 0, a_1 + a_2 = 0, a_2 = 0$, which implies that $a_1 = a_2 = 0$, contradicting with the definition of linearly dependent.

Definition We say a square matrix $Q = (q_1, \cdots, q_n)$ is an orthogonal matrix if $Q^T Q = QQ^T = I$.

Note: all columns of Q are orthogonal unit vectors and all rows of Q are orthogonal unit vectors.

Example,

$$Q = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$$

$$Q = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Definition The rank of a matrix $A(n \times p)$ is defined as the maximum number of linearly independent rows (columns) in A. Note that row ranks and column ranks equal to each other, which is not obvious. It can be shown that the number of independent rows always equal to the number of independent columns.

Consider a matrix $A(n \times p)$. Some properties about ranks:

1. $0 \le rank(A) \le minp(n, p)$
2. $rank(A) = rank(A^T)$

3. $rank(A+B) \le rank(A) + rank(B)$
4. $rank(AB) \le min(rank(A), rank(B))$
5. $rank(A^TA) = rank(AA^T) = rank(A)$
6. For any matrix A, if B and C are conformable and non-singular then $rank(BAC) = rank(A) = rank(BA) = rank(AC)$

Proof of (4): Consider $A_{n\times p}$ and $B_{p\times m}$.

- Let
$$B = \begin{pmatrix} b_1^T \\ \cdots \\ b_p^T \end{pmatrix}$$
Then
$$AB = \begin{pmatrix} \sum a_{1i}b_i^T \\ \cdots \\ \sum a_{ni}b_i^T \end{pmatrix}$$
Thus, the rows of AB are linear combinations of the rows of B and the number of linearly independent rows of AB is less than or equal to those of B, which is less than or equal to that of B.

  For example, suppose $n = 3, p = 2$. We can find c_1, c_2 such that
$$\begin{aligned} c_1 * (a_{11} + a_{12}) &= a_{31} \\ c_2 * (a_{21} + a_{22}) &= a_{32} \end{aligned}$$
As a result, we can express $a_{31}b_1^T + a_{32}b_2^T$ using a linear combination of $a_{11}b_1^T + a_{12}b_2^T$ and $a_{21}b_1^T + a_{22}b_2^T$.

- Let
$$A = (a_1, \cdots, a_p)$$
We have
$$AB = (\sum a_j b_{j1}, \cdots, \sum a_j b_{jm})$$
Thus, the columns of AB are linear combinations of the columns of A and the number of linearly independent columns of AB is less than or equal to that of A, which is less than or equal to the rank of A.

The above implies that $rank(AB) \le min(rank(A), rank(B))$.

Proof of (5): see Appendix A.2. of the text book.

Proof of (6): (show the proof in class) $rank(A) \ge rank(AC) \ge rank(ACC^{-1}) = rank(A)$, indicating $rank(A) = rank(AC)$. $rank(AC) \ge rank(BAC) \ge rank(B^{-1}BAC) = rank(AC)$, indicating $rank(AC) = rank(BAC)$. Thus $rank(A) = rank(BA) = rank(AC) = rank(BAC)$.

Note: if $Ax = Bx$ for all x, then $A = B$. (Proof: let $x = (1, 0, ..., 0)^T$, then the first column of A is identical to the first column of B. etc.)

1.2 Vector Space

Definition A set of vectors which are closed under addition and scalar multiplication is called a vector space.

Definition The column space of a matrix A is the vector space generated by the columns of A. In other words, the column space of A is the set of all possible linear combinations of its column vectors. Let $A_{n\times p} = (a_1, \cdots, a_p)$, then the column space of A, denoted by $C(A)$ is given by

$$C(A) = \{y : y = \sum_{i=1}^{p} c_i a_i\}$$

for scalars $c_i, i = 1, \cdots, p$. Alternatively, $y \in C(A)$ iff $\exists\ c \in R^p$ s.t. $y = Ac$.

Clearly the column space of an $n \times p$ matrix is a subspace of n-dimensional Euclidean space.

Proposition 1.1

1. *If $A = BC$, then $C(A) \subset C(B)$. This is true because: for $\forall a \in C(A)$, $\exists d$ s.t. $a = Ad = BCd = B(Cd) \in C(B)$.*
2. *If $C(A) \subset C(B)$, then there exists a matrix C such that $A = BC$. Proof: (skip the proof in class) let $a_i = A(0, \cdots, 1, \cdots, 0)^T = (a_{1i}, a_{2i}, \cdots, a_{ni})^T$ for $i = 1, \cdots, p$. Note a_i is the ith column of A. Clearly all a_i's are in $C(A)$ thus in $C(B)$. Therefore there exist c_i's s.t. $a_i = Bc_i$. This implies that $A = B(c_1|c_2|\cdots|c_p)$.*

Example:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$$

Is $(1, 1, 1)^T \in C(A)$? (yes) Is $(1, 1, 0)^T \in C(A)$? (no)

Theorem 1.2 $C(AA^T) = C(A)$

Proof: by the definition of $C(AA^T)$, we have $C(AA^T) = \{y : y = AA^Tc = A(A^Tc)\}$, i.e., $C(AA^T) \subset C(A)$. Because the ranks of AA^T and A are the same, the two column spaces have the same number of linearly independent column vectors, which implies that $C(AA^T) = C(A)$.

Definition Let A be and $n \times n$ matrix. A nonzero vector x of length n is called an eigenvector of A if $Ax = \lambda x$ for some scalar λ. The scalar λ is called the eigenvalue of A corresponding to the eigenvector x.

Note: For a square matrix A, $det(A) = \prod \lambda_i$ and the $\lambda_i's$ satisfy $det(\lambda I - A) = 0$; in other words, $\lambda_i's$ are the roots of $det(\lambda I - A) = 0$.

1.2.1 Trace

Definition The trace of a matrix $A = \{a_{ij}\}_{p\times p}$ is $tr(A) = \sum a_{ii}$.

Some properties related to trace

1. $tr(A + B) = tr(A) + tr(B)$
2. $tr(AB) = tr(BA)$. e.g., $x \in R^n$ and $A_{n\times n}$. Then $tr(Axx^T) = tr(x^T Ax) = x^T Ax$.
3. If A is a square matrix with eigenvalues $\lambda_i, i = 1, \cdots, n$, then $tr(A) = \sum \lambda_i$ and $det(A) = \prod \lambda_i$.
4. If A is any $n \times n$ matrix and P is any $n \times n$ nonsingular matrix, then $tr(P^{-1}AP) = tr(A)$.
5. If A is any $n \times n$ matrix and C is any $n \times n$ orthogonal matrix, then $tr(C^T AC) = tr(A)$, as an orthogonal matrix C satisfies $CC^T = C^T C = I$.

Proof of (2): (skip the proof in class) Let $A_{n\times p}$ and $B_{p\times n}$. Then

$$(AB)_{ii} = \sum_{k=1}^{p} a_{ik}b_{ki} \Rightarrow trace(AB) = \sum_{i=1}^{n}\sum_{k=1}^{p} a_{ik}b_{ki}$$

$$(BA)_{jj} = \sum_{k=1}^{n} b_{jk}a_{kj} \Rightarrow trace(BA) = \sum_{j=1}^{p}\sum_{k=1}^{n} b_{jk}a_{kj} = \sum_{i=1}^{n}\sum_{k=1}^{p} a_{ik}b_{ki}$$

Example. $A = \begin{pmatrix} 1 \\ \cdots \\ 1 \end{pmatrix}, B = (1, \cdots, 1)$. $tr(AB) = tr(BA) = n$.

Proof of (3): (skip the proof in class) By the definition of eigenvalues, λ_i's are the n roots to $det(\lambda I_n - A) = 0$. In other words, $det(\lambda I_n - A) = \prod(\lambda - \lambda_i) = \lambda^n - \lambda^{n-1}(\lambda_1 + \cdots + \lambda_n) + \cdots + (-1)^n \lambda_1 \cdots \lambda_n$.

Note that

$$det(\lambda I_n - A) = |\begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \lambda - a_{nn} \end{pmatrix}|$$

So we have

$$\lambda^n - \lambda^{n-1}(\lambda_1 + \cdots + \lambda_n) + \cdots + (-1)^n \lambda_1 \cdots \lambda_n = |\begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \lambda - a_{nn} \end{pmatrix}|$$

Compare the coefficient of $\lambda^(n-1)$ of both sides, we can conclude that

$$-(\lambda_1 + \cdots + \lambda_n) = -(a_{11} + \cdots + a_{nn}),$$

indicating that $tr(A) = \sum \lambda_i$.

Let $\lambda = 0$ and compare both sides we have

$$(-1)^n \lambda_1 \cdots \lambda_n = det(-A) = (-1)^n det(A),$$

indicating that $det(A) = \prod \lambda_i$.

Proof of (4): (show the proof in class) $tr(P^{-1}AP) = tr(APP^{-1}) = tr(A)$.

1.3 symmetric, Positive-definite and positive-semidefinite, idempotent, projection

1.3.1 Symmetric

Definition We say a square matrix A is symmetric if $A = A^T$.

Some properties of a symmetric matrix. If $A_{n\times n}$ is a symmetrix matrix, then

1. there exists an orthogonal matrix $\Gamma = (\gamma_1, \cdots, \gamma_n)$ and a diagonal matrix $\Lambda = diag(\lambda_1, \cdots, \lambda_n$ such that $A = \Gamma\Lambda\Gamma^T$. This factorization is known as the spectral decomposition of a symmetric matrix.
2. $A = \sum_{i=1}^n \lambda_i \gamma_i \gamma_i^T$.
3. $A\Gamma = \Gamma\Lambda$, which implies that $A\gamma_i = \lambda_i\gamma_i$. Thus, $\lambda_i' s$ are eigenvalues and $\gamma_i' s$ are the correponing eigenvectors.
4. $rank(A)$ is equal to the number of non-zero eigenvalues.
5. $tr(A^s) = \sum \lambda_i^s$
6. If A is also non-singular, then $tr(A^{-1}) = \sum \lambda_i^{-1}$

Proof of (1): The proof is length and requires basic knowledge of eigenvalues and eigenvectors. We omit the proof.

Proof of (4): Recall that for any matrix A and conformable and nonsingular B and C we have $rank(BAC) = rank(A)$. Therefore, $rank(A) = rank(\Gamma\Lambda\Gamma^T) = rank(\Lambda) = \sum I(\lambda_i \neq 0)$

Proof of (5): $tr(A^s) = tr((\Gamma\Lambda\Gamma^T)\cdots(\Gamma\Lambda\Gamma^T)) = tr(\Gamma(\Lambda)^s\Gamma^T) = \sum \lambda_i^s$

Proof of (6): When A is nonsingular, the eigenvalues of A^{-1} is λ_i^{-1}. Thus, there exists Γ such that $A^{-1} = \Gamma(\Lambda)^{-1}\Gamma^T$ and the result follows immediately.

Example. $A = \begin{pmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ 2/\sqrt{5} & -1/\sqrt{5} \end{pmatrix}$

Theorem 1.3 *Singular value decomposition: If A is an $n \times p$ matrix of rank r, then A can be written as*

$$A_{n\times p} = U_{n\times r} L_{r\times r} V^T_{r\times p}$$

where $U(n\times r)$ and $V(p\times r)$ are column orthogonal matrices ($U^TU = V^TV = I_r$ and L is a diagonal matrix with positive elements).

Note:$AA^T = UL^2U^T$, and $A^TA = VL^2V^T$.

This result is closely related to the spectral decomposition theorem.

Theorem 1.4 *For any symmetric matrix A, there exists an orthogonal transformation $y = \Gamma^T x$ such that*

$$x^T Ax = \sum \lambda_i y_i^2$$

Proof: Consider the spectral decomposition of $A = \Gamma\Lambda\Gamma^T$ and the transformation $y = \Gamma^T x$ Then

$$x^T Ax = x^T\Gamma^T\Lambda\Gamma x = y^T\Lambda y = \sum \lambda_i y_i^2$$

1.3.2 Positive definite and positive-semidefinitive

Definition A quadratic form of a vector x is a function of the form

$$Q(x) = x^T Ax = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_i x_j$$

where A is a symmetric matrix; that is

$$Q(x) = a_{11}x_1^2 + \cdots + a_{pp}x_p^2 + 2a_{12}x_1x_2 + \cdots + 2a_{p-1,p}x_{p-1}x_p$$

Example.

1. $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $x^TAx = x_1^2 + 2x_1x_2 + x_2^2$. Note that for any $x \neq 0$, $x^TAx \geq 0$.
2. $A = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$, $x^TAx = x_1^2 - x_1x_2 + x_2^2 = (x_1 - x_2/2)^2 + 3x_2^2/4$. Note that for any $x \neq 0$, $x^TAx > 0$.
3. $A = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$, $x^TAx = x_1^2 - 4x_1x_2 + x_2^2 = (x_1 - x_2)^2 - 2x_1x_2$.

Definition $Q(x)$ is called a positive definite (p.d.) quadratic form if $Q(x) > 0$ for all $x \neq 0$; it is call positive semi-definite (p.s.d.) if $Q(x) \geq 0$ for all $x \neq 0$. A symmetric matrix A is called p.d. (or p.s.d.) if $Q(x)$ is p.d. (or p.s.d.). We use the notation > 0 for positive definite and ≥ 0 for positive semi-definite.

Example, let $x = (x_1, x_2)^T$. $Q(x) = x_1^2 + x_2^2$ is p.d., $Q(x) = (x_1 - x_2)^2$ is p.s.d.

Properties of p.s.d matrices. Let $A_{p\times p}$ be p.s.d. Then

1. $\lambda_i \geq 0$ for $i = 1, \cdots, p$. The result implies that if $A \geq 0$, then $tr(A) \geq 0$. Proof: $A \geq 0$ implies that $0 \leq x^T Ax = x^T \Gamma\Lambda\Gamma^T x = \lambda_1 y_1^2 + \cdots + \lambda_p y_p^2$ (Thm 1.5). Take $y_1 = 1, y_2 = 0, \cdots, y_p = 0$ we have $\lambda_1 \geq 0$, etc.

2. There exists a unique p.s.d matrix B such that $A = B^2$. Proof of existence. Consider the spetral decomposition of A: $A = \Gamma\Lambda\Gamma^T$. Let $B = \Gamma\Lambda^{1/2}\Gamma^T$. B is called the square root matrix of A.

   Example. $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, A^{1/2} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$

3. Let C be any $p \times n$ matrix. We have $C^T AC \geq 0$. Proof: Since $A \geq 0$, for any n-vector $x \neq 0$,
$$x^T C^T ACx = (Cx)^T A(Cx) \geq 0$$
   so $C^T AC \geq 0$.

4. $a_{ii} \geq 0$. Proof: take $x = (1, 0, ..., 0)^T$, because A is p.s.d., then $x^T Ax \geq 0$. But $a_{11} = x^T Ax$. So $a_{11} \geq 0$. Similarly, $a_{ii} \geq 0$ for $i = 1, \cdots, p$.

Properties of p.d. matrices. Lat A be p.d.

1. $rank(CAC^T) = rank(C)$. Proof: $rank(CAC^T) = rank(CBBC^T) = rank(CB(CB)^T) = rank(CB) = rank(C)$. The last step is true because we know that for any A, when P and$Q$ are conformable and nonsingular, $rank(PAQ) = rank(A) = rank(PA) = rank(AQ)$.

2. If $C(n \times p)$ has rank n, then CAC^T is p.d. Proof: $x^T CAC^T x = y^T Ay \geq 0$ with equality iff $y = 0$ iff $C^T x = 0$ iff $x = 0$ (since the columns of C^T are linearly independent). Hence $x^T CAC^T x > 0$ for all $x \neq 0$.

3. If $B(n \times p)$ of rank p, then $B^T B$ is p.d. Proof: $x^T B^T Bx = y^T y \geq 0$ with equality iff $Bx = 0$ iff $x = 0$ (since the columns of B are linearly independent)

4. The diagonal elements of a p.d. matrix are all positive. Proof: For i=1, take $x = (1, 0, ..., 0)^T$, then $x^T Ax = a_{11} > 0$. Similarly, you can prove for other i's.

5. $x^T Ax = 0$ implies $Ax = 0$. Proof: A can be written as B^2. Thus $x^T Ax = (Bx)^T (Bx)$, equal to 0 iff $Bx = 0$, which implies $Ax = 0$.

Homework

1. Let $A_{p\times p}$ be a positive definite matrix and let $B_{k\times p}$ be a matrix of rank $k \leq p$. Then BAB^T is positive definite.

2. Let $A_{p\times p}$ be a positive defnite matrix and let $B_{k\times p}$ a matrix. If $k > p$ or if $rank(B) = r < p$, then BAB^T is positive semidefinite.

Theorem 1.5 *A symmetric matrix A is p.d. if and only if there exists a nonsingular matrix P such that $A = P^T P$.*

proof: $\Rightarrow$: By the spectral decomposition, $A = \Gamma\Lambda\Gamma^T$, where Γ is orthogonal and $\Lambda =$ is diagonal. Because A is p.d., $\Gamma^T A\Gamma = \Lambda$ is also p.d.. Thus, we can write Λ to $\Lambda^{1/2}\Lambda^{1/2}$, which leads to $A = P^T P$.

$\Leftarrow$: Since P is nonsingular, its columns are linearly independent. Therefore, for any $x \neq 0$, $Px \neq 0$, and $x^T Ax = x^T P^T Px = (Px)^T(Px) > 0$, i.e., A is positive definite.

Lemma 1.6 *For conformable matrices, the nonzero eigenvalues of AB are the same as those of BA. The eigenvalues are identical for square matrices.*

Proof: Let λ be a nonzero eigenvalue of AB. Then there exists $u(u \neq 0)$ such that $ABu = \lambda u$; that is, $BABu = \lambda Bu$. Hence $BAv = \lambda v$, where $v = Bu \neq 0$ (as $ABu \neq 0$), and λ is an eigenvalue of BA. The argument reverses by interchanging the roles of A and B. For square matrices AB and BA have the same number of zero eigenvalues.

Theorem 1.7 *Let A and B be two symmetric matrices. Suppose that $B > 0$. Then the maximum (minimum) of $x^T Ax$ given $x^T Bx = 1$ is attained when x is the eigenvector of $B^{-1}A$ corresponding to the largest (smallest) eigenvalue of $B^{-1}A$. Thus, if λ_1 and λ_p are the largest and the smallest eigenvalue of $B^{-1}A$, respectively, then subject to the constrain $x^T Bx = 1$,*

$$max_x x^T Ax = \lambda_{max}(B^{-1}A), min_x x^T Ax = \lambda_{min}(B^{-1}A)$$

This is a homework problem. Hint: consider $y = B^{1/2}x$

The above theorem can also be written as

$$max_{x\neq 0}\frac{x^TAx}{x^TBx} = \lambda_{max}(B^{-1}A), min_{x\neq 0}\frac{x^TAx}{x^TBx} = \lambda_{min}(B^{-1}A)$$

A useful result from the theorem: Let $B = I$, then

$$\begin{aligned} max_{x,x\neq 0}\left\{\frac{x^TAx}{x^Tx}\right\} &= \lambda_{max} \\ min_{x,x\neq 0}\left\{\frac{x^TAx}{x^Tx}\right\} &= \lambda_{min} \end{aligned}$$

where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of A, respectively. These values occur when x is the eigenvector corresponding to the eigenvalues λ_{min} and λ_{max}, respectively.

A similar theorem:

Theorem 1.8 *If L is p.d., then for any b,*

$$max_{h,h\neq 0}\left\{\frac{(h^Tb)^2}{h^TLh}\right\} = b^TL^{-1}b$$

Proof: The theorem can be proved by using the previous theorem because we can rewrite $(h^Tb)^2$ to h^Tbb^Th. Thus, the maximum is

$$\lambda_{max}(L^{-1}bb^T) = \lambda_{max}(b^TL^{-1}b) = b^TL^{-1}b$$

The last step is true because $b^TL^{-1}b$ is a scalar; the second to the last step is true because the nonzero eigenvalues of AB are the same as those of BA.

1.3.3 Idempotent matrices and projection matrices

Definition A matrix P is <u>idempotent</u> if $P^2 = P$.

Example. $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, A^2 = A.$

Theorem 1.9 *If P is idempotent, so is $I - P$.*

Proof: $(I-P)^2 = I - 2P + P^2 = I - 2P + P = I - P$.

Definition A symmetric and idempotent matrix is called a <u>projection</u> matrix.

Example. Example. $A = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, A = A^T, A^2 = A.$

Properties of projection matrices:

1. If P is symmetric, then P is idempotent with rank r if and only if it has r eigenvalues equal to 1 and $n-r$ eigenvalues equal to zero.
2. If P is a projection matrix, then $tr(P) = rank(P)$.
3. Projection matrices are positive-semidefinite.
4. If $P_i (i = 1, 2)$ is a projection matrix and $P_1 - P_2$ is p.s.d., then
   - $P_1P_2 = P_2P_1 = P_2$ (note, both P_1P_2 and P_2P_1 are projection matrices).
   - $P_1 - P_2$ is a projection matrix.

Proof of (1): Let λ be an eigenvalue of P, then we have $Px = \lambda x$, where x is the corresponding eigenvector. Since $P^2 = P$, we have $\lambda x = Px = P(Px) = \lambda Px = \lambda^2 x$, which implies that λ equals either zero or 1. For a symmetric matrix, its rank equals the number of nonzero eigenvalues. Thus $rank(P)$ equals the number of 1's.

On the other hand, suppose P has r eigenvalues equal to 1 and $n-r$ eigenvalues equal to 0. Without loss of generality, we assume that the first r eigenvalues are 1. By the spectral decomposition of the symmetric matrix P, we have $P = \Gamma\Lambda\Gamma^T$, where the first r diagonal elements of Λ are 1 and all the other elements are zero. It is obvious that $\Lambda^2 = \Lambda$. Thus $P^2 = \Gamma\Lambda\Gamma^T\Gamma\Lambda\Gamma^T = \Gamma\Lambda\Gamma = P$.

Proof of (2): Based on (1), P has $rank(P)$ eigenvalues equal to 1 and $n - rank(P)$ eigenvalues equal to 0. We also know that $tr(P) = \sum \lambda_i = rank(P)$.

Proof of (3): $x^T Px = X^T PPx = (Px)^T(Px) \geq 0$.

Proof of (4): Clearly the following are true:

$$((I - P_1)y)^T(\text{P1} - \text{P2})((I - P_1)y) \geq 0$$
$$((I - P_1)y)^T P_1((I - P_1)y) = 0$$

which implies that $[(I - P_1)y]^T P_2[(I - P_1)y] \leq 0$. However, we also know that $P_2 \geq 0$. Therefore, it must be true that

$$[(I - P_1)y]^T(P_2 - P_1)[(I - P_1)y] = 0$$

which implies that

$$\begin{aligned}(I - P_1)P_2(I - P_1) = 0 & \\ & \Rightarrow (I - P_1)P_2P_2(I - P_1) = 0 \\ & \Rightarrow (I - P_1)P_2 = 0 \\ & \Rightarrow P_2 = P_1P_2\end{aligned}$$

A projection matrix must be a symmetric matrix. Therefore

$$P_2 = (P_2)^T = (P_1P_2)^T = (P_2)^T(P_1)^T = P_2P_1$$

Last, $(P_1 - P_2)^2 = P_1 - 2P_1P_2 + P_2 = P_1 - 2P_2 + P_2 = P_1 - P_2$.

1.4 Inverse and Generalized Inverse

Definition A full-rank square matrix is called a nonsingular matrix. A nonsingular matrix A has a unique inverse, denoted by A^{-1}:

$$AA^{-1} = A^{-1}A = I.$$

Some properties of A^{-1}:

1. It is unique
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^T)^{-1} = (A^{-1})^T$

Definition For a matrix $A_{n\times p}$, A^- is call a g-inverse (generalized inverse) of A if

$$AA^-A = A$$

A generalized inverse always exists although in general it is not unique.

1.5 Matrix Differentiation

Definition We define the derivative of $f(X)$ with respect to $X_{n\times p}$ as

$$\frac{\partial f(X)}{\partial X} = \left(\frac{\partial f(X)}{\partial x_{ij}}\right)_{n\times p}$$

The following results are useful:

1. $$\frac{\partial a^T x}{\partial x} = a$$

2. $$\frac{\partial x^T x}{\partial x} = 2x, \frac{\partial x^T Ax}{\partial x} = (A + A^T)x, \frac{\partial x^T Ay}{\partial x} = Ay$$

Proof of (2):

$$
\begin{aligned}
\frac{\partial x^T A x}{\partial x_k} &= \frac{\partial}{\partial x_k}\left(\sum_i \sum_j a_{ij} x_i x_j\right) \\
&= \frac{\partial}{\partial x_k}(a_{kk} x_k^2 + \sum_{j \neq k} a_{kj} x_k x_j + \sum_{i \neq k} a_{ik} x_i x_k) \\
&= 2a_{kk} x_k + \sum_{j \neq k} a_{kj} x_j + \sum_{i \neq k} a_{ik} x_i \\
&= \sum_i a_{ki} x_i + \sum_i a_{ik} x_i \\
&= (Ax)_k + (A^T x)_k \\
&= ((A + A^T)x)_k
\end{aligned}
$$