Linear Algebra – Part II

A summary for MIT 18.06SC

Jiaxin Hu

July 6, 2020

1 Orthogonality

1.1 Orthogonal vectors and subspaces

Definition 1 (*Orthogonal vectors*)**.** Suppose two vectors $x, y \in \mathbb{R}^n$. The vectors x and y are orthogonal **iff** $x^T y = y^T x = 0$, denoted $x \perp y$.

Definition 2 (*Orthogonal subspaces*)**.** Suppose two subspaces S, T. The subspaces S and T are orthogonal **iff** for any $s \in S$ and for any $t \in T$, $s^T t = t^T s = 0$, denoted $S \perp T$.

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, there are four subspaces related to $\boldsymbol{A}$: column space $C(\boldsymbol{A})$, row space $C(\boldsymbol{A}^T)$, nullspace $N(\boldsymbol{A})$, and left nullspace $N(\boldsymbol{A}^T)$. Suppose the matrix rank of $\boldsymbol{A}$ is $rank(\boldsymbol{A}) = r$, the dimensions of these subspaces are:

$$dim(C(\boldsymbol{A})) = dim(C(\boldsymbol{A}^T)) = r, \quad dim(N(A)) = n - r, \quad dim(N(\boldsymbol{A}^T)) = m - r.$$

Theorem 1.1 (Orthogonality of matrix subspaces)**.** *Suppose a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. The row space $C(\boldsymbol{A}^T)$ and the nullspace $N(\boldsymbol{A})$ are orthogonal; the column space $C(\boldsymbol{A})$ and the left nullspace $N(\boldsymbol{A}^T)$ are orthogonal i.e.*

$$C(\boldsymbol{A}^T) \perp N(\boldsymbol{A}); \quad C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T).$$

Proof. Consider the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. For any vector $x \in N(\boldsymbol{A})$, we have $\boldsymbol{A}x = 0$.

$$\boldsymbol{A}x = \begin{bmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $a_i \in \mathbb{R}^n, i \in [m]$ is the i-th row of $\boldsymbol{A}$. By the definition 1, x is orthogonal with the rows in matrix $\boldsymbol{A}$. For any vector $v \in C(\boldsymbol{A}^T)$, v is a linear combination of the rows, i.e. $v = c_1 a_1 + ... + c_m a_m$, where $c_i, i \in [m]$ are constants. Multiplying vectors v and x,

$$v^T x = c_1 a_1^T x + \cdots + c_m a_m^T x = 0.$$

Therefore, $v \perp x$, and $N(\boldsymbol{A}) \perp C(\boldsymbol{A}^T)$.
Similarly, for any $x \in N(\boldsymbol{A}^T)$, we have $\boldsymbol{A}^T x = 0$, which implies $N(\boldsymbol{A}^T) \perp C(\boldsymbol{A})$. □

Theorem 1.2 (Relationship between $\boldsymbol{A}^T\boldsymbol{A}$ and $\boldsymbol{A}$)**.** *Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times}$. We have*

$$N(\boldsymbol{A}^T\boldsymbol{A}) = N(\boldsymbol{A}); \quad rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}).$$

Proof. Consider the matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$.
If $x \in N(\boldsymbol{A})$, then $\boldsymbol{A}x = 0 \Rightarrow \boldsymbol{A}^T 0 = 0$, which implies that for any $x \in N(\boldsymbol{A})$, the vector $x \in N(\boldsymbol{A}^T\boldsymbol{A})$. Therefore, we only need to prove that for any $x \in N(\boldsymbol{A}^T\boldsymbol{A})$, the vector $x \in N(\boldsymbol{A})$. We prove this by contradiction.
Suppose a vector $x \in N(\boldsymbol{A}^T\boldsymbol{A}^T)$ but $x \notin N(\boldsymbol{A})$. We have

$$\boldsymbol{A}x = b \neq 0, \quad \boldsymbol{A}^T\boldsymbol{A}x = 0 \quad \Rightarrow \quad \boldsymbol{A}^T b = 0.$$

By the first equation $b \in C(\boldsymbol{A})$, and by the third equation $b \in N(\boldsymbol{A}^T)$. This contradicts the theorem 1.1, i.e. $C(\boldsymbol{A}) \perp N(\boldsymbol{A}^T)$.
Next, given $N(\boldsymbol{A}^T\boldsymbol{A}) = N(\boldsymbol{A})$, the rank of matrix $rank(\boldsymbol{A}^T\boldsymbol{A}) = n - dim(N(\boldsymbol{A}^T\boldsymbol{A})) = n - dim(N(\boldsymbol{A})) = rank(\boldsymbol{A})$. □

Corollary 1 (Invertibility of $\boldsymbol{A}^T\boldsymbol{A}$)**.** If $\boldsymbol{A}$ has independent columns, then $\boldsymbol{A}^T\boldsymbol{A}$ is invertible.

Proof. If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has independent columns, then $rank(\boldsymbol{A}) = n$. By the theorem 1.2, $rank(\boldsymbol{A}^T\boldsymbol{A}) = rank(\boldsymbol{A}) = n$. Since $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a square matrix, $\boldsymbol{A}^T\boldsymbol{A}$ is invertible. □

1.2 Projections onto subspaces

Definition 3 (*Projection and projection matrix*)**.** Consider a vector $x \in \mathbb{R}^m$ and a matrix $\boldsymbol{A}^{m\times n}$ that has independent columns. Suppose a vector $p \in C(\boldsymbol{A})$, such that

$$(x - p) \perp C(\boldsymbol{A}). \tag{1}$$

The vector p is the projection of vector x onto the space $C(\boldsymbol{A})$. Since $p \in C(\boldsymbol{A})$, there exists a vector $\hat{x}$ such that $p = \boldsymbol{A}\hat{x}$. By equation (1), we have

$$\boldsymbol{A}^T(x - p) = \boldsymbol{A}^T(x - \boldsymbol{A}\hat{x}) \quad \Rightarrow \quad \hat{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T x, \quad p = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T x.$$

The matrix $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is called projection matrix.

Theorem 1.3 (Properties of projection matrix)**.** *Consider a projection matrix $\boldsymbol{P}$ to the column space $C(\boldsymbol{A})$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a matrix. Then,*

$$\boldsymbol{P}^T = \boldsymbol{P}; \quad \boldsymbol{P}^2 = \boldsymbol{P}.$$

Proof. Since the projection matrix $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ and $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric, then

$$\boldsymbol{P}^T = (\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T)^T = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{P}.$$

$$\boldsymbol{P}^2 = \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \boldsymbol{P}.$$

□

Corollary 2 (Projection onto $N(\boldsymbol{A}^T)$)**.** Suppose $\boldsymbol{P}$ is a projection matrix in theorem 1.3 , then $I - \boldsymbol{P}$ is also a projection matrix to the left nullspace $N(\boldsymbol{A}^T)$.

Proof. For any $x \in \mathbb{R}^m$, we have $x - \boldsymbol{P}x \perp C(\boldsymbol{A})$

$$x - \boldsymbol{P}x \perp C(\boldsymbol{A}) \quad \Rightarrow \quad (I - \boldsymbol{P})x \perp C(\boldsymbol{A}) \quad \Rightarrow \quad (I - \boldsymbol{P})x \in N(\boldsymbol{A}^T) \text{ and } (x - (I - \boldsymbol{P})x) \perp N(\boldsymbol{A}^T).$$

Therefore, $I - \boldsymbol{P}$ is a projection matrix to the left nullspace $N(\boldsymbol{A}^T)$. □

1.3 Projection matrices and least squares

Given observation vector $y \in \mathbb{R}^n$ and the design matrix $\boldsymbol{X} \in \mathbb{R}^{n\times(k+1)}$, we propose the linear regression model

$$y = \boldsymbol{X}\beta + \epsilon,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_k)$ are coefficients of our interests and ϵ is the noise. The least square estimate of β minimizes the loss

$$\hat{\beta}_{LS} = \underset{\beta\in\mathbb{R}^{k+1}}{\arg\min} \|y - \boldsymbol{X}\beta\|^2 ,$$

where $\|\cdot\|$ is the euclidean norm. The vector $\boldsymbol{X}\hat{\beta}_{LS}$ can be considered as a the projection of y onto the column space of $\boldsymbol{X}$. Therefore, we may use projection tools to solve the least square estimate. The projection $\boldsymbol{X}\hat{\beta}_{LS}$ satisfies

$$\boldsymbol{X}^T(y - \boldsymbol{X}\hat{\beta}_{LS}) = 0 \quad \Rightarrow \quad \hat{\beta}_{LS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T y.$$

The estimate $\hat{\beta}_{LS}$ is identical to the estimates solved by other methods using the derivative.

1.4 Orthogonal matrices and Gram-Schimidt

Definition 4 (*Orthonormal vectors*)**.** The vectors $q_1, ..., q_n$ are orthonormal if

$$q_i^T q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

Orthonormal vectors are always independent.

Definition 5 (Orthonormal matrix and orthogonal matrix)**.** Consider a matrix $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$. If the columns of $\boldsymbol{Q}$ are orthonormal, the matrix $\boldsymbol{Q}$ is an orthonormal matrix. If $m = n$, the square matrix $\boldsymbol{Q}$ is a orthogonal matrix.

If $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ is an orthonormal matrix, $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n$. If $\boldsymbol{Q}$ is an orthogonal matrix, $\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$. For orthonormal matrix $\boldsymbol{Q}$, the projection matrix to $C(\boldsymbol{P})$ becomes $\boldsymbol{P} = \boldsymbol{I}_m$.

Definition 6 (Gram-Schimidt Process and QR decomposition)**.** Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ with $rank(\boldsymbol{A}) = n$. Gram-Schimidt process finds the orthonormal basis for $C(\boldsymbol{A})$. Let $a_i \in \mathbb{R}^m, i \in [m]$ be the columns of matrix $\boldsymbol{A}$.

$$\begin{aligned}
u_1 &= a_1, & e_1 &= \frac{u_1}{\|u_1\|} \\
u_2 &= a_2 - \frac{u_1^T a_2}{u_1^T u_1} u_1, & e_2 &= \frac{u_2}{\|u_2\|} \\
u_3 &= a_3 - \frac{u_1^T a_3}{u_1^T u_1} u_1 - \frac{u_2^T a_3}{u_2^T u_2} u_2, & e_2 &= \frac{u_3}{\|u_3\|} \\
&\vdots
\end{aligned}$$

The vectors $e_1, ..., e_n$ are orthonormal basis of the $C(\boldsymbol{A})$. By matrix operations, we obtain a decomposition of matrix $\boldsymbol{A}$

$$\boldsymbol{A} = [a_1, ..., a_n] = [e_1, ..., e_2] \begin{bmatrix} e_1^T a_1 & e_1^T a_2 & \cdots & e_1^T a_n \\ 0 & e_2^T a_2 & \cdots & e_2^T a_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e_n^T a_n \end{bmatrix} = \boldsymbol{QR} \tag{2}$$

where $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ is an orthonormal matrix and $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ is a upper triangular matrix. We call the matrix decomposition as equation (2) QR decomposition.