

Linear Algebra

A summary for MIT 18.06SC

Jiaxin Hu

June 27, 2020

1 Matrices & Spaces

1.1 Basic concepts

"a" linear combination

- Given vectors v_1, \dots, v_n and scalars c_1, \dots, c_n , the sum $c_1v_1 + \dots + c_nv_n$ is called the *linear combination* of v_1, \dots, v_n .
- The vectors v_1, \dots, v_n are *linearly independent* (or just *independent*) if $c_1v_1 + \dots + c_nv_n = 0$ holds only when all $c_1, \dots, c_n = 0$. If the vectors v_1, \dots, v_n are *dependent*, there exist scalars c_1, \dots, c_n which are not all equal to 0 ~~and satisfy~~ such that $c_1v_1 + \dots + c_nv_n = 0$.
- Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$, the multiplication \mathbf{Ax} is a linear combination of the columns of \mathbf{A} and $x^T \mathbf{A}$ is a linear combination of the rows of \mathbf{A} .
- Matrix multiplication is not communicative, i.e. $\mathbf{AB} \neq \mathbf{BA}$.
Q: When is matrix multiplication commutative?
Hint: characterize the conditions in terms of spectral property.
- Suppose \mathbf{A} is a square matrix. The matrix \mathbf{A} is *invertible* or *non-singular* if there exists a \mathbf{A}^{-1} such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} = \mathbf{I}$. Otherwise, the matrix \mathbf{A} is singular, i.e. its determinant is 0 and **does not have inverse matrix**.
ambiguous subject
- The inverse of a matrix product \mathbf{AB} is $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. The product of invertible matrices is still invertible.
- The transpose of a matrix product \mathbf{AB} is $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. For any invertible matrix \mathbf{A} , $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.
- A matrix \mathbf{Q} is orthogonal if $\mathbf{Q}^T = \mathbf{Q}^{-1}$. A matrix \mathbf{Q} is unitary if $\mathbf{Q}^* = \mathbf{Q}^{-1}$, where \mathbf{Q}^* is the conjugate transpose of \mathbf{Q} .

1.2 Permutation of matrices

For any matrix \mathbf{A} , we can swap its rows by multiplying a *permutation matrix* \mathbf{P} on the left of \mathbf{A} . For example,

$$\mathbf{PA} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_3 \\ a_1 \\ a_2 \end{bmatrix}$$

where a_k refers to the k -th row of \mathbf{A} . The inverse of permutation matrix \mathbf{P} is $\mathbf{P}^{-1} = \mathbf{P}^T$, which implies the orthogonality of permutation matrix. For an $n \times m$ matrix, there are $n!$ different row permutation ~~matrix~~
~~matrices~~ and these permutation matrices form a *multiplicative group*.

Similarly, we can also swap the columns of the matrix \mathbf{A} by multiplying a permutation matrix on the right of \mathbf{A} .

1.3 Elimination of matrices

Elimination is an important technique in linear algebra. We eliminate the matrix by multiplications and subtractions. Take a 3-by-3 matrix \mathbf{A} as example.

~~an example~~

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{\text{step 1}} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{\text{step 2}} \mathbf{U} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}$$

In step 1, we choose the number 1 in row 1 column 1 as a *pivot*, then we recopy the first row and multiply an appropriate number (in this case, 3) and subtract those values from the numbers in the second row. We have thus eliminated 3 in row 2 column 1. Similarly, in step 2, we choose 2 in row 2 column 2 as a pivot and eliminate the number 4 in row 3 column 2. The number 5 in row 3 column 3 is also a pivot. The matrix \mathbf{U} is an upper triangular matrix.

The *elimination matrix* used to eliminate the entry in row m column n is denoted \mathbf{E}_{mn} . In previous example,

$$\mathbf{E}_{21}\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix}; \quad \mathbf{E}_{32}(\mathbf{E}_{21}\mathbf{A}) = \mathbf{U}.$$

Pivots can not be 0. If there is a 0 in the pivot position, we must exchange the row with one below to get a non-zero value in pivot position. If there is not non-zero value below the 0 pivot, then we skip this column and find a pivot in next column.

Since matrix multiplication is associative, we can write $\mathbf{E}_{32}(\mathbf{E}_{21}\mathbf{A}) = (\mathbf{E}_{32}\mathbf{E}_{21})\mathbf{A} = \mathbf{U}$. Let \mathbf{E} denote the product of all elimination matrices. If we need to permute the rows during the process, we have $\mathbf{E}\mathbf{P}\mathbf{A} = \mathbf{U}$, where \mathbf{P} is the product of all the needed permutation matrices.

We also prove the invertibility of the elimination matrix.

~~multiplies~~ ~~means multiplying~~ ~~a scalar $-c$ to the j -th row and subtracting~~ ~~the row from i -th row, where $i \neq j$. Then, \mathbf{E}_{ij} is invertible.~~

Proof. The elimination matrix can be written as:

$$\mathbf{E}_{ij} = \mathbf{I}_n + ce_i e_j^T,$$

~~(missing article) no such term called “identity vector” “basis vector”~~

where $e_i \in \mathbb{R}^n$ is ~~identity vector~~ with value 1 on the i -th entry and value 0 on the other entries. Note that $e_i^T e_j = 0$ because $i \neq j$. Therefore, we have

$$(\mathbf{I}_n + ce_i e_j^T)(\mathbf{I}_n - ce_i e_j^T) = \mathbf{I}_n - c^2 e_i e_j^T e_i e_j^T = \mathbf{I}_n; \quad (\mathbf{I}_n - ce_i e_j^T)(\mathbf{I}_n + ce_i e_j^T) = \mathbf{I}_n \quad \text{punctuation}$$

Thus, $\mathbf{I}_n - ce_i e_j^T$ is the inverse of \mathbf{E}_{ij} and \mathbf{E}_{ij} is invertible. \square

Highlighted sentences have either wrong or bad grammar.

1.4 Gauss-Jordan Elimination

Consider an invertible matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, one of the effective ways to find the inverse of \mathbf{A} uses elimination.

The inverse of \mathbf{A} , \mathbf{A}^{-1} , satisfies $\mathbf{AA}^{-1} = \mathbf{I}_n$. Suppose there is an elimination \mathbf{E} such that $\mathbf{EA} = \mathbf{I}_n$. Multiplying \mathbf{E} on the both side of the equation, we have $\mathbf{EAA}^{-1} = \mathbf{A}^{-1} = \mathbf{E}$. To obtain an such \mathbf{E} , we do elimination to the *augmented matrix* $[\mathbf{A}|\mathbf{I}_n]$ until \mathbf{A} becomes \mathbf{I}_n .

We call the elimination process of finding \mathbf{E} as *Gauss-Jordan Elimination*.

1.5 Factorization of matrices

you should type ``....'' in latex

By elimination, for any square matrix \mathbf{A} , we have $\mathbf{EPA} = \mathbf{U}$ where \mathbf{U} is an upper triangular matrix. By "canceling" the elimination matrix \mathbf{E} , we get $\mathbf{PA} = \mathbf{E}^{-1}\mathbf{U}$. Because \mathbf{E} is invertible, the inverse \mathbf{E}^{-1} exists. Note that \mathbf{E} is lower triangular matrix, the inverse of \mathbf{E} is also a lower triangular matrix. We use \mathbf{L} to denote \mathbf{E}^{-1} . Therefore, we can decompose an arbitrary square matrix \mathbf{A} as:

$$\mathbf{PA} = \mathbf{LU},$$

where \mathbf{U} is an upper triangular matrix with pivots on the diagonal, \mathbf{L} is lower triangular matrix with ones on the diagonal, and \mathbf{P} is a permutation matrix. Note that, there may exist other ways to decompose \mathbf{PA} into \mathbf{LU} , where \mathbf{U} and \mathbf{L} have different settings.

1.6 Time complexity of elimination

ambiguity.. see page 7 in phd-writing-Cochrane.pdf

For an n -by- n matrix, multiplying one row and then subtracting it from another row require $2n$ operations. There are n rows in the matrix, so the total number of operations used in elimination the first column is $2n^2$. The second row and column are shorter, which may cost $2(n-1)^2$ operations and so on. Therefore, the time complexity to factorize \mathbf{A} into \mathbf{LU} is about $\mathcal{O}(n^3)$:

$$1^2 + 2^2 + \cdots + (n-1)^2 + n^2 = \sum_i^n i^2 \approx \int_0^n x^2 dx = \frac{1}{3}n^3.$$

1.7 Reduced row echelon form of matrices

By continuing to use the method of elimination, we can convert \mathbf{U} to a matrix \mathbf{R} in reduced row echelon form, with pivots equal to 1 and zeros above and below the pivots. The matrix \mathbf{R} is called the *reduced row echelon form of matrices*(RREF) of \mathbf{A} . In previous example,

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix} \xrightarrow{\text{make pivots = 1}} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{0 \text{ above and below pivots}} \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For an another example,

$$\mathbf{U} = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{make pivots = 1}} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{0 \text{ above and below pivots}} \mathbf{R} = \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

With proper permutation, the matrix \mathbf{R} can be written in form $[\mathbf{I} \quad \mathbf{F}]$, $\begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix}$, $\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$, or just \mathbf{I} , where \mathbf{F} can be arbitrary matrix in proper dimension. The columns in \mathbf{A} which correspond to the identity matrix \mathbf{I} are called *pivot columns* and the other columns are *free columns*.

“only” should immediately precede the noun that being modified

1.8 Vector space, Subspace and Column space

- *Vector space* is a collection of vectors that are closed under linear combination (addition and multiplication by any real number); i.e. for any vectors in the collection, all the combinations of these vectors are still in the collection.
- *Subspaces of the vector space* is a vector space that is contained inside of another vector space.

Note that any vector space or subspace must include an origin. For a vector space \mathcal{A} , the subspace of \mathcal{A} can be \mathcal{A} itself or can only contain a zero vector.

grammar mistake I have pointed out before (during NeurIPS submission)

- Vectors v_1, \dots, v_n span a space that consists all the combination of those vectors.
- *Column space* of matrix \mathbf{A} is the space spanned by the columns of \mathbf{A} . Let $C(\mathbf{A})$ denote the column space of \mathbf{A} .

Note that if v_1, \dots, v_n span a space \mathcal{S} , then \mathcal{S} is the smallest space that contain those vectors.

- *Basis* of a vector space is a sequence of vectors v_1, \dots, v_n satisfies: (1) v_1, \dots, v_n are independent; (2) v_1, \dots, v_n span the space.
- *Dimension* of the space is the number of vectors in a basis of the space.

1.9 Matrix rank

The *rank* of a matrix \mathbf{A} is defined as the dimension of the columns space of \mathbf{A} . Rank is also equal to the number of pivot columns of \mathbf{A} . That means:

$$\text{rank}(\mathbf{A}) = \# \text{ of pivot columns of } \mathbf{A} = \text{dimension of } C(\mathbf{A}).$$

We use r to denote the rank of \mathbf{A} . If $\mathbf{A} \in \mathbb{R}^{m \times n}$, then we have $r \leq m, r \leq n$. We say the matrix is full rank if $r = n$ or $r = m$.
could you put this two inequality in one?

The rank of a square matrix is closely related to its invertibility.

Lemma 2 (Full rankness and invertibility). A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is full rank, if and only if \mathbf{A} is an invertible matrix.
Avoid “can” in formal writing. Please correct throughout the note.

Proof. First, assume \mathbf{A} is full rank, we prove that \mathbf{A} has an inverse. We can get find a RREF(\mathbf{A}) by eliminations and permutations. There are \mathbf{E} and \mathbf{P} such that

$$\mathbf{EPA} = \mathbf{R}.$$

Since \mathbf{A} is full rank, \mathbf{A} have n pivots columns and that implies $\mathbf{R} = \mathbf{I}_n$. By lemma 1, \mathbf{E} is invertible. The permutation matrix \mathbf{P} is also invertible. Therefore, \mathbf{EP} is invertible and $\mathbf{AEP} = \mathbf{I}_n$. This implies \mathbf{A} is invertible.

Second, assume \mathbf{A} has an inverse, we prove that \mathbf{A} is full rank. We show this by contradiction. Assume \mathbf{A} has an inverse \mathbf{A}^{-1} and \mathbf{A} is not full rank. Because the rank of \mathbf{A} is equal to the dimension of $C(\mathbf{A})$, the columns of \mathbf{A} are linearly dependent without full rankness. So, there exist a non-zero vector v such that

(1) delete unnecessary logit terms such as “So”. (2) correct grammar mistakes

$$\mathbf{Av} = 0.$$

Multiplying \mathbf{A}^{-1} on the both sides of the equation, we have

$$v = \mathbf{A}^{-1}0 = 0.$$

However, it contradicts to non-zeroness of v . Therefore, \mathbf{A} must be full rank. \square

The rank of \mathbf{A} also effects the number of solutions to the system $\mathbf{Ax} = b$. We will discuss it in next section.

2 Solving $\mathbf{Ax} = b$

Here we discuss the solution situation of the linear system $\mathbf{Ax} = b$. Without specific explanation, the matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and the vector $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

2.1 Solving $\mathbf{Ax} = 0$: Nullspace

The nullspace of matrix \mathbf{A} is the collection of all solutions x to the system $\mathbf{Ax} = 0$. Let $N(\mathbf{A})$ denote the nullspace of \mathbf{A} .

Lemma 3 (Nullspace). *The nullspace of matrix \mathbf{A} is a vector space.*

Proof. To show the $N(\mathbf{A})$ is a vector space, we need to show $N(\mathbf{A})$ is close to linear combination. For any integer k , take arbitrary vectors $v_1, \dots, v_k \in N(\mathbf{A})$ and arbitrary scalars c_1, \dots, c_k . We have,

$$\mathbf{A}(c_1v_1 + \dots + c_kv_k) = c_1\mathbf{Av}_1 + \dots + c_k\mathbf{Av}_k = 0.$$

Therefore, the linear combination $(c_1v_1 + \dots + c_kv_k) \in N(\mathbf{A})$. Then $N(\mathbf{A})$ is a vector space. \square

Lemma 4 (The rank of nullspace). *Suppose the rank of \mathbf{A} is r , then the rank of $N(\mathbf{A})$ is $n - r$.*

“then” is an adv.
Proof. Let \mathbf{R} denote the RREF(\mathbf{A}). The matrix \mathbf{R} can be written in form $\mathbf{R} = \begin{bmatrix} \mathbf{I}_r & \mathbf{F} \\ 0 & 0 \end{bmatrix}$ where $\mathbf{F} \in \mathbb{R}^{r \times (n-r)}$ and 0 are zero matrices with proper dimensions. Let $\mathbf{X} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{I}_{n-r} \end{bmatrix}$. Then we have

$$\mathbf{RX} = \begin{bmatrix} \mathbf{I}_r & \mathbf{F} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -\mathbf{F} \\ \mathbf{I}_{n-r} \end{bmatrix} = 0$$

Therefore, each column \mathbf{X} is a special solution to the system $\mathbf{Ax} = 0$. Next, we want to show other solutions to $\mathbf{Ax} = 0$ are linear combinations of those special solutions.

Suppose there is a $x = (x_1, x_2) \in N(\mathbf{A})$. Then

$$\mathbf{Rx} = \begin{bmatrix} \mathbf{I}_r & \mathbf{F} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + \mathbf{Fx}_2 \\ 0 \end{bmatrix} = 0.$$

That implies $x_1 = -\mathbf{Fx}_2$ and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{I}_{n-r} \end{bmatrix} x_2 = \mathbf{X} x_2$. Therefore, arbitrary $x \in N(\mathbf{A})$ is a linear combination of special solutions, i.e. $C(\mathbf{X}) = N(\mathbf{A})$.

Since \mathbf{X} contains an identity matrix, the special solutions are independent and the rank of $C(\mathbf{X})$ is $n - r$. Therefore, the rank of $N(\mathbf{A})$ is $n - r$. \square

Recall that the columns in \mathbf{A} correspond to the \mathbf{I}_r in \mathbf{R} are called pivot columns and other columns are free columns. In $\mathbf{Ax} = b$, the variables in x that correspond to pivot columns are called *pivot variables* and others are *free variables*. As the proof shows, the way to find special solutions of $\mathbf{Ax} = 0$ is to let one of the free variables is 1 while other free variables are 0 and then solve the equation $\mathbf{Ax} = 0$. There are total $n - r$ special solutions.

2.2 Solving $\mathbf{A}x = b$: complete solutions

Lemma 5 (Solvability of $\mathbf{A}x = b$). *The system $\mathbf{A}x = b$ is solvable only when $b \in C(\mathbf{A})$.*

Proof. For any x , $\mathbf{A}x \in C(\mathbf{A})$. Therefore, to satisfy $\mathbf{A}x = b$, $b \in C(\mathbf{A})$. □

Lemma 6 (Complete solution). *The complete solution of $\mathbf{A}x = b$ is given by $x_{\text{comp}} = x_p + x_n$, where x_p is a particular solution that $\mathbf{A}x_p = b$ and $x_n \in N(\mathbf{A})$.*

such that add period before "and". → common mistakes in your note
Proof. Suppose $x = x_p + x_0$ is an arbitrary solution to $\mathbf{A}x = b$. Then we have

$$\mathbf{A}x - \mathbf{A}x_p = \mathbf{A}(x - x_p) = \mathbf{A}x_0 = 0.$$

Then $x_0 \in N(\mathbf{A})$. □

One way to find a particular solution is to let all the free variables be 0 and solve the equations.

~~Here is a summary table that discusses~~ the rank of \mathbf{A} , the form of \mathbf{R} and the situation about the solutions.

~~The following table discusses ==> Common mistakes. Please use the writing principle to edit the note.~~

	$r = m = n$	$r = n < m$	$r = m < n$	$r < m, r < n$
\mathbf{R}	\mathbf{I}	$\begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix}$	$[\mathbf{I} \quad \mathbf{F}]$	$\begin{bmatrix} \mathbf{I} & \mathbf{F} \\ 0 & 0 \end{bmatrix}$
dimension of $N(\mathbf{A})$	0	0	$n - r$	$n - r$
# solutions to $\mathbf{A}X = b$	1	0 or 1	infinitely many	0 or infinitely many