Linear Algebra Recap

Ang Zi Qing

April 8, 2020

This write up serves as a quick review of Linear Algebra and aims to trigger my memory of Gilbert Strang's 18.06 Linear Algebra course when its ideas are needed for research purposes or possibly for future endeavours. That being said, go through the following key questions and see if they ring any bells. If it does not work, either return to the videos or read the book and then populate this note. I intend to continuously update this note hence it is highly possible that I might go beyond Strang's syllabus in the future. This version is at my best understanding of Linear Algebra at the time of writing.

1. What is the role of $\mathbf{A}$ in $\mathbf{Ax} = \mathbf{b}$?
2. What is the difference between dot product (inner product) and cross product (outer product) and what do they mean?
3. What does it mean by the linear combination of columns of $\mathbf{A}$?
4. What are the all the combinations of columns of $\mathbf{A}$, when the columns are dependent/independent of each other?
5. What does it mean when matrix $\mathbf{A}$ is singular?
6. What do we get from elimination?
7. What happens when elimination is a "temporary"/"complete" failure?
8. What is back substitution?
9. Recall elimination and permutation matrix, what does both the matrices do together?
10. Recall getting $\mathbf{U}$ back to $\mathbf{A}$, as in inverse, when does $\mathbf{A}^{-1}$ exist? If it does exist, how would you find it?
11. What is the significance of checking the invertibility of $\mathbf{A}$?
12. Can a rectangular matrix be invertible?
13. What is the difference of $\mathbf{Ax}$ and $\mathbf{x}^T\mathbf{A}^T$? What is the difference of $\mathbf{x}^T\mathbf{y}$ and $\mathbf{xy}^T$?
14. What is the inverse and transpose of a permutation matrix?
15. How does symmetry produces a special relationship between $\mathbf{L}$ and $\mathbf{U}$?
16. Recall factorization into $\mathbf{A} = \mathbf{LU}$, what is $\mathbf{L}$? What is in the diagonal of $\mathbf{L}$ and $\mathbf{U}$? How is $\mathbf{LDU}$ different from $\mathbf{LU}$? What happens when there is no row exchange? What is the factorization with row exchanges?
17. What does $\mathbf{Ax} = \mathbf{0}$ tell us about $\mathbf{A}$? What happens if only $\mathbf{x} = \mathbf{0}$? What happens otherwise? What does this tells us about $\mathbf{A}^T\mathbf{A}$?

18. Connect a vector space and subspace, what is the relationship between subspace and vector space?
19. Explain the following: independence, span, basis, rank, dimension.
20. Explain the four fundamental subspaces.
21. Explain orthogonal vectors and subspaces.
22. Explain projections onto subspaces.
23. When matrix $\mathbf{A}$ is square and independent, we can solve for $\mathbf{x}$ in $\mathbf{Ax} = \mathbf{b}$. How to solve for $\mathbf{x}$ when $\mathbf{A}$ is rectangular?
24. Recall the connection between orthogonality, projection matrices and least squares. [Linear Regression].
25. What are orthogonal matrices, what do they do?
26. Recall the ten properties of determinants.
27. Connect determinant and volume.
28. What are eigenvalues, eigenvectors and eigenspace?
29. What is in the null space of $\mathbf{A} = \lambda\mathbf{I}$? What is this space?
30. Recall differential equations and the matrix exponential e^{At}.
31. Recall Markov matrices and the Fourier Series.
32. What does it mean when $\mathbf{Q}^{-1} = \mathbf{Q}^T$?
33. Why do we look for positive definite symmetric matrices?
34. Why are the eigenvalues of symmetric matrices real? What to do if a vector or a matrix is complex?
35. What does $\mathbf{A} = \mathbf{Q\Lambda Q}^T$ say about symmetric matrices? What is the Spectral Theorem, $\mathbf{Q\Lambda Q}^{-1} = \mathbf{Q\Lambda Q}^T$?
36. What happens when a matrix has repeating eigenvalues?
37. What happens when the vectors and matrices are complex? What do you do when the numbers become complex numbers?[Hermition]
38. How to tell if a matrix is positive definite and what does it mean? Why do we look for positive definite matrices? [Minima of matrix]
39. Why do we look for the minima (completing the square to get positive values)? What happens if the matrix is not positive definite?
40. Why do similar matrices share the same eigenvalues? What do similar matrices have in common? What makes them different?
41. What makes $\mathbf{A} = \mathbf{Q\Lambda Q}^T = \mathbf{S\Lambda S}^{-1}$?
42. What does it mean to have an SVD? How to find the three components of an SVD of a matrix $\mathbf{A}$ such that $\mathbf{A} = \mathbf{U\Sigma V}^T$?
43. In real application, the first vector of $\mathbf{V}$ and $\mathbf{U}$, as well as the biggest singular value of the decomposition of a matrix $\mathbf{A}$ that are the most important, why? Why these components make the best combination to find correlation between desired variables and hence carry the most information?

1 What is the role of A in $\mathbf{Ax} = \mathbf{b}$?

Considering the two linear equations with two unknowns,

$$2x - y = 0 \tag{1}$$
$$-x + 2y = 3 \tag{2}$$

we can express these equations in matrix form $\mathbf{Ax} = \mathbf{b}$ such that

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \tag{3}$$

where the matrix $\mathbf{A}$ of size $m \times n$ is the coefficient matrix, the column vector $\mathbf{x}$ with m entries is the vector of unknowns and the column vector $\mathbf{b}$ with n entries is the result of matrix $\mathbf{A}$ acting on vector $\mathbf{x}$ (Recall the difference between the row picture and the column picture). Comparing to the equations, the coefficient matrix $\mathbf{A}$ consists of all the coefficients of the variables x and y in the set of linear equations. A right combination of these coefficients by finding some values for x and y would give us the vector $\mathbf{b}$.

The matrix $\mathbf{A}$ plays a crucial role in determining whether $\mathbf{Ax} = \mathbf{b}$ has a solution for $\mathbf{b}$ since a solution for the set of linear equations exists only if $\mathbf{b}$ is a linear combination of the columns of $\mathbf{A}$. To check if $\mathbf{Ax} = \mathbf{b}$ has a solution for every $\mathbf{b}$, we look for a pivot in every row of $\mathbf{A}$. If $m = n$ and $\mathbf{A}$ has a pivot in every row,

- all the combinations fill the entire space, that is, every column vector $\mathbf{b}$ is a linear combination of column vectors of $\mathbf{A}$
- columns of $\mathbf{A}$ are independent
- columns of $\mathbf{A}$ spans the column space $\mathbb{R}^m$, where $\mathbb{R}^m$ also contains all the combinations of columns of $\mathbf{A}$
- the null space has only the zero vector, that is $\mathbf{x} = \mathbf{0}$ (trivial solution) is the only solution to get $\mathbf{b} = \mathbf{0}$.

Solving $\mathbf{Ax} = \mathbf{b}$ is the same as solving the system described by the augmented matrix. We will eventually see that finding the pivots of an invertible $n \times n$ matrix $\mathbf{A}$ leads us to determining bases of finite-dimensional vector spaces, where a chosen basis (a sequence of independent vectors that span the space) become columns of scalars. If we know the number of pivots of $\mathbf{A}$, then the rank and hence the dimension of the column space of $\mathbf{A}$ will subsequently be known, which also gives us the number of vectors in any basis as well as the vectors that span the space. Different choices of basis gives different columns, hence also give different matrices, which corresponds to linear transformations. Ultimately the matrix becomes a tool for us to find the determinant, trace and eigenvalues of the linear transformation, which is a structure-preserving map.

2 What does it mean when matrix A is singular?

Consider the following set of linear equations (1) and (2) and the matrix form (3), we can also express in terms of the column picture

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}. \tag{4}$$

It is clear that the combination of one of vector $\begin{bmatrix} 2 & -1 \end{bmatrix}^T$ and two of vector $\begin{bmatrix} -1 & 2 \end{bmatrix}^T$ gives us the RHS $\begin{bmatrix} 0 & 3 \end{bmatrix}^T$. This demonstrates that $\mathbf{Ax}$ is a combination of columns of $\mathbf{A}$, and the vector $\mathbf{x} = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$.

Consider another set of linear equations,

$$x + 2y = 3 \tag{5}$$
$$2x + 4y = 7, \tag{6}$$

expressing these equations in matrix form gives

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} \tag{7}$$

and in the column picture,

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}. \tag{8}$$

This is a situation where it is impossible to produce some $\mathbf{b}$ out of the columns of $\mathbf{A}$ or their combinations. Notice that the columns of $\mathbf{A}$ are not independent since

$$2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix},$$

which means that the columns lie in the same plane, and their combinations will also lie in the same plane. Nevertheless, we still can get some $\mathbf{b}$ from the plane, for example, when $x = 1$ and $y = 1$,

$$\mathbf{b} = 1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix},$$

but most other $\mathbf{b}$ would be out of the plane and unreachable. The only $\mathbf{b}$s that we can get would be the ones on that plane only. We call this kind of cases SINGULAR CASES such that:

- at least one of the column vectors of $\mathbf{A}$ is a linear combination of the rest
- the columns of $\mathbf{A}$ do not span the entire space
- there will not be a solution for every $\mathbf{b}$
- matrix $\mathbf{A}$ will not be invertible, since not all columns of $\mathbf{A}$ has a pivot. Some pivots are zero in row reduced form after elimination operation, which is the case of complete failure for elimination. [Temporary failures can be resolved by exchanging rows].
- There exist a vector $\mathbf{x}$ other than $\mathbf{x} = \mathbf{0}$ such that $\mathbf{Ax} = \mathbf{0}$, that is, the null space of $\mathbf{A}$ has other vectors other than the zero vector.

3 What is the significance of checking the invertibility of A?

Finding the inverse $\mathbf{A}^{-1}$ is equivalent to solving the system of linear equations. Suppose $\mathbf{E}_{ij}$ is the elimination matrix to fix position (i, j) in order to produce zeros below or above the pivots. To obtain a row echelon form for a 3×3 matrix, for example, without row exchanges,

$$\mathbf{E}_{32}\mathbf{E}_{31}\mathbf{E}_{21}\mathbf{A} = \mathbf{U}, \tag{9}$$

where $\mathbf{U}$ is the row echelon form of $\mathbf{A}$ and is an upper triangular matrix. To reverse the elimination operation,

$$\mathbf{A} = \mathbf{E}_{21}^{-1}\mathbf{E}_{31}^{-1}\mathbf{E}_{32}^{-1}\mathbf{U} \tag{10}$$
$$= \mathbf{LU}, \tag{11}$$

where $\mathbf{L}$ is a lower triangular matrix and connects $\mathbf{A}$ to $\mathbf{U}$. From here we can observe that $\mathbf{L}$ is the product of inverses of the elimination matrices. Let $\mathbf{E} = \mathbf{E}_{32}\mathbf{E}_{31}\mathbf{E}_{21}$. In Gauss-Jordan elimination,

$$\mathbf{E}(\mathbf{AI}) = (\mathbf{IA}^{-1}), \tag{12}$$

then $\mathbf{EA} = \mathbf{I}$ tells us $\mathbf{E} = \mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}$ envelops the steps of elimination to reach $\mathbf{U}$. Therefore, if an inverse of $\mathbf{A}$ exists, a square matrix $\mathbf{A}$ has a full set of pivots and will always have a two sided inverse such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{AA}^{-1}. \tag{13}$$

If $\mathbf{A}^{-1}$ does not exist, then there exists a vector $\mathbf{x}$, where $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{Ax} = \mathbf{0}$, for example

$$\mathbf{Ax} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since multiplying $\mathbf{A}$ by some other matrix gives a resulting matrix with columns being multiples of columns of $\mathbf{A}$, there is no way to obtain the identity matrix, because the identity matrix is not a combination of those columns since they lie on the same line in every combination. If $\mathbf{A}^{-1}$ is multiplied on both sides,

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{0}, \tag{14}$$

then $\mathbf{x} = \mathbf{0}$, which is not true. Hence if $\mathbf{A}^{-1}$ exists, then matrix $\mathbf{A}$ is invertible, columns of $\mathbf{A}$ are independent, and there exist a solution for every $\mathbf{b}$.

4 Can a rectangular matrix be invertible?

A matrix $\mathbf{A}$ is invertible if and only if its column vectors are independent. Not all square matrices can fulfill this condition, hence not all square matrices are invertible, for example, equation (7). In the case of rectangular matrices, not all of them have dependent column vectors. If the column vectors of a matrix $\mathbf{A}$ are independent, then either the matrix is square (i.e. $m = n$), or there are fewer columns than rows (i.e. $m > n$). When $m < n$, the columns cannot be linear independent and hence this rectangular matrix cannot be invertible. So a rectangular matrix $\mathbf{A}$ might have a left inverse or a right inverse, but cannot have a two sided inverse, since a rectangular matrix is not one-to-one and onto.

5 How to solve for a rectangular matrix?

The general solution for both square and rectangular matrix $\mathbf{Ax} = \mathbf{b}$ exists if the matrix $\mathbf{A}$ has a full column rank $r = n$ and/or full row rank $r = m$ such that

$$\mathbf{x} = \mathbf{x}_{\mathbf{particular}} + \mathbf{x}_{\mathbf{special}} \tag{15}$$

For a square and invertible matrix $\mathbf{A}$, its rank $r = m = n$ and $\mathbf{Ax} = \mathbf{b}$ has only one solution such that $\mathbf{x} = \mathbf{x}_{\mathbf{particular}} + \mathbf{0}$ where $\mathbf{x}_{\mathbf{particular}}$ gets the scalar values from $\mathbf{b}$ of the augmented matrix after reduced row echelon form is obtained from elimination. There is no special solution since there are no free variables in the reduced row echelon form.

For a short and wide matrix $\mathbf{A}$ that has more columns than rows, it has a full row rank $r = m$ and $r \leq n$. Recall the 4 subspaces, a full rank matrix $\mathbf{A}$ means that the rows are independent but the columns are not, so the left null space has only the zero vector but the null space has other nonzero vectors in addition to the zero vector. So the particular solution $\mathbf{x}_{\mathbf{particular}}$ can be obtained from $\mathbf{b}$ of the augmented matrix after reduced row echelon form is obtained from elimination. The particular solution will have free variable(s), which gives special solution(s) to $\mathbf{Ax} = \mathbf{0}$. Take an example with $m = 2$ and $n = 3$, that is, two equations and three unknowns with rank $r = m = 2$. The two planes in xyz space intersects in a line of solutions (can be obtained from elimination) and the particular solution is one point on the line. The special solution(s) gives the nullspace vector(s), which adding to the particular solution will move along the solution line, hence this kind of matrix has infinite solutions. (See Figure 3.3 of Strang's book)

For a tall and thin matrix $\mathbf{A}$ that has more rows than columns, it has a full column rank $r = n$ and $r \leq m$. For a solution to exists, zero rows in reduced row echelon form must also be zero in $\mathbf{b}$ of the augmented matrix after elimination. Row reduction puts $\mathbf{I}$ at the top of the augmented matrix when $\mathbf{A}$ is reduced to $\mathbf{R}$ with rank n. All columns of $\mathbf{A}$ are pivot columns and there are no free variables or special solutions. Since all columns are pivot columns, the null space contains only the zero vector. Therefore, this kind of matrix has only one or no solution. We will see later that a matrix with full column rank has a nice property such that multiplying $\mathbf{A}^T$ to both sides of $\mathbf{Ax} = \mathbf{b}$ so that $\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\hat{\mathbf{b}}$ where $\mathbf{A}^T\mathbf{A}$ is square and invertible.

If $r < n$ and $r < m$, the system has zero or infinitely many solutions.

While answering this question, we can also see the significance of knowing the rank of $\mathbf{A}$. Strang brings up three levels of understanding for the definition of rank:

- Level 1: Rank tells us number of pivots.
- Level 2: Rank tells us number of independent rows and columns.
- Level 3: Rank tells us the dimensions of the column and the row spaces, which combining with the information of m and n tells us also the dimensions of the null space and the left-null space.

6 What is the meaning behind least squares? (projection and orthogonality)

To solve $\mathbf{Ax} = \mathbf{b}$ is to express $\mathbf{b}$ as a combination of columns of $\mathbf{A}$. Therefore the system $\mathbf{Ax} = \mathbf{b}$ is solvable if and only if $\mathbf{b}$ is in the column space of $\mathbf{A}$, $C(\mathbf{A})$, that is, $\mathbf{b}$ is a combination of the columns. Suppose $\mathbf{Ax} = \mathbf{b}$ has no solution, we can choose a vector $\mathbf{p}$ in the column space that is closest to $\mathbf{b}$ and solve $\mathbf{A}\hat{\mathbf{x}} = \mathbf{p}$ instead such that $\hat{\mathbf{x}}$ is the best possible combination of the columns of $\mathbf{A}$, that is, we project $\mathbf{b}$ into the column space $C(\mathbf{A})$.

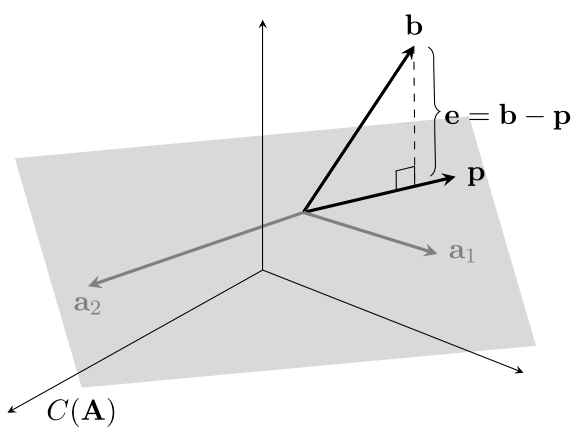

Figure 1: Projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$, $C(\mathbf{A})$.

Consider

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix},$$

then

$$\mathbf{p} = \hat{x}_1\mathbf{a}_1 + \hat{x}_2\mathbf{a}_2 = \mathbf{A}\hat{\mathbf{x}}.$$

From Figure 17, we observe that

$$\mathbf{e} = \mathbf{b} - \mathbf{p} \perp \mathbf{a}_1$$
$$\mathbf{e} = \mathbf{b} - \mathbf{p} \perp \mathbf{a}_2,$$

so

$$\mathbf{a}_1^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0$$
$$\mathbf{a}_2^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = 0.$$

Expressing in matrix form, we have

$$\begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \end{bmatrix}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus we have $\mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0}$, which tells us that $\mathbf{e}$ is in the null space of $\mathbf{A}^T$. Recalling the orthogonality of the four subspaces, this also tells us that $\mathbf{e}$ is perpendicular to $C(\mathbf{A})$. Rearraging the terms, we have

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^T\mathbf{b},$$

$$\hat{\mathbf{x}} = (\mathbf{A}^T\mathbf{A}^{-1})\mathbf{A}^T\mathbf{b}. \tag{16}$$

So

$$\mathbf{p} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \tag{17}$$

and the projection matrix $\mathbf{P}$ that is multiplying $\mathbf{b}$ to give the projection is

$$\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{18}$$

The projection matrix has two properties:

- $\mathbf{P}$ is symmetric such that $\mathbf{P}^T = \mathbf{P}$
- $\mathbf{P}$ is idempotent such that $\mathbf{P}^2 = \mathbf{P}$, which means a second projection does not change the first projection.

If $\mathbf{A}$ is a square and invertible matrix, $C(\mathbf{A})$ is the whole of $\mathbb{R}^n$, then the projection matrix that projects back into the whole space is the identity matrix $\mathbf{I}$. Since $\mathbf{b}$ is already in the column space, $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{I}$. On the other hand, if $\mathbf{A}$ is not a square matrix, then there is no inverse $\mathbf{A}^{-1}$ and $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{I}$ is not allowed. So we have to stay with Equation (17) to project into a subspace. When $\mathbf{A}$ has independent columns, $\mathbf{A}^T\mathbf{A}$ is square, symmetric, and invertible (see page 212 of Strang's book for proof), so a problem that involves a rectangular matrix almost always leads to $\mathbf{A}^T\mathbf{A}$ in order to look for its solution.

To be continued...

7 Symmetric matrices and Positive definiteness

8 Singular Value Decomposition (SVD)

9 Principal Component Analysis (PCA)

10 Linear Transformations