

Applied Linear Algebra Notes

Joe Hollander

Summer 2025

1

1.1

Elementary Row Operations

1. Interchange two rows.
2. Multiply the elements of a row by a nonzero constant.
3. Add a multiple of the elements of one row to the corresponding elements of another.

Gauss-Jordan elimination is a systematic way to eliminate variables to arrive at a solution matrix. The equation is solved when the matrix is diagonalized. Create zeros in each column until finished. Finally, normalize the final element ((3,3) for a 3x3 matrix, etc.). Interchange rows if a diagonal element is zero. Final matrix is called the reduced echelon form. The formula can be used for multiple constants with the same coefficient matrix. To do this, just stack the constants in each column:

$$\begin{aligned}x - y + 3z &= b_1 \\ 2x - y + 4z &= b_2 \\ -x + 2y - 4z &= b_3\end{aligned}$$

for

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 11 \\ -11 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$$

would produce the augmented matrix

$$\begin{bmatrix} 1 & -1 & 3 & 8 & 0 & 3 \\ 2 & -1 & 4 & 11 & 1 & 3 \\ -1 & 2 & -4 & -11 & 2 & 4 \end{bmatrix}$$

which simplifies to

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & -2 \\ 0 & 1 & 0 & -1 & 3 & 1 \\ 0 & 0 & 1 & 2 & 1 & 2 \end{bmatrix}.$$

1.2

A matrix is in reduced echelon form if:

1. Any rows consisting entirely of zeros are grouped at the bottom of the matrix.
2. The first nonzero element of each other row is 1. This element is called a leading 1.
3. The leading 1 of each row after the first is positioned to the right of the leading 1 of the previous row.
4. All other elements in a column that contain a leading 1 are zero.

The reduced echelon of a matrix is unique. To express multiple solutions, write the leading variables in each equation (variables with the lowest subscript: x_1 , etc.) in terms of the remaining free variables. If last row has all zeros apart from constant, there are no solutions for the system. A system of linear equations is said to be homogeneous if all the constants are zero. A homogeneous system of linear equations always has a solution of all zeros. A homogenous system of linear equations that has more variables than equations (more rows than columns) has many solutions.

1.7

Curve fitting a polynomial involves estimating variables, which can be done by solving a system of linear equations. Kirchhoff's Laws state:

1. All the current flowing into a junction must flow out of it
2. The sum of the IR terms (I denotes current, R resistance) in any direction around a closed path is equal to the total voltage in the path in that direction.

Using these laws a system of linear equations can be formed to determine the current flowing through certain points. The same procedure can be used with traffic.

1.3

\mathbb{R}^n is closed under addition and scalar multiplication if the result is also in \mathbb{R}^n . To determine if \mathbf{r} is a linear combination of \mathbf{u} , \mathbf{v} , and \mathbf{w} , determine if this system has a solution:

$$c_1 \mathbf{u}_1 + c_2 \mathbf{v}_1 + c_3 \mathbf{w}_1 = \mathbf{r}_1$$

$$c_1 \mathbf{u}_2 + c_2 \mathbf{v}_2 + c_3 \mathbf{w}_2 = \mathbf{r}_2$$

$$c_1 \mathbf{u}_3 + c_2 \mathbf{v}_3 + c_3 \mathbf{w}_3 = \mathbf{r}_3.$$

The norm of a vector \mathbf{r} is $\|\mathbf{r}\|$, and it denotes the magnitude of a vector. In \mathbb{R}^n , $\|\mathbf{r}\| = \sqrt{\mathbf{r}^n}$.

1.4

Subspaces are subsets of \mathbb{R}^n that have all the same properties (subset must be closed under addition and multiplication). An example for demonstrating a subset in \mathbb{R}^3 with the form $(a, 2a, 3a)$ is closed for addition:

$$\begin{aligned}(a, 2a, 3a) + (b, 2b, 3b) &= (a + b, 2a + 2b, 3a + 3b) \\ &= (a + b, 2(a + b), 3(a + b)),\end{aligned}$$

and multiplication:

$$\begin{aligned}k(a, 2a, 3a) &= (ka, 2ka, 3ka) \\ &= (ka, 2(ka), 3(ka)).\end{aligned}$$

When every vector of a vector space can be written as a linear combination of a set of vectors, the vectors span the space. For example, the subspace in \mathbb{R}^3 defined by $(a, b, a + b)$ is spanned by the vectors $(1, 0, 1)$ and $(0, 1, 1)$ because $(a, b, a + b) = a(1, 0, 1) + b(0, 1, 1)$. Generally, vectors that make up a line in \mathbb{R}^n are subspaces. The set of solutions to any homogenous system of linear equations is a subspace.

1.5

The basis is the finite set of vectors that span a whole space. The basis must also be linearly independent. The vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are linearly independent if the identity $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n = \mathbf{0}$ is only true for $c_1 = 0, c_2 = 0, \dots, c_n = 0$. If the vectors are not linearly independent, they are linearly dependent. Vectors can span the space but not be linearly independent. For example, the vectors $(0, 1), (1, 0)$, and $(1, 1)$ span \mathbb{R}^2 , but the set of the vectors is not a basis for \mathbb{R}^2 because they are not linearly independent. Additionally, \mathbb{R}^2 is not a subspace of \mathbb{R}^3 . All subspaces must pass through the origin because they must contain the $\mathbf{0}$ vector.

2

2.1

Matrix notation:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

The main diagonal consists of a_{11}, a_{22} , and a_{33} . Addition involves adding the corresponding elements of each matrix. Both matrices must be the same size. Scalar multiplication involves multiplying every element. Matrix multiplication ($C = AB$) requires that the number of columns in matrix A , is the same as the

number of rows in matrix B . The product is equal to the product of the rows of A and the columns of B :

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

For example,

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 5 & 0 & 1 \\ 3 & -2 & 6 \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} & \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} & \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix} \\ \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} & \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -2 \end{bmatrix} & \begin{bmatrix} 2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} (1)(5) + (3)(3) & (1)(0) + (3)(-2) & (1)(1) + (3)(6) \\ (2)(5) + (0)(3) & (2)(0) + (0)(-2) & (2)(1) + (0)(6) \end{bmatrix} \\ &= \begin{bmatrix} 14 & -6 & 19 \\ 10 & 0 & 2 \end{bmatrix}. \end{aligned}$$

Matrix multiplication is not commutative, and BA does not exist. If A is an $m \times r$ matrix and B is an $r \times n$ matrix, then AB is an $m \times n$ matrix. Three classes of matrices are notable: The zero matrix O which contains all zeros, the diagonal matrix which contains zeros everywhere except the main diagonal, and the identity matrix I which is a diagonal matrix with ones on the main diagonal. If B_1 is column one of matrix B , then $AB = [AB_1, AB_2, \dots, AB_n]$. Matrices can be partitioned into blocks, and these blocks can be multiplied.

2.2

For matrix multiplication,

1. $A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(A + B)C = AC + BC$

The product of a chain of matrices will have the same number of rows as the first matrix and the same number of columns as the last matrix. For matrices, $AB = AC$ doesn't imply $B = C$, and $PQ = O$ doesn't imply $P = O$ or $Q = O$. If A is an $n \times n$ square matrix, and r and s are nonnegative integers, then

1. $A^r A^s = A^{r+s}$
2. $(A^r)^s = A^{rs}$
3. $A^0 = I_n$

Consider a homogenous system of linear equations $AX = 0$. Let X_1 and X_2 be solutions of the system. Then $AX_1 = 0$ and $AX_2 = 0$, so $AX_1 + AX_2 = A(X_1 + X_2) = 0$, and $X_1 + X_2$ is also a solution. Furthermore, if c is any scalar, then $AX_1 = 0$ and $c(AX_1) = A(cX_1) = 0$, so cX_1 is also a solution.

The corresponding homogeneous system of a matrix is one in which the right-hand side of the equations is zero, and the solution of the corresponding nonhomogeneous system is the solution of the homogeneous system plus a particular solution of the nonhomogeneous system.

2.3

The transpose of a matrix A , denoted A^T , is the matrix whose columns are the rows of the given matrix A . Let A and B be matrices and c be a scalar. There are several properties of the transpose:

1. $(A + B)^T = A^T + B^T$
2. $(cA)^T = cA^T$
3. $(AB)^T = B^T A^T$
4. $(A^T)^T = A$

A symmetrical matrix is a matrix that is equal to its transpose. The trace of A , denoted $tr(A)$, is the sum of the elements on the main diagonal. There are several properties of the trace:

1. $tr(A + B) = tr(A) + tr(B)$
2. $tr(AB) = tr(BA)$
3. $tr(cA) = ctr(A)$
4. $tr(A^T) = tr(A)$

Matrices can also contain complex numbers, and the conjugate of a matrix A is denoted \bar{A} . Furthermore, the conjugate transpose of a matrix A is denoted A^* , and a square matrix C is called Hermitian if $C = C^*$. There are several properties of the conjugate transpose:

1. $(A + B)^* = A^* + B^*$
2. $(zA)^* = \bar{z}A^*$
3. $(AB)^* = B^* A^*$
4. $(A^*)^* = A$

In archeological seriation, a matrix is used to represent the relative ages of artifacts, with the definition

$$a_{ij} = \begin{cases} 1 & \text{if grave } i \text{ contains pottery type } j \\ 0 & \text{if grave } i \text{ does not contain pottery type } j \end{cases}$$

The element g_{ij} of the matrix $G = AA^T$ is equal to the number of types of pottery common to both grave i and grave j . Then, create a map of the ages of the graves by starting with a great element from this matrix, and create connections depending on the elements. For example, consider the matrix

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2 \end{bmatrix}$$

Since $g_{12} = 1$, graves 1 and 2 share one type of pottery, so they are connected, and our current map is 1–2. Similarly, $g_{25} = 1$, so graves 2 and 5 are connected, and our map is now 1–2–5. Again, $g_{53} = 1$, so graves 5 and 3 are connected, and our map is now 1–2–5–3. Finally, $g_{34} = 1$, so graves 3 and 4 are connected, and our final map is now 1–2–5–3–4. This means that grave 1 is the oldest or that grave 4 is the oldest. In a similar method, the matrix $P = A^T A$ can be constructed, and its elements detail how many graves contain both the i th and j th types of pottery. The chronological order of the pottery can then also be determined using a map.

2.4

Let A be an $n \times n$ matrix. If a matrix B can be found such that $AB = BA = I_n$, then A is said to be invertible, and B is called the inverse of A . If A is invertible, then the unique inverse of A is denoted A^{-1} . A^{-k} is defined as $(A^{-1})^k$. The inverse of matrix can be found by solving the system of linear equations $AX = I_n$. For example, to determine the inverse of the matrix

$$A = \begin{bmatrix} 1 & -1 & -2 \\ 2 & -3 & -5 \\ -1 & 3 & 5 \end{bmatrix},$$

determine the reduced echelon form of the augmented matrix

$$[A : I_n] = \begin{bmatrix} 1 & -1 & -2 & 1 & 0 & 0 \\ 2 & -3 & -5 & 0 & 1 & 1 \\ -1 & 3 & 5 & 0 & 0 & 1 \end{bmatrix}.$$

There are several properties of the matrix inverse:

1. $(A^{-1})^{-1} = A$

2. $(cA)^{-1} = \frac{1}{c}A^{-1}$
3. $(AB)^{-1} = B^{-1}A^{-1}$
4. $(A^n)^{-1} = (A^{-1})^n$
5. $(A^T)^{-1} = (A^{-1})^T$

Let $AX = Y$ be a system of n linear equations in n variables. If A is invertible, then $X = A^{-1}Y$. This method is often not used, for it is inefficient.

An elementary matrix is one that can be obtained from the identity matrix by a single elementary row operation. There are three types of elementary matrices:

1. Interchange rows
2. Multiply a row by a nonzero constant
3. Add a multiple of one row to another row

These matrices can be used to apply elementary row operations to matrices through matrix multiplication. Each elementary matrix is square and invertible, and if A and B are row equivalent matrices (one can be obtained from the other by elementary row operations) and A is invertible, then B is also invertible. These matrices are used in the LU factorization of a matrix.

2.5

A transformation T of \mathbb{R}^n into \mathbb{R}^m is a rule that assigns to each vector \mathbf{u} in \mathbb{R}^n a unique vector \mathbf{v} in \mathbb{R}^m , and is denoted by $T(\mathbf{u}) = \mathbf{v}$. \mathbb{R}^n is the domain of T , and \mathbb{R}^m is the codomain of T . There are three main types of transformations:

1. Dilation: $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$
2. Reflection: $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$
3. Rotation: $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

Transformations can also be combined.

An orthogonal matrix A is an invertible matrix that has the property $A^{-1} = A^T$. An orthogonal transformation is one where $T(\mathbf{u}) = A\mathbf{u}$ for some orthogonal matrix A . Orthogonal transformations preserve norms, differences in angles, and distances (rigid motions).

A translation is a transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ of the form $T(\mathbf{u}) = \mathbf{u} + \mathbf{v}$, where \mathbf{v} is a fixed vector. An affine transformation is a transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ of the form $T(\mathbf{u}) = A\mathbf{u} + \mathbf{v}$.

2.6

Linear transformations imply that $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ and $T(c\mathbf{u}) = cT(\mathbf{u})$ for any vectors \mathbf{u} and \mathbf{v} and any scalar c . Affine transformations are not linear transformations. To prove that a transformation is linear, it is sufficient to show that the transformation preserves vector addition and scalar multiplication. Any linear transformation can be represented by its standard matrix a matrix whose columns are the images of the standard basis vectors:

$$A = [T(\mathbf{e}_1) \ \dots \ T(\mathbf{e}_n)]$$

To determine the transformation matrix then, apply the transformation to the standard basis vectors.

Typically, translations cannot be used in combination with linear transformations as a product. However if an extra dimension is added, the translation can be represented as a linear transformation. This is called an homogeneous coordinate system. For example, a point is represented as

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix},$$

and a translation is represented as

$$\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix},$$

where a and b are the translation amounts.

Fractals can be generated using multiple affine transformations given various probabilities. These parameters can be written as rows of a iterated function system matrix.

2.8

A stochastic matrix is a square matrix whose elements are probabilities and whose columns sum to 1. If A and B are stochastic matrices, then AB is also a stochastic matrix.

3

3.1

The determinant of a 2×2 matrix A is denoted by $|A|$ or $\det(A)$ and is given by

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

This determinant is given by the difference of the products of the two diagonals of the matrix.

Let A be a square matrix. The minor of the element a_{ij} is the determinant of the matrix obtained by removing the i th row and j th column from A . The cofactor of the element a_{ij} is denoted C_{ij} and is given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

Now, a new definition. The determinant of a square matrix is the sum of the products of the elements of the first row and their cofactors. For example, if A is a 3×3 matrix, then $|A| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$. More generally, the determinant of a square matrix is the sum of the products of the elements of any row or column and their cofactors.

$$\text{ith row expansion: } |A| = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The computation in evaluating a determinant can be minimized by expanding in terms of the row or column that contains the most 0's. The determinant of a 3×3 matrix can also be found using diagonals.

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$$

3.2

Let A be a square matrix and c be a nonzero scalar.

1. If a matrix B is obtained from A by multiplying a row or column by c , then $|B| = c|A|$.
2. If a matrix B is obtained from A by interchanging two rows or columns, then $|B| = -|A|$.
3. If a matrix B is obtained from A by adding a multiple of one row or column to another, then $|B| = |A|$.

A square matrix A is singular if $|A| = 0$, and A is nonsingular if $|A| \neq 0$. Let A be a square matrix. A is singular if

1. all the elements of a row or column are zero
2. two rows or columns are equal
3. two rows or columns are proportional

Let B be a matrix with the same size as A . AB being singular implies that either one of or both of A and B are singular, but the converse is not true.

Let A and B be $n \times n$ matrices and c be a nonzero scalar.

1. $|cA| = c^n|A|$

2. $|AB| = |A||B|$
3. $|A^T| = |A|$
4. $|A^{-1}| = \frac{1}{|A|}$

The determinant of a triangular matrix is equal to the product of the diagonal elements. When simplifying to find the determinant, if at any stage one of the diagonal elements is zero, and all elements below it are zero, then the determinant is zero.

3.3

The matrix whose elements are cofactors of A is called the matrix of cofactors of A . The transpose of this matrix is called the adjoint of A and is denoted $\text{adj}(A)$. Let A be a nonsingular square matrix.

$$A^{-1} = \frac{1}{|A|} \text{adj}(A)$$

Furthermore, A^{-1} exists if and only if $|A| \neq 0$.

Let $AX = B$ be a system of n linear equations in n variables. If A is nonsingular, then $X = A^{-1}B$, but if A is singular, then there are many or no solutions. Cramer's rule states that if A is nonsingular, then the system has a unique solution given by

$$x_1 = \frac{|A_1|}{|A|}, \quad \dots, \quad x_n = \frac{|A_n|}{|A|}$$

where A_i is the matrix obtained by replacing column i of A with B . Since a nonsingular matrix can only have one solution, a homogeneous system of linear equations can only have a solution apart from the trivial solution if the determinant of the matrix of coefficients is zero.

3.4

Let A be a $n \times n$ matrix. A scalar λ is called an eigenvalue of A if there exists a nonzero vector \mathbf{x} in \mathbb{R}^n such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

The vector \mathbf{x} is called an eigenvector of A corresponding to λ . Solving $|A - \lambda I_n| = 0$ for λ leads to all the eigenvalues of A . The expansion of $A - \lambda I_n$ gives the characteristic polynomial, a polynomial of degree n in λ that has λ as a root. Inputting the found λ back in to $(A - \lambda I_n)\mathbf{x}$ gives the corresponding eigenvectors (there are an infinite amount for each eigenvalue). The set of all eigenvectors corresponding to a given eigenvalue forms a subspace of \mathbb{R}^n called the eigenspace of A . The number of times an eigenvalue appears as a root of the characteristic polynomial is called the multiplicity of the eigenvalue.

In a triangular matrix, the eigenvalues are the diagonal elements of the matrix.

3.5

The transition matrix P of a Markov chain is said to be regular if for some power of P all the components are positive. The chain is then called a regular Markov chain. Define the limit of a regular Markov chain with transition matrix P as \mathbf{x} . Then, \mathbf{x} is an eigenvector of P corresponding to $\lambda = 1$, and the transition matrix also approaches a stochastic matrix with identical columns, each being an eigenvector of P corresponding to $\lambda = 1$.

Leslie matrices are a discrete, age-structured model of population growth. They also have eigenvectors that represent the long-term population distribution and eigenvalues that represent the long-term growth rate.

5.3

B is similar to A if there exists an invertible matrix C such that $B = C^{-1}AC$. Similar matrices have the same eigenvalues. Let A be an $n \times n$ matrix.

1. If A has n linearly independent eigenvectors, it is diagonalizable. The matrix C whose columns consist of n linearly independent eigenvectors can be used in a similarity transformation to give a diagonal matrix D . This diagonal elements of D will be the eigenvalues of A .
2. If A is diagonalizable, then it has n linearly independent eigenvectors.

Let

$$A = \begin{bmatrix} -4 & -6 \\ 3 & 5 \end{bmatrix}.$$

This matrix has eigenvalues 2 and -1, corresponding to eigenvectors $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$. The diagonal elements of the diagonalized matrix are the eigenvalues

$$D = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix},$$

and the column vectors of the diagonalizing matrix are linearly independent eigenvectors

$$C = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}.$$

The order of the eigenvalues in the diagonalized matrix D is the same as the order of the eigenvectors in the diagonalizing matrix C .

This makes computation easy because $A^k = CD^kC^{-1}$.

Let A be an $n \times n$ symmetric matrix.

1. If A is real, all of the eigenvalues are real, and it has n linearly independent eigenvectors.

2. The dimension of the eigenspace corresponding to an eigenvalue is equal to the multiplicity of the eigenvalue.
3. The eigenspaces of A are orthogonal.