Summary of Linear Algebra I

Martin Herschend

Contents

1 Matrices and vectors

1.1 Matrices

An $n \times m$ matrix A is a rectangular array of real numbers with n rows and m columns. It is usual to denote the entry of A in row i and column j by a_{ij} so that

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}.$$

The identity matrix of size n is the $n \times n$ matrix

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

A matrix with all entries equal to zero is called a zero matrix.

1.2 Matrix operations

- The sum of two $n \times m$ matrices is defined entrywise:

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1m}+b_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & \cdots & a_{nm}+b_{nm} \end{bmatrix}$$

- Let $k \in \mathbb{R}$. To multiply a matrix by the scalar k we multiply each entry of the matrix by k:

$$k \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} ka_{11} & \cdots & ka_{1m} \\ \vdots & \ddots & \vdots \\ ka_{n1} & \cdots & ka_{nm} \end{bmatrix}$$

- Let A be an $n \times m$ matrix and B an $m \times l$ matrix, i.e.,

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \text{ and } B = \begin{bmatrix} b_{11} & \cdots & b_{1l} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{ml} \end{bmatrix}.$$

  Then the product of A and B is the $n \times l$ matrix AB whose entry in row i and column j is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots a_{im}b_{mj}.$$

  Notice that the product of A and B is only defined when the number of columns in A is the same as the number of rows in B.

- The transpose of an $n \times m$ matrix

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

  is the $m \times n$ matrix A^T whose entry in row i and column j is a_{ji}, i.e.,

$$A^T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix}$$

Addition of matrices and multiplication of a matrix by a scalar satisfies the following equalities:

- $(A+B)+C = A+(B+C)$
- $A+B = B+A$

- $k(A+B) = kA + kB$
- $(k+l)A = kA + lA$
- $k(lA) = (kl)A$

Matrix multiplication satisfies the following equalities (here I denotes the identity matrix of appropriate size):

- $A(BC) = (AB)C$
- $AI = A$
- $IA = A$
- $A(B+C) = AB + AC$
- $(A+B)C = AC + BC$
- $(kA)B = A(kB) = k(AB)$

The transpose has the following properties

- $(A^T)^T = A$
- $(A+B)^T = A^T + B^T$
- $(kA)^T = k(A^T)$
- $(AB)^T = B^T A^T$

1.3 Vectors

A (column) vector is matrix consisting of a single column

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Its entries v_i are called the components of $\vec{v}$. The set of all vectors with n components is denoted $\mathbb{R}^n$. Addition of vectors, multiplication of a vector by a scalar and multiplication of a vector by a matrix is defined as for matrices.

We interpret the vector $\vec{v}$ geometrically as the arrow pointing from the origin to the point with coordinates $(v_1, v_2, \ldots, v_n)$ in n-dimensional space. Adding two vectors to each other can be visualized as putting one vector after the other. Multiplying a vector $\vec{v}$ with a real number k can be visualized as scaling $\vec{v}$ by k.

1.4 Dot product

The dot product of two vectors

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \text{ and } \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

is the real number

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

In other words it is the single entry in the matrix $\vec{v}^T \vec{w}$.

- The length of a vector $\vec{v}$ is the real number

$$||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}}.$$

- Geometrically we interpret the dot product by

$$\vec{v} \cdot \vec{w} = ||\vec{v}|| \cos\theta ||\vec{w}||,$$

  where θ denotes the angle between $\vec{v}$ and $\vec{w}$.

- We say that two vectors $\vec{v}$ and $\vec{w}$ are orthogonal if

$$\vec{v} \cdot \vec{w} = 0,$$

  which geometrically means that they are at right angles to each other.

2 Systems of linear equations

2.1 Linear equations

A linear equation in the variables $x_1, \ldots, x_n$ is an equation of the form

$$a_1 x_1 + \cdots a_n x_n = b \tag{1}$$

where $a_1, \ldots, a_n$ and b are some real numbers. The numbers $a_1, \ldots, a_n$ are called coefficients and b is called the right hand side of the equation.

We can interpret Equation (1) as

$$\vec{a} \cdot \vec{x} = b$$

where

$$\vec{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \text{ and } \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Thus the solutions to Equation (1) are all vectors $\vec{x}$ satisfying that the dot product of $\vec{a}$ and $\vec{x}$ equals b.

Assume that $\vec{a} \neq 0$. If $b = 0$, then Equation (1) can be interpreted geometrically as follows.

- In two dimensions ($n = 2$) Equation (1) describes a line through the origin. It consists of all vectors that are orthogonal to $\vec{a}$.
- In three dimensions ($n = 3$) Equation (1) describes a plane through the origin. Again it consists of all vectors that are orthogonal to $\vec{a}$.

If $b \neq 0$, then Equation (1) describes a similar line or plane but translated away from the origin. The vector $\vec{a}$ is called a normal vector of the line or plane.

2.2 Systems of linear equations

A system of linear equations is a collection of linear equations with common variables:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1m}x_m = b_1, \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m = b_n. \end{cases} \tag{2}$$

It is equivalent to the matrix equation

$$A\vec{x} = \vec{b}$$

where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}, \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

The matrices A and $\vec{b}$ are respectively called the coefficient matrix and the right hand side of System 2. Combined they form the augmented matrix

$$\left[A \mid \vec{b}\right].$$

In two or three dimensions ($m = 2$ or $m = 3$), the equations of System (2) describe either lines or planes. Therefore its solutions describe the intersection of these lines or planes.

2.3 Row operations

There are three elementary row operations that can be performed on a matrix:

- Multiply a row by a non-zero number k.
- Add a row times a number k to another row.
- Swap two rows.

For two matrices A and A' of the same size, we say that they are row equivalent if A' can be obtained from A by a sequence of elementary row operations. In that case we write $A \sim A'$

Performing a row operation on the augmented matrix $\left[A \mid \vec{b}\right]$ of a system of linear equations does not change the solutions of the system. We conclude that if

$$\left[A \mid \vec{b}\right] \sim \left[A' \mid \vec{b}'\right],$$

then the systems corresponding to $\left[A \mid \vec{b}\right]$ and $\left[A' \mid \vec{b}'\right]$ have the same solutions.

2.4 Reduced row-echelon form

The first non-zero entry in a non-zero row in a matrix is called a leading entry. A matrix is said to be in reduced row-echelon form (rref) if

- All leading entries are equal to 1.
- If a column contains a leading entry, then all other entries in that column are zero.
- If a row has a leading entry, then each row above it has a leading entry further to the left.

2.5 Gauss-Jordan elimination

Gauss-Jordan elimination is an algorithm that starting from any matrix, produces a row equivalent matrix in reduced row-echelon form. The algorithm operates row by row from top to bottom. At row i it takes the following steps.

- If row i and all rows below it are zero, then the algorithm stops. Now consider the case when at least one of these rows is non-zero.
- If row i does not have leading entry furthest to the left among these rows then swap it for the first row that does.
- Divide row i by its leading entry so that it turns into a 1.
- Eliminate all entries in the same column as the leading entry of row i.

- If i is the last row the algorithm stops. Otherwise proceed to row $i+1$.

The result of applying Gauss-Jordan elimination to an $n \times m$ matrix A is denoted $\operatorname{rref}(A)$. It is the unique matrix in reduced row-echelon form that is row equivalent to A. The rank of A is defined to be the number of leading entries in $\operatorname{rref}(A)$ and denoted by $\operatorname{rank} A$. In particular $\operatorname{rank} A \leq m$ and $\operatorname{rank} A \leq n$.

2.6 Solutions to systems of linear equations

Consider the augmented matrix $\left[A \mid \vec{b}\right]$ of a system of n linear equations in m variables. By applying Gauss-Jordan elimination we may assume that A is in reduced row-echelon form. Then the solutions of the system can be found in the following way:

- The variables corresponding to columns with leading entries are called leading variables.
- If there is a zero row in A and the corresponding component of $\vec{b}$ is non-zero, then the system has no solutions.
- Otherwise solve each non-zero equation for its leading variable. The non-leading variables can be chosen freely, i.e., they may be considers as parameters.

The rank of the coefficient matrix A is related to the number of solutions in the following way:

- If $\operatorname{rank} A = n$, then there is at least one solution, since each equation has a leading variable.
- If $\operatorname{rank} A = m$, then there is at most one solution, since each variable is leading.
- If $\operatorname{rank} A < m$, then there is either no solution or infinitely many solutions, since there is at least one non-leading variable.

In two or three dimensions the solutions to a system of linear equations forms the intersection of the corresponding lines or planes. Considering the number of parameters in the solution we have the following interpretation.

- If there is no solution, then the lines/planes do not intersect.
- If there is a unique solution, then the intersection is a point.
- If the solution has one parameter, then the intersection is a line.
- If the solution has two parameters, then the intersection is a plane.

3 Linear Transformations

A function $T : \mathbb{R}^m \to \mathbb{R}^n$ is called a linear transformation if

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$$
$$T(k\vec{x}) = kT(\vec{x})$$

for all vectors $\vec{x}, \vec{y} \in \mathbb{R}^m$ and $k \in \mathbb{R}$. This is equivalent to that there is an $n \times m$ matrix A such that

$$T(\vec{x}) = A\vec{x}.$$

In that case the matrix A is uniquely determined by T. In fact

$$A = \begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_m) \\ | & & | \end{bmatrix}$$

where

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \vec{e}_m = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

We call A the matrix of T.

3.1 Examples of linear transformations

- The identity function $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(\vec{x}) = \vec{x}$ is linear and its matrix is the identity matrix I_n.
- Let $k \in \mathbb{R}$. The function $T : \mathbb{R}^n \to \mathbb{R}^n$, that scales each vector by k, i.e., $T(\vec{x}) = k\vec{x}$ is linear and its matrix is kI_n.
- Let $\alpha \in \mathbb{R}$. The function $T : \mathbb{R}^2 \to \mathbb{R}^2$, that rotates each vector in the plane by the angle α (counterclockwise) is linear. Its matrix is

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.$$

- Let $n = 2$ or $n = 3$, and ℓ be a line through the origin with direction vector $\vec{v} \in \mathbb{R}^n$. Any vector $\vec{x}$ can be written uniquely as $\vec{x} = \vec{x}^{||} + \vec{x}^{\perp}$ where $\vec{x}^{||}$ is parallel to ℓ and $\vec{x}^{\perp}$ is orthogonal to ℓ. The vector $\vec{x}^{||}$ is called the orthogonal projection of $\vec{x}$ onto ℓ. The function $\mathrm{proj}_\ell : \mathbb{R}^n \to \mathbb{R}^n$, defined by $\mathrm{proj}_\ell(\vec{x}) = \vec{x}^{||}$ is linear and satisfies

$$\mathrm{proj}_\ell(\vec{x}) = \left(\frac{\vec{x} \cdot \vec{v}}{\vec{v} \cdot \vec{v}}\right)\vec{v}.$$

- The reflection of $\vec{x} \in \mathbb{R}^n$ in the line ℓ is denoted $\mathrm{ref}_\ell(\vec{x})$. It defines a linear transformation and satisfies

$$\mathrm{ref}_\ell(\vec{x}) = 2\,\mathrm{proj}_\ell(\vec{x}) - \vec{x}.$$

- Let $n = 3$ and V be a plane through the origin with normal vector $\vec{n}$. Consider the line ℓ through the origin in direction $\vec{n}$. The orthogonal projection of $\vec{x} \in \mathbb{R}^n$ onto V is denoted $\mathrm{proj}_V(\vec{x})$. It defines a linear transformation and satisfies

$$\mathrm{proj}_V(\vec{x}) = \vec{x} - \mathrm{proj}_\ell(\vec{x}).$$

  The reflection of $\vec{x} \in \mathbb{R}^n$ in V is denoted $\mathrm{ref}_V(\vec{x})$. It defines a linear transformation and satisfies

$$\mathrm{ref}_V(\vec{x}) = \vec{x} - 2\,\mathrm{proj}_\ell(\vec{x}).$$

3.2 Composition of linear transformations

If $S : \mathbb{R}^l \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^n$ are linear transformations, then the composition

$$T \circ S : \mathbb{R}^l \to \mathbb{R}^n,\ (T \circ S)(\vec{x}) = T(S(\vec{x}))$$

is a linear transformation.

If A is the matrix of T and B is the matrix of S, then the matrix of $T \circ S$ is AB.

3.3 The inverse of a linear transformation

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. If $m \neq n$, then T is not invertible.

Now assume $m = n$, so that $T : \mathbb{R}^n \to \mathbb{R}^n$. If T is invertible, then the inverse T^{-1} is a linear transformation.

We say that an $n \times n$ matrix A is invertible if the linear map $T : \mathbb{R}^n \to \mathbb{R}^n$, $T(\vec{x}) = A\vec{x}$ is invertible. In that case we define the inverse A^{-1} of A to be the matrix of T^{-1}.

The following conditions are equivalent for an $n \times n$ matrix A

- A is invertible.
- The matrix equation $A\vec{x} = \vec{b}$ has a unique solution for each $\vec{b} \in \mathbb{R}^n$.
- $\operatorname{rref}(A) = I_n$.
- $\operatorname{rank} A = n$.
- There is an $n \times n$ matrix B such that $AB = I_n$.
- There is an $n \times n$ matrix B such that $BA = I_n$.

In the last two conditions it follows that $B = A^{-1}$.

Assume that A and B are invertible $n \times n$-matrices. Then the following identities hold

- $A^{-1}A = I_n = AA^{-1}$
- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$

3.3.1 Row transformations and invertible matrices

Let A be an $n \times m$ matrix. Then the following conditions are equivalent:

- The matrix A is row equivalent to A'.
- There is an invertible $n \times n$ matrix C such that $A' = CA$.

3.3.2 Finding the inverse of an invertible matrix

Let A be an $n \times n$ matrix. To find the inverse of A (if it exists) one can take the following steps.

- Apply Gauss-Jordan elimination to find $\operatorname{rref}\big([A \mid I_n]\big)$.
- If $\operatorname{rref}\big([A \mid I_n]\big) = \big[I_n \mid B\big]$ for some B, then A is invertible and $A^{-1} = B$.
- If $\operatorname{rref}\big([A \mid I_n]\big)$ is not of the above form, then $\operatorname{rank}(A) < n$ and so A is not invertible.

4 Determinants

Let A be an $n \times n$ matrix. The determinant of A is a real number denoted $\det A$. It is defined in several steps:

- A pattern P_A in A is a choice of n entries in A such that exactly one entry is chosen from each row and column.
- The product of the entries in a pattern P_A is denoted $\operatorname{prod} P_A$.
- Two entries in a pattern P_A form an inversion if one is above and to the right of the other in the matrix A.

- The sign of a pattern P_A is the number
$$\operatorname{sign} P_A = (-1)^N,$$
where N is the number of inversions in P_A.
- The determinant of A is the sum with terms
$$(\operatorname{sign} P_A)(\operatorname{prod} P_A),$$
where P_A varies over all patterns in A.

4.1 2×2 and 3×3 determinants

From the definition one finds that
$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$
and
$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

4.2 Properties of the determinant

Let A be an $n \times n$ matrix. The determinant has the following properties:

- The determinant does not change when taking the transpose:
$$\det(A^T) = \det A$$
- The determinant is linear in every column. More precisely, fix $\vec{v}_1, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n \in \mathbb{R}^n$. Then the function $T : \mathbb{R}^n \to \mathbb{R}$ defined by
$$T(\vec{x}) = \det \begin{bmatrix} | & & | & | & | & & | \\ \vec{v}_1 & \cdots & \vec{v}_{i-1} & \vec{x} & \vec{v}_{i+1} & \cdots & \vec{v}_n \\ | & & | & | & | & & | \end{bmatrix}$$
is a linear transformation.
- Similarly, the determinant is linear in every row.
- If A' is obtained from A by swapping two rows, then
$$\det A' = -\det A.$$
- If A' is obtained from A by multiplying a row by a constant k, then
$$\det A' = k \det A.$$
- If A' is obtained from A by adding a row times a constant to another row, then
$$\det A' = \det A.$$
- If A has a zero row or two equal rows, then $\det A = 0$.
- The matrix A is invertible if and only if $\det A \neq 0$. In that case
$$\det(A^{-1}) = \frac{1}{\det A} = (\det A)^{-1}.$$
- The determinant is compatible with matrix multiplication:
$$\det(AB) = (\det A)(\det B).$$
- If all entries of A below the diagonal are zero then $\det A$ is the product of the diagonal elements of A. The same holds if all entries of A above the diagonal are zero.

4.3 Laplace expansion

Let A be an $n \times n$ matrix. The $(n-1) \times (n-1)$ matrix obtained by omitting row i and column j in A is denoted A_{ij}. The determinant of A_{ij} is called a minor of A. Minors can be used to calculate $\det A$ as in the following formulas.

- Expansion along column j:

$$\det A = (-1)^{1+j} a_{1j} \det A_{1j} + \cdots + (-1)^{n+j} a_{nj} \det A_{nj} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

- Expansion along row i:

$$\det A = (-1)^{i+1} a_{i1} \det A_{i1} + \cdots + (-1)^{i+n} a_{in} \det A_{in} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

This is particularly useful when the chosen column or row has only one non-zero entry as all terms except for one in the sum vanish.

4.4 Cramer's rule

Let A be an invertible $n \times n$ matrix and $\vec{b} \in \mathbb{R}^n$. Let $A_{\vec{b},i}$ be the matrix obtained from A by replacing column i by $\vec{b}$. Then the matrix equation

$$A\vec{x} = \vec{b}$$

has a unique solution

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

where

$$x_i = \frac{\det(A_{\vec{b},i})}{\det A}.$$

4.5 The adjoint of a matrix

Let A be an $n \times n$ matrix. The classical adjoint of A is the $n \times n$ matrix $\operatorname{adj}(A)$ whose entry in row i and column j is

$$(-1)^{i+j} \det(A_{ji}).$$

It satisfies the formulas

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = (\det A) I_n$$

In particular, if A is invertible, then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj}(A).$$

4.6 Geometric interpretation of the determinant

In two and three dimensions the determinant can be interpreted as an area and a volume respectively.

- Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^2$ and set

$$A = \begin{bmatrix} | & | \\ \vec{v} & \vec{w} \\ | & | \end{bmatrix}.$$

  Then the area of the parallelogram spanned by $\vec{v}$ and $\vec{w}$ is $|\det A|$,

- Let $\vec{u}$, $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^3$ and set

$$A = \begin{bmatrix} | & | & | \\ \vec{u} & \vec{v} & \vec{w} \\ | & | & | \end{bmatrix}.$$

Then the volume of the parallelepiped spanned by $\vec{u}$, $\vec{v}$ and $\vec{w}$ is $|\det A|$.