Notes for Linear algebra

Sophie Marques

Sunday 6th March, 2016

Contents

Chapter 1

Linear equation in Linear algebra

1.1 System of linear equations

1.1.1 Definitions and notations

Definition 1.1.1. *A **linear equation** in the variable $x_1, \cdots, x_n$ is an equation that can be written in the form*

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where b and the coefficients $a_1, \cdots, a_n$ are real or complex numbers, usually known in advance. The subscript n may be any positive integer.
Please note that the name of the variable is not relevant. This is just a generic way to write an equation.

Remarque 1.1.2. *For instance, the equation $2x = 3$ is linear, $2x + 3y = 6$ is linear, $2x_1 + (2 + \sqrt{3})(x_2 + \sqrt{4}) = \sqrt{2}$ is linear because one can write as in the definition. But, be careful, $2\sqrt{x_1} + 7x_2 = 8$ is non-linear, $8xy = 4$ and $3x_1x_2 + x_3 = x_4$ is not linear.*

Definition 1.1.3. *A **system of linear equation** (or **linear system**) is a collection of one or more linear equation involving the same variables, say $x_1, \cdots, x_n$.*

*A **solution** of the system is a list $(s_1, s_2, \cdots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \cdots, s_n$ are subsituted for $x_1, \cdots, x_n$, respectively.*

*The set of all possible solutions is called the **solution set** of the linear system.*

*Two linear systems are called **equivalent** if they have the same solution set. A system of linear equations is said to be* **consistent** *if it has either one solution or infinitely many solutions; a system is* **inconsistent** *if has no solution.*

Theorem 1.1.4. *A system of linear equation has either*

1. *no solution, or*
2. *exactly one solution or*
3. *infinitely many solutions.*

Proof. Later. □

Example 1.1.5.

$$\begin{cases} 3x_1 + 2x_2 & = & 7 \\ x_1 + 4x_3 & = & 9 \end{cases}$$

is a system of linear equations. $x_1 = 1$*,* $x_2 = 2$ *and* $x_3 = 2$ *is a solution of this system of equation.*

The graphs of linear equations in two variables are lines in the planes. Geometrically the solution of a system of linear equations in two variables are the intersection point if any between these lines.

Example 1.1.6. *1. The system*

$$\begin{cases} 2x_1 - 4x_2 & = & -2 \\ -3x_1 + 6x_2 & = & 9 \end{cases}$$

as no solutions. The corresponding lines are disjoint right parallel.

2. *The system*

$$\begin{cases} x_1 + 5x_2 & = & 9 \\ 2x_1 + 10x_2 & = & 18 \end{cases}$$

has infinitely many solutions. The corresponding lines are the mingled.

3. *The system*

$$\begin{cases} 2x_1 - 4x_2 & = & -2 \\ x_1 - 3x_2 & = & -3 \end{cases}$$

has exactly one solution $x_1 = 3$ *and* $x_2 = 2$*. The corresponding lines intersect in exactly one point with coordinate* $(3, 2)$*.*

1.1.2 From system of equation to matrices

Definition 1.1.7. *A linear system can be compactly recorded in a rectangle array called a* **matrix** *with the coefficient of each variable aligned in columns. Given the system*

$$\begin{cases} a_{1,1}x_1 + \cdots + a_{1,n}x_n & = & b_1 \\ \cdots & & \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n & = & b_m \end{cases}$$

The matrix

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \cdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

is called the **coefficient matrix** *(or* **matrix of coefficients***) of the system and the matrix*

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} & b_1 \\ & \cdots & & \\ a_{m,1} & \cdots & a_{m,n} & b_m \end{pmatrix}$$

is called the **augmented matrix** *of the system. The* **size** *of the matrix tells how many rows and columns it has. For instance, the coefficient matrix is an* $m \times n$ *is a rectangular array of numbers with* m *rows and* n *columns.*

The matrix notation might simplify the calculations as we will see.

1.1.3 Solving a linear system

We will describe an algorithm for solving linear system. The basic strategy is to replace one system with an equivalent system (i.e. one with the same solution set) that is easier to solve.
For this observe the following system

$$\begin{cases} 2x_1 & +4x_2 & +5x_3 & = & 2 \\ & x_2 & -3x_3 & = & -3 \\ & & 2x_3 & = & 4 \end{cases}$$

Note that the matrix coefficient corresponding has a special form

$$\begin{pmatrix} 2 & 4 & 5 \\ 0 & 1 & -3 \\ 0 & 0 & 2 \end{pmatrix}$$

This is called an **upper triangular matrix**. Note also that it is very easy to solve a system of this form, indeed the last equation gives us $x_3 = 2$ substituting x_3 in the second equation we get $x_2 = 3$ and substituting x_3 and x_2 in the first equation, we get $x_1 = 2$. So if there was a way to transform any system in an equivalent upper triangular one, then it would be easy to solve the system of equations.
For this let us first describe the operation on the linear equation allowed such that the system obtained is equivalent and leading to such form after reiterating them.

Theorem 1.1.8. *We obtain an equivalent system (i.e. same solution set), if we*

1. *(INTERCHANGE) Interchange two linear equations, we write $R_i \leftrightarrow R_j$;*
2. *(REPLACEMENT) Replace one equation by the sum of itself and a multiple of another equation, we write $R_i \leftarrow R_i + \lambda R_j$,*
3. *(SCALING) Multiply an equation by a nonzero constant, we write $R_i \leftarrow \lambda R_i$, with $\lambda \neq 0$*

where R_i denote the different linear equations and λ is a constant.

Proof. It is clear that if we have a set solution for a system of linear equation, this solution set stay unchanged when we swap two of these linear equation, by scaling or by replacement. To be convince of it, take a solution of the system before making the operation and double check this solution is still a solution of the new system and vice versa. □

Note that if one consider the augmented matrix corresponding we are then doing operations on the rows. Note also that the row operation are reversible. Indeed, if you swap twice two rows you get the same system as in the beginning. If you replace $R_i \leftarrow R_i + \lambda R_j$ and then you replace $R_i \leftarrow R_i - \lambda R_j$ you also go back to the initial system. Finally, if you scale by λ, $R_i \leftarrow \lambda R_i$ and then you scale by $1/\lambda$, $R_i \leftarrow 1/\lambda R_i$ you also come back to the initial system.

The goal is then to find a succession of such a row operation so that we get a upper triangular linear system. For this after swapping two rows if necessary, we can use the x_1 of the first row to eliminate the x_1 in the second row, then use the x_2 in the second row to eliminate the x_2 in the third row etc...

Example 1.1.9. *Let's solve the system:*

$$\left\{\begin{array}{rrrcrl} & x_2 & +5x_3 & = & -4 & R_1 \\ x_1 & +4x_2 & +3x_3 & = & -2 & R_2 \\ 2x_1 & +7x_2 & +x_3 & = & 4 & R_3 \end{array}\right.$$

We will work with and without the matrix notation.
Solution: *The augmented matrix corresponding to the system is*

$$\begin{pmatrix} 0 & 1 & 5 & -4 \\ 1 & 4 & 3 & -2 \\ 2 & 7 & 1 & 4 \end{pmatrix}$$

We can exchange $R_1 \leftrightarrow R_2$ so that we get the equivalent system

$$\left\{\begin{array}{rrrcrl} x_1 & +4x_2 & +3x_3 & = & -2 & R_1 \\ & x_2 & +5x_3 & = & -4 & R_2 \\ 2x_1 & +7x_2 & +x_3 & = & 4 & R_3 \end{array}\right.$$

The augmented matrix corresponding to the system is now

$$\begin{pmatrix} 1 & 4 & 3 & -2 \\ 0 & 1 & 5 & -4 \\ 2 & 7 & 1 & 4 \end{pmatrix}$$

We can replace $R_3 \leftarrow R_3 - 2R_1$ so that we get the equivalent system

$$\left\{\begin{array}{rrrcrl} x_1 & +4x_2 & +3x_3 & = & -2 & R_1 \\ & x_2 & +5x_3 & = & -4 & R_2 \\ & -x_2 & -5x_3 & = & 8 & R_3 \end{array}\right.$$

The augmented matrix corresponding to the system is now

$$\begin{pmatrix} 1 & 4 & 3 & -2 \\ 0 & 1 & 5 & -4 \\ 0 & -1 & -5 & 8 \end{pmatrix}$$

We can replace $R_3 \leftarrow R_3 + R_2$ so that we get the equivalent system

$$\left\{\begin{array}{rrrcrl} x_1 & +4x_2 & +3x_3 & = & -2 & R_1 \\ & x_2 & +5x_3 & = & -4 & R_2 \\ & & 0 & = & 4 & R_3 \end{array}\right.$$

The augmented matrix corresponding to the system is now

$$\begin{pmatrix} 1 & 4 & 3 & -2 \\ 0 & 1 & 5 & -4 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

The system is inconsistent, because the last row would require that $0 = 4$ *if there were a solution. The solution set is empty.*

1.1.4 Existence and uniqueness question

When solving a linear system, you need to determine first if the system is consistent not, that is there any possible solution? If there are is this solution unique? And you need to give the full set of solutions.
Let's see other examples, we just that a system could be inconsistent.

Example 1.1.10. *Let's solve the system:*

$$\begin{cases} 2x_1 & & & -4x_4 & = & -10 & R_1 \\ & 3x_2 & +3x_3 & & = & 0 & R_2 \\ & & x_3 & +4x_4 & = & -1 & R_3 \\ -3x_1 & +2x_2 & +3x_3 & +x_4 & = & 5 & R_4 \end{cases}$$

Solution: *We scale* $R_1 \leftarrow 1/2R_1$ *and* $R_2 \leftarrow 1/3R_2$,

$$\begin{cases} x_1 & & & -2x_4 & = & -5 & R_1 \\ & x_2 & +x_3 & & = & 0 & R_2 \\ & & x_3 & +4x_4 & = & -1 & R_3 \\ -3x_1 & +2x_2 & +3x_3 & +x_4 & = & 5 & R_4 \end{cases}$$

We replace $R_4 \leftarrow R_4 + 3R_1$,

$$\begin{cases} x_1 & & & -2x_4 & = & -5 & R_1 \\ & x_2 & +x_3 & & = & 0 & R_2 \\ & & x_3 & +4x_4 & = & -1 & R_3 \\ & 2x_2 & +3x_3 & -5x_4 & = & -10 & R_4 \end{cases}$$

We replace $R_4 \leftarrow R_4 - 2R_2$,

$$\begin{cases} x_1 & & & -2x_4 & = & -5 & R_1 \\ & x_2 & +x_3 & & = & 0 & R_2 \\ & & x_3 & +4x_4 & = & -1 & R_3 \\ & & x_3 & -5x_4 & = & -10 & R_4 \end{cases}$$

We replace $R_4 \leftarrow R_4 - R_3$,

$$\begin{cases} x_1 & & & -2x_4 & = & -5 & R_1 \\ & x_2 & +x_3 & & = & 0 & R_2 \\ & & x_3 & +4x_4 & = & -1 & R_3 \\ & & & -9x_4 & = & -9 & R_4 \end{cases}$$

We then find that the system is consistent and has exactly one solution. From R_4, *we get* $x_4 = 1$, *then* R_3 *gives* $x_3 = -5$, R_2 *gives* $x_2 = 5$, *and* R_1 *gives* $x_1 = -3$. *So the unique solution is* $(-3, 5, -5, 1)$.

Example 1.1.11. *Let's solve the system:*

$$\begin{cases} 2x_1 & & -4x_3 & = & -10 & R_1 \\ & 3x_2 & & = & 0 & R_2 \end{cases}$$

Solution: *From R_2, $x_2 = 0$ and from R_1, we get $x_1 = 2x_3 - 5$. So the system is consistent and we have an infinite amount of solutions. The solution set is*

$$\{(2t-5, 0, t), t \in \mathbb{R}\}$$

setting $t = x_3$.

Exercise 1.1.12. *Is $(1, -4)$ a solution of the following linear system*

$$\begin{cases} 4x_1 - x_2 & = & -5 \\ x_1 + 5x_2 & = & 0 \end{cases}$$

Solution *No, since $1 + 5 \times (-4) \neq 0$.*

Exercise 1.1.13. *For what values of h is the following system consistent:*

$$\begin{cases} x_1 + hx_2 & = & 4 & R_1 \\ 3x_1 + 6x_2 & = & 8 & R_2 \end{cases}$$

Solution *We replace $R_2 \leftarrow R_2 - 3R_1$*

$$\begin{cases} x_1 + hx_2 & = & 4 & R_1 \\ (6-3h)x_2 & = & -4 & R_2 \end{cases}$$

If $6 - 3h = 0$, that is $h = 2$ then R_2 becomes impossible $0 = -4$, so the system is inconsistent. When $h \neq 2$, then $6h - 3 \neq 0$, we get $x_2 = -4/(6-3h)$ and $x_1 = 4 + 4h/(6-3h)$, so the system is consistent and the solution is unique. The solution is $(-4/(6-3h), (24-8h)/(6-3h))$.

1.2 Row reduction and echelon forms

Definition 1.2.1.
1. *A rectangular matrix is in echelon form (or* **row echelon form***) if it has the following properties:*
   (a) *All nonzero rows are above any rows of all zeros*
   (b) *Each leading entry of a row is in a column to the right of the leading entry of the row above it.*
   (c) *All entries in a column below a leading entry are zeros. (Note that it follows from (b))*
2. *If a matrix in echelon form satisfies the following additional condition, then it is in* **reduced echelon form** *(or* **reduced row echelon form***):*
   (a) *The leading entry in each nonzero row is 1;*
   (b) *Each leading 1 is the only nonzero entry in its column.*

3. *An* **echelon matrix** *(respectively,* **reduced echelon matrix***) is one that is in echelon form (respectively, reduced echelon form).*
4. *A matrix A is* **row equivalent** *to an other matrix B if there is a sequence of row operations (switch, scaling, replacement) transforming A into B. If a matrix A is row equivalent to an echelon matrix U, we call U* **an echelon form (or row echelon form) of** *A; if U is in reduced echelon form, we call U* **the reduced echelon form of** *A.*

Any nonzero matrix may be row reduced (that is, transformed by elementary row operations) into more than one matrix in echelon form, using a different sequences of row operations (Computer algorithm can put a matrix in echelon form). However, the reduced echelon form one obtains from a matrix is unique.

Example 1.2.2. *1. The matrix*

$$\begin{pmatrix} 1 & 2 & 3 & 7 \\ 0 & 8 & 11 & 23 \\ 0 & 0 & 65 & 45 \end{pmatrix}$$

is in echelon form. Also,

$$\begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 23 \\ 0 & 0 & 1 & 450 \end{pmatrix}$$

is in echelon form and even more it is in reduced echelon form.

2. *The following matrices are in echelon form. The leading entries ($\square$) may have any nonzero value, the starred entries ($*$) may have any value (including zero). For instance,*

$$\begin{pmatrix} \square & * & * & * \\ 0 & \square & * & * \\ 0 & 0 & \square & * \end{pmatrix}$$

or

$$\begin{pmatrix} 0 & \square & * & * & * & * & * & * \\ 0 & 0 & 0 & \square & * & * & * & * \\ 0 & 0 & 0 & 0 & \square & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \square & * \end{pmatrix}$$

The following matrices are in reduced echelon form because the leading entries are $1'$s, and there are $0'$s bellow and above each leading 1. For instance,

$$\begin{pmatrix} 1 & 0 & 0 & * \\ 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \end{pmatrix}$$

or

$$\begin{pmatrix} 0 & 1 & * & 0 & 0 & * & 0 & * \\ 0 & 0 & 0 & 1 & 0 & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{pmatrix}$$

Theorem 1.2.3 (Uniqueness of the reduced echelon form)**.** *Each matrix is row equivalent to one and only one reduced matrix.*

Proof. See section 4.3. □

Definition 1.2.4. *A* **pivot position** *in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A* **pivot column** *is a column of A that contains a pivot position. A* **pivot** *is a nonzero number in a pivot position that is used as needed to create zeros via rows operations.*

Remarque 1.2.5. *Once we get the echelon reduced equivalent matrix, then further operation does not change the positions of the leading entries. Again, the reduced form is unique, in particular the leading entries corresponding to the leading $1's$ in the reduced echelon form.*

Here, how to proceed if you want to row reduced a matrix. Repeat these steps as much as necessary until you get a matrix in echelon form. This will happen in a finite number of steps.

1. Find the first nonzero column starting in the left. This is the pivot column. The pivot position is at the top.
2. Choose a nonzero entry in the pivots column (if possible: 1, or one which will make computations easier, practice will tell you this). If necessary, interchange rows to move this entry into the pivot position.
3. Use the row at the pivot position, to create zeros in all positions below the pivot, until you get only zeros above the pivot position.
4. Do the same process with the "submatrix" obtained if you remove the pivot row and the pivot column.
5. You are done when you get an echelon matrix.
6. In order to create a reduced echelon form, scale your rows in order to obtain 1's in all the pivots positions and create 0 above the pivot thanks to row replacement using the pivot from the bottom to the top.

Exercise 1.2.6. *Row reduce the following matrix.*

$$\begin{pmatrix} 1 & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix}$$

Solution:

$$\begin{array}{c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} \\ \begin{pmatrix} \boxed{1} & 2 & 4 & 5 \\ 2 & 4 & 5 & 4 \\ 4 & 5 & 4 & 2 \end{pmatrix} \begin{array}{l} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

C_1 is the pivot column. The pivot number is the 1. boxed One creates zero above the pivot by doing the row operation: $R_3 \leftarrow R_3 - 4R_1$ and $R_2 \leftarrow R_2 - 2R_1$. One obtains the equivalent matrix:

$$\begin{array}{c c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & 2 & 4 & 5 \\ 0 & 0 & -3 & -6 \\ 0 & -3 & -12 & -18 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

Now the second pivot column is C_2. We exchange $C_2 \leftrightarrow C_3$ in order to have the nonzero pivot in the top.

$$\begin{array}{c c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & 2 & 4 & 5 \\ 0 & \boxed{-3} & -12 & -18 \\ 0 & 0 & \boxed{-3} & -6 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

We notice we are fine, since we have an echelon matrix. But since we want a reduced one, we scale $R_2 \leftarrow -1/3R_2$ and $R_3 \leftarrow -1/3R_3$.

$$\begin{array}{c c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & 2 & 4 & 5 \\ 0 & \boxed{1} & 4 & 6 \\ 0 & 0 & \boxed{1} & 2 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

We create 0 above the pivot in the pivot column C_3 with $R_1 \leftarrow R_1 - 4R_3$ and $R_2 \leftarrow R_2 - 4R_3$

$$\begin{array}{c c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & 2 & 0 & -3 \\ 0 & \boxed{1} & 0 & -2 \\ 0 & 0 & \boxed{1} & 2 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

Now we create 0 above the pivot in the pivot column C_2 with $R_1 \leftarrow R_1 - 2R_2$.

$$\begin{array}{c c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & 0 & 0 & 1 \\ 0 & \boxed{1} & 0 & -2 \\ 0 & 0 & \boxed{1} & 2 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

And this matrix is the reduced echelon matrix of the initial matrix.

1.3 Back to the linear system

Let's remember that we have made correspond to a linear system two matrices: the coefficient matrix and the augmented matrix. For now we will work with the augmented matrix and we will see later why and how the coefficient matrix comes also to the picture. We have seem that it is sometimes really straightforward to solve a system and when it is the case we have in reality that the augmented corresponding matrix in echelon form. We have just seen an algorithm which permits us to get systematically a matrix into a echelon form. Each column except the last column correspond to one of the unknown variables.

Definition 1.3.1. *The variable corresponding to pivot columns are called* **basic variable**. *The other variables are called* **free variables**.

Whenever the system is consistent, the solution set can be described explicitly by solving the reduced system of equations for the basic variables in terms of the free variables. This operation is possible because the reduced echelon form places each variables in one and only one equation. The system can be consistent and have no free variable. Once we get the augmented matrix associated to a system in (reduced) echelon form, one can find the solutions of the equations by back-substitution. More precisely, using the existence of an echelon form for any matrix, one can prove:

Theorem 1.3.2. *A linear system is consistent if and only if the right most column of the augmented matrix is not pivot column, that is, if and only if an echelon form of the augmented matrix has no row of the form*

$$[0, 0, 0, \cdots, 0, 0, b] \text{ with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either

1. *a unique solution, when there are no free variable;*
2. *infinitely many solution, when there is at least one free variable.*

When there are free variables, by (arbitrary) convention one choose the free variable as parameters and express the set of solutions (which is then infinite) in terms of these parameters. One can give then a parametric description of the infinite solution set. For each different choice of the parameters we have then a new solution.
Here how to use row reduction in order to solve a linear system:

1. Write the augmented matrix associated to the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form.
3. Decide whether the system is consistent or not. If no solution, stop, otherwise go to next step.
4. Continue row reduction to obtain the reduced echelon form.
5. Write the system of equations corresponding to the matrix obtain in the previous step.
6. Express using back-substitution from the bottom to the top, any basic variable in term of the free variables.

Exercise 1.3.3. *Find the general solution of the systems whose augmented matrices are given by:*

$$\begin{pmatrix} 3 & -2 & 4 & 0 \\ 9 & -6 & 12 & 0 \\ 6 & -4 & 8 & 0 \end{pmatrix}$$

Solution:

$$\begin{array}{cc} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{3} & -2 & 4 & 0 \\ 9 & -6 & 12 & 0 \\ 6 & -4 & 8 & 0 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

The pivot column is C_1. The pivot number is boxed. (If this number is not ideal for you, you can exchange two rows.) Here the pivot is great. So we use it in order to create zeros above it with $R_2 \leftarrow R_2 - 3R_1$, $R_3 \leftarrow R_3 - 2R_1$.

$$\begin{array}{cc} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{3} & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

The matrix is now in echelon form. We reduce it to obtain the reduced echelon form.

$$\begin{array}{cc} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} & \\ \begin{pmatrix} \boxed{1} & -2/3 & 4/3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} R_1 \\ R_2 \\ R_3 \end{array} \end{array}$$

The corresponding system is then

$$\begin{cases} x_1 - 2/3x_2 + 4/3x_3 & = & 0 \\ 0 & = & 0 \\ 0 & = & 0 \end{cases}$$

The basis variable is then x_1 and the free variable are x_2 and x_3. The solution set is then

$$\{(2/3x_2 - 4/3x_3, x_2, x_3), x_2, x_3 \in \mathbb{R}\}$$

1.4 Vector equations

1.4.1 Vector in $\mathbb{R}^n$

Definition 1.4.1. *A matrix with only one column is called a* **column vector** *(or simply a* **vector***). ' If n is a positive integer, $\mathbb{R}^n$ (read "r-n") denotes the collection of all list* *(or ordered n-tuples) of n real numbers, usually written as $n \times 1$ column matrices, such as*

$$u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{pmatrix}$$

The vector whose entries are all zero is called the **zero vector** *and is denoted by* **0**.

Example 1.4.2.

$$u = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \; v = \begin{pmatrix} .2 \\ .3 \\ .4 \end{pmatrix}, \; w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{pmatrix}$$

where w_n are any real numbers.

The set of all vectors with two entries is denoted by $\mathbb{R}^n$ (read "r-n"). The $\mathbb{R}$ stands for the real numbers that appears as entries in vectors, and the exponent n indicates that each vector contains n entries.

Definition 1.4.3.

1. *Two vectors are* **equal** *if and only if their have same size and their corresponding entries are equal. Vectors in $\mathbb{R}^n$ are ordered pairs of real numbers.*
2. *Given two vector u and v in $\mathbb{R}^n$, their* **sum** *is the vector $u + v$ is the vector of $\mathbb{R}^n$ obtained by summing the corresponding entries of each vector u and v. (We can only sum vectors of same size. It does not make sense to sum a vector of $\mathbb{R}^n$ and a vector of $\mathbb{R}^m$, for $m \neq n$.)*
3. *Given a vector u and a real number c, the* **scalar multiple of u by $c \in \mathbb{R}$** *is the vector $c \cdot u$ obtained by multiplying each each entry in u by c.*

Example 1.4.4.

1. *For instance* $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$ *and* $\begin{pmatrix} -1 \\ 3 \end{pmatrix}$ *are not equal.*
2. $$\begin{pmatrix} 3 \\ -1 \end{pmatrix} + \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$$
3. $$5 \cdot \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 15 \\ -5 \end{pmatrix}$$

4. *We can also do linear combination*

$$4 \cdot \begin{pmatrix} 3 \\ -1 \end{pmatrix} + 2 \cdot \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 22 \\ 8 \end{pmatrix}$$

Note that for space purpose, you might find in books a vector denote $(-3,1)^t$ *(or even* $(-3,1)$ *instead of* $\begin{pmatrix} -3 \\ 1 \end{pmatrix}$*. But please if you start with a notation keep up with it. Indeed, keep in mind that*

$$\begin{pmatrix} -3 \\ 1 \end{pmatrix} \neq (-3,1)$$

because the matrices have different number of rows and columns. So, be careful and do not write equalities which are not equalities at all.

These operations on vectors have the following properties, which can be verified directly from the corresponding properties for real numbers.

Lemma 1.4.5 (Algebraic properties of $\mathbb{R}^n$). *For all* $u, v, w \in \mathbb{R}^n$ *and all scalars* c *and* d*:*

1. $u + v = v + u$ *(commutativity)*
2. $(u + v) + w = u + (v + w)$*(associativity)*
3. $u + 0 = 0 + u = u$ *(zero element)*
4. $u + (-u) = -u + u = 0$ *where* $-u$ *denotes* $(-1) \cdot u$ *(inverse)*
5. $c(u + v) = cu + cv$ *(distributivity)*
6. $(c + d)u = cu + du$ *(distributivity)*
7. $c(du) = (cd) \cdot u$. *(associativity)*
8. $1 \cdot u = u$. *(identity element)*

For simplicity of notation, a vector as $u + (-1)v$ is often written as $u - v$.

Definition 1.4.6. *Given vectors* $v_1, v_2, \cdots, v_r$ *in* $\mathbb{R}^n$ *and given scalars* $c_1, c_2, \cdots, c_r$*, the vector* y *defined by*

$$y = c_1 v_1 + \cdots + c_r v_r$$

is called **a linear combination** *of* $v_1, \cdots, v_r$ *with* **weights** $c_1, \cdots, c_r$ *. Associativity above permits us to omit parentheses when forming such a linear combination. The weights in a linear combination can be any real numbers, including* 0.

Example 1.4.7. $3v_1 + (-7)v_2$, $1/3v_1 = 1/3v_1 + 0v_2$ *and* $0 = 0v_1 + 0v_2$*;*

Definition 1.4.8. *A vector equation*

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = b$$

has the same solution set as the linear system whose augmented matrix is

$$[a_1, a_2, \cdots, a_n, b]$$

In particular, b can be generated by a linear combination of $a_1, \cdots, a_n$, if and only if there exists a solution to the linear system corresponding to the matrix

$$[a_1, a_2, \cdots, a_n, b]$$

One of the Key of linear algebra is to study the set of all vectors that can be generated or written as a linear combination of a fixed set $\{v_1, \cdots, v_p\}$ of vectors.

Definition 1.4.9. *If $v_1, \cdots, v_r$ are in $\mathbb{R}^n$, then the set of all linear combinations of $v_1, \cdots, v_r$ is denoted by $Span\{v_1, \cdots, v_r\}$ and is called the* **subset of $\mathbb{R}^n$ spanned** *(or* **generated***) by $v_1, \cdots, v_r$. That is, $Span\{v_1, \cdots, v_r\}$ is the collection of all vectors that can be written in the form*

$$c_1v_1 + c_2v_2 + \cdots + c_rv_r$$

with $c_1, \cdots, c_r$ scalars.

Asking whether a vector b is in $Span\{v_1, \cdots, v_r\}$ amounts to asking whether the vector equation

$$x_1v_1 + x_2v_2 + \cdots + x_rv_r = b$$

has a solution, or, equivalent, asking whether the linear system with augmented matrix $[v_1, \cdots, v_p, b]$ has a solution.
Note that $Span\{v_1, \cdots, v_r\}$ contains every scalar multiple of v_1 (for example), since

$$cv_1 = cv_1 + 0v_2 + \cdots + 0v_r$$

In particular, the zero vector must be in $Span\{v_1, \cdots, v_r\}$.

Exercise 1.4.10. *Determine if $b = \begin{pmatrix} 11 \\ -5 \\ 9 \end{pmatrix}$ is a linear combination of $a_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $a_2 = \begin{pmatrix} -2 \\ 3 \\ -2 \end{pmatrix}$ and $a_3 = \begin{pmatrix} -6 \\ 7 \\ 5 \end{pmatrix}$.*

Solution: *This is equivalent to the question does the vector equation*

$$x_1a_1 + x_2a_2 + x_3a_3 = b$$

have a solution.
The equation

$$x_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 3 \\ -2 \end{pmatrix} + x_3 \begin{pmatrix} -6 \\ 7 \\ 5 \end{pmatrix} = \begin{pmatrix} 11 \\ -5 \\ 9 \end{pmatrix}$$

has the same solution set as the linear system whose augmented matrix is

$$M = \begin{pmatrix} 1 & -2 & -6 & 11 \\ 0 & 3 & 7 & -5 \\ 1 & -2 & 5 & 9 \end{pmatrix}$$

Row reduce M until the pivot position are visible (show your work):

$$M \sim \begin{pmatrix} \boxed{1} & -2 & -6 & 11 \\ 0 & \boxed{3} & 7 & -5 \\ 0 & 0 & \boxed{11} & -2 \end{pmatrix}$$

The linear system corresponding to M has a solution, so the initial vector equation has a solution and therefore b is a linear combination of a_1, a_2 and a_3.

1.5 The matrix equation $Ax = b$

A [1] fundamental idea in linear algebra is to view a linear combination of vectors as a product of a matrix and a vector.

Definition 1.5.1. *If A is an $m \times n$ matrix, with columns $a_1, \cdots, a_n$ and if x is in $\mathbb{R}^n$, then the **product of** A **and** x, denoted by Ax, is the* **linear combination of the columns of** A *using the corresponding entries in x as weights; that is*

$$Ax = (a_1, \cdots, a_n) \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = x_1 a_1 + \cdots + x_n a_n$$

This is called a **matrix equation***.*
Ax is a column vector whose size is equal to the number of row of the matrix A.

Note that Ax is defined (can be computed) if and only if the number of columns of A equals the number of entries in x.

Example 1.5.2. *1.*

$$\begin{pmatrix} 1 & 0 & 2 & 1 \\ 2 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 2 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

2.

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 10 \\ 20 \end{pmatrix} = 10 \cdot \begin{pmatrix} 2 \\ 4 \end{pmatrix} + 20 \begin{pmatrix} 3 \\ 5 \end{pmatrix} = \begin{pmatrix} 20 \\ 40 \end{pmatrix} + \begin{pmatrix} 60 \\ 100 \end{pmatrix} = \begin{pmatrix} 80 \\ 140 \end{pmatrix}$$

Example 1.5.3. *For $u_1, u_2, u_3 \in \mathbb{R}^m$, write the linear combination $4u_1 + 5u_2 + 9u_3$ as a matrix times a vector.*

1. Lecture 4

Solution: *Place u_1, u_2, u_3 into the columns of a matrix A and place the weights $4, 5$ and 9 onto a vector v. That is,*

$$4u_1 + 5u_2 + 9u_3 = (u_1, u_2, u_3)\begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix} = Ax$$

The following example will lead to a more efficient method for calculating the entries in Ax when working problems by hand.

Example 1.5.4. *Compute Ax, where $A = \begin{pmatrix} 2 & 4 & 7 \\ 8 & 9 & -11 \end{pmatrix}$ and $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$.*

Solution: *From the definition,*

$$\begin{aligned} \begin{pmatrix} 2 & 4 & 7 \\ 8 & 9 & -11 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 2 \\ 8 \end{pmatrix} x_1 + \begin{pmatrix} 4 \\ 9 \end{pmatrix} x_2 + \begin{pmatrix} 7 \\ -11 \end{pmatrix} x_3 \\ &= \begin{pmatrix} 2x_1 \\ 8x_1 \end{pmatrix} + \begin{pmatrix} 4x_2 \\ 9x_2 \end{pmatrix} + \begin{pmatrix} 7x_3 \\ -11x_3 \end{pmatrix} \\ &= \begin{pmatrix} 2x_1 + 4x_2 + 7x_3 \\ 8x_1 + 9x_2 - 11x_3 \end{pmatrix} \end{aligned}$$

The first entry in the product Ax is a sum of products (sometimes called a dot product), using the firs row of A and the entries in x. That is,

$$\begin{pmatrix} 2 & 4 & 7 \\ & & \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 + 4x_2 + 7x_3 \\ \end{pmatrix}$$

This shows how to compute the first entry in Ax directly, without writing done the vector computations. Similarly, the second entry in Ax can be calculated once by multiplying the entries in the second row of A by the corresponding entries x and then summing the resulting products:

$$\begin{pmatrix} & & \\ 8 & 9 & -11 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \\ 8x_1 + 9x_2 - 11x_3 \end{pmatrix}$$

Lemma 1.5.5. *If the product Ax is defined, then the i^{th} entry in Ax is the sum of the products of corresponding entries from row i of A and from the vector x.*

Example 1.5.6. 1.5.1 Geometric interpretation:

1. $\mathbb{R}^2$*: Consider a rectangular coordinate system in the plane. Because each point in the plane is determined by an ordered pair of numbers, we can identify a geometric point (a, b) with the column vector $\begin{pmatrix} a \\ b \end{pmatrix}$. So we may regard $\mathbb{R}^2$ as the set of all*

the points in the plane. The geometric visualization of a vector such as $\begin{pmatrix} a \\ b \end{pmatrix}$ *is often aided by including an arrow (directed line segment) from the origin* $(0,0)$ *to the point* $\begin{pmatrix} a \\ b \end{pmatrix}$.

The sum of two vectors has a useful geometric representation.

Lemma 1.5.7 (Parallelogram rule for addition)**.** *If u and v in $\mathbb{R}^2$ are represented as points in the plane, then $u + v$ corresponds to the fourth vertex of the parallelogram whose other vertices are u, 0 and v.*

Exercise 1.5.8. *(a) Let* $u = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$ *and* $v = \begin{pmatrix} -3 \\ 4 \end{pmatrix}$ *then* $u+v = \begin{pmatrix} 2 \\ 10 \end{pmatrix}$. *Place these vectors in a plane and check that in fact you obtain a parallelogram.*

(b) Let $u = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$. *Display the vectors u, $1/2u$ and $-3v$ on a graph. Note that* $1/2u = \begin{pmatrix} 5/2 \\ 3 \end{pmatrix}$ *and* $-3\begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} -15 \\ -18 \end{pmatrix}$. *The arrow $1/2u$ is half as long as the arrow u and in the same direction. The arrow for $-3v$ is 3 times the length of the arrow of u and in opposite direction. More generally, the arrow for cu is $|c|$ times the length of the arrow for u. (Recall that the length of the line segment from $(0,0)$ to (a,b) is $\sqrt{a^2+b^2}$.)*

2. $\mathbb{R}^3$*: Vector in $\mathbb{R}^3$ are 3×1 column matrices with three entries. They are represented geometrically by points in a three-dimensional coordinate space, with arrows from the origin sometimes included for visual clarity.*
   Let v be a nonzero in $\mathbb{R}^3$. Then $Span\{v\}$ is the set of all scalar multiples of v which is the set of points on the line in $\mathbb{R}^3$ through v and 0.
   If u and v are nonzero vectors in $\mathbb{R}^3$, with v not a multiple of u, then $Span\{u,v\}$ is the plane in $\mathbb{R}^3$ that contains u, v and 0. In particular, $Span\{u,v\}$ contains the line $\mathbb{R}^3$ through u and 0 and the line through v and 0.

1.5.2 Concrete application

Here also a more concrete application of vectors:

Exercise 1.5.9. *A mining company has two mines. One day's operation at mine $\sharp 1$ produces ore that contains 30 metric tons of copper and 600 kilograms of silver, while one day's operation at mine $\sharp 2$ produces ore that contains 40 metric tons of copper and 380 kilograms of silver. Let* $v_1 = \begin{pmatrix} 30 \\ 600 \end{pmatrix}$ *and* $v_2 = \begin{pmatrix} 40 \\ 380 \end{pmatrix}$. *Then v_1 and v_2 represent the "output per day" of mine $\sharp 1$ and mine $\sharp 2$, respectively.*

1. *What physical interpretation can be given to the vector $5v_1$?*
2. *Suppose the company operates mine $\sharp 1$ for x_1 days and mine $\sharp 2$ for x_2 days. Write a vector equation whose solution gives the number of days each mine should operate in order to produce 240 tons of copper and 2824 kilograms of silver. Solve the equation.*

Solution:

1. *$5v_1$ is the output of 5 days' operation of mine $\sharp 1$.*
2. *The total output is $x_1v_1 + x_2v_2$, so x_1 and x_2 should satisfy*

$$x_1v_1 + x_2v_2 = \begin{pmatrix} 240 \\ 2824 \end{pmatrix}$$

Reduce the augmented matrix (explain your work)

$$\begin{pmatrix} 30 & 40 & 240 \\ 600 & 380 & 2824 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1.73 \\ 0 & 1 & 4.70 \end{pmatrix}$$

Operate mine $\sharp 1$ for 1.73 days and mine $\sharp 2$ for 4.70 (This is an approximate solution.)

Definition 1.5.10. *The matrix $n \times n$ with $1's$ on the diagonal and $0's$ elsewhere is called an* **identity matrix** *and is denoted by I_n.*

One can show that $I_n x = x$ for all $x \in \mathbb{R}^n$.

Theorem 1.5.11. *If A is an $m \times n$ matrix, u and v in $\mathbb{R}^n$, and c is a scalar, then:*

1. *$A(u + v) = Au + Av$;*
2. *$A(cx) = c(Ax)$.*

Proof. Let $A = (a_1, a_2, \cdots, a_n)$, $u = \begin{pmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{pmatrix} \in \mathbb{R}^m$ and c is a scalar, then:

1.

$$\begin{aligned} A(u+v) &= (a_1, a_2, \cdots, a_n) \begin{pmatrix} u_1 + v_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n + v_n \end{pmatrix} \\ &= a_1(u_1 + v_1) + \cdots + a_n(u_n + v_n) \\ &= (a_1u_1 + \cdots + a_nu_n) + (a_1v_1 + \cdots + a_nv_n) \\ &= Au + Av \end{aligned}$$

2.

$$\begin{aligned} A(cu) &= (a_1, a_2, \cdots, a_n)(c \cdot \begin{pmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{pmatrix}) \\ &= (a_1, a_2, \cdots, a_n) \begin{pmatrix} cu_1 \\ \cdot \\ \cdot \\ \cdot \\ cu_n \end{pmatrix} \\ &= a_1(cu_1) + a_2(cu_2) + \cdots + a_n(cu_n) \\ &= c(a_1u_1 + a_2u_2 + \cdots + a_nu_n) \\ &= c(Av) \end{aligned}$$

□

Theorem 1.5.12. *If A is an $m \times n$ matrix, with columns $a_1, \cdots, a_n$, and if $b \in \mathbb{R}^m$, the matrix equation*

$$Ax = b$$

has the same solution set as the vector equation

$$x_1a_1 + x_2a_2 + \cdots + x_na_n = b$$

which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$(a_1, a_2, \cdots, a_n, b)$$

This theorem provides a powerful tool for gaining insight into problems in linear algebra. Indeed, a system of linear equations may now be viewed in three different but equivalent ways: as a matrix equation, as a vector equation , or as a system of linear equations. Whenever you construct a mathematical model of a problem in real life, you are free to choose whichever viewpoint is most natural. Then you may switch from one formulation of a problem to another whenever it is convenient. In any case, the matrix equation, the vector equation and the system equations can be all solved in the same way: row reduction of the augmented matrix. Note that the matrix appearing in the matrix equation is the coefficient matrix. Other methods in order to solve system of linear equation will be discussed later.

Example 1.5.13. *As an illustration of the previous theorem, we have*

$$\begin{cases} 2x_1 + 4x_2 + 7x_3 &= 6 \\ 8x_1 + 9x_2 - 11x_3 &= -9 \end{cases}$$

is equivalent to

$$\begin{pmatrix} 2 \\ 8 \end{pmatrix} x_1 + \begin{pmatrix} 4 \\ 9 \end{pmatrix} x_2 + \begin{pmatrix} 7 \\ -11 \end{pmatrix} x_3 = \begin{pmatrix} 6 \\ -9 \end{pmatrix}$$

is equivalent to

$$\begin{pmatrix} 2 & 4 & 7 \\ 8 & 9 & -11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ -9 \end{pmatrix}$$

Definition 1.5.14. *The columns of A span $\mathbb{R}$ means that every b in $\mathbb{R}^m$ is a linear combination of the columns of A. In general, a set of vectors $\{v_1, \cdots, v_r\}$ in $\mathbb{R}^m$* **spans** *(or* **generates***) $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ is a linear combination of $v_1, \cdots, v_r$, that is, if $Span\{v_1, \cdots, v_r\} = \mathbb{R}^m$.*

A harder existence problem is to determine whether the equation $Ax = b$ is consistent for all possible b:

Theorem 1.5.15. *Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.*

1. *For each b in $\mathbb{R}^m$, the equation $Ax = b$ has a solution.*
2. *Each b in $\mathbb{R}^m$ is a linear combination of the column of A.*
3. *The columns of A span $\mathbb{R}^m$.*
4. *A has a pivot position in every row.*

Note that the theorem is about the coefficient matrix, not an augmented matrix. If an augmented matrix (A, b) has a pivot position in every row, then the equation $Ax = b$ may or maybe not be consistent.

Proof. The statement 1., 2., 3. are equivalent because of the definition of Ax and what it means for a set of vectors to span $\mathbb{R}^m$. So it suffices to show (for an arbitrary matrix A) that 1. and 4. are either both true or both false. That will tie all four statement together.
Suppose 4. is true. Let U be an echelon form of A. Given b in $\mathbb{R}^m$, we can row reduce the augmented matrix (A, b) to an augmented matrix (U, d), for some d in $\mathbb{R}^m$:

$$(A, b) \sim \cdots \sim (U, d)$$

If the statement is true, then each row of U contains a pivot position and there can be no pivot in the augmented column. So $Ax = b$, has a solution for any b, and 1. is true. If 4. is false, the last row of U is all zeros. Let d be any vector with a 1 in its last entry. Then (U, d) represents an inconsistent system. Since row operations are reversible (U, d) can be transformed into the form (A, b). The new system $Ax = b$ is also inconsistent, and 1. is false. □

Example 1.5.16. *Let $A = \begin{pmatrix} 3 & -1 \\ -9 & 3 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. Describe the set of the possible b such that the system $Ax = b$ is consistent.*

Solution: *The augmented matrix for $Ax = b$ is $\begin{pmatrix} 3 & -1 & b_1 \\ -9 & 3 & b_2 \end{pmatrix}$, which is row equivalent to $\begin{pmatrix} 3 & -1 & b_1 \\ 0 & 0 & b_2 + 3b_1 \end{pmatrix}$.*

This shows that the equation $Ax = b$ is not consistent when $3b_1 + b_2$ is nonzero. The set of b for which the equation is consistent is a line through the origin, the set of all points (b_1, b_2) satisfying $b_2 = -3b_1$.

1.5.3 Solution set of linear systems

Solutions sets of linear system are important objects of study in linear algebra. They will appear later in several context. Thus sections uses vector notation to give explicit and geometric descriptions of such solution sets.

Definition 1.5.17. *A system of linear equation is said to be* **homogeneous** *if it can be written in the form $Ax = 0$, where A is an $m \times n$ matrix and 0 is the zero vector in $\mathbb{R}^m$. Such a system $Ax = 0$ always has at least a solution, namely, $x = 0$ (the zero vector in $\mathbb{R}^n$). This zero solution is usually called the* **trivial solution**. *For a given equation $Ax = 0$.*

The existence and uniqueness leads immediately to the following fact.

Corollary 1.5.18. *The homogeneous equation $Ax = 0$ has a non trivial solution if and only if the equation has at least one free variable.*

Example 1.5.19. *Describe the solutions of the following system in parametric vector form. Also, give a geometric description of the solution set.*

$$\begin{cases} 2x_1 + 2x_2 + 4x_3 &= 0 \\ -4x_1 - 4x_2 - 8x_3 &= 0 \\ -3x_2 - 3x_3 &= 0 \end{cases}$$

Solution: *Row reduce the augmented matrix for the system (write the row operation you are doing):*

$$\begin{pmatrix} 2 & 2 & 4 & 0 \\ -4 & -4 & -8 & 0 \\ 0 & -3 & -3 & 0 \end{pmatrix} \sim \begin{pmatrix} 2 & 2 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -3 & -3 & 0 \end{pmatrix} \sim \begin{pmatrix} 2 & 2 & 4 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The corresponding system is

$$\begin{cases} x_1 + x_3 &= 0 \\ x_2 + x_3 &= 0 \\ 0 &= 0 \end{cases}$$

Thus $x_1 = -x_3$, $x_2 = -x_3$, and x_3 is free. In parametric vector form,

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_3 \\ -x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_3 \\ -x_3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

The solution set is the line passing through the origin and $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$.

Note that a nontrivial solution x can have some nonzero entries as soon as not all its entries are zeros.

Example 1.5.20. *A single linear equation can be treated as a very simple system of equations. Describe all solution of the homogeneous "system":*

$$5x_1 + 10x_2 + 15x_3 = 0$$

The general solution is $x_1 = 2x_2 + 3x_3$ *with* x_2 *and* x_3 *free. As a vector, the general solution is*

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_2 + 3x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$$

with x_2 *and* x_3 *free. This calculation shows that every solution of the linear equation is a linear combination of the vectors* u *and* v*. That is the* $Span\{u, v\}$*. Since neither* u *nor* v *are scalar of the other, the solution set is a plane through the origin.*

As we seen in the example, the solution set of a homogeneous equation $Ax = 0$ can always be expressed explicitly as $span\{u_1, \cdots, u_r\}$ for suitable vectors $u_1, \cdots, u_r$. If the only solution is the zero vector then the solution set is $Span\{0\}$. If the equation $Ax = 0$ has only one free variable, then the solution set is a line through the origin. A plane through the origin provides a good mental image for the solution set of $Ax = 0$ when there are two free variable. Note, however that a similar figure can be used to visualize $Span\{u, v\}$ even when u and v do not arise as solution of $Ax = 0$. The initial equation of the previous example is an **implicit description**. Solving an equation amounts to finding an explicit description of the plane as the set spanned by u and v. The description of the solution as a Span, written as $x = su + tv$, s, t in $\mathbb{R}$, is called a **parametric vector equation** of the plane. We say that the solution is in **parametric vector form**.

When a nonhomogeneous system has many solutions, the general solution can be written in parametric vector form as one vector plus an arbitrary linear combination of vectors that satisfy the corresponding homogeneous system.

Example 1.5.21. *Describe the solutions of the following system in parametric vector form. Also, give a geometric description of the solution set.*

$$\begin{cases} 2x_1 + 2x_2 + 4x_3 &= 8 \\ -4x_1 - 4x_2 - 8x_3 &= -16 \\ -3x_2 - 3x_3 &= 12 \end{cases}$$

Solution: *Row reduce the augmented matrix for the system (write the row operation* *you are doing):*

$$\begin{pmatrix} 2 & 2 & 4 & 8 \\ -4 & -4 & -8 & -16 \\ 0 & -3 & -3 & 12 \end{pmatrix} \sim \begin{pmatrix} 2 & 2 & 4 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & -3 & -3 & 12 \end{pmatrix} \sim \begin{pmatrix} 2 & 2 & 4 & 8 \\ 0 & -3 & -3 & 12 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & 2 & 4 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 8 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The corresponding system is

$$\begin{cases} x_1 + x_3 &= 8 \\ x_2 + x_3 &= -4 \\ 0 &= 0 \end{cases}$$

Thus $x_1 = 8 - x_3$*,* $x_2 = -4 - x_3$*, and* x_3 *is free. In parametric vector form,*

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 - x_3 \\ -4 - x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ -4 \\ 0 \end{pmatrix} + \begin{pmatrix} -x_3 \\ -x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ -4 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$$

The solution set is the line through $\begin{pmatrix} 8 \\ -4 \\ 0 \end{pmatrix}$ *parallel to the line passing through the origin and* $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$ *(that is the the line obtained when we solved the corresponding homogeneous system in the previous example. .*

To describe the solution set of $Ax = b$ geometrically, we can think of vector addition as a translation. Given v and p in $\mathbb{R}^2$ and $\mathbb{R}^3$, the effect of adding p to v is to move v in direction parallel to the line through p and 0. We say that v **is translated by** p to $v + p$. If each point on a line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ is translated by a vector p, the result is a line parallel to L.
Suppose L is the line through 0 and v, described by the equation $x = tv$, $(t \in \mathbb{R})$. Adding p to each point of L produces the translated line described by the equation $x = p + tv$ $(t \in \mathbb{R})$. Note that p is on the line of the later equation. We call the equation $x = p + tv$ $(t \in \mathbb{R})$ **the equation of the line through p parallel to v**.
More generally, we have:

Theorem 1.5.22. *Suppose the equation* $Ax = b$ *is consistent for some given* b*, and* p *be a solution. Then the solution set of* $Ax = b$ *is the set of all the vectors of the form* $w = p + v_h$*, where* v_h *is any solution of the homogeneous equation* $Ax = 0$*.*

Note that this theorem apply only to an equation $Ax = b$ that has at least one nonzero solution p. When $Ax = b$ has no solution, the solution set is empty.
As a summary, if you need to describe the solution set of a system, you can:

1. Row reduce the augmented matrix to reduced echelon form.

2. Express each basic variables in terms of any free variables appearing in an equation.
3. Write a typical solution x as a vector whose entries depend on the free variables, if any.
4. Decompose x into a linear combination of vectors (with numeric entries) using the free variables as parameters.

1.5.4 Application to real life

Suppose a nation's economy is divided into many sectors, such as various manufacturing, communication, entrainment, and service industries. Suppose that for each sector, we know its total output for one year and we know exactly how this output is divided or "exchanged" among the other sectors of the economy. Let the total dollar value of a sector's output be called the **price** of that output. Leontief proved the following result:

There exist equilibrium prices that can be assigned to the total outputs of the various sectors in such a way that the income of each sector exactly balances its expenses.

The following example shows how to find the equilibrium prices.

Example 1.5.23. *Suppose an economy has only two sectors: Goods and Services. Each year, Goods sells* 80% *of its output to Services and keeps the rest, while Services sells* 70% *of its output to Goods and retains the rest. Find equilibrium prices for the annual outputs of the Goods and Services sectors that makes each sector's income match its expenditures.*

Solution: *We will fill what we call a exchange table (fill in the exchange table one column at a time). The entries in a column describe where a sectors's output goes. The decimal fractions in each column sum to* 1.

Output	*from*	
Goods	*Services*	*Purchased by*
0.2	0.7	*Goods*
0.8	0.3	*Services*

Denote the total annual output (in dollar) of the sector p_G *and* p_S. *From the first row, the total input to the Goods is* $0.2p_G + 0.7p_S$. *The Goods sector must pay for that. So the equilibrium prices must satisfy*

$$\begin{array}{lcl} \textit{income} & & \textit{expenses} \\ p_G & = & 0.2p_G + 0.7p_S \end{array}$$

From the second row, the input (that is, the expense) of the Services sector is $0.8p_G + 0.3p_S$. *The total input for the Services sectors is* $0.8p_G + 0.3p_S$. *The equilibrium equation* *for the Services sector is*

$$\begin{array}{lcl} \textit{income} & & \textit{expenses} \\ p_G & = & 0.8p_G + 0.3p_S \end{array}$$

Move all variables to the left side and combine like terms:

$$\begin{cases} 0.8p_G - 0.7p_S & = 0 \\ -0.8p_G + 0.7p_S & = 0 \end{cases}$$

Row reduce the augmented matrix:

$$\begin{pmatrix} 0.8 & -0.7 & 0 \\ -0.8 & 0.7 & 0 \end{pmatrix} \sim \begin{pmatrix} 0.8 & -0.7 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & -0.875 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The general solution is $p_G = 0.875p_S$, with p_S free. One equilibrium solution is $p_S = 1000$ and $p_G = 875$. If one uses fractions instead of decimals in the calculations, the general solution would be written $p_G = (7/8)p_S$, and a natural choice of prices might be $p_S = 80$ and $p_G = 70$. Only the ratio of the prices is important: $p_G = 0.875p_S$. The economic equilibrium is unaffected by proportional change in prices.

Chemical equations describe the quantities of substances consumed and produced by chemical reactions.

Example 1.5.24. *Balance the following chemical equations. Aluminum oxide and carbon react to create elemental aluminum and carbon dioxide:*

$$Al_2O_3 + C \to Al + CO_2$$

(For each compound, construct a vector that lists the numbers of atoms of aluminum, oxygen, and carbon.)
Solution: *The following vectors list the numbers of atoms of aluminum (Al), oxygen (O), and carbon (C):*

$$\begin{array}{cccc} Al_2O_3 & C & Al & CO_2 \\ \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \end{array}$$

The coefficients in the equation

$$x_1Al_2O_3 + x_2C \to x_3Al + x_4CO_2$$

satisfy

$$x_1 \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$$

Move the right terms to the left side (change the sign of each entry in the third and *fourth vectors) and row reduce the augmented matrix of the homogeneous system:*

$$\begin{pmatrix} 2 & 0 & -1 & 0 & 0 \\ 3 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1/2 & 0 & 0 \\ 3 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & -1/2 & 0 & 0 \\ 0 & 0 & 3/2 & -2 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1/2 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 3/2 & -2 & 0 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 0 & -1/2 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -4/3 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & -2/3 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -4/3 & 0 \end{pmatrix}$$

The general solution is $x_1 = (2/3)x_4$, $x_2 = x_4$, $x_3 = (4/3)x_4$ with x_4 free. Take for instance, $x_4 = 3$. Then $x_1 = 2$, $x_2 = 3$ and $x_3 = 4$. The balanced equation is

$$2Al_2O_3 + 3C \to 4Al + 3CO_2$$

1.6 Linear independence

The [2] homogeneous equations can be studied from a different perspective by writing them as a vector equations. In this way, the focus shifts from the unknown solutions of $Ax = 0$ to the vectors that appear in the vector equations. And one of the question arising was wether we had only one solution of infinitely many solution.
This question is related to the following definition:

Definition 1.6.1. *1. An indexed set of vectors $\{v_1, \cdots, v_r\}$ in $\mathbb{R}^n$ is said to be* **linearly independent** *if the vector equation*

$$x_1v_1 + x_2v_2 + \cdots + x_rv_r = 0$$

has only the trivial solution.

2. The set $\{v_1, \cdots, v_r\}$ is said to be **linearly dependent** *if there exist weights $c_1, \cdots, c_r$, not all zero such that*

$$c_1v_1 + c_2v_2 + \cdots + c_rv_r = 0$$

The equation of 2. is called a **linear dependence relation** among $v_1, \cdots, v_r$ when the weights are not zero. An indexed set is linearly dependent if and only if it is not linearly independent. In short, we may say that $v_1, \cdots, v_r$ are linearly dependent when we mean that $\{v_1, \cdots, v_r\}$ is a linearly dependent set. We use analogous terminology for linearly independent sets.

Example 1.6.2. *Determine if the vectors are linearly dependent. Justify each answer*

$$\begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 7 \\ 2 \\ -6 \end{pmatrix}, \begin{pmatrix} 9 \\ 4 \\ -8 \end{pmatrix}$$

2. Lecture 5

Solution: *In order to answer to the question, by definition of linear independence, we need to know wether the system*

$$x_1 \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 7 \\ 2 \\ -6 \end{pmatrix} + x_3 \begin{pmatrix} 9 \\ 4 \\ -8 \end{pmatrix} = 0 \ (S)$$

in x_1, x_2, x_3 *is consistent or not. For this, we consider the augmented matrix associated to the system*

$$\begin{pmatrix} 5 & 7 & 9 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & -6 & -8 & 0 \end{pmatrix} \sim \begin{pmatrix} \boxed{5} & 7 & 9 & 0 \\ 0 & \boxed{2} & 4 & 0 \\ 0 & 0 & \boxed{4} & 0 \end{pmatrix}$$

There are no free variables. So the homogeneous equation (S) *has only the trivial solution. By definition, this implies that the vectors are linearly independent.*

Suppose that we begin with a matrix $A = [a_1, \cdots, a_n]$ instead of a set vectors. The matrix equation $Ax = 0$ can be written as

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = 0$$

Each linear dependence relation among the columns of A corresponds to a nontrivial solution of $Ax = 0$. Thus we have the following important fact.

Fact 1.6.3. *The column of a matrix* A *are linearly independent if and only if the equation* $Ax = 0$ *has only the trivial solution.*

Example 1.6.4. *Determine if the column of the matrix* $A = \begin{pmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{pmatrix}$ *are linear independent.*

Solution: *Observe that the third column is the sum of the first two columns. That is* $c_1 + c_2 - c_3 = 0$ *if* c_i *represent the vector corresponding to the column* i*. So that the column of the matrix are not linearly independent.*

Fact 1.6.5. *A set containing only one vector- say* v*- is linearly independent if and only if* v *is not the zero vector.*

Proof. The vector equation $x_1 0 = 0$ is linearly dependent because $x_1 0 = 0$ has many nontrivial solutions. □

Fact 1.6.6. *A set of two vectors* $\{v_1, v_2\}$ *is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other. In geometric terms two vectors are linearly dependent if and only if they lie on the same line through the origin.*

Proof. If the set of vectors $\{v_1, v_2\}$ is linearly dependent then there is not both zero r, s such that

$$rv_1 + sv_2 = 0$$

Without loss of generality, one can suppose that $r \neq 0$ (if not exchange the role of v_1 and v_2).
Then

$$v_1 = -\frac{s}{r}v_2$$

and thus in this case v_1 is a multiple of v_2 $\square$

Example 1.6.7. *1. Are* $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ *linearly independent?*

Solution: *Yes, since they are not multiple on of the other, indeed there is no* x *such that* $v_1 = xv_2$ *neither* $v_2 = xv_1$.

2. Are $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ *linearly independent?*

Solution: *No, since* $v_2 = 2v_1$.

The following theorem is a generalization of the two previous facts:

Theorem 1.6.8 (Characterization of linearly dependent sets)**.** *An indexed set* $S = \{v_1, \cdots, v_r\}$ *of two vectors is linearly dependent if and only if one of the vectors in* S *is a linear combination of the others. In fact, if* S *is linearly dependent and* $v_1 \neq 0$ *then some* v_j *(with* $j > 1$*) is a linear combination of the preceding vectors* $v_1, \cdots, v_{j-1}$.

Be careful, this theorem do not say that every vector in a linearly dependent set is a linear combination of the preceding vectors. A vector in a linearly dependent set may fail to be a linear combination of the other vectors. You will need to read thoroughly several times a section to absorb important concept such as linear independence. And even come back to it and reread again at a later state of the course. The following proof is worth reading carefully because it shows how the definition of linear independence can be used.

Proof. If some v_j in S equals a linear combination of the other vectors, then v_j can be subtracted from both sides of the equation, producing a linear dependence relation with a nonzero weight (-1) on v_j. Thus S is linearly dependent.
Conversely, suppose S is linearly dependent. If v_1 is zero, then it is a (trivial) linear combination of the other vectors in S. Otherwise, $v_1 \neq 0$ and there exist weights $c_1, \cdots, c_r$ not all zero, such that

$$c_1v_1 + c_2v_2 + \cdots + c_rv_r = 0$$

Let j be the largest subscript for which $c_j \neq 0$. If $j = 1$, then $c_1v_1 = 0$, which is impossible because $v_1 \neq 0$. So $j > 1$, and

$$c_1v_1 + \cdots + c_jv_j + 0v_{j+1} + \cdots 0v_r = 0$$

Then

$$c_jv_j = -c_1v_1 - \cdots - c_{j-1}v_{j-1}$$

Finally,

$$v_j = (-\frac{c_1}{c_j})v_1 + \cdots + (-\frac{c_{j-1}}{c_j})v_{j-1}$$

$\square$

Fact 1.6.9. *Let u and v be two linearly independent vector. The vector w is in $Span\{u, v\}$ if and only if $\{u, v, w\}$ is linearly dependent. Geometrically, that means that the set $\{u, v, w\}$ is linearly dependent if and only if w is in the plane spanned by u and v.*

Proof. For the last assertion, remember that $Span\{u, v\}$ of two linearly independent vectors of $\mathbb{R}^3$ is a plane through the origin whose direction is given by u and v.
If w is a linear combination of u and v, then $\{u, v, w\}$ is linearly dependent, by the previous theorem on the characterization of linearly dependent sets.
Conversely, suppose that $\{u, v, w\}$ are linearly dependents. By the previous theorem on the characterization of linearly dependent sets, some vector in $\{u, v, w\}$ is a linear combination of the preceding vectors (since $u \neq 0$). That vector must be w since v is not a multiple of u and vice versa. So w is in $Span\{u, v\}$. □

The next two theorem describe special cases in which the linear dependence of a set is automatic.

Theorem 1.6.10. *If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{v_1, \cdots, v_r\}$ in $\mathbb{R}^n$ is linearly dependent if $r > n$.*

Be careful! This theorem says nothing about the case in which the number of vectors in the set does not exceed the number of entries in each vector.

Proof. Let $A = [v_1, \cdots, v_r]$. Then A is $n \times p$, and the equation $Ax = 0$ corresponds to a system of n equations in r unknowns. If $r > n$, there are more variables than equations, so there must be free variable. Hence, $Ax = 0$ has a non trivial solution, and the column of A are linearly dependent. □

Example 1.6.11. *Just applying the previous theorem, one can say that for instance $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ are linearly dependent. Notice, however that none of the vector is multiple of the other, the theorem stating this is true only when two vectors are involved.*

Theorem 1.6.12. *If a set $S = \{v_1, \cdots, v_r\}$ in $\mathbb{R}^n$ contains the zero vector, the the set is linearly dependent.*

Proof. By renumbering the vectors, we may suppose that $v_1 = 0$. Then the equation $1v_1 + 0v_2 + \cdots + 0v_r = 0$ show that S is linearly dependent. □

Example 1.6.13.

1. *$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ are linearly dependent since we have more vectors than the number of entry in each vector.*
2. *$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is linearly dependent since one of the vector in the set of the give vectors is the zero vector.*

3. $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ *and* $\begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$ *are linearly dependent since the second vector equal to 2 times the first one, so they are multiple of each other.*

1.7 Introduction to linear transformations

1.7.1 Definition and first properties

If you think about the matrix multiplication we define earlier, you can realize that given a matrix A $m \times n$ we have assigned to a vector $x \in \mathbb{R}^n$ a vector Ax in $\mathbb{R}^m$. You can then see the correspondence from x to Ax as a function from one set of vector $\mathbb{R}^n$ to another $\mathbb{R}^m$. More precisely,

Definition 1.7.1. *A* **transformation** *(or* **function** *or* **mapping***) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector x in $\mathbb{R}^n$ a vector $T(x)$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the* **domain** *of T, and $\mathbb{R}^m$ is called the* **codomain** *of T. The notation $T : \mathbb{R}^n \to \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For x in $\mathbb{R}^n$, the vector $T(x)$ in $\mathbb{R}^m$ is called the* **image of x** *(under the action of T). The set of all images $T(x)$ is called the range of T.*

The notion of transformation is an important notion as it permits to relate different set of vectors but also to build mathematical models of physical systems that evolve over time.

Definition 1.7.2. *A* **matrix transformation** *is a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ which sends $x \in \mathbb{R}^n$ to a $Ax \in \mathbb{R}^m$, for a given $m \times n$ matrix A. We denote this transformation by $x \mapsto Ax$. As noted, the domain of T is $\mathbb{R}^n$ when A has n columns and the codomain of T is $\mathbb{R}^m$ when each column of A has m entries. The range of T is the set of all linear combinations of the columns of A, because each image $T(x)$ is of the form Ax.*

Example 1.7.3. *Let* $A = \begin{pmatrix} 1 & 0 & -3 \\ -3 & 1 & 6 \\ 2 & -2 & -1 \end{pmatrix}$, $b = \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix}$, $u = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ *and* $c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. *Define a transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x) = Ax$.*

1. *Find $T(u)$, the image of u under the transformation T.*
2. *Find the image of* $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ *by T.*
3. *Find an $x \in \mathbb{R}^3$ whose image is b under T. Is it unique?*
4. *Determine if c is in the range of the transformation T.*

Solution:

1. *We compute the image of* u *under* T, *that is*

$$T(u) = Au = \begin{pmatrix} 1 & 0 & -3 \\ -3 & 1 & 6 \\ 2 & -2 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$$

2. *We compute the image of* $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ *by* T *to be*

$$T(x) = Ax = \begin{pmatrix} x_1 - 3x_3 \\ -3x_1 + x_2 + 6x_3 \\ 2x_1 - 2x_2 - x_3 \end{pmatrix}$$

3. *We want to find* $x \in \mathbb{R}^3$ *whose image is* b *under* T *that is an* $x \in \mathbb{R}^3$ *satisfying* $T(x) = b$. *For this we can row reduce the augmented matrix corresponding to this system:*

$$\begin{pmatrix} 1 & 0 & -3 & -2 \\ -3 & 1 & 6 & 3 \\ 2 & -2 & -1 & -1 \end{pmatrix} \sim_{R_2 \leftarrow R_2 + 3R_1 \text{ and } R_3 \leftarrow R_3 - 2R_1} \begin{pmatrix} 1 & 0 & -3 & -2 \\ 0 & 1 & -3 & -3 \\ 0 & -2 & 5 & 3 \end{pmatrix}$$
$$\sim_{R_3 \leftarrow R_3 + 2R_2} \begin{pmatrix} 1 & 0 & -3 & -2 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & -1 & -3 \end{pmatrix} \sim_{R_3 \leftarrow -R_3} \begin{pmatrix} 1 & 0 & -3 & -2 \\ 0 & 1 & -3 & -3 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$
$$\sim_{R_2 \leftarrow R_2 + 3R_3 \text{ and } R_1 \leftarrow R_1 - 3R_3} \begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & 3 \end{pmatrix}$$

As a consequence, since the system has no free variable, there is only one $x \in \mathbb{R}^3$ *whose image is* b *under* T *this* $x = \begin{pmatrix} 7 \\ 6 \\ 3 \end{pmatrix}$.

4. $c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ *is in the range of* T *if and only if the system* $Ax = c$ *has at least a solution. But as we have seen in the previous question the coefficient matrix associated to this system has a reduced echelon form:*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As we can observe that each row of this matrix contains a pivot position we know that the system $Ax = c$ *has a solution for every* c, *indeed it has a solution exchanging* c *by any vector in* $\mathbb{R}^3$.

Example 1.7.4. *If* $A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ *then the transformation* $x \mapsto Ax$ *projects points in* $\mathbb{R}^3$ *onto the* x_2x_3*-plane because*

$$Ax = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix}$$

Important transformations are obtained when one wants that the transformation considered respect some algebraic properties of the domain and codomain. The most important transformation for us are the linear transformation that are the one which preserve the operations of vector addition and scalar multiplication. In other terms,

Definition 1.7.5. *A transformation (or mapping) if* T *is* **linear** *if:*

1. $T(u + v) = T(u) + T(v)$ *for all* u, v *in the domain of* T*;*
2. $T(cu) = cT(u)$ *for all scalars* c *and all* u *in the domain of* T*.*

As we have proven earlier we have $A(u + v) = Au + Av$ and $A(cu) = cA(u)$, for all u, v in the domain of T and all scalars c, so that every matrix transformation is a linear transformation.

Fact 1.7.6. T *is a linear transformation if and only if* $T(cu + dv) = cT(u) + dT(v)$ *for all* u, v *in domain of* T *and* c, d *scalars.*

Proof. $\Rightarrow$ Suppose T is a linear transformation. Let u, v in domain of T, then

$$\begin{aligned} T(cu + dv) &= T(cu) + T(dv) \text{ by property 1.} \\ &= cT(u) + dT(v) \text{ by property 2.} \end{aligned}$$

$\Leftarrow$ Suppose T satisfies $T(cu + dv) = cT(u) + dT(v)$ for all u, v in domain of T and c, d scalars. Then we obtain the assertion 1., by taking $c = 1$ and $d = 1$ and assertion **2.** by taking $d = 0$. So, T is a linear transformation. □

Fact 1.7.7. *If* T *is a linear transformation, then* $T(0) = 0$*.*

Proof. For this take $c = 0$ in the assertion 2.. □

Repeated application of produces a useful generalization:

Fact 1.7.8. *For* $v_1, \cdots, v_r$ *in the image of* T*,* $c_1, \cdots, c_r$ *in* $\mathbb{R}$*,*

$$T(c_1v_1 + \cdots + c_rv_r) = c_1T(v_1) + \cdots + c_rT(v_r)$$

Example 1.7.9. *Given a scalar* r*, define* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $T(x) = rx$*.* T *is called a* **contraction** *when* $0 \leqslant r \leqslant 1$ *and a* **dilation** *when* $r > 1$*. Show that in any cases* T

defines a linear transformation.
Solution: *Let u, v be in $\mathbb{R}^2$ and c, d be scalar. Then*

$$\begin{aligned} T(cu+vc) &= r(cu+dv) \text{ by definition of } T \\ &= rcu+rdv = c(ru)+d(rv) \text{ for vectors properties} \\ &= cT(u)+dT(v) \end{aligned}$$

Thus, T is a linear transformation.

Example 1.7.10. *Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by*

$$T(x) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$$

Find the images of $u = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $v = \begin{pmatrix} 4 \\ 7 \end{pmatrix}$ and $u + v$.
Solution: *The images are:*

$$T(u) = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

$$T(v) = \begin{pmatrix} -7 \\ 4 \end{pmatrix}$$

$$T(u+v) = T(u) + T(v) = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + \begin{pmatrix} -7 \\ 4 \end{pmatrix} = \begin{pmatrix} -10 \\ 5 \end{pmatrix} \text{ by linearity.}$$

It appear that T rotated the entire parallelogram determined by u and v into the one determined by $T(u)$ and $T(v)$, by a rotation counterclockwise about the origin through 90 *degrees.*

1.7.2 The matrix of a linear transformation

Whenever a linear transformation T arises geometrically or is described in words, we usually want a "formula" for $T(x)$. The discussion that follows shows that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $x \mapsto Ax$ and the important properties of T are intimately related to familiar properties of A. The key to finding A is to observe that T is completely determined by what it does to the column of the $n \times n$ identity matrix I_n.

Theorem 1.7.11. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that $T(x) = Ax$ for all x in $\mathbb{R}^n$. In fact, A is the $m \times n$ matrix whose j column is the vector $T(e_j)$ where e_j is the jth column of the identity matrix in $\mathbb{R}^n$:*

$$A = [T(e_1), \cdots, T(e_n)]$$

The matrix A is called **the standard matrix for the linear transformation** T.

Proof. Write $x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = [e_1, \cdots, e_n]x = x_1e_1 + \cdots + x_ne_n$, and use the linearity of T to compute

$$T(x) = T(x_1e_1 + \cdots + x_ne_n) = x_1T(e_1) + \cdots + x_nT(e_n) = [T(e_1), \cdots, T(e_n)] \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = Ax$$

The uniqueness of A will be given as an exercise in the homework. □

Note that in particular, the image of the column of the identity matrix completely determine a linear transformation.
We know now that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be viewed as a matrix transformation, and vice versa. The term linear transformation focuses on a property of a mapping, while matrix transformation describes how such a mapping is implemented. Here some examples,

Example 1.7.12. *Find the standard matrix A for the dilatation transformation $T(x) = rx$, for $x \in \mathbb{R}^n$.*
Solution: *Since $T(e_i) = re_i$ for all $i \in \{1, \cdots, n\}$, then*

$$A = rI_n$$

Example 1.7.13. *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point $\mathbb{R}^2$ about the origin through an angle ϕ, with counterclockwise rotation for a positive angle. One can show that such a transformation is linear. Find the standard matrix A of this transformation.*
Solution: *Note that*

$$T(e_1) = \begin{pmatrix} cos(\phi) \\ sin(\phi) \end{pmatrix}$$

and

$$T(e_2) = \begin{pmatrix} -sin(\phi) \\ cos(\phi) \end{pmatrix}$$

So that the standard matrix A of this transformation is

$$\begin{pmatrix} cos(\phi) & -sin(\phi) \\ sin(\phi) & cos(\phi) \end{pmatrix}$$

In the table below, you can find some transformation in $\mathbb{R}^2$ with the geometric and matrix description.

Transformation	**Standard matrix**	**Action on** $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$
Reflection through the x_1-axis	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$
Reflection through the x_2-axis	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}$
Reflection through the line $x_1 = x_2$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$
Reflection through the line $x_1 = -x_2$	$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -x_2 \\ -x_1 \end{pmatrix}$
Reflection through the origin	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}$
Horizontal contraction and expansion	$\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} kx_1 \\ x_2 \end{pmatrix}$
Vertical contraction and expansion	$\begin{pmatrix} 1 & 0 \\ 0 & k \end{pmatrix}$	$\begin{pmatrix} x_1 \\ kx_2 \end{pmatrix}$
Horizontal shear	$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} x_1 + kx_2 \\ x_2 \end{pmatrix}$
Vertical shear	$\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$	$\begin{pmatrix} x_1 \\ x_2 + kx_1 \end{pmatrix}$
Projection onto the x_1-axis	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} x_1 \\ 0 \end{pmatrix}$
Projection onto the x_2-axis	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ x_2 \end{pmatrix}$

Definition 1.7.14. *A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be* **onto** $\mathbb{R}^m$ *if each b in $\mathbb{R}^m$ is the image of at least one x in $\mathbb{R}^n$.*

Fact 1.7.15. *A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto if and only if the range of T is all the codomain $\mathbb{R}^m$. That is, also equivalent to say that each b in the codomain $\mathbb{R}^m$, there exist at least one solution of $T(x) = b$. The mapping T is not onto when there is some b in in $\mathbb{R}^m$ for which the equation $T(x) = b$ has no solution.*

Does T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$? is an existence question.

Definition 1.7.16. *A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be* **one-to-one** *if each b in $\mathbb{R}^m$ is the image of at most one x in $\mathbb{R}^n$.*

Fact 1.7.17. *T is a one-to-one if, each b in $\mathbb{R}^m$, the equation $T(x) = b$ has either a unique solution or none at all. The mapping T is not one-to-one when some b in $\mathbb{R}^m$ is the image of more than one vector in $\mathbb{R}^n$. If there is no such a b, then T is one-to-one.*

The question " Is T one-to-one?" is a uniqueness question,

Example 1.7.18. *Let T be the linear transformation whose standard matrix is*

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Do T map $\mathbb{R}^2$ onto $\mathbb{R}^2$? Is T a one-to-one mapping?
Solution: *Since the matrix has not a pivot position in each row that means that from a theorem of the class, that the equation $Ax = b$ is not consistent, for each b in $\mathbb{R}^3$. In other word, the linear transformation T does not map $\mathbb{R}^2$ onto $\mathbb{R}^2$. Also, since the equation $Ax = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ has a free variable, then $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is the image of more than one x so that T is not one-to-one.*

Theorem 1.7.19. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(x) = 0$ has only the trivial solution.*

Proof. $\Rightarrow$ Suppose T is one-to-one. We want to prove that the equation $T(x) = 0$ has only one solution. For this, we consider the equation $T(x) = 0$. Since T is one-to-one, then 0 is the image of only one element of $\mathbb{R}^n$, since we have already seen that $T(0) = 0$. Then we know that the trivial solution is this only solution.
$\Leftarrow$ Suppose $T(x) = 0$ has only one solution. We want to prove that T is one-to-one that is that any $b \in \mathbb{R}^m$, the system $Ax = b$ has only one solution. For this one can suppose that there are two solutions u and $v \in \mathbb{R}^n$ of these solutions, that is $T(u) = b$ and $T(v) = b$. We need to prove that $u = v$.
We have that $b = T(u) = T(v)$ by assumption. Thus, by linearity we obtain $T(u) - T(v) = T(u - v) = 0$ since the equation $T(x) = 0$ has only the trivial solution. Then $u - v = 0$ that is $u = v$. We have then proven that T is one-to-one. $\square$

Theorem 1.7.20. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let A be the standard matrix for T. Then:*

1. *T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$ (this is also equivalent to every vector of $\mathbb{R}^m$ is a linear combination of the columns of A) ;*
2. *T is one-to-one if and only if the columns of A are linearly independent.*

Proof. This come from the definitions and an earlier theorem seen in class. $\square$

Example 1.7.21. *Let T be defined as $T(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} 2x_1 + x_2 \\ x_1 \\ x_2 \end{pmatrix}$. Show that T is a one-to-one linear transformation. Does T map $\mathbb{R}^2$ onto $\mathbb{R}^3$?*
Solution: *Note that*

$$T(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \begin{pmatrix} 2x_1 + x_2 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

The columns of A are not multiple one of the others, that means that they are linearly independent and thus by the previous theorem we have that T is one-to-one. Moreover since the column of A span $\mathbb{R}^3$ if and only if they have a pivot position in each row, we see that this is impossible since A has only 2 columns. So the columns of A do not span $\mathbb{R}^3$, and the associated linear transformation is not onto $\mathbb{R}^3$.

Chapter 2

Matrix algebra

2.1 Matrix operation

Definition 2.1.1. *1. If A is a $m \times n$ matrix, that is, a matrix with m rows and n columns, then the scalar entry in the ith row and jth column of A is denoted by $a_{i,j}$ and is called the (i,j) entry of A. Each column of A is a list of m real numbers, which identifies with a vector in $\mathbb{R}^m$. Often, these columns are denoted by $a_1, \cdots, a_n$ and the matrix A is written as*

$$A = [a_1, \cdots, a_n]$$

Observe that the number $a_{i,j}$ is the ith entry (from the top) of the jth column vector a_j. The **diagonal entries** *in an $m \times n$ matrix $A = [a_{i,j}]$ are $a_{1,1}$, $a_{2,2}$, $a_{3,3}$, $\cdots$ and they form the* **main diagonal** *of A. A* **diagonal matrix** *is a square $n \times n$ whose non diagonal entries are zero. An example is the $n \times n$ identity matrix. An $m \times n$ matrix whose entries are all zero is a* **zero matrix** *and is written as 0. The size of a zero matrix is actually clear from the context.*

2. **(Equality)** *Two matrix are* **equal** *if they have same size (i.e same number or row and same number of column) and if their corresponding columns are equals which amounts to saying that their corresponding entries are equal.*

3. **(Sum)** *The* **sum** *of two matrix A and B is defined if and only if the two matrices HAVE THE SAME SIZE. The sum of two matrix $m \times n$ is a matrix $m \times n$ whose columns is the sum of the corresponding columns. That is, the entries of $A + B$ is the sum of the corresponding entries in A and B.*

4. **(Scalar multiplication)** *If r is a scalar and A is a matrix $m \times n$, then the* **scalar multiple** *rA is the matrix whose the columns are r times the corresponding column in A. We denote $-A$ for $(-1)A$ and $A - B = A + (-1)B$.*

Example 2.1.2. *Let* $A = \begin{pmatrix} 2 & -3 & 0 & 9 \\ 1 & 3 & 4 & 2 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & -2 & 7 & 0 \\ 10 & 2 & 3 & 7 \end{pmatrix}$ *and* $C = \begin{pmatrix} 7 & 5 \\ 2 & 7 \end{pmatrix}$. *Then*

1.

$$A+B=\begin{pmatrix} 2 & -3 & 0 & 9 \\ 1 & 3 & 4 & 2 \end{pmatrix}+\begin{pmatrix} 1 & -2 & 7 & 0 \\ 10 & 2 & 3 & 7 \end{pmatrix}=\begin{pmatrix} 3 & -5 & 7 & 9 \\ 11 & 5 & 7 & 9 \end{pmatrix}.$$

Be careful, $A+C$ *or* $B+C$ *are not well defined.*

2. $2A=\begin{pmatrix} 4 & -6 & 0 & 18 \\ 2 & 6 & 8 & 4 \end{pmatrix}.$

3. $2A-B=\begin{pmatrix} 3 & -4 & -7 & 18 \\ -8 & 4 & 5 & -3 \end{pmatrix}.$

Here some important algebraic properties of the matrices similar to the one we see on vectors.

Theorem 2.1.3. *Let* A, B *and* C *3* $m\times n$ *matrices and let* r *and* s *be scalar. Then*

1. $A+B=B+A$ *(commutativity);*
2. $(A+B)+C=A+(B+A)$ *(associativity);*
3. $A+0=A$ *(zero element),*
4. $r(A+B)=rA+rB$ *(distributivity)*
5. $(r+s)A=rA+rB$ *(distributivity)*
6. $r(sA)=(rs)A$ *(associativity).*

Proof. This is a consequence of the definitions of matrix operation and also applications of the properties, we prove proved for vectors. □

Definition 2.1.4. *1. Let* A *be a matrix* $m\times n$ *and* B *be a matrix* $n\times p$ *with column* b_1, b_2, $\cdots$, b_p, *then the* **product** AB *is defined and it is a* $m\times p$ *matrix whose column are* Ab_1, Ab_2, $\cdots$, Ab_p. *That is,*

$$AB=A[b_1,\cdots,b_p]=[Ab_1,\cdots,Ab_p]$$

Each column of AB *is a linear combination of the column of* A *using the weights from the corresponding column of* B. *We have that the* (i,j) *entry of the matrix* AB *is*

$$(ab)_{i,j}=\sum_{k=1}^{n}a_{i,k}b_{k,j}$$

for each $i=1,\cdots,m$ *and* $j=1,\cdots,p$.
Be careful, in order for the multiplication to be defined it is necessary that the NUMBER OF COLUMNS OF A equals the NUMBER OF ROWS IN B. Also the matrix AB *has size* $m\times p$. *The number of rows is equal to* m *(number of row of* A*) and the number of column is equal to* p *(number of column of* B*).*
Let $row_i(A)$ *denotes the* i*th row of the matrix* A *and* $row_i(AB)$ *denotes the* i*th row of the matrix* AB. *Then,*

$$row_i(AB)=row_i(A)\cdot B$$

2. **(Powers)** *If A is an $n \times n$ matrix and if k is a positive integer, then A^k denotes the product of copies of A*
$$A^k = A \cdots A$$
If $k = 0$, $A^0 = I_n$.

Theorem 2.1.5 (Column-Row expansion of AB). *If A is $m \times n$ and B is $n \times p$, then*

$$\begin{aligned} AB &= \begin{pmatrix} Col_1(A) & Col_2(A) & \cdots & Col_n(A) \end{pmatrix} \begin{pmatrix} Row_1(B) \\ Row_2(B) \\ \cdots \\ Row_n(B) \end{pmatrix} \\ &= Col_1(A)Row_1(A) + \cdots + Col_n(A)Row_n(B) \end{aligned}$$

Proof. Left as an exercise. □

Example 2.1.6. *Let $A = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & -3 & 0 & 9 \\ 1 & 3 & 4 & 2 \end{pmatrix}$. Compute AB and BA if possible.*

$$\begin{aligned} AB &= \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -3 & 0 & 9 \\ 1 & 3 & 4 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 \times 1 + 1 \times 0 & -3 \times 1 + 0 \times 3 & 0 \times 1 + 0 \times 4 & 9 \times 1 + 0 \times 2 \\ -1 \times 2 + 2 \times 1 & (-3) \times (-1) + 3 \times 2 & 0 \times (-1) + 2 \times 4 & (-1) \times 9 + 2 \times 2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -3 & 0 & 9 \\ 0 & 9 & 8 & -5 \end{pmatrix} \end{aligned}$$

Note that BA is not define, you cannot make this product indeed the number of columns of A is not equals to the number of rows of B.

Here, some important algebraic properties for the matrix multiplication:

Theorem 2.1.7. *Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined. Then*

1. *$I_m x = x$, for all $x \in \mathbb{R}^m$,*
2. *$A(BC) = (AB)C$, (associativity law of multiplication); (Note that from this property, from now on the product ABC makes sense and more generally the product between r matrices).*
3. *$A(B + C) = AB + AC$ (left distributive law);*
4. *$(B + C)A = BA + CA$ (right distributive law);*
5. *$r(AB) = (rA)B = A(rB)$, for any scalar r;*
6. *$I_m A = A = AI_n$, (Identity for matrix multiplication).*

Proof. 1., 3., 4., 5. and 6 are left for you as a good exercise.
We will prove 2..
Let $C = [c_1, \cdots, c_p]$ $r \times p$ matrix and $B = [b_1, \cdots, b_r]$ $n \times r$ matrix. By definition of matrix multiplication,
$$BC = [Bc_1, \cdots, Bc_p]$$

$$\begin{aligned} A(BC) &= [A(Bc_1), \cdots, A(Bc_p)] = [A(b_1c_{1,1} + \cdots + b_rc_{1,r}), \cdots, A(b_1c_{p,1} + \cdots + b_rc_{p,r})] \\ &= [(Ab_1)c_{1,1} + \cdots + (Ab_r)c_{r,1}, \cdots, (Ab_1)c_{1,p} + \cdots + (Ab_r)c_{r,p}] \\ &= [(AB)c_1, \cdots, (AB)c_p] \\ &= (AB)C \end{aligned}$$

□

Note that, the matrix AB is the matrix of the linear transformation R defined by $x \mapsto (AB)x$. Also, if T is a transformation defined by $x \mapsto Bx$ and S defined by $x \mapsto Ax$, note that $R = S \circ T$. Here, extremely important facts to be careful and to remember:

Fact 2.1.8. (WARMINGS)

1. *Be careful, in general $AB \neq BA$. We say that A and B* **commute** *with another if $AB = BA$.DO NOT FORGET THIS IS NOT TRUE IN GENERAL.*
2. *The cancelation laws DO NOT hold for matrix multiplication. That is $AB = AC$, then it is NOT TRUE in general that $B = C$. You could have $AB = AC$ but still $B \neq C$.*
3. *If a product AB is the zero matrix, you CANNOT conclude in general that either $A = 0$ or $B = 0$. You could have $AB = 0$ but still $A \neq 0$ or $B \neq 0$.*

Example 2.1.9. 1. *Let $A = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 9 \\ 1 & 3 \end{pmatrix}$. Show that for these matrix $AB \neq BA$.*

Solutions:

$$AB = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 9 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 9 \\ 2 & -3 \end{pmatrix}$$

and

$$BA = \begin{pmatrix} 0 & 9 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} -9 & 18 \\ -2 & 6 \end{pmatrix}$$

So, we have that $AB \neq BA$.

2. *Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and compute A^2.*

$$A^2 = AA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

But $A \neq 0$.

2.2 Transpose

Definition 2.2.1. *Given an $m \times n$ matrix A, the* **transpose** *of A is the $n \times m$ matrix, denoted by A^T whose columns are formed from the corresponding rows of A.*

Example 2.2.2. 1. *The transpose of $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $A^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$.*

2. *The transpose of* $B = \begin{pmatrix} 2 & -3 & 0 & 9 \\ 1 & 3 & 4 & 2 \end{pmatrix}$ *is* $B^T = \begin{pmatrix} 2 & 1 \\ -3 & 3 \\ 0 & 4 \\ 9 & 2 \end{pmatrix}$

Theorem 2.2.3. *Let A and B denote the matrices whose sizes are appropriate for the following sums and products.*

1. $(A^T)^T = A$;
2. $(A+B)^T = A^T + B^T$;
3. *For any scalar r,* $(rA)^T = rA^T$;
4. $(AB)^T = B^T A^T$. *(The transpose of a product of matrices equal the product of their transposes in the REVERSE order. Be careful,* $(AB)^T = A^T B^T$.*)*

Proof. Left as an exercise. □

2.3 The inverse of a matrix

We want to have a matrix which play the role of the inverse of a real number for the multiplication.

$$6 \cdot 6^{-1} = 6 \cdot 1/6 = 6^{-1} \cdot 6 = 1$$

We can compute the inverse only for SQUARE matrices $n \times n$.

Definition 2.3.1. *An $n \times n$ matrix A is said to be* **invertible** *if there is an $n \times n$ matrix denoted A^{-1} such that*

$$A^{-1}A = AA^{-1} = I_n$$

where I_n is the $n \times n$ identity matrix. The matrix A^{-1} is called the **inverse** *of A.*
A matrix that is not invertible is sometimes called a **singular matrix**, *and an invertible is called a* **non singular** *matrix.*

Example 2.3.2. *Let* $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ *and* $B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$. *Since*

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$BA = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

As a consequence, B is the inverse of A and $B = A^{-1}$.

Here is a simple formula for the inverse of a 2×2 matrix, along with a test to tell if the inverse exists.

Theorem 2.3.3. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *If* $ad - bc \neq 0$, *then* A *is invertible and*

$$A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

If $ad - bc = 0$, *then* A *is not invertible. The quantity* $ad - bc$ *is called the* **determinant** *of* A, *and we write*

$$det(A) = ad - bc$$

Proof. Left in exercise. □

Definition 2.3.4. *The quantity* $ad - bc$ *is called the* **determinant** *of* A, *and we write*

$$det(A) = ad - bc$$

Note that the theorem can now be retranslated as A *is invertible if and only if* $det(A) \neq 0$.

Example 2.3.5. *Find the inverse of*

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}.$$

Solution: *We compute the determinant:*

$$det(A) = 3 \times 2 - 1 \times 1 = 5 \neq 0$$

So A *is invertible and*

$$A^{-1} = \frac{1}{5} \begin{pmatrix} -2 & 1 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} -2/5 & 1/5 \\ -1/5 & 3/5 \end{pmatrix}$$

Invertible matrix are indispensable in linear algebra see the next theorem but also for mathematical model of a real-life situation.

Theorem 2.3.6. *If* A *is invertible* $n \times n$ *matrix, then for each* b *in* $\mathbb{R}^n$, *the equation* $Ax = b$ *has the unique solution* $x = A^{-1}b$.

Proof. Take any $b \in \mathbb{R}^n$. A solution exists indeed $A^{-1}b$ is one. Since

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

We still need to prove that the solution is unique. Indeed, if u is a solution, we will prove that $u = A^{-1}b$.
Then, let u be a solution, then $Au = b$. By multiplying the equality by A^{-1}, we get

$$A^{-1}(Au) = A^{-1}b$$

And

$$A^{-1}(Au) = (A^{-1}A)u = I_n u = u$$

So that,

$$u = A^{-1}b$$

□

The formula of the previous theorem is not really useful to solve system, since it takes more time computing the inverse usually than just doing the row reduction we have learnt in the previous chapters. But nevertheless it is faster to use the inverse formula for 2×2 matrix in order to solve a system.

Example 2.3.7. *Solve the system*

$$\begin{cases} 3x_1 + & x_2 = 1 \\ x_1 + & 2x_2 = 0 \end{cases}$$

using the formula for the inverse we have seen.
Note that this system is equivalent to

$$\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

We have see that the inverse of

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$$

is

$$A^{-1} = \begin{pmatrix} -2/5 & 1/5 \\ -1/5 & 3/5 \end{pmatrix}$$

As a consequence we know that the system has a unique solution given by

$$u = \begin{pmatrix} -2/5 & 1/5 \\ -1/5 & 3/5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -2/5 \\ -1/5 \end{pmatrix}$$

Theorem 2.3.8. *1. If A is an invertible matrix, then A^{-1} is invertible and*

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

2. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

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

(The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.)

3. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is,

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

Proof. 1. Since $A^{-1}A = AA^{-1} = I_n$, by definition of the inverse of A^{-1}, we have that $(A^{-1})^{-1} = A$.

2. Also note that

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

Similarly

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

So by definition of the inverse,

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

3. Finally, note that

$$A^T(A^{-1})^T = (A^{-1}A)^T = I_n$$

Also

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

So

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

□

2.4 Elementary matrix and inverse

Definition 2.4.1. *An* **elementary matrix** *is one that is obtained by performing a single elementary row operation on the identity matrix.*

Example 2.4.2. *Let* $E_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$, $E_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $E_3 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ *and* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Compute* E_1A, E_2A *and* E_3A *and describe how these products can be seen as elementary row operation on* A.

Solutions: *We have*

$$E_1A = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c-2a & d-2b \end{pmatrix}$$

This multiplication is by E_1 *is equivalent to do* $R_2 \leftarrow R_2 - 2R_1$.

$$E_2A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}$$

This multiplication is by E_2 *is equivalent to do* $R_2 \leftrightarrow R_2$.

$$E_3A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix}$$

This multiplication is by E_3 *is equivalent to do* $R_2 \leftarrow 2R_2$.

Fact 2.4.3. *If an elementary row operation is performed on an* $m \times n$ *matrix* A, *the resulting matrix can be written as* EA, *where the* $m \times n$ *matrix* A, *the resulting matrix can be written as* EA, *where* $m \times n$ *matrix can be written as* EA, *where the* $m \times m$ *matrix* E *is created by performing the same row operation on* I_m.

Fact 2.4.4. *Every elementary matrix E is invertible. The inverse of E os the elementary matrix of the same type that transforms E back into I.*

Proof. Indeed, this can be obtained by noticing that every row operation is revertible. □

Example 2.4.5. *Find the inverse of* $E_1 = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$.

Solution: *To transform* E_1 *into identity, add* $+2$ *times row* 1 *to row* 2*. The elementary matrix that does this is*

$$E_1^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

Theorem 2.4.6. *An* $n \times n$ *matrix* A *is invertible if and only if* A *is row equivalent to* I_n *and in this case. any sequence of elementary row operation that reduces* A *to* I_n *also transform* I_n *into* A^{-1}.

Proof. Suppose that A is invertible. Then since the equation $Ax = b$ has a solution for each b, then A has a pivot position in each row. Because A is square, the n pivot positions must be on the diagonal, which implies the reduced echelon form of A is I_n. That is, $A \sim I_n$.
Now, suppose $A \sim I_n$. Then there is a sequence of row operation that transforms A into I_n, that is the same as the existence of elementary matrix E_i such that $E_1 \cdots E_p A = I_n$. So that $A = (E_1 \cdots E_p)^{-1}$ and $A^{-1} = E_1 \cdots E_p$. So that A is invertible. □

Fact 2.4.7 (Algorithm for finding A^{-1})**.** *Row reduce the augmented matrix* $[A, I]$*. If* A *is row equivalent to* I*, then* $[A, I]$ *is row equivalent to* $[I, A^{-1}]$*, Otherwise,* A *does not have an inverse.*

Example 2.4.8. *Find the inverse of* $A = \begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix}$ *using the algorithm above.*

Solution:

$$[A, I_n] = \begin{pmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & 1 \end{pmatrix} \sim_{R_2 \leftarrow R_2 + 2R_1} \begin{pmatrix} 1 & 0 & 1 & 3 \\ 2 & 1 & 0 & 7 \end{pmatrix}$$
$$\sim_{R_2 \leftarrow 1/7 R_2} \begin{pmatrix} 1 & 0 & 1 & 3 \\ 2/7 & 1/7 & 0 & 1 \end{pmatrix}$$
$$\sim_{R_1 \leftarrow R_1 - 3R_2} \begin{pmatrix} 1/7 & -3/7 & 1 & 0 \\ 2/7 & 1/7 & 0 & 1 \end{pmatrix}$$

So, A *is invertible, since* $A \sim I$ *So that*

$$A^{-1} = \begin{pmatrix} 1/7 & -3/7 \\ 2/7 & 1/7 \end{pmatrix}$$

(You can double check your result by making sure that $AA^{-1} = A^{-1}A = I$*)*

Theorem 2.4.9 (The invertible matrix theorem)**.** *Let* A *be a square* $n \times n$ *matrix. Then the following statements are equivalent. That is, for a given* A*, the statement are either all true or all false.*

1. *A is an invertible matrix*
2. *A is row equivalent to the $n \times n$ identity matrix.*
3. *A has n pivot positions.*
4. *The equation $Ax = 0$ has only the trivial solution*
5. *The columns of A form a linearly independent set.*
6. *The linear transformation $x \mapsto Ax$ is one-to-one.*
7. *The equation $Ax = b$ has at least one solution for each b in $\mathbb{R}^n$.*
8. *The equation $Ax = b$ has a unique solution for each b in $\mathbb{R}^n$.*
9. *The columns of A span $\mathbb{R}^n$.*
10. *The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.*
11. *There is an $n \times n$ matrix C such that $CA = I$.*
12. *There is an $n \times n$ matrix D such that $AD = I$.*
13. *A^T is an invertible matrix.*

Proof. Left as an exercise. □

Fact 2.4.10. *Let A and B be square matrices. If $AB = I$, then A and B are both invertible, with $B = A^{-1}$ and $A = B^{-1}$.*

Exercise 2.4.11. *Use the invertible matrix theorem to decide if A is invertible*

$$A = \begin{pmatrix} 2 & 0 & -4 \\ 6 & 0 & -12 \\ 0 & 1 & 0 \end{pmatrix}$$

Solution: *We row reduce this matrix:*

$$A = \begin{pmatrix} 2 & 0 & -4 \\ 6 & 0 & -12 \\ 0 & 1 & 0 \end{pmatrix} \sim^{R_2 \leftarrow R_2 - 3R_1} \begin{pmatrix} 2 & 0 & -4 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \sim^{R_2 \leftrightarrow R_3} \begin{pmatrix} 2 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Since A do not have a pivot position in each row we know by The invertible matrix theorem that A is not invertible.

Definition 2.4.12. *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is* **invertible** *if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$S \circ T = T \circ S = I_n$$

That is, for all $x \in \mathbb{R}^n$, $S(T(x)) = x$ and $T(S(x)) = x$.
The next theorem shows that if such an S exists, it is unique and must be a linear transformation. We call S the **inverse** *of T and write it as T^{-1}.*

Theorem 2.4.13. *Let $T : \mathbb{R} \to \mathbb{R}^n$ be a linear transformation and let A be the standard matrix for T. Then T is invertible if and only if A is an invertible matrix. In that case the linear transformation S given by $S(x) = A^{-1}x$ is the unique function satisfying*

$$S \circ T = T \circ S = I_n$$

Proof. Suppose that T is invertible. Then in particular, the map $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. And A is invertible by the invertible matrix theorem.
Conversely, if A is invertible, let $S : \mathbb{R}^n \to \mathbb{R}^n$ be the linear transformation defined by $x \mapsto A^{-1}x$. And, for all $x \in \mathbb{R}^n$, $S(T(x)) = S(Ax) = A^{-1}Ax = x$. Similarly, $T(S(x)) = x$ and T is invertible.
The proof that the inverse is unique is left as an exercise.

Example 2.4.14.

What can you say about a one to one linear transformation T from $\mathbb{R}^n$ into $\mathbb{R}^n$?
Solution: The column of the standard matrix A of T are linearly independent. So A is invertible, by the Invertible matrix theorem, and T maps $\mathbb{R}^n$ onto $\mathbb{R}^n$. Also, T is invertible. □

2.5 Matrix factorizations

Definition 2.5.1. *A* **factorization** *of a matrix A is an equation that expresses A as a product of two or more matrices. Whereas matrix multiplication involves a* **synthesis** *of data (combining the effects of two or more linear transformations into a single matrix), matrix factorization is an* **analysis** *of data. In the language of computer science, the expression of A as a product amounts to a* **preprocessing** *of the data in A, organizing that data into two or more parts whose structures are more useful in some way, perhaps more accessible for computation.*

2.5.1 The LU factorization

The LU factorization, described below is motivated by the fairly common industrial and business problem of solving a sequence of equations, all with the same coefficient matrix

$$Ax = b_1, \ Ax = b_2, \ \cdots \ , Ax = b_p$$

When A is invertible, one could compute A^{-1} and then compare $A^{-1}b_1$, $A^{-1}b_2$ and so on. However it is more efficient to solve the first equation sequence

$$Ax = b_1, \ Ax = b_2, \ \cdots \ , Ax = b_p$$

by row reduction and obtain an LU factorization of A at the same time. Thereafter, the remaining equations in the previous sequence are solved with the LU factorization.

Definition 2.5.2. *At first, assume that A is and $m \times n$ matrix that can be row reduced to an echelon form without row interchanges. Then A can be written in the form $A = LU$ where L is an $m \times m$ lower triangular matrix with $1's$ on the diagonal U is an $m \times n$ echelon form of A. Such a factorization is called an LU* **factorization** *of A. The* *matrix L is invertible and is called a* **unit lower triangular matrix**.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{pmatrix} \begin{pmatrix} \square & * & * & * & * \\ 0 & \square & * & * & * \\ 0 & 0 & 0 & \square & * \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = LU$$

Why the LU decomposition is it even useful? Well, take $A = LU$, the equation $Ax = b$ can be written as $L(Ux) = b$. Writing y for Ux, we can find x by solving the pair of equations

$$Ly = b, \; Ux = y$$

First, solve $Ly = b$ for y, and the solve $Ux = y$ for x. Each equation is easy to solve because L and U is triangular.

The computational efficiency of the LU factorization depends on knowing L and U. The next algorithm shows that the row reduction of A to an echelon form U amounts to an LU factorization because it produces L with essentially no extra work. After the first row reduction, L and U are available for solving additional equations whose coefficient matrix is A.

Algorithm 2.5.3. *(An LU factorization algorithm) Suppose A can be reduced to an echelon form U using only row replacements that add a multiple of one row to another row below it. In this case, there exist unit lower triangular elementary matrices $E_1, \cdots, E_p$ such that*

$$E_p \cdots E_1 A = U$$

Then

$$A = (E_p \cdots E_1)^{-1} U = LU$$

where L $= (E_p \cdots E_1)^{-1}$ is a unit lower triangular. (Indeed, it can be proven that products and inverse of units lower triangular matrices are also unit lower triangular). Note that thus L is invertible.

Note that the row operations which reduces A to U, also reduce the L to I, because

$$(E_p \cdots E_1)L = (E_p \cdots E_1)(E_p \cdots E_1)^{-1} = I$$

This observation is the key to constructing L.
Of course, it is not always possible to put A in an echelon form only with row replacement, but when it is, the argument above shows that an LU factorization exists.
In practical work, row interchanges are nearby always needed, because partial pivoting is used for high accuracy. (Recall that this procedure selects among the possible choices for a pivot, an entry in the column having the largest absolute value.) To handle row interchanges, the LU factorization above can be modified easily to produce an L that is permuted lower triangular, in the sense that a rearrangement (called a permutation) of the rows of L can make L (unit) lower triangular. The resulting permuted LU factorization solves $Ax = b$ in the same way as before, except that the reduction of $[L, b]$ to $[I, y]$ follows the order of the pivots in L from left to right, starting with the pivot in

the first column. A reference to an "LU factorization" usually includes the possibility that L might be permuted lower triangular.

Example 2.5.4. *Let* $A = \begin{pmatrix} 3 & -7 & -2 \\ -3 & 5 & 1 \\ 6 & -4 & 0 \end{pmatrix}$ *and* $b = \begin{pmatrix} -7 \\ 5 \\ 2 \end{pmatrix}$.

1. *Find an LU factorization for A.*
2. *Solve the equation $Ax = b$ using the previous question.*

Solution:

1.
$$A = \begin{pmatrix} 3 & -7 & -2 \\ -3 & 5 & 1 \\ 6 & -4 & 0 \end{pmatrix} \sim_{R_2 \to R_2 + R_1 \text{ and } R_3 \to R_3 - 2R_1} \begin{pmatrix} 3 & -7 & -2 \\ 0 & -2 & -1 \\ 0 & 10 & 4 \end{pmatrix}$$
$$\sim_{R_3 \to R_3 + 5R_2} \begin{pmatrix} 3 & -7 & -2 \\ 0 & -2 & -1 \\ 0 & 0 & -1 \end{pmatrix} = U$$

Then one can obtain L by performing the reverse operation in reverse order to I, we get
$$L = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -5 & 1 \end{pmatrix}$$
One can easily check that we have
$$A = LU = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & -5 & 1 \end{pmatrix} \begin{pmatrix} 3 & -7 & -2 \\ 0 & -2 & -1 \\ 0 & 0 & -1 \end{pmatrix}$$

2. *In order to solve the system using the matrix factorization, we solve first the system $Ly = b$ and then $Ux = y$.*
The augmented matrix corresponding to the system is
$$[L, b] = \begin{pmatrix} 1 & 0 & 0 & -7 \\ -1 & 1 & 0 & 5 \\ 2 & -5 & 1 & 2 \end{pmatrix} \sim_{R_2 \to R_2 + R_1 \text{ and } R_3 \to R_3 - 2R_1} \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -2 \\ 0 & -5 & 1 & 16 \end{pmatrix}$$
$$\sim_{R_3 \to R_3 + 5R_2} \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 6 \end{pmatrix}$$

As a consequence $y = \begin{pmatrix} -7 \\ -2 \\ 6 \end{pmatrix}$. *Next, solve $Ux = y$, using back-substitution (with* *matrix notation).*

$$
\begin{aligned}
[U, y] = \begin{pmatrix} 3 & -7 & -2 & -7 \\ 0 & -2 & -1 & -2 \\ 0 & 0 & -1 & 6 \end{pmatrix} &\sim_{R_3 \to -R_3} \begin{pmatrix} 3 & -7 & -2 & -7 \\ 0 & -2 & -1 & -2 \\ 0 & 0 & 1 & -6 \end{pmatrix} \\
&\sim_{R_2 \to R_2 + R_2} \begin{pmatrix} 3 & -7 & 0 & -19 \\ 0 & -2 & 0 & -8 \\ 0 & 0 & 1 & -6 \end{pmatrix} \\
&\sim_{R_2 \to -1/2 R_2} \begin{pmatrix} 3 & -7 & 0 & -19 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -6 \end{pmatrix} \\
&\sim_{R_1 \to R_1 + 7R_2} \begin{pmatrix} 3 & 0 & 0 & 9 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -6 \end{pmatrix} \\
&\sim_{R_1 \to 1/3 R_1} \begin{pmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -6 \end{pmatrix}
\end{aligned}
$$

So

$$x = \begin{pmatrix} 3 \\ 4 \\ -6 \end{pmatrix}.$$

2.5.2 A matrix factorization in electrical engineering

Matrix factorization is intimately related to the problem of constructing an electrical network with specified properties. The following discussion gives just a glimpse of the connection between factorization and circuit design. Consider an electric circuit, with an imput and an output. Record the imput voltage and output. Record the input voltage and current by $\begin{pmatrix} v_1 \\ i_1 \end{pmatrix}$ (with voltage v in volts and currents i in amps), and record the output voltage and current by $\begin{pmatrix} v_2 \\ i_2 \end{pmatrix}$. Frequently, the transformation $\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} \mapsto \begin{pmatrix} v_2 \\ i_2 \end{pmatrix}$ is linear. That is, there is a matrix A, called the transfer matrix, such that

$$\begin{pmatrix} v_2 \\ i_2 \end{pmatrix} = A \begin{pmatrix} v_1 \\ i_1 \end{pmatrix}$$

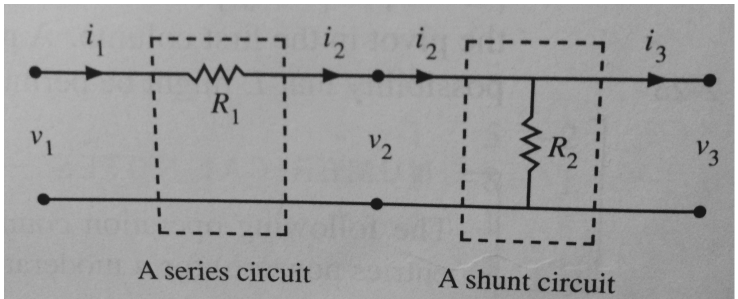

FIGURE 4 A ladder network.

The figure above shows a **ladder network**, where two circuits (there could be more) are connected in series, so that the output of one circuit becomes the input of the next circuit. The left circuit is called a **series circuit**, with resistance R_1 (in ohms).
The right circuit is a **shunt circuit**, with resistance R_2. Using Ohm's law and Kirchhoff's law, one can show that the transfer matrices of the series and shunt circuits, respectively, are

$$\begin{pmatrix} 1 & -R_1 \\ 0 & 1 \end{pmatrix}$$

This is the transfer matrix of series circuit, and

$$\begin{pmatrix} 1 & 0 \\ -1/R_2 & 1 \end{pmatrix}$$

This is the transfer matrix of shunt circuit.

Example 2.5.5. *1. Compute the transfer matrix of the ladder network of the above figure above.*

2. Design a ladder network whose transfer whose transfer matrix is

$$\begin{pmatrix} 1 & -2 \\ -0.2 & 7/5 \end{pmatrix}$$

Solution:

1. *Let A_1 and A_2 be the transfer matrices of the series and shunt circuit, respectively. Then an input vector x is transformed first into A_1x and then into $A_2(A_1x)$. The second connection of the circuits corresponds to composition of linear transformations, and the transfer matrix of the ladder network is (note order)*

$$A_2A_1 = \begin{pmatrix} 1 & 0 \\ -1/R_2 & 1 \end{pmatrix}\begin{pmatrix} 1 & -R_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -R_1 \\ -1/R_2 & 1+R_1/R_2 \end{pmatrix}$$

2. *To factor the matrix $\begin{pmatrix} 2 & -2 \\ -0,2 & 7/5 \end{pmatrix}$ into a product of transfer matrices, from the previous question we see that we need to have*

$$\begin{pmatrix} 1 & -R_1 \\ -1/R_2 & 1+R_1/R_2 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -0.2 & 7/5 \end{pmatrix}$$

From the $(1,2)$-entry, $R_1 = 2$ and from the $(2,1)$-entry we get $R_2 = 5$. With these values, the network has the desired transfer matrix.

2.6 The Leontief input-output model

Linear algebra played an essential role in the Nobel prize-winning work of Wassily Leontief. The economic model described in this section is the basis for more elaborate models used in many part of the world. Suppose a nation's economy is divided into n sectors that produce goods or services, and let x be a **production vector** in $\mathbb{R}^n$ that lists the output of each sector for one year. Also, suppose another part of the economy (called the open sector) does not produce goods or services but only consume them, and let d be a **final demand vector** (or **bill or final demands**) that lists the values of the goods and services demanded from various sectors by the nonproductive part of the economy. The vector d can represent consumer demand, government consumption, surplus production, exports, or other external demands.

As the various sectors produce goods to meet consumer demand, the producers themselves create additional **intermediate demand** for goods they need as inputs for their own production. The interrelations between the sectors are very complex, and the connection between the final demand and the production is unclear. Leontief asked if there is a production level x such that the amount produced (or "supplied") will exactly balance the total demand for that production, so that

$$\{\text{ Amount produced : x}\} = \{\text{ Intermediate demand }\} + \{\text{ Final demand : d }\}$$

The basic assumption of Leontief's input-output model is that for each section sector, there is a **unit consumption vector** in $\mathbb{R}^n$ that lists the inputs needed per unit of output of the sector. All input and output units are measured in millions of dollars, rather than in quantities such as tons or bushels. (Prices of good and services are held constant.)

As a simple example, suppose the economy is divided into three sectors -manufacturing, agriculture, and services. For each unit of output, manufacturing requires $0,10$ unit from other companies in that sector, 0.30 unit from agriculture and 0.30 unit from services. For each unit of output, agriculture uses 0.20 unit of its own output, $0,60$ unit from manufacturing, and $0,10$ unit from services. For each unit of output, the services sector consumes 0.10 unit from services, 0.60 unit from manufacturing, but no agricultural products.

Purshased from :	*Input consumed per unit of output* *Manufacturing*	*Agriculture*	*Services*
Manufacturing	0.10	0.60	0.60
Agriculture	0.30	0.20	0
Services	0.30	0.10	0.10
	c_1	c_2	c_3

1. What amounts will be consumed by the manufacturing sector if it decides to produce 10 units?

Solution: Compute:

$$10c_1 = 10\begin{pmatrix} 0.10 \\ 0.30 \\ 0.30 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}$$

To produce 10 units, manufacturing will order (i.e. "demand") and consume 1 unit from the other parts of the manufacturing sector, 3 units from agriculture, and 3 unit from services.

If manufacturing decides to produce x_1 units of output, then x_1c_1 represents the intermediate demands of manufacturing, because the amounts in x_1c_1 will be consumed in the process of creating the x_1 unit of output. Likewise, if x_2 and x_3 denote the planned output of the agriculture and services sectors, x_2c_2 and x_3c_3 list their corresponding intermediate demands. The total intermediate demand from all three sectors is given by

$$\{\text{ intermediate demand }\} = x_1c_1 + x_2c_2 + x_3c_3 = Cx$$

where C is the **consumption matrix** $[c_1, c_2, c_3]$, namely,

$$C = \begin{pmatrix} 0.10 & 0.60 & 0.60 \\ 0.30 & 0.20 & 0 \\ 0.30 & 0.10 & 0.10 \end{pmatrix}$$

Equations yield Leontief's model :

The Leontief input-output model, or production equation

$$\{\text{Amount produced}\} = \{\text{ Intermediate demand }\} + \{final\ demand\}$$

which is also

$$x = Cx + d$$

This can be rewritten as $Ix - Cx = d$, or $(I - C)x = d$.

2. Construct the consumption matrix for this economy and determine what intermediate demands are created if agriculture plans to produce 100 units and the others nothing.

   Solution: The answer to this exercise will depend upon the order in which the student chooses to list the sectors. The important fact to remember is that each column is the unit consumption vector for the appropriate sector. If we order the sector manufacturing, agriculture, and services, then the consumption matrix is as we have said above:

$$C = \begin{pmatrix} 0.10 & 0.60 & 0.60 \\ 0.30 & 0.20 & 0 \\ 0.30 & 0.10 & 0.10 \end{pmatrix}$$

The intermediate demands created by the production vector x are given by Cx. Thus in this case the intermediate demand is

$$Cx = \begin{pmatrix} 0.10 & 0.60 & 0.60 \\ 0.30 & 0.20 & 0 \\ 0.30 & 0.10 & 0.10 \end{pmatrix} \begin{pmatrix} 0 \\ 100 \\ 0 \end{pmatrix} = \begin{pmatrix} 60 \\ 20 \\ 10 \end{pmatrix}$$

3. Determine the production levels needed to satisfy a final demand of 20 units for agriculture, with no final demand for the other sectors. (Do not compute an inverse matrix).

   Solution: Solve the equation $x = Cx + d$ for $d = \begin{pmatrix} 0 \\ 20 \\ 0 \end{pmatrix}$.

$$d = \begin{pmatrix} 0 \\ 20 \\ 0 \end{pmatrix} = x - Cx = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \begin{pmatrix} 0.10 & 0.60 & 0.60 \\ 0.30 & 0.20 & 0 \\ 0.30 & 0.10 & 0.10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.9x_1 - 0.6x_2 - 0.6x_3 \\ -0,3x_1 + 0.8x_2 \\ -0.3x_1 - 0.1x_2 + 0.9x_3 \end{pmatrix}$$

   This system of equations has the augmented matrix:

$$\begin{pmatrix} 0.90 & -0.60 & -0.60 & 0 \\ -0.30 & 0.80 & 0.00 & 0.20 \\ -0.30 & -0.10 & 0.90 & 0 \end{pmatrix} \sim \text{ Row reduce } \sim \begin{pmatrix} 1 & 0 & 0 & 37.03 \\ 0 & 1 & 0 & 38.89 \\ 0 & 0 & 1 & 16.67 \end{pmatrix}$$

   So, the production level needed is

$$\begin{pmatrix} 37.03 \\ 38.89 \\ 16.67 \end{pmatrix}$$

If the matrix $I - C$ is invertible, then we know by a theorem we have proven earlier, with A replaces by $(I - C)$, and from the equation $(I - C)x = d$ obtain $x = (I - C)^{-1}d$.

2.7 Application to computer graphics

Computer graphics are images displayed or animated on a computer screen. Application of computer graphics are widespread and growing rapidly. For instance, computed aided design (CAD) is an integral part of many engineering processes, such as the aircraft design process. The entertainment industry has made the most spectacular use of computer graphics, from the special effects in the Matrix to playstation 2 and the Xbox.

Here we will examine some of the basic mathematic used to manipulate and display graphical images such as wire-frame model of an airplane. Such an image (or picture) consists of a number of points, connecting lines and curves, and information about how to fill in close regions bounded by the lines and curves. Often, curved lines are approximated by short straight-line segment, and a figure is defined mathematically by a list of points.

Among the simplest $2D$ graphics symbols are letters used for labels on the screen. Some letters are stored as wire-frame objects; others that have curved portions are stored with additional mathematical formula for the curves.

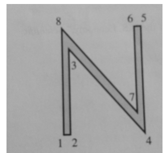

FIGURE 1
Regular N.

Example 2.7.1. *The capital letter N is determined by eight points, or vertices. The coordinates of the points can be stored in a data matrix D.*

$$D = \begin{pmatrix} 0 & 0.5 & 0.5 & 6 & 6 & 5.5 & 5.5 & 0 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{pmatrix}$$

Where the first row is the x-coordinate and the second row is the y-coordinate. And the column C_i correspond to the vertex i of figure 1.
In addition to D, it is necessary to specify which vertices are connected by lines, but we omit this detail.

The main reason graphical object are described by collections of straight line segment is that the standard transformations in computer graphics map line segments onto other line segments. Once the vertices that describe an object have been transformed, their images can be connected with the appropriate straight lines to produce the complete image of the original object.

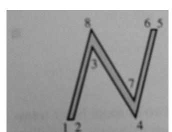

FIGURE 2
Slanted N.

Example 2.7.2. *Given $A = \begin{pmatrix} 1 & 0.25 \\ 0 & 1 \end{pmatrix}$, describe the effect of the shear transformation $x \mapsto Ax$ on the letter N.*
Solution: *By definition of matrix multiplication, the columns of the product AD contain the images of the vertices of the letter N*

$$D = \begin{pmatrix} 0 & 0.5 & 2.105 & 6 & 8 & 7.5 & 5.895 & 2 \\ 0 & 0 & 6.42 & 0 & 8 & 8 & 1.58 & 8 \end{pmatrix}$$

The transformed vertices are plotted in Fig. 2, along with connecting line segments that correspond to those in the original figure.

The italic N in Fig. 2, looks a bit too wide. To compensate, shrink the width by a scale transformation that affects the x-coordinates of the points.

Example 2.7.3. *Compute the matrix of transformation that performs the shear transformation of the previous example and then scale all x coordinates by a factor of 0.75.*
Solution: *The matrix that multiplies the x-coordinate of a point by 0.75 is*

$$S = \begin{pmatrix} 0.75 & 0 \\ 0 & 1 \end{pmatrix}$$

So the matrix of the composite transformation is

$$SA = \begin{pmatrix} 0.75 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0.25 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.75 & 0.1875 \\ 0 & 1 \end{pmatrix}$$

The result of this composite is shown in

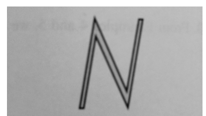

FIGURE 3
Composite transformation of *N*.

2.7.1 Homogeneous coordinate

The mathematics of computer graphics is intimately connected with matrix multiplication. Unfortunately, translating an object on a screen does not correspond directly to matrix multiplication because translation is not a linear transformation. The standard way to avoid this difficulty is to introduce what are called **homogeneous coordinates**.

Definition 2.7.4. *Each point (x, y) in $\mathbb{R}^2$ can be identified with the point $(x, y, 1)$ on the plane $\mathbb{R}^3$ that lies one unit above the xy-plane. We say that (x, y) has* **homogeneous coordinates** *$(x, y, 1)$. Homogeneous coordinates for points are not added or multiplied by scalars, but they can be transformed via multiplication by 3×3 matrices.*

Example 2.7.5. *A translation of the form $(x, y) \mapsto (x+h, y+k)$ is written in homogeneous coordinates as $(x, y, 1) \mapsto (x+h, y+k, 1)$. This transformation can be computed via matrix multiplication:*

$$\begin{pmatrix} 1 & 0 & h \\ 0 & 1 & k \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x+h \\ y+k \\ 1 \end{pmatrix}$$

Some of the newest and most exciting work in computer graphics is connected with molecular modeling. With $3D$ graphics, a biologist can examine simulated protein molecule and search for active sites that might accept a drug molecule.
By analogy with the $2D$ case, we say that $(x, y, z, 1)$ are homogeneous coordinates for the point (x, y, z) in $\mathbb{R}^3$. In general, (X, Y, Z, H) are **homogeneous coordinates** for (x, y, z), if $H \neq 0$ and

$$x = X/H, \;\; y = Y/H, \text{ and } z = Z/H$$

Each nonzero scalar multiple of $(x, y, z, 1)$ gives a set of homogeneous coordinates for points (x, y, z).
The next example illustrates the transformations used in molecular modeling to move a drug into a protein molecule.

Example 2.7.6. *Give* 4×4 *matrices for the following transformations:*

1. *Rotation about the* y*-axis through an angle of* θ*. (By convention, a positive angle is the counterclockwise direction when looking toward the origin from the positive half of the axis of rotation, in this case, the* y *axis.)*
2. *Translation by the vector* $p = (1, 5, 8)$*.*

Solution:

1. *First, construct the* 3×3 *matrix for the rotation. The vector* e_1 *rotates down toward the negative* z*-axis, stopping at* $(cos(\theta), 0, , -sin(\theta))$*, the vector* e_2 *on the* y*-axis does not move, but* e_3 *on the* z*-axis rotates down toward the positive* x*-axis, stopping at* $(sin(\theta), 0, cos(\theta))$*. The standard matrix for this rotation is*

$$\begin{pmatrix} cos(\theta) & 0 & -sin(\theta) \\ 0 & 1 & 0 \\ sin(\theta) & 0 & cos(\theta) \end{pmatrix}$$

So the rotation matrix for homogeneous coordinates is

$$\begin{pmatrix} cos(\theta) & 0 & -sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ sin(\theta) & 0 & cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. *We want to map* $(x, y, z, 1)$ *to map to* $(x+1, y+5, z+8, 1)$*. The matrix that does this is*

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Chapter 3

Determinant

3.1 Introduction to determinants

Definition 3.1.1. *For $n \leqslant 2$, the determinant of an $n \times n$ matrix $A = [a_{i,j}]$ is the sum of n terms of the form $\pm det(A_{1,j})$, with plus and minus signs alternating, where the entries $a_{1,1}, a_{1,2}, \cdots, a_{1,n}$ are from the first row of A and the matrix $A_{1,j}$ is obtained by crossing out the first row and the j column. In symbols,*

$$\begin{array}{rcl} det(A) & = & a_{1,1}det(A_{1,1}) - a_{1,2}det(A_{1,2}) + \cdots + (-1)^{n+1}a_{1,n}det(A_{1,n}) \\ & = & \sum_{j=1}^{n}(-1)^{1+j}a_{1,j}det(A_{1,j}) \end{array}$$

Notation: One can write $|a_{i,j}|$ instead of $det([a_{i,j}])$.

Given $A = [a_{i,j}]$, the (i,j)**-cofactor** of A is the number $C_{i,j}$ given by:

$$C_{i,j} = (-1)^{i+j}det(A_{i,j})$$

Then

$$det(A) = a_{1,1}C_{1,1} + a_{1,2}C_{1,2} + \cdots + a_{1,n}C_{1,n}$$

The formula that follows is called a **cofactor expansion across the first row of** A.

Theorem 3.1.2. *The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactors is:*

$$det(A) = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n}$$

The cofactor expansion down the jth column is

$$det(A) = a_{1,j}C_{1,j} + a_{2,,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}$$

The plus minus sign in the (i,j) depends on the position of $a_{i,j}$ in the matrix regardless of the sign of $a_{i,j}$ itself. The factor $(-1)^{i+j}$ determines the following checkerboard pattern of sign

$$\begin{pmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Example 3.1.3. *Compute the determinant of*

$$A = \begin{pmatrix} 5 & -2 & 4 \\ 0 & 3 & -5 \\ 2 & -4 & 7 \end{pmatrix}$$

Solution:
Compute

$$det(A) = a_{1,1}det(A_{1,1}) - a_{1,2}det(A_{1,2}) + a_{1,3}det(A_{1,3})$$

$$\begin{aligned} det(A) &= 5 \cdot det(\begin{pmatrix} 3 & -5 \\ -4 & 7 \end{pmatrix}) - 2det(\begin{pmatrix} 0 & -5 \\ 2 & 7 \end{pmatrix}) + 4det(\begin{pmatrix} 0 & 3 \\ 2 & -4 \end{pmatrix}) \\ &= 5 \cdot \begin{vmatrix} 3 & -5 \\ -4 & 7 \end{vmatrix} + 2\begin{vmatrix} 0 & -5 \\ 2 & 7 \end{vmatrix} + 4\begin{vmatrix} 0 & 3 \\ 2 & -4 \end{vmatrix} \\ &= 5(3 \times 7 - (-5) \times (-4)) + 2(0 \times 7 - (-5) \times 2) + 4(0 \times (-4) - 2 \times 3) \\ &= 5 + 20 - 24 = 1 \end{aligned}$$

Example 3.1.4. *Compute* $det(A)$*, where*

$$A = \begin{pmatrix} 2 & 4 & 3 & 5 \\ 0 & 4 & 5 & 7 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Solution: *We compute this determinant using the cofactor expansion down the* $1rst$ *column is*

$$\begin{aligned} det(A) &= 2C_{1,1} + 0C_{2,1} + 0C_{3,1} + 0C_{4,1} = 2det(\begin{pmatrix} 4 & 5 & 7 \\ 0 & 1 & 8 \\ 0 & 0 & 3 \end{pmatrix}) \\ &= 2[4det(\begin{pmatrix} 1 & 8 \\ 0 & 3 \end{pmatrix}) + 0det(\begin{pmatrix} 5 & 7 \\ 0 & 3 \end{pmatrix})) + 0det(\begin{pmatrix} 5 & 7 \\ 1 & 8 \end{pmatrix}) \\ &= 2 \cdot 4 \cdot 1 \cdot 3 = 24 \end{aligned}$$

Note that this determinant is equal to the product of the terms on the diagonal.

More generally, with a similar expansion method we used in the previous example one can prove that

Theorem 3.1.5. *If A is a triangular matrix, then the* $det(A)$ *is the product of the entries on the main diagonal of A.*

Remarque 3.1.6. *By today's standard, a* 25×25 *matrix is small. Yet it would be impossible to calculate a* 25×25 *determinant by the cofactor expansion. In general, a cofactor expansion requires over* $n!$ *multiplications and* $25!$ *is approximately* $1.5 \cdot 10^{25}$ *. If a computer performs one trillion multiplications per second, it would have to run for* 500000 *year to compute a* 25×25 *determinant by this method. Fortunately there are faster methods that we shall see later.*

3.2 Properties of the determinants

An important property of determinants is that we can describe then easily after a elementary row operation change.

Theorem 3.2.1. *Let A be a square matrix.*

1. *If a multiple of one row of A is added to another row to produce a matrix B, then $det(B) = det(A)$.*
2. *If two row of A are interchanged to produce B, then $det(B) = -det(A)$.*
3. *If one row of A is multiplied by k to produce B, then $det(B) = k \cdot det(A)$.*

One can reformulate these assertions as follow using the elementary matrix properties: If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$det(EA) = det(E)det(A)$$

where

$$det(E) = \begin{cases} 1 \; if \; E \text{ is a row replacement} \\ -1 \; if \; E \text{ is an interchange} \\ r \; if \; E \text{ is a scale by } r \end{cases}$$

Proof. We will argue by induction on the size of A. I will propose the case 2×2 for you as an exercise. Suppose that the theorem has been verified for determinants of $k \times k$ matrices with $k \geqslant 2$, let $n = k+1$, and let A be a matrix of size $n \times n$. The action of E on A involves either two rows or only one. So we can expand $det(EA)$ across a row that is unchanged by the action of E, say row i since $n > 2$. Let $A_{i,j}$ (respectively $B_{i,j}$) be the matrix obtained by deleting row i and column j from A respectively EA.Then the row of $B_{i,j}$ are obtained from the rows of $A_{i,j}$ by the same type of elementary operation that E performs on A. Since these submatrices are only $k \times k$, the induction assumption implies that

$$det(B_{i,j}) = \alpha \cdot det(A_{i,j})$$

where $\alpha = 1, -1, or\ r$, depending on the nature of E. The cofactor expansion across row i is

$$\begin{array}{rcc} det(EA) & = & a_{i,1}(-1)^{i+1}det(B_{i,1}) + \cdots + a_{i,n}(-1)^{i+n}det(B_{i,n}) \\ & = & \alpha a_{i,1}(-1)^{i+1}det(A_{i,1}) + \cdots + \alpha a_{i,n}(-1)^{i+n}det(A_{i,n}) \\ & = & \alpha \cdot det(A) \end{array}$$

In particular, taking $A = I_n$, we see that I_n, we see that $det(E) = 1, -1,$ or r, depending on the nature of E. Thus the theorem is true for n, thee truth of this theorem for one value of n implies it truth for the next value of n. By the principle of induction, the theorem must be true for $n \geqslant 2$. The theorem is trivially true for $n = 1$. □

Suppose a square matrix A has been reduced to an echelon form U by row replacements and row interchange. (This is always possible.). If there is r interchanges, then from the previous theorem, shows that

$$det(A) = (-1)^r det(U)$$

Since U is an echelon form, it is triangular, and so $det(U)$ is the product of the diagonal entries $u_{i,i}$ are all pivots (because $A \sim I_n$ and the $u_{i,i}$ have not been scaled to $1's$). Otherwise, at least $u_{n,n}$ is zero, and the product $u_{1,1}, \cdots, u_{n,n}$ is zero. Thus,

$$det(A) = \begin{cases} (-1)^r \cdot (\text{ product of pivots in } U) & \text{when } A \text{ is invertible.} \\ 0 & \text{when } A \text{ is not invertible.} \end{cases}$$

It is interesting to note that although the echelon form U described above is not unique (because it is not completely row reduced), and the pivots are not unique, the product of the pivots is unique, except for a possible minus sign.
A common use of the previous theorem 3. in hand calculations is to factor out a common multiple of one row of a matrix. For intance,

$$\begin{vmatrix} * & * & * \\ ak & bk & ck \\ * & * & * \end{vmatrix} = k \begin{vmatrix} * & * & * \\ a & b & c \\ * & * & * \end{vmatrix}$$

where the starred entries are unchanged.

Example 3.2.2. *Compute the determinant using the previous theorem:*

$$\begin{vmatrix} 0 & 5 & -4 \\ 1 & 2 & 3 \\ 3 & 7 & 4 \end{vmatrix}$$

Solution:

$$\begin{aligned} \begin{vmatrix} 0 & 5 & -4 \\ 1 & 2 & 3 \\ 3 & 7 & 4 \end{vmatrix} &=_{R_1 \leftrightarrow R_2} - \begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \\ 3 & 7 & 4 \end{vmatrix} \\ &=_{R_3 \to R_3 - 3R_1} - \begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \\ 0 & 1 & -5 \end{vmatrix} \\ &=_{R_3 \to R_3 - 5R_2} - \begin{vmatrix} 1 & 2 & 3 \\ 0 & 5 & -4 \\ 0 & 0 & 15 \end{vmatrix} = -5 \cdot 15 = -75 \end{aligned}$$

Example 3.2.3. *Compute the determinant:*

$$\begin{vmatrix} 1 & 5 & 1 \\ 2 & 4 & 6 \\ 3 & 6 & 27 \end{vmatrix}$$

Solution:

$$\begin{aligned}
\begin{vmatrix} 1 & 5 & 1 \\ 2 & 4 & 6 \\ 3 & 6 & 27 \end{vmatrix} &= 2 \cdot 3 \cdot \begin{vmatrix} 1 & 5 & 1 \\ 1 & 2 & 3 \\ 1 & 2 & 9 \end{vmatrix} \\
&= {}_{R_2 \leftarrow R_2 - R_1 \text{ and } R_3 \leftarrow R_3 - R_1} 6 \begin{vmatrix} 1 & 5 & 1 \\ 0 & -3 & 2 \\ 0 & -3 & 8 \end{vmatrix} \\
&= {}_{R_3 \leftarrow R_3 - R_2} 6 \begin{vmatrix} 1 & 5 & 1 \\ 0 & -3 & 2 \\ 0 & 0 & 6 \end{vmatrix} \\
&= 6 \cdot 1 \cdot (-3) \cdot 6 = -108
\end{aligned}$$

Theorem 3.2.4. *A square matrix A is invertible if and only if $det(A) \neq 0$.*

We can perform operations on the columns of the matrix in a way that is analogous to the row operations we have considered. The next theorem shows that column operations have the same effects on determinants as row operations.
The theorem adds the statement $det(A) \neq 0$ to the Invertible Matrix Theorem. A useful corollary is that $det(A) = 0$ when the column of A are linearly dependent. Also, $det(A) = 0$ when the rows of A are linearly dependent. (Rows of A are columns of A^T, and linearly dependent columns of A^T make A^T singular. When A^T is singular, so is A, by the Invertible Matrix Theorem.) In practice, linear dependence is obvious when two rows or two columns are a multiple of the others. (See next theorem to see why column and rows can be both studied.)

Theorem 3.2.5. *If A is an $n \times n$ matrix, then $det(A^T) = det(A)$.*

Proof. The theorem is obvious for $n = 1$. Suppose the theorem is true for $k \times k$ determinants and let $n = k + 1$. Then the cofactor $a_{1,j}$ in A equals the cofactor of $a_{j,1}$ in A^T, because the cofactor involve $k \times k$ determinants. Hence, the cofactor expansion of $det(A)$ along the first row equals the cofactor expansion of $det(A^T)$ down the first column. That is, A and A^T have equal determinants. Thus the theorem is true for $n = 1$, and the truth of the theorem for one value of n implies its truth for the next step of $n + 1$. By the principle of induction, the theorem is true for all $n \geqslant 1$. □

Example 3.2.6. *Compute the determinant:*

$$\begin{vmatrix} 1 & 5 & 1 & 2 \\ -2 & -10 & -2 & -4 \\ 2 & 6 & 7 & 9 \\ 1 & 2 & 5 & 8 \end{vmatrix}$$

Solution: *Note that the second row is a equal to the double of the first row. So the row are linearly dependent so $det(A) = 0$.*

Computer can also handle large "sparse" matrices, with special routines that takes advantage of the presence of many zeros. Of course, zero entries can speed hand computations, too. The calculation in the next example combine the power of row operations with the strategy of using zero entries in cofactor expansions.

Example 3.2.7. *Compute the determinant:*

$$\begin{vmatrix} 0 & 5 & 1 & 2 \\ -2 & -10 & -2 & -4 \\ 0 & 6 & 7 & 9 \\ 0 & 2 & 0 & 0 \end{vmatrix}$$

Solution:

$$\begin{vmatrix} 0 & 5 & 1 & 2 \\ -2 & -10 & -2 & -4 \\ 0 & 6 & 7 & 9 \\ 0 & 2 & 0 & 0 \end{vmatrix} = -2\begin{vmatrix} 5 & 1 & 2 \\ 6 & 7 & 9 \\ 2 & 0 & 0 \end{vmatrix} = -2 \cdot 2 \cdot \begin{vmatrix} 1 & 2 \\ 7 & 9 \end{vmatrix} = -4 \cdot (7 - 18) = 44$$

Theorem 3.2.8 (Multiplicative property)**.** *If A and B are $n \times n$ matrices, then*

$$det(AB) = det(A) \cdot det(B)$$

Proof. If A is not invertible, then neither is AB (Exercise). In this case, $det(AB) = (det(A))(det(B))$, because both sides are zero. If A is invertible, then A and the identity matrix I_n are row equivalent by the Invertible Matrix Theorem. So there exist elementary matrices $E_1, \cdots, E_p$ such that

$$A = E_p E_{p-1} \cdots E_1 \cdot I_n = E_p E_{p-1} \cdots E_1$$

Then repeated application of the first theorem of this section, as rephrased above, shows that

$$|AB| = |E_p E_{p-1} \cdots E_1 B| = |E_p||E_{p-1} \cdots E_1 B| = \cdots = |E_p||E_{p-1}| \cdots |E_1||B| = |E_p E_{p-1} \cdots E_1| \cdot |B| = |A||B|$$

□

3.3 Cramer's rule, volume and linear transformations

3.3.1 Cramer's rule

Cramer's rule is needed in a variety of theoretical calculations. For instance, it can be used to study how the solution of $Ax = b$ is affected by changes in the entries of b. However, the formula is inefficient for hand calculations, except for 2×2 or perhaps 3×3 matrices.
For any $n \times n$ matrix A and any b in $\mathbb{R}^n$, let $A_i(b)$ be the matrix obtained from A by replacing column i by the vector b.

$$A_i(b) = [a_1, \cdots, b, \cdots, a_n]$$

Theorem 3.3.1. *Let A be an invertible $n \times n$ matrix. For any b in $\mathbb{R}^n$, the unique solution x of $Ax = b$ has entries given by*

$$x_i = \frac{det(A_i(b))}{det(A)}, \; i = 1, 2, \cdots, n$$

Proof. Denote the columns of A by $a_1, \cdots, a_n$ and the columns of the $n \times n$ identity matrix I by $e_1, \cdots, e_n$. If $Ax = b$, the definition of matrix multiplication shows that

$$A \cdot I_i(x) = A[e_1, \cdots, x, \cdots, e_n] = [Ae_1, \cdots, Ax, \cdots, Ae_n] = [a_1, \cdots, b, \cdots, a_n] = A_i(b)$$

By the multiplicative property of determinants.

$$(det(A)det(I_i(x))) = det(A_i(b))$$

The second determinant on the left is simply x_i. (Make a cofactor expansion along the *ith* row.) Hence,

$$(det(A)) \cdot x_i = det(A_i(b))$$

This proves the theorem because A is invertible. □

Example 3.3.2. *Use the Cramer's rule to solve the system*

$$\begin{cases} 5x_1 + 2x_2 = 2 \\ 2x_1 + x_2 = 1 \end{cases}$$

Solution: *View the system as $Ax = b$, with:*

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}, b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Using the notation introduced above.

$$A_1(b) = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}, A_2(b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

Since $det(A) = 1$, the system has a unique solution. By Cramer's rule.

$$x_1 = \frac{det(A_1(b))}{det(A)} = 0$$

$$x_2 = \frac{det(A_2(b))}{det(A)} = 1$$

So that the solution of the system is

$$x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

3.3.2 Application to Engineering

A number of important engineering problems, particularly in electrical engineering and control theory, can be analyzed by Laplace transforms. This approach converts an appropriate system of linear differential equations into a equations whose coefficients involve a parameter. The next example illustrates the type of algebraic system that may arise.

Example 3.3.3. *Consider the following system in which* s *is an unspecified parameter. Determine the values of* s *for which the system has a unique solution, and use Cramer's rule to describe the solution of*

$$\begin{cases} 5x_1 + 2x_2 = 2 \\ 2x_1 + sx_2 = 1 \end{cases}$$

Solution: *View the system as* $Ax = b$ *with,*

$$A = \begin{pmatrix} 5 & 2 \\ 2 & s \end{pmatrix}, b = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Using the notation introduced above.

$$A_1(b) = \begin{pmatrix} 2 & 2 \\ 1 & s \end{pmatrix}, A_2(b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$$

Since $det(A) = 5s - 4$, *the system has a unique solution. By Cramer's rule.*

$$x_1 = \frac{det(A_1(b))}{det(A)} = \frac{2s-2}{5s-4} = \frac{2(s-1)}{5s-4}$$

$$x_2 = \frac{det(A_2(b))}{det(A)} = \frac{1}{5s-4}$$

So that the solution of the system is

$$x = \begin{pmatrix} \frac{2(s-1)}{5s-4} \\ \frac{1}{5s-4} \end{pmatrix}$$

Using Cramer rule to solve big system is hopeless and inefficient the best way being the row reduction.

3.3.3 A formula for A^{-1}

Cramer's rule leads easily to a general formula for the inverse of an $n \times n$ matrix A. The jth column of A^{-1} is a vector x that satisfies $Ax = e_j$ where e_j is the jth column of the identity matrix, and the ith entry of x is the (i, j)-entry of A^{-1}. By Cramer's rule,

$$\{(i, j) - entry\ of\ A^{-1}\} = x_i = \frac{det(A_i(e_j))}{det(A)}$$

Recall that $A_{j,i}$ denotes the submatrix of A formed by deleting row j and column i. A cofactor expansion down column i of $A_i(e_j)$ shows that

$$det(A_i(e_i)) = (-1)^{i+j} det(A_{ji}) = C_{ji}$$

where C_{ji} is a cofactor of A. The (i, j)-entry of A^{-1} is the cofactor C_{ji} divided by $det(A)$. (Note that the subscripts on C_{ji} are the reverse of (i, j).] Thus

$$A^{-1} = \frac{1}{det(A)} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ \cdots & \cdots & & \cdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix}$$

The matrix of cofactors on the right side is called the **adjugate** (or **classical adjoint**) of A, denoted by $adj(A)$.

Theorem 3.3.4 (An inverse formula)**.** *Let A be an invertible $n \times n$ matrix. Then*

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

Example 3.3.5. *Find the inverse of the matrix*

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{pmatrix}$$

Solution:
The nine cofactors are

$$C_{11} = +\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} = -2,\ C_{12} = -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} = 3,\ C_{13} = -\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} = 5$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} = 14,\ C_{22} = -\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -7,\ C_{23} = -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} = -7$$

$$C_{31} = -\begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 4,\ C_{32} = -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = 1,\ C_{33} = -\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} = -3$$

The adjugate matrix is the transpose of the matrix of cofactors. (For instance, C_{12} goes in the position $(2, 1)$ position.) Thus

$$adj(A) = \begin{pmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{pmatrix}$$

We could compute $det(A)$ directly, but the following computation provides a check on the calculations above and produces $det(A)$:

$$(adj(A)) \cdot A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{pmatrix} \begin{pmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{pmatrix} = \begin{pmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{pmatrix} = 14I$$

Since $(adj(A))A = 14I$, then we know from the inverse formula theorem that $det(A) = 14$ and

$$A^{-1} = \frac{1}{14} \begin{pmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{pmatrix} = \begin{pmatrix} -1/7 & 1 & 2/7 \\ 3/14 & -1/2 & 1/14 \\ 5/14 & -1/2 & -3/14 \end{pmatrix}$$

3.3.4 Determinants as area or volume

We verify the geometric interpretation of determinant described.

Theorem 3.3.6. *If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|det(A)$.*

Proof. The theorem is obviously true for any 2×2 diagonal matrix:

$$|det(\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix})| = |ad| = \{area\ of\ rectangle\}$$

It will suffice to show that any 2×2 matrix $A = [a_1, a_2]$ such that a_1 and a_2 are not scalar multiple of the other, can be transformed into a diagonal matrix in a way that changes neither the area of the associated parallelogram nor $|det(A)|$. We know that the absolute value of the determinant is unchanged when two columns are interchanged or a multiple of one column is added to another. And it is easy to see that such operations suffice to transform A into a diagonal matrix. Column interchanges do not change the parallelogram at all. So it suffices to prove the following simple geometric observation that applies to vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$:
Let a_1 and a_2 be nonzero vectors. Then for any scalar c, the area of the parallelogram determined by a_1 and a_2 equals the area of the parallelogram determined by a_1 and $a_2 + ca_1$.
To prove this statement, we may assume that a_2 is not a multiple of a_1, for otherwise the two parallelograms would be degenerate and have zero area. If L is the line through 0 and a_1, then $a_2 + ca_1$ have the same perpendicular distance to L. Hence the two parallelograms have the same area, since they share the base from 0 to a_1. This complete the proof for $\mathbb{R}^2$
The proof for $\mathbb{R}^3$ is similar and left to you. □

Example 3.3.7. *Calculate the area of the parallelogram determined by the points $(-2,-2)$, $(0,3)$, $(4,-1)$ and $(6,4)$.*
Solution: *First translate the parallelogram to one having the origin as a vertex. For example, subtract the vertex $(-2,-2)$ from each of the four vertices. The new parallelogram has the same area, and its vertices are $(0,0)$, $(2,5)$, $(6,1)$ and $(8,9)$. This parallelogram is determined by the columns of*

$$A = \begin{pmatrix} 2 & 6 \\ 5 & 1 \end{pmatrix}$$

Since $|det(A)| = |-28|$, the area of the parallelogram is 28.

3.3.5 Linear transformations

Determinant can be used to describe an important geometric property of linear transformations in the plane and in $\mathbb{R}^3$. If T is a linear transformation and S is a set in the domain of T, let $T(S)$ denote the set of images of points in S. We are interested in how the area (or volume) of $T(S)$ compares with the area (or volume) of the original set S. For convenience, when S is a region bounded by a parallelogram, we also refer to S as a parallelogram.

Theorem 3.3.8. *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then*

$$\{area\ of\ T(S)\} = |det(A)| \cdot \{area\ of\ S\}$$

If T is determined by 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then

$$\{volume\ of\ T(S)\} = |det(A)| \cdot \{volume\ of\ S\}$$

Proof. Consider the 2×2 case, with $A = [a_1, a_2]$. A parallelogram at the origin in $\mathbb{R}^2$ determined by vectors b_1 and b_2 has the form

$$S = \{s_1 b_1 + s_2 b_2 : 0 \leqslant s_1 \leqslant 1, 0 \leqslant s_2 \leqslant 1\}$$

The image of S under T consists of points of the form:

$$T(s_1 b_1 + s_2 b_2) = s_1 T(b_1) + s_2 T(b_2) = s_1 A b_1 + s_2 A b_2$$

where $0 \leqslant s_1 \leqslant 1,\ 0 \leqslant s_2 \leqslant 1$. It follows that $T(S)$ is the parallelogram determined by the columns of the matrix $[Ab_1, Ab_2]$. This matrix can be written as AB, where $B = [b_1, b_2]$. By the previous theorem and the product theorem for determinants,

$$\{area\ of\ T(S)\} = |det(AB)| = |det(A)||det(B)| = |det(A)| \cdot \{area\ of\ S\}$$

An arbitrary parallelogram has the form $p + S$, where p is a vector and S is a parallelogram at the origin, as above. It is easy to see that T transform $p + S$ into $T(p) + T(S)$. Since transformation does not affect the area of a set,

$$\{area\ of\ T(p+S)\} = \{area\ of\ T(p)+T(S)\} = \{area\ of\ T(S)\} = |det(A)|\{area\ of\ S\} = |det(A)|\{area\ of\ p+S\}$$

This shows the theorem holds for a parallelogram in $\mathbb{R}^2$. The proof for the 3×3 case is analogous. □

Example 3.3.9. *Let a and b be positive numbers. Find the area of the region E bounded by the ellipse whose equation is*

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

Solution: *We claim that E is the image of the unit disk D under the linear transformation T determined by the matrix $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, because if $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $x = Au$, then*

$$u_1 = \frac{x_1}{a} \text{ and } u_2 = \frac{x_2}{b}$$

It follows that u is the unit disk, with $u_1^2 + u_2^2 \leqslant 1$, if and only if x is in E, with $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \leqslant 1$. Then,

$$\{area\ of\ ellipse\} = \{area\ of\ T(D)\} = |det(A)|\{area\ of\ D\} = ab\pi(1)^2 = \pi ab$$

Chapter 4

Vector spaces

4.1 Vector spaces and subspace

The algebraic properties of $\mathbb{R}^n$ are shared by many other system in mathematics.

Definition 4.1.1. *A* **vector space** *is nonempty set V of objects, called vectors, on which are defined two operations, called addition and multiplication by scalars (real numbers), subject to the ten axioms (or rules) listed below. The axioms must hold for all vectors u, v and w in V and for all scalar c and d.*

1. *The sum of u and v, denoted by $u+v$, is in V.*
2. *$u+v=v+u$ (commutativity of addition)*
3. *$(u+v)+w=u+(v+w)$ (associativity of addition)*
4. *There is a zero vector 0 in V such that $u+0=u$. (zero element)*
5. *For each u in$V$, there is a vector $-u$ in V such that $u+(-u)=0$.*
6. *The scalar multiple of u by c, denoted by cu, is in V.*
7. *$c(u+v)=cu+cv$ (distributivity)*
8. *$(c+d)u=cu+du$. (distributivity)*
9. *$c(du)=(cd)u$.*
10. *$1u=u$. (unit element)*

Using only these axioms, one can prove, for each u in V and scalar c,

1. the the zero vector is unique
2. the inverse $-u$ called the **negative** of an element is unique for each u in V.
3. $0u=0$
4. $c0=0$
5. $-u=(-1)u$

Example 4.1.2. *The space $\mathbb{R}^n$, where $n \geqslant 1$, are the first example of vector spaces.*

Example 4.1.3. *Let V be the set of all arrows (directed line segments) in three-dimensional space, with two arrows regarded as equal if they have the same length and point the same direction. Define addition by the parallelogram rule and for each v in V, define cv to be the arrow whose length is $|c|$ times the length of v, pointing in the same direction as v if $c \geqslant 0$ and otherwise pointing in the opposite direction. Show that V is a vector space. This space is a common model in physical problems for various forces.*

Example 4.1.4. *Let $\mathbb{S}$ be the space of all doubly infinite sequences of numbers (usually written in a row rather than a column):*

$$\{y_k\} = (\cdots, y_{-2}, y_{-1}, y_0, y_1, y_2, \cdots)$$

If $\{z_k\}$ is another element of $\mathbb{S}$ then the sum $\{y_k\} + \{z_k\}$ is the sequence $\{x_k + y_k\}$ formed by adding corresponding terms of $\{y_k\}$ and $\{z_k\}$. The scalar multiple $c\{y_k\}$ is the sequence $\{cy_k\}$. The vector space space axioms are verify in the same way as for $\mathbb{R}^n$. Elements of $\mathbb{S}$ arise in engineering, for example, whenever a signal is measured (or sampled) at discrete times. A signal might be electrical, mechanical, optical, and so on. For convenience, we will call $\mathbb{S}$ the space of (discrete-time) **signal**.

Example 4.1.5. *For $n \geqslant 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n consist of all polynomials of the form*

$$p(t) = a_0 + a_1 t + \cdots + a_n t^n$$

where the coefficients $a_0, \cdots, a_n$ and the variable t are real numbers. The degree of p is higher power of t whose coefficient is non zero. If $p(t) = a_0 \neq 0$, the degree of p is zero. If all the coefficients are zero, p is called the zero polynomial. The zero polynomial is included in $\mathbb{P}_n$ even though its degree, for technical reasons, is not defined.
If $p(t) = a_0 + a_1 t + \cdots + a_n t^n$ and $q(t) = b_0 + b_1 t + \cdots + b_n t^n$ then the sum $p + q$ is defined by

$$(p+q)(t) = p(t) + q(t) = p(t) = (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n$$

The scalar multiple cp is the polynomial defined by

$$(cp)(t) = cp(t) = ca_0 + ca_1 t + \cdots + ca_n t^n$$

$\mathbb{P}_n$ is a vector space. (Not hard to prove)

Example 4.1.6. *Let V be the set of all real valued functions defined on a set $\mathbb{D}$. (Typically, $\mathbb{D}$ is the set of real numbers or some interval on the real line.) Functions are added in the usual way: $f + g$ is the function is the function whose value at t in the domain $\mathbb{D}$ is $f(t) + g(t)$. Likewise, for a scalar c and an f in V, the scalar multiple cf is the function whose value at t is $cf(t)$. For instance, if $\mathbb{D} = \mathbb{R}$, $f(t) = 1 + sin(2t)$ and $g(t) = 2 + 0.5t$, then*

$$(f+g)(t) = 3 + sin(2t) + 0.5t \text{ and } (2g)(t) = 4 + t$$

Two function in V are equal if and only if their values are equal for every t in $\mathbb{D}$. Hence the zero vector in V is the function that is identically zero, $f(t) = 0$ for all t, and the negative of f is $(-1)f$. V is a vector space not hard to prove.

The question is now when a subset of a vector space is also a vector space. This leads to the definition of subspace.

Definition 4.1.7. *A* **subspace** *of a vector space H is a subset V that has three properties*

1. *The zero vector of V is in H.*
2. *H is closed under the vector addition. That is, for each u and v in H, the sum $u + v$ is in H.*
3. *H is closed under multiplication by scalars. That is, for each u in H and each scalar c, the vector cu is in H.*

The other properties of a vector space are automatically satisfied since V is a subspace of a vector space H.

Example 4.1.8. *The set consisting of only the zero vector in a vector space V is a subspace of V, called the* **zero subspace** *and written as $\{0\}$.*

Example 4.1.9. *Let $\mathbb{P}$ be the set of all polynomials with real coefficients, with operations in $\mathbb{P}$ defined as for functions. Then $\mathbb{P}$ is a subspace of the space of all real-valued function defined as for functions defined on $\mathbb{R}$. Also, for each $n \geqslant 0$, $\mathbb{P}_n$ is a subset of $\mathbb{P}$ that contains the zero polynomial, the sum of the two polynomials in $\mathbb{P}^n$ is also in $\mathbb{P}_n$, and a scalar multiple of a polynomial in $\mathbb{P}_n$ is also in $\mathbb{P}_n$. So, $\mathbb{P}_n$ is a subspace of $\mathbb{P}$*

Example 4.1.10. *The vector space $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. (The vector in $\mathbb{R}^3$ all have three entries, whereas the vectors in $\mathbb{R}^2$ have only two.) The set*

$$H = \{\begin{pmatrix} 0 \\ s \\ t \end{pmatrix} : s \text{ and } t \text{ are real}\}$$

is a subset of $\mathbb{R}^3$ that "looks" and "acts" like $\mathbb{R}^2$, although it is logically distinct from $\mathbb{R}^2$. Show that H is a subspace of $\mathbb{R}^3$.
Solution: *The zero vector is in H and H is closed under vector addition and scalar multiplication because these operations on vectors in$H$ always produce vectors whose first entries are zero (and so belong to H). Thus H is a subspace of $\mathbb{R}^3$.*

Example 4.1.11. *Given v_1 and v_2 in a vector space V, let $H = Span\{v_1, v_2\}$. Show that H is a subspace of V.*
Solution: *Let $x, y \in V$ so that there is x_1, x_2, y_1, y_2 scalars such that $x = x_1v_1 + x_2v_2$ and $y = y_1v_1 + y_2v_2$ and c a scalar.*

1. $0 = 0v_1 + 0v_2 \in H$.
2. $x + y = (x_1 + y_1)v_1 + (x_2 + y_2)v_2 \in V$
3. $cx = (cx_1)v_1 + (cx_2)v_2 \in V$

It is not hard to generalize this argument and obtain the proof of the following theorem:

Theorem 4.1.12. *If $v_1, \cdots, v_p$ are in a vector space V, then $Span\{v_1, \cdots, v_p\}$ is a subspace of V.*

We call $Span\{v_1, \cdots, v_p\}$ **the subspace spanned** (or **generated**) by $\{v_1, \cdots, v_p\}$. Given any subspace of H of V, a **spanning** (or **generating**) **set** for H is a set $\{v_1, \cdots, v_p\}$ in H such that $H = Span\{v_1, \cdots, v_p\}$.

Example 4.1.13. *Let H be a set of all vectors of the form $(x+2y, x, y)$ where x, y are arbitrary scalars. That is,*

$$H = \{(x+2y, x, y), x, y \in \mathbb{R}\}$$

Show that H is a subspace of $\mathbb{R}^3$.
Solution: *Note that, a general vector is of the form:*

$$\begin{pmatrix} x+2y \\ x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

with x, y scalars. So that

$$H = Span\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}\}$$

and H is a subspace of $\mathbb{R}^3$.

Example 4.1.14. *Is $H = \{(x+2y, 1+x, y), x, y \in \mathbb{R}\}$ a vector space?*
Solution: *If it is a vector space if $(0,0,0) \in H$. That is, there is x, y scalar such that*

$$(x+2y, 1+x, y) = (0,0,0)$$

Then $y = 0$, $x = 0$ but we get $1 = 0$ impossible. So $(0,0,0)$ is not in H and H is not a subspace of $\mathbb{R}^3$.

4.2 Null spaces, columns spaces, and linear transformation

Definition 4.2.1. *The* **null space** *of an $m \times n$ matrix A, written as $Nul(A)$, is the set of all solutions of the homogeneous equation $Ax = 0$. In set notation,*

$$Nul(A) = \{x : x \text{ is in } \mathbb{R}^n \text{ and } Ax = 0\}$$

You can also see the set of all x in $\mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^n$ via the linear transformation $x \mapsto Ax$. Note that finding the null space is equivalent to find the solution set of $Ax = 0$.

Theorem 4.2.2. *The null space of $m \times n$ matrix is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $Ax = 0$ of m homogeneous linear equation in n unknown is a subspace of $\mathbb{R}^n$.*

Proof. Certainly $Nul(A)$ is a subset of $\mathbb{R}^n$ because A has n columns. We must show that $Nul(A)$ satisfies the three properties of a subspace. Of course,0 is in $Nul(A)$. Next, let u and v represent any two vector $Nul(A)$. Then

$$Au = 0 \text{ and } Av = 0$$

To show that $u+v$ is in $Nul(A)$, we must show that $A(u+v) = 0$. Using the properties of the matrix multiplication, compute

$$A(u+v) = Au + Av = 0 + 0 = 0$$

Finally, if c is any scalar, then

$$A(cu) = c(Au) = 0$$

which shows that cu is in $Nul(A)$. Thus $Nul(A)$ is a subspace of $\mathbb{R}^n$. □

It is important that the equation is homogeneous, otherwise the solution set would not be a subspace since 0 would not be on it since 0 is only solution for homogeneous systems.
There is no obvious relation between the vector in $Nul(A)$ and the entries of A. We say that $Nul(A)$ is defined implicitly, because it is defined by a condition that must be checked. No explicit list or description of the elements in $Nul(A)$ is given. However, solving the equation $Ax = 0$ amounts to producing an explicit description of $Nul(A)$.

Example 4.2.3. *Find a spanning set for the null space of the matrix*

$$A = \begin{pmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Solution: *The system $Ax = 0$ can be rewritten as*

$$\begin{cases} x_1 - 2x_2 - x_4 + 3x_5 = 0 \\ x_3 + 2x_4 - 2x_5 = 0 \end{cases}$$

We express then the basic variable in terms of the free variable x_2, x_4, x_5, then the general form of a solution is:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_2 + 2x_5 \\ x_4 \\ x_5 \end{pmatrix} = x_2 \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix} = x_2 u + x_3 v + x_4 w$$

with $u = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $v = \begin{pmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ *and* $w = \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$. *So that every linear combination of*

u, v and w is an element of $Nul(A)$, and vice versa. Thus, $\{u, v, w\}$ is a spanning set for $Nul(A)$.

The two following points apply to all problem of this type where $Nul(A)$ contains nonzero vectors:

1. The spanning set obtained in the previous example is automatically linearly independent. Because one can prove that $x_2u + x_4v + x_5w$ can be 0 only if the weights x_2, x_4 and x_5 are all zero.
2. When $Nul(A)$ contains nonzero vectors, the number of vectors in the spanning set for $Nul(A)$ equals the number of free variables in the equation $Ax = 0$.

Definition 4.2.4. *The* **column space** *of an $m \times n$ matrix A, written as $Col(A)$, is the set of all linear combinations of the columns of A. If $A = [a_1, \cdots, a_n]$, then*

$$Col(A) = Span\{a_1, \cdots, a_n\}$$

Theorem 4.2.5. *The column space of $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.*

Note that a typical vector in $Col(A)$ can be written as Ax for some x because the notation Ax stands for a linear combination of the columns of A. That is

$$Col(A) = \{b : b = Ax \ forsome \ x \ in \ \mathbb{R}^n\}$$

The notation Ax for vectors in $Col(A)$ also shows that $Col(A)$ is the range of the linear transformation $x \mapsto Ax$.

Fact 4.2.6. *The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $Ax = b$ has a solution for each $b \in \mathbb{R}^m$.*

Definition 4.2.7. *A* **linear transformation** *T from a vector space V into a vector space W is a rule that assigns to each vector x in V a unique vector $T(x)$ in W, such that*

1. *$T(u + v) = T(u) + T(v)$, for all u, v in V, and*
2. *$T(cu) = cT(u)$, for all u in V and all scalars c.*

The **kernel** *(or* **null space***) of such a T is the set of all u in V such that $T(u) = 0$. (The zero vector in W.) The* **range** *of T is the set of all vectors in W of the form $T(x)$ for some x in V. If T happens to arise as a matrix transformation-say, $T(x) = Ax$ for some matrix A, then the kernel and the range of T are just the null space and the column space of A, as defined earlier.*

One can prove that $ker(T)$ the kernel of T and $Range(T)$ the range of T are subspaces (Exercise).

Example 4.2.8. *Let V be the vector space of all real-valued functions f defined on an interval $[a, b]$ with the property that they are differentiable and their derivatices are continuous functions on $[a, b]$. Let W be the vector space $C[a, b]$ of all continuous* *functions on $[a,b]$, and let $D : V \to W$ be the transformations that changes f in V into its derivative f'. In calculus, two simple differentiation rules are*

$$D(f+g) = D(f) + D(g) \text{ and } D(cf) = cD(f)$$

That is, D is a linear transformation. It can be shown that the kernel of D is the set of the constant functions on $[a,b]$ and the range of D is the set W of all continuous function $[a,b]$.

Example 4.2.9. *The differential equation*

$$y'' + w^2 y = 0$$

where w is a constant, is used to describe a variety of physical systems, such as the vibration of a weighted spring, the movement of a pendulum, and the voltage in a inductance-capacitance electrical circuit. The set of solutions of this equation is precisely the kernel of the linear transformation that maps a function $y = f(t)$ into the function $f''(t) + w^2 f(t)$. Finding an explicit description of this vector space is a problem in differential equations. (Exercise)

4.3 Linearly independent sets; bases

Definition 4.3.1. *An indexed set of vectors $\{v_1, \cdots, v_p\}$ in V is said to be* **linearly independent** *if the vector equation*

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0$$

has only the trivial solution, $c_1 = 0, \cdots, c_p = 0$.
The set $\{v_1, \cdots, v_p\}$ is said to be linearly dependent if the vector equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0$$

has a non trivial solution, that is if there are some weights, $c_1, \cdots, c_p$ not all zero, such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0$$

holds. In such a case this equation is called a **linear dependence relation** *among $v_1, \cdots, v_p$.*

The properties that we have seen about linear independence about vectors in $\mathbb{R}^n$ are still true. For instance,

1. a set containing a single vector v is linearly independent if and only if $v \neq 0$.
2. Two vectors are linearly dependent if and only if one of the vectors is a multiple of the other.
3. A set containing the zero vector is linearly dependent.
4. An indexed set $\{v_1, \cdots, v_p\}$ of two or more vectors, with $v_1 \neq 0$ is linearly dependent if and only if some v_j (with $j > 1$) is a linear combination of the preceding vectors $v_1, \cdots, v_{j-1}$.

The main difference between linear dependence in $\mathbb{R}^n$ and in a general vector space is that when the vectors are not n-tuples, the vector equation usually cannot be written as a system of n linear equations. That is, the vectors cannot be made into the columns of a matrix A in order to study $Ax = 0$. We must rely instead on the definition of linear dependence.

Example 4.3.2. *Let $p_1(t) = 1 + t$, $p_2(t) = t^2$ and $p_3(t) = 3 + 3t + 4t^2$. Then $\{p_1, p_2, p_3\}$ are linearly dependent in $\mathbb{P}$ since $p_3 = 3p_1 + 4p_2$.*

Example 4.3.3. *The set $\{sin(t), cos(t)\}$ is linearly independent in $C[0,1]$, the space of all continuous function on $0 \leqslant t \leqslant 1$, because $sin(t)$ and $cos(t)$ are not multiple of one another as vectors in $C[0,1]$. That is, there is no scalar c such that $cos(t) = c \cdot sin(t)$ for all $t \in [0,1]$. (Look at the graphs of $sin(t)$ and $cos(t)$.) However, $\{sin(t)cos(t), sin(2t)\}$ is a linearly dependent because of the identity $sin(2t) = 2sin(t)cos(t)$, for all t.*

Definition 4.3.4. *Let H be a subspace of a vector space V. An indexed set of vectors $\mathcal{B} = \{b_1, \cdots, b_p\}$ in V is a* **basis** *for H if*

1. *$\mathcal{B}$ is a linear independent set, and*
2. *the subspace spanned by $\mathcal{B}$ coincides with H; that is*

$$H = Span\{b_1, \cdot, b_p\}$$

The definition of a basis applies to the case when $H = V$, because any vector space is a subspace of itself. Thus a basis of V is a linearly independent set that spans V. Observe that when $H \neq V$, the second condition includes the requirement that each of the vectors $b_1, \cdots, b_p$ must belong to H, because $Span\{b_1, \cdots, b_p\}$ contains $b_1, \cdots, b_p$.

Example 4.3.5. *Let A be an invertible $n \times n$ matrix, say $A = [a_1, \cdots, a_n]$. Then the columns of A form a basis for $\mathbb{R}^n$ because they are linearly independent and they span $\mathbb{R}^n$, by the Invertible Matrix Theorem.*

Example 4.3.6. *Let $e_1, \cdots, e_n$ be the columns of the identity matrix I_n. That is,*

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}, \cdots, e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 1 \end{pmatrix}$$

The set $\{e_1, \cdots, e_n\}$ is called the **standard basis** *for $\mathbb{R}^n$.*

Example 4.3.7. *Let $v_1 = \begin{pmatrix} 3 \\ 0 \\ -6 \end{pmatrix}, v_2 = \begin{pmatrix} -4 \\ 1 \\ 7 \end{pmatrix}$, and $v_3 = \begin{pmatrix} -2 \\ 1 \\ 5 \end{pmatrix}$. Determine if $\{v_1, v_2, v_3\}$ is a basis for $\mathbb{R}^3$.*

Solution: *Since there are exactly three vectors here in $\mathbb{R}^3$, we can use any of several* *methods to determine if the matrix $A = [v_1, v_2, v_3]$ is invertible. For instance, two row replacements reveal that A has three pivot positions. Thus A is invertible. As proven before, we know then that the column of A form a basis for $\mathbb{R}^3$.*

Example 4.3.8. *Let $S = \{1, t, t^2, \cdots, t^n\}$. Verify that S is a basis for $\mathbb{P}_n$. This basis is called the* **standard basis** *for $\mathbb{P}_n$.*
Solution: *Certainly S span $\mathbb{P}_n$, by definition of $\mathbb{P}_n$. To show that S is linearly independent suppose that $c_0, \cdots, c_n$ satisfy*

$$c_0 \cdot 1 + c_1 t + c_2 t^2 + \cdots + c_n t^n = 0(t)$$

This equality means that the polynomial on the left has the same values as the zero polynomial on the right. A fundamental theorem in algebra says that the only polynomial in $\mathbb{P}_n$ with more than n zero is the zero polynomial. So the equation holds for all t only if $c_0 = \cdots = c_n = 0$. This proves that S is linearly independent and hence is a basis for $\mathbb{P}_n$.

Theorem 4.3.9 (The spanning set theorem)**.** *Let $S = \{v_1, \cdots, v_p\}$ be a set in V, and let $H = Span\{v_1, \cdots, v_p\}$.*

1. *If one of the vector in S- say, v_k- is a linear combination of the remaining vectors in S, then the set formed from S by removing v_k still spans H.*
2. *If $H \neq 0$, some subset of S is a basis for H.*

Proof. 1. By rearranging the list of vectors in S, if necessary, we may suppose that v_p is a linear combination of $v_1, \cdots, v_{p-1}$- say,

$$v_p = a_1 v_1 + \cdots + a_{p-1} v_{p-1}$$

Given any x in H, we may write

$$x = c_1 v_1 + \cdots + c_{p-1} v_{p-1} + c_p v_p$$

for suitable scalars $c_1, \cdots, c_p$. Substituting the expression for v_p, that is $v_p = a_1 v_1 + \cdots + a_{p-1} v_{p-1}$ one can see that x is a linear combination of $v_1, \cdots, v_{p-1}$. Thus, $\{v_1, \cdots, v_{p-1}\}$ spans H, because x was an arbitrary element of H.

2. If the original spanning set S is linearly independent then it is already a basis for H. Otherwise one of the vectors in S depends on the others and can be deleted, by the first part. So long as there are two or more vectors in the spanning set, we can repeat this process until the spanning set is linearly independent and hence is a basis for H. If the spanning set is eventually reduced to one vector, that vector will be nonzero (and hence linearly independent) because $H \neq 0$.

□

Recall that any linear dependence relationship among the columns of A can be expressed in the form $Ax = 0$, where x is a column of weights. (If the columns are not involved in a particular dependence relation then their weights are zero.) When A is row reduced to a matrix B, the columns of B are often totally different from the columns of A. However, the equation $Ax = 0$ and $Bx = 0$ have exactly the same solution set of solutions If $A = [a_1, \cdots, a_n]$ and $B = [b_1, \cdots, b_n]$, then the vector equations

$$x_1 a_1 + \cdots + x_n a_n = 0 \text{ and } x_1 b_1 + \cdots + x_n b_n = 0$$

also have the same solution set. That is, the columns of A have exactly the same linear dependence relationships as the columns of B.

Theorem 4.3.10. *The pivot columns of a matrix A form a basis for $Col(A)$.*

Proof. Let B be the reduced echelon form of A. The set of pivot columns of B is linearly independent, for no vector in the set is a linear combination of the vectors that precede it. Since A is row equivalent to B, the pivot columns of A are linearly independent as well, because any linear dependence relation among the columns of A corresponds to a linear dependence relation among the column of B. For this same reason, every non pivot columns of A is a linear combination of the pivot columns of A. Thus the non pivot columns of A may be discarded from the spanning set for $Col(A)$, by the spanning Set Theorem. This leaves the pivot columns of A as a basis for $Col(A)$. □

WARMING: The pivot columns of a matrix A are evident when A has been reduced only to echelon form. But, be careful to use the pivot columns of A itself for the basis of $Col(A)$. Row operation can change the column space of A.

When the spanning set theorem is used, the deletion of vectors from the spanning set must stop, when the set becomes linearly independent.If an additional vector is deleted, it will not be a linear combination of the remaining vectors, and hence the smaller set will no longer span V. Thus a basis is a spanning set that is as small as possible. A basis is also a linearly independent set that is as large as possible. If S is a basis for V, and if S is enlarged by one vector -say, w- from V, then the new set cannot be linearly independent, because S spans V, and w is therefore a linear combination of the elements in S.

Example 4.3.11. *The set* $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix} \right\}$ *is a linearly independent but does not span* $\mathbb{R}^3$.

The set $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \right\}$ *is a basis of* $\mathbb{R}^3$.

The set $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} \right\}$ *is a Span of* $\mathbb{R}^3$ *but is linearly dependent.*

4.4 Coordinate systems

Theorem 4.4.1 (The unique representation theorem). *Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis for a vector space V. Then for each x in V, there exists a unique set of scalars $c_1, \cdots, c_n$*

such that

$$x = c_1b_1 + \cdots + c_nb_n$$

Proof. Since $\mathcal{B}$ spans V, there exist scalars such that

$$x = c_1b_1 + \cdots + c_nb_n$$

holds. Suppose x also has the representation

$$x = d_1b_1 + \cdots + d_nb_n$$

for scalar $d_1, \cdots, d_n$. Then subtracting, we have

$$(c_1 - d_1)b_1 + \cdots + (c_n - d_n)b_n = x - x = 0$$

Since $\mathcal{B}$ is linearly independent, the weights must be all 0. That is $c_i = d_i$ for all $1 \leqslant i \leqslant n$. $\square$

Definition 4.4.2. *Suppose $\mathcal{B} = \{b_1, \cdots, b_n\}$ is a basis for V and x is in V. The* **coordinates of x relative to the basis $\mathcal{B}$** *(or the* **$\mathcal{B}$-coordinate of x***) are the weights $c_1, \cdots, c_n$ such that $x = c_1b_1 + \cdots + c_nb_n$.*

If $c_1, \cdots, c_n$ are the $\mathcal{B}$-coordinates of x, then the vector in $\mathbb{R}^n$

$$[x]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{pmatrix}$$

is the **coordinate vector of x (relative to $\mathcal{B}$)**, or the **$\mathcal{B}$-coordinate vector of x** mapping $x \mapsto [x]_{\mathcal{B}}$ is the **coordinate mapping (determined by $\mathcal{B}$)**.

Example 4.4.3. *Consider a basis $\mathcal{B} = \{b_1, b_2\}$ for $\mathbb{R}^2$, where $b_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $b_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Suppose an x in $\mathbb{R}^2$ has the coordinate vector $[x]_{\mathcal{B}} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$. Find x.*
Solution: *The $\mathcal{B}$-coordinates of x tell how to build x from the vectors in $\mathcal{B}$. That is,*

$$x = (-2)b_1 + 3b_2 = -2\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$$

Example 4.4.4. *The entries in the vector $x = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$ are the coordinates of x relative to the standard basis $\epsilon = \{e_1, e_2\}$, since*

$$\begin{pmatrix} 1 \\ 6 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 6\begin{pmatrix} 0 \\ 1 \end{pmatrix} = 1e_1 + 6e_2$$

If $\epsilon = \{e_1, e_2\}$, then $[x]_\epsilon = x$.

An analogous change of coordinates can be carried out in $\mathbb{R}^n$ for a basis $\mathcal{B} = \{b_1, \cdots, b_n\}$. Let

$$P_{\mathcal{B}} = [b_1, \cdots, b_n]$$

Then the vector equation

$$x = c_1 b_1 + c_2 b_2 + \cdots + c_n b_n$$

is equivalent to

$$x = P_{\mathcal{B}}[x]_{\mathcal{B}}$$

We call $P_{\mathcal{B}}$ the **change-of-coordinates matrix from $\mathcal{B}$** to the standard basis in $\mathbb{R}^n$. Left-multiplication by $P_{\mathcal{B}}$ transforms the coordinate vector $[x]_{\mathcal{B}}$ into x. The change of coordinates equations is important.
Since the columns of $P_{\mathcal{B}}$ form a basis for $\mathbb{R}^n$, $P_{\mathcal{B}}$ is invertible (by the Invertible Matrix Theorem), Left-multiplication by $P_{\mathcal{B}}^{-1}$ converts x into its $\mathcal{B}$-coordinate vector:

$$P_{\mathcal{B}}^{-1} x = [x]_{\mathcal{B}}$$

The correspondence $x \mapsto [x]_{\mathcal{B}}$, produced here by $P_{\mathcal{B}}^{-1}$ is the coordinate mapping mentioned earlier. Since $P_{\mathcal{B}}^{-1}$ is an invertible matrix, the coordinate mapping is a one-to-one linear transformation from $\mathbb{R}^n$ onto $\mathbb{R}^n$, by the Invertible Matrix Theorem. The property of the coordinate mapping is also true in a general vector space that has a basis,as we shall see.

Theorem 4.4.5. *Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis for a vector space V. Then the coordinate mapping $x \mapsto [x]_{\mathcal{B}}$ is a one-to-one linear transformation from V onto $\mathbb{R}^n$.*

Proof. Take two typical vectors in V, say,

$$u = c_1 b_1 + \cdots + c_n b_n$$

$$w = d_1 b_1 + \cdots + d_n b_n$$

Then, using vector operations,

$$u + v = (c_1 + d_1) b_1 + \cdots + (c_n + d_n) b_n$$

It follows that

$$[u+v]_{\mathcal{B}} = \begin{pmatrix} c_1 + d_1 \\ . \\ . \\ . \\ c_n + d_n \end{pmatrix} = \begin{pmatrix} c_1 \\ . \\ . \\ . \\ c_n \end{pmatrix} + \begin{pmatrix} d_1 \\ . \\ . \\ . \\ d_n \end{pmatrix} = [u]_{\mathcal{B}} + [v]_{\mathcal{B}}$$

So the coordinate mapping preserves addition. If r is any scalar, then

$$ru = r(c_1 b_1 + \cdots + c_n b_n) = (rc_1) b_1 + \cdots + (rc_n) b_n$$

So,

$$[ru]_{\mathcal{B}} = \begin{pmatrix} rc_1 \\ . \\ . \\ . \\ rc_n \end{pmatrix} = r \begin{pmatrix} c_1 \\ . \\ . \\ . \\ c_n \end{pmatrix} = r[u]_{\mathcal{B}}$$

Thus the coordinate mapping also preserves scalar multiplication and hence is a linear transformation. The proof of one-to-one and onto are left as exercises. □

The linearity of the coordinate mapping extends to linear combinations,. If $u_1, \cdots, u_p$ are in V and if $c_1, \cdots, c_p$ are scalars, then

$$[c_1u_1 + \cdots + c_pu_p]_{\mathcal{B}} = c_1[u_1]_{\mathcal{B}} + \cdots + c_p[u_p]_{\mathcal{B}}$$

In words, this means that the $\mathcal{B}$-coordinate vector of a linear combination of $u_1, \cdots, u_p$ is the same linear combination of their coordinate vectors.
The coordinate mapping is an important example of an isomorphism from V to $\mathbb{R}^n$. In general, a one-to-one linear transformation from a vector space V onto a vector space W is called an **isomorphism** from V onto W (iso from Greek for "the same" and morph from the Greek for "form" or "structure"). The notation and terminology for V and W may differ, but two spaces are indistinguishable as vector spaces. Every vector space calculation in V is accurately reproduced in W, and vice versa. In particular, any real vector space with a basis of n vectors is indistinguishable from $\mathbb{R}^n$. Linear independence, basis, span are conserved by isomorphisms.

Example 4.4.6. *Let $\mathcal{B}$ be the standard basis of the space $\mathbb{P}_3$ of polynomials; that is, let $\mathcal{B} = \{1, t, t^2, t^3\}$. A typical element p of $\mathbb{P}^3$ has the form*

$$p(t) = a_0 + a_1t + a_2t^2 + a_3t^3$$

Since p is already displayed as a linear combination of the standard basis vectors, we conclude that

$$[p]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}$$

Thus the coordinate mapping $p \mapsto [p]_{\mathcal{B}}$ is an isomorphism from $\mathbb{P}^3$ onto $\mathbb{R}^4$. All vector space operations in $\mathbb{P}^3$, correspond to operation in $\mathbb{R}^4$.

Example 4.4.7. *Use coordinate vectors to verify that the polynomials $1+2t^2$, $4+t+5t^2$ and $3+2t$ are linearly dependent in $\mathbb{P}^2$.*
Solution: *The coordinate mapping produces the coordinate vectors $(1,0,2)$, $(4,1,5)$ and $(3,2,0)$, respectively. Writing these vectors as the columns of a matrix A, we can determine their independence by row reducing the augmented matrix for $Ax = 0$*

$$\begin{pmatrix} 1 & 4 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 2 & 5 & 0 & 0 \end{pmatrix} \sim \textit{Row reduce} \sim \begin{pmatrix} 1 & 4 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The columns of A are linearly dependent, so the corresponding polynomials are linearly dependent. In fat, it is easy to check that columns 3 of A is 2 times column 2 minus 5 times columns 1. The corresponding relation for polynomials is

$$3 + 2t = 2(4 + t + 5t^2) - 5(1 + 2t^2)$$

Example 4.4.8. *Let* $v_1 = \begin{pmatrix} 3 \\ 6 \\ 2 \end{pmatrix}$, $v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ *and* $x = \begin{pmatrix} 3 \\ 12 \\ 7 \end{pmatrix}$.
and $\mathcal{B} = \{v_1, v_2\}$ is a basis for $H = Span\{v_1, v_2\}$. Determine if x is in H, and if it is, find the coordinate vector of x relative to $\mathcal{B}$.
Solution: *If x is in H, then the following vector equation in c_1, c_2 is consistent:*

$$c_1 v_1 + c_2 v_2 = x$$

The scalar c_1 and c_2, if they exist are the $\mathcal{B}$-coordinates of x, Using the row operations on the augmented matrix corresponding to the system:

$$\begin{pmatrix} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{pmatrix} \sim \textit{ Row reduce } \sim \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus $c_1 = 2$, $c_2 = 3$ *and* $[x]_{\mathcal{B}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

4.5 The dimension of vector space

The first theorem generalize a well-Known result about the vector space $\mathbb{R}^n$.

Theorem 4.5.1. *If a vector space V has a basis $\mathcal{B} = \{b_1, \cdots, b_n\}$, then any set in V containing more than n vectors must be linearly dependent.*

Proof. Let $\{u_1, \cdots, u_p\}$ be a set in V with more than n vectors. The coordinated vectors $[u_1]_{\mathcal{B}}, \cdots, [u_p]_{\mathcal{B}}$ form a linearly dependent set in $\mathbb{R}^n$, because there are more vectors (p) than entries (n) in each vector. So there exist scalars $c_1, \cdots, c_p$ not all zero, such that

$$c_1[u_1]_{\mathcal{B}} + \cdots + c_p[u_p]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

Since the coordinate mapping is a linear transformation

$$[c_1 u_1 + \cdots + c_p u_p]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

The zero vector on the right displays the n weights needed to build the vector $c_1u_1 + \cdots + c_pu_p$ from the basis in $\mathcal{B}$. That is, $c_1u_1 + \cdots + c_pu_p = 0 \cdot b_1 + \cdots + 0 \cdot b_n = 0$. Since the c_i are not all zero, $\{u_1, \cdots, u_p\}$ is linearly dependent. $\square$

As a consequence of the previous theorem, if a vector space V has a basis $\mathcal{B} = \{b_1, \cdots, b_n\}$, then each linearly independent set in V has no more than n vectors.

Theorem 4.5.2. *If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.*

Proof. Let $\mathcal{B}_1$ be a basis of n vectors and $\mathcal{B}_2$ be any other basis (of V). Since $\mathcal{B}_1$ is a basis and $\mathcal{B}_2$ is linearly independent, $\mathcal{B}_2$ has no more than n vectors, by the previous theorem. Also, since $\mathcal{B}_2$ is a basis and $\mathcal{B}_1$ is linearly independent, $\mathcal{B}_2$ has at least n vectors. So, $\mathcal{B}_2$ consists of exactly n vectors. $\square$

If a nonzero vector space V is spanned by a finite set S, then a subset of S is a basis for V, by the Spanning Set Theorem. The previous theorem ensures that the following definition makes sense.

Definition 4.5.3. *If V is spanned by a finite set, then V is said to be* **finite-dimensional**, *and the* **dimension** *of V, written as $dim(V)$, is the number of vectors in a basis for V. The dimension of the zero vector $\{0\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be* **infinite-dimensional**.

Example 4.5.4. *The standard basis for $\mathbb{R}^n$ contains n vectors, so $dim(\mathbb{R}^n) = n$. The standard polynomial basis $\{1, t, t^2\}$ shows that $dim(\mathbb{P}_2) = 3$. In general, $dim(\mathbb{P}_n) = n+1$. The space $\mathbb{P}$ of all polynomials is infinite-dimensional.*

Example 4.5.5. *Find the dimension of the subspace*

$$H = \{ \begin{pmatrix} s-2t \\ s+t \\ 3t \end{pmatrix} : s, t \text{ in } \mathbb{R}\}$$

Solution: *It is easy to see that H is the set of all linear combinations of the vectors;*

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, v_2 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$$

Clearly, $v_1 \neq 0$, v_2 is not a multiples and hence are linearly independent. Thus, $dim(H) = 2$.

Example 4.5.6. *The subspaces of $\mathbb{R}^3$ can be classified by dimension.*

— 0-*dimensional subspaces. Only the zero subspace.*
— 1-*dimensional subspaces. Any subspace spanned by a single nonzero vector. Such subspaces are lines through the origin.*
— 2-*dimensional subspaces. Any subspace spanned by two linearly independent vectors. Such subspaces are planes through the origin.*
— 3-*dimensional subspaces. Only $\mathbb{R}^3$ itself. Any three linearly independent vectors in $\mathbb{R}^3$ span all of $\mathbb{R}^3$, by the Invertible Matrix Theorem.*

4.6 Subspaces of a finite-dimensional space.

Theorem 4.6.1. *Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and*

$$dim(H) \leqslant dim(V)$$

Proof. If $H = \{0\}$, then certainly $dim(H) = 0 \leqslant dim(V)$. Otherwise, let $S = \{u_1, \cdots, u_n\}$ be any linearly independent set in H. If $S = Span(H)$, then S is a basis for H. Otherwise, there is some u_{k+1} in H that is not $Span(S)$. But then $\{u_1, \cdots, u_k, u_{k+1}\}$ will be linearly independent, because no vector in the set can be a linear combination of vectors that precede it.
So long as the new set does not span H, we can continue this process of expanding S to a larger linearly independent set in H. But the number of vectors in a linearly independent expansion of S can never exceed the dimension of V, by a proven theorem. So eventually the expansion of S will span H and hence will be a basis for H, and $dim(H) \leqslant dim(V)$. □

When we know the dimension of a vector space or subspace, one can apply the following theorem.

Theorem 4.6.2 (The basis Theorem)**.** *Let V be a p dimensional vector space $p \geqslant 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.*

4.7 The dimensional of $Nul(A)$ and $Col(A)$

Since the pivot columns of matrix A form a basis for $Col(A)$, we know the dimension of $Col(A)$ as soon as we know the pivot columns. The dimension of $Nul(A)$ usually takes more time than a basis for $Col(A)$.
Let A be an $m \times n$ matrix, and suppose the equation $Ax = 0$ has k free variables. We know that the standard method of finding a spanning set for $Nul(A)$, will produce exactly k linearly independent vectors-say, $u_1, \cdots, u_k$-one for each free variable. So $\{u_1, \cdots, u_k\}$ is a basis for $Nul(A)$, and the number of free variable, determines the size of the basis. Let us summarize these facts for future reference.

Fact 4.7.1. *The dimension of $Nul(A)$ is the number of free variables in the equation $Ax = 0$ and the dimension of $Col(A)$ is the number of basic variables in the equation $Ax = 0$ and the dimension of $Col(A)$ is the number of pivot columns in A.*

Example 4.7.2. *Find the dimension of the null space and the column space of*

$$A = \begin{pmatrix} 1 & -6 & 9 & 0 & -2 \\ 0 & 1 & 2 & -4 & 5 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Solution: *There are two free variable x_3 and x_5 corresponding to the non-pivot columns. Hence the dimension of $Nul(A)$ is 2 and there are 3 pivots columns so $dim(Col(A)) = 3$.*

4.8 Rank

4.8.1 The row space

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the vectors is called the **row space** of A and is denoted by $Row(A)$. Each row has n entries, so $Row(A)$ is a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write $Col(A^T)$ in place of $Row(A)$.

Theorem 4.8.1. *If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form of A, the nonzero rows of B form a basis for the row space of A as well as for that of B.*

Proof. If B is obtained from A by row operations, the rows of B are linear combinations of the rows of A. It follows that any linear combination of the rows of B is automatically a linear combination of the rows of A. Thus the row space of B is contained in the row space of A. Since row operations are reversible, the same argument shows that the row space of A is a subset of the row space of B. So the two row spaces are the same. If B is in echelon form, its nonzero rows are linearly independent because no nonzero row is a linear combination of the nonzero rows below it. Thus the nonzero rows of B form a basis of the (common) row space of B and A. □

Example 4.8.2. *Find bases for the row space, the column space and the null space of the matrix:*

$$A = \begin{pmatrix} 1 & -4 & 9 & -7 \\ -1 & 2 & -4 & 1 \\ 5 & -6 & 10 & 7 \end{pmatrix}$$

knowing that it is row equivalent to

$$B = \begin{pmatrix} 1 & 0 & -1 & 5 \\ 0 & -2 & 5 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Solution: *From the previous theorem, a basis for the row space is given by the non zero rows when reduce to an echelon form, so the basis is*

$$\{\begin{pmatrix} 1 & 0 & -1 & 5 \end{pmatrix}, \begin{pmatrix} 0 & -2 & 5 & -6 \end{pmatrix}\}$$

For the basis of $Col(A)$ we have proven that the pivot column of A form a basis of $Col(A)$. Through the echelon form we know that the pivot column are column 1 and 2. So a basis for $Col(A)$ is

$$\{\begin{pmatrix} 1 \\ -1 \\ 5 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \\ -6 \end{pmatrix}\}$$

Finally, in order to find the basis of $Nul(A)$ we remember that $Ax = 0$ if and only if $Bx = 0$, that is equivalent to

$$\begin{cases} x_1 - x_3 + 5x_4 & = & 0 \\ -2x_2 + 5x_3 - 6x_4 & = & 0 \\ 0 & = & 0 \end{cases}$$

So, $x_1 = x_3 - 5x_4$ and $x_2 = \frac{5x_3 - 6x_4}{2}$, with x_3, x_4 free variables. So the solution set is

$$Nul(A) = \{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_3 - 5x_4 \\ \frac{5x_3 - 6x_4}{2} \\ x_3 \\ x_4 \end{pmatrix} = x_3 \begin{pmatrix} 1 \\ 5/2 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -5 \\ -3 \\ 0 \\ 1 \end{pmatrix} : x_3, x_4 \in \mathbb{R} \}$$

So a basis for $Nul(A)$ is

$$\{ \begin{pmatrix} 1 \\ 5/2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -5 \\ -3 \\ 0 \\ 1 \end{pmatrix} \}$$

Observe that, unlike the basis for $Col(A)$ the bases for $Row(A)$ and $Nul(A)$ have no simple connection with the entries in A itself.

Remarque 4.8.3.

1. *Note that you could also row reduce A^T is order to find a basis of $Row(A)$ as you do for $Col(A)$.*
2. *Be careful: Even if the first two row of B are linearly independent we cannot conclude that the two first row of A are. This is not true in general.*

4.9 The rank theorem

Definition 4.9.1. *The* **rank** *of A is the dimension of the column space of A.*

Since $Row(A)$ is the same as $Col(A^T)$, the dimension of the row space of A is the rank of A^T. The dimension of the null space is sometimes called the **nullity** of A, though we will not use this term.

Theorem 4.9.2 (The rank theorem)**.** *The dimensions of the column space and the row space of an $m \times n$ matrix A are equal. This common dimension, the rank of A, also equals the number of pivot positions in A and satisfies the equation*

$$rank(A) + dim(Nul(A)) = n$$

Proof. $rank(A)$ is the number of pivot columns in A. Equivalently, $rank(A)$ is the number of pivot positions in an echelon form B of A. Furthermore, since B has a nonzero row for each pivot, and since these rows forma basis for the row space of A, the rank of A is also the dimension of the row space.

The dimension of $Nul(A)$ equals the number of free variables in the equation $Ax = 0$ equal to the number of the non pivot column. Obviously,

$$\{\text{number of pivot columns}\} + \{\text{number of nonpivot columns}\} = \{\text{number of columns}\}$$

This proves the theorem. $\square$

Example 4.9.3. *1. If A is a 5×3 matrix with a two-dimensional null space, what is the rank of A?*

2. Could a 4×8 matrix have a 2 dimensional null space?

Solution:

1. *Since A has 3 columns,*
$$(rank(A)) + 2 = 3$$
and hence $rank(A) = 1$.
2. *Since there is 8 columns, if the null space has dimension 2. Then*
$$(rank(A)) + 2 = 8$$
So, $rank(A) = 6$. But the columns of A are vectors in $\mathbb{R}^4$, and so the dimension of $Col(A)$ cannot exceed 4.

4.10 Applications to system of equations

The rank theorem is a powerful tool for processing information about system of linear equations. The next example simulates the way a real-life problem using linear equations might be started, without explicit mention of linear algebra terms such as matrix, subspace, and dimension.

Example 4.10.1. *A scientist has found two solutions to a homogeneous system of 40 equations in 42 variables. The two solutions are not multiples, and all other solutions can be constructed by adding together appropriate multiples of these two solutions. Can the scientist be certain that an associated nonhomogeneous system (with the same coefficients) has a solutions?*

Solution: *Yes. Let A be the 40×42 coefficient matrix of the system. The given information implies that the two solutions are linearly independent and span $Nul(A)$. So $dim(Nul(A)) = 2$. By the Rank Theorem, $dim(Col(A)) = 42 - 2 = 40$. Since $\mathbb{R}^{40}$ is the only subspace of $\mathbb{R}^{40}$ whose dimension is 40, $Col(A)$ must be all of $\mathbb{R}^{40}$. This means that every nonhomogeneous equation $Ax = b$ has a solution.*

4.11 Rank and the invertible matrix theorem

Theorem 4.11.1 (The invertible matrix theorem (continued)). *Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.*

1. *The columns of A form a basis of $\mathbb{R}^n$*

2. $Col(A) = \mathbb{R}^n$
3. $dim(Col(A)) = n$
4. $rank(A) = n$
5. $Nul(A) = \{0\}$
6. $dim(Nul(A)) = 0$.

We refrained from adding to the Invertible Matrix Theorem obvious statement about the row space of A, because the row space is the column space of A^T. Recall that A is invertible if and only if A^T is invertible. To do so would the length of the theorem and produce a list of over 30 statements.

4.12 Change of basis

When a basis $\mathcal{B}$ is chosen for an n-dimensional vector space V, the associated coordinate mapping onto $\mathbb{R}^n$ provides a coordinate system for V. Each x in V is identified uniquely by its $\mathcal{B}$-coordinate vector $[x]_{\mathcal{B}}$.
In some applications, a problem is described initially using a basis $\mathcal{B}$, but the problem's solution is aided by changing $\mathcal{B}$ to a new basis $\mathcal{C}$. Each vector is assigned a new $\mathcal{C}$ coordinate vector. In this section, we study how $[x]_{\mathcal{C}}$ and $[x]_{\mathcal{B}}$ are related for each x in V.

Theorem 4.12.1. *Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ and $\mathcal{C} = \{c_1, \cdots, c_n\}$ be bases of a vector space V. Then there is a unique $n \times n$ matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ such that*

$$[x]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[x]_{\mathcal{B}}$$

The column of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are the $\mathcal{C}$ coordinate vectors of the vectors in the basis $\mathcal{B}$. That is,

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[b_1]_{\mathcal{C}}, [b_2]_{\mathcal{C}}, \cdots, [b_n]_{\mathcal{C}}]$$

The matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is called the **change-of-coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$**. Multiplication by $P_{\mathcal{C}\leftarrow\mathcal{B}}$ converts $\mathcal{B}$-coordinates into $\mathcal{C}$-coordinates.
The columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are linearly independent because they are the coordinate vectors of the linearly independent set $\mathcal{B}$. Since $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is square, it must be invertible by the Invertible Matrix Theorem. Left-multiplying both sides of the equation

$$[x]_{\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[x]_{\mathcal{B}}$$

by $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}$ yields

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}[x]_{\mathcal{C}} = [x]_{\mathcal{B}}$$

Thus $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}$ is the matrix that converts $\mathcal{C}$-coordinates into $\mathcal{B}$-coordinates. That is,

$$(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$$

In $\mathbb{R}^n$, if $\mathcal{B} = \{b_1, \cdots, b_n\}$ and ϵ is the standard basis $\{e_1, \cdots, e_n\}$ in $\mathbb{R}^n$, then $[b_1]_\epsilon = b_1$, and likewise for the other vectors in $\mathcal{B}$. In this case, $P_{\epsilon\leftarrow\mathcal{B}}$ is the same as the change of coordinates matrix $P_{\mathcal{B}}$ introduced earlier namely

$$P_{\mathcal{B}} = [b_1, \cdots, b_n]$$

To change coordinate between two nonstandard bases in $\mathbb{R}^n$, the theorem shows that we need the coordinate vectors of the old basis relative to the new basis.

Example 4.12.2. *Let* $b_1 = \begin{pmatrix} -9 \\ 1 \end{pmatrix}$, $b_2 = \begin{pmatrix} -5 \\ -1 \end{pmatrix}$, $c_1 = \begin{pmatrix} 1 \\ -4 \end{pmatrix}$, *and* $c_2 = \begin{pmatrix} 3 \\ -5 \end{pmatrix}$, *and consider the bases for* $\mathbb{R}^2$ *given by* $\mathcal{B} = \{b_1, b_2\}$ *and* $\mathcal{C} = \{c_1, c_2\}$*. Find the change of coordinates matrix from* $\mathcal{B}$ *to* $\mathcal{C}$*.*

Solution: *The matrix* $P_{\mathcal{C}\leftarrow\mathcal{B}}$ *involve the* $\mathcal{C}$*-coordinate vectors of* b_1 *and* b_2*. Let* $[b_1]_{\mathcal{C}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ *and* $[b_2]_{\mathcal{C}} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$*. Then, by definition,*

$$[c_1, c_2]\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = b_1 \text{ and } [c_1, c_2]\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = b_2$$

To solve both systems simultaneously, augment the coefficient matrix with b_1 *and* b_2 *and row reduce:*

$$[c_1, c_2|b_1, b_2] = \left(\begin{array}{cc|cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{array}\right)$$

Thus

$$[b_1]_{\mathcal{C}} = \begin{pmatrix} 6 \\ -5 \end{pmatrix} \text{ and } [b_2]_{\mathcal{C}} = \begin{pmatrix} 4 \\ -3 \end{pmatrix}$$

The desired change of coordinates matrix is therefore

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[b_1]_{\mathcal{C}}, [b_2]_{\mathcal{C}}] = \begin{pmatrix} 6 & 4 \\ -5 & -3 \end{pmatrix}$$

Analogous procedure works for finding the change of coordinates matrix between any two bases in $\mathbb{R}^n$.

Another description of the change of coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ uses the change of coordinate matrices $P_{\mathcal{B}}$ and $P_{\mathcal{C}}$ that convert standard coordinate into $\mathcal{B}$-coordinates and $\mathcal{C}$ coordinates. Recall that for each $x \in \mathbb{R}^n$,

$$P_{\mathcal{B}}[x]_{\mathcal{B}} = x, \; P_{\mathcal{C}}[x]_{\mathcal{C}} = x, \text{ and } [x]_{\mathcal{C}} = (P_{\mathcal{C}})^{-1}x$$

Thus,

$$[x]_{\mathcal{C}} = (P_{\mathcal{C}})^{-1}x = (P_{\mathcal{C}})^{-1}P_{\mathcal{B}}[x]_{\mathcal{B}}$$

In $\mathbb{R}^n$, the change of coordinates matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ may be computed as $(P_{\mathcal{C}})^{-1}P_{\mathcal{B}}$. But it is faster to compute directly $P_{\mathcal{C}\leftarrow\mathcal{B}}$ instead as explained above.

4.13 Application to difference equations

The vector space $\mathbb{S}$ of discrete-time signals was introduce earlier. A **signal** in $\mathbb{S}$ is a function defined only on the integers and is visualized as a sequence of numbers say, $\{y_k\}$. Signals arise from electrical and control system engineering or also biology, economics, demography and many other areas, wherever a process is measured or sampled at

discrete time intervals. When a process begins at a specific time, it is sometimes convenient to write a signal as a sequence of the form $(y_0, y_1, y_2, \cdots)$. The terms y_k for $k < 0$ either are assumed to be zero or are simply omitted.
To simplify the notation we will consider a set of three signal (all these could be generalize for n signal). Let $\{u_k\}, \{v_k\}$ and $\{w_k\}$ be those signal. They are linearly independent precisely when the equation

$$c_1 u_k + c_2 v_k + c_3 w_k = 0, \textit{ for all } k \in \mathbb{Z}$$

implies that $c_1 = c_2 = c_3 = 0$.
If this is true this is obviously true for three consecutive value of k, say k, $k+1$ and $k+2$. Thus c_1, c_2, c_3 satisfy

$$\begin{pmatrix} u_k & v_k & w_k \\ u_{k+1} & v_{k+1} & w_{k+1} \\ u_{k+2} & v_{k+2} & w_{k+2} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \textit{ for all } k \in \mathbb{Z}$$

The coefficient matrix in this system is called the **Casorati matrix** of the signal, and the determinant of the matrix is called the **Casoratian** of $\{u_k\}, \{v_k\}$ and $\{w_k\}$. If the Casorati matrix is invertible for at least one value of k, then we will have $c_1 = c_2 = c_3 = 0$, which will prove that the three signal are linearly independent.

Example 4.13.1. *Verify that 1^k, $(-2)^k$ and 3^k are linearly independent signals.*
Solution: *The Casorati matrix is*

$$\begin{pmatrix} 1^k & (-2)^k & 3^k \\ 1^{k+1} & (-2)^{k+1} & 3^{k+1} \\ 1^{k+2} & (-2)^{k+2} & 3^{k+2} \end{pmatrix}$$

Row operations can show fairly easily that this matrix is always invertible. However it is faster to substitute a value for k, say $k = 0$ and row reduce the numerical matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \sim \textit{ Row reduce } \sim \begin{pmatrix} 1 & 1 & 1 \\ 0 & -3 & 2 \\ 0 & 0 & 10 \end{pmatrix}$$

The Casorati matrix is invertible for $k = 0$. So 1^k, $(-2)^k$ and 3^k are linearly independent.

If the Casorati matrix is not invertible, the associated signals being tested may or may not be linearly dependent. However, it can be shown that if signals are all solutions of the same homogeneous difference equation, then either the Casorati matrix is invertible for all k and the signals are linearly independent or else the Casorati matrix is not invertible for all k and the signal are linearly dependent.

Definition 4.13.2. *Given scalars $a_0, \cdots, a_n$, with a_0 and a_n nonzero, and given a signal $\{z_k\}$, the equation*

$$a_0 y_{k+n} + \cdots + a_n y_k = z_k, \textit{ for all } k \in \mathbb{Z}$$

is called a **linear difference equation** *(or* **linear recurrence relation***) of order n. For simplicity, a_0 is often taken to 1. If $\{z_k\}$ is the zero sequence, the equation is*

homogeneous; *otherwise, the equation is* **nonhomogeneous**.
In digital signal processing, such a difference equation is called a **linear filter**, *and* $a_0, \cdots, a_n$ *are called the* **filter coefficient**.

In many applications, a sequence $\{z_k\}$ is specified for the right side of a difference equation and a $\{y_k\}$ that satisfied this equation is called a **solution** of the equation.

Solution of a homogeneous difference equation often have the form $y_k = r^k$ for some r. A nonzero signal r^k satisfies the homogeneous different equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0, \ for\ all\ k$$

if and only if r is a root of the **auxiliary equation**

$$r^n + a_1 r^{n-1} + \cdots + a_{n-1} r + a_n = 0$$

We will not consider the case in which r is a repeated root of the auxiliary equation. When the auxiliary equation has a complex root, the difference equation has solutions of the form $s^k cos(kw)$ and $s^k sin(kw)$, for constants s and w.

Given $a_1, \cdots, a_n$, consider the mapping $T : \mathbb{S} \to \mathbb{S}$ that transforms a signal $\{y_k\}$ into a signal $\{w_k\}$ given by

$$w_k = y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k$$

It is readily checked that T is a linear transformation. This implies that the solution set of the homogeneous equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0, \ for\ all\ k$$

is the kernel of T (the set of signals that T maps into the zero signal), and hence the solution set is a subspace of $\mathbb{S}$. Any linear combination of solutions is again a solution. The next theorem, a simple but basic result will lead to more information about the solution sets of difference equations.

Theorem 4.13.3. *If* $a_n \neq 0$ *and if* $\{z_k\}$ *is given, the equation*

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = z_k, \ for\ all\ k$$

has a unique solution whenever $y_0, \cdots, y_{n-1}$ *are specified.*

Proof. If $y_0, \cdots, y_{n-1}$ are specified, we can use the equation to define

$$y_n = z_0 - [a_1 y_{n-1} + \cdots + a_{n-1} y_1 + a_n y_0]$$

And now that $y_1, \cdots, y_n$ are specified use again the equation to define y_{n+1}. In general, use the recurrence relation

$$y_{n+k} = z_k - [a_1 y_{k+n-1} + \cdots + a_n y_k]$$

to define y_{n+k} for $k \geqslant 0$. To define y_k for $k < 0$, use the recurrence relation

$$y_k = \frac{1}{a_n} z_k - \frac{1}{a_n}[y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1}]$$

This produces a signal that satisfies the initial equation. Conversely, any signal that satisfies the equations satisfies

$$y_{n+k} = z_k - [a_1 y_{k+n-1} + \cdots + a_n y_k]$$

and

$$y_k = \frac{1}{a_n} z_k - \frac{1}{a_n}[y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1}],$$

so the solution is unique. □

Theorem 4.13.4. *The set H of all solutions of the nth-order homogeneous linear difference equation*

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0, \text{ for all } k$$

is an n-dimensional vector space.

Proof. As was pointed out earlier, H is a subspace of $\mathbb{S}$ because H is the kernel of a linear transformation. For $\{y_k\}$ in H, let $F\{y_k\}$ be the vector in $\mathbb{R}^n$ given by $(y_0, y_1, \cdots, y_{n-1})$. It is readily verified that $F : H \to \mathbb{R}^n$ is a linear transformation. Given any vector $(y_0, \cdots, y_{n-1})$ in $\mathbb{R}^n$, the previous theorem says that there is a unique signal $\{y_k\}$ in H such that $F\{y_k\} = (y_0, \cdots, y_{n-1})$. This means that F is a one-to-one linear transformation of H onto $\mathbb{R}^n$; that is, F is an isomorphism. Thus $dim(H) = dim(\mathbb{R}^n) = n$. □

The standard way to describe the general solution of a homogeneous difference equation is to exhibit a basis for the subspace of all solutions. Such a basis is usually called a **fundamental set of solution**. In practice, if you can find n linearly independent signals that satisfy the equation, they will be automatically span the n-dimensional solution space.
The general solution of the nonhomogeneous difference equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = z_k$$

can be written as one particular solution plus an arbitrary linear combination of a fundamental solutions of the corresponding homogeneous equation. This fact is analogous showing that the solution setsof $Ax = b$ and $Ax = 0$ are parallel. Both results have the same explanation: The mapping $x \mapsto Ax$ is linear, and the mapping that transforms the signal $\{y_k\}$ into the signal $\{z_k\}$ is linear.

Example 4.13.5. 1. *Verify that the signal $y_k = k^2$ satisfies the difference equation*

$$y_{k+2} - 2y_{k+1} + 3y + k = -4k \text{ for all } k$$

2. *Solutions of a homogeneous difference equation often have the form $y_k = r^k$ for some r. Find solution of this equations of this form.*

3. Find a description of all the solution for this equation

Solution:

1. Substitute k^2 for y_k on the left side:

$$(k+2)^2-4(k+1)^2+3k^2 = (k^2+4k+4)-4(k^2+2k+1) = (k^2+4k+4)-4(k^2+2k+1)+3k^2 = -4k$$

So k^2 is indeed a solution.

2. We know that r^k is solution to this equation if and only if r satisfies the auxiliary equation is

$$r^2 - 4r + 3 = (r-1)(r-3) = 0$$

The roots are $r = 1, 3$. So two solutions of the homogeneous different equation are 1^k and 3^k. They are obviously not multiple of each other, so they are linearly independent signal.

3. By the theorem on the dimension of the solution set of a difference equation, we know that the solution space is two dimensional, so 3^k and 1^k for a basis for the set of solution of this equation. Translating that set by a particular solution of the nonhomogeneous equation, we obtain the general solution of the nonhomogeneous initial equation:

$$k^2 + c_1 1^k + c_2 3^k$$

with c_1, c_2 scalars. Those form a basis for the solution set of this equation.

A modern way to study a homogeneous nth-order linear difference equation is to replace it by an equivalent system of first order difference e quations, written in the form

$$x_{k+1} = Ax_k, \; for\ all\ k$$

where vectors x_k are in $\mathbb{R}^n$ and A is an $n \times n$ matrix.

Example 4.13.6. *Write the following difference equation as a first-order system:*

$$y_{k+3} - 2y_{k+2} - 5y_{k+1} + 6y_k = 0, \; for\ all\ k$$

Solution: *For each k set*

$$x_k = \begin{pmatrix} y_k \\ y_{k+1} \\ y_{k+2} \end{pmatrix}$$

The difference equation says that $y_{k+3} = -6y_k + 5y_{k+1} + 2y_{k+2}$, so

$$x_{k+1} = \begin{pmatrix} y_{k+1} \\ y_{k+2} \\ y_{k+3} \end{pmatrix} = \begin{pmatrix} 0 + y_{k+1} + 0 \\ 0 + 0 + y_{k+2} \\ -6y_k + 5y_{k+1} + 2y_{k+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & 5 & 2 \end{pmatrix} \begin{pmatrix} y_k \\ y_{k+1} \\ y_{k+2} \end{pmatrix}$$

That is,

$$x_{k+1} = Ax_k \; for\ all\ k$$

where

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & 5 & 2 \end{pmatrix}$$

In general, the equation

$$y_{k+n} + a_1 y_{k+n-1} + \cdots + a_{n-1} y_{k+1} + a_n y_k = 0, \text{ for all } k$$

can be rewritten as $x_{k+1} = Ax_k$ for all k, where

$$x_k = \begin{pmatrix} y_k \\ y_{k+1} \\ y_{k+2} \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}$$

4.14 Applications to markov chains

The Markov chains described in this section are used as mathematical models of a wide variety of situations in biology, business, chemistry, engineering, physics and elsewhere. In each case, the model is used to describe an experiment or measurement that is performed many times in the same way, where the outcome of each trial of experiment will be one of several specified possible outcomes, and where the outcome of one trial depends only on the immediately preceding trial.
A vector with non negative entries that add up to 1 is called a **probability vector**. A **stochastic matrix** is a square matrix whose columns are probability vectors. A **Markov chain** is a sequence of probability vectors $x_0, x_1, x_2, \cdots$ together with a stochastic matrix P such that

$$x_1 = Px_0, \ x_2 = Px_1, \ x_3 = Px_2, \cdots$$

When a Markov chain of vectors in $\mathbb{R}^n$ describes a system or a sequence of experiments, the entries in x_k list, respectively, the probabilities that the system is in each n possible states, or the probabilities that the outcome of the experiment is one of n possible outcomes. For this reason, x_k is often called a **state vector**. The most interesting aspect about a Markov chains is the study of chain's long-term behavior.

Example 4.14.1. *Let* $P = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.8 & 0.3 \\ 0.2 & 0 & 0.4 \end{pmatrix}$ *and* $x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. *Consider a system whose state is described by the Markov chain* $x_{k+1} = Px_k$, *for* $k = 0, 1, \cdots$. *What happens to the system as time passes? Compute the state vectors* $x_1, \cdots, x_{15}$ *to find out.*
Solution:

$$x_1 = Px_0 = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.2 \end{pmatrix}$$

$$x_2 = Px_1 = \begin{pmatrix} 0.37 \\ 0.45 \\ 0.18 \end{pmatrix}$$

$$x_3 = Px_2 = \begin{pmatrix} 0.329 \\ 0.525 \\ 0.146 \end{pmatrix}$$

The results of further calculations are shown below, with entries rounded to four or five significant figures.

$$x_4 = \begin{pmatrix} 0.3133 \\ 0.5625 \\ 0.1242 \end{pmatrix}, x_5 = \begin{pmatrix} 0.3064 \\ 0.5813 \\ 0.1123 \end{pmatrix}, x_6 = \begin{pmatrix} 0.3032 \\ 0.5906 \\ 0.1062 \end{pmatrix}, x_7 = \begin{pmatrix} 0.3016 \\ 0.5953 \\ 0.1031 \end{pmatrix}$$

$$x_8 = \begin{pmatrix} 0.3008 \\ 0.5977 \\ 0.1016 \end{pmatrix}, x_9 = \begin{pmatrix} 0.3004 \\ 0.5988 \\ 0.1008 \end{pmatrix}, x_{10} = \begin{pmatrix} 0.3002 \\ 0.5994 \\ 0.1004 \end{pmatrix}, x_{11} = \begin{pmatrix} 0.3001 \\ 0.5997 \\ 0.1002 \end{pmatrix}$$

$$x_{12} = \begin{pmatrix} 0.30005 \\ 0.59985 \\ 0.10010 \end{pmatrix}, x_{13} = \begin{pmatrix} 0.30002 \\ 0.59993 \\ 0.10005 \end{pmatrix}, x_{14} = \begin{pmatrix} 0.30001 \\ 0.59996 \\ 0.10002 \end{pmatrix}, x_{15} = \begin{pmatrix} 0.30001 \\ 0.59998 \\ 0.10001 \end{pmatrix}$$

These vectors seem to be approaching $q = \begin{pmatrix} 0.3 \\ 0.6 \\ 0.1 \end{pmatrix}$. *The probabilities are hardly changing from one value of* k *to the next. Observe that the following calculation is exact (with no rounding error)* $Pq = q$ *When the system is in state* q*, there is no change in the system from one measurement to the next.*

Definition 4.14.2. *If* P *is a stochastic matrix then a* **steady-state vector** *(or* **equilibrium vector***) for* P *is a probability vector* q *such that*

$$Pq = q$$

It can be shown that every stochastic matrix has a steady-state vector.

Example 4.14.3. *Let* $P = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{pmatrix}$. *Find a steady-state vector for* P.

Solution: *The steady-state for* P *is the vector* x *such that* $Px = x$. *That is* $Px - x = 0$, *which is equivalent to* $Px - Ix = 0$. *That is* $(P - I)x - 0$.
For P *above,*

$$P - I = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -0.4 & 0.3 \\ 0.4 & -0.3 \end{pmatrix}$$

To find all solutions of $(P - I)x = 0$, *row reduce the augmented matrix corresponding to this system :*

$$\begin{pmatrix} -0.4 & 0.3 & 0 \\ 0.4 & -0.3 & 0 \end{pmatrix} \sim \textit{Row reduce} \sim \begin{pmatrix} 1 & -3/4 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Reconverting the reduced form of the augmented matrix into a system, we get $x_1 = 3/4x_2$ *and* x_2 *free. The general solution is of the form* $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2 \begin{pmatrix} 3/4 \\ 1 \end{pmatrix}$ *with* x_2 *scalar. Clearly* $\{\begin{pmatrix} 3/4 \\ 1 \end{pmatrix}\}$ *is linearly independent so a basis for the solution set. Another basis is obtained by multiplying this vector by* 4*, thus* $\{w = \begin{pmatrix} 3 \\ 4 \end{pmatrix}\}$ *is also a basis for the solution set of this equation.*
Finally, find a probability vector in the set of all the solutions of $Px = x$*. This is easy, since every solution is a multiple* w *above. Divide* w *by the sum of its entries and obtain*

$$q = \begin{pmatrix} 3/7 \\ 4/7 \end{pmatrix}$$

You can check that effectively we have $Pq = q$*.*

Example 4.14.4. *We say that a stochastic matrix is* **regular** *if some matrix power* P^k *contains only strictly positive entries. Also we say that a sequence of vectors* $\{x_k : k = 1, 2, \cdots\}$ **converges** *to a vector* q *as* $k \to \infty$ *if the entries in* x_k *can be made as close as desired to the corresponding entries in* q *by taking* k *sufficiently large.*

Theorem 4.14.5. *If* P *is an* $n \times n$ *regular stochastic matrix, then* P *has a unique steady-state vector* q*. Further, if* x_0 *is any initial state and* $x_{k+1} = Px_k$*, for* $k = 0, 1, 2, \cdots$*, then the Markov chain* $\{x_k\}$ *converges to* q *as* $k \to \infty$*.*

This theorem is proved in standard text on Markov Chains. The amazing part of the theorem is that the initial state has no effect on the long-term behavior of the Markov Chain.

Chapter 5

Eigenvectors and eigenvalues

5.1 Eigenvectors and eigenvalues

Definition 5.1.1. *An* **eigenvector** *of an $n \times n$ matrix A is a nonzero vector x such that $Ax = \lambda x$ for some scalar λ. A scalar λ is called an eigenvalue of A if there is a non trivial solution x of $Ax = \lambda x$ such an x is called an eigenvector corresponding to λ.*

Warming: row reduction on A cannot be used to find eigenvalues.

We have that the equation $Ax = \lambda x$ is equivalent to $(A - \lambda I)x = 0$. Thus λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation $(A - \lambda I)x$ has a nontrivial solution.

Definition 5.1.2. *The set of all solution of this later equation is just the null space of the matrix $A - \lambda I$. So this set is a subspace of $\mathbb{R}^n$ and is called the* **eigenspace** *of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.*

Theorem 5.1.3. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

Proof. For simplicity consider the 3×3 case. If A is upper triangular, then $A - \lambda I$ has the form

$$A - \lambda I = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} & a_{2,3} \\ 0 & 0 & a_{3,3} \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix} = \begin{pmatrix} a_{1,1} - \lambda & a_{1,2} & a_{1,3} \\ 0 & a_{2,2} - \lambda & a_{2,3} \\ 0 & 0 & a_{3,3} - \lambda \end{pmatrix}$$

The scalar λ is an eigenvalue of A if and only if the equation $(A - \lambda I)x = 0$ has a nontrivial solution, that is, if and only if the equation has a free variable. Because of the zero entries in $A - \lambda I$, it is easy to see that $(A - \lambda I)x = 0$ has a free variable if and only if at least one of the entries on the diagonal of $A - \lambda I$ is zero. This happens if and only if λ equals one of the entries $a_{1,1}$, $a_{2,2}$ and $a_{3,3}$ in A. For the case in which A is lower triangular, the reasoning is very similar. $\square$

Note that 0 is a eigenvalue for A if and only if $Ax = 0$ has a non trivial solution that means in particular that A is not invertible.

Theorem 5.1.4 (Invertible matrix theorem (continued))**.** *Let A be an $n \times n$ matrix. Then A is invertible if and only if 0 is not an eigenvalue of A.*

Theorem 5.1.5. *If $v_1, \cdots, v_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \cdots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{v_1, \cdots, v_r\}$ is linearly independent.*

Proof. Suppose $\{v_1, \cdots, v_r\}$ is linearly dependent, by the linearly dependent vector theorem, we know that one of the vector is a linear combination of the preceding vectors. Let p be the least index such that v_{p+1} is a linear combination of the preceding (linearly independent) vectors. Then there exists scalars $c_1, \cdots, c_p$ such that

$$c_1 v_1 + \cdots + c_p v_p = v_{p+1}$$

Multiplying both sides by A and using the fact that $Av_k = \lambda_k v_k$ for each k, we obtain

$$c_1 A v_1 + \cdots + c_p A v_p = A v_{p+1}$$

$$c_1 \lambda_1 v_1 + \cdots + c_p \lambda_p v_p = \lambda_{p+1} v_{p+1}$$

Combining the preceding equations we obtain:

$$c_1(\lambda_1 - \lambda_{p+1}) v_1 + \cdots + c_p(\lambda_p - \lambda_{p+1}) v_p = 0$$

Since $\{v_1, \cdots, v_p\}$ is linearly independent, the weights in the later equation are all zero. But none of the factors $\lambda_i - \lambda_{p+1}$ are zero, because the eigenvalues are distinct. Hence, $c_i = 0$ for $i = 1, \cdots, p$. But then $v_{p+1} = 0$ which is impossible by the definition of an eigenvector. Hence $\{v_1, \cdots, v_r\}$ are linearly independent. □

Let's go back to difference equations, we want to solve an equation of the kind $Ax_k = x_{k+1}$, $k = 0, 1, 2, \cdots$ with A a $n \times n$ matrix. This equation is called a recursive description of the sequence $\{x_k\}$ in $\mathbb{R}^n$. A **solution** is an explicit description of $\{x_k\}$ whose formula for each x_k does not depend directly on A or on the preceding terms in the sequence other than the initial term x_0.
The simplest way to build a solution is to take an eigenvector x_0 and its corresponding eigenvalue λ and let $x_k = \lambda^k x_0$, $(k = 1, 2, \cdots)$.

This sequence is a solution because

$$Ax_k = A(\lambda^k x_0) = \lambda^k (Ax_0) = \lambda^k (\lambda x_0) = \lambda^{k+1} x_0 = x_{k+1}$$

Linear combinations of solutions are solutions too!

5.2 Characteristic equation and similarity

We know that the equation $(A - \lambda I)x = 0$ has a nontrivial solution if and only if A is not invertible, by the Invertible matrix theorem.

Definition 5.2.1. *The scalar equation* $det(A - \lambda I) = 0$ *is called the* **characteristic equation** *of* A.

Example 5.2.2. *Find the characteristic equation of*

$$A = \begin{pmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Solution: *Form* $A - \lambda I$

$$det(A - \lambda I) = \begin{vmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{vmatrix} = (5-\lambda)(3-\lambda)(5-\lambda)(1-\lambda)$$

The characteristic equation is

$$(5-\lambda)^2(3-\lambda)(1-\lambda)$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$

It can be shown that if A is an $n \times n$ matrix, then $det(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A. The **(algebraic) multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

Example 5.2.3. *The characteristic polynomial of a* 6×6 *matrix is* $\lambda^6 - 4\lambda^5 - 12\lambda^4$. *Find the eigenvalues and their multiplicities.*
Solution: *Factor the polynomial*

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

The eigenvalues are 0 *(multiplicity* 4*),* 6 *(multiplicity* 1*), and* -2 *(multiplicity* 1*).*

Because the characteristic equation for an $n \times n$ matrix involves an nth degree polynomial, the equation has exactly n roots, counting multiplicities, provided complex root are allowed. Such a complex roots are called complex eigenvalues.

Definition 5.2.4. *If* A *and* B *are* $n \times n$ *matrices, then* A *is* **similar to** B *if there is an invertible matrix* P *such that* $P^{-1}AP = B$ *or equivalently* $A = PBP^{-1}$. *Writing* $Q = P^{-1}$, *we have* $Q^{-1}BQ = A$, *so* B *is also similar to* A. *Thus, we say simply that* A *and* B **are similar**. *Changing* A *into* $P^{-1}AP$ *is called a* **similarity transformation**.

Theorem 5.2.5. *If* $n \times n$ *matrices* A *and* B *are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with same multiplicities).*

Proof. If $B = P^{-1}AP$ then

$$B - \lambda I = P^{-1}AP - \lambda PIP^{-1} = P^{-1}(A - \lambda I)P$$

Using the multiplicative property of the determinant, we have

$$det(B - \lambda I) = det(P^{-1}(A - \lambda I)P) = det(P^{-1})det(A - \lambda I)det(P) = det(A - \lambda I)$$

Since $det(P^{-1}P) = det(I) = 1$. So A and B have same characteristic polynomial. □

Remarque 5.2.6. *Be careful!*

1. *Two matrices can have the same eigenvalues but still not be similar, for instance*

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \text{ and } \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

   are not similar even though they have the same eigenvalues.

2. *Similarity is not the same as row equivalence. (If A is row equivalent to B, then $B = EA$ for some invertible matrix E.) Row operations on a matrix usually change its eigenvalues.*

5.3 Diagonalization

Definition 5.3.1. *A square matrix is said to be diagonalizable if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D.*

The next theorem gives a characterization of diagonalizable matrices and tells how to construct a suitable factorization.

Theorem 5.3.2 (The diagonalization theorem)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In fact, $A = PDP^{-1}$ with D a diagonal matrix, if and only if the column of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.*
In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** *of $\mathbb{R}^n$.*

Proof. First observe that is P is any $n \times n$ matrix with columns $v_1, \cdots, v_n$ and if D is any diagonal matrix with diagonal entries $\lambda_1, \cdots, \lambda_n$, then

$$AP = A[v_1, \cdots, v_n] = [Av_1, \cdots, Av_n]$$

while

$$PD = P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = [\lambda_1 v_1, \cdots, \lambda_n v_n]$$

Now, suppose A is diagonalizable and $A = PDP^{-1}$. Then right-multiplying this relation by P, we have $AP = PD$. So that from the previous computation we get

$$[Av_1, \cdots, Av_n] = [\lambda_1 v_1, \cdots, \lambda_n v_n]$$

Equating columns, we find that

$$Av_1 = \lambda_1 v_1,\ Av_2 = \lambda_2 v_2, \cdots, Av_n = \lambda_n v_n$$

Since P is invertible, its columns $v_1, \cdots, v_n$ must be linearly independent. Also, since these columns are non zero, that means that $\lambda_1, \cdots, \lambda_n$ are eigenvalues and $v_1, \cdots, v_n$ are corresponding eigenvectors. This argument proves the "only if" parts of the first and second statement, along the third statement of the theorem.
Finally, given any eigenvalues $\lambda_1, \cdots, \lambda_n$ to construct D. Following the computation above we have $AP = PD$. This is true without any condition on the eigenvectors. If, in fact, the eigenvectors are linearly independent, then P is invertible (by the Invertible Matrix Theorem), and $AP = PD$ implies $A = PDP^{-1}$. □

The following theorem provides a sufficient condition for a matrix to be diagonalizable.

Theorem 5.3.3. *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

Proof. Let $v_1, \cdots, v_n$ be eigenvectors corresponding to the n distinct eigenvalues of a matrix A. Then $\{v_1, \cdots, v_n\}$ is linearly independent, using the theorem about distinct eigenvalues.
Hence A is diagonalizable by the previous theorem. □

It is not necessary for an $n \times n$ matrix to have n distinct eigenvalues in order to be diagonalizable.
If an $n \times n$ matrix A has n distinct eigenvalues with corresponding eigenvector $v_1, \cdots v_n$ and if $P = [v_1, \cdots, v_n]$, then P is automatically invertible because its columns are linearly independent. When A is diagonalizable, but has fewer that n distinct eigenvalues, it is still possible to build P in a way that makes P automatically invertible.

Theorem 5.3.4. *Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \cdots, \lambda_p$.*

1. *For $1 \leqslant k \leqslant p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalues λ_k.*
2. *The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if the characteristic polynomial factors completely into linear factors and the dimension of eigenspace for each λ_k equals the multiplicity of λ_k.*
3. *If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \cdots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.*

The proof of this theorem is not hard but lengthy, you can find it in S. Friedberg, A. Insel and L. Spence, Linear Algebra, 4th edition, Section 5.2.

Example 5.3.5. *Determine if the following matrix is diagonalizable*

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 9 & 0 \\ 3 & 0 & 1 \end{pmatrix}$$

Solution: *This is easy! Since that lower triangular, we know from a theorem of this class that the elements of the main diagonal are eigenvalues, thus* $7, 9, 3$ *are eigenvalues. Since* A *is a* 3×3 *matrix with three distinct eigenvalues,* A *is diagonalizable.*

Example 5.3.6. *Diagonalize if possible*

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 2 & 2 & 0 \\ 2 & 2 & 2 \end{pmatrix}$$

Solution: *Since* A *is lower triangular, we already know its eigenvalues which are the reals of the main diagonal so the only eigenvalue in* 2 *of multiplicity* 3*. In order to know if we can diagonalize* A*, we need to find the dimension of the eigenspace that is* $Nul(A - 2I)$*.*
Note that

$$A - 2I = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 2 & 2 & 0 \end{pmatrix}$$

In order to find the dimension of the eigenspace, we can row reduce the augmented matrix associated to the system $(A - 2I)x = 0$*, that is*

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix} \sim \text{ Row reduce } \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Observing the reduce form of the augmented matrix associated to the system $(A-2I)x = 0$*, we see that there is only one free variable, so that we know that* $dim(Nul(A-2I)) = 1$*. As a consequence, the dimension of the eigenspace which is* 1 *is not equal to multiplicity. That implies that* A *is not diagonalizable.*

Example 5.3.7. *Diagonalize the matrix if possible*

$$A = \begin{pmatrix} 0 & -1 & -1 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{pmatrix}$$

and compute A^8*.*
Solution: *Since there is no easy way to find the eigenvalues, one has to compute the characteristic polynomial since we know that its roots are exactly the eigenvalues. The characteristic polynomial of* A *is*

$$det(A - xI) = \begin{vmatrix} 0 - x & -1 & -1 \\ 1 & 2 - x & 1 \\ -1 & -1 & 0 - x \end{vmatrix}$$

We can use the cofactor expansion along the first row to compute the determinant for instance:

$$\begin{array}{rcl} det(A-xI) & = & -x\begin{vmatrix} 2-x & 1 \\ -1 & 0-x \end{vmatrix} + \begin{vmatrix} 1 & 1 \\ -1 & -x \end{vmatrix} - 1\begin{vmatrix} 1 & 2-x \\ -1 & -1 \end{vmatrix} \\ & = & -x[(2-x)(-x)+1]+[-x+1]-[-1+(2-x)] \\ & = & -x[(2-x)(-x)+1]+[-x+1]-[-1+(2-x)] \\ & = & -x[-2x+x^2+1]-x+1+1-2+x \\ & = & +2x^2-x^3-x=-x(x^2-2x+1)=-x(x-1)^2 \end{array}$$

As a consequence, the eigenvalues are 0 *of multiplicity* 1 *and* 1 *of multiplicity* 2.
Then, in order to see if A *is diagonalizable or not we need to figure out if the dimension of the eigenspaces is or not equal to the multiplicity of the eigenvalues.*
The eigenspace of A *corresponding to* 0 *is* $Nul(A)$. *So we need to find the dimension of the solution set of the equation* $Ax = 0$. *For this we can row reduce the corresponding augmented matrix:*

$$\begin{pmatrix} 0 & -1 & -1 & 0 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 0 \end{pmatrix} \sim \text{ Row reduce } \sim \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We observe the reduced form of the augmented matrix and see that there is one free variable so that the dimension of the eigenspace of 0 *is* 1 *and equal to its multiplicity. The system corresponding to the reduced form of the augmented matrix is*

$$\begin{cases} x_1 - x_3 = 0 \\ x_2 + x_3 = 0 \end{cases}$$

So that

$$Nul(A) = \{t\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, t \text{ scalar}\}$$

And $\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\}$ *is a basis for* $Nul(A)$.
Now, we compute the dimension of the eigenspace of 1 *for* A, *that is the* $Nul(A-I)$. *Note that*

$$A - I = \begin{pmatrix} -1 & -1 & -1 \\ 1 & 1 & 1 \\ -1 & -1 & -1 \end{pmatrix}$$

So, we need to find the dimension of the solution set of $(A-I)x = 0$, *for this we can row reduce the augmented matrix associated to this system*

$$\begin{pmatrix} -1 & -1 & -1 & 0 \\ 1 & 1 & 1 & 0 \\ -1 & -1 & -1 & 0 \end{pmatrix} \sim \text{ Row reduce } \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We observe that we have two free variable, thus $dim(Nul(A-I)) = 2$.

The system associated to the reduced form is $x_1 + x_2 + x_3 = 0$.
Thus a general solution of $(A-I)x = 0$ *is*

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

x_2, x_3 *scalars.*
So $\{\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\}$ *form a basis for the eigenspace of* 1. *So that* A *is diagonalizable and* $A = PDP^{-1}$ *with*

$$P = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $A^8 = (PDP^{-1})^8 = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = (PD^8P^{-1}) = PDP^{-1} = A$.

5.4 Eigenvectors and linear transformations

The goal of this section is to understand the matrix factorization $A = PDP^{-1}$ as a statement about linear transformations. We shall see that the transformation $x \mapsto Ax$ is essentially the same as the very simple mapping $u \mapsto Du$, when viewed from the proper perspective. A similar interpretation will apply to A and D even when D is not a diagonal matrix. Recall that we have proven that any linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be implemented via left multiplication by a matrix A, called the standard matrix of T. Now, we need the same sort of representation for any linear transformation between two finite dimensional vector space.
Let V be an n-dimensional vector space, let W be an m-dimensional vector space and let T be any linear transformation from V to W. To associate a matrix with T, choose (ordered) bases $\mathcal{B}$ and $\mathcal{C}$ for V and W, respectively. Given any x in V, the coordinate vector $[x]_{\mathcal{B}}$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[T(x)]_{\mathcal{C}}$ is in $\mathbb{R}^m$.
The connection between $[x]_{\mathcal{B}}$ and $[T(x)]_{\mathcal{C}}$ is easy to find. Let $\{b_1, \cdots, b_n\}$ be the basis $\mathcal{B}$ for V. If $x = r_1b_1 + \cdots + r_nb_n$, then

$$[x]_{\mathcal{B}} = \begin{pmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_n \end{pmatrix}$$

and
$$T(x) = T(r_1b_1 + \cdots + r_nb_n) = r_1T(b_1) + \cdots + r_nT(b_n)$$
because T is linear. Now, since the coordinate mapping from W to $\mathbb{R}^n$ is linear the previous equation leads to
$$[T(x)]_{\mathcal{C}} = M[x]_{\mathcal{B}}$$
where
$$M = [[T(b_1)]_{\mathcal{C}}, [T(b_2)]_{\mathcal{C}}, \cdots, [T(b_n)]_{\mathcal{C}}]$$
The matrix M is a matrix representation of T, called the **matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$**.
So for as coordinate vectors are concerned, the action of T on x may be viewed as left-multiplication by M.

Example 5.4.1. *Suppose $\mathcal{B} = \{b_1, b_2\}$ is a basis for V and $\mathcal{C} = \{c_1, c_2, c_3\}$ is a basis for W. Let T be a linear transformation with the property that*
$$T(b_1) = c_1 + c_2 \text{ and } T(b_2) = c_2 - c_3$$
Find the matrix M for T relative to $\mathcal{B}$ and $\mathcal{C}$.
Solution: *The $\mathcal{C}$-coordinate vectors of the images of b_1 and b_2 are*
$$[T(b_1)]_{\mathcal{C}} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and } [T(b_2)]_{\mathcal{C}} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$
Hence, the matrix M for T relative to $\mathcal{B}$ and $\mathcal{C}$ is
$$M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix}.$$

If $\mathcal{B}$ and $\mathcal{C}$ are bases for the same space V and if T is the identity transformation $T(x) = x$ for x in V, then the matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$ is just a change of coordinates matrix.
In the common case where W is the same as V and the basis $\mathcal{C}$ is the same as $\mathcal{B}$, the matrix M relative to $\mathcal{B}$ and $\mathcal{C}$ is called the **matrix for T relative to $\mathcal{B}$**, or simply the **$\mathcal{B}$-matrix for T**, and is denoted by $[T]_{\mathcal{B}}$.
The $\mathcal{B}$-matrix for $T : V \to V$ satisfies
$$[T(x)]_{\mathcal{B}} = [T]_{\mathcal{B}}[x]_{\mathcal{B}}, \text{ for all } x \text{ in } V$$

Example 5.4.2. *The mapping $T : \mathbb{P}_2 \to \mathbb{P}_2$ defined by*
$$T(a_0 + a_1t + a_2t^2) = a_1 + 2a_2t$$
is a linear transformation. (You can see that this is just the differentiation operator.)

1. *Find the $\mathcal{B}$-matrix for T, when $\mathcal{B}$ is the basis $\{1, t, t^2\}$.*
2. *Verify the $[T(p)]_{\mathcal{B}} = [T]_{\mathcal{B}}[p]_{\mathcal{B}}$, for each p in $\mathbb{P}_2$.*

Solution:

1. *Compute the images of the basis vectors:*

$$T(1) = 0,\ T(t) = 1,\ T(t^2) = 2t$$

Then write the $\mathcal{B}$ coordinate vectors $T(1), T(t), T(t^2)$ (which are found by inspection in this example) and place them together as the $\mathcal{B}$-matrix for T:

$$[T(1)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},\ [T(t)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, [T(t^2)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

$$[T]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

2. *For a general $p(t) = a_0 + a_1 t + a_2 t^2$,*

$$[T(p)]_{\mathcal{B}} = [a_1 + 2a_2 t]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ 2a_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

In an applied problem involving $\mathbb{R}^n$, a linear transformation T usually appears first as a matrix transformation, $x \mapsto Ax$. If A is diagonalizable, then there is a basis $\mathcal{B}$-matrix for T is diagonal. Diagonalizing A amounts to finding a diagonal matrix representation of $x \mapsto Ax$. If A is diagonalizable, then there is a basis $\mathcal{B}$ for $\mathbb{R}^n$ consisting of eigenvectors of A. In this case, the $\mathcal{B}$-matrix for T is diagonal.

Theorem 5.4.3 (Diagonal Matrix Representation)**.** *Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the $\mathcal{B}$-matrix for the transformation $x \mapsto Ax$.*

Proof. Denote the columns of P by $b_1, \cdots, b_n$, so that $\mathcal{B} = \{b_1, \cdots, b_n\}$ and $P = [b_1, \cdots, b_n]$. In this case, P is the change-of-coordinates matrix $P_{\mathcal{B}}$ where

$$P[x]_{\mathcal{B}} = x \text{ and } [x]_{\mathcal{B}} = P^{-1}x$$

If $T(x) = Ax$ for x in $\mathbb{R}^n$, then

$$\begin{aligned} [T]_{\mathcal{B}} &= [[T(b_1)]_{\mathcal{B}}, \cdots, [T(b_n)]_{\mathcal{B}}] \\ &= [[A(b_1)]_{\mathcal{B}}, \cdots, [A(b_n)]_{\mathcal{B}}] \\ &= [P^{-1}Ab_1, \cdots, P^{-1}Ab_n] \\ &= P^{-1}A[b_1, \cdots, b_n] \\ &= P^{-1}AP \end{aligned}$$

Since $A = PDP^{-1}$, we have $[T]_{\mathcal{B}} = P^{-1}AP = D$. □

Example 5.4.4. *Define* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *by* $T(x) = Ax$, *where* $A = \begin{pmatrix} 7 & 2 \\ -4 & 1 \end{pmatrix}$. *Find a basis* $\mathcal{B}$ *for* $\mathbb{R}^2$ *with the property that the* $\mathcal{B}$*-matrix for* T *is a diagonal matrix.*
Solution: *We can prove that* $A = PDP^{-1}$, *where*

$$P = \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}$$

and

$$D = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$$

The columns of P, *call then* b_1 *and* b_2, *are eigenvectors of* A. *Thus, we have proven that* D *is the* $\mathcal{B}$*-matrix for* T *when* $\mathcal{B} = \{b_1, b_2\}$. *The mapping* $x \mapsto Ax$ *and* $u \mapsto Du$ *describe the same linear transformation, relative to different bases.*

If A is similar to a matrix C, with $A = PCP^{-1}$, then C is the $\mathcal{B}$-matrix for the transformation $x \mapsto Ax$, when the basis $\mathcal{B}$ is formed from the columns of P. Conversely, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(x) = Ax$ and if $\mathcal{B}$ is any basis for $\mathbb{R}^n$, then the $\mathcal{B}$-matrix for T is similar to A. In fact the calculation on the proof of the previous theorem show that if P is the matrix whose columns come from the vectors in $\mathcal{B}$, then $[T]_{\mathcal{B}} = P^{-1}AP$. Thus, the set of all matrices similar to a matrix A coincides with the set of all matrix representations of the transformation $x \mapsto Ax$.

Example 5.4.5. *Let* $A = \begin{pmatrix} 4 & -9 \\ 4 & 8 \end{pmatrix}$, $b_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ *and* $b_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. *The characteristic polynomial of* A *is* $(\lambda + 2)^2$ *but the eigenspace for the eigenvalues* -2 *is only one-dimensional; so* A *is not diagonalizable. However, the basis* $\mathcal{B} = \{b_1, b_2\}$ *has the property that the* $\mathcal{B}$*-matrix for the transformation* $x \mapsto Ax$ *is a triangular matrix called the Jordan form of* A. *Find* $\mathcal{B}$*-matrix.*
Solution: *If* $P = [b_1, b_2]$, *then the* $\mathcal{B}$*-matrix is* $P^{-1}AP$. *Compute*

$$AP = \begin{pmatrix} -6 & -1 \\ -4 & 0 \end{pmatrix}$$

$$P^{-1}AP = \begin{pmatrix} -2 & 1 \\ 0 & -2 \end{pmatrix}$$

Notice that that the eigenvalue of A *is on the diagonal.*

5.5 Discrete dynamical systems

Eigenvalues and eigenvectors provides the key to understanding the long-term behavior or evolution, of a dynamical system described a difference equation $x_{k+1} = Ax_k$. Such an equation was used to model population movement, various Markov chains.... The vector x_k give information about the system as time (denoted by k) passes. The applications in this section focus on ecological problems because they are easier to state and explain than, say, problems in physics or engineering. However, dynamical system

arise in many scientific fields. For instance, standard undergraduate courses in control systems discuss several aspect of dynamical systems. The modern state-space design method in such courses relies heavily on matrix algebra. The steady-state response of a control system is the engineering equivalent of what we call here the "long-term behavior" of the dynamical system $x_{k+1} = Ax_k$.
We assume that A is diagonalizable with n linearly independent eigenvectors $v_1, \cdots, v_n$, and corresponding eigenvalues, $\lambda_1, \cdots, \lambda_n$. For convenience, assume the eigenvectors are arranged so that $|\lambda_1| \geqslant \cdots \geqslant |\lambda_n|$. Since $\{v_1, \cdots, v_n\}$ is a basis for $\mathbb{R}^n$, any initial vector x_0 can be written uniquely as

$$x_0 = c_1 v_1 + \cdots + v_n v_n$$

This eigenvector decomposition of x_0 determines what happens to the sequence $\{x_k\}$. Since the v_1 are eigenvectors,

$$x_1 = Ax_0 = c_1 A v_1 + \cdots + c_n A v_n = c_1 \lambda_1 v_1 + \cdots + c_n \lambda_n v_n$$

In general,

$$x_k = c_1(\lambda_1)^k v_1 + \cdots + c_n(\lambda_n)^k v_n, (k = 0, 1, 2, \cdots)$$

The examples that follow illustrate what can happen as $k \to \infty$.

Deep in the redwood forest of California, dusky-footed wood rats provide up to 80% of the diet for the spotted owl, the main predator of the wood rat. The next example uses a linear system to model the physical system of the owls and the rats. (Admittedly, the model is unrealistic in several respects, but it can provide a starting point for the study of more complicated nonlinear models used by environmentals scientists.)

Example 5.5.1. *Denote the owl and wood rat populations at time k by $x_k = \begin{pmatrix} O_k \\ R_k \end{pmatrix}$ where k is the time in months, O_k is the number of owls in the region studied, and R_k is the number of rats (measured in thousands). Suppose*

$$O_{k+1} = 0.5O_k + 0.4R_k$$

$$R_{k+1} = -pO_k + 1.1R_k$$

where p is a positive parameter to be specified. The $0.5O_k$ in the first equation says that with no wood rats for food, only half of the owls will survive each month, while for $1.1R_k$ in the second equation says that with no owls as predators, the rat population will grow by 10% per month. If rats are plentiful, then $0.4R_k$ will tend to make the owl population rise, while the negative term $-pO_k$ measures the deaths of rats due to predation by owls. (In fact, $1000p$ is the average number of rats eaten by one owl is predation by owls in one month.) Determine the evolution of this evolution of this system when the predation parameter p is 0.104.
Solution: *When $p = 0.104$, the eigenvalues of the coefficient matrix A for the system of equation turn out to be $\lambda_1 = 1.02$ and $\lambda_2 = 0.58$. Corresponding eigeinvectors are*

$$v_1 = \begin{pmatrix} 10 \\ 13 \end{pmatrix}, \; v_2 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

An initial x_0 can be written as

$$x_0 = c_1(1.02)^k v_1 + c_2(0.58)^k v_2 = c_1(1.02)^k \begin{pmatrix} 10 \\ 13 \end{pmatrix} + c_2(0.58)^k \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

As $k \to \infty$, $(0.58)^k$ rapidily approches to zero. Assume $c_1 > 0$. Then, for all sufficiently large k, x_k is approximately the same as $c_1(1,02)^k v_1$, and we write

$$x_k \simeq c_1(1.02)^k \begin{pmatrix} 10 \\ 13 \end{pmatrix}$$

The approximation above improves as k increases, and so for large k.

$$x_{k+1} \simeq c_1(1.02)^{k+1} \begin{pmatrix} 10 \\ 13 \end{pmatrix} \simeq 1.02 x_k$$

The approximation says that eventually both entries of x_k (the numbers of owls and rats) grow by a factor of almost 1.02 each month, a 2% monthly growth rate. x_k is approximately a multiple of $\begin{pmatrix} 10 \\ 13 \end{pmatrix}$, so the entries in x_k are nearly in the same ratio as 10 to 13. That is, for 10 owls there are about 13 thousand rats.

When A is 2×2, algebraic calculations can be supplemented by a geometric description of a system's evolution. We can view the equation $x_{k+1} = Ax_k$ as a description of what happens to an initial point x_0 in $\mathbb{R}^2$ as it is transformed repeatedly by the mapping $x \mapsto Ax$. The graph of $x_0, x_1, \cdots$ is called a **trajectory** of the dynamical system.

Example 5.5.2. *Plot several trajectories of the dynamical system $x_{k+1} = Ax_k$, when*

$$A = \begin{pmatrix} 0.80 & 0 \\ 0 & 0.64 \end{pmatrix}$$

Solution: *The eigenvalues of A are 0.8 and 0.64, with eigenvector $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. If $x_0 = c_1 v_1 + c_2 v_2$, then*

$$x_k = c_1(0.8)^k \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2(0.64)^k \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Of course, x_k tends 0 because $(0.8)^k$ and $(0.64)^k$ both approach 0 as $k \to \infty$. But the way x_k goes toward 0 is interesting.

The origin is called an **attractor** of the dynamical system because all trajectories tend toward 0. This occurs whenever both eigenvalues are less than 1 in magnitude. The direction of greatest attraction is along the line through 0 and the eigenvector v_2 for the eigenvalue of smaller magnitude. In the next example, both eigenvalues of A are larger than 1 in magnitude, and 0 is called a **repeller** of the dynamical system. All solutions of $x_k = Ax_k$, except the (constant) zero solution are unbounded and tend away from the origin.

Example 5.5.3. *Plot several typical solutions of the equation $x_{k+1} = Ax_k$, where*

$$A = \begin{pmatrix} 1.44 & 0 \\ 0 & 1.2 \end{pmatrix}$$

Solution: *The eigenvalues of A are 1.44 and 1.2. If $x_0 = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, then*

$$x_k = c_1(1.44)^k \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2(1.2)^k \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

both term grow in size but the first term grows faster. So the direction of greatest repulsion is the line through 0 and the eigenvector for the eigenvalue of larger magnitude.

In the next example, 0 is called a saddle point because the origin attracts solutions from some directions and repels them in other directions. This occurs whenever one eigenvalue is greater attraction is determined by an eigenvector for the eigenvalue is greater than 1 in magnitude and the other is less than 1 in magnitude. The direction of greatest repulsion is determined by an eigenvector for the eigenvalue of greater magnitude.

Example 5.5.4. *Plot several typical solutions of the solutions of the equation $y_{k+1} = Dy_k$, where*

$$D = \begin{pmatrix} 2.0 & 0 \\ 0 & 0.5 \end{pmatrix}$$

(We write D and y here instead of A and x because this example will be used lated.) Show that a solution $\{y_k\}$ is unbounded if its unbounded if its initial point is not on the x_2-axis.

Solution: *The eigenvalues of D are 2 and 0.5. If $y_0 = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$, then*

$$y_k = c_1 2^k \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2(0.5)^k \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

If y_0 is on the x_2-axis, then $c_1 = 0$ and $y_k \to 0$ as $k \to \infty$. But if y_0 is not on the x_2 axis, then the first term in the sum for y_k becomes arbitrarily large, and so $\{y_k\}$ is unbounded.

The previous examples were about diagonal matrices. To handle the non diagonal matrices , we return for a moment to the $n \times n$ case in which eigenvectors of A form a basis $\{v_1, \cdots, v_n\}$ for $\mathbb{R}^n$. Let $P = [v_1, \cdots, v_n]$, and let D be the diagonal matrix with the corresponding eigenvalues on the diagonal. Given a sequence $\{x_k\}$ satisfying $x_{k+1} = Ax_k$, define a new sequence $\{y_k\}$ by

$$y_k = P^{-1}x_k \; orequivalently, \; x_k = Py_k$$

Substituting these relations into the equation $x_{k+1} = Ax_k$ and using the fact that $A = PDP^{-1}$, we find that

$$Py_{k+1} = APy_k = (PDP^{-1})Py_k = PDy_k$$

Left-multiplying both sides by P^{-1}, we obtain

$$y_{k+1} = Dy_k$$

If we write y_k as $y(k)$ then

$$\begin{pmatrix} y_1(k+1) \\ \cdots \\ y_n(k+1) \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & & & \cdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

The change of variable from x_k to y_k has decoupled the system of difference equation. The evolution of $y_1(k)$ is unaffected by for example, for what happens to $y_2(k), \cdots, y_n(k)$ because $y_1(k+1) = \lambda_1 y_1(k)$, for each k.
The equation $x_k = Py_k$ says that y_k is the coordinate vector of x_k with respect to the eigenvector basis $\{v_1, \cdots, v_n\}$. W can decouple the system $x_{k+1} = Ax_k$ by making calculations in the new eigenvector coordinate system. When $n = 2$, this amounts to using graph paper with axes in the directions of the two eigenvectors.

Chapter 6

Orthogonality and least squares

6.1 Inner product, lenght and orthogonality

Geometric concepts of length, distance and perpendicularity, which are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$, are defined here for $\mathbb{R}^n$.These concept provide powerful geometric tools for solving many applied problems, including the least squares problems. All three notions are defined in terms of the inner product of two vectors.

Definition 6.1.1. *If*

$$u = \begin{pmatrix} u_1 \\ \cdots \\ u_n \end{pmatrix}, \text{ and } v = \begin{pmatrix} v_1 \\ \cdots \\ v_n \end{pmatrix}$$

then the **inner product** *or* **dot product** *of u and v denoted by $u \cdot v$ is*

$$u \cdot v = u^T v = (u_1, \cdots, u_n) \begin{pmatrix} v_1 \\ \cdots \\ v_n \end{pmatrix} = u_1 v_1 + \cdots + u_n v_n$$

Theorem 6.1.2. *Let u, v, and w be vectors in $\mathbb{R}^n$, and let c be a scalar. Then*

1. $u \cdot v = v \cdot u$
2. $(u + v) \cdot w = u \cdot w + v \cdot w$
3. $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
4. $u \cdot u \geqslant 0$ *and* $u \cdot u = 0$ *if and only if* $u = 0$.

The proof is left as an exercise.
Properties 2. and 3. can be combined several times to produce the following useful rule:

$$(c_1 u_2 + \cdots + c_p u_p) \cdot w = c_1(u_1 \cdot w) + \cdot + c_p(u_p \cdot w)$$

Definition 6.1.3. *The* **length** *(or* **norm***) of v is the nonnegative scalar $||v||$ defined by*

$$||v|| = \sqrt{v \cdot v} = \sqrt{v_1^2 + \cdots + v_n^2}$$

and

$$||v||^2 = v \cdot v$$

Suppose v is in $\mathbb{R}^2$, say $v = \begin{pmatrix} a \\ b \end{pmatrix}$. If we identify v with a geometric point in the plane, as usual, then $||v||$ coincides with the standard notion of the length of the line segment from the origin to v. This follows from the Pythagorean Theorem.
A similar calculation with the diagonal of a rectangular box shows that the definition of length of a vector v in $\mathbb{R}^3$ coincides with the usual notion of length.
For any scalar c, the length of cv is $|c|$ times the length of v. That is,

$$||cv|| = |c|||v||$$

(To see this, compute

$$||cv||^2 = (cv) \cdot (cv) = c^2 v \cdot v = c^2||v||^2$$

and take square roots).

Definition 6.1.4. *A vector whose length is* 1 *is called a* **unit vector**. *If we divide a nonzero vector* v *by its length- that is, multiply by* $1/||v||$*- we obtain a unit vector* u *because the length of* u *is* $(1/||v||)||v||$. *The process of creating* u *from* v *is sometimes called the* **normalizing** v, *and we say that* u *is in the same direction as* v.

Example 6.1.5. *Let* $v = (1, -2, 2, 0)$. *Find a unit vector* u *in the same direction as* v.
Solution: *First, compute the length of* v*:*

$$||v||^2 = v \cdot v = (1)^2 + (-2)^2 + 2^2 + 0^2 = 9$$

$$||v|| = \sqrt{9} = 3$$

Then v *by* $1/||v||$ *to obtain*

$$u = (1/||v||)v = 1/3v = \begin{pmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{pmatrix}$$

To check that $||u|| = 1$, *it suffice to show that* $||u||^2 = 1$

$$||u||^2 = u \cdot u = 1/9 + 4/9 + 4/9 = 1$$

We are ready now to describe how close one vector is to another. Recall that if a and b are real numbers, the distance on the number between a and b is the number $|a - b|$.

Definition 6.1.6. *For* u *and* v *in* $\mathbb{R}^n$, *the* **distance between** u **and** v, *written as* $dist(u, v)$ *is the length of the vector* $u - v$. *That is,*

$$dist(u, v) = ||u = v||$$

Example 6.1.7. *Compute the distance between the vectors* $u = (7, 1)$ *and* $v = (3, 2)$.
Solution:

$$u - v = \begin{pmatrix} 4 \\ -1 \end{pmatrix}$$

$$||u - v|| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

Example 6.1.8. *If* $u = (u_1, u_2, u_3)$ *and* $v = (v_1, v_2, v_3)$, *then*

$$dist(u, v) = ||u - v|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$$

The concept of perpendicular lines in ordinary Euclidean geometry has an analogue in $\mathbb{R}^n$.
Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors u and v. The two lines shown are geometrically perpendicular if and only if the distance from u to v is the same as the distance from u and $-v$. This is the same as requiring the square of the distances to be the same.

$$(dist(u, -v))^2 = ||u||^2 + ||v||^2 + 2u \cdot v$$

$$(dist(u, v))^2 = ||u||^2 + v||^2 - 2u \cdot v$$

Thus $2u \cdot v = -2u \cdot u$ and so $u \cdot v = 0$.
The following definition generalizes to $\mathbb{R}^n$ this notion of perpendicularity (or orthogonality, as it is commonly called in linear algebra).

Definition 6.1.9. *Two vectors* u *and* v *in* $\mathbb{R}^n$ *are* **orthogonal** *(to each other) if* $u \cdot v = 0$.

Observe that the zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $0^T v = 0$.
The next theorem provide a useful fact about orthogonal vectors. The proof follows immediately from the previous calculation and the definition of orthogonal vectors.

Theorem 6.1.10 (The pythagorean theorem)**.** *Two vectors* u *and* v *are orthogonal if and only if* $||u + v||^2 = ||u||^2 + ||v||^2$.

Definition 6.1.11. *If a vector* z *is orthogonal to every vector in a subspace* W *of* $\mathbb{R}^n$, *then* z *is said to be* **orthogonal to** W. *The set of all vectors* z *that are orthogonal to* W *is called the* **orthogonal complement** *of* W *and is denoted by* W^T *(and read as "W perpendicular" or simply "W perp").*

Example 6.1.12. *Let* W *be a plane through the origin in* $\mathbb{R}^3$, *and let* L *be the line through the origin and perpendicular to* W. *If* z *and* w *are nonzero,* z *is on* L, *and* w *is in* W; *that is,* $z \cdot w = 0$. *So each vector on* L *is orthogonal to every* w *in* W. *In fact,* L *consists of all vectors that are orthogonal to the* $w's$ *in* W, *and* W *consists of all vectors orthogonal to the* $z's$ *in* L. *That is,*

$$L = W^T \text{ and } W = L^T$$

Fact 6.1.13.
1. *A vector* x *is in* W^T *if and only if* x *is orthogonal to every vector in a set that spans* W.
2. W^T *is a subspace of* $\mathbb{R}^n$.

The proof is left as an exercise.

Theorem 6.1.14. *Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:*

$$(Row(A))^T = Nul(A) \text{ and } (Col(A))^T = Nul(A^T)$$

Proof. The row-column for computing Ax shows that if x is in $Nul(A)$, then x is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, x is orthogonal to $Row(A)$. Conversely, if x is orthogonal to $Row(A)$, then x is certainly orthogonal to each row of A, and hence $Ax = 0$. This proves the first statement of the theorem. Since the statement is true for any matrix, it is true for A^T. That is orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because $Row(A^T) = Col(A)$. □

6.2 Orthogonal sets

Definition 6.2.1. *A set of vectors $\{u_1, \cdots, u_p\}$ is said to be an* **othogonal set** *if each pair of distinct vectors from the set is orthogonal, that is, if $u_i \cdot u_j = 0$ whenever $i \neq j$.*

Example 6.2.2. *Show that $\{u_1, u_2, u_3\}$ is an orthogonal set where*

$$u_1 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}, u_2 = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, u_3 = \begin{pmatrix} -1/2 \\ -2 \\ 7/2 \end{pmatrix}$$

Solution: *Consider the three possible pairs of distinct vectors, namely, $\{u_1, u_2\}$, $\{u_1, u_3\}$ and $\{u_2, u_3\}$.*

$$u_1 \cdot u_2 = 3(-1) + 1(2) + 1(1) = 0$$
$$u_1 \cdot u_3 = 3(-1/2) + 1(-2) + 1(7/2) = 0$$
$$u_2 \cdot u_3 = -1(-1/2) + 2(-2) + 1(7/2) = 0$$

Each pair of distinct vectors is orthogonal, and so $\{u_1, u_2, u_3\}$ is an orthogonal set.

Theorem 6.2.3. *If $S = \{u_1, \cdots, u_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.*

Proof. If $0 = c_1u_1 + \cdots + c_pu_p$ for some scalars $c_1, \cdots, c_p$, then

$$\begin{aligned} 0 &= 0 \cdot u_1 = (c_1u_1 + c_2u_2 + \cdots + c_pu_p) \cdot u_1 \\ &= (c_1u_1) \cdot u_1 + (c_2u_2) \cdot u_1 + \cdots + (c_pu_p) \cdot u_1 \\ &= c_1(u_1 \cdot u_1) + c_2(u_2 \cdot u_1) + \cdots + c_p(u_p \cdot u_1) \\ &= c_1(u_1 \cdot u_1) \end{aligned}$$

because u_1 is orthogonal to $u_2, \cdots, u_p$. Since u_1 is nonzero, $u_1 \cdot u_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \cdots, c_p$ must be 0. Thus S is linearly independent. □

Definition 6.2.4. *An* **orthogonal basis** *for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.*

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weigths in a linear combination can be computed easily.

Theorem 6.2.5. *Let $\{u_1, \cdots, u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each y in W, the weights in the linear combination*

$$y = c_1 y_1 + \cdots + c_p u_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \ (j = 1, \cdots, p)$$

Proof. As in the preceding proof, the orthogonality of $\{u_1, \cdots, u_p\}$, shows that

$$y \cdot u_1 = (c_1 u_1 + \cdots + c_p u_p) \cdot u_1 = c_1(u_1 \cdot u_1)$$

Since $u_1 \cdot u_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \cdots, p$, compute $y \cdot u_j$ and solve for c_j. $\square$

We turn to a construction that will become a key step step in many calculation involving orthogonality.
Given a nonzero vector u in $\mathbb{R}^n$, consider the problem of decomposing a vector y in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of u and the other orthogonal to u. We wish to write

$$y = \hat{y} + z \ (*)$$

where $\hat{y} = \alpha u$ for some scalar α and z is some vector orthogonal to u. Given any scalar α, let $z = y - \alpha u$. Then, $y - \hat{y}$ is orthogonal to u if and only if

$$0 = (y - \alpha u) \cdot u = y \cdot u - (\alpha u) \cdot u = y \cdot u - \alpha(u \cdot u)$$

That is $(*)$ is satisfied with z orthogonal to u if and only if $\alpha = \frac{y \cdot u}{u \cdot u}$ and $\hat{y} = \frac{y \cdot u}{u \cdot u} u$.

Definition 6.2.6. *The vector $\hat{y}$ is called the* **orthogonal projection of y onto u**, *and the vector z is called the* **component of y orthogonal to u**.
If c is any nonzero scalar and if u is replaced by cu in the definition of $\hat{y}$, then the orthogonal projection of y onto u is exactly the same as the orthogonal projection of y onto u. Hence this projection is determined by the subspace L spanned by u (the line through u and 0). Sometimes $\hat{y}$ is denoted by $proj_L(y)$ and is called the **orthogonal projection of y onto L**. *That is,*

$$\hat{y} = proj_L(y) = \frac{y \cdot u}{u \cdot u} u$$

Example 6.2.7. *Let $y = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$ and $u = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. Find the orthogonal projection of y onto u. Then write y as the sum of two orthogonal vectors, one in $Span\{u\}$ and one orthogonal to u.*
Solution: *Compute,*

$$y \cdot u = 40$$

$$u \cdot u = 20$$

The orthogonal projection of y onto u is

$$\hat{y} = \frac{y \cdot u}{u \cdot u} u = \frac{40}{20} u = 2 \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix}$$

and the component of y orthogonal to u is

$$y - \hat{y} = \begin{pmatrix} 7 \\ 6 \end{pmatrix} - \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

The sum of these two vector is y.
Note that if the calculation above are correct $\{\hat{y}, y - \hat{y}\}$ will be an orthogonal set. (You can check that $\hat{y} \cdot (y - \hat{y}) = 0$.
Since the line segment between y and $\hat{y}$ is perpendicular to L the line passing through the origin with direction u, by construction of $\hat{y}$, the point identified with $\hat{y}$ is the closest point of L to y. (Assume this is true for $\mathbb{R}^2$.) Then the distance from y to L is the length of $y - \hat{y}$. Thus, this distance is

$$||y - \hat{y}|| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$

Definition 6.2.8. *A set $\{u_1, \cdots, u_p\}$ is an* **orthonormal set** *if it is an orthonormal set of units vectors If W is the subspace spanned by such a set then $\{u_1, \cdots, u_p\}$ is an* **orthonormal basis** *for W, since the set is automatically linearly independent, as we have proven earlier.*

The simplest example of an orthonormal set is the standard basis $\{e_1, \cdots, e_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{e_1, \cdots, e_n\}$ is orthonormal, too.
When the vectors in an orthogonal set of nonzero vectors are normalized to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set. Matrices whose column form an orthonormal set are important in application and in computer algorithms for matrix computations. Here some of their properties.

Theorem 6.2.9. *An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.*

Proof. To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is essentially the same. Let $U = [u_1, u_2, u_3]$ and compute

$$U^T U = \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} (u_1, u_2, u_3) = \begin{pmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{pmatrix}$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$u_1^T u_2 = u_2^T u_1 = 0, u_1^T u_3 = u_3^T u_1 = 0, u_2^T u_3 = u_3^T u_2 = 0$$

The columns of U all have unit length if and only if

$$u_1^T u_1 = 1,\ u_2^T u_2 = 1, u_3^T u_3 = 1$$

And the theorem follows. □

Theorem 6.2.10. *Let U be an $m \times n$ matrix orthonormal columns, and let x and y be in $\mathbb{R}^n$. Then*

1. $||Ux|| = ||x||$;
2. $(Ux) \cdot (Uy) = x \cdot y$;
3. $(Ux) \cdot (Uy) = 0$ *if and only if* $x \cdot y = 0$.

Properties 1. and 3. say that the linear mapping $x \mapsto Ux$ preserves lengths and orthogonality. These properties are crucial for many computer algorithms. The proof is left as an exercise.

Example 6.2.11. *Let* $U = \begin{pmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{pmatrix}$ *and* $x = \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix}$. *Notice that U has orthogonal columns and $U^T U = I_2$.*
Verify that $||Ux|| = ||x||$.
Solution:

$$Ux = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$$

$$||Ux|| = \sqrt{9+1+1} = \sqrt{11}$$

$$||x|| = \sqrt{2+9} = \sqrt{11}$$

Definition 6.2.12. *An* **orthogonal matrix** *is a square invertible matrix such that $U^{-1} = U^T$.*

Such a matrix has orthonormal columns. It is easy to see that any square matrix with orthonormal column is orthogonal. Surprisingly, such a matrix must have orthogonal rows too. (Exercise).

6.3 Orthogonal projection

The orthogonal projection of a point in $\mathbb{R}^2$ has an important analogue in $\mathbb{R}^n$. Given a vector y and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{y}$ in W such that $\hat{y}$ is the unique vector in W for which $y - \hat{y}$ is orthogonal to W, and $\hat{y}$ is the unique vector closest to y. To prepare for the first theorem, observe that whenever a vector y is written as a linear combination of $u_1, \cdots, u_n$ in $\mathbb{R}^n$, the terms in the sum for y can be grouped into two parts so that y can be written as

$$y = z_1 + z_2$$

where z_1 is a linear combination of the some u_i and z_2 is a linear combination of the rest of the u_i. The idea particularly useful when $\{u_1, \cdot, u_n\}$ is an orthogonal basis. Recall that W^T denotes the set of all vectors orthogonal to a subspace W.

Example 6.3.1. *Let $\{u_1, \cdots, u_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let*

$$y = c_1u_1 + \cdots + c_5u_5$$

Consider the subspace $W = Span\{u_1, u_2\}$ and write y as the sum of a vector z_1 in W and a vector z_2 in $W^\perp$.

Solution: *Write*

$$y = c_1u_1 + c_2u_2 + c_3u_3 + c_4u_4 + c_5u_5$$

where

$$z_1 = c_1u_1 + c_2u_2$$

is in $Span\{u_1, u_2\}$, and

$$z_2 = c_3u_3 + c_4u_4 + c_5u_5$$

is in $Span\{u_3, u_4, u_5\}$.

To show that z_2 is in $W^\perp$, it suffices to show that z_2 is orthogonal to the vectors in the basis $\{u_1, u_2\}$ for W. Using properties of the inner product compute

$$z_2 \cdot u_1 = (c_3u_3 + c_4u_4 + c_5u_5) \cdot u_1 = c_3u_3 \cdot u_1 + c_4u_4 \cdot u_1 + c_5u_5 \cdot u_1 = 0$$

because u_1 is orthogonal to u_3, u_4 and u_5. A similar calculation shows that $z_2 \cdot u_2 = 0$. Thus z_2 is in W^T

The next theorem shows that the decomposition $y = z_1 + z_2$ can be computed without having an orthogonal basis for $\mathbb{R}^n$. It is enough to have an orthogonal.

Theorem 6.3.2 (The Orthogonal decomposition theorem)**.** *Let W be a subspace of $\mathbb{R}^n$. Then each y in $\mathbb{R}^n$ can be written uniquely in the form*

$$y = \hat{y} + z$$

where $\hat{y}$ is in W and z is in W^T. In fact, if $\{u_1, \cdots, u_n\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1}u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p}u_p$$

and $z = y - \hat{y}$.

The vector $\hat{y}$ is called the **orhogonal projection of y onto** *W and often is written as $proj_W y$.*

Proof. Let $\{u_1, \cdots, u_p\}$ be any orthogonal basis for W and define

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1}u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p}u_p$$

Then $\hat{y}$ is a linear combination of basis $u_1, \cdots, u_p$. Let $z = y - \hat{y}$. Since u_1 is orthogonal to $u_2, \cdots, u_p$, it follows that

$$z \cdot u_1 = (y - \hat{y}) \cdot u_1 = y \cdot u_1 - \left(\frac{y \cdot u_1}{u_1 \cdot u_1}\right) - 0 = y \cdot u_1 - y \cdot u_1 = 0$$

Thus z is orhtogonal to u_1. Similarly z is orthogonal to each u_j in the basis for W. Hence z is orthogonal to every vector in W. That is, z is in $W^{\perp}$.
To show that the decomposition is unique, suppose y can also be written as $y = \hat{y}_1 + z_1$, with $\hat{y}_1$ in W and z_1 in $W^{\perp}$. Then $\hat{y} + z = \hat{y}_1 + z_1$ (since both side equal y), and so

$$\hat{y} - \hat{y}_1 = z_1 - z$$

This equality shows that the vector $v = \hat{y} - \hat{y}_1$ is in W and in $W^{\perp}$ (because z_1 and and z are both in $W^{\perp}$ and $W^{\perp}$ is a subspace). Hence, $v \cdot v = 0$, which shows that $v = 0$. This proves that $\hat{y} = \hat{y}_1$ and also $z_1 = z$. $\square$

The uniqueness of the decomposition shows that the orthogonal projection $\hat{y}$ depends only on W and not on the particular basis used.

Example 6.3.3. *Let* $u_1 = \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix}, u_2 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$ *and* $y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. *Observe that* $\{u_1, u_2\}$ *is an orthogonal basis for* $W = Span\{u_1, u_2\}$. *Write* y *as the sum of a vector in* W *and a vector orthogonal to* W.
Solution: *The orthogonal projection of* y *onto* W *is*

$$\begin{aligned} \hat{y} &= \tfrac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \tfrac{y \cdot u_2}{u_2 \cdot u_2} u_2 \\ &= 9/30 \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + 3/6 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix} \end{aligned}$$

Also,

$$y - \hat{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 7/5 \\ 0 \\ 14/5 \end{pmatrix}$$

By the previous theorem we know that $y - \hat{y}$ *is in* $W^{\perp}$. *To check the calculations, however, it is a good idea to verify that* $y - \hat{y}$ *is orthogonal to both* u_1 *and* u_2 *and hence to all* W. *The desired decomposition of* y *is*

$$y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix} + \begin{pmatrix} 7/5 \\ 0 \\ 14/5 \end{pmatrix}$$

If $\{u_1, \cdots, u_p\}$ is an orthogonal basis for W and if y happens to be in W, then $proj_W(y) = y$.

Theorem 6.3.4 (The best approximation theorem)**.** *Let W be a subspace of $\mathbb{R}^n$ and let $\hat{y}$ be the orthogonal projection of y into W. Then $\hat{y}$ is the closest point in W to y, in the sense that*

$$||y - \hat{y}|| < ||y - v||$$

for all v in W distinct from $\hat{y}$.

The vector $\hat{y}$ is called **the best approximation to y be elements of** W.The distance from y to v is given by $||y - v||$, can be regarded as the "error" of using v in place of y. The previous theorem says that this error is minimized when $v = \hat{y}$.
The previous theorem leads to a new proof that $\hat{y}$ does not depend on the particular orthogonal basis used to compute it. If a different orthogonal basis for W were used to construct an orthogonal projection of y, then this projection would also be the closest point in W to y, namely, $\hat{y}$.

Proof. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W. By the Orthogonal Decomposition Theorem, $y - \hat{y}$ is orthogonal to W. In particular, $y - \hat{y}$ is orthogonal to $\hat{y} - v$ (which is in W). Since

$$y - v = (y - \hat{y}) + (\hat{y} - v)$$

the Pythagorean Theorem gives

$$||y - v||^2 = ||y - \hat{y}||^2 + ||\hat{y} - v||^2$$

Now, $||\hat{y} - v||^2 > 0$ because $\hat{y} - v \neq 0$, and so inequality

$$||y - \hat{y}|| < ||y - v||$$

follows immediately. □

Example 6.3.5. *The distance from a point y in $\mathbb{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W. Find the distance from y to $W = Span\{u_1, u_2\}$ where*

$$y = \begin{pmatrix} -1 \\ -5 \\ 10 \end{pmatrix}, u_1 = \begin{pmatrix} 5 \\ -2 \\ 1 \end{pmatrix}, u_2 = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

Solution : *By the Best Approximation theorem, the distance from y to W is $||y - \hat{y}||$, where $\hat{y} = proj_W(y)$. Since $\{u_1, u_2\}$ is orthogonal basis for W,*

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \begin{pmatrix} -1 \\ -8 \\ 4 \end{pmatrix}$$

$$y - \hat{y} = \begin{pmatrix} 0 \\ 3 \\ 6 \end{pmatrix}$$

$$||y - \hat{y}||^2 = 3^2 + 6^2 = 45$$

The distance from y to W is $\sqrt{45} = 3\sqrt{5}$.

The final theorem shows how the formula for $proj_W(y)$ is simplified when the basis for W is an orthonormal set.

Theorem 6.3.6. *If $\{u_1, \cdots, u_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then*

$$proj_W(y) = (y \cdot u_1)u_1 + \cdots + (y \cdot u_p)u_p$$

If $U = [u_1, \cdots, u_p]$, then

$$proj_W(y) = UU^T y, \ forall\ y\ in\ \mathbb{R}^n$$

Proof. The first formula is a immediate consequence of the previous results. This formula shows that $proj_W(y)$ is a linear combination of the columns of U using the weights $y \cdot u_1, \cdots, y \cdot u_p$. The weigths can be written as $u_1^T y, \cdots, u_p^T y$, showing that they are the entries in $U^T y$ and proving the second formula. □

Suppose U is an $n \times p$ matrix with orthogonal columns, and let W be the column space of U. Then

$$U^T U x = I_p x = x \ for\ all\ x\ in\ \mathbb{R}^p$$

$$UU^T y = proj_W(y) = proj_W(y) \ for\ all\ y\ in\ \mathbb{R}^n$$

We will use the formula of the last theorem only for theory in practice we rather use the previous one.

6.4 The Gram-Schmith process

Theorem 6.4.1 (The Gram-Schmith Process)**.** *Given a basis $\{x_1, \cdots, x_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define*

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

$$\cdots$$

$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \cdots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

Then $\{v_1, \cdots, v_p\}$ is an orthogonal basis for W. In addition

$$Span\{v_1, \cdots, v_k\} = Span\{x_1, \cdots, x_k\} \ for\ 1 \leqslant k \leqslant p$$

Proof. For $1 \leqslant k \leqslant p$, let $W_k = Span\{x_1, \cdots, x_k\}$. Set $v_1 = x_1$, so that $Span\{v_1\} = Span\{x_1\}$. Suppose, for some $k < p$, we have constructed $v_1, \cdot, v_k$ so that $\{v_1, \cdots, v_k\}$ is an orthogonal basis for W_k. Define

$$v_{k+1} = x_{k+1} - proj_{W_k}(x_{k+1})$$

By the Orthogonal Decomposition Theorem, v_{k+1} is orthogonal to W_k. Note that $proj_{W_k}(x_{k+1})$ is in W_k and hence also in W_{k+1}, so is v_{k+1} (because W_{k+1} is a subspace and is closed under substraction). Furthermore, $v_{k+1} \neq 0$ because x_{k+1} is not in $W_k = Span\{x_1, \cdots, x_k\}$. Hence $\{v_1, \cdots, v_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1)$-dimensional space W_{k+1}. By the Basis Theorem, this set is an orthogonal basis for W_{k+1}. Hence, $W_{k+1} = Span\{v_1, \cdots, v_{k+1}\}$. When $k+1 = p$, the process stops. □

The previous theorem shows that any nonzero subspace W of $\mathbb{R}^n$ has an orthogonal basis, because an ordinary basis $\{x_1, \cdots, x_p\}$ is always available, and the Gram-Schmidt process depends only on the existence of orthogonal projections onto subspaces of W that already have orthogonal bases.

An orthonormal basis is constructed easily from an orthogonal basis $\{v_1, \cdots, v_p\}$ simply normalize (i.e. "scale") all the v_k.

Example 6.4.2. *Let* $x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, *and* $x_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$.
Then $\{x_1, x_2, x_3\}$ *is clearly linearly independent and thus is a basis for a subspace* W *of* $\mathbb{R}^4$. *Construct an orthogonal basis for* W.
Solution: *We apply the Gram-Schmith process.*
Let $v_1 = x_1$ *and* $W_1 = Span\{x_1\} = Span\{v_1\}$.
Let v_2 *be the vector produced by subtracting from* x_2 *its projection onto the subspace* W_1. *That is, let*

$$v_2 = x_2 - proj_{W_1}(x_2) = \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

v_2 *is the component of* x_2 *orthogonal to* x_1 *and* $\{v_1, v_2\}$ *is an orthogonal basis for the subspace* W_2 *spanned by* x_1 *and* x_2.
Let v_3 *be the vector produced by subtracting* x_3 *its projection onto the subspace* W_2. *Use the orthogonal basis* $\{v_1, v_2'\}$ *to compute this projection onto* W_2*:* $v_3 = x_3 - proj_{W_2}x_3 = \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}$.
Thus, we have obtained an orthogonal basis $\{v_1, v_2, v_3\}$, *applying the Gram-Schmith process.*

6.5 QR factorization of matrices

If an $m \times n$ matrix A has linearly independent columns $x_1, \cdots, x_n$, then applying, the Gram-Schmidt process (with normalizations) to $x_1, \cdots, x_n$ amounts to factoring A, as described in the next theorem. This factorization is widely used in computer algorithms for various computation.

Theorem 6.5.1 (The QR factorization)**.** *If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $Col(A)$ and R is an $m \times n$ upper triangular invertible matrix with positive entries on the diagonal.*

Proof. The columns of A form a basis $\{x_1, \cdots, x_n\}$ for $Col(A)$. Construct an orthonormal basis $\{u_1, \cdots, u_n\}$ for $W = Col(A)$. This basis may be constructed by the Gram-Schmith process or some other means. Let

$$Q = [u_1, \cdots, u_n]$$

For $k = 1, \cdots, n$, x_k is in $Span\{x_1, \cdots, x_k\} = Span\{u_1, \cdots, u_k\}$, so there are constants, $r_{1k}, \cdots, r_{k,k}$, such that

$$x_k = r_{1k}u_1 + \cdots + r_{kk}u_k + 0 \cdot u_{k+1} + \cdots + 0 \cdot u_n$$

We may assume that $r_{kk} \geqslant 0$. (If $r_{kk} < 0$, multiply both r_{kk} and u_k by -1). This shows that x_k is a linear combination of the columns of Q using as weights the entries in the vector

$$r_k = \begin{pmatrix} r_{1k} \\ \cdot \\ \cdot \\ r_{kk} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

That is, $x_k = Qr_k$, for $k = 1, \cdots, n$. Let $R = [r_1, \cdots, r_n]$. Then

$$A = [x_1, \cdots, x_n] = [Qr_1, \cdots, Qr_n] = QR$$

The fact R is invertible follows easily from the fact that the columns of A are linearly independent. Since R is clearly upper triangular, its nonnegative diagonal entries must be positive. □

Example 6.5.2. *Find a QR factorization of*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Solution: *The columns of A are the vectors x_1, x_2 and x_3. An orthogonal basis for $Col(A) = Span\{x_1, x_2, x_3\}$ was found:*

$$\{v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix}\}$$

Then, normalize the three vectors to obtain u_1, u_2 and u_3 and use these vectors as the columns of Q:

$$Q = \begin{pmatrix} 1/2 & -3\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix}$$

By construction, the first k columns of Q are an orthonormal basis of $Span\{x_1, \cdots, x_k\}$. From the previous theorem, $A = QR$ for some R. To find R, observe that $Q^TQ = I$ because the columns of Q are orthonormal. Hence

$$Q^TA = Q^T(QR) = IR = R$$

and

$$R = \begin{pmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{pmatrix}$$

6.6 Least-Square problems

Inconstent systems $Ax = b$ arise often in application, though usually not with such an enormous coefficient matrix. When a solution is demanded and none exist, the best one can do is to find an x that makes Ax as close as possible to b.
Think of Ax as an approximation to b. The smaller the distance between b and Ax given by $||b - Ax||$, the better the approximation. The **general least-square problem** is to find an x that makes $||b - Ax||$ as small as possible. The adjective "least-square" arises from the fact that $||b - Ax||$ is the square root of a sum of squares.

Definition 6.6.1. *If A is $m \times n$ and b is in $\mathbb{R}^m$, a* **least-square solution** *of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that*

$$||b - A\hat{x}|| \leqslant ||b - Ax||$$

for all x in $\mathbb{R}^n$.

Of course if b happens to be in $Col(A)$, then b is equal to Ax for some x and such an x is a "least-square solution".
Given A an b as above, apply the Best Approximation Theorem to the column space $Col(A)$. Let

$$\hat{b} = proj_{Col(A)}(b)$$

Because $\hat{b}$ is in the column space of A, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbb{R}^n$ such that

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

Since $\hat{b}$ is the closest point in $Col(A)$ to b, a vector $\hat{x}$ is a least-square solution of $Ax = b$ if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$. Such an $\hat{x}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{b}$ out of the columns of A. There are many solutions of $A\hat{x} = \hat{b}$ is the equation has free variables.

Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$. By the Orthogonal Decomposition Theorem, the projection $\hat{b}$ has the property that $b - \hat{b}$ is orthogonal to $Col(A)$, $b - A\hat{x}$ is orthogonal to each column of A. If a_j is any column of A then $a_j \cdot (b - A\hat{x}) = 0$. Since each a_j^T is a row of A^T,

$$A^T(b - A\hat{x}) = 0$$

Thus,

$$A^T b - A^T A\hat{x} = 0$$

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

These calculation show that each least-square solution of $Ax = b$ satisfies the equation

$$A^T Ax = A^t b$$

The matrix equation $A^T Ax = A^t b$ represents a system of equation called **normal equations for** $Ax = b$. A solution of $A^T Ax = A^t b$ is often denoted by $\hat{x}$.

Theorem 6.6.2. *The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^T Ax = A^T b$.*

Proof. As shown above the set of least-squares solutions is non empty and each least-squares solution $\hat{x}$ satisfies the normal equations. Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$. Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$ which shows that $b - A\hat{x}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A. Since the columns of A span $Col(A)$, the vector $b - A\hat{x}$ is orthogonal to all of $Col(A)$. Hence the equation

$$b = A\hat{x} + (b - A\hat{x})$$

is a decomposition of b into the sum of a vector in $Col(A)$. By the uniqueness of the orthogonal decomposition, $A\hat{x}$ must be the orthogonal projection of b onto $Col(A)$. That is, $A\hat{x} = \hat{b}$, and $\hat{x}$ is a least-squares solution. □

The next theorem gives useful criterion for determining when there is only one least-squares solution of $Ax = b$. Of course, the orthogonal projection $\hat{b}$ is always unique.

Theorem 6.6.3. *Let A be an $m \times n$ matrix. The following statements are logically equivalent:*

1. *The equation $Ax = b$ has a unique least-squares solution for each b in $\mathbb{R}^m$.*
2. *The columns of A are linearly independent.*
3. *The matrix $A^T A$ is invertible.*

When these statements are true, the least-square solution $\hat{x}$ is given by

$$\hat{x} = (A^T A)^{-1} A^T b$$

The proof is left as exercise. The formula $\hat{x} = (A^T A)^{-1} A^T b$ for $\hat{x}$ is useful mainly for theoretical purposes and for hand calculations when $A^T A$ is a 2×2 invertible matrix. When a least squares solution $\hat{x}$ is used to produce $A\hat{x}$ as an approximation to b, the distance from b to $A\hat{x}$ is called the **least-squares error** of this approximation.

Example 6.6.4. *Find a least-squares solution of the inconsistent system $Ax = b$ for*

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}$$

and

$$b = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$$

Then, determine the least-squares error in the least square solution of $Ax = b$.
Solution: *To use the normal equation, compute*

$$A^T A = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}$$

$$A^T b = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

Then the equation $A^T Ax = A^T b$ becomes

$$\begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

Solving this equation the way you prefer you should get

$$\hat{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Then

$$A\hat{x} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix}$$

Hence,

$$b - A\hat{x} = \begin{pmatrix} -2 \\ -4 \\ 8 \end{pmatrix}$$

and

$$||b - A\hat{x}|| = \sqrt{84}$$

The least square error is $\sqrt{84}$. For any x in $\mathbb{R}^2$, the distance between b and the vector Ax is at least $\sqrt{84}$.

The next example shows how to find a least-squares solution of $Ax = b$ when the columns of A are orthogonal. Such matrices are often used in linear regression problems.

Example 6.6.5. *Find a least-squares solutions of $Ax = b$ for*

$$A = \begin{pmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{pmatrix}$$

and

$$b = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix}$$

Solution: *Because the columns a_1 and a_2 of A are orthogonal, the orthogonal projection of b onto $Col(A)$ is given by*

$$\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b_1 \cdot a_2}{a_2 \cdot a_2} a_2 = \begin{pmatrix} -1 \\ 1 \\ 5 \\ 2 \\ 11/2 \end{pmatrix}, \ (\square)$$

Now that $\hat{b}$ is known, we can solve $A\hat{x} = \hat{b}$. But this is trivial, since we already know what weights to place on the columns of A to produce $\hat{b}$. It is clear from $(\square)$ than

$$\hat{x} = \begin{pmatrix} 2 \\ 1/2 \end{pmatrix}$$

In some cases, the normal equations for a least-squares problem can be illconditioned; that is, small errors in the calculations of the entries of $A^T A$ can sometimes cause relatively large errors in the solution $\hat{x}$. If the columns of A are linearly independent, the least-squares solution can often be computed more reliably through a QR factorizationof A.

Theorem 6.6.6. *Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A. Then, for each b in $\mathbb{R}^m$, the equation $Ax = b$ has a unique least-squares solution, given by*

$$\hat{x} = R^{-1} Q^T b$$

Proof. Let $\hat{x} = R^{-1} Q^T b$. Then

$$A\hat{x} = QR\hat{x} = QRR^{-1} Q^T b = QQ^T b$$

The column of Q form an orthogonal basis for $Col(A)$. Hence, $QQ^T b$ is the orthogonal projection $\hat{b}$ of b onto $Col(A)$, Then $A\hat{x} = \hat{b}$, which shows that $\hat{x}$ is a least-squares solution of $Ax = b$. The uniqueness of $\hat{x}$ is a least-squares solution of $Ax = b$. The uniqueness of $\hat{x}$ follows, from the previous theorem. $\square$

Example 6.6.7. *Find the least-squares solution of $Ax = b$ for*

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}, \ b = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}$$

Solution: *The QR factorization of A can be obtained as explained before:*

$$A = QR = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{pmatrix} \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

Then

$$Q^T b = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix}$$

The least-square solution $\hat{x}$ satisfies $Rx = Q^T b$; that is $\hat{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ such that

$$\begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix}$$

This equation is solved easily and yields $\hat{x} = \begin{pmatrix} 10 \\ -6 \\ 2 \end{pmatrix}$.

6.7 Applications to linear models

A common task in science an engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least-squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of $Ax = b$, we write $X\beta = y$ and refer to X as the **design matrix,** β as the **parameter vector** and y as the **observation vector**.

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), \cdots, (x_n, y_n)$ that, when graded, seem to lie close to a line. We want to determine the parameters β_0 and β_1, that make the line as "close" to the points as possible. .

Suppose β_0 and β_1 are fixed, and consider the line $y = \beta_0 + \beta_1 x$. Corresponding to each data point (x_j, y_j), there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. We call y_j the **observed value of** y and $\beta_0 + \beta_1 x_j$ the **predicted y-value** (determined by the line). The difference between an observed y-value and a predicted y-value is called a **residual**.

There are several ways to measure how "close the line is to the data. The usual choice (primarily because the mathematical calculations are simple) is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of square of the residual. The line is also call a **line of regression of y on x** because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called (linear) **regression coefficients**.

If the data point were on the line, the parameters β_0 and β_1 would satisfy the equations:

$$\beta_0 + \beta_1 x_i = y_i$$

where we have the predicted y-value on the left hand side and the observed y-value on the right side.

$$X\beta = y$$

where

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ 1 & x_n \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \ y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}$$

Of course, if the data points don't lie on a line, then there are no parameters β_0 and β_1, for which the predicted y-values in $X\beta$ equal the observed y values in y, and $X\beta = y$ has no solution. This is a least-squares problem, $Ax = b$ with different notation!

The square of the distance between the vectors $X\beta$ and y is precisely the sum of the squares of the residuals. The β that minimizes this sum also minimizes the distance between $X\beta$ and y. Computing the least squares solution of $X\beta = y$ is equivalent to finding the β that determines the least-squares line.

Example 6.7.1. *Find the equation $y = \beta_0 + \beta_1 x$ of the least-square line that best fits the data points $(2,1)$, $(5,2)$, $(7,3)$ and $(8,3)$.* **Solution:** *Use the x-coordinates of the data to build the design matrix X and the y-coordinates to build the observation vector y:*

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix}, \ y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}$$

For the least-squares solution of $X\beta = y$, obtain the normal equations (with the new notation):

$$X^T X\beta = X^T y$$

That is compute

$$X^T X = \begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 9 \\ 57 \end{pmatrix}$$

The normal equation is

$$\begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 9 \\ 57 \end{pmatrix}$$

Hence

$$\beta = \begin{pmatrix} 2/7 \\ 5/14 \end{pmatrix}$$

Thus the least-squares line has the equation

$$y = 2/7 + 5/14x$$

A common practice before computing a least squares line is to compute the average $\bar{x}$ of the original x-values an form a new variable $x^* = x - \bar{x}$. The new x-data are said to be **mean-deviation form**. In this case, the two columns of the design matrix will be orthogonal. Solution of the normal equations is simplified.
In some application, it is necessary to fit data points with something other than a straight line. In the example that follow, the matrix equation is still $X\beta = y$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** ϵ, defined by $\epsilon = y - X\beta$, and write

$$y = X\beta + \epsilon$$

Any equation of this form is referred to as a **linear model**. Once X and y are determined, the goal is to minimize the length of ϵ, which amounts to finding a least-squares solution of $X\beta = y$. In each case, the least-squares solution $\hat{\beta}$ is a solution of the normal equations

$$X^T X\beta = X^T y$$

When the data $(x_1, y_1), \cdots, (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y. The next example shows how to fit the data by curves that have the general form

$$y = \beta_0 f_0(x) + \cdots + \beta_k f_k(x)$$

where $f_0, \cdots, f_k$ are known functions and $\beta_0, \cdots, \beta_k$ are parameters that must be determined. As we will see, the previous equation describes a linear model because it is linear in the unknown parameters. For a particular value of x, this equation gives a predicted or "fitted" value of y. The difference between the observed value and the predicted value is the residual. The parameters $\beta_0, \cdots, \beta_k$ must be determined so as minimize the sum of the squares of the residuals.

Example 6.7.2. *Suppose that data points $(x_1, y_1), \cdots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at opens upward. In ecology, a parabolic curve that opens downward is used to model the* *net primary production of nutrients in a plant, as a function of the surface area of the foliage. Suppose we wish to approximate that the data by an equation of the form*

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Describe the linear model that produces a "least-squares fit" of the data by the previous equation.

Solution: *The previous equation describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the data point (x_i, y_i) satisfies an equation of the form $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$. It is a simple matter to write this system of equation in the form $y = X\beta + \epsilon$. To find X, inspect the first few row of the system and look at the pattern.*

$$y = X\beta + \epsilon$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_1 & x_1^2 \\ . & . & . \\ . & . & . \\ 1 & x_n & x_n^2 \end{pmatrix}, \; \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \; \epsilon = \begin{pmatrix} \epsilon_1 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

Suppose an experiment involves two independent variables- say, u and v, and one dependent variable, y. A simple equation for predicting y from u and v has the form

$$y = \beta_0 + \beta_1 u + \beta_2 v$$

A more general prediction equation might have the form

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2$$

This equation is used in geology, for instance, to model erosion surfaces glacial cirques, soil pH and other quantities. In such cases, the least-squares fit is called a trend surface. The two previous equation both lead to a linear model because they are linear in the unknown parameters (even though u and v and multiplied). In general, a linear model will arise whenever y is to be predicted by an equation of the form

$$y = \beta_0 f_0(u, v) + \cdots + \beta_k f_k(u, v)$$

with $f_0, \cdots, f_k$ any sort of known functions and $\beta_0, \cdots, \beta_k$ unknown weights.

6.8 Inner product spaces

Notions of length, distance, and orthogonality are often important in applications involving a vector space. For $\mathbb{R}^n$, these concepts were based on the properties of the inner product. For other spaces, we need analogues of the inner product with the same properties.

Definition 6.8.1. *An* **inner product** *on a vector space V is a function that , to each pair of vectors u and v in V, associates a real number $< u,v >$ and satisfies the following axioms, for all u,v,w in V and all scalar c:*

1. $< u,v >=< v,u >$
2. $< u+v,w >=< u,w > + < v,w >$
3. $< cu,v >= c < u,v >$
4. $< u,u >\geqslant 0$ *and* $< u,u >= 0$ *if and only if* $u = 0$.

A vector space $\mathbb{R}^n$ with the standard inner product is an inner product space, and nearly everything discussed for $\mathbb{R}^n$ carries over to any inner product spaces.

Definition 6.8.2. *Let V be an inner product space, with the inner product denoted by $< u,v >$. Just as in $\mathbb{R}^n$, we define the* **length** *or* **norm** *of a vector v to be the scalar*

$$||v|| = \sqrt{< v,v >}$$

Equivalently, $||v||^2 =< v,v >$. (This definition makes sense because $< v,v >\geqslant 0$, but the definition does not say that $< v,v >$ is a sum of squares, because v need not be an element of $\mathbb{R}^n$.)
A **unit vector** *is one whose length is 1. The* **distance between u and v** *is $||u-v||$. Vectors u and v are orthogonal if $< u,v >= 0$.*

Example 6.8.3. *Fix any two positive numbers say 4 and 5 and for vectors $u = (u_1,u_2)$ and $v = (v_1,v_2)$ in $\mathbb{R}^2$, set*

$$< u,v >= 4u_1v_1 + 5u_2v_2$$

Show that this defines an inner product.
Solution: *Let $u = (u_1,u_2), v = (v_1,v_2)$ and $w = (w_1,w_2) \in \mathbb{R}^2$,*

$$< u,v >= 4u_1v_1 + 5u_2v_2 = 4v_1u_1 + 5v_2u_2 =< v,u >$$

$$< u+v,w >= 4(u_1+v_1)w_1+5(u_2+v_2)w_2 = 4u_1w_1+5u_2w_2+4v_1w_1+5v_2w_2 =< u,w > + < v,w >$$

$$< cu,v >= 4(cu_1)v_1 + 5(cu_2)v_2 = c(4u_1v_1 + 5u_2v_2) = c < u,v >$$

$$< u,u >= 4u_1^2 + 5v_1^2 \geqslant 0$$

and thus $4u_1^2 + 5u_2^2 = 0$ only if $u_1 = u_2 = 0$, that is, if $u = 0$. Also, $< 0,0 >= 0$, So $<,>$ define an inner product.

Inner products similar to the one of the example can be defined on $\mathbb{R}^n$. They arise naturally in connection with "weighted least squares" problems, in which weights are assigned to the various entries in the sum for the inner product in such way that more importance is given to the more reliable measurements.
From now on, when an inner product space involves polynomials or others functions, it is important to remember that each function is a vector when it is treated as an element of a vector space.

Example 6.8.4. *Let $t_0, \cdots t_n$ be distinct real numbers For p and q in $\mathbb{P}_n$ define*

$$< p, q >= p(t_0)q(t_0) + \cdots + p(t_n)q(t_n).$$

Prove that this define an inner product on $\mathbb{P}_n$ and compute the lengths of the vectors $p(t) = 12t^2$ and $q(t) = 2t - 1$.
Inner product Axioms $1 - 3$ are readily checked. For axiom 4, note that

$$< p, p >= [p(t_0)]^2 + [p(t_1)]^2 + \cdots + [p(t_n)]^2 \geqslant 0$$

Also $< \mathbf{0}, \mathbf{0} >= 0$. The boldface zero here denotes the zero polynomial, the zero vector in $\mathbb{P}_n$. If $< p, p >= 0$, then p must vanish at $n+1$ points: $t_0, \cdots, t_n$. This is possible if and only if p is the zero polynomial, because the degree of p is less than $n+1$. Thus $<, >$ defines an inner product on $\mathbb{P}_n$.

$$||p||^2 =< p, p >= [p(0)]^2 + [p(1/2)]^2 + [p(1)]^2 = 153$$

So that

$$||p|| = \sqrt{153}$$

The existence of orthogonal bases for finite-dimensional subspaces of an inner product space can be established by the Gram-Schmith process, just as in $\mathbb{R}^n$. Certain orthogonal bases that arise frequently in applications can be constructed by this process.
The orthogonal projection of a vector onto a subspace W with an orthogonal basis can be constructed as usual. The projection does not depend on the choice of orthogonal basis and it has the properties described in the Orthogonal Decomposition Theorem and the Best approximation Theorem.
A common problem in applied mathematics involves a vector space V whose elements are functions. The problem is to approximate a function f in V by a function g from a specified subspace W of V. The "closeness" of the approximation of f depends on the way $||f - g||$ is defined. We will consider only the case in which the distance between f and g is determined by an inner product. In this case, the best approximation of f by functions in W is the orthogonal projection of f onto the subspace W.

Example 6.8.5. *Let $C[a, b]$ the set of all continuous function on an interval $a \leqslant t \leqslant b$, with an inner product that we will describe.*
For f, g in $C[a, b]$, set

$$< f, g >= \int_a^b f(t)g(t)dt$$

Show that this defines an inner product on $C[a, b]$.
Solution: *Inner product Axiom $1 - 3$ follow from elementary properties of definite integrals. For Axioms 4, observe that*

$$< f, f >= \int_a^b [f(t)]^2 dt \geqslant 0$$

The function $[f(t)]^2$ is continuous and nonnegative on $[a, b]$. If the definite integral of $[f(t)]^2$ is zero, then $[f(t)]^2$ must be identically zero on $[a, b]$, by a theorem in advanced calculus, in which case f is the zero function. Thus $< f, f >= 0$ implies that f is the zero function on $[a, b]$. So $<, >$ defines an inner product on $C[a, b]$.

Example 6.8.6. *Let V be the space $C[0, 1]$ with the previous inner product, and W be the subspace spanned by the polynomials $p_1(t) = 1$, $p_2(t) = 2t - 1$ and $p_3(t) = 12t^2$. Use the Gram-Schmith process to find an orthogonal basis for W.*
Solution: *Let $q_1 = p_1$, and compute*

$$< p_2, q_1 >= \int_0^1 (2t-1)1dt = 0$$

. So, p_2 is already orthogonal to q_1, and we can take $q_2 = p_2$. For the projection of p_2 onto $W_2 = Span\{q_1, q_2\}$, compute

$$< p_3, q_1 >= \int_0^1 12t^2 \cdot 1dt = 4$$

$$< q_1, q_1 >= \int_0^1 1 \cdot 1dt = 1$$

$$< p_3, q_2 >= \int_0^1 12t^2(2t-1)dt = 2$$

$$< q_2, q_2 >= \int_0^1 (2t-1)^2 dt = 1/3$$

Then

$$proj_{W_2} p_3 = \frac{< p_3, q_1 >}{< q_1, q_1 >} q_1 + \frac{< p_3, q_2 >}{< q_2, q_2 >} q_2 = 4q_1 + 6q_2$$

and

$$q_3 = p_3 - proj_{W_2}(p_3) = p_3 - 4q_1 - 6q_2$$

As a function $q_3(t) = 12t^2 - 4 - 6(2t-1) = 12t^2 - 12t + 2$. The orthogonal basis for the subspace W is $\{q_1, q_2, q_3\}$.

Given a vector v in the product space V and given a finite-dimensional subspaces W, we may apply the Pythagorean Theorem to the orthogonal decomposition of v with respect to W and obtain

$$||v||^2 = ||proj_W(v)||^2 + ||v - proj_w(v)||^2$$

In particular, this shows that the norm of the projection of v onto W does not exceed the norm of v itself. This simple observation leads to the following important inequality.

Theorem 6.8.7 (The Cauchy-Schwartz Inequality)**.** *For all u, v in V,*

$$| < u, v > | \leqslant ||u|| \; ||v||$$

Proof. If $u = 0$, then both sides of the inequality are zero, and hence the inequality is true in this case. If $u \neq 0$, let W be the subspace spanned by u. Recall that $||cu|| = |c| \cdot ||u||$, for any scalar c. Thus,

$$||proj_W(v)|| = ||\frac{< v, u >}{< u, u >}u|| = \frac{| < v, u > |}{| < u, u > |}||u|| = \frac{| < u.v > |}{||u||}$$

Since $||proj_W(v)|| \leqslant ||v||$, we have

$$\frac{| < u.v > |}{||u||} \leqslant ||v||$$

which gives the Cauchy-Schwartz inequality. □

The Cauchy-Schwartz inequality is useful in many branches of mathematics. Our main need for this inequality here is to another fundamental inequality involving norms of vectors.

Theorem 6.8.8 (The triangle inequality)**.** *For all u, v in V.*

$$||u + v|| \leqslant ||u|| + ||v||$$

Proof.

$$||u+v||^2 =< u+v, u+v >=< u, u > +2 < u, v > + < v, v >\leqslant ||u||^2+2||u||||v||+||v||^2 = (|u||+||v||)^2$$

The triangle inequality follows immediately by taking square roots of both sides. □

Chapter 7

Symmetric matrices and quadratic forms

7.1 Diagonalization of symmetric matrices

Definition 7.1.1. *A* **symmetric matrix** *is a matrix A such that $A^T = A$, Such a matrix is necessarily square. Its main diagonal entries are arbitrary, but other entries occur in pairs, on opposite side of the diagonal.*

Example 7.1.2. *The following matrix is symmetric:*

$$A = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

The following matrix is not symmetric:

$$B = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

Theorem 7.1.3. *If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.*

Proof. Let v_1 and v_2 be eigenvectors that correspond to distinct eigenvalues, say, λ_1 and λ_2. To show that $v_1 \cdot v_2 = 0$, compute

$$\begin{array}{rcl} \lambda_1 v_1 \cdot v_2 & = & (\lambda_1 v_1)^T v_2 = (Av_1)^T v_2 \\ & = & (v_1^T A^T) v_2 = v_1^T (A v_2) \text{ since } A^T = A \\ & = & v_1^T (\lambda_2 v_2) \\ & = & v_1^T (\lambda_2 v_2) \\ & = & \lambda_2 v_1^T v_2 = \lambda_2 v_1 \cdot v_2 \end{array}$$

Hence, $(\lambda_1 - \lambda_2) v_1 \cdot v_2 = 0$. But, $\lambda_1 - \lambda_2 \neq 0$, so $v_1 \cdot v_2 = 0$. $\square$

Definition 7.1.4. *An $n \times n$ matrix A is said to be orthogonally diagonalizable if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that*

$$A = PDP^T = PDP^{-1}$$

Such a diagonalizable matrix requires n linearly independent and orthonormal eigenvectors.

When this is possible? If A is orthogonally diagonalizable, then

$$A^T = (PDP^T)^T = P^{TT}D^TP^T = PDP^T = A$$

Thus A is symmetric! The theorem below shows that conversely every symmetric matrix is orthogonally diagonalizable.The proof is much harder and is omitted; the main idea for the proof will be given in the proof of the next theorem.

Theorem 7.1.5. *An $n \times n$ matrix A is orthogonally diagonalizable if and only if A is symmetric matrix.*

Example 7.1.6. *Orthogonally diagonalizable the matrix* $A = \begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$*, where the characteristic equation is*

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

Solution: *The usual calculations produce bases for the eigenspaces:*

$$\lambda = 7: \; v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix}, \; \lambda = -2: \; v_3 = \begin{pmatrix} -1 \\ -1/2 \\ 1 \end{pmatrix}$$

Although v_1 and v_2 are linearly independent, they are not orthogonal. We have proven the projection of v_2 onto v_1 is $\frac{v_2 \cdot v_1}{v_1 \cdot v_1} v_1$, and the component of v_2 orthogonal to v_1 is

$$z_2 = v_2 - \frac{v_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix} - \frac{-1/2}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/4 \\ 1 \\ 1/4 \end{pmatrix}$$

Then $\{v_1, z_2\}$ is an orthogonal set in the eigenspace for $\lambda = 7$. Note that z_2 is a linear combination of the eigenvectors v_1 and v_2, so z_2 is in the eigenspace. This construction of z_2 is just Gram-Schmidt process). Since the eigenspace is two dimensional (with basis v_1, v_2), the orthogonal set $\{v_1, z_2\}$ is an orthogonal basis for the eigenspace, by the Basis Theorem.

Normalize v_1 and z_2 to obtain the following orthonormal basis for the eigenspace for $\lambda = 7$:

$$u_1 = \begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \; u_2 = \begin{pmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ \sqrt{18} \end{pmatrix}$$

An orthonormal basis for the eigenspace for $\lambda = -2$ is

$$u_3 = \frac{1}{||2v_3||} 2v_3 = 1/3 \begin{pmatrix} -2 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}$$

By the first theorem of this section, u_3 is orthogonal to the other eigenvectors u_1 and u_2. Hence $\{u_1, u_2, u_3\}$ is an orthonormal set . Let

$$P = [u_1 \; u_2 \; u_3] = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{pmatrix}, \; D = \begin{pmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

Then P orthogonally diagonalizes A, and $A = PDP^T = PDP^{-1}$.

Definition 7.1.7. *The set of eigenvalues of a matrix A is sometimes called the spectrum of A, and the following description of the eigenvalues is called a* **spectral theorem***.*

Theorem 7.1.8 (The spectral theorem for symmetric matrices)**.** *A $n \times n$ symmetric matrix A has the following properties:*

1. *A has n real eigenvalues, counting multiplicities.*
2. *The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.*
3. *The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.*
4. *A is orthogonally diagonalizable.*

Proof. 1. follows from the fact that other the complex number any polynomial factor though linear factor and that if we have a complex eigenvalue for a symmetric matrix then the real part and the imaginary part are also eigenvalues (exercise).
2. follows easily when 4. is proven.
3. follows from the first theorem of this section.
4. can be deduced from Schur decomposition and prove that in the upper triangular matrix in the Schur decomposition is actually a diagonal matrix, when A is symmetric. (Exercise.) □

Suppose $A = PDP^{-1}$, where the columns of P are orthonormal eigenvectors $u_1, \cdots, u_n$ of A and the corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$ are in the diagonal matrix D. Then, since $P^{-1} = P^T$,

$$A = PDP^T = [u_1, \cdot, u_n] \begin{pmatrix} \lambda_1 & \cdot & \cdot & 0 \\ & \cdot & \cdot & \\ & \cdot & \cdot & \\ 0 & \cdot & \cdot & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix}$$

Using the column-row expansion of a product, we can write

$$A = \lambda_1 u_1 u_1^T + \cdots + \lambda u_n u_n^T$$

This representation of A is called a **spectral decomposition** of A because it breaks up A into pieces determined by the spectrum (eigenvalues) of A. Each term is an $n \times n$ matrix of rank 1. For example, every columns of $\lambda_1 u_1 u_1^T$ is a multiple of u_1. Furthermore, each matrix $u_j u_j^T$ is a **projection matrix** in the sense that for each x in $\mathbb{R}^n$, the vector $(u_j u_j^T)x$ is the orthogonal projection of x onto the subspace spanned by u_j.

Example 7.1.9. *Construct a spectral decomposition of the matrix A that has the orthogonal diagonalization.*

$$A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}$$

Solution: *Denote the columns of P by u_1 and u_2. Then the spectral decomposition is*

$$A = 8u_1u_1^T + 3u_2u_2^T$$

You can also verify the decomposition of A if you want to.

7.2 Quadratic forms

Definition 7.2.1. *A* **quadratic form** *on $\mathbb{R}^n$ is a function Q defined on $\mathbb{R}^n$ can be computed by an expression of the form $Q(x) = x^TAx$, where A is an $n \times n$ symmetric matrix. The matrix A is called the* **matrix of the quadratic form**.

The simplest example of a nonzero quadratic form is $Q(x) = x^TIx = ||x||^2$.

Example 7.2.2. *Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Compute $x^TAx$for the following matrices:*

1. $A = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}$
2. $A = \begin{pmatrix} 3 & -2 \\ -2 & 7 \end{pmatrix}$

Solution:

1.
$$x^TAx = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 4x_1 \\ 3x_2 \end{pmatrix} = 4x_1^2 + 3x_2^2$$

2.
$$x^TAx = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ -2 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{pmatrix} = 3x_1^2 - 4x_1x_2 + 7x_2^2$$

Observe that there is a cross product x_1x_2 when A is not diagonal while it is not there when it is.

Example 7.2.3. *For x in $\mathbb{R}^3$, let $Q(x) = 5x_1^2 + 3x_2^2 + 2x_3^2 - x_1x_2 + 8x_2x_3$. Write this quadratic form as x^TAx. Compute $Q(x_0)$ with $x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.*

Solution: *The coefficient of x_1^2, x_2^2, x_3^2 go the diagonal of A. To make A symmetric,* *the coefficient of $x_i x_j$ for $i \neq j$ must be split evenly between the (i,j) and (j,i) entries in A. The coefficient of x_1x_3 is 0. It is readily checked that*

$$Q(x) = x^TAx = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

and

$$Q(x_0) = 5 \cdot 1^2 + 3 \cdot 0^2 + 2 \cdot 0^2 - 1 \cdot 0 + 80 \cdot 0 = 5$$

If x represents a variable vector in$\mathbb{R}^n$, then a **change of variable** is an equation of the form

$$x = Py, \text{ or equivalently, } y = P^{-1}x$$

where P is an invertible matrix and y is a new variable vector in $\mathbb{R}^n$. Hence y is the coordinate vector of x relative to the basis of $\mathbb{R}^n$ determined by the columns of P.
If the change of variable is made in a quadratic form x^TAx, then

$$x^TAx = (Py)^TA(Py) = y^TP^TAPy = y^T(P^TAP)y$$

and the new matrix of the quadratic form is P^TAP. Since A is symmetric, so we know that there is an orthogonal matrix P such that P^TAP is a diagonal matrix D and the quadratic form becomes y^TDy.

Example 7.2.4. *Make a change of variable that transform the quadratic form defined by*

$$A = \begin{pmatrix} 1 & -4 \\ -4 & -5 \end{pmatrix}$$

into a quadratic form $Q(y) = y^TDy$ with no cross product.
Solution: *The first step is to orthogonally diagonalize A. Its eigenvalues turn out to be $\lambda = 3$ and $\lambda = -7$. Associated unit eigenvectors are*

$$\lambda = 3 : \begin{pmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{pmatrix}, \ \lambda = -7 : \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix}$$

These vectors are automatically orthogonal (because they correspond to distinct eigenvalues) and so provide an orthonormal basis for $\mathbb{R}^2$. Let

$$P = \begin{pmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{pmatrix}, \ D = \begin{pmatrix} 3 & 0 \\ 0 & -7 \end{pmatrix}$$

Then $A = PDP^{-1}$ and $D = P^{-1}AP = P^TAP$, as pointed out earlier. A suitable change of variable is

$$x = Py, \text{ where } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

and

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Then

$$Q(x) = x^TAx = (Py)^TA(Py) = y^TP^TAPy = y^TDy = 3y_1^2 - 7y_2^2$$

Theorem 7.2.5 (The principal axes theorem). *Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $x = Py$, that transforms the quadratic form $x^T Ax$ into a quadratic form $y^T Dy$ with no cross-product term.*

We can prove the theorem easily just as a direct generalization of what is done in the previous example. The column of P in the theorem are called the **principal axes** of the quadratic form $x^T Ax$. The vector y is the coordinate vector of x relative to the orthogonal basis of $\mathbb{R}^n$ given by these principal axes.

Suppose $Q(x) = x^T Ax$, where A is an invertible 2×2 symmetric matrix, and let c be a constant. If can be shown that the set of all x in $\mathbb{R}^2$ that satisfy

$$x^T Ax = c$$

either corresponds to an ellipse (or circle), a hyperbola, two intersecting lines, or a single point, or contains no points at all. If A is a diagonal matrix, the graph is in standard position. If A is not a diagonal matrix, the graph is rotated out of the standard position. Finding, the principal axes (determined by the eigenvector of A) amount to finding a new coordinate system with respect to which the graph is in standard position.

Definition 7.2.6. *A quadratic form Q is:*

1. **positive definite** *if $Q(x) > 0$ for all $x \neq 0$,*
2. **negative definite** *if $Q(x) < 0$ for all $x \neq 0$,*
3. **indefinite** *if $Q(x)$ assumes both positive and negative values.*

Also Q is said to be **positive semidefinite** if $Q(x) \geqslant 0$ for all x, and to be **negative semidefinite** if $Q(x) \leqslant 0$ for all x. The following theorem characterizes some quadratic forms in terms of eigenvalues.

Theorem 7.2.7 (Quadratic forms and eigenvalues). *Let A be an $n \times n$ symmetric matrix. Then a quadratic form $x^T Ax$ is*

1. *positive definite if and only if the eigenvalues of A are all positive.*
2. *negative definite if and only if the eigenvalues of A are all negative, or*
3. *indefinite if and only if A has both positive and negative eigenvalues.*

Proof. By the principal axes theorem, there exists an orthogonal change of variable $x = Py$ such that

$$Q(x) = x^T Ax = y^T Dy = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

where $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of A. Since P is invertible, there is a one-to-one correspondence between all nonzero x and all nonzero y. Thus the values of $Q(x)$ for $x \neq 0$ coincide with the values of the expression on the right side of the equality, which is obviously controlled by the signs of eigenvalues $\lambda_1, \cdots, \lambda_n$, in three ways described in the theorem. □

Example 7.2.8. *Is* $Q(x) = 3x_1^2 + 2x_2^2 + x_3^2 + 4x_1x_2 + 4x_2x_3$ *positive definite?*
Solution: *Because of all the plus signs, this form "looks" positive definite. But the matrix of the form is*

$$A = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 1 \end{pmatrix}$$

and the eigenvalues of A *turn out to be* 5, 2 *and* -1. *So* Q *is an indefinite quadratic form, not positive definite.*

The classification of a quadratic form is often carried over to the matrix of the form. Thus a **positive definite matrix** A is a symmetric matrix for which the quadratic form x^TAx is positive definite. Other terms, such as **positive semidefinite matrix**, are defined analogously.

7.3 Singular value decomposition

The diagonalization theorems play a part in many interesting applications. Unfortunately as we know, not all matrices can be factorized as $A = PDP^{-1}$ with D diagonal. However a factorization $A = QDP^{-1}$ is possible for any $m \times n$ matrix A! A special factorization of this type, called singular value decomposition, is one of the most useful matrix factorizations in applied linear algebra.
The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices: The absolute values of the eigenvalues of a symmetric matrix A mesure the amounts that A stretches or shrinks certain vectors (the eigeinvectors). If $Ax = \lambda x$ and $||x|| = 1$, then

$$||Ax|| = ||\lambda x|| = |\lambda|||x|| = |\lambda|$$

If λ_1 is the eigenvalue with the greatest magnitude, then a corresponding unit vector v_1 identifies a direction in which the stretching effect of A is greatest. That is, the length of Ax is maximized when $x = v_1$ and $||Av_1|| = |\lambda_1|$. This description of v_1 and $|\lambda_1|$. This description of v_1 and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.
Let A be an $m \times n$ matrix. Then A^TA is symmetric and can be orthogonally diagonalized. Let $\{v_1, \cdots, v_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA. Then for $1 \leqslant i \leqslant n$, let $\lambda_1, \cdots, \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \leqslant i \leqslant n$,

$$||Av_i||^2 = (Av_i)^TAv_i = v_i^TA^TAv_i = v_i^T(\lambda_iv_i) = \lambda_i$$

since v_i is an eigenvector of A^TA and v_i is a unit vector. So the **singular values of** A^TA are all nonnegative. By renumbering, if necessary, we may assume that the eigeinvalues are arranged in decreasing order. The singular values of A are the square roots of the eigenvalues of A^TA denoted by $\sigma_1 m \cdots, \sigma_n$ and they are in decreasing order. That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \leqslant i \leqslant n$. By equation above, the singular values of A are the lengths of the vectors $Av_1, \cdots, Av_n$.

Theorem 7.3.1. *Suppose $\{v_1, \cdots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvector of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ and suppose A has r nonzero singular values. Then $\{Av_1, \cdots, Av_r\}$ is an orthogonal basis for $Col(A)$ and $rank(A) = r$*

Proof. Because v_i and $\lambda_j v_j$ are orthogonal for $i \neq j$

$$(Av_i)^T(Av_j) = v_i^T A^T A v_j = v_i^T(\lambda_j v_j) = 0$$

Thus $\{Av_1, \cdots, Av_n\}$ is an orthogonal set. Furthemore, since the lengths of the vectors $Av_1, \cdots, Av_n$ are singular values of A and since there are nonzero singular values, $Av_i \neq 0$ if and only if $1 \leqslant i \leqslant r$. So $Av_1, \cdot, Av_r$ are linearly independent vectors and they are in $Col(A)$. Finally, for any y in $Col(A)$, say $y = Ax$ we can write $x = c_1v_1 + \cdots c_nv_n$ and

$$y = Ax = c_1Av_1 + \cdots c_rAv_r + c_{r+1}Av_{r+1} + \cdots c_nAv_n = c_1Av_1 + \cdots + c_rAv_r + 0 \cdots + 0$$

Thus y is in $Span\{Av_1, \cdots, Av_r\}$, which shows that $\{Av_1, \cdots, Av_r\}$ is an (orthogonal) basis for $Col(A)$. Hence $rank(A) = dim(Col(A)) = r$. □

The decomposition of A involves an $m \times n$ "diagonal" matrix Σ of the form

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n. (If r equals m or n or both, some or all of the zero matrices do not appear.)

Theorem 7.3.2 (The singular value decomposition)**.** *Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix Σ for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that*

$$A = U\Sigma V^T$$

Any factorization $A = U\Sigma V^T$ with U and V orthogonal, σ as before and positive diagonal entries in D, is called **singular value decomposition** (or **SVD**) of A. The matrices U and V are not uniquely determined by A , but the diagonal entries of Σ are necessarily the singular values of A. The columns of U in such a decomposition are called **left singular vectors**. The columns of V are called **right singular vectors** of A.

Proof. Suppose $\{v_1, \cdots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvector of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, so that $\{Av_1, \cdots, Av_r\}$ is an orthogonal basis for $Col(A)$. Normalize each Av_i to obtain an orthonormal basis $\{u_1, \cdots, u_r\}$, where

$$u_i = \frac{1}{||Av_i||}Av_i = \frac{1}{\sigma_i}Av_i$$

and

$$Av_i = \sigma_i u_i \ (1 \leqslant i \leqslant r)$$

Now extend $\{u_1, \cdots, u_r\}$ to an orthonormal basis $\{u_1, \cdot, u_m\}$ of $\mathbb{R}^m$, and let

$$U = [u_1, u_2, \cdots, u_m] \text{ and } V = [v_1, \cdots, v_n]$$

By construction, U and V are orthogonal matrices. Also,

$$AV = [Av_1, \cdots, Av_r, 0, \cdots, 0] = [\sigma_1 u_1, \cdots, \sigma_r u_r, 0, \cdots, 0]$$

Let D be the diagonal matrix with diagonal entries $\sigma_1, \cdots, \sigma_r$ and let Σ be as above. Then

$$U\Sigma = [u_1, \cdots, u_m] \left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \cdot \quad \cdot & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & 0 \end{array} \right) = [\sigma_1 u_1, \cdots, \sigma_r u_r, 0 \cdots, 0] = AV$$

Since V is an orthogonal matrix,

$$U\Sigma V^T = AVV^T = A$$

$\square$

Example 7.3.3. *Find a singular value decomposition of* $A = \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}$.

Solution: *First, compute* $A^TA = \begin{pmatrix} 9 & -9 \\ -9 & 9 \end{pmatrix}$. *The eigenvalues of* A^TA *are* 18 *and* 0 *with corresponding eigenvectors*

$$v_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}, v_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

These unit vectors form the columns of V*:*

$$V = [v_1, v_2] = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

The singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ *and* $\sigma_2 = 0$. *Since there is only one nonzero singular value, the "matrix"* D *may be written as a single number. That is* $D = 3\sqrt{2}$. *The matrix* Σ *is the same size as* A*, with* D *in its upper left corner:*

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

To construct U*, first construct*

$$Av_1 = \begin{pmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{pmatrix}$$

and

$$Av_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

As a check on the calculations, verify that $||Av_1|| = \sigma_1 = 3\sqrt{2}$. *Of course,* $Av_2 = 0$ *because* $||Av_2|| = \sigma_2 = 0$. *The only column found for* U *so far is*

$$u_1 = \frac{1}{3\sqrt{2}} Av_1 = \begin{pmatrix} 1/3 \\ -2/3 \\ 2/3 \end{pmatrix}$$

The other columns of U *are found by extending the set* $\{u_1\}$ *to an orthonormal basis for* $\mathbb{R}^3$. *In this case we need two orthogonal unit vectors* u_2 *and* u_3 *that are orthogonal to* u_1. *Each vector must satisfy* $u_1^T x = 0$ *which is equivalent to the equation* $x_1 - 2x_2 + 2x_3 = 0$. *A basis for the solution set of this equation is*

$$w_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \; w_2 = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

(Check that w_1 *and* w_2 *are each orthogonal to* u_1*.) Apply the Gram-Schmidt process (with normalization) to* $\{w_1, w_2\}$, *and obtain*

$$u_2 = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{pmatrix}, u_3 = \begin{pmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{pmatrix}$$

Finally, set $U = [u_1, u_2, u_3]$, *take* Σ *and* V^T *from above, and write*

$$A = \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = [u_1, u_2, u_3] \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} [v_1, v_2]$$

The SVD is often used to estimate the rank of a matrix. But we have also others application.

Example 7.3.4. *Given an SVD for an* $m \times n$ *matrix* A, *let* $u_1, \cdot, u_m$ *be the left singular vectors,* $v_1, \cdots, v_n$ *the right singular vectors, and let* r *be the rank of* A. *We have seen that* $\{u_1, \cdots, u_n\}$ *is an orthonormal basis for* $Col(A)$. *Recall that* $(Col(A))^\perp = Nul(A^T)$. *Hence* $\{u_{r+1}, \cdots, u_m\}$ *is an orthonormal basis for* $Nul(A^T)$. *Since* $||Av_i||\sigma_i$ *for* $1 \leqslant in$, *and* σ_i *is* 0 *if and only if* $i > r$, *the vectors* $v_{r+1}, \cdots, v_n$ *span a subspace of* $Nul(A)$ *of dimension* $n - r$. *By the Rank theorem,* $dim(Nul(A)) = n - rank(A)$. *It follows that* $\{v_{r+1}, \cdots, v_n\}$ *is an orthonormal basis for* $Nul(A)$ *by the Basis theorem.*
The orthogonal complement of $Nul(A^T)$ *is* $Col(A)$. *Interchanging* A *and* A^T, *note that* $(Nul(A))^\perp = Col(A^T) = Row(A)$. *Hence,* $\{v_1, \cdots, v_r\}$ *is an orthonormal basis for* $Row(A)$.
Explicit orthonormal bases for the four fundamental spaces determined by A *are useful in some calculations particularly in constrained optimization problems.*

The four fundamental subspace and the concept of singular values provides the final statements of the invertible matrix theorem. (Recall that statements about A^T have been omitted from the nearly doubling the number of statements.

Theorem 7.3.5 (The Invertible Matrix Theorem (concluded)). *Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement A is an invertible matrix*

1. $(Col(A))^{\perp} = \{0\}$
2. $(Nul(A))^{\perp} = \mathbb{R}^n$
3. $Row(A) = \mathbb{R}^n$
4. *A has n nonzero singular values.*