

Linear Algebra 1

Alp Ozen

Fall 2019 - **EPFL**

Contents

1	Linear equations	2
1.1	Basics	2
1.2	Solving a linear system	2
1.2.1	Basics	2
1.2.2	Reducing a system to reduced row echelon form	4
1.2.3	High school basics	4
2	Introducing vector spaces	5
2.1	Basics	5
2.2	Linear combinations,span and matrix theorems	5
2.3	Homogeneous system of equations	6
2.4	Linear Independence	6
2.5	Lec 05 notes	7
2.5.1	Matrix transformations	7
2.6	Linear transformations	7
3	Matrix algebra	8
3.1	Basics	8
3.2	Computing inverses	9
3.3	Every matrix is the product of elementary matrices with the identity	9
3.4	Characterizing inverse matrices	9
4	Determinants	10
4.1	Basics	10
5	More on vector spaces	12
5.1	Basics	12
5.2	Some problems from week 6	13
5.3	More theorems	13
5.4	Theorems of Rank, Col(A), Row(A) and Null(A)	15
5.5	Change of basis	15
5.6	Eigenvectors and eigenvalues	16
5.7	Orthogonality and inner products	19
6	Factorizations and applications	22

1 Linear equations

1.1 Basics

A linear equation is any equation of form: $a_1x_1 + \dots + a_nx_n = b$ where the a are 'scalars' that belong to a field and the x belong to the vector set.

A *system of linear equations* is simply a collection of linear equations. The solution of this system, if there is any, is an ordered list (s_1, \dots, s_n) where each s_i is the value of each x_i .

A system consisting of 2 unknowns and two linear equations is generally the intersection of two lines on a cartesian plane. Note that the lines may be parallel or even colinear.

In general, a system will have:

- no solution
- unique solution
- infinite solutions

We may choose to represent a system of linear equations as an *augmented matrix*. A matrix is called $n \times m$ if it is of form:

$$\begin{bmatrix} a_{11} & a_{12} & \dots \\ a_{21} & a_{22} & \dots \end{bmatrix}$$

1.2 Solving a linear system

1.2.1 Basics

When solving a system, our goal is to replace each linear equation with an equivalent set (one that has the same solution) and obtain single linear equations which are trivial to solve.

In solving a system, we use the elementary row operations which are:

- Interchange two rows (or columns).
- Multiply each element in a row (or column) by a non-zero number.
- Multiply a row (or column) by a non-zero number and add the result to another row (or column).

Our goal is to transform our matrix into echelon or row reduced echelon form. A matrix is in echelon form if it looks like this:

$$\begin{bmatrix} \blacktriangledown & \blacktriangle & \blacktriangle \\ 0 & \blacktriangledown & \blacktriangle \\ 0 & 0 & \blacktriangledown \end{bmatrix}$$

Here, each \blacktriangledown and \blacktriangle may take on any value from the set the vector space is defined on.

To obtain this form, we first arrange our matrix into a form where the row with the least amount of trailing zeros is placed uptop. Then we ensure the *pivot position* (meaning first non-zero entry) has only 0 in its own column. When done, we move on the second row, find the pivot position and repeat. We repeat this process for all rows.

If we have a system where the number of unknowns exceeds the number of equations, we obtain a parametric solution. Consider this:

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Which means:

$$x_1 - 5x_3 = 1$$

$$x_2 + x_3 = 4$$

Now, we can express both x_1 and x_2 in terms of x_3 . We call x_1 and x_2 basic variables and x_3 the free variable. We call x_3 a free variable as we are free to choose any value for it.

Most importantly, these are the conditions for row echelon and reduced row echelon form.

Row echelon form if:

- all nonzero rows are above zero rows
- a leading entry is the right column of a leading entry above it
- all entries in a column below a leading entry are 0

Reduced row echelon also if:

- leading entry in each nonzero row is 1
- the entries in the column of the leading entry are 0

To obtain reduced row echelon form, starting from the lowest row, the terms in the column with the leading term are made zero and the leading term scaled to 1. We repeat this process.

As a general remark, we obtain the following.

Theorem 1.2.1. *A system is consistent iff, the rightmost column has no pivot. More simply, iff the lowest row is not of the form: $\begin{bmatrix} 0 & 0 \dots & b \end{bmatrix}$ where b is nonzero.*

A common confusion when solving a linear system is why the Gaussian algorithm works? To help answer this, the idea is to visualize a Cartesian plane with two random lines intersecting. Now, there can be multiple lines that intersect at the same point. But the intersection of something like $x = 2$ and $y = 5$ is obviously $(2, 5)$ and requires less effort to solve. Thus, we try to replace our equations with those that are easier to solve and have the same solution set.

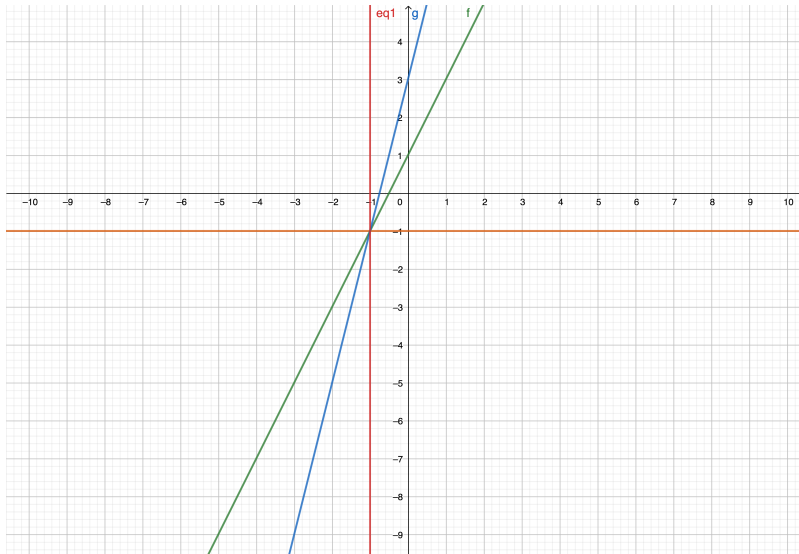


Figure 1: Equivalent systems

1.2.2 Reducing a system to reduced row echelon form

$$\text{Given } \begin{pmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{pmatrix}$$

$$\text{We first realign to get } \begin{pmatrix} 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{pmatrix}$$

After applying elementary row operations, we get the reduced row echelon form of

$$\begin{pmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}$$

Now anything which is not a pivot column is called a free variable.

$$x_5 = 4$$

$$x_2 = 2x_3 - 2x_4 - 7$$

$$x_1 = 2x_3 + 2x_4 - 24$$

1.2.3 High school basics

When solving systems of 2 unknowns or 3 unknowns, we simply are considering the system of lines or planes. Sticking to 3 dimensions, let's quickly recall how we obtain the formula's of a line and plane in 3-d.

A line in 3-d may be described in Vector form as:

$$l[1] : \begin{bmatrix} A \\ B \\ C \end{bmatrix} + \lambda \begin{bmatrix} D \\ E \\ F \end{bmatrix} + \psi \begin{bmatrix} G \\ H \\ I \end{bmatrix}$$

or may be simplified to cartesian form as:

$$k = \frac{x + A}{B} = \dots$$

Now, the equation of a plane is most commonly written in the form:

$$Ax + By + Cz \text{ where } (A, B, C) \text{ is the normal vector to the plane.}$$

2 Introducing vector spaces

2.1 Basics

A vector space is an algebraic structure consisting of two sets, a scalar set and the vector set. The scalar set happens to be field and the vector set an additive abelian group. We also define a 'scalar multiplication' between the scalars and vectors to satisfy the following:

- $r_1 r_2 (v_1) = (r_1 r_2) v_1$ **associativity of scalar multiplication**
- $r_1 (v_1 + v_2) = r_1 v_1 + r_1 v_2$ and the opposite as well, thus multiplication distributes over addition.
- $1_r v_1 = v_1$

Property 1. *From the above properties, we can neatly deduce also that $0_r v_1 = 0$*

$$0_r v_1 = (1 - 1) v_1$$

$$1 v_1 - 1 v_1 = 0$$

The most common vector space is that of \mathbb{R}^n . It's vector and scalar set are the same.

2.2 Linear combinations, span and matrix theorems

A **linear combination** of vectors is simply a sum of vectors multiplied by scalars.

$$v_i = a_1 v_1 + \dots + a_k v_k$$

A **span** for given vector set $\{v_1 \dots v_k\}$ is simply a set that consists of all the possible outputs of $\sum_{i=1}^k a_i v_i$

A common idea in linear algebra is to view a linear combination as the product of two matrices. We can't help but notice that:

$$a_1 v_1 + \dots + a_k v_k = [v_1 \dots v_k] \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$$

An important theorem that follows from the fact that $Ax = b$ is a set of equivalent statements (meaning they always have the same truth value.)

Theorem 2.2.1. *The following are equivalent:*

1. $\forall b \in \mathbb{R}^m, Ax = b$ has a solution
2. $\forall b \in \mathbb{R}^m$ b is a linear combination of the columns of A .
3. columns of A span \mathbb{R}^m
4. A (the coefficient matrix!) has a pivot position in each row.

2.3 Homogeneous system of equations

Any system of linear equations that can be written of the form $Ax = 0$ is called **Homogeneous**. Clearly, this system has at least one solution which is when x is equal to the 0 vector. The non-trivial solution is if x is not equal to 0 in which case the solution is described in terms of a parameter. Now, a theorem that follows is, for a non-homogeneous system $Ax = b$ with solution p , the solution set of $Ax = b$ is the set of vectors of the form $w = p + v_h$ where v_h is any solution to $Ax = 0$.

We now list some useful theorems.

Theorem 2.3.1. *A homogeneous equation $Ax = 0$ has a nontrivial solution iff it has at least one free variable.*

Theorem 2.3.2. *Columns of A are linearly dependent iff $Ax = 0$ has a trivial solution.*

Theorem 2.3.3. *Suppose the equation $Ax = b$ has a solution p . Then the set of possible solutions to $Ax = b$ are of the form $w = p + v_h$ where v_h is a solution of $Ax = 0$. (Ofcourse assuming $Ax = b$ is consistent) That is to say, if we knew two solutions to $Ax = b$, we could find a solution to $Ax = 0$.*

2.4 Linear Independence

A set of vectors is linearly independent if the solution to:

$$a_1v_1 + \dots + a_kv_k = 0$$

is only the trivial solution, that is the $0_v \in \mathbb{R}^k$.

We now present some common statements concerning linear independence.

Theorem 2.4.1. *Whenever $Ax = 0$ has a nontrivial solution, the columns of A are linearly dependent.*

Theorem 2.4.2. *A set of vectors is linearly dependent if at least one of the vectors is a scalar multiple of the other.*

Theorem 2.4.3. *If a set contains the zero vector, then it is linearly dependent.*

Theorem 2.4.4. *If a set of vectors contain more vectors than vector entries, the set is linearly dependent.*

Theorem 2.4.5. *A set of vectors is linearly dependent if the row echelon form of A to $Ax = 0$ has at least one free variable.*

Theorem 2.4.6. *A set of vectors is linearly dependent iff at least one of the vectors can be written as a linear combination of the others.*

Corollary 2.4.1. *Two vectors are linearly dependent iff one is a scalar multiple of the other*

Let's now prove the most useful out of these theorems.

Proof. Let's first prove 2.4.6.

(\rightarrow) Now suppose that for $S = \{v_1, \dots, v_k\}$, $v_1 = a_2v_2 + \dots + a_kv_k$ Then:

$$0 = a_2v_2 + \dots - v_1$$

In which case v_1 has a nonzero coefficient implying a nontrivial solution.

(\leftarrow) Let j represent the vectors that have a nonzero coefficient in $a_1v_1 + \dots + a_jv_j + \dots + a_kv_k$

Now, WLOG, we get for all subscripts less than or equal to j :

$$v_1 = -\frac{a_2}{a_1}v_2 - \dots - \frac{a_k}{a_1}v_k$$

□

2.5 Lec 05 notes

Property 2. A linear T maps linear line segments in \mathbb{R}^n to linear line segments in \mathbb{R}^m

2.5.1 Matrix transformations

In general, the columns of a matrix are the dimensions of the domain set and the rows of the matrix are the codomain.

Theorem 2.5.1. Let T be a square matrix, that is $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Then, T is surjective iff it is injective.

Proof. Proving both directions:

(\rightarrow) (T is injective iff T is surjective)

Now whenever T is surjective, we have that there are no rows of 0. It also means that each row has a pivot. Given this, we check the condition that $T(v_1) = T(v_2) \rightarrow v_1 = v_2$. Now we would essentially obtain this in trying to solve it:

$$T \times v = a$$

where $a \in \mathbb{R}^n$. Now because the reduced matrix has no 0 row and as many pivots as the number of variables, we get that there is only one such v resulting in $v_1 = v_2$

(\leftarrow) (T is surjective iff T is injective) From T is injective, we get that there is only one solution to $T \times v = a$ and considering the homogeneous $Tv = 0$ we get that only 0 would be the solution implying columns of T are a basis hence must span \mathbb{R}^n

□

2.6 Linear transformations

Definition 2.6.1. A linear map is function between $\mathbb{R}^n \rightarrow \mathbb{R}^m$ such that:

- i) $cT(v) = T(cv) \forall c \in \mathbb{R}, \forall v \in \mathbb{R}^n$
- ii) $T(v + u) = T(v) + T(u) \forall v, u \in \mathbb{R}^n$

A condition equivalent to a linear map is:

[Proposition] 1. Whenever we can show that for some T , $T(cv_1 + dv_2) = cT(v_1) + dT(v_2)$, we have that T is a linear map.

Proof. (We must show both conditions i) and ii) defined above)

Now showing i) is simple. Simply fix $d = 0$. Then we get that $T(cv_1 + 0) = cT(v_1)$.

To show ii), fix c, d to be 1 and there we have it!

□

Theorem 2.6.1. For a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, there exists a unique $m \times n$ matrix A such that $T(x) = Ax$

Proof. Now, we can write x as $I_n x$ from which we get $[e_1 \dots e_n] x = x_1 e_1 + \dots + x_n e_n$. Taking $T(x)$ we obtain $x_1 T(e_1) + \dots + x_n T(e_n)$ which is equal to

$$[T(e_1) \dots T(e_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

□

And here are some examples to capturing linear transforms as matrices mainly in \mathbb{R}^2

1. Anti-clockwise rotation by θ degrees. $A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$
2. $T(x) = 3x$ for $x \in \mathbb{R}^2$, then $A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$
3. Reflection through the y-axis $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
4. Contraction or an expansion depending on k $A = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
5. Shear $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
6. Projection $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

And now we present some theorems relating to surjectivity and injectivity:

Theorem 2.6.2. *Taking $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ we have that T is injective iff the homogeneous $T(x) = 0$ has the trivial solution.*

Proof. (To show this, we show that the two statements always have the same truth value, that they are equivalent)

Now when $T(x)$ is injective, it means that $T(x) = 0$ can have only a unique solution by definition of injectivity. Now when $T(x)$ is not injective, it means that $\exists b$ s.t. $T(u) = b$ and $T(v) = b$ where $u \neq v$. Now $T(u) - T(v) = b - b = 0 = T(u - v)$ and since $u \neq v$ we have that 0 is not the only solution. \square

3 Matrix algebra

3.1 Basics

Property 3. *Properties of the matrix transpose:*

$$\begin{aligned} (A^T)^T &= A \\ (A + B)^T &= A^T + B^T \\ (rA)^T &= rA^T \\ (AB)^T &= B^T A^T \end{aligned}$$

Property 4.

$$\begin{aligned} A + B &= B + A \\ (A + B) + C &= A + (B + C) \\ A + 0 &= A \\ r(A + B) &= rA + rB \\ (r + s)A &= rA + sA \\ rs(A) &= r(sA) \end{aligned}$$

And here is the general formula for the $i - j^{th}$ entry of AB whenever defined:

$$(AB)_{ij} = a_{i1}b_{1j} + \dots + a_{in}b_{nj}$$

We note that Matrices whose entries belong to \mathbb{R} form a vector space.

And now some warnings about matrix multiplication. We note that in general if $AB = AC$ then it does not hold that $B = C$. And if $AB = 0$ we can not conclude that A or B is 0. We also note that

whenever a matrix has entries only along its main diagonal, then to take any power, we simply take the power of each diagonal entry.

A transpose of a matrix is defined as swapping the rows of A with its columns.

Theorem 3.1.1. *A matrix A is invertible if $\exists C$ s.t. $AC = I, CA = I$ meaning that the inverse is unique. Non-invertible matrices are called **singular**. And note that only square matrices are invertible. The reason for this is that for a matrix to be invertible, it must be bijective which can only be the case if it is square*

Proof. (Proving that the inverse matrix is unique) Suppose $\exists B, C$ s.t. $AB = I, AC = I$. We want to show that $B = C$.

$$B = BI = BAC = IC = C$$

□

3.2 Computing inverses

The way we find inverses is best done by Gaussian elimination. We are simply looking for a solution to the equation $Ax = I$ hence we write this as an augmented matrix $[A|I]$ and solve.

Theorem 3.2.1. *Elementary matrices are obtained by only performing one elementary row operation on an identity matrix, thus all elementary matrices are invertible.*

Theorem 3.2.2. *Because inverses are only defined on bijective mappings, whenever a square matrix is invertible, we have that $Ax = b$ has a unique solution $x = A^{-1}b$*

3.3 Every matrix is the product of elementary matrices with the identity

Consider the matrix $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$ and $A = E_1 = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ Then notice that $E_1 A = E_1 =$

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g - 4a & h - 4b & i - 4c \end{bmatrix}$$

Which goes on to show that every row operation can be represented as the product of elementary matrices.

3.4 Characterizing inverse matrices

(Theorem on inverse matrices)

- a. A is an invertible matrix.
- b. A is row equivalent to the $n \times n$ identity matrix.
- c. A has n pivot positions.
- d. The equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- e. The columns of A form a linearly independent set.
- f. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.
- g. The equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each \mathbf{b} in \mathbb{R}^n .
- h. The columns of A span \mathbb{R}^n .
- i. The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps \mathbb{R}^n onto \mathbb{R}^n .
- j. There is an $n \times n$ matrix C such that $CA = I$.
- k. There is an $n \times n$ matrix D such that $AD = I$.
- l. A^T is an invertible matrix.

PROOF If statement (a) is true, then A^{-1} works for C in (j), so $(a) \Rightarrow (j)$. Next, $(j) \Rightarrow (d)$ by Exercise 23 in Section 2.1. (Turn back and read the exercise.) Also, $(d) \Rightarrow (c)$ by Exercise 23 in Section 2.2. If A is square and has n pivot positions, then the pivots must lie on the main diagonal, in which case the reduced echelon form of A is I_n . Thus $(c) \Rightarrow (b)$. Also, $(b) \Rightarrow (a)$ by Theorem 7 in Section 2.2. This completes the circle in Figure 1.

Next, $(a) \Rightarrow (k)$ because A^{-1} works for D . Also, $(k) \Rightarrow (g)$ by Exercise 24 in Section 2.1, and $(g) \Rightarrow (a)$ by Exercise 24 in Section 2.2. So (k) and (g) are linked to the circle. Further, (g) , (h) , and (i) are equivalent for any matrix, by Theorem 4 in Section 1.4 and Theorem 12(a) in Section 1.9. Thus, (h) and (i) are linked through (g) to the circle.

Since (d) is linked to the circle, so are (e) and (f) , because (d) , (e) , and (f) are all equivalent for *any* matrix A . (See Section 1.7 and Theorem 12(b) in Section 1.9.) Finally, $(a) \Rightarrow (l)$ by Theorem 6(c) in Section 2.2, and $(l) \Rightarrow (a)$ by the same theorem with A and A^T interchanged. This completes the proof. ■

4 Determinants

4.1 Basics

First of all, determinant is only defined on an $n \times n$ matrix and is denoted $\det(A)$ or simply $|A|$. The most useful formula for a determinant is

$$\det(A) = \sum_{j=1}^n a_{1j} C_{1j}$$

or equivalently

$$\det(A) = \sum_{i=1}^n a_{i1} C_{i1}$$

where C_{ij} which is known as a *cofactor* of matrix A is defined as:

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

wherein A_{ij} denotes a submatrix of A obtained by cancelling out the i th row and j th column of A .

Now the above formula is one way to calculate the determinant. An alternative method comes from the realization that whenever a matrix is triangular, its determinant is the product of the terms along its diagonal. Hence given a matrix A we may row reduce it to its echelon form and simply find the product of all pivots. But noticing how

$$A = E_1 \dots E_n I$$

we get that

$$\det(A) = \det(E_1) \dots \det(E_n)$$

which means that because after row reduction we obtain $[I|A^{-1}]$, $\det(A) = \det(E_1) \dots \det(E_n) \cdot \det(I)$. This brings us to the following properties:

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

And now a very important theorem is:

Theorem 4.1.1. *A matrix is invertible iff $\det(A) \neq 0$*

Proof. When a matrix is invertible, we have that the solution to $Ax = 0$ is unique implying linear independence. And linear independence implies that $\det(A) \neq 0$. \square

Theorem 4.1.2. *If A is $n \times n$, then $\det(A^T) = \det(A)$*

Proof. This is obvious. We have that the cofactors along the columns of A^T are equal to cofactors of A along rows and by definition of the transpose we have that $a_{1j} = a_{j1}$ which implies the coefficients are also equal. This also suggests why the earlier formula for $\det(A)$ comes in two forms. \square

And now yet another theorem:

Theorem 4.1.3. $\det(AB) = \det(A)\det(B)$

And now we come to Cramer's rule. It provides us with an alternative method for solving a system of equations using determinants. It states the following:

Theorem 4.1.4. (Cramer's Rule)

Where A is an $n \times n$ matrix, the solution to $Ax = b$ is of form

$$x_i = \frac{|A_i(b)|}{|A|} \quad i = 1, 2, 3 \dots$$

where $A_i(b)$ is obtained by replacing the i th column of A with the vector b .

Proof. Again quite simple. Notice that

$$A \cdot I_i(b) = A[e_1 \dots b \dots e_n] = [a_1 \dots \underbrace{b}_{\text{since } Ax=b} \dots a_n] = A_i(b)$$

Taking determinants we get

$$\det(A) \underbrace{\det(I_i(b))}_{=x_i} = \det(A_i(b))$$

to give

$$x_i = \frac{\det(A_i(b))}{\det(A)}$$

\square

And now an important theorem is the following:

Theorem 4.1.5. *The area of the parallelogram determined by a_1 and a_2 is the same as that determined by a_1 and $a_2 + ca_1$*

And furthermore:

Theorem 4.1.6. *For a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ we have that whenever S is the generic shape in this vector space than the new area after T is given by:*

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

5 More on vector spaces

5.1 Basics

Definition 5.1.1. A **Vector space** is an algebraic structure satisfying:

1. For some vector set V we have that V is an abelian group under addition.
2. Scalar multiplication between members of a scalar field and vector elements is defined and closure holds.
3. Scalar multiplication of a vector associates
4. $\forall v \in V$ we have that $1 \cdot v = v$
5. Multiplication of a scalar with two vectors distributes and multiplication of a vector with two scalars distributes.

Now three important facts that follow from these axioms are:

$$0v = 0 \quad \forall v \in V \quad (5.1.1)$$

$$c0 = 0 \quad \forall c \in F \quad (5.1.2)$$

$$-u = (-1)u \quad \forall u \in V \quad (5.1.3)$$

Let's prove these solely using the axioms.

Proof. **5.1.1**

$$0v = (a - a)v = \underbrace{av - av}_{\text{additive inverse}} = 0$$

5.1.2

$$c0 = c(0 - 0) = c0 - c0 = 0$$

5.1.3

$$\begin{aligned} 0u &= (1 - 1)u \\ 0 &= u + (-1)u \\ -u &= \underbrace{-u + u + (-1)u}_{\text{right additive inverse}} \end{aligned}$$

□

Definition 5.1.2. For a vector space V , S is said to be a subspace if:

1. $0_v \in S$
2. S is closed under scalar multiplication and vector addition

Definition 5.1.3. (Null space)

For some matrix A , the **Null space** is such that:

$$\text{Nul}A = \{v \in V | Av = 0_v\}$$

[Proposition] 2. *NulA is a subspace of V. Now clearly $0 \in V$ as 0 is the trivial solution to $Ax = 0$. We now need that $\forall u, v \in \text{Nul}A, u + v \in \text{Nul}A$.*

$$A(u + v) = \underbrace{Au + Av}_{\text{matrices are linear maps}} = 0 + 0 = 0$$

We now also need $\forall u \in \text{Nul}A, cu \in A$

$$A(cu) = cA(u) = c0 = 0$$

Theorem 5.1.1. *The span of any single vector space elements is a subspace.*

Definition 5.1.4. *(Column space)*

The column space $\text{Col}A$ is the set of all possible linear combinations of the columns of a matrix A and is also a subspace of \mathbb{R}^n where $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$.

And now some examples of vector spaces:

- \mathbb{R}^n is the most common example equipped with \mathbb{R} as scalars.
- Set of polynomials of degree at most n denoted P_n
- set of all real valued functions. This one is a little more subtle. Our 'vectors' are functions $f(x)$ and for instance the 0 vector is the map $g(x) = 0$. We define addition as $f(x) + g(x) = (f + g)(x)$ and scalar multiplication as $cf(x) = f(cx)$

Under light of vector spaces, a **linear transformation** is now much more clear as being a mapping that respects the two operations of vector spaces.

5.2 Some problems from week 6

Claim:

Suppose matrix B is the reduced echelon form of the columns of A . Then the pivot columns of B are a basis of $\text{Col}(A)$

It is easy to see that this is false. Suppose A is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$ We then have that B is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ but clearly pivot columns of B do not span $\text{Col}(A)$

5.3 More theorems

Theorem 5.3.1. *Let $S \subseteq V$ where S is non-empty. And let $H = \langle S \rangle$. Whenever one element of S is l.d. on the other vectors, removing this element from S , we have that S still spans H . And whenever $H \neq \{0_v\}$, some subset of S is a basis for H .*

Theorem 5.3.2. *The pivot columns of a matrix A form a basis for $\text{Col}(A)$*

Warning: *This does not mean that the pivot columns of the echelon form of A are a basis to $\text{Col}(A)$*

And now we come to a beautiful theorem of linear algebra, the unique representation theorem.

Theorem 5.3.3. *Any $v \in V$ is represented as the unique linear combination of the bases of V*

Proof. We have that:

$$v = \sum_{i=0}^k a_i x_i$$

Now suppose also that:

$$v = \sum_{i=0}^k b_i x_i$$

To give

$$\underbrace{v - v}_{\text{additive inverse}} = 0 = \sum_{i=0}^k a_i x_i - \sum_{i=0}^k b_i x_i = (a_0 - b_0)x_0 + \dots + (a_k - b_k)x_k$$

Now since our x_i are linearly independent only the trivial solution holds hence all $a_i - b_i = 0$ meaning that $a_i = b_i$ \square

Definition 5.3.1. Suppose $B = \{b_1 \dots\}$ is a basis for V and $x \in V$. Then the **coordinates** of x are

simply weights c_i such that $x = \sum_{i=1}^k c_i b_i$ and are denoted $[x]_b = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix}$

Example 5.3.1. (Change of coordinate matrix)

The matrix equation

$$x = P_b x_b$$

transforms a vector x with coordinates relative to b into its coordinates in the standard bases. For instance

suppose that $x = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ and that $b_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $b_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ Then

$$\begin{bmatrix} 4 \\ 5 \end{bmatrix} = c_1 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = P_b x_b$$

Now crucially note that because $\text{Col}(P_b)$ is a basis for \mathbb{R}^n we have that it is invertible.

Theorem 5.3.4. (Basis conversion theorem) To obtain coordinates of a given vector relative to some basis B we do:

$$[x]_b = P_b^{-1} x$$

Another important property of vector spaces is that although for different bases, the coordinate vectors differ, a vector remains always the same. And now we come to present the coordinate mapping theorem.

Theorem 5.3.5. (Coordinate mapping theorem)

The vector x is mapped to $[x]_B$. This is an example of a mapping from some random vector space onto \mathbb{R}^n where $n = \dim(V)$ From this we get that any finite-dimensional vector space has a linear bijection (aka. isomorphism) to \mathbb{R}^n by simply using the coordinate mapping

And another important theorem is the linear independence theorem:

Theorem 5.3.6. If a vector space has a basis set with n elements, then another set with more than n elements has linearly dependent elements.

Proof. Let $\{u_1 \dots u_p\}$ be the set with more than n vectors. Now we have that

$$c_1 [u_1]_B + \dots + c_p [u_p]_B = 0$$

has a non-trivial solution in \mathbb{R}^n since we have more vectors than entries. Now by linearity of the coordinate map we have:

$$[c_1 u_1]_B + \dots + [c_p u_p]_B = 0$$

where at least one of the c_i is not zero. \square

Theorem 5.3.7. *The dimension of $\text{Nul}(A)$ is the number of free variables to the equation $Ax = 0$ which makes sense since every other vector in $\text{Nul}(A)$ is linearly dependent on a number of vectors equal to the number of free variables. The dimension of $\text{Col}(A)$ is the number of pivot columns*

As an aside now, we state that differentiation of polynomials which generally maps $\mathbb{P}_n \rightarrow \mathbb{P}_n$ can be represented as a matrix transform. That is suppose $v \in \mathbb{P}_2 = a_0 + a_1x + a_2x^2$ Now we have that $T(v) = \frac{dv}{dx}$ The matrix mapping of this is:

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

Definition 5.3.2. *The rank of a matrix A is simply equal to $\dim(\text{Col}(A))$*

A nice theorem making use of rank is:

Theorem 5.3.8. *If A is an $m \times n$ matrix then A has rank 1 iff there are non-zero vectors $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ such that $A = uv^T$*

Proof. The forward direction says that given A with rank one, we have that any i elements $\in \text{Col}(A)$ are linearly dependent. But notice that uv^T is equivalent to:

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \begin{bmatrix} v_1 & \dots & v_m \end{bmatrix}$$

which creates a matrix where every column is a scalar multiple of the other as the columns of uv^T look like;

$$v_1 \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, v_2 \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \dots$$

From this the backward direction trivially follows as well. □

5.4 Theorems of Rank, Col(A), Row(A) and Null(A)

Theorem 5.4.1. *(Rank nullity theorem for matrices)*

This theorem states that for an $m \times n$ matrix A :

$$\dim(\text{Col}(A)) + \dim(\text{Null}(A)) = n$$

Proof. Intuitively this should be obvious because $\dim(\text{Col}(A))$ gives us the number of pivot columns in echelon form of A and similarly any remaining columns give us the dimension of $\text{Nul}(A)$. Summing the two we get n . □

Theorem 5.4.2. *For a matrix A with echelon form B , the row space of A and B are the same. The column space of A has the same dimension as the column space of B yet is not equivalent to it. However the column indices for which B has pivot columns form a basis for A .*

5.5 Change of basis

Recall that we have already stated that every finite dimensional vector space is isomorphic to \mathbb{R}^n through $x \rightarrow [x]_B$ known as the coordinate mapping. Given this coordinate mapping we now consider what it means to change basis.

Example 5.5.1. Suppose that we have two bases B_i and C_i for some 2-dimensional V . And suppose that $b_1 = 4c_1 + c_2$ and $b_2 = -6c_1 + c_2$. Given that $[x]_B = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ we want to find $[x]_C$. Now we have:

$$[x]_B = [3b_1 + b_2]_C = 3[b_1]_C + [b_2]_C$$

which simplifies to the matrix multiplication:

$$\begin{bmatrix} (b_1)_C & (b_2)_C \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = [x]_C$$

From this we obtain the change of basis formula which is:

$$[x]_C = \underbrace{P_{B \rightarrow C}}_{\text{known as change of basis matrix}} [x]_B$$

Now since the change of basis matrix is a non-zero square matrix, it follows that it is also invertible and in particular:

$$P_{B \rightarrow C}^{-1} = P_{C \rightarrow B}$$

Example 5.5.2. Consider bases $B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $C = \left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$. And suppose that we want to find $P_{B \rightarrow C}$. Now this is equivalent to asking what the coordinate vectors of the basis B are in the basis C . This results in having to solve 3 separate systems of the form:

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + b \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} + c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

which we simplify to the augmented matrix:

$$\left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 2 & 2 & 0 & 1 & 1 & 0 \\ 3 & 0 & 1 & 0 & 0 & 1 \end{array} \right]$$

We generalize our above example to that, to find $P_{B \rightarrow C}$ solve the matrix:

$$[P_C \mid P_B] \rightarrow [I \mid P_{B \rightarrow C}]$$

5.6 Eigenvectors and eigenvalues

Definition 5.6.1. An **eigenvector** of a square matrix is a vector such that Ax can be written in form λx where λ is a scalar. Of course when we state $Ax = \lambda x$, x must be a non-trivial solution.

From the defining equation, we obtain that:

$$(A - \lambda I)x = 0$$

and the set of all solutions belongs to the null space of the matrix $(A - \lambda I)$ and is called the **eigenspace** of A corresponding to λ .

Theorem 5.6.1. The eigenvalues of a triangular matrix are the entries along its main diagonal.

Now in general, what follows about determining if a matrix has an eigenvector is shown by the following example:

Property 5. Suppose that $Ax = \lambda x$ where $x \neq 0_v$. Then it must follow that:

$$(A - \lambda I)x = 0$$

has a nontrivial solution. Now since $A - \lambda I$ is a square matrix, we have that it will have a nontrivial solution x iff. its columns are linearly dependent. But note that this is equivalent to saying that $\det\{A - \lambda I\} = 0$. The fact the determinant is 0 iff. the matrix columns are linearly dependent follows from the fact that the homogeneous equation of a square matrix in its echelon form will have at least one zero row to admit a nonzero solution and given that determinant of a triangular matrix is the product of entries in main diagonal we have that it will have determinant equal to zero.

Theorem 5.6.2. If the matrix $(A - \lambda I)$ has no linearly dependent columns then it has no eigenvectors.

Now note that we cannot use the equation $(A - \lambda I) = 0$ to solve for λ

Let's continue our exploration of eigenspaces noting that an eigenspace is a subspace of V_n :

Example 5.6.1. Suppose $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$ and we are given that 2 is an eigenvalue of A . Now we have

that $A - 2I = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$ and in reduced homogeneous form we get:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which gives a general solution: $x = y - 6z$ and in parametric form:

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

Thus telling us also that the eigenspace for 2 is a two-dimensional subspace of \mathbb{R}^3 . Visually this is to say that matrix multiplication by A has the same effect as scalar vector addition.

Given this, we now prove our earlier theorem 5.6.1.

Proof. Note that a matrix has eigenvectors iff $A - \lambda I$ has determinant 0. Now determinant of a square triangular matrix is the entries along its diagonal hence we get:

$$(a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) = 0$$

which completes our proof. □

And now we come to an interesting theorem of eigenvectors.

Theorem 5.6.3. Suppose that for a square matrix A , v_i are eigenvectors corresponding to eigenvalues λ_i . Then we have that the v_i are linearly independent.

Proof. Now suppose that p is the least index for which some $\underbrace{v_{p+1}}_{v_i \text{ are all nonzero}}$ can be written as a linear combination of vectors with smaller indices. Then we have that:

$$v_{p+1} = a_1 v_1 + \dots + a_p v_p$$

where at least one of the $a_i \neq 0$. Now using the fact that $Av_k = \lambda_k v_k$ we multiply the above by A and replace the Av_k to get:

$$\lambda_{p+1} v_{p+1} = \lambda_1 a_1 v_1 + \dots + \lambda_p a_p v_p$$

As a final step we multiply our first equation with λ_{p+1} and subtract it from the above to get :
our goal in doing is to obtain 0 on LHS

$$a_1(\lambda_1 - \lambda_{p+1})v_1 + \dots + a_p(\lambda_p - \lambda_{p+1})v_p = 0$$

Now because our v_i up to p are linearly independent we have that the coefficients must all be 0. Since no $\lambda_i - \lambda_{i+1}$ can equal 0, we have that all of our a_i must be 0. But then v_{p+1} would be 0 which it can't be since it is an eigenvector. \square

Now to present our next theorem we present the reminder that if we reduce square matrix A to U only using r many row exchanges and row addition then:

$$\det\{A\} = \begin{cases} (-1)^r \det\{U\} \\ 0 \text{ if not invertible} \end{cases}$$

Now given our findings of eigenvectors we extend the invertible matrix theorem:

Theorem 5.6.4. (*Extending invertible matrix theorem*)

1. A is invertible iff. 0 is not an eigenvalue of A This holds because than the echelon form of A will not have determinant 0 implying columns of A are linearly independent
2. $\det\{A\}$ is not 0.

And this brings us to the characteristic equation theorem.

Theorem 5.6.5. (*Characteristic equation theorem*)

λ is a eigenvalue of square matrix A iff. it satisfies:

$$\det\{A - \lambda I\} = 0$$

The expansion of $\det\{A - \lambda I\}$ is known as the **characteristic polynomial**. The **multiplicity** of some λ value is in how many repeated roots λ_1 appears.

Moving on, two square matrices are said to be similar if both:

$$\begin{aligned} P^{-1}AP &= B \\ PBP^{-1} &= A \end{aligned}$$

The point of two matrices being similar is that whenever two matrices are similar, they essentially represent the same linear transformation between spaces however only in different bases.

Theorem 5.6.6. *If A and B are similar matrices, then they have the same characteristic polynomial hence the same eigenvalues.*

Theorem 5.6.7. (Representation of one matrix in another basis) Consider how we might represent one matrix in another basis B . That is suppose $A = \begin{bmatrix} -1 & 4 \\ -2 & 3 \end{bmatrix}$ and we are given a basis

$B = \left\{ \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ Now it should be somewhat obvious that we must first apply the transformation to the bases and then find the coordinate of each result in the given basis. The matrix formed by these is then simply the coordinates of the bases B after the transformation. Doing this follows as first finding the product AB since for instance $A\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \text{first column of } AB = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$ And after we will have to solve $\begin{bmatrix} 5 \\ 0 \end{bmatrix} = a \begin{bmatrix} 3 \\ 2 \end{bmatrix} + b \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ Formalizing this a little more we get that finding A in basis B follows as solving:

$$[B \mid AB] \rightarrow [I \mid A_B]$$

In fact we can be even more formal. The matrix A_B is nothing more than a matrix that takes as input a coordinate vector in basis B and applies transformation A on it. We obtain that:

$$A_B = \underbrace{B^{-1}}_{\text{obtain coordinates of bases after transformation}} AB$$

Diagonalization, a useful computational tool

Now the first and foremost theorem of diagonalization is:

Theorem 5.6.8. *An $n \times n$ matrix A is diagonalizable iff. it has n linearly independent eigenvectors.*

Proof. We claim that $AP = PD$. Now

$$AP = A \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}$$

and similarly

$$PD = P \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ & \lambda_2 & \dots & 0 \\ \vdots & \dots & & \lambda_n \end{bmatrix}$$

Now by vector equality we have that $Av_i = v_i \lambda_i$. Since now P is an invertible matrix, we have that the v_i are linearly independent. \square

Now to diagonalize a matrix (ofcourse if possible only), we first find all eigenvalues by solving the characteristic polynomial. After we find the bases for each eigenvalue. Finally, P is constructed as the matrix where each column is a basis corresponding to a respective eigenvalue found in the column vectors of matrix D .

5.7 Orthogonality and inner products

Definition 5.7.1. Let $u, v \in \mathbb{R}^n$. Then we define the inner product of u and v as the matrix multiplication $u^T v$ such that

$$u \cdot v = u^T v = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \dots + u_n v_n$$

Definition 5.7.2. The **norm** of a vector is defined in terms of inner product as $|v| = \sqrt{v \cdot v}$

Definition 5.7.3. Distance in \mathbb{R}^n is described by:

$$\text{dist}(u, v) = |u - v|$$

Definition 5.7.4. u and v are orthogonal if $u \cdot v = 0$

Definition 5.7.5. The orthogonal complement of some subspace L is denoted W^\perp . It contains all vectors that are orthogonal to all vectors in L .

Theorem 5.7.1. $\text{Row}(A)^\perp = \text{Nul}A$ and also $\text{Col}(A)^\perp = \text{Nul}A^T$

Proof. For $\text{Row}(A)^\perp = \text{Nul}A$, we have that every vector in $\text{Nul}A$ has inner product 0 (aka. orthogonal) to every row of A .

For $\text{Col}(A)^\perp = \text{Nul}A^T$ we use the fact that our above theorem is always true and given that $\text{Col}A = \text{Nul}A^T$ we have $\text{Col}(A)^\perp = \text{Nul}A^T$. \square

Definition 5.7.6. A set of vectors is an orthogonal set if each vector in the set is orthogonal to every other vector. Similarly, a **orthogonal basis** is a basis which is also an orthogonal set.

Theorem 5.7.2. The usefulness of an orthogonal basis is in the fact that it is easy to find coordinates. We have that:

$$c_i = \frac{v \cdot u_i}{u_i \cdot u_i}$$

Proof.

$$v \cdot u_i = (c_1 u_1 + \dots + c_n u_n) \cdot u_i = c_i (u_i \cdot u_i)$$

□

Definition 5.7.7. A set is orthonormal if it is an orthogonal set of unit vectors.

Theorem 5.7.3. An $m \times n$ matrix has orthogonal columns iff. $U^T U = I$

And yet another useful theorem with its proof:

Theorem 5.7.4. For $x, w \in \mathbb{R}^n$ and A an $m \times n$ matrix we have: $w \cdot (Ax) = x \cdot (A^T w)$

Proof. To see why this is true consider $U^T U$. Now this is equivalent to:

$$\begin{bmatrix} \dots u_1 \dots \\ \dots u_2 \dots \\ \dots u_3 \dots \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$$

Therefore it should be obvious that only the product of vectors with same indices should be the same producing the identity matrix.

□

Proof.

$$\begin{aligned} w \cdot (Ax) &= \underbrace{(Ax) \cdot w}_{\text{by commutativity of dot product}} \\ &= (Ax)^T w = (x^T A^T) w \\ &= x \cdot (A^T w) \end{aligned}$$

□

Theorem 5.7.5. For a matrix with orthonormal columns we have:

1. $|Ux| = |x|$
2. $(Ux) \cdot (Uy) = x \cdot y$
3. $(Ux) \cdot (Uy) = 0$ iff. $x \cdot y = 0$

And now we come to a very useful theorem:

Theorem 5.7.6. Orthogonal decomposition theorem

Let $y \in \mathbb{R}^n$. Then $y = y' + z$ where:

$$y' = \frac{y \cdot u_1}{u_1 \cdot u_1} + \dots + \frac{y \cdot u_p}{u_p \cdot u_p}$$

with an orthogonal basis for $\{u_1, \dots, u_p\}$ $W \leq \mathbb{R}^n$. And we also have that $z = y - y'$

Proof. Define a basis $\{u_1, \dots, u_p\}$ for W . We then have that

$$y' = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \dots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p$$

Now we want to show that $z \in W^\perp$. For this to be true, z must be orthogonal to all the basis vectors u_i . Given that $z = y - y'$ we have that:

$$(y - y') \cdot u_1 = y \cdot u_1 - y' \cdot u_1 = y \cdot u_1 - \left(\frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \dots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \right) \cdot u_1 = y \cdot u_1 - y \cdot u_1 = 0$$

By the very same argument, we have that z is orthogonal to all the other u_i hence we have that $z \in W^\perp$

Finally we show uniqueness:

Suppose that $y = y' + z$ and also that $y = x' + w$ Then we have that:

$$y' + z = x' + w$$

to further give:

$$y' - x' = w - z$$

which implies that $v = y' - x' \in W, W^\perp$ which is only possible if $v = 0$. This implies that $y' = x'$ and also that $w = z$. \square

And now comes a theorem about measurement in \mathbb{R}^n .

Theorem 5.7.7. *We have that:*

$$|y - y'| < |y - v| \text{ for any } v \text{ distinct from } y' \text{ and } y' \text{ defined as } \text{proj}_W y$$

Proof. We have that:

$$y - v = (y - y') + (y' - v)$$

and by the Pythagorean theorem we rather obviously have that:

$$|y - v| > |y - y'|$$

\square

Theorem 5.7.8. *Given that $U = [u_1 \ u_2 \ \dots \ u_p]$ and U has orthonormal column vectors we have that:*

$$y' = UU^T y$$

The above is clear considering that when we have an orthonormal basis $y' = (y \cdot u_1)u_1 + \dots$ that the weights $y \cdot u_i$ are equivalent to $u_i^T y$ which is equivalent to $U^T y$

Theorem 5.7.9. *For a subspace W of \mathbb{R}^n we have that:*

$$\dim(W) + \dim(W^\perp) = n$$

Proof. Put the basis for W and W^\perp into one set. Now this is an orthogonal set which implies linearly independent. Similarly, we have that every $y \in \mathbb{R}^n$ can be written as a linear combination of the elements in this set. Hence we are done. \square

Theorem 5.7.10. Uniqueness and existence theorem *For some $Ax = b$ we have that the equality has a solution iff. $b \cdot y = 0 \ \forall y \in \text{Nul}(A^T)$*

Proof. Firstly we have that $\mathbb{R}^n = \underbrace{W}_{\text{Col}(A)} \oplus W^\perp$ and also that $b = y' + z$ Now for $Ax = b$ to have a solution,

we need that $b \in W$ this is only possible if $z = 0$ which implies that $\forall z \in \text{Nul}(A^T)$ we have $b \cdot z = 0$ \square

6 Factorizations and applications

Gram-Schmidt We find an orthonormal basis for any vector space as follows:

Let x_i form a basis. Now define:

$$\begin{aligned} v_1 &= x_1 \\ v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1^2} v_1 \\ &\vdots \end{aligned}$$

That is, the v_i we find this way form an orthonormal basis.

QR factorisation

Theorem 6.0.1. *If A is an $m \times n$ matrix with linearly independent columns we have that $A = QR$ where columns of Q are an orthonormal basis for $\text{Col}(A)$ and R is upper triangular.*

Proof. Now we have that $X = A = [x_1 \ \dots \ x_n]$ Let $U = [u_1 \ \dots \ u_n]$ be the matrix with each column being an orthonormal vector. Now by rewriting Gram-Schmidt we have that each x_i is a linear combination of the v_i . Hence the matrix R is the matrix simply containing these weights and we are done. \square

Definition 6.0.1. Least squares problem When we try to estimate say a plot for some experiment data, the most natural approach is to minimize the distance between our chosen trendline and outlier experiment points. This generally leads to an equation of the form $b - Ax$ where our goal is to minimize this difference. For instance a polynomial estimation of degree 2 leaves us with:

$$b - (a_0 + a_1x^1 + a_2x^2) = b - \begin{bmatrix} x^0 & x^1 & x^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

Thus we have that a least squares solution is:

$$|b - Ax'| \leq |b - Ax| \quad \forall x \in \mathbb{R}^n$$

Theorem 6.0.2. The general solution to the least squares problem This turns out to be the projection of the vector b onto the columns space of A . Given that $b = b' + z$ we have that $b - b'$ is orthogonal to the column space of A which implies $A^T(b - Ax) = 0$ Thus we conclude that the least square solution must always satisfy:

$$A^T b = A^T A x$$

Thus the general way to find a least squares solution is to solve for x by finding $(A^T A)^{-1} A^T b$. Yet we note that this is only the case if $A^T A$ is invertible which happens when **the columns of $A^T A$ are invertible**.

Finally, if our A happens to have linearly independent columns then the least squares solution to $Ax = b$ is even easier to find. We have that $A = QR$ and given that for orthonormal matrices $Q^T = Q^{-1}$ we get:

$$x' = R^{-1} Q^T b$$

Theorem 6.0.3. *If A is a symmetric matrix, then any two eigenvectors from a different eigenspace are orthogonal. We omit the proof but the idea is to consider $\lambda_1 v_1 \cdot v_2$ get that it is equal to $\lambda_2 v_1 \cdot v_2$ and conclude.*

By the same idea we know that A and A^T have the same eigenvalues and we obtain that any two eigenvectors from a different eigenspace are orthogonal.

Theorem 6.0.4. *A square matrix A is orthogonally diagonalizable that is P being an orthogonal matrix iff. A is symmetric.*

Matrix multiplication can be viewed as a series of outer products We define an outer product as:

$$x \otimes y = xy^T$$

A useful tool: Spectral decomposition

Given our above theorem, we have that if A is square and symmetric, then it is orthogonally diagonalizable. So let $A = PDP^{-1}$. Taking P to be an orthonormal matrix we obtain $A = PDP^T$. Now this simplifies to:

$$\begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \dots \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix}$$

to finally give:

$$A = \lambda_1 u_1 u_1^T + \dots + \lambda_n u_n u_n^T = \sum_{i=1}^n \lambda_i u_i \otimes u_i$$

Definition 6.0.2. Quadratic form

A quadratic form is a function that maps \mathbb{R}^n to \mathbb{R} and is of form $Q(x) = x^T A x$ where A is a square symmetric matrix. For instance if $A = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$ then $Q(x) = 4x^2 + 6xy + 2y^2$

We may also introduce the change of variable $x = Py$ where we represent our vector in \mathbb{R}^n relative to some other basis in columns of P and y the coordinate vector in this basis. Then we have:

$$\begin{aligned} & x^T A x \\ & (Py)^T A Py \\ & y^T (P^T A P) y \end{aligned}$$

Now since A is a symmetric matrix, by 6.0.3 we have that there is a diagonal matrix $D = P^T A P$ hence $x^T A x = y^T D y$ meaning that our quadratic form may be expressed without a cross product term.

Theorem 6.0.5. Principal axis theorem Our above result is now formalized to state that every quadratic form has a change of variable $x = Py$ such that $x^T A x = y^T D y$ where the second term has no cross product terms in it. Note that to find a principal axis, because we know that A is symmetric, it has an orthogonal diagonalization hence we find out D which is the matrix with eigenvalues as column entries. Then P which is the principal axis is simply the matrix with eigenvectors corresponding to each eigenvalue.

Definition 6.0.3. $Q(x)$ is positive definite if > 0 negative definite if < 0 and indefinite if both for $x \in \mathbb{R}^n$

Theorem 6.0.6. $Q(x)$ is positive definite iff. all eigenvalues of A are positive (resp. definite). This is clear given that:

$$x^T A x = y^T D y = \lambda y_1^2 + \dots$$

Definition 6.0.4. Constrained optimization Suppose we want to find a maximum value to the quadratic

form $Q(x) = 9x^2 + 4y^2 + 3z^3$ with the constraint that our solution is a unit vector that is with $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

and $|X^T X| = 1$. Since:

$$9x^2 + 4y^2 + 3z^3 \leq 9(x + y + z) = 9$$

and since $Q(x) = 9$ has a solution, 9 is our maximum.

Theorem 6.0.7. Let A be a symmetric matrix and define:

$$m = \min\{x^T A x : |x| = 1\} \quad M = \max\{x^T A x : |x| = 1\}$$

Then we have that m is the smallest eigenvalue of A and M the largest eigenvalue of A .

With eigenvalues $a \geq b \geq c$ as eigenvalues, the idea of the proof is to consider that $|x| = |y|$ then write out $y^T D y = ay_1^2 + by_2^2 + cy_3^2 \leq ay_1^2 + ay_1^2 + ay_1^2 = a(|y|) = a$ hence $M \leq a$ but clearly with $y = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ hence $a = M$

Theorem 6.0.8. In the previous case, our only constraint was that $|x| = 1$ namely that $x^T x = 1$. Now if we add also that $x^T u_1 = 0$ meaning that the first component of x is 0. Then it follows that the maximum possible value with this constraint is the second largest eigenvalue. Clearly then if we also add $x^T u_2 = 0$ the largest value becomes the third eigenvalue and so on.

Theorem 6.0.9. $A^T A$ is symmetric.

Theorem 6.0.10. In our previous case, we had that A was symmetric hence finding a maximum was easy because we could always find a principal axis P that led to $x^T A x = y^T D y$ making it easy to find a maximum. Now however, we consider the case where A is not symmetric and we still want to find a maximum. I.e. such a mapping is the matrix that maps the unit sphere to some ellipse. We notice that the maximum of $|Ax|$ also maximises $|Ax|^2$ which is much easier to study as $|Ax|^2 = (Ax)^T (Ax) = x^T A^T A x$ and fortunately, $A^T A$ is always square which means we can apply 6.0.7 and know that the greatest eigenvalue λ_i is the greatest value of $|Ax|^2 = x^T A^T A x$. Thus we obtain that the greatest value of a non-symmetric Ax is given by $\sqrt{\lambda}$ and the x satisfying this value, that is the element in the domain is any eigenvector v_i with eigenvalue λ_i . Hence, we say that a **singular value** of A is $|Av_i|$

Theorem 6.0.11. Suppose that for A of dimension $m \times n$, v_1, \dots, v_n is an orthogonal basis for \mathbb{R}^n where each v_i is an eigenvector of λ_i where the λ_i are eigenvalues of $A^T A$. Now, when A has r non-zero singular values, that is r times next maxima, we have that $\{Av_1, \dots, Av_r\}$ are a basis for $\text{Col}(A)$ and that $\text{rank} A = r$.

Theorem 6.0.12. Singular value decomposition When we have an $m \times n$ matrix A with rank r , we may write it as $U \Sigma V^T$ where Σ is the diagonal matrix with r many eigenvalues of A in decreasing order. By considering $A^T A = V \Sigma^T \Sigma V^T$, we see that the entries of Σ are also singular values of A .

Here is how we proceed to find the unique singular value decomposition.

1. Find non-zero eigenvalue of $A^T A$.
2. Construct Σ diagonally from these eigenvalues not forgetting to take square-root!
3. Construct each column of V^T as eigenvectors corresponding to columns of Σ
4. Finally, columns of U are Av_i and we are done!