CO145 - Mathematical Methods

Prelude

The content discussed here is part of CO145 - Mathematical Methods (Computing MEng); taught by Michael Huth, and Mario Berta, in Imperial College London during the academic year 2018/19. The notes are written for my personal use, and have no guarantee of being correct (although I hope it is, for my own sake). This should be used in conjunction with the lecture notes. This module differs as there isn’t as much new content, but it requires practice - as such, I will likely be including worked examples for my own benefit (which are probably incorrect).

Also, the formatting of this document will be a mess. I’d waste too much time trying to figure out how to adjust line spacing otherwise.

Sequences

Formal Definition of a Limit

A sequence an, for n ≥ 1, converges to some limit l ∈ R if, and only if, we can prove ∀ε > 0[∃Nε ∈ N[∀n>Nε[|an − l| < ε]]]. To show l = 0. absolute, can take an integer, it proves convergence We can now for the sequence an attempt to find some = Nn1ε. , we As thus it’s sufficient to reciprocals on both sides, hence it follows Nε that a limit exists.

= find ⌈to 1ε⌉. n get such For n any > that 1ε. value However, n 1need to first make a guess for the limit - suppose < n 1ε. − 0 Since is positive for all n ∈ N, both are positive (hence we can drop the non-zero), we we are restricted by the fact that n must be of ε, we can get some Nε with the function, thus

Common Converging Sequences

Note that for all of these, we are implicitly saying n→∞lim , and that an → 0.

an condition Nε 1n1c ccn n 1for some c ∈ R+ for some c ∈ R, such that |c| > 1 for some c ∈ R, such that |c| < 1 ⌈log⌈log⌈ εc1cc⌉

((ε)⌉ 1ε)⌉

n!1 ln(n) n > 1 ⌈e

1ε⌉

Combining Sequences

Suppose that an → a, and bn → b, as n→∞lim ;

• n→∞lim λan = λa given λ ∈ R

• n→∞lim (an + bn) = a + b

• n→∞lim (anbn) = ab

• n→∞lim abn n= ab given b = 0

For example, the sequence an However, if we divide every term = 4n2+3n

7n2+3n−2, it’s trivial to find by n2, we end up with an = the limit 7+ 4+ n3− n

3n22 as n → ∞ by inspection as 47. , which we can break into an = bcnn, where bn sequences =4+ (but I n3really , and cn cannot =7+ be n 3bothered − n22. Using the rules from above, we can further break down the

to do so), to a point where we get a = 7+0−0 4+0

= 47.

1

Sandwich Theorem

In the sandwich theorem, where we want to prove that lim n→∞an = l, we need two sequences that form upper, and lower bounds for an, namely un, and ln. If such sequences exist, and satisfy ∃N ∈ N[∀n ≥ N[ln ≤ an ≤ un]], and both lim n→∞un = lim n→∞ln = l, then we get lim n→∞an = l. For example, consider the sequence an = cos(n)

n . We know that −1 ≤ cos(n) ≤ 1, therefore ln = − 1n ≤ an ≤ 1n = un. However, as both un → 0, and ln → 0, when n → ∞, it follows that lim n→∞an = 0. The sandwich theorem can be proven by finding Nεl, and Nεu for ln, and un respectively. As they both converge to the same limit, we can justify that for some Nε = max(Nεl,Nεu), any n>Nε, we have |ln−l| < ε, and |un−l| < ε. By removing the modulus signs, we get −ε<ln−l<ε, and −ε<ln−l<ε. By rearranging this, and knowing ln < un by definition, we have l−ε<ln < un < l+ε. In order to apply the sandwich theorem, we have to assume ln ≤ an ≤ un, hence it follows l − ε<ln ≤ an ≤ un < l + ε. This can then be arranged to get −ε<an − l<ε, thus for all n>Nε, |an − l| < ε.

Ratio Tests

The follow tests allow us to check if a sequence an converges to 0, or diverges. If we want to verify a sequence lim n→∞bn = l, where l = 0, we can reformulate it as an = bn − l, and prove it for an. If |an+1

an | ≤ c < 1, for c ∈ R, where n is sufficiently large, then lim n→∞an = 0. Conversely, if |an+1

an | ≥ c > 1, then an diverges. Suppose an = 2−n, in order to show it converges, we want to find some c, such that |2−(n+1)

2−n | ≤ c < 1. By arithmetic, we can say |2−(n+1)

2−n | = 12 ≤ c < 1. As there exists real c = 12 < 1, we can conclude that an converges to 0. However, we cannot prove that an = 1n! converges, despite it being a fairly straightforward proof with just the standard ration tests. For example, we would end up with |an+1

an | = 1n+1. While this is fairly conclusive that it’s less than 1, we don’t get a constant c which sits between it, and 1. Instead of analysing the behaviour of consecutive terms, we take the limit as r = lim n→∞|an+1

an |. If r < 1, it converges to 0, but if r > 1, it then diverges. Applying this to the previous factorial example, we can clearly see that lim n→∞ 1n+1 = 0, therefore an converges to 0.

Manipulating Absolute Values

A useful decomposition with absolute limits is |x| < a ⇔ −a<x<a. Further useful properties are;

• |xy| = |x|·|y|

• |xy| = |x| |y|

• |x + y|≤|x| + |y| triangle inequality

• |x − y| ≥ ||x|−|y||

Properties of Real Numbers

Suppose we have some set S of real numbers, then the following properties hold;

• u is an upper bound on S if ∀s ∈ S[u ≥ s]

• l is a lower bound on S if ∀s ∈ S[l ≤ s]

• S is bounded above if it has an upper bound, bounded below if it has a lower bound, and bounded if it has both

• a set may have no bounds, or many, therefore we say sup(S) is the least upper bound of S, and inf(S) to denote the greatest upper bound of S2

• if a set has no upper bound, then sup(S) = ∞, and similarly if there is no lower bound, then inf(S) = −∞

The fundamental axiom of analysis states that if an increasing sequence of reals is bounded above, then it must converge.

Series

A series is a sum of a sequence - we consider an infinite series as a summation in the form S =

∑∞n=1an.

We can also take a partial sum, where we sum from 1 to n instead of ∞ as Sn =

∑ni=1ai. Since we’re dealing with finite numbers at this point, we can consider Sn as a sequence - thus whether a series converges, or diverges, depends on whether the sequence of partial sums converge, or diverge. If we know that ∀i ≥ 1[ai > 0], then the partial sum is an increasing sequence where S1 < S2 < S3 < ..., if this is bounded above, then it must converge (from the axiom of analysis).

For a series to have a chance of converging, insufficient for concluding a series converges. we It’s require important n→∞lim ato n note = 0. However, just proving that is that we don’t really care about the first values of a series, only how it behaves on large numbers, as it tends to infinity.

If S converges, or diverges, then

∑∞n=N an also converges, or diverges, as it’s equivalent to S − SN−1.

Geometric Series

To determine whether a limit G exists for the geometric series, we need to determine that the partial sums converge. Hence we do the proof on the right hand side. It’s important to note that the only term that changes based on n is xn+1, therefore we can conclude that G must converge when |x| < 1 (from earlier results).

Let us represent the geometric series as G =

∑∞n=1xn. Should a limit G exist - see the left side;

G =

∑∞xn

Gn =

∑nxi

n=1

= x +

∑∞xn

i=1

= x +

∑nxi

n=2

i=2

= x + x

∑∞xn

= x + x

∑n−1xi

n=1

i=1 = x + xG

= x + x(Gn − xn) = 1 − x

x

= x 1 − − xn+1

x

Harmonic Series

The harmonic series is often used in comparison tests to prove that another series diverges.

Let harmonic series be written as S =

∑∞n=1

1n = 1+ 12 +(} 13 {{ + 14} ) > 14 + 14

+(} 15 + 16 {{ + 17 + 18} ) > 18 + 18+ 18+ 18

+... > 1+ 12 + 12 + 12 +...

Therefore, with partial sums, we get Sn =

∑2ni=1

1i > 1 + n2. As Sn clearly diverges, S must also diverge.

3

Inverse Squares

Suppose we have the sum of squares S =

∑∞n=1

n12, and an auxiliary partial sum Tn =

i=1 ∑n1

i(i+1).

To should inverse prove be squares that spotted). S are converges, bounded. Hence we n→∞lim It’s should Tn trivial = 1. consider Using to say this, Tthe n = we following 1− can n+1 1(write out the first argue how the partial inequality; 1

i(i+1) few terms and a pattern sums of the series of < 1i2 < i(i−1)1

. If we sum this inequality from i = 2 to n, we can then justify the following - it’s important to note that the sums on the left, and right, differ by only the first, and last term (therefore we can reduce it to a partial sum);

12(3) + 3(4) 1+ 4(5) 1+ ... + n(n 1

+ 1) <

∑i=2 n1i2 < 1(2) 1+ 2(3) 1+ 3(4) 1+ ... + n(n 1

− 1)

⇔ Tn − 12 <

∑i=2 n1i2 < Tn−1

⇔ Tn + 12 <

∑i=1 n1i2 < Tn−1 + 1 ⇔ 32 + n + 1

1 < Sn < 2 + n 1Therefore, suggests it follows that the partial sums the sequence converges, therefore are bounded the series must by also 32, and 2. converge.

The existence of an upper bound

Important Series

These examples will be useful for the comparison tests covered later on. We can use these results to aid us in proving whether some other series converges, or diverges.

name harmonic series

S ∑∞condition diverges or converges

n=1 1harmonic primes ∑

n diverges

p:prime

1p diverges

geometric series

∑∞n=1xn |x| ≥ 1 diverges geometric series

∑∞n=1xn |x| < 1 converges inverse squares series

∑∞n=1n2 converges ???

∑∞n=1

1nc c > 1 converges

Convergence Tests

For brevity, suppose we have the following;

• S =

∑∞i=1ai a series we want to reason about, and ai ≥ 0

•

∑∞i=1ci a series we have already established converges to c

•

∑∞i=1di a series we have already established diverges

4

In the comparison test, we have some λ > 0, and a N ∈ N. If ∀i>N[ai ≤ λci], then S converges. On the other hand, if ∀i>N[ai ≥ λdi], then S diverges. To avoid having to find such a λ, we can use the limit comparison test, in which we can say if i→∞lim aiexists, then S converges, but if i→∞lim ci diai exists, then S diverges. Another useful test for finding convergence, without needing a pre-established series, is D’Alembert’s ratio test. Once again, we consider this from some point N ∈ N, where the cases are;

1. ∀i ≥ N[ai+1

ai ≥ 1], then S diverges 2. there exists a k such that ∀i ≥ N[ai+1

ai ≤ k < 1], then S converges This is similar to the comparison test for sequences, and once again the requirement of finding some k between the ratio, and 1 can be avoided by taking limits. This leads to D’Alembert’s limit ratio test, in which we have the following cases;

1. i→∞lim ai+1

ai > 1, then S diverges 2. i→∞lim ai+1

ai = 1, then it is inconclusive 3. i→∞lim ai+1

ai < 1, then S converges The second case, in which it is inconclusive, can be illustrated by setting an = 1, which obviously causes S to diverge, or by setting an = test of 1.

1n2 which causes S to converge. However, both lead to a limit

The integral test for a sequence, where an = f(n) is a decreasing function relies on the following idea;

∑∞n=1an+1 < ∫ 1 ∞f(x)dx <

n=1∑∞an

In practice, if the integral were to diverge once evaluated, the series would also diverge by the right hand inequality, but if it were to converge, then the series would converge with the left hand inequality.

Absolute Convergence

Consider a series with negative an values, and another series S =

∑∞n=1|an|. If S converges, so does S. However, the same isn’t true the converges to ln(2), however, doing other way around, for this with the absolute example, when we have series, S , it would diverge, an = (−1)as if it were n−1 n1, the S

normal harmonic series. In order to test for absolute convergence - we have the following cases;

1. if n→∞lim |an+1

an | > 1 then

∑∞n=1an diverges

2. if n→∞lim |an+1

an | = 1 then

∑∞n=1an can converge, or diverge

3. if n→∞lim |an+1

an | < 1 then

∑∞n=1an converges absolutely, and therefore also converges

Power Series, and Radius of Convergence

We can represent a function as a power series; f(x) =

∑∞n=0anxn.

Consider the case where f(x) = 1−xx, which has the expansion

∑∞n=1xn, which only converges for |x| < 1.

5

∞∑n=1Take the power series S =

n2xn, and apply the D’Alembert ratio test, we’d get the following;

n→∞lim |aan+1

n | = n→∞lim |(n + n1)2x2n xn+1 = n→∞lim |x(1 + n1)2|

|

= |x|

As we want to establish convergence, we set the result |x| < 1, therefore it follows that the ratio of convergence for this particular power series is 1.

Power Series

As previously mentioned, a function can be represented in a series expansion on x; f(x) =

∑∞n=0anxn.

Maclaurin Series

Let some function f(x) = a0 +a1x+a2x2+a3x3+..., then we can easily work out a0, by setting x = 0, such that f(0) = a0. By differentiating the function, we get f (x) = a1 +2a2x+3a3x2 +..., therefore it follows that f (0) = a1. By differentiating again,we get f (x)=2a2 + 2 · 3a3x+ ..., therefore it follows that f (0)

2! = a2. Hence, generally an = f(n)n! (0)

, where n ≥ 0.

A Maclaurin series is the representation of f(x) =

∑∞i=0aixi.

This function can then be differentiated n times to get f(n)(x) =

∑∞i=nai(i − 1)...(i − n + 1)xi−n.

From this, we can see that f(n)(0) = n!an, hence f(x) =

∑∞n=0fn(0)xn! n. For example, if we take f(x) = ln(1 + x), and differentiate repeatedly;

f(x) = ln(1 + x) f(0) = 0 f (x) = (1 + x)−1 f (0) = 1 f (x) = (−1)(1 + x)−2 f (0) = −1 f (x) = (−2)(−1)(1 + x)−3 f (0) = 2 f (n)(x) = (−1)n−1(n−1)!

(1+x)n f(n)(0) = (−1)n−1(n − 1)! an = (−1)n

n−1

Therefore, we can justify that f(x) = ln(1 + x) = get −∞, and the power series would become the harmonic x1 − x2 2series, + x33 − which x4 4+ .... By setting x = −1, we’d we know to diverge. However, by setting x = 1, we get ln(2), is the convergence of the alternating harmonic series (as previously mentioned, in the ratio test example).

Taylor Series

When we’re interested in values around a certain point, we can use a more generalised version of a Maclaurin series. The general result, where we apply the series around a is;

f(x) = f(a) + f(1)1! (a)

(x − a)1 + f(2)2! (a)

(x − a)2 + ... =

n=0 ∑∞f(n)n! (a)

(x − a)n

For example, if we take f(x) = ln(x) around 2, and differentiate repeatedly;

f(x) = ln(x) f(2) = ln(2)

6

f (x) = 1x f (2) = 12 f (x) = (−1) x12 f (2) = −14 f (x) = (−2)(−1) x13 f (2) = 14 f (n)(x) = (−1)n−1(n−1)!

xn f(n)(2) = (−1)n−1(n−1)!

2n an = (−1)n−1

n2n

For example, we can approximate ln(x) around 2 as; ln(x) = ln(2) +

∑∞n=1

(−1)n−1

n2n (x − 2)2. It’s important to note the bounds for this, as we start from n = 1, and manually state the constant term, since that doesn’t fall within the ”pattern” we had. We use the same technique to calculate the radius of convergence, as follows;

n→∞lim |an+1

an | = (x−2)n+1 = n→∞lim n→∞ lim |

(n+1)2nn+1 n+1 (x−2)n2· n |x−2|

n

2

|

= |x−2|

2

And then, by radius is limited setting by the |x−2|

closest 2 < 1,we have a radius of convergence of 2, around x = 2. Generally, the singularity (where a point is not defined, or isn’t well-behaved).

Taylor Series Error Term

In order to ensure termination, instead of summing to ∞, we often truncate at some arbitrary kth term.

f(x) =

∑∞n=0

f(n)n! (a)

(x − a)n =

n=0 ∑kf(n)n! (a)

(x − a)n + } f(k+1)! k+1(c)

(x {{ − a)k+1 } Lagrange error term

The Lagrange error term replaces the tail of the infinite series, from, which would otherwise be a sum from k + 1 to ∞. The constant c lies between x, and a. By taking the bound a<c<x, or x<c<a, it can be used to generate the worst-case error.

The mean value theorem states that there is some therefore we can get f(x) = f(a)+(x − a)f (c).

point c between x, and a, such that f (c) = f(x)−f(a)

x−a ,

Cauchy Error Term

Define the set of partial sums of the Taylor series as Fk(t) =

∑kn=0

f(n)n! (t)

(x − t)n, where t is the offset.

It follows that Fk(x) = f(x) for all k (justified as every term becomes 0, other than the first term, n after = 0, k terms;

where it’s f(x)

0! ). Therefore the error term is derived as follows, where Rk(x) is the remainder

Fk(x) − Fk(a) = f(x) −

∑kn=0

f(n)n! (a)

(x − a)n = Rk(x)

Power Series Solutions of ODE

When series for we y. have By a differentiating differential equation term by in term, the form we can dx dy= compare ky, for a coefficients;

constant k, we can consider the power

dydx ≡ dx

d∑∞i=0aixi ≡

∑∞i=1aiixi−1 ≡ ky ≡ k

∑∞i=0aixi ≡ k

∑∞i=1ai−1xi−1 ≡

∑∞i=1kai−1xi−1

7

By matching coefficients of xi−1, where i ≥ 1, we have kai−1 = aii, therefore relation ai = kiai−1. By ”expanding” this, we end up with ai = kiai−1 = ki · k

i−1we end up with ai−2 = ... = ki! ia0.

the

If we were to have some boundary condition, such as when x = 0, y = 1, we know that y(0) = a0, hence a0 = 1. Therefore, this can be solved to the following;

y =

∑∞i=0

(kx)i! i

= ekx

Linear Algebra

Linear Equation Systems

For brevity, instead of writing; ax1 + bx2 + cx3 = j dx1 + ex2 + fx3 = k gx1 + hx2 + ix3 = l

(1) (2) (3)

We will write A x = y, where A =

a b c d e f

g h i, x =

xx12x3, and y =

jkl In general, for a real linear equation system, we have the following possible outcomes;

• no solution

A =

1 1 1 1 −1 2

2 0 3, and y =

32 1By adding equations (1), and (2), we get 2x1+3x3 = 5, which contradicts equation (3), hence it is inconsistent, and therefore has no solutions.

• unique solution

A =

1 1 1 1 −1 2

0 1 1, and y =

32 2By doing equation (1) - equation (3), we get x1 = 1. Adding equations (1), and (2), we end up with 2x1+3x3 = 5, however, we know the value of x1, therefore it follows that x3 = 1. Finally, by using this value in equation (3), we have x2 = 1, therefore the only solution is;

x =

11 1• infinite solutions

A =

1 1 1 1 −1 2

2 0 3, and y =

32 5Note that this is the same A from the first example. However, this time adding equations (1), and (2), is consistent with equation (3). If we let x3 = λ, such that x3 is a free variable, with λ ∈ R, we can get both x1, and x2 in terms of λ. By doing equations (1) - (2), we get 2x2 − x3 = 1, hence x2 leads to x1 = 52 − 32λ. As = such, 12 + 12we λ. have By doing an infinite equations set of (1) solutions, + (2), we get 2x1 + 3x3 = 5, which

given λ ∈ R;

x =

53102 + λ−11

232

8

Groups

With a set G, and some operation ⊗ : G → G, the criteria for (G,⊗) being a group are (1 - 4), and (1 - 5) for an Abelian group;

1. closure ∀x, y ∈ G[x ⊗ y ∈ G] 2. associativity ∀x,y,z ∈ G[(x ⊗ y) ⊗ z = x ⊗ (y ⊗ z)] 3. neutral element ∃e ∈ G[∀x ∈ G[x ⊗ e = x ∧ e ⊗ x = x]] 4. inverse element ∀x ∈ G[∃y ∈ G[x ⊗ y = e ∧ y ⊗ x = e]] often written as y = x−1, note that this does not mean y = x1, defined with respect to ⊗ 5. commutative ∀x, y ∈ G[x ⊗ y = y ⊗ x]

For example, we can say (Z,+) is a group, but (N,+) is not, since the latter doesn’t have an inverse element; for example −5 /∈ N, so 5 doesn’t have an inverse. While (R\{0},·) is an Abelian group, (R,·) isn’t a group as 0 doesn’t have an inverse element (¬∃x ∈ R[0 · x = 1]). (Rn,+) is an Abelian groups under addition, as long as addition is defined component-wise.

Matrices

We can defined the set of all real-valued m × n matrices (m, n ∈ N), where m is the number of rows, and n is the number of columns, as Rm×n. We can refer to each element as ai,j, where i ∈ [1,m], and j ∈ [1,n], and ai,j ∈ R; thus giving us the following rectangular matrix;

A =

a1,1 a1,2 ··· a1,n a2,1 ... a2,2 ... ··· ... a2,n ... am,1 am,2 ··· am,n

A R1×n matrix is referred to as a row (vector), and a Rm×1 matrix is referred to as a column (vector), with component-wise addition, we are able to say (Rm×n,+) is a group.

Multiplication on matrices, A ∈ Rm×n,B ∈ Rn×k,C ∈ Rm×k, where C = AB, is defined as;

ci,j =

∑nl=1ai,lbl,j, with i ∈ [1,m], and j ∈ [1,k]

In general, matrix multiplication is not commutative, hence AB = BA. However, it is associative, and distributive (given that the dimensions match).

The identity matrix, written as In ∈ Rn×n, is all 0s, other than ai,i = 1, where i ∈ [1,n]. The identity matrix has the following property; ImA = A = AIn, where A ∈ Rm×n. If a square matrix (A ∈ Rn×n) is non-singular, we can find an inverse for it, A−1 ∈ Rn×n, such that AA−1 = A−1A = In. However, not every matrix has an inverse. We also have the result (AB)−1 = B−1A−1.

We can define the transpose of A ∈ Rn×m as B = A ∈ Rm×n, where bj,i = ai,j. On the rectangular matrix representation, we are essentially mirroring it on the leading diagonal. A square matrix is symmetric ⇔ A = A . The general results are (A ) = A, (A + B) = A + B , and (AB) = B A . Also, if A is non-singular, then (A−1) = (A )−1.

A multiplication by scalar λ ∈ R, means that for A ∈ Rm×n, and B = λA, it follows that bi,j = λai,j. This operation is distributive, and also associative.9

Solving Linear Equation Systems

Given the general form of an equation system as;

a1,1x1 + a1,2x1 + ··· + a1,nxn = b1 a2,1x... 1 + a2,2x... 2 + ··· + a2,nx... n = b... 2 am,1x1 + am,2x2 + ··· + am,nxn = bm

We can easily represent it as A x = b, where A ∈ Rm×n, x ∈ Rn(×1), and b ∈ Rm(×1).

For example, consider a simple linear equation system as follows;

A x =

[1 0 8 −4 0 1 2 12]x1x2x3x4 =

[428

]

Let us also say A = [ c1 c2 c3 c4], where ci represents the ith column of A, hence b =

∑4i=1xi ci. In order to find a particular solution, we can easily create that b = 42 c1 + 8 c2 + 0 c3 + 0 c4, therefore we can get a b as a particular linear solution; combination x = [42 of 8 c1, 0 and 0] c2, .

such

This doesn’t capture all the solutions of this equation system - if there are more unknowns than equations, then there are an infinite number of solutions. Recall that b + 0 = b, hence we can attempt to find non-trivial 0s to generate the set of all solutions.

Using the fact that A is already in row echelon form, we can try say that c3 = 8 c1 + 2 c2, therefore, it [8 follows 2 hence x −1 = that [−4 0] 0=8 c3 + 2 c2 − c3 + 0 c4, as such we’ve managed to create a zero solution when x = . We can 12 0 −1do ] the is same also with c4, such that c4 = −4 c1+12 c2 ⇔ 0 = −4 c1+12 c2+0 c3− c4, a zero solution. With these two 0s, we can say that the general solution is as follows (where λ1,λ2 ∈ R); x =

4280 + λ1

−4120  −1In this solution, we first find a particular solution for A x = b, and then combine this with all solutions to A x = 0. This was fairly trivial to solve, as the matrix was already in a special form. Generally, equation systems won’t be given to us like this, however we can transform it into this form by Gaussian elimination.

Elementary Transformations

We are allowed to apply the following operations, which let us transform the equation system into a simpler state, but preserve the solution set;

• swap two equations (which means we can swap rows in the matrix)

• multiply an equation (matrix row) by a non-zero constant λ ∈ R\{0}

• add (or subtract) an equation (row) to / from another equation (row)

We can say that a matrix is in row echelon form (REF), if all zero rows are below the non-zero rows, and the pivot of each row is strictly to the right of the pivot in the row above. The pivot of a row is defined as the first non-zero element in a given row, reading left-to-right.

Consider the following system of equations, and the result of it after Gaussian elimination;

10

 + λ2 0

82−10

−2x1 + 4x2 − 2x3 − x4 + 4x5 = −3 4x1 − 8x2 + 3x3 − 3x4 + x5 = 2 x1 − 2x2 + x3 − x4 + x5 = 0 x1 − 2x2 + 0x3 − 3x4 + 4x5 = a

x1 − 2x2 + x3 − x4 + x5 = 0

x3 − x4 + 3x5 = −2

x4 − 2x5 = 1

0 = a + 1

We can now work on the augmented matrix to transform it into REF;

−2 4 1 4 −8 −2 −2 −1 4 −3 3 −3 1 2 1 −1 1 0 1 −2 0 −3 4 a

−4R+2R1 1

1 0 0 −2 0 0 1 −1 0 −1 1 1 −3 −3 6 0 2 −3 −R1 0 0 −1 −2 3 a

↔ R3 ↔ R1



1 −2 1 −1 1 0 4 −8 3 −3 1 2 −2 4 −2 −1 4 −3 1 0 0 1 −2 0 0 −2 −1 1 0 0 −3 4 a −1 1 0 1 −3 2 −3 6 −3 1 0 0 −2 0 0 1 −1 0 −1 1 1 −3 −3 6 0

−R2 0 0 0 −3 6 a − 2−R3 2 −3 0 0 0 0 0 a + 1 · · − − 1 13

1 0 0 −2 0 0 1 −1 1 1 −1 3 0 1 −2 0 −2 1 0 0 0 0 0 a + 1

It follows that there is only a solution when a = −1. To clarify some definitions, we refer to a variable as being basic, if it corresponds to the pivot element of any given row, hence in this case, the basic variables are x1,x3,x4, and the rest of the variables are free (x2, and x5 in our case). The pivot columns in our case (written as pi corresponding to xi) are;

p1 =

100, p3 =

110, and p4 =

−1−11 000

We implicitly set the free variables to 0 when we find our particular solution using this method.

x1 p1 + x3 p3 + x4 p4 = b ⇔ x1



We can get x4 = 1 straight away, and from that we can get x3 = −1, and also x1 = 2. By using the same method as before, where we try create the non-pivot columns as a linear combination of the pivot columns, we can get the following general solution (∀λ1,λ2 ∈ R) - note that we could just as easily negate all the values in the column vectors, since our λ1,λ2 can take any real value;

x =

100 + x3

110 + x4

−1−11

 =



 ⇔ x = 000

0−210

20−110

 −2−20120−11 + λ1  −111 −2−100 + λ2 0

0

Minus-1 Trick

We say that some equation system is in reduced row echelon form (RREF) (this may also be referred to as row-reduced echelon form, or row canonical form in other texts), if it is already in row echelon form, every pivot item is 1, and the pivot is the only non-zero entry in the column. Consider the [0 following ··· 0 matrix, −1 0 ··· with 0], the to pivots coloured purple, and the matrix after adding rows of the form create a square matrix. By reading the columns containing the new -1s, we can get all solutions for A x = 0, once again ∀λ1,λ2 ∈ R;

A =

1 0 0 0 3 −1 0 0 0 0 3 0 0 0 1 0 9 0 1 −4 0 0 0 0 −1 x = λ1 

 + λ2  309−4−1

Gaussian Elimination to Calculate the Inverse

This is a fairly simple technique for calculating the inverse matrix, where we essentially augment the original side; [A matrix In] with ··· 1 3 0 0 3

3−10 0 1 0 9

 A = 00 0 0 1 −400

the [Iidentity, n A−1].

and apply Gaussian elimination to get the identity on the left hand

This is one of the parts in this module that are fairly easy to understand, but just requires practice for the exams. This technique is applied to a simple A ∈ R2×2;

[1 2 1 0

][1 2 1 0 3 4 0 1−3R1

0 −2 −3 1]· − 12

[1 2 1 0

0 1 32 −12]−2R2 [1 0 −2 1

0 1 32 −12]

Vector Spaces

We define a real-valued vector space (R-vector space) is a set V with two operations;

• + : V × V → V inner operation

• · : R × V → V outer operation

The following properties hold;

• (V,+) is an Abelian group

• the outer operation is distributive

λ · ( x + y) = λ · x + λ · y ∀λ ∈ R, x, y ∈ V (λ + ψ) · x = λ · x + ψ · x ∀λ, ψ ∈ R, x ∈ V

• the outer operation is associative

λ · (ψ · x)=(λ · ψ) · x ∀λ, ψ ∈ R, x ∈ V

• the outer operation has a neutral element of 1

1 · x = x ∀ x ∈ V

Examples of vector spaces are V = Rn, V = Rm×n, and V = C (the standard representation of the complex numbers).

Vector Subspaces

Suppose we have U ⊂ V , where V is a vector space, then U is a vector / linear subspace if it is a vector space with the same outer, and inner operations, but restricted to U × U, and R × U, respectively.

For us to show that U is a vector subspace, we must show the following properties;

• U = ∅, especially have to show 0 ∈ U

12

• closure of U

outer operation: λ x ∈ U ∀λ ∈ R, x ∈ U inner operation: x + y ∈ U ∀ x, y ∈ U

We have trivial vector subspaces of the vector space V , namely V itself, and { 0} (the subspace con- taining only the zero vector). The intersection of an arbitrary number of subspaces is also a subspace, and the intersection of all subspaces of V is referred to as the linear hull of V .

The solution space for a homogeneous linear equation system A x = 0, with n unknowns, is a subspace of Rn. Note that a homogeneous equation means that it has the zero vector on the right hand side. The same also applies in reverse; such that for a vector subspace U ⊂ V , there is a homogeneous system which characterises U. In contrast, an inhomogeneous linear equation system A x = b, where b = 0 does not have a vector subspace for its solutions. Assume that it does have some subspace U, therefore, 0 ∈ U by definition. However, 0 doesn’t satisfy the equation, since we know b = 0.

Linear Independence

We can define a vector space V , with x1, x2, ..., xk, as being linearly dependent if, and only if, there exists a non-trivial solution (such that there exists at least one λi = 0) in the following equation;

0 =

∑ki=1λi xi

By the closure of vector spaces, a set of linearly dependent vectors contains some redundancy.

We A = can [ x1 check x2 if vectors ··· xk]. x1, Once x2, ..., put xk into ∈ V row are echelon linearly form, independent if all by constructing an augmented matrix; the columns are pivot columns, then it follows that it is linearly independent, otherwise the non-pivot columns can be made from the pivot columns to the left of it (therefore it is linearly dependent).

Consider the following vectors in R4; x1 = 

, x2 =

110, and x3 = 2−1−211 

We essentially want to solve λ1 x1 + λ2 x2 + λ3 x3 = 0, which means we can set this up as follows;



12−34

1 1 −1 2 1 −2 −3 0 1 0 0 0 −2R+3R1 1 4 2 1 0−4R1

1 0 0 1 −1 3 −1 0 −2 0 0 0 0 −2 5 0+3R2 −2R2 1 0 0 1 −1 1 0 0 1 0

0 0 1 1 −1 0 −1 0 0 0 −2 0 0 5 0 0 0 0 · · − − 1 12

0 0 5 0−5R3

1 0 0 1 −1 1 0 0 1 0 0 0 0 0 0 0

As every column here is a pivot column, such that every column is linearly independent of the ones to its left, it follows that there isn’t a non-trivial solution, as the only solution is λ1 = λ2 = λ3 = 0, therefore the vectors are linearly independent.

Linear Combinations (or Something?)

Suppose there is some vector space V , with k linearly independent vectors b1, b2, ..., bk, and m linear combinations, expressed as follows (with j ∈ [1,m]);

13

k∑i=1λi,j bi

We can then apply the same approach to test for linear independence;

0 =

xj =

∑mμj xj =

∑m∑mμjλi,j) bi j=1j=1j=1This means that x1, x2,..., xm only has linear independence if, and only if the following column vectors are linearly independent;

λ1 =

μj

∑ki=1λi,j bi =

∑mj=1

∑ki=1μjλi,j bi =

∑k( i=1λ1,m... λk,m ∈ Rk

Consider the following example, with a set of linearly independent vectors b1, b2, b3, b4 ∈ R3, and;

b1 − 2 b2 + b3 − b4 = x1 −4 b1 − 2 b2 + 0 b3 + 4 b4 = x2 2 b1 + 3 b2 − b3 − 3 b4 = x3 17 b1 − 10 b2 + 11 b3 + b4 = x4

Now, to check whether x1, x2, x3, x4 are linearly independent, we need to find the reduced row echelon form of the coefficient matrix; 

λ1,1 ...λk,1 ∈ Rk, ..., λm =

1 −4 0 1 0 4 2 17 − −3 10 7−−6

2410 −4R2 0 0 1 −181 −4 2 17 0 0 1 0 1 −4 2 17 −2 −2 3 −10 1 0 −1 11

− − 10 10 72−2410 1 0 0 −4 −10 4 2 17 7 24 −3 −6 0 0 −1 18+2R−R1

1

· − 110 −1 4 −3 1

+R1

· − 1 360 0 1 −1810 + 102R4 1 0 0 −4 1 0 2 − 0 10 7−17 0 2410 ↔ R4

1 0 0 −4 2 1 0 − 1 10 7−18 −17

2410  0 0 1 −18↔ R3

0 0 0 0

−2R3 + 107R3

1 −4 0 53 0 1 0 −15 0 0 1 −18 0 0 0 0

+4R2 1 0 0 0 1 0 0 0 1 −15

−18 −7

 0 0 0 0

Since we have a non-pivot column, it follows that it’s solvable non-trivially, and therefore it is linearly dependent. Therefore, the x vectors are linear dependent since x4 lies in the span of x1, x2, x3.

Basis

Given a finite number of vectors x1, x2, ..., xk ∈ V , we can say that any vector of the following form, is a linear combination of the vectors x1, x2, ..., xk (given λ1,λ2, ..., λk ∈ R);

v =

∑ki=1λi xi ∈ V

Suppose we have A ⊂ V , and A = { x1, x2, ..., xk}, if we can generate every vector v ∈ V , as a linear combination of vectors in A, then A is a generating set / span of the vector space V . This is denoted as V = [A]=[ x1, x2, ..., xk]. We can define a generating set of V as a basis of V , if is linearly independent, and minimal. We say a generating set of V is minimal if there isn’t a smaller set which spans V . With a real vector space

14

V , and B ⊂ V , B = ∅, B is a basis of V ⇔ B is a minimal generating set ⇔ B is a maximal linearly independent subset of V .

It is also equivalent to say that every v ∈ V is a unique linear combination of vectors from B;

v =

∑ki=1λi bi =

∑ki=1μi bi ⇒ ∀i ∈ [1,k][μi = λi]

While bases aren’t unique, the number of elements (or dim(V )) is constant for a vector space V . In R3, we refer to the standard basis as the following set of vectors, thus dim(R3) = 3, obviously;

1B = [0,001,000] 1With a vector subspace U ⊂ V , we can set an upper bound on the dimensions as dim(U) ≤ dim(V ), and dim(U) = dim(V ) ⇔ U = V .

Consider the following subspace U ⊂ R5, spanned by;

x1 =

 12−1−1, x2 = −1

2−112−2, x3 =  3−435, and x4 = −3−1−5−681 

As we’re once again trying to determine linear independence, we end up with another homogeneous equation system, that can be represented by taking the augmented matrix of the vectors, and then apply Gaussian elimination on it, down to reduced row echelon form (still doing all of it, for practice);

 1 2 −1 −1 2 −1 1 2 3 −4 3 5 −1 −5 −6 8

−2R1 +R1 +R1 +R1

1 0 0 0 2 −5 3 4 3 −10 6 8 −1 10 −6 −7 0 0 0 0

·13

1 0 0 0 2 −5 1 4 3 −10 2 8 −1 10 −2 −7 0 0 0 0

+5R3 −4R3 −1 1 0 0 0 −2 2 3 0 0 1 2 0 0 −3 −1 0 −2 1

1 0 0 0 0

→ R4 → R2 → R3

1 0 0 0 2 3 −1 1 2 −2 0 0 1 0 0 0 0 0 0 0



1 0 0 0 2 3 0 1 2 0 0 0 1 0 0 0 0 0 0 0−2R2 1 0 0 0 0 −1 1 2 0 0 0 0 0 0 1 0 0 0 0 0

From this, we can see that the vectors corresponding to the pivot columns are x1, x2, and x4, hence we can say that [ x1, x2, x4] forms a basis of U. On the other hand, we can try to find a simple basis of U, which means we want as many coordinates equal to 0 as possible. We can do this by writing the vectors as rows in the matrix, and then apply the same Gaussian elimination; 

+R3 +2R3

−2R−3R1 1 +R1

1 0 0 2 −5 −10 −1 −1 −1 3 4 0 6 8 0 0 10 −6 −7 0 1 2 −1 −1 −1 2 −1 1 2 −2 3 −4 3 5 −3 −1 8 −5 −6 1

+R3

15

1 0 0 1 0 0 2 −5 0 2 −1 −1 −1 −5 3 4 0 −10 0 0 0 0 0 0 1 0 −1 −1 3 4 6 8 −1 0 0

+2R2 0 0 0 1 0

 ↔ · − R15 4

1 0 0 2 −1 1 0 −0 35 −1 −1 −1 45 0 0 ↔ R3 0 0 0 0 0



1 0 0 2 −1 1 0 −0 35 0 −1 0 0 1 0 0 0 0 0 0

−2R2

1 0 0 0 1 0 1−0 5 35 0 −1 0 0 1 0 0 0 0 0 0

+R3 +45R3

 U = [

10105−1,

,00010]

By reading the rows with the leading 1s, we have a simple basis; U = [ b1, b2, b3]

Rank

The rank of a matrix is the number of linearly independent columns, denoted rk(A), where A ∈ Rm×n. We are also given that the column rank is the same as the row rank, hence rk(A) = rk(A ).

The columns of A span a subspace U ∈ Rm, with dim(U) = rk(A), and the rows span a subspace W ∈ Rn, with dim(W) = rk(A). A basis of U, and W, can be found by applying Gaussian elimination to A, and A respectively.

A ∈ Rn×n is invertible (regular) if, and only if, rk(A) = n. For all A ∈ Rm×n, and all b ∈ Rm, the equation system A x = b can only be solved if rk(A) = rk(A | b). The subspace of solutions for A x = 0 (kernel / nullspace) has dimension n − rk(A). A matrix is said to have full rank if rk(A) = min(m, n), otherwise it is said to be rank deficient.

For example, to calculate the rank of A, as below, we apply Gaussian elimination to row echelon form, and look at the number of pivot columns (which corresponds to the linearly independent columns);

A =

01−0350

1 2 1  −2 −3 1

3 5 0+2R2 −3R2

1 2 1 0 1 3

0 −1 −3+R2

1 2 1 0 1 3  0 0 0As we have two pivot columns, it follows that rk(A) = 2.

Subspace Intersection

Suppose we have two subspaces U1 = [ b1, b2, ..., bk] ⊂ V , and U2 = [ c1, c2, ..., cl] ⊂ V . We’re then interested in finding the basis of U1 ∩ U2, which is all the x ∈ V , such that x ∈ U1 ∧ x ∈ U2. For something to be in U1, or U2, it must be a linear combination of the basis vectors of the respective subspaces; hence we can get the following;

∑ki=1λi bi = x =

∑lj=1μj cj ⇔

∑ki=1λi bi −

∑lj=1μj cj = 0 This gives a homogeneous equation system, where we want to solve;

[ b1 b2 ··· bk − c1 − c2 ··· − cl]λ...1λkμ...1μl

= 0

16

However, we only need to determine λ1,λ2, ..., λk, or μ1,μ2, ..., μl, to find the intersection of U1, and U2. Consider the example, with the following vector subspaces;

U1 = [110,0011,0001] ⊂ R4, and also U2 = [−11210

,0100] ⊂ R4

By generating the matrix shown above, we can now apply Gaussian elimination to reduced row echelon form;1 1 0 0 0 1 1 0 −1 1 1 −2 0 −1 0 0 0 1 0 0

−R1

1 0 0 0 0 1 1 0 −2 1 1 −2 0 −1 0 0 0 1 0 0

−R2

1 0 0 0 0 1 1 0 −2 0 1 0 0 −1 1 0 0 1 0 0

−R4

1 0 0 0 0 1 1 0 −2 0 0 0 0 −1 1 0 0 1 0 0

↔ R4 ↔ R3

1 0 0 0 0 1 1 0 −2 0 1 0 0 −1 0 0 0 0 0 1



From this, we can determine that μ2 = 0, and that μ1 is a free variable, since it doesn’t have a corresponding pivot column. Therefore, it follows that x = μ1 c1, hence we’ve determined that the basis of the intersection is as follows;

U1 ∩ U2 = [−1120 ]

Linear Mappings

A linear mapping preserves the structure of vector spaces. Therefore, for two R-vector spaces V , W, and a mapping Φ (note that it can also be called a vector space homomorphism);

Φ : V → W is linear ⇔ ∀λ, μ ∈ R∀ x, y ∈ V [Φ(λ x + μ y) = λΦ( x) + μΦ( y)]

Special cases of this are;

• isomorphism Φ : V → W is linear, and bijective

finite dimensional R-vector spaces V , and W are isomorph ⇔ dim(V ) = dim(W)

• endomorphism Φ : V → V is linear

• automorphism Φ : V → V is linear, and bijective

• identity mapping idV : V → V such that x ↦→ x

For example, consider the mapping Φ : R2 → C, where Φ( x) = x1 + ix2. We can show that this is a linear mapping, by taking arbitrary x, y ∈ R2, and also arbitrary λ ∈ R;

Φ([xx12]

+

[y1y2]) = Φ([xx1 2 + + y1

y2]) = (x1 + y1) + i(x2 + y2) = = xΦ(1 [+ xxix12]2 ) + + yΦ(1 + [yyiy12]2 )

Φ(λ[x1x2]) = Φ([λxλx1

2]) = (λx1) + i(λx2) = = λ(xλΦ(1 [+ xx12ix])

2)

17

Image, and Kernel / Null Space

For the mapping Φ : V → W, we define the kernel / null space, and image / range respectively;

• ker(Φ) { v ∈ V | Φ( v) = 0W} = Φ−1({ 0W})

this is the set of vectors in V that maps to the neutral element 0W in W

• Im(Φ) { w ∈ W | ∃ v ∈ V [Φ( v) = w]} = Φ(V )

this is the set of vectors in W that can be reached by Φ from a vector in V

Note that Φ( 0V) = 0W, hence the unique 0V ∈ ker(Φ), therefore the null space = ∅. By the definitions above, we can easily see that Im(Φ) ⊂ W, and also ker(Φ) ⊂ V , hence they are subspaces of their respective vector spaces. A mapping, Φ, is injective ⇔ ker(Φ) = { 0}.

A matrix A ∈ Rm×n represents the mapping Φ : Rn → Rm, where Φ : x ↦→ A x. If we consider the column vectors of A, such that we have a1, a2, ..., an, we can apply the following;

Im(Φ) = {A x | x ∈ Rn} = {

∑ni=1xi ai | x1, ..., xn ∈ R} = [ a1, ..., an] ⊂ Rm

Which tells us that the image is the span of the columns of A.

Consider the following example, of a mapping represented by a matrix;

Φ : R4 ↦→ R2, where

xxx123x4 ↦→

[1 2 −1 0 ]1 0 0 1xxx123x4

By what we’ve stated above, we can simply get the image / range, by taking the column vectors;

Im(Φ) = [[11],[20],[−10

],[01]]

To compute the kernel of Φ, we solve A x = 0, which can be done fairly easily with Gaussian elimination into reduced row echelon form, and then applying the Minus-1 Trick;

[1 2 −1 0

]↔ R2 1 0 0 1↔ R1

[1 0 0 1

1 2 −1 0]−R1

[1 0 0 1

]0 2 −1 −1·12

[1 0 0 1

0 1 −12 −12]

Now that the matrix is in RREF, we can apply the Minus-1 Trick to compute the basis of the kernel;

ker(Φ) = [

−−1012,

−1012−1] 0

For a linear mapping between two vector spaces; Φ : V → W, it holds dim(ker(Φ)) + Im(ker(Φ)) = dim(V ). This is referred to as the Rank-Nullity Theorem.

The following properties also hold for R-vector spaces V,W,X, and the linear mappings Φ : V → W, Ψ : W → X, and Λ : V → W;

• Ψ ◦ Φ : V → X is also a linear mapping

• if Φ is an isomorphism (bijective), then it follows that the inverse, Φ−1 : W → V , is also an isomorphism

• Φ + Λ, and λΦ are also linear, for λ ∈ R

18

Matrix Representations

Any finite n-dimensional R-vector space V , is isomorphic to another n-dimensional R-vector space W, especially Rn, if the dimensions match, such that dim(V ) = dim(W). This allows us to use Rn without loss of generality. If we define an ordered basis [ b1, b2, ..., bn] of V , we can construct such an isomorphism.

Suppose there is some R-vector space V , with the ordered basis B, as denoted above. For x ∈ V , we can obtain a unique linear combination, same as before;

x =

∑ni=1αi bi

Therefore, we take the n-tuple given by αi, to be the coordinates of x with respect to B, and have it written as a vector. Consider R2, where the standard basis is [ e1, e2]=[[10],[01]], and x =

[2]

3with respect to [ e1, e2]. However, this same vector doesn’t have to be represented with respect to the standard basis. If we were to define another ordered basis;

B = [[ −11],[11]], then the same vector is

[−5122

]

with respect to the new basis

For some n-dimensional vector space V , with ordered basis B, the mapping Φ : Rn → V , with Φ( ei) = bi is linear, and also an isomorphism.

Suppose we have vector spaces V , and W with ordered bases B = [ b1, b2, ..., bn], and C = [ c1, c2,..., cm] respectively. Then we can consider the representation of each of the basis vectors in B with respect to C; for j ∈ [1,n];

Φ( bj) =

∑mi=1αi,j ci, where we have AΦ ∈ Rm×n (αi,j)

The coordinates of Φ( bj) with respect to the ordered basis C of W are the jth column of AΦ. The rank of the transformation matrix, is the same as the dimension of the image of the transformation, such that rk(AΦ) = dim(Im(Φ)). If we have ˆx as the coordinate vector of x ∈ V , which is defined with respect to B, and ˆy being the coordinate vector of y = Φ( x) ∈ W with respect to C, where Φ : V → W, then;

ˆy = AΦ ˆx

Thus, we’re able to map coordinates with respect to some ordered basis in V , to coordinates with respect to some other ordered basis in W, by using the transformation matrix we found.

Consider a homomorphism Φ : V → W, with ordered bases B = [ b1, b2, b3], and C = [ c1, c2, c3, c4] respectively.

Φ( b1) = c1 − c2 + 3 c3 − c4 Φ( b2)=2 c1 + c2 + 7 c3 + 2 c4 Φ( b3)=3 c2 + c3 + 4 c4

We can then represent that as the transformation matrix AΦ, which also gives us the coordinate vectors αj of Φ( bj), with respect to C.

AΦ = [ α1 α2 α3] =

 1 −1 3 2 1 7 0 3 1  −1 2 419

Suppose the coordinate vector ˆx representing x ∈ V with respect to B is;

ˆx =

xx12x3 = x1 b1 + x2 b2 + x3 b3

Therefore, the coordinate vector ˆy representing Φ( x) ∈ W with respect to C is;

ˆy = Φ( ˆx)

= x1Φ( b1) + x2Φ( b2) + x3Φ( b3) = x1( c1 − c2 + 3 c3 − c4) + x2(2 c1 + c2 + 7 c3 + 2 c4) + x3(3 c2 + c3 + 4 c4) = AΦ ˆx

The above can be done since we know that Φ is a linear map, and that ˆx is a real valued vector.

Basis Change

Continuing on from the previous spaces V , and W, and the mapping Φ : V → W, suppose we have another ordered basis of V , and W, respectively;

B  ̄= [ b ̄1, b ̄2, ..., b ̄n], and also C  ̄= [ ̄c1, ̄c2, ..., ̄cm] Like before, because we know that B, and C are both bases, we can construct anything in B,  ̄and C  ̄as a linear combination of vectors in their respective original basis;

b ̄j =

∑ni=1si,j bi, and also ̄ck =

∑ml=1tl,k cl, where j ∈ [1,n], and k ∈ [1,m]

Therefore, coordinates we with can respect define to regular B,  ̄matrices S ∈ Rn×n (si,j), and onto coordinates with respect to B, T ∈ Rm×m and similar for (tC  ̄l,k), to which C. maps Refer to 3Blue1Brown’s Change of basis | Essence of linear algebra, chapter 13, for a more intuitive perspective on this. Like before, the jth column of S, and T, are the coordinate representations of bˆj, and ˆcj with respect to B, and C, respectively. Suppose The elements we have of Asome  ̄Φ A ̄Φ, which represents the mapping Φ : V are represented with ̄α.

→ W, with bases B,  ̄and C  ̄respectively.

Φ( b ̄j) =

∑mk=1  ̄αk,j ̄ck the matrix which we’re trying to find =

∑mk=1  ̄αk,j

∑mi=1ti,k ci by our definition of ̄ck =

∑m( ∑mi=1k=1ti,k ̄αk,j) ci rearranging ∑

To solve for A ̄Φ, we consider how we represent b ̄j with respect to B, and how we represent each Φ( bk);

Φ( b ̄j) = Φ(

∑nk=1sk,j bk) by our definition of b ̄j =

∑nk=1sk,jΦ( bk) by linearity of Φ =

∑nk=1sk,j

∑mi=1αi,k ci by the mapping of Φ =

∑m( ∑ni=1k=1αi,ksk,j) ci rearranging ∑

20

Looking at the final lines of the above equations, it follows that;

Φ( b ̄j) =

∑m(

∑mti,k ̄αk,j) ci =

∑m(

∑nαi,ksk,j) ci ⇔

∑m∑nαi,ksk,j i=1k=1i=1k=1k=1k=1Therefore, it follows that we mapping composition; Φ C  ̄B  ̄mappings, and the respective = have transformation Φ CC  ̄T◦ A ̄ΦΦ CB = A◦ ΦΦS matrices B B  ̄⇔ = A ̄ΦΦ −1C in = C  ̄T−1AΦS. Note ◦ ΦCB ◦ ΦB B ̄. violet;

that we can represent this as The diagram below shows the

V W

B C

B  ̄C

̄ti,k ̄αk,j = vector spaces Φordered bases ΦB B  ̄S ΦCBAΦ

T−1 Φ CC ̄= Φ−1CC  ̄Φ C  ̄B  ̄We say that two T ∈ Rm×m, such regular matrix S matrices, that A  ̄∈ Rn×n, A, A  ̄∈ Rm×n are equivalent = such T−1that AS. A  ̄We = say they are S−1AS. if similar there exists if A, A  ̄regular ∈ matrices S ∈ Rn×n, and Rn×n, has a corresponding Therefore, it follows that similarity is a stronger property than equivalence, since all similar matrices are equivalent, but not vice versa.

With R-vector spaces V,W,X, mappings Φ : V → W,Ψ: W → X; we define the composed mapping Ψ ◦ Φ : V → X, to have the transformation matrix AΨ◦Φ = AΨAΦ

Determinants

Determinants are only defined for square matrices A ∈ Rn×n, and will be denoted as det(A), or |A|. There are also some properties about determinants of smaller matrices, but I’m skipping that.

A =

A ̄Φ

a0 1,1 ... a1,2 ··· a1,n a.2,2 .. ··· ... a2,n ... 0 ··· 0 an,n is an upper triangular matrix, therefore det(A) =

i=1∏nai,i

We can also apply the following properties of determinants, without proof;

• det(AB) = det(A) · det(B)

• det(A)=0 ⇔ A is singular (hence det(A) = 0 ⇔ if it is regular)

• det(A) = det(A )

• if A is regular; det(A−1) = det(A)−1

• similar matrices have the same determinant; therefore all AΦ of Φ : V → V have the same determinant

In order to get the determinant of a matrix by using the triangular form, we can use Gaussian elimi- nation, with the following row operation rules;

• adding a multiple of a column/row to another doesn’t change det(A)

• multiplying a column/row by λ ∈ R scales det(A) by λ, hence det(λA) = λndet(A)

• swapping two columns/rows changes the sign of det(A)

For example, consider the following sequence of row operations;

21

∣∣∣∣∣∣∣∣∣∣2 0 0 0 0 0 −1 1 0 0 1 −1 2 3 −1 2 0 −1 1 1 2 1 2 −1 1∣∣∣∣∣∣∣∣∣∣2 2 0 +R1 −2 2 0 −1 1 0 0 1 2 0 1 2 1 2 −1 0 1 ∣∣∣∣∣∣∣∣∣∣

2 0 1 2 1∣∣∣∣∣∣∣∣∣∣−R1 +R1 −R1

∣∣∣∣∣∣∣∣∣∣2 0 0 0 0 0 −1 0 0 0 1 2 0 −1 −1 1 1 0 3 3 1 2 ∣∣∣∣∣∣∣∣∣∣2 0 0 0 0 −1 −1 10 −1 0 0 0 1 2 0

−1 −1 1 1 0 3 0 1 −7 0 −1 4

∣∣∣∣∣∣∣∣∣∣+R4

∣∣∣∣∣∣∣∣∣∣2 0 0 0 0 ∣∣∣∣∣∣∣∣∣∣−3R+R3 3 0 −1 0 0 0 1 −1 1 0 0 2 −1 0 1 0 0 1 3 −7 −3∣∣∣∣∣∣∣∣∣∣

= 2 · −1 · 1 · 1 · −3=6

Another method for calculating the determinant of a matrix is via Laplace Expansion. We define a matrix A ∈ Rn×n = (ai,j), and another matrix Ai,j ∈ R(n−1)×(n−1) as A without its ith row, nor its jth column.

• det(A) =

∑n(−1)k+jai,kdet(Ai,k) expansion along row i k=1• det(A) =

∑n(−1)k+jak,jdet(Ak,j) expansion along column j k=1The chequerboard pattern formed by the signs are as follows, for A ∈ Rn×n;

n is odd;

+ − − ... + ... ··· ··· ... + − ...

, or when n is even; + − ··· ++ − ... − + ... ··· ··· ... − + ...  − + ··· +It’s better to expand along rows/columns with as many 0s as possible, as that would reduce the amount of work that needs to be done.

Eigenvalues, and Eigenvectors

We define λ as an eigenvalue for an endomorphism Φ : V → V , ⇔ ∃ v ∈ V 0[ v = 0 ∧ Φ( v) = λ v]. The set of all eigenvectors corresponding to some λ in AΦ is defined as the eigenspace of AΦ, with respect to λ, and is written as Eλ. The set of all eigenvalues of AΦ is called the spectrum of AΦ. We can also say that λ is an eigenvalue of A, when rk(A − λIn) < n, hence det(A − λIn) = 0. From the latter definition, we can define the characteristic polynomial p(λ) det(A−λIn). By setting p(λ) = 0, the roots of our equations are the eigenvalues of A.

To determine the eigenspace of AΦ, with eigenvalue λ, we take ker(AΦ − λIn). Since we defined an eigenvalue to satisfy AΦ x = λ x ⇔ (AΦ − λIn) x = 0, so we’re simply solving a homogeneous system of equations. We can say that two similar matrices A, A  ̄∈ Rn×n, have the same characteristic polynomial;

det(A  ̄− λIn) = det(S−1AS − λIn) by definition of similar matrices = det(S−1AS − λS−1InS) S−1S = In = det(S−1(A − λIn)S) distributivity of matrix multiplication = det(S−1) · det(A − λIn) · det(S) property of determinants = det(S)−1 · det(A − λIn) · det(S) property of determinants = det(A − λIn)

There are some worked examples in the notes, but this is something that needs to be practised. The

22

concept itself isn’t too difficult, but being able to apply it quickly needs some practice.

Diagonalization

The Panopto recording for this lecture wasn’t very legible, so it is mostly based from the notes, and some intuition from 3Blue1Brown.

As an ”exercise”; consider taking powers of A =

[0 1 1

1], diagonalize it, and find the nth power. A ∈ Rn×n is diagonalizable when it is similar to a diagonal matrix. We say a matrix is diagonal if it is in the following form, such that anything not on the leading diagonal is 0;

c0 ... ... 1 0 ··· c2 0 ... ··· 0 ··· ... ... ... cn−1 0 0 ... 0 0 0 ··· 0 cn

When we have an endomorphism Φ of an n-dimensional R-vector space V , if Φ is diagonalizable, it follows that the transformation matrix AΦ is also diagonalizable, and there is a basis in V consisting of the eigenvectors (hence there must be n eigenvectors; the sum of the dimensions of the eigenspaces is n). Φ is only diagonalizable; if the following conditions hold;

• it’s characteristic polynomial is in the form p(λ)=(−1)n(λ − c1)r1 ···(λ − ck)rk; ri ∈ N, ci ∈ R

the characteristic equation decomposes into linear factors

• i ∈ [1,k] - dim(Im(Φ − ciidV )) = n − ri ⇔ dim(ker(Φ − ciidV )) = ri rank-nullity the dimension of the eigenspace times an eigenvalue appears in the characteristic Eci must correspond to the algebraic multiplicity (how many

polynomial) ri of the eigenvalue in p(λ) each eigenvalue ci will appear ri times on the diagonal

Suppose we have a diagonalizable matrix A ∈ Rn×n, and the ordered eigenbasis [ b1, ..., bn] of A, with A bi = ci bi, with i ∈ [1,n] then;

S =

c0 ... ... 1 0 c2 0 ... ··· ··· 0 ··· ... ... ... cn−1 0 0 ... 0 0 0 ··· 0 cn[ b1 ··· bn], and S−1AS = D = 

Therefore, D is the transformation matrix of x ↦→ A x, with respect to the ordered eigenbasis.

Consider the following example;

A =

3 2 −1 2 6 −2 0 0 2



23

To check if it’s diagonalizable, we first find the characteristic polynomial p(λ) = det(A − λI3);

p(λ) = det(A − λI3)

=

∣∣∣∣∣∣3 − 2 0 λ 2 −1 6 − λ −2

0 2 − λ∣∣∣∣∣∣ = (2 − λ)∣∣∣∣3 − 2 λ 6 − 2

λ∣∣∣∣ = (2 − λ)[(6 − λ)(3 − λ) − 4] = (2 − λ)(λ2 − 9λ + 14) = (2 − λ)(λ − 7)(λ − 2) = (−1)3(λ − 2)2(λ − 7)

The eigenspace E2 can be computed by doing Gaussian elimination on A − 2I3;

1 2 −1 2 4 −2 0 0 0

−2R1

1 2 −1 0 0 0 0 0 0



−2By the Minus-1 Trick, we get E2 = [10

,10] 1Likewise, the eigenspace E7 can be computed by doing another Gaussian elimination on A − 7I3;

−4 2 −1 2 −1 −2

0 0 −5+· − 12R15 1

−4 2 −1 0 0 0 0 0 1

 ↔ · − R14 3 ↔ R2

1 −12

10 0 1 4 0 0 0−14R2 1 0 −0 12 0 1  0 0 01By the Minus-1 Trick (and also multiplying to get rid of the fraction), we get E7 = [20] Looking at the characteristic polynomial, it’s clear that the algebraic multiplicities matches the dimen- sions of the eigenspaces, therefore we can combine the spectrum as an ordered eigenbasis;

S =

−2 1 1 1 0 2

, and S−1AS = D = 0 1 02 0 0 0 2 0

0 0 7 Putting a matrix into this form has a great benefit when computing powers; as we are able to say Ak = (SDS−1)k = SDkS−1.

Cayley-Hamilton Theorem

The Cayley-Hamilton Theorem states that the endomorphism Φ : V → V , with a transformation matrix AΦ, on n-dimensional R-vector space V , and the characteristic polynomial p, then p(Φ) = 0 (the zero mapping), and p(AΦ) = 0 (the zero matrix). The proof for this is non-trivial (pretty much the only time you’ll hear that in this degree).

Consider the matrix A =

[1 −1 2 1

], which has the characteristic polynomial p(λ) = λ2−2λ+3 therefore;

A2 − 2A + 3I2 = 0 ⇔ − A2 + 2A = 3I2 ⇔ A13(−A + 2I2) = I2 ⇔ A−1 = 13(2I2 − A)

24

Inner Products

Inner products introduce rigour into intuitive geometric properties.

For example, the n-dimensional scalar product is defined for x, y ∈ Rn, as x y =

∑ni=1xiyi. However, generally, suppose we have a vector space V , and a bilinear mapping β : V × V → R (both arguments are linear);

• β is symmetric ∀ x, y ∈ V [β( x, y) = β( y, x)]

• β is positive definite ∀ x ∈ V [ x = 0 ⇒ β( x, x) > 0], and β( 0, 0) = 0

• if it is both, then it is defined as an inner product on V , and written as 〈 x, y〉

• (V,〈·,·〉) is called an inner product space, amongst other things - if it’s the dot / scalar product, then we call it a Euclidean vector space

Suppose we have some inner product space (V,〈·,·〉), then || x|| √〈 x, x〉 is the length / norm of x ∈ V . The mapping is || · || : V → R, such that || · || : x ↦→ || x||. When this is in an Euclidean vector space, the norm is the geometric length of the vector. The norm has the following properties;

• ∀ x ∈ V [|| x|| ≥ 0], and || x|| = 0 ⇔ x = 0

• ∀ x ∈ V ∀λ ∈ R[||λ x|| = |λ| · || x||]

• ∀ x, y ∈ V [|| x + y|| ≤ || x|| + || y||] Minkowski inequality

With the same inner product space, we define d( x, y) || x− y||, as the distance of x, y ∈ V , with the mapping d : V ×V → R, and d : ( x, y) ↦→ d( x, y). The mapping is called metric. A metric d is positive definite, symmetric, and satisfies the triangular inequality, such that d( x, z) ≤ d( x, y) + d( y, z).

We define two vectors x, y as orthogonal if 〈 x, y〉 = 0, and is often denoted as x ⊥ y.

With an inner product space (V,〈·,·〉), and x, y, z ∈ V , we can use the following properties;

1. |〈 x, y〉| ≤ || x|| · || y|| Cauchy-Schwarz inequality 2. || x + y|| ≤ || x|| + || y|| Minkowski inequality 3. d( x, z) ≤ d( x, y) + d( y, z) triangular inequality 4. || x + y|| + || x − y|| = 2|| x||2 + 2|| y||2 parallelogram law 5. 4〈 x, y〉 = || x + y||2 − || x − y||2 6. x ⊥ y ⇔ || x + y||2 = || x||2 + || y||2

By property 1, we can define θ ∈ [0,π), as long as x = 0, and y = 0;

−1 ≤ 〈 x, y〉

|| x||·|| y|| ≤ 1, and therefore cos(θ) = 〈 x, y〉

|| x||·|| y||

Orthogonal Projections

Suppose we have an R-vector space V , and W ⊂ V , we can define a linear mapping π : V → W as a projection if π2 = π ◦ π = π. As π is a linear map, we can define a projection matrix, Pπ, which has the and E1 = property Im(Pπ). Pπ To = derive P2π. This causes all eigenvalues of Pπ to be 0, or 1, and therefore E0 = ker(Pprojections in (Rn,〈·,·〉), we assume the scalar product 〈 x, y〉 = x y

π),

For the 1-dimensional case (projection to a line), let there be a subspace W ⊂ Rn, spanned by only b. When we project some n-dimensional vector x to W, we write p = πW( x). Geometrically, the projection point p is closest to x, such that we minimise || x − p|| - therefore p − x is orthogonal to W, and therefore to the basis of W. In one dimension, this is fairly easy to solve, as we simply have 〈 p − x, b〉 = ( p − x) b = 0.

25

However, we know that p must lie on the span of W, and with a single basis vector, we have p = λ b, where λ ∈ R. x − λ b ⊥ b

⇔ 〈 x − λ b, b〉 = 0

⇔ 〈 x, b〉 − λ〈 b, b〉 = 0

⇔ λ = 〈 x, b〉

〈 b, b〉 = 〈 x, b〉 || b||2

By using the dot product, we get p = || b b||x

2 b, however, knowing that p = Pπ x, we get Pπ = b b

|| b||2 However, we want to generalise this to more than one dimension. We’re still working with x ∈ Rn, but instead we’re mapping to a general subspace W ⊂ Rn, which is spanned by the ordered basis;

[ b1, b2, ..., bm], which can be expressed as a matrix B =

[ b1 ··· bm]

We can then express the projection in the following form;

πW( x) = p ∈ W ⇒ p =

∑mi=1λi bi = B λ

Since we want the minimal distance, the line between p ∈ W, and x ∈ Rn, must be orthogonal to all basis vectors; and we get the following homogeneous system of equations (assuming we’re in an Euclidean vector space);

∀i ∈ [1,m][〈 bi, x − B λ〉 = bi ( x − B λ) = 0]

 b1 ... b1

( x − A λ) = 0 ⇔ B ( x − A λ) = 0 ⇔ B B λ = B x

However, since we know that the rows / columns of B are linearly independent (since they were constructed from basis vectors), we can invert the matrix, such that λ = (B } B){{ −1B }

pseudo-inverse of B x. Since we’ve established p = B λ, it follows that πW( x) = p = B(B B)−1B x. We also get the projection matrix Pπ = B(B B)−1B .

Affine Subspaces

Suppose we have a vector space V , and x0 ∈ V , and a subspace U ⊂ V , then the following is an affine subspace;L = }{{}

x0 support point+ }{{}

Udirection (space)

{ x0 + u | u ∈ U} = { v | ∃ u ∈ U[ v = x0 + u]} ⊂ V

As we cannot get 0, unless x0 ∈ U, an affine subspace V L . ⊂ Suppose L  ̄⇔ U ⊂ we U  ̄∧( have x0 − affine  ̄x0) ∈ subspaces U.  ̄L = x0 + U, and generally L  ̄= ̄xisn’t 0 + U a  ̄(linear) of a In general, for a k-dimensional vector subspace vector space V , for then affine subspace L = x0 +U, where U has the ordered basis; [ b1, b2, ..., bk], then every x ∈ L can be uniquely described in a parametric form (which is also the parametric equation of L);

x = x0 + λ1 b1 + ... + λk bk = x0 +

∑ki=1λi bi

26

A one-dimensional affine subspace is called a line, a two-dimensional affine subspace is called a plane. In an n-dimensional vector space, an (n−1)-dimensional affine subspace is called a hyperplane, which divides the vector space in two. The solution space of an inhomogeneous linear equation system A x = b, where A ∈ Rm×n, and b ∈ Rm is either ∅, or an affine subspace of Rn, which has dimension n − rk(A).

Intersection of Affine Subspaces

The method for solving this is extremely similar to the one for solving linear subspaces (obviously), except we’re solving for the parameters in an inhomogeneous equation system. Let us define the two affine subspaces L1, and L2, and U1,U2 ⊂ V ;

L1 = x1 + U1, and U1 has ordered basis [ b1, b2, ..., bk] L2 = x2 + U2, and U2 has ordered basis [ c1, c2,..., cm] Now suppose we want to find the intersection, L1 ∩ L2. Let x ∈ L1 ∩ L2 ⇔ x ∈ L1 ∧ x ∈ L2, hence;

x = x1 +

∑ki=1λi bi = x2 +

∑kj=1μj cj ⇔

∑ki=1λi bi −

∑kj=1μj cj = x2 − x1 Which means we can solve for λ1, ..., λk,μ1, ..., μm;

⇔

[ b1 ··· bk − c1 ··· − cm]λ...1λkμ...1μm

= x2 − x1

For example, suppose we have the following affine subspaces;

L1 =

200 + [110,10011,0001], and also L2 =

13100 + [−1120

,0100] We then assume there exists some x ∈ L1 ∩ L2, and solve;

x =

200 + λ1 11100 + λ2

0110 + λ3

001 = 13100 + μ1

−1120

 + μ2

0100 We then solve the inhomogeneous equation system, for λ1,λ2,λ3,μ1,μ2. From the reduced row echelon form, we can try to make the solution from the pivot columns (and implicitly set the non-pivot column to 0), and apply the Minus-1 Trick for the zeroes of the solution; λ1λ2λ3μ1μ2 =



 + α

, for any α ∈ R

For simplicity, let us take just the μ solutions, hence μ1 = −α, and μ2 = 1. Therefore, we get the solution space for x;

x =

11−101

1−20−10

310 − α0−1120

 +

010 = 03200 + ψ

−1120

, for all ψ ∈ R. Therefore L1 ∩ L2 =

320 + [−11200

]

This solution is a line in R4. The full Gaussian elimination is done below;

27

1 1 0 0 0 1 1 0 −1 1 1 −2 0 −1 0 1 1 0 0 0 1 0 0 −1−R1

1 0 0 0 0 1 1 0 −2 1 1 −2 0 −1 0 1 0 0

−R2

1 0 0 0 0 1 1 0 −2 0 1 0 0 −1 1 1 0 0 0 0 1 0 0 −10 0 1 0 0 −1−R3 1 0 0 0 0 1 1 0 −2 0 1 0 0 −1 1 1 0 0 1 0 0 0 0 1 1 0 −2 0 1 0 0 −1 1 1

0 0 −R+R4 4 0 0 0 0 −1 −1· − 1 0 0 0 0 1 11 0 0 0 0 1 1 0 −2 0 1 0 0 1

0 1 0 −1 0 0 0 0 1 1



Parallel Affine Subspaces

We say two affine subspaces L1, and L2 are parallel (L1 L2), if either U1 ⊆ U2, or U2 ⊆ U1 (not sure if I should use the strict ordering here). If the subspaces don’t contain each other, then L1 ∩ L2 = ∅. Suppose we have lines g = x1 + U1, and h = x2 + U2, in R2, such that U1,U2 are 1-dimensional subspaces; if g ∩ h = ∅. If dim(U1 ∩ U2) = 0, then we have a single point of intersection, however if dim(U1 ∩ U2) = 1, then U1 = U2, so g = h. On the other hand, if we have g ∩ h = ∅, then dim(U1 ∩ U2) = 1, hence U1 = U2, therefore g h. We cannot have dim(U1 ∩ U2)=0in R2, since it would suggest no intersection for two non-parallel lines. Consider the same lines in higher dimensions Rn, where n ≥ 3. If g ∩h = ∅, then they either intersect in a single point, or they are the same line. However, if g ∩ h = ∅, the same case for dim(U1 ∩ U2)=1 applies, but dim(U1 ∩ U2) = 0 now has the definition that they are skew - such that there exists no plane containing the two lines.

The (actual) notes contain examples for hyper-planes in higher dimensions, and is worth a read. How- ever, all kinds of intersections are possible for hyper-planes in Rn, where n > 4.

Affine Mappings

Suppose we have two R-vector spaces V , and W, and a linear mapping Φ : V → W, and φ : x ↦→ a+ Φ( x), where a ∈ W. This is an affine mapping, and a is called a translation vector. It follows that φ = τ ◦ Φ, where τ : W → W, and is a translation such that τ : x ↦→ a + x.

Rotations

The start of this will be skipped, since this should’ve been covered mostly in A Level mathematics. The general rotation matrix in 2-dimensions for θ anticlockwise is;

R(θ) =

[cos(θ) −sin(θ) sin(θ) cos(θ)

]

The general convention we use is to stand on top of the vector, and look towards the origin (picture it or something, I don’t know).

In n dimensions, we can generalise it such that we have V , an n-dimensional Euclidean vector space, with Φ : V → V , as an automorphism. We can easily justify that a rotation is a linear mapping, as we’re doing matrix multiplication (which we can generalise to be a linear mapping). The idea in two dimensions is to fix n − 2 dimensions, and restrict such a rotation to a two-dimensional plane in n-dimensional spaces.

We can generalise the matrix for a rotation restricted to the i, j plane as; Ri,j(θ). This is the same as In, but ri,i = cos(θ), ri,j = −sincos(θ), rj,i = sin(θ), and ri,i = cos(θ). There is a special case in

28

3-dimensions, where we restrict it to the 1,3 plane (y-axis fixed), where r1,3 = sin(θ), r3,1 = −sin(θ) (we apparently rotate in the other direction, or something).

We can also say the following properties about rotations;

• we can compose rotations as the sum R(φ)R(θ) = R(φ + θ)

• rotations preserve length || x|| = ||R(θ) x||

• rotations preserve distances || x − y|| = ||R(θ) x − R(θ) y||

• rotations in Rn, where n ≥ 3, aren’t commutative (even around the same point) - they are only commutative in two dimensions

29